

ΠΑΝΕΠΙΣΤΗΜΙΟ ΚΡΗΤΗΣ
ΣΧΟΛΗ ΘΕΤΙΚΩΝ ΚΑΙ ΤΕΧΝΟΛΟΓΙΚΩΝ ΕΠΙΣΤΗΜΩΝ
ΤΜΗΜΑ ΒΙΟΛΟΓΙΑΣ
ΚΑΙ
ΙΝΣΤΙΤΟΥΤΟ ΜΟΡΙΑΚΗΣ ΒΙΟΛΟΓΙΑΣ ΚΑΙ ΒΙΟΤΕΧΝΟΛΟΓΙΑΣ

ΔΙΔΑΚΤΟΡΙΚΗ ΔΙΑΤΡΙΒΗ

*«Φυσικός χάρτης προς αλληλούχηση της περιοχής
q23.3-q25.3 του χρωμοσώματος 10 του ανθρώπου
και ανάλυση του γονιδίου FRA10AC1»*

ΘΕΟΛΟΓΙΑ ΣΑΡΑΦΙΔΟΥ

Εργαστήριο Μοριακής Γενετικής Ανθρώπου

Επιβλέπων Καθηγητής: Ν. Κ. Μοσχονάς

Ηράκλειο, Νοέμβριος 2005

UNIVERSITY OF CRETE
SCHOOL OF SCIENCE AND ENGINEERING
DEPARTMENT OF BIOLOGY
AND
INSTITUTE OF MOLECULAR BIOLOGY AND BIOTECHNOLOGY

Ph. D Thesis

***«Sequence-ready physical map of human
chromosomal region 10q23.3-q25.1 and analysis of
FRA10AC1 gene »***

Theologia Sarafidou

Human Molecular Genetics Lab

Advisor: Prof. N. K. Moschonas

Heraklion, November 2005

Στους γονείς μου Δημήτρη & Ελένη

Επιβλέπων Καθηγητής

N. K. Μοσχονάς (Καθηγητής)

Τριμελής Συμβουλευτική Επιτροπή

N. K. Μοσχονάς (Καθηγητής)

Γ. Μαυροθαλασσίτης (Αν. Καθηγητής)

Κ. Μαμαλάκη (Ερευνήτρια, Ινστ. Μοριακής Βιολογίας και Βιοτεχνολογίας, ΙΤΕ)

Επταμελής Εξεταστική Επιτροπή

Δ. Αλεξανδράκη (Αν. Καθηγήτρια, Τμήμα Βιολογίας, Παν/μιο Κρήτης)

N. Ανάγνου (Καθηγητής, Ιατρική Σχολή, Παν/μιο Αθηνών)

Δ. Καρδάσης (Αν. Καθηγητής, Τμήμα Ιατρικής, Παν/μιο Κρήτης)

Γ. Μαυροθαλασσίτης (Αν. Καθηγητής, Τμήμα Ιατρικής, Παν/μιο Κρήτης)

N. Μοσχονάς (Καθηγητής, Τμήμα Βιολογίας, Παν/μιο Κρήτης)

Ι. Παπαματθαϊάκης (Καθηγητής, Τμήμα Βιολογίας, Παν/μιο Κρήτης)

Γ. Χαλεπάκης (Αν. Καθηγητής, Τμήμα Βιολογίας, Παν/μιο Κρήτης)

Λίγες μέρες πριν από την παρουσίαση της διατριβής μου θέλω να ευχαριστήσω θερμά τους ανθρώπους που με βοήθησαν, με ποικίλους τρόπους και συνέβαλαν στην ολοκλήρωσή της.

Το πιο μεγάλο ευχαριστώ θέλω να το πω στον καθηγητή μου Νικόλαο Μοσχονά που με δέχτηκε στην ομάδα του, αρχικά στα πλαίσια του rotation και στη συνέχεια για το master και το διδακτορικό μου. Μπορώ να πω πολλά για τη βοήθεια και τη συμπαράστασή του αλλά μου φαίνονται λίγο τετριμμένα. Αυτό που νιώθω είναι περηφάνια που ήμουν μέλος της ομάδας του και που ολοκλήρωσα το διδακτορικό μου έχοντάς τον επιβλέποντα καθηγητή.

Ευχαριστώ ιδιαίτερα τους κ. Γιώργο Μαυροθαλασσίτη και κ. Κλειώ Μαμαλάκη, μέλη της τριμελούς μου επιτροπής, για το ενδιαφέρον τους για την πρόδοό μου. Επίσης, ευχαριστώ τα μέλη της επταμελούς μου επιτροπής: την κ. Δέσποινα Αλεξανδράκη που επιπλέον με βοήθησε στο πείραμα των δύο-υβριδίων, κ. Νικόλαο Ανάγνου που δέχτηκε να συμμετέχει στην εξεταστική επιτροπή, κ. Δημήτριο Καρδάση, κ. Κλειώ Μαμαλάκη, κ. Γιώργο Μαυροθαλασσίτη, κ. Ιωσήφ Παπαματθαϊάκη και κ. Γιώργο Χαλεπάκη που εκτός από μέλος της επιτροπής ήταν πάντα πρόθυμος να με βοηθήσει.

Ευχαριστώ θερμά τον Dr. Παναγιώτη Δελούκα από το Sanger Center για την φιλοξενία του στο εργαστήριό του και για την ευκαιρία που μου έδωσε να εργαστώ σε ένα πολύ ενδιαφέρον περιβάλλον. Επίσης ευχαριστώ την Lisa French για την άψογη συνεργασία μας.

Η Μαρία Κοκκινάκη είναι αυτή με την οποία πρωτοσυνεργάστηκα όταν πήγα στο εργαστήριο και γι' αυτό ήμουν πολύ τυχερή. Βρίσκω αξιοθαύμαστη την υπομονή και την διακριτικότητά της και την ευχαριστώ για την πετυχημένη πολύχρονη «συγκατοίκησή» μας. Ευχαριστώ επίσης την Αγγέλα Πασπαράκη, με την οποία συνεργαστήκαμε κατά την κατασκευή του φυσικού χάρτη, για την βοήθειά της στις κυτταροκαλλιέργειες, για τις άπειρες κρουτομές αλλά κυρίως για την συνέπεια και την ευγένειά της. Από τα παλαιότερα μέλη της ομάδας μας ήταν ο Αλέξανδρος Αργυροκαστρίτης. Θα τον θυμάμαι πάντα για την καλή του πρόθεση να βοηθήσει, την έμφυτη κομπόνητά του και το πηγαίο χιούμορ του.

Σημαντική ήταν η βοήθεια του Γιώργου Φραγκιαδάκη για την παραγωγή αντισωμάτων, που ήταν και ο «επιβλέπων» μου κατά την πτυχιακή μου εργασία, και του Λάμπρου Παναγή, συμφοιτητή μου, σε πολλές απορίες που είχα στην νευροανατομία του εγκεφάλου.

Σε αρκετά στάδια της διατριβής μου συνεργάστηκα, στα πλαίσια πτυχιακών εργασιών και rotations, με πολλά παιδιά, τα οποία και ευχαριστώ γιατί αυτές οι συνεργασίες ήταν εξίσου διδακτικές και για μένα. Θα θυμάμαι την άριστη συνεργασία μου με την Μαρία Παπαδάκη, τον Γιάννη Παπαευθυμίου και τον Δάκη Σκοπελίτη, κατά την φυσική χαρτογράφηση, την Δέσποινα Αποστολοπούλου και την Μαρκέλλα Κατίδου που συνέβαλαν στην παραγωγή αντισωμάτων και σε σειρά υποκλωνοποιήσεων και την Ευμορφία Τζαγκαράκη και τον Χρήστο Ανδρεάδη που συνεισέφεραν σημαντικά στα πειράματα των RNA *in situ* υβριδοποιήσεων.

Επίσης, και εκτός Βιολογίας, θέλω να ευχαριστήσω την Γιώτα για την φιλία της. Ένα μεγάλο ευχαριστώ στις αδελφές μου Μελίνα και Λήδα που με στηρίζουν πάντα και με κάνουν να εκτιμώ την μεγάλη οικογένεια και στον παππού και τη γιαγιά μου που έχουν τον ίδιο ενθουσιασμό με μένα. Τέλος, ευχαριστώ τους γονείς μου Δημήτρη και Ελένη για την απέραντη αγάπη τους και την υποστήριξή τους στις επιλογές μου.

1. ΕΙΣΑΓΩΓΗ

1.1. Γενικά	1
1.2. Ιστορική αναδρομή	4
1.3. Χαρτογράφηση	9
➤ Γενετική χαρτογράφηση	9
➤ Φυσική χαρτογράφηση	11
Κυτταρογενετική χαρτογράφηση	
Χάρτες κυτταρικών υβριδίων	
Χάρτες γενωμικών κλώνων	
1.4. Αλληλούχηση του ανθρώπινου γονιδιώματος	14
➤ Φυσικός χάρτης του ανθρώπινου γονιδιώματος	16
1.5. Αλληλούχηση γονιδιωμάτων οργανισμών-μοντέλων	20
1.6. Το χρωμόσωμα 10 του ανθρώπου	25
➤ Γονίδια του χρωμοσώματος 10 που σχετίζονται με νοσήματα	26
➤ Κυτταρογενετικά χαρακτηριστικά	26
➤ Αυτοσωμική επικρατής επιληψία του κροταφικού λοβού (ADLTE)	28
Οι επιληψίες	
Η ADLTE	
➤ Εύθραυστη χρωμοσωμική θέση FRA10A	30
Οι εύθραυστες χρωμοσωμικές θέσεις	
FRA10A	
1.7. Αντικείμενο της διατριβής	35

2. ΥΛΙΚΑ ΚΑΙ ΜΕΘΟΔΟΙ

2.1. Θρεπτικά	36
2.2. Βιβλιοθήκες DNA	37
2.3. Βάσεις δεδομένων και προγράμματα επεξεργασίας αλληλουχιών	38
2.4. Εκκινητές	39
2.5. Βακτηριακά στελέχη	39
2.6. Κυτταρικές σειρές θηλαστικών	40
2.7. Αντισώματα	41
2.8. Δείγματα DNA	41
2.9. Πλασμιδιακοί φορείς	41
2.10. Πλασμιδιακές κατασκευές	43
2.11. Αντιδραστήρια	47
2.12. Απομόνωση νουκλεϊκών οξέων	58
Απομόνωση πλασμιδιακού DNA από <i>E.coli</i> .	
Απομόνωση γενωμικού DNA από πλακούντα	
Απομόνωση ολικού RNA από ιστό	
2.13. Τεμαχισμός DNA με περιοριστικές ενδονουκλεάσες	60
2.14. Ηλεκτροφόρηση	60
Ηλεκτροφόρηση DNA σε πήκτωμα αγαρόζης	
Ηλεκτροφόρηση DNA σε πήκτωμα πολυακρυλαμίδης (PAGE)	
Ηλεκτροφόρηση πρωτεϊνών σε αποδιατακτικό πήκτωμα πολυακρυλαμίδης (SDS-PAGE)	
2.15. Αλληλούχηση DNA	62
2.16. Μετασχηματισμός βακτηρίων	63
Παρασκευή δεκτικών βακτηρίων και μετασχηματισμός με θερμικό σοκ	
Παρασκευή δεκτικών βακτηρίων και μετασχηματισμός με εφαρμογή ηλεκτρικού πεδίου	
2.17. Παρασκευή τυφλών άκρων	65
2.18. Αντίδραση σύνδεσης (ligation)	65
2.19. Ραδιοσήμανση του DNA	65
➤ Ραδιοσήμανση δίκλωνων μορίων	65
Με τη μέθοδο των τυχαίων εκκινητών	
Με PCR	

➤ Ραδιοσήμανση ολιγονουκλεοτιδίων	68
2.20. Εκτίμηση της ειδικής ενεργότητας ανιχνευτή με κατακρήμνιση με TCA	68
2.21. Υβριδοποίηση νουκλεϊκών οξέων	69
Υβριδοποίηση γενωμικών βιβλιοθηκών με STSs	
Υβριδοποίηση αποικιών ακινητοποιημένων σε μεμβράνη	
2.22. Χρώση πηκτωμάτων	71
2.23. Υγρή ηλεκτρομεταφορά	71
2.24. Ανάλυση Western	71
2.25. Αντίδραση αντίστροφης μεταγραφής	72
2.26. Υβριδοποίηση in situ σε τομές ιστών με ανιχνευτή RNA (RISH)	72
Προετοιμασία και μονιμοποίηση ιστών	
Στερεοποίηση και τεμαχισμός του ιστού	
Παρασκευή ανιχνευτή	
Υβριδοποίηση	
Πλυσίματα και επώαση με το αντίσωμα anti-DIG	
Πλυσίματα και αντίδραση αποκάλυψης	
2.27. Σύστημα δύο υβριδίων του σακχαρομύκητα	77
Έλεγχοι καταλληλότητας της βιβλιοθήκης cDNA	
Επιλογή του στελέχους του σακχαρομύκητα	
Κλωνοποίηση του cDNA του FRA10AC1 στον pGBT9	
Έλεγχος αυτόνομης ενεργοποίησης	
Τιτλοδότηση της βιβλιοθήκης	
Μετασχηματισμός κυττάρων που περιέχουν το δόλωμα με DNA της βιβλιοθήκης και καλλιέργεια σε θρεπτικό μέσο επιλογής	
Επιβεβαίωση της ενεργοποίησης των γονιδίων αναφοράς	
Εξάλειψη του δολώματος	
Απομόνωση και ταυτοποίηση του θηράματος	
2.28. Κυτταροκαλλιέργεια	83
Διατήρηση κυτταροκαλλιιεργειών	
Κρυοδιατήρηση κυττάρων	
Επανακαλλιέργεια	
2.29. Παροδική διαμόλυνση κυττάρων θηλαστικών με πλασμιδιακό DNA	85
2.30. Καθαρισμός συγγένειας πολυκλωνικού αντισώματος	86
2.31. Ανοσοεντοπισμός πρωτεϊνών με μικροσκοπία φθορισμού	86
2.32. Παραγωγή και απομόνωση πρωτεϊνών	87
Παραγωγή σε βακτήρια και απομόνωση με χρωματογραφία συγγένειας	
Παραγωγή πρωτεΐνης σε κύτταρα θηλαστικών	
2.33. Συζευγμένη μεταγραφή-μετάφραση in vitro	89
2.34. Δοκιμή συνκατακρήμνισης πρωτεϊνών in vitro (GST pull-down)	90
2.35. Συν-ανοσοκατακρήμνιση πρωτεϊνών	91

3. ΑΠΟΤΕΛΕΣΜΑΤΑ

3.1. Κατασκευή του φυσικού χάρτη των χρωμοσωμικών περιοχών 10q23.31-q25.3 και 10q26.2-qter	92
➤ Διερεύνηση γενωμικής βιβλιοθήκης BAC	94
α. Επιλογή ανιχνευτών	
β. Παρασκευή ανιχνευτών	
γ. Υβριδοποίηση και επιλογή κλώνων BAC	
➤ Κατασκευή και διερεύνηση γενωμικής βιβλιοθήκης του χρωμοσώματος 10	98
Αντιστοίχιση κλώνων BAC με STSs	
➤ Χαρτογράφηση contigs και χρωμοσωμικό περπάτημα	109
α. Φυσική χαρτογράφηση των προκαταρκτικών contigs	
β. Χρωμοσωμικό περπάτημα	
3.2. Η αξιοποίηση του χρωμοσωμικού χάρτη 10q23.3 για την ταυτοποίηση και ανάλυση γονιδίων που σχετίζονται με παθολογικούς φαινότυπους	118
A. Ταυτοποίηση και μοριακή ανάλυση γονιδίων που χαρτογραφούνται στην κρίσιμη περιοχή της ADLTE	118
➤ Γονίδιο #44050	122

➤	Γονίδιο #3077	123
➤	Γονίδιο #AJ312051	124
➤	Γονίδιο TMEM10	126
➤	Γονίδιο LG11	127
B.	Ταυτοποίηση και χαρακτηρισμός του γονιδίου που συνδέεται με την εύθραυστη θέση FRA10A	137
➤	Προσδιορισμός και κλωνοποίηση του cDNA του γονιδίου FRA10AC1	137
➤	Προσδιορισμός της γενωμικής οργάνωσης του FRA10AC1	142
➤	Γονοτυπική ανάλυση της τρινουκλεοτιδικής επανάληψης CGG σε άτομα ελέγχου και σε φορείς της FRA10A	144
➤	Το πρότυπο έκφρασης του γονιδίου FRA10AC1	147
	Ανάλυση Northern	
	Πειράματα RT- PCR	
	Μελέτη του προτύπου έκφρασης του Fra10Ac1 κατά την εμβρυϊκή ανάπτυξη και στον εγκέφαλο του ποντικού	
➤	Ανάλυση εναλλακτικών μεταγράφων και ισομορφές της FRA10AC1	152
➤	Δομικά χαρακτηριστικά της πρωτεΐνης FRA10AC1	153
➤	Έλεγχος της ειδικότητας του αντισώματος αντι-FRA10AC1	160
➤	Υποκυτταρική τοπολογική ανάλυση της ανασυνδυσασμένης και της ενδογενούς πρωτεΐνης FRA10AC1 σε κυτταρικές σειρές θηλαστικών	161
➤	Διερεύνηση πρωτεϊνικών αλληλεπιδράσεων της FRA10AC1	166
	α. Προσδιορισμός αλληλεπιδράσεων με χρήση του συστήματος δύο υβριδίων του σακχαρομύκητα	166
	Κατασκευή και έλεγχος του δολώματος	
	Έλεγχος και πιλοδότηση της βιβλιοθήκης	
	Διερεύνηση της βιβλιοθήκης	
	Απομόνωση, ανάλυση και επαλήθευση θετικών κλώνων	
	β. Επιβεβαίωση πρωτεϊνικών αλληλεπιδράσεων in vitro	169
	Χαρτογράφηση των περιοχών αλληλεπίδρασης των πρωτεϊνών FRA10AC1 και DGCR14	
	γ. Επιβεβαίωση πρωτεϊνικών αλληλεπιδράσεων in vivo	174
➤	Διερεύνηση της υποκυτταρικής τοπολογίας των πρωτεϊνών FRA10AC1, DGCR14 και SAP145	175
➤	Συγκριτική μελέτη του προτύπου έκφρασης των γονιδίων Fra10Ac1, Dgcr14 και Sap145 κατά την εμβρυϊκή ανάπτυξη και στον εγκέφαλο του ποντικού	176

4. ΣΥΖΗΤΗΣΗ

4.1.	Γονιδιωματική ανάλυση του χρωμοσώματος 10	180
➤	Χάρτης κλώνων-αλληλουχίας	180
➤	Σύγκριση γενετικού-φυσικού χάρτη	180
➤	Επαναλαμβανόμενες αλληλουχίες	182
➤	Αριθμός γονιδίων	184
➤	Δομικά χαρακτηριστικά των γονιδίων και κατανομή τους	186
➤	Περιεχόμενο σε GC - Νησίδια CpG	188
➤	Χαρακτηρισμός γονιδίων in silico	191
➤	Εναλλακτική συναρμογή	192
➤	Αλληλοεπικαλυπτόμενα γονίδια	193
➤	Ψευδογονίδια	193
➤	microRNA	194
4.2.	Συγκριτική γονιδιωματική	195
➤	Συγκριτική γονιδιωματική ανάλυση του ανθρώπου με άλλα θηλαστικά	196
	α. Συγκριτική ανάλυση γονιδιωμάτων ανθρώπου και ποντικού-αρουραίου	
	Συνταινία	

Συντήρηση σε γονίδια	
Συντήρηση σε ψευδογονίδια	
Συντήρηση σε μη κωδικές περιοχές	
β. Συγκριτική ανάλυση γονιδιωμάτων ανθρώπου-χιμπατζή (<i>Pan troglodytes</i>)	
➤ Συγκριτική γονιδιωματική ανάλυση του ανθρώπου-ασπονδύλων	203
➤ Πληροφορίες για ορθόλογες περιοχές <i>in silico</i>	204
4.3. Μοριακή βάση της σπάνιας αυτοσωμικής εύθραυστης θέσης FRA10A	207
➤ Πρωτεΐνες που αλληλεπιδρούν με την FRA10AC1	211
Αντίδραση συναρμογής	
Το spliceosome	
Σχηματισμός των υπομονάδων U snRNP	
Σχηματισμός του spliceosome	
Πολυπλοκότητα του spliceosome	
Η πρωτεΐνη SAP145	215
Η πρωτεΐνη DGCR14	217
➤ Η πρωτεΐνη FRA10AC1 και η συμμετοχή της στο spliceosome	225
Ασθενής σύνδεση στο spliceosome	
Εξειδικευμένη δράση	
Συσχέτιση με εναλλακτική συναρμογή	
Η πρωτεΐνη FRA10AC1 ως πιθανός παράγοντας σύζευξης της συναρμογής με άλλες διαδικασίες	
Συσχέτιση με μεταγραφή	
Συσχέτιση με μεταφορά του mRNA στο κυτταρόπλασμα	
4.4 Προοπτικές	232
ΒΙΒΛΙΟΓΡΑΦΙΑ	233
ΔΗΜΟΣΙΕΥΣΕΙΣ	

ΚΑΤΑΛΟΓΟΣ ΠΙΝΑΚΩΝ

	Σελίδες
Πιν. 1: Τα χρωμοσώματα για τα οποία έχει ολοκληρωθεί η αλληλούχηση και η ανάλυσή τους	9
Πιν. 2: Τα γονίδια στα οποία έχουν ταυτοποιηθεί μεταλλαγές στις κύριες μορφές ιδιοπαθών γενικευμένων και μερικών επιληψιών	29
Πιν. 3: Διαλύματα αντιβιοτικών	36
Πιν. 4: Οι εκκινητές των STSs της χρωμοσωμικής περιοχής 10q23.3-q25.3 και 10q26.2-qter	48
Πιν. 5: Αλληλουχίες εκκινητών	57
Πιν. 6: Αντισώματα και συνθήκες εφαρμογής	58
Πιν. 7: Πρόελευση των STSs που χρησιμοποιήθηκαν στην φυσική χαρτογράφηση	94
Πιν. 8: Αποτελέσματα της διερεύνησης της γενωμικής βιβλιοθήκης BAC	99
Πιν. 9: Προκαταρκτικά contigs που χαρτογραφήθηκαν στις περιοχές 10q23.31-q25.3 και 10q26.2-qtel	112
Πιν. 10: Νέα STSs από ακραίες αλληλουχίες κλώνων BAC από τις περιοχές 10q23.31-q25.3 και q26.2-qter.	116
Πιν. 11: Γενωμικοί κλώνοι του χρωμοσώματος 10 που περιέχουν την επανάληψη CGG.	139
Πιν. 12: Γονοτυπική ανάλυση για την μελέτη του πολυμορφισμού της επανάληψης (CGG)n.	146
Πιν. 13: Χαρακτηριστικά των χρωμοσωμάτων	187
Πιν. 14: Συγκριτική ανάλυση ομοιότητας ορθολόγων γονιδίων ανθρώπου-τροπικών	197

ΚΑΤΑΛΟΓΟΣ ΕΙΚΟΝΩΝ

Εικ. 1: Στρατηγική της κλωνοποίησης γονιδίων βάσει της χρωμοσωμικής τοπογραφίας τους (ανατύπωση από Schuler GD <i>et al</i> , 1996)	2
Εικ. 2: Εκθετική αύξηση του ρυθμού κλωνοποίησης των γονιδίων που σχετίζονται με νοσήματα (στοιχεία από το National Human Genome Research Institute, NIH, http://www.genome.gov/).	3
Εικ. 3: Η διακήρυξη της Valencia	7
Εικ. 4: Τα εξώφυλλα των περιοδικών Nature και Science τον Φεβρουάριο του 2001, που περιγράφουν την ανάλυση της πρώτης έκδοσης της αλληλουχίας του ανθρώπινου γονιδιώματος από το HGP και την εταιρία Celera Genomics αντίστοιχα.	8
Εικ. 5: Ιεραρχική shotgun αλληλούχηση. Ανατύπωση από Lander ES <i>et al</i> , 2001	15
Εικ. 6: Ενδεικτική ηλεκτροφόρηση σε πήκτωμα αγαρόζης των προϊόντων τεμαχισμού των κλώνων BAC.	17
Εικ. 7: Τα γονίδια του χρωμοσώματος 10 που έχουν συσχετιστεί με γενετικά νοσήματα	27
Εικ. 8: Η εύθραυστη θέση FRAXA	31
Εικ. 9: Σχηματική απεικόνιση του φορέα pACT2.	38
Εικ. 10: Σχηματική αναπαράσταση της κατασκευής των ριβοανιχνευτών με αντίδραση μεταγραφής <i>in vitro</i> .	75
Εικ. 11: Σχηματική αναπαράσταση της κατασκευής των ριβοανιχνευτών με αντίδραση μεταγραφής <i>in vitro</i> .	76
Εικ. 12: Αρχή του συστήματος δύο υβριδίων.	79
Εικ. 13: Πορεία κατασκευής του φυσικού χάρτη.	93
Εικ. 14: Χάρτες του χρωμοσώματος 10 από τους οποίους επιλέχθηκαν τα STSs για την διερεύνηση της βιβλιοθήκης BAC.	95
Εικ. 15: Προϊόντα PCR για 12 STSs της χρωμοσωμικής περιοχής 10q23.31-q25.3.	97
Εικ. 16: Διάταξη και αυτοραδιογραφία της γενωμικής βιβλιοθήκης BAC.	98
Εικ. 17: Ενδεικτική καταχώρηση των αποτελεσμάτων της διερεύνησης της γενωμικής βιβλιοθήκης BAC στην βάση δεδομένων 10ace.	107
Εικ. 18: Καταχώρηση των αποτελεσμάτων αντιστοίχισης ενός STS με τα BACs που το περιέχουν.	108
Εικ. 19: Προκαταρκτικό contig που χαρτογραφείται στην χρωμοσωμική περιοχή 10q23.33.	109

Εικ. 20: Χαρτογράφηση των contigs των περιοχών 10q23.31-q25.3 και 10q26.2-qter βάσει του περιεχομένου τους σε STSs.	111
Εικ. 21: Ο φυσικός χάρτης του χρωμοσώματος 10.	117
Εικ. 22: Ο φυσικός χάρτης του χρωμοσώματος 10 στην περιοχή 84-86Mb.	118
Εικ. 23: Γονιδιακός χάρτης της κρίσιμης περιοχής της ADLTE.	121
Εικ. 24: Σχηματική απεικόνιση της γενωμικής οργάνωσης, των εναλλακτικών μεταγράφων και της πρωτοταγούς δομή των αντίστοιχων πολυπεπτιδίων του γονιδίου 44050.	123
Εικ. 25: Σχηματική απεικόνιση της γενωμικής οργάνωσης, των εναλλακτικών μεταγράφων και της πρωτοταγούς δομή των αντίστοιχων πολυπεπτιδίων του γονιδίου 3077.	124
Εικ. 26: Σχηματική απεικόνιση της γενωμικής οργάνωσης, των εναλλακτικών μεταγράφων και της πρωτοταγούς δομή των αντίστοιχων πολυπεπτιδίων του γονιδίου AJ312051.	125
Εικ. 27: Σχηματική απεικόνιση της γενωμικής οργάνωσης και της πρωτοταγούς δομής της πρωτεΐνης TMEM10.	126
Εικ. 28: Προσδιορισμός του προτύπου έκφρασης των γονιδίων 44050 και 3077 με ανάλυση Northern.	126
Εικ. 29: Υποκυτταρικός εντοπισμός των ανασυνδυασμένων πρωτεϊνών 3077 και AJ312051 σε κύτταρα COS-7.	127
Εικ. 30: Δομή της πρωτεΐνης LGI1.	128
Εικ. 31: Οι μεταλλαγές που έχουν ταυτοποιηθεί στο γονίδιο <i>LGI1</i> σε ασθενείς με ADLTE.	129
Εικ. 32: Σχηματική απεικόνιση των ριβοανιχνευτών που χρησιμοποιήθηκαν στην <i>in situ</i> υβριδοποίηση.	131
Εικ. 33: Αλληλοεπικάλυψη των 3' άκρων των γονιδίων <i>Lgi4</i> και <i>Fxyd3</i> στο χρωμόσωμα 7 του ποντικού.	131
Εικ. 34A: Συνέκφραση των γονιδίων <i>Lgi1-4</i> στον εγκέφαλο.	134
B: Περιοχές υψηλής μεταγραφικής ενεργότητας των γονιδίων <i>Lgi1-4</i> στον εγκέφαλο.	135
Γ: Έκφραση των γονιδίων <i>Lgi1-4</i> σε διακριτές περιοχές του εγκεφάλου.	136
Εικ. 35: Χαρτογράφηση των γενωμικών κλώνων στο χρωμόσωμα 10 και της επανάληψης (CGG) _n σε αυτούς.	
Εικ. 36: Κατασκευή του cDNA πλήρους μήκους <i>in silico</i> .	141
Εικ. 37: Η νουκλεοτιδική αλληλουχία (1344bp) του cDNA του γονιδίου <i>FRA10AC1</i> .	143
Εικ. 38: Γενωμική οργάνωση του γονιδίου <i>FRA10AC1</i> .	144
Εικ. 39: Θέση υβριδοποίησης των εκκινητών F1 και R1 στο γενωμικό DNA εκατέρωθεν της επανάληψης CGG.	145
Εικ. 40: Γονοτυπική ανάλυση σε δείγματα DNA οικογενειών CEPH για την μελέτη του πολυμορφισμού της επανάληψης CGG.	145
Εικ. 41: Προσδιορισμός του προτύπου έκφρασης του γονιδίου <i>FRA10AC1</i> με ανάλυση κατά Northern.	147
Εικ. 42: Προσδιορισμός του προτύπου έκφρασης του γονιδίου <i>FRA10AC1</i> με RT-PCR και κλωνοποίηση εναλλακτικών μεταγράφων.	149
Εικ. 43: Συγκριτική γενωμική οργάνωση του γονιδίου <i>FRA10AC1</i> στον άνθρωπο και τον ποντικό.	150
Εικ. 44: Προσδιορισμός του προτύπου έκφρασης του <i>Fra10ac1</i> με RNA <i>in situ</i> υβριδοποίηση σε οβελιαίες τομές εμβρύων ποντικού ηλικίας E12.5, E13.5 και E14.5.	151
Εικ. 45: Προσδιορισμός του προτύπου έκφρασης του <i>Fra10ac1</i> με RNA <i>in situ</i> υβριδοποίηση σε στεφανιαίες τομές εγκεφάλου ενήλικου ποντικού.	152
Εικ. 46: Γραφική αναπαράσταση της πρωτεΐνης FRA10AC1.	154
Εικ. 47: Πρόβλεψη πιθανών φωσφορυλιώσεων στην FRA10AC1 και στις εναλλακτικές ισομορφές γ1 και γ3, βάσει του προγράμματος NetPhos.	155
Εικ. 48: Προβλεπόμενα ορθόλογα μόρια της FRA10AC1 στο ζωικό και φυτικό βασίλειο.	156
Εικ. 49: Συγκριτική ανάλυση της αλληλουχίας της πρωτεΐνης FRA10AC1 με τα πιθανά ορθόλογα μόρια.	158
Εικ. 50: Το ορθόλογο γονίδιο <i>FRA10AC1</i> στον <i>C.elegans</i> (F26C11.1.).	160
Εικ. 51: Έλεγχος της εξειδίκευσης του αντισώματος anti-FRA10AC1.	162

Εικ. 52: Έλεγχος εξειδίκευσης αντισώματος anti-FRA10AC1.	163
Εικ. 53: Η ανασυνδυασμένη FRA10AC1-EGFP εντοπίζεται στον πυρήνα των κυττάρων HeLa με διάχυτη ή στικτή κατανομή.	164
Εικ. 54: Η ενδογενής FRA10AC1 εντοπίζεται στον πυρήνα.	164
Εικ. 55: Το τμήμα της FRA10AC1 που περιλαμβάνει τα αμινοξέα 147-253 είναι υπεύθυνο για την πυρηνική τοπολογία του μορίου.	165
Εικ. 56: Γραφική απεικόνιση των «δολωμάτων» που χρησιμοποιήθηκαν στο σύστημα δύο υβριδίων.	167
Εικ. 57: Σχηματική αναπαράσταση της διαδικασίας που εφαρμόστηκε στο σύστημα δύο υβριδίων και των μορίων που ταυτοποιήθηκαν.	169
Εικ. 58: Αλληλεπίδραση της FRA10AC1 με τις DGCR14 και SAP145 <i>in vitro</i> .	171
Εικ. 59: Χαρτογράφηση των περιοχών αλληλεπίδρασης των πρωτεϊνών FRA10AC1 και DGCR14.	173
Εικ. 60: Η ενδογενής πρωτεΐνη FRA10AC1 συν-ανοσοκατακρημνίζεται με την ανασυνδυασμένη DGCR14-EGFP.	174
Εικ. 61: Υποκυτταρικός συνεντοπισμός των πρωτεϊνών FRA10AC1, DGCR14 και SAP145.	176
Εικ. 62: Συγκριτική μελέτη του προτύπου έκφρασης των γονιδίων <i>Fra10ac1</i> , <i>Dgcr14</i> και <i>Sap145</i> με RNA <i>in situ</i> υβριδοποίηση σε οβελιαίες τομές εμβρύων ποντικού ηλικίας E12.5, E13.5 και E14.5 και σε στεφανιαίες τομές εγκεφάλου ενήλικου ποντικού.	179
Εικ. 63: Ο χάρτης της αλληλουχίας του χρωμοσώματος 10, η συνταινία με τον ποντικό, τον αρουραίο και τους ιχθύες και τα γονίδια που κωδικοποιούν πρωτεΐνες.	180
Εικ. 64: Σύγκριση γενετικού-φυσικού χάρτη κατά μήκος του χρωμοσώματος 10.	181
Εικ. 65: Σύγκριση γενετικού-φυσικού χάρτη στη χρωμοσωμική περιοχή 10q23.31-q25.3.	183
Εικ. 66: Ταξινόμηση των γονιδίων του χρωμοσώματος 10 βάσει της ολοκληρωμένης αλληλουχίας του (Deloukas P <i>et al</i> , 2004).	185
Εικ. 67: Κατανομή των γονιδίων κατά μήκος του χρωμοσώματος.	189
Εικ. 68: Νοσήματα που οι κρίσιμες περιοχές τους χαρτογραφούνται στο χρωμόσωμα 10 και για τα οποία δεν έχει ταυτοποιηθεί το υπεύθυνο γονίδιο.	190
Εικ. 69: Συνάρτηση του αριθμού των νησιδίων CpG ανά Mb σε σχέση με τον αριθμό των γονιδίων ανά Mb, με βάση τον χάρτη GeneMap 98, για κάθε χρωμόσωμα (ανατύπωση από Lander ES <i>et al</i> , 2001).	191
Εικ. 70: Ταξινόμηση του πρωτεϊνώματος του χρωμοσώματος 10 και σύγκριση με το συνολικό πρωτεϊνώμα.	192
Εικ. 71: Απεικόνιση των συνταινικών χρωμοσωμικών περιοχών μεταξύ ανθρώπου, αρουραίου και ποντικού.	197
Εικ. 72: Συγκριτική ταξινόμηση των πρωτεϊνών του ανθρώπου και του ποντικού σε λειτουργικές ομάδες	199
Εικ. 73: Απεικονίσεις από τη βάση δεδομένων <i>Ensembl</i> .	206
Εικ. 74: Κυτταρογενετική χαρτογράφηση και ανάλυση FISH της εύθραυστης θέσης FRA10A.	207
Εικ. 75: Επέκταση και μεθυλίωση της τρινουκλεοτιδικής επανάληψης CGG σε φορείς της εύθραυστης θέσης FRA10A.	209
Εικ. 76: Πυκνότητα των διαφορετικών τύπων τρινουκλεοτιδικών επαναλήψεων στα χρωμοσώματα του ανθρώπου.	210
Εικ. 77: Μεταγραφική καταστολή του ενός αλληλομόρφου του γονιδίου <i>FRA10AC1</i> στους φορείς της εύθραυστης θέσης FRA10A.	210
Εικ. 78: Τα δύο καταλυτικά βήματα της αντίδρασης συναρμογής	211
Εικ. 79: Απλοποιημένο σχήμα της βιογένεσης των U snRNPs	214
Εικ. 80: Σχηματισμός ενδιάμεσων συμπλόκων κατά την αντίδραση συναρμογής	215
Εικ. 81: Χαρακτηριστική κρανιοπροσωπική δυσμορφία σε ασθενή με σύνδρομο DiGeorge	217
Εικ. 82: Χρωμοσωμικά ελλείμματα στην περιοχή 22q11 σε ασθενείς με σύνδρομο DiGeorge και σε ποντίκια knockout.	219
Εικ. 83: Πρότυπο έκφρασης του γονιδίου <i>DGCR14</i> στον άνθρωπο και τον ποντικό με ανάλυση Northern.	222

ΠΕΡΙΛΗΨΗ

Στα πλαίσια του προγράμματος της χαρτογράφησης του γονιδιώματος του ανθρώπου, στο πρώτο μέρος αυτής της διατριβής κατασκευάσαμε το φυσικό χάρτη υψηλής ευκρίνειας των πλούσιων σε γονίδια, περιοχών q23.3-q25 και q26.2-qter του χρωμοσώματος 10, μήκους ~35 Mb που ισοδυναμεί με περίπου το $\frac{1}{4}$ του μήκους του χρωμοσώματος. Για την κατασκευή του χάρτη, αρχικά διερευνήσαμε μια ολική γενωμική βιβλιοθήκη κλώνων BAC χρησιμοποιώντας κατά ομάδες, ~500 μοναδιαία τμήματα DNA (STS) από τις ανωτέρω χρωμοσωμικές περιοχές ως ανιχνευτές, και απομονώσαμε 3350 διακριτούς χρωμοσωμο-ειδικούς κλώνους. Οι κλώνοι αυτοί, σε συνδυασμό με εκείνους που με την ίδια προσέγγιση είχαν απομονωθεί από τους συνεργάτες μας στο Sanger Centre, απετέλεσαν την χρωμοσωμο-ειδική βιβλιοθήκη του 10, αποτελούμενη συνολικά από ~18.000 BACs. Η διερεύνηση αυτής της βιβλιοθήκης με κάθε ένα από τα παραπάνω STS, ξεχωριστά, μας οδήγησε στην λεπτομερή και απόλυτα αξιόπιστη αντιστοίχιση κάθε BAC με μια ομάδα από STSs. Τα δεδομένα μας, σε συνδυασμό με αυτά που προέκυψαν από τον προσδιορισμό του προτύπου ενζυμικού τεμαχισμού των αντίστοιχων BACs (Sanger Centre UK, Washington University St. Louis US) χρησιμοποιήθηκαν για την κατασκευή κατ' αρχήν, επιμέρους contigs (ομάδες αλληλοεπικαλυπτόμενων κλώνων) τα οποία, βάσει των STS που περιείχαν, χαρτογραφούνταν κατά μήκος των ανωτέρω χρωμοσωμικών περιοχών. Για την επέκταση και σταδιακή ενοποίηση των contigs, καθορίσαμε με χρωμοσωμικό «περπάτημα», περίπου 80 νέους μοριακούς δείκτες από τα άκρα των τελικών/εξωτερικών κλώνων των contigs με τελικό αποτέλεσμα την κατασκευή ενός ενιαίου, υψηλής ευκρίνειας φυσικού χάρτη για την περιοχή 10q23.3-q25, μήκους περίπου 28 Mb. Η αξιοπιστία της σχετικής τοπογραφίας κάθε γενωμικής θέσης προκύπτει από το «βάθος» του contig στη περιοχή, όπως καθορίζεται από τον αριθμό των επικαλυπτόμενων κλώνων που επιβεβαιώνουν τη σχετική θέση. Σε κάθε περίπτωση, τουλάχιστον 8 επικαλυπτόμενα, λεπτομερώς χαρακτηρισμένα BACs, καθόριζαν κάθε χρωμοσωμική θέση. Με παρόμοια προσέγγιση, συγκροτήθηκε ο φυσικός χάρτης της υποτελομερικής περιοχής αποτελούμενος από δύο contigs μεγέθους 4,6Mb και 2Mb. Η ολοκληρωμένη αλληλούχηση των περιοχών 10q23.3-q25 και 10q26.2-qter από τους συνεργάτες μας (εργαστήριο αυτόματης αλληλούχησης DNA, Sanger Centre UK) έγινε βάσει των παραπάνω χαρτών. Για το σκοπό αυτό, επιλέχθηκε και αλληλουχήθηκε μία σειρά κλώνων BAC (~300), οι οποίοι διατρέχουν τις εν λόγω χρωμοσωμικές περιοχές καθ' όλο το μήκος τους, εμφανίζουν την μέγιστη αξιοπιστία ως προς την τοπογραφία τους και παρουσιάζουν την ελάχιστη δυνατή αλληλοεπικάλυψη (minimal tiling path). Πλέον, η πλήρης αλληλουχία στις αντίστοιχες περιοχές, που περιέχουν 272 από τα 816 γονίδια του χρωμοσώματος που κωδικοποιούν πρωτεΐνες και όπου χαρτογραφούνται 40 γενετικοί τόποι νοσημάτων, είναι σχεδόν ενιαία και ελεύθερα προσβάσιμη στο διαδίκτυο.

Στο δεύτερο μέρος της διατριβής, αξιοποιήσαμε τον φυσικό χάρτη και την αλληλουχία του χρωμοσώματος 10 για να προσδιορίσουμε *in silico*, τον γονιδιακό χάρτη της περιοχής

10q23.3, μήκους 3Mb με στόχο τη ταυτοποίηση γονιδίων που σχετίζονται με παθολογικούς φαινοτύπους. Εστιαστήκαμε σε αυτή την υποπεριοχή λόγω της σύνδεσης της, εκτός άλλων, με α) το γενετικό τόπο της αυτοσωμικής επικρατούς μερικής επιληψίας του πλάγιου κροταφικού λοβού (ADLTE) και β) την έκφραση της εύθραυστης χρωμοσωμικής θέσης FRA10A. Μετά το χαρακτηρισμό μιας σειράς γονιδίων της περιοχής, στα πλαίσια του ευρωπαϊκού consortium για την ADLTE, δείξαμε ότι μεταλλαγές στο γονίδιο *LG1*, ευθύνονται για την εκδήλωση της ασθένειας. Το *LG1* μαζί με τρία ακόμη γονίδια, συγκροτεί μια διασκορπισμένη στο γονιδίωμα, οικογένεια με κοινά χαρακτηριστικά τις πρωτεϊνικές επαναλήψεις EPT και το μερικώς επικαλυπτόμενο πρότυπο έκφρασης στον εγκέφαλο. Η γενετική ετερογένεια της ADLTE ή άλλες νευρολογικές ασθένειες, ίσως εξηγούνται από την ύπαρξη αυτών των παράλογων γονιδίων *LG1*.

Επίσης, με *in silico* διερεύνηση της αλληλουχίας του χρωμοσώματος για τρινουκλεοτιδικές επαναλήψεις (CGG)_n, ταυτοποιήσαμε το γονίδιο *FRA10AC1* που σχετίζεται με την εύθραυστη θέση FRA10A, και προχωρήσαμε στο χαρακτηρισμό του. Η κυτταρογενετική έκφραση της FRA10A οφείλεται στην επέκταση κατά ~200 φορές, της πολυμορφικής επανάληψης (CGG)_n της 5' UTR του γονιδίου. Η επέκταση του (CGG)_n οδηγεί στην υπερμεθυλίωση της περιοχής (αποτελέσματα από συνεργαζόμενο εργαστήριο) και συνεπάγεται την μεταγραφική καταστολή του γονιδίου. Το *FRA10AC1* είναι μεταγραφικά ενεργό σε όλους τους ιστούς παρουσιάζοντας υψηλότερα επίπεδα έκφρασης σε όργανα με υψηλή μεταγραφική δραστηριότητα. Η ύπαρξη 5 εναλλακτικών μεταγράφων στις γονάδες, πιθανόν να συνεπάγεται αντίστοιχο αριθμό ισομορφών της πρωτεΐνης FRA10AC1, ετερογενών στο καρβοξυ-τελικό άκρο. Η πρωτεΐνη εντοπίζεται στον πυρήνα του κυττάρου και εμφανίζεται συντηρημένη στους πολυκύτταρους ευκαρυωτικούς οργανισμούς υποδηλώνοντας συμμετοχή σε μια βασική λειτουργία του κυττάρου. Με στόχο τη διερεύνηση του βιοχημικού ρόλου της πρωτεΐνης FRA10AC1, αξιοποιήσαμε το σύστημα των δύο υβριδίων στο σακχαρομύκητα που αποκάλυψε την αλληλεπίδρασή της με τις πρωτεΐνες SAP145 και DGCR14. Οι πρωτεϊνικές αλληλεπιδράσεις της FRA10AC1 επιβεβαιώθηκαν με προσεγγίσεις *in vitro* και σε κυτταροκαλλιέργειες. Και οι δύο πρωτεΐνες που ταυτοποιήθηκαν, φαίνεται να συμμετέχουν σε κοινή βιολογική λειτουργία, στη διαδικασία συναρμογής του mRNA. Πράγματι, η SAP145 αποτελεί υπομονάδα του παράγοντα συναρμογής SF3b, ο οποίος συμμετέχει στο σχηματισμό του U2 snRNP, συστατικό του μείζονος spliceosome και η DGCR14 έχει απομονωθεί βιοχημικά ως πρωτεΐνη-συστατικό του ίδιου συμπλόκου. Καθώς και οι δύο σχετίζονται με τη διαδικασία της συναρμογής του mRNA, τα αποτελέσματά μας συνηγορούν στην συμμετοχή της FRA10AC1 σε αυτήν την διαδικασία ή εναλλακτικά σε άλλες λειτουργικά συσχετιζόμενες, διαδικασίες σύνθεσης, επεξεργασίας ή/και διακίνησης του mRNA.

ABSTRACT

In the first part of this thesis, we constructed a high resolution, sequence-ready, physical map of the gene-rich regions q23.3-q25 and q26.2-qter of human chromosome 10. These regions comprise about 35 Mb, i.e. about $\frac{1}{4}$ of the length of the chromosome. For the map construction, we initially screened a grided human BAC library, using about 500 distinct sequence tagged sites (STSs) from the above regions as probes, divided in groups, and isolated 3350 distinct, chromosome 10-specific BACs. Our clones, together with those similarly isolated by our Sanger Centre collaborators, formed the human chromosome 10 specific library, consisting of ~18.000 BACs. Screening of this library with each one of the above STSs, led us to the detailed and highly confident correlation of each one of the BACs with a group of STSs. We used these data, in combination with those derived from the BAC fingerprinting analysis (Sanger Center UK, Washington University St. Louis US) for the construction of the preliminary contigs (groups of overlapping clones) mapped along the above chromosomal regions. For the extension and progressive merging of the contigs, we determined by chromosome walking, about 80 new STS-markers, derived from the end-sequences of the most distal clones of each contig. This task gradually resulted in the construction of a unique, high resolution physical map of 10q23.3-q25, with a length of approximately 28 Mb. Contig depth, as defined by the number of overlapping clones in every map position, confidently supports the relative topography of each genomic site. In our map, mapping of each position was supported by at least 8 overlapping, well-characterized, BACs. Similarly, in terms of strategy and confidence, the physical map of the subtelomeric region was constructed, consisting of two contigs, 4.6 Mb and 2 Mb in length. Our mapping data were exploited by the collaborating laboratory of automated DNA sequencing (Sanger Center, UK), initially for the construction of a minimal tiling path of about 300 BACs mapped with the highest confidence, and subsequently, for the complete sequencing of the above regions. The corresponding finished sequence, currently available freely through the network, has a total size of about 35 Mb, consists of 272 out of 816 chromosome 10 protein coding genes and is associated with 40 disease-related genetic loci.

With the aim to identify disease-related genes, in the second part of this thesis, we exploited both physical mapping and sequencing data to construct *in silico*, the genic map of a 10q23.3 region, 3 Mb in length. Among others, this area is associated with i) the genetic locus of the autosomal dominant lateral temporal lobe epilepsy (ADLTE), and ii) the expression of the cytogenetic fragile site FRA10A. In collaboration with other labs, altogether forming the European Consortium of ADLTE, we characterized several ADLTE candidate genes. Among those, *LG1* was determined to be directly associated with the disease, since it was found mutated in patients of ADLTE families. *LG1* a gene of unknown function, belongs to a novel gene family consisting of four members (*LG1-4*) dispersed in the genome. All *LG1* genes are active in the brain displaying a partially overlapping pattern of expression; the encoded proteins

are characterized by a variable number of tandemly linked EPT repeats possibly involved in protein-protein interactions. It is reasonable to hypothesize that the characteristic genetic heterogeneity of ADLTE or other neurological diseases may be related with members of the *LGI* family.

Finally, by *in silico* screening of the nucleotide sequence of chromosome 10 for (CGG)_n repeats, we identified a novel gene, *FRA10AC1*, carrying a polymorphic (CGG)_n repeats in its 5' UTR and showed its direct involvement with the manifestation of FRA10A fragile site. Indeed, FRA10A carriers display expansion of the (CGG)_n repeat by approximately 200-fold. This expansion, as shown by a collaborating group, results in hypermethylation of the region and the consequent transcriptional silencing of the gene. *FRA10AC1* transcripts are detected ubiquitously in human tissues, being more abundant in organs with highly transcriptional activity. The presence of *FRA10AC1* alternative transcripts in gonads, suggested the existence of five protein isoforms displaying structural heterogeneity at their carboxy-terminus. FRA10AC1 protein is accumulated in the nucleoplasm and remains well-conserved in multi-cellular eukaryotes, implying participation of the FRA10AC1 in a fundamental cellular function. In order to investigate its biochemical role, using the yeast two-hybrid system, we showed that FRA10AC1 interacts with two proteins, SAP145 and DGCR14 and verified our finding by *in vitro* and cell culture assays. According to the recent literature, both interacting proteins are involved in the process of pre-mRNA splicing. Indeed, SAP145 is a subunit of the SF3b splicing factor participating in the formation of U2 snRNP, a component of the major spliceosome. Furthermore, DGCR14 has been repeatedly isolated by biochemical approaches, as a component of the same complex. Therefore, it is reasonable to believe that our data support a functional role of FRA10AC1 in the pre-mRNA splicing machinery or, its involvement in functionally linked procedures, i.e. synthesis, processing and nucleo-cytoplasmic transport of mRNA.

1

ΕΙΣΑΓΩΓΗ

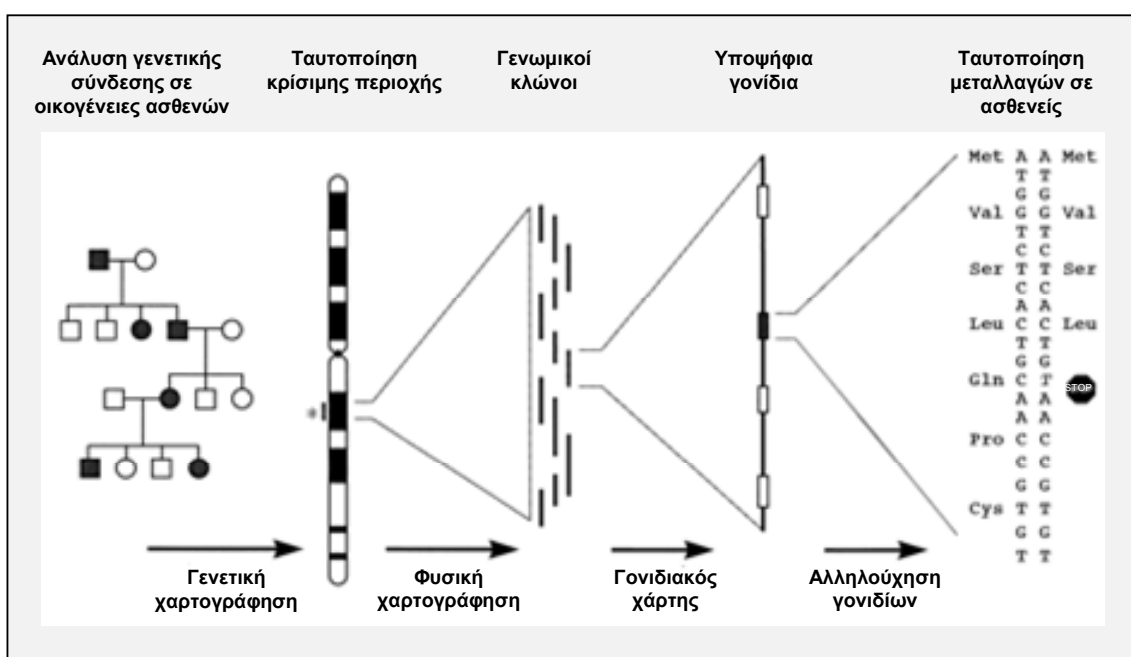
1.1. Γενικά

Η Μοριακή Γενετική του Ανθρώπου μελετάει την γενετική ποικιλότητα και την κληρονομικότητα στον άνθρωπο έχοντας βασικό στόχο εκτός των άλλων, την κατανόηση της μοριακής βάσης των νοσημάτων, τη σύνδεση δηλαδή του φαινοτύπου με τη γενετική σύσταση του ατόμου και την αξιοποίηση αυτής της γνώσης για τη βελτίωση των μεθόδων διάγνωσης και αποτελεσματικής θεραπείας. Τα γενετικά νοσήματα αποτελούν τις πιο εμφανείς και τις πιο ακραίες εκδηλώσεις των γενετικών παρεκκλίσεων, δεσπόζοντας πάνω σε μία ευρύτατη και καθ' όλα φυσιολογική γενετική ποικιλότητα.

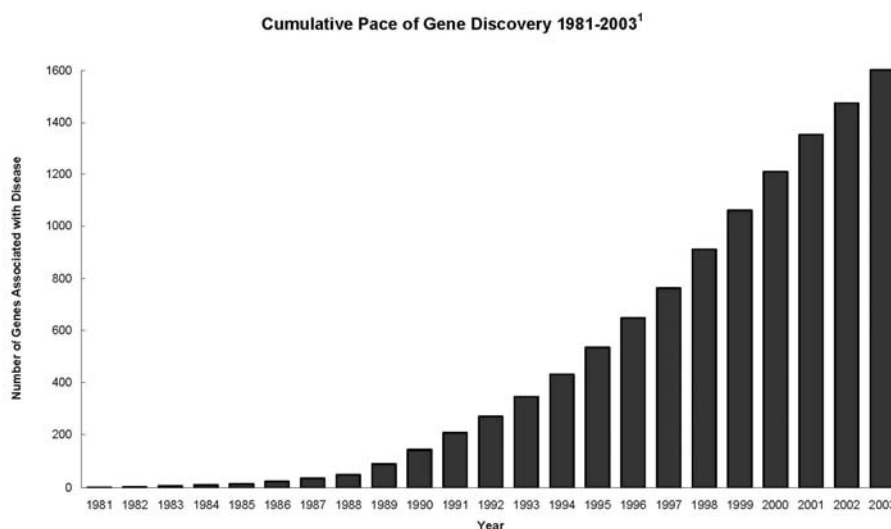
Για τα κληρονομικά νοσήματα που είναι γνωστή η βιοχημική τους βάση, είναι εφικτή συνήθως η ταυτοποίηση της δυσλειτουργικής πρωτεΐνης και στη συνέχεια ο χαρακτηρισμός του αντίστοιχου υπεύθυνου γονιδίου. Κλασικά πλέον παραδείγματα είναι οι διάφορες μορφές θαλασσαιμίας, η αιμορροφιλία Α, η φαιнуλκετονουρία κ.λπ. Οι περισσότερες όμως ασθένειες οφείλονται σε μεταλλαγές γονιδίων που το προϊόν τους και η λειτουργία του είναι άγνωστη ή, με άλλα λόγια, η κλινική εικόνα δεν μπορεί να συνδεθεί με κάποιο γνωστό βιοχημικό μονοπάτι ή βιολογική λειτουργία. Σ' αυτές τις περιπτώσεις όπου δεν υπάρχουν βιοχημικά ή άλλα δεδομένα που να οδηγούν σε ένα πιθανό υποψήφιο γονίδιο, η προσέγγιση που ακολουθείται είναι αυτή της κλωνοποίησης του υπεύθυνου γονιδίου βάσει της τοπογραφίας του στο χρωμόσωμα. Η προσέγγιση αυτή είναι πολύ αποτελεσματική για νοσήματα που εμφανίζουν μεντελικό τύπο κληρονόμησης και συνεπώς οφείλονται ως επί το πλείστον, σε μεταλλαγές ενός γονιδίου που παρουσιάζει μεγάλη διεισδυτικότητα. Η διαδικασία περιλαμβάνει την ταυτοποίηση της κρίσιμης περιοχής μέσω ανάλυσης γενετικής σύνδεσης σε οικογένειες ασθενών, την ταυτοποίηση των γονιδίων που χαρτογραφούνται στη συγκεκριμένη χρωμοσωμική περιοχή, την επιλογή ενός ή περισσότερων υποψηφίων γονιδίων και τη διερεύνηση μεταλλαγών σε ασθενείς (Εικ. 1). Βασική προϋπόθεση γι' αυτήν την πορεία είναι η ύπαρξη λεπτομερών γενετικών και φυσικών χρωμοσωμικών χαρτών.

Πριν μία δεκαετία, τα παραπάνω στάδια ήταν πολύ επίπονα και χρονοβόρα καθώς οι αρχικοί γενετικοί χάρτες ήταν χαμηλής ευκρίνειας και σημαντικό ποσοστό των γονιδίων του ανθρώπου δεν είχε χαρακτηριστεί. Σταδιακά, οι διαδικασίες άρχισαν να επιταχύνονται κυρίως χάρη στη συνεισφορά των αποτελεσμάτων του Προγράμματος Χαρτογράφησης και Ανάλυσης του Ανθρώπινου Γονιδιώματος (HGP). Στα πλαίσια αυτού του προγράμματος, κατασκευάστηκαν χρωμοσωμικοί χάρτες με σταδιακά αυξανόμενη διακριτική ικανότητα που συνεισέφεραν στην ελαχιστοποίηση του εύρους των κρίσιμων χρωμοσωμικών περιοχών. Επιπλέον, με την κλωνοποίηση, τη χαρτογράφηση και αλληλούχηση μεγάλων χρωμοσωμικών

περιοχών και την άμεση διοχέτευση των πληροφοριών σε δημόσιες βάσεις δεδομένων, επιταχύνθηκε κατά πολύ η ταυτοποίηση των γονιδίων που σχετίζονται με μεντελικά νοσήματα. Πλέον, με διαθέσιμη την ολοκληρωμένη έκδοση της αλληλουχίας του DNA του γονιδιώματος του ανθρώπου, ένα μεγάλο μέρος της εργαστηριακής έρευνας μεταφέρθηκε *in silico*. Έτσι, όχι μόνο είναι δυνατός ο εντοπισμός των γονιδίων που συνδέονται με παθολογικούς φαινοτύπους, αλλά επιπλέον υπάρχουν διαθέσιμες πληροφορίες για τον πιθανό βιοχημικό τους ρόλο, τα πιθανά ορθόλογα και παράλογα γονίδια, τους φαινοτύπους που συνδέονται με μεταλλάξεις τους, τις αλληλεπιδράσεις των πρωτεϊνικών προϊόντων τους κ.λπ. Η συνεισφορά των γενετικών και γονιδιωματικών πληροφοριών που προέκυψαν από το HGP, φαίνεται από την εκθετική αύξηση του αριθμού των γονιδίων που έχουν συσχετιστεί με νοσήματα τα τελευταία χρόνια (Εικ. 2). Ωστόσο, μέχρι σήμερα η πλειονότητα των γονιδίων που έχουν συσχετισθεί με νοσήματα αφορά μονογονιδιακούς παθολογικούς φαινοτύπους που η συχνότητα εμφάνισής τους στον πληθυσμό είναι πολύ μικρότερη (2-3 τάξεις μεγέθους) από αυτήν των συνήθων πολυπαραγοντικών νοσημάτων όπως ο διαβήτης, το άσθμα, ο καρκίνος, η σχιζοφρένεια, η παχυσαρκία κ.λπ. Ο εντοπισμός των γονιδίων που συνδέονται με αυτά τα πολυπαραγοντικά νοσήματα είναι μία από τις μεγαλύτερες προκλήσεις στην Ιατρική Γενετική καθώς περισσότερα του ενός γονίδια – ή καλύτερα, συγκεκριμένοι συνδυασμοί αλληλομόρφων – σε συνδυασμό με περιβαλλοντικούς παράγοντες, συμμετέχουν στην εκδήλωση του παθολογικού φαινοτύπου.



Εικ. 1: Στρατηγική της κλωνοποίησης γονιδίων βάσει της χρωμοσωμικής τοπογραφίας τους (ανατύπωση από Schuler GD *et al*, 1996).



Εικ. 2: Εκθετική αύξηση του ρυθμού κλωνοποίησης των γονιδίων που σχετίζονται με νοσήματα (στοιχεία από το National Human Genome Research Institute, NIH, <http://www.genome.gov/>).

Η ανάλυση του ανθρώπινου γονιδιώματος, δεν συμβάλλει μόνο στον προσδιορισμό του γονιδιακού χάρτη αλλά προσφέρει στοιχεία για την οργάνωση και τη ρύθμιση της ενεργότητας των γονιδίων στα πλαίσια της χρωμοσωμικής αλληλουχίας DNA. Παράλληλα, με την ανάλυση των γονιδιωμάτων οργανισμών-μοντέλων, η συγκριτική γονιδιωματική ανάλυση παρέχει πληροφορίες για βασικούς κοινούς βιολογικούς μηχανισμούς αλλά και για τις λεπτομέρειες της γονιδιακής δομής, λειτουργίας και ρύθμισης όταν η συγκριτική ανάλυση γίνεται σε κοντινούς, εξελικτικά, οργανισμούς. Επιπλέον, στην μετα-γονιδιωματική εποχή ανοίγονται νέες δυνατότητες για τη λειτουργική διερεύνηση των γονιδίων σε γενικευμένη κλίμακα (high throughput analysis).

Στα πλαίσια του προγράμματος Χαρτογράφησης και Αλληλούχησης του Γονιδιώματος του Ανθρώπου, το εργαστήριό μας την τελευταία δεκαετία είχε ως κύριο ερευνητικό πρόγραμμα την γονιδιωματική χαρτογράφηση του χρωμοσώματος 10. Στα επόμενα κεφάλαια, μετά από μία σύντομη αναδρομή στην πορεία του HGP, περιγράφεται η διαδικασία κατασκευής των γενετικών και φυσικών χαρτών, η δημιουργία των χρωμοσωμικών χαρτών που χρησιμοποιήθηκαν για την αλληλούχηση, τα βασικά συμπεράσματα από την εκ παραλλήλου αλληλούχηση των γονιδιωμάτων οργανισμών-μοντέλων και εισαγωγικά στοιχεία για τα γονίδια και τα νοσήματα του χρωμοσώματος 10.

1.2. Ιστορική αναδρομή

Η πρώτη συζήτηση για το ενδεχόμενο της αλληλούχησης του ανθρώπινου γονιδιώματος πραγματοποιήθηκε το 1985 σε επιστημονικό συνέδριο που συγκλήθηκε από τον Robert Sinsheimer από το UCSC (University of California, Santa Cruz). Καθώς είχε περάσει μόλις μία δεκαετία από την ανάπτυξη της τεχνικής της αλληλούχησης του DNA από τους Alan Maxam και Walter Gilbert και τον Fredrick Sanger, η ιδέα αυτή αντιμετωπίστηκε από πολλούς ως ακραία και πρώιμη. Ωστόσο απέκτησε και θερμούς υποστηρικτές οδηγώντας έτσι την επόμενη χρονιά το υπουργείο Ενέργειας των Ηνωμένων Πολιτειών, να λάβει την απόφαση της χρηματοδότησης αυτού του προγράμματος. Επισήμως, το 1988 το Κογκρέσο των ΗΠΑ κήρυξε την έναρξη του Προγράμματος Χαρτογράφησης και Αλληλούχησης του Ανθρώπινου Γονιδιώματος με προϋπολογισμό 200 εκατομμύρια δολάρια ετησίως για τα επόμενα 15 χρόνια. Σύμφωνα με το αρχικό σχέδιο, το πρόγραμμα θα εξελισσόταν σε τρία στάδια.

Η πρώτη φάση (μέχρι το 1993) περιλάμβανε: 1) την κατασκευή των πρωταρχικών γενετικών και φυσικών χαρτών που θα επιτάχυναν την ταυτοποίηση γονιδίων που σχετίζονται με νοσήματα, 2) την ανάλυση των γονιδιωμάτων οργανισμών-μοντέλων (*Escherichia coli*, *Saccharomyces cerevisiae*, *Caenorhabditis elegans*, *Drosophila melanogaster* και *Mus musculus*), 3) την ανάπτυξη νέων τεχνολογιών, αυτοματοποίηση μεθόδων και ανάπτυξη λογισμικού κατάλληλου για την επεξεργασία των αποτελεσμάτων και 4) την εξέταση των ηθικών, νομικών και κοινωνικών επιπτώσεων του προγράμματος (πρόγραμμα ELSI, Ethical, Legal, and Social Issues), (<http://www.genome.gov/10001477>).

Για την επόμενη πενταετή φάση (1993-1998) ο προγραμματισμός περιλάμβανε: 1) την αύξηση της διακριτικής ικανότητας των χρωμοσωμικών χαρτών, 2) την βελτιστοποίηση της τεχνολογίας της αλληλούχησης DNA και της ταυτοποίησης γονιδίων, 3) την αλληλούχηση 80Mb από όλους τους οργανισμούς, 4) την ολοκλήρωση της αλληλούχησης του γονιδιώματος της *E.coli* και του *S.cerevisiae*, του μεγαλύτερου μέρους του *C. elegans* και την έναρξη του αντίστοιχου προγράμματος για την *D. melanogaster* και 5) την χαρτογράφηση 10.000 STSs (Sequence Tag Sites, μοναδιαίες αλληλουχίες-σηματοδότες) του ποντικού (Collins F & Galas D, 1993).

Για το τρίτο στάδιο του προγράμματος (1998-2003) είχε σχεδιαστεί: 1) η ολοκλήρωση της χαρτογράφησης του ανθρώπινου γονιδιώματος, 2) η παραγωγή της ολοκληρωμένης αλληλουχίας για το 1/3 του ανθρώπινου γονιδιώματος μέχρι το τέλος του 2001 και για το υπόλοιπο μέχρι το τέλος του 2003, με ρυθμό παραγωγής 500Mb

ετησίως και με κόστος μικρότερο από 0,25 δολάρια ανά βάση, 3) η χαρτογράφηση 100.000 SNPs (Single Nucleotide Polymorphisms, πολυμορφισμοί ενός νουκλεοτιδίου), 4) η απομόνωση cDNAs πλήρους μήκους, 5) η ολοκλήρωση της αλληλούχησης του γονιδιώματος του *C. elegans* μέχρι το τέλος του 1998 και της *D. melanogaster* έως το 2002 και 6) η παραγωγή της πρώτης έκδοσης της αλληλουχίας του γονιδιώματος του *M. musculus* πριν το 2005 και στη συνέχεια η ολοκληρωμένη έκδοση (Collins FS *et al*, 1998a).

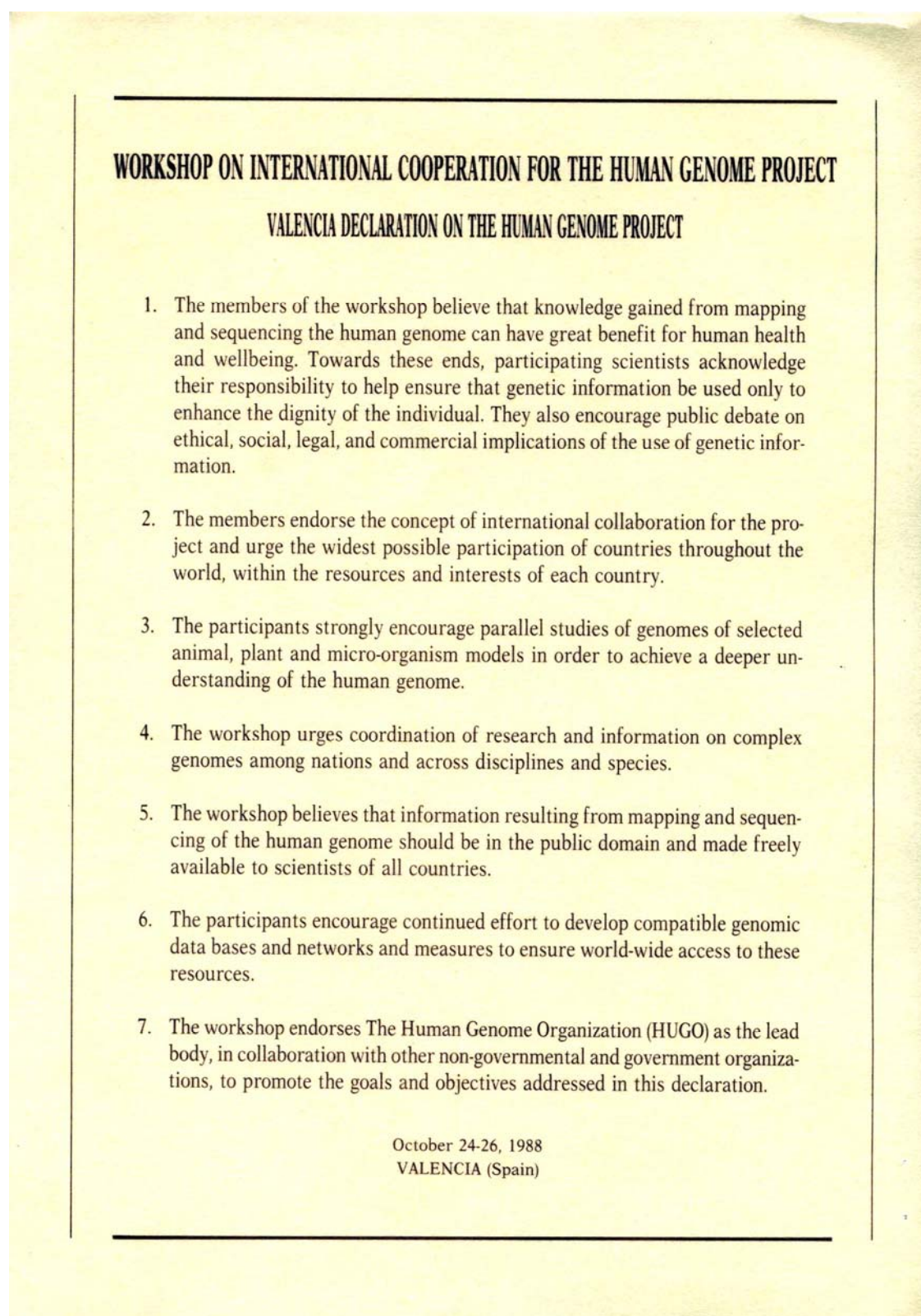
Παρόλο που η έναρξη του προγράμματος έγινε στις ΗΠΑ, εξίσου σημαντικό ρόλο στην πρόοδο του έπαιξε και η Ευρωπαϊκή Ένωση, ξεκινώντας το 1990 το "Human Genome Analysis Programme" (HGAP), η Ιαπωνία αλλά και άλλες χώρες που συνέπραξαν στην πορεία. Έτσι, το 1988 πραγματοποιήθηκε το πρώτο συνέδριο για τον διεθνή, πλέον, συντονισμό της έρευνας και την θέσπιση των αρχών που διέπουν το HGP. Στο συνέδριο, που έλαβαν μέρος επιστήμονες από 24 χώρες, συντάχθηκε το κείμενο για τα καθήκοντα και τις υπευθυνότητες των ατόμων που συμμετέχουν στο πρόγραμμα, γνωστό ως «Διακήρυξη της Valencia» («Valencia Declaration», Εικ. 3) και προτάθηκε ομοφώνως το HUGO (Human Genome Organization) ως επικεφαλής οργανισμός με κύρια ευθύνη για την επιμέλεια των αποτελεσμάτων του Προγράμματος και την προάσπιση της «Διακήρυξης της Valencia». Πρώτος πρόεδρος εκλέχτηκε ο Victor McKusick (Johns Hopkins University, Baltimore, Maryland και «πατέρας» της OMIM). Ένα πολύ σημαντικό γεγονός, στα πλαίσια πάλι της δεοντολογίας του προγράμματος ήταν η σύνταξη του κειμένου των «αρχών της Βερμούδας» (Bermuda principles) το 1996 σε μία συνάντηση ερευνητικών ομάδων από όλο τον κόσμο. Συνοπτικά, αυτές περιλάμβαναν: 1) την άμεση και χωρίς περιορισμούς απελευθέρωση των αποτελεσμάτων της αλληλούχησης που θα προέκυπταν, εφόσον οι αλληλουχίες ήταν μεγαλύτερες από 1kb, εντός 24 ωρών από την παραγωγή τους, 2) την άρνηση δικαιωμάτων ευρεσιτεχνίας και 3) τα κριτήρια πιστότητας προσδιορισμού των αλληλουχιών. Κατά τα επόμενα χρόνια πραγματοποιούνταν περιοδικά παρόμοιες συναντήσεις για τον συντονισμό της συνεργασίας των διάφορων ομάδων (Guyer M, 1998).

Στα αρχικά στάδια του προγράμματος υπήρξε έντονη αντιπαράθεση για το αν πράγματι αξίζει να αλληλουχηθεί ολόκληρο το ανθρώπινο γονιδίωμα. Η μία πλευρά υποστήριζε ότι το κόστος ήταν πολύ υψηλό σε σχέση με το όφελος που θα προέκυπτε, καθώς το ποσοστό των κωδικών περιοχών του γονιδιώματος είναι πολύ μικρό. Η εναλλακτική και οικονομικότερη πρόταση ήταν η στόχευση μόνο των κωδικών περιοχών με τη μορφή των ESTs και cDNAs και η αλληλούχηση μικρότερων γονιδιωμάτων, κατάλληλων οργανισμών-μοντέλων όπως του *Fugu*

rubripes (Brenner S *et al*, 1993). Η άλλη πλευρά όμως υποστήριζε ότι ήταν αδύνατο να εντοπιστούν όλα τα γονίδια στηριζόμενοι μόνο στην τεχνολογία ταυτοποίησης των cDNAs. Επίσης ότι πολλές σημαντικές πληροφορίες θα προκύψουν και από τις μη-κωδικές περιοχές του γονιδιώματος, όπως οι ρυθμιστικές περιοχές και τα διάφορα άλλα δομικά στοιχεία. Παράλληλα, η πλήρης αλληλούχηση θα επέτρεπε την διερεύνηση μιας υψηλότερης τάξης οργάνωσης του γονιδιώματος καθώς και την συγκριτική ανάλυση και τις εξελικτικές σχέσεις των ειδών σε μεγάλη κλίμακα. Τελικά αποφασίστηκε ότι, τουλάχιστον για συγκεκριμένους οργανισμούς, συμπεριλαμβανομένου όμως του ανθρώπου, είναι απαραίτητη η αλληλούχηση ολόκληρου του γονιδιώματός τους.

Το χρονοδιάγραμμα του προγράμματος ακολουθήθηκε αποτελεσματικά και σε αρκετές περιπτώσεις τα αποτελέσματα παρήχθησαν ταχύτερα από το προβλεπόμενο (Collins FS, 1995; Collins FS *et al*, 1998a). Το 1993 η πρόοδος στην χαρτογράφηση του γονιδιώματος οδήγησε στην αναδιαμόρφωση των στόχων. Το νέο σχέδιο αφορούσε την επιτάχυνση των επιδιώξεων κυρίως στο πεδίο της ανάπτυξης της τεχνολογίας για την ταυτοποίηση γονιδίων και της χαρτογράφησης (Collins F & Galas D, 1993). Το 1998, ένα ιδιαίτερα σημαντικό γεγονός στην πορεία ανάλυσης του ανθρώπινου γονιδιώματος ήταν η ίδρυση της ιδιωτικής εταιρείας Celera Genomics από τον Craig Venter, σε σύμπραξη με την Perkin-Elmer Corp. Η πρόταση του Venter στόχευε στην αλληλούχηση του ανθρώπινου γονιδιώματος σε 3 χρόνια, όμως με σχετικά περιορισμένη πιστότητα και με σκοπό την κατοχύρωση της αλληλουχίας ως ευρεσιτεχνία και την εμπορική εκμετάλλευση των πληροφοριών. Ο τρόπος προσέγγισης του C.Venter δεν περιελάμβανε χρωμοσωμική/φυσική χαρτογράφηση, αλλά πλήρη αλληλούχηση μικρών τεμαχίων DNA (shotgun αλληλούχηση) και στη συνέχεια στοίχιση όλων των αλληλουχιών αξιοποιώντας ισχυρά υπολογιστικά προγράμματα. Οι βασικές αντιρρήσεις του HGP, εκτός της διαφωνίας για την εμπορική εκμετάλλευση της αλληλουχίας, ήταν στο κατά πόσον είναι εφικτή αυτή η μεθοδολογία λαμβάνοντας υπόψη το πολύ μεγάλο ποσοστό των επαναλαμβανόμενων αλληλουχιών του ανθρώπινου γονιδιώματος. Υπό την «πίεση» της Celera, το 1999 ανακοινώθηκε από το HGP ότι σε πρώτη φάση και πολύ γρηγορότερα, μέχρι το 2001, θα παραγόταν η «πρώτη έκδοση» της αλληλουχίας, καλύπτοντας το 90% του γονιδιώματος και με ανώτατο όριο πιστότητας «διαβάσματος», τις 4-5 φορές ανά νουκλεοτίδιο, λίγο χαμηλότερο από το αρχικώς προταθέν (7-8 φορές). Έτσι και σε συνδυασμό με γενναία αύξηση της χρηματοδότησης, τη νέα γενιά αντιδραστηρίων (φθορίζουσες χρωστικές, πολυμεράσες) και αυτόματων συσκευών (ρομποτικά συστήματα για παρασκευή

δειγμάτων, ηλεκτροφόρηση σε τριχοειδή) επιταχύνθηκε κατά πολύ η απόδοση του HGP.



Εικ. 3: Η διακήρυξη της Valencia

Το αποτέλεσμα ήταν τον Φεβρουάριο του 2001 να δημοσιευθεί η «πρώτη έκδοση» των χρωμοσωμικών χαρτών και της αλληλουχίας του ανθρώπινου γονιδιώματος (Lander ES *et al*, 2001; Εικ. 4). Η αλληλουχία κάλυπτε το 90% της ευχρωματίνης και το μέσο «βάθος» διαβάσματός της ήταν 4,5 φορές. Ταυτόχρονα, την ίδια ημερομηνία δημοσιεύτηκε η αντίστοιχη ανάλυση από την εταιρεία Celera Genomics (Venter JC *et al*, 2001; Εικ. 4). Ωστόσο, το αποτέλεσμα της προσέγγισης της απευθείας αλληλούχησης, στο επίπεδο πολυπλοκότητας και μεγέθους του ανθρώπινου γονιδιώματος, αμφισβητήθηκε και θεωρήθηκε ότι δεν θα ήταν εφικτό χωρίς την ενσωμάτωση των δεδομένων του HGP (ιδίως των χρωμοσωμικών χαρτών) που ήταν ελεύθερα διαθέσιμα στις δημόσιες βάσεις δεδομένων (Waterston RH *et al*, 2002α; Myers EW *et al*, 2002; Green P, 2002). Σταδιακά, η «ολοκληρωμένη αλληλουχία» (finished), που παρέχει κάλυψη των περισσότερων χασμάτων που εμφάνιζε η «πρώτη έκδοση» και που η πιστότητά της υπερβαίνει το 99,999%, άρχισε να δημοσιεύεται το 2004, ένα χρόνο δηλαδή νωρίτερα από τον αρχικό στόχο από το HGP. Η εκτενής περιγραφή των χαρακτηριστικών της πραγματοποιείται μεμονωμένα για κάθε χρωμόσωμα και μέχρι σήμερα (Ιούλιος 2005) έχει δημοσιευθεί η ανάλυση για 16 από τα 24 χρωμοσώματα (Πιν. 1).



Εικ. 4: Τα εξώφυλλα των περιοδικών Nature και Science τον Φεβρουάριο του 2001, που περιγράφουν την ανάλυση της πρώτης έκδοσης της αλληλουχίας του ανθρώπινου γονιδιώματος από το HGP και την εταιρία Celera Genomics αντίστοιχα.

Πιν. 1: Τα χρωμοσώματα για τα οποία έχει ολοκληρωθεί η αλληλούχηση και η ανάλυσή τους

# Χρωμόσωμα	Έτος	Βιβλιογραφία
2	2005	Hillier LW <i>et al</i>
4	2005	Hillier LW <i>et al</i>
5	2004	Schmutz J <i>et al</i>
6	2003	Mungall AJ <i>et al</i>
7	2003	Hillier LW <i>et al</i>
9	2004	Humphray SJ <i>et al</i>
10	2004	Deloukas P <i>et al</i>
13	2004	Dunham A <i>et al</i>
14	2003	Heilig R <i>et al</i>
16	2004	Martin J <i>et al</i>
19	2004	Grimwood J <i>et al</i>
20	2001	Deloukas P <i>et al</i>
21	2000	Hattori M <i>et al</i>
22	1999	Dunham I <i>et al</i>
X	2005	Ross MT <i>et al</i>
Y	2003	Skaletsky H <i>et al</i>

1.3. Χαρτογράφηση

Η κατασκευή χρωμοσωμικών χαρτών ήταν ένα σημαντικό βήμα στην ανάλυση του γονιδιώματος καθώς παρέχουν έναν συμβατό τρόπο απεικόνισης, τμηματοποίησης και διάταξης των στοιχείων του. Υπάρχουν δύο κατηγορίες χρωμοσωμικών χαρτών: οι γενετικοί και οι φυσικοί που το κοινό στοιχείο τους είναι ότι και οι δύο αναπαριστάνουν την πιθανή διάταξη των γενετικών/μοριακών δεικτών κατά μήκος του χρωμοσώματος. Ωστόσο, ο γενετικός χάρτης παρέχει μία έμμεση εκτίμηση της φυσικής απόστασης δύο δεικτών ενώ ο φυσικός απεικονίζει την πραγματική τους απόσταση, μετρημένη σε μήκος ζευγών βάσεων (bp).

➤ Γενετική χαρτογράφηση

Οι γενετικοί χάρτες στηρίζονται στην πιθανότητα ανασυνδυασμού μεταξύ δύο δεικτών κατά την μείωση, η οποία, γενικά, είναι ανάλογη της απόστασής τους. Η μονάδα μέτρησης της απόστασης είναι το cM (centiMorgan) που αντιπροσωπεύει την συχνότητα ανασυνδυασμού. Έτσι, δύο δείκτες απέχουν 1cM αν σε ανάλυση 100 μειώσεων έχει παρατηρηθεί σε μία από αυτές, ανασυνδυασμός μεταξύ τους. Οι αλληλουχίες που χρησιμοποιούνται ως δείκτες πρέπει να μπορούν να προσδιοριστούν αμφιμονοσήμαντα στο γονιδίωμα και να είναι πολυμορφικές στον πληθυσμό, έτσι ώστε να είναι δυνατή η διάκριση και η παρακολούθησή τους από γενεά σε γενεά. Ένα άτομο για να είναι πληροφοριακό για συγκεκριμένο δείκτη, πρέπει να είναι ετερόζυγο γι' αυτόν.

Ο περιοριστικός παράγοντας για την κατασκευή των πρώτων γενετικών χαρτών ήταν η έλλειψη μεγάλου αριθμού χαρακτηρισμένων πολυμορφικών δεικτών.

Αυτό ξεπεράστηκε εν μέρει με την ταυτοποίηση των RFLPs (restriction fragment length polymorphisms, Kan YW & Dozy AM, 1978) που είναι πολυμορφισμοί στο πρότυπο τεμαχισμού του γενωμικού DNA με περιοριστικές ενδονουκλεάσες λόγω νουκλεοτιδικής ποικιλότητας στην αλληλουχία αναγνώρισης, και που θεωρήθηκαν κατάλληλοι για την κατασκευή γενετικού χάρτη (Botstein D *et al*, 1980). Έτσι κατασκευάστηκε ο πρώτος γενετικός χάρτης (Donis-Keller H *et al*, 1987) που στηρίχτηκε σε ανάλυση 403 πολυμορφικών δεικτών εκ των οποίων οι 393 ήταν RFLPs και οι υπόλοιποι VNTRs (βλ. παρακάτω). Ωστόσο, η διακριτική ικανότητα ήταν περιορισμένη λόγω της μειωμένης αφθονίας των RFLPs, της ανομοιογένειας στην κατανομή τους και της περιορισμένης ετεροζυγωτίας που εμφανίζουν (μέγιστη τιμή 50%). Τα RFLPs εμπλουτίστηκαν με νέους, τους μινι-δορυφόρους και τα VNTRs (variable number of tandem repeats, Jeffreys AJ *et al*, 1985; Nakamura Y *et al*, 1987) απλά επαναλαμβανόμενα στη σειρά τμήματα DNA, μήκους 11-60bp, που ορίζονται αμφίπλευρα από μοναδιαίες αλληλουχίες και εμφανίζουν πολλαπλά αλληλόμορφα λόγω ποικιλότητας στον περιεχόμενο αριθμό επαναλήψεων. Η μεγάλη ώθηση στην χρήση των πολυμορφικών δεικτών ήταν το PCR (Saiki RK *et al*, 1985) που εφαρμόστηκε ευρέως στην ανάλυση μιας νέας κατηγορίας δεικτών, των μικρο-δορυφόρων που, όπως και οι μινι-δορυφόροι, αποτελούν συνεχόμενες επαναλήψεις απλών μικρών αλληλουχιών 2-4 νουκλεοτιδίων (π.χ. GT, GATA). Το συνολικό μήκος της επαναλαμβανόμενης περιοχής είναι μικρότερο από 100bp και καθώς φέρουν εκατέρωθεν των επαναλήψεων, μοναδιαίες αλληλουχίες είναι δυνατός ο πολλαπλασιασμός τους με PCR. Οι μικροδορυφόροι εμφανίζουν υψηλή πολυμορφικότητα, είναι άφθονοι και σχετικά ομοιόμορφα κατανομημένοι στο γονιδίωμα. Η ανάπτυξη της τεχνολογίας μαζικού προσδιορισμού των μικροδορυφόρων ήταν ένα βήμα-κλειδί προς την κατασκευή πυκνών, υψηλής ευκρίνειας γενετικών χαρτών.

Με αυτά τα εργαλεία, σε μικρό χρονικό διάστημα κατασκευάστηκε μία σειρά χρωμοσωμικών γενετικών χαρτών για όλο το γονιδίωμα με αυξανόμενη πυκνότητα δεικτών και συνεπώς, διακριτική ικανότητα (π.χ. Weissenbach J *et al*, 1992; Spurr NK *et al*, 1994; Gyapay G *et al*, 1994; Murray JC *et al*, 1994; Dib C *et al*, 1996; Kong A *et al*, 2002). Παράλληλα, κατασκευάστηκαν γενετικοί χάρτες για τον ποντικό (Dietrich WF *et al*, 1996; Rhodes M *et al*, 1998).

Οι χάρτες αυτοί, εκτός του ότι αποτελούν απαραίτητα εργαλεία για την χαρτογράφηση γενετικών τόπων που συνδέονται με μεντελικές ασθένειες μέσω γονοτυπικής ανάλυσης σε οικογένειες με πάσχοντα μέλη, χρησίμευσαν και στην κατασκευή των φυσικών χαρτών των χρωμοσωμάτων του ανθρώπου. Επίσης, πριν από την αλληλούχηση του γονιδιώματος χρησιμοποιήθηκαν επικουρικά και για την

εκτίμηση του μεγέθους και της φυσικής απόστασης στα χρωμοσώματα. Καθώς το ανθρώπινο γονιδίωμα είναι ~3000cM και αποτελείται από ~3000Mb, το 1cM αντιστοιχεί κατά προσέγγιση σε 1 Mb. Βέβαια αυτή η αντιστοίχιση βασίζεται στην παραδοχή ότι ο ανασυνδυασμός είναι τυχαίος, δηλ. ότι η πιθανότητα ανασυνδυασμού είναι η ίδια καθ' όλο το μήκος κάθε χρωμοσώματος, προϋπόθεση που δεν ισχύει για όλες τις περιοχές. Έτσι, υπάρχουν περιοχές του γονιδιώματος που εμφανίζουν πολύ χαμηλό ποσοστό ανασυνδυασμού, οι «έρημοι» και άλλες πολύ υψηλό, οι «ζούγκλες» (Yu A *et al*, 2001) ενώ το μέγιστο ποσοστό παρατηρείται στις υποτελομερικές περιοχές (Kong A *et al*, 2002; Jensen-Seaman MI *et al*, 2004).

Ένας γενετικός χάρτης νέας γενιάς είναι αυτός που αξιοποιεί ως δείκτες SNPs, πολυμορφισμούς δηλαδή ενός νουκλεοτιδίου οι οποίοι αντιπροσωπεύουν το 90% των νουκλεοτιδικών πολυμορφισμών του ανθρώπου (Collins FS *et al*, 1998b; International HapMap Consortium, 2003).

➤ **Φυσική χαρτογράφηση**

Υπάρχουν τρία είδη φυσικών χαρτών: οι κυτταρογενετικοί, των σωματικών υβριδίων και οι χάρτες αλληλοεπικαλυπτόμενων κλώνων/αλληλουχίας που διαφέρουν μεταξύ τους, στη διακριτική ικανότητα που παρέχει ο καθένας. Το πρότυπο ζώνωσης των χρωμοσωμάτων μπορεί να θεωρηθεί η πιο πρώιμη και με την μικρότερη διακριτική ικανότητα, μορφή του φυσικού χάρτη, ενώ η πλήρης νουκλεοτιδική αλληλουχία του γονιδιώματος αντιπροσωπεύει το άλλο άκρο.

Κυτταρογενετική χαρτογράφηση

Η κυτταρογενετική χαρτογράφηση και η δημιουργία καρυοτυπικού προτύπου, αποτέλεσε τον ιστορικά παλαιότερο τρόπο χαρακτηρισμού των ανθρώπινων χρωμοσωμάτων και υπήρξε η βάση για τον αρχικό χαρακτηρισμό διάφορων χρωμοσωμικών ανωμαλιών (διπλασιασμών, ελλείψεων, μετατοπίσεων) που σχετίζονται με κλινικούς φαινότυπους. Κατ' αυτήν, παρασκευάσματα μεταφασικών χρωμοσωμάτων χρωματίζονται με Giemsa και έτσι είναι ορατό ένα πρότυπο ζώνωσης χαρακτηριστικό για κάθε χρωμόσωμα (Craig JM & Bickmore WA, 1993). Οι ανοιχτόχρωμες R (reverse) και οι σκούρες G (Giemsa) ζώνες που σχηματίζονται εξαρτώνται από την νουκλεοτιδική σύσταση και έχουν συσχετιστεί με διακυμάνσεις στην γονιδιακή πυκνότητα, τον χρόνο αντιγραφής του DNA, την πυκνότητα των επαναλαμβανόμενων αλληλουχιών και την συμπύκνωση της χρωματίνης (Holmquist GP, 1992; Craig JM & Bickmore WA, 1993). Σε επίπεδο γονιδιώματος, με αυτήν την χρώση, είναι ορατές ~850 διακριτές ζώνες, τιμή που περιορίζει και την διακριτική ικανότητα της μεθόδου στο 1-5Mb (Bickmore WA & Sumner AT, 1989).

Επανάσταση στην κυτταρογενετική χαρτογράφηση αποτέλεσε η χρήση σημασμένων ανιχνευτών και η ανίχνευσή τους στα χρωμοσώματα είτε με αυτοραδιογραφία είτε με τεχνικές φθορισμού (Fluorescent in situ Hybridization, FISH). Επιπλέον η FISH εφαρμόστηκε και σε μεσοφασικούς πυρήνες όπου, παρόλο που δεν είναι ορατή η ζώνωση, η λιγότερο συμπυκνωμένη δομή της χρωματίνης επιτρέπει αυξημένη διακριτική ικανότητα της τάξεως του 0,1-1Mb, μεταξύ δύο ανιχνευτών (Wilke CM *et al*, 1994). Η διακριτική ικανότητα γίνεται ακόμη μεγαλύτερη όταν το μεσοφασικό DNA υποστεί επιπλέον μηχανική επεξεργασία (fiber FISH) και προσεγγίζει τις 5kb (Heiskanen M *et al*, 1994), οπότε και επικαλύπτει τα όρια της φυσικής χαρτογράφησης με κλωνοποιημένο DNA σε φορείς βακτηριοφάγου λ.

Χάρτες κυτταρικών υβριδίων

Η κατασκευή αυτών των χαρτών βασίστηκε στην ανάπτυξη της μεθοδολογίας για την παρασκευή υβριδίων σωματικών κυττάρων (Goss SJ & Harris H, 1975) που περιείχαν ολόκληρα χρωμοσώματα ή τμήματα χρωμοσωμάτων (Cox DR *et al*, 1990). Τα υβρίδια προκύπτουν από την σύντηξη σωματικών κυττάρων του ανθρώπου (δότες) που έχουν ακτινοβοληθεί με δόσεις ακτίνων X που οδηγούν στην θραύση των χρωμοσωμάτων με κυττάρων τρωκτικού (δέκτες) που φέρουν έναν πρωτεϊνικό δείκτη επιλογής (selection marker).

Όπως και στους γενετικούς χάρτες, έτσι και σε αυτούς, όσο πλησιέστερα είναι δύο δείκτες τόσο μικρότερη είναι η πιθανότητα να δημιουργηθεί χρωμοσωμική θραύση μεταξύ τους. Έτσι, χρησιμοποιώντας μία ομάδα κυτταρικών υβριδίων που περιέχει μία σειρά αλληλοεπικαλυπτόμενων περιοχών του ίδιου χρωμοσώματος ή του συνόλου του γονιδιώματος, ελέγχεται η παρουσία ή η απουσία ενός δείκτη σε κάθε κυτταρικό κλώνο, καθιστώντας εφικτή τη διάταξη των δεικτών κατά μήκος του χρωμοσώματος. Η μονάδα μέτρησης είναι το 1cR (centiRay) που αντιστοιχεί σε 1% πιθανότητα θραύσης μεταξύ δύο δεικτών, χρησιμοποιώντας μία συγκεκριμένη δόση ακτινοβολίας. Συνεπώς η συσχέτιση μεταξύ των μονάδων cR και της φυσικής απόστασης σε bp ποικίλλει μεταξύ διαφορετικών ομάδων υβριδίων και καθορίζεται κατά αναλογία. Όσο υψηλότερη ακτινοβολία χρησιμοποιείται τόσο μεγαλύτερος είναι ο αριθμός των θραυσμάτων που δημιουργούνται, με αποτέλεσμα την αύξηση της διακριτικής ικανότητας.

Σε αντίθεση με τον γενετικό χάρτη στον οποίο ένας γενετικός δείκτης, όταν είναι σε ομοζυγωτία, δεν είναι πληροφοριακός, οι δείκτες των κυτταρικών υβριδίων είναι είτε θετικοί (περιέχονται) είτε αρνητικοί (δεν περιέχονται) για ένα συγκεκριμένο τμήμα DNA. Έτσι κάθε μοναδιαία αλληλουχία του γονιδιώματος που μπορεί να πολλαπλασιαστεί με PCR μπορεί να αξιοποιηθεί σαν δείκτης (STS, EST,

πολυμορφικός δείκτης). Στην τεχνική αυτή βασίστηκε η κατασκευή ομάδων κυτταρικών υβριδίων για όλο το γονιδίωμα, π.χ. Genebridge4 (GB4, Gyapay G *et al*, 1996), Stanford G3 panel (Stewart EA *et al*, 1997), TNG panel (<http://www-shgc.stanford.edu/resources/humanrhmap.html>) που διαφέρουν μεταξύ τους στην δόση της ακτινοβολίας που χρησιμοποιήθηκε για την κατασκευή τους και συνεπώς και στη διακριτική ικανότητά τους. Βάσει αυτών, κατασκευάστηκαν χάρτες του γονιδιώματος υψηλής διακριτικής ικανότητας (Hudson TJ *et al*, 1995; Schuler GD *et al*, 1996; Deloukas P *et al*; 1998, ηλεκτρονική ανασκόπηση το 1999, <http://www.ncbi.nih.nlm.gov/genemap99>). Οι χάρτες WI-RH και SHGC-G3 περιλαμβάνουν 6.193 και 11.458 STSs χαρτογραφημένα στα Genebridge4 (GB4) και Stanford G3 panel αντίστοιχα. Ο GeneMap99 περιλαμβάνει 41,664 ESTs που αντιστοιχούν σε 30181 cDNAs και ~700 πολυμορφικούς δείκτες βάσει των οποίων γίνεται η στοίχιση με τους γενετικούς χάρτες.

Χάρτες γενωμικών κλώνων

Ο χάρτης αυτός αποτελείται από μία ή περισσότερες ομάδες διευθετημένων αλληλοεπικαλυπτόμενων κλώνων ή αλληλουχιών, που στο σύνολό τους συνιστούν ένα contig, όρος που πρωτοχρησιμοποιήθηκε από τον R. Staden (Staden R, 1980). Οι αποστάσεις εκφράζονται σε ζεύγη βάσεων (bp).

Οι φυσικοί χάρτες μεγάλων γονιδιωμάτων, όπως του ανθρώπου, κατασκευάστηκαν αρχικά από κλώνους YAC (Burke DT *et al*, 1987) και αργότερα από PACs (Ioannou PA *et al*, 1994) και BACs (Shizuya H *et al*, 1992).

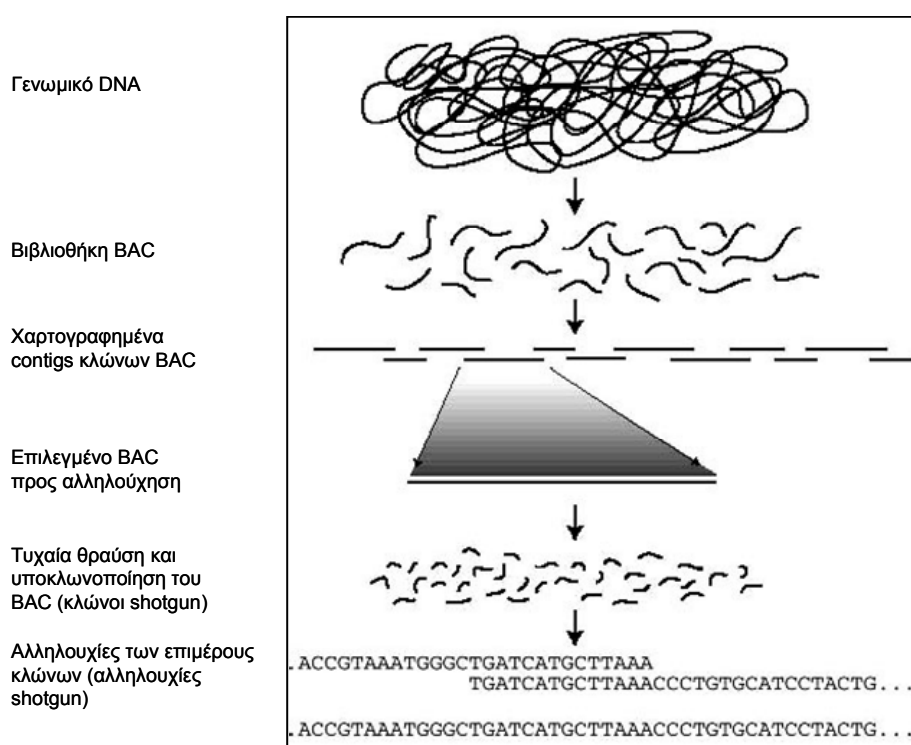
Το κύριο πλεονέκτημα της χρήσης κλώνων YAC είναι το μεγάλο μέγεθος των ενθεμάτων τους (μέχρι 1Mb) που επιτρέπει την κάλυψη μεγάλων περιοχών με σχετικά λιγότερους κλώνους (Chumakov IM *et al*, 1995). Οι πρώτοι χάρτες κατασκευάστηκαν σε περιοχές που σχετίζονται με γενετικά νοσήματα, όπως στην περιοχή του γονιδίου CFTR (Green ED & Olson MV, 1990), της δυστροφίνης (Coffey AJ *et al*, 1992), της φραταξίνης (Wilkes D *et al*, 1991) και αργότερα εκτείνονταν σε ολόκληρα χρωμοσώματα, όπως στο χρωμόσωμα 3 (Gemmill RM *et al*, 1995), το 7 (Bouffard GG *et al*, 1997), το 16 (Doggett NA *et al*, 1995), το 21 (Chumakov I *et al*, 1992; Nizetic D *et al* 1994), το 22 (Collins JE *et al*, 1995), το X (Nagaraja R *et al*, 1997) και το Y (Foote S *et al*, 1992) ή κάλυπταν όλο το γονιδίωμα (Cohen D *et al*, 1993; Chumakov IM *et al*, 1995; Hudson TJ *et al*, 1995). Ωστόσο, οι κλώνοι YAC παρουσιάζουν βασικά μειονεκτήματα που είναι η δομική αστάθεια (ελλείμματα και αναδιατάξεις), ο χιμαιρισμός (Green ED *et al*, 1991; Nagaraja R *et al*, 1994; Selleri L *et al*, 1992) και η δυσκολία στην απομόνωση του DNA και στον διαχωρισμό του από το DNA του ξενιστή (χρωμοσώματα σακχαρομύκητα).

Με την ανάπτυξη των φορέων PAC και BAC που πολλαπλασιάζονται σε βακτήρια, οι παραπάνω δυσκολίες ξεπεράστηκαν. Το μέσο μέγεθος των ενθεμάτων τους είναι ~200kb, δηλαδή ~5 φορές μεγαλύτερο από αυτό των κοσμιδίων (Collins J & Hohn B, 1978) και των φοσμιδίων (Kim UJ *et al*, 1992). Επιπλέον, οι διαθέσιμες βιβλιοθήκες PAC και BAC του ανθρώπινου γονιδιώματος παρέχουν κάλυψη κατά περίπου 60 φορές (<http://www.chori.org/bacpac/>). Στην τελική φάση του προγράμματος χαρτογράφησης του ανθρώπινου γονιδιώματος χρησιμοποιήθηκαν ως επί το πλείστον τα BACs, λόγω του μεγαλύτερου μεγέθους των ενθεμάτων τους σε σχέση με τα PACs και της αυξημένης σταθερότητάς τους. Αυτή οφείλεται στο ότι το σύστημα βασίζεται στον παράγοντα F της *E.coli* με αποτέλεσμα ο αριθμός των αντιγράφων του πλασμιδίου να είναι 1-2 ανά κύτταρο, γεγονός που ελαχιστοποιεί την δυνατότητα ανασυνδυασμού μεταξύ των ενθεμάτων.

1.4. Αλληλούχηση του ανθρώπινου γονιδιώματος

Η στρατηγική που ακολουθήθηκε από το HGP ήταν η χαρτογράφηση να προηγηθεί της αλληλούχησης. Η προσέγγιση αυτή ονομάζεται «ιεραρχική shotgun αλληλούχηση» και περιλάμβανε την δημιουργία χαρτών αλληλοεπικαλυπτόμενων κλώνων BAC, υψηλής πιστότητας, από τους οποίους επιλέγονταν αντιπροσωπευτικοί κλώνοι για τον προσδιορισμό της αλληλουχίας DNA. Η αλληλούχηση των επιλεγμένων κλώνων έγινε με τη μέθοδο shotgun, ορολογία που πρωτοχρησιμοποιήθηκε από τον F. Sanger. Κατά αυτήν, το DNA ενός συγκεκριμένου κλώνου θραύεται τυχαία, υποκλωνοποιείται και προσδιορίζεται η αλληλουχία κάθε κλώνου. Από την στοίχιση των αλληλουχιών των επιμέρους κλώνων, αναπαράγεται η αλληλουχία ολόκληρου του BAC (Εικ. 5).

Αντίθετα η στρατηγική που ακολουθήθηκε από την εταιρία Celera Genomics ήταν της shotgun αλληλούχησης σε επίπεδο γονιδιώματος. Γενικά, κατ' αυτήν ολόκληρο το γονιδίωμα θραύεται τυχαία σε μικρά κομμάτια, τα οποία υποκλωνοποιούνται σε πλασμίδια. Στη συνέχεια προσδιορίζεται η αλληλουχία των ανασυνδυασμένων μορίων και από την στοίχιση των επιμέρους αλληλουχιών, προκύπτει η αλληλουχία του γονιδιώματος. Αυτή η στρατηγική εξαλείφει το στάδιο της φυσικής χαρτογράφησης που περιλαμβάνεται στην μέθοδο του HGP. Η shotgun αλληλούχηση σε επίπεδο γονιδιώματος είχε χρησιμοποιηθεί επιτυχώς σε μικρότερα απλούστερα γονιδιώματα όπως σε ιούς, βακτήρια και την *D. melanogaster*. Ωστόσο, το ανθρώπινο γονιδίωμα με μήκος ~3Gb ήταν τουλάχιστον κατά 25 φορές μεγαλύτερο από οποιοδήποτε άλλο γονιδίωμα είχε προσδιοριστεί προηγουμένως.



Εικ. 5: Ιεραρχική shotgun αλληλούχηση. Ανατύπωση από Lander ES *et al*, 2001

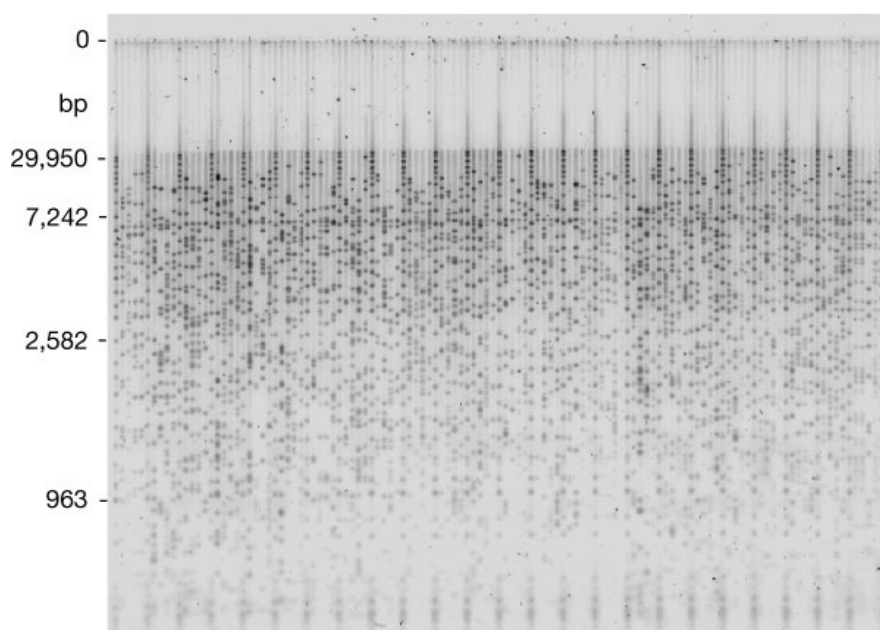
Επιπλέον, ποσοστό >50% αντιστοιχεί σε επαναλαμβανόμενες αλληλουχίες, τιμή κατά πολύ μεγαλύτερη από αυτήν των γονιδιωμάτων των βακτηρίων όπου το αντίστοιχο ποσοστό είναι 1,5% και της *D. melanogaster* στην οποία οι επαναλαμβανόμενες αλληλουχίες καταλαμβάνουν μόλις το 3% του γονιδιώματός της. Η υπεραντιπροσώπηση των επαναλαμβανόμενων αλληλουχιών στο ανθρώπινο γονιδίωμα ήταν και ο βασικός λόγος που το HGP επέλεξε την ιεραρχική αλληλούχηση. Τέτοιες επαναλαμβανόμενες αλληλουχίες είναι αυτές που προέρχονται από μεταθετά στοιχεία, τα ψευδογονίδια, οι απλές επαναλαμβανόμενες αλληλουχίες (SSRs), τα κεντρομερή, τα τελομερή και οι διπλασιασμοί, που περιλαμβάνουν μεγάλα τμήματα, μεγέθους 50-500kb με πολύ υψηλή νουκλεοτιδική ταυτότητα (98-99,9%). Όλες οι κατηγορίες των επαναλαμβανόμενων αλληλουχιών και ιδιαίτερα οι διπλασιασμοί περιπλέκουν την στοίχιση της γενωμικής αλληλουχίας. Με την προσέγγιση της ιεραρχικής αλληλούχησης, η στοίχιση των νουκλεοτιδικών αλληλουχιών περιορίστηκε στο επίπεδο του ενός κλώνου, μειώνοντας στο ελάχιστο δυνατό, τις ψευδείς στοίχισεις. Επιπλέον, με την προσέγγιση του HGP ήταν δυνατή η επιβεβαίωσή της σωστής στοίχισης των επιμέρους αλληλουχιών, επειδή είχε ήδη καθοριστεί το μέγεθος των ενθεμάτων των κλώνων, μέσω τεμαχισμού τους. Επίσης, μειώθηκε στο ελάχιστο η πλεονάζουσα αλληλούχηση καθώς ο φυσικός χάρτης επέτρεψε την επιλογή αντιπροσωπευτικών, αξιόπιστα χαρτογραφημένων κλώνων με

την ελάχιστη επικάλυψη. Όσον αφορά το κόστος των δύο εναλλακτικών προσεγγίσεων εκτιμάται ότι αυτό είναι συγκρίσιμο (Lander ES *et al*, 2001). Η ιεραρχική στρατηγική έχει υψηλότερο κόστος στο στάδιο της φυσικής χαρτογράφησης ενώ η shotgun αλληλούχηση σε επίπεδο γονιδιώματος απαιτεί μεγαλύτερη προσπάθεια και έξοδα στο τελικό στάδιο, της παραγωγής της ολοκληρωμένης αλληλουχίας. Τέλος, η ιεραρχική αλληλούχηση είναι συμβατή με ένα ερευνητικό πρόγραμμα στο οποίο συμμετέχουν πολλά διαφορετικά εργαστήρια γιατί κάνει ευκολότερη την κατανομή των διαφόρων πειραμάτων και υπευθυνότητων.

➤ **Φυσικός χάρτης του ανθρώπινου γονιδιώματος**

Η μέθοδος που ακολουθήθηκε από το HGP για την κατασκευή του χάρτη των κλώνων BAC για τα περισσότερα χρωμοσώματα, περιλάμβανε δύο διαφορετικές αλλά συμπληρωματικές προσεγγίσεις: α) το πρότυπο τεμαχισμού τους σε επίπεδο γονιδιώματος και β) τον καθορισμό του περιεχομένου τους σε STSs.

Το πρότυπο τεμαχισμού των BACs χρησιμοποιήθηκε για τον προσδιορισμό του βαθμού αλληλοεπικάλυψής τους. Κλώνοι που προέρχονται από την ίδια περιοχή του γονιδιώματος, παράγουν πρότυπα τεμαχισμού που εμφανίζουν κοινά τεμάχια, το ποσοστό των οποίων είναι ενδεικτικό του βαθμού αλληλοεπικάλυψης. Η αλληλοεπικάλυψη κατά μήκος πολυάριθμων κλώνων χρησιμοποιείται στη συνέχεια, για να διαταχθούν οι κλώνοι σε contigs. Ο μαζικός τεμαχισμός όλων των κλώνων BAC της βιβλιοθήκης RPCI-11 (~300.000 κλώνοι) ξεκίνησε τον Οκτώβριο του '98 (McPherson JD *et al*, 2001). Ο τεμαχισμός γινόταν με το ένζυμο *HindIII* και τα μεγέθη των ενθεμάτων τους υπολογίζονταν μετά από ηλεκτροφόρηση πηκτώματος αгарόζης (Εικ. 6). Βάσει του μέσου μεγέθους των ενθεμάτων και του μεγέθους του ανθρώπινου γονιδιώματος, ο τεμαχισμός αυτού του αριθμού των κλώνων οδηγεί σε κάλυψη του γονιδιώματος κατά 15 φορές. Οι περισσότεροι κλώνοι προέρχονταν από τα τρία πρώτα μέρη της βιβλιοθήκης RPCI-11. Η συλλογή συμπληρώθηκε αργότερα με κλώνους από τις βιβλιοθήκες RPCI-13 και CalTech D. Ο ρυθμός τεμαχισμού των κλώνων ήταν περίπου 20.000/βδομάδα. Όλα τα δεδομένα που προέκυπταν καταχωρούνταν στην βάση δεδομένων FPC. Η συναρμολόγηση των contigs από τα δεδομένα που προκύπτουν από τον τεμαχισμό τους στηρίζεται σε πολύπλοκη στατιστική ανάλυση και γίνεται με το πρόγραμμα FPC (FingerPrinted Contigs, Soderlund C *et al*, 2000), που για την ώρα είναι το μοναδικό αυτοματοποιημένο λογισμικό κατασκευής φυσικών χαρτών.



Εικ. 6: Ενδεικτική ηλεκτροφόρηση σε πήκτωμα αγαρόζης των προϊόντων τεμαχισμού των κλώνων BAC. Ο τεμαχισμός πραγματοποιήθηκε με *HindIII*, ενώ ανά πέντε δείγματα παρεμβάλλεται δείκτης μοριακών μεγεθών (ανατύπωση από McPherson JD *et al*, 2001).

Το περιεχόμενο των κλώνων σε STSs, καθορίστηκε με υβριδοποίηση γενωμικών βιβλιοθηκών κλώνων BAC (κυρίως της RCPI-11) με ομάδες STSs. Αυτά ήταν πολυμορφικοί μικροδορυφόροι από γενετικούς χάρτες, ESTs από τον GeneMap99, διάφορα STSs από την GDB και STSs που προέκυψαν μετά από διαχωρισμό ροής (flow sorting) των χρωμοσωμάτων. Σταδιακά και καθώς ολοκληρώνονταν οι αλληλουχίες των κλώνων, το περιεχόμενό τους σε γνωστούς δείκτες γινόταν με άμεση διερεύνηση της αλληλουχίας τους με ηλεκτρονικό PCR (e-PCR). Αυτό πραγματοποιείται με λογισμικό που ανιχνεύει σε μία αλληλουχία, την ύπαρξη ενός STS, ταυτοποιώντας τους αντίστοιχους εκκινητές του δείκτη στον σωστό προσανατολισμό και σε σωστή απόσταση μεταξύ τους (Schuler GD, 1998).

Οι δύο προσεγγίσεις που περιγράφηκαν, πραγματοποιούνται ταυτόχρονα και αλληλοσυμπληρώνονται. Το πρότυπο τεμαχισμού των κλώνων δεν επαρκεί για να στηριχτεί ο φυσικός χάρτης γιατί το ανθρώπινο γονιδίωμα περιέχει πολλές επαναλαμβανόμενες αλληλουχίες που μπορεί με τον τεμαχισμό των κλώνων να παράγουν ζώνες ταυτόσημου μεγέθους οδηγώντας σε ψευδείς αλληλοεπικαλύψεις, παρόλο που οι στοιχίσεις των κλώνων πραγματοποιούνται με μεγάλη αυστηρότητα. Από την άλλη, το περιεχόμενο των κλώνων σε STSs δεν επαρκεί γιατί η κατανομή των δεικτών δεν είναι ομοιόμορφη κατά μήκος των χρωμοσωμάτων και υπάρχουν περιοχές που χαρακτηρίζονται από χαμηλή πυκνότητα. Με συνδυασμό και επεξεργασία των πληροφοριών που προκύπτουν από τις δύο προσεγγίσεις,

ελέγχεται η διάταξη των κλώνων μέσα στα contigs και των contigs κατά μήκος των χρωμοσωμάτων.

Μετά τον σχηματισμό και την χαρτογράφηση των αρχικών contigs, αυτά επεκτάθηκαν και συνενώθηκαν. Αρχικά, αυτό πραγματοποιήθηκε με χρωμοσωμικό περπάτημα κατά το οποίο χρησιμοποιήθηκαν οι αλληλουχίες των άκρων των BAC που χαρτογραφούνται στα άκρα των contigs για επαναληπτική διερεύνηση των βιβλιοθηκών και την ταυτοποίηση επιπλέον κλώνων. Στην περίπτωση αυτή συγκρίνονταν τα πρότυπα τεμαχισμού των υποψήφιων κλώνων και εφόσον συμφωνούσαν και τα αποτελέσματα του περιεχομένου τους σε STSs, πραγματοποιούνταν οι ενοποιήσεις των contigs. Αργότερα, με την πρόοδο της αλληλούχησης, η ενοποίηση γινόταν με ταυτοποίηση της αλληλοεπικάλυψης ακριανών κλώνων γειτονικών contigs σε επίπεδο αλληλουχίας. Από τα contigs επιλέγονταν αντιπροσωπευτικοί κλώνοι, οι οποίοι συνιστούν το ελάχιστο αλληλοεπικαλυπτόμενο μονοπάτι, για προσδιορισμό της αλληλουχίας τους. Η θέση αυτών επιβεβαιωνόταν επιπλέον και με ανάλυση FISH.

Στην έκδοση του γονιδιώματος του 2001 (Lander ES *et al*, 2001) ο φυσικός χάρτης αποτελούνταν από 1246 contigs και περιείχε 147821 χάσματα. Αυτά ήταν τριών κατηγοριών: 1) μέσα σε έναν κλώνο επειδή δεν έχει ολοκληρωθεί η αλληλούχηση του (145514 τέτοια χάσματα εκτιμώμενου μεγέθους ~80Mb συνολικά ή 3% της ευχρωματίνης), 2) μέσα στο ίδιο «contig κλώνων» (~4076, συνολικού μεγέθους ~150Mb, ή 5% της ευχρωματίνης) και 3) τα χάσματα μεταξύ διαφορετικών «contigs κλώνων» που αποτελούσαν ~4% της ευχρωματίνης. Συνολικά είχαν αλληλουχηθεί 29298 αλληλοεπικαλυπτόμενοι κλώνοι, κυρίως BACs, με συνολικό μήκος 4,26Gb. Η αλληλουχία κατά το ¼ ήταν «ολοκληρωμένη» (finished), βάσει των εξής κριτηρίων: α) λιγότερες από 1 στις 10.000 βάσεις να είναι λάθος, β) περισσότερο από 95% της ευχρωματίνης να έχει αλληλουχηθεί και γ) τα χάσματα να είναι μικρότερα από 150kb το καθένα (Collins FS *et al*, 1998a). Βάσει του μεγέθους των contigs, των χασμάτων και της ετεροχρωματίνης, το μέγεθος του γονιδιώματος εκτιμήθηκε σε 3,2Gb και της ευχρωματίνης σε 2,9Gb.

Η διαδικασία ολοκλήρωσης του χάρτη και της αλληλουχίας, που πραγματοποιήθηκε μετά το 2001, περιλάμβανε εκτός από την επιβεβαίωση των αλληλοεπικαλύψεων των κλώνων, την γεφύρωση των χασμάτων. Αυτή έγινε με χρωμοσωμικό περπάτημα είτε πειραματικά είτε *in silico*. Κατά την πειραματική προσέγγιση γινόταν εξαντλητική διερεύνηση διαφορετικών γενωμικών βιβλιοθηκών με STSs από τα άκρα των contigs ή με αλληλουχίες που πιθανώς χαρτογραφούνταν μέσα στα χάσματα (π.χ. STSs, mRNAs και αλληλουχίες του ποντικού από συνταινικές περιοχές). Επίσης, πραγματοποιούνταν συγκρίσεις *in silico* με την βάση

δεδομένων των αλληλουχιών των άκρων των BACs (Zhao S, 2000) για ταυτοποίηση αλληλοεπικαλύψεων. Κάποια χάσματα δεν ήταν δυνατόν να καλυφθούν είτε γιατί στάθηκε αδύνατη η ανίχνευση κάποιου κλώνου είτε γιατί παρόλο που ταυτοποιούνταν διάφοροι κλώνοι δεν ήταν δυνατή η εξακρίβωση της σχετικής τους αλληλοεπικάλυψης, λόγω της παρουσίας εκτεταμένων, υψηλά επαναλαμβανόμενων αλληλουχιών ή σχεδόν όμοιων μεταξύ τους διπλασιασμών. Στον φυσικό χάρτη ενσωματώθηκαν και οι περισσότερες τελομερικές περιοχές των χρωμοσωμάτων χρησιμοποιώντας ένα ειδικό σύστημα κλωνοποίησης ('half-YACs vectors', Riethman HC *et al*, 2001).

Στην τελική έκδοση του γονιδιώματος (International Human Genome Sequencing Consortium, 2004), ο αριθμός των χασμάτων περιορίστηκε στα 341 από τα οποία τα 33 αντιστοιχούν σε ετεροχρωματίνη και έχουν συνολικό μήκος 198Mb. Τα υπόλοιπα 308 χάσματα που αντιστοιχούν σε ευχρωματίνη έχουν συνολικό μήκος 28Mb και περιλαμβάνουν 35 στα όρια ευχρωματίνης-ετεροχρωματίνης και 273 μέσα στην ευχρωματίνη. Από τα τελευταία, τα 251 είναι χάσματα μεταξύ των contigs των κλώνων. Σημαντικό ποσοστό αυτών βρίσκεται σε περιοχές που περιέχουν πολλαπλούς διπλασιασμούς, στις περικεντρομερικές και στις υποτελομερικές περιοχές. Συνολικά αλληλουχήθηκαν 59208 κλώνοι (ολικό μήκος 5,84Gb). Το ελάχιστο αλληλοεπικαλυπτόμενο μονοπάτι αποτελούνταν από 26720 κλώνους και η ενοποιημένη αλληλουχία από 2,851,330,913bp. Το μέγεθος του γονιδιώματος εκτιμήθηκε σε 3,08Gb, από τα οποία τα 2,88Gb αντιστοιχούν σε ευχρωματίνη.

Η πιστότητα της αλληλουχίας και η κάλυψη του γονιδιώματος που παρέχει η ολοκληρωμένη έκδοση είναι ιδιαίτερα υψηλή (Schmutz J *et al*, 2004). Πλέον, έχει προσδιοριστεί η διάταξη και ο προσανατολισμός αλληλουχιών που σε κάποιες περιοχές του γονιδιώματος ήταν αμφιλεγόμενος, ενώ διορθώθηκαν διάφοροι τεχνητοί διπλασιασμοί τμημάτων που εμφανίζονταν στην πρώτη έκδοση. Για τον έλεγχο της πιστότητας της ολοκληρωμένης αλληλουχίας εφαρμόστηκαν διάφοροι έλεγχοι, που έδειξαν ότι το ποσοστό λάθους ήταν μικρότερο από 0,001% (1 λάθος ανά 100.000bp), μία τάξη μεγέθους καλύτερο από τα πρότυπα που είχαν θεσπιστεί για την «ολοκληρωμένη» αλληλουχία το 1997 (Bermuda standards, <http://www.gene.ucl.ac.uk/hugo/bermuda2.htm>) και αφορούσαν κυρίως αντικαταστάσεις ενός νουκλεοτιδίου και τον αριθμό των συνεχόμενων αντιγράφων επαναλήψεων (Schmutz J *et al*, 2004). Ο έλεγχος επίσης έδειξε ότι οι στοιχίσεις του ελάχιστου αλληλοεπικαλυπτόμενου μονοπατιού των κλώνων ήταν πολύ ακριβείς και ότι συνολικά η γενωμική αλληλουχία περιέχει μόνο 50-100 πιθανά ελλείμματα μέσου μήκους 5kb λόγω λαθών στη διαδικασία των στοιχίσεων ή μεταλλαγών κατά τον πολλαπλασιασμό των γενωμικών κλώνων στα βακτήρια. Επιπλέον, εκτιμήθηκε ότι

το ποσοστό της ευχρωματινικής αλληλουχίας που λείπει από την «ολοκληρωμένη» έκδοση είναι <1%.

1.5. Αλληλούχηση γονιδιωμάτων οργανισμών-μοντέλων

Η ανάλυση και η σύγκριση γονιδιωμάτων από διαφορετικά είδη είναι το αντικείμενο της Συγκριτικής Γονιδιωματικής. Η συγκριτική ανάλυση μπορεί να αποκαλύψει τα γονίδια που είναι απαραίτητα για τη ζωή καθώς και αυτά που καθορίζουν ένα είδος, να διαλευκάνει τους εξελικτικούς μηχανισμούς και τους μηχανισμούς της γονιδιακής ρύθμισης. Επιπλέον, ένα μεγάλο μέρος της γνώσης της λειτουργίας γονιδίων του ανθρώπου στηρίζεται στη γνώση που υπάρχει για τα ορθόλογα γονίδια σε έναν άλλο απλούστερο οργανισμό-μοντέλο.

Παράλληλα με το HGP, εξελίσσονταν και η γενωμική ανάλυση μιας ποικιλίας οργανισμών-μοντέλων. Αρχικά, αναλύθηκαν σχετικά μικρά γονιδιώματα, που εκτός από τις πληροφορίες που αποκάλυψαν για την γενετική σύσταση των αντίστοιχων οργανισμών, παρείχαν και την στρατηγική και τις τεχνικές πάνω στις οποίες βασίστηκε η μελέτη πολυπλοκότερων οργανισμών, συμπεριλαμβανομένου του ανθρώπου. Σταδιακά, η ανάλυση προχώρησε σε πολυπλοκότερα γονιδιώματα και εκτός από τους πέντε οργανισμούς-μοντέλα που είχαν προταθεί από το HGP, επιλέχθηκαν από την επιστημονική κοινότητα επιπλέον οργανισμοί για αλληλούχηση των γονιδιωμάτων τους. Η επιλογή τους βασίστηκε σε συγκεκριμένα κριτήρια όπως π.χ. στη φυλογενετική απόσταση από άλλους οργανισμούς-μοντέλα, στο μέγεθος του γονιδιώματος, στη δυνατότητα μεταλλαξογένεσης και μετασχηματισμού, στη διαθεσιμότητα πειραματικού υλικού κ.λπ. Καθώς υπάρχουν πλέον διαθέσιμα αντιπροσωπευτικά γονιδιώματα από διάφορες φυλογενετικές ομάδες ως αναφορά, η προσέγγιση που κυρίως ακολουθείται για την αλληλούχηση είναι αυτή του πλήρους *shotgun*. Οι μελέτες πραγματοποιούνται με χρήση υπολογιστικών προγραμμάτων που στοιχίζουν και συγκρίνουν πολλαπλές αλληλουχίες.

Το πρώτο γονιδίωμα που αλληλουχήθηκε ήταν του βακτηριοφάγου φX174 (Sanger F *et al*, 1977; 1978) και ακολούθησε αυτό του βακτηριοφάγου λ (Sanger F *et al*, 1982), στον οποίο πρωτοεφαρμόστηκε η προσέγγιση της *shotgun* αλληλούχησης. Η ανάλυση των γονιδιωμάτων των πέντε οργανισμών-μοντέλων (*S. cerevisiae*, *E. coli*, *C. elegans*, *D. melanogaster*, *M. musculus*) που ήταν στους στόχους του HGP, έχει ήδη ολοκληρωθεί. Το 1989 είχε ξεκινήσει η ανάλυση του ***S. cerevisiae***, του πρώτου ευκαρυωτικού οργανισμού με μέγεθος γονιδιώματος 12,5Mb. Στην ανάλυση συμμετείχαν 74 ευρωπαϊκά εργαστήρια και η στρατηγική που ακολουθήθηκε, λόγω της μεγαλύτερης πολυπλοκότητας του γονιδιώματος αυτού σε σχέση με αυτά που είχαν αναλυθεί μέχρι τότε, περιλάμβανε αρχικά την κατασκευή

ενός φυσικού χάρτη γενωμικών κλώνων φάγου λ και κατόπιν την shotgun αλληλούχηση επιλεγμένων κλώνων (Olson MV *et al*, 1986; Goffeay A *et al*, 1996). Καθώς στον σακχαρομύκητα, μόνο το 4% των γονιδίων που κωδικοποιούν πρωτεΐνες περιέχουν ιντρόνια (Goffeay A *et al*, 1996), η ταυτοποίησή τους *in silico* ήταν σχετικά άμεση και οδήγησε στον προσδιορισμό ~5900 γονιδίων. Η μελέτη του οργανισμού αυτού σε ευρεία κλίμακα είναι σε πολύ προχωρημένο στάδιο και γι' αυτήν χρησιμοποιείται μία ποικιλία προσεγγίσεων, όπως τυχαία (Burns N *et al*, 1994; Ross-MacDonald P *et al*, 1999) ή κατευθυνόμενη (Winzeler EA *et al*, 1999; Giaever G *et al*, 2002) στόχευση των γονιδίων και ακόλουθη φαινοτυπική ανάλυση των μεταλλαγμένων κυττάρων, ανάλυση της γονιδιακής έκφρασης με SAGE (serial analysis of gene expression, Velculescu VE *et al*, 1997) και με μικροσυστοιχίες (DeRisi JL *et al*, 1997), ταυτοποίηση των πρωτεϊνών με ηλεκτροφόρηση δύο-διαστάσεων (Futcher B *et al*, 1999; <http://www.ibgc.u-bordeaux2.fr/YPM/>), του υποκυτταρικού εντοπισμού τους (Kumar A *et al*, 2002; Huh WK *et al* 2003), των αλληλεπιδράσεών τους (Uetz P *et al*, 2000) κ.λπ.

Το 1997 ολοκληρώθηκε η αλληλούχηση του γονιδιώματος του βακτηρίου *E. coli* (Blattner FR *et al*, 1997) που, όπως και στον σακχαρομύκητα, βασίστηκε σε φυσικό χάρτη. Το μέγεθος του γονιδιώματός του είναι ~4,6Mb και ποσοστό ~88% καλύπτεται από τα 4288 γονίδια που φέρει που κωδικοποιούν πρωτεΐνες.

Ο πρώτος πολυκύτταρος οργανισμός που αναλύθηκε το γονιδιώμα του ήταν ο νηματώδης *C. elegans* (C. elegans sequencing consortium, 1998). Ο οργανισμός αυτός είναι πλήρως διαυγής σε όλα τα στάδια της ζωής του, επιτρέποντας έτσι την άμεση παρατήρηση των κυτταρικών διαιρέσεων, των μεταναστεύσεων και της διαφοροποίησης. Επίσης, παρόλο που η ανατομία του είναι απλή, τα 959 σωματικά κύτταρα του ενήλικου αντιπροσωπεύουν τους περισσότερους κύριους διαφοροποιημένους κυτταρικούς τύπους. Επιπλέον, είναι το μοναδικό ζώο για το οποίο έχουν χαρακτηριστεί οι κυτταρικές γενεαλογίες από το ζυγωτό ως το ενήλικο καθώς και το πλήρες νευρωνικό δίκτυο (Wood WB, 1988). Το πρόγραμμα ξεκίνησε με την κατασκευή ενός φυσικού χάρτη (Coulson A *et al*, 1991; Coulson A *et al* ; 1988) κοσμιδιακών κλώνων, που ήταν το σύστημα με τα μεγαλύτερα ενθέματα (~40kb) εκείνη την εποχή, και ταυτόχρονη απομόνωση ESTs (Waterston R *et al*, 1992; McCombie WR *et al*, 1992) και cDNAs (Reboul J *et al*, 2003). Το συνολικό μήκος του γονιδιώματος είναι 97Mb και η ανάλυση της αλληλουχίας, που ήταν σαφώς πολυπλοκότερη από αυτήν του σακχαρομύκητα και της *E. coli* λόγω της παρουσίας ιντρονίων, της μικρότερης γονιδιακής πυκνότητας και της ύπαρξης επαναλαμβανόμενων αλληλουχιών, οδήγησε στην ταυτοποίηση ~19.000 γονιδίων που κωδικοποιούν για πρωτεΐνες, περίπου τρεις φορές περισσότερα από αυτά του

σακχαρομύκητα. Στα πλαίσια επιβεβαίωσης των «προβλεπόμενων» γονιδίων μέσω κλωνοποίησης των αντίστοιχων cDNAs, επιβεβαιώθηκε, μέχρι στιγμής, η δομή των 12.000 από αυτά (ORFeome version 1.1, Reboul J *et al*, 2003). Στον *C. elegans* η ανάλυση, σε επίπεδο γονιδιώματος έχει προχωρήσει αρκετά και περιλαμβάνει δεδομένα από μελέτες της γονιδιακής έκφρασης με χρήση μικροσυστοιχιών (Reinke V *et al*, 2000) ή με χρήση γονιδίων-αναφοράς (Hope IA *et al*, 1996; <http://129.11.204.86:591/default.htm>). Επίσης, έχει εφαρμοστεί με επιτυχία σε ευρεία κλίμακα (>80% γονιδίων), σε επίπεδο οργανισμού, η προσέγγιση της γονιδιακής σίγησης μέσω RNAi που δίνει τη δυνατότητα μελέτης του φαινοτύπου, ελλείψει συγκεκριμένου μεταγράφου (Fraser AG *et al*, 2000; Gonczy P *et al*, 2000; Piano F *et al*, 2000; Maeda I *et al*, 2001; Kamath RS *et al*, 2003, Simmer F *et al*, 2003). Επιπλέον, υπάρχουν πληροφορίες για πλήθος πρωτεϊνικών αλληλεπιδράσεων βάσει πειραμάτων των δύο υβριδίων (Walhout AJ *et al*, 2000; Li S *et al*, 2004; <http://vidal.dfci.harvard.edu/>).

Η *D. melanogaster* αποτελεί αντικείμενο έρευνας για σχεδόν έναν αιώνα και συνιστά ένα ιδανικό μοντέλο για γενετικές και αναπτυξιακές μελέτες. Το 2000 δημοσιεύτηκε η πρώτη έκδοση της αλληλουχίας της ευχρωματινικής περιοχής (120Mb) του γονιδιώματος (180Mb). Η προσέγγιση που ακολουθήθηκε ήταν κυρίως αυτή του πλήρους shotgun σε επίπεδο γονιδιώματος, που χρησιμοποιήθηκε εν μέρει ως απόδειξη της δυνατότητας αλληλούχησης και του ανθρώπινου γονιδιώματος με τον ίδιο τρόπο (Adams MD *et al*, 2000). Ωστόσο, το ανθρώπινο γονιδίωμα είναι μεγαλύτερο κατά τουλάχιστον 15 φορές, έχει πολύ μικρότερη γονιδιακή πυκνότητα και περιλαμβάνει σημαντικότερο ποσοστό επαναλαμβανόμενων αλληλουχιών. Στο πρόγραμμα αλληλούχησης συμμετείχαν τα Berkeley, European, και Canadian *Drosophila* Genome Projects που παρείχαν ένα λεπτομερή χρωμοσωμικό χάρτη (Hoskins RA *et al*, 2000) και αλληλουχία μήκους 28Mb ενώ το υπόλοιπο 75% της αλληλουχίας παρήχθη από συνεργασία της εταιρίας Celera Genomics και του Berkeley *Drosophila* Genome Project. Η ανάλυση του γονιδιώματος έδειξε την ύπαρξη ~13.600 γονιδίων που κωδικοποιούν πρωτεΐνες, αριθμός κατά πολύ μικρότερος από αυτόν του *C. elegans*, παρόλο που ο αριθμός των κυττάρων της *D. melanogaster* είναι σχεδόν δεκαπλάσιος. Αυτό δείχνει, όπως και στην περίπτωση του ανθρώπου, ότι η πολυπλοκότητα ενός οργανισμού δεν «χτίζεται» αποκλειστικά από τον αριθμό των γονιδίων. Μετά την ολοκλήρωση της αλληλούχησης, έχει ξεκινήσει και η λειτουργική ανάλυση σε επίπεδο γονιδιώματος. Πλέον, εκτός των φυσικών και τεχνητών μεταλλαγμένων στελεχών που είχαν απομονωθεί και δημιουργηθεί από τους γενετιστές κατά τα προηγούμενα χρόνια, υπάρχει διαθέσιμη πληθώρα στελεχών που φέρουν ενθέσεις μεταθετών στοιχείων σε συγκεκριμένες

θέσεις του γονιδιώματος, που αποτελούν υλικό για τη δημιουργία χρωμοσωμικών ελλειμμάτων. Υπολογίζεται ότι τα στελέχη αυτά καλύπτουν τα 2/3 των γονιδίων, από την άποψη δυνατότητας μεταλλαξογένεσης (Thibault ST *et al*, 2004; Ryder E & Russell S 2003, Bellen HJ *et al* 2004; <http://flypush.imgen.bcm.tmc.edu/tagbase/>). Η διαθεσιμότητα αυτών των στελεχών (από τα Bloomington *Drosophila* Stock Center, Szeged *Drosophila* Stock Center, Baylor College of Medicine και Harvard Medical School) σε συνδυασμό με τη σύντομη διάρκεια των γενεών σε αυτόν τον οργανισμό, κάνουν δυνατή την ανάλυση των μεταλλαγμένων φαινοτύπων σε σχετικά σύντομο χρονικό διάστημα. Επιπλέον, στον οργανισμό αυτόν έχει εφαρμοστεί με επιτυχία η τεχνολογία του RNAi σε κυτταροκαλλιέργειες (Pollard TD, 2003). Έτσι, έχουν πραγματοποιηθεί διάφορες διερευνήσεις (screens) με σκοπό το χαρακτηρισμό βιολογικών φαινοτύπων σε επίπεδο κυττάρου, όπως είναι η κυτοκίνηση (Somma MP *et al*, 2002; Eggert US *et al*, 2004; Echard A *et al*, 2004), η κυτταρική μορφολογία (Kiger AA *et al*, 2003), η ανάπτυξη και η βιωσιμότητα (Boutros M *et al*, 2004).

Ο ποντικός (*Mus musculus*) αποτελεί τον κύριο οργανισμό-μοντέλο για την μελέτη της γενετικής και της φυσιολογίας των θηλαστικών. Τα πλεονεκτήματά του έναντι των άλλων θηλαστικών είναι η διαθεσιμότητα πληθώρας αιμομικτικών (inbred) στελεχών, η σχετική ευχέρεια στην τροποποίηση και ανάλυση του γονιδιώματός του, η ικανότητα γρήγορης αναπαραγωγής με ταυτόχρονη δημιουργία πολλών απογόνων και το σχετικά χαμηλό κόστος συντήρησής του. Από τα πιο σημαντικά επιτεύγματα ήταν η δημιουργία διαγονιδιακών και "knock out" ποντικών, αρκετά από τα οποία αποτελούν μοντέλα για νοσήματα του ανθρώπου, καθώς εμφανίζουν παρόμοια παθολογία. Χαρακτηριστικά παραδείγματα είναι τα ποντίκια από τα οποία λείπει το γονίδιο p53, τα οποία, όπως και στην αντίστοιχη κατάσταση στον άνθρωπο, εμφανίζεται προδιάθεση για καρκίνους, τα ποντίκια που αποτελούν μοντέλο για το σύνδρομο Down κ.λπ. Κατά την πορεία της ανάλυσης του γονιδιώματος αυτού του οργανισμού, δημιουργήθηκαν διάφοροι γενετικοί (Dietrich WF *et al*, 1996; Rhodes M *et al*, 1998) και φυσικοί χάρτες (Nusbaum C *et al*, 1999; Van Etten WJ *et al*, 1999; Gregory SG *et al*, 2002) μεγάλης ευκρίνειας και απομονώθηκε πληθώρα ESTs (Marra M *et al*, 1999) και cDNAs (Carninci P *et al*, 2003). Το 2002 δημοσιεύτηκε η πρώτη έκδοση της αλληλουχία του γονιδιώματος του στελέχους C57BL/6J (Black 6) με «βάθος διαβάσματος» ~7 φορές και η προκαταρκτική ανάλυση της αλληλουχίας (Waterston RH *et al*, 2002α). Η στρατηγική που ακολουθήθηκε συνδύασε τα χαρακτηριστικά της ιεραρχικής- βάσει φυσικού χάρτη- shotgun αλληλούχησης και της shotgun αλληλούχησης σε επίπεδο γονιδιώματος. Το γονιδίωμα του ποντικού έχει περίπου το ίδιο μέγεθος με του ανθρώπου, περιέχει παρόμοιο αριθμό γονιδίων που

τα περισσότερα εμφανίζουν ορθόλογα στον άνθρωπο και εμφανίζει εκτεταμένη συνταινία (συντηρημένη γονιδιακή διάταξη).

Εκτός των οργανισμών που περιγράφησαν, έχουν ολοκληρωθεί ή είναι σε εξέλιξη η αλληλούχηση πλήθους άλλων οργανισμών-μοντέλων, λόγω βιοϊατρικού, εξελικτικού ή οικονομικού ενδιαφέροντος. Μέχρι σήμερα, έχει ολοκληρωθεί η αλληλούχηση των γονιδιωμάτων περισσότερων των 200 προκαρυωτικών οργανισμών ενώ είναι σε εξέλιξη τουλάχιστον 500 επιπλέον. Από τους ευκαρυωτικούς οργανισμούς, έχει ολοκληρωθεί η ανάλυση σχετικά λίγων και σε εξέλιξη είναι η αλληλούχηση ~300 γονιδιωμάτων (www.genomesonline.org). Από τους σημαντικότερους οργανισμούς που η ανάλυσή τους πρόκειται να συνεισφέρει στην κατανόηση της βιολογίας του ανθρώπου είναι:

α) Ο αρουραίος (***Rattus norvegicus***) που αποτελεί το σημαντικότερο μοντέλο για τη γενετική χαρτογράφηση πολυπαραγοντικών φαινοτύπων (James MR & Lindpaintner K, 1997) όπως η αρθρίτιδα, η υπέρταση, ο διαβήτης και οι καρδιοαγγειακές παθήσεις. Την τελευταία δεκαετία έχουν ταυτοποιηθεί ~ 150 γενετικοί τόποι που σχετίζονται με πολυπαραγοντικά χαρακτηριστικά (<http://ratmap.gen.gu.se>). Η πρώτη έκδοση της αλληλουχίας του γονιδιώματος δημοσιεύτηκε το 2004 (Gibbs RA *et al*, 2004) και επιπλέον, υπάρχουν διαθέσιμοι γενετικοί, φυσικοί και μεταγραφικοί χάρτες μεγάλης ευκρίνειας (Bihoreau MT *et al*, 1997 & 2001; Watanabe TK *et al*, 1999; McCarthy LC *et al*, 2000; Scheetz TE *et al*, 2001).

β) Οι ιχθείς ***Fugu rubripes***, ***Tetraodon nigroviridis*** και ***Danio rerio*** (zebrafish).

Το ***Fugu rubripes*** προτάθηκε σαν μοντέλο σπονδυλωτού οργανισμού (Brenner S *et al*, 1993) λόγω του «συμπαγούς» γονιδιώματός του, το οποίο παρόλο που είναι μόνο 365Mb, έχει σημαντικό ποσοστό ομολογίας στο επίπεδο των γονιδίων με τον άνθρωπο. Στην επιλογή του ως μοντέλο βοήθησε επιπλέον, το μικρό μέγεθος των ιντρονίων του και το σχετικά μικρό ποσοστό (~15%) επαναλαμβανόμενων αλληλουχιών (Brenner S *et al*, 1993). Τα παραπάνω χαρακτηριστικά καθιστούν τον οργανισμό αυτό μία απλή και οικονομική προσέγγιση που μπορεί, λόγω της μεγάλης σχετικά εξελικτικής απόστασης από τον άνθρωπο, να συνεισφέρει στην ταυτοποίηση σημαντικών συντηρημένων στοιχείων (Elgar G *et al*, 1996). Η πρώτη έκδοση της αλληλουχίας ολοκληρώθηκε το 2002 (Aparicio S *et al*, 2002), ενώ υπάρχουν διαθέσιμα πολλά ESTs (Clark MS *et al*, 2003) και ένας πρόσφατος γενετικός χάρτης (Kai W *et al*, 2005). Το ***Tetraodon nigroviridis*** που απέχει εξελικτικά 20-30 εκατομύρια χρόνια από το ***Fugu rubripes*** προτάθηκε σαν μοντέλο που θα συνεισφέρει στην ταυτοποίηση λειτουργικών αλληλουχιών που εμφανίστηκαν μετά την απόκλιση ανθρώπου-τελεόστεων (Cernogorac-Jurcevic T *et al*, 1997). Η πρώτη έκδοση της αλληλουχίας του γονιδιώματός του (~350Mb), που είναι εξίσου

«συμπαγές» με αυτό του *Fugu rubripes*, δημοσιεύτηκε το 2004 (Jaillon O *et al*, 2004). Το **Danio rerio** (zebrafish), παρόλο το μεγάλο μέγεθος του γονιδιώματός του (~1,7Gb) παρουσιάζει σημαντικά πλεονεκτήματα και γι' αυτό και χρησιμοποιείται ως μοντέλο. Το σημαντικότερο είναι ότι πρόκειται για διαυγή οργανισμό γεγονός που κάνει δυνατή την παρατήρηση *in vivo*. Επιπλέον, καθώς η επαγωγή νέων μεταλλαγών είναι σχετικά εύκολη, έχει πραγματοποιηθεί μεταλλαξογένεση σε ευρεία κλίμακα για την ταυτοποίηση γονιδίων που συμμετέχουν σε συγκεκριμένα βιολογικά μονοπάτια, π.χ. στην ανάπτυξη του νευρικού συστήματος. Στα πλαίσια αυτού, έχουν καταγραφεί ~4000 φαινότυποι εμβρυικής θνησιμότητας («big screen», Driever W *et al*, 1996; Haffter P *et al*, 1996). Η αλληλούχηση του γονιδιώματος του οργανισμού αυτού είναι σε εξέλιξη ενώ παράλληλα έχουν κατασκευαστεί και είναι διαθέσιμοι γενετικοί και φυσικοί χάρτες μεγάλης διακριτικής ικανότητας.

γ) Ο χιμπατζής (***Pan troglodytes***) που απέχει εξελικτικά από τον άνθρωπο ~6 εκατομμύρια χρόνια, αποτελεί σημαντικό εργαλείο για την κατανόηση της εξέλιξης των πρωτεύοντων. Ο κύριος στόχος είναι η ταυτοποίηση των κρισιμότερων διαφορών μεταξύ των δύο οργανισμών που οδηγούν στις παρατηρούμενες διαφορές στη φυσιολογία και την συμπεριφορά. Σ' αυτήν την μελέτη θα βοηθήσουν τα γονιδιώματα άλλων πρωτεύοντων, π.χ. του ουραγοτάγκου (*Pongo pygmaeus*) που η ανάλυσή του είναι ήδη σε εξέλιξη. Η πρώτη έκδοση της αλληλουχίας του χιμπατζή, με «βάθος διαβάσματος» ~4 φορές, ολοκληρώθηκε πρόσφατα (Chimpanzee Sequencing and Analysis Consortium, 2005).

1.6. Το χρωμόσωμα 10 του ανθρώπου

Το χρωμόσωμα 10 είναι ένα μεσαίου μεγέθους υπομετακεντρικό χρωμόσωμα που αντιστοιχεί περίπου στο 4,2% του συνολικού γονιδιώματος. Το φυσικό του μήκος παλαιότερα είχε εκτιμηθεί σε 144Mb αλλά πλέον έχει καθοριστεί στις ~135Mb (Deloukas P *et al*, 2004). Στα πλαίσια του HGP, το εργαστήριό μας την τελευταία δεκαετία εστίαστηκε στην γονιδιωματική χαρτογράφηση αυτού του χρωμοσώματος. Μέχρι την έναρξη αυτής της διατριβής είχαν κατασκευαστεί ο γενετικός χάρτης (EUROGEM, Marzella R *et al*, 1997), ο χάρτης των σημείων μειωτικού ανασυνδυασμού (meiotic breakpoint, Cox SA *et al*, 1996) και ο φυσικός χάρτης κυτταρικών υβριδίων ανθρώπου-τροκτικού (Marzella R *et al*, 1997).

Στο πρώτο μέρος αυτής της εργασίας θα περιγράψουμε την συνεισφορά μας στον φυσικό χάρτη των γενωμικών κλώνων που χρησιμοποιήθηκε στην αλληλούχηση του χρωμοσώματος. Το αποτέλεσμα όλων αυτών των χαρτογραφικών προσεγγίσεων προσφέρει σημαντικά εργαλεία για τη χαρτογράφηση νοσημάτων και την ταυτοποίηση γονιδίων. Στο δεύτερο μέρος της εργασίας και με εφόδια τον

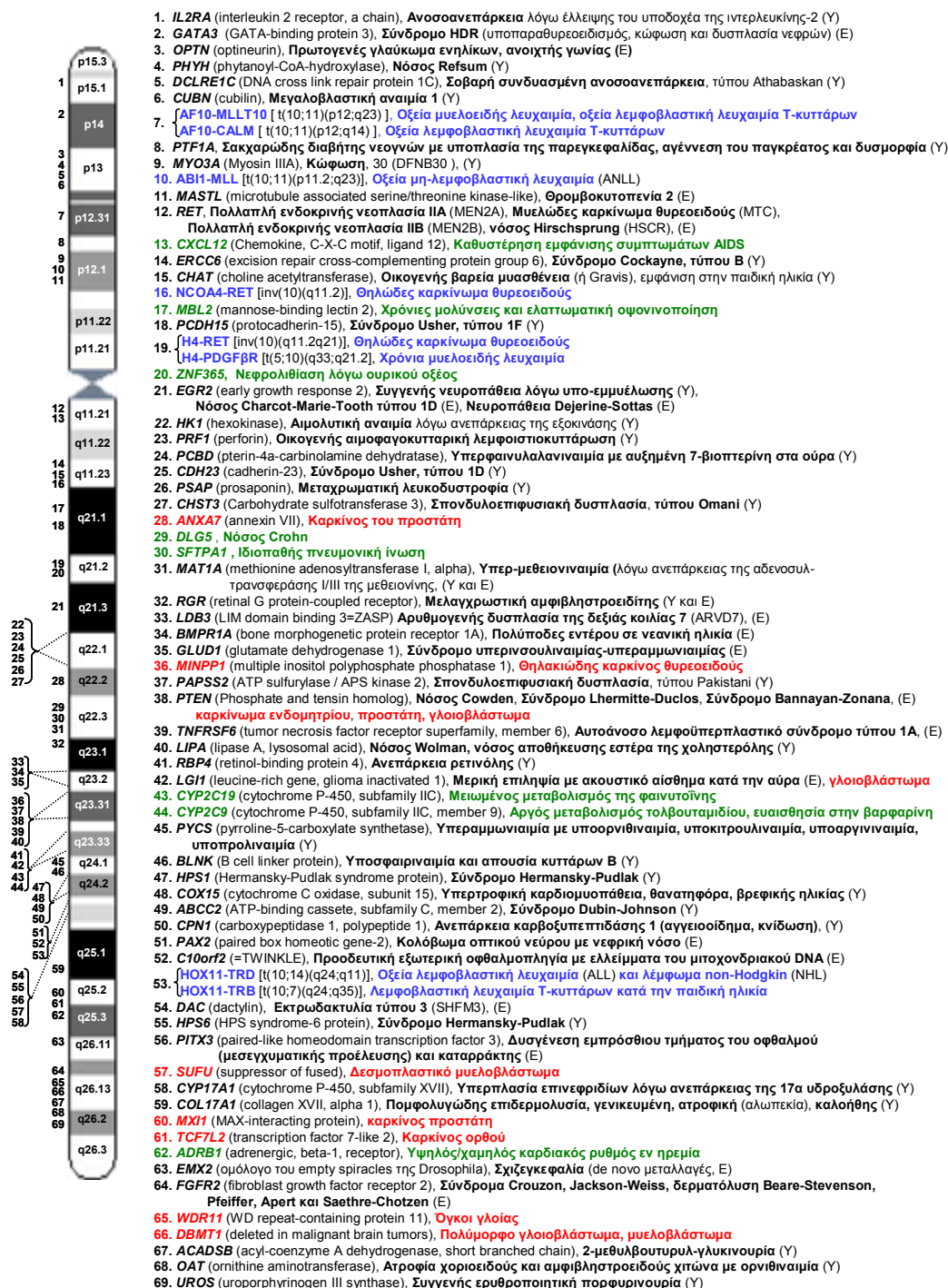
ολοκληρωμένο φυσικό χάρτη και την πλήρη αλληλουχία του χρωμοσώματος, προχωρήσαμε στην κλωνοποίηση μιας σειράς νέων γονιδίων βάσει της χρωμοσωμικής τους τοπογραφίας. Στόχοι μας ήταν α) η ταυτοποίηση του γονιδίου που σχετίζεται με τη μερική επιληψία του κροταφικού λοβού (ADLTE), που εμφανίζει αυτοσωμικό επικρατή τρόπο κληρονομησης και χαρτογραφείται στη ζώνη 10q23-q24 και β) η κλωνοποίηση της εύθραυστης θέσης FRA10A, που ανήκει στην κατηγορία των σπάνιων θέσεων που επάγονται απουσία φυλλικού οξέος (βλ. παρακάτω).

➤ **Γονίδια του χρωμοσώματος 10 που σχετίζονται με νοσήματα**

Μέχρι σήμερα περίπου 70 γονίδια του χρωμοσώματος 10 έχουν συσχετιστεί με την εμφάνιση γενετικών διαταραχών, μεταξύ των οποίων συγκαταλέγονται: 1) διάφοροι τύποι καρκίνου (π.χ. η πολλαπλή ενδοκρινής νεοπλασία τύπου 2A, MEN2A, το πολύμορφο γλοιοβλάστωμα), 2) ανοσοανεπάρκειες (π.χ. η οξεία ανοσοανεπάρκεια RS-SCID), 3) αναπτυξιακές ανωμαλίες (π.χ. η εκτρωδακτυλία τύπου III, SHFM3, η δυσπλασία του συζευκτικού χόνδρου των σπονδύλων SEMD), 4) ενζυμικές ανεπάρκειες (π.χ. η αιμολυτική αναιμία λόγω ανεπάρκεια της εξοκινάσης, η υπερινσουλιναίμια και υπεραμμωνιαίμια λόγω ανεπάρκειας της γλουταμικής αφυδρογονάσης), 5) διαταραχές του μεταβολισμού (π.χ. η νόσος Wolman) και 7) νευρολογικές διαταραχές (π.χ. η μερική επιληψία πλάγιου κροταφικού λοβού ADLTE, η νόσος Charcot-Marie-Tooth τύπου I), Εικ. 7. Επιπλέον, με μελέτες γενετικής σύνδεσης, απώλειας ετεροζυγωτίας και κυτταρογενετικής ανάλυσης, έχουν χαρτογραφηθεί στο χρωμόσωμα αυτό, οι κρίσιμες περιοχές ~35 γενετικών νοσημάτων. Τα μισά περίπου από αυτά εμφανίζουν μεντελικό τύπο κληρονομησης (π.χ. σπαστική παραπληγία τύπου 9 και 27) ενώ τα υπόλοιπα είναι πολυπαραγοντικά νοσήματα (σακχαρώδης διαβήτης, παχυσαρκία, νόσος Alzheimer, σχιζοφρένεια) ή οφείλονται σε ελλείμματα τμημάτων του χρωμοσώματος (π.χ. το σύνδρομο DiGeorge 2).

➤ **Κυτταρογενετικά χαρακτηριστικά**

Μέχρι τώρα έχουν αναφερθεί 46 σημεία θραύσης του χρωμοσώματος 10. Σ' αυτά συμπεριλαμβάνονται χαρακτηρισμένες χρωμοσωμικές μετατοπίσεις που οδηγούν σε διαγονιδιακές συντήξεις οι οποίες έχουν συσχετιστεί με διάφορες μορφές λευχαιμίας, π.χ. η $t(10;11)(p12;q23)$ που αφορά τη σύντηξη των γονιδίων AF10 και MMLT10 και συνδέεται με την οξεία μυελοειδή λευχαιμία και την οξεία λεμφοβλαστική λευχαιμία των T-κυττάρων (Εικ. 7). Επίσης, μία αναστροφή της περιοχής 10q11.2q21 συνδέεται με την εμφάνιση θηλοειδούς καρκινώματος του θυρεοειδούς.



Εικ. 7: Τα γονίδια του χρωμοσώματος 10 που έχουν συσχετιστεί με γενετικά νοσήματα
 Απεικονίζονται: 1) τα γονίδια στα οποία έχουν ταυτοποιηθεί μεταλλάξεις σε μονογονιδιακά νοσήματα (μαύρο) με επικρατή (E) ή υπολειπόμενο (Y) τύπο κληρονομιάς, 2) τα γονίδια που έχουν συσχετιστεί με πολυπαραγοντικά νοσήματα (πράσινο), 3) με καρκίνους λόγω σωματικών μεταλλάξεων σε αυτά (κόκκινο) και 4) τα γονίδια που συμμετέχουν σε χρωμοσωμικές μετατοπίσεις (μπλε). Η χαρτογράφηση τους κατά μήκος του χρωμοσώματος στηρίχθηκε σε στοιχεία της βάσης δεδομένων Ensembl (http://www.ensembl.org/Homo_sapiens/index.html). Είναι αξιοσημείωτο το γεγονός ότι το 40% των γονιδίων που συνδέονται με νοσήματα στο 10, χαρτογραφείται στην περιοχή q23-q24 το μέγεθος της οποίας αντιστοιχεί σε λιγότερο από το 20% του χρωμοσώματος. γεγονός που είναι συνέπεια της υψηλής γονιδιακής πυκνότητάς τους. Τα στοιχεία προέρχονται από την ανασκόπηση για το χρωμόσωμα (Moschosas NK, 2003) και την βάση δεδομένων OMIM (<http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=OMIM>), (Ιούλιος 2005).

Ακόμη, διάφορα ελλείμματα που περιλαμβάνουν τμήματα είτε του μικρού είτε του μεγάλου βραχίονα έχουν ταυτοποιηθεί σε γλοιώματα ενώ η έλλειψη του ακραίου 10q έχει συσχετιστεί με τον καρκίνο του ενδομητρίου (Εικ. 7). Η έλλειψη της περιοχής p14-p13 έχει παρατηρηθεί σε άτομα με κλινικό φαινότυπο παρόμοιο με αυτόν του συνδρόμου DiGeorge και χαρακτηρίζεται ως DiGeorge 2. Ακόμη, η μονοσωμία της περιοχής 10q25-qter έχει συσχετισθεί με σοβαρή διανοητική καθυστέρηση και πολλαπλές συγγενείς ανωμαλίες ενώ της 10q26-qter με ήπια διανοητική καθυστέρηση και διαταραχές στην συμπεριφορά. Ένας ουδέτερος κλινικά πολυμορφισμός αφορά την αναστροφή της περικεντρομερικής περιοχής p11.2-q21.2. Επίσης, έχουν καταγραφεί 7 περιπτώσεις σχηματισμού δακτυλιοειδούς χρωμοσώματος r(10)(p15q26) που σχετίζονται με ήπια διανοητική καθυστέρηση. Τέλος, στο μεγάλο βραχίονα έχουν ταυτοποιηθεί 7 εύθραυστες θέσεις από τις οποίες οι 5 ανήκουν στις συνήθεις (FRA10G, FRA10C, FRA10D, FRA10E, FRA10F) και οι 2 στις σπάνιες (FRA10A, FRA10B).

➤ **Αυτοσωμική επικρατής επιληψία του κροταφικού λοβού (ADLTE)**

Οι επιληψίες

Η επιληψία είναι μία νευρολογική κατάσταση που μπορεί να είναι το χαρακτηριστικό μιας νόσου ή να αποτελεί χαρακτηριστικό ενός συνδρόμου. Στην επιληψία, τα άτομα εμφανίζουν κρίσεις που προκαλούνται από διαταραχές στην ηλεκτρική δραστηριότητα του εγκεφάλου και εκδηλώνονται με διαταραχές στις αισθήσεις, στη συνείδηση ή την συμπεριφορά οι οποίες παρουσιάζονται ξαφνικά και σταματάνε από μόνες τους. Οι επιληψίες ταξινομούνται: 1) σε συμπτωματικές όταν η αιτία που τις προκαλεί είναι διακριτή, π.χ. κάποιο τραύμα ή δομική ανωμαλία του εγκεφάλου (π.χ. όγκος, δυσπλασία των αγγείων), τυχόν μόλυνση κ.λπ. Αυτές υπολογίζεται ότι ευθύνονται για το 40% περίπου των περιπτώσεων και 2) σε ιδιοπαθείς, όταν στο άτομο που εμφανίζει τη νόσο δεν παρατηρούνται ανατομοπαθολογικές βλάβες. Οι ιδιοπαθείς επιληψίες φαίνεται να έχουν ισχυρή γενετική βάση ενώ ο τρόπος κληρονόμησής τους είναι είτε απλός μεντελικός είτε συχνότερα, πολυπαραγοντικός. Περαιτέρω, διακρίνονται σε γενικευμένες και σε μερικές ή εστιακές: οι πρώτες περιλαμβάνουν και τα δύο ημισφαίρια του εγκεφάλου και κατά συνέπεια οι ενδεχόμενες κινητικές διαταραχές εμφανίζονται και στις δύο πλευρές του σώματος. Αντίθετα, οι εστιασμένες επιληπτικές κρίσεις προέρχονται από εντοπισμένα σημεία του εγκεφάλου και, στην περίπτωση που εμφανίζονται κινητικές διαταραχές, αυτές συνήθως εντοπίζονται σε συγκεκριμένη ομάδα μυών. Τα συμπτώματα εμφανίζονται μόνο κατά την διάρκεια της επιληπτικής κρίσης ενώ κατά

τα μεσοδιαστήματα το άτομο είναι απόλυτα φυσιολογικό. Την τελευταία δεκαετία εντοπίστηκαν μεταλλαγές σε μία σειρά από γονίδια, οι οποίες αποτελούν την αιτία ιδιοπαθών επιληψιών που κληρονομούνται με μεντελικό τρόπο. Σχεδόν όλα τα γονίδια που ταυτοποιήθηκαν κωδικοποιούν για κανάλια ιόντων, δηλαδή για πρωτεΐνες της κυτταρικής μεμβράνης που ρυθμίζουν την ιοντική ροή και τη διεγερσιμότητα των νευρώνων (Πιν. 2). Ο συσχετισμός της επιληψίας με τα κανάλια ιόντων είναι τόσο ισχυρός που συχνά οι επιληψίες ονομάζονται καναλο-πάθειες.

Πιν. 2: Τα γονίδια στα οποία έχουν ταυτοποιηθεί μεταλλαγές στις κύριες μορφές ιδιοπαθών γενικευμένων και μερικών επιληψιών

ΓΟΝΙΔΙΟ	ΣΥΝΔΡΟΜΟ	ΒΙΒΛΙΟΓΡΑΦΙΑ
Κανάλια		
Κανάλια ιόντων Na		
SCN1A	GEFS+ & SMEI	#1-9
SCN2A	BFNIS & GEFS+	#10-11
SCN1B	GEFS+	#12-13
Κανάλια ιόντων K		
KCNA1	εστιακές κρίσεις	#14
KCNQ2	BFNS & μυοκυμία	#15-17
KCNQ3	BFNS	#18-19
Κανάλια ιόντων Cl		
CLCN2	IGE	#20
Υποδοχείς GABA		
GABRA1	ADJME	#21
GABRG2	CAE, FS & GEFS+	#22-25
Νικοτινικοί υποδοχείς της ακετυλοχολίνης		
CHRNA4	ADNFLE	#26-31
CHRNA2	ADNFLE	#32-33
Άλλες κατηγορίες γονιδίων		
MASS1	GEFS+	#34
LGI1**	ADLTE	#35-36

Σημ.: Συντομογραφίες: **ADJME** (autosomal dominant juvenile myoclonic epilepsy), **ADNFLE** (autosomal dominant nocturnal frontal lobe epilepsy), **ADLTE** (autosomal dominant lateral temporal lobe epilepsy), **BFNIS** (benign familial neonatal-infantile seizures), **BFNS** (benign familial neonatal seizures), **CAE** (childhood absence epilepsy), **FS** (febrile seizures), **GEFS+** (generalized epilepsy with febrile seizures plus), **IGE** (idiopathic generalized epilepsy), **SMEI** (severe myoclonic epilepsy of infancy). **Βιβλιογραφία:** Escayg A *et al*, 2000 (#1) & 2001 (#2), Wallace RH *et al*, 2001a (#3), Sugawara T *et al*, 2001a (#4) & 2002 (#5), Claes L *et al*, 2001 (#6) & 2003 (#7), Ohmori I *et al*, 2002 (#8), Fujiwara T *et al*, 2003 (#9), Sugawara T *et al*, 2001b (#10), Heron SE *et al*, 2002 (#11), Wallace RH *et al*, 1998 (#12), Wallace RH *et al*, 2002 (#13), Zuberi SM *et al*, 1999 (#14), Biervert C *et al*, 1998 (#15), Singh NA *et al*, 1998 (#16), Dedek K *et al*, 2001 (#17), Charlier C *et al*, 1998 (#18), Hirose S *et al*, 2000 (#19), Haug K *et al*, 2003 (#20), Cossette P *et al*, 2002 (#21), Baulac S *et al*, 2001 (#22), Wallace RH *et al*, 2001b (#23), Harkin LA *et al*, 2002 (#24), Kananura C *et al*, 2002 (#25), Steinlein OK *et al*, 1995 (#26) & 1997 (#27), Hirose S *et al*, 1999 (#28), Saenz A *et al*, 1999 (#29), McLellan A *et al*, 2003 (#30), Phillips HA *et al*, 2000 (#31), De Fusco M *et al*, 2000 (#32), Phillips HA *et al*, 2001 (#33), Nakayama J *et al*, 2002 (#34), Morande-Redolat JM *et al*, 2002 (#35), Kalachikov S *et al*, 2002 (#36), **βλ. και αποτελέσματα.

H ADLTE

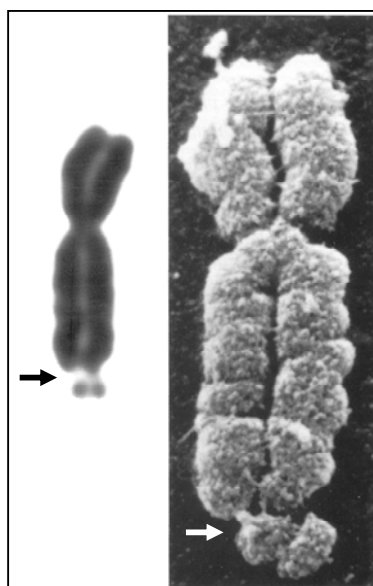
Η αυτοσωμική επικρατής επιληψία του πλάγιου κροταφικού λοβού (autosomal dominant lateral temporal lobe epilepsy, ADLTE) είναι μία ιδιοπαθής εστιακή επιληψία κατά την οποία οι κρίσεις εντοπίζονται στην πλάγια περιοχή του κροταφικού λοβού. Μέχρι τώρα έχουν περιγραφεί ~20 οικογένειες (Ottman R *et al*, 1995; Poza JJ *et al*, 1999; Michelucci R *et al*, 2000; Mautner VF *et al*, 2000; Ikeda A *et al*, 2000; Brodtkorb E *et al*, 2002; Winawer MR *et al*, 2002; Pizzuti A *et al*, 2003; Kanemoto K & Kawasaki J, 2000; Michelucci R *et al*, 2003) στις οποίες η συγκεκριμένη μορφή επιληψία εμφανίζει αυτοσωμικό επικρατή τύπο κληρονόμησης, με διεισδυτικότητα που εκτιμάται στο 70%. Τα συμπτώματα της διαταραχής εκδηλώνονται συνήθως στη δεύτερη ή τρίτη δεκαετία της ζωής ωστόσο υπάρχουν και περιπτώσεις πρωτοεμφάνισης της διαταραχής σε νήπια (Brodtkorb E *et al*, 2002) ή σε άτομα μέσης ηλικίας (Mautner VF *et al*, 2000; Michelucci R *et al*, 2003). Συνήθως, οι επιληπτικές κρίσεις εμφανίζονται μετά από ακουστικά σήματα, που έχουν περιγραφεί σαν βόμβος ή κουδούνισμα ή προκαλούνται από ξαφνικό θόρυβο όπως ήχο τηλεφώνου ή ραδιοφώνου (Michelucci R *et al*, 2003). Στο μεγαλύτερο ποσοστό των ατόμων, η εστιακή κρίση ακολουθείται από δευτερογενείς γενικευμένες τονικές-κλονικές κρίσεις, δηλαδή μία μόνιμη μυϊκή σύσπαση που διαρκεί τουλάχιστον μερικά δευτερόλεπτα ή αμφοτερόπλευρα κλονικά τινάγματα αντίστοιχα. Επίσης, αναφέρθηκαν άτομα που εμφανίζουν αφασία, δηλαδή αδυναμία λόγου (Brodtkorb E *et al*, 2002; Kanemoto K & Kawasaki J, 2000; Ikeda A *et al*, 2000; Winawer MR *et al*, 2002; Pizzuti A *et al*, 2003; Michelucci R *et al*, 2003). Γενικά, τα κλινικά κριτήρια που χρησιμοποιούνται για την ταξινόμηση των περιπτώσεων στην συγκεκριμένη μορφή επιληψίας είναι: α) αυτοσωμικός επικρατής τύπος κληρονόμησης, β) κρίσεις που η σημειολογία τους δείχνει ως προέλευση τον πλάγιο κροταφικό λοβό, γ) απουσία νευρολογικών διαταραχών και δ) απουσία δομικών ανωμαλιών του εγκεφάλου. Η κατάσταση είναι ελεγχόμενη εφόσον χορηγηθεί κατάλληλη αντιεπιληπτική αγωγή. Με ανάλυση γενετικής σύνδεσης, ο υπεύθυνος γενετικός τόπος χαρτογραφήθηκε στην χρωμοσωμική περιοχή 10q23-q24. Αρχικά, η κρίσιμη περιοχή εκτείνονταν σε μήκος ~10cM (Ottman R *et al*, 1995) και κατόπιν περιορίστηκε στα ~3cM (Poza JJ *et al*, 1999).

➤ **Εύθραυστη χρωμοσωμική θέση FRA10A**

Οι εύθραυστες χρωμοσωμικές θέσεις

Εύθραυστες θέσεις ονομάζονται τα κενά ή οι ασυνέχειες στη χρωμοσωμική δομή που παρατηρούνται κυτταρογενετικά σε παρασκευάσματα μεταφασικών

κυττάρων που έχουν καλλιεργηθεί κάτω από συγκεκριμένες συνθήκες (Εικ.8). Το μέγεθος των θέσεων αυτών ποικίλλει ενώ σε μερικές περιπτώσεις μπορούν να οδηγήσουν σε χρωμοσωμική θραύση στην περιοχή. Οι εύθραυστες θέσεις εκδηλώνονται σε ποσοστό των χρωμοσώματα των κυττάρων μιας κυτταροκαλλιέργειας που ποικίλλει από >90% έως και <10% (FRA16B και FRA3B αντίστοιχα). Ταξινομούνται στις συνήθεις, που υπάρχουν στα χρωμοσώματα όλων σχεδόν των ατόμων και αποτελούν συστατικό της δομής των φυσιολογικών χρωμοσωμάτων και τις σπάνιες, που εμφανίζονται σε ποσοστό μικρότερο του 5% του πληθυσμού και κληρονομούνται με μεντελικό τρόπο. Περαιτέρω ταξινόμηση γίνεται βάσει του τρόπου επαγωγής τους, δηλαδή των συνθηκών κάτω από τις οποίες καλλιεργούνται τα κύτταρα ώστε να εμφανιστούν οι θέσεις αυτές. Ως επί το πλείστον, οι συνθήκες κυτταροκαλλιέργειας είναι τέτοιες που ελαττώνουν το ρυθμό αντιγραφής του DNA χωρίς όμως να αναστέλλουν την κυτταρική διαίρεση. Στο ανθρώπινο γονιδίωμα, μέχρι τώρα έχουν περιγραφεί κυτταρογενετικά ~80 συνήθεις και 31 σπάνιες εύθραυστες θέσεις.



Εικ. 8: Η εύθραυστη θέση **FRA(X)**, όπως φαίνεται κυτταρογενετικά με χρώση Giemsa (αριστερά) και με ανάλυση με ηλεκτρονικό μικροσκόπιο σάρωσης (δεξιά), (ανατύπωση από Griffiths MJ & Strachan MC, 1991).

Οι **συνήθεις εύθραυστες θέσεις** καταλαμβάνουν μεγάλη έκταση που μπορεί να φτάσει έως τις 4,5Mb (FRA7E, Zlotorynski E *et al*, 2003) και χαρτογραφούνται σε περιοχές που χαρακτηρίζονται από γενωμική αστάθεια (ελλείμματα, μετατοπίσεις) σε καρκινικά κύτταρα. Μέχρι τώρα έχουν κλωνοποιηθεί 8 θέσεις αυτής της κατηγορίας, οι FRA3B, FRA16D, FRA7G, FRAXB, FRA6F, FRA7H, FRA7E (Zlotorynski E *et al*, 2003) και FRA7I (Ciullo M *et al*, 2002). Τρεις από αυτές, οι FRA3B, FRA16D και

FRA7G χαρτογραφούνται μέσα ή περιλαμβάνουν πιθανά ογκοκατασταλτικά γονίδια (Ohta M *et al*, 1996; Ried K *et al*, 2000). Η τρέχουσα υπόθεση για την εκδήλωσή τους είναι ότι οι συνήθεις εύθραυστες θέσεις έχουν ενδογενή χαρακτηριστικά που καθυστερούν την αντιγραφή του DNA και καθώς ήδη πρόκειται για περιοχές που αντιγράφονται αργά στην φάση S, τουλάχιστον όπως έχει δείχθει για τις FRA3B και FRA7H, παραμένουν με ατελώς αντιγραμμένο DNA που δεν μπορεί να συμπυκνωθεί πλήρως. Εφόσον, η συμπύκνωση δεν έχει ολοκληρωθεί μέχρι και το τέλος της φάσης G2, οδηγεί στην εκδήλωση της εύθραυστης θέσης. Σε φυσιολογικά κύτταρα, η χρήση αναστολέων της αντιγραφής του DNA, επηρεάζει πιθανώς περισσότερο συγκεκριμένες περιοχές, που είναι και αυτές στις οποίες εμφανίζονται οι εύθραυστες θέσεις. Η διερεύνηση των αλληλουχιών των θέσεων που έχουν ήδη κλωνοποιηθεί, εκτός από το σχετικά υψηλό περιεχόμενο σε A/T, δεν έδειξε την ύπαρξη κάποιων επαναλαμβανόμενων αλληλουχιών όπως στις περιπτώσεις των σπάνιων εύθραυστων θέσεων (βλ. παρακάτω). Ωστόσο, με δομικές μελέτες φάνηκε ότι οι περιοχές αυτές έχουν την δυνατότητα, κατά το ξετύλιγμα της διπλής έλικας, να σχηματίσουν δευτεροταγείς δομές ικανών να διαταράξουν την πρόοδο της αντιγραφής. Το πρώτο κύριο μοριακό μονοπάτι που σχετίζεται με τον έλεγχο της έκφρασης των συνήθων εύθραυστων θέσεων ταυτοποιήθηκε πρόσφατα και περιλαμβάνει την κινάση ATR, έναν κύριο ρυθμιστή του κυτταρικού κύκλου (Bentley NJ *et al*, 1996; Cimprich KA *et al*, 1996). Η μελέτη σε βλαστοκύστες ποντικών από τα οποία απουσιάζει το ATR έδειξε γενικευμένη γενωμική θραύση (Brown EJ & Baltimore D, 2000) ενώ σε κύτταρα, απουσία του ATR, εμφανίζεται αυξημένο ποσοστό εμφάνισης των συνήθων εύθραυστων θέσεων (5-20 φορές) ακόμα και χωρίς αναστολείς της αντιγραφής του DNA (Casper AM *et al*, 2002). Αυτές οι παρατηρήσεις, σε συνδυασμό με το ότι η ATR ενεργοποιείται κυρίως από φουρκέτες αντιγραφής που έχουν σταματήσει, οδήγησε στην υπόθεση ότι οι εύθραυστες θέσεις οφείλονται σε ατελώς αντιγραμμένο DNA λόγω σταματημένων φουρκετών αντιγραφής. Ο λόγος που το σταμάτημα συμβαίνει στις συγκεκριμένες περιοχές δεν είναι γνωστός αλλά πιθανόν να είναι θέμα αλληλουχίας ή δυναμικής των περιοχών αυτών. Απουσία της ATR, ακόμη και χωρίς χρήση αναστολέων της αντιγραφής του DNA, οι εύθραυστες θέσεις εκδηλώνονται γιατί δεν είναι δυνατή η σωστή επεξεργασία των σταματημένων φουρκετών της αντιγραφής. Επιπλέον, αυξημένο ποσοστό έκφρασης των συνήθων εύθραυστων θέσεων εμφανίζεται και σε κύτταρα από τα οποία απουσιάζει το BRCA1 (Arlt MF *et al*, 2004) ή το SMC1 (structural maintenance of chromosome), μόρια που αποτελούν στόχους της ATR (Musio A *et al*, 2005). Οι θέσεις αυτές είναι συντηρημένες σε όλα τα θηλαστικά που εξετάστηκαν (άλογο, αγελάδα, χοίρος, σκύλος, γάτα, ποντικός, αρουραίος), επάγονται με τον ίδιο

τρόπο (Elder FF & Robinson TJ, 1989; Stone DM *et al*, 1991 & 1993; Stone DM & Stephens KE, 1993; Yang MY & Long SE, 1993; McAllister BF & Greenbaum IF, 1997) ενώ για κάποιες από αυτές έχει δειχθεί ότι είναι πράγματι συνταϊνικές και ορθόλογες με αυτές του ανθρώπου (Glover TW *et al*, 1998; Shiraishi T *et al*, 2001; Krummel KA *et al*, 2002; Rozier I *et al*, 2004).

Από τις 31 **σπάνιες εύθραυστες θέσεις** που έχουν περιγραφεί κυτταρογενετικά, μέχρι τώρα έχουν κλωνοποιηθεί οι 7: 4 αυτοσωμικές (FRA16A, FRA11B, FRA10B, FRA16B) και τρεις στο χρωμόσωμα X (FRAXA, FRAXE, FRAXF). Η πλειονότητα των σπάνιων θέσεων επάγεται όταν απουσιάζει από το θρεπτικό καλλιέργειας των κυττάρων το φυλλικό οξύ, που λειτουργεί ως συμπαραγοντας για τη σύνθεση της θυμιδίνης. Η μοριακή βάση της εμφάνισης των θέσεων αυτής της κατηγορίας είναι η δραματική επέκταση της τρινουκλεοτιδικής επανάληψης CGG, που στον φυσιολογικό πληθυσμό είναι πολυμορφική, αλλά με περιορισμένο μέγεθος. Η πιο γνωστή και καλά μελετημένη είναι η FRAXA που η εμφάνισή της συνδέεται με το σύνδρομο του εύθραυστου X, την πιο συχνή αιτία διανοητικής καθυστέρησης μετά το σύνδρομο Down. Η επανάληψη CGG εντοπίζεται στο 5'UTR του γονιδίου *FMR1* που κωδικοποιεί την FMRP, η οποία συνδέεται με μόρια mRNA, συμπεριλαμβανομένου του *FMR1*, και λειτουργεί σαν καταστολέας της μετάφρασης (Ashley CT *et al*, 1993; Jin P & Warren ST, 2003). Ο αριθμός των επαναλήψεων στα φυσιολογικά χρωμοσώματα είναι 5-55, στην κατάσταση της «προ-μεταλλαγής» μεταξύ 55 και 230 ενώ στην πλήρη μεταλλαγή υπερβαίνει τις 230. Κατ' αυτήν η περιοχή που περιβάλλει την επανάληψη είναι υπερμεθυλιωμένη και το γονίδιο καταστέλλεται μεταγραφικά (Chiurazzi P & Neri G, 2001). Η τρέχουσα υπόθεση γι' αυτό, είναι ότι το μεθυλιωμένο DNA αναγνωρίζεται από διάφορες πρωτεΐνες (π.χ. MeCP2, MBD2) που με τη σειρά τους στρατολογούν απακετυλάσες των ιστονών HDACs και ακολούθως πρωτεΐνες-καταστολείς (π.χ. ISWI ATPάση) οι οποίες ξεκινούν τοπικά τη συμπύκνωση της χρωματίνης (Bird AP & Wolffe AP, 1999; Ng HH & Bird A, 1999). Τα άτομα με προμεταλλαγή είναι κλινικά φυσιολογικά, εκτός από μικρό ποσοστό που εμφανίζει αταξία κατά την μέση ηλικία (Leehey MA *et al*, 2003; Hagerman RJ *et al*, 2004; Hagerman PJ & Hagerman RJ, 2004) ή ήπια νευροψυχιατρική διαταραχή (Steyaert J *et al*, 2003).

Η επέκταση τρινουκλεοτιδικών επαναλήψεων έχει επίσης συσχετιστεί και με διάφορες άλλες νευρολογικές διαταραχές (π.χ. νόσο του Huntington). Οι περισσότερες από αυτές εμφανίζουν κακή πρόβλεψη (*anticipation*), δηλαδή επιδείνωση του κλινικού φαινοτύπου με παράλληλη μείωση της ηλικίας εκδήλωσης της νόσου, από γενιά σε γενιά. Τόσο στην περίπτωση του εύθραυστου X όσο και σε άλλες νευρολογικές διαταραχές ή εύθραυστες θέσεις που οφείλονται σε επεκτάσεις

επαναλήψεων, εμφανίζεται διαφοροποίηση στην αστάθεια της επανάληψης ανάλογα με τον γονέα μέσω του οποίου αυτή κληρονομείται. Συγκεκριμένα, εκτός από την FRAXA το μήκος της επανάληψης αυξάνεται όταν η μεταβίβαση γίνεται μέσω της μητέρας και στις περιπτώσεις της αταξίας Friedreich (FRDA), της νωτιοπαρεγκεφαλιδικής αταξίας 8 (SCA8), της εύθραυστης θέσης FRA16A, και της μυοτονικής δυστροφίας 1 (DM1). Αντίθετα, το μήκος της επανάληψης αυξάνεται όταν η μεταβίβαση γίνεται μέσω του πατέρα, στα γονίδια για τη νόσο του Huntington, των νωτιοπαρεγκεφαλιδικών αταξιών SCA1, SCA2, SCA3, SCA7, στην νευροεκφυλιστική νόσο DRPLA (dentatorubral-pallidoluysian atrophy), στην εύθραυστη θέση FRA11B και στην προοδευτική μυοκλωνική επιληψία τύπου 1 EPM1 (Pearson CE *et al*, 2005).

Τελομερικά και πλησίον της FRAXA χαρτογραφείται η θέση FRAXE, που σχετίζεται με μη-συνδρομική διανοητική καθυστέρηση, δηλαδή με καθυστέρηση που δεν συνοδεύεται από κάποιον άλλο παθολογικό φαινότυπο και η FRAXF που δεν έχει συσχετισθεί με κλινικό φαινότυπο. Όπως και στην περίπτωση της FRAXA, η εκδήλωσή τους οφείλεται σε επεκτάσεις τρινουκλεοτιδίων CGG που υπάρχουν στο 5' άκρο των γονιδίων *FMR2* και *FAM11A* αντίστοιχα (Gecz J *et al*, 1996; Gu Y *et al*, 1996; Shaw MA *et al*, 2002), τα οποία καταστέλλονται μεταγραφικά στην περίπτωση της πλήρους μεταλλαγής. Από τις αυτοσωμικές σπάνιες εύθραυστες θέσεις που έχουν κλωνοποιηθεί, οι FRA16A και FRA11B οφείλονται στην επέκταση της επανάληψης CGG ενώ οι FRA16B και FRA10B σε επεκτάσεις AT μήκους 33bp και 42bp αντίστοιχα. Η FRA11B είχε προσελκύσει το ενδιαφέρον των ερευνητών γιατί χαρτογραφείται στο 5' UTR του ογκογονιδίου *CBL2* και σε ασθενείς με σύνδρομο Jacobsen έχουν ταυτοποιηθεί χρωμοσωμικά ελλείμματα κοντά στην περιοχή της FRA11B. Εκτενέστερη μελέτη όμως έδειξε ότι η πλειονότητα των ασθενών αυτών φέρει τα ελλείμματα πιο τελομερικά, στο 11q (Michaelis RC *et al*, 1998). Οι υπόλοιπες δεν έχουν συνδεθεί με γονίδιο ή με κλινικό φαινότυπο. Μάλιστα για τις FRA16B, FRA17A και FRA10B έχουν ταυτοποιηθεί και ομόζυγα φυσιολογικά άτομα (Hocking T *et al*, 1999; Berg JM *et al*, 1969; Sutherland GR, 1981).

FRA10A

Η FRA10A ανήκει στις σπάνιες εύθραυστες χρωμοσωμικές θέσεις που ανιχνεύονται κυτταρογενετικά απουσία φυλλικού οξέος και χαρτογραφείται στην περιοχή 10q23.3. Είναι μία από τις συχνότερα εμφανιζόμενες σπάνιες θέσεις με ποσοστό που εκτιμάται στο 1/500. Η ανάλυση οικογενειών στις οποίες εκδηλώνεται η FRA10A, έδειξε ότι η έκφρασή της εξαρτάται από τον γονέα μέσω του οποίου μεταβιβάζεται, φαινόμενο που έχει παρατηρηθεί και σε άλλες θέσεις αυτής της κατηγορίας. Η διυσδεικτικότητα ήταν πλήρης όταν η μεταβίβαση γινόταν μέσω της

μητέρας, αλλά μόνο 50% μέσω του πατέρα (Sherman SL & Sutherland GR, 1986). Παλαιότερες κυτταρογενετικές μελέτες έχουν δείξει ότι η FRA10A είναι η πιο συχνά εμφανιζόμενη εύθραυστη αυτοσωμική θέση σε άτομα με διανοητική καθυστέρηση (Petit P *et al*, 1986). Επίσης, πρόσφατα, στα εργαστήρια των R. Siebert (Kiel, Germany) και J. Gecz (Adelaide, Australia) χαρακτηρίστηκαν οικογένειες κυτταρογενετικά θετικές για την FRA10A. Μέλη των οικογενειών-φορείς της FRA10A εμφανίζουν ήπια διανοητική καθυστέρηση (δυσχέρεια ομιλίας, αυτισμό) ή αναπτυξιακές ανωμαλίες. Ωστόσο, δεν είναι τελείως ξεκάθαρη η συσχέτιση της εύθραυστης θέσης FRA10A με την εκδήλωση παθολογικού φαινοτύπου καθώς έχουν ταυτοποιηθεί και άτομα-φορείς που έχουν φυσιολογικό δείκτη νοημοσύνης. Συνεπώς είτε πρόκειται για μεροληψία της δειγματοληψίας είτε πράγματι υπάρχει συσχέτιση με τον παθολογικό φαινότυπο άλλα με μειωμένη διεισδυτικότητα.

1.7. Αντικείμενο της διατριβής

Κατά την έναρξη της διατριβής, η ομάδα μας, στα πλαίσια του Προγράμματος Χαρτογράφησης και Αλληλούχησης του Γονιδιώματος του Ανθρώπου, είχε κατασκευάσει τον γενετικό χάρτη και τον χάρτη κυτταρικών υβριδίων του χρωμοσώματος 10. Τα αποτελέσματα αυτής της προσπάθειας αξιοποιήθηκαν περαιτέρω κατά την διατριβή αυτή, για την κατασκευή του φυσικού χάρτη υψηλής ευκρίνειας ο οποίος χρησιμοποιήθηκε για την αλληλούχηση του χρωμοσώματος αυτού. Οι πληροφορίες από την χαρτογράφηση και την αλληλούχηση ήταν τα εργαλεία μας για να προχωρήσουμε στην κλωνοποίηση και το χαρακτηρισμό νέων γονιδίων από την περιοχή 10q23.3 που συσχετίζονται με παθολογικές καταστάσεις. Ειδικότερα, το αντικείμενό μας ήταν:

1. *Η συμβολή στην κατασκευή του φυσικού χάρτη υψηλής ευκρίνειας γενωμικών κλώνων των χρωμοσωμικών περιοχών 10q23.3-q25.1 και 10q26.2-qtel, συνολικού μήκους ~35Mb.*
2. *Η αξιοποίηση του φυσικού χάρτη και της γενωμικής αλληλουχίας για την in silico κατασκευή του γονιδιακού χάρτη της κρίσιμης περιοχής για την αυτοσωμική επικρατή μερική επιληψία του κροταφικού λοβού (ADLTE) στο 10q23.3 και η δομική ανάλυση νέων υποψηφίων γονιδίων της περιοχής αυτής και*
3. *Η κλωνοποίηση, ο χαρακτηρισμός και η λειτουργική ανάλυση του γονιδίου FRA10AC1 που χαρτογραφείται στην εύθραυστη χρωμοσωμική θέση FRA10A.*

2

ΥΛΙΚΑ ΚΑΙ ΜΕΘΟΔΟΙ

ΥΛΙΚΑ

2.1. Θρεπτικά

Όλα τα θρεπτικά υλικά έγιναν με υπερκάθαρο νερό και αποστειρώθηκαν είτε σε κλίβανο ατμού είτε με διήθηση μέσω ηθμού με διάμετρο πόρων 0,22μm. Κατά περίπτωση και μετά την αποστείρωση προστίθονταν το κατάλληλο αντιβιοτικό (Πιν. 3).

Πιν. 3: Διαλύματα αντιβιοτικών

Αντιβιοτικό	Διαλύτης	Στοκ	Τελική συγκέντρωση
αμπικιλίνη	H ₂ O	10mg/ml	100μg/ml, 50μg/ml (pcDNA1)
καναμυκίνη	H ₂ O	10mg/ml	30μg/ml
χλωραμφαινικόλη	απόλυτη αιθανόλη	25mg/ml	12,5μg/ml
τετρακυκλίνη	απόλυτη αιθανόλη	5mg/ml	10μg/ml

➤ Καλλιέργεια σακχαρομύκητα

YEP: 10gr yeast extract, 20 gr bactopeptone, 20gr γλυκόζη ανά λίτρο. Ρύθμιση του pH στο 7-7,5 με διάλυμα NaOH

SC: 6,7 gr yeast nitrogen base χωρίς αμινοξέα, 20gr γλυκόζη, 2gr σκόνη SC ανά λίτρο. Ρύθμιση του pH στο 7-7,5 με διάλυμα NaOH. Για την παρασκευή στερεού θρεπτικού προστίθενται πριν την αποστείρωση 20gr άγαρ / λίτρο.

Σκόνη SC: ομοιογενές μίγμα 18 αμινοξέων εξαιρουμένων εκείνων που παράγονται από τους φορείς του συστήματος των δύο υβριδίων, στη συγκεκριμένη περίπτωση της τρυπτοφάνης (pGBT9) και της λευκίνης (pACT2).

M9: Για την παρασκευή 1lt θρεπτικού, αναμιγνύονται 870ml H₂O με 20gr άγαρ και αποστειρώνονται. Όταν η θερμοκρασία του διαλύματος προσεγγίσει τους 55°C, προστίθενται 100ml διαλύματος αλάτων M9, 20ml γλυκόζης 20% v/v (#), 40μl 2,5M CaCl₂ (#), τα απαραίτητα διαλύματα αμινοξέων και νουκλεοτιδίων και το κατάλληλο αντιβιοτικό. Τα συστατικά αναμιγνύονται και το υλικό μοιράζεται σε τρυβλία, όπου και στερεοποιείται.

Διάλυμα αλάτων M9: 6gr Na₂HPO₄, 0,5gr KH₂PO₄, 1gr NH₄Cl ανά 100ml.

Διαλύματα αμινοξέων και νουκλεοτιδίων: Σε όλα το pH ρυθμίζεται στο 7 με αραιό διάλυμα NaOH. Χρησιμοποιήθηκαν διαλύματα αδενίνης x50 (12,5mM), ουρακίλης x100 (20mM), ιστιδίνης x300 (100mM) και τρυπτοφάνης x100 (55mM, φυλάσσεται στους 4°C σε σκοτάδι).

➤ **Καλλιέργεια βακτηρίων**

LB: 10gr bacto-tryptone, 5gr bacto yeast extract, 10gr NaCl ανά λίτρο. Ρύθμιση του pH στο 7-7,5 με διάλυμα NaOH (*). Για την παρασκευή στερεού θρεπτικού προστίθενται πριν την αποστείρωση 15gr άγαρ / λίτρο.

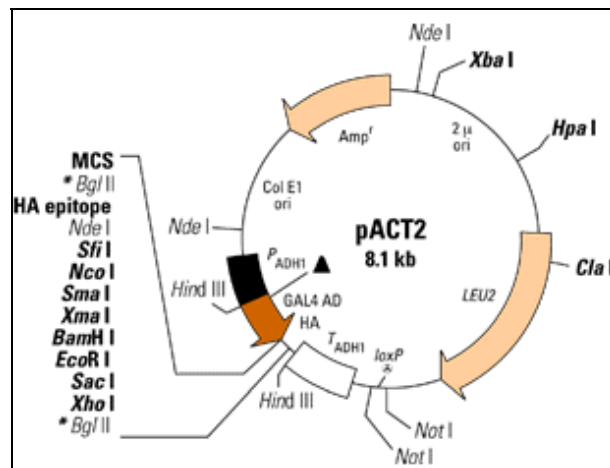
➤ **Καλλιέργεια κυττάρων θηλαστικών**

Θρεπτικό μέσο DMEM: Dulbecco's Modified Eagle's Medium με 4 mM L-glutamine, 0.45% glucose, 0.15% sodium bicarbonate, 50μg/ml gentamycin (δεν είναι απαραίτητο για την ανάπτυξη των κυττάρων).

2.2. Βιβλιοθήκες DNA

- **Γενωμική βιβλιοθήκη RPCI-11:** Γενωμική βιβλιοθήκη που χρησιμοποιήθηκε από όλα τα εργαστήρια που μετείχαν στη χαρτογράφηση και την αλληλούχηση του ανθρώπινου γονιδιώματος. Κατασκευάστηκε από DNA λεμφοκυττάρων άρρενος στον φορέα pBACe 3.6 (~8kb) που είναι κατάλληλος για λήψη μεγάλων ενθεμάτων, παράγει υψηλό αριθμό αντιγράφων και παρέχει ανθεκτικότητα σε χλωραμφαινικόλη. Το μέσο μέγεθος των ενθεμάτων εκτιμάται σε ~170kb και η αντιπροσώπευση του ανθρώπινου γονιδιώματος είναι >10 φορές. Χρησιμοποιήθηκε στη διευθετημένη (gridded) της μορφή, σε 36 νάιλον φίλτρα (8x 12cm) που περιέχουν ~220,000 κλώνους. Τα φίλτρα μας παραχωρήθηκαν από το συνεργαζόμενο Sanger Centre. Περισσότερες πληροφορίες υπάρχουν διαθέσιμες στο: <http://bacpac.chori.org/hmale11.htm>
- **Γενωμική χρωμοσωμοειδική βιβλιοθήκη:** Η βιβλιοθήκη αυτή κατασκευάστηκε στο Sanger Centre με γενωμικούς κλώνους κυρίως της RPCI-11 που εμφανίστηκαν θετικοί, βάσει υβριδοποίησης, για STSs του χρωμοσώματος 10. Χρησιμοποιήθηκε στη διευθετημένη της μορφή σε τρία νάιλον φίλτρα (SC10BACpoly_B_1, SC10BACpoly_B_2 και SC10BACpoly_K_1, 8 x 12cm) που περιέχουν ~18,000 κλώνους.
- **Βιβλιοθήκη cDNA ήπατος ανθρώπου (MATCHMAKER, Clontech cat#HL4024AH):** Χρησιμοποιήθηκε στο πείραμα των δύο υβριδίων. Ο φορέας κλωνοποίησης είναι ο pACT2 (Εικ. 9) που παρέχει υψηλά επίπεδα έκφρασης στον σακχαρομύκητα λόγω του συστατικού υποκινητή ADH1. Οι χιμαιρικές πρωτεΐνες παράγονται σε σύντηξη με την περιοχή ενεργοποίησης της μεταγραφής του παράγοντα GAL4 (AD, 114 αμινοξέα) και κατευθύνονται στον πυρήνα λόγω σήματος πυρηνικού εντοπισμού που υπάρχει ανοδικά της AD. Έχει κατασκευαστεί από

polyA⁺ RNA με συνδυασμό ολιγο-dT και τυχαίων εκκινήτων. Το μέσο μέγεθος των ενθεμάτων είναι 1,3kb και το ποσοστό των ανασυνδυασμένων κλώνων ~90%. Ο φορέας αντιγράφεται αυτόνομα στην *E. coli* και στον *S. cerevisiae*. Φέρει το γονίδιο *bla* (για ανθεκτικότητα στην αμπικιλίνη στην *E. coli*) και το *LEU2* που επιτρέπει στα μετασχηματισμένα κύτταρα του σακχαρομύκητα να αναπτύσσονται σε θρεπτικό μέσο από το οποίο απουσιάζει η λευκίνη.



Εικ. 9: Σχηματική απεικόνιση του φορέα pACT2. Σημειώνεται η περιοχή ενεργοποίησης της μεταγραφής του παράγοντα GAL4 (GAL4 AD) και τα γονίδια *bla* (*amp^r*) και *LEU2*.

- **Βιβλιοθήκη cDNA εγκεφάλου ανθρώπου** (# A900-09, Invitrogen): Έχει παραχθεί από polyA⁺ RNA εγκεφάλου ενήλικος με εκκινήτη ολιγο-dT και χρησιμοποιήθηκε σε μορφή μετασχηματισμένων βακτηρίων (*E. coli*, MC1061/P3). Ο φορέας κλωνοποίησης είναι ο pCDNA1, που προσφέρει ανθεκτικότητα στην αμπικιλίνη. Το μέσο μέγεθος των ενθεμάτων είναι 1,3kb. Πριν την διερεύνησή της, προσδιορίστηκε ο τίτλος της σε 10⁷ (αποικίες/ml).

2.3. Βάσεις δεδομένων και προγράμματα επεξεργασίας αλληλουχιών

Αλληλουχία χρωμοσώματος 10: <http://www.sanger.ac.uk/HGP/Chr10/>

και διερεύνησή της <http://www.sanger.ac.uk/cgi-bin/blast/submitblast/chr10>

Blastn, Blastp: <http://www.ncbi.nlm.nih.gov/BLAST/>

Blast 2 sequences: <http://www.ncbi.nlm.nih.gov/blast/bl2seq/wblast2.cgi>

CEPH: <http://locus.umdnhj.edu/nigms/ceph/ceph.html>

Clustlw: <http://www.ebi.ac.uk/clustalw/index.html>

dbEST: <http://www.ncbi.nlm.nih.gov/dbEST/index.html>

Ensembl: <http://www.ensembl.org/>

Genbank: <http://www.ncbi.nlm.nih.gov/Genbank/index.html>

GeneMap99: <http://www.ncbi.nlm.nih.gov/genemap99/map.cgi?CHR=10>

Genethon: http://www.genlink.wustl.edu/genethon_frame/

Interpro: <http://www.ebi.ac.uk/interpro/>

Marshfield: http://research.marshfieldclinic.org/genetics/Map_Markers/maps/IndexMapFrames.html

NetPhos: <http://www.cbs.dtu.dk/services/NetPhos/>

Pfam: <http://www.sanger.ac.uk/Software/Pfam/>

ProtParam: <http://au.expasy.org/tools/protparam.html>

SHGC-G3: <http://www-shgc.stanford.edu/resources/humanrhmap.html>

SMART: <http://smart.embl-heidelberg.de/>

PSORT II: <http://www.psort.org/>

Sanger Centre <http://www.sanger.ac.uk/>

SUMOplot: <http://www.abgent.com/doc/sumoplot>

TMpred: http://www.ch.embnet.org/software/TMPRED_form.html

UniSTS: <http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=unists>

Φυσικός χάρτης χρωμοσώματος 10: <http://www.sanger.ac.uk/cgi-bin/humace/fpcwebmap.cgi>

WI-RH: http://www.broad.mit.edu/cgi-bin/contig/phys_map

2.4. Εκκινητές

Οι εκκινητές των STSs (Πιν. 4) που χρησιμοποιήθηκαν κατά την κατασκευή του φυσικού χάρτη παρασκευάστηκαν στο Sanger Centre. Οι υπόλοιποι (Πιν. 5) επιλέχθηκαν ώστε το μήκος τους να υπερβαίνει τα 17 νουκλεοτίδια, να υπάρχει C ή G στο 3' άκρο τους ενώ αποφευχθήκαν επαναλήψεις του ίδιου νουκλεοτιδίου, παλίνδρομα τμήματα και περιοχές συμπληρωματικές μέσα στον ίδιο εκκινητή ή στο ζευγάρι του. Επιπλέον, για τα ζεύγη των εκκινητών και όταν υπήρχε δυνατότητα, επιλέχθηκαν παρόμοια Tm. Η σύνθεσή τους έγινε στην MWG-Biotech.

2.5. Βακτηριακά στελέχη

DH5a: Χρησιμοποιήθηκαν σε μετασχηματισμό κατά την κλωνοποίηση τεμαχίων DNA σε πλασμιδιακούς φορείς καθώς η απόδοση είναι υψηλή και υπάρχει δυνατότητα επιλογής των ανασυνδυασμένων πλασμιδίων λόγω της μεταλλαγής που φέρουν στο γονίδιο *LacZ* που εμποδίζει την σύνδεση των υπομονάδων (και την ενεργότητα) της β-γαλακτοσιδάσης, η οποία όμως συμπληρώνεται από φορείς που παράγουν το α-πολυπεπτιδίο. Έτσι τα μη ανασυνδυασμένα πλασμίδια αποκαθιστούν στα μετασχηματισμένα βακτήρια την ενεργότητα του ενζύμου, ενώ αυτά που έχουν δεχθεί ένθεμα, όχι. Κατ' επέκταση οι αντίστοιχες αποικίες είναι λευκές παρουσία X-gal.

KC8: Είναι αυξότροφα σε λευκίνη, τρυπτοφάνη, ουρακίλη, ιστιδίνη και αδενίνη κάνοντας δυνατή την επιλογή των πλασμιδίων που φέρουν τα αντίστοιχα γονίδια, όταν αναπτύσσονται σε ελάχιστο θρεπτικό μέσο M9, ή συγκεκριμένων πλασμιδίων αν προστεθούν τα κατάλληλα συστατικά στο M9. Χρησιμοποιήθηκαν για την εξάλειψη του «δολώματος», καθώς είχαν καλλιεργηθεί παρουσία τρυπτοφάνης, μετά από τον μετασχηματισμό τους με το DNA των θετικών κλώνων του πειράματος των δύο υβριδίων. Το DNA που απομονώνεται από αυτά είναι μεθυλιωμένο γιατί τα κύτταρα φέρουν τη μεταλλαγή *hsdR*. Επίσης εμφανίζουν ανθεκτικότητα στην καναμυκίνη λόγω του μεταθετού στοιχείου *Tn5* που περιέχουν.

BL21 (DE3) CodonPlus: Έχουν σε λυσιγονία τον βακτηριοφάγο DE3 (παράγωγο του βακτηριοφάγου λ) που φέρει το γονίδιο της T7 RNA πολυμεράσης κάτω από τον έλεγχο του υποκινητή *lacUV5*, ο οποίος επάγεται με IPTG. Έτσι, η προσθήκη IPTG σε κύτταρα μετασχηματισμένα με ένα πλασμίδιο με υποκινητή T7, οδηγεί στην παραγωγή της T7 RNA πολυμεράσης που με τη σειρά της μεταγράφει το γονίδιο που έχει κλωνοποιηθεί στο πλασμίδιο. Επιπλέον, έχουν το πλασμίδιο RIL που περιέχει γονίδια για τα tRNAs *argU*, *ileW*, *leuY* και *proL* (που είναι σπανιότερα στην *E.coli* και που πιθανόν να απαιτούνται συχνότερα για την σύνθεση ετερόλογων πρωτεϊνών) και γονίδιο για ανθεκτικότητα σε χλωραμφαινικόλη.

2.6. Κυτταρικές σειρές θηλαστικών

Όλες οι κυτταρικές σειρές που χρησιμοποιήσαμε καλλιεργήθηκαν σε θρεπτικό υλικό DMEM με 10% (CV-1, COS-7, HeLa) ή 20% FCS (CaCO₂) ή 8% CS (NIH-3T3) και 50μg/ml gentamycin στους 37°C παρουσία 5% ή 8% (NIH-3T3) CO₂. Όλα αναπτύσσονται σχηματίζοντας μονοστοιβάδες. Χρησιμοποιήθηκαν σε πειράματα υποκυτταρικού εντοπισμού. Επιπλέον τα COS-7 αξιοποιήθηκαν για την παραγωγή πρωτεϊνών που συμμετείχαν σε δοκιμές συν- και ανοσο- κατακρήμνισης.

CV1 (ATCC #CCL 70): Προέρχονται από νεφρό αφρικανικού πράσινου πιθήκου.

COS-7 (ATCC #CRL 1651): Προέρχονται από τα CV-1 μετά από μετασχηματισμό τους με τον ιό SV40 που είναι ενσωματωμένος στο γονιδίωμα και έχει καταστεί ανενεργός όσον αφορά την αντιγραφή του. Έτσι, τα COS-7 εκφράζουν το T αντιγόνο που είναι η μόνη πρωτεΐνη που απαιτείται για την έναρξη της αντιγραφής του SV40. Ο ιός δεν μπορεί να αντιγραφεί αλλά οι πλασμιδιακές κατασκευές με τις οποίες διαμολύνονται τα κύτταρα αυτά, φέρουν τη θέση έναρξης της αντιγραφής του και έτσι πολλαπλασιάζεται ο αριθμός των αντιγράφων τους (10^4 - 10^5 ανά κύτταρο) που συνεπάγεται υψηλά επίπεδα έκφρασης των αντίστοιχων πρωτεϊνών.

HeLa (CCL-2): Επιθηλιακά κύτταρα ανθρώπου από αδενοκαρκίνωμα τραχήλου

HepG2 (ATCC #HB-8065): Κύτταρα από ηπατοκαρκίνωμα ανθρώπου

CaCO2 (ATCC #HTB 37): Κύτταρα ανθρώπου από αδενοκαρκίνωμα εντέρου.

NIH-3T3 (CRL-1658): Εμβρυϊκοί ινοβλάστες ποντικού.

2.7. Αντισώματα

Όλα τα αντισώματα που χρησιμοποιήθηκαν σε πειράματα ανοσοφθορισμού, Western και ανοσοκατακρήμνισης καθώς και οι συνθήκες επώασης και η προέλευσή τους παρατίθενται στον Πιν. 6.

2.8. Δείγματα DNA

Κατά την κατασκευή του γενετικού χάρτη του χρωμοσώματος 10 (Moschonas N.K, 1996) στα πλαίσια του προγράμματος EUROGEM (European Gene Mapping Project), μας παραχωρήθηκαν από το δείγματα γενωμικού DNA από τις οικογένειες αναφοράς CEPH (Centre d'Etude du Polymorphisme Humain). Στην εργασία αυτή, τα δείγματα χρησιμοποιήθηκαν για την γονοτυπική ανάλυση της επανάληψης CGG του *FRA10AC1* και για την παρασκευή των STSs του φυσικού χάρτη.

2.9. Πλασμιδιακοί φορείς

Χρησιμοποιήθηκαν φορείς κλωνοποίησης τεμαχίων DNA για προσδιορισμό της αλληλουχίας τους και για πραγματοποίηση αντιδράσεων μεταγραφής *in vitro* καθώς και φορείς έκφρασης γονιδίων σε βακτήρια, σακχαρομύκητα και κύτταρα θηλαστικών.

➤ Απλοί φορείς κλωνοποίησης

- **pGEM-T Easy** (Promega, #A3600): Παρέχεται σαν γραμμικός φορέας και είναι κατάλληλος για την κλωνοποίηση προϊόντων PCR καθώς στα 3' άκρα του φέρει προεξέχοντα νουκλεοτίδια θυμιδίνης που υβριδοποιούνται με τα προεξέχοντα νουκλεοτίδια αδενίνης που προστίθενται από πολλές Taq DNA πολυμεράσες στα 5' άκρα των προϊόντων PCR. Επιπλέον, έχει πολλαπλές επιλογές στον polylinker του και περιλαμβάνει τους υποκινητές T7 και SP6 που επιτρέπουν την αλληλούχηση με τη χρήση των αντίστοιχων εκκινητών και τις αντιδράσεις μεταγραφής *in vitro* με τις αντίστοιχες RNA πολυμεράσες. Επίσης, περιλαμβάνει το γονίδιο *lacZ* για χρωματική επιλογή των ανασυνδυασμένων κλώνων.
- **pBluescript II KS +** (Stratagene, #212207): Χρησιμοποιήθηκε για την κλωνοποίηση ενθεμάτων DNA καθώς έχει πολλαπλές επιλογές ένθεσης DNA στον polylinker του, φέρει το γονίδιο *lacZ*, καθιστώντας εύκολη την επιλογή των ανασυνδυασμένων κλώνων και περιλαμβάνει τους υποκινητές T3 και T7 που

επιτρέπουν την αλληλούχηση με τη χρήση των αντίστοιχων εκκινητών και τις αντιδράσεις μεταγραφής *in vitro*.

➤ **Φορείς έκφρασης σε βακτήρια**

- **pGEX** (Amersham Biosciences, #27-4803-01): Φορέας επαγόμενης έκφρασης χιμαιρικών πρωτεϊνών με την GST στο N-τελικό τους άκρο. Η έκφραση των πολυπεπτιδίων ελέγχεται από τον υποκινητή *lac* στον οποίο είναι προσδεδμένος ο καταστολέα $IacI^q$, που παράγεται επίσης από τον φορέα. Σε περίπτωση επαγωγής με IPTG, ο $IacI^q$ συνδέεται με αυτό, απελευθερώνοντας έτσι τον υποκινητή. Η ανασυνδυασμένη πρωτεΐνη ανιχνεύεται με το anti-GST και απομονώνεται με χρωματογραφία συγγένειας με σεφαρόζη-γλουταθειόνη.
- **pET-26b(+)**(Novagen# 69862-3): Η έκφραση των χιμαιρικών πρωτεϊνών είναι υπό τον έλεγχο του υποκινητή T7. Το παραγόμενο πολυπεπτίδιο φέρει στο N-τελικό του άκρο την αλληλουχία *pelB*, που είναι σήμα για τον εντοπισμό του στον περιπλασμικό χώρο, και στο C-τελικό του άκρο τον επίτοπο 6-His, που επιτρέπει την ανίχνευση με το αντίστοιχο αντίσωμα και την απομόνωση με σεφαρόζη-Ni-NTA.

➤ **Φορείς έκφρασης σε σακχαρομύκητα**

- **pGBT9** (BD Bioscience Clontech, #K1605-A): Τα χιμαιρικά μόρια παράγονται σε σύντηξη με την περιοχή πρόσδεσης στο DNA (DNA-BD, 147 αμινοξέα) του παράγοντα GAL4 στο N-τελικό τους άκρο. Η υβριδική πρωτεΐνη εκφράζεται στα κύτταρα του σακχαρομύκητα σε σχετικά χαμηλά επίπεδα από τμήμα του συστατικού υποκινητή ADH1 και κατευθύνεται στον πυρήνα λόγω σήματος πυρηνικού εντοπισμού που φέρει η αλληλουχία DNA-BD. Ο φορέας αντιγράφεται αυτόνομα στην *E. coli* και στον *S. cerevisiae*. Φέρει το γονίδιο *bla* (για ανθεκτικότητα στην αμπικιλίνη στην *E. coli*) και το *TRP1* που επιτρέπει στα μετασχηματισμένα κύτταρα του σακχαρομύκητα να αναπτύσσονται σε θρεπτικό μέσο από το οποίο απουσιάζει η τρυπτοφάνη.

➤ **Φορείς έκφρασης σε κύτταρα θηλαστικών**

- **pEGFP** (BD Bioscience Clontech, #6084-1): Οι χιμαιρικές πρωτεΐνες φέρουν την EGFP είτε στο αμινοτελικό τους άκρο (αν χρησιμοποιηθούν οι φορείς C1-3) είτε στο καρβοξυτελικό τους άκρο (για τους φορείς N1-3) επιτρέποντας έτσι τον υποκυτταρικό εντοπισμό τους και την ανίχνευσή τους με το αντίστοιχο αντίσωμα. Η EGFP είναι ιδιαίτερα σταθερή πρωτεΐνη και ο φθορισμός της διατηρείται για μεγάλα χρονικά διαστήματα. Ο φορέας έχει τον υποκινητή CMV που παρέχει υψηλά επίπεδα

έκφρασης και την θέση έναρξης της αντιγραφής του ιού SV40 που του επιτρέπει να πολλαπλασιάζεται σε κύτταρα θηλαστικών που εκφράζουν το αντιγόνο T.

- **pcDNA3.1 His** (Invitrogen, #V38520): Οι χιμαιρικές πρωτεΐνες εκφράζονται σε υψηλά επίπεδα λόγω του υποκινητή CMV σε σύντηξη με τους επιτόπους 6-His και Xpress στο αμινοτελικό τους άκρο. Έτσι είναι εφικτός ο εντοπισμός τους είτε υποκυτταρικά είτε *in vitro* με χρήση των αντίστοιχων αντισωμάτων. Επιπλέον, ο φορέας έχει την θέση έναρξης της αντιγραφής του ιού SV40 και έτσι πολλαπλασιάζεται σε κύτταρα θηλαστικών που εκφράζουν το αντιγόνο T.
- **pcDNA3.1 (+) myc-His** (Invitrogen, #V80020): Έχει τα ίδια χαρακτηριστικά με τον pcDNA3.1 His με αντικατάσταση του επιτόπου Xpress με το Myc.

2.10. Πλασμιδιακές κατασκευές

Παρατίθενται τα ανασυνδυασμένα πλασμίδια που χρησιμοποιήθηκαν και ο τρόπος κατασκευής τους, ταξινομημένα ανά γονίδιο που αναλύσαμε:

➤ **Lgi1-4**

- **Lgi1/pBluescript KS+**: PCR σε 50ng γενωμικού DNA ποντικού με τους εκκινητές mLGI1F-mLGI1R και κλωνοποίηση του προϊόντος (434bp) στη *SmaI* του pBluescript KS+. Το ένθεμα αντιστοιχεί στα νουκλεοτίδια 1267-1700 του cDNA (NM_020278).
- **Lgi2/pBluescript KS+**: Τεμαχισμός του κλώνου IMAGE #313745 (W10952) με *XmnI*, απομόνωση του ενθέματος (380bp) και κλωνοποίησή του στη θέση *SmaI* του pBluescript KS+. Το ένθεμα αντιστοιχεί στα νουκλεοτίδια 1463-1844 του cDNA (NM_144945).
- **Lgi3/pBluescript KS+**: PCR σε 50ng γενωμικού DNA ποντικού με τους εκκινητές mLGI3F-mLGI3R1 και κλωνοποίηση του προϊόντος (366bp) στη *SmaI* του pBluescript KS+. Το ένθεμα αντιστοιχεί στα νουκλεοτίδια 1690-2055 του cDNA (NM_145219).
- **Lgi4/pBluescript KS+**: PCR σε 50ng γενωμικού DNA ποντικού με τους εκκινητές mLGI4F και mLGI4R. Το προϊόν (733bp) κλωνοποιήθηκε στον pGEM T-Easy από όπου και εξήχθη με *SacII* και *Sall* για να κλωνοποιηθεί στις αντίστοιχες θέσεις του pBluescript KS+. Το ένθεμα αντιστοιχεί στα νουκλεοτίδια 1680-2412 του cDNA (NM_144556).

➤ **3077**

- **3077/ pcDNA 3.1 Myc.His:** PCR με τους εκκινητές MycFnew-MycR, στον κλώνο που απομονώθηκε από την βιβλιοθήκη και αντιστοιχεί στο μετάγραφο 1 (Εικ.11), τεμαχισμό του προϊόντος με *BamHI* και *EcoRI* (θέσεις στους εκκινητές) και κλωνοποίηση στις αντίστοιχες θέσεις του pcDNA 3.1 Myc.His. Το ένθεμα αντιστοιχεί στα νουκλεοτίδια 234-1373 του cDNA (AJ294945).

➤ **AJ312051**

- **AJ312051/ pBluescript KS+:** PCR με τους εκκινητές F_ *BamHI* και R_ *Kpn* σε RT από νεφρό ανθρώπου (MTC cDNA panels, Clontech), τεμαχισμό του προϊόντος με *BamHI* και *Kpn* (θέσεις στους εκκινητές) και κλωνοποίηση στις αντίστοιχες θέσεις του pBluescript KS+. Το ένθεμα αντιστοιχεί στα νουκλεοτίδια 352-1359 του cDNA (AJ312051).
- **AJ312051/ pEGFPC1:** Τεμαχισμός του κλώνου AJ312051/ pBluescript KS+ με *BamHI* και *KpnI*, απομόνωση του ενθέματος και κλωνοποίησή του στις αντίστοιχες θέσεις του pEGFPC1. Το ένθεμα αντιστοιχεί στα νουκλεοτίδια 352-1359 του cDNA (AJ312051).

➤ **FRA10AC1**

- **FRA10AC1_pGEM_T_Easy:** PCR με τους εκκινητές F και R4 σε RT από όρχεις ανθρώπου (MTC cDNA panels, Clontech) και κλωνοποίηση του προϊόντος στον pGEM_T_Easy (προσανατολισμός: το 5' προς τον T7). Το ένθεμα αντιστοιχεί στα νουκλεοτίδια 189-1137 του cDNA.
- **FRA_full_pGEX_4T_1:** τεμαχισμός του κλώνου FRA10AC1_pGEM_T_Easy με *BamHI* και *EcoRI* και κλωνοποίηση του ενθέματος στις αντίστοιχες θέσεις του pGEX-4T-1. Το ένθεμα αντιστοιχεί στα νουκλεοτίδια 189-1137 του cDNA.
- **FRA_BgIII_EcoRI_pGEX_4T2:** τεμαχισμός του κλώνου FRA10AC1_ pcDNA 3.1_His B με *BgIII* και *EcoRI*, αντίδραση γεμίματος και κλωνοποίηση στη *Sma* του pGEX-4T-2. Το ένθεμα αντιστοιχεί στα νουκλεοτίδια 633-1138 του cDNA.
- **FRA_BbsI_Stu_pGEX4T_1:** τεμαχισμός του κλώνου FRA_full_pEGFP_N1 με *BbsI* και *StuI*, απομόνωση του ενθέματος, γέμισμα των άκρων και κλωνοποίησή του στη *SmaI* του pGEX-4T-1. Το ένθεμα αντιστοιχεί στα νουκλεοτίδια 479-960 του cDNA.
- **FRA_full_pET26b+:** PCR με τους εκκινητές F5 και R5 στον κλώνο FRA10AC1_pGEM-T Easy, τεμαχισμό του προϊόντος με *BamHI* και *EcoRI* (θέσεις στους εκκινητές) και κλωνοποίησή του στις αντίστοιχες θέσεις του pET26b+. Το ένθεμα αντιστοιχεί στα νουκλεοτίδια 197-1138 του cDNA.

- **FRA_full_pGBT9:** Τεμαχισμός του κλώνου FRA10AC1_pGEM_T_Easy με *EcoRI*, απομόνωση του ενθέματος και κλωνοποίησή του στην αντίστοιχη θέση του pGBT9. Το ένθεμα αντιστοιχεί στα νουκλεοτίδια 189-1137 του cDNA.
- **FRA_N-terminal_pGBT9:** Τεμαχισμός του κλώνου FRA_full_pGBT9 με *PstI*, απομόνωση του ανασυνδυασμένου φορέα με το παραμένον τμήμα του FRA10AC1 (νουκλεοτίδια: 189-953 του cDNA) και επανακυκλοποίησή του.
- **FRA_C-terminal_pGBT9:** Τεμαχισμός του κλώνου FRA_full_pGBT9 με *PstI*, απομόνωση του θραύσματος 185bp και κλωνοποίησή του στη θέση *PstI* του pGBT9. Το ένθεμα αντιστοιχεί στα νουκλεοτίδια 953-1137 του cDNA.
- **FRA_full_pEGFP_N1:** Τεμαχισμός του κλώνου FRA10AC1_pGEM_T_Easy με *BamHI* και *EcoRI*, απομόνωση του ενθέματος και κλωνοποίησή του σε *BglII-EcoRI* του pEGFP_N1. Το ένθεμα αντιστοιχεί στα νουκλεοτίδια 189-1137 του cDNA.
- **FRA_1_BglII_pEGFP_C1:** Τεμαχισμός του κλώνου FRA_full_pET26b+ με *BamHI* και *BglII*, απομόνωση του θραύσματος ~450bp (νουκλεοτίδια 198-636 του cDNA), αντίδραση γεμίματος των προεξέχοντων άκρων και κλωνοποίησή του στον pEGFPC1 που έχει τεμαχιστεί με *Xho* και έχει υποστεί αντίδραση γεμίματος. Το ένθεμα αντιστοιχεί στα νουκλεοτίδια 197-636 του cDNA.
- **FRA_BglII_Pst_pEGFP_C1:** Τεμαχισμός του κλώνου FRA_BglII_EcoRI_pGEX_4T2 με *BamHI* και *PstI*, απομόνωση του τεμαχίου ~320bp (νουκλεοτίδια 633-954 του cDNA) και κλωνοποίηση στις θέσεις *BglII* και *PstI* του pEGFPC1.
- **FRA_BbsI_Stu_pEGFPC1:** Τεμαχισμός του κλώνου FRA_BbsI_Stu_pGEX_4T1 με *Sall*, αντίδραση γεμίματος, τεμαχισμός με *BamHI*, απομόνωση του ενθέματος και κλωνοποίησή του στις θέσεις *BglII* και *SmaI* του pEGFPC1. Το ένθεμα αντιστοιχεί στα νουκλεοτίδια 479-960 του cDNA.
- **FRA_BbsI_BglII_pEGFP_C1:** Τεμαχισμός του κλώνου FRA_BbsI_Stu_pEGFPC1 με *BglII* και *BamHI* απομόνωση του τεμαχίου ~5kb που περιλαμβάνει τον φορέα και το τμήμα του cDNA που αντιστοιχεί στα νουκλεοτίδια 479-636 και επανακυκλοποίησή του.
- **FRA_Pst_end_pEGFP_C1:** Το μετάγραφο «α» του FRA10AC1 που ορίζεται από τους εκκινητές F3 και R2 είχε κλωνοποιηθεί στον pGEM_T_Easy. Από εκεί απομονώθηκε με *EcoRI* και κλωνοποιήθηκε στην ίδια θέση του pEGFPC1. Το ένθεμα αντιστοιχεί στα νουκλεοτίδια 951-1252 του cDNA.
- **FRA10AC1_pcDNA 3.1_His B:** Τεμαχισμός του κλώνου FRA_full_pET26b+ με *BamHI* και *EcoRI* και κλωνοποίηση στις αντίστοιχες θέσεις του pcDNA 3.1 His B. Το ένθεμα αντιστοιχεί στα νουκλεοτίδια 197-1138 του cDNA.

- ***Fra10ac1_pBluescript KS+***: PCR με τους εκκινητές mFRAF-mFRAR σε 50ng DNA βιβλιοθήκης cDNA εμβρύου ποντικού αναπτυξιακού σταδίου E10.5 και κλωνοποίηση του προϊόντος (494bp) στη *SmaI* του pBluescript KS+. Το ένθεμα αντιστοιχεί στα νουκλεοτίδια 250-743 του cDNA του ποντικού (BN000292).
- ***DGCR14***
 - ***DGCR14_pcDNA3.1_HisA***: Τεμαχισμός του κλώνου DGCR14_pACT2 με HindIII, αντίδραση γεμίματος, τεμαχισμός με BamHI, απομόνωση του ενθέματος και κλωνοποίησή του στις θέσεις BamHI και EcoRV του pcDNA3.1_HisA. Το ένθεμα αντιστοιχεί στα νουκλεοτίδια 83-1422 του cDNA (Genbank Id: NM_022719).
 - ***DGCR14_pEGFP_C1***: Τεμαχισμός του κλώνου DGCR14_pACT2 με BamHI και BglII, απομόνωση του ενθέματος, αντίδραση γεμίματος και κλωνοποίηση στη θέση *SmaI* του φορέα pEGFP_C1. Το ένθεμα αντιστοιχεί στα νουκλεοτίδια 83-1698 του cDNA (Genbank Id: NM_022719).
 - ***DGCR14₂₅₋₄₇₂_pGEX 3X***: Τεμαχισμός του κλώνου DGCR14_pACT2 με HindIII, αντίδραση γεμίματος, τεμαχισμός με BamHI, απομόνωση του ενθέματος και κλωνοποίησή του στις θέσεις BamHI και *SmaI* του pGEX 3X. Το ένθεμα αντιστοιχεί στα νουκλεοτίδια 83-1422 του cDNA (Genbank Id: NM_022719).
 - ***DGCR14₂₅₋₂₂₂_pGEX 4T-2***: Τεμαχισμός του κλώνου DGCR14₂₅₋₄₇₂_pGEX 3X με EcoRI, απομόνωση του τεμαχίου ~600bp και κλωνοποίησή του στη θέση *EcoRI* του pGEX 4T-2. Το ένθεμα αντιστοιχεί στα νουκλεοτίδια 83-673 του cDNA (Genbank Id: NM_022719).
 - ***DGCR14₂₂₃₋₄₇₂_pGEX 3X***: Τεμαχισμός του κλώνου DGCR14₂₅₋₄₇₂_pGEX 3X με EcoRI, απομόνωση του τεμαχίου ~750bp και κλωνοποίησή του στη θέση *EcoRI* του pGEX 3X. Το ένθεμα αντιστοιχεί στα νουκλεοτίδια 674-1422 του cDNA (Genbank Id: NM_022719).
 - ***DGCR14₂₉₄₋₄₇₂_pGEX 3X***: Τεμαχισμός του κλώνου DGCR14₂₅₋₄₇₂_pGEX 3X με *XhoI* και *Sall*, απομόνωση του τεμαχίου ~5,4kb, που περιλαμβάνει τον φορέα και μέρος του ενθέματος, αντίδραση γεμίματος και ακόλουθη αντίδραση επανακυκλοποίησης. Το ένθεμα αντιστοιχεί στα νουκλεοτίδια 889-1422 του cDNA (Genbank Id: NM_022719).
 - ***DGCR14₂₁₈₋₃₁₆_pGEX 3X***: PCR χρησιμοποιώντας τους εκκινητές DGSI_F1_BamHI και DGSI_R2_EcoRI και υπόστρωμα DNA της πλασμιδιακής κατασκευής DGCR14₂₅₋₄₇₂_pGEX 3X. Το προϊόν, μετά τον τεμαχισμό του με *BamHI* και *EcoRI* (θέσεις στους εκκινητές) κλωνοποιήθηκε στις αντίστοιχες θέσεις του φορέα pGEX 3X. Το ένθεμα αντιστοιχεί στα νουκλεοτίδια 679-957 του cDNA (Genbank Id: NM_022719).

- **DGCR14₃₉₀₋₄₇₂_pGEX 3X:** Τεμαχισμός του κλώνου DGCR14₂₅₋₄₇₂_pGEX 3X με *Sall*, αντίδραση γεμίματος, τεμαχισμός με *StuI*, απομόνωση του τμήματος ~5,2kb και επανακυκλοποίησή του με αντίδραση σύνδεσης. Το ένθεμα αντιστοιχεί στα νουκλεοτίδια 1176-1422 του cDNA (Genbank Id: NM_022719).
- **Dgcr14_pBluescript KS+:** Τεμαχισμός του κλώνου IMAGE 3708646 (Genbank ID BC013711) με *Pst*, απομόνωση του τεμαχίου ~1,1kb και κλωνοποίησή του στην αντίστοιχη θέση του φορέα pBluescript KS+. Το ένθεμα αντιστοιχεί στα νουκλεοτίδια 1425-2587 του cDNA BC013711.

➤ **SAP145**

- **SAP145/pcDNA3.1 HisA:** Τεμαχισμός του κλώνου SAP145_pACT2 με *BamHI*, απομόνωση του ενθέματος και κλωνοποίησή του στην *BamHI* του pcDNA3.1 HisA. Το ένθεμα αντιστοιχεί στα νουκλεοτίδια 508-2484 του cDNA (Genbank Id: U41371)
- **Sap145_pBluescript KS+:** Τεμαχισμός του κλώνου IMAGE 3582542 (Genbank ID BC029745) με *StuI* και *BamHI*, απομόνωση του τεμαχίου ~720bp και κλωνοποίησή του στις θέσεις *EcoRV* και *BamHI* του pBluescript KS+. Το ένθεμα αντιστοιχεί στα νουκλεοτίδια 2345-3069 του cDNA BC029745.

2.11. Αντιδραστήρια

Τα χημικά που χρησιμοποιήθηκαν παρασκευάστηκαν από τις εταιρίες Merck, Sigma και Scharlaw.

Πιν. 4: Οι εκκινητές των STSs της χρωμοσωμικής περιοχής 10q23.3-q25.3 (A) και 10q26.2-qter (B)

A. 10q23.3-q25.3
1.ΧΑΡΤΗΣ GM99-GB4

STS	ΑΛΛΗΛΟΥΧΙΑ	ΕΚΚΙΝΗΤΩΝ	ΜΕΓΕΘΟΣ	ΘΕΣΗ
	FORWARD	REVERSE		
Z40596	GGCCATGACTTAATCATAC	GAGAGGGAGGAGAGAGA	66	428.51
F09841	CCCATGACTTACCCTAAATCTCC	AATGTCTCATCCATGCACAG	126	428.86
stSG39758	AAAAGCCCACATTTGAGGTG	CGGACCCATAGTGGAAATGT	158	428.86
stSG41319	TGTTGAAAGGGATTTTGAAGG	CAGGCTTTATTTCAACCAGACA	154	428.86
stSG32258	CTTGACACAGAGCAACCCAA	CACCGTCAACATCAGACACC	168	428.86
stSG21645	CACAACAAACAGAGCCCTT	GCGGGCATACCAAGTGTC	129	428.86
stSG29905	GAGCAGTTGTTTCGCTACAGG	GACACGCTGTGGCTCATCT	170	428.86
T92735	CATGCCGAAGCTATAAGTGAG	AATGGTGACTGCCTAAACAAC	125	428.86
H84912	CTACAAGACCGTTAGACTAGAGAG	ATATTGACTTTTTGCCCTTCATAG	160	428.86
stSG45558	GCTTTGTTTCGAAGAATGGTG	TTCCATGAAGTTGATGCCAA	120	428.86
WI-18437	CAATTTTCAGTGGGTTTTATAATCA	GATCTCTCAGGGATATCCATCC	150	428.86
H92921	AGGTTGCCTCAGTTTACTGG	CCCTCCGAGTTTTACAAGAA	177	428.86
R86942	TTACGCCAAGAGGACTGTAA	CCTAAATGAGACATCTGAGGACATA	198	428.86
stSG3666	AACTTTGTTAAGCAGCTGAGGG	GTTTTGTAATCTGTGAGCACC	166	429.01
stSG30644	GGATAATCAAGTTTTGGCACA	AATTCCTCACACCTTGCACC	154	429.01
stSG16023	CCTTAGAGTATGCAAGCACGC	TGTGGCTGTCTCCCAATGT	207	429.18
stSG41741	TCTGACACCTCAACCCATGA	TCCCATGCCACAGAATATGA	131	429.18
WI-11734	CATTTATTGAGTGCTTACATTTTGC	TCTTTTATAGCCCAAGGTAGATGG	125	429.18
SGC38164	GGTGGGAAGTGGAGAAACA	GATTCTTTGGGAGCCTTCC	252	430.71
stSG15752	CATCAGAAGCAACATCTCTTGC	ATTCTCCAGGGTCACACAG	214	430.81
WI-17128	TAAGGCCCGTGTCTGTTTTTC	ACCCATACCTTAAACATTTTGTCC	125	430.81
stSG2551	TCTTTCAATTGACAGAGATGCC	GCTAGCTCCTCCCTGGTTG	166	432.51
WI-16578	AAAGTCAAAATTACAGAAAGAAGCA	TAAAGAATTATCCGTTTGTTTTTCC	101	432.51
A008Y13	CAGGGAATCTTGTATATTTTTG	CATATCCTCTTTTCCAATATTGT	107	432.54
stSG46580	ATCAACAAAACCAACAACCTCCC	TGTTGTTGGGAAAATAATGTG	129	432.65
L25676	TATGTGACTTGCATCGTGG	AACCCCTGGGAACCATC	187	432.92
A004R39	TCAAAGAAACTACATTAGAGAAG	TAGATACTCCAGGAAGAGTTAA	139	433.13
WI-11615	TAGAGTTTATTTGAGCACAAAGCTT	TGCAACCCTGTTTACTGCCT	127	433.13
stSG26781	TTGTCTAAGTGAGAGATTGCGA	TGCTTAGTTCTCTGTCTGCA	149	433.33
stSG22113	GCACCTGATGACCTCAAAAT	GGAAAATGTTGCCTGTTGG	186	433.33
W73720	TGCATAGCATTTACACACAGAG	ACAAAAGTGCCTCGTTTACC	262	433.63
D10S2211	ATCTGGGGACAGGGCTTC	TGTCTGTATTTGGCAGATGAGG	298-299	433.63
stSG32004	ATGGGGACTCACTATGTTCCC	ATTTGGTATTTTCACGCACG	164	433.68
L16782	GCCAAGTGATTTAATGGACC	ACAGGATGCAAGTCCAATG	148	433.73
stSG2721	CTTTCACGGCATTGACAGA	TTGGACAATTAGCAACAAGTCA	190	433.83
stSG21644	TGGTACTGAGTGACTTTGGGC	CATTCATTGAGGGCTGTTT	162	433.83
D10S2197	GAATGTTCTGGCATTATTTCCG	AGAGAGTCAGAGGCTGTGCC	275-276	434.34
stSG12880	GAATGAGTGGGCGAGTGACT	TGAGGGAAGAAGTGAAGTCA	112	434.75
stSG8438	AATGAAATATGATGTCTGGGGG	AGCCATGCTATCTAATATGGCA	125	434.75
stSG47193	AAATGTATTTAACCCAGCAGCA	AGCCCCAGAGACAGAAATT	146	436.66
stSG39120	CTGTGGGTGGAGAGACATT	TTCCGGTTGATATTCGGTA	130	438.37
H10009=stSG55704	TTGTTACTTACGTTGCTTTATTTTC	CTAACTCCCCATTGACATTG	266	438.68
SHGC-35401	GGCATTTTGAAGGCATGG	CCAGATGGATGATCATGAAGG	222	439.09
A007J42	ATAGACTATTGATTTCTAGCTCC	TATCTCTACAGTGAAGAAACCT	207	440.02
stSG31377	GGAGGCATTCATTCAAGGA	TCCCCTTGACCTCAGGATAC	136	440.02
N37073	AGTCAGGTCTCCTGGAAGC	CTCGGTGGTACAATTTATGC	184	440.02
stSG15491	ATTTAGGATTACCCTTCCCACC	GAAAGCTTCTTACTGGCCTCA	191	440.02
stSG8384	GTTTATTGCAGGTTGGAGAAGG	GGAGACCCTCTGAGGAGCTT	175	440.02
stSG48047	TTGAGCAAATCAATTCTGCA	TACCCAAGTGAAGAGAACGTCA	87	440.02
NIB351	GCTCACACAAAGGAAACTT	ACCTTAAAGCTCATTCTGG	193	440.64
WI-16054	TTTTTTTGCAGTTCTAAAACAGATT	TTTAAAACACAGCATTTACACTGAA	126	440.85
WI-14114	TTAATATACACGGGAAATTTGACA	TATGTGATTATAATCTGGCTTCTCA	103	441.05
stSG35189	CTGCCAGGATGTTATATTTCCC	AAATATTGGAAGGAGTGGCC	121	441.26
stSG44990	ACACTGTGATCAATTTGCTCCC	ATATGTGCCCTTCAATACCC	160	441.26
stSG31710	TGTTCCATGAACATGGTACATC	AGCAAGCAACAAAATCCAGG	124	441.36
stSG50791	TGCCAATCATTCCTGTTTA	GGAGAAGATGCTTGACTTGATG	121	441.36
stSG30011	GTGAGCAAACATCTCTGTTTCG	CAACAAATCTTTGTTGCTGAGG	137	441.36
Z40200	TGCACAGCTCTTTTCAG	TTGGAATTTGCTAAATAGA	85	441.43
stSG51671	CCAGTATGGAGCATGCC	AGGTGCGCTTATGGAGAGAA	161	441.43
stSG12559	TGGAAAAGTTGGAATGTGTGA	GCTTCTGAGTGGGAACAGG	115	441.46
A006X22	ACCAAAAATAAGGCAGAT	ACAAAATTTCTGAACCTCTTT	207	441.46

Πιν. 4: (συνέχεια)

STS	ΑΛΛΗΛΟΥΧΙΑ	ΕΚΚΙΝΗΤΩΝ	ΜΕΓΕΘΟΣ	ΘΕΣΗ
	FORWARD	REVERSE		
stSG2803	GGACCTGTGTTACCCAGGG	TAAATAAATGTGCCCCCAA	130	441.46
stSG35957	GTACAGCTGAGGTCTCGATGGTTG	CATGCCCTGTGCATACATATG	133	441.56
stSG22466	GGCCAATAACATTGATAGAGC	GGTGAGCCTCTTAGCCTCCT	183	443.37
stSG13323	CAGGTAAGGGAGGGTGGG	TCCTAGCCGCAATAGCATG	113	443.37
stSG30492	AAAACCCACACAGAGGATTC	AGTGTGGAACTCTTCCCT	173	443.37
WI-16821	TATTGTCAATGGTTTGTTCATGTT	TATTTCTGTGGGAATCAAAGTGG	125	443.37
L16499	CCTGTTTGACCAAGGTGTT	CCACGGACGGCTTTAT	238	443.37
stSG25928	CCAGTCAATGCAACACATACA	TGCACAATTAAGCCTGTATT	190	443.37
A006W20	TCCAGAGATGTACATATAATTCT	CTTTATCCTCATTGTGGA	238	443.37
stSG9628	CGTAAAACTTCTAATCATACAG	AGCATTACACAGTATTTATCAGT	103	443.37
U37426	TTCTGCTCAGCATGAGTTTGTAG	ATGCTGACATCTGATTTTGC	178	443.37
A008Z38	TAAGTAAGAAACAAGGTATCGCT	TGTTTGTGTTTGTCCACTATAT	246	445.08
stSG29640	GATTTTTCTGTGCTCATGCC	TTACTTTCTTGATGATGCTGCC	129	445.19
D29519	CATTTAGATGGTCCAAGCTG	CATACTGGCTGTGCACATGC	164	445.19
stSG50462	GTACCTAAAGCCCATGAAGTGC	TCTTGGATACAGGCTGATGTG	122	445.19
stSG21346	TGGCAGCAAACTCTAAGG	GCATTTCCCAATGAAGAAGG	148	445.19
SGC30793	CGTTGAGGCAAGAAAAGGCT	CAAGTTCAAATCCTTACCAGG	150	445.19
A009B31	TAAAAATGCTACTATTGGGA	TCTTGTCTTTCAAATCAAA	169	445.19
H90420	CATGTAGAATTTGAGAACTGAAC	TACAAATAAGCATTGAAGTTAGCTC	202	445.19
stSG66850=L07093	ACAGATGGTCTGGCTGCT	TGAGGGTGAATGTTCACT	138	445.19
N29485	CCTAGATGGTCTCTTACAGC	AATGGTTATTATCAGAGATTTTCC	123	445.29
stSG15866	AACATGCAGAAAAGAAATGCC	CCTCCATTACGGAGAGTTTCC	191	445.29
WI-16274	TATACAGAAGCCCTGAAAAGGCA	CTTGGACTGAGCCTGATTTCC	150	445.29
stSG43195	TCATCACGTAACAGGATCTTCA	AAATTTGGCACCAAAATGTACC	144	445.29
stSG45302	TCATCACGTAACAGGATCTTCA	AAATTTGGCACCAAAATGTACC	144	445.29
stSG3745	CGCCTTCATTGGAATCTAGC	GAAGTGGAGGGACATTCTTCC	123	445.29
N22119	GGAACTTGATCTGAAGATTG	AAATTAGAACTTGAAAACCTACAC	121	445.29
A004S47	GAGTTATCCGATAGTACTTTTC	GTAGGATCTTAAGAAAAGAGTC	156	445.29
H03262	AAATCTGAAATAGCTCATTGGA	GGCAGAGACTTTGGTAGGAA	175	445.39
stSG21190	ACCAAAAGAAAAGCAGAGATGC	CTACCCAGCAGGACTGAGTTG	150	445.39
stSG22312	TCTGCATAGGACATGCTCTCA	TTTGGTTTTTGGTTTTTGC	177	445.91
A002E05	CTCTCTTGAAAGTTCTCTG	GTGTCAGACACGTTATATTTG	243	446.81
stSG35884	GACGGGCAAATTTGGTGAGGCATAC	TTTTTCAGGGGTGCTTCTTG	160	446.81
stSG57727=1284	CTTAGCCTGAGGTCCGGAAG	CTATTCCTGGATACCTGCAG	76	446.96
SGC33162	ATGGTGTACTATAGGGATTACGACA	TTGAATTGCCTTGGCAACCT	125	446.96
stSG31812	GAGCAGATCGTAGCACACCA	CATTGTCTTGTCTCTCCAAA	179	446.96
stSG35992	GCCAAGGAAGCTTCACACTCGTC	ATATGCTGGCTGGAGTGAGG	125	446.96
stSG47335	CAACCTACCACCTAGAATTGC	GGAAAGACCTTTGTCTGAGGG	137	447.69
stSG66857=SGC30780	TAAAGCATGCCTACTCTACAGC	CTACTGAGAATACACAGTGTAGGGC	125	448.35
WI-17750	GTTGTATTGGTAGATCTTGGTTGC	GGATGGTGGTGTGAAATA	150	448.35
N27530	TGGATGGAGCCTTCACT	TATGGGTCTCTTGCCTATCA	159	448.52
WI-16989	TTGACTCGTAATCACACAGCTT	CGCCCTTGACACGTAGGG	150	448.52
stSG46053	TTTTTTTGCAGAAGAACACATGG	CCCCAACTTTTATCTTTCCA	155	448.63
stSG4	GTGCTGGCCCTCATAGTGT	AGGTTACAGCTGCTCCTGGA	138	448.63
A006A33	ATACTTGAAAGGTAAGAAA	TCAGGTTATACTACCTGACT	127	448.63
M21941	TCCCTGTCTGAAGAATGCT	GCAGTGACCTGAACAACCTCTC	177	448.63
N95450	TGAAAGTACTAAAGCCCATGAAC	GCCTTACCAAGACAAAAGATATCC	196	448.63
N32898	TCTCTGTTTTCATCTAAGGACTG	TCTTAATCTATGGAAATAATGC	126	448.63
H68323	GCGCAAGGATGACATATAC	TGGCCCAACATATACAGG	181	448.63
SGC31032	CAGCTCTTTCGGGAACTG	CGCTGTTTACATAACCTCC	150	449.25
W86789	ACAGTGTTTGACCAGCTCC	ATTGGCTTGATCTCCACAG	182	452.37
H95763	CAACTTTATCGGTGATTCTGG	TTCTGAACCCATGTTAGAGG	208	452.37
stSG35515	AGGGCAGTTGTAGGGTCAA	TTGGAAGGGACTGTGGC	178	453.69
stSG45548	TCTCTGGTCATGAGCAGGTG	TGCTCAGTCTCTCCCATCTC	167	453.90
A004O25	TATAAATAATTCAGTAGGTCTGG	ATGATAATGTCATCTAGACCTTT	127	454.00
stSG462	TGAGGAGAAGAACTGCTTTCC	TTTGGAACAGAAAACCTGCTGG	106	454.00
SHGC-9747	GGTCTTATTGAAATGCAGATTGC	TATGCAACAGTCATAAACATGGC	110-111	454.84
AA022645	GCATATGTGAAAAGCAAAAATTACA	AGAACCAGAACCACAGCAGC	182	454.84
stSG2844	TGTAAGTTACAGTCTAGGTTTGG	CCTGGGGCTATTTCTCCCTA	121	455.67
stSG55875=F10974	AAAGGCAACCTTGTCTCTG	CCATGTCTGGGACAGATAGA	82	455.78
WI-15944	TGAAACTGAAACGTATTTCTCTCC	AACCATCCTTCAGGTGTGATG	135	455.78
stSG46829	GAAGGCCTGGTGTCTAATTCC	TGTTAGGGGTTATTCTGCACG	150	455.99

Πιν. 4: (συνέχεια)

STS	ΑΛΛΗΛΟΥΧΙΑ	ΕΚΚΙΝΗΤΩΝ	ΜΕΓΕΘΟΣ	ΘΕΣΗ
	FORWARD	REVERSE		
stSG2296	AAATGACAGAAATGCTGAACCT	GCATGTGACTAAGCCATGTAGC	144	456.09
stSG1032	GTGGGACTCTGTGATTGTGT	AGAAACTGCTTAATGGGGAC	134	456.20
A002G19	TAGAAATATGGACTCTTATGTTT	ATTACTAACTGTTGAGCCA	133	457.13
stSG38844	AGCTCAGAACTGGCGACG	CACCAGCCTGAGTAGAACTGC	124	457.44
stSG52092	CAAACCTCAGAGGCACTGAGG	GAATTCTGCAGAGCCTGGAC	150	457.75
WI-18064	CAGCTGCCAATCATCTCTCA	TTTCTGCCTCCCAGGTACC	138	457.86
stSG2761	CCCCATCCCCACTTTACAC	TCTCATAACAGGGGACCCGTA	150	458.27
stSG52467	CTGACAAGCCACTGGGCT	CCCTCCACACTGCAAGCT	130	460.07
stSG53348	GGAGGGTCACATTTTCTACTCG	CATGTGATGGACACTCCAGG	186	460.07
stSG47398	ATCTCACCTGCGCTGTGTG	GATGCTGTGCCCAAATCC	135	460.07
stSG44050	GAGCATTTGGCTCACTGTAGC	GAGGTGGGGGACGTATTTG	135	460.07
A009E24	AACTTTTTAAATCCTCCTT	ATTTTCTCATGCCATTAAC	148	460.07
R89840	GGCCAAAGTCTCCTTAGTGTTCAG	TGCATTCCAGCCACAACG	204	460.07
stSG21208	CTTGAGCCTGGAAGGTTGAG	TTACCTTTGGGTGGATGGAG	185	460.07
stSG51949	GCCTCATCTCTGCCAATTGT	CTCTGGTTTTCTGCTCAGGG	188	461.29
stSG52273	GTGGGTTCTTTCTGGTTAGC	CACGGAAGAACTGATGAGCA	170	461.29
stSG49136	TTACAAACAAGGAGTGCATTGG	AATCTAGGTGGCTTTTGGACTG	129	461.29
stSG45428	CCCTAGTAGGGAACCATGTCA	CCCATTTGAAACCAAACCTG	131	461.29
stSG32285	CTCTGACTTGTCTTAGGCCG	AGCAGGGGAAAAGACACATG	138	461.29
stSG30920	CTGGGCCTCAGTTAGCTCC	AGGGCCCTAATGATAGTGGG	144	461.29
stSG4569	AAACAATGAGGCACATCGGT	CCAAGACTACTCTGGGCAGC	153	461.29
WI-15882	TCTTTAATCGTCACATCAGCC	AAGAAAGTAAAAGCCTCTGTAGTC	150	461.29
stSG22607	CCTCCTTCTGGTTGCTCTTG	TTTGGCATTCTCTCATGC	142	461.29
WI-13377	GTGACTCAATGTCATGTTACAT	TGGGGGTTTCTGTAGTCGTC	150	461.29
stSG9987	GCCTAACCTAAAGGTAATTTTAT	GTTAAGTCCTTATGTGTGAATAA	217	461.29
WI-11858	CTTGATGCAAATGTTTTATTTG	ATTTACCCTCATGCTACTAATCA	140	461.46
stSG3077	TGTGAGAAGTGGCAGTGACC	TTGGCCACCACCTCTTCTAC	121	461.46
A005C16	CAGAGAATGTACAAAATTCT	TTATAGACAGTCTTTGTCAT	119	461.46
stSG4306	CATGTGTGTGGATGCAATGA	GTCACAGAGCGGATATTTAGCC	148	461.49
N20191	GTAGAAACCTCTCAGCCCC	CATGTGGCACAAGTAATGG	122	461.90
stSG15181	CCACAAAGACGTATCCAGAATG	CTGGCCAAAAGATGTCATT	207	462.62
SGC32394	AAAGGCTGTTAGTTTTGTTTTGT	TTTAGGTGGTATTCTTCAGCC	150	462.62
A008I22	TAATAAGAGCTTTCAAATGGAC	TGATCTAGTCAGAAGCTACTTCA	107	463.66
N35618	GAACCTTGATTTTATGCTTTG	GACTGGGAACCTGGATTTG	202	465.29
stSG46524	ATTTCTTGAGCATTGTGGTGG	TGAACCAGCTCTGACCCAG	121	465.39
T40917	AAGTAGAACACAATAGAATGGCTCA	TGTGAATGTGGGTGTCGG	256-257	465.39
stSG28313	ACAGCACAGGCCAGAAGC	AACCACCCTCAGAGTTGTG	169	465.39
stSG40840	GCCTCCAACCTCCTGCTCAA	AATTAACCCAAGGGCCATC	156	465.49
stSG26670	GCACAGGATCAGATGGCC	TTCTAGTTCCCCAGCCC	132	465.49
SGC34176	TTTATCATTAAAGCACACTGAAGGA	TACCTACGCCACCACAATCA	128	466.30
WI-13735	TTTTTTTTGCTGAAGACCAG	TGGATGATTTCTACGTGGATG	150	466.30
stSG8267	GTGTACATGGTGAGAGAGAGC	GAGGTGGCCTAGTAGGAGG	176	466.30
stSG30988	TGGGTGATAAAAATGAGGCC	GGTTAGTTGTGGCCAGCTGT	167	466.30
stSG44588	CAGGGGGACAGAGAATTTGA	AGCCATCGCATGGTCTGT	163	466.30
WI-16766	CATGAATGAGGAAGTCTGATGC	TTATTCAACAGATCTCAGCTTAGGA	102	467.33
stSG38907	AGTAAGCCCTGTGAGGAGAGC	GGCACTGTGGGTTTCAGG	128	467.38
stSG27214	GTGGCAATCGTCACCTCC	GATGGGCCCTCTTGTTTAT	129	467.38
A005X19	AATTTAGACTCAAGATTACAGTG	TCATTCTAGAGATGTTTTATAG	230	467.38
WI-8488	CATTATTTCTGAAAATGCTGAAA	CCTGGTTTTGCTGATCCAAT	226-227	467.38
stSG35096	TGAGACACCATTCTCTGGC	CAATCTTAGGCCTCTTGTCTG	123	467.38
SGC34978	AATAGTCTATGGCTACGGGCC	GGACCCCATTTTTCTCTCTG	126	467.43
A007K11	TTATTATAGTTGAGGAAACTGAG	CAATCTAGGATAGTGTAAAAGG	143	467.43
stSG3691	TTCTCTGAAACCTTGGGG	TGCTGCTGATGATGAGAGG	145	467.43
stSG1671	GTGGCAATGCGTTGTTTATG	TGTGTCCAGATGCTGTCAT	120	467.43
stSG16199	AATCCTCCACTCAGGCC	GGGGGTGTGCTGAAAACCTTA	152	467.43
stSG50801	AATTTTCATGGTTTCCATTCC	TGACATTCAAAGCCAGATTCC	137	467.98
H87438	GGAGAACTAAAGAACACAAAATTC	CCCCTGGGCCTTATCCTATA	116	469.37
W37897	TTTGAAAGAGCTACACTTGAGAAGG	TGCCAAAGGTTTACTTCCAGC	215	471.99
stSG49129	GGACTTTCACAGAAAAGGAAGG	AACATCCTCTCTGGTCTTTTC	147	471.99
stSG1953	TCAAGCAGGCATCTGATGAC	GGCATCCTTGTCTTTGAGC	147	472.60
stSG47056	TCTTGAGAGGCCTGAAGTT	TGTGATGAGGGAATCCCTTC	128	477.54
M14564	AGCTATTTGCCCTTCGGAG	CAGCCTTTAGGTGCTACCC	243	478.42

Πιν. 4: (συνέχεια)

STS	ΑΛΛΗΛΟΥΧΙΑ	ΕΚΚΙΝΗΤΩΝ	ΜΕΓΕΘΟΣ	ΘΕΣΗ
	FORWARD	REVERSE		
A004B36	GCAGTCAGTACTTAAGAGTCTAG	TAACAGCTTTCTCAAAGC	163	478.56
A006J02	GTTGAAGTGGTCTACTACAA	ATCTTTGAAACCAGAGAAG	229	478.63
stSG27406	ACCATGGAATGTACCTTTTGG	CTTGGAGATAGGCCCTAGGG	180	478.63
stSG48072	AACGAGGCCATAAACCTGG	GTTTGAGCTCTGCAGCCTCT	132	478.66
stSG47005	CGAGCTTCTCCTGGGGTAG	ATCCAAGGAGTCCCTGGC	125	478.66
stSG50987	GGCGTGGGATGAGACATC	CAGCCTCACTGGTACCTTCTG	151	478.66
A006D27	TCTCTAAGAACCTACATCT	TCAAATCTCTTTGGGATA	127	478.66
stSG2605	CATGCTCCTGTGGCAGTCC	TGTTTCCCACTCCATTAGCCCTG	140	478.66
stSG3084	AATCACATTTCCCAAGCAAATG	AAGTCTGTCACTACTGTGGGG	142	478.66
stSG51013	AGGCAAGGACCCAACTCC	TTGCCAATGATAGTGACCACA	147	478.66
stSG43839	TCCAAATAAAATATGGGCTTGG	GCTCTCAGAGGTTCCCACTG	143	478.66
stSG46535	TGTGTTTCTCCCACTGCTACC	ATCACAGTCCCACTCCCC	80	478.66
stSG42773	TCCAAATAAAATATGGGCTTGG	GCTCTCAGAGGTTCCCACTG	143	478.66
A004M06	ATCTGTATTTACTTCTAAGAGC	TACTCATGAACAGTGTGCG	104	478.66
stSG50235	CATGTGCTATGGTGGTTTGC	AATAAGGCCAGGCATGATTG	166	478.66
N39580	GCTTCAACAAGTCTGGGTC	AGAAACCTGGGCTGCTG	151	478.66
WI-16946	CAATGCTTTTTATTACACACCCA	TTTCTGGGCTCCATCAAT	131	478.66
H41409	TTGAACATTTAATAATGTGCTGAGC	GGACCATGGACAGTGTCTT	137	478.66
R01689	CAGTCATCTCACTAGCAGCTC	ACCCAGCACCTGTAAAG	186	478.66
AA034033	TGGTTAAAGACCAACTGCG	GTCCATAGGCCCATACTAGAGTTC	221	480.47
stSG38872	AATTGCCAAGCCAAGATC	CTCACGGCTTGACAGCAATA	128	480.47
A005K38	AATCTTACAGTCTAGAAGCA	TAGCATTTTTAATAGTGGA	177	480.47
N31939	GGACCCATTCTCTAAGGC	AAGGCTTGATCCAGCC	158	480.47
H82635	TCCACCCAGAATGCGATG	TTCACCAGCAAGGACCCGT	147	480.47
N37090	CGTAGACTCCCAACAAGCTG	TAAGAGCATTTCCAGGGC	132	480.57
R45921	AGAAGCCAGTCTCTGATGC	ATCTAACTTTTCGGGTGGA	132	482.15
W80406	TGTCATTTTCTTGTCTGCCC	ATCTAGATGCGCCACCC	222	482.20
stSG42602	TCCATATGTGGCCAGTACACA	TCCAGATTGAAATCATGGCA	122	482.20
WI-22164	TACACAACCTTTTGGGAGGCC	CATGCTGCCTGTTTAAAGGA	268	482.20
H19386	CAACTCTAGGATGTCTATACAGA	CAGCACAGCTAATTTTTTC	195	482.20
stSG2653	CCAGCTAAATTTTATTTTCGG	GTGGTTTCAGGGTTCAGTCTG	158	482.20
stSG48287	GAGGGTAAGCTAAATGGGACG	ATCTACCCAGAGATGCCCT	132	482.20
stSG22765	GAGACAGGGCTCCACCAG	ATGGCAGGATTCTCAGATGG	149	482.20
stSG3501	GTTAACTGCATAGAGTACCATAA	ATCAGAAGGGGATTTTTA	168	482.20
stSG40082	TCCCAAGTCGAAATCACCTC	CTCCCTATTGACCAATGACA	184	482.20
stSG4995	AGAGGCAGTTACATGGAAAACC	GTCCACAGCTTTCTGGAAGC	159	482.20
stSG12760	AGCAAACATGATGAAAATTCCA	AAGCAGGAAAGGCATCAGAA	105	482.20
A008M23	CGTCTGCTTATTAGAAGTTAAAC	TATTTAGTCTTGTGTTGGGTTAGC	127	482.20
WI-6938=stSG51146	GATCCTTACATCCAGCTTTT	TTTTCACTGCAGGAATAAAGAGG	155-156	482.20
G26427	AGAGACGTCTAAATTATACATTG	CCAGTCTTACTTCTTTTATT	169	482.20
U36223	TCACGGAGATTGTGCTG	AAGTGGACCTCACGCTG	136	482.20
WI-15962	TTTAGAGGCCGTGTGGGG	CTCCGTCCACCCTTCTAGTG	111	482.20
Z38716	GGTGAGATTCTGTGACTTTTAA	AGTTGCCCCGTCTCTGTTT	212-213	482.20
W86982	AAACGACATCTAAGTGAATCACCC	TCTGGTCTTCCCTTTCTGGC	242	482.20
stSG2699	ATTTGGGCTTCTCACAGTGG	GCACGTTACCTTTGTGCTG	153	482.20
stSG4193	CATGGTCCCAACGCTAGTTT	CTCCTGGAAAGGGCTTCTCT	148	482.20
stSG51021	CTCCTGACTTGTAAACAGGGACC	AGTTGCAAGCAGGGAGAAAA	155	482.26
stSG30235	GCCCCGCCTATTACTTTTC	ACTTTCCCTTCCCTTTTCCC	135	482.26
D60282	ACCTTTACTGTGCCATGGA	AGATCCCAGAGGGATGCAG	130	482.26
AA917820	CTCCTGACTTGTAAACAGGGACC	AGTTGCAAGCAGGGAGAAAA	155bp	482.26
WI-9147	AGGTCAGGGGTGACACAGAC	TGTTGTAGCTGTTGGCTGTTG	116-117	482.26
stSG46822	TTATATCCTTTGCCTCCCC	AGGAGGGACTGGAGTCCG	188	482.26
stSG26135	TGGCCCATGATTAATGGAAT	TTATCGAGTGTGTTGGGTGCA	130	482.26
stSG15820	GCCAGGTA AAAACAATTGAAAGG	CTTACCAACAACCTTGAGATGG	206	482.26
SGC33233	ACCTTTACTGTGCCATGGA	AGATCCCAGAGGGATGCAG	130	482.26
stSG13279	ATTGCTGGCAAATCAGCTCT	TCAAAGTATAGCAAGGATGTGG	106	482.26
AA913870	TCTCAGTTTGGGAAGGGATG	CGAGTCTCCGGGACATTCT	176	482.26
stSG50971	TCTCAGTTTGGGAAGGGATG	CGAGTCTCCGGGACATTCT	176	482.26
stSG44115	ATCTGGCCCTGGTTTGAAC	TTCCCTGTCTGCTTGTCT	173	482.32
stSG48239	TCTGTTTCTCTCCAGATGG	TGCCAGTCTTCTCCGAC	187	482.32
stSG42868	ATCTGGCCCTGGTTTGAAC	TTCCCTGTCTGCTTGTCT	173	482.32
stSG47726	TTTATTGAGCTCTCCATATGCC	TGGAACATAACCAGCATCCTG	89	482.32

Πιν. 4: (συνέχεια)

STS	ΑΛΛΗΛΟΥΧΙΑ	ΕΚΚΙΝΗΤΩΝ	ΜΕΓΕΘΟΣ	ΘΕΣΗ
	FORWARD	REVERSE		
SGC31986	CATGTGTCAATCCCATGATTG	GTGCTCCTTCTCTGGTTTGC	127	482.32
T34337=A004R30	TGTATTACACAGATATAAAGTG	TAAAACCTGAGATGAGGA	201	483.85
stSG916=D10S540	CTCACAAAGCAAAACAAGCTT	CTGCTGAATACCCAATGAAG	169	484.04
SGC30899	ATTCTGTGTGGAAGCTGGTTGG	CCCTATGATTATACCAGAAAACATG	133	484.14
stSG47299	CCCTCCCATCACATGAGC	TTCTCCTCATTGACAATTCCG	178	484.14
stSG31562	CCTGGGAGAGAAAAGCTTGTG	TTGTTGTTTTAATTGCAGGGC	120	486.37
stSG47144	AGTTACAGCTGCCCTGCCTA	GCACACAGCTGGAATAAGCA	182	486.37
stSG48141	TTTCCATGTTTTTGTATCACC	CATGAAACAAAAGTAGCCAGAA	122	486.37
stSG27142	TTTACTGTGCCCATGGAGG	CAGCCAGAAAGTACCGCTTC	189	486.47
stSG31529	AAGCCCTACAGTGCAGAGGA	CACAGCGTGTCTGCCAC	162	486.47
W69567	AATGTCAAATGCACCTTACTGGG	GTTCCTGCTATGGGAAAGG	239	486.57
SGC31936	CACATCTTAAAAATAGCTCCAGATG	ACACAGTGATCAGGGAAGGG	100	486.67
stSG42688	TCCGGGTAGGAGACAAAGG	CTCAATAATCCGCTCTTTCC	179	487.07
R43868	TCAAAGTTTTATTTTGCCTCTTGC	TCATCACTCTAGGATCATGAAGTG	150	487.10
G27616	GAATCTGCCTAAATTTCTCCCA	TTTGAATGTGTGACTTAGAAAAGAA	132	487.10
stSG26359	CCCACCGAGTTCTGTGTCT	GTTTTGCTTTGAAGAGACCCC	161	487.20
Z38397	GTGGGAAAAAGGGTGTTA	TTTTGTAGCTGAGAACGTGGA	221-222	487.48
stSG9207	GTCTGAGGTAACAGTTTTCCG	TGTGGAGGCCAAGAGCTC	124	487.93
stSG34995	CACATTGAAACCTGGCAAA	TCTCTTAAGGTTGAAGGCACA	131	488.21
M85683	CTAAGCCTGTGTGCCTCAG	GGGAGATTACCAACAAAAT	102	488.31
stSG13162	ATAGCAAACGACCCATTGGA	ACCTAATGCCAGTTGAACACC	106	488.31
stSG36084	GTACAACATGCCTGAACCAACAG	GATCCAAAGTGGGATGGTTGTGG	132	488.42
WI-19473	AGGTAGGATCCATGCCTGC	GGGGCTTCAAATTAGAGAC	204	488.42
stSG3345	TGTAGAGAAAGTGGTTAAAAAC	AAAAAAGAGGACAGATATGAG	214	488.42
F10278	TGGAAAAAGCTCTATAAAATGTCC	GGTTGCTTGGTCATAGAGCC	125	488.42
WI-14107	TCTTGATAATATCATACCACATGGC	TTTTCGTAGTTATCTAAGGGGTTGG	100	488.42
H40105	TGACTTTACTCAGGAAATATGACCA	GAGACTTCAAGGCTTGGCC	131	488.57
stSG35254	CGCACAGAAAGCCTTCC	TGAACATCTTTCCGCC	183	489.05
stSG40669	GGCTGTACAAAGAGCAAGGC	TGGCTTGGACAGAGACCAG	181	489.05
stSG48312	TTGGCAGTGAAGATTTGCAG	GTCTTTGAAGACAAGGGACCC	145	489.58
stSG46999	TTTTTGCAGCTTTGTGTGG	GAGGATCCCACAGTTCTCCA	126	489.58
H05479	CTATGCTGTGGTCTTCAAATCC	ACTCCAAAATCCTGAAAAGGC	104	489.58
stSG42504	CCAGAACTGCCCAAATGACT	GGAAAGGTATCCCGATGAAT	146	490.96
Z38786	TAAGTCGGCAAATGAGAG	AGAAGTTCCAACAGCAAC	103	496.65
SGC31312	AGGTAACGTAGGAGCAATCACC	TAGAGCTGGCCGTGTCCAC	150	497.17
W32419	CACCTGGAACATGACCATC	GCCAAGAATAAAATTTGTCTGC	133	499.85
stSG41370	CACAGAGGTTCAATGGCTCA	ACCTGTGAACATGCTTTTCCA	162	501.93
stSG16259	GAGCTTCAGTTTTTAAGGCAAA	AAAGTAGAAACCAGCTGTTGGC	229	503.14
WI-16106	AACTTAGCTGTCGTTCAAGTAAAGC	AGCTAGATCAATTTTAGAGAATGCC	127	504.45
WI-15620	CATTTTCCAAGTTTTCTATCATGG	TTCCATCCACAAAACATTGTC	150	506.37
AA243510	TTACTAGCATTGCAGACCCC	TCAGTGGCGTATCATGTCTAG	146	506.37
NIB1357	CTGTCAATAGCAACATGTGC	AGGGCAGAGACTTTTGTGTTT	119	506.37
stSG49291	AACTCATCCCGGACTCTG	CACCACCTTTCTTTGGTAGTCC	157	506.37
SGC35716	TTTGTAGAAAAATAGATGTCCAGCC	TTTAGGGGGTAAAGTTAAGCTGC	215	506.79
stSG26986	TCCCCTTTGTTGAAAAATAAGG	CCTGAAGTAACCTTTGTTTGCCA	106	506.79
stSG48063	TAATCATGAGGCCCAAGACC	ATCCAACCTGAAGACAGCTCTCC	120	506.99
stSG8954	CTTGCATAAAGTGTGTTGGG	TTGTCATTACTGGTGGGATCTG	125	506.99
stSG9160	TGTAATCTTGGGTGCTCAACC	CTGCCCATGGTAGAATGTT	159	506.99
AA284841	TCACATTCACACGGGTCTTAG	CGGTGCAAACTGGTTTCC	215	506.99
stSG12737	CATCCATCATCAAACCATTC	TTTAAGGCATCCCTCGGAC	111	507.09
stSG22433	CAAACCCAGGGAAGGGTAAT	TTGTTATTCTCATGGCCTTCG	163	507.09
stSG42559	TTTCGGAGGTAGTGTCAATGG	CCGAGGTGTTGTGAGGATTT	139	508.31
stSG50884	CCCTGACTTTGGCAGTTCAT	AAAAGACAAATGGCCCTGC	131	508.31
A005M36	GGTAGAGCTGTAATAAACAA	GCTGTAAAACCTATTGTAGAT	131	509.27
AA040321	CTTCCAAGTTATCAGGCCACC	GGCCCGAACTGGAGGACC	226	512.60
T60418	AGATCTTAGCAACCTCCATATACGA	TAGCTGCCTCAGTGACCAGA	221	512.66
T97004	GGGGACTCTGTCTGGTTTGG	AAGGCCCTGAAAAGACTTGC	192	514.18
A006C33	AAATCTGAGGTTATACTGAA	TCCATCCCAAGTAAGTTA	204	515.11
stSG8843	CTCCCACTGTCTTCTCCAGC	GAAAACCTTCCCCCTCTCC	168	515.32
stSG26044	AATTCTGCAACTTCCAAAAGG	CAAAAGGGCTTCTCCAACCTG	138	515.32
H17634	CTGAGACTTTTTGGCACTTACG	CATGAGCCTACCCATGCTG	131	515.54
WI-14968	TATGCAAAAATAACTCCAGTTGA	AATCCAAAATTCAAAATGCC	134	515.54

Πιν. 4: (συνέχεια)

STS	ΑΛΛΗΛΟΥΧΙΑ	ΕΚΚΙΝΗΤΩΝ	ΜΕΓΕΘΟΣ	ΘΕΣΗ
	FORWARD	REVERSE		
stSG39063	CATGCATGATCCCTCATCC	AAATGCAATCCCTGGATTTG	147	515.54
stSG49550	GAGCTGATAATCAAATCCCAA	TGAGAATGTTTTCTGCCTATG	93	515.54
stSG31639	ACCTCACTGAAGAGTTGGCA	CCCAGCAGCTGAGTCATGT	124	515.71
stSG12596	TCCAAGAGCTGTTAGTCATTGA	CGTGTCTTTACTGAGTCCCTCC	113	515.75
FB3F3	TTAGAAGGCTGATCAATCCA	TACGCATTATGGGAAAGTT	252	515.85
stSG45253	TCATACATTGCATTCCCAA	CAAAGAGCTGCCTGCTTTG	123	515.85
stSG49455	CCCGAAAATGCTTTTTCT	GAGTGTGGGTGAAATGCCT	136	516.07
SGC31141	GACTCTCAACCAAGAGAAAATCA	CCACCATGTTGGTTTGTG	150	516.07
stSG29677	GGAGGATGGGGATGGATC	ACTGGGAAGCCAGATTCTT	127	516.25
stSG47767	GAACATCACAAAACCACTTCCA	TGTATCTGCTTTCCAGAGGACA	169	516.25
stSG47370	CACAGTTTAACTTTTCCCTG	CATGAGGGATGGATATTGGC	138	516.46
stSG48444	TCACTGTGGTAGCCTGAAACA	GACAAGCACACATTTTTCTTGC	127	516.46
WI-9487	TCTATTTATCACTGTGGTAGCCTG	AGTTGGTAAACCATGTGCTCTG	249-250	516.46
WI-17804	TAATTTTCATAAAAATAAGGGCATGC	ATGATGAAACTGCCTAAAGATTCA	132	516.56
A007148	CTAGCTCTAAAGAAGTGTGAA	AAAATGCAGACTGTCATATC	254	516.56
A006J06	AGTAGAACAACATGTGTTA	TGAACTTTTCCAGTAAA	188	516.88
A008S06	TCTAAATGACACAATGTAGCAGA	TAACAATTTATAGGCTCAATGTG	192	516.88
stSG15535	TTCTTTACATTTCCCGGAG	CCCAAACACCTTCTTAAAGG	206	517.09
T91988	CAGTAGTTAGGCCACATTGAGAGC	CTGAGGCACAGTGAAGTTAGCC	107	517.39
stSG47984	TACCCTCCTTGCCAGCTG	CTCCACTGTGCCAGTCCC	85	517.39
A006F09	TATCCGAACTACATAACAGT	AATATGTAGATGAGAGAAGA	119	517.39
stSG28889	TAACAAGGGCTCCCACTTG	TCTCCTTCCCTCCCTGCC	140	517.39
stSG36024	GCCCTCCAGGTGATACTTGAGC	CTAGGGGTTCCCTTACAGCAGT	166	517.39
SHGC-33545	CCCGTAGTGAGAAAGTTGGA	CGATTCCCTGCACTTCTTC	150	517.42
stSG46271	AGCATGTGCTCAGCATTACG	CTGTCTAAGCATGGTGGGT	131	517.52
N38956	GGCACCAGGGGTGTATATG	GAAAGTGCTGTTCTTGGG	139	517.52
stSG42986	ACAAAACATGTTTCATCGTAGCG	GCCCTGGATCACAAAGTGAAT	156	517.52
stSG30264	GTCCAGTGTGGACTTAACCACA	AGTGGCAGGTGGTGAAGAC	139	517.52
stSG36017	TTGGCATTGTTTAAATCCAGGCTT	TTGTCTTGTACCCATGGCCGTC	194	517.81
stSG22417	TTGATAAGCACCCATGCTGA	AGGTTTCTACATGCCTTCTTGG	126	517.91
A004M41	ACACTGTAATGTTTACAACCTCT	TTGCCTGAATACTTAAAGTAGTT	140	519.34
U29690	AGCCCCACCTTGCTTTC	CATGTGCAGTAATCTGAACTGTTTC	262	519.76
A006A44	ACTTCTCTTTGCTAGTAAAG	AGTGTTCATGTAGACAGAT	157	519.76
WI-16878	TCACCAAAGTATGAAGCCATATG	GTGATCAAGATCCATAGATTTGC	150	519.76
stSG43429	AGTGAGTGCCTGCTGGCAAC	ACTGAGCAGGTGTGTGTTGG	132	519.86
stSG22723	CATGCCAGTTTCTACATGG	AGGATGGAGAGTAAGAGGAGGG	132	521.17
N32594	CATAGCAAAGTTCCAAGTGAAT	GTAAAGAATGCGGCTCATAA	126	521.27
stSG45467	CCTTAAAATGGCCTGACCA	GCAGGCCTGAATCAACTGTC	189	521.27
T87905	ACATGCTCTTCTGCAATGA	TATATGACACACTGGTGTCTTTT	212	521.37
stSG3554	CTCCTGTATATTCTGTACTGTTT	TAGTGGTAAAAACTGACTTTATC	206	523.37

2. XAPTHΣ GM99-G3 Map

STS	ΑΛΛΗΛΟΥΧΙΑ	ΕΚΚΙΝΗΤΩΝ	ΜΕΓΕΘΟΣ	ΘΕΣΗ
	FORWARD	REVERSE		
stSG57715=SHGC-526	TTCCGTTATGAAACACAGATTGCC	TTGAAAATGACCAACCAGACCTCC	192	4246
SHGC-17231	CAGCCTCCCAACTAAAGCTG	TACCCTGGTGGTCTTCCAAC	132	4742
SGC33917	TCACTGCATCAAAGACAGTGC	GAATGATTCTCTTCCAGCGC	125	4790
SHGC-4169=stSG55997	GACAGAGGTAATGAGATGATGG	ACTGTCCCACTATGGGTCTC	153	4908

3. XAPTHΣ MARSHFIELD

STS	ΑΛΛΗΛΟΥΧΙΑ	ΕΚΚΙΝΗΤΩΝ	ΜΕΓΕΘΟΣ	ΘΕΣΗ
	FORWARD	REVERSE		
D10S2470	CCTCCTAGCTCCTCAAGCTT	CAGGACAGATTTCTGTGGT	243-271	112.58
D10S1240	TCACCTGTCCCAGCTATCTG	AAAATTACTGGTACACTTGATAGCT	197-221	125.41
D10S1429	GCTCGTAATAGCTTTGTCCA	ATGAAACCATATATGTGACTTTTTG	248-260	130.90

Πιν. 4: (συνέχεια)

4. XAPTHΣ GENETHON

STS	ΑΛΛΗΛΟΥΧΙΑ		ΜΕΓΕΘΟΣ	ΘΕΣΗ
	FORWARD	REVERSE		
D10S541	AAGCAAGTGAAGTCTTAGAACACC	CCACAAGTAACAGAAAGCCTGTCTC	247-273	114.30
D10S1765	ACACTTACATAGTGTCTTCTGCG	CAGCCTCCCAAAGTTGC	166-184	114.30
D10S579	CCGATCAATGAGGAGTGCC	ATACACCCAGCCAATGCTGC	260-276	114.30
D10S1735	CCTGATTTGGATGCCATGT	TGACTCAGTGGGGCCTAGA	173-177	114.80
stSG34710=D10S1739	CTGGAAAAACAACAGAGGTG	GCTGTCTAAATCAAGGAATGTC	227-243	115.50
stSG835=D10S564	TCTCTCCTCTCCATCTCTCC	AGCTCTAACATAGAGGCCAGA	125	119.00
D10S1753	CTGCTGCCACCAACCTAA	CAAGTGAGACTCGATGACA	223-279	119.00
D10S1755	CCATAGTATCAATGTGAGGGTG	TGGCTAATTTTTAAGGTGTGT	299-311	121.10
D10S583	TCTGACCAAAATACCAAAAAGAAC	AGAGACTCCAGATGTTTGATGA	201-219	122.00
stSG472=D10S185	CATCCTATGCTTTCATTGTC	GCTAGGGTCATATTCCTTATCC	201	123.30
D10S571	CAGTCTTGCTTTGTCAGAA	ATGTACACTGTGTCTGTAATCCCT	187-213	124.20
D10S1736	GGTCTGACCCTGAGATTG	AGACTGTGAGTGTGCCG	128-146	124.30
D10S1690	CATTCTGCCATGATGACA	TGCTCCAGGTGACAAAGA	147-165	124.50
D10S1680	AGCCTGAGCAACATATCGAA	TCCCGAAGCAGAGATACCT	201-219	124.60
D10S1758	ATAATGCACACGCTGCC	AATCCCCTGGAGGTTTT	181-197	126.00
D10S577	TTGCACACGACCTAAG	GCCCAAGAGTTGAGAC	199-213	126.10
D10S1709	GTGAGTCCAGAATCACCCC	CAGTGAATGGCTCATTG	158-174	126.80
D10S1726	ACCCTTGCCATTTG	GTGAGGTGCTAAGCAGGTC	188-208	129.70
stSG575=D10S198	TGGTCTAGAAATGGAGAAGG	CTATTGGCCTTTCCCTAGA	255	129.80
D10S603	GCTGGATTATCCTCGGTAAC	GTTACTCTCCCAAGGCAA	232-256	131.00
stSG538=D10S192	ATGATTATACAGAGAACACAGACAG	CCAGGGAAGCCAAAAGATTGAAC	241	131.20
D10S1266	CTCACCTAGTCCAGGATTTTC	CAGCCAACGATTTCTGC	141-167	131.20
D10S1778	CTTGGTTATGATCTCACATGGTCT	CTGCTCTGGATTGAATGTTT	216-226	131.20
D10S1265	TCCATTGCTGACAGGGA	GCCAGTACAACAACCCCTAA	253-271	131.20
D10S1268	AGCTACCATGTAGATGCTACAATAT	TAAAACCACTCATTCTGATCCT	123-161	131.90
D10S1692	GCTGAGATTGTGCCATTG	CCAGGAAAGCCAGATGTT	182-211	131.90
D10S1738	GGTACTGTCTTATTGGG	TCATCTACATTCACCAAGTTG	183-213	131.90
D10S1697	GCTGCTTCGATGGAAC	AACCTGTGTCGGCTGC	167-201	131.90
D10S1267	AGGATGTCAAACATAACTGGGTAG	GGAGGCTGGAGGTAAGG	175-213	131.90
D10S1668	ATAACATCACCTTCCTGGG	TGTTCTTGAAAAGTCAGCAG	216-226	131.90
D10S222	TGGAACCTACCGAATGGA	TCTAACTGTGATTGAAGCGAC	264-276	131.90
D10S1671	ATGCAATGAGTCTCCCT	TGATGATGCTCAAATGATGG	92-110	135.20
D10S1760	GCGAGACTCCATCTCCATAG	CCATATAGTGGGTGGCTTAAA	112-155	136.30
D10S1663	TATCAAGCAATTTACAATCTGTG	AGCCATACCATAGTCAAACCTG	205-229	136.30
stSG34711=D10S1741	GAGTAGTGTGATGGCTCCCT	AGATGTTCAAGTCTTGGCT	210-254	136.30
D10S597	GAATGAAGACATCCAGAGG	GCAAGTATCAGAAACCCAA	206-222	137.60
D10S554	GGAGGACTCATGCAGACTT	CCTACCTTTAATTCAGCCCT	150-162	139.20
D10S1269	CTCAGATCGCGGAGTAGA	TGCTGTATTATTACCAGTCTTG	222-232	140.20
D10S1681	AATGCACTGTATGCACATTTG	CCAGCAGAGGCTATGGG	227-237	141.80
D10S1748	TGAATTGCTGTTACTCTCCAAG	ATAAATCGCCACAAGGTCC	212-266	142.70
D10S1731	TACTAGCAGGAGCAATCAGAC	GTTTGTCTAACAGGTTGTG	173-191	142.70
D10S1773	CTCAGAATGTGACCTTATTCAGAAA	ATCTGTGCCTTCATGTTACT	170-210	142.70

5. XAPTHΣ SHGC-G3

STS	ΑΛΛΗΛΟΥΧΙΑ		ΜΕΓΕΘΟΣ	ΘΕΣΗ
	FORWARD	REVERSE		
stSG57454=SHGC-14535	ATTAAGCCATGGACTTAAGAGCC	CGTCCTAACTAGCCATTCCC	196	4226
G17536	CAGTAGCTTATAGCCATCCACC	CAAGGAATCCAGAATGGAA	102	4920
G13660	GTACGCGCCATAGTGCAGTA	TTCAGCATTTCCTAGACTAAGC	192	5019
G13658	TCATTTTTGAAGACATTTTTGC	TTTTTTAACAGAAACTTGACAGGC	148	5040
G13676	TGACAGCCCCAAATAGTTCC	ATCCAGGGTAAATTTGTGTTTGC	94	5077
G14345	CCTCTCAGAAAGGCTCCATCT	GGTGAATCTCTTTGAGTCACTG	96	5143
stSG57440=SHGC-14061	CATAGGCATTTGAAGCCTTACC	GGCCAGAATGCTTAGACGAG	106	5174
D10S1207	CGCAGATCTTAAACTGCAAA	CTTGGACACCCATACAGAGG	108	5202

Πιν. 4: (συνέχεια)

6. ΧΑΡΤΗΣ WI-RH

STS	ΑΛΛΗΛΟΥΧΙΑ		ΜΕΓΕΘΟΣ	ΘΕΣΗ
	FORWARD	REVERSE		
D10S1242	CAAGCATACTGTATTAGTTAGGGC	GGACCCCATGATCATGTATG	139-140	513.9
stSG58057=AFMa348zg9	AATTGCTTGAGCCAGG	TGACGGTGAATAAGTCTCTCG	156bp	531.7
D10S1562	TGCATTGTTGTGAACTGTTAGA	GCATGGTGGTAACAGTTTACTCA	264-265	543.3
D10S1257	CGTTTTCTTCTACATCTTTGGGG	TTTGGAAAAGACAGCCCATC	256-257	545.7
WI-5915	AATCTTGGTATGCCTTCTGTCTT	CATGTTAACACACCTATAGGTCACA	142	549.7
WI-4209	TACATCGACCAAGAGCAACG	CAGTAACCCCACTCCTCCCT	259-260	555.3
stSG58023=AFM205tg7	GTCACAGTTACCAGTTTATATCCCA	GAAAACAATGGCGGCT	93	558.4
WI-5502	CAGAGTTACCATGTTCCAGGC	CCCTTCAGATATTTGTCAGTTTTG	127-128	567.9
stSG58054= AFMa301wf1	GCAAAAAGTTGGCTACTAGAG	ACATACATACATACATACGTGTGTG	254	567.9
WI-4865	TGTTCTCGAGGATGTGCAAG	CAGAAGAAAAGGAACAGCATCC	185-186	568.0
stSG58067=AFM183xb12	TTGTCCAGAATTAGAAGTACAGTGG	GGTGTGGTGGTTCATACCAA	255	575.5
stSG55936=WI-14791	TTTTTTTACATCATCAACATTGCA	AAGCTGCAATTTTATCCCTCC	102	585.9
WI-3109	GGTGGGATGGGTAACATAACC	AGGGAATCTGGGAGTGGC	257-258	591.0
WI-4379	GCTGAAGAATGGACTTTGGA	CCTCACAACCTCCATGGGG	142-143	595.0
D10S1245	GGTGGTCCCATTTATTTCTC	GAAGCATTTCAAGAGCCAAA	222-223	595.0
D10S1434	TCCACCTTAAATCCTTGCTT	GGGAAGTGAGATTCTGTCTCA	149-150	595.0
WI-1905	CAGGACTGAAATCTCTCATCTGG	TTATTTTACCTTTTCTGCCTTGG	215-216	601.1
stSG58042=AFMa053yg5	TGAGCCTGTTGTTCAAAG	CCCTGGTTTCTATGC	134	604.1
stSG58064=AFMb363xg5	TCCTGCTTTTGAAGTATCA	AATGTGCCTGTATCGTGTG	247	604.1
stSG58043=AFMa071xh1	CACTGGTTGGCTAGTCAGAG	TGAGGCCTGATTTCAAATG	274	604.1
WI-4434	GCTCTGGACATCACTCCCTC	AGAATGAGTCAATGGCCTG	185-186	604.1
WI-4132	TTGATGCAAAGGAAATTGCTG	CTCAGAATGCTGCTTCCTCC	182-183	608.6
WI-5976	TAAACAAGGGCACAAGATTGG	TGTCTTAAACCATTTTCTTATTGG	106-107	619.2
stSG58028=AFM265yc9	CCTGGCCCCAGACAGGTATTATATT	CCAAATTCAGGACTTTTAGCCTC	277	620.4

B. 10q26.2-qter

1. ΧΑΡΤΗΣ GM99-GB4

STS	ΑΛΛΗΛΟΥΧΙΑ		ΜΕΓΕΘΟΣ
	FORWARD	REVERSE	
WI-17697	TTTTATGTAATTATCATGGGCTGC	ATCTAACTGGCTTTCAAACAACA	150
WI-17432	GTTTTCACTGACAATGAGAAAATG	CATTTTGGAAAAGACAGAAACAGC	126
stSG1460	TCTGTGTCTTGAAGCAGCC	CCCCTGTCAACAGTGACATG	194
stSG32124	TCCTGGGCTCAAGTGATCTT	GAGTTTCAGAAGGTGGCAGC	186
A008Y06=55747	TACATAAAGGCCTTATTTGACAT	TCTATCAGAACATAATGTGTCC	125
WI-20848	CATTTTATTTTTCGTTGCAATGA	TGAGAACTCCAAAGTCCTGAAA	166
stSG1566	AGACACTGGGAAACAAACG	GTCCTCTCTGTGTTCCCAA	185
stSG8192	TGTACTGCGTGTGTAAGACGC	CAGAAAAAACCAAGTGCCTC	142
stSG48184	GGTGTGAGTTTAGTTAGGGCG	TGTTTCATCAGCAGTGGCTC	122
SGC33916=stSG66860	GGCTGCACCAGCTGTATTTTC	GGGGAGCTCTCGATGTC	138
H99205	ACCATTTCTAACAGGAACATAAAA	TGTGCTTGAGCCTCGT	121
stSG4213	GCCACCCATTGACTTTGC	GCATCAGAAAAGTTAATGTGGC	127
R96892	CCCATATCTGTTAGTGTGGATCA	GGCTTAGAGCCAGGTGTTA	130
W15548	TCTACTAGGATTTGTCACGGC	GATCTGTTTCGGTTAAGCTG	230
A008M13	GAATTGGTCTATGAATTTCAACT	TCCTTAAAATAATGGTTGTTACC	186
D10S1561=WI-6793	GAACCCACCAGGCTCTGTTA	GGGATGCAGTTGCCTTTTTA	258
stSG28017	CACCATTTCCAAGGGCAG	TGTGCAGGAAAAAAAATTGC	138
stSG36110	CAGACGCTCCCTTAGAGCCTCC	GGGAGGATCAGGTGCTGGATCC	178
stSG47386	CCAGAGAAAGATCTCCAAGTGG	CTCCTCCTCCACGTGAAGTC	134
stSG46609	TCATGACACTACTCGTTACC	GACCTGAGCGGAAAGTTGAG	159
stSG47380	TGGTCACTCGTAGCAGCATC	ATTCTTCCCAGCCCTTG	145
stSG2750	CCAGGGTTGACTAATGCTTAAA	CAGTGCCACTTGAGTTCACA	153
stSG47732	TCTCCTGTTCTGCAGGTGTG	AAGGTGGCCTACCAGGTTG	135
stSG2795	AGGCCTGTGAATACTGCTTAGC	AAAACCTAACCCGGCCTGTTT	129
stSG30731	GCACAATCCCTACCAAGGA	AGCTCTCTTGCCAGGGTACA	189
A009F48=stSG57527	TTGTAGATCACCATACAAAACCTA	TGCCAGGTAGTGTGTGTA	123
AA036862	GAAAGAGTGTGTAATAGCCCTGG	AGCCTCCTTTCGAATGTTCC	187
stSG41791	AAGGGTCACCTTGTGACGAG	ACCTGTGGAGGGCATTTC	177

Πιν. 4: (συνέχεια)

STS	ΑΛΛΗΛΟΥΧΙΑ ΕΚΚΙΝΗΤΩΝ		ΜΕΓΕΘΟΣ
	FORWARD	REVERSE	
AA010383	ATTCCTCAAGGGTGAACGC	TGGTCTTGGCAATTCTAGCC	238
W67510	AAAGAGCGGTATCTGACAATAAC	CACTGCTTCCCAAAGGT	124
WI-22094	TTTTTGAACAACCCACCT	AAATTCTGCCACACTTTGC	257
stSG46145	GAAAAAGAAAAACACCCTGGG	GGACACCTGGTATTTACAGTG	151
SGC32344=stSG66859	GAAATTGGTGCCAGGACG	GTATCTAACAGACAGTCTCAGTGC	126
WI-19193	TACTCCCATGCCGTTCTTC	TTGTCTGTAGCGGAGCAATG	274
stSG45059	CCTGGTGATTTGGCATCAC	GAAGAAGCCACGAGGTTTACG	189
T48398	AGGCTGGGAGCTGCCACT	TACGGAAACCTCATCCTCAGGAAG	153
stSG43417	AGAGACTCCCAGCAGCAGAG	CTTAGAGGCCAGGGGTCTCT	157
stSG45453	AGAGACTCCCAGCAGCAGAG	CTTAGAGGCCAGGGGTCTCT	157
WI-14945	GACGCGTCTTTATCAGCT	AGAAAACCTTTGCAATGCTG	150
stSG4581	AGCTGCTAATGCCATTTGCT	TAGACTCAGAAAAGGAGTGGGG	150
stSG4790	GAGCAGCTTTAACACACAGGC	CCTCTTGAGAAATGGATGCTG	138
stSG43280	CTGCAAAGACAGAACACAGAGC	ACCCAGGGTCCGAGTCCAC	128
stSG42519	ACACAGGGCCAGCTTCAG	GATCATGCCCTCTGTGCC	122
AA191340	TTTCCGTCCGAGCAGAG	AGCCTTCGTCACCCAGC	109
stSG1303	ATCTTGTGAAAGCGGGG	CCTCAGCCCTATACATATTGGG	115
stSG27915	CTGTCCCCTGTAGGGGTG	CTGTCCAATCCTTAACGGTAGC	135
Cda01F10=d10s1328E=stSG55715	AGCCTACTCTGTCCACCC	GCTTTATGTGTCAGCATCTCC	217
55725=A007H40	AGGTAGCAGTCCACATTAT	GAAGGTAAACTCCTGCT	253

2. ΧΑΡΤΗΣ MARSHFIELD

STS	ΑΛΛΗΛΟΥΧΙΑ ΕΚΚΙΝΗΤΩΝ		ΜΕΓΕΘΟΣ
	FORWARD	REVERSE	
D10S1222	CCACCTTTCAACTGTCCATT	CATCACACACCCGGGGACT	221-245
D10S1439	TCAGGTCAGGGGACTTAAACA	TAAAGATCGACCTCTGCCTG	173-193
D10S1134	CAGCCTGGTGACAGAGCCA	AAAATGCAGACGTCCTGGAC	326
D10S1248	GGAATAAGTGCAGTCTTGG	ACCAATCTGGTCACAACCAT	241-261
D10S505	GGGCAACACAAGACCCTGC	TATTGGGGGCTGGTCCCC	247
D10S169	GATCTGTGACTGCCTTCT	AAGAGGAGGAGTCCATTACG	99-117

3. ΧΑΡΤΗΣ GENETHON

STS	ΑΛΛΗΛΟΥΧΙΑ ΕΚΚΙΝΗΤΩΝ		ΜΕΓΕΘΟΣ
	FORWARD	REVERSE	
stSG17459=D10S1703	GTATCTTCCATTGAAATCTATGC	TGTGGGTACAATAAGCCATTATCT	126-144
D10S186	TTTGAGAAATGTGATGAAGGCTTGG	TGCATCTATGAGGACAGGGTGGG	147-167
D10S1782	CATGCAAGGATTCATCTCAA	ACATGATGTCTTTGCTTTGT	263
stSG17460=D10S1727	TCATGTGATGACAATCTGACAGT	ACCAAATGTGGCCTGCT	233-287
stSG725=D10S217	GCAGAGGTTTTCCCTTTCATT	CTAGGAGGGTGTCTAACC	185
D10S1676	CCACCACTCAGAGGTAAGG	AATTGTATGATCCCAACTTTGTC	156-168
D10S1655	TGCCAGACTCCATGATGGTATGAAC	TAACAAACATCTGAGCACAGCCCC	209-257
D10S1770	CATGAGATTTGAGTGGGG	AACAGGCTGTGGCCTC	198
D10S1651	TCGATTTGACGCAAGTTT	AGCACCCAGGGCTTCT	166-200
D10S1675	AAGAGGCAGTGTCTGTCG	ACCTGTTGGAGCCACC	205-251
D10S1711	CTATGATGCTGAGGGCAGTG	ATGCTACCATGCCTGGCTAC	174
stSG987=D10S555	CTGGATATGGAATGGAGTATCA	TTCTCTTCCCTCATTCTG	136
D10S212	GAAGTAAAGCAAGTCTATCCACG	TCTGTGTACGTTGAAAATCCC	189-201

4. ΧΑΡΤΗΣ SHGC-G3

STS	ΑΛΛΗΛΟΥΧΙΑ ΕΚΚΙΝΗΤΩΝ		ΜΕΓΕΘΟΣ
	FORWARD	REVERSE	
SHGC-14552	ACTCTTACTTCTGCCTTGGC	GAGGCCCTTTATGAAATGCA	174

Πιν. 4: (συνέχεια)

5. ΧΑΡΤΗΣ WI-RH

STS	ΑΛΛΗΛΟΥΧΙΑ ΕΚΚΙΝΗΤΩΝ		ΜΕΓΕΘΟΣ
	FORWARD	REVERSE	
D10S2229=WI-3219	ACTGGCAAAAAGGATGGTTG	TAGAATCTTATCCTGCCTTCCC	126
D10S1593	ATCTTTGTTCATTGGTGACAGGG	CAGTGTGTGAGTGCAGGAGC	99
D10S1256	CCGTGTGATGCAGCTTTG	AAGAGGCACAAGAGTAAAACGC	232
D10S1585	GAAACGATGGTTGAAGAAAAGC	ATGCCGAAGATTAGATTTTAATCC	138
D10S1408	TTGACTGGAATCTGTTGCTAGA	AGCACTGATGCTGTTTACCAC	274
D10S1201	CCCCACCATGCGTACA	TAGCACTGATTGATGCCGAG	281
D10S2490	CTCCAAGGGTAACATTGAAC	CGTAAACGTTGTGAGATGGT	160
D10S1588	TGCTCAAGACCCCATTTAT	CACCCTGGTGGCAGTAGG	129

Σημ.: Οι κωδικοί των STSs προέρχονται από την βάση δεδομένων UniSTS. Παρατίθενται οι αλληλουχίες των εκκινήτων που χρησιμοποιήθηκαν στην PCR, το μέγεθος των παραγόμενων προϊόντων και οι συντεταγμένες τους στους διάφορους χάρτες. Οι μονάδες για τους γενετικούς χάρτες Genethon και Marshfield είναι σε cM ενώ για τους χάρτες κυτταρικών υβριδίων GM99-GB4, GM99-G3, WI-RH και SHGC-G3 σε cR.

Πιν. 5: Αλληλουχίες εκκινήτων

Γονίδιο	Κωδικός	Αλληλουχία	Εφαρμογή
3077	1F	CTTAGGGCTTACAGGGC	Διερεύνηση μεταλλαγών
3077	1R	CAGAAGACACAGATGCGC	Διερεύνηση μεταλλαγών
3077	2F	ATTGGCACTGGTCAGTCC	Διερεύνηση μεταλλαγών
3077	2R	GACATGTCCTCTGCTCC	Διερεύνηση μεταλλαγών
3077	3F	GGGAACTCCGAGTCTCC	Διερεύνηση μεταλλαγών
3077	3R	TACTCTCCCACTCCTGG	Διερεύνηση μεταλλαγών
3077	4F	TCCTCACAGGAGCACTGG	Διερεύνηση μεταλλαγών
3077	4R	CAAGGTTAGGTTTTGCC	Διερεύνηση μεταλλαγών
3077	5F	ACTGCAGAGCTGGGTTTG	Διερεύνηση μεταλλαγών
3077	5R	TCCAGTGTCTCTGCAGG	Διερεύνηση μεταλλαγών
3077	6F	GTGACTGGAAGCCTGAGG	Διερεύνηση μεταλλαγών
3077	6R	ACCAGCCCATTCCTTAG	Διερεύνηση μεταλλαγών
3077	7F	GTGAGAACTTCGGGATG	Διερεύνηση μεταλλαγών
3077	7R	CCCAGATGCCTTAGCAAC	Διερεύνηση μεταλλαγών
3077	8F	TCCTCTGTGGCTTTACCC	Διερεύνηση μεταλλαγών
3077	8R	AAAGTGTGGCTCTCTGC	Διερεύνηση μεταλλαγών
3077	9F	CATGCCTGGCTTCACATC	Διερεύνηση μεταλλαγών
3077	9R	ATGCTTGCCTAGTCAGC	Διερεύνηση μεταλλαγών
3077	10F	GTAACAGAGGCCATGGG	Διερεύνηση μεταλλαγών
3077	10R	CAGTATTGGCCTCAAGCC	Διερεύνηση μεταλλαγών
3077	11F	TCATGATTGACCCGAGG	Διερεύνηση μεταλλαγών
3077	11R	CTTGACCCTGACTTTCCC	Διερεύνηση μεταλλαγών
3077	MycFnew	CGGGATCCTCAGGAACCATGCGAGGC	Με τον MycR για υποκυτταρικό εντοπισμό πρωτεΐνης
3077	MycR	CCGAATTCACCTGCCATCACAGAGGC	Με τον MycFnew για υποκυτταρικό εντοπισμό πρωτεΐνης
AJ312051	F_BamHI	CGGGATCCCCTGCAGAGATGCTGGGTC	Με τον R_Kpn για την κλωνοποίηση του cDNA
AJ312051	R_Kpn	GGGGTACCATTTTCTGGGAGTACAAG	Με τον F_BamHI για την κλωνοποίηση του cDNA
AJ312051	F_SpeI	GGACTAGTCCCTGCAGAGATGCTGGGTC	Με τον R_Kpn για RT-PCR
FRA10AC1	F	CGGGATCCGAAAGGATGCATGGTCATG	Με τον R4 για την κλωνοποίηση του cDNA
FRA10AC1	F1	GCCATTTAGATGTAACACGTG	Με τον R1 για γονοτυπική ανάλυση
FRA10AC1	F2	TCAGGATTTGTTCTATGAGACG	Με τον R2 για παρασκευή ανιχνευτή Northern
FRA10AC1	F3	GCAGAAGAGGCCCTCAAG	Με τον R2 για RT-PCR
FRA10AC1	F5	CGGGATCCGCATGGTCATGGAGGCTATG	Με τον R5 για κλωνοποίηση στον pET26b+
FRA10AC1	R1	GTGTCGTTTTCTTCTTTCC	Με τον F1 για γονοτυπική ανάλυση
FRA10AC1	R2	CCACACAAGCTGTAACCTGACG	Με τον F2 για Northern, με τον F3 για RT-PCR
FRA10AC1	R4	CGGAATTCGAAACAATCCTGAAAATACTC	Με τον F για την κλωνοποίηση του cDNA
FRA10AC1	R5	GGAAATTCAGAAACAATCCTGAAAATACTC	Με τον F5 για κλωνοποίηση στον pET26b+
Fra10ac1	mFRAF	AGATGAGTTGCTGCTTAC	Με τον mFRAF για παρασκευή ανιχνευτή για RISH
Fra10ac1	mFRAR	CGAAGTTGACTTCCCAGC	Με τον mFRAF για παρασκευή ανιχνευτή για RISH
Lgi1	mLG11F	TGAAATAGCCAGACCAC	Με τον mLG11R για παρασκευή ανιχνευτή για RISH
Lgi1	mLG11R	TGGATACGTGTGTGAATG	Με τον mLG11F για παρασκευή ανιχνευτή για RISH
Lgi3	mLG13F	GCGACAGAAATTTGTAC	Με τον mLG13R1 για παρασκευή ανιχνευτή για RISH
Lgi3	mLG13R1	TGGTGACATACGTAGGTG	Με τον mLG13F για παρασκευή ανιχνευτή για RISH
Lgi4	mLG14F	GAATTGGATCTCAGTGCC	Με τον mLG14R για παρασκευή ανιχνευτή για RISH
Lgi4	mLG14R	AAGCTCCTGCCTTGAGC	Με τον mLG14F για παρασκευή ανιχνευτή για RISH
pACT2	pACT2F	TACGATGTTCCAGATTAC	Αλληλούχηση θηραμάτων πειράματος δύο υβριδίων
pACT2	pACT2R	TTTCAGTATCTACGATTC	Αλληλούχηση θηραμάτων πειράματος δύο υβριδίων

Πιν. 6: Αντισώματα και συνθήκες εφαρμογής

ΑΝΤΙΣΩΜΑ	ΣΥΝΘΗΚΕΣ	ΠΡΟΕΛΕΥΣΗ
A. ΠΡΩΤΟΓΕΝΗ		
anti-Xpress (mouse)	W: o/n 4°C, 1/4000, TBSTM I: 1h RT, 1/250	Invitrogen (Cat. No. R910-25)
anti-SAP145 (rabbit)	W: o/n 4°C, 1/1000, TBSTM, I: 1h RT, 1/250	R. Reed, Harvard Medical School
anti-SAP145 (mouse)	W: 2h, RT, 1/1000, TBSTM	Santa Cruz (Cat.no. sc-14279)
anti-lamin A/C (mouse)	I: 1h RT, 1/500	Santa Cruz (Cat.no. sc-7292)
anti-DIG-AP (sheep)	T: o/n 4°C, 1/1000	Roche (Cat.no. 1 093 274)
anti-FRA10AC1 (rabbit)	W: o/n 4°C, 1/100, TBSTM, I: 2h RT, 1/10	Δ. Αποστολοπούλου, διπλ. Εργασία, Παν. Κρήτης
anti-GST (mouse)	W: 1h RT, 1/5000, TBSTM	Santa Cruz (Cat.no. sc-138)
anti-His, C-term (mouse)	W: o/n 4°C, 1/2500, PBSTM	Invitrogen (Cat.no R930-25)
anti-GFP (rabbit)	W: o/n 4°C, 1/5000, PBSTM	Invitrogen (Cat.no R970-01)
anti-GFP (mouse)	W: 1h RT, 1/2500, PBSTM	Santa Cruz (Cat.no.sc-9996)
B. ΔΕΥΤΕΡΟΓΕΝΗ		
anti-mouse-HRP (donkey)	W: 1-2h RT, 1/2500	Jackson Immunoresearch (Cat.no. 715-035-150)
anti-mouse HRP (goat)	W: 1-2h RT, 1/1000	Chemicon (Cat.no. AP124P)
anti-rabbit-HRP (goat)	W: 1h RT, 1/13000	Jackson Immunoresearch (Cat.no. 111-005-003)
Alexa Fluor 488 goat anti-mouse	I: 1h RT, 1/1000	Molecular Probes (Cat.no. A-11029)
Alexa Fluor 568 goat anti-mouse	I: 1h RT, 1/1000	Molecular Probes (Cat.no. A-11031)
Alexa Fluor 568 goat anti-rabbit	I: 1h RT, 1/1000	Molecular Probes (Cat.no. A-11036)

Σημ.: Στην δεύτερη στήλη παρατίθενται οι συνθήκες που χρησιμοποιήθηκαν σε ανάλυση Western (W), σε ανοσοφθορισμό (I) και σε υβριδοποίηση σε τομές ιστών με ριβοανιχνευτή (T).

ΜΕΘΟΔΟΙ

Οι μέθοδοι που περιγράφονται πραγματοποιήθηκαν σύμφωνα με τα εργαστηριακά εγχειρίδια των Sambrook J *et al*, 1989 και των Ausubel FM *et al*, 2000.

2.12. Απομόνωση νουκλεϊκών οξέων

➤ Απομόνωση πλασμιδιακού DNA από *E.coli*.

Σε μικρή κλίμακα

Η απομόνωση πλασμιδιακού DNA σε μικρή κλίμακα πραγματοποιείται με αλκαλική λύση. Μοναδιαίες αποικίες μετασχηματισμένων βακτηρίων αναπτύσσονται σε ~3ml LB με το κατάλληλο αντιβιοτικό στους 37°C ολονύκτια υπό ανάδευση. Τα κύτταρα συλλέγονται με φυγοκέντρηση (15min, 3000rpm σε RT) και επαναιωρούνται σε 300μl διαλύματος P1 (15mM Tris-Cl pH 8, 10mM EDTA pH 8). Ακολουθεί επώαση με λυσοζύμη σε τελική συγκέντρωση 300 μg/ml για 15min σε πάγο, προσθήκη 300μl διαλύματος αποδιάταξης P2 (200mM NaOH, 1% SDS), ήπια ανάδευση και επώαση σε RT για 5-15min (μέχρι το διάλυμα να γίνει διαυγές). Η εξουδετέρωση γίνεται με προσθήκη 300μl διαλύματος P3 (3M οξικό κάλιο pH 5,5), ισχυρή ανάδευση και επώαση για 20min σε πάγο. Το μίγμα φυγοκεντρείται στις 12000rpm στους 4°C για 15min, το υπερκείμενο συλλέγεται και ακολουθεί επαναληπτική φυγοκέντρηση στις ίδιες συνθήκες. Το πλασμιδιακό DNA κατακρημνίζεται με προσθήκη ισοπροπανόλης

(0,8 του όγκου) και συλλέγεται μετά από επώαση για 15min σε πάγο και φυγοκέντρηση στις 12000rpm για 15min στους 4°C. Η πελλέτα ξεπλένεται με 70% αιθανόλη, στεγνώνει και επαναδιαλύεται σε 50μl TE (10mM Tris.Cl pH 8, 1mM EDTA pH 8). Στην περίπτωση που ακολουθεί τεμαχισμός χρησιμοποιείται το 1/3-1/10 του DNA.

Σε μεγάλη κλίμακα για αντίδραση αλληλούχησης ή σε παροδική διαμόλυνση

Σε αυτήν την περίπτωση η διαδικασία έχει όπως πριν αυξάνοντας όμως τον όγκο της βακτηριακής καλλιέργειας στα 2lt, χρησιμοποιώντας 200ml των διαλυμάτων P1, P2 και P3 και εφαρμόζοντας επιπλέον επεξεργασία. Συγκεκριμένα, μετά την κατακρήμνιση του πλασμιδιακού DNA ακολουθούν εκχυλίσσεις με φαινόλη/χλωροφόρμιο και επανακατακρήμνιση με CH₃COONa και απόλυτη αιθανόλη. Το DNA συλλέγεται με φυγοκέντρηση στις 12000rpm για 15min στους 4°C, η πελλέτα ξεπλένεται με 70% αιθανόλη και μετά το στέγνωμά της επαναδιαλύεται σε ~3ml TE. Το RNA απομακρύνεται με επώαση με RNAση σε τελική συγκέντρωση 100μg/ml για 1h στους 37°C και ακολουθεί επώαση με πρωτεϊνάση K σε τελική συγκέντρωση 100μg/ml για 2h στους 55°C σε κατάλληλο ρυθμιστικό διάλυμα (1x: 50mM Tris.Cl pH 7,5, 100mM NaCl, 0,02% SDS, 2mM EDTA). Ακολουθεί εκχύλιση με φαινόλη/χλωροφόρμιο. Η απομόνωση του DNA γίνεται με προσθήκη απόλυτης αιθανόλης, ψάρεμα των νηματίων που σχηματίζονται, ξέπλυμα με 70% αιθανόλη και τελική επαναδιάλυση σε 2-3ml T.E. Με την μέθοδο αυτή απομονώνονται 1-3mg πλασμιδιακού DNA.

➤ **Απομόνωση γενωμικού DNA από πλακούντα**

Ο ιστός, αμέσως μετά την απομόνωσή του, πλένεται με φυσιολογικό ορό και διατηρείται στους -80°C σε τεμάχια όγκου 25cm³ περίπου. Για την απομόνωση γενωμικού DNA ο ιστός κονιορτοποιείται παρουσία υγρού N₂. Ακολουθεί προσθήκη διαλύματος TSE (150mM NaCl, 100mM EDTA pH 8, 20mM Tris.Cl pH 8, ~100 ml ανά τεμάχιο) και SDS σε τελική συγκέντρωση 1%. Ο κονιορτοποιημένος ιστός διαλυτοποιείται με θέρμανση στους 65°C. Στη συνέχεια το δείγμα επωάζεται με πρωτεϊνάση K τελικής συγκέντρωσης 100μg/ml, στους 65°C ολονύκτια και εκχυλίζεται με φαινόλη, φαινόλη/ χλωροφόρμιο, χλωροφόρμιο/ισοαμυλική αλκοόλη (v/v: 24/1). Τα νουκλεϊκά οξέα κατακρημνίζονται με CH₃COONa (0,25M) και απόλυτη αιθανόλη και συλλέγονται με την μορφή ινιδίων. Μετά την εξάτμιση της υγρής φάσης γίνεται επαναδιάλυση σε 20 ml διαλύματος T.E στους 4°C με ήπια συνεχή ανάδευση για δύο μέρες. Ακολουθεί επεξεργασία με υπέρηχους. Εμπειρικά, ο ιδανικός χρόνος για την παραγωγή τεμαχίων DNA μεγέθους 200 – 1500bp, υπολογίζεται σε τρεις

επαναλήψεις των 10sec η καθεμία (40% cycle actif, 5 microtip limit), με ανάμιξη και πάγωμα του διαλύματος πριν από κάθε επανάληψη. Το RNA απομακρύνεται με επώαση με RNase A στους 37°C σε τελική συγκέντρωση 50μg /ml για 3h η οποία στη συνέχεια απομακρύνεται με εκχύλιση με μίγμα φαινόλης/ χλωροφορμίου. Η τελική κατακρήμνιση και επαναδιάλυση γίνεται όπως αναφέρθηκε παραπάνω, ρυθμίζοντας τη συγκέντρωση του DNA στα 10mg/ml. Από την επεξεργασία ενός τεμαχίου ιστού παράγονται ~100mg DNA-φορέα (carrier-DNA).

➤ **Απομόνωση ολικού RNA από ιστό**

Για την εκχύλιση ολικού RNA από εγκέφαλο ανθρώπινου εμβρύου χρησιμοποιήθηκε το αποδιατακτικό μέσο ισοθειοκυανική γουανιδίνη όπως περιγράφεται από τους Chomczynski P και Sacchi N 1987. Αναλυτικά, ένα κομμάτι του ιστού κονιορτοποιήθηκε παρουσία υγρού N₂ και στη συνέχεια προστέθηκε διάλυμα D (4M ισοθειοκυανική γουανιδίνη, 25mM κιτρικό νάτριο, 0.5% Sarcosyl και 0.1M 2-μερκαπτοαιθανόλη). Ο όγκος του ρυθμίστηκε ώστε το ιξώδες να είναι το ελάχιστο δυνατό και ακολούθως προστέθηκε διάλυμα CH₃COONa pH 4.0 σε τελική συγκέντρωση 0.2M και όξινη φαινόλη ίσου όγκου με το ομογενοποίημα. Στη συνέχεια, μετά από ισχυρή ανάδευση (vortex) για τουλάχιστον 5min, προστέθηκε μίγμα χλωροφορμίου και ισοαμυλικής αλκοόλης (v/v: 24:1), όγκου ίσου με το 1/5 του δείγματος και ακολούθησε ισχυρή ανάδευση (5min), επώαση στους 4°C για 15min και φυγοκέντρηση στις 12000rpm για 20min στους 4°C. Η υδατική φάση εκχυλίστηκε επανειλημμένως με μίγμα ουδέτερης φαινόλης/χλωροφορμίου στις ίδιες συνθήκες και το RNA κατακρήμνίστηκε με 2.5 όγκους απόλυτης αιθανόλης.

2.13. Τεμαχισμός DNA με περιοριστικές ενδονουκλεάσες

Ο τεμαχισμός του DNA με περιοριστικές ενδονουκλεάσες πραγματοποιήθηκε με περίσσεια ενζύμου (3-5units/μg DNA) σε κατάλληλο ρυθμιστικό διάλυμα και στην συνιστώμενη θερμοκρασία (βάση των οδηγιών των εταιριών Minotech και NEB) τουλάχιστον για 3h, όταν ήταν επιθυμητή η πλήρης πέψη ή για 1h για διαγνωστικές πέψεις.

2.14. Ηλεκτροφόρηση

➤ **Ηλεκτροφόρηση DNA σε πήκτωμα αγαρόζης**

Η ταχύτητα με την οποία το DNA μετακινείται κατά την ηλεκτροφόρησή του σε πήκτωμα αγαρόζης εξαρτάται από το μέγεθός του, την διαμόρφωσή του

(υπερελικωμένο, nicked, γραμμικό), τη συγκέντρωση της αγαρόζης, την τάση που εφαρμόζεται, την παρουσία βρωμιούχου αιθιδίου (EtBr) και την σύσταση και την ιοντική ισχύ του διαλύματος ηλεκτροφόρησης. Η συγκέντρωση των πηκτωμάτων κυμαινόταν μεταξύ 0,8 και 2%. Η ηλεκτροφόρηση, αναλυτική και παρασκευαστική, πραγματοποιήθηκε σε ρυθμιστικό διάλυμα 1xTAE (50xTAE: 200mM Tris, 200mM οξικό οξύ, 1mM EDTA pH 8.0) με EtBr σε συγκέντρωση 0,1μg/ml. Η διάρκεια και η τάση που εφαρμόστηκε κυμάνθηκε μεταξύ 1-24h και μεταξύ 1-7V/cm αντίστοιχα. Πριν την ηλεκτροφόρηση, τα δείγματα θερμαίνονται στους 65°C για 5min παρουσία διαλύματος Orange G (0,25% Orange G, 15% Ficoll σε T.E).

Απομόνωση DNA από πήκτωμα αγαρόζης

Η απομόνωση του DNA από πήκτωμα αγαρόζης πραγματοποιείται με αυτοσχέδια κατασκευή εκχύλισης. Το τεμάχιο του πηκτώματος κόβεται, πολτοποιείται με κατάλληλο έμβολο και τοποθετείται σε σωληνάκι erpendorf 0,5ml στον πάτο του οποίου έχει τοποθετηθεί υαλοβάμβακας και έχει ανοιχτεί μία μικρή οπή. Αυτή η διάταξη τοποθετείται σε σωληνάκι erpendorf 1,5ml. Ακολουθούν φυγοκεντρήσεις διάρκειας 2min στις 10000rpm στους 4°C κατά τις οποίες το διάλυμα του DNA συλλέγεται στο σωληνάκι υποδοχής ενώ η αγαρόζη συγκρατείται από τον υαλοβάμβακα. Το DNA εκχυλίζεται με χλωροφόρμιο και κατακρημνίζεται με CH₃COONa και απόλυτη αιθανόλη. Όσο μικρότερο είναι το μέγεθος του DNA και όσο αραιότερο είναι το πήκτωμα, τόσο υψηλότερη είναι η απόδοση της παραπάνω μεθόδου, που κυμαίνεται μεταξύ 20 και 70%.

➤ **Ηλεκτροφόρηση DNA σε πήκτωμα πολυακρυλαμίδης (PAGE)**

Τα πηκτώματα πολυακρυλαμίδης είναι απαραίτητα όταν απαιτείται ο διαχωρισμός μορίων που διαφέρουν κατά ένα νουκλεοτίδιο και αυτή η ιδιότητά τους αξιοποιήθηκε κατά την ηλεκτροφόρηση αντιδράσεων αλληλούχησης και προϊόντων PCR της γονοτυπικής ανάλυσης. Χρησιμοποιήθηκαν αποδιατακτικά πηκτώματα συγκέντρωσης 8M ουρίας, 6% διαλύματος ακρυλαμίδης: δισακρυλαμίδης 38:2, 0,5% TBE (10xTBE: 0,9M Tris-Cl, 0,9M βορικό οξύ, 20mM EDTA, pH 8.0) τα οποία πολυμερίζονται με APS και TEMED σε τελική συγκέντρωση 0,02% και 0,1% αντίστοιχα. Η ηλεκτροφόρηση πραγματοποιήθηκε σε ρυθμιστικό διάλυμα 0,5xTBE με σταθερή ισχύ πεδίου 80W ενώ είχε προηγηθεί θέρμανση των δειγμάτων στους 80°C για 2min παρουσία stop mix (6x: 95% Formamide, 20mM EDTA pH 8, 0,05% bromophenol blue, 0,05% xylene cyanol). Μετά το πέρας της ηλεκτροφόρησης τα πηκτώματα μεταφέρονται σε διηθητικό χαρτί Watman 3MM, αποξηραίνονται με θέρμανση σε κενό αέρος και υφίστανται αυτοραδιογραφία.

➤ **Ηλεκτροφόρηση πρωτεϊνών σε αποδιατακτικό πήκτωμα πολυακρυλαμίδης (SDS-PAGE)**

Κατά την ηλεκτροφόρηση των πρωτεϊνών σε αποδιατακτικό πήκτωμα πολυακρυλαμίδης, τα μόρια διαχωρίζονται βάσει του μοριακού τους μεγέθους. Αυτό οφείλεται στην παρουσία του ανιονικού απορρυπαντικού SDS στο πήκτωμα, στο διάλυμα ηλεκτροφόρησης και στο διάλυμα Laemmli που προστίθεται στα δείγματα. Τα πολυπεπτίδια προσδένουν SDS, σε ποσότητα ανάλογη με το μέγεθός τους, φορτίζονται αρνητικά και παρουσία ηλεκτρικού πεδίου μετακινούνται προς την άνοδο. Επιπλέον, το SDS διασπά τους υδρογονικούς δεσμούς, τις υδρόφοβες αλληλεπιδράσεις και ουσιαστικά αποδιατάσσει τις πρωτεΐνες. Πλήρης αποδιάταξη επιτυγχάνεται αφενός με την επίδραση της μερκαπτοαιθανόλης (συστατικό του διαλύματος Laemmli) αφετέρου με τη θέρμανση των δειγμάτων στους 100°C. Το πήκτωμα αποτελείται από δύο φάσεις: την ανώτερη στην οποία πραγματοποιείται η ευθυγράμμιση των πρωτεϊνών σε ενιαίο μέτωπο και την υποκείμενη, δηλαδή το πήκτωμα διαχωρισμού. Οι συγκεντρώσεις των συστατικών του πηκτώματος ευθυγράμμισης είναι: 4,5% ακρυλαμίδη/δισακρυλαμίδη (29:1), 0,125M Tris-Cl pH 6,8, 0,1% SDS, 0,1% APS, 0,2% TEMED και του πηκτώματος διαχωρισμού 0,375M Tris.Cl pH 8,8, 0,1% SDS, 0,1% APS, 0,05% TEMED ενώ η συγκέντρωσή του σε ακρυλαμίδη/δισακρυλαμίδη κυμαίνεται μεταξύ 8% και 12% και επιλέγεται βάσει των μοριακών μεγεθών των πολυπεπτιδίων. Τα πρωτεϊνικά δείγματα αποδιατάσσονται με προσθήκη διαλύματος 5xLaemmli (5x: 250mM Tris-Cl pH 6,8, 10% SDS, 50% γλυκερόλη, 25% μερκαπτοαιθανόλη, 0,5% bromophenol blue) και θέρμανση στους 100°C για 10min. Όλα τα πηκτώματα που χρησιμοποιήθηκαν, αναλυτικά και παρασκευαστικά, είχαν πάχος 1,5mm και διαστάσεις 8x 7,3cm. Η ηλεκτροφόρηση πραγματοποιήθηκε σε ρυθμιστικό διάλυμα (10x: 250mM Tris, 1,92M γλυκίνη, 1% SDS, ρύθμιση του pH στο 8,3 με HCl) με σταθερή ένταση πεδίου 35-45mA και μετά το πέρας της ακολουθούσε χρώση ή ηλεκτρομεταφορά των πολυπεπτιδίων σε μεμβράνη.

2.15. Αλληλούχηση DNA

Ο προσδιορισμός της αλληλουχίας των ενθεμάτων πλασμιδιακών κατασκευών και γραμμικών προϊόντων PCR στηρίχθηκε στη μέθοδο Sanger. Ως υπόστρωμα χρησιμοποιήθηκαν 250ng DNA ανά kb μεγέθους και 25ng ραδιοσημασμένου εκκινητή. Οι αντιδράσεις πραγματοποιήθηκαν σε θερμοανακυκλωτή με υλικά του SequiTherm EXCEL™ II DNA Sequencing Kit σύμφωνα με τις οδηγίες της εταιρείας Epicentre (<http://www.epicentre.com/>). Τα προϊόντα αναλύθηκαν σε πήκτωμα πολυακρυλαμίδης μετά από αυτοραδιογραφία.

Κατά την αλληλούχηση των θετικών κλώνων του συστήματος δύο υβριδίων, χρησιμοποιήθηκαν οι εκκινητές pACT2F ή pACT2R που υβριδοποιούνται στον polylinker του pACT2 στις εξής συνθήκες: 95°C 5min/ 95°C 45sec, 55°C 45sec, 70°C 1min, για 30 κύκλους. Εναλλακτικά, η αλληλούχηση έγινε στο εργαστήριο Γονιδιωματικής Ανάλυσης στο IMBB (PE-ABI377 sequencers).

2.16. Μετασχηματισμός βακτηρίων

Με τον μετασχηματισμό επιτυγχάνεται η εισαγωγή και ο πολλαπλασιασμός πλασμιδιακού DNA σε βακτήρια. Τα βακτήρια καθίστανται ικανά να δεχτούν DNA εξωγενώς μετά από ειδική κατεργασία και πλέον ονομάζονται δεκτικά. Ο μετασχηματισμός τους γίνεται είτε με εφαρμογή ηλεκτρικού πεδίου είτε με θερμικό σοκ, μετά από κατάλληλη χημική επεξεργασία. Η απόδοση του μετασχηματισμού, όσον αφορά τη δεκτικότητα των κυττάρων, κυμαίνεται μεταξύ 10^6 και 10^7 (αποικίες ανά μg υπερελικωμένου DNA). Χρησιμοποιήθηκαν κύτταρα *E.coli* στελέχους DH5a, BL21 DE3 codon+ και KC8.

➤ Παρασκευή δεκτικών βακτηρίων και μετασχηματισμός με θερμικό σοκ

Με αυτόν τον τρόπο τα βακτήρια καθίστανται δεκτικά (competent) μετά από χημική επεξεργασία. Αρχικά γίνεται εμβολιασμός 100ml LB με μοναδιαία αποικία βακτηρίων από φρέσκο επιστρωμένο τρυβλίο και επώαση ολονύκτια με ανάδευση στους 37°C. Ακολουθεί εμβολιασμός 100ml φρέσκου θρεπτικού μέσου LB με 1ml από την πρώτη καλλιέργεια και επώαση στις ίδιες συνθήκες μέχρι την αρχή της εκθετικής φάσης ανάπτυξης ($O.D_{560nm} \sim 0,45$). Σε όλα τα βήματα που ακολουθούν τόσο τα κύτταρα όσο και τα διαλύματα διατηρούνται σε πάγο και οι φυγοκεντρήσεις πραγματοποιούνται στους 4°C. Τα βακτήρια συλλέγονται με φυγοκέντρηση στις 2500rpm για 15min και επαναιωρούνται σε 20ml διαλύματος TFB1 (100mM RbCl₂, 50mM MnCl₂, 30mM CH₃COOK, 10mM CaCl₂, 15% γλυκερόλη, ρύθμιση του pH στο 5,8 με αραιό CH₃COOH). Μετά από παραμονή τους για 20min σε πάγο, επανασυλλέγονται με φυγοκέντρηση στις ίδιες συνθήκες και επαναιωρούνται σε 3ml διαλύματος TFB2 (10mM MOPS, 10mM RbCl₂, 75mM CaCl₂, 15% γλυκερόλη, ρύθμιση του pH στο 7 με διάλυμα 1M NaOH). Ακολουθεί επώαση για 20min σε πάγο, διαίρεση του κυτταρικού διαλύματος σε δείγματα των 150μl (σε eppendorfs που έχουν προψυχθεί με υγρό άζωτο) και τελική αποθήκευση στους -80°C.

Μετασχηματισμός: Ένα δείγμα μεταφέρεται από τους -80°C σε πάγο. Μετά την τήξη προστίθεται το DNA και αναδεύεται ήπια. Ακολουθεί επώαση για 20min σε πάγο, κατά τη διάρκεια της οποίας το DNA έρχεται σε επαφή με την κυτταρική μεμβράνη,

θερμικό σοκ στους 42°C για 90sec και άμεση μεταφορά σε πάγο, όπου παραμένει για 1min. Προστίθεται 1ml LB και μετά από επώαση για 1h στους 37°C, τα κύτταρα συλλέγονται με φυγοκέντρηση στις 3000rpm για 15min σε RT και επιστρώνονται σε τρυβλίο.

➤ **Παρασκευή δεκτικών βακτηρίων και μετασχηματισμός με εφαρμογή ηλεκτρικού πεδίου**

Η προετοιμασία των κυττάρων συνίσταται στην μείωση της ιοντικής ισχύς του βακτηριακού αιωρήματος με διαδοχικές εκπλύσεις του με απιονισμένο H₂O. Αρχικά, γίνεται εμβολιασμός 100ml LB με μοναδιαία αποικία βακτηρίων από φρέσκο επιστρωμένο τρυβλίο και επώαση ολονύκτια με ανάδευση στους 37°C. Ακολουθεί εμβολιασμός 500ml φρέσκου θρεπτικού μέσου LB με 5ml από την πρώτη καλλιέργεια και επώαση στις ίδιες συνθήκες μέχρι την εκθετική φάση ανάπτυξης ($O.D_{560nm} \sim 0,65$). Η καλλιέργεια τοποθετείται σε πάγο για 30min και όλα τα επόμενα βήματα πραγματοποιούνται στους 4°C. Τα κύτταρα συλλέγονται με φυγοκέντρηση στις 2500rpm για 15min και επαναιωρούνται σε 500ml παγωμένου ddH₂O. Ακολουθεί επανάληψη της συλλογής και της επαναιώρησης σε 250, 50 και 10ml παγωμένου ddH₂O διαδοχικά. Η τελική επαναιώρηση των κυττάρων γίνεται σε 1,5ml παγωμένης γλυκερόλης 10% και το δείγμα διαιρείται σε δείγματα των 50μl που αποθηκεύονται σε σωληνάκια erpendorfs που έχουν προψυχθεί σε υγρό άζωτο, στους -80°C.

Μετασχηματισμός: Η εισαγωγή του DNA επιτυγχάνεται εφαρμόζοντας υψηλό ηλεκτρικό πεδίο για πολύ μικρό χρονικό διάστημα ώστε να δημιουργηθούν παροδικές οπές στο τοίχωμα των βακτηρίων, μέσω των οποίων εισέρχεται το πλασμιδιακό DNA. Με την μέθοδο αυτή μεγάλο ποσοστό των κυττάρων, της τάξεως του 50-70%, πεθαίνουν, ωστόσο η απόδοση του μετασχηματισμού είναι υψηλότερη σε σχέση με την προηγούμενη μέθοδο. Η διαδικασία που ακολουθείται έχει ως εξής: Ένα δείγμα μεταφέρεται από τους -80°C σε πάγο όπου, μετά την τήξη του (~10min), προστίθεται το DNA και αναδεύεται ήπια. Ο όγκος του δείγματος που προστίθεται είναι όσο το δυνατόν μικρότερος και η αλατότητα τελικά δεν υπερβαίνει το 1mM. Αν το δείγμα προέρχεται από αντίδραση σύνδεσης, είτε προστίθεται το 1/10 αυτής είτε έχει προηγηθεί καθαρισμός και επαναδιάλυση σε H₂O. Αν πρόκειται για υπερελικωμένο DNA προστίθεται ~1ng. Ακολουθεί μεταφορά του δείγματος σε παγωμένη κυψελίδα (BTX P/N 610, 1cm x 1cm με κενό χώρο πάχους 1mm) και υποβολή του σε ηλεκτρική εκκένωση (1600 Volts για 4-6msec) στην συσκευή Electro Cell Manipulator 600. Αμέσως τα κύτταρα μεταφέρονται σε 1ml παγωμένου LB, επωάζονται για 1h στους 37°C, συλλέγονται με φυγοκέντρηση στις 3000rpm για 15min σε RT και επιστρώνονται σε τρυβλίο με στερεό LB που περιέχει το κατάλληλο αντιβιοτικό.

2.17. Παρασκευή τυφλών άκρων

Η κατασκευή τυφλών άκρων σε μόρια που έχουν 5' προεξέχοντα άκρα πραγματοποιήθηκε με την υπομονάδα Klenow της DNA πολυμεράσης. Ως υπόστρωμα χρησιμοποιείται 1,5μg DNA, που ο όγκος του ρυθμίζεται στα 30μl και ακολούθως προστίθενται 4μl ρυθμιστικού διαλύματος 10xEcoPol (NEB, 10x:100mM Tris-Cl pH 7,5, 50mM MgCl₂, 75mM DTT), 5μl dNTPs (1mM) και 1μl (5units) ενζύμου. Ακολουθεί επώαση 15min σε RT, εκχύλιση με φαινόλη/χλωροφόρμιο και κατακρήμνιση με CH₃COONa και απόλυτη αιθανόλη.

2.18. Αντίδραση σύνδεσης (ligation)

Για τη σύνδεση γραμμικών μορίων DNA χρησιμοποιείται η λιγάση του DNA του βακτηριοφάγου T4. Η αντίδραση πραγματοποιείται σε τελικό όγκο 10-20μl και σε αυτήν συμμετέχουν το ένθεμα και ο φορέας με μοριακή αναλογία 10:1 αντίστοιχα, το ρυθμιστικό διάλυμα (10x: 300mM Tris-Cl pH 7.8, 100mM MgCl₂, 100mM DTT, 10mM ATP) και 6 Weiss units ενζύμου. Η αντίδραση επωάζεται είτε 2h σε RT είτε ολονύκτια στους 16°C. Πριν το μετασχηματισμό βακτηρίων, η λιγάση απενεργοποιείται με θέρμανση στους 70°C για 10min.

2.19. Ραδιοσήμανση του DNA

➤ Ραδιοσήμανση δίκλωνων μορίων

Με τη μέθοδο των τυχαίων εκκινήτων

Αυτή η μέθοδος ραδιοσήμανσης χρησιμοποιήθηκε για μόρια DNA >500bp. Στηρίζεται στη χρήση 6μερών εκκινήτων, ετερογενών όσον αφορά την αλληλουχία τους, για την έναρξη της σύνθεσης του DNA χρησιμοποιώντας ως υπόστρωμα μονόκλωνο DNA. Λόγω της ετερογένειάς τους οι εκκινήτες σχηματίζουν υβρίδια με διαφορετικές θέσεις του υποστρώματος και συνεπώς στο τελικό προϊόν αντιπροσωπεύονται εξίσου όλα σχεδόν τα τμήματα. Η πολυμεράση που χρησιμοποιείται είναι η υπομονάδα Klenow της DNA πολυμεράσης I της *E. coli*, ενώ τα δύο από τα τέσσερα νουκλεοτίδια που χρησιμοποιούνται για τη σύνθεση του ανιχνευτή είναι ραδιοσημασμένα. Σε κάθε αντίδραση χρησιμοποιούνται 50-100ng υποστρώματος DNA του οποίου ο όγκος ρυθμίζεται στα 7,5μl. Το DNA αποδιατάσσεται με βρασμό για 5min, τοποθετείται σε πάγο, φυγοκεντρείται στιγμιαία και προστίθενται σε αυτό: 11,5μl διαλύματος 2xLS {25μl Hepes pH 6,6, 25μl DTM2 [0,1mM dGTP, 0,1mM dTTP σε TM (250mM Tris.Cl pH 8, 25mM MgCl₂, 50mM β-μερκαπτοαιθανόλη)], 7μl O.L (90units/ml σε 1mM Tris.Cl pH 7,5, 1mM EDTA pH

7,5}}, 1μl BSA (10mg/ml), 1μl Klenow (5u/μl), 2 μl [α - 32 P] dATP και 2 μl [α - 32 P] dCTP που αντιστοιχούν σε 20μCi το καθένα, ειδικής ενεργότητας 3000Ci/mmmole (Amersham) σε συγκέντρωση 10mCi/ml. Η αντίδραση πραγματοποιείται στους 37°C για 3h. Ακολουθεί καθαρισμός των προϊόντων με χρωματογραφία μοριακής διήθησης μέσω Sephadex G-50 και προσδιορισμός της ειδικής ενεργότητας του ανιχνευτή. Για την αποφυγή της ραδιόλυσής του, ο ραδιοσημασμένος ανιχνευτής χρησιμοποιείται, μετά την αποδιάταξή του, άμεσα ή φυλάσσεται στους -80°C.

Με PCR

Ο τρόπος αυτός ενδείκνυται για μόρια μικρού μεγέθους (60-200bp) καθώς παράγονται ανιχνευτές ιδιαίτερα υψηλής ενεργότητας. Αυτή η μέθοδος εφαρμόστηκε για τη ραδιοσήμανση όλων των ανιχνευτών (STSs) με τους οποίους έγινε η διερεύνηση: α) της γενωμικής βιβλιοθήκης RPCI-11 και β) της υποβιβλιοθήκης του χρωμοσώματος 10. Επιπλέον, με την μέθοδο αυτή παρήχθησαν οι ανιχνευτές για την διερεύνηση της cDNA βιβλιοθήκης εγκεφάλου και για την ανάλυση Northern.

1) Ανιχνευτές για την γενωμική βιβλιοθήκη RPCI-11

Αρχικά, πραγματοποιείται PCR με υπόστρωμα γενωμικό ανθρώπινο DNA ή DNA κυττάρων σωματικών υβριδίων ανθρώπου – κρικητού (hamster), που περιέχει μόνο το μεγάλο βραχίονα του χρωμοσώματος 10 (υβρίδιο #175, βλ. Marzella R, Kokkinaki K *et al*, 1997) και όλο το γονιδίωμα του τρωκτικού. Η τελευταία επιλογή έγινε στις περιπτώσεις που εμφανιζόταν στα προϊόντα PCR πολυπλοκότητα μεγαλύτερη της αναμενόμενης. Η ποσότητα του υποστρώματος ήταν 50-100ng και στις δύο περιπτώσεις και ο τελικός όγκος 10μl. Σε κάθε αντίδραση συμμετείχαν επιπλέον 100ng από τον κάθε εκκινητή, 1χρυσμιστικό διάλυμα Taq πολυμεράσης (10x: 100mM Tris.Cl pH 8.5, 500mM KCl, 1% Triton X-100, 15mM MgCl₂), 100μg/ml BSA, 1mM dNTPs (μίγμα των τεσσάρων) και 1u Taq πολυμεράση (Minotech). Οι συνθήκες έχουν ως εξής: 94°C για 4min / 94°C για 45sec, 54°C έως 62°C για 45sec (αναλόγως των εκκινητών), 72°C για 30sec επανάληψη 35 κύκλων / 72°C - 5min. Οι αλληλουχίες όλων των εκκινητών που χρησιμοποιήθηκαν, οι κωδικοί, τα μεγέθη των αντίστοιχων STSs, οι χάρτες και οι συντεταγμένες τους παρατίθενται στον Πιν. 4.

Το προϊόν της αρχικής αντίδρασης ηλεκτροφορείται σε πήκτωμα αγαρόζης 2% και μετά την απομόνωσή του τοποθετείται σε 50-150μl H₂O, ολονύκτια στους 4°C, όπου σημαντικό ποσοστό του εκλούεται παθητικά στην υδατική φάση, μέρος της οποίας (1-5 μl) χρησιμοποιείται για επαναληπτική PCR. Αρχικά γίνεται επαναληπτική δοκιμαστική PCR στις ίδιες συνθήκες με αυτές της αρχικής αντίδρασης. Το παραγόμενο προϊόν ελέγχεται με ηλεκτροφόρηση με κριτήριο να εμφανίζεται ως

μοναδική ζώνη στο αναμενόμενο μέγεθος. Ακολουθεί PCR ραδιοσήμανσης, χρησιμοποιώντας 0,5μl φρέσκο [α - 32 P] dCTP (5μCi), ειδικής ενεργότητας 3000Ci/mmol (Amersham) σε συγκέντρωση 10mCi/ml. Τα τρία υπόλοιπα νουκλεοτίδια χρησιμοποιούνται μη σημασμένα σε τελική συγκέντρωση 0,125mM, το καθένα. Για τον υπολογισμό της ειδικής ενεργότητάς τους λαμβάνονταν υπ' όψιν: 1) η ποσότητα του [α - 32 P] dCTP που προστίθεται σε μία PCR (περίπου 0,5ng), 2) το ποσοστό ενσωμάτωσής του (περίπου 80% όπως προκύπτει μετά από μοριακή διήθηση σε στήλη Sephadex G-50) και 3) τα cpm που μετρούνται μετά από κατακρήμνιση με TCA σε δείγμα 1% του ραδιοσημασμένου προϊόντος (η συνήθης τιμή είναι 7-8 x10⁴ cpm). Βάσει των παραπάνω η ειδική ενεργότητα εκτιμάται ότι υπερβαίνει τα 10⁹ cpm/μg. Στις περιπτώσεις που η δοκιμαστική επαναληπτική PCR εμφάνιζε πολυπλοκότητα μεγαλύτερη της αναμενόμενης, εκτός από την βελτιστοποίηση των συνθηκών, πραγματοποιούνταν και έλεγχος του προϊόντος με ηλεκτροφόρηση (5% της ποσότητας) σε πήκτωμα αγαρόζης και ακόλουθη αυτοραδιογραφία.

2) Ανιχνευτές για την υποβιβλιοθήκη του χρωμοσώματος 10

Για τα STSs που είχαν χρησιμοποιηθεί και στην διερεύνηση ολόκληρης της βιβλιοθήκης, τα υποστρώματα στην PCR ραδιοσήμανσης ήταν τα ίδια με προηγούμενων. Για τα καινούρια STSs που δημιουργήθηκαν από τα άκρα των BACs τα υποστρώματα παρήχθησαν με PCR σε: α) γενωμικό DNA ανθρώπου, β) γενωμικό DNA κρικτηύ και γ) DNA κυτταρικών υβριδίων που περιείχε το χρωμόσωμα 10. Κάθε αντίδραση πραγματοποιούνταν σε δύο διαφορετικές θερμοκρασίες υβριδοποίησης (55 και 60°C). Ο τελικός όγκος της αντίδρασης ήταν 20μl και περιλάμβανε 100ng από τον κάθε εκκινητή, 1x ρυθμιστικό διάλυμα Taq πολυμεράσης [10x: 450mM Tris.Cl pH 8,8, 110mM (NH₄)₂SO₄, 15mM MgCl₂, 420μg/ml Cresol Red (από διάλυμα 845μg/ml σε T0,1E)], 0,25mM dNTPs (μίγμα των τεσσάρων), 13mM β-μερκαπτοαιθανόλη, 12% σουκρόζη (από διάλυμα 34,6%) και 1u Taq πολυμεράσης. Η συμμετοχή της σουκρόζης και της Cresol Red αντικαθιστούσε την προσθήκη του διαλύματος Orange G πριν την ηλεκτροφόρηση. Ακολουθούσε ηλεκτροφόρηση (σε πήκτωμα αγαρόζης 2,5% σε διάλυμα 1xTBE) και σε όσες αντιδράσεις προέκυπτε προϊόν του αναμενόμενου μεγέθους με τα υποστρώματα «α» και «γ», όχι όμως με το «β», απομονώνονταν το αντίστοιχο τεμάχιο του πηκτώματος το οποίο τοποθετούνταν σε 50μl T0,1E όπου και εκλούονταν ολονύκτια στους 4°C.

Για όλα τα STSs η αντίδραση ραδιοσήμανσης γινόταν σε τελικό όγκο 10μl και περιλάμβανε 100ng από τον κάθε εκκινητή, 0,6mM dNTPs (χωρίς dCTP), 1x ρυθμιστικό διάλυμα Taq πολυμεράσης (10x: 0,5M KCl, 100mM Tris pH 8,3, 15mM

MgCl₂), 1u Ταq πολυμεράσης, 0,3μl [³³P] dCTP (3μCi), ειδικής ενεργότητας 3000Ci/mmole (Amersham) σε συγκέντρωση 10mCi/ml. Η ποσότητα του ραδιενεργού ισοτόπου ανά αντίδραση ραδιοσήμανσης μειώθηκε στα 3μCi αντί των 5μCi [³²P] dCTP που είχε χρησιμοποιηθεί στην διερεύνηση ολόκληρης της βιβλιοθήκης, καθώς η υβριδοποίηση γινόταν σε μικρότερο όγκο. Η διαφορά των δύο ισοτόπων είναι ο χρόνος ημιζωής (14,3 μέρες του ³²P έναντι 25,4 μέρες του ³³P), η εμβέλειά τους στον αέρα (790cm του ³²P έναντι 49cm του ³³P) και η εκπομπή τους (1,709 MeV του ³²P έναντι 0,249 MeV του ³³P).

➤ **Ραδιοσήμανση ολιγονουκλεοτιδίων**

Τα ολιγονουκλεοτίδια που δεν είναι φωσφορυλιωμένα στο 5' άκρο τους (είτε επειδή έχουν αποφωσφορυλιωθεί είτε επειδή έχουν συντεθεί χημικά) μπορούν να ραδιοσημανθούν αν προσλάβουν μία ραδιοσημασμένη φωσφορική ομάδα. Για το σκοπό αυτό χρησιμοποιείται το ένζυμο T4 πολυνουκλεοτιδική κινάση, που μεταφέρει τη γ-³²P του [γ-³²P] ATP στο υδροξύλιο που υπάρχει στο 5' άκρο του πολυνουκλεοτιδίου. Στην αντίδραση χρησιμοποιούνται 10-20pmoles του ολιγονουκλεοτιδίου και ο όγκος ρυθμίζεται στα 15μl. Επίσης προστίθενται 2μl 10x ρυθμιστικού διαλύματος T4 πολυνουκλεοτιδικής κινάσης (10x:700 mM Tris-Cl pH 7.6, 100 mM MgCl₂, 50 mM DTT), 1μl (5units) ενζύμου και 2μl [γ-³²P] ATP (3000Ci/mmole, 10μCi/μl). Η αντίδραση επωάζεται στους 37°C για 45min.

2.20. Εκτίμηση της ειδικής ενεργότητας ανιχνευτή με κατακρήμνιση με TCA

Η ειδική ενεργότητα του ανιχνευτή (cpm/μg) πρέπει να υπερβαίνει τα 3x10⁸ και είναι μία σημαντική παράμετρος που ελέγχεται καθώς από αυτή εξαρτάται η ευαισθησία της υβριδοποίησης. Η μέτρηση γίνεται μετά την απομάκρυνση των μη ενσωματωμένων ραδιοσημασμένων πρόδρομων νουκλεοτιδίων, με χρωματογραφία μοριακής διήθησης. Για τον υπολογισμό των κρούσεων 1μl δείγματος, τοποθετείται σε φίλτρο Whatman μαζί με 10μg DNA - φορέα. Έτσι, μόρια DNA, μεγαλύτερα των 50 νουκλεοτιδίων, κατακρημνίζονται στην επιφάνεια του φίλτρου, το οποίο στη συνέχεια ξεπλένεται διαδοχικά με 10% κρύο TCA για 10 min, ώστε να απομακρυνθούν τα ελεύθερα ραδιοσημασμένα νουκλεοτίδια και με παγωμένη ακετόνη για 10min. Μετά το στέγνωμά του γίνεται μέτρηση των κρούσεων με σπινθηρομετρητή. Για τη μέτρηση των συνολικών κρούσεων του δείγματος χρησιμοποιείται ένα φίλτρο Whatman με 1μl ραδιοσημασμένου μορίου, που δεν έχει υποστεί κατακρήμνιση TCA. Η ένδειξη του μετρητή πολλαπλασιάζεται με 2,5,

συντελεστής που χρησιμοποιείται επειδή η μέτρηση γίνεται απουσία υγρού σπινθηρομέτρησης (τολουένιο).

2.21. Υβριδοποίηση νουκλεϊκών οξέων

➤ Υβριδοποίηση γενωμικών βιβλιοθηκών με STSs

Προϋβριδοποίηση

Η προϋβριδοποίηση των μεμβρανών, γίνεται για την αποφυγή της μη ειδικής σύνδεσης του ανιχνευτή, τόσο στα ίδια τα φίλτρα, όσο και στα μόρια που περιέχουν και είναι παρατεταμένη εξαιτίας της ύπαρξης μεγάλου αριθμού επαναλαμβανόμενων αλληλουχιών στο ανθρώπινο γονιδίωμα. Αυτό επιτυγχάνεται αφενός με τα συστατικά του διαλύματος, αφ'ετέρου με την προσθήκη DNA-φορέα. Η σύσταση των διαλυμάτων που χρησιμοποιήθηκαν, εναλλακτικά, είναι: 1) Για την γενωμική βιβλιοθήκη, το διάλυμα A (7% SDS, 100mM NaH₂PO₄, 50mM Na₄P₂O₇, 10mM EDTA, 1% BSA) και 2) για την υποβιβλιοθήκη του χρωμοσώματος, το διάλυμα B [6x SSC, 2 mg/ml Ficoll 400-DL, 2 mg/ml polyvinyl pyrrolidone 40, 2 mg/ml BSA (fraction V), 1% N-lauroyl-sarcosine, 50 mM Tris-HCl (pH 7.4), 10% w/v dextran sulphate]. Η προυβριδοποίηση πραγματοποιήθηκε είτε σε γυάλινους κυλίνδρους είτε σε falcons 15ml όπου προστίθενται ~15ml προθερμασμένου στους 65°C διαλύματος που περιέχει 100μg/ml αποδιατεταγμένο DNA-φορέα από ανθρώπινο πλακούντα και στην περίπτωση πολυμορφικών δεικτών 100ng/ml poly (dA-dC)-(dG-dT). Η διάρκειά της κυμαίνονταν μεταξύ 3h έως ολονύκτια στους 65°C.

Υβριδοποίηση

Για την διερεύνηση της βιβλιοθήκης RPCI-11 χρησιμοποιήθηκαν ομάδες των 3-22 STSs ανά υβριδοποίηση (pools), ενώ στη χρωμοσωμειδική βιβλιοθήκη κάθε ανιχνευτής εξετάζονταν μεμονωμένα. Η συγκέντρωση κάθε ανιχνευτή ήταν ~3x10⁵ cpm/ ml διαλύματος υβριδοποίησης και η ειδική ενεργότητά του υπερέβαινε τα 10⁹ cpm/μg. Το διάλυμα υβριδοποίησης είναι ίδιο με αυτό της προϋβριδοποίησης, με την επιπλέον προσθήκη του αποδιατεταγμένου ανιχνευτή. Στην περίπτωση που στους ανιχνευτές συμπεριλαμβάνονταν πολυμορφικοί δείκτες, αυτοί αποδιατάσσονταν παρουσία 1μl poly (dA-dC)-(dG-dT) (1mg/ml, Pharmacia) σε 4xSSC και μετά από επώαση στους 65°C για 3h, προστίθονταν στην υβριδοποίηση. Η διάρκεια της υβριδοποίησης κυμαίνονταν μεταξύ 16 και 20h.

Πλύσιμο φίλτρου

Το πλύσιμο των μεμβρανών πραγματοποιούνταν στις παρακάτω συνθήκες: 1) Για το διάλυμα υβριδοποίησης A: 5% SDS, 0.3M NaCl, 20mM NaH₂PO₄, 10mM Na₄P₂O₇, 2mM EDTA, αλλαγή του διαλύματος ανά 20min και παρακολούθηση των

κρούσεων στην επιφάνεια των μεμβρανών. Τα πλυσίματα έγιναν σε θερμοκρασία δωματίου, με διαλύματα που είχαν προθερμανθεί στους 65°C.

2) Για το διάλυμα υβριδοποίησης B: αρχικά με 2x SSC σε RT για 10min, ακολούθως με προθερμασμένο 0,5x SSC, 1% sarcosyl στους 65 °C για 30min (2 φορές) και τέλος με 0,2x SSC σε RT. Όλα τα πλυσίματα πραγματοποιήθηκαν με συνεχή ανάδευση.

Αυτοραδιογραφία

Η έκθεση των φίλτρων εξαρτάται από την αναμενόμενη ένταση του σήματος. Ενδεικτικά, η διάρκεια για την γενωμική βιβλιοθήκη ήταν 1-2 μέρες παρουσία ειδικής οθόνης ενίσχυσης σήματος σε θερμοκρασία -80 °C ενώ για την χρωμοσωμοειδική βιβλιοθήκη 2-6 μέρες (καθώς ο ανιχνευτής ήταν ραδιοσημασμένος με α-³³P) σε RT με φιλμ που είχε φωτιστεί στιγμιαία.

Έκπλυση φίλτρου

Εφαρμόζεται για την απομάκρυνση των μορίων του ανιχνευτή από την επιφάνεια των μεμβρανών, ώστε να επαναυβριδοποιηθούν. Πραγματοποιήθηκε με επώαση των μεμβρανών είτε σε 0.4N NaOH για 20min και στη συνέχεια σε 5x SSC για 10min σε RT (γενωμική βιβλιοθήκη) είτε σε 0.4N NaOH για 30min και στη συνέχεια σε 0,2M Tris.Cl pH 7,4, 0,1xSSC, 0,1% SDS για 20min στους 42°C (βιβλιοθήκη του χρωμοσώματος), υπό συνεχή ανάδευση. Στη συνέχεια οι μεμβράνες διατηρούνται στους 4°C, παρουσία ελάχιστης υγρασίας.

➤ **Υβριδοποίηση αποικιών ακινητοποιημένων σε μεμβράνη**

Η μέθοδος αυτή (Grunstein M & Hogness DS, 1975) εφαρμόστηκε για την ταυτοποίηση μοναδιαίων βακτηριακών αποικιών που φέρουν ανασυνδυασμένα πλασμίδια. Περιλαμβάνει τη μεταφορά των αποικιών σε φίλτρο νιτροκυτταρίνης, την λύση τους, την σύνδεση του DNA τους στη μεμβράνη και την ακόλουθη υβριδοποίηση των μεμβρανών με κατάλληλους, ραδιοσημασμένους, ανιχνευτές.

Μεταφορά των βακτηριακών αποικιών σε φίλτρο νιτροκυτταρίνης: Η μεμβράνη τοποθετείται στο τρυβλίο με τις βακτηριακές αποικίες μέχρι να υγρανθεί τελείως. Για λόγους στοίχισης τόσο η μεμβράνη όσο και το υποκείμενο θρεπτικό μέσο σημαδεύονται ασύμμετρα με βελόνα.

Λύση των βακτηρίων και σύνδεση του DNA τους στη μεμβράνη: Οι μεμβράνες σηκώνονται από το θρεπτικό υλικό και τοποθετούνται με τις αποικίες προς τα πάνω σε χαρτί Whatman 3MM εμποτισμένο μέχρι κορεσμού με διάλυμα αποδιάταξης (1,5M NaCl, 0,5M NaOH, pH 12-12,5) για 5min, ακολούθως σε διάλυμα ουδετεροποίησης (1M Tris, 1,5M NaCl, ρύθμιση του pH στο 7 με HCl) για 7min και τελικά σε διάλυμα 2xSSC για 15min. Τέλος, οι μεμβράνες τοποθετούνται σε στεγνό χαρτί όπου και

παραμένουν μέχρι να στεγνώσουν τελείως (>30 min) και υφίστανται ακτινοβολία με U.V. ώστε να πραγματοποιηθεί η ομοιοπολική σύνδεση μεταξύ DNA- μεμβράνης. Η υβριδοποίηση πραγματοποιήθηκε όπως και στην περίπτωση της γενωμικής βιβλιοθήκης.

2.22. Χρώση πηκτωμάτων

Η χρώση και η μονιμοποίηση των πολυπεπτιδίων πραγματοποιήθηκε με διάλυμα Coomassie brilliant blue [45% v/v μεθανόλη, 10% v/v οξικό οξύ, 0,25% χρωστικής Coomassie Brilliant blue R250 (διαλυτή σε μεθανόλη)], που ανιχνεύει 30-100ng πρωτεΐνης. Μετά την ηλεκτροφόρηση το πήκτωμα τοποθετείται στο διάλυμα της χρωστικής και είτε αναδεύεται ήπια από 2h έως ολονύκτια σε RT είτε θερμαίνεται σε φούρνο μικροκυμάτων για λίγα sec (3 φορές, μέχρι λίγο πριν το βρασμό). Στη συνέχεια το πήκτωμα ξεπλένεται με νερό και μεταφέρεται σε διάλυμα αποχρωματισμού (30% v/v μεθανόλη, 10% v/v οξικό οξύ και 3% v/v γλυκερόλη στην περίπτωση που ακολουθεί αποξήρανση) όπου και επωάζεται από 2h έως ολονύκτια, ώστε να απομακρυνθεί η χρωστική που δεν έχει δεσμευτεί στα πολυπεπτίδια.

2.23. Υγρή ηλεκτρομεταφορά

Μετά την ηλεκτροφόρηση των δειγμάτων, τα πολυπεπτίδια του πηκτώματος μεταφέρονται σε μεμβράνη νιτροκυτταρίνης (PROTRAN, μέγεθος πόρων 0,2 μm) με εφαρμογή ηλεκτρικού πεδίου σταθερής έντασης 110mA, σε διάλυμα ηλεκτρομεταφοράς (ίδιο με το διάλυμα ηλεκτροφόρησης της SDS-PAGE με 20% μεθανόλη), ολονύκτια στους 4°C. Η διάταξη περιλαμβάνει την τοποθέτηση στην ειδική κασέτα των εξής: Σπόγγος, χαρτί Whatman 3MM (x2)-πήκτωμα-μεμβράνη-χαρτί Whatman 3MM (x2)-σπόγγος και την εφαρμογή πεδίου ώστε οι πρωτεΐνες να μεταφέρονται προς την άνοδο.

2.24. Ανάλυση Western

Μετά την ηλεκτρομεταφορά, η μεμβράνη επωάζεται σε παρεμποδιστικό διάλυμα TBSTM (20mM Tris-Cl pH 7,6, 137mM NaCl, 0,1% Tween-20, 5% άπαχο γάλα σκόνη) ή PBSTM (1xPBS, 0,1% Tween-20, 5% άπαχο γάλα σκόνη) για τουλάχιστον 1h, ώστε να καλυφθούν οι ελεύθερες θέσεις της. Ακολουθεί επώαση με το πρωτεύον αντίσωμα, σε κατάλληλη αραιώση σε παρεμποδιστικό διάλυμα. Στη συνέχεια απομακρύνεται η περίσσεια του αντισώματος με τρία διαδοχικά πλυσίματα με διάλυμα TBST ή PBST (TBSTM / PBSTM χωρίς γάλα) διάρκειας 10min το καθένα και προστίθεται το κατάλληλο δευτερεύον αντίσωμα που είναι συζευγμένο με HRP. Μετά την απομάκρυνση του αντισώματος με πλυσίματα όπως και προηγουμένως

προστίθενται τα υποστρώματα της HRP που παρουσία του ενζύμου παράγουν φως το οποίο και αποτυπώνεται σε φιλμ. Χρησιμοποιήθηκε το σύστημα χημειοφωταύγειας της PIERCE σύμφωνα με τις οδηγίες της εταιρίας. Όλη η διαδικασία πραγματοποιείται σε RT υπό ανάδευση, εκτός από την περίπτωση που υπάρχει ολονύκτια επώαση που γίνεται στους 4°C. Τα αντισώματα, οι αραιώσεις, η προέλευσή τους και οι συνθήκες που εφαρμόστηκαν περιγράφονται στον Πιν.6.

2.25. Αντίδραση αντίστροφης μεταγραφής

Η μέθοδος αυτή, σε συνδυασμό με την PCR, χρησιμοποιήθηκε για τον έλεγχο της έκφρασης γονιδίων σε εγκέφαλο. Πριν την αντίδραση αντίστροφης μεταγραφής, το RNA που είχε απομονωθεί από ιστό επώαστηκε DNase I (Boehringer Mannheim). Χρησιμοποιήθηκαν 20 units ενζύμου/μg RNA, παρουσία RNasin (40 units) στο κατάλληλο ρυθμιστικό διάλυμα (200mM Tris-Cl pH 8.0, 50mM NaCl και 30mM MgCl₂) στους 37°C για 1h. Ακολούθησε εκχύλιση του RNA με φαινόλη/χλωροφόρμιο, κατακρήμνιση με 3.5 όγκους μίγματος αιθανόλης και οξικού αμμωνίου 7.5M (αναλογία μίγματος 6:1) και επαναδιαλυτοποίηση σε H₂O. Για την αντίδραση χρησιμοποιήθηκαν ως υπόστρωμα 200ng ολικού RNA από εγκέφαλο εμβρύου ανθρώπου. Αρχικά, το δείγμα, μετά την ρύθμιση του όγκου του στα 11,5μl, θερμάνθηκε για 5min στους 65°C και ακολούθως προστίθεται σε αυτό: 3μl (100ng/l) εκκινητής ολιγο-dT, 2μl ρυθμιστικού διαλύματος, 2μl dNTPs (2,5mM το καθένα), 0,5μl αντίστροφης μεταγραφάσης (Sensiscript RT, Qiagen) και 1μl (10units) RNAsin. Η αντίδραση επώαστηκε για 2h στους 37°C και καθαρίστηκε από τα μη ενσωματωμένα νουκλεοτίδια και τα άλατα με μοριακή διήθηση μέσω Sephadex G-50. Στην PCR που ακολούθησε χρησιμοποιήθηκε ως υπόστρωμα το 5% της αντίδρασης.

2.26. Υβριδοποίηση *in situ* σε τομές ιστών με ανιχνευτή RNA (RISH)

Με την μέθοδο αυτή είναι δυνατός ο προσδιορισμός της μεταγραφικής ενεργότητας ενός γονιδίου σε επίπεδο ιστού. Η μέθοδος πρωτοπεριγράφηκε με σκοπό τον εντοπισμό υβριδικών μορίων DNA:RNA σε κυτταρολογικά παρασκευάσματα (Gall & Pardue, 1969). Έκτοτε, παρουσιάστηκαν διάφορες τροποποιήσεις και βελτιώσεις της. Σημαντική καινοτομία ήταν η χρήση μονόκλωνων μορίων RNA ως ανιχνευτών, των ριβοανιχνευτών (Cox *et al*, 1984), που εμφανίζονται πιο ειδικοί και με μεγαλύτερη ευαισθησία σε σχέση με ανιχνευτές cDNA, ή ολιγονουκλεοτίδια.

Η μέθοδος περιλαμβάνει την προετοιμασία του ιστού, την σταθεροποίησή του σε εκμαγείο, τον τεμαχισμό του, την σύνθεση του ανιχνευτή, την υβριδοποίηση και την ανίχνευση του σήματος. Η υβριδοποίηση έγινε σε τομές εγκεφάλου ενηλίκου

ποντικού και εμβρύων ποντικού αναπτυξιακού σταδίου E12.5-E14.5. Οι ιστοί μονιμοποιήθηκαν με PFA και τεμαχίστηκαν σε κρουτόμο. Οι ριβοανιχνευτές που χρησιμοποιήθηκαν ήταν συμπληρωματικοί του mRNA (antisense probe) και παρήχθησαν με μεταγραφή *in vitro* χρησιμοποιώντας ως υπόστρωμα πλασμιδιακή κατασκευή σε φορέα που έχει, εκατέρωθεν των θέσεων κλωνοποίησης, αλληλουχίες πρόσδεσης της RNA πολυμεράσης (υποκινητή T7, T3). Πριν την μεταγραφή η πλασμιδιακή κατασκευή γραμμοποιείται ενώ κατά την αντίδραση χρησιμοποιείται UTP σημασμένο με DIG. Ο σημασμένος ανιχνευτής υβριδοποιείται στις τομές των ιστών όπου και αναγνωρίζει το στόχο-mRNA και ακολούθως εντοπίζεται με το αντίσωμα anti-DIG το οποίο είναι συζευγμένο με το ένζυμο αλκαλική φωσφατάση. Έτσι, όταν γίνει επώαση με τα υποστρώματα του ενζύμου εμφανίζεται χρώση στις περιοχές όπου εντοπίζεται το mRNA.

➤ **Προετοιμασία και μονιμοποίηση ιστών**

Με την μονιμοποίηση διατηρείται η μορφολογία των ιστών και συγκρατείται το mRNA μέσα στα κύτταρα. Ως μονιμοποιητικό μέσο επιλέχθηκε η PFA καθώς το ποσοστό του στόχου-mRNA που παγιδεύεται στα κύτταρα είναι σημαντικά μεγαλύτερο σε σχέση με άλλες μεθόδους ενώ παράλληλα διατηρείται καλύτερα η κυτταρική μορφολογία (Lawrence JB & Singer RH, 1985). Όσο περισσότερο διαρκεί η μονιμοποίηση, τόσο καλύτερα διατηρείται η μορφολογία του ιστού, παράλληλα όμως μειώνεται η προσβασιμότητα του στόχου- mRNA. Γενικά για ιστούς έως 2cm η μονιμοποίηση διαρκεί ολονυκτία και πραγματοποιείται στους 4°C με ήπια ανάδευση. Ο όγκος του μονιμοποιητικού μέσου ξεπερνά κατά πολύ (τουλάχιστον x30) τον όγκο του ιστού. Αμέσως μετά την λήψη των ιστών, τα δείγματα ξεπλένονται με 1xPBS (40mM Na₂HPO₄, 5mM NaH₂PO₄, 150mM NaCl, pH 7.4) και επωάζονται σε διάλυμα 4% PFA με πολύ ήπια ανάδευση στους 4°C για 12-36h. Ακολουθεί ξέπλυμά τους με 1xPBS (3x10min) και για λόγους κρουοπροστασίας ακολουθεί επώαση στους 4°C σε διάλυμα σουκρόζης 30% μέχρι τα παρασκευάσματα να καθιζάνουν στον πάτο του διαλύματος (24-36h). Αυτό το βήμα πραγματοποιείται επειδή ακολουθεί τομή των ιστών σε κρουτόμο.

➤ **Στερεοποίηση και τεμαχισμός του ιστού**

Μετά την απομάκρυνση της περίσσειας σουκρόζης, τα παρασκευάσματα τοποθετούνται σε πλαστικό εκμαγείο, που περιέχει παγωμένο μέσο Polyfreeze-Tissue Freezing που είναι κατάλληλο για τομές σε κρουστάτη και καλύπτονται πλήρως με το υλικό αυτό. Η τοποθέτησή τους εξαρτάται από τον άξονα κατά τον οποίο πρόκειται να τεμαχιστεί το δείγμα (στεφανιαίες, εγκάρσιες, οβελιαίες τομές).

Στη μορφή αυτή τα παρασκευάσματα παγώνουν στους -80°C όπου και μπορούν να διατηρηθούν ως τον τεμαχισμό τους. Οι τομές πραγματοποιήθηκαν σε κρυοτόμο (-20°C), είχαν πάχος 16 μm και μετά την τοποθέτησή τους σε αντικειμενοφόρες πλάκες Superfrost/Plus διατηρήθηκαν στους -80°C . Το πλεονέκτημα των αντικειμενοφόρων αυτού του τύπου είναι ότι δεν απαιτείται κάποια επεξεργασία πριν τη χρήση τους.

➤ **Παρασκευή ανιχνευτή**

Οι ανιχνευτές που χρησιμοποιήθηκαν είναι συμπληρωματικοί με το mRNA του γονιδίου ενώ σε όλα τα πειράματα χρησιμοποιήθηκε ως αρνητικός έλεγχος και ανιχνευτής ίδιος με το mRNA. Για την παρασκευή τους κλωνοποιήθηκαν τμήματα των cDNAs των γονιδίων *Lgi1*, *Lgi2*, *Lgi3*, *Lgi4*, *Fra10ac1*, *Dgcr14* και *Sap145* του ποντικού (Εικ. 10, 11) στον φορέα pBluescript II KS +. Το τμήμα του cDNA επιλέγεται συνήθως από το 3' UTR του γονιδίου γιατί η περιοχή αυτή είναι η πιο ειδική. Επιπλέον, αποφεύγονται αλληλουχίες μικρής πολυπλοκότητας, δηλαδή συνεχόμενες επαναλήψεις του ίδιου νουκλεοτιδίου. Το μέγεθος του τμήματος του cDNA ήταν μεταξύ 300bp και 1000bp, εύρος που αφενός επιτρέπει αποτελεσματική υβριδοποίηση και αφετέρου ικανοποιητική πρόσβαση στο στόχο-mRNA. Τέλος, αποφεύγονται αλληλουχίες από τον polylinker του φορέα, καθώς συχνά περιέχουν τμήματα πλούσια σε GC που τείνουν να υβριδοποιούνται μη ειδικά με το rRNA.

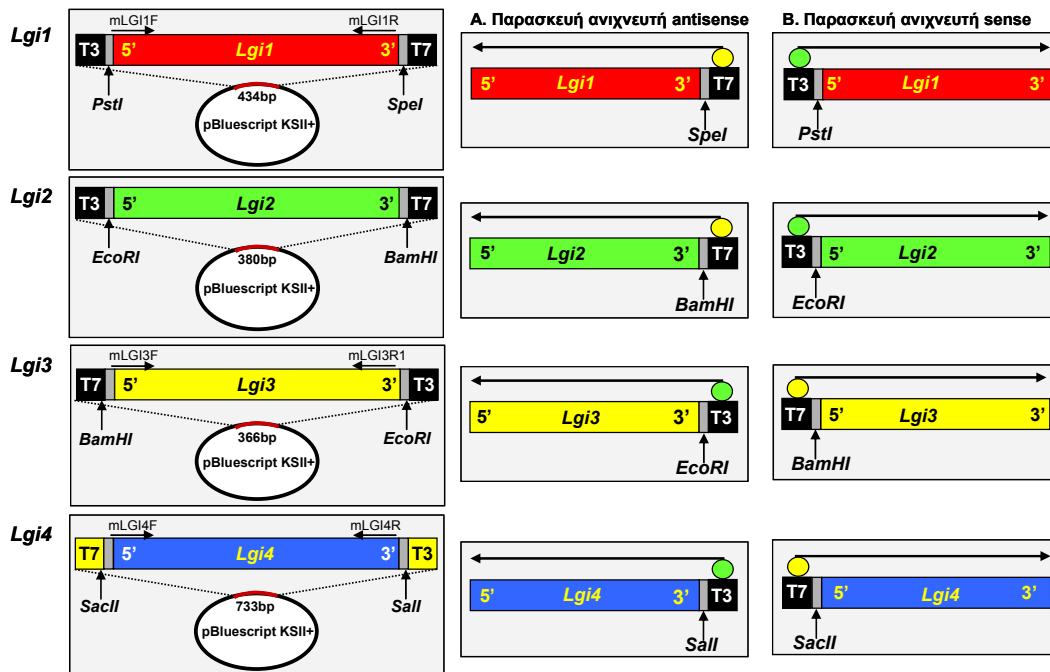
Οι ανιχνευτές παρασκευάστηκαν με αντίδραση μεταγραφής *in vitro* χρησιμοποιώντας ως υπόστρωμα 1 μg πλασμιδιακού DNA που είχε προηγουμένως γραμμοποιηθεί πλήρως με το κατάλληλο ένζυμο περιορισμού (Εικ.10, 11). Το γραμμοποιημένο DNA εκχυλίζεται με φαινόλη/χλωροφόρμιο, κατακρημνίζεται με CH_3COONa και απόλυτη αιθανόλη και επαναδιαλύεται σε TE. Πριν την προσθήκη των συστατικών της αντίδρασης, ο όγκος του υποστρώματος ρυθμίζεται στα 12 μl και το δείγμα θερμαίνεται για 10min στους 65°C . Ακολούθως προστίθενται: 2units RNAsin (2units/ μl) που δρα σαν αναστολέας των RNAσών, 4 μl ρυθμιστικού διαλύματος μεταγραφής 5x (5x: 400mM Tris-Cl pH 8, 60mM MgCl_2 , 100mM DTT, 20mM σπερμιδίνη), 2 μl μίγματος σήμανσης (DIG-RNA labeling mix, 10x:10mM ATP, 10mM CTP, 10mM GTP, 6,5mM UTP, 3,5mM DIG-11-UTP, pH 7,5), και 20units T3 ή T7 RNA πολυμεράσης (20units / μl). Αναλόγως με την φορά που έχει κλωνοποιηθεί το cDNA στον pBluescript II KS +, χρησιμοποιείται η κατάλληλη πολυμεράση (T3 ή T7) ώστε να παραχθεί ο συμπληρωματικός και ο μη συμπληρωματικός ανιχνευτής. Η αντίδραση επωάζεται για 2h στους 37°C και μετά από στιγμιαία φυγοκέντρηση προστίθενται 20units DNase I η οποία δρα για 15min στους 37°C ώστε να καταστρέψει το υπόστρωμα. Η αντίδραση σταματάει με την προσθήκη 1 μl EDTA

0,5M pH 8 και το mRNA κατακρημνίζεται με προσθήκη 2,5μl LiCl 4M και απόλυτης αιθανόλης (2,5 όγκους). Η πελέτα του mRNA επαναδιαλύεται σε 50μl H₂O και αποθηκεύεται στους -80°C

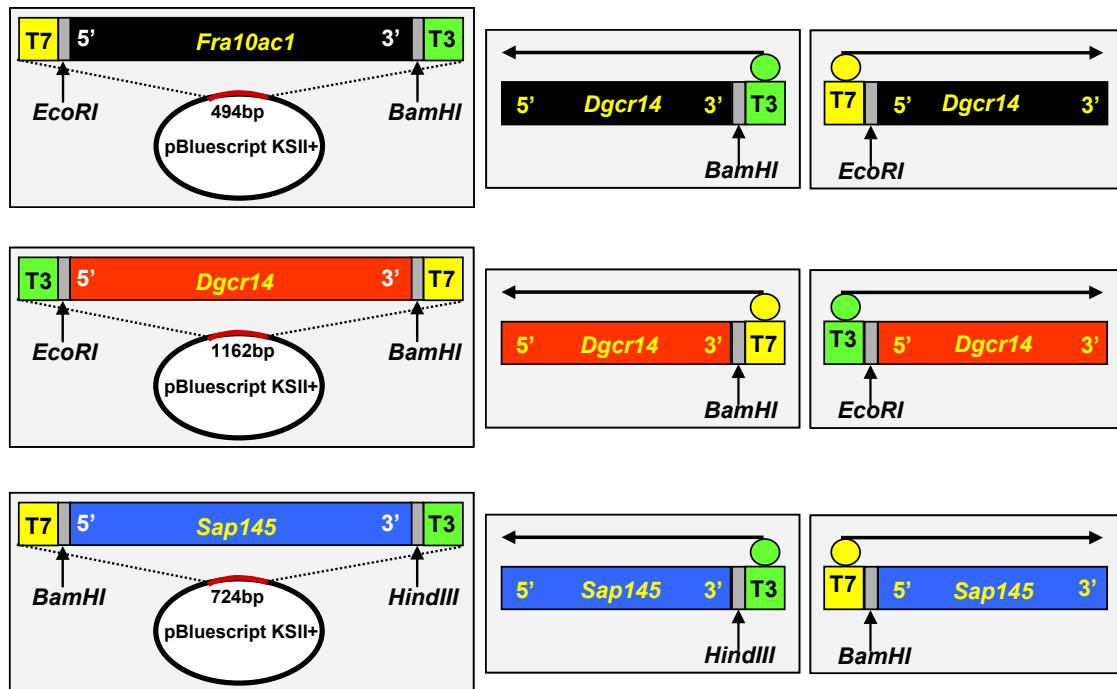
➤ Υβριδοποίηση

Οι τομές μεταφέρονται από τους -80°C σε RT όπου και στεγνώνουν (2-3h).

Για την υβριδοποίηση, προστίθεται σε κάθε αντικειμενοφόρο ο ανιχνευτής, σε ~150μl προθερμασμένου στους 65°C διαλύματος υβριδοποίησης [50% v/v φορμαμίδη, 5xSSC (20x: 3M NaCl, 0,3M κιτρικό νάτριο, ρύθμιση του pH στο 5,5 με κιτρικό οξύ), 5xDenhardts (50x: 1% Ficoll, 1% Polyvinyl pyrrolidone, 1% BSA Fraction V) 500μg/ml herring sperm DNA, 250μg/ml tRNA] σε συγκέντρωση 0,2-1μg /ml. Τα παρασκευάσματα καλύπτονται με parafilm και επωάζονται στους 65°C, ολονυκτία σε περιβάλλον κορεσμένο σε φορμαμίδη.



Εικ. 10: Σχηματική αναπαράσταση της κατασκευής των ριβοανιχνευτών με αντίδραση μεταγραφής *in vitro*. Ο ανασυνδυασμένος πλασμιδιακός φορέας γραμμοποιείται, με τεμαχισμό με ενδονουκλεάσες περιορισμού που βρίσκονται στον polylinker και όσον το δυνατόν πλησιέστερα στο ένθεμα. Η ενδονουκλεάση πλησίον του 5' του cDNA χρησιμοποιείται για την κατασκευή του υποστρώματος για την παραγωγή του συμπληρωματικού (antisense) ανιχνευτή, και αντιστρόφως. Ανάλογα με τον προσανατολισμό του ενθέματος, χρησιμοποιείται η κατάλληλη RNA πολυμεράση (T3 ή T7). Απεικονίζονται οι πλασμιδιακές κατασκευές στον φορέα pBluescript των cDNAs των *Lgi1-4*, ο προσανατολισμός και το μήκος των ενθεμάτων, οι ενδονουκλεάσες που χρησιμοποιήθηκαν για την γραμμοποίηση των ανασυνδυασμένων πλασμιδίων και η RNA πολυμεράση με την οποία έγινε η μεταγραφή σε κάθε περίπτωση, για την παρασκευή του συμπληρωματικού και του μη-συμπληρωματικού ανιχνευτή.



Εικ. 11: Σχηματική αναπαράσταση της κατασκευής των ριβοανιχνευτών με αντίδραση μεταγραφής *in vitro*. Απεικονίζονται οι πλασμιδιακές κατασκευές στον φορέα pBluescript των cDNAs των γονιδίων *Fra10ac1*, *Dgcr14* και *Sap145*, ο προσανατολισμός και το μήκος των ενθεμάτων, οι ενδονουκλεάσες που χρησιμοποιήθηκαν για την γραμμοποίηση των ανασυνδυασμένων πλασμιδίων και η RNA πολυμεράση με την οποία έγινε η μεταγραφή σε κάθε περίπτωση, για την παρασκευή του συμπληρωματικού και του μη-συμπληρωματικού ανιχνευτή.

➤ Πλυσίματα και επώαση με το αντίσωμα *anti-DIG*

Αρχικά, τα πλυσίματα γίνονται σε ειδικά γυάλινα δοχεία με τις αντικειμενοφόρους τοποθετημένες κατακόρυφα, με συνεχή ήπια ανάδευση στους 65°C. Χρησιμοποιούνται διαδοχικά τα διαλύματα I, II και III (και τα τρία περιέχουν 50% v/v φορμαμίδη, 0,1% v/v Tween-20 και SSC σε μειούμενη συγκέντρωση, 5x στο διάλυμα I, 1x στο II και 0,1x στο III), κάνοντας 3 επαναλαμβανόμενα πλυσίματα με το καθένα διάρκειας 30min. Ακολουθεί πλύσιμο με διάλυμα 1x MABT (5x: 100mM μαλαϊκό οξύ, 150mM NaCl, 0,5% Tween-20, ρύθμιση του pH στο 7,5 με NaOH) σε RT, μία φορά για 20min με τις αντικειμενοφόρες κάθετα με ήπια ανάδευση και δύο φορές επιπλέον σε οριζόντια θέση ώστε να μειωθεί το ποσοστό αποκόλλησης των ιστών. Από το στάδιο αυτό και μετά τα παρασκευάσματα τοποθετούνται σε δοχείο με περιβάλλον κορεσμένο σε H₂O. Ακολουθεί επώαση με παρεμποδιστικό διάλυμα B2 (10% FCS σε 1x MABT) για 2h με τα παρασκευάσματα σε οριζόντια θέση και καλυμμένα με parafilm. Στον ίδιο χρόνο πραγματοποιείται και η επώαση του anti-DIG-AP (Roche) σε διάλυμα B2 (αραίωση αντισώματος: 1/1000) το οποίο και προστίθεται στη συνέχεια στις τομές. Τα δείγματα επωάζονται ολονύκτια στους 4°C (καλυμμένα με parafilm).

➤ Πλυσίματα και αντίδραση αποκάλυψης

Μετά την επώαση με το αντίσωμα, ακολουθεί πλύσιμο με διάλυμα 1x MABT, μία φορά για 30min σε RT με τις αντικειμενοφόρες τοποθετημένες κατακόρυφα και ήπια ανάδευση. Στη συνέχεια και μέχρι το τέλος του πειράματος, τα παρασκευάσματα τοποθετούνται σε οριζόντια θέση και πλένονται δύο φορές από 30min με 1x MABT. Ακολουθούν δύο πλυσίματα, 30min το καθένα, με διάλυμα B3 (100mM Tris-Cl pH 9,5, 100mM NaCl, 50mM MgCl₂) που λόγω του αλκαλικού pH του προετοιμάζει τα δείγματα για την αντίδραση της αλκαλικής φωσφατάσης. Το διάλυμα B3 αντικαθίσταται από το διάλυμα της αντίδρασης αποκάλυψης B4 [B3 με NBT και BCIP σε αναλογία όγκων 9/7 (χρησιμοποιήθηκε B4 με 4,5μl NBT και 3,5μl BCIP ανά ml)], τα δείγματα καλύπτονται με parafilm και επωάζονται ολονύκτια στους 4°C. Η ένταση του σήματος ελέγχεται στο μικροσκόπιο και αν απαιτείται περαιτέρω χρώση πραγματοποιείται ξέπλυμα με διάλυμα B3 και επώαση μέχρι και 48h επιπλέον με φρέσκο διάλυμα B4 (ίδιας ή διπλάσιας συγκέντρωσης σε NBT και BCIP) στους 4°C (εναλλακτικά σε RT ή στους 37°C). Καθ' όλη τη διάρκεια της αντίδρασης, τα παρασκευάσματα βρίσκονται σε περιβάλλον κορεσμένο σε H₂O και σε σκοτάδι. Όταν η ένταση του σήματος κριθεί ικανοποιητική γίνεται πλύσιμο με 1xPBS σε κάθετη τοποθέτηση (3 επαναλήψεις από 10min), μονιμοποίηση με 4% PFA για 30min, ξέπλυμα με 1xPBS (2 επαναλήψεις από 10min) και με H₂O για 5min. Τέλος, στα δείγματα τοποθετούνται καλυπτρίδες, στις οποίες έχει προστεθεί μικρή ποσότητα Aquamount, και μετά από ολονύκτια επώαση σε απαγωγό εστία, στεγανοποιούνται και παρατηρούνται σε μικροσκόπιο. Χρησιμοποιήθηκε μικροσκόπιο τύπου Leica με διερχόμενο φως και οπτικά Nomarski. Η φωτογράφιση έγινε με ψηφιακή μηχανή τύπου Spot προσαρμοσμένη στο μικροσκόπιο.

2.27. Σύστημα δύο υβριδίων του σακχαρομύκητα

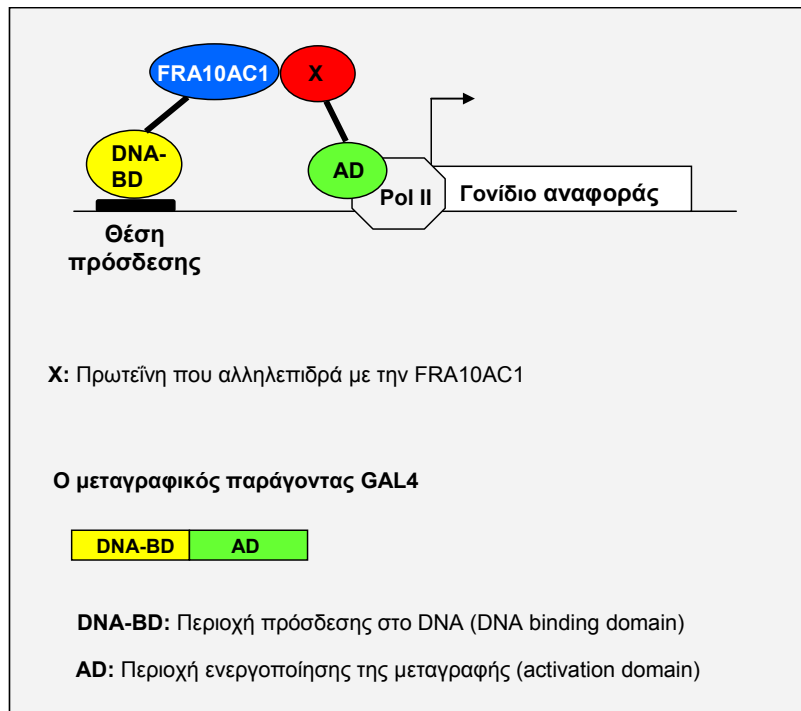
Οι πρωτεϊνικές αλληλεπιδράσεις είναι θεμελιώδεις σε όλες τις κυτταρικές λειτουργίες ενώ συχνά είναι δυνατόν να αντλήσουμε χρήσιμα δεδομένα για την λειτουργία μιας άγνωστης πρωτεΐνης μέσω ταυτοποίησης των πρωτεϊνών με τις οποίες αλληλεπιδρά. Η προσέγγιση που χρησιμοποιείται ευρέως για την μελέτη των πρωτεϊνικών δυαδικών αλληλεπιδράσεων είναι το σύστημα δύο υβριδίων του σακχαρομύκητα, μία ευαίσθητη *in vivo* μέθοδος. Το κύριο πλεονέκτημά της έναντι των εναλλακτικών *in vitro* προσεγγίσεων είναι ότι οι πρωτεΐνες που μελετώνται βρίσκονται σε περιβάλλον που προσομοιάζει με το φυσιολογικό τους, γεγονός που επιτρέπει την πραγματοποίηση μετα-μεταφραστικών τροποποιήσεων και την απόκτηση της σωστής τρισδιάστατης δομής του μορίου.

Το σύστημα στηρίζεται στην ύπαρξη δύο ανεξάρτητα λειτουργικά περιοχών σε αρκετούς μεταγραφικούς παράγοντες, μίας περιοχής πρόσδεσης στο DNA και μία περιοχή ενεργοποίησης της μεταγραφής, μέσω αλληλεπίδρασης με το σύμπλοκο έναρξης της μεταγραφής. Για να δράσουν οι δύο παραπάνω περιοχές πρέπει να είναι η μία πλησίον της άλλης, δεν είναι όμως αναγκαίο να ανήκουν στην ίδια πολυπεπτιδική αλυσίδα. Στο σύστημα δύο υβριδίων, οι δύο διαφορετικές περιοχές ενός μεταγραφικού παράγοντα γεινιάζουν όταν παραχθούν σε σύντηξη με δύο πρωτεΐνες που αλληλεπιδρούν και με αυτόν τον τρόπο επιτυγχάνεται η έκφραση ενός γονιδίου-αναφοράς. Η πρώτη εκδοχή της μεθόδου (Fields S & Song O, 1989) βασίστηκε στην ανασύσταση του μεταγραφικού παράγοντα GAL4. Κατά τα επόμενα χρόνια το σύστημα εξελίχθηκε και τροποποιήθηκε, χρησιμοποιώντας εναλλακτικές περιοχές πρόσδεσης στο DNA και περιοχές ενεργοποίησης της μεταγραφής (Gyuris J *et al*, 1993) ενώ για να αυξηθεί η αυστηρότητα της μεθόδου αναπτύχθηκαν καινούρια στελέχη σακχαρομύκητα που φέρουν διάφορα γονίδια-αναφοράς, το καθένα με διαφορετικό υποκινητή.

Στην παρούσα εργασία, χρησιμοποιήθηκε το σύστημα του μεταγραφικού παράγοντα GAL4. Συνοπτικά, το cDNA του *FRA10AC1* (δόλωμα) κλωνοποιήθηκε στον φορέα pGBT9 στο ίδιο πλαίσιο ανάγνωσης με την περιοχή πρόσδεσης στο DNA (DNA-BD) του GAL4. Το χιμαιρικό αυτό μόριο χρησιμοποιήθηκε για την διερεύνηση βιβλιοθήκης cDNA ήπατος ανθρώπου η οποία έχει κατασκευαστεί στον φορέα pACT2 που περιέχει την περιοχή ενεργοποίησης της μεταγραφής (AD) του GAL4 (θήραμα), Εικ. 12. Ο μετασχηματισμός εφαρμόστηκε σε κύτταρα σακχαρομύκητα στελέχους ρJ69-4α όπου και οι δύο χιμαιρικές πρωτεΐνες κατευθύνονται στον πυρήνα. Εκεί και εφόσον γίνει ανασύσταση του λειτουργικού παράγοντα, μέσω αλληλεπίδρασης της *FRA10AC1* με μία από τις πρωτεΐνες της βιβλιοθήκης, ενεργοποιούνται τα γονίδια αναφοράς. Η ενεργοποίησή τους ελέγχθηκε με καλλιέργεια σε θρεπτικά μέσα επιλογής. Οι θετικές αποικίες συλλέχθηκαν και μετά την εξάλειψη του δολώματος, απομονώθηκε και ταυτοποιήθηκε το θήραμα.

➤ **Έλεγχοι καταλληλότητας της βιβλιοθήκης cDNA**

Η βιβλιοθήκη cDNA που διερευνήθηκε ελέγχθηκε με PCR για την επιβεβαίωση της παρουσίας του *FRA10AC1*. Ο αντίστοιχος φορέας pACT2 περιέχει την περιοχή ενεργοποίησης της μεταγραφής του GAL4 (114 αμινοξέα). Η υβριδική πρωτεΐνη εκφράζεται σε υψηλά επίπεδα από τον συστατικό υποκινητή ADH1 και κατευθύνεται στον πυρήνα. Ο pACT2 φέρει το γονίδιο *LEU2* που επιτρέπει στα κύτταρα του σακχαρομύκητα να αναπτύσσονται σε θρεπτικό μέσο από το οποίο απουσιάζει η λευκίνη.



Εικ. 12: Αρχή του συστήματος δύο υβριδίων. Ως δόλωμα χρησιμοποιήθηκε η χιμαιρική πρωτεΐνη DNA-BD/FRA10AC1. Εφόσον αυτή αλληλεπιδρά με κάποια από τις πρωτεΐνες που κωδικοποιούνται από τους κλώνους της βιβλιοθήκης (X), παρατηρείται ενεργοποίηση των γονιδίων-δεικτών του σακχαρομύκητα

➤ **Επιλογή του στελέχους του σακχαρομύκητα**

Το στέλεχος που χρησιμοποιήθηκε ήταν το ρJ69-4α που περιέχει τρία επαγόμενα γονίδια-δείκτες κάτω από τον έλεγχο διαφορετικών υποκινητών. Συγκεκριμένα, στο στέλεχος αυτό: α) το γονίδιο HIS3 υπό τον έλεγχο του υποκινητή GAL1, έχει ενσωματωθεί στο 3' του γονιδίου LYS2, β) το γονίδιο ADE2 υπό τον έλεγχο του υποκινητή GAL2, έχει αντικαταστήσει τον γενετικό τόπο ADE2 και 3) το γονίδιο lacZ υπό τον έλεγχο του υποκινητή GAL7, ενσωματώθηκε στην κωδική περιοχή του MET2. Έτσι, τα κύτταρα για να αναπτυχθούν χρειάζονται στο θρεπτικό τους υλικό His και Ade. Οι τρεις παραπάνω διαφορετικοί υποκινητές παρουσιάζουν ομοιότητα μεταξύ τους μόνο στην περιοχή των εκφυλισμένων θέσεων πρόσδεσης του GAL4 με πλεονέκτημα η ενεργοποίηση των γονιδίων αναφοράς από ψευδώς θετικά μόρια να μην εξαρτάται από έναν μόνο υποκινητή.

➤ **Κλωνοποίηση του cDNA του FRA10AC1 στον ρGBT9**

Το cDNA (ελλιμματικό και πλήρους μήκους) του FRA10AC1 κλωνοποιήθηκε στον φορέα ρGBT9 στο ίδιο αναγνωστικό πλαίσιο με την DNA-BD (147 αμινοξέα) του GAL4. Ο ρGBT9, όπως συμβαίνει και με τους περισσότερους φορείς νέας γενιάς, παράγει την χιμαιρική πρωτεΐνη σε σχετικά χαμηλά επίπεδα, καθώς περιέχει μόνο μέρος (410 από τις 1500bp) του υποκινητή ADH1, για την αποφυγή φαινομένων

τοξικότητας. Επίσης φέρει το γονίδιο *TRP1* που επιτρέπει στα κύτταρα του σακχαρομύκητα να αναπτύσσονται σε θρεπτικό μέσο χωρίς τρυπτοφάνη.

➤ **Έλεγχος αυτόνομης ενεργοποίησης**

Το στάδιο ελέγχου της αυτόνομης ενεργοποίησης είναι απαραίτητο πριν την ανάλυση των δύο υβριδίων (Walhout AJ & Vidal M, 2001). Αυτόνομη ενεργοποίηση είναι η έκφραση των γονιδίων-αναφοράς του σακχαρομύκητα, παρουσία μόνο του δολώματος. Αν η αυτόνομη ενεργοποίηση είναι ασθενής και περιορισμένη μόνο στο γονίδιο *HIS3*, το πρόβλημα είναι πιθανόν να ξεπεραστεί χρησιμοποιώντας υψηλότερες συγκεντρώσεις 3-AT στο θρεπτικό (βλ. παρακάτω). Στην περίπτωση που η αυτόνομη ενεργοποίηση είναι ισχυρή και αφορά όλα τα γονίδια-αναφοράς, είναι απαραίτητο να αφαιρεθεί το τμήμα της πρωτεΐνης που ευθύνεται γι' αυτό.

Ο έλεγχος έγινε με μετασχηματισμό του στελέχους rJ69-4a με το δόλωμα και με τον φορέα pACT2 και καλλιέργεια των κυττάρων σε θρεπτικό μέσο επιλογής. Παράλληλα, ως αρνητικός έλεγχος, πραγματοποιήθηκε μετασχηματισμός με μίγμα DNA των δύο φορέων. Τα μετασχηματισμένα κύτταρα καλλιεργήθηκαν αρχικά σε πλήρες στερεό θρεπτικό υλικό (SC με 1xUra, 1/20xAde, 1xHis) στους 30°C για 2 μέρες και στη συνέχεια απομονώθηκαν 6 μοναδιαίες αποικίες που επανακαλλιεργήθηκαν σε στοίχιση στις ίδιες συνθήκες. Έτσι κατασκευάστηκε το πρωτότυπο πιάτο (master plate) το οποίο και αντιγράφηκε (replicas) σε τρυβλία που περιείχαν πλήρες θρεπτικό υλικό (θετικός έλεγχος ανάπτυξης) και θρεπτικό υλικό επιλογής. Χρησιμοποιήθηκαν τρία θρεπτικά υλικά επιλογής για να ελεγχθεί η αυτόνομη ενεργοποίηση των γονιδίων-αναφοράς *HIS3*, *ADE2* και *lacZ*, ως εξής:

<u>Γονίδιο-δείκτης</u>	<u>Θρεπτικό υλικό</u>
<i>HIS3</i>	SC+ Ura, Ade, 3/5/10/15/20mM AT
<i>ADE2</i>	SC+ Ura, His
<i>LacZ</i>	SC +Ura, Ade, His, 70mM K _x PO ₄ , 100μg/ml X-gal

Από όλα τα θρεπτικά υλικά απουσίαζαν η τρυπτοφάνη και η λευκίνη ώστε να επιλέγονται τα διπλά μετασχηματισμένα κύτταρα. Επίσης, στα θρεπτικά υλικά επιλογής για την ενεργοποίηση του *HIS3*, προστέθηκε 3-AT σε συγκεντρώσεις από 3-20mM, που δρα σαν συναγωνιστικός αναστολέας του ενζύμου που παράγεται από το γονίδιο *HIS3* και το οποίο παρουσιάζει μία βασική χαμηλή ενεργότητα.

Μετασχηματισμός σακχαρομύκητα

Ο μετασχηματισμός έγινε με την μέθοδο του CH_3COOLi . Κατά αυτήν τα κύτταρα καθίστανται δεκτικά με επεξεργασία με ιόντα λιθίου και PEG και υφίστανται θερμικό σοκ κατά το οποίο εισάγουν το DNA. Η απόδοση του μετασχηματισμού είναι της τάξεως του 10^5 κύτταρα/μg DNA. Αναλυτικότερα, 50ml θρεπτικού μέσου YEP εμπλουτισμένου με 0,25mM αδενίνη εμβολιάστηκε με μοναδιαία αποικία κυττάρων ρJ69-4α και επώαστηκε με ανάδευση στους 30°C ολονύκτια. Βάσει της $\text{O.D}_{550\text{nm}}$, η καλλιέργεια αραιώθηκε με φρέσκο θρεπτικό μέσο (μέχρι την τιμή 0,2) και ο όγκος της ρυθμίστηκε βάσει του αριθμού των επιθυμητών μετασχηματισμών, λαμβάνοντας υπ' όψιν ότι για καθένα απαιτούνται 20ml. Η επώαση εξακολούθησε στις ίδιες συνθήκες μέχρι η O.D. να φτάσει την τιμή 0,8. Τα κύτταρα συλλέχθηκαν με φυγοκέντρηση (2500rpm σε RT για 15min) ξεπλύθηκαν με ίσο όγκο ddH_2O , επανασυλλέχθηκαν με στις ίδιες συνθήκες και τελικά επαναιωρήθηκαν σε διάλυμα οξικού λιθίου 100mM (300μl/μετασχηματισμό). Το διάλυμα μετά την μεταφορά του σε erpendorf (ξεχωριστό για κάθε μετασχηματισμό) φυγοκεντρήθηκε στιγμιαία και στην πελλέτα των κυττάρων προστέθηκε με την ακόλουθη σειρά: α) 240 μl διαλύματος 50% PEG (MW 3350) στο οποίο επαναιωρήθηκαν τα κύτταρα, β) 25 μl αποδιατεταγμένου DNA-φορέα (2μg/μl), γ) 50 μl διαλύματος DNA. Χρησιμοποιήθηκε 0,5 μg πλασμιδιακού DNA για κάθε μετασχηματισμό το οποίο δεν είχε επωαστεί με RNAση, καθώς το RNA συμβάλλει στον αποδοτικότερο μετασχηματισμό, δ) 36 μl διαλύματος οξικού λιθίου 1M. Ακολούθησε επώαση στους 30°C για 30min, ανάδευση και θερμικό σοκ στους 42°C για 25min. Στη συνέχεια τα κύτταρα μεταφέρθηκαν σε πάγο όπου επώαστηκαν για 10min και συλλέχθηκαν με στιγμιαία φυγοκέντρηση. Στη συνέχεια επαναιωρήθηκαν σε 100 μl θρεπτικού μέσου SC και επιστρώθηκαν σε τρυβλία με SC εμπλουτισμένο με τα κατάλληλα αμινοξέα ή νουκλεοτίδια, ανάλογα με την επιλογή. (Πριν την καλλιέργεια των κυττάρων, τα τρυβλία τοποθετήθηκαν σε απαγωγό για ~1h και ακτινοβολήθηκαν με U.V για 15min). Οι αποικίες των κυττάρων σχηματίζονται μετά από επώαση στους 30°C για 2 μέρες.

Αντιγραφή πρωτότυπου τρυβλίου

Στο πρωτότυπο τρυβλίο οι αποικίες έχουν αναπτυχθεί κατά πολύ (διάμετρο ~3mm) και είναι διατεταγμένες σε στοίχιση. Κατά την αντιγραφή του, το πιάτο τοποθετείται πάνω σε ένα κομμάτι βελούδο, που είναι προσαρμοσμένο στην ειδική συσκευή, έτσι ώστε μέρος των αποικιών να μεταφερθεί σε αυτό. Στη συνέχεια τοποθετούνται στο βελούδο, διαδοχικά, τα τρυβλία με τα θρεπτικά μέσα επιλογής επιτυγχάνοντας έτσι την μεταφορά των αρχικών αποικιών (αντίγραφα). Η ανάπτυξή τους πραγματοποιείται με επώαση στους 30°C για 2-4 μέρες.

➤ **Τιτλοδότηση της βιβλιοθήκης**

Σκοπός της τιτλοδότησης είναι ο προσδιορισμός της ελάχιστης ποσότητας του DNA που, αναλογικά, έχει μέγιστη απόδοση στο μετασχηματισμό κυττάρων που περιέχουν ήδη το δόλωμα.

Αρχικά, πραγματοποιήθηκε μετασχηματισμός του στελέχους ρJ69-4α με το δόλωμα και καλλιέργειά του σε στερεό θρεπτικό μέσο SC εμπλουτισμένου με 1x Ade, 1x His, 1x Ura και 1x Leu. Η τιτλοδότηση περιλάμβανε την ανάπτυξη μοναδιαίας αποικίας σε υγρό θρεπτικό μέσο SC με 1/20x Ade, 1x His, 1x Ura, 1x Leu και μετασχηματισμό των κυττάρων με διαφορετικές ποσότητες DNA της βιβλιοθήκης (0,02-5μg). Τα διπλά μετασχηματισμένα κύτταρα καλλιεργήθηκαν σε στερεό θρεπτικό μέσο SC με 1/20x Ade, 1x His, 1x Ura, όπου και καταμετρήθηκε ο αριθμός των αποικιών σε κάθε περίπτωση και επιλέχθηκε η αποδοτικότερη ποσότητα του DNA.

➤ **Μετασχηματισμός κυττάρων που περιέχουν το δόλωμα με DNA της βιβλιοθήκης και καλλιέργεια σε θρεπτικό μέσο επιλογής**

Κατά την διερεύνηση της βιβλιοθήκης είναι κρίσιμο να υπάρχει καλή αντιπροσώπευση των κλώνων, ακόμα και εκείνων που αντιστοιχούν σε γονίδια με χαμηλή μεταγραφική ενεργότητα. Για τη διερεύνηση μέχρι κορεσμού, λήφθηκε υπ' όψιν το ποσοστό των μη ανασυνδυασμένων κλώνων και το μέσο μέγεθος των ενθεμάτων. Βάσει της τιτλοδότησης επιλέχθηκε η ποσότητα του DNA με την βέλτιστη απόδοση αναλογικά, και επαναλήφθηκε ο μετασχηματισμός με την συγκεκριμένη ποσότητα, τόσες φορές ώστε συνολικά να προκύπτουν 5×10^6 αποικίες. Τα διπλά μετασχηματισμένα κύτταρα καλλιεργήθηκαν σε 2 τρυβλία διαστάσεων 23x23cm με θρεπτικό μέσο SC εμπλουτισμένου με 1/20x Ade, 1x Ura, 3mM AT. Η επώαση πραγματοποιήθηκε στους 30°C και η συλλογή των θετικών αποικιών έγινε μετά από 4-7 μέρες.

➤ **Επιβεβαίωση της ενεργοποίησης των γονιδίων αναφοράς**

Οι αποικίες που αναπτύχθηκαν, επανακαλλιεργήθηκαν σε στερεό θρεπτικό μέσο SC με 1/20x Ade, 1x Ura, 3mM AT και αποτέλεσαν το πρωτότυπο τρυβλίο. Από αυτό δημιουργήθηκαν πιάτα-αντίγραφα που περιείχαν είτε το ίδιο θρεπτικό μέσο (θετικός έλεγχος ανάπτυξης) είτε θρεπτικά μέσα επιλογής, ώστε να επιβεβαιωθεί η ενεργοποίηση των γονιδίων-αναφοράς. Χρησιμοποιήθηκαν τα εξής θρεπτικά υλικά:

<u>Γονίδιο-δείκτης</u>	<u>Θρεπτικό υλικό</u>
HIS3	SC+ 1xUra, 1/20x Ade, 3 και 10mM AT
ADE2	SC+ 1xUra, 1xHis
LacZ	SC +1xUra, 1/20x Ade, 1xHis, 70mM K_2PO_4 , 100μg/ml X-gal

➤ **Εξάλειψη του δολώματος**

Κάθε αποικία, που πέρασε τον έλεγχο ανάπτυξης στα θρεπτικά υλικά επιλογής, αναπτύχθηκε σε 1,5ml υγρού θρεπτικού μέσου SC εμπλουτισμένου με Ade, His, Ura και από την κορεσμένη καλλιέργεια απομονώθηκε DNA (μίγμα γενωμικού και πλασμιδιακού). Από αυτό το 1/10 χρησιμοποιήθηκε για μετασχηματισμό *E. coli* στελέχους KC8 που, καθώς φέρουν μεταλλαγή στο γονίδιο *trpC*, δεν μπορούν να παράγουν ενδογενώς τρυπτοφάνη. Τα κύτταρα καλλιεργήθηκαν σε στερεό θρεπτικό μέσο M9 που περιείχε Ura, Ade, His, Trp και 100μg/ml αμπικιλίνη στους 37°C για ~20h. Η παρουσία της τρυπτοφάνης είχε ως συνέπεια την εξάλειψη του δολώματος.

Απομόνωση γενωμικού και πλασμιδιακού DNA από σακχαρομύκητα

Θρεπτικό μέσο SC (~2ml) με 0,25mM Ade, 0,33mM His και 0,2mM Ura εμβολιάζεται με μοναδιαία αποικία κυττάρων rJ69-4α και επωάζεται στους 30°C υπό ανάδευση για 16-20h. Τα κύτταρα συλλέγονται με φυγοκέντρηση στις 2000 rpm για 10min, ξεπλένονται με H₂O και επανασυλλέγονται στις ίδιες συνθήκες. Στην πελλέτα των κυττάρων προστίθενται 200μl διαλύματος διάρρηξης (2% v/v Triton X-100, 1% v/v SDS, 100mM NaCl, 10mM Tris-Cl pH 8, 1mM EDTA pH 8), 200μl γυάλινων σφαιριδίων και 200μl μίγματος φαινόλης/χλωροφορμίου. Ακολουθεί ισχυρή ανάδευση (vortex) για 3-4min και προσθήκη 200μl H₂O. Το δείγμα φυγοκεντρείται για 5min στις 12000 rpm σε RT, το υπερκείμενο της φυγοκέντρησης συλλέγεται και το δείγμα επανεκχυλίζεται. Τα νουκλειικά οξέα κατακρημνίζονται με CH₃COONa και απόλυτη αιθανόλη και επαναδιαλύονται σε 40μl TE.

➤ **Απομόνωση και ταυτοποίηση του θηράματος**

Μία μοναδιαία αποικία από κάθε μετασχηματισμό των κυττάρων KC8, καλλιεργήθηκε σε 2ml θρεπτικού μέσου LB με 100μg/ml αμπικιλίνη. Από κάθε καλλιέργεια απομονώθηκε πλασμιδιακό DNA με το οποίο μετασχηματίστηκε το στέλεχος DH5α. Από αυτό απομονώθηκε πλασμιδιακό DNA, που επεξεργάστηκε ώστε να είναι κατάλληλο για να χρησιμοποιηθεί σε αντίδραση αλληλούχησης.

2.28. Κυτταροκαλλιέργεια

Οι κυτταρικές σειρές COS7, HeLa, CV1, NIH-3T3 χρησιμοποιήθηκαν για την παραγωγή πρωτεϊνών και για την μελέτη του υποκυτταρικού εντοπισμού ενδογενών ή ανασυνδυασμένων πρωτεϊνών. Τα κύτταρα αναπτύσσονται σχηματίζοντας μονοστοιβάδες, στους 37°C παρουσία 5% CO₂.

➤ **Διατήρηση κυτταροκαλλιιεργειών**

Η κυτταροκαλλιέργεια διατηρείται σε φλάσκες (T-75 ή T-25) ενώ οι πειραματικές διαδικασίες πραγματοποιούνται σε τρυβλία διαμέτρου 3 ή 10cm. Τα κύτταρα διαχωρίζονται και συλλέγονται σε εκθετική φάση ανάπτυξης, δηλαδή όταν έχουν καλύψει ~70% της διαθέσιμης επιφάνειας. Κατά προσέγγιση, η ανανέωση της καλλιέργειας πραγματοποιείται κάθε 3-4 μέρες. Η αποκόλληση των κυττάρων από την επιφάνεια καλλιέργειας γίνεται με τρυψίνη, ως εξής: Μετά την απομάκρυνση του θρεπτικού μέσου, τα κύτταρα ξεπλένονται δύο φορές με ικανή ποσότητα διαλύματος 1xPBS και επωάζονται με διάλυμα τρυψίνης (σε 1xPBS) προθερμασμένου στους 37°C (~4ml για φιάλη T-75). Μόλις αρχίσει η αποκόλληση των κυττάρων (2-10min), προστίθεται θρεπτικό μέσο DMEM (προθερμασμένο στους 37°C) εμπλουτισμένου με FBS που αναστέλλει της δράσης της τρυψίνης (8-10ml για φιάλη T-75). Το 1/6-1/8 των επαναιωρημένων κυττάρων παραμένει στη φιάλη και ο όγκος συμπληρώνεται με θρεπτικό υλικό, το οποίο ανανεώνεται μετά την προσκόλληση των κυττάρων (~3-4h). Εφόσον ακολουθεί καλλιέργεια σε τρυβλία, μετά τον υπολογισμό του αριθμού των κυττάρων (με αιμοκυτταρόμετρο) γίνεται η συλλογή τους με φυγοκέντρηση (7min, 1200rpm, RT) και η επαναιώρησή τους, αρχικά σε ~0,5ml και ακολούθως στον κατάλληλο όγκο DMEM (+FBS). Το κυτταρικό εναιώρημα μεταφέρεται σε τρυβλία σε ποσότητα εξαρτώμενη από τον επιθυμητό αριθμό κυττάρων.

➤ **Κρυοδιατήρηση κυττάρων**

Τα κύτταρα πρέπει να φτάσουν σταδιακά στη τελική θερμοκρασία κατάψυξης, ώστε να προηγηθεί η αφυδάτωση του ενδοκυτταρικού σχηματισμού κρυστάλλων. Ως κρυοπροστατευτικό μέσο χρησιμοποιείται το DMSO (dimethylsulfoxide, cell culture grade, Sigma ή ATCC). Η κατάψυξή τους γίνεται όταν βρίσκονται σε εκθετική φάση ανάπτυξης, (~70% κάλυψη της διαθέσιμης επιφάνειας για κύτταρα που αναπτύσσονται σχηματίζοντας μονοστοιβάδα). Μετά το ξέπλυμα, την τρυψινόλυση και την συλλογή τους, επαναιωρούνται σε DMEM με FBS σε συγκέντρωση 10^6 - 10^7 /ml. Στη συνέχεια προστίθεται ίσος όγκος διαλύματος 20% DMSO (σε DMEM με 10% FBS). Η προσθήκη του διαλύματος αυτού γίνεται στάγδην ενώ το εναιώρημα των κυττάρων περιδινείται συνεχώς ώστε να αποφευχθεί το οσμωτικό σοκ και να ελαχιστοποιηθεί η αύξηση της θερμοκρασίας που προκαλείται από την διάλυση του DMSO. Ακολουθεί μεταφορά του δείγματος σε κατάλληλο φιαλίδιο που, για λόγους μόνωσης, τυλίγεται με χαρτί και επωάζεται για 20-30min στους -20°C. Στη συνέχεια το δείγμα μεταφέρεται στους -80°C όπου και παραμένει ολονύκτια και τελικά αποθηκεύεται σε δεξαμενή υγρού αζώτου (-196 °C ως -156 °C).

➤ **Επανακαλλιέργεια**

Τα κύτταρα που έχουν καταψυχθεί σε υγρό άζωτο και πρόκειται να επανακαλλιεργηθούν, μεταφέρονται σε υδατόλουτρο 37°C όπου και αναδεύονται ήπια μέχρι την πλήρη τήξη του θρεπτικού μέσου (1-2min). Η ταχεία τήξη εμποδίζει τον σχηματισμό ενδοκυτταρικών κρυστάλλων κατά τη διαδικασία της ενυδάτωσης τους. Ακολουθεί μεταφορά στη φιάλη καλλιέργειας και προσθήκη DMEM (+FBS). Μετά την προσκόλληση των κυττάρων (3-4h) το θρεπτικό μέσο αντικαθίσταται.

2.29. Παροδική διαμόλυνση κυττάρων θηλαστικών με πλασμιδιακό DNA

Με την παροδική διαμόλυνση, πραγματοποιείται εισαγωγή αντιγράφων του DNA στο κύτταρο χωρίς να γίνεται ενσωμάτωση στο γονιδίωμα. Με αυτήν τη μέθοδο επιτυγχάνονται υψηλά επίπεδα έκφρασης του διαγονιδίου, η παραγωγή όμως του προϊόντος είναι παροδική. Ωστόσο, τα επίπεδα των πρωτεϊνών είναι ικανοποιητικά ώστε να χρησιμοποιηθούν σε βιοχημικές μελέτες, σε πειράματα υποκυτταρικού εντοπισμού, σε ανίχνευση πρωτεϊνικών αλληλεπιδράσεων κ.τ.λ.

Η μέθοδος παροδικής διαμόλυνσης που ακολουθήθηκε στηρίζεται στην συνκατακρήμνιση του DNA με $\text{Ca}_3(\text{PO}_4)_2$ (Graham FL & Van der Eb AJ, 1973). Κατά αυτήν, τα αντίστοιχα μόρια cDNA κλωνοποιούνται σε κατάλληλους φορείς έκφρασης ώστε να βρίσκονται υπό τον έλεγχο υποκινητών που έχουν δυνατότητα ισχυρής ενεργοποίησης (π.χ. CMV, SV40). Η μέθοδος στηρίζεται σε θετικά φορτισμένα μόρια-φορείς που σχηματίζουν σύμπλοκα-υπό μορφή λεπτόκοκκου ίζηματος-με το DNA και στη συνέχεια προσκολλώνται στην κυτταρική μεμβράνη και εισέρχονται στο κύτταρο με ενδοκύτωση. Το ποσοστό των διαμολυσμένων κυττάρων κυμαίνεται μεταξύ 5% και 20% ενώ η απόδοση εξαρτάται κυρίως από την καθαρότητα του πλασμιδιακού DNA και την καλή κατάσταση των κυττάρων. Η διαμόλυνση πραγματοποιείται 12-24h μετά την επίστρωση των κυττάρων ενώ η κάλυψη της διαθέσιμης επιφάνειας καλλιέργειας είναι ~50% και ενώ το θρεπτικό μέσο έχει ανανεωθεί 2-4h προτούτερα. Για κύτταρα που αναπτύσσονται σε τρυβλίο διαμέτρου 10cm προστίθενται 9ml θρεπτικού μέσου. Στο διάλυμα του DNA, αφού ρυθμιστεί με H_2O στα 450μl (τρυβλίο 10cm), προστίθενται 50μl διαλύματος CaCl_2 2,5M και μετά από καλή ανάμιξη ακολουθεί σταδιακή προσθήκη με ταυτόχρονη περιδίνηση 500μl διαλύματος 2xHEBS (280mM NaCl, 50mM Hepes, 5mM Na_2HPO_4 , pH 7,12). Ακολουθεί επώαση για 5-15min σε RT έτσι ώστε να σχηματιστεί λεπτόκοκκο κρυσταλλικό ίζημα $\text{Ca}_3(\text{PO}_4)_2$ -DNA, το οποίο και προστίθεται στάγδην στα κύτταρα. Στην περίπτωση χρήσης τρυβλίου 3cm, ο όγκος του θρεπτικού μέσου είναι 2,7ml και ο τελικός όγκος του μίγματος διαμόλυνσης 300μl, διατηρώντας τις αναλογίες των συστατικών σταθερές.

Μετά από επώαση 8-14h τα κύτταρα ξεπλένονται με διάλυμα 1xPBS ή HBSS (2 φορές) και επωάζονται περαιτέρω για 24-48h με νέο θρεπτικό μέσο.

2.30. Καθαρισμός συγγένειας πολυκλωνικού αντισώματος

Ο ορός που λαμβάνεται από το ανοσοποιημένο ζώο περιέχει πλήθος πρωτεϊνών και πιθανώς άλλα αντισώματα πλην του επιθυμητού. Επιπλέον, αν η ανασυνδυασμένη πρωτεΐνη-αντιγόνο ήταν σε σύντηξη με έναν επίτοπο, ο ορός περιέχει και αντισώματα έναντι αυτού. Εξ αιτίας αυτών εφαρμόστηκε καθαρισμός συγγένειας ώστε να απομονωθούν από τον ορό τα ειδικά αντισώματα (Roth J *et al*, 1985). Κατά αυτόν ο ορός επωάζεται με το αντιγόνο, το οποίο έχει ακινητοποιηθεί σε μεμβράνη νιτροκυτταρίνης και αφού απομακρυνθεί ότι δεν έχει δεσμευτεί στην μεμβράνη, γίνεται έκλουση του αντισώματος. Η διαδικασία εφαρμόστηκε για τον καθαρισμό του πολυκλωνικού ορού anti-FRA10AC1 και καθώς το αντιγόνο ήταν σε σύντηξη με την GST (Αποστολοπούλου Δ., Διπλωματική Εργασία, Πανεπιστήμιο Κρήτης 2003), η ανασυνδυασμένη πρωτεΐνη που χρησιμοποιήθηκε κατά τον καθαρισμό συγγένειας, έφερε τον επίτοπο His (βλ. Πλασμιδιακές κατασκευές).

Συγκεκριμένα, μετά την SDS-PAGE της ανασυνδυασμένης His-FRA10AC1 (~1mg), ένα μικρό κομμάτι του πηκτώματος κόπηκε και χρωματίστηκε με διάλυμα Coomassie brilliant blue ενώ στο υπόλοιπο εφαρμόστηκε υγρή ηλεκτρομεταφορά σε μεμβράνη. Στη συνέχεια και βάσει της χρώσης του πηκτώματος, η μεμβράνη κόπηκε ώστε να περιληφθεί μόνο η ανασυνδυασμένη πρωτεΐνη. Ακολούθησε επώαση, αρχικά με παρεμποδιστικό διάλυμα 1xTBSM για 1h σε RT και στη συνέχεια με τον ορό. Για την επεξεργασία 3ml δείγματος χρησιμοποιήθηκαν δύο τμήματα μεμβράνης που το καθένα περιείχε ~1mg πρωτεΐνης, ενώ η επώαση έγινε ολονύχτια στους 4°C. Ο ορός απομακρύνθηκε και η μεμβράνη ξεπλύθηκε με διάλυμα 1xTBS (20mM Tris-Cl pH 7,6, 137mM NaCl, 3x10min) σε RT. Η έκλουση του αντισώματος πραγματοποιήθηκε με επώαση με διάλυμα γλυκίνης 0,2M pH 1,9 για 20sec και στη συνέχεια για 3min σε RT. Υπό αυτές τις συνθήκες pH, το αντίσωμα αποδεσμεύεται και αμέσως στο διάλυμα έκλουσης προστέθηκε ίσος όγκος διαλύματος 1M Tris.Cl pH 8. Ακολούθησε έλεγχος της ποιότητας και του τίτλου του αντισώματος. Η πρώτη έκλουση χρησιμοποιήθηκε για την ανίχνευση της ενδογενούς πρωτεΐνης σε κύτταρα θηλαστικών και η δεύτερη σε ανάλυση Western και ανοσοκατακρήμνιση.

2.31. Ανοσοεντοπισμός πρωτεϊνών με μικροσκοπία φθορισμού

Ο υποκυτταρικός εντοπισμός των πρωτεϊνών αποτελεί σημαντική ένδειξη του βιολογικού τους ρόλου και είναι εφικτός με πειράματα ανοσοφθορισμού. Κατά αυτά, ανιχνεύεται με το κατάλληλο πρωτεύον αντίσωμα η ενδογενής πρωτεΐνη ή ο επίτοπος -στην περίπτωση ανασυνδυασμένης πρωτεΐνης με επισήμανση (tag)- και

ακολουθεί η αναγνώριση από το κατάλληλο δευτερεύον αντίσωμα που φέρει προσδεδωμένες φθορίζουσες ουσίες. Εναλλακτικά, ο υποκυτταρικός εντοπισμός των πρωτεϊνών που έχουν παραχθεί μετά από παροδική διαμόλυνση κυτταρικών σειρών, μπορεί να είναι άμεσος αν οι πρωτεΐνες παράγονται σε σύντηξη με την GFP. Επιπλέον, είναι δυνατή η ταυτόχρονη, συγκριτική μελέτη των προτύπων έκφρασης περισσότερων πρωτεϊνών, εφόσον χρησιμοποιηθούν τα κατάλληλα πρωτογενή και δευτερογενή αντισώματα.

Η μέθοδος εφαρμόστηκε για την μελέτη του υποκυτταρικού εντοπισμού ανασυνδυασμένων και ενδογενών πρωτεϊνών σε κύτταρα θηλαστικών COS7, HeLa, CV1 και NIH-3T3 που αναπτύχθηκαν πάνω σε καλυπτρίδες. Και στις δύο περιπτώσεις η διαδικασία περιλάμβανε τα εξής στάδια: τα κύτταρα ξεπλύθηκαν με 1x PBS και μονιμοποιήθηκαν με 2% ή 4% PFA (σε 1x PBS) για 5min. Όλα τα πλυσίματα της διαδικασίας επαναλαμβάνονταν τρεις φορές και η διάρκεια του καθενός ήταν 10min. Μετά την απομάκρυνση της PFA ακολούθησε ξέπλυμα με PBS και επώαση με διάλυμα 20mM γλυκίνης (διαλύτης: 1x PBS) για 10min, ώστε να δεσμευτούν τυχόν ελεύθερες αλδευδομάδες (quenching). Ακολούθως τα κύτταρα ξεπλύθηκαν με PBS επώαστηκαν με παραμποδιστικό διάλυμα (0.2% Triton-X, 2mM MgCl₂, 1-3% BSA ή 1% fish gelatin σε 1x PBS, Sigma) για 20min και στη συνέχεια με το πρωτεύον αντίσωμα (σε παρεμποδιστικό διάλυμα). Η αραίωση και ο χρόνος επώασης των αντισωμάτων που χρησιμοποιήθηκαν παρατίθενται στον Πιν.6. Μετά την απομάκρυνση της περίσσειας του αντισώματος με πλύσιμο με παρεμποδιστικό διάλυμα, πραγματοποιήθηκε η επώαση με το δευτερεύον φθορίζον αντίσωμα, στο ίδιο διάλυμα, στο σκοτάδι. Τα κύτταρα ξεπλύθηκαν με PBS και κατά περίπτωση, το DNA τους βαφόταν με διάλυμα DAPI (σε 1xPBS) σε τελική συγκέντρωση 5μg/ml για 10min. Μετά από πλύσιμο με PBS η καλυπτρίδα μεταφέρθηκε, με την πλευρά που ήταν προσκολλημένα τα κύτταρα, σε αντικειμενοφόρο στην οποία είχε προστεθεί μία σταγόνα διαλύματος 1% n-propyl-gallate και 80% γλυκερόλης (σε 1xPBS). Τα δείγματα στεγανοποιούνταν και φυλάσσονταν στους -20°C. Η παρατήρηση τους έγινε σε μικροσκόπιο φθορισμού ενώ στην περίπτωση που ήταν επιθυμητή η λήψη οπτικών τομών, σε συνεστιακό μικροσκόπιο.

2.32. Παραγωγή και απομόνωση πρωτεϊνών

➤ Παραγωγή σε βακτήρια και απομόνωση με χρωματογραφία συγγένειας

Για την παραγωγή ανασυνδυασμένων πρωτεϊνών σε βακτήρια χρησιμοποιείται το στέλεχος *E.coli* BL21 (DE3) CodonPlus. Δεκτικά κύτταρα μετασχηματίζονται με πλασμιδιακές κατασκευές σε φορέα pGEX ή pET26b+ και μετά την καλλιέργειά τους σε στερεό θρεπτικό υλικό, απομονώνεται μοναδιαία αποικία με

την οποία εμβολιάζεται υγρό θρεπτικό μέσο που επωάζεται ολονύκτια στους 37°C υπό ανάδευση. Από την κορεσμένη καλλιέργεια, εμβολιάζεται φρέσκο θρεπτικό μέσο (αραίωση 1/100) το οποίο και επωάζεται μέχρι την εκθετική φάση ανάπτυξης ($O.D_{600nm} \sim 0,6$). Στο στάδιο αυτό απομακρύνονται 2ml από την καλλιέργεια, τα κύτταρα συλλέγονται με φυγοκέντρηση για να αποτελέσουν τον αρνητικό έλεγχο του πειράματος. Ακολουθεί προσθήκη IPTG σε συγκέντρωση 0,1-1mM και επώαση για 2-30h σε θερμοκρασία 20-37°C. Τα κύτταρα συλλέγονται με φυγοκέντρηση στις 3000rpm, ένα κλάσμα τους ελέγχεται για την έκφραση της πρωτεΐνης και η υπόλοιπη βακτηριακή πελέττα είτε αποθηκεύεται στους -20 °C είτε χρησιμοποιείται άμεσα για την απομόνωση της πρωτεΐνης. Συνήθως ακολουθείται η διαδικασία της άμεσης απομόνωσης.

Χρωματογραφία συγγένειας με υπόστρωμα σεφαρόζη -γλουταθειόνη

Η μέθοδος αυτή εφαρμόστηκε για την απομόνωση των ανασυνδυασμένων πρωτεϊνών που είχαν παραχθεί σε σύντηξη με την GST. Εφόσον η πρωτεΐνη εντοπίζεται στο διαλυτό κλάσμα, η βακτηριακή πελέτα επαναιωρείται σε διάλυμα δέσμησης (50mM Tris pH 8, 100mM NaCl, 1mM EDTA, 0,1% NP-40, 50ml/lit καλλιέργειας) και επωάζεται με λυσοζύμη σε τελική συγκέντρωση 300μg/ml σε πάγο για 30min. Ακολουθεί επεξεργασία με υπέρηχους (40% cycle actif, 5 microtip limit) για 2-3min (κάθε 20sec το διάλυμα τοποθετείται σε πάγο για 1min), φυγοκέντρηση στις 12000rpm στους 4°C για 20min, συλλογή του υπερκείμενου, επαναληπτική φυγοκέντρηση και τελική συλλογή του δείγματος. Εάν η πρωτεΐνη εντοπίζεται στο αδιάλυτο κλάσμα, μετά την επώαση με λυσοζύμη ακολουθεί προσθήκη DTT 1M σε τελική συγκέντρωση 5mM και επώαση για 5min σε πάγο. Ακολούθως προστίθεται sarcosyl σε τελική συγκέντρωση 1,5% (από διάλυμα 10%) παρουσία του οποίου το δείγμα επωάζεται για 30min σε πάγο. Στη συνέχεια πραγματοποιείται επεξεργασία με υπέρηχους και φυγοκέντρηση όπως περιγράφηκε παραπάνω. Το πρωτεϊνικό εκχύλισμα προστίθεται σε σφαιρίδια Sepharose 4B με γλουταθειόνη (Pharmacia) που έχουν εξισορροπηθεί σε διάλυμα δέσμησης (1ml σφαιριδίων/5mg πρωτεΐνης) και επωάζονται ολονύκτια στους 4°C με περιστροφική ανάδευση. Στην περίπτωση πρωτεϊνών που απομονώθηκαν από το αδιάλυτο κλάσμα, πριν την χρωματογραφία, προστίθεται Triton σε τελική συγκέντρωση 2% (από διάλυμα 10%), (Frangioni JV & Neel BG, 1993). Ακολουθούν διαδοχικές εκπλύσεις της στήλης με διάλυμα δέσμησης κατά τις οποίες τα σφαιρίδια συλλέγονται με φυγοκέντρηση στις 1000rpm για 1min στους 4°C και επαναιωρούνται σε πολλαπλάσιο όγκο διαλύματος.

Χρωματογραφία συγγένειας με υπόστρωμα σεφαρόζη-Ni-NTA

Η χρωματογραφία αυτή εφαρμόστηκε για την απομόνωση της ανασυνδυασμένης His-FRA10AC1₁₋₁₄₆ (βλ. πλασμιδιακές κατασκευές) από το βακτηριακό εκχύλισμα ώστε να χρησιμοποιηθεί στον καθαρισμό του πολυκλωνικού αντισώματος. Τα βακτήρια συλλέχθηκαν με φυγοκέντρηση και στην πελλέτα προστέθηκε διάλυμα U (8M ουρία, 10mM Tris.Cl pH 8, 100mM Na_xPO₄ pH 7, 10mM ιμιδαζόλιο, 10mM μερκαπτοαιθανόλη) όγκου ίσου με το 1/10 της αρχικής καλλιέργειας. Η επιλογή των αποδιατακτικών συνθηκών στηρίχτηκε στον εντοπισμό της πρωτεΐνης στο αδιάλυτο κλάσμα. Ακολούθησε επεξεργασία με υπέρηχους και φυγοκέντρηση στις 8000rpm για 20min στους 4°C. Το υπερκείμενο, μετά την προσθήκη NP-40 σε τελική συγκέντρωση 1% προστέθηκε σε σφαιρίδια Ni-NTA-σεφαρόζης CL-6B (Qiagen, 1ml σφαιριδίων/5mg πρωτεΐνης), τα οποία είχαν εξισορροπηθεί με διάλυμα U. Η επώαση έγινε σε RT για 3h με περιστροφική ανάδευση και ακολούθησαν διαδοχικές εκπλύσεις της στήλης με το ίδιο διάλυμα.

➤ Παραγωγή πρωτεΐνης σε κύτταρα θηλαστικών

Για την παραγωγή πρωτεϊνών σε κύτταρα θηλαστικών πραγματοποιήθηκε παροδική διαμόλυνση κυττάρων COS-7. Μετά την διαμόλυνση, την επώαση και το πλύσιμό τους τα κύτταρα αποκολλούνται μηχανικά με πλαστική ξύστρα και συλλέγονται με φυγοκέντρηση για 5min στις 1000rpm σε RT. Η πελλέτα επαναδιαλύεται ήπια σε διάλυμα RIPA (150mM NaCl, 50mM Tris-Cl pH 8, 1% NP-40, 0,5% DOC, 0,1% SDS) και ακολουθεί επώαση για 15min στους 4°C σε τροχιακό αναδευτήρα. Στη συνέχεια το δείγμα φυγοκεντρείται για 30min στις 12000rpm στους 4°C και συλλέγεται το υπερκείμενο. Για ένα πείραμα ανοσοκατακρήμνισης ή GST-pull down χρησιμοποιήθηκε το πρωτεϊνικό εκχύλισμα από ένα τρυβλίο διαμέτρου 10cm στο οποίο και προστίθενται 300-400μl διαλύματος RIPA. Η συγκέντρωση του εκχυλίσματος σε πρωτεΐνες είναι 2-3mg/ml.

2.33. Συζευγμένη μεταγραφή-μετάφραση *in vitro*

Η παραγωγή πρωτεϊνών *in vitro* πραγματοποιήθηκε με συζευγμένη μεταγραφή/μετάφραση (T_NT Quick Coupled Transcription/Translation System, Promega). Σε κάθε αντίδραση συμμετείχαν 0,5μg υπερελικωμένου πλασμιδιακού DNA, 20μl του μίγματος που περιέχει το εκχύλισμα των δικτυοερυθροκυττάρων, την RNA πολυμεράση, τα νουκλεοτίδια, τα αμινοξέα και την RNasin και 1μl ³⁵S-Met (1,000Ci/mmol, 10mCi/ml) σε τελικό όγκο 25μl. Η αντίδραση επωάζεται για 90min στους 30°C και ελέγχεται με SDS-PAGE. Το πήκτωμα μονιμοποιείται με επώαση σε διάλυμα αποχρωματισμού για 30min και ακολουθεί επεξεργασία με 1M σαλικυλικό

νάτριο για 1h που δρα ως ενισχυτικό του σήματος. Οι επώσεις πραγματοποιούνται με ήπια ανάδευση σε θερμοκρασία δωματίου και ακολουθεί ξήρανση του πηκτώματος και φθοριογραφία. Με αυτήν την μέθοδο παρήχθησαν οι πρωτεΐνες SAP145 και DGCR14 ως ραδιοσημασμένα μόρια και χρησιμοποιήθηκαν σε δοκιμή συνκατακρήμνισης *in vitro*.

2.34. Δοκιμή συνκατακρήμνισης πρωτεϊνών *in vitro* (GST pull-down)

Η δοκιμή συνκατακρήμνισης είναι μία *in vitro* μέθοδος για τον έλεγχο της άμεσης αλληλεπίδρασης δύο πρωτεϊνών. Η μία πρωτεΐνη παράγεται σε βακτήρια σε σύντηξη με την GST ενώ η δεύτερη είτε παράγεται με μεταγραφή-μετάφραση *in vitro* ως ραδιοσημασμένο μόριο είτε προέρχεται από πρωτεϊνικό εκχύλισμα κυττάρων θηλαστικών. Η χιμαιρική πρωτεΐνη, μέσω της GST συνδέεται με γλουταθειόνη που είναι συζευγμένη με σφαιρίδια σεφαρόζης, αναμιγνύεται με την άλλη ραδιοσημασμένη πρωτεΐνη ή με το πρωτεϊνικό εκχύλισμα σε κατάλληλο ρυθμιστικό διάλυμα και απομακρύνεται επιλεκτικά ώστε να ελεγχθεί αν συν-παρασύρει και την δεύτερη πρωτεΐνη.

Η μέθοδος αυτή χρησιμοποιήθηκε για να ελεγχθεί η άμεση αλληλεπίδραση μεταξύ της FRA10AC1, η οποία παρήχθη σε βακτήρια σε σύντηξη με την GST και των SAP145 και DGCR14 που προέκυψαν με μεταγραφή-μετάφραση *in vitro* ως ραδιοσημασμένα μόρια. Κατά την δοκιμή χρησιμοποιήθηκαν 2-5μg της χιμαιρικής GST-πρωτεΐνης και 10μl της αντίδρασης μεταγραφής/μετάφρασης σε 200μl διαλύματος δέσμευσης (50mM Tris pH 8, 100mM NaCl, 1mM EDTA, 0,1% NP-40). Η επώαση πραγματοποιήθηκε ολονύκτια στους 4°C, υπό ανάδευση. Ακολούθησαν εκπλύσεις με το ίδιο διάλυμα (3 επαναλήψεις με 5 όγκους διαλύματος κάθε φορά) και έλεγχος των πρωτεϊνών που είναι συνδεδεμένες με τα σφαιρίδια. Η διαδικασία αυτή περιλαμβάνει την SDS-PAGE και την επεξεργασία του πηκτώματος όπως και στην περίπτωση ελέγχου των ραδιοσημασμένων μορίων των αντιδράσεων μεταγραφής/μετάφρασης. Επιπλέον, η δοκιμή συνκατακρήμνισης εφαρμόστηκε στην χαρτογράφηση των περιοχών αλληλεπίδρασης των παραπάνω μορίων. Στην περίπτωση αυτή η πλήρους μήκους FRA10AC1 ή τμήματά της παρήχθησαν σε σύντηξη με την EGFP σε κύτταρα θηλαστικών ενώ η παραγωγή τμημάτων των SAP145 και DGCR14, σε σύντηξη με την GST, πραγματοποιήθηκε σε βακτήρια. Η δοκιμή πραγματοποιήθηκε σε διάλυμα RIPA (150mM NaCl, 50mM Tris-Cl pH 8, 1% NP-40, 0,5% DOC, 0,1% SDS) στο οποίο είχαν εξισορροπηθεί και τα σφαιρίδια γλουταθειόνης-σεφαρόζης στις συνθήκες που προαναφέρθηκαν. Η ανίχνευση των πρωτεϊνών έγινε με ανάλυση Western με το αντίσωμα α-GFP.

2.35. Συν-ανοσοκατακρήμιση πρωτεϊνών

Η μέθοδος αυτή εφαρμόζεται για την ανίχνευση πρωτεϊνικών συμπλόκων που υπάρχουν *in vivo* σε κύτταρα θηλαστικών. Σε κάθε δοκιμή χρησιμοποιήθηκε ~1mg πρωτεϊνικού εκχυλίσματος κυττάρων (βλ. παραπάνω). Σε αυτό προστίθεται ~1μg αντισώματος, ειδικού για την πρωτεΐνη προς κατακρήμιση και ακολουθεί επώαση για 2-3h στους 4°C σε τροχιακό αναδευτήρα. Στη συνέχεια προστίθονται 8-15μl πρωτεϊνών A και G που είναι συζευγμένες με σφαιρίδια αγαρόζης (protein A/G Plus Agarose, Santa Cruz Biotechnology, Cat.no. sc-2003), οι οποίες αναγνωρίζουν την περιοχή Fc των αντισωμάτων από μία μεγάλη ποικιλία ειδών και ακολουθεί ολονύκτια επώαση στις παραπάνω συνθήκες. Τα πρωτεϊνικά σύμπλοκα-σφαιρίδια αγαρόζης, απομονώνονται με φυγοκέντρηση για 20sec στις 3000rpm στους 4°C και ξεπλένονται τρεις φορές με 1ml διαλύματος RIPA στις ίδιες συνθήκες. Στην τελευταία φυγοκέντρηση το διάλυμα απομακρύνεται πλήρως. Ακολουθεί επαναιώρηση σε ~20μl διαλύματος, ηλεκτροφόρηση και ανάλυση Western.

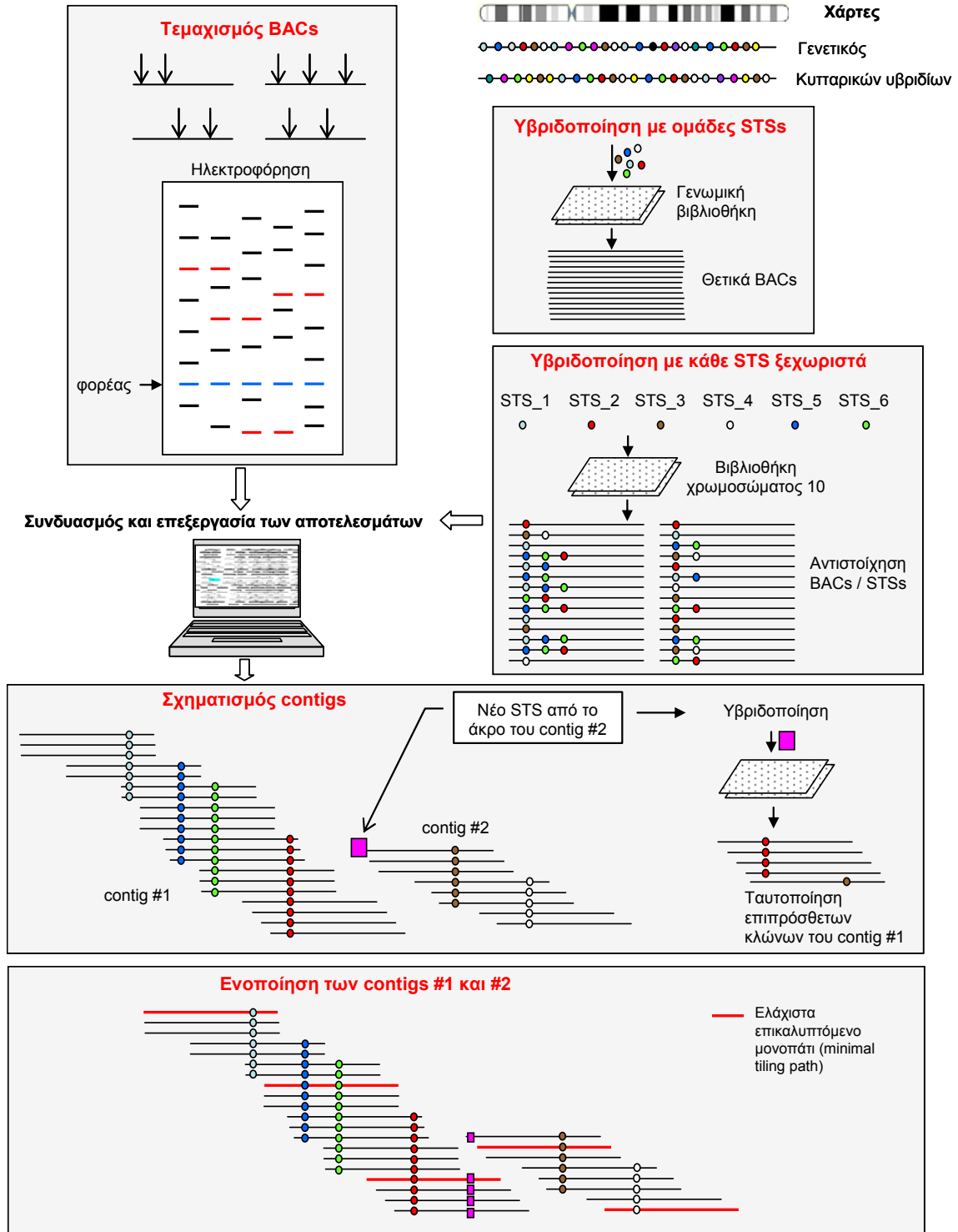
3

ΑΠΟΤΕΛΕΣΜΑΤΑ

3.1. Κατασκευή του φυσικού χάρτη των χρωμοσωμικών περιοχών 10q23.31-q25.3 και 10q26.2-qter

Στα πλαίσια του Προγράμματος Χαρτογράφησης του Ανθρώπινου γονιδιώματος, κατασκευάστηκε ο φυσικός χάρτης των περιοχών 10q23.31-q25.3 και 10q26.2-qter. Οι περιοχές αυτές καταλαμβάνουν γενετική απόσταση ~30cM και ~15cM αντίστοιχα και σε μονάδες γενετικού χάρτη, αποτελούν αθροιστικά το 1/4 του χρωμοσώματος. Η αντίστοιχη φυσική απόσταση για την πρώτη περιοχή εκτιμάται στις 25Mb, σύμφωνα με την αναλογία 1,2cM/1Mb για όλο το ανθρώπινο γονιδίωμα. Το μήκος της υποτελομερικής περιοχής εκτιμήθηκε ότι είναι αρκετά μικρότερο από όσο υποδηλώνει ο γενετικός χάρτης καθώς το ποσοστό ανασυνδυασμού στις περιοχές πλησίον των τελομερών είναι υψηλότερο σε σχέση με το υπόλοιπο χρωμόσωμα, όπως είχε αξιολογηθεί πριν την αλληλούχηση του ανθρώπινου γονιδιώματος και επιβεβαιώθηκε κατόπιν (Payseur BA & Nachman MW, 2000; Yu A *et al*, 2001; Kong A *et al*, 2002; Jensen-Seaman MI *et al*, 2004).

Συνοπτικά, η διαδικασία (Εικ. 13) περιλάμβανε την ταυτοποίηση κλώνων BAC του χρωμοσώματος χρησιμοποιώντας όλα τα υπάρχοντα STSs (~500) από διάφορους χάρτες, την παρασκευή χρωμοσωμοειδικής γενωμικής βιβλιοθήκης, τον ακόλουθο προσδιορισμό του περιεχομένου κάθε κλώνου σε STSs και τον συνδυασμό αυτών των πληροφοριών με τα αποτελέσματα του προτύπου τεμαχισμού των κλώνων. Η αλληλοεπικάλυψη κατά μήκος πολυάριθμων κλώνων χρησιμοποιήθηκε για να διαταχθούν οι κλώνοι σε contigs τα οποία χαρτογραφήθηκαν βάσει του περιεχομένου τους σε STSs. Επιπλέον, οι πληροφορίες που προέκυψαν από τα STSs αξιοποιήθηκαν για τον έλεγχο των αποτελεσμάτων των αυτοματοποιημένων στοιχίσεων των κλώνων ώστε να αποφευχθούν ψευδείς αλληλοεπικαλύψεις ή για να ενοποιηθούν γειτονικά contigs. Στη συνέχεια επιλέγονταν αντιπροσωπευτικοί κλώνοι, οι οποίοι συνιστούσαν το ελάχιστο επικαλυπτόμενο μονοπάτι (minimal tiling path), για τον προσδιορισμό της αλληλουχίας τους. Η περαιτέρω επέκταση και ενοποίηση των επιμέρους contigs πραγματοποιήθηκε με χρωμοσωμικό περπάτημα, κατά το οποίο χρησιμοποιήθηκαν οι ακραίες αλληλουχίες των εξωτερικών BACs, καθώς και με ταυτοποίηση αλληλοεπικάλυψης, σε επίπεδο αλληλουχίας των ακραίων κλώνων γειτονικών contigs.



Εικ. 13. Πορεία κατασκευής του φυσικού χάρτη.

➤ **Διερεύνηση γενωμικής βιβλιοθήκης BAC**

α. Επιλογή ανιχνευτών

Για την διερεύνηση της βιβλιοθήκης BAC χρησιμοποιήθηκαν όλα τα STSs που χαρτογραφούνται στις δύο περιοχές. Η επιλογή έγινε από τους γενετικούς χάρτες Genethon (Dib C *et al*, 1996) και Marshfield (Broman KW *et al*, 1998) και από τους χάρτες κυτταρικών υβριδίων GeneMap99 (Deloukas P *et al*, 1998), WI-RH (Hudson TJ *et al*, 1995) και SHGC-G3 (Stewart EA *et al*, 1997). Συνολικά, επιλέχθηκαν 506 STSs: 430 από την περιοχή 10q23.31-q25.3 και 76 από την υποτελομερική περιοχή (Πιν. 7). Η πυκνότητα των δεικτών είναι ιδιαίτερα υψηλή και ανέρχεται σε 14 δείκτες/cM και 5 δείκτες/cM αντίστοιχα, με την τελευταία αναλογία να είναι πιθανότατα υποτιμημένη λόγω του μικρότερου φυσικού μήκους της υποτελομερικής περιοχής.

Πιν. 7: Προέλευση των STSs που χρησιμοποιήθηκαν στην φυσική χαρτογράφιση

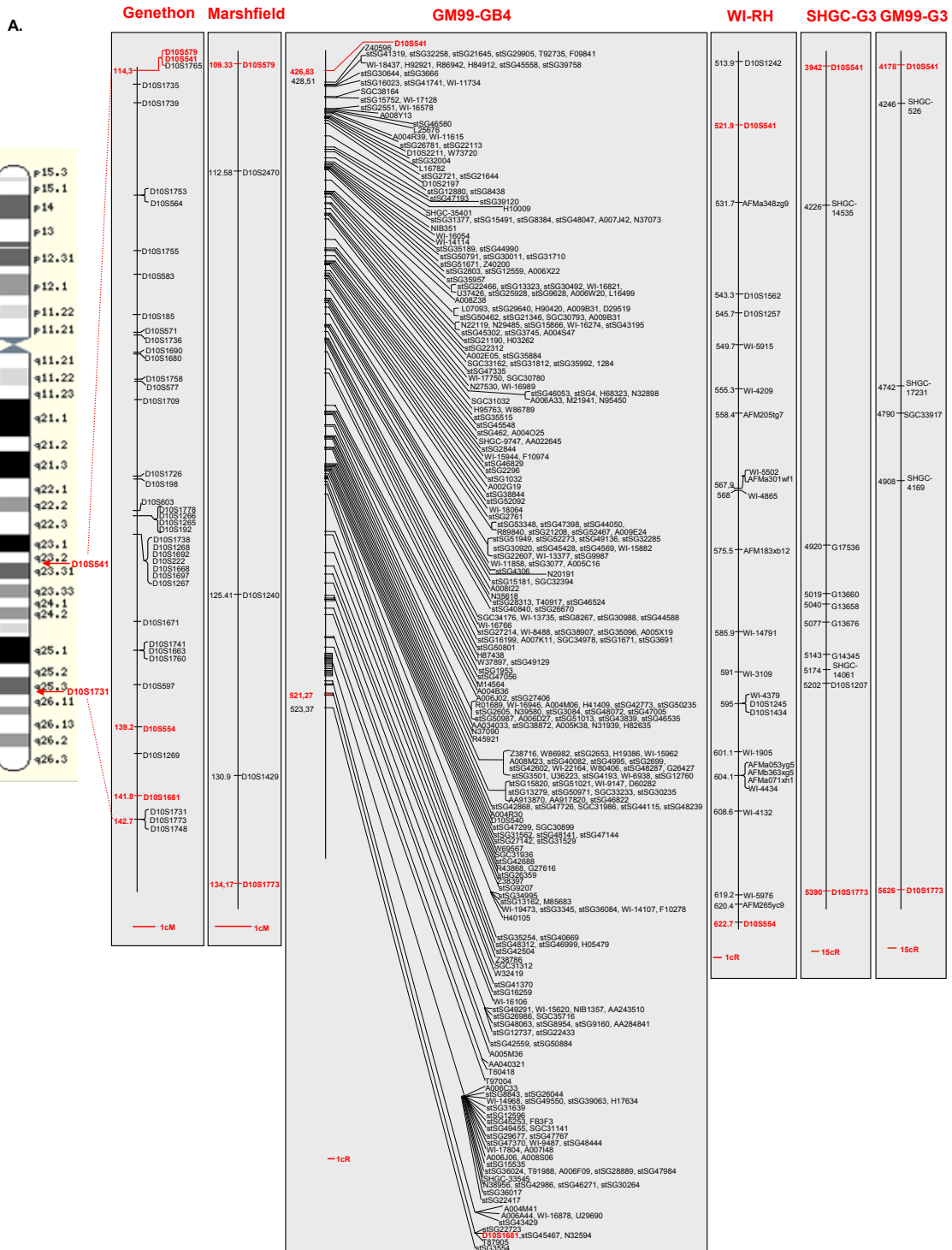
	Genethon	Marshfield	GeneMap99	WI-RH	SHGC-G3	ΣΥΝΟΛΟ
			Genebridge4	Genebridge3		
10q23.31-q25.3	42	3	349	4	24	430
10q26.2-qter	13	6	48		8	76

Σημ.: Αναγράφεται ο αριθμός των STSs που χρησιμοποιήθηκαν από κάθε χάρτη (οριζόντια) για κάθε χρωμοσωμική περιοχή ξεχωριστά (κάθετα). Ο μεταγραφικός χάρτης Genemap '99 περιλαμβάνει δύο panels: Το Genebridge4 και το Genebridge3.

Κάποια από τα ESTs που χρησιμοποιήθηκαν προέρχονταν από την ίδια ομάδα Unigene. Αυτό συνέβη γιατί, καθώς αυτή η βάση δεδομένων εξελίσσεται συνεχώς, κάποιες ομάδες της ενοποιοούνται. Ωστόσο αυτό δεν θεωρείται πλεονασμός καθώς σε αρκετές περιπτώσεις ένα γονίδιο εκτείνεται σε περισσότερα του ενός BACs. Στην Εικ. 14 παριστάνονται τα τμήματα των παραπάνω χαρτών που αντιστοιχούν στις συγκεκριμένες χρωμοσωμικές περιοχές και η χαρτογράφιση των STSs (για τις ακριβείς συντεταγμένες τους βλ. Πιν. 4).

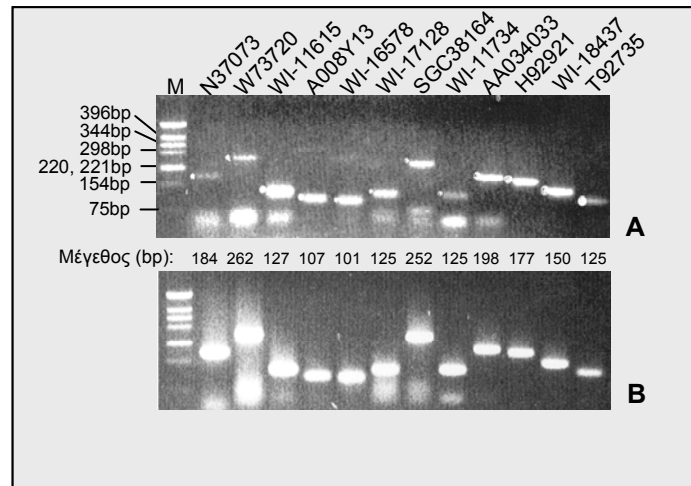
β. Παρασκευή ανιχνευτών

Για τα 506 επιλεγμένα STSs παρασκευάστηκαν οι κατάλληλοι εκκινητές (οι αλληλουχίες τους παρατίθενται στον Πιν. 4) και πραγματοποιήθηκε PCR με υπόστρωμα γενωμικό DNA ή DNA κυττάρων σωματικών υβριδίων ανθρώπου – κρινητού που περιείχε μόνο το μεγάλο βραχίονα του χρωμοσώματος 10 (βλ. Μεθόδους). Οι θερμοκρασίες υβριδοποίησης των εκκινητών κυμαίνονταν μεταξύ 54°-62°C. Ενδεικτικά, στην Εικ. 15Α φαίνονται τα προϊόντα της PCR για 12 STSs.



Εικ. 14: Χάρτες του χρωσώματος 10 από τους οποίους επιλέχθηκαν τα STSs για την διερεύνηση της βιβλιοθήκης BAC. Απεικονίζονται αποκλειστικά οι χρωμοσωμικές περιοχές 10q23.31-q25.3 μεταξύ των γενετικών δεικτών D10S541-D10S1731 (Εικ. 14A) και 10q26.2-qter τελομερικά του δείκτη D10S1703 (Εικ. 14B). Από αριστερά προς τα δεξιά: ιδεόγραμμα του χρωσώματος, οι γενετικοί χάρτες Genethon και Marshfield και οι χάρτες κυτταρικών υβριδίων GM99-GB4, WI-RH, SHGC-G3 και GM99-G3 (μόνο στην 14A). Δεν έχουν συμπεριληφθεί οι κοινοί δείκτες, εκτός αυτών που καθορίζουν τα όρια των παραπάνω περιοχών (με κόκκινο). Οι ακριβείς συντεταγμένες των STSs για την πρώτη περιοχή παρατίθενται στον Πιν. 4 (Μέθοδοι). Επίσης, φαίνεται η κλίμακα που χρησιμοποιήθηκε, σε cM για τους γενετικούς και σε cR για τους υπόλοιπους χάρτες.

συμμετείχαν ~10 STSs ενώ τουλάχιστον για έναν από τους ανιχνευτές προσδιοριζόταν η ειδική ενεργότητα (βλ. Μεθόδους). Στις περιπτώσεις που η επαναληπτική PCR παρουσίαζε δυσκολίες, το αντίστοιχο ραδιοσημασμένο προϊόν ελέγχονταν με ηλεκτροφόρηση και ακόλουθη αυτοραδιογραφία. Οι ανιχνευτές χρησιμοποιούνταν άμεσα για υβριδοποίηση εφόσον η ειδική ενεργότητα είχε εκτιμηθεί ότι υπερβαίνει τα 10^9 cpm/μg.



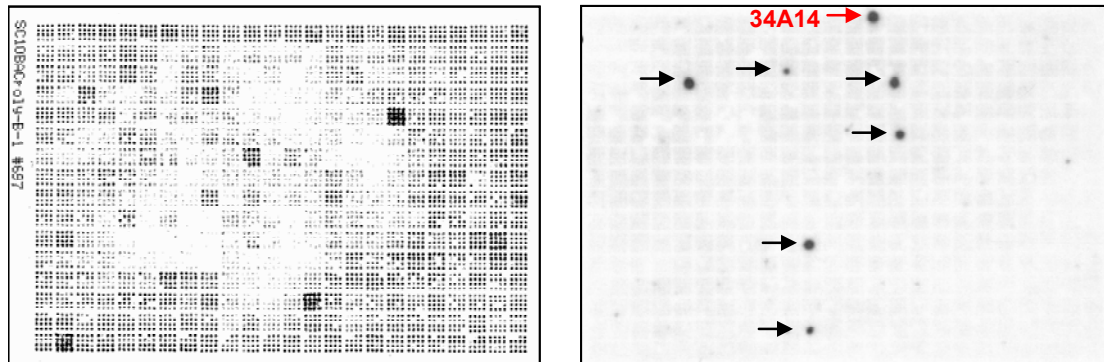
Εικ. 15: Προϊόντα PCR για 12 STSs της χρωμοσωμικής περιοχής 10q23.31-q25.3.
A. Ηλεκτροφόρηση προϊόντων PCR για 12 STSs. Παρατίθενται οι κωδικοί τους (πάνω) και τα μεγέθη των προϊόντων (στη μέση). Επίσης, σημειώνονται οι ζώνες που απομονώθηκαν από το πήκτωμα (με τελείες). Η θερμοκρασία υβριδοποίησης των εκκινήτων ήταν 57°C .
B. Επαναληπτική PCR με υπόστρωμα 3μl της υδατικής φάσης στην οποία εκλούστηκαν τα προϊόντα της Εικ. A. Η σειρά των δειγμάτων και οι συνθήκες της αντίδρασης είναι οι ίδιες με την Εικ. A. Ο δείκτης (M) που χρησιμοποιήθηκε είναι ο pBR322/Hinfl (A και B) τα μεγέθη του οποίου αναγράφονται (αριστερά).

γ. Υβριδοποίηση και επιλογή κλώνων BAC

Η διερεύνηση της βιβλιοθήκης γινόταν σε ομάδες των 3 έως 22 STSs (pools) κάθε φορά χρησιμοποιώντας $2 - 4,5 \times 10^5$ cpm ανιχνευτή ανά ml διαλύματος υβριδοποίησης. Τα 36 φίλτρα της βιβλιοθήκης τοποθετούνταν σε τρεις κυλίνδρους υβριδοποίησης με 15ml διαλύματος στον καθένα. Μετά την αυτοραδιογραφία ακολουθούσε η ταυτοποίηση των κλώνων όπως παρουσιάζεται στην Εικ. 16. Συνολικά πραγματοποιήθηκαν 50 υβριδοποιήσεις της βιβλιοθήκης και τα αποτελέσματα που προέκυψαν παρατίθενται στον Πιν. 8.

Κατά μέσο όρο κάθε STS αντιστοιχούσε σε 11 κλώνους, αναλογία που αντικατοπτρίζει το βάθος της κάλυψης αυτής της γονιδιωματικής περιοχής. Τα μοναδιαία BACs που ταυτοποιήθηκαν ήταν ~2700 για την περιοχή 10q23.31-q25.3 και 650 για την υποτελομερική περιοχή. Λαμβάνοντας υπ' όψιν το μέσο μέγεθος των ενθεμάτων των BACs (~175kb), οι κλώνοι αυτοί παρέχουν μέση υπερκάλυψη κατά ~16 φορές και τουλάχιστον κατά 7,5 φορές αντίστοιχα.

Όλα τα αποτελέσματα έχουν κατατεθεί στην χρωμοσωμειδική βάση δεδομένων ACeDB, την 10ace με την ονομασία pools 1L-50L. Ενδεικτικά, στην Εικ. 17 φαίνεται η καταχώρηση των αποτελεσμάτων της διερεύνησης της βιβλιοθήκης με τα STSs του pool 30.



Εικ. 16: Διάταξη και αυτοραδιογραφία της γενωμικής βιβλιοθήκης BAC.

A. Απεικόνιση ενός εκ των 36 φίλτρων της βιβλιοθήκης. Κάθε φίλτρο περιλαμβάνει 6144 κλώνους που είναι διατεταγμένοι ανά 16 (στη μορφή 4x4) σε 16 σειρές (A, B, C...P) και 24 στήλες (1-24). Το όνομα κάθε κλώνου αποτελείται από έναν αριθμό που δηλώνει το φίλτρο στο οποίο ανήκει και την θέση του στη διάταξη 4x4, ακολουθούμενο από ένα γράμμα που δείχνει την σειρά και έναν τελικό αριθμό που υποδηλώνει την στήλη.

B. Αυτοραδιογραφία στην οποία ταυτοποιούνται 7 θετικοί κλώνοι (βέλη). Ενδεικτικά, το όνομα του κλώνου που σημειώνεται με κόκκινο βέλος είναι 34A14. Το πρόθεμα 34 δείχνει ότι το συγκεκριμένο φίλτρο είναι το τρίτο της βιβλιοθήκης και ότι πρόκειται για τον δεύτερο κλώνο στην διάταξη 4x4. Η διάταξη στο πρώτο φίλτρο περιλαμβάνει τους αριθμούς 1-16, στο επόμενο 17-32 κ.λπ.

➤ **Κατασκευή και διερεύνηση γενωμικής βιβλιοθήκης του χρωμοσώματος 10**

Οι παραπάνω 3350 διαφορετικοί κλώνοι που ταυτοποιήθηκαν, σε συνδυασμό με τα BACs που με την ίδια προσέγγιση χαρακτηρίστηκαν ως γενωμικοί κλώνοι του χρωμοσώματος 10 (από την ομάδα του Π. Δελούκα στο Sanger Center) αποτέλεσαν μία γενωμική, χρωμοσωμειδική υποβιβλιοθήκη (βλ. Μεθόδους) που διατάχτηκε σε 3 φίλτρα υψηλής πυκνότητας (polygrids).

Πιν. 8: Αποτελέσματα της διερεύνησης της γενωμικής βιβλιοθήκης BAC

POOL	STSs	BACs
1	sISG9207	2F13, 16H23, 31L23, 26I17, 31C2, 37L21, 34D21, 38F19, 47K14, 47A8, 55P24, 55H24, 63G20, 73I21, 68M5, 85J3, 87A18
	WI-9147	92K19, 91E4, 91A6, 105N15, 124A14, 121B1, 124G18, 121A1, 121B2, 114F1, 135I2, 131P16, 131N13, 135G2, 135G3,
	Z38716	147B13, 148E14, 155M3, 161J3, 166D12, 177N16, 190J1, 208K8, 205K9, 217B13, 215H21, 216K21, 239A7, 225G22,
	H05479	229H16, 225H22, 277F21, 294B4, 320K6, 308M12, 310C11, 326I7, 326J7, 325G20, 345B6, 344B9, 338F22, 364M5,
	WI-16946	381E23, 372K16, 372L16, 381N11, 377N7, 388D6, 390A9, 390A22, 400D18, 399L6, 399P6, 397F19, 403K22, 416N2,
	SGC31986	427G23, 426P21, 418G13, 418E13, 440H6, 462G17, 450K22, 449L8, 474E14, 465A7, 491G16, 493O5, 499D21, 517G4,
	SGC30899	514D16, 513A17, 524G11, 522K10, 539A22, 529I10, 544J23, 537L19, 530L13, 560B7, 550H22, 561B19
	G27616	
	W86982	
	2	Z38397
SGC33233		110M24, 135I2, 135G2, 135G3, 133L23, 138J24, 138J21, 155M3, 154K6, 154E18, 170J3, 173J5, 172G11, 166D12,
SGC31936		167C14, 203B23, 214C13, 223B13, 213C11, 230H19, 229N4, 228C22, 256H10, 252J11, 256P15, 243L20, 263M9,
A006J02		268F5, 271I13, 260L13, 277F21, 285D15, 273N6, 303C5, 299I5, 301L10, 298P19, 302K17, 301F16, 301H14, 299C16,
M85683		317J21, 346A7, 345B6, 353B24, 353I17, 363D8, 360E9, 367M7, 382M19, 381N11, 393G6, 403K22, 412G16, 416K23,
F10278		426A8, 422K11, 418P17, 446H13, 442J21, 446K23, 446K18, 451M19, 474A18, 472D22, 468M23, 471P23, 496G1,
H05479		482E11, 489E24, 503I5, 509I15, 497N19, 511O21, 520F2, 523M11, 524L15, 523F17, 513A17, 541N10, 555A18,
Z38786		550H22, 548N18, 549M18, 556D1, 555K2, 559P8, 571M9
W32419		
3		H87438
	SGC33917	116D23, 116E5, 117P24, 121I4, 121J23, 133I19, 142H5, 157O7 159K16, 160I11, 165O6, 175B20, 188N23, 190O12,
	SGC34176	190O13, 190N2 227A18, 260L2 274I6, 283C24, 285F16, 287B3, 296C3, 316M21, 325E17, 328F4, 330J1, 338O11,
	T40917	344E13, 345F3, 352M13, 361J5, 363N16, 383K22, 395E12, 399H7, 407A10, 411B6, 417D19, 417F17, 420P24, 428K4,
	WI-8488	433A14, 436A6, 436E6, 436A7, 438A21, 443A5, 455H10, 460C19, 469H13, 470C8, 478N5, 483F11, 483B11, 490D4,
	WI-13735	494O2, 504A6, 509A14, 520E2, 539M3, 553H1, 565E10, 565H24, 572B16
	WI-16766	
	sISG1671	
	sISG3691	
	4	A004M06
sISG27214		369M5 377N7 393A6 393G6 416K23 427G23 426A8 419K13 436B8 452K1 453H6 461F9 465H10 471P23,
sISG2653		496G1, 489E24, 501J12, 508J18, 550G18, 573E23
G26427		
H19386		
WI-15962		
D60282		
5	sISG51146	6I23, 6I24, 15I4, 29L1, 37K12, 73I21, 68M5, 91A6, 87A18, 91E4, 95J12, 85J7, 81F18, 82F10, 105N15, 131P16,
	R43868	148E14, 177N16, 190J1, 239A7, 270D20, 279B8, 300I11, 293M13, 312F17, 325G20, 349G14, 356H19, 369M5,
	WI-14107	394B24, 390C1, 401G3, 413E5, 413I9, 430D8, 438O13, 443N17, 438K21, 460P6, 503H15, 503B2, 502B6, 509B9,
	H41409	529I10, 530L13, 544J23, 537L19, 539A22
	A004B36	
	U36223	
6	T34337=A004R30	37L21, 37D18, 34A22, 62C24, 52H22, 88C2, 85A6, 91K19, 107L7, 155M3, 155N10, 155M10, 203B23, 286K1, 277F21,
	R01689	286O21, 299C16, 293M13, 242B12, 249E8, 243G8, 321C9, 345B6 360E9 360M7, 370K10, 381N11, 393A6, 396N9,
	A006J02	416K23, 419K13, 436B8, 437A23, 461F9, 453H6, 452K1, 465H10, 471P23, 501J12, 513A17, 550H22, 203B23
	A008M23	
	SGC31936	
	M14564	
7	sISG34995	2A2, 10B18, 16H23, 18I14, 18O12, 30F12, 30H12, 29L1, 38P1, 40M1, 70E24, 91A6, 91E4, 99N20, 99O20, 117P7,
	sISG13162	131N13, 131P16, 162B11, 170J3, 190J1, 207D20, 208K8, 216L6, 229H16, 230H19, 239M11, 250J20, 252J18,
	SGC36084	301F16, 301H14, 301L10, 308D20, 313D6, 332C10, 346A7, 346N21, 348N5, 363L19, 363N23, 366K23, 369M5,
	sISG16259	371A11, 373N18, 403L6 409D24, 426A8, 445I22, 468J18, 468N18, 468B22, 472D22, 474E6, 475M16, 491K17,
	sISG40082	502B19, 501N10, 506E3, 521L8, 524L15, 525P19, 530A9, 530A20, 530G15, 530G2, 550G18, 561B19, 571D14,
	sISG27142	572F16, 573E23, 574J10
	sISG42868	
9	sISG28313	3L6, 16H23, 17P13, 20D13, 31L23, 37L21, 37G12, 42I2, 46I2, 60A9, 69F17, 77F18, 85A1, 92K19, 115D23, 135G2,
	sISG40840	135G3, 135I2, 129J12, 149A13, 155M3, 157O7, 155K9, 159K16, 166D12, 190O12, 190O13, 178O18, 193B23,
	sISG8267	203B23, 235N10, 264B12, 287G8, 277F21, 301O4, 357 (358) I15, 370D14, 381N11, 383K22, 394E12, 407A10, 415K21,
	sISG30988	403K22, 418J9, 436E6, 433A14, 441O15, 454H10, 469H13, 467N8, 478N5, 482F11, 514N1, 513A17, 514G18, 515N17,
	sISG38907	515N18, 533O14, 550H22, 561B19, 565H24, 573L13
	sISG27406	
AA034033		
10	sISG38872	16H23, 21N23, 24K12, 25P11, 27L23, 37L21, 40M1, 55N17, 55D2, 70E24, 75K16, 79C21, 81E5, 86M22, 88C2,
	sISG42773	89G15, 92K19, 99N20, 99O20, 101J13, 107(108)L7, 123L1, 127L20, 129I16, 131N13, 131L23, 134M6, 135G2,
	sISG4995	135G3, 135I2, 142K6, 155M10, 155M3, 155N10, 155E8, 158G3, 162B21, 167D12, 170C16, 186B2, 192G6, 199D9,
	sISG2699	203B23, 205K9, 211D4, 229H16, 242B12, 249E8, 250J20, 251D12, 263I21, 277F21, 287G20, 289J19, 291K18,
	sISG31529	295H12, 299C16, 302K17, 341G3, 345B6, 366K1, 366K23, 373N18, 373D9, 381N11, 382P22, 387O2, 388D12,
	sISG35254	388(392)C16, 395M19, 396N9, 403K22, 409D24, 437A23, 459H15, 471P23, 474E6, 475M16, 498G15, 499K15, 502B19,
	sISG42504	507G20, 513A17, 524G11, 525P19, 530A9, 530A20, 530G15, 550H22, 561B19, 565E8, 569F9, 570J10, 572F16
	sISG26359	
13	Z40596	9C23, 13K2, 21F19, 22A6, 27D5, 28G22, 38I8, 40O16, 47K24, 46M17, 51J20, 54J21, 56N18, 57C13, 62J15, 71E2,
	T92735	71O3, 80H5, 85C14, 112G4, 135H15, 149A14, 159F21, 149I23, 161N7, 168O10, 181I8, 179M9, 204L12, 218E3,
	WI-18437	236H7, 256F2, 249G24, 287B17, 276D24, 300K2, 291L24 338H11, 338M20, 375N21, 400M3, 388M4, 397M7,
	H92921	404P11, 406N15, 409D17, 410C16, 425M17, 425C21, 437M10, 474I15, 498G4, 485B24, 502C7, 518A11, 524M7,
	R86942	522P8, 528M22, 523M4, 576H10, 40K13, 40N12, 62B8, 145B5, 148D8, 149I9, 151J19, 175I5, 186O14, 298J5, 349G6,
	WI-11734	354E22, 428L9, 473F1, 528P2
	WI-17128	
	SGC38164	
	WI-16578	
	A008Y13	

Πιν. 8: (συνέχεια)

POOL	STSs	BACs	
14	sISG2605	181F14, 18O12, 261I7, 31C2, 38P1, 47K14, 68M5, 73I21, 87A18, 105N15, 124A14, 161J3, 177N16, 230H19, 240A7,	
	H40105	301F16, 320K6, 326J7, 346A7, 426A8, 426P21, 465A7, 472D22, 524L15, 522K10, 529I10, 530L13, 537L19, 544J23,	
	N39580	539A22, 148E14, 147B13, 326I7, 301H14, 411D17	
15	H84912	10B9, 10L2, 11K21, 11P23, 27I3, 30G13, 27L16, 35C6, 41C12, 43D10, 39E8, 46M17, 71F8, 70C15, 68J7, 67L13,	
	sISG45558	98P18, 103A2, 105F14, 98K17, 119F19, 130B8, 157H10, 146I16, 149I23, 158J11, 149J19, 168M11, 164M14, 168O10,	
	sISG46580	188C5, 204L12, 201M14, 228A3, 239G4, 231I18, 239G24, 256F2, 253L6, 254N8, 245D15, 249G24, 53N19, 267E5,	
	WI-11615	262D21, 275G24, 287K24, 303A10, 304L3, 293L10, 320F15, 319B14, 309B15, 364F9, 360H20, 362O4, 380G5	
	W73720	377E16, 382P11, 375N21, 388M4, 410L2, 401P6, 402D21, 430L14, 440B12, 437J2, 441G9, 440F21, 448L14, 443N19,	
	sISG47193	459I18, 460M16, 462N15, 460N21, 474H4, 474I15, 478L15, 489D4, 496G4, 496D14, 496H23, 491K20, 481N24, 512J3,	
	N37073	510B24, 505K23, 501M24, 513B1, 517O8, 542J7, 534O5, 532L6, 536G17, 538C20, 539D21, 560G4, 558K13, 567B5,	
	sISG48047	575C13, 568I23, 16A13, 60O1, 73L23, 95N9, 98K15, 98P18, 124D13, 123J3, 136P13, 155M6, 163J10, 180I12, 180I13,	
	sISG44990	192D22, 222G3, 223O16, 229A6, 265L24, 281C3, 286B11, 304E17, 294J20, 315O7, 311O12, 322M19, 346J9, 337B24,	
	sISG50791	363B10, 379K18, 383L20, 399O19, 410H2, 415J12, 414M22, 460N22, 524I13, 535B19, 570H15, 570L24, 571O5	
	16	WI-16054	4L8, 11P23, 25O20, 61N7, 61K23, 65C4, 67L13, 71F8, 81C11, 95J18, 106M14, 124A3, 139A1, 137B2, 137B3,
		WI-14114	159J10, 171H14, 168O13, 181C9, 184C22, 212N8, 239J6, 248I14, 251F13, 252D12, 266P20, 280C8, 316G19,
		A007J42	326I7, 348J12, 353L10, 362O4, 366I13, 358I19, 360N18, 413E18, 416G12, 442J15, 464B1, 474K16, 481O14,
A006X22		503J12, 510B24, 539D21, 543N17, 558K13, 562K6, 17P9, 18F13, 29A1, 38H10, 50E3, 156M2, 164M14, 229P13,	
sISG22466		292K9, 320F15, 397J18, 413M23, 427B15, 439P3, 438D17, 460I14, 481N24, 499K21, 532L6, 536G17, 555D2, 567B5,	
sISG13323		565P20	
sISG30492			
WI-16821			
A004S47			
L16499			
17	sISG25928	16J1, 17P9, 22B1, 27I3, 27L16, 28G21, 44J24, 55M22, 57F1, 59F22, 70C15, 81N7, 88A22, 90J1, 95N9, 98P18,	
	A008Z38	99P7, 102H21, 108M11, 111H11, 119K6, 119H10, 127J3, 136P15, 149I15, 150M17, 159J10, 160G7, 161K8, 166H9,	
	sISG50462	173C8, 179F1, 189N16, 192O5, 208C17, 243O21, 266B14, 288I20, 305E9, 313N18, 316L13, 325D11, 326M8,	
	sISG21346	329G11, 341B18, 349I10, 350H8, 350F20, 353H23, 360G10, 361K9, 369K19, 373I20, 377B24, 404A17, 411A11,	
	SGC30793	416G12, 419M21, 423F23, 428L7, 430L14, 435N8, 437J2, 441G9, 441E11, 459I18, 460C14, 460B23, 461N7,	
	N22119	466J14, 469C1, 474K16, 484B9, 492A19, 493H16, 505K23, 507M2, 509E6, 512J3, 515D2, 521L12, 523H23,	
	N29485	550J1, 550N4, 554E22, 556C2, 556I23, 562K6, 562L21, 568I23	
	sISG15866		
	WI-16274		
	sISG43195		
18	sISG45302	5M14, 6E23, 16J1, 16M4, 17P9, 20M23, 27I3, 27L16, 31J1, 37B7, 44M16, 44D15, 48I5, 55L15, 58O16, 58E21, 61M17,	
	sISG21190	63F3, 64H1, 67K6, 70C15, 76P2, 78B19, 79M5, 80J24, 81N7, 88A22, 90J1, 90O1, 95N9, 98P18, 99P7, 102L6, 109J16,	
	H03262	113F11, 116H10, 119K6, 121A3, 122J21, 122N12, 126H3, 127B10, 127J3, 136P15, 137B2, 137B3, 139A1, 143J17,	
	A002E05	147C3, 149I15, 160G7, 160P20, 162B9, 162K11, 174E12, 175A5, 179B12, 188L18, 189N16, 190H7, 190I7, 191L5,	
	sISG35884	192O5, 198K1, 199A10, 199C11, 200M10, 205N7, 207E20, 208C17, 213B12, 217C3, 230B23, 236E21, 236G12, 237F10,	
	SGC33162	241J14, 247I14, 248J23, 251O1, 251A15, 266L15, 272G5, 276E20, 280E20, 284A18, 287M10, 288I20, 293J1, 300H5,	
	sISG31812	300O15, 301A3, 301B8, 303N10, 305E9, 310E22, 313P2, 315C16, 319D1, 319I6, 325D11, 326M8, 329B8, 341J4,	
	sISG35992	349I10, 350H8, 353H23, 355A11, 355J6, 358I4, 361B6, 361K9, 362N20, 368O6, 372K18, 373I20, 374N23, 375N10,	
	sISG47335	377B24, 380F23, 391H15, 391J2, 393F1, 393F22, 396J14, 397J18, 400G3, 402D21, 404A17, 406M15, 409N24,	
	A004O25	411A11, 413M23, 414D4, 419M21, 420P15, 423F23, 427B15, 427L6, 428I7, 429G19, 429H9, 437J2, 440D6, 441E11,	
	WI-17750	441G9, 447G19, 450K13, 452B15, 460I14, 463J11, 466J14, 466M22, 467D11, 476E15, 484B9, 491P14, 505K23, 506F1,	
	N27530	512J3, 512B11, 513F19, 516J19, 521L12, 523H23, 530A1, 533B2, 534O13, 536H2, 551O13, 556C2, 557F6, 561I23,	
	WI-16989	562E21, 566H2, 568I23, 570M19, 570D9	
	sISG46053		
	sISG4		
	M21941		
	N32898		
	N95450		
	H68323		
	SGC31032		
19	H95763	1D22, 3N15, 7D5, 12J7, 12J23, 26K19, 30K2, 31J1, 31P18, 32P3, 34E5, 41E18, 48D17, 50J11, 50J12, 50P12,	
	W86789	53P16, 61M17, 65P8, 66E6, 81K15, 83C5, 87A19, 93C15, 105H12, 114O20, 116M10, 122H11, 122N12, 123G19,	
	sISG45548	125L1, 136O21, 141A20, 145P10, 147G3, 164I17, 164O1, 165A22, 165M16, 168H20, 168I22, 169C12, 170O19, 171K9,	
	sISG35515	175O19, 178P8, 179I2, 190H7, 191J13, 196N24, 199D17, 201K4, 205N7, 208M2, 220P4, 225L22, 230C8, 246D12,	
	sISG2844	251C15, 254B23, 265D9, 270N12, 271L15, 272M22, 278B4, 280E20, 300H5, 308C11, 309A11, 311D21, 313P2,	
	sISG55875=F10974	314I12, 315C16, 318M12, 319I6, 323I21, 328A2, 331E7, 335G20, 336G18, 342K23, 348C12, 355J6, 360H7, 369E13,	
	WI-15944	369I23, 374N23, 375N9, 376N20, 386M4, 387C15, 396K3, 402P15, 408G17, 408K20, 409N13, 411E24, 414D4, 414B12,	
	sISG46829	418H18, 425I21, 426P1, 427L6, 429D12, 430D21, 434I6, 442I19, 445I23, 450O3, 452K12, 453D9, 453D2, 453E2,	
	sISG2236	456I7, 461N22, 470M7, 471M3, 476E15, 477F21, 481J2, 483E22, 484E21, 488I2, 488J3, 493C7, 493N13, 494G17,	
	sISG38844	495F4, 496G18, 502I24, 504P21, 507K5, 511O2, 514I23, 514M5, 516H12, 528B11, 536A15, 541P18, 548K23,	
	sISG2761	560L16, 561I23, 562L13, 569G1, 570M19	
	sISG53348		
	sISG47398		
	sISG44050		
	sISG51949		
	sISG52273		
	R89840		
	sISG49136		
	sISG21208		
	sISG45428		

Πιν. 8: (συνέχεια)

POOL	STSs	BACs	
20	sISG462	7D5, 16F10, 17M15, 22L14, 22K16, 26K19, 29D10, 31M16, 31P18, 44K20, 47D13, 49O21, 50K24, 53D22, 54G9,	
	WI-18064	59D6, 61M16, 61M17, 62P21, 65P8, 70A7, 72D12, 72N20, 74K9, 83C5, 84H9, 85N18, 85P16, 88L4, 89L2, 91K3,	
	sISG32285	101E20, 102D7, 104K10, 105H12, 106B18, 107J10, 110B6, 111L4, 112K3, 114M15, 122N12, 123G19, 128I23, 135P6,	
	sISG30920	140O6, 140M12, 141A20, 153O14, 153P14, 170O19, 172B21, 173C15, 175O19, 183H21, 183J3, 183J21, 186D15, 198N23,	
	sISG4569	204L10, 207B4, 207C21, 208A8, 208M2, 218F18, 223M20, 226C16, 228L9, 238F24, 241J4, 245I6, 246D12, 248D10,	
	WI-15882	251C15, 251L13, 265D9, 267A2, 270N12, 271I24, 275E13, 278B4, 280E20, 286O3, 289N4, 292D10, 298J6, 306O4,	
	A002G19	308C11, 315F22, 318M12, 319I6, 319P24, 336G18, 343A14, 348C12, 352O10, 362C13, 374N23, 387N22, 388D13,	
	WI-11858	390C10, 396H12, 396N12, 404P18, 408G17, 408K20, 409N13, 414B12, 414D4, 414P23, 417P24, 418H18, 422J4, 424I22,	
	A004O25	424P4, 426P1, 427L6, 429D12, 431M23, 434I6, 441E22, 442C10, 445I23, 448N5, 452K12, 453D2, 453E2, 462E18,	
		466I22, 466K21, 468E14, 473J7, 475D20, 479M18, 480H19, 481F6, 486K23, 495L14, 499G18, 499H13, 507K5, 510J14,	
		510A21, 510E24, 510K9, 511O2, 528B11, 529D3, 532K20, 532O20, 537M13, 541O18, 548K23, 557C11, 561I23,	
		562L13, 569I12, 570M19	
	21	N20191	2M13, 12B22, 13I20, 17B3, 17I22, 18I21, 19C6, 19K9, 20D13, 22A21, 34A14, 34D15, 36D5, 42K11, 42O11, 43F15, 54L24,
sISG15181		58N21, 69A9, 71F4, 80H11, 85A1, 114F7, 116E5, 116D23, 121I4, 144B21, 145P11, 157O7, 159K16, 160I11, 172M13,	
SGC32394		175B20, 178D2, 179L19, 179N20, 189B8, 190O12, 190O13, 196K8, 202P6, 256E14, 260L2, 281J21, 285F16, 291P4,	
sISG46524		299I7, 306O4, 316M21, 325P4, 327C15, 331G8, 348N4, 359I5, 361J5, 383K22, 384I8, 392H5, 395E12, 404I20, 405D22,	
N35618		407A10, 411B6, 411G6, 422N23, 425D3, 433A14, 436E6, 437J4, 439D8, 447I9, 454C22, 455H10, 458I22, 459F3,	
A00CX19		460C19, 469H13, 475F16, 478N5, 478O3, 480G19, 483F11, 483J1, 507P18, 519H9, 520E2, 528M17, 531G10,	
A007K11		546L22, 550C5, 553C6, 553H1, 565A11, 565H24, 565I10, 569B10, 572F2	
SGC34978			
sISG50801			
W37897			
22	sISG49129	2A2, 2P7, 11K21, 12M16, 13F20, 15I4, 16A13, 18I14, 18O12, 21N23, 26I17, 29L1, 30F12, 30H12, 31C2, 31L23,	
	sISG30384	35C6, 37K12, 37L21, 38P1, 40M1, 47K14, 55H24, 55P24, 56D2, 56N17, 58N21, 70E24, 75D10, 75K16, 76G2, 81E5,	
	sISG47056	82M1, 86M22, 89G15, 92K19, 94I8, 95E22, 99N20, 99O20, 101J13, 105F14, 108L7, 112M13, 119L24, 124A14, 127L20,	
	sISG48072	130B8, 131P16, 135G2, 135G3, 135I2, 136P13, 140N14, 148K21, 148D8, 155E8, 157H10, 158G3, 161J3, 163J10,	
	sISG47005	165K8, 166D12, 170J3, 176A24, 179B2, 181L7, 186B2, 190J1, 195N2, 203B23, 208A18, 211D4, 212G21, 212F24,	
	sISG50987	216L6, 223M3, 225G22, 225H22, 226E19, 226C19, 229N4, 230H19, 242B12, 249E8, 250J20, 252D12, 260B2, 262D21,	
	A006D27	267E5, 269N23, 275A16, 277F21, 279B8, 285I3, 287G20, 293M13, 296A11, 296H12, 298H1, 299H5, 299C16,	
	N37090	299B22, 300L4, 301L10, 301H14, 301F16, 302K17, 303A14, 303A24, 307E17, 310K2, 311D11, 319L14, 320K6, 324G3,	
	R45921	326I7, 326J7, 330M2, 332C10, 334C18, 345B6, 346A7, 352J8, 354F7, 356L5, 360E9, 366K23, 369M5, 371A11,	
	sISG12760	381N11, 387O2, 389D5, 393G6, 393G12, 395M19, 396N9, 401D10, 402C8, 403K22, 407C2, 409D24, 409M21, 410C10,	
		415B23, 416K23, 418O7, 422D3, 422H16, 426A8, 426P21, 430M18, 432I2, 432F13, 432H13, 444A13, 447B23, 447F7,	
		449D20, 456L10, 457H4, 459H15, 465A7, 468B22, 468J18, 468N18, 469D8, 471P23, 472D22, 473A22, 474E6, 475M16,	
		476K10, 489E24, 492H3, 493P17, 496G1, 497P24, 502B19, 505A14, 506E3, 507G20, 513A17, 522D15, 522K10, 523E3,	
		524L15, 525P19, 527G8, 528J4, 530G15, 532M13, 541N11, 542J7, 547M21, 550D14, 550G18, 550H22, 551E2, 552C1,	
		556D1, 556I14, 557K9, 558F14, 565O18, 571D14, 572F16, 573E23, 574J10	
	23	WI-16106	1A9, 1P12, 10B18, 24H16, 42F3, 42H3, 44L23, 45A2, 45H3, 46C9, 49O23, 68H13, 68J11, 69E21, 76F6, 90J14,
		sISG9160	93E16, 98B4, 98C4, 98D4, 98F2, 103H19, 110A14, 110G14, 117P7, 123J14, 130E11, 133L11, 133L23, 144M12,
AA243510		145C12, 148M22, 151E21, 154K6, 154E18, 162D11, 162B11, 172G11, 173J5, 177C23, 182C2, 182P7, 189H20, 192H14,	
WI-15620		193K6, 193L6, 195H22, 201L1, 207D20, 211A2, 212L6, 215C21, 225A6, 225B6, 239K17, 246C11, 252B18, 252O7,	
sISG49291		258P15, 260L13, 263P6, 265F22, 265L17, 271H13, 273N6, 274M24, 280F23, 284D14, 284H20, 287P16, 288K8,	
sISG26986		292H1, 293H15, 300E3, 303C5, 303D5, 306H13, 308D20, 309L19, 313D8, 317J21, 318O16, 321D16, 324G5, 328L2,	
sISG48063		331J3, 331J4, 340U7, 344A11, 346N21, 347O8, 348N5, 353B24, 353C13, 353H17, 354J17, 363N23, 388O15, 387F22,	
SGC35716		387L20, 387N21, 388A22, 389B17, 403L6, 407P24, 408B4, 408D4, 422K11, 430P13, 432H14, 436N19, 443A12,	
sISG8954		444J17, 445L22, 449O21, 451M19, 465G12, 476J18, 476P13, 478F14, 482J10, 485K6, 487I24, 491J16, 491K17,	
sISG12737		497J5, 503I6, 508E18, 511O21, 514J6, 521L8, 534A1, 538M10, 549L6, 554A2, 556K19, 558F17, 564F21, 572J23,	
sISG22433			
sISG42559			
24	sISG42602	2A2, 2M16, 13F20, 21N23, 26I17, 29L1, 30F12, 30H12, 37G16, 38P1, 39H10, 40M1, 43P8, 47K14, 49I2, 55H24, 55P24,	
	WI-22164	59C8, 59H20, 59I3, 70E24, 75D10, 85J3, 86G10, 94I8, 95J12, 99N20, 99O20, 107I14, 117B6, 119K16, 119L24,	
	sISG13279	119O18, 119P22, 121B1, 121B2, 123H24, 131P16, 164H18, 169I9, 169K8, 186B20, 190J1, 216L6, 217B13,	
	sISG48239	217G12, 218J13, 220N10, 229N4, 229P4, 230H19, 239M11, 241N12, 242B12, 245B2, 246K17, 250J20, 251E22,	
	sISG50971	252A13, 259N20, 264P19, 268G17, 300I11, 310C11, 313B5, 324L3, 327O9, 329C7, 332C10, 336L8, 344O15, 346A7,	
	sISG47299	349G14, 353K14, 357C7, 362B11, 363L19, 364M5, 366K23, 367C4, 369M5, 371A11, 372K16, 372L16, 373N18, 380K2,	
	sISG31562	388D6, 390A9, 395A22, 399P6, 399L6, 400D18, 400F12, 403G18, 405C3, 407C21, 409D24, 413E5, 416N2, 426A8,	
	sISG44115	430D8, 438M6, 440H6, 456B4, 459M17, 462J17, 468B22, 468J18, 468N18, 472D22, 471G17, 474E6, 474E14, 475M16,	
		479O18, 489E24, 490L17, 491G16, 496G1, 501N10, 502B19, 506E3, 507E12, 511F4, 514D16, 515C1, 517G4, 524L15,	
		525P19, 530G2, 530A20, 530A20, 530G15, 550G18, 560B7, 570F15, 571D14, 571B10, 572F16, 573E23, 574J10,	
25-qtel	sISG1460	1N13, 4K14, 5A13, 7K2, 7J13, 8F15, 11F17, 14C22, 14O10, 17M13, 21M8, 25L8, 26E23, 26F22, 29J6, 30A2, 30E4,	
	sISG1566	31A20, 31D20, 32H11, 32P21, 33D11, 35C24, 41F7, 48A2, 48E17, 48E18, 48P16, 50E23, 53G11, 58N21, 62N18,	
	sISG8192	67B9, 67N15, 76K8, 77F5, 82H11, 85C15, 88B12, 90B19, 90O13, 91E2, 96C10, 98H19, 106C7, 108K14, 109A6,	
	sISG48184	110E23, 112D7, 116J12, 120N8, 122A23, 122E13, 122K13, 124N12, 125H18, 126E16, 126G14, 128P8, 134G23,	
	sISG27915	134G24, 134H23, 135J12, 140A10, 142I9, 146F20, 151O15, 153D15, 156N9, 165K1, 167C6, 168C9, 168K18,	
	sISG4213	169E24, 170F24, 170G20, 170M2, 176D10, 180P1, 184D1, 184I20, 184L4, 196O13, 203B14, 206H3, 213D18,	
	sISG2795	210E22, 210N17, 214B10, 215A22, 219N9, 220B11, 223P11, 224A13, 226G7, 226H12, 228A16, 229C15, 234G16,	
	sISG30731	240E3, 241I12, 248G9, 251P10, 252H16, 253C14, 253D14, 258L6, 265C14, 268H4, 267I9, 267I10, 267K7,	
	sISG41791	271P15, 274B7, 274I7, 276C12, 283A4, 283G7, 283N21, 288A5, 288F1, 289A20, 290N15, 300B2, 301I21, 303B24,	
	sISG46145	303O3, 305E21, 305O23, 305J13, 306D12, 312B12, 312G12, 317M5, 318M10, 319A15, 319M21, 322B10, 327M9,	
	sISG43280	328H8, 330A10, 330E16, 331B17, 332P17, 337J4, 337O17, 340H11, 341P12, 343L20, 350K23, 354M21, 357L4,	
	sISG45453	357M20, 358B21, 360C11, 365A5, 371J2, 373M15, 373M17, 380E16, 380J17, 381E4, 384J24, 387D6, 388D9,	
	sISG43417	390D4, 390F8, 393C9, 395A24, 398D5, 400C15, 400F15, 402D14, 410M10, 413F20, 419N19, 423K12, 425F5,	
	sISG45353	425G21, 427M8, 430C24, 431G18, 434P18, 435P9, 436N17, 442B6, 443L19, 446E5, 448I7, 451B5, 452N9, 453A9,	
	A008M13	458A5, 463E4, 463I4, 465K2, 466F21, 467B2, 479E11, 479H17, 486K3, 483D18, 483I5, 487G22, 489J4, 492O7,	
	D10S1561	496I21, 500B2, 500G10, 502M15, 507K21, 510F23, 511B10, 516O11, 526K3, 532H18, 535A6, 536B18, 536H19,	
	R96892	534G20, 543G19, 547P11, 550I5, 553B4, 554F19, 560J16, 564D19, 571H21, 576C4	
W67510			
AA036862			
WI-14945			
WI-17432			

Πιν. 8: (συνέχεια)

POOL	STSs	BACs	
26	sISG50884	6B11, 7K1, 8A24, 10B18, 13F9, 17C18, 25C19, 26C3, 30C3, 31B9, 31O24, 41L19, 55B17, 58C12, 66H11, 67A18,	
	A005M36	68F10, 69J14, 69P21, 77E11, 78N18, 79A18, 81G7, 105G3, 110F10, 110L9, 114I19, 114M19, 118I9, 124C6,	
	T97004	129M16, 135E9, 137J2, 137F2, 139K1, 140G2, 140I2, 145B17, 145A20, 145J21, 154N23, 158I7, 160F2, 165P12,	
	sISG8843	166G14, 167G10, 170K14, 173D4, 180J12, 181K20, 184J10, 185B1, 187L7, 188A21, 194C21, 196B19, 197O13,	
	sISG26044	200N20, 200M18, 203L6, 204H19, 206G23, 206I22, 207E3, 207G2, 208I17, 209D1, 209C2, 214N15, 219K6, 220M20,	
	WI-14968	222O1, 230P2, 230N1, 238E2, 239E22, 245J24, 250C23, 252A11, 252B10, 255C14, 256J1, 260I4, 265I9, 273E11,	
	sISG49550	281E6, 290K19, 297A11, 300N9, 304N19, 310D11, 311D2, 312P6, 316M11, 317C1, 318F6, 319I23, 323N19, 339P18,	
	sISG39063	342C13, 348N5, 351G21, 352G2, 352F16, 354F10, 354L10, 366J24, 367E10, 369D14, 373P12, 377L17, 379H3,	
	H17634	381G3, 389A6, 389E6, 391J18, 395B13, 396N23, 397C11, 398I2, 398M2, 400M13, 406G10, 406K10, 407E10, 435O11,	
	sISG12596	441L1, 443K9, 446A19, 448G4, 450O14, 451G7, 451K6, 453B24, 463C8, 463G17, 466I19, 467I14, 468L8, 479A21,	
	sISG49455	479B21, 483J5, 487I21, 492I22, 498J9, 513N5, 514M1, 514P11, 514N10, 516B6, 527C19, 530A17, 531J4, 539I5,	
	SGC31141	540A24, 542I14, 542J15, 543O13, 546C7, 546I12, 547A12, 550D7, 551E1, 552C9, 558K18, 559O6, 568M10, 572B12,	
	sISG29677	573M9, 575B12.	
	A006J06		
	WI-17804		
	27	sISG57727=1284	2F13, 12M16, 18I14, 18O12, 30F12, 30H12, 31P18, 38P1, 46D14, 47A8, 49I2, 51A7, 58N21, 63G20, 67K17, 68M5,
		sISG22607	69H10, 73I21, 83C5, 85J3, 87A18, 90J14, 91A6, 94I8, 95J12, 105N15, 110A14, 110G14, 121B1, 123G19, 124G18,
sISG33077		127B10, 131P16, 147B13, 148E14, 149N17, 161J3, 165H14, 175O19, 177N16, 179B12, 182P7, 190J1, 198K1,	
sISG15820		200M10, 213B12, 215C21, 215H21, 216K21, 216L6, 217B13, 230H19, 239A7, 246D12, 247I14, 248J23, 252A13,	
sISG51021		252O7, 256O1, 264P19, 272G5, 289P19, 294B4, 300I11, 300O15, 301F16, 302K17, 303N10, 306I13, 308M12,	
W69567		310C11, 318M12, 320K6, 331J4, 324L3, 325G20, 332C10, 336G18, 344B9, 346A7, 348C12, 349G14, 361B6, 361F5,	
sISG42688		364M5, 369M5, 372K16, 372L16, 377N7, 381E23, 387N21, 388D6, 389J6, 399P6, 390A9, 395A22, 400D18, 408K20,	
WI-19473		412G16, 414B12, 426A8, 426P1, 427G23, 428I7, 429D12, 429G19, 430P13, 434I6, 435O9, 435P11, 440H6, 445I23,	
sISG3345		452K12, 453D2, 453E2, 462J17, 468J18, 468N18, 472D22, 474E14, 483B5, 491G16, 501N10, 506E3, 508E18,	
sISG40669		514D18, 517G4, 524L15, 529I10, 530L13, 530G2, 537L19, 544J23, 550G18, 560A23, 560B7, 571D14, 573E23,	
sISG48312			
sISG46999			
sISG41370			
28		A008S06	5G2, 6B11, 12B23, 17C18, 20F16, 21D4, 27I10, 28O6, 29P5, 31B9, 38I17, 41L19, 53A23, 57H14, 86E10, 87N15,
	sISG3554	90E14, 96N16, 98I3, 106M7, 108K1, 120N11, 124C6, 129M16, 137F2, 137J2, 139K1, 142B2, 143O23, 152P20, 158D9,	
	sISG47370	161N5, 163E8, 163G8, 163L15, 165P12, 167M7, 184J10, 187E5, 187M3, 188A21, 190F19, 195A24, 196I15, 197K7,	
	sISG47767	200N20, 203A7, 206G17, 218G17, 249K21, 253A7, 255L20, 255L21, 256E17, 276N23, 281C5, 288M17, 296E24, 300N9,	
	A007I48	302K2, 304N19, 317C1, 320J17, 324N10, 327A11, 328D18, 328O18, 336A8, 336B7, 354F10, 354L10, 354M20, 361G9,	
	sISG15535	363M6, 377L17, 385O16, 391G21, 393E20, 393G20, 403B19, 413J5, 415G16, 421C7, 422E6, 446A19, 451L24, 453B24,	
	sISG36024	460M19, 466I19, 467I14, 468L8, 481H12, 482G14, 517K7, 518K6, 537G20, 539D18, 539E17, 540A24, 548K24, 551C2,	
	N38956	551E14, 556E2, 559H10, 564G1, 566K12, 567B18	
	sISG42986		
	sISG46271		
	A006A44		
WI-16878			
29	D10S222	3G2, 30G2, 31I15, 37O22, 40N8, 58N21, 62G6, 66C4, 67K6, 71N16, 78B19, 79M5, 83C5, 85J3, 86D19, 87P8, 107A9,	
	D10S571	126A8, 136O21, 147C3, 161J3, 177E2, 186B4, 198O21, 205N7, 211E23, 225F22, 225G22, 244A2, 271L15, 300E21,	
	D10S577	307N12, 310C11, 315C16, 320K6, 326J7, 351P19, 355J6, 362N20, 364M5, 366I13, 372K16, 372L16, 375N9, 383E10,	
	D10S583	390F23, 399P6, 400D18, 400G14, 402N18, 409N24, 414B12, 416N2, 426P1, 439P3, 440H6, 462J17, 463J11, 472O8,	
	D10S597	474C17, 474E14, 476E15, 476L1, 491G16, 506F1, 514D16, 516J19, 517G4, 550A24, 551O13, 560B7, 567P23	
30	sISG575=D10S198	2A10, 2M13, 10G20, 11P23, 34D15, 35H23, 39E8, 43F15, 46G13, 49H17, 55F23, 56O4, 58C10, 58N21, 68J7, 69F17,	
	sISG835=D10S564	70L24, 71F8, 71N16, 77F18, 103A2, 144P18, 146I16, 148C23, 151I18, 160I11, 164M14, 168M11, 184C6, 193B23, 195F21,	
	sISG34710=D10S1739	198A9, 223O16, 245D15, 247E7, 264B12, 269A20, 283I21, 293L10, 294C1, 304L3, 320F15, 323K21, 360H20, 362O4,	
	sISG34711=D10S1741	363H4, 368I8, 393N22, 401P6, 415J12, 415K21, 426G23, 433K7, 411O15, 450E20, 454C21, 456L15, 460K16, 460M16,	
	D10S603	460N21, 460N22, 465H6, 473K12, 474H4, 480L6, 481N24, 482A10, 488C11, 491K20, 492L19, 496D14, 496H23, 510B24,	
	D10S1681	511I16, 514G18, 517O8, 523D19, 526M21, 532L6, 533O14, 535B2, 536G17, 539D21, 557J23, 558K13, 560G4, 567B5,	
		573B17, 573L13	
31	F09841	7K1, 8D20, 12J23, 13F9, 14H22, 17I22, 19C6, 19K9, 20F16, 25C19, 25O20, 27B9, 28G21, 30J8, 35J23, 37L21, 41D21,	
	L25676	44J24, 45G10, 47K24, 53A23, 61K23, 62J15, 65C4, 67A18, 68F10, 69J14, 71O3, 80H5, 81C11, 84A2, 86E10, 86P17,	
	A006W20	87L9, 90E14, 93C15, 95J18, 101I14, 105G3, 108L7, 110M24, 110P1, 111H11, 115P8, 118I9, 119I23, 120N11, 143J17,	
	D29519	145A20, 145B17, 148P19, 151A19, 153P22, 154G4, 168I22, 168O13, 169C12, 169E12, 172M13, 173C8, 178D2, 179F1,	
	sISG9628	181C9, 181I8, 182A20, 186B4, 189B24, 191J13, 194D11, 195A3, 196I15, 199D17, 200I21, 202P6, 211E23, 212N8,	
	Z40200	213C11, 216C18, 218E3, 219K6, 222O1, 223B13, 228B15, 230N1, 230F2, 236H7, 245K20, 251F13, 253A7, 256H10,	
	A009B31	256J1, 260I4, 266B14, 266P20, 268F5, 272M22, 273E11, 280C8, 281J21, 290K19, 297A11, 299I5, 300E21, 301B8,	
	H90420	304C14, 307N12, 311D21, 316L13, 322G22, 324N10, 326I17, 327G20, 334D12, 336A8, 336B7, 343P6, 348J12,	
	D10S2197	348N5, 350F20, 352C22, 353H23, 353L10, 360E9, 366I13, 369D14, 369K19, 370K23, 373P12, 382N12, 391J18,	
	SHGC-9747	393E20, 393G20, 395B13, 396N23, 400G3, 400G14, 400M3, 411G6, 425C21, 430D21, 439P3, 441L19, 442I15, 443K9,	
	NIB1357	446H13, 451D9, 452B15, 459F3, 465E22, 466J14, 469C1, 474A18, 475F16, 476L1, 479A21, 479B21, 481O14,	
	sISG66850=L07093	483E22, 483J5, 492A19, 492I20, 492I22, 499K21, 509C7, 514M1, 514M10, 515D2, 520F2, 523F17, 523H23, 523M11,	
	U29690	531G10, 533B2, 539D18, 539I5, 540G5, 542I14, 543N17, 547M11, 548N18, 550H22, 554E22, 557F6, 558K18, 564G1,	
	T60418	565I10, 572B12, 575B12	
	T91988		
	W80406		
	SGC31312		
WI-9487			

Πιν. 8: (συνέχεια)

POOL	STSs	BACs
32	sISG52092	2D2, 2F13, 3H12, 5G5, 7K1, 12J7, 14H13, 15I4, 17C21, 18I14, 18O12, 29L1, 31I24, 31L23, 31P18, 34C10, 34E5, 37K12,
	A009E24	37L21, 39D23, 43P6, 47A8, 49I2, 50H21, 51A7, 55H24, 56P24, 57E4, 57E5, 58C12, 66H11, 67A18, 67C5, 67K10,
	A005C16	68F10, 68L16, 69P21, 73H14, 75D10, 76B23, 81G7, 83C5, 83C14, 83C15, 85J3, 88C2, 92K19, 95J12, 95P8, 100D5,
	sISG51013	100D6, 100G4, 106M7, 106N20, 107I14, 108I5, 108L7, 114F1, 116M10, 118J3, 119L24, 119O18, 121B1, 123L1, 131P16,
	N31939	134L1, 135G2, 135G3, 135I2, 136O21, 143H15, 145P10, 146L10, 148C15, 148G20, 150M19, 154M18, 155M3, 156N4,
	sISG43839	156M10, 156N10, 159K22, 160D2, 161N5, 162J6, 162L8, 163D2, 164E11, 166D12, 169I19, 170J3, 175O19, 181E16,
	sISG4193	190J1, 200B6, 200K5, 203A7, 203B23, 206G17, 206I24, 208I17, 214N15, 215H21, 216K21, 217B13, 220E18, 220P4,
	sISG46535	221I22, 225G22, 225H22, 225K21, 228K19, 229N4, 230C8, 230N1, 230P2, 233G24, 245E19, 245H16, 246D12, 247K23,
	sISG48287	248P14, 252A13, 258A12, 260I4, 260O2, 264P19, 268G17, 271L15, 273C24, 277F21, 279B8, 280F22, 293M13, 294B4,
	sISG30235	296L23, 300I11, 301F16, 301H14, 301L10, 304P6, 310C11, 310F23, 310K2, 313B5, 322O10, 323P10, 324L3, 328O18,
	sISG47726	336G18, 339P18, 342K23, 343C14, 344B9, 344O15, 345B6, 348C12, 349G14, 357H24, 360E9, 364M5, 369D14, 369M5,
	sISG46822	370B4, 372K16, 372L16, 375N9, 377N7, 381E23, 381N11, 381N23, 382I13, 385D3, 388D6, 390A9, 391J18, 393A21,
	sISG48141	395A22, 396N23, 399L6, 399P6, 400D18, 403H18, 403K22, 406K1, 408K20, 409L17, 413E5, 414B12, 414J11, 415G16,
	AA284841	416K23, 416N2, 417J24, 419G15, 424I8, 426A8, 426E5, 426F8, 426P1, 427G23, 430D8, 430G5, 432G2, 434I6, 437A23,
	A006C33	438M6, 440H6, 450N3, 450P3, 451E6, 451H23, 451K6, 451O7, 452K12, 453D2, 454L22, 454P14, 456I7, 456L24, 459A17,
	sISG30264	460M19, 460O23, 462J17, 470M7, 471A7, 471P23, 474E14, 474M2, 476A4, 487K11, 488G3, 489E24, 490E24, 491G16,
	sISG45253	492O21, 494G7, 495C3, 495B24, 495F4, 495G15, 496G1, 508J18, 513A17, 513G4, 513N5, 514D16, 514I23, 515C1,
	sISG48444	524F19, 525A16, 525K18, 527G8, 527G22, 528O2, 529D2, 531J4, 538A3, 539I5, 542H2, 542H18, 542I14, 544D23,
	N32594	550G18, 550H22, 551E2, 551N13, 558K18, 559E3, 560B7, 560L1, 560L16, 560P1, 561D21, 566K12, 572B12, 573E23
	A006F09	
33	D10S579	5J1, 11O21, 15N11, 21I24, 24A3, 30G13, 41C12, 58N21, 64K2, 66I10, 66J10, 77F13, 79A15, 95A18, 98M21, 105A10,
	D10S1735	106D15, 110K7, 113H4, 115J15, 115N17, 122D20, 124B18, 125J8, 132P22, 136F21, 137E12, 138J21, 138J24, 141N20,
	D10S1753	146P19, 149J19, 155M6, 156C9, 162L12, 164E3, 165M8, 167H16, 169F20, 170O19, 176F8, 186B4, 188C5, 195A12,
	D10S1755	199A4, 201M14, 204K3, 208M2, 210E13, 211E23, 219J23, 220I11, 229C23, 231I18, 234L12, 243L2, 245A22, 252J11,
	D10S1758	253L6, 263M9, 264D5, 264E2, 264I7, 268A23, 268P24, 271I14, 278B4, 281C3, 284E3, 285D15, 287K24, 300E21,
	D10S1765	302E2, 302H1, 302L4, 307N12, 309B15, 315J11, 318H2, 319B14, 325O23, 329L12, 332A11, 337H16, 338H11, 342N15,
	D10S1778	344O15, 348E17, 352E7, 363D8, 364F9, 367M7, 370I7, 379G17, 379K18, 382M19, 383D9, 386E23, 390O19, 400G14,
	D10S1268	406C13, 408G17, 413A9, 418H18, 421C21, 428M16, 433G8, 439J3, 442F10, 442I21, 446K18, 446G11, 447I14, 448L14,
		453D2, 453E2, 454J2, 466H6, 468M23, 468N21, 469M1, 477P1, 481P1, 482E11, 491K20, 497H5, 497N19, 498I6, 509I15,
		510B16, 513K14, 517O8, 524I13, 538C20, 541N10, 545G20, 547J17, 555A18, 558P19, 559D10, 559P8, 571M9, 574M1
34	D10S1266	2F13, 3N15, 16E4, 29H7, 34E5, 47A8, 54F15, 58N21, 71N16, 88C2, 96N16, 11D21, 113H17, 116E5, 116M10, 121I4,
	D10S1663	140L19, 140M21, 145P10, 146P20, 148C15, 152P20, 156M10, 156N10, 156G16, 163L15, 170I16, 183D1, 191C24,
	WI-16578	192G22, 196K8, 215H21, 216K21, 226J23, 226L19, 230C8, 233L9, 236K20, 247P13, 252I9, 256O18, 276E9, 281C5,
	D10S1692	291P4, 293M13, 294B8, 303J11, 334G1, 342K23, 344B9, 349G2, 353I12, 359I5, 361J5, 369I23, 376P12, 377N7, 389F17,
	D10S1726	402P15, 404I22, 409L17, 411B6, 427G23, 433D10, 437A23, 454H24, 455M18, 460C19, 470M7, 483J1, 484I13, 491M7,
	D10S1731	491O9, 514I23, 516H12, 517K7, 517M19, 517M20, 525O3, 527G8, 528M17, 530J7, 542C22, 548K24, 550C5, 556E2,
	D10S1736	555E13, 558D11, 559H10, 560L16, 570A6, 570C6, 572E11, 573F20, 576M4
	D10S1748	
35	D10S1760	2H2, 3D17, 6B11, 8A4, 8C2, 17P13, 18E7, 20J11, 21N23, 27B9, 31O24, 35P4, 38P1, 39A18, 42J22, 57H14, 57I8,
	sISG58067=AFM183xb12	58N21, 59C8, 59H20, 59I3, 69J14, 71N16, 78N18, 83A2, 118A12, 118I9, 132K21, 133O4, 137F2, 137J2, 140N3, 154G4,
	sISG58023=AFM205tg7	154K6, 166G14, 166O7, 173D4, 178O18, 179B2, 179C23, 180G19, 186P6, 190E24, 193L19, 197K23, 197L23, 200O3,
	D10S1265	206L1, 208A18, 215P6, 216B24, 218D3, 220M20, 230H19, 238E2, 242B12, 250C23, 252F4, 256O18, 265J14, 273N6,
	sISG58028=AFM265yc9	285O13, 287G8, 290K19, 290L19, 291F12, 301O4, 302K14, 303K17, 306M4, 308B24, 309L21, 309N13, 316M21, 320C9,
	D10S1269	320M9, 322O12, 328K1, 340K4, 340O6, 346A7, 349F9, 353I12, 353I17, 356L4, 360E9, 362F20, 370K23, 376E11,
	D10S1773	376P12, 377L17, 378E5, 383K22, 389P11, 390E10, 395E12, 398I2, 398M2, 400M13, 407A10, 407C2, 407C21, 416K23,
	sISG58042=AFMa053yg5	418P17, 426A8, 431N8, 436E6, 441L19, 442B7, 446K23, 448G1, 450O14, 472D22, 473F13, 474E10, 476B11, 480N19,
	sISG58043=AFMa071xh1	488D23, 491B20, 492I22, 494E20, 506D10, 506P9, 510M9, 514M10, 515N17, 515N18, 517C17, 517M22, 523E14,
		524L15, 526O24, 538E15, 540G5, 544O22, 546I12, 546M19, 547M11, 566F5, 568F21, 569J22, 575K10
36	sISG22312	1A17, 4J3, 4G4, 9M11, 10L13, 17B3, 18I21, 26I17, 31C2, 36D5, 47K14, 49L20, 49N22, 52G13, 54L24, 55H24, 55P24,
	sISG57715=SHGC-526	56I23, 56M3, 62G6, 62P14, 63G20, 67K6, 71N16, 76J6, 78B19, 79M5, 81A19, 82G19, 86D19, 90J11, 90J14, 98E23,
	G13435	108E7, 110A14, 110G14, 114P2, 117B16, 124A14, 125P11, 125P13, 127D1, 130E11, 144B21, 159E9, 160P20, 161J3,
	G13658	166J24, 182P7, 183J8, 188L18, 203J9, 207D11, 215C21, 225G22, 225H22, 229N4, 231L9, 234I13, 240N1, 252O7,
	G13660	257E5, 264C3, 266B24, 276B11, 283E22, 287N2, 299I7, 303E22, 306F6, 306I13, 308A20, 315C16, 316I10, 320K6,
	G13676	324L23, 326J7, 326N10, 327C15, 328K14, 330A24, 331G8, 331J4, 346G12, 347D24, 350M21, 351F9, 357J4, 357I15,
	sISG57440=SHGC-14061	358P5, 362N20, 369I8, 375N10, 378E9, 380M16, 381K7, 384N23, 387N21, 392H5, 393C18, 393G6, 402N18, 406A12,
	G17536	408A7, 409N24, 412H17, 413G23, 413I23, 419N22, 420K10, 426P21, 430P13, 434B2, 435O20, 437J4, 443C8, 446J3,
	sISG57454=SHGC-14535	447A23, 449F24, 450K13, 455O17, 463J11, 465A7, 467N8, 468L18, 474C17, 476E15, 489E24, 490O10, 495N22,
	SHGC-17231	495P22, 496G1, 506F1, 508E18, 514B24, 514N1, 516J19, 520A14, 522K10, 525M20, 527D11, 531L10, 532L19,
	sISG26135	532N19, 540K12, 542B19, 545O17, 551O13, 565E5, 567L6, 567P23, 570F15
	sISG22765	
sISG3501		
37	U37426	2A10, 2M13, 3D11, 7O22, 10G20, 12J7, 12M16, 13F20, 14H13, 16H23, 26K19, 29H7, 30F12, 30H12, 34D15, 34E5,
	sISG29640	36N22, 40M1, 43F15, 46G13, 57H14, 57I8, 58C10, 58I15, 65P8, 67I16, 69P21, 70E24, 76P2, 77B12, 83A2, 84A2,
	AA022645	86P17, 89F21, 94I8, 95N2, 99N20, 99O20, 100A12, 101J13, 110M24, 110P1, 118A12, 122H11, 127L20, 132K21,
	sISG52467	133O4, 134L1, 136M15, 137F2, 137H1, 137J2, 141A20, 142H5, 144P18, 146L10, 146P21, 148C23, 151I18, 155E8,
	sISG16199	160I11, 172J20, 184C6, 186B2, 186B4, 187J7, 187P9, 189B24, 190I7, 199D17, 200O3, 203A7, 205K9, 206G17, 207J24,
	sISG50235	209G13, 211E23, 216C18, 216L6, 220P4, 220M20, 223B13, 238E2, 247K23, 250C23, 250J20, 252D12, 256O18, 260L2,
	sISG47144	265D9, 265J14, 270N12, 273C24, 283C24, 285F16, 287G20, 296H12, 300E21, 300O15, 304C14, 304P6, 307N12,
	FB3F3	308B24, 309N13, 316M21, 320C9, 320M9, 324M2, 324O2, 332C10, 340J13, 344E13, 353I12, 354N23, 361B6, 363H4,
	sISG22417	366K23, 368I8, 371A11, 376P12, 383C6, 385D3, 387A2, 393F22, 393N22, 396J13, 397H12, 398I2, 398M2, 400G14,
	A004M41	400M13, 409D24, 417D19, 419E10, 426A18, 426E5, 431N8, 433K7, 434N18, 439P3, 442I19, 446H13, 450A14, 456I7,
	sISG45467	455M18, 461B20, 462K21, 465H6, 468J18, 468N18, 469M11, 472L17, 473K12, 473N24, 474E6, 474E10, 476M16, 476L1,
	sISG35096	488C11, 490N12, 495F4, 501N10, 502B19, 503G18, 506E3, 506P9, 507G20, 511I16, 519F1, 520E2, 520F2, 523D19,
	D10S1257	524G11, 525M14, 525P19, 528B11, 528L11, 530A9, 530A20, 530C3, 530E22, 530G2, 530G15, 530I22, 539M3,
	D10S1562	543C2, 544D23, 546I12, 546M19, 548K23, 548B6, 548N18, 557J23, 561B19, 568F21, 571D14, 571N6, 572B16,
	WI-5976	572E11, 572F16, 574J10, 576G6
	sISG5936=WI-14791	
	WI-4379	
	WI-4434	
	WI-4865	
	WI-5915	

Πιν. 8: (συνέχεια)

POOL	STSs	BACs	
38	sISG28889	1A17, 2A10, 9C23, 10I4, 11D15, 12B22, 13I20, 20D13, 21L23, 27I3, 30E16, 30J7, 31P18, 31O24, 32I9, 34A14, 35O17,	
	sISG43429	37E23, 38I17, 42K11, 42O11, 46M17, 49I2, 53O21, 56J20, 57H14, 57N23, 58C10, 59N2, 60P18, 62L2, 62C21, 63O11,	
	SHGC-4169	64L22, 67D22, 70C15, 71F4, 71O6, 78N18, 80H11, 83A2, 85A1, 85J3, 89H7, 93O9, 98I3, 100C1, 103I2, 107L7, 107I14,	
	WI-13377	108K1, 114O20, 116D23, 119O18, 121B1, 121B4, 123F7, 124A14, 126H3, 127J3, 135H15, 137A14, 137F2, 137J2,	
	sISG66857=SGC30780	145P11, 153G4, 157O7, 159K16, 160K16, 161J3, 162B9, 162L22, 166G14, 168O10, 170L14, 170N14, 173D4, 174E12,	
	NIB351	175O19, 178P8, 179H24, 179N20, 187E5, 187M3, 190O12, 190O13, 191N1, 198C20, 198K1, 204L12, 204N5, 209K21,	
	A008I22	211P20, 215N21, 217B13, 220M20, 223P17, 224N18, 225G22, 225H22, 226O18, 230N8, 233G17, 236O9, 238E2, 240N1,	
	WI-1905	241P4, 246A13, 246D12, 247I14, 248J23, 250C23, 251A15, 251C15, 256E14, 257E5, 268G17, 272G5, 281G19, 281M10,	
	WI-3109	282P13, 283C2, 283O16, 288G10, 292G8, 298P3, 298H8, 300O15, 303N10, 304L15, 307A13, 308C11, 310C11, 312E1,	
	WI-4132	312P2, 313B5, 317F20, 318C21, 319P10, 320K6, 320C9, 320J17, 323F7, 325B17, 326J7, 329B8, 329P8, 331E7, 336G18,	
	WI-4209	339P21, 348C12, 353A20, 355H23, 355O19, 364M5, 368O6, 372K16, 372L16, 375N21, 376N20, 385H8, 386K9, 387C15,	
	WI-5502	388M4, 393N22, 394H20, 398I2, 399P6, 399M10, 400M13, 400D18, 402D21, 407F5, 407N7, 408K20, 409K6, 411E24,	
	sISG47984	412I14, 413E5, 415L13, 416N2, 420E3, 422L5, 422N23, 424D3, 425D3, 428F14, 428F15, 430G5, 430D8, 431H7, 431N8,	
	sISG9987	434I6, 435J16, 438M6, 439A24, 440H6, 441G1, 441G9, 442B7, 447I10, 449B2, 450O14, 450P1, 453F3, 453J9, 455H10,	
	sISG39758	455J5, 455O17, 456P10, 458C22, 458D2, 458I22, 462J17, 465P12, 467L15, 467O24, 469H13, 469M2, 469P8, 471M3,	
	sISG51671	473O18, 474E10, 474E14, 474I15, 478O3, 478N5, 479A21, 479B21, 483F11, 483P13, 486O17, 487G1, 488C11, 489M10,	
	sISG38017	491G16, 491I24, 492I22, 495D11, 495O12, 496P3, 496P21, 499C13, 505K23, 506N5, 507K5, 507P18, 508I18, 511F11,	
		514M10, 514D16, 515C1, 515E4, 517G4, 520D16, 526I24, 530C19, 533H5, 533F12, 534O13, 535D1, 535B11, 539B7, 539N6	
		540O12, 542B19, 542H18, 546I12, 546L23, 548I13, 548K16, 553L4, 554J21, 555F2, 560B7, 562L13, 563D6, 565H24, 568I23	
	39-qtel	sISG17459	240E3, 25E1, 377E5, 542E1, 91E2, 1N13, 110E23, 122E13, 126E23, 50E23, 5A13, 126E16, 330E16, 6H1,
		sISG32124	7J13, 9N19, 10B10, 17M13, 18M14, 21F1, 21M8, 26F22, 27F2, 32P21, 35L1, 45A17, 45P6, 52A3, 55O19,
		sISG2750	56O21, 58L14, 58N21, 67N15, 69A11, 69J5, 71N16, 75C14, 77F5, 82L9, 88P13, 90B19, 90O13, 92A10,
		sISG28017	96B11, 101K5, 102O13, 106C7, 108K14, 109A6, 122K13, 124N12, 125H18, 126G14, 127G14, 134G23,
		sISG47732	134G24, 134H23, 136J21, 139I20, 144F10, 146F20, 149H12, 152N21, 155G11, 170G20, 178B8, 182H18,
		sISG45059	182I15, 185G13, 185K17, 194M18, 196O13, 197M1, 198B2, 204C3, 205J9, 211O2, 220J19, 229C15,
		sISG1303	231J23, 234G16, 241I12, 247B17, 253D7, 263B14, 265C14, 266H4, 271P15, 274I7, 286H1, 286N21,
		sISG4790	289A20, 292B19, 296A18, 298M15, 303B24, 305O23, 306D12, 312G12, 317M5, 322B20, 324F1, 326F7,
		SHGC-14552	327B20, 330A10, 331B17, 337O17, 339B9, 357M20, 357O7, 358A16, 358B21, 373M15, 373M17, 384J24,
D10S1256		387D6, 390D4, 398D5, 400F15, 407G17, 410M10, 411O14, 413A9, 413F20, 419N19, 425G21, 429B1,	
WI-17697		431G18, 435A3, 435H16, 448I7, 455A19, 456J1, 458C16, 475K23, 489J4, 500G10, 502M15, 511B10,	
WI-20848		527B6, 528K3, 532H18, 533P22, 534O23, 535A6, 540N6, 543P10, 552L24, 563I5, 571H21	
H99205			
WI15548			
AA010383			
WI-22094			
Cda01F10			
D10S2490			
40-qtel		D10S2229	463E4, 8E6, 177E10, 224E20, 23E10, 122E13, 50E23, 5A13, 330E16, 17C5, 17M13, 17M15, 29J6, 33G5, 38L2, 41F7,
		D10S1585	41G17, 45A17, 48A2, 49M8, 55O19, 57O18, 64I4, 64P19, 67B9, 71N16, 72L6, 75C7, 77F5, 85C15, 96C10, 112C7,
		D10S1588	112D7, 116J12, 120N7, 122K13, 125H18, 134G23, 134G24, 134H23, 142I9, 154K24, 162J15, 167C6, 173D3,
		D10S1593	176L2, 178I10, 181P20, 182K19, 190O5, 198I18, 208A8, 215A22, 220K7, 224A13, 231O17, 235D14, 241P21,
		D10S1408	247L13, 248G16, 252C1, 258L6, 259G12, 259J10, 273H7, 274I7, 283A18, 283F16, 284J10, 285C4, 288A5,
		A009F48	288G11, 289H13, 298L11, 306D12, 311C17, 317C21, 317M5, 319M21, 331K4, 341M3, 341P12, 343D22,
		T48398	358B21, 361G17, 363I7, 365A5, 380J17, 387D6, 389H15, 392J15, 400C15, 400F15, 402D14, 422D7, 424P4,
		AA191340	432J24, 432N6, 442B6, 442O18, 448I7, 450I22, 451B5, 452H13, 455M1, 463I4, 466F21, 470M14, 478F9,
		A008Y06	486K3, 487G22, 488F19, 489J4, 499J6, 502H20, 509A4, 509B23, 509J2, 510K7, 511B10, 513N19, 516O11,
		SGC33916	531F13, 540F8, 540N16, 544I4, 546M18, 552A16, 553O9, 554F19, 571H21, 572H15, 576C4
	WI-19193		
	AA010383		
	WI-22094		
	sISG47380		
	sISG4581		
	sISG42519		
	sISG66859=SGC32344		
	41-qtel	D10S1201	5J1, 10O9, 10L14, 11E21, 15K2, 17E23, 27F2, 30C18, 52K1, 56O21, 58L14, 69J5, 74B4, 82N3, 92A10, 108O12,
		D10S1675	111P12, 113K3, 113I21, 125O1, 130D4, 137C6, 138I1, 139I20, 151A20, 155P2, 158J6, 165B10, 187B3, 197H20, 211O2,
		D10S1711	220J19, 229C15, 245K15, 246K3, 272F24, 293O18, 293P15, 308D13, 322I17, 339B3, 377E5, 407F13, 411N16, 417M4,
		D10S1770	419B4, 422M14, 427A1, 435A3, 436P1, 455A19, 468N21, 479P10, 480N19, 486O2, 500C8, 513D17, 518I7, 525A18,
		D10S1851	528M19, 529M22, 539D17, 539O12, 539N19, 540N6, 542E1, 548O1, 548P5, 549M19, 552P1, 553N1, 555N22,
		D10S1782	559M23, 562M1, 565F22, 575L1
		D10S1676	
		42-qtel	sISG28017
	sISG36110		134G23, 134G24, 134H23, 136L23, 136P22, 142F3, 142I9, 146F20, 151O22, 162J15, 167C6, 169E24, 170G20,
	sISG55725=AA7H40		173C23, 17M13, 182K19, 189A5, 18M14, 194M18, 196O13, 197M1, 199O1, 1N13, 211P15, 215A22, 220J19,
	sISG47732		224A13, 234G16, 235D14, 241M19, 246J4, 246L23, 248G18, 251G22, 256F24, 258B5, 258L6, 260B18, 263B14,
sISG46609	266H4, 26E23, 26F22, 274I7, 274M3, 283A18, 286G11, 293F16, 299G9, 2L16, 300E6, 303B24, 304J23, 305O23,		
sISG32124	309K14, 312A16, 312G12, 319M21, 322B20, 330A10, 335E23, 337O17, 341P12, 356D1, 358B21, 35C24,		
sISG45059	373M15, 376M2, 377E5, 380J17, 383F5, 384J24, 390D4, 390G18, 400C15, 402D14, 413F20, 414N24, 419N19,		
sISG4581	41F7, 431G18, 435P9, 448I7, 451B5, 463E4, 466F21, 468J1, 474A5, 486K3, 487G22, 489J4, 48A2, 502H20,		
sISG1303	502M15, 50E23, 516O11, 51B10, 528K3, 535A6, 536C3, 540N6, 542E1, 54K7, 552L24, 554F19, 55O19, 561G22,		
sISG4790	566G7, 576C4, 58N21, 5A13, 67B9, 67N15, 69J5, 6C2, 77F5, 7J13, 85C15, 90B19, 90O13, 91E2, 9N19		
55747=A008Y06			
sISG42519			
sISG66860=SGC33916			
sISG47380			
Cda01F10=D10S1328E			
T48398			
AA191340			
WI15548			
AA010383			
SGC32344			
WI-22094			
WI-17697			
WI-20848			
WI-19193			

Πιν. 8: (συνέχεια)

POOL	STSs	BACs
43-qtel	sISG4790	2L16, 6O4, 7J13, 17M13, 26E23, 26F22, 32P21, 46H6, 50E23, 58P3, 67B9, 69J5, 77F5, 91E2, 106C7, 108K14,
	sISG4581	110E23, 116J12, 118E15, 122E13, 122K13, 126E16, 134G23, 134G24, 134H23, 139I20, 142F3, 165L12, 167C6,
	sISG28017	170G20, 183A22, 186G7, 197M1, 201O13, 211O2, 220J19, 258B5, 263A14, 271P15, 271G22, 274I7, 278B1,
	sISG36110	305O23, 306D12, 312G12, 317M5, 318M10, 319M21, 322B20, 330E16, 337O17, 341P12, 354M21, 359J21,
	sISG45059	390D4, 396E6, 413F20, 419N19, 431G18, 435A3, 435P9, 436M11, 441D21, 447G24, 447E21, 448I7, 460G7,
	sISG47732	469N13, 487G22, 489J4, 500G10, 502F15, 502M15, 506F7, 507C9, 516O11, 526J7, 528K3, 533O20, 540N6,
	sISG27750	552L24, 554F19
	WI-17697	
	WI-20848	
	H99205	
	AA010383	
	WI-22094	
	Cda01F10	
	T48398	
	AA191340	
	A008Y06	
44	D10S1207	24A3, 30G13, 50P4, 55F23, 82N7, 94M14, 98M21, 106D15, 134D17, 137E12, 149J19, 164E3, 173N12, 173N13, 177P2,
	D10S1242	181P6, 188C5, 195F21, 232M15, 242P6, 243P14, 246L14, 252A23, 269A20, 284E3, 294C1, 298A12, 315J11, 319B14,
	D10S1429	323A21, 352E7, 362P3, 364F9, 366F23, 370I7, 386E23, 406C13, 439J3, 439L16, 448L14, 454C21, 456L15, 480L6,
	D10S1245	482A10, 501D5, 502P5, 511J1, 513K14, 526M21, 535B2, 556I16
	D10S1240	
	D10S1434	
D10S2470		
45	D10S541	17P9, 26K19, 46G13, 49I2, 56A21, 58C10, 58N21, 71N16, 93P18, 121B1, 137B2, 139A1, 148C23, 151I18, 159J10,
	D10S554	167C14, 173C23, 182I15, 184C6, 193A9, 216K10, 217B13, 229C23, 252J11, 270L24, 306A4, 310C11, 319E5, 335E23,
	D10S1709	344A15, 345C24, 345D20, 348J9, 363H4, 364M5, 367M7, 372K16, 372L16, 388D6, 400D18, 413M23, 416N2, 426L18,
	D10S1668	433K7, 446K18, 446L8, 458G8, 460I14, 462F4, 462M19, 465H6, 468M23, 473K12, 474K16, 482E11, 488G5, 491G16,
	D10S1680	491K2, 497H5, 498K17, 511I16, 514D16, 523D19, 536C3, 548K23, 556N15, 557J23, 560B7, 560E12, 562K6
	D10S1738	
	sISG58064=AFMb363xg5	
46qtel	sISG17460	1N13, 21M8, 23N8, 27F2, 31A20, 33B17, 35C24, 42K2, 51K1, 53G11, 55*19, 58N21, 58L14, 69J5, 71N16, 82L9,
	sISG987=D10S555	88B12, 90K8, 92A10, 106L6, 116C8, 136P22, 137C6, 139I20, 142I8, 146F20, 151O15, 156N9, 158J6, 170J5,
	sISG725	175F7, 176D10, 178I10, 182I15, 195E8, 197O1, 199O1, 211O2, 220B11, 220J19, 226G7, 229C15, 245K15,
	D10S1676	246K3, 270L24, 275G4, 283N21, 286H1, 292B19, 312B12, 317I13, 318M10, 322I17, 327M9, 328H8, 328J8,
	D10S1655	339B9, 340L9, 360C11, 365A5, 368P24, 377E5, 400C15, 410M10, 417E15, 435A3, 443L19, 455A19, 467J19,
	D10S1651	469C6, 479F1, 479O20, 483I5, 486K3, 500B2, 513D17, 532H18, 536B1, 540N6, 542E1
	D10S1770	
	D10S1711	
	D10S212	
	D10S1201	
	D10S1675	
D10S1782		
47qtel	D10S1222	11A8, 23A10, 26I8, 26I18, 37A16, 40N2, 40O12, 41G17, 51A6, 51A23, 55O19, 64A3, 72C14, 75A5, , 75G6, 76A15,
	D10S1439	81D6, 82A14, 82J5, 86I5, 88C24, 89A22, 94A21, 107K5, 119H22, 134O17, 136A3, 138H15, 144A11, 148C1, 149E2,
	D10S1248	151E18, 156I8, 156E12, 157E12, 159A14, 165N2, 170C12, 171C12, 171P8, 177B5, 183P8, 185K19, 187P15,
	D10S1134	188J10, 189G19, 193B2, 194P18, 197B13, 199B23, 200K21, 205A21, 214F17, 215O24, 216H22, 221E21, 223P5,
	D10S505	224F22, 226P18, 230P21, 231E13, 239N18, 241E22, 246O18, 247N18, 253L4, 255F6, 263C9, 270D1, 271P7,
	sISG47386	287B5, 295I16, 296H24, 296L5, 302E21, 303E18, 303A22, 307A6, 318H22, 320I19, 332P5, 341D12, 342M14,
	sISG46609	383B19, 383C23, 384E19, 387D4, 397A19, 401N14, 407A6, 419F15, 420P8, 421D3, 445A21, 446A15, 446C12,
	D10S186	449E19, 462C5, 464F4, 467C9, 471O9, 472G2, 475B2, 479M17, 494G22, 495L20, 496G17, 497A2, 516A22,
	D10S169	520A12, 526L14, 535A6, 538B5, 538K21, 540E24, 549I13, 551A24, 553B19, 572A2, 574E24, 574D4
	48	sISG472=D10S185
sISG538=D10S192		47H17, 57M24, 58N21, 68P15, 77B12, 77A13, 85N8, 87E14, 98K12, 106N20, 106P20, 112J9, 116M10, 122J21,
sISG916		129C16, 129J4, 135P8, 141B2, 141B3, 142H5, 143H15, 145A20, 145P10, 145P11, 146L10, 148C15, 150M19, 156N4,
sISG1032		157A7, 160I11, 161N5, 162L18, 163D2, 165P9, 169I19, 189B8, 192L3, 203A7, 206G17, 217C3, 222A1, 225K21,
sISG1953		230C8, 231M20, 236H19, 237F16, 246F7, 247N16, 248P14, 252D12, 256E14, 256J1, 258A12, 270P10, 273C24,
sISG3745		283C24, 285F16, 293C18, 299E9, 304P6, 305F18, 310F23, 318B21, 319C14, 328A18, 337C17, 341B18, 341L11,
sISG4306		342K23, 344E13, 345F3, 348I19, 348N4, 348N5, 357H24, 357J17, 359H22, 359I5, 367N8, 369I23, 370B4, 374I6,
sISG22723		375L18, 379A11, 379P10, 385D3, 393A21, 395B13, 405D22, 409L17, 409A21, 411D18, 413M23, 414J11, 417D19,
sISG26670		417F17, 419K19, 422N23, 425D3, 426A9, 427B15, 429H9, 435N8, 437A18, 439D8, 447I9, 450N17, 451A7, 458C22,
sISG31639		458I22, 459A17, 460C14, 462A12, 469I13, 470M7, 476A4, 478A3, 479A21, 479B21, 483J5, 487F18, 487K11, 488G3,
sISG44588		490I21, 494G7, 499L13, 501A15, 501A3, 502I12, 507P18, 509N21, 514I23, 516H12, 519G9, 519B8, 520E2, 528M17,
SHGC-33545		536H2, 539M3, 551N13, 553H17, 554P13, 557H1, 560L16, 561D21, 565A11, 566K12
AA040321		
D10S2211		
T87905		
D10S1671		
D10S1690		
sISG58054=AFMa301wf1		
sISG58057=AFMa348zg9		

Πιν. 8: (συνέχεια)


POOL	STSs	BACs
49	D10S222	2F13, 15I4, 29L1, 30F12, 30H12, 18H14, 18O12, 31L23, 21N23, 47A8, 37G16, 37L21, 37K12, 40M1, 38P1, 59C8, 49I2,
	WI-14791	59I3, 63G20, 59H20, 55H24, 55P24, 75D10, 70E24, 85J3, 86G10, 92K19, 95J12, 94I8, 107I14, 101J13, 108L7, 121B1,
	N37090	124A14, 124G18, 119L24, 119O18, 131P16, 135I2, 155E8, 154M18, 155M32, 166D12, 161J3, 170J3, 169I9, 162J6,
	R45921	169K8, 186B2, 190J1, 203B23, 217B13, 220E18, 215H21, 216K21, 217G12, 216L6, 220N10, 230H19, 225G22, 225H22,
	sISG12760	250J20, 251E22, 252D12, 252A13, 42B12, 268G17, 264P19, 260O2, 279B8, 277F21, 287G20, 301F16, 302K17, 289P19,
	sISG42602	296H12, 300I11, 301L10, 293M13, 313B5, 310C11, 310K2, 320K6, 326J7, 324L3, 332C10, 346A7, 345B6, 344B9,
	W80406	349G14, 344O15, 367C4, 360E9, 364M5, 363K23, 369M5, 377N7, 381N11, 382I13, 372K16, 382L16, 395A22, 400D18,
	D10S1692	390A9, 388D6, 399P6, 399L6, 387O2, 407C21, 413E5, 412G16, 416N2, 403K22, 416K23, 426A8, 430D8, 426F8,
	D10S1267	427G23, 440H6, 438M6, 462J17, 474E6, 474E14, 475M16, 471P23, 472D22, 468B22, 496G1, 491G16, 495G15, 489E24,
	WI-22164	506E3, 501N10, 502B19, 507G20, 515C1, 517G4, 527G8, 513A17, 514D16, 524L15, 525P19, 530G2, 530A9, 530A20,
	sISG13279	560B7, 551E2, 550G18, 550H22, 574J10, 571D14, 572F16, 573E23
	sISG15820	
	AA913870	
	AA917820	
	A005K38	
	N31939	
	sISG43839	
	sISG46535	
	sISG48287	
	sISG30235	
	sISG47726	
sISG46822		
50	sISG55997=SHGC-4169	2F13, 47A8, 62C24, 52H22, 85A6, 88C2, 95J12, 101J13, 107I14, 127L20, 119O18, 123L1, 123C1, 124A14, 131P16,
	H82835	135P1, 155E8, 161J3, 175H16, 186B2, 200K5, 206L22, 206L23, 217B13, 215H21, 252A13, 252D12, 243G8, 244M5,
	A008M23	268G17, 264P19, 287G20, 279B8, 286K1, 283K10, 288K13, 286K17, 286A21, 296H12, 300I11, 313B5, 310C11, 310K2,
	G26427	311A4, 320K6, 321C9, 324L3, 344B9, 342L24, 360A19, 366K23, 359L2, 374A19, 377N7, 369M5, 374P24, 393A6,
	sISG916	387A2, 390A19, 395A22, 416N2, 413E5, 430D8, 427G23, 436B8, 436M6, 440H6, 446J8, 461F9, 452K1, 474E6, 491G16,
	U36223	507G20, 508G18, 501J12, 514D16, 515C1, 517G4, 527G8, 542H18, 535D21, 551E2, 560B7
	sISG4193	
	D10S1738	
	D10S1697	
	WI-6938	

Σημ.: Στην πρώτη στήλη φαίνεται η αρίθμηση των pools όπως έχουν κατατεθεί στην βάση δεδομένων 10ace και σημειώνονται αυτά που προέρχονται από την υποτελομερική περιοχή (qtel). Στην δεύτερη παρατίθενται τα STSs που χρησιμοποιήθηκαν σε κάθε υβριδοποίηση και στην τρίτη οι κλώνοι που ταυτοποιήθηκαν. Τα ονόματα των ανιχνευτών είναι συμβατά με την βάση δεδομένων UniSTS. Στις περιπτώσεις που σε κάποια STSs υπάρχουν δύο ονόματα, αυτό με το πρόθεμα sISG αντιστοιχεί στον κωδικό του Sanger Centre. Δεν έχει συμπεριληφθεί το πρόθεμα "bA" στους κλώνους BACs. Η παρασκευή 32 ανιχνευτών και η διερεύνηση της βιβλιοθήκης με αυτά (pools 8, 11 και 12) πραγματοποιήθηκαν από την Μ. Κοκκινάκη (Διδακτορική διατριβή) και δεν παρατίθενται εδώ. Επιπλέον οι 76 ανιχνευτές της υποτελομερικής περιοχής και οι υβριδοποιήσεις της βιβλιοθήκης με αυτούς πραγματοποιήθηκαν από την Α. Πασπαράκη.


ο Αντιστοίχιση κλώνων BAC με STSs

Το επόμενο βήμα της ταυτοποίησης του περιεχομένου των κλώνων σε STSs, πραγματοποιήθηκε με διερεύνηση της βιβλιοθήκης του χρωμοσώματος, με κάθε ανιχνευτή ξεχωριστά. Ταυτόχρονα πραγματοποιούνταν τουλάχιστον 24 υβριδοποιήσεις της βιβλιοθήκης, χρησιμοποιώντας ραδιοσημασμένα προϊόντα PCR ενώ η διαδικασία πραγματοποιήθηκε όπως περιγράφηκε στις Μεθόδους.

Από τα 506 STSs που χρησιμοποιήθηκαν στην διερεύνηση ολόκληρης της βιβλιοθήκης, τα 491 αντιστοιχίστηκαν με συγκεκριμένα BACs, 421 από την περιοχή 10q23.31-q25.3 και 70 από την υποτελομερική περιοχή. Κατά μέσο όρο ένα STS αναλογούσε σε 10 κλώνους. Όλα τα αποτελέσματα κατατέθηκαν στην βάση δεδομένων 10ace (ενδεικτικά βλ. Εικ. 18.).



Sanger Centre: Webace



[Info](#) | [HGP](#) | [Projects](#) | [Database Searches](#) | [Software](#) | [Teams](#) | [Search](#)

Webace view of [acedb10](#) [Configure](#) | [Help](#)

Pool : pool 30L
General Remark hyb data from Sarafidou-04-05-99
Positive Positive BAC [[Collapse 78](#)]

bA2A10	bA144P18	bA36204	bA496D14
bA2M13	bA146I16	bA363H4	bA496H23
bA10G20	bA148C23	bA368I8	bA510B24
bA11P23	bA151I18	bA393N22	bA511I16
bA34D15	bA160I11	bA401P6	bA514G18
bA39E8	bA164M14	bA415J12	bA51708
bA43F15	bA168M11	bA433K7	bA523D19
bA46G13	bA184C6	bA441O15	bA526M21
bA49H17	bA193B23	bA454C21	bA532L6
bA55F23	bA195F21	bA460M16	bA533O14
bA58C10	bA223O16	bA460N21	bA535B2
bA58N21	bA245D15	bA460N22	bA536G17
bA68J7	bA247E7	bA465H6	bA539D21
bA69F17	bA264B12	bA473K12	bA557J23
bA70L24	bA269A20	bA474H4	bA558K13
bA71F8	bA293L10	bA481N24	bA560G4
bA71N16	bA294C1	bA482A10	bA567B5
bA77F18	bA304L3	bA488C11	bA573B17
bA103A2	bA320F15	bA491K20	bA573L13
	bA360H20	bA492L19	

Positive_BAC_weak [[Collapse 11](#)]

bA35H23	
bA56O4	bA426G23
bA198A9	bA450E20
bA283I21	bA456L15
bA323K21	bA460K16
bA415K21	bA480L6


Contains STS

- [strRH816](#)
- [strRH1014](#)
- [stAFMb362yg5](#)
- [stAFMc005xh5](#)
- [stAFM350wa5](#)
- [stAFMa272zd1](#)

Enter search word:


[Do full search](#)

Εικ. 17: Ενδεικτική καταχώρηση των αποτελεσμάτων της διερεύνησης της γενωμικής βιβλιοθήκης BAC στην βάση δεδομένων 10ace. Στο πάνω μέρος αναγράφεται το όνομα του pool (30L) και ακολουθούν τα BACs που ταυτοποιήθηκαν, χωρισμένα σε δύο ομάδες: σ' αυτά που έδωσαν ισχυρό σήμα κατά την υβριδοποίηση (Positive_Positive_BAC) και σ' αυτά με ασθενέστερο σήμα (Positive_BAC_weak) και τέλος τα STSs που συμμετείχαν στην διερεύνηση. Όλα τα παραπάνω στοιχεία (με μπλε) είναι επιλέξιμα.



Sanger Centre: Webace

| Info | HGP | Projects | Database Searches | Software | Teams | Search |



Webace view of [acedb10](#)
[[Configure](#) | [Help](#)]

STS : stSG29640

General RHdb_Gb4 [RH48066](#)
 RHace_name [stSG29640](#)
 Parent_sequence [Em:AA131255](#)
 Sequence_type [3'cDNA](#)
 Class [UNIEST](#)
 Remark [Primers sent to Sarafidou-24-11-98](#)

Order_info Ordered [1997-03-17](#)
 Supplier [Genset](#)

PCR Oligo_1 [GATTTTCTGTGCTCTATGCCC](#)
 Oligo_2 [TTACTTTCTTGATGATGCTGCC](#)
 Expected_product_length [129](#)
 STS length [129](#)
 Annealing_temp_time_cycle [62C](#)
 PCR_buffer [SC_file7](#)

PRESCREEN Prescreen_status [Passed + DP](#)
 Experiment_prescreen [exp_60_stSG29640](#) [1997-04-28](#) [Passed](#)
[exp_55_stSG29640](#) [1997-04-28](#) [Passed](#)
[exp_50_stSG29640](#) [1997-04-28](#) [Passed](#)
[exp_62C_stSG29640](#) [1997-05-09](#) [Passed](#)

RH_project [RH_ASSIGNMENT](#) Experiment_assignment [exp_assign1_stSG29640](#)
[1997-05-15](#) [Passed](#)

Chrom_assigned [10](#)
 Assignment_status [Assigned In_house](#)

RH_MAPPER Positioned [SANGER chrom10 rhmap 26 02 98](#) [P1.01 10:1](#)
[Transcript map 98](#) [P1.01 10:1](#)
[SANGER chrom10 rhmap 03 11 98](#) [P1.01 10:1](#)

RH_Links RHalloc [SC35152](#)
 RHallocation [done](#)

Position Map [SANGER chrom10 rhmap 26 02 98](#) Position [438.28](#)
[Transcript map 98](#) Position [438.08](#)
[SANGER chrom10 rhmap 03 11 98](#) Position [438.14](#)
[Chr 10GSC](#) Position [113261](#)
[Chr 10](#) Position [111071](#)
[Chr 10ctg15](#) Position [638](#)

Filter_number [SC10BACpoly_B_1#264 ls2 1999-07-28](#) [No_positives](#)
[SC10BACpoly_B_2#264 ls2 1999-07-28](#)
[SC10BACpoly_K_1#264 ls2 1999-07-28](#)

Positive Positive_BAC [bA146P21](#) [SC10BACpoly_K_1](#) [[Collapse 9](#)]
[bA419E10](#) [SC10BACpoly_K_1](#)
[bA426A18](#) [SC10BACpoly_K_1](#)
[bA396J13](#) [SC10BACpoly_K_1](#)
[bA393F22](#) [SC10BACpoly_B_2](#)
[bA20M23](#) [SC10BACpoly_B_2](#)
[bA3D11](#) [SC10BACpoly_B_2](#)
[bA190I7](#) [SC10BACpoly_B_2](#)
[bA76P2](#) [SC10BACpoly_B_2](#)

Sequence [Em:AA384580](#)
 In_pool [pool 37L](#)

Enter search word:

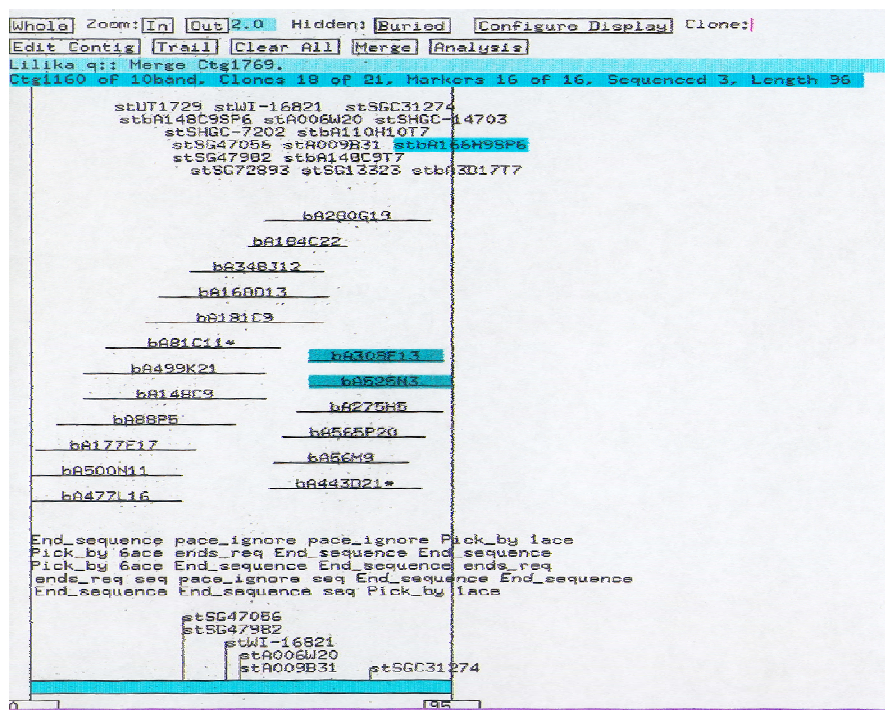
Do full search]

Εικ. 18: Καταχώρηση των αποτελεσμάτων αντιστοίχισης ενός STS με τα BACs που το περιέχουν. Στη σελίδα σημειώνονται: 1) ο κωδικός του STS (stSG29640), 2) στην κατηγορία “General” τα εναλλακτικά του ονόματα και ότι αντιστοιχεί σε EST, 3) στην παράγραφο “PCR” οι αλληλουχίες των εκκινητών, το μέγεθος του προϊόντος και η θερμοκρασία υβριδοποίησης, 4) στην παράγραφο “Position Map” οι συντεταγμένες του στους αντίστοιχους χάρτες συμπεριλαμβανομένου του φυσικού (ctg15), 5) στην παράγραφο “Positive Positive_BAC” οι κλώνοι που το περιέχουν και 6) ο κωδικός του pool στο οποίο συμμετείχε (pool 37L) κατά την διερεύνηση ολόκληρης της βιβλιοθήκης. Όλα τα στοιχεία με μπλε χρώμα είναι επιλέξιμα.

➤ **Χαρτογράφηση contigs και χρωμοσωμικό περπάτημα**

α. Φυσική χαρτογράφηση των προκαταρκτικών contigs

Ο συστηματικός τεμαχισμός όλων των κλώνων της γενωμικής βιβλιοθήκης RPCI-11 με *HindIII* και ο υπολογισμός των μεγεθών των επιμέρους τεμαχίων μετά από ηλεκτροφόρηση πραγματοποιήθηκε στο Washington University (St. Louis, US) και στο Sanger Centre UK, (Marra MA *et al*, 1997 Soderlund C *et al*, 2000). Τα δεδομένα συλλέγονταν με το πρόγραμμα IMAGE, που επεξεργάζεται πρότυπα τεμαχισμού κλώνων και κατατίθονταν στην βάση δεδομένων FPC (fingerprinting contig, Soderlund C *et al*, 1997). Κλώνοι που προέρχονται από την ίδια περιοχή του γονιδιώματος παράγουν πρότυπα που εμφανίζουν κοινά τεμάχια, το ποσοστό των οποίων είναι ενδεικτικό του βαθμού αλληλοεπικάλυψης. Βάσει των αποτελεσμάτων του περιεχομένου των κλώνων σε STSs, επιλέγονταν οι κλώνοι του χρωμοσώματος και καταχωρούνταν στην αντίστοιχη βάση δεδομένων. Από τον συνδυασμό των δεδομένων, σταδιακά αναπτύσσονταν contigs (Εικ. 19) που τοποθετούνταν κατά μήκος του χρωμοσώματος.

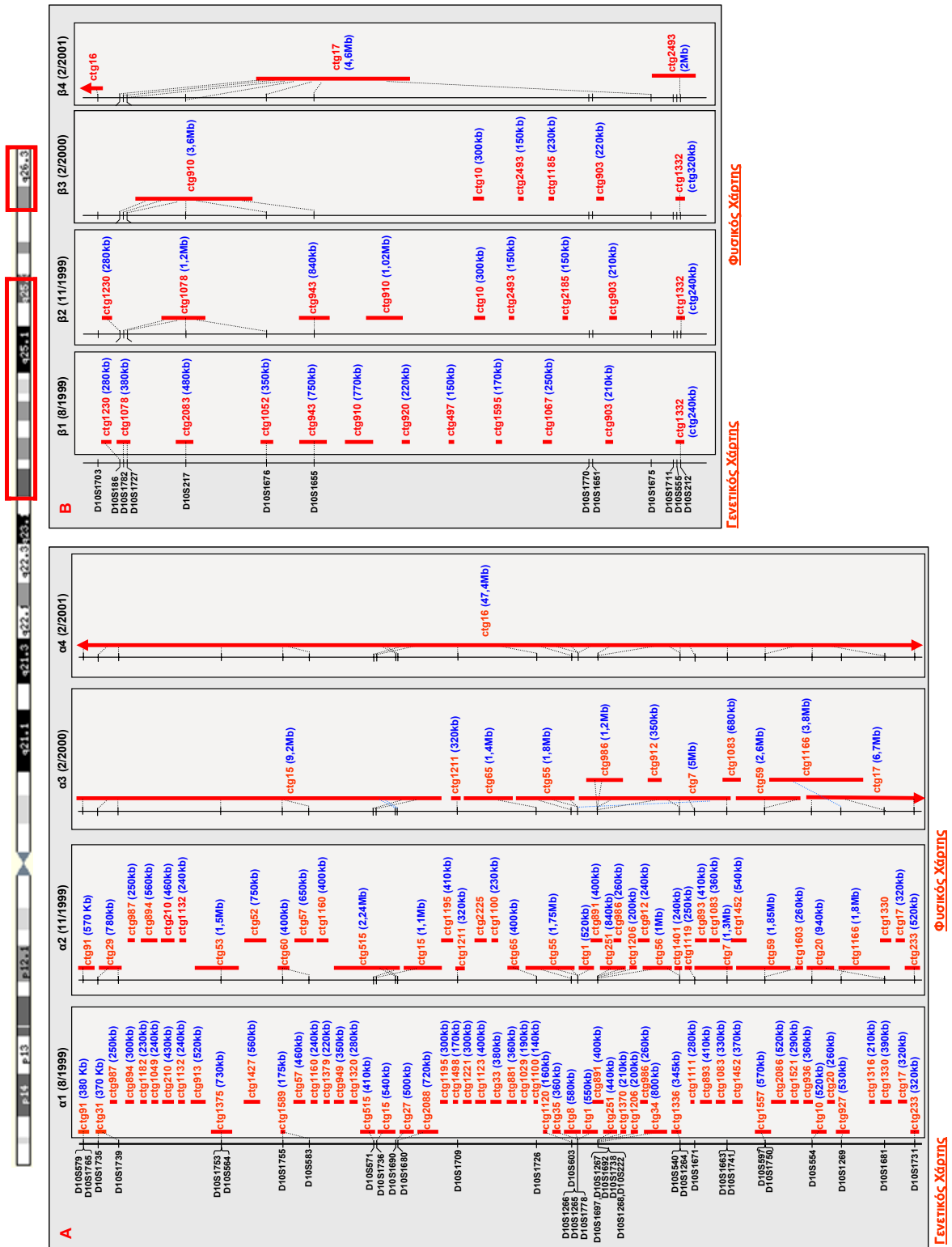


Εικ. 19: Προκαταρκτικό contig που χαρτογραφείται στην χρωμοσωμική περιοχή 10q23.33. Οι οριζόντιες γραμμές αναπαριστούν τους κλώνους όπως τοποθετούνται στο contig βάσει της σχετικής τους αλληλοεπικάλυψης. Στο πάνω μέρος φαίνονται τα STSs που περιέχονται στο συγκεκριμένο contig. Όλα τα στοιχεία είναι επιλέξιμα, π.χ. επιλέγοντας το STS που σημειώνεται με γαλάζιο, εμφανίζονται με το ίδιο χρώμα οι κλώνοι που το περιέχουν. Το συγκεκριμένο contig έχει μέγεθος 380kb, περιέχει 21 κλώνους από τους οποίους 3 επιλέχθηκαν για αλληλούχηση και 16 STSs, που περισσότερα από τα μισά τοποθετούνται με βεβαιότητα. Τα στοιχεία αυτά σημειώνονται σε σειρά πάνω από τα STSs.

Μία ενδιαμέση εικόνα του φυσικού χάρτη όπως διαμορφώθηκε μετά την ολοκλήρωση της αντιστοίχισης κλώνων/STSs φαίνεται στην Εικ. 20 (α1, β1). Στην περιοχή 10q23.31-q25.3 χαρτογραφήθηκαν συνολικά 54 contigs με μήκος 140-800kb που αντιπροσώπευαν γενωμική αλληλουχία ~20Mb που αντιστοιχούσε τουλάχιστον στα 2/3 της περιοχής. Αντίστοιχα, η υποτελομερική περιοχή περιλάμβανε 12 contigs μήκους 150-770kb που συνολικά κάλυπταν φυσική απόσταση 4,25Mb. Συνολικά, τα contigs και των δύο περιοχών περιλάμβαναν 465 STSs τοποθετημένα με μεγάλη αξιοπιστία και 176 με αρκετή αξιοπιστία (Πιν. 9). Σ' αυτά περιλαμβάνονταν και STSs που είχαν προκύψει μετά από διαλογή ροής (flow sorting) των χρωμοσωμάτων, κατά την οποία τα χρωμοσώματα 9, 10, 11 και 12 συλλέγονται ως μία ομάδα. Στη συνέχεια μετά από τεμαχισμό και υποκλωνοποίησή τους, είχε προσδιοριστεί η αλληλουχία των ενθεμάτων, βάσει της οποίας σχεδιάστηκαν νέα STSs. Η χρωμοσωμική προέλευση αυτών των STSs καθορίστηκε με PCR σε μονοχρωμοσωματικά κυτταρικά υβρίδια της συλλογής Coriell Cell Repositories (panel 2). Η παραγωγή, η χαρτογράφηση και η διερεύνηση της βιβλιοθήκης με αυτούς τους ανιχνευτές πραγματοποιήθηκε στο Sanger Centre.

Σ' αυτήν την φάση κατασκευής του χάρτη, από τα αρχικά STSs που είχαν χρησιμοποιηθεί για την διερεύνηση ολόκληρης της βιβλιοθήκης, ποσοστό >60% είχαν τοποθετηθεί στα contigs με μεγάλη αξιοπιστία. Τα παραπάνω contigs έφεραν επιπλέον πολλά STSs που όμως περιείχονταν σε έναν μόνο κλώνο. Τα τελευταία αγνοούνταν στην περίπτωση που ο αντίστοιχος κλώνος ήταν εσωτερικός του contig γιατί κατά πάσα πιθανότητα προέρχονταν από ψευδώς θετικό σήμα κατά την υβριδοποίηση, είτε από ταυτοποίηση λάθος κλώνου, είτε από λανθασμένη τοποθέτηση του κλώνου. Όταν όμως επρόκειτο για ακραίο κλώνο εξετάζονταν η περίπτωση αλληλοεπικάλυψης του contig αυτού, με άλλα που περιείχαν το ίδιο STS με επανεξέταση των προτύπων τεμαχισμού.

Σ' αυτήν την φάση κατασκευής του φυσικού χάρτη είχαν επιλεγεί ~100 αντιπροσωπευτικοί κλώνοι για προσδιορισμό της αλληλουχίας τους, 83 και 18 από τις δύο περιοχές αντίστοιχα (Πιν. 9) που αντιστοιχούσαν σε αλληλουχία DNA μήκους 17,5Mb. Η επιλογή τους έγινε βάσει της αξιοπιστίας της θέσης τους, που προέκυπτε τόσο από το πρότυπο τεμαχισμού τους όσο και από τα STSs που περιείχαν. Επιπλέον, αυτοί οι κλώνοι επιλέχθηκαν έτσι ώστε να έχουν την μικρότερη δυνατή αλληλοεπικάλυψη (minimal tiling path). Η αλληλούχηση των κλώνων αυτών καθώς και όλων όσων επιλέχθηκαν στη συνέχεια, πραγματοποιήθηκε στο εργαστήριο αυτόματης αλληλούχησης DNA, Sanger Centre.



Εικ. 20: Χαρτογράφηση των contigs των περιοχών 10q23.31-q25.3 (A) και 10q26.2-qter (B) βάσει του περιεχομένου τους σε STSs. Στο ιδεόγραμμα του χρωμοσώματος, στο πάνω μέρος της εικόνας, σημειώνονται οι δύο περιοχές (A και B). Αναπαριστώνται τρία ενδιάμεσα (α1-α3 και β1-β3) και το τελικό στάδιο (α4, β4) κατασκευής του φυσικού χάρτη όπου φαίνεται η σταδιακή μείωση του αριθμού των contigs με την παράλληλη αύξηση του μήκους τους. Η διάταξή τους πραγματοποιήθηκε βάσει του γενετικού χάρτη Genethon, ενώ για όσα δεν περιείχαν γενετικούς δείκτες η τοποθέτησή τους βασίστηκε στην χαρτογράφηση των STSs τους σε άλλους χάρτες. Τα contigs συμβολίζονται με κόκκινες κάθετες γραμμές και η ονομασία τους περιλαμβάνει το πρόθεμα ctg ακολουθούμενο από έναν αριθμό.

Πιν. 9: Προκαταρκτικά contigs που χαρτογραφήθηκαν στις περιοχές 10q23.31-q25.3 (A) και 10q26.2-qtel (B)

A.

Κωδικός contig	Μέγεθος	STSs # 1	STSs # 2	Επιλεγμένοι κλώνοι για αλληλούχηση
ctg31	~370Kb	AFM269xf9, stSG226781, stSG15752, stSG32258, stSG45558, stSG72793	AFMb362yg5, WI-1936	ba399019
ctg1375	~730Kb	stc10-gata1, SHGC-35401, stSG72271, stA007J42, AFM029xh12, stN37073, stSG48047, SHGC-35401, stRH1014, stH10009, stSG39120, stSG47193, AFM287ze1, GATA66C04.P17780, stbA103A2SP6, stbA103A2T7, stbA15K3T7	stSG72472, ATA115E01.P34519	dJ156B16, ba30G13, dJ23G1
ctg949	~350Kb	stRH2, stN27530, stA006A33, stSG21190, stSG35189, stSG68316, stUT925 (=D10S520), stSG50462, stA008Z38, stN95450, stSG67985, stbA146P21SP6, stbA99P7T7	stSG67985	ba119K6
ctg15	~540Kb	WI-2054, stSG2761, stA005C16, stSG1032, stSG4306, AFMb354xe1, stSG70002, stSG2296, SHGC-9747, WI-3011	stSG21208	ba34E5, ba53P16, ba35J23
ctg1120	~160kb	stSG16199, AFM350wa5, stSG21522, stSG35096, stSG1671, stSG27214, SHGC-34978	stSG1953	ba34D15
ctg881	~360Kb	stA005X19, stSG3691, stSG71105, SHGC-32598, AFM205tg7, stSG38907, stSG68655, SHGC-2753, SHGC-5941, stSG50801, SHGC-17231	stSG21522, stSG27214, stSG35096	ba36D5, ba316M21
ctg35	~360Kb	stSG68573, AFMa132yd9, stW37897, A007K11, stSG68960, SHGC-30060	stRH627, st68525, stSG50275, WI-5905	ba411B6
ctg34	~800Kb	stSG71846, stSG42504, stAA034033, stSG34995, stSG38872, SHGC-31986, stRH75650, SHGC-4169, AFMa064za5, stSG31562, stSG47005, stSG71715, stSG47726, stSG42688, stSG50987, AFMa217yg5, stN39580, st1265/1266, stA005K38, stSG9207, stSG48239, stSG3501, stSG26135	stSG21180, WI-4865, stN31939, stSG42602, stSG70050, stSG68435, stSG70674	ba416N2, ba16H23, ba382M19, ba225H22, ba541N10
ctg1452	~370Kb	stSG22765, stA006D27, stSG47299, stSG48287, stSG30899, stU36223, WI-14063, stSG36084, stbA91A6T7	stSG46999, stSGC32623, SHGC-14701	ba229N4, ba91E4
ctg251	~440Kb	AFMa301xe1, AFMb001wb9, stSG51021, stSG51013, SHGC-4000, stSG71703, WI-15962, stR45921, stSG2605, stSG21225, stSG30235, SHGC-14679, AFM183xb12, stSG13279, stSG40669, stSG27142, stSGC33233, stSG9369, stSG72852	stA002G45, stSG2653, stSG68738, stSG69794	ba181I4, ba170J3, ba346A7
ctg1336	~345Kb	stSG916, stSG44115, WI-14791, stSG42868, stN37090, stSG26359, stSG31529, stbA40M1A, stbA101J13T7, stbA2A2SP6	AFM205tg7, stSG13162, stSG38907, stSG68655, SHGC-2753	ba127L20, ba99N20
ctg2086	~520Kb	stSG72862, AFMa053yg5, stSG48063, WI-15620, WI-1936	stSG16199, stRH816,	ba271113, ba451M19
ctg27	~500Kb	stSG29640, stSG46053, stN32898, stSG70045, stH68323, stSG72339, stbA5M14T7, stbA146P21SP6	stc10-gt4, stNIB1686, AFMa302wh1, stSG69752, stbA146P21SP6	ba121A3, ba76P200, ba162K11
ctg33	~380Kb	SHGC-34176, stA008I22, WI-8488, stSG72394, stN35618, SHGC-15696, stSG30988	stSG44588, stSG50801, stSGH8267	ba483F11, ba85A1
ctg893	~410Kb	stSG69764, stSGC31312, 18621/PrG-18622, WI-4379, stCda0pc11, stSG68907, stSG70478, stSG71740		ba446H13, ba86P17
ctg936	~360Kb	stNIB1357, stSG49291, stSG12737, stAA243510, stSG8954	stAA040321, stA005M36, stSG50884	ba313D6
ctg1160	~240Kb	SHGC-7202, WI-16821	stA006W20, stA009B31	ba148C9
ctg927	~530Kb	stSG399, stRH76567, AFM274xf9, WI-4436, stSG67568, WI-14968, WI-5976, stSG28889, WI-2469, AFM265yc9, stSG71828, stSG2378, LC.ATA1F04.P8234	T97004, stSG36024, WI-7467	ba25C19, ba431N8
ctg1182	~230Kb	stA004R39, WI-17128, stSG39758	R86942, WI-11615, Z40596	ba168O10
ctg1049	~240Kb	stA008Y13, stSG2721, stSG70345	AFM282yc1, stSG2551	ba57C13
ctg1427	~560Kb	stSG15491, stA004S47, stSG22466, stSG71918, AFMa348zg9, stSG71997, stSG8384, stNIB351, WI-16054, SHGC-12255	stSG12559, stSG31710, stSG35957, stSG44990, stH03262, stSG68001	ba106M14, ba251A15

Πιν. 9: (συνέχεια)

Κωδικός contig	Μέγεθος	STSs # 1	STSs #2	Επιλεγμένοι κλώνοι για αλληλούχηση
ctg2088	~720Kb	stSG30011, SHGC-1146, stSG51671, stSG70501, SHGC-170, stSG43195, AFMa244vg9, WI-14114, stSG45302, WI-16989, stN29485, stIB1099, stSG3745, stSG50791, WI-16274, stSG68723, stN22119, stZ40200, stSG68212	SHGC-11716, stSG72015, stSG472, stSG70600	bA30E16, bA437J2, bA397J18, bA562L21
ctg894	~300Kb	H92921	WI-16578	bA425M17
ctg1320	~280Kb	L07093, stSG15866, stSG21346	stSGC33162, AFM330va5	bA400G3
ctg987	~250Kb	F09841, stSG16023, stSG41319, stSG41741, stSG68511, stZ40596, WI-11734	stSG68080	bA80H5
ctg1370	~210Kb	stSG2699, stSG48141, WI-6938	stA002G45, stSG2653, stSG51013, stSG68738, stM14564, AFMa343yg5	bA156M10
ctg1100	~140Kb	SHGC-32394, WI-9639		bA19C6
ctg1123	~400Kb	stSG46524, stSG72293, WI-4209	stSG15181, stSG26670, WI-9985, stSG72469, stSG71164	bA34A14, bA439D8
ctg1316	~210Kb	AA235311	stSG32285, stSG71013	bA431M23
ctg8	~580Kb	GATA64b01.P17619, stSG4193, WI-9147, stSG68457, SHGC-32265, WI-16299, stSG69104, AFM206yb2, stSG50971, A006J02, stSGC32923, stSG27406, stSG48072, WI-19473, stSG42773, stSG67769, stSG43839, A004R30	stSG27142, WI-6372, stSG47056, stSG46535, stRH76747	bA108L7, bA324L3, bA179B2, bA107114.
ctg986	~260Kb	stSG3345, stSG40082, stSG47144, stSG3084	stSG68214, WI-22164, stSG50235	bA30H12
ctg891	~400Kb	stSG4995, stSG49129	stAA243510, stSG8954, stSG12737, stSG16259, stSG49291, stSGC35716	bA158G3, bA81E5
ctg1083	~330Kb	A008M23	stA008Y13, stSG67891, stSG2551, stSG2721, stSG70436, SGC33428, stSG21644, stSG26026	bA34A22
ctg1195	~300Kb	stSG71719, stSG35992, stSGC30780, stSG72314, st1283/1284, stA004O25, stAA022645, stSG68110	stSG71098, AFMb361wb9	bA429G19
ctg17	~320Kb	stSG76769, stSG2649, A004Q04, stSG29677	WI-17804, stSG67802	bA389E6, bA140G2
ctg1	~550Kb	stSG68999, AFMa051we5, stSG70882		bA219J23
ctg1330	~390Kb	stN32594, stSG42986, stT87905, stA004M41	stSG35058, stSG68795, SHGC-8483	bA203A7, bA106M7
ctg1521	~290Kb	CHLC.ATC3.P10215, stSG68859	stAA284841	bA73H14
ctg1557	~570Kb	AFMa071xh1, GGAA22D12.P17515, WI-4434, stbA166O7T7, stc10-gata4, SHGC-8449	AFM331xa9, SHGC-10882, stSG69022, stSG72183	bA469M11, bA328K1, dJ317G21
ctg10	~520Kb	stSG70628, stSG41412, AFM240yc7, stSG43429, stSG8778, stSG71116, stSG15535, stSG45467, stSG69054, stSG36017, stSG71701	stRH40293, WI-16878, AFMa272zd1, stSG67552, stSG70243	bA411P18, bA317F20
ctg20	~260Kb	stRH40293, WI-16878, stA006A44, AFMb361wb9		bA393E20
ctg91	~380Kb	A006H32, AFM282yc1 stAA021529, stSG3666, stSG31703, stSG1447, AFM337xt9, stR42168, stSG47123	stSG50381, stA009C28	bA165M8, bA77F13
ctg210	~430Kb	stSG2551, stSG70436, stSG9043, stSG26026, stSG67891, stSG3472, stSG21644, stSG46580, A009V47, stSG49624, A008Y13, WI9874, stSGC33428, stSG70643, stSG49433	stSG71805, stSG69617, stbA2A2SP6, stbA101J13177, stbA40M1A	bA119F19
ctg1132	~240Kb	W73720		
ctg913	~520Kb	stSG8438, stSG12880, WI-4264	stSG70841, stbA15K3T7	bA7F3
ctg1589	~175Kb	AFM289zh5, stSG70637, AFM292vg1, stH90420, stL16499, stU37426		
ctg1379	~220Kb	stA006X22, stSG68762, SHGC-30793	AFM079ya1	
ctg57	~460Kb	stSG70363, stSG25928, stSG73477, stSG9628	stbA166O7T7, stSG70780	bA152G7
ctg1206	~200Kb	WI-6938, AFMb359zg5, stSG12760, stSG41841, stA006S48, SHGC-9157, stSG72253		
ctg233	~320Kb	stSG22723, stSG58001	AFMb343zc1, stbA103A2SP6	bA87E14
ctg515	~410Kb	stSG68867, W86789, H95763, stSG67633, SHGC-15732	stRH190, stSG45548, WI-17750, stSG22312, stSG47335, SHGC-31032, AFM269xg9, stSG67673	
ctg1221	~300Kb	WI-15944, stSG35355, WI-13377, WI-5915, stSG69060, stSG72349, stSG22607, SHGC-17121, stSG72269	CHLC.GAAT4C01.P8513, A009V05, stSG9987, stSG50275	bA175O19
ctg1029	~190Kb	stSG1671, stSG1953, stSG16199, WI-5852		

Πιν. 9: (συνέχεια)

Κωδικός contig	Μέγεθος	STSs # 1	STSs # 2	Επιλεγμένοι κλώνοι για αλληλούχηση
ctg1111	~280Kb	stSGC32623, stA004B36, W69567, WI-15144, WI-16946, stSG72304, stbA91A6T7		bA68M5
ctg1498	~170Kb	N20191, stSG71780, WI-9639, stSG72889	stT85373, SHGC-32394	

B.

Κωδικός contig	Μέγεθος	STSs # 1	STSs # 2	Επιλεγμένοι κλώνοι για αλληλούχηση
ctg2083	~ 480 Kb	WI-3219, A008Y06, stSG1566, stSG8192, stSGC33916, stbA31A20A, stSG69976, stSG68220, stSG70306, stSG48184, stSG71267, D10S217, stSG69706, stbA48A2B, stbA48A2A	stSG68773, WI-4342, stSG72409	bA380J17, bA31A20
ctg497	~150 Kb	stSG71885, WI-2006, AA010383, stSG2750, W67510, WI-22094.	stSG68171, stSG70234	bA45A17
ctg903	~ 210 Kb	stSG4790, stSG43417, stSG45453, st10QTE124, stSG1303, stSG27915, SHGC-10984, WI-14945, stSG70571		bA108K14
ctg910	~770 Kb	stSG36110, D10S1201, WI-3778, stSG47386, WI-6793, H99205, stSG41412, D10S505, W15548, AA036862, stSG68341, stbA92A10SP6, stbA165L12T7, stSG72104, stSG4213,	stSG69123, stSG67900	bA500G10, bA339B9, bA400F15, bA435P9, bA218C11, bA109A6
ctg920	~220Kb	D10S169, stSG67569, stSG47732	stSG72427, stSG47909, stbA58L14A	bA113P9
ctg943	~750Kb	D10S1655, SHGC-7516, A007H40, stSG30731, WI-4197, stbA168C9T7, GGAT14G01.P33618, WI-2391	GGAA23C05.P17625, stSG71832, stSG67835	bA343D22, bA396A24, bA231D11
ctg1052	~350Kb	stSG67746, D10S1676, stSG70252, stbA48A2B	stSG70252, stSG1566, stSG33916, WI-2469, stSG71793, WI-2469, WI-7467	
ctg1067	~250 Kb	R96892		bA8F15
ctg1078	~ 380 Kb	WI-20848, stSG17460 (D10S1727), stSG1460, D10S1782, stSG32124	WI-4342	bA82L9, bA21M8
ctg1230	~280 Kb	SHGC-2731, WI-17432, stSG68986, stSG71743, stSG71743,	RH1112	bA48E18
ctg1332	~240Kb	WI-19193, AFM198zb4	AFMb027wc9, stSG71217, stSG46145	
ctg1595	~170Kb	stSG2795, stSG41791	stSG12596	

Σημ.: Στην πρώτη στήλη αναγράφεται το όνομα του contig και στη δεύτερη το μήκος του. Οι επόμενες δύο στήλες περιλαμβάνουν τα STSs που έχουν τοποθετηθεί με μεγάλη (# 1) ή αρκετή (# 2) βεβαιότητα αντίστοιχα. Τέλος, παρατίθενται οι αντιπροσωπευτικοί κλώνοι που είχαν επιλεγεί από κάθε contig για αλληλούχηση, σε αυτήν την φάση του φυσικού χάρτη. Το πρόθεμα "bA" αντιστοιχεί σε BACs ενώ το "dJ" σε PACs που είχαν απομονωθεί παλαιότερα (Θ. Σαραφίδου, Μεταπτυχιακός τίτλος ειδίκευσης, Μ. Κοκκινάκη, Διδακτορική Διατριβή).

β. Χρωμοσωμικό περπάτημα

Καθώς η κατανομή των STSs κατά μήκος του χρωμοσώματος, δεν είναι ομοιόμορφη, ήταν αναμενόμενο ότι μετά την διερεύνηση της γενωμικής βιβλιοθήκης και τον σχηματισμό των προκαταρκτικών contigs, θα παρέμεναν αρκετά χάσματα στον φυσικό χάρτη. Η επέκταση και η ενοποίηση των contigs έγινε με την δημιουργία νέων δεικτών από τα άκρα των εξωτερικών κλώνων των contigs. Οι νέοι αυτοί δείκτες είτε ταυτοποιούσαν υπάρχουσες αλληλοεπικαλύψεις μεταξύ γειτονικών contigs είτε χρησιμοποιήθηκαν για την απομόνωση επιπλέον κλώνων από την χρωμοσωμοειδική και από άλλες γενωμικές βιβλιοθήκες (RPC-11, RPC-13).

Οι αλληλουχίες των άκρων των BACs υπάρχουν καταχωρημένες στη βάση δεδομένων Genbank και η διερεύνησή τους γίνεται με τον χαρακτηριστικό κωδικό αριθμό του κλώνου (π.χ. 103A2) μετά το πρόθεμα RPCI-11-. Για τους κλώνους που χαρτογραφούνταν στα άκρα των προκαταρκτικών contigs και για τους οποίους υπήρχαν διαθέσιμες τέτοιες αλληλουχίες, σχεδιάστηκαν εκκινητές, παρασκευάστηκαν

οι αντίστοιχοι ανιχνευτές (βλ. Μεθόδους) και χρησιμοποιήθηκαν στην διερεύνηση της χρωμοσωμοειδικής βιβλιοθήκης. Συνολικά, σχεδιάστηκαν 96 ζεύγη εκκινητών και απομονώθηκαν 78 προϊόντα PCR. Από αυτά, τα 58 αναγνώριζαν εκτός του πατρικού και επιπρόσθετους κλώνους (Πιν. 10). Επιπλέον, αρκετά προϊόντα PCR ταυτοποιούσαν κλώνους που προέρχονταν από κάποιο άλλο contig, γεγονός που αποτελούσε ένδειξη αλληλοεπικάλυψης. Σ' αυτές τις περιπτώσεις επανεξετάζονταν τα πρότυπα τεμαχισμού των ακραίων κλώνων των αντίστοιχων contigs. Αυτά τα αποτελέσματα, σε συνδυασμό με νέα πρότυπα τεμαχισμού κλώνων, οδήγησαν στην επέκταση και ενοποίηση αρκετών contigs. Ενδεικτικά, στον Πιν.10 φαίνονται τα STSs stbA1P12SP6, stbA1P12T7, stbA133L23SP6 και stbA260L13T7 που αναγνωρίζουν κλώνους προερχόμενους από τα ctg2086 και ctg1557 που δεν είχαν προηγουμένως κάποιο κοινό STS. Η επαναξέταση των προτύπων τεμαχισμού των ακραίων κλώνων τους βάσει του νέου δεδομένου, οδήγησε στην ενοποίησή τους και στην παραγωγή ενός contig μήκους 1,27Mb. Πλέον, στην περιοχή 10q23.31-q25.3 ο χάρτης περιλάμβανε 39 contigs από τα οποία τα 7 είχαν μήκος >1 Mb ενώ το μεγαλύτερο ήταν μήκους 2,24Mb. Αντίστοιχα, στην υποτελομερική περιοχή τα contigs μειώθηκαν στα 9, το μεγαλύτερο είχε μήκος 1,2Mb (Εικ. 20, α2 και β2) και περιείχε τους μικροδορυφορικούς δείκτες D10S1782 και D10S1676 που στον γενετικό χάρτη απέχουν 4cM. Το γεγονός αυτό επιβεβαίωνε την εκτίμηση ότι το φυσικό μήκος της υποτελομερικής περιοχής ήταν κατά πολύ μικρότερο από όσο υποδήλωνε ο γενετικός χάρτης, λόγω υψηλού ποσοστού ανασυνδυασμών.

Σε αυτήν την φάση εκτιμήθηκε ότι η κάλυψη των χασμάτων δεν απαιτούσε μεγάλο πλήθος επιπλέον κλώνων. Η υπόθεση αυτή στηρίχτηκε στην ολική γενωμική κάλυψη των δύο περιοχών με τα επιμέρους contigs (26Mb και 4,4Mb αντίστοιχα) καθώς και στην διάταξή τους κατά μήκος του χρωμοσώματος, όπως καθορίστηκε βάσει των περιεχόμενων δεικτών. Η προσέγγιση που επιλέχθηκε στη συνέχεια, αφενός λόγω του μικρού μεγέθους των χασμάτων αφετέρου λόγω χρονικών περιορισμών στην επιλογή των κλώνων προς αλληλούχηση, εφαρμόστηκε για όλο το χρωμόσωμα από την ομάδα του Sanger Centre και περιλάμβανε την σύγκριση του προτύπου τεμαχισμού των ακραίων κλώνων των επιμέρους contigs με αυτά των κλώνων όλου του γονιδιώματος είτε με τη συνήθη είτε με λίγο χαμηλότερη αυστηρότητα. Μ' αυτόν τον τρόπο αυξήθηκε η κάλυψη, κυρίως χάρη σε κλώνους που προέρχονταν από το δεύτερο μισό της βιβλιοθήκης BAC (segment 2, plates 289-576) το οποίο δεν είχε χρησιμοποιηθεί στη δική μας διερεύνηση. Έτσι ο αριθμός των contigs στην περιοχή 10q23.31-q25.3 μειώθηκε σε 10, ενώ το μεγαλύτερο (που προέκυψε από την ενοποίηση των ctg91, 29, 987, 894, 53, 52, 60, 57, 1160, 515, 15 και 1195) έφτασε σε μήκος τα 9,2Mb (Εικ. 20, α3). Στην υποτελομερική περιοχή το

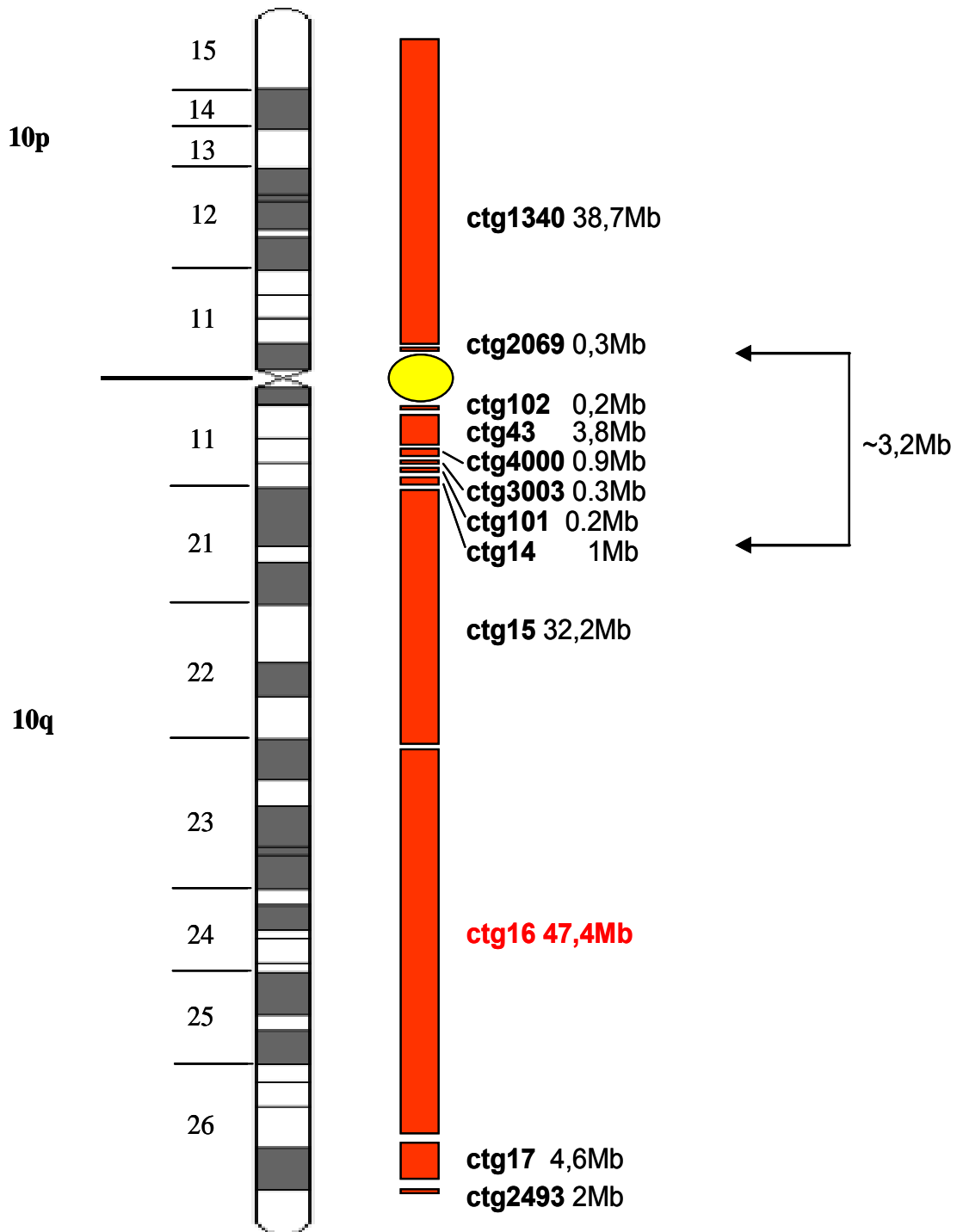
κύριο πλέον contig (που προέκυψε από την συνένωση των ctg1230, ctg1078, ctg943 και ctg910) είχε μήκος 3,6Mb και αντιστοιχούσε σε γενετική απόσταση 13,5cM (Εικ. 20, β3). Επιπλέον από τα άκρα των νέων κλώνων που επιλέχθηκαν σχεδιάστηκαν καινούρια STSs για την επαναδιερεύνηση της βιβλιοθήκης και το κλείσιμο και των τελικών χασμάτων. Το τελικό αποτέλεσμα για την περιοχή 10q23.31-q25.3 ήταν η κάλυψή της από ένα ενιαίο contig, που εκτείνεται από το q22.3 ως και το q26.2, μεγέθους 47,4Mb. Η φυσική απόσταση μεταξύ των ακραίων δεικτών που καθόριζαν τα όρια της περιοχής που μελετήθηκε ήταν ~28Mb. Όσον αφορά την υποτελομερική περιοχή, ένα μικρό τμήμα της (<200kb) περιλαμβάνεται στο παραπάνω contig, ενώ η υπόλοιπη καλύπτεται από δύο contigs μεγέθους 4,6Mb και 2Mb αντίστοιχα (Εικ. 20, α4 και β4). Το μέγεθος των δύο χασμάτων, όπως εκτιμήθηκε με ανάλυση FISH σε εκτεταμένα ινίδια DNA (fiber FISH) με ανιχνευτές τους γενωμικούς κλώνους εκατέρωθεν κάθε χασματος, είναι 50kb και 10kb αντίστοιχα (Deloukas P *et al*, 2004).

Ο φυσικός χάρτης, όπως διαμορφώθηκε τελικά (Εικ. 21), υπάρχει διαθέσιμος στο διαδίκτυο, όπου απεικονίζονται οι κλώνοι, οι σχετικές τους αλληλοεπικαλύψεις, το tiling path, τα STSs με σχετικές συνδέσεις σε περαιτέρω πληροφορίες και επιπλέον ιστοσελίδες (Εικ. 22).

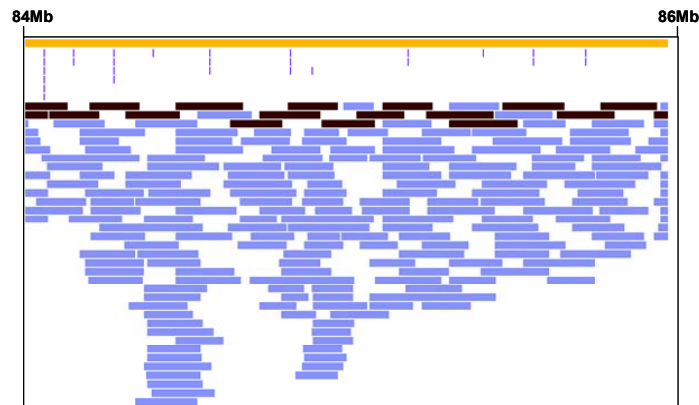
Πιν. 10: Νέα STSs από ακραίες αλληλουχίες κλώνων BAC από τις περιοχές 10q23.31-q25.3 (A) και q26.2-qter (B).

STS		contig	STS		contig
stbA1P12SP6	(stSG80453)	ctg2086, ctg1557	stbA103A2T7	(stSG79811)	ctg1375
stbA1P12T7	(stSG80452)	ctg2086, ctg1557	stbA166O7T7	(stSG79834)	ctg1557, ctg57
stbA2A2SP6	(stSG80371)	ctg1336	stbA146P21SP6	(stSG79823)	ctg27
stbA7O22SP6	(stSG80455)	ctg1330	stbA5M14T7	(stSG79855)	ctg27, ctg515
stbA7O22T7	(stSG80456)	ctg1330	stbA108L7SP6	(stSG79819)	ctg8
stbA20D13SP6	(stSG80459)	ctg33	stbA196I15T7	(stSG80426)	ctg20
stbA24H9SP6	(stSG81050)	ctg1007	stbA142B2SP6	(stSG80420)	ctg20
stbA40M1A	(stSG80357)	ctg1336			
stbA56D2A	(stSG80363)	ctg891	stbA31A20A	(stSG80358)	ctg2083
stbA56N17A	(stSG80362)	ctg891	stbA48A2A	(stSG80365)	ctg2083
stbA91A6T7	(stSG80372)	ctg1452	stbA48A2B	(stSG80366)	ctg2083
stbA99P7SP6	(stSG81048)	ctg949, ctg515	stbA67N15A	(stSG81047)	ctg910
stbA99P7T7	(stSG80388)	ctg949, ctg515	stbA90O13SP6	(stSG80373)	ctg910
stbA101J13T7	(stSG80380)	ctg1336	stbA142I9SP6	(stSG80423)	ctg2083
stbA110D19SP6	(stSG80396)	ctg38, ctg7	stbA142I9T7	(stSG80424)	ctg2083
stbA111D21T7	(stSG80390)	ctg38, ctg7	stbA67N15A	(stSG81047)	ctg910
stbA114F7T7	(stSG80403)	ctg33	stbA168C9T7	(stSG79843)	ctg943
stbA115J15T7	(stSG80393)	ctg31	stbA92A10SP6	(stSG79814)	ctg910
stbA133L23SP6	(stSG80397)	ctg2086, ctg1557	stbA165L12T7	(stSG79840)	ctg910
stbA165O6T7	(stSG80413)	ctg1007	stbA7J13T7		903
stbA170J3SP6	(stSG80425)	ctg251	stbA7J13SP6		903
stbA179B2SP6	(stSG80427)	ctg8	stbA119L6T7		1078
stbA196K8SP6	(stSG81049)	ctg35, ctg1120	stbA21M8T7		1078
stbA260L13T7	(stSG80429)	ctg2086, ctg1557	stbA165N2T7		910
stbA15K3T7	(stSG79861)	ctg1375, ctg913	stbA343J20T7		ctg943, ctg1078
stbA103A2SP6	(stSG79810)	ctg1375			

Σημ.: Το όνομα κάθε νέου STS που δημιουργήθηκε από τα άκρα των BACs, φέρει το πρόθεμα stbA, ακολουθούμενο από το όνομα του BAC από το οποίο προέρχεται (π.χ. 1P12) και την ένδειξη SP6 ή T7, που αναφέρεται στα δύο άκρα κάθε κλώνου. Στις παρενθέσεις αναγράφονται τα εναλλακτικά ονόματά τους (κωδικός Sanger Centre). Στην δεύτερη στήλη φαίνονται τα contigs στα οποία το συγκεκριμένο STS τοποθετείται με μεγάλη βεβαιότητα. Όλα τα STSs που παρατίθενται αναγνώριζαν τον πατρικό τους κλώνο. Στις περιπτώσεις που στη δεύτερη στήλη αναγράφονται περισσότερα του ενός contig, επανεξετάζονταν τα πρότυπα τεμαχισμού των αντίστοιχων ακραίων κλώνων τους για ταυτοποίηση αλληλοεπικάλυψης.



Εικ. 21: Ο φυσικός χάρτης του χρωμοσώματος 10. Ο χάρτης αποτελείται από 12 contigs (έγχρωμες κάθετες γραμμές) που καλύπτουν συνολικά 136,6Mb και αντιστοιχούν στο 99,4% της ευχρωματίνης. Το ελάχιστο επικαλυπτόμενο μονοπάτι συνίσταται από 1144 κλώνους. Η περιοχή q23.3-q25.3 καλύπτεται από ένα ενιαίο contig, το ctg16, μήκους 47,4Mb και η υποτελομερική από δύο, συνολικού μήκους 6,6Mb. (Bentley DR *et al*, 2001).



Εικ. 22: Ο φυσικός χάρτης του χρωμοσώματος 10 στην περιοχή 84-86Mb. Ενδεικτικά απεικονίζονται 2Mb του χάρτη. Φαίνονται όλοι οι αλληλοεπικαλυπτόμενοι κλώνοι (οριζόντιες γραμμές) και σημειώνονται αυτοί (μαύροι) που συνιστούν το ελάχιστο επικαλυπτόμενο μονοπάτι (tiling path). Είναι εμφανές το μεγάλο βάθος της κάλυψης του γονιδιώματος. Επίσης παριστάνονται τα STSs (οριζόντιες γραμμές, πάνω) που πλέον η τοποθέτησή τους βασίζεται στην αλληλουχία των κλώνων. Όλα τα στοιχεία είναι επιλέξιμα και ο χάρτης είναι διαθέσιμος στο διαδίκτυο (<http://www.sanger.ac.uk/cgi-bin/humace/fpcwebmap.cgi>).

3.2. Η αξιοποίηση του χρωμοσωμικού χάρτη 10q23.3 για την ταυτοποίηση και ανάλυση γονιδίων που σχετίζονται με παθολογικούς φαινότυπους

Ένα από τα πιο σημαντικά οφέλη που προέκυψε από την κατασκευή χρωμοσωμικών χαρτών υψηλής ευκρίνειας και την αλληλούχηση του ανθρώπινου γονιδιώματος, είναι η δραματική επιτάχυνση του χρόνου που απαιτείται για τη κλωνοποίηση γονιδίων που σχετίζονται με κληρονομικά νοσήματα, με βάση την τοπογραφία του στα χρωμοσώματα (positional cloning). Η αρχική διαδικασία της επιλογής υποψηφίων γονιδίων, γίνεται ακόμη πιο αποτελεσματική, στις περιπτώσεις όπου ο παθολογικός φαινότυπος έχει ήδη συσχετισθεί με μια συγκεκριμένη χρωμοσωμική περιοχή, με γενετική ανάλυση σύνδεσης. Αξιοποιώντας το χρωμοσωμικό χάρτη και τα «μοριακά εργαλεία» που δημιουργήσαμε για τη περιοχή 10q23.3, εστιαστήκαμε στη μοριακή ανάλυση δύο υποπεριοχών που συνδέονται: 1) με την αυτοσωμική επικρατή μερική επιληψία πλάγιου κροταφικού λοβού (Autosomal Dominant Lateral Temporal Epilepsy, ADLTE) και 2) με την εύθραυστη χρωμοσωμική θέση FRA10A.

A. Ταυτοποίηση και μοριακή ανάλυση γονιδίων που χαρτογραφούνται στην κρίσιμη περιοχή της ADLTE

Τον Φεβρουάριο του 2001 με πρωτοβουλία της δικής μας ομάδας και της ομάδας του Dr. C.Nobile (CNR-Institute of Neurosciences, Section of Padua, Italy), δημιουργήθηκε το «European Collaborative Consortium for the Study of Autosomal Dominant Lateral Temporal Epilepsy» με σκοπό τη συλλογική ανάλυση υποψηφίων γονιδίων της περιοχής και τον προσδιορισμό του υπεύθυνου γονιδίου για την ADLTE. Εκτός των ανωτέρω ομάδων, το consortium αποτελείται από τις ομάδες των

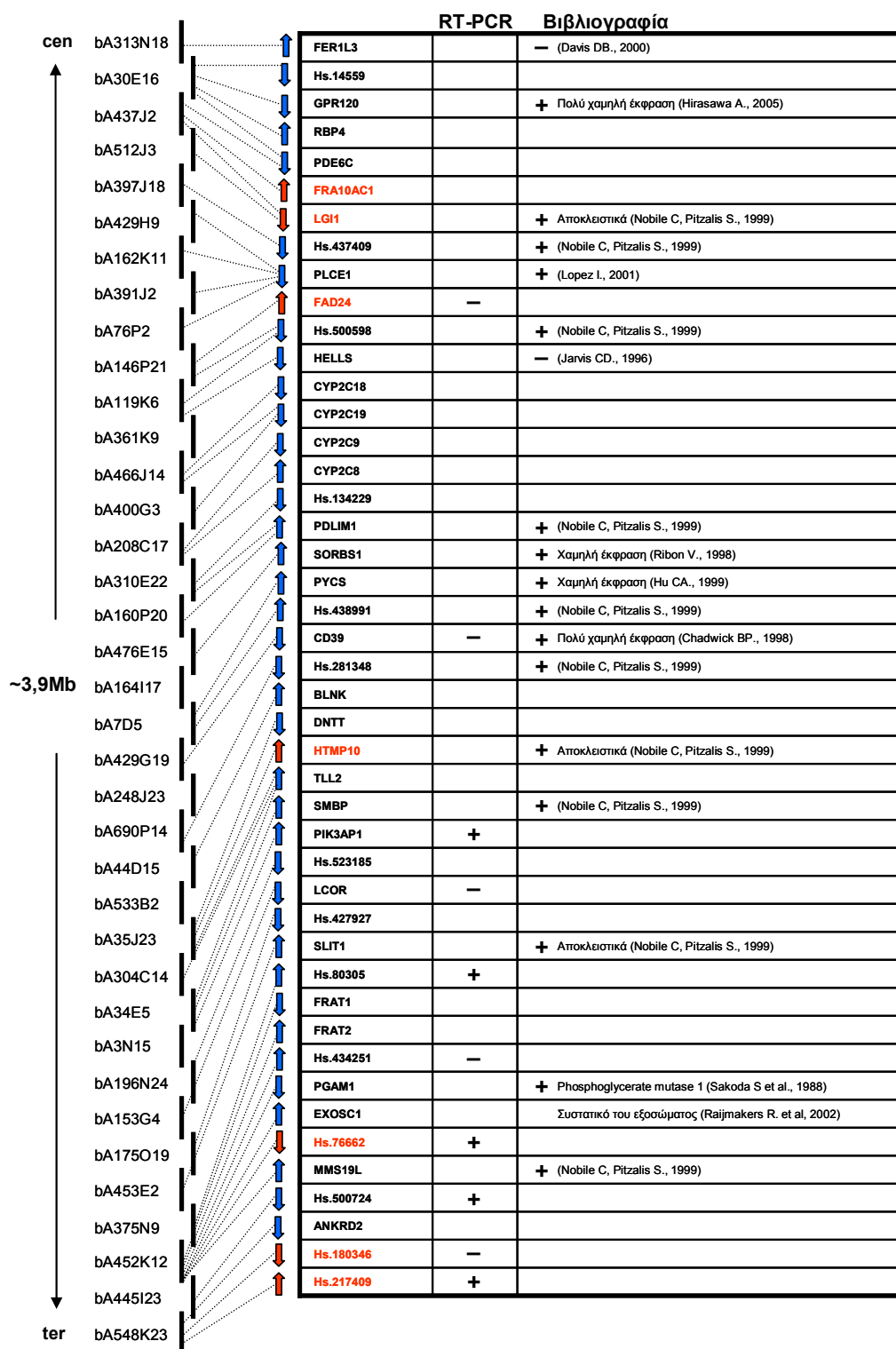
Dr. J. Perez-Tur (Institut de Biomedicina de Valencia-CSIC, Valencia, Spain), Dr. R. Siebert (Institute of Human Genetics, University Hospital Schleswig-Holstein Campus Kiel, Kiel, Germany), Dr. P. Deloukas (The Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, UK) και αρκετά εργαστήρια κλινικής νευροβιολογίας.

Η προσέγγιση που ακολουθήθηκε ήταν αυτή του υποψήφιου γονιδίου και στηρίχτηκε σε γονιδιακό χάρτη υψηλής ευκρίνειας που κατασκευάσαμε αξιοποιώντας όλα τα διαθέσιμα εργαλεία για το χαρακτηρισμό των γενωμικών αλληλουχιών *in silico* (βλ. Μεθόδους). Από την ανάλυση αυτήν προέκυψε ο γονιδιακός χάρτης της κρίσιμης περιοχής (Εικ. 23). Η περιοχή καλύπτεται από 37 κλώνους BAC, έχει μήκος ~3,9Mb και περιλαμβάνει 46 γονίδια. Ο χάρτης αυτός αντιστοιχεί στη σημερινή μορφή του, όπου έχει ολοκληρωθεί η αλληλούχηση των σχετικών γενωμικών κλώνων και είναι πλέον διαθέσιμα στις βάσεις δεδομένων πλήθος, ως επί το πλείστον πλήρους μήκους, cDNAs. Όταν άρχισε η αναζήτηση του υπεύθυνου γονιδίου για την ADLTE, μόλις είχε ολοκληρωθεί η πρώτη έκδοση της αλληλουχίας του χρωμοσώματος 10 που περιλάμβανε πολλά χάσματα σε επίπεδο αλληλουχίας, είτε μεταξύ διαδοχικών κλώνων είτε μέσα στον ίδιο γενωμικό κλώνο BAC. Επιπλέον, οι διαθέσιμοι κλώνοι cDNA πλήρους μήκους ήταν περιορισμένοι καθώς τα κύρια ερευνητικά προγράμματα απομόνωσής τους δεν είχαν ξεκινήσει ακόμα ή ήταν στα πρώτα στάδια υλοποίησης. Ενδεικτικά, στην αρχή του 2002 υπήρχαν κατατεθειμένες 2000 τέτοιες αλληλουχίες από το πρόγραμμα Kazusa (<http://www.kazusa.or.jp/huge>) και οι καταχωρήσεις από την MGC (Mammalian Gene Collection, <http://mgc.nci.nih.gov/>) ήταν λιγότερες από 700 (Φεβρουάριος 2001). Επίσης, η γονιδιακή ανάλυση (annotation) της γενωμικής αλληλουχίας χιζόταν σταδιακά. Καθώς η ανάλυση αυτή ήταν αρχικά πλήρως αυτοματοποιημένη είχε εγγενείς περιορισμούς, δηλαδή παρουσίαζε ψευδώς θετικές και ψευδώς αρνητικές προβλέψεις για εξόνια. Για όλους τους παραπάνω λόγους, σε πολλές περιπτώσεις, ήταν απαραίτητη η στοίχιση πολλαπλών ESTs ώστε να προκύψει το πλήρους μήκους cDNA για κάθε προβλεπόμενο γονίδιο. Όμως, επειδή η πιστότητα της αλληλουχίας των ESTs δεν είναι άριστη, ήταν απαραίτητη η προβολή του cDNA στην αντίστοιχη γενωμική αλληλουχία και ο σωστός καθορισμός των ορίων ιντρονίων-εξονίων, ώστε να χρησιμοποιηθεί η γενωμική πλέον αλληλουχία για τον καθορισμό του ανοικτού πλαισίου ανάγνωσης και την πρωτοταγή δομή του αντίστοιχου πολυπεπτιδίου. Στη συνέχεια γινόταν η πρόβλεψη της πιθανής λειτουργίας των μορίων αυτών βάσει δομικών ομολογιών και με χρήση υπολογιστικών προγραμμάτων πρόβλεψης δομής πρωτεϊνών (βλ. Μεθόδους).

Αρχικά, με βάση τον γονιδιακό χάρτη, έγινε διερεύνηση της κρίσιμης χρωμοσωμικής περιοχής *in silico* για τον εντοπισμό γονιδίων με βάση τα εξής

κριτήρια: α) γονίδια που κωδικοποιούν κανάλια ιόντων, β) υποδοχείς νευροδιαβιβαστών, γ) γονίδια που παρουσιάζουν ομολογία με τις κατηγορίες (α) ή (β), καθώς μέχρι τότε οι μεταλλαγές που είχαν εντοπιστεί σε ιδιοπαθείς μορφές επιληψίας εντοπιζόνταν σε τέτοιας δομής γονίδια. Η διερεύνηση έδειξε ότι στην περιοχή αυτή δεν χαρτογραφείτο κάποιο γονίδιο με τις ανωτέρω προδιαγραφές που θα ήταν προφανώς υποψήφιο. Στη συνέχεια αξιοποιήθηκε η γνωστή ή η συναγόμενη λειτουργία των γονιδίων της περιοχής, για όσα αυτό ήταν δυνατόν, ώστε να επιλεγούν τα γονίδια που θα μπορούσαν να συσχετιστούν με τον παθολογικό φαινότυπο. Στη δεύτερη φάση διερεύνησης αξιολογήθηκαν γονίδια που, με προγράμματα πρόβλεψης δομής, εμφανίζονταν να κωδικοποιούν πολυπεπτίδια με πολλές διαμεμβρανικές περιοχές ικανά να θεωρηθούν υποψήφιοι κυτταρικοί υποδοχείς. Επιπλέον, κάποια γονίδια χαρακτηρίστηκαν ως υποψήφια βάσει του προτύπου έκφρασής τους εστιάζοντας σε γονίδια που εκφράζονται στον εγκέφαλο, έστω και όχι κατ' αποκλειστικότητα. Γονίδια με περιορισμένη έκφραση στο νευρικό σύστημα είναι πιθανόν να έχουν συμμετοχή σε θεμελιώδεις νευρικές διαδικασίες και πιθανόν οι μεταλλαγές τους να σχετίζονται με την εκδήλωση νευρολογικών διαταραχών. Ενδείξεις για το πρότυπο έκφρασης ενός γονιδίου αντλήθηκαν από τη βάση δεδομένων Unigene (expression profile db), η οποία παρέχει μία χονδρική εκτίμηση των επιπέδων έκφρασης ενός γονιδίου σε διάφορους ιστούς. Θα πρέπει να σημειωθεί ότι η μελέτη αυτή είναι μεροληπτική καθώς αρκετά ESTs προέρχονται από κανονικοποιημένες βιβλιοθήκες cDNA. Επιπρόσθετα και μετά την κατασκευή του γονιδιακού χάρτη, ελέγχθηκε η έκφραση αρκετών γονιδίων της περιοχής, στον εγκέφαλο με RT-PCR. Η αντίστροφη μεταγραφή έγινε σε ολικό RNA από εγκέφαλο εμβρύου ανθρώπου με εκκινητή oligo-dT και η PCR με τους εκκινητές των αντίστοιχων ESTs (Πιν. 5). Η δοκιμή πραγματοποιήθηκε για 10 ESTs που πληρούσαν τα ανωτέρω κριτήρια. Ειδικότερα, για όλα υπήρχαν πληροφορίες (expression profile db) ότι εκφράζονται σε εγκέφαλο ενώ για 6 από αυτά ότι τα προϊόντα τους περιλαμβάνουν διαμεμβρανικές περιοχές (πρόγραμμα TMpred). Τα 5 από τα ESTs που εξετάστηκαν έδωσαν θετικό σήμα (Εικ. 23) και θεωρήθηκαν πιθανά υποψήφια.

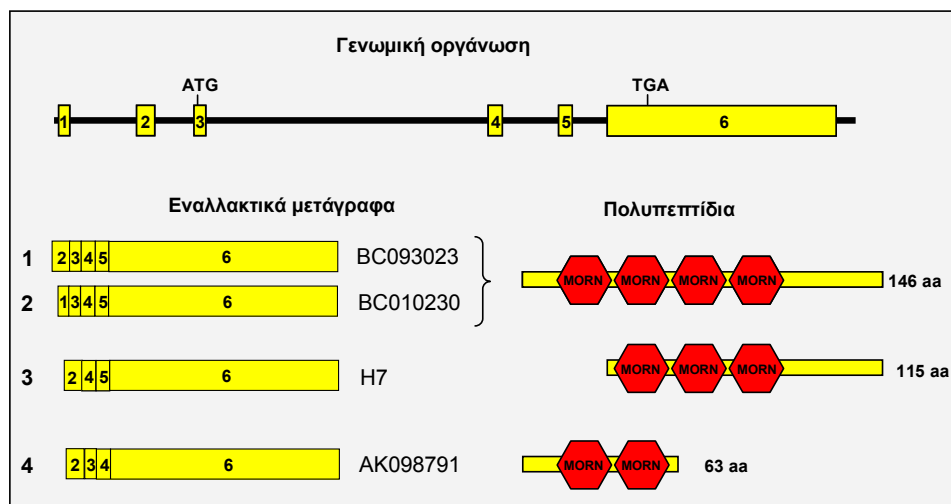
Για τα γονίδια που θεωρήθηκαν υποψήφια και βάσει της γενωμικής τους οργάνωσης, εξετάστηκαν τα εξόνια και τα σημεία συναρμογής ιντρονίων-εξονίων για τον εντοπισμό μεταλλαγών. Η μελέτη αρχικά έγινε σε DNA ενός ασθενούς από καθεμία από τρεις οικογένειες στις οποίες κληρονομείται με μεντελικό τρόπο η ADLTE. Στη συνέχεια παρουσιάζεται η μοριακή ανάλυση τεσσάρων υποψηφίων γονιδίων που αποδείχθηκε ότι δεν σχετίζονται με την ADLTE και τέλος η μοριακή ανάλυση του *LG11*, του γονιδίου που ευθύνεται για το εν λόγω νόσημα.



Εικ. 23: Γονιδιακός χάρτης της κρίσιμης περιοχής της ADLTE. Η περιοχή περιλαμβάνει 37 κλώνους BAC (κάθετες μαύρες γραμμές) που περιέχουν 46 γονίδια (βέλη). Για κάθε γονίδιο συμβολίζεται η φορά της μεταγραφής του (προσανατολισμός του βέλους) και ο/οι κλώνοι που αντιστοιχεί (σύνδεση κλώνων/γονιδίων με διακεκομμένες γραμμές). Όσα γονίδια έχουν τουλάχιστον μία βιβλιογραφική αναφορά, συμβολίζονται με το όνομά τους, ενώ τα υπόλοιπα με τον κωδικό της ομάδας της Unigene (Hs.) που ανήκουν (πρώτη στήλη του πίνακα). Στη δεύτερη στήλη του πίνακα η ένδειξη (+) συμβολίζει θετικό αποτέλεσμα στην RT-PCR σε εγκέφαλο, ενώ στην τελευταία στήλη παρατίθενται συνοπτικά πληροφορίες για τα γονίδια και η ένδειξη (+) για όσα από αυτά υπάρχουν δεδομένα από την βιβλιογραφία για έκφραση στον εγκέφαλο.

➤ **Γονίδιο #44050**

Το γονίδιο αυτό επιλέχθηκε ως υποψήφιο λόγω της έκφρασής του σε εγκέφαλο βάσει RT-PCR (Εικ. 23) και των ομολογιών που εμφανίζει το πρωτεϊνικό προϊόν του (βλ. παρακάτω). Η μεταγραφική ενεργότητά του σε εγκέφαλο επιβεβαιώθηκε με ανάλυση Northern κατά την οποία ταυτοποιήθηκε ένα μετάγραφο ~2,5kb που εκτός από τον εγκέφαλο εμφανίζει υψηλή έκφραση σε καρδιά και νεφρό (Εικ. 28). Με διερεύνηση βιβλιοθήκης cDNA εγκεφάλου εμβρύου ανθρώπου με ανιχνευτή το ραδιοσημασμένο προϊόν PCR του EST stSG44050, απομονώθηκαν 4 κλώνοι των οποίων και προσδιορίστηκε η αλληλουχία τους. Βάσει αυτής, σε συνδυασμό με ανάλυση των αντίστοιχων ESTs και cDNAs (Blast: nr db και EST_human db) προσδιορίστηκε η γενωμική οργάνωση του γονιδίου και τα πιθανά εναλλακτικά μετάγραφα (Εικ. 24). Με *in silico* μετάφραση του ανοικτού πλαισίου ανάγνωσης προκύπτει πολυπεπτίδιο 146 αμινοξέων που περιέχει 4 μοτίβα MORN (Membrane Occupation and Recognition Nexus, Εικ. 24). Τα μοτίβα αυτά υπάρχουν σε πολλαπλά αντίγραφα σε πολλές πρωτεΐνες από τα θηλαστικά ως και τα βακτήρια αλλά ο ρόλος τους δεν έχει προσδιοριστεί. Ωστόσο, στον άνθρωπο, χαρακτηρισμένες λειτουργικά πρωτεΐνες που φέρουν μοτίβα MORN είναι οι συνδετοφιλίνες (junctophilins) και η αλσίνη (alsin) που μεταλλαγές των αντίστοιχων γονιδίων έχουν ταυτοποιηθεί σε νευρολογικές διαταραχές. Συγκεκριμένα, στην συνδετοφιλίνη-3 (JP-3) ταυτοποιήθηκαν ετερόζυγες μεταλλαγές σε άτομα που εμφανίζουν μία ασθένεια παρόμοια της Huntington (Huntington disease-like2, HDL2, Holmes SE *et al*, 2001). Οι συνδετοφιλίνες αποτελούν μία οικογένεια πρωτεϊνών που περιλαμβάνει μέχρι τώρα 4 μέλη (JP-1, JP-2, JP-3 και JP-4) και συμμετέχουν στο σχηματισμό των συμπλόκων που συνδέουν την κυτταρική μεμβράνη με την μεμβράνη του ενδοπλασματικού/ σαρκοπλασματικού στα διεγερόμενα κύτταρα. Μέσω αυτών των συμπλόκων πραγματοποιείται η επικοινωνία μεταξύ των καναλιών που υπάρχουν στις αντίστοιχες μεμβράνες. Η αλσίνη κωδικοποιείται από το γονίδιο ALS2 που ομόζυγες μεταλλαγές του έχουν ταυτοποιηθεί σε τρεις διακριτές οικογενείς διαταραχές, που εμφανίζουν όμως κοινά κλινικά συμπτώματα, κατά τις οποίες εμφανίζεται προοδευτική παράλυση εξαιτίας δυσλειτουργίας των κινητικών νευρώνων: στην πλάγια μυοτροφική σκλήρυνση (ALS, amyotrophic lateral sclerosis), την πρωτογενή πλάγια σκλήρυνση κατά την παιδική ηλικία (JPLS, juvenile primary lateral sclerosis) και σε μία μορφή σπαστικής παραπληγίας (Yang Y *et al*, 2001; Eymard-Pierre E *et al*, 2002). Η ανασυνδυασμένη πρωτεΐνη εντοπίζεται τόσο στον πυρήνα όσο και στο κυτταρόπλασμα και την κυτταρική μεμβράνη (Οικονόμου Κ., Διπλωματική εργασία, Πανεπιστήμιο Κρήτης 2002) ενώ πιθανά ορθόλογα μόρια ταυτοποιήθηκαν σε θηλαστικά, πτηνά, ψάρια, έντομα και φυτά.

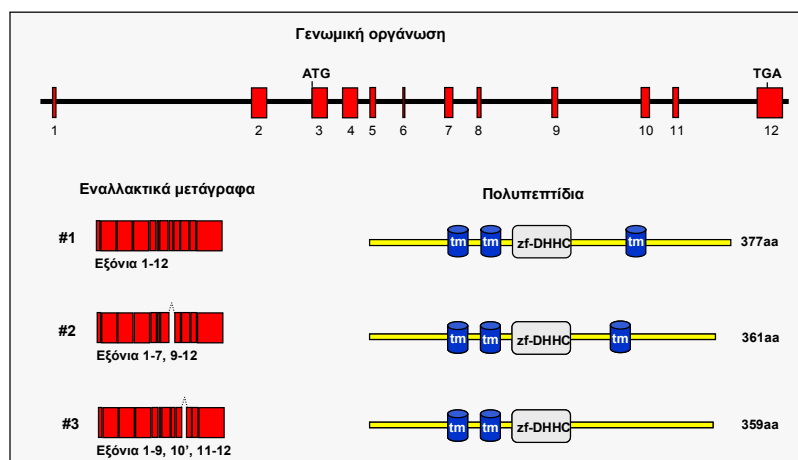


Εικ. 24: Σχηματική απεικόνιση της γενωμικής οργάνωσης, των εναλλακτικών μεταγράφων και της πρωτοταγούς δομή των αντίστοιχων πολυπεπτιδίων του γονιδίου 44050. Το γονίδιο περιλαμβάνει 6 εξόνια και εκτείνεται σε γενωμική απόσταση ~20kb στον κλώνο bA548K23. Βάσει ανάλυσης ESTs και cDNAs το πιο άφθονο είναι το μετάγραφο «1» που περιλαμβάνει τα εξόνια 2-6 και κωδικοποιεί πολυπεπτιδίο 146 αμινοξέων, που περιλαμβάνει 4 μοτίβα MORN (aa: 14-35, 37-58, 60-81, 83-104). Με διερεύνηση των νουκλεοτιδικών βάσεων δεδομένων ταυτοποιήθηκε το εναλλακτικό μετάγραφο «2» (BC010230) στο οποίο το εξόνιο 1 συνδέεται με τα εξόνια 3-6. Την δομή αυτή ακολουθούν ένα επιπλέον cDNA (BC040015) και τρία ESTs (BM453881, BI598383, BI596911). Το μετάγραφο αυτό έχει το ίδιο πλαίσιο ανάγνωσης με το «1» και διαφοροποιείται στο 5'UTR. Ο κλώνος H7 απομονώθηκε με διερεύνηση cDNA βιβλιοθήκης εγκεφάλου εμβρύου αλλά δεν έχει καθοριστεί αν είναι πλήρης ως προς το 5' άκρο του. Σ' αυτόν, το εξόνιο 2 συνδέεται με τα εξόνια 4-6 και αυτήν τη δομή φέρει και το EST BI918954, το οποίο επίσης προέρχεται από βιβλιοθήκη από εγκέφαλο. Το προβλεπόμενο πολυπεπτιδίο έχει 114 αμινοξέα και περιλαμβάνει 3 μοτίβα MORN. Τέλος, το εναλλακτικό μετάγραφο «4» στηρίζεται αποκλειστικά στο cDNA AK098791, που περιλαμβάνει τα εξόνια 2-4 σε σύνδεση με το 6 και κωδικοποιεί πολυπεπτιδίο 63 αμινοξέων με 2 μοτίβα MORN.

➤ Γονίδιο #3077

Η επιλογή αυτού του γονιδίου βασίστηκε στην έκφρασή του στον εγκέφαλο βάσει RT-PCR (Εικ. 23) και των προβλεπόμενων πολλαπλών διαμεμβρανικών περιοχών που έχει το αντίστοιχο πολυπεπτιδίο. Ανάλυση Northern έδειξε την ύπαρξη ενός μεταγράφου ~2kb κυρίως σε εγκέφαλο, καρδιά, σκελετικό μυ, νεφρό, ήπαρ και πλακούντα (Εικ. 28). Με διερεύνηση βιβλιοθήκης cDNA εγκεφάλου ανθρώπινου εμβρύου με ανιχνευτή το ραδιοσημασμένο προϊόν PCR του EST stSG3077, απομονώθηκαν 6 κλώνοι των οποίων και προσδιορίστηκε η αλληλουχία. Ο μεγαλύτερος από αυτούς είχε ένθεμα μήκους 1,85kb και περιείχε όλο το ανοικτό πλαίσιο ανάγνωσης. Η αλληλουχία του κατατέθηκε στην βάση δεδομένων Genbank με κωδικό πρόσβασης AJ294945. Με σύγκριση με τη γενωμική αλληλουχία προσδιορίστηκε η οργάνωση του γονιδίου, που περιλαμβάνει 12 εξόνια (Εικ. 25) και κωδικοποιεί πολυπεπτιδίο 377 αμινοξέων. Σ' αυτό, βάσει πρόβλεψης δομής με υπολογιστικά προγράμματα (SMART, TMpred), εντοπίστηκε το συντηρημένο μοτίβο

zf-DHHC και τουλάχιστον 3 διαμεμβρανικές περιοχές (Εικ. 25). Το μοτίβο αυτό περιλαμβάνει 30 περίπου αμινοξέα και εμφανίζεται πολύ συντηρημένο σε πρωτεΐνες διάφορων οργανισμών, από τον σακχαρομύκητα μέχρι τον άνθρωπο (βλ. Συζήτηση). Από ανάλυση των διαφορετικών κλώνων cDNA που απομονώθηκαν, ταυτοποιήθηκαν δύο εναλλακτικά μεταγράφα με μικρή διαφορά μεγέθους μεταξύ τους (Εικ. 25). Η ανασυνδυασμένη πρωτεΐνη εντοπίζεται στο κυτταρόπλασμα, πιθανώς στο ενδοπλασματικό δίκτυο καθώς και στην κυτταρική μεμβράνη (Εικ. 29). Για την ανίχνευση μεταλλαγών σε ασθενείς με ADLTE, σχεδιάστηκαν εκκινητές εκατέρωθεν κάθε εξονίου (Πιν. 5), έγινε βελτιστοποίηση των συνθηκών της PCR και στάλθηκαν στο εργαστήριο του Dr. C.Nobile για διερεύνηση μεταλλαγών αλλά το αποτέλεσμα ήταν αρνητικό.

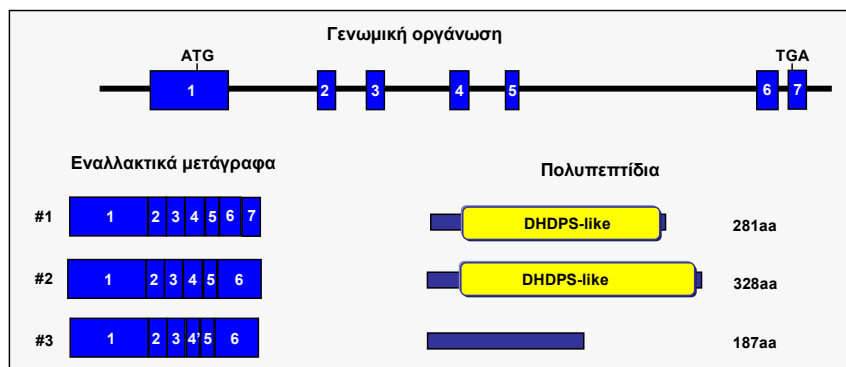


Εικ. 25: Σχηματική απεικόνιση της γενωμικής οργάνωσης, των εναλλακτικών μεταγράφων και της πρωτοταγούς δομή των αντίστοιχων πολυπεπτιδίων του γονιδίου 3077. Το γονίδιο περιλαμβάνει 12 εξόνια και εκτείνεται περί τα ~11kb στον κλώνο bA452K12. Το μετάγραφο #3077.1 απομονώθηκε μετά από διερεύνηση cDNA βιβλιοθήκης εγκεφάλου εμβρύου, έχει μήκος 1851 νουκλεοτίδια, περιλαμβάνει και τα 12 εξόνια και κωδικοποιεί πολυπεπτίδιο 377 αμινοξέων. Βάσει του προγράμματος SMART, το πολυπεπτίδιο φέρει τρία διαμεμβρανικά (tm) τμήματα (aa: 81-103, 115-137, 251-273) και το μοτίβο zf-DHHC. Τα εναλλακτικά μεταγράφα #3077.2 και #3077.3 ταυτοποιήθηκαν μετά από PCR σε DNA της παραπάνω βιβλιοθήκης με τους εκκινητές MycR-3077867 (βλ. Πιν.5) και προσδιορισμό της αλληλουχίας τους. Το μετάγραφο #3077.2 στερείται το εξόνιο 8 (νουκλεοτίδια 933-980 της καταχώρησης AJ294945), διατηρεί όμως το ίδιο αναγνωστικό πλαίσιο. Στο μετάγραφο #3077.3 το εξόνιο 9 συνδέεται με τμήμα του εξονίου 10. Συγκεκριμένα, κατά την συναρμογή του χρησιμοποιείται μία συντηρημένη αλληλουχία "ag" που υπάρχει μέσα στο εξόνιο 10, με αποτέλεσμα να μην συμπεριλαμβάνονται στο μετάγραφο τα 54 πρώτα νουκλεοτίδια του εξονίου (νουκλεοτίδια 1067-1120 της καταχώρησης AJ294945). Παρόλο που το αναγνωστικό πλαίσιο διατηρείται, το παραγόμενο πολυπεπτίδιο (359 αμινοξέων) στερείται της τρίτης διαμεμβρανικής περιοχής.

➤ **Γονίδιο #AJ312051**

Για το γονίδιο αυτό η RT-PCR ήταν αρνητική (Εικ. 23), σε αντίθεση με την ανάλυση των ESTs (Expression profile db, Unigene). Ωστόσο, το γονίδιο αυτό επιλέχθηκε ως υποψήφιο λόγω της ομοιότητας που εμφανίζει το παραγόμενο πολυπεπτίδιο με το ένζυμο λύση του N-ακετυλ-νευραμινικού οξέος, που είναι το

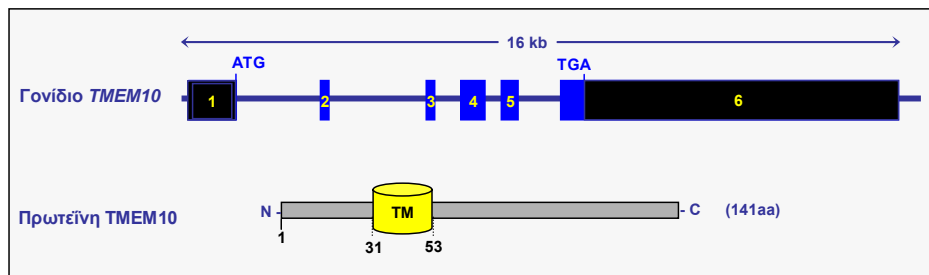
τελικό ένζυμο του καταβολισμού των σιαλικών οξέων και δεν έχει χαρακτηριστεί ακόμα στον άνθρωπο. Η επιλογή στηρίχθηκε στη συσχέτιση διαταραχών του μεταβολισμού των σιαλικών οξέων με διάφορες νόσους, συμπεριλαμβανομένου και νευρολογικών διαταραχών, όπως στην περίπτωση της νόσου Salla μίας ψυχοκινητικής διαταραχής κατά την οποία σημαντικό ποσοστό των ασθενών εμφανίζει και επιληψία (Erikson A *et al*, 2002). Το cDNA κατασκευάστηκε *in silico* με στοίχιση πολλαπλών ESTs και απομονώθηκε με RT-PCR από νεφρό ανθρώπου. Η αλληλουχία επιβεβαιώθηκε και κατατέθηκε στη νουκλεοτιδική βάση δεδομένων με κωδικό πρόσβασης AJ312051. Το γονίδιο αποτελείται από 7 εξόνια, όπως προσδιορίστηκε *in silico* και εκτός από το κύριο μετάγραφο, παράγει δύο επιπλέον εναλλακτικά μετάγραφα, που απομονώθηκαν με RT-PCR από ήπαρ ανθρώπου (Εικ. 26). Η ανασυνδυασμένη πρωτεΐνη εντοπίζεται στο κυτταρόπλασμα (Εικ. 29), τοπολογία που είναι σε συμφωνία με τον θεωρητικό εντοπισμό της λύσης (Bulai *et al*, 2002). Η ανάλυση Northern έδειξε ότι εκφράζεται κυρίως σε ήπαρ, νεφρό και σκελετικό μυ και γι' αυτό δεν συμπεριελήφθη στην ανάλυση μεταλλαγών.



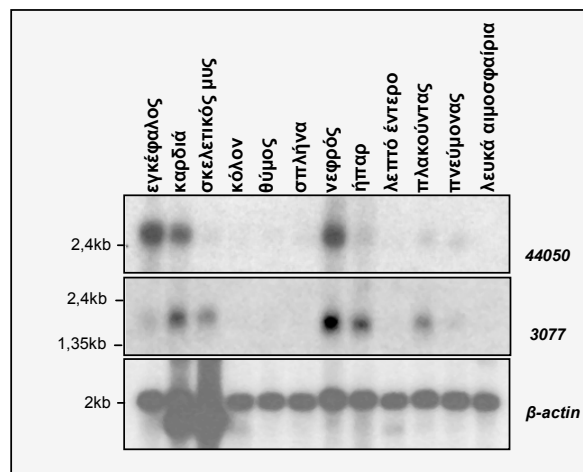
Εικ.26. Σχηματική απεικόνιση της γενωμικής οργάνωσης, των εναλλακτικών μεταγράφων και της πρωτοταγούς δομή των αντίστοιχων πολυπεπτιδίων του γονιδίου AJ312051. Το γονίδιο περιλαμβάνει 7 εξόνια και εκτείνεται περί τα ~18kb στον κλώνο bA548K23. Μετά τον προσδιορισμό της δομής του γονιδίου *in silico*, σχεδιάστηκαν κατάλληλοι εκκινητές (F_BamHI-R_Kpn) και πραγματοποιήθηκε PCR σε 5ng αντίδρασης αντίστροφης μεταγραφής από νεφρό (MTC, cDNA pannels, Clontech). Το προϊόν κλωνοποιήθηκε σε φορέα *pBluescript* και μετά τον προσδιορισμό της αλληλουχίας του, βρέθηκε ότι περιλαμβάνει και τα 7 εξόνια. Το αντίστοιχο πολυπεπτίδιο αποτελείται από 281 αμινοξέα και εμφανίζει σημαντική ομοιότητα (30% ταυτότητα, 50% ομολογία) καθ' όλο το μήκος του με τα ένζυμα λύση του N-ακετυλ-νευραμινικού οξέος και διυδροδιπικολινική συνθάση. Τα εναλλακτικά μετάγραφα #AJ312051.2 και # AJ312051.3 ταυτοποιήθηκαν μετά από RT-PCR σε ήπαρ (MTC, cDNA pannels, Clontech) με τους εκκινητές F_SpeI-R_Kpn, κλωνοποίησή τους σε φορέα *pBluescript* και προσδιορισμό της αλληλουχίας τους. Το μετάγραφο # AJ312051.2 περιλαμβάνει επιπλέον και το ιντρόνιο μεταξύ των εξονίων 6 και 7, που φέρει ανοικτό πλαίσιο ανάγνωσης, με αποτέλεσμα να παράγεται πολυπεπτίδιο 328 αμινοξέων. Στο μετάγραφο # AJ312051.3, το εξόνιο 3 συνδέεται με τμήμα του εξονίου 4. Συγκεκριμένα, κατά την συναρμογή του χρησιμοποιείται μία συντηρημένη αλληλουχία "ag" που υπάρχει μέσα στο εξόνιο 4, με αποτέλεσμα να μην συμπεριλαμβάνονται στο μετάγραφο τα 38 πρώτα νουκλεοτίδια του εξονίου (νουκλεοτίδια 829-866 της καταχώρησης AJ312051). Το αποτέλεσμα είναι ένα πρώιμο κωδικόνιο λήξης στο τέλος του εξονίου 4' που οδηγεί στην παραγωγή πολυπεπτιδίου 187 αμινοξέων.

➤ Γονίδιο *TMEM10*

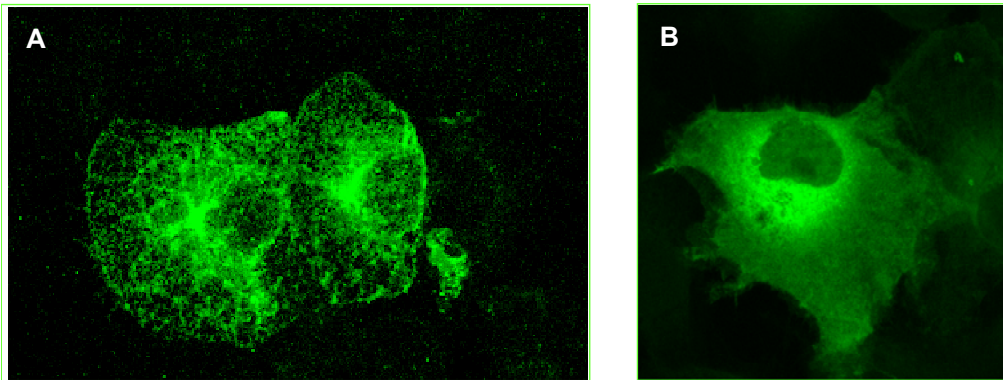
Το γονίδιο αυτό μελετήθηκε λόγω της υψηλής και αποκλειστικής του έκφρασης στον εγκέφαλο (Nobile C & Pitzalis S, 1999), συμπεριλαμβανομένου του κροταφικού λοβού. Αποτελείται από 6 εξόνια και κωδικοποιεί πρωτεΐνη 141 αμινοξέων, πιθανώς διαμεμβρανική, που εμφανίζει σημαντική αμινοξική ταυτότητα αποκλειστικά με μόρια των θηλαστικών. Σε ασθενείς με ADLTE δεν ανιχνεύτηκαν μεταλλαγές στην κωδική περιοχή του γονιδίου (Nobile C *et al*, 2002). Η δική μας συνεισφορά αφορούσε την *in silico* μελέτη των δομικών χαρακτηριστικών του γονιδίου (Εικ. 27).



Εικ. 27: Σχηματική απεικόνιση της γενωμικής οργάνωσης και της πρωτοταγούς δομής της πρωτεΐνης *TMEM10*. Το γονίδιο περιλαμβάνει 6 εξόνια και εκτείνεται περί τα ~16kb στον κλώνο bA35J23. Το μοναδικό μετάγραφο που έχει ανιχνευθεί (AF367761) έχει μήκος 3555nt. Το αντίστοιχο πολυπεπτιδίο (AAK62349) αποτελείται από 141 αμινοξέα και περιέχει μία διαμεμβρανική περιοχή (TM), με βάση την πρόβλεψη του προγράμματος SMART.



Εικ. 28: Προσδιορισμός του προτύπου έκφρασης των γονιδίων 44050 και 3077 με ανάλυση Northern. Με την ανάλυση ανιχνεύθηκε ένα κύριο μετάγραφο μεγέθους ~1,4kb για το 3077 και ~3kb για το 44050. Ο ποιοτικός και ποσοτικός έλεγχος του RNA έγινε με επαναυβριδοποίηση της μεμβράνης με β-ακτίνη (σύμφωνα με τις οδηγίες της Clontech). Στην αυτοραδιογραφία της ακτίνης, στα δείγματα της καρδιάς και του σκελετικού μυ εμφανίζονται δύο ζώνες που είναι πιθανότατα αποτέλεσμα υβριδοποίησης του ανιχνευτή με άλλες ισομορφές του mRNA, τις α ή γ ακτίνες. Ως ανιχνευτές και για τα δύο γονίδια χρησιμοποιήθηκαν ραδιοσημασμένα προϊόντα PCR (βλ. Μεθόδους) με εκκινητές αυτούς των αντίστοιχων ESTs (stSG44050, stSG3077, βλ. Πιν., Μεθόδους). Η υβριδοποίηση πραγματοποιήθηκε με ~5x10⁵cpm ανιχνευτή ανά ml διαλύματος υβριδοποίησης, στους 66oC για 3h μετά από προυβριδοποίηση της μεμβράνης για 30min παρουσία φορέα-DNA με το διάλυμα υβριδοποίησης ExpressHybTM. Η ειδική ενεργότητα των ανιχνευτών ήταν ~109cpm/μg. Τα πλυσίματα της μεμβράνης έγιναν με διάλυμα 2XSSC, 0,05% SDS.



Εικ. 29: Υποκυτταρικός εντοπισμός των ανασυνδυασμένων πρωτεϊνών 3077 (A) και AJ312051 (B) σε κύτταρα COS-7. Η πρωτεΐνη 3077 συσσωρεύεται στο κυτταρόπλασμα (με κατανομή που πιθανώς αντιστοιχεί στο ενδοπλασματικό δίκτυο) και στην κυτταρική μεμβράνη. Η AJ312051 εμφανίζεται διάχυτη στο κυτταρόπλασμα. Τα κύτταρα διαμολύνθηκαν παροδικά με τις κατάλληλες πλασμιδιακές κατασκευές που περιλαμβάνουν τα cDNA πλήρους μήκους στους φορείς pcDNA 3.1 Myc.His και pEGFPc1 (βλ. Πλασμιδιακές κατασκευές). Για την ανίχνευση της ανασυνδυασμένης πρωτεΐνης 3077 χρησιμοποιήθηκε το πρωτογενές αντίσωμα anti-Myc και ακολούθως το δευτερογενές anti-mouse FITC. Οι εικόνες προέρχονται από συνεστιακό μικροσκόπιο φθορισμού.

➤ **Γονίδιο *LG1***

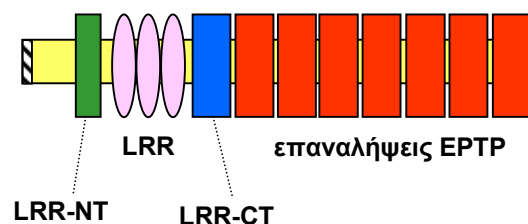
Το γονίδιο *LG1* (Leucine-rich gene, Glioma Inactivated 1) είχε κλωνοποιηθεί παλαιότερα βάσει της χρωμοσωμικής του θέσης στο σημείο θραύσης της αμοιβαίας χρωμοσωμικής μετατόπισης t(10;19)(q24;q13) από μία κυτταρική σειρά γλοιοβλαστώματος και είχε προταθεί ως ογκοκατασταλτικό γονίδιο (Chernova OB *et al*, 1998). Ως υποψήφιο υπεύθυνο για την ADLTE θεωρήθηκε: 1) επειδή χαρτογραφείται στην κρίσιμη περιοχή (βλ. παραπάνω) και 2) εξαιτίας της υψηλής και σχεδόν αποκλειστικής έκφρασής του στον εγκέφαλο (Nobile C & Pitzalis S, 1999). Η πρωτεΐνη *LG1* περιλαμβάνει ένα N-τελικό πεπτιδίο-σινιάλο, τρεις πλούσιες σε Leu επαναλήψεις LRR (leucine-rich repeats) και 7 συνεχόμενες, ειδικές για την πρωτεΐνη, επαναλήψεις EPTP (Εικ. 30) και ορθόλογά της εντοπίζονται μόνο σε σπονδυλωτά (Gu W *et al*, 2005).

Μεταλλαγές του γονιδίου *LG1* που συγκληρονομούνται με το φαινότυπο της ADLTE προσδιορίστηκαν ανεξάρτητα και σχεδόν ταυτόχρονα, από το ευρωπαϊκό consortium για τη ADLTE (Morante-Redolat JM *et al*, 2002) και την ομάδα της R. Ottman (U. of Columbia, US, Kalachikov S *et al*, 2002). Οι μεταλλαγές που ταυτοποιήθηκαν προκαλούν: 1) πρόωρο τερματισμό της μετάφρασης, 2) αμινοξικές αντικαταστάσεις και 3) νουκλεοτιδικές αλλαγές σε συντηρημένα σημεία συναρμογής εξονίων-ιντρονίων (Εικ. 31). Οι μεταλλαγές αυτές χαρτογραφούνται καθ' όλο το μήκος της πρωτεΐνης και παρά τη λεπτομερή περιγραφή των κλινικών φαινοτύπων δεν είναι ακόμη εφικτή η συσχέτιση γονοτύπου-φαινοτύπου. Επίσης, μετά από διερεύνηση μίας σειράς γενωμικών βάσεων δεδομένων και *in silico* ανάλυση

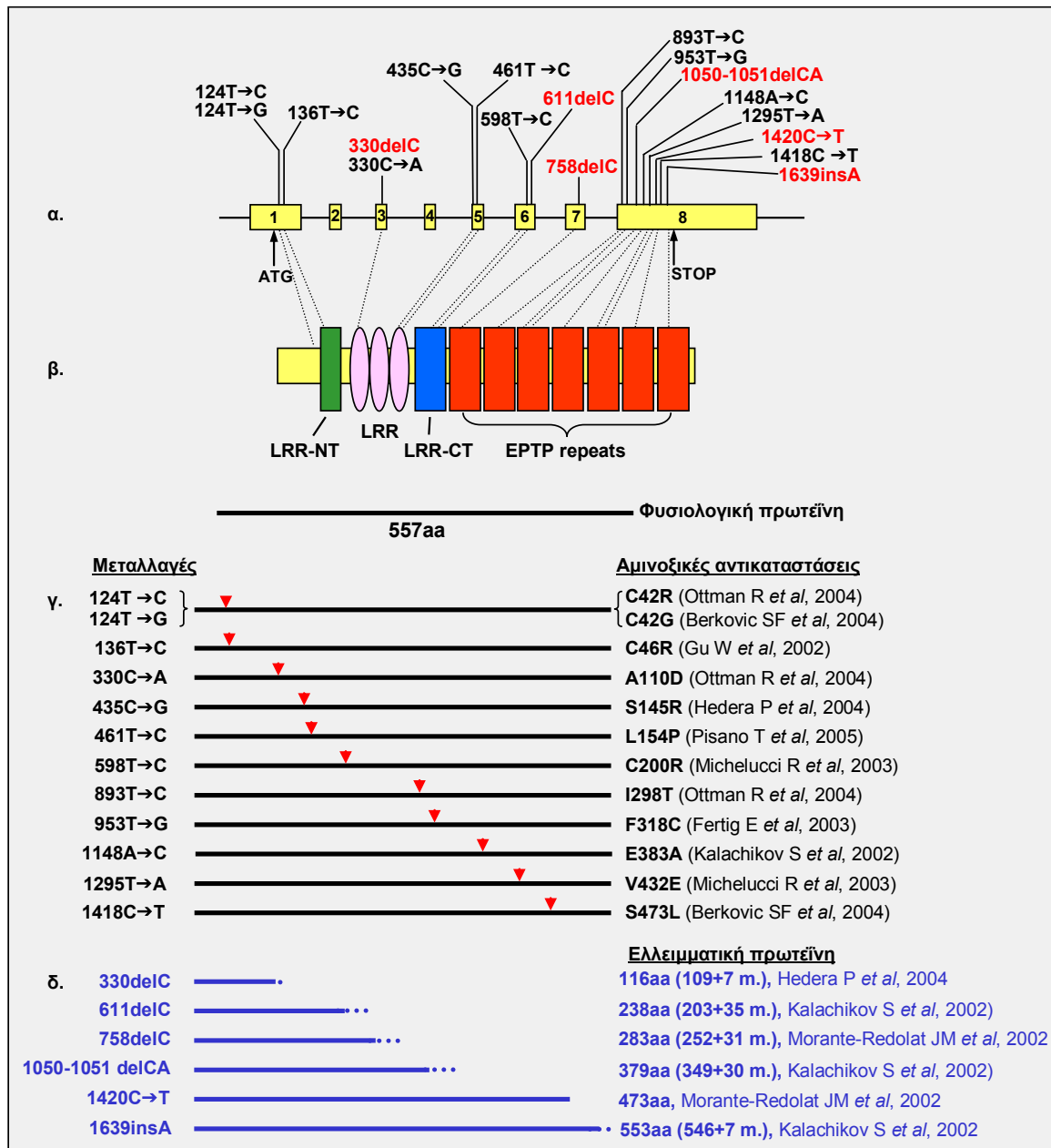
ταυτοποιήθηκαν τρία επιπλέον γονίδια, με παρόμοια δομικά χαρακτηριστικά που ονομάστηκαν *LG1* 2, 3 και 4, που χαρτογραφούνται στις περιοχές 4p15.31, 8p21.3 και 19q13.12 αντίστοιχα δημιουργώντας έτσι την οικογένεια των γονιδίων με επαναλήψεις EPT (Εικ. 30, Staub E *et al*, 2002). Οι αντίστοιχες πρωτεΐνες εμφανίζουν μεταξύ τους ταυτότητα που κυμαίνεται μεταξύ 65% και 75%.

Μετά την ταυτοποίηση μεταλλαγών στο γονίδιο *LG1* σε οικογένειες με επιληψία ADLTE το ενδιαφέρον εστιάστηκε σε δύο κυρίως κατευθύνσεις: 1) στην διερεύνηση του βιοχημικού ρόλου της πρωτεΐνης και τη συσχέτισή της με τον παθολογικό φαινότυπο και 2) στην εξέταση των παραλόγων γονιδίων *LG12-LG14* ως πιθανά υποψήφια για την συγκεκριμένη ή άλλες μορφές επιληψίας.

Το *LG1* είναι από τις ελάχιστες εξαιρέσεις μορίων που σχετίζονται με επιληπτικούς φαινοτύπους χωρίς να είναι κανάλι ιόντων ή υποδοχέας νευροδιαβιβαστών. Όσον αφορά την αρχική θεώρησή του ως ογκοκατασταλτικό γονίδιο που σχετίζεται με την ανάπτυξη γλοιομάτων (Chernova OB *et al*, 1998) δεν υπάρχουν επιπλέον ενδείξεις και εξάλλου το γονίδιο εκφράζεται κυρίως σε νευρικά κύτταρα και όχι σε γλοία (Kalachikov S *et al*, 2002, Morande-Redolat *et al*, 2002) ενώ δεν έχουν αναφερθεί περιπτώσεις εμφάνισης γλοιοβλαστώματος στις οικογένειες με ADLTE (Ottman R *et al*, 2004, Brodtkorb E *et al*, 2003). Πρόσφατα πειράματα δείχνουν ότι πιθανώς οι πρωτεΐνες *LG1-4* είναι εκκρινόμενες (Senechal KR *et al*, 2005) και ότι διάφορες μεταλλαγές της *LG1*, που εντοπίστηκαν σε ασθενείς, έχουν ως αποτέλεσμα μείωση στην ποσότητα της εκκρινόμενης πρωτεΐνης. Παρόλο που οι συγκεκριμένες μελέτες πραγματοποιήθηκαν σε ινοβλάστες (κύτταρα 293T) και αναφέρονται στα ανασυνδυασμένα μόρια, τα αποτελέσματά τους σχετικά με την τοπολογία της πρωτεΐνης *LG1* είναι σε συμφωνία με παλαιότερες προβλέψεις (Morande-Redolat *et al*, 2002).



Εικ. 30: Δομή της πρωτεΐνης *LG1*. Η πρωτεΐνη φέρει στο αμινοτελικό της άκρο ένα πεπτιδίο-σινιάλο και 3 επαναλήψεις πλούσιες σε λευκίνη (LRR) που εκατέρωθεν τους έχουν δύο περιοχές πλούσιες σε κυστεΐνες (LRR-NT, LRR-CT). Στο καρβοξυτελικό άκρο της πρωτεΐνης ταυτοποιήθηκε, σε 7 αντίγραφα, το μοτίβο EPTP που περιλαμβάνει 50 αμινοξέα. Την ίδια ακριβώς δομή παρουσιάζουν και οι πρωτεΐνες *LG12*, *LG13* και *LG14*.



Εικ. 31: Οι μεταλλαγές που έχουν ταυτοποιηθεί στο γονίδιο *LG11* σε ασθενείς με **ADLTE.** Απεικονίζεται η θέση των μεταλλαγών σε επίπεδο γονιδίου (α) και πρωτεΐνης (β). Η αρίθμηση που ακολουθείται θεωρεί ως +1 το A του κωδικονίου έναρξης της μετάφρασης, ενώ σε όσες βιβλιογραφικές αναφορές ακολουθείται διαφορετική αρίθμηση, αυτή προσαρμόστηκε κατάλληλα. Μέχρι σήμερα (Ιούλιος 2005), έχουν ταυτοποιηθεί 12 μεταλλαγές που προκαλούν αμινοξικές αντικαταστάσεις στην πρωτεΐνη (γ) και 6 μεταλλαγές που οδηγούν στην παραγωγή ελλειμματικού προϊόντος (δ). Επιπλέον, έχουν εντοπιστεί δύο μεταλλαγές, οι IVS3(-3)C> A (Kalachikov S *et al*, 2002) και IVS7(-2)A> G (Kobayashi E *et al*, 2003) σε συντηρημένες θέσεις των σημείων συναρμογής ιντρονίων-εξονίων. Στην περίπτωση που τα αντίστοιχα ιντρόνια παραμένουν στο ώριμο μετάγραφο, προκύπτει ελλειμματική πρωτεΐνη λόγω πρώιμου τερματισμού της μετάφρασης.

Ως συμβολή στην ανάλυση της οικογένειας των πρωτεϊνών EPT, εξετάσαμε λεπτομερώς το πρότυπο έκφρασης των ορθόλογων γονιδίων *Lgi1-4* του ποντικού (Somerville RP *et al*, 2000; Staub E *et al*, 2002; Gu W *et al*, 2002; Scheel H *et al*, 2002), με RNA *in situ* υβριδοποίηση σε τομές εγκεφάλου ενήλικου ποντικού. Η μελέτη αυτή είχε σκοπό τον έλεγχο της έκφρασης των γονιδίων αυτών σε συγκεκριμένες περιοχές του εγκεφάλου που θα μπορούσαν να συσχετιστούν με κλινικούς φαινότυπους επιληψίας και στηρίχθηκε στις εξής παρατηρήσεις:

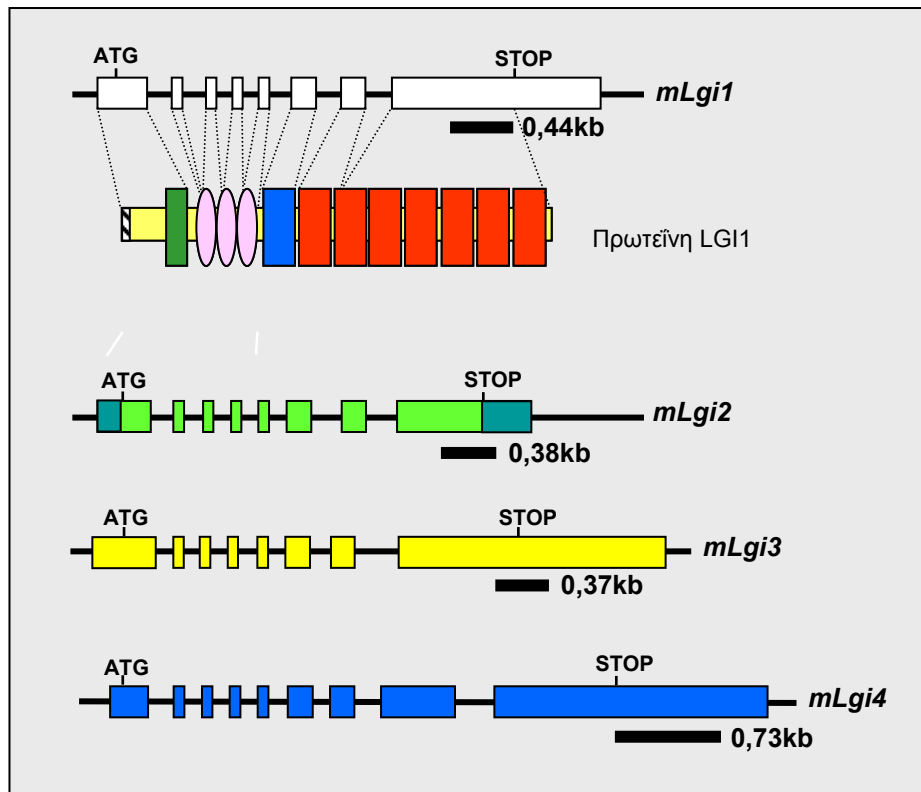
1) Στη γενετική ετερογένεια που παρουσιάζει η ADLTE

Αρκετές οικογένειες που έχουν περιγραφεί, ενώ έχουν φαινότυπο κλινικά παρόμοιο με αυτόν της ADLTE, δεν φέρουν μεταλλαγές στο *LG11*. Για αρκετές από αυτές τις οικογένειες δεν ήταν δυνατόν να εφαρμοστεί η ανάλυση γενετικής σύνδεσης λόγω του μικρού μεγέθους τους. Οπότε, σ' αυτά τα άτομα, καθώς και σε σποραδικές περιπτώσεις ιδιοπαθούς επιληψίας με παρόμοια συμπτώματα, τα γονίδια *LG12-4*, θα μπορούσαν να θεωρηθούν ως υποψήφια και να ελεγχθούν περαιτέρω για μεταλλαγές, εφόσον ταυτοποιηθεί ότι είναι μεταγραφικά ενεργά στον κροταφικό λοβό.

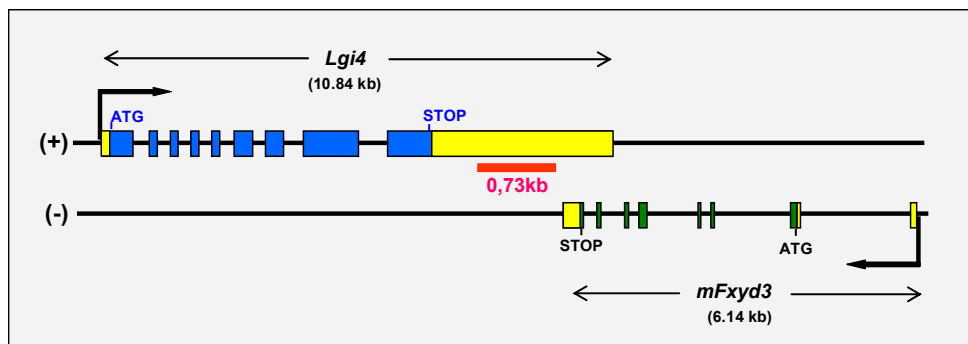
2) Στη χαρτογράφηση του *LG14*

Το *LG14* χαρτογραφείται στην χρωμοσωμική περιοχή 19q13, που έχει συσχετιστεί με μία μορφή ιδιοπαθούς μερικής επιληψίας, την BFIC (benign familial infantile convulsions), που όπως και η ADLTE εμφανίζει αυτοσωμικό επικρατή τύπο κληρονομησης. Επειδή σ' αυτήν την διαταραχή, οι επιληπτικές κρίσεις σχετίζονται με περιοχές του ινιακού και του βρεγματικού λοβού, σκοπός των πειραμάτων ήταν ο έλεγχος της έκφρασης του *Lgi4* σε αυτές τις περιοχές.

Η διαδικασία της υβριδοποίησης πραγματοποιήθηκε όπως περιγράφηκε στις μεθόδους. Οι ριβοανιχνευτές (Εικ. 10, Μέθοδοι) παρήχθησαν με μεταγραφή *in vitro* μετά από γραμμοποίηση με κατάλληλα ένζυμα, για την παρασκευή του συμπληρωματικού και του μη-συμπληρωματικού ανιχνευτή. Η επιλογή των ριβοανιχνευτών έγινε έτσι ώστε να αντιστοιχούν σε γονιδιο-ειδικές περιοχές που δεν παρουσιάζουν συμπληρωματικότητα με τα mRNAs των άλλων μελών της υποοικογένειας (Εικ. 32). Επίσης, καθώς τμήμα του τελευταίου εξονίου του *Lgi4* αλληλοεπικαλύπτεται με τα δύο τελευταία εξόνια του γονιδίου *Fxyd3*, το οποίο μεταγράφεται με αντίθετο προσανατολισμό (Εικ. 33, Runkel F *et al*, 2003), ο αντίστοιχος ριβοανιχνευτής σχεδιάστηκε εκτός της περιοχής αλληλοεπικάλυψης.



Εικ. 32: Σχηματική απεικόνιση των ριβοανιχνευτών που χρησιμοποιήθηκαν στην *in situ* υβριδοποίηση. Παριστάνονται οι ριβοανιχνευτές (οριζόντιες μαύρες γραμμές), το μήκος τους και η περιοχή στην οποία αντιστοιχούν, σε σχέση με την γενωμική οργάνωση των γονιδίων *Lgi1-4*. Η γενωμική οργάνωση καθορίστηκε με βάση τις αλληλουχίες των cDNAs AF246818 (*Lgi1*), AY841361 που καθώς περιλαμβάνει μόνο το ανοικτό πλαίσιο ανάγνωσης, επεκτάθηκε *in silico* (σκούρο πράσινο) βάσει αλληλουχιών από ESTs (*Lgi2*), NM_145219 (*Lgi3*) και AJ487521 (*Lgi4*). Επιπλέον φαίνεται η αντιστοίχιση των μοτίβων τις πρωτεΐνης LGI1 (συμβολίζονται όπως και στην Εικ.30) με τα εξόνια του γονιδίου από τα οποία κωδικοποιούνται. Η ίδια αντιστοιχία ισχύει και για τα γονίδια *Lgi2* και *Lgi3*. Το γονίδιο *Lgi4* έχει ένα εξόνιο επιπλέον και οι επαναλήψεις EPTP 2-7 της πρωτεΐνης LGI4 κωδικοποιούνται από τα δύο τελευταία εξόνια.



Εικ. 33: Αλληλοεπικάλυψη των 3' άκρων των γονιδίων *Lgi4* και *Fxyd3* στο χρωμόσωμα 7 του ποντικού. Σημειώνεται η θέση του ριβοανιχνευτή (κόκκινη γραμμή) που χρησιμοποιήθηκε στα πειράματα *in situ* υβριδοποίησης.

Από τη συγκριτική μελέτη του προτύπου έκφρασης διαπιστώθηκε ότι τα γονίδια αυτής της υποοικογένειας γενικά εμφανίζουν εκτεταμένη έκφραση στον εγκέφαλο, παρουσιάζοντας τόσο αλληλοεπικαλυπτόμενες όσο και διακριτές περιοχές μεταγραφικής ενεργότητας.

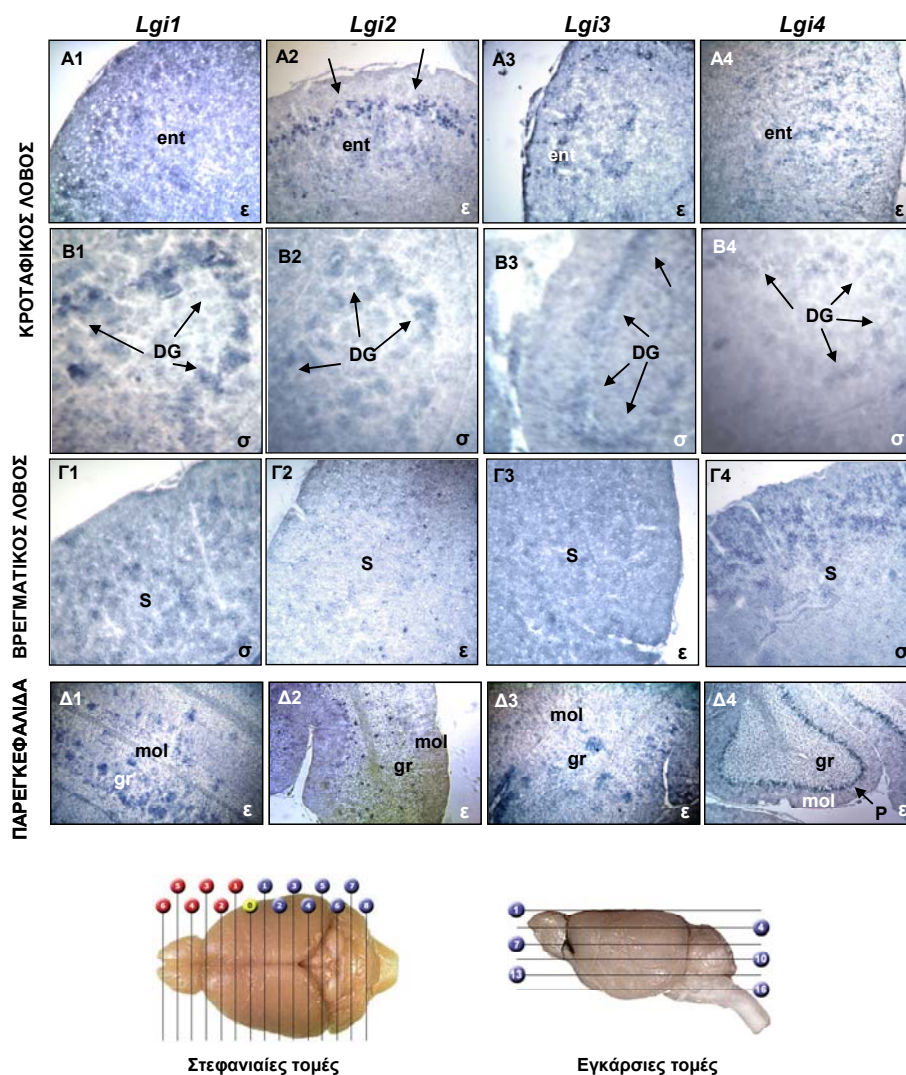
Ειδικότερα, ταυτοποιήθηκαν περιοχές του κροταφικό λοβού όπου όλα τα γονίδια είναι μεταγραφικά ενεργά. Έτσι, παρατηρήθηκε συνεντοπισμός των μεταγράφων των *Lgi1-4* στον ενδορινικό φλοιό και τον ιππόκαμπο (Εικ. 34A). Στον πρώτο, διαπιστώθηκε έκφραση σε όλες τις στοιβάδες του, εκτός από το *Lgi2*, που εμφανίζει ειδικότερη υψηλή έκφραση σε συγκεκριμένη στοιβάδα. Ο ενδορινικός φλοιός αποτελεί ένα σημαντικό κέντρο του εγκεφάλου που σχετίζεται με την μνήμη. Αποτελεί την είσοδο στον ιππόκαμπο και είναι υπεύθυνος για την επεξεργασία των σημάτων εισόδου. Στον ιππόκαμπο παρουσιάζεται συνεντοπισμός των μεταγράφων στην οδοντωτή έλικα (*dentate gyrus*, Εικ. 34A). Η περιοχή αυτή περιέχει κοκκώδη κύτταρα που προεκβάλλουν στα πυραμιδικά κύτταρα των ζωνών CA2-4 του ιπποκάμπου. Τα κοκκώδη κύτταρα αποτελούν διεγερτικούς νευρώνες που δέχονται σήματα από περιοχές του νεοφλοιού. Η οδοντωτή έλικα είναι από τις λίγες περιοχές του εγκεφάλου όπου γίνεται νευρογένεση. Συνέκφραση και των τεσσάρων γονιδίων διαπιστώθηκε στον σωματοαισθητικό φλοιό του βρεγματικού λοβού και την παρεγκεφαλίδα (Εικ. 34A). Ωστόσο, στην τελευταία, ενώ τα *Lgi1* και *Lgi3* εκφράζονται τόσο στην μοριακή όσο και στην κοκκοειδή στοιβάδα, το *Lgi2* είναι μεταγραφικά ενεργό κυρίως στην κοκκοειδή στοιβάδα και το *Lgi4* αποκλειστικά στα κύτταρα Purkinje.

Επιπλέον, ταυτοποιήθηκαν περιοχές όπου συγκεκριμένα μόνο μέλη της υποοικογένειας εμφανίζουν αλληλοεπικαλυπτόμενη έκφραση (Εικ. 34B). Ειδικότερα, στο μεταιχμιακό σύστημα διαπιστώθηκε συνέκφραση των: α) *Lgi2*, 3, 4 στο διάφραγμα (*septum*) έναν πυρήνα που σχετίζεται με το θυμό και την ευχαρίστηση, β) *Lgi1*, 2 και 4 στον απιοειδή φλοιό (*piriform cortex*), γ) των *Lgi1*, 2 και 3 στην περιοχή CA3 του ιπποκάμπου και σε πυρήνες του θαλάμου. Το μεταιχμιακό σύστημα περιλαμβάνει περιοχές από όλους τους λοβούς των εγκεφαλικών ημισφαιρίων και φυλογενετικά αποτελεί ένα από τα πιο πρωτόγονα μέρη του εγκεφάλου. Διαδραματίζει κεντρικό ρόλο στη μνήμη, τη μάθηση, τα συναισθήματα, τη νευροενδοκρινική λειτουργία και σε αυτόνομες δραστηριότητες ενώ διάφορες κλινικές καταστάσεις συσχετίζονται με διαταραχές αυτού του συστήματος, όπως επιληψίες, ψυχιατρικές διαταραχές, άνοια κ.λπ. Οι παραπάνω δομές του συστήματος που περιγράφηκαν, πλην του θαλάμου, ανήκουν όλες στον κροταφικό λοβό. Επιπλέον, στον μετωπιαίο λοβό, ταυτοποιήθηκε συνέκφραση των γονιδίων *Lgi3* και *Lgi4* στον κινητικό φλοιό.

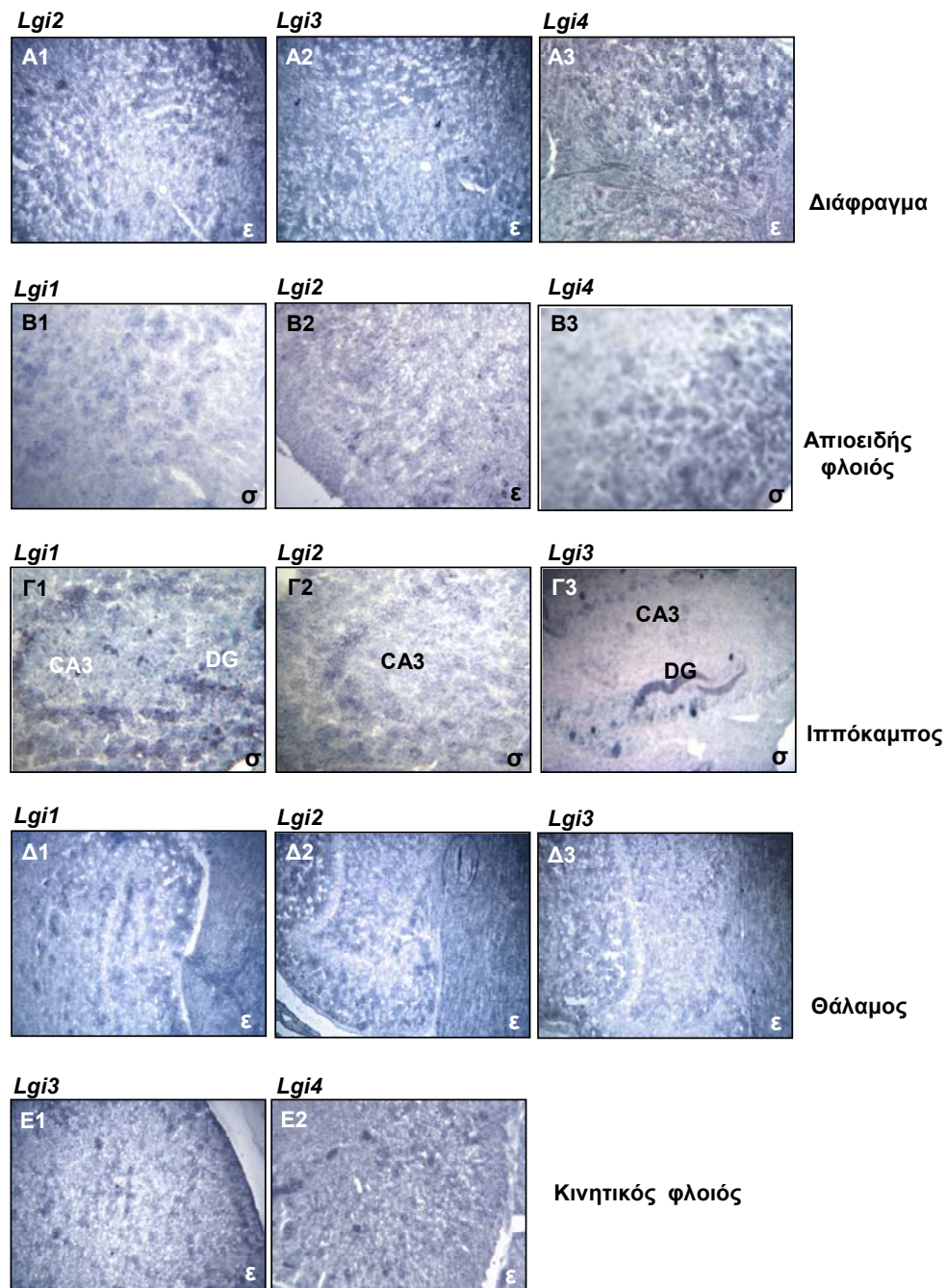
Τέλος, ταυτοποιήθηκαν δομές όπου μόνο ένα γονίδιο της υποοικογένειας εμφάνιζε υψηλή μεταγραφική ενεργότητα (Εικ. 34Γ). Έτσι, στον κροταφικό λοβό το *Lgi1* παρουσιάζει υψηλά επίπεδα έκφρασης στον ακουστικό φλοιό, στην στοιβάδα Oriens, στον κοιλιακό σύνδεσμο (ventral commissure) και στο υπόθεμα (subicullum) του ιπποκάμπου, καθώς και στα αμύγδαλα. Επιπλέον, παρατηρήθηκε έκφραση του *Lgi1* στον οσφρητικό και τον οπτικό φλοιό. Υψηλά επίπεδα μεταγράφων του *Lgi2* εντοπίστηκαν στον κάτω παρεγκεφαλιδικό μίσχο (inferior cerebellar peduncle), μία μεγάλη δέσμη ινών που συνδέει τον προμήκη μυελό με την παρεγκεφαλίδα. Για το *Lgi3* ταυτοποιήθηκε υψηλή έκφραση στον φλοιό του προσαγωγίου (cingulate cortex) και σε αιθουσαίους πυρήνες ενώ για το *Lgi4* στην συνδετική περιοχή (association area) του κροταφικού φλοιού.

Συμπερασματικά, το πρότυπο έκφρασης του *Lgi1* είναι σε συμφωνία με τον κλινικό φαινότυπο της συγκεκριμένης μερικής επιληψίας που εκπορεύεται από δυσλειτουργία του κροταφικού λοβού και του περιβάλλοντος ακουστικού φλοιού. Επίσης, καθώς παρατηρήθηκαν περιοχές του κροταφικού λοβού όπου τα γονίδια *Lgi2-4* εμφανίζουν πρότυπο έκφρασης παρόμοιο με αυτό του *Lgi1*, είναι δυνατόν να θεωρηθούν ως υποψήφια για περιπτώσεις οικογενειών, που παρουσιάζουν παρόμοια συμπτώματα και στις οποίες δεν ταυτοποιήθηκαν μεταλλαγές στο *LG1*. Η ADLTE εμφανίζει επικρατή τρόπο κληρονόμησης, γεγονός που μπορεί να σημαίνει ότι οφείλεται σε κέρδος λειτουργίας της πρωτεΐνης. Επίσης, και με βάση τα αποτελέσματα των Gu W *et al* (2005) όπου διάφορα μεταλλάγματα της ανασυνδυασμένης πρωτεΐνης LG1 συσσωρεύονται σε μεγάλο βαθμό ενδοκυτταρικά αντί να εκκρίνονται, μπορούμε να υποθέσουμε ότι μεταλλαγές στα άλλα παράλογα γονίδια μπορεί να οδηγούν στο ίδιο αποτέλεσμα με συνέπεια την εκδήλωση παρόμοιου κλινικού φαινοτύπου. Ωστόσο, η ανάλυση των κωδικών αλληλουχιών των παραλόγων γονιδίων σε 12 οικογένειες και σε 38 σποραδικές περιπτώσεις που εμφανίζουν κλινικά συμπτώματα παρόμοια με της ADLTE χωρίς να φέρουν μεταλλαγές στο *LG1*, δεν οδήγησε στην ταυτοποίηση μεταλλαγών (αδημοσίευτα αποτελέσματα Ayerdi-Izquierdo A *et al*). Τα αποτελέσματα αυτά δείχνουν ότι τα γονίδια *LG2-4* δεν σχετίζονται, τουλάχιστον σε μεγάλο βαθμό, με την ADLTE αν και η τελική απάντηση για συμμετοχή των παραλόγων στην ανάπτυξη της νόσου, απαιτεί περαιτέρω μελέτες σε μεγαλύτερο δείγμα. Με βάση την απουσία μεταλλαγών στα παράλογα γονίδια στην παραπάνω μελέτη και με το γεγονός ότι το πρότυπο έκφρασης των τεσσάρων γονιδίων φαίνεται να διαφοροποιείται σε αρκετές περιοχές του εγκεφάλου, μπορούμε να υποθέσουμε ότι η λειτουργία τους έχει αποκλίσει κατά την εξέλιξη. Αυτό είναι σε συμφωνία με αντίστοιχη πρόσφατη μελέτη στο zebrafish κατά την οποία εντοπίστηκαν σημαντικές διαφορές στις περιοχές μεταγραφικής

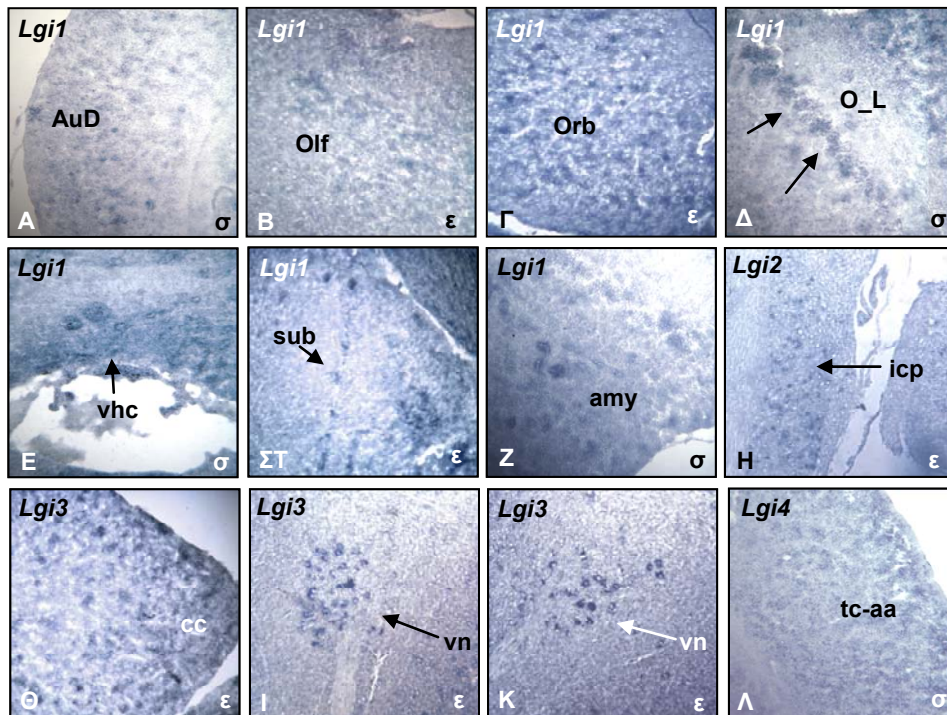
ενεργότητας μεταξύ των παραλόγων γονιδίων *Lgi* σε αυτόν τον οργανισμό (Gu W *et al*, 2005). Ωστόσο, και στο zebrafish παρατηρήθηκε έκφραση σε νευρικό ιστό. Ανοικτό παραμένει το ερώτημα εάν τα παράλογα γονίδια μπορούν να συσχετιστούν με άλλες μορφές ιδιοπαθούς επιληψίας. Η ανάλυση του προτύπου έκφρασής τους έδειξε ότι αυτά είναι μεταγραφικά ενεργά στον εγκέφαλο, απαραίτητη προϋπόθεση για να θεωρηθούν ως υποψήφια. Επιπλέον, σημαντικό είναι ότι μεταξύ των περιοχών όπου ανιχνεύτηκαν μετάγραφα του *Lgi4* συμπεριλαμβάνεται ο σωματοαισθητικός φλοιός, γεγονός που σε συνδυασμό με την χαρτογράφηση του γονιδίου αυτού στην κρίσιμη περιοχή της BFIC, θεωρούνται σημαντικές ενδείξεις ώστε να διερευνηθεί η παρουσία μεταλλαγών σ' αυτό, σε άτομα με την συγκεκριμένη διαταραχή.



Εικ. 34A: Συνέκφραση των γονιδίων *Lgi1-4* στον εγκέφαλο. Η συγκριτική μελέτη των προτύπων έκφρασης έδειξε ότι όλα τα γονίδια είναι μεταγραφικά ενεργά σε συγκεκριμένες περιοχές του κροταφικού λοβού: στον ενδορινικό φλοιό (*ent*, A1-A4) και στην οδοντωτή έλικα (DG: *dentate gyrus*, B1-B4). Επίσης διαπιστώθηκε συνέκφραση των γονιδίων στον σωματοαισθητικό φλοιό (S, Γ1-Γ4) του βρεγματικού λοβού και στην παρεγκεφαλίδα. Σημειώνεται η κοκκοειδής (*gr*) και η μοριακή (*mol*) στοιβάδα ενώ στο *Lgi4* (Δ4) το βέλος δείχνει τη στοιβάδα των κυττάρων Purkinje, όπου εντοπίζεται η έκφραση. Επιπλέον, συμβολίζονται οι στεφανιαίες (σ) και οι εγκάρσιες τομές (ε).



Εικ. 34B: Περιοχές υψηλής μεταγραφικής ενεργότητας των γονιδίων *Lgi1-4* στον εγκέφαλο. Συνέκφραση των γονιδίων *Lgi2* (A1), *Lgi3* (A2) και *Lgi4* (A3) στο διάφραγμα (septum), των *Lgi1* (B1), *Lgi2* (B2) και *Lgi4* (B3) στον απιοειδή φλοιό (piriform cortex), των *Lgi1* (Γ1), *Lgi2* (Γ2) και *Lgi3* (Γ3) στην περιοχή CA3 του ιπποκάμπου, των *Lgi1* (Δ1), *Lgi2* (Δ2) και *Lgi3* (Δ3) στον θάλαμο και των *Lgi3* (E1) και *Lgi4* (E2) στον κινητικό φλοιό σε στεφανιαίες (σ) και εγκάρσιες τομές (ε).



Εικ. 34Γ: Έκφραση των γονιδίων *Lgi1-4* σε διακριτές περιοχές του εγκεφάλου. Το γονίδιο *Lgi1* εκφράζεται στον ακουστικό (Α, AuD), τον οσφρητικό (Β, Olf) και τον οπτικό (Γ, Orb) φλοιό, στη στοιβάδα Oriens (Δ, O_L), τον κοιλιακό σύνδεσμο (Ε, vhc) και το υπόθεμα (ΣΤ, sub) του ιπποκάμπου, καθώς και στα αμύγδαλα (Ζ, amy). Υψηλή έκφραση του *Lgi2* ταυτοποιήθηκε στον κάτω παρεγκεφαλιδικό μίσχο (Η, icp), του *Lgi3* στον φλοιό του προσαγωγίου (Θ, cc) και σε αισθησιαίους πυρήνες (Ι, Κ, vn) και του *Lgi4* στην συνδετική περιοχή (association area) του κροταφικού φλοιού (Λ, tc-aa). Συμβολίζονται οι στεφανιαίες (σ) και οι εγκάρσιες τομές (ε).

B. Ταυτοποίηση και χαρακτηρισμός του γονιδίου που συνδέεται με την εύθραυστη θέση FRA10A

➤ Προσδιορισμός και κλωνοποίηση του cDNA του γονιδίου FRA10AC1

Η FRA10A ανήκει στις εύθραυστες χρωμοσωμικές θέσεις που επάγονται με καλλιέργεια των κυττάρων ελλείπει φυλλικού οξέος. Σε μοριακό επίπεδο έχουν χαρακτηριστεί 5 από τις 24 θέσεις αυτής της κατηγορίας (FRA16A, FRA11B, FRAXA, FRAXE, FRAXF, βλ. Εισαγωγή) και η έκφρασή τους έχει συσχετιστεί με την επέκταση της επανάληψης CGG. Με πιθανή εξαίρεση την FRA16A, οι υπόλοιπες εντοπίζονται μέσα σε γονίδια.

Με σκοπό τον εντοπισμό της FRA10A και του αντίστοιχου πιθανού γονιδίου, διερευνήθηκε το σύνολο της νουκλεοτιδικής αλληλουχίας του χρωμοσώματος 10 για την ταυτοποίηση επαναλήψεων τύπου (CGG)_n. Η διερεύνηση έγινε με ανιχνευτή την αλληλουχία (CGG)₈ χρησιμοποιώντας το πρόγραμμα blastn, χωρίς φιλτράρισμα των περιοχών χαμηλής πολυπλοκότητας και των επαναλαμβανόμενων αλληλουχιών. Με αυτόν τον τρόπο εντοπίστηκαν διάφοροι γενωμικοί κλώνοι και διαχωρίστηκαν βάσει του αριθμού των επαναλήψεων που περιείχαν (Πιν. 11). Η πρώτη ομάδα κλώνων έφερε την αλληλουχία (CGG)_{≥8}, η δεύτερη ομάδα την (CGG)₇, ενώ οι υπόλοιποι κλώνοι περιείχαν λιγότερες ή ατελείς επαναλήψεις. Ακολούθως, εξετάστηκε η θέση τους στο φυσικό χάρτη του χρωμοσώματος (Πιν. 11) με προτεραιότητα τους κλώνους των δύο πρώτων ομάδων.

Εφόσον η FRA10A είχε χαρτογραφηθεί στο 10q23.3, επιλέχθηκαν οι κλώνοι αυτής της περιοχής για περαιτέρω μελέτη: οι bA103A2, bA380G5 και bA165M8 που χαρτογραφούνται στο 10q23.31 και ο bA437J2 στο 10q23.33 (Εικ. 35).

Ο bA103A2 φέρει την αλληλουχία (CGG)₉ στον υποκινητή του γονιδίου *HTR7* (Εικ. 35) που ανήκει στην οικογένεια των υποδοχέων της σεροτονίνης. Ο *HTR7* εντοπίζεται τόσο στον εγκέφαλο όσο και στην περιφέρεια και έχει προταθεί ότι συμμετέχει στη λειτουργία της θερμορύθμισης, βάσει μελετών σε ποντίκια knockout (Hedlund PB, 2003). Παρά την θέση όπου χαρτογραφείται αυτή η επανάληψη, ο φαινότυπος των ποντικών με απενεργοποιημένο το *HTR7* είναι φυσιολογικός, τόσο μορφολογικά όσο και από άποψη συμπεριφοράς και συνεπώς δεν φαίνεται να υπάρχει σύνδεση της λειτουργίας αυτού του γονιδίου με τον φαινότυπο διανοητικής καθυστέρησης που έχει συσχετιστεί με την FRA10A.

Ο bA380G5 φέρει τις επαναλήψεις (CGG)₈, (CGG)₆ και (CGG)₅ στο 5'UTR του γονιδίου *PTEN* (Εικ. 35). Οι δύο πρώτες είναι ατελείς καθώς παρεμβάλλονται οι αλληλουχίες CCG και CAG αντίστοιχα, ενώ για την πρώτη έχει δειχθεί ότι είναι πολυμορφική με ποσοστό ετεροζυγωτίας 27% (Wren JD *et al*, 2000). Το *PTEN* είναι

ένα ογκοκατασταλτικό γονίδιο και μεταλλαγές του έχουν ταυτοποιηθεί σε σποραδικούς καρκίνους γλοιοβλαστώματος ενδομητρίου και προστάτη, καθώς και σε κληρονομήσιμα νοσήματα (νόσος Cowden, σύνδρομο Bannayan-Zonana, νόσος Lhermitte-Duclos) που χαρακτηρίζονται από πολλαπλά αμαρτώματα και αυξημένη προδιάθεση εμφάνισης καρκίνου. Η πρωτεΐνη έχει δράση φωσφατάσης και συμμετέχει σε σηματοδοτικά μονοπάτια που ρυθμίζουν βασικές κυτταρικές λειτουργίες, όπως τον κυτταρικό κύκλο, την απόπτωση και την διαφοροποίηση. Τα ομόζυγα knockout ποντίκια πεθαίνουν πολύ νωρίς κατά την εμβρυογένεση, ενώ τα ετερόζυγα αναπτύσσουν ένα ευρύ φάσμα όγκων (μαστού, θυρεοειδή, ενδομητρίου, προστάτη, T-λεμφοκυττάρων). Η βιοχημική λειτουργία του *PTEN* και ο φαινότυπος (υπερπλασία, αταξία, επιληπτικές κρίσεις, πρόωρος θάνατος) των ποντικών με απενεργοποιημένο το γονίδιο ειδικά σε συγκεκριμένες περιοχές του εγκεφάλου δεν μπορεί μάλλον να συσχετιστεί με το φαινότυπο των φορέων της FRA10A (Kishimoto H *et al*, 2003; Stiles B *et al*, 2004). Ο bA165M8 είναι αλληλοεπικαλυπτόμενος του bA380G5 και περιέχει τις ίδιες επαναλήψεις (Εικ. 35).

Ο bA437J2 φέρει την αλληλουχία (CGG)₈ στο μέσο του και περιλαμβάνει το 5' άκρο (τα δύο πρώτα εξόνια) του γονιδίου *LG11* και το 3' άκρο (τα επτά τελευταία εξόνια) του γονιδίου *PDE6C* (Εικ. 35). Για να ταυτοποιηθεί ο κλώνος BAC που προσδιορίζει την FRA10A, έγινε ανάλυση FISH στους ανωτέρω κλώνους από την συνεργαζόμενη ομάδα του Dr. R. Seibert. Το αποτέλεσμα της ανάλυσης υπέδειξε ως θετικό τον κλώνο bA437J2 (βλ. Συζήτηση).

Καθώς δεν είχε ταυτοποιηθεί κάποιο γονίδιο παράπλευρα της επανάληψης, αξιοποιήθηκαν οι διαθέσιμες αλληλουχίες των ESTs για να εξεταστεί η ύπαρξη ανοικτού πλαισίου ανάγνωσης στην περιοχή. Συγκεκριμένα, διερευνήθηκε η βάση δεδομένων dbEST του ανθρώπου χρησιμοποιώντας ως ανιχνευτή την αλληλουχία του κλώνου εκατέρωθεν της επανάληψης με το πρόγραμμα blastn. Το αποτέλεσμα ήταν η ταυτοποίηση 9 αλληλοεπικαλυπτόμενων ESTs (BM272089, AA114024, BF680539, BF794692, CB266740, BF033688, BI763499, AL702137 και BM545030) που χρησιμοποιήθηκαν για επαναληπτικό κύκλο διερεύνησης. Με αυτόν τον τρόπο εμπλουτίστηκε η συλλογή των ESTs και από την στοίχισή τους κατασκευάστηκε ένα πλήρους μήκους cDNA (Εικ. 36), 1344bp που περιέχει ένα ανοιχτό πλαίσιο ανάγνωσης 945bp. Με βάση τον κανόνα Kozak (Kozak M, 1986) το κωδικόνιο έναρξης της μετάφρασης εντοπίζεται στο nt:195 και το κωδικόνιο λήξης στο nt:1140 (Εικ. 37). Η επανάληψη (CGG)_n εντοπίζεται στο 5'UTR του cDNA ενώ η ανάλυση της αλληλουχίας των ESTs υπέδειξε την πιθανή πολυμορφικότητά της. Το 5' UTR προσδιορίστηκε βάσει του EST BF794692 και το μέγεθος του εκτιμάται σε ≥194bp. Το 3' UTR έχει μέγεθος 202bp και από την στοίχιση των ESTs φαίνεται πως φέρει

δύο λειτουργικά εναλλακτικά σήματα πολυαδενυλίωσης, στο nt:1265 (AATAAA) και στο nt:1320 (ATTAAG), Εικ.36. Τα ESTs που χρησιμοποιούν το σήμα AATAAA είναι τριπλάσια αυτών που χρησιμοποιούν το εναλλακτικό, φαινόμενο που έχει παρατηρηθεί στα περισσότερα mRNAs του ανθρώπου που φέρουν πολλαπλά σήματα πολυαδενυλίωσης, γεγονός που πιθανώς αντικατοπτρίζει την μεγαλύτερη σταθερότητα ή την αποτελεσματικότερη ωρίμανση των μορίων με κανονικό σήμα (Beaudoing E *et al*, 2000).

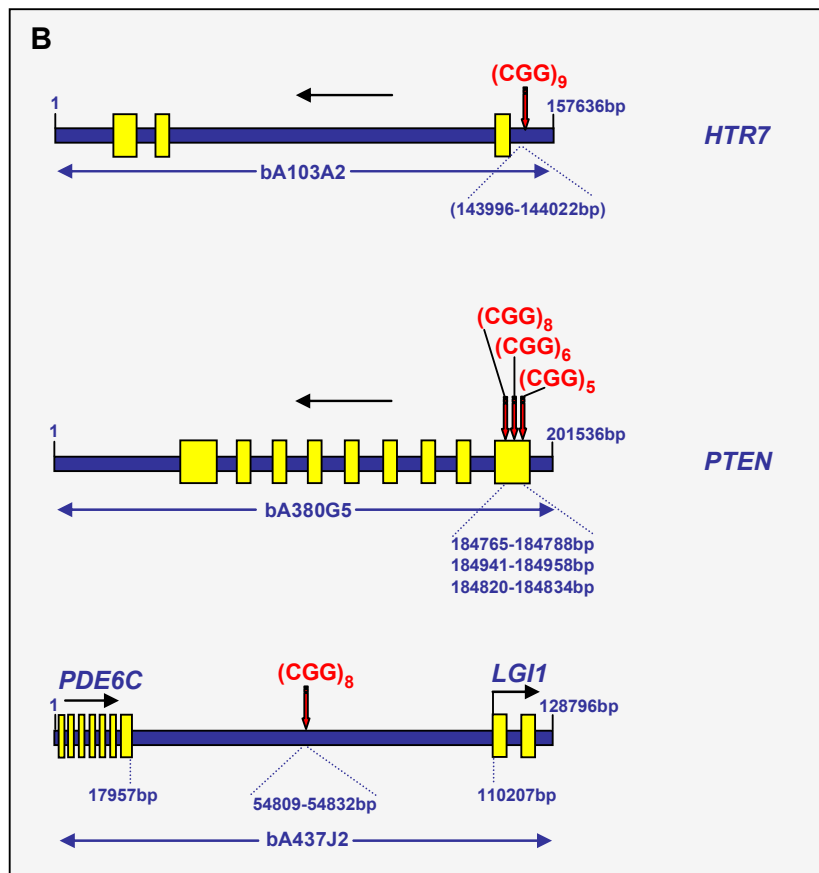
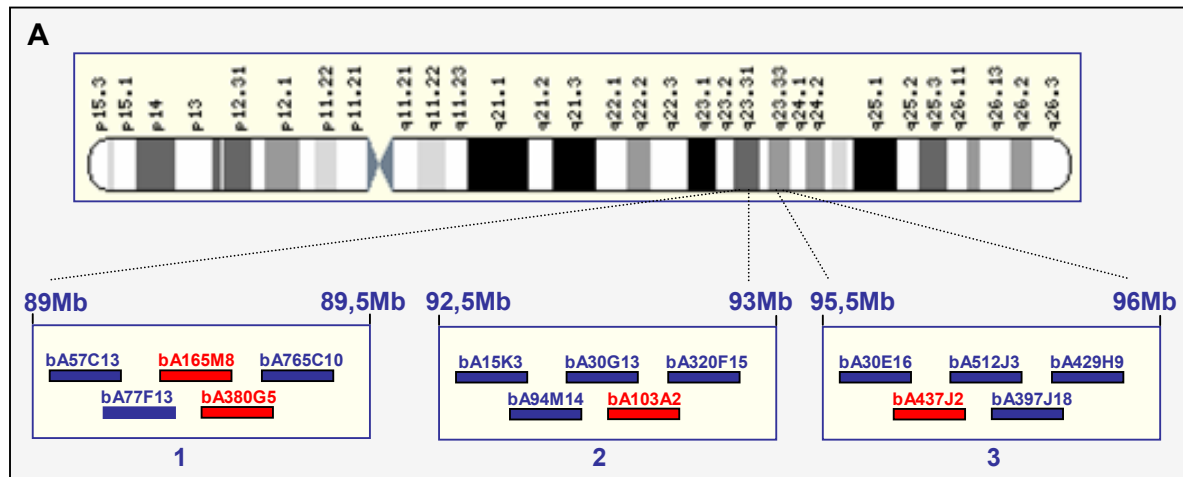
Πιν. 11: Γενωμικοί κλώνοι του χρωμοσώματος 10 που περιέχουν την επανάληψη CGG.

ΟΜΑΔΑ 1	ΧΡΩΜΟΣΩΜΙΚΗ ΘΕΣΗ
bA406G10	q26.11
bA96B5	q11.23-q21.1
bA506M13	q22.3
bA567J24	q26.11
bA342M3	q22.3
bA19C6	q24.2
bA103A2*	q23.31
bA273H7	q26.3
bA437J2*	q23.33
bA85A1	q24.2
bA418C1	p12.31
bA149I8	p13
bA466G24	p12.1
bA298J20	q26.13

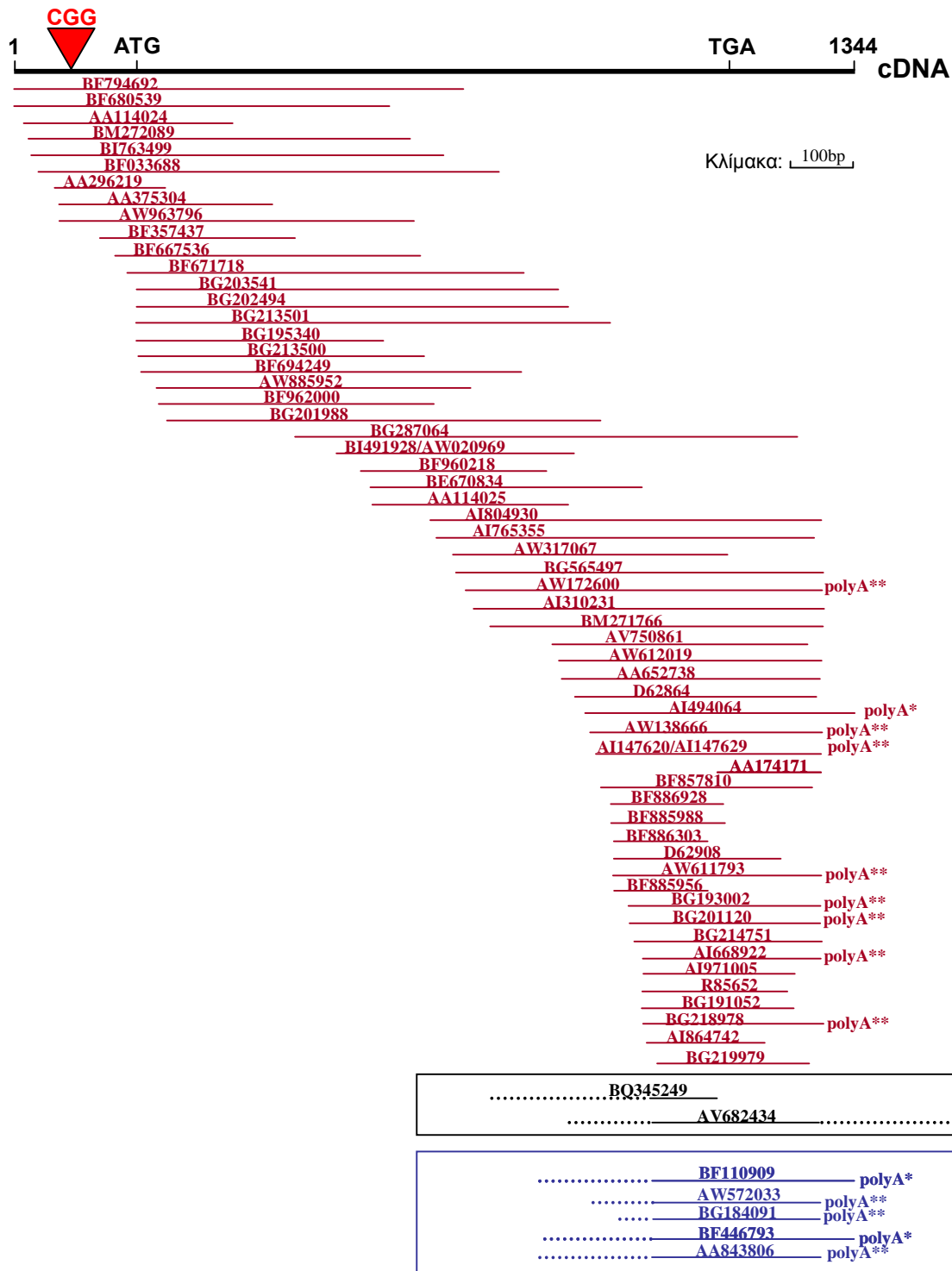
ΟΜΑΔΑ 2	ΧΡΩΜΟΣΩΜΙΚΗ ΘΕΣΗ
bA131C15	q22.3
bA36D19	q22.3
bA529C24	q22.3

ΟΜΑΔΑ 3	ΧΡΩΜΟΣΩΜΙΚΗ ΘΕΣΗ
bA428P16	q22.3
bA380G5*	q23.31
bA399K21	q22.2
bA86E10	q25.3
bA165M8*	q23.31
bA445K13	p13
bA161K20	p12.33

Σημ. :Η πρώτη ομάδα κλώνων φέρει την επανάληψη τουλάχιστον 8 φορές, η επόμενη ομάδα 7 φορές, ενώ στους κλώνους της τελευταίας ομάδας η επανάληψη υπάρχει 6 φορές ή είναι ατελής. Η θέση των κλώνων (δεξιά στήλη) προσδιορίστηκε βάσει του φυσικού χάρτη. Με αστερίσκο (*) σημειώνονται οι κλώνοι που χαρτογραφούνται στην ζώνη q23.3.



Εικ. 35: Χαρτογράφηση των γενωμικών κλώνων στο χρωμόσωμα 10 και της επανάληψης (CGG)_n σε αυτούς. Α. Απεικονίζονται τρία τμήματα του φυσικού χάρτη (1, 2, 3,) μήκους 0,5Mb το καθένα, που περιέχουν τους κλώνους bA437J2, bA103A2, bA165M8 και bA380G5 (με κόκκινο). Πάνω από το κάθε τμήμα σημειώνονται οι συντεταγμένες σε Mb. Συμβολίζονται μόνο οι κλώνοι που συνιστούν το tiling path. **Β.** Γενωμική οργάνωση των γονιδίων *PTEN* και *HTR7* στους αντίστοιχους κλώνους και τμημάτων των *LGI1* και *PDE6C* στον bA437J2 (τα παραλληλόγραμμα αντιστοιχούν στα εξόνια). Με κόκκινα βέλη σημειώνεται η επανάληψη CGG και τα νουκλεοτίδια όπου εντοπίζεται. Τα μαύρα βέλη δείχνουν την φορά μεταγραφής.



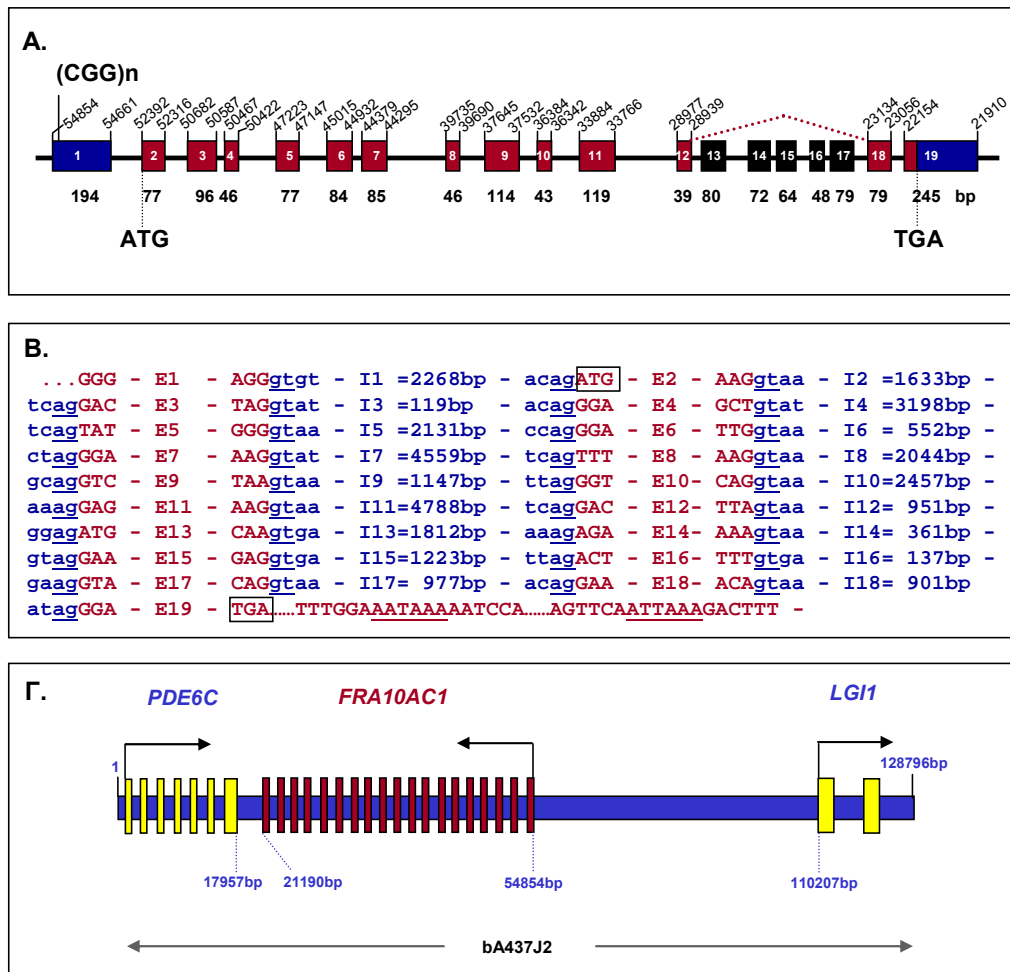
Εικ. 36: Κατασκευή του cDNA πλήρους μήκους *in silico*. Παρουσιάζεται το cDNA πλήρους μήκους (μαύρη γραμμή) και τα ESTs που χρησιμοποιήθηκαν για την κατασκευή του (κόκκινες γραμμές) τα οποία έχουν τοποθετηθεί βάσει της σχετικής τους αλληλοεπικάλυψης. Δεκατέσσερα ESTs φέρουν ουρά polyA και παρέχουν ένδειξη για ύπαρξη δύο εναλλακτικών σημάτων: στο nt:1289 (***) και στο nt:1344 (*). Στο μαύρο πλαίσιο απεικονίζονται τα ESTs AV682434 και BQ345249 που πιθανώς αντιστοιχούν σε εναλλακτικά μετάγγραφα (βλ. παρ.2.2.6.). Τα ESTs στο μπλε πλαίσιο δεν χρησιμοποιήθηκαν στην κατασκευή του cDNA, καθώς θεωρήθηκαν προϊόντα ατελούς ωρίμανσης του πρόδρομου mRNA.

Για την κλωνοποίηση του cDNA σχεδιάστηκαν οι εκκινητές F και R4 (Εικ. 37) και πραγματοποιήθηκε PCR (συνθήκες: 94°C 5min/ 94°C 45sec, 52°C 45sec, 72°C 1min, για 5 κύκλους/ 94°C 45sec, 57°C 45sec, 72°C 1min, για 30 κύκλους/ 72°C 5min), χρησιμοποιώντας ως υπόστρωμα 5ng προϊόντος αντίδρασης μεταγραφής από όρχεις ανθρώπου (MTC, Clontech). Το προϊόν είχε μέγεθος 949bp και περιλάμβανε όλο το ανοικτό πλαίσιο ανάγνωσης (πλην του τελευταίου κωδικονίου και του κωδικονίου λήξης). Μετά την απομόνωσή του από πήκτωμα αγαρόζης, ακολούθησε επαναληπτική PCR και κλωνοποίηση στον φορέα pGEM-T Easy όπου και προσδιορίστηκε η αλληλουχία του DNA, που επιβεβαίωσε την πρόβλεψη *in silico*. Η στοίχιση των δύο αλληλουχιών δείχνει πέντε νουκλεοτιδικές αλλαγές, που πιθανώς αντιστοιχούν σε πολυμορφισμό νουκλεοτιδίου (SNPs), τις T284C, A869G, G241A, C427G και G435A (Εικ. 37). Οι δύο πρώτες δεν οδηγούν σε αμινοξική αλλαγή ενώ οι υπόλοιπες προκαλούν τις αντικαταστάσεις R16H, T78R και V81I αντίστοιχα οι οποίες συναντώνται τουλάχιστον σε ένα ορθόλογο μόριο (βλ. παρακάτω). Από τις παραπάνω νουκλεοτιδικές αλλαγές, οι T284C, G241A και C427G υπάρχουν καταχωρημένες ως SNPs και στην αντίστοιχη βάση δεδομένων (<http://www.ncbi.nlm.nih.gov/projects/SNP/>) με κωδικούς πρόσβασης rs11187597, rs726817 και rs2275438 αντίστοιχα.

Το cDNA κατατέθηκε στη βάση δεδομένων Genbank με κωδικό πρόσβασης AJ431721 και ονομάστηκε *FRA10AC1*, λόγω της χαρτογράφησης του στην εύθραυστη θέση FRA10A.

➤ **Προσδιορισμός της γενωμικής οργάνωσης του *FRA10AC1***

Ο προσδιορισμός της γενωμικής οργάνωσης του *FRA10AC1* πραγματοποιήθηκε με στοίχιση του cDNA με την γενωμική αλληλουχία του bA437J2 (Genbank ID. AL157396) χρησιμοποιώντας το πρόγραμμα Blast 2 sequences. Από την ανάλυση προέκυψε ότι το γονίδιο εκτείνεται περί τα 33kb και ότι το cDNA στοιχίζεται με τη γενωμική αλληλουχία σε 14 τμήματα (Εικ. 38) που αντιστοιχούν στα εξόνια 1-12 και 18-19. Η αρίθμηση αυτή χρησιμοποιείται εδώ για να είναι συμβατή με τα επιπλέον, εναλλακτικά εξόνια που ταυτοποιήθηκαν πειραματικά (βλ. παρακάτω). Το μέγεθος των εξονίων κυμαίνεται μεταξύ 39 και 245bp ενώ των ιντρονίων μεταξύ 119 και 4788bp. Όλα τα σημεία συναρμογής εξονίων-ιντρονίων φέρουν την συντηρημένη αλληλουχία ag/gt. Το κωδικόνιο έναρξης της μετάφρασης εντοπίζεται στην αρχή του δευτέρου εξονίου ενώ το κωδικόνιο λήξης βρίσκεται στο τελευταίο (#19) εξόνιο. Η επανάληψη (CGG)_n χαρτογραφείται στο εξόνιο 1 που αντιστοιχεί στο 5' UTR.

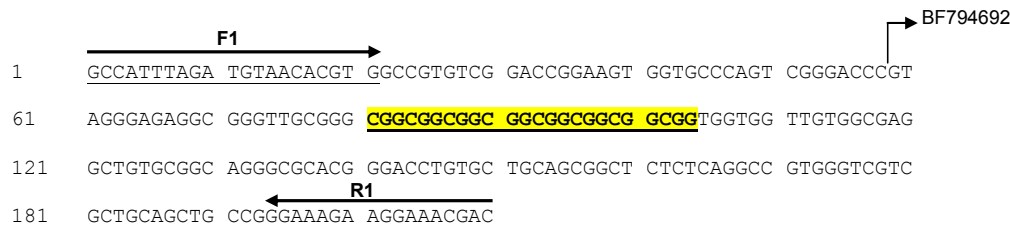


Εικ. 38: Γενωμική οργάνωση του γονιδίου *FRA10AC1*. **A.** Το γονίδιο περιλαμβάνει 19 εξόνια (κουτιά). Στο κύριο μετάγραφο το εξόνιο 12 συνδέεται με το εξόνιο 18, ενώ τα εξόνια 13-17 (μαύρα) περιλαμβάνονται μόνο στα εναλλακτικά μετάγραφα (παρ. 2.2.6.). Η επανάληψη CGG χαρτογραφείται στο 5'UTR (μπλε). Επίσης αναγράφεται το μήκος των εξονίων και τα νουκλεοτίδια της γενωμικής αλληλουχίας AL157396 (αντίστροφη συμπληρωματική της κωδικής αλυσίδας) με τα οποία στοιχίζονται. **B.** Οι αλληλουχίες στα όρια εξονίων (E1-E19, κεφαλαία γράμματα)-ιντρονίων (I1-I18, μικρά γράμματα) είναι συντηρημένες (ag-gt). Επιπλέον, αναγράφεται το μήκος κάθε ιντρονίου ενώ έχουν υπογραμμιστεί τα δύο πιθανά εναλλακτικά σήματα πολυαδενυλίωσης. **Γ.** Χρωμοσωμική οργάνωση του γονιδίου *FRA10AC1* και των γειτονικών του *LGI1* και *PDE6C*. Τα μαύρα βέλη δείχνουν την φορά της μεταγραφής.

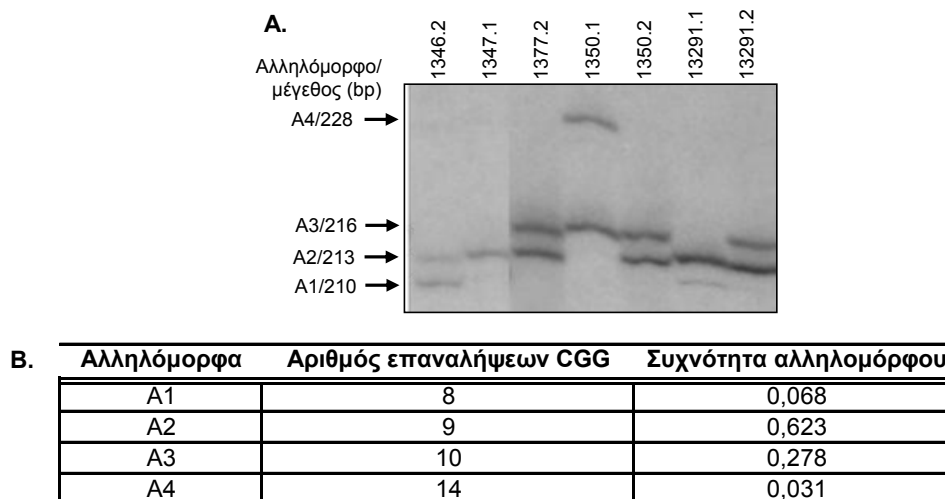
➤ **Γονοτυπική ανάλυση της τρινουκλεοτιδικής επανάληψης CGG σε άτομα ελέγχου και σε φορείς της *FRA10A***

Η πολυμορφικότητα της επανάληψης CGG σε γονίδια, αποτελεί ένδειξη δυνητικής επέκτασής της. Για το γονίδιο *FRA10AC1* αυτό εξετάστηκε με γονοτυπική ανάλυση DNA σε 81 μη συγγενή άτομα από 38 οικογένειες αναφοράς CEPH. Η ανάλυση έγινε με PCR χρησιμοποιώντας τους εκκινητές F1 (ραδιοσημασμένος) και R1 που υβριδοποιούνται εκατέρωθεν της αλληλουχίας (CGG)_n του γονιδίου (Εικ. 39). Τα προϊόντα της αντίδρασης διαχωρίστηκαν με ηλεκτροφόρηση σε αποδιατακτικό πήκτωμα 6% ακρυλαμίδης-ουρίας και ταυτοποιήθηκαν με αυτοραδιογραφία. Από την

ανάλυση προέκυψε ότι η επανάληψη είναι πράγματι πολυμορφική στον πληθυσμό (Πιν. 12) και ταυτοποιήθηκαν 4 αλληλόμορφα (A1-A4) μεγέθους 210, 213, 216 και 228bp, τα οποία φέρουν 8, 9, 10 και 14 CGG αντίστοιχα (Εικ. 40). Βάσει των αποτελεσμάτων, υπολογίστηκε η συχνότητα κάθε αλληλομόρφου (Εικ. 40) και προσδιορίστηκε ότι το πιο σύνηθες είναι το A2, που φέρει την αλληλουχία (CGG)₉, ενώ το ποσοστό ετεροζυγωτίας ανέρχεται σε 47%. Η γονοτυπική ανάλυση της επανάληψης σε φορείς της FRA10A έδειξε ότι στα άτομα αυτά η επανάληψη (CGG)_n επεκτείνεται δραματικά, γεγονός που οδηγεί στην υπερμεθυλίωση της περιοχής και την μεταγραφική καταστολή του αντίστοιχου αλληλομόρφου του γονιδίου (αποτελέσματα από τα συνεργαζόμενα εργαστήρια των Dr. R. Seibert και Jozef Gecz, Sarafidou T *et al*, 2004, βλ. Συζήτηση).



Εικ. 39: Θέση υβριδοποίησης των εκκινήτων F1 και R1 στο γενωμικό DNA εκατέρωθεν της επανάληψης CGG. Η PCR σε γενωμικό DNA με τους εκκινήτες F1 και R1 (βέλη) παράγει προϊόν μεγέθους 210bp, βάσει του αριθμού των επαναλήψεων (κίτρινη σκίαση) που φέρει η γενωμική αλληλουχία AL157396 (Εικ. 2). Οι συνθήκες που χρησιμοποιήθηκαν κατά την αντίδραση είναι: 95°C 5min/ 95°C 45sec, 56°C 45sec, 72°C 35sec, για 35 κύκλους/ 72°C 3min. Σημειώνεται η θέση από όπου ξεκινάει το EST BF794692 βάσει του οποίου καθορίστηκε το μέγεθος της 5'UTR (Εικ. 2).



Εικ. 40: Γονοτυπική ανάλυση σε δείγματα DNA οικογενειών CEPH για την μελέτη του πολυμορφισμού της επανάληψης CGG. **A.** Στην αυτοραδιογραφία φαίνονται τα 4 αλληλόμορφα (A1-A4) που ανιχνεύθηκαν (μέγεθος 210, 213, 216 και 228bp, αντίστοιχα). Παρατίθενται τα αποτελέσματα από 7 αντιπροσωπευτικά άτομα (ο κωδικός κάθε ατόμου συμβολίζεται στο πάνω μέρος) **B.** Η συχνότητα των 4 αλληλομόρφων και ο αντίστοιχος αριθμός, όπως προέκυψε από τη γονοτυπική ανάλυση (συνεργασία με M. Κοκκινάκη).

Πιν. 12: Γονοτυπική ανάλυση για την μελέτη του πολυμορφισμού της επανάληψης (CGG)_n.

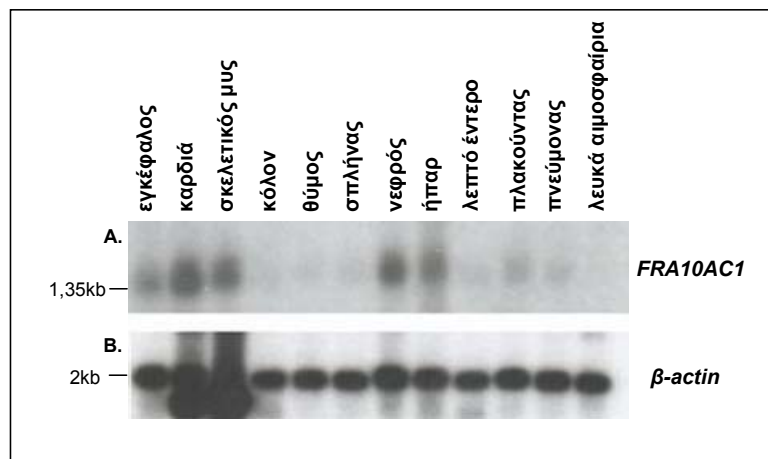
ΑΤΟΜΟ	ΑΛΛΗΛΟΜΟΡΦΑ	ΑΤΟΜΟ	ΑΛΛΗΛΟΜΟΡΦΑ	ΑΤΟΜΟ	ΑΛΛΗΛΟΜΟΡΦΑ			
102.1*	A2	A3	1334.2	A2	A2	1421.1	A4	A4
102.2	A2	A2	1340.1	A2	A2	1421.2*	A2	A4
104.13	A2	A2	1340.2*	A2	A3	1423.11*	A2	A3
104.14	A2	A2	1341.1*	A1	A3	1423.13	A2	A2
104.15	A2	A2	1341.13	A2	A2	1423.14	A2	A2
104.16	A2	A2	1341.14*	A2	A3	1424.1	A3	A3
12.1	A2	A2	1344.1*	A2	A3	1424.2	A2	A2
12.2	A2	A2	1344.2*	A2	A3	17.1	A1	A1
13291.1	A2	A2	1345.1*	A2	A3	17.2*	A2	A3
13291.2*	A2	A3	1345.2*	A2	A3	2.18*	A2	A3
13292.1	A2	A2	1346.1*	A2	A3	2.2	A2	A2
13292.2	A2	A2	1346.2*	A1	A2	21.1*	A2	A3
13293.1	A2	A2	1347.1	A2	A2	21.2*	A2	A3
13293.2	A2	A2	1347.2*	A2	A3	23.1	A1	A1
13294.1	A2	A2	1350.1*	A3	A4	23.2*	A1	A2
13294.2	A2	A2	1350.2*	A2	A3	28.1*	A2	A3
1331.12	A2	A2	1375.1	A2	A2	28.2*	A1	A2
1331.13	A3	A3	1375.2	A2	A2	35.1	A2	A2
1331.14	A3	A3	1377.1*	A1	A3	35.2*	A2	A3
1331.15	A2	A2	1377.2*	A2	A3	37.1	A2	A2
1332.13*	A2	A3	1408.1	A2	A2	37.2*	A2	A3
1332.14*	A2	A3	1408.2*	A2	A3	45.1	A2	A2
1332.15	A2	A2	1413.2*	A2	A3	45.2*	A1	A3
1332.16*	A2	A3	1416.2	A2	A2	66.1*	A1	A3
1333.12*	A2	A3	1418.1*	A2	A4	66.2	A3	A3
1333.13	A2	A2	1418.2	A2	A2	884.1	A3	A3
1334.1*	A2	A3	1420.1*	A2	A3	884.2	A3	A3

Σημ.: Κάθε άτομο συμβολίζεται με δύο αριθμούς που χωρίζονται με τελεία. Ο πρώτος αντιστοιχεί στην οικογένεια και ο δεύτερος σε ένα συγκεκριμένο άτομο της οικογένειας αυτής. Δίπλα σε κάθε άτομο φαίνονται τα αλληλόμορφα (A1, A2, A3, A4) που φέρει. Από τα 81 άτομα που εξετάστηκαν, τα 37 ήταν ετερόζυγα (*), που αντιστοιχεί σε ποσοστό 47%.

➤ **Το πρότυπο έκφρασης του γονιδίου *FRA10AC1***

Ανάλυση Northern

Για την ταυτοποίηση των ιστών όπου το γονίδιο είναι μεταγραφικά ενεργό και τον παράλληλο προσδιορισμό του μεγέθους του μεταγράφου, πραγματοποιήθηκε ανάλυση Northern σε δείγματα από διάφορους ιστούς. Ως ανιχνευτής χρησιμοποιήθηκε τμήμα του τελευταίου εξονίου (nt: 1124-1252), μεγέθους 129bp, που κατά το μεγαλύτερο μέρος του αντιστοιχεί στην 3' UTR ο οποίος παρήχθει με PCR ραδιοσήμανσης χρησιμοποιώντας τους εκκινητές F2-R2 (Εικ. 37). Από την ανάλυση προέκυψε ένα κύριο μετάγραφο μεγέθους ~1,4kb, που συμφωνεί με το εκτιμώμενο μέγεθος του cDNA (Εικ. 41). Το μετάγραφο ανιχνεύεται σε όλους τους ιστούς, γεγονός που υποδηλώνει καθολική έκφραση του *FRA10AC1*. Ιδιαίτερα έντονη έκφραση του γονιδίου παρατηρήθηκε στον εγκέφαλο, την καρδιά, τους σκελετικούς μύες, τα νεφρά και το ήπαρ ενώ χαμηλά επίπεδα έκφρασης εντοπίστηκαν στο θύμο, το σπλήνα, τον πλακούντα, τους πνεύμονες, το παχύ και το λεπτό έντερο. Στα λευκοκύτταρα εμφανίστηκε η χαμηλότερη έκφραση του γονιδίου, η οποία και ανιχνεύτηκε οριακά με τη μέθοδο αυτή.



Εικ. 41: Προσδιορισμός του προτύπου έκφρασης του γονιδίου *FRA10AC1* με ανάλυση κατά Northern. A. Με την ανάλυση ανιχνεύτηκε ένα κύριο μετάγραφο μεγέθους ~1,4kb που εμφανίζει μεγαλύτερη αφθονία σε εγκέφαλο, καρδιά, σκελετικό μυ, νεφρό και ήπαρ. Σημειώνεται η ζώνη μεγέθους 1,35kb του μάρτυρα. Ο ανιχνευτής για το *FRA10AC1* παρήχθη με PCR με τους εκκινητές F2 και R2 (Εικ. 4) με υπόστρωμα 50ng DNA από βιβλιοθήκη cDNA ήπατος ανθρώπου (βλ. Μεθόδους). Οι συνθήκες της αντίδρασης ήταν οι ακόλουθες: 95°C 5min/ 95°C 45sec, 56°C 45sec, 72°C 35sec επανάληψη 35 κύκλων/ 72°C 3min. Το προϊόν χρησιμοποιήθηκε ως υπόστρωμα για την παρασκευή του ανιχνευτή, με PCR ραδιοσήμανσης. Η υβριδοποίηση πραγματοποιήθηκε με 5x10⁵cpm ανιχνευτή ανά ml διαλύματος υβριδοποίησης, στους 66°C για 3h μετά από προυβριδοποίηση της μεμβράνης για 30min παρουσία φορέα-DNA με το διάλυμα υβριδοποίησης ExpressHybTM (Clontech). Η ειδική ενεργότητα του ανιχνευτή εκτιμήθηκε σε 10⁹cpm/μg. Τα πλυσίματα της μεμβράνης έγιναν με διάλυμα 2XSSC, 0,05% SDS. **B.** Ποιοτικός και ποσοτικός έλεγχος του RNA με επαναυβριδοποίηση με β-ακτίνη (σύμφωνα με τις οδηγίες της Clontech). Στα δείγματα της καρδιάς και του σκελετικού μυ εμφανίζονται δύο ζώνες που είναι πιθανότατα αποτέλεσμα υβριδοποίησης του ανιχνευτή με άλλες ισομορφές του mRNA, τις α ή γ.

Πειράματα RT- PCR

Το πρότυπο έκφρασης μελετήθηκε περαιτέρω με PCR σε αντιδράσεις RT, χρησιμοποιώντας RNA από μία ποικιλία ιστών ανθρώπου. Η PCR πραγματοποιήθηκε κάτω από μη-κορεσμένες συνθήκες έτσι ώστε να εκτιμηθούν συγκριτικά, κατά προσέγγιση, τα επίπεδα των μεταγράφων του *FRA10AC1* στους διάφορους ιστούς. Με την ανάλυση, αφενός επιβεβαιώθηκαν τα αποτελέσματα του Northern, αφετέρου διαπιστώθηκε ότι το γονίδιο είναι επίσης μεταγραφικά ενεργό σε πάγκρεας, όρχεις, ωθήκες και σε χαμηλότερα επίπεδα, στον προστάτη. Επίσης, επιβεβαιώθηκε η έκφραση του γονιδίου, σε χαμηλά επίπεδα, στα λευκοκύτταρα (Εικ. 42). Συμπερασματικά, και με τις δύο μεθόδους που εφαρμόστηκαν, παρατηρήθηκε καθολική μεταγραφική ενεργότητα του γονιδίου. Επιπλέον, η RT-PCR οδήγησε στον εντοπισμό, την κλωνοποίηση και την δομική ανάλυση εναλλακτικών μεταγράφων (βλ. παρακάτω)

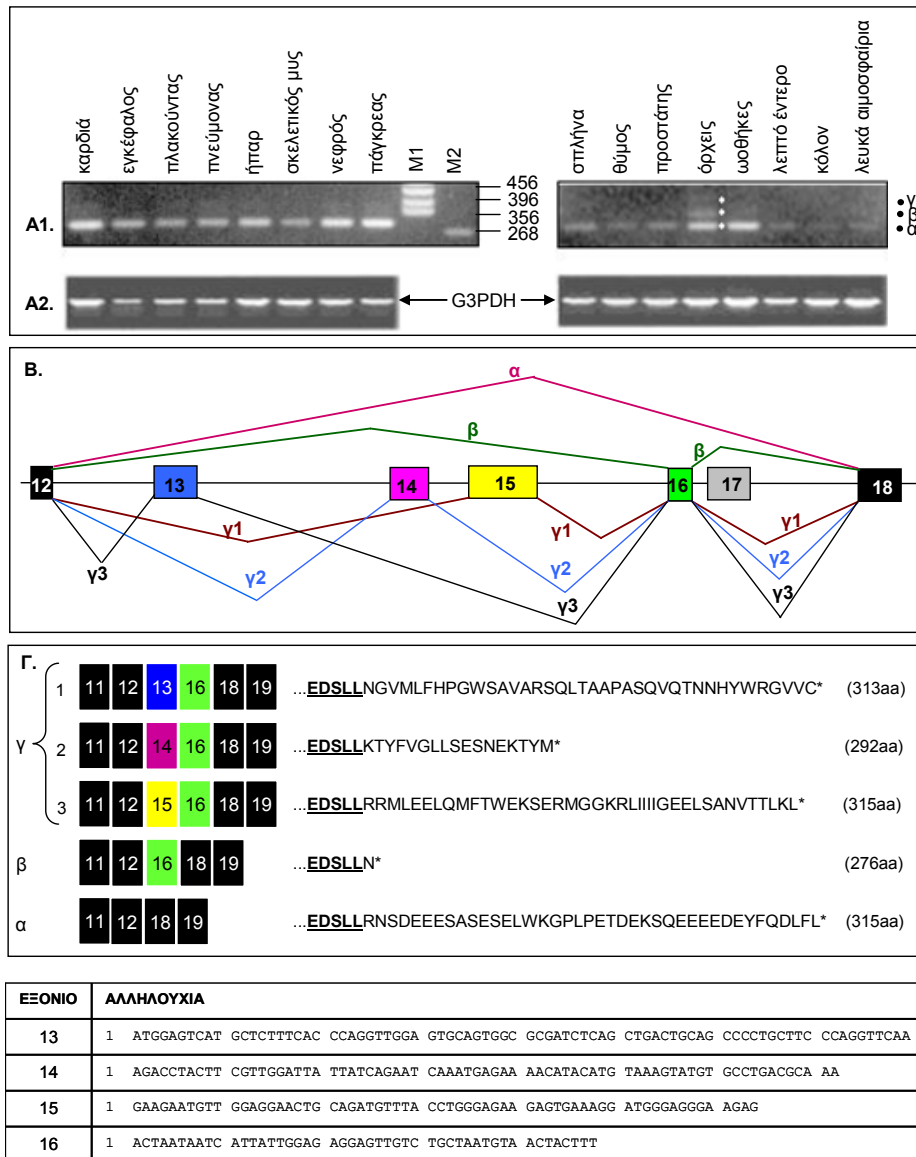
Μελέτη του προτύπου έκφρασης του *Fra10Ac1* κατά την εμβρυική ανάπτυξη και στον εγκέφαλο του ποντικού

Η διερεύνηση του προτύπου έκφρασης ενός γονιδίου και ο εντοπισμός των ιστών όπου παρατηρούνται υψηλά επίπεδα μεταγράφων παρέχει σημαντικές πληροφορίες για την ενδεχόμενη συμμετοχή του στη λειτουργία των αντίστοιχων οργάνων και τη συσχέτισή του με άλλα συνεκφραζόμενα γονίδια. Ως μοντέλο για τη μελέτη του προτύπου έκφρασης του γονιδίου *FRA10AC1*, σε επίπεδο ιστού, επιλέξαμε τον ποντικό. Για την πραγματοποίηση του πειράματος ταυτοποιήθηκε το ορθόλογο γονίδιο *Fra10ac1*.

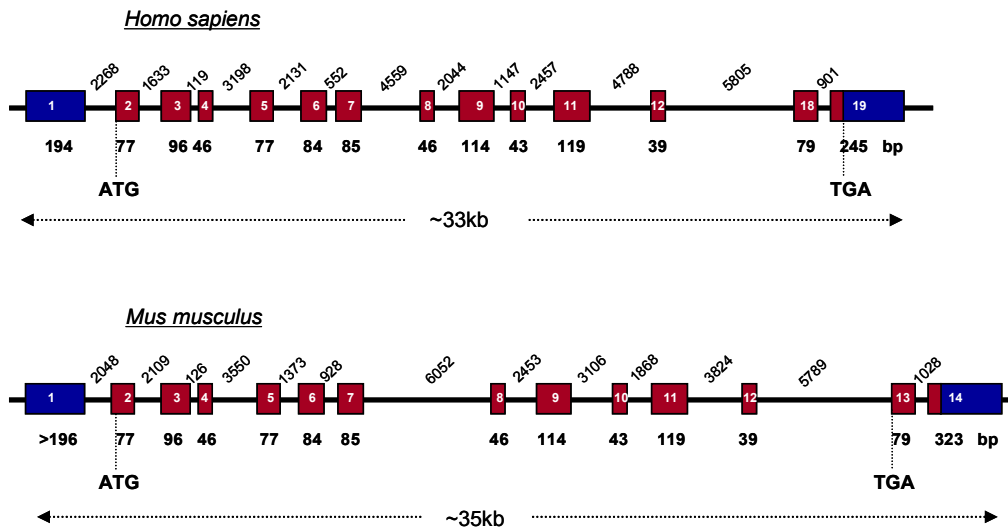
Ταυτοποίηση του ορθόλου γονιδίου *Fra10ac1* στον ποντικό

Με ανιχνευτή το cDNA του *FRA10AC1* του ανθρώπου, διερευνήσαμε την νουκλεοτιδική βάση δεδομένων του ποντικού (nr και dbEST) και ταυτοποιήσαμε διάφορα ESTs και τον γενωμικό κλώνο BAC RP24-93G10 του ποντικού. Από την στοίχιση των ESTs και με την μεθοδολογία που είχαμε ακολουθήσει και για το γονίδιο του ανθρώπου, κατασκευάσαμε *in silico* το cDNA του ποντικού. Το μήκος του είναι 1380bp και η αλληλουχία του καταχωρήθηκε στην βάση δεδομένων Genbank με κωδικό πρόσβασης BN000292. Η γενωμική οργάνωση καθορίστηκε με σύγκριση του cDNA με την αλληλουχία του γενωμικού κλώνου και από την ανάλυση διαπιστώθηκε ότι τα δύο ορθόλογα γονίδια έχουν παρόμοια δομικά χαρακτηριστικά (Εικ. 43). Ειδικότερα, τα μεγέθη των εξονίων είναι ακριβώς τα ίδια, εκτός από το εξόνιο 1 για το οποίο δεν μπορεί να καθοριστεί με ακρίβεια το 5' άκρο του και το τελευταίο εξόνιο, που είναι μικρότερο στον άνθρωπο. Επίσης, τα γονίδια έχουν παρόμοιο μήκος και χαρτογραφούνται σε συνταϊνικές περιοχές (10q23 του ανθρώπου και C3 του χρωμοσώματος 19 του ποντικού). Το κωδικόνιο έναρξης της μετάφρασης και στο

ορθόλογο γονίδιο του ποντικού χαρτογραφείται στην αρχή του εξονίου 2. Σημαντική συντήρηση παρατηρείται και στα ιντρόνια τα οποία, παρόλο που έχουν διαφορετικό μήκος στους δύο οργανισμούς, διατηρούν την σχετική αναλογία μεγεθών.



Εικ. 42: Προσδιορισμός του προτύπου έκφρασης του γονιδίου *FRA10AC1* με RT-PCR και κλωνοποίηση εναλλακτικών μεταγράφων. A1: μετάγραφα του γονιδίου ανιχνεύθηκαν σε όλους τους ιστούς που μελετήθηκαν με RT-PCR. Στην PCR χρησιμοποιήθηκε ως υπόστρωμα 5ng προϊόντος RT (Multiple Tissue cDNA rannels, Clontech) και οι εκκινητές F3 και R2 (Εικ.4) που υβριδοποιούνται στα εξόνια 11 και 19 αντίστοιχα. Οι συνθήκες της αντίδρασης ήταν: 94°C 5min/ 94°C 45sec, 57°C 45sec, 72°C 30sec επανάληψη 35 κύκλων/ 72oC 5min. Απεικονίζονται οι δείκτες μοριακών μεγεθών *pBluescript/Hinfl* (M1) και *λEcoRV* (M2). **A2.** PCR ελέγχου με εκκινητές (Clontech) για το cDNA του *G3PDH*. Η θερμοκρασία υβριδοποίησης ήταν 55°C και εκτελέστηκαν 27 κύκλοι πολλαπλασιασμού. **B.** Σχηματική απεικόνιση της εναλλακτικής συναρμογής του *FRA10AC1*. **Γ.** Σχηματική απεικόνιση των εναλλακτικών μεταγράφων α, β, γ1, γ2 και γ3 (αριστερά), των προβλεπόμενων ισομορφών του πολυπεπτιδίου (δεξιά) και των ισοηλεκτρικών σημείων των εναλλακτικών τμημάτων. Τα αμινοξέα που κωδικοποιούνται από το εξόνιο 12, που είναι κοινό σε όλα τα μεταγράφα του γονιδίου, είναι υπογραμμισμένα. Η αλληλουχία των εναλλακτικών εξονίων 13-17 παρουσιάζεται στον πίνακα στο κάτω μέρος της εικόνας. Το εξόνιο 17, παρόλο που δεν εντοπίστηκε σε κάποιο μετάγραφο, είχε προβλεφθεί *in silico*, βάσει του EST AV682434.

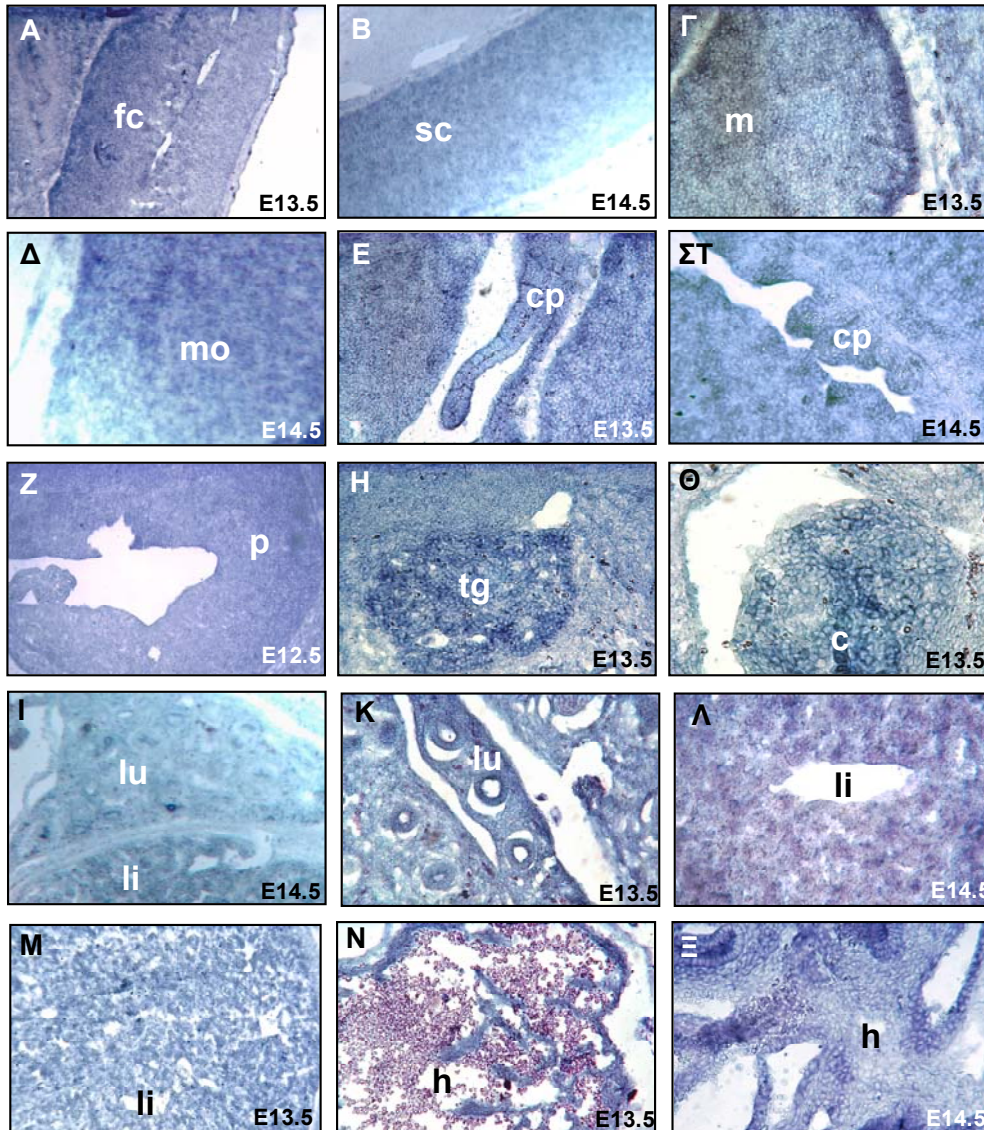


Εικ. 43: Συγκριτική γενωμική οργάνωση του γονιδίου *FRA10AC1* στον άνθρωπο και τον ποντικό.

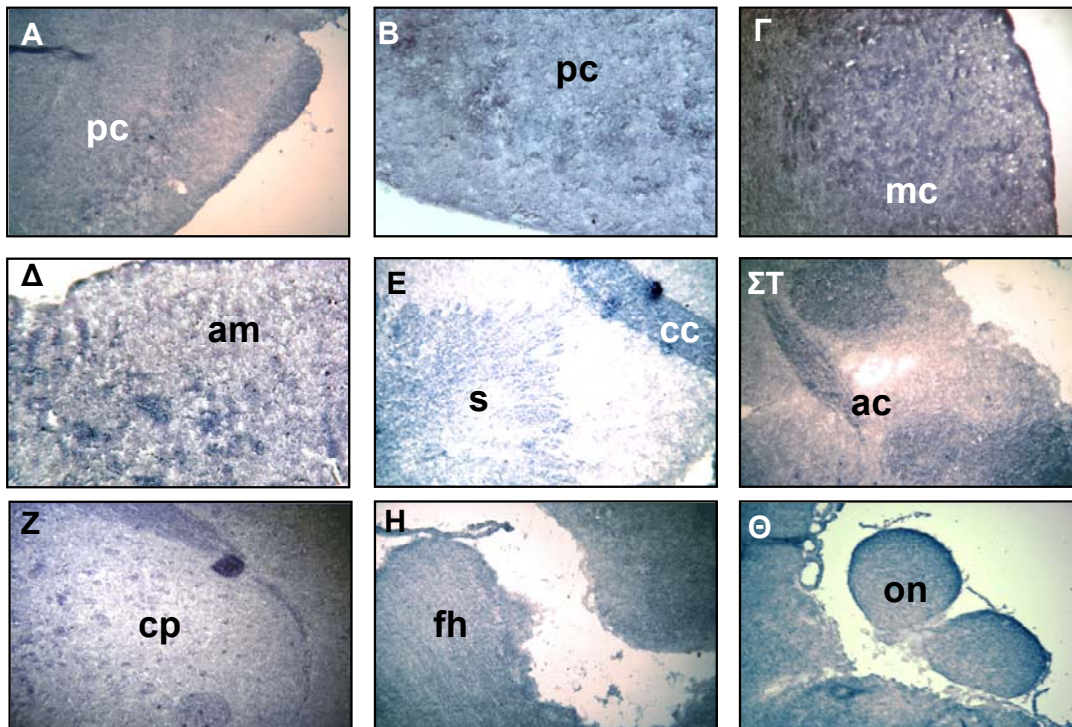
Η μελέτη του προτύπου έκφρασης του *Fra10ac1* πραγματοποιήθηκε με πειράματα υβριδοποίησης *in situ* σε οβελιαίες τομές ολόκληρων εμβρύων αναπτυξιακών σταδίων E12.5, E13.5 και E14.5 και σε στεφανιαίες τομές εγκεφάλου ενήλικου ποντικού. Για την παρασκευή του ριβοανιχνευτή αξιοποιήθηκε η αλληλουχία του cDNA και σχεδιάστηκαν οι εκκινητές mFRAF-mFRAR που υβριδοποιούνται στα εξόνια 3 και 9 αντίστοιχα. Μετά από PCR σε βιβλιοθήκη cDNA εμβρύου ποντικού αναπτυξιακού σταδίου E10.5 το αντίστοιχο προϊόν (~500bp) κλωνοποιήθηκε στον φορέα pBluescript KS+ (βλ. Πλασμιδιακές κατασκευές). Η πλασμιδιακή κατασκευή (*Fra10ac1/Bl.Script*) γραμμοποιήθηκε κατάλληλα και χρησιμοποιήθηκε ως υπόστρωμα για αντίδραση μεταγραφής *in vitro*, με την οποία παρασκευάστηκε ο συμπληρωματικός ανιχνευτής (antisense) (Εικ. 10). Για τον αρνητικό έλεγχο πραγματοποιήθηκαν πειράματα με ανιχνευτή RNA που ήταν συγγραμικός με το ενδογενές μετάγραφο καθώς και απουσία ανιχνευτή.

Από την μελέτη προέκυψε ότι το γονίδιο εμφανίζει ευρύτατη έκφραση τόσο στο αναπτυσσόμενο έμβryo όσο και στον εγκέφαλο του ενήλικου, παρουσιάζοντας ωστόσο υψηλότερη μεταγραφική ενεργότητα σε συγκεκριμένα όργανα και δομές. Συγκεκριμένα, στο έμβryo διαπιστώθηκε υψηλότερη έκφραση στο νευρικό ιστό όπως στον πρόσθιο φλοιό, στο εγκεφαλικό στέλεχος (προμήκη μυελό, μεσεγκέφαλος, γέφυρα), στο χοριοειδές πλέγμα, στο τρίδυμο νεύρο, κατά μήκος του νωτιαίου μυελού, στον κοχλία αλλά και σε εσωτερικά όργανα όπως σε καρδιακό ιστό, ήπαρ και πνεύμονες (κυρίως στο επιθήλιο των βρόγχων και των βρογχιολίων και λιγότερο στο εσωτερικό παρέγχυμα), Εικ. 44. Στον εγκέφαλο των ενήλικων ποντικών, υψηλά

επίπεδα μεταγράφων ανιχνεύτηκαν στον απιοειδή και τον κινητικό φλοιό, σε δύο από τις βασικότερες περιοχές του μεταιχμιακού συστήματος (παρυφές του ιπποκάμπου και αμυγδαλοειδές σώμα), στον πρόσθιο σύνδεσμο, το μεσολόβιο και το διάφραγμα (Εικ. 45). Η παρατηρούμενη έκφραση στο κεντρικό νευρικό σύστημα είναι ενδεικτική της συμμετοχής του γονιδίου στην ανάπτυξη και τη λειτουργία του εγκεφάλου ενώ η ευρύτερου φάσματος μεταγραφική ενεργότητα στα εσωτερικά όργανα, υποδηλώνει ενδεχομένως ένα σημαντικό ρόλο του γονιδίου σε βασικές κυτταρικές λειτουργίες τόσο κατά την ανάπτυξη του εμβρύου όσο και στο ενήλικο.



Εικ. 44: Προσδιορισμός του προτύπου έκφρασης του *Fra10ac1* με RNA *in situ* υβριδοποίηση σε οβελιαίες τομές εμβρύων ποντικού ηλικίας E12.5, E13.5 και E14.5. Υψηλή μεταγραφική ενεργότητα διαπιστώθηκε στον πρόσθιο φλοιό (fc:frontal cortex, A), στο νωτιαίο μυελό (sp:spinal cord, B), στον μεσεγκέφαλο (m:midbrain, Γ), στον προμήκη μυελό (mo:medulla oblongata, Δ), στο χοριοειδές πλέγμα (cp : choroid plexus, E και ΣΤ), στη γέφυρα (p:pons, Z), στο τρίδυμο νεύρο (tg : trigeminal nerve, H), στον κοχλία (c:cochlea, Θ), στους πνεύμονες (lu:lung, I και K), στο ήπαρ (li:liver, I, Λ και M) και στην καρδιά (h:heart, N και Ξ).



Εικ. 45: Προσδιορισμός του προτύπου έκφρασης του *Fra10ac1* με RNA *in situ* υβριδοποίηση σε στεφανιαίες τομές εγκεφάλου ενήλικου ποντικού. Υψηλότερη έκφραση του γονιδίου παρατηρήθηκε στον απιοειδή (pc:piriform cortex, A και B) και τον κινητικό φλοιό (mc: motor cortex, Γ), στα αμύγδαλα (am: amygdala, Δ), στο μεσολόβιο (cc: corpus callosum, E), στο διάφραγμα (s: septum pellucidum, E), στον πρόσθιο σύνδεσμο του εγκεφάλου (ac: anterior commissure, ΣΤ), στο ραβδωτό σώμα (cp: caudate putamen, Z), σε νευρικές ίνες του ιπποκάμπτου (fh:fibriia hippocampus, H) και στο οπτικό νεύρο (on: optical nerve, Θ).

➤ **Ανάλυση εναλλακτικών μεταγράφων και ισομορφές της *FRA10AC1***

Στην πλειονότητα των ιστών που εξετάστηκαν με RT-PCR, εκτός του κυρίου προϊόντος (ζώνη “α”, Εικ. 42Α1), ανιχνεύτηκε, σε μικρή ποσότητα, και ένα επιπρόσθετο μεγαλύτερο προϊόν (ζώνη “β”, Εικ. 42Α1). Επιπλέον, στο δείγμα των όρχεων εμφανίστηκε και ένα τρίτο προϊόν, ακόμη μεγαλύτερου μεγέθους (ζώνη “γ”, Εικ. 42Α1). Η παρουσία των επιπρόσθετων προϊόντων στην PCR πιθανώς υποδηλώνει την ύπαρξη εναλλακτικών μεταγράφων. Για να επιβεβαιωθεί η παραπάνω υπόθεση, οι ζώνες α, β και γ απομονώθηκαν από το πήκτωμα και χρησιμοποιήθηκαν για προσδιορισμό της νουκλεοτιδικής τους αλληλουχίας. Η αλληλούχηση έδειξε ότι η ζώνη α, που είναι και η επικρατέστερη, αντιστοιχεί στο cDNA που είχε κλωνοποιηθεί, στο οποίο το εξόνιο 12 συνδέεται άμεσα με το εξόνιο 18 (Εικ. 42). Το προϊόν β φέρει ένα επιπλέον εξόνιο, το 16, μήκους 48 bp, μεταξύ των εξονίων 12 και 18 (Εικ. 42). Το εξόνιο αυτό περιλαμβάνεται και στο EST BQ345249 σε σειρά με τα εξόνια 7 έως 12 και ακολούθως με το 18 όπου και σταματάει η διαθέσιμη αλληλουχία του (Εικ. 36). Η αλληλούχηση της ζώνης γ δεν ήταν δυνατή, ωστόσο από την αυτοραδιογραφία φάνηκε ότι πιθανότατα αυτό οφειλόταν σε

ετερογένεια του προϊόντος. Έτσι το προϊόν γ κλωνοποιήθηκε στον φορέα pGEM-T Easy και επιλέχθηκαν διαφορετικοί κλώνοι στους οποίους και έγινε προσδιορισμός της αλληλουχίας τους. Από την ανάλυση προέκυψε ότι το προϊόν γ της PCR πράγματι είναι ετερογενές, αποτελώντας μίγμα τριών, τουλάχιστον, εναλλακτικών μεταγράφων, των γ1, γ2 και γ3 (Εικ. 42). Και τα τρία μετάγραφα περιέχουν το εξόνιο 16 και ένα επιπρόσθετο εξόνιο: το μετάγραφο γ1 το εξόνιο 15 (64bp), το γ2 το εξόνιο 14 (72 bp) και το γ3 το εξόνιο 13 (80bp). Οι αλληλουχίες των εναλλακτικών εξονίων (Εικ. 42) κατατέθηκαν στην βάση δεδομένων Genbank με τους κωδικούς πρόσβασης AJ431725, AJ431724, AJ431723 και AJ431722 για τα μετάγραφα β, γ1, γ2 και γ3 αντίστοιχα. Τα προβλεπόμενα πολυπεπτίδια που παράγονται από τα πέντε εναλλακτικά μετάγραφα παρουσιάζουν αξιοσημείωτη ετερογένεια στο καρβοξυτελικό τους άκρο (Εικ.42) και σημαντική διαφορά στα φυσικοχημικά χαρακτηριστικά τους. Έτσι, ενώ το καρβοξυτελικό άκρο του κύριου πολυπεπτιδίου είναι ιδιαίτερα όξινο (pI 3,8), οι δύο μεγαλύτερες εναλλακτικές ισομορφές γ1 και γ3, εμφανίζονται βασικές έχοντας αντίστοιχα τιμές pI 9,5 και 8,5 αντίστοιχα. Η απομόνωση των εναλλακτικών μεταγράφων από όρχεις πιθανόν σχετίζεται με την λεπτή ρύθμιση της λειτουργίας του γονιδίου και συμφωνεί με παλαιότερες μελέτες όπου διαπιστώθηκαν πολύ υψηλά επίπεδα εναλλακτικών μεταγράφων σε αυτόν τον ιστό, πιθανότατα επειδή ο πολλαπλασιασμός και η διαφοροποίηση των γαμετικών κυττάρων απαιτεί δραματικές μορφολογικές αλλαγές και πολλαπλά ρυθμιστικά βήματα (Χυ Q *et al*, 2002; Yeo G *et al*, 2004).

Εκτός των μεταγράφων που απομονώθηκαν, υπάρχει ένδειξη για την παρουσία ενός επιπλέον βάσει του EST AV682434 (Εικ. 36) που περιλαμβάνει, εκτός των εξονίων 16, 18 και 19, τα nt:24121-24190 του γενωμικού κλώνου. Αυτά πιθανώς αντιστοιχούν στο εναλλακτικό εξόνιο 17 καθώς τα σημεία συναρμογής του φέρουν την συντηρημένη αλληλουχία ag/gt. Πιθανώς το μετάγραφο αυτό να περιλαμβάνονταν σε μικρότερη αναλογία στο μίγμα των μορίων που αποτελούσαν την ζώνη γ και η ταυτοποίησή του απαιτούσε απομόνωση επιπλέον κλώνων.

➤ **Δομικά χαρακτηριστικά της πρωτεΐνης FRA10AC1**

Στα πλαίσια της διερεύνησης *in silico* της δομής της πρωτεΐνης FRA10AC1, αξιοποιήσαμε μία σειρά βάσεων δεδομένων και υπολογιστικών προγραμμάτων με στόχο την αδρή περιγραφή των φυσικοχημικών χαρακτηριστικών της.

Το μοριακό βάρος της πρωτεΐνης FRA10AC1 είναι 37,55 kDa και το ισοηλεκτρικό σημείο 8,25, όπως προσδιορίστηκαν από το πρόγραμμα ProtParam. Το πολυπεπτίδιο είναι ιδιαίτερα πλούσιο σε λυσίνες και σε γλουταμικό οξύ με ποσοστά 15,2% και 12,7% αντίστοιχα (πρόγραμμα ProtParam). Η πιο βασική περιοχή

βρίσκεται μεταξύ των αμινοξέων 200-275 (pI 10,0) ενώ η C-τελική περιοχή (αμινοξέα 271-315) εμφανίζεται ιδιαίτερα όξινη (pI 3,7). Η ανάλυση της αμινοξικής αλληλουχίας με τις βάσεις δεδομένων Pfam, SMART και InterPro δεν υπέδειξε συντηρημένα μοτίβα, ωστόσο ταυτοποιήθηκε ένα διμερές σήμα πυρηνικού εντοπισμού (bipartite NLS) που περιλαμβάνει τα αμινοξέα 223-238 και 245-262 (Εικ. 46). Σε συμφωνία με το ανωτέρω εύρημα είναι και η πρόβλεψη (87% πιθανότητα) του προγράμματος PSORT II για τον πυρηνικό εντοπισμό της πρωτεΐνης.



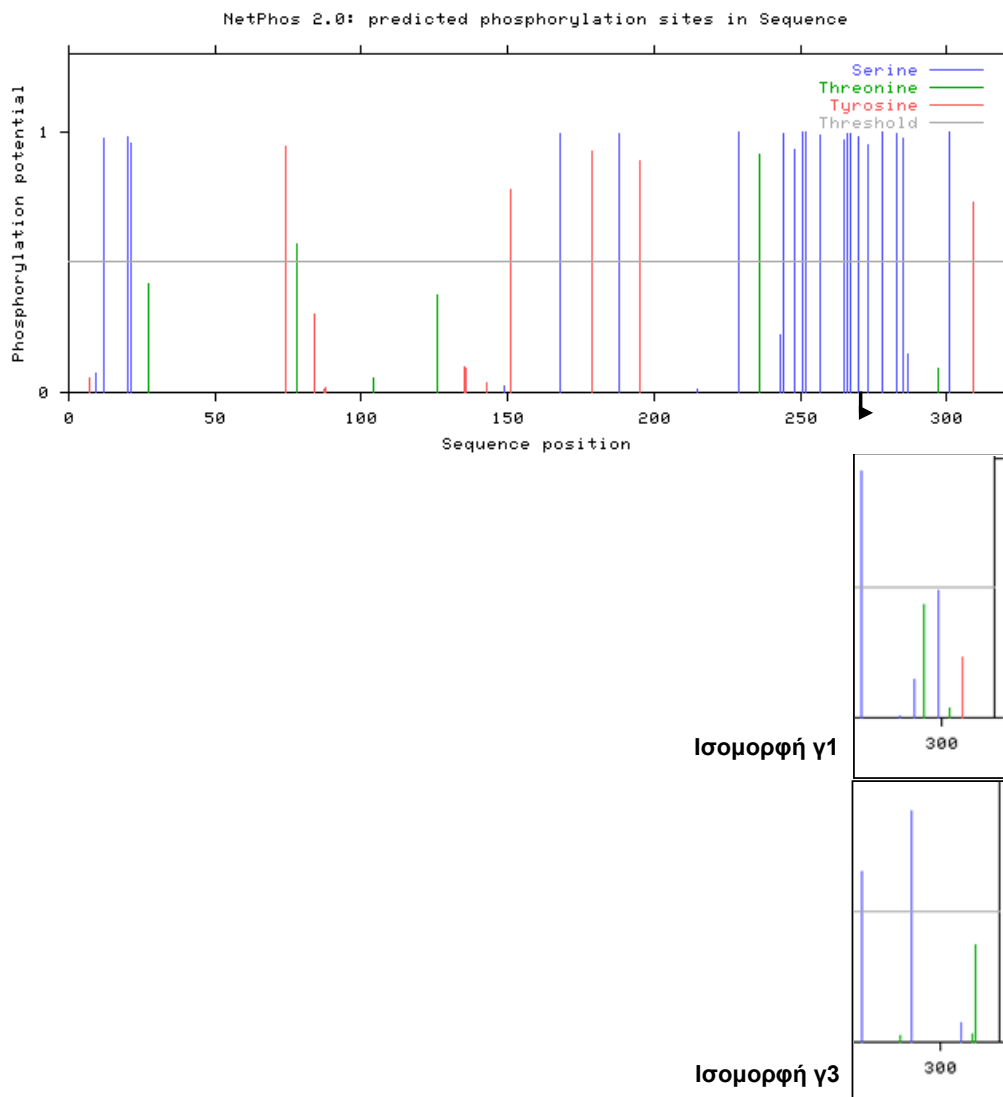
Εικ. 46: Γραφική αναπαράσταση της πρωτεΐνης FRA10AC1. Η πρωτεΐνη αποτελείται από 315 αμινοξέα και περιέχει ένα διμερές σήμα πυρηνικού εντοπισμού (NLS, αμινοξέα 223-238 και 245-262) που ταυτοποιήθηκε με διερεύνηση της βάσης δεδομένων Interpro.

Επίσης, εντοπίστηκαν αρκετές θέσεις φωσφορυλίωσης των αμινοξέων (Ser, Tyr, Thr) κυρίως στο καρβοξυτελικό άκρο του μορίου (πρόγραμμα NetPhos, Εικ. 47). Αυτό το πρότυπο των φωσφορυλίωσεων φαίνεται πως τροποποιείται σημαντικά στις εναλλακτικές ισομορφές του μορίου. Έτσι, ενώ η πρωτεΐνη έχει στο καρβοξυτελικό της άκρο 5 πιθανές θέσεις φωσφορυλίωσης σε σερίνη και μία σε τυροσίνη, τα εναλλακτικά πολυπεπτίδια γ1 (313aa) και γ3 (315aa) έχουν μία και δύο πιθανές θέσεις φωσφορυλίωσης σε σερίνη αντίστοιχα. Τέλος, ταυτοποιήθηκε μία πιθανή θέση σουμοϋλίωσης στην λυσίνη 91 (πρόγραμμα SUMOplot).

Επιπλέον, με σύγκριση της FRA10AC1 με την πρωτεϊνική βάση δεδομένων (πρόγραμμα Blastp) εντοπίστηκαν πολυπεπτίδια με σημαντική αμινοξική ταυτότητα καθ' όλο το μήκος τους, προερχόμενα από είδη του ζωικού αλλά και του φυτικού βασιλείου. Οι ορθόλογες πρωτεΐνες ταυτοποιήθηκαν αποκλειστικά σε πολυκύτταρους ευκαρυωτικούς οργανισμούς. Οι αντιπρόσωποι του ζωικού βασιλείου προέρχονταν τόσο από το υπόφυλο των σπονδυλωτών όσο και των ασπόνδυλων. Η ανάλυση αυτή, όπως και η αντίστοιχη ανάλυση σε επίπεδο DNA (πρόγραμμα Blastn) έδειξε ότι το γονίδιο είναι μοναδιαίο και δεν έχει παράλογα στον άνθρωπο.

Ειδικότερα, στο υπόφυλο των σπονδυλωτών, ορθόλογα εντοπίστηκαν σε θηλαστικά, ιχθύες, αμφίβια και πτηνά. Οι πρωτεΐνες των θηλαστικών (*Pan troglodytes*, *Canis familiaris*, *Rattus norvegicus* και *Mus musculus*) εμφανίζουν αμινοξική ταυτότητα 99%, 87%, 86% και 86% αντίστοιχα. Στους τελεόστεους ιχθύες (*Fugu rubripes*, *Danio rerio* και *Tetraodon nigroviridis*) η ταυτότητα ανέρχεται σε 63%, 65% και 62% αντίστοιχα. Τα πολυπεπτίδια που ταυτοποιήθηκαν στον εκπρόσωπο

των αμφιβίων (*Xenopus laevis*) και των πτηνών (*Gallus gallus*) παρουσιάζουν ταυτότητα με την ανθρώπινη πρωτεΐνη 65% και 70% αντίστοιχα.

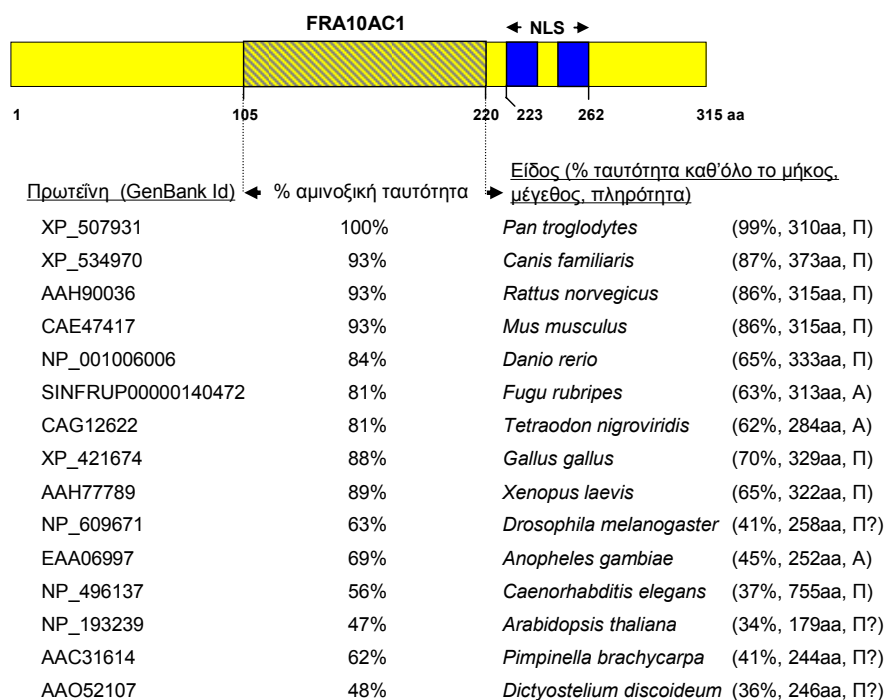


Εικ. 47: Πρόβλεψη πιθανών φωσφορυλιώσεων στην FRA10AC1 και στις εναλλακτικές ισομορφές γ1 και γ3, βάσει του προγράμματος NetPhos. Ο οριζόντιος άξονας αντιστοιχεί στα αμινοξικά κατάλοιπα και ο κάθετος στην πιθανότητα φωσφορυλίωσής τους με τον ουδό να απεικονίζεται με τη γκρι οριζόντια γραμμή. Φαίνονται οι πιθανές φωσφορυλιώσεις σε σερίνη, θρεονίνη και τυροσίνη (μπλε, πράσινες και κόκκινες γραμμές αντίστοιχα). Η πλειονότητα των πιθανών θέσεων φωσφορυλίωσης εντοπίζεται στο καρβοξυτελικό άκρο του μορίου.

Στα ασπόνδυλα, ταυτοποιήθηκαν πρωτεΐνες σε έντομα, στον *Caenorhabditis elegans* και στο *Dictyostelium discoideum*. Τα ορθόλογα πολυπεπτίδια στη *Drosophila melanogaster* και τον *Anopheles gambiae* παρουσιάζουν ταυτότητα 41% και 45% αντίστοιχα. Η πρωτεΐνη του νηματώδη έχει μήκος 755 αμινοξέα και η ομοιότητα της με την FRA10AC1 (37%) περιορίζεται στο N-τελικό τμήμα της. Υψηλή είναι η αμινοξική ταυτότητα με την πρωτεΐνη της αμοιβάδας *Dictyostelium discoideum*

(36%) που είναι και ο απλούστερος οργανισμός στον οποίο ταυτοποιήθηκε ορθόλογο μόριο. Στο φυτικό βασίλειο εντοπίστηκαν ορθόλογα μόρια στην *Arabidopsis thaliana* και *Pimpinella brachycarpa* με ταυτότητα 34% και 41% αντίστοιχα.

Στην Εικ. 48 συνοψίζονται οι οργανισμοί και τα πολυπεπτίδια που εντοπίστηκαν με την παραπάνω ανάλυση. Η στοίχιση όλων των μορίων (πρόγραμμα Clustw, Εικ. 49) έδειξε υψηλότερη συντήρηση στο κεντρικό τμήμα της πρωτεΐνης (αμινοξέα: 105-220), γεγονός που πιθανώς υποδηλώνει ότι το συγκεκριμένο τμήμα ενέχεται σε μία σημαντική λειτουργία (Εικ. 48). Από τα 116 αμινοξέα που περιλαμβάνει η περιοχή αυτή, τα 22 είναι τα ίδια σε όλους τους οργανισμούς που μελετήθηκαν, ενώ σε 31 οι αμινοξικές αλλαγές που παρατηρούνται είναι συντηρητικές. Στο C-τελικό τμήμα της πρωτεΐνης που περιέχει και το διμερές NLS, η αμινοξική συντήρηση είναι κατά πολύ μικρότερη. Αυτό είναι αναμενόμενο γιατί υπάρχει μεγάλη ποικιλία στις αλληλουχίες που λειτουργούν σαν διμερή σήματα πυρηνικού εντοπισμού, με κοινό παρονομαστή την παρουσία βασικών αμινοξέων (Fontes MR *et al*, 2003).



Εικ. 48: Προβλεπόμενα ορθόλογα μόρια της FRA10AC1 στο ζωικό και φυτικό βασίλειο.

Σχηματική απεικόνιση της FRA10AC1 και του κεντρικού τμήματός της (γραμμοσκιασμένο, αμινοξέα 105-220), όπου παρατηρείται και η υψηλότερη αμινοξική ταυτότητα. Παρατίθενται οι οργανισμοί στους οποίους ταυτοποιήθηκαν πολυπεπτίδια με σημαντική αμινοξική ομοιότητα, ο κωδικός πρόσβασής τους (αριστερή στήλη), το ποσοστό αμινοξικής ταυτότητας στο κεντρικό τμήμα των πολυπεπτιδίων (δεύτερη στήλη), το οποίο είναι κατά πολύ υψηλότερο από την ταυτότητα κατά μήκος ολόκληρου του μορίου (δεξιά στήλη, στην παρένθεση). Επιπλέον φαίνεται το μέγεθος των πολυπεπτιδίων και η πληρότητα (Π) ή όχι (A) του μορίου (στην παρένθεση). Η ταυτοποίηση όλων των μορίων πραγματοποιήθηκε με διερεύνηση της πρωτεϊνικής βάσης δεδομένων με το πρόγραμμα Blastp εκτός του *Fugu rubripes* για το οποίο η διερεύνηση έγινε στη βάση δεδομένων Ensembl, (NLS: σινιάλο πυρηνικού εντοπισμού).

<i>Homo sapiens</i>	LVKLRLCQECSTIKLNFHHRRKEIKSKRRKDKTKKDCSESSHKKSRLLSSA-EEASKKKD--	261
<i>Pan troglodytes</i>	LVKLRLCQECSTIKLNFHHRRKEIKSKRRKDKTKKDCSESSHKKSRLLSSA-EEASKKKD--	261
<i>Canis familiaris</i>	LVKLRLCQECSTIKLNFHHRRKEIKSKRRKDKTKKDCSESSHKKSRLLSSA-EETSCKKD--	260
<i>Rattus norvegicus</i>	LVKLRLCQECSTIKLNFHHRRKEIKSTKRSKTKTESDESPHKNSRSSSS-EEASQGGK--	261
<i>Mus musculus</i>	LVKLRLCQECSTIKLNFHHRRKEIKSTKRSKTKTESDESPHKNSRSSSS-EEASKGKD--	261
<i>Gallus gallus</i>	LVKLRLCPECSYKLNFFHHRKEVKASKRRGTAVGNSKQPKIKKAKLSHGKSKKKT--	272
<i>Xenopus laevis</i>	LVKLRLCPECSYKLNFFHHRKEIQPTTKRRAKTEREESRKKKKTSHS-QKNTNTKD--	265
<i>Tetraodon nigroviridis</i>	LVKLRLCPECSYKLNFFHHRKEVKAKAKSQRLSKENKLSHKKKRRKSSSSSHSKKQLG	235
<i>Fugu rubripes</i>	LVKLRLCPECSYKLNFFHHRKEVKAKPKTQ--GEENEKPSHKKKRRKSSSSSHSKKHK--	257
<i>Danio rerio</i>	LVKLRLCPECSYKLNFFHHRKEVTAKRFR--SKENLDVSKKSKSKSRKNKDKRKHK--	269
<i>Drosophila melanogaster</i>	LVKLRLCPECSYKLNFFHHRKEVTAKRFR--SKENLDVSKKSKSKSRKNKDKRKHK--	202
<i>Anopheles gambiae</i>	LVKLRLCPECSYKLNFFHHRKEVTAKRFR--SKENLDVSKKSKSKSRKNKDKRKHK--	192
<i>Caenorhabditis elegans</i>	LVKLRLCPECSYKLNFFHHRKEVTAKRFR--SKENLDVSKKSKSKSRKNKDKRKHK--	262
<i>Arabidopsis thaliana</i>	LVKLRLCPECSYKLNFFHHRKEVTAKRFR--SKENLDVSKKSKSKSRKNKDKRKHK--	133
<i>Pimpinella brachycarpa</i>	LVKLRLCPECSYKLNFFHHRKEVTAKRFR--SKENLDVSKKSKSKSRKNKDKRKHK--	204
<i>Dictyostelium discoideum</i>	LVKLRLCPECSYKLNFFHHRKEVTAKRFR--SKENLDVSKKSKSKSRKNKDKRKHK--	216
	*** * :	
<i>Homo sapiens</i>	-----KGHSKSSKSEDSLLR---NSDEEE---SASESELWKGPLP--ETDEKSEQEEFDE	308
<i>Pan troglodytes</i>	-----KGHSKSSKSEDSLLR---NSDEEE---SASESELWKGPLP--ETDEKSEQEEFDE	304
<i>Canis familiaris</i>	-----KGNS-----DEE---GTSDELWKGPLP--ETDEKSEQEEFDE	287
<i>Rattus norvegicus</i>	-----EGHSKSSKSEDSRNR---NAGEED---SASDELWKGPLP--ETDEKSEQEEFDD	308
<i>Mus musculus</i>	-----EGHSKSSKSEDSRNR---NAGEED---SASDELWKGPLP--ETDEKSEQEEFDD	308
<i>Gallus gallus</i>	-----HGDQVSSSEDSSTSDKSDNSDVQD---GPSADDFWKGPLQ--EPDEKSRREEEFDE	322
<i>Xenopus laevis</i>	-----HPTKHTNSSETDSSDSDTGGEGGD---DKTDFEFWKGATH--ETDEKTRREEFDD	315
<i>Tetraodon nigroviridis</i>	RDRDRSSSSCSDDSDQEPDKDAQASDEFE---GQSEADHWRGAP--VVVEKSR-----	284
<i>Fugu rubripes</i>	---RDRSSPS--SSEDSQESDK---AGEEFE---GQSEADHWRGAP--VVVEKSRREEEFDE	306
<i>Danio rerio</i>	---KRRKEHSSSSSEESQESDKGSDDDDDDEDADGPSESEHWKGPAP--AVEEKSRREEEFDE	326
<i>Drosophila melanogaster</i>	-----PVEPASESHPEASKEPTDSKDAT---TADEQIQQQPD--ISREQASREQEFER	251
<i>Anopheles gambiae</i>	---TFVATATQESTVPEEVEPAVGEQDGS---CAPEVAGGSWTGHEVEEKTREEEFDE	245
<i>Caenorhabditis elegans</i>	-----VELNQHEAKNQETKVQPSTL---SASTDIWEGPPP--AETKTVDEIDE	309
<i>Arabidopsis thaliana</i>	-----HSEYDTEEDRRKGTTRKSK---LE-----SADREKGDDEFDE	169
<i>Pimpinella brachycarpa</i>	-----PTEYKGNKDSRKMKTST---LGDQKAGDTENFDE	237
<i>Dictyostelium discoideum</i>	-----KNDNSEDEDEDEDEQYKQDE-----EYEDDEDDN	246
	:	
<i>Homo sapiens</i>	YFQDLFL---	315
<i>Pan troglodytes</i>	-VQNLCI---	310
<i>Canis familiaris</i>	-----	
<i>Rattus norvegicus</i>	YFQDLFL---	315
<i>Mus musculus</i>	YFQDLFL---	315
<i>Gallus gallus</i>	YFQDLFL---	329
<i>Xenopus laevis</i>	YFQDMFL---	322
<i>Tetraodon nigroviridis</i>	-----	
<i>Fugu rubripes</i>	YFEDLFL---	313
<i>Danio rerio</i>	YFEDMFL---	333
<i>Drosophila melanogaster</i>	YLEDLIF---	258
<i>Anopheles gambiae</i>	FLEDLLL---	252
<i>Caenorhabditis elegans</i>	FLDDLFL---	316
<i>Arabidopsis thaliana</i>	YMEGMFPGNG	179
<i>Pimpinella brachycarpa</i>	FLEGMFL---	244
<i>Dictyostelium discoideum</i>	-----	

Εικ. 49: Συγκριτική ανάλυση της αλληλουχίας της πρωτεΐνης FRA10AC1 με τα πιθανά ορθόλογα μόρια. Υψηλότερη αμινοξική ταυτότητα εμφανίζεται στο κεντρικό τμήμα της πρωτεΐνης (αμινοξέα 105-220). Από τα 116 αμινοξέα που περιλαμβάνει η περιοχή αυτή τα 22 είναι απaráλλακτα σε όλους τους οργανισμούς (πράσινα), ενώ σε 31 οι αμινοξικές αλλαγές που παρατηρούνται είναι συντηρημένες (κίτρινα). Στη στοίχιση παραλήφθηκαν τα 439 C-τελικά και τα 76 N-τελικά αμινοξέα των πρωτεϊνών του *C. elegans* και του *Canis familiaris* αντίστοιχα.

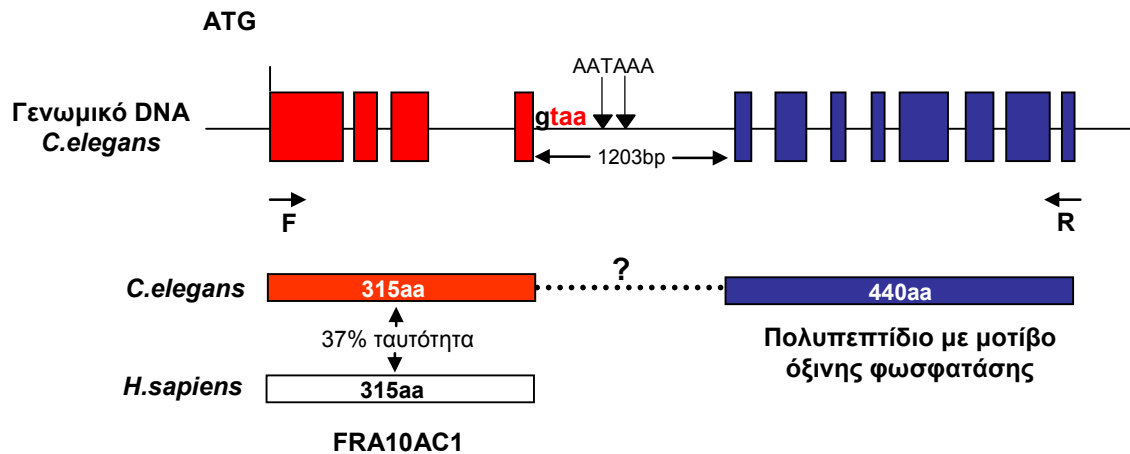
Κατά τη στοίχιση των ορθόλογων μορίων (Εικ. 48, 49) παραλείφθηκαν τμήματα των πολυπεπτιδίων του *C. familiaris* και του *C. elegans*. Στον *C. familiaris*, το αντίστοιχο μόριο έχει 373 αμινοξέα και φέρει 76 αμινοξέα στο N-τελικό του άκρο που απουσιάζουν από την πρωτεΐνη του ανθρώπου. Ωστόσο, η πρωτεϊνική αλληλουχία βασίζεται σε ένα μόνο EST (CO590869, 689bp) το οποίο δεν στηρίζει τα επιπλέον αμινοξέα. Στο *C. elegans* η πρωτεϊνική καταχώρηση αποτελείται από 755

αμινοξέα και έχει προκύψει από υπολογιστικά προγράμματα ανάλυσης της γενωμικής αλληλουχίας ενώ δεν στηρίζεται σε κάποιο μόριο cDNA. Η μεγάλη διαφορά στο μέγεθος των πολυπεπτιδίων του ανθρώπου και του *C. elegans* μπορεί να οφείλεται είτε σε φαινόμενο γονιδιακής σύντηξης (Enright AJ *et al*, 1999 & 2001) είτε σε λάθος στο χαρακτηρισμό του αντίστοιχου γονιδίου *in silico* (annotation). Υπέρ του τελευταίου συνηγορούν τα εξής (Εικ. 50): Το προβλεπόμενο γονίδιο (F26C11.1) αποτελείται από 12 εξόνια, εκ των οποίων τα 4 πρώτα κωδικοποιούν το τμήμα της πρωτεΐνης που εμφανίζει ομοιότητα με την FRA10AC1, ενώ τα υπόλοιπα κωδικοποιούν πολυπεπτίδιο που φέρει μοτίβο όξινης φωσφατάσης ιστιδίνης. Το ιντρόνιο 4 έχει μέγεθος 1203bp, που είναι κατά πολύ μεγαλύτερο από το μέσο μέγεθος των ιντρονίων του νηματώδη (370bp, Yu J *et al*, 2002) ενώ όλα τα υπόλοιπα είναι <120bp πλην του ιντρονίου 3 που είναι 503bp. Αμέσως μετά το τέλος του εξονίου 4 (οι πρώτες τέσσερις βάσεις του ιντρονίου 4) έχουν στο ίδιο αναγνωστικό πλαίσιο, κωδικόνιο λήξης που, αν χρησιμοποιηθεί, παράγει πολυπεπτίδιο 315 αμινοξέων, ίδιου μήκους και κατά 37% ταυτόσημο με αυτό του ανθρώπου. Επίσης, στην αρχή του ιντρονίου 4 (nt:19-24 και nt:313-318, μετρώντας από την αρχή του ιντρονίου), υπάρχουν δύο συντηρημένες αλληλουχίες σημάτων πολυαδενυλίωσης (AATAAA). Επιπλέον, στα πλαίσια προσπάθειας κλωνοποίησης και συνεπώς επιβεβαίωσης 12.000 προβλεπόμενων cDNA του *C. elegans* (Lamesch P *et al*, 2004) η PCR σε βιβλιοθήκη cDNA όπου χρησιμοποιήθηκε ένας εκκινητής στο πρώτο εξόνιο του γονιδίου F26C11.1 σε συνδυασμό με εκκινητή στο εξόνιο 12 ήταν ανεπιτυχής (http://www.wormbase.org/db/seq/pcr?name=mv_F26C11.1;class=PCR_product). Η διερεύνηση των ESTs του *C.elegans* δεν βοήθησε να εξαχθεί κάποιο συμπέρασμα καθώς δεν υπάρχουν ESTs προερχόμενα από αυτό το γονίδιο, γεγονός που συμβαίνει σε ποσοστό ~40% των προβλεπόμενων γονιδίων του νηματώδη (Lamesch P *et al*, 2004).

Τέλος, η σύγκριση όλων των ορθολόγων πρωτεϊνών έδειξε ότι οι αμινοξικές αλλαγές R16H, T78R και V81I, που προέρχονται από τα SNPs που ταυτοποιήθηκαν στην κωδική περιοχή του γονιδίου του ανθρώπου, είναι συντηρημένες τουλάχιστον σε έναν άλλο οργανισμό. Έτσι, το αμινοξύ 16 της πρωτεΐνης του *Pan troglodytes* είναι ιστιδίνη, το αμινοξύ 78 του *Gallus gallus* είναι αργινίνη και το αμινοξύ 81 στους οργανισμούς *Anopheles gambiae*, *Caenorhabditis elegans*, *Pimpinella brachycarpa* και *Dictyostelium discoideum* είναι ιντερλευκίνη (Εικ. 10).

Αξίζει να σημειωθεί ότι κανένα από τα πιθανά ορθόλογα μόρια που ταυτοποιήθηκαν δεν έχει μελετηθεί πειραματικά με συνέπεια να μην υπάρχουν διαθέσιμες πληροφορίες για τον πιθανό βιοχημικό ρόλο της FRA10AC1. Η ύπαρξη αρκετών από αυτά στηρίζεται αποκλειστικά σε γονιδιακές προβλέψεις

προγραμμάτων επεξεργασίας γενωμικών αλληλουχιών. Ωστόσο, η παρουσία ορθολόγων αποκλειστικά σε πολυκύτταρους ευκαρυωτικούς οργανισμούς υποδηλώνει ενδεχομένως την συμμετοχή του μορίου σε βιολογικές διαδικασίες που χαρακτηρίζουν αυτούς τους οργανισμούς.



Εικ. 50: Το ορθόλογο γονίδιο *FRA10AC1* στον *C. elegans* (F26C11.1.). Απεικονίζεται η γενωμική οργάνωση του γονιδίου *FRA10AC1* στον *C. elegans* βάσει *in silico* χαρακτηρισμού της γενωμικής αλληλουχίας. Ωστόσο, καθώς: 1) η ομοιότητα (37% ταυτότητα) με την πρωτεΐνη *FRA10AC1* του ανθρώπου περιορίζεται στο αμινοτελικό μέρος του προβλεπόμενου πολυπεπτιδίου, 2) υπάρχει πιθανό κωδικόνιο λήξης (taa) και σήματα πολυαδενυλίωσης στην αρχή του ιντρονίου 4 και 3) οι εκκινητές F και R δεν παράγουν προϊόν με PCR σε cDNA βιβλιοθήκη, είναι πιθανόν το προβλεπόμενο γονίδιο του *C. elegans* να αντιστοιχεί σε δύο διαφορετικά γονίδια.

➤ Έλεγχος της ειδικότητας του αντισώματος αντι-*FRA10AC1*

Κατά τον υποκυτταρικό εντοπισμό της ενδογενούς πρωτεΐνης *FRA10AC1* και στα πειράματα ανοσοκατακρήμνησης που περιγράφονται παρακάτω, χρησιμοποιήθηκε το ειδικό αντίσωμα για την *FRA10AC1* που απομονώθηκε από τον πολυκλωνικό ορό ανοσοποιημένου κουνελιού (βλ. Μεθόδους). Η ειδικότητα του αντισώματος ελέγχθηκε: 1) με ανοσοεντοπισμό της πρωτεΐνης *FRA10AC1* σε παροδικά διαμολυσμένα κύτταρα HeLa που εκφράζουν την ανασυνδυασμένη *FRA10AC1*-EGFP, όπου παρατηρείται πλήρης ταύτισή της με το σήμα που προκύπτει από την χρήση του αντισώματος (Εικ. 51), 2) με ανάλυση Western κατά την οποία το ειδικό αντίσωμα αναγνωρίζει την ανασυνδυασμένη πρωτεΐνη His-*FRA10AC1* που έχει παραχθεί σε βακτήρια ενώ δεν ανιχνεύει πλέον την πρωτεΐνη GST, αποτέλεσμα που είναι επιθυμητό καθώς το αντιγόνο που είχε χρησιμοποιηθεί στην ανοσοποίηση ήταν σε σύντηξη με την GST, 3) με ανάλυση Western σε πρωτεϊνικό εκχύλισμα κυττάρων COS-7 κατά την οποία το αντίσωμα αναγνωρίζει μία ζώνη αναμενόμενου μεγέθους, που αντιστοιχεί στην ενδογενή πρωτεΐνη, 4) με ανάλυση Western σε πρωτεϊνικό εκχύλισμα παροδικά διαμολυσμένων κυττάρων

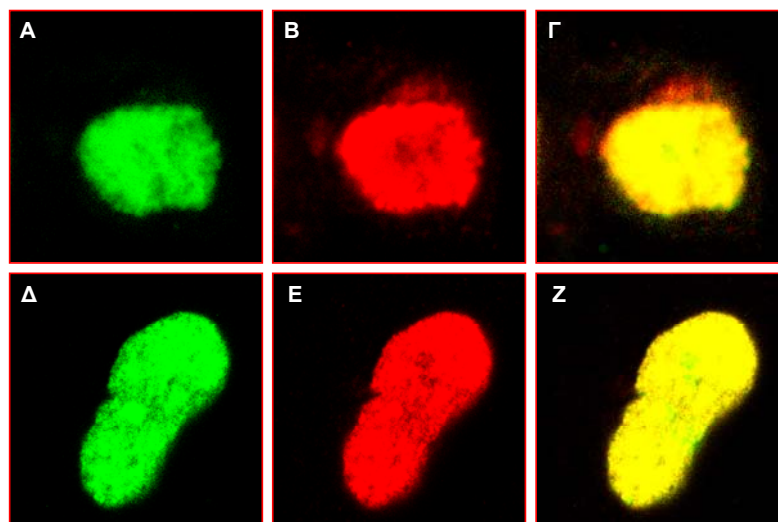
COS-7 που εκφράζουν την ανασυνδυσασμένη FRA10AC1-EGFP, όπου η πρωτεΐνη που αναγνωρίζεται από το ειδικό αντίσωμα ταυτίζεται με αυτήν που αναγνωρίζεται από το αντίσωμα αντι-GFP και 5) με δοκιμή ανοσοκατακρήμνισης της χιμαιρικής πρωτεΐνης FRA10AC1-EGFP από πρωτεϊνικό εκχύλισμα παροδικά διαμολυσμένων κυττάρων COS-7 και ακόλουθη ανάλυση Western όπου φαίνεται ότι το αντίσωμα είναι ικανό να ανοσοκατακρήμνισει την ανασυνδυσασμένη πρωτεΐνη.

➤ **Υποκυτταρική τοπολογική ανάλυση της ανασυνδυσασμένης και της ενδογενούς πρωτεΐνης FRA10AC1 σε κυτταρικές σειρές θηλαστικών**

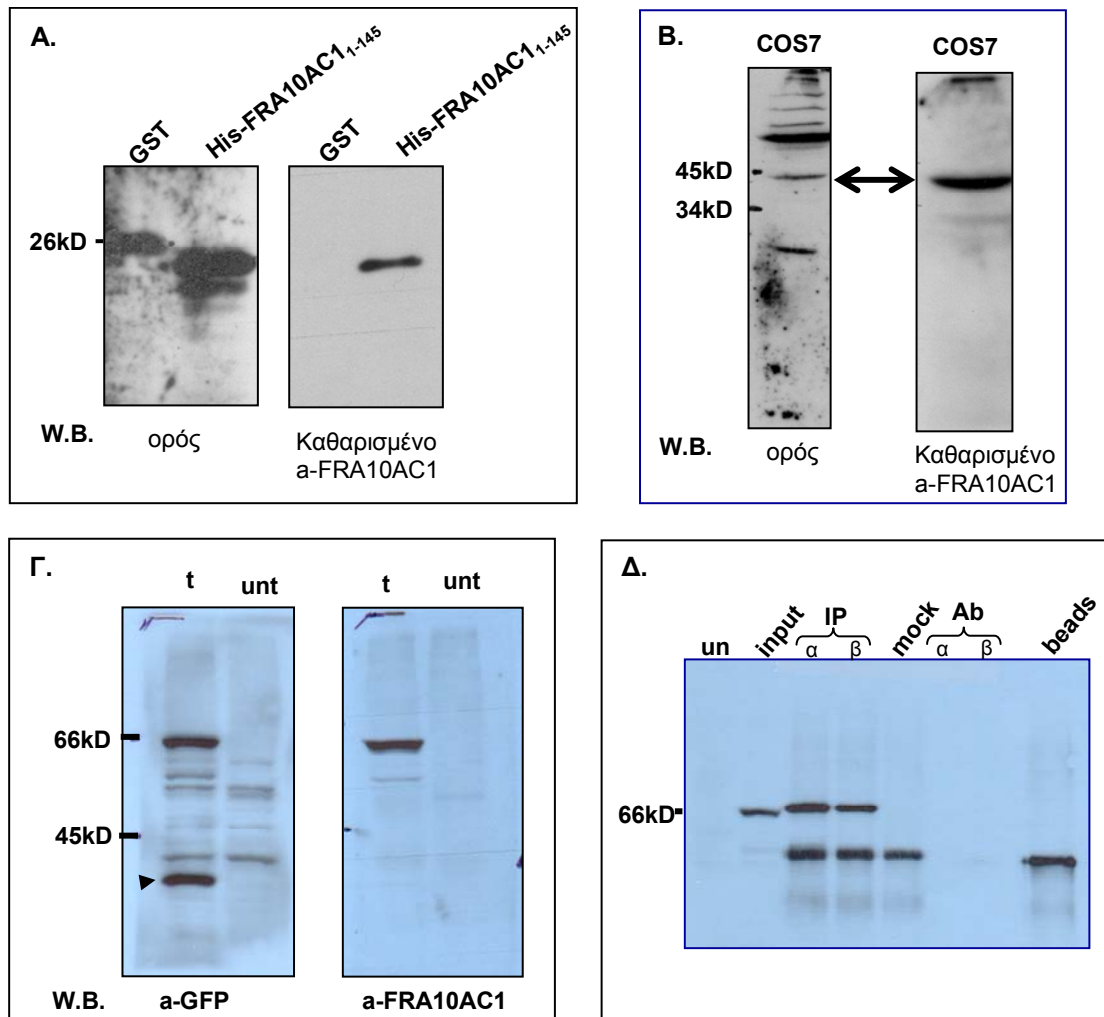
Η εξέταση της υποκυτταρικής τοπολογίας της πρωτεΐνης FRA10AC1 και μέσω αυτής ο έλεγχος της λειτουργικότητας του θεωρούμενου NLS, πραγματοποιήθηκε αρχικά με παρατήρηση με μικροσκοπία φθορισμού σε κύτταρα HeLa που είχαν διαμολυνθεί παροδικά με πλασμιδιακή κατασκευή που κωδικοποιούσε χιμαιρικό πολυπεπτίδιο αποτελούμενο από την FRA10AC1 πλήρους μήκους σε σύντηξη με την EGFP στο C-τελικό της άκρο. Διαπιστώθηκε ότι η χιμαιρική πρωτεΐνη ανιχνεύεται στον πυρήνα των κυττάρων, εξαιρουμένου του πυρηνίσκου, πιθανώς οδηγημένη από το διμερές NLS που διαθέτει. Η κατανομή της στην πλειοψηφία των κυττάρων είναι διάχυτη (Εικ. 53). Ωστόσο, σε ποσοστό 5-10% των κυττάρων, η πρωτεΐνη συσσωρεύεται σε διακριτές εστίες, μικρότερου ή μεγαλύτερου μεγέθους, δημιουργώντας ένα σικτικό πρότυπο (Εικ. 53). Ο συγκεκριμένος υποκυτταρικός εντοπισμός είναι σαφώς διακριτός από τον μη ειδικό και διάχυτο εντοπισμό της EGFP σε όλο το κύτταρο (Εικ. 55). Στα πειράματά μας, τα όρια του πυρήνα καθοριζόταν είτε με παράλληλη ανοσοανίχνευση της λαμίνης A/C, που εντοπίζεται στην πυρηνοπλασματική πλευρά της πυρηνικής μεμβράνης, είτε με χρώση με DAPI. Η περιγραφείσα υποκυτταρική τοπολογία παρατηρήθηκε και σε διάφορες άλλες κυτταρικές σειρές που εξετάστηκαν: HepG2 (ηπατικό καρκίνωμα), CaCO2 (αδενοκαρκίνωμα εντέρου) καθώς και στα νεφρικά κύτταρα πιθήκου CV1 και COS-7. Ταυτόσημο πυρηνικό πρότυπο παρατηρήθηκε και κατά την ανίχνευση της ενδογενούς πρωτεΐνης (Εικ. 54) σε διάφορες κυτταρικές σειρές. Κατά την ανίχνευση του ενδογενούς μορίου, σε καμία περίπτωση δεν παρατηρήθηκε η σικτώδης πυρηνική κατανομή που εμφανίζει η ανασυνδυσασμένη πρωτεΐνη, σε ένα μικρό ποσοστό των κυττάρων. Αυτό μπορεί να οφείλεται στην υπερέκφρασή της η οποία στα παροδικά διαμολυσμένα κύτταρα είτε σχηματίζει συσσωματώματα, είτε «αποθηκεύεται» σε συγκεκριμένα πυρηνικά διαμερίσματα που εμφανίζουν παρόμοια δομή, όπως π.χ. τα σωμάτια PML, τα speckles, τα σωμάτια Cajal. Ενδιαφέρον είναι ότι η σικτώδης κατανομή παρατηρήθηκε τόσο σε κύτταρα COS-7 όσο και σε CV-1

που, καθώς δεν εκφράζουν το T-αντιγόνο, εμφανίζουν χαμηλότερα επίπεδα έκφρασης των ανασυνδυασμένων πρωτεϊνών.

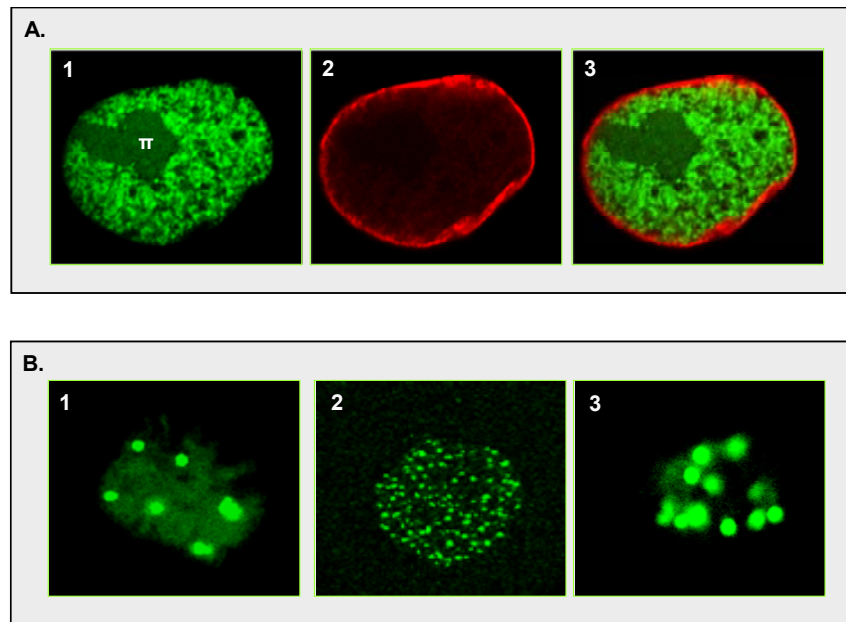
Για τον καθορισμό της περιοχής που είναι υπεύθυνη για τον πυρηνικό εντοπισμό της FRA10AC1, πραγματοποιήθηκαν πειράματα διαμολύνσεων με τη χρήση μιας σειράς χιμαιρικών πολυπεπτιδίων, αποτελούμενων από υποπεριοχές της πρωτεΐνης σε σύνδεση με την EGFP (Εικ. 55). Η μελέτη έγινε σε κύτταρα HeLa. Από την ανάλυση φάνηκε ότι το ανασυνδυασμένο μόριο που περιλαμβάνει τα αμινοξέα 147-253 εντοπίζεται στον πυρήνα και άρα περιλαμβάνει λειτουργικό NLS. Αντίθετα τα τμήματα της πρωτεΐνης που αντιστοιχούν στα αμινοξέα 1-146 και 257-315 εμφανίζουν διάχυτη κατανομή σε όλο το κύτταρο, παρόμοια με της EGFP. Το διμερές NLS που είχε προβλεφθεί από υπολογιστικά προγράμματα περιλάμβανε τα αμινοξέα 223-238 και 245-265. Με το πείραμα που περιγράφηκε φάνηκε ότι το δεύτερο τμήμα δεν είναι απαραίτητο να υπάρχει αυτούσιο για να οδηγηθεί η πρωτεΐνη στον πυρήνα. Επιπλέον, το ελάχιστο τμήμα της πρωτεΐνης που εμφανίζει πυρηνικό εντοπισμό είναι ικανό και για τον εντοπισμό της FRA10AC1 στις σικτώδεις δομές που περιγράφηκαν παραπάνω. Μάλιστα το ποσοστό των κυττάρων που εμφάνιζαν σικτώδεις πυρηνικό πρότυπο κατανομής ήταν συγκρίσιμο με αυτό που παρατηρήθηκε για την πλήρους μήκους πρωτεΐνη.



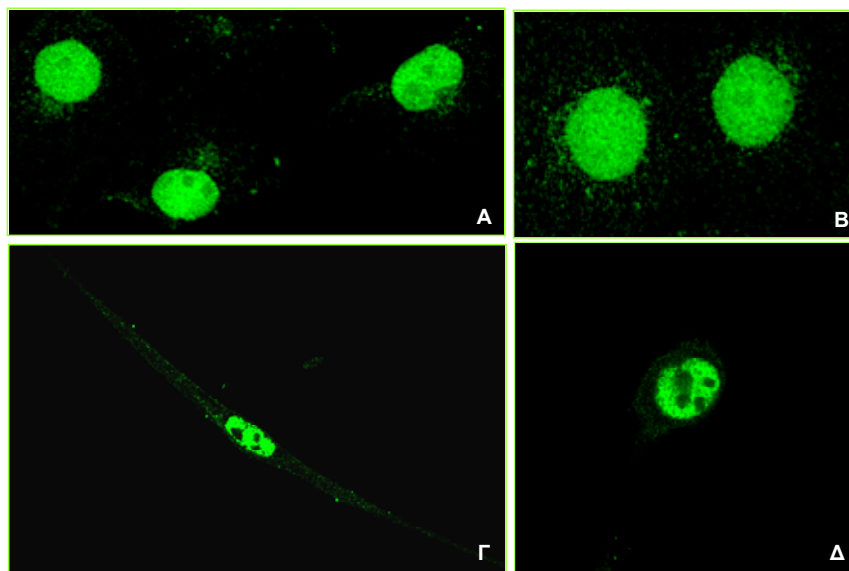
Εικ. 51. Έλεγχος της εξειδίκευσης του αντισώματος anti-FRA10AC1. Οπτικές τομές από δύο ενδεικτικά κύτταρα με συνεστιακό μικροσκόπιο φθορισμού που εκφράζουν παροδικά την ανασυνδυασμένη πρωτεΐνη FRA10AC1-EGFP (A, Δ) και στα οποία πραγματοποιήθηκε ανοσοανίχνευση με το αντίσωμα anti-FRA10AC1 (B, E). Ως δευτερογενές αντίσωμα χρησιμοποιήθηκε το anti-rabbit TRITC. Με ταυτόχρονη προβολή των εικόνων παρατηρείται πλήρης αλληλοεπικάλυψη των σημάτων φθορισμού (Γ, Ζ). Τα όρια του πυρήνα καθορίστηκαν με χρώση με DAPI, κατά την οποία φάνηκε ότι η πρωτεΐνη εντοπίζεται αποκλειστικά στον πυρήνα.



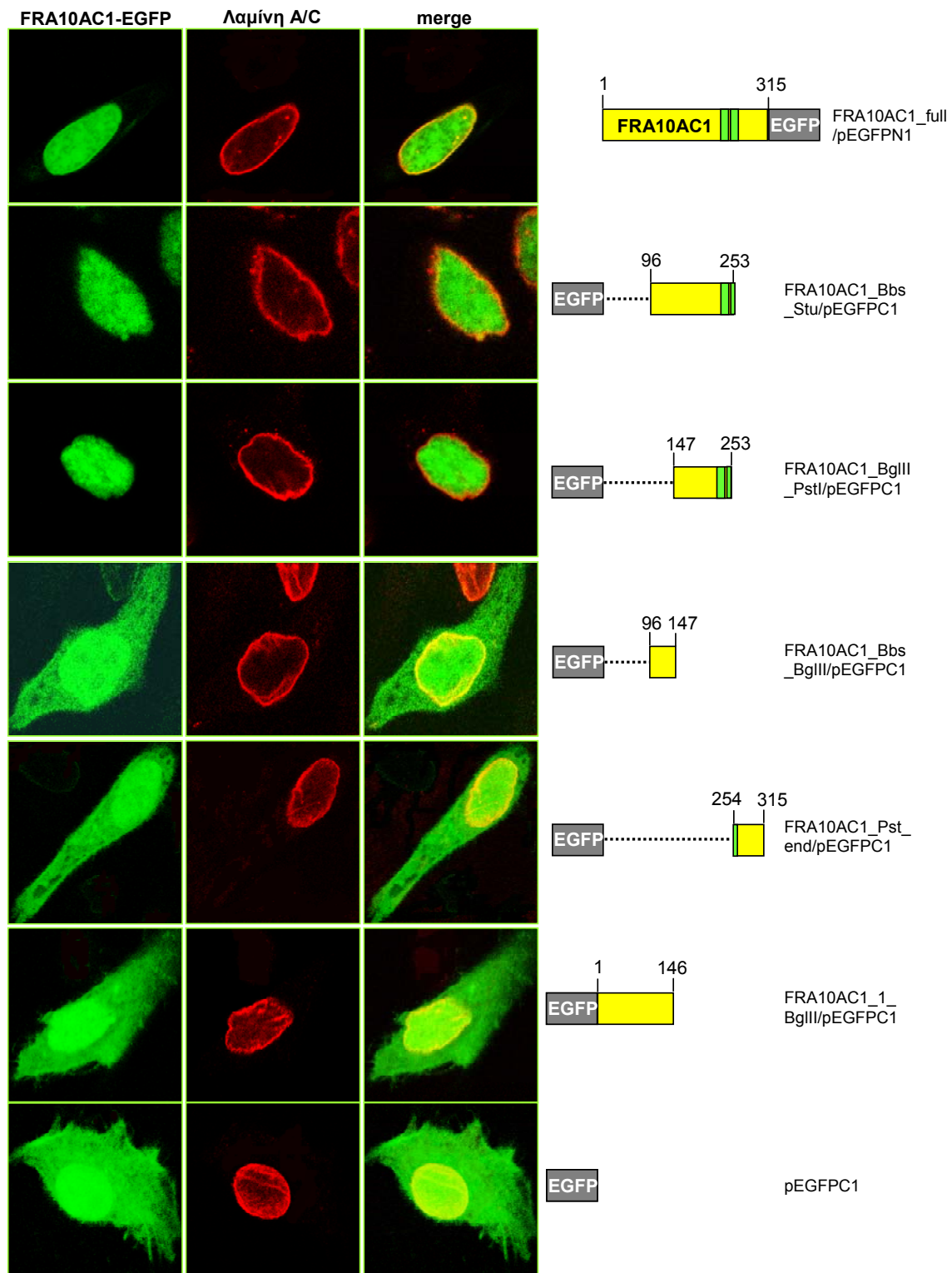
Εικ. 52: Έλεγχος εξειδίκευσης αντισώματος anti-FRA10AC1. **A.** Ο αντιορός (πριν τον καθαρισμό) αναγνωρίζει, εκτός από την ανασυνδυασμένη πρωτεΐνη His-FRA10AC11-145 και την GST, καθώς το αντιγόνο που είχε χρησιμοποιηθεί κατά την ανοσοποίηση ήταν σε σύντηξη με την GST. Μετά τον καθαρισμό (αριστερά), το αντίσωμα δεν ανιχνεύει πλέον την GST ενώ εξακολουθεί να ανιχνεύει την His-FRA10AC11-145 (ποσότητα 10ηg ολικής πρωτεΐνης). **B.** Σε πρωτεϊνικά εκχυλίσματα κυττάρων COS-7, ο ορός ανιχνεύει τουλάχιστον 6 ζώνες ενώ μετά τον καθαρισμό εμφανίζεται μία κύρια ζώνη αναμενόμενου μεγέθους (100μg πρωτεϊνικού εκχυλίσματος). **Γ.** Ανάλυση Western με αντίσωμα a-GFP (αριστερά) και a-FRA10AC1 (δεξιά) σε πρωτεϊνικά εκχυλίσματα κυττάρων COS-7 παροδικά διαμολυσμένων (t) και μη (unt) με πλασμιδιακή κατασκευή που κωδικοποιεί την πλήρους μήκους FRA10AC1 σε σύντηξη με την EGFP. Από την ανάλυση φαίνεται ότι και τα δύο αντισώματα αναγνωρίζουν την ίδια πρωτεΐνη και ότι το a-GFP ανιχνεύει επιπλέον και μία πρωτεΐνη μοριακού βάρους ~30kD που πιθανότατα αντιστοιχεί στην EGFP (βέλος), που δεν αναγνωρίζεται από το a-FRA10AC1. **Δ.** Το αντίσωμα a-FRA10AC1 ανοσοκατακρημνίζει την χιμαιρική πρωτεΐνη FRA10AC1-EGFP. Η ανοσοκατακρήμνιση πραγματοποιήθηκε σε πρωτεϊνικά εκχυλίσματα κυττάρων COS-7 που είχαν διαμολυνθεί παροδικά με πλασμιδιακή κατασκευή που κωδικοποιεί την πλήρους μήκους FRA10AC1 σε σύντηξη με την EGFP και η ανάλυση Western με a-GFP. Un=πρωτεϊνικό εκχύλισμα COS-7 μη διαμολυσμένων (100μg), input= το 1/10 του δείγματος που χρησιμοποιήθηκε στην ανοσοκατακρήμνιση, IP=ανοσοκατακρήμνιση με ~1μg (α) και 0,7μg (β) αντισώματος, mock=παράλληλο πείραμα στις ίδιες συνθήκες απουσία αντισώματος, Ab=~2μg και 1μg αντισώματος, beads=σφαιρίδια αгарόζης συζευγμένα με τις πρωτεΐνες A και G που χρησιμοποιήθηκαν στις ανοσοκατακρήμνισεις.



Εικ. 53. Η ανασυνδυασμένη FRA10AC1-EGFP εντοπίζεται στον πυρήνα των κυττάρων HeLa με διάχυτη ή σπικτή κατανομή. **A.** Στο μεγαλύτερο ποσοστό των παροδικά διαμολυσμένων κυττάρων η ανασυνδυασμένη πρωτεΐνη ανιχνεύεται σε όλο το πυρηνόπλασμα (1) εξαιρουμένου του πυρηνίσκου (π). Τα όρια του πυρήνα καθορίστηκαν με ανοσοανίχνευση της λαμίνης A/C (2). Η εικόνα 3 προέρχεται από ταυτόχρονη προβολή των 1 και 2. **B.** Σε ποσοστό 5-10% των κυττάρων η πρωτεΐνη συσσωρεύεται σε μεγάλες (1, 3) ή μικρές (2) σπικτώδεις δομές. Όλες οι εικόνες προέρχονται από οπτικές τομές σε συνεστιακό μικροσκόπιο φθορισμού.



Εικ. 54. Η ενδογενής FRA10AC1 εντοπίζεται στον πυρήνα. Η ανίχνευση της πρωτεΐνης πραγματοποιήθηκε με το αντίσωμα anti-FRA10AC1 και ακολούθως με anti-rabbit FITC, σε κύτταρα ανθρώπου (COS-7, A, B) και ποντικού (NIH-3T3, Γ, Δ). Η μονιμοποίηση πραγματοποιήθηκε είτε με 2% (A) είτε με 4% (B) PFA. Τα όρια του πυρήνα καθορίστηκαν με χρώση με DAPI.



Εικ. 55. Το τμήμα της FRA10AC1 που περιλαμβάνει τα αμινοξέα 147-253 είναι υπεύθυνο για την πυρηνική τοπολογία του μορίου. Η εξέταση του υποκυτταρικού εντοπισμού της ανασυνδυασμένης πρωτεΐνης και των επιμέρους τμημάτων της, έγινε μετά από παροδική διαμόλυνση κυττάρων HeLa με μία σειρά πλασμιδιακών κατασκευών στον φορέα pEGFP, ταυτόχρονη ανοσοανίχνευση της λαμίνης A/C και παρατήρηση με συνθεσιακό μικροσκόπιο φθορισμού. Δεξιά, αναγράφονται οι πλασμιδιακές κατασκευές που χρησιμοποιήθηκαν σε κάθε περίπτωση και απεικονίζονται σχηματικά τα αντίστοιχα τμήματα της πρωτεΐνης που κωδικοποιούν καθώς και το NLS (πράσινο). Στις τελευταίες εικόνες, τα κύτταρα έχουν διαμολυνθεί παροδικά μόνο με τον φορέα (αρνητικός έλεγχος).

➤ **Διερεύνηση πρωτεϊνικών αλληλεπιδράσεων της FRA10AC1**

Στα κύτταρα οι πρωτεΐνες μέσω των αλληλεπιδράσεών τους συμμετέχουν σε εκτεταμένα πρωτεϊνικά δίκτυα, τις πολυπρωτεϊνικές «μηχανές», που είναι δυναμικές δομές που συναρμολογούνται και αποσυναρμολογούνται σε ανταπόκριση σε εξω-, δια- και ενδοκυτταρικά ερεθίσματα. Ένα προκαταρκτικό βήμα στην κατανόηση του βιολογικού ρόλου μιας πρωτεΐνης είναι η ταυτοποίηση των πρωτεϊνών με τις οποίες αλληλεπιδρά, ώστε να είναι δυνατή η ένταξή της σε ένα ή περισσότερα βιοχημικά μονοπάτια.

α. Προσδιορισμός αλληλεπιδράσεων με χρήση του συστήματος δύο υβριδίων του σακχαρομύκητα

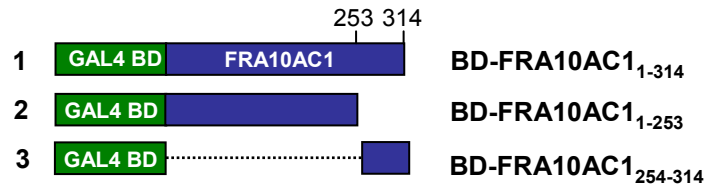
Καθώς στην FRA10AC1 δεν αναγνωρίζεται κάποια γνωστή λειτουργική περιοχή που θα μπορούσε να μας υποδείξει πιθανή λειτουργία της και επειδή για κανένα από τα ορθόλογα μόρια δεν είχε προσδιοριστεί ο βιοχημικός του ρόλος, εφαρμόστηκε το σύστημα δύο υβριδίων του σακχαρομύκητα ώστε να ταυτοποιηθούν πρωτεΐνες που αλληλεπιδρούν μαζί της. Το σύστημα που χρησιμοποιήθηκε στηρίζεται στην ανασυγκρότηση του μεταγραφικού παράγοντα GAL4.

Κατασκευή και έλεγχος του δολώματος

Ως δόλωμα χρησιμοποιήθηκε η πρωτεΐνη FRA10AC1 πλήρους μήκους ή τμήματά της που περιλαμβάνουν τα 253 N-τελικά ή τα 62 C-τελικά αμινοξέα της, σε σύντηξη με την DNA-BD του GAL4 (Εικ. 56).

Οι ανωτέρω πλασμιδιακές κατασκευές ελέγχθηκαν για αυτόνομη ενεργοποίηση των γονιδίων-δεικτών. Ο έλεγχος πραγματοποιήθηκε με μετασχηματισμό του στελέχους ρJ69-4α με κάθε μία από τις τρεις πλασμιδιακές κατασκευές σε συνδυασμό με τον φορέα της βιβλιοθήκης ρACT2 και με επακόλουθη καλλιέργεια των κυττάρων σε θρεπτικό μέσο επιλογής. Το πλήρους μήκους πολυπεπίδιο και τα 62 C-τελικά αμινοξέα του ενεργοποιούσαν αυτόνομα και τα τρία γονίδια αναφοράς (HIS3, ADE2, lacZ). Το γονίδιο HIS3 ενεργοποιούνταν ισχυρότατα, καθώς τα κύτταρα αναπτύσσονταν ακόμα και σε θρεπτικό με συγκέντρωση 20mM AT. Τα γονίδια ADE2 και lacZ ενεργοποιούνταν λιγότερο αλλά μετά από 5 μέρες παρατηρούνταν ανάπτυξη των αποικιών. Συνεπώς οι κατασκευές αυτές κρίθηκαν ακατάλληλες και η διερεύνηση πραγματοποιήθηκε με τα 253 N-τελικά αμινοξέα του μορίου.

Το φαινόμενο της αυτόνομης ενεργοποίησης παρατηρείται συχνά. Στην περίπτωση μας μπορεί να οφείλεται στην ενδογενή ικανότητα της FRA10AC1 να δρα ως μεταγραφικός ενεργοποιητής ή η ενεργοποίηση να είναι μη ειδική και να προκαλείται εξαιτίας του ιδιαίτερα όξινου C-τελικού άκρου της FRA10AC1.



Εικ. 56: Γραφική απεικόνιση των «δολωμάτων» που χρησιμοποιήθηκαν στο σύστημα δύο υβριδίων. Ως δόλωμα χρησιμοποιήθηκε το τμήμα της FRA10AC1 που περιλάμβανε τα 253 N-τελικά αμινοξέα (“2”), σε σύντηξη με την περιοχή πρόσδεσης στο DNA του GAL4. BD: περιοχή πρόσδεσης στο DNA

Έλεγχος και τιτλοδότηση της βιβλιοθήκης

Η cDNA βιβλιοθήκη που διερευνήθηκε προέρχεται από ήπαρ ανθρώπου, όπου το γονίδιο εμφανίζει υψηλά επίπεδα έκφρασης όπως δείχθηκε με ανάλυση Northern και RP-PCR. Επιπλέον, η παρουσία μεταγράφων του επιβεβαιώθηκε με PCR, με υπόστρωμα 100ng DNA και τα ζεύγη των εκκινητών F2-R2 και F3-R2 (Εικ. 37). Η θερμοκρασία υβριδοποίησης των εκκινητών είναι, για το πρώτο ζευγάρι 55°C ενώ για το δεύτερο 52°C για τους 5 αρχικούς κύκλους και 56°C για τους επόμενους 30 κύκλους. Από την αντίδραση προέκυψαν τα αναμενόμενα προϊόντα μεγέθους 129bp και 302bp αντίστοιχα.

Στη συνέχεια η βιβλιοθήκη τιτλοδοτήθηκε ώστε να καθοριστεί, αναλογικά, η ποσότητα του DNA με την καλύτερη απόδοση. Έτσι, μετά από μετασχηματισμό με 0.02, 0.2, 0.6, 2.0 και 5.0μg DNA προσδιορίστηκε ότι τα 2μg είναι η βέλτιστη ποσότητα, από την οποία προκύπτουν 2×10^5 αποικίες.

Διερεύνηση της βιβλιοθήκης

Βάσει της τιτλοδότησης πραγματοποιήθηκαν 25 μετασχηματισμοί (σε κύτταρα που είχαν ήδη μετασχηματιστεί με το δόλωμα) διατηρώντας την ποσότητα του DNA σταθερή στα 2μg, ώστε να διερευνηθούν 5×10^6 αποικίες, που θεωρήθηκαν αρκετές ώστε να υπάρχει ικανοποιητική αντιπροσώπευση.

Τα κύτταρα καλλιεργήθηκαν σε θρεπτικό υλικό επιλογής για το γονίδιο HIS3 και επωάστηκαν στους 30°C. Συνολικά αναπτύχθηκαν 81 αποικίες οι οποίες συλλέχθηκαν μετά από 4-8 μέρες. Ακολούθησε επανακαλλιέργειά τους σε πλήρες θρεπτικό υλικό που αποτέλεσε το πρωτότυπο πιάτο (master plate) από το οποίο δημιουργήθηκαν πιάτα-αντίγραφα για έλεγχο της ανάπτυξης των αποικιών στα θρεπτικά υλικά επιλογής. Για την επιλογή για το HIS3 παρασκευάστηκαν δύο αντίγραφα με 3mM και 10mM AT αντίστοιχα. Κατά την επιλογή απορρίφθηκαν 7 κλώνοι που δεν αναπτυσσόταν στο θρεπτικό υλικό όπου απουσίαζε η Ade. Έτσι επιλέχθηκαν 74 κλώνοι για περαιτέρω ανάλυση.

Απομόνωση, ανάλυση και επαλήθευση θετικών κλώνων

Από τους 74 θετικούς κλώνους απομονώθηκε DNA με το οποίο μετασχηματίστηκαν κύτταρα *E.coli* στελέχους KC8 ώστε να εξαλειφθεί το πλασμίδιο που κωδικοποιεί το δόλωμα. Ακολούθησε απομόνωση πλασμιδιακού DNA από το στέλεχος αυτό και τελικός μετασχηματισμός σε στέλεχος DH5a από όπου και απομονώθηκε το πλασμίδιο-θήραμα. Για την εξάλειψη των ομοειδών κλώνων, αυτοί τεμαχίστηκαν με τα ένζυμα *BamHI* και *XhoI*, που αναγνωρίζουν θέσεις στον polylinker του φορέα. Με βάση το πρότυπο τεμαχισμού, επιλέχθηκαν 42 κλώνοι, πιθανώς διαφορετικοί και παράλληλα προσδιορίστηκε κατά προσέγγιση το μέγεθος των ενθεμάτων τους (0,5-2,3kb). Σε όλους τους επιλεγμένους κλώνους προσδιορίστηκε η αλληλουχία των ενθεμάτων τους χρησιμοποιώντας τους εκκινητές pACT2F και pACT2R (βλ. Μεθόδους). Για την ταυτοποίηση των αλληλουχιών διερευνήθηκε η νουκλεοτιδική βάση δεδομένων (Blastn). Οι 22 κλώνοι απορρίφθηκαν είτε γιατί ήταν εκτός πλαισίου ανάγνωσης είτε γιατί είχαν καταγραφεί από προηγούμενες καταχωρήσεις ως ψευδώς-θετικά μόρια που εμφανίζονται συχνά σε ανάλογα πειράματα διερεύνησης (<http://www.uib.no/aasland/two-hybrid.html>).

Οι υπόλοιποι 20 κλώνοι αντιστοιχούσαν σε 2 γονίδια: 2 όμοιοι κλώνοι στο *SAP145* (spliceosome-associated protein **145**, GDB Id: Q13435) και 18, χωρισμένοι σε δύο αλληλοεπικαλυπτόμενες ομάδες, στο *DGCR14* (DiGeorge syndrome critical region protein **14**, GDB Id: NP_073210). Οι κλώνοι του *SAP145* κωδικοποιούν τα αμινοξέα 154-872 του μορίου, που έχει συνολικό μήκος 872 αμινοξέα ενώ οι κλώνοι του *DGCR14* διακρίνονται σε δύο ομάδες που αντιστοιχούν στα αμινοξέα 25-476 και 207-476 από τα 476 συνολικά του μορίου (Εικ. 57). Οι αλληλεπιδράσεις που ταυτοποιήθηκαν επιβεβαιώθηκαν με επαναληπτικό μετασχηματισμό κυττάρων με το δόλωμα και τα θηράματα και καλλιέργειά τους σε θρεπτικά υλικά επιλογής. Επίσης διαπιστώθηκε ότι ο ρυθμός ανάπτυξης των αποικιών όπου εκφραζόταν το ζεύγος FRA10AC1-DGCR14 ήταν σαφώς ταχύτερος σε σχέση με το FRA10AC1- SAP145. Η διαφορά αυτή μπορεί να αντανάκλα την ισχύ της αλληλεπίδρασης ή το ποσοστό των μορίων που συμμετέχουν σ' αυτήν. Επιπλέον, δεν παρατηρήθηκε διαφορά μεταξύ των δύο αλληλοεπικαλυπτόμενων τμημάτων της DGCR14, αποτέλεσμα που υποδηλώνει ότι η C-τελική περιοχή του μορίου είναι ικανή για την αλληλεπίδραση με την FRA10AC1. Είναι ενδιαφέρον το γεγονός ότι και οι δύο πρωτεΐνες που ταυτοποιήθηκαν φαίνεται να συμμετέχουν σε κοινό βιολογικό μονοπάτι, στη διαδικασία συναρμογής του mRNA. Συγκεκριμένα, η SAP145 αποτελεί υπομονάδα του παράγοντα συναρμογής SF3b. Το πρωτεϊνικό αυτό σύμπλοκο συμμετέχει στο σχηματισμό τόσο του U2 snRNP όσο και του U11/U12 δι-snRNP, συστατικά του μείζονος και του ελάσσονος spliceosome αντίστοιχα (Will CL *et al*, 1999) και είναι

μετασχηματίστηκαν είτε με τον φορέα pGEX-4T1 είτε με την πλασμιδιακή κατασκευή FRA₁₈₉₋₁₁₃₇/pGEX 4T-1 που περιέχει ολόκληρη την κωδική περιοχή του FRA10AC1 στο ίδιο αναγνωστικό πλαίσιο με την GST (Εικ. 58). Η GST παρήχθη μετά από επαγωγή με 1mM IPTG στους 30°C για 3h σε καλλιέργεια 50ml. Υπό αυτές τις συνθήκες, αφενός δεν επιτεύχθηκε υπερέκφραση της χιμαιρικής GST -FRA10AC1 αφετέρου το μόριο αποικοδομούταν, όπως φάνηκε μετά από ανάλυση Western με anti-GST. Επιπλέον παρατηρήθηκε τοξικότητα όπως φάνηκε με βάση τη δραματική μείωση της O.D. της βακτηριακής καλλιέργειας μετά την επαγωγή. Έτσι, έγινε βελτιστοποίηση των συνθηκών σε καλλιέργειες μικρής κλίμακας (3ml), δοκιμάζοντας διαφορετικούς χρόνους και θερμοκρασίες επαγωγής, παρουσία ή απουσία γλυκόζης στο θρεπτικό καλλιέργειας, ενώ τα επίπεδα έκφρασης και η ακεραιότητα του πολυπεπτιδίου ελέγχθηκαν με Western. Μετά τον καθορισμό των βέλτιστων συνθηκών (28°C για 19h παρουσία γλυκόζης), πραγματοποιήθηκε η παραγωγή της πρωτεΐνης σε μεγάλη κλίμακα (200ml). Καθώς η χιμαιρική πρωτεΐνη εντοπιζόταν στο αδιάλυτο κλάσμα του βακτηριακού εκχυλίσματος πραγματοποιήθηκε κατάλληλη επεξεργασία που οδήγησε σε διαλυτοποίηση του μεγαλύτερου μέρους της.

Τα πολυπεπτίδια απομονώθηκαν από το εκχύλισμα με χρωματογραφία συγγένειας με σφαιρίδια σεφαρόζης (0,15ml) συζευγμένα με γλουταθειόνη και ποσοτικοποιήθηκαν βάσει σύγκρισης με πρωτεϊνικό δείκτη γνωστών μεγεθών και ποσοτήτων μετά από SDS-PAGE και χρώση με Coomassie blue.

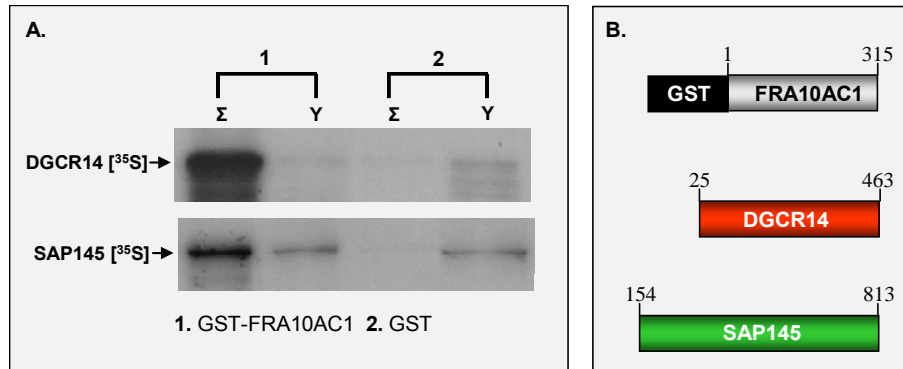
Σύνθεση των πρωτεϊνών SAP145 και DGCR14

Οι ανασυνδυασμένες πρωτεΐνες SAP145 (154-813aa) και DGCR14 (25-463aa) παρήχθησαν με συζευγμένη μεταγραφή/μετάφραση *in vitro*. Ως υπόστρωμα χρησιμοποιήθηκαν οι πλασμιδιακές κατασκευές DGCR14/pCDNA3.1 HisA και SAP145/pCDNA3.1 HisA (βλ. Πλασμιδιακές κατασκευές) ενώ η αντίδραση πραγματοποιήθηκε παρουσία ³⁵S-Met. Τα προϊόντα της αντίδρασης εμφανίζονταν ως μία ζώνη του αναμενόμενου μεγέθους μετά από ηλεκτροφόρηση και αυτοραδιογραφία.

Η FRA10AC1 αλληλεπιδρά *in vitro* με τις πρωτεΐνες SAP145 και DGCR14

Στην δοκιμή συνκατακρήμνισης συμμετείχαν 5μg GST-FRA10AC1 (συζευγμένης με γλουταθειόνη-σεφαρόζη) και 10μl (40%) της αντίδρασης μεταγραφής/μετάφρασης των άλλων πρωτεϊνών. Η επώαση πραγματοποιήθηκε σε 200 μl διαλύματος δέσμησης, ολονύκτια στους 4°C, υπό ανάδευση. Ακολούθησαν εκπλύσεις με το ίδιο διάλυμα, SDS-PAGE και αυτοραδιογραφία. Για τον αρνητικό έλεγχο, ακολουθήθηκε η ίδια διαδικασία με την GST μόνη της. Τα αποτελέσματα επιβεβαιώνουν την άμεση αλληλεπίδραση της πλήρους μήκους FRA10AC1 με τις SAP145 και DGCR14 (Εικ. 58). Επιπλέον, φαίνεται πως η αλληλεπίδραση με την

DGCR14 είναι ισχυρότερη από αυτήν με την SAP145, φαινόμενο που είχε παρατηρηθεί και κατά το σύστημα δύο υβριδίων. Βέβαια υπάρχει η πιθανότητα αυτό να οφείλεται στην ελλειμματικότητα του κλώνου του *SAP145* που ταυτοποιήθηκε στο σύστημα δύο υβριδίων ο οποίος στερείται τα 153 αμινοτελικά αμινοξέα, που αντιστοιχούν στο ~15% του μορίου.



Εικ. 58: Αλληλεπίδραση της FRA10AC1 με τις DGCR14 και SAP145 *in vitro*.

A. Για την δοκιμή χρησιμοποιήθηκαν ισομοριακές ποσότητες της GST και της GST-FRA10AC1 συζευγμένων με σφαιρίδια σεφαρόζης-γλουταθειόνης και οι ραδιοσημασμένες πρωτεΐνες DGCR14 και SAP145 που παρήχθησαν με μεταγραφή/μετάφραση *in vitro*. Ο έλεγχος των πρωτεϊνών που είναι συνδεδεμένες με τα σφαιρίδια έγινε μετά από SDS-PAGE σε πήκτωμα 10%, σταθεροποίηση, ξήρανση και αυτοραδιογραφία του μετά από ολονύκτια έκθεση. Σ: πρωτεΐνες συνδεδεμένες στα σφαιρίδια, Υ: υπερκείμενο- το 10% του συνολικού **B.** Σχηματική απεικόνιση των πολυπεπτιδίων που συμμετείχαν στην δοκιμή: η FRA10AC1 πλήρους μήκους, η DGCR14 (25-463aa) και η SAP145 (154-813aa).

- Χαρτογράφηση των περιοχών αλληλεπίδρασης των πρωτεϊνών FRA10AC1 και DGCR14

Η αλληλεπίδραση των πρωτεϊνών FRA10AC1 και DGCR14 μελετήθηκε περαιτέρω ώστε να χαρτογραφηθούν οι υπεύθυνες περιοχές. Ο έλεγχος πραγματοποιήθηκε με δοκιμή συνκατακρήμνισης κατά την οποία ελέγχθηκαν τμήματα της DGCR14 που παρήχθησαν σε βακτήρια σε σύντηξη με την GST με περιοχές της FRA10AC1 που παρήχθησαν σε κύτταρα COS7 σε σύντηξη με την EGFP.

Παραγωγή και απομόνωση τμημάτων της πρωτεΐνης DGCR14 σε βακτήρια

Οι χιμαιρικές πρωτεΐνες παρήχθησαν στο στέλεχος BL21 (DE3) CodonPlus της *E.coli* μετά από μετασχηματισμό των κυττάρων με τις πλασμιδιακές κατασκευές: 1) DGCR14₂₅₋₄₇₂_pGEX 3X, 2) DGCR14₂₅₋₂₂₂_pGEX 4T-2, 3) DGCR14₂₂₃₋₄₇₂_pGEX 3X, 4) DGCR14₂₉₄₋₄₇₂_pGEX 3X, 5) DGCR14₂₁₈₋₃₁₆_pGEX 3X, 6) DGCR14₃₉₀₋₄₇₂_pGEX 3X και 7) pGEX 3X. Η επαγωγή της έκφρασής τους πραγματοποιήθηκε με 1mM IPTG στους 28°C για 16h ή με 0,1mM IPTG για 2h (GST- DGCR14₂₅₋₄₇₂ και GST-DGCR14₂₉₄₋₄₇₂) ή με 1mM IPTG για 4h (GST- DGCR14₂₅₋₂₂₂) σε καλλιέργεια 200ml. Ακολούθησε κατάλληλη επεξεργασία του βακτηριακού εκχυλίσματος ώστε να

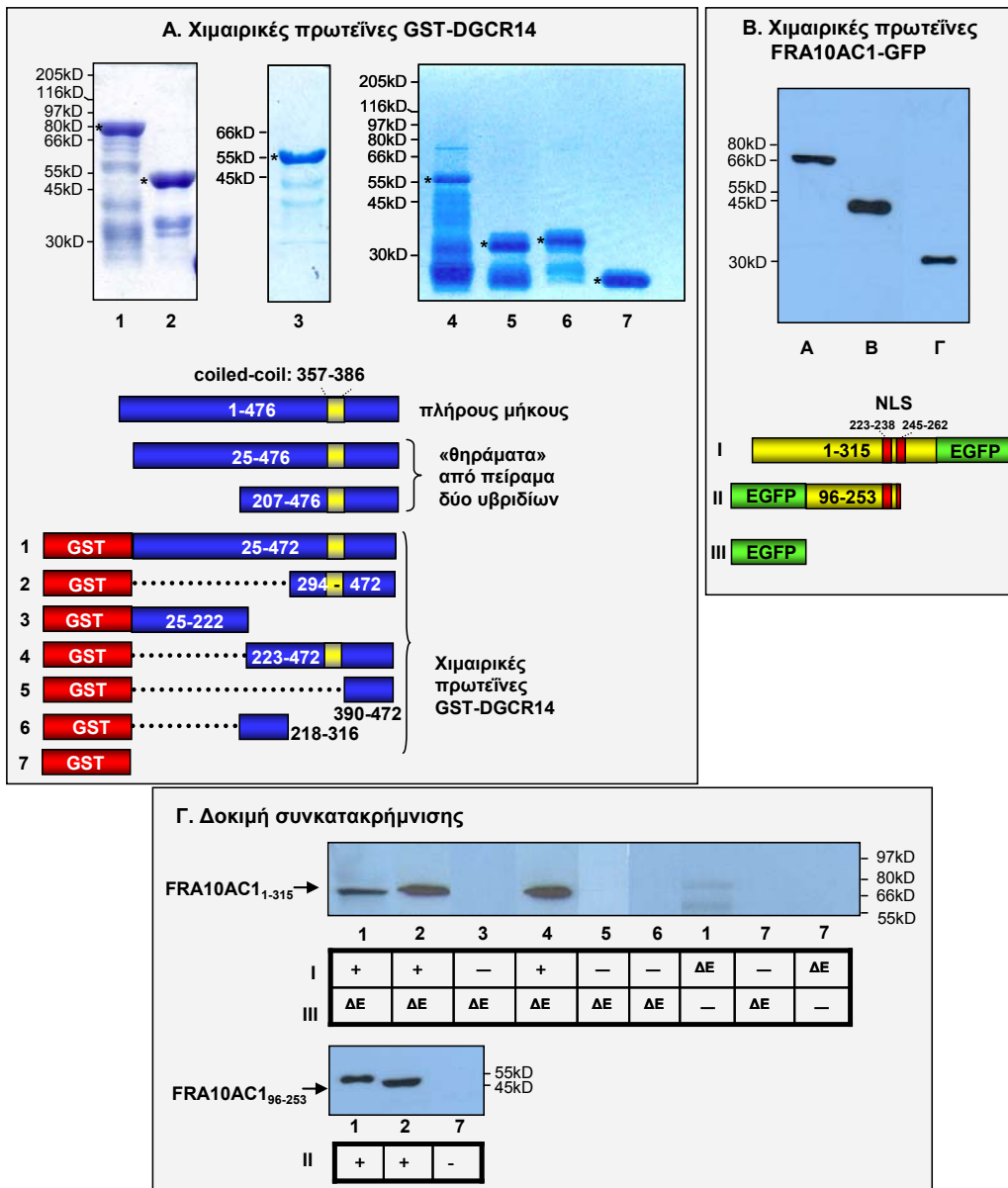
διαλυτοποιηθούν οι χιμαιρικές πρωτεΐνες. Το αντίστοιχο εκχύλισμα της GST εντοπίζεται στο διαλυτό κλάσμα. Στη συνέχεια, τα πολυπεπτίδια απομονώθηκαν με χρωματογραφία συγγένειας χρησιμοποιώντας σφαιρίδια σεφαρόζης (0,15ml) συζευγμένα με γλουταθειόνη. Μετά την ποσοτικοποίησή τους (Εικ. 59), εξισορροπήθηκαν σε διάλυμα RIPA και χρησιμοποιήθηκαν στη δοκιμή συνκατακρήμισης.

Παραγωγή τμημάτων της πρωτεΐνης FRA10AC1 σε κύτταρα θηλαστικών COS-7

Οι χιμαιρικές πρωτεΐνες παρήχθησαν σε κύτταρα COS-7 μετά από παροδική διαμόλυνση με τις πλασμιδιακές κατασκευές FRA10AC1_full/pEGFPN1, FRA10AC1_Bbs_Stu/pEGFPC1 και pEGFPC1 που κωδικοποιούν το πλήρους μήκους πολυπεπτίδιο, το κεντρικό τμήμα του (aa:96-253) και την EGFP, αντίστοιχα (Εικ.59). Τα κύτταρα συλλέχθηκαν μετά από επώαση 36h, λύθηκαν με διάλυμα RIPA και το εκχύλισμα χρησιμοποιήθηκε στη δοκιμή συνκατακρήμισης.

Δοκιμή συνκατακρήμισης

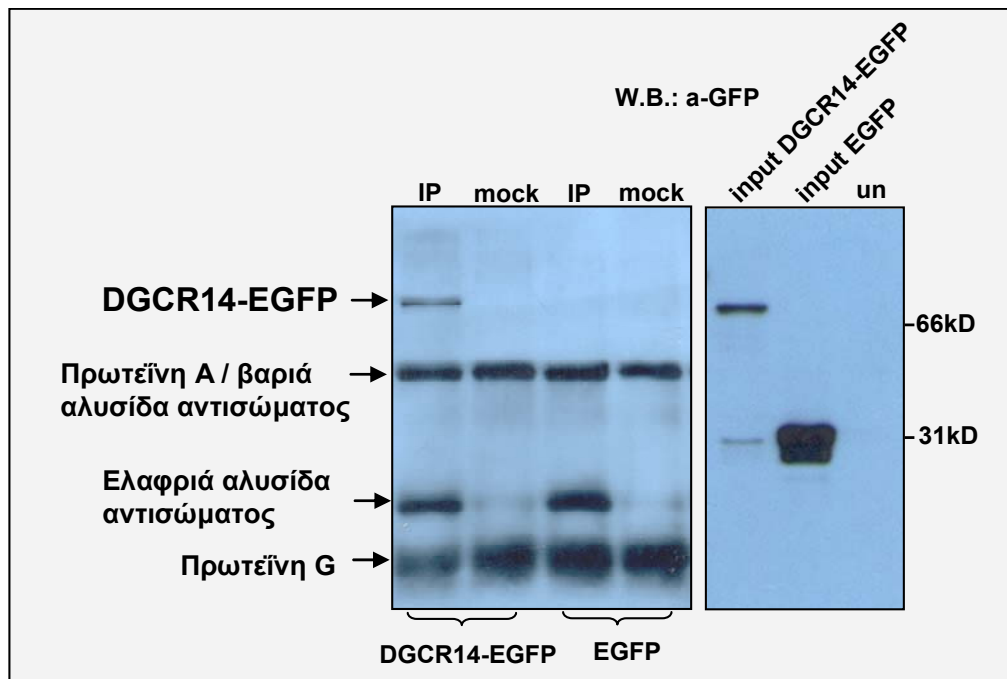
Για τη δοκιμή συνκατακρήμισης χρησιμοποιήθηκε ~1μg των χιμαιρικών πρωτεϊνών GST-DGCR14 ή της GST (συζευγμένων με γλουταθειόνη-σεφαρόζη) και το πρωτεϊνικό εκχύλισμα κυττάρων COS-7 που είχαν διαμολυνθεί παροδικά με τις κατάλληλες πλασμιδιακές κατασκευές. Η επώαση πραγματοποιήθηκε σε 300 μl διαλύματος RIPA, στους 4°C υπό ολονύκτια ανάδευση. Ακολούθησαν εκπλύσεις με το ίδιο διάλυμα, SDS-PAGE, υγρή ηλεκτρομεταφορά των πολυπεπτιδίων σε μεμβράνη και Western με το αντίσωμα α-GFP (Εικ. 59). Κατά την αρχική διερεύνηση, ελέγχθηκε η αλληλεπίδραση της FRA10AC1 πλήρους μήκους με διαφορετικά τμήματα της DGCR14. Το αποτέλεσμα έδειξε ότι το τμήμα της πρωτεΐνης DGCR14 που περιλαμβάνει τα αμινοξέα 294-472 είναι ικανό και αναγκαίο για την παρατηρούμενη αλληλεπίδραση. Επιπλέον, φάνηκε ότι δύο υποπεριοχές αυτού του τμήματος, που αντιστοιχούν στα αμινοξέα 218-316 και 390-472, δεν είναι ικανές να αλληλεπιδράσουν με την FRA10AC1, γεγονός που συνεπάγεται ότι το τμήμα της πρωτεΐνης μεταξύ των αμινοξέων 317-390 είναι απολύτως απαραίτητο (Εικ. 59). Πρέπει να σημειωθεί ότι σ' αυτό το τμήμα, μήκους ~70 αμινοξέων, χαρτογραφείται το μοτίβο coiled coil, που έχει συσχετιστεί με πρωτεϊνικές αλληλεπιδράσεις. Στη συνέχεια και με δεδομένη την υψηλή συντήρηση που εμφανίζει η κεντρική περιοχή της πρωτεΐνης FRA10AC1 στα ορθόλογα μόρια, μελετήθηκε η αλληλεπίδραση αυτού του τμήματος (αμινοξέα: 96-253) με την περιοχή της DGCR14 που ταυτοποιήθηκε προηγουμένως (αμινοξέα 294-472). Πράγματι, με τη δοκιμή συνκατακρήμισης, φάνηκε ότι η άμεση αλληλεπίδραση μεταξύ των πρωτεϊνών FRA10AC1 και DGCR14 επιτυγχάνεται μέσω των συγκεκριμένων τμημάτων (Εικ. 59).



Εικ. 59: Χαρτογράφηση των περιοχών αλληλεπίδρασης των πρωτεϊνών FRA10AC1 και DGCR14. **A.** Η πρωτεΐνη πλήρους μήκους DGCR14 και τμήματά της παρήχθησαν σε βακτήρια σε σύντηξη με την GST. Οι ανασυνδυασμένες πρωτεΐνες ελέγχθηκαν με SDS-PAGE (10%) και ακόλουθη χρώση με Coomassie blue (πάνω). Τα πολυπεπτιδία # 1-7 απεικονίζονται σχηματικά και συγκρίνονται με την πλήρους μήκους πρωτεΐνη καθώς και με τα τμήματά της που αντιστοιχούν στους κλώνους που ταυτοποιήθηκαν στο πείραμα δύο υβριδίων. **B.** Η πρωτεΐνη πλήρους μήκους FRA10AC1 και το κεντρικό της τμήμα (96-256aa) παρήχθησαν σε κύτταρα COS-7 σε σύντηξη με την EGFP. Φαίνεται το αποτέλεσμα της ανάλυσης Western των ανασυνδυασμένων πρωτεϊνών χρησιμοποιώντας το αντίσωμα α-GFP (πάνω) και η σχηματική απεικόνιση των πολυπεπτιδίων. **Γ.** Δοκιμή συνκατακρήμνισης μεταξύ των διαφορετικών τμημάτων των πρωτεϊνών. Σε κάθε δοκιμή χρησιμοποιήθηκαν ~1μg των χιμαιρικών πρωτεϊνών GST-DGCR14. Η πρωτεΐνη FRA10AC1 πλήρους μήκους αλληλεπιδρά με τα τμήματα #1, #2 και #4 της DGCR14, όπως φαίνεται από την ανάλυση Western με το α-GFP (πάνω). Οι αριθμοί 1-7 κάτω από το φιλμ αντιστοιχούν στα τμήματα της DGCR14 που χρησιμοποιήθηκαν σε κάθε δοκιμή, βάσει της αρίθμησης στην Εικ. (B) και το I και III αντιστοιχούν στην FRA10AC1 πλήρους μήκους και την EGFP (αρνητικός έλεγχος) αντίστοιχα. Το (+) συμβολίζει την επιβεβαίωση της αλληλεπίδρασης, (ΔΕ: δεν πραγματοποιήθηκε). Στο κάτω μέρος φαίνεται η επιβεβαίωση της αλληλεπίδρασης μεταξύ του τμήματος II της FRA10AC1 και των τμημάτων 1 και 2 της DGCR14, όπως απεικονίζονται στις Εικ. (A) και (B). Το τμήμα II της FRA10AC1 δεν αλληλεπιδρά με την GST (αρνητικός έλεγχος).

γ. Επιβεβαίωση πρωτεϊνικών αλληλεπιδράσεων *in vivo*

Η ταυτόχρονη παρουσία των πρωτεϊνών FRA10AC1 και DGCR14 στο ίδιο πρωτεϊνικό σύμπλοκο ελέγχθηκε με πείραμα συν-ανοσοκατακρήμισης. Ειδικότερα, πραγματοποιήθηκε ανοσοκατακρήμιση της ενδογενούς FRA10AC1 με το αντίσωμα anti-FRA10AC1 σε εκχυλίσματα κυττάρων COS-7, στα οποία είχε επιτευχθεί υπερέκφραση της χιμαιρικής πρωτεΐνης EGFP-DGCR14 (Εικ. 60). Η υπερέκφραση πραγματοποιήθηκε με παροδική διαμόλυνση με την πλασμιδιακή κατασκευή DGCR14_pEGFP_C1 που κωδικοποιεί σχεδόν την πλήρους μήκους πρωτεΐνη (στερείται τα N-τελικά 24 αμινοξέα). Η παρουσία της EGFP-DGCR14 στο ανοσοκατακρήμισμα επιβεβαιώθηκε με ανάλυση Western με το αντίσωμα α-GFP. Για τον αρνητικό έλεγχο, ακολουθήθηκε η ίδια διαδικασία, σε κυτταρικά εκχυλίσματα όπου είχε υπερεκφραστεί η πρωτεΐνη EGFP μόνη της (Εικ. 60).



Εικ. 60: Η ενδογενής πρωτεΐνη FRA10AC1 συν-ανοσοκατακρημνίζεται με την ανασυνδυασμένη DGCR14-EGFP. Η ανοσοκατακρήμιση πραγματοποιήθηκε με ~1μg αντισώματος α-FRA10AC1 σε πρωτεϊνικό εκχύλισμα κυττάρων COS-7 στα οποία είχαν εκφραστεί παροδικά οι πρωτεΐνες DGCR14-EGFP και EGFP (αρνητικός έλεγχος) και ελέγχθηκε με ανάλυση Western με το αντίσωμα α-GFP (δεξιά). Η DGCR14-EGFP ανιχνεύεται στο ανοσοκατακρήμισμα (βέλος) ενώ δεν ανιχνεύεται η EGFP. IP=συν-ανοσοκατακρήμιση, mock=παράλληλη επανάληψη του πειράματος απουσία αντισώματος. Επίσης σημειώνεται η προέλευση των επιπλέον ζωνών που ταυτοποιούνται. Δεξιά, φαίνεται το αποτέλεσμα ανάλυσης Western με α-GFP στο 1/10 του πρωτεϊνικού εκχυλίσματος που χρησιμοποιήθηκε στην συν-ανοσοκατακρήμιση.

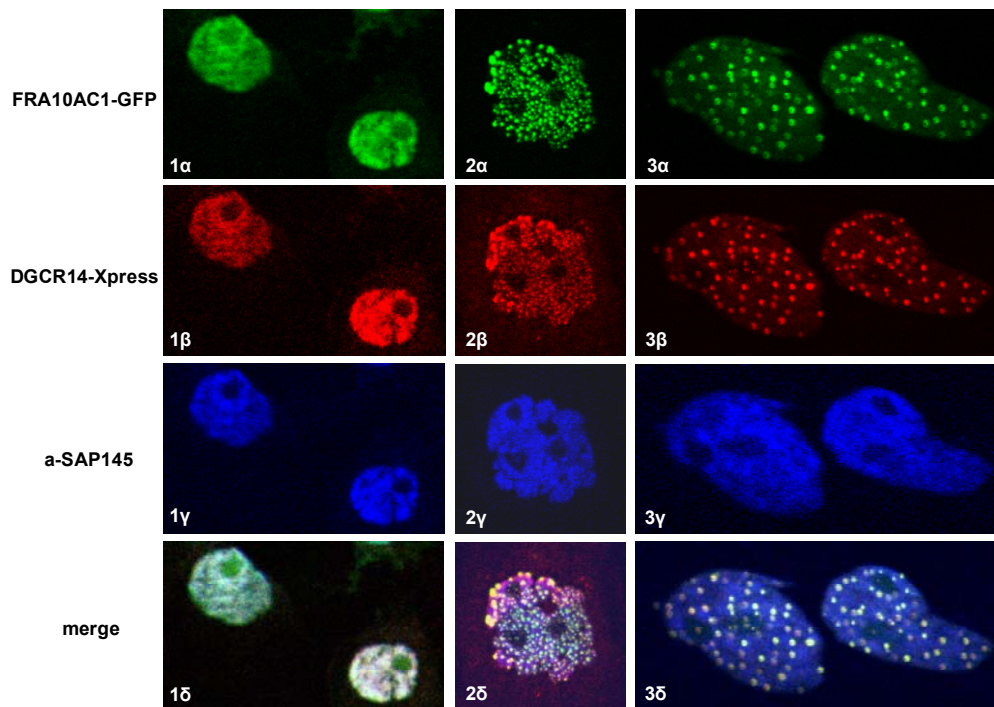
➤ **Διερεύνηση της υποκυτταρικής τοπολογίας των πρωτεϊνών FRA10AC1, DGCR14 και SAP145**

Για την πρωτεΐνη FRA10AC1 δείξαμε ότι εντοπίζεται στον πυρήνα ενώ για τις πρωτεΐνες SAP145 και DGCR14 είναι γνωστό από παλαιότερες μελέτες ότι είναι πυρηνικές στα θηλαστικά (Das BK *et al*, 1999; Lindsay EA *et al*, 1998). Το ερώτημα που τέθηκε και που αποτελεί συνέχεια της επιβεβαίωσης των μεταξύ τους αλληλεπιδράσεων, είναι εάν οι τρεις πρωτεΐνες συνεντοπίζονται και υποκυτταρικά.

Η συγκριτική μελέτη του προτύπου έκφρασής τους πραγματοποιήθηκε με μικροσκοπία φθορισμού σε κύτταρα HeLa στα οποία ανιχνεύσαμε την ενδογενή SAP145 και τις χιμαιρικές FRA10AC1-EGFP και Xpress-DGCR14 (Εικ. 61). Τα κύτταρα διαμολύνθηκαν παροδικά με τις πλασμιδιακές κατασκευές FRA_full_pEGFP_N1 που κωδικοποιεί το πλήρους μήκους πολυπεπτιδίο και DGCR14_pcDNA3.1_HisA, που κωδικοποιεί την σχεδόν πλήρους μήκους πρωτεΐνη (25-472aa). Για τον εντοπισμό των πρωτεϊνών χρησιμοποιήθηκαν το πολυκλωνικό αντίσωμα a-SAP145 (δώρο της Dr. R.Reed) και το μονοκλωνικό αντίσωμα a-Xpress που αναγνωρίζει τον επίτοπο επισήμανσης της DGCR14.

Από την ανάλυση (Εικ. 61) φάνηκε ότι οι πρωτεΐνες FRA10AC1 και DGCR14 έχουν ταυτόσημο πρότυπο κατανομής, τόσο σε κύτταρα όπου η συσσώρευσή τους είναι σχετικά ομοιογενής όσο και σε κύτταρα όπου οι πρωτεΐνες εντοπίζονται σε στικτώδεις δομές στον πυρήνα. Αυτές οι έντονα διακριτές εστίες δεν παρατηρήθηκαν στην περίπτωση της πρωτεΐνης SAP145, ωστόσο είναι δυνατή η διάκριση πυρηνικών περιοχών όπου η συσσώρευσή της είναι μεγαλύτερη. Σε αυτές τις περιοχές, σε σημαντικό ποσοστό των κυττάρων, παρατηρήθηκε και αυξημένη συσσώρευση των πρωτεϊνών FRA10AC1 και DGCR14. Ο παρατηρούμενος συνεντοπισμός αποτελεί θετική ένδειξη για τη λειτουργική συσχέτιση των μορίων.

Η υπόθεση για την ταυτότητα των στικτών δομών είναι ότι είτε πρόκειται για συσσωματώματα είτε για πυρηνικά διαμερίσματα-χώρους συσσώρευσης-αποθήκευσης των πρωτεϊνών, που σχηματίζονται σε κατάσταση υπερ-αφθονίας τους, όπως συμβαίνει κατά την παροδική διαμόλυνση. Πιθανώς, αυτός είναι ο λόγος που η SAP145 δεν εμφανίζει έντονα στικτό πρότυπο καθώς ανιχνεύουμε την ενδογενή πρωτεΐνη, που τα επίπεδά της είναι σαφώς χαμηλότερα από αυτά των ανασυνδυασμένων FRA10AC1 και DGCR14.



Εικ. 61: Υποκυτταρικός συνεντοπισμός των πρωτεϊνών FRA10AC1, DGCR14 και SAP145. Σε όλα τα κύτταρα που εξετάστηκαν, οι ανασυνδυασμένες πρωτεΐνες FRA10AC1-EGFP (πράσινο) και DGCR14-Xpress (κόκκινο) εμφανίζουν ταυτόσημη πυρηνική κατανομή (σύγκριση 1α-1β, 2α-2β, 3α-3β), διάχυτη (1α-1β) ή στικτή (2α-2, 3α-3β). Στην πλειονότητα των κυττάρων (~90%) και οι τρεις πρωτεΐνες εμφανίζουν σχετικά διάχυτη πυρηνική κατανομή (1α, 1β, 1γ) και η σύγκριση των προτύπων τους (1δ) παρουσιάζει σημαντική αλληλοεπικάλυψη (άσπρο). Όλες οι εικόνες προέρχονται από οπτικές τομές σε συνεστιακό μικροσκόπιο φθορισμού.

➤ **Συγκριτική μελέτη του προτύπου έκφρασης των γονιδίων *Fra10Ac1*, *Dgcr14* και *Sap145* κατά την εμβρυική ανάπτυξη και στον εγκέφαλο του ποντικού**

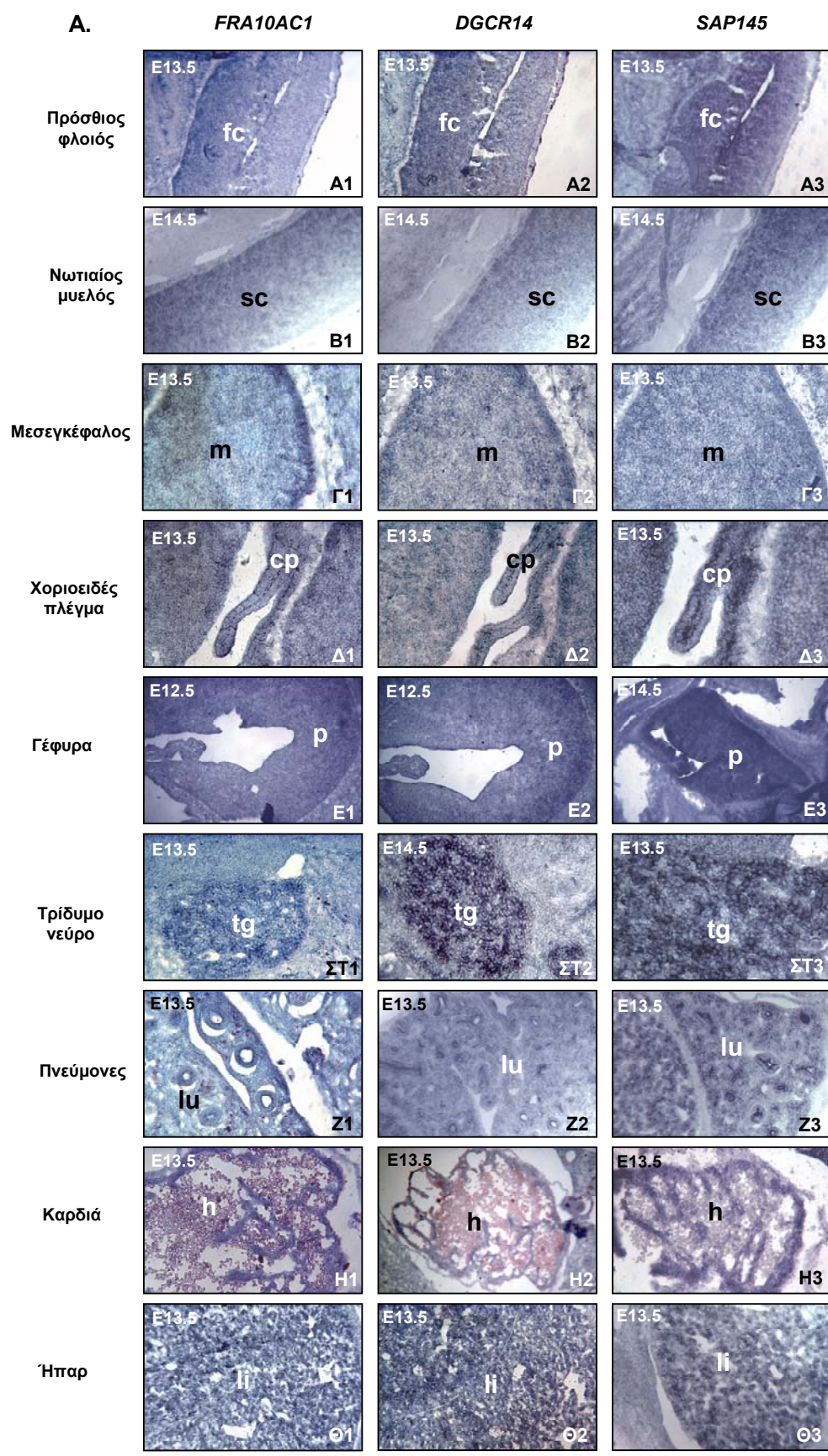
Απαραίτητη προϋπόθεση για την λειτουργική συσχέτιση των πρωτεϊνών FRA10AC1, DGCR14 και SAP145, εκτός από την επιβεβαίωση των πρωτεϊνικών αλληλεπιδράσεων τους και τον υποκυτταρικό συνεντοπισμό τους, είναι η επικαλυπτόμενη μεταγραφική ενεργότητα των αντίστοιχων γονιδίων σε επίπεδο οργανισμού. Επιπλέον, ο τοπολογικός και χρονικός προσδιορισμός της έκφρασης των γονιδίων προσφέρουν πολύτιμες πληροφορίες για την συνεισφορά τους στην οντογένεση του οργανισμού.

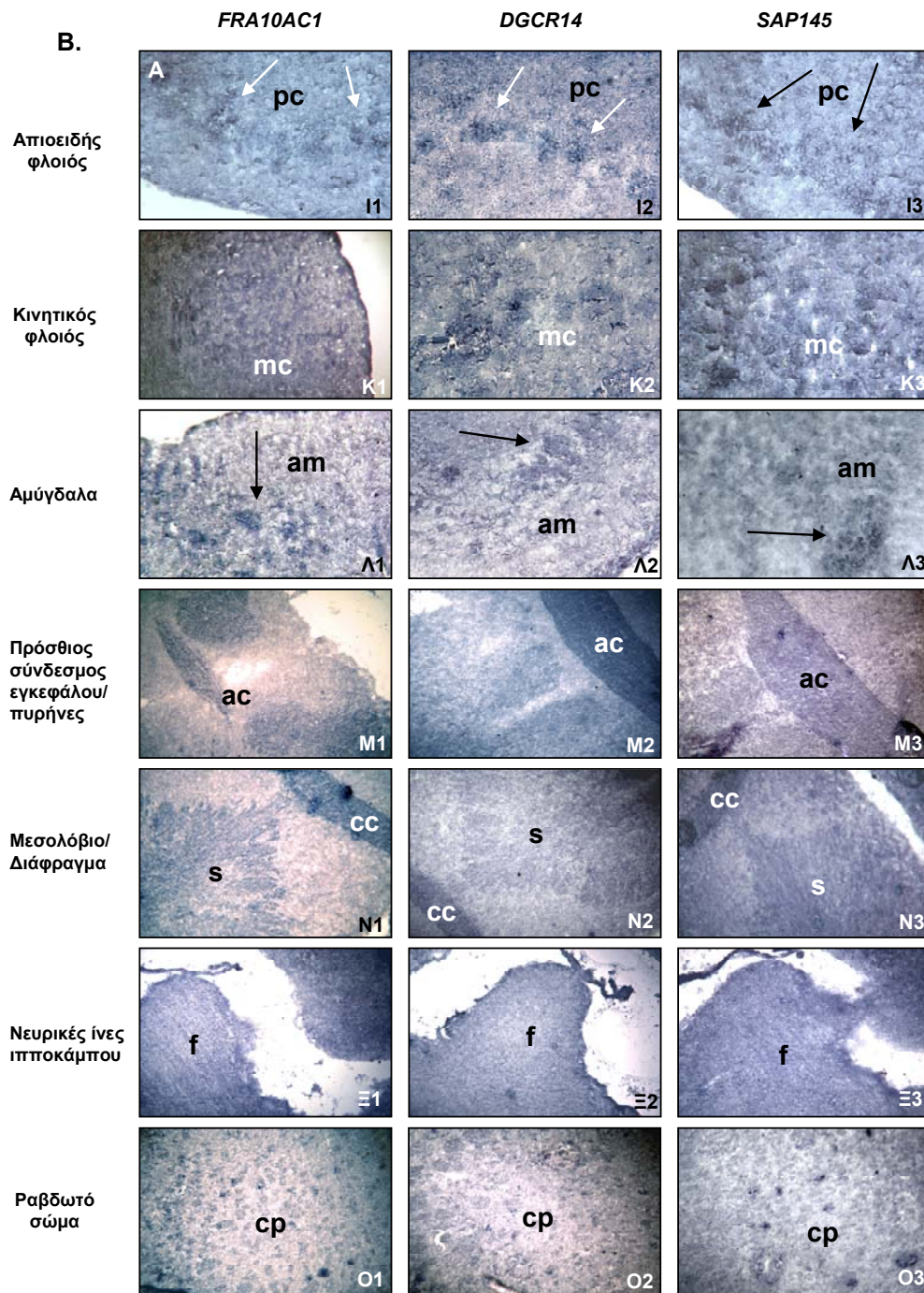
Ο συγκριτικός έλεγχος του προτύπου έκφρασής τους πραγματοποιήθηκε με πειράματα υβριδοποίησης *in situ* σε οβελιαίες τομές ολόκληρων εμβρύων αναπτυξιακών σταδίων E12.5, E13.5 και E14.5 και σε στεφανιαίες τομές εγκεφάλου ενήλικου ποντικού και με αναφορά τους ιστούς και τις δομές όπου είχαμε παρατηρήσει υψηλή μεταγραφική ενεργότητα του γονιδίου *Fra10ac1*. Οι ριβοανιχνευτές σχεδιάστηκαν με βάση τα ορθόλογα γονίδια του ποντικού, αντιστοιχούν κατά το μεγαλύτερο μέρος τους στις περιοχές 3'UTR και παρήχθησαν

με μεταγραφή *in vitro* μετά από γραμμοποίηση των πλασμιδιακών κατασκευών *Fra10ac1_pBluescript KS+*, *Dgcr14_pBluescript KS+* και *Sap145_pBluescript KS+* (Εικ. 11).

Από την ανάλυση (Εικ. 62) προέκυψε ότι τα γονίδια *Sap145* και *Dgcr14* είναι μεταγραφικά ενεργά σε όλους σχεδόν τους ιστούς που ελέγχθηκαν συμπεριλαμβανομένου και αυτών όπου είχαμε ανιχνεύσει υψηλά επίπεδα μεταγράφων του *Fra10ac1*. Ειδικότερα, εντοπίστηκε συνέκφραση των τριών γονιδίων τόσο στο νευρικό σύστημα [πρόσθιος φλοιός, κατά μήκος του νωτιαίου μυελού, εγκεφαλικό στέλεχος (μεσεγκέφαλος, γέφυρα), χοριοειδές πλέγμα, τρίδυμο νεύρο] όσο και σε εσωτερικά όργανα του εμβρύου (πνεύμονες, καρδιά, ήπαρ), Εικ. 62Α. Επιπλέον, σημαντική αλληλοεπικάλυψη παρατηρήθηκε σε δομές του εγκεφάλου του ενηλίκου ποντικού [φλοιός (απιοειδής, κινητικός), αμύγδαλα, πρόσθιος σύνδεσμος εγκεφάλου, μεσολόβιο, διάφραγμα, νευρικές ίνες ιπποκάμπτου, ραβδωτό σώμα], Εικ. 62Β. Επίσης, παρατηρήθηκε υψηλότερη έκφραση του *Sap145* και συγκρίσιμη των *Fra10ac1* και *Dgcr14*.

Το συγκριτικό πρότυπο έκφρασης των τριών γονιδίων συμφωνεί με τα αποτελέσματα της ανάλυσης της αλληλεπίδρασης των αντίστοιχων πρωτεϊνών. Όσον αφορά τον βιολογικό τους ρόλο, υποδεικνύει πρωτεΐνες που είναι απαραίτητες από τα πρώιμα αναπτυξιακά στάδια σε όλους τους ιστούς, υποδηλώνοντας πρωτεΐνες «γενικών καθηκόντων».





Εικ. 62: Συγκριτική μελέτη του προτύπου έκφρασης των γονιδίων *Fra10ac1*, *Dgcr14* και *Sap145* με RNA *in situ* υβριδοποίηση σε οβελιαίες τομές εμβρύων ποντικού ηλικίας E12.5, E13.5 και E14.5 (A) και σε στεφανιαίες τομές εγκεφάλου ενήλικου ποντικού (B). Στα έμβρυα, διαπιστώθηκε συνέκφραση των τριών γονιδίων στον πρόσθιο φλοιό (fc: frontal cortex, A1-A3), το νωτιαίο μυελό (sc: spinal cord, B1-B3), το μεσεγκέφαλο (m: midbrain, Γ1-Γ3), το χοριοειδές πλέγμα (cp: choroid plexus, Δ1-Δ3), τη γέφυρα (p: pons, E1-E3), το τρίδυμο νεύρο (tg : trigeminal nerve, ΣΤ1-ΣΤ3), τους πνεύμονες (lu: lung, Z1-Z3), την καρδιά (h: heart, Η1-H3) και το ήπαρ (li: liver, Θ1-Θ3). Στον εγκέφαλο συνέκφραση παρατηρήθηκε στον απιοειδή (pc: piriform cortex, I1-I3), τον κινητικό φλοιό (mc: motor cortex, K1-K3), τα αμύγδαλα (am: amygdala, Λ1-Λ3), τον πρόσθιο σύνδεσμο του εγκεφάλου (ac: anterior commissure, M1-M3), το μεσολόβιο (cc: corpus callosum, N1-N3), το διάφραγμα (s: septum pellucidum, N1-N3), σε νευρικές ίνες του ιπποκάμπτου (f: fimbria hippocampus, Ξ1-Ξ3) και το ραβδωτό σώμα (cp: caudate putamen, O1-O3).

4

ΣΥΖΗΤΗΣΗ

4.1. Γονιδιωματική ανάλυση του χρωμοσώματος 10

Το πρώτο μέρος της εργασίας συνέβαλε στην κατασκευή του φυσικού χάρτη των περιοχών 10q23.31-q25.3 και 10q26.2-qter που αθροιστικά αποτελούν 35Mb, αντιστοιχώντας στο 25% του χρωμοσώματος. Ο χάρτης αυτός αποτέλεσε τη βάση για τον προσδιορισμό της αλληλουχίας των αντίστοιχων χρωμοσωμικών περιοχών. Η τελική αλληλουχία και ο χαρακτηρισμός της *in silico* ολοκληρώθηκαν το 2004 (Deloukas P *et al*, 2004), καθιστώντας ιστορικά το χρωμόσωμα αυτό, το όγδοο κατά χρονολογική σειρά χρωμόσωμα που αναλύεται πλήρως (Εικ. 63).

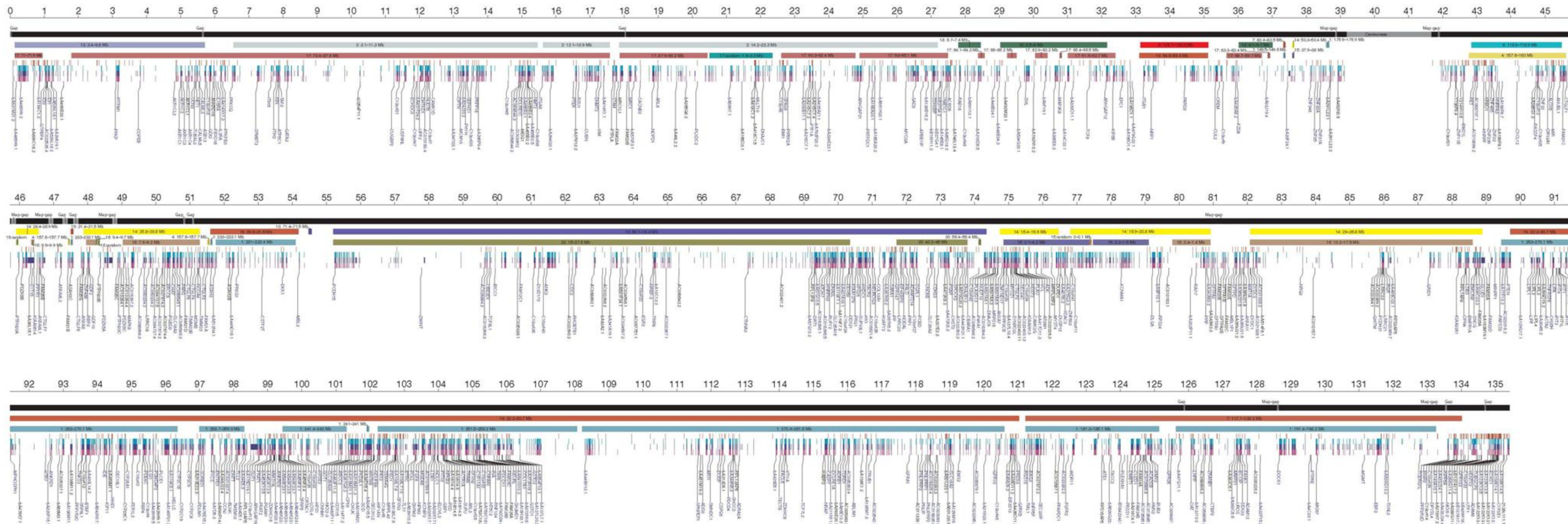
➤ Χάρτης κλώνων-αλληλουχίας

Το φυσικό μήκος του χρωμοσώματος είναι 135Mb που αντιστοιχεί στο 4,3% του γονιδιώματος (3,08Gb). Το ελάχιστο αλληλοεπικαλυπτόμενο μονοπάτι συνίσταται από 1,144 κλώνους, ως επί το πλείστον BACs (1095). Ο μικρός βραχίονας καλύπτεται ολόκληρος από ένα ενιαίο contig κλώνων που περιέχει το όριο ευχρωματίνης-ετεροχρωματίνης, ενώ τα ακραία 250kb φέρουν τις περικεντρομερικές δορυφορικές επαναλήψεις (Εικ. 21, Εικ. 63). Στο μεγάλο βραχίονα, ο φυσικός χάρτης εμφανίζει 9 χάσματα, 5 από τα οποία χαρτογραφούνται σε μία περιοχή ~4Mb, κοντά στο κεντρομερές (Εικ. 63). Η παρουσία τους σ' αυτήν την θέση οφείλεται στην ύπαρξη εκτεταμένων διπλασιασμών ενώ το μέγεθός τους, όπως εκτιμήθηκε με FISH, δεν ξεπερνά τα 50kb το καθένα. Συνολικά, το μέγεθος όλων των ευχρωματινικών χασμάτων είναι μικρότερο από 840kb. Όσον αφορά τα τελομερή, στο μικρό βραχίονα, ο ακραίος κλώνος απέχει 20kb από τις τελομερικές επαναλήψεις ενώ στο μεγάλο βραχίονα ο ακραίος κλώνος περιέχει μέρος αυτών των επαναλήψεων. Η τελική αλληλουχία είναι 131.666.441bp, καλύπτει το 99,4% της ευχρωματίνης και η πιστότητά της ξεπερνά το 99,99%. Περιλαμβάνει όλα τα γνωστά mRNAs και STSs του χρωμοσώματος και διακόπτεται από 17 μόνο χάσματα (Εικ. 63). Το μεγαλύτερο contig αλληλουχίας έχει μήκος 44,7Mb και καλύπτει την περιοχή 10q23.1-q26.13 (Εικ. 21, Εικ. 63). Η αλληλούχηση ποσοστού ~60% αυτού του contig, στηρίχτηκε στον φυσικό χάρτη που περιγράψαμε.

➤ Σύγκριση γενετικού-φυσικού χάρτη

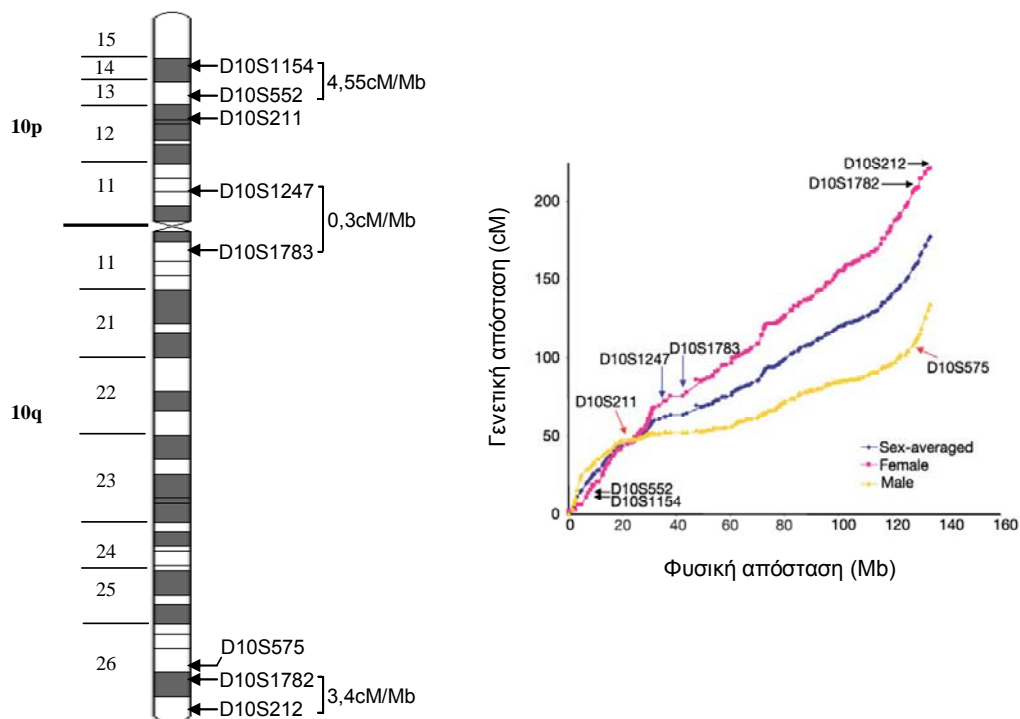
Ο ανασυνδυασμός είναι μία θεμελιώδης διαδικασία της εξέλιξης που σε συνδυασμό με τις μεταλλαγές ευθύνεται για τα πρότυπα της γενετικής ποικιλομορφίας που παρατηρούνται στους πληθυσμούς. Το ποσοστό των ανασυνδυασμών υπολογίζεται μέσω σύγκρισης του γενετικού με το φυσικό χάρτη, που πλέον αντιπροσωπεύεται από την γενωμική αλληλουχία. Το ποσοστό αυτό στο χρωμόσωμα 10, κατά μέσο όρο, είναι 1,32cM/Mb. Ωστόσο και σε συμφωνία με το

Chromosome 10



Εικ. 63: Ο χάρτης της αλληλουχίας του χρωσώματος 10, η συνταινία με τον ποντικό, τον αρουραίο και τους ιχθύες και τα γονίδια που κωδικοποιούν πρωτεΐνες. Απεικονίζονται από πάνω προς τα κάτω: 1) οι συντεταγμένες σε Mb, 2) ο χάρτης της αλληλουχίας του χρωσώματος (μαύρη γραμμή) και τα χάσματα του φυσικού χάρτη και της αλληλουχίας (γκρι), 3) οι συνταινικές περιοχές στον ποντικό και τον αρουραίο. Κάθε χρωμάσωμα συμβολίζεται με διαφορετικό χρώμα. Επίσης αναγράφεται η αρίθμηση των χρωσωμάτων και οι συντεταγμένες των συνταινικών περιοχών σε Mb, 4) τα προβλεπόμενα νησίδα CrG (καφέ), 5) οι περιοχές ομολογίας στο *Fugu* (μπλε), το *zebrafish* (σκούρο μπλε) και το *Tetraodon* (ροζ) και 6) τα γονίδια που κωδικοποιούν πρωτεΐνες. Τα γνωστά γονίδια συμβολίζονται με σκούρο μπλε και τα «νέα» με μαύρο. Ανατύπωση από Deloukas P *et al*, 2004

υπόλοιπο γονιδίωμα (Yu A *et al*, 2001), παρουσιάζει σημαντικές διακυμάνσεις κατά μήκος του. Έτσι, το ποσοστό ανασυνδυασμών στην περικεντρομερική περιοχή εμφανίζεται μειωμένο μέχρι και δύο φορές λιγότερο σε σχέση με άλλες περιοχές. Ειδικότερα, μεταξύ των δεικτών D10S1247-D10S1783 (Εικ. 64), που χαρτογραφούνται εκατέρωθεν του κεντρομερούς στις ζώνες 10p11.22 και 10q11.21 αντίστοιχα, η αναλογία γενετικού / φυσικού μήκους είναι 0,3cM/Mb. Χαμηλό ποσοστό ανασυνδυασμών (0,72cM/Mb) παρατηρείται γενικά σε όλη την χρωμοσωμική περιοχή 10q11. Αντίθετα, άλλες περιοχές του χρωμοσώματος χαρακτηρίζονται από ιδιαίτερα αυξημένο ποσοστό ανασυνδυασμών. Το υψηλότερο ποσοστό παρατηρείται στις υποτελομερικές περιοχές και ειδικότερα μεταξύ των δεικτών D10S1782-D10S212 (3,4cM/Mb) και D10S1154-D10S552 (4,55cM/Mb) στις ζώνες q26.2 –q26.3 και p15.3-p15.1 αντίστοιχα (Εικ. 64). Το φαινόμενο αυτό παρατηρείται σε όλα τα χρωμοσώματα (Yu A *et al*, 2001; Kong A *et al*, 2002; Jensen-Seaman MI *et al*, 2004) και είχε επισημανθεί και πριν την αλληλούχηση του ανθρώπινου γονιδιώματος (Payseur BA & Nachman MW, 2000).

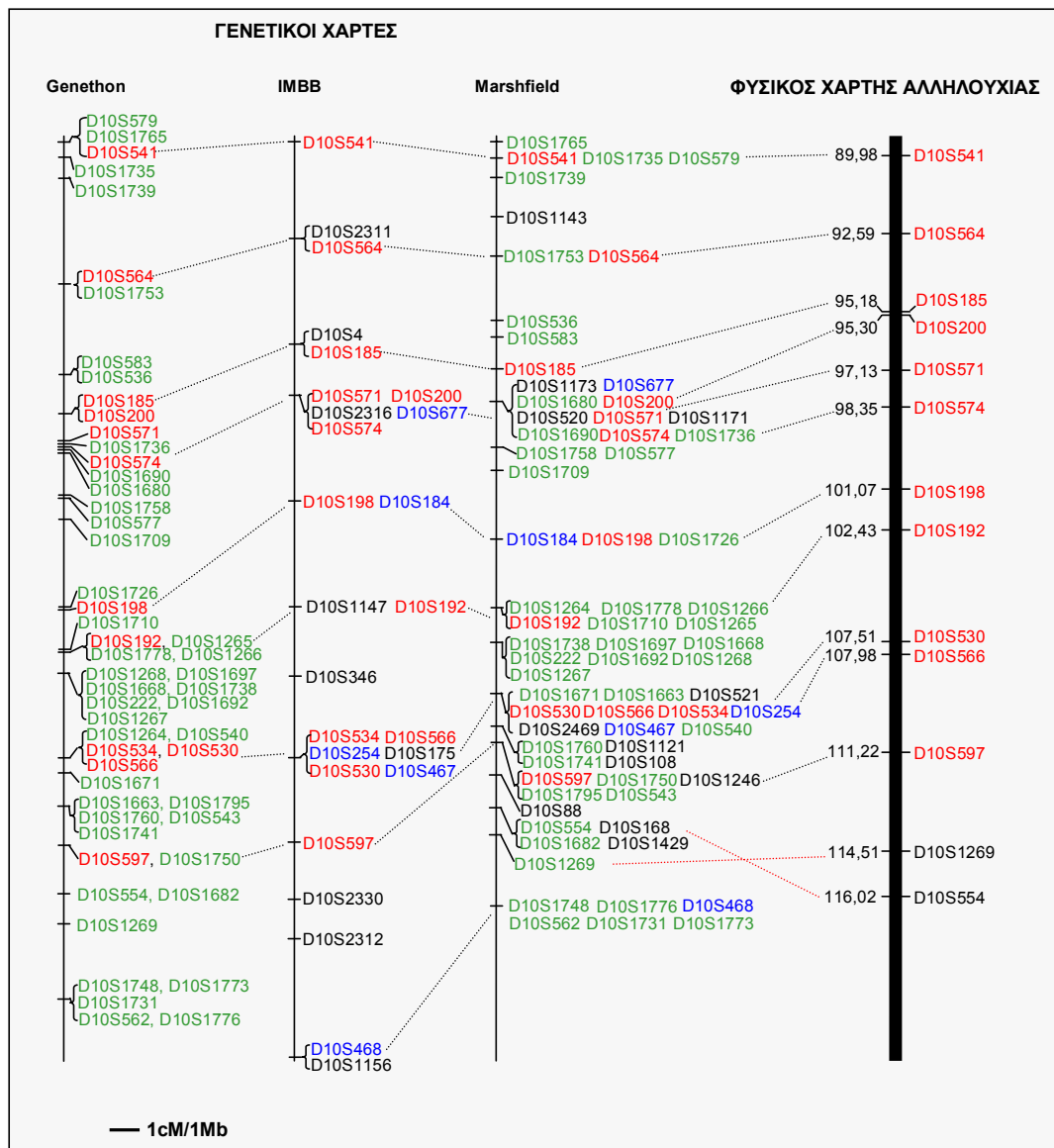


Εικ. 64: Σύγκριση γενετικού-φυσικού χάρτη κατά μήκος του χρωμοσώματος 10. Στο διάγραμμα απεικονίζεται η στοιχισή του γενετικού χάρτη (deCODE) των αρσενικών ατόμων (κίτρινο), των θηλυκών ατόμων (κόκκινο) και του συνδυαστικού χάρτη των δύο φύλων με το φυσικό χάρτη κατά μήκος του χρωμοσώματος. Σημειώνονται οι γενετικοί δείκτες που καθορίζουν τα όρια των περιοχών όπου παρατηρείται ιδιαίτερα χαμηλό (D10S1247-D10S1783) ή υψηλό ποσοστό ανασυνδυασμού (D10S1154-D10S552, D10S1782-D10S212). Επίσης απεικονίζεται η περιοχή μεταξύ των γενετικών δεικτών D10S211 και D10S575, κατά μήκος της οποίας το αντίστοιχο ποσοστό είναι αυξημένο στο χάρτη των θηλυκών ατόμων. Η θέση όλων των παραπάνω γενετικών δεικτών φαίνεται στο ιδεόγραμμα του χρωμοσώματος (αριστερά).

Διερευνώντας την αναλογία cM /Mb στην περιοχή 10q23.31-q25.3 (Εικ. 65), με βάση τους γενετικούς χάρτες των δύο φύλων Genethon, IMBB και Marshfield (Dib C *et al*, 1996; Marzella R *et al*, 1997; Broman KW *et al*, 1998) παρατηρούμε ότι το ποσοστό ανασυνδυασμών παρουσιάζει γενικά ομοιόμορφη κατανομή καθ' όλο το μήκος των ~25Mb και η αναλογία γενετικής/φυσικής απόστασης είναι, κατά μέσο όρο για τους τρεις γενετικούς χάρτες, ~1cM/Mb. Επίσης, η διατάξη των κοινών γενετικών δεικτών οι οποίοι αποτελούσαν σημεία αναφοράς κατά την κατασκευή του φυσικού χάρτη, είναι συμβατή με τη διάταξή τους στην αλληλουχία του χρωμοσώματος (Εικ. 65). Ωστόσο, σε κάποιες περιπτώσεις η σειρά μεμονωμένων γενετικών δεικτών είναι ανεστραμμένη, πχ. η διάταξη του D10S1269 σε σχέση με τον D10S554 στον χάρτη Marshfield (Εικ. 65). Το φαινόμενο αυτό οφείλεται στο ότι η κατασκευή των γενετικών χαρτών Genethon και Marshfield στηρίχτηκε στη γονοτυπική ανάλυση 8 οικογενειών αναφοράς CEPH, αριθμός που σε μερικές περιπτώσεις δεν είναι αρκετός. Τέλος, διαφορές παρατηρούνται και μεταξύ των γενετικών χαρτών των δύο φύλων. Στα θηλυκά άτομα το ποσοστό ανασυνδυασμού είναι γενικά υψηλότερο καθ' όλο το μήκος του χρωμοσώματος (D10S211-D10S575, Εικ. 64) πλην των τελομερών, όπου συμβαίνει το αντίστροφο.

➤ **Επαναλαμβανόμενες αλληλουχίες**

Οι επαναλαμβανόμενες αλληλουχίες αποτελούν τουλάχιστον το 50% του ανθρώπινου γονιδιώματος (Lander ES *et al*, 2001). Σε αυτές ανήκουν τα μεταθετά στοιχεία, τα ψευδογονίδια, οι απλές επαναλαμβανόμενες αλληλουχίες SSRs [simple sequence repeats π.χ. (A)_n, (CA)_n, (CGG)_n], οι διπλασιασμοί, οι επαναλήψεις των κεντρομερών και των τελομερών, οι μικροί βραχίονες των ακροκεντρικών χρωμοσωμάτων και οι ομάδες των ριβοσωμικών γονιδίων. Τα μεταθετά στοιχεία αντιπροσωπεύουν το μεγαλύτερο ποσοστό (45%). Στο χρωμόσωμα 10 οι επαναλαμβανόμενες αλληλουχίες αντιστοιχούν στο 45% του μήκους και τα μεταθετά στοιχεία αποτελούν επίσης την μεγαλύτερη ομάδα. Αυτά διακρίνονται σε αυτά που μετατίθενται μέσω σχηματισμού ενδιάμεσου RNA (RNA mediated) που είναι τα LINES, SINES και τα LTRs και σε αυτά που μετατίθενται κατευθείαν σαν DNA (DNA mediated) τα οποία αντιστοιχούν στο 2,84% του χρωμοσώματος (3% σε επίπεδο γονιδιώματος). Το μήκος τους κυμαίνεται μεταξύ 0,1kb (SINES) και 11kb (LTRs). Από τις παραπάνω κατηγορίες ενεργά είναι τα LINE1, τα Alu (SINES) και τα ERV (endogenous retroviruses, LTR).



Εικ. 65: Σύγκριση γενετικού-φυσικού χάρτη στη χρωμοσωμική περιοχή 10q23.31-q25.3. Απεικονίζονται τα τμήματα των γενετικών χαρτών Genethon, IMBB και Marsfield που αντιστοιχούν στην χρωμοσωμική περιοχή 10q23.31-q25.3. Για τους κοινούς δείκτες (κόκκινοι) προσδιορίστηκε η διάταξή τους στην αλληλουχία του χρωμοσώματος (δεξιά), (αναγράφονται οι συντεταγμένες τους σε Mb). Η σχετικά υψηλή πιστότητα και των τριών γενετικών χαρτών προκύπτει από το γεγονός ότι η διάταξη των κοινών δεικτών είναι συμβατή μεταξύ γενετικού-φυσικού χάρτη (διακεκομμένες γραμμές). Σημειώνεται (κόκκινες διακεκομμένες γραμμές) η αναστροφή μεταξύ των δεικτών D10S1269 και D10S554 στον χάρτη Marshfield σε σχέση με τον χάρτη αλληλουχίας.

Ιδιαίτερο ενδιαφέρον παρουσιάζει η συσσώρευση των Alu (ενεργά SINES) σε πλούσιες σε γονίδια περιοχές του χρωμοσώματος 10. Το φαινόμενο αυτό ισχύει για όλο το γονιδίωμα και η τρέχουσα υπόθεση για την παρατηρούμενη υψηλή πυκνότητα υποστηρίζει ότι δρουν προς όφελος του οργανισμού, είναι δηλαδή συμβιωτικά στοιχεία. Η θεωρία αυτή στηρίχτηκε στην παρατήρηση ότι σε κάποια είδη, τα SINES, σε κατάσταση στρες, ενεργοποιούν την μετάφραση. Αυτό επιτυγχάνεται με την μεταγραφή τους υπό τέτοιες συνθήκες και με την ακόλουθη σύνδεση των RNAs τους

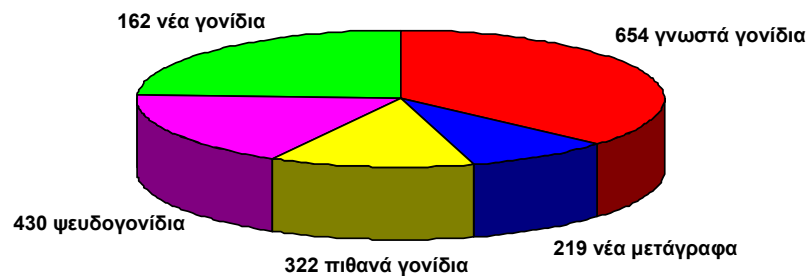
με την κινάση PKR, εμποδίζοντάς την έτσι να αναστείλει την μετάφραση (Schmid CW, 1998; Chu WM *et al*, 1998; Li T *et al*, 1999; Liu WM *et al*, 1995). Ο ρόλος της ρύθμισης της μετάφρασης ταιριάζει στα SINES καθώς μπορούν να μεταγράφονται γρήγορα σε πολύ μεγάλες ποσότητες από χιλιάδες αντίγραφα τους και να λειτουργούν σε επίπεδο RNA. Βάσει αυτού μπορεί να εξηγηθεί η θετική επιλογή τους σε περιοχές πλούσιες σε γονίδια.

Ως διπλασιασμοί χαρακτηρίζονται οι περιοχές που δεν είναι αντίγραφα μεταθετών στοιχείων, έχουν μήκος >1kb και εμφανίζουν νουκλεοτιδική ταυτότητα >90%, που σημαίνει ότι ο διπλασιασμός συνέβη τα προηγούμενα 40 εκατομμύρια χρόνια. Οι διπλασιασμοί διακρίνονται σε δια-χρωμοσωμικούς, κατά τους οποίους το αντίστοιχο τμήμα μπορεί να μεταφέρθηκε σε περισσότερες από μία θέσεις του γονιδιώματος, όχι απαραίτητα γειτονικές και σε ενδο-χρωμοσωμικούς. Οι διπλασιασμοί παρουσιάζουν ιδιαίτερο ενδιαφέρον από ιατρική άποψη καθώς συχνά σχετίζονται με ελλείμματα και χρωμοσωμικές αναδιατάξεις που οδηγούν σε φαινοτυπικές ανωμαλίες, όπως π.χ. το σύνδρομο Williams (7q), DiGeorge (22q), Charcot–Marie–Tooth (17p). Επιπλέον, η γνώση της φύσης και της έκτασής τους είναι σημαντική για την κατανόηση της εξέλιξης του ανθρώπινου γονιδιώματος. Η ανάλυσή τους, που κατέστη δυνατή με βάση την ολοκληρωμένη αλληλουχία, έδειξε εκτεταμένους ενδο-χρωμοσωμικούς διπλασιασμούς ενός τμήματος ~5Mb από το 10q11 στις περιοχές 10q22, 10q23.1 και 10q23.2 που εκτιμώνται ότι πραγματοποιήθηκαν πριν 30-40 εκ. χρόνια (Crosier M *et al*, 2002; Deloukas P *et al*, 2004). Επίσης από διπλασιασμό των περιοχών 10q11 και 10q23 και ταυτόχρονη παρακεντρομερική αναστροφή τους πιστεύεται ότι έχει προκύψει ολόκληρη η περιοχή 10q22. Σε συμφωνία με όλες τις παραπάνω χρωμοσωμικές αναδιατάξεις είναι τα αθροίσματα των παραλόγων γονιδίων που ταυτοποιήθηκαν στις αντίστοιχες περιοχές. Εκτός των 3 παραλόγων GLUD (Deloukas P *et al*, 1993) και των 3 CTSL (Bryce SD *et al*, 1994) που είχαν χαρακτηριστεί παλαιότερα, ταυτοποιήθηκαν επιπλέον 3 και 4 παράλογα ψευδογονίδια αντίστοιχα. Επίσης εντοπίστηκαν 7 παράλογα γονίδια της οικογένειας BMS1L, 11 της CTGLF1, 7 της FAM25A και 6 της ANXA. Οι παραπάνω αναδιατάξεις φαίνεται πως είχαν αντίκτυπο στην καταγεγραμμένη γονιδιακή πυκνότητα του χρωμοσώματος, καθώς το 31% των λειτουργικών γονιδίων του συγκεντρώνεται στις περιοχές 10q11 και 10q23 (Εικ. 67).

➤ **Αριθμός γονιδίων**

Η ταυτοποίηση των γονιδίων του χρωμοσώματος πραγματοποιήθηκε *in silico* χρησιμοποιώντας δύο κύριες ομάδες υπολογιστικών προγραμμάτων. Η πρώτη ομάδα ανιχνεύει γονίδια βάσει χαρακτηριστικών που προσδιορίζουν την αρχή και το

τέλος τους (προγράμματα «*ab initio*») ενώ η δεύτερη βασίζεται σε συγκρίσεις με γνωστά γονίδια και πρωτεΐνες (συγκριτική ανίχνευση γονιδίων). Από την επεξεργασία των αυτοματοποιημένων στοιχίσεων (εξέταση της γενωμικής οργάνωσης των γονιδίων, των αλληλουχιών των ESTs και cDNAs, κ.λπ.) ταυτοποιήθηκαν 1787 γονίδια (http://vega.sanger.ac.uk/Homo_sapiens/, Εικ. 63). Αυτά ταξινομήθηκαν βάσει των κριτηρίων που χρησιμοποιούνται σε όλα τα χρωμοσώματα που έχουν αναλυθεί και από τις αντίστοιχες βάσεις δεδομένων, ως εξής (Εικ. 66): α) 816 γονίδια που κωδικοποιούν πρωτεΐνες. Από αυτά, τα 654 χαρακτηρίζονται ως «γνωστά», αντιστοιχούν δηλαδή σε χαρακτηρισμένα cDNAs ή πρωτεΐνες και τα 162 ως «νέα», έχουν δηλαδή ανοικτό πλαίσιο ανάγνωσης και στοιχίζονται πλήρως με ESTs του ανθρώπου τα οποία αποτελούνται από 2 ή περισσότερα εξόνια και / ή έχουν ομολογία με γνωστά γονίδια ή πρωτεΐνες από άλλα είδη, β) 219 «νέα μετάγραφα», τα οποία έχουν τα χαρακτηριστικά των «νέων γονιδίων» αλλά δεν μπορεί να καθοριστεί γι' αυτά ένα συγκεκριμένο ανοικτό πλαίσιο ανάγνωσης, γ) 322 «πιθανά γονίδια», που είναι αλληλουχίες ταυτόσημες με ESTs του ανθρώπου τα οποία συναρμολογούνται από 2 ή περισσότερα εξόνια αλλά στερούνται ανοικτού πλαισίου ανάγνωσης και δ) 430 ψευδογονίδια, που είναι αλληλουχίες ομόλογες με γνωστά γονίδια και πρωτεΐνες αλλά δεν έχουν ανοικτό πλαίσιο ανάγνωσης. Τα τελευταία διακρίνονται περαιτέρω σε επεξεργασμένα (371) και μη (59).



Εικ. 66: Ταξινόμηση των γονιδίων του χρωμοσώματος 10 βάσει της ολοκληρωμένης αλληλουχίας του (Deloukas P *et al*, 2004).

Στην παραπάνω κατάταξη και εξαιρουμένων των ψευδογονιδίων, είναι φανερό ότι υπάρχει μία ασάφεια όσον αφορά τον συνολικό αριθμό των γονιδίων του χρωμοσώματος, καθώς ποσοστό μεγαλύτερο του 20% των γονιδίων έχουν την ένδειξη «πιθανά». Η ίδια ασάφεια ισχύει και στο υπόλοιπο γονιδίωμα. Έτσι, η εκτίμηση του συνολικού αριθμού των γονιδίων του ανθρώπου που κωδικοποιούν πρωτεΐνες εμφανίζει ένα μεγάλο εύρος διακύμανσης 5000 γονιδίων. Ωστόσο, παρόλο που δεν είναι γνωστός ο ακριβής συνολικός αριθμός των γονιδίων του ανθρώπου, η

ταυτοποίησή τους έχει βελτιστοποιηθεί κατά πολύ και σε αυτό, εκτός από την ολοκληρωμένη αλληλουχία, συνεισέφεραν τα προγράμματα απομόνωσης cDNAs πλήρους μήκους, η γενωμική αλληλουχία άλλων οργανισμών-μοντέλων και η βελτιστοποίηση του λογισμικού επεξεργασίας τους. Η τρέχουσα εκτίμηση (International Human Genome Sequencing Consortium, 2004) είναι ότι ο αριθμός αυτός κυμαίνεται μεταξύ 20.000 και 25.000, ενώ για τα 19.599 έχει επιβεβαιωθεί η ύπαρξή τους. Ο αριθμός των γονιδίων κατέπληξε πολλούς, καθώς θεωρήθηκε πολύ μικρός σε σύγκριση τόσο με άλλους οργανισμούς (π.χ. με τα ~19.000 γονίδια του *C. elegans*) όσο και με τις προηγούμενες εκτιμήσεις, οι οποίες ποίκιλλαν κατά πολύ και βασίστηκαν σε διάφορα κριτήρια. Οι αρχικές εκτιμήσεις στηρίχτηκαν σε μελέτες κινητικής επαναδιάταξης και προσδιόριζαν τον αριθμό των γονιδίων στις 40.000 (Lewin B, 1990). Βάσει του αριθμού των νησιδίων CpG και το ποσοστό αυτών που σχετίζονται με γονίδια ο συνολικός αριθμός των γονιδίων είχε υπολογιστεί στα ~80.000 (Antequera F & Bird A, 1993). Ακολούθησαν εκτιμήσεις βάσει των ομαδοποιήσεων των ESTs που κυμαίνονταν από 35.000 (Ewing B & Green P, 2000), ή 60.000-70.000 (Fields C *et al*, 1994) έως και 120.000 (Liang F *et al*, 2000). Άλλες προσεγγίσεις που περιλάμβαναν συγκριτικές μελέτες ανθρώπου-*Tetraodon nigroviridis* προσδιόρισαν τον αριθμό στις 28,000-34,000 (Roest Crollius H *et al* 2000) ενώ οι κατ' αναλογία εκτιμήσεις βάσει του αντίστοιχου αριθμού των γονιδίων των χρωμοσωμάτων 20, 21 και 22 ήταν 31,500 (Deloukas P *et al*, 2001), 44.000 (Hattori M *et al*, 2000) και 45.000 (Dunham I *et al*, 1999), αντίστοιχα. Ακόμα και η εκτίμηση βάσει της πρώτης έκδοσης της αλληλουχίας του γονιδιώματος είχε μεγάλη απόκλιση από τη σημερινή. Η ανάλυση του 2001 θεωρούσε ότι υπάρχουν 30.000-40.000 γονίδια που κωδικοποιούν πρωτεΐνες (Lander ES *et al*, 2001; Venter JC *et al*, 2001). Η διαφορά μεταξύ της αρχικής και της τωρινής πρόβλεψης οφείλεται κυρίως στο φαινόμενο της τεχνητής κατάτμησης των γονιδίων λόγω ατελειών των υπολογιστικών προγραμμάτων πρόβλεψης, όπου, στην παλαιότερη έκδοση, ένα γονίδιο εμφανίζονταν ως δύο.

➤ **Δομικά χαρακτηριστικά των γονιδίων και κατανομή τους**

Εξαιρουμένων των ψευδογονιδίων, τα υπόλοιπα 1357 γονίδια του χρωμοσώματος 10, καλύπτουν συνολικά ~66Mb, που συνεπάγεται ότι μεταγράφεται το ~50% της αλληλουχίας του χρωμοσώματος. Είναι ενδιαφέρον ότι το ποσοστό αυτό είναι μεγαλύτερο από όλα σχεδόν τα χρωμοσώματα που έχουν αναλυθεί μέχρι τώρα (Πιν. 13). Ωστόσο, η γονιδιακή πυκνότητά του το κατατάσσει στα μεσαίας γονιδιακής πυκνότητας χρωμοσώματα (~10 γονίδια/Mb) (Πιν. 13). Αυτό εξηγείται από το γεγονός ότι το μέσο μέγεθος των γονιδίων του (~51Kb) είναι κατά πολύ μεγαλύτερο σε σχέση

με το αντίστοιχο μέσο μέγεθος για το υπόλοιπο γονιδίωμα (27Kb, Lander *et al*, 2001). Ωστόσο, το μέγεθος μεμονωμένων γονιδίων παρουσιάζει σημαντικές διακυμάνσεις. Τα δύο ακραία παραδείγματα είναι τα γονίδια *CTNNA3* και *CALML5* που έχουν μήκος 1,7Mb και 0,85kb αντίστοιχα. Το μεγάλο μέγεθος των γονιδίων οφείλεται κυρίως στα μεγάλα ιντρόνια, καθώς τα εξόνια αντιπροσωπεύουν μόνο το 2.3% (Πιν.13) και έχουν μέσο μήκος 313bp, παρόλο που ταυτοποιήθηκαν εξόνια μεγέθους ~9,7kb (γονίδιο *SH3MD1*) και 3bp (*CDH23*).

Πιν. 13: Χαρακτηριστικά των χρωμοσωμάτων

	Ποσοστό κάλυψης με γονίδια (%)	Γονιδιακή πυκνότητα (γονίδια/Mb)	Ποσοστό κάλυψης με εξόνια (%)	Περιεχόμενο σε GC (%)	Περιεχόμενο σε επανάληψεις (%)	Νησίδια CpG (ανά Mb)	Μήκος (Mb)
1		12,8*					245*
2		6,1**		40,2		7	243**
3		5,8**					199,5**
4		4,5**		38,2		5,4	191,4**
5		5,4**		39,5	46,3		177
6	42,2	9,2	2,2	40	43,9	13,1	170,9
7	46,5	7,5	1,4	41	45	9,5	158,6**
8		5,4**					146,3**
9	47	10,5	2,5	41,4	46,1		140,3
10	50,6	10	2,3	41,6	43,7		135
11		10,6**					134,4**
12		8,3**					132,4**
13	37	6,5	1,3	38,5	42,3	5,4	114
14	43,6	9,7	2,3	40,9	46,2	20,2	106,3
15		7,6**					100,3**
16		10,8**		44,7	47,8		88,2*
17		15,9**					78,8**
18		4,8*					76*
19	50	26	6,4	48	55	28,9	63,8
20	42,4	12,2	2,4	44,1	42	11,1	62,4
21	31	6,7		40,9	40,1		46,9**
22	51	16,3	5	47,8	41,9	16,5	49,5
X	33	7,1	1,7	39	56	5,2	154,9
Y		1,8*				2,9"	57,7
γονιδίωμα		6,5-8 [^]		41"	>50"	10-16 [^]	3090

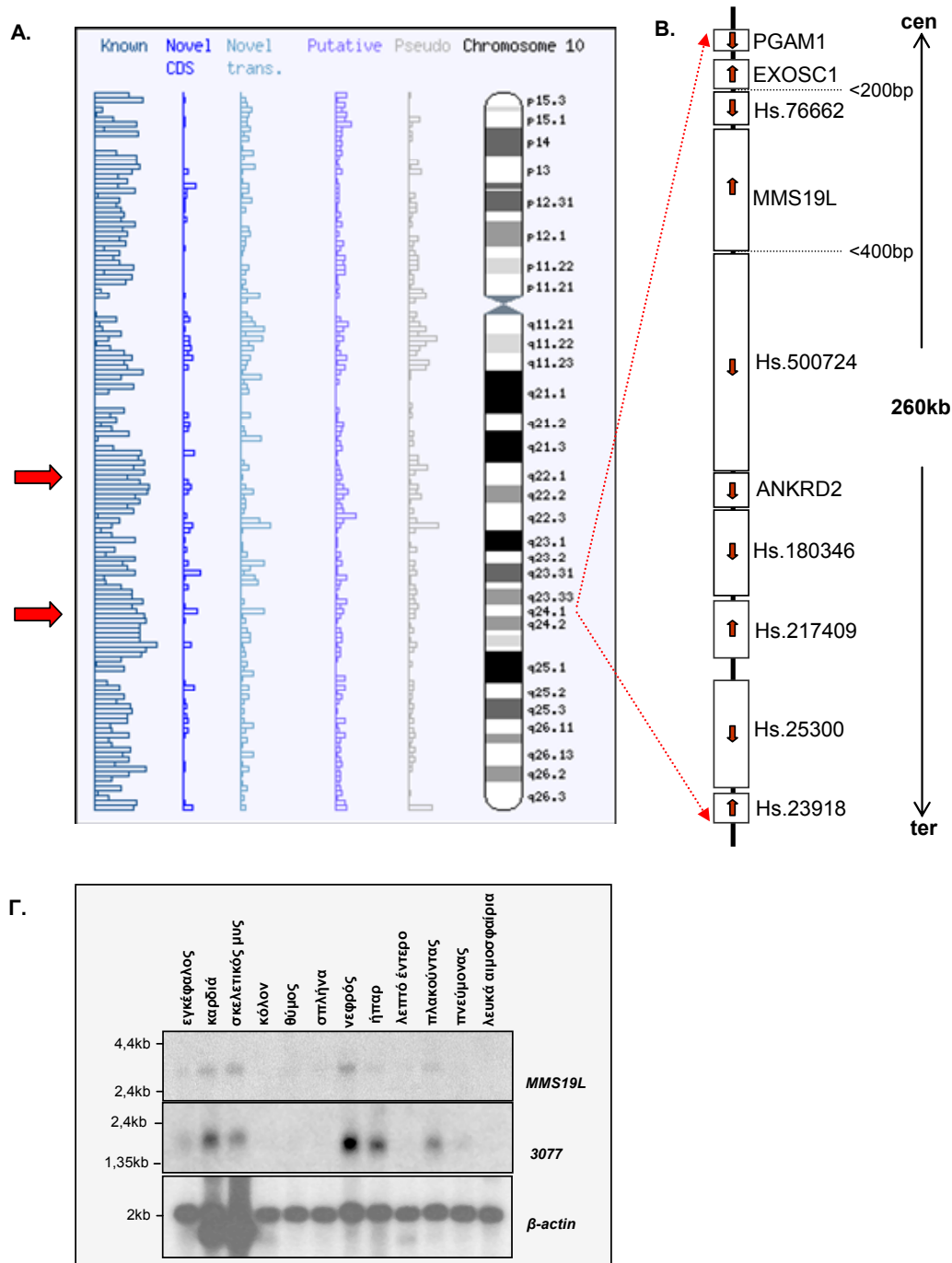
Σημ. : Τα στοιχεία προέρχονται από την ανάλυση της αρχικής (") και της ολοκληρωμένης (^) έκδοσης του γονιδιώματος, τις αντίστοιχες αναλύσεις για τα χρωμοσώματα 2, 4, 5, 6, 7, 9, 10, 13, 14, 16, 19, 20, 21, 22, X, Y και τις βάσεις δεδομένων Vega (*) και Ensembl (**). Για όσα στοιχεία δεν καταγράφονται στον πίνακα δεν υπάρχουν ακόμη διαθέσιμες οι αντίστοιχες πληροφορίες.

Η κατανομή των γονιδίων κατά μήκος του χρωμοσώματος είναι ανομοιογενής (Εικ. 63), όπως άλλωστε παρατηρείται και στα υπόλοιπα χρωμοσώματα. Ο μεγάλος βραχίονας εμφανίζεται σημαντικά πλουσιότερος σε γονίδια, παρουσιάζοντας μεγαλύτερη πυκνότητα σε δύο κυρίως περιοχές, την q22 και την q23.3-q25.1 (Εικ. 67, Εικ. 63). Ενδεικτικά, στην Εικ. 67 απεικονίζεται ένα τμήμα της χρωμοσωμικής περιοχής 10q24.1, το οποίο εμφανίζει ιδιαίτερα υψηλή γονιδιακή πυκνότητα. Ειδικότερα, σε μία περιοχή μήκους 260kb, η οποία είναι συνταινική στον ποντικό, χαρτογραφούνται 10 γονίδια. Αυτά τα γονίδια δεν φαίνονται να σχετίζονται λειτουργικά, υπό την έννοια ότι δεν συμμετέχουν σε κοινό βιολογικό μονοπάτι π.χ. σε αυτά συμπεριλαμβάνεται το ένζυμο φωσφογλυκερική μουτάση (PGAM1), ένα

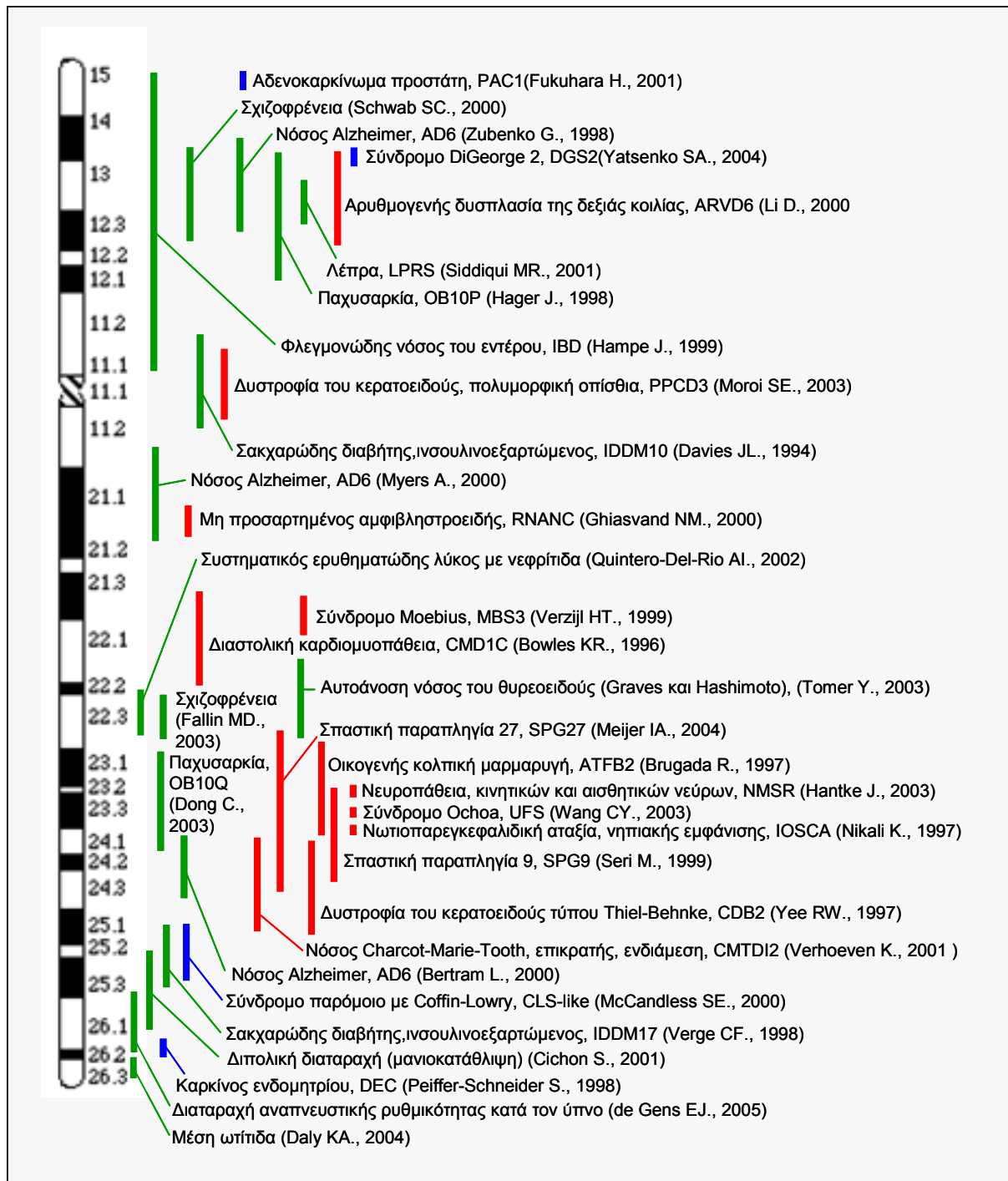
συστατικό του εξωσώματος (EXOSC1), μία πρωτεΐνη που συμμετέχει στην επιδιόρθωση του DNA (MMS19L) κ.λπ. Ωστόσο, οι ιδιαίτερα μικρές διαγονιδιακές αποστάσεις τους πιθανώς υποδηλώνουν συσχέτιση σε επίπεδο κοινής ρύθμισης. Σε αυτό συνηγορεί και το ταυτόσημο πρότυπο των επιπέδων έκφρασης στους διάφορους ιστούς των γειτονικών γονιδίων Hs.76662 και MMS19L, όπως δείξαμε πειραματικά με ανάλυση Northern (Εικ. 67). Η υπόθεση αυτή συμφωνεί με πρόσφατες μελέτες που δείχνουν ότι η διάταξη των γονιδίων στους ευκαρυωτικούς οργανισμούς δεν είναι τυχαία, αλλά τα γονίδια που έχουν παρόμοια πρότυπα έκφρασης τείνουν να ομαδοποιούνται (Hurst LD *et al*, 2004). Αποτέλεσμα της αυξημένης γονιδιακής πυκνότητας είναι και η συσσώρευση των γενετικών τόπων που σχετίζονται με νοσήματα στις εν λόγω περιοχές. Έτσι, από τα ~70 γονίδια του χρωμοσώματος που έχουν συσχετιστεί με την εμφάνιση γενετικών διαταραχών, ποσοστό ~40% αυτών χαρτογραφείται στην περιοχή q23-q24 (Εικ. 7). Επιπλέον, στην ίδια περιοχή χαρτογραφείται και το ~30% των κρίσιμων περιοχών, νοσημάτων για τα οποία το υπεύθυνο γονίδιο δεν έχει ακόμη ταυτοποιηθεί (Εικ. 68).

➤ **Περιεχόμενο σε GC - Νησίδια CpG**

Σε ακολουθία με τη γονιδιακή κατανομή είναι και το περιεχόμενο του χρωμοσώματος σε νουκλεοτίδια με GC, που κατά μέσο όρο αντιστοιχούν στο 41,6% των βάσεων, τιμή παρόμοια με το μέσο όρο του γονιδιώματος (41%). Ωστόσο και σε συμφωνία με παλαιότερες εκτιμήσεις, η κατανομή παρουσιάζει σημαντικές διακυμάνσεις. Έτσι, σε επίπεδο γονιδιώματος, το ποσοστό κυμαίνεται μεταξύ 36% και 47% όταν εξετάζονται μεγάλα τμήματα δεκάδων Mb, ενώ η διακύμανση είναι εντονότερη μεταξύ περιοχών μικρότερου μήκους, π.χ. σε γειτονικές περιοχές μήκους 300kb η τιμή μεταβάλλεται από 33% σε 59%. Επίσης, με βάση την αλληλουχία του γονιδιώματος, επιβεβαιώθηκε η στενή συσχέτιση του περιεχομένου σε GC με τις ζώνες G και R των χρωμοσωμάτων: η πλειονότητα των κλώνων που χαρτογραφούνται κυτταρογενετικά στις ζώνες G, έχουν χαμηλό ποσοστό GC (37% κατά μέσο όρο) και αντιστρόφως. Επιπλέον, επιβεβαιώθηκε η συσχέτιση των νησιδίων CpG με τα γονίδια. Τα νησίδια αυτά είναι μη μεθυλιωμένες περιοχές του γονιδιώματος με αναμενόμενη συχνότητα δινουκλεοτιδίων CpG, που σχετίζονται με το 5' άκρο του ~60% των γονιδίων (Bird AP, 1986; Larsen F *et al*, 1992; Antequera F & Bird A, 1993). Στις περιοχές εκτός νησιδίων, το δινουκλεοτίδιο αυτό υποαντιπροσωπείται (κατά 1/5) λόγω μεθυλίωσης στην C και αυθόρμητης απαμίνωσης που την μετατρέπει σε T. Αντίθετα, η αυθόρμητη απαμίνωση μη μεθυλιωμένων C οδηγεί σε U, οι οποίες στη συνέχεια αναγνωρίζονται και επιδιορθώνονται από μηχανισμούς του κυττάρου.

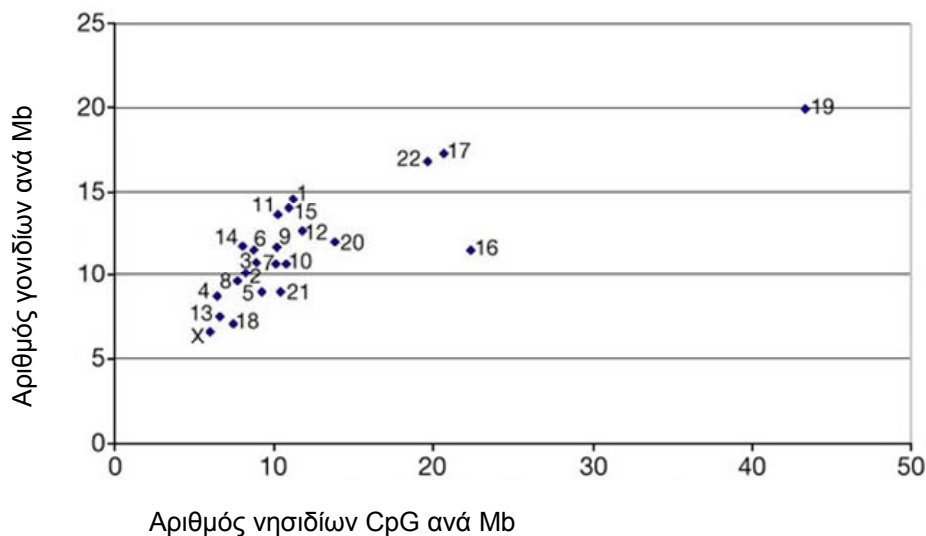


Εικ. 67: Κατανομή των γονιδίων κατά μήκος του χρωμοσώματος. Α. Απεικονίζεται η κατανομή για τα γνωστά (Known) και τα νέα γονίδια (Novel CDS), τα νέα μετάγραφα (Novel trans.), τα πιθανά γονίδια (Putative) και τα ψευδογονίδια (Pseudo). Οι περιοχές με τη μεγαλύτερη γονιδιακή πυκνότητα είναι οι q22 και η q23.3-q25.1 (βέλη). Τα στοιχεία προέρχονται από την βάση δεδομένων Vega (http://vega.sanger.ac.uk/Homo_sapiens/). **Β.** Ενδεικτικό τμήμα της περιοχής 10q24.1 που εμφανίζει ιδιαίτερα υψηλή γονιδιακή πυκνότητα. Η υποπεριοχή αυτή έχει μήκος 260kb και περιλαμβάνει 10 γονίδια. Οι αποστάσεις είναι σε κλίμακα και τα βέλη συμβολίζουν την φορά μεταγραφής των γονιδίων. Όσα γονίδια δεν ανήκουν στα «γνωστά», συμβολίζονται με τον κωδικό της βάσης δεδομένων Unigene. **Γ.** Ταυτόσημο πρότυπο των επιπέδων έκφρασης των γονιδίων Hs.76662 και MMS19L, βάση ανάλυσης Northern.



Εικ. 68: Νοσήματα που οι κρίσιμες περιοχές τους χαρτογραφούνται στο χρωμόσωμα 10 και για τα οποία δεν έχει ταυτοποιηθεί το υπεύθυνο γονίδιο. Απεικονίζονται οι κρίσιμες περιοχές (οριζόντιες γραμμές): 1) μονογονιδιακών νοσημάτων που εμφανίζουν μεντελικό τύπο κληρονόμησης (κόκκινο), 2) πολυπαραγοντικών νοσημάτων (πράσινο) και 3) νοσημάτων που οφείλονται σε χρωμοσωμικά ελλείμματα (μπλε). Τα στοιχεία προέρχονται από την ανασκόπηση για το χρωμόσωμα (Moschonas NK, 2003) και την βάση δεδομένων OMIM (<http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=OMIM>) (Ιούλιος 2005).

Στο χρωμόσωμα 10 ταυτοποιήθηκαν 1.025 πιθανά νησίδα και η κατανομή τους ακολουθεί την αντίστοιχη των γονιδίων (Εικ. 63). Αντίστοιχη κατανομή εμφανίζεται και σε επίπεδο γονιδιώματος. Έτσι, τα περισσότερα χρωμοσώματα έχουν 5-15 νησίδα/Mb, το χρωμόσωμα 19, που είναι το πλουσιότερο σε γονίδια, έχει και την μεγαλύτερη πυκνότητα νησιδίων ενώ το Y έχει την μικρότερη πυκνότητα νησιδίων (2,9/Mb) και γονιδίων (Εικ. 69). Συνολικά, με ανάλυση της πρώτης έκδοσης της αλληλουχίας του γονιδιώματος ο αριθμός των νησιδίων εκτιμήθηκε μεταξύ 29.000 και 50.000 (με ή χωρίς φιλτράρισμα των επαναλαμβανόμενων αλληλουχιών).



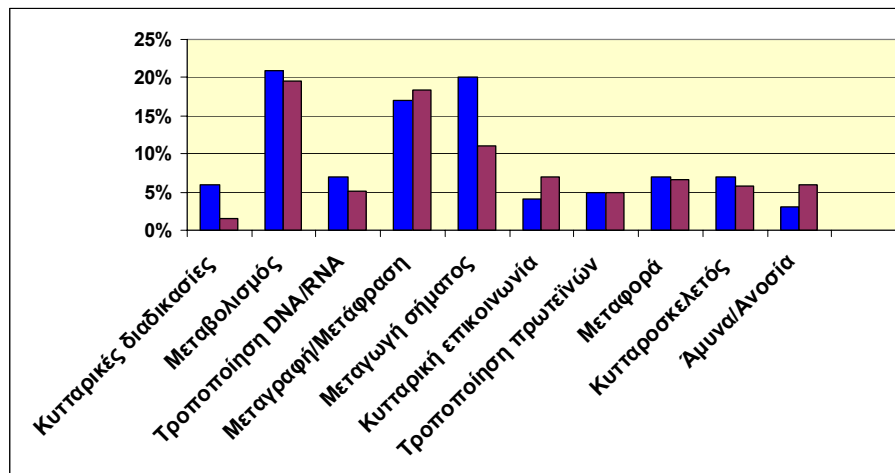
Εικ. 69: Συνάρτηση του αριθμού των νησιδίων CpG ανά Mb σε σχέση με τον αριθμό των γονιδίων ανά Mb, με βάση τον χάρτη GeneMap 98, για κάθε χρωμόσωμα (ανατύπωση από Lander ES *et al*, 2001).

➤ Χαρακτηρισμός γονιδίων *in silico*

Τα γονίδια που ταυτοποιήθηκαν στο χρωμόσωμα 10, χαρακτηρίστηκαν *in silico* για την παρουσία αναγνωρίσιμων μοτίβων, σύμφωνα με τη βάση δεδομένων InterPro (<http://www.ebi.ac.uk/interpro/index.html>). Από την ανάλυση προέκυψε ότι το πιο σύνηθες μοτίβο είναι ο δακτύλιος ψευδαργύρου (20 γονίδια) που είναι και το δεύτερο πιο συχνά απαντώμενο στο γονιδίωμα. Επιπλέον, φάνηκε ότι το χρωμόσωμα είναι εμπλουτισμένο, σε σχέση με το υπόλοιπο γονιδίωμα, σε γονίδια που κωδικοποιούν λιπάσες, αλδο/κετο-αναγωγάσες και υδρολάσες. Τα παραπάνω συχνότερα απαντώμενα μοτίβα κωδικοποιούνται από γονίδια που χαρτογραφούνται σε αθροίσματα. Χαρακτηριστικά παραδείγματα είναι αυτά που κωδικοποιούν πρωτεΐνες με δακτυλίους ψευδαργύρου στο 10p11.21 και στο 10q11.21 που αποτελούνται από 4 και 5 γονίδια αντίστοιχα και οι ομάδες έξι γονιδίων στο 10q23.31 και τεσσάρων γονιδίων στο 10q25.3, που κωδικοποιούν λιπάσες. Πιθανώς, η

συγκεκριμένη χωρική τοποθέτησή τους είναι αποτέλεσμα διπλασιασμών και εξυπηρετεί την κοινή ρύθμιση της έκφρασής τους.

Μία προκαταρκτική ανάλυση ταξινόμησης των γονιδίων του χρωμοσώματος, σε 10 ευρύτερες λειτουργικές κατηγορίες, βάσει βιβλιογραφικών αναφορών ή μοτίβων που περιέχουν, οδήγησε στην κατηγοριοποίηση ~300 γονιδίων (Moschonas NK, 2003, Εικ. 70). Η κατανομή τους ακολουθεί γενικά αυτήν του υπόλοιπου γονιδιώματος (Εικ. 70), σύμφωνα με την ανάλυση κατά την οποία ταξινομήθηκαν ~40% των πρωτεϊνών του ανθρώπου και τα ποσοστά τους συγκρίθηκαν με τα αντίστοιχα του *S.cerevisiae*, του *C.elegans*, της *D.melanogaster* και του *A. thaliana*. Η σύγκριση αυτή, έδειξε ότι ο άνθρωπος, σε σχέση με τα δύο ασπόνδυλα, έχει πολλές περισσότερες πρωτεΐνες που σχετίζονται με τον κυτταροσκελετό, την μεταγραφή και την μετάφραση καθώς και με λειτουργίες χαρακτηριστικές της φυσιολογίας των σπονδυλωτών, όπως η άμυνα και η ανοσία.



Εικ. 70: Ταξινόμηση του πρωτεϊνώματος του χρωμοσώματος 10 (κόκκινο) και σύγκριση με το συνολικό πρωτεϊνώμα (μπλε).

Ο μικρός αριθμός γονιδίων του ανθρώπου, σε σχέση με τον αναμενόμενο, έδειξε ότι η γενετική πολυπλοκότητα δεν καθορίζεται αποκλειστικά από αυτόν τον παράγοντα. Σημαντικές πηγές γενετικής πολυπλοκότητας του ανθρώπινου γονιδιώματος που αντισταθμίζουν τον σχετικά μικρό αριθμό γονιδίων που διαθέτει, είναι η εναλλακτική συναρμογή, η χρήση εναλλακτικών υποκινητών, οι πολυπληθείς μετα-μεταφραστικές τροποποιήσεις των πρωτεϊνών και πιθανώς η ρύθμιση της γονιδιακής έκφρασης μέσω αλληλοεπικαλυπτόμενων μεταγράφων, ψευδογονιδίων, micro- και siRNAs.

➤ **Εναλλακτική συναρμογή**

Περισσότερο μελετημένος από τους παραπάνω μηχανισμούς είναι η εναλλακτική συναρμογή που πλέον φαίνεται ότι αποτελεί τον κανόνα και όχι την

εξαίρεση. Στο χρωμόσωμα 10 ταυτοποιήθηκαν 4204 μετάγραφα που αντιστοιχούν στα 1357 γονίδια, χωρίς να συμπεριληφθούν αυτά που προκύπτουν από χρήση εναλλακτικών σημάτων πολυαδενυλίωσης. Ποσοστό 73% των γονιδίων με ανοικτό πλαίσιο ανάγνωσης παράγουν περισσότερα του ενός μετάγραφα με μέσο όρο 5,8 μετάγραφα/γονίδιο. Ένα ακραίο παράδειγμα είναι το γονίδιο *ADD3* για το οποίο εντοπίστηκαν 22 εναλλακτικά μετάγραφα. Επιπλέον, διαπιστώθηκε ότι η εναλλακτική συναρμογή οδηγεί κατά κανόνα στη σύνθεση διαφορετικών πρωτεϊνικών ισομορφών, καθώς από τα 342 γονίδια του χρωμοσώματος που παράγουν τουλάχιστον δύο μετάγραφα πλήρους μήκους, τα 312 (91%) συνθέτουν τουλάχιστον δύο διαφορετικά πολυπεπτίδια. Σε επίπεδο γονιδιώματος, διάφορες μελέτες εκτιμούν ότι τουλάχιστον το 40-60% των γονιδίων υπόκεινται σε εναλλακτική συναρμογή (Modrek B & Lee C, 2002). Το ποσοστό αυτό σε άλλες μελέτες, που συνδυάζουν δεδομένα ESTs και μικροσυστοιχιών για περισσότερα από 10.000 γονίδια, ανέρχεται σε 73% (Johnson JM *et al*, 2003). Παρόμοιες μελέτες για τα χρωμοσώματα 21 και 22 έδειξαν ότι η εναλλακτική συναρμογή συμβαίνει σε >80% των γονιδίων (Kampa D *et al*, 2004).

➤ **Αλληλοεπικαλυπτόμενα γονίδια**

Η διερεύνηση της αλληλουχίας του χρωμοσώματος 10 έδειξε την εκτεταμένη παρουσία αλληλοεπικαλυπτόμενων γονιδίων. Τέτοια γονίδια έχουν παρατηρηθεί από τους ιούς (Barrell BG *et al*, 1976) μέχρι τα θηλαστικά (Williams T & Fried M, 1986; Adelman JP *et al*, 1987; Kiyosawa H & Abe K, 2002; Ohinata Y *et al*, 2002; Zhou C & Blumberg B, 2003). Τα συμπληρωματικά μετάγραφα φαίνεται πως συμμετέχουν σε διάφορα επίπεδα της ρύθμισης της γονιδιακής έκφρασης και μέχρι τώρα έχουν διαπιστωθεί τρεις μηχανισμοί: η μεταγραφική παρεμβολή (transcriptional interference), η κάλυψη του RNA (RNA masking), και μηχανισμοί που δρουν σε δίκλινα μόρια RNA (Lavorghna G *et al*, 2004). Συνολικά εντοπίστηκαν 101 ζεύγη αλληλοεπικαλυπτόμενων γονιδίων και σε 38 περιπτώσεις το ένα από τα γονίδια περιέχεται σε ένα ιντρόνιο του άλλου ενώ ως επί το πλείστον μεταγράφεται με αντίθετη φορά. Από τα υπόλοιπα, ποσοστό 34% εμφανίζουν αλληλοεπικάλυψη στο 5' άκρο τους. Σε μία μελέτη όλου του γονιδιώματος, ταυτοποιήθηκαν 774 τέτοια ζεύγη γονιδίων (Veeramachaneni V *et al*, 2004). Ωστόσο, πιθανώς αυτή η τιμή είναι υποεκτίμηση καθώς δεν έχει ολοκληρωθεί ακόμη η ανάλυση όλων των χρωμοσωμάτων.

➤ **Ψευδογονίδια**

Σημαντική συμβολή στη μελέτη της εξελικτικής ιστορίας των γονιδιωμάτων προέρχεται από την μελέτη των ψευδογονιδίων. Η ανάλυσή τους έχει αρχίσει να

προσελκύει όλο και περισσότερο ενδιαφέρον, καθώς πρόσφατα δεδομένα δείχνουν ότι ορισμένα ψευδογονίδια παίζουν ρόλο στην ρύθμιση της γονιδιακής έκφρασης, επάγοντας την αποικοδόμηση των μεταγράφων (RNA decay) των αντίστοιχων λειτουργικών γονιδίων (Hirotsume S *et al*, 2003; Yano Y *et al*, 2004). Στο χρωμόσωμα 10 εντοπίστηκαν 430 ψευδογονίδια. Τα περισσότερα (86%) έχουν προκύψει από ρετρομετάθεση επεξεργασμένων μεταγράφων λειτουργικών γονιδίων και είναι ανέκανα να εκφραστούν λόγω έλλειψης των απαραίτητων ρυθμιστικών στοιχείων στη νέα τους θέση αλλά και λόγω εκφυλισμού της νουκλεοτιδικής τους αλληλουχίας από συσσώρευση μεταλλαγών. Αντίθετα, τα μη επεξεργασμένα ψευδογονίδια έχουν προκύψει από διπλασιασμούς γονιδίων ή άνισο ανασυνδυασμό χρωμοσωμάτων και σε αρκετές περιπτώσεις διατηρούν την οργάνωση εξονίων/ιντρονίων του αρχικού γονιδίου. Η απενεργοποίησή τους συμβαίνει κατά την εξέλιξη με ταυτόχρονη συσσώρευση μεταλλαγών. Η παρουσία ψευδογονιδίων είναι πολύ συχνή στο γονιδίωμα των θηλαστικών. Με βάση μελέτες *in silico*, ο αριθμός τους στο ανθρώπινο γονιδίωμα εκτιμάται μεταξύ 11.000 και 20.000, ανάλογα με τα κριτήρια που χρησιμοποιούνται για τον εντοπισμό τους, με το μεγαλύτερο ποσοστό από αυτά να είναι επεξεργασμένα (Zhang Z & Gerstein M, 2004), ενώ εκτιμάται ότι ο αριθμός τους θα υπερβεί τον αριθμό των γονιδίων που κωδικοποιούν πρωτεΐνες (International Human Genome Sequencing Consortium, 2004).

➤ **microRNAs**

Τα microRNAs είναι μη-κωδικοποιήσιμα RNAs μήκους ~22 νουκλεοτιδίων που μεταγράφονται από την RNA πολυμεράση II (Lee Y *et al*, 2004) και καταστέλλουν την μετάφραση γονιδίων-στόχων μέσω σύνδεσής τους με τα αντίστοιχα mRNAs, ρυθμίζοντας έτσι την γονιδιακή έκφραση. Για λίγα από αυτά είναι γνωστοί οι στόχοι τους, π.χ. για τα γονίδια HOX (Yekta S *et al*, 2004), ενώ υπάρχουν ενδείξεις ότι ένα σημαντικό ποσοστό τους υπάρχει αποκλειστικά στα πρωτεύοντα (Bentwich I *et al* 2005). Μέχρι τώρα (Ιούνιος 2005), η βάση δεδομένων των microRNAs (miRBase, <http://microrna.sanger.ac.uk/>) περιλαμβάνει 321 καταχωρήσεις για τον άνθρωπο από τις οποίες οι 223 έχουν επιβεβαιωθεί πειραματικά ενώ οι υπόλοιπες στηρίζονται σε ομολογίες με microRNAs που έχουν κλωνοποιηθεί στον ποντικό, τον αρουραίο και το zebrafish (Griffiths-Jones S, 2004; Ambros V *et al*, 2003). Σε αυτές περιλαμβάνονται 6 microRNAs στο χρωμόσωμα 10.

4.2. Συγκριτική γονιδιωματική

Η ολοκλήρωση της αλληλούχησης του ανθρώπινου γονιδιώματος είναι το πρώτο βήμα για την ανάλυση της λειτουργίας των γονιδίων αλλά και των υπόλοιπων συστατικών του γονιδιώματος. Για αυτήν την ανάλυση απαραίτητη είναι η προσέγγιση της σύγκρισης πληροφοριών από την αποκάλυψη άλλων γονιδιωμάτων, που σήμερα είναι διαθέσιμες από μία μεγάλη ποικιλία οργανισμών.

Η σύγκριση των γονιδιωμάτων των οργανισμών μπορεί να απαντήσει σε μία σειρά από πολύ ενδιαφέροντα ερωτήματα ανάλογα με την εξελικτική απόσταση που τους χωρίζει. Έτσι, από τη σύγκριση του γονιδιώματος του ανθρώπου με τα γονιδιώματα των μονοκύτταρων οργανισμών, εντοπίζονται γονίδια υπεύθυνα για βασικές κυτταρικές λειτουργίες που παρουσιάζονται συντηρημένες κατά την εξέλιξη των ειδών ενώ με τα γονιδιώματα του *C. elegans* και της *D. melanogaster* για πιο πολύπλοκους αναπτυξιακούς μηχανισμούς. Η συγκριτική γονιδιωματική ανάλυση με οργανισμούς με σημαντικές ομοιότητες στη φυσιολογία και τη συμπεριφορά, όπως ο ποντικός, συνεισφέρει τόσο στον προσδιορισμό της οργάνωσης των γονιδίων όσο και στην ταυτοποίηση λειτουργικά σημαντικών ρυθμιστικών στοιχείων. Στην ταυτοποίηση των τελευταίων σημαντικότερη είναι η συμβολή των συγκριτικών αναλύσεων με τα γονιδιώματα των ιχθύων, των πτηνών και των αμφιβίων, καθώς είναι μεγαλύτερη η εξελικτική απόστασή τους από τον άνθρωπο, σε σχέση με τον ποντικό, γεγονός που αυξάνει την διακριτική ικανότητα του εντοπισμού λειτουργικών συντηρημένων περιοχών σε σχέση με συντηρημένων λόγω ανεπαρκούς εξελικτικής απόστασης. Τέλος, οι συγκρίσεις με γονιδιώματα οργανισμών πολύ κοντινών εξελικτικά, όπως ο χιμπατζής, έχουν σκοπό την ταυτοποίηση των ειδοποιών διαφορών που πιθανώς σχετίζονται με τις διαφορές στην μορφολογία, τη φυσιολογία και την συμπεριφορά.

Η σκοπιμότητα της επιλογής κάθε οργανισμού-μοντέλου, η στρατηγική που ακολουθήθηκε για την αλληλούχηση των γονιδιωμάτων τους και οι προσεγγίσεις που εφαρμόζονται για την λειτουργική ανάλυση σε επίπεδο γονιδιώματος για συγκεκριμένους οργανισμούς αναλύθηκαν εκτενώς σε προηγούμενο κεφάλαιο (βλ. Εισαγωγή). Στη συνέχεια, περιγράφονται τα βασικά συμπεράσματα από την συγκριτική γονιδιωματική ανάλυση του ανθρώπου με τα γονιδιώματα άλλων θηλαστικών και των ασπονδύλων.

➤ **Συγκριτική γονιδιωματική ανάλυση του ανθρώπου με άλλα θηλαστικά**

α. Συγκριτική ανάλυση γονιδιωμάτων ανθρώπου και ποντικού-αρουραίου

Εξελικτικά, ο κοινός πρόγονος του ανθρώπου και των δύο αυτών τρωκτικών εκτιμάται ότι έζησε πριν ~75 εκατομμύρια χρόνια. Η απόκλιση των αντίστοιχων γονιδιωμάτων που συνέβη σ' αυτό το χρονικό διάστημα επιτρέπει τόσο την εύκολη στοίχιση των ορθόλογων γονιδίων όσο και την αναγνώριση αρκετών λειτουργικώς σημαντικών μη-κωδικών περιοχών λόγω του υψηλού βαθμού συντήρησής τους.

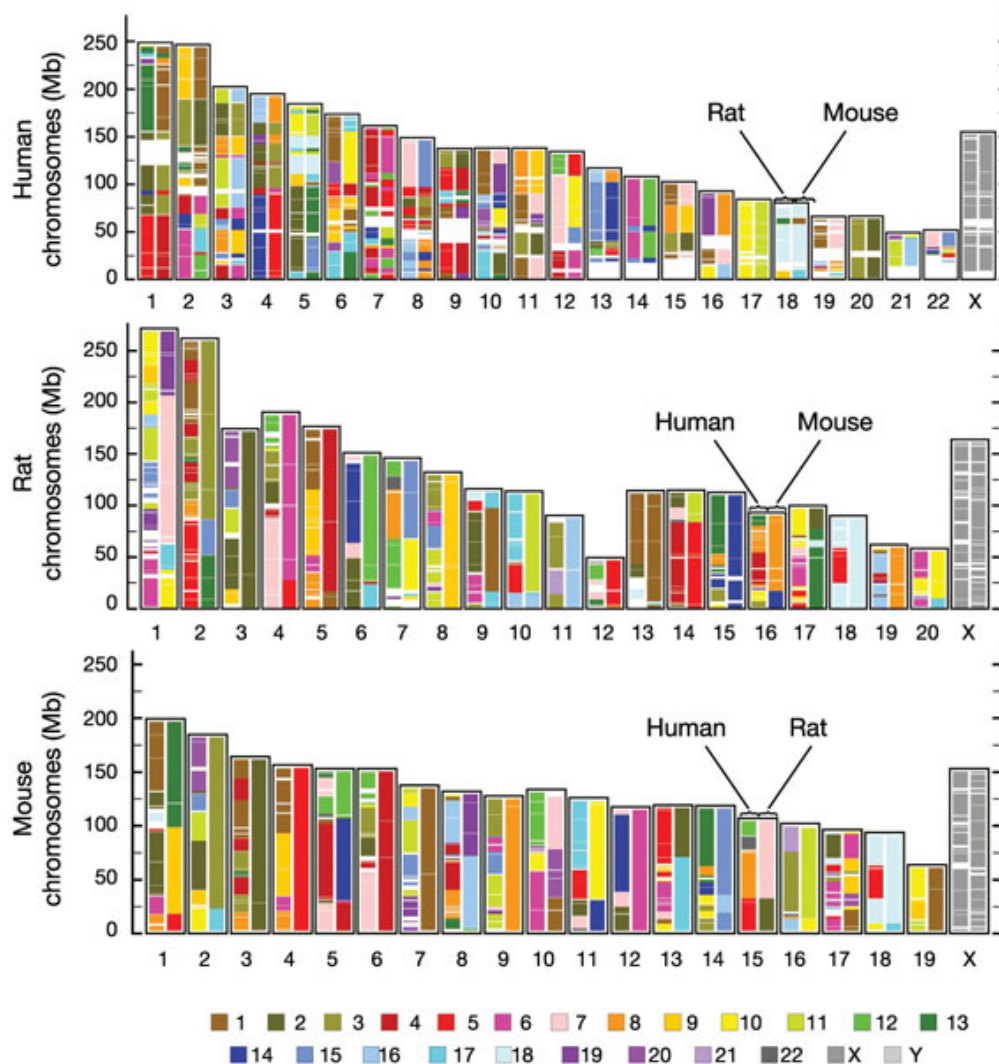
Συνταινία

Παρά την διαφορά στο μήκος του γονιδιώματος μεταξύ του ανθρώπου (~2,9Gb) και του αρουραίου (2,75Gb, Gibbs RA *et al*, 2004) ή του ποντικού (~2,6Gb, NCBI build 32), η χρωμοσωμική οργάνωση των τριών αυτών οργανισμών εμφανίζει σημαντικές ομοιότητες. Αποτέλεσμα αυτού, είναι η δυνατότητα ταυτοποίησης συνταινικών περιοχών, δηλαδή χρωμοσωμικών τμημάτων που η γονιδιακή σύσταση και διάταξη έχει διατηρηθεί. Έτσι, ταυτοποιήθηκαν ~280 συνταινικές περιοχές ανθρώπου/αρουραίου ή ποντικού με μέγεθος >1Mb η καθεμία και στις οποίες εκτιμάται ότι αντιστοιχεί το ~90% του γονιδιώματος του ανθρώπου (Gibbs RA *et al*, 2004, Εικ. 71).

Συντήρηση σε γονίδια

Η ανάλυση των γονιδιωματικών αλληλουχιών του ανθρώπου και των τρωκτικών έδειξε ότι ποσοστό ~5% των γονιδιωμάτων εμφανίζει σημαντική συντήρηση (Waterston RH *et al*, 2002; Gibbs RA *et al*, 2004). Από αυτό, εκτιμάται ότι το 1,5% αντιστοιχεί σε γονίδια που κωδικοποιούν πρωτεΐνες. Ο αριθμός των γονιδίων του ανθρώπου είναι παρόμοιος με αυτόν των τρωκτικών ενώ το ποσοστό των γονιδίων του ανθρώπου για τα οποία μέχρι τώρα δεν έχει εντοπιστεί ομολογία με κάποιο γονίδιο στον ποντικό, και το αντίστροφο, είναι μικρότερο από 1%. Το ποσοστό ομοιότητας κατά μήκος των γονιδίων που κωδικοποιούν πρωτεΐνες ποικίλλει. Έτσι, με δεδομένα που προέρχονται από ~11.000 γονίδια που κωδικοποιούν πρωτεΐνες διαπιστώθηκε ότι το ποσοστό ομοιότητας στις κωδικές περιοχές των γονιδίων, το οποίο είναι παρόμοιο μεταξύ ανθρώπου-αρουραίου και ανθρώπου-ποντικού, είναι σε αμινοξικό επίπεδο, κατά μέσο όρο, 88% και κυμαίνεται μεταξύ 74,4% και 96,4% ενώ σε επίπεδο νουκλεοτιδικής ομοιότητας είναι κατά μέσο όρο 85,1% και κυμαίνεται μεταξύ 77,4% και 90% (Gibbs RA *et al*, 2004, Πιν. 14).

Οι περιοχές 5'UTR και 3'UTR εμφανίζουν μικρότερη συντήρηση από τις κωδικές περιοχές. Ειδικότερα, κατά τη συγκριτική ανάλυση 3165 ζευγών ορθόλογων γονιδίων ανθρώπου-ποντικού το ποσοστό νουκλεοτιδικής ταυτότητας σε αυτές τις περιοχές είναι 75% (Waterston RH *et al*, 2002).



Εικ. 71: Απεικόνιση των συνταινικών χρωμοσωμικών περιοχών μεταξύ ανθρώπου, αρουραίου και ποντικού. Σε κάθε διάγραμμα, που αντιστοιχεί σε έναν οργανισμό, απεικονίζονται τα χρωμοσώματα (άξονας χ) με το μικρό βραχίονα να είναι προς τη βάση του διαγράμματος. Σε κάθε χρωμόσωμα αντιστοιχούν δύο στήλες οι οποίες είναι χρωματισμένες βάσει της συνταινίας με τα χρωμοσώματα των δύο άλλων οργανισμών. Ο ίδιος χρωματικός κώδικας ισχύει για κάθε χρωμόσωμα και για τα τρία είδη (κάτω). Π.χ. οι πρώτες 30 Mb του χρωμοσώματος 15 του ποντικού είναι συνταινικές με τμήμα του χρωμοσώματος 5 του ανθρώπου (συμβολίζεται με κόκκινο στην αριστερή στήλη) και με τμήμα του χρωμοσώματος 2 του αρουραίου (συμβολίζεται με λαδί στην δεξιά στήλη), (ανατύπωση από Gibbs RA *et al*, 2004).

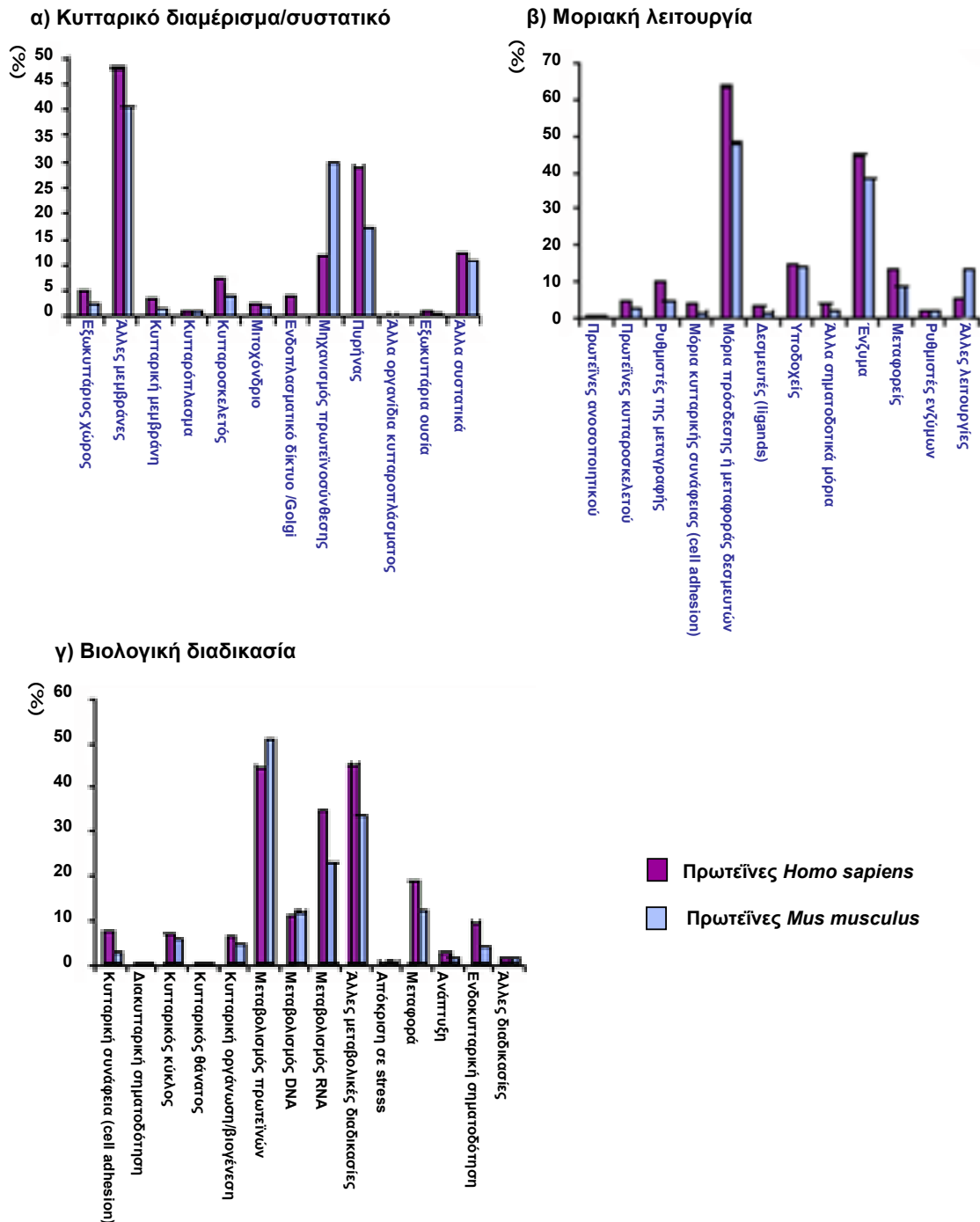
Πιν. 14: Συγκριτική ανάλυση ομοιότητας ορθολόγων γονιδίων ανθρώπου-τρωκτικών

	Άνθρωπος-ποντικός	Άνθρωπος-αρουραίος
Ζεύγη ορθολόγων γονιδίων	11.084	10.066
Μέση τιμή αμινοξικής ταυτότητας (%)	88,0% (74,4-96,3%)	88,3% (75,9-96,4%)
Μέση τιμή νουκλεοτιδικής ταυτότητας (%)	85,1% (77,4-90,0%)	85,1% (77,8-89,9%)

Στα ιντρόνια δεν παρατηρήθηκε σημαντική νουκλεοτιδική ταυτότητα. Επιπλέον, διαπιστώθηκε ότι το μέσο μέγεθος των ιντρονίων του ποντικού (3.9Kb) είναι μικρότερο από αυτό του ανθρώπου (4.6Kb), γεγονός που εν μέρει δικαιολογεί και τη μικρή διαφορά μεγέθους των δύο γονιδιωμάτων. Ωστόσο, φαίνεται πως μεταξύ των δύο οργανισμών διατηρείται η αναλογία των μεγεθών των διαφορετικών ιντρονίων ενός γονιδίου. Επίσης, φαίνεται πως διατηρείται και η κατηγορία των ιντρονίων. Η παρατήρηση αυτή βασίστηκε στην σύγκριση 6352 ορθόλογων ιντρονίων ανθρώπου-ποντικού-αρουραίου από 976 γονίδια (Gibbs RA *et al*, 2004). Κατ' αυτήν δεν εντοπίστηκε καμία μετατροπή ενός ιντρονίου τύπου U2 σε U12 και το αντίστροφο. Επιπλέον, μεταξύ των ιντρονίων τύπου U12 δεν ταυτοποιήθηκε καμία μετατροπή μεταξύ των κύριων υποτύπων (AT-AC και GT-AG), γεγονός που έχει παρατηρηθεί όταν εξετάζονται οργανισμοί εξελικτικά μακρινότεροι (Burge CB *et al*, 1998). Αντίθετα, ταυτοποιήθηκαν μετατροπές μεταξύ των υποτύπων των ιντρονίων τύπου U2, του κανονικού GT-AG και του μη-κανονικού GC-AG. Μόνο το ~70% των ιντρονίων GC-AG είναι συντηρημένα μεταξύ του ανθρώπου και των τρωκτικών ενώ μεταξύ του αρουραίου και του ποντικού το αντίστοιχο ποσοστό είναι 90%.

Σημαντική συντήρηση διαπιστώνεται και στον αριθμό των κωδικών εξονίων μεταξύ ορθόλογων γονιδίων ανθρώπου-ποντικού. Ειδικότερα, η σύγκριση 1506 ζευγών ορθόλογων γονιδίων (RefSeq db), έδειξε ότι το 86% των ζευγών έχουν ίδιο αριθμό κωδικών εξονίων ενώ το αντίστοιχο ποσοστό μειώνεται στο 62% όταν συμπεριληφθούν στην ανάλυση και τα μη κωδικά εξόνια (Waterston RH *et al*, 2002).

Οι πρωτεΐνες του ανθρώπου και του ποντικού ταξινομούνται σε λειτουργικές ομάδες σύμφωνα με το σύστημα ταξινόμησης πρωτεϊνών 'Gene Ontology' (GO, www.geneontology.org/), κατά το οποίο λαμβάνεται υπ' όψιν το κυτταρικό διαμέρισμα εντοπισμού της κάθε πρωτεΐνης, ο μοριακός της ρόλος και η βιολογική λειτουργία στην οποία συμμετέχει (Waterston RH *et al*, 2002). Συγκρίνοντας την κατανομή των πρωτεϊνών του ανθρώπου και του ποντικού σε αυτές τις ομάδες διαπιστώνονται σημαντικές ομοιότητες στους δύο οργανισμούς (Εικ. 72). Αναλυτικότερα, και στα δύο είδη παρατηρούνται υψηλά ποσοστά (40-50%) μεμβρανικών πρωτεϊνών, ενώ οι περισσότερες πρωτεΐνες σχετίζονται με μεταβολικές διαδικασίες. Τέλος, οι πρωτεΐνες του ανθρώπου και του ποντικού που προσδένουν ή μεταφέρουν μόρια-συνδέτες (ligands) αντιστοιχούν σε ποσοστό 65% και 50% αντίστοιχα ενώ αυτές που εμφανίζουν ενζυμική δράση σε ποσοστό 40-45% (Waterston RH *et al*, 2002). Σημαντική ομοιότητα παρατηρείται επίσης και στον αριθμό των μελών στις οικογένειες πρωτεϊνών και των λειτουργικών περιοχών (*domains*) μεταξύ των δύο ειδών.



Εικ. 72: Συγκριτική ταξινόμηση των πρωτεϊνών του ανθρώπου και του ποντικού σε λειτουργικές ομάδες, ανάλογα με το κυτταρικό διαμέρισμα/συστατικό στο οποίο ανήκει η πρωτεΐνη (α), τη μοριακή της λειτουργία (β) και τη βιολογική διαδικασία στην οποία συμμετέχει (γ). Η διαφορά που παρατηρείται στις πρωτεΐνες που αποτελούν συστατικά του μηχανισμού πρωτεϊνοσύνθεσης (στο α), θεωρείται πλασματική λόγω του μεγάλου αριθμού ψευδογονιδίων ριβοσωμικών πρωτεϊνών του ποντικού που εσφαλμένα συγκαταλέγονται μεταξύ των γονιδίων βάσει των προγραμμάτων ανίχνευσης κωδικών αλληλουχιών (Waterston RH *et al*, 2002).

Όσον αφορά τις διαφορές μεταξύ των δύο ειδών, οι κυριότερες εντοπίζονται σε οικογένειες πρωτεϊνών που σχετίζονται με το ανοσοποιητικό και το αναπαραγωγικό σύστημα. Ενδεικτικά, στο γονιδίωμα του ποντικού εντοπίστηκαν 25

οικογένειες πρωτεϊνών, οι οποίες στον άνθρωπο αντιπροσωπεύονται από ένα μόνο μέλος. Σε αυτές ανήκουν 14 οικογένειες που συμμετέχουν στην αναπαραγωγή των τρωκτικών και περιλαμβάνουν μόρια που ρυθμίζονται μεταγραφικά από ανδρογόνα και συμμετέχουν στη βιοσύνθεση ή στον μεταβολισμό ορμονών, στην ανάπτυξη του πλακούντα, στην ωρίμανση των γονάδων και των σπερματοζωαρίων και στην σηματοδότηση από φερομόνες για την επιλογή συντρόφου. Το γεγονός ότι οι περισσότερες οικογένειες πρωτεϊνών που εμφανίζουν επιλεκτική επέκταση στον ποντικό σχετίζονται με την αναπαραγωγή, πιθανόν αντικατοπτρίζει τις φυσιολογικές διαφορές των δύο ειδών, που είναι πολύ πιο εμφανείς στο συγκεκριμένο σύστημα και αφορούν κυρίως τη δομή του πλακούντα, τα μεγέθη των νεογέννητων και την περίοδο κύησης.

Παρά τις σημαντικές ομοιότητες που διαπιστώθηκαν στα γονίδια που κωδικοποιούν πρωτεΐνες μεταξύ ανθρώπου-ποντικού, διάφορες συγκριτικές γονιδιωματικές μελέτες έδειξαν ότι μία ειδο-ειδική πηγή πολυπλοκότητας και διαφορετικότητας αποτελεί η διαφορική εναλλακτική συναρμογή των μεταγράφων στους δύο οργανισμούς. Έτσι, παρά τη διακύμανση στα ποσοστά ειδο-ειδικής συναρμογής που εκτιμήθηκαν ανάλογα με την μεθοδολογία που χρησιμοποιήθηκε, φαίνεται ότι το ποσοστό της συντηρημένης εναλλακτικής συναρμογής είναι χαμηλό (Modrek B & Lee CJ, 2003; Nurtdinov RN *et al*, 2003).

Συντήρηση σε ψευδογονίδια

Από τα ψευδογονίδια που ταυτοποιήθηκαν στον άνθρωπο και τον ποντικό εκτιμάται ότι ποσοστό ~60% των επεξεργασμένων ψευδογονιδίων είναι ειδικά της κάθε γενεαλογίας, δηλαδή δημιουργήθηκαν μετά τον τελευταίο κοινό πρόγονό τους. Ωστόσο, φαίνεται ότι και στους δύο οργανισμούς η πλειονότητα των επεξεργασμένων ψευδογονιδίων έχουν προκύψει από παρόμοιες κατηγορίες γονιδίων, με τα γονίδια των ριβοσωμικών πρωτεϊνών να αποτελούν την μεγαλύτερη ομάδα (Zhang Z *et al*, 2004).

Συντήρηση σε μη κωδικές περιοχές

Ένα χαρακτηριστικό των γονιδιωμάτων των θηλαστικών, το οποίο αναγνωρίστηκε μόλις πρόσφατα, είναι η ιδιαίτερα υψηλή συντήρηση που παρατηρείται σε μικρές μη-γονιδιακές περιοχές τους, που δεν ανήκουν στις επαναλαμβανόμενες αλληλουχίες. Οι περιοχές αυτές ονομάζονται ECR (Evolutionarily Conserved Region) και εναλλακτικά CNG (Conserved Nongenic Sequence) ή MCS (Multi-species Conserved Sequence), και αντιστοιχούν σε ποσοστό 1-2% των γονιδιωμάτων του ανθρώπου, του αρουραίου και του ποντικού (Mallon AM *et al*, 2000; Dermitzakis ET *et al*, 2003; Margulies EH *et al*, 2003). Εκτιμάται ότι υπάρχουν ~327.000 περιοχές ECR στο γονιδίωμα του ανθρώπου από

τις οποίες το 35% χαρτογραφείται σε ιντρόνια και το υπόλοιπο 65% σε περιοχές μακράν και εκτός γονιδίων (Dermitzakis ET *et al*, 2005). Η κατανομή των τελευταίων είναι γενικά ομοιόμορφη ενώ φαίνεται πως ο αριθμός τους είναι μεγαλύτερος σε περιοχές του γονιδιώματος που είναι φτωχές σε γονίδια. Οι ECRs έχουν γίνει αντικείμενο εκτεταμένης έρευνας τελευταία καθώς πιστεύεται ότι πιθανώς αντιστοιχούν σε λειτουργικές περιοχές του γονιδιώματος. Κάποια από αυτά τα ECRs χαρακτηρίζονται ως υπερ-συντηρημένα (Bejerano G *et al*, 2004) καθώς είναι εντελώς ίδια μεταξύ ανθρώπου-αρουραίου και ποντικού.

Για κάποια από τα μη-κωδικά ECRs υπάρχουν ενδείξεις ότι εμφανίζουν χαρακτηριστικά ρυθμιστικών περιοχών (Dermitzakis ET *et al*, 2003), ενώ έχουν αρχίσει να προκύπτουν πειραματικά δεδομένα για την λειτουργικότητα ορισμένων. Ειδικότερα, ταυτοποιήθηκαν ECRs που δρουν σαν μακρινοί ενισχυτές του γονιδίου DACH (dachshund) (Nobrega MA *et al*, 2003) καθώς και των γονιδίων SOX21, PAX6, HLXB9, και SHH (Woolfe A *et al*, 2005). Επιπλέον ταυτοποιήθηκαν ECRs κοντά στο γονίδιο SIM2 (single-minded 2) που *in vitro*, σε συνδυασμό με τον υποκινητή του γονιδίου, εμφανίζουν ικανότητα μεταγραφικής ενεργοποίησης (Frazer KA *et al* 2004). Ωστόσο, φαίνεται ότι κάποια ECRs πιθανόν στερούνται λειτουργίας. Έτσι, η έλλειψη δύο περιοχών μήκους 1817 Kb και 983 Kb στα χρωμοσώματα 3 και 19 αντίστοιχα στον ποντικό, οι οποίες περιέχουν εκατοντάδες ECRs, αλλά στερούνται γονιδίων, αφενός δεν είχε επίδραση στη βιωσιμότητα αφετέρου δεν οδηγεί σε κάποιο ανιχνεύσιμο φαινότυπο (Nobrega MA *et al*, 2004). Αν και υπάρχει περίπτωση τα ποντίκια αυτά να παρουσιάζουν κάποιες αλλαγές σε σχέση με τα αγρίου τύπου, οι οποίες δεν ανιχνεύτηκαν κατά την ανάλυση, τα αποτελέσματα συγκλίνουν στην υπόθεση ότι πιθανώς κάποια ECRs αλλά και κάποιες περιοχές του γονιδιώματος που δεν περιέχουν γονίδια (gene deserts) δεν έχουν λειτουργικό ρόλο.

Στην βελτιστοποίηση της αναγνώρισης των πραγματικά λειτουργικών συντηρημένων αλληλουχιών συνεισφέρει σημαντικά η συγκριτική γονιδιωματική ανάλυση με εξελικτικά μακρινότερους οργανισμούς, όπως οι ιχθύες. Σε αυτά τα πλαίσια, μία πρόσφατη συγκριτική ανάλυση των γονιδιωμάτων του ανθρώπου και του zebrafish οδήγησε στην ταυτοποίηση ~5000 συντηρημένων μη-γονιδιακών περιοχών, σημαντικό ποσοστό των οποίων χαρτογραφούνται σε χρωμοσωμικές περιοχές του ανθρώπου που δεν περιέχουν γονίδια (Shin JT *et al*, 2005). Ο προκαταρκτικός έλεγχος 16 τέτοιων περιοχών για δράση ως *cis*- ρυθμιστικά στοιχεία *in vivo* με χρήση διαγονιδιακών zebrafish, έδειξε ότι τα 10 από αυτά έχουν την ικανότητα ρύθμισης της ιστοειδικής έκφρασης γονιδίων-αναφοράς (Shin JT *et al*, 2005).

β. Συγκριτική ανάλυση γονιδιωμάτων ανθρώπου-χιμπατζή (*Pan troglodytes*)

Πολύ πρόσφατα (Σεπτέμβριος 2005) ολοκληρώθηκε η πρώτη έκδοση του γονιδιώματος του χιμπατζή που καλύπτει το 94% του γονιδιώματος και έχει βάθος διαβάσματος ~3,6 φορές για τα αυτοσωμικά και ~1,8 φορές για τα φυλετικά χρωμοσώματα (Chimpanzee Sequencing and Analysis Consortium, 2005). Επιπλέον, σε προχωρημένο στάδιο είναι η αλληλούχηση του γονιδιώματος του Rhesus macaque, ενός αντιπροσώπου του κλάδου των πιθήκων του παλαιού Κόσμου, μία πρώτη έκδοση του γονιδιώματος του οποίου έγινε διαθέσιμη σε δημόσιες βάσεις δεδομένων τον Μάιο 2005 (<http://www.genome.gov/11008262>; <http://www.hgsc.bcm.tmc.edu/projects/rmacaque>), ενώ προγραμματίζεται ότι τα επόμενα χρόνια θα ολοκληρωθεί η αλληλούχηση του γονιδιώματος του ουραγοτάγκου και θα ξεκινήσει το αντίστοιχο πρόγραμμα για τον *Callithrix jacchus* (common marmoset), ενός πιθήκου του Νέου Κόσμου. Εν τω μεταξύ υπάρχουν διαθέσιμες αλληλουχίες για αρκετά πρωτεύοντα από ESTs, cDNAs, το πρόγραμμα ENCODE (ENCODE Project Consortium, 2004) και για διάφορους μεμονωμένους γενετικούς τόπους.

Σε αντίθεση με τις συγκρίσεις του γονιδιώματος του ανθρώπου με τον αρουραίο, τον ποντικό, τους ιχθύες και το κοτόπουλο κατά τις οποίες γίνεται προσπάθεια ταυτοποίησης περιοχών που εμφανίζουν ασυνήθιστα υψηλή ομοιότητα, κατά τη στοίχιση των γονιδιωμάτων ανθρώπου και χιμπατζή, επειδή η εξελικτική απόσταση είναι μόλις ~6 εκ. χρόνια, γίνεται το αντίθετο. Δηλαδή η ανάλυση εστιάζεται στις διαφορές που παρουσιάζουν τα δύο γονιδιώματα παρά στις ομοιότητες, καθώς η πλειονότητα των βάσεων είναι ίδιες. Έτσι, εκτιμήθηκε ότι οι νουκλεοτιδικές διαφορές εμφανίζονται κατά μέσο όρο σε ποσοστό 1,23% αλλά όταν ληφθούν υπόψη και οι πολυμορφισμοί μέσα σε κάθε είδος, το αντίστοιχο ποσοστό απόκλισης γίνεται 1,06% ή και λιγότερο. Η ανάλυση των τμημάτων που έχουν εντεθεί ή εκλείψει από το ένα ή το άλλο είδος και τα οποία χαρακτηρίζονται ως indel (=insertions-deletions), έδειξε ότι ~40Mb αλληλουχίας υπάρχουν αποκλειστικά στον άνθρωπο. Από αυτές οι ~32Mb αντιστοιχούν σε τμήματα <15kb και η πλειονότητά τους έχει μήκος <80bp και οι υπόλοιπες ~8Mb (>15kb) περιλαμβάνουν 163 περιοχές, από τις οποίες οι 34 χαρτογραφούνται σε εξόνια γνωστών γονιδίων (Chimpanzee Sequencing and Analysis Consortium, 2005). Επιπλέον, επιβεβαιώθηκαν παλαιότερες κυτταρογενετικές μελέτες που αφορούσαν μεγάλες χρωμοσωμικές αναδιατάξεις και οι οποίες είχαν δείξει ότι τα χρωμοσώματα του ανθρώπου και του χιμπατζή διαφέρουν κατά ένα φαινόμενο χρωμοσωμικής σύντηξης και τουλάχιστον 9 περιφερειακές αναστροφές (Yunis JJ *et al*, 1980). Ειδικότερα, το χρωμόσωμα 2 του ανθρώπου

προέκυψε από την σύντηξη δύο αρχαίων χρωμοσωμάτων, των 12 και 13, τα οποία παρέμειναν διαχωρισμένα στον χιμπατζή.

Από τις παραπάνω διαφορές που εντοπίστηκαν μεταξύ των δύο γονιδιωμάτων, αυτές που είναι κρίσιμες από την άποψη ότι οδηγούν στις παρατηρούμενες διαφορές στη φυσιολογία, την μορφολογία και την συμπεριφορά μεταξύ των δύο οργανισμών, αποτελούν την μειονότητα. Συνεπώς ο επόμενος στόχος που είναι και ο δυσκολότερος είναι η ταυτοποίηση των λειτουργικά σημαντικών διαφορών. Στην προσπάθεια αυτή, ο πλήρης κατάλογος όλων των διαφορών που προέκυψε από την αλληλούχηση του γονιδιώματος του χιμπατζή, αποτελεί τη βάση για την συστηματική διερεύνηση και εντοπισμό αυτών που καθορίζουν τα δύο είδη. Ήδη, έχουν ξεκινήσει προκαταρκτικές μελέτες του προτύπου έκφρασης των γονιδίων των ορθόλογων χρωμοσωμάτων σε διάφορους ιστούς (Khaltovich P *et al*, 2005) από τις οποίες έχουν προκύψει σημαντικές ενδείξεις για διαφορές στα επίπεδα έκφρασης. Αυτά τα προκαταρκτικά αποτελέσματα είναι σε συμφωνία με παλαιότερες υποθέσεις κατά τις οποίες η γενετική βάση των διαφορών ανθρώπου-χιμπατζή είναι η διαφοροποίηση στην γονιδιακή έκφραση και όχι οι κωδικές αλληλουχίες (King MC & Wilson AC, 1975). Η ολοκλήρωση των γονιδιωμάτων των άλλων πρωτεύοντων θα συνεισφέρει στη διελεύκανση των διαφορών που ταυτοποιήθηκαν και που δεν σχετίζονται με τα χαρακτηριστικά που εμφανίζονται στο ανθρώπινο είδος.

➤ **Συγκριτική γονιδιωματική ανάλυση του ανθρώπου-ασπονδύλων**

Από τη σύγκριση των πρωτεϊνωμάτων του ανθρώπου με του *C. elegans* και της *D. melanogaster* διαπιστώνεται ότι το 60% των πρωτεϊνών του ανθρώπου έχουν ορθόλογες στους άλλους δύο οργανισμούς και αντίστοιχα, το 61% των πρωτεϊνών της *D. melanogaster* και το 43% των πρωτεϊνών του *C. elegans* εμφανίζουν ομολογία με ανθρώπινες πρωτεΐνες. Στις ορθόλογες πρωτεΐνες μεταξύ των τριών ειδών συγκαταλέγονται μόρια που ελέγχουν θεμελιώδεις λειτουργίες, όπως λειτουργίες ενδοκυτταρικής ή διακυτταρικής σηματοδότησης, βασικοί αναπτυξιακοί μηχανισμοί, απόπτωση και μεταγραφική ρύθμιση γονιδίων.

Τα βασικά συμπεράσματα που προέκυψαν από τη συγκριτική πρωτεωμική ανάλυση σπονδυλωτών-ασπονδύλων είναι ότι:

1) ο ρυθμός ανακάλυψης νέων πρωτεϊνών ή νέων μοτίβων (domains) για κάθε φυλογενετική ομάδα είναι χαμηλός Έτσι η σύγκριση των μοτίβων των πρωτεϊνών του ανθρώπου με του νηματώδη και της δροσόφιλας δείχνει ότι το 90% των μοτίβων των ανθρώπινων πρωτεϊνών αντιπροσωπεύονται και στα άλλα δύο είδη,

2) η μεγαλύτερη ποικιλία στην αρχιτεκτονική και στα λειτουργικά χαρακτηριστικά των πρωτεϊνών του ανθρώπου οφείλεται κατά πολύ στον διαφορετικό συνδυασμό αυτών των μοτίβων. Έτσι, ενώ υπάρχει χαμηλή συχνότητα 'νεοεμφανιζόμενων' πρωτεϊνών στα σπονδυλωτά, διαπιστώνεται υψηλή συχνότητα πρωτεϊνών με καινούρια αρχιτεκτονική, οι οποίες εμφανίζουν νέους συνδυασμούς των προϋπάρχοντων συντηρημένων μοτίβων. Αυτή η διαφορά είναι περισσότερο εμφανής στην πρόσφατη εξέλιξη νέων διαμεμβρανικών ή εξωκυττάρων δομών στις πρωτεΐνες του ανθρώπου και του ποντικού,

3) στα σπονδυλωτά υπάρχει μεγαλύτερη τάση επέκτασης γονιδιακών οικογενειών, γεγονός που υποδηλώνει ότι οι διπλασιασμοί των γονιδίων αποτέλεσαν σημαντικό εξελικτικό μηχανισμό για αυτούς τους οργανισμούς. Πολλές από τις οικογένειες πρωτεϊνών που με αυτόν τον μηχανισμό έχουν αποκτήσει περισσότερα μέλη στα σπονδυλωτά σε σχέση με τη δροσόφιλα ή τον νηματώδη, συμμετέχουν σε διαδικασίες που χαρακτηρίζουν τη φυσιολογία των σπονδυλωτών. Ενδεικτικό παράδειγμα είναι τα πολυάριθμα μέλη της οικογένειας των πρωτεϊνών με μοτίβο ανοσοσφαιρίνης (Ig) στον άνθρωπο και στον ποντικό, σε σχέση με την απουσία της συγκεκριμένης οικογένειας από τον σακχαρομύκητα και τα φυτά και την παρουσία του *μοτίβου* Ig μόνο σε ορισμένα μόρια στην επιφάνεια των κυττάρων των ασπονδύλων. Επιπλέον, στον άνθρωπο και τον ποντικό εντοπίζονται ~30 μέλη της οικογένειας του αυξητικού παράγοντα ινοβλαστών (*Fibroblast Growth Factor*, FGF), σε αντίθεση με τα δύο μέλη της οικογένειας αυτής στη δροσόφιλα και τον νηματώδη.

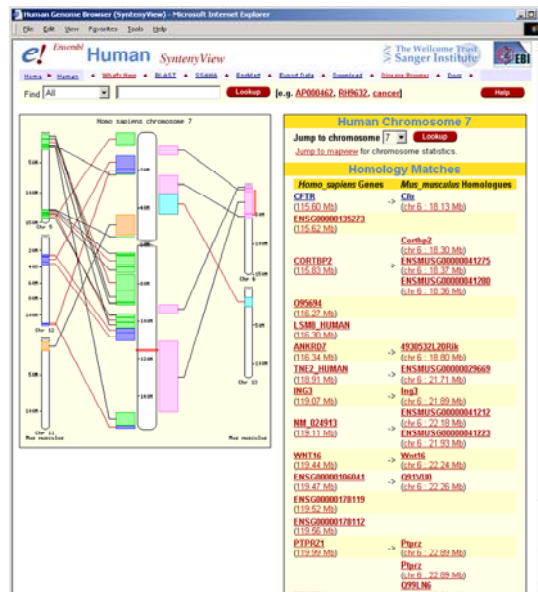
➤ **Πληροφορίες για ορθόλογες περιοχές *in silico***

Πλήθος πληροφοριών για τα ορθόλογα γονίδια του ανθρώπου και των άλλων σπονδυλωτών υπάρχουν διαθέσιμες στην βάση δεδομένων Ensembl (www.ensembl.org/, Εικ. 73). Αυτή η βάση δεδομένων δίνει τη δυνατότητα περιήγησης κατά μήκος των χρωμοσωμάτων καθώς και επιλογής συγκεκριμένων χρωμοσωμικών περιοχών και ταυτόχρονης προβολής τους με τις συντακτικές τους σε άλλους οργανισμούς. Για τα γνωστά γονίδια χρησιμοποιούνται οι κωδικοί πρόσβασης εξωτερικών βάσεων δεδομένων (πχ HUGO, EMBL, Swissprot) αλλά ταυτόχρονα παρατείνονται και όλα τα εναλλακτικά ονόματα του γονιδίου. Για τα νέα γονίδια χρησιμοποιείται η ονοματολογία της Ensembl.

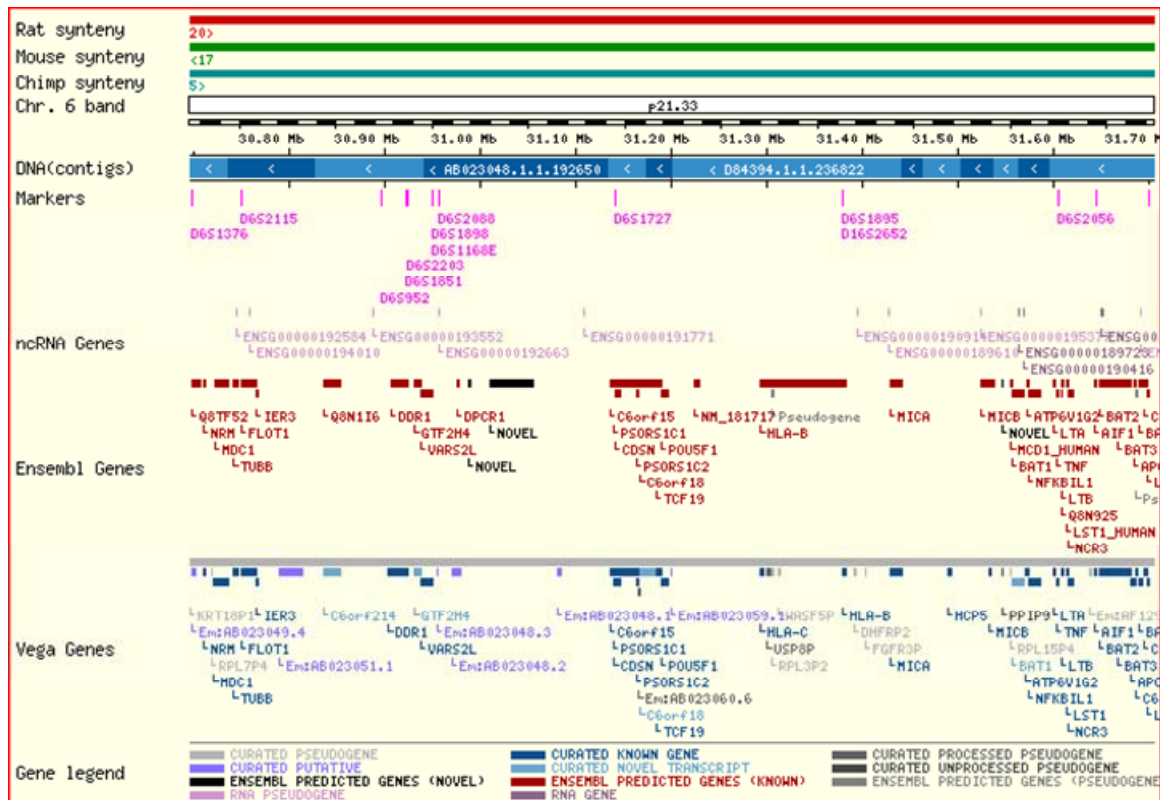
Επιπλέον, σε συγκεκριμένη χρωμοσωμική περιοχή είναι δυνατή η απεικόνιση: 1) του αντίστοιχου κυτταρογενετικού προτύπου ζώνωσης και του συσχετισμού του με τη νουκλεοτιδική αλληλουχία και τους γενωμικούς κλώνους σε έναν ή περισσότερους οργανισμούς ταυτόχρονα, 2) της γενωμικής οργάνωσης κάθε γονιδίου και των

ορθολόγων του καθώς και των πιθανών εναλλακτικών μεταγράφων με καταγραφή των ESTs και cDNAs που τα στηρίζουν, 3) της δομής της κωδικοποιούμενης πρωτεΐνης, 4) των προβλεπόμενων μη κωδικοποιών RNAs, 5) των προβλεπόμενων νησιδίων CpG, 6) των STSs, 7) των SNPs κ.λπ. Επιπλέον, παρέχεται η δυνατότητα νουκλεοτιδικών ή αμινοξικών συγκρίσεων, με χρήση των αντίστοιχων αλγορίθμων.

A. Περιοχές συνταϊνίας (Synteny blocks, ContigView/CytoView)



B.



Γ. Πληροφορίες για γονίδια και τα ορθόλογά τους

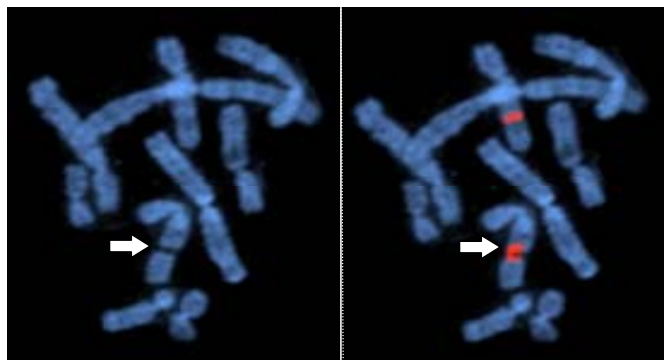
Συνδέσεις για πιθανά ορθόλογα γονίδια

Εικ. 73: Απεικονίσεις από τη βάση δεδομένων *Ensembl*. Παρουσιάζονται ενδεικτικά το χρωμόσωμα 7 του ανθρώπου και η συνταινιά του με χρωμοσώματα του ποντικού (Α), μία συγκεκριμένη χρωμοσωμική περιοχή με τα γονίδια, τα ncRNAs και τα STSs που περιέχει και οι συνταινικές στον ποντικό, τον αρουραίο και τον χιμπατζή (Β), αναλυτικά οι πληροφορίες για ένα γονίδιο με σχετικές συνδέσεις για τα πιθανά ορθόλογά του (Γ) και η συγκριτική γενωμική οργάνωση ενός γονιδίου και του ορθόλογου του στον ποντικό (Δ).

4.3. Μοριακή βάση της σπάνιας αυτοσωμικής εύθραυστης θέσης FRA10A

Στα πλαίσια της διερεύνησης της μοριακής βάσης της αυτοσωμικής σπάνιας εύθραυστης θέσης FRA10A και αξιοποιώντας τον φυσικό χάρτη υψηλής ευκρίνειας και την αλληλουχία του χρωμοσώματος 10, ταυτοποιήσαμε και χαρακτηρίσαμε το γονίδιο *FRA10AC1*. Το γονίδιο επιλέχθηκε βάσει της χαρτογράφησης του στη χρωμοσωμική περιοχή 10q23.3 και της τρινουκλεοτιδικής επανάληψης CGG που φέρει στο 5'UTR του.

Η χαρτογράφηση του γονιδίου στην εύθραυστη θέση FRA10A επιβεβαιώθηκε με ανάλυση FISH σε μεταφασικά χρωμοσώματα τεσσάρων ατόμων-φορέων, χρησιμοποιώντας ως ανιχνευτή το γενωμικό κλώνο BAC bA437J2 του γονιδίου. Από την ανάλυση αποδείχθηκε ότι ο κλώνος αυτός αναγνωρίζει την εν λόγω εύθραυστη θέση και εκτείνεται κατά μήκος της (αποτέλεσμα από την συνεργαζόμενη ομάδα του Dr. R. Seibert, Sarafidou T *et al*, 2004, Εικ. 74).



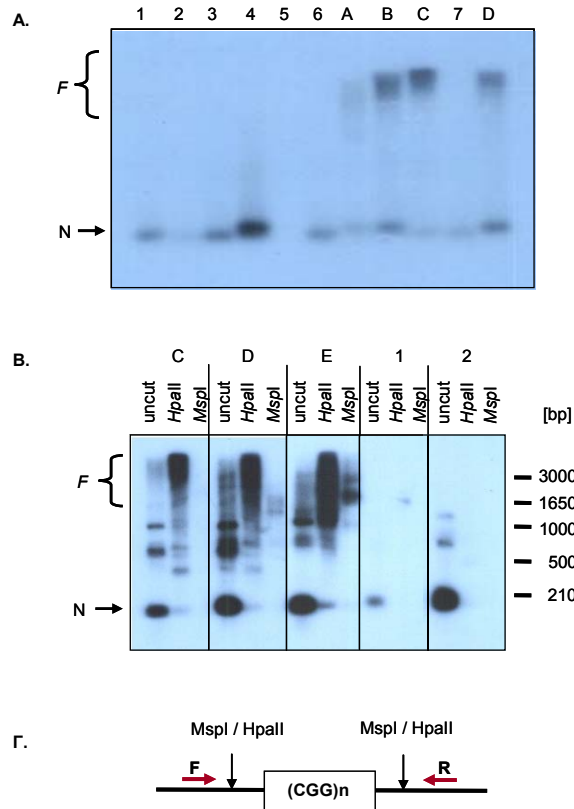
Εικ. 74: Κυτταρογενετική χαρτογράφηση και ανάλυση FISH της εύθραυστης θέσης FRA10A. Μεταφασικά χρωμοσώματα ενός ατόμου-φορέα της FRA10A μετά από χρώση με DAPI (αριστερά) και μετά από FISH με ανιχνευτή το γενωμικό κλώνο του γονιδίου *FRA10AC1* (BAC 437J2, δεξιά). Με κόκκινο, φαίνεται το σήμα εκατέρωθεν της ασυνέχειας, αποτέλεσμα που δείχνει ότι ο κλώνος εκτείνεται κατά μήκος της εύθραυστης θέσης.

Η τρινουκλεοτιδική επανάληψη CGG που εντοπίζεται στο 5'UTR του γονιδίου *FRA10AC1* είναι πολυμορφική στον φυσιολογικό πληθυσμό, ενώ στους φορείς της FRA10A επεκτείνεται δραματικά, περισσότερο από 200 φορές. Ο έλεγχος της επέκτασης στους φορείς, πραγματοποιήθηκε με PCR σε γενωμικό DNA με εκκινητές εκατέρωθεν της επανάληψης και με ακόλουθη ανάλυση των προϊόντων κατά Southern, χρησιμοποιώντας ως ανιχνευτή ολιγονουκλεοτίδια GCC/GGC. Έτσι ανιχνεύτηκαν, εκτός του φυσιολογικού αλληλομόρφου, προϊόντα μεγέθους 600-800bp που καταδεικνύουν την επέκταση της επανάληψης στα άτομα αυτά κατά 200-300 φορές (αποτελέσματα από τα συνεργαζόμενα εργαστήρια των Dr. R. Seibert και Dr. Jozef Gecz, Sarafidou T *et al*, 2004, Εικ. 75). Ο συσχετισμός της επέκτασης των επαναλήψεων CGG με την κυτταρογενετική εμφάνιση εύθραυστης θέσης σε

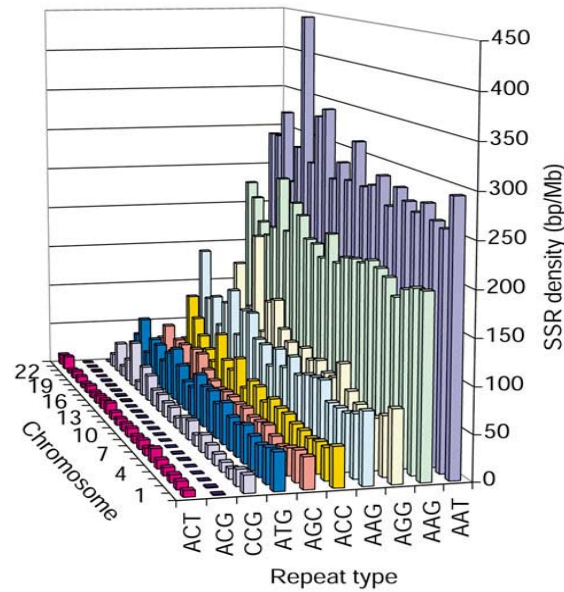
συνθήκες κυτταροκαλλιέργειας απουσία φυλλικού οξέος, έχει αποδειχθεί και στις άλλες πέντε εύθραυστες τις ίδιας κατηγορίας που έχουν κλωνοποιηθεί (FRAXA, FRAXE, FRAXF, FRA11B, FRA16A). Κατ' επαγωγή, πιθανότατα ο ίδιος μηχανισμός ισχύει και στις άλλες θέσεις αυτού του τύπου που έχουν περιγραφεί κυτταρογενετικά. Ωστόσο, ο συνολικός αριθμός αυτών των εύθραυστων θέσεων, που δεν ξεπερνά τις 30, αντιστοιχεί σε ένα μικρό ποσοστό των τρινουκλεοτιδικών επαναλήψεων CGG του ανθρώπινου γονιδιώματος. Σε μία γονιδιωματική *in silico* μελέτη ταυτοποιήθηκαν συνολικά σε όλο το γονιδίωμα 3351 θέσεις που έχουν 4 ή περισσότερες συνεχόμενες επαναλήψεις CGG. Από αυτές οι 170 αντιστοιχούν στο χρωμόσωμα 10 (Subramanian S *et al*, 2003). Στην μελέτη που πραγματοποιήσαμε στην αλληλουχία του χρωμοσώματος αυτού για τον εντοπισμό επαναλήψεων μεγαλύτερου μήκους, που είχαν 7 ή περισσότερες τέλειες επαναλήψεις CGG, ταυτοποιήσαμε 17 τέτοιες θέσεις (βλ. Αποτελέσματα). Ωστόσο, μόνο μία εύθραυστη θέση της κατηγορίας του φυλλικού οξέος χαρτογραφείται στο χρωμόσωμα 10, η FRA10A. Κατ' αναλογία, και με βάση ότι η πυκνότητα των διαφορετικών χρωμοσωμάτων σε επαναλήψεις CGG είναι σχετικά συγκρίσιμη, εκτός από ορισμένες εξαιρέσεις (Subramanian S *et al*, 2003, Εικ. 76) φαίνεται ότι η γενωμική αστάθεια που οδηγεί στην εκδήλωση εύθραυστης θέσης, εμφανίζεται σε συγκεκριμένες μόνο περιοχές του γονιδιώματος, χωρίς να είναι γνωστή η αιτία γι' αυτό.

Η επέκταση της επανάληψης CGG συσχετίζεται με την επιγενετική καταστολή της ενεργότητας του γονιδίου ή/και της ευρύτερης περιοχής λόγω υπερμεθυλίωσης. Η υπερμεθυλίωση του αλληλομόρφου που φέρει την επέκταση ελέγχθηκε με τεμαχισμό γενωμικού DNA των φορέων με διαφορεική χρήση περιοριστικών ενδονουκλεασών που αναγνωρίζουν ή όχι μεθυλιωμένα CpG (*MspI* και *HpaII* αντίστοιχα) και ακολούθως PCR με εκκινητές εκατέρωθεν των θέσεων αναγνώρισης των ενζύμων και ανάλυση κατά Southern με ανιχνευτή ολιγονουκλεοτίδια GCC/GGC. Το αποτέλεσμα ήταν η ανίχνευση του αλληλομόρφου που φέρει την επέκταση αποκλειστικά στα δείγματα που είχαν τεμαχιστεί με *HpaII* και όχι με *MspI*, γεγονός που συνεπάγεται ότι το αλληλόμορφο που φέρει την επέκταση είναι υπερμεθυλιωμένο (αποτελέσματα από τα συνεργαζόμενα εργαστήρια των Dr. R. Seibert και J.Gecz, Sarafidou T *et al*, 2004, Εικ. 75). Για τον έλεγχο της πιθανής μεταγραφικής καταστολής του αλληλομόρφου που φέρει την επέκταση αξιοποιήθηκαν τα SNPs στα nt:284 και 427 (βλ. Αποτελέσματα) για τα οποία ένας φορέας της FRA10A ήταν ετερόζυγος (γονότυπος: 284T/C-427C/G). Έτσι, πραγματοποιήθηκε RT-PCR σε mRNA από κυτταρική σειρά αυτού του ατόμου και ακόλουθη αλληλούχηση του προϊόντος που έδειξε ότι είναι δυνατή η ανίχνευση αποκλειστικά του ενός αλληλο-μόρφου (284C-427G). Βάσει αυτού προκύπτει η

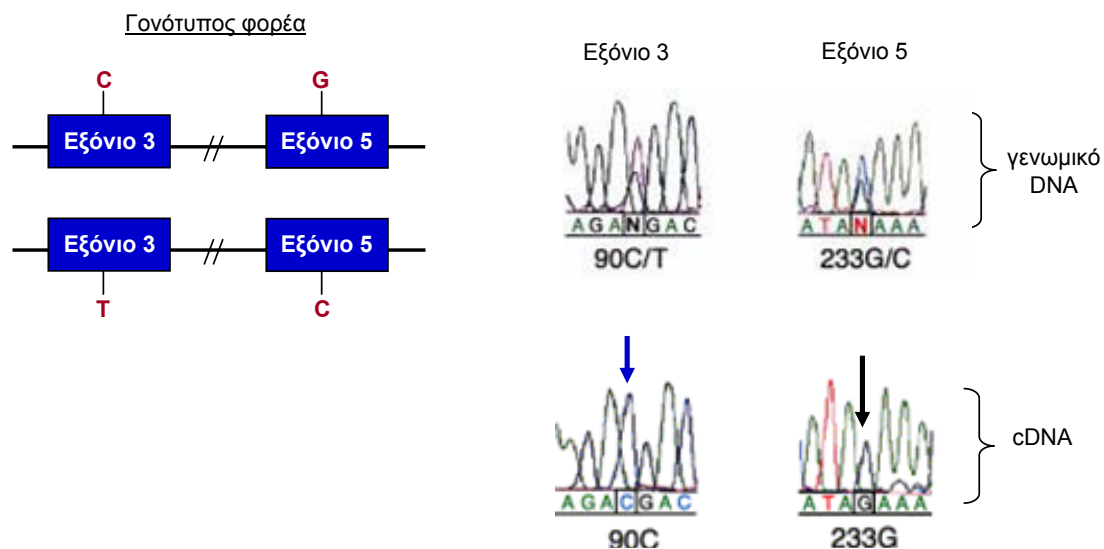
υπόθεση ότι το ένα αλληλόμορφο του γονιδίου έχει κατασταλεί μεταγραφικά λόγω υπερμεθυλίωσης στην περιοχή της επανάληψης, που με τη σειρά της οφείλεται στην επέκτασή της (αποτέλεσμα από το συνεργαζόμενο εργαστήριο του Dr. J.Gecz, Sarafidou T *et al*, 2004, Εικ. 77).



Εικ. 75: Επέκταση και μεθυλίωση της τρινουκλεοτιδικής επανάληψης CGG σε φορείς της εύθραυστης θέσης FRA10A. **A.** Αυτοραδιογραφία μετά από ανάλυση Southern και υβριδοποίηση, προϊόντων PCR με υπόστρωμα γενωμικό DNA ατόμων ελέγχου (διαδρομές 1-7) και φορέων της FRA10A (A, B, C και D) και εκκινητές που υβριδοποιούνται εκατέρωθεν της επανάληψης. Στους φορείς, εκτός από το φυσιολογικό αλληλόμορφο (N) εμφανίζονται επιπλέον προϊόντα PCR μεγαλύτερου μεγέθους (F), γεγονός που δείχνει επέκταση της επανάληψης. **B.** Για να ελεγχθεί εάν το αλληλόμορφο που φέρει την εκτεταμένη επανάληψη υφίσταται μεθυλίωση, πραγματοποιήθηκε τεμαχισμός γενωμικού DNA ατόμων ελέγχου (1, 2) και φορέων της FRA10A με *Hpa II* ή *MspI*, το οποίο ακολούθως χρησιμοποιήθηκε ως υπόστρωμα σε PCR με εκκινητές που υβριδοποιούνται εκατέρωθεν της επανάληψης (βλ. σχηματική απεικόνιση στην Εικ. Γ). Παράλληλα, πραγματοποιήθηκε PCR σε γενωμικό DNA που δεν είχε υποστεί τεμαχισμό. Στα προϊόντα της αντίδρασης, πραγματοποιήθηκε ανάλυση κατά Southern και υβριδοποίηση με ανιχνευτή ολιγονουκλεοτίδια GCC/GGC. Η αυτοραδιογραφία δείχνει ότι οι εκτεταμένες επαναλήψεις ανιχνεύονται μόνο στο δείγμα του DNA που δεν είχε υποστεί τεμαχισμό και στο γενωμικό DNA που είχε τεμαχιστεί με *HpaII*, ενώ απουσιάζουν στα δείγματα των φορέων που είχαν τεμαχιστεί με *MspI*. Στα άτομα ελέγχου ταυτοποιείται αποκλειστικά το φυσιολογικό αλληλόμορφο. Τα αποτελέσματα δείχνουν ότι η περιοχή γύρω από την επανάληψη CGG είναι υπερμεθυλιωμένη. Τα αλληλόμορφα με την επέκταση δεν εμφανίζονται σαν διακριτές ζώνες πιθανότατα λόγω δυσκολιών κατά τον πολλαπλασιασμό κατά μήκος της επανάληψης.



Εικ. 76: Πυκνότητα των διαφορετικών τύπων τρινουκλεοτιδικών επαναλήψεων στα χρωμοσώματα του ανθρώπου. Από το διάγραμμα φαίνεται ότι οι επαναλήψεις του τύπου CGG είναι από τις σχετικά σπάνιες στο ανθρώπινο γονιδίωμα και η κατανομή τους είναι γενικά ομοιόμορφη.



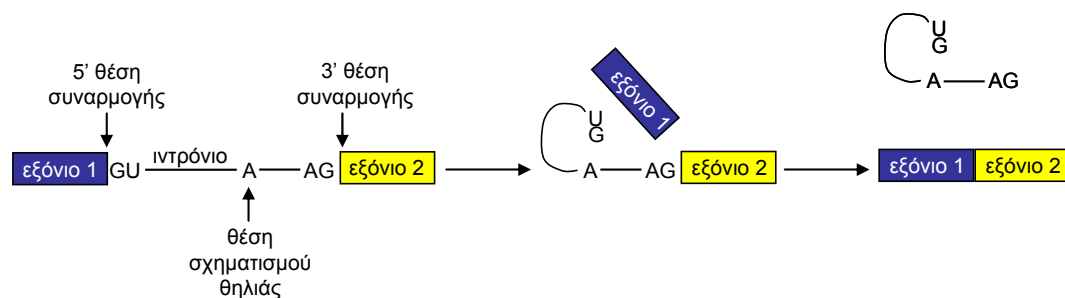
Εικ. 77: Μεταγραφική καταστολή του ενός αλληλομόρφου του γονιδίου *FRA10AC1* στους φορείς της εύθραυστης θέσης *FRA10A*. **A.** Απεικονίζεται σχηματικά ο γονότυπος ενός φορέα της *FRA10A*, ο οποίος είναι ετερόζυγος σε δύο SNPs της κωδικής αλληλουχίας: το 284C/T στο εξόνιο 3 και το 427G/C στο εξόνιο 5 του cDNA. **B.** Αποτέλεσμα αλληλούχησης στην περιοχή των SNPs σε γενωμικό DNA του φορέα (πάνω) και σε αντίδραση RT-PCR σε mRNA από κυτταρική σειρά του φορέα (κάτω). Ενώ η γενωμική αλληλουχία δείχνει ετεροζυγωτία για τα αλληλόμορφα T/C και G/C και στις δύο θέσεις, η αλληλουχία του cDNA δείχνει την παρουσία του ενός μόνο αλληλομόρφου.

➤ **Πρωτεΐνες που αλληλεπιδρούν με την FRA10AC1**

Στα πλαίσια της διερεύνησης του βιοχημικού ρόλου της πρωτεΐνης FRA10AC1 εφαρμόσαμε το σύστημα δύο υβριδίων του σακχαρομύκητα και ταυτοποιήσαμε ως αλληλεπιδρώντα μόρια, τις πρωτεΐνες SAP145 και DGCR14. Το γεγονός ότι και οι δύο πρωτεΐνες έχουν συσχετιστεί με τη διαδικασία της συναρμογής του πρόδρομου mRNA, μας οδηγούν στην υπόθεση ότι πιθανώς και η FRA10AC1 συμμετέχει στο ίδιο βιολογικό μονοπάτι.

Αντίδραση συναρμογής

Η συναρμογή του πρόδρομου mRNA είναι μία καταλυτική διαδικασία που περιλαμβάνει δύο αντιδράσεις τρανς-εστεροποίησης κατά τις οποίες απομακρύνονται τα ιντρόνια ενώ παράλληλα συνδέονται τα διαδοχικά εξόνια. Κατά την πρώτη αντίδραση, το 2'-OH της αδενοσίνης της θέσης σχηματισμού θηλιάς (branch site) του ιντρονίου, συνδέεται με τη φωσφορική ομάδα του νουκλεοτιδίου της 5' θέσης συναρμογής (πρώτο νουκλεοτίδιο του ιντρονίου). Έτσι δημιουργούνται τα ενδιάμεσα μόρια, δηλαδή το ελεύθερο 5' εξόνιο και το ιντρόνιο σε μορφή βρόχου ενωμένο με το 3' εξόνιο. Στη δεύτερη αντίδραση, πραγματοποιείται η σύνδεση του 3'-OH του ελεύθερου εξονίου (5' εξόνιο) με τη φωσφορική ομάδα του νουκλεοτιδίου της 3' θέσης συναρμογής με αποτέλεσμα την απομάκρυνση του ιντρονίου και τη σύνδεση των ελεύθερων πλέον άκρων των δύο εξονίων (Εικ. 78).



Εικ. 78: Τα δύο καταλυτικά βήματα της αντίδρασης συναρμογής

Το spliceosome

Η συναρμογή πραγματοποιείται από ένα υπερμοριακό σύμπλοκο, το spliceosome που αποτελείται από U snRNAs (uridine-rich small nuclear RNAs) και πρωτεΐνες. Τα συστατικά αυτά σχηματίζουν επιμέρους ριβονουκλεοπρωτεϊνικές δομές, τα U snRNPs, που ονομάζονται βάσει του αντίστοιχου U snRNAs που περιέχουν. Στα θηλαστικά, οι πρωτεΐνες των U snRNPs ταξινομούνται σε δύο κατηγορίες: α) στις snRNPs (μικρές πυρηνικές ριβονουκλεοπρωτεΐνες) και β) στις μη-

snRNPs που ονομάζονται και παράγοντες συναρμογής. Ο παραπάνω διαχωρισμός βασίζεται στο κατά πόσο οι πρωτεΐνες αυτές παραμένουν ή όχι συνδεδεμένες με ένα συγκεκριμένο U snRNA κατά τη διάρκεια των βιοχημικών κλασμάτων. Το spliceosome διακρίνεται στο μείζον και το ελάσσον. Το πρώτο είναι υπεύθυνο για την απομάκρυνση των περισσότερων ιντρονίων, των τύπου U2, από τα πρόδρομα mRNAs ενώ το ελάσσον απομακρύνει τα ιντρόνια τύπου U12, που απαντώνται σπάνια. Το μείζον spliceosome περιλαμβάνει τις υπομονάδες snRNP U1, U2, U4, U5 και U6 και το μικρό τις U11, U12, U4atac και U6atac.

Σχηματισμός των υπομονάδων U snRNP

Τα snRNPs σχηματίζονται σταδιακά από ένα μόριο U snRNA, 7 πρωτεΐνες Sm και επιπλέον πρωτεΐνες, ειδικές για κάθε snRNP (Εικ. 79). Τα snRNAs μεταγράφονται από την RNA pol II, πλην των U6 και U6atac που μεταγράφονται από την RNA pol III. Όλα, σχηματίζονται αρχικά ως πρόδρομα μόρια που περιέχουν επιπλέον νουκλεοτίδια στο 3' τους άκρο και μία μονομεθυλιωμένη δομή cap, m7GpppG (m7G). Στη συνέχεια, εξέρχονται στο κυτταρόπλασμα, μέσω ενός πολυπρωτεϊνικού συμπλόκου εξόδου, όπου και συνδέονται με τις πρωτεΐνες Sm. Οι τελευταίες, που βρίσκονται στη μορφή τριών ετερομερών συμπλόκων (D1-D2, E-F-G και B-D3 ή B'-D3) συνδέονται σε μία πολύ συντηρημένη θέση των RNAs, που ονομάζεται θέση Sm (PuAU₄₋₆GPu). Ακολουθεί η ωρίμανση των πρόδρομων snRNAs (απομάκρυνση των επιπλέον νουκλεοτιδίων στο 3' άκρο τους και μετατροπή της δομής m7G σε 2,2,7 τριμεθυλιωμένη γουανοσίνη, m3G), η είσοδος του ενδιάμεσου συμπλόκου στον πυρήνα, μέσω των πρωτεϊνών Snurportin-1 και Importin β και τέλος η σύνδεση των ειδικών για κάθε snRNP πρωτεϊνών.

Σχηματισμός του spliceosome

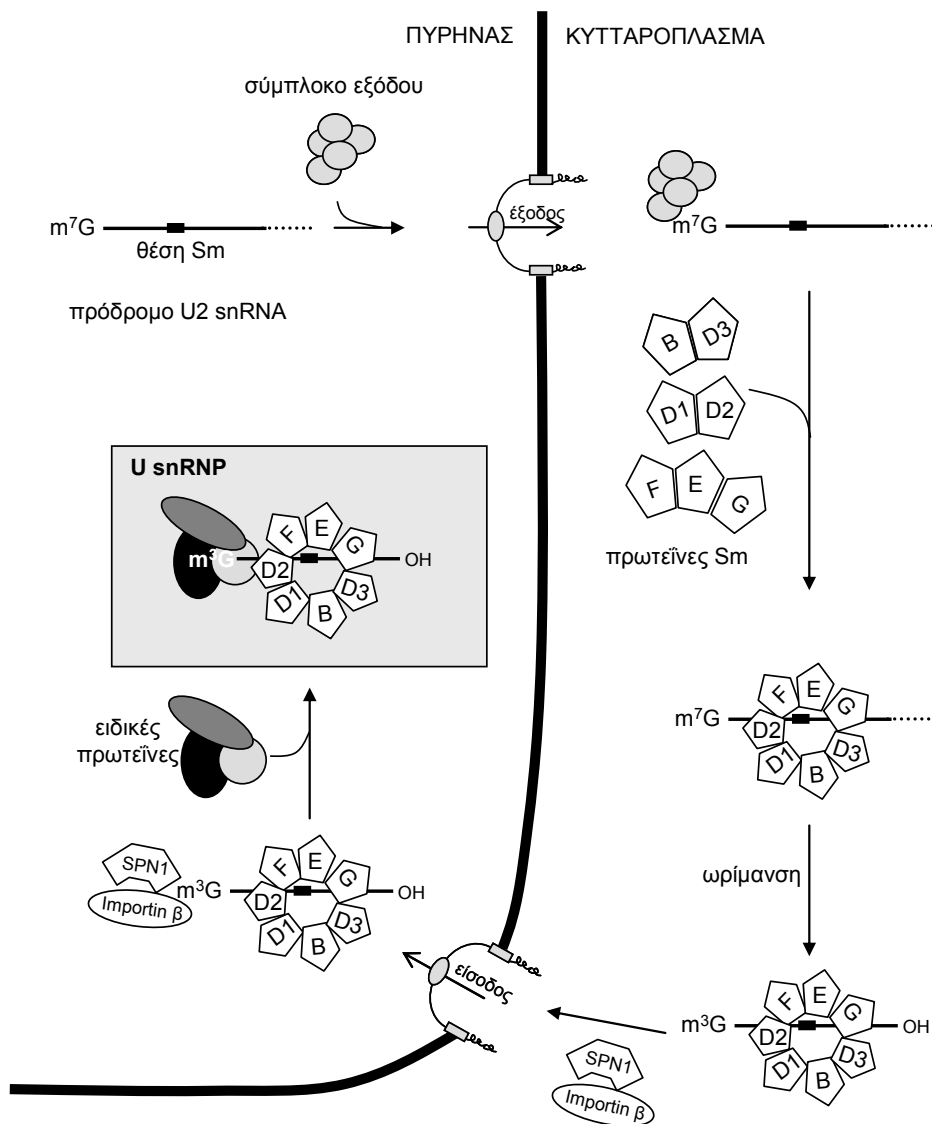
Η βασική λειτουργία των υπομονάδων U snRNP είναι να αναγνωρίζουν τα όρια εξονίων-ιντρονίων και να στρατολογήσουν τους παράγοντες που καταλύουν την συναρμογή του πρόδρομου mRNA. Αυτή η στρατολόγηση γίνεται με συγκεκριμένη ακολουθία και συνεπώς ο σχηματισμός του spliceosome επιτυγχάνεται με την ιεραρχημένη αλληλεπίδραση των συστατικών του (Εικ. 80). Τα πολυπρωτεϊνικά σύμπλοκα που δημιουργούνται σταδιακά, έχουν συγκεκριμένη ονοματολογία: έτσι σχηματίζεται διαδοχικά το spliceosome E, A, B και C. Η συναρμογή του πρόδρομου mRNA είναι μία δυναμική διαδικασία κατά την οποία συμβαίνουν πολλές ανακατατάξεις στις αλληλεπιδράσεις των μορίων που συμμετέχουν σε κάθε στάδιο σχηματισμού και αποσχηματισμού του spliceosome και στα καταλυτικά βήματα της διαδικασίας. Όσον αφορά το μείζον spliceosome, το πρώτο βήμα για τον σχηματισμό

του είναι η αναγνώριση της θέσης συναρμογής 5' του ιντρονίου από την υπομονάδα U1 snRNP. Η αναγνώριση αυτή βασίζεται στη συμπληρωματικότητα των βάσεων της θέσης συναρμογής 5' με τμήμα του U1 snRNA, σε πρωτεϊνικές αλληλεπιδράσεις και σε αλληλεπιδράσεις μεταξύ πρωτεϊνών και πρόδρομου mRNA που τελικά οδηγούν στο σχηματισμό του συμπλόκου E. Σύμφωνα με το κλασικό μοντέλο, ακολουθεί η στρατολόγηση της υπομονάδας U2 snRNP που αναγνωρίζει τη θέση σχηματισμού θηλιάς (branch site) στο πρόδρομο mRNA. Η διαδικασία αυτή απαιτεί ενέργεια και βασίζεται στη συμπληρωματικότητα των βάσεων των μορίων RNA που συμμετέχουν καθώς και σε αλληλεπιδράσεις μεταξύ πρωτεϊνών και πρόδρομου mRNA. Στις αλληλεπιδράσεις αυτές συμμετέχουν οι υπομονάδες SF3a και SF3b που ανήκουν στις ειδικές πρωτεΐνες του U2 snRNP καθώς και παράγοντες συναρμογής όπως ο U2AF και ο SF1/BBP. Το αποτέλεσμα των αλληλεπιδράσεων αυτών είναι ο σχηματισμός του συμπλόκου A που ονομάζεται και pre-spliceosome. Το επόμενο βήμα είναι η στρατολόγηση των υπομονάδων U4, U5 και U6 snRNP, που έχουν ήδη συνδεθεί μεταξύ τους και έχουν σχηματίσει το tri-snRNP U4/U6.U5 και η αλληλεπίδραση του τελευταίου με το πρόδρομο mRNA. Η διαδικασία αυτή απαιτεί ενέργεια και το αποτέλεσμα είναι ο σχηματισμός του συμπλόκου B. Κατά τη διάρκεια σχηματισμού αυτού του συμπλόκου λαμβάνουν χώρα πολλές δομικές αλλαγές π.χ. το U6 snRNA παύει να αλληλεπιδρά με το U4 snRNA και συνδέεται με το U2 snRNA και την θέση συναρμογής 5', το U5 snRNA συνδέεται με αλληλουχίες του εξονίου στη θέση συναρμογής 5' και αργότερα με αλληλουχίες του εξονίου στη θέση 3' κ.λπ. Όλες οι δομικές αλλαγές που συμβαίνουν οδηγούν στο σχηματισμό του συμπλόκου C κατά το οποίο το πρόδρομο mRNA υπόκειται το πρώτο καταλυτικό βήμα. Μετά το δεύτερο καταλυτικό βήμα, το spliceosome αποσυναρμολογείται και απελευθερώνεται το ώριμο πλέον mRNA και τα ιντρόνια.

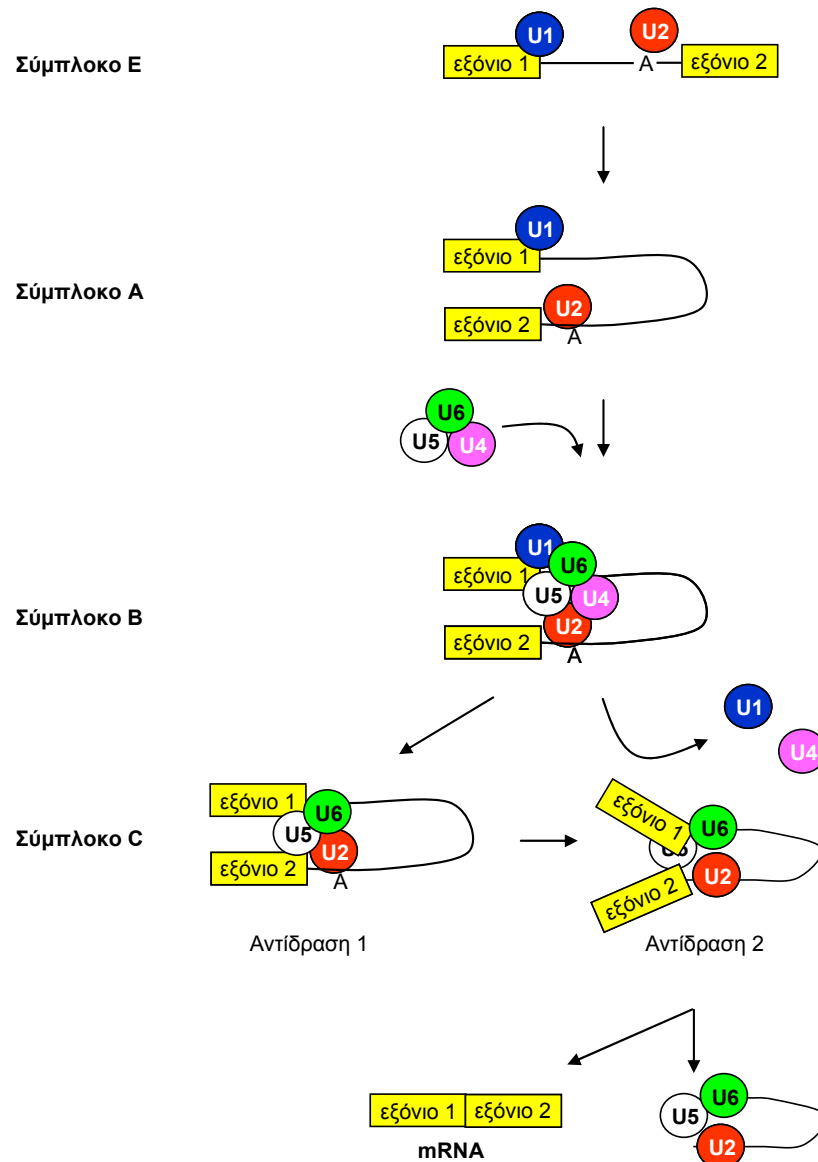
Πολυπλοκότητα του spliceosome

Από βιοχημικά πειράματα απομόνωσης του spliceosome και ταυτοποίησης των υπομονάδων του φάνηκε ότι πρόκειται για ένα σύμπλοκο πολύ πιο πολύπλοκο από ότι θεωρούνταν παλαιότερα, καθώς αποτελείται από τουλάχιστον 300 πρωτεΐνες. Ένα βασικό ερώτημα και με δεδομένο ότι η αντίδραση της τρανσεστεροποίησης μπορεί να πραγματοποιηθεί και απουσία πρωτεϊνών (Cech TR, 1993), είναι η ταυτοποίηση του ρόλου αυτής της πληθώρας των πρωτεϊνών στο spliceosome. Μία πιθανή παράμετρος που συνεισφέρει στην πολυπλοκότητα του συμπλόκου, είναι η ακρίβεια με την οποία πραγματοποιείται η συναρμογή. Κάθε σήμα συναρμογής ελέγχεται πολυάριθμες φορές από διαφορετικούς παράγοντες ώστε να επιβεβαιωθεί ο σωστός σχηματισμός του spliceosome τόσο πριν όσο και

μετά το πρώτο καταλυτικό βήμα (Staley JP & Guthrie C, 1998). Επιπλέον, η παρατηρούμενη πολυπλοκότητα μπορεί να ερμηνευτεί και με βάση τη λειτουργική σύνδεση της συναρμογής με άλλα στάδια της γονιδιακής έκφρασης. Παρόλο που κάθε βήμα της γονιδιακής έκφρασης πραγματοποιείται από διαφορετική πρωτεϊνική «μηχανή», πρόσφατα δεδομένα δείχνουν ότι οι μεμονωμένες πρωτεϊνικές «μηχανές» συνδέονται λειτουργικά, δημιουργώντας πρωτεϊνικά «εργοστάσια», κατοχυρώνοντας πιθανώς έτσι την πιστότητα, την επαναληψιμότητα και το συγχρονισμό της όλης διαδικασίας *in vivo*. Ειδικότερα, εκτός από τη σύζευξη της συναρμογής με τα άλλα στάδια ωρίμανσης του πρόδρομου mRNA (προσθήκη δομής cap στο 5' άκρο, πολυαδενυλίωση), υπάρχουν σημαντικές ενδείξεις για τη λειτουργική συσχέτιση της συναρμογής με τη μεταγραφή καθώς και με την έξοδο του ώριμου mRNA από τον πυρήνα.



Εικ. 79: Απλοποιημένο σχήμα της βιογένεσης των U snRNPs



Εικ. 80: Σχηματισμός ενδιάμεσων συμπλόκων κατά την αντίδραση συναρμογής

Η πρωτεΐνη SAP145

Η πρωτεΐνη SAP145 (Spliceosome Associated Protein, 145kD) αποτελεί έναν λειτουργικά χαρακτηρισμένο παράγοντα συναρμογής. Το πολυπεπτίδιο αποτελείται από 895 αμινοξέα και έχει εκτιμώμενο μοριακό βάρος ~100kD. Ωστόσο, λόγω του ηλεκτροφορητικού του προτύπου σε μεγαλύτερα μεγέθη, ονομάστηκε υπομονάδα 145kD. Αποτελεί συστατικό του συμπλόκου SF3b (Gozani O *et al*, 1996) που με τη σειρά του συμμετέχει στο σχηματισμό τόσο του U2 snRNP όσο και του U11/U12 snRNP (Will CL *et al*, 1999). Ο SF3b είναι ένα ιδιαίτερα σταθερό σύμπλοκο (Brosi R *et al*, 1993), απαραίτητο για το σχηματισμό του pre-spliceosome (Brosi R *et al*, 1993) και παραμένει στο spliceosome μέχρι και την ολοκλήρωση της διαδικασίας της

συναρμογής (Champion-Arnaud P & Reed R, 1994). Εκτός της SAP145, ο SF3b περιλαμβάνει τουλάχιστον 7 επιπλέον πρωτεΐνες έχοντας συνολικό μοριακό βάρος 450kD (Will CL *et al* 2001 & 2002). Η πρωτεΐνη SAP145 συμμετέχει στη σύνδεση του U2 snRNA με τη θέση σχηματισμού θηλιάς του ιντρονίου και αλληλεπιδρά άμεσα με την SAP49, μία από τις άλλες υπομονάδες του SF3b (Champion-Arnaud P & Reed R, 1994). Η πρωτεΐνη είναι συντηρημένη στους ευκαρυωτικούς οργανισμούς, ζωικούς και φυτικούς και ορθόλογό της υπάρχει και στον σακχαρομύκητα. Σημαντικό είναι ότι στην *D. melanogaster* και τον σακχαρομύκητα είναι συντηρημένη η αλληλεπίδραση της SAP145 με την SAP49, γεγονός που δείχνει και την λειτουργική σημασία της (Flygrid db, Igel H *et al*, 1998).

Ιδιαίτερο ενδιαφέρον παρουσιάζει ο φαινότυπος του ποντικού με απενεργοποιημένο το ένα αλληλόμορφο του γονιδίου *SAP145* (αναφέρεται σαν αδημοσίευτα αποτελέσματα στους Isono K *et al*, 2005). Τα ποντίκια αυτά παρουσιάζουν μεταμόρφωση του τελευταίου οσφυϊκού σπονδύλου σε ιερό, με διεισδυτικότητα 71%. Η ίδια αναπτυξιακή ανωμαλία εμφανίζεται και σε ετερόζυγα knock-out ποντίκια για το γονίδιο *SAP155*, το προϊόν του οποίου αποτελεί, όπως και η *SAP145*, υπομονάδα του συμπλόκου SF3b. Τα ποντίκια αυτά εμφανίζονται επιπλέον σκελετικές ανωμαλίες κατά μήκος του πρόσθιο-οπίσθιο άξονα. Ειδικότερα, παρατηρείται μεταμόρφωση του έβδομου αυχενικού σπονδύλου σε θωρακικό, του πρώτου θωρακικού σε δεύτερο και του τελευταίου θωρακικού σε οσφυϊκό. Επιπλέον, στα ποντίκια *SAP155 +/-* παρατηρείται εκτοπική έκφραση των γονιδίων *Hoxc6*, *Hoxb8*, *Hoxb6* (Isono K *et al*, 2005), τα οποία ανήκουν στην οικογένεια των ομοιωτικών γονιδίων που φυσιολογικά εμφανίζουν καθορισμένα πρότυπα έκφρασης και παίζουν κρίσιμο ρόλο στην εμβρυϊκή ανάπτυξη δρώντας ως μεταγραφικοί ρυθμιστές. Ο παρατηρούμενος φαινότυπος των ετερόζυγων ποντικών *SAP155 +/-* μοιάζει με αυτόν που προκύπτει από απενεργοποίηση γονιδίων της οικογένειας Polycomb. (Akasaka T *et al*, 1996). Οι πρωτεΐνες Polycomb σχηματίζουν πολυπρωτεϊνικά σύμπλοκα που γενικά καταστέλλουν τα ομοιοτικά γονίδια. Ο φαινότυπος σε συνδυασμό με την ταυτοποίηση της άμεσης αλληλεπίδρασης μεταξύ της πρωτεΐνης *SAP155* και των *Zfp144* και *Rnf2*, που αποτελούν μέλη της οικογένειας των γονιδίων Polycomb, υποδηλώνει την πιθανή συμμετοχή της *SAP155* και ίσως και της *SAP145*, στην ρύθμιση των γονιδίων *Hox* μέσω ή συνεργατικά με μέλη αυτής της οικογένειας. Η πρωτεΐνη *SAP155* είχε και παλαιότερα συσχετιστεί έμμεσα με την οικογένεια Polycomb, λόγω της αλληλεπίδρασης του παράγοντα συναρμογής NIPP1, ο οποίος συνδέεται άμεσα με την *SAP155*, με την πρωτεΐνη EED (embryonic ectoderm development), μέλος της παραπάνω οικογένειας (Jin Q *et al*, 2003). Η ομόζυγη έλλειψη του γονιδίου *SAP155* προκαλεί θάνατο των ποντικών

πριν το στάδιο της εμφύτευσης, στο στάδιο των 16-32 κυττάρων, φαινότυπος συμβατός με τον κεντρικό ρόλο της πρωτεΐνης στην διαδικασία της συναρμογής.

Σύμφωνα με τα αποτελέσματα των ποντικών knock-out, θα μπορούσε η SAP155 και πιθανώς και η SAP145, να αποτελεί μία πρωτεΐνη Polycomb η οποία συμμετέχει στο σχηματισμό κατασταλτικών συμπλόκων μαζί με άλλα μέλη της ίδιας οικογένειας. Μία εναλλακτική υπόθεση είναι ότι η αλληλεπίδραση της SAP155 με μέλη της οικογένειας Polycomb αποτελεί μέρος ενός μηχανισμού που στοχεύει στην διατήρηση των απαιτούμενων επιπέδων των μεταγράφων των γονιδίων Hox. Έτσι, τα mRNAs που έχουν μεταγραφεί έκτοπα μπορεί να δεσμεύονται από την SAP155, που είναι συνδεδεμένη με το σύμπλοκο Polycomb, ώστε να ανασταλεί η συναρμογή τους. Σε αυτήν την υπόθεση, οι πρωτεΐνες Polycomb παίζουν τον ρόλο των ανιχνευτών των επιπέδων των μεταγράφων στις περιοχές όπου τα γονίδια Hox δύνανται να μεταγραφούν (Isono K *et al*, 2005).

Η πρωτεΐνη DGCR14

Το γονίδιο *DGCR14* του ανθρώπου είχε κλωνοποιηθεί βάσει της τοπογραφίας του στη χρωμοσωμική περιοχή 22q11, που έχει συσχετιστεί με το σύνδρομο DiGeorge (Rizzu P *et al*, 1996; Gong W *et al*, 1997; OMIM #188400). Το σύνδρομο εκδηλώνεται με αναπτυξιακή ανωμαλία του εμβρυικού φαρυγγικού συστήματος, μιας παροδικής δομής από την οποία προκύπτουν διάφοροι ιστοί της κεφαλής, του λαιμού και του θώρακα. Τα κύρια κλινικά χαρακτηριστικά του συνδρόμου είναι οι καρδιαγγειακές δυσπλασίες, η υποπλασία ή απλασία του θύμου με ανεπάρκεια των κυττάρων T και ο υποπαραθυρεοειδισμός ενώ σε αρκετές περιπτώσεις παρουσιάζεται κраниοπροσωπική δυσμορφία, αναπτυξιακή και διανοητική καθυστέρηση, σχιζοφρένεια, υπερωϊοσχιστία κ.λπ.

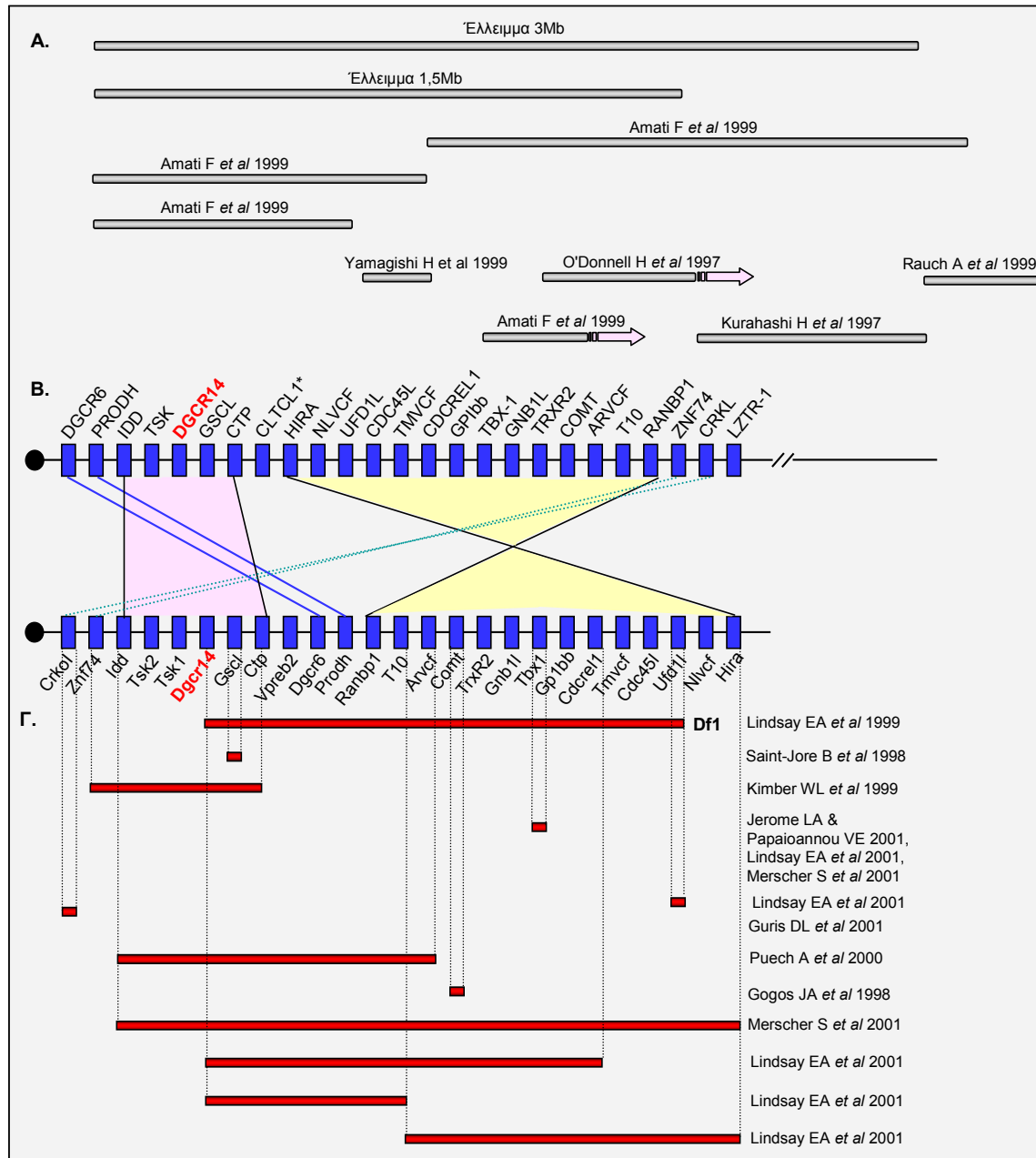


Εικ. 81: Χαρακτηριστική κраниοπροσωπική δυσμορφία σε ασθενή με σύνδρομο DiGeorge

Στην πλειονότητα των ασθενών (~85%) ο φαινότυπος σχετίζεται με ετερόζυγο, συνήθως de novo, έλλειμμα ενός τμήματος ~3Mb της περιοχής 22q11.2, στην οποία χαρτογραφούνται περίπου 30 γονίδια, συμπεριλαμβανομένου του *DGCR14* (Εικ. 82). Επιπλέον, ~8% των ασθενών, φέρει ένα μικρότερο έλλειμμα μήκους ~1,5Mb, υποπεριοχή των 3Mb με κοινό το 5' σημείο θραύσης. Το τμήμα αυτό περιλαμβάνει 24 γονίδια, συμπεριλαμβανομένου του *DGCR14* (Lindsay EA *et al*, 1995; Carlson C *et al*, 1997; Shaikh TH *et al*, 2000, Εικ. 82). Η δημιουργία των παραπάνω γενετικών ελλειμμάτων φαίνεται πως σχετίζεται με την παρουσία επαναλαμβανόμενων αλληλουχιών στα σημεία χρωμοσωμικής θραύσης αυτών των περιοχών (Halford S *et al*, 1993; Edelmann L *et al*, 1999a & 1999b), που πιθανόν οδηγούν σε γενωμική αστάθεια λόγω μη φυσιολογικού ομόλογου ανασυνδυασμού και ακολούθως χρωμοσωμικό έλλειμμα κατά τη μείωση (Saitta SC *et al*, 2004). Λόγω της χρωμοσωμικής του θέσης, το γονίδιο *DGCR14* θεωρήθηκε ως πιθανό υποψήφιο για το σύνδρομο. Ωστόσο, από πειράματα δημιουργίας knock-out ποντικών φαίνεται ότι μάλλον δεν σχετίζεται, τουλάχιστον με τα κύρια κλινικά χαρακτηριστικά του συνδρόμου. Ειδικότερα, κατά την πορεία δημιουργίας του πρώτου knock-out (*Df1*)-μοντέλου (Lindsay E *et al*, 1999; Εικ. 82), που έφερε ετερόζυγο έλλειμμα της κρίσιμης περιοχής των 1,5Mb, δημιουργήθηκε και ένα ποντίκι στο οποίο υπάρχει μόνο ένα λειτουργικό αλληλόμορφο του γονιδίου *Dgcr14*. Τα ποντίκια αυτά δεν εμφάνιζαν δυσπλασία της καρδιάς, γεγονός που δείχνει ότι η ετερόζυγη μεταλλαγή του *Dgcr14* δεν αρκεί για την πρόκληση του κύριου χαρακτηριστικού του συνδρόμου. Ωστόσο, τα ομόζυγα ποντίκια με απενεργοποιημένο το γονίδιο αυτό δεν επιβιώνουν πέραν της εμβρυικής μέρας E9,5, γεγονός που δείχνει την αναντικατάστατη λειτουργία του κατά την πρώιμη εμβρυική ανάπτυξη. Το γεγονός αυτό είναι συμβατό και με το ότι το γονίδιο είναι μοναδιαίο στο ανθρώπινο γονιδίωμα, όπως φάνηκε: α) με διερεύνηση της νουκλεοτικής βάσης δεδομένων του ανθρώπου και β) με ανάλυση Southern σε γενωμικό DNA και σε DNA μονοχρωμοσωμικών κυτταρικών υβριδίων χρησιμοποιώντας ως ανιχνευτή το cDNA (Gong W *et al*, 1997).

Μετά τη δημιουργία πολλών διαφορετικών στελεχών ποντικών τα οποία είχαν ετερόζυγα ελλείμματα τμημάτων της κρίσιμης περιοχής ή είχαν απενεργοποιημένα συγκεκριμένα γονίδια (Εικ. 82), φάνηκε ότι το κύριο υπεύθυνο γονίδιο για τον παθολογικό φαινότυπο είναι το *TBX-1*, που κωδικοποιεί έναν μεταγραφικό παράγοντα της οικογένειας των πρωτεϊνών T-box (Jerome LA & Papaioannou VE, 2001; Lindsay EA *et al*, 2001; Merscher S *et al*, 2001). Ο κρίσιμος ρόλος του *Tbx-1* στηρίζεται στον φαινότυπο των ποντικών με απενεργοποιημένο το ένα αλληλόμορφό του, τα οποία εμφανίζουν τα τρία κύρια κλινικά χαρακτηριστικά του συνδρόμου με διεισδυτικότητα 40-50% και στην ταυτοποίηση μεταλλαγών στην κωδική περιοχή του

γονιδίου σε 5 ασθενείς στους οποίους δεν ανιχνεύτηκαν καθόλου χρωμοσωμικά ελλείμματα (Yagi H *et al*, 2003).



Εικ. 82: Χρωμοσωμικά ελλείμματα στην περιοχή 22q11 σε ασθενείς με σύνδρομο DiGeorge και σε ποντίκια knockout. Φαίνονται: A) οι κρίσιμες περιοχές 3Mb και 1,5Mb καθώς και τα διαφορετικά χρωμοσωμικά ελλείμματα που έχουν ταυτοποιηθεί σε ασθενείς με σύνδρομο DiGeorge. Με ροζ απεικονίζονται ελλείμματα που δεν έχει καθοριστεί το τελομερικό τους όριο, B) τα γονίδια που χαρτογραφούνται στην κρίσιμη περιοχή των 1,5Mb στον άνθρωπο (πάνω) και η αντίστοιχη συνταϊνική περιοχή στον ποντικό (κάτω). Όλα τα γονίδια, πλην του CLTCL1 (*) έχουν ορθόλογα στον ποντικό, χωρίς όμως η διάταξή τους να είναι πλήρως συντηρημένη. Η διάταξη του γονιδίου *DGCR14* (κόκκινο) και των γειτονικών του έχει διατηρηθεί (ροζ) ενώ μία περιοχή που περιλαμβάνει 14 γονίδια έχει αναστραφεί στον ποντικό (κίτρινο), Γ) τα ελλείμματα που έχουν δημιουργηθεί σε ποντίκια knockout. Πρώτο απεικονίζεται το knockout Df1.

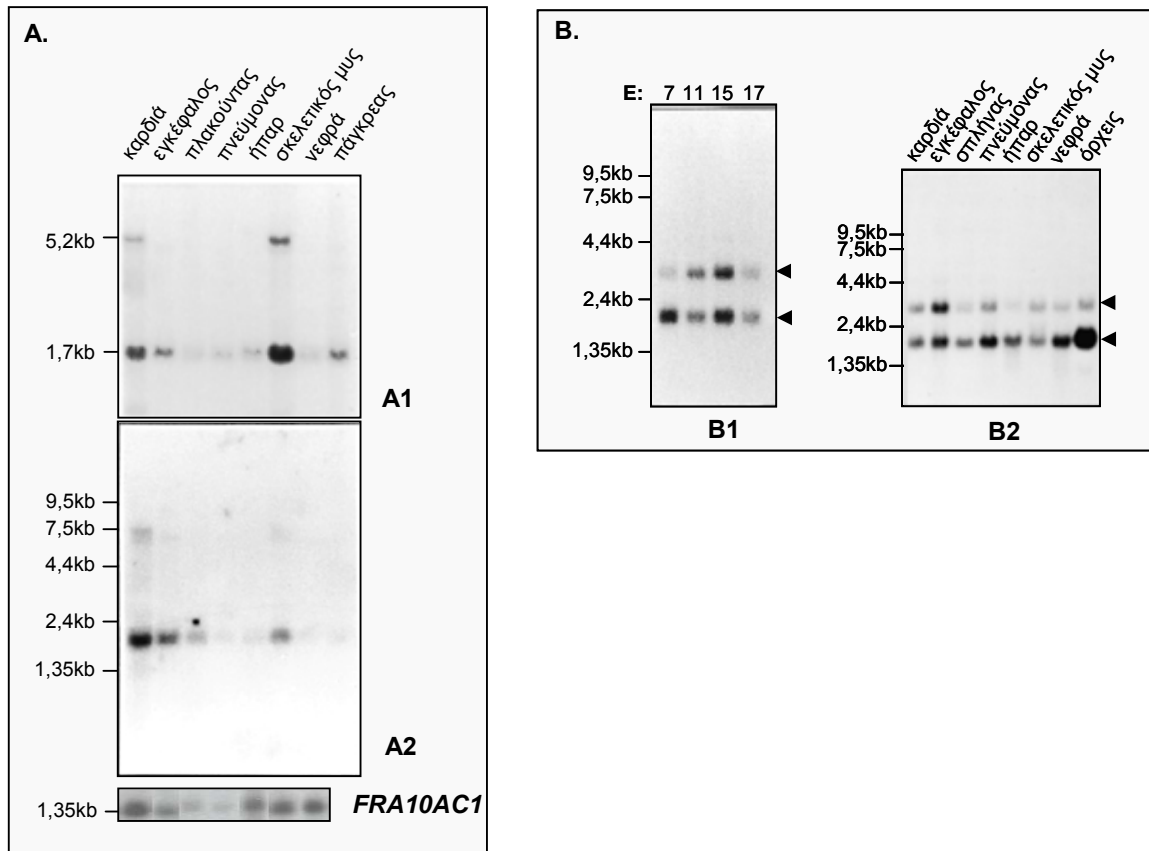
Ωστόσο υπάρχουν δεδομένα που δείχνουν ότι η μοριακή βάση του συνδρόμου είναι αρκετά πιο πολύπλοκη. Συγκεκριμένα, υπάρχουν περιπτώσεις ασθενών, που συνολικά αντιστοιχούν στο ~5% του συνόλου, στους οποίους παρατηρείται αξιοσημείωτη γενετική ετερογένεια. Έτσι, σε τέτοιους ασθενείς, ταυτοποιήθηκαν διαφορετικά χρωμοσωμικά, μη-επικαλυπτόμενα ελλείμματα στην περιοχή 22q11 (Amati F *et al*, 1999; Kurahashi H *et al*, 1997; Rauch A *et al*, 1999; Yamagishi H *et al*, 1999; O'Donnell H *et al*, 1997; Εικ. 82), κάποια από τα οποία δεν περιλαμβάνουν το γονίδιο *TBX-1*. Επίσης, έχουν αναφερθεί τουλάχιστον 200 περιπτώσεις ασθενών στους οποίους δεν ανιχνεύθηκαν καθόλου χρωμοσωμικά ελλείμματα και που δεν έχουν μεταλλάξεις στο *TBX-1* (Gong W *et al*, 2001; Lindsay EA *et al*, 2001; Chieffo C *et al*, 1997), καθώς και περιπτώσεις με πολλά κοινά κλινικά χαρακτηριστικά με αυτά του συνδρόμου, που φέρουν όμως ελλείμματα στο μικρό βραχίονα του χρωμοσώματος 10 (10p13-p14, Daw SC *et al*, 1996) ή στην περιοχή 18q21.33 (Greenberg F *et al*, 1988). Επιπλέον, ποντίκια με ομόζυγο έλλειμμα του γονιδίου *Crlk* (Guris DL *et al*, 2001), που επίσης χαρτογραφείται στην κρίσιμη περιοχή (Εικ. 82), παρουσιάζουν φαινοτυπικές ανωμαλίες που μοιάζουν με αυτές του συνδρόμου, ενώ ομόζυγα ποντίκια knockout για το γονίδιο *COMT* (Gogos JA *et al*, 1998, Εικ. 82) εμφανίζουν, μεταξύ άλλων, και διαταραχές σε θέματα συμπεριφοράς που θα μπορούσαν πιθανώς να συσχετιστούν με την σχιζοφρένεια που αποτελεί ένα από τα χαρακτηριστικά του συνδρόμου. Η ταυτοποίηση μη επικαλυπτόμενων χρωμοσωμικών ελλειμμάτων σε ασθενείς με σύνδρομο DiGeorge θα μπορούσε πιθανόν να οφείλεται σε αποτέλεσμα θέσης (position effect), όπως συμβαίνει στην περίπτωση του γονιδίου *SOX-9* (Wunderle VM *et al*, 1998; Pfeifer D *et al*, 1999; Velagaleti GV *et al*, 2005) που σχετίζεται με ένα σύνδρομο σκελετικής δυσπλασίας (OMIM #114290). Επιπλέον, η παρατηρούμενη γενετική ετερογένεια, σε συνδυασμό με την απουσία των δευτερογενών χαρακτηριστικών του συνδρόμου στα ποντίκια με απενεργοποιημένο το *Tbx-1*, ίσως εξηγείται με την συμμετοχή επιπλέον γονιδίων-τροποποιητών του φαινοτύπου. Έτσι, η έλλειψη κάποιων άλλων γονιδίων τόσο της κρίσιμης περιοχής όσο και άλλων περιοχών, θα μπορούσε να επηρεάζει το ίδιο αναπτυξιακό μονοπάτι με αποτέλεσμα ο φαινότυπος των ποντικών *Tbx-1*+/- να επιδεινώνεται ή η διεισδυτικότητα και η εκφραστικότητα να αυξάνεται. Επίσης, το ότι η ετερόζυγη έλλειψη κάποιων γονιδίων στον ποντικό δεν συνοδεύεται από τον συγκεκριμένο παθολογικό φαινότυπο, όπως στην περίπτωση του *Dgcr14*, δεν συνεπάγεται απαραίτητα ότι το ίδιο ισχύει και για τον άνθρωπο, ο οποίος ίσως έχει διαφορετική ευαισθησία στα επίπεδα διαφόρων πρωτεϊνών. Εξάλλου, για τα συγκεκριμένα ποντίκια που φέρουν ένα μόνο λειτουργικό αλληλόμορφο του *Dgcr14*, εκτός του ότι δεν εμφανίζουν τα κύρια κλινικά χαρακτηριστικά του συνδρόμου δεν έχει

ελεγχθεί ο φαινότυπος τους όσον αφορά διαταραχές στη μνήμη, τη μάθηση και την συμπεριφορά τους. Οι φαινότυποι αυτοί, εκτός του ότι συναντώνται σε κάποιο ποσοστό των ασθενών με σύνδρομο DiGeorge, θα μπορούσαν ίσως να συσχετιστούν και με τον φαινότυπο της ήπιας διανοητικής καθυστέρησης των ατόμων-φορέων της FRA10A. Δηλαδή, πιθανώς ο παρατηρούμενος φαινότυπος των φορέων να μην σχετίζεται άμεσα με την ετερόζυγη απενεργοποίηση του *FRA10AC1* αλλά με την πρωτεΐνη DGCR14 που για να λειτουργήσει σωστά θα πρέπει να αλληλεπιδράσει με την FRA10AC1.

Στα πλαίσια της λειτουργικής μελέτης του *Dgcr14* ως υποψήφιο για το σύνδρομο DiGeorge προέκυψαν πληροφορίες που αφορούν κυρίως το πρότυπο έκφρασής του. Στον άνθρωπο, το γονίδιο είναι μεταγραφικά ενεργό σε όλους τους ιστούς που εξετάστηκαν εμφανίζοντας υψηλότερη έκφραση σε καρδιά, εγκέφαλο και σκελετικό μυ, όπως δείχθηκε με ανάλυση Northern (Rizzu P *et al*, 1996; Gong W *et al*, 1997; Εικ. 83). Το πρότυπο έκφρασής του είναι παρόμοιο με αυτό του *FRA10AC1* (Εικ. 83), που αποτελεί θετική ένδειξη για την λειτουργική αλληλεπίδρασή τους. Από την ανάλυση Northern ταυτοποιήθηκε η παρουσία δύο επιπλέον, πιθανώς εναλλακτικών, μεταγράφων σε καρδιά και σκελετικό μυ, μήκους 5,2kb (Gong W *et al*, 1997) ή ~7kb (Rizzu P *et al*, 1996) των οποίων όμως δεν ταυτοποιήθηκε η φύση τους. Πιθανώς, πρόκειται για τα ίδια μετάγραφα και η διαφορά στο μέγεθός τους, μεταξύ των δύο διαφορετικών αναλύσεων, οφείλεται στις συνθήκες ηλεκτροφόρησης ή προσδιορισμού των μεγεθών. Παρόλο που η ανάλυση των ESTs του ανθρώπου δεν συνεισέφερε στην ταυτοποίηση ενός πιθανού εναλλακτικού μεταγράφου, είναι ενδιαφέρον ότι για το ορθόλογο γονίδιο του ποντικού (Gong W *et al*, 1997; Lindsay EA *et al*, 1998), ταυτοποιήθηκαν και απομονώθηκαν δύο εναλλακτικά μετάγραφα ~2kb και ~3kb που διαφέρουν στο 3'UTR τους λόγω χρήσης εναλλακτικών σημάτων πολυαδενυλίωσης (Lindsay EA *et al*, 1998). Στον ποντικό το γονίδιο είναι μεταγραφικά ενεργό από τα πρώιμα στάδια της εμβρυογένεσης, τουλάχιστον από την μέρα E7 (Εικ. 83), και στο ενήλικο εμφανίζει καθολική μεταγραφική ενεργότητα, που είναι ιδιαίτερα υψηλή στους όρχεις (Lindsay E *et al*, 1998; Εικ. 83).

Το πρότυπο έκφρασής του με RNA *in situ* υβριδοποίηση είχε μελετηθεί παλαιότερα, ωστόσο η μελέτη είχε περιοριστεί σε έμβρυα και είχε γίνει με γνώμονα το αν η παρατηρούμενη έκφραση συμβαδίζει με την θεώρησή του ως υποψήφιο για το σύνδρομο DiGeorge. Το αποτέλεσμα ήταν η ταυτοποίηση υψηλών επιπέδων έκφρασης σε εγκέφαλο και ιδιαίτερα στη γέφυρα, στα φαρυγγικά τόξα, σε θύμο και νεφρά, όργανα που σχετίζονται με τον φαινότυπο του εν λόγω συνδρόμου (Lindsay EA *et al*, 1998). Τα αποτελέσματα αυτά είναι συγκρίσιμα με την αντίστοιχη δική μας

ανάλυση η οποία επεκτάθηκε και σε εγκέφαλο ενήλικα και έγινε συγκριτικά με την ενεργότητα του *FRA10AC1*.



Εικ. 83: Πρότυπο έκφρασης του γονιδίου *DGCR14* στον άνθρωπο και τον ποντικό με ανάλυση Northern. **A.** Απεικονίζεται το αποτέλεσμα της ανάλυσης σε ιστούς ανθρώπου (A1: Gong W et al 1997, A2: Rizzu P et al 1996) και συγκριτικά η ανάλυση Northern του γονιδίου *FRA10AC1*. Και τα δύο γονίδια εκφράζονται σε όλους τους ιστούς και εμφανίζουν υψηλά επίπεδα σε καρδιά, εγκέφαλο και σκελετικό μυ. **B.** Στον ποντικό κατά την εμβρυική ανάπτυξη (B1) το γονίδιο είναι μεταγραφικά ενεργό από την ημέρα E7 και παραμένει ενεργό τις ημέρες E11, E15 και E17. Σε ενήλικο (B2), μετάγραφα ανιχνεύονται σε όλους τους ιστούς (Lindsay E et al 1998) που μελετήθηκαν. Τα βέλη δείχνουν τα δύο εναλλακτικά μετάγραφα που ταυτοποιήθηκαν.

Η πρωτεΐνη DGCR14 αποτελείται από 476 αμινοξέα, έχει εκτιμώμενο μοριακό βάρος 52,5kD και ισοηλεκτρικό σημείο 7.06. Στο C-τελικό της άκρο φέρει μία περιοχή coiled-coil, που έχει συσχετιστεί με πρωτεϊνικές αλληλεπιδράσεις και που όπως δείξαμε συμμετέχει στην αλληλεπίδραση με την *FRA10AC1*. Πιθανά ορθόλογα μόρια εντοπίστηκαν αποκλειστικά σε ευκαρυωτικούς οργανισμούς, κυρίως πολυκύτταρους όπως και στην περίπτωση της *FRA10AC1*. Από τους λίγους μονοκύτταρους που ταυτοποιήθηκαν είναι ο *Schizosaccharomyces pombe*, ενώ δεν ταυτοποιήθηκε πολυπεπίδιο με σημαντική αμινοξική ομολογία στον *Saccharomyces cerevisiae*. Η πρωτεΐνη φέρει πολλές πιθανές θέσεις φωσφορυλίωσης στο C-τελικό της άκρο. Ως πιθανές κινάσες μπορούν να θεωρηθούν: α) η MPK-1 βάσει της αλληλεπίδρασης

που ταυτοποιήθηκε μεταξύ της ορθόλογης πρωτεΐνης του *C.elegans* (F42H10.7) με την MPK-1 σε πείραμα δύο υβριδίων (mitogen-activated protein kinase, F43C1.2, Walhout AJ *et al*, 2000; Li S *et al*, 2004; <http://vidal.dfci.harvard.edu/>) και β) η κινάση cyclin-cdk2, βάσει της παρουσίας 10 διπεπτιδίων TP που βρίσκονται συγκεντρωμένα στο C-τελικό τμήμα και της ταυτόχρονης παρουσίας δύο μοτίβων RXL (όπου X ένα υδρόφοβο αμινοξύ). Τα δύο παραπάνω μοτίβα είναι χαρακτηριστικά των υποστρωμάτων των κινασών cyclin-cdk2. Επιπλέον ταυτοποιήθηκαν 3 πιθανές θέσεις σουμοϋλίωσης που πιθανώς συσχετίζονται και με την άμεση αλληλεπίδραση που ταυτοποιήθηκε μεταξύ της ορθόλογης της DGCR14 στον *Schizosaccharomyces pombe* (*Bis1*) και του *Pli1p*, μιας χαρακτηρισμένης λειτουργικά E3-SUMO λιγάσης (Xhemalce B *et al*, 2004). Αυτές οι πιθανές μετα-μεταφραστικές τροποποιήσεις θα μπορούσαν να ερμηνεύσουν το ηλεκτροφορητικό πρότυπο της ανασυνδυασμένης και της ενδογενούς πρωτεΐνης (δεν παρατίθενται αποτελέσματα) σε SDS-PAGE, όπου ανιχνεύεται σε μεγέθη μεγαλύτερα του αναμενόμενου.

Μία πολύ σημαντική πληροφορία για το βιολογικό μονοπάτι στο οποίο πιθανώς συμμετέχει η πρωτεΐνη DGCR14 προέρχεται από μελέτες βιοχημικής απομόνωσης του spliceosome. Ειδικότερα, η πρωτεΐνη DGCR14 έχει ταυτοποιηθεί σαν συστατικό του spliceosome από τρεις διαφορετικές ομάδες (Zhou Z *et al*, 2002; Rappsilber J *et al*, 2002; Makarova O *et al*, 2004). Στις δύο περιπτώσεις η ανάλυση πραγματοποιήθηκε με απομόνωση του spliceosome με χρωματογραφία συγγένειας και ταυτοποίηση των υπομονάδων του με φασματοσκοπία μάζας. Από την ανάλυση ταυτοποιήθηκαν 145 (Zhou Z *et al*, 2002) και 292 (Rappsilber J *et al*, 2002) πρωτεΐνες, συμπεριλαμβανομένης και στις δύο περιπτώσεις της DGCR14. Η συσχέτιση της DGCR14 με την διαδικασία της συναρμογής ενισχύεται από το γεγονός ότι ταυτοποιήθηκε σε spliceosome που είχε συσταθεί *in vitro* με υπόστρωμα δύο διαφορετικά πρώιμα mRNAs. Οι Makarova O *et al* (2004) ταυτοποίησαν τα συστατικά του spliceosome μέσω ανοσοκατακρήμνισης και ακολούθως με φασματοσκοπία μάζας των πρωτεϊνών. Σε αυτήν την περίπτωση χρησιμοποιήθηκαν αντισώματα έναντι συγκεκριμένων παραγόντων που συνδέονται ισχυρά αλλά παροδικά με το spliceosome, ώστε να απομονωθούν ξεχωριστά τα συστατικά καθενός σταδίου συναρμογής καθώς και του ενεργού spliceosome. Σε αυτήν την μελέτη η πρωτεΐνη DGCR14 ταυτοποιήθηκε σαν συστατικό του συμπλόκου C, ενώ δεν συμμετείχε στο σύμπλοκο B ή το B* (ενεργοποιημένο B). Φαίνεται δηλαδή ότι αποτελεί υπομονάδα του καταλυτικού συστατικού του spliceosome, χωρίς αυτό να σημαίνει ότι λαμβάνει μέρος στην κατάλυση καθώς στο σύμπλοκο αυτό συμμετέχει και πληθώρα πρωτεϊνών που σχετίζονται με άλλες διαδικασίες, όπως π.χ. οι πρωτεΐνες του EJC (exon junction complex). Εξετάζοντας τα ανωτέρω δεδομένα,

μέσα από το πρίσμα της συμμετοχής της πρωτεΐνης στη διαδικασία της συναρμογής, βλέπουμε ότι αυτά είναι συμβατά. Στην περίπτωση ενός απαραίτητου παράγοντα συναρμογής, είναι λογικό η έλλειψή του να μην είναι συμβατή με την ζωή, όπως παρατηρήθηκε στα ποντίκια που φέρουν ομόζυγη απενεργοποίηση του γονιδίου *Dgcr14*. Επίσης, συμβατή είναι η καθολική έκφραση του γονιδίου στους διάφορους ιστούς και η πυρηνική υποκυτταρική τοπολογία της πρωτεΐνης, όπως δείξαμε για την ανασυνδρασμένη DGCR14-Xpress σε κύτταρα HeLa, αποτέλεσμα που επιβεβαιώνει παλαιότερες μελέτες έκφρασης της ανασυνδρασμένης πρωτεΐνης *Dgcr14*-GFP του ποντικού σε κύτταρα NIH-3T3 (Lindsay E *et al*, 1998) αλλά και της ορθόλογης πρωτεΐνης *Bis1*-YFP σε κύτταρα του *S. pombe* (Taricani L *et al*, 2002). Επίσης, πάλι με βάση την υπόθεση της συμμετοχής της DGCR14 στην διαδικασία της συναρμογής, η παρουσία ορθόλογου μορίου στον *S. pombe* και όχι στον *S. cerevisiae* δεν προκαλεί έκπληξη, καθώς είναι γνωστό ότι ένα σημαντικό ποσοστό των γονιδίων που έχουν χαθεί ή έχουν αποκλίνει τόσο ώστε δεν είναι αναγνωρίσιμα στον *S. cerevisiae*, σε σχέση με τον *S. pombe*, αντιστοιχούν σε πρωτεΐνες που σχετίζονται με την επεξεργασία του mRNA (Aravind L *et al*, 2000), π.χ. η U2AF³⁵ (Abonich N, 1994). Αυτό πιθανώς σχετίζεται με τον κατά πολύ μικρότερο συνολικό αριθμό των ιντρονίων του *S. cerevisiae*, που εκτιμάται σε ~260 (Grate L & Ares M. Jr, 2002) σε σχέση με τον αντίστοιχο αριθμό στον *S. pombe*, που υπερβαίνει τα 4500 (Wood VR *et al*, 2002), ή με την απόκλιση που παρατηρείται στην συντηρημένη αλληλουχία της θέσης σχηματισμού θηλιάς, που στον *S. cerevisiae* είναι UACUAAC (Berglund JA, 1997) ενώ στον *S. pombe* είναι η πιο εκφυλισμένη CURAY (R=πουρίνη, Zhang MQ & Marr TG, 1994).

Τέλος, η λειτουργική σημασία της αλληλεπίδρασης των πρωτεϊνών FRA10AC1 και DGCR14 τονίζεται και επιβεβαιώνεται από το γεγονός ότι η ορθόλογη πρωτεΐνη της DGCR14 στην *D. melanogaster* (CG1474) αλληλεπιδρά με την ορθόλογη της FRA10AC1 (CG16890) βάσει πειράματος δύο υβριδίων που πραγματοποιήθηκε στα πλαίσια χαρτογράφησης των αλληλεπιδράσεων σε επίπεδο πρωτεϊνώματος (Giot L *et al*, 2003). Μάλιστα, η συγκεκριμένη αλληλεπίδραση χαρακτηρίζεται σαν υψηλής πιστότητας (Flygdid db). Δηλαδή, φαίνεται πως η αλληλεπίδραση των πρωτεϊνών FRA10AC1-DGCR14 είναι εξελικτικά συντηρημένη, γεγονός που καθιστά την *D. melanogaster* ένα εφαρμόσιμο μοντέλο ελέγχου της συμμετοχής των πρωτεϊνών αυτών στη διαδικασία της συναρμογής *in vivo*.

➤ **Η πρωτεΐνη FRA10AC1 και η συμμετοχή της στο spliceosome**

Παρόλο που η πρωτεΐνη FRA10AC1 αλληλεπιδρά άμεσα με τις SAP145 και DGCR14 και μάλιστα με ιδιαίτερα μεγάλη συγγένεια με την δεύτερη, δεν έχει απομονωθεί βιοχημικά, τουλάχιστον μέχρι τώρα, ως συστατικό του spliceosome. Η απουσία της FRA10AC1 από το απομονωμένο σύμπλοκο δεν σημαίνει απαραίτητα ότι δεν ενέχεται λειτουργικά σε κάποιο στάδιο της διαδικασίας της συναρμογής. Έτσι, με βάση διάφορα στοιχεία, που αναλύονται εκτενέστερα στη συνέχεια, μπορούμε να υποθέσουμε ότι η FRA10AC1: α) αποτελεί μιν συστατικό του spliceosome συνδέεται όμως ασθενώς με αυτό, β) έχει εξειδικευμένη δράση στη συναρμογή υποσυνόλου των mRNAs ή στην εναλλακτική συναρμογή, γ) λειτουργεί ως συζευκτικός παράγοντας της συναρμογής με την μεταγραφή ή με την έξοδο των ώριμων mRNAs από τον πυρήνα.

Ασθενής σύνδεση στο spliceosome

Σύμφωνα με την πρώτη υπόθεσή μας υπάρχει περίπτωση, κατά τη διάρκεια της βιοχημικής απομόνωσης του spliceosome, λόγω ασθενούς σύνδεσής της FRA10AC1, να αποδεσμεύεται αυτή από το σύμπλοκο. Αυτό φαίνεται να ισχύει στην περίπτωση της πρωτεΐνης Prp18. Η Prp18 ενώ είναι ένας απαραίτητος παράγοντας του δεύτερου βήματος της συναρμογής όπως δείχθηκε κατά την *in vitro* αντίδραση με εκχυλίσματα κυττάρων HeLa από τα οποία είχε αφαιρεθεί ειδικά η πρωτεΐνη, δεν ανιχνεύτηκε στο απομονωμένο σύμπλοκο C του spliceosome ούτε με χρωματογραφία μάζας ούτε με ανάλυση Western (Horowitz DS & Krainer AR, 1997; Jurica MS *et al*, 2002).

Εναλλακτικά, θα μπορούσε η σύνδεση της FRA10AC1 με τις SAP145 και DGCR14 να εξαρτάται σε μεγάλο βαθμό από φωσφορυλιώσεις των μορίων και καθώς δεν χρησιμοποιούνται αναστολείς φωσφατασών κατά την διαδικασία της απομόνωσης του spliceosome, η FRA10AC1 να αποδεσμεύεται από το σύμπλοκο. Η υπόθεση αυτή βασίζεται στην ταυτοποίηση πολλαπλών πιθανών θέσεων φωσφορυλίωσης, τόσο στην FRA10AC1 όσο και στις πρωτεΐνες DGCR14 και SAP145 καθώς και λόγω της παρουσίας της τελευταίας στο ίδιο πρωτεϊνικό σύμπλοκο με την φωσφατάση PP1 σε πυρηνικά εκχυλίσματα ηπατικών κυττάρων (Tran HT *et al*, 2004). Είναι χαρακτηριστικό το παράδειγμα της πρωτεΐνης NIPP1 που, ενώ αποτελεί απαραίτητο παράγοντα συναρμογής και αλληλεπιδρά ειδικά με την φωσφορυλιωμένη SAP155, υπομονάδα του SF3b, δεν έχει απομονωθεί βιοχημικά από παρασκευάσματα του spliceosome, λόγω πιθανών αποφωσφορυλιώσεων κατά την απομόνωσή τους (Boudrez A *et al*, 2002).

Εξειδικευμένη δράση

Εναλλακτικά, η απουσία της FRA10AC1 από το απομονωμένο spliceosome θα μπορούσε να ερμηνευθεί βάσει της πιθανής συμμετοχής της πρωτεΐνης στη συναρμογή συγκεκριμένων μόνο πρόδρομων mRNAs. Εναλλακτικά, θα μπορούσε να συμμετέχει σε spliceosomes εναλλακτικής συναρμογής και όχι στο «συστατικό» spliceosome, βάσει της υπόθεσης ότι τα mRNAs που υφίστανται εναλλακτική συναρμογή έχουν διαφορετικές ή επιπρόσθετες ανάγκες για ρυθμιστικούς παράγοντες συναρμογής.

Υπάρχουν ενδείξεις για διαφορεικές απαιτήσεις των πρόδρομων mRNAs για παράγοντες συναρμογής *in vivo*. Πιθανώς αυτές οι ανάγκες εξαρτώνται από τη δομή των ιντρονίων των γονιδίων. Οι ενδείξεις αυτές προέρχονται από γονιδιωματικές μελέτες στον σακχαρομύκητα, όπου συσχετίζεται η ρύθμιση της συναρμογής συγκεκριμένων γονιδίων με διαφορετικούς παράγοντες, με τη χρήση κατάλληλων μικροσυστοιχιών που ανιχνεύουν την αποτελεσματικότητα της συναρμογής (Clark TA *et al*, 2002). Η μελέτη αυτή πραγματοποιήθηκε στον σακχαρομύκητα επειδή ο χαρακτηρισμός (annotation) του γονιδιώματός του και ιδιαίτερα των λιγοστών ιντρονίων του είναι πολύ ακριβής. Στο σακχαρομύκητα, τα γονίδια που περιέχουν ιντρόνια είναι λιγότερα από 250, με μόλις περίπου 10 από αυτά, να περιέχουν περισσότερα από ένα ιντρόνια. Ειδικότερα, ελέγχθηκε η αποτελεσματικότητα της συναρμογής όλων των γονιδίων που περιέχουν ιντρόνια, απουσία καθενός από 18 παράγοντες συναρμογής που δεν είναι απαραίτητοι για την βιωσιμότητα του οργανισμού. Συνοπτικά, εντοπίστηκε μία ομάδα γονιδίων που επηρεάζονται από την έλλειψη όλων σχεδόν των παραγόντων, λίγα σχετικά γονίδια που εξαρτώνται από δύο μόνο παράγοντες (Ppr17, Ppr18) και μία ομάδα γονιδίων που ενώ δεν επηρεάζεται από τους παραπάνω παράγοντες, εξαρτάται από τρεις άλλους (Snu66, Brr1, Ms11). Επιπλέον, φάνηκε ότι τα ιντρόνια που επηρεάζονται από τον παράγοντα συναρμογής Ppr17 έχουν στην πλειονότητά τους μέγεθος >200bp ενώ αντίθετα, δεν επηρεάζονται ιντρόνια στα οποία η απόσταση μεταξύ της 3' θέσης συναρμογής και της θέσης σχηματισμού θηλιάς είναι μικρότερη από 14 νουκλεοτίδια (Sapra AK *et al*, 2004). Μέχρι τώρα, δεν υπάρχουν παρόμοιες μελέτες σε επίπεδο γονιδιωμάτων άλλων οργανισμών ούτε πληροφορίες από τα *in vitro* συστήματα. Έτσι, κατά την *in vitro* ανασύσταση του spliceosome με διαφορετικά πρόδρομα mRNAs-υποστρώματα, είναι πιθανόν να υπάρχουν μεν κοινές πρωτεΐνες που αποτελούν τον καταλυτικό πυρήνα του συμπλόκου, όμως παράλληλα, να περιέχεται διαφορετικός συνδυασμός βοηθητικών παραγόντων. Το θέμα αυτό δεν έχει μελετηθεί γιατί μόνο ένας περιορισμένος αριθμός πρόδρομων mRNAs λειτουργεί ικανοποιητικά σε πειράματα συναρμογής *in vitro*. Έτσι, οι περισσότερες ερευνητικές ομάδες που πραγματοποιούν

in vitro ανασύσταση του spliceosome ή επιμέρους συμπλόκων του, χρησιμοποιούν παράγωγα του ιντρονίου 1 του AdML (Adenovirus Major Late). Με βάση τα παραπάνω, μπορούμε να θεωρήσουμε την FRA10AC1 ως πιθανό ρυθμιστή της συναρμογής συγκεκριμένων μεταγράφων, θεωρία που όμως χρειάζεται να ελεγχθεί με πειραματική προσέγγιση ανάλογη με αυτήν των Clark TA *et al* (2002).

Συσχέτιση με εναλλακτική συναρμογή

Στα ίδια πλαίσια, η FRA10AC1 θα μπορούσε να ρυθμίζει την εναλλακτική συναρμογή, μηχανισμό που πλέον φαίνεται ότι αποτελεί τον κανόνα στα γονίδια του ανθρώπου, συμβάλλοντας στην πολυπλοκότητα του πρωτεϊνώματός του (Modrek B & Lee C, 2002; Johnson JM *et al*, 2003; Kamra D *et al*, 2004). Παρόλο που η ρύθμιση της εναλλακτικής συναρμογής δεν έχει ακόμη μελετηθεί εκτενώς, φαίνεται ότι οι βασικοί συντελεστές της είναι trans-ρυθμιστικοί παράγοντες και cis-στοιχεία των μεταγράφων στα οποία αυτοί δρουν. Τα εναλλακτικά εξόνια συνήθως είναι μικρά και έχουν μη συντηρημένες θέσεις συναρμογής. Ωστόσο, φαίνεται ότι, συνήθως, έχουν βοηθητικές cis-αλληλουχίες που βελτιστοποιούν την αναγνώριση των θέσεων συναρμογής και ρυθμίζουν την επιλογή τους. Αυτά τα cis-στοιχεία είναι μικρά σε μήκος, εκφυλισμένα και υπάρχουν σε πολλαπλά αντίγραφα καθώς όταν είναι μεμονωμένα, η δράση τους είναι ασθενής. Συνήθως, είναι συντηρημένα εξελικτικά και πιθανώς παρόμοια μεταξύ γονιδίων που ρυθμίζονται με παραπλήσιο τρόπο. Χαρτογραφούνται τόσο σε εξόνια όσο και σε ιντρόνια. Βρίσκονται ανοδικά, καθοδικά ή και εκατέρωθεν του εναλλακτικού εξονίου, συνήθως σε μικρή απόσταση από αυτό. Τέλος, τα cis-στοιχεία μπορεί να υποβοηθούν ή να καταστέλλουν την επιλογή συγκεκριμένων θέσεων συναρμογής. Με βάση την χαρτογράφησή τους και το αποτέλεσμα που έχουν, ονομάζονται εξονιακοί / ιντρονιακοί ενισχυτές / καταστολείς της συναρμογής (ESE, ISE, ESS, ISS), αντίστοιχα (Ladd AN & Cooper TA, 2002; Fairbrother WG *et al*, 2002).

Όσον αφορά τους trans-παράγοντες, μέχρι πρόσφατα, ως ρυθμιστές της εναλλακτικής συναρμογής είχαν χαρακτηριστεί αποκλειστικά μέλη των οικογενειών των πρωτεϊνών SR (serine and arginine-rich proteins) και hnRNP (heterogenous nuclear ribonucleoproteins). Σύμφωνα με το γενικό μοντέλο δράσης τους, οι πρωτεΐνες SR συνδέονται με αλληλουχίες-ενισχυτές της συναρμογής υποβοηθώντας έτσι τον σχηματισμό συμπλόκων που διευκολύνουν την αναγνώριση των εξονίων και την επιλογή των θέσεων συναρμογής (Bourgeois CF *et al*, 2004; Zheng ZM, 2004). Αντίθετα, οι πρωτεΐνες hnRNP θεωρούνται γενικά ως καταστολείς δρώντας ανταγωνιστικά στις πρωτεΐνες SR ή / και σε συστατικούς παράγοντες συναρμογής (Chabot B *et al*, 2003; Eperon IC *et al*, 2000). Η τρέχουσα θεωρία για την ρύθμιση της εναλλακτικής συναρμογής μέσω των παραπάνω παραγόντων είναι ότι αυτή

πραγματοποιείται μέσω του ανταγωνισμού τους για σύνδεση σε ρυθμιστικές αλληλουχίες ή/και για αλληλεπίδραση με συστατικούς παράγοντες. Έτσι, γεγονότα όπως η συμμετοχή ή όχι ενός εξονίου στο ώριμο mRNA, εξαρτάται από την ισχύ των αλληλεπιδράσεων μεταξύ των πρωτεϊνών SR και των αλληλουχιών-στόχων καθώς και των συστατικών παραγόντων συναρμογής. Ωστόσο, πρόσφατα ταυτοποιήθηκαν και πρωτεΐνες-ρυθμιστές της εναλλακτικής συναρμογής που δεν ανήκουν στις παραπάνω οικογένειες. Ειδικότερα, ελέγχθηκε η εναλλακτική συναρμογή τριών γονιδίων, που δεν σχετίζονται λειτουργικά, τα οποία παράγουν συνολικά 19 εναλλακτικά εξόνια, απουσία 250 πρωτεϊνών-πιθανών ρυθμιστών της διαδικασίας. Οι πρωτεΐνες αυτές επιλέχθηκαν ώστε να έχουν μοτίβο σύνδεσης για RNA ή επειδή αποτελούν συστατικά του spliceosome της *D. melanogaster*. Ο έλεγχος έγινε με RT-PCR σε κύτταρα S2 μετά από RNAi για κάθε παράγοντα (Park JW *et al*, 2004). Έτσι ταυτοποιήθηκαν 47 πρωτεΐνες που επηρεάζουν την εναλλακτική συναρμογή τουλάχιστον ενός εξονίου ενώ σε καμία περίπτωση δεν επηρεάστηκε η συναρμογή του GAPDH mRNA, αποτέλεσμα που δείχνει ότι η έλλειψη αυτών των πρωτεϊνών δεν επηρεάζει τη γενική-συστατική συναρμογή. Ιδιαίτερα ενδιαφέρον είναι το αποτέλεσμα ότι μεταξύ των πρωτεϊνών αυτών περιλαμβάνονται και μόρια που δεν έχουν μοτίβο σύνδεσης με RNA καθώς και πρωτεΐνες που αποτελούν βασικούς παράγοντες του spliceosome, όπως συστατικά του U1 snRNP, οι δύο υπομονάδες του παράγοντα U2AF, ένα συστατικό του U2 snRNP και μία πρωτεΐνη του U4/U6 snRNP. Οι παραπάνω γενικοί παράγοντες δεν επηρεάζουν ταυτόχρονα και τα τρία γονίδια που ελέγχθηκαν, ούτε επηρεάζουν την συναρμογή του GAPDH mRNA, γεγονός που δείχνει ότι η μείωση των επιπέδων τους, μέσω RNAi, ρυθμίζει εξειδικευμένα την εναλλακτική συναρμογή συγκεκριμένων εξονίων. Αυτές οι διαφορές στις σχετικές συγκεντρώσεις των συστατικών του spliceosome μπορεί να αλλάζουν την κινητική της ανασύστασης του συμπλόκου ή την επιλογή των θέσεων συναρμογής. Έτσι οι διακυμάνσεις στις συγκεντρώσεις ή στην ενεργότητα των βασικών παραγόντων του spliceosome μπορεί να αποτελεί έναν κοινό μηχανισμό ρύθμισης της συναρμογής. Στην παραπάνω μελέτη δεν συμπεριλήφθηκε η πρωτεΐνη SAP145, παρόλο ότι αποτελεί βασικό παράγοντα συναρμογής ούτε η DGCR14 ως συστατικό του spliceosome.

Ωστόσο και με βάση ότι: 1) πρωτεΐνες χωρίς μοτίβα σύνδεσης στο RNA μπορούν να ρυθμίζουν την εναλλακτική συναρμογή δρώντας, προφανώς, έμμεσα και 2) δεν ανιχνεύονται ορθόλογα μόρια των πρωτεϊνών FRA10AC1 και DGCR14 στον σακχαρομύκητα στον οποίο η εναλλακτική συναρμογή είναι ένα πολύ σπάνιο γεγονός, μπορούμε να διατυπώσουμε την υπόθεση ότι η FRA10AC1 και η DGCR14 συμμετέχουν στην εναλλακτική συναρμογή συγκεκριμένων μεταγράφων.

Η πρωτεΐνη FRA10AC1 ως πιθανός παράγοντας σύζευξης της συναρμογής με άλλες διαδικασίες

Από μία άλλη οπτική γωνία, η FRA10AC1 θα μπορούσε να μην αποτελεί συστατικό του καταλυτικού spliceosome αλλά να συμμετέχει στη σύζευξη της συναρμογής με άλλες διαδικασίες επεξεργασίας του mRNA, όπως η μεταγραφή ή η έξοδος του ώριμου mRNA από τον πυρήνα.

ο *Συσχέτιση με μεταγραφή*

Για την υπόθεση της συμμετοχής της πρωτεΐνης FRA10AC1 στην σύζευξη της συναρμογής με την μεταγραφή, συνάδουν: 1) η παρατήρηση της αυτόνομης ενεργοποίησης των γονιδίων αναφοράς του σακχαρομύκητα, κατά το πείραμα δύο υβριδίων με την πρωτεΐνη FRA10AC1 (βλ. Αποτελέσματα). Το φαινόμενο της αυτόνομης ενεργοποίησης παρατηρείται συχνά σε μόρια που φυσιολογικά λειτουργούν σαν μεταγραφικοί ενεργοποιητές διατηρώντας την ιδιότητα αυτήν και στον σακχαρομύκητα (π.χ. Du W *et al*, 1996) και 2) Η ταυτοποίηση άμεσης αλληλεπίδρασης της πρωτεΐνης DGCR14 της *D. melanogaster* και των πρωτεϊνών Ada2b και Bar60, συστατικών των συμπλόκων αναδιαμόρφωσης της χρωματίνης SAGA και Brahma αντίστοιχα, καθώς και τριών πρωτεϊνών που φέρουν χαρακτηριστικά μοτίβα μεταγραφικών παραγόντων (homeobox και helix-loop-helix, Flygrid db).

Επεκτείνοντας τους παραπάνω συλλογισμούς, μπορούμε να υποθέσουμε ότι η FRA10AC1 ενέχεται στην διαδικασία της μεταγραφής μέσω της αλληλεπίδρασής της με την πρωτεΐνη DGCR14 και συμμετέχει στην στρατολόγηση του spliceosome στο πρόδρομο mRNA. Πλέον, υπάρχουν πολλά δεδομένα που υποστηρίζουν την λειτουργική σύνδεση της μεταγραφής με την συναρμογή. Έτσι, εκτός από την μετα-μεταγραφική συναρμογή του πρόδρομου mRNA, σε πολλές περιπτώσεις η συναρμογή πραγματοποιείται ενώ ακόμα η μεταγραφή είναι σε εξέλιξη (Beyer AL & Osheim YN, 1988). Ωστόσο, για την πλειονότητα των ιντρονίων δεν είναι γνωστό ποιο μοντέλο ακολουθούν ούτε εάν ένα συγκεκριμένο ιντρόνιο ακολουθεί πάντα το ίδιο πρότυπο. Σε συγκεκριμένες περιπτώσεις φαίνεται ότι το πρότυπο απομάκρυνσης ενός ιντρονίου από το πρόδρομο mRNA, εξαρτάται από τη θέση του. Έτσι, για παράδειγμα, στο γονίδιο BR1 (Balbiani ring 1), το ιντρόνιο 3 που βρίσκεται 3kb από το 5' άκρο του πρόδρομου mRNA (~40kb), απομακρύνεται κατά τη διάρκεια της μεταγραφής, ενώ το ιντρόνιο 4 που χαρτογραφείται 0,6kb ανοδικά της θέσεως πολυαδενυλίωσης, απομακρύνεται, στην πλειονότητα των μορίων, μετα-μεταγραφικά (Bauren G & Wieslander L, 1994). Η σύζευξη των δύο διαδικασιών βοηθάει στην εστίαση της πολυπρωτεϊνικής μηχανής της συναρμογής στην περιοχή που βρίσκεται το πρόδρομο mRNA κατά την παραγωγή του, διευκολύνοντας πιθανώς την

αναγνώριση των εξονίων, με αποτέλεσμα την μείωση της κινητικής της αλληλεπίδρασης μεταξύ του spliceosome και του πρόδρομου mRNA. Η φυσική σύνδεση των δύο διαδικασιών δείχθηκε αρχικά με συνανοσοκατακρήμνιση της RNA Pol II, των snRNPs και των ενδιάμεσων μορίων mRNA της συναρμογής (Chabot B *et al*, 1995; Vincent M *et al*, 1996). Η επιβεβαίωση της λειτουργικής συσχέτισής τους προέρχεται από επιπλέον δεδομένα που στηρίζονται σε *in vitro* και *in vivo* βιοχημικά, κυτταρολογικά και λειτουργικά πειράματα. Σημαντικό ρόλο στη σύζευξη διαδραματίζει η καρβοξυτελική περιοχή της μεγαλύτερης υπομονάδας της Pol II (CTD), λόγω της δυνατότητάς της να συνδέεται ή να μεσολαβεί στη σύνδεση διαφόρων ρυθμιστικών παραγόντων. Η περιοχή αυτή περιλαμβάνει 52 ή 26 επαναλήψεις ενός επταπεπτιδίου στα θηλαστικά και τον σακχαρομύκητα αντίστοιχα και υφίσταται πολλαπλές φωσφορυλιώσεις. Το πρότυπο αυτών των φωσφορυλιώσεων αλλάζει κατά την εξέλιξη της μεταγραφής και φαίνεται πως έχει ρυθμιστικό ρόλο στην επεξεργασία του mRNA, συμπεριλαμβανομένης και της συναρμογής. Έτσι, σε πειράματα *in vitro* και *in vivo*, η προσθήκη της φωσφορυλιωμένης Pol II ενισχύει σημαντικά την αποτελεσματικότητα της συναρμογής (Hirose Y *et al*, 1999; Corden JL & Patturajan M, 1997) ενώ συγκεκριμένα ελλείμματά της περιοχής CTD έχουν το αντίθετο αποτέλεσμα (McCracken S *et al*, 1997). Επιπλέον, η Pol II αλληλεπιδρά άμεσα με διάφορους παράγοντες συναρμογής, π.χ. τον Prp40 (Morris DP & Greenleaf AL, 2000). Η συσχέτιση των δύο διαδικασιών φάνηκε και από την σύσταση του απομονωμένου spliceosome, στο οποίο συμμετέχουν και μεταγραφικοί παράγοντες π.χ. ο μεταγραφικός συμπάραγοντας TAT-SF1 και οι μεταγραφικοί παράγοντες CA150, XAB2, and SKIP. Ειδικότερα, υπάρχουν σημαντικές ενδείξεις για την συσχέτιση της συναρμογής με την έναρξη και την επιμήκυνση της μεταγραφής. Έτσι, ο TFIIH, ένα πολυπρωτεϊνικό σύμπλοκο που σχετίζεται με την έναρξη της μεταγραφής αλληλεπιδρά ειδικά με το U1snRNA (Kwek KY *et al* 2002). Επίσης, ταυτοποιήθηκαν αλληλεπιδράσεις μεταξύ του συμπλόκου έναρξης της μεταγραφής και πρωτεϊνών SR ενώ υπάρχουν παραδείγματα μεταγραφικών παραγόντων που στρατολογούν πρωτεΐνες SR στην πρωτεϊνική μηχανή της μεταγραφής (Ge H *et al*, 1998). Επιπλέον, ταυτοποιήθηκαν αλληλεπιδράσεις μεταξύ του παράγοντα επιμήκυνσης της μεταγραφής TAT-SF1 και παραγόντων συναρμογής και δείχθηκε ότι, *in vitro*, η μηχανή της συναρμογής που στρατολογείται από τον TAT-SF1 ενεργοποιεί ισχυρά την επιμήκυνση της μεταγραφής (Fong YW & Zhou Q, 2001). Επίσης, ένας άλλος παράγοντας επιμήκυνσης της μεταγραφής, ο TFIIIS, ταυτοποιήθηκε σε πολυπρωτεϊνικό σύμπλοκο που εκτός από την RNA pol II περιέχει και διάφορους παράγοντες συναρμογής (Robert F *et al*, 2002). Τέλος, σε πειράματα βιοχημικής απομόνωσής του πολυπρωτεϊνικού συμπλόκου STAGA του ανθρώπου,

που είναι το λειτουργικό ομόλογο του συμπλόκου SAGA του σακχαρομύκητα και δρα ως ενεργοποιητής της μεταγραφής μέσω ακετυλίωσης των ιστονών, ταυτοποιήθηκε μεταξύ των άλλων υπομονάδων, η πρωτεΐνη SAP130, που αποτελεί υπομονάδα του παράγοντα συναρμογής SF3b (Martinez E *et al*, 2001). Πιθανώς, η SAP130 λειτουργεί σαν πρωτεΐνη-συνδέτης που βοηθάει στη στρατολόγηση του spliceosome σε μεταγραφικά ενεργά γονίδια, υπόθεση που ενισχύεται από το γεγονός ότι είναι η μόνη υπομονάδα του SF3b που δεν συνδέεται με το πρόδρομο mRNA μετά από επίδραση υπερϊώδους ακτινοβολίας (Martinez E *et al*, 2001).

Έτσι, η μεταγραφική ενεργότητα που εμφανίζει η FRA10AC1 στον σακχαρομύκητα και εφόσον αυτή επιβεβαιωθεί και σε άλλα συστήματα, θα μπορούσε να στηρίξει την ενδεχόμενη συμμετοχή της πρωτεΐνης στην σύζευξη της μεταγραφής με την συναρμογή. Πιθανώς και η πρωτεΐνη DGCR14 παίζει παρόμοιο συζευκτικό ρόλο, καθώς αλληλεπιδρά με διάφορους παράγοντες που ενέχονται στην διαδικασία της μεταγραφής.

ο *Συσχέτιση με μεταφορά του mRNA στο κυτταρόπλασμα*

Η υπόθεση της συμμετοχής της FRA10AC1 στη λειτουργική συσχέτιση της συναρμογής με την έξοδο από τον πυρήνα είναι πιθανή λόγω της ισχυρής της σύνδεσης με την πρωτεΐνη DGCR14 και λόγω της παρουσίας της τελευταίας στα τελικά στάδια της συναρμογής (συμπλόκο C), (Makarova O *et al*, 2004). Ένα άλλο στοιχείο που συνηγορεί υπέρ αυτής της υπόθεσης είναι οι πρωτεϊνικές αλληλεπιδράσεις που ταυτοποιήθηκαν σε οργανισμούς-μοντέλα. Ειδικότερα, η Bis1, η ορθόλογη πρωτεΐνη της DGCR14 στον *S. rombe* αλληλεπιδρά άμεσα με τις SPCC162.08C και SPBC14F5.03c οι οποίες φέρουν χαρακτηριστικά μοτίβα της οικογένειας των καρυοφερινών και παρουσιάζουν σημαντική ομοιότητα καθ' όλο το μήκος τους με μέλη της εν λόγω οικογένειας (Taricani L *et al*, 2002). Επιπλέον, στα πλαίσια γονιδιωματικής ανάλυσης των πρωτεϊνικών αλληλεπιδράσεων στην *D. melanogaster* ταυτοποιήθηκε η άμεση αλληλεπίδραση της ορθόλογης της DGCR14 (CG1474) με την νουκλεοπορίνη Nup170 (Flygrid db, Lutzmann M *et al*, 2005). Οι παραπάνω αλληλεπιδράσεις της DGCR14 πιθανώς σχετίζονται με την πυρηνοπλάσματική μεταφορά των mRNAs καθώς έχει αποδειχθεί η συμμετοχή των καρυοφερινών και συγκεκριμένα της Kap2b, στην έξοδο των ώριμων mRNAs από τον πυρήνα (Shamsher MK *et al*, 2002).

4.4. Προοπτικές

Η ολοκλήρωση της αλληλούχησης του γονιδιώματος του ανθρώπου αποτελεί το πρώτο βήμα στην πορεία για την κατανόηση των γενετικών πληροφοριών και της βιολογίας του οργανισμού. Παρόλο που ο πλήρης κατάλογος των γονιδίων του

ανθρώπου δεν έχει τελειοποιηθεί, είναι προφανής η δραματική συνεισφορά της διαθέσιμης αλληλουχίας στην επιτάχυνση της κλωνοποίησης γονιδίων που σχετίζονται με νοσήματα βάσει την χρωμοσωμική τοπογραφία τους. Μέχρι σήμερα, η πλειονότητα αυτών των γονιδίων αφορά μεντελικά κληρονομήσιμα νοσήματα (1700 γονίδια) που είναι σπάνια στον πληθυσμό (1:10.000-1:100.000). Ωστόσο, πιστεύεται ότι στο μέλλον θα σημειωθεί σημαντική πρόοδος στην εξακρίβωση των γενετικών παραμέτρων και στα πολύ πιο συχνά πολυπαραγοντικά νοσήματα. Σε αυτά τα πλαίσια, ξεκίνησε το 2002 και η πρώτη φάση ολοκληρώθηκε μόλις πρόσφατα (Οκτώβριος 2005), το Πρόγραμμα Χαρτογράφησης Απλοτύπων (International HarMap Project) που δημιούργησε τον κατάλογο της γενετικής ποικιλότητας στο γονιδίωμα, καταγράφοντας περισσότερους από ένα εκατομμύριο πολυμορφισμούς ενός νουκλεοτιδίου (SNPs). Αυτοί είναι ιδιαίτερα χρήσιμοι σε μελέτες γενετικές συσχέτισης των πολυπαραγοντικών νοσημάτων, κατά τις οποίες ελέγχεται κατά πόσο ένα SNP υπεραντιπροσωπεύεται σε ένα δείγμα ασθενών σε σχέση με κατάλληλα επιλεγμένα άτομα ελέγχου.

Στην παρούσα διατριβή, αξιοποιήσαμε το φυσικό χάρτη και τη διαθέσιμη αλληλουχία του χρωμοσώματος 10 και κλωνοποιήσαμε βάσει της τοπογραφίας του το γονίδιο *FRA10AC1* που συνδέεται με την εύθραυστη χρωμοσωμική θέση FRA10A. Η εμφάνιση της εύθραυστης αυτής θέσης έχει συσχετιστεί με την εκδήλωση ήπιας μορφής διανοητικής καθυστέρησης. Ωστόσο, η συσχέτιση αυτή δεν έχει επαληθευτεί καθώς δεν έχει γίνει έλεγχος για ταυτοποίηση μεταλλαγών σε άτομα με διανοητική καθυστέρηση και οι μελέτες δεν περιλαμβάνουν μεγάλο αριθμό δειγμάτων. Για να εξακριβώσουμε την πιθανή συσχέτιση του γονιδίου με τον παθολογικό φαινότυπο, είναι ήδη σε εξέλιξη από την ομάδα μας η δημιουργία ποντικών knockout τα οποία θα ελεγχθούν σε διάφορες δοκιμές συμπεριφοράς, μάθησης και μνήμης.

Στα πλαίσια της διερεύνησης του βιοχημικού ρόλου της πρωτεΐνης *FRA10AC1* και για να ελέγξουμε την πιθανή συσχέτισή της με τη διαδικασία της συναρμογής του mRNA είναι απαραίτητα περαιτέρω βιοχημικά πειράματα, όπου θα ελεγχθεί: α) η *in vivo* αλληλεπίδραση της πρωτεΐνης *FRA10AC1* με τα snRNPs και πιθανόν με επιπλέον πρωτεΐνες, β) η πιθανή σύνδεση της *FRA10AC1* με κάποιο/α από τα σύμπλοκα του spliceosome, γ) η παρεμπόδιση της πορείας της συναρμογής *in vitro* από τα ειδικά αντισώματα της *FRA10AC1* και ε) η αναγκαιότητα της *FRA10AC1* στην συναρμογή *in vitro* ή η επίδρασή της στην κινητική της αντίδρασης.

ΒΙΒΛΙΟΓΡΑΦΙΑ

1. Abou-Khalil B, Ge Q, Desai R, Ryther R, Bazyk A, Bailey R, Haines JL, Sutcliffe JS, George AL Jr (2001). Partial and generalized epilepsy with febrile seizures plus and a novel SCN1A mutation. *Neurology*, 57, (12): 2265-72.
2. Abovich N, Liao XC, Rosbash M (1994). The yeast MUD2 protein: an interaction with PRP11 defines a bridge between commitment complexes and U2 snRNP addition. *Genes Dev.*, 8, (7): 843-54.
3. Adams MD, Celniker SE, Holt RA, Evans CA, Gocayne JD, Amanatides PG, Scherer SE, Li PW, Hoskins RA, Galle RF, George RA, Lewis SE, Richards S, Ashburner M, Henderson SN, Sutton GG, Wortman JR, Yandell MD, Zhang Q, Chen LX, Brandon RC, Rogers YH, Blazej RG, Champe M, Pfeiffer BD, Wan KH, Doyle C, Baxter EG, Helt G, Nelson CR, Gabor GL, Abril JF, Agbayani A, An HJ, Andrews-Pfannkoch C, Baldwin D, Ballew RM, Basu A, Baxendale J, Bayraktaroglu L, Beasley EM, Beeson KY, Benos PV, Berman BP, Bhandari D, Bolshakov S, Borkova D, Botchan MR, Bouck J, Brokstein P, Brottier P, Burtis KC, Busam DA, Butler H, Cadieu E, Center A, Chandra I, Cherry JM, Cawley S, Dahlke C, Davenport LB, Davies P, de Pablos B, Delcher A, Deng Z, Mays AD, Dew I, Dietz SM, Dodson K, Doup LE, Downes M, Dugan-Rocha S, Dunkov BC, Dunn P, Durbin KJ, Evangelista CC, Ferraz C, Ferriera S, Fleischmann W, Fosler C, Gabrielian AE, Garg NS, Gelbart WM, Glasser K, Glodek A, Gong F, Gorrell JH, Gu Z, Guan P, Harris M, Harris NL, Harvey D, Heiman TJ, Hernandez JR, Houck J, Hostin D, Houston KA, Howland TJ, Wei MH, Ibegwam C, Jalali M, Kalush F, Karpen GH, Ke Z, Kennison JA, Ketchum KA, Kimmel BE, Kodira CD, Kraft C, Kravitz S, Kulp D, Lai Z, Lasko P, Lei Y, Levitsky AA, Li J, Li Z, Liang Y, Lin X, Liu X, Mattei B, McIntosh TC, McLeod MP, McPherson D, Merkulov G, Milshina NV, Mobarry C, Morris J, Moshrefi A, Mount SM, Moy M, Murphy B, Murphy L, Muzny DM, Nelson DL, Nelson DR, Nelson KA, Nixon K, Nusskern DR, Pacleb JM, Palazzolo M, Pittman GS, Pan S, Pollard J, Puri V, Reese MG, Reinert K, Remington K, Saunders RD, Scheeler F, Shen H, Shue BC, Siden-Kiamos I, Simpson M, Skupski MP, Smith T, Spier E, Spradling AC, Stapleton M, Strong R, Sun E, Svirskas R, Tector C, Turner R, Venter E, Wang AH, Wang X, Wang ZY, Wassarman DA, Weinstock GM, Weissenbach J, Williams SM, Woodage T, Worley KC, Wu D, Yang S, Yao QA, Ye J, Yeh RF, Zaveri JS, Zhan M, Zhang G, Zhao Q, Zheng L, Zheng XH, Zhong FN, Zhong W, Zhou X, Zhu S, Zhu X, Smith HO, Gibbs RA, Myers EW, Rubin GM, Venter JC (2000). The genome sequence of *Drosophila melanogaster*. *Science*, 287, (5461): 2185-95.
4. Adelman JP, Bond CT, Douglass J, Herbert E (1987). Two mammalian genes transcribed from opposite strands of the same DNA locus. *Science*, 235, (4795): 1514-7.
5. Akasaka T, Kanno M, Balling R, Mieza MA, Taniguchi M, Koseki H (1996). A role for mel-18, a Polycomb group-related vertebrate gene, during theanterior-posterior specification of the axial skeleton. *Development*, 122, (5): 1513-22.
6. Amati F, Conti E, Novelli A, Bengala M, Diglio MC, Marino B, Giannotti A, Gabrielli O, Novelli G, Dallapiccola B (1999). Atypical deletions suggest five 22q11.2 critical regions related to the DiGeorge/velo-cardio-facial syndrome. *Eur J Hum Genet.*, 7, (8): 903-9.
7. Ambros V, Bartel B, Bartel DP, Burge CB, Carrington JC, Chen X, Dreyfuss G, Eddy SR, Griffiths-Jones S, Marshall M, Matzke M, Ruvkun G, Tuschl T (2003). A uniform system for microRNA annotation. *RNA*, 9, (3): 277-9.
8. Anderson S (1981). Shotgun DNA sequencing using cloned DNase I-generated fragments. *Nucleic Acids Res.*, 9, (13): 3015-27.
9. Annesi G, Gambardella A, Carrideo S, Incorpora G, Labate A, Pasqua AA, Civitelli D, Polizzi A, Annesi F, Spadafora P, Tarantino P, Ciro Candiano IC, Romeo N, De Marco EV, Ventura P, LePiane E, Zappia M, Aguglia U, Pavone L, Quattrone A (2003). Two novel SCN1A missense mutations in generalized epilepsy with febrile seizures plus. *Epilepsia*, 44, (9): 1257-8.
10. Antequera F, Bird A (1993). Number of CpG islands and genes in human and mouse. *Proc Natl Acad Sci U S A*, 90, (24): 11995-9.
11. Aparicio S, Chapman J, Stupka E, Putnam N, Chia JM, Dehal P, Christoffels A, Rash S, Hoon S, Smit A, Gelpke MD, Roach J, Oh T, Ho IY, Wong M, Detter C, Verhoef F, Predki P, Tay A, Lucas S, Richardson P, Smith SF, Clark MS, Edwards YJ, Doggett N, Zharkikh A, Tavtigian SV, Pruss D, Barnstead M, Evans C, Baden H, Powell J, Glusman G, Rowen L, Hood L, Tan YH, Elgar G, Hawkins T, Venkatesh B, Rokhsar D, Brenner S (2002). Whole-genome shotgun assembly and analysis of the genome of *Fugu rubripes*. *Science*, 297, (5585): 1301-10.

12. Aravind L, Watanabe H, Lipman DJ, Koonin EV (2000). Lineage-specific loss and divergence of functionally linked genes in eukaryotes. *Proc Natl Acad Sci U S A*, 97, (21): 11319-24.
13. Arlt MF, Xu B, Durkin SG, Casper AM, Kastan MB, Glover TW (2004). BRCA1 is required for common-fragile-site stability via its G2/M checkpoint function. *Mol Cell Biol.*, 24, (15): 6701-9.
14. Ashley CT, Jr Wilkinson KD, Reines D, Warren ST (1993). FMR1 protein: conserved RNP family domains and selective RNA binding. *Science*, 262, (5133): 563-6.
15. Audenaert D, Claes L, Ceulemans B, Lofgren A, Van Broeckhoven C, De Jonghe P (2003). A deletion in SCN1B is associated with febrile seizures and early-onset absence epilepsy. *Neurology*, 61, (6): 854-56.
16. Barrell BG, Air GM, Hutchison CA 3rd (1976). Overlapping genes in bacteriophage phiX174. *Nature*, 264, (5581): 34-41.
17. Baulac S, Huberfeld G, Gourfinkel-An I, Mitropoulou G, Beranger A, Prud'homme JF, Baulac M, Brice A, Bruzzone R, LeGuern E (2001). First genetic evidence of GABA(A) receptor dysfunction in epilepsy: a mutation in the gamma2-subunit gene. *Nat Genet.*, 28, (1): 46-8.
18. Bauren G, Wieslander L (1994). Splicing of Balbiani ring 1 gene pre-mRNA occurs simultaneously with transcription. *Cell*, 76, (1): 183-92.
19. Beaudoin E, Freier S, Wyatt JR, Claverie JM, Gautheret D (2000). Patterns of variant polyadenylation signal usage in human genes. *Genome Res.*, 10, (7): 1001-10.
20. Bejerano G, Pheasant M, Makunin I, Stephen S, Kent WJ, Mattick JS, Haussler D. (2004). Ultraconserved elements in the human genome. *Science*, 304, (5675): 1321-5.
21. Bellen HJ, Levis RW, Liao G, He Y, Carlson JW, Tsang G, Evans-Holm M, Hiesinger PR, Schulze KL, Rubin GM, Hoskins RA, Spradling AC (2004). The BDGP gene disruption project: single transposon insertions associated with 40% of Drosophila genes. *Genetics*, 167, (2): 761-81.
22. Bentley DR, Deloukas P, Dunham A, French L, Gregory SG, Humphray SJ, Mungall AJ, Ross MT, Carter NP, Dunham I, Scott CE, Ashcroft KJ, Atkinson AL, Aubin K, Beare DM, Bethel G, Brady N, Brook JC, Burford DC, Burrill WD, Burrows C, Butler AP, Carder C, Catanese JJ, Clee CM, Clegg SM, Copley V, Coffey AJ, Cole CG, Collins JE, Conquer JS, Cooper RA, Culley KM, Dawson E, Dearden FL, Durbin RM, de Jong PJ, Dhami PD, Earthrowl ME, Edwards CA, Evans RS, Gillson CJ, Ghori J, Green L, Gwilliam R, Halls KS, Hammond S, Harper GL, Heathcote RW, Holden JL, Holloway E, Hopkins BL, Howard PJ, Howell GR, Huckle EJ, Hughes J, Hunt PJ, Hunt SE, Izmajlowicz M, Jones CA, Joseph SS, Laird G, Langford CF, Lehtvaslaihio MH, Leversha MA, McCann OT, McDonald LM, McDowall J, Maslen GL, Mistry D, Moschonas NK, Neocleous V, Pearson DM, Phillips KJ, Porter KM, Prathalingam SR, Ramsey YH, Ranby SA, Rice CM, Rogers J, Rogers LJ, Sarafidou T, Scott DJ, Sharp GJ, Shaw-Smith CJ, Smink LJ, Soderlund C, Sotharan EC, Steingruber HE, Sulston JE, Taylor A, Taylor RG, Thorpe AA, Tinsley E, Warry GL, Whittaker A, Whittaker P, Williams SH, Wilmer TE, Wooster R, Wright CL (2001). The physical maps for sequencing human chromosomes 1, 6, 9, 10, 13, 20 and X. *Nature*, 409, (6822): 942-3.
23. Bentley NJ, Holtzman DA, Flaggs G, Keegan KS, DeMaggio A, Ford JC, Hoekstra M, Carr AM (1996). The *Schizosaccharomyces pombe* rad3 checkpoint gene. *EMBO J.*, 15, (23): 6641-51.
24. Berg JM, Fauch JA, Pendrey MJ, Penrose LS, Ridler MA, Shapiro A (1969). A homozygous chromosomal variant. *Lancet*, 1, (7593): 531.
25. Berglund JA, Chua K, Abovich N, Reed R, Rosbash M (1997). The splicing factor BBP interacts specifically with the pre-mRNA branchpoint sequence UACUAAC. *Cell*, 89, (5): 781-7.
26. Berkovic SF, Izzillo P, McMahon JM, Harkin LA, McIntosh AM, Phillips HA, Briellmann RS, Wallace RH, Mazarib A, Neufeld MY, Kocczyn AD, Scheffer IE, Mulley JC (2004). LGI1 mutations in temporal lobe epilepsies. *Neurology.*, 62, (7): 1115-9.
27. Beyer AL, Osheim YN (1988). Splice site selection, rate of splicing, and alternative splicing on nascent transcripts. *Genes Dev.*, 2, (6): 754-65.

28. Bickmore WA, Sumner AT (1989). Mammalian chromosome banding--an expression of genome organization. *Trends Genet.*, 5, (5): 144-8.
29. Biervert C, Schroeder BC, Kubisch C, Berkovic SF, Propping P, Jentsch TJ, Steinlein OK (1998). A potassium channel mutation in neonatal human epilepsy. *Science*, 279, (5349): 403-6.
30. Bihoreau MT, Gauguier D, Kato N, Hyne G, Lindpaintner K, Rapp JP, James MR, Lathrop GM (1997). A linkage map of the rat genome derived from three F2 crosses. *Genome Res.*, 7, (5): 434-40.
31. Bihoreau MT, Sebag-Montefiore L, Godfrey RF, Wallis RH, Brown JH, Danoy PA, Collins SC, Rouard M, Kaisaki PJ, Lathrop M, Gauguier D (2001). A high-resolution consensus linkage map of the rat, integrating radiation hybrid and genetic maps. *Genomics*, 75, (1-3): 57-69.
32. Bird AP (1986). CpG-rich islands and the function of DNA methylation. *Nature*, 321, (6067): 209-13.
33. Bird AP, Wolffe AP (1999). Methylation-induced repression-belts, braces, and chromatin. *Cell*, 99, (5): 451-4.
34. Blattner FR, Plunkett G 3rd, Bloch CA, Perna NT, Burland V, Riley M, Collado-Vides J, Glasner JD, Rode CK, Mayhew GF, Gregor J, Davis NW, Kirkpatrick HA, Goeden MA, Rose DJ, Mau B, Shao Y (1997). The complete genome sequence of *Escherichia coli* K-12. *Science*, 277, (5331): 1453-74.
35. Blom N, Gammeltoft S, Brunak S (1999). Sequence- and structure-based prediction of eukaryotic protein phosphorylation sites. *J Mol Biol.*, 294, (5): 1351-62.
36. Botstein D, White RL, Skolnick M, Davis RW (1980). Construction of a genetic linkage map in man using restriction fragment length polymorphisms. *Am J Hum Genet.*, 32, (3): 314-31.
37. Boudrez A, Beullens M, Waelkens E, Stalmans W, Bollen M (2002). Phosphorylation-dependent interaction between the splicing factors SAP155 and NIPP1. *J Biol Chem.*, 277, (35): 31834-41.
38. Bouffard GG, Idol JR, Braden VV, Iyer LM, Cunningham AF, Weintraub LA, Touchman JW, Mohr-Tidwell RM, Peluso DC, Fulton RS, Ueltzen MS, Weissenbach J, Magness CL, Green ED (1997). A physical map of human chromosome 7: an integrated YAC contig map with average STS spacing of 79 kb. *Genome Res.*, 7, (7): 673-92.
39. Bourgeois CF, Lejeune F, Stevenin J (2004). Broad specificity of SR (serine/arginine) proteins in the regulation of alternative splicing of pre-messenger RNA. *Prog Nucleic Acid Res Mol Biol.*, 78: 37-88.
40. Boutros M, Kiger AA, Armknecht S, Kerr K, Hild M, Koch B, Haas SA, Consortium HF, Paro R, Perrimon N (2004). Genome-wide RNAi analysis of growth and viability in *Drosophila* cells. *Science*, 303, (5659): 832-5.
41. Brenner S, Elgar G, Sandford R, Macrae A, Venkatesh B, Aparicio S (1993). Characterization of the pufferfish (*Fugu*) genome as a compact model vertebrate genome. *Nature*, 366, (6452): 265-8.
42. Brodtkorb E, Gu W, Nakken KO, Fischer C, Steinlein OK (2002). Familial temporal lobe epilepsy with aphasic seizures and linkage to chromosome 10q22-q24. *Epilepsia*, 43, (3): 228-35.
43. Broman KW, Murray JC, Sheffield VC, White RL, Weber JL (1998). Comprehensive human genetic maps: individual and sex-specific variation in recombination. *Am J Hum Genet.*, 63, (3): 861-9.
44. Brosi R, Hauri HP, Kramer A (1993). Separation of splicing factor SF3 into two components and purification of SF3a activity. *J Biol Chem.*, 268, (23): 17640-6.
45. Brown EJ, Baltimore D (2000). ATR disruption leads to chromosomal fragmentation and early embryonic lethality. *Genes Dev.*, 14, (4): 397-402.

46. Bryce SD, Lindsay S, Gladstone AJ, Braithwaite K, Chapman C, Spurr NK, Lunec J (1994). A novel family of cathepsin L-like (CTSL) sequences on human chromosome 10q and related transcripts. *Genomics*, 24, (3): 568-76.
47. Bulai T, Bratosin D, Artenie V, Montreuil J (2002). Characterization of a sialate pyruvate-lyase in the cytosol of human erythrocytes. *Biochimie.*, 84, (7): 655-60.
48. Burge CB, Padgett RA, Sharp PA (1998). Evolutionary fates and origins of U12-type introns. *Mol Cell.*, 2, (6): 773-85.
49. Burke DT, Carle GF, Olson MV. (1987). Cloning of large segments of exogenous DNA into yeast by means of artificial chromosome vectors. *Science*, 236, (4803): 806-12.
50. Burn J, Goodship J (1996). Developmental genetics of the heart. *Curr Opin Genet Dev.*, 6, (3): 322-5.
51. Burns N, Grimwade B, Ross-Macdonald PB, Choi EY, Finberg K, Roeder GS, Snyder M (1994). Large-scale analysis of gene expression, protein localization, and gene disruption in *Saccharomyces cerevisiae*. *Genes Dev.*, 8, (9): 1087-105
52. Carey AH, Kelly D, Halford S, Wadey R, Wilson D, Goodship J, Burn J, Paul T, Sharkey A, Dumanski J, et al (1992). Molecular genetic study of the frequency of monosomy 22q11 in DiGeorge syndrome. *Am J Hum Genet.*, 51, (5): 964-70.
53. Carlson C, Sirotkin H, Pandita R, Goldberg R, McKie J, Wadey R, Patanjali SR, Weissman SM, Anyane-Yeboah K, Warburton D, Scambler P, Shprintzen R, Kucherlapati R, Morrow BE (1997). Molecular definition of 22q11 deletions in 151 velo-cardio-facial syndrome patients. *Am J Hum Genet.*, 61, (3): 620-9.
54. Carninci P, Waki K, Shiraki T, Konno H, Shibata K, Itoh M, Aizawa K, Arakawa T, Ishii Y, Sasaki D, Bono H, Kondo S, Sugahara Y, Saito R, Osato N, Fukuda S, Sato K, Watahiki A, Hirozane-Kishikawa T, Nakamura M, Shibata Y, Yasunishi A, Kikuchi N, Yoshiki A, Kusakabe M, Gustincich S, Beisel K, Pavan W, Aidinis V, Nakagawara A, Held WA, Iwata H, Kono T, Nakauchi H, Lyons P, Wells C, Hume DA, Fagiolini M, Hensch TK, Brinkmeier M, Camper S, Hirota J, Mombaerts P, Muramatsu M, Okazaki Y, Kawai J, Hayashizaki Y (2003). Targeting a complex transcriptome: the construction of the mouse full-length cDNA encyclopedia. *Genome Res.*, 13, (6B): 1273-89.
55. Casper AM, Nghiem P, Artl MF, Glover TW (2002). ATR regulates fragile site stability. *Cell*, 111, (6): 779-89.
56. C. elegans Sequencing Consortium (1998). Genome sequence of the nematode *C. elegans*: a platform for investigating biology. *Science*, 282, (5396): 2012-8.
57. Chabot B, Bisotto S, Vincent M (1995). The nuclear matrix phosphoprotein p255 associates with splicing complexes as part of the [U4/U6.U5] tri-snRNP particle. *Nucleic Acids Res.*, 23, (16): 3206-13.
58. Chabot B, LeBel C, Hutchison S, Nasim FH, Simard MJ (2003). Heterogeneous nuclear ribonucleoprotein particle A/B proteins and the control of alternative splicing of the mammalian heterogeneous nuclear ribonucleoprotein particle A1 pre-mRNA. *Prog Mol Subcell Biol.*, 31, (59-88):
59. Chadwick BP, Frischauf AM (1998). The CD39-like gene family: identification of three new human members (CD39L2, CD39L3, and CD39L4), their murine homologues, and a member of the gene family from *Drosophila melanogaster*. *Genomics*, 50, (3): 357-67.
60. Champion-Arnaud P, Reed R (1994). The prespliceosome components SAP 49 and SAP 145 interact in a complex implicated in tethering U2 snRNP to the branch site. *Genes Dev.*, 8, (16): 1974-83.
61. Charlier C, Singh NA, Ryan SG, Lewis TB, Reus BE, Leach RJ, Leppert M (1998). A pore mutation in a novel KQT-like potassium channel gene in an idiopathic epilepsy family. *Nat Genet.*, 18, (1): 53-5.
62. Chernova OB, Somerville RP, Cowell JK (1998). A novel gene, LG11, from 10q24 is rearranged and downregulated in malignant brain tumors. *Oncogene*, 17, (22): 2873-81.

63. Chieffo C, Garvey N, Gong W, Roe B, Zhang G, Silver L, Emanuel BS, Budarf ML (1997). Isolation and characterization of a gene from the DiGeorge chromosomal region homologous to the mouse Tbx1 gene. *Genomics*, 43, (3): 267-77.
64. Chimpanzee Sequencing and Analysis Consortium (2005). Initial sequence of the chimpanzee genome and comparison with the human genome. *Nature*, 437, (7055): 69-87.
65. Chiurazzi P, Neri G.(2001). Pharmacological reactivation of inactive genes: the fragile X experience. *Brain Res Bull.*, 56, (3-4): 383-7.
66. Chu WM, Ballard R, Carpick BW, Williams BR, Schmid CW (1998). Potential Alu function: regulation of the activity of double-stranded RNA-activated kinase PKR. *Mol Cell Biol.*, 18, (1): 58-68.
67. Chumakov I, Rigault P, Guillou S, Ougen P, Billaut A, Guasconi G, Gervy P, LeGall I, Soularue P, Grinas L *et al* (1992). Continuum of overlapping clones spanning the entire human chromosome 21q. *Nature*, 359, (6394): 380-7.
68. Chumakov IM, Rigault P, Le Gall I, Bellanne-Chantelot C, Billaut A, Guillou S, Soularue P, Guasconi G, Poullier E, Gros I *et al* (1995). A YAC contig map of the human genome. *Nature*, 377, (6547 Suppl): 175-297.
69. Cimprich KA, Shin TB, Keith CT, Schreiber SL (1996). cDNA cloning and gene mapping of a candidate human cell cycle checkpoint protein. *Proc Natl Acad Sci U S A*, 93, (7): 2850-5.
70. Ciullo M, Debily MA, Rozier L, Autiero M, Billaut A, Mayau V, El Marhomy S, Guardiola J, Bernheim A, Coullin P, Piatier-Tonneau D, Debatisse M (2002). Initiation of the breakage-fusion-bridge mechanism through common fragile site activation in human breast cancer cells: the model of PIP gene duplication from a break at FRA71. *Hum Mol Genet.*, 11, (23): 2887-94.
71. Claes L, Del-Favero J, Ceulemans B, Lagae L, Van Broeckhoven C, De Jonghe P (2001). De novo mutations in the sodium-channel gene SCN1A cause severe myoclonic epilepsy of infancy. *Am J Hum Genet.*, 68, (6): 1327-32.
72. Claes L, Ceulemans B, Audenaert D, Smets K, Lofgren A, Del-Favero J, Ala-Mello S, Basel-Vanagaite L, Plecko B, Raskin S, Thiry P, Wolf NI, Van Broeckhoven C, De Jonghe P (2003). De novo SCN1A mutations are a major cause of severe myoclonic epilepsy of infancy. *Hum Mutat.*, 21, (6): 615-21.
73. Clark MS, Edwards YJ, Peterson D, Clifton SW, Thompson AJ, Sasaki M, Suzuki Y, Kikuchi K, Watabe S, Kawakami K, Sugano S, Elgar G, Johnson SL (2003). Fugu ESTs: new resources for transcription analysis and genome annotation. *Genome Res.*, 13, (12): 2747-53.
74. Clark TA, Sugnet CW, Ares M. Jr (2002). Genomewide analysis of mRNA processing in yeast using splicing-specific microarrays. *Science*, 296, (5569): 907-10.
75. Coffey AJ, Roberts RG, Green ED, Cole CG, Butler R, Anand R, Giannelli F, Bentley DR (1992). Construction of a 2.6-Mb contig in yeast artificial chromosomes spanning the human dystrophin gene using an STS-based approach. *Genomics*, 12, (3): 474-84.
76. Cohen D, Chumakov I, Weissenbach J (1993). A first-generation physical map of the human genome. *Nature*, 366, (6456): 698-701.
77. Collins F, Galas D (1993). A new five-year plan for the U.S. Human Genome Project. *Science*, 262, (5130): 43-6.
78. Collins FS (1995). Ahead of schedule and under budget: the Genome Project passes its fifth birthday. *Proc Natl Acad Sci U S A*, 92, (24): 10821-3.
79. Collins FS, Patrinos A, Jordan E, Chakravarti A, Gesteland R, Walters L (1998a). New goals for the U.S. Human Genome Project: 1998-2003. *Science*, 282, (5389): 682-9.
80. Collins FS, Brooks LD, Chakravarti A (1998b). A DNA polymorphism discovery resource for research on human genetic variation. *Genome Res.*, 8, (12): 1229-31.

81. Collins J, Hohn B (1978). Cosmids: a type of plasmid gene-cloning vector that is packageable in vitro in bacteriophage lambda heads. *Proc Natl Acad Sci U S A*, 75, (9): 4242-6.
82. Collins JE, Cole CG, Smink LJ, Garrett CL, Leversha MA, Soderlund CA, Maslen GL, Everett LA, Rice KM, Coffey AJ *et al* (1995). A high-density YAC contig map of human chromosome 22. *Nature*, 377, (6547 Suppl): 367-79.
83. Corden JL, Patturajan M (1997). A CTD function linking transcription to splicing. *Trends Biochem Sci.*, 22, (11): 413-6.
84. Cossette P, Liu L, Brisebois K, Dong H, Lortie A, Vanasse M, Saint-Hilaire JM, Carmant L, Verner A, Lu WY, Wang YT, Rouleau GA (2002). Mutation of GABRA1 in an autosomal dominant form of juvenile myoclonic epilepsy. *Nat Genet.*, 31, (2): 184-9.
85. Coulson A, Kozono Y, Lutterbach B, Shownkeen R, Sulston J, Waterston R (1991). YACs and the *C. elegans* genome. *Bioessays*, 13, (8): 413-7.
86. Coulson A, Waterston R, Kiff J, Sulston J, Kohara Y (1988). Genome linking with yeast artificial chromosomes. *Nature*, 335, (6186): 184-6.
87. Cox DR, Burmeister M, Price ER, Kim S, Myers RM (1990). Radiation hybrid mapping: a somatic cell genetic method for constructing high-resolution maps of mammalian chromosomes. *Science*, 250, (4978): 245-50.
88. Craig JM, Bickmore WA (1993). Chromosome bands-flavours to savour. *Bioessays*, 15, (5): 349-54.
89. Crnogorac-Jurcevic T, Brown JR, Lehrach H, Schalkwyk LC (1997). Tetraodon fluviatilis, a new puffer fish model for genome studies. *Genomics*, 41, (2): 177-84.
90. Crosier M, Viggiano L, Guy J, Misceo D, Stones R, Wei W, Hearn T, Ventura M, Archidiacono N, Rocchi M, Jackson MS (2002). Human paralogs of KIAA0187 were created through independent pericentromeric-directed and chromosome-specific duplication mechanisms. *Genome Res.*, 12, (1): 67-80
91. Davis DB, Delmonte AJ, Ly CT, McNally EM (2000). Myoferlin, a candidate gene and potential modifier of muscular dystrophy. *Hum Mol Genet.*, 9, (2): 217-26.
92. Daw SC, Taylor C, Kraman M, Call K, Mao J, Schuffenhauer S, Meitinger T, Lipson T, Goodship J, Scambler P (1996). A common region of 10p deleted in DiGeorge and velocardiofacial syndromes. *Nat Genet.*, 13, (4): 458-60.
93. De Fusco M, Becchetti A, Patrignani A, Annesi G, Gambardella A, Quattrone A, Ballabio A, Wanke E, Casari G (2000). The nicotinic receptor beta 2 subunit is mutant in nocturnal frontal lobe epilepsy. *Nat Genet.*, 26, (3): 275-6.
94. Dedek K, Kunath B, Kananura C, Reuner U, Jentsch TJ, Steinlein OK (2001). Myokymia and neonatal epilepsy caused by a mutation in the voltage sensor of the KCNQ2 K⁺ channel. *Proc Natl Acad Sci U S A*, 98, (21): 12272-7.
95. Deloukas P, Dauwerse JG, Moschonas NK, van Ommen GJ, van Loon AP (1993). Three human glutamate dehydrogenase genes (GLUD1, GLUDP2, and GLUDP3) are located on chromosome 10q, but are not closely physically linked. *Genomics*, 17, (3): 676-81.
96. Deloukas P, Schuler GD, Gyapay G, Beasley EM, Soderlund C, Rodriguez-Tome P, Hui L, Matisse TC, McKusick KB, Beckmann JS, Bentolila S, Bihoreau M, Birren BB, Browne J, Butler A, Castle AB, Chiannilkulchai N, Clee C, Day PJ, Dehejia A, Dibling T, Drouot N, Duprat S, Fizames C, Fox S, Gelling S, Green L, Harrison P, Hocking R, Holloway E, Hunt S, Keil S, Lijnzaad P, Louis-Dit-Sully C, Ma J, Mendis A, Miller J, Morissette J, Muselet D, Nusbaum HC, Peck A, Rozen S, Simon D, Slonim DK, Staples R, Stein LD, Stewart EA, Suchard MA, Thangarajah T, Vega-Czarny N, Webber C, Wu X, Hudson J, Auffray C, Nomura N, Sikela JM, Polymeropoulos MH, James MR, Lander ES, Hudson TJ, Myers RM, Cox DR, Weissenbach J, Boguski MS, Bentley DR (1998). A physical map of 30,000 human genes. *Science*, 282, (5389): 744-6.
97. Deloukas P, Matthews LH, Ashurst J, Burton J, Gilbert JG, Jones M, Stavrides G, Almeida JP, Babbage AK, Bagguley CL, Bailey J, Barlow KF, Bates KN, Beard LM, Beare DM, Beasley OP,

- Bird CP, Blakey SE, Bridgeman AM, Brown AJ, Buck D, Burrill W, Butler AP, Carder C, Carter NP, Chapman JC, Clamp M, Clark G, Clark LN, Clark SY, Clee CM, Clegg S, Cobley VE, Collier RE, Connor R, Corby NR, Coulson A, Coville GJ, Deadman R, Dhami P, Dunn M, Ellington AG, Frankland JA, Fraser A, French L, Garner P, Grafham DV, Griffiths C, Griffiths MN, Gwilliam R, Hall RE, Hammond S, Harley JL, Heath PD, Ho S, Holden JL, Howden PJ, Huckle E, Hunt AR, Hunt SE, Jekosch K, Johnson CM, Johnson D, Kay MP, Kimberley AM, King A, Knights A, Laird GK, Lawlor S, Lehtasalo MH, Leversha M, Lloyd C, Lloyd DM, Lovell JD, Marsh VL, Martin SL, McConnachie LJ, McLay K, McMurray AA, Milne S, Mistry D, Moore MJ, Mullikin JC, Nickerson T, Oliver K, Parker A, Patel R, Pearce TA, Peck AI, Phillimore BJ, Prathalingam SR, Plumb RW, Ramsay H, Rice CM, Ross MT, Scott CE, Sehra HK, Shownkeen R, Sims S, Skuce CD, Smith ML, Soderlund C, Steward CA, Sulston JE, Swann M, Sycamore N, Taylor R, Tee L, Thomas DW, Thorpe A, Tracey A, Tromans AC, Vaudin M, Wall M, Wallis JM, Whitehead SL, Whittaker P, Willey DL, Williams L, Williams SA, Wilming L, Wray PW, Hubbard T, Durbin RM, Bentley DR, Beck S, Rogers J (2001). The DNA sequence and comparative analysis of human chromosome 20. *Nature*, 414, (6866): 865-71.
98. Deloukas P, Earthworn ME, Grafham DV, Rubinfeld M, French L, Steward CA, Sims SK, Jones MC, Searle S, Scott C, Howe K, Hunt SE, Andrews TD, Gilbert JG, Swarbreck D, Ashurst JL, Taylor A, Battles J, Bird CP, Ainscough R, Almeida JP, Ashwell RI, Ambrose KD, Babbage AK, Bagguley CL, Bailey J, Banerjee R, Bates K, Beasley H, Bray-Allen S, Brown AJ, Brown JY, Burford DC, Burrill W, Burton J, Cahill P, Camire D, Carter NP, Chapman JC, Clark SY, Clarke G, Clee CM, Clegg S, Corby N, Coulson A, Dhami P, Dutta I, Dunn M, Faulkner L, Frankish A, Frankland JA, Garner P, Garnett J, Gribble S, Griffiths C, Grocock R, Gustafson E, Hammond S, Harley JL, Hart E, Heath PD, Ho TP, Hopkins B, Horne J, Howden PJ, Huckle E, Hynds C, Johnson C, Johnson D, Kana A, Kay M, Kimberley AM, Kershaw JK, Kokkinaki M, Laird GK, Lawlor S, Lee HM, Leongamornlert DA, Laird G, Lloyd C, Lloyd DM, Loveland J, Lovell J, McLaren S, McLay KE, McMurray A, Mashreghi-Mohammadi M, Matthews L, Milne S, Nickerson T, Nguyen M, Overton-Larty E, Palmer SA, Pearce AV, Peck AI, Pelan S, Phillimore B, Porter K, Rice CM, Rogosin A, Ross MT, Sarafidou T, Sehra HK, Shownkeen R, Skuce CD, Smith M, Standring L, Sycamore N, Tester J, Thorpe A, Torcasso W, Tracey A, Tromans A, Tsolas J, Wall M, Walsh J, Wang H, Weinstock K, West AP, Willey DL, Whitehead SL, Wilming L, Wray PW, Young L, Chen Y, Lovering RC, Moschonas NK, Siebert R, Fechtel K, Bentley D, Durbin R, Hubbard T, Doucette-Stamm L, Beck S, Smith DR, Rogers J (2004). The DNA sequence and comparative analysis of human chromosome 10. *Nature*, 429, (6990): 375-81.
99. DeRisi JL, Iyer VR, Brown PO (1997). Exploring the metabolic and genetic control of gene expression on a genomic scale. *Science*, 278, (5338): 680-6.
100. Dermitzakis ET, Reymond A, Scamuffa N, Ucla C, Kirkness E, Rossier C, Antonarakis SE (2003). Evolutionary discrimination of mammalian conserved non-genic sequences (CNGs). *Science*, 302, (5647): 1033-5.
101. Dermitzakis ET, Reymond A, Antonarakis SE (2005). Conserved non-genic sequences - an unexpected feature of mammalian genomes. *Nat Rev Genet.*, 6, (2): 151-7.
102. Dib C, Faure S, Fizames C, Samson D, Drouot N, Vignal A, Millasseau P, Marc S, Hazan J, Seboun E, Lathrop M, Gyapay G, Morissette J, Weissenbach J (1996). A comprehensive genetic map of the human genome based on 5,264 microsatellites. *Nature*, 380, (6570): 152-4.
103. Dietrich WF, Miller J, Steen R, Merchant MA, Damron-Boles D, Husain Z, Dredge R, Daly MJ, Ingalls KA, O'Connor TJ (1996). A comprehensive genetic map of the mouse genome. *Nature*, 380, (6570): 149-52.
104. Digilio MC, Marino B, Giannotti A, Dallapiccola B (1997). Familial deletions of chromosome 22q11. *Am J Med Genet.*, 28, (73): 95-6.
105. Doggett NA, Goodwin LA, Tesmer JG, Meincke LJ, Bruce DC, Clark LM, Altherr MR, Ford AA, Chi HC, Marrone BL et al (1995). An integrated physical map of human chromosome 16. *Nature*, 377, (6547 Suppl): 335-65.
106. Donis-Keller H, Green P, Helms C, Cartinhour S, Weiffenbach B, Stephens K, Keith TP, Bowden DW, Smith DR, Lander ES et al (1987). A genetic linkage map of the human genome. *Cell*, 51, (2): 319-37.
107. Driever W, Solnica-Krezel L, Schier AF, Neuhauss SC, Malicki J, Stemple DL, Stainier DY, Zwartkruis F, Abdelilah S, Rangini Z, Belak J, Boggs C (1996). A genetic screen for mutations affecting embryogenesis in zebrafish. *Development*, (123): 37-46.

108. Driscoll DA, Salvin J, Sellinger B, Budarf ML, McDonald-McGinn DM, Zackai EH, Emanuel BS (1993). Prevalence of 22q11 microdeletions in DiGeorge and velocardiofacial syndromes: implications for genetic counselling and prenatal diagnosis. *J Med Genet.*, 30, (10): 813-7.
109. Du W, Vidal M, Xie JE, Dyson N (1996). RBF, a novel RB-related gene that regulates E2F activity and interacts with cyclin E in *Drosophila*. *Genes Dev.*, 10, (10): 1206-18.
110. Dunham A, Matthews LH, Burton J, Ashurst JL, Howe KL, Ashcroft KJ, Beare DM, Burford DC, Hunt SE, Griffiths-Jones S, Jones MC, Keenan SJ, Oliver K, Scott CE, Ainscough R, Almeida JP, Ambrose KD, Andrews DT, Ashwell RI, Babbage AK, Bagguley CL, Bailey J, Bannerjee R, Barlow KF, Bates K, Beasley H, Bird CP, Bray-Allen S, Brown AJ, Brown JY, Burrill W, Carder C, Carter NP, Chapman JC, Clamp ME, Clark SY, Clarke G, Clee CM, Clegg SC, Cobley V, Collins JE, Corby N, Coville GJ, Deloukas P, Dhami P, Dunham I, Dunn M, Earthrowl ME, Ellington AG, Faulkner L, Frankish AG, Frankland J, French L, Garner P, Garnett J, Gilbert JG, Gilson CJ, Ghori J, Grafham DV, Gribble SM, Griffiths C, Hall RE, Hammond S, Harley JL, Hart EA, Heath PD, Howden PJ, Huckle EJ, Hunt PJ, Hunt AR, Johnson C, Johnson D, Kay M, Kimberley AM, King A, Laird GK, Langford CJ, Lawlor S, Leongamornlert DA, Lloyd DM, Lloyd C, Loveland JE, Lovell J, Martin S, Mashreghi-Mohammadi M, McLaren SJ, McMurray A, Milne S, Moore MJ, Nickerson T, Palmer SA, Pearce AV, Peck AI, Pelan S, Phillimore B, Porter KM, Rice CM, Searle S, Sehra HK, Shownkeen R, Skuce CD, Smith M, Steward CA, Sycamore N, Tester J, Thomas DW, Tracey A, Tromans A, Tubby B, Wall M, Wallis JM, West AP, Whitehead SL, Willey DL, Wilming L, Wray PW, Wright MW, Young L, Coulson A, Durbin R, Hubbard T, Sulston JE, Beck S, Bentley DR, Rogers J, Ross MT (2004). The DNA sequence and analysis of human chromosome 13. *Nature*, 428, (6982): 522-8.
111. Dunham I, Shimizu N, Roe BA, Chisoe S, Hunt AR, Collins JE, Bruskiewich R, Beare DM, Clamp M, Smink LJ, Ainscough R, Almeida JP, Babbage A, Bagguley C, Bailey J, Barlow K, Bates KN, Beasley O, Bird CP, Blakey S, Bridgeman AM, Buck D, Burgess J, Burrill WD, O'Brien KP *et al* (1999). The DNA sequence of human chromosome 22. *Nature*, 402, (6761): 489-95.
112. Echard A, Hickson GR, Foley E, O'Farrell PH (2004). Terminal cytokinesis events uncovered after an RNAi screen. *Curr Biol.*, 14, (18): 1685-93.
113. Edelmann L, Pandita RK, Spiteri E, Funke B, Goldberg R, Palanisamy N, Chaganti RS, Magenis E, Shprintzen RJ, Morrow BE (1999). A common molecular basis for rearrangement disorders on chromosome 22q11. *Hum Mol Genet.*, 8, (7): 1157-67.
114. Edelmann L, Pandita RK, Morrow BE. (1999). Low-copy repeats mediate the common 3-Mb deletion in patients with velo-cardio-facial syndrome. *Am J Hum Genet.*, 64, (4): 1076-86.
115. Eggert US, Kiger AA, Richter C, Perlman ZE, Perrimon N, Mitchison TJ, Field CM (2004). Parallel chemical genetic and genome-wide RNAi screens identify cytokinesis inhibitors and targets. *PLoS Biol.*, (12): e379.
116. Elder FF, Robinson TJ (1989). Rodent common fragile sites: are they conserved? Evidence from mouse and rat. *Chromosoma*, 97, (6): 459-64.
117. Elgar G, Sandford R, Aparicio S, Macrae A, Venkatesh B, Brenner S (1996). Small is beautiful: comparative genomics with the pufferfish (*Fugu rubripes*). *Trends Genet.*, 12, (4): 145-50.
118. ENCODE Project Consortium (2004). The ENCODE (ENCyclopedia Of DNA Elements) Project. *Science*, 306, (5696): 636-40.
119. Enright AJ, Iliopoulos I, Kyrpides NC, Ouzounis CA (1999). Protein interaction maps for complete genomes based on gene fusion events. *Nature*, 402, (6757): 86-90.
120. Enright AJ, Ouzounis CA (2001). Functional associations of proteins in entire genomes by means of exhaustive detection of gene fusions. *Genome Biol.*, 2, (9): RESEARCH0034.
121. Eperon IC, Makarova OV, Mayeda A, Munroe SH, Caceres JF, Hayward DG, Krainer AR (2000). Selection of alternative 5' splice sites: role of U1 snRNP and models for the antagonistic effects of SF2/ASF and hnRNP A1. *Mol Cell Biol.*, 20, (22): 8303-18.
122. Erikson A, Aula N, Aula P, Mansson JE (2002). Free sialic acid storage (Salla) disease in Sweden. *Acta Paediatr.*, 91, (12): 1324-7.

123. Escayg A, MacDonald BT, Meisler MH, Baulac S, Huberfeld G, An-Gourfinkel I, Brice A, LeGuern E, Moulard B, Chaigne D, Buresi C, Malafosse A (2000). Mutations of SCN1A, encoding a neuronal sodium channel, in two families with GEFS+2. *Nat Genet.*, 24, (4): 343-5.
124. Escayg A, Heils A, MacDonald BT, Haug K, Sander T, Meisler MH. (2001). A novel SCN1A mutation associated with generalized epilepsy with febrile seizures plus and prevalence of variants in patients with epilepsy. *Am J Hum Genet.*, 68, (4): 866-73.
125. Ewing B, Green P (2000). Analysis of expressed sequence tags indicates 35,000 human genes. *Nat Genet.*, 25, (2): 232-4.
126. Eymard-Pierre E, Lesca G, Dollet S, Santorelli FM, di Capua M, Bertini E, Boespflug-Tanguy O (2002). Infantile-onset ascending hereditary spastic paralysis is associated with mutations in the alsin gene. *Am J Hum Genet.*, 71, (3): 518-27.
127. Fairbrother WG, Yeh RF, Sharp PA, Burge CB (2002). Predictive identification of exonic splicing enhancers in human genes. *Science*, 297, (5583): 1007-13.
128. Fertig E, Lincoln A, Martinuzzi A, Mattson RH, Hisama FM (2003). Novel LGI1 mutation in a family with autosomal dominant partial epilepsy with auditory features. *Neurology*, 60, (10): 1687-90.
129. Fields C, Adams MD, White O, Venter JC (1994). How many genes in the human genome? *Nat Genet.*, 7, (3): 345-6.
130. Fields S, Song O (1989). A novel genetic system to detect protein-protein interactions. *Nature*, 340, (6230): 245-6.
131. Fong YW, Zhou Q (2001). Stimulatory effect of splicing factors on transcriptional elongation. *Nature*, 414, (6866): 929-33.
132. Foote S, Vollrath D, Hilton A, Page DC (1992). The human Y chromosome: overlapping DNA clones spanning the euchromatic region. *Science*, 258, (5079): 60-6.
133. Fontes MR, Teh T, Jans D, Brinkworth RI, Kobe B (2003). Structural basis for the specificity of bipartite nuclear localization sequence binding by importin-alpha. *J Biol Chem.*, 278, (30): 27981-7.
134. Fraser AG, Kamath RS, Zipperlen P, Martinez-Campos M, Sohrmann M, Ahringer J (2000). Functional genomic analysis of *C. elegans* chromosome I by systematic RNA interference. *Nature*, 408, (6810): 325-30.
135. Frazer KA, Tao H, Osoegawa K, de Jong PJ, Chen X, Doherty MF, Cox DR (2004). Noncoding sequences conserved in a limited number of mammals in the SIM2 interval are frequently functional. *Genome Res.*, 14, (3): 367-72.
136. Fujiwara T, Sugawara T, Mazaki-Miyazaki E, Takahashi Y, Fukushima K, Watanabe M, Hara K, Morikawa T, Yagi K, Yamakawa K, Inoue Y (2003). Mutations of sodium channel alpha subunit type 1 (SCN1A) in intractable childhood epilepsies with frequent generalized tonic-clonic seizures. *Brain*, 126, (3): 531-46.
137. Futcher B, Latter GI, Monardo P, McLaughlin CS, Garrels JI (1999). A sampling of the yeast proteome. *Mol Cell Biol.*, 19, (11): 7357-68.
138. Ge H, Si Y, Wolffe AP (1998). A novel transcriptional coactivator, p52, functionally interacts with the essential splicing factor ASF/SF2. *Mol Cell*, 2, (6): 751-9.
139. Gecz J, Gedeon AK, Sutherland GR, Mulley JC (1996). Identification of the gene FMR2, associated with FRAXE mental retardation. *Nat Genet.*, 13, (1): 105-8.
140. Gemmill RM, Chumakov I, Scott P, Waggoner B, Rigault P, Cypser J, Chen Q, Weissenbach J, Gardiner K, Wang H *et al* (1995). A second-generation YAC contig map of human chromosome 3. *Nature*, 377, (6547 Suppl): 299-319.

141. Gennaro E, Veggiotti P, Malacarne M, Madia F, Cecconi M, Cardinali S, Cassetti A, Cecconi I, Bertini E, Bianchi A, Gobbi G, Zara F (2003). Familial severe myoclonic epilepsy of infancy: truncation of Nav1.1 and genetic heterogeneity. *Epileptic Disord.*, 5, (1): 21-5.
142. Giaever G, Chu AM, Ni L, Connelly C, Riles L, Veronneau S, Dow S, Lucau-Danila A, Anderson K, Andre B, Arkin AP, Astromoff A, El-Bakkoury M, Bangham R, Benito R, Brachat S, Campanaro S, Curtiss M, Davis K, Deutschbauer A, Entian KD, Flaherty P, Foury F, Garfinkel DJ, Gerstein M, Gotte D, Guldener U, Hegemann JH, Hempel S, Herman Z, Jaramillo DF, Kelly DE, Kelly SL, Kotter P, LaBonte D, Lamb DC, Lan N, Liang H, Liao H, Liu L, Luo C, Lussier M, Mao R, Menard P, Ooi SL, Revuelta JL, Roberts CJ, Rose M, Ross-Macdonald P, Scherens B, Schimmack G, Shafer B, Shoemaker DD, Sookhai-Mahadeo S, Storms RK, Strathern JN, Valle G, Voet M, Volckaert G, Wang CY, Ward TR, Wilhelmy J, Winzeler EA, Yang Y, Yen G, Youngman E, Yu K, Bussey H, Boeke JD, Snyder M, Philippsen P, Davis RW, Johnston M (2002). Functional profiling of the *Saccharomyces cerevisiae* genome. *Nature*, 418, (6896): 387-91.
143. Gibbs RA, Weinstock GM, Metzker ML, Muzny DM, Sodergren EJ, Scherer S, Scott G, Steffen D, Worley KC, Burch PE, Okwuonu G, Hines S, Lewis L, DeRamo C, Delgado O, Dugan-Rocha S, Miner G, Morgan M, Hawes A, Gill R, Celera, Holt RA, Adams MD, Amanatides PG, Baden-Tillson H, Barnstead M, Chin S, Evans CA, Ferriera S, Fosler C, Glodek A, Gu Z, Jennings D, Kraft CL, Nguyen T, Pfannkoch CM, Sitter C, Sutton GG, Venter JC, Woodage T, Smith D, Lee HM, Gustafson E, Cahill P, Kana A, Doucette-Stamm L, Weinstock K, Fechtel K, Weiss RB, Dunn DM, Green ED, Blakesley RW, Bouffard GG, De Jong PJ, Osoegawa K, Zhu B, Marra M, Schein J, Bosdet I, Fjell C, Jones S, Krzywinski M, Mathewson C, Siddiqui A, Wye N, McPherson J, Zhao S, Fraser CM, Shetty J, Shatsman S, Geer K, Chen Y, Abramzon S, Nierman WC, Havlak PH, Chen R, Durbin KJ, Egan A, Ren Y, Song XZ, Li B, Liu Y, Qin X, Cawley S, Worley KC, Cooney AJ, D'Souza LM, Martin K, Wu JQ, Gonzalez-Garay ML, Jackson AR, Kalafus KJ, McLeod MP, Milosavljevic A, Virk D, Volkov A, Wheeler DA, Zhang Z, Bailey JA, Eichler EE, Tuzun E, Birney E, Mongin E, Ureta-Vidal A, Woodwark C, Zdobnov E, Bork P, Suyama M, Torrents D, Alexandersson M, Trask BJ, Young JM, Huang H, Wang H, Xing H, Daniels S, Gietzen D, Schmidt J, Stevens K, Vitt U, Wingrove J, Camara F, Mar Alba M, Abril JF, Guigo R, Smit A, Dubchak I, Rubin EM, Couronne O, Poliakov A, Hubner N, Ganten D, Goesele C, Hummel O, Kreitler T, Lee YA, Monti J, Schulz H, Zimdahl H, Himmelbauer H, Lehrach H, Jacob HJ, Bromberg S, Gullings-Handley J, Jensen-Seaman MI, Kwitek AE, Lazar J, Pasko D, Tonellato PJ, Twigger S, Ponting CP, Duarte JM, Rice S, Goodstadt L, Beatson SA, Emes RD, Winter EE, Webber C, Brandt P, Nyakatura G, Adetobi M, Chiaromonte F, Elnitski L, Eswara P, Hardison RC, Hou M, Kolbe D, Makova K, Miller W, Nekrutenko A, Riemer C, Schwartz S, Taylor J, Yang S, Zhang Y, Lindpaintner K, Andrews TD, Caccamo M, Clamp M, Clarke L, Curwen V, Durbin R, Eyraas E, Searle SM, Cooper GM, Batzogluou S, Brudno M, Sidow A, Stone EA, Venter JC, Payseur BA, Bourque G, Lopez-Otin C, Puente XS, Chakrabarti K, Chatterji S, Dewey C, Pachter L, Bray N, Yap VB, Caspi A, Tesler G, Pevzner PA, Haussler D, Roskin KM, Baertsch R, Clawson H, Furey TS, Hinrichs AS, Karolchik D, Kent WJ, Rosenbloom KR, Trumbower H, Weirauch M, Cooper DN, Stenson PD, Ma B, Brent M, Arumugam M, Shteynberg D, Copley RR, Taylor MS, Riethman H, Mudunuri U, Peterson J, Guyer M, Felsenfeld A, Old S, Mockrin S, Collins F; Rat Genome Sequencing Project Consortium. (2004). Genome sequence of the Brown Norway rat yields insights into mammalian evolution. *Nature*, 428, (6982): 493-521.
144. Giot L, Bader JS, Brouwer C, Chaudhuri A, Kuang B, Li Y, Hao YL, Ooi CE, Godwin B, Vitols E, Vijayadamodar G, Pochart P, Machineni H, Welsh M, Kong Y, Zerhusen B, Malcolm R, Varrone Z, Collis A, Minto M, Burgess S, McDaniel L, Stimpson E, Spriggs F, Williams J, Neurath K, Ioime N, Agee M, Voss E, Furtak K, Renzulli R, Aanensen N, Carrola S, Bickelhaupt E, Lazovatsky Y, DaSilva A, Zhong J, Stanyon CA, Finley RL Jr, White KP, Braverman M, Jarvie T, Gold S, Leach M, Knight J, Shimkets RA, McKenna MP, Chant J, Rothberg JM (2003). A protein interaction map of *Drosophila melanogaster*. *Science*, 302, (5651): 1727-36.
145. Glover TW, Hoge AW, Miller DE, Ascara-Wilke JE, Adam AN, Dagenais SL, Wilke CM, Dierick HA, Beer DG (1998). The murine *Fhit* gene is highly similar to its human orthologue and maps to a common fragile site region. *Cancer Res.*, 58, (15): 3409-14.
146. Goffeau A, Barrell BG, Bussey H, Davis RW, Dujon B, Feldmann H, Galibert F, Hoheisel JD, Jacq C, Johnston M, Louis EJ, Mewes HW, Murakami Y, Philippsen P, Tettelin H, Oliver SG (1996). Life with 6000 genes. *Science*, 274, (5287): 546, 63-7.

147. Gogos JA, Morgan M, Luine V, Santha M, Ogawa S, Pfaff D, Karayiorgou M (1998). Catechol-O-methyltransferase-deficient mice exhibit sexually dimorphic changes in catecholamine levels and behavior. *Proc Natl Acad Sci U S A*, 95, (17): 9991-6.
148. Gonczy P, Echeverri C, Oegema K, Coulson A, Jones SJ, Copley RR, Duperon J, Oegema J, Brehm M, Cassin E, Hannak E, Kirkham M, Pichler S, Flohrs K, Goessen A, Leidel S, Alleaume AM, Martin C, Ozlu N, Bork P, Hyman AA (2000). Functional genomic analysis of cell division in *C. elegans* using RNAi of genes on chromosome III. *Nature*, 408, (6810): 331-6.
149. Gong W, Emanuel BS, Galili N, Kim DH, Roe B, Driscoll DA, Budarf ML (1997). Structural and mutational analysis of a conserved gene (DGS1) from the minimal DiGeorge syndrome critical region. *Hum Mol Genet.*, 6, (2): 267-76.
150. Gong W, Gottlieb S, Collins J, Blescia A, Dietz H, Goldmuntz E, McDonald-McGinn DM, Zackai EH, Emanuel BS, Driscoll DA, Budarf ML (2001). Mutation analysis of TBX1 in non-deleted patients with features of DGS/VCFCS or isolated cardiovascular defects. *J Med Genet.*, 38, (12): E45.
151. Goodship J, Cross I, Scambler P, Burn J (1995). Monozygotic twins with chromosome 22q11 deletion and discordant phenotype. *J Med Genet.*, 32, (9): 746-8.
152. Goshima G, Vale RD (2003). The roles of microtubule-based motor proteins in mitosis: comprehensive RNAi analysis in the *Drosophila* S2 cell line. *J Cell Biol.*, 162, (6): 1003-16.
153. Goss SJ, Harris H (1975). New method for mapping genes in human chromosomes. *Nature*, 255, (5511): 680-4.
154. Gozani O, Feld R, Reed R (1996). Evidence that sequence-independent binding of highly conserved U2 snRNP proteins upstream of the branch site is required for assembly of spliceosomal complex A. *Genes Dev.*, 10, (2): 233-43.
155. Graham FL, van der Eb AJ (1973). Transformation of rat cells by DNA of human adenovirus 5. *Virology*, 54, (2): 536-9.
156. Grate L, Ares M. Jr (2002). Searching yeast intron data at Ares lab Web site. *Methods Enzymol.*, 350: 380-92.
157. Green ED, Olson MV (1990). Chromosomal region of the cystic fibrosis gene in yeast artificial chromosomes: a model for human genome mapping. *Science*, 250, (4977): 94-8.
158. Green ED, Riethman HC, Dutchik JE, Olson MV (1991). Detection and characterization of chimeric yeast artificial-chromosome clones. *Genomics*, 11, (3): 658-69.
159. Green P (2002). Whole-genome disassembly. *Proc Natl Acad Sci U S A*, 99, (7): 4143-4.
160. Greenberg F, Elder FF, Haffner P, Northrup H, Ledbetter DH (1988). Cytogenetic findings in a prospective series of patients with DiGeorge anomaly. *Am J Hum Genet.*, 43, (5): 605-11.
161. Gregory SG, Sekhon M, Schein J, Zhao S, Osoegawa K, Scott CE, Evans RS, BurrIDGE PW, Cox TV, Fox CA, Hutton RD, Mullenger IR, Phillips KJ, Smith J, Stalker J, Threadgold GJ, Birney E, Wylie K, Chinwalla A, Wallis J, Hillier L, Carter J, Gaige T, Jaeger S, Kremitzki C, Layman D, Maas J, McGrane R, Mead K, Walker R, Jones S, Smith M, Asano J, Bosdet I, Chan S, Chittaranjan S, Chiu R, Fjell C, Fuhrmann D, Girn N, Gray C, Guin R, Hsiao L, Krzywinski M, Kutsche R, Lee SS, Mathewson C, McLeavy C, Messervier S, Ness S, Pandoh P, Prabhu AL, Saeedi P, Smailus D, Spence L, Stott J, Taylor S, Terpstra W, Tsai M, Vardy J, Wye N, Yang G, Shatsman S, Ayodeji B, Geer K, Tsegaye G, Shvartsbeyn A, Gebregeorgis E, Krol M, Russell D, Overton L, Malek JA, Holmes M, Heaney M, Shetty J, Feldblyum T, Nierman WC, Catanese JJ, Hubbard T, Waterston RH, Rogers J, de Jong PJ, Fraser CM, Marra M, McPherson JD, Bentley DR (2002). A physical map of the mouse genome. *Nature*, 418, (6899): 743-50.
162. Griffiths-Jones S (2004). The microRNA Registry. *Nucleic Acids Res.*, 32, (Database issue): D109-11.
163. Griffiths MJ, Strachan MC (1991). A single lymphocyte culture for fragile X induction and prometaphase chromosome analysis. *J Med Genet.*, 28, (12): 837-9.

164. Grimwood J, Gordon LA, Olsen A, Terry A, Schmutz J, Lamerdin J, Hellsten U, Goodstein D, Couronne O, Tran-Gyamfi M, Aerts A, Altherr M, Ashworth L, Bajorek E, Black S, Branscomb E, Caenepeel S, Carrano A, Caoile C, Chan YM, Christensen M, Cleland CA, Copeland A, Dalin E, Dehal P, Denys M, Detter JC, Escobar J, Flowers D, Fotopulos D, Garcia C, Georgescu AM, Glavina T, Gomez M, Gonzales E, Groza M, Hammon N, Hawkins T, Haydu L, Ho I, Huang W, Israni S, Jett J, Kadner K, Kimball H, Kobayashi A, Larionov V, Leem SH, Lopez F, Lou Y, Lowry S, Malfatti S, Martinez D, McCready P, Medina C, Morgan J, Nelson K, Nolan M, Ovcharenko I, Pitluck S, Pollard M, Popkie AP, Predki P, Quan G, Ramirez L, Rash S, Retterer J, Rodriguez A, Rogers S, Salamov A, Salazar A, She X, Smith D, Slezak T, Solovyev V, Thayer N, Tice H, Tsai M, Ustaszewska A, Vo N, Wagner M, Wheeler J, Wu K, Xie G, Yang J, Dubchak I, Furey TS, DeJong P, Dickson M, Gordon D, Eichler EE, Pennacchio LA, Richardson P, Stubbs L, Rokhsar DS, Myers RM, Rubin EM, Lucas SM (2004). The DNA sequence and biology of human chromosome 19. *Nature*, 428, (6982): 529-35.
165. Grunstein M, Hogness DS (1975). Colony hybridization: a method for the isolation of cloned DNAs that contain a specific gene. *Proc Natl Acad Sci U S A*, 72, (10): 3961-5.
166. Gu W, Brodtkorb E, Steinlein OK (2002). LGI1 is mutated in familial temporal lobe epilepsy characterized by aphasic seizures. *Ann Neurol.*, 52, (3): 364-7.
167. Gu W, Wevers A, Schroder H, Grzeschik KH, Derst C, Brodtkorb E, de Vos R, Steinlein OK (2002). The LGI1 gene involved in lateral temporal lobe epilepsy belongs to a new subfamily of leucine-rich repeat proteins. *FEBS Lett.*, 519, (1-3): 71-6.
168. Gu W, Gibert Y, Wirth T, Elischer A, Bloch W, Meyer A, Steinlein OK, Begemann G (2005). Using Gene-History and Expression Analyses to Assess the Involvement of LGI Genes in Human Disorders. *Mol Biol Evol.*, 22, (11): 2209-16.
169. Gu Y, Shen Y, Gibbs RA, Nelson DL (1996). Identification of FMR2, a novel gene associated with the FRAXE CCG repeat and CpG island. *Nat Genet.*, 13, (1): 109-13.
170. Guris DL, Fantes J, Tara D, Druker BJ, Imamoto A (2001). Mice lacking the homologue of the human 22q11.2 gene CRKL phenocopy neurocristopathies of DiGeorge syndrome. *Nat Genet.*, 27, (3): 293-8.
171. Guyer M (1998). Statement on the rapid release of genomic DNA sequence. *Genome Res.*, 8, (5): 413.
172. Gyapay G, Morissette J, Vignal A, Dib C, Fizames C, Millasseau P, Marc S, Bernardi G, Lathrop M, Weissenbach J (1994). The 1993-94 Genethon human genetic linkage map. *Nat Genet.*, 7, (2 Spec No): 246-339.
173. Gyapay G, Schmitt K, Fizames C, Jones H, Vega-Czarny N, Spillett D, Muselet D, Prud'homme JF, Dib C, Auffray C, Morissette J, Weissenbach J, Goodfellow PN (1996). A radiation hybrid map of the human genome. *Hum Mol Genet.*, 5, (3): 339-46.
174. Gyuris J, Golemis E, Chertkov H, Brent R (1993). Cdi1, a human G1 and S phase protein phosphatase that associates with Cdk2. *Cell*, 75, (4): 791-803.
175. Haffter P, Granato M, Brand M, Mullins MC, Hammerschmidt M, Kane DA, Odenthal J, van Eeden FJ, Jiang YJ, Heisenberg CP, Kelsh RN, Furutani-Seiki M, Vogelsang E, Beuchle D, Schach U, Fabian C, Nusslein-Volhard C (1996). The identification of genes with unique and essential functions in the development of the zebrafish, *Danio rerio*. *Development*, (123): 1-36.
176. Hagerman PJ, Hagerman RJ (2004). Fragile X-associated tremor/ataxia syndrome (FXTAS). *Ment Retard Dev Disabil Res Rev.*, 10, (1): 25-30.
177. Hagerman RJ, Leavitt BR, Farzin F, Jacquemont S, Greco CM, Brunberg JA, Tassone F, Hessler D, Harris SW, Zhang L, Jardini T, Gane LW, Ferranti J, Ruiz L, Leehey MA, Grigsby J, Hagerman PJ (2004). Fragile-X-associated tremor/ataxia syndrome (FXTAS) in females with the FMR1 premutation. *Am J Hum Genet.*, 74, (5): 1051-6.
178. Halford S, Lindsay E, Nayudu M, Carey AH, Baldini A, Scambler PJ (1993). Low-copy-number repeat sequences flank the DiGeorge/velo-cardio-facial syndrome loci at 22q11. *Hum Mol Genet.*, 2, (2): 191-6.

179. Harkin LA, Bowser DN, Dibbens LM, Singh R, Phillips F, Wallace RH, Richards MC, Williams DA, Mulley JC, Berkovic SF, Scheffer IE, Petrou S (2002). Truncation of the GABA(A)-receptor gamma2 subunit in a family with generalized epilepsy with febrile seizures plus. *Am J Hum Genet.*, 70, (2): 530-6.
180. Hattori M, Fujiyama A, Taylor TD, Watanabe H, Yada T, Park HS, Toyoda A, Ishii K, Totoki Y, Choi DK, Groner Y, Soeda E, Ohki M, Takagi T, Sakaki Y, Taudien S, Blechschmidt K, Polley A, Menzel U, Delabar J, Kumpf K, Lehmann R, Patterson D, Reichwald K, Rump A, Schillhabel M, Schudy A, Zimmermann W, Rosenthal A, Kudoh J, Schibuya K, Kawasaki K, Asakawa S, Shintani A, Sasaki T, Nagamine K, Mitsuyama S, Antonarakis SE, Minoshima S, Shimizu N, Nordsiek G, Hornischer K, Brant P, Scharfe M, Schon O, Desario A, Reichelt J, Kauer G, Blocker H, Ramser J, Beck A, Klages S, Hennig S, Riesselmann L, Dagand E, Haaf T, Wehrmeyer S, Borzym K, Gardiner K, Nizetic D, Francis F, Lehrach H, Reinhardt R, Yaspo ML; Chromosome 21 mapping and sequencing consortium. (2000). The DNA sequence of human chromosome 21. *Nature*, 405, (6784): 311-9.
181. Haug K, Warnstedt M, Alekov AK, Sander T, Ramirez A, Poser B, Maljevic S, Hebeisen S, Kubisch C, Rebstock J, Horvath S, Hallmann K, Dullinger JS, Rau B, Haverkamp F, Beyenburg S, Schulz H, Janz D, Giese B, Muller-Newen G, Propping P, Elger CE, Fahlke C, Lerche H, Heils A (2003). Mutations in *CLCN2* encoding a voltage-gated chloride channel are associated with idiopathic generalized epilepsies. *Nat Genet.*, 33, (4): 527-32.
182. Hedera P, Abou-Khalil B, Crunk AE, Taylor KA, Haines JL, Sutcliffe JS (2004). Autosomal dominant lateral temporal epilepsy: two families with novel mutations in the *LG11* gene. *Epilepsia*, 45, (3): 218-22.
183. Hedlund PB, Danielson PE, Thomas EA, Slanina K, Carson MJ, Sutcliffe JG (2003). No hypothermic response to serotonin in 5-HT7 receptor knockout mice. *Proc Natl Acad Sci U S A*, 100, (3): 1375-80.
184. Heilig R, Eckenberg R, Petit JL, Fonknechten N, Da Silva C, Cattolico L, Levy M, Barbe V, de Berardinis V, Ureta-Vidal A, Pelletier E, Vico V, Anthouard V, Rowen L, Madan A, Qin S, Sun H, Du H, Pepin K, Artiguenave F, Robert C, Cruaud C, Bruls T, Jaillon O, Friedlander L, Samson G, Brottier P, Cure S, Segurens B, Aniere F, Samain S, Crespeau H, Abbasi N, Aiach N, Boscus D, Dickhoff R, Dors M, Dubois I, Friedman C, Gouyvenoux M, James R, Madan A, Mairey-Estrada B, Mangenot S, Martins N, Menard M, Oztas S, Ratcliffe A, Shaffer T, Trask B, Vacherie B, Bellemere C, Belser C, Besnard-Gonnet M, Bartol-Mavel D, Boutard M, Briez-Silla S, Combette S, Dufosse-Laurent V, Ferron C, Lechaplais C, Louesse C, Muselet D, Magdelenat G, Pateau E, Petit E, Sirvain-Trukniewicz P, Trybou A, Vega-Czarny N, Bataille E, Bluet E, Bordelais I, Dubois M, Dumont C, Guerin T, Haffray S, Hammadi R, Muanga J, Pellouin V, Robert D, Wunderle E, Gauguet G, Roy A, Sainte-Marthe L, Verdier J, Verdier-Discala C, Hillier L, Fulton L, McPherson J, Matsuda F, Wilson R, Scarpelli C, Gyapay G, Wincker P, Saurin W, Quetier F, Waterston R, Hood L, Weissenbach J (2003). The DNA sequence and analysis of human chromosome 14. *Nature*, 421, (6923): 601-7.
185. Heiskanen M, Karhu R, Hellsten E, Peltonen L, Kallioniemi OP, Palotie A. (1994). High resolution mapping using fluorescence in situ hybridization to extended DNA fibers prepared from agarose-embedded cells. *Biotechniques*, 17, (5): 928-9, 32-3.
186. Heron SE, Crossland KM, Andermann E, Phillips HA, Hall AJ, Bleasel A, Shevell M, Mercho S, Seni MH, Guiot MC, Mulley JC, Berkovic SF, Scheffer IE (2002). Sodium-channel defects in benign familial neonatal-infantile seizures. *Lancet*, 360, (9336): 851-2.
187. Hillebrand G, Siebert R, Simeoni E, Santer R (2000). DiGeorge syndrome with discordant phenotype in monozygotic twins. *J Med Genet.*, 37, (9): E23.
188. Hillier LW, Fulton RS, Fulton LA, Graves TA, Pepin KH, Wagner-McPherson C, Layman D, Maas J, Jaeger S, Walker R, Wylie K, Sekhon M, Becker MC, O'Laughlin MD, Schaller ME, Fewell GA, Delehaunty KD, Miner TL, Nash WE, Cordes M, Du H, Sun H, Edwards J, Bradshaw-Cordum H, Ali J, Andrews S, Isak A, Vanbrunt A, Nguyen C, Du F, Lamar B, Courtney L, Kalicki J, Ozersky P, Bielicki L, Scott K, Holmes A, Harkins R, Harris A, Strong CM, Hou S, Tomlinson C, Dauphin-Kohlberg S, Kozlowicz-Reilly A, Leonard S, Rohlfing T, Rock SM, Tin-Wollam AM, Abbott A, Minx P, Maupin R, Strowmatt C, Latreille P, Miller N, Johnson D, Murray J, Woessner JP, Wendl MC, Yang SP, Schultz BR, Wallis JW, Spieth J, Bieri TA, Nelson JO, Berkowicz N, Wohldmann PE, Cook LL, Hickenbotham MT, Eldred J, Williams D, Bedell JA, Mardis ER, Clifton SW, Chissoe SL, Marra MA, Raymond C, Haugen E, Gillett W, Zhou Y, James R, Phelps K, Iadonoto S, Bubb K, Simms E, Levy R, Clendenning J,

- Kaul R, Kent WJ, Furey TS, Baertsch RA, Brent MR, Keibler E, Flicek P, Bork P, Suyama M, Bailey JA, Portnoy ME, Torrents D, Chinwalla AT, Gish WR, Eddy SR, McPherson JD, Olson MV, Eichler EE, Green ED, Waterston RH, Wilson RK (2003). The DNA sequence of human chromosome 7. *Nature*, 424, (6945): 157-64.
189. Hillier LW, Graves TA, Fulton RS, Fulton LA, Pepin KH, Minx P, Wagner-McPherson C, Layman D, Wylie K, Sekhon M, Becker MC, Fewell GA, Delehaunty KD, Miner TL, Nash WE, Kremitzki C, Oddy L, Du H, Sun H, Bradshaw-Cordum H, Ali J, Carter J, Cordes M, Harris A, Isak A, van Brunt A, Nguyen C, Du F, Courtney L, Kalicki J, Ozersky P, Abbott S, Armstrong J, Belter EA, Caruso L, Cedroni M, Cotton M, Davidson T, Desai A, Elliott G, Erb T, Fronick C, Gaige T, Haakenson W, Haglund K, Holmes A, Harkins R, Kim K, Kruchowski SS, Strong CM, Grewal N, Goyea E, Hou S, Levy A, Martinka S, Mead K, McLellan MD, Meyer R, Randall-Maher J, Tomlinson C, Dauphin-Kohlberg S, Kozlowicz-Reilly A, Shah N, Swearengen-Shahid S, Snider J, Strong JT, Thompson J, Yoakum M, Leonard S, Pearman C, Trani L, Radionenko M, Waligorski JE, Wang C, Rock SM, Tin-Wollam AM, Maupin R, Latreille P, Wendl MC, Yang SP, Pohl C, Wallis JW, Spieth J, Bieri TA, Berkowicz N, Nelson JO, Osborne J, Ding L, Meyer R, Sabo A, Shotland Y, Sinha P, Wohldmann PE, Cook LL, Hickenbotham MT, Eldred J, Williams D, Jones TA, She X, Ciccarelli FD, Izaurralde E, Taylor J, Schmutz J, Myers RM, Cox DR, Huang X, McPherson JD, Mardis ER, Clifton SW, Warren WC, Chinwalla AT, Eddy SR, Marra MA, Ovcharenko I, Furey TS, Miller W, Eichler EE, Bork P, Suyama M, Torrents D, Waterston RH, Wilson RK (2005). Generation and annotation of the DNA sequences of human chromosomes 2 and 4. *Nature*, 434, (7034): 724-31.
190. Hirasawa A, Tsumaya K, Awaji T, Katsuma S, Adachi T, Yamada M, Sugimoto Y, Miyazaki S, Tsujimoto G (2005). Free fatty acids regulate gut incretin glucagon-like peptide-1 secretion through GPR120. *Nat Med.*, 11, (1): 90-4.
191. Hirose S, Iwata H, Akiyoshi H, Kobayashi K, Ito M, Wada K, Kaneko S, Mitsudome A (1999). A novel mutation of CHRNA4 responsible for autosomal dominant nocturnal frontal lobe epilepsy. *Neurology*, 53, (8): 1749-53.
192. Hirose S, Zenri F, Akiyoshi H, Fukuma G, Iwata H, Inoue T, Yonetani M, Tsutsumi M, Muranaka H, Kurokawa T, Hanai T, Wada K, Kaneko S, Mitsudome A (2000). A novel mutation of KCNQ3 (c.925T->C) in a Japanese family with benign familial neonatal convulsions. *Ann Neurol.*, 47, (6): 822-6.
193. Hirose Y, Tacke R, Manley JL (1999). Phosphorylated RNA polymerase II stimulates pre-mRNA splicing. *Genes Dev.*, 13, (10): 1234-9.
194. Hirotsune S, Yoshida N, Chen A, Garrett L, Sugiyama F, Takahashi S, Yagami K, Wynshaw-Boris A, Yoshiki A (2003). An expressed pseudogene regulates the messenger-RNA stability of its homologous coding gene. *Nature*, 423, (6935): 91-6.
195. Hocking T, Feichtinger W, Schmid M, Haan EA, Baker E, Sutherland GR (1999). Homozygotes for FRA16B are normal. *Chromosome Res.*, 7, (7): 553-6.
196. Holmes SE, O'Hearn E, Rosenblatt A, Callahan C, Hwang HS, Ingersoll-Ashworth RG, Fleisher A, Stevanin G, Brice A, Potter NT, Ross CA, Margolis RL (2001). A repeat expansion in the gene encoding junctophilin-3 is associated with Huntington disease-like 2. *Nat Genet.*, 29, (4): 377-8.
197. Holmquist GP (1992). Chromosome bands, their chromatin flavors, and their functional features. *Am J Hum Genet.*, 51, (1): 17-37.
198. Hope IA, Albertson DG, Martinelli SD, Lynch AS, Sonnhammer E, Durbin R (1996). The *C. elegans* expression pattern database: a beginning. *Trends Genet.*, 12, (9): 370-1.
199. Horowitz DS, Krainer AR (1997). A human protein required for the second step of pre-mRNA splicing is functionally related to a yeast splicing factor. *Genes Dev.*, 11, (1): 139-51.
200. Hoskins RA, Nelson CR, Berman BP, Laverty TR, George RA, Ciesiolka L, Naeemuddin M, Arenson AD, Durbin J, David RG, Tabor PE, Bailey MR, DeShazo DR, Catanese J, Mammoser A, Osoegawa K, de Jong PJ, Celniker SE, Gibbs RA, Rubin GM, Scherer SE (2000). A BAC-based physical map of the major autosomes of *Drosophila melanogaster*. *Science*, 287, (5461): 2271-4.

201. Hu CA, Lin WW, Obie C, Valle D (1999). Molecular enzymology of mammalian Delta1-pyrroline-5-carboxylate synthase. Alternative splice donor utilization generates isoforms with different sensitivity to ornithine inhibition. *J Biol Chem.*, 274, (10): 6754-62.
202. Hudson TJ, Stein LD, Gerety SS, Ma J, Castle AB, Silva J, Slonim DK, Baptista R, Kruglyak L, Xu SH, Hu X, Colbert AM, Rosenberg C, Reeve-Daly MP, Rozen S, Hui L, Wu X, Vestergaard C, Wilson KM, Bae JS, Maitra S, Ganiatsas S, Evans CA, DeAngelis MM, Ingalls KA, Nahf RW, Horton LT Jr, Anderson MO, Collumore AJ, Ye W, Kouyoumjian V, Zemsteva IS, Tam J, Devine R, Courtney DF, Renaud MT, Nguyen H, O'Connor TJ, Fizames C, Faure S, Gyapay G, Dib C, Morissette J, Orlin JB, Birren BW, Goodman N, Weissenbach J, Hawkins TL, Foote S, Page DC, Lander ES (1995). An STS-based map of the human genome. *Science*, 270, (5244): 1945-54.
203. Huh WK, Falvo JV, Gerke LC, Carroll AS, Howson RW, Weissman JS, O'Shea EK (2003). Global analysis of protein localization in budding yeast. *Nature*, 425, (6959): 686-91.
204. Humphray SJ, Oliver K, Hunt AR, Plumb RW, Loveland JE, Howe KL, Andrews TD, Searle S, Hunt SE, Scott CE, Jones MC, Ainscough R, Almeida JP, Ambrose KD, Ashwell RI, Babbage AK, Babbage S, Bagguley CL, Bailey J, Banerjee R, Barker DJ, Barlow KF, Bates K, Beasley H, Beasley O, Bird CP, Bray-Allen S, Brown AJ, Brown JY, Burford D, Burrill W, Burton J, Carder C, Carter NP, Chapman JC, Chen Y, Clarke G, Clark SY, Clee CM, Clegg S, Collier RE, Corby N, Crosier M, Cummings AT, Davies J, Dhami P, Dunn M, Dutta I, Dyer LW, Earthrowl ME, Faulkner L, Fleming CJ, Frankish A, Frankland JA, French L, Fricker DG, Garner P, Garnett J, Ghori J, Gilbert JG, Glison C, Grafham DV, Gribble S, Griffiths C, Griffiths-Jones S, Grocock R, Guy J, Hall RE, Hammond S, Harley JL, Harrison ES, Hart EA, Heath PD, Henderson CD, Hopkins BL, Howard PJ, Howden PJ, Huckle E, Johnson C, Johnson D, Joy AA, Kay M, Keenan S, Kershaw JK, Kimberley AM, King A, Knights A, Laird GK, Langford C, Lawlor S, Leongamornlert DA, Leversha M, Lloyd C, Lloyd DM, Lovell J, Martin S, Mashreghi-Mohammadi M, Matthews L, McLaren S, McLay KE, McMurray A, Milne S, Nickerson T, Nisbett J, Nordsiek G, Pearce AV, Peck AI, Porter KM, Pandian R, Pelan S, Phillimore B, Povey S, Ramsey Y, Rand V, Scharfe M, Sehra HK, Shownkeen R, Sims SK, Skuce CD, Smith M, Steward CA, Swarbreck D, Sycamore N, Tester J, Thorpe A, Tracey A, Tromans A, Thomas DW, Wall M, Wallis JM, West AP, Whitehead SL, Willey DL, Williams SA, Wilming L, Wray PW, Young L, Ashurst JL, Coulson A, Blocker H, Durbin R, Sulston JE, Hubbard T, Jackson MJ, Bentley DR, Beck S, Rogers J, Dunham I (2004). DNA sequence and analysis of human chromosome 9. *Nature*, 429, (6990): 369-74.
205. Hurst LD, Pal C, Lercher MJ (2004). The evolutionary dynamics of eukaryotic gene order. *Nat Rev Genet.*, 5, (4): 299-310.
206. Igel H, Wells S, Perriman R, Ares M. Jr (1998). Conservation of structure and subunit interactions in yeast homologues of splicing factor 3b (SF3b) subunits. *RNA*, 4, (1): 1-10.
207. Ikeda A, Kunieda T, Miyamoto S, Fukuyama H, Shibasaki H (2000). Autosomal dominant temporal lobe epilepsy in a Japanese family. *J Neurol Sci.*, 176, (2): 162-5.
208. International HapMap Consortium (2003). The International HapMap Project. *Nature*, 426, (6968): 789-96.
209. International Human Genome Sequencing Consortium (2004). Finishing the euchromatic sequence of the human genome. *Nature*, 431, (7011): 931-45.
210. Ioannou PA, Amemiya CT, Ganes J, Kroisel PM, Shizuya H, Chen C, Batzer MA, de Jong PJ (1994). A new bacteriophage P1-derived vector for the propagation of large human DNA fragments. *Nat Genet.*, 6, (1): 84-9.
211. Isono K, Mizutani-Koseki Y, Komori T, Schmidt-Zachmann MS, Koseki H (2005). Mammalian polycomb-mediated repression of Hox genes requires the essential spliceosomal protein Sf3b1. *Genes Dev.*, 19, (5): 536-41.
212. Jaillon O, Aury JM, Brunet F, Petit JL, Stange-Thomann N, Mauceli E, Bouneau L, Fischer C, Ozouf-Costaz C, Bernot A, Nicaud S, Jaffe D, Fisher S, Luffalla G, Dossat C, Segurens B, Dasilva C, Salanoubat M, Levy M, Boudet N, Castellano S, Anthonard V, Jubin C, Castelli V, Katinka M, Vacherie B, Biemont C, Skalli Z, Cattolico L, Poulain J, De Berardinis V, Cruaud C, Duprat S, Brottier P, Coutanceau JP, Gouzy J, Parra G, Lardier G, Chapple C, McKernan KJ, McEwan P, Bosak S, Kellis M, Volf JN, Guigo R, Zody MC, Mesirov J, Lindblad-Toh K, Birren B, Nusbaum C, Kahn D, Robinson-Rechavi M, Laudet V, Schachter V, Quetier F, Saurin W,

- Scarpelli C, Wincker P, Lander ES, Weissenbach J, Roest Crolius H (2004). Genome duplication in the teleost fish *Tetraodon nigroviridis* reveals the early vertebrate proto-karyotype. *Nature*, 431, (7011): 946-57.
213. James MR, Lindpaintner K (1997). Why map the rat? *Trends Genet.*, 13, (5): 171-3.
214. Jarvis CD, Geiman T, Vila-Storm MP, Osipovich O, Akella U, Candeias S, Nathan I, Durum SK, Muegge K. (1996). A novel putative helicase produced in early murine lymphocytes. *Gene*, 169, (2): 203-7.
215. Jeffreys AJ, Wilson V, Thein SL (1985). Hypervariable 'minisatellite' regions in human DNA. *Nature*, 314, (6006): 67-73.
216. Jensen-Seaman MI, Furey TS, Payseur BA, Lu Y, Roskin KM, Chen CF, Thomas MA, Haussler D, Jacob HJ (2004). Comparative recombination rates in the rat, mouse and human genomes. *Genome Res.*, 14, (4): 528-38.
217. Jerome LA, Papaioannou VE (2001). DiGeorge syndrome phenotype in mice mutant for the T-box gene, *Tbx1*. *Nat Genet.*, 27, (3): 286-91.
218. Jin Q, van Eynde A, Beullens M, Roy N, Thiel G, Stalmans W, Bollen M (2003). The protein phosphatase-1 (PP1) regulator, nuclear inhibitor of PP1 (NIPP1), interacts with the polycomb group protein, embryonic ectoderm development (EED), and functions as a transcriptional repressor. *J Biol Chem.*, 278, (33): 30677-85.
219. Jin P, Warren ST (2003). New insights into fragile X syndrome: from molecules to neurobehaviors. *Trends Biochem Sci.*, 28, (3): 152-8.
220. Johnson JM, Castle J, Garrett-Engele P, Kan Z, Loerch PM, Armour CD, Santos R, Schadt EE, Stoughton R, Shoemaker DD (2003). Genome-wide survey of human alternative pre-mRNA splicing with exon junction microarrays. *Science*, 302, (5653): 2141-4.
221. Jurica MS, Licklider LJ, Gygi SR, Grigorieff N, Moore MJ (2002). Purification and characterization of native spliceosomes suitable for three-dimensional structural analysis. *RNA*, 8, (4): 426-39.
222. Kai W, Kikuchi K, Fujita M, Suetake H, Fujiwara A, Yoshiura Y, Ototake M, Venkatesh B, Miyaki K, Suzuki Y (2005). A genetic linkage map for the tiger pufferfish, *Takifugu rubripes*. *Genetics*, 171(1):227-38
223. Kalachikov S, Evgrafov O, Ross B, Winawer M, Barker-Cummings C, Martinelli Boneschi F, Choi C, Morozov P, Das K, Teplitskaya E, Yu A, Cayanis E, Penchaszadeh G, Kottmann AH, Pedley TA, Hauser WA, Ottman R, Gilliam TC (2002). Mutations in *LGI1* cause autosomal-dominant partial epilepsy with auditory features. *Nat Genet.*, 30, (3): 335-41.
224. Kamath RS, Fraser AG, Dong Y, Poulin G, Durbin R, Gotta M, Kanapin A, Le Bot N, Moreno S, Sohrmann M, Welchman DP, Zipperlen P, Ahringer J (2003). Systematic functional analysis of the *Caenorhabditis elegans* genome using RNAi. *Nature*, 421, (6920): 231-7.
225. Kampa D, Cheng J, Kapranov P, Yamanaka M, Brubaker S, Cawley S, Drenkow J, Piccolboni A, Bekiranov S, Helt G, Tammana H, Gingeras TR (2004). Novel RNAs identified from an in-depth analysis of the transcriptome of human chromosomes 21 and 22. *Genome Res.*, 14, (3): 331-42.
226. Kan YW, Dozy AM (1978). Polymorphism of DNA sequence adjacent to human beta-globin structural gene: relationship to sickle mutation. *Proc Natl Acad Sci U S A*, 75, (11): 5631-5.
227. Kananura C, Haug K, Sander T, Runge U, Gu W, Hallmann K, Rebstock J, Heils A, Steinlein OK (2002). A splice-site mutation in *GABRG2* associated with childhood absence epilepsy and febrile convulsions. *Arch Neurol.*, 59, (7): 1137-41.
228. Kanemoto K, Kawasaki J (2000). Familial aphasic episodes: another variant of partial epilepsy with simple inheritance? *Epilepsia*, 41, (8): 1036-8.
229. Khaitovich P, Hellmann I, Enard W, Nowick K, Leinweber M, Franz H, Weiss G, Lachmann M, Paabo S (2005). Parallel patterns of evolution in the genomes and transcriptomes of humans and chimpanzees. *Science*, 309, (5742): 1850-4.

230. Kiger AA, Baum B, Jones S, Jones MR, Coulson A, Echeverri C, Perrimon N (2003). A functional genomic analysis of cell morphology using RNA interference. *J Biol.*, 2, (4): 27.
231. Kim UJ, Shizuya H, de Jong PJ, Birren B, Simon MI (1992). Stable propagation of cosmid sized human DNA inserts in an F factor based vector. *Nucleic Acids Res.*, 20, (5): 1083-5.
232. Kimber WL, Hsieh P, Hirotsune S, Yuva-Paylor L, Sutherland HF, Chen A, Ruiz-Lozano P, Hoogstraten-Miller SL, Chien KR, Paylor R, Scambler PJ, Wynshaw-Boris A (1999). Deletion of 150 kb in the minimal DiGeorge/velocardiofacial syndrome critical region in mouse. *Hum Mol Genet.*, 8, (12): 2229-37.
233. King M.C, Wilson AC (1975). Evolution at two levels in Humans and Chimpanzees. *Science*, 188: 107-16.
234. Kishimoto H, Hamada K, Saunders M, Backman S, Sasaki T, Nakano T, Mak TW, Suzuki A (2003). Physiological functions of Pten in mouse tissues. *Cell Struct Funct.*, 28, (1): 11-21.
235. Kiyosawa H, Abe K (2002). Speculations on the role of natural antisense transcripts in mammalian X chromosome evolution. *Cytogenet Genome Res.*, 99, (1-4): 151-6.
236. Kobayashi E, Santos NF, Torres FR, Secolin R, Sardinha LA, Lopez-Cendes I, Cendes F (2003). Magnetic resonance imaging abnormalities in familial temporal lobe epilepsy with auditory auras. *Arch Neurol.*, 60, (11): 1546-51.
237. Kong A, Gudbjartsson DF, Sainz J, Jonsdottir GM, Gudjonsson SA, Richardsson B, Sigurdardottir S, Barnard J, Hallbeck B, Masson G, Shlien A, Palsson ST, Frigge ML, Thorgeirsson TE, Gulcher JR, Stefansson K (2002). A high-resolution recombination map of the human genome. *Nat Genet.*, 31, (3): 241-7.
238. Kozak M (1986). Point mutations define a sequence flanking the AUG initiator codon that modulates translation by eukaryotic ribosomes. *Cell*, 44, (2): 283-92.
239. Krummel KA, Denison SR, Calhoun E, Phillips LA, Smith DI (2002). The common fragile site FRA16D and its associated gene WWOX are highly conserved in the mouse at Fra8E1. *Genes Chromosomes Cancer.*, 34, (2): 154-67.
240. Kumar A, Agarwal S, Heyman JA, Matson S, Heidtman M, Piccirillo S, Umansky L, Drawid A, Jansen R, Liu Y, Cheung KH, Miller P, Gerstein M, Roeder GS, Snyder M (2002). Subcellular localization of the yeast proteome. *Genes Dev.*, 16, (6): 707-19.
241. Kurahashi H, Tsuda E, Kohama R, Nakayama T, Masuno M, Imaizumi K, Kamiya T, Sano T, Okada S, Nishisho I (1997). Another critical region for deletion of 22q11: a study of 100 patients. *Am J Med Genet.*, 72, (2): 180-5.
242. Kwek KY, Murphy S, Furger A, Thomas B, O'Gorman W, Kimura H, Proudfoot NJ, Akoulitchev A (2002). U1 snRNA associates with TFIIF and regulates transcriptional initiation. *Nat Struct Biol.*, 9, (11): 800-5.
243. Ladd AN, Cooper TA (2002). Finding signals that regulate alternative splicing in the post-genomic era. *Genome Biol.*, 3, (11): reviews0008.
244. Lamesch P, Milstein S, Hao T, Rosenberg J, Li N, Sequerra R, Bosak S, Doucette-Stamm L, Vandenhoute J, Hill DE, Vidal M (2004). C. elegans ORFeome version 3.1: increasing the coverage of ORFeome resources with improved gene predictions. *Genome Res.*, 14, (10B): 2064-9.
245. Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, Devon K, Dewar K, Doyle M, FitzHugh W, Funke R, Gage D, Harris K, Heaford A, Howland J, Kann L, Lehoczky J, LeVine R, McEwan P, McKernan K, Meldrim J, Mesirov JP, Miranda C, Morris W, Naylor J, Raymond C, Rosetti M, Santos R, Sheridan A, Sougnez C, Stange-Thomann N, Stojanovic N, Subramanian A, Wyman D, Rogers J, Sulston J, Ainscough R, Beck S, Bentley D, Burton J, Clee C, Carter N, Coulson A, Deadman R, Deloukas P, Dunham A, Dunham I, Durbin R, French L, Grafham D, Gregory S, Hubbard T, Humphray S, Hunt A, Jones M, Lloyd C, McMurray A, Matthews L, Mercer S, Milne S, Mullikin JC, Mungall A, Plumb R, Ross M, Shownkeen R, Sims S, Waterston RH, Wilson RK, Hillier LW, McPherson JD, Marra MA, Mardis ER, Fulton LA, Chinwalla AT, Pepin KH, Gish WR, Chissole SL, Wendl MC, Delehaunty

- KD, Miner TL, Delehaunty A, Kramer JB, Cook LL, Fulton RS, Johnson DL, Minx PJ, Clifton SW, Hawkins T, Branscomb E, Predki P, Richardson P, Wenning S, Slezak T, Doggett N, Cheng JF, Olsen A, Lucas S, Elkin C, Uberbacher E, Frazier M, Gibbs RA, Muzny DM, Scherer SE, Bouck JB, Sodergren EJ, Worley KC, Rives CM, Gorrell JH, Metzker ML, Naylor SL, Kucherlapati RS, Nelson DL, Weinstock GM, Sakaki Y, Fujiyama A, Hattori M, Yada T, Toyoda A, Itoh T, Kawagoe C, Watanabe H, Totoki Y, Taylor T, Weissbach J, Heilig R, Saurin W, Artiguenave F, Brottier P, Bruls T, Pelletier E, Robert C, Wincker P, Smith DR, Doucette-Stamm L, Rubenfield M, Weinstock K, Lee HM, Dubois J, Rosenthal A, Platzer M, Nyakatura G, Taudien S, Rump A, Yang H, Yu J, Wang J, Huang G, Gu J, Hood L, Rowen L, Madan A, Qin S, Davis RW, Federspiel NA, Abola AP, Proctor MJ, Myers RM, Schmutz J, Dickson M, Grimwood J, Cox DR, Olson MV, Kaul R, Raymond C, Shimizu N, Kawasaki K, Minoshima S, Evans GA, Athanasiou M, Schultz R, Roe BA, Chen F, Pan H, Ramsay J, Lehrach H, Reinhardt R, McCombie WR, de la Bastide M, Dedhia N, Blocker H, Hornischer K, Nordsiek G, Agarwala R, Aravind L, Bailey JA, Bateman A, Batzoglou S, Birney E, Bork P, Brown DG, Burge CB, Cerutti L, Chen HC, Church D, Clamp M, Copley RR, Doerks T, Eddy SR, Eichler EE, Furey TS, Galagan J, Gilbert JG, Harmon C, Hayashizaki Y, Haussler D, Hermjakob H, Hokamp K, Jang W, Johnson LS, Jones TA, Kasif S, Kasprzyk A, Kennedy S, Kent WJ, Kitts P, Koonin EV, Korfi I, Kulp D, Lancet D, Lowe TM, McLysaght A, Mikkelsen T, Moran JV, Mulder N, Pollara VJ, Ponting CP, Schuler G, Schultz J, Slater G, Smit AF, Stupka E, Szustakowski J, Thierry-Mieg D, Thierry-Mieg J, Wagner L, Wallis J, Wheeler R, Williams A, Wolf YI, Wolfe KH, Yang SP, Yeh RF, Collins F, Guyer MS, Peterson J, Felsenfeld A, Wetterstrand KA, Patrino A, Morgan MJ, de Jong P, Catanese JJ, Osoegawa K, Shizuya H, Choi S, Chen YJ; International Human Genome Sequencing Consortium (2001). Initial sequencing and analysis of the human genome. *Nature*, 409, (6822): 860-921.
246. Larsen F, Gundersen G, Lopez R, Prydz H (1992). CpG islands as gene markers in the human genome. *Genomics*, 13, (4): 1095-107.
247. Lavorgna G, Dahary D, Lehner B, Sorek R, Sanderson CM, Casari G (2004). In search of antisense. *Trends Biochem Sci.*, 29, (2): 88-94.
248. Lawrence JB, Singer RH (1985). Quantitative analysis of in situ hybridization methods for the detection of actin gene expression. *Nucleic Acids Res.*, 13, (5): 1777-99.
249. Leana-Cox J, Pangkanon S, Eanet KR, Curtin MS, Wulfsberg EA (1996). Familial DiGeorge/velocardiofacial syndrome with deletions of chromosome area 22q11.2: report of five families with a review of the literature. *Am J Med Genet.*, 65, (4): 309-16.
250. Lee Y, Kim M, Han J, Yeom KH, Lee S, Baek SH, Kim VN (2004). MicroRNA genes are transcribed by RNA polymerase II. *EMBO J.*, 23, (20): 4051-60.
251. Leehey MA, Munhoz RP, Lang AE, Brunberg JA, Grigsby J, Greco C, Jacquemont S, Tassone F, Lozano AM, Hagerman PJ, Hagerman RJ (2003). The fragile X premutation presenting as essential tremor. *Arch Neurol.*, 60, (1): 117-21.
252. Lewin, B. *Genes IV* 466–481 (Oxford University Press, Oxford, 1990).
253. Li S, Armstrong CM, Bertin N, Ge H, Milstein S, Boxem M, Vidalain PO, Han JD, Chesneau A, Hao T, Goldberg DS, Li N, Martinez M, Rual JF, Lamesch P, Xu L, Tewari M, Wong SL, Zhang LV, Berriz GF, Jacotot L, Vaglio P, Reboul J, Hirozane-Kishikawa T, Li Q, Gabel HW, Elewa A, Baumgartner B, Rose DJ, Yu H, Bosak S, Sequerra R, Fraser A, Mango SE, Saxton WM, Strome S, Van Den Heuvel S, Piano F, Vandenhaute J, Sardet C, Gerstein M, Doucette-Stamm L, Gunsalus KC, Harper JW, Cusick ME, Roth FP, Hill DE, Vidal M (2004). A map of the interactome network of the metazoan *C. elegans*. *Science*, 303, (5657): 540-3.
254. Li T, Spearow J, Rubin CM, Schmid CW (1999). Physiological stresses increase mouse short interspersed element (SINE) RNA expression in vivo. *Gene*, 239, (2): 367-72.
255. Liang F, Holt I, Pertea G, Karamycheva S, Salzberg SL, Quackenbush J (2000). Gene index analysis of the human genome estimates approximately 120,000 genes. *Nat Genet.*, 25, (2): 239-40.
256. Lindsay EA, Goldberg R, Jurecic V, Morrow B, Carlson C, Kucherlapati RS, Shprintzen RJ, Baldini A (1995). Velo-cardio-facial syndrome: frequency and extent of 22q11 deletions. *Am J Med Genet.*, 57, (3): 514-22.

257. Lindsay EA, Harvey EL, Scambler PJ, Baldini A (1998). ES2, a gene deleted in DiGeorge syndrome, encodes a nuclear protein and is expressed during early mouse development, where it shares an expression domain with a Goosecoid-like gene. *Hum Mol Genet.*, 7, (4): 629-35.
258. Lindsay EA, Botta A, Jurecic V, Carattini-Rivera S, Cheah YC, Rosenblatt HM, Bradley A, Baldini A (1999). Congenital heart disease in mice deficient for the DiGeorge syndrome region. *Nature*, 401, (6751): 379-83.
259. Lindsay EA, Vitelli F, Su H, Morishima M, Huynh T, Pramparo T, Jurecic V, Ogunrinu G, Sutherland HF, Scambler PJ, Bradley A, Baldini A (2001). Tbx1 haploinsufficiency in the DiGeorge syndrome region causes aortic arch defects in mice. *Nature*, 410, (6824): 97-101.
260. Liu WM, Chu WM, Choudary PV, Schmid CW (1995). Cell stress and translational inhibitors transiently increase the abundance of mammalian SINE transcripts. *Nucleic Acids Res.*, 23, (10): 1758-65.
261. Lopez I, Mak EC, Ding J, Hamm HE, Lomasney JW (2001). A novel bifunctional phospholipase c that is regulated by Galpha 12 and stimulates the Ras/mitogen-activated protein kinase pathway. *J Biol Chem.*, 276, (4): 2758-65.
262. Lutzmann M, Kunze R, Stangl K, Stelter P, Toth KF, Bottcher B, Hurt E (2005). Reconstitution of Nup157 and Nup145N into the Nup84 complex. *J Biol Chem.*, 280, (18): 18442-51.
263. Maeda I, Kohara Y, Yamamoto M, Sugimoto A (2001). Large-scale analysis of gene function in *Caenorhabditis elegans* by high-throughput RNAi. *Curr Biol.*, 11, (3): 171-6.
264. Makarova OV, Makarov EM, Urlaub H, Will CL, Gentzel M, Wilm M, Luhrmann R (2004). A subset of human 35S U5 proteins, including Prp19, function prior to catalytic step 1 of splicing. *EMBO J.*, 23, (12): 2381-91.
265. Malhotra JD, Kazen-Gillespie K, Hortsch M, Isom LL (2000). Sodium channel beta subunits mediate homophilic cell adhesion and recruit ankyrin to points of cell-cell contact. *J Biol Chem.*, 275, (15): 11383-8.
266. Mallon AM, Platzer M, Bate R, Gloeckner G, Botcherby MR, Nordsiek G, Strivens MA, Kioschis P, Dangel A, Cunningham D, Straw RN, Weston P, Gilbert M, Fernando S, Goodall K, Hunter G, Greystrom JS, Clarke D, Kimberley C, Goerdes M, Blechschmidt K, Rump A, Hinzmann B, Mundy CR, Miller W, Poustka A, Herman GE, Rhodes M, Denny P, Rosenthal A, Brown SD (2000). Comparative genome sequence analysis of the Bpa/Str region in mouse and Man. *Genome Res.*, 10, (6): 758-75.
267. Margulies EH, Blanchette M, Haussler D, Green ED; NISC Comparative Sequencing Program (2003). Identification and characterization of multi-species conserved sequences. *Genome Res.*, 13, (12): 2507-18.
268. Marra M, Hillier L, Kucaba T, Allen M, Barstead R, Beck C, Blistain A, Bonaldo M, Bowers Y, Bowles L, Cardenas M, Chamberlain A, Chappell J, Clifton S, Favello A, Geisel S, Gibbons M, Harvey N, Hill F, Jackson Y, Kohn S, Lennon G, Mardis E, Martin J, Mila L, McCann R, Morales R, Pape D, Person B, Prange C, Ritter E, Soares M, Schurk R, Shin T, Steptoe M, Swaller T, Theising B, Underwood K, Wylie T, Yount T, Wilson R, Waterston R (1999). An encyclopedia of mouse genes. *Nat Genet.*, 21, (2): 191-4.
269. Marra MA, Kucaba TA, Dietrich NL, Green ED, Brownstein B, Wilson RK, McDonald KM, Hillier LW, McPherson JD, Waterston RH (1997). High throughput fingerprint analysis of large-insert clones. *Genome Res.*, 7, (11): 1072-84.
270. Martin J, Han C, Gordon LA, Terry A, Prabhakar S, She X, Xie G, Hellsten U, Chan YM, Altherr M, Couronne O, Aerts A, Bajorek E, Black S, Blumer H, Branscomb E, Brown NC, Bruno WJ, Buckingham JM, Callen DF, Campbell CS, Campbell ML, Campbell EW, Caoile C, Challacombe JF, Chasteen LA, Chertkov O, Chi HC, Christensen M, Clark LM, Cohn JD, Denys M, Detter JC, Dickson M, Dimitrijevic-Bussod M, Escobar J, Fawcett JJ, Flowers D, Fotopulos D, Glavina T, Gomez M, Gonzales E, Goodstein D, Goodwin LA, Grady DL, Grigoriev I, Groza M, Hammon N, Hawkins T, Haydu L, Hildebrand CE, Huang W, Israni S, Jett J, Jewett PB, Kadner K, Kimball H, Kobayashi A, Krawczyk MC, Leyba T, Longmire JL, Lopez F, Lou Y, Lowry S, Ludeman T, Manohar CF, Mark GA, McMurray KL, Meincke LJ, Morgan J, Moyzis RK, Mundt MO, Munk AC, Nandkeshwar RD, Pitluck S, Pollard M, Predki P, Parson-

- Quintana B, Ramirez L, Rash S, Retterer J, Ricke DO, Robinson DL, Rodriguez A, Salamov A, Saunders EH, Scott D, Shough T, Stallings RL, Stalvey M, Sutherland RD, Tapia R, Tesmer JG, Thayer N, Thompson LS, Tice H, Torney DC, Tran-Gyamfi M, Tsai M, Ulanovsky LE, Ustaszewska A, Vo N, White PS, Williams AL, Wills PL, Wu JR, Wu K, Yang J, Dejong P, Bruce D, Doggett NA, Deaven L, Schmutz J, Grimwood J, Richardson P, Rokhsar DS, Eichler EE, Gilna P, Lucas SM, Myers RM, Rubin EM, Pennacchio LA (2004). The sequence and analysis of duplication-rich human chromosome 16. *Nature*, 432, (7020): 988-94.
271. Martinez E, Palhan VB, Tjernberg A, Lyman ES, Gamper AM, Kundu TK, Chait BT, Roeder RG (2001). Human STAGA complex is a chromatin-acetylating transcription coactivator that interacts with pre-mRNA splicing and DNA damage-binding factors in vivo. *Mol Cell Biol.*, 21, (20): 6782-95.
272. Marzella R, Kokkinaki MA, Kapsetaki M, Ricco A, Argyrokastritis A, Kamakari S, Sarafidou T, Archidiacono N, Roussou A, Pasparaki A, Rocchi M, Moschonas NK (1997). Map integration at human chromosome 10: molecular and cytogenetic analysis of a chromosome-specific somatic cell hybrid panel and genomic clones, based on a well-supported genetic map. *Cytogenet Cell Genet.*, 79, (3-4): 257-65.
273. Mautner VF, Lindenau M, Gottesleben A, Goetze G, Kluwe L (2000). Supporting evidence of a gene for partial epilepsy on 10q. *Neurogenetics*, 3, (1): 31-4.
274. McAllister BF, Greenbaum IF (1997). How common are common fragile sites: variation of aphidicolin-induced chromosomal fragile sites in a population of the deer mouse (*Peromyscus maniculatus*). *Hum Genet.*, 100, (2): 182-8.
275. McCarthy LC, Bihoreau MT, Kiguwa SL, Browne J, Watanabe TK, Hishigaki H, Tsuji A, Kiel S, Webber C, Davis ME, Knights C, Smith A, Critcher R, Huxtall P, Hudson JR Jr, Ono T, Hayashi H, Takagi T, Nakamura Y, Tanigami A, Goodfellow PN, Lathrop GM, James MR (2000). A whole-genome radiation hybrid panel and framework map of the rat genome. *Mamm Genome.*, 11, (9): 791-5.
276. McCombie WR, Adams MD, Kelley JM, FitzGerald MG, Utterback TR, Khan M, Dubnick M, Kerlavage AR, Venter JC, Fields C (1992). *Caenorhabditis elegans* expressed sequence tags identify gene families and potential disease gene homologues. *Nat Genet.*, 1, (2): 124-31.
277. McCracken S, Fong N, Yankulov K, Ballantyne S, Pan G, Greenblatt J, Patterson SD, Wickens M, Bentley DL (1997). The C-terminal domain of RNA polymerase II couples mRNA processing to transcription. *Nature*, 385, (6614): 357-61.
278. McLean SD, Saal HM, Spinner NB, Emanuel BS, Driscoll DA (1993). Velo-cardio-facial syndrome. Intrafamilial variability of the phenotype. *Am J Dis Child.*, 147, (11): 1212-6.
279. McLellan A, Phillips HA, Rittey C, Kirkpatrick M, Mulley JC, Goudie D, Stephenson JB, Tolmie J, Scheffer IE, Berkovic SF, Zuberi SM (2003). Phenotypic comparison of two Scottish families with mutations in different genes causing autosomal dominant nocturnal frontal lobe epilepsy. *Epilepsia*, 44, (4): 613-7.
280. McPherson JD, Marra M, Hillier L, Waterston RH, Chinwalla A, Wallis J, Sekhon M, Wylie K, Mardis ER, Wilson RK, Fulton R, Kucaba TA, Wagner-McPherson C, Barbazuk WB, Gregory SG, Humphray SJ, French L, Evans RS, Bethel G, Whittaker A, Holden JL, McCann OT, Dunham A, Soderlund C, Scott CE, Bentley DR, Schuler G, Chen HC, Jang W, Green ED, Idol JR, Maduro VV, Montgomery KT, Lee E, Miller A, Emerling S, Kucherlapati, Gibbs R, Scherer S, Gorrell JH, Sodergren E, Clerc-Blankenburg K, Tabor P, Naylor S, Garcia D, de Jong PJ, Catanese JJ, Nowak N, Osoegawa K, Qin S, Rowen L, Madan A, Dors M, Hood L, Trask B, Friedman C, Massa H, Cheung VG, Kirsch IR, Reid T, Yonescu R, Weissenbach J, Bruls T, Heilig R, Branscomb E, Olsen A, Doggett N, Cheng JF, Hawkins T, Myers RM, Shang J, Ramirez L, Schmutz J, Velasquez O, Dixon K, Stone NE, Cox DR, Haussler D, Kent WJ, Furey T, Rogic S, Kennedy S, Jones S, Rosenthal A, Wen G, Schilhabel M, Gloeckner G, Nyakatura G, Siebert R, Schlegelberger B, Korenberg J, Chen XN, Fujiyama A, Hattori M, Toyoda A, Yada T, Park HS, Sakaki Y, Shimizu N, Asakawa S, Kawasaki K, Sasaki T, Shintani A, Shimizu A, Shibuya K, Kudoh J, Minoshima S, Ramser J, Seranski P, Hoff C, Poustka A, Reinhardt R, Lehrach H; International Human Genome Mapping Consortium. (2001). A physical map of the human genome. *Nature*, 409, (6822): 934-41.

281. McQuade L, Christodoulou J, Budarf M, Sachdev R, Wilson M, Emanuel B, Colley A (1999). Patient with a 22q11.2 deletion with no overlap of the minimal DiGeorge syndrome critical region (MDGCR). *Am J Med Genet.*, 86, (1): 27-33.
282. Merscher S, Funke B, Epstein JA, Heyer J, Puech A, Lu MM, Xavier RJ, Demay MB, Russell RG, Factor S, Tokooya K, Jore BS, Lopez M, Pandita RK, Lia M, Carrion D, Xu H, Schorle H, Kobler JB, Scambler P, Wynshaw-Boris A, Skoultschi AI, Morrow BE, Kucherlapati R (2001). TBX1 is responsible for cardiovascular defects in velo-cardio-facial/DiGeorge syndrome. *Cell*, 104, (4): 619-29.
283. Michaelis RC, Velagaleti GV, Jones C, Pivnick EK, Phelan MC, Boyd E, Tarleton J, Wilroy RS, Tunnacliffe A, Tharapel AT (1998). Most Jacobsen syndrome deletion breakpoints occur distal to FRA11B. *Am J Med Genet.*, 76, (3): 222-8.
284. Michelucci R, Passarelli D, Pitzalis S, Dal Corso G, Tassinari CA, Nobile C (2000). Autosomal dominant partial epilepsy with auditory features: description of a new family. *Epilepsia*, 41, (8): 967-70.
285. Michelucci R, Poza JJ, Sofia V, de Feo MR, Binelli S, Bisulli F, Scudellaro E, Simionati B, Zimbello R, D'Orsi G, Passarelli D, Avoni P, Avanzini G, Tinuper P, Biondi R, Valle G, Mautner VF, Stephani U, Tassinari CA, Moschonas NK, Siebert R, Lopez de Munain A, Perez-Tur J, Nobile C (2003). Autosomal dominant lateral temporal epilepsy: clinical spectrum, new epitempin mutations, and genetic heterogeneity in seven European families. *Epilepsia*, 44, (10): 1289-97.
286. Modrek B, Lee C (2002). A genomic view of alternative splicing. *Nat Genet.*, 30, (1): 13-9.
287. Modrek B, Lee CJ (2003). Alternative splicing in the human, mouse and rat genomes is associated with an increased frequency of exon creation and/or loss. *Nat Genet.*, 34, (2): 177-80.
288. Morante-Redolat JM, Gorostidi-Pagola A, Piquer-Sirerol S, Saenz A, Poza JJ, Galan J, Gesk S, Sarafidou T, Mautner VF, Binelli S, Staub E, Hinzmann B, French L, Prud'homme JF, Passarelli D, Scannapieco P, Tassinari CA, Avanzini G, Marti-Masso JF, Kluwe L, Deloukas P, Moschonas NK, Michelucci R, Siebert R, Nobile C, Perez-Tur J, Lopez de Munain A (2002). Mutations in the LGI1/Epitempin gene on 10q24 cause autosomal dominant lateral temporal epilepsy. *Hum Mol Genet.*, 11, (9): 1119-28.
289. Morris DP, Greenleaf AL (2000). The splicing factor, Prp40, binds the phosphorylated carboxyl-terminal domain of RNA polymerase II. *J Biol Chem.*, 275, (51): 39935-43.
290. Moschonas NK (2003). Chromosome 10. *Nature Encyclopedia of the Human Genome*, Edit. D.N. Cooper, Nature Publishing Group (ISBN 0-333-80386-8)
291. Mungall AJ, Palmer SA, Sims SK, Edwards CA, Ashurst JL, Wilming L, Jones MC, Horton R, Hunt SE, Scott CE, Gilbert JG, Clamp ME, Bethel G, Milne S, Ainscough R, Almeida JP, Ambrose KD, Andrews TD, Ashwell RI, Babbage AK, Bagguley CL, Bailey J, Banerjee R, Barker DJ, Barlow KF, Bates K, Beare DM, Beasley H, Beasley O, Bird CP, Blakey S, Bray-Allen S, Brook J, Brown AJ, Brown JY, Burford DC, Burrill W, Burton J, Carder C, Carter NP, Chapman JC, Clark SY, Clark G, Clee CM, Clegg S, Copley V, Collier RE, Collins JE, Colman LK, Corby NR, Coville GJ, Culley KM, Dhami P, Davies J, Dunn M, Earthrowl ME, Ellington AE, Evans KA, Faulkner L, Francis MD, Frankish A, Frankland J, French L, Garner P, Garnett J, Ghori MJ, Gilby LM, Gillson CJ, Glithero RJ, Grafham DV, Grant M, Gribble S, Griffiths C, Griffiths M, Hall R, Halls KS, Hammond S, Harley JL, Hart EA, Heath PD, Heathcote R, Holmes SJ, Howden PJ, Howe KL, Howell GR, Huckle E, Humphray SJ, Humphries MD, Hunt AR, Johnson CM, Joy AA, Kay M, Keenan SJ, Kimberley AM, King A, Laird GK, Langford C, Lawlor S, Leongamornlert DA, Lerversha M, Lloyd CR, Lloyd DM, Loveland JE, Lovell J, Martin S, Mashreghi-Mohammadi M, Maslen GL, Matthews L, McCann OT, McLaren SJ, McLay K, McMurray A, Moore MJ, Mullikin JC, Niblett D, Nickerson T, Novik KL, Oliver K, Overton-Larty EK, Parker A, Patel R, Pearce AV, Peck AI, Phillimore B, Phillips S, Plumb RW, Porter KM, Ramsey Y, Ranby SA, Rice CM, Ross MT, Searle SM, Sehra HK, Sheridan E, Skuce CD, Smith S, Smith M, Spraggon L, Squares SL, Steward CA, Sycamore N, Tamlyn-Hall G, Tester J, Theaker AJ, Thomas DW, Thorpe A, Tracey A, Tromans A, Tubby B, Wall M, Wallis JM, West AP, White SS, Whitehead SL, Whittaker H, Wild A, Willey DJ, Wilmer TE, Wood JM, Wray PW, Wyatt JC, Young L, Younger RM, Bentley DR, Coulson A, Durbin R, Hubbard T, Sulston JE, Dunham I, Rogers J, Beck S (2003). The DNA sequence and analysis of human chromosome 6. *Nature*, 425, (6960): 805-11.

292. Murray JC, Buetow KH, Weber JL, Ludwigsen S, Scherpbier-Heddema T, Manion F, Quillen J, Sheffield VC, Sunden S, Duyk GM et al (1994). A comprehensive human linkage map with centimorgan density. Cooperative Human Linkage Center (CHLC). *Science*, 265, (5181): 2049-54.
293. Musio A, Montagna C, Mariani T, Tilenni M, Focarelli ML, Brait L, Indino E, Benedetti PA, Chessa L, Albertini A, Ried T, Vezzoni P (2005). SMC1 involvement in fragile site expression. *Hum Mol Genet.*, 14, (4): 525-33.
294. Myers EW, Sutton GG, Smith HO, Adams MD, Venter JC (2002). On the sequencing and assembly of the human genome. *Proc Natl Acad Sci U S A*, 99, (7): 4145-6.
295. Nabbout R, Gennaro E, Dalla Bernardina B, Dulac O, Madia F, Bertini E, Capovilla G, Chiron C, Cristofori G, Elia M, Fontana E, Gaggero R, Granata T, Guerrini R, Loi M, La Selva L, Lispi ML, Matricardi A, Romeo A, Tzolas V, Valseriati D, Veggiotti P, Vigevano F, Vallee L, Dagna Bricarelli F, Bianchi A, Zara F (2003). Spectrum of SCN1A mutations in severe myoclonic epilepsy of infancy. *Neurology*, 60, (12): 1961-7.
296. Nagaraja R, Kere J, MacMillan S, Masisi MJ, Johnson D, Molini BJ, Halley GR, Wein K, Trusgnich M, Eble B et al (1994). Characterization of four human YAC libraries for clone size, chimerism and X chromosome sequence representation. *Nucleic Acids Res.*, 22, (16): 3406-11.
297. Nagaraja R, MacMillan S, Kere J, Jones C, Griffin S, Schmatz M, Terrell J, Shomaker M, Jermak C, Hott C, Masisi M, Mumm S, Srivastava A, Pilia G, Featherstone T, Mazzarella R, Kesterson S, McCauley B, Railey B, Burough F, Nowotny V, D'Urso M, States D, Brownstein B, Schlessinger D (1997). X chromosome map at 75-kb STS resolution, revealing extremes of recombination and GC content. *Genome Res.*, 7, (3): 210-22.
298. Nakamura Y, Leppert M, O'Connell P, Wolff R, Holm T, Culver M, Martin C, Fujimoto E, Hoff M, Kumlin E et al (1987). Variable number of tandem repeat (VNTR) markers for human gene mapping. *Science*, 235, (4796): 1616-22.
299. Nakayama J, Fu YH, Clark AM, Nakahara S, Hamano K, Iwasaki N, Matsui A, Arinami T, Ptacek LJ (2002). A nonsense mutation of the MASS1 gene in a family with febrile and afebrile seizures. *Ann Neurol.*, 52, (5): 654-7.
300. Ng HH, Bird A (1999). DNA methylation and chromatin modification. *Curr Opin Genet Dev.*, (9): 2.
301. Nizetic D, Gellen L, Hamvas RM, Mott R, Grigoriev A, Vatcheva R, Zehetner G, Yaspo ML, Dutriaux A, Lopes C et al (1994). An integrated YAC-overlap and 'cosmid-pocket' map of the human chromosome 21. *Hum Mol Genet.*, 3, (5): 759-70.
302. Nobile C, Pitzalis S (1999). Expression analysis of 21 transcripts physically anchored within the chromosomal region 10q24. *Genomics*, 62, (1): 86-9.
303. Nobile C, Hinzmann B, Scannapieco P, Siebert R, Zimbello R, Perez-Tur J, Sarafidou T, Moschonas NK, French L, Deloukas P, Ciccodicola A, Gesk S, Poza JJ, Lo Nigro C, Seri M, Schlegelberger B, Rosenthal A, Valle G, Lopez de Munain A, Tassinari CA, Michelucci R (2002). Identification and characterization of a novel human brain-specific gene, homologous to *S. scrofa* tmp83.5, in the chromosome 10q24 critical region for temporal lobe epilepsy and spastic paraplegia. *Gene*, 282, (1-2): 87-94.
304. Nobrega MA, Ovcharenko I, Afzal V, Rubin EM (2003). Scanning human gene deserts for long-range enhancers. *Science*, 302, (5644): 413.
305. Nobrega MA, Zhu Y, Plajzer-Frick I, Afzal V, Rubin EM. (2004). Megabase deletions of gene deserts result in viable mice. *Nature*, 431, (7011): 988-93.
306. Nurtdinov RN, Artamonova II, Mironov AA, Gelfand MS (2003). Low conservation of alternative splicing patterns in the human and mouse genomes. *Hum Mol Genet.*, 12, (11): 1313-20.
307. Nusbaum C, Slonim DK, Harris KL, Birren BW, Steen RG, Stein LD, Miller J, Dietrich WF, Nahf R, Wang V, Merport O, Castle AB, Husain Z, Farino G, Gray D, Anderson MO, Devine R, Horton LT Jr, Ye W, Wu X, Kouyoumjian V, Zemsteva IS, Wu Y, Collymore AJ, Courtney DF, Tam J, Cadman M, Haynes AR, Heuston C, Marsland T, Southwell A, Trickett P, Strivens MA,

- Ross MT, Makalowski W, Xu Y, Boguski MS, Carter NP, Denny P, Brown SD, Hudson TJ, Lander ES (1999). A YAC-based physical map of the mouse genome. *Nat Genet.*, 22, (4): 388-93.
308. O'Donnell H, McKeown C, Gould C, Morrow B, Scambler P (1997). Detection of an atypical 22q11 deletion that has no overlap with the DiGeorge syndrome critical region. *Am J Hum Genet.*, 60, (6): 1544-8.
309. Ohinata Y, Sutou S, Kondo M, Takahashi T, Mitsui Y (2002). Male-enhanced antigen-1 gene flanked by two overlapping genes is expressed in late spermatogenesis. *Biol Reprod.*, 67, (6): 1824-31.
310. Ohmori I, Ouchida M, Ohtsuka Y, Oka E, Shimizu K (2002). Significant correlation of the SCN1A mutations and severe myoclonic epilepsy in infancy. *Biochem Biophys Res Commun.*, 295, (1): 17-23.
311. Ohta M, Inoue H, Cotticelli MG, Kastury K, Baffa R, Palazzo J, Siprashvili Z, Mori M, McCue P, Druck T, Croce CM, Huebner K (1996). The FHIT gene, spanning the chromosome 3p14.2 fragile site and renal carcinoma-associated t(3;8) breakpoint, is abnormal in digestive tract cancers. *Cell*, 84, (4): 587-97.
312. Ohtsuka Y, Ohmori I, Ogino T, Ouchida M, Shimizu K, Oka E (2003). Paroxysmal movement disorders in severe myoclonic epilepsy in infancy. *Brain Dev.*, 25, (6): 401-5.
313. Olson MV, Dutchik JE, Graham MY, Brodeur GM, Helms C, Frank M, MacCollin M, Scheinman R, Frank T (1986). Random-clone strategy for genomic restriction mapping in yeast. *Proc Natl Acad Sci U S A*, 83, (20): 7826-30.
314. Ottman R, Risch N, Hauser WA, Pedley TA, Lee JH, Barker-Cummings C, Lustenberger A, Nagle KJ, Lee KS, Scheuer ML *et al* (1995). Localization of a gene for partial epilepsy to chromosome 10q. *Nat Genet.*, 10, (1): 56-60.
315. Ottman R, Winawer MR, Kalachikov S, Barker-Cummings C, Gilliam TC, Pedley TA, Hauser WA (2004). LGI1 mutations in autosomal dominant partial epilepsy with auditory features. *Neurology*, 62, (7): 1120-6.
316. Park JW, Parisky K, Celotto AM, Reenan RA, Graveley BR (2004). Identification of alternative splicing regulators by RNA interference in *Drosophila*. *Proc Natl Acad Sci U S A*, 101, (45): 15974-9.
317. Payseur BA, Nachman MW (2000). Microsatellite variation and recombination rate in the human genome. *Genetics*, 156, (3): 1285-98.
318. Pearson CE, Edamura KN, Cleary JD (2005). Repeat instability: mechanisms of dynamic mutations. *Nat Rev Genet.*, 6, (10): 729-42.
319. Petit P, Fryns JP, van den Berghe H, Hecht F (1986). Population cytogenetics of autosomal fragile sites. *Clin Genet.*, 29, (2): 96-100.
320. Pfeifer D, Kist R, Dewar K, Devon K, Lander ES, Birren B, Korniszewski L, Back E, Scherer G (1999). Campomelic dysplasia translocation breakpoints are scattered over 1 Mb proximal to SOX9: evidence for an extended control region. *Am J Hum Genet.*, 65, (1): 111-24.
321. Phillips HA, Favre I, Kirkpatrick M, Zuberi SM, Goudie D, Heron SE, Scheffer IE, Sutherland GR, Berkovic SF, Bertrand D, Mulley JC (2001). CHRN2 is the second acetylcholine receptor subunit associated with autosomal dominant nocturnal frontal lobe epilepsy. *Am J Hum Genet.*, 68, (1): 225-31.
322. Piano F, Schetter AJ, Mangone M, Stein L, Kempthues KJ (2000). RNAi analysis of genes expressed in the ovary of *Caenorhabditis elegans*. *Curr Biol.*, 10, (24): 1619-22.
323. Pisano T, Marini C, Brovedani P, Brizzolara D, Pruna D, Mei D, Moro F, Cianchetti C, Guerrini R (2005). Abnormal phonologic processing in familial lateral temporal lobe epilepsy due to a new LGI1 mutation. *Epilepsia*, 46, (1): 118-23.

324. Pizzuti A, Flex E, Di Bonaventura C, Dottorini T, Egeo G, Manfredi M, Dallapiccola B, Giallonardo AT (2003). Epilepsy with auditory features: a LGI1 gene mutation suggests a loss-of-function mechanism. *Ann Neurol.*, 53, (3): 396-9.
325. Pollard TD (2003). Functional genomics of cell morphology using RNA interference: pick your style, broad or deep. *J Biol.*, 2, (4): 25.
326. Poza JJ, Saenz A, Martinez-Gil A, Cheron N, Cobo AM, Urtasun M, Marti-Masso JF, Grid D, Beckmann JS, Prud'homme JF, Lopez de Munain A (1999). Autosomal dominant lateral temporal epilepsy: clinical and genetic study of a large Basque pedigree linked to chromosome 10q. *Ann Neurol.*, 45, (2): 182-8.
327. Puech A, Saint-Jore B, Merscher S, Russell RG, Cherif D, Sirotkin H, Xu H, Factor S, Kucherlapati R, Skoultschi AI (2000). Normal cardiovascular development in mice deficient for 16 genes in 550 kb of the velocardiofacial/DiGeorge syndrome region. *Proc Natl Acad Sci USA*, 97, (18): 10090-5.
328. Raijmakers R, Noordman YE, van Venrooij WJ, Pruijn GJ (2002). Protein-protein interactions of hCsl4p with other human exosome subunits. *J Mol Biol.*, 315, (4): 809-18.
329. Rappsilber J, Ryder U, Lamond AI, Mann M (2002). Large-scale proteomic analysis of the human spliceosome. *Genome Res.*, 12, (8): 1231-45.
330. Rauch A, Pfeiffer RA, Leipold G, Singer H, Tigges M, Hofbeck M (1999). A novel 22q11.2 microdeletion in DiGeorge syndrome. *Am J Hum Genet.*, 64, (2): 659-66.
331. Reboul J, Vaglio P, Rual JF, Lamesch P, Martinez M, Armstrong CM, Li S, Jacotot L, Bertin N, Janky R, Moore T, Hudson JR Jr, Hartley JL, Brasch MA, Vandenhoute J, Boulton S, Endress GA, Jenna S, Chevet E, Papatotiropoulos V, Tolia PP, Ptacek J, Snyder M, Huang R, Chance MR, Lee H, Doucette-Stamm L, Hill DE, Vidal M (2003). *C. elegans* ORFeome version 1.1: experimental verification of the genome annotation and resource for proteome-scale protein expression. *Nat Genet.*, 34, (1): 35-41.
332. Reinke V, Smith HE, Nance J, Wang J, Van Doren C, Begley R, Jones SJ, Davis EB, Scherer S, Ward S, Kim SK (2000). A global profile of germline gene expression in *C. elegans*. *Mol Cell*, 6, (3): 605-16.
333. Rhodes M, Straw R, Fernando S, Evans A, Lacey T, Dearlove A, Greystrom J, Walker J, Watson P, Weston P, Kelly M, Taylor D, Gibson K, Mundy C, Bourgade F, Poirier C, Simon D, Brunialti AL, Montagutelli X, Gu'enet JL, Haynes A, Brown SD (1998). A high-resolution microsatellite map of the mouse genome. *Genome Res.*, 8, (5): 531-42.
334. Ribon V, Printen JA, Hoffman NG, Kay BK, Saltiel AR (1998). A novel, multifunctional c-Cbl binding protein in insulin receptor signaling in 3T3-L1 adipocytes. *Mol Cell Biol.*, 18, (2): 872-9.
335. Ried K, Finnis M, Hobson L, Mangelsdorf M, Dayan S, Nancarrow JK, Woollatt E, Kremmidiotis G, Gardner A, Venter D, Baker E, Richards RI (2000). Common chromosomal fragile site FRA16D sequence: identification of the FOR gene spanning FRA16D and homozygous deletions and translocation breakpoints in cancer cells. *Hum Mol Genet.*, 9, (11): 1651-63.
336. Riethman HC, Xiang Z, Paul S, Morse E, Hu XL, Flint J, Chi HC, Grady DL, Moyzis RK (2001). Integration of telomere sequences with the draft human genome sequence. *Nature*, 409, (6822): 948-51.
337. Rizzu P, Lindsay EA, Taylor C, O'Donnell H, Levy A, Scambler P, Baldini A (1996). Cloning and comparative mapping of a gene from the commonly deleted region of DiGeorge and Velocardiofacial syndromes conserved in *C. elegans*. *Mamm Genome*, 7, (9): 639-43.
338. Robert F, Blanchette M, Maes O, Chabot B, Coulombe B (2002). A human RNA polymerase II-containing complex associated with factors necessary for spliceosome assembly. *J Biol Chem.*, 277, (11): 9302-6.
339. Roest Crollius H, Jaillon O, Bernot A, Dasilva C, Bouneau L, Fischer C, Fizames C, Wincker P, Brottier P, Quetier F, Saurin W, Weissenbach J (2000). Estimate of human gene number provided by genome-wide analysis using *Tetraodon nigroviridis* DNA sequence. *Nat Genet.*, 25, (2): 235-8.

340. Rogers SL, Wiedemann U, Stuurman N, Vale RD (2003). Molecular requirements for actin-based lamella formation in *Drosophila* S2 cells. *J Cell Biol.*, 162, (6): 1079-88.
341. Ross MT, Grafham DV, Coffey AJ, Scherer S, McLay K, Muzny D, Platzer M, Howell GR, Burrows C, Bird CP, Frankish A, Lovell FL, Howe KL, Ashurst JL, Fulton RS, Sudbrak R, Wen G, Jones MC, Hurler ME, Andrews TD, Scott CE, Searle S, Ramser J, Whittaker A, Deadman R, Carter NP, Hunt SE, Chen R, Cree A, Gunaratne P, Havlak P, Hodgson A, Metzker ML, Richards S, Scott G, Steffen D, Sodergren E, Wheeler DA, Worley KC, Ainscough R, Ambrose KD, Ansari-Lari MA, Aradhya S, Ashwell RI, Babbage AK, Bagguley CL, Ballabio A, Banerjee R, Barker GE, Barlow KF, Barrett IP, Bates KN, Beare DM, Beasley H, Beasley O, Beck A, Bethel G, Blechschmidt K, Brady N, Bray-Allen S, Bridgeman AM, Brown AJ, Brown MJ, Bonnin D, Bruford EA, Buhay C, Burch P, Burford D, Burgess J, Burrill W, Burton J, Bye JM, Carder C, Carrel L, Chako J, Chapman JC, Chavez D, Chen E, Chen G, Chen Y, Chen Z, Chinault C, Ciccodicola A, Clark SY, Clarke G, Clee CM, Clegg S, Clerc-Blankenburg K, Clifford K, Copley V, Cole CG, Conquer JS, Corby N, Connor RE, David R, Davies J, Davis C, Davis J, Delgado O, Deshazo D, Dhani P, Ding Y, Dinh H, Dodsworth S, Draper H, Dugan-Rocha S, Dunham A, Dunn M, Durbin KJ, Dutta I, Eades T, Ellwood M, Emery-Cohen A, Errington H, Evans KL, Faulkner L, Francis F, Frankland J, Fraser AE, Galgoczy P, Gilbert J, Gill R, Glockner G, Gregory SG, Gribble S, Griffiths C, Grocock R, Gu Y, Gwilliam R, Hamilton C, Hart EA, Hawes A, Heath PD, Heitmann K, Hennig S, Hernandez J, Hinzmann B, Ho S, Hoffs M, Howden PJ, Huckle EJ, Hume J, Hunt PJ, Hunt AR, Isherwood J, Jacob L, Johnson D, Jones S, de Jong PJ, Joseph SS, Keenan S, Kelly S, Kershaw JK, Khan Z, Kioschis P, Klages S, Knights AJ, Kosiura A, Kovar-Smith C, Laird GK, Langford C, Lawlor S, Leversha M, Lewis L, Liu W, Lloyd C, Lloyd DM, Loulseged H, Loveland JE, Lovell JD, Lozado R, Lu J, Lyne R, Ma J, Maheshwari M, Matthews LH, McDowall J, McLaren S, McMurray A, Meidl P, Meitinger T, Milne S, Miner G, Mistry SL, Morgan M, Morris S, Muller I, Mullikin JC, Nguyen N, Nordsiek G, Nyakatura G, O'Dell CN, Okwuonu G, Palmer S, Pandian R, Parker D, Parrish J, Pasternak S, Patel D, Pearce AV, Pearson DM, Pelan SE, Perez L, Porter KM, Ramsey Y, Reichwald K, Rhodes S, Ridler KA, Schlessinger D, Schueler MG, Sehra HK, Shaw-Smith C, Shen H, Sheridan EM, Shownkeen R, Skuce CD, Smith ML, Sothoran EC, Steingruber HE, Steward CA, Storey R, Swann RM, Swarbreck D, Tabor PE, Taudien S, Taylor T, Teague B, Thomas K, Thorpe A, Timms K, Tracey A, Trevanion S, Tromans AC, d'Urso M, Verduzco D, Villasana D, Waldron L, Wall M, Wang Q, Warren J, Warry GL, Wei X, West A, Whitehead SL, Whiteley MN, Wilkinson JE, Willey DL, Williams L, Williamson A, Williamson H, Wilming L, Woodmansey RL, Wray PW, Yen J, Zhang J, Zhou J, Zoghbi H, Zorilla S, Buck D, Reinhardt R, Poustka A, Rosenthal A, Lehrach H, Meindl A, Minx PJ, Hillier LW, Willard HF, Wilson RK, Waterston RH, Rice CM, Vaudin M, Coulson A, Nelson DL, Weinstock G, Sulston JE, Durbin R, Hubbard T, Gibbs RA, Beck S, Rogers J, Bentley DR (2005). The DNA sequence of the human X chromosome. *Nature*, 434, (7031): 325-37.
342. Ross-Macdonald P, Coelho PS, Roemer T, Agarwal S, Kumar A, Jansen R, Cheung KH, Sheehan A, Symoniatis D, Umansky L, Heidtman M, Nelson FK, Iwasaki H, Hager K, Gerstein M, Miller P, Roeder GS, Snyder M (1999). Large-scale analysis of the yeast genome by transposon tagging and gene disruption. *Nature*, 402, (6760): 413-8.
343. Rozier L, El-Achkar E, Apiou F, Debatisse M (2004). Characterization of a conserved aphidicolin-sensitive common fragile site at human 4q22 and mouse 6C1: possible association with an inherited disease and cancer. *Oncogene*, 23, (41): 6872-80.
344. Ruden DM, Ma J, Li Y, Wood K, Ptashne M (1991). Generating yeast transcriptional activators containing no yeast protein sequences. *Nature.*, 350, (6315): 250-2.
345. Ruden DM (1992). Activating regions of yeast transcription factors must have both acidic and hydrophobic amino acids. *Chromosoma*, 101, (5-6): 342-8.
346. Runkel F, Michels M, Franz T (2003). Fxyd3 and Lgi4 expression in the adult mouse: a case of endogenous antisense expression. *Mamm Genome*, 14, (10): 665-72.
347. Ryder E, Russell S (2003). Transposable elements as tools for genomics and genetics in *Drosophila*. *Brief Funct Genomic Proteomic*, 2, (1): 57-71.
348. Saenz A, Galan J, Caloustian C, Lorenzo F, Marquez C, Rodriguez N, Jimenez MD, Poza JJ, Cobo AM, Grid D, Prud'homme JF, Lopez de Munain A (1999). Autosomal dominant nocturnal frontal lobe epilepsy in a Spanish family with a Ser252Phe mutation in the CHRNA4 gene. *Arch Neurol.*, 56, (8): 1004-9.

349. Saiki RK, Scharf S, Faloona F, Mullis KB, Horn GT, Erlich HA, Arnheim N (1985). Enzymatic amplification of beta-globin genomic sequences and restriction site analysis for diagnosis of sickle cell anemia. *Science*, 230, (4732): 1350-4.
350. Saint-Jore B, Puech A, Heyer J, Lin Q, Raine C, Kucherlapati R, Skoultchi AI (1998). Goosecoid-like (Gsc), a candidate gene for velocardiofacial syndrome, is not essential for normal mouse development. *Hum Mol Genet.*, 7, (12): 1841-9.
351. Saitta SC, Harris SE, Gaeth AP, Driscoll DA, McDonald-McGinn DM, Maisenbacher MK, Yersak JM, Chakraborty PK, Hacker AM, Zackai EH, Ashley T, Emanuel BS (2004). Aberrant interchromosomal exchanges are the predominant cause of the 22q11.2 deletion. *Hum Mol Genet.*, 13, (4): 417-28.
352. Sakoda S, Shanske S, DiMauro S, Schon EA (1988). Isolation of a cDNA encoding the B isozyme of human phosphoglycerate mutase (PGAM) and characterization of the PGAM gene family. *J Biol Chem.*, 263, (32): 16899-905.
353. Sanger F, Air GM, Barrell BG, Brown NL, Coulson AR, Fiddes CA, Hutchison CA, Slocombe PM, Smith M (1977). Nucleotide sequence of bacteriophage phi X174 DNA. *Nature*, 265, (5596): 687-95.
354. Sanger F, Coulson AR, Friedmann T, Air GM, Barrell BG, Brown NL, Fiddes JC, Hutchison CA 3rd, Slocombe PM, Smith M (1978). The nucleotide sequence of bacteriophage phiX174. *J Mol Biol.*, 125, (2): 225-46.
355. Sanger F, Coulson AR, Hong GF, Hill DF, Petersen GB (1982). Nucleotide sequence of bacteriophage lambda DNA. *J Mol Biol.*, 162, (4): 729-73.
356. Sapra AK, Arava Y, Khandelvia P, Vijayraghavan U (2004). Genome-wide analysis of pre-mRNA splicing: intron features govern the requirement for the second-step factor, Prp17 in *Saccharomyces cerevisiae* and *Schizosaccharomyces pombe*. *J Biol Chem.*, 279, (50): 52437-46.
357. Sarafidou T, Kahl C, Martinez-Garay I, Mangelsdorf M, Gesk S, Baker E, Kokkinaki M, Talley P, Maltby EL, French L, Harder L, Hinzmann B, Nobile C, Richkind K, Finnis M, Deloukas P, Sutherland GR, Kutsche K, Moschonas NK, Siebert R, Gecz J; European Collaborative Consortium for the Study of ADLTE (2004). Folate-sensitive fragile site FRA10A is due to an expansion of a CGG repeat in a novel gene, FRA10AC1, encoding a nuclear protein. *Genomics*, 84, (1): 69-81.
358. Scambler PJ (2000). The 22q11 deletion syndromes. *Hum Mol Genet.*, 9, (16): 2421-6.
359. Scheel H, Tomiuk S, Hofmann K (2002). A common protein interaction domain links two recently identified epilepsy genes. *Hum Mol Genet.*, 11, (15): 1757-62.
360. Scheetz TE, Raymond MR, Nishimura DY, McClain A, Roberts C, Birkett C, Gardiner J, Zhang J, Butters N, Sun C, Kwitek-Black A, Jacob H, Casavant TL, Soares MB, Sheffield VC (2001). Generation of a high-density rat EST map. *Genome Res.*, 11, (3): 497-502.
361. Schmid CW (1998). Does SINE evolution preclude Alu function? *Nucleic Acids Res.*, 26, (20): 4541-50.
362. Schmutz J, Martin J, Terry A, Couronne O, Grimwood J, Lowry S, Gordon LA, Scott D, Xie G, Huang W, Hellsten U, Tran-Gyamfi M, She X, Prabhakar S, Aerts A, Altherr M, Bajorek E, Black S, Branscomb E, Caoile C, Challacombe JF, Chan YM, Denys M, Detter JC, Escobar J, Flowers D, Fotopulos D, Glavina T, Gomez M, Gonzales E, Goodstein D, Grigoriev I, Groza M, Hammon N, Hawkins T, Haydu L, Israni S, Jett J, Kadner K, Kimball H, Kobayashi A, Lopez F, Lou Y, Martinez D, Medina C, Morgan J, Nandkeshwar R, Noonan JP, Pitluck S, Pollard M, Predki P, Priest J, Ramirez L, Retterer J, Rodriguez A, Rogers S, Salamov A, Salazar A, Thayer N, Tice H, Tsai M, Ustaszewska A, Vo N, Wheeler J, Wu K, Yang J, Dickson M, Cheng JF, Eichler EE, Olsen A, Pennacchio LA, Rokhsar DS, Richardson P, Lucas SM, Myers RM, Rubin EM (2004). The DNA sequence and comparative analysis of human chromosome 5. *Nature*, 431, (7006): 268-74.
363. Schmutz J, Wheeler J, Grimwood J, Dickson M, Yang J, Caoile C, Bajorek E, Black S, Chan YM, Denys M, Escobar J, Flowers D, Fotopulos D, Garcia C, Gomez M, Gonzales E, Haydu L,

- Lopez F, Ramirez L, Retterer J, Rodriguez A, Rogers S, Salazar A, Tsai M, Myers RM (2004). Quality assessment of the human genome sequence. *Nature*, 429, (6990): 365-8.
364. Schuler GD, Boguski MS, Stewart EA, Stein LD, Gyapay G, Rice K, White RE, Rodriguez-Tome P, Aggarwal A, Bajorek E, Bentolila S, Birren BB, Butler A, Castle AB, Chiannikulchai N, Chu A, Clee C, Cowles S, Day PJ, Dibling T, Drouot N, Dunham I, Duprat S, East C, Edwards C, Fan JB, Fang N, Fizames C, Garrett C, Green L, Hadley D, Harris M, Harrison P, Brady S, Hicks A, Holloway E, Hui L, Hussain S, Louis-Dit-Sully C, Ma J, MacGilvery A, Mader C, Maratukulam A, Matise TC, McKusick KB, Morissette J, Mungall A, Muselet D, Nusbaum HC, Page DC, Peck A, Perkins S, Piercy M, Qin F, Quackenbush J, Ranby S, Reif T, Rozen S, Sanders C, She X, Silva J, Slonim DK, Soderlund C, Sun WL, Tabar P, Thangarajah T, Vega-Czarny N, Vollrath D, Voyticky S, Wilmer T, Wu X, Adams MD, Auffray C, Walter NA, Brandon R, Dehejia A, Goodfellow PN, Houlgatte R, Hudson JR Jr, Ide SE, Iorio KR, Lee WY, Seki N, Nagase T, Ishikawa K, Nomura N, Phillips C, Polymeropoulos MH, Sandusky M, Schmitt K, Berry R, Swanson K, Torres R, Venter JC, Sikela JM, Beckmann JS, Weissenbach J, Myers RM, Cox DR, James MR, Bentley D, Deloukas P, Lander ES, Hudson TJ (1996). A gene map of the human genome. *Science*, 274, (5287): 540-6.
365. Schuler GD (1998). Electronic PCR: bridging the gap between genome mapping and genome sequencing. *Trends Biotechnol.*, 16, (11): 456-9.
366. Selleri L, Eubanks JH, Giovannini M, Hermanson GG, Romo A, Djabali M, Maurer S, McElligott DL, Smith MW, Evans GA (1992). Detection and characterization of "chimeric" yeast artificial chromosome clones by fluorescent in situ suppression hybridization. *Genomics*, 14, (2): 536-41.
367. Senechal KR, Thaller C, Noebels JL (2005). ADPEAF mutations reduce levels of secreted LGI1, a putative tumor suppressor protein linked to epilepsy. *Hum Mol Genet.*, 14, (12): 1613-20.
368. Shaikh TH, Kurahashi H, Saitta SC, O'Hare AM, Hu P, Roe BA, Driscoll DA, McDonald-McGinn DM, Zackai EH, Budarf ML, Emanuel BS (2000). Chromosome 22-specific low copy repeats and the 22q11.2 deletion syndrome: genomic organization and deletion endpoint analysis. *Hum Mol Genet.*, 9, (4): 489-501.
369. Shaikh TH, Kurahashi H, Emanuel BS (2001). Evolutionarily conserved low copy repeats (LCRs) in 22q11 mediate deletions, duplications, translocations, and genomic instability: an update and literature review. *Genet Med.*, 3, (1): 6-13.
370. Shamsher MK, Ploski J, Radu A (2002). Karyopherin beta 2B participates in mRNA export from the nucleus. *Proc Natl Acad Sci U S A*, 99, (22): 14195-9.
371. Shaw MA, Chiurazzi P, Romain DR, Neri G, Gecz J (2002). A novel gene, FAM11A, associated with the FRAXF CpG island is transcriptionally silent in FRAXF full mutation. *Eur J Hum Genet.*, 10, (11): 767-72.
372. Sherman SL, Sutherland GR (1986). Segregation analysis of rare autosomal fragile sites. *Hum Genet.*, 72, (2): 123-8.
373. Shin JT, Priest JR, Ovcharenko I, Ronco A, Moore RK, Burns CG, MacRae CA (2005). Human-zebrafish non-coding conserved elements act in vivo to regulate transcription. *Nucleic Acids Res.*, 33, (17): 5437-45.
374. Shiraishi T, Druck T, Mimori K, Flomenberg J, Berk L, Alder H, Miller W, Huebner K, Croce CM (2001). Sequence conservation at human and mouse orthologous common fragile regions, FRA3B/FHIT and Fra14A2/Fhit. *Proc Natl Acad Sci U S A*, 98, (10): 5722-7.
375. Shizuya H, Birren B, Kim UJ, Mancino V, Slepak T, Tachiiri Y, Simon M (1992). Cloning and stable maintenance of 300-kilobase-pair fragments of human DNA in *Escherichia coli* using an F-factor-based vector. *Proc Natl Acad Sci U S A.*, 89, (18): 8794-7.
376. Simmer F, Moorman C, van der Linden AM, Kuijk E, van den Berghe PV, Kamath RS, Fraser AG, Ahringer J, Plasterk RH (2003). Genome-wide RNAi of *C. elegans* using the hypersensitive rrf-3 strain reveals novel gene functions. *PLoS Biol.*, (1): E12.
377. Singh NA, Charlier C, Stauffer D, DuPont BR, Leach RJ, Melis R, Ronen GM, Bjerre I, Quattlebaum T, Murphy JV, McHarg ML, Gagnon D, Rosales TO, Peiffer A, Anderson VE,

- Leppert M (1998). A novel potassium channel gene, KCNQ2, is mutated in an inherited epilepsy of newborns. *Nat Genet.*, 18, (1): 25-9.
378. Skaletsky H, Kuroda-Kawaguchi T, Minx PJ, Cordum HS, Hillier L, Brown LG, Repping S, Pyntikova T, Ali J, Bieri T, Chinwalla A, Delehaunty A, Delehaunty K, Du H, Fewell G, Fulton L, Fulton R, Graves T, Hou SF, Latrielle P, Leonard S, Mardis E, Maupin R, McPherson J, Miner T, Nash W, Nguyen C, Ozersky P, Pepin K, Rock S, Rohlffing T, Scott K, Schultz B, Strong C, Tin-Wollam A, Yang SP, Waterston RH, Wilson RK, Rozen S, Page DC (2003). The male-specific region of the human Y chromosome is a mosaic of discrete sequence classes. *Nature*, 423, (6942): 825-37.
379. Soderlund C, Longden I, Mott R (1997). FPC: a system for building contigs from restriction fingerprinted clones. *Comput Appl Biosci.*, 13, (5): 523-35.
380. Soderlund C, Humphray S, Dunham A, French L (2000). Contigs built with fingerprints, markers, and FPC V4.7. *Genome Res.*, 10, (11): 1772-87.
381. Somerville RP, Chernova O, Liu S, Shoshan Y, Cowell JK (2000). Identification of the promoter, genomic structure, and mouse ortholog of LGI1. *Mamm Genome*, 11, (8): 622-7.
382. Somma MP, Fasulo B, Cenci G, Cundari E, Gatti M (2002). Molecular dissection of cytokinesis by RNA interference in *Drosophila* cultured cells. *Mol Biol Cell*, 13, (7): 2448-60.
383. Spurr NK, Bryant SP, Attwood J, Nyberg K, Cox SA, Mills A, Bains R, Warne D, Cullin L, Povey S, Sebaoun JM, Weissenbach J, Cann HM, Lathrop M, Dausset J, Marcadet-Troton A, Cohen D (1994). European Gene Mapping Project (EUROGEM): genetic maps based on the CEPH reference families. *Eur J Hum Genet.*, 2, (3): 193-203.
384. Srinivasan J, Schachner M, Catterall WA (1998). Interaction of voltage-gated sodium channels with the extracellular matrix molecules tenascin-C and tenascin-R. *Proc Natl Acad Sci U S A*, 95, (26): 15753-7.
385. Staden R (1980). A new computer method for the storage and manipulation of DNA gel reading data. *Nucleic Acids Res.*, 8, (16): 3673-94.
386. Staley JP, Guthrie C (1998). Mechanical devices of the spliceosome: motors, clocks, springs, and things. *Cell*, 92, (3): 315-26.
387. Staub E, Perez-Tur J, Siebert R, Nobile C, Moschonas NK, Deloukas P, Hinzmann B (2002). The novel EPTP repeat defines a superfamily of proteins implicated in epileptic disorders. *Trends Biochem Sci.*, 27, (9): 441-4.
388. Steinlein OK, Mulley JC, Propping P, Wallace RH, Phillips HA, Sutherland GR, Scheffer IE, Berkovic SF (1995). A missense mutation in the neuronal nicotinic acetylcholine receptor alpha 4 subunit is associated with autosomal dominant nocturnal frontal lobe epilepsy. *Nat Genet.*, 11, (2): 201-3.
389. Steinlein OK, Magnusson A, Stoodt J, Bertrand S, Weiland S, Berkovic SF, Nakken KO, Propping P, Bertrand D (1997). An insertion mutation of the CHRNA4 gene in a family with autosomal dominant nocturnal frontal lobe epilepsy. *Hum Mol Genet.*, 6, (6): 943-7.
390. Stewart EA, McKusick KB, Aggarwal A, Bajorek E, Brady S, Chu A, Fang N, Hadley D, Harris M, Hussain S, Lee R, Maratukulam A, O'Connor K, Perkins S, Piercy M, Qin F, Reif T, Sanders C, She X, Sun WL, Tabar P, Voyticky S, Cowles S, Fan JB, Cox DR *et al* (1997). An STS-based radiation hybrid map of the human genome. *Genome Res.*, 7, (5): 422-33.
391. Steyaert J, Legius E, Borghgraef M, Fryns JP (2003). A distinct neurocognitive phenotype in female fragile-X premutation carriers assessed with visual attention tasks. *Am J Med Genet A.*, 116, (1): 44-51.
392. Stiles B, Groszer M, Wang S, Jiao J, Wu H (2004). PTENless means more. *Dev Biol.*, 273, (2): 175-84.
393. Stone DM, Jacky PB, Hancock DD, Prieur DJ (1991). Chromosomal fragile site expression in dogs: I. Breed specific differences. *Am J Med Genet.*, 40, (2): 214-22.

394. Stone DM, Stephens KE (1993). Bromodeoxyuridine induces chromosomal fragile sites in the canine genome. *Am J Med Genet.*, 46, (2): 198-202.
395. Stone DM, Stephens KE, Doles J (1993). Folate-sensitive and aphidicolin-inducible fragile sites are expressed in the genome of the domestic cat. *Cancer Genet Cytogenet.*, 65, (2): 130-4.
396. Subramanian S, Madgula VM, George R, Mishra RK, Pandit MW, Kumar CS, Singh L (2003). Triplet repeats in human genome: distribution and their association with genes and other genomic regions. *Bioinformatics*, 19, (5): 549-52.
397. Sugawara T, Mazaki-Miyazaki E, Ito M, Nagafuji H, Fukuma G, Mitsudome A, Wada K, Kaneko S, Hirose S, Yamakawa K (2001a). Nav1.1 mutations cause febrile seizures associated with afebrile partial seizures. *Neurology*, 57, (4): 703-5.
398. Sugawara T, Tsurubuchi Y, Agarwala KL, Ito M, Fukuma G, Mazaki-Miyazaki E, Nagafuji H, Noda M, Imoto K, Wada K, Mitsudome A, Kaneko S, Montal M, Nagata K, Hirose S, Yamakawa K (2001b). A missense mutation of the Na⁺ channel alpha II subunit gene Na(v)1.2 in a patient with febrile and afebrile seizures causes channel dysfunction. *Proc Natl Acad Sci U S A*, 98, (11): 6384-9.
399. Sugawara T, Mazaki-Miyazaki E, Fukushima K, Shimomura J, Fujiwara T, Hamano S, Inoue Y, Yamakawa K (2002). Frequent mutations of SCN1A in severe myoclonic epilepsy in infancy. *Neurology*, 58, (7): 1122-4.
400. Sutherland GR (1981). Heritable fragile sites on human chromosomes. VII. Children homozygous for the BrdU-requiring fra(10)(q25) are phenotypically normal. *Am J Hum Genet.*, 33, (6): 946-9.
401. Taricani L, Tejada ML, Young PG (2002). The fission yeast ES2 homologue, Bis1, interacts with the Ish1 stress-responsive nuclear envelope protein. *J Biol Chem.*, 277, (12): 10562-72.
402. Thibault ST, Singer MA, Miyazaki WY, Milash B, Dompe NA, Singh CM, Buchholz R, Demsky M, Fawcett R, Francis-Lang HL, Ryner L, Cheung LM, Chong A, Erickson C, Fisher WW, Greer K, Hartouni SR, Howie E, Jakkula L, Joo D, Killpack K, Laufer A, Mazzotta J, Smith RD, Stevens LM, Stuber C, Tan LR, Ventura R, Woo A, Zakrajsek I, Zhao L, Chen F, Swimmer C, Kopczyński C, Duyk G, Winberg ML, Margolis J (2004). A complementary transposon tool kit for *Drosophila melanogaster* using P and piggyBac. *Nat Genet.*, 36, (3): 283-7.
403. Tran HT, Ulke A, Morrice N, Johannes CJ, Moorhead GB (2004). Proteomic characterization of protein phosphatase complexes of the mammalian nucleus. *Mol Cell Proteomics*, 3, (3): 257-65.
404. Uetz P, Giot L, Cagney G, Mansfield TA, Judson RS, Knight JR, Lockshon D, Narayan V, Srinivasan M, Pochart P, Qureshi-Emili A, Li Y, Godwin B, Conover D, Kalbfleisch T, Vijayadamar G, Yang M, Johnston M, Fields S, Rothberg JM (2000). A comprehensive analysis of protein-protein interactions in *Saccharomyces cerevisiae*. *Nature*, 403, (6770): 623-7.
405. Velagaleti GV, Bien-Willner GA, Northup JK, Lockhart LH, Hawkins JC, Jalal SM, Withers M, Lupski JR, Stankiewicz P (2005). Position effects due to chromosome breakpoints that map approximately 900 Kb upstream and approximately 1.3 Mb downstream of SOX9 in two patients with campomelic dysplasia. *Am J Hum Genet.*, 76, (4): 652-62.
406. Van Etten WJ, Steen RG, Nguyen H, Castle AB, Slonim DK, Ge B, Nusbaum C, Schuler GD, Lander ES, Hudson TJ (1999). Radiation hybrid map of the mouse genome. *Nat Genet.*, 22, (4): 384-7.
407. Veeramachaneni V, Makalowski W, Galdzicki M, Sood R, Makalowska I (2004). Mammalian overlapping genes: the comparative perspective. *Genome Res.*, 14, (2): 280-6.
408. Velculescu VE, Zhang L, Zhou W, Vogelstein J, Basrai MA, Bassett DE Jr, Hieter P, Vogelstein B, Kinzler KW (1997). Characterization of the yeast transcriptome. *Cell*, 88, (2): 243-51.
409. Venter JC, Adams MD, Myers EW, Li PW, Mural RJ, Sutton GG, Smith HO, Yandell M, Evans CA, Holt RA, Gocayne JD, Amanatides P, Ballew RM, Huson DH, Wortman JR, Zhang Q, Kodira CD, Zheng XH, Chen L, Skupski M, Subramanian G, Thomas PD, Zhang J, Gabor Miklos GL, Nelson C, Broder S, Clark AG, Nadeau J, McKusick VA, Zinder N, Levine AJ,

- Roberts RJ, Simon M, Slayman C, Hunkapiller M, Bolanos R, Delcher A, Dew I, Fasulo D, Flanigan M, Florea L, Halpern A, Hannenhalli S, Kravitz S, Levy S, Mobarry C, Reinert K, Remington K, Abu-Threideh J, Beasley E, Biddick K, Bonazzi V, Brandon R, Cargill M, Chandramouliswaran I, Charlab R, Chaturvedi K, Deng Z, Di Francesco V, Dunn P, Eilbeck K, Evangelista C, Gabrielian AE, Gan W, Ge W, Gong F, Gu Z, Guan P, Heiman TJ, Higgins ME, Ji RR, Ke Z, Ketchum KA, Lai Z, Lei Y, Li Z, Li J, Liang Y, Lin X, Lu F, Merkulov GV, Milshina N, Moore HM, Naik AK, Narayan VA, Neelam B, Nusskern D, Rusch DB, Salzberg S, Shao W, Shue B, Sun J, Wang Z, Wang A, Wang X, Wang J, Wei M, Wides R, Xiao C, Yan C, Yao A, Ye J, Zhan M, Zhang W, Zhang H, Zhao Q, Zheng L, Zhong F, Zhong W, Zhu S, Zhao S, Gilbert D, Baumhueter S, Spier G, Carter C, Cravchik A, Woodage T, Ali F, An H, Awe A, Baldwin D, Baden H, Barnstead M, Barrow I, Beeson K, Busam D, Carver A, Center A, Cheng ML, Curry L, Danaher S, Davenport L, Desilets R, Dietz S, Dodson K, Doup L, Ferriera S, Garg N, Gluecksmann A, Hart B, Haynes J, Haynes C, Heiner C, Hladun S, Hostin D, Houck J, Howland T, Ibegwam C, Johnson J, Kalush F, Kline L, Koduru S, Love A, Mann F, May D, McCawley S, McIntosh T, McMullen I, Moy M, Moy L, Murphy B, Nelson K, Pfannkoch C, Pratts E, Puri V, Qureshi H, Reardon M, Rodriguez R, Rogers YH, Romblad D, Ruhfel B, Scott R, Sitter C, Smallwood M, Stewart E, Strong R, Suh E, Thomas R, Tint NN, Tse S, Vech C, Wang G, Wetter J, Williams S, Williams M, Windsor S, Winn-Deen E, Wolfe K, Zaveri J, Zaveri K, Abril JF, Guigo R, Campbell MJ, Sjolander KV, Karlak B, Kejariwal A, Mi H, Lazareva B, Hatton T, Narechania A, Diemer K, Muruganujan A, Guo N, Sato S, Bafna V, Istrail S, Lippert R, Schwartz R, Walenz B, Yooseph S, Allen D, Basu A, Baxendale J, Blick L, Caminha M, Carnes-Stine J, Caulk P, Chiang YH, Coyne M, Dahlke C, Mays A, Dombroski M, Donnelly M, Ely D, Esparham S, Fosler C, Gire H, Glanowski S, Glasser K, Glodek A, Gorokhov M, Graham K, Gropman B, Harris M, Heil J, Henderson S, Hoover J, Jennings D, Jordan C, Jordan J, Kasha J, Kagan L, Kraft C, Levitsky A, Lewis M, Liu X, Lopez J, Ma D, Majoros W, McDaniel J, Murphy S, Newman M, Nguyen T, Nguyen N, Nodell M, Pan S, Peck J, Peterson M, Rowe W, Sanders R, Scott J, Simpson M, Smith T, Sprague A, Stockwell T, Turner R, Venter E, Wang M, Wen M, Wu D, Wu M, Xia A, Zandieh A, Zhu X (2001). The sequence of the human genome. *Science*, 291, (5507): 1304-51.
410. Vincent M, Lauriault P, Dubois MF, Lavoie S, Bensaude O, Chabot B (1996). The nuclear matrix protein p255 is a highly phosphorylated form of RNA polymerase II largest subunit which associates with spliceosomes. *Nucleic Acids Res.*, 24, (23): 4649-52.
411. Vincent MC, Heitz F, Tricoire J, Bourrouillou G, Kuhlein E, Rolland M, Calvas P (1999). 22q11 deletion in DGS/VCFS monozygotic twins with discordant phenotypes. *Genet Couns.*, 10, (1): 43-9.
412. Walhout AJ, Sordella R, Lu X, Hartley JL, Temple GF, Brasch MA, Thierry-Mieg N, Vidal M (2000). Protein interaction mapping in *C. elegans* using proteins involved in vulval development. *Science*, 287, (5450): 116-22.
413. Walhout AJ, Vidal M (2001). High-throughput yeast two-hybrid assays for large-scale protein interaction mapping. *Methods*, 24, (3): 297-306.
414. Wallace RH, Wang DW, Singh R, Scheffer IE, George AL Jr, Phillips HA, Saar K, Reis A, Johnson EW, Sutherland GR, Berkovic SF, Mulley JC (1998). Febrile seizures and generalized epilepsy associated with a mutation in the Na⁺-channel β 1 subunit gene SCN1B. *Nat Genet*, 19, (4): 366-70.
415. Wallace RH, Scheffer IE, Barnett S, Richards M, Dibbens L, Desai RR, Lerman-Sagie T, Lev D, Mazarib A, Brand N, Ben-Zeev B, Goikhman I, Singh R, Kremmidiotis G, Gardner A, Sutherland GR, George AL Jr, Mulley JC, Berkovic SF (2001a). Neuronal sodium-channel alpha1-subunit mutations in generalized epilepsy with febrile seizures plus. *Am J Hum Genet.*, 68, (4): 859-65.
416. Wallace RH, Marini C, Petrou S, Harkin LA, Bowser DN, Panchal RG, Williams DA, Sutherland GR, Mulley JC, Scheffer IE, Berkovic SF (2001b). Mutant GABA(A) receptor gamma2-subunit in childhood absence epilepsy and febrile seizures. *Nat Genet.*, 28, (1): 49-52.
417. Wallace RH, Scheffer IE, Parasivam G, Barnett S, Wallace GB, Sutherland GR, Berkovic SF, Mulley JC (2002). Generalized epilepsy with febrile seizures plus: mutation of the sodium channel subunit SCN1B. *Neurology*, 58, (9): 1426-9.
418. Wallace RH, Hodgson BL, Grinton BE, Gardiner RM, Robinson R, Rodriguez-Casero V, Sadleir L, Morgan J, Harkin LA, Dibbens LM, Yamamoto T, Andermann E, Mulley JC, Berkovic SF,

- Scheffer IE (2003). Sodium channel alpha1-subunit mutations in severe myoclonic epilepsy of infancy and infantile spasms. *Neurology*, 61, (6): 765-9.
419. Watanabe TK, Bihoreau MT, McCarthy LC, Kiguwa SL, Hishigaki H, Tsuji A, Browne J, Yamasaki Y, Mizoguchi-Miyakita A, Oga K, Ono T, Okuno S, Kanemoto N, Takahashi E, Tomita K, Hayashi H, Adachi M, Webber C, Davis M, Kiel S, Knights C, Smith A, Critcher R, Miller J, Thangarajah T, Day PJ, Hudson JR Jr, Irie Y, Takagi T, Nakamura Y, Goodfellow PN, Lathrop GM, Tanigami A, James MR (1999). A radiation hybrid map of the rat genome containing 5,255 markers. *Nat Genet.*, 22, (1): 27-36.
420. Waterston R, Martin C, Craxton M, Huynh C, Coulson A, Hillier L, Durbin R, Green P, Shownkeen R, Halloran N *et al* (1992). A survey of expressed genes in *Caenorhabditis elegans*. *Nat Genet.*, 1, (2): 114-23.
421. Waterston RH, Lander ES, Sulston JE (2002a). On the sequencing of the human genome. *Proc Natl Acad Sci U S A*, 99, (6): 3712-6.
422. Waterston RH, Lindblad-Toh K, Birney E, Rogers J, Abril JF, Agarwal P, Agarwala R, Ainscough R, Alexandersson M, An P, Antonarakis SE, Attwood J, Baertsch R, Bailey J, Barlow K, Beck S, Berry E, Birren B, Bloom T, Bork P, Botcherby M, Bray N, Brent MR, Brown DG, Brown SD, Bult C, Burton J, Butler J, Campbell RD, Carninci P, Cawley S, Chiaromonte F, Chinwalla AT, Church DM, Clamp M, Clee C, Collins FS, Cook LL, Copley RR, Coulson A, Couronne O, Cuff J, Curwen V, Cutts T, Daly M, David R, Davies J, Delehaunty KD, Deri J, Dermitzakis ET, Dewey C, Dickens NJ, Diekhans M, Dodge S, Dubchak I, Dunn DM, Eddy SR, Elnitski L, Emes RD, Eswara P, Eyas E, Felsenfeld A, Fewell GA, Flicek P, Foley K, Frankel WN, Fulton LA, Fulton RS, Furey TS, Gage D, Gibbs RA, Glusman G, Gnerre S, Goldman N, Goodstadt L, Grafham D, Graves TA, Green ED, Gregory S, Guigo R, Guyer M, Hardison RC, Haussler D, Hayashizaki Y, Hillier LW, Hinrichs A, Hlavina W, Holzer T, Hsu F, Hua A, Hubbard T, Hunt A, Jackson I, Jaffe DB, Johnson LS, Jones M, Jones TA, Joy A, Kamal M, Karlsson EK, Karolchik D, Kasprzyk A, Kawai J, Keibler E, Kells C, Kent WJ, Kirby A, Kolbe DL, Korf I, Kucherlapati RS, Kulbokas EJ, Kulp D, Landers T, Leger JP, Leonard S, Letunic I, Levine R, Li J, Li M, Lloyd C, Lucas S, Ma B, Maglott DR, Mardis ER, Matthews L, Mauceli E, Mayer JH, McCarthy M, McCombie WR, McLaren S, McLay K, McPherson JD, Meldrim J, Meredith B, Mesirov JP, Miller W, Miner TL, Mongin E, Montgomery KT, Morgan M, Mott R, Mullikin JC, Muzny DM, Nash WE, Nelson JO, Nhan MN, Nicol R, Ning Z, Nusbaum C, O'Connor MJ, Okazaki Y, Oliver K, Overton-Larty E, Pachter L, Parra G, Pepin KH, Peterson J, Pevzner P, Plumb R, Pohl CS, Poliakov A, Ponce TC, Ponting CP, Potter S, Quail M, Reymond A, Roe BA, Roskin KM, Rubin EM, Rust AG, Santos R, Sapojnikov V, Schultz B, Schultz J, Schwartz MS, Schwartz S, Scott C, Seaman S, Searle S, Sharpe T, Sheridan A, Shownkeen R, Sims S, Singer JB, Slater G, Smit A, Smith DR, Spencer B, Stabenau A, Stange-Thomann N, Sugnet C, Suyama M, Tesler G, Thompson J, Torrents D, Trevaskis E, Tromp J, Ucla C, Ureta-Vidal A, Vinson JP, Von Niederhausern AC, Wade CM, Wall M, Weber RJ, Weiss RB, Wendl MC, West AP, Wetterstrand K, Wheeler R, Whelan S, Wierzbowski J, Willey D, Williams S, Wilson RK, Winter E, Worley KC, Wyman D, Yang S, Yang SP, Zdobnov EM, Zody MC, Lander ES; Mouse Genome Sequencing Consortium (2002). Initial sequencing and comparative analysis of the mouse genome. *Nature*, 420, (6915): 520-62.
423. Weissenbach J, Gyapay G, Dib C, Vignal A, Morissette J, Millasseau P, Vaysseix G, Lathrop M (1992). A second-generation linkage map of the human genome. *Nature*, 359, (6398): 794-801.
424. Wilke CM, Guo SW, Hall BK, Boldog F, Gemmill RM, Chandrasekharappa SC, Barcroft CL, Drabkin HA, Glover TW (1994). Multicolor FISH mapping of YAC clones in 3p14 and identification of a YAC spanning both FRA3B and the t(3;8) associated with hereditary renal cell carcinoma. *Genomics*, 22, (2): 319-26.
425. Wilkes D, Shaw J, Anand R, Riley J, Winter P, Wallis J, Driesel AG, Williamson R, Chamberlain S (1991). Identification of CpG islands in a physical map encompassing the Friedreich's ataxia locus. *Genomics*, 9, (1): 90-5.
426. Will CL, Schneider C, Reed R, Luhrmann R (1999). Identification of both shared and distinct proteins in the major and minor spliceosomes. *Science*, 284, (5422): 2003-5.
427. Will CL, Schneider C, MacMillan AM, Katopodis NF, Neubauer G, Wilm M, Luhrmann R, Query CC (2001). A novel U2 and U11/U12 snRNP protein that associates with the pre-mRNA branch site. *EMBO J.*, 20, (16): 4536-46.

428. Will CL, Urlaub H, Achsel T, Gentzel M, Wilm M, Luhrmann R (2002). Characterization of novel SF3b and 17S U2 snRNP proteins, including a human Prp5p homologue and an SF3b DEAD-box protein. *EMBO J.*, 21, (18): 4978-88.
429. Williams T, Fried M (1986). A mouse locus at which transcription from both DNA strands produces mRNAs complementary at their 3' ends. *Nature*, 322, (6076): 275-9.
430. Winawer MR, Martinelli Boneschi F, Barker-Cummings C, Lee JH, Liu J, Mekios C, Gilliam TC, Pedley TA, Hauser WA, Ottman R (2002). Four new families with autosomal dominant partial epilepsy with auditory features: clinical description and linkage to chromosome 10q24. *Epilepsia*, 43, (1): 60-7.
431. Winzeler EA, Shoemaker DD, Astromoff A, Liang H, Anderson K, Andre B, Bangham R, Benito R, Boeke JD, Bussey H, Chu AM, Connolly C, Davis K, Dietrich F, Dow SW, El Bakkoury M, Foury F, Friend SH, Gentalen E, Giaever G, Hegemann JH, Jones T, Laub M, Liao H, Liebundguth N, Lockhart DJ, Lucau-Danila A, Lussier M, M'Rabet N, Menard P, Mittmann M, Pai C, Rebischung C, Revuelta JL, Riles L, Roberts CJ, Ross-MacDonald P, Scherens B, Snyder M, Sookhai-Mahadeo S, Storms RK, Veronneau S, Voet M, Volckaert G, Ward TR, Wysocki R, Yen GS, Yu K, Zimmermann K, Philippsen P, Johnston M, Davis RW (1999). Functional characterization of the *S. cerevisiae* genome by gene deletion and parallel analysis. *Science*, 285, (5429): 901-6.
432. Wood WB (1988). Determination of pattern and fate in early embryos of *Caenorhabditis elegans*. *Dev Biol (N Y 1985)*, 5, (57-78):
433. Woolfe A, Goodson M, Goode DK, Snell P, McEwen GK, Vavouri T, Smith SF, North P, Callaway H, Kelly K, Walter K, Abnizova I, Gilks W, Edwards YJ, Cooke JE, Elgar G (2005). Highly conserved non-coding sequences are associated with vertebrate development. *PLoS Biol.*, 3, (1): e7.
434. Wren JD, Forgacs E, Fondon JW 3rd, Pertsemliadis A, Cheng SY, Gallardo T, Williams RS, Shohet RV, Minna JD, Garner HR (2000). Repeat polymorphisms within gene regions: phenotypic and evolutionary implications. *Am J Hum Genet.*, 67, (2): 345-56.
435. Wunderle VM, Critcher R, Hastie N, Goodfellow PN, Schedl A (1998). Deletion of long-range regulatory elements upstream of SOX9 causes campomelic dysplasia. *Proc Natl Acad Sci USA*, 95, (18): 10649-54.
436. Xhemalce B, Seeler JS, Thon G, Dejean A, Arcangioli B (2004). Role of the fission yeast SUMO E3 ligase Pli1p in centromere and telomere maintenance. *EMBO J.*, 23, (19): 3844-53.
437. Xiao ZC, Ragsdale DS, Malhotra JD, Mattei LN, Braun PE, Schachner M, Isom LL (1999). Tenascin-R is a functional modulator of sodium channel beta subunits. *J Biol Chem.*, 274, (37): 26511-7.
438. Xu Q, Modrek B, Lee C (2002). Genome-wide detection of tissue-specific alternative splicing in the human transcriptome. *Nucleic Acids Res.*, 30, (17): 3754-66.
439. Yagi H, Furutani Y, Hamada H, Sasaki T, Asakawa S, Minoshima S, Ichida F, Joo K, Kimura M, Imamura S, Kamatani N, Momma K, Takao A, Nakazawa M, Shimizu N, Matsuoka R (2003). Role of TBX1 in human del22q11.2 syndrome. *Lancet*, 362, (9393): 1366-73.
440. Yamagishi H, Ishii C, Maeda J, Kojima Y, Matsuoka R, Kimura M, Takao A, Momma K, Matsuo N (1998). Phenotypic discordance in monozygotic twins with 22q11.2 deletion. *Am J Med Genet.*, 78, (4): 319-21.
441. Yamagishi H, Garg V, Matsuoka R, Thomas T, Srivastava D (1999). A molecular pathway revealing a genetic basis for human cardiac and craniofacial defects. *Science*, 283, (5405): 1158-61.
442. Yang MY, Long SE. (1993). Folate sensitive common fragile sites in chromosomes of the domestic pig (*Sus scrofa*). *Res Vet Sci.*, 55, (2): 231-5.
443. Yang Y, Hentati A, Deng HX, Dabbagh O, Sasaki T, Hirano M, Hung WY, Ouahchi K, Yan J, Azim AC, Cole N, Gascon G, Yagmour A, Ben-Hamida M, Pericak-Vance M, Hentati F, Siddique T (2001). The gene encoding alsin, a protein with three guanine-nucleotide exchange

- factor domains, is mutated in a form of recessive amyotrophic lateral sclerosis. *Nat Genet.*, 29, (2): 160-5.
444. Yano Y, Saito R, Yoshida N, Yoshiki A, Wynshaw-Boris A, Tomita M, Hirotsune S (2004). A new role for expressed pseudogenes as ncRNA: regulation of mRNA stability of its homologous coding gene. *J Mol Med.*, 82, (7): 414-22.
445. Yekta S, Shih IH, Bartel DP (2004). MicroRNA-directed cleavage of HOXB8 mRNA. *Science*, 304, (5670): 594-6.
446. Yeo G, Holste D, Kreiman G, Burge CB (2004). Variation in alternative splicing across human tissues. *Genome Biol.*, 5, (10): R74.
447. Yu A, Zhao C, Fan Y, Jang W, Mungall AJ, Deloukas P, Olsen A, Doggett NA, Ghebranious N, Broman KW, Weber JL (2001). Comparison of human genetic and sequence-based physical maps. *Nature*, 409, (6822): 951-3.
448. Yu J, Yang Z, Kibukawa M, Paddock M, Passey DA, Wong GK (2002). Minimal introns are not "junk". *Genome Res.*, 12, (8): 1185-9.
449. Yunis JJ, Sawyer JR, Dunham K (1980). The striking resemblance of high-resolution G-banded chromosomes of man and chimpanzee. *Science*, 208, (4448): 1145-8.
450. Zhang MQ, Marr TG (1994). Fission yeast gene structure and recognition. *Nucleic Acids Res.*, 22, (9): 1750-9.
451. Zhang Z, Gerstein M (2004). Large-scale analysis of pseudogenes in the human genome. *Curr Opin Genet Dev.*, 14, (4): 328-35.
452. Zhang Z, Carriero N, Gerstein M (2004). Comparative analysis of processed pseudogenes in the mouse and human genomes. *Trends Genet.*, 20, (2): 62-7.
453. Zhao S (2000). Human BAC ends. *Nucleic Acids Res.*, 28, (1): 129-32.
454. Zheng ZM (2004). Regulation of alternative RNA splicing by exon definition and exon sequences in viral and mammalian gene expression. *J Biomed Sci.*, 11, (3): 278-94.
455. Zhou C, Blumberg B (2003). Overlapping gene structure of human VLCAD and DLG4. *Gene*, 305, (2): 161-6.
456. Zhou Z, Licklider LJ, Gygi SP, Reed R (2002). Comprehensive proteomic analysis of the human spliceosome. *Nature*, 419, (6903): 182-5.
457. Zlotorynski E, Rahat A, Skaug J, Ben-Porat N, Ozeri E, Hershberg R, Levi A, Scherer SW, Margalit H, Kerem B (2003). Molecular basis for expression of common and rare fragile sites. *Mol Cell Biol.*, 23, (20): 7143-51.
458. Zuberi SM, Eunson LH, Spauschus A, De Silva R, Tolmie J, Wood NW, McWilliam RC, Stephenson JP, Kullmann DM, Hanna MG (1999). A novel mutation in the human voltage-gated potassium channel gene (Kv1.1) associates with episodic ataxia type 1 and sometimes with partial epilepsy. *Brain*, 122, (Pt 5): 817-25.

ΔΗΜΟΣΙΕΥΣΕΙΣ

Folate-sensitive fragile site *FRA10A* is due to an expansion of a CGG repeat in a novel gene, *FRA10AC1*, encoding a nuclear protein ☆☆☆

Theologia Sarafidou,^{a,1} Christina Kahl,^{b,1} Isabel Martinez-Garay,^{c,1} Marie Mangelsdorf,^{d,1} Stefan Gesk,^b Elizabeth Baker,^d Maria Kokkinaki,^a Polly Talley,^e Edna L. Maltby,^e Lisa French,^f Lana Harder,^b Bernd Hinzmann,^g Carlo Nobile,^h Kathy Richkind,ⁱ Merran Finnis,^d Panagiotis Deloukas,^f Grant R. Sutherland,^d Kerstin Kutsche,^c Nicholas K. Moschonas,^a Reiner Siebert,^b and Jozef Gécz^{d,*}

European Collaborative Consortium for the Study of ADLTE²

^aDepartment of Biology, University of Crete, and Institute of Molecular Biology and Biotechnology(IMBB), Foundation of Research and Technology (FORTH-GR), P.O. Box 2208, 714 09 Heraklion, Crete, Greece

^bInstitute of Human Genetics, University Hospital Schleswig-Holstein Campus Kiel, Schwanebergweg 24, D-24105 Kiel, Germany

^cInstitute of Human Genetics, University Hospital Hamburg-Eppendorf, Butenfeld 42, D-22529 Hamburg, Germany

^dDepartment of Genetic Medicine, Women's and Children's Hospital, 72 King William Road, North Adelaide, SA 5006 Adelaide, Australia, and Department of Paediatrics, Adelaide University, Adelaide, Australia

^eNorth Trent Cytogenetics Department, Sheffield Children's NHS Trust, Sheffield S10 2TH, UK

^fThe Wellcome Trust Sanger Institute, Hinxton, Cambridge CB10 1SA, UK

^gmetaGen Pharmaceuticals GmbH, Oudenarder Strasse 16, D-13347 Berlin, Germany

^hCNR Institute of Neurosciences, Section of Padua, Viale G. Colombo 3, 35121 Padua, Italy

ⁱGenzyme, Genzyme Genetics, Santa Fe, NM 87505, USA

Received 8 October 2003; accepted 31 December 2003

Available online 23 April 2004

Abstract

Fragile sites appear visually as nonstaining gaps on chromosomes that are inducible by specific cell culture conditions. Expansion of CGG/CCG repeats has been shown to be the molecular basis of all five folate-sensitive fragile sites characterized molecularly so far, i.e., *FRAXA*, *FRAXE*, *FRAXF*, *FRA11B*, and *FRA16A*. In the present study we have refined the localization of the *FRA10A* folate-sensitive fragile site by fluorescence in situ hybridization. Sequence analysis of a BAC clone spanning *FRA10A* identified a single, imperfect, but polymorphic CGG repeat that is part of a CpG island in the 5'UTR of a novel gene named *FRA10AC1*. The number of CGG repeats varied in the population from 8 to 13. Expansions exceeding 200 repeat units were methylated in all *FRA10A* fragile site carriers tested. The *FRA10AC1* gene consists of 19 exons and is transcribed in the centromeric direction from the *FRA10A* repeat. The major transcript of ~1450 nt is ubiquitously expressed and codes for a highly conserved protein, FRA10AC1, of unknown function. Several splice variants leading to alternative 3' ends were identified (particularly in testis). These give rise to FRA10AC1 proteins with altered COOH-termini. Immunofluorescence analysis of full-length, recombinant EGFP-tagged FRA10AC1 protein showed that it was present exclusively in the nucleoplasm. We show that the expression of *FRA10A*, in parallel to the other cloned folate-sensitive fragile sites, is caused by an expansion and subsequent methylation of an unstable CGG

☆ The gene symbol *FRA10AC1* was approved by the HUGO Gene Nomenclature Committee for the gene name fragile site, folic acid type, rare, fra(10)(q23.3) candidate 1, alias *C10orf4*.

☆☆ Supplementary data for this article may be found on ScienceDirect.

* Corresponding author. Fax: +61-8-8161-7342.

E-mail address: jozef.gecz@adelaide.edu.au (J. Gécz).

¹ These authors contributed equally to this work.

² A list of members of the Consortium is given under Acknowledgments.

trinucleotide repeat. Taking advantage of three cSNPs within the *FRA10A* gene we demonstrate that one allele of the gene is not transcribed in a *FRA10A* carrier. Our data also suggest that in the heterozygous state *FRA10A* is likely a benign folate-sensitive fragile site. © 2004 Elsevier Inc. All rights reserved.

Keywords: Folate-sensitive fragile site; Triplet repeat expansion; Novel gene; Nuclear protein

Chromosomal fragile sites are loci that show gaps or breaks in metaphase spreads of cells grown under specific cell culture conditions. In general, common and rare fragile sites are distinguished. To date, more than 80 common fragile sites have been identified in the human genome [1]. Some of these, like the most frequently seen fragile site, *FRA3B* in 3p14.2, have been proposed to be regions of genomic instability implicated in, e.g., breakage events leading to recurrent chromosomal aberrations in cancer cells [for review see 2 and 3]. Although some of these common fragile sites were sequenced, no recognizable sequence elements have been identified that would be sufficient to confer fragility to these chromosomal sites.

According to the culture conditions under which fragile sites are induced, rare fragile sites are classified into folate-sensitive and distamycin A- or BrdU-requiring fragile sites. Twenty-two folate-sensitive fragile sites have been cytogenetically described, the best known being *FRAXA* in Xq27.3. Expression of *FRAXA* is associated with the fragile X syndrome (MIM309550), the most frequent familial form of mental retardation [4]. Similarly, expression of another rare gonosomal folate-sensitive fragile site, *FRAXE*, leads to mild to borderline mental retardation (MIM309548) [5]. To date, two autosomal folate-sensitive fragile sites, *FRA11B* and *FRA16A*, have been cloned [6,7], but only the former has been associ-

ated with a human disease, Jacobsen syndrome, which is caused by terminal deletion of part of the long arm of chromosome 11 [8].

The common molecular basis of the currently characterized rare fragile sites is the expansion of repetitive DNA elements. Expression of the distamycin A-inducible *FRA16B* site in 16q22 and of the BrdU-requiring *FRA10B* site in 10q25 is caused by expansion of 33- and 42-bp AT-rich minisatellite repeats, respectively [9,10]. In contrast, expression of all five folate-sensitive rare fragile sites cloned so far, i.e., *FRAXA*, *FRAXE*, and *FRAXF* on the X chromosome as well as *FRA16A* in 16p13.11 and *FRA11B* in 11q23.3, is uniformly based on expansion of CCG/CGG-trinucleotide repeats. The majority of characterized folate-sensitive fragile site repeats are located in the regulatory or 5' untranslated regions (UTRs) of transcribed genes, probably with the exception of *FRA16A* for which an associated gene has not yet been reported [7]. All of these repeats follow the paradigm of expansion and hypermethylation with subsequent gene silencing as a result [for review see 11]. The CGG-repeat expansion is the only known molecular mechanism underlying expression of folate-sensitive fragile sites. It is yet unknown why this kind of genomic instability leading to chromosomal fragility affects only a very limited subset of all CGG-repeats within the human genome.

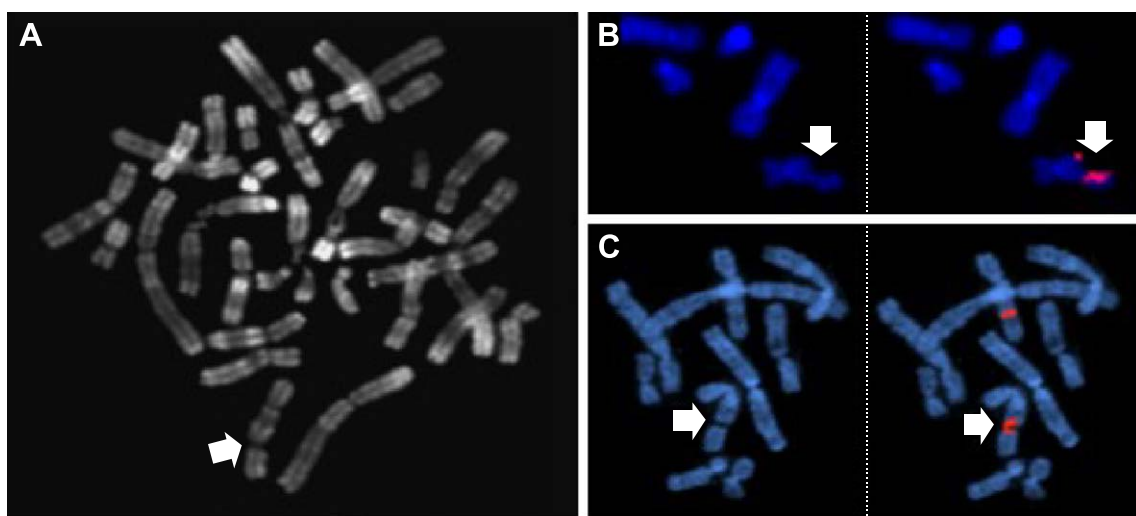


Fig. 1. Cytogenetic and FISH-mapping of the *FRA10A* site. (A) Fluorescence R-banded metaphase of case I expressing the *FRA10A* as a clear gap (arrow) in chromosomal band 10q23.3. (B) Partial metaphase of case I stained with DAPI (left) and hybridized with YAC 912C7 (right). The red signal from the YAC probe is split due to expression of the *FRA10A* site (arrow) and can be seen on both the centromeric and the telomeric fragment, indicating that YAC 912C7 spans the fragile site. (C) Partial metaphase of case III stained with DAPI (left) and hybridized with BAC 437J2 (right), which produces a signal proximal and distal to the gap, indicating that this clone is spanning the *FRA10A* site (arrow).

To learn more about the molecular basis of rare fragile site expression, we have characterized *FRA10A* in 10q23.3, the most prevalent among the rare folate-sensitive autosomal fragile sites in the human genome [12]. It has been estimated that about 1 in 500 individuals manifest *FRA10A*. Similar to the other autosomal folate-sensitive fragile sites, no distinct phenotype has been associated with expression of *FRA10A*. Nevertheless, some studies have proposed that fragile site expression in 10q23–q24 might be associated with mental retardation [12,13], tumorigenesis [14,15], or neurological disorders, including schizophrenia [16]. It is likely that these associations can be attributed to ascertainment bias. The 10q23–q24 chromosomal region has also been shown to contain tumor suppressor genes [17,18]; however their relation to the *FRA10A* is not clear.

In the present work we have refined the localization of the *FRA10A* fragile site and identified its molecular basis. We show that expression of *FRA10A* is caused by an

expansion of a polymorphic CGG repeat within the 5'UTR of a ubiquitously expressed gene, *FRA10AC1*, encoding a novel nuclear protein.

Results

Fluorescence in situ hybridization (FISH) mapping of the FRA10A site

To narrow down the locus containing *FRA10A*, a sequential FISH approach was applied to metaphase spreads of the index patient (case I). In this patient, cytogenetic expression of *FRA10A* resulted in chromosomal breakage on one homolog at 10q23.3 in two-thirds of cells (Fig. 1A). In a first step, double-color FISH was performed with YAC clones from a recently established contig of 10q23–q24 using YAC 790H10 as an internal control that hybridizes to the *MXI1* gene in 10q25 ([18,19] and R.S., unpublished

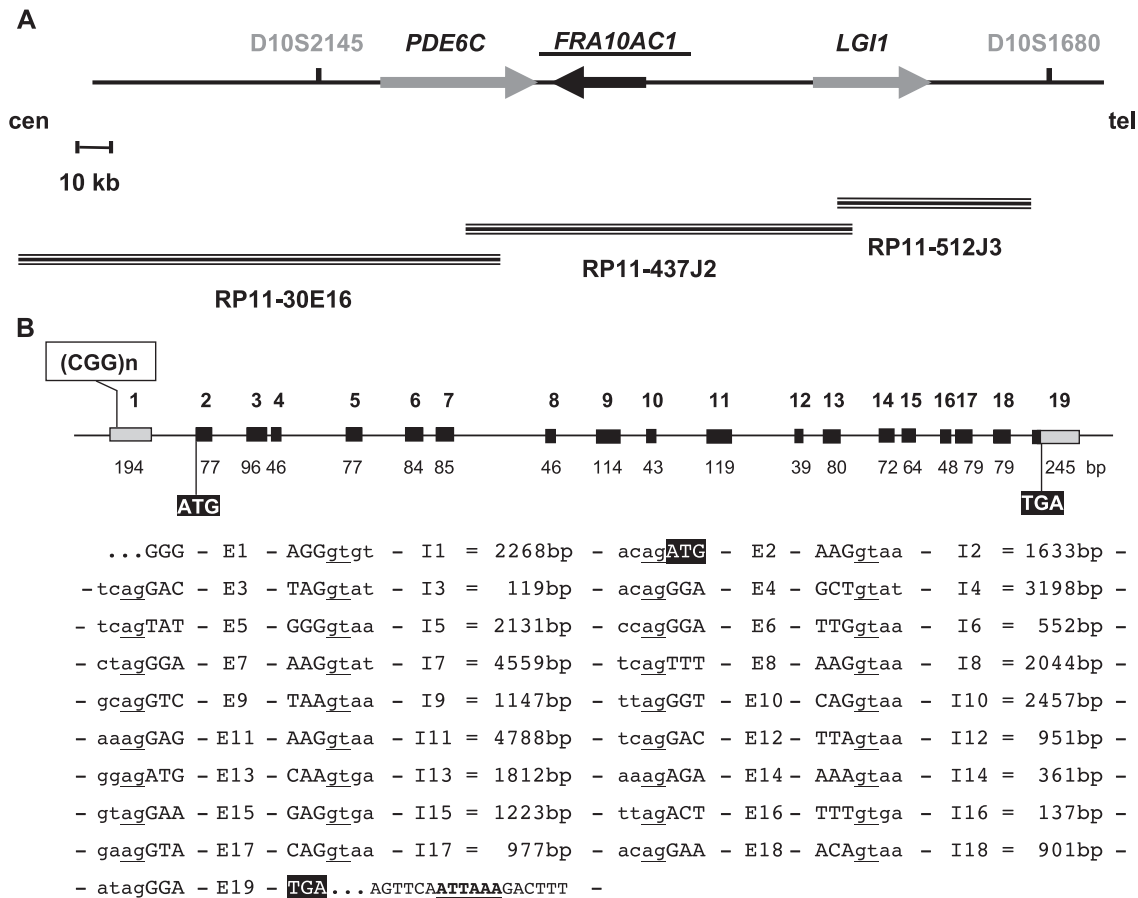


Fig. 2. Mapping of the *FRA10A* locus and genomic organization of the *FRA10AC1* gene. (A) Schematic representation of the genomic region containing the genes *FRA10AC1*, *PDE6C*, and *LGI1* as well as two microsatellite markers. Genes are indicated by large arrows. The 5' → 3' orientation of each gene is given by the orientation of the arrow. BAC clones are represented by black lines. (B) Exon–intron structure of *FRA10AC1*. In the upper part, the genomic structure of *FRA10AC1* is shown schematically (not drawn to scale). Exons are indicated by boxes. Coding regions of exons are shown in black, whereas noncoding regions are in gray. The number of each exon is depicted above and the size below the exon. Start as well as stop codons and the (CGG)_n repeat are indicated. In the lower part, partial sequences of the exon–intron boundaries and the intron sizes are given. Exon sequences are given in uppercase letters and intron sequences in lowercase. Exon is abbreviated by “E” and intron by “I”. The ATG and TGA codons are shown in black boxes. The conserved “ag” and “gt” bases at the intron boundaries are underlined. The polyadenylation signal ATAAA is in bold and underlined.

data). YAC clones 905F4 and 912C7 showed signals distal and proximal to the fragile site, indicating that these clones span the fragile site in the index case (Fig. 1B and Supplementary Table S1).

Based on the available physical map of 10q23–q24 that is partially covered by the inserts of the breakpoint-spanning YACs 905F4 and 912C7, contiguous BAC clones of the chromosome 10 sequence tile path [12] were selected to delineate the fragile site more precisely. FISH analysis showed that BAC clones RP11-341B18 and RP11-30E16 hybridized proximally, whereas RP11-562K6, -237M6, -76P2, -121A3, -162K11, and -397J18 were distal to the fragile site. Clones RP11-568I23, -437J2, -512J3, -306O10, and -367C3 spanned the fragile site in the index case (Supplementary Table S2).

To verify mapping of the *FRA10A* fragile site, metaphase spreads of three additional *FRA10A* carriers (cases II–IV) were analyzed by FISH (Supplementary Table S3). As in the index case, BAC 437J2 was shown to span the fragile site in all of them (Fig. 1C and Supplementary Table S3).

Sequence analysis of the *FRA10A* locus

FISH mapping refined the *FRA10A* locus to an ~50-kb region. Interestingly, BAC clone RP11-437J2 spans the entire *FRA10A* site (EMBL Accession No. AL157396). The total length of the RP11-437J2 clone insert is 162,840 bp. The true right end of this clone is at position 128,796 in the reported sequence (AL157396). The true left end of RP11-437J2 is at position 129,920 of AL356214 (RP11-30E16). Database searches identified the 3' end of the *PDE6C* gene encoding phosphodiesterase 6C as well as the first two 5' exons of the gene encoding the leucine-rich, glioma inactivated 1 protein, *LGII* (Fig. 2), within this *FRA10A* spanning BAC. A single (CGG)₈ repeat was identified in the very central part of the sequence (kb 54.809–54.832 of AL157396). Adjacent to it lies an incomplete and invariable CCA repeat. Both repeats are part of a CpG island of a novel gene, named *FRA10AC1* (fragile site, folic acid type, rare, fra(10) (q23.3) candidate 1), encoding a protein with unknown function (see below).

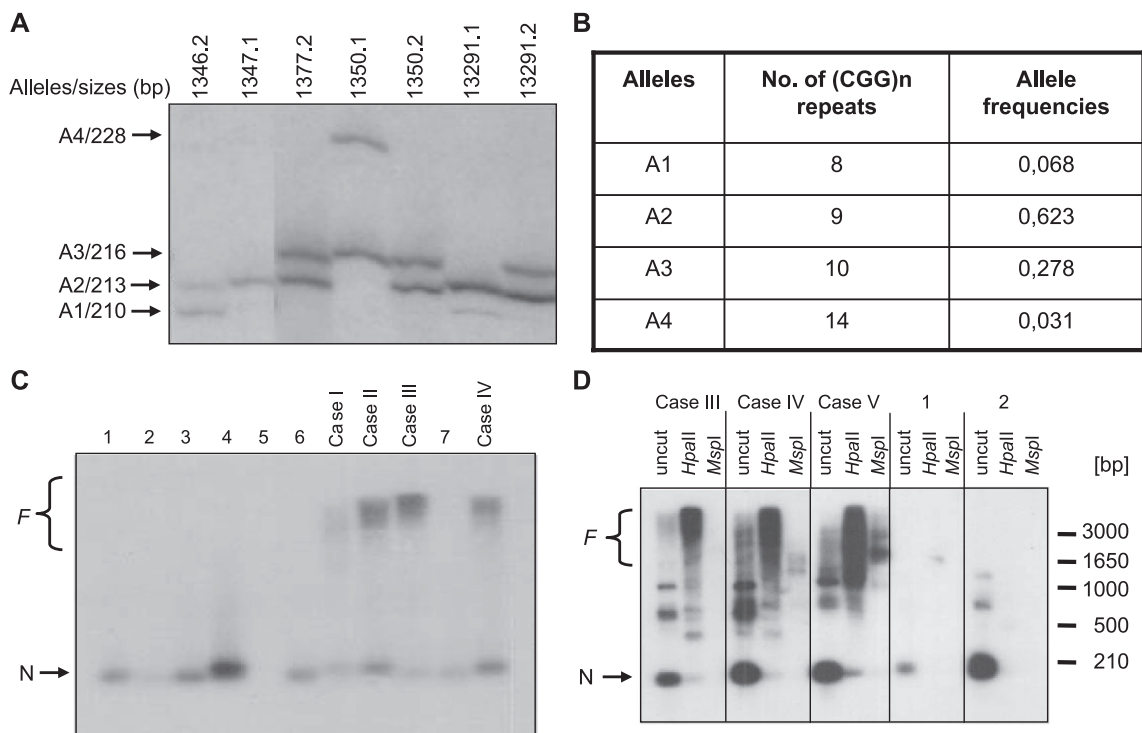


Fig. 3. CGG-repeat polymorphism in *FRA10A* carriers and controls. (A) Analysis of the CGG polymorphism in the 5'UTR of *FRA10AC1* in individuals of the CEPH families. Autoradiogram showing the four alleles (A1–A4) of the repeat detected by PCR. Seven representative genotypes are shown. (B) Allele distribution of the polymorphic *FRA10AC1* repeats in 81 unrelated individuals of the CEPH families. (C) Analysis of the (CGG)_n trinucleotide repeat in controls (lanes 1–7) and individuals expressing the *FRA10A* site (cases I, II, III, IV, and V). In addition to a normal allele (labeled with “N”) individuals who cytogenetically express the fragile site show PCR fragments of larger size (labeled with “F”), indicating repeat expansion. (D) Methylation analysis of the *FRA10A* CGG repeats. DNA from *FRA10A* heterozygous individuals (cases III, IV, and V) and controls (1 and 2) were digested with *HpaII* or *MspI* restriction enzyme, prior to PCR. The (CGG)_n repeat was then PCR-amplified from the restricted and unrestricted genomic DNA (uncut). Hybridization signals were obtained only for the expanded (CGG)_n repeats in individuals expressing *FRA10A* after digestion with *HpaII*, indicating that the region around the (CGG)_n repeat is hypermethylated. Approximate sizes of DNA marker bands are indicated on the right.

Analysis of the CGG repeat in controls and carriers of the FRA10A site

PCR-based amplification of the identified CGG repeat in the 5'UTR of *FRA10AC1* in various individuals revealed this repeat to be polymorphic. Testing of 81 unrelated individuals from 40 CEPH families identified four alleles (A1–A4) with repeat size $n = 8, 9, 10,$ and 14 . The most common allele, A2, contained nine CGG repeats (Figs. 3A and 3B). In contrast, individuals cytogenetically expressing the fragile site showed much larger PCR fragments (detected as a smear of higher molecular weight products), indicating expansion of the repeat (Fig. 3C). One normal CGG allele was detected in all *FRA10A* carriers, which is consistent with the heterozygous status of the repeat expansion. To test whether this expanded CGG repeat is subject to methylation (as is the case for the other cloned folate-sensitive fragile sites), we digested genomic DNA from *FRA10A* carriers and controls with a methylation-sensitive (*HpaII*) or -insensitive (*MspI*) restriction enzyme. Following digestion, PCR, Southern blot, and hybridization with a GCC/GGC probe, the expanded CGG repeats were detected only from the undigested or *HpaII*-digested genomic DNA and not from the *MspI*-digested DNA of *FRA10A* carriers (cases III, IV, and V in Fig. 3D). These data indicate that the expanded CGG repeat is hypermethylated. Control individuals showed the presence of normal allele only. In all experiments, the expanded alleles presented as a smear rather as a distinct

band, suggesting some difficulties for the *Taq* polymerase to amplify across the expanded CGG repeat. The size of the expanded alleles in *FRA10A* carriers exceeded that of the normal alleles by at least 600–800 bp (up to several kilobases). Thus, the number of CGG repeat units in *FRA10A* can be estimated to be at least 200.

Full-length cDNA sequence and genomic organization of the FRA10AC1 gene

By combining in silico data with results obtained from RT-PCRs, we assembled the full-length *FRA10AC1* cDNA sequence of 1344 bp that contains an open reading frame encoding a protein of 315 amino acids (GenBank Accession No. AJ431721). The genomic organization of *FRA10AC1* was established by BLASTN searches of the complete cDNA sequence with the genomic sequence of clone RP11-437J2 (GenBank Accession No. AL157396). In total, this novel gene consists of 19 exons (Fig. 2B). Nonetheless, 4 exons (exons 13, 14, 15, and 16) are alternatively spliced in (see below), and exon 17 has been identified at least once, e.g., in the EST clone GKBBABE07 (GenBank Accession No. AV682434). *FRA10AC1* spans a region of 33 kb and is transcribed from telomere to centromere (Fig. 2A). All intron boundaries follow the AG/GT rule of the consensus splice site. The (CGG) n repeat is embedded in exon 1, within the putative 5'UTR of the *FRA10AC1* gene, and is part of the CpG island. The ATG codon at position 195 (in

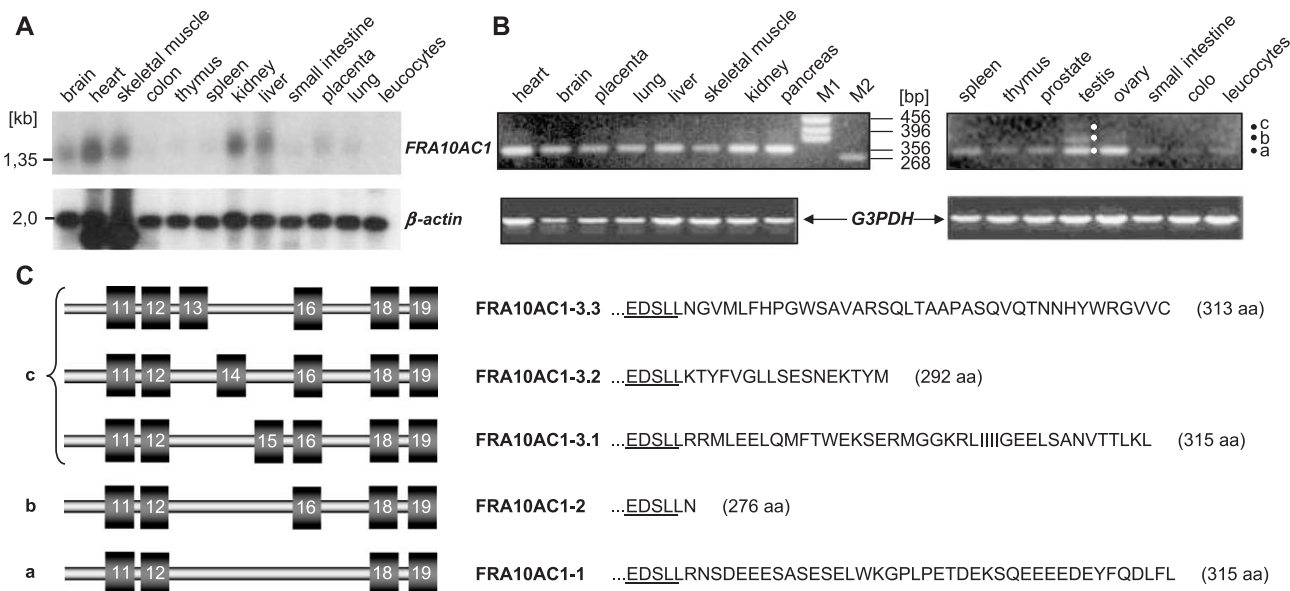


Fig. 4. Expression pattern of *FRA10AC1* in various human tissues and identification of alternative splicing variants. (A) Northern blot analysis of *FRA10AC1*. Hybridization of a *FRA10AC1* exon 19-containing probe to a human multiple-tissue Northern blot containing total RNA. A single transcript of approximately 1450 nt was detectable in numerous tissues (top). The 1.35-kb band of the marker is indicated. The same blot was reprobed with a β -actin probe as control (bottom). (B) RT-PCRs with human MTC cDNA panels. *FRA10AC1*-specific PCR product spanning exons 11 to 19 was amplified from all tissues tested (top). White and black dots indicate the three different transcript variants “a”, “b”, and “c”. pBluescript/*HinfI* (M1) and λ EcoRV (M2) were used as DNA size markers; sizes of marker bands are indicated. Control RT-PCR amplification of a *G3PDH* product was done from the same cDNA samples (bottom). (C) Identification of alternative splice variants of the *FRA10AC1* gene. Schematic representation of the alternatively spliced *FRA10AC1* variants a, b, and c (left) and their derived amino acid sequences (right). Exons are indicated by dark gray boxes and are numbered. Introns are depicted as thick gray lines. Amino acids encoded by exon 12, common to all *FRA10AC1* variants, are underlined.

exon 2) was considered the most likely translation start site conforming to the Kozak rule. The translation termination codon, TGA, was identified in-frame with this start codon at position 1140 bp in the cDNA sequence (in exon 19). By using the ExPASy proteomics tools, the encoded protein was calculated to have a theoretical pI of 8.25 and a calculated molecular mass of 37.548 kDa. Its most basic region lies between amino acids 200 and 275 (theoretical $pI \sim 10.0$), whereas the carboxy-terminus (aa 280–315) represents the most acidic region ($pI \sim 3.70$). Various serine residues, particularly at the C-terminus (aa 250–313), are potential sites for phosphorylation [20]. By database searches using the Pfam and Smart software tools, we found no significant similarities of FRA10AC1 to any known protein domains or repeats. However, a bipartite nuclear localization signal (NLS; at aa positions 223–238 and 245–262), which is embedded within a larger lysine-rich region (aa 129–269), was identified.

Expression analysis of FRA10AC1

For Northern blot analysis, we used a FRA10AC1-specific probe containing the first 616 bp of the 5' coding region and a probe covering part of exon 19. We obtained identical hybridization patterns for both probes. A single transcript of approximately 1450 bp, which corresponds well to the estimated cDNA size, was detected in all adult tissues analyzed (Fig. 4A), indicating ubiquitous expression of FRA10AC1. Strong expression of FRA10AC1 was found in brain, heart, skeletal muscle, kidney, and liver, whereas low expression was found in colon, thymus, spleen, small intestine, placenta, and lung. In leukocytes, the transcript is barely detectable (Fig. 4A). To investigate further the expression profile of the FRA10AC1 gene, we employed RT-PCR on normalized first-strand cDNAs of a panel of human tissues using primers spanning exons 11 to 19. Amplification was done under nonsaturating conditions allowing semiquantitative estimation of the relative FRA10AC1 transcript levels. In general, ubiquitous expression of FRA10AC1 was confirmed (Fig. 4B).

In addition to the tissues tested by Northern blot, FRA10AC1-specific transcripts were detected in pancreas, testis, ovary, and, at lower levels, prostate and leukocytes (Fig. 4B). In the majority of the tissues examined, a fainter and larger band ("b" in Fig. 4B) was found in addition to the abundant FRA10AC1 transcript ("a" in Fig. 4B) indicating the presence of a splice variant. Moreover, in ovarian and particularly in testicular tissue, a third band with slower mobility (variant "c" in Fig. 4B) appeared. Sequence analysis showed that these transcript variants are due to an alternative splicing of exons 13–16 (Fig. 4C). In the abundant variant a, exon 12 is spliced directly to exon 18 (GenBank Accession No. AJ431721). The ubiquitously expressed larger transcript variant b contains the 48-bp exon 16 spliced between exons 12 and 18 (GenBank Accession No. AJ431725). The largest amplicon, c, turned out to be a

mixture of three different FRA10AC1 transcript variants, in which exon 13, 14, or 15 is individually spliced in, in addition to exon 16 (GenBank Accession Nos. AJ431722, AJ431723, and AJ431724). Comparison of the derived amino acid sequences of the five splicing variants showed that the encoded C-terminal regions are different (Fig. 4C).

cSNP/RT-PCR analysis of FRA10AC1 in a FRA10A carrier individual

To investigate that one allele of the FRA10AC1 gene is transcriptionally silenced by the methylation at the FRA10A fragile site, we performed cSNP/RT-PCR on cDNA from one FRA10A carrier (case I, Fig. 3). First, at least three cSNPs were detected within the FRA10AC1 open reading frame by analysis of available ESTs (Hs.303727): one at position 47 in exon 2 (c.47G → A, 17 ESTs with the G and 11 with the A allele), the second at position 90 in exon 3 (c.90T → C, 22 ESTs with T and 8 with C), and the third at position 233 in exon 5 (c.233G → C, 21 ESTs with G and 9 with C). Second, FRA10A carriers for whom we had cell lines available were tested for these three cSNPs and one was found informative with genotype g.47A–g.90T/C–g.233C/G (case I). When the exon 2/5 RT-PCR product (primers available upon request) of this FRA10A carrier was sequenced, only one allele was detected: c.47A–c.90C–c.233G (see Fig. 5). This result demonstrated that one FRA10AC1 allele (c.47A–c.90T–c.233C in this case) was not transcribed. That the silenced allele is the one with the FRA10A fragile site is highly likely but was not formally established (due to the lack of available family material).

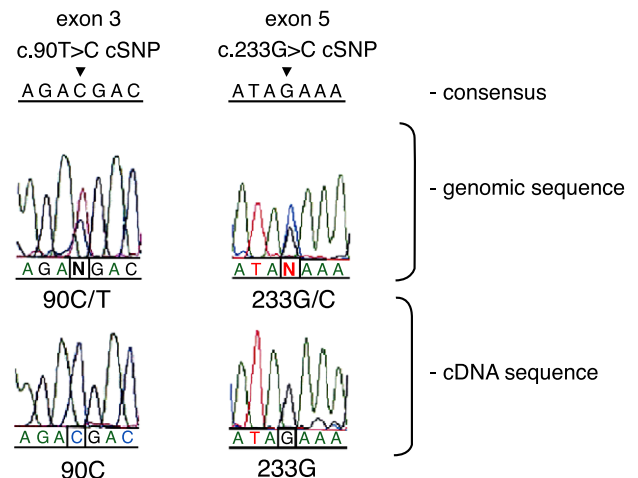


Fig. 5. cSNP/RT-PCR on FRA10A carrier individual. Partial genomic and cDNA sequence of exon 3 (c.90T → C) and exon 5 (c.233G → C) cSNPs of the FRA10AC1 gene of a FRA10A carrier individual (case I) is shown. While the genomic sequence shows heterozygosity for T/C and G/C alleles at both positions, the cDNA sequence clearly shows the presence of only one allele, c.90C–c.233G. This individual was not informative for the c.47G → A cSNP. The consensus FRA10AC1 sequence (GenBank Accession No. AJ431721) is also shown.

The *FRA10AC1* protein is located to the nucleus

To examine the subcellular localization of the *FRA10AC1* protein and, thereby, the functional potential of the predicted NLS, COS-7 cells were transiently transfected with an enhanced green fluorescent protein (EGFP)–*FRA10AC1* expression plasmid. The cloned *FRA10AC1* cDNA corresponds to the most abundant splice variant a (Fig. 4C). Upon transient transfection the green fluorescence was diffusely detected in the nucleus (Figs. 6A and B-a). However, in about 10% of the transfected cells a punctate pattern was found (Fig. 6B-d). Coimmunostaining of cells with lamin A/C antibody, which marks

the nuclear envelope, suggested that *FRA10AC1* is located exclusively in the nucleoplasm (Fig. 6B-f).

FRA10AC1 is a conserved protein

We used the human *FRA10AC1* cDNA sequence to search for ortholog ESTs in the databases. Several overlapping ESTs were found. Three of them (GenBank Accession Nos. BQ939866, AA840143, and BE946023) were selected for the assembly of the mouse *FRA10AC1* ortholog (size 1380 bp; GenBank Accession No. BN000292). *Fra10Ac1* maps to mouse chromosome 19C2–C3, between the genes *Pde6c* and *Lgi1*, in a

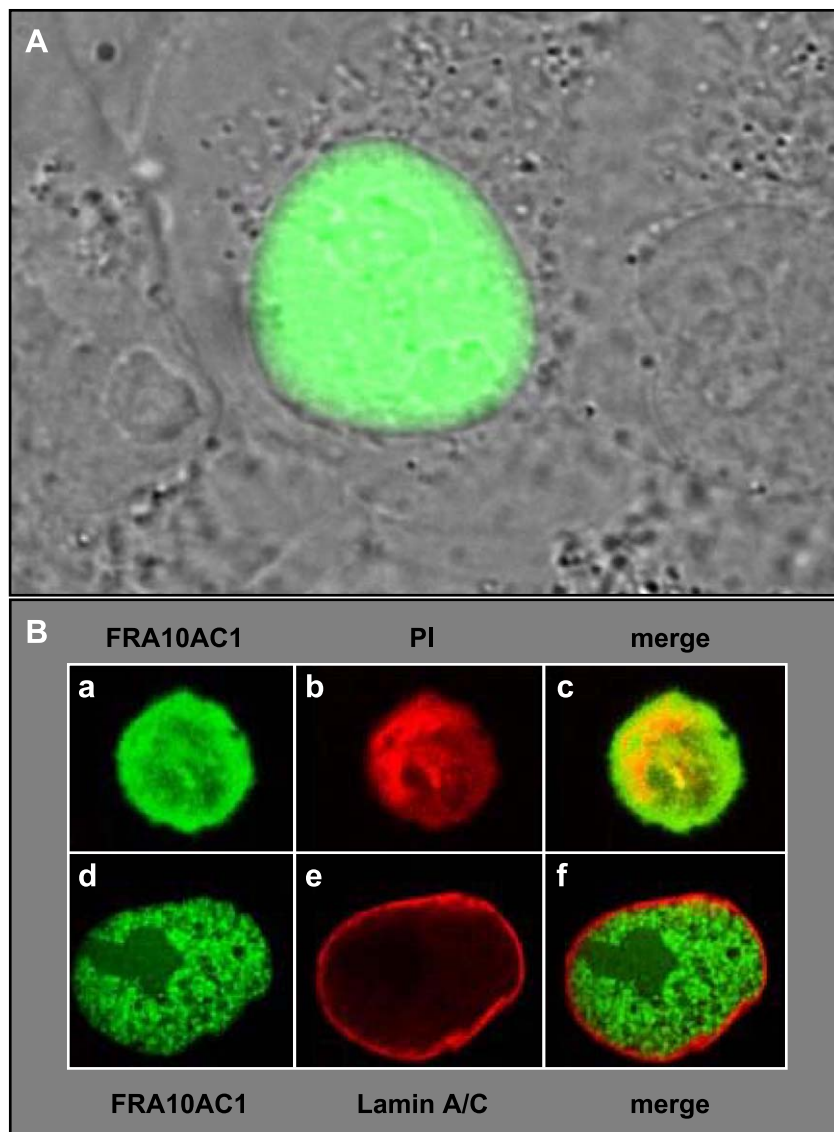


Fig. 6. Subcellular localization of *FRA10AC1*. (A) Subcellular localization of the EGFP-tagged *FRA10AC1* fusion protein (GATEWAY compatible) detected by fluorescence microscopy. Representative COS-7 cells from transient transfections were photographed. More than 95% of transfected cells showed identical patterns of subcellular localization. (B) The EGFP-tagged *FRA10AC1* protein is exclusively detected in the nucleoplasm of transfected COS-7 cells. (a) and (d) show fluorescence photomicrographs of nuclei with the green signal of *FRA10AC1*. (b) Propidium iodide staining of the same cell as shown in (a). (c) Merge of (a) and (b). (e) Labeling of the nuclear envelope of the same cell as shown in (d) with a lamin A/C antibody. (f) Merge of (d) and (e).

region syntenic to human chromosome 10q23–q24. *Fra10Ac1* encodes a protein of the same size (315 aa) compared to human and both proteins share a sequence identity of 86% and similarity of 92% (Fig. 7A). BLASTP analysis revealed that numerous FRA10AC1 orthologs exist, in both animal and plant species (Fig. 7B). The FRA10AC1 proteins identified in plants (*Arabidopsis thaliana* and *Pimpinella brachycarpa*) and *Caenorhabditis elegans* show about 35% identity to the human protein compared to 58% in insects (*Anopheles gambiae*) (Fig. 7B). These data suggest that a common ancestor gene had existed before plants and animals diverged. It is interesting to note that the central portion (aa 105–240 of human FRA10AC1; Fig. 7B) of the

protein is 45–67% identical among all orthologs analyzed, suggesting a highly conserved domain with an important function.

Discussion

In the present study we have elucidated the molecular basis of the most prevalent autosomal folate-sensitive fragile site, *FRA10A* [12], to be an unstable CGG repeat. The *FRA10A* repeat is similar to those identified at the other five folate-sensitive fragile sites, *FRAXA* [21,22], *FRAXE* [23], *FRAXF* [24,25], *FRA11B* [8], and *FRA16A* [7], that have been cloned so far. Overall, the *FRA10A*

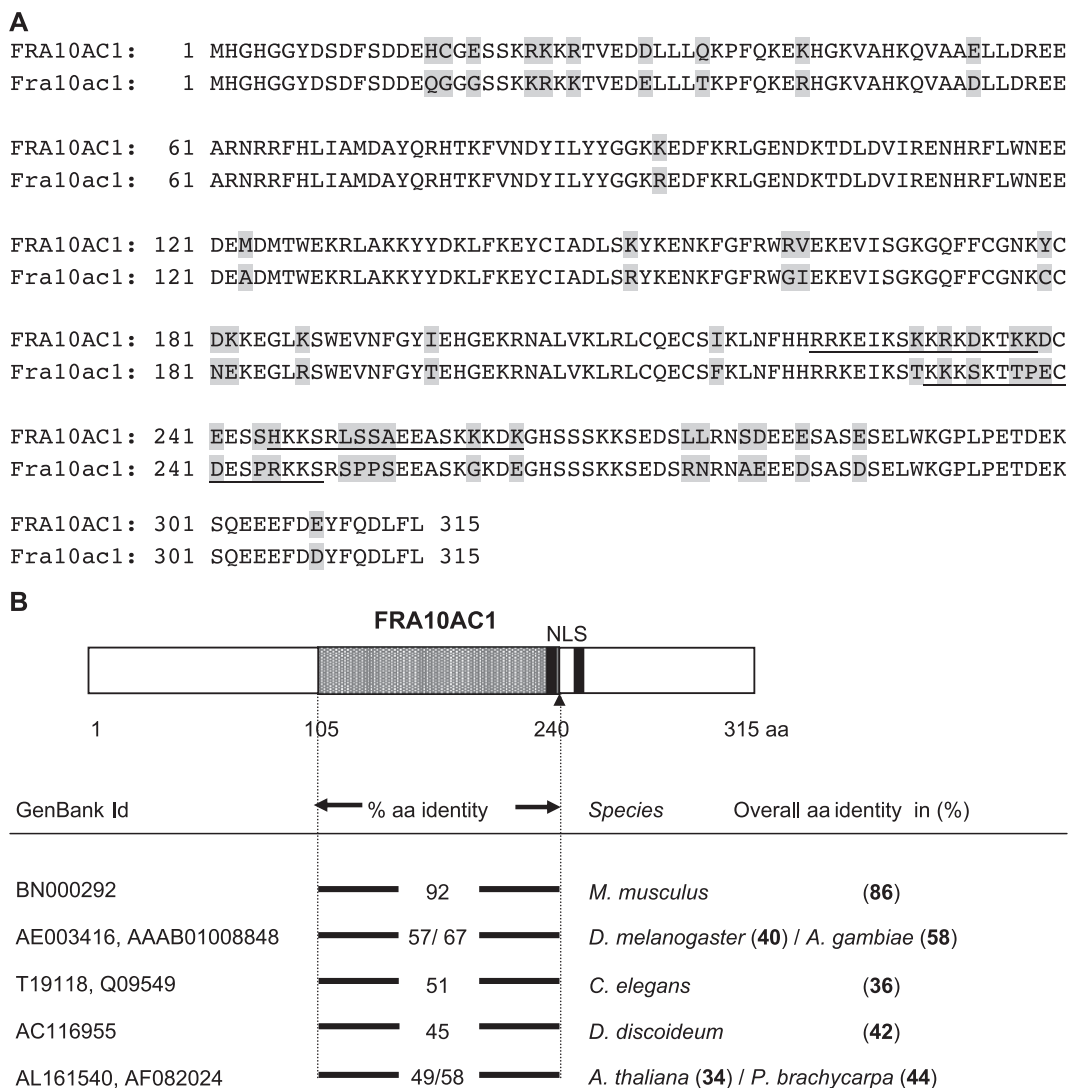


Fig. 7. Conservation of the FRA10AC1 protein. (A) Amino acid sequence alignment of the predicted human and mouse FRA10AC1 polypeptides. Amino acids that differ between human and mouse are highlighted in gray. The bipartite nuclear localization signals (NLS) are underlined. Numbers at the left and right indicate amino acid positions. (B) Schematic representation and conservation of the human FRA10AC1 protein. Top: The central conserved region (aa 105–240) is depicted as a gray box. The bipartite NLS (aa 223–238 and 245–262) is indicated by two black bars. Bottom: On the left, GenBank accession numbers of the orthologs are given. Amino acid identity within the conserved central part of FRA10AC1 is given in the middle, whereas the overall protein identity is shown on the right.

CGG repeat does not differ from the other known folate-sensitive fragile site-associated CCG/CGG repeats. The repeat is normally polymorphic with four alleles of 8, 9, 10, and 14 repeat units. In *FRA10A* carriers the repeat expands beyond 200 repeats (600–800 bp) and in that size it becomes also methylated. It appears to be somatically stable, although this was not rigorously tested. Among the 81 individuals (162 chromosomes) tested so far we have not seen any premutations.

Similar to the other folate-sensitive fragile sites, perhaps with the exception of *FRA16A* [7], *FRA10A* is associated with a transcription unit. To investigate that the expansion and methylation of the CGG repeat in the 5'UTR of the *FRA10AC1* gene results in the loss of its expression we tested RNA of one *FRA10A* carrier. For this purpose we took advantage of the presence of three cSNPs in the *FRA10AC1* mRNA. While genomic PCR showed heterozygosity for cSNPs at positions 90 (c.90T → C) and 233 (c.233G → C), only one *FRA10AC1* mRNA allele (c.90C and c.233G) has been detected by RT-PCR. This result clearly demonstrated that one allele of the *FRA10AC1* gene is not transcribed in this individual. It is highly likely that this *FRA10AC1* allele is transcriptionally silenced as a direct consequence of the presence of the *FRA10A* fragile site on this chromosome. To our knowledge this region of chromosome 10 is not imprinted. Whether the lack of one *FRA10AC1* mRNA allele expression is compensated for at the posttranscriptional level was not investigated.

While for the three X-linked folate-sensitive fragile sites the association with a particular phenotype is more (*FRAXA* and *FRAXE*) or less (*FRAXF*) resolved, the situation is not clear-cut for the two autosomal sites, *FRA11B* and *FRA16A*. Jones et al. [26] characterized an additional eight CCG repeats from the 40-Mb region of Jacobsen patient deletion breakpoints. These authors concluded that the nonrandom clustering of chromosome deletion breakpoints with CCG repeats suggests that they may play an important role in a common mechanism of chromosome breakage, perhaps including also the *FRA11B* CCG repeat. The great majority of Jacobsen syndrome patients, however, do not have their chromosome 11 deletion breakpoint at or near *FRA11B* [27]. As for the *FRA16A* site [28], there is no phenotype associated with this fragile site reported yet.

Carrier individuals of *FRA10A* analyzed as part of this study do have a clinical phenotype broadly defined as learning difficulties or developmental delay. However, this phenotype most likely cannot be attributed to the expression of the fragile site and likely haploinsufficiency of the *FRA10AC1* gene expression, but rather to an ascertainment bias (these individuals would have been primarily screened for *FRAXA*). Additional evidence supporting ascertainment bias as a likely cause comes (where available) from the evaluation of the parents of *FRA10A* carriers who did carry the fragile site, but did not display any clinical phenotype similar to that found in their

offspring. On the other hand it is intriguing that lack of one copy of the *FRA10AC1* gene does not give any obvious phenotype. The *FRA10AC1* gene is highly conserved and widely expressed and has no obvious paralogs, which could indicate some redundancy of its function. Recently, Shaw et al. [29] identified the *FAM11A* gene, whose expression was silenced as a consequence of the presence of the *FRAXF* fragile site. Only one individual with *FRAXF* was analyzed for *FAM11A* expression, which was extinguished. Although this individual did not present with any obvious *FRAXF*-related phenotype it is too early to speculate that the lack of *FAM11A* does not cause any discernible phenotype [29]. This also applies for *FRA10A* for which additional cases together with further functional studies of the *FRA10AC1* gene will probably help to resolve this issue. Similar early controversy about the existence of an associated phenotype, debated for the *FRAXE* site and mental retardation, has later been resolved by postulating a minor and/or partly redundant function for the *FMR2* gene [for review see 5]. For *FRA10AC1* it is not clear yet whether haploinsufficiency would have any pathologic consequences in tissues where the gene is expressed. It is interesting to observe that the highest *FRA10AC1* expression is seen in metabolically active tissues like heart, skeletal muscle, brain, kidney, and liver. Whether the single copy of the *FRA10AC1* gene in *FRA10A* carriers can be sufficiently up-regulated in times of high demand needs to be addressed.

Contrary to the fact that *FRA10A* is one of the most frequently seen folate-sensitive fragile sites so far, to our knowledge, there is no *FRA10A* homozygote reported. Such a case would be very desirable (if not in utero lethal) in addressing the function and associated phenotype (if any) of the *FRA10AC1* gene. However, this could be achieved by manipulating *FRA10AC1* expression in any of the model organisms with *FRA10AC1* orthologs, including fly, worm, or mouse.

The *FRA10A* fragile site lies about 50 kb from the 5' end of the *LGII* gene, which was recently implicated in autosomal-dominant partial epilepsy with auditory features [30,31]. Moreover, down-regulation of *LGII* has been shown during malignant progression of astrocytic gliomas due to a hitherto unknown mechanism [32]. Although the presence of some long-range regulatory elements of the *LGII* gene affected by the expression of the *FRA10A* fragile site can be speculated, such hypothesis has not been tested. Neither an increased incidence of tumors nor such type of epilepsy has been noticed in *FRA10A* carriers. There does not seem to be any other gene present in the 50-kb region containing *FRA10AC1* and *LGII* (Fig. 2A).

In conclusion, *FRA10A* represents another unstable CCG trinucleotide repeat-triggered folate-sensitive fragile site. Although no phenotype can be associated with the fragile site as yet, it is intriguing that the *FRA10A*-associated *FRA10AC1* gene is highly conserved and, thus,

should serve an important function. Identification and characterization of the *FRA10A* site and its associated gene will allow the elucidation of any possible genotype–phenotype correlation in the future.

Materials and methods

Chromosome analysis

All cytogenetic studies were performed after obtaining informed consent as part of a diagnostic workup. In the index patient (case I), chromosome analysis was initiated because of speech delay and autistic features and revealed expression of *FRA10A* in 66% of the cells. The expression of the fragile site in 10q23 resulted in chromosomal breakage in the majority of the patient's cells, leading to an acentric fragment of the distal part of 10q, which was gained or lost in part of the metaphases. Another patient (case II), a 1-week-old boy, who presented with mildly dysmorphic features, expressed *FRA10A* in 5 of 20 metaphases. Three other unrelated *FRA10A*-positive individuals (cases III, IV, and V) were detected when tested for the cause of their developmental delay (case V) or developmental delay in another family member (cases III and IV). In two instances the index case (case III and case IV) did not present with *FRA10A*, while one of the parents did. Follow-up studies on the parents and/or siblings of the remaining *FRA10A*-positive patients showed that the *FRA10A* site was inherited from one of the parents, in each case, who was not affected. Chromosome analyses from peripheral blood cells suitable for the demonstration of folate-sensitive fragile sites as well as harvesting, slide preparation, and G or R banding were performed according to standard methods.

YAC and BAC clones

YAC clones were derived from the CEPH-mega-YAC library. A detailed STS-based map of 10q23–q24 containing the YAC clones used in the present study has been published previously [19,33]. BAC clones were derived from the RPCI-11 library (<http://www.chori.org/bacpac/home.htm>). They were part of the BAC contig covering chromosome region 10q23–q24, which was assembled by means of STS-content mapping and fingerprinting (<http://www.sanger.ac.uk/cgi-bin/humace/fpcwebmap.cgi?mode=map&map=bac.10.1.html>) [34,35]. Shotgun sequencing and finishing of clones from this contig including the clone RP11-437J2 (GenBank Accession No. AL157396) was performed as described previously [36].

FISH

BAC and YAC DNA was prepared using the Perfect-Prep Plasmid Maxi Kit (Eppendorf) or the Nucleon MiY

kit (Amersham Pharmacia). Indirect or direct labeling of BAC DNA and *Alu*-PCR products from YACs was performed with the BioPrime DNA labeling system (Invitrogen) or by nick translation. Indirect and direct FISH to metaphase spreads was performed as described previously [37]. Chromosomes were counterstained with DAPI.

Northern blot and RT-PCR analyses

The expression pattern of *FRA10AC1* in human tissues was studied using multiple-tissue Northern blots (BD Biosciences Clontech). The probes used were a 129-bp PCR product representing bp 1124–1252 of exon 19 and a 617-bp PCR product flanking the ATG start codon. The exon 19 probe was radiolabeled directly by addition of [α -³²P]dCTP (Amersham) into the PCR. Fifty nanograms of DNA from a human fetal liver cDNA library (Invitrogen) was amplified using primers F2 (5'-TCAG-GATTTGTTTCTATGAGAC-3') and R2 (5'-CCACACAAGCTGTAAGTTCAGCA-3'). The 5' coding region probe was amplified with primers N19F (5'-GATGCATGGTTCATGGAGGCTATG-3') and N19R (5'-CAAGTGCATTTCTTCTCACCATG-3') and radiolabeled by random priming with hexanucleotides (Invitrogen; Amersham Biosciences) and [α -³²P]dCTP (Amersham Biosciences). Blots were hybridized according to the manufacturer's instructions, washed at high stringency, and exposed to X-ray film. To control for RNA integrity and loading amount, the blots were rehybridized with an α -³²P-labeled β -actin probe as described in the protocol provided. For the analysis of the *FRA10AC1* expression profile, primers F3 (5'-GCAGAAGAGGCTCCAA-3') and R2 were employed in PCRs using cDNA samples (1 ng) from a multiple-tissue panel of normalized first-strand cDNAs (MTC, BD Biosciences Clontech). PCR products were separated by 1.5% agarose gel electrophoresis, visualized using ethidium bromide, and sequence analyzed either by direct sequencing or after subcloning into vector pGEM-T Easy (Promega).

cSNP/RT-PCR analysis of *FRA10AC1* in a *FRA10A* carrier individual

Putative cSNPs within the *FRA10AC1* open reading frame were identified by sequence comparison of available ESTs (Hs.303727). The presence of cSNPs in *FRA10A* carriers was tested by sequencing of their genomic DNA. The following primers were used: for c.47G → A cSNP, E2Fi, 5'-CCTGGACAATTGGCAACATAG-3', and E2Ri, 5'-CATTCACTAGAGCAATTGTCTTC-3', 254-bp amplicon; for c.90T → C cSNP, E3Fi, 5'-ATCTCATTAAAGATCAGT-GAGTC-3', and E3Ri, 5'-AATCCGAAGCAACT-TAAAC-CAAG-3', 283-bp amplicon; and for c.233G → C cSNP, E5Fi, 5'-GACACCAAACAGCTCCTAAGAG-3', and E5Ri, 5'-CATAACCCACATTCCCTTCTATG-3', 170-bp ampli-

con. cDNA was prepared from EBV-transformed lymphoblastoid cell lines of *FRA10A* carriers using standard techniques and tested for the presence of the cSNPs by direct sequencing of an RT-PCR product encompassing exons 2–5. The following primers were used: E2F, 5'-ATGCATGGT-CATGGAGGCTATG-3', and E5R, 5'-CCCAAACGCTT-GAAGTCTTCT-3', 296-bp amplicon. Dye terminator (Big Dye Terminator) sequencing was performed according to the instructions of the supplier of the kit (Perkin–Elmer). Exon PCR products were purified using the UltraClean PCR Clean-up DNA purification kit (MoBio Laboratories, Inc.). All PCR products were sequenced in both forward and reverse directions.

Construction of green fluorescence protein-tagged *FRA10AC1* cDNA expression plasmids

A cDNA fragment corresponding to the 315-aa-long *FRA10AC1* protein was obtained by PCR using as template 1 ng of human testis first-strand cDNA (MTC, BD Biosciences Clontech) with primers F, 5'-CGGGATCCGAAAGGATG CATGGTCATG-3' (*Bam*HI site and ATG start codon underlined), and R4, 5'-CGGAATTCGAAACAAATCCTGAAAATACTC-3' (*Eco*RI site underlined). The 964-bp agarose gel-purified PCR product was cloned into pGEM-T Easy (Promega), sequence verified, excised as a *Bam*HI/*Eco*RI fragment, and ligated to pEGFP-N1 (BD Biosciences Clontech) digested with the same enzymes. The resulting pEGFP construct encoded a GFP-tagged *FRA10AC1* polypeptide. In an alternative approach, an amplicon of the *FRA10AC1* ORF compatible with the GATEWAY cloning system (Invitrogen) was generated from fetal brain cDNA with primers N19C3F (5'-GGGGACAAGTTTGTACAAAAAAGCAGGCTTGCATGGT-CATGGAGGCTATGATTC-3') and N19C3R (5'-GGGGACCCTTTGTACAAGAAAGCTGGGTGT-TACTATAGAAACAAATCCTGAAAATACTCATC-3') using cDNA Advantage polymerase (BD Biosciences Clontech). PCR products were purified according to the instructions. Plasmid pEGFP-C3 (BD Biosciences Clontech) was digested with *Sma*I and purified, and 5' phosphates were removed by alkaline phosphatase. After a second purification step by phenol/chloroform extraction, the plasmid was ligated to the GATEWAY Reading Frame Cassette C (RfC.1) of the GATEWAY Vector Conversion System (Invitrogen) according to the manufacturer's instructions. The clones obtained represent GATEWAY-compatible "destination vectors" that were sequenced for integrity and correct orientation of the cassette with primers 5'attR1 (5'-CACATTATACGAGCCGGAAGCAT-3') and 3'attR2 (5'-CAGTGTGCCGGTCTCCGTTATCG-3'). These vectors were propagated in *Escherichia coli* host DB3.1 (Invitrogen) to compensate for expression of the *ccdB* gene present on the cassette. Construction of the EGFP-*FRA10AC1* plasmid with the GATEWAY system expressing EGFP fused to the N-terminus was carried out as previously described [38].

Cell culture, transfection, and immunocytochemistry

COS-7 cells were grown in DMEM containing 10% fetal calf serum and gentamycin. About 3×10^4 cells grown on glass coverslips were transfected with 0.5–1.0 μ g of GFP-tagged *FRA10AC1* plasmid DNA using the calcium phosphate coprecipitation method [39]. Thirty-six hours after transfection, cells were washed with $1 \times$ phosphate-buffered saline (PBS) and fixed for 10 min with 4% paraformaldehyde followed by three washes with PBS. After fixation, cells were treated for 5 min with 20 mM glycine followed by three washes with PBS. Cells were permeabilized with blocking buffer (0.2% Triton X, 2 mM $MgCl_2$, 1% fish gelatin in PBS) for 5 min and incubated with a lamin A/C mouse monoclonal antibody (Santa Cruz Biotechnology) for 1 h at room temperature, then washed three times in blocking buffer, and incubated for 1 h with fluorescence-conjugated secondary antibody, Alexa Fluor 568 goat anti-mouse IgG conjugate (Molecular Probes). Primary and secondary antibodies were used at dilutions of 1:500 and 1:1000, respectively. After incubation, cells were washed three times with blocking buffer and once with PBS. Cell staining with propidium iodide was accomplished after treatment of the cells with RNase A for 30 min at room temperature. Processed cells were visualized with a Leica SP confocal microscope.

For transient expression of the EGFP-*FRA10AC1* GATEWAY-compatible plasmid, COS-7 cells were transfected and analyzed as described in Martinez-Garay et al. [38].

PCR analysis of the *FRA10AC1* CGG repeat

Genotyping of the *FRA10AC1* CGG repeat in 81 unrelated individuals, parents, or grandparents of the 40 CEPH families was performed by amplification of 100 ng of genomic DNA using 1 U of *Taq* polymerase (Minotech) with primers F1 (5'-CCATTTAGATGTAACACGTG-3') and R1 (5'-GTCGTTTCCTTCTTTCCC-3'). Primer F1 was added to the reaction after 5' end-labeling with [α - 32 P]dATP. PCR products were separated on a denaturing acrylamide–urea DNA sequencing gel and visualized by X-ray autoradiography.

To amplify across the CGG repeat in *FRA10A* carriers, primers 5'-TCGGACCGGAAGTGGTGCCC-3' and 5'-TCCCTGTGTGCCAAGTTCGC-3' were applied in a 20- μ l reaction containing 100 ng of genomic DNA; $1 \times$ buffer (Expand Long Template Buffer 2; Roche); 200 μ M each dCTP, dTTP, and dATP; 500 μ M 7-deaza dGTP; 10% (v/v) DMSO; 20 pmol of each oligonucleotide; and 0.4 μ l enzyme mix (Expand Long Template; Roche). Ten microliters of each PCR was run on a 1.5% agarose gel with $1 \times$ TBE. The gel was then blotted on a nylon membrane (Amersham) and hybridized with a GCC/GGC oligonucleotide probe.

Labeling and hybridization using the GCC/GGC probe

This was done essentially as described for the poly(AC)_n/(GT)_n probe by Thompson et al. [40]. Briefly, reagents as for random priming were used; however, the random hexamers were omitted. Instead, two complementary synthetic CCG oligonucleotides (30-mers; 10 copies of the CGG/CCG repeat) were used. They prime off of each other to generate the probe. Fifty nanograms of each oligonucleotide was used in the labeling reaction followed by conventional hybridization at 65°C and washing at 42–65°C. After a final stringent wash (0.1× SSC at 65°C for 20 min), the X-ray films were developed after an overnight exposure at room temperature. For the PCR studies to assess the methylation status of the expanded *FRA10AC1* CGG repeat, 1 µg of genomic DNA was digested with *HpaII* or *MspI* restriction enzyme before PCR amplification. Undigested and digested DNA was PCR-amplified using *FRA10AC1* flanking oligonucleotides (see above), run on a 1.5% agarose gel, blotted, and hybridized using the GCC/GGC probe.

Sequence analysis

Analysis of DNA and protein sequences was performed using BLAST (<http://www.ncbi.nlm.nih.gov/blast/>), SMART (<http://smart.embl-heidelberg.de/>), <http://www.ebi.ac.uk/interpro/scan.html>, Pfam (<http://www.sanger.ac.uk/Software/Pfam/search.shtml>), Ensembl (<http://www.ensembl.org/>), GoldenPath (<http://genome.ucsc.edu/>), NetPhos (www.cbs.dtu.dk/services/NetPhos), and ExpASY (<http://us.expasy.org/tools/#proteome>) servers.

Acknowledgments

This study was supported by grants from the Deutsche Krebshilfe (10-1643-Si1), the Hensel-Stiftung Kiel, the National Health and Medical Research Council of Australia, the Deutsche Forschungsgemeinschaft (GRK336), and the Greek Ministry of Development–Secretariat for Research and Technology (No. 236/234/603 and 6th Session/37) to N.K.M. We thank A. Pasparki (IMBB-FORTH) and Dorit Schuster for excellent technical assistance. Other members of the European Collaborative Consortium for the Study of ADLTE are J. Galan, A. Gorostidi, L. Kluwe, A. López de Munain, J.F. Martí Massó, V.F. Mautner, R. Michelucci, J.M. Morante-Redolat, J. Perez-Tur, J.J. Poza, J.F. Prud'homme, A. Rosenthal, A. Sáenz, E. Staub, U. Stephani, and C.A. Tassinari.

References

- [1] G.R. Sutherland, F. Hecht, Fragile Sites on Human Chromosomes, Oxford Univ. Press, New York, 1985.
- [2] Y. Pekarsky, A. Palamarchuk, K. Huebner, C.M. Croce, *FHIT* as

- tumor suppressor: mechanisms and therapeutic opportunities, *Cancer Biol. Ther.* 1 (2002) 232–236.
- [3] M.F. Arlt, A.M. Casper, T.W. Glover, Common fragile sites, *Cytogenet. Genome Res.* 100 (2003) 92–100.
- [4] B. Bardoni, J.L. Mandel, G.S. Fisch, *FMR1* gene and fragile X syndrome, *Am. J. Med. Genet.* 97 (2000) 153–163.
- [5] J. Gécz, The *FMR2* gene, *FRAXE* and non-specific X-linked mental retardation: clinical and molecular aspects, *Ann. Hum. Genet.* 64 (2000) 95–106.
- [6] C. Jones, et al., Physical linkage of the fragile site *FRA11B* and a Jacobsen syndrome chromosome deletion breakpoint in 11q23.3, *Hum. Mol. Genet.* 3 (1994) 2123–2130.
- [7] J.K. Nancarrow, et al., Implications of *FRA16A* structure for the mechanism of chromosomal fragile site genesis, *Science* 264 (1994) 1938–1941.
- [8] C. Jones, et al., Association of a chromosome deletion syndrome with a fragile site within the proto-oncogene *CBL2*, *Nature* 376 (1995) 145–149.
- [9] S. Yu, et al., Human chromosomal fragile site *FRA16B* is an amplified AT-rich minisatellite repeat, *Cell* 88 (1997) 367–374.
- [10] D.R. Hewett, et al., *FRA10B* structure reveals common elements in repeat expansion and chromosomal fragile site genesis, *Mol. Cell* 1 (1998) 773–781.
- [11] G.R. Sutherland, Rare fragile sites, *Cytogenet. Genome Res.* 100 (2003) 77–84.
- [12] P. Petit, J.P. Fryns, H. van den Berghe, F. Hecht, Population cytogenetics of autosomal fragile sites, *Clin. Genet.* 29 (1986) 96–100.
- [13] A. Mavrou, M. Syrou, C. Tsenghi, C. Metaxotou, Autosomal folate sensitive fragile sites in normal and mentally retarded individuals in Greece, *Am. J. Med. Genet.* 38 (1991) 437–439.
- [14] L. Lacombe, et al., Microsatellite instability and deletion analysis of chromosome 10 in human prostate cancer, *Int. J. Cancer* 69 (1996) 110–113.
- [15] S.K. Kim, et al., Identification of two distinct tumor-suppressor loci on the long arm of chromosome 10 in small cell lung cancer, *Oncogene* 17 (1998) 1749–1753.
- [16] B.J. Mowry, et al., Second stage of a genome scan of schizophrenia: study of five positive regions in an expanded sample, *Am. J. Med. Genet.* 96 (2000) 864–869.
- [17] O.B. Chernova, R.P. Somerville, J.K. Cowell, A novel gene, *LGII*, from 10q24 is rearranged and downregulated in malignant brain tumors, *Oncogene* 17 (1998) 2873–2881.
- [18] R. Siebert, et al., Deletions in the long arm of chromosome 10 in lymphomas with t(14;18): a pathogenetic role of the tumor suppressor genes *PTEN/MMAC1* and *MXI1*? *Blood* 92 (1998) 4487–4489.
- [19] C. Nobile, et al., A refined physical and EST map spanning 7.4 Mb of 10q24, a region involved in neurological disorders, *Mamm. Genome* 9 (1998) 835–837.
- [20] N. Blom, S. Gammeltoft, S. Brunak, Sequence- and structure-based prediction of eukaryotic protein phosphorylation sites, *J. Mol. Biol.* 294 (1999) 1351–1362.
- [21] E.J. Kremer, et al., Mapping of DNA instability at the fragile X to a trinucleotide repeat sequence p(CCG)_n, *Science* 252 (1991) 1711–1714.
- [22] A.J. Verkerk, et al., Identification of a gene (*FMR-1*) containing a CCG repeat coincident with a breakpoint cluster region exhibiting length variation in fragile X syndrome, *Cell* 65 (1991) 905–914.
- [23] S.J.L. Knight, et al., Trinucleotide repeat amplification and hypermethylation of a CpG island in *FRAXE* mental retardation, *Cell* 74 (1993) 127–134.
- [24] J.E. Parrish, et al., Isolation of a GCC repeat showing expansion in *FRAXF*, a fragile site distal to *FRAXA* and *FRAXE*, *Nat. Genet.* 8 (1994) 229–235.
- [25] R.J. Ritchie, et al., The cloning of *FRAXF*: trinucleotide repeat expansion and methylation at a third fragile site in distal Xqter, *Hum. Mol. Genet.* 3 (1994) 2115–2121.
- [26] C. Jones, et al., Co-localisation of CCG repeats and chromosome

- deletion breakpoints in Jacobsen syndrome: evidence for a common mechanism of chromosome breakage, *Hum. Mol. Genet.* 9 (2000) 1201–1208.
- [27] R.C. Michaelis, et al., Most Jacobsen syndrome deletion breakpoints occur distal to *FRA11B*, *Am. J. Med. Genet.* 76 (1998) 222–228.
- [28] J.K. Nancarrow, et al., Molecular basis of p(CCG)_n repeat instability at the *FRA16A* fragile site locus, *Hum. Mol. Genet.* 4 (1995) 367–372.
- [29] M.A. Shaw, P. Chiurazzi, D.R. Romain, G. Neri, J. Géczy, A novel gene, *FAM11A*, associated with the *FRAXF* CpG island is transcriptionally silent in *FRAXF* full mutation, *Eur. J. Hum. Genet.* 10 (2002) 767–772.
- [30] S. Kalachikov, et al., Mutations in *LGII* cause autosomal-dominant partial epilepsy with auditory features, *Nat. Genet.* 30 (2002) 335–341.
- [31] J.M. Morante-Redolat, et al., Mutations in the *LGII*/Epitempin gene on 10q24 cause autosomal dominant lateral temporal epilepsy, *Hum. Mol. Genet.* 11 (2002) 1119–1128.
- [32] R. Besleaga, M. Montesinos-Rongen, J. Perez-Tur, R. Siebert, M. Deckert, Expression of the *LGII* gene product in astrocytic gliomas: downregulation with malignant progression, *Virchows Arch.* 443 (2003) 561–564.
- [33] I.C. Gray, et al., An integrated physical and genetic map spanning chromosome band 10q24, *Genomics* 43 (1997) 85–88.
- [34] D.R. Bentley, et al., The physical maps for sequencing human chromosomes 1, 6, 9, 10, 13, 20 and X, *Nature* 409 (2001) 942–943.
- [35] J.D. McPherson, et al., A physical map of the human genome, *Nature* 409 (2001) 934–941.
- [36] E. Lander, et al., Initial sequencing and analysis of the human genome, *Nature* 409 (2001) 860–921.
- [37] B. Schlegelberger, S. Metzke, S. Harder, R. Zühlke-Jenisch, Y. Zhang, R. Siebert, Classical and molecular cytogenetics of tumor cells, in: R.D. Wegner (Ed.), *Diagnostic Cytogenetics*, Springer-Verlag, Berlin, 1999, pp. 151–185.
- [38] I. Martinez-Garay, et al., A new gene family (*FAM9*) of low-copy repeats in Xp22.3 expressed exclusively in testis: implications for recombinations in this region, *Genomics* 80 (2002) 259–267.
- [39] F.L. Graham, A.J. van der Eb, A new technique for the assay of infectivity of human adenovirus 5 DNA, *Virology* 52 (1973) 456–467.
- [40] A.D. Thompson, et al., Isolation and characterisation of (AC)_n microsatellite genetic markers from human chromosome 16, *Genomics* 13 (1992) 402–408.

The DNA sequence and comparative analysis of human chromosome 10

P. Deloukas¹, M. E. Earthrowl¹, D. V. Grafham¹, M. Rubenfield^{2,3}, L. French¹, C. A. Steward¹, S. K. Sims¹, M. C. Jones¹, S. Searle¹, C. Scott¹, K. Howe¹, S. E. Hunt¹, T. D. Andrews¹, J. G. R. Gilbert¹, D. Swarbreck¹, J. L. Ashurst¹, A. Taylor¹, J. Battles², C. P. Bird¹, R. Ainscough¹, J. P. Almeida¹, R. I. S. Ashwell¹, K. D. Ambrose¹, A. K. Babbage¹, C. L. Bagguley¹, J. Bailey¹, R. Banerjee¹, K. Bates¹, H. Beasley¹, S. Bray-Allen¹, A. J. Brown¹, J. Y. Brown¹, D. C. Burford¹, W. Burrill¹, J. Burton¹, P. Cahill², D. Camire², N. P. Carter¹, J. C. Chapman¹, S. Y. Clark¹, G. Clarke¹, C. M. Clew¹, S. Clegg¹, N. Corby¹, A. Coulson¹, P. Dhami¹, I. Dutta¹, M. Dunn¹, L. Faulkner¹, A. Frankish¹, J. A. Frankland¹, P. Garner¹, J. Garnett¹, S. Gribble¹, C. Griffiths¹, R. Grocock¹, E. Gustafson^{2,3}, S. Hammond¹, J. L. Harley¹, E. Hart¹, P. D. Heath¹, T. P. Ho², B. Hopkins¹, J. Horne², P. J. Howden¹, E. Huckle¹, C. Hynds², C. Johnson¹, D. Johnson¹, A. Kana², M. Kay¹, A. M. Kimberley¹, J. K. Kershaw¹, M. Kokkinaki⁴, G. K. Laird¹, S. Lawlor¹, H. M. Lee², D. A. Leongamornlert¹, G. Laird¹, C. Lloyd¹, D. M. Lloyd¹, J. Loveland¹, J. Lovell¹, S. McLaren¹, K. E. McLay¹, A. McMurray¹, M. Mashreghi-Mohammadi¹, L. Matthews¹, S. Milne¹, T. Nickerson¹, M. Nguyen², E. Overton-Larty¹, S. A. Palmer¹, A. V. Pearce¹, A. I. Peck¹, S. Pelan¹, B. Phillimore¹, K. Porter¹, C. M. Rice¹, A. Rogosin^{2,3}, M. T. Ross¹, T. Sarafidou⁴, H. K. Sehra¹, R. Shownkeen¹, C. D. Skuce¹, M. Smith¹, L. Standing², N. Sycamore¹, J. Tester¹, A. Thorpe¹, W. Torcasso², A. Tracey¹, A. Tromans¹, J. Tsolas^{2,3}, M. Wall¹, J. Walsh², H. Wang², K. Weinstock², A. P. West¹, D. L. Willey¹, S. L. Whitehead¹, L. Wilming¹, P. W. Wray¹, L. Young¹, Y. Chen⁵, R. C. Lovering⁶, N. K. Moschonas⁴, R. Siebert⁷, K. Fichtel², D. Bentley¹, R. Durbin¹, T. Hubbard¹, L. Doucette-Stamm^{2,3}, S. Beck¹, D. R. Smith^{2,3} & J. Rogers¹

¹The Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton CB10 1SA, UK

²Genome Therapeutics Corporation, 100 Beaver Street, Waltham, Massachusetts 02453, USA

³Agencourt Bioscience Corporation, 100 Cummings Center, Beverly, Massachusetts 01915, USA

⁴Department of Biology, University of Crete & Institute of Molecular Biology and Biotechnology, Foundation of Research and Technology, PO Box 2208, 71409 Heraklion, Crete, Greece

⁵European Bioinformatics Institute (EBI), Wellcome Trust Genome Campus, Hinxton CB10 1SD, UK

⁶HUGO Gene Nomenclature Committee, Department of Biology, University College London, London NW1 2HE, UK

⁷Institute of Human Genetics, University Hospital Schleswig-Holstein Campus Kiel, Schwanenweg 24, D-24105 Kiel, Germany

The finished sequence of human chromosome 10 comprises a total of 131,666,441 base pairs. It represents 99.4% of the euchromatic DNA and includes one megabase of heterochromatic sequence within the pericentromeric region of the short and long arm of the chromosome. Sequence annotation revealed 1,357 genes, of which 816 are protein coding, and 430 are pseudogenes. We observed widespread occurrence of overlapping coding genes (either strand) and identified 67 antisense transcripts. Our analysis suggests that both inter- and intrachromosomal segmental duplications have impacted on the gene count on chromosome 10. Multispecies comparative analysis indicated that we can readily annotate the protein-coding genes with current resources. We estimate that over 95% of all coding exons were identified in this study. Assessment of single base changes between the human chromosome 10 and chimpanzee sequence revealed nonsense mutations in only 21 coding genes with respect to the human sequence.

We report here the final clone map and sequence assembly of human chromosome 10. This metacentric chromosome accounts for 4.5% of the genome and is best known for harbouring the *PTEN* tumour suppressor gene and the *RET* proto-oncogene.

With the human genome sequence in hand, the task ahead is to identify the different units of genetic information embedded in the sequence and understand their function both at the molecular and cellular level. In this study we address the former by reporting a comprehensive annotation of manually inspected gene structures and their correlation to sequence variation and other features of the genomic sequence. The annotation process is assessed by comparative analysis to the genome sequence of two rodents, *Mus musculus* and *Rattus norvegicus*, and three fishes, *Tetraodon nigroviridis*, *Fugu rubripes* and *Danio rerio*. Finally, we report our preliminary findings on the distribution of single base differences along human chromosome 10 in comparison to the chimpanzee genome.

The clone map and finished sequence

A clone map spanning the euchromatic regions of the short (p) and long (q) arm of human chromosome 10 was assembled by restriction fingerprint and sequence-tagged site (STS) content analysis¹. We identified clones by screening approximately 85 genomic equivalents of P1-derived artificial chromosome (PAC), bacterial

artificial chromosome (BAC), yeast artificial chromosome (YAC), cosmid and fosmid libraries. The tiling path consists of 1,144 minimally overlapping clones (Supplementary Table S1) organized into 12 contigs (Table 1). Contig-1340 spans the entire p arm and harbours the boundary between euchromatin and heterochromatin with the proximal 250 kilobases (kb) extending into pericentromeric satellite repeats. In a detailed study of this region two further clones carrying satellite 3 sequences, contig-2069, were identified and mapped by pulse field gel electrophoresis (PFGE) ~50 kb proximal to contig-1340 (ref. 2). Similarly, contig-43 harbours the q-arm boundary with the proximal 240 kb composed of satellite repeats³. In addition, clone RP11-745D9, 94% of which consists of α -satellite repeats, was arbitrarily placed proximal to contig-43 as it was suggested to map to human chromosome 10 (E. Eichler, personal communication). A 9.75-megabase (Mb) PFGE map spanning the chromosome 10 centromere⁴ places the core α -satellite block (D10Z1) ~0.2 Mb distal to contig-102 and 1 Mb proximal to contig-43.

In contrast with the p arm, nine gaps remain in the clone map of the q arm. Five of them are clustered within a ~4-Mb region (Table 1 and Fig. 1 (rollfold); see also Supplementary Fig. S1 for a detailed view). Our inability to walk across these five gaps was due to the extensive segmental duplications in this region (Fig. 2);

however, we obtained a size estimate by fluorescent *in situ* hybridization (FISH) of RP11-172C24 (AL512595) and RP11-13E1 (AC013284) on metaphase chromosomes. We sized the remaining four gaps by FISH with clones immediately flanking each gap to extended DNA fibres. No gap was estimated to be larger than 50 kb in size. Altogether the euchromatic gaps account for no more than 840 kb (Table 1). Finally, we defined the location of both telomeres. Clone RP11-631M1 (AL713922) ends ~20 kb away from the telomeric repeats of the p arm based on the telomeric half-YAC XX-YAC22O3 (<http://www.wistar.upenn.edu/Riethman/>). At the end of the q arm (qtel), clone XX-YAC2136 (BX322534) contains part of the telomeric repeat block.

Each clone of the tiling path was subjected to random subcloning and sequencing at either the Genome Therapeutics Corporation (GENE) or the Sanger Institute—the initial draft sequence of a few clones was carried out by other centres that are credited in the corresponding submissions to the EMBL/GenBank/DBJ databases. We finished clones according to the international finishing standard (<http://genome.wustl.edu/Overview/g16stand.php>). Of the 1,144 clones in the human chromosome 10 tiling path, 221 and 913 were finished at GENE and the Sanger Institute, respectively, and three elsewhere (Supplementary Fig. S1). The remaining seven clones show persistent deletion of internal fragments. In total, we finished 131,666,441 base pairs (bp) in 18 sequence contigs; euchromatic coverage is estimated at 99.4%. Sequence accuracy was estimated as described in ref. 5 and found to exceed 99.99%. The sequence assembly comprises all known chromosome 10 messenger RNAs (RefSeq set) and STS markers from available radiation hybrid⁶ and genetic maps^{7,8} (T. Furey, personal communication). In addition, the integrity of the sequence map was independently assessed at the University of California, Santa Cruz, by alignment of fosmid and BAC paired end sequences (<http://genome.cse.ucsc.edu/>). Table 1 lists the size of each sequence contig, with the largest one spanning 44,693,577 bp.

Table 1 Clone and sequence contigs on chromosome 10

Clone contig	Sequence contig	Size (bp)	Gap size estimate (bp)
Telomere	—	—	20,000
1340	BX32259-BX276080	5,574,992	10,000
	AL365356-BX255924	12,337,510	10,000
	AL928729-AL133216	20,793,917	—
Gap	—	—	50,000
2069	AL133173-AL590623	286,100	—
Gap	—	—	50,000
Centromere	—	—	2,380,000
Gap	—	—	50,000
102	BX322613	191,752	—
Gap	—	—	50,000
43	AL031601-AC012044	3,830,268	—
Gap*	—	—	ND*
4000	AL831769-BX649215	952,201	—
Gap*	—	—	ND*
3003	AL450388-AL603965	263,307	—
Gap*	—	—	ND*
101	AL954360	163,321	—
Gap*	—	—	ND*
14	BX547991-AL731733	989,826	—
Gap*	—	—	ND*
15	AL603966-AL731572	1,941,839	10,000
	AL672187	211,435	10,000
	AL589822-AL132656	30,068,948	—
Gap	—	—	10,000
16	AC068139-AL731667	44,693,577	10,000
	AL606510-AL772134	2,696,534	—
Gap	—	—	50,000
17	BX470155-AL607044	4,615,046	—
Gap	—	—	10,000
2493 (includes telomere)	BX294094-BX511297	246,123	10,000
	AL732395-BX322534	1,809,745	—

*Cumulative gap size estimate of 750,000 for all five of the indicated gaps together. ND, not determined.

The gene and protein index

The Sanger Institute has established a standardized annotation pipeline (outlined at <http://vega.sanger.ac.uk/>) in which gene structures are drawn on the basis of human interpretation of the combined supportive evidence generated during sequence analysis. Annotation of the human chromosome 10 sequence resulted in a total of 1,787 gene structures that we then classified, as described in ref. 9, into: (1) 654 ‘known’ genes; (2) 162 ‘novel genes’; (3) 219 ‘novel transcripts’; (4) 322 ‘putative genes’; and (5) 430 ‘pseudogenes’. Pseudogenes were further subdivided into processed (371) and unprocessed (59).

Excluding the pseudogenes, human chromosome 10 is a chromosome with an average gene density (10.4 genes Mb⁻¹). The 1,357 genes span 66,309,730 bp in total (mean 51,335 bp per gene). Therefore, 50.6% of the analysed sequence is transcribed, matching the figure reported for chromosome 22 (51%), which is gene-rich, but appearing elevated in comparison with chromosomes 6, 7, 14 and 20 (42.2%, 46.5%, 43.6% and 42.4%, respectively), which have gene densities similar to chromosome 10. The latter suggests that the human chromosome 10 genes have on average a larger genomic span than those on chromosomes 6, 7, 14 and 20. Gene size along human chromosome 10 varies enormously, with the two extremes being *CTNNA3* (1,776,209 bp) and *CALML5* (859 bp). Exons account for only 2.3% of the sequence and the mean exon size is 313 bp. The longest and shortest exons annotated in this study have a length of 9,763 (*SH3MD1*) and 3 bp (*CDH23*), respectively. *CDH23* is also the gene with most exons (69) on this chromosome. Table 2 summarizes the features of each gene class.

Alternative splicing is a major contributor to the complexity of the human transcriptome. We annotated a total of 4,204 transcripts for 1,357 gene structures (Table 2). No splice variants were annotated on the basis of alternative polyadenylation sites. Approximately 73% of the protein-coding genes have more than one transcript and 5.8 on average. For *ADD3* we annotated 22 variants. Note that the use of partial expressed sequences (for example, expressed sequence tags (ESTs)) may result in the annotation of more than one transcript per splice variant. Given this caveat, our analysis suggests a significantly higher level of alternative splicing compared with previous estimates¹⁰. Approximately 50% of the 3,456 transcripts (known and novel) do not seem to encode a protein. Annotation of these transcripts is largely based on ESTs, many of which may correspond to aberrant transcripts. Their precise role is largely unknown but they may be part of the machinery of transcriptional regulation (for example, via nonsense-mediated decay). Nevertheless, there are 1,837 transcripts with an open reading frame (ORF). Of the 342 genes with at least two transcripts having a complete ORF (that is, possess both a 5' and a 3' untranslated region (UTR)), 312 encode at least two distinct peptides.

Identification of transcription start sites and promoter regions remains a challenge in the annotation process. First, we scanned the human chromosome 10 unmasked sequence and predicted a total of

Figure 1 The sequence map of chromosome 10 and its features (see rollfold). The current sequence assembly (v1.1) and that used in the analysis presented in this study (v1.0) are available at <http://www.sanger.ac.uk/HGP/Chr10>. Tracks from top to bottom are: (1) the scale bar (in Mb); (2) the sequence map of human chromosome 10 represented by a black solid bar interrupted by clone and sequence gaps (grey); (3) syntenic blocks in the mouse (top track) and the rat (bottom track) where blocks are colour-coded per chromosome and labelled with the chromosome number and coordinates (Mb) (for example, 2: 3.1–11.3 Mb; unordered sequence contigs are tagged as random); (4) predicted CpG islands (brown); (5) regions of sequence homology to *Fugu* (blue), zebrafish (dark blue) and *Tetraodon* (dark pink); and (6) protein-coding genes. Gene names are in dark blue for the known and black for the novel CDS.

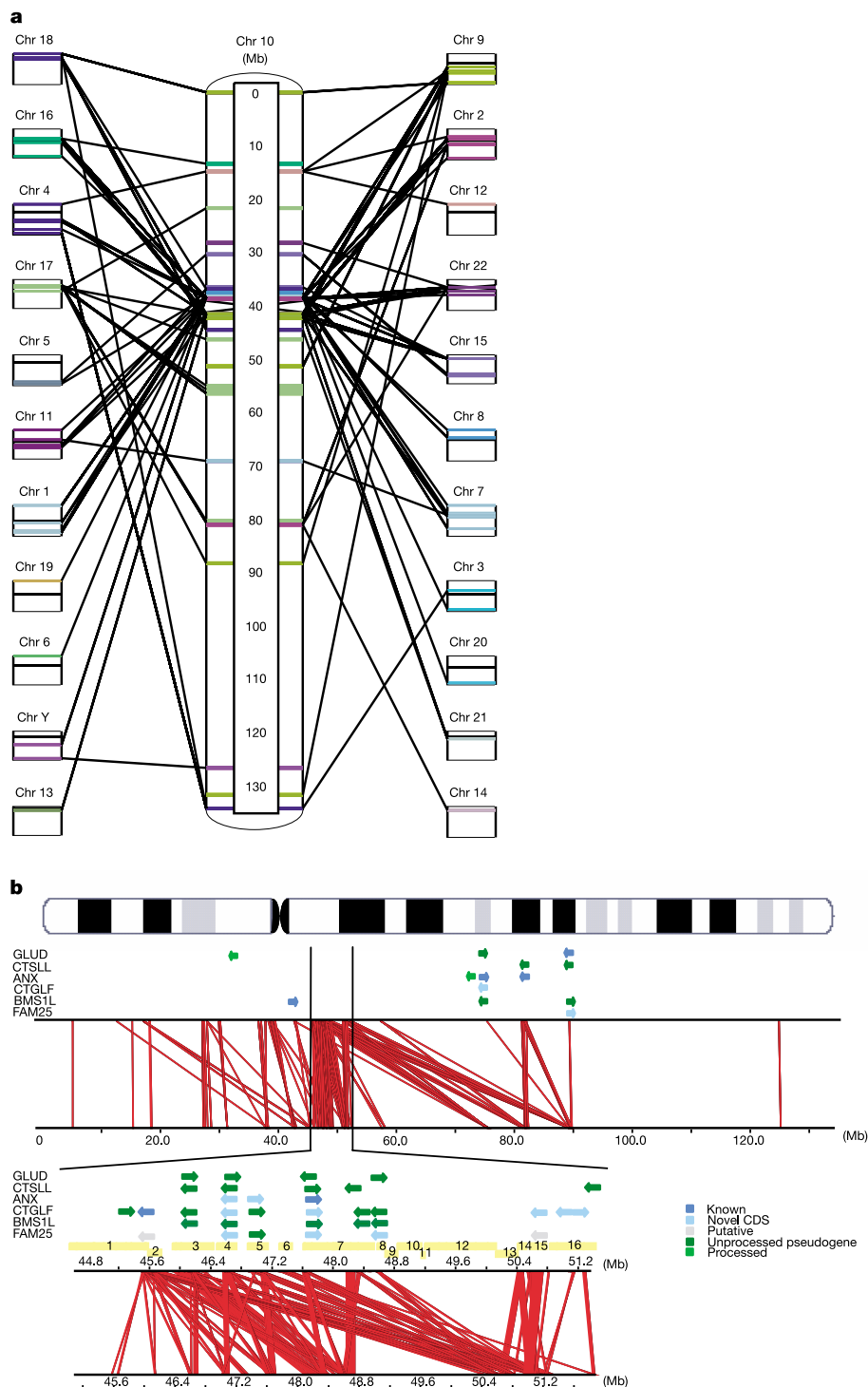


Figure 2 Chromosome 10 inter- and intrasegmental duplications. **a**, Interchromosomal duplications across chromosome 10 showing blocks of 10 kb or greater. Duplicated regions are colour-coded per chromosome and indicated as lines (thickness is proportional to physical distance). Each chromosome other than 10 is represented by an open black rectangle with a black line representing the centromere. **b**, 10q11:10q22:10q23.1:10q23.3 intrasegmental duplications. Top row, ideogram of chromosome 10; second row, positions of members of the six main gene families outside 10q11 (colour-coded per gene class; arrows indicate transcription); third row, intrachromosomal duplications across the whole chromosome showing blocks of 10 kb or greater; bottom row, enlarged view of the 10q11 region. Yellow bars represent sequence (bottom row for sequences submitted in reverse orientation) contigs between AL358394 and AL589794 (complete clone list in Supplementary Fig. S1). From left to right the

members (where parentheses indicate members only appearing in the enlarged section) of each family are: GLUD family, *GLUDP5*, (*GLUDP7*, *GLUDP8*, *GLUDP6*, *GLUDP2*), *GLUDP3*, *GLUD1*; CTSLL family, (*CTSLL5*, *CTSLL7*, *CTSLL2*, *CTSLL3*, *CTSLL4*), *CTSLL6*, *CTSLL1*; ANXA family, (*ANXA8L1*, *ANXA8L2*, *ANXA8*), *ANXA2P3*, *ANXA7*, *ANXA11*; CTGLF family, (*CTGLF10P*, *CTGLF1*, *CTGLF13P*, *CTGLF7*, *CTGLF11P*, *CTGLF6*, *CTGLF9P*, *CTGLF12P*, *CTGLF5*, *CTGLF4*, *CTGLF3*), *CTGLF2*; BMS1L family, *BMS1L*, (*BMSILP1*, *BMSILP2*, *BMSILP6*, *BMSILP5*, *BMSILP7*), *BMSILP4*, *BMSILP3*; FAM25 family, (*FAM25E*, *FAM25B*, *FAM25HP*, *FAM25G*, *FAM25C*, *FAM25D*), *FAM25A*. *ANXA7* and *ANXA11* were included owing to their proximity to the 10q22 and 10q23.1 locus, respectively. Seven *CTGLF1* paralogues (Supplementary Table S3) were annotated as novel genes for consistency but probably represent expressed pseudogenes.

Table 2 Structural characteristics of annotated gene structures on chromosome 10

Gene structure class	Number of genes	Total transcribed length (bp)	Mean gene length (bp)	Mean exon length (bp)	Mean transcripts per gene	Mean exons per gene
Known genes	654	56,312,218	86,870	316	4.76	11.02
Novel CDS	162	3,615,155	22,435	344	2.12	4.96
Total protein coding	816	59,482,121	74,078	320	4.24	9.81
Novel transcript	219	6,123,588	29,166	272	1.77	4.06
Putative genes	322	2,798,239	8,780	287	1.12	2.15
Processed pseudogenes	371	328,131	887	638	1.00	1.19
Unprocessed	59	982,662	16,655	176	1.00	6.95
Total structures	1,787	67,353,917	39,717	322	2.59	5.84

Note that transcribed length is not additive because some genes overlap. CDS, coding sequence.

1,025 CpG islands (Fig. 1), which are known to be associated with the 5' end of an estimated 60% of human genes¹¹. We then used Eponine¹² to predict transcription start sites and FirstEF¹³ to predict regions that encompass the promoter and 5' exon. In total, FirstEF and Eponine predicted 1,801 (rank = 1, score ≥0.8) and 2,800 features, respectively (Supplementary Fig. S1). Notably, 62% of FirstEF and 96% of Eponine features directly overlapped CpG islands, suggesting a heavy bias towards this feature in both algorithms. The distribution of CpG islands, FirstEF and Eponine hits was also assessed relative to the first exon of each of the 4,635 annotated transcripts using a window of 5,000 bp upstream and 1,000 bp downstream of the exon. Table 3 summarizes the results obtained per gene class. For example, in the 'known' class, 1,544 (49.6%) and 1,124 (36.1%) transcripts had a FirstEF and Eponine feature, respectively. Not surprisingly, 89.9% of FirstEF and 96.7% of Eponine transcripts were also associated with a CpG island. Note that Eponine predicts multiple transcription start sites per transcript (4.24 on average).

Regulation of gene expression by antisense transcription is a recognized mechanism with examples reported in species ranging from bacteria to mammals^{14–16}. We observed widespread occurrence of overlapping coding genes (either strand) in human chromosome 10 (101 pairs in total). In 38 cases one of the genes is fully contained within an intron of the other, typically on the opposite strand (34). For example, the second intron of the splice variant LIPA-004 encompasses four members of the IFIT gene cluster (Supplementary Table S3 and Fig. S2) and appears to be transcribed from the same bidirectional promoter as *IFIT5* (IFIT cluster member). Interestingly, the LIPA-001 and -002 variants do not overlap any IFIT gene, whereas LIPA-003 and LIPA-005 both overlap with *IFIT2* and *IFIT4* (Supplementary Fig. S2). Mutations in *LIPA* can cause Wolman and cholesteryl ester storage disease (Online Mendelian Inheritance in Man (OMIM) 278000). Among partially overlapping pairs (opposite strands), 34% involve the respective 5' exons, which is indicative in each case of a bidirectional promoter. We also searched for non-coding transcripts located on the opposite strand of coding genes. There are 67 antisense transcripts overlapping 63 coding genes. The two most common patterns observed were intragenic with partial overlap of one exon, and partial overlap with the most 5' exon of the coding gene. For *ZNF32*, we found two antisense transcripts (*ZNF32OS1* and *ZNF32OS2*).

Finally, we looked at the distribution of known protein domains

in both human chromosome 10 (this study) and the whole genome (Ensembl v.17.33.1) proteome using InterProScan. At the gene level, 70.6% of the human chromosome 10 peptides have at least one InterPro match with a Pfam domain and 32% are multidomain (1.37 distinct InterPros on average). Supplementary Table S2 shows the top 24 domains in chromosome 10 alongside their genome-wide ranking, suggesting that this chromosome is enriched in peptides with a lipase (IPR000734), aldo/keto reductase (IPR001395) or alpha/beta hydrolase (IPR000073) domain. BLASTP analysis (*e*-values <10⁻¹⁵) showed that all six genes encoding the peptides carrying the IPR001395 domain are clustered (AKRIC cluster) at position 4.8–5.3 Mb, whereas there are two lipase clusters, LIP (90.0–91.0 Mb; six members) and PNLIP (117.9–118.1 Mb; four members). In total, we found 42 gene clusters along human chromosome 10 (Supplementary Table S3).

Genomic landscape

The average G+C and repeat content of human chromosome 10 are 41.58% and 43.66%, respectively. The distribution of the main classes of repeats (Supplementary Table S4), the G+C and CpG density plots all seem to follow the known genome-wide trends. For example, the G+C content fluctuates along the chromosome, peaks at the qtel and is positively correlated to gene density (Supplementary Fig. S1). Large genes tend to be located adjacent to or within gene-poor regions, for example *PRKG1* and *PCDH15* (interval 52.0–59.0 Mb), whereas regions of high gene density seem enriched in short interspersed elements (SINES).

Segmental duplications are an important feature of the genomic landscape, being an integral part of the evolutionary machinery. Figure 2a illustrates the interchromosomal duplications along human chromosome 10 (see also ref. 17 and <http://humanparalogy.cwru.edu/SDD>). Figure 2b shows the extensive segmental duplications within a 5 Mb region at 10q11 with sequences further dispersed at 10q22, 10q23.1 and 10q23.2. Using the draft genome sequence Crosier and colleagues¹⁸ reported that 10q11 has been subject to multiple rounds of local duplications in at least the last 30–40 million years (Myr); they also showed that a 10q11:10q23 paracentric inversion occurred after the divergence of orang-utan from other great apes and hypothesized that the 10q22 locus resulted from chromosome-specific duplicative transposition. Bryce and colleagues¹⁹ characterized three cathepsin-L-like paralogues, which are expressed pseudogenes, and reported FISH signals

Table 3 Correlation of CpG island, FirstEF and Eponine features with annotated transcripts

Gene class	All Transcripts		FirstEF transcripts		Eponine transcripts	
	Complete	Incomplete	Complete	Incomplete	Complete	Incomplete
Known	1,490 (580)	1,623 (1,018)	593 (531)	951 (877)	421 (412)	703 (682)
Novel CDS	130 (57)	213 (75)	45 (43)	62 (57)	38 (38)	43 (41)
Novel transcript	388 (119)	0 (0)	120 (105)	0 (0)	83 (79)	0 (0)
Processed pseudogene	371 (32)	0 (0)	29 (22)	0 (0)	21 (19)	0 (0)
Unprocessed pseudogene	59 (11)	1 (2)	14 (10)	1 (1)	9 (8)	0 (0)
Putative	360 (74)	0 (0)	87 (64)	0 (0)	57 (50)	0 (0)

Values in parentheses are for transcripts associated with a CpG island.

at 10q11, 10q22 and 10q23, a pattern previously seen with the *GLUD* paralogues²⁰ (Fig. 2b). The duplication of the *CTSL* locus between chromosomes 9 and 10 was estimated to have occurred some 40 Myr ago¹⁹. We identified four and three additional *CTSL* and *GLUD* pseudogenes, respectively (Fig. 2b), consistent with the local duplication events involving *BMS1L* (known as *KIAA0187*)¹⁸ and *CTGLF1* (this study). In total, we annotated 7 *BMS1L* and 11 *CTGLF1* paralogues (Fig. 2b; functional genes in dark blue). Retro-position of a truncated *KIAA1099* (*CENTG2*; chromosome 2) mRNA gave rise to a processed pseudogene on chromosome 10 (ref. 18). However, this pseudogene forms the 3' exon of *CTGLF1*, suggesting that this gene resulted from a fusion between an ancestral gene and the *CENTG2* pseudogene, and the retroposition event predated all segmental duplications. Note that there is also evidence of transcripts combining exons of *CTGLF1* and *BMS1L* paralogues. Interestingly, we predicted a novel gene, *FAM25A* (based on mouse complementary DNA AK008614 and without similarity to any known protein), with seven human chromosome 10 paralogues (Fig. 2b) that follow the pattern of *GLUD*, *BMS1L*, *CTGLF* and *CTSL*. In addition to these five types of low copy repeats, the segmental duplications in 10q11:10q22:10q23 seem to have impacted on the number of genes on this chromosome. Notably, 31% of all the functionally related gene clusters are located within 10q11 and 10q23 (Supplementary Fig. S1 and Table S3).

The average recombination rate across the chromosome is 1.32 cM Mb⁻¹ (Fig. 3, sex average). Note that ref. 7 used the draft human chromosome 10 sequence (inflated by ~8%) and thus obtained a lower figure (1.21 cM Mb⁻¹). The rate of male recombination is higher than the female rate near the telomeres, whereas between D10S211 and D10S575 the female rate is higher (Fig. 3). This comparison also indicates the presence of two female-specific recombination hotspots (Fig. 3, arrows). As expected, pericentromeric regions display a low rate of recombination, more than twofold below the chromosome average. In particular, the region between D10S1247 and D10S1783 has a rate of 0.3 cM Mb⁻¹ and contains the only human chromosome 10 recombination 'desert'²¹. The region of extensive segmental duplications at 10q11 shows a low

rate of recombination (0.72 cM Mb⁻¹). Thus, meiotic recombination is unlikely to have been the driving force in generating these duplications. Our analysis confirmed the recombination 'jungle' between D10S1782–D10S1651, which we extended to D10S212 (3.4 cM Mb⁻¹); we refuted the one between D10S189 and D10S1728 (2.3 cM Mb⁻¹), and identified a new one between D10S1154 and D10S552 (4.55 cM Mb⁻¹).

Comparative sequence analysis

Many of the functional units in present-day vertebrate genomes have been conserved through evolution. Coding regions are highly conserved across all vertebrates, whereas non-coding regions are conserved among more closely related species. The genome sequences of three fishes (*Tetraodon*, *Fugu* and zebrafish) and two rodents (mouse and rat) were publicly available at the time of analysis. We compared the sequence of human chromosome 10 to each of the above species and searched for conserved regions with coding potential—distinguishing functional non-coding elements above background sequence conservation requires the use of additional genomes²². We then correlated the obtained hits in each species with the annotated genes. As expected, each of the fish genomes showed higher specificity (>0.82; that is, 82% of sequences conserved in fish overlapped human annotated exons) than the rodents (>0.29), whereas the highest specificity was obtained using all genomes together (0.96). Sensitivity was higher using a rodent genome (0.7; that is, 70% of all annotated exons had matches in rodent) than a fish genome (0.4).

Of the 1,787 gene structures (16,765 exons), 84% (78% of exons) had at least one exon supported by a conserved region in one of the other genomes and 52% (32% of exons) in all genomes. Note that the figures given for exons are underestimated owing to lower sequence conservation in the untranslated regions. Protein-coding genes are highly conserved (98.5%; 85% of exons). In contrast, 61% (29% of exons) and 45% (27% of exons) of novel transcripts and putative genes, respectively, have at least one match. Furthermore, only 21% of novel transcripts and 8% of putative genes show conservation with a fish (91% for protein-coding genes). Typically, novel transcripts were annotated on the basis of solid experimental evidence (that is, human mRNA) and may represent either genes that have evolved more rapidly or non-coding RNAs.

On the basis of specificity, regions conserved in all six species can serve as a measure of completeness of the gene annotation process that occurred independently of the comparative analysis. We found 5,604 such evolutionarily conserved regions (ECRs) of which 5,292 mapped inside annotated exons (including pseudogenes). Of the remaining 312 ECRs, 142 were intergenic and 170 intragenic. On inspection, we found 79 of these ECRs with supportive evidence to annotate a missed exon, most of which were part of a pseudogene (79%). The remaining 233 ECRs provide the basis to estimate that we have annotated at least 95.8% of all conserved coding exons on human chromosome 10. This is a conservative estimate as 131 of these ECRs are located in introns and may represent conserved non-exonic sequences. Interestingly, 54 (41%) of them are associated with just four genes: *C10orf11* (26; also known as *CDA017*), *EBF3* (20), *TCF7L2* (5) and *PAX2* (3). All but *C10orf11* (unknown function) are transcription factors. Figure 4 shows a MultiPip-Maker²³ alignment of the orthologous *EBF3* loci and the relative position of ECRs in the human gene. Note that sequence identity is often higher in ECRs than in exons.

Sequence variation

During the human chromosome 10 project we discovered 35,882 single nucleotide polymorphisms (SNPs) by sequence alignment in regions of clone overlaps. In total, we mapped 143,364 SNPs (dbSNP release 115) to the chromosome 10 sequence. Supplementary Fig. S1 shows the density plots for randomly discovered²⁴ and all SNPs across the chromosome.

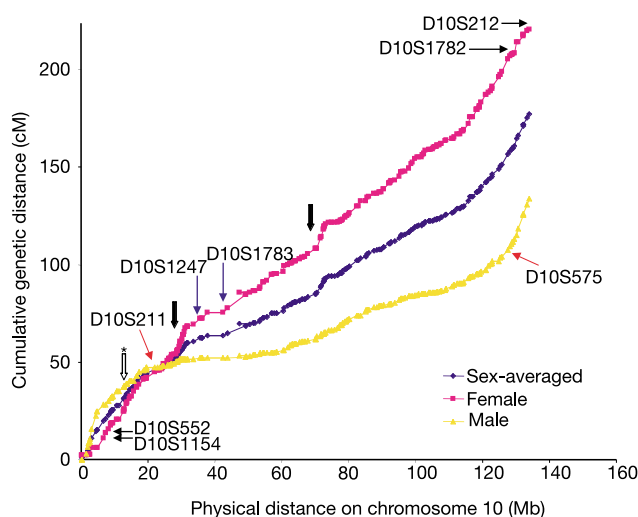


Figure 3 Alignment of the deCODE genetic map of chromosome 10 to the physical map from the telomeric end of the short arm to the telomeric end of the long arm. The position of each genetic marker on the female, male and sex-averaged genetic map is indicated. Female-specific recombination hotspots are indicated by thick arrows (left, D10S1732–D10S208; right, D10S599–D10S676). The location of markers flanking recombination deserts (blue arrows) and jungles (black arrows) is shown. The asterisk indicates the location of the refuted jungle (D10S189–D10S1728).

There are 5,864 (4.1%) exonic and 65,973 (46%) intronic SNPs. Of the 1,821 SNPs in coding exons 984 are non-synonymous. *MSMB* has the most polymorphic coding region with 43 SNPs kb⁻¹; it encodes a protein with inhibin-like activity and its expression is decreased in prostate cancer²⁵.

We also considered 729,553 human–chimpanzee single base differences (SBDs) remapped on the current assembly of human chromosome 10. These were high-confidence sequence differences originally identified by aligning 14 million shotgun reads of the chimpanzee genome, generated jointly by the Whitehead Institute and Washington University Genome Centers, to the human genome sequence assembly (NCBI build 31). We first removed all human–chimpanzee SBDs that co-localized with known human SNPs. Supplementary Fig. S1 shows the density plot of the remaining 703,338 SBDs. Of those, 55.3% are intergenic, 42.9% intronic and 1.8% exonic. The highest density of human–chimpanzee SBDs, fourfold greater than the average level, was observed in a 200-kb gene-poor region at 19.43–19.63 Mb. We then examined the 12,710 human–chimpanzee SBDs that lie in exons of the 816 human coding genes. Of those, 3,972 were in coding regions and can be subdivided further into 2,273 synonymous, 1,678 non-synonymous and 21 nonsense with respect to the human sequence. For each gene we calculated the rate of evolution of non-synonymous (K_a) and synonymous (K_s) changes, and the ratio K_a/K_s , which provides a measure of evolutionary selection. Supplementary Table S5 lists the 1,413 transcripts with at least one coding human–chimpanzee SBD sorted on the K_a/K_s value. There are only 29 transcripts (21 genes) that have a K_a/K_s value ≥ 1 , whereas there are 484 without non-synonymous SBDs. Note that several caveats apply in this type of analysis owing to the incomplete nature of both the chimpanzee data and the list of human SNPs; we used the number of intronic human–chimpanzee SBDs per base in comparison to the chromosome average of 0.005 as a possible estimate of coverage. The gene with most non-synonymous human–chimpanzee SBDs is *MKI67*, an antigen identified by monoclonal antibody Ki-67, which appears to be fast evolving in humans ($K_a/K_s = 1.038507$; SNP data). The expression pattern of *MKI67* in gastric and other cancers is under investigation as this gene is expressed in proliferating cells. Interestingly, a nonsense human–chimpanzee SBD is present in both of its coding transcripts. Among the 21 genes with nonsense human–chimpanzee SBDs notable examples are the serotonin receptor *HTR7* (the neurotransmitter serotonin is thought to be involved

in cognition and behaviour), *PSAP* (prosaposin; involved in variant Gaucher's disease and metachromatic leukodystrophy) and the developmental gene *NODAL*.

Medical implications

At the time of writing there were 85 disease loci reported on human chromosome 10 (<http://www.ncbi.nlm.nih.gov/omim/>), a 47% increase since 1999 (ref. 26). Several of these loci account for multiple disease phenotypes caused by mutations in the same gene; notable examples are *FGFR2* (OMIM 176943), *PTEN* (OMIM 601728) and the proto-oncogene *RET* (OMIM 164761). Since *PTEN* was first shown to be mutated in brain, breast and prostate cancers²⁷, there has been an explosion of reported mutations (110 germline and 332 somatic mutations)²⁸ and disease phenotypes²⁹. Human chromosome 10 harbours several other genes involved in tumorigenesis; for example, deregulation of *TLX1*, *NFKB2* or *BMI1* caused by chromosomal translocations or amplifications has been detected in lymphatic neoplasms. Mapping of allelic imbalances and functional studies suggest the presence of additional tumour-related genes. The finished and annotated sequence is key in the process of cloning these and other hitherto unidentified disease-associated genes.

The prompt release of both the clone and sequence map resources throughout the project has accelerated the cloning of many disease-causing genes. To this end we recently showed as part of the European ADLTE consortium that mutations in the *LGII* gene cause autosomal dominant lateral temporal epilepsy³⁰. Notably, we found that the *FRA10A* folate-sensitive fragile site is located close to *LGII* and its expression is associated with the expansion of a polymorphic CGG repeat located at the 5' UTR of *FRA10AC1*, a gene encoding a novel nuclear protein³¹. There are seven fragile sites mapping to human chromosome 10 (ref. 32).

The challenge ahead is to unravel the molecular basis of common disease. An increasing number of susceptibility loci for complex diseases is being mapped to human chromosome 10, including metabolic diseases such as type I diabetes (IDDM10 and IDDM17), psychiatric disorders such as schizophrenia, or neurodegenerative illnesses such as Alzheimer's disease^{26,32}. In a case control study of morbidly obese and healthy individuals Boutin and colleagues³³ identified a SNP in the *GAD2* gene that increases the risk for obesity as well as a protective haplotype. Studies so far have mainly focused on candidate genes. The construction of a comprehensive haplotype

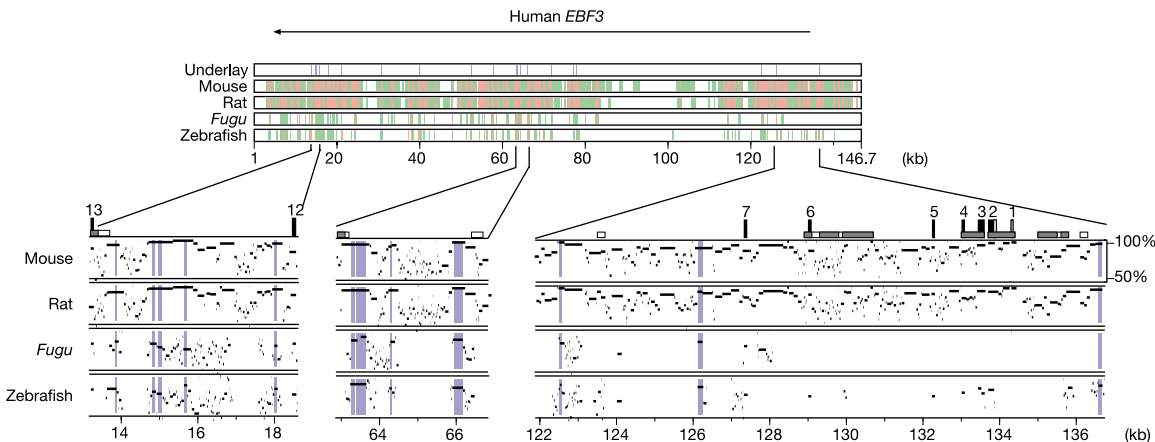


Figure 4 Multispecies alignment of orthologous *EBF3* loci. The human early B-cell factor 3 (*EBF3*) gene is represented by the arrow at the top. Alignments are displayed using MultiPipMaker²³. In the top panel, the first track shows the location of the ECRs (blue lines) across the human locus, whereas the following four tracks show regions conserved in mouse, rat, *Fugu* and zebrafish, respectively (green, aligned regions; orange, aligned

regions with at least 70% nucleotide identity and no gap over 100 bp). The bottom panel shows a detailed view of the three regions with the highest number of ECRs. Vertical black and grey numbered boxes represent coding and UTR exons, respectively. The scale at the right indicates the percentage of sequence identity. Physical distance is given in kilobases (kb).

map of the human genome is well underway³⁴, making it possible to undertake a systematic approach to scanning the genome for associations to disease-related and other phenotypes. □

Methods

The mapping and sequencing methods used in the assembly of the bacterial clone and sequence map of chromosome 10, respectively, as well as the tools of the gene annotation pipeline are described in refs 1 and 35 (see ref. 36 for detailed protocols). Manual annotation of gene structures followed the guidelines agreed in the human annotation workshop (HAWK; <http://www.sanger.ac.uk/HGP/havana/hawk.shtml>), whereas gene symbols were approved where possible by the HUGO Gene Nomenclature Committee³⁷. Protein translations were analysed with InterProScan (<http://www.ebi.ac.uk/interpro/scan.html>), which was run via the Ensembl protein annotation pipeline, to obtain Pfam, Prosite, Prints and Profiles domain matches.

Alignments for inter- and intrachromosomal duplications were performed with WU-BLASTN (<http://blast.wustl.edu>) using the current sequence assembly of chromosome 10 and the NCBI34 build for the rest of the genome. All sequences were repeat masked with RepeatMasker (<http://repeatmasker.genome.washington.edu>) and low-quality alignments (e -value $> 10^{-30}$, sequence identity $< 90\%$, length < 80 bp) were discarded. For intrachromosomal duplications, self matches were discarded. For interchromosomal duplications, the sequence was split into 400-kb segments. Adjacent matches in the same orientation were joined together as described by ref. 38. Only blocks of 10 kb or greater were retained.

The following sequence assemblies were used for comparative analysis: *M. musculus* NCBI build 30 (Mouse Genome Sequencing Consortium; http://www.ensembl.org/Mus_musculus/resources.html); *R. norvegicus* version 2.0 (Rat Genome Sequencing Consortium; <http://www.hgsc.bcm.tmc.edu/projects/rat>); *D. rerio* Assembly version 1 (Sanger Institute; http://www.sanger.ac.uk/Projects/D_rerio); *F. rubripes* version 2 (International Fugu Genome Consortium; <http://www.fugu-sg.org/project/info.html>); and *T. nigroviridis* version 6 (Genoscope and Whitehead Institute for Genome Research; <http://www.genoscope.cns.fr/externe/tetraodon/Ressource.html>). The repeat-masked sequence of chromosome 10 was aligned to the mouse and rat genome sequences using BLASTZ³⁹ and the resulting matches were post-processed by axtBest and subsetAxt (W. J. Kent; <http://www.soe.ucsc.edu/~kent/src>) as described elsewhere³⁵. Alignment of the three fish genome sequences to chromosome 10 was performed with WU-TBLASTX using the scoring matrix, parameters and filtering strategy applied by Exofish⁴⁰. Overlapping alignments to different sequences were merged to produce contiguous regions of sequence conservation, analogous to the ECRs or 'Ecores', reported by Exofish.

SNPs in sequence overlaps were identified using a modification of the SSAHA software⁴¹. The chromosome 10 SNPs (dbSNP release 115) were mapped to the sequence assembly of this chromosome (this study) first with SSAHA and then Cross-Match. Of the approximately 14 million chimpanzee reads (<http://www.genome.gov/11008056>) mapped onto the human sequence assembly (NCBI build 31), those mapping to chromosome 10 were remapped to our sequence assembly and used to identify human-chimpanzee SBDs.

Received 17 December 2003; accepted 9 March 2004; doi:10.1038/nature02462.

1. Bentley, D. R. *et al.* The physical maps for sequencing human chromosomes 1, 6, 9, 10, 13, 20 and X. *Nature* **409**, 942–943 (2001).
2. Guy, J. *et al.* Genomic sequence and transcriptional profile of the boundary between pericentromeric satellites and genes on human chromosome arm 10p. *Genome Res.* **13**, 159–172 (2003).
3. Guy, J. *et al.* Genomic sequence and transcriptional profile of the boundary between pericentromeric satellites and genes on human chromosome arm 10q. *Hum. Mol. Genet.* **9**, 2029–2042 (2000).
4. Jackson, M. S., See, C. G., Mulligan, L. M. & Lauffart, B. F. A 9.75-Mb map across the centromere of human chromosome 10. *Genomics* **33**, 258–270 (1996).
5. Felsenfeld, A., Peterson, J., Schloss, J. & Guyer, M. Assessing the quality of the DNA sequence from the Human Genome Project. *Genome Res.* **9**, 1–4 (1999).
6. Deloukas, P. *et al.* A physical map of 30,000 human genes. *Science* **282**, 744–746 (1998).
7. Kong, A. *et al.* A high-resolution recombination map of the human genome. *Nature Genet.* **31**, 241–247 (2002).
8. Broman, K. W., Murray, J. C., Sheffield, V. C., White, R. L. & Weber, J. L. Comprehensive human genetic maps: individual and sex-specific variation in recombination. *Am. J. Hum. Genet.* **63**, 861–869 (1998).
9. Deloukas, P. *et al.* The DNA sequence and comparative analysis of human chromosome 20. *Nature* **414**, 865–871 (2001).
10. International Human Genome Sequencing Consortium, Initial sequencing and analysis of the human genome. *Nature* **409**, 860–921 (2001).
11. Antequera, F. & Bird, A. Number of CpG islands and genes in human and mouse. *Proc. Natl Acad. Sci. USA* **90**, 11995–11999 (1993).
12. Down, T. A. & Hubbard, T. J. Computational detection and location of transcription start sites in mammalian genomic DNA. *Genome Res.* **12**, 458–461 (2002).
13. Davuluri, R. V., Grosse, I. & Zhang, M. Q. Computational identification of promoters and first exons in the human genome. *Nature Genet.* **29**, 412–417 (2001).

14. Wagner, E. G. & Simons, R. W. Antisense RNA control in bacteria, phages, and plasmids. *Annu. Rev. Microbiol.* **48**, 713–742 (1994).
15. Eddy, S. R. Non-coding RNA genes and the modern RNA world. *Nature Rev. Genet.* **2**, 919–929 (2001).
16. Kiyosawa, H. *et al.* Antisense transcripts with FANTOM2 clone set and their implications for gene regulation. *Genome Res.* **13**, 1324–1334 (2003).
17. Bailey, J. A. *et al.* Recent segmental duplications in the human genome. *Science* **297**, 1003–1007 (2002).
18. Crosier, M. *et al.* Human paralogs of KIAA0187 were created through independent pericentromeric-directed and chromosome-specific duplication mechanisms. *Genome Res.* **12**, 67–80 (2002).
19. Bryce, S. D. *et al.* A novel family of cathepsin L-like (CTSLL) sequences on human chromosome 10q and related transcripts. *Genomics* **24**, 568–576 (1994).
20. Deloukas, P., Dauwerse, J. G., Moschonas, N. K., van Ommen, G. J. & van Loon, A. P. Three human glutamate dehydrogenase genes (GLUD1, GLUDP2, and GLUDP3) are located on chromosome 10q, but are not closely physically linked. *Genomics* **17**, 676–681 (1993).
21. Yu, A. *et al.* Comparison of human genetic and sequence-based physical maps. *Nature* **409**, 951–953 (2001).
22. Thomas, J. W. *et al.* Comparative analyses of multi-species sequences from targeted genomic regions. *Nature* **424**, 788–793 (2003).
23. Schwartz, S. *et al.* PipMaker—a web server for aligning two genomic DNA sequences. *Genome Res.* **10**, 577–586 (2000).
24. Sachidanandam, R. *et al.* A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. *Nature* **409**, 928–933 (2001).
25. Vanaja, D. K., Chevillat, J. C., Iturria, S. J. & Young, C. Y. Transcriptional silencing of zinc finger protein 185 identified by expression profiling is associated with prostate cancer progression. *Cancer Res.* **63**, 3877–3882 (2003).
26. Deloukas, P., French, L., Meitinger, T. & Moschonas, N. K. Report of the third international workshop on human chromosome 10 mapping and sequencing 1999. *Cytogenet. Cell Genet.* **90**, 1–12 (2000).
27. Li, J. *et al.* PTEN, a putative protein tyrosine phosphatase gene mutated in human brain, breast, and prostate cancer. *Science* **275**, 1943–1946 (1997).
28. Bonneau, D. & Longy, M. Mutations of the human PTEN gene. *Hum. Mutat.* **16**, 109–122 (2000).
29. Eng, C. PTEN: one gene, many syndromes. *Hum. Mutat.* **22**, 183–198 (2003).
30. Morante-Redolat, J. M. *et al.* Mutations in the LGI1/Epitempin gene on 10q24 cause autosomal dominant lateral temporal epilepsy. *Hum. Mol. Genet.* **11**, 1119–1128 (2002).
31. Sarafidou, T. *et al.* Folate-sensitive fragile site *FRA10A* is due to an expansion of a CGG-repeat in a novel gene *FRA10AC1*, encoding a nuclear protein. *Genomics* (in the press).
32. Moschonas, N. K. in *Encyclopedia of the Human Genome* Vol. 1, 618–625 (Nature Publishing Group, London, 2003).
33. Boutin, P. *et al.* *GAD2* on chromosome 10p12 is a candidate gene for human obesity. *PLoS Biol.* **1**, 1–11 (2003).
34. The International HapMap Consortium. The International HapMap project. *Nature* **426**, 789–796 (2003).
35. Mungall, A. J. *et al.* The DNA sequence and analysis of human chromosome 6. *Nature* **425**, 805–811 (2003).
36. Dunham, I. (ed.) *Genome Mapping and Sequencing* (Horizon Scientific, Wymondham, UK).
37. Wain, H. M., Lush, M., Ducluzeau, F. & Povey, S. Genew: the human gene nomenclature database. *Nucleic Acids Res.* **30**, 169–171 (2002).
38. Cheung, J. *et al.* Genome-wide detection of segmental duplications and potential assembly errors in the human genome sequence. *Genome Biol.* **4**, R25 (2003).
39. Schwartz, S. *et al.* Human–mouse alignments with BLASTZ. *Genome Res.* **13**, 103–107 (2003).
40. Roest Crollius, H. *et al.* Estimate of human gene number provided by genome-wide analysis using *Tetraodon nigroviridis* DNA sequence. *Nature Genet.* **25**, 235–238 (2000).
41. Ning, Z., Cox, A. J. & Mullikin, J. C. SSAHA: a fast search method for large DNA databases. *Genome Res.* **11**, 1725–1729 (2001).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank the *Fugu*, Mouse and Rat Genome Sequencing Consortia; Genoscope and Whitehead Institute for Genome Research (*Tetraodon* genome data); the members of the zebrafish project at the Sanger Institute; the groups at the Washington University Genome Sequencing Center and the Whitehead Institute for early access to the chimpanzee whole-genome shotgun data; the laboratories of H. Riethman and M. S. Jackson for the provision of telomeric and pericentromeric clones, respectively; H. M. Wain, E. A. Bruford, C. C. Talbot, V. K. Khodiyar, M. J. Lush, M. W. Wright and S. Povey for assigning official gene symbols; the EMBL, GenBank and Ensembl database groups; I. Dunham for critical reading of the manuscript; and the Wellcome Trust for financial support.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to P.D. (panos@sanger.ac.uk). All DNA sequences reported in this study have been deposited in the EMBL or GenBank databases and the accession numbers can be found in Supplementary Fig. S1.

The physical maps for sequencing human chromosomes 1, 6, 9, 10, 13, 20 and X

D. R. Bentley*, **P. Deloukas***, **A. Dunham***, **L. French***, **S. G. Gregory***, **S. J. Humphray***, **A. J. Mungall***, **M. T. Ross***, **N. P. Carter***, **I. Dunham***, **C. E. Scott***, **K. J. Ashcroft***, **A. L. Atkinson***, **K. Aubin***, **D. M. Beare***, **G. Bethel***, **N. Brady***, **J. C. Brook***, **D. C. Burford***, **W. D. Burrill***, **C. Burrows***, **A. P. Butler***, **C. Carder***, **J. J. Catanese†**, **C. M. Clee***, **S. M. Clegg***, **V. Cobley***, **A. J. Coffey***, **C. G. Cole***, **J. E. Collins***, **J. S. Conquer***, **R. A. Cooper***, **K. M. Culley***, **E. Dawson***, **F. L. Dearden***, **R. M. Durbin***, **P. J. de Jong‡**, **P. D. Dhami***, **M. E. Earthrowl***, **C. A. Edwards***, **R. S. Evans***, **C. J. Gillson***, **J. Ghor†**, **L. Green***, **R. Gwilliam***, **K. S. Halls***, **S. Hammond***, **G. L. Harper***, **R. W. Heathcott***, **J. L. Holden***, **E. Holloway***, **B. L. Hopkins***, **P. J. Howard***, **G. R. Howell***, **E. J. Huckler***, **J. Hughes***, **P. J. Hunt***, **S. E. Hunt***, **M. Izmajlowicz***, **C. A. Jones***, **S. S. Joseph***, **G. Laird***, **C. F. Langford***, **M. H. Lehvaslaiho***, **M. A. Leversha***, **O. T. McCann***, **L. M. McDonald***, **J. McDowall***, **G. L. Masien‡**, **D. Mistry***, **N. K. Moschonas‡**, **V. Neocleous§**, **D. M. Pearson***, **K. J. Phillips***, **K. M. Porter***, **S. R. Prathalingam***, **Y. H. Ramsey***, **S. A. Ranby***, **C. M. Rice***, **J. Rogers***, **L. J. Rogers***, **T. Sarafidou‡**, **D. J. Scott***, **G. J. Sharp***, **C. J. Shaw-Smith***, **L. J. Smink***, **C. Soderlund***, **E. C. Sotheran***, **H. E. Steingruber***, **J. E. Sulston***, **A. Taylor***, **R. G. Taylor***, **A. A. Thorpe***, **E. Tinsley***, **G. L. Warry***, **A. Whittaker***, **P. Whittaker***, **S. H. Williams***, **T. E. Wilmer***, **R. Wooster*** & **C. L. Wright***

* The Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA, UK

‡ Department of Biology, University of Crete and Institute of Molecular Biology and Biotechnology, PO Box 2208, 71409 Heraklion, Crete, Greece

§ Neurogenetic Laboratory, The Cyprus Institute of Neurology and Genetics, 6, International Airport Avenue, PO Box 23462, 1683 Nicosia, Cyprus

† Children's Hospital-BACPAC Resources, 747 52nd Street, Oakland, California 94609, USA

We constructed maps for eight chromosomes (1, 6, 9, 10, 13, 20, X and (previously) 22), representing one-third of the genome, by building landmark maps, isolating bacterial clones and assembling contigs. By this approach, we could establish the long-range organization of the maps early in the project, and all contig extension, gap closure and problem-solving was simplified by containment within local regions. The maps currently represent more than 94% of the euchromatic (gene-containing) regions of these chromosomes in 176 contigs, and contain 96% of the chromosome-specific markers in the human gene map. By measuring the remaining gaps, we can assess chromosome length and coverage in sequenced clones.

The task of sequencing the 3,200 megabase (Mb) human genome

can be subdivided into individual chromosome projects ranging in size from 263 Mb (chromosome 1)¹ to about 35 Mb (chromosomes 21q and 22q)^{2,3}. Our strategy, in common with other groups⁴⁻⁶, was to map selected chromosomes individually, and then to combine the results with those of whole-genome mapping studies into a single map of the human genome⁷. Chromosome maps were constructed as follows (see Supplementary Information). First, we constructed a landmark map for each chromosome. Second, we identified bacterial clones (bacterial- or P1-derived artificial chromosomes (BACs or PACs)) from genomic libraries using the chromosome-specific landmarks, and assembled them into contigs on the basis of shared restriction enzyme fingerprints and landmark content. Third, contigs were extended and joined by chromosome walking. Walking was carried out using BAC end sequences generated in house from clones selected at the ends of contigs, or from the publicly available resources; joins were also made by identification of overlaps using genomic sequence data.

Clones that were representative of each contig were selected regularly for inclusion in the 'tiling path' (a set of minimally overlapping clones) for genomic sequencing⁸. Clones from the tiling path of each chromosome were deposited in the 'HumanMap' database at the Genome Sequencing Centre⁷ (<http://genome.wustl.edu/gsc/human/mapping/>), for integration with the data obtained by whole-genome fingerprinting. New clones were identified from this integrated dataset to assist the extension and closure of the chromosome maps (Table 1). A detailed description of tiling paths and the underlying clone contigs is available as Supplementary Information and will continue to be updated at www.sanger.ac.uk; all clones are publicly available.

A key question is the extent of coverage of each chromosome in the map. We analysed the coverage of the euchromatic regions on the basis of the assumptions given below and the approximate estimates that were available for chromosome length. Heterochromatic regions, which are estimated to comprise 3-15% of each chromosome (Table 1), are absent from the contigs analysed here. On the basis of the fingerprint bands in the maps (converted to Mb as described in Supplementary Information), 176 contigs represent 927 Mb, or 94% of the estimated total of 981 Mb euchromatin. Half of this (485 Mb) is in fifteen contigs of 22-62 Mb, that illustrates the extent of continuity obtained. We also analysed the representation of human gene markers in the map. The clone map contained 96% of the unique markers that were previously mapped to these chromosomes in GeneMap99 (ref. 9), and 90% were present in the genome sequence. As expected, this figure is higher for individual chromosomes 20 (99%) and 22q (98%), which are nearly or completely finished (Table 1).

Independent corroboration of the coverage of each chromosome required identification of the boundaries between euchromatic sequence and the centromeric, telomeric, and other heterochromatic repeat sequences, as well as measurement of the remaining gaps in the map. We have measured 52 of the 57 remaining gaps in

Table 1 Status of chromosome maps

Chromosome	Estimated length		Physical map			Gene map		
	Total (Mb)	Euchromatin (Mb)	Contigs (n)	Euchromatin covered (Mb)	(%)	Markers analysed (n)	Markers in contigs (%)	Markers in sequence (%)
1	263	238	78	226	95	4,769	94	87
6	183	178	13	170	96	2,942	96	92
9	145	123	21	113	92	1,771	98	90
10	144	139	14	124	89	2,022	99	91
13	114	98	7	102	105	1,028	99	90
20	72	67	6	61	91	1,159	99	99
22(q12-tel)	25.5	25.5	8	24.8	97	777	98	98
X(p21-q27)	115	112	29	106	94	925	93	88
Total	1,062	981	176	927	94	15,393	96	90

See methods in Supplementary Information for details.

the maps of chromosomes 6, 9, 10, 13 and 20 by fluorescent *in situ* hybridization (FISH), in addition to the gaps previously measured on chromosome 22. Clones immediately flanking each gap were hybridized to extended DNA fibres, interphase nuclei or metaphase chromosome spreads.

An example of this analysis is given for chromosome 10 (Fig. 1). Fourteen contigs account for 124.4 Mb of DNA. They contain pericentromeric satellite sequences, and also a subtelomeric sequence on the short arm which is <0.1 Mb from the telomere; the most distal clone on the long arm is <0.15 Mb from the telomere¹⁰. From FISH analysis, the total extent of euchromatic gaps is ≤ 4.2 Mb. From these measurements, the estimated coverage of euchromatin in the map can be revised from 89% (based on Morton's previous estimate; Table 1) to 96.7% (124.4 of 128.6 Mb). A 9.75-Mb restriction map that spans the centromere has been constructed for chromosome 10 (ref. 11). By anchoring this to the clone maps on both chromosome arms, we estimate that the size of the gap across the centromeric region is around 4.5 Mb. This analysis provides a new estimate of 133 Mb for the total length of the chromosome (Fig. 1), compared with the previous value of 144 Mb. In a similar analysis of chromosome 13, seven contigs span 102 Mb and the remaining gaps in the euchromatic region total a maximum of 5 Mb. This leads to a new, increased estimate of the size of the euchromatic region of 107 Mb, and the current coverage in the map is 95%.

Our strategy may form the basis for finishing the map and sequence of every human chromosome. Accurate measurement of all gap sizes assists our continued efforts to bridge the remaining gaps in clones, using other cloning systems (for example yeast artificial chromosomes (YACs), plasmids and bacteriophage

lambda), and characterization of the centromeric satellites and the remaining heterochromatic regions, including the short arms of the acrocentric chromosomes (13, 14, 15, 21 and 22), will enable us to determine the true physical extent of DNA in the human genome. □

Received 28 November; accepted 21 December 2000.

1. Morton, N. E. Parameters of the human genome. *Proc. Natl Acad. Sci. USA* **88**, 7474–7476 (1991).
2. The chromosome 21 mapping and sequencing consortium. The DNA sequence of human chromosome 21. *Nature* **405**, 311–319 (2000).
3. Dunham, I. *et al.* The DNA sequence of human chromosome 22. *Nature* **402**, 489–495 (1999).
4. Tilford, C. A. *et al.* A physical map of the human Y chromosome. *Nature* **409**, 943–945 (2001).
5. Montgomery, K. T. A high-resolution map of human chromosome 12. *Nature* **409**, 945–946 (2001).
6. Bruls, T. *et al.* A physical map of human chromosome 14. *Nature* **409**, 947–948 (2001).
7. The International Human Genome Mapping Consortium. A physical map of the human genome. *Nature* **409**, 934–941 (2001).
8. International Human Genome Sequencing Consortium. Initial sequencing and analysis of the human genome. *Nature* **409**, 860–921 (2001).
9. Deloukas, P. *et al.* A physical map of 30,000 human genes. *Science* **282**, 744–746 (1998).
10. Riethman, H. C. *et al.* Integration of telomere sequences with the draft human genome sequence. *Nature* **409**, 948–951 (2001).
11. Jackson, M. S., See, C. G., Mulligan, L. M. & Lauffart, B. F. A 9.75-Mb map across the centromere of human chromosome 10. *Genomics* **33**, 258–270 (1996).

Supplementary information is available on Nature's World-Wide Web site (<http://www.nature.com>) or as paper copy from the London editorial office of Nature.

Acknowledgements

We thank M. Sekhon, A. Chinwalla, J. McPherson and staff of the Genome Sequencing Centre, St. Louis for their assistance; the many collaborators who have contributed reagents and information to assist map construction on individual chromosomes; M. Jackson for helpful discussion; T. Cox, R. Pettett and the web team; and the Wellcome Trust for support.

Correspondence and requests for material should be addressed to D.R.B (e.mail: drb@sanger.ac.uk).

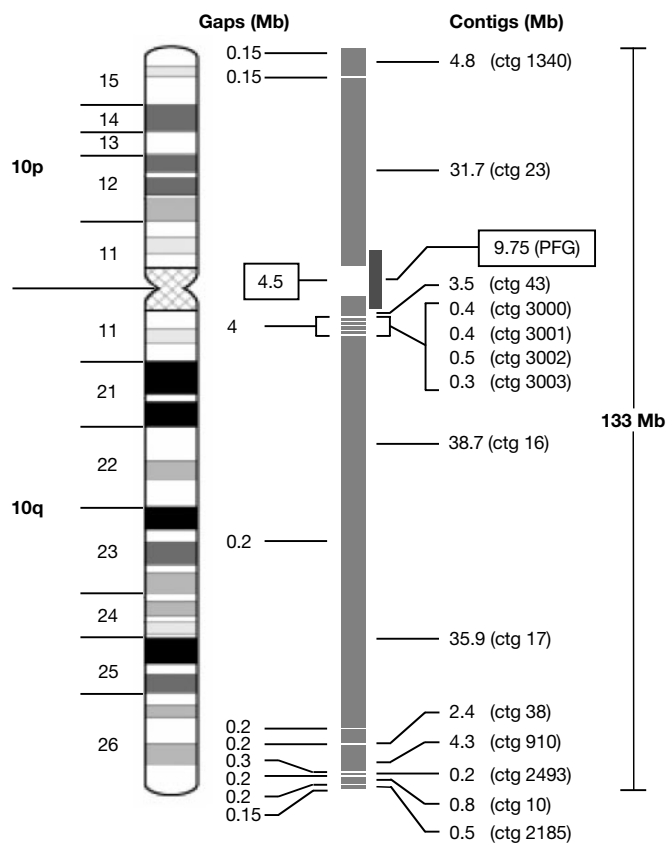


Figure 1 Physical map of chromosome 10. Clone contigs cover 124.4 Mb, euchromatic gaps cover 4.2 Mb, and the gap across centromeric satellites is 4.5 Mb¹¹. The separation between contigs 43 and 16 was determined as ~4 Mb on the basis of FISH of metaphase chromosome spreads, from which the sum of contigs 3000–3003 (1.6 Mb) was subtracted for a net gap coverage of 2.4 Mb in this region of 10q11.

A physical map of the human Y chromosome

Charles A. Tilford*, Tomoko Kuroda-Kawaguchi*, Helen Skaletsky*, Steve Rozen*, Laura G. Brown*, Michael Rosenberg*, John D. McPherson†, Kristine Wylie†, Mandeep Sekhon†, Tamara A. Kucaba†, Robert H. Waterston† & David C. Page*

* Howard Hughes Medical Institute, Whitehead Institute, and Department of Biology, Massachusetts Institute of Technology, 9 Cambridge Center, Cambridge, Massachusetts 02142, USA

† Genome Sequencing Center, Department of Genetics, Washington University School of Medicine, 4444 Forest Park Boulevard, St. Louis, Missouri 63108, USA

The non-recombining region of the human Y chromosome (NRY), which comprises 95% of the chromosome, does not undergo sexual recombination and is present only in males. An understanding of its biological functions has begun to emerge from DNA studies of individuals with partial Y chromosomes, coupled with molecular characterization of genes implicated in gonadal sex reversal, Turner syndrome, graft rejection and spermatogenic failure^{1,2}. But mapping strategies applied successfully elsewhere in the genome have faltered in the NRY, where there is no meiotic recombination map and intrachromosomal repetitive sequences are abundant³. Here we report a high-resolution physical map of the euchromatic, centromeric and heterochromatic regions of the NRY and its construction by unusual methods, including genomic clone subtraction⁴ and dissection of sequence family variants⁵. Of the map's 758 DNA markers, 136 have multiple locations in the NRY, reflecting its unusually repetitive sequence composition. The markers anchor 1,038 bacterial artificial chromosome clones, 199 of which form a tiling path for sequencing.

Mutations in the *LGI1/Epitempin* gene on 10q24 cause autosomal dominant lateral temporal epilepsy

José M. Morante-Redolat^{1†}, Ana Gorostidi-Pagola^{1,2†}, Salomé Piquer-Sirerol¹, Amets Sáenz², Juan J. Poza², Juan Galán³, Stefan Gesk⁴, Theologia Sarafidou⁵, Victor-F. Mautner⁶, Simona Binelli⁷, Eike Staub⁸, Bernd Hinzmann⁸, Lisa French⁹, Jean-F. Prud'homme¹⁰, Daniela Passarelli¹¹, Paolo Scannapieco¹², Carlo A. Tassinari¹³, Giuliano Avanzini⁸, José F. Martí-Massó², Lan Kluwe¹⁴, Panagiotis Deloukas⁹, Nicholas K. Moschonas⁵, Roberto Michelucci¹³, Reiner Siebert⁴, Carlo Nobile¹⁵, Jordi Pérez-Tur^{1,*} and Adolfo López de Munain²

¹Unitat de Genètica Molecular, Institut de Biomedicina de València–CSIC, València, Spain, ²Unidad Experimental and Servicio de Neurología, Hospital Donostia, Donostia, Spain, ³Servicio de Neurología, Hospital Valme, Sevilla, Spain, ⁴Institute of Human Genetics, University Hospital Kiel, Kiel, Germany, ⁵Institute of Molecular Biology and Biotechnology, Foundation of Research and Technology, Heraklion, Greece, ⁶Laboratory of Brain Tumor Biology, Department of Neurosurgery, University of Hamburg, University Hospital Eppendorf, Hamburg, Germany, ⁷Istituto Nazionale Neurologico 'Carlo Besta', Milan, Italy, ⁸MetaGen Gesellschaft für Genomforschung mbH, Berlin, Germany, ⁹The Wellcome Trust Sanger Institute, Hinxton, Cambridge, UK, ¹⁰Généthon, Evry, France, ¹¹Division of Neurology, Infermi Hospital, Faenza, Italy, ¹²CRIBI Biotechnology Centre, University of Padua, Padua, Italy, ¹³Department of Neurological Sciences, Division of Neurology, Bellaria Hospital and University of Bologna, Bologna, Italy, ¹⁴Department of Neurology, Klinikum Nord Oschenscholl, Hamburg, Germany and ¹⁵CNR–Centro di Studio per la Biologia e Fisiopatologia Muscolare, Dipartimento di Scienze Biomediche Sperimentali, Padova, Italy

Received January 25, 2002; Revised and Accepted February 26, 2002 DDBJ/EMBL/GenBank accession nos[‡]

Autosomal dominant lateral temporal epilepsy (EPT; OMIM 600512) is a form of epilepsy characterized by partial seizures, usually preceded by auditory signs. The gene for this disorder has been mapped by linkage studies to chromosomal region 10q24. Here we show that mutations in the *LGI1* gene segregate with EPT in two families affected by this disorder. Both mutations introduce premature stop codons and thus prevent the production of the full-length protein from the affected allele. By immunohistochemical studies, we demonstrate that the *LGI1* protein, which contains several leucine-rich repeats, is expressed ubiquitously in the neuronal cell compartment of the brain. Moreover, we provide evidence for genetic heterogeneity within this disorder, since several other families with a phenotype consistent with this type of epilepsy lack mutations in the *LGI1* gene.

INTRODUCTION

Autosomal dominant lateral temporal epilepsy (EPT) refers to an epileptic syndrome characterized by focal seizures located in the lateral area of the temporal lobe, and is also termed lateral temporal epilepsy with acoustic aura or autosomal dominant partial epilepsy with auditory features. Ottman and colleagues

(1) provided the first description of this disorder. Clinically, this epileptic syndrome is characterized, in part, by auditory hallucinations, although the extent of patients experiencing these varies amongst the reported pedigrees (1,2). Other sensory symptoms are also present, such as visual hallucinations. The onset of the disease is quite variable even within pedigrees, although it usually affects young individuals. Seizures are

*To whom correspondence should be addressed at: Unitat de Genètica Molecular, Institut de Biomedicina de València–CSIC, Jaume Roig 11, E-46010 València, Spain. Fax: + 34 96 369 0800; Email: jpereztur@ibv.csic.es

†The authors wish it to be known that, in their opinion, the first two authors should be regarded as joint First Authors.

‡AF055636 (*LGI1/Epitempin*), AF473548 (splice form of *LGI1/Epitempin*) and AB067503 (*KIAA1916*)

scarce, usually occur during sleep periods and are well controlled with standard anti-epileptic treatments.

In 1995, a kindred with this type of focal epilepsy was linked to a 10 cM interval on 10q23–q24 (1), this linkage was later confirmed, in 1999, by our group in a second large pedigree (2), further refining the linked area to a 3 cM interval between markers D10S185 and D10S577. In this paper, we report the identification of mutations in a gene, *LGI1*, that segregate with the disease in two different pedigrees affected by EPT. In light of the proposed relationship between this gene and EPT, we propose a new, more descriptive, name for this gene: *LGI1/Epitempin* (approved by the HUGO Nomenclature Committee; symbol *LGI1*).

RESULTS

After further defining the region linked with EPT to a 3 cM interval, our group undertook both a positional cloning strategy

and a candidate gene approach to identify the gene responsible of this form of focal epilepsy (3,4). Our candidate gene strategy took into account, in addition to the known or inferred function of a gene, the expression pattern of the genes mapped to this region. Thus we were interested in genes expressed (exclusively or not) in the brain and, within the central nervous system, at least in the lateral part of the temporal lobe. One of these genes selected for screening was *LGI1*, described previously by Chernova and colleagues (5).

Mutation screening of *LGI1*

In silico analysis of public databases showed that *LGI1* consists of 8 coding exons that were analyzed by direct sequencing of PCR products. In family ADLTE01, we found a 1 bp deletion in position 758 in exon 7 (c758delC, numbering from the start codon). This deletion changes the reading frame after its occurrence and introduces a new mutant-specific sequence of

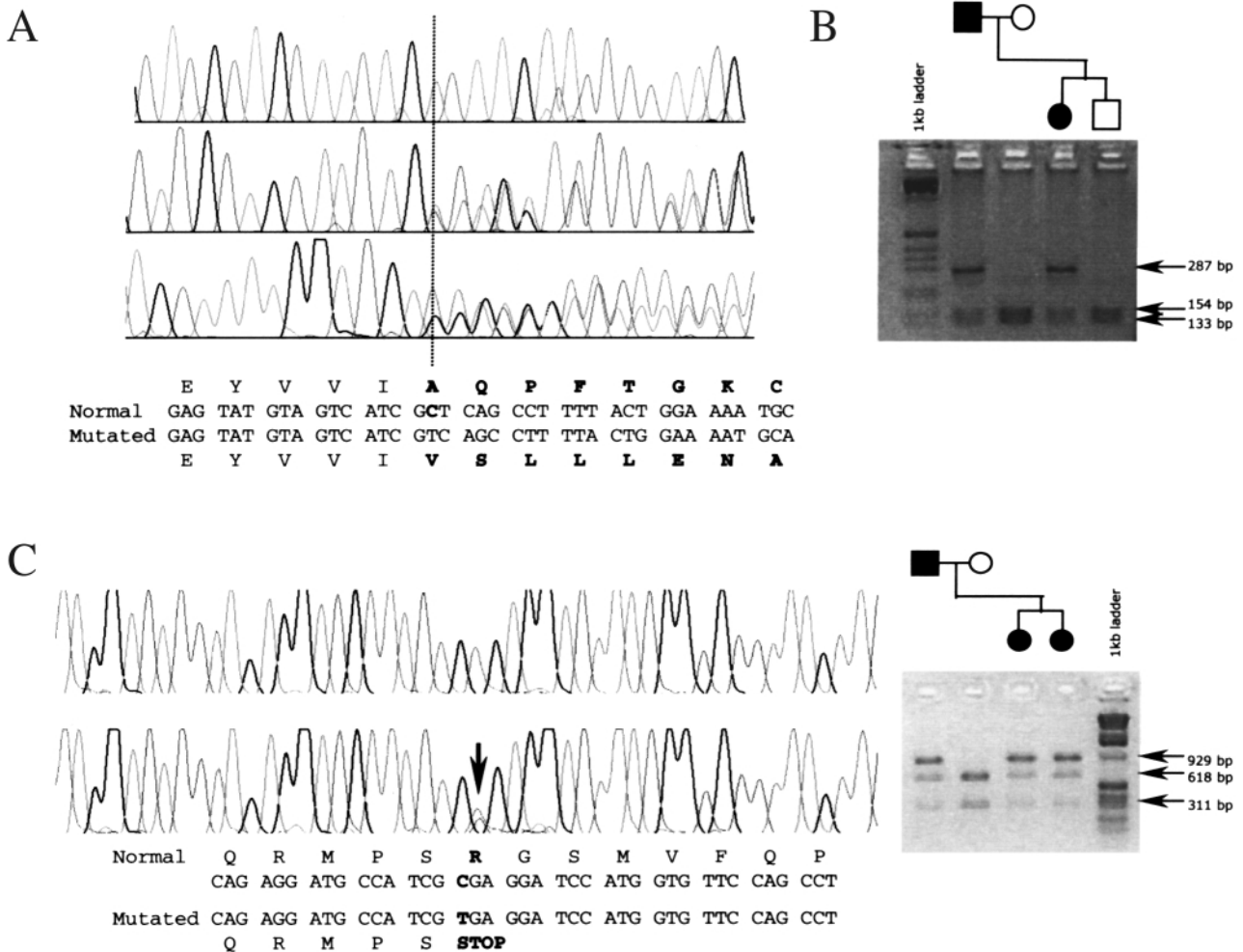


Figure 1. Mutations found in *LGI1*. (A) 758delC in family ADLTE01. Electropherograms from healthy (top) and affected (bottom, in both senses) individuals showing the deletion. The normal and mutated sequences are shown below the electropherograms. (B) Segregation analysis of 758delC with the disease within ADLTE01; only a nuclear family of the pedigree is shown. An RFLP assay was developed and applied to the complete pedigree. Mutant alleles are detected by the destruction of the DdeI site that remains intact in wild-type alleles (see the Materials and Methods section). (C) C1320T transition in family ADLTE03. Electropherograms from healthy (top) and an affected (bottom) individuals from family ADLTE03 and segregation analysis of the disease with an RFLP created by the presence of the mutation in the family (the complete pedigree tree is shown; see the Materials and Methods section for further details).

31 amino acids before the first in-frame stop codon is found. The change in the reading frame of the protein occurs at positions 862–927 of the cDNA, counting from the first codon. If translated, the mutated allele would produce a soluble protein of 283 amino acids instead of the 557 amino acids in the normal allele. This change segregated with the disease in all affected and obligate carriers of this family, whereas it was absent in 120 chromosomes from a healthy population with the same ethnic background (Fig. 1A, B). A second mutation was identified in family ADLTE03: a C→T transition found in position 1320 in exon 8. This change introduced a premature stop codon deleting the last 80 amino acids of the protein. In this nuclear family, the change also segregated with the disease and was not found in a population of similar ethnic background (Fig. 1C), strongly supporting LGI1 as the gene causing the disease. We did not find evidence suggesting the existence of other mutations in the remaining families analyzed (see the Materials and Methods section for a description of the families collected).

Alternative splicing of LGI1

As previously reported, LGI1 is mainly expressed in brain tissue, with no particular differences amongst different areas within the central nervous system (5). In contrast with Chernova and co-workers (5), we observed three different species by northern blot (Fig. 2A): the main form of about 2.4 kb and two additional species – one of 1.6 kb and an additional band at about 3.2 kb. By PCR amplification from a cDNA library, we cloned the two major species and characterized them as alternatively spliced isoforms of the gene arising from the usage of a splicing site located within exon 8 (Fig. 2B; GenBank accession no. AF473548). This splicing event seems to be highly regulated, since the relative intensity of the two mRNA species in the northern (5) (Fig. 2A) is different in brain and in skeletal muscle. The spliced protein is 292 amino acids long, and the usage of the alternative splicing site changes the reading frame of the C-terminal part of the protein, modifying the last 12 amino acids prior to the

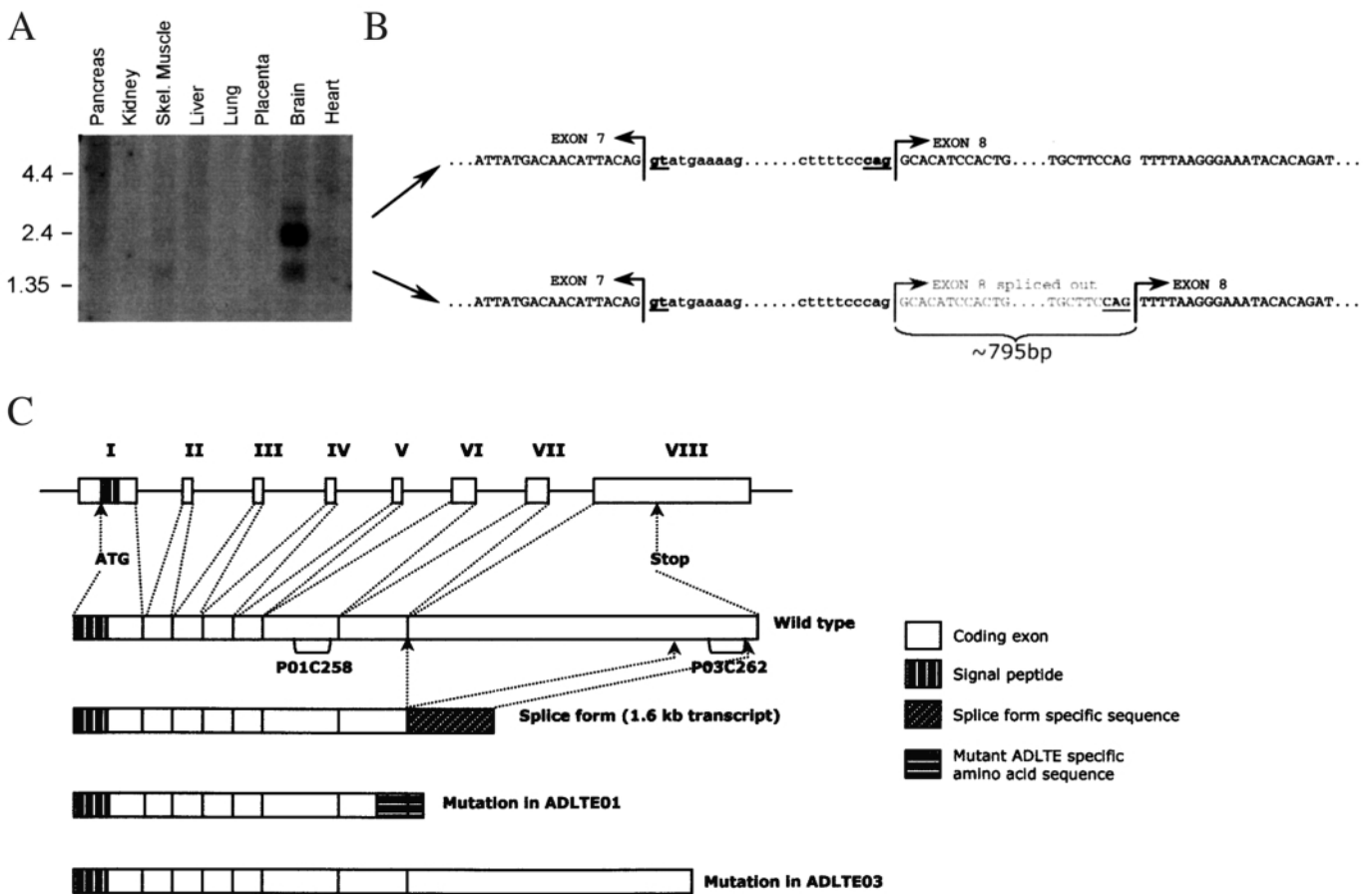


Figure 2. Alternative splicing of LGI1. (A) Northern blot of several tissues using IMAGE clone 178022. A strong signal is only observed in brain tissue, whereas a much weaker labeling of skeletal muscle can also be observed. (B) Alternative usage of a splicing site within exon 8. Capital letters represent exonic sequences, lower-case letters denote intronic sequences. Gray letters indicate the spliced-out portion of exon 8. (C) A graphical representation of LGI1 and the effect of mutations in the translated protein. P01C258 and P03C262 indicate the epitopes of the polyclonal antibodies raised in rabbit (see the Materials and Methods section for further details).

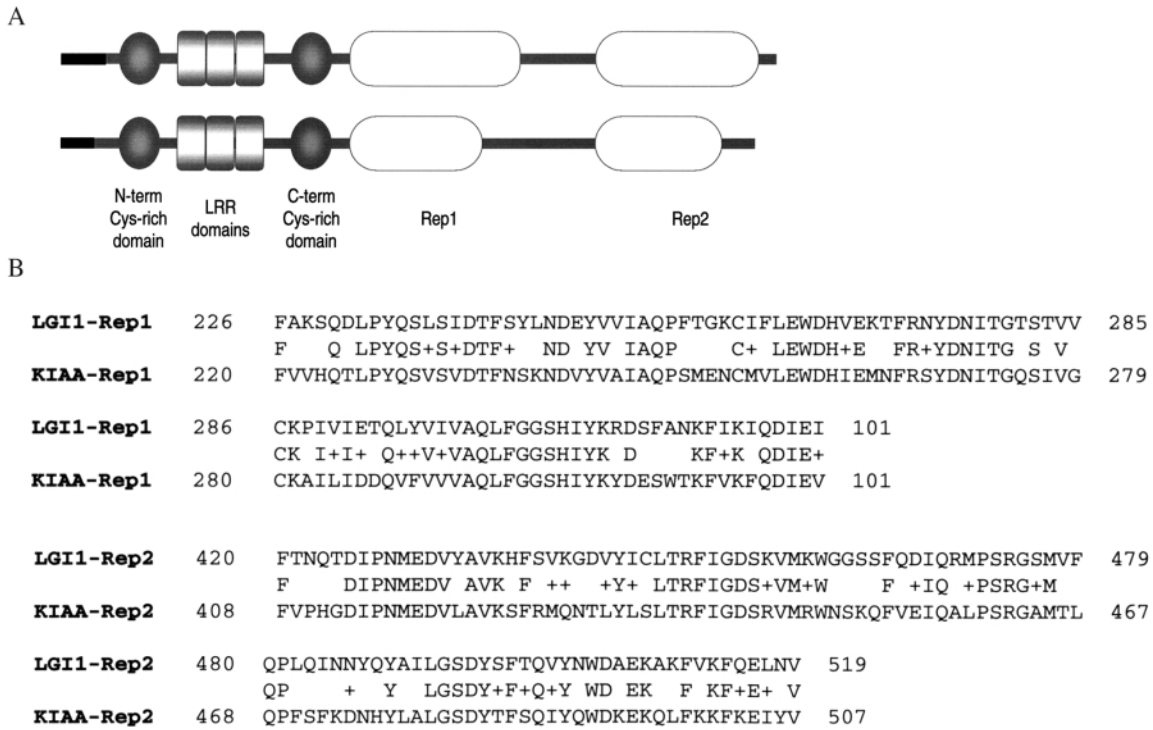


Figure 3. Domain structure of LGI1 and KIAA1916. (A) Schematic representation of the domain structures of LGI1 and KIAA1916, showing the similarities in domain composition and organisation. Black boxes, signal peptide; Gray circles, N-terminal and C-terminal Cys-rich domains; light gray squares, LRR domains; white boxes, internal repeats 1 and 2. (B) Alignment of the two internal repeats found in LGI1, with their homologous counterparts in KIAA1916.

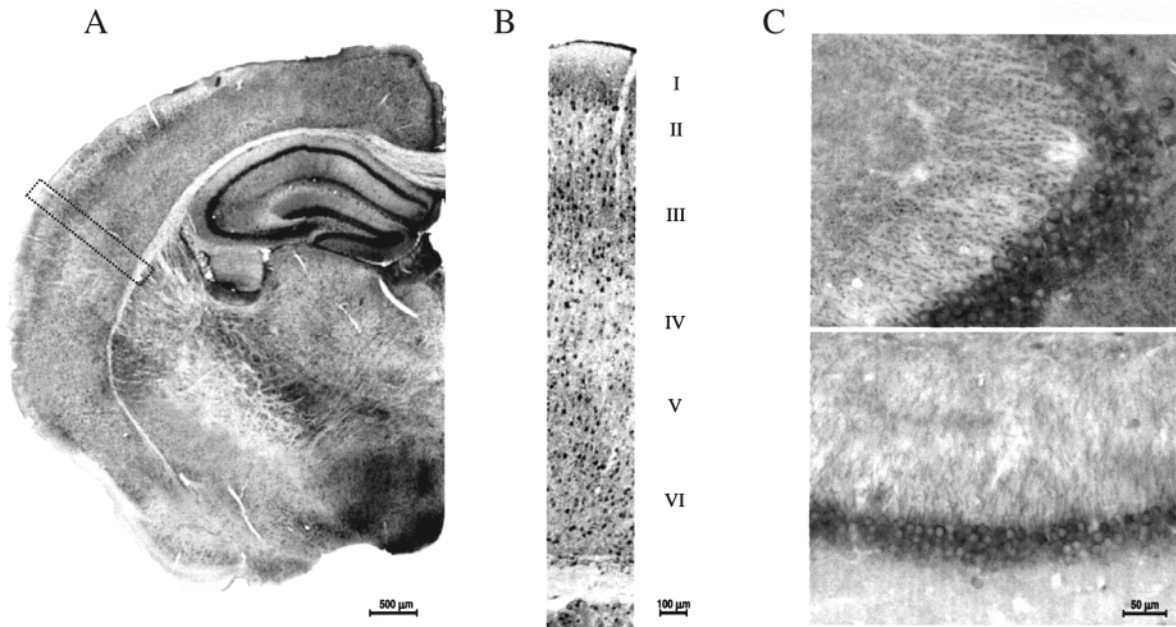


Figure 4. Expression of *Lgi1* in adult mouse brain. Antibodies P01C258, which recognizes both the full-length and the spliced isoforms, and P03C262, which only recognizes the full-length isoform, were used to stain coronal sections of adult mouse brain (for further details, see the Materials and Methods section). (A) Coronal image of half brain showing staining with P01C258 in all areas. (B) A representative image of cortical staining (inset in A) with P01C258 in the cortex of adult mouse brain. All cortical layers show staining of neuronal bodies. (C) Close-up images of representative sections from hippocampus stained with antibodies P01C258 (top panel) and P03C262 (bottom panel). Note that only a few interneurons are labeled in the molecular layer. The same images were obtained for the equivalent antibodies raised in different animals against the same epitopes (see the Materials and Methods section). Scale bars indicate magnification.

first stop codon in frame. The nature of the largest band remains to be elucidated, although it might be due to the usage of alternative transcription or a polyadenylation signal.

Domain architecture of LGI1

In order to provide further insight into the putative function of LGI1, we performed an analysis of the primary structure of the protein using the SMART web-tool. In addition to the described signal peptide and the three leucine-rich repeats and flanking cysteine-rich domains (5), within the SMART web-tool, PROSPERO identified two copies of an internal repeated sequence of 136 and 130 amino acids at positions 226–361 and 420–549 respectively. Interestingly, the first of these repeats includes the region previously identified as encoding a transmembrane domain (5). To ascertain whether this repeated sequence was an exclusive feature of LGI1 or whether it was also found in other human proteins as well, we performed a similarity search using the Advanced BLAST2 Search Service (see the Materials and Methods section) and identified a putative paralogue of LGI1 encoded by KIAA1916 (GenBank accession no. AB067503) mapped to chromosome 4p15.1 (Fig. 3). *In silico* analysis of the currently available human genomic and EST sequences (Golden Path website, <http://genome.ucsc.edu>) suggested that the KIAA1916 gene is organized in 8 exons and that the corresponding putative polypeptide is 545 amino acids long. It has a similar size to LGI1, and shares significant similarity in domain architecture, exhibiting an identity of 57% and a similarity of 73% extended throughout the polypeptide molecule (Fig. 3).

Expression of LGI1

LGI1 cDNA was initially cloned based on its rearrangements in malignant brain tumors (5). In order to obtain further insight on the physiology of the gene product, we raised antibodies against regions of LGI1 conserved between human and mouse (6).

Immunohistochemistry performed on adult mouse brain sections showed *Lgi1* to be widely expressed throughout the adult brain, mainly in neuronal cells, with little expression in glial cells (Fig. 4A). In this respect, there is an apparent higher level of expression of *Lgi1* in several areas of the brain, such as the hippocampus and the piriform cortex, although this is likely caused by the higher density of cell bodies located in those areas in contrast with the neocortex. Apparently, not all neurons are labeled; projection neurons are readily stained by both antibodies, whereas interneurons are infrequent (Fig. 4C). On the other hand, little labeling of glial cells can be observed throughout the brain – only in the white matter and other regions poor in neurons can a few cells be observed. Antibodies directed against the N-terminal part of LGI1 and against the C-terminal end of LGI1 yielded the same pattern of staining. In addition, polyclonal antibodies raised against the same antigens in different rabbits yielded comparable results (data not shown).

In mouse brain sections, in addition to cell bodies, we also observed punctuated staining characteristic of synaptic terminals (Fig. 5A). We also performed immunocytochemistry

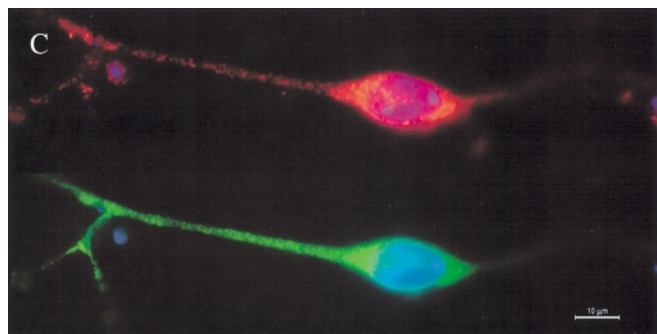
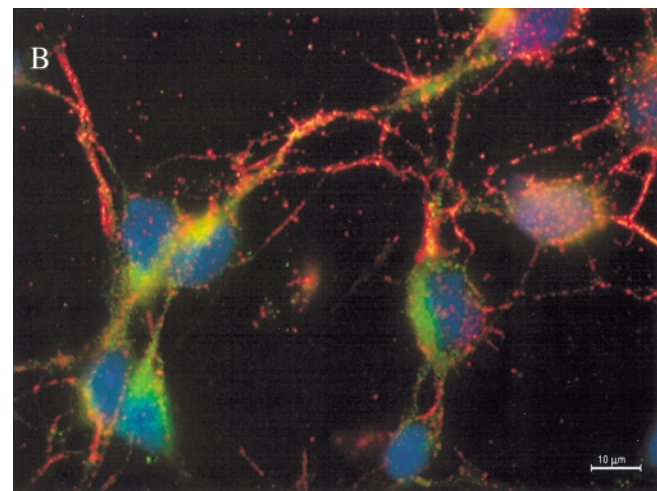
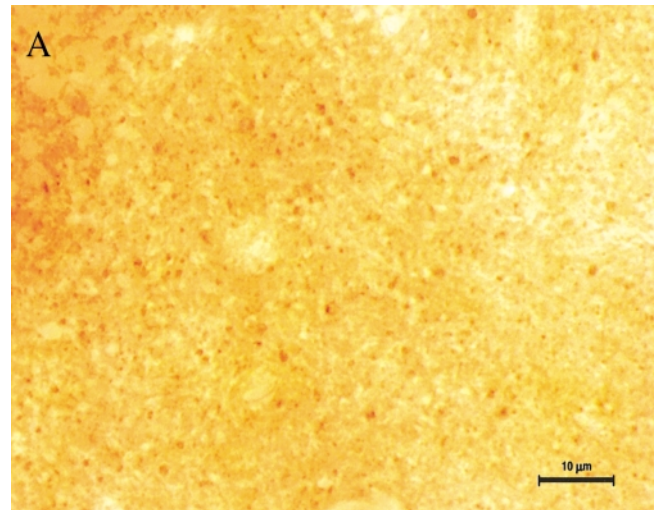


Figure 5. *Lgi1* in synaptic terminals. (A) Adult mouse brain section stained with antibody P01C258, showing a typical pattern of staining in synaptic terminals throughout the neuropil. (B) A composite immunostaining image of primary cultures from embryonic neurons with P01C258 (red), anti-map2 (green) and DAPI (blue). DAPI staining shows the localization of the nucleus, whereas anti-map-2 staining is used to reveal the cell body. (C) Composite image of P01C258 (red) with DAPI (left panel) and of anti-rab3 (green) with DAPI (right panel). P01C258 shows the punctuated pattern of staining typical of vesicle labelling within cell bodies and neurites, very similar to what is found with the anti-rab3 antibody. Scale bars indicate magnification.

experiments in primary cultures of embryonic cortical neurons. Cultured neurons show Lgi1 to be present in terminals, in cell bodies and in growing neurites (Fig. 5B). The pattern of staining resembles that obtained with antibodies recognizing synaptic proteins such as rab3 (Fig. 5C). Taken together, our results on cell culture indicates that the pattern of expression of LGI1 is consistent with it being a vesicle protein.

DISCUSSION

EPT is a mild form of epilepsy with an early onset, mild and scarce seizures, and excellent response to treatment. This genetic form of focal epilepsy, in the reported pedigrees, reaches penetrances of 70–80% (1,2). Here we report the identification of mutations in LGI1 in two unrelated Spanish families and provide further insight in the biology of this gene. Chernova and colleagues (5) previously identified LGI1 in glioblastoma cell lines as a tumor suppressor gene with a role in malignant brain tumors. Several candidate genes have been proposed in 10q23–q26 to account for some glioblastoma multiforme tumors (5). In the mutation carriers of the families analyzed here, neither the appearance of malignant glioblastoma nor an otherwise increased incidence of malignancies has been observed.

A signal peptide (positions 1–34), susceptible to direct LGI1 to the plasma membrane, and a single transmembrane domain have also been previously proposed by Chernova and collaborators (5). Although the prediction of a LGI1 transmembrane domain was obtained using computer programs that were state-of-the-art at the time they were employed (5), it was recently shown that they produce a rather high level of false-positive predictions (7). When we re-analyzed LGI1 using different computer programs shown to offer a lower false-positive rate (7), the significance values for the predicted transmembrane domain did not pass the program threshold. This suggests that LGI1 may enter the secretory pathway merely because of the presence of the signal peptide. The most remarkable feature of the protein is the presence of three leucine-rich repeats (LRR) in the predicted N-terminal domain of the protein, flanked by conserved cysteine-rich clusters. This type of repeat is the signature of a superfamily of proteins characterized by the presence of a relatively conserved sequence of 20–40 amino acids repeated a variable number of times. The functions of LRR-containing proteins are variable, and include signal transduction, cell adhesion and DNA repair, as well as extracellular matrix attachment (8) and plant steroid binding (9).

Recently, the mouse orthologue has been identified (6), although no other orthologues have been found in public databases in other model organisms, more distant in the evolutionary tree, such as *Drosophila* and *Caenorhabditis elegans*. Only weak similarities are found with other proteins harboring LRRs.

On the other hand, the presence of internal repeats also found in one other protein indicates that LGI1 could form part of a new family of proteins, with at least two members, combining both the LRR domains and the new type of repeats. The functional relationships between the two genes remains to be

elucidated. The new type of repeat was submitted to Pfam, and will appear in the next release under the name EPT.

The observed mutations in EPT families introduce premature stop codons in LGI1 by generating a frameshift in the protein or by a genomic transition translated into a nonsense change in the amino acid sequence. Surprisingly, these remarkable alternations in the protein sequence are not accompanied by a severe phenotype, as is the case in other genetic disorders with similar genetic mutations (see for example 10). This might reflect the fact that the mutated alleles, if translated, will produce proteins that are only slightly different from their normal counterparts. The mutation observed in ADLTE01 introduces a premature stop codon in exon 7 that generates a protein closely resembling the naturally occurring spliced isoform of LGI1 (Fig. 2C). This mutation could exert its effect by altering the isoform ratio in the brain. As shown in Figure 2A, the abundance of each isoform is not the same in brain as it appears to be in skeletal muscle. The appearance of a premature stop codon in the mRNA encoding LGI1 could either deviate the mutant allele towards the RNA degradation machinery or produce a truncated protein closely resembling the spliced isoform. In any of those cases, the regulated ratio between both isoforms in the brain would be distorted. On the other hand, the second pedigree bears a mutation that also truncates the protein – though near its C-terminal end. In both cases, the functional relevance of the mutations is subtle enough for the disease to be relatively mild but important enough, with respect to its function, to account for the high penetrance in families with mutations.

Despite this ubiquitous presence of LGI1 throughout the brain, the staining pattern shown by the antibodies indicates that the labeling is mainly due to projection neurons, such as pyramidal neurons from the cortex, hippocampal neurons and Purkinje cells (data not shown), whereas little labeling is seen in other neuronal types, such as interneurons (Fig. 4C). The fact that not all neurons seem to express LGI1 brings up the question of which subpopulations are involved and how this relates to the disease process. In addition to this, although some labelling of glial cells is also observed, especially in white matter, these cell types do not seem to express LGI1 in significant numbers. This finding might be responsible for the obvious but hitherto unexplained lack of LGI1 expression in malignant glioblastomas, which consist predominantly of glial cells rather than neuronal cells. We have also shown that the protein is located in synaptic terminals (Fig. 5), suggesting a role in synaptic development, integrity or maintenance. In addition, in the absence of a putative transmembrane domain, LGI1 appears to be a protein associated with soluble compartments in neurons – the punctuate staining of the synaptic terminals and of cell bodies supporting this interpretation.

No other families show definite linkage to 10q24. Even though a few pedigrees provide some evidence for the genetic involvement of this region (11,12), linkage data are not conclusive in most cases. Our mutation-detection strategy by direct sequencing does not detect all coding mutations in the gene; for example, deletions of coding exons could not be detected. Nevertheless, it is also possible that another gene or genes account for the familial aggregation of lateral temporal epilepsy, especially in those kindreds responding to the

'atypical' form of the disorder, with a heterogeneous phenotype or reduced penetrance (J.J. Poza, R. Michelucci and collaborators, in preparation).

Most of the genes described to date as causing autosomal dominant focal or generalized epilepsies are directly implicated in neuronal transmission: several subunits of ion channels as well as of neurotransmitter receptors are mutated in genetic forms of generalized epilepsy (for a review, see 13). In contrast, the gene product responsible for EPT does not seem to have such a clear role, suggesting that a new mechanism leading to an epileptic phenotype is taking place in our families. The presence of the LRR domains points towards a role in protein-protein interaction, possibly mediating either cell adhesion or interactions with the extracellular matrix. Nevertheless, we cannot rule out the possibility of LGI1 forming complexes with membrane components and interacting with membrane receptors or channels, somehow modulating their function or acting downstream in the transmission of the nervous impulse.

In this paper, we have described the finding of mutations in LGI1 in two pedigrees and provided further insight in the biology of this gene. Despite its ubiquitous neuronal expression, the phenotype caused by mutations in this gene is mainly observed in the lateral temporal lobe, where seizures start. How this regional specificity is achieved, the way in which the variation in the ratio between isoforms affects LGI1 function, and the influence of KIAA1916 in EPT are open questions that require further investigation. The identification of the interacting partners of LGI1 is likely to provide valuable information as to the pathogenic mechanism involved in EPT as well as to its localized effect.

After submission of our manuscript, Kalachikov and co-workers described similar findings in five families with EPT (14).

MATERIALS AND METHODS

Human samples

We have analyzed a total of six index cases from Spanish pedigrees (ADLTE01-06), one German family (family G) and four pedigrees of Italian origin (pedigrees I1-I4). The index

cases were clinically studied by at least one of us (J.J.P., J.G., J.F.M.M. or A.L.M. in Spain, V.-F.M. or L.K. in Germany, and D.P., R.M., C.A.T., S.B. or G.A. in Italy). Details on ADLTE01, family G and I1 are available elsewhere (2,11,12). Clinical phenotypes were consistent with lateral temporal epilepsy with auditory features, with the exception of family I4, which shows a clinical picture consistent with a mesial origin of the disease. Informed consent for genetic analysis was obtained for all samples, and the ethical committees from the respective institutions involved in sample collection approved all the protocols. Family ADLTE02 corresponds to the 'atypical' form of EPT (J.J. Poza, R. Michelucci and collaborators, in preparation). Most of the collected pedigrees consist of small families with a reduced number of affected individuals, and showed a pattern of inheritance consistent with an autosomal dominant trait. When available, genetic analysis of 10q24 showed results that could not rule out this locus as causative for the disorder.

Sequencing of LGI1

The genomic structure of LGI1 was obtained from information derived from the human genome project. LGI1 cDNA sequence (GenBank accession no. AF055636) was blasted against the human genome sequence draft and a finished BAC (AL358154) was identified that contained the genomic sequence of LGI1. The intron/exon boundaries were deduced *in silico* from the alignment of those two sequences, and intronic primers were designed to amplify and sequence the coding exons. Primer sequences and PCR conditions used are shown in Table 1.

Mutation screening

PCR products were purified (Concert nucleic acid purification system, Invitrogen) and subjected to cycle sequencing using the BigDye v2.0 dye terminator kit and an ABIPrism 3100 Genetic Analyzer (Applied Biosystems).

Segregation and population analysis of mutations

In order to ascertain segregation of the mutations within each family, we designed an RFLP assay for each one. The mutation

Table 1. Primers and PCR conditions used for the mutation screening of LGI1

Exon	Forward	Reverse	Size (bp)	Annealing temperature ^a
1	GGTGGACTCCTATGTGACCTG	TCTCTCTCCATGCCCTTCTAC	438	60°C
2	GCTAAACCCGGATTAACATAAGG	GGCTTATCCAAATATATGCC	219	48°C ^b
3	ACAATCTGTCAGTTTCACC	TTGAGAGTACTGTCTCCTGAATC	171	48°C
4	CAGGACAAGTACTCTCAAGTTC	ACAGGTGATCAAACCTGCATTG	197	60TD50
5	TCACTACAGTTTACATCACC	AGGCTTCCTTGTTAATGA	172	50°C
6	CGTGGGTAGGGTCCTTGTAC	TTTGAGGTGGAATGATGATG	249	60TD50
7	GGAAAGAGGTATTAGCTCAC	CACTCATCTTCTTGTGTC	287	50°C
8	TGGCCACACAATAATCTCTCC	TCATCCGGGTCATGTACTG	929	60TD50 ^b

^aAfter the initial denaturation step, all PCR profiles were performed with 35 cycles, with 30 s at 94°C; 30 s at the indicated temperature and a 45 s elongation step, followed by 5 min of final elongation at 72°C. 60TD50 denotes a touchdown protocol where, after 4 cycles at 60°C, the annealing temperature decreases by 0.5°C/cycle until it reaches 50°C. Then the remaining cycles up to a total of 35 are performed at this lowest temperature. All PCR reactions were performed with a Taq polymerase from Qiagen.

^bThe PCR requires the usage of the Q additive, which is part of the polymerase system.

in the ADLTE01 family, a 1 bp deletion, destroyed a DdeI site. Digestion of the amplified exon 7 with this enzyme at 37°C according to the manufacturer's instructions (MBI Fermentas) yields two fragments from the wild-type allele at 180 and 185 bp, whereas a single band at 365 bp appeared when the mutation is present.

On the other hand, the C→T transition in family ADLTE03 destroyed a NruI site. Digestion of the wild-type allele with NruI (MBI Fermentas) yields two fragments of 618 and 311 bp, whereas the undigested mutant allele yields a fragment of 929 bp.

To ascertain whether the changes found in the coding sequence of LGI1 were pathogenic mutations or polymorphisms, we analyzed two populations of about 120 chromosomes each of similar ethnic background as each of the families on which the mutations were found. ADLTE01 is a family of Basque descent, and 60 unrelated healthy individuals were typed for the presence of the 1 bp deletion in exon 7. Similarly, ADLTE03 is a family originally from southern Spain, and we analyzed a sample of 60 healthy individuals from the south-east part of Spain when looking for the C→T transition in exon 8 of LGI1.

Northern blot analysis

A nylon membrane containing mRNA from the tissues expressed was purchased from Clontech and hybridized with a probe obtained by random primer labeling of IMAGE clone 178022. Clone 178022 contains a 770 bp insert with sequence homologous to the 5' end of LGI1. The same probe was used to detect expression of LGI1 in different brain areas (data not shown).

Sequence analysis

The primary sequence of LGI1 was analyzed using the SMART server (Simple Modular Architecture Research Tool; <http://smart.embl-heidelberg.de>) for the annotation of the sequence (15,16). Analysis of the sequence to predict transmembrane domains was performed by TMHMM v2.0 (<http://www.cbs.dtu.dk/services/TMHMM/>). A search for possible paralogues was performed using the internal repeats detected by PROSPERO as query with the assistance of the Advanced BLAST2 Search Service of EMBL. In silico analysis of human genomic sequences using GoldenPath (<http://genome.ucsc.edu>) allowed us to identify a putative start codon upstream from the publicly available KIAA1916 sequence.

Antibody production

Polyclonal antibodies P01C258 and P03C262 were raised in rabbits against synthetic peptides coupled to KLH (DiverDrugs SL). Antibody P01C258 was raised against the peptide sequence found in the N-terminal region of LGI1 (positions 201–220), EGPPEYKRRKINSLSKDFD. On the other hand, antibody P03C262 was raised against the peptide DAEKAKFVKFQELNVQAPR (positions 505–522) located in the C-terminal domain of LGI1. Of these antibodies, P01C258 recognizes both the full-length protein and the spliced isoform, whereas P03C262 only recognizes the full-length isoform,

since this epitope is located in the spliced-out region in the shorter isoform. Two animals were immunized with each peptide, and the results of the experiments shown here are equivalent with the antibodies obtained in the corresponding animals (data not shown). After taking preimmune serum from each animal, rabbits were immunized every three weeks with Freund's adjuvant (complete or incomplete, Sigma-Aldrich); after the third injection, the immunization level of the animals was assayed using the same immunogens in an ELISA assay. Three more immunizations took place before the animals were bled. Immune sera were then obtained for each rabbit and the IgGs were purified by chromatography on a ProtA–Sepharose column (Amersham Biosciences).

Immunohistochemistry

Adult CD1 mice were anesthetized with chloral hydrate and transcardially perfused with 4% paraformaldehyde in 0.1 M phosphate buffer (pH 7.3; PB). Coronal sections of thickness 30 µm through the brain were obtained with a vibratome, quenched in 3% hydrogen peroxide in PB, and blocked with 10% normal goat serum in PB (blocking buffer). Sections were incubated overnight with primary antibodies P01C258 and P03C262, and diluted at 1:400–1:1000 in blocking buffer, and detection was performed using the ABC-peroxidase method (Pierce) following the manufacturer's instructions. PreadSORption of primary antibodies with an excess of the corresponding non-KLH-coupled peptide prior to their addition to the sections resulted in complete absence of all labeling (data not shown).

Cell culture and immunocytochemistry

Neurons dissociated from cortices of E15.5 embryos were plated on poly-L-lysine and cultured for three days in BME supplemented with 5% horse serum and 5% fetal bovine serum. Cultures were fixed with 4% paraformaldehyde in PB and incubated with either mouse anti-MAP2 (1:100; Chemicon) or mouse anti-rab3 (1:200; Synaptic systems) and rabbit P01C258 antibody (1:400) in blocking buffer, followed by Cy2-conjugated anti-mouse, Cy3-conjugated anti-rabbit (Jackson Immunochemicals) and DAPI (Sigma).

ACKNOWLEDGEMENTS

The authors acknowledge the collaboration of all family members who made possible this work. Dr Isabel Fariñas and her group are warmly acknowledged for their fruitful advice. Technical assistance from Mr Jorge Juan Sellés Martínez is also acknowledged. A.S. and A.G.-P. are fellows of the Department of Education of the Basque Government. J.M.M.-R. is a recipient of a FPU fellowship from the Spanish Ministry of Education. S.P.-S. is a recipient of a fellowship from the Fondo de Investigaciones Sanitarias. S.G. work was supported through a grant to R.S. from the Deutsche Krebshilfe 10-1643-Si1. P.S. was supported through a grant from Telethon Foundation ONLUS, Italy (no. B57). C.N. and R.M. were supported with grants from the Italian League Against Epilepsy, and N.K.M. and C.N. with a grant from the Joint Italian–Greek Committee for Scientific and Technological Cooperation (6th session,

project no. 37). This work has also been supported in part by grants from the Department of Health of the Basque Country (A.L.M.) and by Grants 00/0900 from the Fondo de Investigaciones Sanitarias and GV99-20-1-12 from the Generalitat Valenciana (J.P.-T.). Part of this work was presented at the December 2001 Meeting of the Spanish Neurological Society in Barcelona.

REFERENCES

- Ottman, R., Risch, N., Hauser, W.A., Pedley, T.A., Lee, J.H., Barker-Cummings, C., Lustenberger, A., Nagle, K.J., Lee, K.S., Scheuer, M.L. et al. (1995) Localization of a gene for partial epilepsy to chromosome 10q. *Nature Genet.*, 10, 56–60.
- Poza, J.J., Saenz, A., Martinez-Gil, A., Cheron, N., Cobo, A.M., Urtasun, M., Marti-Masso, J.F., Grid, D., Beckmann, J.S., Prud'homme, J.F. et al. (1999) Autosomal dominant lateral temporal epilepsy: clinical and genetic study of a large Basque pedigree linked to chromosome 10q. *Ann. Neurol.*, 45, 182–188.
- Nobile, C. and Pitzalis, S. (1999) Expression analysis of 21 transcripts physically anchored within the chromosomal region 10q24. *Genomics*, 62, 86–89.
- Nobile, C., Hinzmann, B., Scannapieco, P., Siebert, R., Zimbelloc, R., Perez-Tur, J., Sarafidou, T., Moschonas, N., French, L., Deloukas, P. et al. (2002) Identification and characterization of a novel human brain-specific gene, homologous to *S. scrofa* tnp83.5, in the chromosome 10q24 critical region for temporal lobe epilepsy and spastic paraplegia. *Gene*, 282, 87–94.
- Chernova, O.B., Somerville, R.P. and Cowell, J.K. (1998) A novel gene, LGI1, from 10q24 is rearranged and downregulated in malignant brain tumors. *Oncogene*, 17, 2873–2881.
- Somerville, R.P., Chernova, O., Liu, S., Shoshan, Y. and Cowell, J.K. (2000) Identification of the promoter, genomic structure, and mouse ortholog of LGI1. *Mamm. Genome.*, 11, 622–627.
- Möller, S., Croning, M.D.R. and Apweiler, R. (2001) Evaluation of methods for the prediction of membrane spanning regions. *Bioinformatics*, 17, 646–653.
- Andrade, M.A., Perez-Iratxeta, C. and Ponting, C.P. (2001) Protein repeats: structures, functions, and evolution. *J. Struct. Biol.*, 134, 117–131.
- Friedrichsen, D. and Chory, J. (2001) Steroid signaling in plants: from the cell surface to the nucleus. *Bioessays*, 23, 1028–1036.
- Yang, Y., Hentati, A., Deng, H.-X., Dabbagh, O., Sasaki, T., Hirano, M., Hung, W.-Y. and Ouahchi, K. (2001) The gene encoding alsin, a protein with three guanine-nucleotide exchange factor domains, is mutated in a form of recessive amyotrophic lateral sclerosis. *Nature Genet.*, 29, 160–165.
- Mautner, V.F., Lindenau, M., Gottesleben, A., Goetze, G. and Kluwe, L. (2000) Supporting evidence of a gene for partial epilepsy on 10q. *Neurogenetics*, 3, 31–34.
- Michelucci, R., Passarelli, D., Pitzalis, S., Dal Corso, G., Tassinari, C.A. and Nobile, C. (2000) Autosomal dominant partial epilepsy with auditory features: description of a new family. *Epilepsia*, 41, 967–970.
- Steinlein, O.K. (2001) Genes and mutations in idiopathic epilepsy. *Am. J. Med. Genet.*, 106, 139–145.
- Kalachikov, S., Evgrafov, O., Ross, B., Winawer, M., Barker-Cummings, C., Martinelli Boneschi, F., Choi, C., Morozov, P., Das, K., Teplitskaya, E. et al. (2002) Mutations in LGI1 cause autosomal-dominant partial epilepsy with auditory features. *Nature Genet.*, 30, 335–341.
- Schultz, J., Milpetz, F., Bork, P. and Ponting, C. (1998) SMART, a simple modular architecture research tool: identification of signaling domains. *Proc. Natl Acad. Sci. USA*, 95, 5857–5864.
- Letunic, I., Goodstadt, L., Dickens, N.J., Doerks, T., Schultz, J., Mott, R., Ciccarelli, F., Copley, R.R., Ponting, C.P. and Bork, P. (2002) Recent improvements to the SMART domain-based sequence annotation resource. *Nucleic Acids Res.*, 30, 242–244.

Identification and characterization of a novel human brain-specific gene, homologous to *S. scrofa tmp83.5*, in the chromosome 10q24 critical region for temporal lobe epilepsy and spastic paraplegia

C. Nobile^{a,*}, B. Hinzmann^b, P. Scannapieco^c, R. Siebert^d, R. Zimbello^c, J. Perez-Tur^e, T. Sarafidou^f, N.K. Moschonas^f, L. French^g, P. Deloukas^g, A. Ciccodicola^h, S. Gesk^d, J.J. Pozaⁱ, C. Lo Nigroⁱ, M. Seriⁱ, B. Schlegelberger^d, A. Rosenthal^b, G. Valle^c, A. Lopez de Munain^j, C.A. Tassinari^k, R. Michelucci^k

^aCNR-Centro di Studio per la Biologia e Fisiopatologia Muscolare, Dipartimento di Scienze Biomediche Sperimentali, Viale G. Colombo 3, 35121 Padova, Italy

^bmetaGen Gesellschaft für Genomforschung mbH, Berlin, Germany

^cCRIBI, University of Padua, Padova, Italy

^dInstitute of Human Genetics, University Hospital Kiel, Kiel, Germany

^eUnitat de Genètica Molecular, Institut de Biomedicina de València (CSIC), València, Spain

^fInstitute of Molecular Biology and Biotechnology, Heraklion, Greece

^gSanger Centre, Hinxton Hall, Hinxton, Cambridge, UK

^hIstituto Internazionale di Genetica e Biofisica del CNR, Napoli, Italy

ⁱLaboratorio di Genetica Molecolare, Istituto Giannina Gaslini, Genova, Italy

^jDepartment of Neurology, Hospital Ntra Sra de Aranzazu, San Sebastian, Spain

^kDivisione di Neurologia, Ospedale Bellaria, Bologna, Italy

Received 11 June 2001; received in revised form 29 October 2001; accepted 20 November 2001

Received by R. Di Lauro

Abstract

We describe the structure, genomic organization, and some transcription features of a human brain-specific gene previously localized to the genomic region involved in temporal lobe epilepsy and spastic paraplegia on chromosome 10q24. The gene, which consists of six exons disseminated over 16 kb of genomic DNA, is highly homologous to the porcine *tmp83.5* gene and encodes a putative transmembrane protein of 141 amino acids. Unlike its porcine homolog, from which two mRNAs with different 5'-sequences are transcribed, the human gene apparently encodes three mRNA species with 3'-untranslated regions of different sizes. Mutation analysis of its coding sequence in families affected with temporal lobe epilepsy or spastic paraplegia linked to 10q24 do not support the involvement of this gene in either diseases. © 2002 Elsevier Science B.V. All rights reserved.

Keywords: Conserved gene; Brain expression; Transmembrane domain; Partial epilepsy

1. Introduction

Genes selectively expressed in only one tissue or organ presumably encode proteins with specialized functions. In particular, brain-specific genes are likely to play a role in fundamental neurological processes and their mutations may cause or predispose to cerebral disorders.

Autosomal dominant temporal lobe epilepsy has recently been recognized as a distinct form of idiopathic partial epilepsy (Berkovic et al., 1996), though it appears to be heterogeneous, both clinically and genetically. A clinical subtype (EPT; OMIM No. 600512) characterized by sensory

Abbreviations: bp, base pair(s); cDNA, DNA complementary to RNA; dNTP, deoxyribonucleoside triphosphate; kb, kilobase(s) or 1000 bp; EST, expressed sequence tag; UTR, untranslated region; PCR, polymerase chain reaction; RACE, rapid amplification cDNA ends; ORF, open reading frame; poly-A tail, polyadenylation tail; ADLTE, autosomal dominant lateral temporal epilepsy; SPG9, spastic paraplegia type 9; *tmp83.5*, transmembrane protein 83.5 (*S. scrofa*); MEMP, membrane protein (*S. scrofa*); *sp83.5*, small protein 83.5 (*S. scrofa*); *TLL2*, tollid-like 2 gene; *DNTT*, deoxynucleotidyltransferase gene; *HTMP10*, human transmembrane protein 10 gene

* Corresponding author. Tel.: +39-49-827-6214; fax: +39-49-827-6209.

E-mail address: carlo@telethon.bio.unipd.it (C. Nobile).

(auditory or visual) symptoms has been mapped to chromosome band 10q24 in two large pedigrees (Ottman et al., 1995; Poza et al., 1999). Because the characterizing sensory symptoms suggest that seizures originate from the lateral area of the temporal lobe cortex, this subtype is also named autosomal dominant lateral temporal epilepsy (ADLTE; Poza et al., 1999). On the other hand, partial epilepsy of probable mesial temporal (limbic) origin was not found to link to 10q24 in several families (Berkovic et al., 1996).

The 10q24 critical region for EPT/ADLTE spans about 3 cM, between markers D10S185 and D10S577 (Ottman et al., 1995; Poza et al., 1999). It overlaps the larger genomic region, spanning 12 cM at 10q23.3–q24.2, recently linked to a form of autosomal dominant spastic paraplegia (SPG9; Seri et al., 1999). To identify candidate genes for these diseases, we constructed a YAC map of the 10q24 region (Gray et al., 1997; Siebert et al., unpublished) and placed on it numerous brain expressed sequence tags (ESTs; Nobile et al., 1998; Lo Nigro et al., 2000). Subsequent expression analysis of these ESTs led to the identification of several putative genes selectively or predominantly expressed in the brain (Nobile and Pitzalis, 1999), which may be involved in neurological disorders. In this paper, we describe the structure, genomic organization, and some expression features of one of these genes.

2. Materials and methods

2.1. Sequencing

Sequencing of the IMAGE cDNA clone 31056, PCR and RACE PCR products was performed using the dye terminator chemistry (Big Dye Terminator Cycle Sequencing kit – ABI Prism Perkin Elmer, Noewalk, CT, USA) following the user manual instructions. The electrophoresis of the cycle-sequencing products was carried out in an ABI 377 automatic sequencer (Applied Biosystems, Foster City, CA, USA).

2.2. Similarity searches, EST alignments and protein prediction

BLAST searches in dbEST (Boguski et al., 1993), assembly of ESTs and automatic extension of consensus sequence was performed by using the program AUTEX (Schmitt et al., 1999) until no additional matching ESTs could be detected. Genomic similarity searches were performed at http://www.sanger.ac.uk/HGP/blast_server.shtml against the chromosome 10 database. Exon prediction from genomic sequences was obtained by Genscan (Burge and Karlin, 1997) and MZEF (Zhang, 1997) programs. The trace file resource of the mouse shotgun genome sequencing project (<http://trace.ensembl.org/>) was used to correct and extend mouse EST consensus sequences. Analysis of coding sequences was performed by using the ORF Finder program

(<http://www.ncbi.nlm.nih.gov/gorf/gorf.html>). Protein sequence similarity searches were performed at NCBI (<http://www.ncbi.nlm.nih.gov/BLAST/>) against non-redundant protein databases. Prediction of transmembrane domains in deduced amino acid sequences was accomplished by TMpred (Persson and Argos, 1994).

2.3. RACE PCR analysis and Northern hybridization

5'-RACE PCR experiments were performed using the human brain Marathon cDNA amplification kit (Clontech, Palo Alto, CA, USA), according to the manufacturer's instructions, and the following primers: 5R: 5'-CATGGCCTCAATGCTGCTTCTTCTTC-3'; 5RN (nested): 5'-GTATGCCCGCCGCTAATCCAAGAGA-3'. PCR amplifications were performed in a thermal cycler (MJ Research, Waltham, MA, USA) under the following conditions: 1 min at 94°C, followed by 30 s at 94°C and 2 min at 72°C (five cycles), 30 s at 94°C and 2 min at 70°C (five cycles), and 20 s at 94°C and 2 min at 68°C (25 or 20 cycles with primers 5R and 5RN, respectively). Amplification products were fractionated on a 2% agarose gel, gel-eluted, and sequenced.

Northern blots Human Brain II (Clontech, Palo Alto, CA, USA) were probed with PCR fragments amplified from human brain Marathon cDNA (Clontech, Palo Alto, CA, USA) across the putative coding sequence (probe A, primer L: AAATGAAAGACACAAGCCAATAAAG; R: CTTTCCTCTCCAAGTACAAAACC) and a portion of the 3'-untranslated region (probe B, primer L: AACCAAGGGT-TTTGTAAGTTGGAG; R: CACATGGTTTCTACCCTGAGA) of the gene under study and with the insert of the IMAGE cDNA 31056. DNA fragments were ³²P-labeled and Northern-hybridized according to standard protocols. Following hybridization, membranes were exposed under X-ray films at –70°C for 2–4 days.

2.4. Mutation analysis

PCR amplification of DNA fragments containing each individual exon from patient and control DNAs were performed in 10 µl 1X PCR buffer, 10 pmol primers, 200 mM dNTPs, and 0.5 units Taq polymerase in a thermal cycler (MJ Research, Waltham, MA, USA). PCR amplification conditions were as follows: 5 min at 94°C, followed by 33 cycles (1 min at 94°C, 1 min at 58°C, 1 min at 72°C). PCR products were fractionated by electrophoresis on a 2% agarose gel, gel-eluted and sequenced. The following exon-specific primer pairs were used (5' – > 3'): E1L: GAGATGGAGGTATAGCCAAGAAAG; E1R: GAAAGGGAACATAGGTGAAACATC; E2L: TTCAACAGTTTTCTGTC-TCAAAGG; E2R: GTGAGGGGTTTATGTGAGTTCTG; E3L: CCTAATTCTTTTCTCCTTGTAGCC; E3R: ATC-TGTGATTTTGCATGTTCTCC; E4L: ACGTTTATGATCCATGCAGATTTT; E4R: GTTACCCCAAATACCATCTTTGTC; E5L: ATTTCTTGGCTTAGGTATGGGTTT; E5R: CTCATACTGTGCTTCTTTTGGTGA; E6L: CTC-

CTTGCCTGTTGTCTAGTGTCT; E6R: TCTTTCCTC-TCCAAGTACAAAACC.

The same primers were used to perform semiquantitative PCR analysis of patient and control DNAs. The most appropriate number of replication rounds was determined preliminarily by amplifying 50 and 100 ng of genomic DNA for a number of cycles varying from 17 to 21 under the conditions described above. We then chose to PCR amplify 100 ng of DNA from one patient and one healthy member of each family for 18 cycles. PCR products were fractionated by electrophoresis on a 8% acrylamide gel and visualized by silver staining.

3. Results

3.1. Transcript analysis

Among the 10q24 brain ESTs previously analyzed at the expression level, one, stSG2296, showed selective and abundant expression in the brain (Nobile and Pitzalis, 1999). To isolate the corresponding full-length transcript, we first sequenced the insert of the IMAGE cDNA clone 31056, containing the stSG2296 marker. BLASTN searches in dbEST (Boguski et al., 1993) were initially performed with this sequence. Several significantly matching human EST sequences were detected and subsequently assembled by using the GAP program (Bonfield et al., 1995). The resulting new consensus sequence was then analyzed by the automatic iterative BLAST search and assembly procedure (AUTEX; Schmitt et al., 1999) until no additional matching ESTs could be detected. The resulting assembly contained two IMAGE cDNA clones, 23653 and c-3ma02, at the 3' end, the 5'-readings of which were available in dbEST and had the same 5'-start sequence. We sequenced the complete 2.8-kb insert of clone 23653. A BLASTX search of the resulting consensus versus the non-redundant protein database GenBank (<http://www.ncbi.nlm.nih.gov:80/BLAST/>) revealed striking homology to two porcine proteins (see below, this section) in the first 45 bp. BLASTN

search versus the chromosome-10 genomic sequence database showed that the human transcript was located on BACs RP11-35J23 (accession no. AL136181) and RP11-304C14 (AL391136), the sequences of which were then subjected to further investigation. The exon-prediction programs Genscan (Burge and Karlin, 1997) and MZEF (Zhang, 1997) predicted together five exons in a region where also a TBLASTN search of the porcine peptide sequences gave significant hits, allowing us to establish a gene model. Thus, by combining EST analysis with exon prediction and homology searches, we assembled a consensus sequence of 3235 bp, incorporating EST yc93b02.r1 (A008W15), the 5' portion of which was confirmed by sequencing PCR products amplified from human brain cDNA. We further extended this transcript sequence towards its 5'-end by PCR analysis of human brain cDNA. Several alternative PCR forward primers designed from the genomic sequence adjacent to our putative first exon were used with the same reverse primer specific for a putative internal exon. Some PCR reactions yielded amplification products, the sequences of which enabled us to extend the size of the transcript to 3554 bp (sequence deposited in the EMBL/DBJ/GenBank databases under the accession no. AF367761). Part of this transcript sequence is shown in Fig. 1.

BLAST analysis of the 3554-bp transcript sequence against nucleotide databases revealed high similarity to two porcine mRNAs: the region from nucleotide 317–994 showed 82% identity to MEMP (GenBank accession No. X89480); and the region from nucleotide 480–994 showed identical degree of similarity to sp83.5 (accession No. Y11026). Both these transcripts derive from the recently identified *S. scrofa* gene *tmp83.5* (accession No. X99893), but they have different 5'-sequences that are generated by alternative promoters (Bangsow et al., 1998). In addition, a similar degree of sequence homology was shown between a 5'-portion of the human transcript and several mouse brain ESTs (for example, accession no. AU035253 and BE651315). By using trace data from the mouse shotgun genome sequencing project (<http://trace.ensembl.org/>), we

```

ccagcctcagcctcatgtaaccattgtttccaaagggctcagcagctgcagaagcccccaagctctcaggccagctgtggctcttttctgctgagatggaggatagaccagaagaagcagcct 120
gaaaagtaagaaggacagaaaaatcaaggcaagagggtgggtgaaccaggccatacacgtcggccatgcagcagcagcccatcttcccagggatggactgagggtctggctactcccc 240
tgaccataaattggcttggccagggtctctctggagccacctctcagcctgcctctgcagctttgtcagtttaactgttcttcaaaatgaaagacacaagcaataaagccatggagaagg 360
agcttaccaaaggcagtgtaacgaagaaggttctctgggagactgtcagaaATGAGTTTTACTGAACTTACCCTGCCGGCGAACACACGTCCTCTCTGTACAGGTGGGAAAGAAAC 480
                                     M S F S L N F T L P A N T T S S P V T G G K E T
GGACTGTGGGCCCTCTCTTGGATTAGCGCGGGCATAACCATTGCTGTTGGCCACAGCCCTGCTGGTGGCTTTACTATTACTTTGATTCCCGAAGAAGAGCAGCATTGAGGCCATGGA 600
D C G P S L G L A A G I P L L V A T A L L V A L L F T L I H R R R S S I E A M E
GGAAAGTGACAGACCATGTGAAATTCAGAAATTGATGACAAATCCCAAGATATCTGAGAATCTAGGAGATCACCCACACATGAGAAGAATACGATGGGAGCACAAAGGGCCACATATA 720
E S D R P C E I S E I D D N P K I S E N P R R S P T H E K N T M G A Q E A H I Y
TGTGAAGACTGTAGCAGGAAGCGAGGAACCTGTGCATGACCGTTACCGTCTACTATAGAAATGAAAGAGGGGATGTGGTGGATTGTGCCAGACTGAGCCTGGAATGAGgcag 840
V K T V A G S E E P V H D R Y R P T I E M E R R R G L W W I V P R L S L E *
cgagtcgaaggagcagcagacctggcactggaacagggttgaaaaccaagggttttctactggagagaaagatgccaagctgcttcttaatacaatccaaatcttcatctacagctctgg 960
aacactttggggctgatcttctctctttaggggacatcccacaacatggttaattccaactcagatctctgtgctcttagttagtacaatgtgactcaccagatggggtccttagatccatatt 1080
cctgtcccagtgaggaaatttctctctcttctgctatttgggaaaggggttgggttctcagtgctctgctctctcatcttttttttccatcccttttccaaaaagccatcagatct 1200
gacttccatggaagtgtgctgaggtcagcctggtgcaagttgggatacaaatgaaactatgcagagatggtgagaggaagcagtaattggttctccgaatctcagggtaggaacct 1320
gtggagccacacattccctgaccacaggaagcactggctcaatcatgtcacacagcagtggaagaatacggactcttaagtcacactaccactgagcagctgtcagactttggagaa 1440
gttgttaacatttcaagcctcagtttttctgttttaaggaggggaaatattgctctatgcatataatgaaaagatataaataagaaataaaggggaagtgctgctactagtgtcc 1560
agtcaaaatgttagtctctctctaccaccttctctcactctctccatattgctgctgataaaacagctaatcaccagcatctgttcccocatagtcacagggccacacaagga 1680
acatttaggacaaacttctccatggcctatgatccaaatgttatcctaaagatgattctaggtgtgctggtgagtagtgaatctccaatctcaggtgtgagctgctccatagat 1800
caggaaaaggcagttctctacaagttccgaattccaaatacacagactggtggtgttacatttaacctaaagatgttaattgtgatggaatctatgttccatataaaacaacacttt 1920

```

Fig. 1. Partial cDNA sequence (from nucleotide 1–1920) and predicted amino acid sequence of *HTMP10*. The poly(A) signals at positions 1535 and 1910 are underlined.


```

agctggtgcccctacaaaatggggagacaaaagcctgcagtgccagtgagagagaagcctgccaagggggcctgtcagaaATGAGTTTTCTACTGAACCTTACACTGCCATCGAATACGACAT 120
M S F S L N F T L P S N T T S
CCTCTCCAGTTGTCACGAGTGCACCAAGCCACAGACTGTGGTCCCTCTATTGGACTGGCAGCCGGAATCCCATCCCTGCTGGCCACAGCCCTGCTGGTGGCTTTGCTGTTTACCTTGATCC 240
S P V V T S A K A T D C G P S I G L A A G I P S L L A T A L L V A L L F T L I Q
AGCGAAGAAGAACCATCGATGATGAGCCGTAGAGGAGACTGAGATACCATGTGAAATTCAGAACTCTATGACAAATCCCAAAATATCTGAGAATCCTGAGCGGTACCACACATGAAA 360
R R R T I D D E P V E E T E I P C E I S E L Y D N P K I S E N P R R S P T H E M
TGAACCCACGGGGCAGCAAGAAGGCCACATATATGTGAAGACTGTCTCCGGGAAGCGAGGAGCCCTGCCGGACACTACCGTCTCTCTGAAGAACTAGAAAGGAGAAGGACTTTGGT 480
N P R G S Q E G H I Y V K T V S G S E E P L P D T Y R P P E E L E R R R G L W W
GGCTTGATACCAGCTGAGCCTAGAAATGATgtcaccgcatcaggggagcagaccagctggatcaggggtgaaaactggaatctgtgctcagggttcagcttggagagggcaggagat 600
L V P S L S L E *
cagctgcctctcactcaacatcaatttcattatagatccaagacacatctggctgccttgcctcttaggggatgggtgcagggcaggttcattcgatctcagaccttcaggggtaatca 720
gcaagggcaccagggtgagtcgactatgtgtccagatgagcctcgtatttctgttctgaatcgagagaagctgccttgggaaacggcttggctcccagatgtcctgcctctctctc 840
tacctgctccaaatgttatcatttaaaaaataataaagctttgagagatggctttttggggggtgcatgacaaaggccagctgggcacaagttaggacaggcataaaaacctaaccaaa 960
gacatgtgacagaagcatttcttctctttgaaatctcaggataaagctcgatggagccgtccatgccccgacacaaggaagatgtggctgatttggctcacaacagtgagaatgag 1080
ctaggcactccctagtcagagctgtgctgagcactttgagaagtcagatgagctctacacctccccgagcctgtttggggtctgtgcttcttggccactggaaaaatcacagca 1200
ggaataaaggaagcaatttccacttagtagacactagagatatacaactctcaccagcttcttccacatctccactctacttatccatccacagagtcacaaagtcgacaagga 1320
aatctaaaccttctctgtggcctgtaactctgaaattttgacatttaaaaaaactctagtggtcactggcagctctcctcactctgactctgacacacatttttccccctgggtgtcaa 1440
ccaagtccaacatcacagtagcagagatctgtgttacacattacatatcaactcttaggatgctcgagctcctggatggtaaatctcactgtcaactcaatgggagtggagaatccaacag 1560
gagacacacctccagaagatttaactgagtaagaagatccacctgaaatgtggccagcactcatccatagcctgggggtctggggaattcaagttgagcagcagaacctcatcactatcc 1680
ttcctgacgacagataaagtggttggccagctacccccacactcctgctgccacaccttcttggactgaacaaaataaatacctccacccttaagttgactttatcatacactgagag 1800
aagtaagctaataataaactgatattcactgttttaatgacttttcttggtaatgagccatttggcttgaagtaccataaagaaaagatggttgtgagggagaggatcaagaggaccaaa 1920
    
```

Fig. 2. Partial cDNA sequence (from nucleotide 1–1920) and predicted amino acid sequence of *Mtmp*.

were able to assemble the genomic sequence bearing the mouse ortholog of our gene. The use of all available ESTs covering our genomic sequence allowed us to assemble a cDNA with a consensus length of 2602 bp, partially shown in Fig. 2, with an open reading frame (ORF) of 429 nucleotides (sequence deposited in the GenBank third party databases (TPA) under the accession no. BK000002).

The largest ORF detected in the human 3554-bp transcript sequence spans 426 bp, between the AUG initiation codon at position 410 and the stop codon at position 833, resulting in a predicted protein of 141 amino acids (Fig. 1). As shown in Fig. 3, this amino acid sequence is only 1-residue shorter than the porcine protein encoded by the MEMP mRNA, the overall amino acid identity being 80.3%, and is highly homologous (76% identity) also to the predicted mouse protein shown in Fig. 2. The porcine protein contains an internal hydrophobic region (positions 31–54; Fig. 3), which may represent a single-spanning membrane domain (see Bangsow et al., 1998). Applying the program TMpred (Persson and Argos, 1994), both the human and mouse protein sequences show a strong transmembrane prediction at the positions 30–53 and 31–54,

respectively, with scores above 2000 where 500 is regarded as the significance threshold. The N-terminus is predicted to be outside the cell, which is in good concordance to Bangsow et al. (1998). The region of the human protein homologous to the smaller porcine isoform sp83.5 is confined to the C-terminus (amino acids 96–141). Thus, based on these nucleotide and amino acid similarities, the human gene under study, here referred to as *HTMP10*, appears to be homologous to the porcine *tmp83.5* gene described by Bangsow et al. (1998) and is likely to have a murine homolog (*Mtmp*).

3.2. Exon-intron organization of *HTMP10*

BLASTN searches in the chromosome 10 database (<http://www.sanger.ac.uk>) with the *HTMP10* transcript sequence identified fully matching genomic sequences in the BAC clones RP11-35J23 (AL136181) and RP11-304C14 (AL391136). Compared with the cDNA sequence, the genomic organization of *HTMP10* consists of six exons (>412, 36, 33, 120, 57 and 2897 bp) and five introns (3603, 3845, 1550, 1359 and 2170 bp), encompassing together 16

1	MSFSLNFTLP	ANTTSSPVVT	SGKGADCGPS	LGLAAGIPSL	VATALLVALL	50	tmp83.5
1	MSFSLNFTLP	ANTTSSPVVTG	-GKETDCGPS	LGLAAGIPLL	VATALLVALL	49	htmp10
1	MSFSLNFTLP	SNTTSSPVVT	SAKATDCGPS	IGLAAGIPSL	LATALLVLL	50	mtmp
51	LILIHRRRRS	S-ESTEEIER	PCEISEIYDN	PRVAENPRRS	PTHEKNIMGA	99	tmp83.5
50	FTLIHRRRSS	I-EAMEESDR	PCEISEIDDN	PKISENPRRS	PTHEKNIMGA	98	htmp10
51	FTLIQRRRTI	DDEPVEETEI	PCEISELYDN	PKISENPRRS	PTHEMNPRGS	100	mtmp
100	EEAHIVKTV	SGSQEPMRDT	YRPAVEMERR	RGLWWLIPRL	SLE	142	tmp83.5
99	QEAHIVKTV	AGSEEPVHDR	YRPTIEMERR	RGLWWLIPRL	SLE	141	htmp10
101	QEGHIVKTV	SGSQEPLPDT	YRPPEELERR	RGLWWLIPSL	SLE	143	mtmp

Fig. 3. Sequence comparison of human (htmp10), mouse (mtmp) and porcine (tmp83.5) putative transmembrane proteins. The deduced human amino acid sequence is aligned with the homologous proteins. Residues conserved in all three species are shown in bold; (-) indicates a gap.

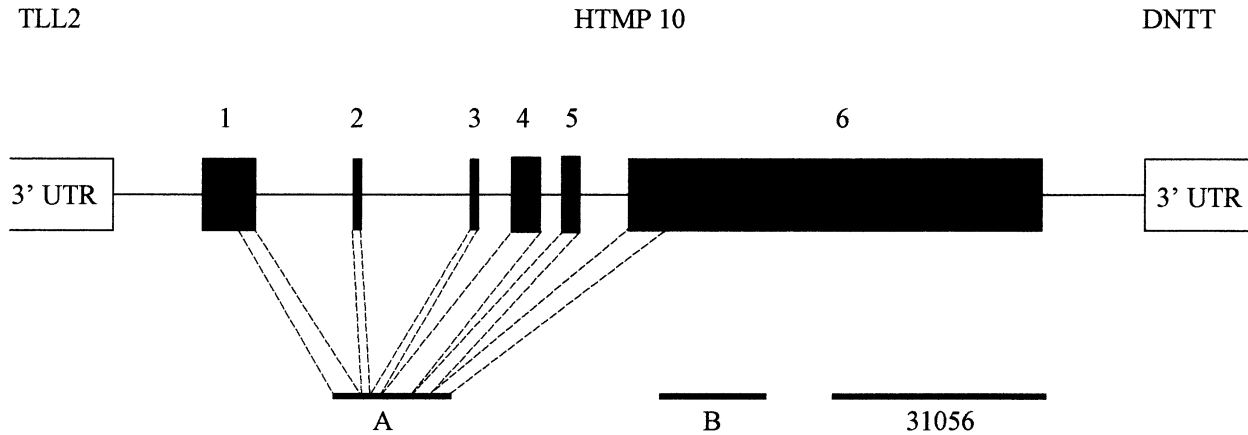


Fig. 4. Exon-intron organization of *HTMP10*. Exons are shown as solid bars and numbered consecutively, introns are shown as thin segments. Brackets on both sides, represent the terminal portions of the 3'-UTR regions of the flanking genes *TLL2* and *DNTT*. Segments A, B, and 31056 correspond to the gene fragments used as probes in Northern experiments.

kb (Fig. 4). All exon-intron boundaries follow the GT/AG rule concerning intronic consensus sequences necessary for splicing (Table 1). We confirmed the presence and size of each intron by PCR analysis of human DNA with primers designed from flanking exon sequences (data not shown). The overall genomic organizations of the human and porcine (MEMP-encoding) homologous genes appear to be precisely conserved. The initiation codon corresponds to the last nucleotide triplet of exon 1, and all exon-intron borders occur at identical positions. Exon sizes are therefore identical in the two species with the exceptions of exon 3, which is 3-bp shorter in the human gene, and exon 6, which is considerably larger in humans (2897 versus 1816 bp). There is an alternate polyadenylation site (AGUAAA) at position -20 upstream from the cDNA end (see EMBL/ DDBJ/GenBank accession no. AF367761). Recent investigations (Beaudoing et al., 2000) show that this hexamer has the highest prevalence after the well known AAUAAA and AUUAAA. All three ESTs (yy60b01.s1, yc93b02.s1, and yg02g05.s1) where trace files could be inspected (ftp:// genome.wustl.edu/pub/gsc1) show a clear poly-A tail, thus defining the precise 3' end of the *HTMP10* gene. At the genomic level, the deoxynucleotidyltransferase (*DNTT*) gene ends approximately 5 kb downstream in a tail to tail orientation (Fig. 4), whereas only 7.2 kb upstream of our putative 5'-end the Tolloid-like 2 (*TLL2*) gene has its 3'-

end. Based on the results of our PCR analysis of the genomic region adjacent to *HTMP10* on the 5'-side (see Section 3.1), the precise start of exon 1 is likely to lie within 100 bp upstream of the first nucleotide of the 3554-bp transcript sequence.

3.3. Expression analysis

The porcine MEMP and sp83.5 transcript isoforms differ in their 5'-regions, the common downstream region starting from the first nucleotide of exon 3 (see Bangsow et al., 1998). To determine whether mRNA isoforms generated by alternative promoters are expressed in humans, we performed 5'-RACE PCR experiments with primers designed from the *HTMP10* transcript sequence. An antisense PCR primer lying within exon 4 (from nucleotide 518–494) amplified a single product of about 200 bp from brain mRNA (data not shown), the sequence of which precisely matched the 5'-coding region and part of the 5'-untranslated region (5'-UTR) of the transcript sequence shown in Fig. 1. Thus, expression of a human sp83.5-like mRNA is not supported by 5'-RACE as well as EST analysis.

To further characterize the expression features of *HTMP10* we performed Northern analysis of specific cerebral regions with cDNA probes spanning different portions of the 3554-bp transcript sequence, as schematically shown

Table 1
Exon-intron boundaries and sizes of exons and introns of *HTMP10*

Exon number	Exon size(bp)	5'-boundary	3'-boundary	Intron number	Intron size(bp)
1	>412	N.D. ^a	AAATGgtgag	1	3603
2	36	tgcagAGTTT	ACACAgtaag	2	3845
3	33	tacagACGTC	AAACGgtaac	3	1550
4	120	gccagGACTG	TGGAGgtgat	4	1359
5	57	ttagGAAAG	CTGAGgtaag	5	2170
6	2897	ttaagAATCC	AACCTaaaa		

^a N.D., not determined.

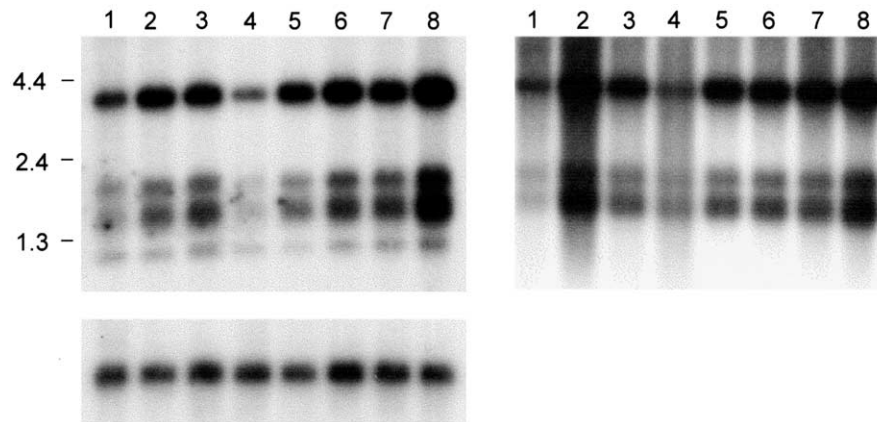


Fig. 5. Northern blot hybridization analysis of *HTMP10*. Northern blot membranes were probed with cDNA fragments spanning the coding sequence (probe A, left), and part of the 3'-UTR (probe B, right). Lanes 1, cerebellum; 2, cerebral cortex; 3, medulla; 4, spinal cord; 5, occipital lobe; 6, frontal lobe; 7, temporal lobe; and 8, putamen. The approximate positions of RNA size markers are indicated on the left. Hybridization signals detected by the human actin probe on the same filter are shown at the bottom of the left panel.

in Fig. 4. A cDNA fragment spanning the predicted coding sequence (probe A, nucleotide 323–913) hybridized with four mRNAs, approximately 3.7, 1.9, 1.5 and 1.0 kb in length (Fig. 5, left). As shown in the right panel of Fig. 5, the three larger bands, 3.7, 1.9 and 1.5 kb, were also detected by the downstream probe B (nucleotide 884–1323). Finally, the insert of the IMAGE cDNA 31056, corresponding to nucleotide 1857 to 3'-end, hybridized only with the 3.7-kb mRNA, as previously shown (Nobile and Pitzalis, 1999). The three isoforms detected by probe B very likely result from alternative 3'-UTR terminations. The 1.5-kb isoform presumably terminates at the canonical AAUAAA site at position 1535 (Fig. 1). Another polyadenylation site at position 1910 (AUUAAA; see Fig. 1) likely serves as termination site for the transcript form of about 1.9 kb, in agreement with the finding of an EST (qy72c10.x1) with a poly-A tail beginning at position 1932. The distance of 21 bp between the hexamer end and the beginning of the poly-A tail is in good conformity to the results of Beaudoin et al. (2000). As indicated in Section 3.2, the third polyadenylation site (AGUAAA) ends at position 3534, exactly 20 bp upstream of the end of the largest transcript. Altogether, these three mRNA isoforms show lower expression levels in cerebellum and, particularly, spinal cord than in other brain regions (Fig. 5). The smaller and weaker mRNA band (about 1.0 kb) detected only by probe A may result from cross-hybridization with a transcript moderately homologous to the *HTMP10* coding sequence.

3.4. Mutation analysis

Because of its brain-specific expression, *HTMP10* can be regarded as a candidate for neurological disorders. We designed exon-specific primer pairs from neighboring intron sequences (see Section 2.4) and amplified each exon from one patient and one healthy member of a large family affected with temporal lobe epilepsy with auditory and/or visual

symptoms linked to 10q24 (Poza et al., 1999). One affected and one unaffected individuals from a smaller family presenting only auditory features (Michelucci et al., 2000) were also tested. In both families, sequence analysis of exon-containing PCR fragments did not reveal any disease-associated mutations in the coding sequence, nor at splice junctions (data not shown). In addition, semiquantitative PCR analysis failed to reveal disease-causing deletions of the *HTMP10* coding region in these families (Fig. 6).

Finally, mutation analysis performed in individuals from two families affected with complicated spastic paraplegia (SPG9) linked to 10q24 (Seri et al. 1999; Lo Nigro et al. 2000) also failed to detect disease-associated mutations.

4. Discussion

We identified a novel human brain-specific gene, *HTMP10*, while searching for genes of neurological interest in the chromosome 10q24 region. This gene is homologous to the *S. scrofa tmp83.5* gene (Bangsow et al., 1998), based

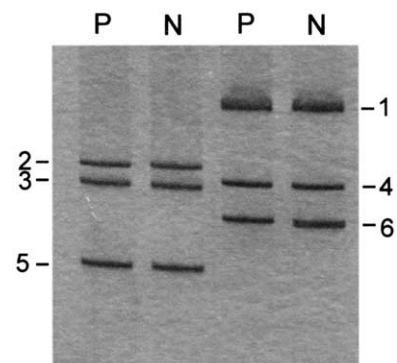


Fig. 6. Example of semiquantitative PCR analysis. DNAs from a patient (P) and a healthy member (N) of the chromosome 10q24-linked temporal epilepsy family described by Poza et al. (1999) were tested with primers specific for individual exons (2–5) or exon coding portions (1 and 6).

on conserved coding sequence and exon-intron organization. The two homologous genes, however, appear to differ at the transcription level. The porcine *tmp83.5* expresses two transcript forms, MEMP, 2674 bp, and sp83.5, 2374 bp, which are generated by alternative promoters and encode proteins of different sizes, the smaller (sp83.5) corresponding to the C-terminal portion of the larger one, encoded by MEMP (Bangsow et al., 1998). On the other hand, *HTMP10* encodes three mRNA isoforms with 3'-UTRs of different sizes, the largest of which, >3554 bp, corresponds to MEMP. Expression of a human sp83.5-like isoform generated by an alternative promoter was not supported by 5'-RACE PCR experiments as well as EST analysis.

The function of the porcine MEMP-encoded protein has not yet been elucidated. Based on the features of its amino acid sequence, Bangsow et al. (1998) suggested that it may be an integral membrane protein composed of a N-terminal extracellular domain (amino acids 1–30), followed by a single transmembrane anchor (amino acids 31–54) and a C-terminal cytoplasmic region of 87 amino acids. Because it is expressed mainly in neurons (Bangsow et al., 1998), the MEMP protein might therefore be a receptor molecule, or subunit of receptor complex, involved in the control of neuronal activity or development. The same structural features, including the putative transmembrane domain, are conserved in the *HTMP10*-encoded protein (Fig. 3), suggesting that it may also be anchored to the neuronal membrane. Such structural similarity may imply similar functions for the human and porcine transmembrane proteins, though expression of the cytoplasmic sp83.5 protein in *S. scrofa* suggests that the *tmp83.5* gene may exert more complex functions than *HTMP10*.

Similarly to the porcine transcripts, the largest *HTMP10* mRNA has an unusually long 3'-UTR, whereas additional transcript forms with shorter 3'-UTRs are encoded by this gene. 3'-UTRs are known to contain signals that control the translation, stability, and location of mRNA within the cell. Therefore, transcription of several mRNA forms with 3'-UTRs of various lengths may represent a way to modulate expression of *HTMP10* by post-transcriptional mechanisms.

Because of its brain specificity, *HTMP10* can be regarded as a candidate for 10q24-linked neurological disorders. Mutation analysis performed in two families affected with ADLTE (Poza et al., 1999; Michelucci et al., 2000) failed to detect disease-associated mutations in the coding sequence of *HTMP10*. Mutations affecting the coding sequence of this gene, however, might occur in other ADLTE families linked to 10q24 (Ottman et al., 1995; Mautner et al., 2000), not analyzed in this work. Moreover, because the 5'-regulatory region of the gene has not yet been investigated, the occurrence of mutations in the promoter region cannot be excluded. Identification of the transcriptional control region will make it possible to understand the regulation of this gene and complete mutation analysis.

Acknowledgements

C. N. thanks Professor G.A. Danieli for encouragement and generous hospitality. This work was supported by the Italian League Against Epilepsy (LICE; grant to C. N. and R. M.), the Deutsche Krebshilfe (Grant 10-1643-Sil), the Joint Italian-Greek Committee for Scientific and Technological Cooperation (6th session, project No. 37) and the Basque Government (grant to A. L. M.).

References

- Bangsow, T., Schepelman, S., Martin, C., May, M., Oberthur, A., Perl, S., Knupfer, E., Zinke, H., Gassen, H.G., 1998. Identification of a gene selectively expressed in the brain, which encodes a putative transmembrane protein and a soluble cytoplasmic isoform. *Eur. J. Biochem.* 256, 24–35.
- Beaudoing, E., Freier, S., Wyatt, J.R., Claverie, J.M., Gautheret, D., 2000. Patterns of variant polyadenylation signal usage in human genes. *Genome Res.* 10, 1001–1010.
- Berkovic, S.F., McIntosh, A., Howell, A.R., Mitchell, A., Sheffield, L.J., Hopper, J.L., 1996. Familial temporal lobe epilepsy: a common disorder identified in twins. *Ann. Neurol.* 40, 227–235.
- Boguski, M.S., Lowe, T.M., Tolstoshev, C.M., 1993. DbEST-Database for 'expressed sequence tags'. *Nat. Genet.* 4, 332–333.
- Bonfield, J.K., Smith, K.F., Staden, R., 1995. A new DNA sequence assembly program. *Nucleic Acids Res.* 23, 4992–4999.
- Burge, C., Karlin, S., 1997. Prediction of complete gene structures in human genomic DNA. *J. Mol. Biol.* 268, 78–94.
- Gray, I.C., Fallowfield, J., Ford, S., Nobile, C., Volpi, E.V., Spurr, N.K., 1997. An integrated physical and genetic map spanning chromosome band 10q24. *Genomics* 43, 85–88.
- Lo Nigro, C., Cusano, R., Scaranari, M., Cinti, R., Forabosco, P., Brescia Morra, V., De Michele, G., Santoro, L., Davies, S., Hurst, J., Devoto, M., Ravazzolo, R., Seri, M., 2000. A refined physical and transcriptional map of the SPG9 locus on 10q23.3–q24.2. *Eur. J. Hum. Genet.* 8, 777–782.
- Mautner, V.F., Lindenau, M., Gottesleben, A., Goetze, G., Kluwe, L., 2000. Supporting evidence of a gene for partial epilepsy on 10q. *Neurogenetics* 3, 31–34.
- Michelucci, R., Passarelli, D., Pitzalis, S., Dal Corso, G., Tassinari, C.A., Nobile, C., 2000. Autosomal dominant temporal lobe epilepsy with auditory features: description of a new family. *Epilepsia* 41, 967–970.
- Nobile, C., Pitzalis, S., 1999. Expression analysis of 21 transcripts physically anchored within the chromosomal region 10q24. *Genomics* 62, 86–89.
- Nobile, C., Manca, A., Pisano, M., Angius, A., Gray, I.C., Spurr, N.K., 1998. A refined physical and EST map spanning 7.4 Mb of 10q24, a region involved in neurological disorders. *Mamm. Genome* 9, 835–837.
- Ottman, R., Risch, N., Hauser, W.A., Pedley, T.A., Lee, J.H., Burkert-Cummings, C., Lustenberger, A., Nagle, K.J., et al., 1995. Localization of a gene for partial epilepsy to chromosome 10q. *Nat. Genet.* 10, 56–60.
- Persson, P., Argos, P., 1994. Prediction of transmembrane segments in proteins utilising multiple sequence alignments. *J. Mol. Biol.* 237, 182–192.
- Poza, J.J., Saenz, A., Martinez-Gil, A., Cheron, N., Cobo, A.M., Urtasun, M., Marti-Masso, J.F., Grid, D., Beckman, J.S., Prud'homme, J.F., Lopez de Munain, A., 1999. Autosomal dominant lateral temporal epilepsy: clinical and genetic study of a large Basque pedigree linked to chromosome 10q. *Ann. Neurol.* 45, 182–188.
- Schmitt, A.O., Specht, T., Beckman, G., Dahl, E., Pilarsky, C.P., Hinzmann, B., Rosenthal, A., 1999. Exhaustive mining of EST libraries

- for genes differentially expressed in normal and tumor tissues. *Nucleic Acids Res.* 27, 4251–4260.
- Seri, M., Cusano, R., Forabosco, P., Cinti, R., Caroli, F., Picco, P., Brescia Morra, V., De Michele, G., Lerone, M., et al., 1999. Genetic mapping to 10q23.3–q24.2, in a large Italian pedigree, of a new syndrome showing bilateral cataracts, gastroesophageal reflux, and spastic paraparesis with amyotrophy. *Am. J. Hum. Genet.* 64, 586–593.
- Zhang, M.Q., 1997. Identification of protein coding regions in the human genome by quadratic discriminant analysis. *Proc. Natl. Acad. Sci. USA* 94, 565–568.

Short sequence-paper

ZNF232: structure and expression analysis of a novel human C₂H₂ zinc finger gene¹, member of the SCAN/LeR domain subfamily

Lampros A. Mavrogiannis^{a,2}, Alexandros Argyrokastritis^a, Nicholas Tzitzikas^a,
Emmanuel Dermitzakis^{a,3}, Theologia Sarafidou^a, Philippos C. Patsalis^b,
Nicholas K. Moschonas^{a,c,*}

^a Department of Biology, University of Crete, P.O. Box 2208, 71409 Heraklion, Greece

^b Cyprus Institute of Neurology and Genetics, P.O. Box 3462, 1683 Nicosia, Cyprus

^c Institute of Molecular Biology and Biotechnology, Foundation for Research and Technology-Hellas, P.O. Box 1527, 71110 Heraklion, Greece

Received 26 October 2000; received in revised form 18 January 2001; accepted 18 January 2001

Abstract

We have identified a novel zinc finger gene, *ZNF232*, mapped to human chromosome 17p12. The coding region of the gene is organized in three exons corresponding to a 417 amino acid long polypeptide containing a SCAN/LeR domain and five C₂H₂-type zinc fingers. *ZNF232* is possibly a nuclear protein, as suggested by expression analysis of GFP/*ZNF232* chimeric constructs. *ZNF232* transcripts were detected in a wide collection of adult human tissues. The gene is possibly subjected to tissue-specific post-transcriptional regulation by means of alternative splicing. © 2001 Elsevier Science B.V. All rights reserved.

Keywords: Zinc finger protein; SCAN/LeR domain; Human erythroleukemia K562 cell; Normalized cDNA library; Chromosome 17; Nucleus

During the past decade, partial cDNA sequences (expressed sequence tags, ESTs) corresponding to novel human genes, obtained through systematic library screening, have been rapidly accumulated and exploited for the construction of an integrated framework genome map [1,2]. In addition to ordinary cDNA libraries, screening of normalized libraries greatly facilitated the identification of relatively rare transcripts and significantly contributed to the enrichment of the EST database [3]. Currently, determination of novel full-length cDNAs and sequence comparisons with already characterized genes frequently reveal well-known protein motifs, thus enabling essential gene annotation, i.e. classification of novel genes to various

gene/protein families [4,5] and incorporation of analyzed genes to the emerging high resolution human genic map.

Human erythroleukemia K562 cells, considered as pluripotent hematopoietic progenitors, have been proven to be useful in studying regulated gene expression. Numerous studies on important biological functions such as developmental stage-specific gene switching, signal transduction and transcriptional regulation have been based on the analysis of genes active in K562 cells [6–9]. In order to systematically search for novel genes expressed in these cells, we generated a normalized cDNA library [10] and isolated a number of partial cDNA clones. Among several ESTs of unknown nature, EST H5IMBB246 (GenBank accession No. X93860), displaying similarities to genes coding for C₂H₂-type zinc finger proteins, was considered for further analysis. In order to obtain a full-length cDNA and avoid possible sequence ambiguities due to normalization, we screened the corresponding native K562 library [10]. A 114 bp ³²P-labeled PCR product, synthesized by using a pair of EST-deduced primers (5'-AGATGGGGT-GCTCATCTTG-3' and 5'-CCGATGCTGACTTAGA-TATG-3', nucleotide (nt) positions 1149–1167 and 1243–1262, respectively), was applied as a probe. PCR conditions included 35 cycles with denaturation at 94°C (1 min), annealing at 57°C (1 min) and extension at 72°C (2 min).

* Corresponding author, at address c. Fax: +30-81-394-408;

E-mail: moschon@imbb.forth.gr

¹ The cDNA sequence of *ZNF232* (name approved by the HUGO/GDB Nomenclature Committee) has been submitted to GenBank and assigned accession No. Y15067; the corresponding genomic DNA sequences are available under accession Nos. AF080169–171.

² Present address: Clinical Genetics Group, Institute of Molecular Medicine, John Radcliffe Hospital, Oxford OX3 9DS, UK.

³ Present address: Institute of Molecular Evolutionary Genetics, Department of Biology, Pennsylvania State University, University Park, PA 16802, USA.

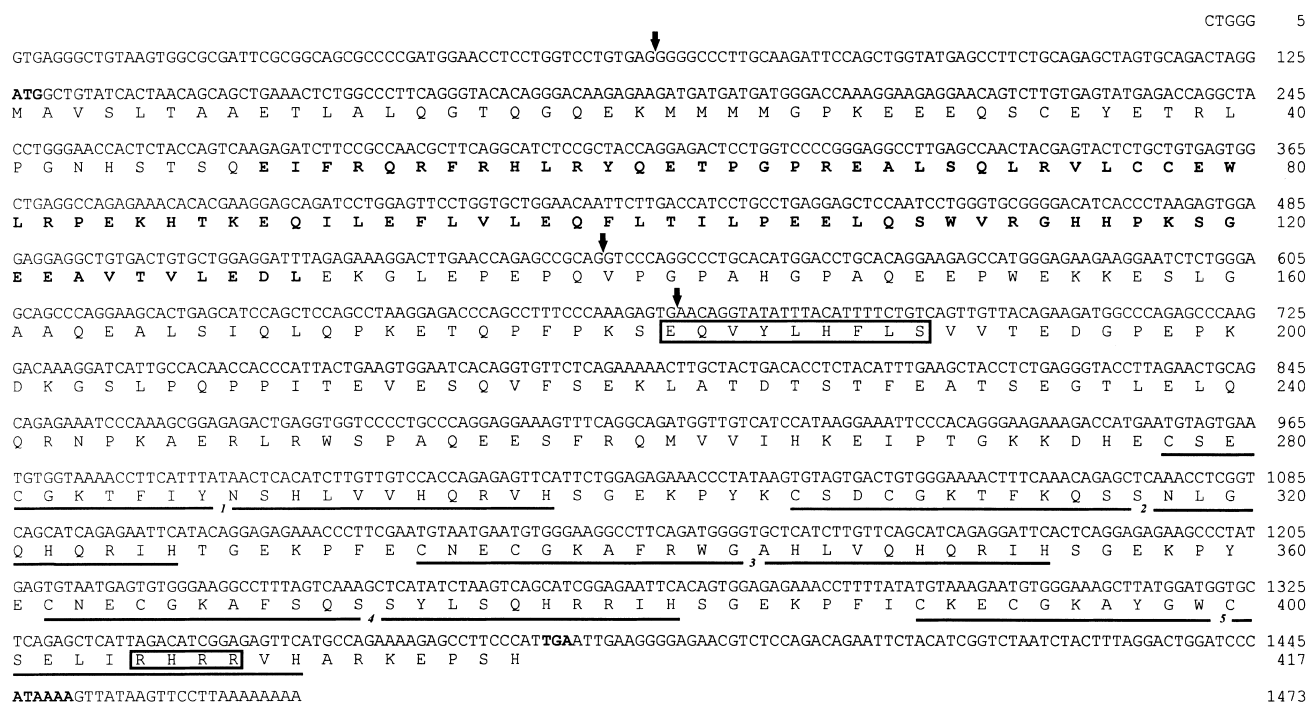


Fig. 1. Nt and aa sequences derived from the analysis of three largely overlapping *ZNF232* cDNA clones, i.e. No. 3a (nt 1–1473), No. 6a (nt 131–1445) and No. 7b (nt 370–1473). The proposed translation start codon (nt 126), the corresponding stop codon (nt 1376) and the SCAN/LeR domain (aa 49–130) are indicated in bold. The five consecutive Cys₂His₂ zinc finger repeats (numbered 1 to 5) and the unique polyadenylation signal are underlined. An internal nine amino acid long peptide (aa 182–190) corresponding to a region of *ZNF232* transcript that is subjected to alternative splicing and the putative nuclear localization signal (RHRR) are framed. Vertical arrows on the cDNA sequence denote the exon/intron boundaries.

Five positive clones with insert sizes ranging from 1.1 to 1.35 kb were isolated, subcloned into pBluescript II KS+ (Stratagene), sequenced (Sequenase V.2 kit, Amersham) and evaluated in silico. A 1473 bp long cDNA corresponding to a novel human zinc finger gene, designated *ZNF232*, was assembled (Fig. 1). An internal 27 bp long coding sequence, nt positions 670–696, corresponding to nine amino acid (aa) residues, was absent in two of the analyzed clones, possibly reflecting an alternative splicing event (Figs. 1 and 2). *ZNF232* cDNA and genomic DNA (a *ZNF232*-specific PAC No. 202L17, see below) sequencing revealed the presence of three potential translation start codons at the 5' end of the gene. AUG at nt position 126 (Fig. 1) was considered the most suitable, as being in optimum accordance to the Kozak rule [11], whereas the others, located 81 (Fig. 1) and 390 bp (not shown) upstream from it, were considered less possible. In addition, an in frame translation stop codon was determined 504 bp upstream from the putative translation start codon (not shown). Moreover, our sequence data combined with recent data from GenBank (No. AC074339) and EST database (EST No. AI344299; EST No. AW952625), suggested a *ZNF232* transcription unit of at least 2.0 kb with a proposed 5' untranslated region of about 0.5 kb. The actual size of the transcript was confirmed by Northern analysis (see below). The putative 417 aa long polypeptide has a predicted molecular mass of 47.6 kDa and may be divided into two distinct regions (Figs. 1 and 2A):

the amino terminal, where a finger associated motif, the α -helical leucine-rich SCAN/LeR domain [12,13] is recognized (aa 49–130), and the carboxy terminal zinc finger region, characterized by five consecutive repeats of the C₂H₂ type (aa 278–410). A putative nucleus localization signal (NLS; aa sequence: -RHRR-) [14] is embedded in the fifth repeat (Fig. 1).

The C₂H₂ zinc finger protein family exemplified by the *Drosophila melanogaster* Krüppel gene product [15], the *Xenopus laevis* TFIIIA [16,17] and the human transcription factor Sp1 [18], constitutes one of the largest protein families in the mammalian genomes [19]. The *ZNF232* zinc finger motifs are closely related; they fit the Krüppel-type consensus [20] and are joined by the conserved S/TGEEKPY/FX linker peptide. Application of the proposed DNA-zinc finger recognition rules [21] suggests that all *ZNF232* fingers, and especially the second, the fourth and the fifth, can, in principle, participate in optimal DNA binding. The SCAN/LeR domain placed about 50 aa downstream from the amino terminus, lies in a region which is conserved in the majority of the members belonging to the respective subfamily. This domain, initially identified in human p18 [12] and *ZNF174* [13] proteins, is shared by a rapidly growing number of mammalian C₂H₂ zinc finger proteins. Although *ZNF232* appears to be overall neutral (deduced $pI=6.73$), the non-finger portion is markedly acidic ($pI=4.85$; aspartate and glutamate content: 17.3%), reminiscent of several transcription

regulators [22]. However, transient expression experiments in COS-7 cells, designed to test the possible influence of a GAL4-ZNF232 chimeric protein on the transcriptional activity of a chloramphenicol acetyltransferase (CAT) reporter gene driven by a thymidine kinase (tk) promoter containing a GAL4 binding site [23], did not result in any significant change of CAT activity (data not shown). Since the presence of the zinc fingers may interfere with chimeric protein folding, thus making a potentially ZNF232 activation domain inaccessible to the transcriptional machinery, we also designed a zinc finger deletion construct, by omitting 614 bp from the 3' end of *ZNF232* cDNA. Similarly, no significant change in CAT activity was observed. Moreover, transfection in mouse NIH-3T3 cells, under optimized conditions [13], also demonstrated no significant effects. Our results are in agreement with previous observations [12,13] and may indicate that ZNF232 is a weak transcriptional regulator or that the SCAN/LeR domain cannot act on a heterologous DNA binding site [12,13], suggesting that it is not an independent transcriptional regulatory element. Interestingly, a human gene encoding an isolated SCAN/LeR domain was reported recently [24]. Although the function of the domain is still unknown in depth, it was demonstrated that SCAN/LeR mediates hetero- and/or homotypic protein

associations and thus modulates the function of transcriptional factors [24–26].

In order to investigate whether *ZNF232* exists in a single or multiple copies in the human genome, we performed genomic DNA blot hybridization analysis. The simple hybridization pattern detected in several restriction digests (data not shown) is consistent with the existence of a unique *ZNF232* gene. To elucidate the exon/intron organization of *ZNF232*, we screened the gridded RPCI-1 human genomic PAC library (a generous gift of UK HGMP-Resource Centre, Hinxton, UK). This was performed by probing with a ³²P-labeled 730 bp *EcoRI* fragment derived from the 5' end of a *ZNF232* cDNA clone spanning nt positions 370–1100. A positive clone, No. 202L17, was isolated. Structural integrity of the clone was confirmed by DNA blot hybridization analysis in parallel with a human genomic DNA sample (data not shown). Subcloning of appropriate PAC clone fragments, sequencing, and PCR using a series of *ZNF232* cDNA-specific primers, enabled us to elucidate the genomic organization of the gene (Fig. 2). Our analysis suggested that *ZNF232* encompasses a genomic region of at least 6 kb and the translated portion of the gene spans three exons, encoding 139, 42 and 236 amino acids, respectively (Figs. 1 and 2). The putative 5' UTR of the gene is interrupted by an intron

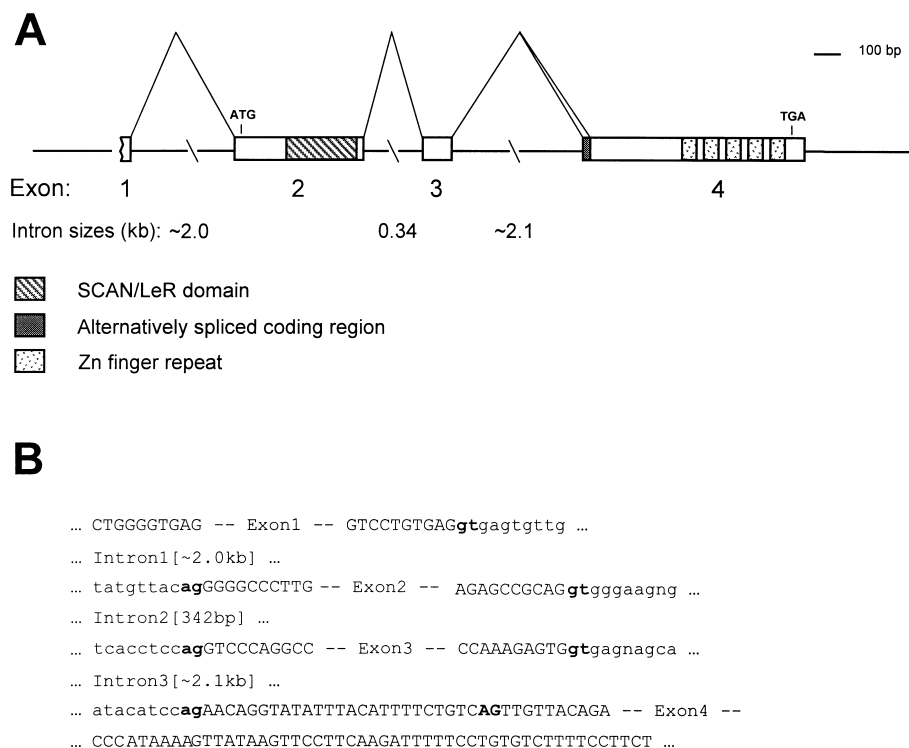


Fig. 2. Exon/intron organization and genomic sequence of *ZNF232*. (A) Schematic illustration of *ZNF232* exon/intron organization. Boxes represent *ZNF232* exons. Relative exon sizes are on scale, while the approximate intron sizes are indicated in kb. The SCAN/LeR and the zinc finger domains as well as the alternatively spliced part of exon 4 are depicted by filled boxes. The translation start and the stop codon are marked on exon 2 and 4, respectively. (B) Genomic sequences flanking *ZNF232* exons. Exons are indicated by uppercase letters; lowercase letters denote intron sequences. The splice sites, all complying with the GT/AG consensus, and the alternative acceptor site in exon 4, are in bold. Intron sizes were estimated by either PCR using primer pairs corresponding to exon sequences (introns 2 and 3), or by Southern analysis exploiting appropriate cDNA fragments as probes (intron 1). Intron 2 was sequenced to its entity.

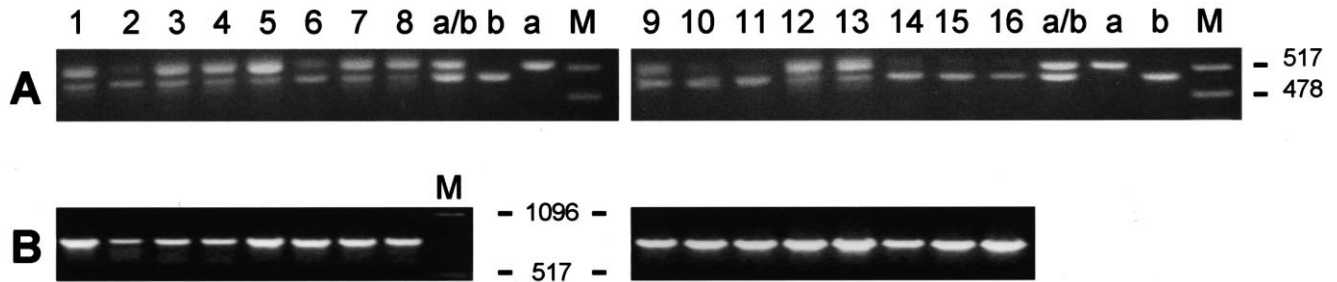


Fig. 3. Expression pattern of *ZNF232* in human tissues. (A) Each of the indicated MTC panel cDNA was used as a PCR template with a *ZNF232*-specific primer pair (see text). After 35 amplification cycles (annealing at 55°C), samples were subjected to 2.2% agarose/EtBr electrophoresis. Lanes: 1–8, heart, brain, placenta, lung, liver, skeletal muscle, kidney, pancreas, respectively; 9–16, spleen, thymus, prostate, testis, ovary, small intestine, colon, leukocytes, respectively; a, PCR product from a *ZNF232* cDNA containing the 27 bp region at nt positions 670–696; b, PCR product from a *ZNF232* cDNA lacking the 27 bp region at nt positions 670–696; a/b, co-electrophoresis of a mixture of the PCR products of lanes a and b. (B) The same as above cDNA samples (lanes 1–8 and 9–16) used as PCR templates with a control glyceraldehyde-3-phosphate dehydrogenase gene (*G3PDH*) primer pair (Clontech) after 27 amplification cycles (annealing at 55°C). Lanes M (both A and B panels), pBluescript/*Hin*I size marker.

of about 2.0 kb. The sequence encoding for the SCAN/LeR domain covers the second half of exon 2, while all five zinc finger repeats are clustered in exon 4. An AG dinucleotide located 27 bp downstream from the 5' end of exon 4 (nt position 696, Figs. 1 and 2B), may represent an appropriate alternative acceptor site, able to be utilized interchangeably by the splicing machinery in vivo. This observation correlates perfectly with the absence of a 27 bp long coding sequence (nt positions 670–696) determined in two of the analyzed *ZNF232* cDNA clones.

RNA blot hybridization analysis using 2–3 µg of poly-(A)⁺RNA from K562 cells, detected a single *ZNF232*-specific transcript of about 2.1 kb (data not shown). This is comparable with the size of the suggested *ZNF232* overall cDNA sequence. Furthermore, in order to explore the *ZNF232* expression profile in various human tissues and simultaneously investigate the possibility of *ZNF232* in

vivo alternative splicing, we employed a multiple tissue panel of normalized first-strand cDNAs (MTC, Clontech) in PCR experiments using a pair of primers (5'-ACTGTGCTGGAGGATTTAGAG-3' and 5'-ATGAACTCTCTGGTGGACAAC-3') flanking cDNA nt positions 670–696. PCR performed under non-saturating conditions (Clontech MTC manual) permitted semi-quantitative estimation of relative *ZNF232* transcript levels. The gene is expressed in all tissues tested (Fig. 3); however, slightly higher expression was observed in liver, testis and ovary. Furthermore, two transcript variants with a size difference comparable to the PCR products obtained by two *ZNF232* cDNA templates differing internally by 27 bp at nucleotide positions 670–696, were detected in all samples (Fig. 3). This suggests that *ZNF232* is subjected to alternative splicing in vivo, a feature shared by other members of the SCAN/LeR subfamily [27–29].

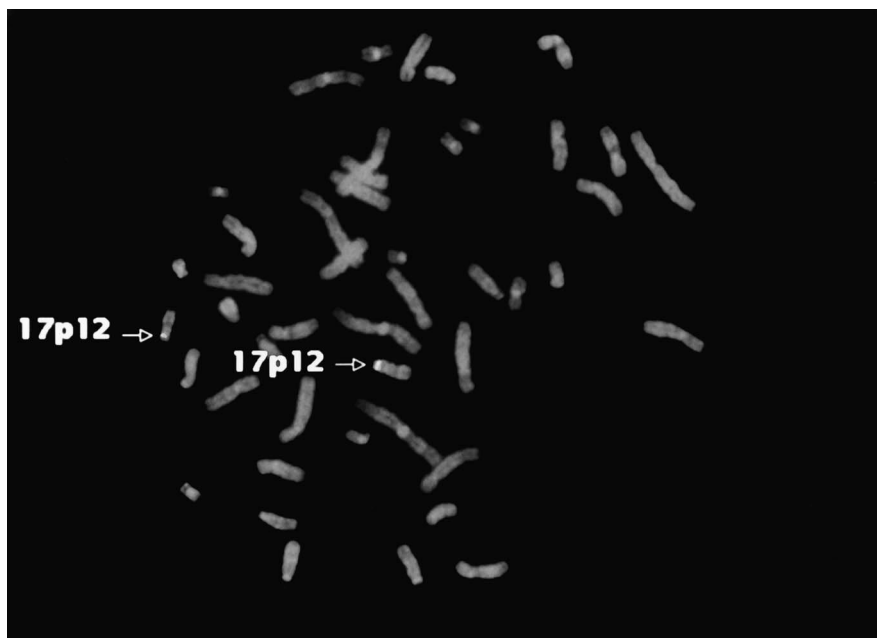


Fig. 4. Fluorescence in situ hybridization on human male metaphase spreads using PAC clone RPCI-1 No. 202L17 as a probe. The unique signal at 17p12 is denoted by an arrow.

Interestingly, the relative abundance of the two splice variants differed among tested samples (Fig. 3), indicating a tissue-specific post-transcriptional control for *ZNF232*. It remains to be seen whether these alternatively spliced transcripts suggesting two *ZNF232* isoforms, differing internally by nine amino acids, are of functional significance.

Chromosomal assignment of *ZNF232* was initially performed using a monochromosomal human-rodent somatic cell hybrid panel (Coriell Institute of Medical Research). DNA samples (50–250 ng) from each hybrid and comparable amounts from the respective human and rodent parental cell lines were used as templates in PCR amplifications with a *ZNF232*-specific primer pair. The experiment suggested that *ZNF232* maps to chromosome 17 (data not shown). Subsequent FISH analysis with fluorescence labeled PAC No. 202L17 as a probe, performed to normal metaphase chromosome spreads according to standard high resolution procedures [30,31], resulted in a strong hybridization signal at 17p12 (Fig. 4). Database searches revealed that a partially sequenced BAC clone (GenBank accession No. AC012146) roughly mapped to chromosome 17, encompasses *ZNF232*. Our FISH data place this BAC to 17p12. The chromosomal distribution of human genes encoding for zinc finger proteins is clearly non-random [32,33]. A disproportionate density characterizes chromosomes 1, 3, 6, 7, 10, 11, 22 and an excessive abundance marks chromosome 19 (Genome Database; www.gdb.org). Moreover, available mapping data for members of the human SCAN/LeR group reveal an apparently clustered genomic organization. The majority of the reported members are mapped to certain chromosomal regions, i.e. 3p21-22, 6p21-22 and 16p13 [34–37]. Apart from *ZNF232*, at least four other zinc finger genes of the C₂H₂ type map to 17p12-p13, i.e. *ZNF18/KOX11*, *ZNF29/KOX26*, *ZNF62* and *ZFP3*, thus pinpointing another zinc finger gene-rich region in the human genome. Structural abnormalities of C₂H₂ zinc finger genes have been found in a variety of genetic diseases. In particular, deletions at 17p12-13 have been associated with gliomas, astrocytomas, colon, and lung tumors [32,38].

Finally, in order to investigate experimentally the sub-

cellular localization of the *ZNF232* polypeptide and the functional role of the predicted NLS, we proceeded in transient expression of a green fluorescent protein-*ZNF232* chimera, in african green monkey COS-7 cells. The construct, pEGFP/*ZNF232*, contained the enhanced green fluorescent protein (GFP) open reading frame combined with the *ZNF232* coding region. This was prepared by co-ligation of the linearized pEGFP-C1 vector (Clontech) with a PCR product (primers: 5'-AGGATGGCTGTATCACTAAC-3' and 5'-GATCCAGTCCTAAAGTAGATTAGAC-3', nt positions 123–142 and 1418–1444, respectively) representing the entire *ZNF232* translated region; the amplified DNA was inserted into the blunt-ended *Hind*III site of the vector polylinker. A deletion derivative, pEGFP/del*ZNF232*, lacking the NLS region was prepared by excision of a *Kpn*I fragment defined by the unique internal (nt position 830, Fig. 1) and the vector *Kpn*I sites (expressed *ZNF232* polypeptide: 1–235 aa). About 10⁶ COS-7 cells were transfected with 2 µg of either pEGFP/*ZNF232* or pEGFP/del*ZNF232* DNA using the calcium phosphate co-precipitation method [39]. Twenty-four to 36 h after transfection, the cells were examined with fluorescence microscopy. The chimeric GFP/*ZNF232* polypeptide was detected specifically in the nucleus; on the contrary, expression of pEGFP/del*ZNF232* resulted in the exclusion of the truncated polypeptide from the nucleus (Fig. 5). These findings suggest that *ZNF232* is a nuclear protein and demonstrate the role of the predicted NLS.

To summarize, this study reports the detailed analysis of a novel chromosome 17p12 zinc finger gene, *ZNF232*, member of the SCAN/LeR subfamily. *ZNF232* is widely expressed in man, encodes a nuclear protein and is subjected to tissue-specific post-transcriptional regulation. The latter may lead to two structural variants differing in relative abundance among tissues. Like other members of the group, *ZNF232* may be associated with the transcriptional machinery in a wide variety of human tissues. It would be interesting to investigate the potential of *ZNF232* as a transcriptional modulator as well as to explore its interaction repertoire.

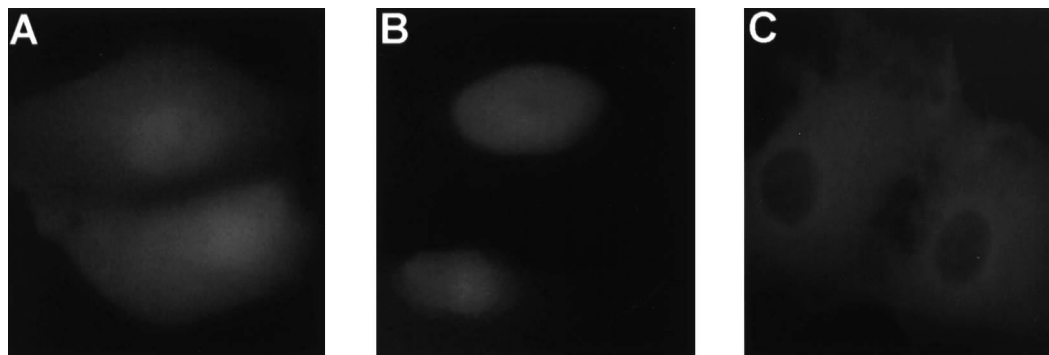


Fig. 5. Fluorescence photomicrographs showing the subcellular localization of GFP-*ZNF232* chimeras in transiently transfected COS-7 cells. Plasmid DNA used: (A) pEGFP-C1 vector (control); the GFP product exhibited passive nucleocytoplasmic diffusion, as expected [40]; (B) pEGFP/*ZNF232*; (C) pEGFP/del*ZNF232*.

The authors wish to thank S. Vrontou for her contribution in the initial screening of the cDNA library, A. Paspaparakis for technical support and Dr. G. Chalepakis for critical reading of the manuscript. This work was supported by grants of the University of Crete and IMBB-FORTH.

References

- [1] V.A. McKusick, *Genomics* 45 (1997) 244–249.
- [2] P. Deloukas, G.D. Schuler, G. Gyapay, E.M. Beasley, C. Soderlund, P. Rodriguez-Tome, L. Hui, T.C. Matisse, K.B. McKusick, J.S. Beckmann, S. Bentolila, M. Bihoreau, B.B. Birren, J. Browne, A. Butler, A.B. Castle, N. Chiannikulchai, C. Clee, P.J. Day, A. Dehejia, T. Dibling, N. Drouot, S. Duprat, C. Fizames, D.R. Bentley et al., *Science* 282 (1998) 744–746.
- [3] M.F. Bonaldo, G. Lennon, M.B. Soares, *Genome Res.* 6 (1996) 791–806.
- [4] S. Henikoff, E.A. Greene, S. Pietrokovski, P. Bork, T.K. Attwood, L. Hood, *Science* 278 (1997) 609–614.
- [5] R.L. Tatusov, E.V. Koonin, D.J. Lipman, *Science* 278 (1997) 631–637.
- [6] M.L. Holmes, J.D. Haley, L. Cerruti, W.L. Zhou, H. Zogos, D.E. Smith, J.M. Cunningham, S.M. Jane, *Mol. Cell. Biol.* 19 (1999) 4182–4190.
- [7] L. Pirkkala, T.P. Alastalo, P. Nykanen, L. Seppa, L. Sistonen, *FASEB J.* 13 (1999) 1089–1098.
- [8] R.P. de Groot, J.A. Raaijmakers, J.W. Lammers, R. Jove, L. Koenderman, *Blood* 94 (1999) 1108–1112.
- [9] H. Asano, X.S. Li, G. Stamatoyannopoulos, *Mol. Cell. Biol.* 19 (1999) 3571–3579.
- [10] A.T. Puzyrev, K. Chroniary, N.K. Moschonas, *Mol. Biol. (Mosk.)* 29 (1995) 97–103.
- [11] M. Kozak, *Mamm. Genome* 7 (1996) 563–574.
- [12] G. Pengue, V. Calabrò, P. Cannada-Bartoli, A. Pagliuca, L. Lania, *Nucleic Acids Res.* 22 (1994) 2908–2914.
- [13] A.J. Williams, L.M. Khachigian, T. Shows, T. Collins, *J. Biol. Chem.* 270 (1995) 22143–22152.
- [14] D. Goerlich, I.W. Mattaj, *Science* 271 (1996) 1513–1518.
- [15] A. Preiss, U.B. Rosenberg, A. Kienlin, E. Seifert, H. Jackle, *Nature* 313 (1985) 27–32.
- [16] J. Miller, A.D. McLachlan, A. Klug, *EMBO J.* 4 (1985) 1609–1614.
- [17] R.S. Brown, C. Sander, P. Argos, *FEBS Lett.* 186 (1985) 271–274.
- [18] J.T. Kadonaga, K.R. Carner, F.R. Masiarz, R. Tjian, *Cell* 51 (1987) 1079–1090.
- [19] E.J. Bellefroid, P.J. Lecocq, A. Benhida, D.A. Poncelet, A. Belayew, J.A. Martial, *DNA* 8 (1989) 377–387.
- [20] R. Schuh, W. Aicher, U. Gaul, S. Cote, A. Preiss, D. Maier, E. Seifert, U. Nauber, C. Schroder, R. Kemler et al., *Cell* 47 (1986) 1025–1032.
- [21] M. Suzuki, M. Gerstein, N. Yagi, *Nucleic Acids Res.* 22 (1994) 3397–3405.
- [22] M. Carey, Y.S. Lin, M.R. Green, M. Ptashne, *Nature* 345 (1990) 361–364.
- [23] I. Sadowski, M. Ptashne, *Nucleic Acids Res.* 17 (1989) 7539.
- [24] C. Schumacher, H. Wang, C. Honer, W. Ding, J. Koehn, Q. Lawrence, C.M. Coulis, L.L. Wang, D. Ballinger, B.R. Bowen, S. Wagner, *J. Biol. Chem.* 275 (2000) 17173–17179.
- [25] A.J. Williams, S.C. Blacklow, T. Collins, *Mol. Cell. Biol.* 19 (1999) 8526–8535.
- [26] T.L. Sander, A.L. Haas, M.J. Peterson, J.F. Morris, *J. Biol. Chem.* 275 (2000) 12857–12867.
- [27] M.J. Peterson, J.F. Morris, *Gene* 254 (2000) 105–118.
- [28] J.F. Prost, D. Negre, F. Cornet-Javaux, J.C. Cortay, A.J. Cozzone, D. Herbage, F. Mallein-Gerin, *Biochim. Biophys. Acta* 1447 (1999) 278–283.
- [29] C. Monaco, M. Helmer Citterich, E. Caprini, I. Vorechovsky, G. Russo, C.M. Croce, G. Barbanti-Brodano, M. Negrini, *Genomics* 52 (1998) 358–362.
- [30] D.E. Rooney, B.H. Czepulkowski, *Human Cytogenetics I: Constitutional Analysis*, Oxford University Press-IRL Press, New York, 1992, pp. 70–74.
- [31] P.C. Patsalis, M.I. Hadjimarcou, V. Velissariou, S. Kitsiou-Tzeli, C. Zera, M. Syrrou, E. Lyberatou, A. Tsezou, A. Galla, N. Skordis, *Clin. Genet.* 51 (1997) 184–190.
- [32] K. Huebner, T. Druck, C.M. Croce, H.J. Thiesen, *Am. J. Hum. Genet.* 48 (1991) 726–740.
- [33] J.M. Hoovers, M. Mannens, R. John, J. Bliiek, V. van Heyningen, D.J. Porteous, N.J. Leschot, A. Westerveld, P.F. Little, *Genomics* 12 (1992) 254–263.
- [34] G. Pengue, V. Calabrò, P. Cannada-Bartoli, P. De Luca, T. Esposito, P. Taillon-Miller, S. LaForgia, T. Druck, K. Huebner, M. D’Urso, L. Lania, *Hum. Mol. Genet.* 2 (1993) 791–796.
- [35] R.M. Attar, M.Z. Gilman, *Mol. Cell. Biol.* 12 (1992) 2432–2443.
- [36] P.L. Lee, T. Gelbart, C. West, M. Adams, R. Blackstone, E. Beutler, *Genomics* 43 (1997) 191–201.
- [37] X. Chen, M. Hamon, Z. Deng, M. Centola, R. Sood, K. Taylor, D.L. Kastner, N. Fischel-Ghodsian, *Biochim. Biophys. Acta* 1444 (1999) 218–230.
- [38] M.F. Rousseau-Merck, K. Huebner, R. Berger, H.J. Thiesen, *Genomics* 9 (1991) 154–161.
- [39] F. Graham, A. van der Eb, *Virology* 52 (1973) 456–467.
- [40] M.A. Powers, D.J. Forbes, *Cell* 79 (1994) 931–934.

Map integration at human chromosome 10: molecular and cytogenetic analysis of a chromosome-specific somatic cell hybrid panel and genomic clones, based on a well-supported genetic map

R. Marzella,^a M.A. Kokkinaki,^{b,c} M. Kapsetaki,^b A. Ricco,^a A. Argyrokastritis,^b S. Kamakari,^b T. Sarafidou,^{b,c} N. Archidiacono,^a A. Roussou,^c A. Paspaparakis,^b M. Rocchi,^a and N.K. Moschonas^{b,c}

^aIstituto di Genetica, Bari (Italy),

^bInstitute of Molecular Biology and Biotechnology, Foundation of Research and Technology, and

^cDepartment of Biology, University of Crete, Crete (Greece)

Abstract. Well-characterized, chromosome-specific somatic cell hybrid panels are powerful tools for the analysis of the human genome. We have characterized a panel of human × hamster somatic cell hybrids retaining fragments of human chromosome 10 by fluorescence in situ hybridization and associated them to genetic markers. Most of the hybrids were generated by the radiation-reduction method, starting from a chromosome 10-specific monochromosomal hybrid, whereas some were collected from hybrids retaining chromosome 10-specific fragments as a result of spontaneous in vitro rearrangements. PCR was used to score the retention of 57 microsatellite markers evenly distributed along a well-supported framework genet-

ic map containing 149 loci uniquely placed at 69 anchor points (odds exceeding 1,000:1), with an average spacing of 2.8 cM. As an additional resource for genomic studies involving human chromosome 10, we report the cytogenetic localization of a series of YAC and PAC clones recognized by at least one genetic marker. Somatic cell hybrids provide a powerful source of partial chromosome paints useful for detailed clinical cytogenetic and primate chromosome evolution investigations. Furthermore, correlation of the above physical, genetic, and cytogenetic data contribute to an emerging consensus map of human chromosome 10.

Human chromosome 10 is a medium-size, submetacentric chromosome corresponding to about 4.6% of the length of the autosomes. To date, more than 40 genetic diseases have been associated with chromosome 10 loci; about 165 genes and more than 880 partial and full-length cDNAs (<http://www.ncbi.nlm.gov/>

[SCIENCE96/](http://www.ncbi.nlm.gov/SCIENCE96/)) have been mapped to this chromosome. Several disease-causing genes have been identified on chromosome 10 and are under intensive investigation, including the RET proto-oncogene responsible for multiple endocrine neoplasia and Hirschsprung disease (see, e.g., Eng et al., 1997) and fibroblast growth factor receptor 2, which has been associated with Apert, Jackson-Weiss, Pfeiffer, Crouzon, and Beare-Stevenson cutis gyrata syndromes (Jabs et al., 1994; Reardon et al., 1994; Rutland et al., 1995; Wilkie et al., 1995; Przylepa et al., 1996). However, a number of other disease-related loci on chromosome 10, including several involved in developmental abnormalities, cancer, and neurological diseases (e.g., split hand/foot malformation type 3 [Nunes et al., 1995], DiGeorge II and velocardiofacial syndromes [Daw et al., 1996], endometrial carcinoma [Nagase et al., 1997], and infantile onset of spi-

Supported in part by the European Gene Mapping Project (EC EUROGEN), a Telethon grant, and AIRC, and by the Hellenic Secretariat for Research and Technology. R.M. and M.A.K. have equally contributed to the project.

Received 21 July 1997; revision accepted 12 January 1998.

Request reprints from Dr. Nicholas K. Moschonas, Institute of Molecular Biology and Biotechnology, Foundation of Research and Technology, PO Box 1527, Heraklion 711 10, Crete (Greece); telephone: 30-81-394-357; fax: 30-81-394-408; e-mail: moschon@imbb.forth.gr.

nocerebellar atrophy [Nikali et al., 1997]), still remain unidentified.

Positional cloning of disease-related genes and advanced molecular and cytogenetic diagnostic tools depend upon detailed integrated chromosome maps and the availability of well-characterized probes with precisely defined genetic, cytogenetic, and physical locations. Such resources, recently used for the construction of high-density, whole genome genetic and physical maps, include collections of microsatellite markers (Dib et al., 1996), YAC clones (Chumakov et al., 1995), and radiation hybrids (e.g., Hudson et al., 1995; Gyapay et al., 1996). To date, available chromosome 10-specific resources include λ phage (Deaven et al., 1986) and cosmid flow-sorted libraries (Ma et al., 1996) and a large collection of cytogenetically mapped YAC and PAC clones (Moschonas et al., 1996; Meitinger et al., 1997).

The radiation hybrid technique (Goss and Harris, 1975) was designed initially as a tool for physical mapping by ordering genetic markers according to the frequency of radiation-induced breakages between them. The numerous highly informative genetic markers available today permit detailed characterization of radiation hybrids, leading to adequate integration of the genetic and physical maps of human chromosomes. Some laboratories have produced chromosome-specific radiation hybrids derived from a lethally irradiated human \times rodent monochromosomal hybrid and fused to a second rodent cell line (Francke et al., 1994). In the present study, we report the construction of a well-supported genetic linkage map of human chromosome 10 and the detailed cytogenetic and molecular characterization of a chromosome 10-specific panel of 21 somatic cell hybrids retaining specific fragments of chromosome 10 by fluorescence in situ hybridization (FISH) and by the use of a set of microsatellite markers evenly distributed along the genetic map. As an example of the applications that such a panel may have in cytogenetic studies, we report the characterization of a paracentric inversion that differentiates human chromosome 10 and its corresponding homolog in the orangutan.

Materials and methods

Characterization of novel microsatellites and YAC library screening

Two Charon 21A-phage libraries (LA10NS01 and LL10NS01), constructed from a chromosome 10-specific human \times hamster somatic cell hybrid (ATCC), were screened with $5'$ - 32 P-labeled (GATA) $_{10}$, (AAAT) $_{10}$, or (CA) $_{20}$ oligonucleotides. A monochromosomal somatic cell hybrid panel (Corriell Cell Repositories Mapping Panel #2) was used in Southern blots to confirm the chromosomal origin of the positive clones. DNA sequencing was performed with either universal primers or complementary primers corresponding to the repeat, using conventional methodologies. Primers flanking each repeat were designed and employed in PCRs using as templates 80 Centre d'Etude du Polymorphisme Humain (CEPH) parental DNA samples. Genotyping of the 40 CEPH families was performed either manually or by using a Pharmacia A.L.F. DNA sequencer. Genotypic data evaluation was performed using ALP version 1.00 software (A. Wright and A. Brown, personal communication). DNA sequencing and genotypic data for nine markers were submitted to EMBL/GenBank database and GDB, respectively. Primer pairs corresponding to each marker were used for PCR screening of the ICI YAC library (supplied by the United Kingdom Human Gene Mapping Project [UK HGMP] Resource Center) in three successive steps, according to the instructions of the Resource Center.

Genetic linkage analysis

Genotypic data generated locally for 36 markers (D10S249, D10S92, D10S33, D10S2314, D10S179, D10S189, D10S547, D10S191, D10S211, D10S89, D10S197, D10S193, GLUDP5, D10S141, ZNF22, GLUDP2, D10S220, D10S589, D10S210, D10S535, D10S109, D10S201, D10S608, D10S2311, D10S2316, D10S192, D10S597, D10S2330, D10S2312, D10S190, D10S186, D10S212, D10S2318, D10S2309, D10S2317, and D10S2310) by screening the 40 CEPH families (Kapssetaki et al., 1994; Cox et al., 1996; this study) were submitted electronically to the European Gene Mapping (EUROGEM) ftp server (Cox et al., 1996) and subsequently passed on to CEPH. For the construction of a genetic map of human chromosome 10, our genotypic data were combined with data (for 117 markers) retrieved from the CEPH version 6 and CHLC version 2 databases, as described by Cox et al. (1996). Linkage analysis was performed by the CRI-MAP version 2.4 program essentially as described by Spurr et al. (1994), using a criterion of at least 1,000:1 odds. Error checking was performed by the CROMPIC option of the program, and potential data errors for markers genotyped locally were checked by retyping or removed from the analysis according to certain guidelines (Cox et al., 1996).

Cell lines

The normal human \times hamster somatic cell hybrids retaining fragments of human chromosome 10 were first reported by Antonacci et al. (1995). Hybrid GM10926D, which retained this chromosome as the only human contribution, was obtained from the Coriell Institute. The orangutan (*Pongo pygmaeus pygmaeus*, PPY) lymphoblastoid cell line was kindly donated by Dr. R.E. Bontrop (The Netherlands).

Radiation hybrids

A 75-cm² flask containing a subconfluent culture of the GM10926D cell hybrid was gamma-irradiated at low dosage (5,000 rad) to obtain hybrids with medium-to-large fragments that would be particularly suited to cytogenetic analysis. After exposure, the cell line was fused to a thymidine kinase-deficient hamster cell line (B14-150, Flow Laboratories) and selected in HAT medium. The hybrid clones were transferred to 24-well plates and subjected to trypsinization at near confluence. An aliquot was seeded in an eight-well chamber slide (Nunc), and the remaining cells were kept in the well. The chamber slides were harvested and hybridized in situ with biotin-labeled total human DNA. The signal was detected using Cy3-conjugated avidin (Amersham). FISH analysis provided information on the presence of human material, on its approximate size, and, particularly, on the homogeneity of the clone. Clones with less than 50% positive nuclei were discarded. Each positive clone was then expanded in a 75-cm² flask for liquid nitrogen storage and in a 150-cm² flask for DNA extraction. Hybrid PCP3B10 was a gift from P. Goodfellow (Cambridge).

Alu-PCR probes

PCR amplification of cell hybrid DNA templates was performed using the following *Alu* primers: 5'-GGA TTA CAG GYR TGA GCC A-3' and 5'-RCC AYT GCA CTC CAG CCT G-3' (where Y = C/T and R = A/G) (Liu et al., 1993). From 50 to 100 ng of DNA was amplified in a volume of 50 μ l. Following an initial denaturation step of 5 min at 95°C, 30 cycles of amplification (1 min at 95°C, 1 min at 65°C, and 4 min at 72°C) were performed, with a final extension step of 10 min at 72°C. A series of conventional *Alu* primers was used for the preparation of YAC-specific probes (Nelson et al., 1989). A mixture of *Alu*-PCR products corresponding to ICI YAC-specific sequences were used as probes for filter hybridization screening of a gridded PAC library (Ioannou et al., 1994), supplied by the UK HGMP Resource Center. Interpretation of the hybridization results was performed according to the scheme provided by the Resource Center.

FISH analysis of the hybrids and the YAC clones

Human metaphase spreads were obtained from PHA-stimulated peripheral blood lymphocytes from a normal human donor, hybridized in situ with *Alu*-PCR products biotinylated by nick translation and detected by Cy3-conjugated avidin. Hybridization was performed essentially as described by Archidiacono et al. (1994). Briefly, 400 ng of labeled DNA from each hybrid were used for each slide; hybridization was performed at 37°C in 2 \times SSC, 50% (v/v) formamide, 10% (w/v) dextran sulfate, 5 μ g Cot-I DNA (Boehringer Mannheim), and 3 μ g sonicated salmon sperm DNA, in a volume of 10 μ l. Posthybridization washing was carried out at 42°C in 2 \times SSC, 50%

formamide (three washes), followed by three additional washes in $0.1 \times \text{SSC}$ at 60°C . In situ hybridization of YAC probes was performed similarly, using 300–500 ng of biotin-labeled yeast DNA. Alternatively, chromosome spreads banded with 4',6-diamidino-2-phenylindole (DAPI) were hybridized to biotin-labeled yeast DNA or *Alu*-PCR products of the clones, essentially as described by Knapp et al. (1995). The signal was detected with FITC-conjugated avidin. Chromosome 10 was identified by cytological banding using DAPI and by cohybridization to a control plasmid clone, D10Z1, which detected chromosome 10-specific repetitive centromeric sequences (ATCC).

Digital imaging

Digital images were obtained using a Leica DMRXA epifluorescence microscope equipped with a CCD camera (Princeton Instruments). The fluorescent Cy3, FITC, and DAPI signals, detected with specific filters, were recorded separately as grayscale images. Pseudocoloring and merging of images were performed using Adobe Photoshop software.

Results

Isolation, characterization, and FISH mapping of novel human chromosome 10 microsatellite markers

To isolate novel polymorphic microsatellites of human chromosome 10, we screened two chromosome 10-specific phage libraries. Screening of about 100,000 plaques, yielded 36 candidate marker loci of the relatively frequent $(\text{CA})_n$ and the rare $(\text{GATA})_n$ and $(\text{AAAT})_n$ repeat type (26, 7, and 3 clones, respectively). The human origin of the positives was tested by differential hybridization to human vs. hamster genomic DNA and their human chromosome 10 origin, by Southern analysis of a somatic cell hybrid panel. Nine microsatellites, viz., D10S2311, D10S2312, D10S2314, and D10S2330 of the $(\text{GATA})_n$ repeat type, D10S2309 and D10S2310 of the $(\text{AAAT})_n$ type, and D10S2316, D10S2317, and D10S2318 of the $(\text{GT})_n$ repeat type, were selected for further analysis. To evaluate their genetic informativeness, we genotyped the 80 CEPH parental DNA samples. Allele numbers ranged from 5 to 11. The observed heterozygosities varied from 0.19 to 0.81. Codominant segregation was observed in all CEPH families genotyped. To determine the cytogenetic localization of the markers, ICI YACs corresponding to each of the markers, were isolated by PCR screening (a total of 21 YACs), and the corresponding *Alu*-PCR products were used in FISH experiments (Table 1). One marker (D10S2314) was mapped to the subtelomeric region of 10p, another one (D10S2318) to 10p12, and the rest to 10q23→q25.

Construction of a genetic map of human chromosome 10

A chromosome 10 genetic linkage map (Fig. 1) was constructed by employing 153 polymorphic loci with an average heterozygosity of 0.7. Genotypic data retrieved from the CEPH and CHLC databases (for 32 and 85 markers, respectively), were combined with data (for 36 markers) generated locally. The map (EUROGEM map) has a total sex-average length of 194.7 cM (male length, 148.2 cM; female length, 243.3 cM) and consists of 69 framework markers (viz., 40 haplotype systems and 29 individual markers) ordered with a likelihood support of at least 1,000:1 (Fig. 1). The average marker spacing on the sex-average map is 2.8 cM. The map extends from D10S249 (pter) to D10S6 (qter). Increased recombination was observed on the distal long arm of chromosome 10 in females and on the

distal short arm in males (Fig. 1). This finding is not unusual for several human chromosomes (Spurr et al., 1994); however, the molecular and biological etiology of this phenomenon remains to be resolved. Four markers which could not be ordered uniquely with odds of greater than 1,000:1 (D10S2318, D10S2309, D10S2317, and D10S2310) were positioned approximately.

FISH and molecular characterization of the somatic cell hybrid panel

Figure 2 summarizes the results of the cytogenetic and molecular characterization of a set of somatic cell hybrids retaining fragments of human chromosome 10. Since the distribution of *Alu* repeats along the human chromosomes correlates with G-negative bands (Korenberg and Rykowski, 1988), inter-*Alu* PCR products do not produce a homogeneous staining but, rather, an R-banding pattern (Baldini and Ward, 1991). Consequently, FISH characterization of fragments with breakpoints occurring in very pale regions could be inaccurate. In addition, small chromosomal deletions are not detectable by FISH because of the chromatin condensation in metaphase chromosomes. Thus, in order to characterize the hybrids at the molecular level and to overcome the limits of the FISH method, we employed 57 microsatellite markers chosen from those used for the construction of the genetic map. Selection of the markers was based on their informativeness and their nearly even distribution on the genetic map. A relatively larger number of markers mapped to the disease-gene-rich region 10q23→q25 was employed. PCR screening of 21 hybrid DNA samples was performed at least twice. In parallel, a positive (human DNA) and a negative (hamster DNA) control was always used. Some regions detected by the markers were not detected by FISH and vice versa, as indicated in Fig. 2. To enrich the physical and cytogenetic mapping resources available for human chromosome 10, we also characterized by FISH a set of CEPH YACs, in addition to those from the ICI library (Table 1). Table 1 also includes markers found by PCR to recognize these YACs, so that linkage and physical mapping data can be correlated. All but one YAC (886-F-3) recognized at least one human chromosome 10 marker. In addition, for YACs mapped to the disease-gene-rich 10q23→q25 region, we isolated several PAC clones recognized by at least one marker participating in our genetic map.

A paradigm of partial chromosome paints use in primate cytogenetic evolution

Partial chromosome paints (PCPs) generated from these hybrids are a powerful tool for cytogenetic investigations, particular for investigating the cytogenetic evolution of primates (Muller et al., 1996). Figure 3 presents an example of the combined use of PCPs and YACs in the fine characterization of a paracentric inversion present in orangutan species. It has been shown that the long arm of orangutan chromosome 7, which is syntenic to human chromosome 10, exhibits a banding pattern different from that of human 10q, suggesting the existence of a paracentric inversion (Yunis and Prakash, 1982). For a detailed characterization of this inversion, we performed cohybridization experiments, using appropriate human chromo-

Table 1. Regional assignment of human chromosome 10-specific YAC and PAC clones and the corresponding markers/STSs

YACs	Markers	Chromosome location	YACs	Markers	Chromosome location ^a	PACs
912-B-3	D10S249	10pter	821-D-2	D10S541	10q23	
946-G-12	D10S533	10p15	12HA1	D10S2309	10q23	
	D10S559			D10S1753		181-E18, 109-L90, 31-J24, 48-P11
	D10S602		24DE9	D10S2309	10q23	
	D10S1745			D10S1753		
	D10S1716		647-G-1	D10S564	10q23	
	D10S1706		8EH2	D10S2311	10q23.3→q24	
10GA2	D10S2314	10pter		D10S564		
10GA7	D10S2314	10pter	10BA1	D10S2311	10q23.3→q24	172-M13, 63-O24, 27-H17, 241-O11
81E1	D10S2314	10pter		D10S1753		23-G1, 135-E5
809-F-9	D10S1729	10p15		D10S564		27-H17, 241-O11, 82-C17, 105-N10, 216-P14, 74-P12
	D10S591		32GE7	D10S2311	10q23.3→q24	
	D10S552			D10S564		
851-H-9	D10S547	10p14	37IF9	D10S2311	10q23.3→q24	
747-H-8	D10S547	10p14		D10S564		
	D10S585			D10S1753		
808-A-2	D10S223	10p14	11DG4	D10S2316	10q23	
	D10S1725		22FG5	D10S2316	10q23	
886-F-3		10p13	3HH7	D10S2317	10q23	
959-G-10	D10S191	10p12	878-F-11	D10S566	10q24	
958-C-9	D10S191	10p12		D10S530		
33GB2	D10S2318	10p12	27BC9	D10S597		30-C6
36HC3	D10S2318	10p12		D10S2330	10q24	30-C6, 171-A21, 171-F5, 75-B20, 119-I13, 317-G21
940-F-9	D10S548	10p12	35GE1	D10S597		30-C6
934-E-11	D10S1660	10p12		D10S2330	10q24	30-C6, 171-A21, 171-F5, 75-B20, 119-I13, 317-G21
	D10S1747		27DD2	D10S2312	10q24→q25.1	32-B12, 320-F13, 59-L4, 241-L17, 84-E23
	D10S1749		14CB8	D10S2312	10q24→q25.1	
875-B-4	D10S197	10p12	19FC12	D10S2312	10q24→q25.1	
	D10S1775		24AC1	D10S2312	10q24→q25.1	
751-D-5	D10S1771	10p11.23→p12.1	22DC2	D10S2310	10q25.1	
	D10S588		966-H-9	D10S543	10q25	
953-E-3	D10S1732	10p11	935-F-11	D10S1748	10q25	
	D10S1684			D10S1657		
	D10S204			D10S544		
	D10S213			D10S545		
	D10S224		937-A-6	D10S209	10q25	
686-D-11	D10S224	10p11	934-C-4	D10S575	10q26	
945-A-12	D10S224	10p11		D10S214		
	D10S193		932-F-11	D10S1727	10q26	
	D10S1674			D10S217		
754-D-10	D10S1781	10p11	<u>799-E-5</u>		10q11.2/ 10q22→q23	
798-C-7	D10S208				10q23/10q11.2	
837-B-5	D10S578	10p11	<u>738-B-12</u>	D10S1687	10q23/10q11.2	
895-D-3	D10S1783	10q11	<u>889-C-10</u>	D10S558	10p15/5p15	
	D10S604			D10S594		
933-A-3	D10S196	10q11	<u>764-E-4</u>	D10S203	10p12/8p23	
	D10S225		<u>760-C-2</u>	D10S593	10p11.2/1q43/ 19p13.1	
	D10S1724				10q21/12q21?	
885-D-10	D10S568	10q11	<u>928-B-3</u>	D10S1767	10q21/12q21?	
905-G-8	D10S599	10q21	<u>665-D-2</u>	D10S556	10q22/7q34	
781-F-5	D10S210	10q21	<u>955-A-4</u>	D10S607	10q23/Yq11.22	
779-E-7	D10S1688	10q21	<u>968-E-6</u>	D10S541	10q23/6q21	
876-B-5	D10S218	10q22	<u>24DE4</u>	D10S2309	10q23/Chr 6	
	D10S188		<u>9CH9</u>	D10S2316	10q23/Chr 4	
804-H-2	D10S202	10q22	<u>649-H-1</u>		10q24/ 2q35→q36.3	
807-D-4	D10S551	10q23			10q24/ centromeres	
	D10S1686		<u>10AE2</u>	D10S2330	10q24→q25.1/ Chr 2	
	D10S1689		<u>38HG5</u>	D10S2312	10q26.1/ multiple sites	
			<u>771-A-11</u>			

^a Fifteen YACs (shown underscored in the lower right half of the table) mapped to more than one chromosome and possibly represent chimeric clones. FISH images of most of the YACs can be viewed at <http://bioserver.uniba.it/fish/rocchi/welcome.html>.

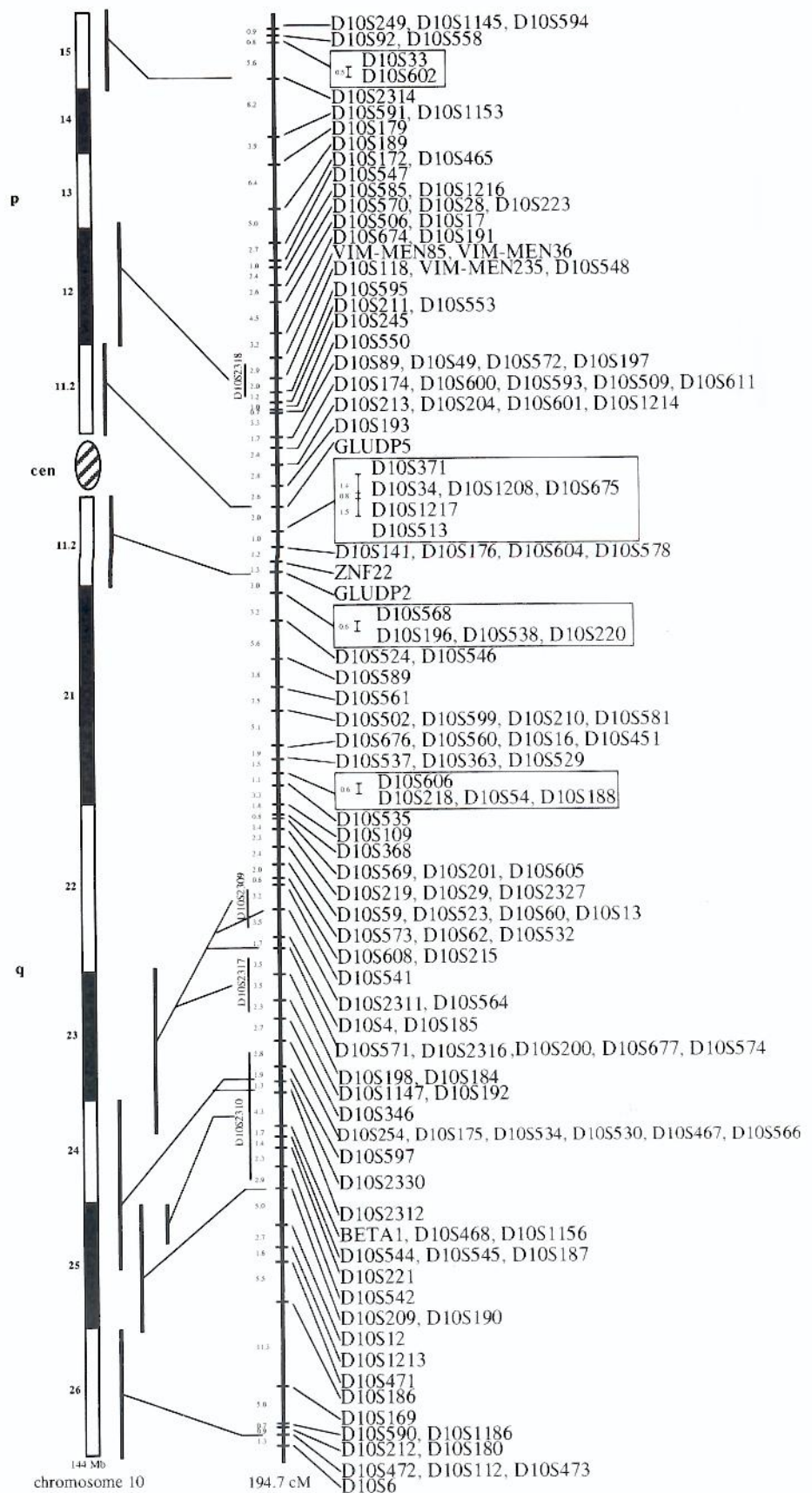


Fig. 1. A framework genetic linkage map of human chromosome 10. The sex-average map and the telomeric regions of the female and male maps, where significantly skewed recombination distances were determined, are presented. The distances between markers are shown to scale. The support intervals of four markers that were placed approximately are indicated by vertical lines. Markers haplotyped together are shown on the same line or as framed groups. The thin vertical lines to the left of framed markers indicate groups of loci for which the order could not be resolved with odds exceeding 1,000:1. The diagram at the left is an ideogram of chromosome 10 showing the cytogenetic localization of 13 of the markers used (Deloukas et al., 1993; Tzimogiorgis et al., 1993; this study). An earlier version of this map is presented in GDB (GDB No. 4584969). The male and female maps are accessible at: <http://www.imbb.forth.gr/papers/malemap.jpg> and <http://www.imbb.forth.gr/papers/femalemap.jpg>, respectively.

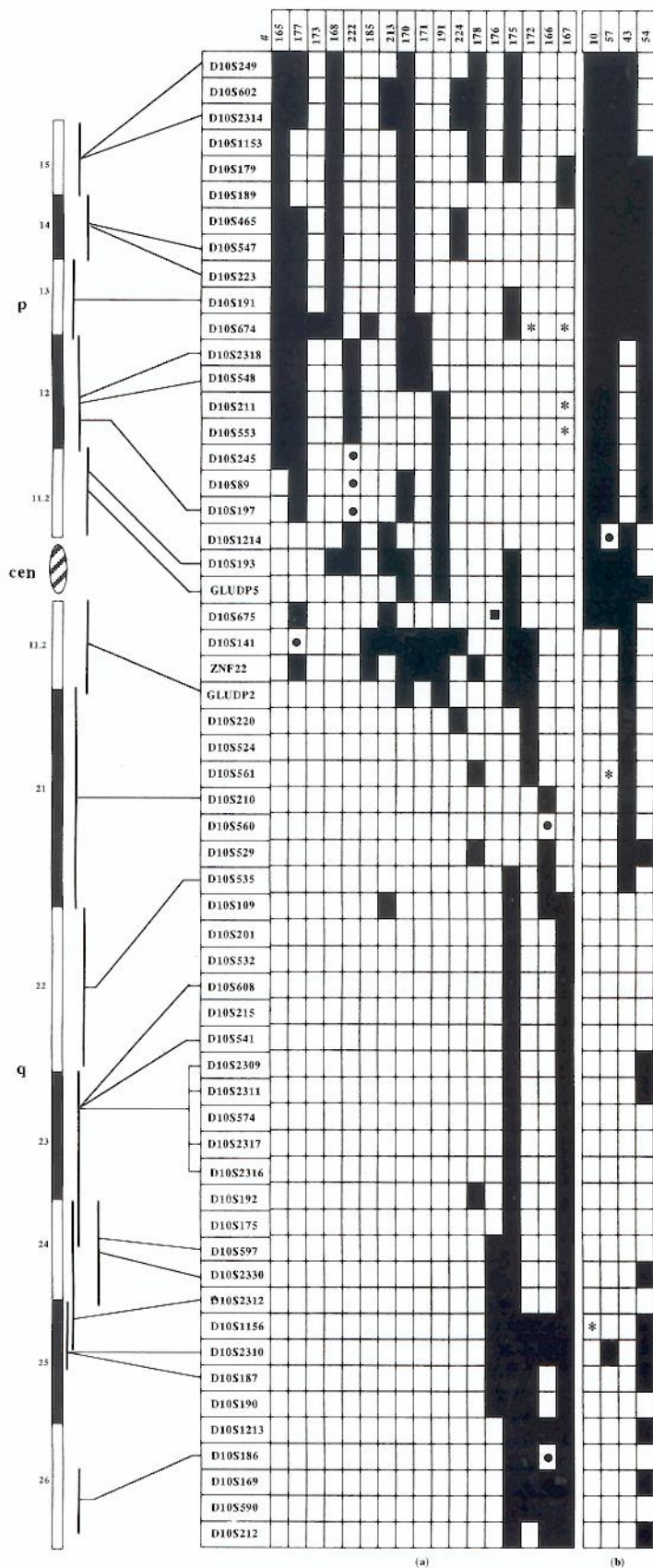


Fig. 2. FISH analysis and molecular characterization of 21 somatic cell hybrids retaining fragments of human chromosome 10, based on the use of 57 microsatellite markers. Shaded boxes indicate retention of the respective marker and detection of chromosomal segments by FISH. Information on the discrepancies between molecular characterization and FISH analysis is also included. Chromosome 10 segments recognized by the markers only are indicated by an asterisk (*), a segment detected only by FISH is marked by a small filled square, and molecularly characterized microdeletions which do not produce a FISH signal are marked by a dot. **(a)** Radiation hybrids. **(b)** Somatic cell hybrids with spontaneous chromosomal rearrangements containing additional human cytogenetic material. The diagram at the left is an ideogram of chromosome 10 showing the cytogenetic localization of 25 of the markers.

some 10 PCPs and YACs. Shown in Fig. 3a is the result of cohybridization of PCP 167, specific for 10q22→qter, with PCP 3B10, specific for 10q11.2 and 10q21.3. The pattern observed in the orangutan clearly indicates the presence of a paracentric inversion with breakpoints at 10q22.1 and 10q11.23. The result of cohybridization of PCP 3B10 and YAC 928B3, which maps to 10q21.1, confirmed the previous observation (Fig. 3b). Several YACs were used to refine the breakpoints. The most informative result was obtained when YAC 933A3 was applied (Fig. 3c). This YAC was clearly disrupted by the inversion, thereby allowing a very precise identification of the rearrangement.

Discussion

In this work, we characterized a somatic cell hybrid panel retaining fragments of human chromosome 10. Hybrids were mapped by FISH and subsequently screened with a group of microsatellite markers distributed along a well-supported genetic map of this chromosome. The map consists of a mixture of novel as well as previously employed microsatellite and fragment length polymorphic markers, thus providing a useful resource for comparisons of genetic distances and marker order. In regard to the Généthon chromosome 10 map (Dib et al., 1996), our genetic map contains 66 markers in common (43% of the markers we employed). With a few minor exceptions, the order of the markers is the same in both maps. This observation confirms the location of the markers along chromosome 10 and supports the credibility of our map. Only a small number of minor order conflicts, within short genetic distances, were observed. For example, within a comparable 3-cM interval, we have placed D10S190 and D10S209 distal to D10S542 (Fig. 1), whereas in the Généthon map, D10S209 is placed distal to D10S542 and D10S190. Similarly, D10S578 and D10S604 are separated by 2 cM in the Généthon map, whereas in our map they co-localize with D10S141 and D10S176. These conflicts may be due to (1) differences in the approach to map construction, (2) the relative amount of genotypic data used, (3) the overall marker content of the two maps, or (4) unchecked genotyping errors. Three reference markers (D10S33, D10S1214, and D10S12) mapped to more than one interval on the Généthon map have been placed at unique positions on our map. Within a map segment of comparable genetic distance (~33 cM) in both maps framed by markers D10S541 and D10S187 at 10q23.3→q25, we have placed seven novel markers, viz., D10S2311, D10S2316, D10S2330, D10S2312, D10S2309, D10S2317, and D10S2310. Interestingly, five of these markers are of the rare tetranucleotide repeat type. Although the sample is small, this could indicate a nonrandom distribution of such markers along the chromosomes. Alternatively, it could indicate a biased representation of the libraries employed for their isolation. Within the disease-gene-rich 10q23→q25 region, still unidentified disease-related genes exist (Meitinger et al. 1997). The high density of markers in the region may be useful for the positional cloning of, as yet, unknown disease-causing genes. The largest gap (11.3 cM) on the genetic map is between D10S186 and D10S169 on the dis-

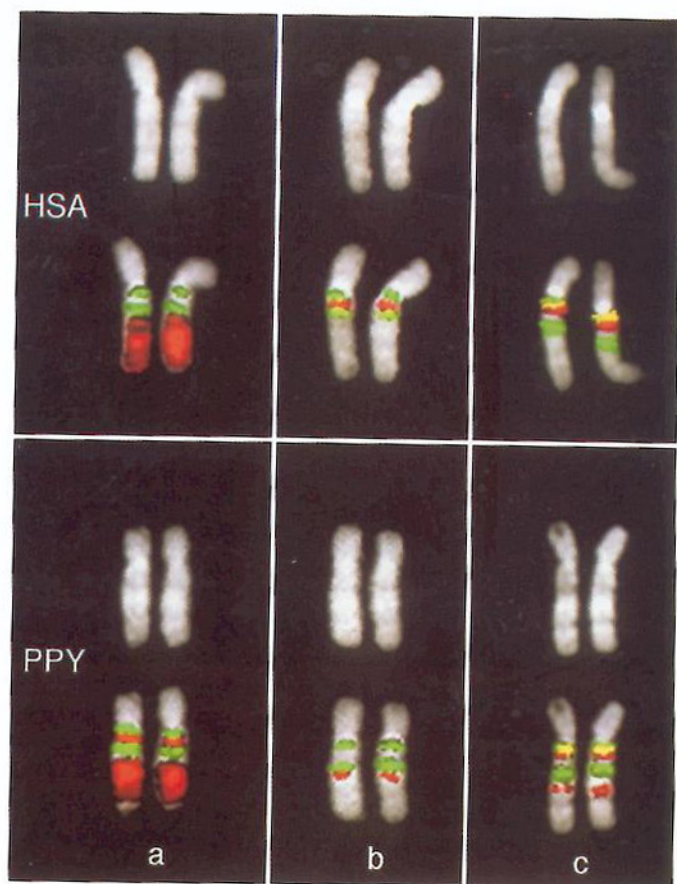


Fig. 3. Conclusive co-hybridization experiments to investigate the rearrangement differentiating the long arm of human chromosome 10 (HSA 10) and its orangutan counterpart (PPY 7). (a) Cohybridization of partial chromosome paint (PCP) 167 (red signal) with PCP 3B10 (green signal). (b, c) Co-hybridization of PCP 3B10 with YAC 928B3 (b) and YAC 933A3 (c). The DAPI-banded chromosomes are shown without hybridization signals (upper row) to display their morphology better. Additional information on the cytogenetic applications of PCPs may be found at <http://bioserver.uni-ba.it/fish/rocchi/welcome.html>.

tal portion of 10q. D10S186 maps to 10q26 (MIT database, Fig. 2). Since D10S169 is distal to D10S186, we conclude that it also maps to 10q26. The observed distance gap could be due to the fact that 10q26 is a subtelomeric region exhibiting a high rate of recombination. However, it should be mentioned that these markers were found to be physically separated in two hybrids (54 and 166), as suggested by both FISH and molecular analysis (Fig. 2).

Highly informative microsatellite markers are the most useful STSs involved in linkage analysis and in somatic cell hybrid characterization. Linkage analysis and molecular screening of the hybrid panel, permitted the dissection of chromosome 10 into 49 regions at an average genetic distance of about 3.9 cM, reflecting an average physical distance of about 2.9 Mb (size of HC10: 144 Mb). A combination of the two approaches provided better resolution for subregional mapping of sequences located on chromosome 10. Accordingly, markers D10S191 and D10S674, positioned together by linkage analysis, were

found physically to be separated in five hybrids, viz., 173, 185, 171, 172, and 167 (Fig. 2). In contrast, other markers grouped together at the genetic map, viz., D10S211/D10S553, D10S89/D10S197, and D10S608/D10S215, were found to be retained together in all hybrids scored, possibly suggesting their close physical linkage. Furthermore, molecular analysis of the hybrids permitted more precise mapping of the markers positioned approximately by linkage analysis. Thus, D10S2310, mapped to an 18.6-cM interval between D10S254 and D10S209 (Fig. 1), was found to be closely linked to D10S1156, as deduced from hybrids 176, 166, and 54 (Fig. 2). Similarly, the marker pattern of hybrids 222 and 171 suggested a close physical linkage of D10S2318 with D10S548.

FISH and molecular characterization of the hybrids revealed complementary and mutually supporting mapping information. FISH is inadequate for detecting microdeletions (present, as expected, mainly in radiation hybrids) because of chromatin condensation and because spreading of the fluorescent signals tends to fill small gaps; however, it is very efficient in detecting small chromosomal fragments. In contrast, the molecular approach offers a more detailed profile of the hybrid. In fact, several microdeletions that were not obvious in the FISH studies were detected by scoring for microsatellite markers. For example, hybrid 222, mapped by FISH to 10p12→p11.2, showed an internal microdeletion between D10S553 and D10S1214. When comparing FISH and PCR screening results, it should be taken into account that the ability of the molecular approach to detect small deletions or remote fragments depends upon both the number and spacing of the markers used for the analysis and the degree of representation of the particular sequence in the hybrid. Chromosome fragments produced by spontaneous rearrangements in cell hybrids are considered as having been created by a single breakage event, thus consisting of an uninterrupted chromosomal segment. Indeed, such hybrids in our panel appear with a considerably smaller number of internal microdeletions (Fig. 2b). These hybrids may be particularly useful in mapping studies, especially when radiation hybrid experiments produce conflicting results. Furthermore, they can be used for generating PCPs via *Alu*-PCR. PCPs offer an alternative approach for basic or clinical research and for diagnostic studies. They are useful for detecting chromosome rearrangements in tumors (Mitelman et al., 1997), where the classical techniques frequently fail due to the poor morphology of cancer cell chromosomes, or in studies of mammalian and, particularly, primate chromosome evolution (Muller et al., 1996). A combination of appropriate human chromosome 10 hybrids and YAC clones were employed for the refinement of cytogenetic and physical mapping of the paracentric inversion distinguishing the orangutan chromosome homolog from human chromosome 10.

As an additional resource for high-resolution map integration and cytogenetic investigations, we characterized 57 human chromosome 10-specific YACs. The physical order of the YAC clones, as determined by FISH, was found to correlate well with the MIT and the EUROGEN genetic linkage map data (results not shown). The majority of the YACs were recognized by markers used for the analysis of the cell hybrid panel, thus enabling the integrated use of YACs and PCPs derived from

the hybrids. In some cases, groups of YACs were recognized by the same marker, e.g., 10GA2, 10GA7, and 8IE1 by D10S2314 and 27DD2, 14CB8, 19FC12, and 24AC1 by D10S2312. *Alu*-PCR fingerprinting and pulsed-field gel electrophoresis restriction analysis revealed that these genomic clones comprise contigs with sizes ranging from 500 to 800 kb (M.A. Kokkinaki et al., unpublished data). Furthermore, markers grouped together in the genetic map turned out to be located, in fact, within short physical distances. Thus, D10S2311 and D10S564 (Fig. 1) were identified on a single YAC (10BA1). YACs 12HA1 and 37IF9, overlapping with 10BA1, contained two additional markers, D10S1753 and D10S2309, suggesting the close physical linkage of all four markers. To improve the resolution of the determined contig, several PAC clones were isolated. *Alu*-PCR fingerprinting and restriction analysis resulted in the selection of 16 PACs (Table 1) consisting a well-defined contig of about 700 kb in size, at a resolution of 50–80 kb (M.A. Kokkinaki et al., unpublished results). In another case, two YACs (27BC9 and 35GE1) and two PACs (171-A21 and 30-C6) revealed the close physical linkage of D10S2330 and D10S597, thus invalidating the estimated genetic distance (1.9 cM) of the two markers (Fig. 1). Approaches as those described, upgrade the integration of the genetic, cytogenetic, and physical maps of human chromosome 10. The molecular analysis of the panel facilitates the choice of appropriate genomic sequences to be used as additional tools for refining the molecular structure of cytogenetic rearrangements. This analysis, combined with the characterization of chromosome 10-specific genomic clones, permit the precise placement of the hybrids to the integrated chromosome 10 map, contribute to resources available for mapping this chromosome, and provide a powerful tool for various analytical cytogenetic methodologies and positional cloning procedures.

Acknowledgements

We are particularly grateful to the YAC Screening Center, DIBIT-HSR, Milan. We thank J. Attwood for assistance with the linkage map building. CEPH DNA filters and samples were provided to one of us (N.K.M.) by Génethon (Paris) in the context of the EUROGEN consortium.

References

- Antonacci R, Marzella R, Finelli P, Lonoce A, Forabosco A, Archidiacono N, Rocchi M: A panel of subchromosomal painting libraries representing over 300 regions of the human genome. *Cytogenet Cell Genet* 68:25–32 (1995).
- Archidiacono N, Antonacci R, Forabosco A, Rocchi M: Preparation of human chromosome painting probes from somatic cell hybrids, in Choo KHA (ed): *Methods in Molecular Biology: In situ Hybridization Protocols*, pp 1–13 (Humana Press, Ottowa 1994).
- Baldini A, Ward DC: In situ hybridisations of human chromosomes with *Alu*-PCR products: a simultaneous karyotype for gene mapping studies. *Genomics* 9:770–774 (1991).
- Chumakov IM, Rigault P, LeGall I, Bellané-Chantelot C, Billault A, Guillou S, Soularue P, Guasconi G, Poullier E, Gros I, Belova M, Sambucy JL, Susini L, Gervy P, Glibert F, Beaufils S, Bui H, Massart C, De Tand MF, Dukasz F, Lekoulant S, Ougen P, Perrot V, Saumier M, Soravito C, Bahouayila R, Cohen-Akenine A, Berillot E, Bertrand S, Codani JJ, Caterina D, Georges I, Laeroix B, Lucotte G, Sabbatou M, Schmit C, Sanguoat M, Tubacher E, Dib C, Fauré S, Fizames C, Gyapay G, Millasseau P, Nguyen S, Muselet D, Vignal A, Morissette J, Menninger J, Lieman J, Desai T, Banks A, Bray-Ward P, Ward D, Hudson T, Gerety S, Foote S, Stein L, Page DC, Lander ES, Weissenbach J, LePaslier D, Cohen D: A YAC contig map of the human genome. *Nature* 377(suppl):175–297 (1995).
- Cox SA, Attwood J, Bryant SP, Bains R, Povey S, Rebello M, Kapssetaki M, Moschonas NK, Grzeschik KH, Otto M, Dixon M, Sudworth HE, Kooy RF, Wright A, Teague P, Terrenato L, Vergnaud G, Monfouilloux S, Weissenbach J, Alibert O, Dib C, Faure S, Bakker E, Pearson NM, Vossen RHAM, Gal A, Mueller-Myhsok, Cann HM, Spurr NK: European Gene Mapping Project (EUROGEM): breakpoint panels for human chromosomes based on the CEPH reference families. *Ann hum Genet* 60:447–486 (1996).
- Daw SC, Taylor C, Kraman M, Call K, Mao J, Schuffenhauer S, Meitinger T, Lipson T, Goodship J, Scrambler P: A common region of 10p deleted in DiGeorge and velocardiofacial syndromes. *Nature Genet* 13:458–460 (1996).
- Deaven LL, Van Dilla MA, Barthold MF, Carrano AV, Cram LS, Fuscoe JC, Gray JW, Hildebrand CE, Moyzis RK, Perlman J: Construction of human chromosome-specific DNA libraries from flow-sorted chromosomes. *Cold Spring Harb Symp quant Biol* 51:159–167 (1986).
- Deloukas P, Dauwerse JG, Moschonas NK, van Ommen G-JB, van Loon APMG: Three human glutamate dehydrogenase genes (GLUD1, GLUDP2 and GLUDP3) are located on chromosome 10q, but are not closely physically linked. *Genomics* 17:676–681 (1993).
- Dib C, Fauré S, Fizames C, Samson D, Drouot N, Vignal A, Millasseau P, Marc S, Hazan J, Seboun E, Lathrop M, Gyapay G, Morissette J, Weissenbach J: A comprehensive genetic map of the human genome based on 5,264 microsatellites. *Nature* 380:152–154 (1996).
- Eng C, Mulligan LM: Mutations of the RET proto-oncogene in the multiple endocrine neoplasia type 2 syndromes, related sporadic tumours, and Hirschsprung disease. *Hum Mutat* 9:97–109 (1997).
- Francke U, Chang E, Comeau K, Geigle E-M, Giacalone J, Li X, Luna J, Moon A, Welch S, Wilgenbus P: A radiation hybrid map of human chromosome 18. *Cytogenet Cell Genet* 66:196–213 (1994).
- Goss SJ, Harris H: New method for mapping genes in human chromosomes. *Nature* 255:680–684 (1975).
- Gyapay G, Schmitt K, Fizames C, Jones H, Vega-Czarny N, Spillet D, Muselet D, Prud'Homme JF, Dib C, Auffray C, Morissette J, Weissenbach J, Goodfellow PN: A radiation hybrid map of the human genome. *Hum molec Genet* 5:339–346 (1996).
- Hudson TJ, Stein LD, Gerety SS, Ma J, Castle AB, Silva JS, Slonim DK, Baptista R, Kruglyak L, Xu SH, Hu X, Colbert AME, Rosenberg C, Reeve-Daly MP, Rozen S, Hui L, Wu X, Vestergaard C, Wilson KM, Bae JS, Maitra S, Ganiatsas S, Evans CA, DeAngelis MM, Ingalls KA, Nahf RW, Horton LT, Anderson MO, Collymore AJ, Ye W, Kouyoumjian V, Zemsteva IS, Tam J, Devine R, Courtney DF, Renaud MT, Nguyen H, O'Connor TJ, Fizames C, Fauré S, Gyapay G, Dib C, Morissette J, Orlin JB, Birren BW, Goodman N, Weissenbach J, Hawkins TL, Foote S, Page DC, Lander ES: An STS-based map of the human genome. *Science* 270:1945–1954 (1995).
- Ioannou PA, Amemiya CT, Garnes J, Kroisel PM, Shizuya H, Chen C, Batzer MA, de Jong PJ: A new bacteriophage P1-derived vector for the propagation of large human DNA fragments. *Nature Genet* 6:84–89 (1994).
- Jabs EW, Li X, Scott AF, Meyers G, Chen W, Eccles M, Mao J, Chamas LR, Jackson CE, Jaye M: Jackson-Weiss and Crouzon syndromes are allelic with mutations in fibroblast growth factor receptor 2. *Nature Genet* 8:275–279 (1994).
- Kapssetaki M, Kokkinaki M, Angelicheva D, Lubyova B, Mavraki H, Argyrokastritis A, Vergnaud G, Ferguson-Smith M, Rubinsztein DC, Lush M, Moschonas NK: The EUROGEM map of human chromosome 10. *Eur J hum Genet* 2:222–223 (1994).
- Knapp RD, Smith LC, Baldini A: Eigenanalysis of DAPI-stained chromosomes: tools and strategies toward computer-assisted analysis of FISH experiments. *Cytogenet Cell Genet* 69:81–86 (1995).
- Korenberg JR, Rykowski MC: Human genome organization: *Alu*, Lines and the molecular structure of metaphase chromosome bands. *Cell* 53:391–400 (1988).
- Liu P, Siciliano J, Seong D, Craig J, Zhao Y, de Jong PJ, Siciliano MJ: Dual *Alu* polymerase chain reaction primers and conditions for isolation of human chromosome painting probes from hybrid cells. *Cancer Genet Cytogenet* 65:93–99 (1993).
- Ma NSF, Zheng C, Benchekroun Y, Deaven LL, Longmire JL, Moir DT, Mao J: Characterization of a flow-sorted human chromosome 10 cosmid library by FISH. *Cytogenet Cell Genet* 74:266–271 (1996).
- Meitinger T, Scharfe C, Call K, Moschonas N: Report of the second international workshop on human chromosome 10 mapping 1997. *Cytogenet Cell Genet* 78:183–196 (1997).
- Mitelman F, Mertens F, Johansson B: A breakpoint map of recurrent chromosomal rearrangements in human neoplasia. *Nature Genet* 15(suppl):417–474 (1997).
- Moschonas NK, Spurr NK, Mao J: Report of the first international workshop on human chromosome 10 mapping 1995. *Cytogenet Cell Genet* 72:99–112 (1996).
- Muller S, Koehler U, Wienberg J, Marzella R, Finelli P, Antonacci R, Rocchi M, Archidiacono N: Comparative fluorescence in situ hybridization mapping of primate chromosomes with *Alu*-PCR generated probes from human/rodent somatic cell hybrids. *Chrom Res* 4:38–42 (1996).
- Nagase S, Yamakawa H, Sato S, Yajima A, Horii A: Identification of a 790-kilobase region of common allelic loss in chromosome 10q25-q26 in human endometrial cancer. *Cancer Res* 57:1630–1633 (1997).
- Nelson DL, Ledbetter SA, Corbo L, Victoria MF, Ramirez-Solis R, Webster TD, Ledbetter DH, Caskey CT: *Alu*-polymerase chain reaction: a method for rapid isolation of human-specific sequences from complex DNA sources. *Proc natl Acad Sci, USA* 86:6686–6690 (1989).
- Nikali K, Isomomppi J, Lonnqvist T, Mao JI, Suomalainen A, Peltonen L: Toward cloning of a novel ataxia gene: refined assignment and physical map of the IOSCA locus (SCA8) on 10q24. *Genomics* 39:185–191 (1997).
- Nunes ME, Schutt G, Kapur RP, Luthardt F, Kukolich M, Byers P, Evans JP: A second autosomal split hand/split foot locus maps to chromosome 10q24-q25. *Hum molec Genet* 4:2165–2170 (1995).
- Przyłępa KA, Paznekas W, Zhang M, Golabi M, Bias W, Bamshad MJ, Carey JC, Hall BD, Stevenson R, Orlow SJ, Cohen Jr MM, Jabs EW: Fibroblast growth factor receptor 2 mutations in Beare-Stevenson cutis gyrata syndrome. *Nature Genet* 13:492–494 (1996).
- Reardon W, Winter RM, Pulleyn LJ, Jones BM, Malcolm S: Mutation in the fibroblast growth factor receptor 2 gene cause Crouzon syndrome. *Nature Genet* 8:98–103 (1994).
- Rutland P, Pulleyn LJ, Reardon W, Baraitser M, Hayward R, Jones B, Malcolm S, Winter RM, Oldridge M, Slaney SF, Poole MD, Wilkie AOM: Identical mutations in the FGFR2 gene cause both Pfeiffer and Crouzon syndrome phenotypes. *Nature Genet* 9:173–176 (1995).
- Spurr NK, Bryant SP, Attwood J, Nyberg K, Cox SA, Mills A, Bains R, Warne D, Cullin L, Povey S, Seboun JM, Weissenbach J, Cann HM, Lathrop M, Dausset J, Marcadet-Troton A, Cohen D: European Gene Mapping Project (EUROGEM): genetic maps based on the CEPH reference families. *Eur J hum Genet* 2:193–252 (1994).
- Tzimogiorgis G, Leversha M, Chroniary K, Goulielmos G, Sargent CA, Ferguson-Smith M, Moschonas NK: Structure and expression analysis of a novel member of the human glutamate dehydrogenase (GLUD) gene family mapped to chromosome 10p11.2. *Hum Genet* 91:433–438 (1993).
- Wilkie AOM, Slaney SF, Oldridge M, Poole MD, Ashworth GJ, Hockley AD, Hayward RD, David DJ, Pulleyn LJ, Rutland P, Malcolm S, Winter RM, Reardon W: Apert syndrome results from localized mutations of FGFR2 and is allelic with Crouzon syndrome. *Nature Genet* 9:165–172 (1995).
- Yunis JJ, Prakash O: The origin of man: a chromosomal pictorial legacy. *Science* 215:1525–1530 (1982).