The Functional Role of Nucleotide Excision Repair in Transcription-associated DNA Damage in Mammals

UNIVERSITY OF CRETE-BIOLOGY DEPARTMENT

INSTITUTE OF MOLECULAR BIOLOGY & BIOTECHNOLOGY, FORTH

Kyriacos Agathangelou

Genome (In)stability and Mammalian Physiology Lab

PhD thesis

Supervisor: Professor George A. Garinis

Heraklion, June 2019

ΤΡΙΜΕΛΗΣ ΣΥΜΒΟΥΛΕΥΤΙΚΗ ΕΠΙΤΡΟΠΗ:

Γαρίνης Γιώργος: Καθηγητής τμήματος Βιολογίας Πανεπιστημίου Κρήτης και συνεργαζόμενο μέλος ΔΕΠ στο Ινστιτούτο Μοριακής Βιολογίας και Βιοτεχνολογίας-ΙΤΕ.

Σπηλιανάκης Χαράλαμπος: Αναπληρωτής Καθηγητής τμήματος Βιολογίας Πανεπιστημίου Κρήτης και συνεργαζόμενο μέλος ΔΕΠ στο Ινστιτούτο Μοριακής Βιολογίας και Βιοτεχνολογίας-ΙΤΕ.

Αλεξανδράκη Δέσποινα: Καθηγήτρια τμήματος Βιολογίας Πανεπιστημίου Κρήτης και συνεργαζόμενο μέλος ΔΕΠ στο Ινστιτούτο Μοριακής Βιολογίας και Βιοτεχνολογίας-ΙΤΕ.

ΕΠΤΑΜΕΛΗΣ ΕΞΕΤΑΣΤΙΚΗ ΕΠΙΤΡΟΠΗ:

Γαρίνης Γιώργος: Καθηγητής τμήματος Βιολογίας Πανεπιστημίου Κρήτης και συνεργαζόμενο μέλος ΔΕΠ στο Ινστιτούτο Μοριακής Βιολογίας και Βιοτεχνολογίας-ΙΤΕ.

Σπηλιανάκης Χαράλαμπος*:* Αναπληρωτής Καθηγητής τμήματος Βιολογίας Πανεπιστημίου Κρήτης και συνεργαζόμενο μέλος ΔΕΠ στο Ινστιτούτο Μοριακής Βιολογίας και Βιοτεχνολογίας-ΙΤΕ.

Αλεξανδράκη Δέσποινα: Καθηγήτρια τμήματος Βιολογίας Πανεπιστημίου Κρήτης και συνεργαζόμενο μέλος ΔΕΠ στο Ινστιτούτο Μοριακής Βιολογίας και Βιοτεχνολογίας-ΙΤΕ.

Τζαμαρίας Δημήτρης: Καθηγητής τμήματος Βιολογίας Πανεπιστημίου Κρήτης.

Δελιδάκης Χρήστος: Καθηγητής τμήματος Βιολογίας Πανεπιστημίου Κρήτης και συνεργαζόμενο μέλος ΔΕΠ στο Ινστιτούτο Μοριακής Βιολογίας και Βιοτεχνολογίας-ΙΤΕ.

Ζάχος Γιώργος: Αναπληρωτής Καθηγητής τμήματος Βιολογίας Πανεπιστημίου Κρήτης.

Κρετσόβαλη Ανδρονίκη: Κύρια Ερευνήτρια, Ινστιτούτο Μοριακής Βιολογίας και Βιοτεχνολογίας-ΙΤΕ.

Funding

The Horizon 2020 ERC Consolidator grant “DeFiNER” (GA64663), the FP7 Marie Curie ITN “aDDRess” (GA316390), “CodeAge” (GA316354), “Marriage” (GA316964), the Horizon 2020 Marie Curie ITN “Chromatin3D (GA GA622934), the ELIDEK grant 1059 and the “Fondation Santé” grant supported this PhD thesis.

CONTENTS

Προσωπικές ευχαριστίες

Καθώς η διδακτορική μου διατριβή πλησιάζει προς ολοκλήρωση, θα ήθελα να αναγνωρίσω όλους όσους συνέβαλαν σε αυτή την πορεία, προσφέροντάς μου τη στήριξή τους τόσο σε ακαδημαϊκό όσο και σε προσωπικό επίπεδο.

Με αφετηρία την εμπιστοσύνη που μου έδειξε ο επιβλέποντας καθηγητής μου κ. Γιώργος Γαρίνης, ξεκίνησα τις διδακτορικές μου σπουδές στο εργαστήριό του το 2013. Ευχαριστώ τον Γιώργο που μου πρόσφερε, πάντοτε και απλόχερα, ένα ευρύ φάσμα επιστημονικής γνώσης και άποψης. Θέτοντας το παράδειγμα, μου έδωσε το κίνητρο να αναπτύξω ερευνητικές δεξιότητες που θα με συντροφεύουν στα επόμενα μου βήματα. Θα είμαι πάντοτε ευγνώμων που μέσα από την άρτια συνεργασία, την αστείρευτη ενέργεια και ενδιαφέρον του με κατεύθυνε σωστά προς την υλοποίηση της ερευνητικής αυτής διατριβής.

Ευχαριστίες οφείλω προς όλα τα μέλη του εργαστηρίου, μέλη που έφυγαν, μέλη που ήρθαν αλλά και μέλη που μας άφησαν. Ιδιαίτερες ευχαριστίες προς την Καλλίνα και την Τζωρτζίνα με τις οποίες η συνεργασία υπήρξε καθοριστική και πολύτιμη μέχρι και τα τελευταία στάδια της πορείας μου αυτής. Ξεχωριστό μέρος στην υπόληψή μου θα έχει πάντοτε ο Θοδωρής, που αφιερώνοντας προσωπικό του χρόνο, μου πρόσφερε καθημερινά τις γνώσεις του στη δημιουργία των ποντικιών, αλλά και τη στήριξή του ως ένας πολύ καλός φίλος.

Θα ήθελα επίσης να ευχαριστήσω τα μέλη της τριμελούς και κατ' επέκταση της επταμελούς επιτροπής μου για την αμεσότητα στην επιστημονική καθοδήγηση και τις συμβουλές τους.

Με χαρά ευχαριστώ το καλό φίλο Μάνο, που παράλληλα με τη συντροφιά που μου πρόσφερε στο καθημερινό μεσημεριάτικο τραπέζι, ήταν πάντα πρόθυμος να με ακούσει και να με στηρίξει. Προσπαθώντας να προβλέψει το αποτέλεσμα των πειραμάτων μου σύμφωνα με τους αστερισμούς και τις φάσεις της σελήνης, χάριζε μια χιουμοριστική πνοή που μου έδινε κουράγιο να συνεχίσω.

Οι τελευταίες και πιο θερμές ευχαριστίες είναι προς την οικογένειά μου, που άντεξαν την πολύχρονη απουσία μου, που πίστευαν πάντα σε μένα και υποστήριζαν τις επιθυμίες μου. Εγκάρδιες ευχαριστίες προς την αδελφή μου Σοφία που ήταν πάντα εκεί για να μοιραστώ τις ανησυχίες μου και για να με συμβουλεύσει κατάλληλα στο κάθε μου βήμα.

ΠΕΡΙΛΗΨΗ

Σε αντίθεση με τις πρωτεΐνες και τα υπόλοιπα μακρομόρια π.χ. τα σάκχαρα ή τα λίπη, το πυρηνικό DNA (η απαρχή του RNA και των πρωτεϊνών) είναι αναντικατάστατο. Παρά το γεγονός ότι η χημική του σύσταση είναι εξαιρετικά ασταθής, το DNA οφείλει να διατηρηθεί αναλλοίωτο καθόλη τη διάρκεια της ζωής του κυττάρου ώστε η γενετική πληροφορία να κληρονομηθεί αυτούσια στα θυγατρικά κύτταρα. Ωστόσο, η ύπαρξη διαφόρων ενδογενών γενοτοξικών παραγόντων προκαλούν τη σταδιακή συσσώρευση πλήθους δομικών αλλοιώσεων και προσβολών (π.χ. υδρόλυση, απαμίνωση βάσεων, διμερισμός πυριμιδινών, δημιουργία θραυσμάτων μονής ή διπλής DNA έλικας κτλ.) στο DNA. Η επακόλουθη γενωμική αστάθεια επιφέρει δραματικές αλλαγές στη φυσιολογία του κυττάρου παρεμποδίζοντας τη φυσιολογική λειτουργία ζωτικών βιολογικών διεργασιών όπως η μεταγραφή ή/και ο αναδιπλασιασμός του DNA. Για να αντιμετωπίσουν την σταδιακή συσσώρευση DNA βλαβών, τα ευκαρυωτικά κύτταρα έχουν αναπτύξει ένα σύνολο αλληλεπικαλυπτόμενων επιδιορθωτικών μηχανισμών, συμπεριλαμβανομένου του μηχανισμού εκτομής νουκλεοτιδίων (Nucleotide Excision Repair, NER), που εντοπίζουν, επιδιορθώνουν και αποκαθιστούν το προσβαλλόμενο DNA στην αρχική του μορφή. Στον άνθρωπο και τα αντίστοιχα πειραματικά μοντέλα ποντικών, η ύπαρξη εγγενών μεταλλαγών σε γονίδια του μονοπατιού NER, προκαλεί ένα ευρύ φάσμα κλινικών συμπτωμάτων που χαρακτηρίζεται από εξαιρετική ετερογένεια, η οποία δε μπορεί να εξηγηθεί αποκλειστικά λόγω της ατελούς επιδιόρθωσης του DNA. Πρόσφατες μελέτες απεκάλυψαν ότι ορισμένες πρωτεΐνες του NER συμμετέχουν, πέραν της επιδιόρθωσης των DNA βλαβών, σε κυτταρικές διεργασίες όπως η έναρξη της μεταγραφής και η αναδιαμόρφωση ή ο σχηματισμός της τρισδιάστατης δομής της χρωματίνης στο χώρο. Για να διαλευκάνουμε το λειτουργικό ρόλο του NER στην ανάπτυξη και τις ασθένειες στα θηλαστικά, αναπτύξαμε την μέθοδο της *in vivo* σήμανσης με βιοτίνη της πρωτεΐνης XPF στον ποντικό. Η προσέγγιση αυτή σε συνδυασμό με μεθοδολογίες αλληλούχισης DNA υψηλής απόδοσης και λειτουργικές προσεγγίσεις απεκάλυψαν ότι το ετεροδιμερές του συμπλόκου ενδονουκλεάσης του NER ERCCI-XPF αλληλεπιδρά με πρωτεϊνικούς παράγοντες που συμμετέχουν στην μεταγραφή και την εξομάλυνση του τοπολογικού φόρτου του DNA κατά την διαδικασία της μεταγραφής. Συγκεκριμένα, ανιχνεύσαμε ότι κατά την επαγωγή της μεταγραφής, η ERCC1-XPF προσδένεται σε υποκινητές, κατά μήκος του γονιδιώματος. Επιπλέον μελέτες απεκάλυψαν ότι η πρόσδεση του συμπλόκου ERCC1-XPF στο DNA, συμπίπτει με την δημιουργία εκτομών διπλού θραύσματος στο

DNA (DNA double strand breaks, DSBs) σε διακριτές περιοχές του DNA. Τα αποτελέσματα της μελέτης αναδεικνύουν τον λειτουργικό ρόλο της ERCC1-XPF στην επιδιόρθωση DNA βλαβών που προκαλούνται κατά την διαδικασία της μεταγραφής και παρέχουν ένα μηχανιστικό μοντέλο που εξηγεί ικανοποιητικά την κλινική ετερογένεια των συνδρόμων NER.

Abstract

DNA is not disposable or recyclable and must be repaired when damaged. All other cellular polymers, including RNA, proteins and polysaccharides, are regularly turned over, broken down and rebuilt based on the blueprint provided by the genetic information stored in DNA. Thus, as the maintenance of DNA integrity is vital to the proper functioning of every cell, so is DNA damage inexorably linked to cellular dysfunction. To meet this challenge, mammalian cells have evolved machineries to maintain telomeres intact as well as overlapping repair pathways to counteract structural DNA modifications (e.g., nicks, gaps, DNA double-strand breaks (DNA DSBs), and the myriad alterations that may block DNA transcription or replication). For bulky helix-distorting damage, such as the main Ultraviolet (UV)-induced lesions, the principal repair mechanism is the evolutionarily conserved Nucleotide Excision Repair (NER) pathway. In humans and the corresponding animal models, inborn mutations in genes involved in NER are associated with a wide range of clinical symptoms whose remarkable heterogeneity cannot be adequately explained by the random accumulation of DNA damage. Recent studies revealed that, in addition to DNA repair, NER factors are also involved in transcription initiation and elongation, chromatin remodeling and the 3D genome architecture. To dissect the functional role of NER during mammalian development and disease, we used an *in vivo* biotinylation-tagging approach and unbiased high-throughput genomics and proteomics approaches techniques in mice. Here, we show that the NER structure-specific endonuclease ERCC1-XPF complex interacts with protein factors involved in the resolution of DNA topological stress that is triggered during the process of transcription. Upon induction of mRNA synthesis, we find that the ERCC1-XPF complex recruits on promoters, genome-wide and that its recruitment on DNA coincides with the presence of DNA DSBs on these genomic regions. Together, our findings point to the involvement of ERCC1-XPF in the repair of transcription-associated DNA damage events, shedding new light into the pathological complications associated with NER-deficient syndromes.

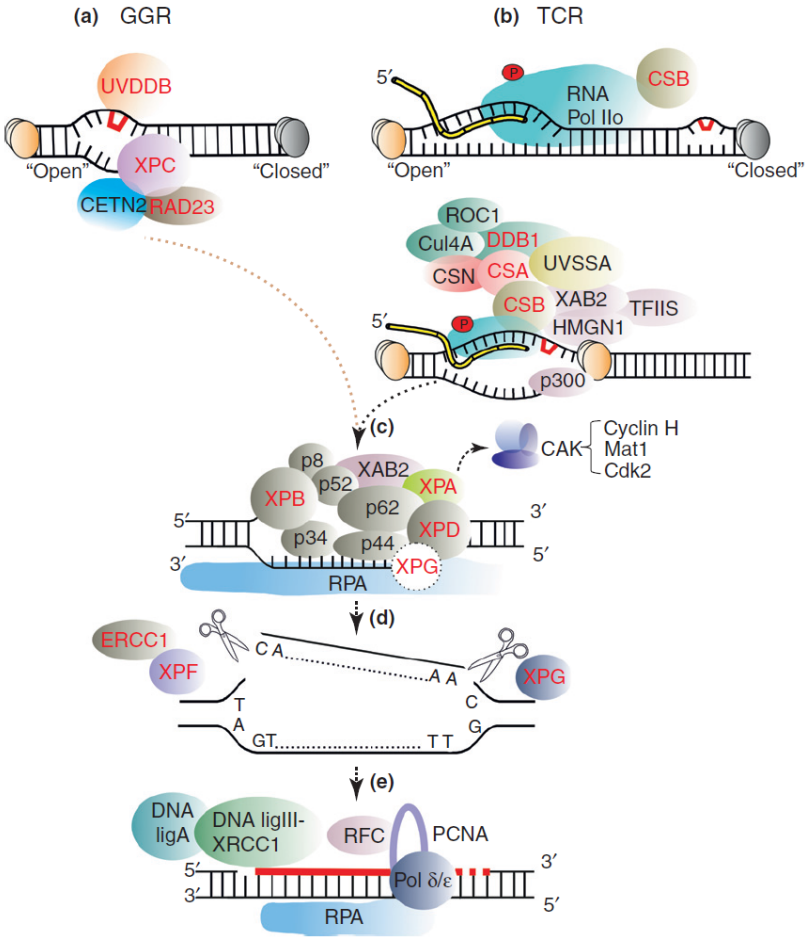
1. Introduction

The DNA double helix is continuously exposed to genotoxic insults from exogenous (e.g. ultraviolet radiation), or endogenous (e.g metabolic byproducts) sources, that threaten its integrity (Agathangelou et al., 2018; Apostolou et al., 2019; Kamileri et al., 2012a). DNA insults may obstruct transcription altering cellular metabolism (Chatzinikolaou et al., 2017; Kamileri et al., 2012c; Karakasilioti et al., 2013) or interfere with proper DNA replication driving mutations and cancer. Mammalian cells have evolved a battery of highly conserved, partially complementary mechanisms aiming at preserving genome integrity. Nucleotide excision repair (NER), is a highly conserved DNA repair machinery that consists of ~30 proteins that detect and repair bulky DNA helix-distorting lesions, such as the UV-induced cyclobutane pyrimidine dimers (CPDs) and pyrimidine-pyrimidone (6-4) photoproducts (6-4PPs) across the genome or only on the actively transcribed part of the genome (Hoeijmakers, 2009).

1.1 The Nucleotide Excision Repair mechanism. In NER, lesion recognition occurs genome-wide (global-genome NER; GG-NER) or in a transcription-coupled mode (transcription-coupled repair; TC-NER) that differ only in how DNA damage is recognized (**Figure 1**). In GG-NER, DNA lesions found on the non-transcribed strand of active genes or on transcriptionally inert genomic regions are detected by the heterotrimeric protein complex XPC-RAD23B-Centrin-2 (Araki et al., 2001; Riedl et al., 2003; Sugasawa et al., 1998). XPC binds the strand opposite the damaged DNA site, thus allowing the recognition of diverse types of DNA insults. Based on a recent series of photobleaching experiments, XPC does not seem to scan the genome for lesions by freely moving through the nucleoplasm. Instead, the protein binds and subsequently dissociates from DNA; its mobility decreases substantially in the presence of UV-induced DNA lesions allowing the remaining NER factors to recruit to sites of damage (Hoogstraten et al., 2008). Once XPC binds to DNA damage, RAD23 dissociates from the complex and becomes dispensable for the remaining NER reaction (Batty et al., 2000; Bergink et al., 2012). Minor DNA distortions, such as those induced by the UV-induced cyclobutane pyrimidine dimers (CPDs), are preferentially recognized by the heterodimer UV-damaged DNA-binding protein UV-DDB. UV-DDB is composed of DDB1 (or XPE binding factor) and DDB2 (called XPE) that are part of the CUL4A-ROC1 ubiquitin ligase complex (Bondar et al., 2006). DDB2 acts as an auxiliary damage recognition factor. It binds directly to UV-induced DNA

lesions and exposes a ssDNA moiety opposite the DNA lesion to facilitate the recruitment of XPC at sites of damage (Scrima et al., 2008). DDB2 also triggers the ATP- and poly(ADP-ribose) (PAR) polymerase (PARP)-dependent unfolding of higher-order chromatin structure in the vicinity of DNA damage in a mechanism that is independent of the CRL4-ubiquitin ligase complex (Luijsterburg et al., 2012). Other ATP-dependent chromatin remodeling factors, such as the INO80 complex (Jiang et al., 2010) or the ATPase BRG1 of the SWI/SNF-like BAF (BRG1/BRM-associated factor) (Zhang et al., 2009) are known to play critical roles in rendering the chromatin accessible during the initial recognition steps of the NER process. Upon binding of the complex to damage, UV-DDB ubiquitinates XPC (Sugasawa et al., 2005), DDB2 (Guerrero-Santoro et al., 2008) and histones (Sugasawa et al., 2005; Wang et al., 2006). XPC-RAD23B-Centrin-2 melts the DNA around the lesion and recruits TFIIH, a 10-subunit pre-incision complex that contains the ATPases/helicases XPB and XPD, GTF2H1 (P62), GTF2H4 (P52) that stimulates XPB, GTF2H2 (P44) that stimulates XPD, GTF2H3 (P34) and (GTF2H5 (P8) in addition to the Cdk-activating-kinase (CAK) complex [22] and XPG (Egly and Coin, 2011). Unlike XPD, however, the helicase activity of XPB is dispensable for NER. Instead, the ATPase activity of XPB is required for the opening and repair of damaged DNA. As with promoters, XPB also uses ATP to keep the two strands of the DNA around the DNA lesion apart facilitating the unwinding of DNA by XPD; the helicase activity of XPD is thought to be required for damage verification (Coin et al., 2007; Sugasawa et al., 2009). The 3-dimensional structure of XPD suggests that it forms a central pore sufficient in size to allow a single-stranded DNA (ssDNA) to pass and likely to be scanned for possible DNA lesions (Koch et al., 2016; Mathieu et al., 2013; Wolski et al., 2008). CAK is then released from the core complex. Following TFIIH assembly on damaged DNA, XPA binds the DNA on the 5'side of the DNA insult and together with replication protein A (RPA) that binds the ssDNA opposite the DNA lesion, it stabilizes the damaged DNA for incision (Scharer, 2013). RPA activates the structure-specific endonucleases ERCC1-XPF and XPG and is released from DNA to initiate new incision events (Overmeer et al., 2011). ERCC1-XPF and XPG cleave the 5' and 3'side of the 24-32-nucleotide fragment, respectively containing the damaged DNA fragment. Following the removal of the DNA lesion, the ssDNA gap is then filled by the DNA pol ε (in replicating cells) or δ (in non-replicating cells) or the translesion DNA polymerase κ. The DNA polymerases are loaded on DNA with the proliferating cell nuclear antigen (PCNA) and the replication factors C (RFC) and RPA that stimulate their polymerase activity (Ogi et al., 2010; Shivji et al., 1992). The nascent DNA fragment is

then sealed by DNA ligase I in replicating cells or by ligase IIIα-XRCC1 in quiescent cells (Araujo et al., 2000; Mocquet et al., 2008; Moser et al., 2007). In TC-NER, the principal DNA damage sensor is the elongating RNA polymerase (RNAPII) itself. RNAPII presumably stalls at damaged DNA sites on the actively transcribed strand of genes (Hanawalt and Spivak, 2008) and recruits Cockayne syndrome B (CSB; also called ERCC6), a DNA-dependent ATPase (Citterio et al., 1998; Eisen et al., 1995; Troelstra et al., 1992) that wraps ~125 bp of DNA around its surface and depending on ATP binding, it translocates as a dimer along template DNA with RNAPII (Beerens et al., 2005). It is possible that the blockage of RNAPII at a lesion increases the affinity of CSB for the polymerase (Marteijn et al., 2014). This and the binding of CSB to CSA (also called ERCC8) that is part of an E3-ubiquitin ligase complex (i.e., DDB1, Cullin 4A, and ROC1/Rbx1) (Groisman et al., 2003) may serve to backtrack RNAPII that slides reversibly along DNA template (Nudler, 2012) rendering the DNA lesion accessible to the remaining NER reaction. CSA recruits together with CSB the nucleosomal binding protein HMGN1 (Birger et al., 2003; Fousteri et al., 2006), the pre-mRNA-splicing factor XAB2 (Nakatsu et al., 2000), the transcription elongation factor TFIIS (Fousteri et al., 2006; Tornaletti et al., 1999) and the UV-stimulated scaffold protein UVSSA that interacts with the elongating form of RNAPII and stabilizes CSB (Schwertman et al., 2012; Zhang et al., 2012) to sites where RNAPII is arrested. Besides UV-induced DNA lesions, CSB and CSA are also recruited to DNA ICLs, DNA DSBs, monoadducts and, for CSB alone, to oxidative DNA lesions (Iyama and Wilson, 2016) which are then removed by the core NER reaction.



Kamileri et al., 2012

***Figure 1. NER machinery**: **a)** Global Genome NER (GG-NER) and **b)** Transcription-Coupled NER (TC-NER), distinguishable by means of lesion recognition while common to the DNA unwinding and damage excision steps of repairing events.*

1.2 NER Syndromes - Clinical Phenotypic Heterogeneity. Genetic pleiotropy refers to a single gene mutation that influences distinct and seemingly disparate phenotypic disorders. Instead, genetic heterogeneity occurs when a single phenotypic trait is linked to mutations on multiple alleles from the same or different loci. In humans, inborn mutations in NER genes give rise to syndromes that are both heterogeneous and pleiotropic. Defects in GG-NER give rise to the skin cancer-prone disorder xeroderma pigmentosum (XP; affected proteins: XPA through XPG) or to a heterogeneous group of progeroid disorders that are phenotypically distinct from XP, including the Cockayne syndrome (CS; affected proteins: CSA, CSB, UVSSA or XPB, XPD and TTDA and certain mutations in the gene encoding XPG) and Trichothiodystrophy (TTD; affected proteins: XPB, XPD, P8) (Bootsma, 1998; Bootsma, 2001). Even though XP, CS, and TTD are derived from mutations in genes involved in the same DNA repair mechanism or even in the same NER

gene (Fan et al., 2008), their clinical outcome is often remarkably diverse. For instance, two siblings with XP as well as a CS patient carrying the same mutation in *CSB* gene were previously shown to be affected differently in terms of the severity of the

Table 1– Phenotypic characteristics of NER syndromes and associated mouse models (Apostolou et al., 2019)**.**

		Clinical features									
		UV sensitivity		Skin cancer		Growth defects		Premature aging		Neurological symptoms	
Protein defect	Syndrome	Human	Mouse	Human	Mouse	Human	Mouse	Human	Mouse	Human	Mouse
XPA	XP	Y	Y	Y	Y	Y	N	NA	N	Y	NA
XPB	XP, XPCS, TTD	Y	Y	Y	N	N	N	N	N	Y	N
XPC	XP	Y	Y	Y	Y	N	N	N	N	N	N
XPD	CS, XPCS, TTD, COFS	Y	Y	Y	Y	Y	Y	NA	Y	Y	N
DDB2	XP	Y	Y	Y	Y	N	N	N	N	N	N
XPF	XP, CS	Y	Y	ND	ND	Y	Y	Y	Y	Y	Y
XPG	XP, XPCS	Y	Y	ND	ND	Y	Y	NA	Y	Y	Y
ERCC1	COFS	Y	Y	ND	ND	Y	Y	Y	Y	Y	Y
TTDA	TTD	Y	Y	NA	NA	Y	NA	NA	NA	NA	NA
CSA	CS, UVsS	Y	Y	N	Y	Y	N	Y	N	Y	N
CSB	CS, UVsS, COFS	Y	Y	N	Y	Y	N	Y	N	Y	Y
CSB/XPA	CS, COFS	NA	Y	NA	NA	NA	Y	NA	Y	NA	Y
CSB/XPC	CS, COFS	NA	Y	NA	NA	NA	Y	NA	Y	NA	Y

disease indicating that a straightforward association between the genetic defect and the observed pathological features is not always guaranteed (Colella et al., 2000). In XP, patients carrying mutations in *XPC* or *XPE* genes show extreme sensitivity to sunlight and present with freckling in sun-exposed skin areas. If left unprotected, XP patients have a high frequency of non-melanoma skin cancer and melanomas that manifest at an early age (~9 and ~22 years, respectively) in UV-exposed areas i.e. skin, eye, and tongue but occasionally in non-UV exposed organs of the central nervous system (DiGiovanna and Kraemer, 2012). However, it is only 60% of XP patients that are photosensitive whereas the remaining 40% show no sunburn reactions but still present with increased freckle-like pigmentation (Mareddy et al., 2013). The time to onset of neurological abnormalities also varies dramatically from the age of two to middle age (Bradford et al., 2011; Mahindra et al., 2008). Moreover, it is only <30% of the patients with defects in XPA, XPB, XPD, or XPG that show neurological deficits of varying severity, including isolated hyperreflexia, progressive mental retardation, sensorineural deafness, spasticity, or seizures (Mareddy et al., 2013). In *XPC* patients, the inactivating mutations are spread throughout the gene giving rise mostly to truncated XPC proteins of variable length, with no indication of hotspots or founder effects (Chavanne et al., 2000). A rather distinct group of XP patients

involves mutations in the gene encoding XPV (also called low fidelity Pol η)(Johnson et al., 1999b); XPV patients are proficient in NER but manifest with similar symptoms seen in XP, including UV-induced mutagenesis and skin cancer (Johnson et al., 1999a; Masutani et al., 1999). Unlike in XP, patients with CS suffer from progressive neurodevelopmental defects along with an early (at birth) or late (late childhood or even adulthood) onset of growth defects and cachexia, microcephaly, retinal degeneration, and a reported life expectancy of ~12 years. The CS pathological symptoms appear more severe in the case of the CS variant Cerebro-oculo-facioskeletal syndrome (COFS; life expectancy <2 years) or considerably milder as in the UV sensitive syndrome where the stability of CSB is compromised (UVsS; affected proteins: CSB, CSA, UVSSA) (Laugel, 2013; Nance and Berry, 1992; Spivak, 2005). Intriguingly, although CS patients manifest with cutaneous photosensitivity, they have no predisposition to skin cancer. The lower incidence of tumors in CS results from the absence of UV-induced mutagenesis rather than from enhanced lethality (Reid-Bayliss et al., 2016). Similar to CS, the clinical phenotypes of TTD patients are remarkably diverse in terms of severity and extent. TTD patients are photosensitive and present with the characteristic Sulphur-deficient, dry and sparse hair and nails of TTD, congenital ichthyosis, physical and mental retardation and signs of premature aging but no cancer (Lambert et al., 2010). In certain instances, the pathological features in TTD are uncoupled from the DNA repair defect. For example, TTD patients carrying mutations in the *TTDN1* gene are DNA repair-proficient, show no signs of photosensitivity and are not cancer-prone. However, they manifest with most other clinical symptoms associated with the photosensitive TTD disorder in addition to delayed bone aging and an over-representation of seizures (Heller et al., 2015; Itin et al., 2001; Nakabayashi et al., 2005). A recent study on a TTD causative TFIIEβ mutation also showed that the GG-NER and TC-NER subpathways of NER were not affected further uncoupling any transcription defects in TTD from defects in DNA repair (Theil et al., 2017).

1.3 The heterogeneity of DNA repair-deficient syndromes: beyond NER. Other DNA repair pathways are not exempted from the bewildering complexity of clinical manifestations seen in NER. Patients with mutations in ataxia telangiectasia-mutated gene (AT; affected gene: *ATM*) that are defective in the repair of DSBs, present with progressive cerebellar ataxia to oculocutaneous telangiectasia, gonadal sterility, growth retardation and a high incidence of lymphoid tumors. In fact, AT is gradually perceived to represent a highly heterogeneous syndrome with dramatic pleomorphic manifestations (Teive et al.,

2015). Likewise, Nijmegen breakage syndrome patients (defective also in DSB repair; affected gene: *NBS1*) are characterized by short stature, microcephaly, and characteristic facial appearances and are particularly prone to respiratory tract infections and highly susceptible to B-cell lymphomas (Chrzanowska et al., 2012). Fanconi anemia patients (defective in the repair DNA ICLs; affected genes *FANC*, *BRCA2*) are a clinically highly heterogeneous group but shares a number of phenotypic similarities with NBS; the great majority of FA patients manifest with major birth defects ranging from short stature and developmental delay to cutaneous, skeletal, craniofacial, and genital anomalies along with bone marrow failure and myelodysplasia. Nearly 80% of children and adults with FA are reported with a least one endocrine defect, including GH deficiency, abnormal glucose or insulin metabolism and dyslipidemia or hypothyroidism (Petryk et al., 2015). Patients with Bloom (BLM; affected gene: *BLM* involved in DNA replication and repair) or Rothmund-Thomson (affected protein: *RECQL4* involved in DNA replication and repair) syndromes all exhibit growth defects and pathological features ranging from cataracts and diabetes to renal dysfunction and immunodeficiency. Instead, patients carrying defects in mismatch repair (as in Lynch syndrome) or BER are -with the exception of colorectal cancer- asymptomatic (Tiwari et al., 2015; Wallace et al., 2012).

1.4 NER-deficient mouse models. Mice with engineered mutations in NER genes mimic in some respects the pleiotropic and heterogeneous pathological symptoms seen in NER syndromes (Hasty et al., 2003). The GG-NER-deficient *Xpc*$^{-/-}$ mouse or the complete NER defective *Xpa*$^{-/-}$ or *Xpc*$^{-/-}$/*Xpa*$^{-/-}$ double mutant animals present with a higher frequency of UV-induced skin cancers but show none of the progeroid CS features or the progressive neurodegenerative complications associated with XP (Nakane et al., 1995; Sands et al., 1995; van der Pluijm et al., 2006). The TC-NER-defective *Csb*$^{m/m}$ animals *associate with a mild phenotype* (van der Horst et al., 1997) (similar to the UVsS syndrome) and need to be crossed with either *Xpa*$^{-/-}$ or *Xpc*$^{-/-}$ mice (i.e., *Csb*$^{m/m}$/*Xpa*$^{-/-}$ or *Csb*$^{m/m}$/*Xpc*$^{-/-}$ mice) to reliably recapitulate the severe CS variant COFS (van der Pluijm et al., 2006). *Xpd*^{TTD} and *Xpd*^{XPCS} mice each carrying a causative point mutation in *Xpd* gene fail to show the complete clinical spectrum of CS features associated with TTD and XP/CS, respectively (Andressoo et al., 2006; de Boer et al., 2002). Interestingly, *Xpf* $^{-/-}$, *Xpg*$^{-/-}$ or *Ercc1*$^{-/-}$ mice are perhaps the only single NER mutant animals (also defective in other DNA repair systems) to reiterate the severe developmental abnormalities and premature aging features of patients with mutations in the corresponding genes (Niedernhofer, 2008; Shiomi et al., 2004; Tian

et al., 2004). Indeed, most other single mutant animals *e.g.* *Xpc*$^{-/-}$, *Csb*$^{m/m}$ or *Xpa*$^{-/-}$ mice result in phenotypically healthy animals with minor to moderate pathological symptoms.

Current models are based on the differential roles of NER factors in *e.g.* TC-NER, GG-NER, ICL or DSB repair and the transcription-coupled-associated BER to explain the complex and heterogeneous constellation of clinical symptoms in NER (Marteijn et al., 2014). Indeed, the myriad of different types of DNA lesions involved as well as the overlapping and distinct functions of NER proteins in other DNA repair systems could provide a valid explanation for the progressive and heterogeneous symptoms seen in NER syndromes. From a genetic point of view, however, it may be difficult to reconcile why the complete NER-defective *Xpa*$^{-/-}$ animals show none of the clinical features seen in CS or XP (with the exception of skin cancer) or why these animals differ substantially from *Csb*$^{m/m}$/*Xpc*$^{-/-}$ that are like *Xpa*$^{-/-}$ mice also defective in both subpathways of NER. It is also puzzling that the causative TC-NER defect in *Csb*$^{m/m}$ mice aggravates immensely the phenotype of the NER-deficient *Xpa*$^{-/-}$ animals as in *Csb*$^{m/m}$/*Xpa*$^{-/-}$ mice. Lastly, the accumulation of random DNA lesions in the mammalian genome may well explain the frequent but unsystematic occurrence of skin tumors seen in *Xpa*$^{-/-}$ animals or the progressive onset of neurological symptoms seen in some, but not all, XPA patients over time (DiGiovanna and Kraemer, 2012; van Steeg et al., 2001). However, the model of indiscriminate DNA damage contrasts with the unvarying onset and severity of developmental abnormalities seen in all isogenic animals carrying the same inborn mutation in NER. The reported frequency (~70.000 DNA lesions per day per cell in humans) (Lindahl and Barnes, 2000) and the presumed indiscriminate nature of DNA insults in the mammalian genome makes it unlikely that all genes poised for transcription are damaged during mammalian development or that the same genes are, at all times, selectively damaged.

Evidently, the genetic-phenotypic paradox seen in NER reflects the diversity of DNA lesions encountered, the position of the DNA damage in the genome, the proliferative status and type of cells involved and likely any putative functions of NER factors outside the canonical NER pathway. Indeed, the wider spectra of recurrent developmental abnormalities seen in some, but importantly not all, defects in NER may be better rationalized when one considers a model where i. NER factors function in DNA damage-associated processes other than or in conjunction with NER, ii. DNA damage events are often programmed and required for the proper execution of *e.g.* developmental gene

expression programs and iii. targeted DNA damage events are topologically restricted to specific genomic regions affecting transcription activation or regulation (Garinis et al., 2008).

1.5. NER factors and transcription. Transcription during mammalian development requires profound and dynamic changes in chromatin organization and structure; in certain instances, such as during zygotic genome activation, the process of mRNA synthesis may be so abrupt leading to genome instability (Butuci et al., 2015). Steric effects on *e.g.* promoter regions or restrictions in the abundance of shared factors amongst the distinct DNA-templated processes suggest that DNA repair and transcription are competing processes. In yeast, whereas transcription is inhibited in the presence of active NER, increased transcription has no detectable effect on NER *in vitro* (You et al., 1998). Transcription factors may also interfere with NER resulting in an increased rate of DNA mutations at the protein binding sites (Sabarinathan et al., 2016). In line, NER is impaired within the DNase I-hypersensitive site center of active promoters leading to a higher density of mutations compared to enhancers (Perera et al., 2016). Conspicuously, however, the relative repair rates of bulky DNA lesions gradually increase in gene promoters reaching a peak at transcription start sites and around ORF structures in yeast (Yu et al., 2016). Nonetheless, direct mechanistic links between NER and transcription are well documented for nearly three decades (Compe and Egly, 2016; Schaeffer et al., 1993). The basal transcription factor TFIIH is known to facilitate initiation, promoter escape, and early elongation steps of RNAPII transcription upon DNA damage (Chalut et al., 1994; de Boer et al., 2002). In TFIIH, the *Xpb* and *Xpd* genes encode two helicase subunits (Oksenych et al., 2009; Schaeffer et al., 1994; Schaeffer et al., 1993) that unwind DNA in the vicinity of DNA damaged sites for repair or on promoters to facilitate the synthesis of the primary transcript (Egly and Coin, 2011). More recently, the NER factors i.e., XPC, CSB, XPA with RPA, XPG and the ERCC1-XPF complex were found to be sequentially recruited to the promoters of nuclear receptor genes upon transcription stimulation *in vitro*; promoter recruitment occurred in the absence of any exposure to exogenous genotoxic insults and was sensitive to transcription inhibitors (Le May et al., 2010a; Le May et al., 2010b). Conspicuously, however, $Csb^{m/m}$ or $Xpa^{-/-}$single mutant livers show –unlike the $Csb^{m/m}/Xpa^{-/-}$ double mutant mice- minimal differences in gene expression (van der Pluijm et al., 2006). In line, a recent study on multiple human cell lines revealed that, for at least XPA, the impact of NER on transcription may only be limited to a small subset of genes

(Manandhar et al., 2017). In developing livers, ERCC1-XPF was found to interact with multiple TFIID subunits i.e. TATA-associated factors and the TBP *in vitro* (Kamileri et al., 2012d) and *in vivo* (Chatzinikolaou et al., 2017) and to assemble together with RNAPII and the basal transcription machinery at promoters of growth genes during postnatal hepatic development. In support, the DNA repair-deficient *Ercc1*$^{-/-}$ animals shared genome-wide gene expression similarities with those of transcription-defective (but otherwise DNA repair-proficient) *Taf10*$^{-/-}$ mice. Using a series chromatographic and purification steps, the XPC/RAD23B/CETN2 was also shown to be part of a multi-subunit stem cell coactivator complex (SCC) required for co-activating Oct4/Sox2-dependent transcription of *Nanog* for stem cell pluripotency (Fong et al., 2011). Unlike what one would expect, the effect of *XPC* or *RAD23B* knockdown was more profound in the reprogramming of mouse embryonic fibroblasts than in the maintenance of ES cells. Similar to HeLa cells (Le May et al., 2010b), the complex in transcription was distinct from that involved in DNA repair; the XPC-SCC could still co-activate *Nanog* transcription even when XPC was mutated abolishing GG-NER and independently of DNA binding likely through interactions with sequence-specific transcription factors, including TFIID. In a follow up work using 3D single-particle reconstruction by electron microscopy, the regions of contact between the XPC complex and OCT4, SOX2, XPA, and TFIIH were proposed (Zhang et al., 2015). More recently, XPC was shown to facilitate TDG-mediated DNA demethylation during somatic cell reprogramming indicating a distinct but complimentary mechanism by which XPC could further impinge on gene expression independently of NER (Ho et al., 2017). Nevertheless, the functional role of XPC in stem cell pluripotency contrasts with the lack of developmental defects seen in *Xpc*$^{-/-}$ mice (Sands et al., 1995). In line, ablation of the C-terminal region of *Xpc* gene abrogating the interaction sites of XPC with RAD23 and CETN2 has minimal impact on gene expression or pluripotency to contribute chimeric embryos (Ito et al., 2014) . Besides transcription initiation, XPG-TFIIH also interacts with factors involved in transcription elongation; *XPG* knockdown dampens epidermal growth factor-induced *FOS* transcription whose mRNA levels are also decreased in XP-G/CS cells carrying a truncated XPG (Narita et al., 2015). Unlike other examples suggesting a role of NER in stimulating transcription, XPB may also inhibit transcription initiation by interfering with promoter melting; inhibition is only relieved by its own ATPase activity (Alekseev et al., 2017) revealing an additional level of transcriptional regulation. NER factors may also affect gene expression in ways beyond their putative role in transcription. For example, earlier studies using a *Xenopus* expression library revealed that XPG may be targeted to

distinct demethylation sites replacing methylated cytosines with unmethylated nucleotides to relieve epigenetic silencing (Barreto et al., 2007). Irrespectively of the proposed mechanisms mentioned above, defects in NER may also selectively interfere with the transcription of genes *in cis*. For example, persistent transcription-blocking DNA lesions may obstruct preferentially the transcription of genes with longer primary transcripts and thus with a greater risk of RNAPII blockage at damaged sites. In line, the majority genes with significantly decreased mRNA levels in *Ercc1*$^{\Delta/-}$ mouse livers also have substantially longer primary transcript lengths (Vermeij et al., 2016). Since it is mostly introns rather than exons contributing to the total gene size, it is thought-provoking that the mRNAs of genes with shorter introns are also more abundant (Castillo-Davis et al., 2002).

1.6 DNA repair factors in transcription: beyond NER. Unanticipated roles in transcription have also been described for proteins involved in the repair of DNA breaks, DNA ICLs, damaged bases or G/T mismatches. Key components of the DSB repair pathway, such as the BRCA1 may repress or stimulate hormone receptor activities at the transcriptional level (Fan et al., 1999; Park et al., 2000; Zheng et al., 2001). Likewise, the monoubiquitinated form of FANCD2 (FANCD2-Ub) involved in the repair of DNA ICLs activates the transcription of tumour suppressor gene TAp63; the recruitment of FANCD2-Ub to the regulatory region of the TAp63 promoter requires SLX4, a subunit of many structure-specific endonucleases that resolves DNA secondary structures generated during DNA repair and recombination (Park et al., 2013a). Proteins involved in BER are no exception; A/P endonuclease Ref-1 is a multifunctional protein that stimulates redox control-dependent DNA binding of transcription factors, including P53, NF-κB, and HIF-1 (Gaiddon et al., 1999; Thakur et al., 2015). Likewise, thymidine DNA glycosylase (TDG) is known to interact with oestrogen receptor alpha to stimulate the transcription of estrogen-regulated genes (Chen et al., 2003). TDG forms a physical and functional complex with the transcriptional coactivators CBP and p300 for both the excision step of repair and histone acetylation (Tini et al., 2002). TDG also interacts with DNA methyl transferase Dnmt3a to inhibit its methylation activity (Li et al., 2007b) and is required for recruiting p300 to retinoic acid-regulated promoters (Cortellino et al., 2011). The multiple roles of TDG in gene regulation could also explain why abrogation of TDG is -unlike other DNA glycosylases- embryonically lethal (Cortazar et al., 2007; Cortazar et al., 2011). In essence, several DNA repair proteins function, similar to NER factors, in DNA-templated processes other than the repair of DNA damage.

1.7 NER factors and the chromatin architecture. The great majority of genes involved in developmental-stage and tissue-specific gene expression programs are found in gene-dense or repressive chromatin regions; their cis-acting regulatory sequences, however, are often located several kilobases apart with one or several unrelated genes in between (de Laat and Grosveld, 2003). The timely expression of a gene relies on the formation of an “active chromatin hub” (de Laat and Grosveld, 2003) to allow the transcription unit to physically communicate with its cis-regulatory elements. Such chromatin hubs depend on protein complexes to loop out any intervening sequences or inactive genes, thereby bringing promoters and *cis*- or *trans*-acting sequences into physical proximity. The CCCTC-binding factor CTCF is a highly conserved 11-zinc finger protein that is involved in transcription activation and repression, chromatin architecture, genome imprinting, X chromatin inactivation and insulation (Phillips and Corces, 2009). Recent data suggest that distinct NER factors associate with chromatin organizers and remodelers to shape chromatin architecture for optimal gene expression, developmental gene silencing or for promoting the accessibility of DNA repair proteins to sites of DNA damage. For instance, XPG and XPF are essential for establishing CTCF-dependent chromatin looping between the promoter and terminator of the activated RARb2 gene (Le May et al., 2012). Abrogation of XPG or XPF due to gene silencing or mutations in their catalytic sites substantially hampered the recruitment of CTCF in chromatin, the formation of appropriate chromatin loops and the fine-tuning of *RARb2* mRNA synthesis. As gene looping and CTCF recruitment occur in parallel with the formation of DNA breaks, it may be attractive to consider a model where XPG and ERCC1-XPF trigger the formation of targeted DNA nicks on promoters to facilitate the access of transcription factors on promoters and the process of mRNA synthesis. Intriguingly, CTCF also interacts with the CSB (Lake et al., 2016). CTCF and CSB regulate each other’s chromatin association in response to oxidative stress; indeed, whereas oxidative stress enhances the CSB-CTCF interaction in 293T cells, CSB itself facilitates CTCF-DNA interactions *in vitro* and regulates CTCF-chromatin interactions in cells treated with menadione, an inducer of oxidative stress. Most recently, an *in vivo* biotinylation tagging approach in mice revealed that the heterodimer ERCC1–XPF complex interacts with CTCF, the cohesin subunits SMC1A and SMC3 and with MBD2 (Chatzinikolaou et al., 2017). The CTCF-cohesin complex colocalized with ERCC1-XPF and the chromatin remodeler ATRX at the promoters and control regions of imprinted genes in the developing liver. Interestingly, abrogation of ERCC1 in *Ercc1*$^{-/-}$ mice

or exposure of primary embryonic fibroblasts to the DNA interstrand crosslinker mitomycin triggered the localization of CTCF to heterochromatin and the dissociation of the CTCF-cohesin complex and ATRX from promoters and imprinted control regions. Besides the possible role of ERCC1, XPF and XPG in chromatin looping and architecture, the findings also suggest that CTCF (and likely other chromatin organizers) are also (in)directly involved in DNA damage repair in ways that are currently undetermined. The crosstalk between NER and chromatin organizers could also explain how, in spite of the low mobility of TFIIH in cortex neurons, the protein is still capable of responding rapidly to new transcription demands or local DNA damage during differentiation or lineage-specific gene expression programs (Giglia-Mari et al., 2009).

Conversely, chromatin architectural proteins whose primary function lies well beyond DNA repair are now known to actively participate in DDR and/or to form an integral part of DNA repair mechanisms. The chromatin-associated scaffold attachment factor SAFB1 was recently linked to DNA damage signaling (Altmeyer et al., 2013). SAFB1 cooperates with histone acetylation allowing the spreading of DNA damage-associated γH2AX marks and for making chromatin permissive to DNA damage signaling. Mobilization of the heterochromatin protein 1 (HP1) involved in chromatin organization, gene silencing, DNA replication, and transcription is also known to initiate DDR (Ayoub et al., 2008; Goodarzi et al., 2008). Interestingly, various HP1 isoforms were recently shown to recruit to sites of UV-induced DNA lesions independently of GG-NER or to photo-oxidative DNA lesions (Luijsterburg et al., 2009; Zarebski et al., 2009). Finally, the cohesin multi-protein complex is perhaps one of the best-known examples of how a protein complex involved in sister chromatid cohesion (Peters et al., 2008) has pivotal roles in transcription and DNA repair (Mehta et al., 2013). In humans, the cohesin consists of four core subunits i.e. SMC1, SMC3, STAG1 and STAG3 that form a ring-like structure able to encircle and keep the two sister chromatids tightly connected. In doing so, the DSB in one sister chromatid is repaired through HR by using the undamaged chromatid as a template. Besides, cohesin is known to activate the DNA damage-induced intra S phase and G2-M checkpoints (Watrin and Peters, 2009; Yazdi et al., 2002). Interestingly, the MRX (Mre11/Rad50/Xrs2) complex that recognizes DSBs promotes the assembly of cohesin at DNA break sites (Unal et al., 2004).

1.8. Physiological DNA damage events during mammalian development. DNA damage occurs often deliberately and for a defined biological purpose during mammalian

development. For instance, DNA DSBs are purposely formed during meiotic homologous recombination (HR), a fundamental process that is conserved in all species. Meiotic HR is to facilitate genetic diversity when cells divide to become specialized gamete cells while maintaining genome stability; the process is initiated by the formation of DNA DSBs during G2/leptotene (Zickler and Kleckner, 2015). During HR, the resection of the 5'-terminated DNA strand is catalyzed by the evolutionary conserved Topo II-like enzyme Spo11 protein and accessory factors which remain covalently attached to the 5'end of the broken DNA (Cejka, 2015). In the homology search and strand invasion step of homologous DNA, the invading 3' end of the broken DNA molecule then "invades" on template DNA duplex (Li and Heyer, 2008). In other instances, DNA breaks are intentionally generated during V(D)J recombination (Boboila et al., 2012); DNA breaks occur in developing B- and T-lymphocytes to provide the basis for the antigen-binding diversity of the immunoglobulin and T-cell receptor proteins. The DNA DSBs are generated at the border of the signal sequence by a site-specific nuclease composed of the RAG1 and RAG2 proteins (Schatz and Swanson, 2011), and the DSBs are subsequently repaired by proteins that also function in the repair of DSBs that have been generated by mutagenic agents. Likewise, "programmed" DNA lesions followed by error-prone DNA repair are required to enhance antibody diversity by triggering somatic hypermutation of immunoglobulin variable genes (Di Noia and Neuberger, 2007). Hypermutation is triggered by the Activation-Induced cytidine Deaminase (AID), a unique enzyme that deaminates cytosines into uracils in Ig genes. DNA replication across the deaminated cytosines (i.e. uracil) results into transition mutations at C:G pairs. Alternatively, UNG uracil-DNA glycosylase may create abasic sites by excising uracil leading to transversions (Di Noia and Neuberger, 2002). Mutations at A:T pairs are thought to result from a mutagenic patch repair mechanism that is triggered by MSH2/MSH6-mediated recognition of the U:G lesion in which polymerase eta plays a major role owing to spontaneous polymerase error leading to base mispairing (Di Noia and Neuberger, 2007). Ironically, scheduled DNA damage events are now also known to play pivotal roles during transcription to ensure the optimal execution of developmental gene expression programs. Indeed, when transcriptional activity is inert, chromosomes are thought to remain condensed preventing the easy access of transcription factories to the DNA template. Instead, when transcription needs to be activated e.g. during the execution of developmental gene expression programs, the formation of scheduled DNA breaks is often required to rapidly relax the DNA double helix to facilitate access of transcription factors and the RNAP II at sites of transcription initiation. TOP2A and TOP2B, the two

human class II topoisomerases efficiently resolve DNA topologic constraints by binding two separate segments of DNA, generating a DNA DSB in one of the segments and passing one DNA segment through the transient DSB of the segment, in an ATP-dependent fashion (Deweese and Osheroff, 2009). TOP2A and TOP2B share ~7-% homology and are encoded by separate genes; whereas TOP2A is required for resolution of topologic constraints arising during replication and chromosomal segregation, TOP2B is expressed independently of the proliferative status of the cell and cannot complement TOP2A function in mammalian cells (Deweese and Osheroff, 2009). Recent findings suggest that TOP2B plays a pivotal role in transcription-induced DSB formation. TOP2B is known to associate with gene promoter regions upon activation of gene transcription by nuclear receptors generating transient, site-specific dsDNA breaks (Ju et al., 2006); TOP2B and poly(ADP-ribose) 1 (PARP-1) co-recruit in the vicinity of DNA break formation on promoters. In addition, TOPO2B is required for efficient androgen-mediated gene expression; androgen signaling triggers the coordinated recruitment of AR and TOPO2B along with the formation of TOP2B-dependent DSBs to the targeted enhancers and promoter regions (Haffner et al., 2010). Together these findings provide a mechanistic link between TOP2B-dependent dsDNA breaks and components of the DNA damage repair machinery in regulating gene transcription. In other instances, chromosomes require to dynamically reorganize themselves for the proper execution of complex transcription programs during development or in response to specific signals. The formation of site-specific DNA DSBs is known to facilitate such chromatin changes; for instance, the androgen receptor recruits the ligand and genotoxic stress-induced enzymes, including the AID and LINE-1 repeat-encoded ORF2 endonuclease to intronic binding sites near specific tumor translocation sites; the latter facilitates the formation of DNA DSBs which are then repaired by the error-prone NHEJ pathway (Lin et al., 2009). Consistent with the notion of programmed DNA damage events aimed for optimal gene expression, DNA repair factors involved in NHEJ or DSB repair, including the Ku70, Ku80, PARP1 and the DNA-PK are known to associate with transcription factors and recruit on regulatory elements of target genes upon transcription induction (Haffner et al., 2011). DNA breaks are not the only documented type of DNA lesions that play a physiological role during transcription initiation; efficient transcription de-repression or induction often requires cyclical DNA demethylation by DNA methylstrasferases upon activation by estrogen (Metivier et al., 2008) or oxidation triggered by H3K9me2 demethylation (Perillo et al., 2008). Besides mammals, recent findings in worms also reveal that chromosomal damage is tightly linked

to gene expression (Butuci et al., 2015); whereas RAD-51 foci are not typically detected on the transcriptionally quiescent X chromosomes, these are prominent when transcription is active on the X. Moreover, unlike with splicing or translation, inhibiting RNAPII by chemical or genetic means led to the significant reduction in RAD-51 foci in Z2/Z3 coupling RNAPII activity to the generation of DSBs at sites of ongoing transcription to facilitate RNAPII promoter escape or efficient elongation. Longer tracts of RNA–DNA hybrids, known as R-loops have also attracted recent attention as they appear to liaise between DNA damage events and transcription regulation (Costantino and Koshland, 2015). R-loops are formed during transcription when the nascent RNA molecule emerging from ongoing transcription hybridizes with the DNA template forming an RNA-DNA hybrid structure that displaces single-stranded DNA. Importantly, R-loops arise naturally in mammalian cells to *e.g.* facilitate immunoglobulin class switching (Yu et al., 2003), to protect promoters from CpG island methylation by DNMT3B1, the primary de novo DNA methyltransferase during early development (Ginno et al., 2012) or to facilitate transcription termination of genes over G-rich pause sites (Skourti-Stathaki et al., 2011). Typically, R-loops are resolved by RNaseH as the enzyme degrades the RNA molecule in RNA-DNA hybrids (Wahba et al., 2011), or by Senataxin, a specialized helicase known to efficiently unwind RNA-DNA hybrids (Mischo et al., 2011; Skourti-Stathaki et al., 2011). The increase of R-loops in cells defective for RNA biogenesis factors support the notion that these proteins suppress the formation of R-loops by binding to the nascent RNA (Huang et al., 2012; Li et al., 2007a). Alternatively, topoisomerase I may ensure genomic instability by preventing the conflict between transcription and DNA replication through the suppression of RNA-DNA hybrids formation; topoisomerase I reduces the torsional stress behind RNAPII, thereby preventing the annealing of the nascent RNA with the DNA template (Tuduri et al., 2009). Naturally occurring R-loops during transcription are thought to promote DNA damage by exposing long stretches of ssDNA leading to spontaneous formation of DSBs but also to transcription-associated mutagenesis (Muers, 2011; Wimberly et al., 2013). Interestingly, transposable genetic elements often alter gene function and promote genomic rearrangements as well as variation and likely also organismal speciation; to do so they utilize DNA or RNA intermediates and possess transposase and endonuclease activity, respectively to create a dsDNA break for the new insertion site (Mills et al., 2006). Taken together, it becomes increasingly profound that DNA damage events may not be as random as previously thought; in fact, they are unavoidable events during mammalian development and –in spite of the obvious risks- vital for organismal survival.

1.9 NER and transcription-associated DNA damage. It may seem perplexing that DNA repair enzymes in NER are recruited on promoters of actively transcribed genes (Le May et al., 2010a; Le May et al., 2010b) or that potentially threatening DNA cutters, such as the ERCC1-XPF or XPG are required for the optimal fine-tuning of gene expression (Kamileri et al., 2012b; Kamileri et al., 2012d). Some, not necessarily exclusive, possibilities exist that may rationalize these findings. For instance, as TFIIH functions in both NER and transcription, the recruitment of NER factors on promoters could reflect the binding or physical proximity of these proteins to TFIIH during transcription. In this scenario, the transient, and in this context unproductive, recruitment of NER proteins to promoters will correspond to the amount of TFIIH bound on these regions and by analogy to the level of transcription activity. A previous genome-wide expression analysis in P15 Xpd^{TTD} livers, however, revealed distinct gene expression changes from those seen in P15 $Csb^{m/m}/Xpa^{-/-}$ or $Ercc1^{-/-}$ mice (Niedernhofer et al., 2006a; van der Pluijm et al., 2006). Moreover, Xpd^{TTD} animals are -unlike the $Csb^{m/m}/Xpa^{-/-}$ or $Ercc1^{-/-}$ mice- not growth-defective (Kamileri et al., 2012d). Nonetheless, Xpd^{TTD} mice are expected to retain some residual XPD activity in transcription and thus this possibility cannot be excluded. The presumed function of NER in transcription also contrasts with the fact that abrogation of XPA, CSB or XPC in mice or human cells has hardly any significant effect on transcription or murine development (Fong et al., 2011; Manandhar et al., 2017; Niedernhofer et al., 2006a; van der Pluijm et al., 2006). As a counterargument, the synergistic function of specific NER proteins in transcription activation -as it is also proposed for the DNA repair function of various double mutant animals *e.g.* $Csb^{m/m}$-$Xpc^{-/-}$ mice cannot be excepted. Moreover, as NER factors are not detected in unrelated genomic regions upstream of the transcription start site, it is unlikely that they may recruit circumstantially to promoters due to their DNA-binding ability. The remaining possibilities suggest that NER factors associate with promoters to facilitate the process of mRNA synthesis or to repair DNA lesions specifically on these sites. Although the relevance of NER to transcription-associated DNA breaks has yet to be shown, it is worth noting that DNA breaks (likely related to R-loops) are detected in UV-irradiated cells of XPD/CS patients and, surprisingly so, at sites distant from the DNA damage itself (Berneburg et al., 2000). Inhibition of RNAPII transcription with transcription inhibitors dramatically reduced the number of UV-induced breaks. Moreover, the DNA breaks were dependent on functional NER pointing to its likely involvement in the induction of these lesions at sites of ongoing transcription (Theron et al., 2005). Besides targeted

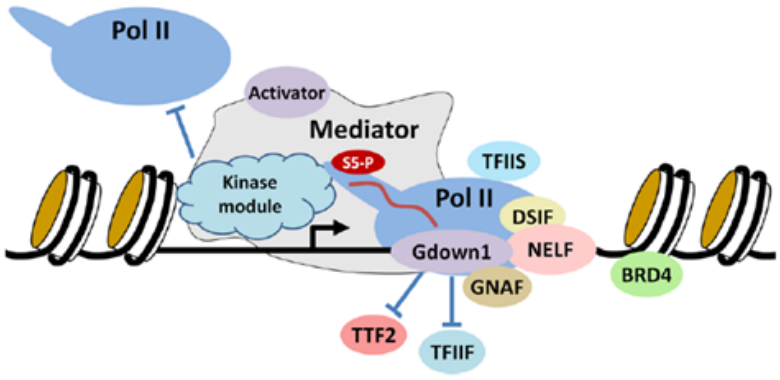
DNA DSBs on promoters, longer tracts of RNA-DNA hybrids, known as “R-loops” liaise between spontaneous DNA damage events and transcription (Costantino and Koshland, 2015). Naturally occurring R-loops generated on promoters or termination regions are frequently formed when a nascent RNA molecule is hybridized with the DNA template before the two strands of the DNA duplex reanneal leaving the non-template DNA single-stranded (Skourti-Stathaki and Proudfoot, 2014; Skourti-Stathaki et al., 2011). R-loops expose long stretches of ssDNA leading to the spontaneous formation of DSBs or to transcription-associated mutagenesis (Muers, 2011; Wimberly et al., 2013). Transient DNA-RNA hybrids are found in organisms from bacteria to humans and play physiological roles in *e.g.* transcription (Aguilera and Garcia-Muse, 2012) or class switching recombination (Yu et al., 2003) but are likely also generated when an RNAPII is stalled at transcription-blocking DNA lesions (Sollier et al., 2014). The NER endonucleases ERCC1-XPF and XPG actively process R-loops (Sollier et al., 2014). Intriguingly, R-loop processing requires CSB (involved in TC-NER) but not XPC (involved in GG-NER). XPF (the nuclease active site of ERCC1-XPF) and XPG are thought to generate nicks or gaps on both the transcribed and non-transcribed strand and at both ends of the R-loop. At present, it remains unclear whether the apparent cleavage of R-loops by XPG is directly related to NER. Similarly, cleavage of R-loops by ERCC1-XPF may require functional NER (where targeting would be mediated by *e.g.* XPA, or TFIIH), or else the ERCC1-XPF may be targeted by MSH2-MSH3 or SLX4 outside the canonical NER mechanism. In this regard, it would be informative to test whether, R-loops are successfully cleaved in *e.g.* XPA-defective cells. Arguably, the notion that R-loops processed by NER factors could potentially trigger the formation of cytotoxic DSBs is counter-intuitive for factors involved in genome maintenance. Nevertheless, R-loops also accumulate in cells deficient in the RNA/DNA helicase Aquarius and XPG (Sollier et al., 2014) suggesting that distinct NER proteins may remove harmful R-loops formed when the ongoing transcription or mRNA processing is compromised (Stirling et al., 2012).

1.10 TFIIS & transcription-blocking DNA lesions. The process of mRNA synthesis requires the action of sequence-specific DNA binding factors, the basal transcription machinery and chromatin remodeling and modification enzymes (Lemon and Tjian, 2000; Orphanides et al., 1996). Together, these factors create a chromatin environment that allows the synthesis of the primary transcript (Ohler and Wassarman, 2010). In eukaryotes, protein expressing genes are transcribed by RNAPII to generate messenger

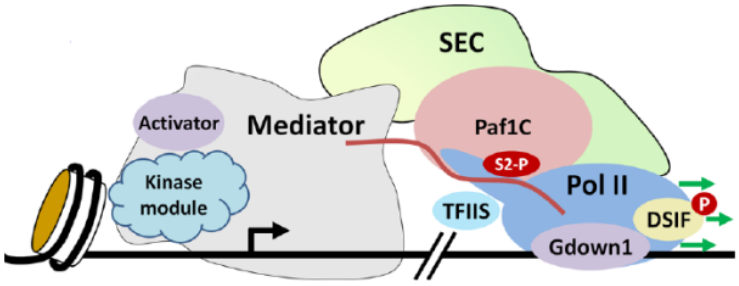
RNAs (mRNAs) that have a rather short half-life and are quantitatively and qualitatively regulated throughout production and degradation (Thomas and Chiang, 2006). RNAPII transcription consists of distinct events from the generation of a single mRNA product up to its recycling (Shandilya and Roberts, 2012. At the beginning of the process, DNA elements, located near the promoter elements, are recognized by the general transcription factors that together constitute the pre-initiation complex (PIC) and form the backbone for RNAPII recruitment (Soutoglou and Talianidis, 2002). This is followed by the Mediator, a multiprotein complex that functions as a transcriptional coactivator in all eukaryotes and serves as a transcription platform (Malik and Roeder, 2005). Transcription bubble formation, by unwinding of the DNA double helix, sets the onset for initiation as the template strand translocates to the polymerase active center for RNA synthesis to commence from the transcription start sites (TSS) (Miller and Hahn, 2006).

Before RNAPII passes onto productive elongation, it is forced to pause at a short distance from the promoter during this intermediate stage of promoter proximal pausing (Adelman and Lis, 2012). Under specific activation conditions, the polymerase evades from the paused state and enters productive elongation through the act of P-TEFB. During productive elongation, RNAPII transcribes with speeds up to 4kb/min, while newly synthesized RNA is polyadenylated and spliced duting the process of maturation (Danko et al., 2013). Termination of transcription then leads to polymerase recycling and to its general release from chromatin (Guo and Price, 2013). Gene expression regulation is the means by which gene functionality is coordinated based on what, where and when it is needed. It occurs in two levels: originally with histone modification events, e.g. phosphorylation and acetylation that grand accessibility to the DNA template (Heintzman et al., 2009) and subsequently with transcription factors (TFs) during the initiation and elongation stages of mRNA synthesis. TFs interact with co-activators or co-repressors, while the Mediator acts as a processing platform for the overall signal generated by these factors (Guo and Price, 2013). TFs are then set for transcription initiation control by recruiting co-activators, including the Mediator complex, P300 and the general TFs (Poss et al., 2013; Soutourina, 2018). Enhancer elements bind to the TFs and act in synergy for gene expression regulation mainly through the identification of genomic regions whosesequence affects protein binding and eventually gene expression (Guo and Price, 2013).

Promoter proximal pausing maintains transcription in an active state by keeping the promoters open for the recruitment of transcription factors (Krumm et al., 1995) (**Figure 2**). TFIIS is one of the most widely characterized elongation factors. Three different forms of TFIIS have been identified in mammals. A ubiquitously expressed form encoded by the TCEA1 gene on mice chromosome 1, a testis-specific form encoded by TCEA2 on chromosome 2 and a third one, encoded by TCEA3 on chromosome 4 (Shema et al., 2011). TCEA2 tissue expression profile manifests a moderate to high protein localization in many tissues including the liver, while differences are observed among the paralogues (Uhlen et al., 2015). Paralogues share a high percentage of homology in their amino-acid sequence; for instance, TCEA2 shares 93% homology with its human orthologue on chromosome 20 (BLAST generated data). TFIIS stimulates arrested RNAPII to cleave the nascent transcript during ongoing transcription. In doing so, the transcription elongation factor TFIIS rescues RNAPII from backtracking by altering the 3-D structure of the polymerase in order for the newly synthesized RNA to be directed in the polymerase active center for cleavage and the formation of a new 3' prime end that will allow rebooting (Donahue et al., 1994). Promoter proximal pausing has a prominent role during developmental gene transcriptional regulation as well as in stimulus-controlled pathways, including the heat shock family of genes (Shao and Zeitlinger, 2017). In doing so, RNAPII pausing may achieve activation of specific sets of genes in a time- and cell-type dependent manner.



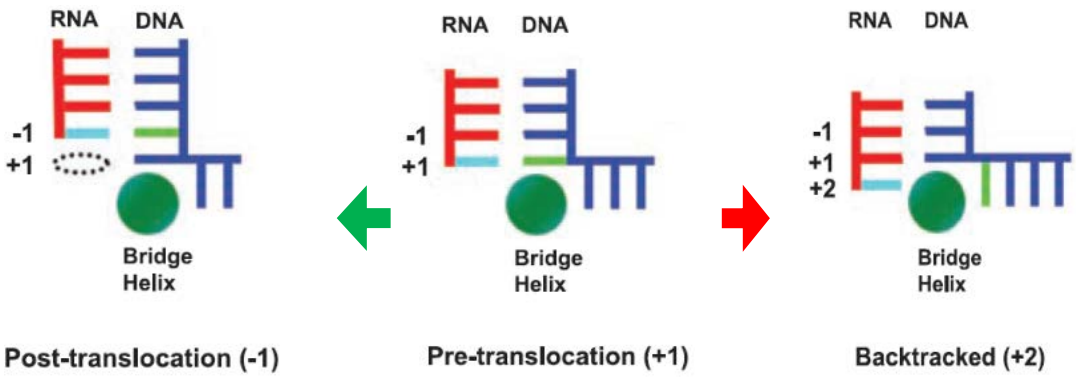
P-TEFB



Guo and Price, 2013

Figure 2. Promoter Proximal Pausing: *Coordinated actions of promoter binding elements, along with the negative elongation factor NELF and DSIF, keep RNAPII in a paused state (Chen and Xie, 2017). Knocking down NELF led to gene expression reduction due to nucleosomes masking promoter regions. Elongation rebooting is triggered by the positive transcription elongation factor PTEF-B that phosphorylates NELF, at the serine 2 of RNAPII carboxy-terminal region that abandons RNAPII, while DSIF is transformed into a positive elongation factor (Price, 2018; Li, et al., 2017).*

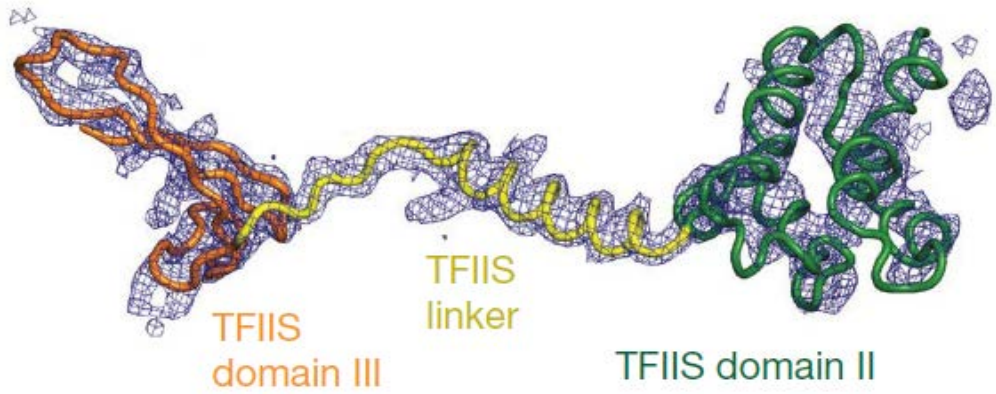
Upon DNA damage, RNAPII stalls and backtracks when it encounters a DNA lesion opening up a time window for the recruitment of the TC-NER factor CSB; in doing so, it also provides the necessary space for the recruitment of the remaining NER factors at DNA damaged sites (Donahue et al., 1994; Selby et al., 1997). Using photolyases, which specifically recognize and repair UV-induced lesions, previous studies have shown that RNAPII backtracking (**Figure 3**) is required for damage recognition (Tornaletti and Hanawalt, 1999). Similar to the promoter proximal pausing, TFIIS results in the cleavage of the newly synthesized RNA following template strand direction towards the polymerase active center (Donahue et al., 1994) (**Figure 4**).



Wang et al., 2009

Figure 3. Elongation complex states*: Three transcribing complexes are observed during elongation. A pre-translocation state, in which the nucleotide just added to the growing RNA chain is still in the nucleotide addition site; from there on the enzyme must move forward on the template, to make the nucleotide addition site available for entry of the next nucleotide at the post-translocation state. In the case of DNA damage or obstacles that prevent RNAPII progression, the enzyme retreats on the template, an action which is called backtracking and lead to the extrusion of the 3'-end of the RNA (Brueckner et al., 2007; Wang et al., 2009). Backtracking by one or a few residues is reversible, whereas backtracking a greater distance (the greatest backtracking distance observed is as far as 7-9 nucleotides), leads to arrest, from which recovery is only possible by cleavage of the transcript in the polymerase active center, which is induced by TFIIS (Wang et al., 2009).*

RNAPII stalling and backtracking is likely part of a wider mechanism in which the polymerase originally stops at the damaged site. If legion bypass is not an option, then the enzyme slides back on the template, allowing time for the recruitment of the repair factors while it also defines the boundaries of the excised DNA fragment. If recruitment of DNA repair factors is not possible due to e.g. defective NER or defective polymerase rescuing, then the enzyme is arrested. In this scenario, RNAPII in the presence of TFIIS recovers from backtracks. Finally, although the role of TFIIS in RNAPII elongation is well established, some studies have suggested that TFIIS may also function during early events in transcription, including PIC formation (Wery et al., 2004)as well as during promoter proximal pausing (Guglielmi et al., 2007; Kim et al., 2007; Reines et al., 1989; Wery et al., 2004).



Cheung and Cramer, 2011

Figure 4. TFIIS domains: *TCEA2 is a member of the TFIIS family of transcription elongation factors along with the two other paralogues, which are genes that derive from the same ancestral gene but now reside at different locations within the same genome. TFIIS has four distinguishable structural domains that fold independently and are common among the paralogues. Domain II is responsible for the TFIIS binding to RNAPII while C-terminal two-thirds of the protein, that is its domain II and III separated by a 15-aa linker, are responsible for the cleavage-stimulating activity (Cheung and Cramer, 2011). Domain III directs the RNA to the RNAPII active site for cleavage*

through its zinc formation. No known functions have been accounted to Domain I (Cheung and Cramer, 2011).

Several studies led by the notion that backtracked RNAPII allows for the damaged site to be revealed and repaired have focused on the coupling of DNA repair and TFISS-mediated rebooting of transcription elongation. Synthetic lethality screens, followed by co-immunoprecipitation studies revealed that BRCA1 interacts genetically with TCEA2 and co-localize at post-UV transcription associated damage foci together with RNAPII; the latter could facilitate damage repair and/or transcription restart (Hill et al., 2014). At present, the functional role of TFIIS in NER has not been fully addressed. Originally, TFIIS was shown to play minimal roles in Transcription Coupled Repair *in vivo* (Verhage et al., 1997). However, more recent crystallography-based structure-function studies have put forward a model in which pyrimidine dimers are rescued by CSB-induced bypass without any conformation changes in RNAPII or exposure of the lesion supporting. These findings indicate that stalled RNAPII complexes are rescued by TFIIS while complexes that stall at a non-bulky lesion are rescued by CSB-induced lesion bypass (Brueckner et al., 2007). Intriguingly, abrogation or overexpression of TFIIS has minimal impact on the growth of yeast cells under normal condition. However, inactivation of mouse TFIIS and, specifically of TCEA1, causes embryonic lethality (Ito et al., 2006). Pull-down and yeast two-hybrid assays revealed direct interaction of tissue-specific transcriptional activator FESTA to TFIIS suggesting that TFIIS participates in tissue- and gene-specific transcription regulation (Saso et al., 2003). TFIIS works in conjunction with the Mediator by altering the catalytic properties of RNAPII in order to advance across the +1 nucleosome (Nock et al., 2012). Given the reported role of TFIIS in relieving RNAPII pausing and facilitating transcriptional elongation in distinct genomic regions, it is interesting that overexpression of TFIIS alone can lead to selective upregulation of RNF20-suppressed genes, such as *Il-8* and *c-myc* (Shema et al., 2011) whereas RNF20 itself, likely through H2B ubiquitylation, is known to repress transcription elongation through inhibition of TFIIS recruitment to chromatin (Shema et al., 2011). Gene regulation activities have also been attributed to the other members of the TFIIS family. For instance, TCEA3 may limit the pluripotency differentiation potential of mouse embryonic stem cells (Park et al., 2013b). Overall, TFIIS-driven adjustments of the RNAPII 3D conformation are essential in transcription activation, elongation and reboot after backtracking. Moreover, the presence of three TFIIS paralogues allows for distinct expression patterns and tissue specificities in mammals.

1.11 Introducing the *in vivo* biotinylation tagging in NER. To gain insight into the contribution of NER factors in the regulation of gene expression during development, we recently isolated and characterized XPF protein complexes during postnatal liver development in mice. Traditionally, the purification of such complexes involves several pre-purification steps making the procedure laborious and costly. We, therefore, established an *in vivo* biotinylation tagging methodology as a simple approach for the efficient direct purification of ERCC1-XPF protein partners from crude nuclear extracts (in use in the Garinis lab) (Driegen et al., 2005; Grosveld et al., 2005). As a starting point, we tagged XPF by fusing the 14aa biotinylation tag to XPF protein and expressing the biotinylation tag-fused XPF protein in established cell lines expressing the BirA biotin ligase knocked-in the Rosa26 locus. BirA specifically recognizes and biotinylates the short tag, thus creating a very high affinity “handle” for isolating tagged XPF-bound factors by binding to streptavidin. Given the success of gene targeting in mice and the highly specific, tightly regulated biotinylation occurring only in the presence of BirA biotin ligase, we next considered more physiological applications of a similar strategy. Specifically, we generated a series of XPF knockin mice, by targeting the endogenous locus with a C terminally tagged version of the gene (Figure 5), which is specifically recognized by the bacterial biotin ligase BirA. XPF knockin mice were crossed with BirA expressing transgenic mice, generating the desired double homozygous i.e. bioXPF mice. The system exploits the advantage of one of the highest affinity interactions known in nature, that of avidin/streptavidin to biotin (Beckett et al., 1999; Cull and Schatz, 2000; de Boer et al., 2003; Driegen et al., 2005) allowing a one-step high throughput isolation of XPF bound protein complexes, while limiting the background noise, since it does not depend on specific antibodies. To its premise, XPF was fused with a small 38 a.a. Flag-Tev-Avidin (FTA) tag. The multiregional FTA tag bears the Avi-tag, which can be efficiently in-vitro and in-vivo biotinylated, at the epsilon nitrogen of its Lysine residue, by the BirA biotin ligase enzyme from E.coli (Cull and Schatz, 2000; de Boer et al., 2003; Driegen et al., 2005). It is also consisted of the Flag-tag, that can be used separately or concurrently to the Avi-tag, further elevating purification capacity (Einhauer and Jungbauer, 2001). Excision of Flag-tag from the eluate is possible due to the TEV proteinase cleavage site sequence, located between the two fusion tags, with the use of a recombinant proteinase from the tobacco etch virus (TEV) (Parks et al., 1994; Shih et al., 2005).

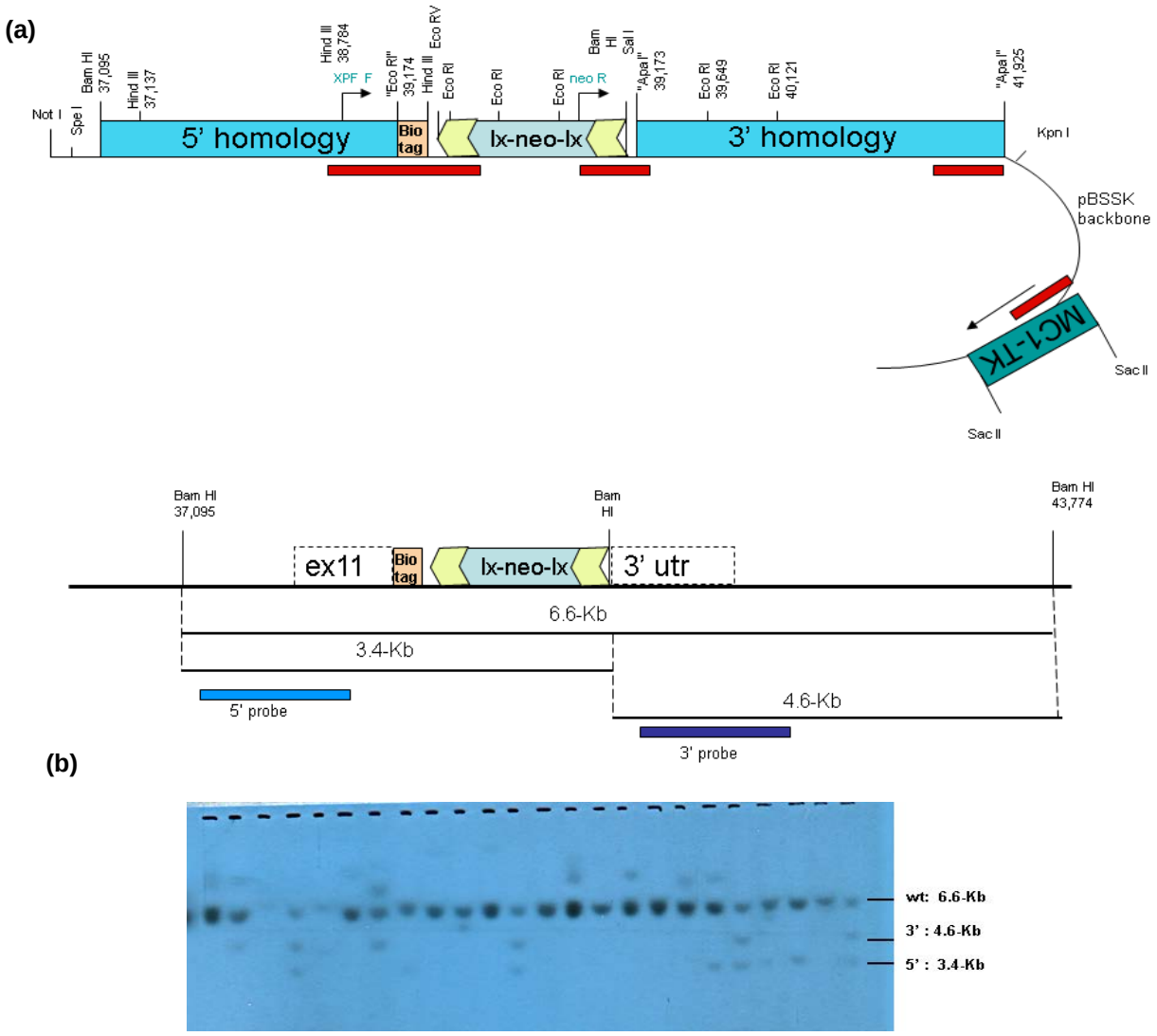


***Figure 5. XPF targeting strategy**: **a)** The Xpf targeting construct is consisted of the 5' and 3' homology regions, the Flag-Tev-Avidin tag (Biotin tag), and two selection cassettes. The homology regions were cloned from 129 mouse Embryonic Stem Cell's genomic DNA. The 2.1 kb 5' region spans the last exon (Ex11) of the Xpf gene up to the stop codon and half of the intron 10-11. The 2.8 kb 3' region spans the rest of Ex11, which corresponds to the 3' untranslated region of the Xpf gene. The synthetically generated Bio-tag, bearing three stop codons, was cloned immediately after the last codon of the XPF's open reading frame, by removing the internal stop codon. Adjacent to the tag a LoxP-NeomycinR-LoxP cassette was cloned that would allow for positive selection of targeted ES clones in the presence of G418 antibiotic. Negative selection cassette MC1-Thimidine-Kinase, cloned outside the targeting region, becomes lethal for the cells that have incorporated the targeting locus in their genome by random integration and not by homologous recombination, in the presence of gancyclovir. **b)** Southern Blot for the selection of positively targeted 129 ES cells.*

Generated *Xpf* targeting constructs were electroporated in 129 ES cells that were selected with G418/gancyclovir for 9 days. Southern Blot analysis of the genomic DNA, digested with BamH, allowed the selection of the correct clones (**Figure 5b**). The use of two 5' and 3' region radioactive probes, were expected to generate three distinguishable bands. Both probes hybridize with the wt. allele and give a 6.6 kb corresponding band, while two 3.4 kb and 4.6 kb bands are present in the targeted allele, to the sum of three, due the extra BamHI site. Further analysis by Western Blot performed on positive ES's nuclear extracts, probed with anti-FLAG antibody, revealing one 110-kDa band corresponding to the Bio-tag fused XPF.

Following injection of selected and expanded ES clone 41 into 22 mouse blastocysts, of the C57 black coat background, they were implanted into two foster mothers that gave rise to three mosaic mice (chimeras) and 8 non-mosaic mice. Since ES cells of the 129 background give rise to brown coat color, because they are wild type for the Agouti allele, the percentage of mosaicism was determined by brown color penetration in the black coat background.

Offspring with a high percentage of chimerism (over 50%) was crossed with C57BL/6 mice, for several generations, in order for the Xpf knock-in allele to be transferred from the 129 genetic backgrounds to the more common C57 background. Germline transmission was determined and positive selection was achieved by genotyping each offspring. No Cre deleter mouse crossing was necessary for the removal of the neomycin cassette, which could influence the expression of the xpf gene (Artelt et al., 1991; Pham et al., 1996), since the ES cells used were from the protamine Cre line that express Cre recombinase only in the germ line (Bunting et al., 1999; O'Gorman et al., 1997).

These inbred XPF knockin mice were crossed with the birA transgenic mice to generate double heterozygous bioXpf$^{+/-}$; birA$^{+/-}$ mice and again with the knockin mice to generate the desired homozygous for the tagged bXPF and birA genotyped bioXPF mice (**Figure 6**). The generation of transgenic mice that express BirA in every cell type (**Figure 7**) and developmental stage, due to the introduction of the gene in the ROSA26 transcription unit that follows this uniformly expressed pattern (Driegen et al., 2005; Friedrich and Soriano, 1991; Zambrowicz et al., 1997), has been decisive for the conception of the project. In a birA biotin ligase expressing background, this approach allows the expression of tagged XPF in physiological levels, as the regulatory elements of the gene remain unchanged.

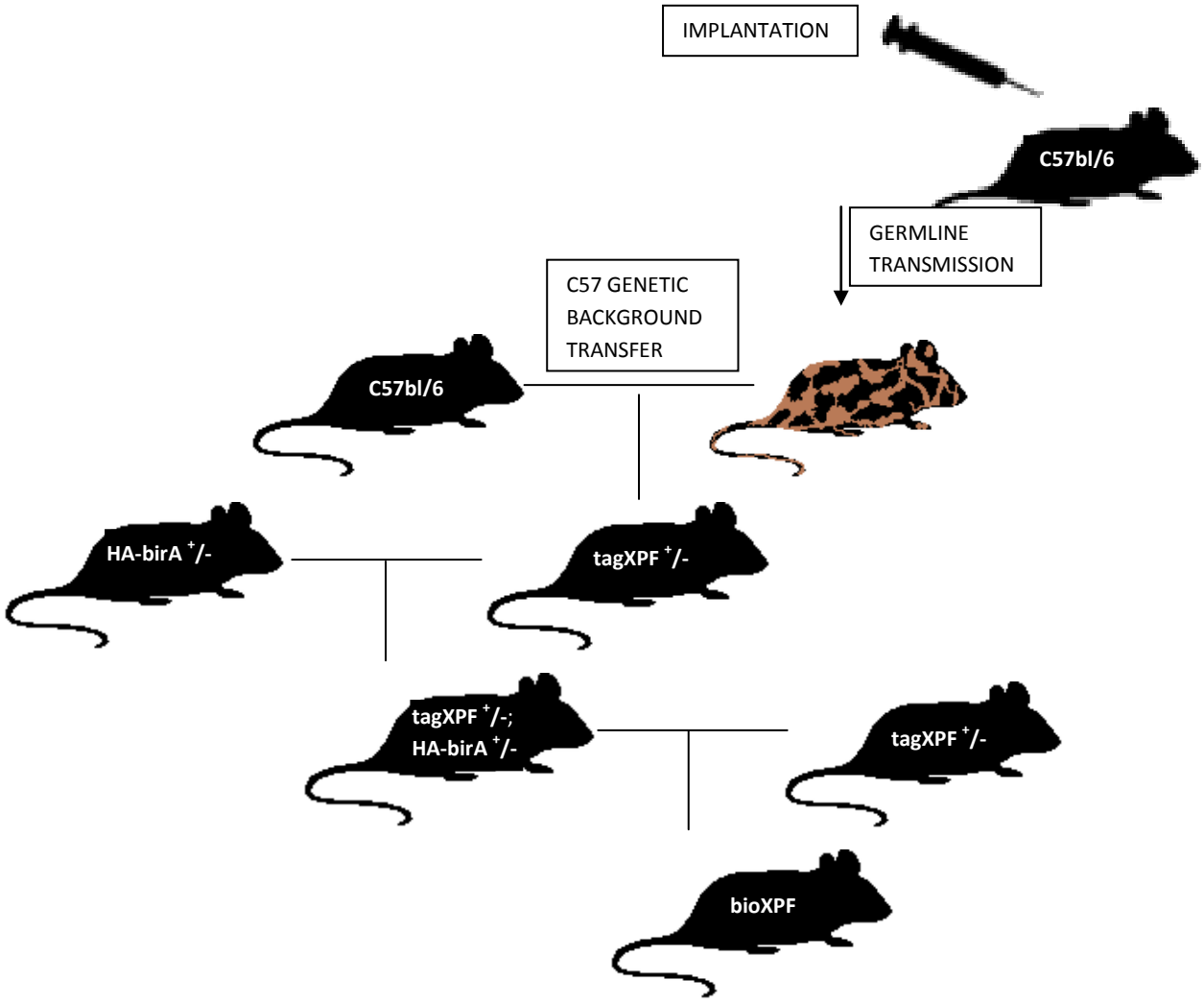
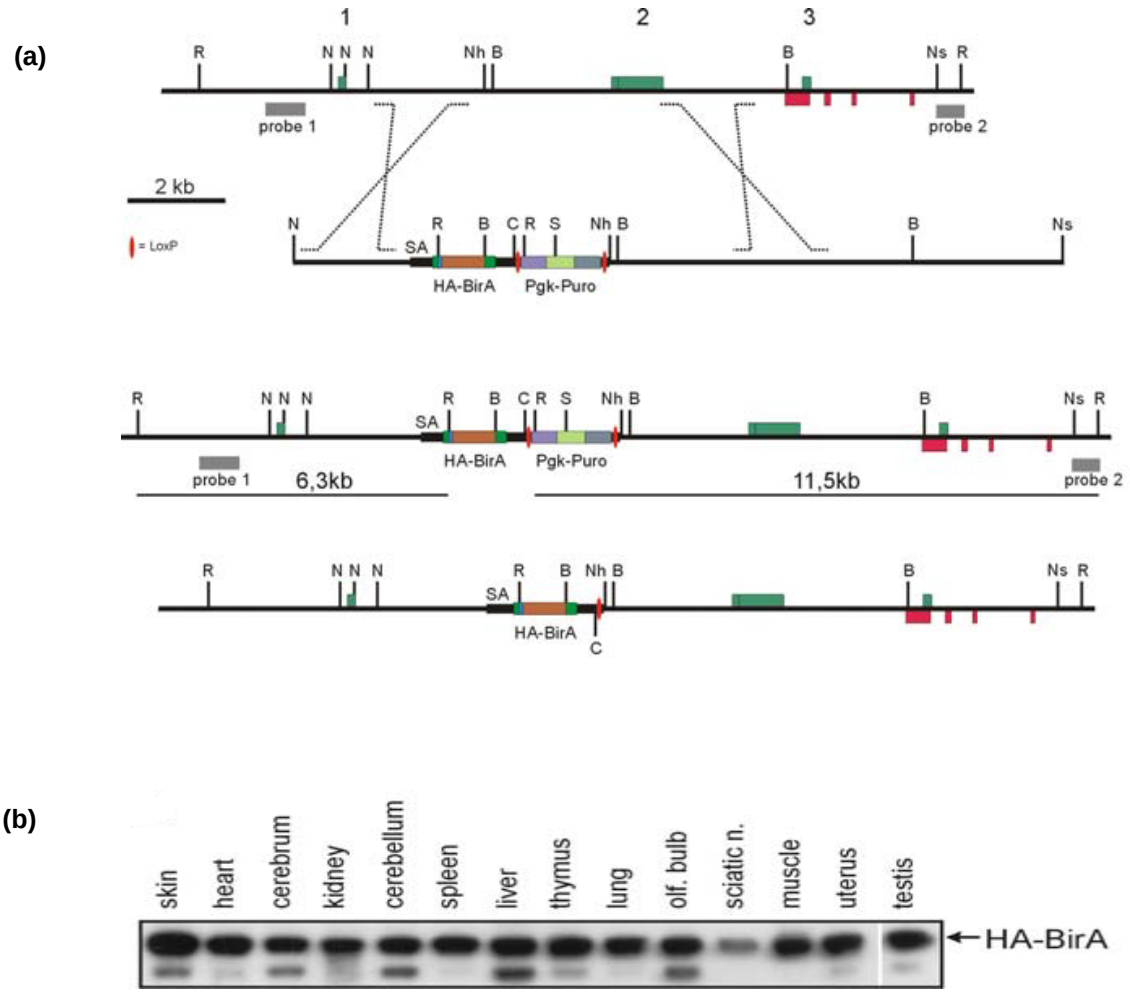


Figure 6. Crossing scheme: *C57 background blastocysts, bearing ES cells with an FTA tagged xpf allele of the 129 background, were implanted into foster mothers. There chimeric offspring were crossed with wt C57bl/6 mice to determine germline transmission of the tagged allele. Generated xpf knocked-in (tagXPF^{+}/-) mice were crossed with C57 mice for two more generations to transfer the allele in the C57 genetic background. Xpf knock-in mice were then crossed with BirA expressing transgenic mice to generate double heterozygotes for the two tagged alleles. There offspring were again crossed with Xpf knock-in mice to generate the homozygous for the two loci bioXPF mouse.*



Adapted from Driegen et al, 2005

***Figure 7. ROSA targeting strategy**: **a)** Initially the triple HA tagged birA gene, that would allow for the identification of the expressed protein via anti-HA, was introduced into the the 3rd exon of the rabbit b-globin gene. The exon was excised and introduced into the ROSA26 targeting construct along with a LoxP-PGKpuro-LoxP positive selection cassette at its 3', covering 13Kb of ROSA26 homologous DNA. Following homologous recombination, the HA birA gene and the selection cassette were inserted into the first intron of the ROSA26 gene and ES cells were probed via Southern Blot. Correct ES cells were injected into mouse blastocysts and implanted into a foster mother. **b)** Their offspring were crossed with a Cre deleter mouse, in order for the puromycin selection cassette to be removed, and Western Blotted for the expression of the 39 kD birA protein in a variety of tissue types.*

2. Aims

Whereas defective NER of damaged DNA has been established as the underlying cause of mutations leading to skin cancer, the links between NER defects and the developmental and metabolic abnormalities seen in NER disorders remain obscure (Garinis et al., 2008; Niedernhofer et al., 2006b; van der Pluijm et al., 2007). Besides DNA repair, earlier studies have shown that distinct NER factors play a role in transcription (Citterio et al., 2000; de Boer et al., 2002) and upon stimulation, certain NER proteins are recruited to active promoters *in vitro* (Le May et al., 2010a; Le May et al., 2010b) and *in vivo* (Kamileri et al., 2012c). These primary observations revealed that the ERCC1-XPF complex functions beyond DNA repair in the transcriptional regulation of genes associated with postnatal growth and differentiation, thus adding a novel piece to the puzzle for understanding the causal basis of complex NER disorders. We, therefore, hypothesize that the so called "segmental" NER progeroid features (Martin, 2005) may reflect the "segmental" transcriptional requirements for certain NER factors during mammalian development.

At present, however, the mechanistic role of NER factors, such as the ERCC1-XPF complex in transcription during development and/or in age-related diseases remains elusive. No solid evidence exists as to whether and how the ERCC1-XPF complex is involved in these fundamental processes, what are the underlying gene networks or which particular cell types are mostly affected by this complex during development or with advancing age. It also remains elusive how NER factors, including the ERCC1-XPF complex and associated protein partners respond to environmental or intrinsic threats, such as the age-related accumulation of DNA damage in our genome. It would be attractive to determine whether, for instance, a progeroid cell carrying defects in genome maintenance or cells exposed to genotoxic threats show a characteristic (aberrant) genome-wide redistribution of NER factors in the mouse genome. Evidently, a comprehensive and multidisciplinary approach is required to decipher the functional relevance of NER factors in transcription beyond DNA repair during development.

In addressing these aims, the present thesis focuses on the following key objectives:

- Identify the ERCC1-XPF-associated protein (sub)complexes and bound-genomic targets in mammalian cells.
- Assess the physiological impact of selected ERCC1-XPF-associated protein partners and bound-genomic targets on transcription.

- Examine the impact of DNA damage on ERCC1/XPF protein complexes and bearing genomic targets in mammals during the process of mRNA synthesis.
- Generate a new series of biotin-tagged TCEA2 knockin animals to dissect the functional role of transcription-blocking DNA lesions in development and disease in mammals.

Until recently, the daunting complexity of the developmental process, the conspicuous lack of tools to study it, and a dearth of experimentally tractable model systems have greatly hindered any testable hypothesis-driven approaches to understand the molecular basis of development, particularly in mammals. Here, we employed a series of functional approaches and state-of-the-art next generation sequencing and proteomics methodologies to provide a pioneering groundwork aiming at dissecting the role of ERCC1-XPF in mammals.

3. Results

3.1 ERCC1-XPF recruits on promoters (genome-wide) upon transcription Induction.

Recent data have shown that NER factors function, in addition to DNA repair, in the transcriptional process by recruiting on specific genomic regions to regulate gene expression (Apostolou et al., 2019; Kamileri et al., 2012a). We recently established the *in vivo* biotinylation tagging of XPF factor in mice (Chatzinikolaou et al., 2017). To gain insight into the genome-wide occupancy of ERCC1-XPF in the mouse genome, we performed a series of chromatin immunoprecipitation assays combined with Illumina-based next

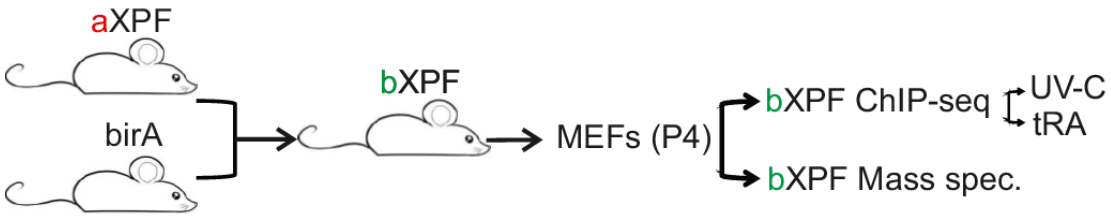


Figure 8. Experimental approach. *Schematic representation of bXPF ChIP-Seq analysis in mouse embryonic fibroblasts (MEFs) derived from bXPF animals expressing the BirA transgene and BirA transgenic mice upon trans-retinoic acid (tRA) or UVC-irradiation.*

generation sequencing approach (ChIP-Seq) on biotin-tagged XPF (bXPF) and the respective BirA control mouse embryonic fibroblasts (MEFs) under basal conditions or upon exposure to UVC-induced DNA damage and transcription stimulation with trans-retinoic acid (tRA), a pleiotropic factor known to activate transcription and regulate gene expression during cell differentiation and embryonic development (Bastien and Rochette-Egly, 2004)(Bastien and Rochette-Egly, 2004) (**Figure 8**). bXPF ChIP-Seq analysis of two independent biological replicates revealed 1100, 1964 and 44 significant peaks in basal conditions, upon transcription activation and UVC irradiation, respectively (**Figure 9A**). The great majority of ERCC1-XPF binding sites mapped on promoters (27%), whereas intergenic (31%) and intronic (24.1%) regions were also substantially represented in basal conditions (**Figure 9Ai**). Upon treatment of cells with tRA, we find that ERCC1-XPF is highly enriched on promoters (44.8%), exons (4.3%) and CpG islands (7.2%), and to a lesser extent on introns (13.9%) and intergenic regions (18.8%) (**Figure 9Aii**) when compared to the untreated control cells. Interestingly, only twenty-two ERCC1-XPF-bound peaks were detected on promoters following 2h post-UVC irradiation (**Figure 10Aiii**). Specifically, 295 and 879 bXPF-bound peaks on proximal promoter regions were identified

in untreated and tRA-treated MEFs respectively, which corresponds to 158 and 683 annotated genes, respectively (**Figure 9B**).

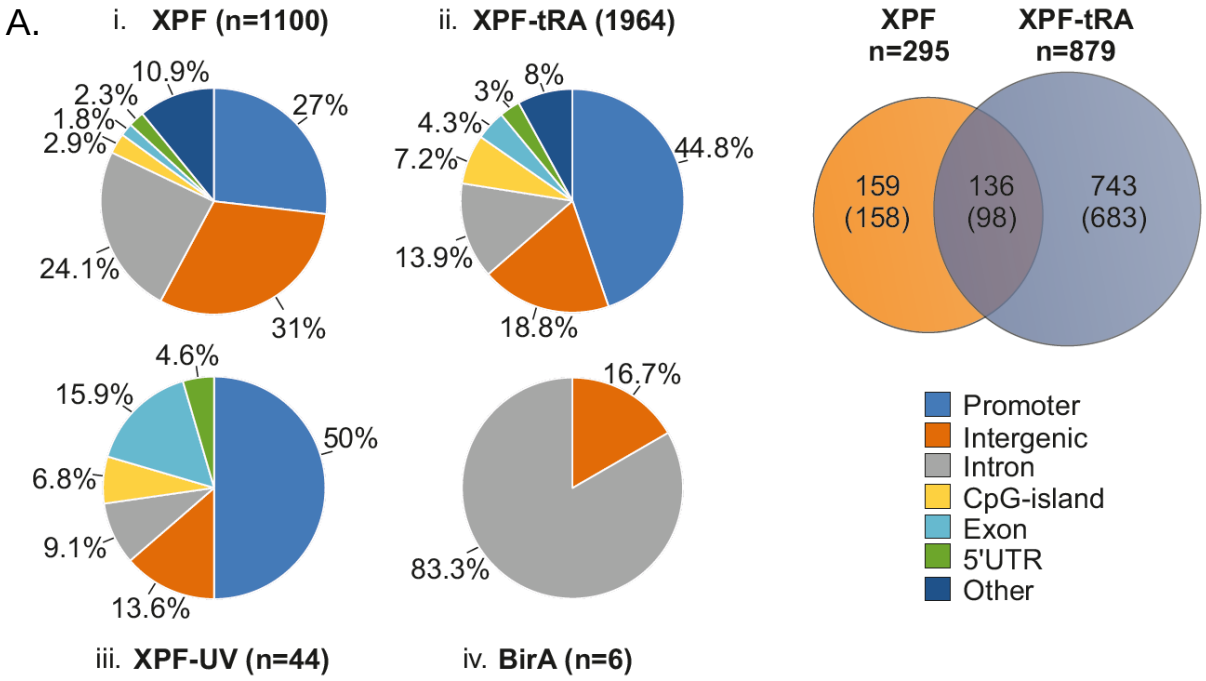


Figure 9. Genome-wide ChIP-Seq analysis of ERCC1-XPF occupancy in MEFs: A*. Pie charts illustrating the genomic distribution of bXPF binding sites in untreated, tRA- and UVC-treated bXPF and BirA (control) MEFs. Peaks occurring within ±2kb of the TSS were considered "promoter regions". **B** Venn diagram of XPF and XPF-tRA ChIP-Seq peaks mapped on promoters and corresponding number of unique genes (parenthesis).*

bXPF ChIP followed by qPCR performed on peak sequences flanking transcription start sites (TSS) of gene promoters (**Figure 10A**) further validated the increased recruitment of ERCC1-XPF on tRA-responsive genes upon transcription activation (**Figure 10B**). Instead, ChIP signals for bXPF were significantly reduced in promoters tested in 2h-post UVC-irradiated MEFs (**Figure 10B**). Thus, upon transcription stimulation, the ERCC1-XPF complex binds preferentially to promoters and is rapidly released upon UVC-induced DNA damage.

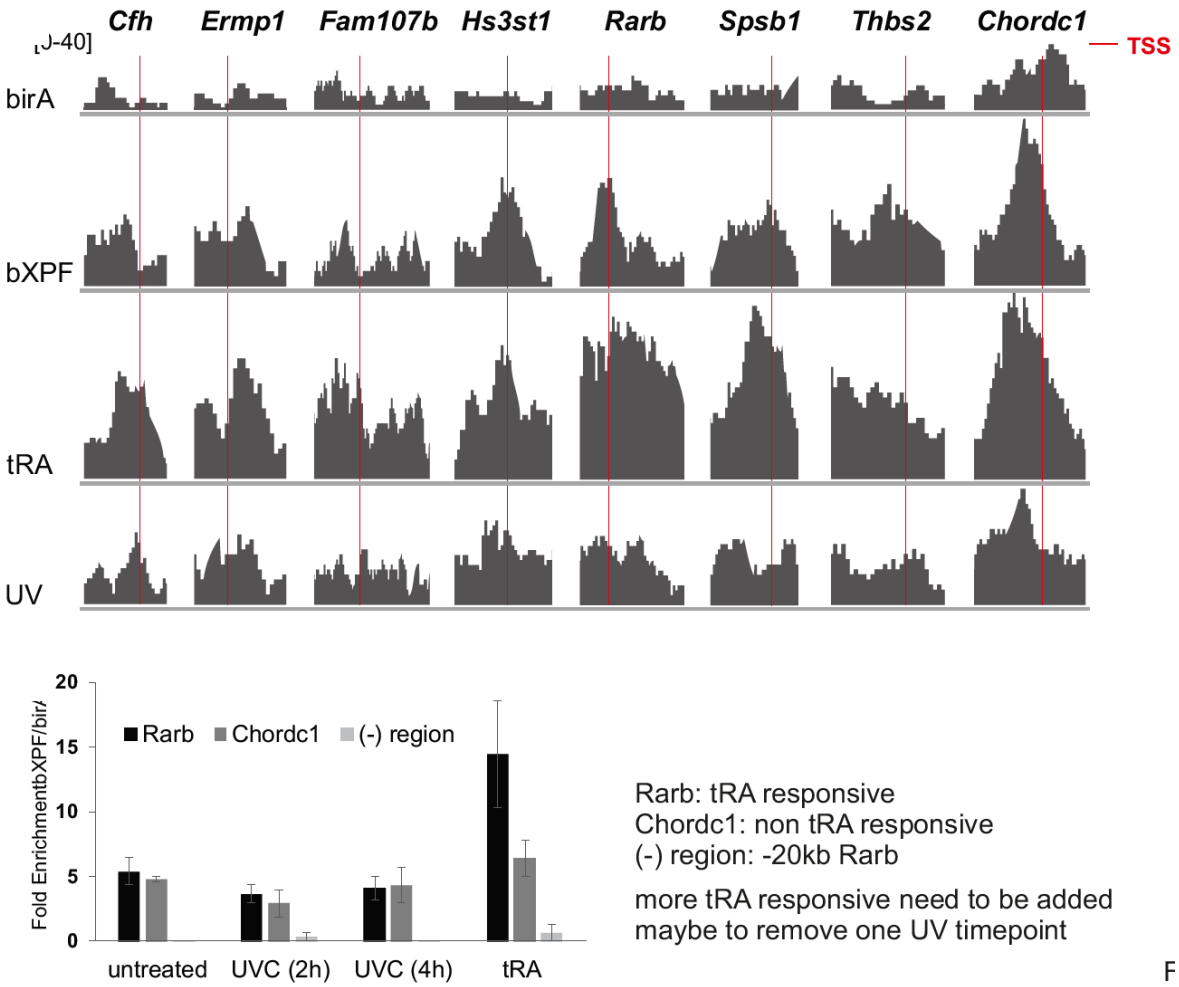


Figure 10. ChIP-Seq distribution: A. *IGV genome browser views depicting bXPF ChIP-Seq signals within ±2kb of the TSS of representative genes under basal conditions and following tRA and UVC treatment.* ***B****. bXPF ChIP signals are shown as fold enrichment of the percentage of input over the percentage of input BirA at the promoter region of the Rarb gene in MEFs.*

3.2 Genome-wide ERCC1-XPF recruitment on DNA & active transcription signatures.

Previous findings revealed that NER factors are recruited with RNAP IIat the promoters of activated genes (Le May 2010). This and the finding that ERCC1-XPF is recruited on promoters, genome-wide, in our ChIP-Seq data prompted us to assess the chromatin state of ERCC1-XPF-bound peaks, as an indication of transcriptional status. To do so, publicly available ChIP-Seq data of transcription markers in MEFs (see Methods) were analyzed. ERCC1-XPF signals upon basal conditions and tRA treatment positively associate with RNAPII binding sites, activating histone H3K4 trimethylation and H3K27 acetylation marks (**Figure 11**). Consistently, the location of ERCC1-XPF peaks does not overlap with the repressive histone H3K4 monomethylation-enriched sites and shows negative correlation with lamina-associated domains (**Figure 11**).

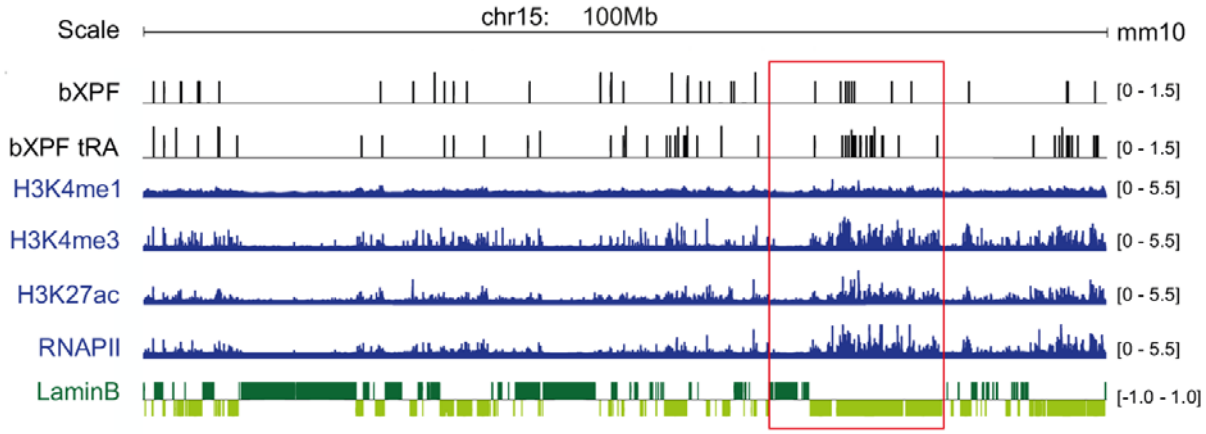


Figure 11. Chromatin state of ERCC1-XPF binding peaks.: *IGV overview of representative genomic region of XPF, XPF-tRA, histone H3 modifications, RNAP II and Lamin B1 ChIP-Seq profiles.*

Moreover, genome-wide correlation analysis between the ChIP-Seq datasets further confirmed a high positive association of ERCC1-XPF with RNAPII and H3K27ac active histone modification signature, whereas a poor correlation score with H3K4me3 and H3K4me1 histone marks was observed (**Figure 12A**). We next compared the genome-wide distribution of ERCC1-XPF peaks, as well as upon tRA, around the transcription start sites (TSS), with the distribution of active H3K4 trimethylation and H3K27 acetylation histone marks near the TSS, as well as of repressive H3K4 histone monomethylation (**Figure 12B**). Consistent to their genome-wide correlation analysis, XPF distribution around the TSS same as upon transcription induction follows active chromatin status. Thus, ChIP signals of ERCC1-XPF upon basal conditions and tRA stimulation are intimately linked to chromatin signatures known to be associated with transcriptional activation

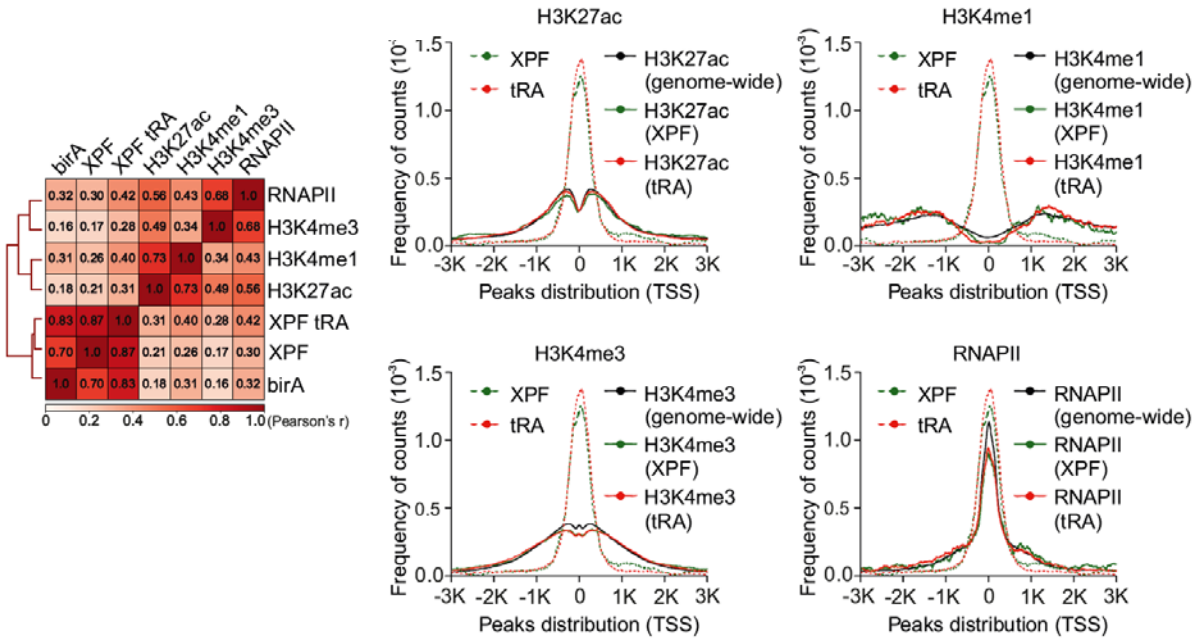
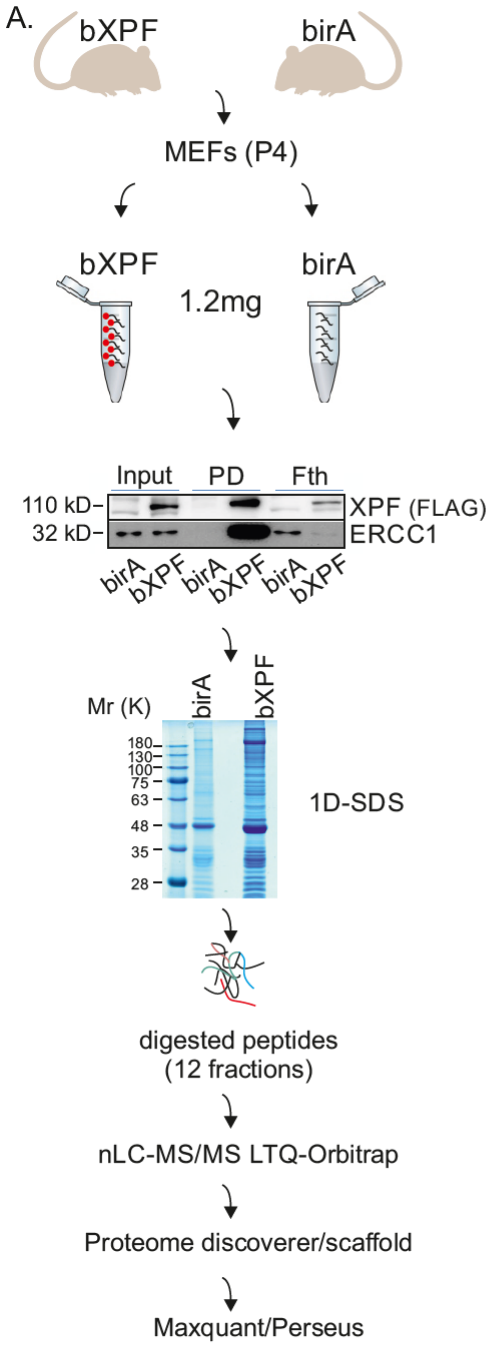


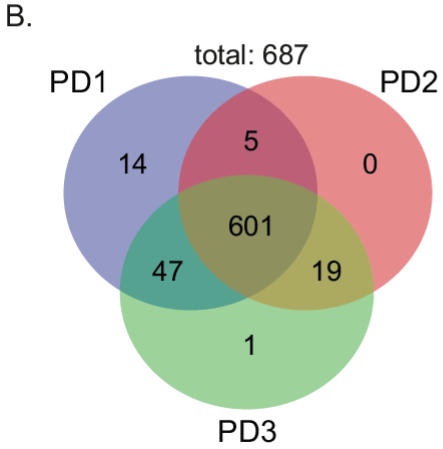
Figure 12. Genome-wide distribution and correlations: A. *Genome-wide Pearson correlation analysis of XPF, XPF-tRA, RNAP II, H3K27ac, H3K4me3 and H3K4me1 epigenetic marks. The coefficient is indicated by the color of each square.* ***B.*** *TSS-centered average ChIP-Seq profile for RNAP II, H3K27ac and H3K4me3 activating histone marks, H3K4me1 repressive histone modification and for genes that bind XPF and XPF-tRA. Dotted lines depict the profiles of XPF and XPF-tRA on whole genome.*

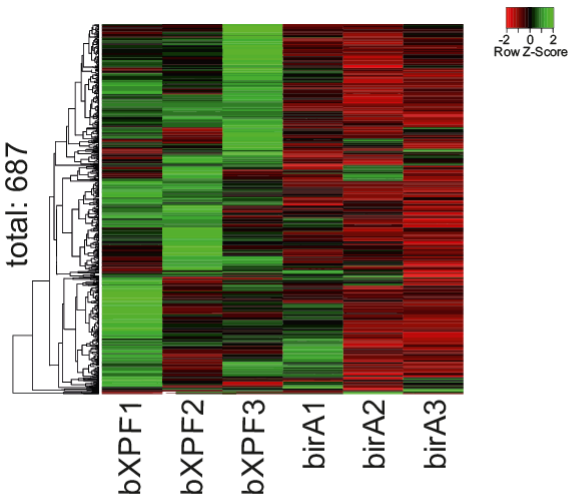
3.3. ERCC1-XPF cooperates/acts/recruits synergistically with DNA topoisomerase IIβ on promoters. To isolate and characterize protein complexes associated with ERCC1-XPF heterodimer in MEFs, we employed the *in vivo* biotinylation tagging approach in combination with a hypothesis-free, high-throughput mass spectrometry approach. To do this, we prepared nuclear extracts from bXPF and birA control MEFs using high-salt extraction conditions (**Figure 13A**). Nuclear extracts were treated with benzonase and RNase A, to ensure that the identified protein interactions are not mediated by DNA or RNA. Nuclear extracts were further incubated with streptavidin-coated beads and bound proteins were eluted and subjected to Western blot analysis confirming that bXPF can still interact with known protein partners involved in NER, that is, ERCC1. Next, we separated the bound protein proteome by one-dimensional SDS-PAGE (~12 fractions) followed by in-gel digestion and peptides were analyzed with high-resolution liquid chromatography-tandem mass spectrometry (nLC MS/MS) on a hybrid linear ion trap Orbitrap instrument (**Figure 13A**). From three biological replicates, we identified a total of 687 proteins with 601 proteins shared in all three measurements under stringent selection criteria (**Figure**

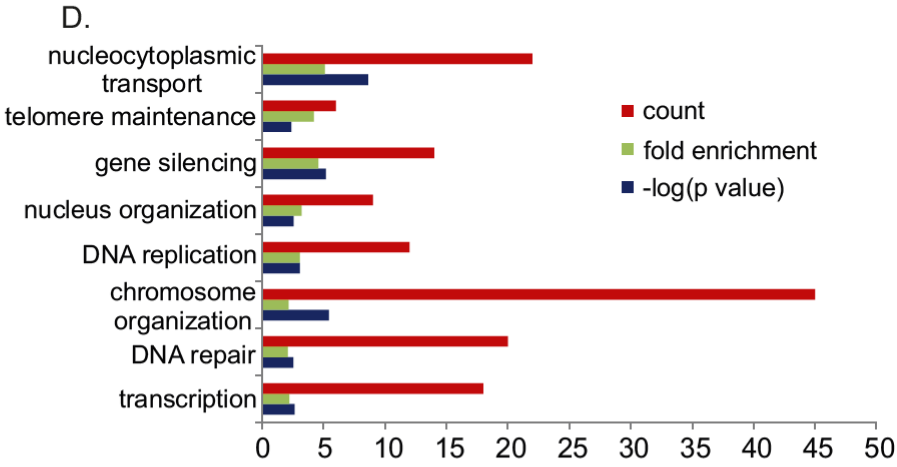
13B; see Methods). Using a hierarchical clustering approach, we confirmed that the 687 bXPF-bound proteins are capable of classifying the bXPF knock-in and BirA transgenic MEFs into the expected groups (**Figure 13C**). To functionally characterize this dataset, we next subjected the 601 shared bXPF-bound proteins to gene ontology (GO) classification. At the confidence interval of [95% (p<0,05)] used, the significant over-represented GO terms found (**Fig. 13D**) involved 101 out of the initial 601 bXPF-bound core proteins. Using this dataset, we were able to discern two major bXPF-associated protein complexes involved in transcriptional regulation and chromosome organization (**Figure 13E and F**). These findings confirm previously documented interactions of ERCC1-XPF with components of the TFIID complex (e.g TAF4A, TAF6, TAF10) and components involved in chromatin organization (e.g CTCF, SMC1A, SMC3, SMARCA5), whilst they reveal interactions of ERCC1-XPF with factors associated with transcriptional dynamics.

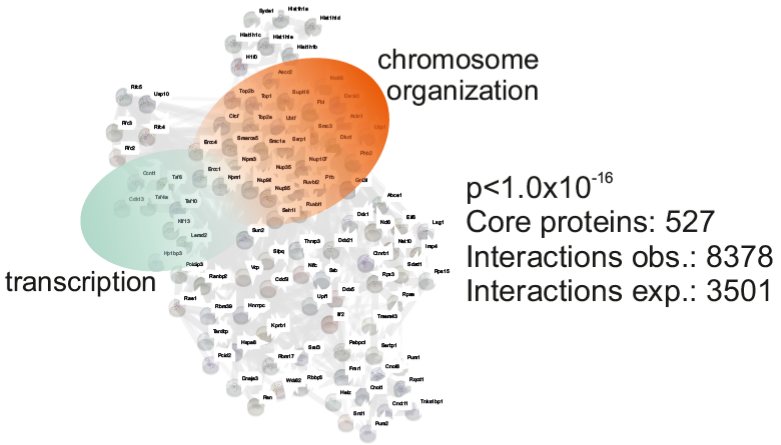
A previous high-throughput mass spectrometry approach on bXPF (MEFs) nuclear extracts revealed that XPF interacts with the TAF4A, TAF6, TAF10 subunits of the TFIID complex as well as with the topoisomerase TOP2B and the chromatin organization factors CTCF, SMC1A, SMC3 and SMARCA5.


A.
bXPF
birA
MEFs (P4)
bXPF
birA
1.2mg
Input
PD
Fth
110 kD–
32 kD–
XPF (FLAG)
ERCC1
birA
bXPF
birA
bXPF
birA
bXPF
Mr (K)
birA
bXPF
180
130
100
75
63
48
35
28
1D-SDS
digested peptides
(12 fractions)
nLC-MS/MS LTQ-Orbitrap
Proteome discoverer/scaffold
Maxquant/Perseus


B.
total: 687
PD1
PD2
PD3
14
5
0
601
47
19
1


C.
-2
0
2
Row Z-Score
total: 687
bXPF1
bXPF2
bXPF3
birA1
birA2
birA3


D.
nucleocytoplasmic transport
telomere maintenance
gene silencing
nucleus organization
DNA replication
chromosome organization
DNA repair
transcription
count
fold enrichment
-log(p value)
0
5
10
15
20
25
30
35
40
45
50


E.
chromosome organization
transcription
p<1.0x10-16
Core proteins: 527
Interactions obs.: 8378
Interactions exp.: 3501

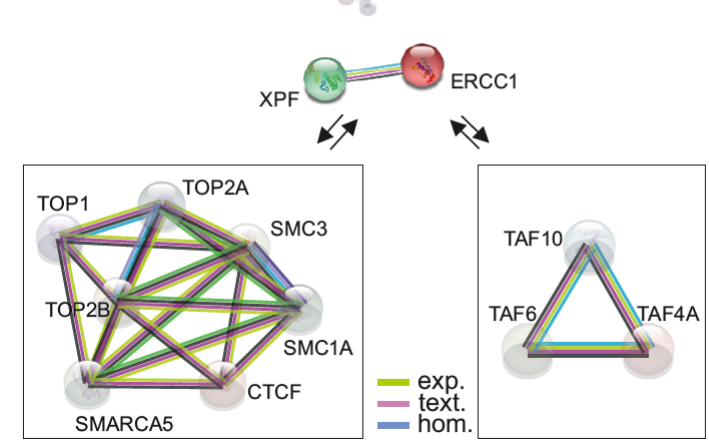

F.
XPF
ERCC1
TOP1
TOP2A
SMC3
TOP2B
SMC1A
CTCF
SMARCA5
exp.
text.
hom.
TAF10
TAF6
TAF4A

Figure 13. ERCC1-XPF interacts with chromatin remodeling and transcription factors. ***(A)****. Schematic representation of the high-throughput MS analysis performed using nuclear extracts from bXPF;BirA and BirA MEFs.* ***(B)****. Venn diagram of bXPF-bound protein factors from three independent pulldowns (PD) and subsequent MS analyses.* ***(C)****. Hierarchical clustering using z-score for normalized value of differentially enriched proteins between the signal intensities of 687 bXPF-bound proteins in three independent bXPF and BirA MEF samples.* ***(D)****. Significantly over-represented biological processes of XPF-bound proteins.* ***(E)****. Number of observed (obs.) and expected (exp.) known protein interactions within the core XPF-bound protein set.* ***(F)****. Schematic representation of two major XPF-bound protein complexes involved based on experimental (exp.) evidence and/or text mining (text) evidence, hom: homologous proteins.*

In this work follow-up pulldown experiments in nuclear extracts of bXPF and control BirA MEFs further confirmed that the endogenous bXPF interacts with TAF4, TAF6, TAF10, TBP and TOP2B (**Figure 14A**). The reciprocity of ERCC1/XPF-TOP2B interaction was further confirmed with co-immunoprecipitation experiments, whereas no interaction of ERCC1-XPF with TOP1 or TOP2A was observed (**Figure 14A and B**). Likewise, TOP2B was able to immunoprecipitate ERCC1, and antibodies against TOP1 and TOP2A verified the absence of the interaction with ERCC1 (**Figure 14C**).

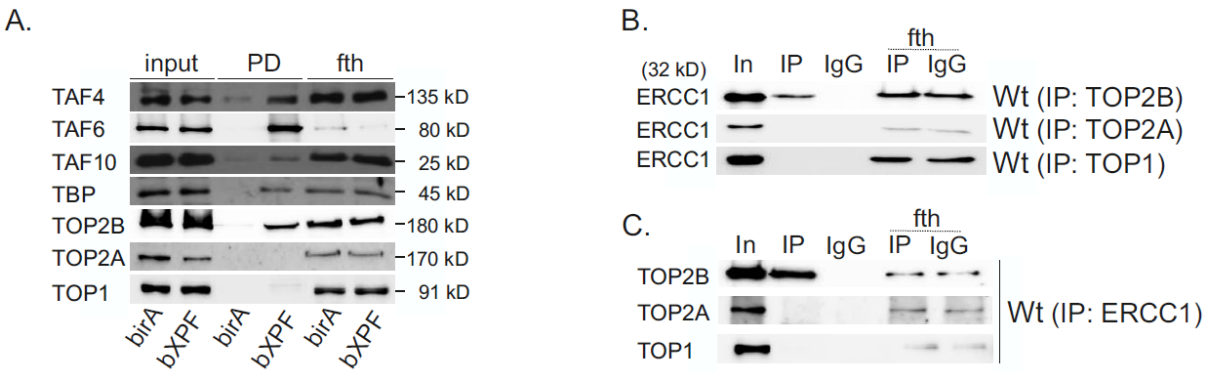
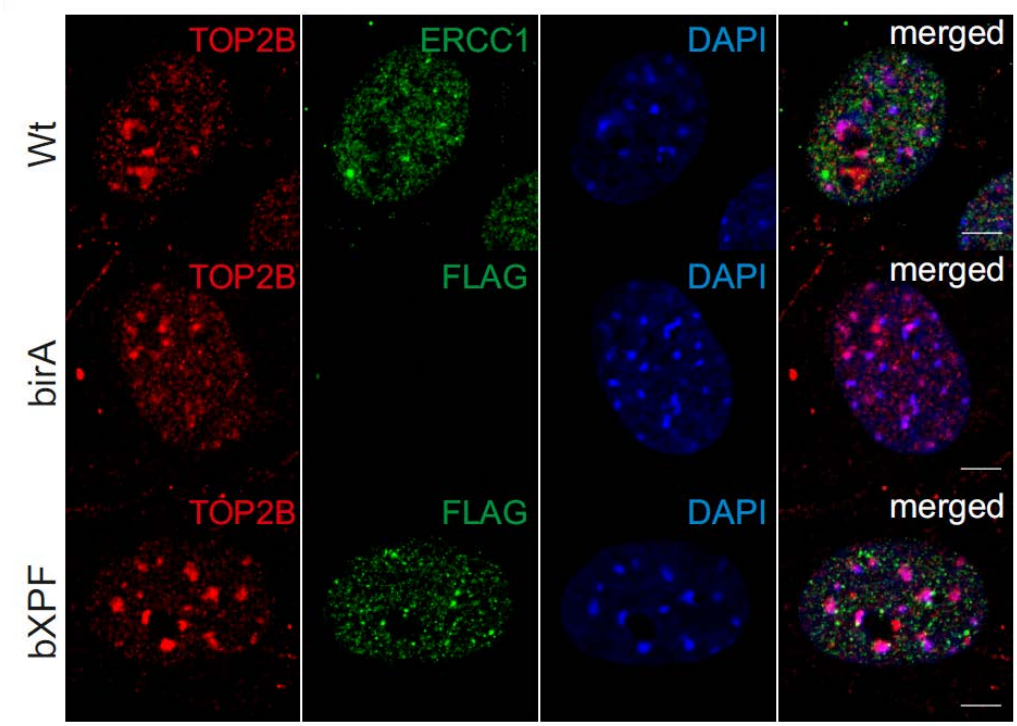


Figure 14. ERCC1-XPF cooperates with DNA topoisomerase IIβ upon transcription activation. ***(A)****. bXPF pulldowns (PD) and western blot with TAF4, TAF6, TAF10, TBP, TOP2B, TOP2A and TOP1 in nuclear extracts from bXPF and BirA MEFs.* ***(B)****. Co-immunoprecipitation experiments using anti-TOP2B, anti-TOP2A or anti-TOP1 in nuclear extracts from wt. MEFs analyzed by western blotting for ERCC1 as indicated.* ***(C)****. Co-immunoprecipitation experiments using anti-ERCC1 in nuclear extracts from wt. MEFs analyzed by western blotting for the indicated protein factors. The input and flow-through are 1/20 of the extract used.*

Confocal microscopy experiments in Wt. MEFs showed the nuclear localization of TOP2B, ERCC1 and Flag-tagged XPF throughout the nucleus, with prominent

distribution of TOP2B on heterochromatic regions (**Figure 15A**), thereby supporting our previous observations. The recruitment of ERCC1-XPF on the promoter regions and our findings confirming the interaction of ERCC1-XPF with TOP2B revealed in proteomics studies prompted us to perform a series of ChIP/re-ChIP experiments in MEFs upon basal conditions and in the presence of tRA or UVC irradiation. ChIP for TOP2B and re-ChIP for ERCC1 and Flag-tagged XPF showed that factors co-occupy the tRA responsive gene promoters (**Figure 15B**). Thus, the ERCC1-XPF complex acts synergistically and assembles with TOP2B on gene promoter regions.



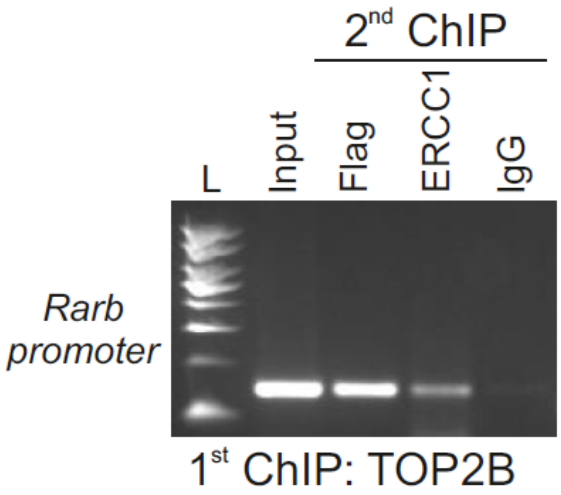


Figure 15. XPF/ERCC1 – TOP2B interaction: A. *Immunofluorescence detection of ERCC1, Flag-tagged XPF and TOP2B in Wt. or bXPF MEFs under basal conditions. B. ChIP with TOP2B antibody and re-ChIP with antibodies raised against Flag-tagged XPF and ERCC1.*

3.4 ERCC1/XPF recruits preferentially on TOP2B-generated DSBs on active gene promoters. Previous findings revealed a profound role of transcription as a source of genome instability. Transcriptional activation has been tightly linked to TOP2B-dependent DSB generation on promoters prompting us to further explore whether transcription-associated damage interferes with proper DDR signaling. Confocal imaging documented the presence of phosphorylated histone H2AX (γ-H2AX) containing foci and 53BP1 DNA damage (response) marker staining throughout the nucleoplasm of tRA-treated primary MEFs (**Figure 16**). Under these conditions, we find comparable immunolocalization patterns of RAD51 (a homologous DNA recombination repair protein), FANCI (an indicator of inter-strand crosslinks, ICLs, presence) and pATM (a DNA damage response marker) across WT and tRA-treated MEFs (**Figure S2**). Inactivation of Ataxia-telangiectasia mutated (ATM) by exposing tRA-treated MEFs to KU-55933, an ATM inhibitor abrogated the accumulation of DNA damage-associated γ-H2AX and 53BP1 foci, whereas Ataxia telangiectasia and Rad3 related (ATR)/Cyclin dependent kinase inhibitor NU6027 showed lack of difference in the presence of γ-H2AX and 53BP1 DNA damage-associated staining compared to tRA-treated cells (**Figure 16**). In agreement with previous data, our findings indicate that transcription activation triggers damage that requires functional ATM-mediated DNA damage response (DDR) signaling.

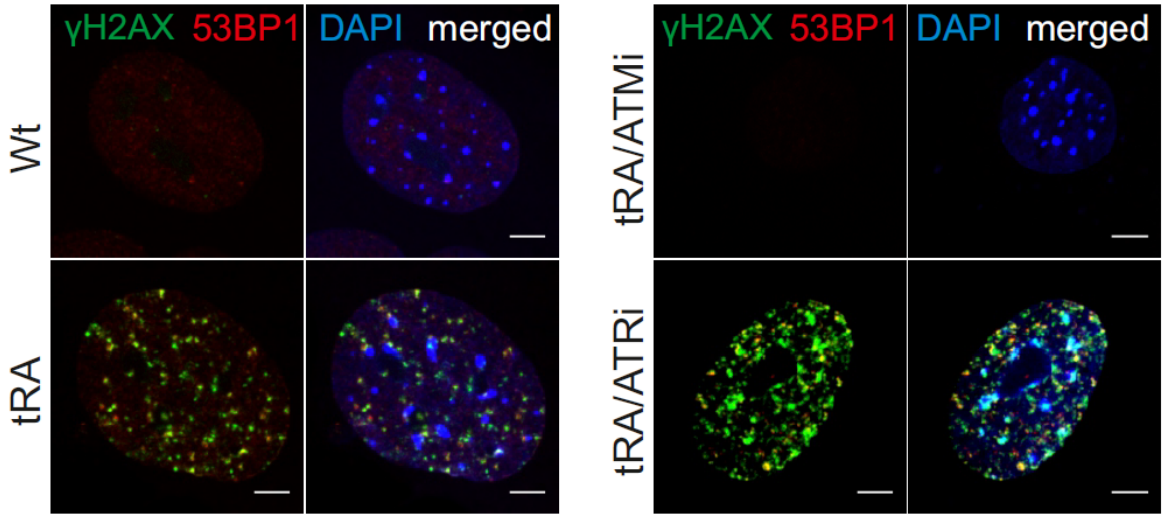


Figure 16. ERCC1-XPF recruits on TOPIIβ-generated DSBs on active promoters. *Immunofluorescence detection of γH2AX and 53BP1 in Wt. MEFs upon basal conditions or tRA treatment and in the presence of ATM or ATR inhibitors.*

Next, we investigated whether the DNA damage events generated during transcription are linked to TOP2B-induced DSB formation. To do so, we treated WT primary MEFs, in the presence of tRA, with merbarone· a DNA topo II catalytic inhibitor that acts by blocking topo II-mediated DNA cleavage without interfering with protein-DNA binding. Unlike in tRA-treated cells, we find a noticeable lack of γH2AX foci in MEFs exposed to merbarone and tRA (**Figure 17**), supporting the notion that the observed transcription-associated damage is TOP2B-mediated.

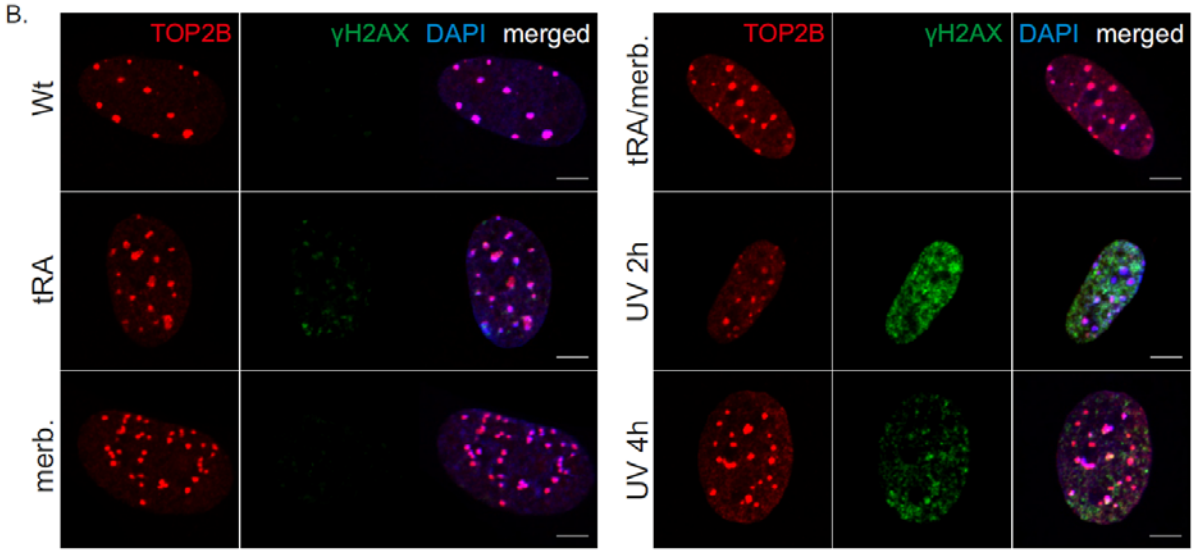


Figure 17. TOP2B mediates transcription-associated damage: *Immunofluorescence detection of γH2AX and TOP2B in Wt. MEFs upon basal conditions or upon exposure to UVC irradiation, tRA and/or merbarone.*

The relevance of transcription-associated DNA breaks to NER remains poorly understood. However, it is intriguing that TOP2B is enriched amongst the bXPF-bound proteins and upon transcription activation are highly co-recruited on promoters. We therefore reasoned whether the recruitment of ERCC1-XPF on the promoters upon transcription activation coincides with the induction of DSBs. Indeed, inhibition of TOP2B-mediated DSB formation using merbarone resulted in reduced ChIP signals for bXPF in promoter regions compared to tRA-treated MEFs (**Figure 18**), indicating the release of ERCC1-XPF from promoter regions upon TOP2B inhibition.

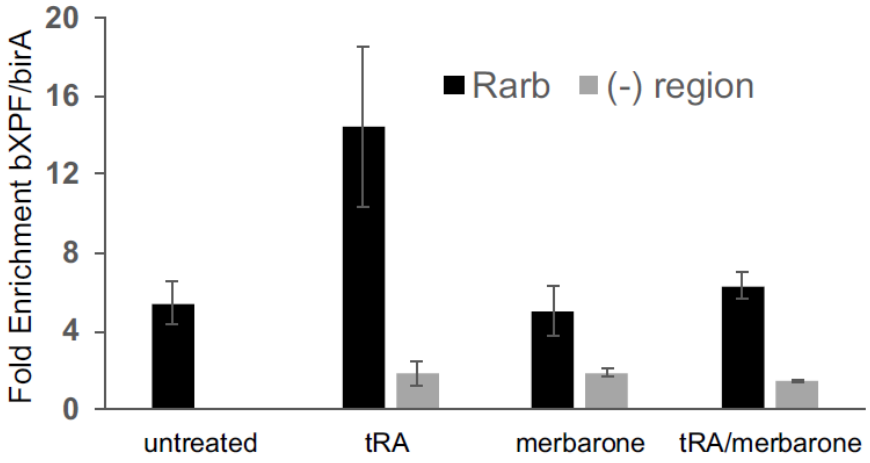


Figure 18. XPF assembly on promoters is TOP2B depended upon transcription induction: *bXPF ChIP signals are shown as fold enrichment of the percentage of input over the percentage of input BirA in the presence of tRA, merbarone and tRA/merbarone at the Rarb promoter and (-) regions.*

Furthermore, immunofluorescence studies revealed that TOP2B is released from heterochromatin and translocates to the nucleoplasm in tRA-treated MEFs (**Figure 19**) compared to the corresponding control cells (**Figure 15**). As with untreated control cells, we find that ERCC1-XPF accumulates predominantly in the nucleoplasm in tRA-treated MEFs (**Figure 19 and 15**). Importantly, we find that the exposure of MEFs to UVC irradiation leads to a punctate staining pattern of TOP2B

in heterochromatic regions, whereas ERCC1 and Flag-tagged XPF are still localized throughout the nucleoplasm (**Figure 19**). Thus, upon transcription stimulation ERCC1-XPF binds preferentially to promoters marked with DSBs generated by TOP2B, suggesting a putative role of ERCC1-XPF in transcription-associated DNA break repair through its involvement in certain HR and NHEJ repair mechanisms.

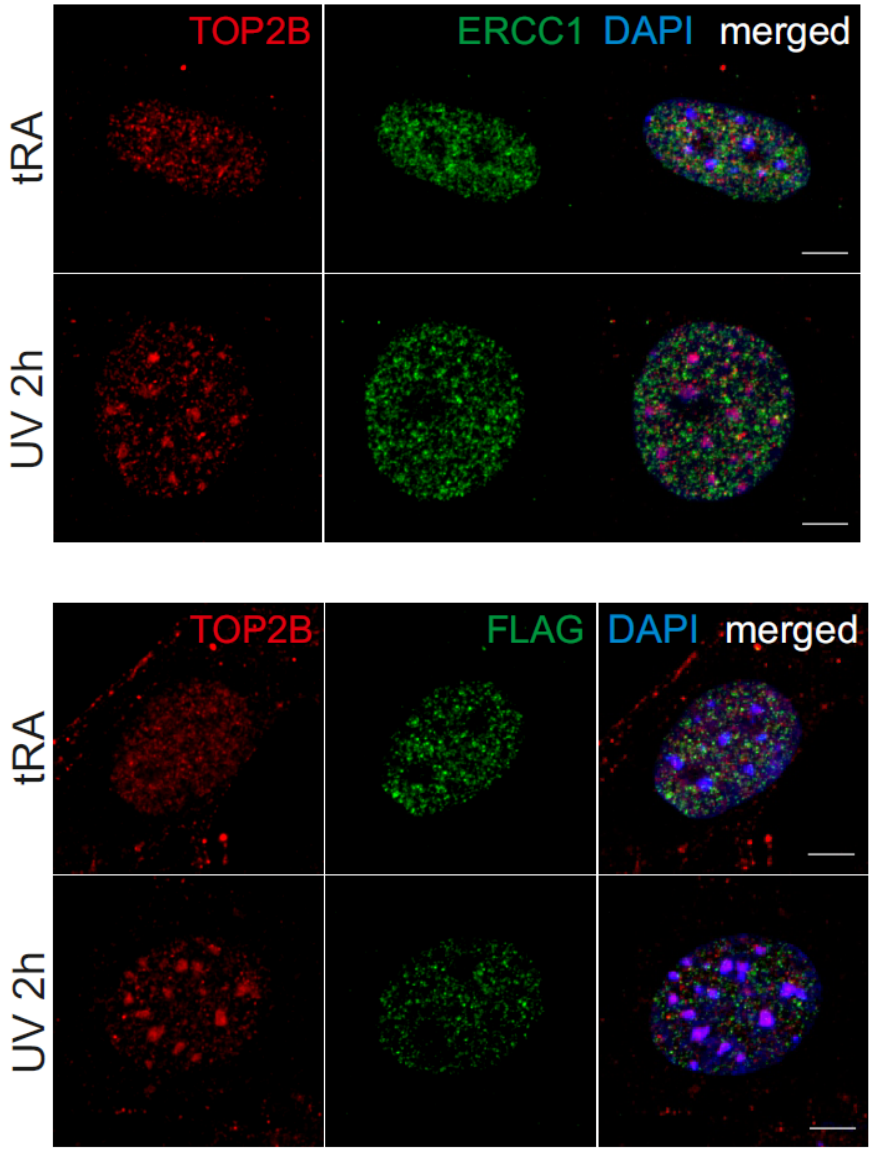


Figure 19. TOP2B and ERCC1/XPF localization upon tRA and the UV: *Immunofluorescence detection of TOP2B, ERCC1 and Flag-tagged XPF in Wt. or bXPF MEFs exposed to tRA or UVC irradiation.*

3.5 Genome-wide identification of DNA DSBs & ERCC1/XPF recruitment on promoters. Having established that the recruitment of ERCC1/XPF on promoters is guided by TOP2B-mediated DSBs, we next compared the ChIP-Seq DNA binding

results with the genome-wide distribution of DSBs in primary mouse MEFs. To identify the endogenous DSBs that are generated upon tRA-mediated transcription activation we performed BLISS (Breaks Labeling *In Situ* and Sequencing), which quantifies DSBs across the genome. A detailed workflow schematic is depicted in **Figure 20**.

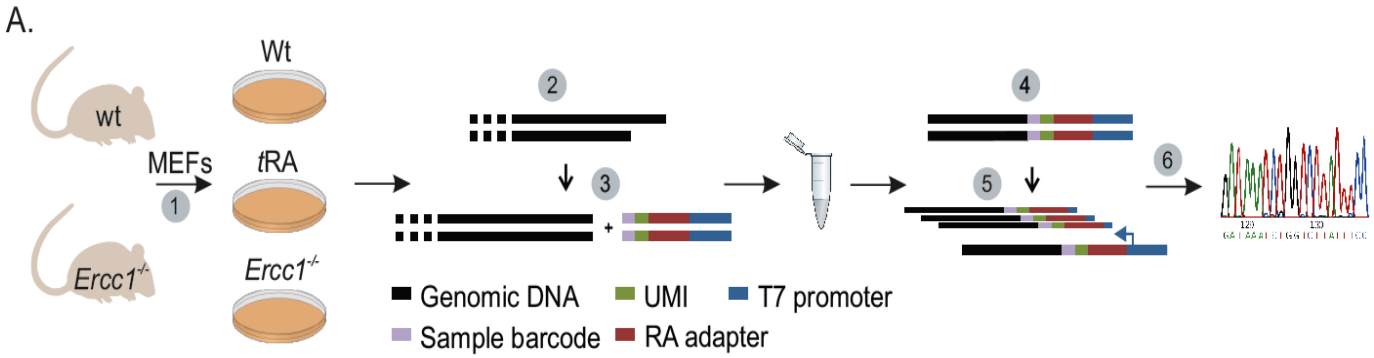


Figure 20. *BLISS methodology scheme: MEFs are extracted from wt. and Ercc1$^{-/-}$ mice, cultured and fixed on plates. Following in situ DSB blunting, DSB ends are ligated to dsDNA adapters, genomic DNA is extracted, sonicated and in vitro transcription isolates the tagged DSB sequences. The resulting DNA fragments are used for illumina library preparation and sequencing.*

By filtering out any PCR duplications, we mapped the DSBs in 100ng of genomic DNA isolated from untreated and tRA-activated wt and *Ercc1*$^{-/-}$ MEFs and found comparable differences in the total amount of DSBs between the three conditions (**Figure 21**).

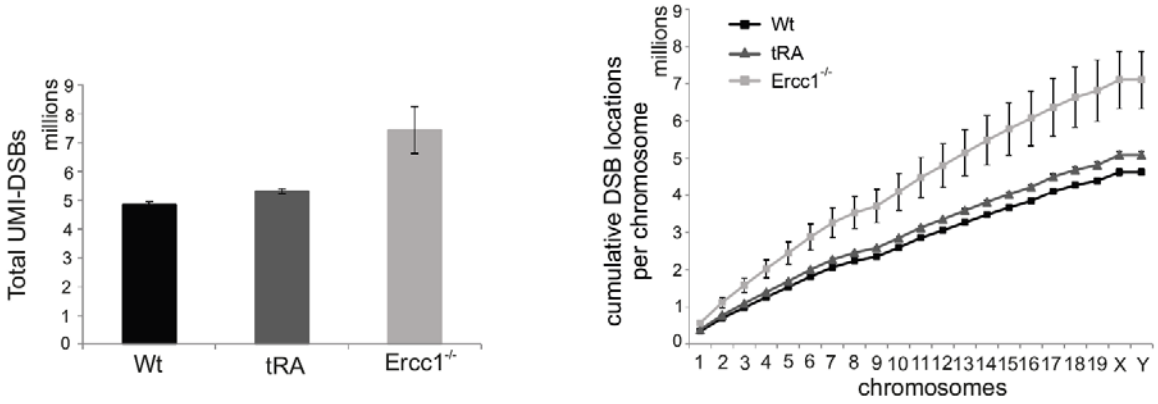
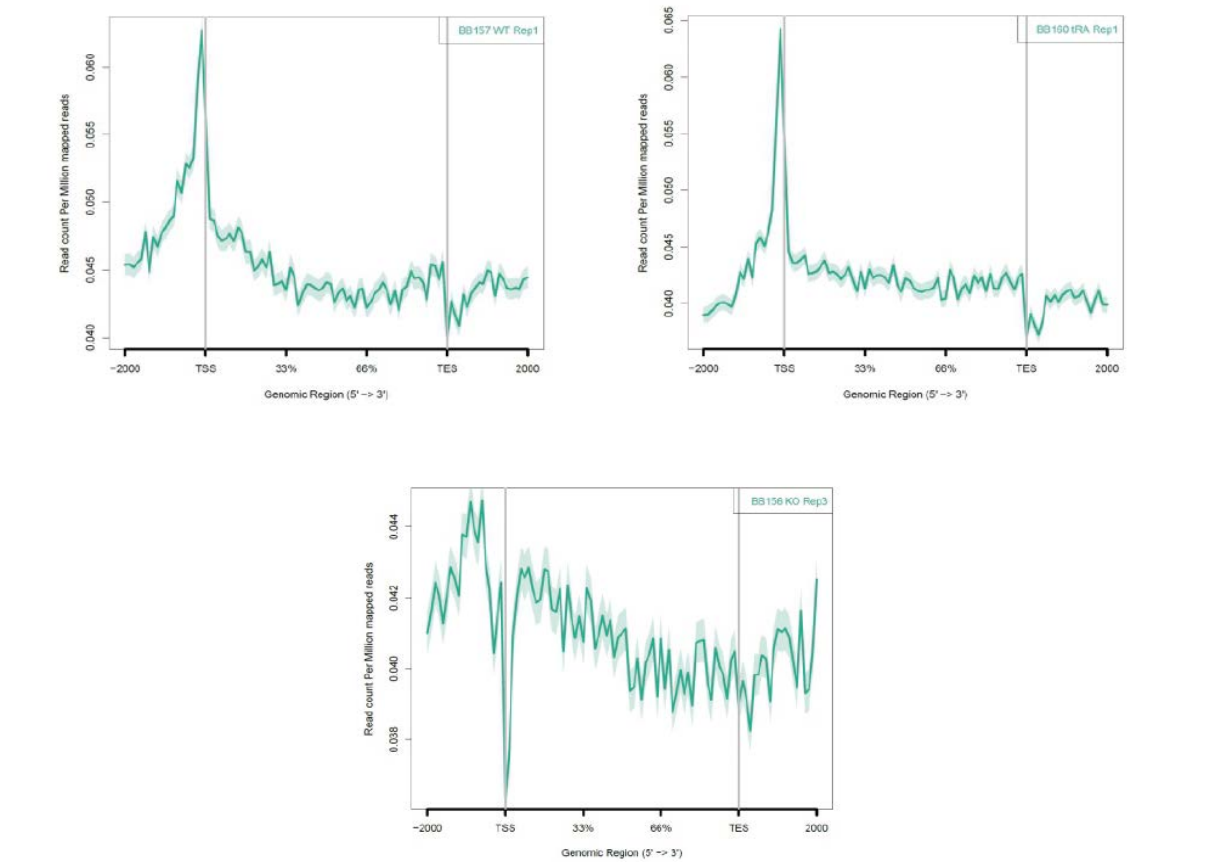


Figure 21. Genome-wide identification of DSBs: *Total unique DSBs were plotted for wt. (untreated), tRA-treated and Ercc1$^{-/-}$ MEFs, showing an increase of the total DSBs upon*

tRA-induced transcription and in the Ercc1$^{-/-}$ mouse. These differences are also evident for different DSB locations plotted for each chromosome among the conditions.

In line with recent findings in different cell types, DSBs were strongly enriched in the vicinity of TSSs as well as along gene bodies and promoters (**Figure 22A and 22B**) in the cell population. Interestingly, tRA stimulation of the cells increased the levels of the generated DSBs specifically on promoters and along gene bodies, compared to wt and *Ercc1*$^{-/-}$ MEFs.



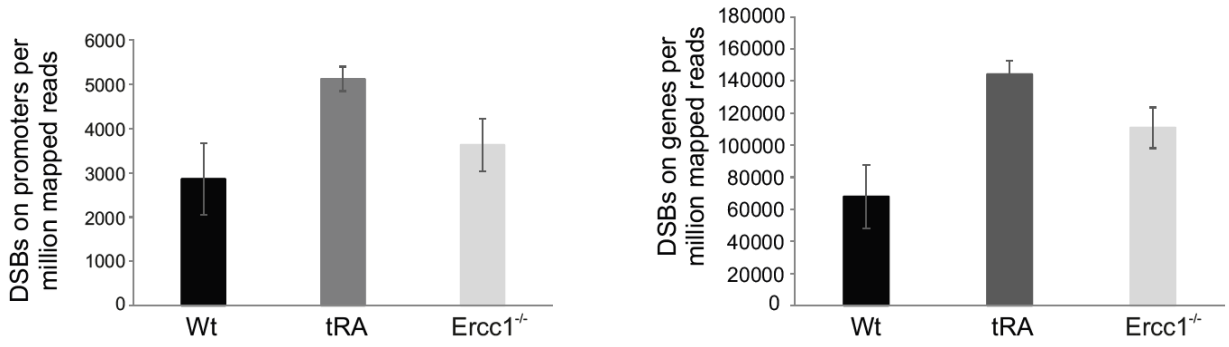


Figure 22. DSBs mapping and quantification across the genome: A *Enrichment of DSBs around TSSs, throughout gene bodies.* ***B.*** *Enrichment of DSBs on promoters and on genes per million mapped reads in untreated, tRA-activated wt. and Ercc1-/- MEFs*.

Further examination of the DSB maps revealed that 95,9% out of 1100 peaks identified in ChIP-Seq to be bound by XPF in basal conditions (**Figure 9B**) had DSBs (n=5803), while tRA treatment increased the number of unique DSBs to 9665 in 91,2% of the XPF-bound DNA regions (**Figure 23**). A detailed annotation analysis showed that 26% of these DSBs were detected preferentially on XPF-bound promoters in wt MEFs, compared to the 42% of promoter-associated DSBs in tRA-treated cells.

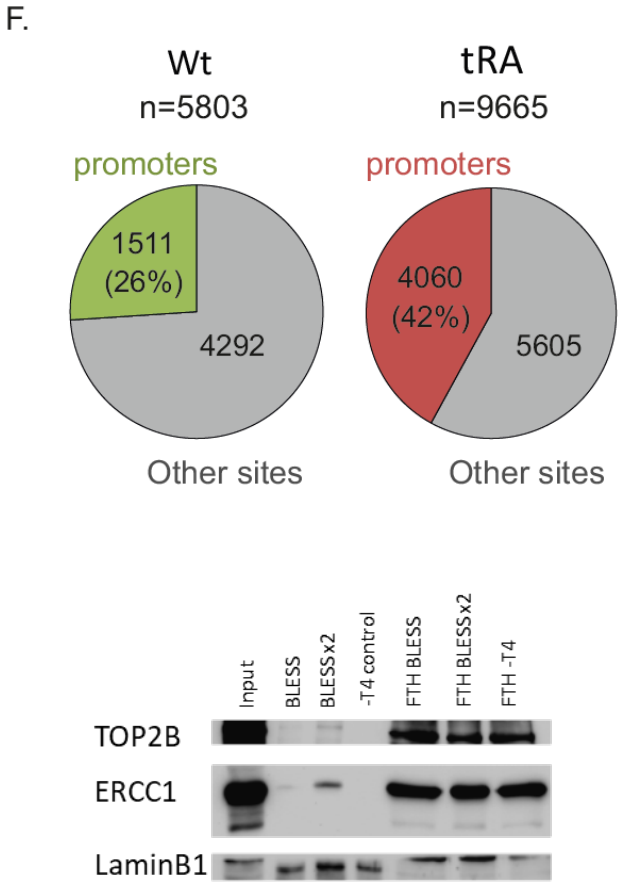


Figure 23. Comparing ChIP with BLISS: A. *Annotation of transcription-associated DNA DSBs on XPF-bound sites.* ***B****. Modified BLESS-Western blotting approach shows the isolation of DSBs with ERCC1 and TOP2B bound on DSB fragments.*

Next, we used BLESS (Breaks Labeling, Enrichment on Streptavidin and Sequencing) coupled to qPCR (**Figure 24**), to further validate the involvement of TOP2B in the transcription-associated generation of DSBs. Indeed, we quantified

DSBs in basal conditions, tRA stimulation and TOP2B inhibition and found that tRA-induced DSBs are mainly generated by TOP2B. Thus, we confirm that the recruitment of ERCC1/XPF on gene promoters coincides with transcription-associated DSB formation genome-wide.

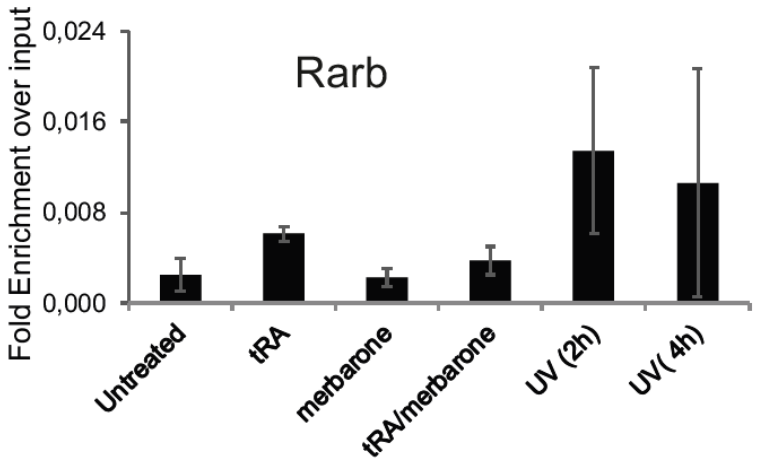


Figure 24. BLESS validation: *BLESS signals quantified by qPCR in the presence of tRA, merbarone and tRA/merbarone at the Rarb promoter region.*

3.6 Generation of TCEA2 knockin and conditional knockout ES lines. The *in vivo* biotinylation tagging approach has also been applied for the generation of TCEA2 knock-in and conditional knockout PC3 and JM8N4 ES lines, by targeting the endogenous locus with a C terminally tagged version of the gene (**Figure 25**), in addition to cooperating a three-loxP system that allows for a practically full knockout. Similarly to the XPF targeting strategy (see available mouse models), TCEA2 knockin ES may be applied for the establishment of knock-in mice by crossing with BirA expressing transgenic mice, generating the desired double homozygous bioTCEA2 genotype. The TCEA2 knockin mice may also be crossed with a CRE-deleter mouse to generate a complete TCEA2 knockout animal.

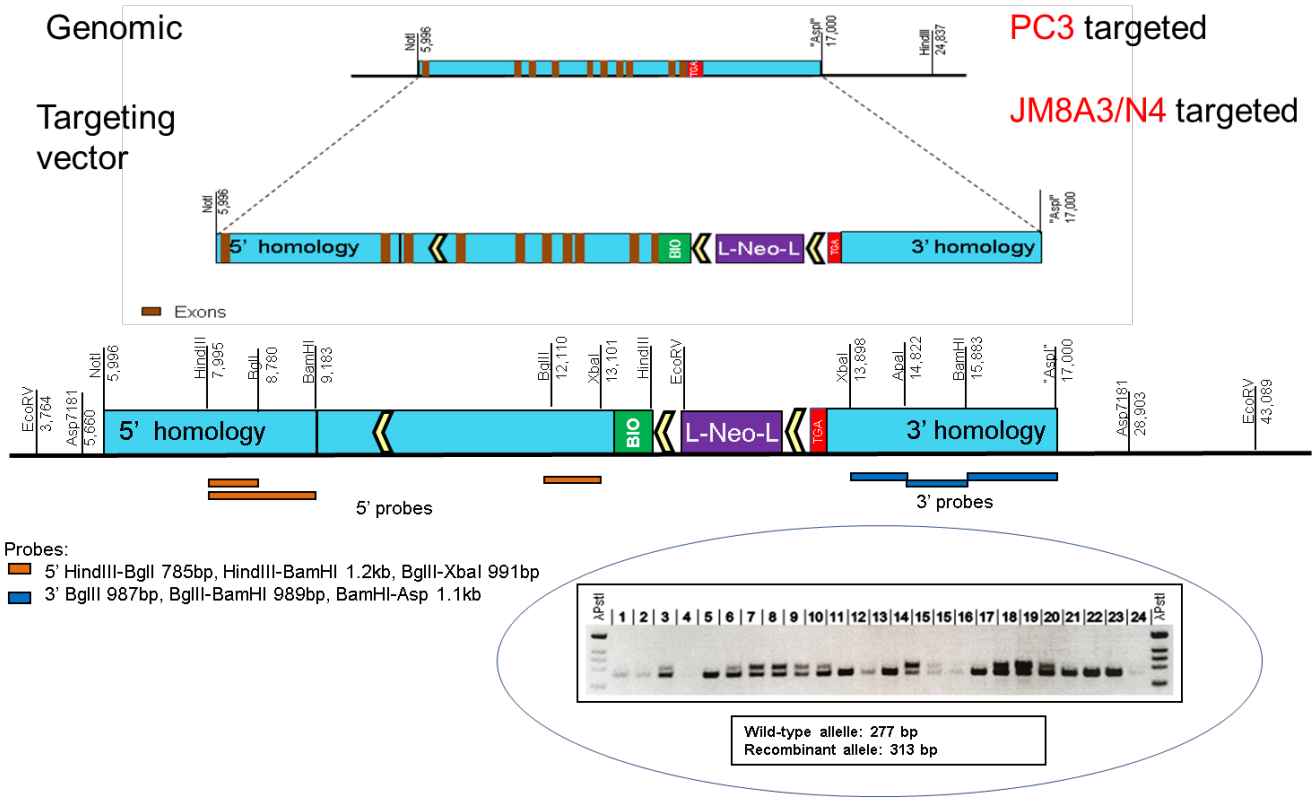


Figure 25. TCEA2 targeting strategy: *The TCEA2 targeting construct is consisted of a 5' homology region and a 3' homology region, the Flag-Tev-Avidin tag (Bio tag), and a selection cassette. The homology regions were cloned from 129 mouse Embryonic Stem Cell's genomic DNA. The 7.8 kb 5' region spans exon Ex1 of the TCEA2 gene up to the last exon's (EX10) stop codon. The 3.4 kb 3' region spans the rest of Ex10 which corresponds to the 3' untranslated region of the TCEA2 gene. The synthetically generated Bio-tag, bearing three stop codons, was cloned immediately after the last codon of the TCEA2 open reading frame, by removing the internal stop codon. Adjacent to the tag a LoxP-NeomycinR-LoxP cassette was cloned that would allow for positive selection of targeted ES clones in the presence of G418 antibiotic. A third LoxP, that will allow for a conditional knock out mouse by crossing with a dermis-specific CRE transgene, has been strategically placed prior the highest conserved region including Ex4 with the same orientation as Neo-LoxP. While allowing a large 4.3Kb region for homologous recombination, its blunt end insertion further diminishes the possibility of exclusion during ES targeting leaving no restriction sites at its ends.*

The TCEA2 targeting constructs were electroporated into JM8A3N4 as well as PC3 ES cells and selected with G418 since a Neo cassette has been employed. Even though previous mouse models had been established with the use of PC3 solely, JM8A3 were chosen for this task to test their efficiency, that literature emprases, according to which they give high rates of germline transmission and are feeders-free, thereby increasing the effciency of cell culturing.

For the genotyping of the clones that survived the selection, few performed a PCR assay for the identification of the third loxP site. The upper band represents the recombinant loxP included allele and the lower is the Wt. allele. Positive clones were then blotted by Southern with 4 different 5' and 3' probes that had been isolated from genomic DNA resulting in several positive clones (**Figure 26**).

Since JM8n4 are from the c57 black coat background whose non-agouti mutation has been restored in one allele, chimeras appear brown spotted when ijected into blastocysts of the C57 black coat background, however color penetration is not a representative of the genetic infiltration.

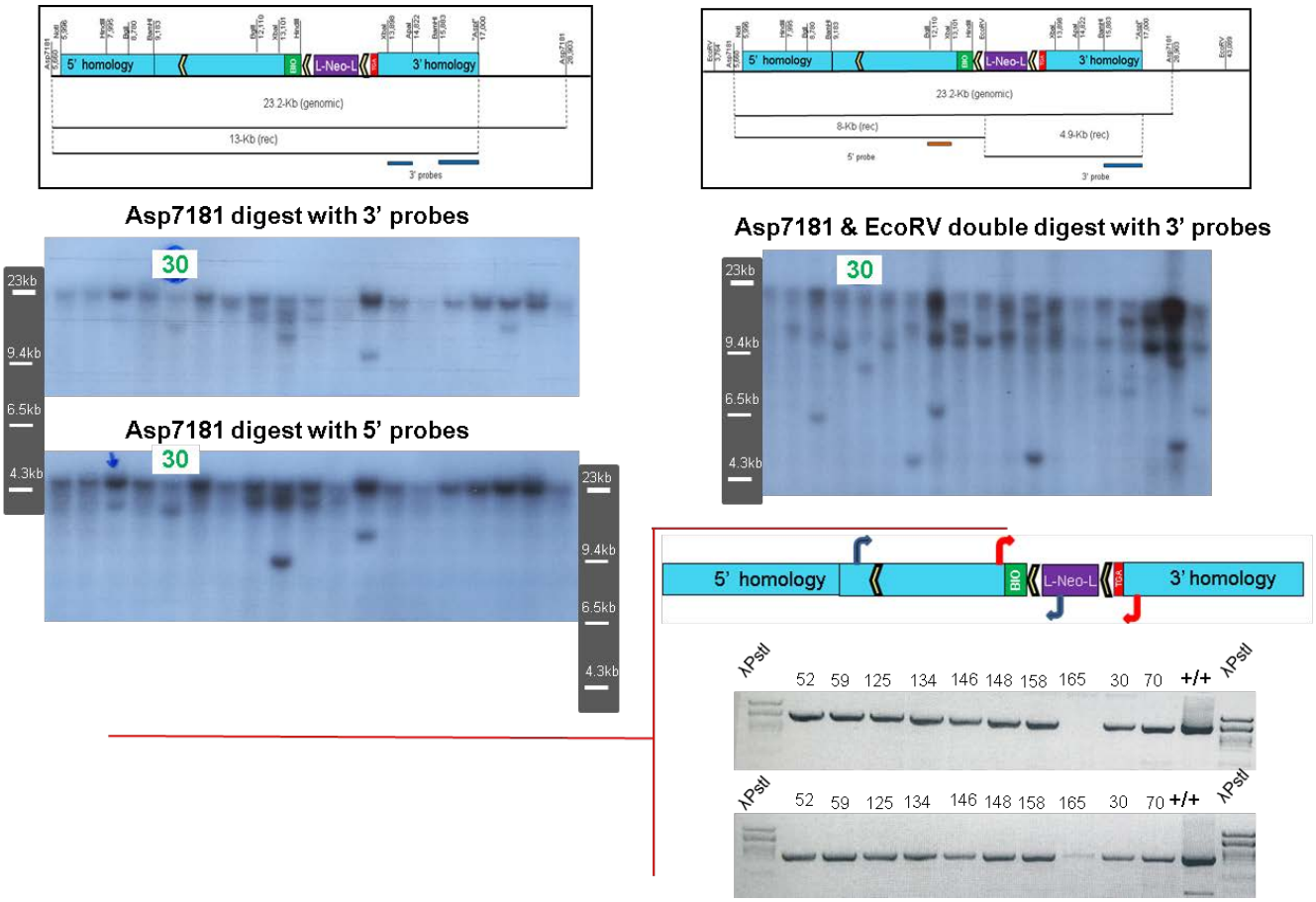


***Figure 26. Targeted TCEA2 genotyping**: Genotyping of selected clones for the identification of the constructs presence was followed by several Southern Blot assays that confirmed proper topology of the construct as a result of homologous recombination (an example of positive clone 30 is illustrated).*

To estimate the efficiency of recombination for the three loxP system, Cre recombinase has also been expressed in vitro with the pIC plasmid under the expression of MCI mammalian promoter by the electroporation of a positive clone (**Figure 27**).

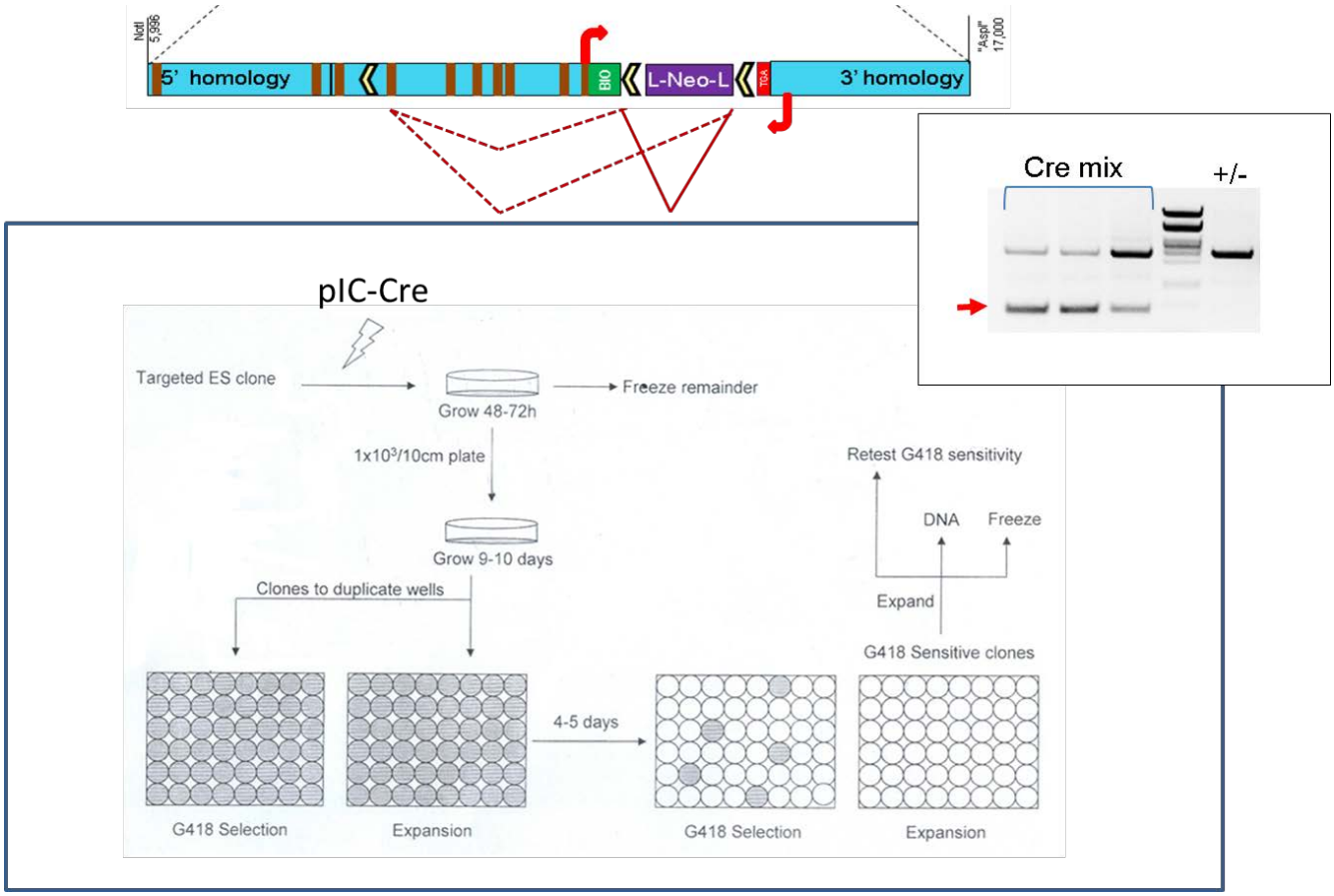


Figure 27. TCEA2 in vitro Cre recombination: *Selection with g418 is undertaken in a replica plate for every clone, three different groups of cells are expected to occur based on the pairing among the three loxPs. DNA was extracted from the mixture of these targeted cells and were screened through PCR for the area enclosing the neo cassette, revealing only the exclusion of Neo in a subset of cells. The process revealed a 40% of efficiency for Cre recombination.*

3.7 Supplementary figures

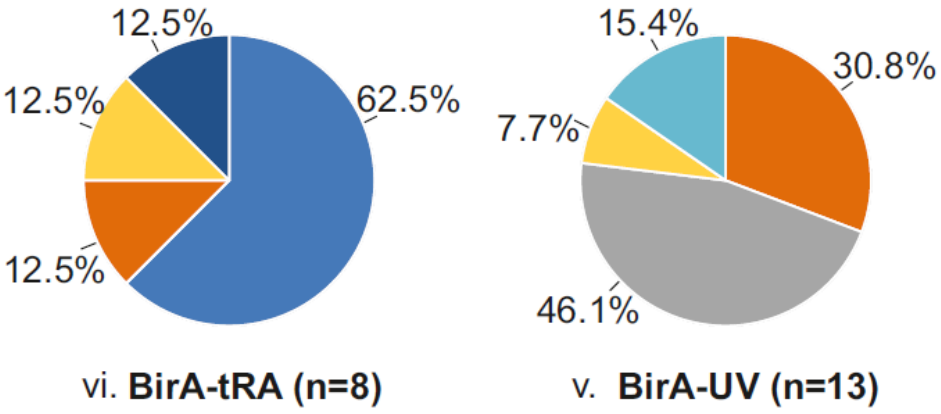
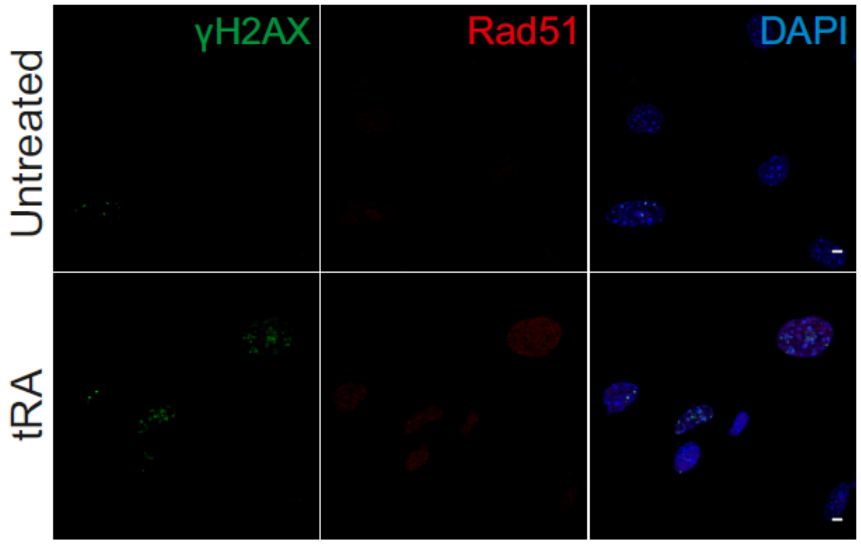


Figure S1. ChIP-Seq BirA controls: *BirA control mouse embryonic fibroblasts (MEFs) under basal conditions, following UVC-induced DNA damage or upon transcription stimulation using trans-retinoic acid (tRA), where used as controls in our ChIP-Seq approach. Generated data where contrasted to the bXPF samples data.*



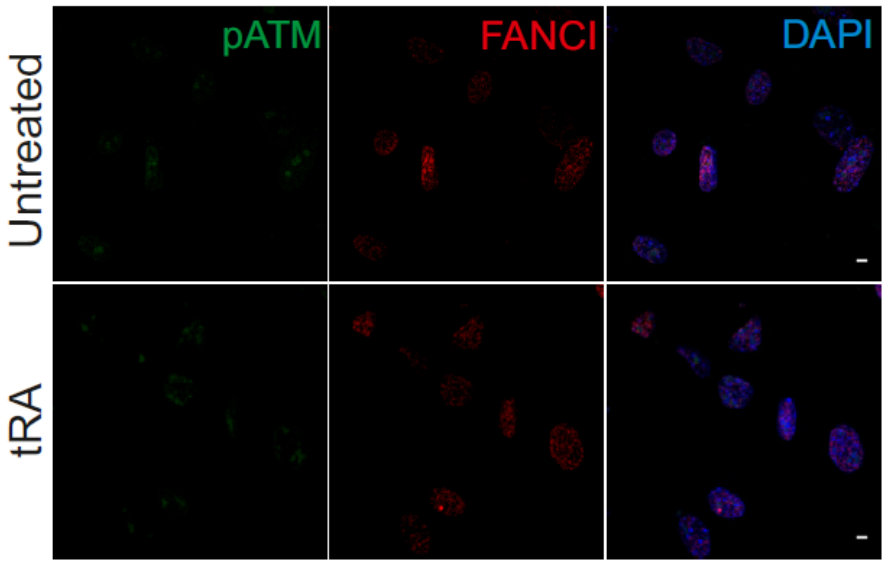


Figure S2. DSB immunolocalization of γH2AX, Rad51 and FANCI proteins: *An increased number of DNA double strand brakes foci have been detected with the use of γH2AX and Rad51 markers upon transcription activation but not with the Fanconi anemia complementation group I (FANCI) marker.*

4. Discussion

Developmental-stage and tissue-specific programs of gene expression require the action of sequence-specific DNA binding factors, the basal transcription machinery and chromatin remodeling and modification enzymes (Lemon and Tjian, 2000; Orphanides et al., 1996). Together, these factors create a chromatin environment that allows the synthesis of the primary transcript (Ohler and Wassarman, 2010). If the transcriptional machinery is defective or the process of transcription is challenged due to *e.g.* transcription-blocking DNA lesions, the process of RNA synthesis halts (Hoeijmakers, 2009). DNA damage events may occur accidentally injuring indiscriminately the DNA helix or else in a scheduled and spatially restricted manner that may often be beneficial to the survival of the cell. If left unrepaired, random DNA lesions gradually build up in the mammalian genome inadvertently interfering with vital DNA-templated transactions, such as transcription, DNA replication or homologous recombination (HR). Irreparable DNA insults may occasionally be fixed into mutations during DNA replication driving cancer (Pages and Fuchs, 2002) or obstruct ongoing transcription leading to cellular malfunction and the premature onset of age-related diseases (Garinis et al., 2009; Garinis et al., 2008). Unlike stochastic DNA insults, physiological DNA lesions are scheduled (in terms of time) or targeted (in terms of genomic location) DNA damage events that are thought to facilitate DNA-dependent transactions during mammalian development or upon exposure to various stimuli. For instance, DNA DSBs are purposely formed during meiotic HR to facilitate genetic diversity in gamete cells (Zickler and Kleckner, 2015). DNA breaks are also required during V(D)J recombination in developing B- and T-lymphocytes to secure antigen-binding diversity of the immunoglobulin and T-cell receptor proteins (Boboila et al., 2012). Moreover, DNA lesions followed by error-prone DNA repair enhance antibody diversity by triggering somatic hypermutation of immunoglobulin variable genes (Di Noia and Neuberger, 2007). Targeted DNA damage events are also an inevitable consequence of extreme shifts in transcription demands (Gaillard and Aguilera, 2016) and, as we will discuss later, they are required for fine-tuning transcription, facilitating access of transcription factors and the RNAPII at promoter and enhancer regions or to enable rapid intra- and inter-chromosomal interactions for DNA-templated transactions. Thus, whereas unscheduled DNA insults pose a threat to cell viability, programmed DNA damage events are unavoidable during mammalian

development and, ironically so, an essential prerequisite for cellular and organismal survival.

Previous studies (also from our lab) have shown that ERCC1-XPF recruits on the promoters of growth genes during postnatal hepatic development (Kamileri et al., 2012c). However, it has been difficult to rationalize the i. functional relevance of an endonuclease complex in the vicinity of promoter regions as well as the ii. putative functional role of ERCC1-XPF in transcription. To dissect the functional relevance of these earlier findings in mammalian cells, we used a unique series of knockin animals carrying a tagged version of bXPF factor. Using these mice and a series of immunoprecipitation, ChIP-Seq and proteomics methodologies, we find that bXPF recruits genome-wide on promoters under native conditions and that this recruitment is further enhanced when cells are treated with the transcription activator tRA. In line, we show that bXPF DNA binding sites are positively correlated with those of RNAPII, histone H3K4 trimethylation and H3K27 acetylation ChIP-Seq profiles that are known to be associated with active transcription. Similar data were observed we interrogated bXPF ChIP-Seq data only on the TSS of genes. Intriguingly, bXPF ChIP-Seq profiles showed no overlap with the repressive histone H3K4 monomethylation sites or with the lamin-associated domains. Subsequent work revealed that bXPF is released from nearly all previously identified bXPF-bound genome targets when cells are exposed to UV irradiation. Together, these findings support the notion that bXPF occupancy on promoters and the process of mRNA synthesis are tightly linked and that DNA damage inadvertently affects the recruitment of ERCC1-XPF complex on promoters.

A high-throughput proteomics approach on nuclear extracts derived from bXPF MEFs revealed a wide range of XPF-bound interacting protein partners associated with transcription, chromosome organization and DNA repair. Using co-immunoprecipitation experiments, we find that ERCC1/XPF interacts with distinct TATA-associated factors (TAFs) implicating its direct involvement in transcription initiation. Further immunoprecipitation experiments revealed that ERCC1/XPF interacts with topoisomerase IIβ (TOP2B) aiming at relieving the topological stress triggered by RNAPII during the process of mRNA synthesis. In support, a series of immunofluorescence experiments revealed that TOP2B, ERCC1 and flag-tagged XPF co-localize in the nucleus. ChIP for TOP2B and re-ChIP-qPCR for ERCC1 and

Flag-tagged XPF further confirmed that ERCC1/XPF occupancy overlaps with that of TOP2B on transcription stimulated promoter regions.

DNA topoisomerases are ubiquitously expressed in all cell types to regulate DNA supercoiling. In particular, TOP2B triggers the relief of torsional stress that occurs during RNAPII-mediated transcription and DNA replication. TOP2B catalyzes the transient breaking and rejoining of two strands of duplex DNA ensuing that the two DNA strands pass through one another, thereby altering the topology of DNA before resealing the breaks. These data and our finding that ERCC1-XPF, an endonuclease complex known to be involved in DSB repair, recruits with TOPIIβ on promoters prompted to further investigate the functional link between TOP2B-dependent DNA DSB generation and the recruitment of ERCC1/XPF complex with TOPIIβ on the promoters of active genes during the process of mRNA synthesis. Immunofluorescence experiments revealed that transcription per se triggers a robust DNA damage response as evidenced by the detection of γ-H2AX and 53BP1 foci followed by ATM signaling in MEFs that was abolished when tRA-treated cells were exposed to merbarone, an inhibitor of TOP2B activity supporting that TOP2B activity is a major cause of transcription-associated DNA damage. Having established the link between transcription, DNA damage and TOP2B, we then investigated the involvement of ERCC1-XPF complex in the repair of transcription-associated NA DSBs. Importantly; we find that the recruitment of ERCC1-XPF on promoters coincides with the induction of DNA DSBs upon transcription activation. Similar to our previous findings, inhibition of the catalytic activity of TOP2B diminished bXPF recruitment at the promoters indicating that the substrate specificity of ERCC1-XPF is the presence of DNA DSBs themselves rather than the transcriptional activity of promoters. Genome-wide mapping of DNA DSBs in untreated or transcription-stimulated (tRA-treated) conditions and in *Ercc1*$^{-/-}$ MEFs, via BLISS, revealed the substantial overlap of XPF binding sites to transcription-associated DSBs. Specifically, DNA DSBs were strongly enriched in the vicinity of TSSs as well as within gene bodies and promoters. Furthermore, upon transcription activation, DNA DSBs were significantly increased on promoters and gene bodies of XPF-bound gene targets. Indeed, we find that more than 95% of XPF binding sites bear DNA DSBs, while more than 91% of tRA-mediated DSB sites were occupied by XPF, of which nearly half are located on promoters. Subsequent work using BLESS-qPCR

quantification confirmed that TOP2B is a major cause of DNA DSBs generated by tRA.

4.1. Concluding remarks and future aspects. Delineating the functional role of protein complexes involved in DNA repair, transcription and chromatin architecture has been fascinating but also challenging owing to the fact that these processes are intrinsically intertwined. A causal link that could keep these otherwise functionally distinct mechanisms together is DNA damage itself. Indeed, stochastic DNA lesions are often a threatening menace during the lifetime of a cell. However, programmed DNA damage events (*e.g.* meiotic recombination) or targeted DNA lesions (*e.g.* on promoters during abrupt switches in transcriptional load) are also an inevitable consequence and a prerequisite for the proper execution of developmental gene expression programs. As with random DNA lesions, programmed or targeted DNA damage events must be repaired or resolved in a timely manner. The latter would require that the DNA repair machinery is in close cooperation with protein complexes involved in transcription or chromatin architecture. Failure to do so (as in the case of the NER-deficient patients and mice), could hamper unequivocally the expression of genes during developmental transitions. The severity and time to onset of developmental defects could, therefore, reflect the individual roles of NER proteins outside DNA repair (*e.g.* transcription), the repair of tightly scheduled, transcription-associated, targeted DNA lesions outside NER (*e.g.* DNA breaks on promoters or R-loops) as well as the time- and cell type-specific requirements of the transcriptional programs involved. This scenario offers a likely explanation for the recurring onset of heterogeneous, but importantly non-random, developmental abnormalities associated with NER syndromes.

In spite of the recent progress, the precise biochemical functions of NER proteins in transcription-associated DNA damage, fine-tuning gene expression or shaping chromatin architecture remain obscure. With the advent of more sophisticated, functional animal models *e.g.* biotin-tagged or cell-specific knockout mice coupled to high-throughput mass spectrometry and next generation sequencing approaches (Chatzinikolaou et al., 2017; Karakasilioti et al., 2013), however, we may soon be able to gain insights into the NER-associated protein complexes and gene targets in distinct cell types at any stage during mammalian development. In this direction, the generation of TFIIS biotin tagged animal models that also carry also a

floxed TCEA2 allele will be decisive to dissect the functional relevance of transcription -associated stress in development and disease in mammals. Moreover, the recent advances in genome-wide chromosome conformation capture methods for characterizing the three-dimensional architectures of genomes (Duan et al., 2012) will allow us to gain insight on the functional relevance of distinct NER factors in genome-wide looping formation. As it is now possible to map DNA breaks at nucleotide resolution (Crosetto et al., 2013), it will also be soon possible to test whether the recruitment of NER-associated protein complexes (other than those shown in this work) on promoters also marks the simultaneous presence of *e.g.* DNA breaks on these genomic regions. Further work will be also necessary to explore how and to what extent *e.g.* DNA breaks or R-loops are functionally linked to transcription demands during mammalian development or with disease onset. It may only be then that we might begin to fully appreciate the impact and relevance of unavoidable, transcription-associated DNA lesions to the developmental abnormalities associated with NER syndromes shedding also some light into the multiple pathological complications associated with old age.

5. Methodology

Animal models and primary cells. The generation and characterization of Biotin-tagged XPF (bXPF) and NER-deficient mice has been previously described (Chatzinikolaou et al., 2017; Kamileri et al., 2012c). Animals were kept on a regular diet and housed at the IMBB animal house, which operates in compliance with the "Animal Welfare Act" of the Greek government, using the "Guide for the Care and Use of Laboratory Animals" as its standard. As required by Greek law, formal permission to generate and use genetically modified animals was obtained from the responsible local and national authorities. All animal studies were approved by independent Animal Ethical Committees at FORTH and BSRC Al. Fleming. Primary MEFs were isolated from E13.5d animals and cultured in standard medium containing Dulbecco's modified Eagle medium (DMEM) supplemented with 10% Fetal Bovine Serum (FBS), 50μg/ml streptomycin, 50 U/ml penicillin (Sigma) and 2mM L glutamine (Gibco). Cells were rinsed with PBS, exposed to UVC irradiation (10 J/m^2), MMC (10 μg/mL) (AppliChem), tRA (10 μM) (Sigma-Aldrich) or merbarone (1 μM) (Sigma-Aldrich) and cultured at 37^{o}C for 1 to 16h prior to subsequent experiments. Pre-incubation with ATM inhibitor (10 μM) and ATR inhibitor (10μM) started 1 h before genotoxic treatments and lasted throughout the experiment.

Immunofluorescence, Antibodies, Westerns blots, ChIP, Co-immunoprecipitation and Chromatin Pull-Down assays. Immunofluorescence experiments were performed as previously described (Chatzinikolaou et al., 2017; Kamileri et al., 2012c; Karakasilioti et al., 2013). Briefly, cells (primary MEFs) were fixed in 4% formaldehyde, permeabilized with 0,5% Triton-X and blocked with 1% BSA. After one-hour incubation with primary antibodies, secondary fluorescent antibodies were added and DAPI was used for nuclear counterstaining. Samples were imaged with SP8 confocal microscope (Leica). For local DNA damage infliction, cells were UV-irradiated (60 J/m^2) through isopore polycarbonate membranes containing 3-μm-diameter pores (Millipore). For co-immunoprecipitation assays, nuclear protein extracts from primary MEFs were prepared as previously described (Chatzinikolaou et al., 2017) using the high-salt extraction method (10mM HEPES-KOH pH 7.9, 380mM KCl, 3mM MgCl2, 0.2mM EDTA, 20% glycerol and protease inhibitors). Nuclear lysates were diluted three-fold by adding ice-cold HENG buffer (10mM HEPES-KOH pH 7.9, 1.5mM MgCl2, 0.25 mM EDTA, 20% glycerol) and

precipitated with antibodies overnight at 4oC followed by incubation for 3 h with protein G Sepharose beads (Millipore). Normal mouse, rabbit or goat IgG (Santa Cruz) was used as a negative control. Immunoprecipitates were washed five times (10mM HEPES-KOH pH7.9, 300mM KCl, 0.3% NP40, 1.5mM MgCl2, 0.25mM EDTA, 20% glycerol and protease inhibitors), eluted and resolved on 10% SDS-PAGE. Pulldowns were performed with 1.2 mg of nuclear extracts using M-280 paramagnetic streptavidin beads (Invitrogen) as previously described (Chatzinikolaou et al., 2017). For ChIP assays, primary cells (MEFs) were cross-linked at R.T. for 2.5 min with 1% formaldehyde. Chromatin was prepared and sonicated on ice 15 min using Covaris S220 Focused-ultrasonicator. Samples were immunoprecipitated with antibodies (5-8 μg) overnight at 4^{o}C followed by incubation for 3 hours with protein G-Sepharose beads (Millipore) and washed sequentially. The complexes were eluted and the crosslinking was heat reversed. Purified DNA fragments were analysed by sequencing or qPCR using sets of primers targeting different regions of tRA-responsive genes. ChIP re-ChIP experiments were performed as described above with the following modifications˙ after the first immunoprecipitation and washing, complexes were eluted with 10 mM DTT, 1% SDS in TE buffer for 30 min. Eluted samples were diluted 1:20 with re-ChIP buffer (10 mM Tris-HCl pH 8, 1 mM EDTA, 150 mM NaCl, 0.01% SDS and 1% Triton X-100) and immunoprecipitated overnight with the second antibody. Antibodies against HA (Y-11, wb: 1:500), ERCC1 (D-10, wb: 1:500, IF: 1:50), TOP2A (C-15, wb: 1:200, IF: 1:50) were from SantaCruz Biotechnology. γH2AX (05-636, IF: 1:12000) was from Millipore. TOP1 (NBP1-30482, wb: 1:1000, IF: 1:50), TOP2B (NB100-40842, wb: 1:1000) and 53BP1 (NB100-304, IF: 1:100) were from Novus Biologicals. TOP2B (20549-I-AP, IF: 1:50) was from Proteintech. TAF-4 (TAF2B9, wb: 1:500, IF:1:50), TAF-6 (TAF2G7, wb: 1:500) and TAF-10 (6TA-2B11, wb: 1:500) were from ProteoGenix. Streptavidin-HRP (wb: 1:12,000) was from Upstate Biotechnology. pATM (wb: 1:1000, IF: 1:1000) was from Rockland. pATR (wb: 1:1000, IF: 1:500) was from Genetex. FLAGM2 (F3165, wb 1:2.000, F1804, IF: 1:1000) was from Sigma-Aldrich.

Mass Spectrometry studies. Proteins eluted from the beads were separated by SDS/PAGE electrophoresis on a 10% polyacrylamide gel and stained with Colloidal blue silver (ThermoFisher Scientific, USA; 70). SDS-PAGE gel lanes were cut into 2-

mm slices and subjected to in-gel reduction with dithiothreitol, alkylation with iodoacetamide and digested with trypsin (sequencing grade; Promega), as described previously. Peptide mixtures were analysed by nLC-ESI-MS/MS on a LTQ-Orbitrap XL coupled to an Easy nLC (Thermo Scientific). The sample preparation and the nLC-ESI-MS/MS analysis were performed as previously described (Aivaliotis et al., 2007) (Rappsilber et al., 2002) with minor modifications. Briefly, the dried peptides were dissolved in 0.5% formic acid aqueous solution, and the tryptic peptide mixtures were separated on a reversed-phase column (Reprosil Pur C18 AQ, Dr. Maisch GmbH), fused silica emitters 100 mm long with a 75 μm internal diameter (ThermoFisher Scientific, USA) packed in-house using a packing bomb (Loader kit SP035, Proxeon). Tryptic peptides were separated and eluted in a linear water-acetonitrile gradient and injected into the MS.

RNA-Seq and Quantitative PCR studies. Total RNA was isolated from cells using a Total RNA isolation kit (Qiagen) as described by the manufacturer. For RNA-Seq studies, libraries were prepared using the Illumina® TruSeq® mRNA stranded sample preparation kit. Library preparation started with 1ug total RNA. After poly-A selection (using poly-T oligo-attached magnetic beads), mRNA was purified and fragmented using divalent cations under elevated temperature. The RNA fragments underwent reverse transcription using random primers. This is followed by second strand cDNA synthesis with DNA Polymerase I and RNase H. After end repair and A-tailing, indexing adapters were ligated. The products were then purified and amplified (14 PCR cycles) to create the final cDNA libraries. After library validation and quantification (Agilent 2100 Bioanalyzer), equimolar amounts of library were pooled. The pool was quantified by using the Peqlab KAPA Library Quantification Kit and the Applied Biosystems 7900HT Sequence Detection System. The pool was sequenced by using a S2 flowcell on the Illumina NovaSeq6000 sequencer and the 2x100nt protocol. Quantitative PCR (Q-PCR) was performed with a Biorad 1000-series thermal cycler according to the instructions of the manufacturer (Biorad) as previously described (Chatzinikolaou et al. 2017). All relevant data and primer sequences for the genes tested by qPCR are available upon request.

sBLISS and BLESS. To map DNA double-strand breaks (DSBs) genome-wide, we applied an adapted set-up of the Breaks labeling *in situ* and sequencing (BLISS) method (Yan et al., 2017). In suspension BLISS (sBLISS), DSB ends are *in situ*

blunted and ligated to specialized BLISS adapters that enable selective linear amplification of the genomic sequences at the DSB ends, via T7-driven *in vitro* transcription. Briefly, after cell treatment and prior to fixation, cells were washed, trypsinized and resuspended in pre-warmed PBS supplied with 10% fetal bovine serum (FBS), ensuring single-cell suspensions. Then the cells were counted and diluted to 10^6 cells/ml and fixed with 4% paraformaldehyde aqueous solution (Electron Microscopy Sciences #15710, Formaldehyde methanol-free) for 10 minutes at room temperature (RT). Formaldehyde was quenched with 2M glycine at a final concentration of 125 mM for 5 minutes at RT, while gently rotating, and for an additional 5 minutes on ice. Fixed cells were washed with ice-cold PBS and pelleted by centrifuging at 100-400g for 10 minutes at 4°C. For *in situ* DSB labeling, 10^6 fixed cells were incubated in a lysis buffer (10mM Tris-HCl, 10 mM NaCl, 1 mM EDTA, and 0.2% Triton X-100 (pH 8)), for 60 min on ice and the nuclei were, then, permeabilized with a pre-warmed permeabilization buffer (10 mM Tris-HCl, 150 mM NaCl, 1 mM EDTA, and 0.3% SDS (pH 8)) for 60 minutes at 37°C. After pelleting, the nuclei were washed twice with pre-warmed 1x CutSmart Buffer (New England Biolabs (NEB) #B7204) supplemented with 0.1% Triton X-100 (1xCS/TX100). To prepare the DNA Double Strand Break ends (DSB ends) for BLISS adapter ligation, the DSBs were blunted with NEB's Quick Blunting Kit (NEB #E1201) according to the manufacturer's instructions in a final volume of 100 µl for 60 minutes at RT. After blunting, the nuclei were washed twice with 1x CS/TX100 before proceeding with *in situ* ligation of BLISS adapters (see below for adapter preparation). Ligation was performed with 25 Weiss units of T4 DNA Ligase (5 U/µl, ThermoFisher Scientific #EL0011) for 20-24h at 16°C in reaction volumes of 100 µl supplemented with BSA (Thermo #AM2616) and ATP (Thermo #R0441). Per preparation of 10^6 cells, 4 µl of the selected BLISS adapter (10 µM) was ligated. Prior to use, BLISS dsDNA adapters were prepared from two complementary HPLC-purified oligonucleotides ordered from Integrated DNA Technologies (IDT). Each dsDNA adapter contains a T7 promoter sequence for *in vitro* transcription (IVT), the RA5 Illumina RNA adapter sequence for downstream sequencing, an 8-nt Unique Molecular Identifier (UMI) sequence generated by random incorporation of the four dNTPs according to IDT's 'Machine mixing' strategy, and a 8-nt sample barcode to enable multiplexing of BLISS libraries. Sense oligos diluted to 10 µM in nuclease-free water were phosphorylated with T4 PNK (NEB #M0201) supplemented with ATP, after which an

equimolar amount of antisense oligo was added. Oligos were annealed in a Thermocycler (5 minutes 95°C, then ramping down to 25°C in steps of 1.5°C per minute), to generate a 10 µM phosphorylated dsDNA adapter. After overnight ligation, nuclei were washed twice with 1x CS/TX100. To reverse crosslinks and extract gDNA, nuclei were resuspended in 100 µl DNA extraction buffer (10 mM Tris-HCl, 100 mM NaCl, 50 mM EDTA, and 1% SDS (pH7.5)), supplemented with 10 µl Proteinase K (800 U/ml, NEB #P8107), and incubated at 55°C for 14-18h while shaking at 800rpm. Afterwards, Proteinase K was heat-inactivated for 10 minutes at 95°C, followed by extraction using Phenol:Chloroform:Isoamyl Alcohol 25:24:1 with 10 mM Tris, pH 8.0, 1 mM EDTA (Sigma-Aldrich/Merck #P2069) and Chloroform (Merck #1024451000), followed by ethanol precipitation. The purified gDNA was resuspended in 100 µl TE and sonicated using a BioRuptor Plus (Diagenode) with settings: 30s ON, 60s OFF, HIGH intensity, 30 cycles. Sonicated DNA was concentrated with Agencourt AMPure XP beads (Beckman Coulter) and fragment sizes were assessed using a BioAnalyzer 2100 (Agilent Technologies) to range from 300bp to 800bp, with a peak around 400-600bp. To selectively and linearly amplify BLISS adapter-tagged genomic DSB ends, 100 ng of sonicated template was used for T7-mediated IVT using the MEGAscript T7 Transcription Kit (ThermoFisher #AMB13345, supplemented with Ribosafe RNAse Inhibitor (Bioline #BIO-65028)), according to the manufacturer's guidelines. Directly after RA3 ligation, reverse transcription was performed with Reverse Transcription Primer (RTP) (Illumina sequence, ordered via IDT) and SuperScript IV Reverse Transcriptase (ThermoFisher #18090050). The manufacturer's protocol was followed extending the incubation time to 50 minutes at 50°C followed by 10-minute heat inactivation at 80°C. Libraries amplification was carried out with NEBNext Ultra II Q5 Master Mix (NEB #M0544), RP1 common primer, and a selected RPIX index primer (Illumina sequences, ordered through IDT). Libraries were amplified for 8 PCR cycles, purified with a 0.8x AMPure XP bead purification, and then amplified for 4 additional PCR cycles. Then, the amplified libraries were cleaned-up according to the two-sided AMPure XP bead purification protocol, aiming at retaining library sizes from ~300–800bp. Final library profiles were assessed and quantified on a BioAnalyzer High Sensitivity DNA chip and using the Qubit dsDNA HS Assay kit (ThermoFisher #Q32851). Sequencing was performed at the Science for Life Laboratory, Sweden, on a NextSeq 500 with NextSeq 500/550 High Output Kit v2 chemistry for SE 1x75

sequencing with an additional 6 cycles for index sequencing. Multiple indexed BLISS libraries were pooled together, aiming to retrieve at least 50 million reads per condition/library. Upon completion of the run, raw sequencing reads were de-multiplexed based on index sequences by Illumina's BaseSpace, after which the generated FASTQ files were downloaded. The Breaks Labeling, Enrichment on Streptavidin and next-generation Sequencing (BLESS) validation experiments were performed previously described (Biernacka et al., 2018). The procedure resembles the sBLISS protocol and includes the *in situ* blunting of DSB ends, after mild fixation of the cells, and ligation to specialized biotinylated BLESS adapters, bearing the RA5 Illumina RNA sequence, that allow the selective affinity capture of DSBs. Upon ligation of the biotinylated adapter on DSBs, genomic DNA is purified and sonicated. Then, streptavidin beads (Dynabeads MyOne C1 #65001) are used to isolate DSB-bearing DNA fragments, followed by blunting of the other end and ligation to a second BLESS adapter containing the RA3 Illumina RNA adapter sequence. PCR amplification was performed according to Illumina's guidelines, for 10 cycles using the RA5 and RA3 adapters, followed by purification and specific-target qPCR amplification. Sequences of the BLISS and BLESS adapters were previously described (Biernacka et al., 2018; Yan et al., 2017) respectively. The RA3, RA5 adapters, RTP primer, and RP1 and RPIX primers, see the sequence information available for the Illumina small RNA library preparation kit.

Data analysis. Statistically significant data were extracted by means of the IBM SPSS Statistics 19 (IBM) and R-statistical package (www.r-project.org). Significant over-representation of pathways and gene networks was determined by Gene Ontology (http://geneontology.org/) and KEGG pathways (https://www.genome.jp/kegg/pathway.html). For mass spectrometry (MS), the MS/MS raw data were loaded in Proteome Discoverer 1.3.0.339 (ThermoFischer Scientific) and run using the Mascot 2.3.02 (Matrix Science) search algorithm against the Mus musculus theoretical proteome (last modified 6 July 2015) containing 46,470 entries in Uniprot. A list of common contaminants was included in the database67. For protein identification, the following search parameters were used: precursor error tolerance 10 ppm, fragment ion tolerance 0.8Da, trypsin full specificity, maximum number of missed cleavages 3 and cysteine alkylation as a fixed modification. The resulting .dat and .msf files were subsequently loaded and merged in Scaffold

(version 3.04.05, Proteome Software) for further processing and validation of the assigned MS/MS spectra. Thresholds for protein and peptide identification were set to 99% and 95% accordingly, for proteins with minimum 1 different peptides identified, resulting in a protein false discovery rate (FDR) of <0.1%. For single peptide identifications, we applied the same criteria in addition to manual validation of MS/MS spectra. Protein lists were constructed from the respective peptide lists through extensive manual curation based on previous knowledge. For label-free relative quantitation of proteins, we applied a label-free relative quantitation method between the different samples (control versus bait) to determine unspecific binders during the affinity purification. All .dat and .msf files created by Proteome Discoverer were merged in Scaffold where label-free relative quantification was performed using the total ion current (TIC) from each identified MS/MS spectra. The TIC is the sum of the areas under all the peaks contained in a MS/MS spectrum and total TIC value results by summing the intensity of the peaks contained in the peak list associated to a MS/MS sample. Protein lists containing the Scaffold-calculated total TIC quantitative value for each protein were exported to Microsoft Excel for further manual processing including categorization and additional curation based on previous knowledge. The fold change of protein levels was calculated by dividing the mean total TIC quantitative value in bait samples with the mean value of the control samples for each of the proteins. Proteins having ≥60% protein coverage, ≥1 peptide in each sample and a fold change ≥1.2 in all three measurements were selected as being significantly enriched in bXPF compared with BirA MEF samples. Proteins that were significantly enriched in bait samples were considered these with P value ≤0.05 and a fold change ≥2. Significant over-representation of pathways, protein-protein interactions and protein complexes were derived by STRING68 (http://string-db.org). The quality of ChIP-Seq raw reads was checked using FastQC software (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/). For both transcription factors (https://www.encodeproject.org/chip-seq/transcription_factor/) and histones (https://www.encodeproject.org/chip-seq/histone/) the appropriate pipelines proposed by ENCODE were adopted. All analyses were performed using as a reference the mm10 mouse genome from UCSC using Kundaje's lab ChIP-Seq pipeline and selecting the conservative set of peaks at the end. Peak annotation was performed using the HOMER Analysis package (Heinz et al., 2010). Peak visualization around TSS was performed using "ChIPSeeker" R package (Yu et al.,

2015). For sBLISS, the generated amplified RNA is sequenced using next-generation sequencing, after which the obtained reads are mapped to the reference genome to identify the genomic locations of the DSBs. As described previously (Yan et al., 2017), a custom-built pipeline was used to keep only those reads that contain the expected prefix of 8nt UMI and 8nt sample barcode, using SAMtools and scan for matches, allowing at most one mismatch in the barcode sequence. The prefixes were then clipped off and stored, and the trimmed reads per condition were aligned to the GRCm38/mm10 reference genome with BWA-MEM. Only those reads with mapping quality scores ≥ 30 were retained. Next, PCR duplicates were identified and removed, by searching for proximal reads (at most 30bp apart in the reference genome) with at most two mismatches in the UMI sequence. Finally, we generated BED files for downstream analyses, comprising a list of DSB end locations and a number of unique UMIs identified at these locations, which we refer to as 'UMI-DSB ends' or unique DSB ends. DSBs from all samples and all replicates have been annotated using HOMER software and a generic genome distribution (Intergenic, 3UTR, miRNA, ncRNA, TTS, pseudo, Exon, Intron, Promoter, 5UTR, snoRNA, rRNA) was created. The BLISS-ChIP Seq comparisons were performed using “bedtools” (https://bedtools.readthedocs.io/en/latest/).

5. REFERENCES

Adelman, K., and Lis, J.T. (2012). Promoter-proximal pausing of RNA polymerase II: emerging roles in metazoans. Nat Rev Genet *13*, 720-731.

Agathangelou, K., Apostolou, Z., and Garinis, G.A. (2018). Nuclear DNA Damage and Ageing. Subcell Biochem *90*, 309-322.

Aguilera, A., and Garcia-Muse, T. (2012). R loops: from transcription byproducts to threats to genome stability. Mol Cell *46*, 115-124.

Aivaliotis, M., Gevaert, K., Falb, M., Tebbe, A., Konstantinidis, K., Bisle, B., Klein, C., Martens, L., Staes, A., Timmerman, E.*, et al.* (2007). Large-scale identification of N-terminal peptides in the halophilic archaea Halobacterium salinarum and Natronomonas pharaonis. J Proteome Res *6*, 2195-2204.

Alekseev, S., Nagy, Z., Sandoz, J., Weiss, A., Egly, J.M., Le May, N., and Coin, F. (2017). Transcription without XPB Establishes a Unified Helicase-Independent Mechanism of Promoter Opening in Eukaryotic Gene Expression. Mol Cell *65*, 504-514 e504.

Altmeyer, M., Toledo, L., Gudjonsson, T., Grofte, M., Rask, M.B., Lukas, C., Akimov, V., Blagoev, B., Bartek, J., and Lukas, J. (2013). The chromatin scaffold protein SAFB1 renders chromatin permissive for DNA damage signaling. Molecular cell *52*, 206-220.

Andressoo, J.O., Mitchell, J.R., de Wit, J., Hoogstraten, D., Volker, M., Toussaint, W., Speksnijder, E., Beems, R.B., van Steeg, H., Jans, J.*, et al.* (2006). An Xpd mouse model for the combined xeroderma pigmentosum/Cockayne syndrome exhibiting both cancer predisposition and segmental progeria. Cancer Cell *10*, 121-132.

Apostolou, Z., Chatzinikolaou, G., Stratigi, K., and Garinis, G.A. (2019). Nucleotide Excision Repair and Transcription-Associated Genome Instability. Bioessays *41*, e1800201.

Araki, M., Masutani, C., Takemura, M., Uchida, A., Sugasawa, K., Kondoh, J., Ohkuma, Y., and Hanaoka, F. (2001). Centrosome protein centrin 2/caltractin 1 is part of the xeroderma pigmentosum group C complex that initiates global genome nucleotide excision repair. J Biol Chem *276*, 18665-18672.

Araujo, S.J., Tirode, F., Coin, F., Pospiech, H., Syvaoja, J.E., Stucki, M., Hubscher, U., Egly, J.M., and Wood, R.D. (2000). Nucleotide excision repair of DNA with recombinant human proteins: definition of the minimal set of factors, active forms of TFIIH, and modulation by CAK. Genes Dev *14*, 349-359.

Artelt, P., Grannemann, R., Stocking, C., Friel, J., Bartsch, J., and Hauser, H. (1991). The prokaryotic neomycin-resistance-encoding gene acts as a transcriptional silencer in eukaryotic cells. Gene *99*, 249-254.

Ayoub, N., Jeyasekharan, A.D., Bernal, J.A., and Venkitaraman, A.R. (2008). HP1-beta mobilization promotes chromatin changes that initiate the DNA damage response. Nature *453*, 682-686.

Barreto, G., Schafer, A., Marhold, J., Stach, D., Swaminathan, S.K., Handa, V., Doderlein, G., Maltry, N., Wu, W., Lyko, F.*, et al.* (2007). Gadd45a promotes epigenetic gene activation by repair-mediated DNA demethylation. Nature *445*, 671-675.

Bastien, J., and Rochette-Egly, C. (2004). Nuclear retinoid receptors and the transcription of retinoid-target genes. Gene *328*, 1-16.

Batty, D., Rapic'-Otrin, V., Levine, A.S., and Wood, R.D. (2000). Stable binding of human XPC complex to irradiated DNA confers strong discrimination for damaged sites. J Mol Biol *300*, 275-290.
Beckett, D., Kovaleva, E., and Schatz, P.J. (1999). A minimal peptide substrate in biotin holoenzyme synthetase-catalyzed biotinylation. Protein science : a publication of the Protein Society *8*, 921-929.
Beerens, N., Hoeijmakers, J.H., Kanaar, R., Vermeulen, W., and Wyman, C. (2005). The CSB protein actively wraps DNA. J Biol Chem *280*, 4722-4729.
Bergink, S., Toussaint, W., Luijsterburg, M.S., Dinant, C., Alekseev, S., Hoeijmakers, J.H., Dantuma, N.P., Houtsmuller, A.B., and Vermeulen, W. (2012). Recognition of DNA damage by XPC coincides with disruption of the XPC-RAD23 complex. J Cell Biol *196*, 681-688.
Berneburg, M., Lowe, J.E., Nardo, T., Araujo, S., Fousteri, M.I., Green, M.H., Krutmann, J., Wood, R.D., Stefanini, M., and Lehmann, A.R. (2000). UV damage causes uncontrolled DNA breakage in cells from patients with combined features of XP-D and Cockayne syndrome. EMBO J *19*, 1157-1166.
Biernacka, A., Zhu, Y., Skrzypczak, M., Forey, R., Pardo, B., Grzelak, M., Nde, J., Mitra, A., Kudlicki, A., Crosetto, N.*, et al.* (2018). i-BLESS is an ultra-sensitive method for detection of DNA double-strand breaks. Commun Biol *1*, 181.
Birger, Y., West, K.L., Postnikov, Y.V., Lim, J.H., Furusawa, T., Wagner, J.P., Laufer, C.S., Kraemer, K.H., and Bustin, M. (2003). Chromosomal protein HMGN1 enhances the rate of DNA repair in chromatin. Embo J *22*, 1665-1675.
Boboila, C., Alt, F.W., and Schwer, B. (2012). Classical and alternative end-joining pathways for repair of lymphocyte-specific and general DNA double-strand breaks. Advances in immunology *116*, 1-49.
Bondar, T., Kalinina, A., Khair, L., Kopanja, D., Nag, A., Bagchi, S., and Raychaudhuri, P. (2006). Cul4A and DDB1 associate with Skp2 to target p27Kip1 for proteolysis involving the COP9 signalosome. Mol Cell Biol *26*, 2531-2539.
Bootsma, D., Kraemer,K.H., Cleaver,J.E., Hoeijmakers,J.H.J. (1998). Nucleotide excision repair syndromes: xeroderma pigmentosum, Cockayne syndrome and trichothiodystrophy. In The genetic basis of human cancer, B. Vogelstein, Kinzler, K.W., ed. (McGraw-Hill New-York), pp. 245-274.
Bootsma, D.K., KH. Cleaver, JE. Hoeijmakers, JHJ. (2001). The Metabolic and Molecular Basis of inherited Disease (New York: McGraw-Hill).
Bradford, P.T., Goldstein, A.M., Tamura, D., Khan, S.G., Ueda, T., Boyle, J., Oh, K.S., Imoto, K., Inui, H., Moriwaki, S.*, et al.* (2011). Cancer and neurologic degeneration in xeroderma pigmentosum: long term follow-up characterises the role of DNA repair. J Med Genet *48*, 168-176.
Brueckner, F., Hennecke, U., Carell, T., and Cramer, P. (2007). CPD damage recognition by transcribing RNA polymerase II. Science *315*, 859-862.
Bunting, M., Bernstein, K.E., Greer, J.M., Capecchi, M.R., and Thomas, K.R. (1999). Targeting genes for self-excision in the germ line. Genes & development *13*, 1524-1528.
Butuci, M., Williams, A.B., Wong, M.M., Kramer, B., and Michael, W.M. (2015). Zygotic Genome Activation Triggers Chromosome Damage and Checkpoint Signaling in C. elegans Primordial Germ Cells. Dev Cell *34*, 85-95.
Castillo-Davis, C.I., Mekhedov, S.L., Hartl, D.L., Koonin, E.V., and Kondrashov, F.A. (2002). Selection for short introns in highly expressed genes. Nat Genet *31*, 415-418.

Cejka, P. (2015). DNA end resection: nucleases team up with the right partners to initiate homologous recombination. J Biol Chem.
Chalut, C., Moncollin, V., and Egly, J.M. (1994). Transcription by RNA polymerase II: a process linked to DNA repair. Bioessays *16*, 651-655.
Chatzinikolaou, G., Apostolou, Z., Aid-Pavlidis, T., Ioannidou, A., Karakasilioti, I., Papadopoulos, G.L., Aivaliotis, M., Tsekrekou, M., Strouboulis, J., Kosteas, T.*, et al.* (2017). ERCC1-XPF cooperates with CTCF and cohesin to facilitate the developmental silencing of imprinted genes. Nat Cell Biol *19*, 421-432.
Chavanne, F., Broughton, B.C., Pietra, D., Nardo, T., Browitt, A., Lehmann, A.R., and Stefanini, M. (2000). Mutations in the XPC gene in families with xeroderma pigmentosum and consequences at the cell, protein, and transcript levels. Cancer Res *60*, 1974-1982.
Chen, D., Lucey, M.J., Phoenix, F., Lopez-Garcia, J., Hart, S.M., Losson, R., Buluwela, L., Coombes, R.C., Chambon, P., Schar, P.*, et al.* (2003). T:G mismatch-specific thymine-DNA glycosylase potentiates transcription of estrogen-regulated genes through direct interaction with estrogen receptor alpha. J Biol Chem *278*, 38586-38592.
Cheung, A.C., and Cramer, P. (2011). Structural basis of RNA polymerase II backtracking, arrest and reactivation. Nature *471*, 249-253.
Chrzanowska, K.H., Gregorek, H., Dembowska-Baginska, B., Kalina, M.A., and Digweed, M. (2012). Nijmegen breakage syndrome (NBS). Orphanet J Rare Dis *7*, 13.
Citterio, E., Rademakers, S., van der Horst, G.T., van Gool, A.J., Hoeijmakers, J.H., and Vermeulen, W. (1998). Biochemical and biological characterization of wild-type and ATPase-deficient Cockayne syndrome B repair protein. J Biol Chem *273*, 11844-11851.
Citterio, E., Van Den Boom, V., Schnitzler, G., Kanaar, R., Bonte, E., Kingston, R.E., Hoeijmakers, J.H., and Vermeulen, W. (2000). ATP-dependent chromatin remodeling by the Cockayne syndrome B DNA repair-transcription-coupling factor. Mol Cell Biol *20*, 7643-7653.
Coin, F., Oksenych, V., and Egly, J.M. (2007). Distinct roles for the XPB/p52 and XPD/p44 subcomplexes of TFIIH in damaged DNA opening during nucleotide excision repair. Mol Cell *26*, 245-256.
Colella, S., Nardo, T., Botta, E., Lehmann, A.R., and Stefanini, M. (2000). Identical mutations in the CSB gene associated with either Cockayne syndrome or the DeSanctis-cacchione variant of xeroderma pigmentosum. Hum Mol Genet *9*, 1171-1175.
Compe, E., and Egly, J.M. (2016). Nucleotide Excision Repair and Transcriptional Regulation: TFIIH and Beyond. Annu Rev Biochem *85*, 265-290.
Cortazar, D., Kunz, C., Saito, Y., Steinacher, R., and Schar, P. (2007). The enigmatic thymine DNA glycosylase. DNA Repair (Amst) *6*, 489-504.
Cortazar, D., Kunz, C., Selfridge, J., Lettieri, T., Saito, Y., MacDougall, E., Wirz, A., Schuermann, D., Jacobs, A.L., Siegrist, F.*, et al.* (2011). Embryonic lethal phenotype reveals a function of TDG in maintaining epigenetic stability. Nature *470*, 419-423.
Cortellino, S., Xu, J., Sannai, M., Moore, R., Caretti, E., Cigliano, A., Le Coz, M., Devarajan, K., Wessels, A., Soprano, D.*, et al.* (2011). Thymine DNA glycosylase is essential for active DNA demethylation by linked deamination-base excision repair. Cell *146*, 67-79.
Costantino, L., and Koshland, D. (2015). The Yin and Yang of R-loop biology. Curr Opin Cell Biol *34*, 39-45.

Crosetto, N., Mitra, A., Silva, M.J., Bienko, M., Dojer, N., Wang, Q., Karaca, E., Chiarle, R., Skrzypczak, M., Ginalski, K.*, et al.* (2013). Nucleotide-resolution DNA double-strand break mapping by next-generation sequencing. Nat Methods *10*, 361-365.
Cull, M.G., and Schatz, P.J. (2000). Biotinylation of proteins in vivo and in vitro using small peptide tags. Methods in enzymology *326*, 430-440.
Danko, C.G., Hah, N., Luo, X., Martins, A.L., Core, L., Lis, J.T., Siepel, A., and Kraus, W.L. (2013). Signaling pathways differentially affect RNA polymerase II initiation, pausing, and elongation rate in cells. Mol Cell *50*, 212-222.
de Boer, E., Rodriguez, P., Bonte, E., Krijgsveld, J., Katsantoni, E., Heck, A., Grosveld, F., and Strouboulis, J. (2003). Efficient biotinylation and single-step purification of tagged transcription factors in mammalian cells and transgenic mice. Proceedings of the National Academy of Sciences of the United States of America *100*, 7480-7485.
de Boer, J., Andressoo, J.O., de Wit, J., Huijmans, J., Beems, R.B., van Steeg, H., Weeda, G., van der Horst, G.T., van Leeuwen, W., Themmen, A.P.*, et al.* (2002). Premature aging in mice deficient in DNA repair and transcription. Science *296*, 1276-1279.
de Laat, W., and Grosveld, F. (2003). Spatial organization of gene expression: the active chromatin hub. Chromosome Res *11*, 447-459.
Deweese, J.E., and Osheroff, N. (2009). The DNA cleavage reaction of topoisomerase II: wolf in sheep's clothing. Nucleic acids research *37*, 738-748.
Di Noia, J., and Neuberger, M.S. (2002). Altering the pathway of immunoglobulin hypermutation by inhibiting uracil-DNA glycosylase. Nature *419*, 43-48.
Di Noia, J.M., and Neuberger, M.S. (2007). Molecular mechanisms of antibody somatic hypermutation. Annu Rev Biochem *76*, 1-22.
DiGiovanna, J.J., and Kraemer, K.H. (2012). Shining a light on xeroderma pigmentosum. J Invest Dermatol *132*, 785-796.
Donahue, B.A., Yin, S., Taylor, J.S., Reines, D., and Hanawalt, P.C. (1994). Transcript cleavage by RNA polymerase II arrested by a cyclobutane pyrimidine dimer in the DNA template. Proc Natl Acad Sci U S A *91*, 8502-8506.
Driegen, S., Ferreira, R., van Zon, A., Strouboulis, J., Jaegle, M., Grosveld, F., Philipsen, S., and Meijer, D. (2005). A generic tool for biotinylation of tagged proteins in transgenic mice. Transgenic research *14*, 477-482.
Duan, Z., Andronescu, M., Schutz, K., Lee, C., Shendure, J., Fields, S., Noble, W.S., and Anthony Blau, C. (2012). A genome-wide 3C-method for characterizing the three-dimensional architectures of genomes. Methods *58*, 277-288.
Egly, J.M., and Coin, F. (2011). A history of TFIIH: two decades of molecular biology on a pivotal transcription/repair factor. DNA Repair (Amst) *10*, 714-721.
Einhauer, A., and Jungbauer, A. (2001). The FLAG peptide, a versatile fusion tag for the purification of recombinant proteins. Journal of biochemical and biophysical methods *49*, 455-465.
Eisen, J.A., Sweder, K.S., and Hanawalt, P.C. (1995). Evolution of the SNF2 family of proteins: subfamilies with distinct sequences and functions. Nucleic Acids Res *23*, 2715-2723.
Fan, L., Fuss, J.O., Cheng, Q.J., Arvai, A.S., Hammel, M., Roberts, V.A., Cooper, P.K., and Tainer, J.A. (2008). XPD helicase structures and activities: insights into the cancer and aging phenotypes from XPD mutations. Cell *133*, 789-800.

Fan, S., Wang, J., Yuan, R., Ma, Y., Meng, Q., Erdos, M.R., Pestell, R.G., Yuan, F., Auborn, K.J., Goldberg, I.D.*, et al.* (1999). BRCA1 inhibition of estrogen receptor signaling in transfected cells. Science *284*, 1354-1356.
Fong, Y.W., Inouye, C., Yamaguchi, T., Cattoglio, C., Grubisic, I., and Tjian, R. (2011). A DNA repair complex functions as an Oct4/Sox2 coactivator in embryonic stem cells. Cell *147*, 120-131.
Fousteri, M., Vermeulen, W., van Zeeland, A.A., and Mullenders, L.H. (2006). Cockayne syndrome A and B proteins differentially regulate recruitment of chromatin remodeling and repair factors to stalled RNA polymerase II in vivo. Mol Cell *23*, 471-482.
Friedrich, G., and Soriano, P. (1991). Promoter traps in embryonic stem cells: a genetic screen to identify and mutate developmental genes in mice. Genes & development *5*, 1513-1523.
Gaiddon, C., Moorthy, N.C., and Prives, C. (1999). Ref-1 regulates the transactivation and pro-apoptotic functions of p53 in vivo. EMBO J *18*, 5609-5621.
Gaillard, H., and Aguilera, A. (2016). Transcription as a Threat to Genome Integrity. Annu Rev Biochem *85*, 291-317.
Garinis, G.A., Uittenboogaard, L.M., Stachelscheid, H., Fousteri, M., van Ijcken, W., Breit, T.M., van Steeg, H., Mullenders, L.H., van der Horst, G.T., and Brüning, J.C. (2009). Persistent transcription-blocking DNA lesions trigger somatic growth attenuation associated with longevity. Nature cell biology *11*, 604-615.
Garinis, G.A., van der Horst, G.T., Vijg, J., and Hoeijmakers, J.H. (2008). DNA damage and ageing: new-age ideas for an age-old problem. Nat Cell Biol *10*, 1241-1247.
Giglia-Mari, G., Theil, A.F., Mari, P.O., Mourgues, S., Nonnekens, J., Andrieux, L.O., de Wit, J., Miquel, C., Wijgers, N., Maas, A.*, et al.* (2009). Differentiation driven changes in the dynamic organization of Basal transcription initiation. PLoS Biol *7*, e1000220.
Ginno, P.A., Lott, P.L., Christensen, H.C., Korf, I., and Chedin, F. (2012). R-loop formation is a distinctive characteristic of unmethylated human CpG island promoters. Molecular cell *45*, 814-825.
Goodarzi, A.A., Noon, A.T., Deckbar, D., Ziv, Y., Shiloh, Y., Lobrich, M., and Jeggo, P.A. (2008). ATM signaling facilitates repair of DNA double-strand breaks associated with heterochromatin. Molecular cell *31*, 167-177.
Groisman, R., Polanowska, J., Kuraoka, I., Sawada, J., Saijo, M., Drapkin, R., Kisselev, A.F., Tanaka, K., and Nakatani, Y. (2003). The ubiquitin ligase activity in the DDB2 and CSA complexes is differentially regulated by the COP9 signalosome in response to DNA damage. Cell *113*, 357-367.
Grosveld, F., Rodriguez, P., Meier, N., Krpic, S., Pourfarzad, F., Papadopoulos, P., Kolodziej, K., Patrinos, G.P., Hostert, A., and Strouboulis, J. (2005). Isolation and characterization of hematopoietic transcription factor complexes by in vivo biotinylation tagging and mass spectrometry. Ann N Y Acad Sci *1054*, 55-67.
Guerrero-Santoro, J., Kapetanaki, M.G., Hsieh, C.L., Gorbachinsky, I., Levine, A.S., and Rapic-Otrin, V. (2008). The cullin 4B-based UV-damaged DNA-binding protein ligase binds to UV-damaged chromatin and ubiquitinates histone H2A. Cancer Res *68*, 5014-5022.
Guglielmi, B., Soutourina, J., Esnault, C., and Werner, M. (2007). TFIIS elongation factor and Mediator act in conjunction during transcription initiation in vivo. Proc Natl Acad Sci U S A *104*, 16062-16067.

Guo, J., and Price, D.H. (2013). RNA polymerase II transcription elongation control. Chem Rev *113*, 8583-8603.
Haffner, M.C., Aryee, M.J., Toubaji, A., Esopi, D.M., Albadine, R., Gurel, B., Isaacs, W.B., Bova, G.S., Liu, W., Xu, J.*, et al.* (2010). Androgen-induced TOP2B-mediated double-strand breaks and prostate cancer gene rearrangements. Nat Genet *42*, 668-675.
Haffner, M.C., De Marzo, A.M., Meeker, A.K., Nelson, W.G., and Yegnasubramanian, S. (2011). Transcription-induced DNA double strand breaks: both oncogenic force and potential therapeutic target? Clin Cancer Res *17*, 3858-3864.
Hanawalt, P.C., and Spivak, G. (2008). Transcription-coupled DNA repair: two decades of progress and surprises. Nat Rev Mol Cell Biol *9*, 958-970.
Hasty, P., Campisi, J., Hoeijmakers, J., van Steeg, H., and Vijg, J. (2003). Aging and genome maintenance: lessons from the mouse? Science *299*, 1355-1359.
Heintzman, N.D., Hon, G.C., Hawkins, R.D., Kheradpour, P., Stark, A., Harp, L.F., Ye, Z., Lee, L.K., Stuart, R.K., Ching, C.W.*, et al.* (2009). Histone modifications at human enhancers reflect global cell-type-specific gene expression. Nature *459*, 108-112.
Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y.C., Laslo, P., Cheng, J.X., Murre, C., Singh, H., and Glass, C.K. (2010). Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. Mol Cell *38*, 576-589.
Heller, E.R., Khan, S.G., Kuschal, C., Tamura, D., DiGiovanna, J.J., and Kraemer, K.H. (2015). Mutations in the TTDN1 gene are associated with a distinct trichothiodystrophy phenotype. J Invest Dermatol *135*, 734-741.
Hill, S.J., Rolland, T., Adelmant, G., Xia, X., Owen, M.S., Dricot, A., Zack, T.I., Sahni, N., Jacob, Y., Hao, T.*, et al.* (2014). Systematic screening reveals a role for BRCA1 in the response to transcription-associated DNA damage. Genes & development *28*, 1957-1975.
Ho, J.J., Cattoglio, C., McSwiggen, D.T., Tjian, R., and Fong, Y.W. (2017). Regulation of DNA demethylation by the XPC DNA repair complex in somatic and pluripotent stem cells. Genes Dev *31*, 830-844.
Hoeijmakers, J.H. (2009). DNA damage, aging, and cancer. N Engl J Med *361*, 1475-1485.
Hoogstraten, D., Bergink, S., Ng, J.M., Verbiest, V.H., Luijsterburg, M.S., Geverts, B., Raams, A., Dinant, C., Hoeijmakers, J.H., Vermeulen, W.*, et al.* (2008). Versatile DNA damage detection by the global genome nucleotide excision repair protein XPC. J Cell Sci *121*, 2850-2859.
Huang, H.S., Allen, J.A., Mabb, A.M., King, I.F., Miriyala, J., Taylor-Blake, B., Sciaky, N., Dutton, J.W., Jr., Lee, H.M., Chen, X.*, et al.* (2012). Topoisomerase inhibitors unsilence the dormant allele of Ube3a in neurons. Nature *481*, 185-189.
Itin, P.H., Sarasin, A., and Pittelkow, M.R. (2001). Trichothiodystrophy: update on the sulfur-deficient brittle hair syndromes. J Am Acad Dermatol *44*, 891-920; quiz 921-894.
Ito, S., Yamane, M., Ohtsuka, S., and Niwa, H. (2014). The C-terminal region of Xpc is dispensable for the transcriptional activity of Oct3/4 in mouse embryonic stem cells. FEBS Lett *588*, 1128-1135.
Ito, T., Arimitsu, N., Takeuchi, M., Kawamura, N., Nagata, M., Saso, K., Akimitsu, N., Hamamoto, H., Natori, S., Miyajima, A.*, et al.* (2006). Transcription elongation factor

S-II is required for definitive hematopoiesis. Molecular and cellular biology *26*, 3194-3203.
Iyama, T., and Wilson, D.M., 3rd (2016). Elements That Regulate the DNA Damage Response of Proteins Defective in Cockayne Syndrome. J Mol Biol *428*, 62-78.
Jiang, Y., Wang, X., Bao, S., Guo, R., Johnson, D.G., Shen, X., and Li, L. (2010). INO80 chromatin remodeling complex promotes the removal of UV lesions by the nucleotide excision repair pathway. Proc Natl Acad Sci U S A *107*, 17274-17279.
Johnson, R.E., Kondratick, C.M., Prakash, S., and Prakash, L. (1999a). hRAD30 mutations in the variant form of xeroderma pigmentosum. Science *285*, 263-265.
Johnson, R.E., Prakash, S., and Prakash, L. (1999b). Efficient bypass of a thymine-thymine dimer by yeast DNA polymerase, Poleta. Science *283*, 1001-1004.
Ju, B.G., Lunyak, V.V., Perissi, V., Garcia-Bassets, I., Rose, D.W., Glass, C.K., and Rosenfeld, M.G. (2006). A topoisomerase IIbeta-mediated dsDNA break required for regulated transcription. Science *312*, 1798-1802.
Kamileri, I., Karakasilioti, I., and Garinis, G.A. (2012a). Nucleotide excision repair: new tricks with old bricks. Trends Genet *28*, 566-573.
Kamileri, I., Karakasilioti, I., and Garinis, G.A. (2012b). Nucleotide excision repair: new tricks with old bricks. Trends in Genetics *28*, 566-573.
Kamileri, I., Karakasilioti, I., Sideri, A., Kosteas, T., Tatarakis, A., Talianidis, I., and Garinis, G.A. (2012c). Defective transcription initiation causes postnatal growth failure in a mouse model of nucleotide excision repair (NER) progeria. Proc Natl Acad Sci U S A *109*, 2995-3000.
Kamileri, I., Karakasilioti, I., Sideri, A., Kosteas, T., Tatarakis, A., Talianidis, I., and Garinis, G.A. (2012d). Defective transcription initiation causes postnatal growth failure in a mouse model of nucleotide excision repair (NER) progeria. Proceedings of the National Academy of Sciences *109*, 2995-3000.
Karakasilioti, I., Kamileri, I., Chatzinikolaou, G., Kosteas, T., Vergadi, E., Robinson, A.R., Tsamardinos, I., Rozgaja, T.A., Siakouli, S., Tsatsanis, C*., et al.* (2013). DNA damage triggers a chronic autoinflammatory response, leading to fat depletion in NER progeria. Cell Metab *18*, 403-415.
Kim, B., Nesvizhskii, A.I., Rani, P.G., Hahn, S., Aebersold, R., and Ranish, J.A. (2007). The transcription elongation factor TFIIS is a component of RNA polymerase II preinitiation complexes. Proc Natl Acad Sci U S A *104*, 16068-16073.
Koch, S.C., Simon, N., Ebert, C., and Carell, T. (2016). Molecular mechanisms of xeroderma pigmentosum (XP) proteins. Q Rev Biophys *49*, e5.
Krumm, A., Hickey, L.B., and Groudine, M. (1995). Promoter-proximal pausing of RNA polymerase II defines a general rate-limiting step after transcription initiation. Genes Dev *9*, 559-572.
Lake, R.J., Boetefuer, E.L., Won, K.J., and Fan, H.Y. (2016). The CSB chromatin remodeler and CTCF architectural protein cooperate in response to oxidative stress. Nucleic Acids Res *44*, 2125-2135.
Lambert, W.C., Gagna, C.E., and Lambert, M.W. (2010). Trichothiodystrophy: Photosensitive, TTD-P, TTD, Tay syndrome. Adv Exp Med Biol *685*, 106-110.
Laugel, V. (2013). Cockayne syndrome: the expanding clinical and mutational spectrum. Mech Ageing Dev *134*, 161-170.
Le May, N., Egly, J.M., and Coin, F. (2010a). True lies: the double life of the nucleotide excision repair factors in transcription and DNA repair. J Nucleic Acids *2010*.

Le May, N., Fradin, D., Iltis, I., Bougneres, P., and Egly, J.M. (2012). XPG and XPF endonucleases trigger chromatin looping and DNA demethylation for accurate expression of activated genes. Mol Cell *47*, 622-632.
Le May, N., Mota-Fernandes, D., Velez-Cruz, R., Iltis, I., Biard, D., and Egly, J.M. (2010b). NER factors are recruited to active promoters and facilitate chromatin modification for transcription in the absence of exogenous genotoxic attack. Mol Cell *38*, 54-66.
Lemon, B., and Tjian, R. (2000). Orchestrated response: a symphony of transcription factors for gene control. Genes Dev *14*, 2551-2569.
Li, X., and Heyer, W.D. (2008). Homologous recombination in DNA repair and DNA damage tolerance. Cell Res *18*, 99-113.
Li, X., Niu, T., and Manley, J.L. (2007a). The RNA binding protein RNPS1 alleviates ASF/SF2 depletion-induced genomic instability. RNA *13*, 2108-2115.
Li, Y.Q., Zhou, P.Z., Zheng, X.D., Walsh, C.P., and Xu, G.L. (2007b). Association of Dnmt3a and thymine DNA glycosylase links DNA methylation with base-excision repair. Nucleic Acids Res *35*, 390-400.
Lin, C., Yang, L., Tanasa, B., Hutt, K., Ju, B.G., Ohgi, K., Zhang, J., Rose, D.W., Fu, X.D., Glass, C.K.*, et al.* (2009). Nuclear receptor-induced chromosomal proximity and DNA breaks underlie specific translocations in cancer. Cell *139*, 1069-1083.
Lindahl, T., and Barnes, D.E. (2000). Repair of endogenous DNA damage. Cold Spring Harb Symp Quant Biol *65*, 127-133.
Luijsterburg, M.S., Dinant, C., Lans, H., Stap, J., Wiernasz, E., Lagerwerf, S., Warmerdam, D.O., Lindh, M., Brink, M.C., Dobrucki, J.W.*, et al.* (2009). Heterochromatin protein 1 is recruited to various types of DNA damage. The Journal of cell biology *185*, 577-586.
Luijsterburg, M.S., Lindh, M., Acs, K., Vrouwe, M.G., Pines, A., van Attikum, H., Mullenders, L.H., and Dantuma, N.P. (2012). DDB2 promotes chromatin decondensation at UV-induced DNA damage. J Cell Biol *197*, 267-281.
Mahindra, P., DiGiovanna, J.J., Tamura, D., Brahim, J.S., Hornyak, T.J., Stern, J.B., Lee, C.C., Khan, S.G., Brooks, B.P., Smith, J.A.*, et al.* (2008). Skin cancers, blindness, and anterior tongue mass in African brothers. J Am Acad Dermatol *59*, 881-886.
Malik, S., and Roeder, R.G. (2005). Dynamic regulation of pol II transcription by the mammalian Mediator complex. Trends Biochem Sci *30*, 256-263.
Manandhar, M., Lowery, M.G., Boulware, K.S., Lin, K.H., Lu, Y., and Wood, R.D. (2017). Transcriptional consequences of XPA disruption in human cell lines. DNA Repair (Amst) *57*, 76-90.
Mareddy, S., Reddy, J., Babu, S., and Balan, P. (2013). Xeroderma pigmentosum: man deprived of his right to light. ScientificWorldJournal *2013*, 534752.
Marteijn, J.A., Lans, H., Vermeulen, W., and Hoeijmakers, J.H. (2014). Understanding nucleotide excision repair and its roles in cancer and ageing. Nat Rev Mol Cell Biol *15*, 465-481.
Martin, G.M. (2005). Genetic modulation of senescent phenotypes in Homo sapiens. Cell *120*, 523-532.
Masutani, C., Kusumoto, R., Yamada, A., Dohmae, N., Yokoi, M., Yuasa, M., Araki, M., Iwai, S., Takio, K., and Hanaoka, F. (1999). The XPV (xeroderma pigmentosum variant) gene encodes human DNA polymerase eta. Nature *399*, 700-704.
Mathieu, N., Kaczmarek, N., Ruthemann, P., Luch, A., and Naegeli, H. (2013). DNA quality control by a lesion sensor pocket of the xeroderma pigmentosum group D helicase subunit of TFIIH. Curr Biol *23*, 204-212.

Mehta, G.D., Kumar, R., Srivastava, S., and Ghosh, S.K. (2013). Cohesin: functions beyond sister chromatid cohesion. FEBS letters *587*, 2299-2312.
Metivier, R., Gallais, R., Tiffoche, C., Le Peron, C., Jurkowska, R.Z., Carmouche, R.P., Ibberson, D., Barath, P., Demay, F., Reid, G.*, et al.* (2008). Cyclical DNA methylation of a transcriptionally active promoter. Nature *452*, 45-50.
Miller, G., and Hahn, S. (2006). A DNA-tethered cleavage probe reveals the path for promoter DNA in the yeast preinitiation complex. Nat Struct Mol Biol *13*, 603-610.
Mills, R.E., Bennett, E.A., Iskow, R.C., Luttig, C.T., Tsui, C., Pittard, W.S., and Devine, S.E. (2006). Recently mobilized transposons in the human and chimpanzee genomes. Am J Hum Genet *78*, 671-679.
Mischo, H.E., Gomez-Gonzalez, B., Grzechnik, P., Rondon, A.G., Wei, W., Steinmetz, L., Aguilera, A., and Proudfoot, N.J. (2011). Yeast Sen1 helicase protects the genome from transcription-associated instability. Molecular cell *41*, 21-32.
Mocquet, V., Laine, J.P., Riedl, T., Yajin, Z., Lee, M.Y., and Egly, J.M. (2008). Sequential recruitment of the repair factors during NER: the role of XPG in initiating the resynthesis step. EMBO J *27*, 155-167.
Moser, J., Kool, H., Giakzidis, I., Caldecott, K., Mullenders, L.H., and Fousteri, M.I. (2007). Sealing of chromosomal DNA nicks during nucleotide excision repair requires XRCC1 and DNA ligase III alpha in a cell-cycle-specific manner. Mol Cell *27*, 311-323.
Muers, M. (2011). Mutation: the perils of transcription. Nat Rev Genet *12*, 156.
Nakabayashi, K., Amann, D., Ren, Y., Saarialho-Kere, U., Avidan, N., Gentles, S., MacDonald, J.R., Puffenberger, E.G., Christiano, A.M., Martinez-Mir, A.*, et al.* (2005). Identification of C7orf11 (TTDN1) gene mutations and genetic heterogeneity in nonphotosensitive trichothiodystrophy. Am J Hum Genet *76*, 510-516.
Nakane, H., Takeuchi, S., Yuba, S., Saijo, M., Nakatsu, Y., Murai, H., Nakatsuru, Y., Ishikawa, T., Hirota, S., Kitamura, Y.*, et al.* (1995). High incidence of ultraviolet-B-or chemical-carcinogen-induced skin tumours in mice lacking the xeroderma pigmentosum group A gene. Nature *377*, 165-168.
Nakatsu, Y., Asahina, H., Citterio, E., Rademakers, S., Vermeulen, W., Kamiuchi, S., Yeo, J.P., Khaw, M.C., Saijo, M., Kodo, N.*, et al.* (2000). XAB2, a novel tetratricopeptide repeat protein involved in transcription-coupled DNA repair and transcription. J Biol Chem *275*, 34931-34937.
Nance, M.A., and Berry, S.A. (1992). Cockayne syndrome: review of 140 cases. Am J Med Genet *42*, 68-84.
Narita, T., Narita, K., Takedachi, A., Saijo, M., and Tanaka, K. (2015). Regulation of Transcription Elongation by the XPG-TFIIH Complex Is Implicated in Cockayne Syndrome. Mol Cell Biol *35*, 3178-3188.
Niedernhofer, L.J. (2008). Nucleotide excision repair deficient mouse models and neurological disease. DNA Repair (Amst) *7*, 1180-1189.
Niedernhofer, L.J., Garinis, G.A., Raams, A., Lalai, A.S., Robinson, A.R., Appeldoorn, E., Odijk, H., Oostendorp, R., Ahmad, A., and Van Leeuwen, W. (2006a). A new progeroid syndrome reveals that genotoxic stress suppresses the somatotroph axis. Nature *444*, 1038-1043.
Niedernhofer, L.J., Garinis, G.A., Raams, A., Lalai, A.S., Robinson, A.R., Appeldoorn, E., Odijk, H., Oostendorp, R., Ahmad, A., van Leeuwen, W.*, et al.* (2006b). A new progeroid syndrome reveals that genotoxic stress suppresses the somatotroph axis. Nature *444*, 1038-1043.
Nock, A., Ascano, J.M., Barrero, M.J., and Malik, S. (2012). Mediator-regulated transcription through the +1 nucleosome. Molecular cell *48*, 837-848.

Nudler, E. (2012). RNA polymerase backtracking in gene regulation and genome instability. Cell *149*, 1438-1445.
O'Gorman, S., Dagenais, N.A., Qian, M., and Marchuk, Y. (1997). Protamine-Cre recombinase transgenes efficiently recombine target sequences in the male germ line of mice, but not in embryonic stem cells. Proc Natl Acad Sci U S A *94*, 14602-14607.
Ogi, T., Limsirichaikul, S., Overmeer, R.M., Volker, M., Takenaka, K., Cloney, R., Nakazawa, Y., Niimi, A., Miki, Y., Jaspers, N.G.*, et al.* (2010). Three DNA polymerases, recruited by different mechanisms, carry out NER repair synthesis in human cells. Mol Cell *37*, 714-727.
Ohler, U., and Wassarman, D.A. (2010). Promoting developmental transcription. Development *137*, 15-26.
Oksenych, V., de Jesus, B.B., Zhovmer, A., Egly, J.M., and Coin, F. (2009). Molecular insights into the recruitment of TFIIH to sites of DNA damage. EMBO J *28*, 2971-2980.
Orphanides, G., Lagrange, T., and Reinberg, D. (1996). The general transcription factors of RNA polymerase II. Genes Dev *10*, 2657-2683.
Overmeer, R.M., Moser, J., Volker, M., Kool, H., Tomkinson, A.E., van Zeeland, A.A., Mullenders, L.H., and Fousteri, M. (2011). Replication protein A safeguards genome integrity by controlling NER incision events. J Cell Biol *192*, 401-415.
Pages, V., and Fuchs, R.P. (2002). How DNA lesions are turned into mutations within cells? Oncogene *21*, 8957-8966.
Park, E., Kim, H., Kim, J.M., Primack, B., Vidal-Cardenas, S., Xu, Y., Price, B.D., Mills, A.A., and D'Andrea, A.D. (2013a). FANCD2 activates transcription of TAp63 and suppresses tumorigenesis. Mol Cell *50*, 908-918.
Park, J.J., Irvine, R.A., Buchanan, G., Koh, S.S., Park, J.M., Tilley, W.D., Stallcup, M.R., Press, M.F., and Coetzee, G.A. (2000). Breast cancer susceptibility gene 1 (BRCAI) is a coactivator of the androgen receptor. Cancer Res *60*, 5946-5949.
Park, K.S., Cha, Y., Kim, C.H., Ahn, H.J., Kim, D., Ko, S., Kim, K.H., Chang, M.Y., Ko, J.H., Noh, Y.S.*, et al.* (2013b). Transcription elongation factor Tcea3 regulates the pluripotent differentiation potential of mouse embryonic stem cells via the Lefty1-Nodal-Smad2 pathway. Stem cells *31*, 282-292.
Parks, T.D., Leuther, K.K., Howard, E.D., Johnston, S.A., and Dougherty, W.G. (1994). Release of proteins and peptides from fusion proteins using a recombinant plant virus proteinase. Analytical biochemistry *216*, 413-417.
Perera, D., Poulos, R.C., Shah, A., Beck, D., Pimanda, J.E., and Wong, J.W. (2016). Differential DNA repair underlies mutation hotspots at active promoters in cancer genomes. Nature *532*, 259-263.
Perillo, B., Ombra, M.N., Bertoni, A., Cuozzo, C., Sacchetti, S., Sasso, A., Chiariotti, L., Malorni, A., Abbondanza, C., and Avvedimento, E.V. (2008). DNA oxidation as triggered by H3K9me2 demethylation drives estrogen-induced gene expression. Science *319*, 202-206.
Peters, J.M., Tedeschi, A., and Schmitz, J. (2008). The cohesin complex and its roles in chromosome biology. Genes & development *22*, 3089-3114.
Petryk, A., Kanakatti Shankar, R., Giri, N., Hollenberg, A.N., Rutter, M.M., Nathan, B., Lodish, M., Alter, B.P., Stratakis, C.A., and Rose, S.R. (2015). Endocrine disorders in Fanconi anemia: recommendations for screening and treatment. J Clin Endocrinol Metab *100*, 803-811.

Pham, C.T., MacIvor, D.M., Hug, B.A., Heusel, J.W., and Ley, T.J. (1996). Long-range disruption of gene expression by a selectable marker cassette. Proceedings of the National Academy of Sciences of the United States of America *93*, 13090-13095.
Phillips, J.E., and Corces, V.G. (2009). CTCF: master weaver of the genome. Cell *137*, 1194-1211.
Poss, Z.C., Ebmeier, C.C., and Taatjes, D.J. (2013). The Mediator complex and transcription regulation. Crit Rev Biochem Mol Biol *48*, 575-608.
Rappsilber, J., Ryder, U., Lamond, A.I., and Mann, M. (2002). Large-scale proteomic analysis of the human spliceosome. Genome research *12*, 1231-1245.
Reid-Bayliss, K.S., Arron, S.T., Loeb, L.A., Bezrookove, V., and Cleaver, J.E. (2016). Why Cockayne syndrome patients do not get cancer despite their DNA repair deficiency. Proc Natl Acad Sci U S A *113*, 10151-10156.
Reines, D., Chamberlin, M.J., and Kane, C.M. (1989). Transcription elongation factor SII (TFIIS) enables RNA polymerase II to elongate through a block to transcription in a human gene in vitro. J Biol Chem *264*, 10799-10809.
Riedl, T., Hanaoka, F., and Egly, J.M. (2003). The comings and goings of nucleotide excision repair factors on damaged DNA. EMBO J *22*, 5293-5303.
Sabarinathan, R., Mularoni, L., Deu-Pons, J., Gonzalez-Perez, A., and Lopez-Bigas, N. (2016). Nucleotide excision repair is impaired by binding of transcription factors to DNA. Nature *532*, 264-267.
Sands, A.T., Abuin, A., Sanchez, A., Conti, C.J., and Bradley, A. (1995). High susceptibility to ultraviolet-induced carcinogenesis in mice lacking XPC. Nature *377*, 162-165.
Saso, K., Ito, T., Natori, S., and Sekimizu, K. (2003). Identification of a novel tissue-specific transcriptional activator FESTA as a protein that interacts with the transcription elongation factor S-II. Journal of biochemistry *133*, 493-500.
Schaeffer, L., Moncollin, V., Roy, R., Staub, A., Mezzina, M., Sarasin, A., Weeda, G., Hoeijmakers, J.H., and Egly, J.M. (1994). The ERCC2/DNA repair protein is associated with the class II BTF2/TFIIH transcription factor. Embo J *13*, 2388-2392.
Schaeffer, L., Roy, R., Humbert, S., Moncollin, V., Vermeulen, W., Hoeijmakers, J.H., Chambon, P., and Egly, J.M. (1993). DNA repair helicase: a component of BTF2 (TFIIH) basic transcription factor. Science *260*, 58-63.
Scharer, O.D. (2013). Nucleotide excision repair in eukaryotes. Cold Spring Harb Perspect Biol *5*, a012609.
Schatz, D.G., and Swanson, P.C. (2011). V(D)J recombination: mechanisms of initiation. Annu Rev Genet *45*, 167-202.
Schwertman, P., Lagarou, A., Dekkers, D.H., Raams, A., van der Hoek, A.C., Laffeber, C., Hoeijmakers, J.H., Demmers, J.A., Fousteri, M., Vermeulen, W*., et al.* (2012). UV-sensitive syndrome protein UVSSA recruits USP7 to regulate transcription-coupled repair. Nat Genet.
Scrima, A., Konickova, R., Czyzewski, B.K., Kawasaki, Y., Jeffrey, P.D., Groisman, R., Nakatani, Y., Iwai, S., Pavletich, N.P., and Thoma, N.H. (2008). Structural basis of UV DNA-damage recognition by the DDB1-DDB2 complex. Cell *135*, 1213-1223.
Selby, C.P., Drapkin, R., Reinberg, D., and Sancar, A. (1997). RNA polymerase II stalled at a thymine dimer: footprint and effect on excision repair. Nucleic Acids Res *25*, 787-793.
Shao, W., and Zeitlinger, J. (2017). Paused RNA polymerase II inhibits new transcriptional initiation. Nat Genet *49*, 1045-1051.

Shema, E., Kim, J., Roeder, R.G., and Oren, M. (2011). RNF20 inhibits TFIIS-facilitated transcriptional elongation to suppress pro-oncogenic gene expression. Mol Cell *42*, 477-488.
Shih, Y.P., Wu, H.C., Hu, S.M., Wang, T.F., and Wang, A.H. (2005). Self-cleavage of fusion protein in vivo using TEV protease to yield native protein. Protein science : a publication of the Protein Society *14*, 936-941.
Shiomi, N., Kito, S., Oyama, M., Matsunaga, T., Harada, Y.N., Ikawa, M., Okabe, M., and Shiomi, T. (2004). Identification of the XPG region that causes the onset of Cockayne syndrome by using Xpg mutant mice generated by the cDNA-mediated knock-in method. Mol Cell Biol *24*, 3712-3719.
Shivji, K.K., Kenny, M.K., and Wood, R.D. (1992). Proliferating cell nuclear antigen is required for DNA excision repair. Cell *69*, 367-374.
Skourti-Stathaki, K., and Proudfoot, N.J. (2014). A double-edged sword: R loops as threats to genome integrity and powerful regulators of gene expression. Genes Dev *28*, 1384-1396.
Skourti-Stathaki, K., Proudfoot, N.J., and Gromak, N. (2011). Human senataxin resolves RNA/DNA hybrids formed at transcriptional pause sites to promote Xrn2-dependent termination. Molecular cell *42*, 794-805.
Sollier, J., Stork, C.T., Garcia-Rubio, M.L., Paulsen, R.D., Aguilera, A., and Cimprich, K.A. (2014). Transcription-coupled nucleotide excision repair factors promote R-loop-induced genome instability. Mol Cell *56*, 777-785.
Soutoglou, E., and Talianidis, I. (2002). Coordination of PIC assembly and chromatin remodeling during differentiation-induced gene activation. Science *295*, 1901-1904.
Soutourina, J. (2018). Transcription regulation by the Mediator complex. Nat Rev Mol Cell Biol *19*, 262-274.
Spivak, G. (2005). UV-sensitive syndrome. Mutat Res *577*, 162-169.
Stirling, P.C., Chan, Y.A., Minaker, S.W., Aristizabal, M.J., Barrett, I., Sipahimalani, P., Kobor, M.S., and Hieter, P. (2012). R-loop-mediated genome instability in mRNA cleavage and polyadenylation mutants. Genes Dev *26*, 163-175.
Sugasawa, K., Akagi, J., Nishi, R., Iwai, S., and Hanaoka, F. (2009). Two-step recognition of DNA damage for mammalian nucleotide excision repair: Directional binding of the XPC complex and DNA strand scanning. Mol Cell *36*, 642-653.
Sugasawa, K., Ng, J.M., Masutani, C., Iwai, S., van der Spek, P.J., Eker, A.P., Hanaoka, F., Bootsma, D., and Hoeijmakers, J.H. (1998). Xeroderma pigmentosum group C protein complex is the initiator of global genome nucleotide excision repair. Mol Cell *2*, 223-232.
Sugasawa, K., Okuda, Y., Saijo, M., Nishi, R., Matsuda, N., Chu, G., Mori, T., Iwai, S., Tanaka, K., Tanaka, K.*, et al.* (2005). UV-induced ubiquitylation of XPC protein mediated by UV-DDB-ubiquitin ligase complex. Cell *121*, 387-400.
Teive, H.A., Moro, A., Moscovich, M., Arruda, W.O., Munhoz, R.P., Raskin, S., and Ashizawa, T. (2015). Ataxia-telangiectasia - A historical review and a proposal for a new designation: ATM syndrome. J Neurol Sci *355*, 3-6.
Thakur, S., Dhiman, M., Tell, G., and Mantha, A.K. (2015). A review on protein-protein interaction network of APE1/Ref-1 and its associated biological functions. Cell Biochem Funct *33*, 101-112.
Theil, A.F., Mandemaker, I.K., van den Akker, E., Swagemakers, S.M.A., Raams, A., Wust, T., Marteijn, J.A., Giltay, J.C., Colombijn, R.M., Moog, U.*, et al.* (2017). Trichothiodystrophy causative TFIIEbeta mutation affects transcription in highly differentiated tissue. Hum Mol Genet.

Theron, T., Fousteri, M.I., Volker, M., Harries, L.W., Botta, E., Stefanini, M., Fujimoto, M., Andressoo, J.O., Mitchell, J., Jaspers, N.G.*, et al.* (2005). Transcription-associated breaks in xeroderma pigmentosum group D cells from patients with combined features of xeroderma pigmentosum and Cockayne syndrome. Mol Cell Biol *25*, 8368-8378.
Thomas, M.C., and Chiang, C.M. (2006). The general transcription machinery and general cofactors. Crit Rev Biochem Mol Biol *41*, 105-178.
Tian, M., Shinkura, R., Shinkura, N., and Alt, F.W. (2004). Growth retardation, early death, and DNA repair defects in mice deficient for the nucleotide excision repair enzyme XPF. Mol Cell Biol *24*, 1200-1205.
Tini, M., Benecke, A., Um, S.J., Torchia, J., Evans, R.M., and Chambon, P. (2002). Association of CBP/p300 acetylase and thymine DNA glycosylase links DNA repair and transcription. Mol Cell *9*, 265-277.
Tiwari, A.K., Roy, H.K., and Lynch, H.T. (2015). Lynch syndrome in the 21st century: clinical perspectives. QJM.
Tornaletti, S., and Hanawalt, P.C. (1999). Effect of DNA lesions on transcription elongation. Biochimie *81*, 139-146.
Tornaletti, S., Reines, D., and Hanawalt, P.C. (1999). Structural characterization of RNA polymerase II complexes arrested by a cyclobutane pyrimidine dimer in the transcribed strand of template DNA. J Biol Chem *274*, 24124-24130.
Troelstra, C., van Gool, A., de Wit, J., Vermeulen, W., Bootsma, D., and Hoeijmakers, J.H. (1992). ERCC6, a member of a subfamily of putative helicases, is involved in Cockayne's syndrome and preferential repair of active genes. Cell *71*, 939-953.
Tuduri, S., Crabbe, L., Conti, C., Tourriere, H., Holtgreve-Grez, H., Jauch, A., Pantesco, V., De Vos, J., Thomas, A., Theillet, C.*, et al.* (2009). Topoisomerase I suppresses genomic instability by preventing interference between replication and transcription. Nature cell biology *11*, 1315-1324.
Uhlen, M., Fagerberg, L., Hallstrom, B.M., Lindskog, C., Oksvold, P., Mardinoglu, A., Sivertsson, A., Kampf, C., Sjostedt, E., Asplund, A.*, et al.* (2015). Proteomics. Tissue-based map of the human proteome. Science *347*, 1260419.
Unal, E., Arbel-Eden, A., Sattler, U., Shroff, R., Lichten, M., Haber, J.E., and Koshland, D. (2004). DNA damage response pathway uses histone modification to assemble a double-strand break-specific cohesin domain. Molecular cell *16*, 991-1002.
van der Horst, G.T., van Steeg, H., Berg, R.J., van Gool, A.J., de Wit, J., Weeda, G., Morreau, H., Beems, R.B., van Kreijl, C.F., de Gruijl, F.R.*, et al.* (1997). Defective transcription-coupled repair in Cockayne syndrome B mice is associated with skin cancer predisposition. Cell *89*, 425-435.
van der Pluijm, I., Garinis, G.A., Brandt, R.M., Gorgels, T.G., Wijnhoven, S.W., Diderich, K.E., de Wit, J., Mitchell, J.R., van Oostrom, C., Beems, R.*, et al.* (2007). Impaired genome maintenance suppresses the growth hormone--insulin-like growth factor 1 axis in mice with Cockayne syndrome. PLoS Biol *5*, e2.
van der Pluijm, I., Garinis, G.A., Brandt, R.M., Gorgels, T.G.F., Wijnhoven, S.W., Diderich, K.E., de Wit, J., Mitchell, J.R., van Oostrom, C., and Beems, R. (2006). Impaired genome maintenance suppresses the growth hormone–insulin-like growth factor 1 axis in mice with Cockayne syndrome. PLoS biology *5*, e2.
van Steeg, H., de Vries, A., van Oostrom, C., van Benthem, J., Beems, R.B., and van Kreijl, C.F. (2001). DNA repair-deficient Xpa and Xpa/p53+/- knock-out mice: nature of the models. Toxicol Pathol *29 Suppl*, 109-116.

Verhage, R.A., Heyn, J., van de Putte, P., and Brouwer, J. (1997). Transcription elongation factor S-II is not required for transcription-coupled repair in yeast. Molecular & general genetics : MGG *254*, 284-290.
Vermeij, W.P., Dolle, M.E., Reiling, E., Jaarsma, D., Payan-Gomez, C., Bombardieri, C.R., Wu, H., Roks, A.J., Botter, S.M., van der Eerden, B.C.*, et al.* (2016). Restricted diet delays accelerated ageing and genomic stress in DNA-repair-deficient mice. Nature *537*, 427-431.
Wahba, L., Amon, J.D., Koshland, D., and Vuica-Ross, M. (2011). RNase H and multiple RNA biogenesis factors cooperate to prevent RNA:DNA hybrids from generating genome instability. Molecular cell *44*, 978-988.
Wallace, S.S., Murphy, D.L., and Sweasy, J.B. (2012). Base excision repair and cancer. Cancer Lett *327*, 73-89.
Wang, D., Bushnell, D.A., Huang, X., Westover, K.D., Levitt, M., and Kornberg, R.D. (2009). Structural basis of transcription: backtracked RNA polymerase II at 3.4 angstrom resolution. Science *324*, 1203-1206.
Wang, H., Zhai, L., Xu, J., Joo, H.Y., Jackson, S., Erdjument-Bromage, H., Tempst, P., Xiong, Y., and Zhang, Y. (2006). Histone H3 and H4 ubiquitylation by the CUL4-DDB-ROC1 ubiquitin ligase facilitates cellular response to DNA damage. Mol Cell *22*, 383-394.
Watrin, E., and Peters, J.M. (2009). The cohesin complex is required for the DNA damage-induced G2/M checkpoint in mammalian cells. The EMBO journal *28*, 2625-2635.
Wery, M., Shematorova, E., Van Driessche, B., Vandenhaute, J., Thuriaux, P., and Van Mullem, V. (2004). Members of the SAGA and Mediator complexes are partners of the transcription elongation factor TFIIS. EMBO J *23*, 4232-4242.
Wimberly, H., Shee, C., Thornton, P.C., Sivaramakrishnan, P., Rosenberg, S.M., and Hastings, P.J. (2013). R-loops and nicks initiate DNA breakage and genome instability in non-growing Escherichia coli. Nat Commun *4*, 2115.
Wolski, S.C., Kuper, J., Hanzelmann, P., Truglio, J.J., Croteau, D.L., Van Houten, B., and Kisker, C. (2008). Crystal structure of the FeS cluster-containing nucleotide excision repair helicase XPD. PLoS Biol *6*, e149.
Yan, W.X., Mirzazadeh, R., Garnerone, S., Scott, D., Schneider, M.W., Kallas, T., Custodio, J., Wernersson, E., Li, Y., Gao, L.*, et al.* (2017). BLISS is a versatile and quantitative method for genome-wide profiling of DNA double-strand breaks. Nat Commun *8*, 15058.
Yazdi, P.T., Wang, Y., Zhao, S., Patel, N., Lee, E.Y., and Qin, J. (2002). SMC1 is a downstream effector in the ATM/NBS1 branch of the human S-phase checkpoint. Genes & development *16*, 571-582.
You, Z., Feaver, W.J., and Friedberg, E.C. (1998). Yeast RNA polymerase II transcription in vitro is inhibited in the presence of nucleotide excision repair: complementation of inhibition by Holo-TFIIH and requirement for RAD26. Mol Cell Biol *18*, 2668-2676.
Yu, G., Wang, L.G., and He, Q.Y. (2015). ChIPseeker: an R/Bioconductor package for ChIP peak annotation, comparison and visualization. Bioinformatics *31*, 2382-2383.
Yu, K., Chedin, F., Hsieh, C.L., Wilson, T.E., and Lieber, M.R. (2003). R-loops at immunoglobulin class switch regions in the chromosomes of stimulated B cells. Nature immunology *4*, 442-451.
Yu, S., Evans, K., van Eijk, P., Bennett, M., Webster, R.M., Leadbitter, M., Teng, Y., Waters, R., Jackson, S.P., and Reed, S.H. (2016). Global genome nucleotide

excision repair is organized into domains that promote efficient DNA repair in chromatin. Genome Res *26*, 1376-1387.
Zambrowicz, B.P., Imamoto, A., Fiering, S., Herzenberg, L.A., Kerr, W.G., and Soriano, P. (1997). Disruption of overlapping transcripts in the ROSA beta geo 26 gene trap strain leads to widespread expression of beta-galactosidase in mouse embryos and hematopoietic cells. Proceedings of the National Academy of Sciences of the United States of America *94*, 3789-3794.
Zarebski, M., Wiernasz, E., and Dobrucki, J.W. (2009). Recruitment of heterochromatin protein 1 to DNA repair sites. Cytometry Part A : the journal of the International Society for Analytical Cytology *75*, 619-625.
Zhang, E.T., He, Y., Grob, P., Fong, Y.W., Nogales, E., and Tjian, R. (2015). Architecture of the human XPC DNA repair and stem cell coactivator complex. Proc Natl Acad Sci U S A *112*, 14817-14822.
Zhang, L., Zhang, Q., Jones, K., Patel, M., and Gong, F. (2009). The chromatin remodeling factor BRG1 stimulates nucleotide excision repair by facilitating recruitment of XPC to sites of DNA damage. Cell Cycle *8*, 3953-3959.
Zhang, X., Horibata, K., Saijo, M., Ishigami, C., Ukai, A., Kanno, S.I., Tahara, H., Neilan, E.G., Honma, M., Nohmi, T.*, et al.* (2012). Mutations in UVSSA cause UV-sensitive syndrome and destabilize ERCC6 in transcription-coupled DNA repair. Nat Genet.
Zheng, L., Annab, L.A., Afshari, C.A., Lee, W.H., and Boyer, T.G. (2001). BRCA1 mediates ligand-independent transcriptional repression of the estrogen receptor. Proc Natl Acad Sci U S A *98*, 9587-9592.
Zickler, D., and Kleckner, N. (2015). Recombination, Pairing, and Synapsis of Homologs during Meiosis. Cold Spring Harb Perspect Biol *7*.