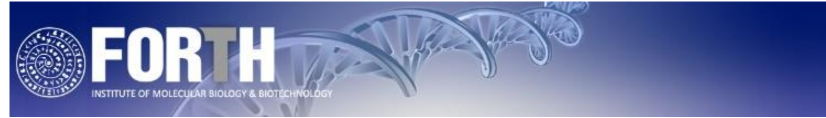


Master of Science thesis

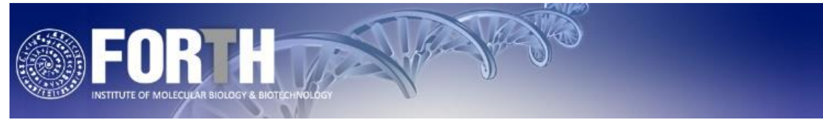
The functional role of XAB2 in Embryonic Stem Cells

ELEFTHERIA LEDAKI

Academic Supervisor: Professor George Garinis

Heraklion, Crete

November, 2016



ΜΕΤΑΠΤΥΧΙΑΚΗ ΔΙΑΤΡΙΒΗ

Ο λειτουργικός ρόλος της XAB2 στα εμβρυϊκά βλαστικά κύτταρα

ΕΛΕΥΘΕΡΙΑ ΛΕΔΑΚΗ

Ακαδημαϊκός Επιβλέπων: Καθηγητής Γεώργιος Γαρίνης

Ηράκλειο, Κρήτη

Νοέμβριος, 2016

Evaluation Committee

Professor George Garinis

Associate Professor Charalampos Spilianakis

Assistant Professor Sotirios Kampranis

Acknowledgements

I would like to thank Professor George Garinis for giving me the opportunity to conduct top-notch research in his laboratory, in a time point in my life were I was experiencing real difficulties. He provided me continuous scientific feedback, advice and guidance on my project as well as showed daily availability even under tight schedule conditions. He is truly a balanced and gifted group leader to whom footsteps many would want to follow. Moreover, I would like to express my gratitude towards the Ph.D candidate and friend, Maria Tsekrekou, who was constantly by my side, guiding me through the difficulties we have faced in our experimental assays and with which I have spent long hours discussing our project. Not once have I heard her complain or deny answering my questions. I truly admire her strength, positivity and knowledge.

In addition, I would like to thank Associate Professor Charalampos Spilianakis and Assistant Professor Sotirios Kampranis. They have mentored me and guided me throughout my Bachelor and MSc studies, advised me on every step of my scientific path, always available, kind and willing to help. Dr. Spilianakis always provided me with the courage to set my goals higher and Dr. Kampranis was always there to reassure me I could make everything possible.

I wish to thank my family who even though not being part of the research field in order to fully understand what my life is about, they have kept supporting my decisions. Especially, my sister Stella-Iro Ledaki who continuously made me feel confident for my scientific searches and reassured me of my future success in the decisions I chose to make. I could not forget my close friends, part of the research field, with whom I have had endless scientific and personal discussions, Dr. Eralda Salataj and MSc student, Dimitra Nikoleri.

Last but not least, the rest of the lab members, as well as the members of the Institute of Molecular Biology and Biotechnology, have provided a friendly, helpful and co-operative environment, which I truly appreciate.

CONTENTS

Acknowledgements 4
Abstract 7
ΠΕΡΙΛΗΨΗ 8
1. Introduction 9
    1. 1 Genome Maintenance and Repair 9
    1.1a Nucleotide Excision Repair (NER) pathway 10
    1.2 Pre-mRNA splicing 13
    1.2a What is pre-mRNA splicing? 13
    1.2b The Spliceosome 13
    1.2c Basic properties and Regulation 14
    1.2d The splicing reaction 16
    1.3 DNA damage affects pre-mRNA splicing 17
    1.3a Post-translational modifications 19
    1.3b Absence of splicing factors causes DNA damage through R-loop formation 20
    1.3c Recruitment or disassociation from sites of damage 21
    1.3d Altering of expression 22
    1.4 XAB2 22
    1.4a XAB2 is important for DNA damage repair and transcription 23
    1.4b XAB2 gene disruption leads to pre-implantation lethality 24
    1.4c XAB2 is involved in pre-mRNA splicing 24
    1.4d XAB2 preferentially binds RNA 25
    1.4e XAB2 involvement in the end resection step of homologous recombination 26
2. Aim of the Study 27
3. Materials and Methods 28
    3.1 Cell culture 28
    3.1.a JM8A3.N1 & bioXAB2;birA mESCs 28
    3.1.b Mouse Embryonic Fibroblasts (MEFs) 29
    Isolation of MEFs from mouse embryos 29
    3.2 Transfection assays 30

3.2a siRNA of XAB2 in different cell types 30
3.2b Overexpression of XAB2 in MEFs 32
3.3 RNA extraction 32
RT-qPCR 32
3.4 Western Blot 33
Protein extraction 33
SDS-PAGE 34
TRANSFER 35
Detection 35
3.5 Immunocytochemistry for Adherent Cells 36
3.6 Immunoprecipitation Assay 37
3.7 RNA-seq data analysis 38
4.1 Efficient knock-down of XAB2 in three different cell types 39
4.2 XAB2 knock-down reduces RNA-DNA hybrids 43
4.3 Overexpression of XAB2 coupled with siRNA rescues S9.6 and RNApolII phenotype .. 46
4.4 XAB2 does not cause damage or nuclear envelope disarrangement 48
4.5 RNA-seq analysis reveals probable XAB2 roles 49
4.6 Validation of XAB2 protein interactors 52
5. Discussion 54
6. References 57

Abstract

Nucleotide Excision Repair is an evolutionary conserved DNA repair pathway which safeguards the genome and ensures its faithful transmission to progeny. However, several proteins participating in NER have distinct roles further to DNA repair which could provide a direct link of NER defects to development and disease. Such a protein is XAB2 involved in transcription, two DNA repair pathways, TC-NER and HR and pre-mRNA splicing. XAB2 gene disruption leads to pre-implantation lethality but its function is yet to be elucidated. We therefore used siRNA assays to study its role and efficiently knocked-down XAB2 in JM8A3.N1, bio XAB2;birA mESCs and MEFs both in RNA and protein levels. Assessing the impact of XAB2 elimination in transcription, showed reduction of RNA-DNA hybrids in the nucleoli of JM8A3.N1 mESCs and MEFs coupled with decrease in RNApolII protein levels. Both phenotypes where rescued when overexpressing XAB2 in MEFs. Experiments conducted to validate its participation in NER found XAB2 not interacting with previously presented factors like RNA polII and XPA and its knock-down not causing damage studied by γH2Ax staining. RNA-seq data from siRNA XAB2 samples presented down-regulated genes targeted to the ribosomal group. These results suggest a probable role of XAB2 in transcription and splicing of ribosomal genes.

Keywords: XAB2, NER, pre-mRNA splicing, ribosome

ΠΕΡΙΛΗΨΗ

Το μονοπάτι επιδιόρθωσης DNA μέσω εκτομής νουκλεοτιδίων (Nucleotide Excision Repair pathway, NER) είναι ένα εξελικτικά συντηρημένο μονοπάτι το οποίο διαφυλάσσει το γονιδίωμα και διασφαλίζει την πιστή μεταβίβασή του στο θυγατρικό κύτταρο. Παρ' όλα αυτά, υπάρχει πληθώρα πρωτεϊνών που ενώ συμμετέχουν στο NER, διαθέτουν διακριτούς ρόλους και πέρα από την επιδιόρθωση του DNA, ρόλοι οι οποίοι μπορεί να συσχετίσουν άμεσα βλάβες στο NER με το ρόλο του μονοπατιού αυτού στην ανάπτυξη και την παθογένεια ασθενειών. Μία τέτοια πρωτεΐνη είναι η XAB2 η οποία εμπλέκεται στη μεταγραφή, δύο μονοπάτια επιδιόρθωσης DNA, το εξαρτώμενο από τη μεταγραφή NER (Transcription Coupled-NER, TC-NER) και αυτό του ομόλογου ανασυνδυασμού (Homologous Recombination, HR) καθώς και στο μάτισμα του προ-mRNA. Η διάσπαση του γονιδίου της XAB2 οδηγεί σε προ-εμφυτική θνησιμότητα αλλά ο ρόλος της παραμένει ακόμα αδιευκρίνιστος. Για αυτό χρησιμοποιήσαμε μεθόδους παρεμβολής RNA (small interfering RNA, siRNA) ώστε να μελετήσουμε το ρόλο της και καταφέραμε να σιγήσουμε επιτυχώς την XAB2 στα JM8A3.N1, bioXAB2;birA εμβρυϊκά βλαστικά κύτταρα ποντικού (Mouse Embryonic Stem Cells, mESCs) και στους εμβρυϊκούς ινοβλάστες ποντικού (Mouse Embryonic Fibroblasts, MEFs) τόσο στα επίπεδα RNA όσο και στα πρωτεϊνικά της επίπεδα. Η μελέτη της σίγησης της XAB2 στη μεταγραφή, έδειξε μείωση των υβριδίων RNA-DNA στους πυρηνίσκους των JM8A3.N1 εμβρυϊκών βλαστικών κυττάρων ποντικού καθώς και στους εμβρυϊκούς ινοβλάστες ποντικού ενώ υπήρχε και μία ταυτόχρονη μείωση των επιπέδων της RNA πολυμεράσης II. Και οι δύο φαινότυποι διασώθηκαν όταν πραγματοποιήθηκε υπερέκφραση της XAB2 στα MEFs. Πειράματα τα οποία διεξάχθηκαν για να αποδειχθεί η συμμετοχή της XAB2 στο NER, έδειξαν ότι δεν αλληλεπιδρά με πρωτεΐνες που είχε αποδειχθεί σε προηγούμενες ερευνητικές δημοσιεύσεις όπως η RNA πολυμεράση II και η XPA και περαιτέρω, η σίγησή της δεν προκαλεί γενετική βλάβη όπως προτείνεται μέσω ανοσοφθορισμού όπου χρησιμοποιήθηκε το αντίσωμα ενάντια στη γH2Ax. Δεδομένα αλληλούχισης RNA που προέκυψαν από siRNA δείγματα της XAB2 παρουσιάζουν υποεκφραζόμενα γονίδια τα οποία είναι στοχευμένα στη ριβοσωμική ομάδα γονιδίων. Κατά αυτόν τον τρόπο, αυτά τα αποτελέσματα προτάσσουν ένα πιθανό ρόλο της XAB2 στη μεταγραφή και το μάτισμα ριβοσωμικών γονιδίων.

Λέξεις Κλειδιά: XAB2, Μονοπάτι επιδιόρθωσης DNA μέσω εκτομής νουκλεοτιδίων (NER), μάτισμα του προ-mRNA, ριβόσωμα

1. Introduction

1. 1 Genome Maintenance and Repair

DNA is a nucleic acid essential for all forms of life **(Watson and Crick, 1953)**. It carries most of the genetic instructions required for growth, development, function and reproduction of all known living organisms and many viruses. This information must be passed faithfully to the daughter cell in each cell division. This is a challenge for the cell as genomic DNA is prone to lesions due to continuous threats. There are various kinds of threats such as products arising from normal cell metabolism, errors coming from DNA replication and environmental agents (i.e UV light, gamma rays or toxic chemicals) **(Harper and Elledge, 2007; Hoeijmakers, 2001).**

The cell needs to protect the genomic DNA from these threats and allow its faithful transmission to the daughter cell. For this purpose, the co-operative mode of action of several processes in cellular metabolism is required. DNA repair systems are responsible to identify and repair DNA lesions occurring throughout the genome **(Friedberg, 2003)** and several proteins work on ensuring that a single round of DNA replication will be completed with minimal errors in each cell cycle through cell cycle checkpoints **(Bell and Dutta, 2002)**. Moreover, in order for chromosomes to be properly segregated during mitosis, spindle must be assembled, kinetochore attached and chromosomes physically separated **(Nasmyth, 2001)**. All of these genome maintenance actions take place under controlled chromatin dynamics where chromatin modifications are strictly regulated **(Campos and Reinberg, 2009)** and coordinated by signalling pathways constituting the DNA damage response (DDR). DDR regulates transcriptional reprogramming, DNA replication, mitosis and DNA repair **(Harper and Elledge, 2007)**. In the case that the cell is severely damaged, it is eliminated by apoptosis or senescence. These processes constitute genome maintenance pathways by preserving genome integrity and preventing disease **(Bansbach and Cortez, 2012) (Figure 1.)**.

DNA repair pathways have the ability to repair different types of DNA lesions. Mismatch repair and base-excision repair, work on simple lesions (i.e single nucleotides) whereas nucleotide excision repair (NER), non-homologous end-joining (NHEJ) and homologous recombination (HR) repair more complex lesions (i.e helix distorting lesions and Double strand breaks). The repair of these lesions is of paramount importance. If they remain in place, they impede replication and

transcription, lead to mutations and chromosomal rearrangements which in turn cause disease.

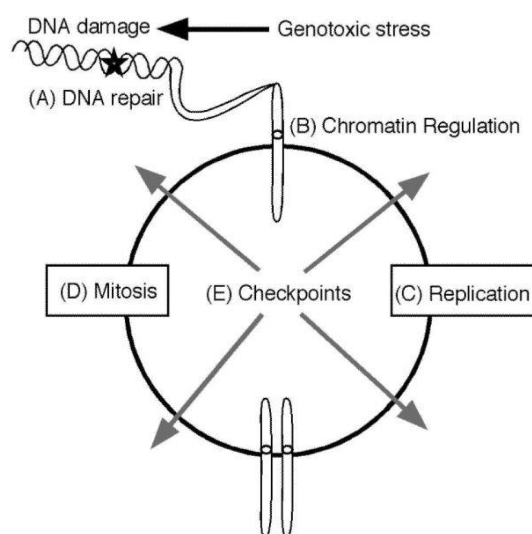


Figure 1. Several processes must be coordinated to achieve genome maintenance. **(A)** Operation of DNA repair mechanisms that remove DNA lesions caused by environmental agents or cellular byproducts. **(B)** DNA repair and metabolic processes take place in the context of chromatin. The modifications of chromatin make DNA accessible to proteins and regulate the function of signaling pathways as a response to DNA damage. **(C)** Chromatin structure and DNA must remain intact while duplicating in DNA replication, once, per division cycle. **(D)** Correct spindle assembly and chromosome segregation are vital during mitosis. **(E)** Cell cycle checkpoints oversee DNA damage, replication and mitosis. **(Adapted from Bansbach and Cortez, 2012)**

1.1a Nucleotide Excision Repair (NER) pathway

NER is an evolutionary conserved DNA repair pathway which ensures that the DNA remains functionally intact and is faithfully transmitted to the daughter cell. It recognizes and removes helix-distorting DNA lesions. These lesions have two main characteristics, thermodynamically destabilizing the DNA duplex and being bulky **(Hess et al., 1997)**. Specifically, the wide range of NER substrates include UV-induced photoproducts (cyclopyrimidine dimmers [CPDs], 6-4 photoproducts [6,4PPs]), adducts that may be formed by environmental mutagens (i.e benzo[α]pyrene, various aromatic amines, specific oxidative endogenous lesions (i.e cyclopurines) and the ones formed by chemotherapeutic drugs for cancer (i.e cisplatin) **(Friedberg et al., 2005; Gillet and Scharer, 2006)**.

The core NER reaction follows four basic steps, DNA damage site is recognized, dual incisions of the damaged DNA strand take place, the oligonucleotide bearing the lesion is removed and the undamaged DNA strand is used as template for a patch to be synthesized and ligated to the contiguous strand **(Figure 2.)**. NER can act through two subpathways: global genome NER or else called GG-NER and transcription-coupled

NER or else known as TC-NER (Hanawalt, 2002). GG-NER can remove lesions throughout the genome, whereas TC-NER does so only in the transcribed strand of active genes.

GG-NER scans and identifies lesions throughout the genome and is initiated by xeroderma pigmentosum complementation group C (XPC)-RAD23B-centrin, EFhand protein, 2 (CETN2) complex and the UV-damaged DNA-binding protein (UV-DDB) complex. These complexes associate with the thermodynamically destabilized duplex DNA **(Masutani et al., 1994; Nishi et al., 2005; Lagerwerf et al., 2011) (Figure 2a)**. RAD23B has the ability to stabilize XPC and helps its delivery to the sites of UV-damage. For that reason, as long as the complex has bound the non-damaged strand opposite the lesion, RAD23B dissociates from XPC and is not further involved in NER **(Bergink et al., 2012)**.

TC-NER identifies lesions at strands of genes being actively transcribed and is initiated by the blocked RNA polymerase II (RNAPII) as the first step of damage recognition **(Figure 2b) (Fousteri et al., 2006; Laine and Egly, 2006)**. Cockayne syndrome group B (CSB) or ERCC6, a transcription elongation factor with DNA-dependent ATPase activity, is recruited by the arrested elongation complex. It strongly binds the stalled RNAPII and recruits the NER factors, CSA complex and p300, a histone acetyltrasferase **(Hasan et al., 2001)**. CSA (ERCC8) participates in the DCX E3-ubiquitin ligase complex which includes DDB1, Culin 4A and ring-box 1 (ROC1/Rbx1) **(Groisman et al., 2003).**

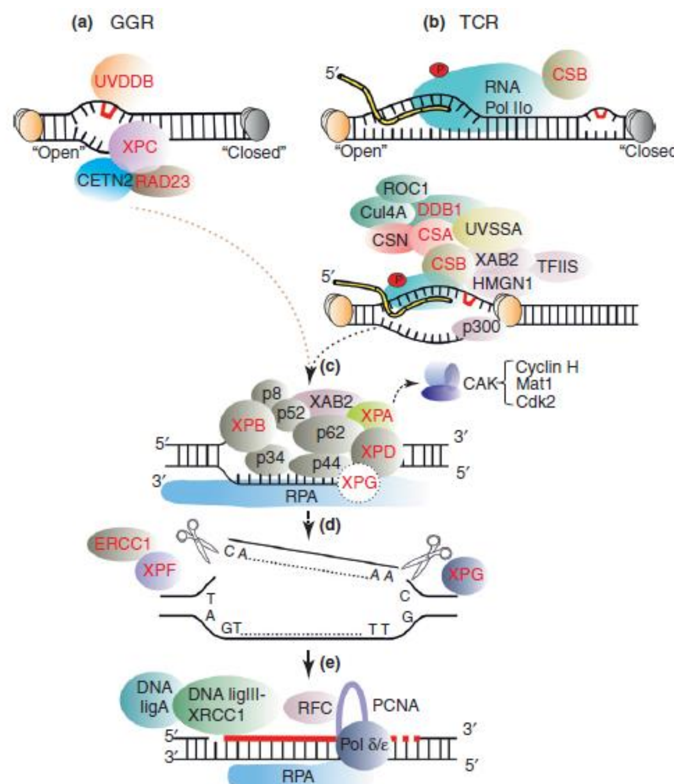


Figure 2. Nucleotide Excision Repair (NER) in mammals, consisting of two sub-pathways, Global Genome –NER (GG-NER) or GGR **(a)** and Transcription Coupled –NER (TC-NER) or TCR **(b)**. The two subpathways differ only at the bulky DNA lesion recognition part and merge into a common stabilization, excision, synthesis and ligation pathway, **(c)**, **(d)** and **(e)** respectively. **(Adapted from Kamileri et al., 2012)**

CSA is also crucial for the recruitment of XPA Binding Protein 2 (XAB2), HMGN1, a nucleosome remodeler and TFIIS at the sites of stalled RNAPII **(Fousteri et al., 2006; Nakatsu et al., 2000)**.

GG-NER and TC-NER both merge in a common way of action when transcription factor II H (TFIIH) is recruited at the site of the lesion. Specifically, TFIIH consists of 10/11 subunits (XPB, p62, p52, p44, p34, p8, XPD, Cdk-activating-kinase (CAK) complex and XPG) **(Figure 2c)**. TFIIH uses its XPB subunit which is an ATPase, to pry DNA open, creating a 27-nucleotide bubble **(Oksenych et al., 2009)**. Moreover, it allows XPD which is a helicase, to track along DNA until it stops at the damaged site to verify the bulkiness of the lesion **(Mathieu et al., 2013).** This blocking allows the formation of the preincision complex which includes XPA, replication protein A (RPA) and XPG. These proteins act synergistically and stabilize the damaged ssDNA. Further on, ERCC1-XPF nucleases are recruited by XPA to create the 5′ incision. Repair synthesis is initiated by the replicative DNA polymerases Polδ/κ/ε and 3′ insicion is made by XPG **(Figure 2d) (Fagbemi et al., 2011)**. Both endonucleases are activated by RPA that afterwards dissociates from DNA **(Overmeer et al., 2011)**. The process is completed by the end of the repair synthesis when sealing of the nick by DNA ligase IIIα/XRCC1 or DNA ligase I takes place **(Moser et al., 2007; Araujo et al., 2000; Mocquet at al., 2008) (Figure 3e)**.

Defects in NER genes can cause diverse disorders and three of them are the most known. Firstly, there are patients suffering from the genetic disorder xeroderma pigmentosum (XP), (XPA through XPG complementation groups). They are extremely sensitive to sunlight and present an over 1000-fold increased risk for skin cancer. This happens as they are not capable of repairing lesions induced by solar UV irradiation **(DiGiovanna and Kraemer, 2012)**. Moreover, there are the Cockayne syndrome and trichothiodystrophy (TTD) disorders, which are not associated with skin cancer predisposition but interestingly, patients bearing these diseases, present various developmental and neurological abnormalities **(Lehmann, 2003)**. The severity of the phenotypes can vary and even individuals with the same mutation can present different clinical outcomes. This diversity and severity of the diseases rising from defects in the same pathway, and the known paradoxical genotypes of NER mutated genes (mutations in the Xpc gene do not present any hotspots or founder effects but are spread throughout the gene sequence **(Chavanne et al., 2000)**), propose that the clinical heterogeneity of NER syndromes cannot be explained only by the DNA repair properties of the NER factors. For this reason, it can be assumed that proteins involved

in NER have multiple functions, linking NER with other processes such as pre-mRNA splicing.

1.2 Pre-mRNA splicing

DNA damage affects pre-mRNA splicing when the cell activates the DNA damage response (DDR) pathway. This pathway leads to a cascade of events which reconfigure gene expression at multiple levels. In this way, splicing control is changed to provide feedback for repair and cell decisions. For this reason, it is crucial to discuss in detail the pre-mRNA splicing properties, regulation and mechanisms, in order to understand the link between these two cell functions as well as the role of the proteins they have in common.

1.2a What is pre-mRNA splicing?

Precursor (pre)-mRNA splicing occurs in eukaryotic genes to remove non-coding sequences (introns) and ligate protein-coding sequences (exons) together. Introns disrupt exons **(Sharp, 1994; Sharp 2005)**. Pre-mRNA splicing can have constitutive and alternative splice sites depending on whether they are always (constitutive) or only sometimes (alternative) recognized as splice sites for the maturation of the pre-mRNA. The alternative splice sites can give rise to different mature mRNAs, resulting in different protein functions. Indeed, even though mouse and human genomes contain almost the same amount of genes, alternative pre-mRNA splicing takes place at >95 to 100% of human genes, whereas ~63% of mouse genes. This difference results in the expansion of the form and function of the human proteome **(Merkin et al., 2012; Barbarosa-Morais et al., 2012)**.

1.2b The Spliceosome

The pre-mRNA splicing reaction is catalyzed by the spliceosome which is a large, highly dynamic ribonucleoprotein (RNP) complex. There are two types of spliceosomes, the major and the minor spliceosome. The major spliceosome (termed U2 spliceosome) consists of five small nuclear ribonucleoproteins (snRNP) U1, U2, U4, U5 and U6 **(Wahl et al., 2009; Will and Lührmann, 2011)** and is responsible for the excision of more than 99% of human introns. The minor spliceosome (termed U12 spliceosome) consists of the U5 snRNP, as well as U11, U12, U4atac and U6atac snRNPs which act as the functional analogues of the corresponding snRNPs of the major spliceosome **(Turunen et al., 2013)**. Each snRNP consists of a short non-coding RNA, an Sm or Sm-like protein complex, which is fundamental for the formation of the mature snRNP complex and proteins that are specific for each snRNP **(Matera and Wang et al., 2014)**.

The fact that the catalytic center of the spliceosome is also composed of RNA, makes it profound that the spliceosome is actually a ribozyme, like the ribosome **(Fica et al., 2013)**.

1.2c Basic properties and Regulation

The splicing reaction, conducted by the spliceosome, requires fidelity and flexibility. The splicing sites must be discriminated between correct and incorrect ones. This is achieved through multistage proofreading of the sequences by different factors. Splicing regulatory factors act synergistically in order to follow the splicing commitment and be able to reinforce or repress splice site selection. Splicing itself may be catalysed by RNA but the proper recognition of splice sites requires RNA–RNA, RNA–protein and protein–protein interactions **(Fica et al., 2013)**.

The proper recognition of the splice sites relies on the main characteristics of an intron. These include four consensus elements: the exon/intron junctions at the 5' and 3' end of the intron, the 5' and 3' splice sites (SS), the branch point sequence (BPS) located upstream of the 3' SS and the polypyrimidine tract located between the BPS and the 3'SS as shown in **Figure 3.**

Moreover, there are additional *cis*-acting regulatory sequences which lay within introns and exons and are recognized by *trans*-acting splicing factors. These sites are capable of guiding splice site recognition by the core spliceosome. Specifically, a variety of splicing regulators, including hnRNP (heterogeneous nuclear ribonucleoproteins) and SR proteins (Serine (S)- Arginine (R)), have the ability to bind to exon or intron splicing enhancers (ESE or ISE, respectively) and to exon or intron splicing silencers (ESS or ISS, respectively) in order to control splice site recognition and utilization **(Singh and Valcarcel et al., 2005)**.

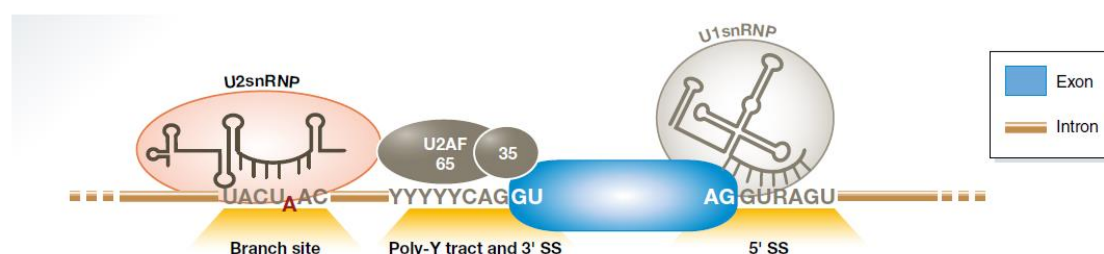


Figure 3. The consensus sequence elements that are required for the splicing complex recruitment. U1 snRNP recognized 5' SS by base-pairing interactions between the 5' end of U1 snRNA. The 3' SS is recognized by the binding of U2AF65/35 to the Poly-Y tract and 3' SS. These interactions lead to the U2 snRNP recruitment to the branch site where U2 snRNP interacts in a base-pairing way with the nucleotides flanking the branch point adenosine. **(Adapted from Daguenet et al., 2015)**

SR proteins are Serine/arginine-rich proteins. They constitute a family of splicing factors that bind pre-mRNA to activate splicing in a sequence dependent manner. HnRNPs on the other hand, have splicing factors in their protein family that mostly repress splicing **(Figure 4.)**. Both families constitute of proteins that are implicated in several cellular processes such as mRNA export, translation and RNA metabolic processes, respectively.

Mass spectrometry analyses have shown that the spliceosome associates with more than 170 proteins and exon recognition computational studies have proposed that hundreds sequence motifs can contribute to the regulation of splicing, suggesting that actually, hundreds of proteins may have splicing regulatory properties **(Jurika and Moore 2003; Barash et al., 2010)**.

However, one factor does not only have an activator or repressor role in splicing. On the contrary, *trans*-acting factors can act either way depending on the position of their binding site in correspondence to the regulated SS. Additionally, major role, in these factors action, plays the context of other cognate sites for other regulatory factors with which they can either co-operate or antagonize **(Fu and Ares et al., 2014)**. This creates a splicing code in which regulatory sequences, positional effects and *trans*-acting factors interactions co-exist and determine splicing fate.

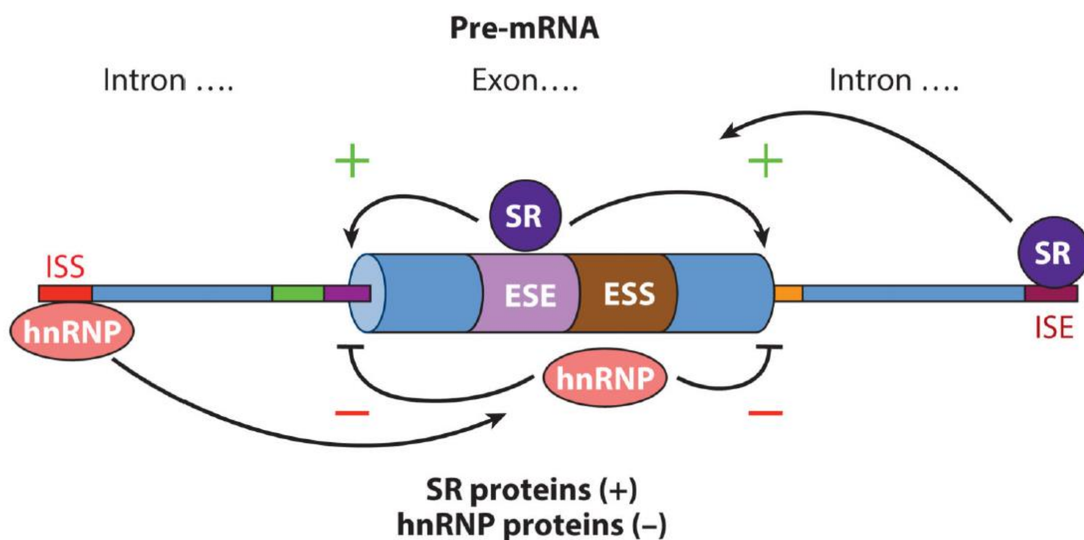


Figure 4. Pre-mRNA splicing can be positively (SR proteins) and negatively (hnRNP proteins) controlled**. (Adapted from Lee and Rio, 2015)**

1.2d The splicing reaction

In order to understand the role and functions of the different proteins needed in pre-mRNA splicing, it is of major importance to study the splicing reaction, conducted by the spliceosome, itself. The spliceosome assembly pathway follows specific steps. Firstly, **Complex E** is created. It consists of U1 snRNP which recognizes the 5' SS, the binding of SF1 splicing factor to the BPS region and the binding of a heterodimer consisting of U2 auxiliary factors (U2AFs) and the U2AF65/U2AF35 to the polypyrimidine tract and 3' SS **(Figure 5.)**. Afterwards, **complex A (pre-spliceosome)** is assembled as U2 snRNP is recruited to the BPS region in an ATP-dependent manner. The BPS adenosine is bulged out and SF1 is displaced. The **pre-catalytic spliceosome** forms when the preassembled U4-U6-U5 snRNP trimer is recruited to the pre-spliceosome and establishes **complex B**. U1 and U4 snRNP are released and after the enzymatic activation of the machinery the final **catalytic active spliceosome (Complex Bact/ B*)** is created. Complex B* and complex C are the ones responsible for the first and second trans-esterification reactions of the splicing reaction. Lastly, the lariat intron is excised and both exons are ligated **(Figure 5.). (Wahl et al., 2009; Will et al., 2011)**

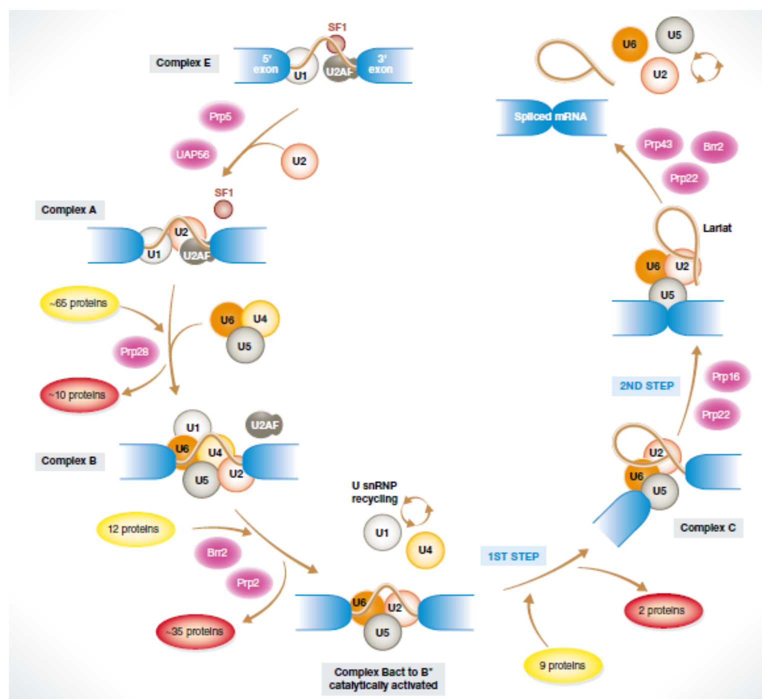


Figure 5. The main steps of pre-mRNA splicing are depicted. At first, complex E is created for the recognition of 5' SS, BPS, polypyrimidine tract and 3' AG. These interactions recruit U2 snRNP to BPS, leading to Complex A formation. Afterwards, the pre-spliceosome is converted to complex B by the addition of pre-assembled tri-snRNP. The machinery is then activated by a series of changes and becomes complex B (Bact, B*), which catalayzes the first trans-esterification reaction and complex C, which catalyzes the second. The lariat intron is then linearized and led to degradation. **(Adapted from Daguenet et al., 2015)**

Several pathologies (i.e myelodysplastic syndromes, cancer, neurodegenerative diseases) are linked with splicing defects. These defects are caused by mutated sequence elements, altered expression or function of splicing factors and effects resulting from nucleotide repeat expansions. Firstly, the mutations that can affect *cis*-acting elements (5'SS, 3'SS, Pol-Y tract, enhancers, silencer sequences) lead to constitutive exon skipping, retained introns, activation of cryptic splice sites and altered exon inclusion. These missplicing events have been shown to lead to a variety of diseases such as spinal muscular atrophy **(Naryshkin et al., 2014)**, cystic fibrosis **(Pagani et al., 2005)** and cancer **(Gnad et al., 2013; Supek et al., 2014)**. Secondly, there have also been found mutations in *trans*-acting factors such as core splicing machinery components, snRNP biogenesis proteins and splicing regulators. The defects in those genes can cause widespread splicing modifications. These modifications are hallmarks of cancer and tissue-specific alterations and lead to, amongst other diseases, cancer, Alzheimer's disease **(Bai et al., 2013)** and Huntington's disease **(MacDonald et al., 1993)**. Thirdly, the nucleotide repeat expansions can cause both loss of function and gain of function mutations. These affect the sequestration and stabilization of splicing factors as well as their splice site recognition. When the stability of the splicing factors changes and the SS is altered, widespread splicing modifications can take place and RNA toxicity mechanisms are activated. This activation is due to the aberrant splicing factors distribution and the production of short spliced RNA isoforms with pathogenic (toxic) properties. These missplicing events could cause a large spectrum of diseases like myotonic dystrophies **(Liquori et al., 2001)** and Fragile X-associated tremor/ataxia syndrome **(Cooper et al., 2009)**. Lastly, it should be mentioned that available mouse models bearing deletions of several pre-mRNA splicing associated factors, present lethal phenotypes, such as BRCA1 **(Savage et al., 2014)**, PRP19 **(Farmer et al., 2005)**, TRA2B **(Storbeck et al., 2014)**, SFRS10 **(Mende et al., 2010)**, Aquarius and XAB2 **(Yonemasu et al., 2005)**.

1.3 DNA damage affects pre-mRNA splicing

In order to repair DNA lesions, a cell activates the DNA damage response (DDR) pathway to help the cell recover from damage or lead it to apoptosis if the damage is excessive. Except for following the distinct repair mechanisms that were previously mentioned, DDR should first sense the DNA lesion's existence and for example block transcription and replication. In this way, DNA damage can cause phosphorylation of the histone variant H2AX, modify canonical histones and recruit poly (ADP-ribose) enzymes.

DDR recruits several other proteins such as ATM and DNA-PK kinases in the case of double-strand breaks (DSBs) whereas single strand breaks (SSBs) recruit the ATR kinase and the signaling cascade goes on with CHK1 and CHK2 **(Ciccia and Elledge,**

2010). All these actions from DDR lead to cell cycle arrest, DNA repair and apoptosis in case of excessive damage. The resulting cascade of events reconfigures gene expression at multiple levels. For this to be achieved, splicing control is changed through the modification, recruitment, relocalization and altering of the expression of splicing and alternative splicing factors which provide feedback for repair and cell decisions **(Figure 6.)**.

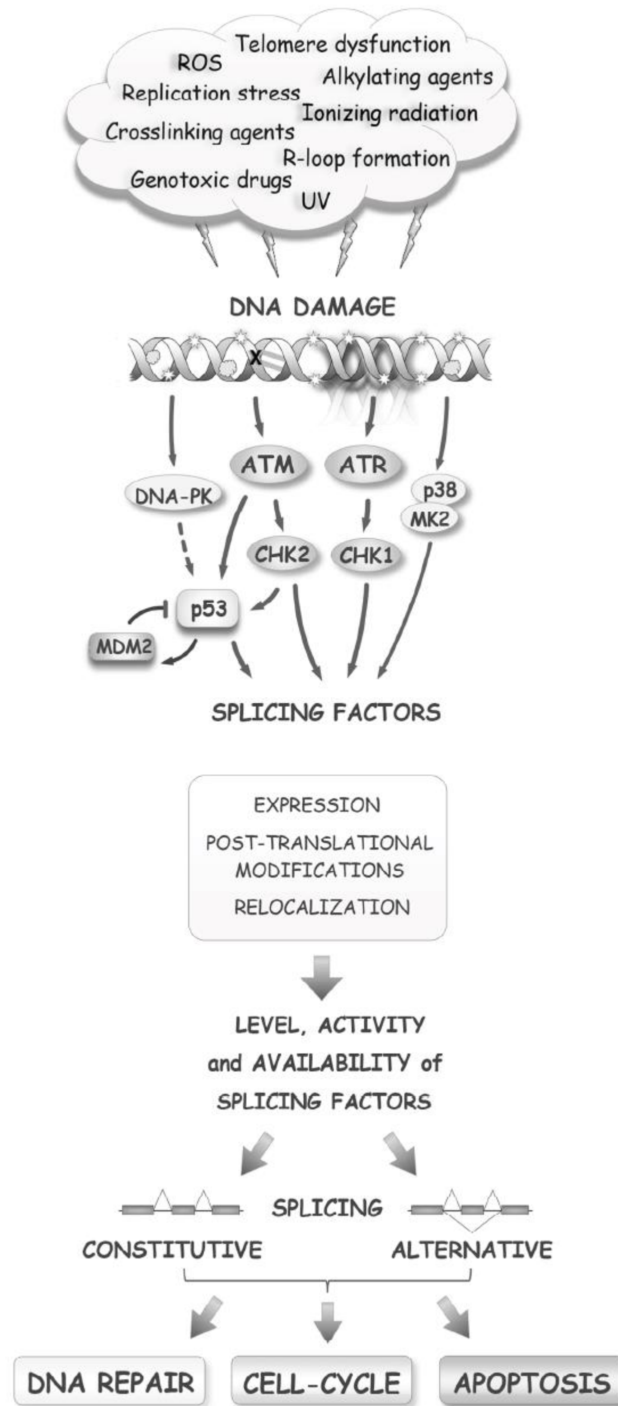


Figure 6. DNA damage affects pre-mRNA splicing in order to provide feedback for DDR. **(Adapted from Shkreta and Chabot, 2015)**

1.3a Post-translational modifications

As mentioned earlier, DDR recruits splicing and alternative splicing factors. Splicing proteins and RNP complexes are known to bind nascent transcripts to protect genome integrity by preventing RNA/DNA duplexes but except for this distinct function, they can also be recruited or removed from sites of damage **(Shkreta and Chabot, 2015**). This recruitment aims in their post-translational modification which regulates their localization and activity. In this way it would be interesting to refer to the alterations on pre-mRNA splicing factors due to DDR events.

Firstly, as an initial step of detecting DNA lesions, DNA damage modifies splicing factors at the post-translational level, in the same way that proteins of the DDR are modified themselves in response to damage stimuli **(Polo and Jackson et al., 2011)**. These changes affect the localization, stability and activity of the proteins. For example, hnRNP proteins are PARylated which inhibits their RNA binding and modifies RNA splicing **(Ji and Tulin et al., 2009)**. Moreover, acetylation has been associated with altering the activity of splicing factors. Specifically, BCLAF1, a BRCA1 interactor that recruits spliceosome components after damage stimuli, has been found rapidly deacetylated after the treatment of U2OS cells with IR **(Bennetzen et al., 2013)**. Furthermore, phosphorylation which is an immediate link between DNA lesion sensing and DDR due to the signaling cascades being activated, seems to affect several splicing factors. For example, PRP19, which is a splicing factor in yeast and has been implicated in transcription elongation when recruiting TREX components in mammals, has been found phosphorylated at ATM/ATR consensus sites in proteomic studies of yeast treated with IR **(Smolka et al., 2007; Chanarat et al., 2012)**. However, these post-translational modifications can also occur on components of the transcription machinery that change the rate or pausing of transcription in order to impact splice selection. A characteristic example is the phosphorylation of RNAPII upon UV which is altered so as to affect transcription elongation and splicing **(Munoz et al., 2009)**. In general, simultaneously with the post-translational modifications (i.e phosphorylation of kinases) of DDR genes, RNA-splicing factors are modified to feedback on these genes themselves (i.e signaling genes). As a result, the required response is amplified or the cell is recovered from damage.

1.3b Absence of splicing factors causes DNA damage through R-loop formation

It has already been mentioned that the DDR considers essential the role of splicing factors to reconfigure gene expression. The same applies the other way around as the absence of splicing factors itself has the ability to cause DNA damage through R-loop formation.

RNA-DNA hybrids can be formed during DNA replication in a 11bp hybrid of an Okazaki fragment, at transcription in a 8bp hybrid found within the RNA polymerase active site and in longer tracts, the so called R-loops. Specifically, R-loops can form whenever an RNA molecule can anneal with its template DNA strand, creating a displaced sense ssDNA. This is based on the thread back mechanism which happens co-transcriptionally (in *cis*). This mechanism proposes that DNA which is behind RNA polymerase and is negatively supercoiled **(Westover et al., 2004)**, has the ability to unwind which may allow the annealing of the template strand with nascent RNA. Moreover, it can also form post-transcriptionally (in *trans*) as a transcribed RNA from one locus, forms an R-loop with a homologous DNA sequence from another locus **(Wahba et al., 2013)**.

Even though R-loops form naturally during transcription, a mechanism evolutionarily conserved, their persistence may threaten genome integrity. So far, five mechanisms have been documented to regulate R-loop formation. Firstly, RNase H1 and RNase H2 act towards the dissolution of R-loops by degrading the RNA in hybrids **(Shaban et al., 2010)**. Secondly, RNA/DNA helicases like Rho, DHX9 and Senataxin are able to unwind the RNA from the R-loop **(Richardson, 2003; Chakraborty and Grosse, 2011; Mischo et al., 2011)**. Thirdly, topoisomerases suppress R-loop formation. They have the ability to resolve the negative torsional stress behind RNA polymerase II in order to prevent the annealing of the nascent RNA with the DNA template **(Tuduri et al., 2009; Sordet et al., 2009)**. Moreover, suppressors of proteins that promote R-loop formation (like Rad51 and Rad52) also regulate R-loop formation. Specifically, Rad51, promotes strand exchange between homologous sequences during HR. Rad52 has strand-annealing properties and can stimulate Rad51 activity. Both of them have been shown to favor hybridization of the RNA to the template DNA during *trans* R-loop formation **(Wahba et al., 2013)**. Lastly, defects in RNA processing factors, which are involved in coating, splicing, exporting and degrading the nascent RNA, elevate the level of R-loops. In the case that any of these five mechanisms is absent, persistent R-loops, that endanger genome stability and cause DNA damage, are created.

The absence of RNA processing causes R-loop formation because basic processes are deficient such as pre-mRNA splicing, in the case of splicing factors absence. The fact

that the depletion of splicing factors causes DNA damage was showed in a specific genome-wide siRNA screen of several splicing factors, including XAB2, in which γH2AX phosphorylation was increased **(Paulsen et al., 2009)**. Specifically, if a nascent RNA sequence is not bound by an RNA binding protein (RBP), R-loops are formed **(Wan et al., 2015)**. This happens because RNA hybridizes to the template strand of the melted DNA. This structure blocks replication, causes fork collapse **(Gan et al., 2011)**, obscures transcription, creates mutations and hyper-recombination having serious impacts on genome stability and leading to DDR activation **(Li and Manley et al., 2006; Aguilera and Garcia-Muse, 2012)**. Further on cell cycle checkpoint is activated and DNA damage and chromosome rearrangements take place. This is due to the fact that the ssDNA is more unstable and prone to transcription associated mutagenesis, recombination (TAR) and DSBs **(Li and Manley, 2005)**. The latter happens because R-loops are cleaved by structure specific endonucleases to create DSBs **(Lin and Pasero et al., 2014)**. This process requires the NER endonucleases XPF and XPG as well as the TCR factor CSB **(Sollier et al., 2014)**. In this way, R loops have been associated with a large number of cancers and human diseases in general.

R-loops can also act positively. Several examples can prove this function. The accumulative R-loop environment leads to genomic rearrangements required for the change of class of antibody. Moreover, R-loops regulate the expression of lncRNA COOLAIR in *Arabidopsis* and protect the promoter of a subset of transcriptionally active, unmethylated CpG islands, from the de novo methyltransferase DNMT3B1 **(Roy et al., 2008; Sun et al., 2013; Ginno et al., 2012)**. The biggest challenge in the field of R-loops would be to explore the mechanisms which could balance the “bright” with the “dark” side of these RNA-DNA hybrids.

1.3c Recruitment or disassociation from sites of damage

DNA damage can either exclude or recruit splicing factors to the site of damage. On the one hand, exclusion may take place at sites of damage because the cell pauses transcription in order to provide access to the repair machinery and decrease the association of splicing factors to limit mRNA processing **(Shanbhag et al., 2010)**. It has been shown that UV-induced transcription blocking DNA lesions decrease U2 and U5/U6 snRNPs localization at irradiated sites proposing exclusion of late-stage spliceosomes **(Tresini et al., 2015)**. This spliceosome displacement has been proposed to lead to R-loop formation which activates ATM and modulates global alternative splicing. On the other hand, even though it has been proposed that splicing components can be recruited to sites of damage, they may not participate in the repair process but coordinate repair with splicing decisions. Sites of damage themselves are poorly defined only by γH2AX stainings and it is not sure which proteins exist so close in the DNA lesion **(Lukas et al., 2011)**. However, there are multifactorial proteins that

play important and direct roles in DNA repair, such as hnRNP C (moves to sites of damage after IR), PRP19 (essential for DNA repair in yeast and activates ATR through RPA interactions) and the other way around BRCA1 that functions in DDR but plays roles in splicing as well **(Anatha et al., 2013; Zhang et al., 2005; Savage et al., 2015)**.

1.3d Altering of expression

DNA damage also alters the expression of splicing factors. This happens in order to modulate splicing processes towards needed mRNA products for translation. These proteins are involved in DNA repair, cell cycle control and apoptotic signaling as well as changing transcription-coupled splicing decisions such as speed of elongation **(Ip et al., 2011; Munoz at al., 2009)**.

1.4 XAB2

XAB2 (XPA Binding Protein 2) is a protein involved in both DNA repair (NER pathway) and pre-mRNA splicing mechanisms. It consists of 855 amino acids with a molecular mass of approximately 100kDa. NCI BLAST homology search reveals that it is highly conserved among lower and higher eukayotes. Specifically, mouse XAB2 shares 99% identity with its human homologue, 82% with that of zebrafish (Danio rerio), 61% with the fruitfly (Drosophila melanogaster), 49% with Caenorhabditis elegans and 28% with yeast (Saccharomyces cerevisiae). It is a predicted acidic protein containing three stretches of acidic residues and belongs to the tetratricopeptide repeat (TPR) superfamily. It comprises of 4 TPR domains **(www.ebi.ac.uk)** throughout the protein molecule, mediating protein interactions, or 11 HAT (Half a TPR) which mediate protein-RNA interactions **(Nakatsu et al., 2000)**. Specifically, TPR structural motifs are 34 amino acid scaffolds which create alpha-helix pair repeats that fold together to produce a single linear solenoid domain, the TPR **(Lamb et al., 1995)**. TPRs have been implicated in cycle control, transcriptional regulation, RNA processing, mitochondrial and peroxisomal protein transport.

In humans the XAB2 gene is located at chromosome 19 (19p13.2) whereas in mice at chromosome 8 (8A1.1) **(www.genecards.org)**. The mouse XAB2 gene produces five transcripts. Two of them are protein coding (855aa, 285aa) while the remaining three do not produce any protein **(www.ensembl.org)**.

1.4a XAB2 is important for DNA damage repair and transcription

XAB2 was initially described by **Nakatsu et al., 2000** where it was found enrolled in TC-NER as it interacts with XPA, in yeast two-hybrid system assays (XPA Binding Protein 2) and CSA, CSB in *in vitro* translation assays and *in vivo* immunoprecipitation of whole cell extracts of SV-40 transformed CSA/B fibroblast lines. It was also shown to interact with RNA-polII through co-Immunoprecipitation experiments in HeLa whole cell extracts. Specifically, **Kuraoka et al., 2008** showed an interaction of the XAB2 complex with the hyperphosphorylated form of RNApolIIo (elongation) which was enhanced upon DNA damage similar to the XPA interaction.

In addition, more evidence suggests the role of XAB2 in TC-NER and normal transcription **(Nakatsu et al., 2000)**. Microinjection of anti-XAB2 antisera in fibroblasts of XPC patients, who lack functional GG-NER and would produce unscheduled DNA synthesis (UDS) only from TC-NER, reduced the UDS in these cells compared to normal human cells that derive UDS mainly from GG-NER. Moreover, the microinjection inhibited the normal RNA synthesis (RS) which is attributed to transcription and the recovery of RNA synthesis after UV irradiation (RRS) in normal human cells, where a failure of RRS is known to be ascribed to defective TC-NER **(Nakatsu et al., 2000)**.

The role of XAB2 in TCR and transcription was also shown by colony forming assays in HeLa cells knocked-down for XAB2. Specifically, these cells were hypersensitive to UV light, suggesting a defect in NER and the RS as well as RRS were decreased **(Yonemasu et al., 2008)**.

Lastly, XAB2 has been found to regulate gene expression by associating with retinoic acid receptor α (RARα) and histone deacetylase 3 (HDAC3). This complex acts as a corepressor which inhibits gene activation by RARα. Binding of retinoic acid (ATRA) to its receptor leads to the dissociation of XAB2 and HDAC3 allowing transcription of RARα associated genes **(Ohnuma-Ishikawa et al., 2007)**.

1.4b XAB2 gene disruption leads to pre-implantation lethality

One of the most intriguing functions of XAB2 is the one in mouse embryogenesis. **Yonemasu et al., 2005** worked on elucidating the *in vivo* function of XAB2 through the introduction of two types of mutations into the *XAB2* gene in mice. One is the deletion of the region encompassing the promoter and exons 1-4, *XAB2*(-/-) and the other one is the deletion of C-terminal 162 amino acids, *XAB2*($-^{cd}/-^{cd}$). The group observed both types of heterozygous mice to appear normal as far as physiology and behavior is concerned, but surprisingly, homozygous mutants were selectively absent in the newborn mice. Having analyzed the embryos at different stages of development, they discovered that *XAB2*-homozygous mutants could only survive until the morula stage and did not have the ability to develop to the blastocyst stage **(Figure 7)**.

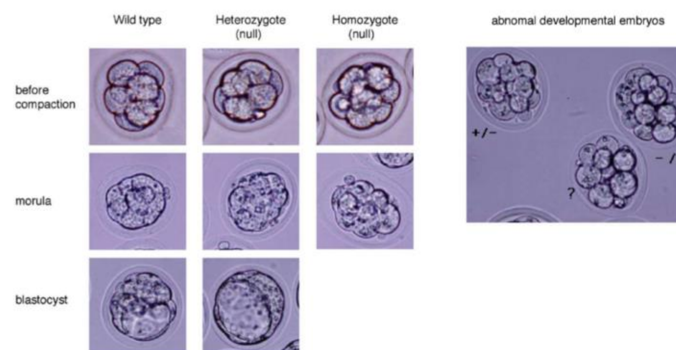


Figure 7. The embryos that have been collected from the *XAB2*(+/-) intercrosses were examined at 3.5 d.p.c and were separated into normal (left panel) and abnormal (right panel) developmental embryos. The abnormal developmental embryos did not manage to reach compaction which normally occurs at the eight-cell stage of development **(Adapted from Yonemasu et al., 2005)**.

1.4c XAB2 is involved in pre-mRNA splicing

XAB2 has been found to be part of a multimeric protein complex from immunoprecipitation assays of HeLa nuclear extracts which stably express FLAG-XAB2 **(Kuraoka et al., 2008)**. Using mass spectrometry analysis, it was shown that the protein complex which was isolated had six subunits, hAquarius (IBP160), XAB2 (hSYF1), hPRP19, CCDC16, hISY1 and peptidyl-prolyl *cis-trans* isomerase E (PPIE). All these proteins have been implicated in pre-mRNA splicing. The fact that XAB2 consists of 4 TPR motifs which are responsible for both inter and intra-molecular protein interactions means that it could simultaneously interact with many proteins. The deletions of various lengths in the FLAG-XAB2 protein in HeLa cells, suggested that the N-terminal regions of 1-469, 1-699, 1-734 are required for the binding of CCDC16, hAquarius and hPRP19, respectively and that the N-terminal region of 1-295 is responsible for hISY1 and PPIE interactions **(Kuraoka et al., 2008)**. Moreover, knock-

down of XAB2 in HeLa cells resulted in the decrease of hAquarius and hISY1 protein levels suggesting that XAB2 is crucial for the expression or stability of the affected subunits **(Kuraoka et al., 2008)**. In order to check whether pre-mRNA splicing is affected in the knock-down cells, Bcl-x mRNA was studied and difference in the expression of the alternative spliced products was detected. XAB2 has also been found to interact with human Debranching enzyme 1 (hDbr1), part of the Intron Large (IL) complex, which acts as an intermediate complex of the intron degradation pathway. This interaction was suggested through immunoprecipitation experiments from total cell lysates of HEK293T with Flag-tagged proteins even though mass spectrometry experiments from the same group did not detect such an interaction **(Masaki et al., 2015)**.

Furthermore, **Hofmann et al., 2010** propose that XAB2 is part of Complex C in the pre-mRNA splicing reaction, associated with the catalysis of the second transesterification reaction. An RNAi screen of pre-mRNA splicing factors essential for mitosis showed that XAB2 knocked-down cells exhibit mitotic delay, M alignment delay and m-phase delay which ultimately leaded to cell death **(Neumann et al., 2006)**. This agrees with the lethal phenotype of XAB2 expressed by **Yonemasu et al., 2005**.

The hPRP19/CDC5L complex (both interact with XAB2) with ~30 more other proteins is part of the Nineteen-containing complex (NTC) and is recruited by the spliceosome to help it maintain its catalytically active RNA network **(Grote et al., 2010)**. It is proposed that NTC is present at the human Complex C but is less tightly associated with the core complex. It has also been found in the 35S-supercomplex, a proposed dissociation product of the splicing machinery that includes U5snRNP, NTC and other accessory proteins. However, given its great stability, it is highly likely that it is recruited earlier, from the stage of step 1 splicing reaction (Complex B) **(Makarova et al., 2004)**.

1.4d XAB2 preferentially binds RNA

Lastly, gel mobility shift assays have proposed that XAB2 has the ability to preferentially bind RNA and not DNA **(Kuraoka et al., 2008)**.

1.4e XAB2 involvement in the end resection step of homologous recombination

XAB2 has also been found to participate in the homologous recombination repair pathway (HR) for the repair of DSBs. Specifically, **Onyango et al., 2016** show that XAB2 is involved in the end resection step that generates the 3’ ssDNA intermediate in the HR **(Figure 8)**. This role emerged from Camptothesin (Cpt) treatment, which induced damage, in U2OS cells. XAB2 forms a complex with ISY1 and PRP19 which localizes to interchromatin granules having adjacent localization with γH2AX. It should be noted that their siXAB2 experiments showed no change in γH2AX foci which contrasts the siRNA study from **Miller et al., 2009**. Lastly, their mass spectrometry analysis upon Cpt treatment was not able to detect any NER interactors, BRCA1 or CtIP with which XAB2 is suggested to associate functionally.

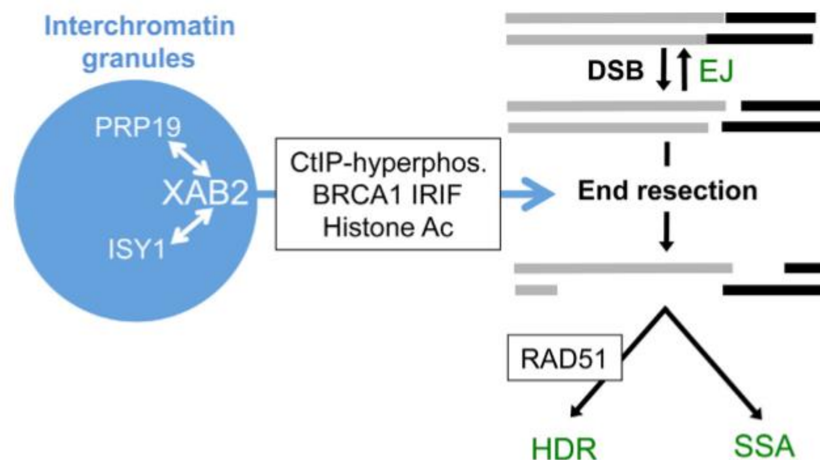


Figure 8. XAB2 role in the end resection step of HR. XAB2 interacts with PRP19 and ISY1 and they are localized in interchromatin granules, separate from DSBs. XAB2 promotes a series of DDR responses that are associated with the end resection step of DSBs which are required for both HDR and SSA but not EJ. **(Adapted from Onyango et al., 2016)**

2. Aim of the Study

The fact that the disruption of the XAB2 gene leads to pre-implantation lethality, makes XAB2 a fascinating protein to study. XAB2 has been shown to participate in several processes, among which are the repair of DNA lesions (NER and HR pathways), pre-mRNA splicing and transcription. However, its direct role and functions *in vivo* are yet to be elucidated. To decipher the role of XAB2, we used RNAi combined with immunostaining, immunoblotting and RNA-seq assays. Immunoprecipitation assays were also conducted to verify previously shown XAB2 interactors in our system and shed light in the complexes it participates.

3. Materials and Methods

3.1 Cell culture

3.1.a JM8A3.N1 & bioXAB2;birA mESCs

JM8A3.N1 is a highly germline competent C57BL/6N mouse embryonic stem cell line. The dominant agouti coat color gene was restored in these cells by targeted repair of the C57BL/6 nonagouti mutation **(Pettitt et al., 2009)**. JM8A3.N1 mESCs were cultured on 0.2% gelatin coated tissue culture vessels, in DMEM (Gibco) supplemented with 15% FBS, 0.5% Penicillin –Streptomycin, 1% L-glutamine, 1% Non-essential Amino Acids, 4.8% LIF (supernatant from CHO cells expressing human recombinant LIF), CHIR (3μM), PD (1μM) and 0.001% β-merkaptoEtOH. Medium was renewed daily and cells were split every two to three days. Plates were incubated at a humidified 37^0C, 5% $C0_2$ incubator.

We have also used mESCs isolated from bioXAB2$^+$/$^+$;birA mice generated in our lab. They bear an Avidin-TEV-Flag (ATF) tag that comprises an Avidin and a Flag affinity moiety linked by a cleavage site which is recognized by the protease TEV. The Avidin tag is recognized and biotinylated by the bacterial BirA ligase creating a very high affinity module in order to isolate the interactors of XAB2. The TEV site allows the elution of the isolated complexes from the streptavidin beads whereas the Flag tag is used for the second round of purification. The targeting vector can be observed at **Figure 9.** BioXAB2;birA mESCs were isolated using standard protocols **(Czechanski et al., 2014)** and were cultured in the same way as JM8A3.N1 cells.

As far as the generation of ATF-tagged XAB2 knock-in mice is concerned, targeted ESCs with the desired vector were injected in blastosysts which were then transferred in pseudopregnant foster mothers. The embryos generated chimera mice which were crossed with C57BL/6 mice. The ATF-tagged knock-in XAB2 offsprings were crossed with BirA transgene mice to produce Biotinylated XAB2 mice.

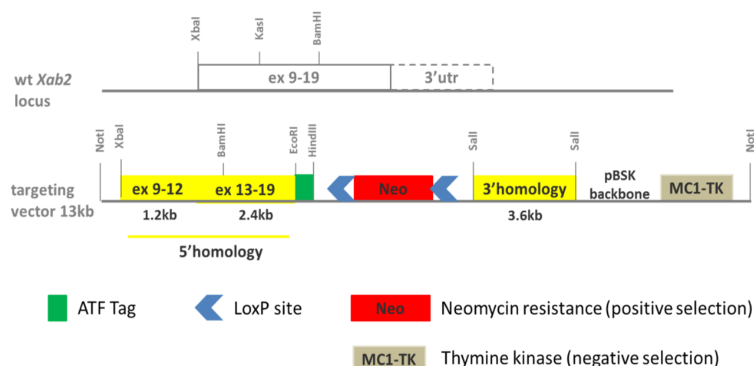


Figure 9. The targeting vector has an Avidin-TEV-Flag (ATF) at the 3' prime of XAB2 gene. **(Adapted and manufactured from Ph.D candidate M. Tsekrekou)**

3.1.b Mouse Embryonic Fibroblasts (MEFs)

Isolation of MEFs from mouse embryos

C57BL/6 MEFs were isolated by standard protocols **(Xu, 2005)**. In brief, MEFs were isolated from C57BL/6 pregnant females at day 13.5 of gestation. The mouse was sacrificed by cervical dislocation, sprayed with 70% Ethanol and subsequnently dissected inside the tissue culture hood. The uterus was dissected out using blunt forceps and sterile dissection scissors, (caution was taken not to damage internal organs or getting contact with the mouse skin) and transferred in a 10-cm tissue culture dish containing 10ml sterile PBS (DULBECCO'S w/o Mg^{2+}, Ca^{2+}). A pair of fine watchmakers forceps was used to hold the uterine wall and cut through one side of the uterine wall to expose individual embryos. The yolk sac was opened and each fetus was dissected out, free of extra-embryonic membranes and placed in a new, clean 10-cm tissue culture dish with sterile PBS where the soft organs and viscera were removed (e.g liver, brain, gut etc.). The brain was either removed and discarded or kept for genotyping in the case that the parental mice were heterozygous. The remaining carcass was cut into small pieces using scissors and placed into 15ml falcon tubes containing 3ml 0.5X Trypsin (GIBCOM 10X) (1 embryo per falcon). It was left at a 37^0C water bath for 20 to 30min depending on the efficiency of the Trypsin and diluted by hand every 5min. Tubes were then vigorously resuspended, to break up the digested tissues into a cell suspension, three to five times. 4ml MEF culture medium was mixed and tubes were centrifuged at 1.200rpm, 5min at room temperature (RT). Cell pellet was resuspended in 1ml MEF medium and homogenized by pipetting up and down and then placed into 6-cm culture plates containing 5ml of MEF medium. Plates were incubated at a humidified 37^oC, 5% $C0_2$ incubator.

MEFs isolated from mouse embryos can be expanded in culture and manipulated experimentally at passages 3 and 4 or cryopreserved in a -80^0C freezer or liquid nitrogen tank when they have reached 80-100% confluency after isolation. They can be used as primary cultured cells or further immortalized to generate cell lines that can be maintained permanently in culture. MEF culture medium consists of DMEM, 10%FBS and 1.5% Penicillin –Streptomycin and is renewed every 2-3 days.

3.2 Transfection assays

3.2a siRNA of XAB2 in different cell types

Small (or short) interfering RNA (siRNA) is a synthetic RNA duplex which has been designed to target a specific mRNA or even a long non-coding RNA (lncRNA). This synthetic RNA is commonly used as a RNA interference (RNAi) tool in order to induce short-term silencing of protein coding genes. This transient knock-down can take place in various cell lines and primary cells and experiments in this matter are limited within 96h after transfection of these synthetic nucleotides.

siRNAs are double stranded: one of the RNA strands is the antisense (or guide) and the other one is the sense (or passenger) strand. These two form a duplex of 19 to 25 nucleotides with 3' dinucleotide overhangs. This dsRNA can be transfected into cells and used by the RNAi machinery which cleaves and degrades the complementary single target mRNA sequence in the cytoplasm **(dharmacon.gelifesciences.com)**.

The dsRNA is loaded onto the RNA-induced silencing complex (RISC). This complex contains Argonaute 2 (Ago-2) which has the ability to cleave and release one strand from the dsRNA. That is the guide siRNA which activates the RISC complex and can direct the specific binding in the target mRNA through complementarity. Afterwards, Ago-2 cleaves the target mRNA between bases 10 and 11 relative to the 5' end of the siRNA antisense strand and in this way causes mRNA degradation and gene silencing **(Wang et al., 2010) (Figure 10.)**.

The dsRNA can be delivered to a cell by transfection (cationic liposomes or polymer based), electroporation (electrical pulse) or viruses (lentivirus, adenovirus, retrovirus). Our siRNA experiments were conducted using Lipofectamine 2000 (Invitrogen), a cationic lipid-based carrier **(de Fougerolles et al., 2008)**. Cationic lipids have three parts, a cationic head group, a lipophilic tail group and a connecting linker. The cationic lipid has the ability to spontaneously form multilamellar structures with negatively charged nucleic acids such as siRNA upon mixing. The lipocomplexes (lipofectamine and dsRNA) can protect the siRNA from being degraded and as a consequence, facilitate the cellular uptake of the dsRNAs through endocytosis **(Wang et al., 2010)**. Moreover, the release of the siRNA from the endosomal/lysosomal entrapment is enhanced in a way that siRNA accumulates in the cytosol and can target the mRNA for cleavage.

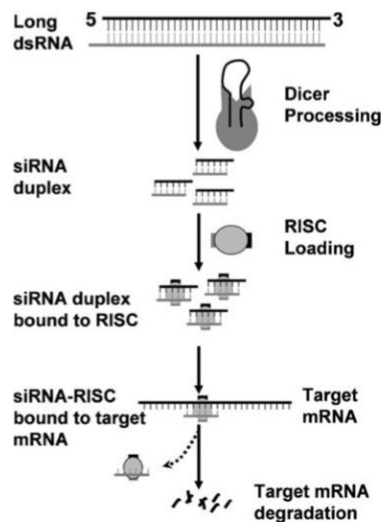


Figure 10. Mechanism of RNAi using dsRNA nucleotides **(Adapted from Wang et al., 2010)**.

The dsRNA nucleotides were designed using the Dharmacon online tool and ordered from Invitrogen. Their names are attributed to the position of the first nucleotide of the XAB2 cDNA that they target. For example, 1685 RNA nucleotide sequence starts from the 1685th nucleotide of the XAB2 cDNA. The sequence as well as the transcript ID target of each dsRNA is shown at **Table 1.**

dsRNA nucleotide	5'→3' prime sequence	TRANSCRIPT TARGET ID
1355	CCTTAAAGCTGCTGCGGAAUU	ENSMUST00000019614.12
1460	GGTCCATGCTTGCCGACTTUU	ENSMUST00000019614.12 ENSMUST00000159548.1 ENSMUST00000159235.1
1685	CCAAATTCATCTCGCGCTAUU	ENSMUST00000019614.12 ENSMUST00000159548.1 ENSMUST00000159235.1
1874	CAGCACAGCAGTACGACATUU	ENSMUST00000019614.12 ENSMUST00000159548.1 ENSMUST00000159235.1

Table 1. Main information on the dsRNA nucleotides used for the siRNA transfection experiments.

The protocol followed in order to knock-down XAB2 in Mouse Embryonic Fibroblasts was a combination **of Singh et al., 2015**, **Ma et al., 2010** and **Lipofectamine 2000 Manufacturer's Protocol**, whereas for the knock-down of XAB2 in JM8A3.N1 and bioXAB2;birA mouse embryonic stem cells, it was a combination of **Ma et al., 2010** and **Lipofectamine 2000 Manufacturer's Protocol (Mouse D3 embryonic stem cells siRNA transfection)**.

To prepare siRNA/lipid solutions for each 6cm plate, 600pmol of siRNA (20pmol/μl) for MEFs or 250pmol of siRNA (20pmol/μl) for mESCs were diluted in 500μl OPTI-MEMI (Invitrogen) and incubated at room temperature for 5min. In a separate tube, 10μl of Lipofectamine 2000 (Invitrogen) was diluted in 500μl OPTI-MEMI and incubation was carried out for 15min at room temperature. The contents of the two tubes were combined by gentle pipetting and incubated at room temperature for 20-40min. The resulting transfection solution (complexes, 1.040mL for MEFs or 1.022mL for mESCs) was used to resuspend a cell pellet of $3x10^5$ MEFs at passage 4 or $1x10^6$ mESCs. The mix was flicked every 2min and incubated for 7-9min. The cells-complexes mix was added dropwise in 6cm tissue culture plates containing 5ml MEFs or mESCs growth medium. Fresh medium (2mL) was added 24h post-transfection. Subsequent analysis (RNA or protein extraction, stainings) was performed 48h or 72h post-transfection. In case that subsequent staining was to be performed, 0.2% gelatin coated plates with coverslips were used. To further improve the knock-down efficiency, the dsRNA concentration was increased up to 1.4 times for MEFs or 1.6 times for mESCs (it will be mentioned in each case). Scrambled (AllStars Negative control, Qiagen), mock transfected (lipofectamine only) and untransfected controls were also included.

3.2b Overexpression of XAB2 in MEFs

On the first day, Passage 3 MEFs were split in 6cm plates of $3x10^5$ cells each. On the second day, 8μg of plasmid DNA (50ng/μl) carrying the XAB2 cDNA fused with the ATF-tag was used to overexpress XAB2 according to the manufacturer's protocol.

3.3 RNA extraction

RNA extraction was conducted either with TRIreagent (Sigma), according to manufacturer's protocol for the evaluation of the knock-down efficiency, or with RNeasy Mini Kit (QIAGEN) for the RNA-seq experiments.

RT-qPCR

For the first strand (cDNA) synthesis a mix of 1μg total RNA (~500ng/μl), 1μl of Oligo dT (Invitrogen) and RNase free water to 11.5μl final volume was prepared and incubated in a thermocycler at 65^0C for 10min and 4^0C for 5min. During the 5min incubation at 4^0C, 8.5μl of the RT mix (5x RT buffer, 10mM dNTPs, 0.1M DTT, RNAse Out (40u/μl) and Superscript II)) was added in each sample to reach a 20μl final volume. Afterwards the samples were incubated at 42^0C for 60min and cooled at 4^0C.

The samples were diluted 5X with ddH_20 and stored in -20^0C or used for qPCR according to the reaction mixture shown in **Table 2**.

5x diluted cDNA	2.5μl
50mM $MgCl_2$	1.25μl
10x Platinum Taq Buffer	2.5μl
SYBR1 (1/2000)	0.75μl
10mM dNTPs	0.4μl
qPCR primer mix	4μl
Platinum Taq Polymerase	0.1μl
H_20	13.5μl
Final volume	**25μl**

PRIMERS SEQUENCE	
XAB2	F: 5'- TCATGGCTTCACAGATGCTC-3'
	R: 5'-GGATGCATCACTCCTCACAA-3'
HPRT	F: 5'-CCCAACATCAACAGGACTCC-3'
	R: 5'CGAAGTGTTGGATACAGGCC-3'

Table 2. Upper: qPCR reaction mixture, Lower: Primers sequence of XAB2 and HPRT cDNA used for qPCR reaction

It should be noted that SYBR Green was added just before adding the mix to the samples. Knock-down efficiency was evaluated by the relative expression of XAB2 cDNA to HPRT cDNA **(Table 2.)**. The PCR program that was followed was: 94^0C for 5min, 94^0C for 15s, 60^0C for 25s, 72^0C for 25s, 78^0C for 1s (x39 cycles from second step), 72^0C for 2min, 72^0C for 1s and 94^0C for 5 s.

3.4 Western Blot

Protein extraction

Adherent cells were washed twice with PBS, trypsinised with 1x Trypsin (GIBCO) and centrifuged at 1.200rpm for 5min, RT. The cell pellet was washed once with PBS and resuspended with 5 volumes of Whole Cell Protein Extraction Buffer (50mM Tris-HCL

pH=7.5, 150mM NaCl, 5% Glycerol, 1% NP-40, 1mM PMSF, 1x PIs (Sigma)). The mix was kept on ice for 30 min and was flicked by hand every 5min. Cellular debris were then pelleted at 14.000rpm for 15min at 4°C. Supernatant was kept as the whole cell protein extract and its protein concentration was determined by Bradford assay. As far as the Bradford assay is concerned, 20μl were used for each BSA standard (0.1μg/μl, 0.25μg/μl, 0.5μg/μl, 0.75μg/μl, 1μg/μl) and 1-4 μl were used from the extracts/WCE buffer as control, depending on cell pellet. 1ml of BIORAD working solution (stock solution diluted 1:5 with ddH_20) was added to each sample and 200μl/sample were transfered in duplicates in a 96-well plate. The absorbance was measured by [Elisa reader] using Chromate Manager application.

SDS-PAGE

Whole protein extracts (50μg) were rediluted with ddH_20 and 5x Loading Dye in a final volume of 50μl and were heated for 10min at 80^0C. The apparatus used for the SDS-PAGE is the Mini-PROTEAN® Tetra Vertical Electrophoresis Cell (BIORAD). Samples were then loaded into an SDS-PAGE gel using gel loading tips. As far as the SDS-PAGE gels are concerned, separating gels of 10% for a M.W range of 20kDa -300kDa and 14% for a M.W range of 3kDa- 100kDa were used. Stacking gel is poured on top of the separating gel and a gel comb is inserted in the stacking gel. The acrylamide percentage in SDS-PAGE is dependent on the size of the target protein in the sample.

10ml separating gel:

10%	**14%**
3.35ml Acrylamide-Bis-acrylamide (29:1)	4.7ml Acrylamide-Bis
2.5ml Tris 1.5M, pH=8.8	2.5ml Tris 1.5M, pH=8.8
100μl 10% SDS	100μl 10% SDS
70μl APS 10%	*70μl APS*
5μl TEMED	*5μl TEMED*
4ml ddH_20	2.6ml ddH_20

10ml stacking gel:

1.35ml Acrylamide-Bis (29:1)
2.5ml Tris-HCl 0.5 M, pH=6.8
100μl 10% SDS
70μl APS
10μl TEMED
6ml ddH_20

Table 3. The volumes of the separating and stacking gels for one 1.5mm thick gel.

The SDS-PAGE runs for an approximate time of 2h (30min in 50V until proteins have exited from the stacking gel and 1h and 30min in 110V for the proteins to be completely separated).

TRANSFER

After electrophoresis, the separated proteins are transferred onto a solid support matrix which in this case is a nitrocellulose membrane (Life Sciences). Specifically, after the apparatus is disassembled, the electrophoresis gel is put in Running Buffer and the stacking gel is removed. The separating gel is then kept in Transfer Buffer until transfer apparatus is ready. Wet electroblotting (Tank Transfer) was used (Mini Trans-Blot® Cell) in which as already mentioned the gel was first equilibrated in transfer buffer. The gel is placed in a "transfer sandwich" (sponge→ 2 whatmann papers→ gel→ nitrocellulose membrane→ 2 whatmann papers→sponge). All these were pressed together using a support grid and the gel sandwich was placed vertically in a tank between stainless steel wire electrodes and filled with transfer buffer. The gels were electrotransfered using 300mA as a constant current for 1h in RT by the addition of an ice pack in the apparatus.

Detection

After the transfer of the proteins from the gel, the nitrocellulose membrane was washed once with PBS-T buffer (PBS, 0.1% Tween) and Ponceau buffer (Sigma) was added in order to detect the proteins that had been transferred. Following two 5min PBS-T washes, the membrane was blocked in 5% milk in PBS-T for 1h, RT in order to prevent non-specific binding of the detection antibodies during subsequent steps. The blocked membrane was probed using the antibodies depicted in **Table 4.**, overnight in 1% milk-PBS-T at 4^0C.

Primary Antibody	Species	Dilution	Company	Secondary Antibody	Dilution
XAB2	Rabbit	1:1000	Abcam	α-Rabbit HRP	1:10.000
RNA-polII	Mouse	1:500	Santa-Cruz	α-mouse HRP	1:10.000
γ-tubulin	Goat	1:500	Santa-Cruz	α-goat HRP	1:5.000
XPB	Rabbit	1:500	Santa-Cruz	α-Rabbit HRP	1:10.000
XPA	Mouse	1:500	Santa-Cruz	α-mouse HRP	1:10.000

Table 4. Information on the antibodies used for immunoblotting are depicted.

Following the overnight incubation with the primary antibody, the membrane was washed 3 times for 5 or 10min with PBS-T and incubated with the secondary antibodies diluted as well, in 1% milk in PBS-T, for 1h RT. HRP (horseradish peroxidase)

conjugated antibodies were used as a label for protein detection and the array of chemiluminescence substrate was available for the use of this enzyme (ECL-Thermo Scientific). The signal is a transient product of the enzyme-substrate reaction and persists as long as the reaction is occurring. The detection was allowed using X-ray films (Fujifilm) in 4 exposure times (instant, 30s, 1min and 3min).

3.5 Immunocytochemistry for Adherent Cells

MEFs, JM8A3.N1 and bioXAB2;birA mESCs were fixed in methanol for 10min in -20^{0}C and 4% formaldehyde fixation freshly made was also conducted for 10min in RT. Two times PBS washes followed. If coverslips were previously kept in -20^{0}C, 3 times PBS washes for 5min each was done.

Blocking and permebilization were conducted for 1h at RT with 1% BSA, 0.1%Triton X in PBS (B1 buffer). Desired primary antibodies **(Table 5.)** were prepared in B1 buffer and mix was added on a parafilm with the coverslips having cells facing down for 1h or 2h RT. Coverslips were then washed with B2 buffer (0.1% Triton X in PBS), cells facing upwards, 3 times, 10min per wash on a rotating platform.

Primary Antibody	Fixation	Species	Dilution	Company	Secondary Antibody Alexa-fluor 488/555	Dilution
XAB2	F/A	Rabbit	1:1000	Abcam	α-rabbit	1:2000
γ-H2AX	F/A	Mouse	1:100	Millipore	α-mouse	1:2000
pATM	F/A	Mouse	1:1000	Santa-Cruz	α-mouse	1:2000
S9.6	methanol	Mouse	1:200	Kerafast	α-mouse	1:2000
Rad51	F/A	Rabbit	1:50	Santa-Cruz	α-rabbit	1:2000
Nucleolin	F/A	Rabbit	1:1000	Abcam	α-rabbit	1:2000
RNA-polII	F/A	Mouse	1:50	Santa-Cruz	α-mouse	1:2000
RPA	F/A	Mouse	1:100	Abcam	α-mouse	1:2000
LaminB1	F/A	Rabbit	1:100	Abcam	α-rabbit	1:2000
XPA	F/A	Mouse	1:50	Santa-Cruz	α-mouse	1:2000
Flag	F/A	Mouse	1:1000	Sigma	α-mouse	1:2000

Table 5. Information on the antibodies used for the immunocytochemistry experiments are depicted.

Following the washes, cells were incubated with secondary antibodies **(Table 5.)** and DAPI (nuclear staining, Thermo Scientific, 1:20000) diluted in B1 buffer for 1h RT. Coverslips with cells facing upwards were washed 3 times with B2 buffer, 10min per wash on a shaking platform. Lastly, the specimens were mounted (cells facing down) with 80% glycerol in PBS on microscope slides and stored in 4^{0}C for a short period of time or -20^{0}C for long term storage. Immunostainings were analyzed using confocal microscopy at the Leica SP8 inverted confocal of the IMBB-FoRTH imaging facility.

3.6 Immunoprecipitation Assay

Co-immunoprecipitation followed by western blot was carried out to validate previously shown XAB2 interactors in nuclear protein extracts from MEFs treated and untreated with UV irradiation.

Day 1.

At the first day of the assay, cells were lysed, cell pellet was washed once with PBS and resuspended in five volumes of NP-40 Lysis Buffer (10mM Tris-HCl pH=7.4, 10mM NaCl, 3mM MgCl2, 0.5% NP-40, 1mM PMSF, 1x PIs- PMSF and PIs were added before use). The mix was rotated for 10min at 4^0C and centrifuged at 1.500rpm, 4^0C for 5min. Supernatant was kept as the cytoplasmic fraction and a second NP-40 wash for 3min followed. The mix was centrifuged at 1.500rpm for 5min at 4^0C and the supernatant was discarded. Pellet was resuspended in five volumes of High Salt Buffer (10mM HEPES pH=7.9, 20% Glycerol, 380mM KCl, 3mM MgCl2, 0.2mM EDTA, 1mM DTT, 1mM PMSF, 1x PIs) and incubated at 4^oC for 1h on a rotating platform. Insoluble nuclear material was precipitated by centrifugation at 13.000rpm for 30min at 4^0C and the supernatant was kept as the nuclear extract. The concentration of the proteins of both the cytoplasmic and the nuclear extracts were measured using Bradford (BIORAD) assay.

Two immunoprecipitation reactions were prepared: one for the antibody of interest and one for the species matched IgG control. For each reaction 500μg of nuclear extract was used and diluted three times with HENG buffer (10mM HEPES-KOH pH=9, 5mM $MgCl_2$, 25mM EDTA, 20% Glycerol, 1mM PMSF (100mM) or PIs) and supplemented with benzonase (1μl/mg of protein), RNAse H (0.2μg/μl), $MgCl_2$ (2mM), 1mM PMSF and PIs (1:1000). All samples were incubated overnight at 4^0C rotating with a minimal volume of 500μl. An input sample was prepared by diluting 50μg of the nuclear extract (1/10 of the IP material) in Laemmli buffer, heated at 80^oC for 10min and stored at -20^oC.

Day 2.

On the second day, sepharose beads (100μl for preclearing and 50μl for each IP of 500μg protein) were prepared: beads were washed with 1mL PBS on a rotating platform for 1h, RT. The beads were precipitated by centrifugation at 6.000rpm for 3min. The washes were repeated for two more times. Beads were then blocked by adding 1ml HENG and 200μg/ml of Chicken Egg Albumin (20mg/mL) for 1h at a rotating wheel at room temperature and centrifuged at 6.000rpm, 3min, RT. Supernatant was discarded and the nuclear extracts that had been incubated overnight with benzonase and RNase were added on the beads to be pre-cleared for

1h at 4^0C. After a centrifugation of 6.000rpm for 3min at 4^0C, supernatants were incubated overnight at a rotator at 4^0C with the desired antibodies (XAB2: 5μg/IP, IgG: 3.5μg/IP).

Day 3.

At the third day, the nuclear extract-antibody mixes were added to the pre-equilibrated and blocked beads (50μl) for 1h- 2h at 4^0C on a rotator. They were afterwards centrifuged at 6.000rpm for 3min, at 4^0C and the flow-throughs were kept on ice. Beads were rinsed twice and washed (10min each, 4^0C, rotator) three times with Wash buffer (HENG, 380mM KCL, 0.3% NP-40, 0.5mM PMSF or PIs). After one rinse with PBS, bead-bound proteins were eluted by boiling in 2X Laemmli buffer for 10min at 80^0C, 800rpm. Flow-throughs were prepared by redilution in 5x of Laemmli buffer. All samples were stored in -20^0C until loading on a polyacrylamide gel (marker-input- IP- IgG control- S/N).

3.7 RNA-seq data analysis

RNA was extracted from 6cm plates of JM8A3.N1 cells from biological triplicates of mock (lipofectamine only) and siRNA for XAB2 samples. 1μg RNA from each sample was ran in the Ion Torrent Next Generation Sequencer. Sequencing data were quality tested using the FastQC quality control application. The reads obtained were aligned to the mouse genome (mm103884). The cufflings algorithm was used in order to detect differentially expressed and spliced genes between the control and XAB2 knock-down samples. False Discovery Rate (FDR <0.05) was used in order to distinguish statistically important differences between mock and siRNA samples.

4. Results

4.1 Efficient knock-down of XAB2 in three different cell types

Our main objective was to identify the cellular phenotype in the absence of XAB2. To address this, we established protocols to efficiently knock-down XAB2 in MEFs and two genotypes of mESCs. Four different 21bp dsRNA nucleotides were designed using the Dharmacon online design tool. The efficiency of RNAi was evaluated for each oligonucleotide RNA (single oligonucleotide experiments), as well as for several combinations, (double or triple oligonucleotide experiments) by mRNA expression and protein levels.

As shown in **Figure 11**, a single oligonucleotide RNA (1685), two double combinations (1874-1685, 1355-1685) and a triple combination of oligonucleotide RNAs (1874-1355-1685) had the highest knock-down in the RNA levels of XAB2, 24h post-transfection, using the untreated cells as a control. It must be noted, as this was the first experiment conducted, cells transfected with lipofectamine and scrambled RNA oligonucleotides were stressed more than the rest of the knock-down experiments which explains the ~40% knock-down of XAB2. We also isolated protein extracts 24h post-transfection, using the TRI (Sigma) protocol, which directed the redilution of the protein pellet in 1% SDS, which did not allow the protein quantification by Bradford assay. Moreover, it seems that our loading control in Western blots, TBP, is an off target of the dsRNA oligonucleotides. However, 1685 RNA oligonucleotide seems to produce the highest decrease in XAB2 protein levels among its counterparts with the same loading control quantity **(Figure 11.)**.

Since 1685 oligonucleotide is sufficient to reduce effectively the XAB2 mRNA and protein levels, it was used in a 12-well plate format in JM8A3.N1 cells to study the RNA and protein levels of XAB2 at 24h, 48h and 72h post-transfection **(Figure 12.).** Xab2 mRNA levels are efficiently reduced at 24.2% of the scrambled control 24h post-transfection, a percentage that is slightly enhanced after another 24h (16.8%). The protein levels, however, are significantly reduced only after 48h. Both the mRNA and protein levels are restored 72h after siRNA transfection which is a consequence of the minimal number of cells still attached on the tissue culture plate surface rather than loss of siRNA activity. This is in agreement with several studies **(Kuraoka et al., 2008; Ohnuma-Ishikawa et al., 2007; Onyango et al., 2016)**. In this way, our experiments were limited to 48h post-transfection, even though in some cases, we managed to obtain RNA and protein, 72h post-transfection **(Figure 12.)**.

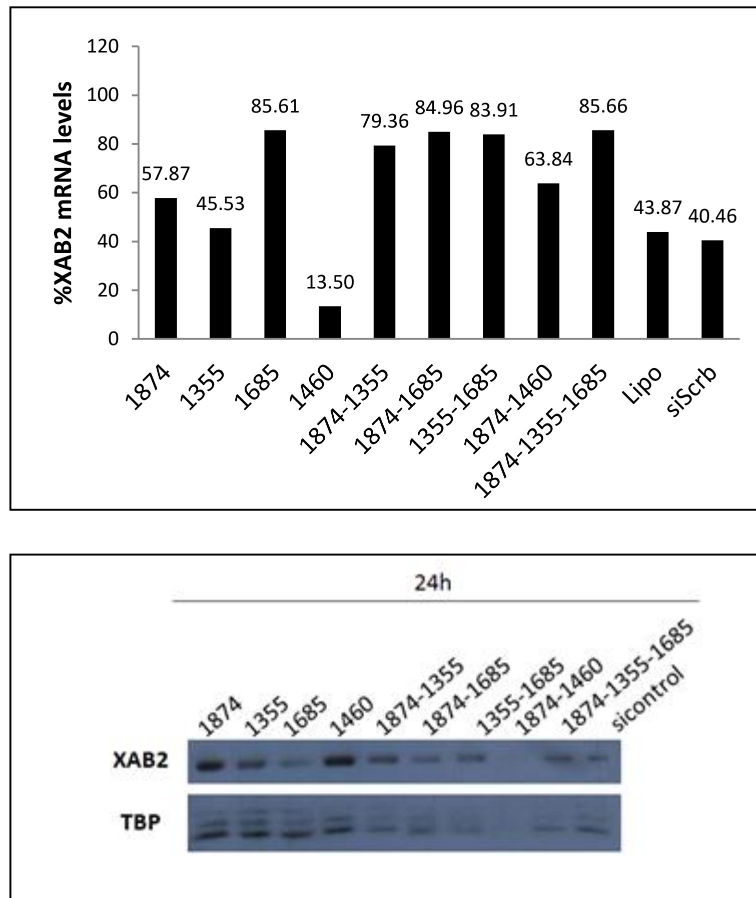


Figure 11. In the upper panel, the percentage of XAB2 mRNA levels, 24h after transfection of JM8A3.N1 cells with the four oligonucleotides (1355, 1460, 1685, 1874) and combination of them, are depicted. In the lower panel, the protein levels of XAB2, 24h post-transfection are shown, where the most efficient knock-down of XAB2 is due to 1685 siRNA oligonucleotide.

Further knock-down experiments of XAB2 were also conducted in bioXAB2;birA mESCs using the most efficient RNA oligonucleotide, 1685. A knock-down of 74.3% in the RNA levels of XAB2 was managed 48h post-transfection, in a 12-well plate format experiment and the decrease of XAB2 was also prominent at its protein levels as well **(Figure 13.)**.

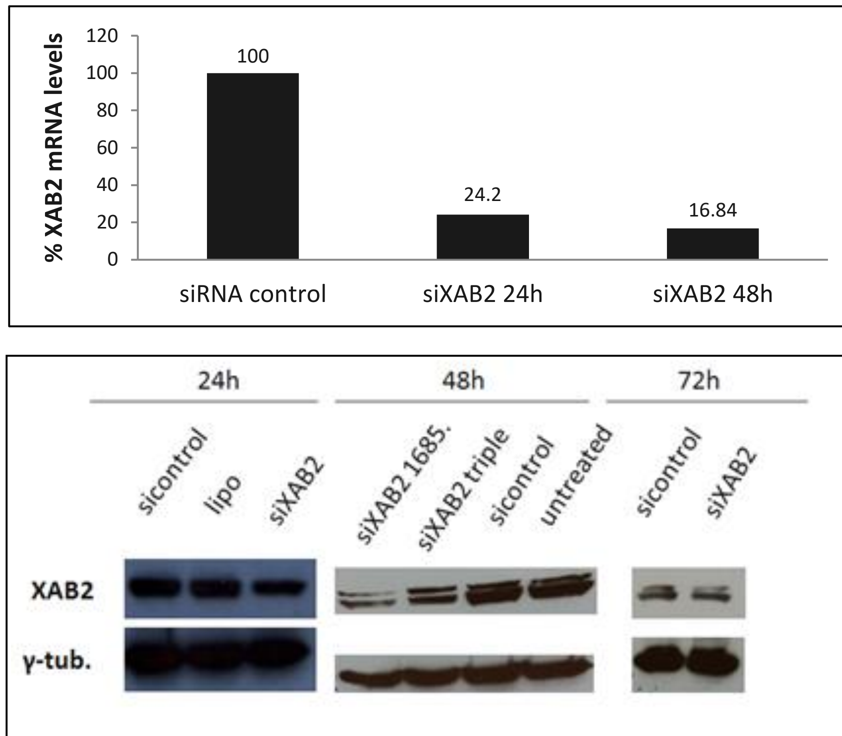


Figure 12. In the upper panel, the low percentage of XAB2 mRNA levels 24h and 48h after transfection of 1685 RNA oligonucleotide in JM8A3.N1 are depicted. In the lower panel, the protein levels of XAB2, 24h, 48h and 72h are shown, where the most efficient knock-down of XAB2 is 48h post-transfection.

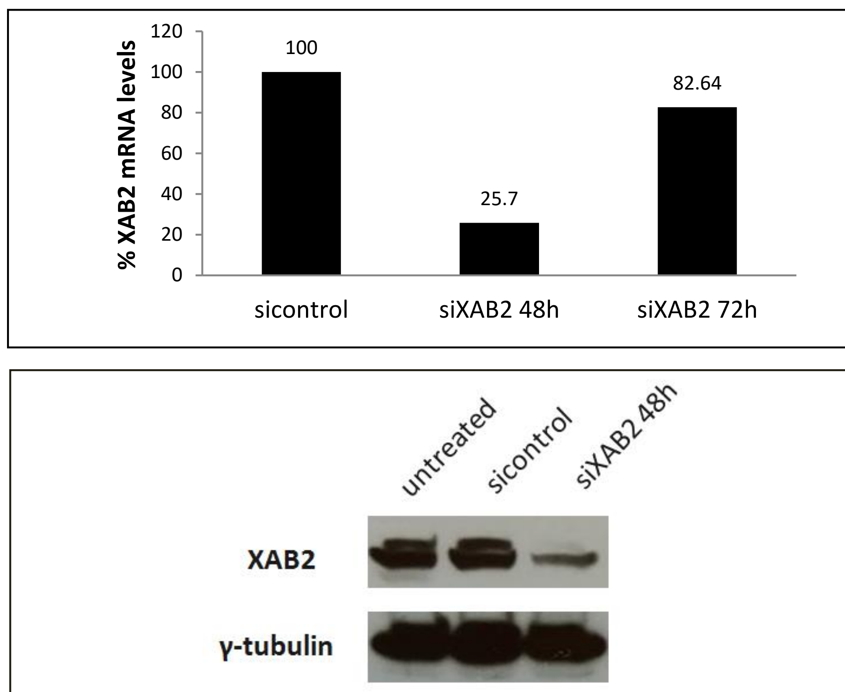


Figure 13. In the upper panel, the XAB2 RT-qPCR analysis shows the mRNA levels 48h after transfection of 1685 RNA oligonucleotide in bioXAB2;birA mESCs. In the lower panel, the protein levels of XAB2 48h post-transfection are shown. The efficient knock-down of XAB2 in 48h post-transfection is obvious both in its RNA and protein levels.

Moreover, the flag-tagged protein can hardly be detected in the immunostaining using a flag antibody, another assay used to verify the efficient knock-down of XAB2 protein levels in bioXAB2;birA knocked-down cells **(Figure 14)**.

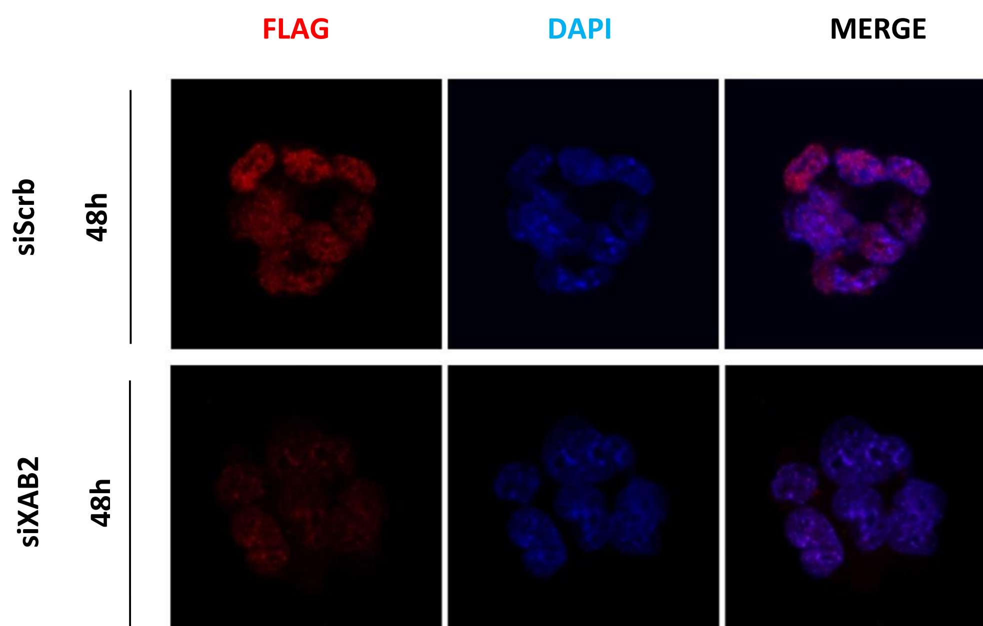


Figure 14. XAB2 staining, using a flag antibody for the flag-tagged XAB2 in the knocked-down bioXAB2;birA mESCs, shows the decrease of XAB2 48h post-transfection, in comparison to the cells transfected with scrambled RNA oligonucleotides.

Mouse Embryonic Fibroblasts were also chosen, as a primary cell type, for the knock-down assays as they do not form colonies (flatter structure) and have larger cytoplasm than mESCs, allowing (easier and safer) immunostaining observation. Passage 4 C57BL/6 MEFs were transfected with the 1685 RNA oligonucleotide in a 6cm plate format and a knock-down of XAB2 both in its RNA (80.05%) and protein levels was achieved 48h post-transfection **(Figure 15.)**.

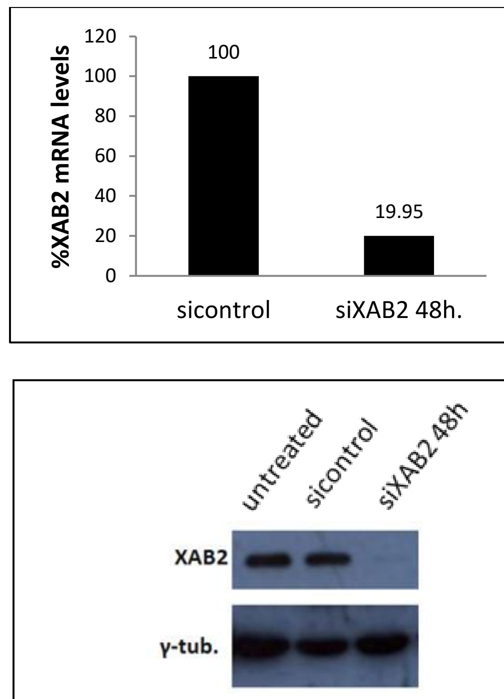


Figure 15. In the upper panel, the XAB2 RT-qPCR analysis shows the mRNA levels 48h after transfection of 1685 RNA oligonucleotide in passage 4 C57BL/6 MEFs, whereas in the lower panel, the protein levels of XAB2 48h post-transfection are shown. The efficient knock-down of XAB2 in 48h post-transfection is obvious both in its RNA and protein levels.

4.2 XAB2 knock-down reduces RNA-DNA hybrids

It is known that deletion of RNA binding factors that are associated with pre-mRNA processing leads to the formation of R-loops which activate the ATM pathway and lead to DSBs **(Li and Manley et al., 2006; Aguilera and Garcia-Muse, 2012)**. Given that XAB2 is considered a pre-mRNA splicing factor, we decided to assess the impact of its elimination on R-loop formation. To address this, we used the S9.6 antibody, which detects RNA-DNA hyrbrids, in mESCs as well as MEFs subjected to siXAB2. In an untreated cell **(Sollier et al., 2014)**, RNA-DNA hybrids are highly detected throughout the cytoplasm and in the nucleolus. However, the staining showed reduction of these hybrids in the nucleoli of JM8A3.N1 mESCs, 72h post-transfection **(Figure 16.)** and absence of these hybrids in the nucleoli of MEFs, 48h post-transfection **(Figure 17.)**. The time difference of the phenotype could be attributed to the difference in dsRNA concentration or stress response among the cell types. The signal was dispersed and faint, in comparison to the cells transfected only with the scrambled RNA oligonucleotides, as detected in **Figure 16.** Such a finding suggests that transcription taking place in the nucleolus which deals with rRNA is less active in the knock-down cells. Moreover, nucleolin, a protein required for rRNA transcription (synthesis and maturation of ribosomes), is more dispersed and present throughout the nucleus

(compared to the concentrated in a few areas phenotype in siScrambled cells), 72h post-transfection in the siXAB2 cells, suggesting nucleolar disruption.

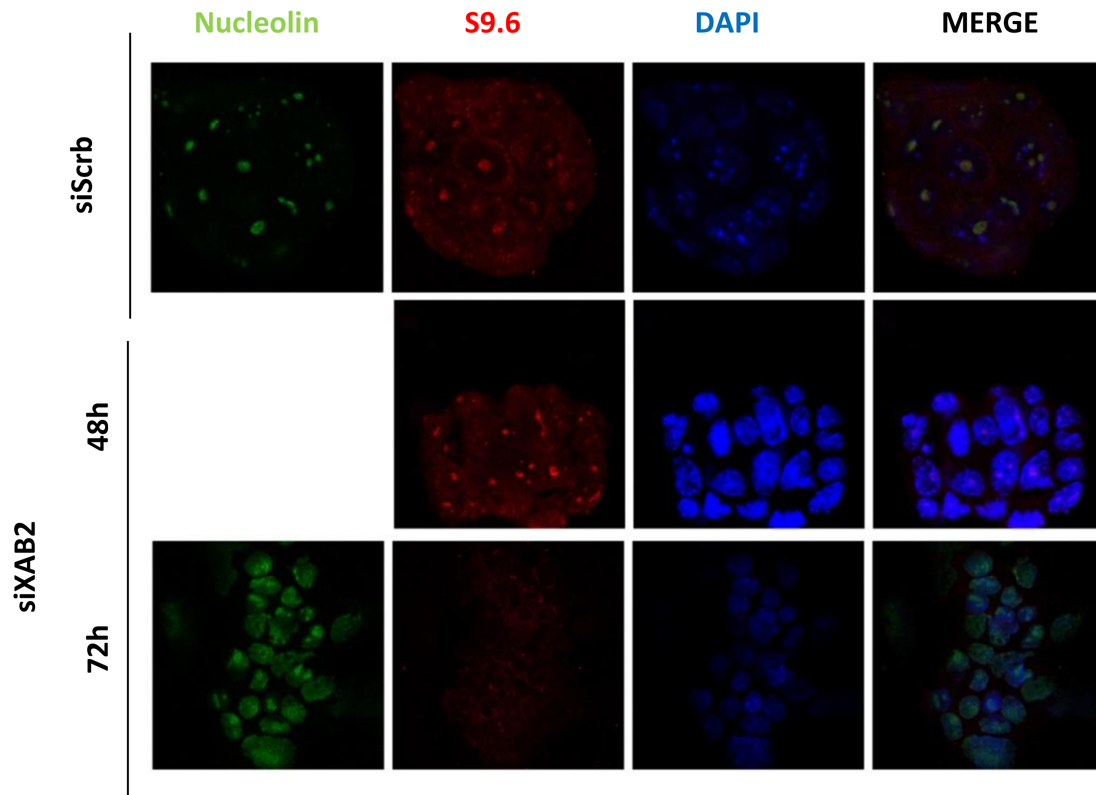


Figure 16. S9.6 staining shows the decrease of RNA-DNA hybrids in the nucleolus of 72h post-transfected siXAB2 mESCs in comparison to siScrambled cells whereas nucleolin seems more dispersed 72h post-transfection in the siXAB2 cells compared to siScrambled cells.

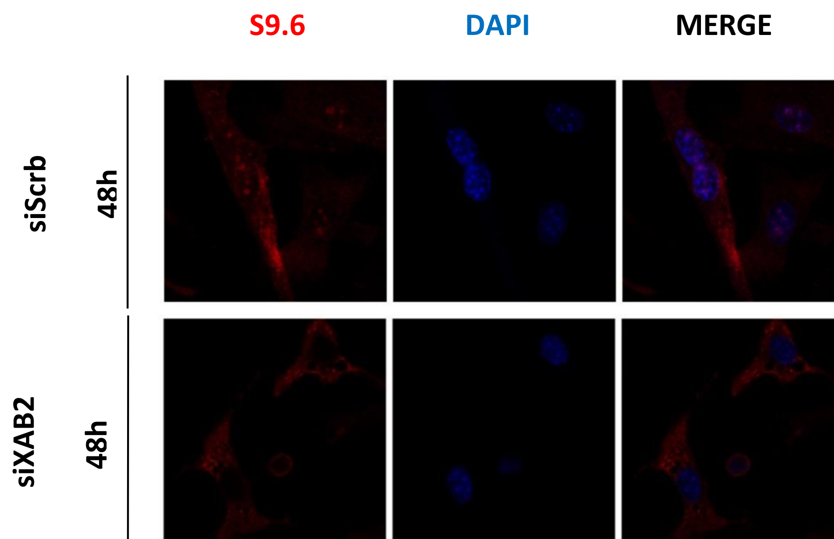


Figure 17. S9.6 staining shows the decrease of RNA-DNA hybrids in the nucleolus of 48h post-transfected siXAB2 MEFs in comparison to siScrambled cells.

Having observed the decrease of the RNA-DNA hybrids in knocked-down JM8A3.N1 mESCs and MEFs, it was concluded that this must be a universal phenomenon as far as the XAB2 knock-down is concerned. Moreover, it could be linked to a deficiency in transcription and this is the reason why RNA polII expression was studied. It seems that this might be the case as RNApolII protein levels are slightly decreased both 24h and 72h after transfection of 1685 RNA oligonucleotide in JM8A3.N1 mESCs and 48h post-transfection in MEFs **(Figure 18.)**. RNA polII levels were also studied through immunostaining in which the signal of RNA polII is less detectable in comparison to the siScrambled in bioXAB2;birA mESCs 72h post-transfection, showing the decrease of this protein in the siXAB2 cells **(Figure 18.)**.

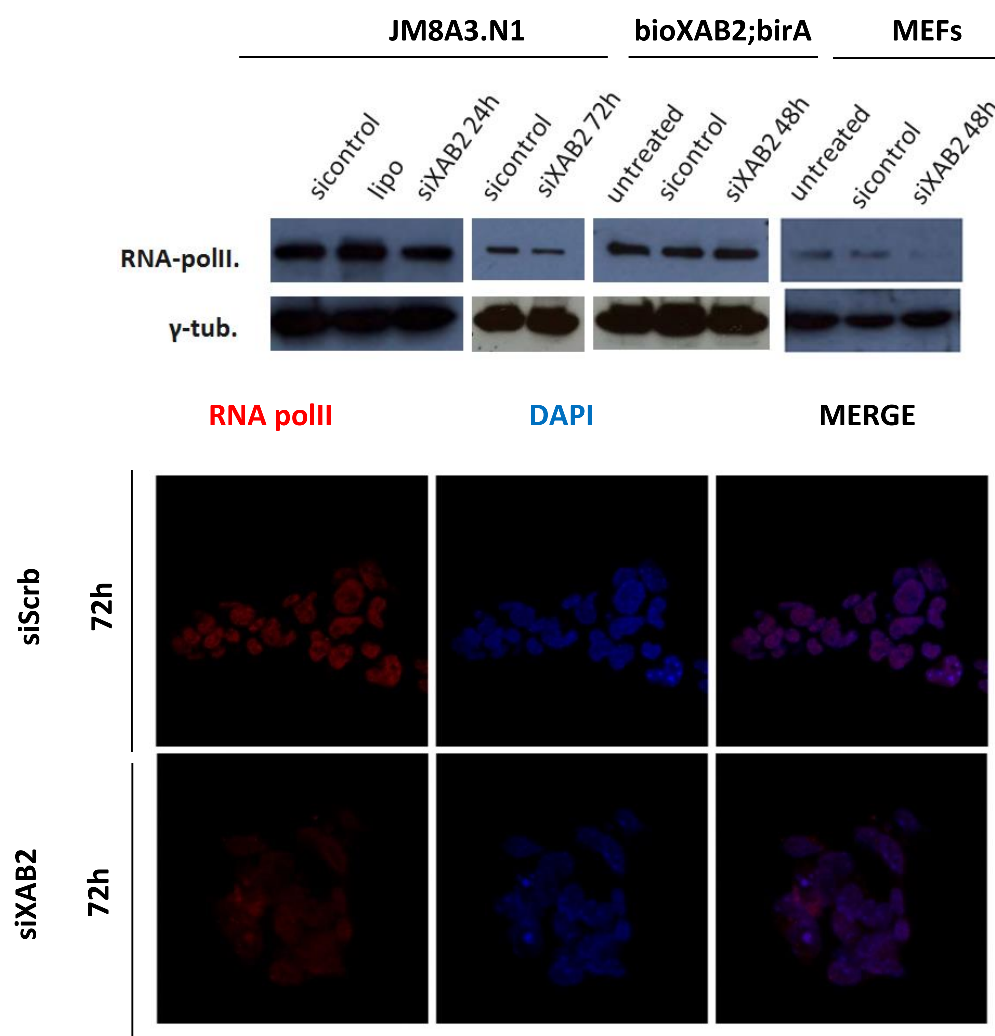
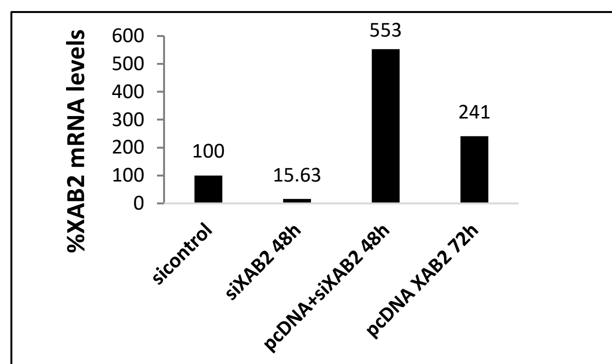
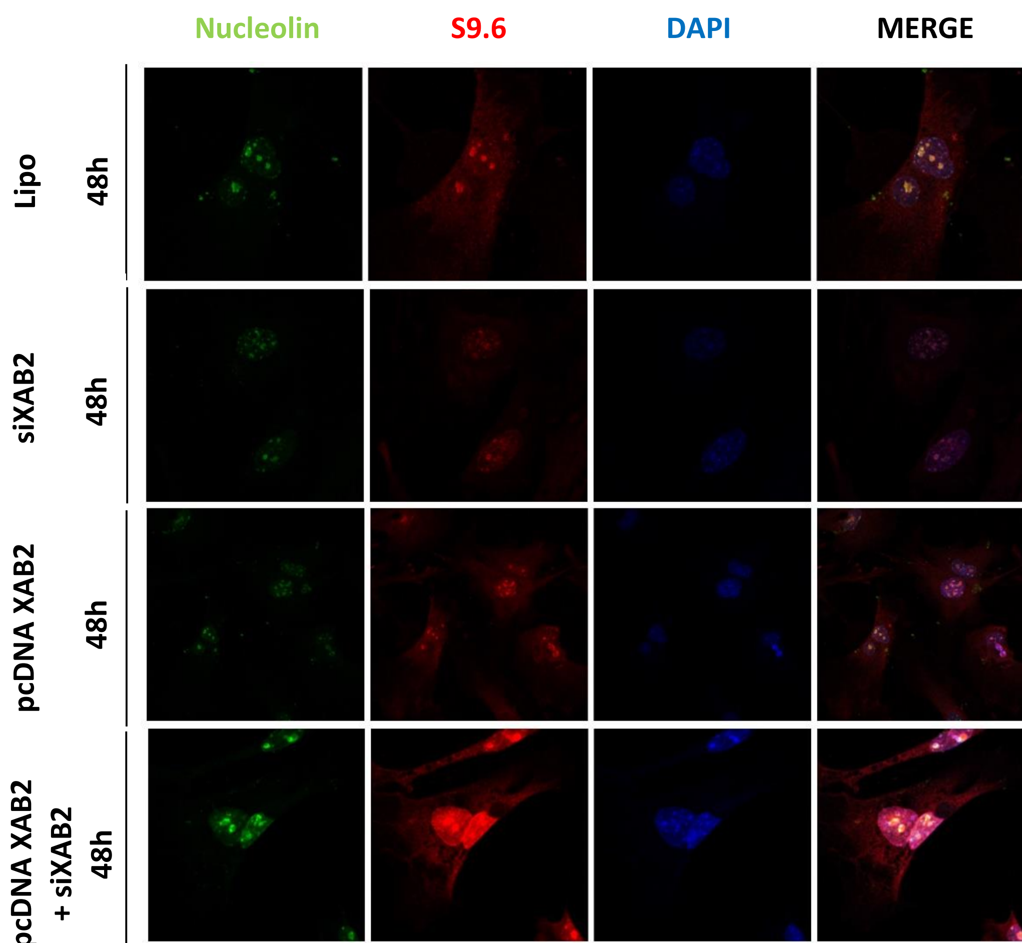


Figure 18. The upper panel shows the protein levels of RNApolII in the siXAB2 of JM8A3.N1 (24h, 72h post-transfection), bioXAB2;birA mESCs and MEFs (48h post-transfection). The lower panel shows the RNA polII staining where the decrease of RNA polII 72h post-transfection in the siXAB2 bioXAB2;birA mESCs cells in comparison to the cells transfected with scrambled RNA oligonucleotides is prominent.

4.3 Overexpression of XAB2 coupled with siRNA rescues S9.6 and RNApolII phenotype

Our next goal was to decipher whether the decrease of the RNA-DNA hybrids in the nucleolus was a direct effect of the XAB2 knock-down. For this reason, we decided to overexpress XAB2 by transfecting a pcDNA vector in which we cloned the ATF-tagged Xab2 cDNA, in MEFs. As it is shown in **Figure 19.**, overexpression of XAB2 72h post-transfection rescues the S9.6 phenotype compared to siXAB2 knock-down cells. The same case stands for the RNApolII expression which even though decreased 48h post-transfection of siXAB2 MEFs, returns to normal staining phenotype in pcDNA XAB2 and pcDNA + siXAB2 transfected cells **(Figure 19.)**.





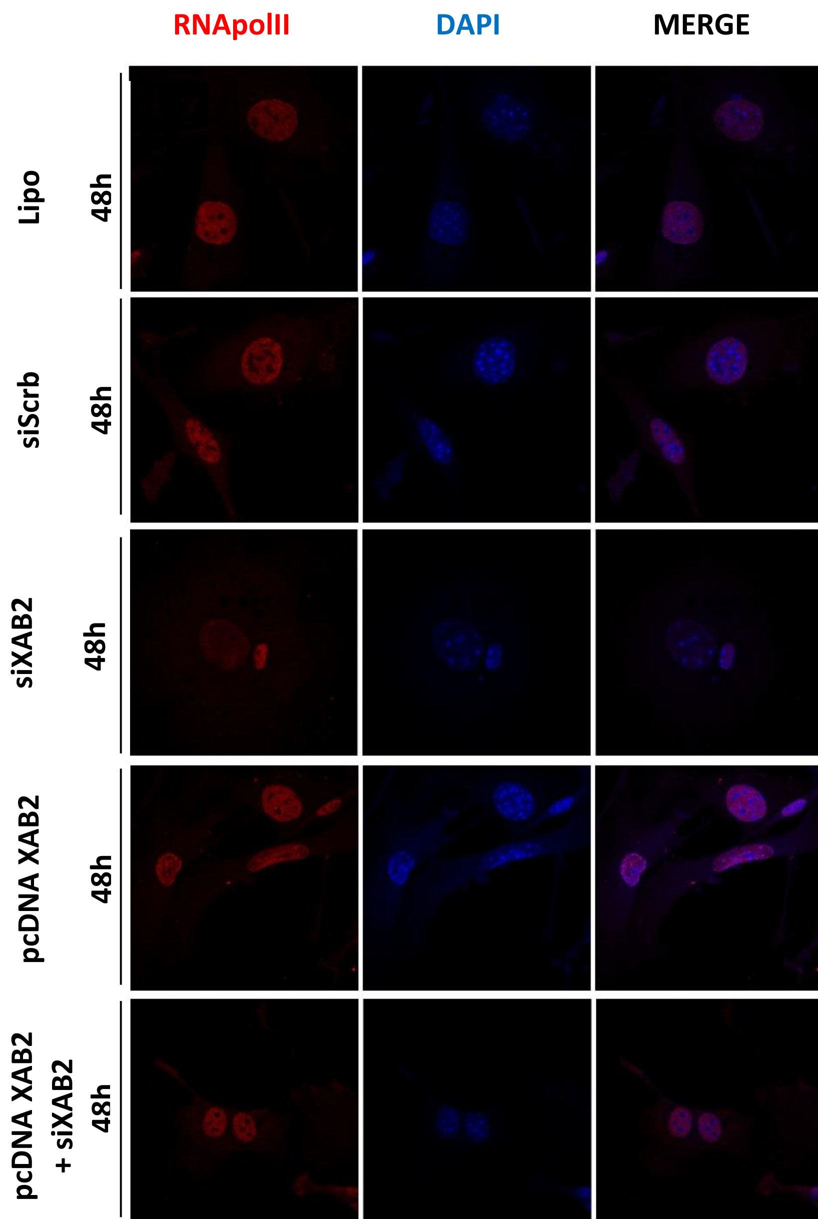


Figure 19. In the upper panel, the XAB2 RT-qPCR analysis shows the mRNA levels 72h after overexpression of pcDNA XAB2 in passage 4 C57BL/6 MEFs and 48h post-transfection of pcDNA XAB2 overexpressed cells with 1685 RNA oligonucleotide. In the lower panel, the rescued phenotype of S9.6 and RNApolII, in pcDNA XAB2 overexpressed cells and pcDNA XAB2 + siXAB2 cells, is depicted.

4.4 XAB2 does not cause damage or nuclear envelope disarrangement

Miller at al., 2009 proposed that XAB2 is a protein causing genome instability by inducing γH2AX foci enrichment, a marker used to detect sites of damage. However, that was not the case in our system as shown by γH2AX staining which was not enriched in siXAB2 MEFS 48h post-transfection **(Figure 20).**

It is known that in eukaryotes transcription, RNA processing and export are tightly linked. As XAB2 is a factor involved in pre-mRNA splicing, it would be possible that the cell death of the siXAB2 cells is attributed to the nuclear envelope being affected. However, the knock-down of XAB2 does not seem to affect nuclear envelope organization, as depicted in the conducted LaminB1 staining **(Figure 20).**

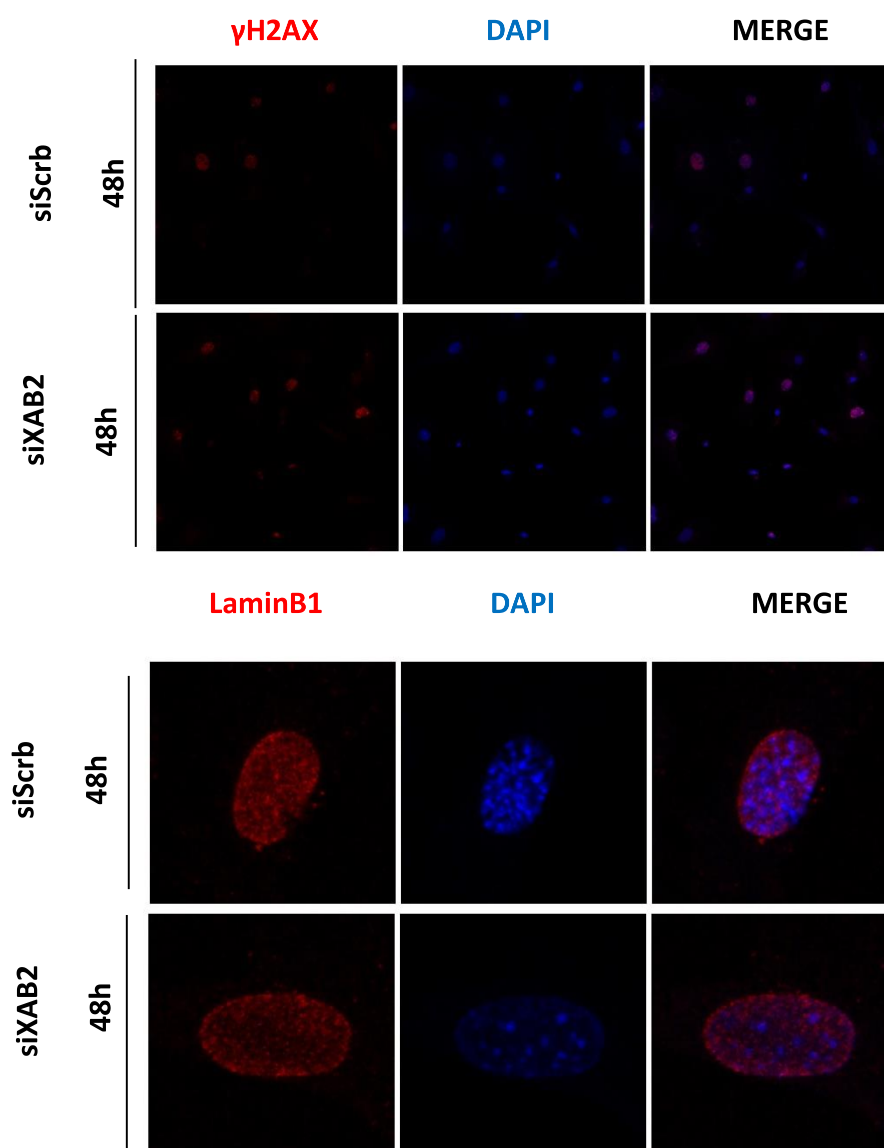
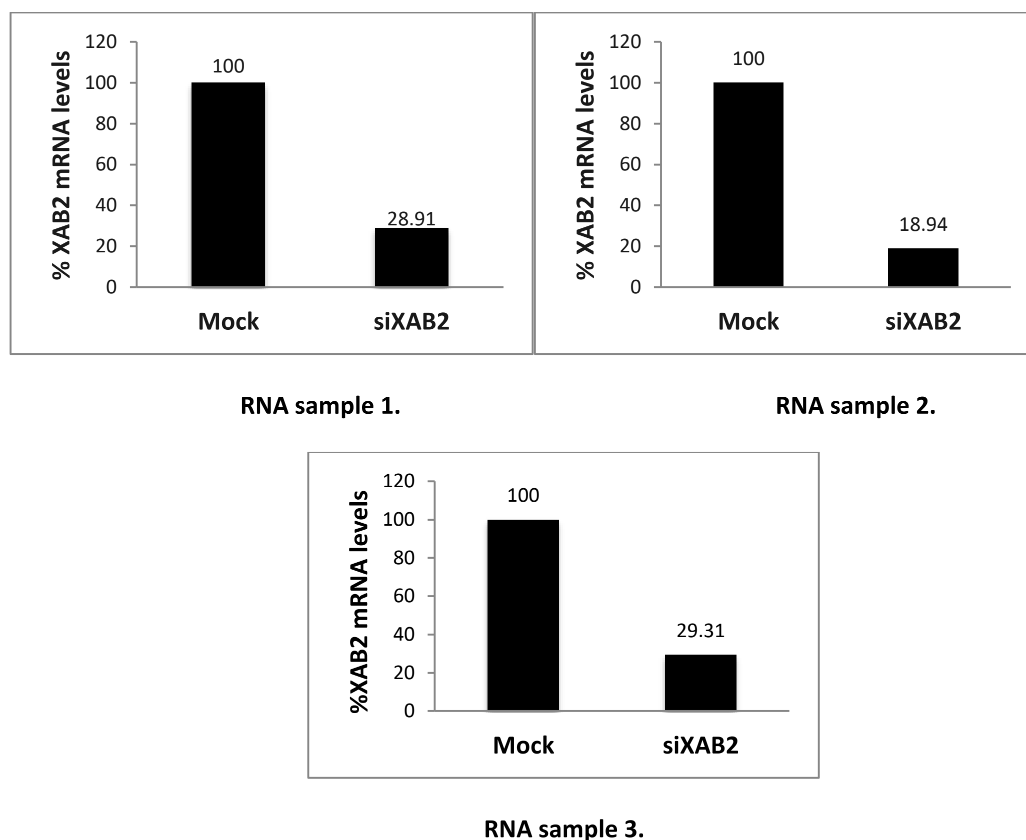


Figure 20. γH2AX staining shows no damage site enrichment and LaminB1 staining shows no changes in the nuclear envelope caused by the XAB2 knock-down in siXAB2 cells 48h post-transfection compared to control.

4.5 RNA-seq analysis reveals probable XAB2 roles

After establishing a protocol to efficiently knock-down XAB2, we proceeded to identify genes that are differentially expressed and/or spliced due to the XAB2 knock-down. For this purpose, three RNA samples (48h post-transfection), from triplicate biological siRNA experiments in 6-cm plate formats of JM8A3.N1 cells, were used for transcriptome analysis as described in the Materials and Methods section. Cells that were transfected with lipofectamine (mock) were used as a control. **(Figure 21.)**.



RNA sample 3.

Figure 21. RT-qPCR analysis shows the decrease of XAB2 RNA levels 48h post-transfection, in the three RNA samples that were used for transcriptome analysis.

The Cufflings algorithm, that was used in order to identify differentially expressed genes between the mock and the siRNA samples, found 42 up-regulated genes and 292 down-regulated genes.

As far as the up-regulated genes are concerned, STRING identified 16 out of the 42 hits but none of these proteins have been found to correlate with its other, in any predicted or experimental way, when using the highest confidence (0.9) as a minimum required interaction score.

As far as the down-regulated genes are concerned, STRING identified 138 annotated genes. Their interaction network, using the highest confidence (0.9) as a minimum

required interaction score, is depicted in **(Figure 22.)**. Specifically, purple strings depict interaction records containing two or more proteins which have been experimentally shown to bind or to be in the same protein complex. Blue strings show annotated proteins that have been shown to be involved in the same pathway, protein complex or biological process whereas black strings, present the co-expression of proteins which can, due to that, be predicted to functionally associate.

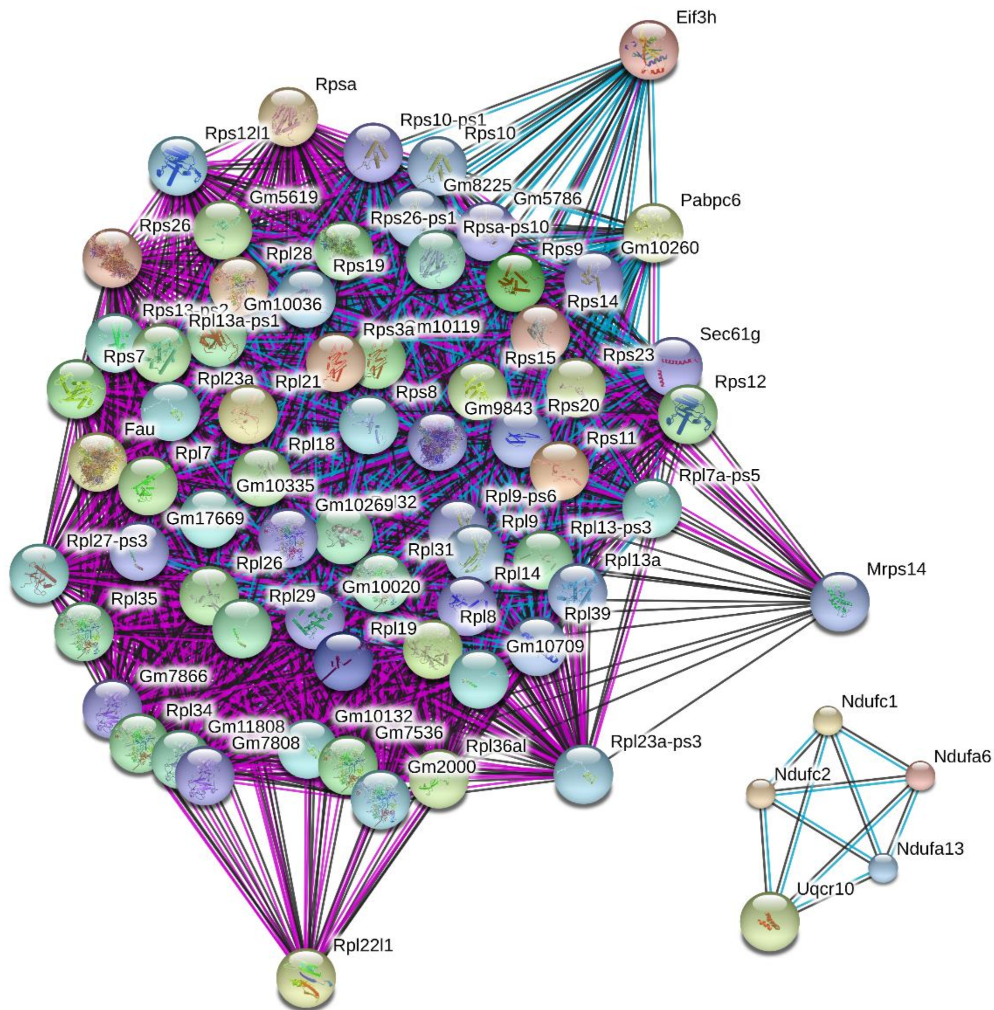


Figure 22. Down-regulated genes functional interaction network (STRING).

Gene ontology analysis, using the geneontology.org application for the 291 down-regulated genes, identified the biological processes to which they are associated. Specifically, there are 103 down-regulated genes which include 29 ribosomal genes and 74 pseudogenes which reside within ribosomal genes (filtered in excel using "begins with rp" e.g ribosomal protein). The 29 identified protein coding genes are associated primarily (p-value) with metabolic and biosynthetic processes i.e translation, peptide biosynthetic process, amide biosynthetic process and peptide metabolic process **(Figure 23)**. Moreover, there are 120 down-regulated genes (filtered in excel using "begins with gm") which include 11 predicted genes (9 of which are members of the ribosomal protein class) and the rest are pseudogenes (109). These are primarily (p-value) involved in metabolic processes i.e biosynthetic process (GO:0009058), nitrogen compound metabolic process (GO:0006807) and primary metabolic process (GO:0044238) including proteolysis (GO: 0006508), translation (GO:0006412) and rRNA metabolic processes (GO:0016072) **(data not shown)**.

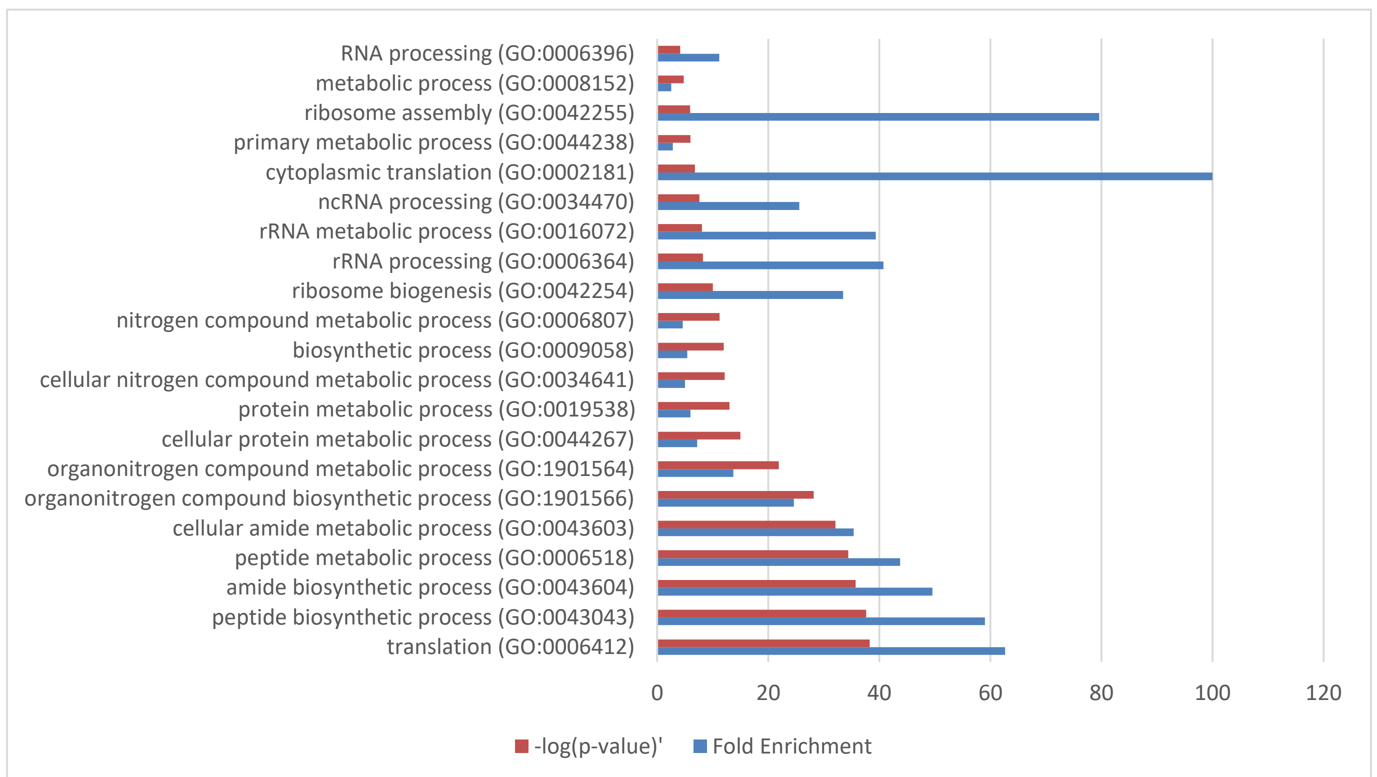


Figure 23. Gene ontology analysis presents the overrepresented biological processes that the 103 (out of 291) down-regulated genes are involved in.

As far as the enriched pathways that the down-regulated genes participate in are concerned, the ribosome and mitochondrion (oxidative phosphorylation) seem to be the ones affected by the XAB2 knock-down **Figure 24**.

KEGG Pathways			
pathway ID	*pathway description*	*count in gene set*	*false discovery rate*
03010	Ribosome	37	5.39e-51
00190	Oxidative phosphorylation	6	0.02

Figure 24. Overrepresented pathways that down-regulated proteins are part of.

4.6 Validation of XAB2 protein interactors

XAB2 has been shown to participate in NER through its interaction with XPA, RNApolII and CSA, CSB **(Nakatsu et al., 2000)**. However, **Onyango et al., 2016** was not able to identify any NER factors in their mass spectrometry analysis, using α-FLAG antibodies from cells being treated with camptothesin and immunoblotting failed to detect CSB as an XAB2 interactor. Investigating XAB2 protein interactors could provide insight in the role and function of XAB2 and this is the reason why we sought out to explore these interactions in our system.

Previous mass spectrometry analysis of XAB2 immunoprecipitation assays in JM8A3.N1 mESCs in our lab, showed no NER factor interacting with XAB2, similar to **Onyango et al., 2016.** Further immunoprecipitation assays were conducted using α-XAB2 in MEFs nuclear extracts with and without UV irradiation (UVC 20J/m^2, 2h recovery). Moreover, the same experiment was conducted with whole cells extracts from MEFs without any treatment. Both assays were once again unable to prove that RNApolII and XPA interact with XAB2 through immunoblotting, whereas enrichment of XAB2 in the IP was confirmed **(Figure 26.)**.

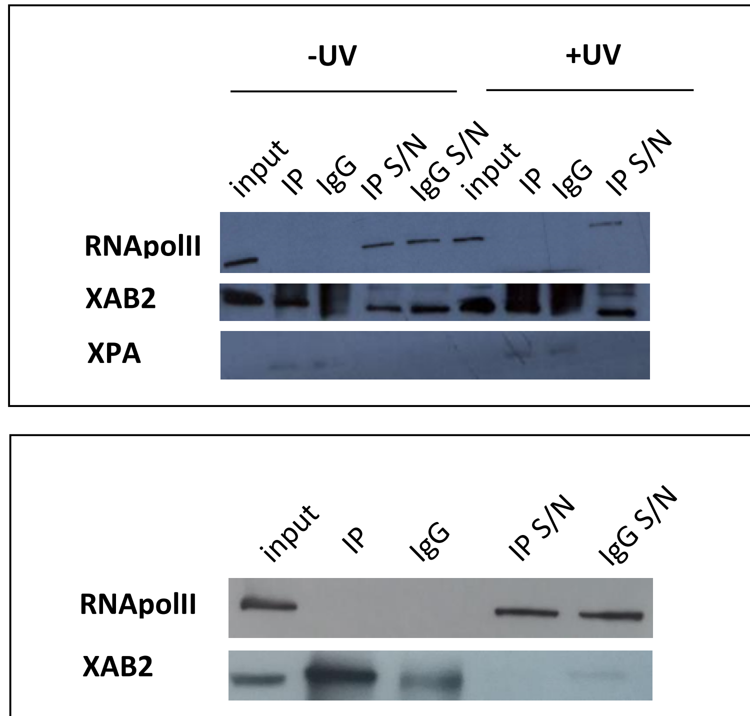


Figure 26. IP of XAB2 from MEFs (+/- UV) nuclear extracts (upper panel) and MEFs (without treatment) whole cell extracts (lower panel) shows no interaction of XAB2 with RNApolII and XPA whereas XAB2 enrichment in the IPs is prominent.

5. Discussion

Previous studies have shown that XAB2 is a multi-functional protein, with a lethal phenotype upon its absence **(Yonemasu et al., 2005)** which also causes cell death in several knock-down cell lines **(Yonemasu et al., 2008; Onyango et al., 2016)**. It participates in transcription, two DNA repair pathways, TC-NER **(Nakatsu et al., 2000)** and HR **(Onyango et al., 2016)** and pre-mRNA splicing **(Kuraoka et al., 2008)**.

Concerning its role in NER and transcription, it has been shown that XAB2 interacts with NER factors XPA, CSA and CSB. However, **Onyango et al., 2016** were not able to reproduce XAB2 interactions with NER factors in mass spectrometry analysis of U2OS cells treated with Cpt. The same case was when immunoblotting for each antibody separately from the IP experiment. We were also not able to show such interactions in IP of nuclear and whole cell extracts of non UV, UV treated and untreated MEF cells. In this way, it would be logical to assume that this interaction may be transient and was lost during the 2h recovery of UV irradiated cells. Moreover, XAB2 has been implicated to have a role in transcription due to the RNAPII interaction and the non-rescue of RNA synthesis in normal UV irradiated cells, microinjected with anti-XAB2 antisera **(Nakatsu et al., 2000)**. However, this interaction could be due to its role in pre-mRNA splicing as transcription and RNA splicing are tightly linked.

It has been shown that depletion of pre-mRNA splicing factors causes DNA damage. XAB2 was recognized as such a mediator of DNA damage by **Miller et al., 2009** in a siRNA screen of factors probably involved in genome instability. However, neither **Onyango et al., 2016** nor we, in our siRNA for XAB2 experiments in 48h post-transfected MEFs, were able to detect increase of γH2AX foci in the knock-down cells compared to control. However, 24h and 72h post-transfection stainings should be conducted, both in MEFs and mESCs, in order to further study this phenotype. Moreover, it is also known that the deletion of RNA binding factors leads to the accumulation of R-loops which activate the ATM pathway and lead to DSBs **(Li and Manley et al., 2006; Aguilera and Garcia-Muse, 2012)**. For this reason, we decided to stain siXAB2 MEFs and JM8A3.N1 with a specific for RNA-DNA hybrids antibody, S9.6, to evaluate whether the concentration of hybrids is altered in the absence of XAB2. To our surprise, RNA-DNA hybrids were clearly decreased. The staining produced a faint, dispersed signal in the nucleoli of 48h post-transfected MEFs and 72h post-transfected JM8A3.N1 mESCs. XAB2 RNA levels were ~80% lower in both cell types. Nucleoli are nuclear compartments where rRNA transcription takes place, however, XAB2 has not been detected in the nucleoli of cells, proposing an indirect effect. The phenotype of S9.6 was rescued by the overexpression of XAB2 in MEFs, suggesting that the decrease of XAB2 protein levels is the reason of R-loop absence and that transcription taking place in the nucleolus which deals with rRNA is less active in the

knock-down cells. Moreover, nucleolin, a protein required for rRNA transcription, is more dispersed three days post-transfection in the siXAB2 cells, proposing nucleolar disruption of knock-down JM8A3.N1 cells. However, it should be noted that because of the observed cell death exhibited in siXAB2 cells, we cannot assume whether the aberrant nucleolar staining (and possibly rRNA transcription defect) is related to cell death or XAB2 function.

In order to study whether the XAB2 knock-down affects transcription in general, we chose to check for RNAPII expression in our knocked-down cells. We observed slight decrease of RNAPII 48h post transfection in siXAB2 MEFs. This decrease was also seen in 24h/72h post-transfection siXAB2 JM8A3.N1 mESCs in immunoblotting and 72h post transfection in staining of bXAB2;birA mESCs. This phenotype was rescued in pcDNA XAB2 and pcDNA + siXAB2 transfected MEFs.

An interesting aspect towards the participation of XAB2 in the core pre-mRNA spliceosome is that similar experiments conducted for AQR (Aquarius), interactor of XAB2, have produced opposite results. siRNA experiments for AQR increase the R-loop formation, a phenotype rescued with RNase H1 treatment **(Sollier et al., 2014)**. The reason why AQR siRNA produces different results than XAB2 in the context of R-loops, even though they have been proposed to be part of the same complex, surely relies on their properties. However, both XAB2 and AQR are fundamental for life, as their knock-out is embryonic lethal.

IBP160 (hAquarius), has been found to couple snoRNP assembly with intron excision **(Hirose et al., 2006)**. SnoRNAs, that are needed for guiding chemical modifications in rRNAs transcribed in the nucleoli of eukaryotic cells, are transcribed from sequences within the introns of several protein-coding and non-coding snoRNA host genes. Therefore, they are released from excised and debranched introns and processed by exonucleolytic trimming. The coding region of these introns associate with snoRNPs (the snRNP complexes of the nucleoli) and their correct assembly is fundamental for their processing. For this reason, it could be possible, that XAB2 which is involved in the intron degradation pathway **(Masaki et al., 2015)**, binds near snoRNA sequences (in Complex C of pre-mRNA splicing reaction) assisting its interactor, Aquarius, to properly recruit snoRNP proteins during pre-mRNA splicing.

Several components of pre-mRNA splicing have lethal phenotypes in their knockout models (PRP19, BRCA1, hAquarius etc.) and these factors have also been implicated to participate in DNA repair. In the same way, XAB2 may function mainly as a pre-mRNA splicing factor but be recruited at the sites of damage when in need, either to be post-translationally modified in order to alter pre-mRNA splicing decisions or participate in DNA damage repair as it has been already proposed for TC-NER. On the other hand, it is worth mentioning that CSB, a NER factor, has been shown to form a complex with RNAPI, TFIIH and XPG in nucleoli in order to stabilize this complex for

efficient rRNA synthesis **(Bradsher et al., 2002)**. This suggests that previously described proteins with DNA repair activities could be associated with rRNA biogenesis.

Moreover, RNA-seq data from our siRNA experiments present a strong and targeted down-regulation of ribosomal protein genes, which could possibly explain the aberrant nucleolar phenotype of the siXAB2 cells, whereas no correlation with DNA repair genes was prominent. This proposes that XAB2 may specifically be responsible for the correct splicing of ribosomal genes, the missplicing of which, could lead to its pre-embryonic lethal phenotype. The analysis of the differentially spliced genes (4.000 transcripts) will shed more light on the XAB2 pre-mRNA splicing specificity.

Our data propose that XAB2 may be involved in ribosomal biogenesis and its knock-down affects vital cellular processes, (as cells die in XAB2 absence) which could be transcription and rRNA biogenesis. Future experiments are needed to further elucidate its role, as due to the cell death observed in the knock-down cells, a phenotype rising due to it (i.e end of ribosome production), could be misled with a direct XAB2 role. However, the fact that the RNA-seq results are highly targeted to the ribosomal group of genes, make us believe that XAB2 is most probably involved in their biogenesis. Firstly, γH2AX, S9.6 and RNApolII stainings need to be conducted in NER factors knock-out mESCs and MEFs, in order to study whether our data are related with other NER factor ablations and damage. Secondly, post-translational modifications of XAB2 need to be checked, in order to find out whether this is the reason for its recruitment in sites of damage **(Onyango et al., 2016)** and TC-NER functions **(Nakatsu et al., 2000)**. Moreover, mass spectrometry analysis, immunoprecipitation and pull-down assays in bioXAB2;birA and mouse livers would verify the complexes in which it participates and shed light into its pre-mRNA splicing of ribosomal genes properties, in embryonic and developmental states. Lastly, RNA immunoprecipitation assays which are already been conducted, will suggest whether XAB2 is involved in snoRNA biogenesis and snoRNP recruitment.

6. References

- Aguilera, A., & García-Muse, T. (2012). R loops: from transcription byproducts to threats to genome stability. Molecular cell, 46(2), 115-124.
- Anantha, R. W., Alcivar, A. L., Ma, J., Cai, H., Simhadri, S., Ule, J., König J., & Xia, B. (2013). Requirement of heterogeneous nuclear ribonucleoprotein C for BRCA gene expression and homologous recombination. PloS one, 8(4), e61368.
- Araújo, S. J., Tirode, F., Coin, F., Pospiech, H., Syväoja, J. E., Stucki, M., Hübscher, U., Egly J. M., & 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 & development, 14(3), 349-359.
- Bai, B., Hales, C. M., Chen, P. C., Gozal, Y., Dammer, E. B., Fritz, J. J., Wang, X., Xia, Q., Duong, D. M., Street, C., Cantero, G., Cheng, D., Jones, D. R., Wu, Z., Li, Y., Diner, I., Heilman, C. J., Rees, H. D., Wu, H., Lin, L., Szulwach, K. E., Gearing, M., Mufson, E. J., Bennett, D. A., Montine, T. J., Seyfried, N. T., Wingo, T. S., Sun, Y. E., Jin, P., Hanfelt, J., Willcock, D. M., Levey, A., Lah, J. J., & Peng J. (2013). U1 small nuclear ribonucleoprotein complex and RNA splicing alterations in Alzheimer's disease. Proceedings of the National Academy of Sciences, 110(41), 16562-16567.
- Bansbach, C. E., & Cortez, D. (2011). Defining genome maintenance pathways using functional genomic approaches. Critical reviews in biochemistry and molecular biology, 46(4), 327-341.
- Barash, Y., Calarco, J. A., Gao, W., Pan, Q., Wang, X., Shai, O., Blencowe, B. J., & Frey, B. J. (2010). Deciphering the splicing code. Nature, 465(7294), 53-59.
- Barbosa-Morais, N. L., Irimia, M., Pan, Q., Xiong, H. Y., Gueroussov, S., Lee, L. J., Slobodeniuc, V., Kutter, C., Watt, S., Çolak, R., Kim, T., Misquitta-Ali, C. M., Wilson, M. D., Kim, P. M., Odom, D. T., Frey, B. J., & Blencowe, B. J. (2012). The evolutionary landscape of alternative splicing in vertebrate species. Science, 338(6114), 1587-1593.
- Bell, S. P., & Dutta, A. (2002). DNA replication in eukaryotic cells. Annual review of biochemistry, 71(1), 333-374.
- Bennetzen, M., Larsen, D., Dinant, C., Watanabe, S., Bartek, J., Lukas, J., & Andersen, J. S. (2013). Acetylation dynamics of human nuclear proteins during the ionizing radiation-induced DNA damage response. Cell cycle, 12(11), 1688-1695.
- Bergink, S., Toussaint, W., Luijsterburg, M. S., Dinant, C., Alekseev, S., Hoeijmakers, J. H., Dantuma, N. P., Houtsmuller, A. B., & Vermeulen, W. (2012). Recognition of DNA damage by XPC coincides with disruption of the XPC–RAD23 complex. The Journal of cell biology, 196(6), 681-688.

- Bradsher, J., Auriol, J., de Santis, L. P., Iben, S., Vonesch, J. L., Grummt, I., & Egly, J. M. (2002). CSB is a component of RNA pol I transcription. Molecular cell, 10(4), 819-829.
- Campos, E. I., Reinberg, D. (2009). Histones: annotating chromatin. Annual review of genetics, 43(1), 559–99.
- Chakraborty, P., & Grosse, F. (2011). Human DHX9 helicase preferentially unwinds RNA-containing displacement loops (R-loops) and G-quadruplexes. DNA repair, 10(6), 654-665.
- Chanarat, S., Burkert-Kautzsch, C., Meinel, D. M., & Sträßer, K. (2012). Prp19C and TREX: interacting to promote transcription elongation and mRNA export. Transcription, 3(1), 8-12.
- Chavanne, F., Broughton, B. C., Pietra, D., Nardo, T., Browitt, A., Lehmann, A. R., & Stefanini, M. (2000). Mutations in the XPC gene in families with xeroderma pigmentosum and consequences at the cell, protein, and transcript levels. Cancer Research, 60(7), 1974-1982.
- Ciccia, A., & Elledge, S. J. (2010). The DNA damage response: making it safe to play with knives. Molecular cell, 40(2), 179-204.
- Cooper, T. A., Wan, L., & Dreyfuss, G. (2009). RNA and disease. Cell, 136(4), 777-793.
- Czechanski, A., Byers, C., Greenstein, I., Schrode, N., Donahue, L. R., Hadjantonakis, A. K., & Reinholdt, L. G. (2014). Derivation and characterization of mouse embryonic stem cells from permissive and nonpermissive strains. Nature protocols, 9(3), 559-574.
- Daguenet, E., Dujardin, G., & Valcárcel, J. (2015). The pathogenicity of splicing defects: mechanistic insights into pre-mRNA processing inform novel therapeutic approaches. EMBO reports, e201541116.
- De Fougerolles, A. R. (2008). Delivery vehicles for small interfering RNA in vivo. Human gene therapy, 19(2), 125-132.
- DiGiovanna, J. J., & Kraemer, K. H. (2012). Shining a light on xeroderma pigmentosum. Journal of Investigative Dermatology, 132, 785-796.
- Fagbemi, A. F., Orelli, B., & Schärer, O. D. (2011). Regulation of endonuclease activity in human nucleotide excision repair. DNA repair, 10(7), 722-729.
- Farmer, H., McCabe, N., Lord, C. J., Tutt, A. N., Johnson, D. A., Richardson, T. B., Santarosa, M., Dillon, K. J., Hickson, I., Knights, C., Martin, N. M. B., Jackson, S. P., Smith, G. C. M., & Ashworth A. (2005). Targeting the DNA repair defect in BRCA mutant cells as a therapeutic strategy. Nature, 434(7035), 917-921.
- Fica, S. M., Tuttle, N., Novak, T., Li, N. S., Lu, J., Koodathingal, P., Dai, Q., Staley J. P. & Piccirilli, J. A. (2013). RNA catalyses nuclear pre-mRNA splicing. Nature, 503(7475), 229-234.

- Fousteri, M., Vermeulen, W., van Zeeland, A. A., & 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. Molecular cell, 23(4), 471-482.
- Friedberg, E. C. (2003). DNA damage and repair. Nature, 421(6921), 436-440.
- Friedberg, E. C., Walker, G. C., Siede, W., & Wood, R. D. (Eds.). (2005). DNA repair and mutagenesis. American Society for Microbiology Press.
- Fromont-Racine, M., Rain, J. C., & Legrain, P. (1997). Toward a functional analysis of the yeast genome through exhaustive two-hybrid screens. Nature genetics, 16(3), 277-282.
- Fu, X. D., & Ares Jr, M. (2014). Context-dependent control of alternative splicing by RNA-binding proteins. Nature Reviews Genetics, 15(10), 689-701.
- Gan, W., Guan, Z., Liu, J., Gui, T., Shen, K., Manley, J. L., & Li, X. (2011). R-loop-mediated genomic instability is caused by impairment of replication fork progression. Genes & development, 25(19), 2041-2056.
- Gillet, L. C., & Schärer, O. D. (2006). Molecular mechanisms of mammalian global genome nucleotide excision repair. Chemical reviews, 106(2), 253-276.
- Ginno, P. A., Lott, P. L., Christensen, H. C., Korf, I., & Chédin, F. (2012). R-loop formation is a distinctive characteristic of unmethylated human CpG island promoters. Molecular cell, 45(6), 814-825.
- Gnad, F., Baucom, A., Mukhyala, K., Manning, G., & Zhang, Z. (2013). Assessment of computational methods for predicting the effects of missense mutations in human cancers. BMC genomics, 14(3), 1.
- Groisman, R., Polanowska, J., Kuraoka, I., Sawada, J. I., Saijo, M., Drapkin, R., Kisselev, A. F., Tanaka, K., & 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(3), 357-367.
- Grote, M., Wolf, E., Will, C. L., Lemm, I., Agafonov, D. E., Schomburg, A., Fischle, W., Urlaub, H., & Lührmann, R. (2010). Molecular architecture of the human Prp19/CDC5L complex. Molecular and cellular biology, 30(9), 2105-2119.
- Hamperl, S., & Cimprich, K. A. (2014). The contribution of co-transcriptional RNA: DNA hybrid structures to DNA damage and genome instability. DNA repair, 19, 84-94.
- Hanawalt, P. C. (2002). Subpathways of nucleotide excision repair and their regulation. Oncogene, 21(58), 8949-8956.
- Harper, J. W., & Elledge, S. J. (2007). The DNA damage response: ten years after. Molecular cell, 28(5), 739-745.
- Hasan, S., Hassa, P. O., Imhof, R., & Hottiger, M. O. (2001). Transcription coactivator p300 binds PCNA and may have a role in DNA repair synthesis. Nature, 410(6826), 387-391.

- Hess, M. T., Schwitter, U., Petretta, M., Giese, B., & Naegeli, H. (1997). Bipartite substrate discrimination by human nucleotide excision repair. Proceedings of the National Academy of Sciences, 94(13), 6664-6669.
- Hirose, T., Ideue, T., Nagai, M., Hagiwara, M., Shu, M. D., & Steitz, J. A. (2006). A spliceosomal intron binding protein, IBP160, links position-dependent assembly of intron-encoded box C/D snoRNP to pre-mRNA splicing. Molecular cell, 23(5), 673-684.
- Hoeijmakers, J. H. (2001). Genome maintenance mechanisms for preventing cancer. nature, 411(6835), 366-374.
- Hofmann, J. C., Husedzinovic, A., & Gruss, O. J. (2010). The function of spliceosome components in open mitosis. Nucleus, 1(6), 447-459.
- Ip, J. Y., Schmidt, D., Pan, Q., Ramani, A. K., Fraser, A. G., Odom, D. T., & Blencowe, B. J. (2011). Global impact of RNA polymerase II elongation inhibition on alternative splicing regulation. Genome research, 21(3), 390-401.
- Ji, Y., & Tulin, A. V. (2009). Poly (ADP-ribosyl) ation of heterogeneous nuclear ribonucleoproteins modulates splicing. Nucleic acids research, gkp218.
- Jurica, M. S., & Moore, M. J. (2003). Pre-mRNA splicing: awash in a sea of proteins. Molecular cell, 12(1), 5-14.
- Kamileri, I., Karakasilioti, I., & Garinis, G. A. (2012). Nucleotide excision repair: new tricks with old bricks. Trends in genetics, 28(11), 566-573.
- Kuraoka, I., Ito, S., Wada, T., Hayashida, M., Lee, L., Saijo, M., Nakatsu, Y., Matsumoto, M., Matsunaga, T., Handa, H., Qin, J., Nakatani, Y., Tanaka, K. (2008). Isolation of XAB2 complex involved in pre-mRNA splicing, transcription, and transcription-coupled repair. Journal of Biological Chemistry, 283(2), 940-950.
- Lagerwerf, S., Vrouwe, M. G., Overmeer, R. M., Fousteri, M. I., & Mullenders, L. H. (2011). DNA damage response and transcription. DNA repair, 10(7), 743-750.
- Lainé, J. P., & Egly, J. M. (2006). Initiation of DNA repair mediated by a stalled RNA polymerase IIO. The EMBO journal, 25(2), 387-397.
- Lamb, J. R., Tugendreich, S., & Hieter, P. (1995). Tetratrico peptide repeat interactions: to TPR or not to TPR?. Trends in biochemical sciences, 20(7), 257-259.
- Lee, Y., & Rio, D. C. (2015). Mechanisms and regulation of alternative pre-mRNA splicing. Annual review of biochemistry, 84, 291.
- Lehmann, A. R. (2003). DNA repair-deficient diseases, xeroderma pigmentosum, Cockayne syndrome and trichothiodystrophy. Biochimie, 85(11), 1101-1111.
- Li, X., & Manley, J. L. (2005). Inactivation of the SR protein splicing factor ASF/SF2 results in genomic instability. Cell, 122(3), 365-378.

- Li, X., & Manley, J. L. (2006). Cotranscriptional processes and their influence on genome stability. Genes & development, 20(14), 1838-1847.
- Lin, Y. L., & Pasero, P. (2014). Caught in the Act: R-loops are cleaved by structure-specific endonucleases to generate DSBs. Molecular cell, 56(6), 721-722.
- Liquori, C. L., Ricker, K., Moseley, M. L., Jacobsen, J. F., Kress, W., Naylor, S. L., Day, J. W., & Ranum, L. P. (2001). Myotonic dystrophy type 2 caused by a CCTG expansion in intron 1 of ZNF9. Science, 293(5531), 864-867.
- Lukas, J., Lukas, C., & Bartek, J. (2011). More than just a focus: The chromatin response to DNA damage and its role in genome integrity maintenance. Nature cell biology, 13(10), 1161-1169.
- Ma, Y., Jin, J., Dong, C., Cheng, E. C., Lin, H., Huang, Y., & Qiu, C. (2010). High-efficiency siRNA-based gene knockdown in human embryonic stem cells. RNA, 16(12), 2564-2569.
- MacDonald, M. E., Ambrose, C. M., Duyao, M. P., Myers, R. H., Lin, C., Srinidhi, L., Barnes, G., Taylor, S. A., James, M., Groot, N., MacFarlane , H., Jenkins, B., Anderson, M. A., Wexler, N. S., & Gusella, J. F. (1993). A novel gene containing a trinucleotide repeat that is expanded and unstable on Huntington's disease chromosomes. Cell, 72(6), 971-983.
- Makarova, O. V., Makarov, E. M., Urlaub, H., Will, C. L., Gentzel, M., Wilm, M., & Lührmann, R. (2004). A subset of human 35S U5 proteins, including Prp19, function prior to catalytic step 1 of splicing. The EMBO journal, 23(12), 2381-2391.
- Masaki, S., Yoshimoto, R., Kaida, D., Hata, A., Satoh, T., Ohno, M., & Kataoka, N. (2015). Identification of the specific interactors of the human lariat RNA debranching enzyme 1 protein. International journal of molecular sciences, 16(2), 3705-3721.
- Masutani, C., Sugasawa, K., Yanagisawa, J., Sonoyama, T., Ui, M., Enomoto, T., Takio, K., Tanaka, K., van der Spek, P. J., & Bootsma, D. (1994). Purification and cloning of a nucleotide excision repair complex involving the xeroderma pigmentosum group C protein and a human homologue of yeast RAD23. The EMBO Journal, 13(8), 1831.
- Matera, A. G., & Wang, Z. (2014). A day in the life of the spliceosome. Nature reviews Molecular cell biology, 15(2), 108-121.
- Mathieu, N., Kaczmarek, N., Rüthemann, P., Luch, A., & Naegeli, H. (2013). DNA quality control by a lesion sensor pocket of the xeroderma pigmentosum group D helicase subunit of TFIIH. Current Biology, 23(3), 204-212.
- Mende, Y., Jakubik, M., Riessland, M., Schoenen, F., Roßbach, K., Kleinridders, A., Köhler, C., Buch, T., & Wirth, B. (2010). Deficiency of the splicing factor Sfrs10 results in early embryonic lethality in mice and has no impact on full-length SMN/Smn splicing. Human molecular genetics, 19(11), 2154-2167.

- Merkin, J., Russell, C., Chen, P., & Burge, C. B. (2012). Evolutionary dynamics of gene and isoform regulation in Mammalian tissues. Science, 338(6114), 1593-1599.
- Mischo, H. E., Gómez-González, B., Grzechnik, P., Rondón, A. G., Wei, W., Steinmetz, L., Aguilera, A., & Proudfoot, N. J. (2011). Yeast Sen1 helicase protects the genome from transcription-associated instability. Molecular cell, 41(1), 21-32.
- Mocquet, V., Lainé, J. P., Riedl, T., Yajin, Z., Lee, M. Y., & Egly, J. M. (2008). Sequential recruitment of the repair factors during NER: the role of XPG in initiating the resynthesis step. The EMBO journal, 27(1), 155-167.
- Moser, J., Kool, H., Giakzidis, I., Caldecott, K., Mullenders, L. H., & Fousteri, M. I. (2007). Sealing of chromosomal DNA nicks during nucleotide excision repair requires XRCC1 and DNA ligase IIIα in a cell-cycle-specific manner. Molecular cell, 27(2), 311-323.
- Muñoz, M. J., Santangelo, M. S. P., Paronetto, M. P., de la Mata, M., Pelisch, F., Boireau, S., Glover-Cutter, K., Ben-Dov, C., Blaustein, M., Lozano, J. J., Bird, G., Bentley, D., Bertrand, E., & Kornblihtt, A. R. (2009). DNA damage regulates alternative splicing through inhibition of RNA polymerase II elongation. Cell, 137(4), 708-720.
- Nakatsu, Y., Asahina, H., Citterio, E., Rademakers, S., Vermeulen, W., Kamiuchi, S., Yeo, J., Khaw, M., Saijo, M., Kodo, M., Matsuda, T., Hoeijmakers, J. H. J., & Tanaka, K. (2000). XAB2, a novel tetratricopeptide repeat protein involved in transcription-coupled DNA repair and transcription. Journal of Biological Chemistry, 275(45), 34931-34937.
- Nasmyth, K. (2001). Disseminating the genome: joining, resolving, and separating sister chromatids during mitosis and meiosis. Annual review of genetics, 35(1), 673-745.
- Neumann, B., Held, M., Liebel, U., Erfle, H., Rogers, P., Pepperkok, R., & Ellenberg, J. (2006). High-throughput RNAi screening by time-lapse imaging of live human cells. Nature methods, 3(5), 385-390.
- Nishi, R., Okuda, Y., Watanabe, E., Mori, T., Iwai, S., Masutani, C., Sugasawa, K., & Hanaoka, F. (2005). Centrin 2 stimulates nucleotide excision repair by interacting with xeroderma pigmentosum group C protein. Molecular and Cellular Biology, 25(13), 5664-5674.
- Ohnuma-Ishikawa, K., Morio, T., Yamada, T., Sugawara, Y., Ono, M., Nagasawa, M., Yasuda, A., Morimoto, C., Ohnuma, K., Dang, N. H., Hosoi, H., Verdin, E., & Mizutani, S. (2007). Knockdown of XAB2 Enhances All-Trans Retinoic Acid–Induced Cellular Differentiation in All-Trans Retinoic Acid–Sensitive and–Resistant Cancer Cells. Cancer research, 67(3), 1019-1029.

- Oksenych, V., De Jesus, B. B., Zhovmer, A., Egly, J. M., & Coin, F. (2009). Molecular insights into the recruitment of TFIIH to sites of DNA damage. The EMBO journal, 28(19), 2971-2980.
- Onyango, D. O., Howard, S. M., Neherin, K., Yanez, D. A., & Stark, J. M. (2016). Tetratricopeptide repeat factor XAB2 mediates the end resection step of homologous recombination. Nucleic acids research, gkw275.
- Overmeer, R. M., Moser, J., Volker, M., Kool, H., Tomkinson, A. E., van Zeeland, A. A., Mullenders, L. H. F., & Fousteri, M. (2011). Replication protein A safeguards genome integrity by controlling NER incision events. The Journal of cell biology, 192(3), 401-415.
- Paulsen, R. D., Soni, D. V., Wollman, R., Hahn, A. T., Yee, M. C., Guan, A., Hesley, J. A., Miller, S. C., Cromwell, E. F., Solow-Cordero, D. E., Meyer, T., & Cimprich, K. A. (2009). A genome-wide siRNA screen reveals diverse cellular processes and pathways that mediate genome stability. Molecular cell, 35(2), 228-239.
- Pettitt, S. J., Liang, Q., Rairdan, X. Y., Moran, J. L., Prosser, H. M., Beier, D. R., Lloyd, K. C., Bradley, A., & Skarnes, W. C. (2009). Agouti C57BL/6N embryonic stem cells for mouse genetic resources. Nature methods, 6(7), 493.
- Polo, S. E., & Jackson, S. P. (2011). Dynamics of DNA damage response proteins at DNA breaks: a focus on protein modifications. Genes & development, 25(5), 409-433.
- Richardson, J. P. (2003). Loading Rho to terminate transcription. Cell, 114(2), 157-159.
- Roy, D., Yu, K., & Lieber, M. R. (2008). Mechanism of R-loop formation at immunoglobulin class switch sequences. Molecular and cellular biology, 28(1), 50-60.
- Savage, K. I., Gorski, J. J., Barros, E. M., Irwin, G. W., Manti, L., Powell, A. J., Pellagatti, A., Lukashchuk, N., McCance, D. J., McCluggage, W. G., Schettino, G., Salto-Tellez, M., Boultwood, J., Richard, D. J., McDade1, S. S., & Harkin, D. P. (2014). Identification of a BRCA1-mRNA splicing complex required for efficient DNA repair and maintenance of genomic stability. Molecular cell, 54(3), 445-459.
- Savage, K. I., & Harkin, D. P. (2015). BRCA1, a ‘complex’ protein involved in the maintenance of genomic stability. FEBS journal, 282(4), 630-646.
- Shaban, N. M., Harvey, S., Perrino, F. W., & Hollis, T. (2010). The structure of the mammalian RNase H2 complex provides insight into RNA· DNA hybrid processing to prevent immune dysfunction. Journal of Biological Chemistry, 285(6), 3617-3624.
- Shanbhag, N. M., Rafalska-Metcalf, I. U., Balane-Bolivar, C., Janicki, S. M., & Greenberg, R. A. (2010). ATM-dependent chromatin changes silence transcription in cis to DNA double-strand breaks. Cell, 141(6), 970-981.

- Sharp, P. A. (1994). Split genes and RNA splicing (Nobel lecture). Angewandte Chemie International Edition in English, 33(12), 1229-1240.
- Sharp, P. A. (2005). The discovery of split genes and RNA splicing. Trends in biochemical sciences, 30(6), 279-280.
- Shkreta, L., & Chabot, B. (2015). The RNA splicing response to DNA damage. Biomolecules, 5(4), 2935-2977.
- Singh, M., Shin, Y. K., Yang, X., Zehr, B., Chakrabarti, P., & Kandror, K. V. (2015). 4E-BPs control fat storage by regulating the expression of Egr1 and ATGL. Journal of Biological Chemistry, 290(28), 17331-17338.
- Singh, R., & Valcárcel, J. (2005). Building specificity with nonspecific RNA-binding proteins. Nature structural & molecular biology, 12(8), 645-653.
- Skourti-Stathaki, K., & Proudfoot, N. J. (2014). A double-edged sword: R loops as threats to genome integrity and powerful regulators of gene expression. Genes & development, 28(13), 1384-1396.
- Smolka, M. B., Albuquerque, C. P., Chen, S. H., & Zhou, H. (2007). Proteome-wide identification of in vivo targets of DNA damage checkpoint kinases. Proceedings of the National Academy of Sciences, 104(25), 10364-10369.
- Sollier, J., Stork, C. T., García-Rubio, M. L., Paulsen, R. D., Aguilera, A., & Cimprich, K. A. (2014). Transcription-coupled nucleotide excision repair factors promote R-loop-induced genome instability. Molecular cell, 56(6), 777-785.
- Sordet, O., Redon, C. E., Guirouilh-Barbat, J., Smith, S., Solier, S., Douarre, C., Conti, C., Nakamura, A. J., Das, B. B., Nicolas, E., Kohn K. W., Bonner, W. M., & Yves Pommier, Y. (2009). Ataxia telangiectasia mutated activation by transcription-and topoisomerase I-induced DNA double-strand breaks. EMBO reports, 10(8), 887-893.
- Storbeck, M., Hupperich, K., Gaspar, J. A., Meganathan, K., Carrera, L. M., Wirth, R., Sachinidis, A., & Wirth, B. (2014). Neuronal-specific deficiency of the splicing factor Tra2b causes apoptosis in neurogenic areas of the developing mouse brain. PloS one, 9(2), e89020.
- Sun, Q., Csorba, T., Skourti-Stathaki, K., Proudfoot, N. J., & Dean, C. (2013). R-loop stabilization represses antisense transcription at the Arabidopsis FLC locus. Science, 340(6132), 619-621.
- Supek, F., Miñana, B., Valcárcel, J., Gabaldón, T., & Lehner, B. (2014). Synonymous mutations frequently act as driver mutations in human cancers. Cell, 156(6), 1324-1335.
- Tresini, M., Warmerdam, D. O., Kolovos, P., Snijder, L., Vrouwe, M. G., Demmers, J. A., van IJcken, W. F. J., Grosveld, F. G., Medema, R. H., Hoeijmakers, J. H. J., Mullenders, L. H. F., Vermeulen, W., & Marteijn, J. A. (2015). The core spliceosome as target and effector of non-canonical ATM signalling. Nature, 523(7558), 53-58.

- Tuduri, S., Crabbé, L., Conti, C., Tourrière, H., Holtgreve-Grez, H., Jauch, A., Pantesco, V., De Vos, J., Thomas, A., Theillet, C., Pommier, Y., Tazi, J., Arnaud Coquelle, A., & Pasero, P., (2009). Topoisomerase I suppresses genomic instability by preventing interference between replication and transcription. Nature cell biology, 11(11), 1315-1324.
- Turunen, J. J., Niemelä, E. H., Verma, B., & Frilander, M. J. (2013). The significant other: splicing by the minor spliceosome. Wiley Interdisciplinary Reviews: RNA, 4(1), 61-76.
- Wahba, L., Gore, S. K., & Koshland, D. (2013). The homologous recombination machinery modulates the formation of RNA–DNA hybrids and associated chromosome instability. Elife, 2, e00505.
- Wahl, M. C., Will, C. L., & Lührmann, R. (2009). The spliceosome: design principles of a dynamic RNP machine. Cell, 136(4), 701-718.
- Wan, Y., Zheng, X., Chen, H., Guo, Y., Jiang, H., He, X., Zhu X., & Zheng, Y. (2015). Splicing function of mitotic regulators links R-loop–mediated DNA damage to tumor cell killing. The Journal of cell biology, 209(2), 235-246.
- Wang, J., Lu, Z., Wientjes, M. G., & Au, J. L. S. (2010). Delivery of siRNA therapeutics: barriers and carriers. The AAPS journal, 12(4), 492-503.
- Watson, J. D., & Crick, F. H. C. (1953). Molecular structure of nucleic acids: A structure for deoxyribose nucleic acid. Nature, 171(737-738), 3-12.
- Westover, K. D., Bushnell, D. A., & Kornberg, R. D. (2004). Structural basis of transcription: nucleotide selection by rotation in the RNA polymerase II active center. Cell, 119(4), 481-489.
- Will, C. L., & Lührmann, R. (2011). Spliceosome structure and function. Cold Spring Harbor perspectives in biology, 3(7), a003707.
- Xu, J. (2005). Preparation, culture, and immortalization of mouse embryonic fibroblasts. Current protocols in molecular biology, 28-1.
- Yonemasu, R., Minami, M., Nakatsu, Y., Takeuchi, M., Kuraoka, I., Matsuda, Y., Higashib, Y., Kondohb, H., & Tanaka, K. (2005). Disruption of mouse XAB2 gene involved in pre-mRNA splicing, transcription and transcription-coupled DNA repair results in preimplantation lethality. DNA repair, 4(4), 479-491.
- Zhang, N., Kaur, R., Lu, X., Shen, X., Li, L., & Legerski, R. J. (2005). The Pso4 mRNA splicing and DNA repair complex interacts with WRN for processing of DNA interstrand cross-links. Journal of Biological Chemistry, 280(49), 40559-40567.
- dharmacon.gelifesciences.com
- www.ebi.ac.uk
- www.ensembl.org
- www.genecards.org
- geneontology.org