The Role of R-loops in the Pathology of Trinucleotide Expansion Diseases

Matthias Groh

St John's College

University of Oxford

A thesis submitted for the degree of Doctor of Philosophy

At the University of Oxford

Michaelmas Term, 2015

Abstract

Friedreich ataxia and fragile X syndrome are among 40 human diseases associated with expansion of repeated DNA sequences. In both disorders repeat expansion leads to gene silencing, the molecular mechanism of which is not well understood. It was proposed that formation of unusual DNA structures such as R-loops over repeat regions may play a role, but their molecular function has not been investigated *in vivo*. R-loops are three-stranded structures, which occur when RNA hybridises to a complementary DNA strand. This leads to formation of an RNA/DNA hybrid and results in displacement of the other DNA strand.

In the first part of this thesis, I show that RNA/DNA hybrids are formed in patient cells on expanded repeats of *FXN* and *FMR1* genes, mutated in Friedreich ataxia and fragile X syndrome. These RNA/DNA hybrids are stable, colocalise with repressive chromatin marks and impede *FXN* gene transcription in patient cells. Furthermore, I studied the relationship between repressive chromatin and RNA/DNA hybrids. I found that increasing RNA/DNA hybrid levels triggers heterochromatin formation and leads to transcriptional repression of the *FXN* gene, providing a direct molecular link between RNA/DNA hybrids and the pathology of expansion diseases.

The current understanding of R-loop biology in health and disease is limited by the small number of proteins known to bind RNA/DNA hybrids or R-loops *in vivo*. In the second part of this thesis I therefore established an affinity purification approach, employing the RNA/DNA hybrid-specific S9.6 antibody followed by mass spectrometry to identify RNA/DNA hybrid-binding proteins in HeLa cells. Using this approach, 469 proteins were identified that constitute the RNA/DNA hybrid interactome by stringent biochemical and statistical criteria. Identified proteins include known R-loop factors senataxin, SRSF1 and topoisomerase I and yet uncharacterised interactors, such as RNA- and DNA-binding proteins, DNA repair and chromatin factors, and helicases. To demonstrate the biological relevance of the RNA/DNA hybrid interactome, I show that the top candidate DHX9 helicase promotes RNA/DNA hybrid resolution *in vivo*. Furthermore, DHX9 and other RNA/DNA hybrid interactome helicases are overexpressed in cancer, revealing a strong link between R-loop-mediated genome instability and human disease.

Für meine Familie.

Acknowledgements

I would first like to sincerely thank my supervisor, Dr Natalia Gromak, for the opportunity to work on this project in her group and for supporting me at every stage of it. She has always been there to teach me all the important details of essential methods and offered excellent advice and guidance during my DPhil from the first day to the last. In addition to the many thought-provoking discussions we have had, she has been a very influential mentor for my own professional development. I am deeply grateful and honoured for having had the chance to pursue a DPhil at the University of Oxford, and in particular with such a highly dedicated supervisor.

I would also like to thank all the past and present members of the Gromak group. It has always been a great joy working together and I am thankful for the very nice atmosphere they have provided. My special thanks go to my bench neighbour Lara Silva Cravo, whose positive attitude made pipetting dozens of qPCR plates a little bit less boring. I would also like to thank Sarah Irmscher and Natalie Braun in particular, who performed countless ChIPs and western blots for some of the more explorative aspects of my two projects during their internships in the Gromak group.

I am very grateful to the scientific support I have received from Prof Dr Nicholas Proudfoot and Prof Dr Chris Norbury and the members of their groups throughout my DPhil in the form of joint group meetings, presentations and discussions over lunch. Special thanks also go to Prof Dr Shabaz Mohammed and Svenja Hester at the Central Proteomics Facility, who were always incredibly helpful and without whom the mass spectrometry project would not have been possible.

I am also grateful for having met all the fantastic people during my time in Oxford, both inside and outside the Sir William Dunn School of Pathology. We have spent an important part of our lives together and we have shared the evenings after good and bad days in the laboratory.

Finally, I would like to thank my parents and my sister for their support, without whom I would not have come this far. My parents in particular have always encouraged me to try and reach for more than the easily-achievable and to commit myself fully to my aims.

Table of Contents

List of Figures

List of Tables

Abbreviations

Listed below are abbreviations used in this thesis. For proteins and genes, recommended symbols and names are used with common alternative names indicated when appropriate. Protein names are from UniProt (http://www.uniprot.org/) and gene names are from the Human Gene Nomenclature Committee (HUGO, http://www.genenames.org) and from the Ensembl genome browser (http://www.ensembl.org/index.html).

Abbreviation	Full name
3'-UTR	3'-untranslated region
5-azadC	5-aza-2'-deoxycytidine
5'-UTR	5'-untranslated region
A	Adenine
A. thaliana	*Arabidopsis thaliana*
Ab	Antibody
ActD	actinomycin D
Ago1/2	Argonaute 1/2
AGS	Aicardi-Goutières syndrome
AID	Activation-induced deaminase
ALS4	Amyotrophic lateral sclerosis type 4
AOA2	Ataxia with oculomotor apraxia type 2
AS	Angelman syndrome
ASO	Antisense oligonucleotide
AT	Ataxia telangiectasia
ATP	Adenosine triphosphate
bp	Base pair
Br	Bromo
BSA	Bovine serum albumin
C	Cytosine
cDNA	Complementary DNA
ChIP	Chromatin immunoprecipitation
CMV	Cytomegalovirus
CPT	Camptothecin
CSR	Class-switch recombination
CTD	Carboxy-terminal domain
CTP	Cytosine triphosphate
DIP	DNA/RNA immunoprecipitation
DMEM	Dulbecco's Modified Eagle's Medium
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
DNase	Deoxyribonuclease
DRIP	DNA/RNA immunoprecipitation

DSB	Double-strand break
ds	Double-stranded
DTT	Dithiothreitol
E. coli	*Escherichia coli*
e.g.	Example given
EBV	Epstein-Barr virus
EDTA	Ethylenediaminetetraacetic acid
EtOH	Ethanol
FA	Fanconi anemia
FM	Full mutation
FMR1	Fragile X mental retardation 1 gene
FRDA	Friedreich ataxia
FXN	Frataxin gene
FXPOI	Fragile X-associated primary ovarian insufficiency
FXS	Fragile X syndrome
FXTAS	Fragile X-associated tremor/ataxia syndrome
g	g-force
GAPDH	Glyceraldehyde 3-phosphate dehydrogenase gene
GFP	Green fluorescent protein
GTP	Guanosine triphosphate
h(s)	Hour(s)
H3	Histone H3
H3K9	Lysine 9 of histone H3
HAT	Histone acetyltransferase
HD	Huntington's disease
HDAC	Histone deacetylase
hESC(s)	Human embryonic stem cell(s)
hnRNP	Heterogeneous nuclear ribonucleoprotein
HP1	Heterochromatin protein 1
HTT	Huntingtin protein
Ig	Immunoglobulin
IP	Immunoprecipitation
iPSC(s)	Induced pluripotent stem cell(s)
kb	Kilobase
kDa	Kilodalton
LC-MS/MS	Liquid chromatography tandem mass spectrometry
MeOH	Methanol
Min	Minute(s)
ml(s)	Millilitre(s)
MLS	Mitochondrial localisation signal
mRNA	Messenger RNA
ncRNA	non-coding RNA
NLS	Nuclear localisation signal

NRO	Nuclear run-on
nt(s)	Nucleotide(s)
NTP	Nucleotide triphosphate
PBS	Phosphate buffered saline
PCR	Polymerase chain reaction
PM	Premutation
PMSF	Phenylmethylsulfonyl fluoride
PWS	Prader-Willi syndrome
qPCR	Quantitative polymerase chain reaction
rDNA	Ribosomal DNA
RISC	RNA-induced silencing complex
RNA	Ribonucleic acid
RNA Pol II	RNA polymerase II
RNAi	RNA interference
RNase	Ribonuclease
rNTP	Ribonucleoside triphosphate
RT-qPCR	Reverse transcription-quantitative polymerase chain reaction
S. cerevisiae	*Saccharomyces cerevisiae*
S. pombe	*Schizosaccharomyces pombe*
SCAN1	Spinocerebellar ataxia with axonal neuropathy 1
SDS	Sodium dodecylsulfate
SDS-PAGE	Sodium dodecylsulfate polyacrylamide gel electrophoresis
SEM	Standard error of the mean
SINQ	Normalized spectral index quantification
siRNA	Small interfering RNA
SILAC	Stable isotope labelling by amino acids in cell culture
SSB	Single-strand break
ss	Single-stranded
SUV39H1/2	Suppressor of variegation 3-9 homologue 1/2
T	Thymine
TAM	Transcription-associated mutagenesis
TAR	Transcription-associated recombination
TAT	Trans-Activator of Transcription, HIV protein
TC-NER	Transcription-coupled nucleotide excision repair
Top1	Topoisomerase I
Top3B	Topoisomerase IIIB
TSS	Transcription start site
TTS	Transcription termination site
U	Uracil
UTP	Uridine 5'-triphosphate
WT	Wild type
X. laevis	*Xenopus laevis*

Chapter 1

Introduction

1.1 Structure and functions of R-loops

1.1.1 The structure of R-loops

R-loops are three-stranded structures, which occur when RNA hybridises to a complementary DNA strand. This leads to formation of an RNA/DNA hybrid, and results in displacement of the other DNA strand in this process (Figure 1.1). R-loops are implicated in many cellular processes in a wide range of organisms including bacteria, yeast, plants, mice and humans (Figure 1.1) (Chapters 1.1.2–1.1.5). The first R-loops were described in 1976, when their formation *in vitro* by annealing of ssRNA to dsDNA in the presence of 70% formamide was visualised by electron microscopy (Thomas et al., 1976) (see diagram of landmark discoveries in R-loop research, Figure 1.2). Since then, the technique of 'R-loop mapping' – hybridising messenger or ribosomal RNA to double-stranded DNA – has been used in over 140 studies to map gene organisation, transcription initiation sites and the direction of transcription, as well as measure the quantities of cellular RNAs (White and Hogness, 1977). Thus, initially R-loop formation was used as a molecular biology tool but it took more than 30 years until endogenous R-loops became the focus of research.

The exceptional *in vitro* stability of R-loops was already noted when they were first described in 1976 (Thomas *et al.*, 1976). Most subsequent studies investigated the double-stranded RNA/DNA hybrid part of R-loops rather than the whole three-stranded structure. *In vitro* data suggest that these RNA/DNA hybrids can be thermodynamically more stable than double-stranded DNA (dsDNA) (Roberts and Crothers, 1992; Thomas *et al.*, 1976). In particular, an extensive study comparing stabilities of oligonucleotides with different sequence composition has demonstrated that guanosine (G)-rich RNA forms particularly stable RNA/DNA hybrids, which are stronger than the corresponding dsDNA but weaker than dsRNA of the same sequence (Lesnik and Freier, 1995). Interestingly, RNA molecules lacking Gs form RNA/DNA hybrids which are considerably weaker than the corresponding dsDNA (Lesnik and Freier, 1995). In addition to the G-richness in the RNA strand, the relative stability of RNA/DNA hybrids is further dictated by their overall length and base composition, especially by the destabilising effect of pyrimidines (Lesnik and Freier, 1995; Shaw and Arya, 2008).

In addition to their stability, RNA/DNA hybrids can also exhibit conformations distinct from dsDNA or dsRNA. While dsDNA adopts the so-called B-form with 10 base pairs per turn, double-helical dsRNA adopts the A-form, which is more compact, with 11 base pairs per turn (Watson *et al.*, 2013). The A-form is further characterised by a narrower major groove and a broader minor groove (Watson *et al.*, 2013). In line with their chimeric nature, RNA/DNA hybrids can exist in an intermediate A/B-form, although some RNA/DNA hybrids can adopt a predominantly A-type helix (Shaw and Arya, 2008). Similar to the stability, the conformation of RNA/DNA hybrids is affected by their length and sequence composition (Shaw and Arya, 2008).

The thermodynamic stability and distinct conformation of RNA/DNA hybrids has important implications for the cellular functions of R-loops. While their unusual stability may in itself contribute to both beneficial and damaging effects *in vivo*, their unique structure may allow specific recognition by cellular proteins which may bind to either the RNA/DNA hybrid itself, the displaced single-stranded DNA (ssDNA) or the single-stranded RNA (ssRNA) in close proximity to the R-loop (Figure 1.1). Thus, it is not surprising that after their initial discovery, many processes have been associated with R-loops (indicated in Figure 1.1). Intriguingly, R-loops play essential roles in many cellular processes, but at the same time they can be the source of detrimental genomic instability (Santos-Pereira and Aguilera, 2015).

The first evidence for R-loop formation in live bacteria was obtained in 1994 (Drolet *et al.*, 1994). This was followed by numerous studies showing that R-loops exist in different organisms (Figure 1.2) (Aguilera and Garcia-Muse, 2012; Hamperl and Cimprich, 2014; Skourti-Stathaki and Proudfoot, 2014). In living cells, the majority of R-loops are thought to form *in cis* during transcription, when nascent RNA hybridises to the DNA template behind the elongating RNA polymerase (Aguilera and Garcia-Muse, 2012). However, in contrast to this more common view of co-transcriptional R-loops, recent studies suggest that RNA transcribed at one locus can hybridise to homologous DNA at another locus, thus leading to R-loop formation *in trans* (Wahba *et al.*, 2013). The following sections will describe the currently known functions of R-loops *in vivo* (Chapters 1.1.2-1.1.5), and their involvement in human disease (Chapters 1.2, 1.3.7). It is important to note that the current methodologies infer the presence of R-loops *in vivo* based on detecting either the

RNA/DNA hybrid (e.g. DNA/RNA immunoprecipitation, DIP) or the non-template strand ssDNA (e.g. using native bisulfite sequencing) present in R-loops. The detection of one substructure does not allow direct inference of the presence of the other substructure or the R-loop. An alternative potential source of RNA/DNA hybrids not present in R-loops are the RNA primers synthesised during DNA replication (Balakrishnan and Bambara, 2013). However, their very short half-life of less than 1 min in eukaryotes makes their detection less probable (DePamphilis and Bell, 2010). Furthermore, there is strong evidence suggesting that RNA/DNA hybrids formed in the genomic context of dsDNA predominantly occur as R-loops, as demonstrated by several studies detecting both RNA/DNA hybrids and ssDNA at the same loci (Ginno *et al.*, 2012; Loomis *et al.*, 2014; Zhang *et al.*, 2014a). Therefore, all current evidence suggests that the presence of R-loops can be correctly deduced from the presence of RNA/DNA hybrids, in particular in transcribed regions.

1.1.2 R-loops in DNA replication

Being amongst the first to investigate R-loops, Itoh and colleagues showed that R-loops are formed around the bacterial ColE1 origin of replication *in vitro* (Itoh and Tomizawa, 1980). Later studies have demonstrated that these R-loops help to unwind the DNA, making it more accessible for the initiation of replication (Baker and Kornberg, 1988; Fitzwater *et al.*, 1992). R-loops also form at the origin of replication of bacteriophage T4 and in mitochondria of *S. cerevisiae* and mammals (Brown *et al.*, 2008; Nossal *et al.*, 2001; Xu and Clayton, 1995, 1996).

During mitochondrial replication, R-loops are cleaved by the mitochondrial RNase MRP to produce RNA primers for DNA polymerase (Lee and Clayton, 1997). It has not been investigated if genuine R-loops form during eukaryotic DNA replication. Short RNA/DNA hybrids occur during replication in eukaryotes and serve as primers for DNA polymerases, but they seem to differ from R-loops in their biogenesis and structure (Mechali, 2010). The importance of R-loops for DNA replication has recently been demonstrated by genome-wide replication mapping in *E. coli* cells lacking RNase H1, which exhibited abnormal replication fork progression (Maduike *et al.*, 2014).

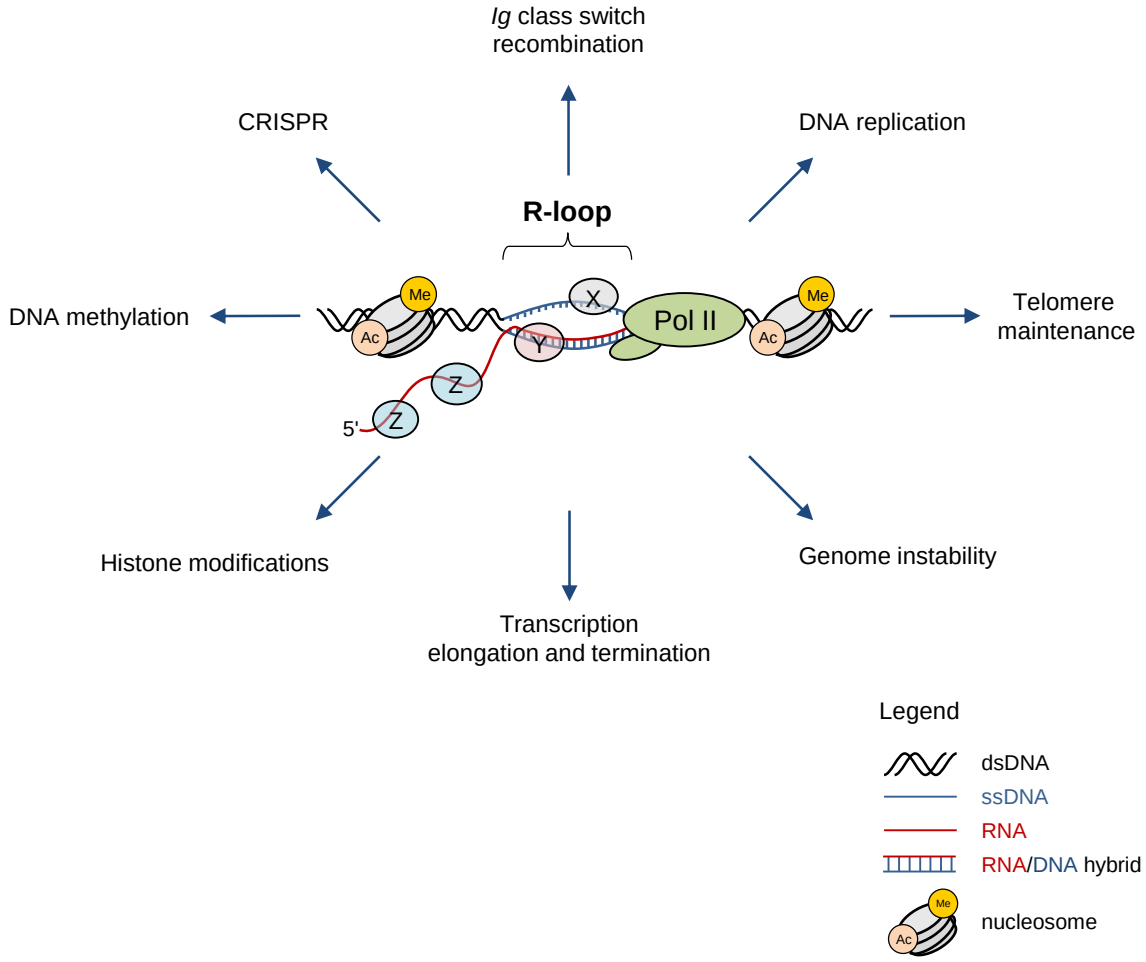


Figure 1.1. Structure and functions of R-loops

Schematic diagram representing the structure of co-transcriptionally formed R-loops. Nascent RNA (red) exiting RNA polymerase II (Pol II) can reanneal with the single-stranded DNA template strand (ssDNA, blue), forming an RNA/DNA hybrid and displacing the non-template DNA strand. R-loops can be associated with a diverse set of proteins. 'Z' depicts proteins binding to RNA, 'X' proteins binding to ssDNA, and 'Y' proteins binding to RNA/DNA hybrids. Ac, acetylation; Me, methylation. Transcripts produced from RNA Pol I and III can also form R-loops. Blue arrows indicate R-loop-associated functions and processes.

1.1.3 R-loops in *Ig* class switch recombination

R-loops have been implicated in immunoglobulin (*Ig*) class switch recombination (CSR) in mammals (Daniels and Lieber, 1995). *Ig* CSR is a highly regulated process involving controlled induction of DNA double-strand breaks (DSBs) in B cells, by which the *Ig* subtype is changed due to an immunological stimulus (Stavnezer *et al.*, 2008). During CSR, two distant gene segments are joined together: the *VDJ* gene encoding the *Ig* light chain and the heavy chain constant gene, which determines the functional *Ig* subtype (e.g., IgA, IgG, and IgM).

G-rich tandem repeat-containing sequences, termed switch (S) regions, lie upstream of each heavy chain constant gene. CSR is initiated by transcription of S regions upon B cell maturation (Stavnezer *et al.*, 2008). This allows access of activation-induced deaminase (AID) to ssDNA, leading to deamination of cytidines and formation of DSBs, which initiates the CSR process (Stavnezer *et al.*, 2008). Finally, the heavy chain constant gene is joined to the upstream *VDJ* gene, enabling the production of different immunoglobulin isotypes (Stavnezer *et al.*, 2008).

It has been shown that transcription of various class switch regions (*Sα, Sγ2b, Sγ3, Sμ*) leads to the formation of R-loops *in vitro* and in activated B cells (Daniels and Lieber, 1995; Reaban and Griffin, 1990; Tian and Alt, 2000; Yu *et al.*, 2003). Interestingly, these R-loops were found to extend up to ~1 kb, thereby exposing large stretches of ssDNA (Kao *et al.*, 2013; Yu *et al.*, 2003). A possible mechanistic involvement of R-loops in CSR has been highlighted by the observation that the activities of some DNA-targeting enzymes depend on the presence of R-loops. In particular, the nucleotide excision repair enzymes XPG and XPF-ERCC1 recognise and cleave at *Sμ* and *Sγ2b* switch regions *in vitro* only in the presence of R-loops (Tian and Alt, 2000). Similarly, AID is only able to deaminate cytidines *in vitro* when the non-template strand is exposed by formation of R-loops (Chaudhuri *et al.*, 2003). Consistent with this, hyper-recombination caused by AID is greatly increased in cells with aberrantly high R-loop formation (Dominguez-Sanchez *et al.*, 2011; Gomez-Gonzalez and Aguilera, 2007; Huertas and Aguilera, 2003). Interestingly, activated B cells lacking the 3'-5' exoribonuclease exosome complex exhibit widespread R-loop accumulation at known hypermutation sites and translocation hotspots close to divergently transcribed RNA Pol II promoters

(Pefanis *et al.*, 2014). Therefore, controlling R-loops is especially important for maintaining genome integrity in B cells which express the mutagenic AID enzyme.

The frequency of cellular R-loops formed during CSR has been estimated at 0.3% (Kao *et al.*, 2013). Although this is likely an underestimate due to the transient nature of R-loops, it suggests that other mechanisms could be important for CSR. When the murine *Sμ* region is replaced by its *X. laevis* equivalent, CSR is only slightly reduced (Zarrin *et al.*, 2004). In contrast to its mammalian counterpart, the *X. laevis Sμ* switch region is AT-rich and does not form R-loops *in vitro* (Zarrin *et al.*, 2004). Moreover, deletion of the *Sμ* region in murine B cells does not completely abolish their ability to undergo CSR (Luby *et al.*, 2001). Furthermore, it would be expected that AID-induced mutations in the class switch region occur preferentially on the exposed non-template DNA strand. In contrast to this, AID-induced mutations in murine B cells have been detected on both DNA strands at equal frequencies (Xue *et al.*, 2006; Yu *et al.*, 2003). As an alternative to an R-loop-dependent mechanism, AID could act on small stretches of ssDNA, potentially exposed by RNA Pol II itself, in cooperation with other proteins such as replication protein A (Chaudhuri *et al.*, 2004; Ronai *et al.*, 2007; Zarrin *et al.*, 2004).

Despite these open questions, recent studies have underlined the importance of R-loops for CSR. R-loops have been more thoroughly characterised at switch regions *in vitro* and *in vivo* and it has been shown that the efficiency of CSR depends on the presence of R-loop-promoting G-rich sequences and AID-targeting sites (Zhang *et al.*, 2014a; Zhang *et al.*, 2014b, 2015). Moreover, mice lacking factors that normally suppress R-loops exhibit higher R-loop levels and recombination rates at the *Ig* switch regions (Yang *et al.*, 2014).

1.1.4 R-loops and transcriptional regulation

First evidence for an involvement of R-loops in transcription was obtained in early studies in *E. coli* cells deficient of topoisomerase 1 (Top1). The growth defect of Top1-deficient *E. coli* cells was linked by genetic evidence to R-loops since overexpression of RNase H1, which targets RNA/DNA hybrids, partially compensates for this growth defect while loss of RNase H1 in Top1-deficient cells was lethal (Drolet *et al.*, 1995). R-loops formed *in vivo* were detected by

electrophoresis of plasmids extracted from Top1-deficient *E. coli*, inferred from differential and RNase H1-sensitive electrophoretic mobility compared to plasmids not forming R-loops (Masse and Drolet, 1999). Negative supercoiling, which occurs naturally behind the elongating RNA polymerase, favours the formation of R-loops *in vivo* and can be relieved by the activity of Top1 (Drolet, 2006; Liu and Wang, 1987).

Increased levels of R-loops lead to aberrant transcription of mRNA and rRNA genes in *E. coli* and *S. cerevisiae* (Baaklini *et al.*, 2008; El Hage *et al.*, 2010; Hraiky *et al.*, 2000; Tous and Aguilera, 2007). Furthermore, various reports have linked the abnormal transcription phenotypes of cells lacking important messenger ribonucleoprotein (mRNP) biogenesis factors to the formation of R-loops. In particular, the first description of the mechanism linking the impaired transcription and hyper-recombination in cells lacking the mRNP biogenesis factor Hpr1 with R-loop accumulation has inspired future work investigating the tight connection between transcription and genome integrity (Huertas and Aguilera, 2003). This and subsequent work from the Aguilera group established the importance of co-transcriptional RNA processing, packaging and export for efficient transcription and genome stability (Santos-Pereira and Aguilera, 2015). To date, this is best understood for the subunits of the mRNP biogenesis and export complexes THO/TREX, which promote removal of nascent RNA from DNA, thus preventing unscheduled R-loop formation and its detrimental consequences for cells (Balk et al., 2013; Bhatia et al., 2014; Castellano-Pozo et al., 2012; Castellano-Pozo et al., 2013; Chan et al., 2014; Dominguez-Sanchez et al., 2011; Gonzalez-Aguilera et al., 2008; Huertas and Aguilera, 2003; Pfeiffer et al., 2013; Stirling et al., 2012)

R-loops have also been implicated in transcriptional termination in mammalian cells, where maintaining correct levels of R-loops is required to prevent transcriptional read-through (Skourti-Stathaki *et al.*, 2011). In particular, R-loops cause pausing of RNA Pol II at a G-rich pause site downstream of the *β-Actin* gene poly(A) signal. These R-loops are resolved by the RNA/DNA helicase senataxin (SETX), which allows the 5'-3' exonuclease XRN2 to bind to the nascent RNA, degrade it and thereby promote RNA Pol II transcriptional termination (Skourti-Stathaki *et al.*, 2011). Furthermore, it has been demonstrated that R-loops at RNA Pol II pause sites promote antisense transcription, leading to dsRNA accumulation and local heterochromatin formation in an RNAi-

dependent way (Skourti-Stathaki *et al.*, 2014). Two recent studies have suggested that SETX may be recruited to human gene termination regions by the interaction with other proteins. In particular, BRCA1 is enriched at termination regions and is required for efficient SETX recruitment and R-loop resolution during transcriptional termination (Hatchi *et al.*, 2015). Moreover, symmetric arginine methylation of the non-consensus RNA Pol II carboxy-terminal domain (CTD) residue R1810 by PRMT5 is recognised by the SMN protein, which promotes SETX recruitment and transcriptional termination (Zhao *et al.*, 2015). It is also becoming increasingly clear that dysregulation of R-loop-dependent transcriptional termination due to loss of SETX, BRCA1, PRMT5 or SMN poses serious threats to genome integrity, which may contribute to human diseases (Chapters 1.2.1, 1.2.4) (Hatchi *et al.*, 2015; Zhao *et al.*, 2015). Interestingly, *S. cerevisiae* mutants of Sen1 (the yeast SETX homologue) and pre-mRNA cleavage factor Pcf11 also accumulate R-loops and exhibit defective transcriptional termination of mRNA and ncRNA genes (Grzechnik *et al.*, 2015; Mischo *et al.*, 2011). This suggests an evolutionary conserved requirement for tightly controlling R-loop levels during transcriptional termination.

The large potential for functional interplay between R-loops and transcription has been further highlighted by a number of genome-wide studies which have found a strong enrichment of R-loops at RNA Pol II gene promoters in human and murine cells (Chen *et al.*, 2015a; Ginno *et al.*, 2012; Nadel *et al.*, 2015). R-loop levels correlate with reduced DNA methylation of CpG dinucleotides at promoters *in vivo* and they were shown to inhibit methylation by DNA methyltransferase DNMT3B1 *in vitro* (Ginno et al., 2012). Thus, R-loops may be important for initiating or maintaining the hypomethylation at a subset of CpG promoters. In line with this, cells from Aicardi-Goutières patients which accumulate R-loops also exhibit icreased hypomethylation at promoters (Lim *et al.*, 2015). This suggests that R-loops formed at promoters may assist transcription initiation by maintaining them in a hypomethylated state accessible for the RNA Pol II initiation complex.

A recent study further supports the notion that promoter-associated R-loops may positively affect transcription. In particular, R-loops promote binding of the Tip60-p400 histone acetyltransferase complex but prevent recruitment of the Polycomb repressive complex 2 (PRC2)

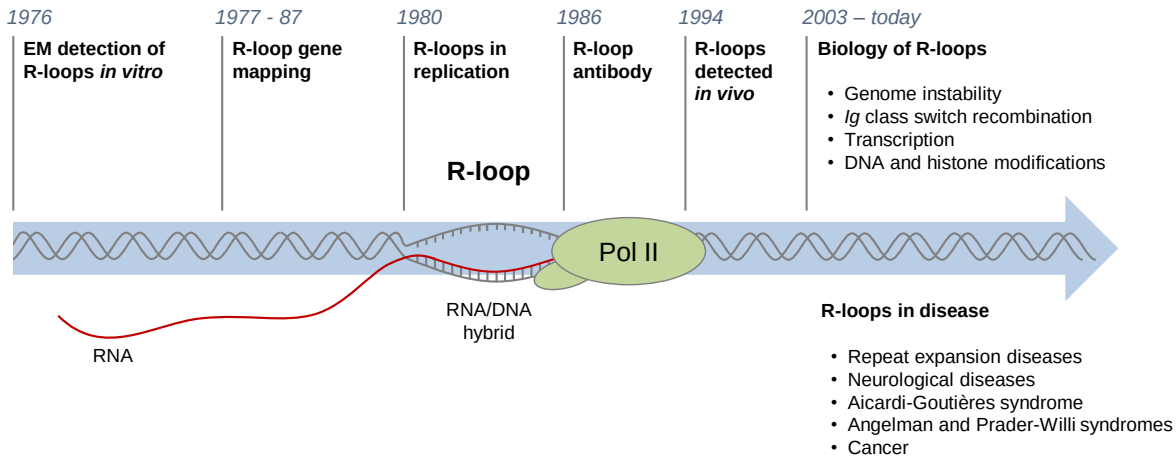


Figure 1.2. Landmark discoveries in R-loop research

The diagram depicts the major developments in the R-loop field since 1976 and diseases associated with R-loop dysregulation. R-loop gene mapping was methodology used to describe the mapping of genes by hybridising messenger or ribosomal RNA to double-stranded genomic DNA. The first genome-wide mapping of endogenous R-loops was published in 2012.

(Chen *et al.*, 2015a). Although it was not investigated if this was mediated by direct interaction between Tip60 or PRC2 with R-loops, this and other studies have demonstrated the intricate interplay between R-loops and epigenetic regulators (Castellano-Pozo *et al.*, 2013; Chen *et al.*, 2015a; Skourti-Stathaki *et al.*, 2014).

As first shown by Skourti-Stathaki and colleagues, R-loops may also influence transcription and chromatin structure by regulating antisense transcription at 3' ends of genes (see above) (Skourti-Stathaki *et al.*, 2014). Since antisense transcription is a common feature of many human promoters and is tightly intertwined with sense transcription (Pelechano and Steinmetz, 2013), there may be connections between R-loops and promoter-associated antisense transcription. Indeed, it has been shown for the *vimentin* (*VIM*) gene that an R-loop is formed by a promoter-associated antisense RNA, termed *VIM-AS* (Boque-Sastre *et al.*, 2015). These antisense R-loops maintained the *VIM* gene promoter in an accessible chromatin conformation, thus allowing binding of sequence-specific transcription factors and efficient transcription of the *VIM* gene (Boque-Sastre *et al.*, 2015).

These data suggest a potentially widespread positive effect of promoter-associated R-loops on transcription. However as mentioned above, R-loops have also been implicated in directly impeding transcription. This has been shown directly *in vitro* and in *S. cerevisiae* (Huertas and Aguilera, 2003; Tous and Aguilera, 2007). Thus, while a certain level of promoter-associated R-loops may be necessary for active transcription, aberrantly high R-loop levels can lead to transcriptional repression, as shown for the *c-MYC* gene (Yang *et al.*, 2014). These seemingly disparate observations suggest that depending on their genomic location and relative abundance R-loops may be associated with different functions. One possible explanation for this is that R-loops may be recognised by different proteins, depending on where and how they are formed.

1.1.5 Proteins interacting with R-loops

The number of proteins associated with R-loop biology has increased in the last few years, reflecting the diversity of R-loop processes (Table 1.1) (Aguilera and Garcia-Muse, 2012; Hamperl and Cimprich, 2014; Skourti-Stathaki and Proudfoot, 2014). Many proteins can regulate cellular R-loop levels either directly or indirectly, mostly by preventing RNA from hybridising to DNA, thus

reducing excessive R-loop accumulation. Among these are proteins required for efficient transcription elongation, termination, polyadenylation, RNA splicing, packaging and export (Herrera-Moyano *et al.*, 2014; Huertas and Aguilera, 2003; Li and Manley, 2005; Santos-Pereira *et al.*, 2013; Skourti-Stathaki *et al.*, 2011; Stirling *et al.*, 2012; Wahba *et al.*, 2011).

DNA topology itself can influence hybridization of RNA to DNA, and topoisomerases consequently play important roles in modulating R-loop levels (Tuduri *et al.*, 2009; Yang *et al.*, 2014). Proteins involved in maintenance of genome integrity can also regulate R-loops, suggesting a dynamic interplay between DNA repair and R-loop formation (Wahba *et al.*, 2013). Importantly, cells possess dedicated enzymes including the members of the RNase H family that specifically degrade the RNA in R-loops (Cerritelli and Crouch, 2009), and helicases that can unwind RNA/DNA hybrids (Mischo *et al.*, 2011; Skourti-Stathaki *et al.*, 2011). The cellular functions of RNase H enzymes and implications for human disease are discussed in Chapter 1.2.2.

As discussed above, R-loops can directly affect many gene expression-associated processes including DNA methylation, post-translational histone modifications and transcription by influencing the function of regulatory proteins (Castellano-Pozo *et al.*, 2013; Ginno *et al.*, 2012; Nakama *et al.*, 2012; Skourti-Stathaki *et al.*, 2011). Despite the growing number of proteins involved in R-loop homeostasis and human disease, many questions still remain unanswered. For many proteins with documented *in vitro* RNA/DNA helicase activity (*e.g.* Pif1, the MCM complex), *in vivo* evidence is generally still lacking (Table 1.1) (Boule and Zakian, 2007; Shin and Kelman, 2006). Moreover, the molecular mechanisms underlying interactions between proteins and R-loops are poorly understood, and in many cases the connections to disease remain obscure.

Table 1.1. Proteins implicated in R-loop biology

This table constitutes a broad list of proteins linked to R-loop biology. In many cases, evidence is based solely on screens and most proteins listed here still warrant detailed further investigation. For multiprotein complexes, only subunits directly implicated in R-loop biology are mentioned in the table. Asterisk (*) indicates that the protein association with R-loops is based on *in vitro* evidence.

Protein	Organism	Reference
Transcription initiation and capping		
Bur2	Yeast	*(Wahba et al., 2011)*
Mediator kinase module (Med12, Med13, Cdk8, CycC)	Yeast	*(Wahba et al., 2011)*
Mediator complex (Med1, Med5)	Yeast	*(Wahba et al., 2011)*
Taf5	Yeast	*(Chan et al., 2014)*
Mot1	Yeast	*(Chan et al., 2014)*
Not5	Yeast	*(Wahba et al., 2011)*
Stb3	Yeast	*(Wahba et al., 2011)*
Capping enzyme	Human	*(Kaneko et al., 2007)**
Transcription elongation		
PAF1 complex (Leo1, Cdc73)	Yeast	*(Wahba et al., 2011; Wahba et al., 2013)*
Spt2	Yeast	*(Sikdar et al., 2008; Wahba et al., 2011)*
FACT complex (Spt16/SPT16, Pob3/SSRP1)	Yeast, Human	*(Herrera-Moyano et al., 2014)*
Transcriptional termination, cleavage and polyadenylation		
CFIA complex (Pcf11, Clp1, Rna15)	Yeast	*(Stirling et al., 2012)*
Rtt103/KUB5-HERA/RPRD1B	Yeast, Human	*(Morales et al., 2014; Stirling et al., 2012)*
Sen1/SETX	Yeast, Mouse, Human	*(Alzu et al., 2012; Becherel et al., 2013; Castellano-Pozo et al., 2013; Chan et al., 2014; Mischo et al., 2011; Skourti-Stathaki et al., 2011; Stirling et al., 2012; Yeo et al., 2014; Yuce and West, 2013)*
CPF complex (Cft2, Fip1/FIP1L1)	Yeast, Human	*(Stirling et al., 2012)*
SMN/GEMIN1	Human	*(Zhao et al., 2015)*
RNA processing and export		
THO/TREX complex (Tho2/THOC2, Hpr1/THOC1, Mft1, Thp2, Sub2/UAP56)	Yeast, Nematode, Mouse, Human	*(Balk et al., 2013; Bhatia et al., 2014; Castellano-Pozo et al., 2012; Castellano-Pozo et al., 2013; Chan et al., 2014; Dominguez-Sanchez et al., 2011; Gonzalez-Aguilera et al., 2008; Huertas and Aguilera, 2003; Pfeiffer et al., 2013; Stirling et al., 2012)*
THSC complex (Thp1, Sac3, Sus1)	Yeast	*(Gonzalez-Aguilera et al., 2008)*
Npl3	Yeast	*(Santos-Pereira et al., 2013; Wahba et al., 2011)*

Dbp6	Yeast	*(Chan et al., 2014)*
Dbp7	Yeast	*(Chan et al., 2014)*
Rpf1	Yeast	*(Chan et al., 2014)*
Imp4	Yeast	*(Chan et al., 2014)*
Xrn1/Kem1	Yeast	*(Wahba et al., 2011; Wahba et al., 2013)*
Srm1	Yeast	*(Stirling et al., 2012)*
Nup133	Yeast	*(Chan et al., 2014)*
Rna1	Yeast	*(Chan et al., 2014)*
Kae1	Yeast	*(Chan et al., 2014)*
TRAMP complex (Trf4, Air1)	Yeast	*(Gavalda et al., 2013; Wahba et al., 2011)*
Exosome complex (Rrp6/EXOSC10, Dis3, EXOSC3)	Yeast, Mouse	*(Chan et al., 2014; Pefanis et al., 2014; Pefanis et al., 2015; Wahba et al., 2011; Wahba et al., 2013)*
Splicing		
snRNP complexes (Prp31, Snu13, Snu114, Snu66)	Yeast	*(Chan et al., 2014)*
Yhc1	Yeast	*(Chan et al., 2014)*
Mud2	Yeast	*(Chan et al., 2014)*
ASF/SF2	Chicken, Mouse, Human	*(Li and Manley, 2005; Tuduri et al., 2009)*
RNPS1	Chicken, Human	*(Li et al., 2007)*
OMCG1	Mouse	*(Houlard et al., 2011)*
AQR (Aquarius)	Human	*(Paulsen et al., 2009; Sollier et al., 2014)*
RNA/DNA hybrid cleavage		
RNase H1	Bacteria, Yeast, Mouse, Human	*(Cerritelli and Crouch, 2009; Tadokoro and Kanaya, 2009)*
RNase H2	Bacteria, Yeast, Mouse, Human	*(Cerritelli and Crouch, 2009; Tadokoro and Kanaya, 2009)*
DNA topology		
Topoisomerase 1	Bacteria, Yeast, Mouse, Human	*(El Hage et al., 2010; Marinello et al., 2013; Masse and Drolet, 1999; Powell et al., 2013; Tuduri et al., 2009)*
Topoisomerase 3B	Mouse, Human	*(Yang et al., 2014)*
DNA and histone modifications		
H3S10p	Yeast, Nematode, Human	*(Castellano-Pozo et al., 2013)*
H3K9me2	Yeast, Nematode, Human	*(Castellano-Pozo et al., 2013; Groh et al., 2014a; Nakama et al., 2012; Pefanis et al., 2014)*
H4R3me2a	Human, Mouse	*(Yang et al., 2014)*
TDRD3	Human, Mouse	*(Yang et al., 2014)*
γH2A/γH2AX	Yeast, Mouse, Rat, Human	*(Chernikova et al., 2012; Herrera-Moyano et al., 2014; Morales et al., 2014; Paulsen et al., 2009; Sordet et al., 2009; Stirling et al., 2012; Yeo et al., 2014)*

CpG methylation (DNMT3B1)	Human	*(Ginno et al., 2012)*
Psh1	Yeast	*(Chan et al., 2014)*
Rpd3C(L) complex (Sin3, Sds3)	Yeast	*(Chan et al., 2014; Wahba et al., 2011; Wahba et al., 2013)*
AID	Yeast, Mouse	*(Gomez-Gonzalez and Aguilera, 2007; Ruiz et al., 2011; Yang et al., 2014; Yu et al., 2003)*
BRE1A/B	Mouse	*(Chernikova et al., 2012)*
PRMT5	Human	*(Zhao et al., 2015)*
DNA repair and genome maintenance		
Esc2	Yeast	*(Chan et al., 2014)*
RecA/Rad51	Bacteria, Yeast	*(Hong et al., 1995; Kirkpatrick and Radding, 1992; Wahba et al., 2013)*
Rad52	Yeast	*(Herrera-Moyano et al., 2014; Wahba et al., 2013)*
Srs2	Yeast	*(Wahba et al., 2013)*
Sts1	Yeast	*(Chan et al., 2014)*
ATM	Mouse	*(Yeo et al., 2014)*
TDP1	Mouse	*(Yeo et al., 2014)*
APTX	Mouse	*(Yeo et al., 2014)*
BUB3	Human	*(Wan et al., 2015)*
ZNF207	Human	*(Wan et al., 2015)*
BRCA1	Human	*(Bhatia et al., 2014; Hatchi et al., 2015)*
BRCA2	Human	*(Bhatia et al., 2014)*
XPF	Human	*(Sollier et al., 2014)*
XPG	Human	*(Sollier et al., 2014)*
ssDNA-binding proteins		
AtNDX	Arabidopsis	*(Sun et al., 2013)*
RNA/DNA hybrid helicases		
NS3	HPV & HPG	*(Du et al., 2002; Gwack et al., 1999)**
UvsW	Bacteriophage T4	*(Dudas and Kreuzer, 2001)*
Rho	Bacteria	*(Boudvillain et al., 2013; Brennan et al., 1987)**
RecG	Bacteria	*(Hong et al., 1995; Vincent et al., 1996)*
Cas3	Bacteria	*(Sinkunas et al., 2011)**
DnaB/MCM complex	Bacteria, Archaea, Yeast	*(Shin and Kelman, 2006)**
Pif1/PIF1	Yeast, Human	*(Boule and Zakian, 2007; Zhang et al., 2006)**
Has1	Yeast	*(Rocak et al., 2005)**
MLE/DHX9	Drosophila, Human	*(Chakraborty and Grosse, 2011; Lee et al., 1997)**
SUV3	Human	*(Shu et al., 2004)**
WRN	Human	*(Chakraborty and Grosse, 2010; Suzuki et al., 1997)**
DDX11	Human	*(Hirota and Lahti, 2000)**

1.2 R-loops in human disease

1.2.1 Amyotrophic lateral sclerosis 4 and Ataxia oculomotor apraxia type 2

The biological importance of R-loops in humans is supported by the fact that mutations in proteins implicated in R-loop resolution cause devastating human diseases, often related to neurodegeneration. Mutations in the putative RNA/DNA helicase senataxin (SETX) cause neurodegenerative diseases, the dominant juvenile form of amyotrophic lateral sclerosis type 4 (ALS4) and a recessive form of ataxia oculomotor apraxia type 2 (AOA2) (Figure 1.3A). These diseases are characterised by progressive degeneration of motor neurons in the brain and spinal cord, muscle weakness and atrophy (Anheim *et al.*, 2009; Chen *et al.*, 2004; Moreira *et al.*, 2004).

In addition to its predicted function as an RNA/DNA helicase, SETX interacts with proteins involved in diverse aspects of RNA metabolism (Suraweera *et al.*, 2009). Moreover, a single amino acid mutation which compromises the function of the yeast homologue, Sen1, dramatically changes the RNA Pol II distribution genome-wide, further supporting the view that SETX/Sen1 functions in the regulation of transcription (Steinmetz *et al.*, 2006). As discussed in Chapter 1.1.4, the role of SETX and Sen1 in transcriptional termination has recently been elucidated (Mischo *et al.*, 2011; Porrua and Libri, 2013; Skourti-Stathaki *et al.*, 2014; Skourti-Stathaki *et al.*, 2011). In light of the importance of transcriptional termination for gene expression, it is possible that this mechanism contributes to the phenotypes of AOA2/ALS4 (Figure 1.3A). In line with its function in R-loop resolution, SETX/Sen1 is also involved in maintaining genome integrity by coordinating transcription, DNA replication and the DNA damage response (Alzu *et al.*, 2012; Suraweera *et al.*, 2007; Yuce and West, 2013). SETX can target the 3'-5' RNA degradation complex, the exosome, to sites of transcription-induced DNA damage (Richard *et al.*, 2013).

Furthermore, SETX protects genome integrity by coordinating meiotic recombination with transcription during spermatogenesis, and gene silencing during meiotic sex chromosome inactivation (Becherel *et al.*, 2013). In particular, *Setx* knock-out mice accumulate DNA double strand breaks and R-loops, and fail to disassemble Rad51 filaments. This results in a failure to cross-over, likely due to collisions between R-loops and Holliday junctions (Becherel *et al.*, 2013). These

defects in *Setx* knock-out mice lead to male infertility, raising the question as to how this relates to the fertility of male AOA2/ALS4 patients.

Studies in neuronal cells have demonstrated a role for SETX in neuronal differentiation through FGF8 signalling, providing one explanation for the effects of loss-of-function AOA2 mutations (Vantaggiato *et al.,* 2011). Surprisingly, overexpression of dominant mutant forms of SETX does not affect neuritogenesis, suggesting that a different function of SETX may be affected in ALS4 patients. However, the interplay between the function of SETX in R-loop resolution, genome maintenance and neuronal differentiation warrants further investigation. In a recent study, Lavin and colleagues examined cells from mice with disrupted *Atm*, *Tdp1, Setx,* or *Aptx* genes, which cause Ataxia telangiectasia (AT), Spinocerebellar ataxia with axonal neuropathy 1 (SCAN1), AOA2 and AOA1 disorders, respectively (Yeo *et al.,* 2014). These diseases are characterised by a defective response to DNA damage, suggesting that R-loops may be implicated in triggering genome instability. Indeed, R-loops were found to be enriched in proliferating testis cells but not in the brain tissues from *Setx*, *Atm*, *Tdp1* or *Aptx* knock-out mice (Yeo *et al.,* 2014). The enrichment of R-loops in testes correlated with high levels of DNA damage and apoptosis. However, the lack of R-loops in brain tissue questions the association between R-loops and neurodegeneration.

This result is surprising because inducible R-loops have been previously detected in neuronal cells at the *Snord116* locus, which is associated with the neurodevelopmental disorder Angelman syndrome (AS), as discussed in Chapter 1.2.3 (Powell *et al.,* 2013). Furthermore, R-loops were implicated in inducing DNA damage in non-proliferating cells and post-mitotic neurons and proposed to contribute to the neurodegeneration seen in AT patients (Sordet *et al.,* 2009). It is possible that R-loops are regulated by different mechanisms in proliferating cells and post-mitotic neurons, thereby leading to different R-loop kinetics and so preventing their detection in some model systems. In particular, R-loop accumulation may reflect collisions between transcription and replication machineries (Helmrich *et al.,* 2013; Helmrich *et al.,* 2011), events which do not occur in post-mitotic neurons. It should be noted that the mouse models currently used may not fully recapitulate all aspects of human neurodegeneration. Patient-derived induced pluripotent stem cells (iPSCs) may provide an alternative model system to study the molecular pathology of AOA2. Indeed,

iPSCs from AOA2 patients and AOA2 iPSC-derived neuronal progenitors exhibit increased DNA damage and R-loop levels (Becherel *et al.*, 2015). Interestingly, this study also reported a significant increase of mitochondrial R-loops in AOA2 iPSC-derived neuronal progenitors, which may contribute to the mitochondrial DNA damage observed in the majority of neurodegenerative diseases, including AOA2 (Becherel *et al.*, 2015).

Additional mechanisms may link the loss-of-function of SETX with AOA2 pathology. SETX has been implicated in promoting premature transcriptional termination, which contributes to regulating the release of promoter-proximally paused RNA Pol II to productive elongation (Jonkers and Lis, 2015; Wagschal *et al.*, 2012). More recently, SETX-dependent premature transcriptional termination was shown to limit the activation of pro-inflammatory genes upon viral infection (Miller *et al.*, 2015). Thus, loss of SETX leads to altered immune responses to viruses and it was suggested that the resulting excessive inflammation may contribute to the progressive deterioration of affected neuronal tissues (Miller *et al.*, 2015). However, it has not been investigated yet if R-loops play a role in SETX-mediated premature transcriptional termination.

1.2.2 Aicardi-Goutières syndrome

In addition to their generation during transcription, RNA/DNA hybrids can arise due to incorporation of ribonucleotides into DNA by DNA polymerases during replication. Ribonucleases H (RNase H) are endonucleases that cleave the RNA of RNA/DNA hybrids in a sequence-independent manner, thus maintaining genome stability by resolving R-loops that form during transcription and by removing misincorporated ribonucleotides from the DNA (Cerritelli and Crouch, 2009).

Eukaryotic cells have two types of these enzymes, RNase H1 and H2, which have different enzymatic and site-specific activities (Chon *et al.*, 2013). In particular, RNase H1 requires a tract of at least four ribonucleotides to cleave the RNA/DNA hybrid, whereas RNase H2 can incise 5' to a single ribonucleotide incorporated within a DNA molecule (Cerritelli and Crouch, 2009; Chon *et al.*, 2013). Therefore, only RNase H2 can process single ribonucleotides incorporated in the DNA but both enzymes are capable of eliminating RNA/DNA hybrids. Unlike in bacteria and unicellular eukaryotic organisms, where RNase H enzymes are dispensable for viability, both RNase H enzymes

are essential in higher eukaryotes. RNase H1 has been implicated in mitochondrial DNA (mtDNA) replication during mouse development, a process likely to be associated with processing of RNA primers during mtDNA replication (Cerritelli *et al.*, 2003).

RNase H2 is composed of three different subunits, the catalytic subunit 2A and two other subunits, 2B and 2C, all of which are required for enzyme activity. RNase H2 has been implicated in recognition and removal of ribonucleotides incorporated into DNA and hydrolysis of Okazaki fragment RNA primers during DNA replication (Cerritelli and Crouch, 2009; Lazzaro *et al.*, 2012; McElhinny *et al.*, 2010; Reijns *et al.*, 2012; Rydberg and Game, 2002). In addition, recent studies point towards a role of RNase H2 in R-loop resolution during transcription *in vivo* (Arana *et al.*, 2012; El Hage *et al.*, 2010; El Hage *et al.*, 2014). In particular, deletion of *S. cerevisiae* RNase H2 imposes transcriptional blocks and R-loop accumulation over rDNA regions in cells depleted of Top1 (El Hage *et al.*, 2010) and transcriptional down-regulation of genes with higher GC content at the promoter regions, which are likely to form stable R-loops (Arana *et al.*, 2012). In humans, mutations in any of the three subunits of RNase H2 cause Aicardi-Goutières syndrome (AGS), a neurological inflammatory disorder, which resembles a congenital viral infection and is associated with accumulation of ribonucleotides in the DNA (Figure 1.3B) (Crow *et al.*, 2006b; Crow and Rehwinkel, 2009).

Interestingly, AGS can also be triggered by mutations in ssDNA 3'-5' exonuclease TREX1 (DNASEIII) (Crow *et al.*, 2006a), dsRNA-editing enzyme ADAR1 (Rice *et al.*, 2012), dNTP triphosphatase SAMHD1 (Rice *et al.*, 2009) and dsRNA receptor IFIH1 (also called MDA5) (Rice *et al.*, 2014). These proteins are involved in diverse pathways of nucleic acid metabolism, although their functions are not yet fully understood. They have been implicated in degrading ssDNA arising from endogenous retroelements or replication stress (TREX1), regulating the intracellular dNTP pool available for replication and reverse transcription of these retroelements (SAMHD1), or altering the immune response to RNA species through RNA editing of retroelements and microRNAs (ADAR1) (Lee-Kirsch *et al.*, 2014). Loss-of-function mutations in these genes are associated with an accumulation of unprocessed nucleic acids or increased recognition by the immune system (Lee-Kirsch *et al.*, 2014; Rabe, 2013). In contrast, IFIH1 normally recognises exogenous viral dsRNA

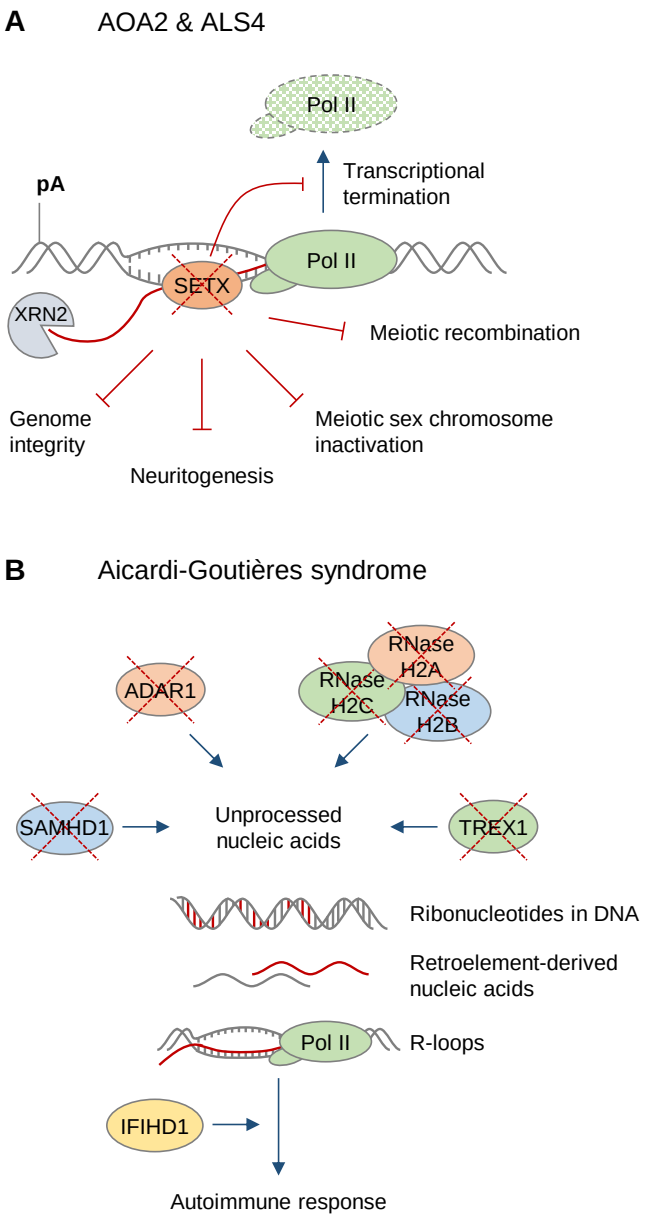


Figure 1.3. R-loops in AOA2/ALS4 and Aicardi-Goutières syndrome (AGS)

A. Mutations in human RNA/DNA helicase senataxin (SETX) are associated with AOA2/ALS4 disorders and lead to R-loop accumulation and defects in transcriptional termination by RNA Pol II, genome maintenance, meiotic recombination during spermatogenesis, gene silencing during meiotic sex chromosome inactivation, and neuronal differentiation.

B. Aicardi-Goutières syndrome (AGS) is associated with mutations in all three subunits of RNase H2, ssDNA 3'-5' exonuclease TREX1 (DNASEIII), dsRNA-editing enzyme ADAR1 and dNTP triphosphatase SAMHD1. Mutations in these proteins trigger accumulation of unprocessed nucleic acids including genomic DNA with incorporated ribonucleotides, R-loops and retroelement-derived nucleic acids, and result in the immune response characteristic of AGS. Mutations in cytosolic dsRNA sensor IFIHD1 decrease its activation threshold, leading to aberrant recognition of endogenous nucleic acids and inflammation.

(Rice *et al.*, 2014). AGS-associated gain-of-function mutations of IFIH1 lower its activation threshold and may lead to hypersensitivity for endogenous nucleic acids and activation of the immune response (Rice *et al.*, 2014).

So far, pathologies linked to AGS mutations in RNase H2 have been mainly attributed to genome instability caused by accumulation of rNTPs in DNA (Hiller *et al.*, 2012; Reijns *et al.*, 2012). The research has been hampered by the difficulty to uncouple the two activities of RNase H2; its ability to remove ribonucleotides from the DNA and to resolve R-loops, both of which are affected when RNase H2 is deleted (Chon *et al.*, 2013; Reijns *et al.*, 2012). Nevertheless, several lines of evidence suggest that R-loops may be involved in AGS pathology. An AGS-related mutation in the yeast RNase H2 enzyme resulted in its reduced RNA/DNA cleavage activity (Chon *et al.*, 2013). Since RNase H2 constitutes ~90% of the total cellular RNA/DNA hybrid cleavage activity, its loss due to AGS mutations may lead to significant accumulation of R-loops (Reijns *et al.*, 2012). In line with a contribution of RNA/DNA hybrids to AGS pathology, recent genome-wide data demonstrated that increased R-loop levels were detected in cells from AGS patients and correlated with pro-inflammatory markers (Lim *et al.*, 2015).

The importance of RNase H2 is further highlighted by the fact that mutations in RNase H1 do not cause AGS, suggesting that RNase H2 may have unique properties to degrade RNA/DNA hybrids (Chon *et al.*, 2013). Indeed, R-loops arising during DNA replication may be exclusively degraded by RNase H2, as they may be inaccessible to RNase H1 (Bubeck *et al.*, 2011; Chon *et al.*, 2013). A recently generated *S. cerevisiae* RNase H2 mutant, which possesses R-loop degrading activity but fails to remove single rNTPs from the DNA (Chon *et al.*, 2013) will be a useful tool in addressing the contribution of unresolved transcription-associated R-loops to AGS pathology.

TREX1, ADAR1 and SAMHD1 process retroelement-derived nucleic acids and help to suppress retroelements expansion in the host genome and their recognition by the immune system (Lee-Kirsch *et al.*, 2014). Interestingly, recent genome-wide studies have demonstrated that RNA/DNA hybrids are particularly enriched at retrotransposon elements in yeast and human cells (Chan *et al.*, 2014; Nadel *et al.*, 2015), suggesting that expansion of retroelements due to mutations in TREX1, ADAR1 or SAMHD1 may lead to increased RNA/DNA hybrid levels, contributing to

autoimmunity in AGS. Indeed, it has recently been demonstrated that RNA/DNA hybrids can be sensed by toll-like receptor 9 (TLR9) to induce pro-inflammatory cytokine and anti-viral interferon production in dendritic cells (Rigby *et al.*, 2014). The relative contribution of ribonucleotides incorporated into DNA and RNA/DNA hybrids in the pathology of AGS will need to be addressed in studies investigating both of these aspects in the future.

1.2.3 Prader-Willi syndrome and Angelman syndrome

Angelman syndrome (AS) and Prader-Willi syndrome (PWS) are imprinted neurodevelopmental disorders that are often caused by large deletions of human chromosome 15q11–q13 over the *Snord116* gene locus, but the deletion differs in its parent-of-origin (Cassidy *et al.*, 2000). In neurons, only the maternal *Ube3a* allele is expressed because the paternal *Ube3a* allele is silenced by expression of the ncRNA *Ube3a-ATS* (Figure 1.4A) (Meng *et al.*, 2012). AS therapies therefore seek to reactivate the silenced but genetically intact paternal *Ube3a* allele. Indeed, a recent study employed antisense oligonucleotides (ASOs) to specifically reduce levels of *Ube3-ATS* (Meng *et al.*, 2015). The resulting sustained reactivation of the paternal *Ube3a* allele led to amelioration of cognitive defects in an AS mouse model, thereby providing a proof of principle for successful therapeutic intervention in AS (Meng *et al.*, 2015).

Interestingly, R-loops were shown to regulate the neuronal expression of the paternal *Ube3a-ATS* transcript (Powell *et al.*, 2013). Treatment with the Top1 inhibitor topotecan increases R-loop levels over the *Snord116* locus, resulting in chromatin decondensation, inhibition of RNA Pol II transcription of *Ube3a-ATS*, and concomitant increase in *Ube3a* expression from the paternal allele (Figure 1.4B). This R-loop-mediated reactivation of paternal *Ube3a* could therefore compensate for the loss of maternal *Ube3a* in AS, and so potentially holds promise for targeted therapies for both AS and PWS.

1.2.4 R-loops in cancer

Genome instability is a hallmark of cancer and it may actively drive hereditary tumour development (Hanahan and Weinberg, 2011; Negrini *et al.*, 2010). Research in the last decade has clearly demonstrated that dysregulation of R-loops can corrupt genome integrity, resulting in

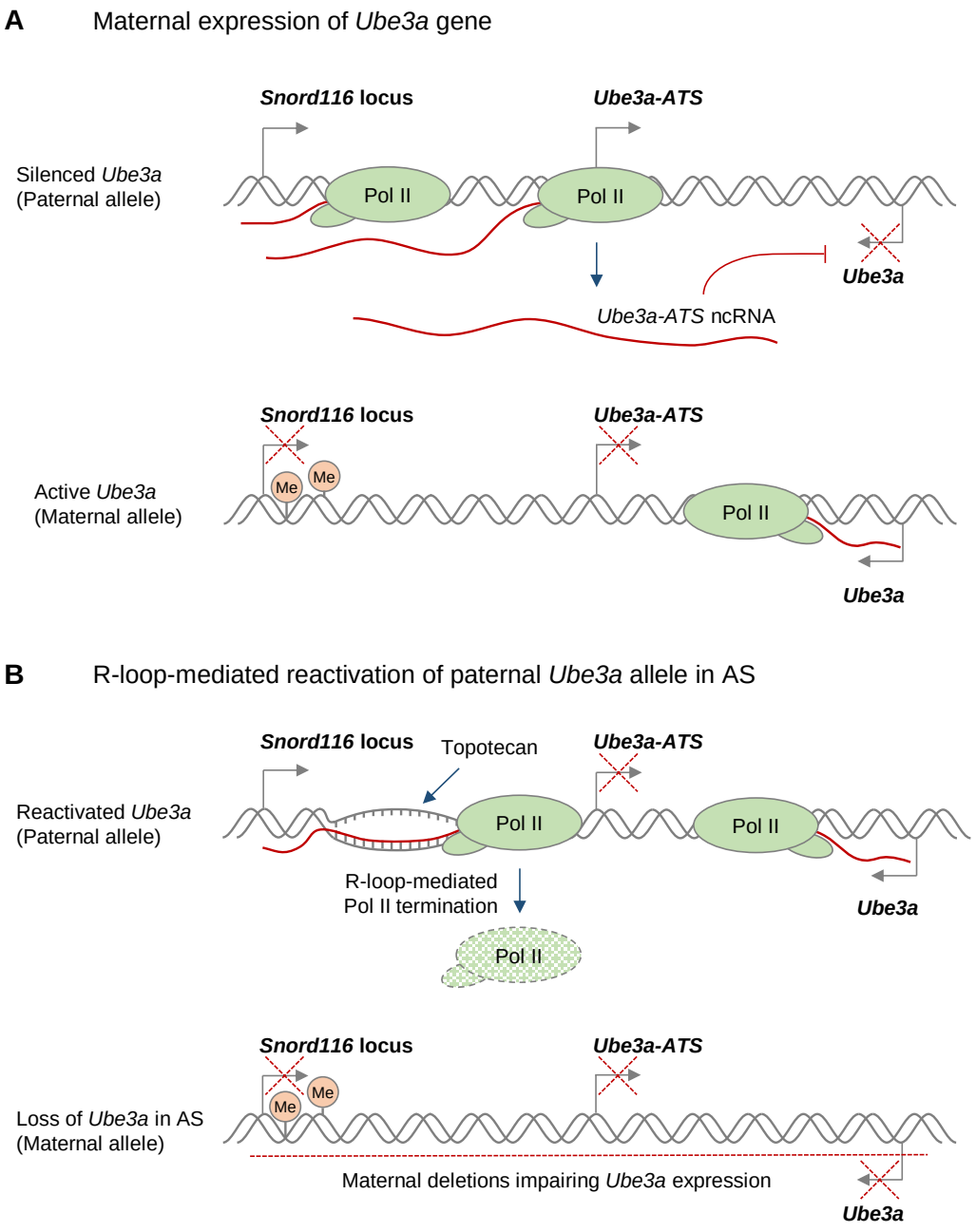


Figure 1.4. R-loops in Angelman syndrome (AS)

A. Neuronal expression of the paternal ncRNA *Ube3a-ATS* represses paternal *Ube3a* gene *in cis*. DNA methylation of the *Snord116* locus on the maternal allele prevents *Ube3a-ATS* transcription, resulting in *Ube3a* expression from the maternal allele. Transcriptional repression is indicated by red crosses.

B. R-loop-mediated reactivation of silent paternal *Ube3a* gene may provide a targeted therapy for AS. Deletion leading to the loss of maternal *Ube3a* expression detected in AS is indicated by the red dashed line. Topotecan treatment increases R-loop levels over the *Snord116* locus, resulting in chromatin decondensation, inhibition of RNA Pol II transcription through *Ube3a-ATS* and increased expression of *Ube3a* from the paternal allele (Powell *et al.*, 2013).

increased DNA sensitivity to damaging agents, formation of DSBs, chromosome breaks, fragile site instability, chromosome loss and recombination events (Hamperl and Cimprich, 2014).

Several mechanisms have therefore evolved to maintain R-loop levels in balance, and alterations in genome caretaker processes can affect R-loop levels and genome stability (Aguilera and Garcia-Muse, 2012). Moreover, mutations in proteins controlling R-loop levels have been identified in tumours (Figure 1.5A). For example, in eosinophilic leukemia, an oncogenic translocation renders cleavage and polyadenylation factor FIP1L1 inactive, which has been previously shown to cause increased R-loop levels, DNA damage and chromosome instability (Figure 1.5A) (Stirling *et al.*, 2012). A similar mechanism was suggested for RNA kinase CLP1, which is associated with a translocation in mixed lineage leukemia (MLL) (Stirling *et al.*, 2012). The histone ubiquitin ligase BRE1 also limits R-loop levels, and its decreased expression may contribute to the high levels of genomic instability observed in testicular seminoma (Chernikova *et al.*, 2012).

The link between R-loops and cancer has been further substantiated by the finding that the tumour suppressors BRCA1 and BRCA2, which are mutated in breast and ovarian cancer, are required to prevent R-loop accumulation and genome instability (Bhatia *et al.*, 2014; Hatchi *et al.*, 2015). Moreover, two recent studies demonstrated that R-loops are a major source of replication stress and DNA damage in proliferating cells and are resolved by factors of the Fanconi anemia (FA) pathway (Garcia-Rubio *et al.*, 2015; Schwab *et al.*, 2015). Therefore, R-loop-mediated genome instability may contribute to the increased cancer predisposition of individuals with mutations in FA proteins (Garcia-Rubio *et al.*, 2015; Schwab *et al.*, 2015).

These observations raise the interesting possibility that R-loops may provide proliferative advantages to tumour cells by promoting genome instability. This will in turn increase the probability of accumulating mutations favourable to tumour growth and metastasis. Intriguingly, recent evidence suggests that human oncogenic viruses may also promote genomic instability through accumulation of R-loops after infection. Kaposi's sarcoma-associated herpesvirus (KSHV), which causes multiple AIDS-related cancers, encodes ORF57 protein, which can sequester the host hTREX complex, important for mRNA processing and export (Jackson *et al.*, 2014). Sequestration of hTREX leads to

KSHV-induced accumulation of R-loops and causes damage to the host DNA, contributing to tumorigenesis (Jackson *et al.*, 2014).

Whilst some proteins suppress R-loop formation, others may promote R-loops and so increase genome instability, leading to tumour development. This unexpected function has been shown in yeast for transcription elongation factor Spt2 and DNA repair protein Rad51 (Sikdar *et al.*, 2008; Wahba *et al.*, 2013). Overexpression of Spt2 leads to transcription-dependent chromosomal rearrangements, which are prevented by RNase H overexpression (Sikdar *et al.*, 2008). Spt2 is structurally related to human HMG1, which is overexpressed in gastric cancers and malignant melanomas (Sikdar *et al.*, 2008). However, it is not clear if increased HMG1 levels promote R-loops and DNA damage in cancer cells.

In contrast to its well-established role in DNA strand exchange during homologous recombination and DNA repair (San Filippo *et al.*, 2008), recent studies have shown that Rad51 can also mediate R-loop formation and genome instability *in trans*, extending the prevailing view that R-loops form co-transcriptionally (Wahba *et al.*, 2013). Similar to HMG1, RAD51 is overexpressed in human cancers (Wahba *et al.*, 2013). However, it remains to be elucidated if RAD51 overexpression in cancers is a consequence of activated DNA repair pathways, or a cause of genome instability (Wahba *et al.*, 2013).

As discussed in Chapter 1.1.3, R-loops are implicated in *Ig* class switch recombination. Although this beneficial process is essential for generation of antibody diversity, AID-mediated mutagenesis has also been implicated in pathological translocations between the *Ig* loci and other transcriptionally active genes, leading to production of fusion proteins or oncogenic gene expression, observed in B cell malignancies (Figure 1.5B) (Robbiani and Nussenzweig, 2013). Interestingly, R-loops are also found in common translocation partners of *Ig* genes, including the oncogene *c-MYC* (Ginno *et al.*, 2012; Yang *et al.*, 2014). Therefore, the simultaneous formation of R-loops in *Ig* and transcribed non-*Ig* genes may induce AID-mediated DSB formation, leading to pathological translocations (Figure 1.5B) (Duquette *et al.*, 2005; Ruiz *et al.*, 2011; Yang *et al.*, 2014). Moreover, overexpression of the APOBEC family of AID-related enzymes in breast cancer have been linked to

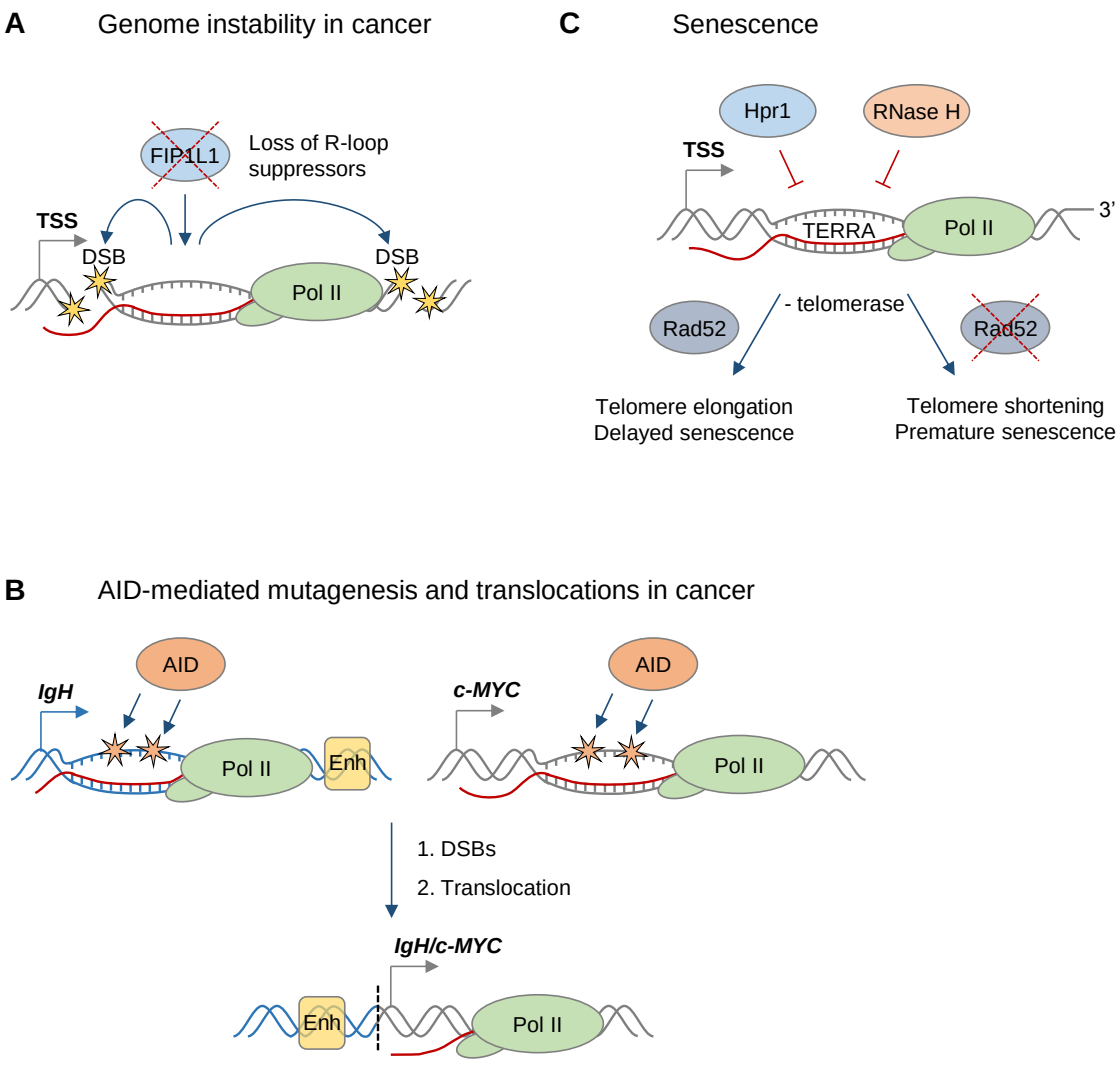


Figure 1.5. Roles of R-loops in genome instability, cancer and senescence

A. Loss of proteins preventing abnormal R-loop accumulation such as FIP1L1 leads to genome instability, one hallmark of cancer. Yellow stars denote double-stranded DNA breaks (DSBs).

B. Single-stranded DNA in R-loops is a substrate for cytidine deamination by activation-induced cytidine deaminase (AID), leading to mutagenesis as indicated by orange stars. These mutations can cause DSB formation, resulting in chromosomal translocations. The *IgH/c-MYC* translocation brings the strong *IgH* enhancers, shown as yellow box, close to *c-MYC*, leading to its overexpression in Burkitt's lymphoma. Transcription of *IgH/c-MYC* starts from a previously inactive promoter downstream of the translocation breakpoint. The *IgH* locus is depicted in blue, *c-MYC* gene is in grey. The translocation breakpoint is indicated by a dashed black line.

C. R-loops formed by the non-coding RNA TERRA accumulate at telomeres in cells deficient of Hpr1 and RNase H. In the absence of telomerase, these R-loops promote Rad52-dependent telomere elongation and delayed senescence. In the absence of telomerase and Rad52, R-loops promote telomere shortening and premature senescence.

genomic mutations, pointing to a potentially broader role of R-loops and AID/APOBEC-mediated genome instability in cancer (Burns *et al.*, 2013).

Despite the clear evidence linking R-loops with genome instability, the underlying mechanism remains poorly understood. Conflicts between replication and transcription have been proposed to serve as major source of R-loop-driven DNA damage (Helmrich *et al.*, 2011; Tuduri *et al.*, 2009). However, R-loop-associated DNA damage also occurs in non-replicating cells (Sordet *et al.*, 2009). The exposure of ssDNA present in R-loops to mutagenic molecules or proteins has been suggested to account for this observation (Skourti-Stathaki and Proudfoot, 2014). Additionally, R-loops themselves may be recognised by DNA repair factors and erroneously processed into DNA breaks, leading to genome instability if left unrepaired (Sollier *et al.*, 2014).

Changes in gene expression are another central aspect of cancer (Hanahan and Weinberg, 2011). In healthy cells, the expression of tumour suppressor genes prevents abnormal proliferation and other aspects of tumorigenesis (Hanahan and Weinberg, 2011). Tumour suppressors are frequently silenced in cancer by excessive promoter DNA methylation (Kulis and Esteller, 2010). It has been proposed that R-loop formation at promoters protects against DNA methylation by *de novo* DNA methyltransferase DNMT3B, thereby keeping genes active (Ginno *et al.*, 2012). Since R-loops have been computationally predicted to form at promoters of tumour suppressor genes *BRCA1*, *RASSF1A* and *CDKN2A* (Wongsurawat *et al.*, 2012), it is important to investigate if R-loop levels at these genes are reduced in cancer and how this relates to the observed DNA hypermethylation.

In contrast to this, efficient transcription of the oncogene *c-MYC* requires that R-loop levels are kept low by the activity of DNA topoisomerase IIIB, which is recruited to arginine-methylated histones by the TDRD3 protein (Yang *et al.*, 2014). This R-loop-mediated mechanism of *c-MYC* gene regulation may be relevant to tumour progression in breast cancer, which frequently shows overexpression of both *c-MYC* and TDRD3 (Hynes and Stoelzle, 2009; Yang *et al.*, 2014). Therefore, it is tempting to speculate that increased TDRD3 levels suppress R-loops in *c-MYC*, thereby allowing its enhanced expression, which correlates with poor cancer prognosis (Hynes and Stoelzle, 2009). However, it still remains to be determined if R-loops play a specific role in transcriptional

dysregulation in cancer and if this process differs from R-loop-mediated transcriptional programmes associated with housekeeping genes.

In conclusion, multiple lines of evidence point to an involvement of R-loops in cancer biology. Yet it still remains to be investigated if R-loop levels are indeed regulated differentially in normal and tumour tissues and if they can directly influence tumorigenesis.

1.2.5 R-loops in cellular senescence

More recently, R-loops have been implicated in cellular senescence, a mechanism protecting against tumour cell proliferation (Hanahan and Weinberg, 2011). In particular, the telomeric ncRNA TERRA forms R-loops which are induced when R-loop suppressors such as RNase H or Thp2 are lost (Balk *et al.*, 2013; Pfeiffer *et al.*, 2013). In the absence of telomerase, telomeric R-loops promote recombination-mediated telomere elongation via Rad52, and this delays the onset of cellular senescence (Balk *et al.*, 2013). In contrast, in Rad52-deficient cells, R-loop accumulation leads to telomere shortening and premature senescence (Balk *et al.*, 2013). Interestingly cells from AOA2 patients with SETX mutations contain shorter telomeres, suggesting a possible involvement of SETX in telomere stability (De Amicis *et al.*, 2011). Telomeric R-loops therefore play a complex and dynamic role in telomere length maintenance and cellular proliferative potential (Figure 1.5C).

1.3 Repeat expansion diseases

1.3.1 Transcriptional dysregulation in expansion diseases

Around forty human diseases are associated with expanded repeat sequences found in coding and non-coding gene regions (Figure 1.6) (Lopez Castel *et al.*, 2010). The expansion can be located in the coding part of a gene, causing the production of a toxic protein containing stretches of glutamines, as in Huntington's disease (HD) (Lopez Castel *et al.*, 2010). If the expansion lies within the noncoding part of a gene, transcription can be affected or toxic RNA species are produced, as observed in fragile X syndrome (FXS) and myotonic dystrophy 1 (DM1), respectively (Kumari *et al.*, 2012). It is important to note that these classic distinctions between coding and noncoding repeat expansion diseases may not be as clear as previously thought, and multiple mechanisms may contribute to the pathology of a single repeat expansion disease, as outlined below.

A hallmark characteristic of repeat expansion diseases is that the size of expansion increases during intergenerational transmission and in somatic tissues and correlates with disease severity. Many of these diseases are neurodegenerative in their nature. Cell division and DNA replication are known to contribute to repeat instability in mitotic cells (Zhang *et al.*, 2012). Interestingly, the size of expansion also increases in post-mitotic neuronal cells in the absence of DNA replication (Gonitel *et al.*, 2008). This observation raises the importance of RNA Pol II transcription through expanded repeats in mediating repeat instability and disease pathology.

The expanded repeats can affect a variety of transcription-associated processes, including RNA splicing, formation of R-loops, production of antisense, short non-coding and bidirectional RNA transcripts (Figure 1.7) (Groh *et al.*, 2014b). Expanded GAA repeats are known to directly inhibit RNA Pol II transcription *in vitro* (Bidichandani *et al.*, 1998; Grabczyk and Usdin, 2000b) and on reporter constructs in bacteria and mammalian cells (Krasilnikova *et al.*, 2007; Ohshima *et al.*, 1998). In line with these *in vitro* data, expanded GAA repeats cause transcriptional repression of the *FXN* gene, the cause of Friedreich ataxia (see Chapter 1.3.6) (Campuzano *et al.*, 1996; Herman *et al.*, 2006; Kim *et al.*, 2011; Kumari *et al.*, 2011). Other diseases, including DM1, characterised by CTG expansion in the *DMPK* gene, and amyotrophic lateral sclerosis (ALS) and frontotemporal dementia

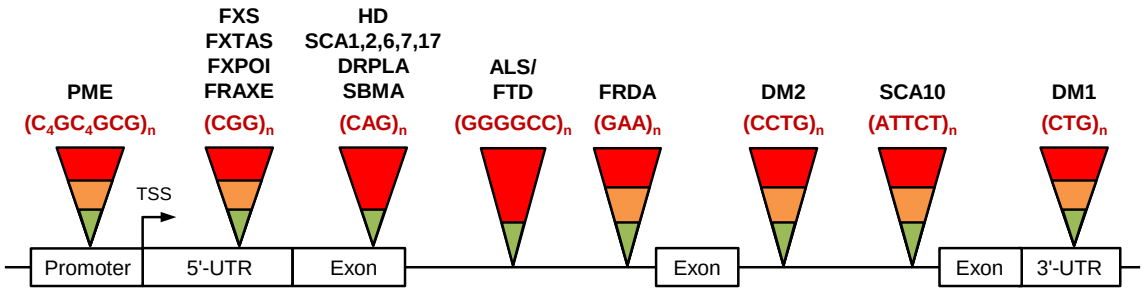


Figure 1.6. Location of repeat expansions causing human diseases

Diagram representing a generic gene indicating the location of disease-causing repeats. Triangles above the gene depict length of repeats: alleles without pathological impact are shown in green, premutation alleles are shown in orange and full mutation alleles are shown in red. Indicated above the triangles are the sequences of the expanded repeats and associated human diseases. PME, progressive myoclonus epilepsy; FXS, fragile X syndrome; FXTAS/FXPOI, fragile X-associated tremor and ataxia syndrome/primary ovarian insufficiency; FRAXE, fragile X E syndrome; HD, Huntington's disease; SCA, spinocerebellar ataxia; DRPLA, dentorubral-pallidoluysian atrophy; SBMA, spinal bulbar muscular atrophy; ALS/FTD, amyotrophic lateral sclerosis/frontotemporal dementia; FRDA, Friedreich ataxia; DM1/2, myotonic dystrophy 1/2. Diagram was adapted from Kumari and Usdin, 2012.

(FTD), triggered by the GGGGCC expansion in *C9orf72* gene, have also been associated with reduced RNA Pol II transcription (Brouwer *et al.*, 2013; Haeusler *et al.*, 2014).

The rate of RNA Pol II transcription elongation through expanded repeats is an important factor contributing to disease pathology. The CAG expansion in the *HTT* gene in Huntington's disease, results in production of toxic huntingtin protein (*mHTT)*. Depletion of Spt4/Supt4, a transcription elongation factor that aids RNA Pol II processivity by reducing dissociation of RNA Pol II from the DNA template, reduces *mHTT* protein in neuronal cells and decreases its aggregation and toxicity (Liu *et al.*, 2012). Moreover, the process of transcription has recently emerged as an essential player in modulating repeat instability in *S. cerevisiae* and mammalian cells (Goula *et al.*, 2012; Zhang *et al.*, 2012). Interestingly, GAA expansions can act as independent promoters and initiate transcription in yeast, potentially resulting in production of antisense or bidirectionally-transcribed RNAs, which can have pathological functions (Zhang *et al.*, 2012).

Expanded repeats can also mediate regulation of transcription *in trans*. In particular, in HD, the *mHTT* protein can bind a number of transcription factors and promote transcriptional down-regulation of many genes, associated with a decrease in histone acetylation, an effect which can be reversed by treatment with HDAC inhibitors (Sadri-Vakili *et al.*, 2007).

1.3.2 Aberrant splicing in expansion diseases

RNA-containing expanded repeats can sequester splicing factors, thereby affecting splicing globally. In the well-studied case of DM1, RNA containing expanded CUG repeats sequesters MBNL1 protein, which leads to changes in splicing of MBNL1-regulated genes such as *IGF1R*, *CLCA1* and *TNNC2*, implicated in the clinical symptoms of DM1 (Echeverria and Cooper, 2012). Similar to CUG repeats, other expanded sequences, associated with SCA3, SCA8 and FXTAS were also shown to affect splicing *in trans* (Echeverria and Cooper, 2012; Krzyzosiak *et al.*, 2012).

It has been demonstrated that the speed of RNA Pol II transcription elongation can trigger changes in alternative splicing (de la Mata *et al.*, 2003). Therefore, transcription elongation defects caused by expanded repeats could affect splicing *in cis*. Furthermore, splicing can be affected by an increase in intron or exon size or by sequence-specific effects (Shishkin *et al.*, 2009). In particular,

expanded GAA repeats inserted in reporter constructs trigger aberrant splicing in a position- and context-dependent manner (Baralle *et al.*, 2008). These GAA-containing RNAs can also bind different splicing factors and promote accumulation of pre-mRNA splicing intermediates, which are not turned over into mature mRNAs (Baralle *et al.*, 2008). Similarly, expanded CAG repeats in the *HTT* gene affect its splicing, possibly through aberrant binding of splicing factor SRSF6, leading to production of short polyadenylated CAG-repeat containing mRNAs, which are translated into small toxic poly-glutamine proteins (Sathasivam *et al.*, 2013).

1.3.3 Antisense and bidirectional transcription over expanded repeats

Recent years have produced growing evidence that antisense transcription is widespread in the human genome and can play an important role in regulating gene expression (Pelechano and Steinmetz, 2013). Antisense transcripts have been detected in many triplet repeat-containing genes and hence may contribute to disease pathology. Consequences of antisense transcription on expanded repeats include regulation of sense transcription by transcriptional interference, production of toxic proteins and small repeat-containing RNAs that lead to post-transcriptional gene silencing.

Antisense RNAs can be important for regulating expression of repeat-containing genes. In particular, an antisense *HTTAS* transcript spanning the CAG repeats negatively regulates sense *HTT* transcript levels, potentially through transcriptional interference (Chung *et al.*, 2011). Similarly, in the *ATXN7* gene a CAG-repeat spanning antisense *SCAANT1* transcript negatively regulates *ATXN7* sense transcript levels *in cis* (Sopher *et al.*, 2011). Toxic properties of antisense transcripts have been demonstrated for Spinocerebellar Ataxia 8 (SCA8) and Huntington disease-like 2 (HDL2). In particular, in SCA8 bidirectional transcription of repeats leads to the production of two independent repeat-containing RNAs with different pathological properties. Expression of CAG-containing *ATXN8* RNA gives rise to a toxic polyQ protein (Moseley *et al.*, 2006). Conversely, transcription of the opposite strand produces long non-coding CUG-containing *ATXN8OS* RNA, leading to RNA-mediated toxicity, similar to DM1 (Echeverria and Cooper, 2012). A CAG-repeat containing antisense RNA described for the *JPH3* gene also leads to a production of a polyQ protein, promoting HDL2 pathology even in the absence of sense transcripts (Wilburn *et al.*, 2011).

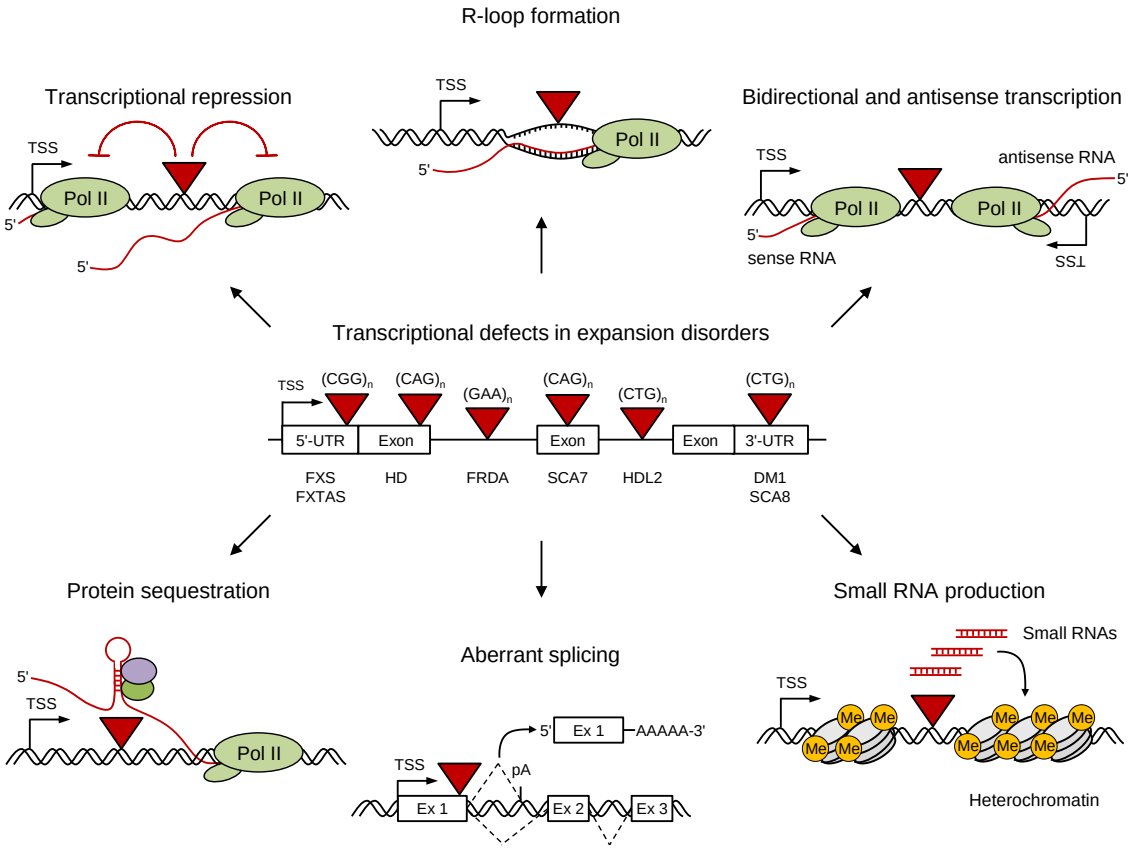


Figure 1.7. Role of transcription in the pathology of repeat expansion diseases

A generic gene diagram shows expanded sequences within 5' and 3'UTRs, exons and non-coding sequences. Introns are depicted by black lines and exons by white boxes. Diseases associated with specific expansions discussed in this Chapter are presented below the gene diagram. The location of the expansion is indicated by a triangle. TSS denotes transcriptional start site. DNA is shown as double helix and RNA is represented as a red line. Transcription through an expansion can result in defects in RNA Pol II elongation and initiation, R-loop formation, antisense and bidirectional transcription, production of small non-coding RNAs, aberrant RNA splicing and production of toxic structured RNAs.

Emerging evidence for DM1, HD and SCA1 suggests that one prominent outcome of bidirectional transcription of triplet repeats is the production of small repeat-containing RNAs. Research into DM1 has primarily focused on the sense transcription, which produces a CUG repeat-containing RNA that can sequester splicing proteins (see Chapter 1.3.2). However, more recently it has been shown that the CTG repeats in *DMPK* are also transcribed in the antisense direction (Cho *et al.*, 2005). This repeat-spanning bidirectional transcription is locally confined by CTCF binding and allows formation of double-stranded RNA, which is processed into small RNAs and is likely to play a role in establishment of heterochromatin (Cho *et al.*, 2005).

Similarly, bidirectional transcription of CAG repeats leads to a Dicer-dependent production of small CAG RNAs, adding to the complexity of HD and SCA1 pathologies, thought to be primarily caused by production of toxic proteins (Lawlor *et al.*, 2011). Repeat-containing small RNAs can also be produced from RNA hairpins, which contain double-stranded stretches. Such repeat-containing small RNAs can also be processed by Dicer *in vitro* and *in vivo,* and their expression levels correlate with expansion size (Banez-Coronel *et al.*, 2012; Handa *et al.*, 2003; Krol *et al.*, 2007; Krzyzosiak *et al.*, 2012).

Irrespective of their origin, small repeat RNAs can either affect the expression of the host gene itself or regulate other genes that contain repeat-complementary sequences. In particular, small repeat RNAs from the *DMPK* gene mediate transcriptional repression *in cis*, an effect which is also observed in HD and SCA1 (Cho *et al.*, 2005; Krol *et al.*, 2007). It is possible that transcriptional repression specific to expanded alleles could be a cellular surveillance mechanism that prevents excessive accumulation of repeat-containing RNA and proteins. However, unwanted targeting of homologous sequences *in trans*, causing dysregulation of other genes, may also contribute to the disease pathology. In line with this, CUG repeat-containing small RNAs have been shown to repress CAG repeat-containing genes such as *Ataxin-2* and *TBP*, while CAG-derived siRNAs can repress CTG-containing genes such as *ADORA2A* and *MEIS2* (Banez-Coronel *et al.*, 2012; Yu *et al.*, 2011).

Interestingly, bidirectional transcription may also play a role in triplet repeat instability. Accordingly, both expanded CAG and CTG repeats exhibit increased DNA instability if transcribed from both strands (Lin *et al.*, 2010; Nakamori *et al.*, 2011). Therefore, the somatic repeat instability

observed in non-dividing cells in DM1, HD and SCA1 could be connected to ongoing bidirectional transcription of expanded repeats.

In a number of triplet expansion disease genes, antisense transcripts have been identified but their role in pathology remains elusive. The *FMR1* gene contains a repeat-spanning *ASFMR1* antisense transcript that shows co-regulation with the sense *FMR1* transcript. Thus, *ASFMR1* transcription is silenced in FXS, while its expression is increased in fragile X/tremor ataxia syndrome (FXTAS), mirroring *FMR1* mRNA levels (Ladd *et al.*, 2007). Although the function of the *ASFMR1* transcript is not clear, increased *FMR1* and *ASFMR1* levels in patients correlate with loss of mitochondrial DNA and Parkinsonism, suggesting that bidirectional transcription of CGG repeats could contribute to cellular toxicity (Loesch *et al.*, 2011).

1.3.5 Fragile X syndrome (FXS)

Pathophysiology of fragile X syndrome

The genetic cause of fragile X syndrome (FXS) is the expansion of a CGG repeat located in the 5'UTR of the *FMR1* gene (Figure 1.8). The *FMR1* gene encodes FMRP protein, an RNA-binding factor which regulates the translation of a large set of neuronal mRNAs (Santoro *et al.*, 2011).

The sizes of CGG repeats are polymorphic in the unaffected population and reach up to 55 repeats without impacting *FMR1* gene function (Kumari *et al.*, 2012). Pathological *FMR1* alleles can be separated in two different classes based on the size of the CGG repeats (Figure 1.8). *FMR1* alleles with 55–200 CGG repeats are referred to as premutation (PM) alleles and are characterised by increased *FMR1* mRNA expression (Tassone *et al.*, 2007). In contrast, full mutation (FM) *FMR1* alleles with >200 CGG repeats lead to transcriptional silencing of the *FMR1* gene and loss of FMRP protein. Consistent with their opposite effects on *FMR1* gene expression, PM and FM alleles confer risks to develop different diseases. Carriers of PM alleles are at risk to developing fragile X-associated tremor/ataxia syndrome (FXTAS) and females can also develop fragile X-associated primary ovarian insufficiency (FXPOI) (Hagerman, 2013). FM alleles are associated with fragile X syndrome, the leading heritable cause of mental retardation and the major genetic cause of autism (Santoro *et al.*, 2011).

Molecular pathology of FXTAS and FXPOI

Extensive evidence supports a gain-of-function effect due to *FMR1* mRNA-mediated toxicity as the primary cause of FXTAS and FXPOI (Figure 1.8) (Hagerman, 2013). It has been shown that the increased expression of *FMR1* mRNA from PM alleles is due to increased transcription, not altered mRNA stability (Tassone *et al.*, 2007). PM alleles exhibit higher levels of histone acetylation (Todd *et al.*, 2010). However, it is less clear how PM alleles promote these epigenetic and gene expression changes. CGG repeats exclude nucleosomes *in vitro*, which could allow higher accessibility for the transcription machinery to the *FMR1* promoter *in vivo* (Wang *et al.*, 1996). Furthermore, unmethylated CpG dinucleotides are recognised by CxxC zinc finger proteins, which can recruit transcriptional activators such as H3K4 methyltransferases (Thomson *et al.*, 2010). The increased number of unmethylated CpGs present in the PM alleles may therefore serve as a binding platform for CxxC transcriptional activators. It has recently been demonstrated that R-loops form in PM alleles (Loomis *et al.*, 2014). Although it has not been investigated yet, these R-loops may contribute to the increased *FMR1* expression of PM alleles, similar to their function in maintaining other G-rich promoters in epigenetically active states (see Chapter 1.1.4) (Chen *et al.*, 2015a; Ginno *et al.*, 2012). However, direct experimental support for any of these mechanisms is currently not available.

Despite these open questions, it has been suggested that the main pathological consequence of accumulated CGG repeat-containing *FMR1* mRNA is the sequestration of several RNA-binding proteins, including hnRNP A2/B1, Purα, Sam68 and DGCR8 (Hagerman, 2013). In particular, the sequestration of Sam68 by expanded CGG repeat RNA leads to changes in alternative splicing (Sellier *et al.*, 2010). Similarly, sequestration of miRNA biogenesis factor DGCR8 by expanded CGG RNA is associated with impaired miRNA processing which leads to globally reduced miRNA levels, causing neuronal toxicity (Sellier *et al.*, 2013).

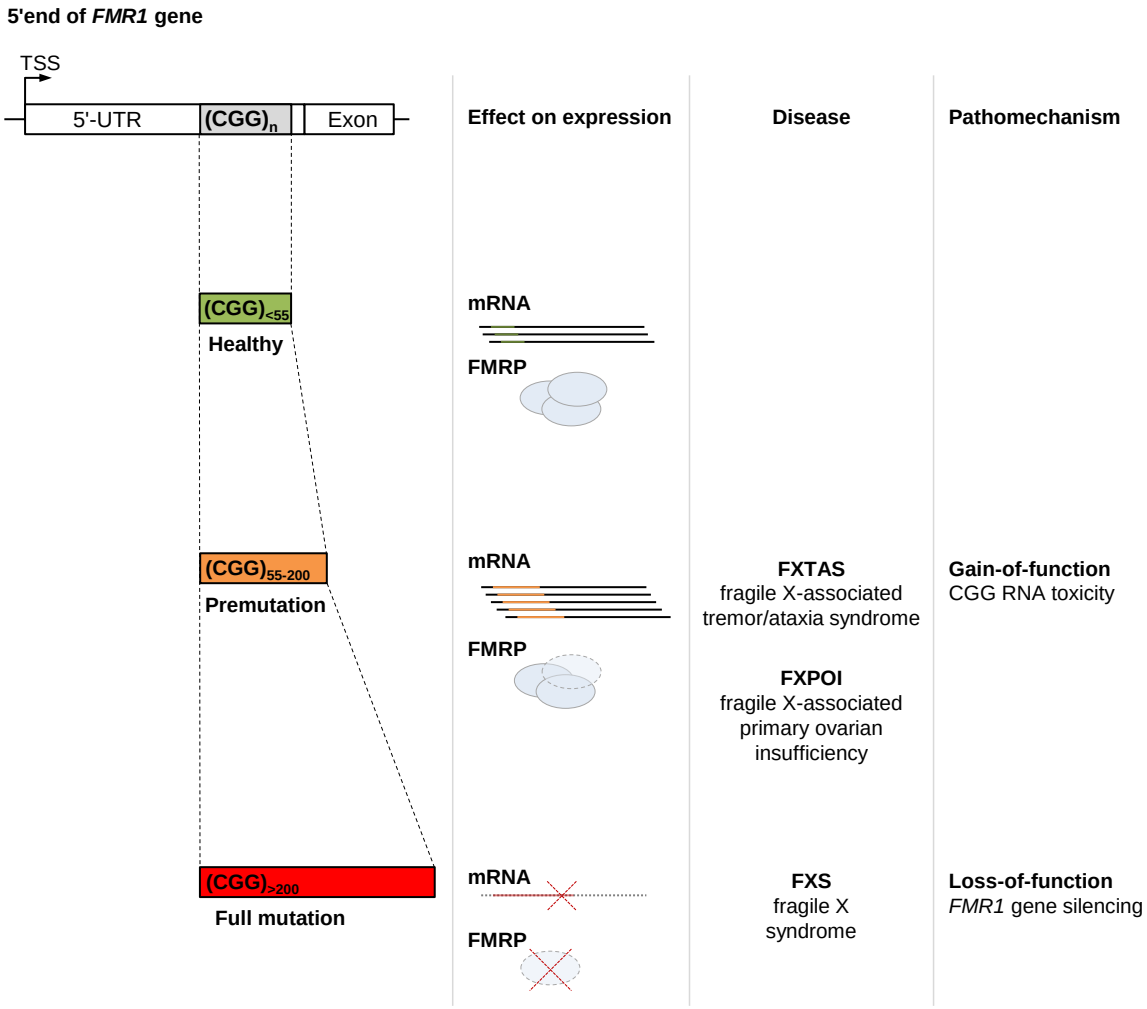


Figure 1.8. Pathomechanism of diseases caused by expanded CGG repeats in *FMR1* gene

An overview of the effects of the size of the CGG repeat located in the 5'UTR of the *FMR1* gene on its expression and human pathology. The diagram depicts the 5' end of the *FMR1* gene. *FMR1* RNA is shown as lines with the location of the CGG repeat indicated in green (unexpanded allele), orange (premutation allele) or red (full mutation allele). Loss of *FMR1* mRNA and FMRP is indicated by red crosses. Premutation alleles lead to overexpression of a toxic CGG repeat-containing RNA that can sequester cellular proteins. Full mutation alleles usually produce little or no mRNA and FMRP, causing translational de-repression of FMRP target mRNAs. Diagram adapted from Kumari and Usdin, 2012.

The contribution of FMRP protein levels to FXTAS and FXPOI pathologies is not fully understood. Despite significantly increased *FMR1* mRNA expression in PM carriers, FMRP levels are either unchanged or decreased compared to unaffected individuals (Hagerman, 2013). It has been suggested that decreased FMRP levels may be associated with additional clinical symptoms in some FXTAS patients (Hagerman, 2013).

Molecular pathology of FXS

In contrast to FXTAS, FM alleles of *FMR1* are hypoacetylated and enriched for repressive epigenetic marks, including promoter DNA methylation, H3K9me2 and H3K27me3 (Coffee *et al.*, 1999; Kumari and Usdin, 2010; Santoro *et al.*, 2011). The sequence of events leading to *FMR1* silencing in FXS has not been fully elucidated yet. However, DNA methylation plays a major role in the process, as DNA methyltransferase inhibitors increase histone acetylation and cause partial transcriptional reactivation of FM alleles (Coffee *et al.*, 1999). It has been shown that directly targeting histone deacetylases such as SIRT1 can also alleviate the transcriptional silencing of FM alleles, without concomitant DNA demethylation (Biacsi *et al.*, 2008).

Interestingly, inhibiting DNA methylation in FXS patient cells leads to recruitment of Polycomb repressive complex 2 (PRC2) and enrichment of H3K27me3 on FM alleles, dependent on the presence of *FMR1* mRNA (Kumari and Usdin, 2014). This may explain the inability to fully and persistently reactivate the FM allele since the CGG expansion-containing *FMR1* mRNA may be part of a negative feedback loop, causing its own repression. In line with this, the knockdown of *FMR1* mRNA in differentiating embryonic stem cells prevents *FMR1* gene silencing (Colak *et al.*, 2014). These findings suggest that the CGG expansion-containing *FMR1* transcript is the initial trigger of the transcriptional silencing, potentially by tethering transcriptional repressors to the *FMR1* gene.

The consequences of *FMR1* gene silencing and the loss of its protein product are most severe in neurons and testes, where FMRP protein levels are highest in unaffected individuals (Santoro *et al.*, 2011). FMRP binds selectively to target mRNAs, mediated by RNA secondary structures including G4 quadruplexes (Santoro *et al.*, 2011). Many of the biologically validated target mRNAs are localised in dendrites and play a role in synaptic transmission and neuronal development (Santoro

et al., 2011). In unaffected individuals, FMRP represses its target mRNAs translationally in non-excited neurons but allows their translation upon activation of the metabotropic glutamate receptor (mGluR) (Santoro *et al.*, 2011). In the absence of FMRP in FXS patients, basal protein levels of FMRP targets are elevated and cannot be further induced upon mGluR signalling. This is thought to be the major cause of the impaired synaptic plasticity associated with learning and memory disabilities observed in FXS patients (Santoro *et al.*, 2011).

1.3.6 Friedreich ataxia (FRDA)

Pathophysiology of Friedreich ataxia

Friedreich ataxia (FRDA) is an autosomal recessive disorder characterised by progressive neurodegeneration and was first described in 1863 by the German pathologist Nikolaus Friedreich (Friedreich, 1863). With a prevalence of 1 in 50000, FRDA is the most common recessive ataxia (Kumari *et al.*, 2012). FRDA patients suffer from early-onset progressive ataxia, sensory loss, muscle weakness and hypertrophic cardiomyopathy, leading to a significantly reduced life expectancy (Marmolino, 2011).

The genetic cause for FRDA is the reduced expression of frataxin protein, due to impaired expression of the frataxin (*FXN*) gene (Campuzano *et al.*, 1996). In the large majority of cases, impaired *FXN* gene expression is due to the homozygous expansion of a GAA repeat present in the first intron of the *FXN* gene (Figure 1.9) (Campuzano *et al.*, 1996). Although repeat lengths of ~35 GAAs already cause FRDA symptoms, most expansion alleles contain 500-1000 GAA repeats (Marmolino, 2011). In a small number of heterozygous patients (~4%), FRDA is caused by a GAA repeat expansion in one *FXN* allele and a loss-of-function mutation in the other (Campuzano *et al.*, 1996). In homozygous FRDA patients, age of onset and severity of the disease correlates with the size of the GAA expansion (Durr *et al.*, 1996). The causative role of expanded GAA repeats in FRDA has recently been verified by nuclease-mediated removal of the repeats, leading to alleviation of the molecular FRDA phenotype (Li *et al.*, 2015). Importantly, all FRDA patients retain minimal levels of frataxin protein and no *FXN* null mutations have been described, suggesting that complete loss of frataxin is embryonically lethal (Marmolino, 2011). The frataxin protein is highly conserved

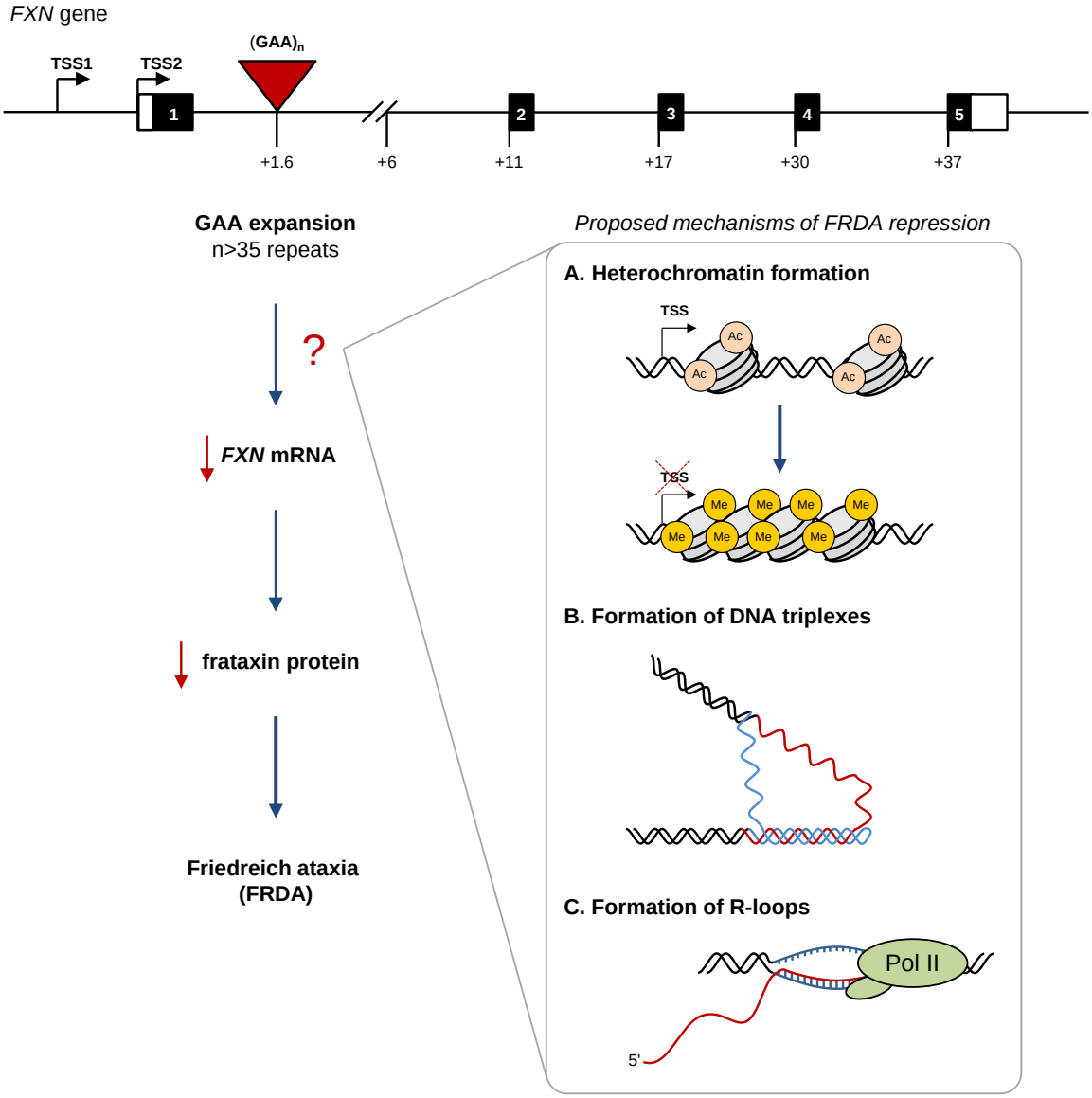


Figure 1.9. Pathomechanism of Friedreich ataxia (FRDA)

The diagram depicts the frataxin (*FXN*) gene. Black boxes are exons, white boxes are 5' and 3'UTRs, lines are introns, red triangle is the $(GAA)_n$ expansion. TSS2 is the major transcriptional start site in lymphoblastoid cells. Numbers indicate the distances to TSS2 in kilobases. Expansion of the GAA repeat leads to *FXN* gene transcriptional repression, loss of frataxin protein, causing FRDA. Mechanisms proposed to link GAA expansion with transcriptional repression are shown in the insert:

A. Formation of repressive heterochromatin. Yellow circles indicate enrichment of repressive histone methylation such as H3K9me2/3 and H3K27me3. Orange circles depict active histone acetylation lost in FRDA.

B. Formation of DNA triplexes by folding back of the non-template strand (containing GAA repeats, blue) to the upstream GAA:TTC dsDNA; the template strand remains single-stranded (red).

C. Formation of R-loops by hybridisation of nascent RNA (red) containing GAA repeats to the DNA (blue).

among eukaryotes and localises to mitochondria where it has been implicated in the biogenesis of iron-sulfur cluster-containing proteins, important for iron homeostasis, DNA replication and oxidative phosphorylation (Kumari *et al.*, 2012; Marmolino, 2011). It is not clear why only specific tissues including neurons of the cerebellum (dentate nuclei) and cardiomyocytes are most affected by the ubiquitous GAA expansion but it is possible that the high metabolic activity of these tissues makes them more susceptible to mitochondrial defects.

While the genotype-phenotype correlation in FRDA is well studied, the molecular mechanism underlying the loss of *FXN* expression due to expanded GAA repeats remains largely unknown.

Molecular pathology of Friedreich ataxia

The expansion of GAA repeats in the *FXN* gene is associated with reduced RNA Pol II transcription initiation and elongation, leading to decreased *FXN* mRNA levels (Kim *et al.*, 2011; Kumari *et al.*, 2011; Punga and Buhler, 2010). In line with a pathomechanism based primarily on aberrant transcription, no difference in the stability of *FXN* mRNA in cells from unaffected or FRDA individuals has been found (Punga and Buhler, 2010). Although initial studies have not detected aberrant splicing of *FXN* mRNA in FRDA (Bidichandani *et al.*, 1998; Punga and Buhler, 2010), it cannot be excluded that more sensitive and unbiased approaches may yet uncover a contribution of defective splicing to FRDA pathology. The possibility of *FXN* splicing defects is highlighted by experiments using splicing reporter systems in human cells and in *S. cerevisiae*, where insertion of expanded GAA repeats leads to complex alternative splicing defects and in some cases reduced mRNA levels (Baralle *et al.*, 2008; Shishkin *et al.*, 2009).

The formation of repressive heterochromatin by expanded GAA repeats has been proposed to be a key step leading to transcriptional repression of the *FXN* gene in FRDA (Figure 1.9) (Kumari *et al.*, 2012). In particular, *FXN* alleles containing expanded repeats are characterised by high levels of repressive histone modifications including H3K9me2/3 and H3K27me3 (Herman *et al.*, 2006; Kim *et al.*, 2011; Kumari *et al.*, 2011; Punga and Buhler, 2010). Additionally, expanded GAA repeats are associated with decreased levels of histone acetylation (Herman *et al.*, 2006). Importantly, the

pathological relevance of histone deacetylation in FRDA has been demonstrated since inhibition of histone deacetylases (HDACs) can partially reactivate *FXN* gene expression in FRDA cells (Chan *et al.*, 2013; Herman *et al.*, 2006; Soragni *et al.*, 2014). Although differential DNA methylation of CpG dinucleotides in the first intron of *FXN* gene has been reported in FRDA patient cells, its functional relevance is not clear (Al-Mahdawi *et al.*, 2008; Greene *et al.*, 2007). Despite growing evidence demonstrating the importance of heterochromatin formation in *FXN* repression, the initial events which trigger the formation of heterochromatin in FRDA have remained elusive. Three major mechanisms have been suggested: the direct recruitment of epigenetic repressors by GAA repeat DNA or RNA and the formation of unusual nucleic acid structures such as DNA triplexes or R-loops (Figure 1.9) (Kumari *et al.*, 2012; Wells, 2008).

1.3.7 R-loops in expansion-associated pathology

R-loops have been proposed to play a role in the pathology of repeat expansion diseases and in particular in FRDA (Grabczyk *et al.*, 2007; Lin *et al.*, 2010; McIvor *et al.*, 2010; Reddy *et al.*, 2011). Remarkably, R-loops are formed following transcription of trinucleotide repeats *in vitro*, in bacteria and on CAG repeats in human cells (Grabczyk *et al.*, 2007; Lin *et al.*, 2010; Reddy *et al.*, 2011). Furthermore, R-loops have recently been detected at premutation and full mutation CGG repeats on the endogenous *FMR1* gene (Colak *et al.*, 2014; Loomis *et al.*, 2014). It has been suggested that the non-template DNA strand in many repetitive sequences can adopt unusual DNA structures, including G-quadruplexes and DNA triplexes, which may further stabilise R-loops (Belotserkovskii *et al.*, 2013). Further substantiating a role for R-loops in these diseases, R-loops formed at CTG repeats promote repeat instability characteristic of these diseases (Lin *et al.*, 2010).

R-loops could contribute to the pathology of expansion diseases in various ways. Similar to R-loops at the 3'ends of human genes, expansion-associated R-loops may form a structural block, directly interfering with RNA Pol II transcription elongation (Huertas and Aguilera, 2003; Skourti-Stathaki *et al.*, 2011). Alternatively, R-loops may nucleate repressive chromatin over the expansion region, by analogy with heterochromatin formation at centromeres in *S. pombe* (Nakama *et al.*, 2012), or promote chromatin compaction associated with histone H3S10 phosphorylation, as

observed in *S. cerevisiae*, *C. elegans* and human cells (Castellano-Pozo *et al.*, 2013). Thus, although several studies have suggested that R-loops could be involved in human trinucleotide expansion disorders, there has not been direct evidence of their contribution to the molecular mechanisms underlying their pathology *in vivo*.

1.4 Aims of this work

Trinucleotide repeat expansion diseases are devastating neurological conditions with a severe impact on the lives of patients and their families. Despite much progress in identifying the mutated genes causing the diseases, the mechanisms linking repeat expansion and the resulting phenotypes are still not fully understood. A common theme in many repeat expansion diseases so far investigated is the dysregulation of transcription, leading to toxic gain-of-function or loss-of-function of the repeat-containing host gene. There is a great need to rationally develop novel therapies to treat repeat expansion diseases but this is hampered by our limited understanding of the underlying pathomechanisms.

1.4.1 Investigation of R-loop-mediated pathology in Friedreich ataxia

The first main aim of this project was to investigate the pathomechanism of Friedreich ataxia (FRDA), associated with a GAA expansion in the first intron of the *FXN* gene. FRDA is a genetically well-characterised noncoding repeat expansion disease, caused by loss of normal *FXN* gene expression. It therefore presented an ideal model system to study the consequences of expanded repeats on transcription *in vivo*.

I initially focused on investigating the regulation of *FXN* gene transcription in control and FRDA cells, which is described in Chapter 3. Furthermore, I describe the detection and characterisation of RNA/DNA hybrids in R-loops formed at expanded GAA repeats on the endogenous *FXN* gene *in vivo* and their enrichment in FRDA patient cells. To study the generality of RNA/DNA hybrid-mediated gene silencing in other expansion diseases, I also investigated RNA/DNA hybrid formation in another repeat expansion disease, fragile X syndrome (FXS).

In Chapter 4, I describe the impact of RNA/DNA hybrid formation on *FXN* gene transcription. I also present my investigation of the interplay of R-loops with repressive heterochromatin and its relevance for FRDA pathology. In particular, I demonstrated that RNA/DNA hybrids associated with expanded GAA repeats trigger heterochromatin formation and promote the transcriptional repression of the *FXN* gene associated with FRDA.

1.4.2 Characterisation of the RNA/DNA hybrid interactome

Proteins recognising RNA/DNA hybrids or R-loops could be potent disease modifiers and potential targets for therapeutic approaches aiming at modulating RNA/DNA hybrid levels. The second main aim of this project was therefore to identify proteins binding to RNA/DNA hybrids *in vivo*.

In Chapter 5, I describe the development and validation of an unbiased method to identify novel proteins binding to RNA/DNA hybrids in human cells, termed RNA/DNA hybrid IP. Using quantitative mass spectrometry, I defined a set of proteins constituting the RNA/DNA hybrid interactome in HeLa cells and characterised their properties bioinformatically. In particular, I found that the RNA/DNA hybrid interactome is distinct from the mRNA interactome of HeLa cells and is enriched for dual RNA/DNA binding proteins. I confirmed the specificity of the RNA/DNA hybrid IP biochemically and validated a range of positive and negative controls. To demonstrate the biological relevance of the RNA/DNA hybrid IP method, I also investigated the biological role of the top RNA/DNA hybrid interactome candidate, helicase DHX9, in RNA/DNA hybrid resolution *in vivo*.

In Chapter 6, I discuss the overall implications of my results described in Chapters 3–5 and provide directions for future research.

1.5 Publications

Parts of the work described in this thesis have been published in peer-reviewed journals.

Chapter 1

Groh, M.; Gromak, N. *Out of balance: R-loops in human disease.* **PLOS Genetics,** 2014. 10, e1004630. http://dx.doi.org/10.1371/journal.pgen.1004630

Groh, M.; Silva, L.M.; Gromak, N. *Mechanisms of transcriptional dysregulation in repeat expansion disorders*. **Biochemical Society Transactions**, 2014, 42, 1123-1128. http://dx.doi.org/10.1042/BST20140049

Chapters 3 and 4

Groh, M.; Lufino, M.P.; Wade-Martins, R.; Gromak, N. *R-loops Associated with Triplet Repeat Expansions Promote Gene Silencing in Friedreich Ataxia and Fragile X Syndrome***. PLOS Genetics**, 2014. 10, e1004318.

Chapter 2

Materials and Methods

2.1 Materials

2.1.1 Bacterial strains

For amplification of plasmid DNA, chemicompetent *E. coli* strains NM554 (*recA13 araD139 Δ(ara-leu)7696 Δ(lac)l7A galU galK hsdR rpsL (Strr) mcrA mcrB*) and DH5α (*F– Φ80lacZΔM15 Δ(lacZYA-argF) U169 recA1 endA1 hsdR17 (rK–, mK+) phoA supE44 λ– thi-1 gyrA96 relA1*) were obtained from Stratagene and Invitrogen, respectively.

2.1.2 Mammalian cell lines

Epstein-Barr virus (EBV)-transformed lymphoblastoid cell lines from healthy (GM15851, GM14926, GM06895), FRDA (GM15850, GM16243, GM16209) and FXS (GM03200) patients were obtained from Coriell Institute for Medical Research. Frataxin (GAA) repeat sizes were 650/1030 (GM15850), 800/800 (GM16209) and 670/1170 (GM16243). (CGG) repeat size in FMR1 gene was 530 (GM03200). HeLa cells were from Dr Natalia Gromak and have been used before (Gromak *et al.*, 2013; Skourti-Stathaki *et al.*, 2011). HEK293 FXN-Luc and FXN-GAA-Luc cells were a kind gift from Dr Michele Lufino (Lufino *et al.*, 2013).

2.1.3 Oligonucleotides and plasmids

Plasmids

The pRNaseH1-GFP plasmid was a kind gift of Prof Dr R. Crouch. It contains the RNaseH1 open reading frame (ORF) without its N-terminal mitochondrial localisation signal (MLS), fused to green fluorescent protein (GFP) at its C-terminus and has been described before (Cerritelli *et al.*, 2003). The pFlag plasmid, containing Flag ORF was from Sigma-Aldrich (E7398).

The pRNaseH1-Flag plasmid was created from pRNaseH1-GFP by PCR site-directed mutagenesis, replacing the C-terminal GFP with a Flag tag using pRNaseH1-GFP plasmid as template and pRNaseH1-Flag(F) and pRNaseH1-Flag(R) primers (Table 2.1). The plasmid was validated by sequencing using RNaseH1ex7(F) and RNaseH1ex8(R) primers.

The *pHIV-βGlobin* plasmid was provided by Dr N. Gromak (University of Oxford, UK). It contains the coding sequence of *β-Globin* under control of the human immunodeficiency virus (HIV)

promoter (pHIV) with the pause element from *β-Actin* gene (Dye and Proudfoot, 1999; Gromak *et al.*, 2006). The pTAT plasmid expressing HIV transactivator of transcription protein (TAT) was provided by Dr N. Gromak and was described previously (Dye and Proudfoot, 1999).

DNA oligonucleotides

All DNA oligonucleotides were synthesised and desalted by Sigma-Aldrich. Dried DNA oligonucleotides were resuspended in ddH_2O to final concentrations of 100 μM and stored at -20°C. All DNA oligonucleotides employed in this work are listed in Table 2.1.

RNA oligonucleotides

All RNA oligonucleotides were synthesised, deprotected and desalted by Dharmacon (GE Healthcare), except stated otherwise. RNA oligonucleotides were resuspended in RNase-free ddH_2O and stored at -20°C. All RNA oligonucleotides are listed in Table 2.2.

2.1.4 Buffers and solutions

Non-commercial buffers and solutions were prepared in-house using standard RNase- and DNase-free sterile techniques. Buffers and solutions were prepared with ddH_2O (18.2 MΩ) and autoclaved. All buffers employed in this work are listed in Table 2.3. Chemicals used to prepare the buffers were of molecular biology grade and obtained from Sigma-Aldrich and Roche.

2.1.5 Enzymes and other materials

Commercial buffers, reagents, materials, kits and enzymes not prepared in the laboratory were used according to the manufacturers' instructions and are listed in Table 2.4.

2.1.6 Antibodies

Antibodies were obtained from commercial suppliers or other sources as listed and stored according to the manufacturers' instructions, either aliquoted at -20°C or at 4°C. All antibodies employed in this work are listed in Table 2.5.

Table 2.1. List of DNA oligonucleotides

Sequences of DNA oligonucleotides are given in 5' to 3' direction. For oligonucleotides not designed in this work, the original publications are indicated.

Name	Sequence (5' to 3')
DNA oligonucleotides for PCR	
FXN-Luc expansion PCR *(Filla et al., 1996)*	
GAA104F	GGCTTAAACTTCCCACACGTGTT
GAA629R	AGGACCATCATGGCCACACTT
RNase H1-Flag cloning	
RNaseH1-Flag (F)	GATTACAAGGATGACGACGATAAGGTTTAAAGCGGCCGCGACTCTAGATCA
RNaseH1-Flag (R)	ATCGATCCCTCCACCGTCTTC
DNA oligonucleotides for RNA/DNA hybrid IP	
(Phillips et al., 2013)	
ssDNA (sense)	5'-CGGTGTGAATCAGAC-3'
ssDNA (antisense)	5'-GTCTGATTCACACCG-3'
DNA oligonucleotides for qPCR	
FXN-Luc	
FXN in4F	GCTGTGCTGTGGAATTACT
FXN ex5R	AGGCTTTAGTGAGCTCTGCG
FXN lucR	TTTATGTTTTTGGCGTCTTCC
FXN *(Punga and Buhler, 2010)*	
FXN A (F)	CCCCACATACCCAACTGCTG
FXN A (R)	GCCCGCCGCTTCTAAAATTC
FXN ex1 (F)	GAGCAGCATGTGGACTCTCG
FXN ex1 (R2)	AGAGTGGGGCCAACTCTGC
FXN B (F)	AAACTGACCCGACCTTTATTCCA
FXN B (R)	GGAATCCCCCAAGGTCACA
FXN C (F)	GAAACCCAAAGAATGGCTGTG
FXN C (R)	TTCCCTCCTCGTGAAACACC
FXN D (F)	CTGGAAAAATAGGCAAGTGTGG
FXN D (R)	CAGGGGTGGAAGCCCAATAC
FXN E (F)	ACTTCAGGCCCAGAAACTATCTAATATT
FXN E (R)	TGCCTAGCTTGAGGCTTCAGA
FXN F (F)	TGGCTTTCAGAGTTCGAACCA
FXN F (R)	CACTCTGCTTTTTGACATTCCAA
FXN G (F)	TGACCAGAGTACTTTGCATGAATGT
FXN G (R)	CCAAACGCAGCTGCTAGAATATT

FXN in1 (F2)	TGCAACTCCCTTCTCTGGTTC
FXN in1 (R2)	ATGTCCCCTTTTCCTTCGGA
FXN spliced ex3 (F)	CAG AGG AAA CGC TGG ACT CT
FXN spliced ex 4 (R)	AGC CAG ATT TGC TTG TTT GG
FXN antisense	
FXN up (F)	CTGACCCGACCTTTATTCCA
FXN D (F2)	TTAAGCCAAGATCGCCCAATG
FXN E (F2)	CAAAGGGGCAGCCAATAATC
FXN F (F2)	TGGTGAAGGGGAGGTGAAAG
FXN G (F2)	GGCTGCTTGGGAACTCTTAC
FMR1	
FMR1 up (F)	ACAGTGGAATGTAAAGGGTTG
FMR1 up (R)	GTGTTAAGCACTTGAGGTTCAT
FMR1 ex1 (F)	GAACAGCGTTGATCACGTGAC
FMR1 ex1(R)	GTGAAACCGAAACGGAGCTGA
FMR1 in1A (F)	TAAATTCAGGAATGCACATGC
FMR1 in1A (R)	CCTGAAGTTTCATGGCATATATT
FMR1 in 1B (F)	CTTGAAGGTGAATGAAGAATAGG
FMR1 in 1B (R)	AGCAATTTGTCTGACACACAC
FMR1 in15 (F)	GAACTTCCAGTAAGCATTTCAG
FMR1 ex16 (R)	CTGTTGTTCTTCCTTTAGCCTCTC
FMR1 ex14 spliced (F)	GGAGCTAGTTCTAGACCACCAC
FMR1 ex15 spliced (R)	GAGTTCGTCTCTGTGGTCAGATTC
GAPDH	
(Gromak et al., 2013)	
GAPDH (F)	ACATCAAGAAGGTGGTGAAG
GAPDH (R3)	GGGTCTTACTCCTTGGAGGC
β-Actin	
(Skourti-Stathaki et al., 2011)	
β-Actin ex5 (F)	GGACATCCGCAAAGACCTGTA
β-Actin ex6 (R)	CTCCAACCGACTGCTGTCACC
β-Actin 5'prom (F)	CCACCTGGGTACACACAGTCT
β-Actin 5'prom (R)	TGTCCTTGTCACCCTTTCTTG
β-Actin prom (F)	GAGGGGAGAGGGGGTAAA
β-Actin prom (R)	AGCCATAAAAGGCAACTTTCG
β-Actin in1 (F)	CGGGGTCTTTGTCTGAGC
β-Actin in1 (R)	CAGTTAGCGCCCAAAGGAC
β-Actin in3 (F)	TAACACTGGCTCGTGTGACAA
β-Actin in3 (R)	AAGTGCAAAGAACACGGCTAA
β-Actin in5 (F)	GGAGCTGTCACATCCAGGGTC
β-Actin in5 (R)	TGCTGATCCACATCTGCTGG
β-Actin 5'pause (F)	TTACCCAGAGTGCAGGTGTG
β-Actin 5'pause (R)	CCCCAATAAGCAGGAACAGA

β-Actin pause (F)	GGGACTATTTGGGGGTGTCT
β-Actin pause (R)	TCCCATAGGTGAAGGCAAAG
β-Actin C (F)	TGGGCCACTTAATCATTCAAC
β-Actin C (R)	CCTCACTTCCAGACTGACAGC
β-Actin D (F)	CAGTGGTGTGGTGTGATCTTG
β-Actin D (R)	GGCAAAACCCTGTATCTGTGA
β-Actin F (F)	CCATCACGTCCAGCCTATTT
β-Actin F (R)	TGTGTGAGTCCAGGAGTTGG
RNase H1	
RNaseH1ex7(F)	TAACTGGGTTCAAGGTTGGAAG
RNaseH1ex8(R)	TCTTCCGATTGTTTAGCTCCTTC
γ-Actin	
(Gromak et al., 2013)	
γ-Actin intr1 (F)	CCGCAGTGCAGACTTCCGAG
γ-Actin intr1 (R)	CGGGCGCGTCTGTAACACGG
γ-Actin spliced (F)	AATCTTGCGGCATCCACGAG
γ-Actin spliced (R)	TCGTACTCCTGCTTGCTAATCC
5S	
(Skourti-Stathaki et al., 2011)	
5S (F)	AGCGTCTACGGCCATACC
5S (R)	GGTATTCCCAGGCGGTCTC
pHIV-βGlobin plasmid	
(Skourti-Stathaki et al., 2011)	
5'ex1(F)	GAGTTAGCTCACTCATTAGGC
5'ex1(R)	CTGTGTGAAATTGTTATCCGC
ex2(F)	TTGGACCCAGAGGTTCTTTG
ex2(R)	GAGCCAGGCCATCACTAAAG
3'p(A)(F)	AAAAGGGAATGTGGGAGGTC
3'p(A)(R)	AGCCTCACCTTCTTTCATGG
5'pause(F)	GGGACTATTTGGGGGTGTCT
5'pause(R)	TCCCATAGGTGAAGGCAAAG
A(F)	TTGCCTTCCTGTTTTTGCTC
A(R)	CCGCTGTTGAGATCCAGTTC

Table 2.2. List of RNA oligonucleotides

Bases complementary to target sequences are shown in uppercase letters. Lowercase letters denote the 3' overhangs added for increased stability and efficiency (da, deoxyadenosine; dt deoxythymidine). For oligonucleotides not designed in this work, the original publications are indicated.

Name	Sequence
RNA oligonucleotides for siRNA	
Control (non-targeting siGENOME) siRNA duplex	sense: 5'-UAGCGACUAAACACAUCAA-3' antisense: 5'-UUGAUGUGUUUAGUCGCUA-3' *(#1D-001210-01-20, Thermo Fisher Scientific)*
RNase H1 siRNA duplex	sense: 5'-CAGACAGUAUGUUUACGAUdtdt antisense: 5'- AUCGUAAACAUACUGUCUGdtda *(s48357, Ambion)*
Topoisomerase 1 siRNA duplex	sense: 5'-GGACUCCAUCAGAUACUAUdtdt-3' antisense: 5'-AUAGUAUCUGAUGGAGUCCdtdt -3' *(custom synthesis, Thermo Fisher Scientific)*
DHX9 siRNA duplex	sense: 5'-GAAGUGCAAGCGACUCUAGdtdt-3' antisense: 5'-CUAGAGUCGCUUGCACUUCdtdt -3' *(custom synthesis, Thermo Fisher Scientific)*
SETX siRNA duplex	sense: 5'-AUUUGACGACGGCUUCCACCCAUUG-3' antisense: 5'-CAAUGGGUGGAAGCCGUCGUCAAAU-3' *(custom synthesis, Invitrogen)*
RNA oligonucleotides for RNA/DNA hybrid IP	
(Phillips et al., 2013)	
ssRNA (sense)	5'-CGGUGUGAAUCAGAC-3'
ssRNA (antisense)	5'-GUCUGAUUCACACCG-3'

Table 2.3. List of Buffers

Buffer name	Composition
ChIP Cell Lysis Buffer	85 mM KCl, 5 mM PIPES pH 8.0, 0.5% NP-40, 0.5 mM PMSF, 0.8 µg/µl pepstatin A, 1 µg/µl leupeptin
ChIP Elution Buffer	100 mM NaHCO3, 1% SDS
ChIP IP Buffer	16.7 mM TRIS·HCl pH 8.0, 1.2 mM EDTA, 167 mM NaCl, 0.01% SDS, 1.1% Triton X-100, 0.5 mM PMSF, 0.8 µg/µl pepstatin A, 1 µg/µl leupeptin
ChIP Nuclear Lysis Buffer	50 mM TRIS·HCl pH 8.0, 5 mM EDTA, 1% SDS, 0.5 mM PMSF, 0.8 µg/µl pepstatin A, 1 µg/µl leupeptin
ChIP Wash Buffer A	20 mM TRIS·HCl, pH 8.0, 2 mM EDTA, 150 mM NaCl, 0.1% SDS, 1% Triton X-100
ChIP Wash Buffer B	20 mM TRIS·HCl, pH 8.0, 2 mM EDTA, 500 mM NaCl, 0.1% SDS, 1% Triton X-100
ChIP Wash Buffer C	10 mM TRIS·HCl, pH 8.0, 1 mM EDTA, 250 mM LiCl, 1% NaDOC, 1% NP-40
ChIP Wash Buffer D	10 mM TRIS·HCl, pH 8.0, 1 mM EDTA
LDS Elution Buffer	4% LDS, 500 mM Tris, pH 8.5, 1 mM EDTA, 100 mM DTT
NRO HLB	10 mM TRIS pH 7.5, 10 mM NaCl, 2.5 mM MgCl2, 0.5% NP-40
NRO HLB+Sucrose	10 mM TRIS pH 7.5, 10 mM NaCl, 2.5 mM MgCl2, 0.5% NP-40, 10% w/v sucrose
NRO RSB100	10 mM TRIS pH 7.4, 100 mM NaCl, 2.5 mM MgCl2, 0.4% Triton-X100
NRO Transcription Buffer	40 mM TRIS pH 7.9, 300 mM KCl, 10 mM MgCl2, 40% glycerol, 2 mM DTT
Ponceau S staining solution	5% v/v acetic acid, 0.1% w/v Ponceau S
RIPA Buffer	25 mM TrisHCl pH 7.6, 150 mM NaCl, 1% NP40, 1% Deoxycholate, 0.1% SDS
R-loop IP RSB	10 mM TRIS pH 7.5, 200 mM NaCl, 2.5mM MgCl2
R-loop IP RSB+DSST	10 mM TRIS pH 7.5, 200 mM NaCl, 2.5mM MgCl2, 0.5% Triton X-100, 0.2% NaDOC, 0.1% SDS, 0.05% NaSarkosyl
R-loop IP RSB+T	10 mM TRIS pH 7.5, 200 mM NaCl, 2.5mM MgCl2, 0.5% Triton X-100
SDS-PAGE Resolving gel	375 mM TRIS pH 8.8, 0.1% SDS, 0.1% APS, 0.1% TEMED, 10-15 % acrylamide mix
SDS-PAGE Running Buffer	24.76 mM TRIS pH 8.3, 192 mM glycine, 1 % w/v SDS
SDS-PAGE Stacking gel	125 mM TRIS pH 6.8, 0.1% SDS, 0.1% APS, 0.1% TEMED, 5% acrylamide mix
Silver Developing Solution	3% w/v Na_2CO_3, 1.85% formaldehyde
Silver Fixative Solution	50% EtOH, 10% acetic acid, 1.85% formaldehyde
Silver Nitrate Solution	0.1% w/v $AgNO_3$
Silver Stopping Solution	50% EtOH, 10% acetic acid
TE	10 mM TRIS·HCl, pH 8.0, 1 mM EDTA

Table 2.4. List of commercial reagents, materials and enzymes

Reagent	Catalogue number	Supplier
5-azadC	A3656	Sigma-Aldrich Ltd.
Acrylamide:bisacrylamide mix (37.5:1)	EC-890	National Diagnostics
Actinomycin D	A9415	Sigma-Aldrich Ltd.
Amersham ECL detection reagent	RPN2209	GE Healthcare
Benzonase nuclease	E8263	Sigma-Aldrich Ltd.
Biotaq DNA polymerase	BIO-21040	Bioline, UK
BIX-01294	B9311	Sigma-Aldrich Ltd.
Bovine serum albumin (BSA, >96%)	A9647	Sigma-Aldrich Ltd.
c0mplete EDTA-free protease inhibitors	5056489001	Roche AG
Camptothecin	C9911	Sigma-Aldrich Ltd.
Dithiothreitol	D9779	Sigma-Aldrich Ltd.
DNA ladder (GeneRuler)	SM1352	Thermo Fisher Scientific Inc.
DNA loading dye, 10x (GeneRuler)	SM1352	Thermo Fisher Scientific Inc.
DNase I, RNase-free	4716728001	Roche AG
Dynabeads Protein A (Invitrogen)	10002D	Life Technologies Ltd.
Formaldehyde (36.5-38%)	F8775	Sigma-Aldrich Ltd.
Glycogen	11017101001	Roche AG
LDS Loading Buffer	NP0007	Life Technologies Ltd.
Leupeptin	11017101001	Roche AG
Lipofectamine2000 (Invitrogen)	11668027	Life Technologies Ltd.
NuPAGE 3-8% Tris-Acetate gel	EA0375BOX	Life Technologies Ltd.
NuPAGE 4-12% Bis-Tris gel	NP0321BOX	Life Technologies Ltd.
NuPAGE antioxidant	NP0005	Life Technologies Ltd.
NuPAGE MOPS Running Buffer	NP0001	Life Technologies Ltd.
NuPAGE Transfer Buffer	NP00061	Life Technologies Ltd.
NuPAGE TrisAcetate Running Buffer	LA0041	Life Technologies Ltd.
Pepstatin	11359053001	Roche AG
Plasmid Maxi Prep Kit	12162	QIAGEN
PMSF	P7626	Sigma-Aldrich Ltd.
Protein A Agarose Beads	16-157	Merck-Millipore, Merck KGaA
Protein molecular weight maker (PageRuler Plus)	26619	Thermo Fisher Scientific Inc.
Proteinase K	3115828001	Roche AG
PVDF membrane (Amersham Hybond-P)	10600023	GE Healthcare
QIAprep Spin Miniprep Kit	27104	QIAGEN
QIAquick PCR Purification Kit	28104	QIAGEN
QuantiTect SYBR Green PCR Master Mix	204145	QIAGEN
RNase A (PureLink, 20mg/ml)	12091021	Life Technologies Ltd.
RNase H	M0297S	New England Biolabs Ltd.
SuperScript III Reverse Transcriptase (Invitrogen)	18080-044	Life Technologies Ltd.
TransFectin Lipid Reagent	1703351	Bio-Rad Laboratories, Inc.
TRIzol reagent	15596026	Life Technologies Ltd.
X-ray film (FujiFilm, RX NIF)	12715325	Fisher Scientific Ltd.

Table 2.5. List of antibodies

The table indicates the targets of all antibodies employed in this work. Dilutions for immunoblotting are given in column 'Western blot dilution'. The amount of antibody used per sample for chromatin immunoprecipitation (ChIP) experiments are also indicated.

Target	Catalogue number	Western blot dilution	ChIP	Supplier
Actin	A2066	1:2000	-	Sigma-Aldrich Ltd.
BrdU	B2531, clone BU-33	-	-	Sigma-Aldrich Ltd.
CBP80	sc-48803	-	-	Santa Cruz Biotechnology
Control IgG2a	M5409	-	-	Sigma-Aldrich Ltd.
DDX1	11357-1-AP	1:1000	-	Proteintech Group Inc.
DDX5	A300-523A	1:1000	-	Bethyl Laboratories
DGCR8	10996-1-AP	1:400	-	Proteintech Group Inc.
DHX9	ab26271	1:8000	-	Abcam Plc.
Drosha	D28B1	1:1000	-	Cell Signaling Technology
Exosc10	ab50558	1:2000	-	Abcam Plc.
Histone H3, total	ab1791	1:10000	5 µg	Abcam Plc.
Histone H3K9me2	ab1220	1:2000	5 µg	Abcam Plc.
Lamin B1	ab16048	1:2000	-	Abcam Plc.
Mouse IgG	anti-Mouse-ECL, A8924	1:8000	-	Sigma-Aldrich Ltd.
Nuclear Pore Complex	ab24609	1:1000	-	Abcam Plc.
Rabbit IgG	anti-Rabbit-ECL, A0545	1:10000	-	Sigma-Aldrich Ltd.
RNA polymerase II	ab817 (8WG16)	1:1000	-	Abcam Plc.
RNA polymerase II	H224	-	3 µg	Santa Cruz Biotechnology
RNA/DNA hybrids	S9.6	1:1000	14.4 µl	Dr Natalia Gromak (University of Oxford, UK)
RNase H1	ab83179	1:800	-	Abcam Plc.
SETX	A301-105A	1:350	-	Bethyl Laboratories
SRSF1	32-4500, clone 96	1:2000	-	Life Technologies Ltd.
Topoisomerase 1	ab109374	1:1000	-	Abcam Plc.
Tubulin	T5168	1:2000	-	Sigma-Aldrich Ltd.
XRN2	11267-1-AP	1:1000	-	Proteintech Group Inc.

2.2 Bacterial techniques

2.2.1 Transformation of bacterial strains

Transformation of bacteria was carried out with standard sterile molecular biology techniques (Sambrook and Russell, 2006). Briefly, 50 µl of frozen chemicompetent *E. coli* stocks were thawed on ice (~20 min), and 10–50 ng plasmid DNA was gently added and incubated on ice for 30 min. Heat-shock was performed for 2.5 min at 37°C, after which cells were immediately incubated on ice for 2 min. Then, 100 µl prewarmed 2xTY (1.6% w/v bacto-tryptone, 1% w/v yeast extract, 0.5% w/v NaCl) was added to the cells and incubated at 37°C for 1 h with gentle shaking. Transformed cells were then plated onto LB-agar plates containing the appropriate antibiotic for selection and incubated at 37°C overnight.

2.2.2 Growth of bacterial strains

For small scale growth of bacteria, colonies were picked from a LB-agar plate (1.5% w/v agar dissolved in 1xLB) grown overnight at 37°C, inoculated into 5 ml of LB (1% w/v bacto-tryptone, 0.5% w/v yeast extract, 1% w/v NaCl) containing the appropriate selective antibiotic and grown overnight at 37°C in a horizontal shaker. For medium scale growth of bacteria, a starter culture was grown by inoculating 5 ml LB including antibiotic with a colony picked from a LB-agar plate by incubating at 37°C for 6–8 h in a horizontal shaker. This starter culture was transferred into 100 ml LB including antibiotic and incubated overnight at 37°C in a horizontal shaker.

For *E. coli* culture media and LB-agar plates, ampicillin (Sigma-Aldrich) was added to 100 µg/ml. Kanamycin (Sigma-Aldrich) was added to 50 µg/ml final concentration.

2.2.3 Small scale preparation of plasmid DNA

For small scale preparation of plasmid DNA, 4 ml of overnight small scale cultures were pelleted and processed according to the manufacturer's instructions (QIAprep Spin Miniprep Kit, QIAGEN). DNA was quantified and stored at -20°C until further use.

2.2.4 Medium scale preparation of plasmid DNA

For large scale preparation of plasmid DNA, the medium scale overnight culture was pelleted at 6000 g for 15 min at 4°C and processed according to the manufacturer's instructions (Plasmid Maxi Kit, QIAGEN). Purified plasmid DNA was quantified using spectrophotometry, sized using agarose gel electrophoresis, and stored at a concentration of 1 µg/µl at -20°C until further use.

2.3 Mammalian cell culture techniques

2.3.1 Propagation of mammalian cell lines

Lymphoblastoid cell lines were grown in Roswell Park Memorial Institute 1640 medium (RPMI 1640, Sigma-Aldrich) supplemented with 15% fetal bovine serum (FBS, from Gibco), 100 U/ml penicillin and 100 µg/ml streptomycin (Gibco, Thermo Fisher Scientific) at 37ºC in 5% CO_2. HeLa cells were grown in Dulbecco's modified Eagle's medium (DMEM, Sigma-Aldrich) supplemented with 10% FBS, 100 U/ml penicillin and 100 µg/ml streptomycin at 37ºC in 5% CO_2. HEK293 FXN-Luc and FXN-GAA-Luc cell lines were grown in DMEM medium supplemented with 10% fetal bovine serum (FBS), 100 µg/ml hygromycin B (Life Technologies), 100 U/ml penicillin (Thermo Fisher Scientific) and 100 µg/ml streptomycin (Thermo Fisher Scientific) at 37ºC in 5% CO_2. All cell lines were split regularly every three days by trypsinisation (HeLa and HEK293 cells) or dilution (lymphoblasts) at approximately 80% confluency. All experiments were performed with early passages of the cell lines.

2.3.2 Pharmacological treatments of mammalian cell lines

1 µM 5-azadC (Sigma-Aldrich) was added to the media for 7 days. The media was replenished every 24 hours by replacement of half of the conditioned media with fresh media and drug. For 5-azadC wash-out experiments, cells were treated with 1 µM 5-azadC for 7 days. On day 7, cell were washed twice with fresh RPMI 1640 medium and cultured in the absence of 5-azadC during the indicated time. 4 µM BIX-01294 (Sigma-Aldrich) was added to the media for a total of 72 hours. The media was replenished every 24 hours by replacement of half of the conditioned media with fresh media and drug. 10 µM camptothecin (CPT, Sigma-Aldrich) was added for 6 hours.

Actinomycin D (Sigma-Aldrich) was added to a final concentration of 5 μg/ml to the media for 6-48 hours.

2.3.3 Storage of mammalian cell lines

To prepare frozen stocks of HeLa and HEK293 FXN-Luc or FXN-GAA-Luc cells, $6 \cdot 10^6$ cells were resuspended in 1 ml freezing medium (80% v/v FBS, 20% DMSO). FRDA patient and control lymphoblasts were frozen by resuspending $1 \cdot 10^7$ cells in 1 ml freezing medium (20% v/v FBS, 6% v/v DMSO in RPMI 1640). All frozen cell stocks were stored at -80°C.

Frozen cells were thawed by gradually warming up to 37°C. The defrosted cells were diluted in complete RPMI 1640 (lymphoblasts) or in complete DMEM (HeLa or FXN-GAA-Luc and FXN-GAA-Luc cells), pelleted and resuspended again in complete medium with antibiotics, followed by standard propagation at 37°C.

2.3.4 siRNA-mediated knockdown

The siRNA-mediated knockdowns were carried out as described (Skourti-Stathaki *et al.*, 2011; Wollerton *et al.*, 2004). Throughout the whole procedure, cells were cultured in the absence of antibiotics. In brief, $1.5 \cdot 10^5$ HEK293 FXN-Luc, FXN-GAA-Luc or HeLa cells were split on day one into 24 well plates. Next day, cells were transfected with 60 pmol siRNA duplex per well using Lipofectamine2000 (Invitrogen) according to the manufacturer's instructions. The following day, cells were split into 6-well plates. On day four, the second siRNA transfection was carried out using 60 pmol siRNA duplex per well as described above. Cells were split the following day to cell densities used for ChIP or DIP and harvested on day six. The duplex sequences used for siRNA are listed in Table 2.1.

2.3.5 Plasmid transfections

FXN-Luc and FXN-GAA-Luc cells were freshly split to $2.5 \cdot 10^6$ cells into 10 cm dishes and transfected on the following day with 10 μg pFlag or pRNaseH1-Flag plasmids (Table 2.1) using TransFectin reagent (BioRad), following the manufacturer's instructions. Cells were harvested 48 hours after transfection. For experiments with the pHIV-βGlobin plasmid in siRNA-treated HeLa

cells, 1 μg of pHIV-βGlobin plasmid and 0.1 μg of pTAT HIV transactivator plasmid was added per well during the second siRNA transfection (Chapter 2.3.4).

2.4 DNA analysis methods

2.4.1 Quantitation of nucleic acids

RNA and DNA nucleic acids were quantified using UV spectrophotometry. 1 μl of the nucleic acid sample was quantified using a Nanodrop 1000 (Thermo Fisher Scientific) and the absorbance at 260 nm was measured to calculate the concentration of RNA or DNA. The ratio of 260/280 nm was used to assess the purity of nucleic acid samples.

2.4.2 Agarose gel electrophoresis

Agarose gel electrophoresis was employed to measure sizes of plasmid constructs, PCR products and sonicated DNA fragments from DIP, ChIP and RNA/DNA hybrid IP experiments. DNA was mixed with 1x bromophenol blue-containing DNA loading dye (GeneRuler, Thermo Fisher Scientific) and separated at 100V using 0.8–2% agarose gels containing 1xTE buffer and 0.005% ethidium bromide. DNA sizes were estimated by comparison with a 100 bp DNA ladder (GeneRuler, Thermo Fisher Scientific) loaded alongside the samples. DNA bands were visualised under UV light.

2.4.3 DNA sequencing

DNA sequencing was carried out to confirm sequences of all plasmid constructs. Sanger sequencing reactions contained 100 ng/μl of purified plasmid Maxi prep DNA and were supplied with 3.2 pmol/μl of the appropriate sequencing primers. All sequencing reactions were performed by Source BioScience UK limited, Department of Biochemistry, University of Oxford, UK.

2.4.4 Purification of mammalian genomic DNA

In order to obtain genomic mammalian DNA to prepare standards for qPCR analysis, confluent lymphoblasts (GM15851) from one T25 cell culture flask were harvested and DNA was isolated following the first steps of the DIP protocol (Chapter 2.6.1). After removal of proteins, genomic DNA was precipitated with 100% isopropanol and washed once in 70% EtOH. Washed

genomic DNA was resuspended in 400 µl ddH_2O, sonicated for 10 s at medium power setting (Bioruptor, Diagenode) and then purified using the QIAquick PCR Purification Kit (QIAGEN).

2.4.5 Polymerase chain reaction (PCR)

GAA expansion sizes in HEK293 FXN-Luc and FXN-GAA-Luc cells were confirmed using PCR (Figure 4.1B). 20 ng of genomic DNA was used as template in a 25 µl PCR reaction, containing 2.5 U Biotaq DNA polymerase (Bioline), 3 mM MgCl2, 0.4 mM dNTPs, 0.4 µM GAA104(F) primer , 0.4 µM GAA629(R) primer in 1xNH4 reaction buffer. Primer sequences are listed in Table 2.1 and have been described before (Filla *et al.*, 1996). Products were amplified using the protocol from (Holloway *et al.*, 2011) with minor modifications. In particular, 5 min at 94°C were followed by 10 cycles of 94°C for 20s, 65°C for 30 s, 72°C for 5 min. This was followed by 20 cycles of 94°C for 20 s, 65°C for 30 s and 72°C for 5 min, with the 72°C step becoming 20 s longer in each cycle. After a final step at 72°C for 10 min, PCR products were resolved on a 1% agarose gel.

2.4.6 Real-time quantitative PCR

Real-time quantitative PCR (qPCR) was employed to amplify and simultaneously quantify specific DNA sequences. In contrast to end-point PCR, real-time PCR measures the quantity of amplified DNA after each PCR cycle. Here, the fluorescent dye SYBR Green, which intercalates into DNA, was employed to monitor amplification of DNA products. For each 10 µl reaction, 2 µl of sample (purified DNA or cDNA) were mixed with 5 µl of 2xQuantiTect PCR SYBR Green master mix (QIAGEN), 0.5 µl forward primer (10 µM), 0.5 µl reverse primer (10 µM), and 2 µl ddH_2O. All qPCR reactions were performed in technical triplicates to reduce technical error. In addition, a non-template control (NTC) and a dilution series of purified genomic DNA was amplified for each primer pair. Genomic DNA standards (1/1–1/400 dilution series) for qPCR had the following concentrations: 16 ng/µl (1/1), 3.2 ng/µl (1/5), 0.64 ng/µl (1/25), 0.16 ng/µl (1/100), 0.08 ng/µl (1/200), 0.04 ng/µl (1/400).

Amplification and analysis was carried out in RotorGene 3000 (Corbett Research) using the manufacturer's software algorithms (Corbett Research Rotor Gene 6.0). Quantification of DNA in samples was based on comparison of the amplification curves of the samples to the amplification of

the genomic DNA dilution series, thereby taking into account variability of amplification efficiencies between different qPCR primer pairs and variation between individual qPCR runs. To confirm the specificity and quality of the amplified qPCR products, a melting curve analysis was performed during each qPCR run. For new qPCR primer pairs, qPCR reactions were separated using agarose gel electrophoresis to confirm the correct size of expected qPCR products.

2.5 RNA analysis methods

2.5.1 Purification of total RNA

RNA was extracted from cells using the TRIzol reagent (Life Technologies) following the manufacturer's instructions. Briefly, tissue culture medium was removed by centrifugation (lymphoblastoid cells) or aspiration (adherent cells) and cells were lysed by addition of 1 ml TRIzol (for up to $5 \cdot 10^6$ cells) and vigorous mixing. 200 µl chlorophorm were added and samples incubated for 5 min at room temperature, followed by centrifugation for 15 min at 16000 g at 4°C. The top phase was transferred to a new tube and RNA precipitated by the addition of an equal volume of isopropanol, incubation for 10 min at room temperature and centrifugation for 25 min at 16000 g at 4°C. In case of expected low RNA yields, 20 µg of glycogen (Roche) was added as during isopropanol precipitation. Pellets were washed with 1 ml 75% EtOH and centrifuged for 10 min at 16000 g at 4°C and air-dried at room temperature. RNA pellets were dissolved in RNase-free ddH_2O (25 µl for $5 \cdot 10^6$ cells). Removal of contaminating genomic DNA was achieved by incubation for 2 h at 37°C in 1x DNase I buffer with 30 U RNase-free DNase I, followed by heat inactivation for 15 min at 75°C.

2.5.2 Br-UTP nuclear run-on (Br-UTP NRO)

The Br-UTP nuclear run-on (Br-UTP NRO) analysis was carried as described before (Skourti-Stathaki *et al.*, 2011). The equivalent of $8 \cdot 10^6$ nuclei from lymphoblastoid cells were used for each Br-UTP NRO reaction. In brief, for each BrUTP-NRO, two batches of cells were harvested, washed in cold PBS (Lonza) and resuspended and lysed in NRO HLB (10 mM TRIS pH 7.5, 10 mM NaCl, 2.5 mM MgCl2, 0.5% NP-40). Nuclei were underlayered with NRO HLB+Sucrose (HLB

buffer + 10% w/v sucrose) and centrifuged for 5 min at 500 g at 4°C. Nuclear pellets were gently resuspended in 50 µl NRO Transcription Buffer (40 mM TRIS pH 7.9, 300 mM KCl, 10 mM MgCl2, 40% glycerol, 2 mM DTT) and ribonucleotides (rATP, rGTP, rCTP) were added to both pellets at 0.6 mM final concentrations. In addition, to one sample (+Br-UTP), 0.2 mM Br-UTP was added, while to the control sample (-Br-UTP) 0.2 mM unlabelled rUTP was added. The NRO was performed for 30 min at 30°C. Nuclear RNA was harvested and purified using TRIzol reagent (Life Technologies) and DNA was degraded by DNase I treatment for 1 h at 37°C (Roche). RNA was incubated with 2 µg α-BrdU-antibody (Sigma-Aldrich) prebound to 60 µl protein A agarose beads for 1 h at 4°C. Immunoprecipitated labelled RNA was washed 3x with NRO RSB100, eluted with TRIzol reagent (Life Technologies) and precipitated as described in section 2.4.1. The amount of Br-UTP-labelled RNA was quantified by RT-qPCR analysis, subtracting the signal of the control (-Br-UTP) sample from the +Br-UTP sample.

2.5.3 Reverse transcription qPCR (RT-qPCR)

1–2 µg of RNA were used for reverse transcription using SuperScript Reverse Transcriptase III (Life Technologies) following the manufacturer's instructions using either 100 ng random hexamer primers (Life Technologies), 50 pmol oligo-dT primer (Life Technologies) or 10 pmol gene-specific primers (Life Technologies). 2 µl of undiluted or 1:2-diluted (random hexamer-primed) cDNA was then analysed by quantitative real-time PCR using primers listed in Table 2.1. For quantitative analysis of ribosomal RNA, cDNA was diluted 1:100 prior to qPCR.

Strand-specific analysis of FXN sense RNA

For the analysis of sense *FXN* RNA (Figure 3.2), reverse transcription was carried out using a specific reverse (R) primer, followed by qPCR with the corresponding forward (F) and reverse (R) primers. All sequences are listed in Table 2.1.

Strand-specific analysis of FXN antisense RNA

To analyse *FXN* antisense RNA (Figure 4.9A), a nested RT-qPCR approach was employed. Additional gene-specific forward primers were designed upstream of the qPCR amplicons shown in Figure 3.1C (Table 2.1). These upstream forward (F) primers were used in reverse transcription and followed by qPCR using the corresponding downstream qPCR primer pair. The combination of upstream antisense primers and qPCR amplicons were as follows: up(F)→A(F)/A(R), A(F)→ex1(F)/ex1(R2), ex1(F)→B(F)/B(R), B(F)→C(F)/C(R), D(F2)→ D(F)/D(R), E(F2)→E(F)/E(R), F(F2)→F(F)/F(R), G(F2)→ G(F)/G(R).

2.6 RNA/DNA hybrid analysis methods

2.6.1 DNA/RNA immunoprecipitation (DIP)

DNA/RNA immunoprecipitation (DIP) analysis was carried out essentially as described previously (Skourti-Stathaki *et al.*, 2011). Cells were split one day before DIP to the following numbers: $5 \cdot 10^6$ lymphoblastoid cells; $2 \cdot 10^6$ FXN-Luc and FXN-GAA-Luc; $1.5 \cdot 10^6$ HeLa cells. The entire DIP procedure was carried out at 4°C.

Cells were harvested and washed twice in ice-cold PBS (Lonza), resuspended in 800 µl ChIP Cell Lysis Buffer (85 mM KCl, 5 mM PIPES pH 8.0, 0.5% NP-40) and incubated for 10 min. Nuclei were collected by centrifugation (5 min at 500 g, 4°C) and resuspended in 800 µl ChIP Nuclear Lysis Buffer (50 mM TRIS·HCl pH 8.0, 5 mM EDTA, 1% SDS). After 10 min, proteins were digested by addition of 5 µl proteinase K (Roche) and incubated for a total of 3 h at 55°C. Another 2 µl proteinase K (Roche) were added after the first hour and second hour of incubation, respectively.

After complete digestion, 200 µl of 5 M KOAc were added and proteins and cell debris were removed by centrifugation (5 min 16000 g, 4°C). DNA was precipitated from the supernatant by addition of 800 µl isopropanol and centrifugation (5 min 16000g, 4°C). DNA pellets were washed once in 75% EtOH, air-dried, resuspended in 100 µl ddH_2O and diluted with 300 µl ChIP IP Buffer (16.7 mM TRIS·HCl pH 8.0, 1.2 mM EDTA, 167 mM NaCl, 0.01% SDS, 1.1% Triton X-100) and sonicated with the following settings for the Diagenode Bioruptor: medium power with 30 s on / 30 s

off intervals for 3 min (lymphoblastoid, FXN-Luc and FXN-GAA-Luc cells) or 12 min (HeLa). After sonication, samples were centrifuged (10 min 16000 g, 4°C) and diluted with 2.6 ml ChIP IP Buffer including protease inhibitors (0.5 mM PMSF, 0.8 µg/µl pepstatin A, 1 µg/µl leupeptin). Preclearing was carried out by addition of 50 µl protein A agarose beads (Merck-Millipore) and rotation at 16 rpm for 1 h at 4°C. Beads were removed by centrifugation (5 min 500 g, 4°C) and 100 µl were taken from the supernatant as 'Input' sample. The remaining material was divided: to 1.34 ml material, 14.4 µl S9.6 antibody was added ('IP' sample) or not ('beads-only' sample). Immunoprecipitation was carried out overnight rotating at 16 rpm at 4°C. Next day, washes, elution and DNA purification steps were identical to those described for ChIP (Chapter 2.7.6), except that the de-crosslinking step was omitted.

DIP analysis of RNase H sensitivity in vitro

DIP was carried out as described and the RNase H treatment performed after isolation of genomic DNA and prior to immunoprecipitation. In particular, genomic DNA from $1 \cdot 10^7$ lymphoblastoid cells was resuspended in 200 µl 1xRNase H digestion buffer (New England Biolabs) and divided equally into a '+RNase H' and 'control' sample. 25 U recombinant RNase H (New England Biolabs) were added to the '+RNase H' sample and both samples were incubated for 6 h in a thermoshaker (Eppendorf) at 450 rpm at 37°C. Following this, the samples were diluted with 300 µl ChIP IP buffer, and processed according to the DIP protocol.

2.6.2 RNA/DNA hybrid slot blot

RNA/DNA hybrids were purified as described in the DIP protocol with minor modifications. Samples were deproteinised for 2 h at 45°C, followed by precipitation and purification of genomic DNA and RNA/DNA hybrids. Purified RNA/DNA hybrids were diluted in 0.5x TE buffer (5 mM TRIS pH 8.0, 0.5 mM EDTA) and spotted on Hybond-N+ membrane (Amersham Biosciences) using a Bio-Dot SF apparatus (Bio-Rad Laboratories) following the manufacturer's instructions. The membrane was probed with S9.6 antibody (diluted 1:1000 in PBS containing 2% dried skimmed milk) and detected as described for western blotting (Chapter 2.7.4).

2.7 Protein analysis methods

2.7.1 Preparation of whole cell extracts

For whole cell extract preparation, cells were washed 2x in ice-cold PBS (Lonza) and cell pellets were resuspended in RIPA buffer (25 mM TRIS·HCl pH 7.6, 150 mM NaCl, 1% NP40, 1% Deoxycholate, 0.1% SDS), supplemented with 50 mM PMSF (Sigma-Aldrich) and 1x cOmplete EDTA-free protease inhibitor cocktail (Roche). After incubation for 20 min on ice, samples were sonicated 3x for 30 s on, 30 s off at medium power setting (Bioruptor, Diagenode) to reduce viscosity and cleared by centrifugation at 16.000 g for 10 min at 4°C. Protein concentration of the supernatant was then determined using the Bradford assay and stored at -20°C until further use.

2.7.2 Quantitation of protein concentration

Protein concentration of mammalian whole cell extracts was determined according to the Bradford method (Bradford, 1976). In brief, for each sample 200 µl 1xBradford dye reagent (Bio-Rad Laboratories) were mixed with 795 µl PBS (Lonza) and 5 µl of protein sample. After 5 min incubation at room temperature, absorbance at 595 nm was measured. The appropriate sample buffer was used as blank control. A standard curve based on at least six BSA standards of known concentrations was used to calculate the protein concentration of the samples based on their absorbance. For each sample, duplicate measurements were performed.

2.7.3 SDS polyacrylamide gel electrophoresis

Proteins were separated according to their molecular weight using SDS-PAGE. Gels were cast in-house, using 5% stacking gels (125 mM TRIS pH 6.8, 0.1% SDS, 0.1% APS, 0.1% TEMED, 5% of 37.5:1 acrylamide:bisacrylamide mix) and 10-15% resolving gels (375 mM TRIS pH 8.8, 0.1% SDS, 0.1% APS, 0.1% TEMED, 10-15% of 37.5:1 acrylamide:bisacrylamide mix). Alternatively, pre-cast 3-8% Tris-Acetate or 4-12% Bis-Tris gradient NuPAGE gels (Life Technologies) were used. Proteins were denatured in the presence of 100 mM DTT in 1xLDS loading buffer (Life Technologies) by heating at 70°C for 10 min. Electrophoresis was carried out in the presence of antioxidant according to the manufacturer's instructions (Life Technologies). In particular, pre-cast 4-12% Bis-Tris gels were run at 200V for 45 min in MOPS SDS running buffer

(Life Technologies). Pre-cast 3-8% Tris-Acetate gels were run at 150V for 60 min in Tris-Acetate SDS Running Buffer (Life Technologies). Gels prepared in-house were run at 150V for 90 min in SDS-PAGE Running Buffer (24.76 mM TRIS pH 8.3, 192 mM glycine, 1% w/v SDS).

Silver staining

For silver staining, gels were washed briefly 3x in ddH_2O and fixed for 1 h at room temperature in freshly prepared Silver Fixative Solution (50% EtOH, 10% acetic acid, 1.85% formaldehyde). After 6x10 min washes in ddH_2O using a horizontal shaker, gels were incubated for 1 h at room temperature in 2.5 μM DTT in ddH_2O. This was followed by 3x brief rinses in ddH_2O. Gels were then incubated in Silver Nitrate Solution (0.1% w/v $AgNO_3$) for 1 h at room temperature, followed by 3 rinses in ddH_2O. Gels were developed by gentle shaking for 3-15 min in Silver Developing Solution (3% w/v Na_2CO_3, 1.85% formaldehyde). Development was stopped by incubation in Silver Stopping Solution (50% EtOH, 10% acetic acid) for 30 min at room temperature under gentle horizontal shaking. Finally, gels were washed in ddH_2O for 30 min and dried for documentation and storage.

2.7.4 Western blotting

Protein samples were separated using SDS-PAGE. After electrophoresis, gels were equilibrated in 2x NuPAGE Transfer Buffer (Life Technologies) containing 10% w/v MeOH for 20 min at room temperature. Amersham Hybond-P PVDF membranes (GE Healthcare) were activated by incubation for 60 s in 100% MeOH, followed by 5 min washes in ddH_2O and equilibration in 2xNuPAGE Transfer Buffer containing 10% MeOH for 15 min at room temperature. Proteins were transferred onto the PVDF membrane by semi-dry transfer at 15V for 2 h using the Trans-Blot SD Semi-Dry transfer cell (Bio-Rad Laboratories).

Ponceau S staining

After semi-dry transfer of proteins onto PVDF membranes, transfer efficiency was assessed by staining of the PVDF membrane with Ponceau S stain. Briefly, the PVDF membrane was incubated for 1–2 min on a horizontal shaker in Ponceau S staining solution (5% v/v acetic acid,

0.1% w/v Ponceau S) at room temperature until bands appeared. Membranes were then washed in ddH_2O to remove background staining and visualise bands.

Antibody probing

After extensive washes in ddH_2O, destained PVDF membranes (Amersham Hybond-P, GE Healthcare) were blocked in PBS (Lonza) containing 2% dried semi-skimmed milk (Tesco) and 0.1% Tween-20 (Sigma-Aldrich) overnight at 4°C and used for western blotting. Membranes were incubated for 90 min at room temperature under gentle horizontal shaking with primary antibodies (Table 2.5), diluted in PBS containing 2% Milk and 0.1% Tween-20. After 3x washes for 10 min in PBS containing 2% Milk and 0.1% Tween-20, membranes were incubated with HRP-coupled secondary antibodies diluted in PBS containing 2% Milk and 0.1% Tween-20 (Sigma-Aldrich), followed by 3x washes for 10 min in PBS containing 0.1% Tween-20. Proteins were detected using chemiluminescence (Amersham ECL detection reagent) by exposure of X-ray films (FujiFilm).

2.7.5 Protein co-immunoprecipitation

HeLa cells at 85% confluency were washed with PBS (Lonza) and lysed in RSB+T (10 mM TRIS pH 7.5, 200 mM NaCl, 2.5mM $MgCl_2$, 0.5% Triton X-100) on ice, followed by sonication for 3 min at medium power setting (Bioruptor, Diagenode). After removal of insoluble material, 1 mg of extracts was incubated with 3 μg of antibody overnight. Immunocomplexes were captured with 50 μl protein A dynabeads (Invitrogen), washed 4x for 5 min in RSB+T at 4°C and 2x for 5 min in RSB (10 mM TRIS pH 7.5, 200 mM NaCl, 2.5mM $MgCl_2$) at 4°C and eluted as described for R-loop IP for analysis of co-immunoprecipitated proteins or RNA/DNA hybrids using RNA/DNA hybrid slot blot.

2.7.6 Chromatin immunoprecipitation (ChIP)

Chromatin immunoprecipitation (ChIP) was carried out as described before with minor modifications (Kadener *et al.*, 2002; West *et al.*, 2004). One day before ChIP, cells were split to the following numbers: $5 \cdot 10^6$ lymphoblastoid cells; $2 \cdot 10^6$ FXN-Luc and FXN-GAA-Luc; $1.5 \cdot 10^6$ HeLa. After 24 h, cells were crosslinked by the addition of formaldehyde to the culture medium at a final

concentration of 0.95% followed by incubation for 15 min at 37°C on a horizontal shaker. Crosslinking was stopped by addition of glycine to a final concentration of 130 mM and incubation for 5 min on a horizontal shaker.

All subsequent steps were carried out at 4°C. Cells were then washed twice with ice-cold PBS (Lonza) including protease inhibitors (0.1 mM PMSF, 1.25 µg/µl pepstatin A, 0.25 µg/µl leupeptin) and lysed by resuspension in 300 µl ChIP Cell Lysis Buffer (85 mM KCl, 5 mM PIPES pH 8.0, 0.5% NP-40, 0.5 mM PMSF, 0.8 µg/µl pepstatin A, 1 µg/µl leupeptin). After 10 min incubation, nuclei were collected by centrifugation (5 min 500 g, 4°C) and resuspended in 400 µl ChIP Nuclear Lysis Buffer (50 mM TRIS·HCl pH 8.0, 5 mM EDTA, 1% SDS, 0.5 mM PMSF, 0.8 µg/µl pepstatin A, 1 µg/µl leupeptin). Following 10 min incubation on ice, samples were sonicated in a Diagenode Bioruptor for 15 min at medium power setting with 30s on/30s off intervals. To remove cell debris and insoluble material, samples were centrifuged (10 min 16000g, 4°C) and supernatants were diluted with 2.6 ml ChIP IP Buffer (16.7 mM TRIS·HCl pH 8.0, 1.2 mM EDTA, 167 mM NaCl, 0.01% SDS, 1.1% Triton X-100, 0.5 mM PMSF, 0.8 µg/µl pepstatin A, 1 µg/µl leupeptin). For preclearing, 50 µl protein A beads (Merck-Millipore) were added and incubated with nuclear extracts rotating at 16 rpm for 1 h. Beads were removed by centrifugation (5 min 500 g, 4°C) and 100 µl were taken from the supernatant as 'Input' samples and stored at -20°C. The remaining sample was divided into two fractions for immunoprecipitation: to 1.7 ml material, 5 µg antibody were added ('IP' sample) or not ('beads-only' sample). For ChIPs of histone modifications, the sample was divided into three fractions of 850 µl each ('histone modification' IP, 'total histone' IP, 'beads-only').

After incubation overnight rotating at 16 rpm, antibody-protein complexes were captured by addition of 60 µl protein A agarose beads and incubation for 1.5 h, rotating at 16 rpm at 4°C. Samples were washed by consecutive incubation of beads in different wash buffers rotating for 5 min at 16 rpm at 4°C and removal of wash buffers by centrifugation (3 min 500 g, 4°C). The washes comprised: 1x ChIP Wash Buffer A (20 mM TRIS·HCl, pH 8.0, 2 mM EDTA, 150 mM NaCl, 0.1% SDS, 1% Triton X-100), 1x ChIP Wash Buffer B (20 mM TRIS·HCl, pH 8.0, 2 mM EDTA, 500 mM NaCl, 0.1% SDS, 1% Triton X-100), 1x ChIP Wash Buffer C (10 mM TRIS·HCl, pH 8.0, 1 mM

EDTA, 250 mM LiCl, 1% NaDOC, 1% NP-40), and 3x ChIP Wash Buffer D (10 mM TRIS·HCl, pH 8.0, 1 mM EDTA).

After complete removal of wash buffers, samples were eluted twice from the beads by addition of 250 µl of freshly-prepared ChIP Elution Buffer (100 mM NaHCO3, 1% SDS) and rotation at 16 rpm for 15 min at room temperature. Samples were collected by centrifugation for 3 min at 16000 g at room temperature. The 500 µl ChIP eluates, alongside with Inputs, were decrosslinked in 300 mM NaCl in the presence of 0.016 µg/µl RNase A (Life Technologies) for 4 h at 65°C. Proteins were then digested in 40 mM TRIS·HCl pH 6.5 and 10 mM EDTA with 0.04 µg/µl proteinase K (Roche) for 2 h at 45°C.

Samples were then purified using the QIAquick PCR Purification Kit (QIAGEN) following the manufacturer's instructions. Briefly, samples were resuspended in 5 volumes of buffer PB and loaded on the spin columns by centrifugation (1 min, 16000 g, room temperature). Columns were washed once with 750 µl buffer PE and dried by an additional spin. DNA was eluted with 60–100 µl ddH_2O and used as template for real-time quantitative PCR using a Rotor-Gene RG-3000 (Corbett Research). The amount of immunoprecipitated material at a particular gene region was calculated as '% of Input' after subtraction the background signal, as determined by the 'beads-only' control.

For ChIP using S9.6 antibody (Figure 5.2A), cells were crosslinked either 3 or 7 minutes. ChIP protocol was followed as described above and 14.4 µl S9.6 were added to 'IP' samples.

2.7.7 RNA/DNA hybrid immunoprecipitation (RNA/DNA hybrid IP)

For RNA/DNA hybrid immunoprecipitation (RNA/DNA hybrid IP), HeLa cells were split one day before the experiment. Non-crosslinked HeLa cells at 85% confluency were harvested after extensive washes with ice-cold PBS (Lonza). Cells were lysed in ChIP Cell Lysis Buffer (85 mM KCl, 5 mM PIPES pH 8.0, 0.5% NP-40, 0.5 mM PMSF, 0.8 µg/µl pepstatin A, 1 µg/µl leupeptin) for 10 min on ice, followed by centrifugation at 500g for 5 min to collect the nuclei. The nuclei were resuspended in RSB+DSST buffer (10 mM TRIS pH 7.5, 200 mM NaCl, 2.5mM MgCl2, 0.5% Triton X-100, 0.2% NaDOC, 0.1% SDS, 0.05% NaSarkosyl), to solubilise nuclear proteins and chromatin. Following 10 min incubation on ice, extracts were sonicated for 10 min using the

Diagenode Bioruptor (Medium setting, 30 s on 30 s off). 100 µl fraction of the extracts diluted 1:4 in RSB+T was removed and stored as 'Input'. The remaining diluted extracts were subjected to immunoprecipitation using 100 µl (~2 µg) S9.6 antibody, bound to protein A dynabeads (Invitrogen), pre-blocked with 0.5% BSA in PBS. As a control, 2 µg of CBP80 (sc-48803, Santa Cruz) or 2 µg of IgG2a (M5409, Sigma) antibodies were used instead. RNase A (PureLink, Invitrogen) was added during IP at 0.1 ng RNase A/µg genomic DNA. After 2 h, RNA/DNA hybrids and associated proteins were collected and extensively washed 4x with RSB+T and 2x with RSB. Washed immuno-complexes were analysed by SDS-PAGE and RNA/DNA hybrid slot blot. Samples for protein analysis were eluted using 2x LDS (Invitrogen) containing 100 mM DTT by incubation for 10 min at 70°C. For RNA/DNA hybrid slot blot analysis, DNA containing RNA/DNA hybrids was eluted at room temperature with elution buffer containing 1% SDS, 0.1 M $NaHCO_3$.

Where indicated, benzonase nuclease (Sigma) was added to nuclear extracts at a concentration of 1 U/µl for 30 min at 37°C prior to IP with S9.6 antibody. Sequences for RNA/DNA hybrid competitors were described (Phillips et al., 2013) and are listed in Table 1. Double-stranded competitors were prepared as described (Rigby *et al.*, 2014). Briefly, annealing reactions containing 40 µM of each complementary oligonucleotide in 60 mM KCl, 50 mM TRIS pH 8.0 were denatured at 95°C for 5 min, followed by slow cooling to room temperature.

2.7.8 Mass spectrometry

Sample preparation for mass spectrometry

For MS analysis, proteins were eluted from beads in LDS elution buffer by heating for 10 min at 70°C, followed by filter-aided sample preparation (FASP) with trypsin (Wisniewski *et al.*, 2009).

Sample processing

Samples were analysed on an Ultimate 3000 RSLCnano HPLC (Dionex, Camberley, UK) system run in direct injection mode coupled to a QExactive Orbitrap mass spectrometer (Thermo Electron, Hemel Hempstead, UK). Protein samples were resolved on a 25 cm by 75 micron inner

diameter picotip analytical column (New Objective, Woburn, MA, USA) which was packed in-house with ProntoSIL 120-3 C18 Ace-EPS phase, 3 µm bead (Bischoff Chromatography, Germany). The system was operated at a flow-rate of 300 nL min-1. A 120 min gradient was used to separate the peptides. The mass spectrometer was operated in a 'Top 20' data-dependent acquisition mode. Precursor scans were performed in the orbitrap at a resolving power of 70,000, from which the twenty most intense precursor ions were selected by the quadrupole and fragmented by HCD at a normalized collision energy of 30%. The quadrupole isolation window was set at 1.6 m/z. Charge state +1 ions and undetermined charge state ions were rejected from selection for fragmentation. Dynamic exclusion was enabled for 27s.

2.8 Bioinformatic analysis

2.8.1 Statistical analysis

Unless otherwise stated, the figures present the average values of at least three independent biological experiments +/- SEM. Asterisks (*) indicate statistical significance (* $p<0.05$; ** $p<0.01$; *** $p<0.001$), based on unpaired, two-tailed distribution Student's t test. Correlations were calculated using Pearson's correlation coefficient (r).

2.8.2 Quantitation of mass spectrometry data: MaxQuant and SINQ methods

Quantitation using MaxQuant

Mass spectrometry data processing for all figures, unless otherwise stated, was carried out using MaxQuant 1.5.0.35 and Andromeda search engine (Cox and Mann, 2008; Cox *et al.*, 2011). Enzyme specificity was set to trypsin/P, allowing a maximum of two missed cleavages. Cysteine carbamidomethylation was selected as fixed and protein N-terminal acetylation and methionine oxidation as variable modifications. Initial mass tolerance of precursor ions was set to 50 ppm. Proteins and peptides were identified with FDR<0.01 with a minimum peptide length of 7 amino acids. Protein identification required one unique peptide to the protein group.

For protein quantification a minimum of two ratio counts were set and the 'requantify' and 'match between runs' function enabled. The initial 848 identified proteins were filtered for

occurrence in at least three samples using Perseus 1.5.2.6. Common contaminants such as keratins and proteins of the large and small ribosomal subunits (RPL and RPS) were filtered out, due to their known contribution as contaminants and unresolved interactions in affinity purification procedures (Mellacheruvu *et al.*, 2013).

Quantitation using SINQ

For additional validation of the RNA/DNA hybrid interactome, an independent mass spectrometry quantitation pipeline was used, which is based on MS/MS spectra rather than ion intensities. For this, data were converted from .RAW to .MGF files using ProteoWizard (Chambers *et al.*, 2012). Data were analysed using the SINQ algorithm by the Central Proteomics Facility Pipeline software (Trudgian *et al.*, 2010). Peptide searches were performed using the InterProphet meta-search combining Mascot, X! Tandem with the k-score plugin, and OMSSA against concatenated target/decoy sequence databases. Proteins were identified with at least two peptide sequences, with at least one unique peptide, FDR<1%. Relative label-free quantitation of proteins was carried out using the normalized spectral index implemented in the SINQ software (Trudgian *et al.*, 2011).

2.8.3 Bioinformatic analysis of mass spectrometry data

Statistical analysis of the mass spectrometry data from RNA/DNA hybrid IP is based on 3 independent biological replicates and was performed essentially as described elsewhere (Castello *et al.*, 2012; Kwon *et al.*, 2013). In short, intensity values were $\log_2$-transformed and missing values were imputed with random numbers from a normal distribution to simulate low abundance values below noise level in Perseus 1.5.2.6 as described (Raschle *et al.*, 2015). Using the limma package in R/Bioconductor, a linear model was fitted to these data to calculate the $\log_2$ enrichment between RNA/DNA hybrid IP and control samples. An empirical Bayes moderated t-test was used to calculate p-values (Smyth, 2004) . P-values were then corrected for multiple testing by Benjamini-Hochberg method. Proteins enriched in RNA/DNA hybrid IP compared to control were included in the "RNA/DNA hybrid Interactome" if their corrected p-values were <0.01.

Cellular compartment analysis

Cellular compartment analysis was based on GO term cellular component analysis. Fisher's exact test was used to calculate statistical enrichment, using Benjamini-Hochberg correction for multiple testing. Proteins were classified into the groups 'nucleus', 'nucleolus', 'nucleoplasm' and 'cytoplasm', with the latter group consisting of proteins that are exclusively cytoplasmic.

Over-representation analysis

Over-representation analysis of protein classes was performed using the PANTHER database (www.pantherdb.org), version 10.0 (Mi *et al.*, 2013). PANTHER protein classes over-represented in the RNA/DNA hybrid IP were ranked according to their Benjamini-Hochberg-corrected p-values (p-value threshold 0.05) and 16 most-significant groups were manually curated for redundancy.

Venn diagram

The Venn diagram representing the overlap between RNA/DNA hybrid interactome and the HeLa mRNA interactome was prepared using Venn Diagram Plotter (http://omics.pnl.gov), with data for the mRNA interactome obtained from (Castello *et al.*, 2012). Proteins identified both in RNA/DNA hybrid and HeLa interactome and proteins exclusively identified in RNA/DNA hybrid interactome were further examined for over-represented protein classes using PANTHER, as described above.

Chromatin probability assignment

Chromatin probability assignment was based on previously published data (Kustatscher *et al.*, 2014) by assigning Interphase Chromatin Probability (ICP) values to the proteins in the RNA/DNA hybrid interactome. As comparison, ICP values are shown for the whole list of 7635 HeLa proteins as provided in (Kustatscher *et al.*, 2014).

Mutation and expression alterations in cancer

Analysis of genetic alterations of helicases identified in RNA/DNA hybrid interactome was carried out using the COSMIC database (cancer.sanger.ac.uk) (Forbes *et al.*, 2015). In brief, total number of copy number variations (either gain or loss) were retrieved for each gene and plotted

alongside two tumour suppressor genes and oncogenes. Differential mRNA expression analysis in cancer was performed using the ONCOMINE platform (www.oncomine.org) (Rhodes *et al.*, 2007), with a p-value threshold of <0.05 and a minimal fold-change of 2 between cancer and matched control samples. Cancer case studies with significant alterations were grouped based on whether the gene was amongst the top 1%, top 5%, or top 10% of all altered genes.

Chapter 3

Transcriptional regulation and R-loop formation in FRDA and FXS

3.1 *FXN* transcription initiation and elongation are impaired in FRDA

The genetic cause of Friedreich ataxia (FRDA) was identified in 1996 in the form of the GAA expansion in the first intron of the frataxin (*FXN*) gene (Campuzano *et al.*, 1996). However until now, the underlying molecular mechanism leading from expanded DNA to reduced frataxin protein levels remains largely elusive. The main objective of the present work therefore was to investigate the transcriptional regulation of the *FXN* gene in control and FRDA patient cells and to identify potential mechanisms linking the presence of expanded triplet repeats to the transcriptional repression of *FXN* gene.

3.1.1 *FXN* mRNA levels are reduced in FRDA

To define the transcriptional regulation of the *FXN* gene in a pathological context, Epstein-Barr virus (EBV)-immortalised lymphoblastoid cells established from either healthy individuals or FRDA patients were purchased from the Coriell Institute for Medical Research (Camden, USA) and used as cellular models (Figure 3.1A). The majority of the following experiments were carried out in lymphoblastoid cell lines from an FRDA patient (Coriell ID: GM15850) and his unaffected sibling (GM15851), thus providing a genetically closely matched model system to study FRDA. Analysis of steady-state mRNA levels showed that *FXN* expression is reduced by approximately 80% in FRDA cells compared to control cells (Figure 3.1B). Importantly, mRNA levels of several housekeeping genes, including *β-Actin, γ-Actin* and *GAPDH* were comparable between control and FRDA cells (Figure 3.1B). The reduction of *FXN* mRNA is consistent with previous observations in primary peripheral mononuclear blood cells, primary buccal cells, and iPSC-derived neurons, thus confirming the validity of this model system to study FRDA (Abrahao *et al.*, 2015; Bird *et al.*, 2014; Coppola *et al.*, 2011; Eigentler *et al.*, 2013; Lazaropoulos *et al.*, 2015; Soragni *et al.*, 2014).

3.1.2. RNA Pol II levels are decreased at the *FXN* promoter in FRDA

In order to investigate the transcriptional and epigenetic landscape of the *FXN* gene in greater detail, a set of eight qPCR primer pairs was employed, spanning the whole *FXN* gene body with two amplicons directly surrounding the GAA triplet repeat: amplicon C (382 bp upstream of the GAA expansion) and amplicon D (214 bp downstream of the GAA repeat) (Figure 3.1C, sequences in

A

Phenotype	Expansion	Coriell ID
Friedreich ataxia	**$(GAA)_n$**	
Control 1	<10	GM15851
Control 2	normal range	GM06895
Control 3	normal range	GM14926
FRDA 1	650/1030	GM15850
FRDA 2	670/1170	GM16243
FRDA 3	800/800	GM16209
Fragile X syndrome	**$(CGG)_n$**	
Control	23	GM06895
FXS	530	GM03200

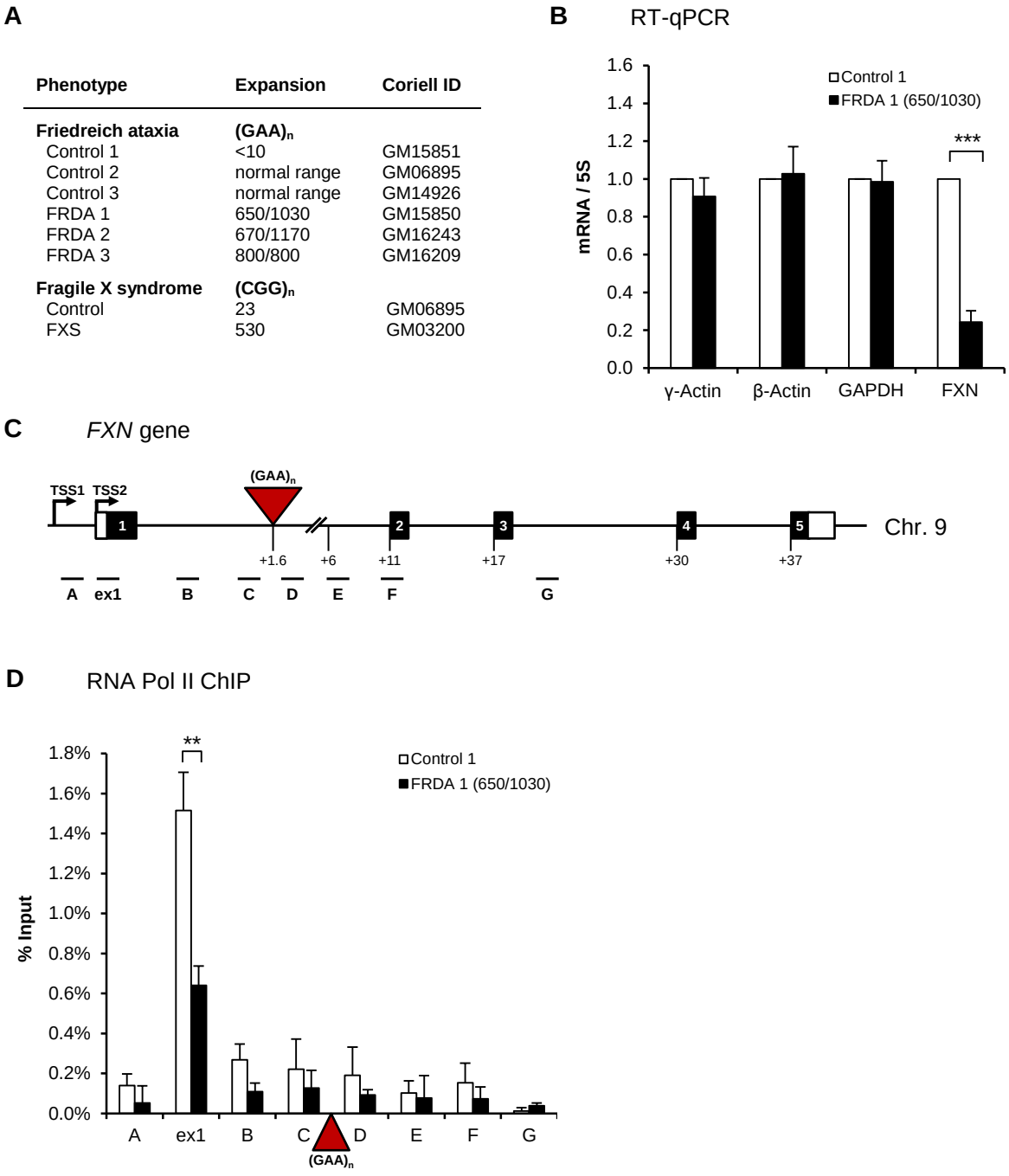


Figure 3.1. *FXN* gene transcription is impaired in FRDA cells

A. Lymphoblastoid cell lines used in the study. The repeat sizes are indicated.

B. RT-qPCR analysis of *γ-Actin*, *β-Actin*, *GAPDH* and *FXN* mRNAs in control 1 (GM15851) and FRDA 1 (GM15850) cells. Values are normalized to 5S rRNA and relative to control cells.

C. Diagram of *FXN* gene. Black boxes are exons, white boxes are 5' and 3'UTRs, lines are introns, red triangle is $(GAA)_n$ expansion. TSS2 is the major transcriptional start site in lymphoblastoid cells. qPCR amplicons are shown below the diagram. Numbers indicate the distances to TSS2 in kilobases.

D. RNA Pol II ChIP in control 1 (GM15851) and FRDA 1 (GM15850) cells, using H-224 antibody recognising total (CTD-modified and unmodified) RNA Pol II.

Table 2.1). Chromatin immunoprecipitation (ChIP) analysis using antibody recognising total RNA Pol II in control lymphoblasts showed that it is highly enriched over the ex1 amplicon, which is located 54 bp downstream of the major transcriptional start site (TSS2) in lymphoblasts (Chutake *et al.*, 2014; Kumari *et al.*, 2011) (Figure 3.1D, white bars). This promoter-proximal peak of RNA Pol II is a common feature of many human genes and may represent paused RNA Pol II which is stalled at the early elongation checkpoint (Jonkers and Lis, 2015). Interestingly, a two-fold reduction of RNA Pol II levels was detected at ex1 in FRDA cells compared to control cells (Figure 3.1D, black bars). This suggests that there is a transcription initiation defect in FRDA cells. Although not statistically significant, RNA Pol II levels were also reduced in the *FXN* gene body in FRDA cells (amplicons B, C, D and F). However, limitations of ChIP sensitivity precluded detailed analysis of potential elongation defects within the *FXN* gene body in FRDA cells.

3.1.3 *FXN* pre-mRNA levels are reduced in FRDA

The above results indicated that transcription initiation of the *FXN* gene may be impaired in FRDA cells. To investigate transcription in the *FXN* gene body, quantitative reverse transcription-PCR (RT-qPCR) was employed, using individual strand specific primers for reverse transcription (Figure 3.1C). As shown in Figure 3.2A, RNA levels were dramatically reduced both close to the TSS2 (ex1 amplicon) and upstream of the GAA repeats (5.5- and 2.6-fold reduction at amplicons B, C). It should be noted that due to its location, ex1 amplicon detects both pre-mRNA and mRNA. Therefore, amplicons B and C are more reliable readouts for pre-mRNA levels upstream of the GAA repeats. This reduction in pre-mRNA levels close to the promoter (amplicons B, C) thus confirmed the defect in transcription initiation as detected by RNA Pol II ChIP (Figure 3.1D).

An even more pronounced reduction in pre-mRNA levels was observed downstream of the triplet repeats (8.4–9.3-fold reduction at amplicons D–G). This was validated in two additional control and FRDA cell lines, suggesting that transcriptional dysregulation of *FXN* gene is a genuine feature of FRDA cells (Figures 3.2B, C). Interestingly, an overall analysis of the *FXN* pre-mRNA levels in all three control and FRDA cell lines highlighted a prominent transcriptional defect

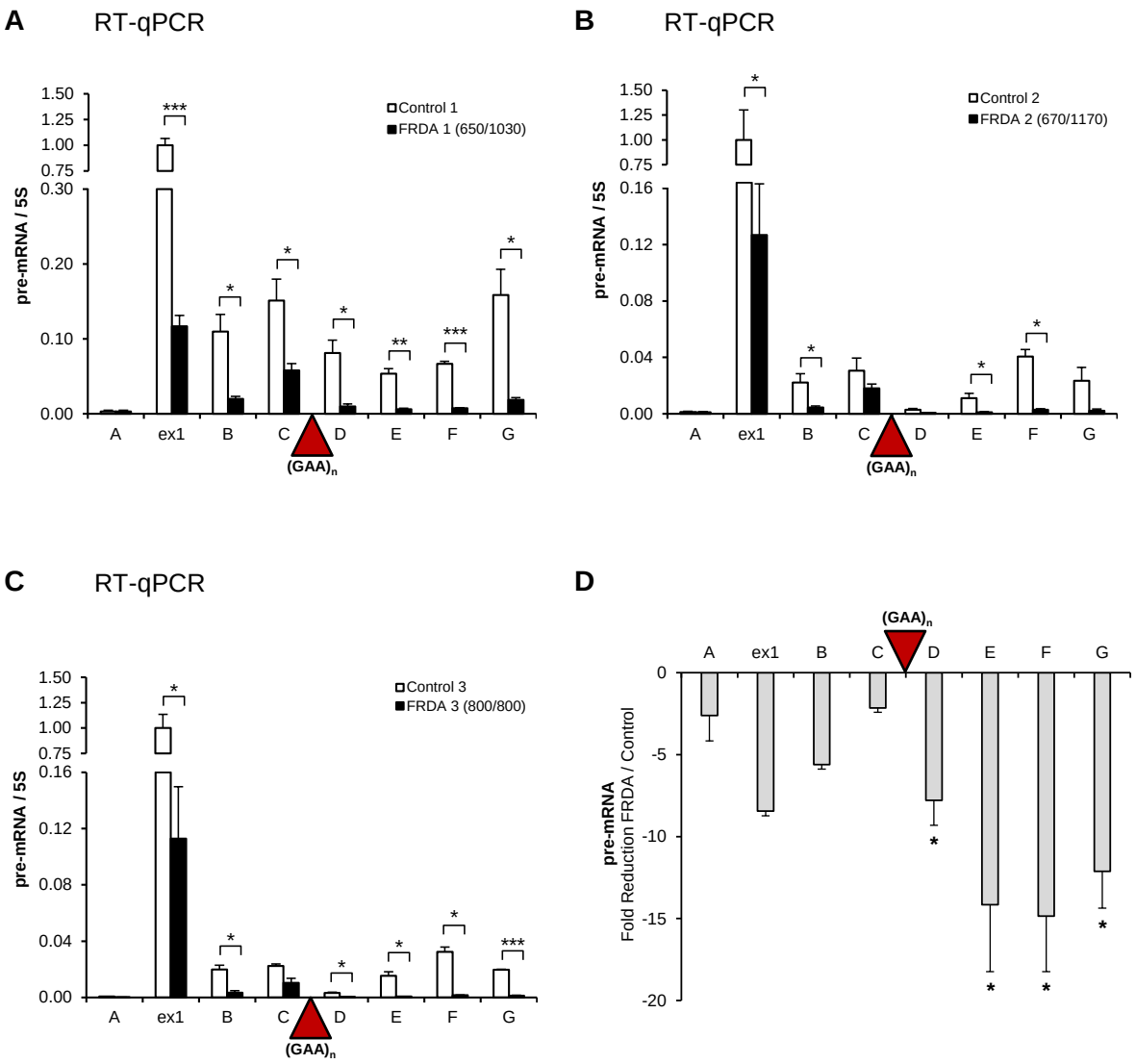


Figure 3.2. Pre-mRNA analysis reveals transcription initiation and elongation defects

A–C. RT-qPCR analysis of *FXN* pre-mRNA in three control (GM15851, GM06895, GM14926) and three FRDA (GM15850, GM16209, GM16243) cell lines, respectively. The values are normalized to 5S rRNA and relative to *FXN* ex1 RNA in control cells.

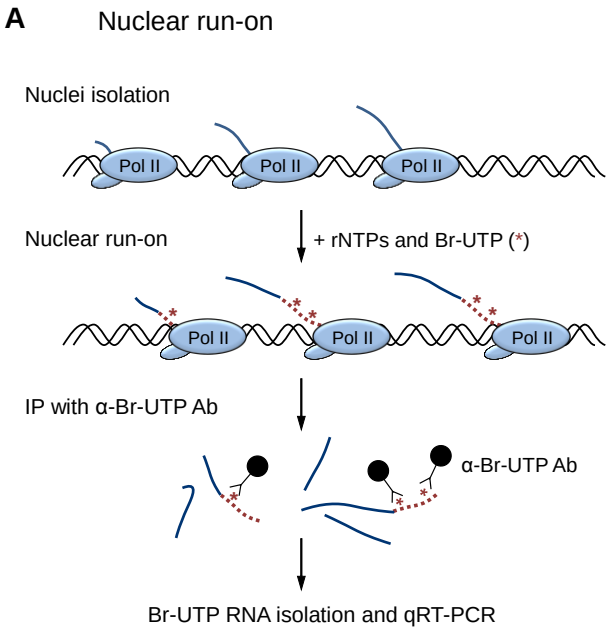
D. Analysis of *FXN* pre-mRNA in three control and three FRDA cell lines. Values represent averages of pre-mRNA in control versus FRDA lymphoblasts. Asterisks indicate statistically significant differences ($p<0.05$) compared to amplicon C, based on Student's unpaired t-test. In contrast to all other *FXN* amplicons, the ex1 amplicon recognises both *FXN* pre-mRNA and *FXN* mRNA.

downstream of expanded repeats, compared to amplicon C just upstream of the GAA repeats (Figure 3.2D). This suggests that a deficiency in transcription elongation may contribute to reduced *FXN* mRNA levels, thus playing an important role in FRDA pathology.

3.1.4 Transcription elongation of *FXN* gene is defective in FRDA

Although RT-qPCR based analysis of *FXN* pre-mRNA suggested decreased RNA levels downstream of expanded GAA repeats, in line with an RNA Pol II elongation defect (Figures 3.1D, 3.2), a contribution of post-transcriptional effects, including RNA stability, could not be fully excluded. It was therefore investigated if the GAA expansion in the *FXN* gene could affect RNA Pol II elongation directly. For this, nuclear run-on (NRO) analysis with Br-UTP labelled nucleotides was employed (Jackson *et al.,* 1998; Skourti-Stathaki *et al.,* 2011). In Br-UTP NRO, intact nuclei are isolated and resuspended in transcription buffer containing all ribonucleotides and Br-UTP (Figure 3.3A). Transcriptionally competent initiated RNA Pol II molecules are then allowed to transcribe for 30 min (run-on), thereby producing Br-UTP-labelled nascent RNA which can be distinguished from pre-existing RNA by selective immunoprecipitation of labelled RNA using anti-Br-UTP antibody, followed by RT-qPCR (Figure 3.3A). Br-UTP NRO therefore specifically detects transcriptionally active RNA Pol II, in contrast to the ChIP, which detects both paused and actively elongating RNA Pol II (Figure 3.1D).

Br-UTP NRO revealed a substantial decrease in active transcription upstream of the GAA expansion (regions in1 and B) in FRDA cells, confirming the initiation defect observed by RNA Pol II ChIP (Figure 3.3B and 3.1D). In control cells, there was a 3.1-fold reduction in active transcription in *FXN* regions D and E, positioned 210 nt and 4.5 kb downstream of the GAA repeats, respectively (Figure 3.3B). It is possible that this decrease in active RNA Pol II is caused by the existence of the 10-30 GAA repeats present in control *FXN* alleles. However, the effect was more pronounced in FRDA cells, demonstrating an 8.8-fold transcriptional decrease downstream of the expanded repeats. This elongation defect is unlikely to be due to the increased distance caused by GAA expansion (~3 kb), since there is no decrease in active transcription observed between regions D to E, separated by ~4.3 kb in both cell lines. Together, the Br-UTP NRO results suggest that expanded repeats directly



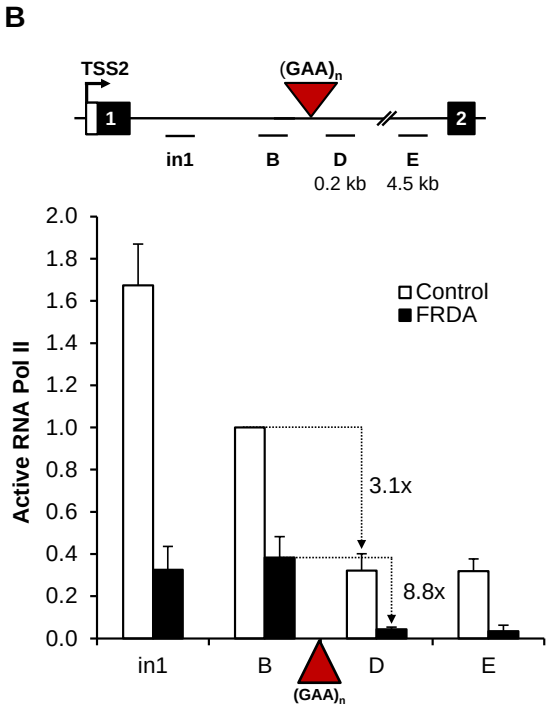


Figure 3.3. Br-UTP nuclear run-on analysis reveals elongation defect in *FXN* gene FRDA

A. Diagram depicting the Br-UTP nuclear run-on (NRO) method.

B. Br-UTP nuclear run-on analysis of *FXN* gene. Values represent averages from two control (GM15851, GM14926) and two FRDA (GM15850 and GM16243) cell lines and were normalised to the region B in control cells. The reduction of Br-UTP NRO signal over GAA repeats is indicated with dotted lines: A 3.1-fold reduction is detected in control cells, while an 8.8-fold reduction is detected in FRDA cells.

interfere with active RNA Pol II transcription elongation in FRDA cells. Additionally, RNA Pol II ChIP and RT-qPCR analysis of pre-mRNA also point to an initiation defect in the presence of expanded GAA repeats.

3.1.5 Heterochromatin formation at expanded GAA repeats

The presence of expanded GAA repeats may influence *FXN* transcription either by directly affecting RNA Pol II or by altering the epigenetic landscape of the *FXN* gene. Indeed, a GAA-expansion-containing transgene has previously been shown to promote heterochromatin formation in murine cells (Saveliev *et al.,* 2003). It was therefore important to investigate if heterochromatin formation due to expanded GAA repeats may also contribute to *FXN* gene repression in human cells. Since one of the major histone modifications associated with facultative heterochromatin is H3K9me2, I investigated its distribution across the *FXN* gene in control and FRDA cells.

H3K9me2 ChIP analysis of *FXN* gene in control cells showed that this modification peaks downstream of the GAA repeats (amplicons C, D, E) and that its levels are significantly higher than over the highly expressed *γ-Actin* gene (Figure 3.4A). H3K9me2 was strongly increased around expanded GAA repeats in FRDA cells (amplicons ex1– D), spreading both up- and downstream of the repeat region (Figure 3.4A). This broad H3K9me2 signal around GAA repeats was not due to technical limitations, as the employed ChIP method has a resolution of ~200 bp. This is evidenced by the clear distinction between the RNA Pol II peak at ex1 amplicon and background levels at the amplicon A, located 200 bp upstream (Figure 3.1D). The H3K9me2 enrichment in FRDA did not spread across the entire *FXN* gene (amplicon G is unaffected, Figure 3.4A) and levels at *γ-Actin* gene were equal between control and FRDA cells, thus confirming the specific and localised increase of this heterochromatin mark in FRDA cells. Importantly, these observations were not due to differences in total histone H3 levels between control and FRDA cells, as shown by H3 ChIP (Figure 3.4B). This analysis also revealed decreased levels of total histone H3 at the ex1 amplicon in control and FRDA cells, indicative of a nucleosome-depleted region often found associated with active promoters (Jonkers and Lis, 2015), thus confirming the use of TSS2 as the major transcriptional start site in lymphoblastoid cells (Figure 3.1C).

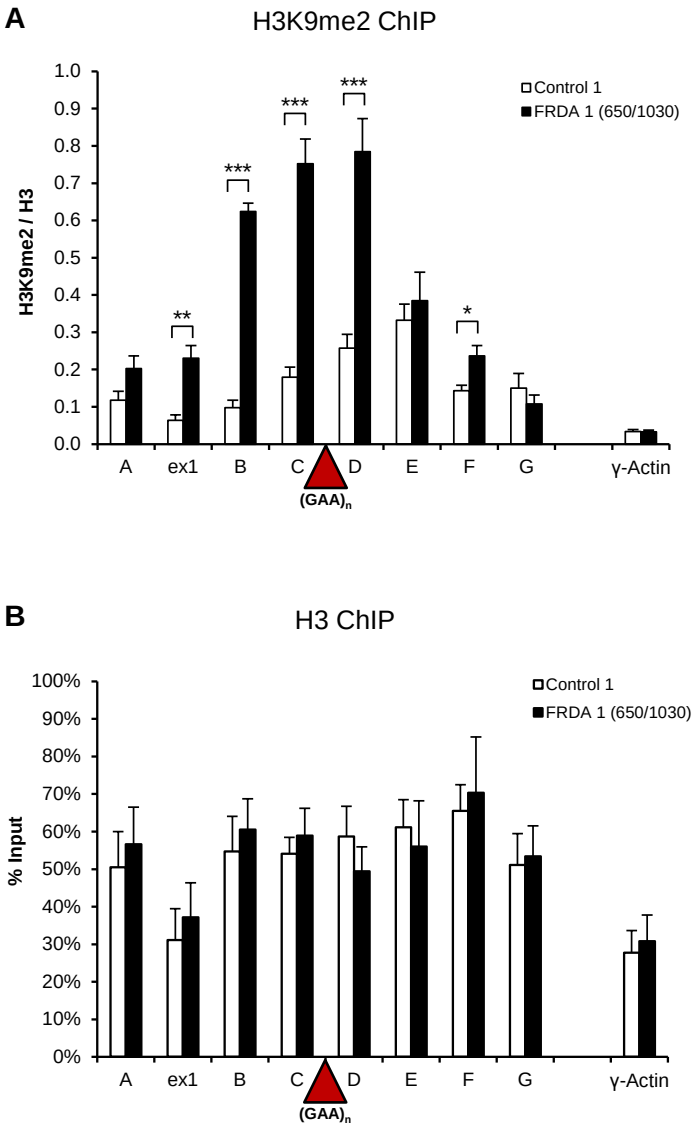


Figure 3.4. Increased heterochromatin formation at expanded GAA repeats in *FXN* gene

A. ChIP for heterochromatin histone modification H3K9me2 in control 1 (GM15851) and FRDA 1 (GM15850) cells. H3K9me2 levels were normalized to the total H3 levels. *γ-Actin* is used as background control.

B. ChIP analysis using antibody recognising total histone H3 in control 1 (GM15851) and FRDA 1 (GM15850) cells.

3.2 R-loop detection at GAA repeats of *FXN* gene *in vivo*

3.2.1 RNA/DNA hybrids are enriched at expanded GAA repeats of *FXN* gene *in vivo*

DNA containing GAA repeats exhibits distinct biochemical properties and it has been demonstrated that it is able to form unusual DNA and DNA/RNA structures such as R-loops *in vitro* (Wells, 2008). It has further been proposed that these structures may contribute to *FXN* gene repression, either in conjunction with or independently of heterochromatin formation (Kumari and Usdin, 2012; McIvor *et al.*, 2010; Wells, 2008). Although R-loop formation at GAA repeats has been demonstrated on plasmid constructs *in vitro* and in *E. coli* (Grabczyk *et al.*, 2007), limitations in methodology appropriate to detect endogenous R-loops in mammalian cells has prevented investigation of their formation *in vivo* up to now.

Previously, native bisulfite sequencing has been used to detect R-loops over endogenous *Ig* class switch regions and CAG repeats *in vivo* (Lin *et al.*, 2010; Yu *et al.*, 2003). This method detects ssDNA as part of R-loops and it is based on the treatment of non-denatured DNA with bisulfite, which leads to the conversion of cytosine to uracil in ssDNA but does not affect cytosines in dsDNA. Since R-loops formed in the sense direction on the GAA repeats would only leave the non-template GAA strand exposed, no cytosines will be available for bisulfite conversion. Thus, this method would be unable to detect sense R-loops at GAA repeats, in addition to the problems arising from the necessary amplification, cloning and sequencing of large repetitive DNA sequences. An alternative method to detect endogenous R-loops at expanded GAA repeats was therefore required.

Recently, the DNA/RNA immuno-precipitation (DIP) method has been established by Dr N. Gromak and Dr K. Skourti-Stathaki (University of Oxford, UK), which allows quantitative detection of RNA/DNA hybrids on any endogenous genomic location in human cells (Skourti-Stathaki *et al.*, 2011). DIP is based on the recognition of RNA/DNA hybrids by the monoclonal mouse S9.6 antibody (Boguslawski *et al.*, 1986). *In vitro*, the S9.6 antibody recognises RNA/DNA hybrids as short as 6 bp in a sequence-independent manner with high affinity (~0.5 nM) (Phillips *et al.*, 2013). Furthermore, since its original establishment in human cells, many independent groups have successfully employed DIP or variations of the method in *S. cerevisiae*, *A. thaliana* and in

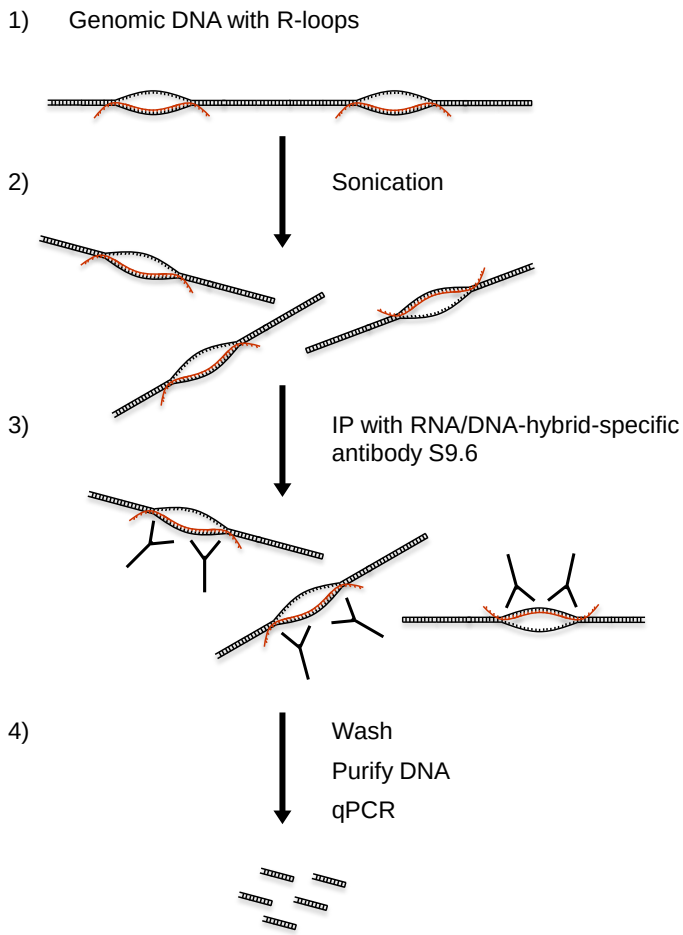


Figure 3.5. Overview of the DIP method

Diagram depicting the DNA/RNA immunoprecipitation (DIP) method. 1) Nuclei are extracted from non-crosslinked cells and proteins are removed by proteinase K treatment. Genomic DNA containing RNA/DNA hybrids (R-loops) is precipitated. 2) Purified genomic DNA with R-loops is sonicated. 3) RNA/DNA hybrids are immunoprecipitated using RNA/DNA-hybrid-specific antibody S9.6. 4) Immuno-precipitated R-loops are extensively washed, followed by DNA purification and qPCR analysis.

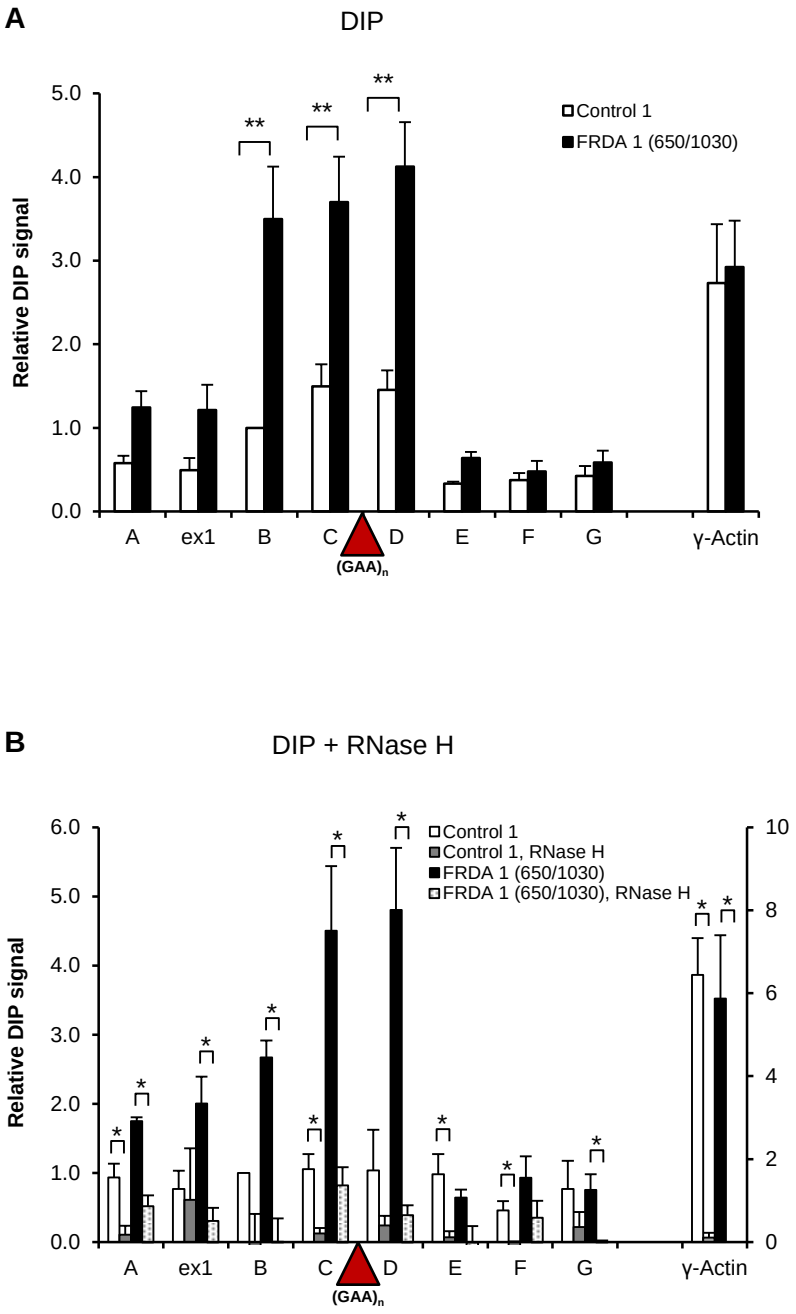


Figure 3.6. RNA/DNA hybrids are formed over expanded repeats of *FXN* gene in FRDA cells

A. DIP on endogenous *FXN* gene in control 1 (GM15851) and FRDA 1 (GM15850) cells. The first intron of *γ-Actin* gene is the positive control. The values are normalized to FXN B amplicon in control cells.

B. RNA/DNA hybrids are sensitive to RNase H digestion. DIP samples were treated with 25U of recombinant *E. coli* RNase H (NEB, M0297S) for 6 hours at 37ºC prior to addition of S9.6 antibody. The *γ-Actin* gene intron 1 probe serves as positive control.

mammalian cells (Arora *et al.*, 2014; Garcia-Rubio *et al.*, 2015; Ginno *et al.*, 2012; Herrera-Moyano *et al.*, 2014; Sun *et al.*, 2013; Yang *et al.*, 2014; Yeo *et al.*, 2014). In DIP, nuclei are extracted from non-crosslinked cells, proteins are removed and genomic RNA/DNA hybrids are immunoprecipitated using the S9.6 antibody, followed by extensive washes, DNA purification and qPCR analysis (Figure 3.5). Importantly, DIP does not require the presence of cytosines in the non-template strand, in contrast to native bisulfite sequencing. Therefore, DIP was employed to investigate RNA/DNA hybrid distribution on the *FXN* gene. The method was adapted for use with human control and FRDA lymphoblastoid cells by reducing sonication duration to 3 minutes to be able to detect RNA/DNA hybrid arising directly over the expanded GAA repeats, which cannot be probed directly by qPCR-based methods. As a positive control the intron 1 region of the *γ-Actin* gene was used, where high levels of RNA/DNA hybrids have previously been predicted (Ginno *et al.*, 2012; Wongsurawat *et al.*, 2012) and experimentally detected in HeLa cells (Dr N. Gromak, pers. comm.).

In control cells, RNA/DNA hybrids peaked around the GAA repeats (amplicons B–D), compared to other regions of the *FXN* gene (Figure 3.6A). This could be due to the presence of 10–30 GAA repeats in control cells. Significantly, a ~3-fold enrichment of RNA/DNA hybrids could be observed over regions B, C and D in the *FXN* intron 1 in FRDA cells, compared to control cells (Figure 3.6A). RNA/DNA hybrids were concentrated over the expanded repeat region and were low in the downstream regions E-G, demonstrating that RNA/DNA hybrid formation at expanded GAA repeats is a highly localised effect. Furthermore, *γ-Actin* RNA/DNA hybrid levels were similar in control and FRDA cells, indicating that increased RNA/DNA hybrid levels were specific for GAA repeats of the *FXN* gene and that they do not represent a global feature of multiple genes in FRDA cells (Figure 3.6A).

3.2.2 RNA/DNA hybrids at expanded GAA repeats are sensitive to RNase H *in vitro* and correlate with GAA expansion size

To confirm the specificity of the DIP signal, the DIP samples were treated prior to immunoprecipitation with recombinant *E. coli* RNase H, which specifically degrades the RNA in RNA/DNA hybrids (Cerritelli and Crouch, 2009). Following RNase H digestion, the signal was

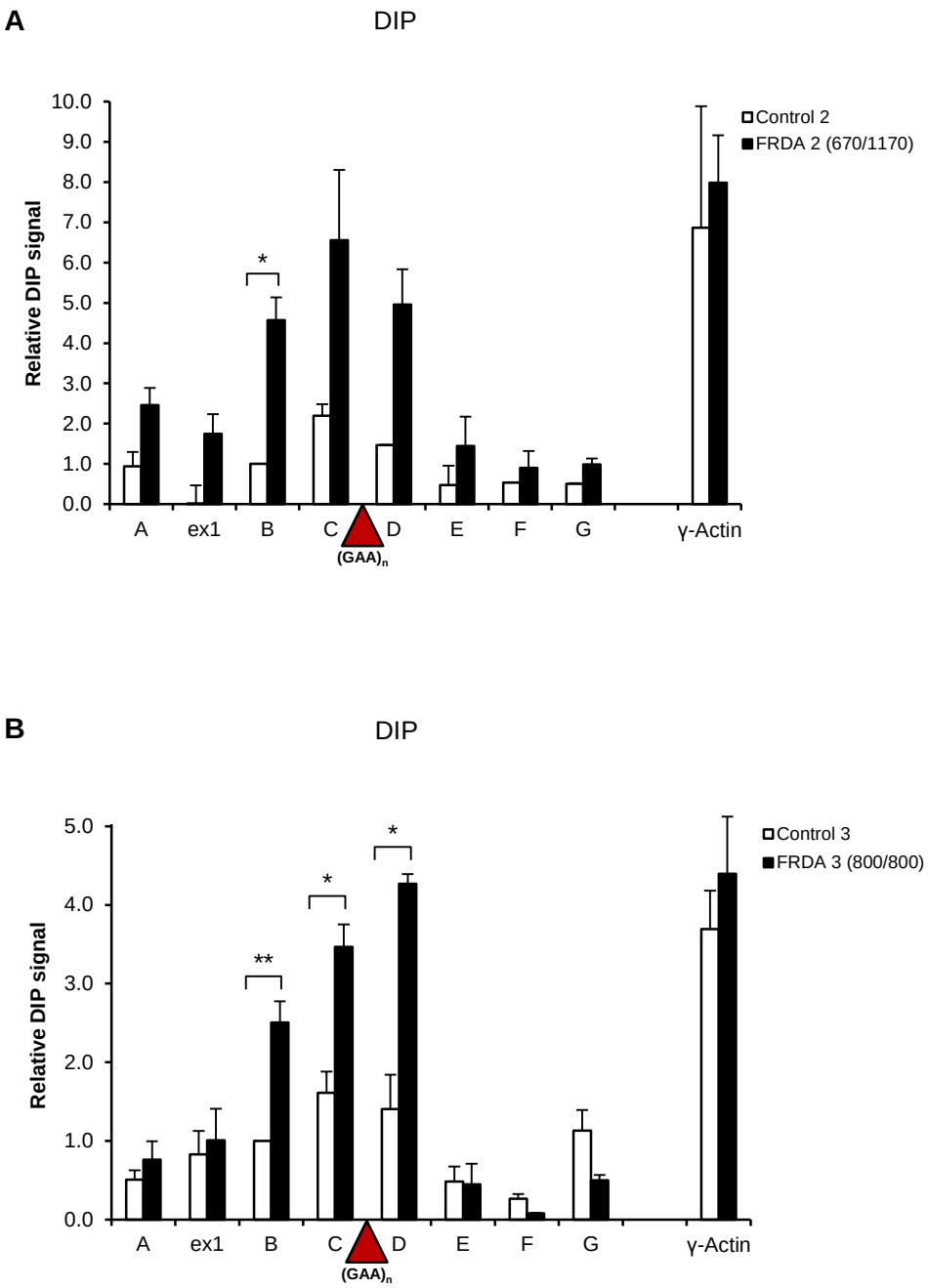


Figure 3.7. Confirmation of RNA/DNA hybrid formation in additional cell lines

A. DIP on endogenous *FXN* gene in control 2 (GM06895) and FRDA 2 (GM16209) cells. The *γ-Actin* gene serves as positive control. The values are normalized to FXN B amplicon in control cells.

B. DIP on endogenous *FXN* gene in control 3 (GM14926) and FRDA 3 (GM16243) cells. The *γ-Actin* gene serves as positive control. The values are normalized to FXN B amplicon in control cells.

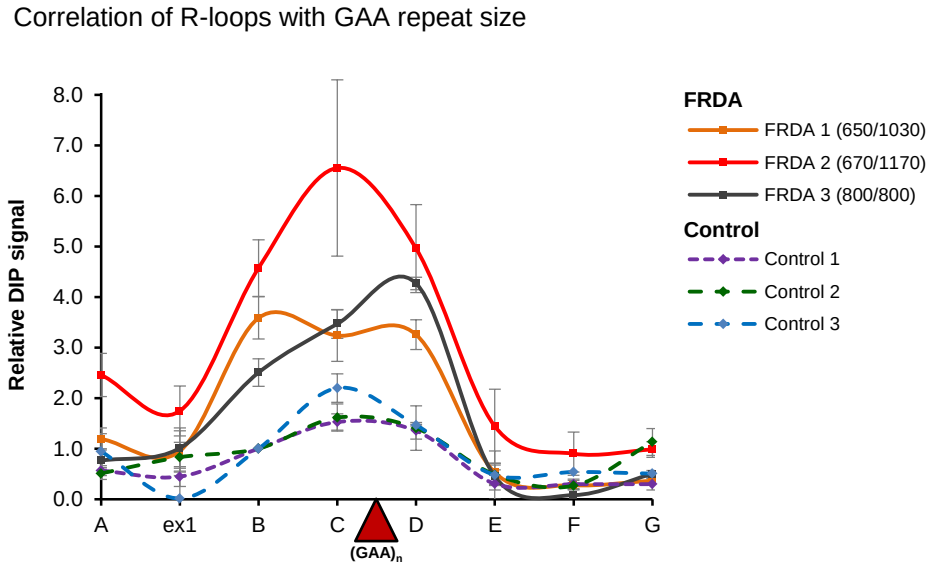


Figure 3.8. RNA/DNA hybrid levels correlate with GAA repeat size in FRDA cells

DIP analyses on endogenous *FXN* gene in three control (GM15851, GM06895, GM14926, indicated by blue, purple and green dashed lines) and three FRDA cell lines (GM15850, GM16209, GM16243, indicated by black, orange and red lines). The values are normalized to FXN B amplicon in control cells. Positions of the qPCR amplicons are on the x-axis; relative DIP signal is on the y-axis. Numbers in brackets indicate *FXN* GAA repeat lengths (small allele/large allele).

strongly reduced for the control *γ-Actin* gene and all *FXN* regions in both control and FRDA cells, suggesting that genuine RNA/DNA hybrids are formed over the expanded GAA repeats (Figure 3.6B).

In patient-derived cell lines, genetic and gene expression variability between individuals may contribute to observed effects (Min *et al.*, 2010). To mitigate this concern, DIP was carried out in two additional control and FRDA cell lines. Importantly, similar RNA/DNA hybrid enrichment over expanded GAA repeats was detected in two additional independent FRDA cell lines compared to control cells (Figure 3.7A, B). Interestingly, comparison of the DIP data from all three studied control and FRDA cell lines indicated that the level of RNA/DNA hybrids correlated with expansion length (Figure 3.8). High levels of RNA/DNA hybrids detected in FRDA cells may result from increased formation frequency due to expanded GAA repeats but it could also suggest that these structures are particularly stable over *FXN* expanded repeats. Therefore, RNA/DNA hybrids could act *in cis* to affect *FXN* gene expression in FRDA cells.

3.2.3 RNA/DNA hybrids at expanded GAA repeats are stable *in vivo*

The formation of RNA/DNA hybrids and concomitantly R-loops could affect transcription in a number of ways (see Chapters 1.1–1.3). In particular, GAA-repeat-associated RNA/DNA hybrids may persist long enough to act as a roadblock to subsequent RNA polymerases, which may lead to RNA Pol II slowing or premature transcriptional termination. Thus, to further characterize the RNA/DNA hybrids over the expanded *FXN* allele, their cellular half-life was studied. For this, cells were treated with the transcriptional inhibitor actinomycin D (Perry and Kelley, 1970). Following the treatment with 5 μg/ml actinomycin D for 21 hours, *γ-Actin* and *FXN* intronic pre-mRNAs were reduced approximately 80% (Figure 3.9A). This is in line with the short half-life of the majority of human introns (>65%) which are spliced out of pre-mRNA and degraded within 5 min of their production (Windhager *et al.*, 2012). Transcriptional inhibition led to an almost complete loss of RNA/DNA hybrid signal in *γ-Actin* suggesting that *γ-Actin* RNA/DNA hybrids are quickly turned over in the cell (Figure 3.9B). In contrast, although *FXN* pre-mRNA decreased following

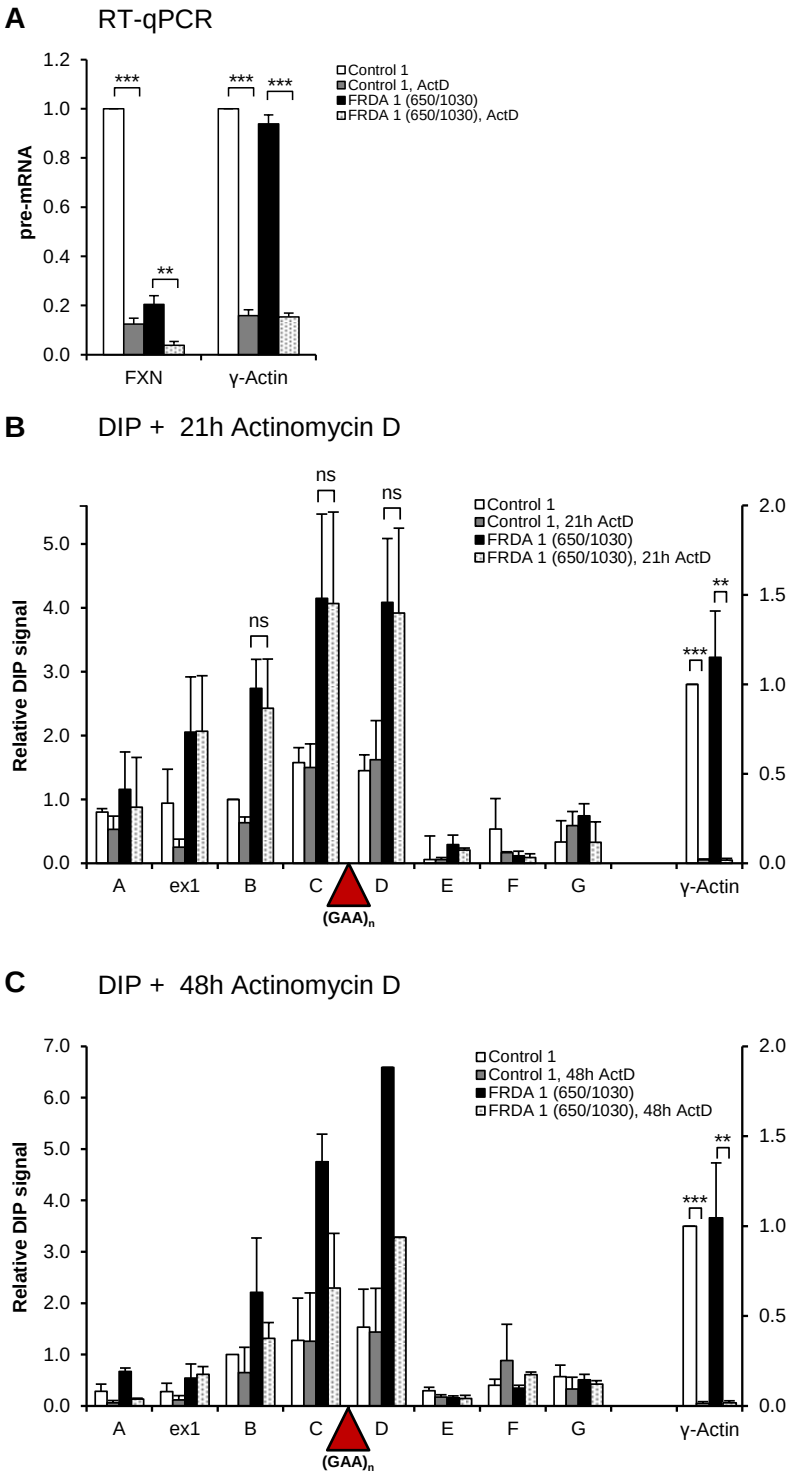


Figure 3.9. RNA/DNA hybrids formed at expanded GAA repeats in *FXN* gene are stable *in vivo*

A. RT-qPCR analysis of nascent *γ-Actin* and *FXN* RNA control 1 (GM15851) and FRDA 1 (GM15850) cells treated with 5 μg/ml of actinomycin D for 21 hours. Values are relative to untreated control cells.

B. DIP on *FXN* gene in control 1 and FRDA 1 cells treated with 5 μg/ml of actinomycin D for 21 hours. *γ-Actin* is positive control.

C. DIP on *FXN* gene in control 1 and FRDA 1 cells treated with 5 μg/ml of actinomycin D for 48 hours. *γ-Actin* is positive control.

actinomycin D treatment, RNA/DNA hybrid levels at expanded GAA repeats of *FXN* gene were not affected (Figure 3.9B). Importantly, the decrease in *FXN* and *γ-Actin* pre-mRNA, followed by reduction in RNA/DNA hybrids at *γ-Actin* but not *FXN* gene was already observed after a shorter treatment of 6 hours (data not shown). A significant decrease in the levels of RNA/DNA hybrids over expanded repeats was finally observed following prolonged treatment with actinomycin D for 48 hours (Figure 3.9C). These results suggest that RNA/DNA hybrids associated with *FXN* expanded repeats are resistant to degradation *in vivo*. Failure of normal cellular mechanisms to remove RNA/DNA hybrids at expanded GAA repeats may directly relate to their pathological potential.

3.2.4 Stable RNA/DNA hybrids correlate with H3K9me2 and gene repression

I next investigated the correlation between stable RNA/DNA hybrids and the repressive histone modification H3K9me2, a hallmark of FRDA (Chapters 1.3.6, 3.2.5). Interestingly, enrichment of stable RNA/DNA hybrids colocalises with the highest peaks of H3K9me2 over the *FXN* regions B-D in FRDA cells (Figure 3.10A). This suggests that RNA/DNA hybrids over expanded GAA repeats may functionally associate with repressive heterochromatin and that they may be involved in mediating transcriptional repression. In contrast to that, RNA/DNA hybrids in the first intron of *γ-Actin* gene, which are quickly turned over (Figure 3.9B), are associated with low levels of H3K9me2 (Figure 3.4A).

Moreover, a quantitative correlation analysis revealed that RNA/DNA hybrid levels in the *FXN* gene are significantly correlated with H3K9me2 levels in FRDA cells (r=0.938), but not in control cells (r=0.149). Both colocalisation and quantitative correlation analyses thus suggest a functional interplay between RNA/DNA hybrids and heterochromatin at expanded GAA repeats which could contribute to FRDA pathology. In line with this, results from both RT-qPCR and Br-UTP NRO analysis of active RNA Pol II indicated the presence of a transcription elongation defect due to expanded GAA repeats, correlating with the highest peaks of the stable RNA/DNA hybrids in FRDA (Figures 3.2, 3.3B, 3.10). Together, the formation of RNA/DNA hybrids in the genomic context of the FRDA gene as detected here suggests that, as a consequence, R-loops may be formed by displacement of the non-template strand as ssDNA. Further studies, including analysis of

alternative structures formed by the non-template strand, will be necessary in the future. The molecular mechanism linking GAA expansion with *FXN* gene transcriptional repression is further investigated in Chapter 4.

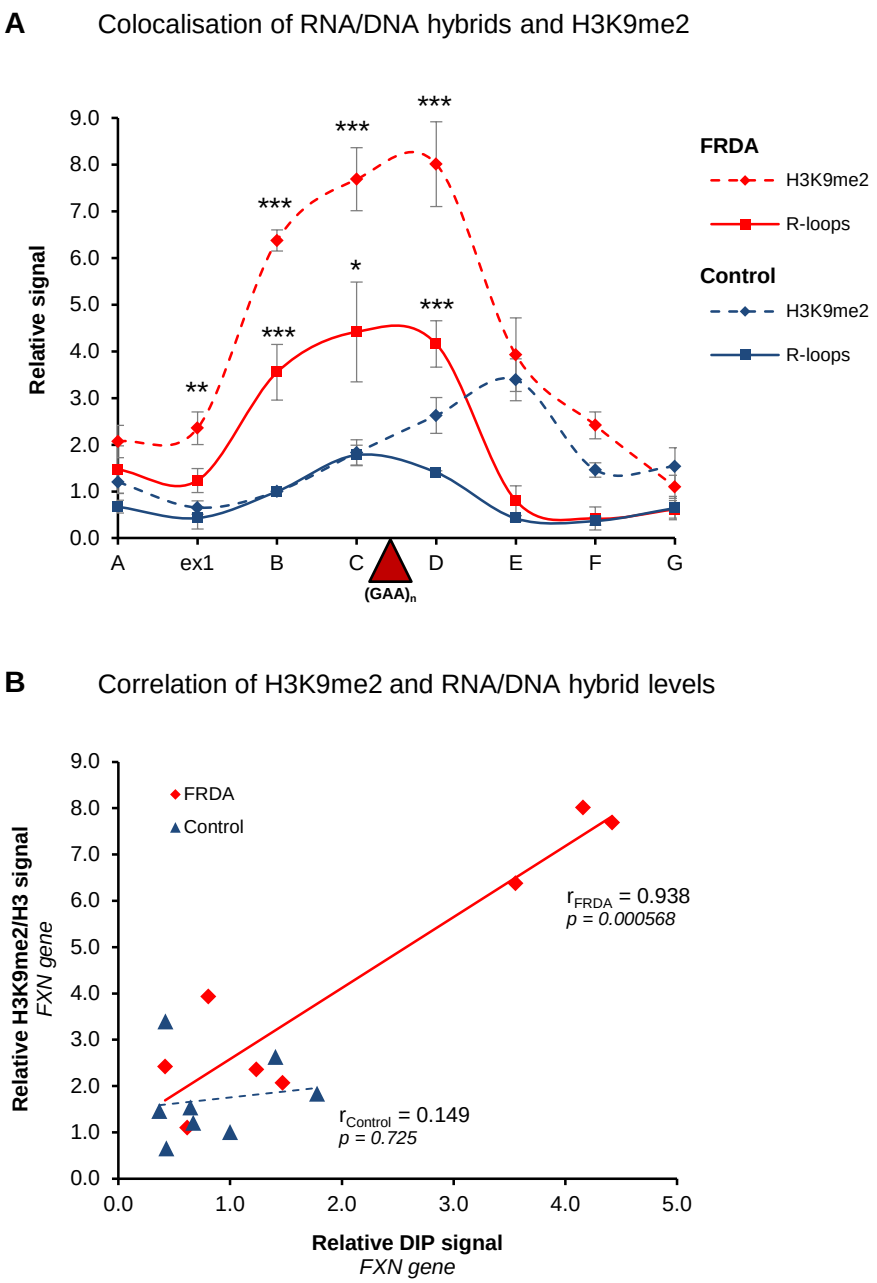


Figure 3.10. H3K9me2 correlates with RNA/DNA hybrids at expanded GAA repeats

A. Colocalisation of RNA/DNA hybrids and H3K9me2 in FRDA. R-loop levels are based on the average DIP signals from three control and three FRDA cell lines (data from Figures 3.6-3.7). H3K9me2/H3 levels are based on ChIP performed in control 1 and FRDA 1 cell lines (data from Figure 3.4). DIP and ChIP signals are normalised to amplicon B in control cells. Asterisks indicate significant enrichment in FRDA cells compared to control cells.

B. Correlation of RNA/DNA hybrids and H3K9me2 in *FXN* gene in FRDA. Individual data points represent the amplicons in *FXN* gene. Average values for DIP and H3K9me2/H3 are as described in A. Lines represent linear regressions and r is Pearson's correlation coefficient. The p-values indicate the significance of correlation.

3.3 R-loops at expanded CGG repeats in fragile X syndrome (FXS)

3.3.1 RNA/DNA hybrids are formed at transcribed expanded CGG repeats of *FMR1* gene in FXS

To test if RNA/DNA hybrids formation is a general feature of trinucleotide expansion diseases, the DIP method was also applied to examine the *FMR1* gene (Figure 3.11A). In fragile X syndrome (FXS) patients, the *FMR1* allele containing a $(CGG)_{n>200}$ expansion in the 5'UTR is fully methylated and transcriptionally silenced (Chapter 1.3.5) (Santoro *et al.*, 2011). Therefore to investigate the potential role of R-loops in FXS, *FMR1* transcription was reactivated by treatment with the DNA methylation inhibitor 5-aza-2'-deoxycytidine (5-azadC), as employed before (Coffee *et al.*, 2002). This resulted in expression of *FMR1* mRNA in FXS cells to 25% of control cells, as previously reported (Coffee *et al.*, 2002). In contrast, *FMR1* expression in control cells was unaffected by 5-azadC treatment, likely due to absence of DNA methylation of the *FMR1* promoter (Figure 3.11B). Using DIP, a low RNA/DNA hybrid signal was detected in control and untreated FXS cells (Figure 3.11C, black bars). Following 5-azadC treatment, a ~4-fold increase in RNA/DNA hybrids was observed over the exon 1 region upstream of the expansion in FXS cells, while no significant changes were detected in control cells, in line with mRNA data (Figures 3.11B, C). The specificity of the DIP signal was confirmed by treatment with recombinant *E. coli* RNase H prior to immunoprecipitation (Figure 3.11D).

The low DIP signal detected in FXS cells prior to 5-azadC reactivation may indicate the technical background or may be due to the ability of the S9.6 antibody to recognise dsRNA, potentially formed by antisense RNA at low affinity (Phillips *et al.*, 2013). However, the consistent and strong increase in RNA/DNA hybrids after 5-azadC treatment suggests that RNA/DNA hybrids in *FMR1* gene are transcriptionally-dependent and localise to the expanded CGG repeat region. Since inhibition of DNA methylation only partially reactivates the expanded *FMR1* allele in FXS cells, it is possible that RNA/DNA hybrids at expanded CGG repeats contribute to the mechanisms which prevent full *FMR1* reactivation.

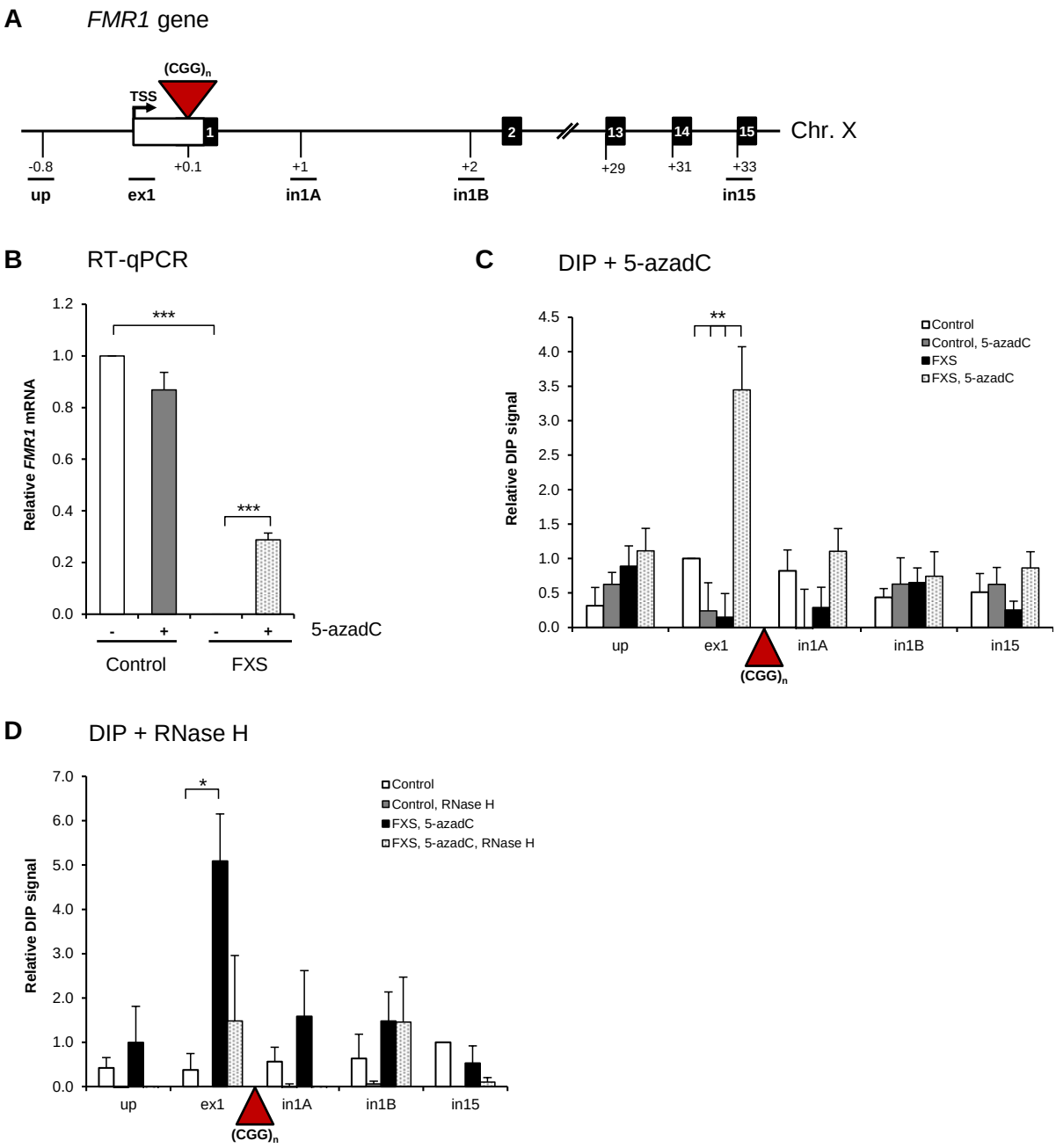


Figure 3.11. RNA/DNA hybrids are formed over CGG expanded repeats of *FMR1* gene

A. Diagram of *FMR1* gene. Black boxes are exons, white box is 5'UTR and lines are introns. Red triangle is $(CGG)_n$ expansion. qPCR amplicons are shown below the diagram. TSS is the transcriptional start site. Numbers indicate the distances to TSS in kilobases.

B. RT-qPCR analysis of *FMR1* mRNA in control and FXS cells, treated with 1 µM 5-azadC for 7 days, normalized to *GAPDH* mRNA levels.

C. DIP analysis on endogenous *FMR1* gene in control and FXS cells, treated with 1 µM 5-azadC for 7 days. Values are relative to ex1 region in control untreated cells.

D. *FMR1* RNA/DNA hybrids are sensitive to RNase H digestion, following the treatment with 25U of *E. coli* RNase H for 6 hours at 37ºC prior to immunoprecipitation. Values are relative to in15 region in control untreated cells.

3.3.2 Kinetic analysis of RNA/DNA hybrid formation and mRNA levels of *FMR1* gene

To further characterise the dynamic relationship between RNA/DNA hybrids and *FMR1* expression, a time course experiment was carried out. In particular, RNA/DNA hybrids and *FMR1* mRNA levels were measured during the process of transcriptional reactivation with 5-azadC treatment (7 days) followed by 5-azadC washout with drug-free media for 28 days (Figure 3.12A, B). As shown in Figure 3.12A, RNA/DNA hybrid levels over the exon 1 of *FMR1* gene stayed at the background during the activation and washout periods in control cells (dotted line). In FXS cells, RNA/DNA hybrids were at their peak during the reactivation procedure with 5-azadC on day 7. After removal of 5-azadC, RNA/DNA hybrid levels gradually diminished and completely disappeared after 14 days. This pattern of RNA/DNA hybrid dynamics correlated with the *FMR1* expression profile (Figure 3.12B), suggesting that RNA/DNA hybrids are associated with *FMR1* gene regulation. Interestingly, the maximum *FMR1* mRNA levels were detected on day 12, while the highest RNA/DNA hybrid peak was reached on day 7. The lag time between RNA/DNA hybrids and *FMR1* mRNA increases after 5-azadC washout may be caused by the additional time required to accumulate full-length spliced *FMR1* mRNA. Similarly, the longer persistence of *FMR1* mRNA after 5-azadC washout may be due to its long half-life of ~12 hours (Feng *et al.*, 1995).

3.4 Summary

Overall, the results obtained in this Chapter demonstrate that RNA/DNA hybrids are formed *in vivo* at endogenous expanded GAA repeats in the *FXN* gene and transcribed expanded CGG repeats in the *FMR1* gene, respectively (Figures 3.6A, 3.11C). The correlation of RNA/DNA hybrids levels with GAA repeat length (Figure 3.8) has important implications for the molecular pathology of FRDA since GAA repeat length determines the age-of-onset and the severity of disease progression, with longer repeats associated with worse clinical prognosis (Koeppen, 2011). The colocalisation of RNA/DNA hybrids with the gene repression-associated histone modification H3K9me2 (Figure 3.10) further suggests a functional interplay between RNA/DNA hybrids and transcriptional and epigenetic *FXN* gene regulation, potentially contributing to FRDA pathology. In conclusion, the results of this Chapter suggest that as a consequence to RNA/DNA hybrid formation, R-loops may be formed at

expanded CGG repeats of *FMR1* gene, supported by other methods which have detected the ssDNA of the displaced non-template strand at CGG repeats *in vitro* and *in vivo* (Loomis *et al.*, 2014; Reddy *et al.*, 2011). In Chapter 4, the mechanistic implications of RNA/DNA hybrids formation in FRDA are investigated, followed by a detailed discussion covering Chapters 3 and 4.

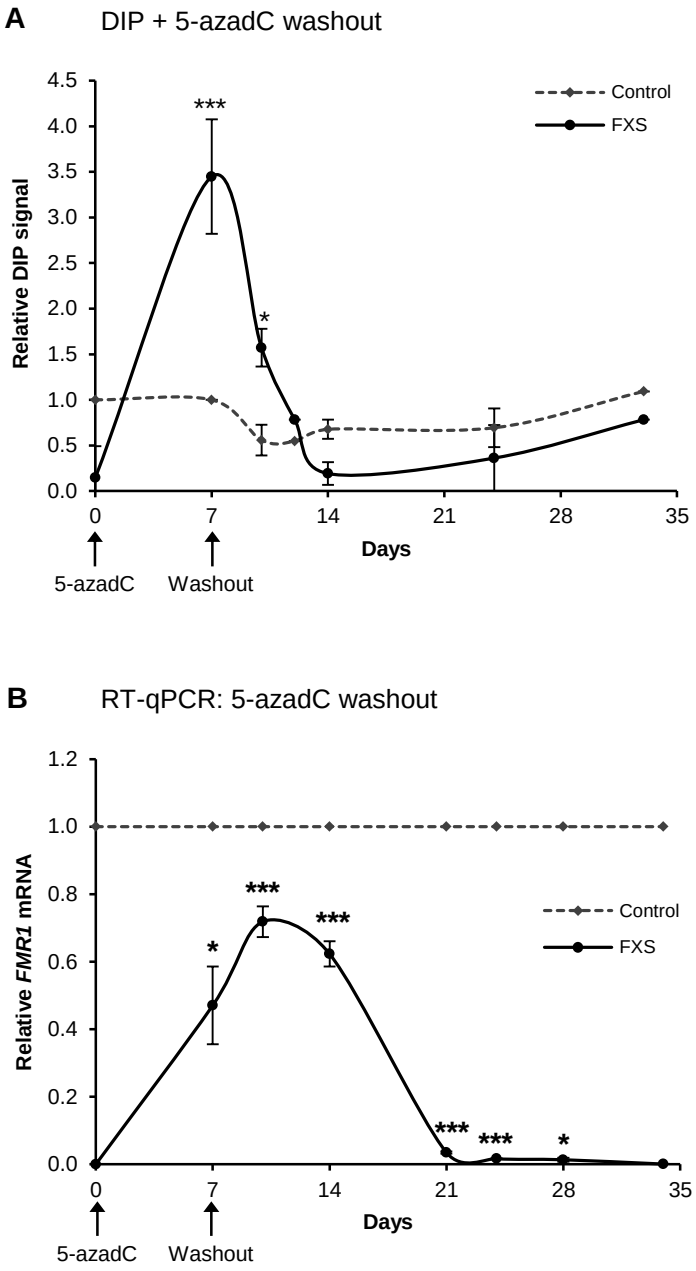


Figure 3.12. Kinetic analysis of RNA/DNA hybrid formation and mRNA levels of *FMR1* gene

A. RNA/DNA hybrid kinetics on exon 1 of *FMR1* gene in control and FXS cells during the process of transcriptional reactivation with 1 µM 5-azadC (7 days) followed by 5-azadC wash out with drug-free media (28 days). Values are relative to ex1 region in control untreated cells on day 7. Asterisks indicate statistically significant differences of DIP levels in *FXS* cells treated with 5-azadC compared to control cells.

B. RT-qPCR analysis of *FMR1* mRNA in control and FXS cells, treated with 1 µM 5-azadC (7 days) followed by 5-azadC wash out with drug-free media (additional 28 days). For all data points, the level of *FMR1* mRNA in control cells was taken as 1. Asterisks indicate statistically significant differences to *FMR1* mRNA levels in FXS cells not treated with 5-azadC.

Chapter 4

R-loops promote transcriptional repression of *FXN* gene in FRDA

4.1 R-loops and heterochromatin at expanded repeats in HEK293 cells

4.1.1 RNA/DNA hybrids are formed at expanded GAA repeats in HEK293 cells

To test if expanded GAA repeats trigger formation of RNA/DNA hybrids and heterochromatin in a different genomic and cellular context, a recently published cellular model system for FRDA based on a transgenic *FXN* gene inserted into HEK293 cells was employed. Using the Flp-In system (Life Technologies), HEK293 clonal cell lines had been generated containing an integrated transgenic copy of the complete human *FXN* gene with either six (*FXN-Luc*) or ~310 (*FXN-GAA-Luc*) GAA repeats, fused to the luciferase reporter gene, on chromosome 1 (Figure 4.1A) (Lufino *et al.*, 2013). It is important to note that in addition to this inserted transgene, the resulting cell lines contained three endogenous *FXN* gene alleles as determined by DNA FISH (Lufino *et al.*, 2013). These endogenous *FXN* genes are indistinguishable from the transgenic *FXN-Luc* and *FXN-GAA-Luc* alleles, except for the *luciferase* sequence.

First, the presence of the GAA expansion was validated using PCR on genomic DNA extracted from these cells. As demonstrated in Figure 4.1B, the FXN-GAA-Luc cell line indeed harboured ~310 expanded repeats. To specifically measure RNA levels of the *FXN-Luc* and *FXN-GAA-Luc* transgenes, reverse transcription was performed using a luciferase-specific primer (lucR), followed by qPCR with in4F/ex5R primers (Figure 4.1A). The presence of GAA repeats caused a ~37% reduction of *FXN-luciferase* (*FXN-Luc*) pre-mRNA levels as determined by RT-qPCR (Figure 4.1C), recapitulating the repression of gene expression seen in FRDA patient cells. The observed smaller reduction in the RNA levels in HEK293 cells can be explained by the lower number of GAA repeats (only ~310) on the integrated *FXN* copy, compared to 650–1170 GAA repeats present in *FXN* alleles of patient lymphoblasts used in this work (Figure 3.1A, 3.2A-C).

Using DIP, the formation of RNA/DNA hybrids in the *FXN-Luc* transgene was investigated. In HEK293 FXN-Luc cells, RNA/DNA hybrids were enriched around the GAA repeats (amplicons C, D) in the *FXN-Luc* gene, compared to regions up- and downstream (amplicons A, E–G) (Figure 4.2A). Furthermore, a two to three-fold increase in the levels of RNA/DNA hybrids over the expanded repeats region (amplicons B, C and D) in FXN-GAA-Luc cells was observed, similar to

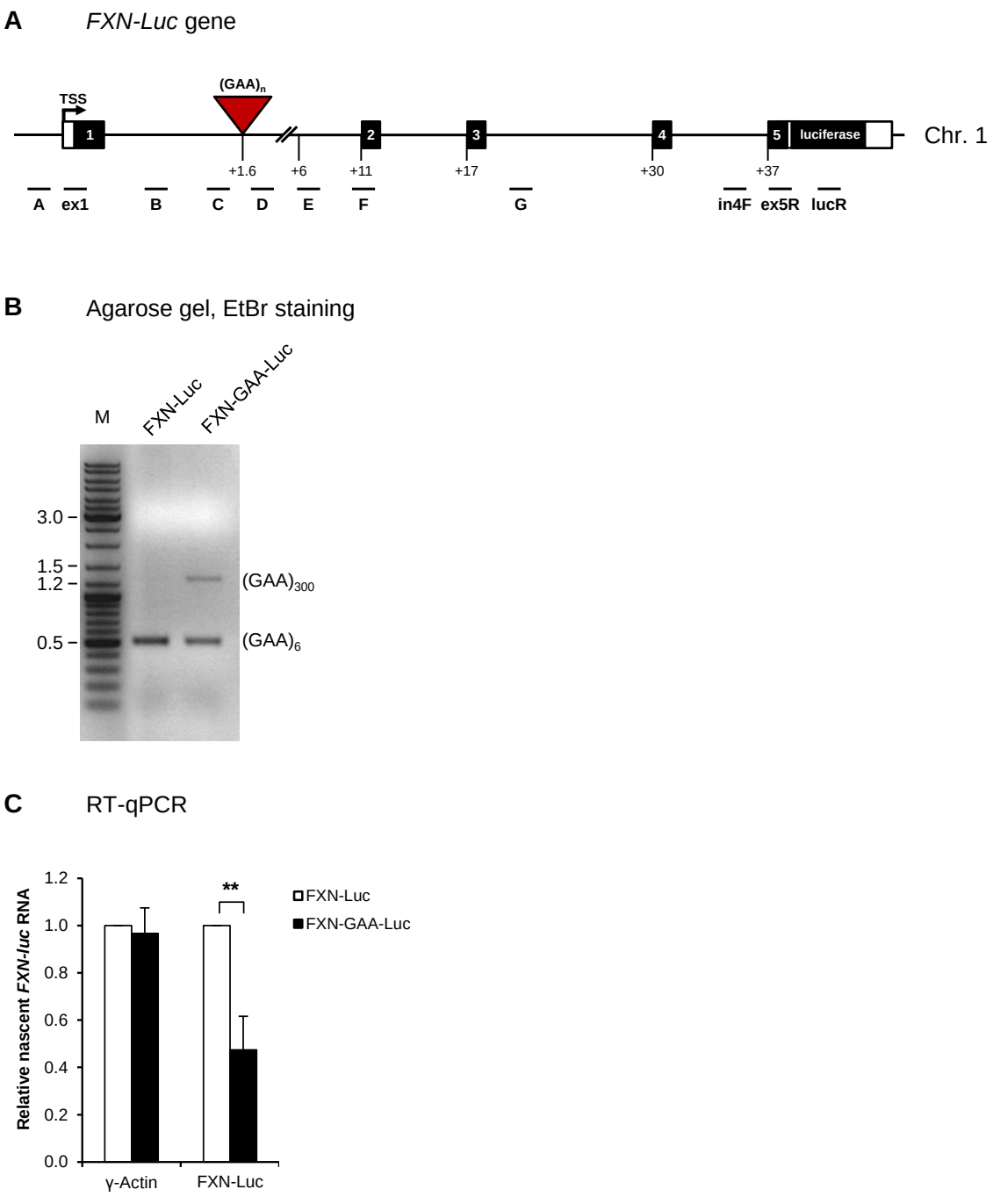


Figure 4.1. GAA-mediated repression in HEK293 cells

A. Diagram of the *FXN-Luc* gene, containing 6 (FXN-Luc cells) or 310 GAA repeats (FXN-GAA-Luc cells), integrated on the chromosome 1 of HEK293 cells. White boxes indicate 5' and 3'UTR, respectively. qPCR amplicons are shown below the diagram. TSS is the transcriptional start site. Numbers indicate the distances to TSS in kilobases.

B. Size of GAA expansion as determined by PCR analysis on genomic DNA from FXN-Luc and FXN-GAA-Luc cell lines.

C. *FXN* and *γ-Actin* pre-mRNA levels in FXN-Luc (white bars) and FXN-GAA-Luc (black bars) HEK293 cells, determined by RT-qPCR and normalized to 5S. LucR primer was used for the reverse transcription reaction. qPCR was carried using in4F and ex5R primers, shown in A.

patient-derived FRDA lymphoblastoid cells (Figures 3.6A, 4.2A). These results suggest that RNA/DNA hybrids are formed on transcribed expanded GAA repeats in the *FXN* gene, independently of their genomic location.

4.1.2 Heterochromatin formation is independent of chromosomal location and cell type

To investigate whether expanded GAA repeats in the integrated *FXN-Luc* transgene also exhibit epigenetic characteristics similar to the endogenous *FXN* gene, ChIP using antibody recognising H3K9me2 was performed. Similar to control lymphoblasts (Figure 3.4A), H3K9me2 peaked downstream of GAA repeats (amplicons D–F) in *FXN-Luc* cells (Figure 4.2B). Moreover, the levels of H3K9me2 were significantly increased at amplicons C, D and E in FXN-GAA-Luc compared to FXN-Luc cells (Figure 4.2B). The relatively small increase in H3K9me2 levels between unexpanded and expanded alleles may be attributed to the small expansion size in the transgene (~310 repeats). Furthermore, standard ChIP as employed here cannot distinguish between endogenous and transgenic *FXN* alleles, so the presence of three endogenous *FXN* alleles in *FXN-GAA-Luc* and *FXN-Luc* cells (Lufino *et al.*, 2013) may lead to an increased background compared to the homozygous expanded GAA alleles in FRDA lymphoblasts. Additionally, it cannot be excluded that the different wider chromosomal context around the transgene insertion site on chromosome 1 may have contributed to the composition of the *FXN-luc* and *FXN-GAA-Luc* epigenetic landscape. Despite these reservations, the above results indicate that irrespective of chromosomal background and cell type, expanded GAA repeats can trigger R-loops, heterochromatin formation and transcriptional repression of *FXN* gene.

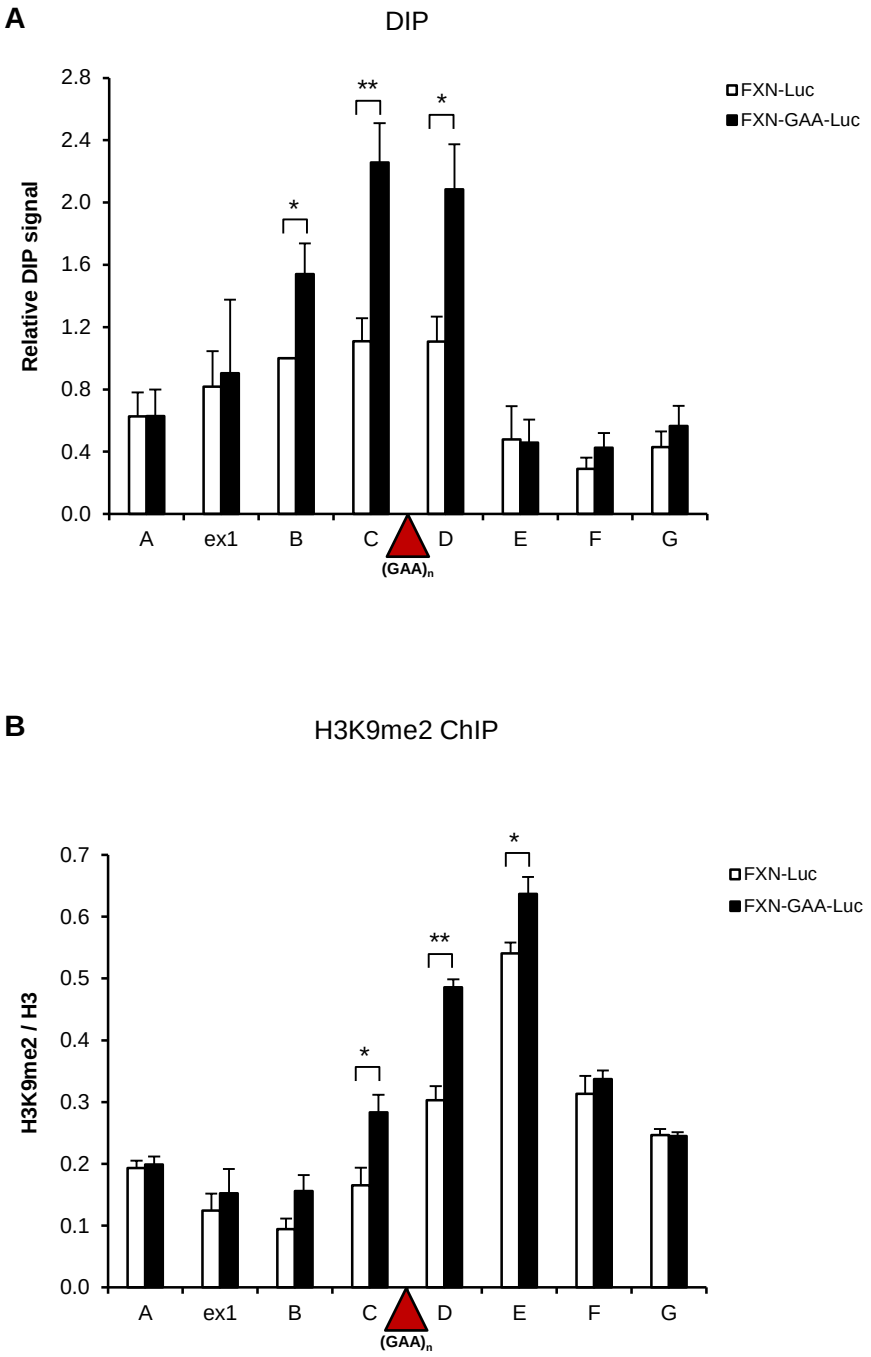


Figure 4.2. H3K9me2 and RNA/DNA hybrid enrichment at GAA expansions in HEK293 cells

A. DIP analysis on *FXN-Luc* gene in FXN-Luc (white bars) and FXN-GAA-Luc (black bars) HEK293 cells using RNA/DNA hybrid-specific S9.6 antibody.

B. H3K9me2 ChIP on *FXN-Luc* gene in FXN-Luc and FXN-GAA-Luc HEK293 cells. H3K9me2 levels were normalized to the total H3 levels.

4.2 RNase H1 resolves RNA/DNA hybrids at expanded GAA repeats *in vivo*

4.2.1 RNA/DNA hybrids in *FXN* gene are targets of endogenous RNase H1

Having established a correlation between enrichment of stable RNA/DNA hybrids at GAA repeats and transcriptional defects of *FXN* gene (see Chapter 3), an outstanding key question was whether RNA/DNA hybrid levels could be directly affected by proteins acting on them *in vivo*. While actinomycin D treatment showed that RNA/DNA hybrids at expanded repeats are turned over slower than RNA/DNA hybrids in the intron 1 of the highly expressed *γ-Actin* (Figure 3.9B), it was still unclear if any endogenous proteins were involved in resolving GAA repeat-associated RNA/DNA hybrids (Figure 3.9C).

Previous evidence in *E. coli* and *S. cerevisiae* suggested that endogenous RNase H enzymes may contribute to controlling cellular RNA/DNA hybrid levels, especially in cells lacking topoisomerase I (Top1) or RNA processing proteins such as Hpr1 (Chan *et al.*, 2014; El Hage *et al.*, 2010; El Hage *et al.*, 2014; Hong *et al.*, 1995; Huertas and Aguilera, 2003). Therefore, the effects of depleting endogenous RNase H1 on RNA/DNA hybrid levels were studied. Human lymphoblastoid cell lines are generally difficult to transfect using transient transfection procedures such as lipofection or electroporation, resulting in low transfection efficiency and high toxicity (data not shown). Therefore, the HEK293 FXN-Luc and FXN-GAA-Luc cells were used instead, in which endogenous RNase H1 was depleted by siRNA treatment.

Following treatment of HEK293 cells with siRNA targeting RNase H1, mRNA levels of RNase H1 were significantly reduced to ~20% of levels in cells treated with a non-targeting control siRNA (Figure 4.3A). DIP analysis in control- and RNase H1-depleted cells showed an increase in RNA/DNA hybrid signal around the GAA repeats both in FXN-Luc and FXN-GAA-Luc cells (Figure 4.3B), suggesting that endogenous RNase H1 can degrade R-loops formed over the *FXN* gene *in vivo*. This further confirms the specificity of the DIP signal at expanded GAA repeats within an independent cellular system and using different methodology (Figures 3.6B, 4.3B).

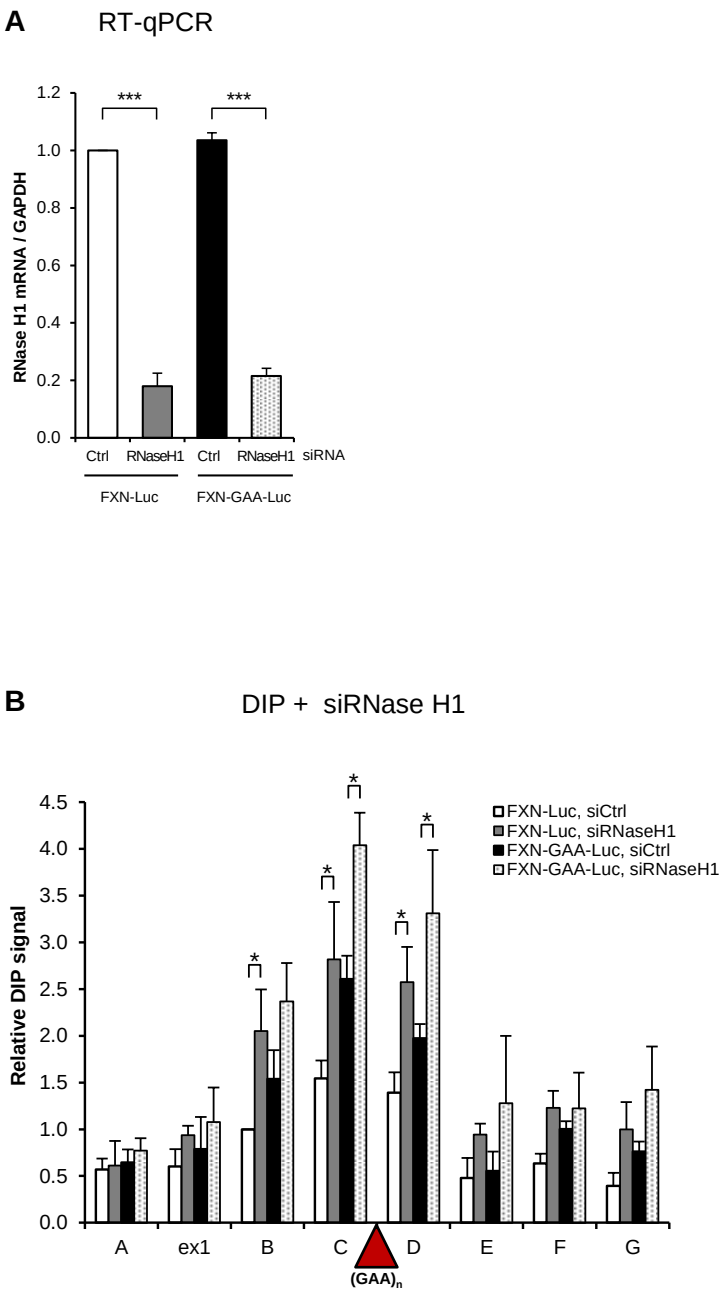


Figure 4.3. RNA/DNA hybrids at expanded GAA repeats are targets of endogenous RNase H1

A. RT-qPCR analysis of RNase H1 mRNA from FXN-Luc and FXN-GAA-Luc cells, treated with control and RNase H1 siRNAs. Values are normalized to *GAPDH* mRNA and are relative to FXN-Luc cells, treated with control siRNA.

B. DIP analysis on *FXN-Luc* gene in FXN-Luc and FXN-GAA-Luc HEK293 cells, treated with control and RNase H1 siRNAs.

4.2.2 RNA/DNA hybrids at GAA repeats can be resolved by RNase H1 overexpression

Numerous studies have used the introduction and overexpression of exogenous RNase H1 into *E. coli*, *S. cerevisiae* or vertebrate cells to investigate specificity and reversibility of RNA/DNA hybrid-associated processes (Drolet *et al.*, 1995; Huertas and Aguilera, 2003; Li and Manley, 2005; Skourti-Stathaki *et al.*, 2011; Stirling *et al.*, 2012; Tuduri *et al.*, 2009; Wahba *et al.*, 2011). Consequently, I tested if RNase H1 overexpression can resolve RNA/DNA hybrids formed at expanded GAA repeats. A commonly used RNase H1 overexpression construct for vertebrate cells lacks the mitochondrial localisation signal (MLS) and contains a C-terminal GFP tag (Cerritelli *et al.*, 2003; Skourti-Stathaki *et al.*, 2011; Sordet *et al.*, 2009). However, the presence of the large GFP tag close to the RNase H domain could potentially interfere with proper RNase H1 recruitment or activity *in vivo*. Therefore, a C-terminally Flag-tagged RNase H1 overexpression plasmid, lacking MLS, was created (Figure 4.4A) (see Chapter 2.1.4).

Transfection of HEK293 FXN-Luc and FXN-GAA-Luc cells with the RNase H1-Flag construct led to a ~200-fold increase in RNase H1 mRNA levels compared to Flag-transfected cells (Figure 4.4B). Protein expression of RNase H1-Flag was confirmed by western blotting analysis (Figure 4.4C). RNase H1-Flag has a slightly lower molecular weight in SDS-PAGE than endogenous RNase H1 due to its lack of the 27-amino acid MLS (Chapter 2.1.4, Figure 4.4C). Following RNase H1-Flag overexpression, RNA/DNA hybrid signals were significantly reduced over the GAA expansion in FXN-GAA-Luc cells (Figure 4.4D). In contrast, RNase H1-Flag overexpression had no significant effect on the DIP signals on the *FXN-Luc* allele, potentially because RNA/DNA hybrids at non-expanded GAA repeats are already controlled by endogenous RNase H1 or RNase H2 (Figure 4.4D). In line with these observations, RNase H1-Flag overexpression resulted in upregulation of *FXN* transcription from the expanded allele (Figure 4.4E). This suggests that RNA/DNA hybrids formed over expanded repeats *in vivo* can be resolved by overexpression of RNase H1 and that removal of RNA/DNA hybrids can lead to an increase in *FXN* gene expression.

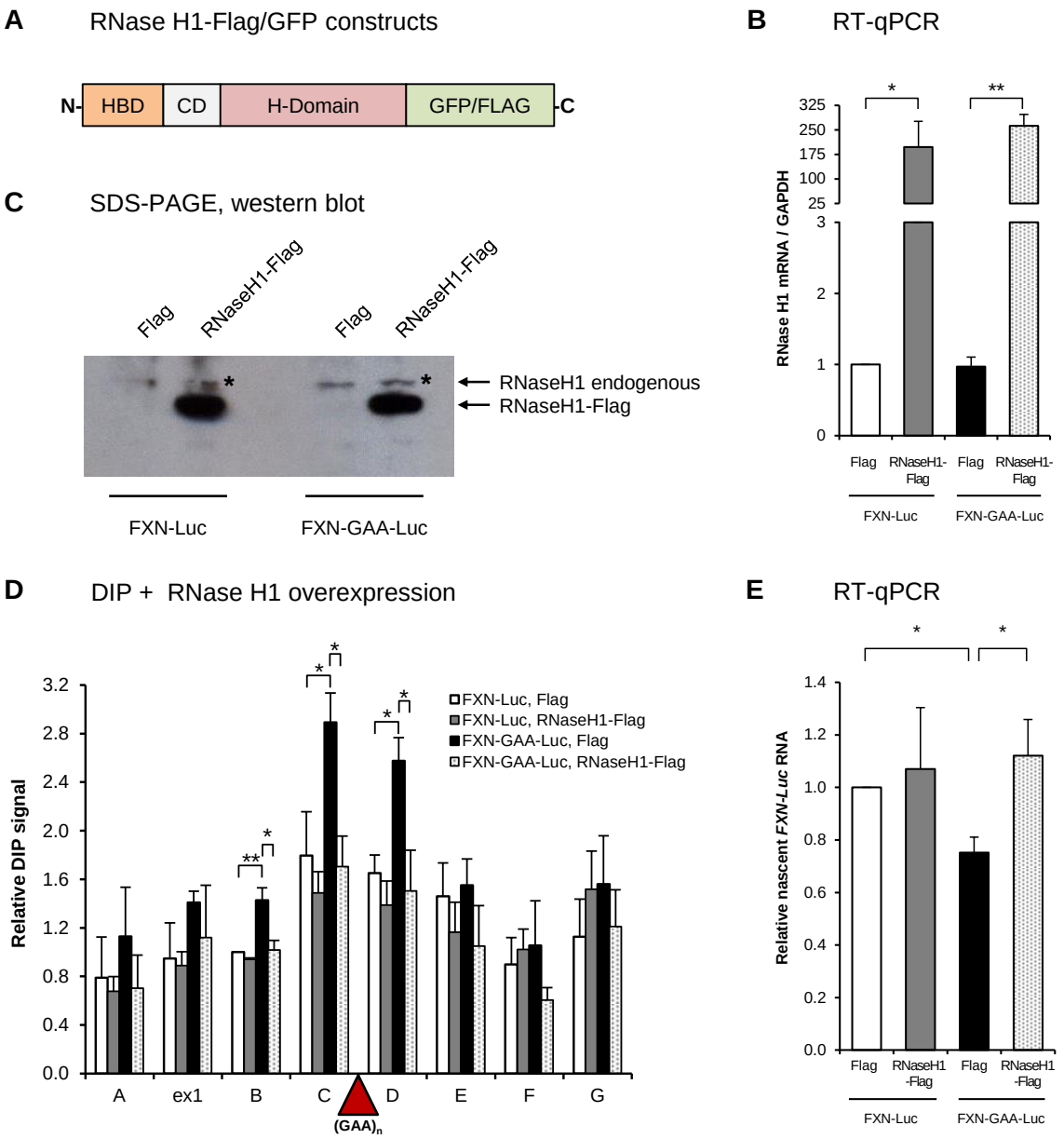


Figure 4.4. RNA/DNA hybrids at GAA repeats are sensitive to RNase H1 overexpression

A. Diagram depicting the RNase H1-GFP/FLAG constructs. HBD denotes the hybrid-binding domain, CD the connection domain, H-Domain the RNase H domain.

B. RT-qPCR analysis of *RNase H1* mRNA in FXN-Luc and FXN-GAA-Luc HEK293 cells, overexpressing Flag and RNase H1-Flag plasmids. Values are normalised to the level of *GAPDH* mRNA and are relative to FXN-Luc cells, overexpressing Flag.

C. Western blot analysis of 50 μg of protein extracts obtained from FXN-Luc and FXN-GAA-Luc cells transfected with Flag and RNase H1-Flag expression plasmids. Western blot was probed with anti-RNase H1 antibody. * denotes endogenous RNase H1 protein.

D. DIP analysis on *FXN-Luc* gene in FXN-Luc and FXN-GAA-Luc HEK293 cells transfected with Flag or RNase H1-Flag expression plasmids.

E. RT-qPCR analysis of *FXN-Luc* pre-mRNA in FXN-Luc and FXN-GAA-Luc HEK293 cells, overexpressing Flag and RNase H1-Flag plasmids, normalized to *γ-Actin* pre-mRNA. Values are relative to FXN-Luc cells, overexpressing Flag plasmid.

4.3 *FXN* gene repression in FRDA is mediated by RNA/DNA hybrids

4.3.1 H3K9me2 is dispensable for transcriptional repression of *FXN* gene in FRDA

The results from the FRDA cellular model system suggested that accumulation of RNA/DNA hybrids at expanded GAA repeats directly impairs *FXN* gene expression. However, the mechanistic relationship between RNA/DNA hybrids and repressive chromatin marks such as H3K9me2 remained to be determined. The present and previous studies have demonstrated that expanded *FXN* GAA repeats are associated with increased levels of heterochromatin marks (Figure 3.4A, 4.2B) (Chan *et al.*, 2013; Greene *et al.*, 2007; Kim *et al.*, 2011; Kumari *et al.*, 2011). So far, only decreased levels of histone acetylation, in particular of histone H3K14, H4K5 and H4K12 were directly linked to *FXN* gene repression in FRDA. This is supported by the fact that inhibition of histone deacetylase enzymes (HDACs) led to increased *FXN* expression from expanded alleles (Herman *et al.*, 2006; Soragni *et al.*, 2008). The strong correlation of RNA/DNA hybrids and H3K9me2 enrichment in the present study suggested a novel functional interplay and a possible role of H3K9me2 in *FXN* gene regulation (Figure 3.10).

To investigate the relationship between RNA/DNA hybrids and the histone modification H3K9me2 formed on the expanded *FXN* allele, the G9a histone methyltransferase inhibitor BIX-01294 was used (Figure 4.5A). BIX-01294 has previously been shown to reduce the level of H3K9me2 *in vivo* and over the *FXN* gene (Kubicek *et al.*, 2007; Punga and Buhler, 2010). Following treatment of control and FRDA lymphoblastoid cells with BIX-01294 for 72 hours, levels of H3K9me2 were significantly reduced (Figure 4.5B), similar to a previous report (Punga and Buhler, 2010). Significantly, following BIX-01294 treatment the level of RNA/DNA hybrids over the expanded repeat region remained unchanged (Figure 4.5C). In line with this, the reduction in H3K9me2 levels in FRDA cells had no effect on the expression of *FXN* pre-mRNA (Figure 4.5D). These data suggest that the H3K9me2 chromatin modification is not directly responsible for the *FXN* transcriptional repression and is likely to be a consequence of the reduced transcription or caused by R-loop formation.

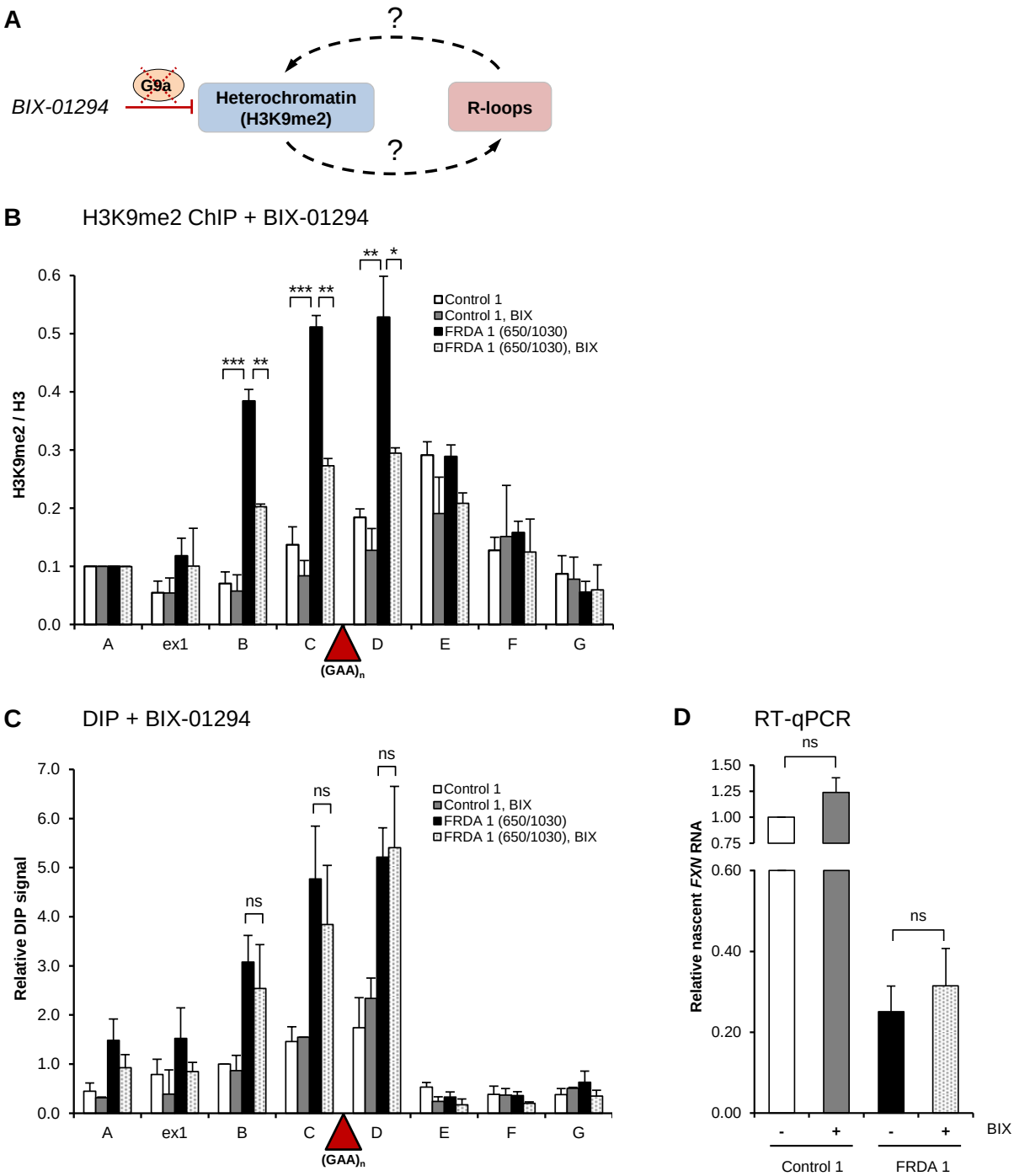


Figure 4.5. RNA/DNA hybrids and *FXN* transcription are not affected by changes in H3K9me2

A. BIX-01294, a specific inhibitor of H3K9me2 methyltransferase G9a, is used to investigate the interplay between RNA/DNA hybrids and H3K9me2 levels.

B. H3K9me2 ChIP in control 1 (GM15851) and FRDA 1 (GM15850) cells, treated with 4 µM BIX-01294 for 72 h. H3K9me2 levels were normalized to the total H3 levels and relative to amplicon FXN A, not affected by the treatment.

C. DIP analysis in control 1 and FRDA 1 cells, treated with 4 µM BIX-01294 for 72 h.

D. RT-qPCR analysis of *FXN* pre-mRNA in control 1 and FRDA 1 cells, treated with 4 µM BIX-01294 for 72 h. Values are relative to untreated control cells, and normalized to *γ-Actin* nascent RNA.

4.3.2 RNA/DNA hybrids trigger H3K9me2 enrichment and repression of *FXN* gene *in vivo*

To further investigate the interplay between RNA/DNA hybrids and H3K9me2 levels in control and FRDA lymphoblasts, it was possible to take advantage of a highly specific inhibitor for Top1, camptothecin (CPT) (Capranico *et al.*, 2010). Top1 relieves topological stress in DNA by counteracting negative supercoiling emerging behind travelling RNA polymerases, which is thought to favour R-loop formation (Masse and Drolet, 1999; Roy *et al.*, 2010; Wu *et al.*, 1988). Top1 has therefore been suggested to belong to the family of cellular proteins naturally keeping R-loop formation in check (Aguilera and Garcia-Muse, 2012; Santos-Pereira and Aguilera, 2015). In line with this, cells lacking Top1 or treated with CPT exhibit increased R-loop formation (El Hage *et al.*, 2010; El Hage *et al.*, 2014; Marinello *et al.*, 2013; Powell *et al.*, 2013; Sordet *et al.*, 2009; Tuduri *et al.*, 2009). Therefore, control or FRDA patient lymphoblastoid cells were treated with 10 μM CPT for 6 hours, followed by DIP analysis (Figure 4.6A). RNA/DNA hybrid formation over the expanded GAA repeat region in FRDA cells (amplicons B–D) was increased by CPT treatment, while RNA/DNA hybrid levels in *FXN* regions E-G remained unchanged (Figure 4.6B). The DIP showed no effect on RNA/DNA hybrid levels in control cells, demonstrating the specificity of CPT treatment to expanded repeats (Figure 4.6B). Interestingly, the increase in the level of RNA/DNA hybrids coincided with a significant increase in the amount of the H3K9me2 histone modification over *FXN* regions B-D surrounding the expansion in FRDA cells (Figure 4.6C). Similarly to the DIP results, H3K9me2 levels were not affected by CPT in control cells. Furthermore, RT-qPCR analysis demonstrated that induction of RNA/DNA hybrids by CPT coincided with a further repression of *FXN* gene transcription while *FXN* RNA levels were not affected in control cells (Figure 4.6D). The same effects of CPT on induction of RNA/DNA hybrids and H3K9me2 were observed in two independent control and FRDA cell lines (Figures 4.7A–D). These results link RNA/DNA hybrid induction to transcriptional repression of *FXN* gene in FRDA cells.

CPT inhibits Top1 by inducing the formation of stable Top1-DNA cleavage complexes (Top1cc), leading to degradation of Top1 by proteolysis (Capranico *et al.*, 2010). An additional control therefore consisted in investigating whether the induction of Top1cc complexes by CPT or

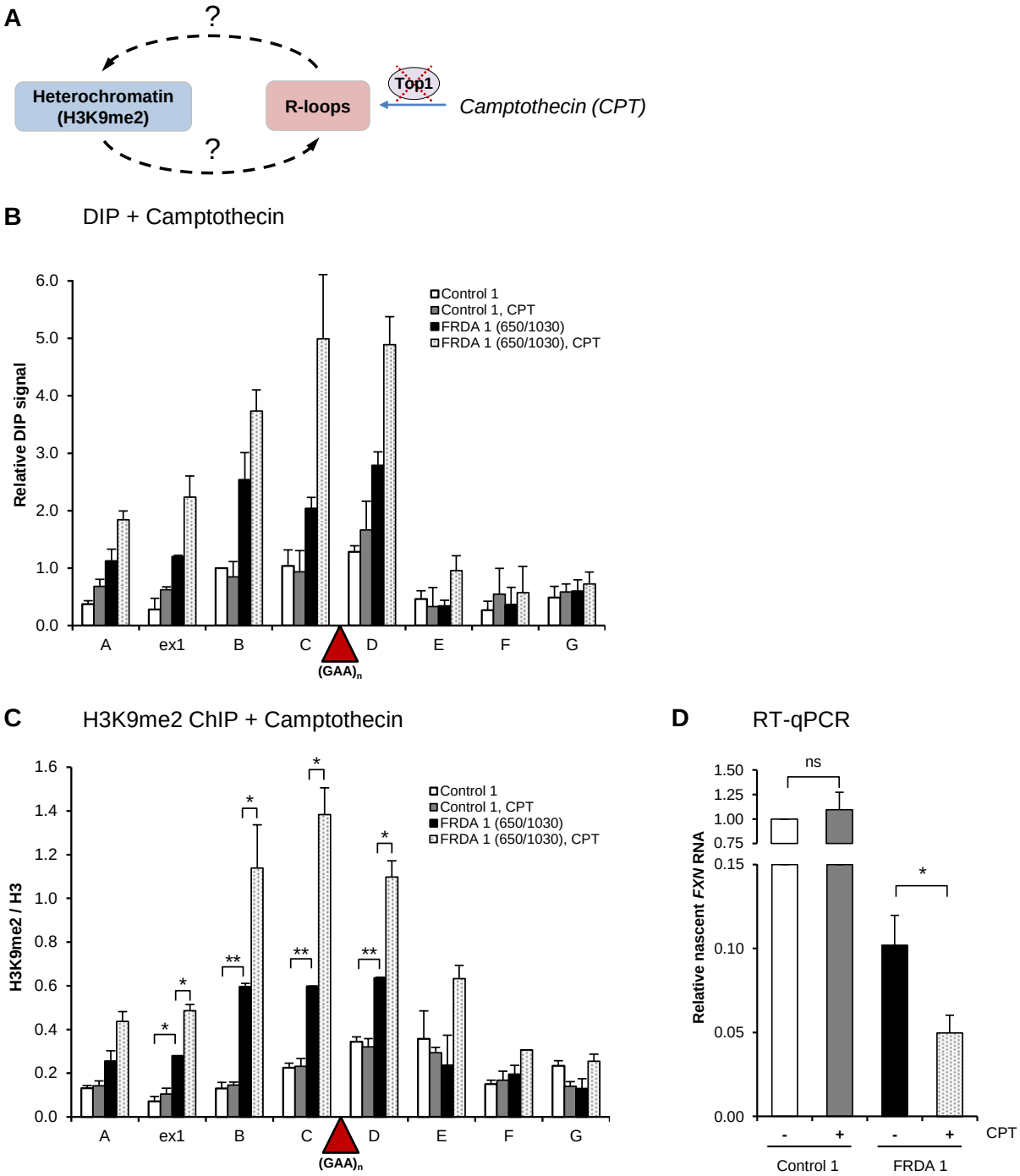


Figure 4.6. R-loops promote H3K9me2 and transcriptional repression of *FXN* gene

A. Camptothecin (CPT), a specific inhibitor of Top1, is used to investigate the interplay between R-loops and H3K9me2 levels.

B. DIP analysis on *FXN* gene in control 1 (GM15851) and FRDA 1 (GM15850) cells, treated with 10 μM CPT for 6 hours.

C. H3K9me2 ChIP on *FXN* gene in control 1 and FRDA 1 cells, treated with 10 μM CPT for 6 hours. H3K9me2 levels were normalized to the total H3 levels.

D. RT-qPCR analysis of *FXN* nascent RNA in control 1 and FRDA 1 cells, treated with 10 μM CPT for 6 hours. Values are relative to untreated control cells and normalized to *γ-Actin* nascent RNA.

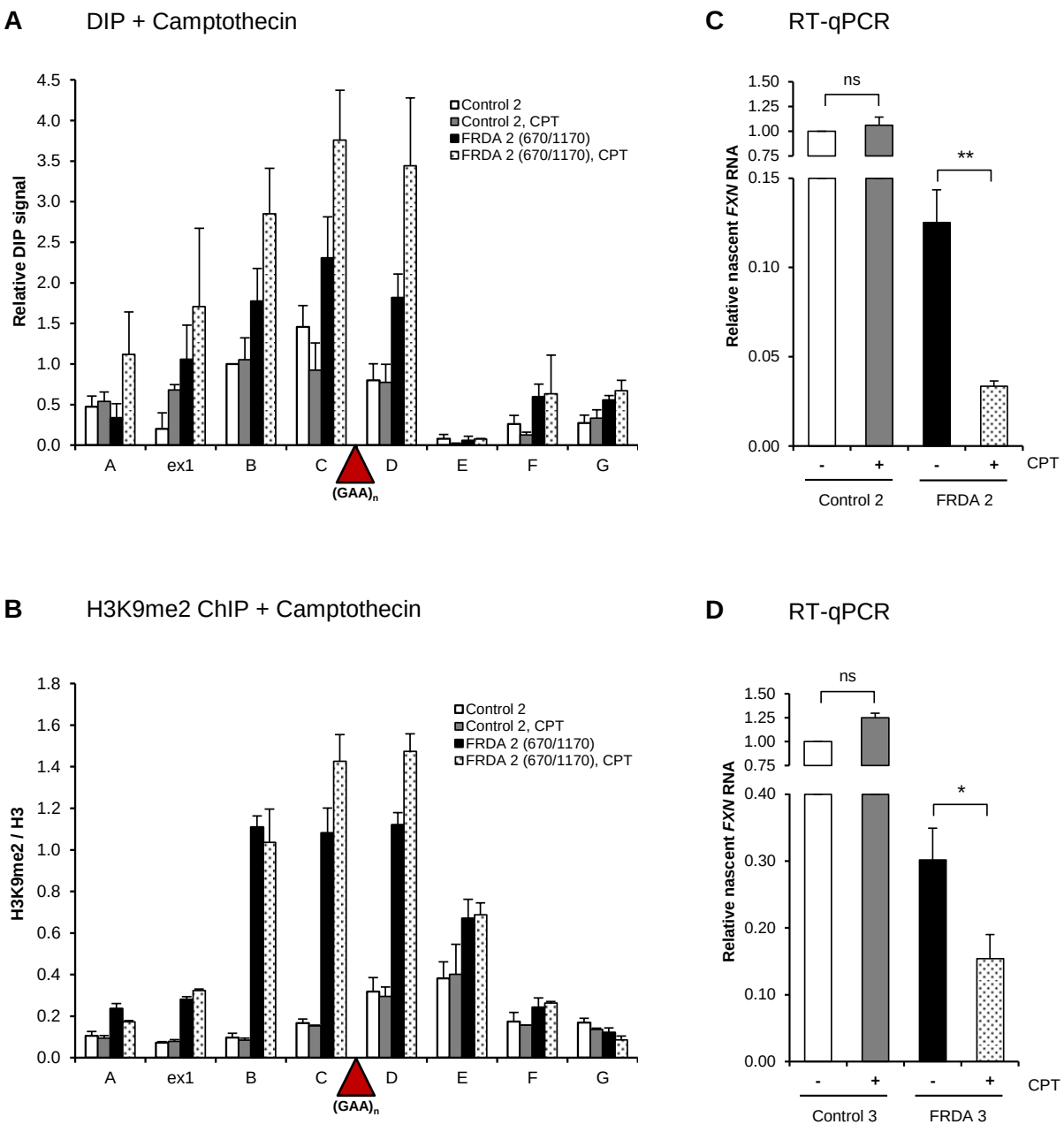


Figure 4.7. Camptothecin experiments in additional cell lines

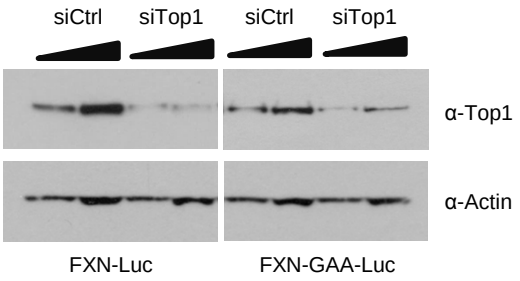
A. DIP on *FXN* gene in control 2 (GM14926) and FRDA 2 (GM16243) cells, treated with 10 μM camptothecin for 6 hours.

B. H3K9me2 ChIP on *FXN* gene in control 2 (GM14926) and FRDA 2 (GM16243) cells, treated with 10 μM camptothecin for 6 hours. H3K9me2 levels were normalized to the total H3 levels.

C. RT-qPCR analysis of *FXN* nascent RNA in control 2 (GM14926) and FRDA 2 (GM16243) cells, treated with 10 μM camptothecin for 6 hours. Values are relative to untreated control cells and normalized to *γ-Actin* nascent RNA.

D. RT-qPCR analysis of *FXN* nascent RNA in control 3 (GM06895) and FRDA 3 (GM16209) cells, treated with 10 μM camptothecin for 6 hours. Values are relative to untreated control cells and normalized to *γ-Actin* nascent RNA.

A SDS-PAGE, western blot



B DIP + Topoisomerase I knock-down

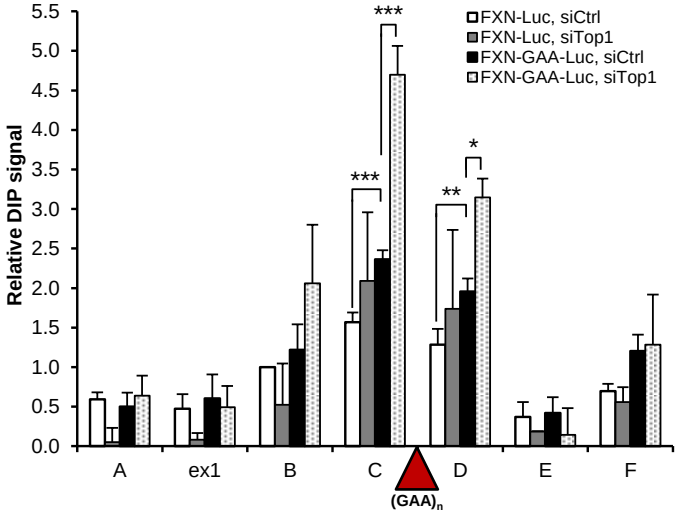


Figure 4.8. R-loops are induced at expanded GAA repeats by Top1 knockdown

A. Western blot analysis of 20 and 40 µg of protein extracts obtained from FXN-Luc and FXN-GAA-Luc cells, treated with control and Top1 siRNAs. Western blot was probed with anti-Top1 and anti-Actin antibody.

B. DIP analysis on *FXN-Luc* gene in FXN-Luc and FXN-GAA-Luc HEK293 cells, treated with control and topoisomerase 1 (Top1) siRNAs.

the loss of Top1 function was the cause of RNA/DNA hybrid formation and increased H3K9me2 levels. For this, Top1 was depleted from FXN-Luc and FXN-GAA-Luc cells using siRNA (Figure 4.8A). Top1 knockdown led to a significant increase in RNA/DNA hybrid levels at expanded GAA repeats, but not in control cells (Figure 4.8B), recapitulating the results from CPT experiments. This suggests that the ability of CPT to increase RNA/DNA hybrid levels was not due to CPT-induced covalent links between Top1 and DNA.

4.3.3 RNAi factors do not correlate with *FXN* repression in FRDA

Different mechanisms have been identified that can trigger heterochromatin formation. Although the exact process depends on genomic location, cell state and organism, RNA components such as noncoding RNAs (ncRNA) and small double-stranded RNAs play pivotal roles (Holoch and Moazed, 2015; Mozzetta *et al.*, 2015). In particular, RNA plays a key role in the establishment of heterochromatin at repetitive elements in mammals, such as pericentromeric and telomeric DNA, and the production of dsRNA have been implicated in this process in other organisms (Bierhoff *et al.*, 2014; Holoch and Moazed, 2015). Convergent transcription producing complementary sense and antisense RNA can lead to the formation of dsRNA, which then recruits components of the RNA interference (RNAi) machinery, including the proteins Dicer, Ago1 and Ago2 (Holoch and Moazed, 2015). Recently, two studies reported that R-loops can regulate the production of dsRNA, leading to the recruitment of RNAi factors and formation of heterochromatin at centromeres in *S. pombe* and at mammalian gene terminators (Nakama *et al.*, 2012; Skourti-Stathaki *et al.*, 2014). It was thus reasoned that RNAi could be involved in mediating *FXN* repression in FRDA, potentially linking R-loop enrichment and heterochromatin formation.

Therefore, I investigated whether components of the RNAi machinery can be recruited at the *FXN* gene and whether there is a correlation with H3K9me2 levels. Using strand-specific reverse transcription followed by qPCR, antisense *FXN* pre-mRNA levels were measured. Interestingly, levels of antisense *FXN* pre-mRNA were highest at the transcription start site and reduced in FRDA cells compared to control cells (Figure 4.9A). Therefore, dsRNA is more likely to form at the TSS2, where both sense and antisense *FXN* RNA levels are highest, than in the *FXN* gene body (Figures

3.2, 4.9A). The RNAi proteins Ago1 and Ago2 were enriched at the promoter of *FXN* gene in control cells but showed reduced levels in FRDA cells (Figure 4.9B, C). Thus, there is a lack of correlation in recruitment of the investigated RNAi components and the H3K9me2 and RNA/DNA hybrid profiles in the *FXN* gene in FRDA cells. This, together with the fact that antisense transcription at transcriptional start sites is a widespread feature of many human genes (Pelechano and Steinmetz, 2013) suggests that RNAi may be involved in regulating *FXN* gene transcription in healthy individuals. However, it is unlikely to be required to mediate heterochromatin formation and *FXN* gene repression in FRDA.

4.3.4 G9a is recruited to expanded GAA repeats in FRDA

To understand the molecular mechanism of heterochromatin formation at expanded GAA repeats, the binding of the histone methyltransferase G9a to *FXN* gene was investigated. G9a is a major source of H3K9me2 on many genes in mammalian cells and the inhibitor BIX-01294 targets G9a and the close homologue GLP (Kubicek *et al.*, 2007; Mozzetta *et al.*, 2015). Interestingly, as demonstrated by ChIP, G9a is enriched close to the expanded GAA repeat region of the *FXN* gene in FRDA cells (Figure 4.10A). This suggests that R-loops may recruit G9a to the expanded repeats, thereby promoting the formation of H3K9me2 marks. Since no correlation of RNA/DNA hybrids was observed with the recruitment of RNAi factors (Chapter 4.3.3.), this result points to an RNAi-independent mechanism of heterochromatin formation, mediated by RNA/DNA hybrids. In line with this mechanism, G9a was shown to be targeted to chromatin by long noncoding RNAs (lncRNAs), such as *Kcnq1ot1* and *DHRS4AS1* (Mozzetta et al., 2015).

4.4 Summary

The data presented in this Chapter demonstrate that stable RNA/DNA hybrids associated with expanded GAA repeats in the first intron of the *FXN* gene promote heterochromatin formation and lead to its transcriptional repression in FRDA cells (Figure 4.10B).

In particular, GAA-associated RNA/DNA hybrids form independently of cell type and chromosomal location (Figures 4.1–4.2) and are targets of endogenous RNase H1 (Figure 4.3). Overexpression of Flag-tagged RNase H1 led to a decrease in RNA/DNA hybrids at expanded GAA

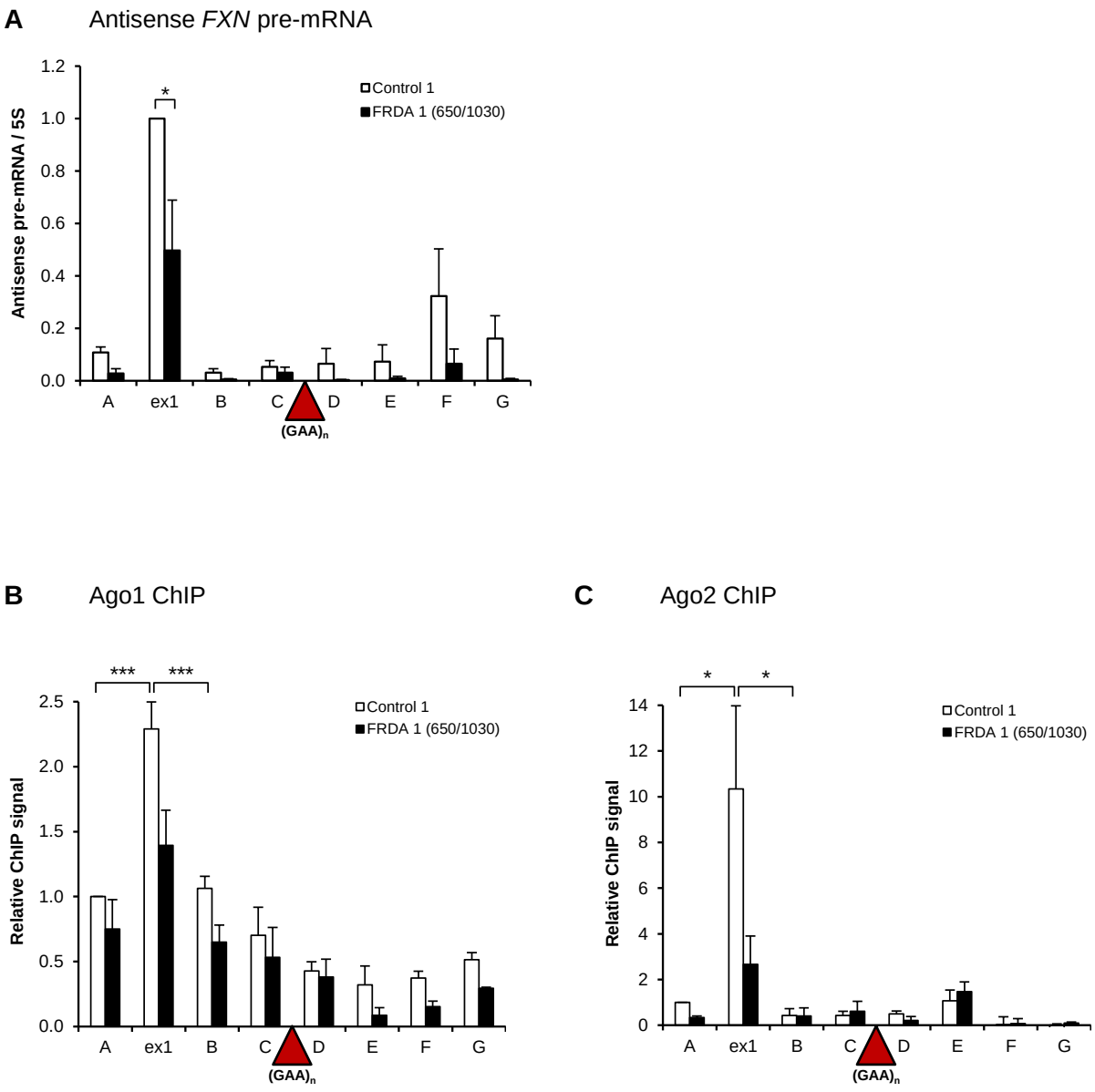


Figure 4.9. RNAi factors are not enriched at expanded GAA repeats of *FXN* gene

A. RT-qPCR analysis of antisense *FXN* pre-mRNA in control 1 (GM15851) and FRDA 1 (GM15850) cells. The values are normalized to to 5S rRNA and relative to ex1 RNA in control cells. Reverse transcription was carried out using specific primer for each of the indicated amplicons.

B. Ago1 ChIP in control 1 and FRDA 1 cells. Values are relative to FXN A amplicon in control cells.

C. Ago2 ChIP in control 1 and FRDA 1 cells. Values are relative to FXN A amplicon in control cells.

repeats and restored *FXN* expression to normal levels, suggesting that RNA/DNA hybrids trigger *FXN* transcriptional repression in FRDA (Figure 4.4). The investigation of the mechanistic link between RNA/DNA hybrids and heterochromatin revealed that reducing levels of the repressive histone modification H3K9me2 did not affect RNA/DNA hybrids (Figure 4.5A–C). Despite its enrichment at expanded GAA repeats, reducing H3K9me2 levels surprisingly did not have a detectable impact on *FXN* RNA levels in lymphoblastoid cells (Figure 4.5D).

In contrast, induction of RNA/DNA hybrids led to increased H3K9me2 levels and further transcriptional repression of the *FXN* gene in FRDA cells, but not in control cells (Figures 4.6–4.8). The enrichment of *FXN* antisense pre-mRNA, Ago1 and Ago2 at the *FXN* promoter and their reduced levels in FRDA cells suggested that heterochromatin is established in an RNAi-independent manner in FRDA (Figure 4.9). The H3K9me2 histone methyltransferase G9a was found to be enriched over expanded GAA repeats in FRDA cells, suggesting that RNA/DNA hybrids may directly recruit repressive epigenetic modifying enzymes including G9a (Figure 4.10A). Although further investigations are required to unequivocally determine the state of the displaced single-stranded non-template strand associated with RNA/DNA hybrids, the data of Chapter 3 and 4 suggest that R-loops form at expanded GAA and CGG repeats of *FXN* and *FMR1* genes, respectively. The identification of novel R-loop interactors will thus help to shed light on the pathomechanism of FRDA (see Chapter 5).

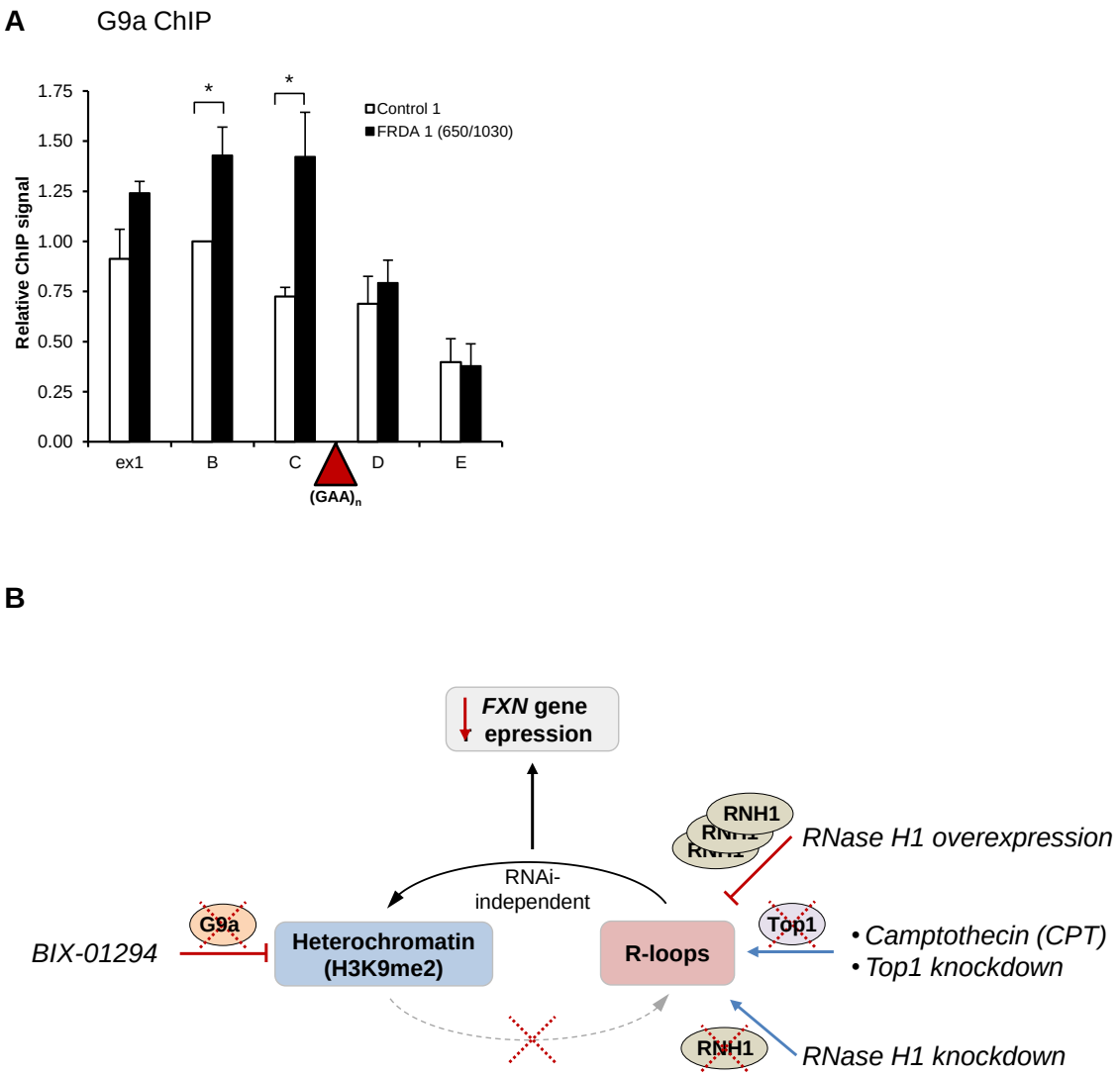


Figure 4.10. H3K9 methyltransferase G9a is enriched at expanded GAA repeats in FRDA

A. G9a ChIP on *FXN* gene in control 1 (GM15851) and FRDA 1 (GM15850) cells. G9a levels are normalised relative to amplicon B in control cells.

B. Mechanism of R-loop-mediated transcriptional repression of *FXN* gene in FRDA. Approaches to investigate the functional interplay between R-loop formation, H3K9me2 levels and *FXN* repression in FRDA are depicted in italics. BIX-01294 and camptothecin experiments were carried out in control and FRDA lymphoblasts, RNase H1 overexpression, Top1 and RNase H1 knockdowns were performed in HEK293 FXN-Luc and FXN-GAA-Luc cells. Reducing H3K9me2 levels had no effect on R-loop levels or *FXN* RNA. Induction of R-loops led to *FXN* repression, while their cleavage by RNase H1 led to *FXN* transcriptional de-repression.

4.5 Discussion of Chapters 3 and 4

The data presented in Chapter 3 and 4 show a direct connection between increased RNA/DNA hybrid (R-loop) formation due to an expanded trinucleotide repeat and the resultant transcriptional repression of the repeat-containing gene *in vivo* (Figure 4.11). Using FRDA as a disease model, this work provides new insight into the underlying molecular pathology of human trinucleotide expansion diseases.

4.5.1 Transcriptional defects of *FXN* gene in FRDA

FXN transcription initiation defect in FRDA

Similar to results presented here (Figures 3.1D, 3.2, 3.3B), several other groups have demonstrated a defect in transcription initiation of *FXN* gene in FRDA cells, using pre-mRNA, histone modifications and CTD-Ser5-phosphorylated RNA Pol II levels as readouts (Kim *et al.*, 2011; Kumari *et al.*, 2011). The initiation defect was further supported more recently by studies demonstrating reduced promoter chromatin accessibility in FRDA, in line with reduced levels of initiating RNA Pol II (Chan *et al.*, 2013; Chutake *et al.*, 2014).

In contrast to the observations mentioned above, two *in vitro* studies failed to detect initiation defects, but detected elongation defects due to expanded GAA repeats (Grabczyk and Usdin, 2000b; Ohshima *et al.*, 1998). This may suggest that the various downstream effects of expanded repeats on different stages of transcription may not necessarily be coupled and can occur independently of each other, potentially being influenced by the context of the GAA repeats and the strength of the host gene promoter. Potential mechanisms linking expanded GAA repeats with defects in transcription initiation are discussed in Chapter 4.5.5.

Deficient FXN transcription elongation in FRDA

The initial characterisation of *FXN* gene transcriptional regulation in this work by RNA Pol II ChIP and RT-qPCR suggested transcription initiation and elongation defects triggered by expanded repeats (Figures 3.1D, 3.2). In line with this, previous reports have demonstrated transcription elongation defects due to expanded GAA repeats *in vitro*, using heterologous reporter constructs

(Grabczyk and Usdin, 2000b; Ohshima *et al.*, 1998). More recent reports have provided more evidence of reduced transcription elongation downstream of expanded GAA repeats in patient cells by analysing pre-mRNA levels and H3K36me3, a post-translational histone modification associated with actively elongating RNA Pol II (Kim *et al.*, 2011; Kumari *et al.*, 2011; Punga and Buhler, 2010). The present work confirms these observations and expands them by directly investigating the levels of active RNA Pol II by Br-UTP nuclear run-on, instead of using indirect read-outs such as histone modifications (Figure 3.3).

Interestingly, several independent groups did not observe an elongation defect due to expanded GAA repeats (Baralle *et al.*, 2008; Rindler and Bidichandani, 2011; Shishkin *et al.*, 2009). In light of the *in vivo* results from this work and others, this disparity may be attributed to the different model systems employed, such as heterologous expression systems, *FXN* minigenes, GAA insertions in yeast introns and most importantly the use of very strong non-*FXN* promoters such as the cytomegalovirus (CMV) promoter (Baralle *et al.*, 2008; Rindler and Bidichandani, 2011; Shishkin *et al.*, 2009). Although more detailed side-by-side experiments will be needed, an interesting implication of these observations is that the effect of expanded GAA repeats on transcription may not be universal and could potentially be overcome by modifying the strength of the endogenous *FXN* promoter.

4.5.2 R-loop formation at expanded GAA repeats *in vivo*

Detection of R-loops at expanded GAA repeats

The present study suggests that R-loops are formed *in vivo* over endogenous expanded GAA and CGG repeats, associated with FRDA and FXS disorders, through the formation of RNA/DNA hybrids and displacement of the single-stranded non-template strand (Figures 3.6, 3.11). These R-loops depend on transcription and correlate with expansion size, while unexpanded repeats show much reduced or background R-loop levels (Figures 3.8–3.12).

Different requirements of R-loop formation *in vivo* have been suggested. Genome-wide analysis of R-loops in human cells showed a particular enrichment of R-loops at GC-skewed promoters, i.e. the G-rich RNA hybridises with the C-rich template strand (Ginno *et al.*, 2012).

Similarly, *in vitro* and *in vivo* work on the immunoglobulin class switch regions has demonstrated that initiation of R-loops requires clusters of GGGG tetranucleotides, while downstream of the initiation site G-rich sequences in the non-template strands are sufficient to sustain R-loops (Roy and Lieber, 2009; Roy *et al.*, 2008; Zhang *et al.*, 2014a; Zhang *et al.*, 2014b). Data for repeated sequences support the requirement for G-richness in the R-loop-forming RNA since RNA/DNA hybrids are formed by $(GAA)_n$ RNA but not antisense $(CUU)_n$ RNA transcribed *in vitro* (Grabczyk *et al.*, 2007; Reddy *et al.*, 2011). The exact mapping of the length and position of the GAA-associated R-loop in the *FXN* gene will require higher resolution techniques than the currently employed DIP. For this, it will be interesting to employ the newly developed DRIP-seq approaches which are based on sequencing of the RNA in RNA/DNA hybrids, thus offering higher resolution and strand-specific analysis (Chen *et al.*, 2015a; Nadel *et al.*, 2015).

Based on the presented data, it is possible that R-loops in the *FXN* first intron are initiated upstream of the GAA repeats due to the presence of the 11 GGGG tetranucleotides (Figure 3.6). After initial formation, R-loops could then be extended through the GAA repeats, which present a sufficiently G-rich RNA. This would explain the detection of RNA/DNA hybrids at GAA repeats in control *FXN* alleles, while they may be further stabilised by the expanded GAA repeat itself, thereby leading to higher RNA/DNA hybrid and potentially R-loop levels.

Formation of GAA-associated R-loops independently of genomic context

Multiple lines of evidence from this work suggest that RNA/DNA hybrids form at expanded GAA repeats in lymphoblastoid cells from FRDA patients in their natural genomic context (Figures 3.6–3.10). Interestingly, a complete *FXN* gene containing ~300 GAA repeats integrated into a different chromosomal background in HEK293 cells also exhibits enhanced R-loop formation (Figure 4.2). Thus, R-loops are not significantly affected by potential epigenetic effects at the new insertion site such as position effect variegation (Sadelain *et al.*, 2012). Expanded and transcribed GAA repeats thus may act as autonomous triggers of RNA/DNA hybrid and consequently R-loop formation and additional downstream effects such as transcriptional repression and heterochromatinisation.

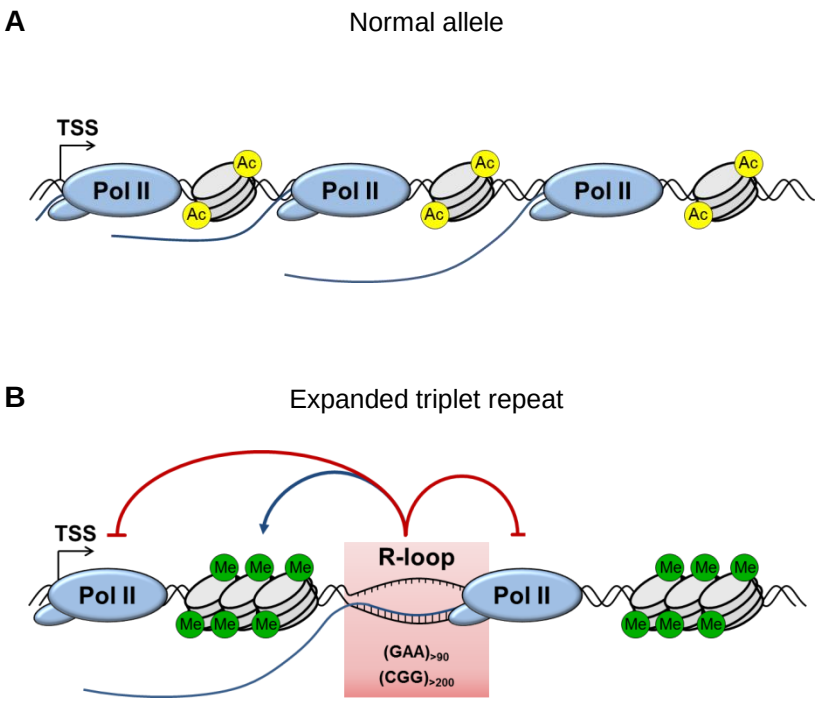


Figure 4.11. Model for the role of R-loops in mediating *FXN* and *FMR1* gene silencing

A. Background R-loop levels on unexpanded type alleles allows efficient transcriptional elongation and gene expression, associated with active chromatin marks (Ac, acetylation).

B. Transcribed $(GAA)_n$ and $(CGG)_n$ expanded repeats form R-loops resulting in decreased initiation and elongation of RNA Pol II. This leads to downregulation of *FXN* and *FMR1* expression, associated with formation of repressive DNA and chromatin marks (Me, methylation).

In support of this, isolated expanded repeats transcribed using non-mammalian RNA polymerases have been reported to form R-loops *in vitro* (Grabczyk *et al.*, 2007; Reddy *et al.*, 2011). Future experiments will need to determine if persistent RNA/DNA hybrid and R-loop formation at GAA repeats is also found in additional cell types, especially the pathologically relevant terminally differentiated neurons and pluripotent stem cells (see Outlook in Chapter 6). In addition, insertion of expanded GAA repeats into different genomic contexts may prove useful as a molecular tool for studying R-loop formation and their consequences for normal gene regulation.

4.5.3 Stability of R-loops formed at expanded GAA repeats

RNA stability and turnover

Interestingly, RNA/DNA hybrids over expanded GAA repeats are very stable in human cells, compared to RNA/DNA hybrids formed over the first intron of the highly expressed housekeeping gene *γ-Actin*, which are quickly turned over after actinomycin D treatment (Figure 3.9B). Since the level of *FXN* RNA was reduced to 20% in this experiment (Figure 3.9A), the persistence of repeat-associated RNA/DNA hybrids suggests that the residual *FXN* RNA may be trapped in these stable RNA/DNA hybrids. One possible cause for the selective stability of GAA-associated RNA/DNA hybrids is that the repeat-containing *FXN* RNA itself may be processed and packaged differently from typical human RNA, thus allowing the RNA to hybridise to the DNA more efficiently. In line with this hypothesis, RNA FISH studies of mammalian cells have shown that several species of nuclear non-coding RNAs (ncRNAs) containing GAA repeats exist in normal cells and that these RNA species are particularly resistant to turnover (Zheng *et al.*, 2010). Although the origin of these ncRNAs was not mapped, this study suggested that GAA-containing RNAs may generally be problematic substrates for cellular RNA degradation machineries (Zheng *et al.*, 2010). This may further be exacerbated in case of RNAs containing large GAA expansions measuring several kilobases, as is the case in FRDA.

Indeed, it has previously been shown that pre-rRNA produced from R-loop-forming rDNA is degraded by the 3'-5' exonuclease exosome complex in *S. cerevisiae* (El Hage *et al.*, 2010). Furthermore, loss of exosome subunits has been shown to lead to accumulation of R-loops and R-

loop-associated DNA damage (Chan *et al.*, 2014; Pefanis *et al.*, 2015; Wahba *et al.*, 2011). In addition to exhibiting intrinsic properties which may prevent cellular degradation, such as unusual secondary structures, repeat-containing RNAs may attract a different subset of RNA-binding proteins, which will influence the kinetics and dynamics of RNA processing. In support of this possibility, synthetic $(GAA)_{10\text{-}15}$ RNA was shown to interact with a distinct set of proteins, including SRSF1, SRSF5, SRSF6, hnRNP A1/A2, hnRNP U and PABPN1, but not PTBP1 or SRSF3 (Baralle *et al.*, 2008; Zheng *et al.*, 2010). Thus, the differential binding of RNA processing factors may affect *FXN* RNA stability, and consequently enhance its ability to form RNA/DNA hybrids and R-loops at expanded GAA repeats. Although these findings suggest a possible contribution of RNA processing to FRDA pathology, this has not yet been investigated *in vivo*.

Alleviating the transcriptional defect by overexpression of RNase H1

Although the precise cause for the inability of endogenous enzymes to fully resolve repeat-associated RNA/DNA hybrids remains to be investigated, this work has shown that overexpression of exogenous RNase H1 can cause a decrease in RNA/DNA hybrid levels at expanded repeats, leading to transcriptional upregulation of *FXN* gene expression *in vivo* (Figure 4.4). This offers the interesting possibility that targeting RNase H1 specifically to repeat-associated R-loops could reverse the transcriptional defect and thus the cause of FRDA pathology in affected tissues (see also discussion in Chapter 6.3.4).

Formation of unusual DNA structures

Unusual DNA structures may form at expanded repeats and stabilise the formation of R-loops, since GAA-repeat-containing DNA can form DNA triplexes or ‘sticky DNA’ *in vitro* (Grabczyk and Usdin, 2000a, b; Potaman *et al.*, 2004; Sakamoto *et al.*, 1999; Sakamoto *et al.*, 2001). The formation of DNA triplexes, leading to folding-back of the GAA non-template strand to upstream sequences could uncover the TTC template strand, which then would be available for hybridisation by the GAA-containing RNA (Figure 1.9B) (Grabczyk and Usdin, 2000b). In addition to promoting RNA hybridisation to the template DNA strand, DNA structures could also directly affect binding of protein factors to the *FXN* gene, thus hampering efficient RNA processing, further

enhancing RNA/DNA hybrid formation. However, the interplay between unusual DNA structures, such as triplexes or G4 quadruplexes with R-loops has not been investigated *in vivo* so far.

In addition to the intrinsic characteristics of repeat-containing underlying RNA or DNA, the high stability of R-loops formed at expanded GAA repeats may be due to a failure of their complete turnover by endogenous enzymes targeting RNA/DNA hybrids. This work has shown that depletion of endogenous RNase H1 leads to increased RNA/DNA hybrid levels, which suggests that RNA/DNA hybrids at GAA repeats are natural targets of endogenous RNase H1 but they cannot be fully degraded in FRDA (Figure 4.3). In support of this, R-loop-forming CAG repeats and common fragile sites exhibit increased transcription-induced instability when RNase H1 is depleted, suggesting that RNA/DNA hybrids in R-loops and R-loop-associated DNA damage are at least partially supressed by endogenous RNase H1 (Helmrich *et al.*, 2011; Lin *et al.*, 2010).

4.5.4 Interplay between R-loops and heterochromatin in FRDA

H3K9me2 and R-loop formation

Recently it was suggested that the repressive chromatin H3K9me2 modification was not directly responsible for the *FXN* transcriptional repression (Punga and Buhler, 2010). Moreover, inhibition of histone deacetylases (HDACs) only leads to partial reactivation of the *FXN* gene in FRDA model systems, usually amounting to a ~2-3-fold increase in *FXN* RNA levels (Chan *et al.*, 2013; Herman *et al.*, 2006). Reducing the levels of the repressive H3K9me2 chromatin mark does not result in decrease of RNA/DNA hybrids on the expanded allele or upregulation of *FXN* RNA in lymphoblastoid cells (Figure 4.5). As shown in Figures 4.6–4.7, increasing RNA/DNA hybrid levels instead leads to an increase in H3K9me2 and subsequent repression of the *FXN* gene expression. This suggests that RNA/DNA hybrids in R-loops trigger heterochromatin formation and transcriptional repression of *FXN* gene in FRDA. Interestingly, recruitment of G9a methyltransferase is enhanced on the expanded *FXN* allele (Figure 4.10A), providing an interesting possibility that R-loops may directly recruit this enzyme to promote deposition of the H3K9me2 histone mark. Thus, although high levels of H3K9me2 in *FXN* gene are a clear marker of FRDA, this histone modification appears to be dispensable for *FXN* repression. These data may furthermore indicate that RNA/DNA hybrid

and R-loop formation is an early event in the process of *FXN* transcriptional gene silencing, which happens prior to the appearance of repressive epigenetic histone modifications. Thus, the inability to fully reactivate *FXN* alleles with expanded GAA repeats using HDAC or histone methyltransferase inhibitors may suggest that R-loops are not affected by these treatments and maintain the *FXN* gene in a repressed state.

Although the increase in H3K9me2 levels follows induction of RNA/DNA hybrids and the transcriptional repression of the *FXN* gene, it is possible that this histone modification still plays an important role in fine-tuning transcriptional regulation or in mediating stable inheritance of the repressed state over generations. It is also possible that H3K9me2 is required only initially to nucleate formation of heterochromatin by recruitment of the H3K9me2-interacting protein HP1α, which allows binding of H3K9 trimethyltransferase such as SUV39H1 (Mozzetta *et al.*, 2015). In line with this, H3K9me3 is highly enriched at expanded GAA repeats in FRDA (Al-Mahdawi *et al.*, 2008; Herman *et al.*, 2006; Kim *et al.*, 2011; Punga and Buhler, 2010; Soragni *et al.*, 2008). Once established, HP1α and SUV39H1 could perpetuate H3K9me3 without the need for H3K9me2 since both proteins can interact with each other and bind to H3K9me3 directly (Mozzetta *et al.*, 2015). Moreover, H3K9me2/3 methyltransferase complexes contain histone deacetylases (HDACs), a class of enzymes known to contribute to *FXN* repression in FRDA (Gottesfeld *et al.*, 2013; Herman *et al.*, 2006; Mozzetta *et al.*, 2015; Soragni *et al.*, 2014).

These links show that there may be a potential cascade of epigenetic silencing including H3K9 dimethylation by G9a/GLP, followed by H3K9 trimethlyation by SUV39H1 and histone deacetylation by HDACs, all initially triggered by RNA/DNA hybrids in R-loops formed at expanded repeats. However, more work is clearly required to uncover these potential links.

Role of RNAi in heterochromatin formation in FRDA

Two recent reports have linked R-loops with heterochromatin formation, mediated by factors of the RNA interference (RNAi) machinery. In *S. pombe*, R-loops were required for induction and maintenance of heterochromatin at centromeric repeats by the RNA-induced transcriptional silencing complex (Nakama *et al.*, 2012). In mammalian cells, R-loops are required for induction of antisense

transcription over terminator regions and subsequent recruitment of the RNAi factors Dicer, Ago1 and Ago2, leading to establishment of local heterochromatin (Skourti-Stathaki *et al.*, 2014). Interestingly, a heterologous expression system in *S. cerevisiae* has shown that expanded GAA repeats can recruit transcription initiation factors and act as an independent promoter in antisense direction (Zhang *et al.*, 2012). In light of these findings, it has recently been proposed that R-loops at GAA repeats allow antisense transcription over GAA repeats, formation of double-stranded RNA, recruitment of Ago1/2 and subsequent heterochromatin formation, similar to the mechanism observed at mammalian termination elements (Butler and Napierala, 2015).

Data obtained in this work however raise doubts about the likelihood of this mechanism. In particular, I observed localisation of antisense transcription, Ago1 and Ago2 proteins over the transcription start site of *FXN* gene rather than the GAA repeats (Figure 4.9). These three components of the RNAi machinery clearly anti-correlated with H3K9me2 levels and gene repression in FRDA patient cells containing expanded GAA repeats (Figures 3.1–3.3, 3.10). This suggests that an RNAi-independent pathway may link R-loops, G9a recruitment, heterochromatin formation and gene repression in FRDA.

4.5.5 Mechanisms of transcriptional repression by R-loops in FRDA

Interference with RNA Pol II transcription initiation

As discussed above, this and work of others strongly support a transcription initiation defect of the *FXN* gene in FRDA (See Chapter 4.5.1). However, it is less clear how expansion of the GAA repeats can affect a process taking place 1.6 kb further upstream, at the gene's promoter. One explanation applicable to all genes is the extensive feedback and crosstalk between various co-transcriptional processes (Moore and Proudfoot, 2009). Therefore, detrimental effects on elongation and *FXN* RNA processing may feed back, leading to reduced transcription initiation efficiency.

In addition, introns frequently contain positive regulatory sequences including enhancers (Rose, 2008). Deletion of one such enhancer element in *FXN* intron 1 reduces the activity of the *FXN* promoter *in vitro* (Greene *et al.*, 2007). The formation of RNA/DNA hybrids in R-loops and heterochromatin within the first *FXN* intron may therefore interfere with enhancer binding to the

gene. Furthermore, spread of repressive histone modifications and loss of activating histone modifications (amongst others histone acetylation) also affects the promoter in FRDA and may thereby influence efficiency of transcription initiation (Al-Mahdawi *et al.*, 2008; Chan *et al.*, 2013; Chutake *et al.*, 2014; De Biase *et al.*, 2009; Kumari *et al.*, 2011). Interestingly, nuclease-mediated excision of GAA repeats and surrounding 1230 bp (334 bp up- and 896 bp downstream of repeats) in control cells did not affect *FXN* expression, demonstrating that GAA repeats themselves are not required for normal *FXN* function (Li *et al.*, 2015). In contrast, excision of expanded GAA repeats alleviated transcriptional defects in FRDA cells, thus clearly demonstrating that expanded GAA repeats are the primary trigger of *FXN* repression (Li *et al.*, 2015). The extension of these *in vivo* gene modification studies using zinc-finger nucleases or CRISPR/Cas9 to additional sequences inside the first intron of the *FXN* gene could provide useful insights into the function of these elements for *FXN* expression in control and FRDA cells.

Direct interference of R-loops with elongating RNA Pol II

Previous work has demonstrated that co-transcriptionally formed RNA/DNA hybrids mediate transcription elongation impairment *in vitro* and in *S. cerevisiae* (Huertas and Aguilera, 2003; Tous and Aguilera, 2007), suggesting that these structures may provide roadblocks for RNA polymerases. RNA/DNA hybrids in R-loops over expanded repeats may form a structural block, directly interfering with RNA Pol II transcription elongation. Similar to R-loops at the 3'ends of human genes (Skourti-Stathaki *et al.*, 2011), expansion-associated R-loops could promote RNA Pol II termination, resulting in reduction of active RNA Pol II downstream of the expansion, as detected in this study. Alternatively, R-loops could lead to a pile-up of RNA Pol II at the expanded repeats, thereby preventing efficient elongation by subsequent RNA Pol II. According to this model, an enrichment of RNA Pol II would be expected at expanded GAA repeats in FRDA. So far, no data supporting this have been found either in this study (Figure 3.1D, 3.3B) or by others (Kim *et al.*, 2011; Kumari *et al.*, 2011; Punga and Buhler, 2010). However, it is possible that RNA Pol II may be ubiquitylated and degraded by the proteasome system after becoming irreversibly stalled at expanded repeats (Wilson *et al.*, 2013). Interestingly, this arrest-induced degradation of RNA Pol II may be a

possible explanation for reduced total and active RNA Pol II around expanded GAA repeats found here and elsewhere.

Aberrant splicing of GAA-repeat-containing FXN RNA

Many repeat expansion diseases are associated with alterations of splicing *in cis* or *in trans* (see Chapter 1.3.2). In line with this, several lines of evidence have pointed towards aberrant splicing caused by expanded GAA repeats. In *S. cerevisiae*, heterologous intronic insertion of GAA repeats led to inefficient splicing, leading to reduced mRNA levels (Shishkin *et al.*, 2009). Similarly, multiple plasmid-based reporter systems in mammalian cells exhibited complex splicing defects when GAA repeats were inserted (Baralle *et al.*, 2008). It was suggested that the binding of multiple splicing factors to GAA-repeat RNA may impair efficient splicing and lead to the observed accumulation of splicing intermediates (see Chapter 4.5.3) (Baralle *et al.*, 2008).

Interestingly, so far no splicing defect has been described for the endogenous *FXN* gene in FRDA patients. The lack of direct data on *FXN* dysregulated splicing in FRDA cells may in part be explained by the choice of methodology. While being very sensitive, the widely employed RT-qPCR approaches only detect specific amplicons and do not necessarily reveal the precise identity or length of the underlying transcript. Other methods that are more appropriate to detect alternative splice forms of *FXN* mRNA including northern blotting and RNase protection assays have been performed in FRDA cells but they did not detect any major alternatively spliced *FXN* mRNAs (Bidichandani *et al.*, 1998; Punga and Buhler, 2010). However, these experiments may have limited sensitivity in detecting splicing isoforms of low abundance. Therefore, a definitive conclusion whether or not a fraction of *FXN* RNA may undergo alternative splicing cannot be drawn with the currently available data. In addition, the presence of truncated GAA-repeat-containing *FXN* RNA species cannot be excluded. In line with this possibility, a previously undetected RNA was characterised for the *huntingtin* (*HTT*) gene (Sathasivam *et al.*, 2013). This RNA consists of the first *HTT* exon including the CAG expansion and part of the first *HTT* intron, where the use of cryptic poly(A)-signals leads to the production of a truncated, stable and translated $(CAG)_n$-RNA (Sathasivam *et al.*, 2013).

A potential truncated low abundance *FXN* RNA species may be particularly resistant to degradation by cellular processes, and could contribute to RNA/DNA hybrid and R-loop formation. Depending on the degree of truncation up- and downstream of the GAA repeats, it may not be readily detected by the RT-qPCR methods as employed in this work. Indirect evidence suggests that splicing of *FXN* may indeed be affected in FRDA. Splicing occurs mostly co-transcriptional and it is strongly affected by the speed of RNA Pol II elongation and the chromatin structure of the underlying DNA sequence (Nieto Moreno *et al.*, 2015). As discussed in Chapter 4.5.1, data from this work and others have demonstrated that RNA Pol II elongation at expanded GAA repeats is significantly impaired. Furthermore H3K9me2, enriched over expanded GAA repeats of the *FXN* gene, has been associated with inclusion of alternative exons (Nieto Moreno *et al.*, 2015). Taken together, it may be possible that R-loops and a highly compact chromatin structure around expanded GAA repeats could lead to slowing of RNA Pol II, which may facilitate alternative splicing or the use of cryptic poly(A) sites, potentially from the *Alu* element associated with the GAA repeat (Clark *et al.*, 2004). It will be interesting to employ unbiased and sensitive approaches such as deep sequencing to investigate potential splicing defects of the *FXN* gene in FRDA.

Intronic FXN RNA as a potential binding platform for Polycomb complexes

A GAA repeat-containing *FXN* intron 1 RNA may also contribute directly to transcriptional repression of *FXN* in FRDA, potentially by recruiting repressive protein factors. In line with this, it was shown that several intronic RNAs can epigenetically repress their host genes *in cis* by directly interacting with and recruiting the Polycomb repressive complex PRC2 (Guil *et al.*, 2012). Interestingly, the repressive histone modification H3K27me3, which is deposited by PRC2, is enriched at the *FXN* promoter and around the GAA repeats in FRDA cells (Chan *et al.*, 2013; Chutake *et al.*, 2014; De Biase *et al.*, 2009; Kim *et al.*, 2011; Soragni *et al.*, 2014). These observations highlight the possibility that repeat-containing intronic *FXN* RNA, anchored to the *FXN* gene due to its integration into R-loops, may act as a binding platform for Polycomb complex recruitment and thus *FXN* gene silencing. It will be important to investigate the potential role of

PRC2 in *FXN* gene repression in FRDA and to decipher its mode of recruitment either by *FXN* RNA or yet unidentified proteins such as sequence-specific transcription factors.

It has been shown that PRC2 can also be recruited by binding to H3K27me1 which is another histone modification that can be established by G9a (Mozzetta *et al.*, 2014), but the distribution of H3K27me1 has not been investigated in FRDA so far. Alternatively, Polycomb complexes can be recruited to target genes by sampling of CpG islands, followed by their stable establishment in response to repressed transcription (Blackledge *et al.*, 2015; Klose *et al.*, 2013). This mechanism could potentially contribute to repression of *FXN* gene, which contains a CpG island promoter and it is conceivable that the initial trigger of transcriptional repression and PRC2 stabilisation may be the formation of R-loops.

4.5.6 R-loops in the pathology of FXS

Reactivation of the silenced FMR1 gene

In FXS cells, the promoter of the *FMR1* gene containing $(CGG)_{n>200}$ expansions is fully methylated and is enriched for several repressive histone modifications (see Chapter 1.3.5). Numerous studies have aimed to reactivate the silenced *FMR1* gene to restore FMRP expression in FXS patient cells. Treatment with the DNA methylation inhibitor 5-azadC leads to almost complete *FMR1* promoter demethylation but restores expression of *FMR1* mRNA to only 10–25% of unexpanded alleles (Pietrobono *et al.*, 2002). In line with this, *FMR1* mRNA expression was restored to ~30% in FXS lymphoblasts treated with 5-azadC in this work (Figure 3.11B). The reactivation is accompanied by increases in *FMR1* promoter histone acetylation and H3K4 dimethylation (Coffee *et al.*, 1999; Tabolacci *et al.*, 2005). However, attempts to reactivate the *FMR1* gene by treatment with histone deacetylase inhibitors have been less successful, leading to no or very low *FMR1* mRNA expression despite significant increases in histone acetylation (Biacsi *et al.*, 2008; Coffee *et al.*, 1999; Tabolacci *et al.*, 2008a). The inability to fully restore *FMR1* expression despite changes in epigenetic marks suggested that additional mechanisms are involved in *FMR1* silencing, including the formation of unusual RNA/DNA structures.

R-loops as mediators of transcriptional silencing of the FMR1 gene

The high G-content of RNA transcribed from CGG repeats allows for the formation of thermodynamically highly stable RNA/DNA hybrids (see Chapter 1.1.1). Indeed, R-loop formation at CGG repeats has been demonstrated *in vitro* (Reddy *et al.*, 2011). As shown in Figure 3.11C, FXS lymphoblasts with a silenced *FMR1* allele exhibit background RNA/DNA hybrid levels, in line with the absence of *FMR1* mRNA (Figure 3.11B). However, upon partial reactivation of *FMR1* transcription by 5-azadC treatment, a significant enrichment of RNase H1-sensitive RNA/DNA hybrids in proximity to the expanded CGG repeats was observed (Figures 3.11C, D). This accumulation of RNA/DNA hybrids suggested that they may be involved in preventing full reactivation of the *FMR1* gene, potentially involving mechanisms similar to FRDA repression (discussed in Chapter 4.5.5). In support of their transcriptionally dependent formation, RNA/DNA hybrids have recently been detected on the *FMR1* gene using DIP in cells from unaffected and CGG premutation carriers (Loomis *et al.*, 2014). In the same study, single-stranded DNA was detected on the non-template DNA strand at CGG repeats using native bisulfite sequencing in the same location as RNA/DNA hybrids, strongly suggesting that indeed the formation of RNA/DNA hybrids is associated with R-loop formation *in vivo* (Loomis *et al.*, 2014).

The possibility of a functional involvement of R-loops in *FMR1* silencing was further investigated in my work with a time course experiment measuring the levels of *FMR1* mRNA and RNA/DNA hybrids during 5-azadC treatment and the subsequent 5-azadC washout. DIP analysis showed that even after removal of 5-azadC, RNA/DNA hybrid levels remained significantly elevated for several days, similar to *FMR1* mRNA levels (Figure 3.12A, B). This further points to the stability of pathological R-loops as observed for GAA-associated R-loops in FRDA. A recent study has found a comparable dynamic response to 5-azadC treatment and washout for the repressive histone modification H3K27me3 (Kumari and Usdin, 2014). Similar to RNA/DNA hybrids in the present work (Figure 3.12A), H3K27me3 levels increase immediately upon 5-azadC treatment of FXS lymphoblasts and then remain at high levels for more than 20 days after removal of the drug (Kumari and Usdin, 2014). Interestingly, this study found that enrichment of H3K27me3 is dependent on transcription and the presence of *FMR1* mRNA and correlates with recruitment of methyltransferase

EZH2, a component of PRC2 (Kumari and Usdin, 2014). Taken together, these data suggest that the formation of R-loops at expanded CGG repeats may serve as a binding platform for PRC2, thus promoting the silencing of the *FMR1* gene. In line with a primary function of R-loops in *FMR1* gene repression, it has recently been demonstrated that the appearance of R-loops during differentiation of FXS-derived iPSCs to neurons triggered heterochromatin formation and *FMR1* gene silencing (Colak *et al.*, 2014).

The role of H3K9me2 in FMR1 gene silencing

Several lines of evidence suggest that H3K9me2 is dispensable for *FMR1* gene silencing in differentiated FXS cells. While reactivation of FM alleles using 5-azadC leads to significant loss of repressive H3K27me2/3 and increase in H3K4me2 levels, H3K9me2 is not affected and remains at high levels (Bar-Nur *et al.*, 2012; Biacsi *et al.*, 2008; Tabolacci *et al.*, 2005). Moreover, rare carriers of fully transcriptionally active FM alleles exhibit high H3K9me2 levels similar to silenced FM alleles, suggesting that this mark is not required for transcriptional silencing in somatic cells (Tabolacci *et al.*, 2008b). This is similar to the observations for the *FXN* gene in FRDA made in this work and by others, indicating that enrichment of H3K9me2 is a marker for expanded repeats but does not necessarily correlate with the transcriptional state of the repeat-containing gene (see Chapters 4.3.1, 4.5.4) (Punga and Buhler, 2010). Although dispensable for maintaining *FMR1* silencing in somatic cells, H3K9me2 may be involved in the initial silencing events occurring during development. In line with this, H3K9me2 levels increase due to the formation of transient *FMR1* R-loops during differentiation of FXS iPSCs to neurons (Colak *et al.*, 2014). Despite this interplay between R-loops and H3K9me2 in FXS, it remains to be determined if H3K9me2 itself is indeed required for *FMR1* silencing.

Other factors influencing FMR1 gene silencing: R-loop modulators?

Silencing of *FMR1* genes in full mutation (FM) carriers is developmentally regulated. Human FXS embryonic stem cells (hESCs) contain unmethylated *FMR1* alleles that are fully transcriptionally active but become silenced upon differentiation (Colak *et al.*, 2014; Eiges *et al.*, 2007). Furthermore, DNA demethylation and reactivation of FM alleles can be achieved by fusing

FXS patient cells with murine pluripotent stem cells (Wohrle *et al.*, 2001). This suggests that *FMR1* gene silencing is not only dictated by transcription of expanded CGG repeats but by the action of other cellular factors.

Interestingly, rare cases of individuals carrying FM alleles without exhibiting any FXS symptoms have been described. In these FXS patients, the FM alleles are not methylated in somatic cells, and H3K4me2 and H3K27me2/3 levels are similar to unexpanded alleles, while histone acetylation is intermediate between unaffected and FXS individuals (Tabolacci *et al.*, 2008b). It is possible that in these individuals, the FM alleles evaded inactivation during development, potentially due to disease-modifying factors. It would be interesting to investigate if in these rare cases, other genetic causes could explain the lack of inactivation. Amongst others, this could be due to reduced levels of repressive epigenetic factors or increased levels of RNA/DNA helicases which could prevent the formation of R-loops essential for *FMR1* gene silencing.

4.5.7 Conclusions

The data presented here propose that R-loops may be a common feature of many trinucleotide expansion diseases, contributing to their pathology *in vivo* (Figure 4.11). The ability of R-loops to trigger transcriptional silencing in trinucleotide expansion diseases makes them an attractive target for future therapeutic approaches to treat these devastating diseases (see Chapter 6.3.4). In addition to contributing novel insight to uncover the molecular mechanisms underlying FRDA and fragile X pathologies, this work also provides interesting implications for R-loop biology.

Taken into consideration this work and the work of others in the field, depending on their genomic location, R-loops may have different functions (Aguilera and Garcia-Muse, 2012; Groh and Gromak, 2014; Skourti-Stathaki and Proudfoot, 2014). Therefore, stable R-loops formed over expanded triplet repeats (this study) may be different from R-loops at the 3'-ends of genes (Ginno *et al.*, 2013; Skourti-Stathaki *et al.*, 2014; Skourti-Stathaki *et al.*, 2011) and R-loops formed over CpG island promoters (Ginno *et al.*, 2013; Ginno *et al.*, 2012). At promoters, R-loops play a protective role against epigenetic silencing (Chen *et al.*, 2015a; Ginno *et al.*, 2012). By contrast, R-loops over *FXN* expanded repeats correlate with a reduction in transcription elongation and the recruitment of

repressive chromatin marks. This suggests that R-loops may be 'sensed' differently, depending on their genomic location and sequence composition.

Understanding the molecular mechanisms that cells use to distinguish 'harmful' from 'useful' R-loops is an important biological question in the study of human diseases. To uncover alternative pathways of R-loop-mediated gene repression, it will be instrumental to identify and characterise proteins that interact with R-loops *in vivo*. These factors could influence R-loop levels at different genomic locations and mediate downstream effects of R-loops on gene expression and genome stability.

Chapter 5

Characterisation of the RNA/DNA hybrid interactome in HeLa cells

5.1 Introduction

As discussed in Chapters 4.5.2–4.5.6, the observation that R-loops mediate diverse functions depending on their genomic location and cellular state may be in part due to the interplay with protein factors. Most of the first descriptions of R-loops closely linked their formation with the loss of particular proteins, including *E. coli* Top1, *S. cerevisiae* Hpr1, and vertebrate SRSF1 (Drolet *et al.*, 1995; Huertas and Aguilera, 2003; Li and Manley, 2005; Masse and Drolet, 1999). Subsequently, many hypothesis-driven studies identified additional factors involved in R-loop biology, in particular subunits of the THO/TREX complexes and proteins regulating RNA metabolism, DNA replication and repair (for a list of proteins implicated in R-loop biology so far, see Table 1.1). These studies provided crucial novel mechanistic insights and significantly advanced the R-loop field but the majority of published R-loop literature is based on hypothesis-driven single-protein and single-gene studies.

More recently, a number of studies aimed to identify proteins involved in R-loop biology using unbiased methodology (Chan *et al.*, 2014; Paulsen *et al.*, 2009; Stirling *et al.*, 2012; Wahba *et al.*, 2011). By analysing R-loop levels or genome instability in cells depleted of a range of factors, these studies proved very valuable in extending the scope of R-loop-associated processes. However, they could not address whether the measured effects were mediated by direct interaction of the protein factors with R-loops. To address this important gap in our understanding of R-loop biology, the main aim of the second part of this work was to establish an unbiased method to identify proteins directly interacting with RNA/DNA hybrids *in vivo*, constituting the RNA/DNA hybrid interactome.

5.2 Establishment of the RNA/DNA hybrid IP method

5.2.1 Choice of methodology

The lack of a proteomic characterisation of factors interacting with R-loops may in part be due to the challenges asociated with proteomics studies of chromatin, including its insolubility in many extraction buffers and the demanding enrichment of specific target sequences or proteins (Guillen-Ahlers *et al.*, 2014). However, recent advances in quantitative mass spectrometry and downstream analysis tools have allowed significant progress in characterising the protein composition of specific chromatin regions *in vivo* (Guillen-Ahlers *et al.*, 2014).

Two classes of approaches have been used to identify proteins directly binding to nucleic acids *in vivo* (Tacheny *et al.*, 2013). The first group of methods is based on the pulldown of proteins with synthetic oligonucleotides from nuclear extracts (Tacheny *et al.*, 2013). Alternatively, endogenous nucleic acids can be targeted and selectively enriched using DNA or RNA hybridisation probes (Simon *et al.*, 2011), endogenous affinity tags (Guillen-Ahlers *et al.*, 2014) or by direct labelling of target RNA or DNA (Haynes, 1999). Although traditionally widely used, the pulldown of proteins using synthetic oligonucleotides has several limitations. Synthetic oligonucleotides may not exhibit the same covalent modifications or topological features as their endogenous counterparts (Nieto Moreno *et al.*, 2015). Furthermore, protein-RNA/DNA interactions *in vivo* are complex and highly regulated, as exemplified by the process of transcription initiation by RNA polymerases (Jankowsky and Harris, 2015; Jonkers and Lis, 2015). To avoid these limitations, I therefore decided to develop a strategy to purify endogenous RNA/DNA hybrids with all associated proteins. To date, a unifying RNA/DNA hybrid-specific motif in sequences forming R-loops *in vivo* has not been described and R-loops occur at many different genomic locations (Costantino and Koshland, 2015; Santos-Pereira and Aguilera, 2015; Skourti-Stathaki and Proudfoot, 2014). This precludes the use of sequence-specific hybridisation approaches previously employed to identify protein interactomes of lncRNAs, heterochromatin and telomeres (Guillen-Ahlers *et al.*, 2014; Simon *et al.*, 2011). The introduction of transgenic affinity tags into DNA or RNA at known R-loop forming sequences could be feasible, especially with the advent of the CRISPR technique, but would only identify the

interactome at one specific location. Although this approach could prove useful to define R-loop-binding proteins at specific chromatin locations, it would only provide a limited insight into the global RNA/DNA hybrid and R-loop interactome.

Two affinity-based methods to recognise RNA/DNA hybrids in cells have been described before. They rely either on the usage of the RNA/DNA hybrid-specific S9.6 antibody (Boguslawski *et al.*, 1986; Phillips *et al.*, 2013) or a catalytically inactive mutant of RNase H1 (or its RNA/DNA hybrid-binding domain), capable of binding but not cleaving RNA/DNA hybrids (Bhatia *et al.*, 2014; Britton *et al.*, 2014; Ginno *et al.*, 2012). We decided to use the S9.6 as it is well characterised and successfully used by many independent groups to study R-loop biology (Arora *et al.*, 2014; El Hage *et al.*, 2010; El Hage *et al.*, 2014; Garcia-Rubio *et al.*, 2015; Ginno *et al.*, 2012; Herrera-Moyano *et al.*, 2014; Sun *et al.*, 2013; Yang *et al.*, 2014; Yeo *et al.*, 2014). Moreover, evidence generated in this and other groups suggested that affinity-tagged overexpressed RNase H1 mutants may not have access to all endogenous RNA/DNA hybrids due to differences in affinity or steric hindrance. This is clearly illustrated in one study when a lower number of RNA/DNA hybrids was identified at promoters using mutant RNase H1 compared to S9.6 antibody (318 versus 1978 gene hits, respectively) (Ginno *et al.*, 2012).

5.2.2 RNA/DNA hybrid slot blot

To optimise conditions for the RNA/DNA hybrid immunoprecipitation method (RNA/DNA hybrid IP), it was important to measure the efficiency of RNA/DNA hybrid pulldown. A method suitable to detect RNA/DNA hybrids based on immunoblotting has been described before (Rigby *et al.*, 2014; Sollier *et al.*, 2014). Therefore, the RNA/DNA hybrid slot-blot method was adapted and more extensively validated.

To test the specificity of the RNA/DNA hybrid slot blot for RNA/DNA hybrids, synthetic short nucleotides were spotted on a positively-charged membrane (Hybond N+, Amersham) and probed with S9.6 antibody. As shown in Figure 5.1A, only RNA/DNA hybrids were detected, while single-stranded or double-stranded DNA or RNA oligonucleotides of the corresponding sequence did not produce a signal. Importantly, R-loops could also be detected in nuclear extracts prepared from

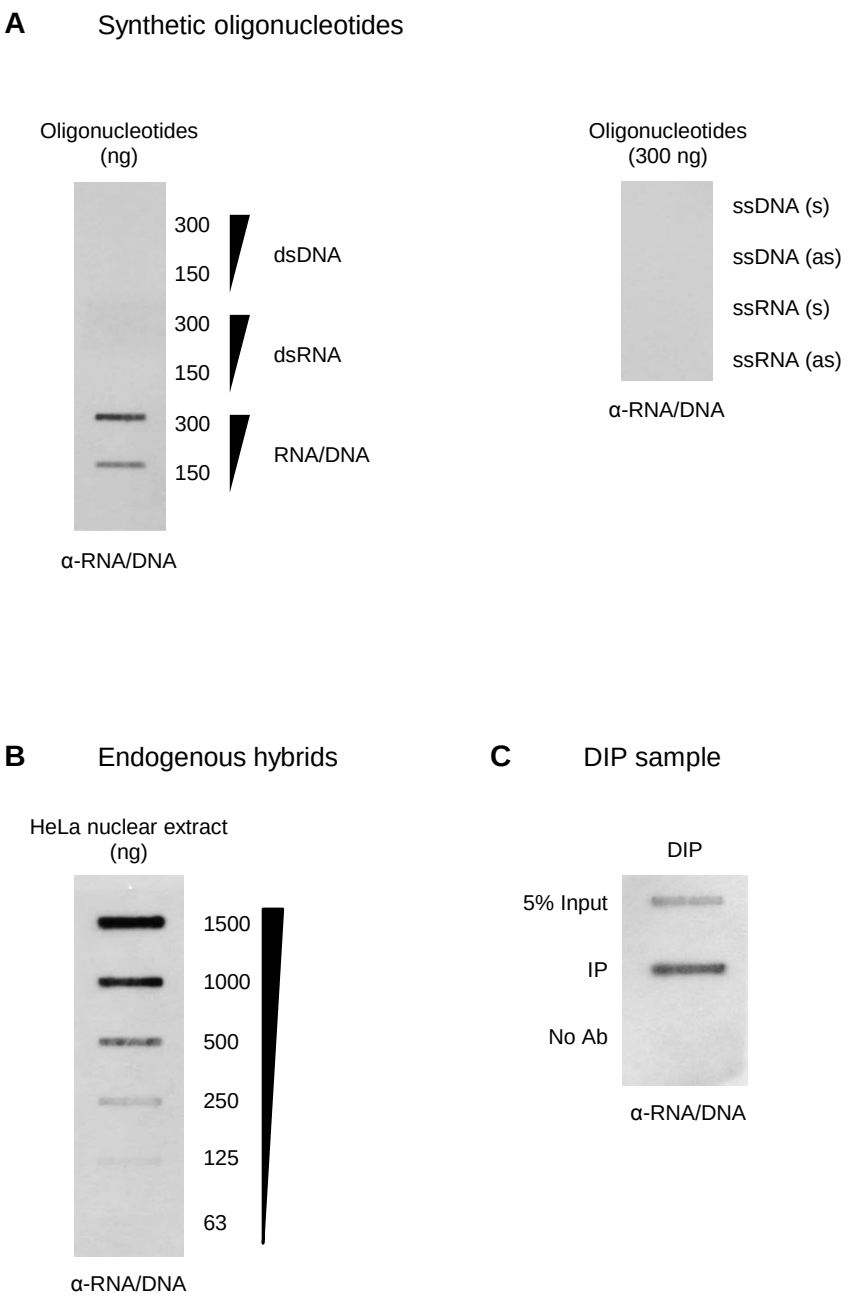


Figure 5.1. S9.6 antibody recognises endogenous and synthetic RNA/DNA hybrids

A. Slot blot analysis with S9.6 antibody. Left panel: 300 and 150 ng of synthetic double-stranded oligonucleotides (dsDNA, dsRNA, RNA/DNA hybrids) were loaded on the slot blot. Right panel: 300 ng of corresponding synthetic sense (s) and antisense (as) single-stranded oligonucleotides were used as negative controls.

B. RNA/DNA hybrid slot blot detects endogenous RNA/DNA hybrids over a wide range of concentrations (63–1500 ng purified genomic DNA from HeLa nuclear extract).

C. RNA/DNA hybrid slot blot detects enrichment of RNA/DNA hybrids in DIP samples. IP: immunoprecipitation with S9.6 antibody, No Ab: control IP without S9.6 antibody.

HeLa cells, with a signal over a wide dynamic range of nucleic acid amounts loaded on the membrane (125–1500 ng) (Figure 5.1B). Employing the RNA/DNA hybrid slot blot to analyse a typical DIP sample clearly showed a strong enrichment of immunoprecipitated RNA/DNA hybrids, while no signal was detected in the 'No Ab' control (Figure 5.1C). These results suggested that the RNA/DNA hybrid slot blot may be used to detect synthetic and endogenous RNA/DNA hybrids.

5.2.3 Optimisation of the RNA/DNA hybrid IP method

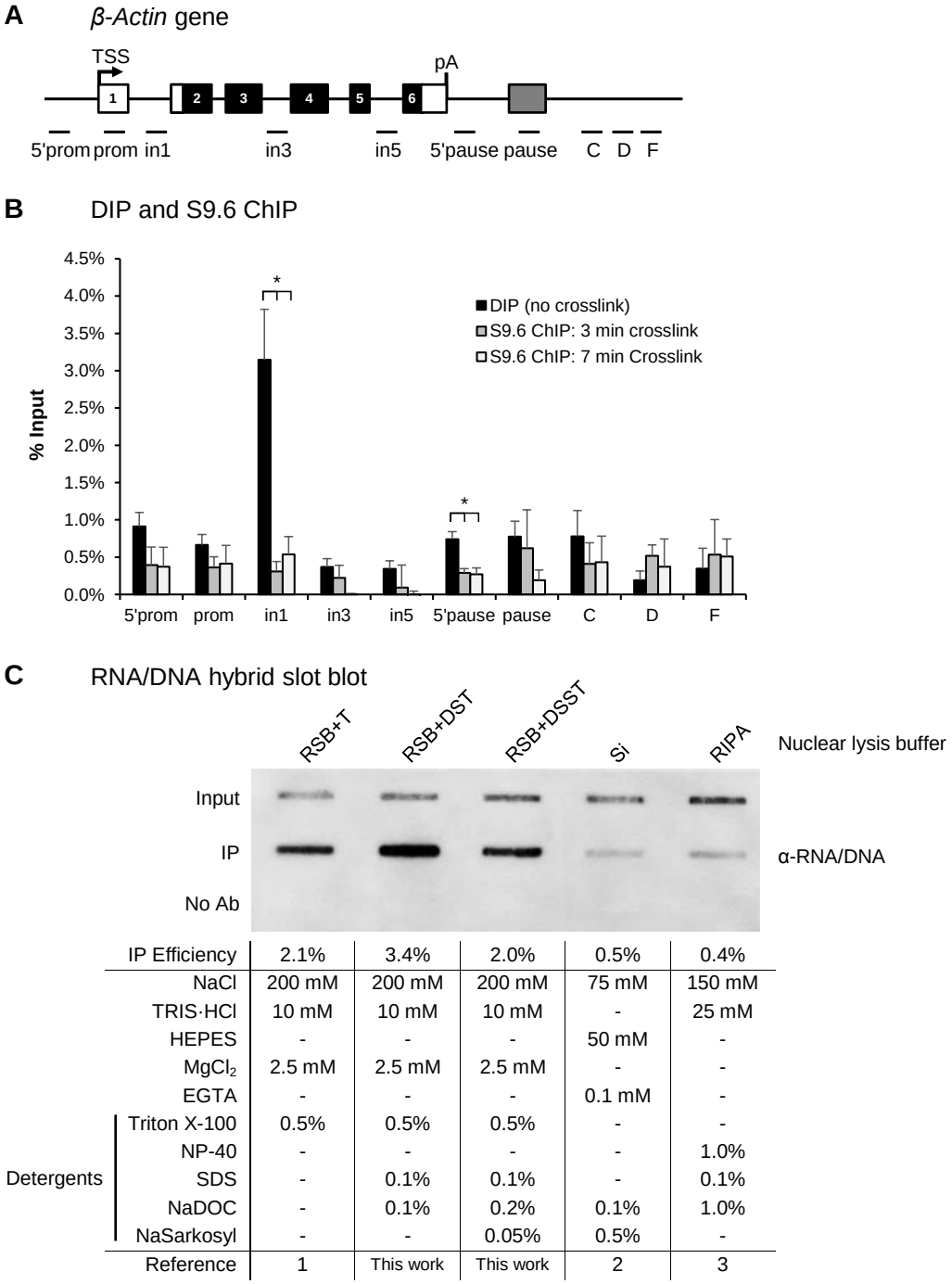
As the starting point for the procedure, Hybrid IP was designed to extract RNA/DNA hybrids and associated proteins from HeLa cell nuclei following nucleo-cytoplasmic fractionation. The main focus of optimisation was to find conditions for nuclear lysis and subsequent immunoprecipitation to ensure high RNA/DNA hybrid enrichment while reducing unspecific protein binding.

Crosslinking versus native purification

Initially, both formaldehyde and UV-induced crosslinking of cells prior to immunoprecipitation were tried. To determine the efficiency of RNA/DNA hybrid pulldown from formaldehyde-crosslinked cells, ChIP was carried out using the S9.6 antibody and compared to the RNA/DNA hybrid profile on the endogenous *β-Actin* gene in non-crosslinked HeLa cells as detected by DIP (Figure 5.2A). S9.6 ChIP from cells crosslinked for 3 or 7 min showed high background with no significant differences between signals in the *β-Actin* gene body (amplicons prom–in5) compared to intergeneric regions (amplicons 5'prom, D, F) (Figure 5.2B). Moreover, previously described RNA/DNA hybrid peaks at in1 and 5'pause were only detected using DIP but not S9.6 ChIP (Figure 5.2B) (Skourti-Stathaki *et al.*, 2011). These results suggested that crosslinking may artificially affect endogenous RNA/DNA hybrid levels or obstruct accessibility of S9.6 to endogenous RNA/DNA hybrids. Therefore, further RNA/DNA hybrid IP method optimisations were carried out using non-crosslinked HeLa cells.

Composition of nuclear lysis buffer

For optimal RNA/DNA hybrid IP, a balance had to be found between efficient extraction of RNA/DNA hybrids from the isolated nuclei and preservation of endogenous RNA/DNA



		RSB+T	RSB+DST	RSB+DSST	Si	RIPA
	IP Efficiency	2.1%	3.4%	2.0%	0.5%	0.4%
	NaCl	200 mM	200 mM	200 mM	75 mM	150 mM
	TRIS·HCl	10 mM	10 mM	10 mM	-	25 mM
	HEPES	-	-	-	50 mM	-
	$MgCl_2$	2.5 mM	2.5 mM	2.5 mM	-	-
	EGTA	-	-	-	0.1 mM	-
Detergents	Triton X-100	0.5%	0.5%	0.5%	-	-
	NP-40	-	-	-	-	1.0%
	SDS	-	0.1%	0.1%	-	0.1%
	NaDOC	-	0.1%	0.2%	0.1%	1.0%
	NaSarkosyl	-	-	0.05%	0.5%	-
	Reference	1	This work	This work	2	3

Figure 5.2. Optimisation of conditions for RNA/DNA hybrid IP

A. Diagram of *β-Actin* gene. Black boxes are exons, white boxes are 5' and 3'UTRs, lines are introns. TSS is the transcriptional start site. Position of the pA site is indicated. Grey box denotes the termination region. qPCR amplicons are shown below the diagram.

B. Comparison of RNA/DNA hybrid profiles on endogenous *β-Actin* gene in HeLa cells detected using DIP method (no crosslink) or S9.6 ChIP (3 min or 7 min crosslink).

C. Nuclear lysis buffer influences RNA/DNA hybrid IP efficiency (IP signal/Input signal), as demonstrated by RNA/DNA hybrid slot blot. Extracts were sonicated for 5 minutes. Composition of buffers is indicated below the membrane with references to sources of recipes: Haynes 1999 (1); Simon *et al.*, 2011 (2); Pierce/Thermo Fisher Scientific (3).

hybrid/protein complexes. To achieve efficient extraction of RNA/DNA hybrids and associated proteins from nuclei, a range of buffers of different stringencies was tested (Haynes, 1999; Simon *et al.*, 2011). Preliminary proteomics data of RNA/DNA hybrid IP employing the mild RSB+T buffer (10 mM TRIS pH 7.5, 200 mM NaCl, 2.5mM $MgCl_2$, 0.5% Triton X-100) for nuclear lysis identified only very few chromatin-associated factors, suggesting that this buffer may not efficiently extract chromatin and RNA/DNA hybrids from nuclei. Very stringent buffers such as RIPA buffer (25 mM TRIS·HCl pH 7.6, 150 mM NaCl, 1% NP40, 1% Deoxycholate, 0.1% SDS) and Si buffer (50 mM HEPES pH 7.6, 75 mM NaCl, 0.1 mM EGTA, 0.5% N-lauroylsarkosine, 0.1 % NaDOC, 5 mM DTT) extracted RNA/DNA hybrids more efficiently than the milder RSB+T (Figure 5.2C, Input lane). However, use of these stringent buffers resulted in reduced RNA/DNA hybrid immunoprecipitation efficiency of ~0.5% compared to ~2.1% for RSB+T (Figure 5.2C, IP lane). This could be due to impaired stability of RNA/DNA hybrids and R-loops in the very stringent buffers or reduced binding of S9.6 to RNA/DNA hybrids under these conditions. Therefore, two buffers with intermediate stringencies were tested: RSB+DST (10 mM TRIS pH 7.5, 200 mM NaCl, 2.5mM $MgCl_2$, 0.5% Triton X-100, 0.1% NaDOC, 0.1% SDS) and RSB+DSST (10 mM TRIS pH 7.5, 200 mM NaCl, 2.5mM $MgCl_2$, 0.5% Triton X-100, 0.2% NaDOC, 0.1% SDS, 0.05% NaSarkosyl). Both buffers extracted RNA/DNA hybrid more efficiently than RSB+T (Figure 5.2C, Input lane) while maintaining similar or higher IP efficiency. Although RSB+DST showed higher IP efficiency (~3.4% vs 2.0%), the more stringent RSB+DSST buffer was employed for all future experiments to reduce unspecific protein-protein interactions (Figure 5.2C, IP lanes). To minimise total duration of RNA/DNA hybrid IP post-lysis, the S9.6 antibody was pre-bound to protein A dynabeads, which were directly added to the lysates and incubated for 2 hours.

Sonication of nuclear extracts

To further reduce the pulldown of unspecific chromatin-binding proteins in the RNA/DNA hybrid IP, sonication of nuclear lysates was optimised. While 5 min sonication left large (>20 kb) fragments of genomic DNA, 10 or 15 min produced DNA fragment sizes around ~700 and ~ 500 bp, respectively (Figure 5.3A). Both 10 and 15 minute sonication led to reduced IP efficiency of

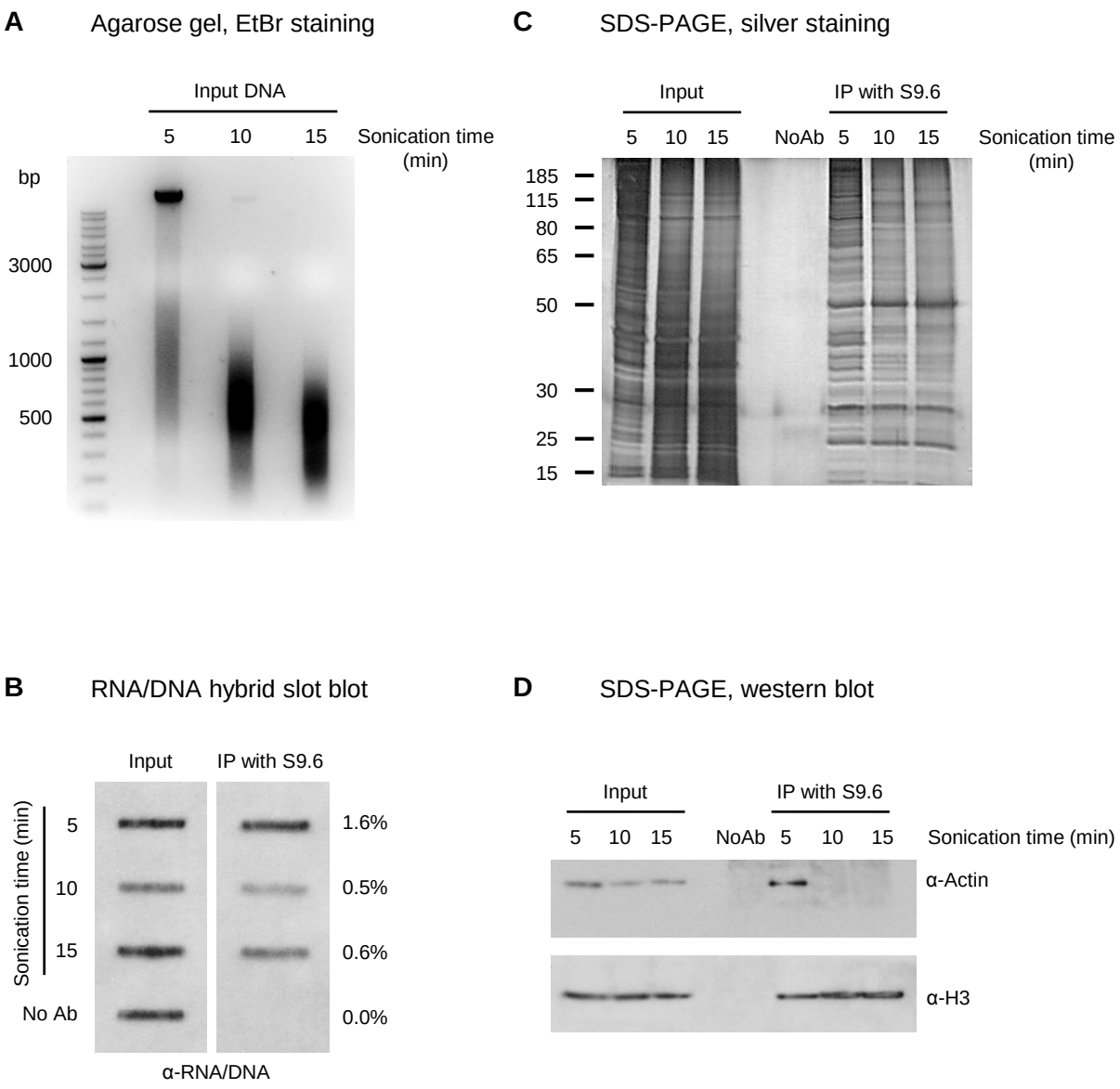


Figure 5.3. Optimisation of sonication for RNA/DNA hybrid IP

A. Agarose gel analysis of DNA fragment sizes of chromatin preparations from RNA/DNA hybrid IP samples, sonicated for 5, 10 and 15 min. DNA size markers are shown to the left of the gel.

B. RNA/DNA hybrid slot blot of RNA/DNA hybid IP samples using RSB+DSST buffer, produced using different sonication times. The blot was probed with S9.6 antibody. IP efficiency ('IP' signal/'Input' signal) is indicated to the right of the blot.

C. Silver staining of protein samples from RNA/DNA hybrid IP carried out with different sonication times.

D. Western blot analysis of protein samples from RNA/DNA hybrid IP carried out with different sonication times.

RNA/DNA hybrids (0.5–0.6% compared to ~1.6%) and co-purifying proteins (Figure 5.3B, C). Thus, it could not be excluded that RNA/DNA hybrid and R-loop integrity was negatively affected by the mechanical disruption during sonication. Co-purification of non-specific proteins at mild sonication conditions was demonstrated by the identification of Actin in the RNA/DNA hybrid IP after 5 min sonication, but not after longer durations (Figure 5.3D). As expected from the chromatin environment in which RNA/DNA hybrids and R-loops are formed *in vivo*, RNA/DNA hybrid IP immunoprecipitated histone H3 under all three sonication conditions (Figure 5.3D). As evident from the above results, a balance between immunoprecipitating RNA/DNA hybrids as efficiently as possible and reducing potential unspecific binding had to be found. Since co-immunoprecipitation of Actin was not detected after 10 min of sonication, and because there was no significant further reduction of DNA fragment size, this was the condition chosen for all further RNA/DNA hybrid IPs.

RNA trimming

It has been shown that the S9.6 antibody can bind to AU-rich dsRNA, albeit with 5.6-fold lower affinity than for RNA/DNA hybrids *in vitro* (Phillips *et al.*, 2013). Furthermore, during extraction procedures ssRNA may artificially hybridise to dsDNA, forming RNA/DNA hybrids in R-loops, or to other ssRNA, forming dsRNA (Zhang *et al.*, 2015). To reduce unspecific RNA-mediated interactions, a mild RNA trimming was included during the RNA/DNA hybrid IP. To find the optimal condition for RNA trimming, HeLa nuclear extracts were incubated with different concentrations of RNase A for 2 h. Analysis of nucleic acids using agarose gel electrophoresis showed that RNase A significantly degraded the most abundant RNA species (28S and 18S rRNA) starting at 0.01 ng RNase A/µg genomic DNA (Figure 5.4). At the highest RNase A concentrations (1–10 ng RNase A/µg genomic DNA), a reduction of the high molecular weight signal was observed, representing genomic DNA and associated RNA. This suggests that RNase A may also degrade genuine genomic RNA/DNA hybrids at these higher concentrations. All future experiments were therefore carried out at 0.1 ng RNase A/µg genomic DNA, where rRNA was efficiently degraded while not affecting the integrity of RNA/DNA hybrids (Figure 5.4).

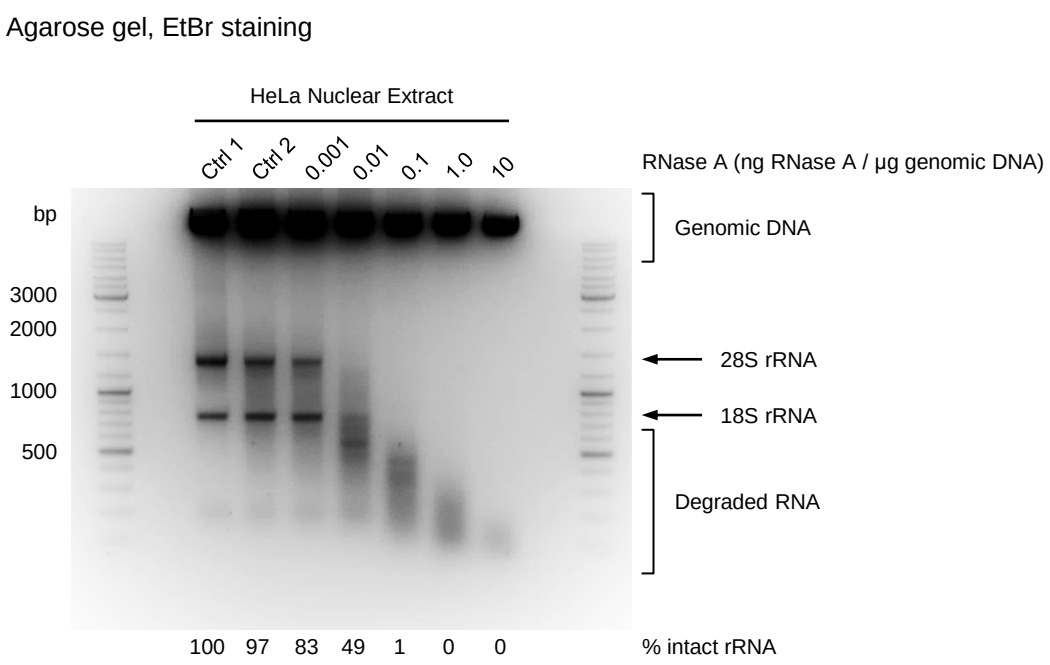


Figure 5.4. Titration of recombinant RNase A for RNA trimming during RNA/DNA hybrid IP

Agarose gel analysis of RNA trimming by recombinant RNase A added during RNA/DNA hybrid IP. HeLa nuclear extracts containing 12 µg of genomic DNA were prepared according to the RNA/DNA hybrid IP procedure. Recombinant RNase A was added to the indicated concentrations and incubated for 2 h. Two samples (Ctrl 1&2) were not treated with RNase A. Samples were deproteinised and extracted nucleic acids were separated on 1.2% agarose gel and stained with EtBr. DNA size markers are shown to the left of the gel. The migration of genomic DNA, ribosomal RNA (18S and 28S rRNA) and degraded RNA is indicated on the right of the gel. The fraction of intact rRNA (in %) is indicated below the gel and is based on the quantification of the EtBr staining signal between 0.6 and 2.0 kb.

5.3 Specificity of the RNA/DNA hybrid IP method

5.3.1 Investigation of candidate R-loop biology factors using RNA/DNA hybrid IP

The optimised RNA/DNA hybrid IP method was carried out and immunoprecipitated samples were analysed by SDS-PAGE followed by silver staining, western blotting, RNA/DNA hybrid slot blotting and mass spectrometry (Figure 5.5A). This optimised RNA/DNA hybrid IP procedure significantly enriched for RNA/DNA hybrids, as shown in Figure 5.5B. Immunoprecipitated RNA/DNA hybrid-interacting proteins were then analysed by SDS-PAGE followed by silver staining. As demonstrated in Figure 5.5C, a complex mixture of RNA/DNA hybrid-binding proteins was observed, which differed significantly from the nuclear input material. Importantly, these proteins were not immunoprecipitated when the S9.6 antibody was omitted from the purification procedure. In addition, when matched mouse isotype control IgG2 antibody was used instead of S9.6 antibody, very few bands and much weaker staining intensity was observed (Figure 5.5C). These results implied that the majority of the proteins co-immunoprecipitated by S9.6 antibody are not due to unspecific binding to the beads matrix or to the constant region of the antibody. Furthermore, the presence of a banding pattern different from the nuclear input suggested that the RNA/DNA hybrid IP selectively enriches for a unique set of interacting proteins.

Although knowledge about interaction of proteins with R-loops *in vivo* is very limited, it was reasoned that RNA/DNA hybrid IP will detect some strong candidates, previously implicated experimentally in R-loop biology. Therefore, these candidates were tested using western blotting (Figure 5.5D). The RNA/DNA hybrid IP co-purified RNA Pol II, in line with RNA/DNA hybrid and R-loop formation during transcription (Aguilera and Garcia-Muse, 2012). Interestingly, RNA/DNA hybrid IP material was enriched for hypophosphorylated Pol II (IIa), which may be related to the predominant formation of R-loops close to transcription start sites (Figure 5.5D) (Chen *et al.*, 2015a; Ginno *et al.*, 2012; Nadel *et al.*, 2015). RNA/DNA hybrid IP also immunoprecipitated the RNA/DNA helicase senataxin (SETX), which was previously shown to resolve RNA/DNA hybrids in R-loops at human termination regions (Hatchi *et al.*, 2015; Skourti-Stathaki *et al.*, 2011). Furthermore, Top1 was identified, corroborating its important role in R-loop biology

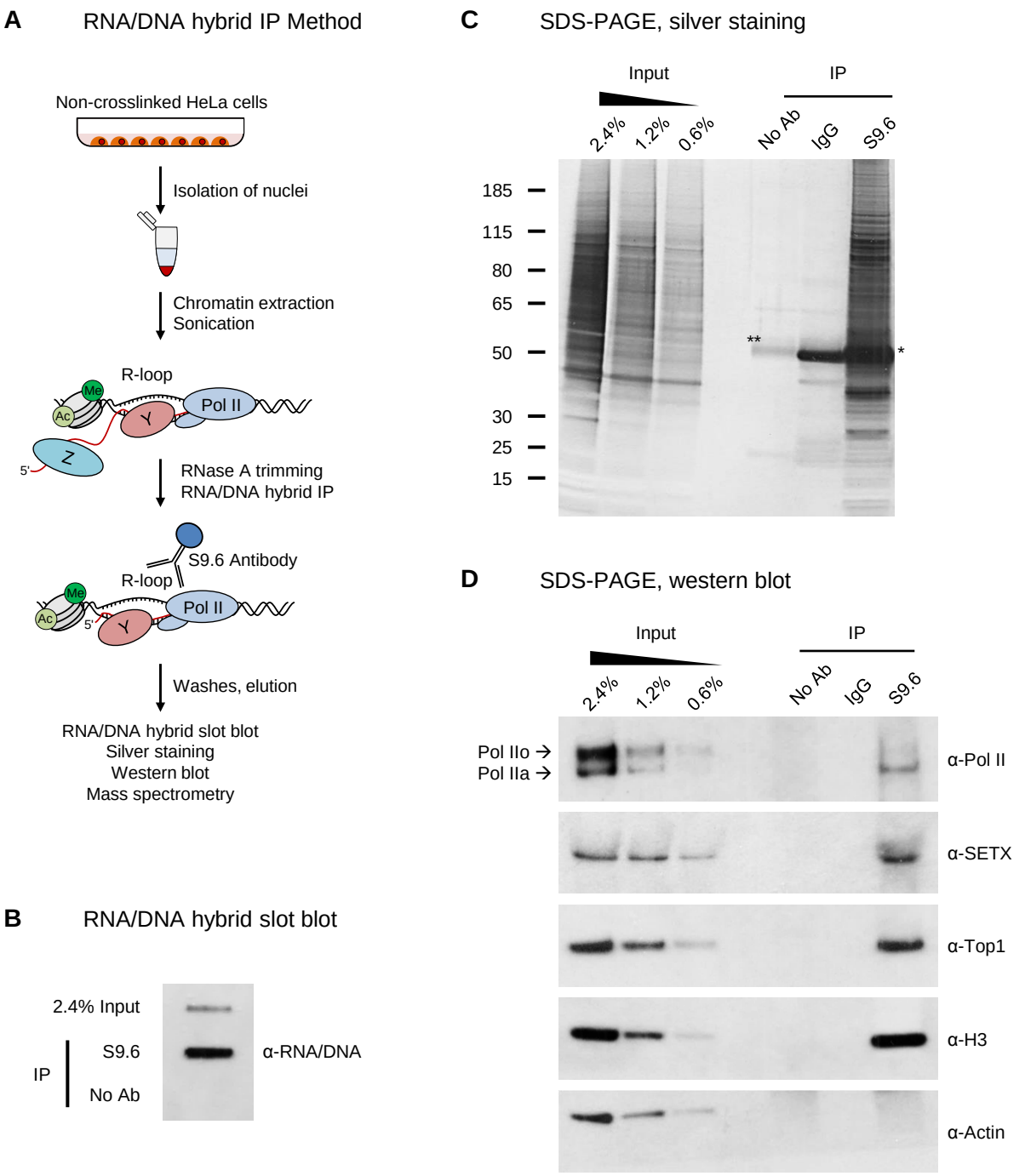


Figure 5.5. Investigation of candidate RNA/DNA hybrid interactors using RNA/DNA hybrid IP

A. Schematic diagram of RNA/DNA hybrid IP method using S9.6 antibody in HeLa cells.

B. RNA/DNA hybrid slot blot of RNA/DNA hybrid IP samples probed with S9.6 antibody.

C. Silver staining of protein samples from RNA/DNA hybrid IP, following SDS-PAGE. No antibody and isotype-matched IgG2a antibody were used as controls. * indicates the heavy chain from the S9.6 and IgG2a antibodies and ** the BSA used for pre-blocking of beads.

D. Western blot analysis of protein samples from RNA/DNA hybrid IP using RNA Pol II, SETX, Top1, H3 and Actin antibodies. 2.4%, 1.2%, and 0.6% of Input was loaded. Arrows indicate the positions of unphosphorylated (IIa) and hyperphosphorylated (IIo) forms of RNA Pol II.

(El Hage *et al.*, 2010; Tuduri *et al.*, 2009). As an integral part of chromatin, RNA/DNA hybrids predictably interacted with histone H3, but not with the abundant Actin (Figure 5.5D).

5.3.2 The S9.6 antibody does not bind proteins unspecifically in RNA/DNA hybrid IP

The affinity of the S9.6 antibody to several nucleic acid species has previously been characterised *in vitro* (Phillips *et al.*, 2013). However, it is not clear whether S9.6 may directly interact with cellular proteins, which could give rise to false negative RNA/DNA hybrid interactors. Therefore, the specificity of the S9.6 antibody for cross-reactivity with proteins in the RNA/DNA hybrid IP was investigated. For this, RNA/DNA hybrid IP was performed following treatment of nuclear extracts with benzonase, which degrades all forms of nucleic acids without affecting proteins (Aygun *et al.*, 2008). Benzonase effectively degraded RNA/DNA hybrids (Figure 5.6A) and this led to almost complete loss of immunoprecipitated proteins in the RNA/DNA hybrids, as shown by silver staining and western blotting (Figure 5.6B, C). Importantly, the S9.6 antibody itself was not affected by benzonase treatment as shown by equal intensities of S9.6 heavy and light chains in untreated and treated samples (bands at 55 kDa and 25 kDa, Figure 5.6B). These results confirmed that immunoprecipitation of proteins in the RNA/DNA hybrid IP is not due to unspecific binding to the S9.6 antibody but rather mediated by their binding to endogenous nucleic acids.

5.3.3 RNA/DNA hybrid IP is specific for RNA/DNA hybrids

To validate that proteins identified in the RNA/DNA hybrid IP specifically bind RNA/DNA hybrids and not other forms of nucleic acids present in nuclear extracts, RNA/DNA hybrid IP was carried out under competitive conditions with addition of synthetic double-stranded oligonucleotides over a wide concentration range (Figure 5.7A). First, a previously characterised synthetic 15 bp long RNA/DNA hybrid was used, which binds S9.6 antibody with 0.54 nM affinity *in vitro* (Phillips *et al.*, 2013). Addition of this artificial RNA/DNA hybrid competed for binding to the S9.6 antibody with the endogenous RNA/DNA hybrids and led to a complete loss of co-immunoprecipitated proteins at the majority of concentrations, apart from the two lowest ones (0.065; 0.013 μM) (Figure 5.7B). It is important to note that the heavy chain of the S9.6 antibody is clearly visible over all competitor

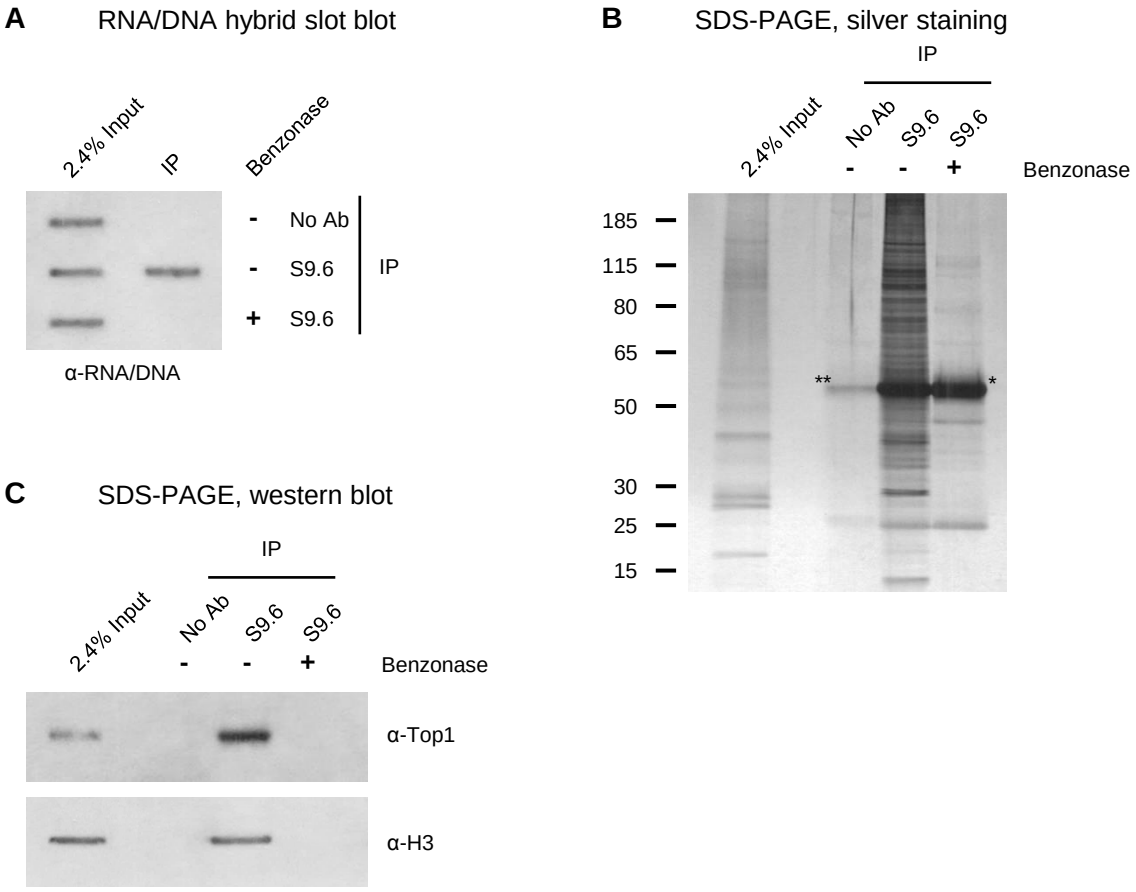


Figure 5.6. Confirmation of specificity of RNA/DNA hybrid IP using benzonase treatment

A. RNA/DNA hybrid slot blot of RNA/DNA hybrid IP samples probed with S9.6 antibody.

B. Silver staining of protein samples from RNA/DNA hybrid IP. * indicates the heavy chain from the S9.6 antibody and ** indicates migration of BSA used for pre-blocking of beads.

C. Western blot analysis of protein samples from RNA/DNA hybrid IP, probed with Top1 and H3 antibodies. A-C. Nuclear extracts were treated with benzonase nuclease (1 U/µl) for 30 min prior to IP with S9.6 antibody. 2.4% of Input was loaded.

concentrations, providing an internal loading control for all experiments (marked by asterisk in Figure 5.7B).

When the RNA/DNA hybrid IP was carried out with corresponding dsRNA or dsDNA competitors of the same sequence, no changes in the RNA/DNA hybrid IP protein profiles were observed (Figure 5.7C, left and right panel). Two unexpected bands at ~28 kDa and ~17 kDa appeared exclusively at the highest concentration of dsRNA competitor. While their origins are unclear it will be interesting to establish their identity using mass spectrometry in the future.

Next, the efficiency of competition by different oligonucleotides was analysed using western blotting. As shown in Figure 5.8A, synthetic RNA/DNA hybrid efficiently competed out the binding of Top1 and histone H3 to RNA/DNA hybrids. In contrast, addition of dsRNA or dsDNA to the RNA/DNA hybrid IP did not affect binding of either Top1 or histone H3 even at the highest concentrations (Figure 5.8A).

5.3.4 Confirmation of RNA/DNA hybrid IP specificity using α-CBP80 antibody

The results described above suggested a high specificity of RNA/DNA hybrid IP for RNA/DNA hybrids in nuclear extracts from non-crosslinked cells (Figures 5.6–5.8). However, the possibility had to be excluded that these observations were due to artefacts of the purification procedure itself, resulting in unspecific protein or DNA aggregation. To mitigate this concern, the RNA/DNA hybrid IP procedure was carried out using an antibody recognising mRNA cap-binding protein CBP80 instead of the S9.6 antibody. As shown in Figure 5.8B, the protein pattern from the CBP80 immunoprecipitation was clearly distinguishable from the one observed for RNA/DNA hybrid IP using S9.6 antibody (Figure 5.5C). To further confirm the specificity of the RNA/DNA hybrid IP approach, double-stranded synthetic oligonucleotides (RNA/DNA, dsRNA, dsDNA) were added during CBP80 immunoprecipitation. As shown in Figure 5.8B, none of these competitors affected CBP80 IP, in line with its role as a bridging factor for the 7-methyl-cytosine-binding protein CBP20 with numerous other RNA processing factors (Gonatopoulos-Pournatzis and Cowling, 2014). In conclusion, the RNA/DNA hybrid IP method exhibits high specificity and therefore efficiently enriches for proteins associated with RNA/DNA hybrids *in vivo*.

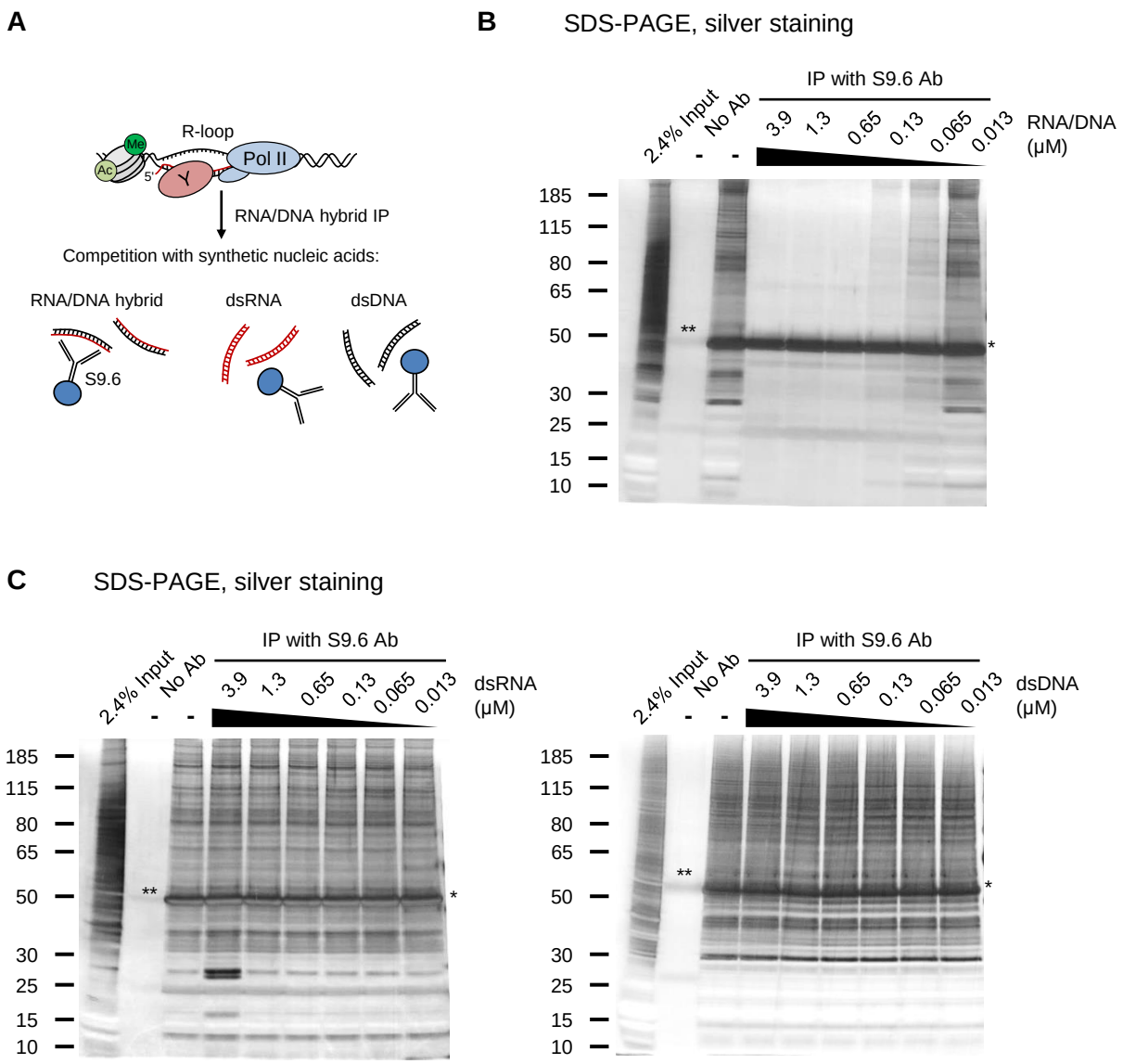


Figure 5.7. RNA/DNA hybrid IP in presence of double-stranded oligonucleotide competitors

A. Schematic of RNA/DNA hybrid IP competition experiments with synthetic RNA/DNA, dsRNA and dsDNA. Competitors were added to the nuclear extracts during IP with S9.6 antibody.

B. Silver staining of protein samples from RNA/DNA hybrid IP with RNA/DNA hybrid competitor added at 3.9-0.013 μM. 2.4% of Input material was loaded on the gel. * indicates the heavy chain from the S9.6 antibody and ** indicates migration of BSA used for pre-blocking of beads.

C. Silver staining of protein samples from RNA/DNA hybrid IP with synthetic dsRNA (left panel) and dsDNA (right panel) competitors added at 3.9-0.013 μM. 2.4% of Input was loaded. * indicates the migration of the heavy chain from the S9.6 antibody and ** indicates migration of BSA used for pre-blocking of beads.

5.4 Mass spectrometry analysis of the RNA/DNA hybrid interactome

5.4.1 Definition of the RNA/DNA hybrid interactome using quantitative mass spectrometry

After careful validation of the RNA/DNA hybrid IP method, a proteomic pipeline was developed based on label-free quantitative LC-MS/MS mass spectrometry to identify novel proteins associated with RNA/DNA hybrids in HeLa cells *in vivo*. For increased specificity of identified proteins, a control IP with S9.6 antibody was performed in the presence of 1.3 μM synthetic RNA/DNA hybrid competitor, as described in Chapter 5.3.3 (Figure 5.7A, B). Three biological replicates of RNA/DNA hybrid IP were carried out with S9.6 antibody in the presence or absence of RNA/DNA hybrid competitor. Eluted proteins were prepared for mass spectrometry by trypsin digestion using filter-assisted sample preparation (FASP) (Wisniewski *et al.*, 2009), followed by online one-dimensional separation on a nano-LC HPLC and analysis on a QExactive mass spectrometer (Thermo Fisher Scientific).

Using the MaxQuant software (Cox and Mann, 2008), relative enrichment for RNA/DNA hybrid IP versus control IP was calculated for each identified protein in each biological replicate. This approach identified 846 proteins in total in all three independent biological replicates of RNA/DNA hybrid IP. Importantly, the RNA/DNA hybrid IP showed high reproducibility, as demonstrated by a strong correlation of relative enrichment of proteins between all three biological replicates (Pearson's correlation coefficients $r>0.76$, Figure 5.9).

To define proteins significantly enriched in RNA/DNA hybrid IP compared to control IP, carried out in the presence of 1.3 μM RNA/DNA hybrid competitor, a stringent statistical analysis was performed. Since estimates of local variance are unstable for experiments with just three biological replicates, a standard Student's t-test would lack power to define the full set of enriched proteins (Castello *et al.*, 2012). The moderated t-test is more powerful than the Student's t-test because it combines the local variance for each protein with a global variance estimate over all proteins by an empirical Bayesian method (Smyth, 2004). To correct for multiple testing over this

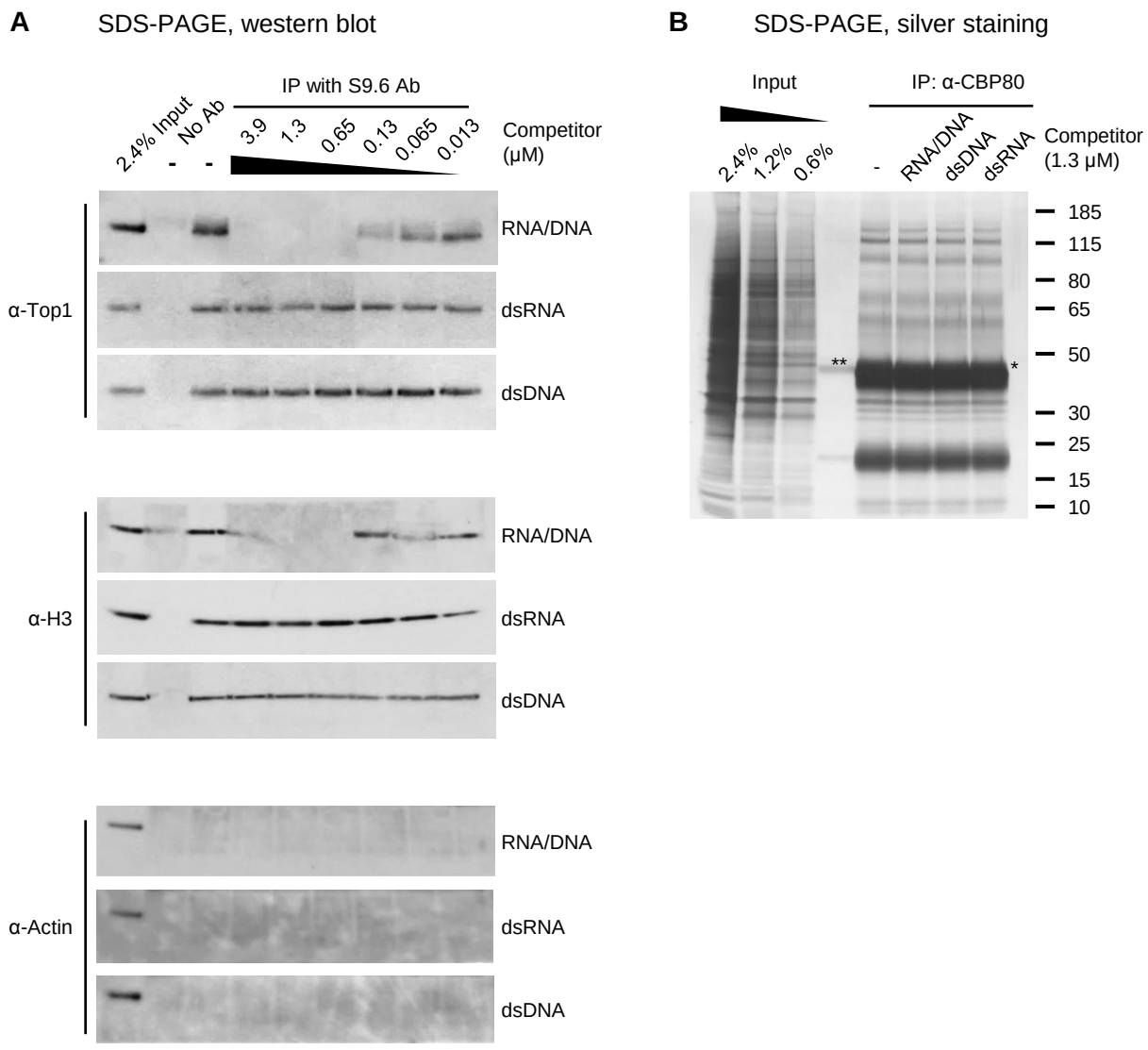


Figure 5.8. Additional validation of RNA/DNA hybrid IP method

A. Western blot analysis of protein samples from RNA/DNA hybrid IP with RNA/DNA, dsRNA or dsDNA competitors added at 3.9-0.013 μM during IP with S9.6 antibody. Western blot was probed with Top1 (top panel), histone H3 (middle panel) and Actin (bottom panel) antibodies. 2.4% of Input was loaded on the gel.

B. Silver staining of protein samples from CBP80 IP with nucleic acid competitors added at 1.3 μM. 2.4%, 1.2%, and 0.6% of Input was loaded on the gel. Molecular weight markers are indicated on the right side of the gel. * indicates the migration of the heavy chain from the S9.6 antibody and ** indicates migration of BSA used for pre-blocking of beads.

large dataset and control the false discovery rate, the p-values from the moderated t-test were corrected using the Benjamini-Hochberg method (Benjamini and Hochberg, 1995).

Employing the moderated t-test, 469 proteins were identified as significantly enriched in the RNA/DNA hybrid IP with Benjamini-Hochberg-corrected p-values <0.01. This collective set of 469 factors is designated as the RNA/DNA hybrid interactome' (Figure 5.10A, orange data points). All proteins significantly enriched in RNA/DNA hybrid IP ($p<0.01$) also exhibited a minimum $\log_2$-fold enrichment in RNA/DNA hybrid IP of greater than 2 (dotted vertical line in Figure 5.10A), while background proteins failed these thresholds (Figure 5.10A, grey data points, n=47).

The RNA/DNA hybrid interactome was further subdivided into three classes according to their corrected p-values, with the top 25% constituting class I, the next 50% class II and the bottom 25% class III (Figure 5.10B). Note that 332 proteins were identified in less than three replicates and were thus filtered out during MaxQuant analysis (Figure 5.10B, light blue). A representative list of proteins identified in all three classes of the RNA/DNA hybrid interactome and involved in various biological processes is given in Table 5.1

5.4.2 Bioinformatic analysis using the SINQ quantitation method

To test the robustness of the MaxQuant proteomics analysis, the mass spectrometry data were also processed with an independent approach using the normalized spectral index (SINQ) method (Trudgian *et al.*, 2011). While the MaxQuant analysis is based on quantitation of peptide ion intensities, quantitation in the SINQ method relies entirely on MS/MS spectra, also called spectral counting (Trudgian *et al.*, 2011). Therefore, the SINQ method represents a complementary analysis using a different set of parameters present in mass spectrometry data.

SINQ quantitation method was followed by a moderated t-test and p-value correction for multiple testing using the Benjamini-Hochberg method to identify significantly enriched proteins in the RNA/DNA hybrid IP (Benjamini and Hochberg, 1995; Smyth, 2004). The SINQ method identified fewer significantly enriched proteins (n=338) than the MaxQuant analysis (Figure 5.11A, orange data points). SINQ also exhibited a higher number of 'background' proteins not significantly enriched in all three replicates (n=102) (Figure 5.11A, grey data points). These observations may be

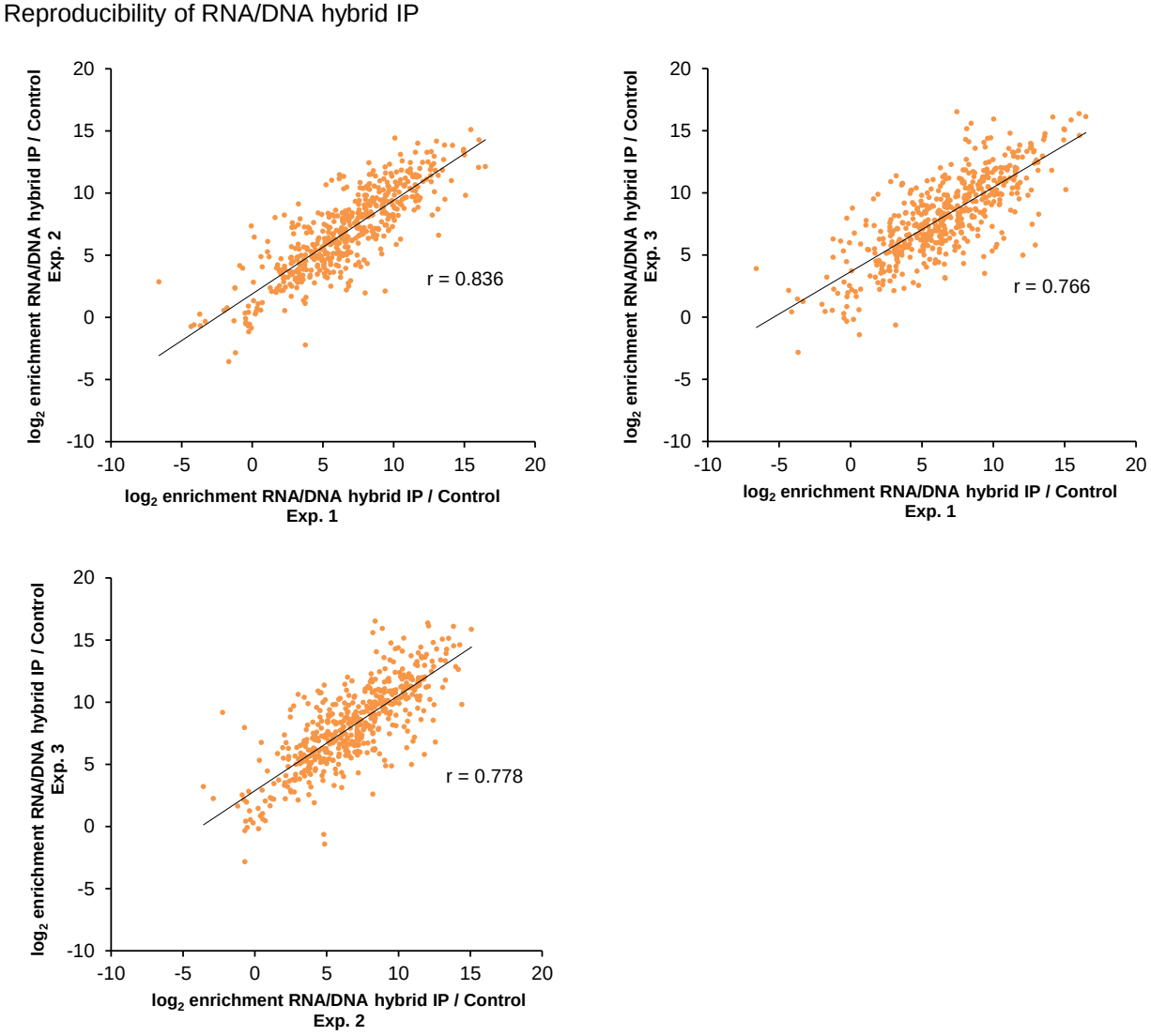


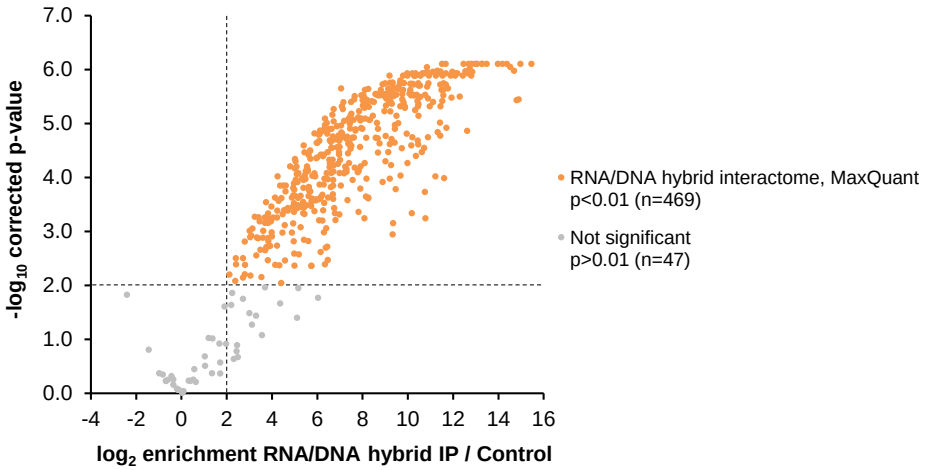
Figure 5.9. The RNA/DNA hybrid methodology is highly reproducible

Analysis of biological and technical reproducibility of RNA/DNA hybrid IP method. The panel shows the correlation between $\log_2$ enrichment RNA/DNA hybrid IP / Control IP, carried out in the presence of 1.3 μM synthetic RNA/DNA hybrid, of proteins quantified in three independent biological replicates by mass spectrometry. The black line represents a linear regression model fitting the data. r is Pearson's correlation coefficient.

attributed to the higher sensitivity of the MaxQuant approach due to its 'match between runs' feature, which allows transfer of peptide identifications and more robust quantification between replicates (Thakur *et al.*, 2011).

Among the proteins identified in both independent analyses, there was a strong correlation of relative enrichment in RNA/DNA hybrid IP/control IP in the three biological replicates between both methods (correlation coefficient r = 0.796) (Figure 5.11B). The higher sensitivity of the MaxQuant analysis is further evidenced by the fact that all proteins identified with the SINQ method were also identified using MaxQuant, while the SINQ method did not identify or failed to demonstrate statistically significant enrichment for 131 proteins. This was particularly clear for less enriched proteins, identified in classes II and III of the RNA/DNA hybrid interactome (Figure 5.11C). In conclusion, using two label-free quantitation pipelines, the robustness of the bioinformatical and statistical definition of the RNA/DNA hybrid interactome was confirmed.

A Definition of the RNA/DNA hybrid interactome using MaxQuant



B Classification of proteins identified in RNA/DNA hybrid interactome using MaxQuant

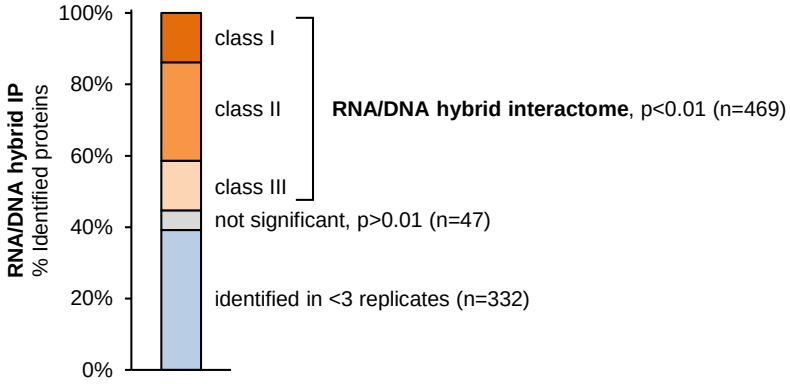


Figure 5.10. Bioinformatical definition of the RNA/DNA hybrid interactome using MaxQuant

A. Volcano plot displaying mass spectrometry results of three biological replicates of RNA/DNA hybrid IP experiments using MaxQuant software. Averaged $\log_2$ ratios between RNA/DNA hybrid IP and Control IP, carried out in the presence of 1.3 µM synthetic RNA/DNA hybrid, are plotted against their Benjamini-Hochberg corrected $-\log_{10}$ p-values calculated across all three biological replicates using a moderated t-test. Proteins significantly enriched in RNA/DNA hybrid IP with a corrected p-value <0.01 are shown in orange, and constitute the RNA/DNA hybrid interactome. Proteins identified with p-values >0.01 are depicted in grey. Dashed lines indicate the significance cutoffs ($\log_2$ enrichment >2 and $-\log_{10}$ p-value >2).

B. Classification of RNA/DNA hybrid interactome proteins. 469 Proteins enriched in RNA/DNA hybrid IP with a corrected p-value <0.01 were considered as the 'RNA/DNA hybrid interactome' (shades of orange). The RNA/DNA hybrid interactome was further subdivided into three classes according to the corrected p-values: top 25% (class I), middle 50% (class II), and bottom 25% (class III). 47 proteins were identified in 3 replicates but not significantly enriched (grey). 332 proteins were identified in less than three biological replicates (light blue).

Table 5.1. Representative list of proteins enriched in RNA/DNA hybrid interactome

Proteins identified in RNA/DNA hybrid interactome are involved in a wide range of cellular processes. A selection of significantly enriched proteins is grouped according to major functions and ranked according to their Benjamini-Hochberg-corrected p-value from the moderated t-test ('p-value') over three biological replicates. The 'Enrichment' column represents the mean $\log_2$ enrichment RNA/DNA hybrid IP / Control IP. The 'Class' denotes classification according to corrected p-value: top 25% (class I), middle 50% (class II), and bottom 25% (class III).

Protein	Description	Enrichment	p-value	Class
Transcription				
DHX9	ATP-dependent RNA helicase A	12.85	1.11E-06	I
DDX5	ATP-dependent RNA helicase DDX5	10.34	1.32E-06	I
ZNF326	DBIRD complex subunit ZNF326	10.09	2.83E-06	I
CTCF	Transcriptional repressor CTCF	8.58	2.88E-06	I
XRN2	5-3 exoribonuclease 2	9.82	4.10E-06	II
DDX1	ATP-dependent RNA helicase DDX1	8.29	4.94E-06	II
MED19	Mediator of RNA polymerase II transcription subunit 19	6.35	8.20E-06	II
TTF1	Transcription termination factor 1	8.33	8.67E-06	II
TAF15	TATA-binding protein-associated factor 2N	5.10	8.70E-05	II
TOP1	DNA topoisomerase 1	4.61	1.83E-04	II
SUPT16H	FACT complex subunit SPT16	6.23	2.12E-04	II
SSRP1	FACT complex subunit SSRP1	4.23	1.86E-03	III
Splicing and Processing				
HNRNPC	Heterogeneous nuclear ribonucleoproteins C1/C2	10.56	1.29E-06	I
SNRPD1	Small nuclear ribonucleoprotein Sm D1	10.95	1.32E-06	I
SYNCRIP	Heterogeneous nuclear ribonucleoprotein Q	11.69	1.91E-06	I
SNRPE	Small nuclear ribonucleoprotein E	8.77	3.04E-06	I
PRPF19	Pre-mRNA-processing factor 19	10.89	3.45E-06	I
HNRNPU	Heterogeneous nuclear ribonucleoprotein U	14.91	3.62E-06	II
HNRNPA1	Heterogeneous nuclear ribonucleoprotein A1	7.50	4.03E-06	II
TRA2A	Transformer-2 protein homolog alpha	9.88	4.97E-06	II
SRPK1	SRSF protein kinase 1	7.07	1.10E-05	II
U2AF1	Splicing factor U2AF 35 kDa subunit	6.33	1.38E-05	II
TARDBP	TAR DNA-binding protein 43	8.17	1.85E-05	II
SRSF9	Serine/arginine-rich splicing factor 9	7.51	2.60E-05	II

SRSF6	Serine/arginine-rich splicing factor 6	4.62	1.42E-04	II
SNRNP70	U1 small nuclear ribonucleoprotein 70 kDa	9.47	1.66E-04	II
U2AF2	Splicing factor U2AF 65 kDa subunit	4.89	2.38E-04	III
FIP1L1	Pre-mRNA 3-end-processing factor FIP1	5.16	2.40E-04	III
XAB2	Pre-mRNA-splicing factor SYF1	5.13	3.71E-04	III
SRSF1	Serine/arginine-rich splicing factor 1	3.03	9.84E-04	III
Epigenetic gene regulation				
H2AFY	Core histone macro-H2A.1	10.27	1.91E-06	I
WHSC1	Histone-lysine N-methyltransferase NSD2	9.78	2.42E-06	I
HP1BP3	Heterochromatin protein 1-binding protein 3	10.27	3.04E-06	I
HDAC2	Histone deacetylase 2	8.09	5.17E-06	II
BAZ1B	Tyrosine-protein kinase BAZ1B	8.97	5.97E-06	II
MBD2	Methyl-CpG-binding domain protein 2	7.33	1.77E-05	II
NAT10	N-acetyltransferase 10	9.15	3.42E-05	II
KMT2A	Histone-lysine N-methyltransferase 2A/MLL	6.17	3.82E-05	II
CDYL	Chromodomain Y-like protein	5.63	3.87E-05	II
BRD7	Bromodomain-containing protein 7	6.73	6.04E-05	II
CBX3	Chromobox protein homolog 3	7.80	8.06E-05	II
RUVBL2	RuvB-like 2 / INO80 complex subunit J	5.40	1.58E-04	II
DNMT1	DNA (cytosine-5)-methyltransferase 1	4.99	1.71E-04	II
HIST2H3A	Histone H3.2	10.79	5.79E-04	III
SUV39H1	Histone-lysine N-methyltransferase SUV39H1	4.01	1.25E-03	III
CBX5	Chromobox protein homolog 5	2.80	1.56E-03	III
SMARCA5	SWI/SNF-related matrix-associated actin-dependent regulator of chromatin subfamily A member 5	5.75	4.41E-03	III
DNA replication and repair				
TOP2A	DNA topoisomerase 2-alpha	7.82	7.76E-06	II
PRKDC	DNA-dependent protein kinase catalytic subunit	7.30	1.94E-05	II
PARP1	Poly [ADP-ribose] polymerase 1	6.47	3.42E-05	II
PARP2	Poly [ADP-ribose] polymerase 2	5.92	9.60E-05	II
PCNA	Proliferating cell nuclear antigen	4.37	1.42E-04	II
DDB1	DNA damage-binding protein 1	5.75	2.87E-04	III
MCM5	DNA replication licensing factor MCM5	5.00	3.71E-04	III

MCM3	DNA replication licensing factor MCM3	3.31	2.81E-03	III
Miscellaneous				
RBMX	RNA-binding motif protein, X chromosome	14.70	1.06E-06	I
MOV10	Putative helicase MOV-10	10.35	1.18E-06	I
SAFB	Scaffold attachment factor B1	10.56	3.23E-06	I
RBM15	Putative RNA-binding protein 15	8.18	4.42E-06	II
TARBP2	RISC-loading complex subunit TARBP2	5.69	3.24E-05	II
AURKB	Aurora kinase B	3.75	1.97E-03	III

5.4.3 Composition of the RNA/DNA hybrid interactome

To gain insights into R-loop protein biology, the RNA/DNA hybrid IP interactome was analysed using a range of bioinformatical tools. Gene ontology analysis for cellular compartment indicated that the RNA/DNA hybrid interactome is highly enriched for nuclear proteins (Figure 5.12A). Only a small, not statistically significant fraction (7%) of proteins in RNA/DNA hybrid IP were annotated as exclusively cytoplasmic, which underscores the purity of the nucleo-cytoplasmic fractionation used for RNA/DNA hybrid IP.

Next, the quantitative enrichment of different factors in the RNA/DNA hybrid interactome was analysed in more detail. For this, the relative abundance of proteins in the RNA/DNA hybrid interactome was compared to their relative abundance in HeLa cells using previously published data (Geiger *et al.*, 2012). This analysis demonstrated a lack of statistical correlation between protein abundance and their enrichment in the RNA/DNA hybrid interactome (Figure 5.12B), indicating that the RNA/DNA hybrid IP selects for a distinct functional subset of proteins.

As shown in Figure 5.13A, the RNA/DNA hybrid interactome was enriched for factors involved in RNA processing such as hnRNPs (38%), SR proteins and splicing factors (12%), DNA-binding proteins (15%), including Top1 and FACT complex subunits SUPT16H and SSRP1, chromatin-associated proteins (4%), including HP1γ and HDAC2, and histones (3%). Notably, various helicases were over-represented (8%), including RNA-directed helicases such as DDX5 and DHX9 and the replicative helicase MCM3. Overall, this protein class analysis indicated a very clear tendency towards nucleic-acid binding proteins, suggesting that the RNA/DNA hybrid IP detects

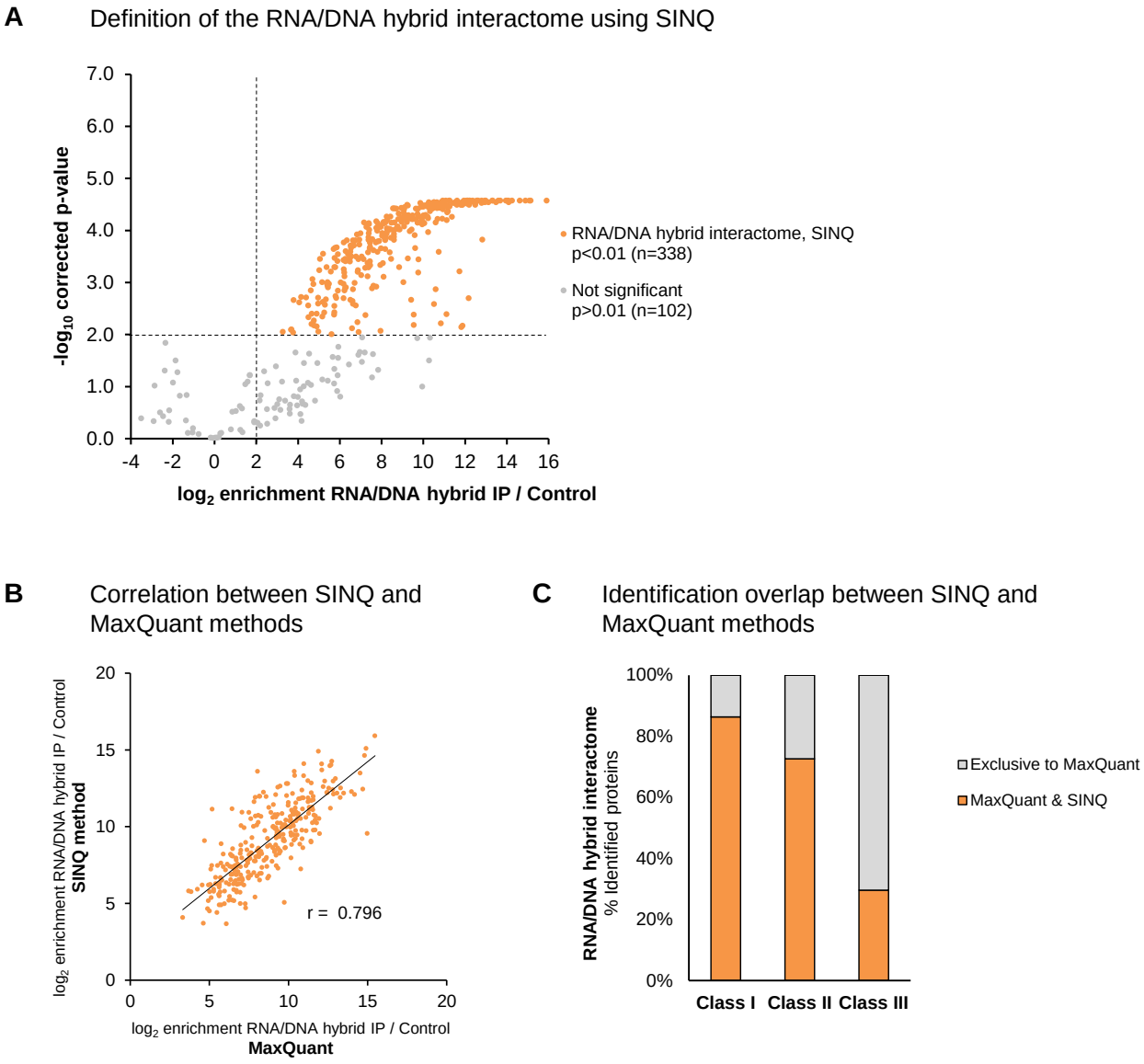


Figure 5.11. Validation of RNA/DNA hybrid interactome using the SINQ analysis method

A. Volcano plot displaying mass spectrometry results of three biological replicates of RNA/DNA hybrid IP experiments, using normalised spectral indexes (SINQ) quantification method (Trudgian *et al.*, 2011). Averaged $\log_2$ ratios between RNA/DNA hybrid IP and Control IP, carried out in the presence of 1.3 μM synthetic RNA/DNA hybrid, are plotted against their Benjamini-Hochberg corrected $-\log_{10}$ p-values calculated across all three biological replicates using a moderated t-test. Proteins significantly enriched in RNA/DNA hybrid IP with a corrected p-value <0.01 are shown in orange, and constitute the RNA/DNA hybrid interactome. Proteins identified with p-values >0.01 are depicted in grey. Dashed lines indicate the significance cutoffs ($\log_2$enrichment >2 and $-\log_{10}$ p-value >2).

B. Correlation between protein enrichment in RNA/DNA hybrid IP / Control IP of proteins quantified by both MaxQuant and SINQ methods. r is Pearson's correlation coefficient.

C. Analysis of proteins identified using MaxQuant or SINQ method. Proteins identified with both methods are shown in orange, those identified by MaxQuant alone in grey.

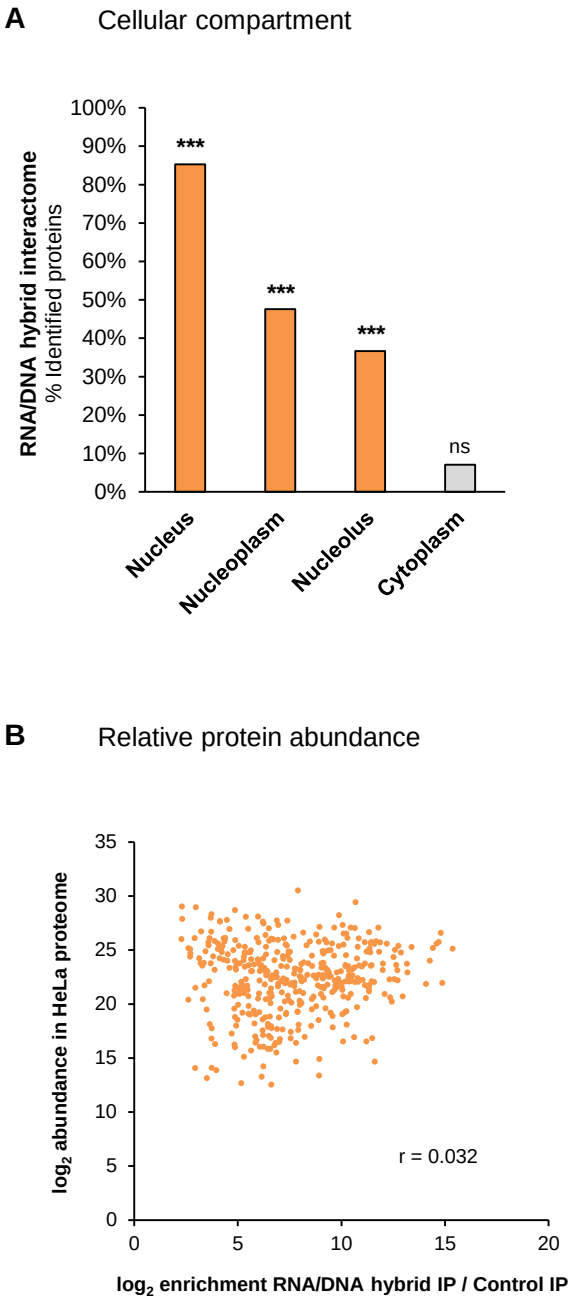


Figure 5.12. Analysis of cellular compartment enrichment and protein abundance

A. Cellular compartment analysis of RNA/DNA hybrid interactome ($p<0.01$). Asterisks (***) indicate highly significant enrichment of the depicted compartments in RNA/DNA hybrid IP as determined by Fisher's exact test (Benjamini-Hochberg corrected p-values of $2 \cdot 10^{-127}$, $1 \cdot 10^{-69}$, $8 \cdot 10^{-113}$, respectively).

B. Relative protein abundance in RNA/DNA hybrid interactome as compared to the total cellular protein abundance in HeLa cells, based on previously published data (Geiger *et al.*, 2012). Log_2 abundance of proteins in RNA/DNA hybrid interactome (x-axis) is plotted against their corresponding log_2 abundance in the total HeLa proteome (y-axis).

direct interactions rather than secondary interactions mediated by protein-protein binding. In line with this, protein domains functioning in nucleic acid binding, including RRM (ssRNA and ssDNA), KH (RNA and ssDNA) and SAP (DNA) domains were highly enriched in the RNA/DNA hybrid interactome (Figure 5.13B).

Recent studies have characterised the mRNA interactome in HeLa, HEK293 and mouse embryonic stem cells (Baltz *et al.*, 2012; Castello *et al.*, 2012; Kwon *et al.*, 2013). Comparing the RNA/DNA hybrid interactome to the corresponding HeLa mRNA interactome dataset, I found that despite an overlap between the two datasets (282 proteins), a significant part of the RNA/DNA hybrid interactome (187 proteins) is unique (Figure 5.14A). While factors common for the mRNA interactome and the RNA/DNA hybrid interactome are enriched for RNA-related functions (Figure 5.14B, top panel), factors unique to the R-loop interactome are involved in DNA and chromatin biology (Figure 5.14B, bottom panel).

Next, I investigated whether the RNA/DNA hybrid interactome is defined by any specific characteristics. An analysis based on chromatin probability index, which measures the likelihood of chromatin association for a particular protein (Kustatscher *et al.*, 2014), showed that the RNA/DNA hybrid interactome indeed exhibits an overall enrichment for proteins with high chromatin probability compared to the whole HeLa proteome (Figure 5.15A). However, proteins highly abundant in the RNA/DNA hybrid interactome exhibit a broad distribution (low and high) of chromatin probabilities, demonstrating that the RNA/DNA hybrid enriches for specific RNA/DNA hybrid-associated proteins over the chromatin background (Figure 5.15B).

The previous analyses have shown that the RNA/DNA hybrid interactome is enriched for RNA- and DNA-binding proteins with a high chromatin probability. Therefore, I analysed if the RNA/DNA hybrid interactome contains proteins with dual RNA/DNA-binding properties, as described before (Hudson and Ortlund, 2014). Interestingly, the RNA/DNA hybrid interactome is significantly enriched for this class of proteins compared to the HeLa proteome (Figure 5.15C), extending the initial observation that the RNA/DNA hybrid interactome unites RNA- and DNA-associated processes. The full list of dual RNA/DNA-binding proteins enriched in RNA/DNA hybrid IP is given in Table 5.2.

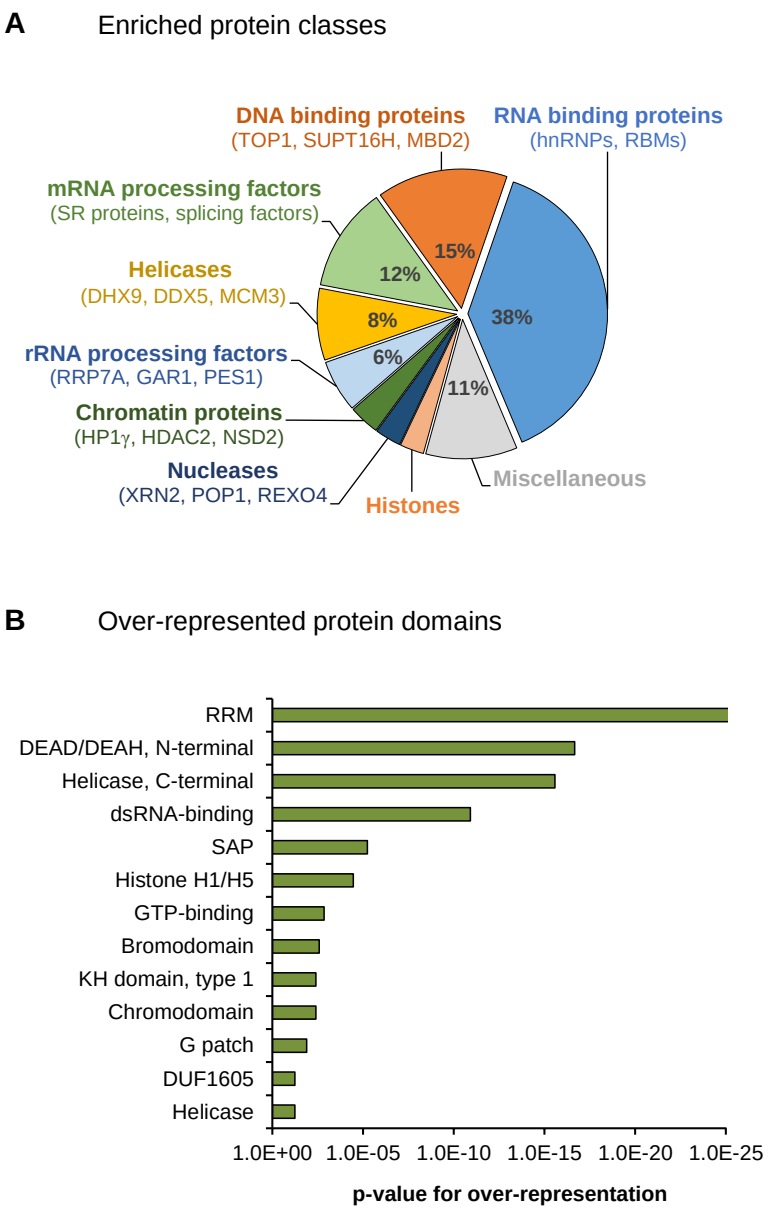


Figure 5.13. RNA/DNA hybrid interactome is enriched for RNA- and DNA-binding proteins

A. Protein classes over-represented in RNA/DNA hybrid interactome ($p<0.01$). All represented protein classes are statistically significantly enriched (Fisher's exact test, Benjamini-Hochberg corrected p- values <0.05). Representative proteins are indicated in brackets.

B. Protein domains over-represented in RNA/DNA hybrid IP. Analysis of enrichment of Pfam InterPro domains in RNA/DNA hybrid interactome using 'Enrichr' software (Chen *et al.*, 2013). Top thirteen significantly over-represented protein domains as determined by Fisher's exact test are shown and ranked according to their Benjamini-Hochberg corrected p-value.

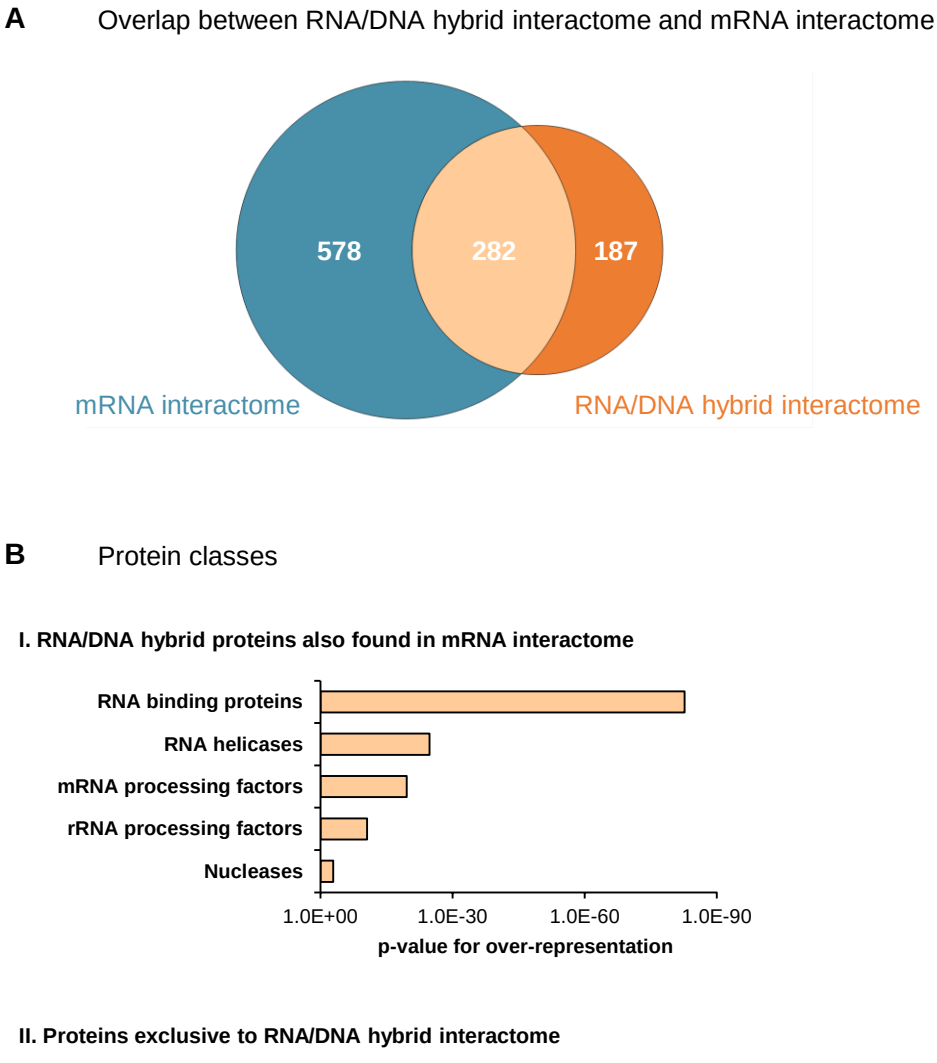


Figure 5.14. Overlap between RNA/DNA hybrid interactome and mRNA interactome

A. Venn diagram depicting overlap between human mRNA interactome (Castello *et al.*, 2012) and the RNA/DNA hybrid interactome ($p<0.01$) in HeLa cells.

B. Analysis of protein classes enriched in both RNA/DNA hybrid interactome and mRNA interactome (I), and proteins unique to RNA/DNA hybrid interactome (II). Benjamini-Hochberg corrected p-values on the x-axis indicate statistical significance of over-representation for the particular protein class.

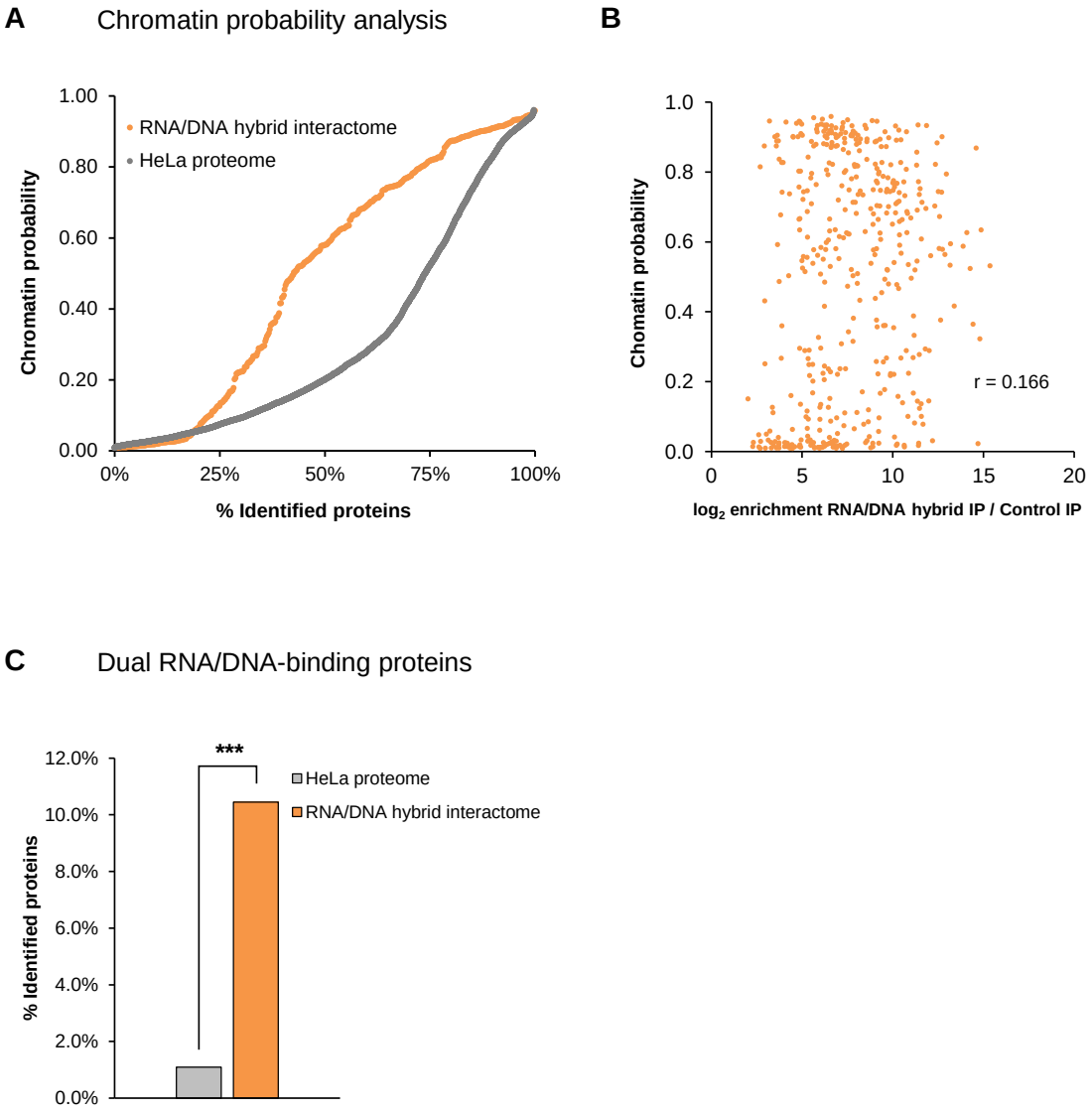


Figure 5.15. High chromatin probability and dual DNA/RNA binding

A. Analysis of chromatin probabilities of 7635 HeLa proteins ('HeLa proteome' as described in Kustatscher *et al.*, 2014) and proteins in the RNA/DNA hybrid interactome ($p<0.01$).

B. Relative protein abundance in RNA/DNA hybrid interactome compared to chromatin probability, based on previously published data (Kustatscher *et al.*, 2014). Abundance of proteins in RNA/DNA hybrid interactome (x-axis) is plotted against their corresponding chromatin probability (y-axis).

C. Enrichment for proteins known to bind both DNA and RNA in the RNA/DNA hybrid interactome ($p<0.01$) compared to HeLa proteome (Hudson and Ortlund, 2014), p-value of $1.6 \cdot 10^{-28}$, Fisher's exact test.

Table 5.2. RNA/DNA hybrid interactome proteins with dual RNA & DNA binding function

The analysis of dual RNA- and DNA-binding is based on data from Hudson and Ortlund, 2014. Asterisks indicate cases where the specific RNA binding has not been determined.

Protein	DNA binding	RNA binding
ADAR	Z-DNA	dsRNA, structured RNA, miRNA
BAZ2A	dsDNA	ssRNA, pRNA
CBX5	dsDNA	telomere RNA, heterochromatin mRNA
CENPC1	dsDNA	ssRNA, centromeric RNA
CEBPZ	dsDNA	mRNA
DDX3X	dsDNA	5'UTRs
DDX1	dsDNA	*
DDX5	dsDNA	*
DEK	dsDNA, junction DNA	mRNA
DHX36	G4	G4, mRNA
DHX9	ssDNA, dsDNA	ssRNA
DNMT1	non-CpG DNA	ecCEBPA RNA
DNTTIP2	ssDNA, dsDNA	mRNA
ENO1	dsDNA	(CUG)n, mRNA
FUS	ssDNA, dsDNA	ncRNA, mRNA
G3BP1	dsDNA	mRNA, dsRNA
HNRNPA1	telomere DNA, dsDNA	telomerase RNA, mRNA, miRNA
HNRNPA2B1	telomere DNA, dsDNA	telomerase RNA, mRNA, miRNA
HNRNPC	ssDNA	5'-UTR, 3'-UTR
HNRNPD	G4, dsDNA	ssRNA, mRNA
HNRNPK	ssDNA, dsDNA	mRNA, lincRNA
HNRNPL	ssDNA	3'UTR
HNRNPLL	DNA	RNA
HNRNPU	ssDNA, dsDNA	3'UTR, lincRNA, snRNA
HNRNPUL2	ssDNA, dsDNA	3'UTR, lincRNA, snRNA
ILF3	ssDNA, dsDNA	dsRNA, mRNA
ILF2	ssDNA, dsDNA	dsRNA, mRNA
NACA	ssDNA, dsDNA	rRNA, tRNA
NAT10	dsDNA, rDNA	snoRNA
NCL	dsDNA	mRNA
NKRF	dsDNA	mRNA
NONO	dsDNA	ssRNA, pre-mRNA
PARP1	ssDNA, dsDNA, DNA damage	ssRNA, mRNA
PCBP1	ssDNA	ssRNA, mRNA
PCBP2/3	ssDNA	ssRNA, mRNA
PGK1	ssDNA	mRNA
PTBP1	ssDNA	mRNA
RBM8A	dsDNA	miRNA, mRNA
RUVBL2	ssDNA, dsDNA	ssRNA

SAFB	scaffold DNA	mRNA
SFPQ	ssDNA, dsDNA	mRNA
SON	dsDNA	mRNA
SRSF1	ssDNA, telomere DNA	mRNA
SURF6	dsDNA	rRNA
TAF15	ssDNA	mRNA
TARDBP	ssDNA, dsDNA	mRNA, miRNA, ssRNA
TOP1	G4, dsDNA, telomere DNA	hairpin RNA, mRNA
YBX1	ssDNA, dsDNA, DNA damage	mRNA, ssRNA
YBX3	ssDNA, dsDNA	mRNA, ssRNA

5.4.4 Confirmation of RNA/DNA hybrid interactome candidates by western blotting

To validate RNA/DNA hybrid interactome proteins identified by mass spectrometry independently, RNA/DNA hybrid IP was followed by western blotting, probing for factors of all three RNA/DNA hybrid interactome classes (Figure 5.16). In particular, the presence of three helicases was confirmed, including DHX9, DDX5 and DDX1, which are all highly enriched in the RNA/DNA hybrid interactome (Table 5.1). The 5'-3' exonuclease XRN2 was also validated, previously implicated in R-loop-mediated transcriptional termination (Skourti-Stathaki *et al.*, 2011). Amongst proteins associated with gene repression such as HP1γ and histone deacetylase HDAC2, the histone modification H3K9me2 was also found in the RNA/DNA hybrid interactome, which may relate to association of R-loops with heterochromatin formation and gene repression (Skourti-Stathaki *et al.*, 2014). The RNA/DNA hybrid interactors Top1 and SRSF1 were also validated, in line with their established role in R-loop biology (Figures 5.5D, 5.6C, 5.8A) (Li and Manley, 2005; Tuduri *et al.*, 2009). The validated proteins span all three RNA/DNA hybrid interactome classes (class I: DHX9, DDX5; class II: DDX1, XRN2, Top1; class III: histone H3, SRSF1). This suggests that candidate RNA/DNA hybrid interactors with low relative enrichment in the mass spectrometry analysis (comprising class III) can represent valid interactors, warranting further investigation. As an important control, several abundant nuclear proteins absent from the mass spectrometry RNA/DNA hybrid IP analysis were investigated by western blotting. These proteins, including nuclear RNA-binding proteins Drosha and DGCR8, Lamin B1 and nuclear pore complex (NPC) protein were not found in the RNA/DNA hybrid IP (Figure 5.16). This further suggests that RNA/DNA hybrid IP indeed enriches for a specific subset of nuclear proteins.

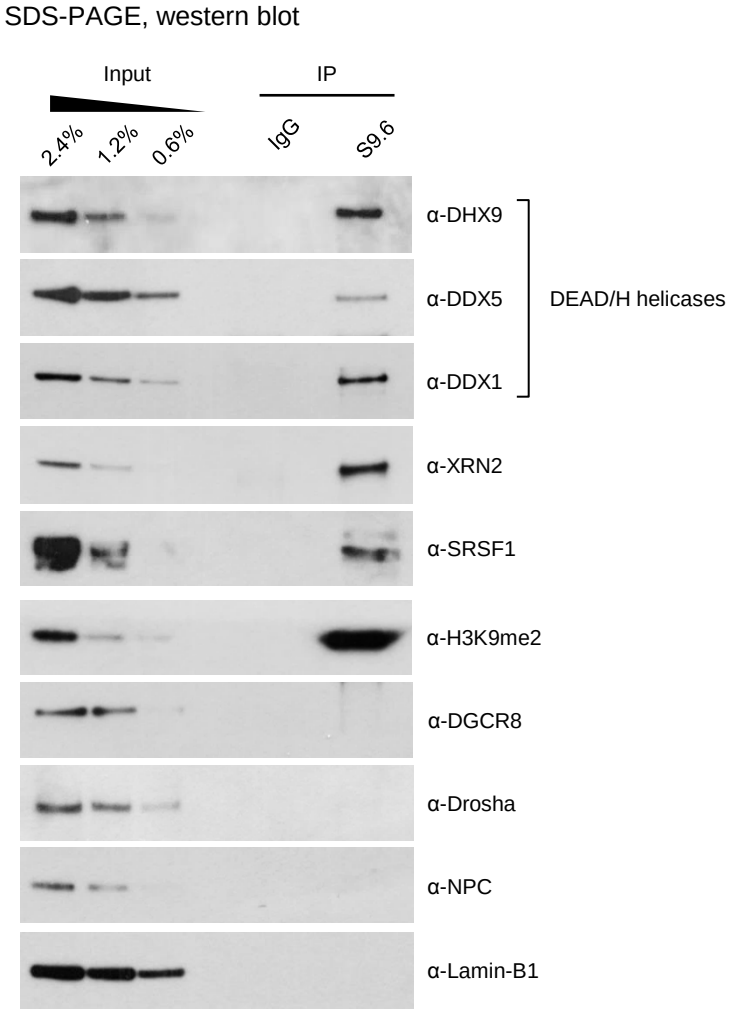


Figure 5.16. Validation of RNA/DNA hybrid interactome factors by western blotting

Confirmation of RNA/DNA hybrid-binding proteins detected by mass spectrometry using western blot analysis. SDS-PAGE was followed by western blotting with DHX9, DDX5, DDX1, XRN2, SRSF1 and H3K9me2 antibodies. DGCR8, Drosha, Nuclear Pore Complex (NPC) and Lamin B1 were negative controls not identified in RNA/DNA hybrid interactome ($p<0.01$). 2.4%, 1.2%, and 0.6% of Input was loaded on the gel.

5.5 Functional analysis of proteins interacting with RNA/DNA hybrids

5.5.1 The top interactome candidate DHX9 resolves RNA/DNA hybrids *in vivo*

One important benchmark of the power of the RNA/DNA hybrid interactome is its relevance in discovering new cellular functions and processes. Therefore, one of the most highly enriched candidates of the RNA/DNA hybrid interactome, DHX9 protein, was chosen for further investigation. DHX9 has been previously shown to possess RNA/DNA helicase activity *in vitro* (Chakraborty and Grosse, 2010, 2011). However, if DHX9 also interacts with RNA/DNA hybrids *in vivo* has not been investigated. Therefore, I performed a number of biochemical assays to establish if DHX9 associates with RNA/DNA hybrids *in vivo*. Indeed, treatment of nuclear extracts with benzonase ablated DHX9- RNA/DNA hybrid interaction (Figure 5.17A). Furthermore, DHX9 was not co-immunoprecipitated in the RNA/DNA hybrid IP in the presence of synthetic RNA/DNA hybrid competitor, while corresponding dsDNA or dsRNA competitors had no effect on the DXH9-RNA/DNA hybrid interaction (Figure 5.17B).

In an independent approach, DHX9 was immunoprecipitated and DHX9-bound nucleic acids were purified. As shown in Figure 5.17C (top panel), DHX9 was efficiently immunoprecipitated. Immunoprecipitation of the abundant Tubulin was carried out as a control for the specificity of RNA/DNA hybrid interaction with DHX9. Using a slot blot assay with S9.6 antibody, a specific RNA/DNA hybrid signal was observed in the DHX9 IP fraction, but was absent in the 'No antibody' sample and in samples immunoprecipitated with anti-Tubulin antibody (Figure 5.17C, bottom panel). These experiments further demonstrated that DHX9 interacts with RNA/DNA hybrids *in vivo*.

Since RNA/DNA hybrids are thought to be mainly formed co-transcriptionally, it was reasoned that inhibition of transcription would prevent co-immunoprecipitation of DHX9 in the RNA/DNA hybrid IP. Transcription was inhibited by treatment of cells with 5 μg/ml actinomycin D for 6 hours, which was previously shown to significantly reduce pre-mRNA levels in HeLa cells (Gromak *et al.*, 2013). Following this treatment, nuclear RNA/DNA hybrids were dramatically reduced, in line with their dependence on ongoing transcription (Figure 5.17D). Although transcriptional inhibition had no effect on overall nuclear DHX9 levels (Figure 5.17E, Input lane),

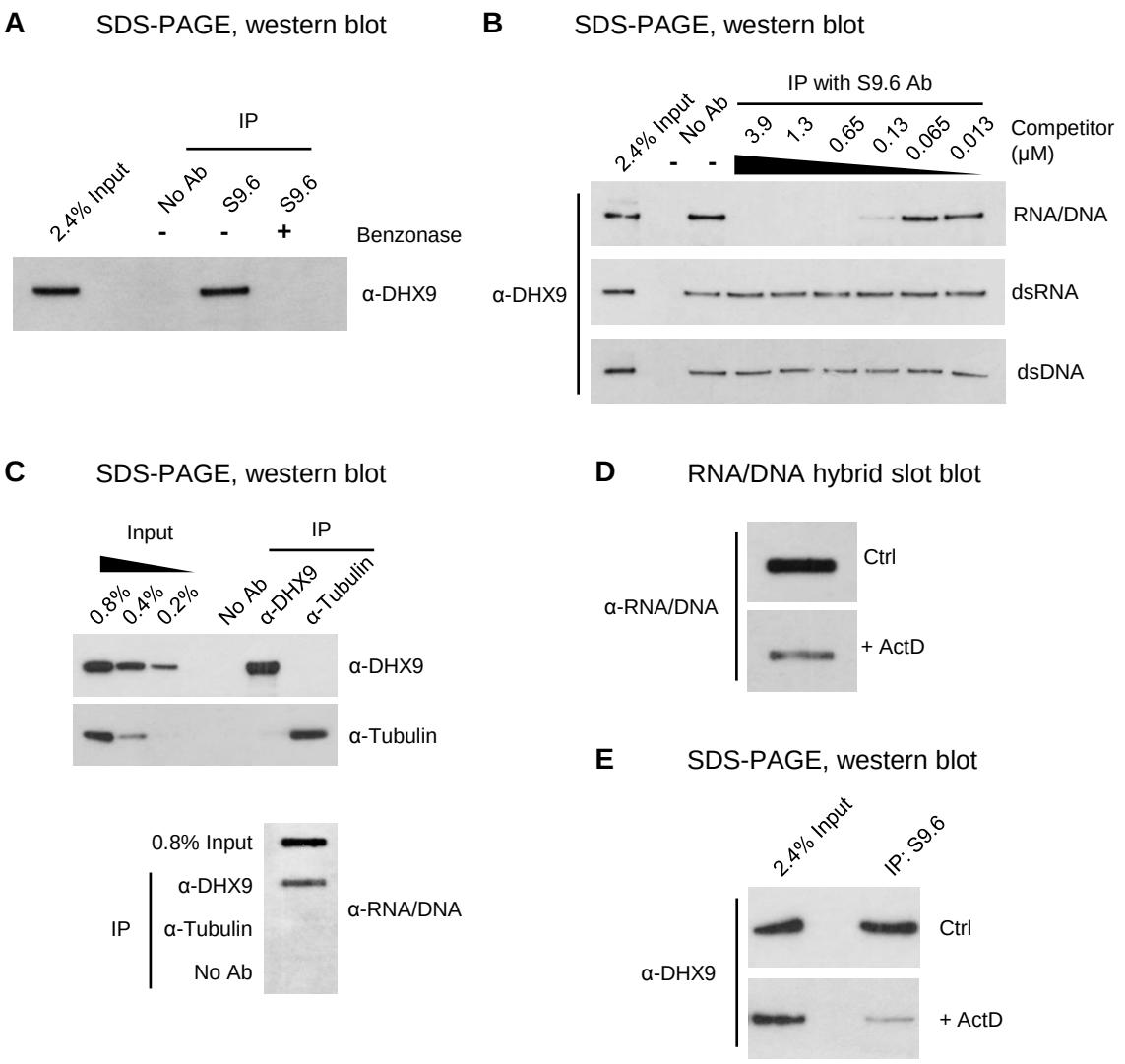


Figure 5.17. DHX9 interacts with RNA/DNA hybrids *in vivo*

A. Western blot analysis of protein samples from RNA/DNA hybrid IP. Nuclear extracts were treated with benzonase (1 U/μl) for 30 min prior to immunoprecipitation with S9.6 antibody. Western blot was probed with DXH9 antibody.

B. Western blot analysis of protein samples from RNA/DNA hybrid IP with RNA/DNA, dsRNA or dsDNA competitors added at 3.9-0.013 μM. Western blot was probed with DHX9 antibody. 2.4% of Input was loaded on the gel.

C. RNA/DNA hybrids are associated with DHX9 protein *in vivo*. IP in HeLa cells was carried out with DHX9 and tubulin antibodies. Immunoprecipitated proteins were probed on western blot (top panel) and nucleic acids were probed on RNA/DNA hybrid slot blot with S9.6 antibody (bottom panel).

D. RNA/DNA hybrid slot blot probed with S9.6 antibody. Nucleic acids were purified from Hela cells, treated with 5 μg/ml actinomycin D (or 96% EtOH, 'Ctrl') for 6 hours.

E. Western blot analysis of protein samples from RNA/DNA hybrid IP in HeLa cells, treated with 5 μg/ml actinomycin D for 6 hours. Nuclear extracts were immunoprecipitated with S9.6 antibody. Western blot was probed with DXH9 antibody.

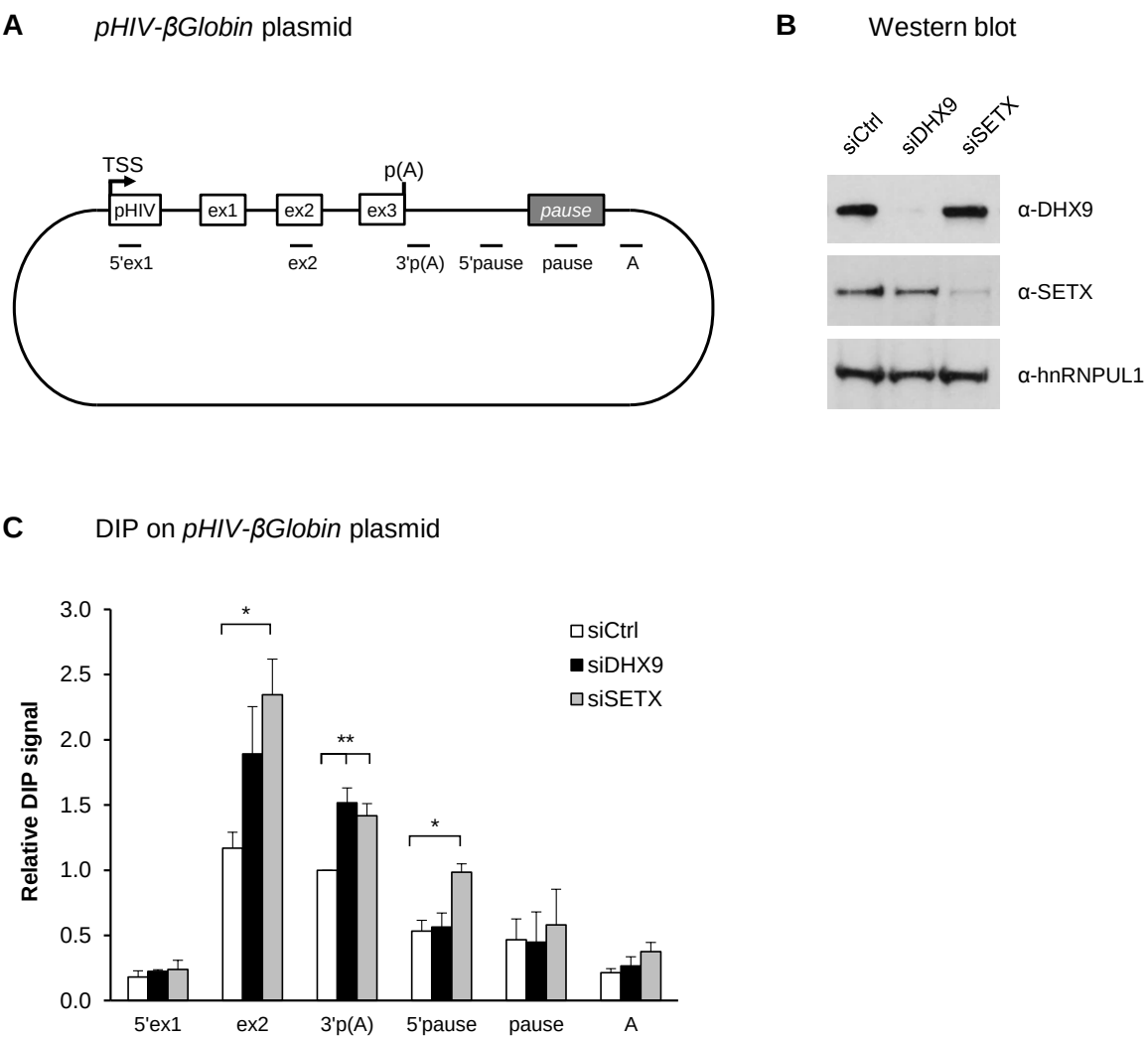


Figure 5.18. Loss of DHX9 leads to RNA/DNA hybrid accumulation on reporter plasmid *in vivo*

A. Diagram depicting the *pHIV-β-Globin* construct. The human immunodeficiency virus (HIV) promoter (pHIV) and the *β-Globin* exons are depicted as white boxes. The grey box shows the *β-Actin* pause element required for efficient transcription termination. Positions of amplicons used for DIP analysis are shown below the diagram.

B. Western blot of 60 µg whole-cell extracts from HeLa cells transfected with a non-targeting siRNA (siCtrl), or siRNAs targeting DHX9 (siDHX9) or SETX (siSETX). Western blot was probed with antibodies against DHX9, SETX, and hnRNPUL1 (which serves as loading control). This western blot was performed by Dr Maiken Kristiansen.

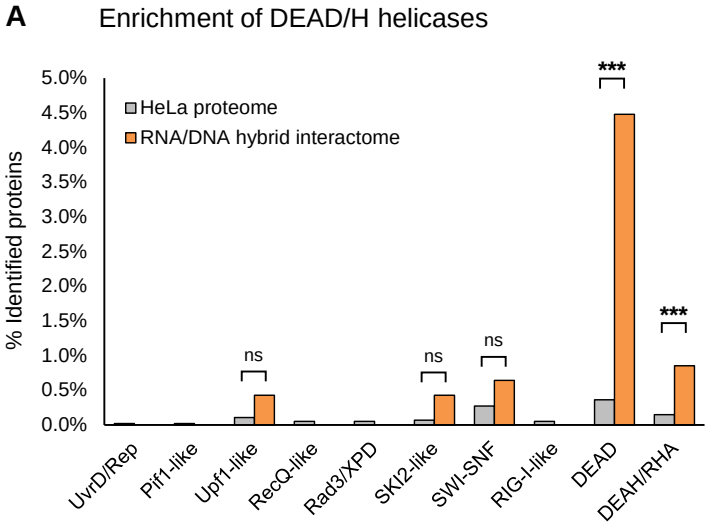
C. DIP analysis on the *pHIV-β-Globin* construct in HeLa cells transfected with a non-targeting siRNA (siCtrl), or siRNAs targeting DHX9 (siDHX9) or SETX (siSETX). All cells were co-transfected with 1 µg *pHIV-β-Globin* plasmid and 0.1 µg pTAT plasmid during the 2nd siRNA hit. Values are relative to probe 3'p(A) in siCtrl cells.

the amount of DHX9 co-immunoprecipitated in RNA/DNA hybrid IP was significantly reduced upon actinomycin D treatment (Figure 5.17E, IP lane). Together, these results suggested that DHX9 may be involved in RNA/DNA hybrid-mediated transcriptional regulation of gene expression.

To further investigate the ability of DHX9 to resolve RNA/DNA hybrids *in vivo*, the reporter plasmid *pHIV-βGlobin* was employed (Dye and Proudfoot, 1999; Skourti-Stathaki *et al.*, 2011). In this construct, the expression of the *β-Globin* gene is driven by the strong human immunodeficiency virus (HIV) promoter (pHIV), which is activated by a co-transfected plasmid, encoding transactivator of transcription protein (pTAT). In addition to this, the plasmid contains the *β-Actin* pause element inserted downstream of the pA signal to mediate efficient RNA Pol II transcriptional termination (Figure 5.18A) (Dye and Proudfoot, 1999; Skourti-Stathaki *et al.*, 2011). Previously it was shown that RNA/DNA hybrid levels on this construct were sensitive to depletion of SETX protein in HeLa cells (Skourti-Stathaki *et al.*, 2011). In line with its previously established function in RNA/DNA hybrid resolution, cells depleted of SETX, which serves as a positive control, showed significantly increased RNA/DNA hybrid levels across the *pHIV-βGlobin* reporter construct (Figure 5.18B, C). Using siRNA-mediated knockdown, DHX9 protein was efficiently depleted in HeLa cells, as confirmed by western blotting (Figure 5.18B). As demonstrated by DIP, following DHX9 knockdown, RNA/DNA hybrid levels were increased in the gene body of the *pHIV-βGlobin* reporter (probes ex2, 3'p(A)), while negative control regions upstream and downstream were unaffected (probes 5'ex1, A) (Figure 5.18C). These results suggested that DHX9 is involved in resolving RNA/DNA hybrids *in vivo*. Furthermore, experiments investigating the effects of DHX9 depletion on endogenous genes performed by Dr M. Kristiansen in the group suggested an involvement of DHX9 in resolution of RNA/DNA hybrids and termination of RNA Pol II transcription *in vivo* (discussed in Chapter 5.6.2).

5.5.2 RNA/DNA hybrid interactome helicases are overexpressed in human cancers

Apart from their involvement in transcriptional termination, R-loops can trigger genome instability, a hallmark of cancer (Aguilera and Garcia-Muse, 2012; Groh and Gromak, 2014; Hamperl and Cimprich, 2014; Skourti-Stathaki and Proudfoot, 2014; Sollier and Cimprich, 2015).



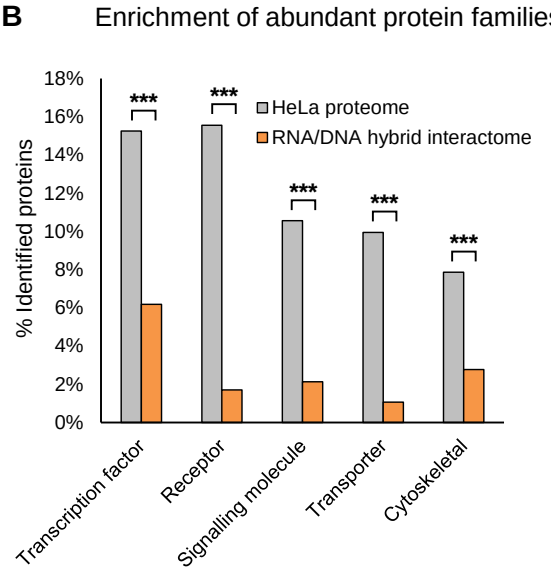


Figure 5.19. Enrichment of DEAD/H helicases in RNA/DNA hybrid interactome

A. Enrichment analysis of DEAD and DEAH/RHA helicases in the RNA/DNA hybrid interactome ($p<0.01$).

B. Enrichment analysis of abundant protein families in the RNA/DNA hybrid interactome ($p<0.01$). A-B. Asterisks (***) indicate statistically significant enrichment based on Fisher's exact test with Benjamini-Hochberg-corrected p-values $p<0.0001$.

Interestingly, in addition to DHX9, other DEAD and DEAH/RHA helicases were significantly enriched in the RNA/DNA hybrid interactome (Figure 5.19A). To exclude the possibility that this enrichment is simply due to the abundance of DEAD/H helicases, enrichment analysis of a number of abundant protein families was performed. These protein classes were significantly under-represented in the RNA/DNA hybrid interactome (Figure 5.19B). In particular, the RNA/DNA hybrid interactome was depleted for transcription factors and receptors (a group also comprising nuclear receptors), suggesting that the RNA/DNA hybrid IP enriches for a specific subset of protein classes. Together, these results suggested a potential role of DEAD/H helicases in RNA/DNA hybrid and R-loop processes.

In light of this enrichment, I therefore explored if DHX9 and other helicases may be linked to human cancer. Using the publically available COSMIC cancer dataset (Forbes et al., 2015), copy number variation analysis showed that DEAD/H helicases identified in the RNA/DNA hybrid interactome are frequently genetically amplified in a range of human cancers, similar to known oncogenes SKP2 and MDM2 (Figure 5.20A). In line with these genetic amplifications, most helicases identified in the RNA/DNA hybrid interactome, including DHX9 and DDX5, showed significant mRNA overexpression in a range of cancers compared to matched controls, based on ONCOMINE database (Rhodes et al., 2007) (Figure 5.20B). With cancer expression profiles similar to oncogenes, helicases identified in the RNA/DNA hybrid interactome may therefore have a role in promoting cancer progression or represent a mechanism to supress excessive genome instability.

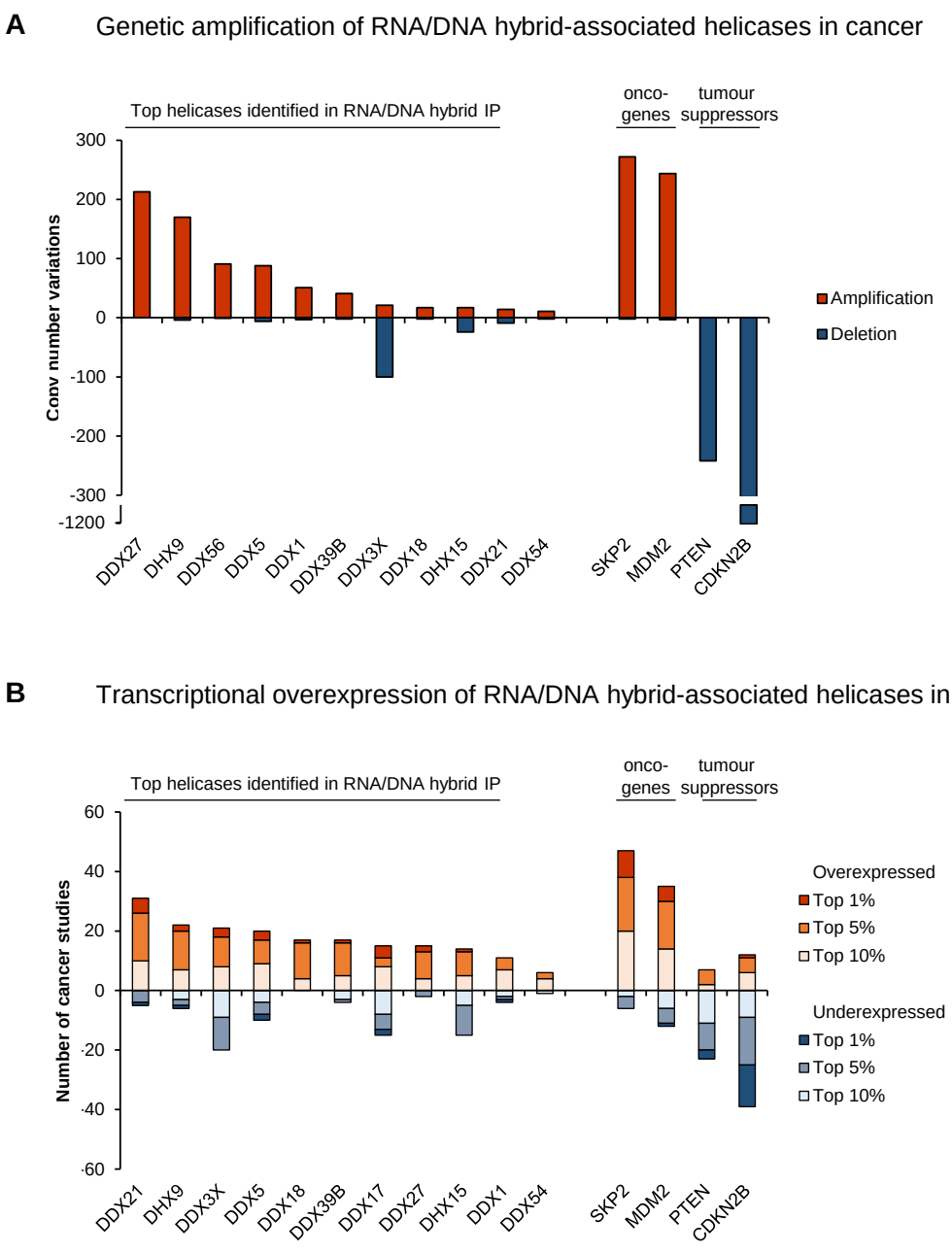


Figure 5.20. RNA/DNA hybrid interactome DEAD/H helicases are overexpressed in cancers

A. Analysis of genetic alterations of top RNA/DNA hybrid-interacting DEAD/H helicases in cancer, based on copy number variations from COSMIC database (Forbes *et al.*, 2015). Amplifications (red) and deletions (blue) are shown. *SKP2* and *MDM2* are oncogenes and *PTEN* and *CDKN2B* are tumour suppressors not identified in RNA/DNA hybrid IP shown as comparison.

B. Transcriptional expression changes of top RNA/DNA hybrid-interacting DEAD/H helicases in cancer versus matched normal tissues based on ONCOMINE database (Rhodes *et al.*, 2007). The y-axis shows the number of cancer studies, in which the depicted helicases are among the top 1%, 5% or 10% over-expressed (shades of orange) and under-expressed (shades of blue) genes. *SKP2* and *MDM2* are oncogenes and *PTEN* and *CDKN2B* are tumour suppressors not identified in RNA/DNA hybrid IP shown as comparison.

5.6 Discussion

5.6.1 Discovery of novel RNA/DNA hybrid interacting proteins

Despite advances in the R-loop field in recent years, our knowledge of direct interactions between RNA/DNA hybrids or R-loops and proteins *in vivo*, and the significance of such interactions in human disease remains limited. This work describes an unbiased mass spectrometry approach to identify proteins binding RNA/DNA hybrids in HeLa cells, using S9.6 antibody, which specifically recognises RNA/DNA hybrids with high affinity (Figure 5.5). This method was optimised to ensure purification of RNA/DNA hybrid-binding factors from the chromatin fraction of non-crosslinked cells (Figures 5.1–5.4). The identified RNA/DNA hybrid interactome of HeLa cells comprises of 469 proteins, based on stringent statistical parameters. Several lines of evidence demonstrate that binding of these factors to RNA/DNA hybrids is dependent on transcription and the presence of RNA/DNA hybrids (Figures 5.6–5.8, 5.17D, E).

Mass spectrometry analysis of RNA/DNA hybrid-interacting proteins demonstrates that in addition to a number of anticipated proteins with established function in R-loop biology, including RNA Pol II, Top I and SETX, novel RNA/DNA hybrid-binding candidates could also be identified (Figures 5.5D, 5.16, Table 5.1). In particular, the RNA/DNA hybrid interactome exhibits a significant enrichment for RNA-and DNA-binding proteins, RNA processing factors, helicases, histone modifiers and proteins involved in DNA damage and repair (Figures 5.13A, 5.15A, Table 5.1). The RNA/DNA hybrid interactome is unique and it differs from the mRNA interactome (Castello et al., 2012), being highly enriched for proteins with dual DNA and RNA-binding properties (Figures 5.14, 5.15, Table 5.2). The RNA/DNA hybrid interactome contains many proteins involved in RNA splicing and processing, including SR proteins (SRSF2, SRSF1), hnRNPs (hnRNPA1, hnRNPU, hnRNPQ, hnRNPC), and XRN2. It has previously been shown by genetic assays that loss of factors involved in these processes leads to accumulation of RNA/DNA hybrids and R-loops (Aguilera and Garcia-Muse, 2012; Skourti-Stathaki and Proudfoot, 2014). Data from this work now suggest that in unperturbed cells, these processes are tightly and directly linked to RNA/DNA hybrids.

5.6.2 DHX9 is an RNA/DNA hybrid helicase *in vivo*

Mass spectrometry analysis identified DHX9 helicase amongst the most enriched proteins of the RNA/DNA hybrid interactome (Table 5.1, Figure 5.16). The specificity of the interaction of DHX9 with RNA/DNA hybrids was confirmed using enzymatic and oligonucleotide competition experiments (Figure 5.17A–B). Reciprocal co-immunoprecipitation experiments verified that DHX9 interacts with RNA/DNA hybrids *in vivo* (Figures 5.17C–D). Moreover, the increase of RNA/DNA hybrids on a reporter construct in cells depleted of DHX9 suggested that the interaction of DHX9 with RNA/DNA hybrids is biologically relevant and that DHX9 may be involved in a range of RNA/DNA hybrid-mediated processes (Figure 5.18).

To uncover more mechanistic details of this RNA/DNA hybrid-protein interplay, work carried out by Dr M. Kristiansen in this group focused on further functional analysis of DHX9 helicase. Dr Kristiansen observed that endogenous DHX9 protein localizes to the nucleus, where it interacts with RNA/DNA hybrids (pers. comm.). Furthermore, she demonstrated that DHX9 binds to terminator regions of endogenous human *β-Actin* and *γ-Actin* genes where it acts to resolve RNA/DNA hybrids, essential for efficient transcriptional termination. This phenotype is similar to the cells depleted for SETX protein, previously shown to be involved in the process of transcriptional termination (Skourti-Stathaki *et al.*, 2011). In line with these results, DHX9 has been implicated in various aspects of RNA metabolism. It interacts with RNA Pol II (Padmanabhan et al., 2012), regulates binding of other factors to RNA Pol II, including BRCA1 and CBP (Anderson et al., 1998; Nakajima et al., 1997) and mediates efficient transcription of genes regulated by EGF receptor and NF-κB (Huo et al., 2010; Tetsuka et al., 2004). Findings from my and Dr Kristiansen's work now suggest that some cellular functions of DHX9 may relate to its ability to resolve RNA/DNA hybrids.

The function of DHX9 helicase in RNA/DNA hybrid resolution and transcriptional termination *in vivo* provides a biological validation to the RNA/DNA hybrid-associated function of proteins identified in the RNA/DNA hybrid interactome. While helicases have been implicated in different aspects of RNA metabolism and gene expression, it is not clear if their biological functions are associated with RNA/DNA hybrid resolution. Currently, for many proteins with documented *in vitro* RNA/DNA helicase activity (e.g. Pif1, the MCM complex), *in vivo* evidence is generally still

lacking (Boule and Zakian, 2007; Shin and Kelman, 2006). Interestingly, another top candidate helicase in the RNA/DNA hybrid interactome, DDX21, is required for expression of RNA Pol II and RNA Pol I-transcribed genes (Calo et al., 2015), but it is currently unclear if RNA/DNA hybrids play a role in this process. This work shows that it will be important to delineate the functions of helicases in RNA/DNA hybrid resolution and transcriptional regulation.

5.6.3 RNA/DNA hybrid-interacting proteins and cancer

This study highlights another important aspect of proteins interacting with RNA/DNA hybrids *in vivo*: their potential role in genome stability (Aguilera and Garcia-Muse, 2012; Sollier and Cimprich, 2015). The RNA/DNA hybrid interactome uncovers DNA repair proteins such as Poly ADP-ribose polymerase 1 (PARP1), DNA-PK catalytic subunit (PRKDC) and DNA damage-binding protein 1 (DDB1), which may be involved in resolving RNA/DNA hybrid-induced DNA damage. Furthermore, helicases are frequently dysregulated in cancer (Robert and Pelletier, 2013). Indeed based on ONCOMINE analysis, helicases identified in the RNA/DNA hybrid IP are genetically amplified and overexpressed in cancer tissues, closely mirroring the behaviour of oncogenes (Figures 5.19–5.20) (Rhodes *et al.*, 2007). DHX9 in particular is overexpressed in lung cancer tissues (Wei et al., 2004). DHX9 promotes upregulation of MDR1, the multidrug-resistant gene in cancer cells (Zhong and Safa, 2004), and it is crucial for drug resistance in lymphoma cells (Mills et al., 2013). Furthermore, interaction of DHX9 with oncogenic fusion protein EWS-FLI1 was shown to be essential for tumour growth (Erkizan et al., 2009; Toretsky et al., 2006). Future work needs to determine how the ability of DHX9 and other helicases to resolve R-loops may drive or support cancer progression.

5.6.4 RNA/DNA hybrid interactors as epigenetic modifiers

In addition to a role of RNA/DNA hybrid interactome factors in transcriptional and post-transcriptional processes, a significant enrichment of epigenetic modifiers of transcription suggests that an additional layer of gene regulation could be facilitated by RNA/DNA hybrids (see Table 5.1). Interestingly, RNA/DNA hybrid IP shows an enrichment for the repressive histone mark H3K9me2, the H3K9 histone methyl transferase SUV39H1 and heterochromatin protein HP1γ, which binds

H3K9me2. These factors constitute a repressive epigenetic module, which has recently been implicated in R-loop-mediated heterochromatin formation at termination regions (Skourti-Stathaki et al., 2014) and over expanded triplet repeats, associated with Friedreich ataxia disorder as discussed in Chapters 3 and 4 of this work (Groh *et al.*, 2014a). In conclusion, this work provides the first proteomic characterisation of factors interacting with R-loops *in vivo*, offering a new perspective on the multitude of cellular RNA/DNA hybrid and R-loop functions. Thus, it constitutes a powerful resource to study R-loop biology in health and disease.

5.6.5 The depth of the RNA/DNA hybrid interactome

Genome- and proteome-wide genetic or interaction screens offer the unique advantage of being comparably unbiased, and they are usually only limited by available technology. This increased discovery potential comes with three caveats, which have to be addressed before any biologically meaningful interpretation of the screen data can be obtained.

Firstly, the experimental setup has to allow the internal definition of 'hits' versus 'non-hits'. This usually means that the technical and biological variability has to be taken into consideration. In proteomic studies, the current widely accepted standard of best practice is to perform at least three biological replicates of the whole experiment (Prof S. Mohammed, pers. comm.). Using previously published bioinformatical pipelines (Castello *et al.*, 2012; Kwon *et al.*, 2013), the RNA/DNA hybrid interactome was thus defined as the set of proteins consistently identified and highly enriched in RNA/DNA hybrid IP over three biological replicates (see Chapter 5.4.1).

The second consideration is that the screen methodology should be validated using an independent read-out. Here, western blotting using specific antibodies recognising proteins identified by mass spectrometry and negative controls was employed. This approach confirmed the overlap between immunoblotting and mass spectrometry read-outs. However, it should be noted that some proteins were detected using western blotting, including SETX and RNA Pol II but not in the mass spectrometry data set (Figure 5.5D). Furthermore, some of the expected candidate RNA/DNA hybrid interactors including RNase H1 and RNase H2A were not identified. This raises the question of the proteome coverage and depth of the currently employed RNA/DNA hybrid IP method. For proteins

such as SETX, RNase H1 and RNase H2A, their low cellular abundance may impede detection by mass spectrometry even though they may be detected by the more sensitive western blotting procedure (Boisvert *et al.*, 2012). In addition, as active enzymes, their dynamic association and dissociation with the RNA/DNA hybrid substrate may be too transient to allow efficient immunoprecipitation in the RNA/DNA hybrid IP carried out under non-crosslinking conditions. With increased sequencing depth and improved RNA/DNA hybrid purification methodology, it is possible that additional RNA/DNA hybrid and R-loop interactors will be identified in the future.

The third consideration is that the biological relevance of screen results needs to be confirmed. This can be done by an extensive secondary screen, often using siRNA-mediated knockdown of candidate hits with an independent high-throughput read-out (Castello *et al.*, 2012). Alternatively, studies have focused on exploring the biology of top candidates *in vivo*. In this work, and in collaboration with Dr Maiken Kristiansen in the group, the latter approach was pursued. As described in Chapters 5.5.1 and 5.5.2, the biology of the highly enriched DHX9 protein was investigated biochemically and functionally.

In light of the large set of proteins enriched in the RNA/DNA hybrid interactome performing a range of biological functions, it is clear that defining the RNA/DNA hybrid interactome has only been the first step to enhance our understanding of R-loop biology. Further studies will focus on proteins identified in this work and unravel their interplay with R-loops and their function in health and disease.

Chapter 6

Conclusions and outlook

6.1 Achievement of initial aims

6.1.1 The role of R-loops in trinucleotide expansion diseases

The first main aim of this work was to investigate mechanisms involved in the transcriptional repression causing the trinucleotide repeat expansion diseases Friedreich ataxia and fragile X syndrome. The working hypothesis was that unusual RNA/DNA structures such as RNA/DNA hybrids in R-loops formed by expanded repeats may play a central role in the pathomechanism of these diseases. In the present work I have characterised the transcriptional and epigenetic regulation of the *FXN* gene in Friedreich ataxia. Furthermore, I have demonstrated that RNA/DNA hybrids are enriched at expanded GAA and CGG trinucleotide repeats *in vivo*. My results show that RNA/DNA hybrids promote heterochromatin formation and transcriptional repression of the *FXN* gene in FRDA and suggest that they may play a similar role in *FMR1* gene silencing in FXS and potentially other trinucleotide expansion diseases.

6.1.2 Characterisation of the RNA/DNA hybrid interactome

The second major aim of this project was to identify novel proteins binding to RNA/DNA hybrids *in vivo*. The motivation for this was that identifying and characterising factors interacting with RNA/DNA hybrids and R-loops would allow gaining insight into R-loop biology in health and disease. In this work, I have established the RNA/DNA hybrid IP, a method to co-immunoprecipitate proteins interacting with RNA/DNA hybrids in HeLa cells. I have combined RNA/DNA hybrid IP with quantitative mass spectrometry-based proteomic and bioinformatic analysis to identify and characterise the RNA/DNA hybrid interactome. Furthermore, I validated top candidates biochemically and demonstrated that the top R-loop interactor, RNA/DNA helicase DHX9 resolves RNA/DNA hybrids *in vivo*.

6.2 Key findings and novelty

6.2.1 R-loops promote gene repression in trinucleotide repeat expansion diseases

Transcriptional characterisation of the FXN gene in FRDA

Since the identification of the GAA expansion in the *FXN* gene as the cause of FRDA, several studies have investigated the transcriptional regulation of the *FXN* gene (Chapter 1.3.6) (Campuzano *et al.*, 1996). The results from the present work confirm previous studies implicating transcription initiation and elongation defects as causes for reduced *FXN* gene expression (Chapters 3.1.1–3.1.2). Previously, most studies have investigated steady-state levels of *FXN* mRNA or relied on indirect read-outs of transcriptional activity such as histone modifications (Al-Mahdawi *et al.*, 2008; Bidichandani *et al.*, 1998; De Biase *et al.*, 2009; Herman *et al.*, 2006). By probing RNA Pol II recruitment and *FXN* pre-mRNA levels at multiple locations up- and downstream of the GAA repeats in several unaffected and FRDA patient-derived cells, this work has expanded the characterisation of *FXN* gene regulation in health and disease. In particular, the quantitative analysis of active RNA Pol II using Br-UTP nuclear run-on directly demonstrated the transcription elongation defect due to expanded GAA repeats for the first time *in vivo* (Chapter 3.1.4).

Accumulation of R-loops in FRDA and FXS

The formation of R-loops on expanded repeats has been demonstrated previously *in vitro* (Grabczyk *et al.*, 2007; Haeusler *et al.*, 2014; Reddy *et al.*, 2011). One study showed R-loop formation on transgenic integrated $(CTG)_{67}$ repeats in HEK293 cells using native bisulfite sequencing but did not investigate repeats of different sizes (Lin *et al.*, 2010). The present work demonstrated the formation of RNA/DNA hybrids at endogenous expanded GAA repeats for the first time *in vivo*, using DIP, consistent with the formation of R-loops (Chapter 3.2). Furthermore, R-loops have now been detected on endogenous CGG repeats of different sizes and in various cell types, including expanded FXS alleles in lymphoblasts (this study), premutation FXTAS alleles in lymphoblasts (Loomis *et al.*, 2014), expanded FXS alleles in differentiating neurons (Colak *et al.*, 2014). Together with the work of others, my results suggest that R-loop formation may be a general feature of repeat expansion diseases.

RNA/DNA hybrids and R-loops as triggers of FRDA and FXS pathologies

The results from this work suggest a functional role of RNA/DNA hybrids and R-loops in the pathology of FRDA for the first time based on *in vivo* data. In particular, I showed that stable RNA/DNA hybrids promote heterochromatin formation and transcriptional repression of the *FXN* gene in FRDA cells (Chapter 4.3). Furthermore, the levels of RNA/DNA hybrids correlate with the GAA expansion size, associated with more dramatic *FXN* repression (Chapters 3.2.1, 3.2.2, 4.5.2). Therefore, increased RNA/DNA hybrid formation may underlie the earlier onset and more severe disease symptoms associated with larger GAA repeat expansions. Moreover, the observation that RNA/DNA hybrids are induced at expanded CGG repeats of *FMR1* gene upon inhibition of DNA methylation may provide a new explanation as to why full mutation *FMR1* alleles can only be partially and transiently reactivated (Chapters 3.3, 4.5.6).

The role of H3K9me2 in repeat-mediated gene silencing

The demonstration that GAA repeats can trigger ectopic heterochromatin formation reminiscent of position effect variegation prompted the investigation of heterochromatin-mediated repression of *FXN* gene in FRDA (Saveliev *et al.*, 2003). I investigated the distribution of the repressive histone modification H3K9me2 and found a strong enrichment around expanded GAA repeats in FRDA cells, confirming previous studies (Chapter 3.1.5) (Herman *et al.*, 2006; Punga and Buhler, 2010). Furthermore, I confirmed that despite its clear correlation with expanded GAA repeats, H3K9me2 is dispensable for *FXN* gene repression (Chapter 4.3.1) (Punga and Buhler, 2010). My functional studies on the involvement of R-loops in *FXN* repression may also provide an explanation for this seemingly paradoxical observation: changes in H3K9me2 levels do not affect RNA/DNA hybrids, which play a key role in *FXN* repression. As discussed in Chapter 4.5.6, H3K9me2 associated with expanded CGG repeats may also be dispensable for *FMR1* silencing in FXS (Bar-Nur *et al.*, 2012; Tabolacci *et al.*, 2008b; Tabolacci *et al.*, 2005) . My results and those of others thus suggest that high H3K9me2 levels at expanded repeats are likely to be a consequence, not a cause of transcriptional repression, while other histone modifications such as acetylation or repressive H3K27me3 may play more central functions in silencing. This may have implications for

the design of future therapies which should be more efficient if they target RNA/DNA hybrids or R-loops instead of downstream epigenetic processes (see also Chapter 6.3.4).

6.2.2 Characterisation of the RNA/DNA hybrid interactome

In this study, I developed a method which allows the unbiased quantitative characterisation of proteins interacting with RNA/DNA hybrids *in vivo* (Chapter 5). Previous studies provided indirect evidence for functions of proteins in R-loop biology using genetic assays (discussed in Chapters 1.1.5, 5.1). Moreover, ChIP experiments have demonstrated RNA/DNA hybrid-dependent recruitment of some factors, including the FACT complex, SETX, BRCA1 and BRCA2 (Bhatia *et al.*, 2014; Gomez-Gonzalez *et al.*, 2011; Hatchi *et al.*, 2015; Herrera-Moyano *et al.*, 2014; Santos-Pereira *et al.*, 2013; Skourti-Stathaki *et al.*, 2011). The proteome-wide identification of candidate RNA/DNA hybrid interactors in this work may complement these functional studies and provide new targets for future in-depth functional characterisations.

Supporting the relevance of the RNA/DNA hybrid interactome as a discovery tool, the biochemical investigation of top candidate DHX9 demonstrated that it interacts with and resolves RNA/DNA hybrids *in vivo*, validating and extending previous evidence of its RNA/DNA helicase activity *in vitro* (Chapter 5.5.1) (Chakraborty and Grosse, 2011). The functional relevance of this work was further supported by work carried out by Dr M. Kristiansen in this group, which suggests that DHX9 is required for efficient transcriptional termination of human genes *in vivo* (discussed in Chapter 5.6.2).

A similar approach to enrich for proteins associated with RNA/DNA hybrids using the S9.6 antibody was recently employed in HEK293T cells (Nadel *et al.*, 2015). Despite differences in the methodology and cell type, this study also identified a number of proteins highly enriched in the HeLa RNA/DNA hybrid interactome including DEAD/H helicases such as DHX9, hnRNPs and SR proteins (Chapter 5.4) (Nadel *et al.*, 2015). However, the results from my work provide a more comprehensive RNA/DNA hybrid interactome data set (469 versus 81 proteins), based on more extensive method validation, greater proteomic coverage and stringent statistical criteria employing biological replicates. Furthermore, the work described in this thesis provides the first bioinformatic

analysis of the RNA/DNA hybrid interactome composition and a functional characterisation of top candidate DHX9.

6.3 Future directions

6.3.1 Interplay between R-loops and heterochromatin

R-loops and histone acetylation in repeat expansion diseases

The present study has provided a mechanistic link between the formation of RNA/DNA hybrids, enrichment of repressive heterochromatin and transcriptional silencing in a human repeat expansion disease, Friedreich ataxia (Chapters 4.5.3–4.5.4). However, many questions remain to be answered. It will be important to investigate how RNA/DNA hybrids and R-loops relate to the decreased histone acetylation associated with expanded GAA and CGG repeats in FRDA and FXS, respectively. Several studies have shown that inhibition of HDACs leads to only partial transcriptional reactivation of expanded alleles in FRDA and FXS (Coffee *et al.*, 1999; Herman *et al.*, 2006; Tabolacci *et al.*, 2008a). It is tempting to speculate that this may be due to persistence of stable R-loops during HDAC inhibition, which could reinforce the defect in transcription elongation. The relevant experiments would be designed similarly to the ones performed in this work (Chapters 4.3.1–4.3.3). In particular, lymphoblastoid cells could be treated with HDAC inhibitors, followed by DIP and RT-qPCR analyses to profile their effects on gene expression. Conversely, R-loops could be modified using CPT treatment, or RNase H1 overexpression and knockdown, followed by histone acetylation ChIP and RT-qPCR analysis of *FXN* and *FMR1* expression.

A potential mechanistic link between R-loops, H3K27me3 and PRC2

The enrichment of H3K27me3 at expanded GAA and CGG repeats suggests that this histone modification may be important for *FXN* and *FMR1* silencing (discussed in Chapters 4.5.4, 4.5.6). Although transcriptionally-dependent recruitment of PRC2 subunit EZH2 to reactivated *FMR1* alleles has been demonstrated (Kumari and Usdin, 2014), future studies need to address if PRC2 and H3K27me3 are indeed required for *FMR1* silencing. This can be investigated by depletion of PRC2 subunits or pharmacological inhibition of the H3K27me3 methyltransferase EZH2 in combination

with DIP and RT-qPCR analysis of *FMR1* gene expression. Furthermore, the potential recruitment of PRC2 by R-loops formed at expanded CGG may be studied by modifying RNA/DNA hybrid levels using CPT treatment and RNase H1 overexpression or depletion. It will be of particular interest to perform these experiments in FXS cells with a silenced *FMR1* gene and cells treated with 5-azadC to determine if H3K27me3 and PRC2 are required for initial establishment of silencing or maintenance, or both. Although less data are available, PRC2 recruitment could potentially also contribute to *FXN* repression in FRDA. Initial studies would need to examine if subunits of PRC2 are enriched at expanded GAA repeats, with subsequent experiments focusing on the interplay with R-loops and other histone modifications such as histone acetylation.

6.3.2 R-loops and repeat instability

Transcription-dependent repeat instability

The formation of RNA/DNA hybrids in R-loops at trinucleotide repeats in FRDA and FXS as described in this work and by others is likely to have implications for our understanding of the intergenerational instability characteristic of these diseases (Chapters 1.3, 4.5.2) (Lopez Castel *et al.*, 2010). It will be a major task to comprehensively investigate the contribution of R-loops to trinucleotide repeat instability in the future.

Several studies employing plasmid-based and integrated reporter systems containing trinucleotide repeats have demonstrated that transcription of expanded repeats contributes to their instability (Ditch *et al.*, 2009; Lin *et al.*, 2006; Soragni *et al.*, 2008). Furthermore, R-loop formation and R-loop-dependent instability was directly shown for expanded CTG repeats using integrated reporter constructs in human cells (Lin *et al.*, 2010). In support of a crucial role of transcription for repeat instability *in vivo*, expanded CGG repeats are highly unstable in the germ line and undifferentiated cells, where the *FMR1* gene is expressed but become stable upon transcriptional *FMR1* silencing (Colak *et al.*, 2014; Lopez Castel *et al.*, 2010; Wohrle *et al.*, 2001). A mechanism for *FMR1* CGG repeat instability has recently been suggested. Replication forks stall at CGG repeats of the endogenous *FMR1* gene in FXS cells, likely due to the formation of DNA hairpins in the single-stranded template DNA (Gerhardt *et al.*, 2014). Moreover, a switch in the origin of replication in

FXS cells may contribute to repeat expansion by changing the direction of replication across the *FMR1* gene, which allows increased formation of stable CGG DNA hairpins, leading to DNA polymerase slippage (Gerhardt *et al.*, 2014). Future studies will need to address if R-loops contribute to the cell type specificity and extent of repeat instability *in vivo*. Furthermore, it will be interesting to investigate if R-loops play a role in the origin of replication switch at *FMR1* gene in FXS and if this can be affected by R-loop-targeting small molecules.

Involvement of protein factors

Interestingly, expanded GAA repeats are stable in differentiated FRDA lymphoblasts and neurons but expand in undifferentiated FRDA iPSCs (Du *et al.*, 2012; Ku *et al.*, 2010). This has been attributed to increased recruitment of the mismatch repair factors MSH2/3/6 in iPSCs (Du *et al.*, 2012; Ku *et al.*, 2010). It will be interesting to investigate if a potential difference in R-loop levels in FRDA iPSCs compared to somatic cells causes this recruitment. Alternatively, different expression levels or activities of DNA repair factors may affect their affinity for expanded GAA repeats. A recent study has highlighted that R-loops can serve as substrates for transcription-coupled nucleotide excision repair (TC-NER) enzymes, leading to their processing into double-stranded breaks (Sollier *et al.*, 2014). In line with their involvement in repeat expansion, depletion of TC-NER and mismatch repair factors reduces transcription-dependent instability of repeats (Lin *et al.*, 2006; Lin and Wilson, 2007).

However, the interplay between trinucleotide repeats and DNA repair factors may be more complex and dependent on cellular context, since mismatch repair factors have more recently been shown to protect against intergenerational instability in mice (Ezzatizadeh *et al.*, 2012). The identification of several proteins involved in DNA damage repair in the RNA/DNA hybrid interactome may uncover new factors with so far uncharacterised functions in repeat instability (Chapter 5.6.3). It will be important to comprehensively investigate the formation of R-loops and their interaction with DNA repair factors in different relevant cell types. A potential study could involve simultaneous proteomics-based quantitation of repair factor protein expression, DIP analyses

and measurements of repeat instability capturing different time points during differentiation of FRDA or FXS iPSCs to neurons.

6.3.3 Further characterisation of the RNA/DNA hybrid interactome

Methodological development of RNA/DNA hybrid IP

The RNA/DNA hybrid IP method established in this work provides a list of 469 proteins identified as candidate components of the RNA/DNA hybrid interactome (Chapter 5.4.1). However, some *bona fide* RNA/DNA hybrid interactors, including low abundance proteins, may not have been identified due to technical limitations (discussed in Chapter 5.6.5). Therefore, further optimisation of the RNA/DNA hybrid IP method may extend the RNA/DNA hybrid interactome in the future. The RNA/DNA hybrid IP could be scaled up, in combination with the usage of alternative sample processing techniques for mass spectrometry. In particular, in the method described here a one-dimensional reverse phase liquid chromatography separation of peptides prior to mass spectrometry was employed (Chapter 5.4.1). For complex samples, the identification depth could be increased by incorporating an additional pre-fractionation step such as SDS-PAGE of proteins or strong cation exchange chromatography of peptides (Marcus, 2012).

A highly accurate way of detecting even small differences in relative protein abundance is stable isotope labelling of amino acids, which is based on adding unique mass tags to amino acids originating from different samples (Zhou *et al.*, 2014). Stable isotopes can either be introduced metabolically by supplementing growth media (stable isotope labelling by amino acids in cell culture, SILAC) or by chemical addition of the mass tags to peptides after cell lysis (Zhou *et al.*, 2014). Since different samples can be mixed and processed as one sample after incorporation of the mass tags, technical variability is greatly reduced compared to label-free quantitation methods (Zhou *et al.*, 2014). Therefore, employing stable isotope labelling for RNA/DNA hybrid IP may provide more accurate and sensitive quantitation. This could be particularly useful in exploring the RNA/DNA hybrid interactome in different biological samples.

Dynamics of the RNA/DNA hybrid interactome

The current work has characterised the RNA/DNA hybrid interactome of proliferating HeLa cells. Work from other groups has demonstrated that the there is considerable overlap between mRNA interactomes of HeLa, HEK293 and mouse embryonic stem cells, but there are also many cell type-specific mRNA interactors (Baltz *et al.*, 2012; Castello *et al.*, 2012; Kwon *et al.*, 2013). Thus, one of the main tasks in the future will be to comprehensively describe the composition of RNA/DNA hybrid interactomes in different cell types and tissues. Since protein factors may play an important role in transcriptional silencing and repeat instability in FRDA and FXS (discussed in Chapters 4.5.6 and 6.3.2), the investigation of RNA/DNA hybrid interactomes in clinically relevant tissues may shed light on the molecular mechanisms of repeat expansion diseases. In particular, it will be interesting to investigate the RNA/DNA hybrid interactomes in FXS throughout development from embryonic stem cells to differentiated neurons, which will offer new insights into the association of DNA repair factors or repressive chromatin complexes with RNA/DNA hybrids. However, since the causative genetic change in FRDA and FXS involves a single gene, it may be necessary to improve the sensitivity of the RNA/DNA hybrid IP method or to develop a novel strategy to specifically purify RNA/DNA hybrids and associated proteins from a single gene locus. Furthermore, the RNA/DNA hybrid IP method may be directly applied to human diseases with globally dysregulated RNA/DNA hybrids, including Aicardi-Goutières syndrome and AOA2/ALS4 (see Chapters 1.2.1– 1.2.2 and 6.3.4) (Lim *et al.*, 2015; Yeo *et al.*, 2014).

6.3.4 R-loop therapies

The present work may have implications for the future development of drugs for repeat expansion disorders. It will be important to target the primary trigger of transcriptional repression; RNA/DNA hybrids or R-loops. Gene reactivation by restoring histone acetylation in FRDA or removing DNA methylation in FXS may be only partial due to the accumulation of RNA/DNA hybrids which could recruit alternative repressive factors to the genes.

Design and specificity of R-loop therapies

Various ligands can target RNA/DNA hybrids, including ethidium bromide, the aminoglycosides neomycin and paramomycin, and the polyamides distamycin and netropsin (Shaw and Arya, 2008). These compounds recognise RNA/DNA hybrids through intercalation and binding to the nucleic acid groove. Despite exhibiting high binding affinities to RNA/DNA hybrids, many of these molecules also bind dsDNA and RNA and are mutagenic, limiting their potential biological applications (Shaw and Arya, 2008). However, recent studies suggest that combining the properties of these ligands can achieve sub-nanomolar affinity for RNA/DNA hybrids (Shaw and Arya, 2008). In particular, this has been demonstrated for ligands linking aminoglycosides to derivatives of ethidium bromide, providing a possible approach for the development of potent and specific RNA/DNA hybrid ligands in future drug design efforts. A recent study highlighted an important aspect of potential R-loop therapies. A small molecule inhibitor shown to prevent R-loop formation at expanded CGG repeats of the *FMR1* gene did not target RNA/DNA hybrids directly but rather stabilised the secondary structure of CGG repeat RNA, preventing its hybridisation with the DNA template strand (Colak *et al.*, 2014). Many repeat-containing RNAs have been proposed to fold into secondary structures such as hairpins and G4 quadruplexes (Krzyzosiak *et al.*, 2012). Furthermore, it has been demonstrated that efficient folding of nascent RNA into secondary structures prevents RNA/DNA hybrid and consequently R-loop formation and genome instability in *S. cerevisiae* (Chen *et al.*, 2015b). Thus, designing inhibitors that stabilise the folding of repeat-associated RNA sequence-specifically may be an approach to reduce R-loop levels while minimising off target effects.

An alternative approach for R-loop therapies is to modify functions of proteins implicated in R-loop biology. The potential for this has been shown for the first time by targeting Top1 with the specific inhibitor topotecan to reactivate the paternally imprinted epigenetically silenced *Ube3a* gene in Angelman syndrome (Figure 1.4, Chapter 1.2.3) (Powell *et al.*, 2013). However, when targeting proteins with many different cellular functions for therapeutical purposes it will be important to control for off target effects. Inhibition of Top1 has been widely used as a tool to induce R-loops in various experimental systems (Chapter 4.3.2) (Marinello *et al.*, 2013; Sordet *et al.*, 2009; Tuduri *et*

al., 2009). In particular, the observation that R-loop induction correlated with widespread DNA damage in these studies suggests that other strategies targeting R-loops may be required. The characterisation of the RNA/DNA hybrid interactome in the present study may provide new targets for potential R-loop therapies (Chapter 5.6.3–5.6.4).

Designing therapies utilising RNA/DNA hybrid-interacting proteins to affect RNA/DNA hybrid and R-loop levels genome-wide may prove more promising in diseases with a widespread dysregulation of R-loops. While this aspect of R-loops in human pathologies has not been investigated in great detail so far, it is conceivable that this approach may be used in at least two disease families: Aicardi-Goutières syndrome and cancer (Chapters 1.2.2, 1.2.4). In particular, the widespread accumulation of RNA/DNA hybrids in cells from AGS patients (Lim *et al.*, 2015) could potentially be targeted by either reactivating the mutated AGS genes or introducing proteins capable of resolving RNA/DNA hybrids such as RNase H1 or RNA/DNA helicases.

In cancer, transcription, DNA replication and DNA repair are dysregulated on a genome-wide scale (Hanahan and Weinberg, 2011). The enrichment of DEAD/H helicases in the RNA/DNA hybrid interactome and their amplification in human cancers suggests that these proteins play an important part in tumorigenesis (Chapter 5.5.2). It will be interesting to investigate whether R-loops are dysregulated in different cancers and if they play a causal role in tumorigenesis. The transition from normal to cancer cells relies on increased genome instability, enough to acquire unlimited growth and metastasis capabilities but not as much as to be harmful to the cancer cells (Hanahan and Weinberg, 2011). Therefore, modifying the levels of R-loops in pre-cancerous or cancerous cells may reduce the tumorigenic potential or kill cancer cells due to excessive DNA damage. Indeed, the Top1 inhibitor camptothecin and its analogues topotecan and irinotecan are used as chemotherapeutic drugs (Capranico *et al.*, 2010). It has been demonstrated that their cellular cytotoxicity is at least partially mediated by induction of R-loop-dependent DNA damage (Marinello *et al.*, 2013; Sordet *et al.*, 2009; Tuduri *et al.*, 2009).

Reactivation therapies for FXS and FRDA

Efforts to alleviate the FMRP deficiency in FXS by reactivating the *FMR1* gene have been hampered by the lack of efficiency of the investigated drugs including DNA methylation and HDAC inhibitors (Chapter 4.5.6). Alternative therapeutical strategies may need to be pursued depending on the disease stage. In particular, while the inhibition of R-loop formation during differentiation of neuronal cells from iPSCs prevents *FMR1* silencing, the treatment does not reactivate silenced *FMR1* alleles in somatic cells (Colak *et al.*, 2014). To fully and permanently reactivate *FMR1* alleles in differentiated cells, results from this and other groups now suggest that it may be necessary to develop a combination therapy employing DNA methylation and R-loop inhibitors (Chapter 4.5.6).

FRDA treatments currently under development are based on HDAC inhibitors which only partially reactivate *FXN* gene expression. Another concern, especially for the long term treatment of FRDA patients are off-target effects since HDACs are involved in epigenetic regulation of many genes (Chapters 4.5.4–4.5.5). Targeting R-loops formed at expanded GAA repeats may therefore allow for more site-specific approaches. Although results from this study demonstrated that resolving RNA/DNA hybrids by overexpressing RNase H1 may alleviate the transcriptional repression of *FXN* gene in FRDA (Chapter 4.2.2), this approach likely targets other RNA/DNA hybrids formed throughout the genome, potentially impairing essential cellular functions including transcription initiation and termination (Chen *et al.*, 2015aChen *et al.*, 2015a; Ginno *et al.*, 2012; Skourti-Stathaki *et al.*, 2011). Therefore, it will be important to develop strategies to recruit RNase H1 specifically to the *FXN* locus, or to design small molecules which target the *FXN* gene sequence-specifically.

References

Abrahao, A., Pedroso, J.L., de Carvalho Aguiar, P.M., and Barsottini, O.G. (2015). Gene Expression Profile in Peripheral Blood Cells of Friedreich Ataxia Patients. Cerebellum.

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

Al-Mahdawi, S., Pinto, R.M., Ismail, O., Varshney, D., Lymperi, S., Sandi, C., Trabzuni, D., and Pook, M. (2008). The Friedreich ataxia GAA repeat expansion mutation induces comparable epigenetic changes in human and transgenic mouse brain and heart tissues. Human molecular genetics *17*, 735-746.

Alzu, A., Bermejo, R., Begnis, M., Lucca, C., Piccini, D., Carotenuto, W., Saponaro, M., Brambati, A., Cocito, A., Foiani, M., *et al.* (2012). Senataxin associates with replication forks to protect fork integrity across RNA-polymerase-II-transcribed genes. Cell *151*, 835-846.

Anheim, M., Monga, B., Fleury, M., Charles, P., Barbot, C., Salih, M., Delaunoy, J.P., Fritsch, M., Arning, L., Synofzik, M., *et al.* (2009). Ataxia with oculomotor apraxia type 2: clinical, biological and genotype/phenotype correlation study of a cohort of 90 patients. Brain *132*, 2688-2698.

Arana, M.E., Kerns, R.T., Wharey, L., Gerrish, K.E., Bushel, P.R., and Kunkel, T.A. (2012). Transcriptional responses to loss of RNase H2 in Saccharomyces cerevisiae. DNA repair *11*, 933-941.

Arora, R., Lee, Y., Wischnewski, H., Brun, C.M., Schwarz, T., and Azzalin, C.M. (2014). RNaseH1 regulates TERRA-telomeric DNA hybrids and telomere maintenance in ALT tumour cells. Nature communications *5*, 5220.

Aygun, O., Svejstrup, J., and Liu, Y. (2008). A RECQ5-RNA polymerase II association identified by targeted proteomic analysis of human chromatin. Proceedings of the National Academy of Sciences of the United States of America *105*, 8580-8584.

Baaklini, I., Usongo, V., Nolent, F., Sanscartier, P., Hraiky, C., Drlica, K., and Drolet, M. (2008). Hypernegative supercoiling inhibits growth by causing RNA degradation. J Bacteriol *190*, 7346-7356.

Baker, T.A., and Kornberg, A. (1988). Transcriptional activation of initiation of replication from the E. coli chromosomal origin: an RNA-DNA hybrid near oriC. Cell *55*, 113-123.

Balakrishnan, L., and Bambara, R.A. (2013). Okazaki fragment metabolism. Cold Spring Harbor perspectives in biology *5*.

Balk, B., Maicher, A., Dees, M., Klermund, J., Luke-Glaser, S., Bender, K., and Luke, B. (2013). Telomeric RNA-DNA hybrids affect telomere-length dynamics and senescence. Nature structural & molecular biology *20*, 1199-1205.

Baltz, A.G., Munschauer, M., Schwanhausser, B., Vasile, A., Murakawa, Y., Schueler, M., Youngs, N., Penfold-Brown, D., Drew, K., Milek, M., *et al.* (2012). The mRNA-bound proteome and its global occupancy profile on protein-coding transcripts. Molecular cell *46*, 674-690.

Banez-Coronel, M., Porta, S., Kagerbauer, B., Mateu-Huertas, E., Pantano, L., Ferrer, I., Guzman, M., Estivill, X., and Marti, E. (2012). A pathogenic mechanism in Huntington's disease involves small CAG-repeated RNAs with neurotoxic activity. PLoS genetics *8*, e1002481.

Bar-Nur, O., Caspi, I., and Benvenisty, N. (2012). Molecular analysis of FMR1 reactivation in fragile-X induced pluripotent stem cells and their neuronal derivatives. Journal of molecular cell biology *4*, 180-183.

Baralle, M., Pastor, T., Bussani, E., and Pagani, F. (2008). Influence of Friedreich ataxia GAA noncoding repeat expansions on pre-mRNA processing. American journal of human genetics *83*, 77-88.

Becherel, O.J., Sun, J., Yeo, A.J., Nayler, S., Fogel, B.L., Gao, F., Coppola, G., Criscuolo, C., De Michele, G., Wolvetang, E., *et al.* (2015). A new model to study neurodegeneration in ataxia oculomotor apraxia type 2. Human molecular genetics *24*, 5759-5774.

Becherel, O.J., Yeo, A.J., Stellati, A., Heng, E.Y., Luff, J., Suraweera, A.M., Woods, R., Fleming, J., Carrie, D., McKinney, K., *et al.* (2013). Senataxin plays an essential role with DNA damage response proteins in meiotic recombination and gene silencing. PLoS genetics *9*, e1003435.

Belotserkovskii, B.P., Mirkin, S.M., and Hanawalt, P.C. (2013). DNA sequences that interfere with transcription: implications for genome function and stability. Chem Rev *113*, 8620-8637.

Benjamini, Y., and Hochberg, Y. (1995). Controlling the False Discovery Rate - a Practical and Powerful Approach to Multiple Testing. J Roy Stat Soc B Met *57*, 289-300.

Bhatia, V., Barroso, S.I., Garcia-Rubio, M.L., Tumini, E., Herrera-Moyano, E., and Aguilera, A. (2014). BRCA2 prevents R-loop accumulation and associates with TREX-2 mRNA export factor PCID2. Nature.

Biacsi, R., Kumari, D., and Usdin, K. (2008). SIRT1 inhibition alleviates gene silencing in Fragile X mental retardation syndrome. PLoS genetics *4*, e1000017.

Bidichandani, S.I., Ashizawa, T., and Patel, P.I. (1998). The GAA triplet-repeat expansion in Friedreich ataxia interferes with transcription and may be associated with an unusual DNA structure. American journal of human genetics *62*, 111-121.

Bierhoff, H., Postepska-Igielska, A., and Grummt, I. (2014). Noisy silence: non-coding RNA and heterochromatin formation at repetitive elements. Epigenetics : official journal of the DNA Methylation Society *9*, 53-61.

Bird, M.J., Needham, K., Frazier, A.E., van Rooijen, J., Leung, J., Hough, S., Denham, M., Thornton, M.E., Parish, C.L., Nayagam, B.A., *et al.* (2014). Functional characterization of Friedreich ataxia iPS-derived neuronal progenitors and their integration in the adult brain. PloS one *9*, e101718.

Blackledge, N.P., Rose, N.R., and Klose, R.J. (2015). Targeting Polycomb systems to regulate gene expression: modifications to a complex story. Nature reviews Molecular cell biology *16*, 643-649.

Boguslawski, S.J., Smith, D.E., Michalak, M.A., Mickelson, K.E., Yehle, C.O., Patterson, W.L., and Carrico, R.J. (1986). Characterization of monoclonal antibody to DNA.RNA and its application to immunodetection of hybrids. Journal of immunological methods *89*, 123-130.

Boisvert, F.M., Ahmad, Y., Gierlinski, M., Charriere, F., Lamont, D., Scott, M., Barton, G., and Lamond, A.I. (2012). A quantitative spatial proteomics analysis of proteome turnover in human cells. Molecular & cellular proteomics : MCP *11*, M111 011429.

Boque-Sastre, R., Soler, M., Oliveira-Mateos, C., Portela, A., Moutinho, C., Sayols, S., Villanueva, A., Esteller, M., and Guil, S. (2015). Head-to-head antisense transcription and R-loop formation

promotes transcriptional activation. Proceedings of the National Academy of Sciences of the United States of America.

Boudvillain, M., Figueroa-Bossi, N., and Bossi, L. (2013). Terminator still moving forward: expanding roles for Rho factor. Current opinion in microbiology *16*, 118-124.

Boule, J.B., and Zakian, V.A. (2007). The yeast Pif1p DNA helicase preferentially unwinds RNA DNA substrates. Nucleic acids research *35*, 5809-5818.

Bradford, M.M. (1976). A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. Analytical biochemistry *72*, 248-254.

Brennan, C.A., Dombroski, A.J., and Platt, T. (1987). Transcription termination factor rho is an RNA-DNA helicase. Cell *48*, 945-952.

Britton, S., Dernoncourt, E., Delteil, C., Froment, C., Schiltz, O., Salles, B., Frit, P., and Calsou, P. (2014). DNA damage triggers SAF-A and RNA biogenesis factors exclusion from chromatin coupled to R-loops removal. Nucleic acids research *42*, 9047-9062.

Brouwer, J.R., Huguet, A., Nicole, A., Munnich, A., and Gourdon, G. (2013). Transcriptionally Repressive Chromatin Remodelling and CpG Methylation in the Presence of Expanded CTG-Repeats at the DM1 Locus. Journal of nucleic acids *2013*, 567435.

Brown, T.A., Tkachuk, A.N., and Clayton, D.A. (2008). Native R-loops persist throughout the mouse mitochondrial DNA genome. The Journal of biological chemistry *283*, 36743-36751.

Bubeck, D., Reijns, M.A., Graham, S.C., Astell, K.R., Jones, E.Y., and Jackson, A.P. (2011). PCNA directs type 2 RNase H activity on DNA replication and repair substrates. Nucleic acids research *39*, 3652-3666.

Burns, M.B., Lackey, L., Carpenter, M.A., Rathore, A., Land, A.M., Leonard, B., Refsland, E.W., Kotandeniya, D., Tretyakova, N., Nikas, J.B., *et al.* (2013). APOBEC3B is an enzymatic source of mutation in breast cancer. Nature *494*, 366-370.

Butler, J.S., and Napierala, M. (2015). Friedreich's ataxia--a case of aberrant transcription termination? Transcription *6*, 33-36.

Campuzano, V., Montermini, L., Molto, M.D., Pianese, L., Cossee, M., Cavalcanti, F., Monros, E., Rodius, F., Duclos, F., Monticelli, A., *et al.* (1996). Friedreich's ataxia: autosomal recessive disease caused by an intronic GAA triplet repeat expansion. Science *271*, 1423-1427.

Capranico, G., Marinello, J., and Baranello, L. (2010). Dissecting the transcriptional functions of human DNA topoisomerase I by selective inhibitors: implications for physiological and therapeutic modulation of enzyme activity. Biochimica et biophysica acta *1806*, 240-250.

Cassidy, S.B., Dykens, E., and Williams, C.A. (2000). Prader-Willi and Angelman syndromes: sister imprinted disorders. American journal of medical genetics *97*, 136-146.

Castellano-Pozo, M., Garcia-Muse, T., and Aguilera, A. (2012). R-loops cause replication impairment and genome instability during meiosis. EMBO reports *13*, 923-929.

Castellano-Pozo, M., Santos-Pereira, J.M., Rondon, A.G., Barroso, S., Andujar, E., Perez-Alegre, M., Garcia-Muse, T., and Aguilera, A. (2013). R loops are linked to histone H3 S10 phosphorylation and chromatin condensation. Molecular cell *52*, 583-590.

Castello, A., Fischer, B., Eichelbaum, K., Horos, R., Beckmann, B.M., Strein, C., Davey, N.E., Humphreys, D.T., Preiss, T., Steinmetz, L.M., *et al.* (2012). Insights into RNA biology from an atlas of mammalian mRNA-binding proteins. Cell *149*, 1393-1406.

Cerritelli, S.M., and Crouch, R.J. (2009). Ribonuclease H: the enzymes in eukaryotes. The FEBS journal *276*, 1494-1505.

Cerritelli, S.M., Frolova, E.G., Feng, C., Grinberg, A., Love, P.E., and Crouch, R.J. (2003). Failure to produce mitochondrial DNA results in embryonic lethality in Rnaseh1 null mice. Molecular cell *11*, 807-815.

Chakraborty, P., and Grosse, F. (2010). WRN helicase unwinds Okazaki fragment-like hybrids in a reaction stimulated by the human DHX9 helicase. Nucleic acids research *38*, 4722-4730.

Chakraborty, P., and Grosse, F. (2011). Human DHX9 helicase preferentially unwinds RNA-containing displacement loops (R-loops) and G-quadruplexes. DNA repair *10*, 654-665.

Chambers, M.C., Maclean, B., Burke, R., Amodei, D., Ruderman, D.L., Neumann, S., Gatto, L., Fischer, B., Pratt, B., Egertson, J., *et al.* (2012). A cross-platform toolkit for mass spectrometry and proteomics. Nature biotechnology *30*, 918-920.

Chan, P.K., Torres, R., Yandim, C., Law, P.P., Khadayate, S., Mauri, M., Grosan, C., Chapman-Rothe, N., Giunti, P., Pook, M., *et al.* (2013). Heterochromatinization induced by GAA-repeat hyperexpansion in Friedreich's ataxia can be reduced upon HDAC inhibition by vitamin B3. Human molecular genetics *22*, 2662-2675.

Chan, Y.A., Aristizabal, M.J., Lu, P.Y., Luo, Z., Hamza, A., Kobor, M.S., Stirling, P.C., and Hieter, P. (2014). Genome-Wide Profiling of Yeast DNA:RNA Hybrid Prone Sites with DRIP-Chip. PLoS genetics *10*, e1004288.

Chaudhuri, J., Khuong, C., and Alt, F.W. (2004). Replication protein A interacts with AID to promote deamination of somatic hypermutation targets. Nature *430*, 992-998.

Chaudhuri, J., Tian, M., Khuong, C., Chua, K., Pinaud, E., and Alt, F.W. (2003). Transcription-targeted DNA deamination by the AID antibody diversification enzyme. Nature *422*, 726-730.

Chen, P.B., Chen, H.V., Acharya, D., Rando, O.J., and Fazzio, T.G. (2015a). R loops regulate promoter-proximal chromatin architecture and cellular differentiation. Nature structural & molecular biology.

Chen, X., Yang, J.R., and Zhang, J. (2015b). Nascent RNA folding mitigates transcription-associated mutagenesis. Genome research.

Chen, Y.Z., Bennett, C.L., Huynh, H.M., Blair, I.P., Puls, I., Irobi, J., Dierick, I., Abel, A., Kennerson, M.L., Rabin, B.A., *et al.* (2004). DNA/RNA helicase gene mutations in a form of juvenile amyotrophic lateral sclerosis (ALS4). American journal of human genetics *74*, 1128-1135.

Chernikova, S.B., Razorenova, O.V., Higgins, J.P., Sishc, B.J., Nicolau, M., Dorth, J.A., Chernikova, D.A., Kwok, S., Brooks, J.D., Bailey, S.M., *et al.* (2012). Deficiency in mammalian histone H2B ubiquitin ligase Bre1 (Rnf20/Rnf40) leads to replication stress and chromosomal instability. Cancer research *72*, 2111-2119.

Cho, D.H., Thienes, C.P., Mahoney, S.E., Analau, E., Filippova, G.N., and Tapscott, S.J. (2005). Antisense transcription and heterochromatin at the DM1 CTG repeats are constrained by CTCF. Molecular cell *20*, 483-489.

Chon, H., Sparks, J.L., Rychlik, M., Nowotny, M., Burgers, P.M., Crouch, R.J., and Cerritelli, S.M. (2013). RNase H2 roles in genome integrity revealed by unlinking its activities. Nucleic acids research *41*, 3130-3143.

Chung, D.W., Rudnicki, D.D., Yu, L., and Margolis, R.L. (2011). A natural antisense transcript at the Huntington's disease repeat locus regulates HTT expression. Human molecular genetics *20*, 3467-3477.

Chutake, Y.K., Costello, W.N., Lam, C., and Bidichandani, S.I. (2014). Altered nucleosome positioning at the transcription start site and deficient transcriptional initiation in Friedreich ataxia. The Journal of biological chemistry *289*, 15194-15202.

Clark, R.M., Dalgliesh, G.L., Endres, D., Gomez, M., Taylor, J., and Bidichandani, S.I. (2004). Expansion of GAA triplet repeats in the human genome: unique origin of the FRDA mutation at the center of an Alu. Genomics *83*, 373-383.

Coffee, B., Zhang, F., Ceman, S., Warren, S.T., and Reines, D. (2002). Histone modifications depict an aberrantly heterochromatinized FMR1 gene in fragile x syndrome. American journal of human genetics *71*, 923-932.

Coffee, B., Zhang, F., Warren, S.T., and Reines, D. (1999). Acetylated histones are associated with FMR1 in normal but not fragile X-syndrome cells. Nature genetics *22*, 98-101.

Colak, D., Zaninovic, N., Cohen, M.S., Rosenwaks, Z., Yang, W.Y., Gerhardt, J., Disney, M.D., and Jaffrey, S.R. (2014). Promoter-bound trinucleotide repeat mRNA drives epigenetic silencing in fragile X syndrome. Science *343*, 1002-1005.

Coppola, G., Burnett, R., Perlman, S., Versano, R., Gao, F., Plasterer, H., Rai, M., Sacca, F., Filla, A., Lynch, D.R., *et al.* (2011). A gene expression phenotype in lymphocytes from Friedreich ataxia patients. Annals of neurology *70*, 790-804.

Costantino, L., and Koshland, D. (2015). The Yin and Yang of R-loop biology. Current opinion in cell biology *34*, 39-45.

Cox, J., and Mann, M. (2008). MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. Nature biotechnology *26*, 1367-1372.

Cox, J., Neuhauser, N., Michalski, A., Scheltema, R.A., Olsen, J.V., and Mann, M. (2011). Andromeda: a peptide search engine integrated into the MaxQuant environment. Journal of proteome research *10*, 1794-1805.

Crow, Y.J., Hayward, B.E., Parmar, R., Robins, P., Leitch, A., Ali, M., Black, D.N., van Bokhoven, H., Brunner, H.G., Hamel, B.C., *et al.* (2006a). Mutations in the gene encoding the 3'-5' DNA exonuclease TREX1 cause Aicardi-Goutieres syndrome at the AGS1 locus. Nature genetics *38*, 917-920.

Crow, Y.J., Leitch, A., Hayward, B.E., Garner, A., Parmar, R., Griffith, E., Ali, M., Semple, C., Aicardi, J., Babul-Hirji, R., *et al.* (2006b). Mutations in genes encoding ribonuclease H2 subunits cause Aicardi-Goutieres syndrome and mimic congenital viral brain infection. Nature genetics *38*, 910-916.

Crow, Y.J., and Rehwinkel, J. (2009). Aicardi-Goutieres syndrome and related phenotypes: linking nucleic acid metabolism with autoimmunity. Human molecular genetics *18*, R130-136.

Daniels, G.A., and Lieber, M.R. (1995). RNA:DNA complex formation upon transcription of immunoglobulin switch regions: implications for the mechanism and regulation of class switch recombination. Nucleic acids research *23*, 5006-5011.

De Amicis, A., Piane, M., Ferrari, F., Fanciulli, M., Delia, D., and Chessa, L. (2011). Role of senataxin in DNA damage and telomeric stability. DNA Repair (Amst) *10*, 199-209.

De Biase, I., Chutake, Y.K., Rindler, P.M., and Bidichandani, S.I. (2009). Epigenetic silencing in Friedreich ataxia is associated with depletion of CTCF (CCCTC-binding factor) and antisense transcription. PloS one *4*, e7914.

de la Mata, M., Alonso, C.R., Kadener, S., Fededa, J.P., Blaustein, M., Pelisch, F., Cramer, P., Bentley, D., and Kornblihtt, A.R. (2003). A slow RNA polymerase II affects alternative splicing in vivo. Molecular cell *12*, 525-532.

DePamphilis, M., and Bell, S. (2010). Genome Duplication (Garland Science).

Ditch, S., Sammarco, M.C., Banerjee, A., and Grabczyk, E. (2009). Progressive GAA.TTC repeat expansion in human cell lines. PLoS genetics *5*, e1000704.

Dominguez-Sanchez, M.S., Barroso, S., Gomez-Gonzalez, B., Luna, R., and Aguilera, A. (2011). Genome instability and transcription elongation impairment in human cells depleted of THO/TREX. PLoS genetics *7*, e1002386.

Drolet, M. (2006). Growth inhibition mediated by excess negative supercoiling: the interplay between transcription elongation, R-loop formation and DNA topology. Molecular microbiology *59*, 723-730.

Drolet, M., Bi, X., and Liu, L.F. (1994). Hypernegative supercoiling of the DNA template during transcription elongation in vitro. The Journal of biological chemistry *269*, 2068-2074.

Drolet, M., Phoenix, P., Menzel, R., Masse, E., Liu, L.F., and Crouch, R.J. (1995). Overexpression of RNase H partially complements the growth defect of an Escherichia coli delta topA mutant: R-loop formation is a major problem in the absence of DNA topoisomerase I. Proceedings of the National Academy of Sciences of the United States of America *92*, 3526-3530.

Du, J., Zhong, X., Bernatavichute, Y.V., Stroud, H., Feng, S., Caro, E., Vashisht, A.A., Terragni, J., Chin, H.G., Tu, A., *et al.* (2012). Dual binding of chromomethylase domains to H3K9me2-containing nucleosomes directs DNA methylation in plants. Cell *151*, 167-180.

Du, M.X., Johnson, R.B., Sun, X.L., Staschke, K.A., Colacino, J., and Wang, Q.M. (2002). Comparative characterization of two DEAD-box RNA helicases in superfamily II: human translation-initiation factor 4A and hepatitis C virus non-structural protein 3 (NS3) helicase. The Biochemical journal *363*, 147-155.

Dudas, K.C., and Kreuzer, K.N. (2001). UvsW protein regulates bacteriophage T4 origin-dependent replication by unwinding R-loops. Molecular and cellular biology *21*, 2706-2715.

Duquette, M.L., Pham, P., Goodman, M.F., and Maizels, N. (2005). AID binds to transcription-induced structures in c-MYC that map to regions associated with translocation and hypermutation. Oncogene *24*, 5791-5798.

Durr, A., Cossee, M., Agid, Y., Campuzano, V., Mignard, C., Penet, C., Mandel, J.L., Brice, A., and Koenig, M. (1996). Clinical and genetic abnormalities in patients with Friedreich's ataxia. The New England journal of medicine *335*, 1169-1175.

Dye, M.J., and Proudfoot, N.J. (1999). Terminal exon definition occurs cotranscriptionally and promotes termination of RNA polymerase II. Molecular cell *3*, 371-378.

Echeverria, G.V., and Cooper, T.A. (2012). RNA-binding proteins in microsatellite expansion disorders: mediators of RNA toxicity. Brain research *1462*, 100-111.

Eigentler, A., Boesch, S., Schneider, R., Dechant, G., and Nat, R. (2013). Induced pluripotent stem cells from friedreich ataxia patients fail to upregulate frataxin during in vitro differentiation to peripheral sensory neurons. Stem cells and development *22*, 3271-3282.

Eiges, R., Urbach, A., Malcov, M., Frumkin, T., Schwartz, T., Amit, A., Yaron, Y., Eden, A., Yanuka, O., Benvenisty, N., *et al.* (2007). Developmental study of fragile X syndrome using human embryonic stem cells derived from preimplantation genetically diagnosed embryos. Cell stem cell *1*, 568-577.

El Hage, A., French, S.L., Beyer, A.L., and Tollervey, D. (2010). Loss of Topoisomerase I leads to R-loop-mediated transcriptional blocks during ribosomal RNA synthesis. Genes & development *24*, 1546-1558.

El Hage, A., Webb, S., Kerr, A., and Tollervey, D. (2014). Genome-Wide Distribution of RNA-DNA Hybrids Identifies RNase H Targets in tRNA Genes, Retrotransposons and Mitochondria. PLoS genetics *10*, e1004716.

Ezzatizadeh, V., Pinto, R.M., Sandi, C., Sandi, M., Al-Mahdawi, S., Te Riele, H., and Pook, M.A. (2012). The mismatch repair system protects against intergenerational GAA repeat instability in a Friedreich ataxia mouse model. Neurobiology of disease *46*, 165-171.

Feng, Y., Lakkis, L., Devys, D., and Warren, S.T. (1995). Quantitative comparison of FMR1 gene expression in normal and premutation alleles. American journal of human genetics *56*, 106-113.

Filla, A., De Michele, G., Cavalcanti, F., Pianese, L., Monticelli, A., Campanella, G., and Cocozza, S. (1996). The relationship between trinucleotide (GAA) repeat length and clinical features in Friedreich ataxia. American journal of human genetics *59*, 554-560.

Fitzwater, T., Yang, Y.L., Zhang, X.Y., and Polisky, B. (1992). Mutations affecting RNA-DNA hybrid formation of the ColE1 replication primer RNA. Restoration of RNA I sensitivity to a copy-number mutant by second-site mutations. Journal of molecular biology *226*, 997-1008.

Forbes, S.A., Beare, D., Gunasekaran, P., Leung, K., Bindal, N., Boutselakis, H., Ding, M., Bamford, S., Cole, C., Ward, S., *et al.* (2015). COSMIC: exploring the world's knowledge of somatic mutations in human cancer. Nucleic acids research *43*, D805-811.

Friedreich, N. (1863). Ueber degenerative Atrophie der spinalen Hinterstraenge. Archiv fuer pathologische Anatomie und Physiologie und fuer klinische Medizin *27*.

Garcia-Rubio, M.L., Perez-Calero, C., Barroso, S.I., Tumini, E., Herrera-Moyano, E., Rosado, I.V., and Aguilera, A. (2015). The Fanconi Anemia Pathway Protects Genome Integrity from R-loops. PLoS genetics *11*, e1005674.

Gavalda, S., Gallardo, M., Luna, R., and Aguilera, A. (2013). R-loop mediated transcription-associated recombination in trf4Delta mutants reveals new links between RNA surveillance and genome integrity. PloS one *8*, e65541.

Geiger, T., Wehner, A., Schaab, C., Cox, J., and Mann, M. (2012). Comparative proteomic analysis of eleven common cell lines reveals ubiquitous but varying expression of most proteins. Molecular & cellular proteomics : MCP *11*, M111 014050.

Gerhardt, J., Tomishima, M.J., Zaninovic, N., Colak, D., Yan, Z., Zhan, Q., Rosenwaks, Z., Jaffrey, S.R., and Schildkraut, C.L. (2014). The DNA replication program is altered at the FMR1 locus in fragile X embryonic stem cells. Molecular cell *53*, 19-31.

Ginno, P.A., Lim, Y.W., Lott, P.L., Korf, I., and Chedin, F. (2013). GC skew at the 5' and 3' ends of human genes links R-loop formation to epigenetic regulation and transcription termination. Genome research *23*, 1590-1600.

Ginno, P.A., Lott, P.L., Christensen, H.C., Korf, I., and Chedin, F. (2012). R-loop formation is a distinctive characteristic of unmethylated human CpG island promoters. Molecular cell *45*, 814-825.

Gomez-Gonzalez, B., and Aguilera, A. (2007). Activation-induced cytidine deaminase action is strongly stimulated by mutations of the THO complex. Proceedings of the National Academy of Sciences of the United States of America *104*, 8409-8414.

Gomez-Gonzalez, B., Garcia-Rubio, M., Bermejo, R., Gaillard, H., Shirahige, K., Marin, A., Foiani, M., and Aguilera, A. (2011). Genome-wide function of THO/TREX in active genes prevents R-loop-dependent replication obstacles. The EMBO journal *30*, 3106-3119.

Gonatopoulos-Pournatzis, T., and Cowling, V.H. (2014). Cap-binding complex (CBC). The Biochemical journal *457*, 231-242.

Gonitel, R., Moffitt, H., Sathasivam, K., Woodman, B., Detloff, P.J., Faull, R.L., and Bates, G.P. (2008). DNA instability in postmitotic neurons. Proc Natl Acad Sci U S A *105*, 3467-3472.

Gonzalez-Aguilera, C., Tous, C., Gomez-Gonzalez, B., Huertas, P., Luna, R., and Aguilera, A. (2008). The THP1-SAC3-SUS1-CDC31 complex works in transcription elongation-mRNA export preventing RNA-mediated genome instability. Molecular biology of the cell *19*, 4310-4318.

Gottesfeld, J.M., Rusche, J.R., and Pandolfo, M. (2013). Increasing frataxin gene expression with histone deacetylase inhibitors as a therapeutic approach for Friedreich's ataxia. Journal of neurochemistry *126 Suppl 1*, 147-154.

Goula, A.V., Stys, A., Chan, J.P., Trottier, Y., Festenstein, R., and Merienne, K. (2012). Transcription elongation and tissue-specific somatic CAG instability. PLoS Genet *8*, e1003051.

Grabczyk, E., Mancuso, M., and Sammarco, M.C. (2007). A persistent RNA.DNA hybrid formed by transcription of the Friedreich ataxia triplet repeat in live bacteria, and by T7 RNAP in vitro. Nucleic acids research *35*, 5351-5359.

Grabczyk, E., and Usdin, K. (2000a). Alleviating transcript insufficiency caused by Friedreich's ataxia triplet repeats. Nucleic acids research *28*, 4930-4937.

Grabczyk, E., and Usdin, K. (2000b). The GAA*TTC triplet repeat expanded in Friedreich's ataxia impedes transcription elongation by T7 RNA polymerase in a length and supercoil dependent manner. Nucleic Acids Res *28*, 2815-2822.

Greene, E., Mahishi, L., Entezam, A., Kumari, D., and Usdin, K. (2007). Repeat-induced epigenetic changes in intron 1 of the frataxin gene and its consequences in Friedreich ataxia. Nucleic acids research *35*, 3383-3390.

Groh, M., and Gromak, N. (2014). Out of balance: R-loops in human disease. PLoS genetics *10*, e1004630.

Groh, M., Lufino, M., Wade-Martins, R., and Gromak, N. (2014a). R-loops associated with triplet repeat expansions promote gene silencing in Friedreich ataxia and Fragile X syndrome. PLoS genetics.

Groh, M., Silva, L.M., and Gromak, N. (2014b). Mechanisms of transcriptional dysregulation in repeat expansion disorders. Biochemical Society Transactions, *in press*.

Gromak, N., Dienstbier, M., Macias, S., Plass, M., Eyras, E., Caceres, J.F., and Proudfoot, N.J. (2013). Drosha regulates gene expression independently of RNA cleavage function. Cell reports *5*, 1499-1510.

Gromak, N., West, S., and Proudfoot, N.J. (2006). Pause sites promote transcriptional termination of mammalian RNA polymerase II. Molecular and cellular biology *26*, 3986-3996.

Grzechnik, P., Gdula, M.R., and Proudfoot, N.J. (2015). Pcf11 orchestrates transcription termination pathways in yeast. Genes & development *29*, 849-861.

Guil, S., Soler, M., Portela, A., Carrere, J., Fonalleras, E., Gomez, A., Villanueva, A., and Esteller, M. (2012). Intronic RNAs mediate EZH2 regulation of epigenetic targets. Nature structural & molecular biology *19*, 664-670.

Guillen-Ahlers, H., Shortreed, M.R., Smith, L.M., and Olivier, M. (2014). Advanced methods for the analysis of chromatin-associated proteins. Physiological genomics *46*, 441-447.

Gwack, Y., Yoo, H., Song, I., Choe, J., and Han, J.H. (1999). RNA-Stimulated ATPase and RNA helicase activities and RNA binding domain of hepatitis G virus nonstructural protein 3. Journal of virology *73*, 2909-2915.

Haeusler, A.R., Donnelly, C.J., Periz, G., Simko, E.A., Shaw, P.G., Kim, M.S., Maragakis, N.J., Troncoso, J.C., Pandey, A., Sattler, R., *et al.* (2014). C9orf72 nucleotide repeat structures initiate molecular cascades of disease. Nature *507*, 195-200.

Hagerman, P. (2013). Fragile X-associated tremor/ataxia syndrome (FXTAS): pathology and mechanisms. Acta neuropathologica *126*, 1-19.

Hamperl, S., and Cimprich, K.A. (2014). The contribution of co-transcriptional RNA:DNA hybrid structures to DNA damage and genome instability. DNA Repair (Amst).

Hanahan, D., and Weinberg, R.A. (2011). Hallmarks of cancer: the next generation. Cell *144*, 646-674.

Handa, V., Saha, T., and Usdin, K. (2003). The fragile X syndrome repeats form RNA hairpins that do not activate the interferon-inducible protein kinase, PKR, but are cut by Dicer. Nucleic acids research *31*, 6243-6248.

Hatchi, E., Skourti-Stathaki, K., Ventz, S., Pinello, L., Yen, A., Kamieniarz-Gdula, K., Dimitrov, S., Pathania, S., McKinney, K.M., Eaton, M.L., *et al.* (2015). BRCA1 Recruitment to Transcriptional Pause Sites Is Required for R-Loop-Driven DNA Damage Repair. Molecular cell *57*, 636-647.

Haynes, S.R. (1999). RNA-Protein Interaction Protocols, Vol 118.

Helmrich, A., Ballarino, M., Nudler, E., and Tora, L. (2013). Transcription-replication encounters, consequences and genomic instability. Nat Struct Mol Biol *20*, 412-418.

Helmrich, A., Ballarino, M., and Tora, L. (2011). Collisions between replication and transcription complexes cause common fragile site instability at the longest human genes. Molecular cell *44*, 966-977.

Herman, D., Jenssen, K., Burnett, R., Soragni, E., Perlman, S.L., and Gottesfeld, J.M. (2006). Histone deacetylase inhibitors reverse gene silencing in Friedreich's ataxia. Nature chemical biology *2*, 551-558.

Herrera-Moyano, E., Mergui, X., Garcia-Rubio, M.L., Barroso, S., and Aguilera, A. (2014). The yeast and human FACT chromatin-reorganizing complexes solve R-loop-mediated transcription-replication conflicts. Genes & development.

Hiller, B., Achleitner, M., Glage, S., Naumann, R., Behrendt, R., and Roers, A. (2012). Mammalian RNase H2 removes ribonucleotides from DNA to maintain genome integrity. The Journal of experimental medicine *209*, 1419-1426.

Hirota, Y., and Lahti, J.M. (2000). Characterization of the enzymatic activity of hChlR1, a novel human DNA helicase. Nucleic acids research *28*, 917-924.

Holloway, T.P., Rowley, S.M., Delatycki, M.B., and Sarsero, J.P. (2011). Detection of interruptions in the GAA trinucleotide repeat expansion in the FXN gene of Friedreich ataxia. BioTechniques *50*, 182-186.

Holoch, D., and Moazed, D. (2015). RNA-mediated epigenetic regulation of gene expression. Nature reviews Genetics *16*, 71-84.

Hong, X., Cadwell, G.W., and Kogoma, T. (1995). Escherichia coli RecG and RecA proteins in R-loop formation. The EMBO journal *14*, 2385-2392.

Houlard, M., Artus, J., Leguillier, T., Vandormael-Pournin, S., and Cohen-Tannoudji, M. (2011). DNA-RNA hybrids contribute to the replication dependent genomic instability induced by Omcg1 deficiency. Cell cycle *10*, 108-117.

Hraiky, C., Raymond, M.A., and Drolet, M. (2000). RNase H overproduction corrects a defect at the level of transcription elongation during rRNA synthesis in the absence of DNA topoisomerase I in Escherichia coli. The Journal of biological chemistry *275*, 11257-11263.

Hudson, W.H., and Ortlund, E.A. (2014). The structure, function and evolution of proteins that bind DNA and RNA. Nature reviews Molecular cell biology *15*, 749-760.

Huertas, P., and Aguilera, A. (2003). Cotranscriptionally formed DNA:RNA hybrids mediate transcription elongation impairment and transcription-associated recombination. Molecular cell *12*, 711-721.

Hynes, N.E., and Stoelzle, T. (2009). Key signalling nodes in mammary gland development and cancer: Myc. Breast cancer research : BCR *11*, 210.

Itoh, T., and Tomizawa, J. (1980). Formation of an RNA primer for initiation of replication of ColE1 DNA by ribonuclease H. Proceedings of the National Academy of Sciences of the United States of America *77*, 2450-2454.

Jackson, B.R., Noerenberg, M., and Whitehouse, A. (2014). A novel mechanism inducing genome instability in Kaposi's sarcoma-associated herpesvirus infected cells. PLoS pathogens *10*, e1004098.

Jackson, D.A., Iborra, F.J., Manders, E.M., and Cook, P.R. (1998). Numbers and organization of RNA polymerases, nascent transcripts, and transcription units in HeLa nuclei. Molecular biology of the cell *9*, 1523-1536.

Jankowsky, E., and Harris, M.E. (2015). Specificity and nonspecificity in RNA-protein interactions. Nature reviews Molecular cell biology *16*, 533-544.

Jonkers, I., and Lis, J.T. (2015). Getting up to speed with transcription elongation by RNA polymerase II. Nature reviews Molecular cell biology *16*, 167-177.

Kadener, S., Fededa, J.P., Rosbash, M., and Kornblihtt, A.R. (2002). Regulation of alternative splicing by a transcriptional enhancer through RNA pol II elongation. Proceedings of the National Academy of Sciences of the United States of America *99*, 8185-8190.

Kaneko, S., Chu, C., Shatkin, A.J., and Manley, J.L. (2007). Human capping enzyme promotes formation of transcriptional R loops in vitro. Proceedings of the National Academy of Sciences of the United States of America *104*, 17620-17625.

Kao, Y.P., Hsieh, W.C., Hung, S.T., Huang, C.W., Lieber, M.R., and Huang, F.T. (2013). Detection and characterization of R-loops at the murine immunoglobulin Salpha region. Molecular immunology *54*, 208-216.

Kim, E., Napierala, M., and Dent, S.Y. (2011). Hyperexpansion of GAA repeats affects post-initiation steps of FXN transcription in Friedreich's ataxia. Nucleic acids research *39*, 8366-8377.

Kirkpatrick, D.P., and Radding, C.M. (1992). RecA protein promotes rapid RNA-DNA hybridization in heterogeneous RNA mixtures. Nucleic acids research *20*, 4347-4353.

Klose, R.J., Cooper, S., Farcas, A.M., Blackledge, N.P., and Brockdorff, N. (2013). Chromatin sampling--an emerging perspective on targeting polycomb repressor proteins. PLoS genetics *9*, e1003717.

Koeppen, A.H. (2011). Friedreich's ataxia: pathology, pathogenesis, and molecular genetics. Journal of the neurological sciences *303*, 1-12.

Krasilnikova, M.M., Kireeva, M.L., Petrovic, V., Knijnikova, N., Kashlev, M., and Mirkin, S.M. (2007). Effects of Friedreich's ataxia (GAA)n*(TTC)n repeats on RNA synthesis and stability. Nucleic Acids Res *35*, 1075-1084.

Krol, J., Fiszer, A., Mykowska, A., Sobczak, K., de Mezer, M., and Krzyzosiak, W.J. (2007). Ribonuclease dicer cleaves triplet repeat hairpins into shorter repeats that silence specific targets. Molecular cell *25*, 575-586.

Krzyzosiak, W.J., Sobczak, K., Wojciechowska, M., Fiszer, A., Mykowska, A., and Kozlowski, P. (2012). Triplet repeat RNA structure and its role as pathogenic agent and therapeutic target. Nucleic acids research *40*, 11-26.

Ku, S., Soragni, E., Campau, E., Thomas, E.A., Altun, G., Laurent, L.C., Loring, J.F., Napierala, M., and Gottesfeld, J.M. (2010). Friedreich's ataxia induced pluripotent stem cells model intergenerational GAATTC triplet repeat instability. Cell stem cell *7*, 631-637.

Kubicek, S., O'Sullivan, R.J., August, E.M., Hickey, E.R., Zhang, Q., Teodoro, M.L., Rea, S., Mechtler, K., Kowalski, J.A., Homon, C.A., *et al.* (2007). Reversal of H3K9me2 by a small-molecule inhibitor for the G9a histone methyltransferase. Molecular cell *25*, 473-481.

Kulis, M., and Esteller, M. (2010). DNA methylation and cancer. Advances in genetics *70*, 27-56.

Kumari, D., Biacsi, R.E., and Usdin, K. (2011). Repeat expansion affects both transcription initiation and elongation in friedreich ataxia cells. The Journal of biological chemistry *286*, 4209-4215.

Kumari, D., Lokanga, R., Yudkin, D., Zhao, X.N., and Usdin, K. (2012). Chromatin changes in the development and pathology of the Fragile X-associated disorders and Friedreich ataxia. Biochimica et biophysica acta.

Kumari, D., and Usdin, K. (2010). The distribution of repressive histone modifications on silenced FMR1 alleles provides clues to the mechanism of gene silencing in fragile X syndrome. Human molecular genetics *19*, 4634-4642.

Kumari, D., and Usdin, K. (2012). Is Friedreich ataxia an epigenetic disorder? Clin Epigenetics *4*, 2.

Kumari, D., and Usdin, K. (2014). Polycomb group complexes are recruited to reactivated FMR1 alleles in Fragile X syndrome in response to FMR1 transcription. Human molecular genetics *23*, 6575-6583.

Kustatscher, G., Wills, K.L., Furlan, C., and Rappsilber, J. (2014). Chromatin enrichment for proteomics. Nature protocols *9*, 2090-2099.

Kwon, S.C., Yi, H., Eichelbaum, K., Fohr, S., Fischer, B., You, K.T., Castello, A., Krijgsveld, J., Hentze, M.W., and Kim, V.N. (2013). The RNA-binding protein repertoire of embryonic stem cells. Nature structural & molecular biology *20*, 1122-1130.

Ladd, P.D., Smith, L.E., Rabaia, N.A., Moore, J.M., Georges, S.A., Hansen, R.S., Hagerman, R.J., Tassone, F., Tapscott, S.J., and Filippova, G.N. (2007). An antisense transcript spanning the CGG repeat region of FMR1 is upregulated in premutation carriers but silenced in full mutation individuals. Human molecular genetics *16*, 3174-3187.

Lawlor, K.T., O'Keefe, L.V., Samaraweera, S.E., van Eyk, C.L., McLeod, C.J., Maloney, C.A., Dang, T.H., Suter, C.M., and Richards, R.I. (2011). Double-stranded RNA is pathogenic in Drosophila models of expanded repeat neurodegenerative diseases. Human molecular genetics *20*, 3757-3768.

Lazaropoulos, M., Dong, Y., Clark, E., Greeley, N.R., Seyer, L.A., Brigatti, K.W., Christie, C., Perlman, S.L., Wilmot, G.R., Gomez, C.M., *et al.* (2015). Frataxin levels in peripheral tissue in Friedreich ataxia. Annals of clinical and translational neurology *2*, 831-842.

Lazzaro, F., Novarina, D., Amara, F., Watt, D.L., Stone, J.E., Costanzo, V., Burgers, P.M., Kunkel, T.A., Plevani, P., and Muzi-Falconi, M. (2012). RNase H and postreplication repair protect cells from ribonucleotides incorporated in DNA. Mol Cell *45*, 99-110.

Lee-Kirsch, M.A., Wolf, C., and Gunther, C. (2014). Aicardi-Goutieres syndrome: a model disease for systemic autoimmunity. Clinical and experimental immunology *175*, 17-24.

Lee, C.G., Chang, K.A., Kuroda, M.I., and Hurwitz, J. (1997). The NTPase/helicase activities of Drosophila maleless, an essential factor in dosage compensation. The EMBO journal *16*, 2671-2681.

Lee, D.Y., and Clayton, D.A. (1997). RNase mitochondrial RNA processing correctly cleaves a novel R loop at the mitochondrial DNA leading-strand origin of replication. Genes & development *11*, 582-592.

Lesnik, E.A., and Freier, S.M. (1995). Relative thermodynamic stability of DNA, RNA, and DNA:RNA hybrid duplexes: relationship with base composition and structure. Biochemistry *34*, 10807-10815.

Li, X., and Manley, J.L. (2005). Inactivation of the SR protein splicing factor ASF/SF2 results in genomic instability. Cell *122*, 365-378.

Li, X., Niu, T., and Manley, J.L. (2007). The RNA binding protein RNPS1 alleviates ASF/SF2 depletion-induced genomic instability. RNA *13*, 2108-2115.

Li, Y., Polak, U., Bhalla, A.D., Rozwadowska, N., Butler, J.S., Lynch, D.R., Dent, S.Y., and Napierala, M. (2015). Excision of Expanded GAA Repeats Alleviates the Molecular Phenotype of Friedreich's Ataxia. Molecular therapy : the journal of the American Society of Gene Therapy *23*, 1055-1065.

Lim, Y.W., Sanz, L.A., Xu, X., Hartono, S.R., and Chedin, F. (2015). Genome-wide DNA hypomethylation and RNA:DNA hybrid accumulation in Aicardi-Goutieres syndrome. eLife *4*.

Lin, Y., Dent, S.Y., Wilson, J.H., Wells, R.D., and Napierala, M. (2010). R loops stimulate genetic instability of CTG.CAG repeats. Proceedings of the National Academy of Sciences of the United States of America *107*, 692-697.

Lin, Y., Dion, V., and Wilson, J.H. (2006). Transcription promotes contraction of CAG repeat tracts in human cells. Nature structural & molecular biology *13*, 179-180.

Lin, Y., and Wilson, J.H. (2007). Transcription-induced CAG repeat contraction in human cells is mediated in part by transcription-coupled nucleotide excision repair. Molecular and cellular biology *27*, 6209-6217.

Liu, C.R., Chang, C.R., Chern, Y., Wang, T.H., Hsieh, W.C., Shen, W.C., Chang, C.Y., Chu, I.C., Deng, N., Cohen, S.N., *et al.* (2012). Spt4 is selectively required for transcription of extended trinucleotide repeats. Cell *148*, 690-701.

Liu, L.F., and Wang, J.C. (1987). Supercoiling of the DNA template during transcription. Proceedings of the National Academy of Sciences of the United States of America *84*, 7024-7027.

Loesch, D.Z., Godler, D.E., Evans, A., Bui, Q.M., Gehling, F., Kotschet, K.E., Trost, N., Storey, E., Stimpson, P., Kinsella, G., *et al.* (2011). Evidence for the toxicity of bidirectional transcripts and mitochondrial dysfunction in blood associated with small CGG expansions in the FMR1 gene in patients with parkinsonism. Genetics in medicine : official journal of the American College of Medical Genetics *13*, 392-399.

Loomis, E.W., Sanz, L.A., Chedin, F., and Hagerman, P.J. (2014). Transcription-associated R-loop formation across the human FMR1 CGG-repeat region. PLoS genetics *10*, e1004294.

Lopez Castel, A., Cleary, J.D., and Pearson, C.E. (2010). Repeat instability as the basis for human diseases and as a potential target for therapy. Nature reviews Molecular cell biology *11*, 165-170.

Luby, T.M., Schrader, C.E., Stavnezer, J., and Selsing, E. (2001). The mu switch region tandem repeats are important, but not required, for antibody class switch recombination. The Journal of experimental medicine *193*, 159-168.

Lufino, M.M., Silva, A.M., Nemeth, A.H., Alegre-Abarrategui, J., Russell, A.J., and Wade-Martins, R. (2013). A GAA repeat expansion reporter model of Friedreich's ataxia recapitulates the genomic context and allows rapid screening of therapeutic compounds. Hum Mol Genet *22*, 5173-5187.

Maduike, N.Z., Tehranchi, A.K., Wang, J.D., and Kreuzer, K.N. (2014). Replication of the Escherichia coli chromosome in RNase HI-deficient cells: multiple initiation regions and fork dynamics. Molecular microbiology *91*, 39-56.

Marcus, K. (2012). Quantitative Methods in Proteomics, Vol 893.

Marinello, J., Chillemi, G., Bueno, S., Manzo, S.G., and Capranico, G. (2013). Antisense transcripts enhanced by camptothecin at divergent CpG-island promoters associated with bursts of topoisomerase I-DNA cleavage complex and R-loop formation. Nucleic acids research *41*, 10110-10123.

Marmolino, D. (2011). Friedreich's ataxia: past, present and future. Brain research reviews *67*, 311-330.

Masse, E., and Drolet, M. (1999). Escherichia coli DNA topoisomerase I inhibits R-loop formation by relaxing transcription-induced negative supercoiling. The Journal of biological chemistry *274*, 16659-16664.

McElhinny, S.A.N., Kumar, D., Clark, A.B., Watt, D.L., Watts, B.E., Lundstrom, E.B., Johansson, E., Chabes, A., and Kunkel, T.A. (2010). Genome instability due to ribonucleotide incorporation into DNA. Nature chemical biology *6*, 774-781.

McIvor, E.I., Polak, U., and Napierala, M. (2010). New insights into repeat instability: role of RNA*DNA hybrids. RNA Biol *7*, 551-558.

Mechali, M. (2010). Eukaryotic DNA replication origins: many choices for appropriate answers. Nature reviews Molecular cell biology *11*, 728-738.

Mellacheruvu, D., Wright, Z., Couzens, A.L., Lambert, J.P., St-Denis, N.A., Li, T., Miteva, Y.V., Hauri, S., Sardiu, M.E., Low, T.Y., *et al.* (2013). The CRAPome: a contaminant repository for affinity purification-mass spectrometry data. Nature methods *10*, 730-736.

Meng, L., Person, R.E., and Beaudet, A.L. (2012). Ube3a-ATS is an atypical RNA polymerase II transcript that represses the paternal expression of Ube3a. Human molecular genetics *21*, 3001-3012.

Meng, L., Ward, A.J., Chun, S., Bennett, C.F., Beaudet, A.L., and Rigo, F. (2015). Towards a therapy for Angelman syndrome by targeting a long non-coding RNA. Nature *518*, 409-412.

Mi, H., Muruganujan, A., Casagrande, J.T., and Thomas, P.D. (2013). Large-scale gene function analysis with the PANTHER classification system. Nature protocols *8*, 1551-1566.

Miller, M.S., Rialdi, A., Ho, J.S., Tilove, M., Martinez-Gil, L., Moshkina, N.P., Peralta, Z., Noel, J., Melegari, C., Maestre, A.M., *et al.* (2015). Senataxin suppresses the antiviral transcriptional response and controls viral biogenesis. Nature immunology *16*, 485-494.

Min, J.L., Barrett, A., Watts, T., Pettersson, F.H., Lockstone, H.E., Lindgren, C.M., Taylor, J.M., Allen, M., Zondervan, K.T., and McCarthy, M.I. (2010). Variability of gene expression profiles in human blood and lymphoblastoid cell lines. BMC genomics *11*, 96.

Mischo, H.E., Gomez-Gonzalez, B., Grzechnik, P., Rondon, A.G., Wei, W., Steinmetz, L., Aguilera, A., and Proudfoot, N.J. (2011). Yeast Sen1 helicase protects the genome from transcription-associated instability. Molecular cell *41*, 21-32.

Moore, M.J., and Proudfoot, N.J. (2009). Pre-mRNA processing reaches back to transcription and ahead to translation. Cell *136*, 688-700.

Morales, J.C., Richard, P., Rommel, A., Fattah, F.J., Motea, E.A., Patidar, P.L., Xiao, L., Leskov, K., Wu, S.Y., Hittelman, W.N., *et al.* (2014). Kub5-Hera, the human Rtt103 homolog, plays dual functional roles in transcription termination and DNA repair. Nucleic acids research.

Moreira, M.C., Klur, S., Watanabe, M., Nemeth, A.H., Le Ber, I., Moniz, J.C., Tranchant, C., Aubourg, P., Tazir, M., Schols, L., *et al.* (2004). Senataxin, the ortholog of a yeast RNA helicase, is mutant in ataxia-ocular apraxia 2. Nature genetics *36*, 225-227.

Moseley, M.L., Zu, T., Ikeda, Y., Gao, W., Mosemiller, A.K., Daughters, R.S., Chen, G., Weatherspoon, M.R., Clark, H.B., Ebner, T.J., *et al.* (2006). Bidirectional expression of CUG and CAG expansion transcripts and intranuclear polyglutamine inclusions in spinocerebellar ataxia type 8. Nature genetics *38*, 758-769.

Mozzetta, C., Boyarchuk, E., Pontis, J., and Ait-Si-Ali, S. (2015). Sound of silence: the properties and functions of repressive Lys methyltransferases. Nature reviews Molecular cell biology *16*, 499-513.

Mozzetta, C., Pontis, J., Fritsch, L., Robin, P., Portoso, M., Proux, C., Margueron, R., and Ait-Si-Ali, S. (2014). The histone H3 lysine 9 methyltransferases G9a and GLP regulate polycomb repressive complex 2-mediated gene silencing. Molecular cell *53*, 277-289.

Nadel, J., Athanasiadou, R., Lemetre, C., Wijetunga, N.A., P, O.B., Sato, H., Zhang, Z., Jeddeloh, J., Montagna, C., Golden, A., *et al.* (2015). RNA:DNA hybrids in the human genome have distinctive nucleotide characteristics, chromatin composition, and transcriptional relationships. Epigenetics & chromatin *8*, 46.

Nakama, M., Kawakami, K., Kajitani, T., Urano, T., and Murakami, Y. (2012). DNA-RNA hybrid formation mediates RNAi-directed heterochromatin formation. Genes Cells *17*, 218-233.

Nakamori, M., Pearson, C.E., and Thornton, C.A. (2011). Bidirectional transcription stimulates expansion and contraction of expanded (CTG)*(CAG) repeats. Hum Mol Genet *20*, 580-588.

Negrini, S., Gorgoulis, V.G., and Halazonetis, T.D. (2010). Genomic instability--an evolving hallmark of cancer. Nature reviews Molecular cell biology *11*, 220-228.

Nieto Moreno, N., Giono, L.E., Cambindo Botto, A.E., Munoz, M.J., and Kornblihtt, A.R. (2015). Chromatin, DNA structure and alternative splicing. FEBS letters *589*, 3370-3378.

Nossal, N.G., Dudas, K.C., and Kreuzer, K.N. (2001). Bacteriophage T4 proteins replicate plasmids with a preformed R loop at the T4 ori(uvsY) replication origin in vitro. Molecular cell *7*, 31-41.

Ohshima, K., Montermini, L., Wells, R.D., and Pandolfo, M. (1998). Inhibitory effects of expanded GAA.TTC triplet repeats from intron I of the Friedreich ataxia gene on transcription and replication in vivo. The Journal of biological chemistry *273*, 14588-14595.

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., *et al.* (2009). A genome-wide siRNA screen reveals diverse cellular processes and pathways that mediate genome stability. Molecular cell *35*, 228-239.

Pefanis, E., Wang, J., Rothschild, G., Lim, J., Chao, J., Rabadan, R., Economides, A.N., and Basu, U. (2014). Noncoding RNA transcription targets AID to divergently transcribed loci in B cells. Nature *514*, 389-393.

Pefanis, E., Wang, J., Rothschild, G., Lim, J., Kazadi, D., Sun, J., Federation, A., Chao, J., Elliott, O., Liu, Z.P., *et al.* (2015). RNA exosome-regulated long non-coding RNA transcription controls super-enhancer activity. Cell *161*, 774-789.

Pelechano, V., and Steinmetz, L.M. (2013). Gene regulation by antisense transcription. Nature reviews Genetics *14*, 880-893.

Perry, R.P., and Kelley, D.E. (1970). Inhibition of RNA synthesis by actinomycin D: characteristic dose-response of different RNA species. Journal of cellular physiology *76*, 127-139.

Pfeiffer, V., Crittin, J., Grolimund, L., and Lingner, J. (2013). The THO complex component Thp2 counteracts telomeric R-loops and telomere shortening. The EMBO journal *32*, 2861-2871.

Phillips, D.D., Garboczi, D.N., Singh, K., Hu, Z., Leppla, S.H., and Leysath, C.E. (2013). The sub-nanomolar binding of DNA-RNA hybrids by the single-chain Fv fragment of antibody S9.6. Journal of molecular recognition : JMR *26*, 376-381.

Pietrobono, R., Pomponi, M.G., Tabolacci, E., Oostra, B., Chiurazzi, P., and Neri, G. (2002). Quantitative analysis of DNA demethylation and transcriptional reactivation of the FMR1 gene in fragile X cells treated with 5-azadeoxycytidine. Nucleic acids research *30*, 3278-3285.

Porrua, O., and Libri, D. (2013). A bacterial-like mechanism for transcription termination by the Sen1p helicase in budding yeast. Nature structural & molecular biology *20*, 884-891.

Potaman, V.N., Oussatcheva, E.A., Lyubchenko, Y.L., Shlyakhtenko, L.S., Bidichandani, S.I., Ashizawa, T., and Sinden, R.R. (2004). Length-dependent structure formation in Friedreich ataxia (GAA)n*(TTC)n repeats at neutral pH. Nucleic acids research *32*, 1224-1231.

Powell, W.T., Coulson, R.L., Gonzales, M.L., Crary, F.K., Wong, S.S., Adams, S., Ach, R.A., Tsang, P., Yamada, N.A., Yasui, D.H., *et al.* (2013). R-loop formation at Snord116 mediates topotecan inhibition of Ube3a-antisense and allele-specific chromatin decondensation. Proceedings of the National Academy of Sciences of the United States of America *110*, 13938-13943.

Punga, T., and Buhler, M. (2010). Long intronic GAA repeats causing Friedreich ataxia impede transcription elongation. EMBO Mol Med *2*, 120-129.

Rabe, B. (2013). Aicardi-Goutieres syndrome: clues from the RNase H2 knock-out mouse. J Mol Med (Berl) *91*, 1235-1240.

Raschle, M., Smeenk, G., Hansen, R.K., Temu, T., Oka, Y., Hein, M.Y., Nagaraj, N., Long, D.T., Walter, J.C., Hofmann, K., *et al.* (2015). DNA repair. Proteomics reveals dynamic assembly of repair complexes during bypass of DNA cross-links. Science *348*, 1253671.

Reaban, M.E., and Griffin, J.A. (1990). Induction of RNA-stabilized DNA conformers by transcription of an immunoglobulin switch region. Nature *348*, 342-344.

Reddy, K., Tam, M., Bowater, R.P., Barber, M., Tomlinson, M., Nichol Edamura, K., Wang, Y.H., and Pearson, C.E. (2011). Determinants of R-loop formation at convergent bidirectionally transcribed trinucleotide repeats. Nucleic acids research *39*, 1749-1762.

Reijns, M.A., Rabe, B., Rigby, R.E., Mill, P., Astell, K.R., Lettice, L.A., Boyle, S., Leitch, A., Keighren, M., Kilanowski, F., *et al.* (2012). Enzymatic removal of ribonucleotides from DNA is essential for mammalian genome integrity and development. Cell *149*, 1008-1022.

Rhodes, D.R., Kalyana-Sundaram, S., Mahavisno, V., Varambally, R., Yu, J., Briggs, B.B., Barrette, T.R., Anstet, M.J., Kincead-Beal, C., Kulkarni, P., *et al.* (2007). Oncomine 3.0: genes, pathways, and networks in a collection of 18,000 cancer gene expression profiles. Neoplasia *9*, 166-180.

Rice, G.I., Bond, J., Asipu, A., Brunette, R.L., Manfield, I.W., Carr, I.M., Fuller, J.C., Jackson, R.M., Lamb, T., Briggs, T.A., *et al.* (2009). Mutations involved in Aicardi-Goutieres syndrome implicate SAMHD1 as regulator of the innate immune response. Nature genetics *41*, 829-832.

Rice, G.I., del Toro Duany, Y., Jenkinson, E.M., Forte, G.M., Anderson, B.H., Ariaudo, G., Bader-Meunier, B., Baildam, E.M., Battini, R., Beresford, M.W., *et al.* (2014). Gain-of-function mutations in IFIH1 cause a spectrum of human disease phenotypes associated with upregulated type I interferon signaling. Nature genetics *46*, 503-509.

Rice, G.I., Kasher, P.R., Forte, G.M., Mannion, N.M., Greenwood, S.M., Szynkiewicz, M., Dickerson, J.E., Bhaskar, S.S., Zampini, M., Briggs, T.A., *et al.* (2012). Mutations in ADAR1 cause Aicardi-Goutieres syndrome associated with a type I interferon signature. Nature genetics *44*, 1243-1248.

Richard, P., Feng, S., and Manley, J.L. (2013). A SUMO-dependent interaction between Senataxin and the exosome, disrupted in the neurodegenerative disease AOA2, targets the exosome to sites of transcription-induced DNA damage. Genes & development *27*, 2227-2232.

Rigby, R.E., Webb, L.M., Mackenzie, K.J., Li, Y., Leitch, A., Reijns, M.A., Lundie, R.J., Revuelta, A., Davidson, D.J., Diebold, S., *et al.* (2014). RNA:DNA hybrids are a novel molecular pattern sensed by TLR9. The EMBO journal *33*, 542-558.

Rindler, P.M., and Bidichandani, S.I. (2011). Role of transcript and interplay between transcription and replication in triplet-repeat instability in mammalian cells. Nucleic acids research *39*, 526-535.

Robbiani, D.F., and Nussenzweig, M.C. (2013). Chromosome translocation, B cell lymphoma, and activation-induced cytidine deaminase. Annual review of pathology *8*, 79-103.

Roberts, R.W., and Crothers, D.M. (1992). Stability and properties of double and triple helices: dramatic effects of RNA or DNA backbone composition. Science *258*, 1463-1466.

Rocak, S., Emery, B., Tanner, N.K., and Linder, P. (2005). Characterization of the ATPase and unwinding activities of the yeast DEAD-box protein Has1p and the analysis of the roles of the conserved motifs. Nucleic acids research *33*, 999-1009.

Ronai, D., Iglesias-Ussel, M.D., Fan, M., Li, Z., Martin, A., and Scharff, M.D. (2007). Detection of chromatin-associated single-stranded DNA in regions targeted for somatic hypermutation. The Journal of experimental medicine *204*, 181-190.

Rose, A.B. (2008). Intron-mediated regulation of gene expression. Current topics in microbiology and immunology *326*, 277-290.

Roy, D., and Lieber, M.R. (2009). G clustering is important for the initiation of transcription-induced R-loops in vitro, whereas high G density without clustering is sufficient thereafter. Molecular and cellular biology *29*, 3124-3133.

Roy, D., Yu, K., and Lieber, M.R. (2008). Mechanism of R-loop formation at immunoglobulin class switch sequences. Molecular and cellular biology *28*, 50-60.

Roy, D., Zhang, Z., Lu, Z., Hsieh, C.L., and Lieber, M.R. (2010). Competition between the RNA transcript and the nontemplate DNA strand during R-loop formation in vitro: a nick can serve as a strong R-loop initiation site. Molecular and cellular biology *30*, 146-159.

Ruiz, J.F., Gomez-Gonzalez, B., and Aguilera, A. (2011). AID induces double-strand breaks at immunoglobulin switch regions and c-MYC causing chromosomal translocations in yeast THO mutants. PLoS genetics *7*, e1002009.

Rydberg, B., and Game, J. (2002). Excision of misincorporated ribonucleotides in DNA by RNase H (type 2) and FEN-1 in cell-free extracts. Proc Natl Acad Sci U S A *99*, 16654-16659.

Sadelain, M., Papapetrou, E.P., and Bushman, F.D. (2012). Safe harbours for the integration of new DNA in the human genome. Nature reviews Cancer *12*, 51-58.

Sadri-Vakili, G., Bouzou, B., Benn, C.L., Kim, M.O., Chawla, P., Overland, R.P., Glajch, K.E., Xia, E., Qiu, Z., Hersch, S.M., *et al.* (2007). Histones associated with downregulated genes are hypo-acetylated in Huntington's disease models. Hum Mol Genet *16*, 1293-1306.

Sakamoto, N., Chastain, P.D., Parniewski, P., Ohshima, K., Pandolfo, M., Griffith, J.D., and Wells, R.D. (1999). Sticky DNA: self-association properties of long GAA.TTC repeats in R.R.Y triplex structures from Friedreich's ataxia. Molecular cell *3*, 465-475.

Sakamoto, N., Ohshima, K., Montermini, L., Pandolfo, M., and Wells, R.D. (2001). Sticky DNA, a self-associated complex formed at long GAA*TTC repeats in intron 1 of the frataxin gene, inhibits transcription. The Journal of biological chemistry *276*, 27171-27177.

Sambrook, J., and Russell, D.W. (2006). Preparation and Transformation of Competent E. coli Using Calcium Chloride. CSH protocols *2006*.

San Filippo, J., Sung, P., and Klein, H. (2008). Mechanism of eukaryotic homologous recombination. Annual review of biochemistry *77*, 229-257.

Santoro, M.R., Bray, S.M., and Warren, S.T. (2011). Molecular mechanisms of fragile X syndrome: a twenty-year perspective. Annu Rev Pathol *7*, 219-245.

Santos-Pereira, J.M., and Aguilera, A. (2015). R loops: new modulators of genome dynamics and function. Nature reviews Genetics *16*, 583-597.

Santos-Pereira, J.M., Herrero, A.B., Garcia-Rubio, M.L., Marin, A., Moreno, S., and Aguilera, A. (2013). The Npl3 hnRNP prevents R-loop-mediated transcription-replication conflicts and genome instability. Genes & development *27*, 2445-2458.

Sathasivam, K., Neueder, A., Gipson, T.A., Landles, C., Benjamin, A.C., Bondulich, M.K., Smith, D.L., Faull, R.L., Roos, R.A., Howland, D., *et al.* (2013). Aberrant splicing of HTT generates the pathogenic exon 1 protein in Huntington disease. Proceedings of the National Academy of Sciences of the United States of America *110*, 2366-2370.

Saveliev, A., Everett, C., Sharpe, T., Webster, Z., and Festenstein, R. (2003). DNA triplet repeats mediate heterochromatin-protein-1-sensitive variegated gene silencing. Nature *422*, 909-913.

Schwab, R.A., Nieminuszczy, J., Shah, F., Langton, J., Lopez Martinez, D., Liang, C.C., Cohn, M.A., Gibbons, R.J., Deans, A.J., and Niedzwiedz, W. (2015). The Fanconi Anemia Pathway Maintains Genome Stability by Coordinating Replication and Transcription. Molecular cell *60*, 351-361.

Sellier, C., Freyermuth, F., Tabet, R., Tran, T., He, F., Ruffenach, F., Alunni, V., Moine, H., Thibault, C., Page, A., *et al.* (2013). Sequestration of DROSHA and DGCR8 by expanded CGG RNA repeats alters microRNA processing in fragile X-associated tremor/ataxia syndrome. Cell reports *3*, 869-880.

Sellier, C., Rau, F., Liu, Y., Tassone, F., Hukema, R.K., Gattoni, R., Schneider, A., Richard, S., Willemsen, R., Elliott, D.J., *et al.* (2010). Sam68 sequestration and partial loss of function are associated with splicing alterations in FXTAS patients. The EMBO journal *29*, 1248-1261.

Shaw, N.N., and Arya, D.P. (2008). Recognition of the unique structure of DNA:RNA hybrids. Biochimie *90*, 1026-1039.

Shin, J.H., and Kelman, Z. (2006). The replicative helicases of bacteria, archaea, and eukarya can unwind RNA-DNA hybrid substrates. The Journal of biological chemistry *281*, 26914-26921.

Shishkin, A.A., Voineagu, I., Matera, R., Cherng, N., Chernet, B.T., Krasilnikova, M.M., Narayanan, V., Lobachev, K.S., and Mirkin, S.M. (2009). Large-scale expansions of Friedreich's ataxia GAA repeats in yeast. Molecular cell *35*, 82-92.

Shu, Z., Vijayakumar, S., Chen, C.F., Chen, P.L., and Lee, W.H. (2004). Purified human SUV3p exhibits multiple-substrate unwinding activity upon conformational change. Biochemistry *43*, 4781-4790.

Sikdar, N., Banerjee, S., Zhang, H., Smith, S., and Myung, K. (2008). Spt2p defines a new transcription-dependent gross chromosomal rearrangement pathway. PLoS genetics *4*, e1000290.

Simon, M.D., Wang, C.I., Kharchenko, P.V., West, J.A., Chapman, B.A., Alekseyenko, A.A., Borowsky, M.L., Kuroda, M.I., and Kingston, R.E. (2011). The genomic binding sites of a noncoding RNA. Proceedings of the National Academy of Sciences of the United States of America *108*, 20497-20502.

Sinkunas, T., Gasiunas, G., Fremaux, C., Barrangou, R., Horvath, P., and Siksnys, V. (2011). Cas3 is a single-stranded DNA nuclease and ATP-dependent helicase in the CRISPR/Cas immune system. The EMBO journal *30*, 1335-1342.

Skourti-Stathaki, K., Kamieniarz-Gdula, K., and Proudfoot, N.J. (2014). R-loops induce repressive chromatin marks over mammalian gene terminators. Nature.

Skourti-Stathaki, K., and Proudfoot, N.J. (2014). A double-edged sword: R loops as threats to genome integrity and powerful regulators of gene expression. Genes & development *28*, 1384-1396.

Skourti-Stathaki, K., Proudfoot, N.J., and Gromak, N. (2011). Human senataxin resolves RNA/DNA hybrids formed at transcriptional pause sites to promote Xrn2-dependent termination. Molecular cell *42*, 794-805.

Smyth, G.K. (2004). Linear models and empirical bayes methods for assessing differential expression in microarray experiments. Statistical applications in genetics and molecular biology *3*, Article3.

Sollier, J., Stork, C.T., Garcia-Rubio, M.L., Paulsen, R.D., Aguilera, A., and Cimprich, K.A. (2014). Transcription-coupled nucleotide excision repair factors promote R-loop-induced genome instability. Molecular cell *56*, 777-785.

Sopher, B.L., Ladd, P.D., Pineda, V.V., Libby, R.T., Sunkin, S.M., Hurley, J.B., Thienes, C.P., Gaasterland, T., Filippova, G.N., and La Spada, A.R. (2011). CTCF regulates ataxin-7 expression through promotion of a convergently transcribed, antisense noncoding RNA. Neuron *70*, 1071-1084.

Soragni, E., Herman, D., Dent, S.Y., Gottesfeld, J.M., Wells, R.D., and Napierala, M. (2008). Long intronic GAA*TTC repeats induce epigenetic changes and reporter gene silencing in a molecular model of Friedreich ataxia. Nucleic acids research *36*, 6056-6065.

Soragni, E., Miao, W., Iudicello, M., Jacoby, D., De Mercanti, S., Clerico, M., Longo, F., Piga, A., Ku, S., Campau, E., *et al.* (2014). Epigenetic therapy for Friedreich ataxia. Annals of neurology *76*, 489-508.

Sordet, O., Redon, C.E., Guirouilh-Barbat, J., Smith, S., Solier, S., Douarre, C., Conti, C., Nakamura, A.J., Das, B.B., Nicolas, E., *et al.* (2009). Ataxia telangiectasia mutated activation by transcription- and topoisomerase I-induced DNA double-strand breaks. EMBO reports *10*, 887-893.

Stavnezer, J., Guikema, J.E., and Schrader, C.E. (2008). Mechanism and regulation of class switch recombination. Annual review of immunology *26*, 261-292.

Steinmetz, E.J., Warren, C.L., Kuehner, J.N., Panbehi, B., Ansari, A.Z., and Brow, D.A. (2006). Genome-wide distribution of yeast RNA polymerase II and its control by Sen1 helicase. Mol Cell *24*, 735-746.

Stirling, P.C., Chan, Y.A., Minaker, S.W., Aristizabal, M.J., Barrett, I., Sipahimalani, P., Kobor, M.S., and Hieter, P. (2012). R-loop-mediated genome instability in mRNA cleavage and polyadenylation mutants. Genes & development *26*, 163-175.

Sun, Q., Csorba, T., Skourti-Stathaki, K., Proudfoot, N.J., and Dean, C. (2013). R-loop stabilization represses antisense transcription at the Arabidopsis FLC locus. Science *340*, 619-621.

Suraweera, A., Becherel, O.J., Chen, P., Rundle, N., Woods, R., Nakamura, J., Gatei, M., Criscuolo, C., Filla, A., Chessa, L., *et al.* (2007). Senataxin, defective in ataxia oculomotor apraxia type 2, is involved in the defense against oxidative DNA damage. J Cell Biol *177*, 969-979.

Suraweera, A., Lim, Y., Woods, R., Birrell, G.W., Nasim, T., Becherel, O.J., and Lavin, M.F. (2009). Functional role for senataxin, defective in ataxia oculomotor apraxia type 2, in transcriptional regulation. Human molecular genetics *18*, 3384-3396.

Suzuki, N., Shimamoto, A., Imamura, O., Kuromitsu, J., Kitao, S., Goto, M., and Furuichi, Y. (1997). DNA helicase activity in Werner's syndrome gene product synthesized in a baculovirus system. Nucleic acids research *25*, 2973-2978.

Tabolacci, E., De Pascalis, I., Accadia, M., Terracciano, A., Moscato, U., Chiurazzi, P., and Neri, G. (2008a). Modest reactivation of the mutant FMR1 gene by valproic acid is accompanied by histone modifications but not DNA demethylation. Pharmacogenetics and genomics *18*, 738-741.

Tabolacci, E., Moscato, U., Zalfa, F., Bagni, C., Chiurazzi, P., and Neri, G. (2008b). Epigenetic analysis reveals a euchromatic configuration in the FMR1 unmethylated full mutations. European journal of human genetics : EJHG *16*, 1487-1498.

Tabolacci, E., Pietrobono, R., Moscato, U., Oostra, B.A., Chiurazzi, P., and Neri, G. (2005). Differential epigenetic modifications in the FMR1 gene of the fragile X syndrome after reactivating pharmacological treatments. European journal of human genetics : EJHG *13*, 641-648.

Tacheny, A., Dieu, M., Arnould, T., and Renard, P. (2013). Mass spectrometry-based identification of proteins interacting with nucleic acids. Journal of proteomics *94*, 89-109.

Tadokoro, T., and Kanaya, S. (2009). Ribonuclease H: molecular diversities, substrate binding domains, and catalytic mechanism of the prokaryotic enzymes. The FEBS journal *276*, 1482-1493.

Tassone, F., Beilina, A., Carosi, C., Albertosi, S., Bagni, C., Li, L., Glover, K., Bentley, D., and Hagerman, P.J. (2007). Elevated FMR1 mRNA in premutation carriers is due to increased transcription. RNA *13*, 555-562.

Thakur, S.S., Geiger, T., Chatterjee, B., Bandilla, P., Frohlich, F., Cox, J., and Mann, M. (2011). Deep and highly sensitive proteome coverage by LC-MS/MS without prefractionation. Molecular & cellular proteomics : MCP *10*, M110 003699.

Thomas, M., White, R.L., and Davis, R.W. (1976). Hybridization of RNA to double-stranded DNA: formation of R-loops. Proceedings of the National Academy of Sciences of the United States of America *73*, 2294-2298.

Thomson, J.P., Skene, P.J., Selfridge, J., Clouaire, T., Guy, J., Webb, S., Kerr, A.R., Deaton, A., Andrews, R., James, K.D., *et al.* (2010). CpG islands influence chromatin structure via the CpG-binding protein Cfp1. Nature *464*, 1082-1086.

Tian, M., and Alt, F.W. (2000). Transcription-induced cleavage of immunoglobulin switch regions by nucleotide excision repair nucleases in vitro. The Journal of biological chemistry *275*, 24163-24172.

Todd, P.K., Oh, S.Y., Krans, A., Pandey, U.B., Di Prospero, N.A., Min, K.T., Taylor, J.P., and Paulson, H.L. (2010). Histone deacetylases suppress CGG repeat-induced neurodegeneration via transcriptional silencing in models of fragile X tremor ataxia syndrome. PLoS genetics *6*, e1001240.

Tous, C., and Aguilera, A. (2007). Impairment of transcription elongation by R-loops in vitro. Biochemical and biophysical research communications *360*, 428-432.

Trudgian, D.C., Ridlova, G., Fischer, R., Mackeen, M.M., Ternette, N., Acuto, O., Kessler, B.M., and Thomas, B. (2011). Comparative evaluation of label-free SINQ normalized spectral index quantitation in the central proteomics facilities pipeline. Proteomics *11*, 2790-2797.

Trudgian, D.C., Thomas, B., McGowan, S.J., Kessler, B.M., Salek, M., and Acuto, O. (2010). CPFP: a central proteomics facilities pipeline. Bioinformatics *26*, 1131-1132.

Tuduri, S., Crabbe, L., Conti, C., Tourriere, H., Holtgreve-Grez, H., Jauch, A., Pantesco, V., De Vos, J., Thomas, A., Theillet, C., *et al.* (2009). Topoisomerase I suppresses genomic instability by preventing interference between replication and transcription. Nature cell biology *11*, 1315-1324.

Vantaggiato, C., Bondioni, S., Airoldi, G., Bozzato, A., Borsani, G., Rugarli, E.I., Bresolin, N., Clementi, E., and Bassi, M.T. (2011). Senataxin modulates neurite growth through fibroblast growth factor 8 signalling. Brain : a journal of neurology *134*, 1808-1828.

Vincent, S.D., Mahdi, A.A., and Lloyd, R.G. (1996). The RecG branch migration protein of Escherichia coli dissociates R-loops. Journal of molecular biology *264*, 713-721.

Wagschal, A., Rousset, E., Basavarajaiah, P., Contreras, X., Harwig, A., Laurent-Chabalier, S., Nakamura, M., Chen, X., Zhang, K., Meziane, O., *et al.* (2012). Microprocessor, Setx, Xrn2, and Rrp6 co-operate to induce premature termination of transcription by RNAPII. Cell *150*, 1147-1157.

Wahba, L., Amon, J.D., Koshland, D., and Vuica-Ross, M. (2011). RNase H and multiple RNA biogenesis factors cooperate to prevent RNA:DNA hybrids from generating genome instability. Molecular cell *44*, 978-988.

Wahba, L., Gore, S.K., and Koshland, D. (2013). The homologous recombination machinery modulates the formation of RNA-DNA hybrids and associated chromosome instability. eLife *2*, e00505.

Wan, Y., Zheng, X., Chen, H., Guo, Y., Jiang, H., He, X., Zhu, X., and Zheng, Y. (2015). Splicing function of mitotic regulators links R-loop-mediated DNA damage to tumor cell killing. The Journal of cell biology *209*, 235-246.

Wang, Y.H., Gellibolian, R., Shimizu, M., Wells, R.D., and Griffith, J. (1996). Long CCG triplet repeat blocks exclude nucleosomes: a possible mechanism for the nature of fragile sites in chromosomes. Journal of molecular biology *263*, 511-516.

Watson, J.D., Baker, T.A., Bell, S.P., Gann, A., Levine, M., and Losick, R. (2013). Molecular Biology of the Gene, Vol 1 (Benjamin Cummings).

Wells, R.D. (2008). DNA triplexes and Friedreich ataxia. FASEB J *22*, 1625-1634.

West, S., Gromak, N., and Proudfoot, N.J. (2004). Human 5' --> 3' exonuclease Xrn2 promotes transcription termination at co-transcriptional cleavage sites. Nature *432*, 522-525.

White, R.L., and Hogness, D.S. (1977). R loop mapping of the 18S and 28S sequences in the long and short repeating units of Drosophila melanogaster rDNA. Cell *10*, 177-192.

Wilburn, B., Rudnicki, D.D., Zhao, J., Weitz, T.M., Cheng, Y., Gu, X., Greiner, E., Park, C.S., Wang, N., Sopher, B.L., *et al.* (2011). An antisense CAG repeat transcript at JPH3 locus mediates expanded polyglutamine protein toxicity in Huntington's disease-like 2 mice. Neuron *70*, 427-440.

Wilson, M.D., Harreman, M., and Svejstrup, J.Q. (2013). Ubiquitylation and degradation of elongating RNA polymerase II: the last resort. Biochimica et biophysica acta *1829*, 151-157.

Windhager, L., Bonfert, T., Burger, K., Ruzsics, Z., Krebs, S., Kaufmann, S., Malterer, G., L'Hernault, A., Schilhabel, M., Schreiber, S., *et al.* (2012). Ultrashort and progressive 4sU-tagging reveals key characteristics of RNA processing at nucleotide resolution. Genome research *22*, 2031-2042.

Wisniewski, J.R., Zougman, A., Nagaraj, N., and Mann, M. (2009). Universal sample preparation method for proteome analysis. Nature methods *6*, 359-362.

Wohrle, D., Salat, U., Hameister, H., Vogel, W., and Steinbach, P. (2001). Demethylation, reactivation, and destabilization of human fragile X full-mutation alleles in mouse embryocarcinoma cells. American journal of human genetics *69*, 504-515.

Wollerton, M.C., Gooding, C., Wagner, E.J., Garcia-Blanco, M.A., and Smith, C.W. (2004). Autoregulation of polypyrimidine tract binding protein by alternative splicing leading to nonsense-mediated decay. Mol Cell *13*, 91-100.

Wongsurawat, T., Jenjaroenpun, P., Kwoh, C.K., and Kuznetsov, V. (2012). Quantitative model of R-loop forming structures reveals a novel level of RNA-DNA interactome complexity. Nucleic acids research *40*, e16.

Wu, H.Y., Shyy, S.H., Wang, J.C., and Liu, L.F. (1988). Transcription generates positively and negatively supercoiled domains in the template. Cell *53*, 433-440.

Xu, B., and Clayton, D.A. (1995). A persistent RNA-DNA hybrid is formed during transcription at a phylogenetically conserved mitochondrial DNA sequence. Molecular and cellular biology *15*, 580-589.

Xu, B., and Clayton, D.A. (1996). RNA-DNA hybrid formation at the human mitochondrial heavy-strand origin ceases at replication start sites: an implication for RNA-DNA hybrids serving as primers. The EMBO journal *15*, 3135-3143.

Xue, K., Rada, C., and Neuberger, M.S. (2006). The in vivo pattern of AID targeting to immunoglobulin switch regions deduced from mutation spectra in msh2-/- ung-/- mice. The Journal of experimental medicine *203*, 2085-2094.

Yang, Y., McBride, K.M., Hensley, S., Lu, Y., Chedin, F., and Bedford, M.T. (2014). Arginine methylation facilitates the recruitment of TOP3B to chromatin to prevent R loop accumulation. Molecular cell *53*, 484-497.

Yeo, A.J., Becherel, O.J., Luff, J.E., Cullen, J.K., Wongsurawat, T., Jenjaroenpoon, P., Kuznetsov, V.A., McKinnon, P.J., and Lavin, M.F. (2014). R-Loops in Proliferating Cells but Not in the Brain: Implications for AOA2 and Other Autosomal Recessive Ataxias. PloS one *9*, e90219.

Yu, K., Chedin, F., Hsieh, C.L., Wilson, T.E., and Lieber, M.R. (2003). R-loops at immunoglobulin class switch regions in the chromosomes of stimulated B cells. Nature immunology *4*, 442-451.

Yu, Z., Teng, X., and Bonini, N.M. (2011). Triplet repeat-derived siRNAs enhance RNA-mediated toxicity in a Drosophila model for myotonic dystrophy. PLoS genetics *7*, e1001340.

Yuce, O., and West, S.C. (2013). Senataxin, defective in the neurodegenerative disorder ataxia with oculomotor apraxia 2, lies at the interface of transcription and the DNA damage response. Molecular and cellular biology *33*, 406-417.

Zarrin, A.A., Alt, F.W., Chaudhuri, J., Stokes, N., Kaushal, D., Du Pasquier, L., and Tian, M. (2004). An evolutionarily conserved target motif for immunoglobulin class-switch recombination. Nature immunology *5*, 1275-1281.

Zhang, D.H., Zhou, B., Huang, Y., Xu, L.X., and Zhou, J.Q. (2006). The human Pif1 helicase, a potential Escherichia coli RecD homologue, inhibits telomerase activity. Nucleic acids research *34*, 1393-1404.

Zhang, Y., Shishkin, A.A., Nishida, Y., Marcinkowski-Desmond, D., Saini, N., Volkov, K.V., Mirkin, S.M., and Lobachev, K.S. (2012). Genome-wide screen identifies pathways that govern GAA/TTC repeat fragility and expansions in dividing and nondividing yeast cells. Mol Cell *48*, 254-265.

Zhang, Z.Z., Pannunzio, N.R., Han, L., Hsieh, C.L., Yu, K., and Lieber, M.R. (2014a). The strength of an Ig switch region is determined by its ability to drive R loop formation and its number of WGCW sites. Cell reports *8*, 557-569.

Zhang, Z.Z., Pannunzio, N.R., Hsieh, C.L., Yu, K., and Lieber, M.R. (2014b). The role of G-density in switch region repeats for immunoglobulin class switch recombination. Nucleic acids research *42*, 13186-13193.

Zhang, Z.Z., Pannunzio, N.R., Hsieh, C.L., Yu, K., and Lieber, M.R. (2015). Complexities due to single-stranded RNA during antibody detection of genomic rna:dna hybrids. BMC research notes *8*, 127.

Zhao, D.Y., Gish, G., Braunschweig, U., Li, Y., Ni, Z., Schmitges, F.W., Zhong, G., Liu, K., Li, W., Moffat, J., *et al.* (2015). SMN and symmetric arginine dimethylation of RNA polymerase II C-terminal domain control termination. Nature.

Zheng, R., Shen, Z., Tripathi, V., Xuan, Z., Freier, S.M., Bennett, C.F., Prasanth, S.G., and Prasanth, K.V. (2010). Polypurine-repeat-containing RNAs: a novel class of long non-coding RNA in mammalian cells. Journal of cell science *123*, 3734-3744.

Zhou, Y., Shan, Y., Zhang, L., and Zhang, Y. (2014). Recent advances in stable isotope labeling based techniques for proteome relative quantification. Journal of chromatography A *1365*, 1-11.

Appendix

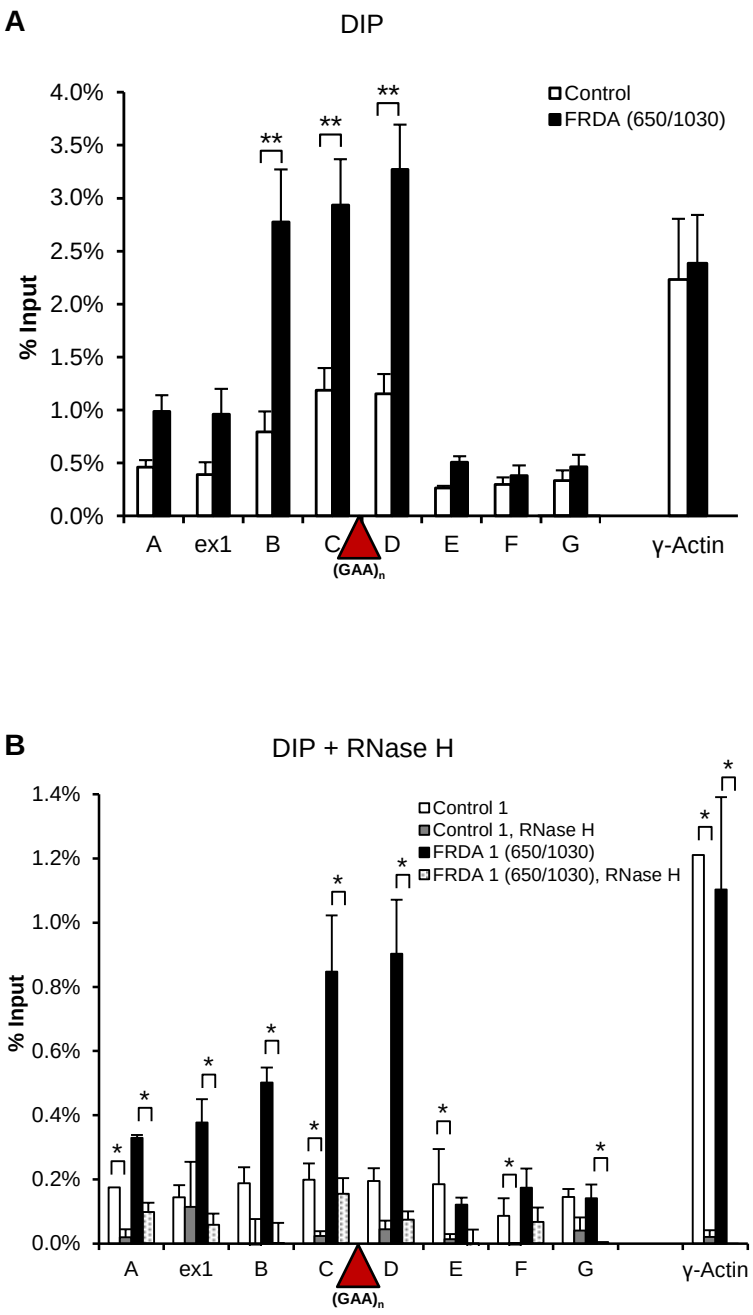


Figure A.1., related to Figure 3.6

RNA/DNA hybrids are formed over expanded repeats of *FXN* gene in FRDA cells

A. DIP on endogenous *FXN* gene in control 1 (GM15851) and FRDA 1 (GM15850) cells. The first intron of *γ-Actin* gene is the positive control.

B. RNA/DNA hybrids are sensitive to RNase H digestion. DIP samples were treated with 25U of recombinant *E. coli* RNase H (NEB, M0297S) for 6 hours at 37ºC prior to addition of S9.6 antibody. The *γ-Actin* gene intron 1 probe serves as positive control.

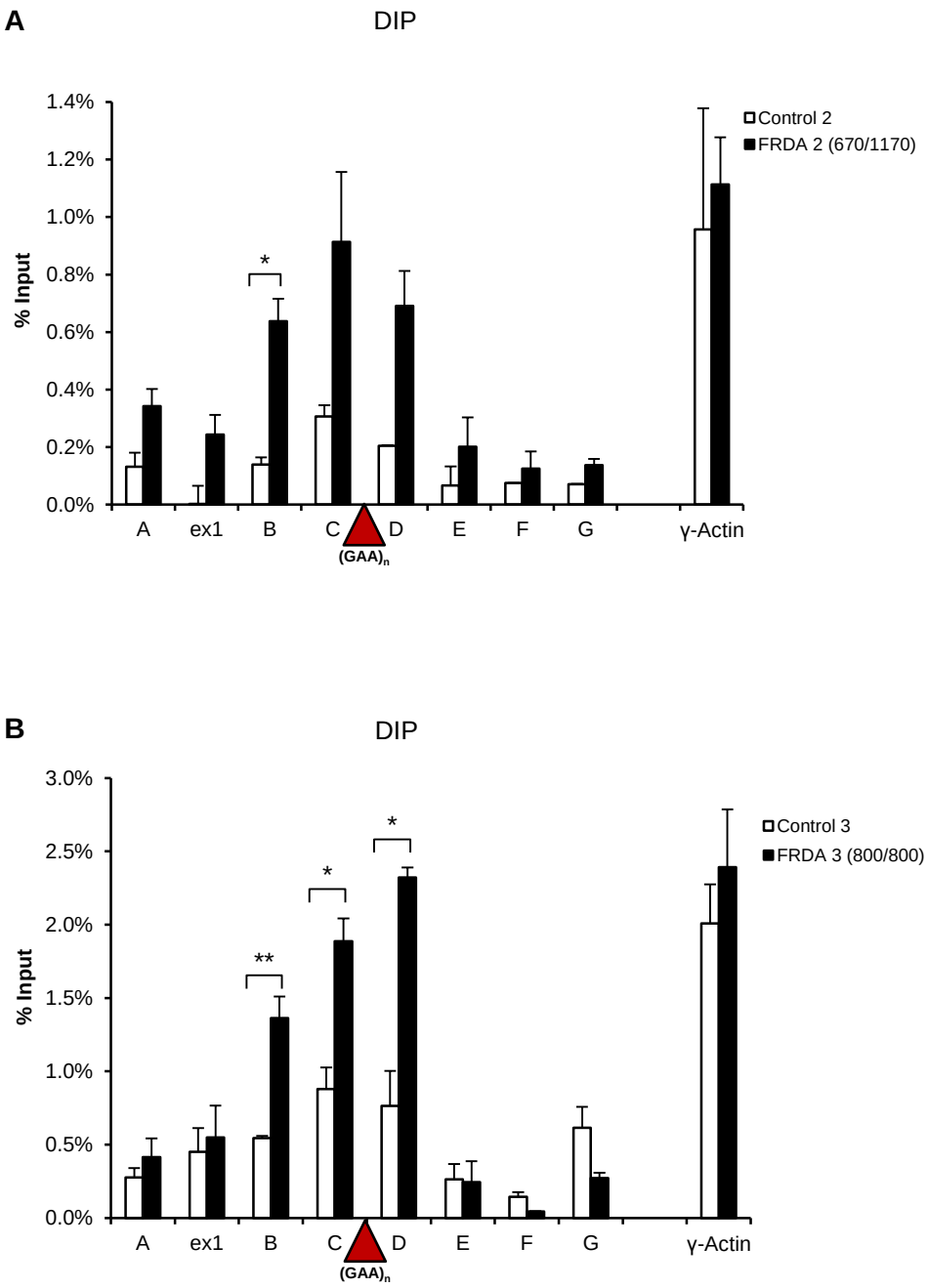


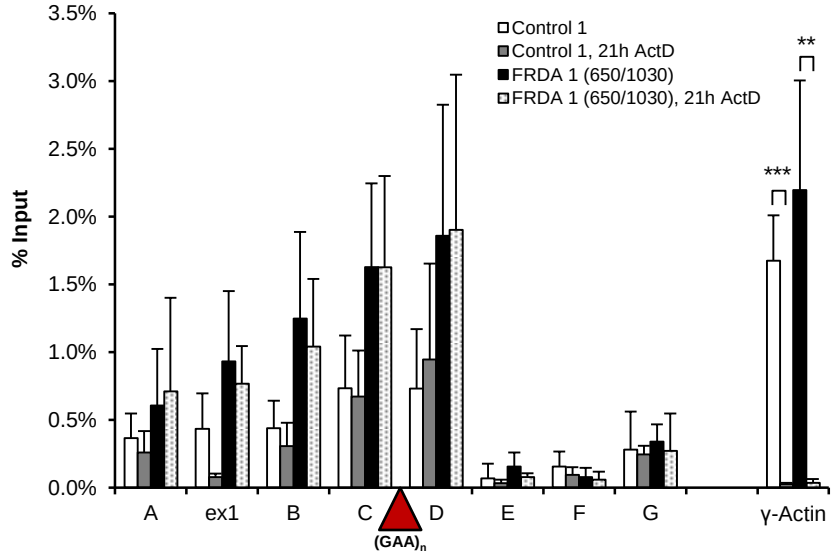
Figure A.2., related to Figure 3.7

Confirmation of RNA/DNA hybrid formation in additional cell lines

A. DIP on endogenous *FXN* gene in control 2 (GM06895) and FRDA 2 (GM16209) cells. The *γ-Actin* gene serves as positive control.

B. DIP on endogenous *FXN* gene in control 3 (GM14926) and FRDA 3 (GM16243) cells. The *γ-Actin* gene serves as positive control.

A DIP + 21h Actinomycin D



B DIP + 48h Actinomycin D

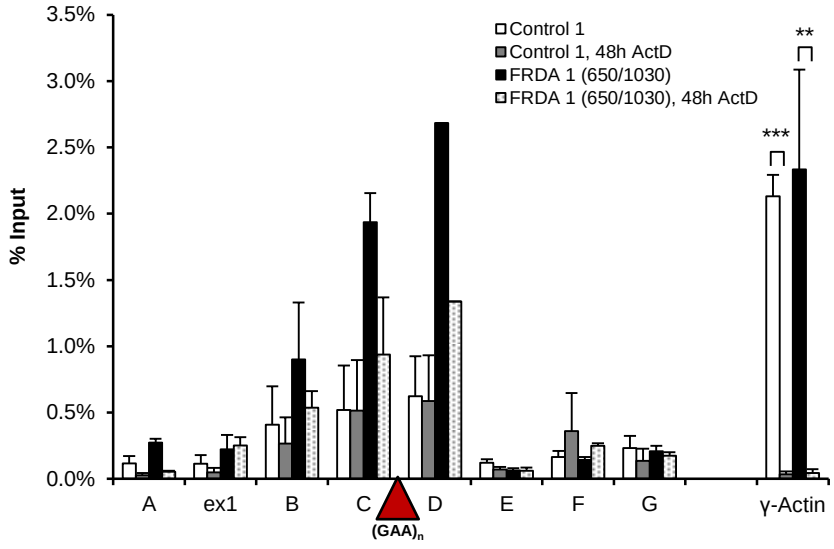


Figure A.3., related to Figure 3.9

RNA/DNA hybrids formed at expanded GAA repeats in *FXN* gene are stable *in vivo*

A. DIP on *FXN* gene in control 1 and FRDA 1 cells treated with 5 μg/ml of actinomycin D for 21 hours. *γ-Actin* is positive control.

B. DIP on *FXN* gene in control 1 and FRDA 1 cells treated with 5 μg/ml of actinomycin D for 48 hours. *γ-Actin* is positive control.

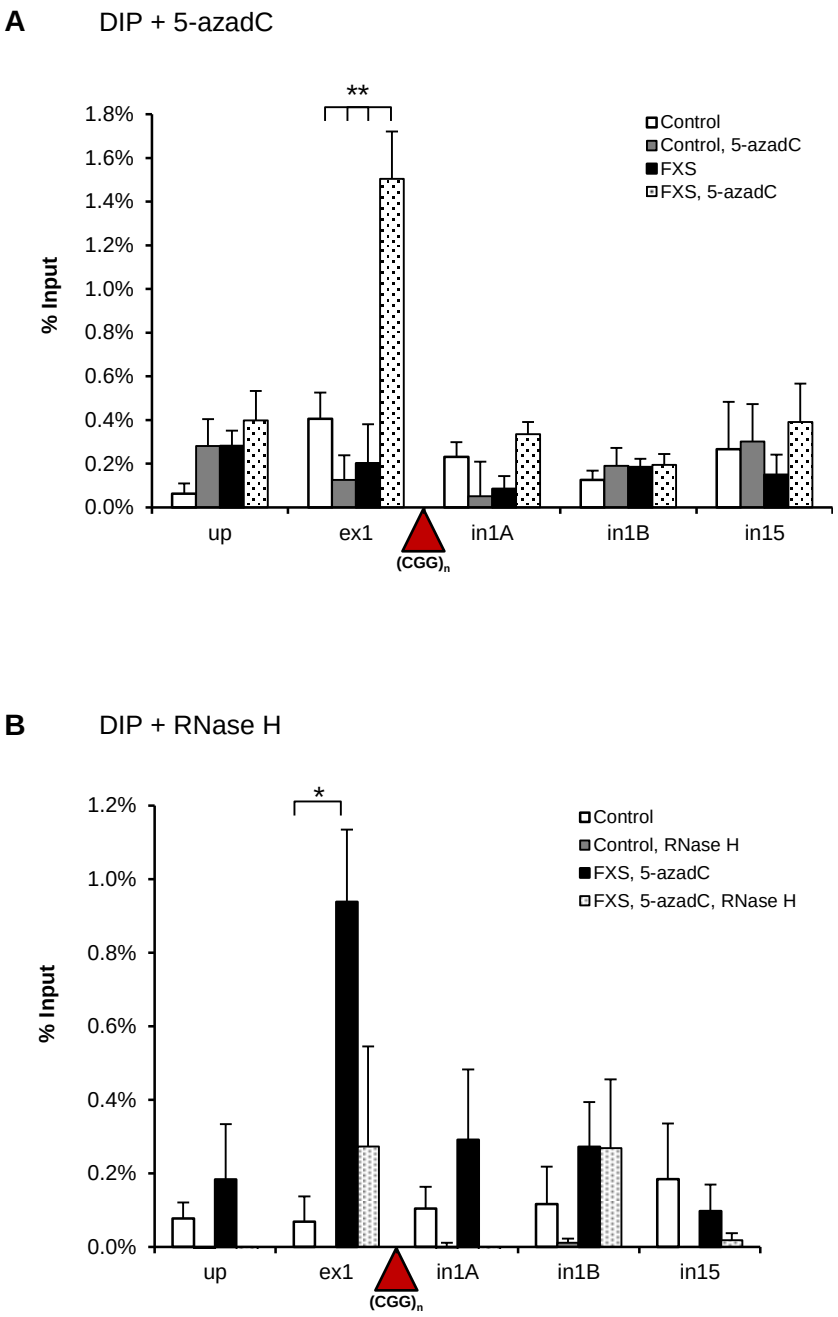


Figure A.4., related to Figure 3.11

RNA/DNA hybrids are formed over CGG expanded repeats of *FMR1* gene

A. DIP analysis on endogenous *FMR1* gene in control and FXS cells, treated with 1 µM 5-azadC for 7 days.

B. *FMR1* RNA/DNA hybrids are sensitive to RNase H digestion, following the treatment with 25U of *E. coli* RNase H for 6 hours at 37ºC prior to immunoprecipitation.

A DIP + 5-azadC washout

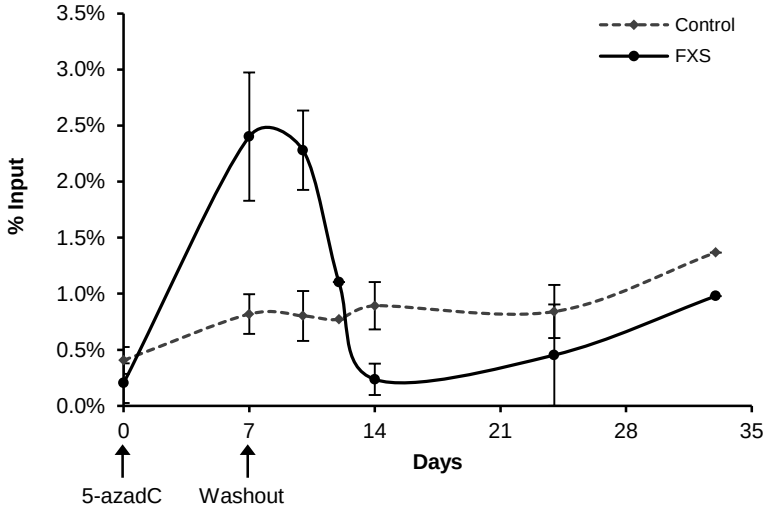


Figure A.5., related to Figure 3.12

Kinetic analysis of RNA/DNA hybrid formation and mRNA levels of *FMR1* gene

A. RNA/DNA hybrid kinetics on exon 1 of *FMR1* gene in control and FXS cells during the process of transcriptional reactivation with 1 µM 5-azadC (7 days) followed by 5-azadC wash out with drug-free media (28 days).

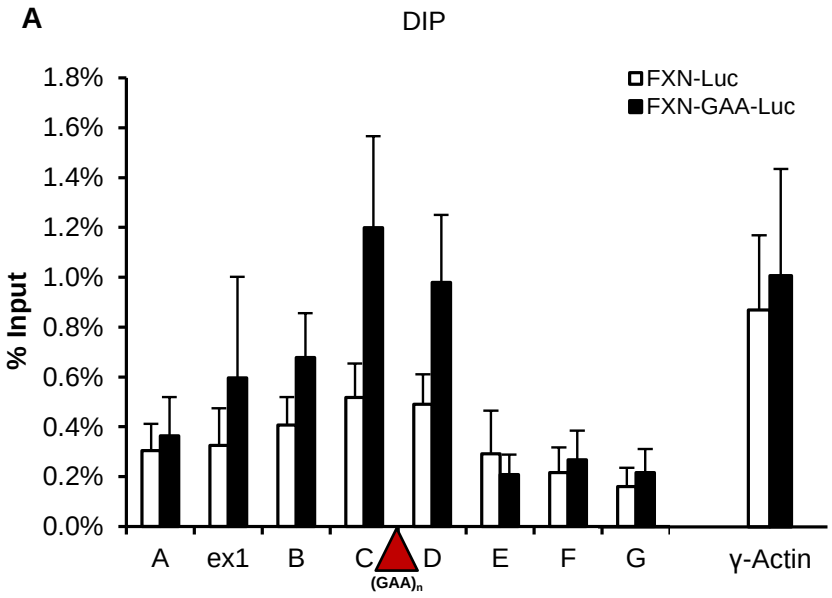


Figure A.6., related to Figure 4.2.

H3K9me2 and RNA/DNA hybrid enrichment at GAA expansions in HEK293 cells

A. DIP analysis on *FXN-Luc* gene in FXN-Luc (white bars) and FXN-GAA-Luc (black bars) HEK293 cells using RNA/DNA hybrid-specific S9.6 antibody.

A DIP + RNase H1 overexpression

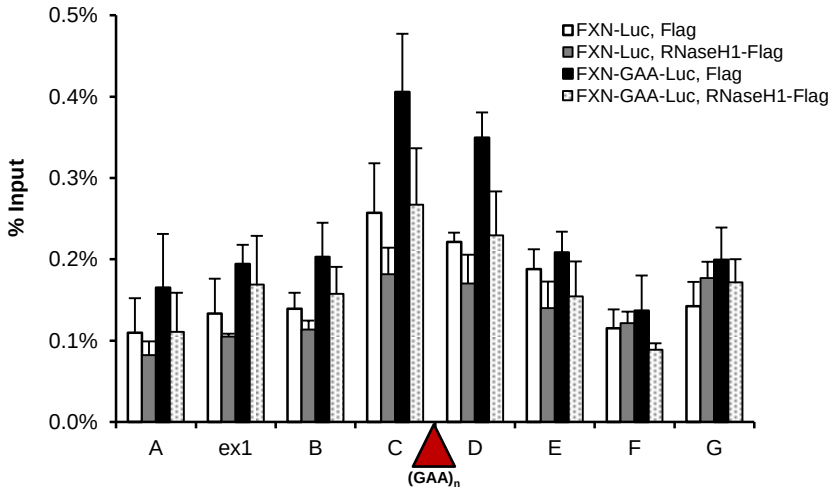


Figure A.7., related to Figure 4.4.

RNA/DNA hybrids at GAA repeats are sensitive to RNase H1 overexpression.

A. DIP analysis on *FXN-Luc* gene in FXN-Luc and FXN-GAA-Luc HEK293 cells transfected with Flag or RNase H1-Flag expression plasmids.

A DIP + BIX-01294

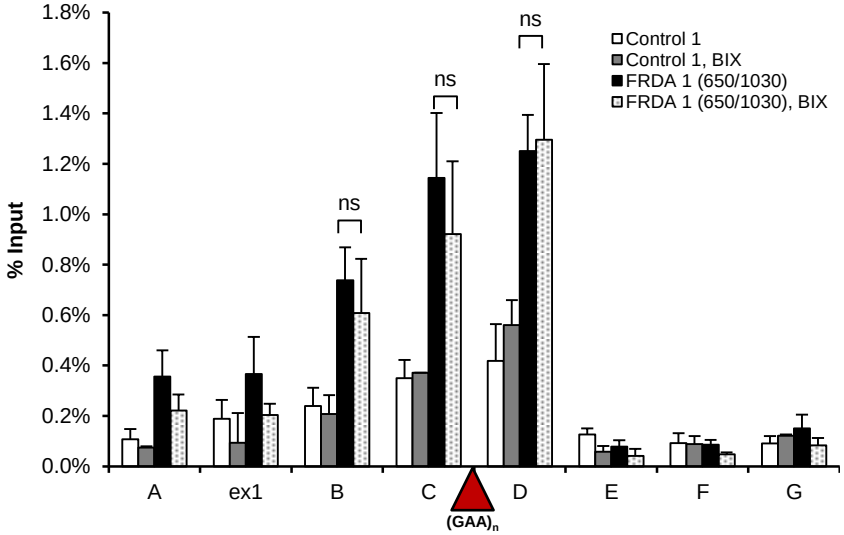


Figure A.8., related to Figure 4.5.

RNA/DNA hybrids and *FXN* transcription are not affected by changes in H3K9me2

A. DIP analysis in control 1 and FRDA 1 cells, treated with 4 µM BIX-01294 for 72 h.

A DIP + Camptothecin

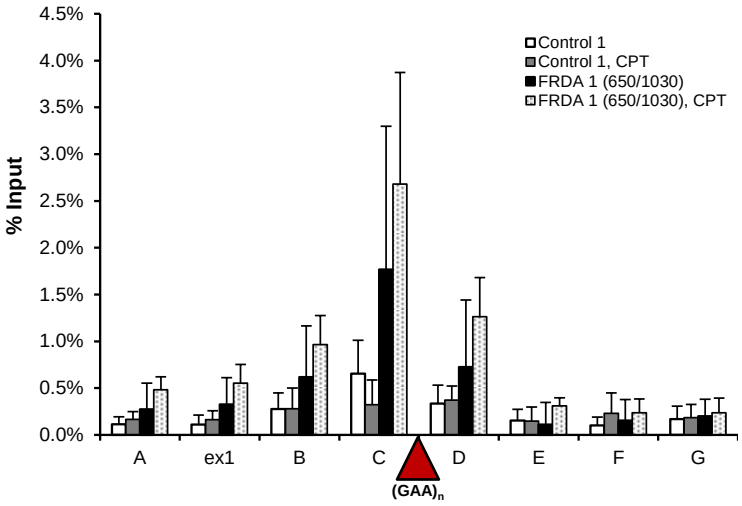


Figure A.9., related to Figure 4.6.

R-loops promote H3K9me2 and transcriptional repression of *FXN* gene

A. DIP analysis on *FXN* gene in control 1 (GM15851) and FRDA 1 (GM15850) cells, treated with 10 μM CPT for 6 hours.

A DIP + Camptothecin

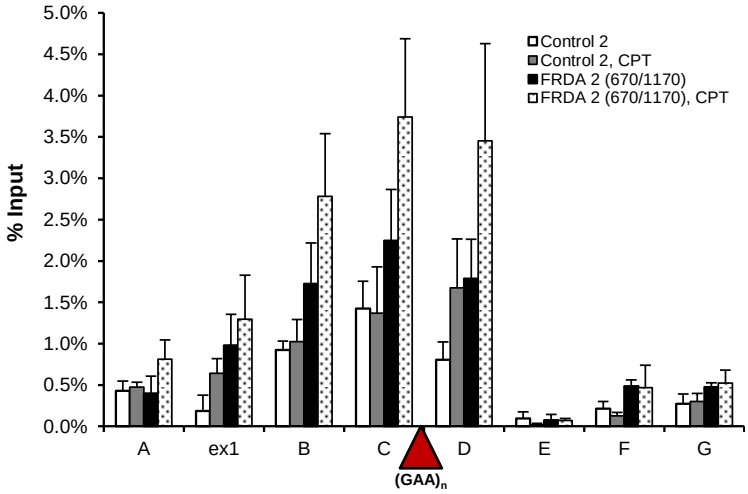


Figure A.10., related to Figure 4.7.

Camptothecin experiments in additional cell lines

A. DIP on *FXN* gene in control 2 (GM14926) and FRDA 2 (GM16243) cells, treated with 10 μM camptothecin for 6 hours.

A DIP on *pHIV-βGlobin* plasmid

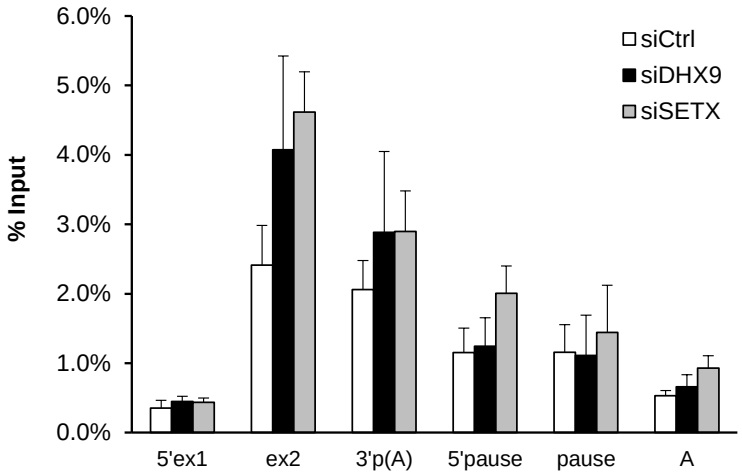


Figure A.11., related to Figure 5.18.

Loss of DHX9 leads to RNA/DNA hybrid accumulation on reporter plasmid *in vivo*

A. DIP analysis on the *pHIV-β-Globin* construct in HeLa cells transfected with a non-targeting siRNA (siCtrl), or siRNAs targeting DHX9 (siDHX9) or SETX (siSETX). All cells were co-transfected with 1 µg *pHIV-β-Globin* plasmid and 0.1 µg pTAT plasmid during the 2nd siRNA hit.