Nucleic acid structures as transcriptional and epigenetic regulators in health and disease

Lara Cravo

St John's College

University of Oxford

A thesis submitted for the Degree of Doctor of Philosophy

Sir William Dunn School of Pathology

Michaelmas Term, 2017

Nucleic acid structures as transcriptional and epigenetic regulators in health and disease

A thesis submitted for the Degree of Doctor of Philosophy, Michaelmas Term, 2017

Lara Cravo, St John's College, University of Oxford

Abstract

DNA and RNA structures are involved in major biological processes, such as regulating gene expression, and can play an important role in cancer and neurodegenerative diseases.

Huntington's disease (HD) is the most common monogenic neurological disorder caused by the expanded and unstable CAG trinucleotide repeat in the first exon of the huntingtin gene (*HTT*). The pathology of HD has been traditionally associated with the mutant HTT protein, which confers toxic gain of function leading to dysfunction and death of neurons. However, recent evidence indicates that at least part of the gain of function may operate at the RNA level.

In the first part of this thesis, the transcriptional and epigenetic regulation of the *HTT* gene in HD is investigated. It is shown that expanded CAG repeats of endogenous *HTT* transcript adopt double-stranded RNA structures that lead to RNA Polymerase II transcriptional repression and heterochromatin formation of the *HTT* gene *in cis*, providing a molecular link between dsRNA and the pathology of expansion diseases.

Double-stranded RNA is a key player in numerous Biological activities in cells, however little is known about the proteins that bind to dsRNA. Therefore, a proteomics approach was established, employing the dsRNA pull-down to identify dsRNA-binding proteins in control and HD fibroblast cells. Analysis of the dsRNA interactome indicated that interaction with the top candidate ENO1 is enriched in HD and promotes dsRNA-mediated transcriptional repression of the *HTT* gene.

This thesis work also elucidates on the importance of RNA/DNA hybrids in cancer as their accumulation is a major cause for genomic instability. The research demonstrated that cancer cells overexpressing oncogenes such as $HRAS^{V12}$, lead to increased RNA synthesis and RNA/DNA hybrid accumulation that can be correlated with replication stress. These events lead to DNA damage, revealing a strong link between R-loop-mediated genome instability and human disease.

Overall, nucleic acid structures and their interactors could determine the ability of a single gene product to perform multiple functions, as potential disease modifiers and future targets to modulate diseases.

Para o Nuno

Acknowledgements

I would first like to thank my supervisor, Professor Matthew Wood for the opportunity to work on this project in his group and for being an influential mentor for my professional development. I am also deeply grateful for having the invaluable support and advice of Dr Miguel Varela throughout the project.

I am thankful to the Graduate Program in Areas of Basic and Applied Biology (GABBA) and the Portuguese Foundation for Science and Technology (FCT) for their generous funding to pursue my DPhil at the University of Oxford.

I would like to thank the members of the Professor Wood group for providing a joyful atmosphere and stimulating discussions. My special thanks to Dr Ashling Holland for precious input and help during thesis writing, to Dr Carlo Rinaldi, Dr Mariana Conceição and Dr Raquel Manzano.

I am also grateful for the scientific support I have received from Professor Sarah Tabrizi group, especially for sharing their insights into Huntington's disease.

I wish to express my gratitude to Professor Matthew Freeman and Professor Anton van der Merwe of Sir William Dunn School of Pathology for their availability and support. I also want to thank Dr Natalia Gromak lab, especially Dr Matthias Groh and Dr Laura Albulescu for advice and help.

To all the friends I have met in Oxford, thanks for sharing cultures and cheerful moments.

I would like to thank my family for their embrace and inspiration in making me take a step forward.

Finally, I want to say thank you, Nuno, for being here, so close to me, in this long voyage – how wonderful it is to travel in unknown seas and discover new worlds with you.

Table of Contents

Abstract ii

Acknowledgements v

Table of Contents vi

List of Figures xii

List of Tables xv

List of Abbreviations xvi

Chapter 1 | Introduction 19

1.1 Genetic diseases of unstable repeat mutations 20

1.2 Repeat expansion diseases 22

1.2.1 Mechanisms of transcriptional dysregulation 22

1.2.2 Aberrant splicing in expansion diseases 23

1.2.3 Antisense and bidirectional transcription over expanded repeats 24

1.2.4 Epigenetic modifications of repeat expansions 28

1.3 Huntington's disease 31

1.3.1 Genetic mutation 31

1.3.2 Disease mechanism and pathology 32

1.3.3 Function of mutant HTT 34

1.3.4 Mutant huntingtin as a potential target for therapy 38

1.4 Genomic instability in cancer 42

1.4.1 Transcription and replication machineries 42

1.4.2 Oncogene-induced DNA damage 43

1.5 Nucleic acid structures and functions 44

1.5.1 The structure of double-stranded RNA (dsRNA) and DNA/RNA hybrids 44

1.5.2 DsRNA in expansion-associated pathology 46

1.5.3 DNA/RNA hybrid in cancer 48

1.5.4 DsRNA and RNA/DNA-binding proteins in human disease 50

1.6 Aims of this work 53

1.6.1 Investigation of the transcriptional and epigenetic regulation in Huntington's disease 53

1.6.2 Investigation of transcriptional and genomic instability processes in cancer disease 53

Chapter 2 | Materials and Methods 55

2.1 Materials 56

2.1.1 Antisense oligonucleotide sequences 56

2.1.2 Antibodies 59

2.1.3 Enzymes, materials and other reagents 60

2.1.4 Buffers 62

2.2 Cell culture 63

2.2.1 Cell lines 63

2.2.3 iPSC-derived neurons 64

2.2.4 Transfection of oligonucleotides 65

2.2.5 Pharmacological treatments 65

2.3 Nucleic acid analysis methods 66

2.3.1 Quantification of nucleic acids 66

2.3.2 Purification of genomic DNA 66

2.3.3 Real-time quantitative polymerase chain reaction (qPCR) 66

2.3.4 RNA extraction and purification 67

2.3.5 Reverse transcription qPCR (RT-qPCR) 68

2.3.6 Agarose gel electrophoresis 68

2.3.7 DNA/RNA hybrid immunoprecipitation (DIP) 69

2.3.8 DNA/RNA hybrid slot blot 70

2.3.9 dsRNA immunoprecipitation (dsRNA IP) 70

2.4 Protein analysis methods 71

2.4.1 Preparation and quantification of protein 71

2.4.2 Western Blotting 72

2.4.3 Chromatin immunoprecipitation (ChIP) 73

2.4.5 Silver Staining 75

2.4.4 MTS cell proliferation assay 75

2.4.6 Mass spectrometry 76

Chapter 3 | Transcriptional and epigenetic alterations of *HTT* gene associated with CAG repeats 79

3.1 Transcriptional dysregulation of *HTT* gene in Huntington's disease 80

3.1.1 *HTT* RNA levels are reduced in HD 80

3.1.2 RNA Pol II levels are decreased at the CAG-repeat region in HD 85

3.1.3 HD tissue-dependent regulation of RNA Pol II dynamics 86

3.2 Transcriptionally impaired *HTT* correlates with increased repressive histone marks 90

3.2.1 Histone acetylation is altered at gene loci in HD 90

3.2.2 Repressive histone marks are involved in the *HTT* gene regulation 92

3.2.3 *HTT* transcript is not affected by methyltransferase inhibitor UNC0638 in HD 97

3.3.4 SRSF6 is enriched at expanded CAG repeats in HD cells 100

3.3 Discussion of Chapter 3 102

Chapter 4 | Double-stranded RNA are formed at triplet repeat expansions and contribute to transcriptional dysregulation in HD 111

4.1 R-loops formed on *HTT* gene are reduced in HD cells and do not interfere with efficient gene expression 112

4.2 dsRNA are enriched at transcribed expanded CAG repeats of *HTT* gene and correlate with expansion size in human HD cells 115

4.3 Investigation of LNA treatments for targeting of Huntingtin 118

4.3.1 RNA expression following treatments to reduce huntingtin expression ..118

4.3.2 Protein expression following treatments to reduce huntingtin expression121

4.3.3 Effects of LNA treatments on *HTT* dsRNA expansion structures 123

4.4 Involvement of dsRNA in transcriptional repression and heterochromatin formation in HD 124

4.4.1 H3K9me2 is important for the formation of *HTT* dsRNA in HD 124

4.4.2 Formation of the hairpin dsRNA correlates with H3K9me2 and *HTT* gene repression.....127

4.5 Summary 129

4.6 Investigation of the dsRNA-binding proteins in human cells 130

4.6.1 The dsRNA IP methodology 131

4.6.2 Mass spectrometry analysis of the dsRNA-binding proteins 134

4.6.3 Characterisation and Biological validation of the dsRNA-binding proteins140

4.6.4 The top candidate alpha-enolase1 (ENO1).....145

4.7 Discussion of Chapter 4 148

Chapter 5 | Increased global transcription leads to RNA/DNA hybrid accumulation and DNA damage in cancer..... 158

Introduction 159

5.1 $HRAS^{V12}$ leads to increased nascent RNA synthesis and transcription activity 160

5.2 $HRAS^{V12}$ leads to increased R-loop formation 163

5.3 Gene expression analysis and RNase A pre-treatment for R-loop DIP in $HRAS^{V12}$ cells 168

5.4 Transcription promotes $HRAS^{V12}$-induced replication stress 171

5.5 Transcription triggers R-loop-associated DNA damage 174

5.6 Discussion of Chapter 5 178

Chapter 6 Conclusions and outlook 185

6.1 Summary of results 186

6.1.1 Transcriptional and epigenetic regulation in Huntington's disease 186

6.1.2 Investigation of transcriptional and genomic instability processes in cancer disease.....189

6.2 Future directions 194

6.2.1 Interplay between dsRNA and heterochromatin 194

6.2.2 Expansion of the CAG locus associated with chromatin and transcription196

6.2.3 Dynamics of the dsRNA interactome.....198

6.2.4 R-loop as potential therapies to cancer disease 198

Concluding remarks 200

References..... 202

Appendix..... 236

List of Figures

Figure 1.1 Trinucleotide repeat expansion disorders caused by triplet repeats in coding and noncoding gene regions. 21

Figure 1.2 Role of transcription in the pathology of repeat expansion diseases. 26

Figure 1.3 Competition between euchromatin and heterochromatin. 30

Figure 1.4 Graphic representing the HTT protein aggregation. 35

Figure 1.5 RNA interference (RNAi) and antisense oligonucleotides (ASOs). 40

Figure 3.1 Human cell lines and diagram of *HTT* gene. 82

Figure 3.2 RT-qPCR analysis of the human *HTT* gene. 83

Figure 3.3 RT-qPCR analysis of housekeeping genes. 84

Figure 3.4 *HTT* gene transcription is impaired in HD cells. 86

Figure 3.5 HD iPSCs-derived neurons at day 36. 88

Figure 3.6 Transcription initiation and elongation defects in HD iPSCs-derived neurons. 89

Figure 3.7 ChIP analysis of H3K9ac histone mark. 92

Figure 3.8 Histone modifications are associated with the transcriptional regulation of the *HTT* gene. 97

Figure 4.1 R-loops formed on *HTT* gene are reduced in HD cells. 115

Figure 4.2 dsRNA are enriched at transcribed expanded CAG repeats of *HTT* gene. .117

Figure 4.3 Investigation of antisense oligonucleotide treatments for *HTT* knockdown. 120

Figure 4.4 Quantification of HTT protein following LNA treatments for *HTT* knockdown. 122

Figure 4.5 Quantification of dsRNA levels following LNA treatments. 124

Figure 4.6 dsRNA are affected by changes in H3K9me2 repressive mark. 126

Figure 4.7 dsRNA promote transcriptional repression of *HTT* gene. 128

Figure 4.8 dsRNA IP profile in control and HD fibroblast cells. 133

Figure 4.9 The dsRNA-binding proteins IP methodology shows strong positive correlation. 136

Figure 4.10 Mass spectrometry results of three Biological replicates of dsRBP IP experiments using MaxQuant software. 138

Figure 4.11 Heat map of differentially expressed proteins in control and HD. 139

Figure 4.12 Analysis of cellular component enrichment of dsRNA interactors in HD. 140

Figure 4.13 Molecular functions of the dsRNA interactors. 142

Figure 4.14 Validation of the top dsRBP factors by western blotting. 144

Figure 4.15 ENO1 interacts with *HTT* dsRNA in HD. 147

Figure 4.16 Schematic representation of HTT mRNA fragments. 151

Figure 5.1 Protein levels of HRAS, pERK1/2, ERK1/2 and β-actin in BJ-HRAS^{V12} fibroblast cells after RAS induction for the times indicated. 161

Figure 5.2 HRAS^{V12} overexpression increases transcription activity. 163

Figure 5.3 RNA/DNA hybrid slot blot of genomic DNA. 164

Figure 5.4 HRAS^{V12} overexpression increases R-loop formation. 168

Figure 5.5 Gene expression analysis of HRAS^{V12} target and control genes. 168

Figure 5.6 RNase A pre-treatment for R-loop DIP in HRAS^{V12} cells. 170

Figure 5.7 HRAS^{V12}-induced replication stress is promoted by ongoing transcription. 173

Figure 5.8 ChIP analysis of γH2AX after Aphidicolin treatment. 176

Figure 5.9 Endogenous protein and RNA levels of RNase H1. 176

Figure 5.10 Model of how $HRAS^{V12}$ causes replication stress via increasing transcription 177

Figure 6.1 Model for the role of CAG expanded repeats of the mutant HTT in Huntington's disease..... 186

List of Tables

Table 2.1 Locked nucleic acid (LNA) antisense oligonucleotides (ASO) 56

Table 2.2 DNA oligonucleotide sequences.....56

Table 2.3 List of antibodies 59

Table 2.4 List of enzymes, materials and other reagents.....60

Table 2.5 List of Buffers.....62

Table 2.6 Cell lines used in the HD study. CAG repeat sizes, genotype and gender are indicated.....64

Table 3.1 Summary of the results described in the Chapter 3……103

Table 4.1 Representative list of proteins enriched in dsRBP in HD patients 139

Table 4.2 Representative list of proteins enriched in dsRBP in HD patients 142

List of Abbreviations

Abbreviation	Full name
3’-UTR	3’-untranslated region
5’-UTR	5’-untransted region
A	Adenine
Ab	Antibody
ACN	Acetonitrile
ACTB	β-Actin
AGC	Automatic gain control
Ago1/2	Argonaute 1/2
ALS	Amyotrophic lateral sclerosis
ASO	Antisense oligonucleotide
HD	Huntington’s disease
ATP	Adenosine triphosphate
Bp	Base pair
BSA	Bovine serum albumin
C	Cytosine
CAA	2-Chloroacetamide
cDNA	Complementary DNA
CNS	Central nervous system
ChIP	Chromatin immunoprecipitation
CPT	Camptothecin
CRISPR	Clustered regularly-interspaced short palindromic repeats
DIP	DNA/RNA immunoprecipitation
DMEM	Dulbecco’s modified eagle’s medium
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
DNase	Deoxyribunuclease
dNTP	Deoxynucleotide triphosphate
dsRNA	Double-stranded RNA
DTT	Dithiothreitol

e.g.	Example given
EBV	Epstein-barr virus
EDTA	Ethylenediaminetetraacetic acid
EtOH	Ethanol
FA	Formic acid
FASP	Filter-aided sample preparation
FBS	Foetal bovine serum
GAPDH	Glyceraldehyde 3-phosphate dehydrogenase gene
GFP	Green fluorescent protein
h	Hour
H3	Histone H3
H3K9ac	Acetylated histone H3 at lysine 9
H3K9me2	Dimethylated histone H3 at lysine 9
H3K9me3	Trimethylated histone H3 at lysine 9
H3K27me3	Trimethylated histone H3 at lysine 27
HAT	Histone acetyltransferase
HCD	Higher-energy collisional dissociation
HD	Huntington's disease
HDAC	Histone deacetylase
hESC	Human embryonic stem cell
HTT	Huntingtin
IP	Immunoprecipitation
iPSC	Induced pluripotent stem cell
kb	Kilobase
kDa	Kilodalton
LNA	Locked nucleic acid
min	Minute
ml	Millilitre
mRNA	Messenger RNA
MS	Mass spectrometry
nt	Nucleotide
PBS	Phosphate buffered saline
PCR	Polymerase chain reaction

PMSF	Phenylmethylsulfonyl fluoride
qPCR	Quantitative polymerase chain reaction
RISC	RNA-induced silencing complex
RNA	Ribonucleic acid
RNA Pol II	RNA polymerase II
RNase	Ribonuclease
RT-qPCR	Reverse transcription-quantitative polymerase chain reaction
SDS	Sodium dodecylsulfate
SDS-PAGE	Sodium dodecylsulfate polyacrylamide gel electrophoresis
SEM	Standard error of mean
T	Thymine
TSS	Transcription start site
U	Uracil
UV	Ultraviolet
WB	Western Blot
WT	Wild type

Chapter 1 | Introduction

1.1 Genetic diseases of unstable repeat mutations

Mutation is the major cause of genetic variation. A mutation is a heritable change in the DNA sequence and comprises deletions, duplication, insertions, and unrepaired DNA damage. Until 1991, the major concept of mutation was the Mendelian transmission of a stable nucleotide sequence change, that altered gene expression (Jeffreys *et al.*, 2004), and humans have developed systems for resisting rapid and harmful variations in the genome (Hegde and Saraph, 2011). However, in the early 1990s, pathogenic mutations that affect nucleotide repeats were first described, when both fragile X syndrome (FRAXA) (Verkerk *et al.*, 1991; Kremer *et al.*, 1991) and spinal bulbar muscular atrophy (SBMA) (La Spada *et al.*, 1991) were characterized as triplet repeat expansions diseases. These unstable DNA triplet sequences (microsatellites) were identified as a new type of mutation, and changed the face of genetics (Hegde and Saraph, 2011). In particular, the trinucleotide repeat expansion has taken special relevance due to its highly unstable nature, altering gene expression and/or function of the gene product, expanding during most parent-child transmissions and during the lifetime (McMurray, 2010; La Spada and Taylor, 2010). Over the years, other disease-associated microsatellite repeats were described, including additional trinucleotide repeats as well as dynamic tetra-, penta- and hexanucleotide repeats (Toth *et al.*, 2000; Mirkin, 2007).

Around forty human diseases are caused by expanded repeat sequences found in coding and non-coding gene regions (Mirkin, 2007; Castel *et al.*, 2010) (Figure 1.1). The majority are human neurological and neuromuscular inherited diseases. Over half of these diseases are caused by expansions in the coding part of a gene, typically associated with expanded CAG repeat, leading to the formation of toxic polyglutamine proteins, as in Huntington's disease (HD), spinocerebellar ataxias (SCAs) type 1, 2, 3, 6, 7 and 17, spinal bulbar muscular atrophy (SBMA), and dentatorubral-pallidoluysian atrophy

(DRPLA) (Orr and Zoghbi, 2007; Castel *et al.*, 2010; Budworth and McMurray, 2013). The other group of diseases that are caused by repeat expansions located in the noncoding region, disrupting transcription, are Friedreich's ataxia (FRDA) or fragile X tremor/ataxia syndrome (FRAXA), through modifying splicing and sequestration of splicing factors in RNA foci, as in myotonic dystrophy type 1 and type 2 (DM1, 2) (Lee and Cooper, 2009; Nelson *et al.*, 2013), which are characterized by much larger repeat expansions than the first group. And most recently, other diseases caused by toxic transcripts with elongated GGGGCC tracts were shown for amyotrophic lateral sclerosis and frontotemporal dementia (ALS/FTD) (Renton *et al.*, 2011; DeJesus-Hernandez *et al.*, 2011).

In all these diseases, the context and the abundance of the expanded repeat, subcellular localization and interactions of the proteins and RNAs that are affected have key roles on the severity of the disease phenotype, and it is exciting that in the last years, identification of molecular pathways altered by these expanded repeats is beginning to reveal potential therapeutic targets.

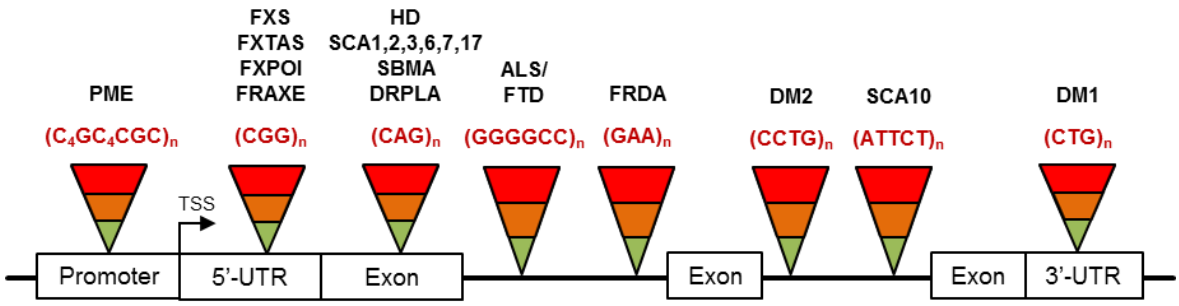


Figure 1.1 Trinucleotide repeat expansion disorders caused by triplet repeats in coding and noncoding gene regions.
Diagram representing a generic gene indicating the location of disease-causing repeats. Triangles above the gene depict length of repeats: alleles without pathological impact are shown 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

1.2.4 Epigenetic modifications of repeat expansions

Epigenetics can be described as any potentially heritable modification that modifies gene expression, but not resultant from direct changes in the primary DNA sequence. Chromatin plays a central role in the mechanisms from gene expression to chromosome dynamics throughout the cell cycle. Chromatin can be largely characterized into two types, that is euchromatin and heterochromatin (Bassett *et al.*, 2009) (Figure 1.3). Euchromatin has an open and decondensed structure, is enriched for transcriptionally active genes and is typically associated to high levels of acetylation on histone H3 and H4 (H3ac, H4ac), and trimethylation of histone H3 at lysine 4 (H3K4). On the other hand, heterochromatin is highly compact and condensed in structure and mostly contains transcriptionally inactive genes. Heterochromatin is characterized by low levels of acetylation and high levels of H3K9me2/3, H3K27me2/3 and H4K20me3 (Li *et al.*, 2007; Bassett *et al.*, 2009).

Recent studies have identified alterations in DNA methylation, histone modification and chromatin structure around expanded repeats that are not present at the correspondent wild type alleles (Dion and Wilson, 2009). It is apparent that in several repeat expansion diseases the transcription is negatively influenced by both the repeat and the epigenetic landscape of the expansion (Dion and Wilson, 2009). In Fragile X syndrome, the *FMR1* gene is transcriptionally silenced and its low expression is caused by DNA hyper-methylation within the CGG repeat and the surrounding DNA region (Sutcliffe *et al.*, 1992; Hornstra *et al.*, 1993). Recent research revealed that this DNA methylation pattern is associated with high levels of repressive marks, including H3K9me2/3, H4K29me3 and H3K27me3, and lower levels of histone acetylation marks, such as acetylated H3 and H4 (Coffee *et al.*, 1999; Coffee *et al.*, 2002; Pietrobono *et al.*, 2005; Kumari and Usdin, 2010). Thus, the expanded CGG repeats are associated with

heterochromatin marks and exhibit reduced levels of euchromatic marks. Similarly, CTG repeats induce DNA methylation throughout *DMPK* gene of DM1 patients that are resistant to methylation sensitive enzymes (Steinbach *et al.*, 1998). The expanded CTG locus is enriched for H3K9me and depleted in H3ac, adopting a heterochromatic conformation (Cho *et al.*, 2005).

In SCA7, the *ataxin-7* regulated gene transcription is due to its contribution in complexes that acetylate nucleosomal histones and function as transcription coactivators (Helmlinger *et al.*, 2004). The polyglutamine expansion of *ataxin-7* disrupts acetylation of histones in nucleosomes, which prevents their histone acetyltransferase (HAT) activity (McMahon *et al.*, 2005), leading to hypoacetylation of H3 histones, causing retinal degeneration (Palhan *et al.*, 2005).

mHTT can modify chromatin regulation and transcription during neural development, contributing to the HD pathology (Kerschbamer and Biagioli, 2015). Deletion of *HTT* leads to decreased levels of H3K27me3 class of promoters (Biagioli *et al.*, 2015) and the expanded CAG repeats are also associated with high levels of H3K4me3, leading to decreased RNA expression (Biagioli *et al.*, 2015).

As stated above, gene expression is modulated through epigenetic changes: acetylation or methylation of histones, modifying the accessibility of chromatin to transcriptional factors. In line with this, histone deacetylase (HDAC) inhibitors that derepress transcription are potential therapies in HD. It has been demonstrated that inhibiting histone deacetylases can ameliorate learning and memory in mice (Guan *et al.*, 2009) and rescue neurons (Fischer *et al.*, 2010). Moreover, HD cells show modifications in DNA methylation that can be improved by HDAC inhibitors (Jia *et al.*, 2015). Interestingly, reduced levels of intragenic histone acetylation H3 are linked with global gene expression downregulation (Bowles *et al.*, 2012; McFarland *et al.*, 2012). In

contrast, decreased levels of histone acetylation H3 and H4 levels are linked with mutant HTT overexpression and this effect can be reversed with HDAC inhibitors, which recover progressive neuronal degeneration induced by the CAG expansion (Steffan *et al.*, 2001).

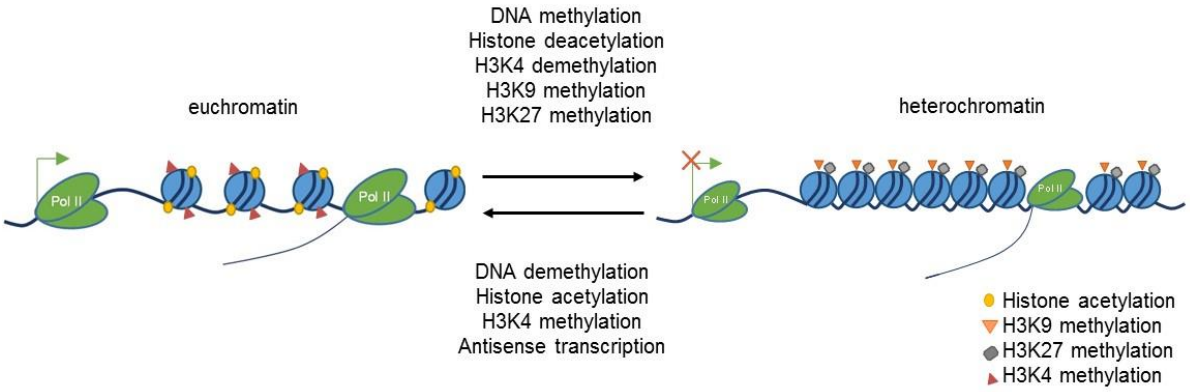


Figure 1.3 Competition between euchromatin and heterochromatin.
Easily accessible euchromatin is associated with DNA demethylation, histone acetylation and H3K4 methylation. On the other hand, inaccessible and tightly packed heterochromatin is linked to DNA methylation, histone deacetylation, H3K4 demethylation, H3K9 and H3K27 methylation.

1.3 Huntington's disease

1.3.1 Genetic mutation

First described in detail by George Huntington in 1872 (Huntington, 2003), Huntington's disease (HD) is the most common monogenic neurological disorder (Evans *et al.*, 2013; Fisher and Hayden, 2014). HD is characterized by a combination of motor dysfunction, psychiatric disturbances, and cognitive decline (Martin and Gusella, 1986; Walker, 2007) with a prevalence of 10.6-13.7 individuals per 100,000 in Western populations (Morrison *et al.*, 2011; Fisher and Hayden, 2014).

HD is caused by the expanded and unstable CAG trinucleotide repeat in the first exon of huntingtin gene (*HTT*) that was first mapped in 1993 (HD Collaborative Research Group, 1993). *HTT* is located at the chromosome 4p16.3, contains 67 exons and encodes a very large 350 kDa protein named huntingtin (HTT) (HD Collaborative Research Group, 1993). The CAG repeat is followed by a 12-base pair sequence encoding two glutamines and two prolines, followed by a CCG repeat. The number of CCG codons vary from 7 to 12 but it has been demonstrated that the expansion does not affect the pathogenesis of the *HTT* allele (Andrew *et al.*, 1994). The unstable CAG repeat is polymorphic with a range of 6-35 repeats in the healthy population, to more than 40 repeats when the mutation is fully penetrant. Penetrance may be incomplete in the range of 36 to 39 CAG repeats, that is only some individuals will develop the disease (Walker, 2007; Dorsey and Invest, 2012). An intermediate expandable repeat length of 27-35 CAG repeats is unstable through meiotic transmission, particularly during spermatogenesis and in such the pathogenesis can be modified (Wheeler *et al.*, 2007).

The length of the longer CAG repeat in *HTT* defines the age of onset in a completely dominant manner (Lee *et al.*, 2012) and also determines the rate of pathogenesis leading to the characteristic motor signs (Duyao *et al.*, 1993; Andrew *et al.*, 1993; Snell *et al.*,

1993), being responsible for about 50-70% of the variation observed in the age at motor onset (Gusella *et al.*, 2014). Effects on age of onset have been described for several other genes located in the genome of the HD patients, designated as genetic modifiers, which can contribute to the remaining variation (Gusella *et al.*, 2014).

Although clinical manifestation of HD occurs in mid-life, molecular changes due to CAG expansion are detectable in humans with HD gene mutation, but with no clinical diagnostics yet, more than a decade before diagnosis (Paulsen *et al.*, 2008) and in induced pluripotent stem cells from HD patients before they show phenotypic variation, including cell toxicity (Consortium, 2012).

1.3.2 Disease mechanism and pathology

The *HTT* gene was identified in 1993 and the causative mutation in HD is an expanded CAG repeat in its first exon (HD Collaborative Research Group, 1993). The *HTT* gene originates from two alternatively spliced transcripts with a difference only in their 3' untranslated regions and leading to the same polyglutamine protein (Lin et al., 1993; Casanova et al., 1996). The CAG repeat is translated to a polyglutamine tract, present at the N-terminus of the huntingtin protein (HTT) that is widely expressed (Li *et al.*, 1993; Jou and Myers, 1995).

RNA toxicity of expanded CAG

The pathology of HD has been traditionally associated with the mHTT, which confers toxic gain of function leading to dysfunction and death of neurons (Vonsattel and DiFiglia, 1998). However, recent evidence indicates that at least part of the gain of function may operate at the RNA level. Long tracts of untranslated CAG repeats were shown to induce toxicity at the RNA level that disrupt splicing, leading to

neurodegeneration in *Drosophila melanogaster* and *Caenorhabditis elegans* (Li *et al*., 2008; Wang *et al*., 2011).

Huntingtin protein fragments are considered to be responsible for the pathogenic mechanism observed in HD (Kuhn *et al.*, 2007; Landles *et al.*, 2010). Until very recently, the N-terminal fragments that have been attributed to the proteolytic cleavage of mHTT can induce toxicity in animal models and in cells (Wellington *et al.*, 2000). However, recent evidence indicates that the smallest *HTT* fragment detected in HD mouse models is produced by aberrant splicing that leads to the production of an exon1 HTT protein that accumulate in neuronal nuclei (Landles *et al.*, 2010; Sathasivam *et al.*, 2013). This aberrant splicing of the mutant *HTT* transcripts is CAG repeat length dependent and is prompted by the binding of SRSF6 splicing factor to the repeats, resulting in a short polyadenylated mRNA (Sathasivam *et al.*, 2013). On translation, this generates an exon1 HTT protein, because there is a stop codon at the start of the intron1. HD mouse models containing only the human HTT exon1 have a rapid onset of HD-like symptoms (Sathasivam *et al.*, 2013).

The expanded CAG repeats could reduce the expression of the antisense *HTT* RNA (*HTTAS*), leading to higher expression of *HTT* (Chung *et al.*, 2011). In addition, in human HD, the *HTT* transcript retained in the nucleus is formed by long CAG repeats that fold into hairpins *in vivo* (de Mezer *et al.*, 2011). These repeat containing hairpins are toxic to cells, can be found in RNA foci and cause alternative splicing defects, as is the case for DM1 (Mykowska *et al.*, 2011). A mechanism of RNA processing could be altered, leading to production of small CAG repeat RNAs (sCAGs) derived from long hairpins (Banez-Coronel *et al.*, 2012). The sCAGs are neurotoxic, correlate with CAG repeat length and are dependent on Dicer and Argonaute activity (Banez-Coronel *et al.*, 2012).

The diseases caused by CAG expanded repeats in coding sequences have only recently been examined for possible RNA-related mechanisms of pathology. In HD the evidence suggests that the *HTT* transcript may contribute to the pathology, and it will need to be considered in the design of therapies, but deserve further studies focus on the mechanism.

1.3.3 Function of mutant HTT

The normal HTT protein (HTT) contains 3,144 amino acids with a molecular weight of about 350 kDa. HTT is ubiquitously expressed but at varying levels depending on the cell type and its normal functions are still being investigated. Some evidence points to the role of HTT in the development of the central nervous system, synaptic activity, anti-apoptotic function, axonal and vesicle transport, and regulation of neurotrophic support, particularly involving brain-derived neurotrophic factor (BDNF) (Auerbach *et al.*, 2001; Schulte and Littleton, 2011; Bates *et al.*, 2015).

The mutant HTT protein (mHTT) contains a long polyglutamine sequence and has been described to have toxic gain of function. mHTT can adopt multiple harmful functions, causing cellular dysfunction and neuronal death (Landles and Bates, 2004). The precise mechanism by which mHTT leads to neuronal degeneration is not completely understood but aggregation of mHTT, which leads to insoluble mHTT inclusions may play the defining pathologic role (Figure 1.4). Indeed, the polyglutamine tract in mHTT has the capacity to form abnormal conformations which increase the likelihood of aggregation and this links HD to other neurodegenerative diseases in which protein aggregation plays a role in pathology (Haass and Selkoe, 2007). Multiple studies have detected mHTT inclusions in HD models and HD patient cells (DiFiglia *et al.*, 1997; Davies *et al.*, 1997; Arrasate *et al.*, 2004). In particular, mHTT aggregates are associated

with impairment of the cerebral vasculature, as well as blood brain barrier leakage, which may lead to the pathophysiology seen in both R6/2 mouse model and in HD patients (Drouin-Ouellet *et al.*, 2015). In terms of subcellular localization, although mHTT inclusions can occur in the cytoplasm, dendrites and axon terminals (Li *et al.*, 2001), they are predominantly localised to the nucleus, which causes toxic effects through transcription dysregulation processes *in trans* (Cha, 2007).

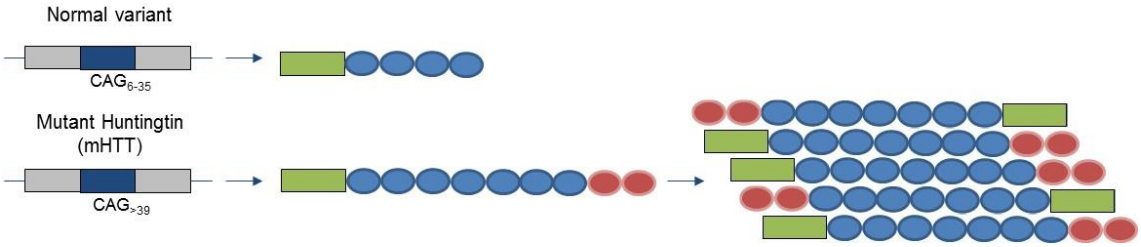


Figure 1.4 Graphic representing the HTT protein aggregation.
In HD patients, the *HTT* gene with a CAG expansion larger than 39 CAG repeats leads to the production of a polyglutamine protein, which forms HTT protein aggregates.

The length of the polyglutamine in HTT affects HD progress and is inversely associated with age of onset in human patients (Andresen *et al.*, 2007; Lee *et al.*, 2012). The most common HD models are R6/2 and R6/1 mice that have about 150 and 115 CAG repeats, respectively. The influence of CAG length is also seen in these mice, with longer CAG R6/2 animals showing an earlier and more severe phenotype (Mangiarini *et al.*, 1996). Intriguingly, when the CAG stretch is significantly long with more than 200 CAG repeats, the toxicity is ameliorated in HD transgenic mice, which might be associated with reduced HTT RNA and protein levels (Dragatsis *et al.*, 2009; Morton *et al.*, 2009).

mHTT is known to affect multiple Biological processes, including neuronal transport, lipid metabolism, mitochondrial efficiency, vesicular trafficking, oxidative stress, autophagy and gene transcription (Landles and Bates, 2004; Lin and Beal, 2006; Cha, 2007; Jin and Johnson, 2010; Jimenez-Sanchez *et al.*, 2012; Di Pardo *et al.*, 2017). The majority of these pathological events occur in the brain region, namely the striatum and cortex show prominent pathology in HD (Hadzi *et al.*, 2012; Pillai *et al.*, 2012). Interestingly, the mHTT aggregates were found in normal and unrelated allografts of striatal tissue transplanted into the brain of HD patients, suggesting that the mutant protein can spread in the brain in monogenic neurodegenerative disorders (Cicchetti *et al.*, 2014). Additionally the hippocampus is severely affected in the earliest stages of HD in patients (Begeti *et al.*, 2016) and overexpression of mHTT in the hippocampus of adult mice leads to the formation of aggregates, and neurogenic alterations associated with spatial memory deficits (Schwab *et al.*, 2017).

Effects of mutant HTT on transcription regulation

mHTT has been shown to affect transcriptional regulation (Cha, 2007) and global gene expression is downregulated in human HD brain (Hodges *et al.*, 2006). In particular, neuronal cell-specific reductions in mRNAs have been described in the mouse model brain (Luthi-Carter *et al.*, 2000; Zucker *et al.*, 2005; Kuhn *et al.*, 2007; Thomas *et al.*, 2011) and human brain (Hodges *et al.*, 2006). In HD cell culture models this abnormal gene expression profile was also observed (Sipione *et al.*, 2002; Kita *et al.*, 2002). The dynamics of the mRNA is mirrored in the different transcription processes, including RNA synthesis, processing, stability and degradation.

The binding of the mHTT to gene promoters is dependent on the length of the polyglutamine (Benn *et al.*, 2008) and mHTT interacts with transcriptionally active

proteins, including CBP, p53 (Steffan *et al.*, 2000), TBP (Huang *et al.*, 1998), Sp1, TAFII130 (Dunah *et al.*, 2002) and REST protein (Zuccato *et al.*, 2003), sequestering them to the nucleus, thus affecting transcription of many genes *in trans*.

Gene expression can also be regulated by epigenetic alterations in the genome, including acetylation and methylation of histones, facilitating or repressing the recruitment of transcriptional factors. In line with this, histone deacetylase inhibitors which increase histone acetylation, promote a relaxed chromatin conformation, making the DNA accessible and derepressing transcription in HD (Glajch and Sadri-Vakili, 2015). Interestingly, inhibition of histone deacetylase is known to ameliorate memory and learning in mouse models (Guan *et al.*, 2009) and demonstrate neuroprotective properties in various animal models (Fischer *et al.*, 2010). In addition, decreased histone acetylation H3 (H3K9K14ac) correlates with downregulated gene expression (McFarland *et al.*, 2012; Bowles *et al.*, 2012). mHTT has been shown to interact with the acetyltransferase domain of the CBP, which plays an important role in mediating neuronal survival response. Therefore, depletion of CBP caused a global histone hypoacetylation, affecting transcription of genes important for neuronal survival (Glajch and Sadri-Vakili, 2015). In transgenic mouse models, hypoacetylation of H4 leads to reduction in expression of heat shock genes (Labbadia *et al.*, 2011). Similarly, expression of mHTT in cultured cells decreases the level of the acetylated histones H3 and H4, leading to hypoacetylation, and increases heterochromatin formation. This reduction can be reversed by therapies using inhibitors of histone deacetylases (Steffan *et al.*, 2001).

Recent investigations point to a role of mHTT in transcriptional regulation, causing protein misfolding and degradation. Moreover, the role of expanded CAG repeats in dysregulation of splicing and chromatin remodelling further underline the importance of investigating RNA-associated transcriptional processes in the pathology of HD.

1.3.4 Mutant huntingtin as a potential target for therapy

The *HTT* mutation affects several molecular and cellular pathways very early in the pathogenic chain of events. It is possible that targeting specific pathways will be insufficient to tackle the pathology. Targeting the mutant gene is a promising therapeutic strategy in HD, particularly reducing mHTT expression with oligonucleotide approaches such as RNA interference (RNAi) and antisense oligonucleotides (ASOs) (Sah and Aronin, 2011) (Figure 1.5).

Antisense oligonucleotide strategies

RNAi is a mechanism of gene silencing whereby siRNA molecules mimic endogenous microRNA (miRNA). miRNAs are small noncoding RNA molecules of ~22 nucleotides which regulate endogenous gene expression by binding to complementary mRNA, usually resulting in gene silencing (Seok *et al.*, 2017). The mature miRNA becomes part of an active RNA-induced silencing complex (RISC), and with successful delivery, the small interfering RNAs (siRNAs) enter the cell and are incorporated into RISC (Leuschner *et al.*, 2006). siRNAs are small ~19 to 25 nucleotide double stranded RNAs which post-transcriptionally regulate gene expression. The RISC complex binds the target mRNA and through Argonaute-2-mediated cleavage leads to mRNA degradation (Meister *et al.*, 2004; Sah and Aronin, 2011). ASOs are single stranded oligonucleotides of ~15 to 25 oligonucleotides that bind to their complementary mRNA and function by inhibiting gene expression either by degradation of target mRNA molecules through an RNase-H-mediated mechanism or by translational block of the target mRNA (Bennett and Swayze, 2010; Sah and Aronin, 2011).

There are also a number of chemical modifications that are regularly used to increase the efficacy of ASOs. Intramolecular linkage of 2'-oxygen of ribose with the 4'-

carbon generates a bridged nucleic acid known as locked nucleic acid (LNA). Introduction of LNA into a DNA oligomer increases the binding affinity for complementary sequences (Braasch and Corey, 2001; Kurreck *et al.*, 2002) and LNAs can function as mixmers and gapmers.

In a mixmer, LNA and DNA nucleosides are intermixed throughout the sequence of the oligonucleotide and may inhibit translation by blocking the binding of the ribosomal subunits to the mRNA. Alternatively mixmers are able to bind close to the 5' end of the pre-mRNA preventing 5'-cap formation and thus inhibiting translation. A mixmer cannot induce cleavage of the target RNA but acts as a steric block to mediate the phenotype without destroying the RNA (Hagedorn *et al.*, 2017).

On the other hand, a gapmer contain LNA bases flanking a central DNA region and induce silencing via recruitment of RNase H to cleave the mRNA of a DNA-RNA duplex (Hagedorn et al., 2017). Generally, gapmers are more potent at inhibiting mRNA expression and protein, as the central DNA recruits the RNA-cleaving enzyme RNase H when the gapmer hybridizes to the mRNA (Braasch and Corey, 2001; Cerritelli and Crouch, 2009). Gapmers can target both exons and introns likewise efficiently and the activity occurs in both the cytoplasm and the nucleus, although mainly in the nucleus (Cerritelli and Crouch, 2009; Hagedorn *et al.*, 2017).

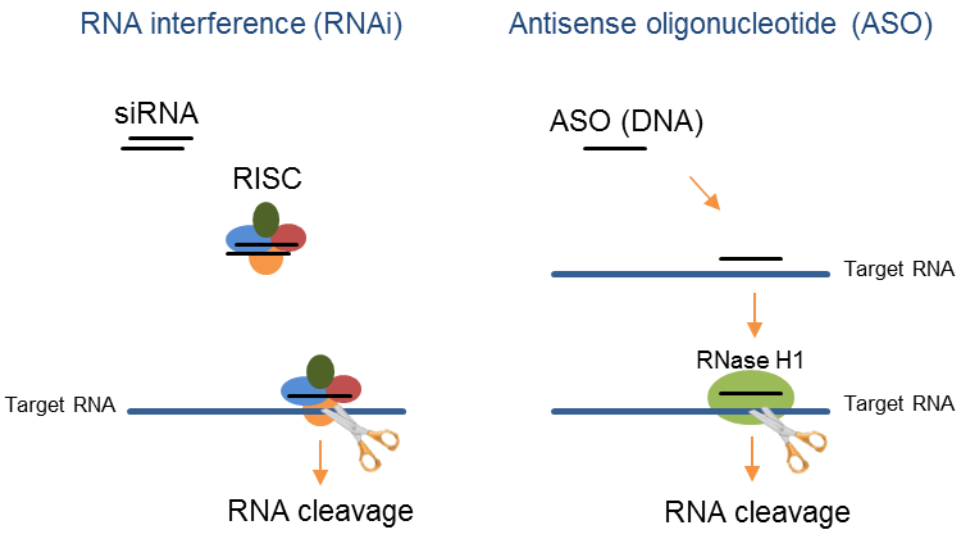


Figure 1.5 RNA interference (RNAi) and antisense oligonucleotides (ASOs).
RNA interference (RNAi) uses the RNAi-induced silencing complex (RISC) containing a siRNA to specifically degrade the target RNA. Antisense oligonucleotides (ASOs) bind to their target RNA and require endogenous RNase H enzyme to cleave the RNA in an RNA/DNA heteroduplex.

For HD, the development of oligonucleotide therapies have shown promise in silencing *HTT* and reducing the pathology, but the challenge of safe and effective delivery need to be fully addressed. HTT has beneficial activities in the mature brain and its loss reduces the ability of neurons to survive and to counteract the effects of the mHTT. In particular, it is known that *HTT* is necessary for mouse development as its silencing results in embryonically lethality (Nasir *et al.*, 1995; Duyao *et al.*, 1995) and decreasing *HTT* expression causes progressive neuronal pathology in mice (White *et al.*, 1997; Dragatsis *et al.*, 2000).

One promising strategy is to selectively target the mutant allele of mutant *HTT* by targeting expanded CAG repeats in mRNAs or target SNPs which co-localise with the *HTT* gene (Pfister *et al.*, 2009; Hu *et al.*, 2009). This could safeguard the wild-type *HTT* allele, ensuring that total HTT levels are maintained at safe levels. In particular, the

clinical development of siRNAs targeting allele-specific SNPs provide therapeutic benefit in HD patients (Pfister *et al.*, 2009). In addition, an allele-specific ASO was predicted to silence *HTT* at SNP sites and could be used in 49% of the HD population (Carroll *et al.*, 2011).

Apart from RNAi and ASOs, there are other methods to silence *HTT* including exon skipping oligonucleotides, the plasmid based strategies clustered regularly interspaced short palindromic repeats interference (CRISPRi) and short hairpin RNA (shRNA). They act at different stages of the gene expression regulation to repress transcription, alter splicing, degrade mRNA or inhibit translation.

There are several studies on HD listed on www.clinicaltrials.gov and the majority investigate the use of drugs for symptomatic treatment. Regarding the antisense strategy, initial study led by Isis/IONIS used antisense oligonucleotides to target the CAG expansion at the mRNA level (Gagnon *et al.*, 2010) to selectively inhibit expression of mHTT. Further work has focused on the development of antisense oligonucleotides targeting SNPs. Using primary cells from patients with HD and the transgenic YAC128 and BACHD mouse lines, they developed antisense oligonucleotide molecules that potently and selectively silence mutant *HTT* at SNP sites both *in vitro* and *in vivo* (Carroll *et al.*, 2011). Recently, a new study sponsored by IONIS has focused on an antisense strategy for HD. The purpose is to investigate the safety, tolerability, pharmacokinetics and pharmacodynamics of antisense oligonucleotide IONIS-HTT_{RX} in HD patients. IONIS-HTT_{RX} is an ASO that binds to *HTT* RNA, inhibiting HTT production, with the aim of slowing disease progression and being safe at the same time (http://hdresearch.ucl.ac.uk).

These findings suggest the existence of a therapeutic window for a safe and efficient approach. However, the long-term effect of silencing normal HTT in the adult

brain is still unknown and the effective delivery of siRNA and ASO therapies to the CNS remains a major challenge because these molecules do not cross the blood-brain barrier. Gene expression modifiers must be directly introduced into the CNS with proven efficacy through a safe and efficient delivery for HD treatment and other neurodegenerative disorders. In addition, the knowledge that HTT silencing mechanisms are a major focus of current therapeutic efforts highlights the need to ensure that the *HTT* gene is thoroughly evaluated.

1.4 Genomic instability in cancer

1.4.1 Transcription and replication machineries

Genomic instability is a hallmark of human cancers (Negrini *et al.*, 2010) characterized by high mutation rates and genomic rearrangements that ultimately lead to aggressiveness and resistance to therapy (McGranahan *et al.*, 2012). One of the mechanisms proposed to cause genomic instability in cancer is DNA replication stress (Halazonetis *et al.*, 2008), which occurs when DNA replication fork progression slows or stalls. It was shown that this leads to the collapse of forks into DNA double-strand breaks (DSBs), as well as incomplete sister chromatid separation in the following mitosis (Kotsantis *et al*., 2015). Markers of spontaneous replication stress are present in tumour *in vitro* and *in vivo* and genomic instability favourably affects specific genomic sites, called common fragile sites (Gorgoulis *et al.*, 2005; Tsantoulis *et al.*, 2008), which are sensitive to DNA replication stress (Durkin and Glover, 2007).

It has been demonstrated that upregulation of transcription in cancer cells can trigger replication stress. Unless transcription and replication are spatially and temporally separated, it is inevitable that the cell cycle of both DNA and RNA polymerases is affected, due to direct collisions or torsional stress between the two protein complexes.

This interference between the two machineries leads to replication fork slowing and genomic instability and has the potential to cause replication stress (French, 1992; Aguilera and Garcia-Muse, 2012). Another important source of replication stress is the collision of the replication machinery with transcriptional RNA/DNA hybrids (R-loops), where the nascent transcript hybridizes with the DNA template (Gan *et al.*, 2011). R-loops affect transcription and genome integrity and are an important source of chromosome fragility, causing cancer and other diseases (Aguilera and Garcia-Muse, 2012).

1.4.2 Oncogene-induced DNA damage

Genomic instability characterizes almost all sporadic human cancers (Di Micco *et al.*, 2006; Dereli-Oz *et al.*, 2011). It has been demonstrated that only a few genes are mutated, deleted and/or amplified at high rates in sporadic human cancers (Negrini *et al.*, 2010), including TP53 tumour suppressor and genes that regulate cell growth either positively – RAS and EGFR – or negatively, as the tumour suppressor genes PTEN and CDKN2A. In line with this, it has been shown that overexpression of oncogenes such as RAS, MOS, MYC CDC25A or CYCLIN E is sufficient to induce replication stress in cells (Di Micco *et al.*, 2006; Dominguez-Sola *et al.*, 2007; Bester *et al.*, 2011; Neelsen *et al.*, 2013). All these oncogenes act in the growth factor signalling pathways that induce proliferation by promoting cell growth and division. Therefore, molecular changes linked to proliferation are important candidates for triggering replication stress. One study suggests that CYCLIN E-induced replication stress is a consequence of increased replication initiation during S phase (Jones *et al.*, 2013). Interestingly, overexpressed c-MYC is able to stimulate transcription through all three RNA polymerases (Lin *et al.*, 2012; Nie *et al.*, 2012). In addition, oncogenic RAS activates transcription through the

mitogen-activated protein kinase extracellular signal-regulated kinase (ERK), which stimulates different transcription factors, including the TATA-box binding protein (TBP) (Johnson *et al.*, 2000; White, 2005). It is not surprising that components of the transcription machinery are found dysregulated in a broad range of cancers (White, 2005; Bywater *et al.*, 2013).

Although the molecular mechanisms underlying genomic instability in cells are still unclear, evidence suggest that targeting transcriptional profile and spontaneous replication stress phenotype is a promising strategy for cancer therapy (Dobbelstein and Sorensen, 2015).

1.5 Nucleic acid structures and functions

1.5.1 The structure of double-stranded RNA (dsRNA) and DNA/RNA hybrids

Double-stranded RNA structure

RNA structure is involved in major Biological processes, such as regulating gene expression (Lee *et al.*, 2003; Li *et al.*, 2006) and plays an important role in diseases, including cancer (Lu *et al.*, 2005), viral diseases (Watts *et al.*, 2009), and neurodegenerative diseases (Sobczak *et al.*, 2003; Chen *et al.*, 2017). The most common secondary structure described in RNA folding prediction is double-stranded RNA (dsRNA), particularly an RNA hairpin which consists of a dsRNA stem, frequently enclosing mismatches and bulges (i.e. unpaired sequences within the stem) and a terminal loop (Sobczak *et al.*, 2003; Michlewski and Krzyzosiak, 2004).

RNA hairpins can be generated by either an a) RNA molecule that is twisted, made as a folded-back template for RNA polymerase, which synthesizes the second strand of the stem or by b) transcription of an inverted repeat DNA causing the RNA folding into a stem-loop structure by the action of DNA dependent RNA polymerase (Svoboda and

Di Cara, 2006). In nature, the first mechanism, which generates a perfect long dsRNA hairpin is not so common and is mostly seen in viral replication (Behrens *et al.,* 1996). In contrast, the second mechanism, about RNA hairpins folded within RNA transcripts, is related to a broad range of Biological processes, including translational regulation. In terms of dynamics, RNA hairpins can adopt different positions within different classes of RNA; they differ in the size of the loop, the length of the stem, the size and number of bulges and, importantly, in the nucleotide sequence (Svoboda and Di Cara, 2006). Thus, the scenarios are immense, making RNA a versatile structure, permitting specific interactions with proteins (Hudson and Ortlund, 2014). In addition, RNA hairpins can regulate gene expression *in cis* or *in trans*, affecting other RNAs or pathways and serve as binding sites for numerous RNA-binding proteins (RBPs) (Johnstone and Lasko, 2001; Marzluff, 2005). These RBPs modulate protein expression by regulating the stability and translation of RNAs (Hudson and Ortlund, 2014). Apart of gene regulation, it is becoming clear that RNA hairpins are involved in subcellular localization, regulation of translation and RNA stability (Marzluff, 2005).

DNA/RNA hybrids (R-loops) structure

R-loops are three-stranded nucleic acid structures, which occur when nascent RNA hybridises to the DNA template, leading to an RNA/DNA hybrid, leaving the non-template DNA single-stranded (Aguilera and Garcia-Muse, 2012). R-loops form naturally during transcription, and even though sometimes their presence is harmful to genome stability, they can be powerful regulators of gene expression (Skourti-Stathaki and Proudfoot, 2014).

2004; Chen *et al.*, 2017). It was shown that repeats are susceptible to form dsRNA structures in the form of hairpins and those are overrepresented in exons, particularly the CNG (where N is any nucleotide), CGA and AGG, whereas CTT, CAT, CAA, TAA and TTA are underrepresented (Kozlowski *et al.*, 2010). The stem of the CNG repeat hairpin was shown to be formed of C-G and G-C base pairs (Sobczak *et al.*, 2003) and the order of stability for CNG repeat hairpins was CGG > CAG > CUG > CCG. The hairpins participate in different Biological processes (Krzyzosiak *et al.*, 2012) and the repeats have a high mutation rate and their abundance may be higher than originally thought due to bidirectional transcription in many genes and intergenic regions (Dinger *et al.*, 2009).

DsRNAs have been proposed to contribute to toxic RNA gain-of-function caused by transcripts formed of expanded CUG, CAG or CGG repeats (Li and Bonini, 2010; Jiang *et al.*, 2004). In line with this, it was shown that the *DMPK* transcript implicated in DM1 pathogenesis adopt ds-like structures formed by CUG repeats. This finding was crucial to the later discovery that muscleblind-like 1 (MBNL1) protein is sequestered by mutant *DMPK* transcript (Sicot *et al.*, 2011; Chen *et al.*, 2017). The contribution of sequences surrounding the repeats to repeat hairpin stabilization was shown in *HTT* and *AR* transcripts, in which the flanking CCG and CUG repeats, respectively, interact with CAG repeat tracts to form different hairpin compositions and conformations (de Mezer *et al.*, 2011).

It is becoming evident that the RNA structures formed by triplet repeat expansions, particularly dsRNA hairpins, greatly contribute to toxicity through an RNA-mediated pathogenic pathway. Additionally, these dsRNA hairpins with toxic gain of function can serve as a binding site to essential proteins and a typical feature is the formation of nuclear foci (Wojciechowska and Krzyzosiak, 2011). In DM1 nuclear retention of toxic aggregates which lead to misregulated alternative splicing is observed (Jiang *et al.*, 2004).

Overall, different studies point to an involvement of R-loops in cancer Biology. However, the specific role of R-loops in transcriptional dysregulation in cancer remains elusive and whether they can directly influence tumorigenesis.

1.5.4 DsRNA and RNA/DNA-binding proteins in human disease

DsRNA-binding proteins

The number of proteins associated with dsRNA Biology has increased in the last decade, reflecting the diversity of dsRNA processes (Hudson and Ortlund, 2014; Wang *et al.*, 2015). DsRNA-binding proteins (dsRBPs) were found to regulate cellular dsRNA levels in several processes, including transcription, translation, gene silencing, microRNA biogenesis, mRNA transport and telomere maintenance (Saunders and Barber, 2003; Hudson and Ortlund, 2014; Wang *et al.*, 2015).

DsRBPs were first identified in the early 1990s and, despite their involvement in diverse Biological functions, interacting with different RNA structures, they are highly conserved in terms of amino acid composition and domain structures (St Johnston *et al.*, 1992; Masliah *et al.*, 2013). For example, in particular, dsRNA structures found in virus activate protein kinase R (PKR), which in turn triggers the downstream antiviral immune signalling (Zheng and Bevilacqua, 2004). Also, it has been shown that dsRNA interacts directly with the proteins RAN and RCC1 and, though the latter, with the mitotic RANGAP1–SUMO1–RANBP2 complex. This binding relies on the double-stranded conformation and altering of dsRNA levels which leads to abnormal mitotic progression and cell death (Portal *et al.*, 2015).

Genetic experiments in mice showed that disruption of ADAR1 and SPNR dsRBPs result in embryonic lethality, proving that dsRNA-protein interactions play vital roles in controlling gene expression (Wang *et al.*, 2000; Pires-daSilva *et al.*, 2001). Some of the

dsRBPs, such as DICER and NFARs, are located in the nucleus and participate in processes of RNA interference (RNAi), mRNA stability, editing, splicing and/or export (Bernstein *et al.*, 2001; Saunders and Barber, 2003). Other dsRBPs, including SPNR, PRPB and TENR are able to interact with the 3' UTR of murine *Prm1* gene, however it is still unclear whether they perform similar functions in the human cell (Schumacher *et al.*, 1995b, 1995a).

In addition, cytoplasmic dsRBPs, including PKR and PACT, have been shown to be involved in the regulation of translation, dsRNA signalling events and host defence (Clemens and Elia, 1997; Patel and Sen, 1998). Evidence shows that interaction with dsRNA leads to PKR autophosphorylation, allowing PKR to catalyse the phosphorylation of target substrates, such as eIF2, required in the initiation of translation, inhibiting protein synthesis in the cell (Joshi *et al.*, 2013). Moreover, studies in mice have shown that PKR is a key component of the early host defence, acting in the innate immune system before the activation of the IFN system and the acquired immune response (Balachandran *et al.*, 2000). Apart from that, it was shown that PKR binds mutant huntingtin transcripts and is present in the brain areas most affected by the disease, suggesting a role in the pathogenesis of HD (Peel *et al.*, 2001). A similar finding was shown in Parkinson's disease (PD), in which a strong induction of phosphorylated-PKR is associated with defects in apoptosis and degeneration in PD and HD (Bando *et al.*, 2005).

DNA/RNA-binding proteins

Interestingly, various proteins can bind and regulate cellular RNA/DNA hybrid (R-loop) levels either directly or indirectly, mostly maintaining R-loop levels in balance. Among these are proteins necessary for efficient transcription elongation,

polyadenylation, termination, RNA splicing, packaging and export (Huertas and Aguilera, 2003; Skourti-Stathaki *et al.*, 2011; Santos-Pereira *et al.*, 2013; Herrera-Moyano *et al.*, 2014).

DNA topology itself can affect hybridization of RNA to DNA, and topoisomerases 1 and 3B play important roles in modulating R-loop levels (Tuduri *et al.*, 2009; Yang *et al.*, 2014). As R-loop accumulation can be harmful, some proteins are involved in the maintenance of genome integrity, such as XPF, XPG, BRCA1 and BRCA2 proteins, suggesting an interplay between DNA repair and R-loop formation (Wahba *et al.*, 2013; Sollier *et al.*, 2014; Bhatia *et al.*, 2014). Importantly, there are enzymes that specifically degrade the RNA in R-loops, as RNase H1 and RNase H2 and other helicases that are able to unwind RNA/DNA hybrids, as Sen1/SETX (Cerritelli and Crouch, 2009; Skourti-Stathaki *et al.*, 2011).

R-loops can also destabilize the function of regulatory proteins, including the THO/TREX complex, H3S10p and H3K9me2 histone modifications, and CpG methylation (DNMT3B1), influencing gene expression regulation, including DNA methylation, post-translational histone modifications and transcription (Nakama *et al.*, 2012; Ginno *et al.*, 2012; Castellano-Pozo *et al.*, 2013). In relation to the RNA/DNA helicases, e.g. the Pif1/PIF1 and DnaB/MCM complexes, the caveat is that for many there is lacking *in vivo* evidence for their *in vitro* functions (Shin and Kelman, 2006; Boule and Zakian, 2007) and the molecular mechanisms underlying interactions between proteins and R-loops are still poorly understood.

1.6 Aims of this work

1.6.1 Investigation of the transcriptional and epigenetic regulation in Huntington's disease

The first main aim of this project was to investigate the pathogenic molecular events, associated with the CAG expansion in the first exon of the *HTT* gene, including transcription, epigenetics and regulation of the *HTT* gene. Moreover, I also investigated the formation of dsRNA structures formed at expanded CAG repeats and their enrichment in HD patient cells. In particular, I explored the role of dsRNA and interacting proteins as potential disease modifiers in regulating *HTT* gene expression.

1.6.2 Investigation of transcriptional and genomic instability processes in cancer disease

Cancer is a disease associated with genomic instability that often results from oncogene activation. In Chapter 5, it was investigated whether the oncogene-induced increase in transcription is a mechanism of endogenous replication stress and genomic instability in cancer cells. It was also explored whether the processes of RNA synthesis and R-loop accumulation interfere with replication fork progression and DNA damage in cancer.

Publications

Parts of the work described in this thesis have been published in scientific journals:

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

Kotsantis, P.; Silva, L.M.; Irmscher, S.; Jones, R.M.; Folkes, L.; Gromak N. & Petermann E. **Increased global transcription activity as a mechanism of replication stress in cancer.** Nature Communications 7, 13087 (2016) doi: 10.1038/ncomms13087.

Chapter 2 | Materials and Methods

2.1 Materials

2.1.1 Antisense oligonucleotide sequences

Table 2.1 Locked nucleic acid (LNA) antisense oligonucleotides (ASO)

All locked nucleic acid (LNA) antisense oligonucleotides (ASO) were designed by Dr Miguel Varela (Wood laboratory) and synthesized by Exiqon. Antisense LNA Gapmers are designed to catalyse RNase H1 dependent degradation of the target RNA. LNA-REP mixmer and LNA-REPH gapmer target the *HTT* repeat.

LNA	**Sequence**
LNA-REP mixmer	TgcTgcTgcTgcTgcTgcT
LNA-REPH gapmer	gcTgcTgctgctgcTgcTg
LNA Negative Control	TgcTatAccAgcGtcGtcA

Table 2.2 DNA oligonucleotide sequences

All DNA oligonucleotide sequences (used for qPCR) are given in 5’ to 3’ direction.

Name	**Sequence (5' to 3')**
HTT gene	
HTT up2 (F)	GGCTGTCCGGGTGAGTATGG
HTT up2 (R)	CCATGCGGAAAGGATCCCCC
HTT up1 (F)	TCGCCTCACCCCATTACAGTC
HTT up1 (R)	GGTTCTGCCTCACACAGCAAG
HTT prom (F)	GCCGCTCAGGTTCTGCTTTT
HTT prom (R)	TCAGCTTTTCCAGGGTCGCC
HTT A (F)	CCGGCTGTGGCTGAGGAG
HTT A (R)	GCAGGGTTACCGCCATCC
HTT B (F)	TTGGTGAACTCGTGCCCTGT
HTT B (R)	TCATTCAGTGAGTGTTTCTCGACT

HTT in1-ex2 (F)	GGTGTAGAACTAAGTGGAGTGGGT
HTT in1-ex2 (R)	ACACGGTCTTTCTTGGTAGCTGA
HTT C (F)	CAGCTACCAAGAAAGACCGTGTG
HTT C (R)	CAAAGTGCAATTACCTGACAGAC
HTT D (F)	AGTGGTTTGCTGGGTTGAGG
HTT D (R)	GGCTCTCACCACCTTGCATC
HTT E (F)	GGCTCCTCTTGACAGTTTCTCTTC
HTT E (R)	CTGCGTCATCACTGCACAGC
HTT F (F)	GGGATGTGTCTTCCATAAATCTC
HTT F (R)	GCTCGAGCTGTAACCTTGGAA
HTT 2181 (F)	ATCTTCCTTCCACCCCTTTCC
HTT 2357 (F)	AGTCTCACATTGTCACCCAGGC
HTT 2556 (F)	TCCTGACCTTGTGATCTGT
HTT in1 3 kb (R)	ATTAGCTGGACATGGTGGTG
HTT 7128 (F)	GAGGACTTTTGGAGATGTAAAGGC
HTT 7169 (F)	GTGGCTGTAGAATGTGCTGGTG
HTT in1 7kb (F)	GCATTTGGGAGGAAGACCTGTG
HTT in1 7kb (R)	ACAGAAATGCTCAACGGCCC
HTT in5 (F)	ACTGGCTTTTCCCTAACAGG
HTT ex6 (R)	GGTCTCTTGCTTGTTCGAGTC
UAP dt18	GGCCACGCGTCGACTAGTAC(T)18
UAPnest	GGCCACGCGTCGACTAGTAC
β-actin gene	
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 5'pause (F)	TTACCCAGAGTGCAGGTGTG
β-actin 5'pause (R)	CCCCAATAAGCAGGAACAGA
β-actin pause (F)	GGGACTATTTGGGGGTGTCT

β-actin pause (R)	TCCCATAGGTGAAGGCAAAG
c-FOS gene	
c-FOS2 (F)	TTCCACGCTTTGCACTGAA
c-FOS2 (R)	GGACTTAAGCCTCGGCTC
c-FOS3 (F)	TGAGCCCGTGACGTTTAC
c-FOS3 (R)	TGCAGATGCGGTTGGAG
c-FOS4 (F)	GGATAGCCTCTCTTACTACCACTCA
c-FOS4 (R)	GGAAGCCAGCCTTACCT
c-FOS nascent (F)	CCATGACAGGAGGCCGA
c-FOS nascent (R)	ACATTCCCAGGAAGAGTA
Intergenic region	
Intergenic region (F)	TCCTCCACCCTCCAAGT
Intergenic region (R)	TGAGGGCTGGTTCGGTC
DUSP6 gene	
DUSP6 prom (F)	GGCTGAGTCCAAGAGATAGC
DUSP6 prom (R)	GCTTCTAATCCCTCCCTCCA
DUSP6 in1 (F)	AACGGGTTCTGCGCCTTCTT
DUSP6 in1 (R)	TGT GCC AAT TTG CAT CCC CAA
SPRY2 gene	
SPRY2 prom (F)	TTCCCATTCGCTCATCTGCCA
SPRY2 prom (R)	AAGCGCACGCGGAGTATTTC
SPRY2 in1A (F)	CGAATTCGGCGCTGAGAG
SPRY2 in1A (R)	ACAGGTTAGAAATGCGGG
SPRY2 in1B (F)	TCTGTGTCCGTTTGCCCCA
SPRY2 in1B (R)	TGAAAGTGCTTTGAAACCCCC
GAPDH gene	
GAPDH prom (F)	AGCTCAGGCCTCAAGACCTTGGGCT
GAPDH prom (R)	GGCTGACTGTCGAACAGGAGC

GAPDH in1 (F)	CCACTAGGCGCTCACTGTTC
GAPDH in1 (R)	TCGTAGACGCGGTTCGG
γ-Actin gene	
γ-Actin prom (F)	GGAAAGATCGCCATATATGGAC
γ-Actin prom (R)	TCACCGGCAGAGAAACGCGAC
γ-Actin in1 (F)	CCGCAGTGCAGACTTCCGAG
γ-Actin in1 (R)	CGGGCGCGTCTGTAACACGG
RNase H1 gene	
RNase H1 (F)	TAACTGGGTTCAAGGTTGGAAG
RNase H1 (R)	TCTTCCGATTGTTTAGCTCCTTC
TBP gene	
TBP (F)	TTCGGAGAGTTCTGGGATTG
TBP (R)	CTCATGATTACCGCAGCAAA

2.1.2 Antibodies

The antibodies listed were stored according to the manufacturer's guidelines, either aliquoted at -20°C or at 4°C. Table 2.3 indicates the amounts and dilutions used per sample to probe for the specific protein.

Table 2.3 List of antibodies

Target	Reference number	WB	ChIP	Supplier
Actin	A2066	1:2000	-	Sigma-Aldrich Ltd.
Histone H3 total	Ab1791	1:10000	5 µg	Abcam
Histone H3K9me2	ab2120	1:2000	5 µg	Abcam
Histone H3K9ac	ab4441	1:2000	5 µg	Abcam

Histone H3K27me3	ab8898	1:2000	5 µg	Abcam
RNA polymerase II	H224	-	3 µg	Santa Cruz Biotechnology
Huntingtin	MAB2166	1:1000	5 µl	Millipore
RNA/DNA hybrids	S9.6	1:1000	14.4 µl	University of Oxford, UK
J2 dsRNA	10010200	-	3 µl	Scicons
Ago1	04-083	-	5 µl	Millipore
Topoisomerase-I	109374	1:1000	10 µl	Abcam
SR	104793	-	5 µl	Abcam
U1 snRNP 70 K	H111	-	10 µl	Millipore
Ago2	32381		5 µl	Abcam
HP1γ	05-690	-	13 µl	Upstate
G9a	40542		5 µl	Abcam
γH2AX	07-164	-	5 µg	Millipore
H2AX	07-627	-	5 µg	Millipore

2.1.3 Enzymes, materials and other reagents

Commercial reagents were used according to the manufacturer's instructions. Non-commercial buffers and solutions were prepared with ddH_2O and autoclaved.

Table 2.4 List of enzymes, materials and other reagents

Reagent	**Reference number**	**Supplier**
AccuPrime Taq DNA Polymerase	12346-086	Thermo Fisher Scientific
AccuPrime GC-Rich DNA Polymerase	12337-016	Thermo Fisher Scientific
Actinomycin D	A9415	Sigma-Aldrich
Amersham ECL detection reagent	RPN2209	GE Healthcare
Aphidicolin	A0781	Sigma-Aldrich
Benzonase nuclease	70746-4	Novagen
Biotaq DNA polymerase	BIO-21040	Biologyine, UK
BIX-01294	B9311	Sigma-Aldrich
Bovine serum albumin	A9647	Sigma-Aldrich

CellTiter 96 Cell Proliferation Assay	G358C	Promega
cOmplete EDTA-free protease inhibitors	5056489001	Roche
Camptothecin	C9911	Sigma-Aldrich
Dithiothreitol	D9779	Sigma-Aldrich
DNA Hyperladder	H2-516107	Biologyine
DNA loading buffer	HLBB-516107	Biologyine
DNase I (RNase free)	4716728001	Roche
Fast Sybr Green Master Mix	4385612	Thermo Fisher Scientific
Formaldehyde 36.5-38%	F8775	Sigma-Aldrich
GlycoBlue	AM9515	Thermo Fisher Scientific
Glycogen	R0561	Roche
High-Capacity cDNA RT Kit	4375222	Applied Biosystems
LDS loading buffer	NP0007	Life Technologies
Leupeptin	11017101001	Roche
Lipofectamine RNAiMAX Reagent	13778-150	Thermo Fisher Scientific
MultiScribe Reverse Transcriptase	4311235	Thermo Fisher Scientific
NuPAGE 3-8% Tris-Acetate gel	EA0375BOX	Life Technologies
NuPAGE 4-12% Bis-Tris gel	NP0321BOX	Life Technologies
NuPAGE antioxidant	NP0005	Life Technologies
NuPAGE MOPS Running Buffer	NP0001	Life Technologies
NuPAGE Transfer Buffer	NP00061	Life Technologies
NuPAGE Tris-Acetate Running Buffer	LA0041	Life Technologies
Opti-MEM Reduced Serum Medium	31985062	Life Technologies
Pepstatin	11359053001	Roche
Pierce Silver Stain Kit	24612	Thermo Fisher Scientific
Plasmid Maxi Prep kit	12162	QIAGEN
Platinum Pfx DNA polymerase	11708-013	Thermo Fisher Scientific
PMSF	P7626	Sigma-Aldrich
Protein A agarose beads	16-157	Merck-Millipore
Protein G agarose beads	16-201	Merck-Millipore
Protein molecular weight marker	26619	Thermo Fisher Scientific
Proteinase K	3115828001	Roche
PVDF membrane	106000023	GE Healthcare

QIAquick PCR Purification Kit	28104	QIAGEN
QuantiTect SYBR Green PCR Master Mix	204145	QIAGEN
RNase A	12091021	Life Technologies
RNase H	M0297S	New England Biologyabs
RNeasy Mini kit	74104	QIAGEN
ShortCut RNase III	M0245S	New England Biologyabs
SYBR Green PCR Master Mix	4309155	Thermo Fisher Scientific
SuperScript III Reverse Transcriptase	18080093	Thermo Fisher Scientific
TRIzol Reagent	15596-018	Thermo Fisher Scientific
UNC0638	U4885	Sigma-Aldrich
X-ray film (FujiFilm)	12715325	Fisher Scientific

2.1.4 Buffers

Table 2.5 List of Buffers

Buffer	**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 $NaHCO_3$, 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, 2 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, 500 mM NaCl, 0.1% SDS, 1% Triton X-100
ChIP Wash Buffer B	20 mM Tris-HCl, pH 8.0, 2 mM EDTA, 150 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
RIPA Buffer	25 mM Tris-HCl pH 7.6, 150 mM NaCl, 1% NP-40, 1% Deoxycholate, 0.1% SDS

2.2 Cell culture

2.2.1 Cell lines

Untransformed fibroblast cell lines and Epstein-Barr virus transformed B-lymphocyte cell lines from healthy individuals and HD patients were acquired from Coriell Institute for Medical Research, New Jersey, USA (see Table 2.6 for cell line description). Human BJ-hTert HRASV12$^{ER\text{-}TAM}$ (Voorhoeve *et al.*, 2006; De Vita *et al.*, 2005) fibroblast cells (ATCC) were induced with 333nM 4OHT (Sigma) and TBP expression was induced with 2 mg ml^{-1} doxycycline (Sigma). 4OHT or doxycycline remained present during all experiments.

Cells were confirmed to be free of mycoplasma infection. Fibroblast cell lines were grown in Dulbecco's Modified Eagle's Medium (DMEM) (Sigma-Aldrich) supplemented with 10% fetal bovine serum (FBS) (Gibco), 100 U/ml penicillin and 100 µg/ml streptomycin (Gibco). Lymphoblastoid cell lines were grown in Roswell Park Memorial Institute 1640 medium (RPMI 1640) (Sigma-Aldrich) supplemented with 15% FBS, 100 U/ml penicillin and 100 µg/ml streptomycin. All cells were incubated in a humidified atmosphere containing 5% CO_2 at 37°C. The cell lines were split frequently every three days either by trypsinisation (fibroblast cells) or by diluting the cells (lymphoblasts) at approximately 80% confluency.

Table 2.6 Cell lines used in the HD study. CAG repeat sizes, genotype and gender are indicated.

Cell line	CAG repeat on each allele	Genotype	Gender
Fibroblasts			
GM04845	16/20	Wild-type	Female
GM04857	40/50	Homozygous	Female
GM04281	17/68	Heterozygous	Female
GM09197	18/173	Heterozygous	Male
Lymphoblasts			
GM04844	16/20	Wild-type	Female
GM04856	40/50	Homozygous	Female
GM14044	18/250	Heterozygous	Male
iPSC-derived neurons			
Control	16/20	Wild-type	Female
HD	21/73	Heterozygous	Female

2.2.3 iPSC-derived neurons

Medium-sized spiny neurons (MSNs) from Human induced pluripotent stem cells (hiPSCs) were differentiated and maintained as described in (Arber *et al.*, 2015) in the laboratory of Prof Dr Sarah Tabrizi, under the supervision of Dr Alison Wood. The cells were maintained on Geltrex in Essential 8 medium. All passaging and expansion of the cell lines were performed via manual dissociation using 0.02% EDTA pH 7.2. To initiate differentiation of 80% confluent cells, the iPSC media was switched to DMEM-F12/Neurobasal media (2:1) supplemented with N2 (Life Technologies) and retinol-free B27 (Life Technologies) (together referred to as N2B27pre). For the first 9 days, cultures were supplemented with SB431542 (10 µM; Tocris), LDN-193189 (100 nM; StemGene) and dorsomorphin (200 nM; Tocris).

Mechanical splitting was performed using 0.02% EDTA onto fibronectin at day 9 and then poly-D-lysine/laminin at day 20. Cells were maintained in N2B27pre supplemented with activin A (25 ng/ml; R&D) from day 9. BDNF, GDNF (10 ng/ml each; Peprotech) were added from day 26 to aid neuronal maturation and survival. In order to obtain a single cell suspension for counting and re-seeding into Geltrex-coated T75 cm2 flasks, an Accutase split was performed at day 28. The MSNs were visualised by light microscopy using the Zeiss AxioVert A1 inverted light microscope, and the pictures were taken with a Zeiss AxioCam ICm1.

2.2.4 Transfection of oligonucleotides

Fibroblast cells from control and Huntington's disease patients were seeded at $4 \cdot 10^6$ cells per well in a 6-well plate and transfected with 100 nM ASO LNAs, using 4 µl of Lipofectamine RNAiMAX (ThermoFisher Scientific) mixed with 180 µl Opti-MEM. In brief, ASO LNAs and then the mixture were added drop-wise to each well and incubated for 15 min at room temperature. After that, $4 \cdot 10^6$ cells per well were seeded and the duration of silencing experiment was 48 hours. After that period, cells were harvested for protein or RNA extraction, ChIP or dsRNA immunoprecipitation.

2.2.5 Pharmacological treatments

10 µM Camptothecin (CPT, Sigma-Aldrich) was added to the media for 3 or 6 hours, depending on the experiment. 250 nM UNC0638 (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 UNC0638 treatment. Actinomycin D (Sigma-Aldrich) was added to a final concentration of 5 µg/ml to the media for 6 hours and the Control is 96% EtOH.

2.3 Nucleic acid analysis methods

2.3.1 Quantification of nucleic acids

DNA and RNA nucleic acids were quantified using a Nanodrop 1000 Spectrophotemeter (Thermo Fisher Scientific). The concentration of DNA or RNA was calculated measuring the absorbance at 260 nm. The ratio 260/280 nm was used to assess the purity in both DNA and RNA extractions.

2.3.2 Purification of genomic DNA

In order to prepare standards for qPCR analysis, genomic DNA was isolated from control lymphoblast cells (GM04844), using the first steps of the DIP protocol (Chapter 2.3.7). After protein digestion, genomic DNA was precipitated with 100% isopropanol and washed with 75% EtOH. Washed genomic DNA was resuspended in ddH_2O, sonicated for 10 s at medium power (Bioruptor, Diagenode) and then purified using the QIAquick PCR Purification Kit (QIAGEN).

2.3.3 Real-time quantitative polymerase chain reaction (qPCR)

Real-time quantitative PCR was performed to measure the quantity of amplified DNA after each PCR cycle. For each 10 µl reaction, 5 µl of 2xQuantiTect SYBR green PCR master mix (QIAGEN) was mixed with 2 µl of sample (purified DNA or cDNA), 0.5 µl forward primer (10 µM), 0.5 µl reverse primer (10 µM) and 2 µl ddH_2O. Primers are listed in the Table 2.2. The data processing is based on purified genomic DNA standards with 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) and 0.04 ng/µl (1/400). All qPCR reactions were performed in technical triplicates to reduce technical inaccuracy or error.

Amplification and analysis was carried out in either the Rotorgene 3000 (Qiagen) or the StepOnePlus Real-time PCR system (Thermo Fisher Scientific). The cycling parameters for Rotorgene 3000 were 95°C for 15 min, followed by 45 cycles at 95°C for 15 s, 58°C or 62°C for 20 s, and 72°C for 20 s. The cycling reaction for the StepOnePlus was at 94°C for 2 min after which there were 40 cycles of 15 s at 94°C and 1 min at 65°C (or 5°C below lowest primer melting temperature).

2.3.4 RNA extraction and purification

Total RNA was extracted from cells using TRIzol reagent (Life Technologies). Briefly, cells were lysed by addition of 1 ml TRIzol and mixing actively. Then, 0.2 ml chloroform was added and samples incubated for 5 min at room temperature, followed by centrifugation for 15 min at 12,000 *g* at 4°C. The colourless upper aqueous phase containing the RNA was transferred to a new tube and RNA precipitated with 0.5 ml isopropanol and 20 µg glycogen. Samples were incubated for 10 min at room temperature and centrifuged for 10 min at 12,000 *g* at 4°C. RNA pellets were washed with 1 ml 75% EtOH, centrifuged for 5 min at 7,500 *g* at 4°C and air-dried at room temperature. Pellets were solubilized in RNase-free ddH_2O (25 µl for $5 \cdot 10^6$ cells). Lastly, DNase treatment was carried out by incubation of 30 U RNase-free DNase I (Roche) for 2 hours at 37 °C to remove contaminating genomic DNA, followed by heat inactivation for 15 min at 75°C.

For transfection experiments with fibroblasts, RNA was extracted using an RNeasy mini kit (Qiagen) according to the manufacturer's instructions. After removal of the medium, cells were washed with pre-warmed PBS and then lysed with Buffer RLT lysis buffer. For 6-well plates, after discarding PBS, 800 µl LRT (supplemented with β-mercapoethanol: 10 µl per 1 ml Buffer RLT) was added per well. A pipette was used to disrupt clumps and guarantee

the lysis of cells. An equal volume (800 µl) of 70% ethanol was added to the lysate and the samples were transferred to an RNeasy Mini spin column in 2 ml collection tubes. These were centrifuged for 15 seconds at 8,000 g. The flow-through was discarded and 700 µl of Buffer RW1 was added to the spin column. Samples were centrifuged for 15 seconds at 8,000 *g* and flow-through discarded. 500 µl of Buffer RPE was added to the column and samples were again spun at 8,000 *g* for 15 seconds. The flow-through was discarded and a second volume of RPE was added to the spin column. Samples were centrifuged at 8,000 *g* for 2 minutes. The RNeasy spin column was placed in a new 2 ml collection tube and spun for 1 minute at full speed to dry the membrane. The spin column was placed in a new 1.5 ml collection tube and 30-50 µl of RNase-free water was added directly to the spin column membrane. After 1 minute, the sample was centrifuged for 1 minute at 8,000 *g* to elute RNA. All spins were performed at 20 °C.

2.3.5 Reverse transcription qPCR (RT-qPCR)

1.5 µg RNA was reverse transcribed with SuperScript Reverse Transcriptase III (Life Technologies) using 100 ng random hexamer primers (Life Technologies), 50 ρmol oligo-dT primer (Life Technologies) or 10 ρmol gene-specific primers (Life Technologies). For further analysis, 2 µl of cDNA was used as a template in quantitative real-time PCR using the primers listed in Table 2.2.

2.3.6 Agarose gel electrophoresis

Agarose gel electrophoresis was carried out to separate and measure DNA fragments from ChIP, DIP and PCR products by size. The approximate length of the DNA fragment was determined by comparison with a 100 bp DNA ladder (GeneRuler, Thermo Fisher Scientific). DNA was mixed with a 1x bromophenol blue-containing

DNA loading dye (GeneRuler, Thermo Fisher Scientific) and run at 100V on a 1 – 2% agarose gel containing 1xTE buffer and 0.005% ethidium bromide. DNA bands were visualized under UV light.

2.3.7 DNA/RNA hybrid immunoprecipitation (DIP)

DNA/RNA immunoprecipitation (DIP) was performed as described previously (Skourti-Stathaki *et al.*, 2011). Cells were split 24 hours before the DIP protocol to $5 \cdot 10^6$ for lymphoblast cells. Cells were harvested and washed in cold PBS, resuspended in 800 µl ChIP Cell Lysis Buffer and incubated for 10 min on ice. Nuclei were collected by centrifugation (5 min, 500 *g* at 4°C) and resuspended in 800 µl ChIP Nuclear Lysis Buffer. Following a 10 min incubation, proteins were digested by addition of 5 µl proteinase K (Roche) and incubated for 3 hours at 55°C. After 1 hour, 3 µl Proteinase K was added and a subsequent 1 hour incubation, 2 µl Proteinase K was added. Post digestion, 200 µl of 5 M KAc was added and proteins were removed by centrifugation (5 min, 16,000 *g* at 4°C). DNA was isolated from the supernatant and precipitated by addition of 800 µl isopropanol and centrifuged (5 min, 16,000 *g* at 4°C). The DNA pellet was then washed with 75% EtOH, air-dried and resuspended in 100 µl ddH_2O. Following addition of 300 µl ChIP IP Buffer, DNA was sonicated for 3 min in medium power with 30 s on/30 s off intervals (Diagenode Bioruptor). After sonication, samples were centrifuged (10 min, 16,000 *g* at 4°C) and diluted with 2.6 ml ChIP IP Buffer. Following the preclearing step, 50 µl protein A agarose beads (Merck Millipore) were added and incubated for 1 hour, 16 rpm in the rotating wheel at 4°C. Beads were removed by centrifugation (5 min, 500 *g* at 4°C) and the supernatant was transferred to a new tube. Post preclearing, 100 µl were taken from the supernatant as the "Input" sample. The remaining material was divided into two samples with 1.34 ml material in each. Next,

14.4 µl S9.6 antibody was added in the "IP" sample and the "No IP" sample had no antibody. Immunoprecipitation was carried out overnight rotating at 16 rpm at 4°C. Finally, washes, elution and DNA purification steps were equal to those described for ChIP (see *Chapter 2.4.3*), except for the de-crosslinking step.

2.3.8 DNA/RNA hybrid slot blot

DNA/RNA hybrids were isolated as described in the DIP protocol with small modifications. Purified DNA/RNA hybrids and genomic DNA were diluted in 0.5x TE buffer 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 incubated with an S9.6 antibody (diluted 1:1000 in PBS containing 2% dried milk) for 1.5 hours and detected in a LI-COR instrument.

2.3.9 dsRNA immunoprecipitation (dsRNA IP)

dsRNA pull-down was performed essentially as described previously (Skourti-Stathaki *et al.*, 2014). Confluent cells were harvested, washed in cold PBS and lysed by resuspension in 300 µl ChIP Cell Lysis Buffer. After 10 min incubation, nuclei were pelleted by centrifugation (5 min, 500 *g* at 4°C) and then resuspended in 400 µl ChIP Nuclear Lysis Buffer. Following 10 min incubation on ice, samples were sheared in a Diagenode Bioruptor for 7 min at medium power, 30 s on/30 s off intervals. To eliminate insoluble material and cell debris, samples were centrifuged (10 min, 16000 *g* at 4°C). In the preclearing step, 50 µl protein G agarose beads (Merck-Millipore) were added and incubated with nuclear extracts rotating at 16 rpm for 1 hour. Beads were pelleted and 100 µl of the supernatant was reserved as the 'Input' sample and stored at -20°C. For dsRNA immunoprecipitation, the remaining supernatant was divided into two fractions,

1.32 ml of material was incubated with 3 µg of J2 anti-dsRNA IgG2a monoclonal antibody (Scicons) ("IP" sample) and the other fraction had no antibody ("No IP" sample). Immunoprecipitation was carried out for 1.5 hours, rotating at 16 rpm at 4°C. Antibody-dsRNA-binding protein complexes were captured by addition of 50 µl protein G agarose beads (Merck-Millipore) and incubation for 1.5 hours, on a rotating wheel at 16 rpm, at 4°C. To remove unbound material, samples were washed by resuspending the 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).

dsRNA was isolated with the TRIzol RNA extraction reagent (Thermo Fisher Scientific) according to manufacturer's instructions. RNA was eluted in RNase-free dH_2O, reverse transcribed to cDNA and analysed by RT-qPCR. The amount of dsRNA pulled-down was calculated as a % of Input after subtracting the background signal (no antibody control). V1 and S1 RNase treatments were carried out for 2 h at 37°C after the dsRNA isolation.

2.4 Protein analysis methods

2.4.1 Preparation and quantification of protein

For extracting the whole cell, cells were washed with PBS and cell pellets resuspended in RIPA buffer supplemented with 1x complete EDTA-free protease inhibitor cocktail (Roche). After incubation for 20 min on ice, samples were sonicated for 3 min, 30 s on/30 s off intervals at medium power (Bioruptor, Diagenode). Samples were centrifuged for 10 min, 16,000 *g* at 4°C. Protein was recovered from the supernatant and concentration was determined using the Bradford assay and stored at -20°C.

Protein concentration of whole cell extracts was determined according to the Bradford method (Bradford, 1976). Briefly, 5 µl of protein samples were mixed with 200

µl of the Bradford dye reagent (Bio-Rad Laboratories) and 795 µl PBS. After 5 min incubation at room temperature, absorbance at 595 nm was measured. The protein concentration was calculated based on a standard curve of seven BSA standards of known concentrations and lysis buffer used as a blank.

2.4.2 Western Blotting

SDS-PAGE separates proteins by their molecular weight. Pre-cast 3-8% Tris-Acetate or 4-12% Bis-Tris gradient NuPAGE gels (Life Technologies) were used. Protein samples were mixed with 100 mM DTT in 1xLDS Buffer (Life Technologies) and incubated on a heat block at 70°C for 10 min to denature proteins. Samples and Mark Pre-stained ladders (Life Technologies) were loaded on the gel and electrophoresis was carried out at 200 V for 45 min in MOPS SDS running buffer (Life Technologies) for pre-cast 4-12% Bis-Tris gels. Pre-cast 3-8% Tris-Acetate gels were run at 150V for 1 hour in Tris-Acetate SDS Running Buffers (Life Technologies).

After electrophoresis, gels were transferred onto Amersham Hybond-P PVDF membranes (GE Healthcare) pre-activated with MeOH for 60 s and equilibrated in 2xNuPAGE Transfer Buffer. Transfer was carried out in 2xNuPAGE Transfer Buffer containing 10% MeOH by semi-dry transfer at 15V for 2 hours using the Trans-Blot SD Semi-Dry transfer cell (Bio-Rad Laboratories). Transfer efficiency was assessed by staining of the PVDF membrane with Ponceau-S for 1 min at room temperature until bands were visible. Membranes were washed in ddH_2O to remove background staining and traces of Ponceau-S.

Membranes were blocked overnight at 4°C in PBS containing 2% dried milk and 0.1% Tween (Sigma-Aldrich). Membranes were incubated for 1.5 hours at room temperature with primary antibodies (Table 2.3), diluted in PBS containing 2% milk and

0.1% Tween-20. Following 3x 7 min washes 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, followed by 3x washes for 10 min each in PBS containing 0.1% Tween-20. All incubations and washes were performed with gentle horizontal shaking. Proteins were detected by chemiluminescence (Amersham ECL detection reagent) by exposure of X-ray films (FujiFilm) or using Li-COR software.

2.4.3 Chromatin immunoprecipitation (ChIP)

Chromatin immunoprecipitation (ChIP) was carried out as described previously with minor modifications (Kadener *et al.*, 2002; West *et al.*, 2004). Cells were split 24 hours before the ChIP protocol. On the first day of ChIP, cells were crosslinked by the addition of 36.5% formaldehyde to the culture medium at a final concentration of 0.95% and then incubated for 15 min at 37°C, rotating at 16 rpm on a horizontal shaker. Crosslinking was quenched by addition of glycine to a final concentration of 130 mM and incubation for 5 min on a horizontal shaker.

The following steps were performed at 4°C. All buffers were prepared freshly on the morning of the ChIP experiment and the protease inhibitors were added directly before use. Cells were washed with cold PBS supplemented with protease inhibitors (0.1 mM PMSF, 0.25 µg/µl leupeptin, and 1.25 µg/µl pepstatin) and lysed by resuspension in 300 µl ChIP Cell Lysis Buffer supplemented with protease inhibitors. After 10 min incubation on ice, nuclei were centrifuged (5 min, 500 *g* at 4°C) and resuspended in 400 µl ChIP Nuclear Lysis Buffer supplemented with protease inhibitors. After 10 min incubation on ice, samples were sonicated for 15 min at medium power with 30 s on/30 s off intervals (Bioruptor, Diagenode). Samples were centrifuged (10 min, 16,000 *g* at 4°C) to remove insoluble material and the supernatant was diluted with 2.6 ml ChIP IP

Buffer supplemented with protease inhibitors. For preclearing, 50 µl protein A agarose beads (Merck-Millipore) were added and incubated with nuclear extracts for 1 hour at 16 rpm at 4°C. Subsequently, the immunoprecipitation step was carried out – beads were removed by centrifugation (5 min, 500 *g* at 4°C) and 100 µl of the "Input" samples were transferred to a new tube and stored at -20°C. The remaining sample was divided into two samples with 1.34 ml each. In the "IP" samples the antibody was added and in the "No IP" samples no antibody was added. For histone modifications, the sample was divided into three fractions of 850 µl each ("histone modification IP", "total histone IP" and "No IP"). Immunoprecipitation was carried out overnight at 4°C, rotating at 16 rpm.

Next day, antibody-protein complexes were captured by addition of 60 µl protein A agarose beads over 1.5 hours and rotating the samples at 16 rpm at 4°C. Samples were washed by incubation of beads in different wash buffers rotating for 5 min, 16 rpm at 4°C and removal of wash buffers by centrifugation (3 min, 500 *g* at 4°C). The washes included ChIP Wash Buffer A, ChIP Wash Buffer B, ChIP Wash Buffer C and ChIP Wash Buffer D.

After complete removal of wash buffer, samples were eluted twice from the beads by addition of 250 µl of ChIP Elution Buffer, rotating at 16 rpm for 15 min at room temperature. After elution, the ChIP eluates and Input samples were decrosslinked in NaCl and RNase A for 4 hours at 65°C. Proteins were digested in ChIP digest buffer (40 mM Tris-HCl pH 6.5 and 10 mM EDTA) and proteinase K for 2 hours at 45°C.

Samples were purified using the QIAquick PCR Purification Kit (QIAGEN), according to the manufacturer's instructions. Briefly, samples were resuspended in 5 volumes of buffer PB and applied on the QIAquick columns by centrifugation (1 min, 16,000 *g* at room temperature). Columns were washed with 750 µl buffer PE and placed

in a clean 1.5 ml centrifuge tube. DNA was eluted with 70 – 100 µl dH_2O and used as a template for qPCR.

2.4.5 Silver Staining

Proteins were detected using the Pierce Silver Stain Kit (Thermo Fisher Scientific). Gels were washed 2x 5 min in dH_2O and fixed 2x 15 min in 30% EtOH:10% acetic acid solution. After 2x 5 min washes in 10% EtOH and 2x 5 min washes in dH_2O, gels were incubated for 1 min in a sensitizer solution (50 µL sensitizer with 25 ml dH_2O), at room temperature, following 2x 1 min wash with dH_2O. Gels were incubated in stain solution for 30 min. Gels were washed 2x 20 s with dH_2O and developed with developer solution (0.5 ml enhancer with 25 ml developer) by gentle shaking for 2-3 min until the bands appear. When the desired band intensity was reached, development was stopped with 5% acetic acid for 10 min, followed by a brief wash with dH_2O.

2.4.4 MTS cell proliferation assay

To determine the number of viable cells following treatments to knock down expression of *HTT*, the CellTiter 96 Aqueous One Solution Cell Proliferation Assay (Promega) was used. Control and Huntington's disease patient derived fibroblasts transfected with the Antisense Oligonucleotides method described in section *2.2.4* were incubated for 48 hours in a 96 well assay plate with 100 µl of cultured medium and treated with 20 µl of the CellTiter Solution Reagent. The plate was incubated for a further 3 hours and absorbance at 490 nm was measured with a 96-well plate reader.

2.4.6 Mass spectrometry

2.4.6.1 Sample preparation for mass spectrometry

The mass spectrometry experiments were carried out at Advanced Proteomics Facility, Department of Biochemistry, University of Oxford, South Parks Road, Oxford, OX1 3QU. SDS-containing protein samples were digested according to the filter-aided sample preparation (FASP) procedure as described previously (Wisniewski, Zougman and Mann, 2009).

2.4.6.2 Liquid chromatography-mass spectrometry analysis

Peptides were analysed by nano-liquid chromatography tandem mass spectrometry (nano-LC/MS/MS) using a UltiMate 3000 UHPLC system (Dionex Thermo Scientific) coupled to an Orbitrap Q Exactive mass spectrometer (Thermo Fisher Scientific) using Higher-energy Collisional Dissociation (HCD) fragmentation. In brief, peptides were loaded on a C18 PepMap100 pre-column (300 µm id x 5 mm particle size, 100Å pore size; Thermo Fisher Scientific) at a flow rate of 12 µl/min in 100% buffer A (0.1% (v/v) formic acid in H_2O). Peptides were then transferred to an in-house packed analytical column heated to 45°C (75 µm id packed with ReproSil-Pur 120 C18-AQ, 1.9 µm, 120 Å, Dr. Maisch GmbH) and separated using a 60 min gradient from 15% to 35% buffer B (0.1% (v/v) formic acid in acetonitrile) at a flow rate of 200 nl/min. Q Exactive survey scans were acquired at 70,000 resolution to a scan range from 350 to 1,500 *m/z*, Automatic Gain Control (AGC) target 3^6, maximum injection time set to 50 ms. The mass spectrometer was operated in a data-dependent mode to automatically switch between MS and MS/MS. The 10 most intense precursor ions were submitted to HCD fragmentation using an MS/MS resolution set to 17,500, a precursor AGC target set to 5^4, a precursor isolation width set to 1.5 Da, and a maximum injection time set to 120 ms.

2.4.6.3 Mass Spec data analysis using MaxQuant and Perseus

Unless otherwise stated, the figures show 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 an unpaired, two-tailed distribution Student's *t*-test. Correlations were calculated using Pearson's correlation coefficient (r).

MS raw files were processed with MaxQuant software, version 1.5.0.0 (Cox and Mann, 2008; Cox *et al.*, 2011) and analysed by Dr Ashling Holland (Wood laboratory). The raw files were searched against the human, *Homo sapiens*, UniProt KB FASTA database (20,219 proteins, August 2017) for MS/MS based peptide identification via Andromeda. Enzyme specificity was set to trypsin and Lys-C, with a minimum number of seven amino acids required for peptide identification, a maximum of two missed cleavages and the FDR was set to <1% at the peptide and protein level.

Default settings were used for variable and fixed modifications (variable modification; acetylation (N-terminus) and methionine oxidation; fixed modifications; carbamidomethylation). Peptide identification was performed with the precursor mass deviation up to 4.5 ppm after time-dependent mass calibration and an allowed fragment mass deviation of 20 ppm. The MaxLFQ algorithm was used for intensity determination and normalisation procedures in label-free protein quantification. This method is fully compatible with peptide and protein separation prior to MS analysis. The minimum ratio count was set to two. Quantitative of high-resolution peptide profiles was based on mass-to-charge (*m/z*), retention time and intensity values.

MaxQuant was used to calculate pairwise protein ratios by taking the median of all pairwise peptide ratios per protein. Only shared identical peptides were considered for each pairwise comparison, with a minimum number of one ratio count required for each

pairwise comparison. A least-squares analysis was used to rebuild the relative abundance profile for individual proteins. A minimum of two ratio counts was set for protein quantification with the 'requantify' and 'match between runs' function enabled. For matching between runs, the retention time alignment window was set to 30 min and the match time window was 1 min.

Perseus (version 1.5.5.3) was used for statistical and annotation enrichment analysis based on 3 independent Biological replicates, freely available open-source bioinformatics platforms were employed. Proteins that were identified only by site modification, contaminants, found in the decoy database or quantified in only one of the replicates were excluded from protein analysis. Sample LFQ intensity values were log2 transformed and samples were assigned to their corresponding groups. Missing values were imputed on the basis of a normal distribution (width = 0.3, downshift = 1.8) to replace missing values with numbers that represent low abundant proteins. A two-side *t*-test statistic was employed for pairwise comparison of proteomes. All other statistical and bioinformatics data analyses were performed using the freely available software STRING version 10.0, PANTHER, version 11.1 and Kyoto Encyclopedia of Genes and Genomes (KEGG), release 77.1. STRING (http:string-db.org/) was used to illustrate known and predicted protein-protein interactions. Data was displayed in confidence view using the high confidence (0.700) parameter for analysis. Differentially expressed protein analysis was performed using the PANTHER (http://pantherdb.org/) to perform GO analysis. To determine which pathways play a potential role in our HD samples, all differentially expressed proteins were subjected to Kyoto Encyclopedia of Genes and Genomes (KEGG).

Chapter 3 | Transcriptional and epigenetic alterations of *HTT* gene associated with CAG repeats

3.1 Transcriptional dysregulation of *HTT* gene in Huntington's disease

The gene responsible for the development of Huntington's disease was identified in 1993 and shown to contain a CAG repeat localised in its first exon with expansion in HD cases (HD Collaborative Research Group, 1993). The gene, named *Huntingtin* (*HTT*), encodes a mutant protein (HTT) known as being the main molecular cause in the pathology of HD (McMurray, 2001). However, the most recent studies have revealed novel pathogenic pathways directly dependent on an RNA toxicity (Chung *et al.*, 2011; de Mezer *et al.*, 2011; Banez-Coronel *et al.*, 2012; Budworth and McMurray, 2013). Moreover, the underlying molecular mechanism leading from the expanded DNA to mutant HTT protein remains largely elusive. It is possible that the *HTT* transcript may contribute to the pathology and so part of the gain of function could be activated at the RNA level. Therefore, the main objective in this chapter was to investigate the role of the CAG expansion in *HTT* gene expression and to identify the potential co-existing mechanisms that may participate in the pathogenesis of HD.

3.1.1 *HTT* RNA levels are reduced in HD

To investigate the mechanism of transcriptional regulation of the *HTT* gene in a pathological context, Epstein-Barr virus (EBV)-immortalized lymphoblastoid and non-immortalized fibroblast cells derived from healthy individuals and HD patients were purchased from Coriell Institute for Medical Research (Camden, USA) and used as cellular models (Figure 3.1A). A set of fourteen qPCR primer pairs was designed, spanning the whole *HTT* gene body with two amplicons directly surrounding the CAG triplet repeat: amplicon prom (347 bp upstream of the CAG expansion) and amplicon A (117 bp downstream of the CAG repeat) (Figure 3.1B).

Expression analysis of the *HTT* gene in these cells showed that spliced *HTT* mRNA (primers ex1-ex2 and ex66-ex67) and nascent RNA, over the exon1-intron1 junction, downstream of the CAG expansion (primers A), are reduced by approximately 40% in fibroblasts cells from HD patients with the short (50 CAG) and long (173 CAG) expansion (Figure 3.2A). In lymphoblast cells, RNA levels were also reduced by ~40% in the cell line with the long expansion (250 CAG), but not in the short expansion (50 CAG) (Figure 3.2B). Importantly, mRNA levels of control housekeeping genes, including *GAPDH*, *5S*, *γ-Actin* and *β-Actin*, were similar between control and HD cells (Figure 3.3A, B). Overall, non-immortalized fibroblast cells show more pronounced transcriptional defects. In line with this, it was shown that immortalized cells are more stable over time, they are less dependent on growth factors and their DNA synthesis is more stable compared to non-immortalized cells (Holley *et al.*, 1976; Bradley *et al.*, 1990). This may explain less dramatic transcriptional defects observed in lymphoblast cells, in the cell line with the intermediate expansion length (40/50 CAG). The reduction of *HTT* RNA was observed previously in human post-mortem HD brain samples (Evers *et al.*, 2015) and transgenic mouse models (Cha, 2007; Benn *et al.*, 2008), thus confirming the validity of this model system to study HD.

A

Cell line	CAG repeat on each allele	Genotype	Gender
Fibroblasts			
GM04845	16/20	Wild-type	Female
GM04857	40/50	Homozygous	Female
GM04281	17/68	Heterozygous	Female
GM09197	18/173	Heterozygous	Male
Lymphoblasts			
GM04844	16/20	Wild-type	Female
GM04856	40/50	Homozygous	Female
GM14044	18/250	Heterozygous	Male

B

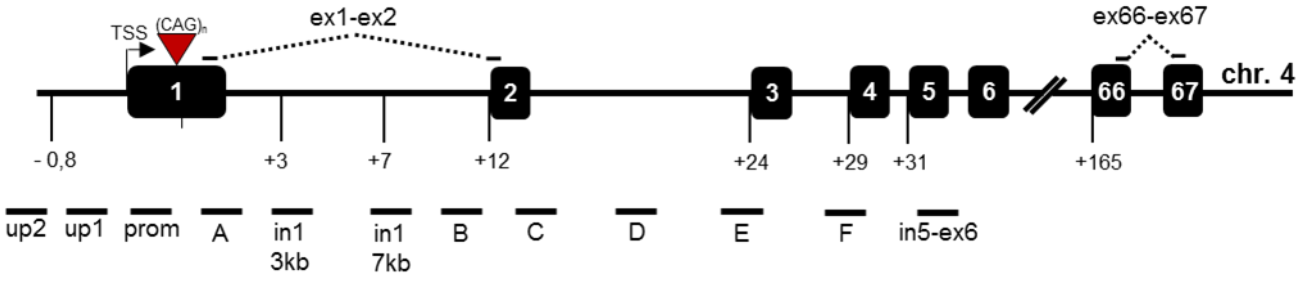
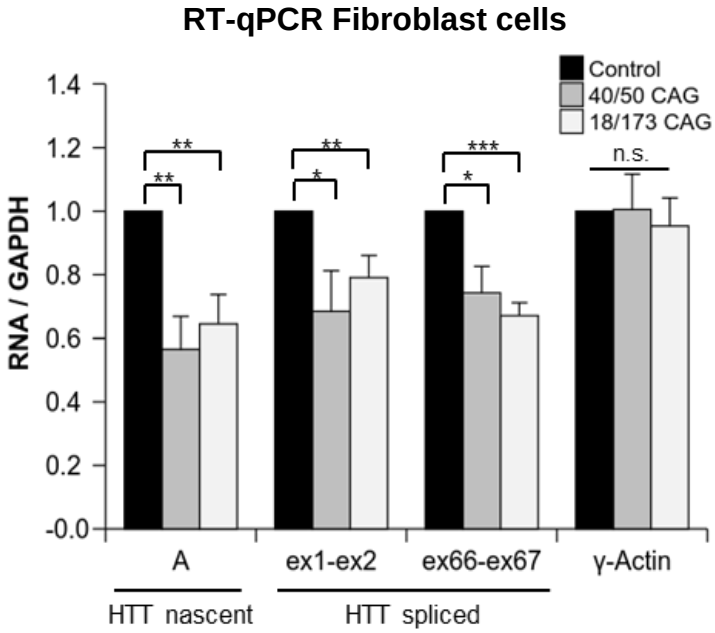


Figure 3.1 Human cell lines and diagram of *HTT* gene.

A. Fibroblast and lymphoblast cell lines used in the study. The repeat sizes, genotype and gender are indicated.

B. Diagram of *HTT* gene. Black boxes represent exons, lines are introns and red triangle is $(CAG)_n$ expansion, n=16-35 in normal population and n>36 in HD patients. TSS is the major transcriptional start site. Numbers indicate distances in kb relative to TSS. Amplicons for qPCR are shown below the gene diagram.

A



B

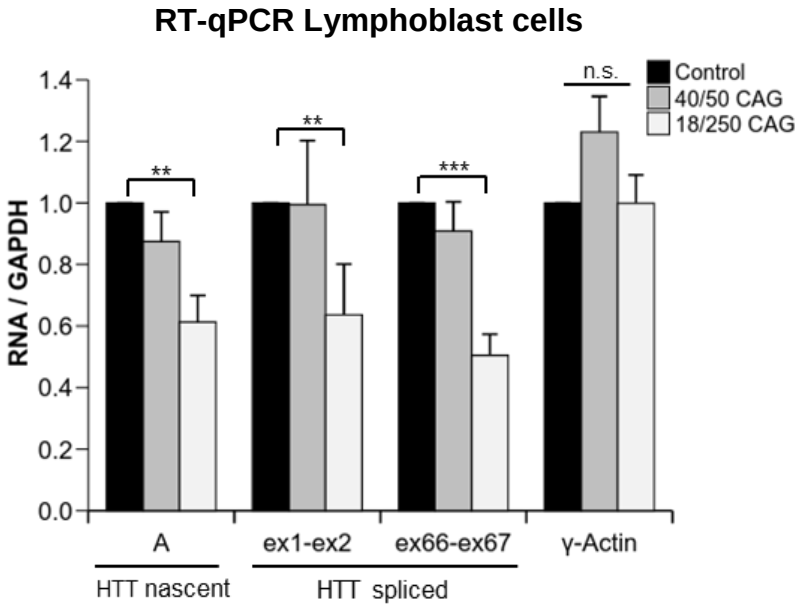


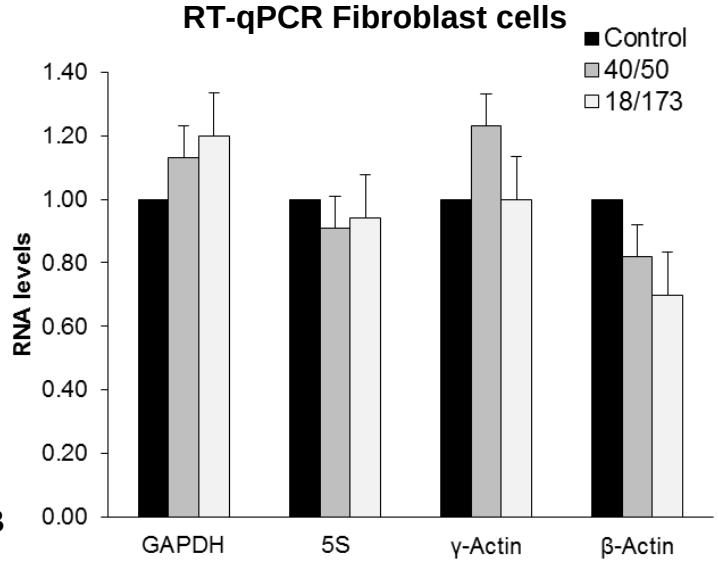
Figure 3.2 RT-qPCR analysis of the human *HTT* gene.

A. RT-qPCR analysis of *γ-Actin* and *HTT* genes in control and HD fibroblast cell lines (n=5). Values are normalised to GAPDH and relative to control cells.

B. RT-qPCR analysis of *γ-Actin* and *HTT* genes in control and HD lymphoblast cell lines (n=5). Values are normalised to GAPDH and relative to control cells.

Asterisks (*) indicate statistical significance (* $p<0.05$; ** $p<0.01$; *** $p<0.001$), based on an unpaired, two-tailed distribution Student's t-test.

A



B

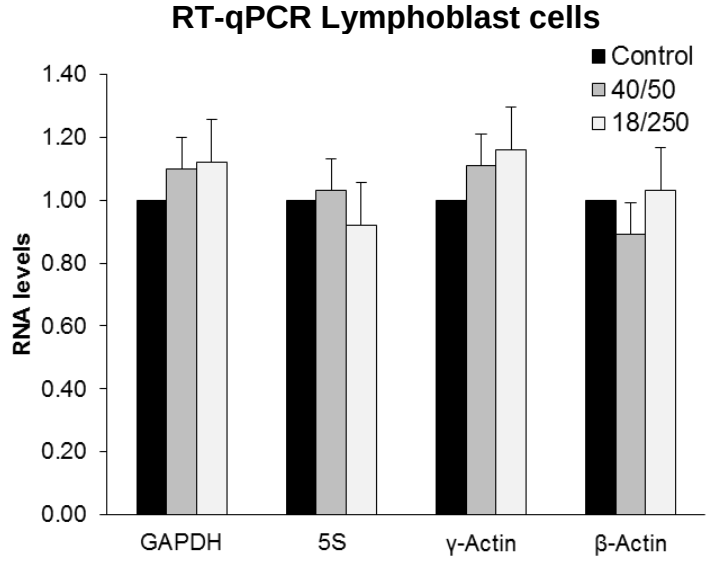


Figure 3.3 RT-qPCR analysis of housekeeping genes.

A. RT-qPCR analysis of *GAPDH*, *5S*, *γ-Actin* and *β-Actin* genes in control and HD fibroblast cell lines (n=5). Values are relative to control cells.

B. RT-qPCR analysis of *GAPDH*, *5S*, *γ-Actin* and *β-Actin* genes in control and HD lymphoblast cell lines (n=5). Values are relative to control cells.

Asterisks (*) indicate statistical significance (* $p<0.05$; ** $p<0.01$; *** $p<0.001$), based on an unpaired, two-tailed distribution Student's t-test.

3.1.2 RNA Pol II levels are decreased at the CAG-repeat region in HD

The transcription of DNA to messenger RNA is one of the most highly regulated processes in the cell. Gene expression depends on the orchestrated regulation of several transcription factors that interact with each other and with the DNA regulatory elements that specify the role of the gene. In order to investigate the mechanism of *HTT* transcriptional regulation, associated with CAG expansion, Chromatin immunoprecipitation (ChIP) using H224 antibody recognising total RNA Polymerase II (Pol II) was performed in lymphoblast cells (Figure 3.4). Here, a set of nine qPCR primer pairs was employed, spanning the *HTT* gene locus with two amplicons directly surrounding the CAG trinucleotide repeat: amplicon prom and amplicon A (Figure 3.1B). ChIP analysis showed that in control lymphoblasts RNA Pol II is highly enriched over the promoter region (prom), which is located 27 bp downstream of the major transcriptional start site (TSS) (HD Collaborative Research Group, 1993). This promoter-proximal peak, found closely downstream of the start site, is a defining feature of RNA Pol II ChIP and RNA-sequence patterns of many genes (Adelman and Lis, 2012; Jonkers and Lis, 2015). Strikingly, RNA Pol II levels were significantly reduced over promoter in HD patient cells compared to control cells (Figure 3.4). This result is in line with nascent RNA profile and suggests a defect in transcription initiation in HD cells. Interestingly, the decrease of Pol II levels was also evident downstream of the CAG repeats (amplicon A). The levels of Pol II across the gene body were lower (amplicons C-F) and although not statistically significant, Pol II levels were also reduced in the *HTT* gene body in HD cells (amplicons B and C), whereas no significant changes were seen in Pol II binding to the control *γ-Actin* gene. Overall, both Pol II and RNA profiles highlighted a prominent transcriptional defect of *HTT* gene, associated with the CAG

expanded repeats. This suggests that a deficiency in transcription initiation may contribute to reduced *HTT* mRNA levels, contributing to the HD pathogenesis.

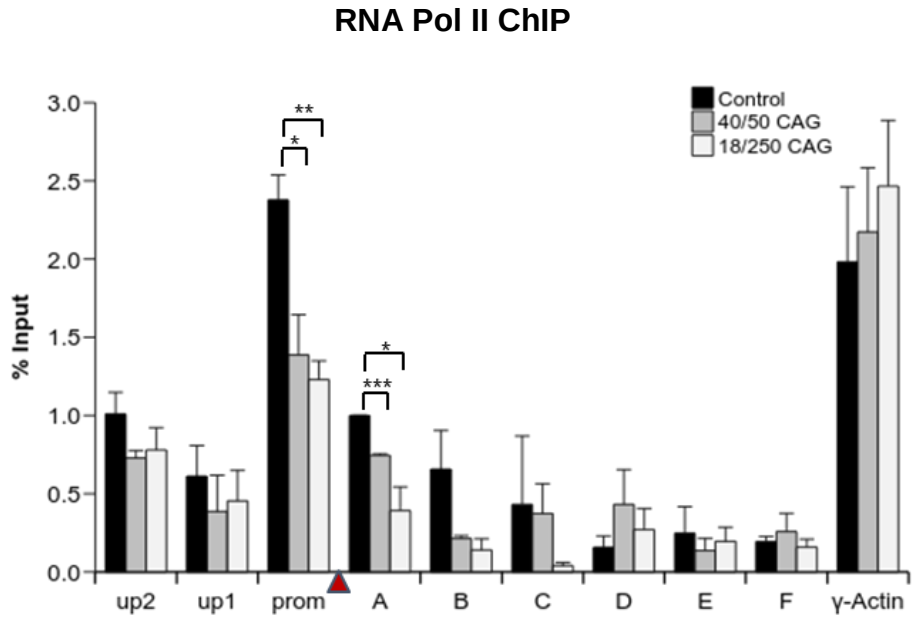


Figure 3.4 RNA Polymerase II levels are decreased at the *HTT* gene in HD cells.
ChIP analysis show that RNA Polymerase II is bound to the *γ-Actin* and *HTT* genes in control and HD lymphoblast cells with short (40/50) and long (18/250) CAG tracts (n=3) using H224 antibody recognising total RNA Pol II.
%Input is calculated based on the ratio (IP – No IP) / Input. IP is the immunoprecipitated material, No IP is the non-immunoprecipitated, similar to the IP sample but without antibody and Input is the total starting material.
Asterisks (*) indicate statistical significance (* $p<0.05$; ** $p<0.01$; *** $p<0.001$), based on an unpaired, two-tailed distribution Student's t-test.

3.1.3 HD tissue-dependent regulation of RNA Pol II dynamics

The above results showed that transcription initiation of the *HTT* gene is impaired in HD lymphoblast cells. In order to investigate whether transcription defects are tissue-specific, RNA Pol II ChIP was employed in human HD patient neurons derived from induced pluripotent stem cells (iPSCs).

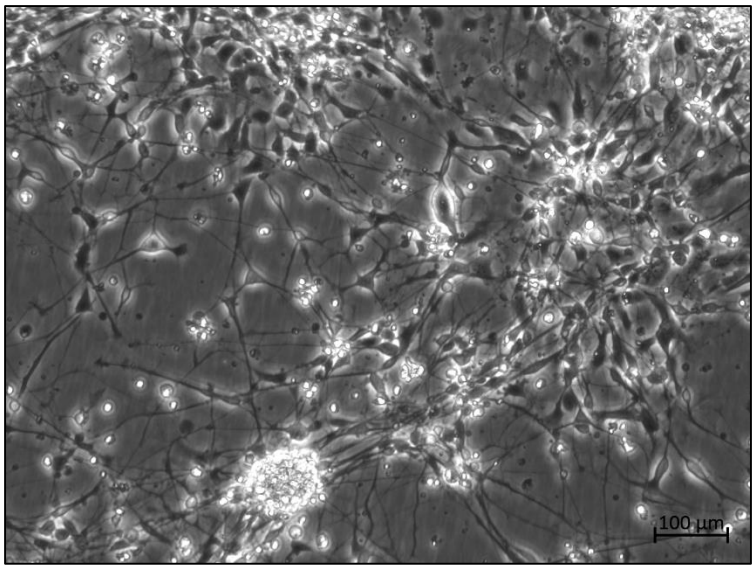
Major pathology occurs most prominently in the striatum and the cerebral cortex (Pillai *et al.*, 2012; Hadzi *et al.*, 2012). To this end, the collaboration with Prof Sarah Tabrizi's laboratory (University College London) allowed us to have differentiated and cultured human iPSCs-derived neurons. The cells were generated using the medium spiny neurons (MSN) differentiation protocol (Arber *et al.*, 2015) from patient-derived fibroblast with 73 CAG repeats and control-derived fibroblasts with 20 CAG repeats, which had been converted to iPSCs.

ChIP experiments were performed using the principal neurons of the striatum, the medium spiny neurons (MSNs) derived from three clones of HD patient-derived human iPSCs and three clones of non-affected control patient-derived human iPSCs (Figure 3.5A, B). As shown in Figure 3.6A, RNA levels were dramatically reduced in the region surrounding the CAG repeats of the *HTT* gene. Interestingly, RNA levels were significantly reduced downstream of the CAG repeats (amplicons A, B, C, D and E). Validation of Biological repeats for control and HD iPSC-derived neurons indicate that transcriptional dysregulation of *HTT* gene is a distinct feature of HD cells.

Moreover, RNA Pol II is enriched over promoter region (prom) and strikingly decreased over amplicons prom and A, 2- and 5-fold reduction, respectively (Figure 3.6B). The prom amplicon detects both pre-mRNA and mRNA and amplicons A-F measure pre-mRNA levels. Overall, our results show a stronger reduction of RNA and Pol II levels in iPSCs-derived neurons when compared to the lymphoblast cells (Figure 3.4). This suggests that transcription initiation is regulated in a tissue-dependent manner that may contribute to the disease pathogenesis. Interestingly, analysis of the *HTT* RNA and Pol II underlined a prominent transcriptional defect downstream of the CAG expansion (Figure 3.6), suggesting that a deficiency in transcription elongation might also play a role in HD pathology. It is interesting that the CAG expansion is affecting the

region at the promoter gene, possibly due to the widespread crosstalk between different co-transcriptional processes, including defects in transcriptional initiation and elongation.

A



B

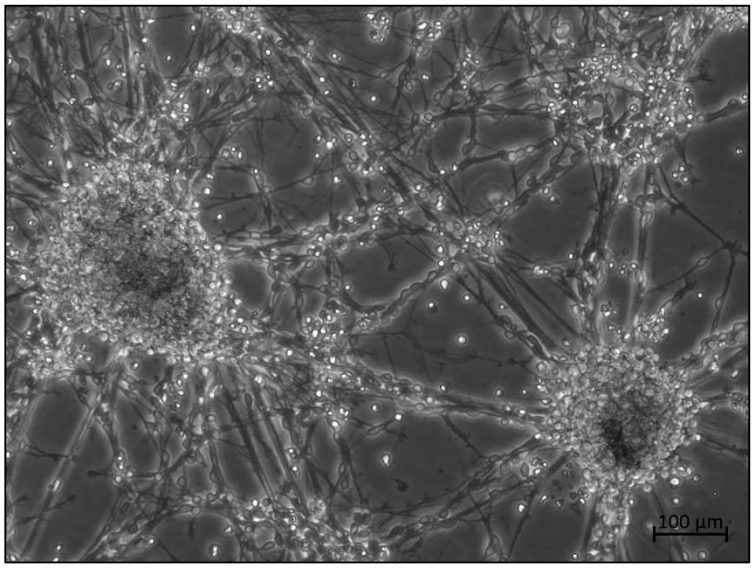


Figure 3.5 HD iPSCs-derived neurons at day 36.

A. Representative image of control medium spiny neurons (with 20 CAG repeats) at day 36. Microscopy was performed by Caroline Casey, PhD student in Sarah Tabrizi's lab, University College London.

B. Representative image of HD medium spiny neurons at day 36. Microscopy was performed by Caroline Casey, PhD student in Sarah Tabrizi's lab, University College London.

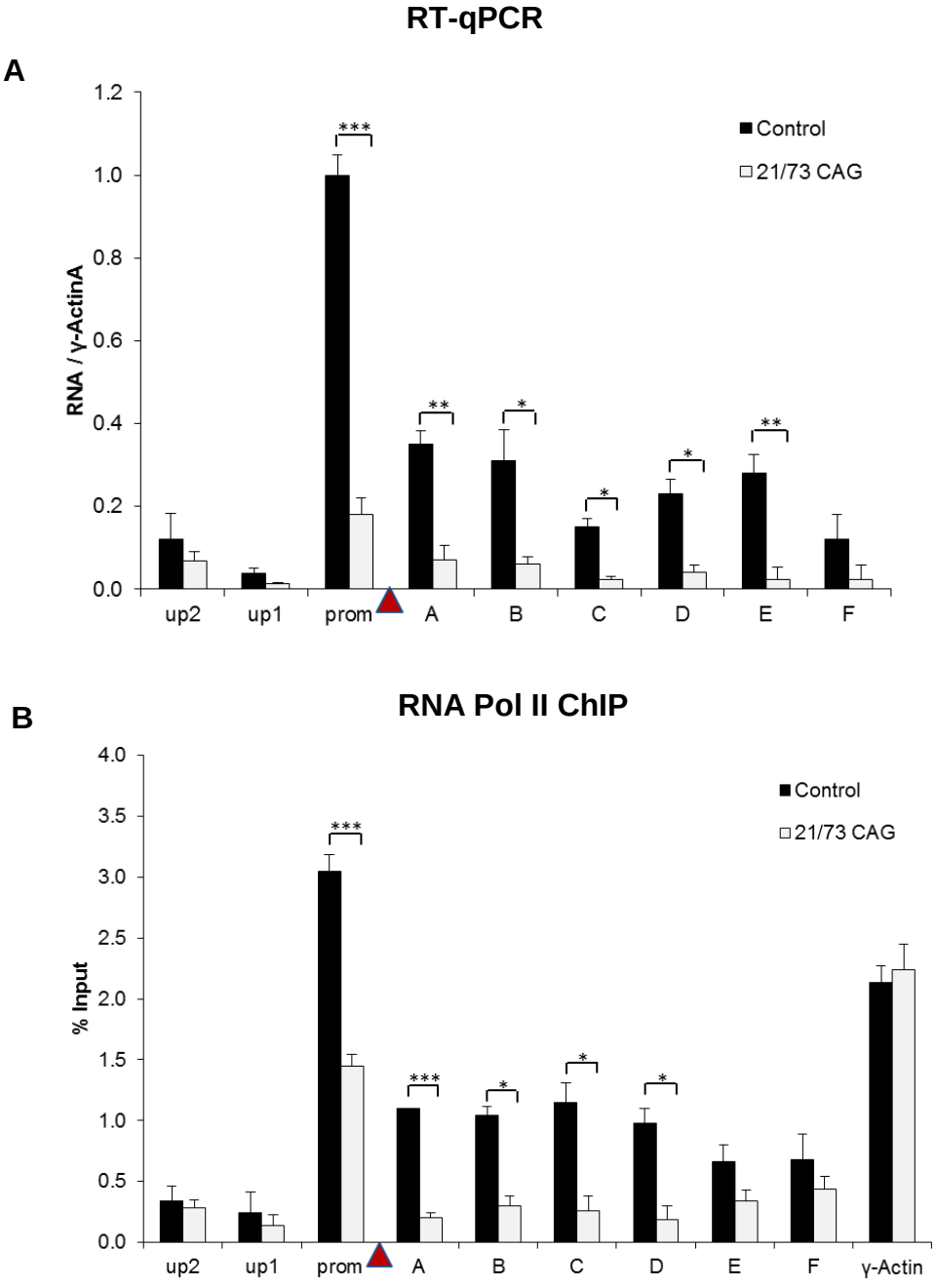


Figure 3.6 Transcription initiation and elongation defects in HD iPSCs-derived neurons.

A. RT-qPCR analysis of *HTT* RNA in HD iPSCs-derived neurons from three HD clones with 21/173 CAG repeats (n=3). The values are normalized to *γ-Actin* and relative to *HTT* prom RNA in control cells.

B. ChIP analysis shows that RNA Polymerase II is transcribing both *γ-Actin* and *HTT* genes in control and HD iPSCs-derived neurons with 21/173 CAG repeats (n=3). The values are relative to *HTT* amplicon A in control cells.

Asterisks (*) indicate statistical significance (* $p<0.05$; ** $p<0.01$; *** $p<0.001$), based on an unpaired, two-tailed distribution Student's t-test.

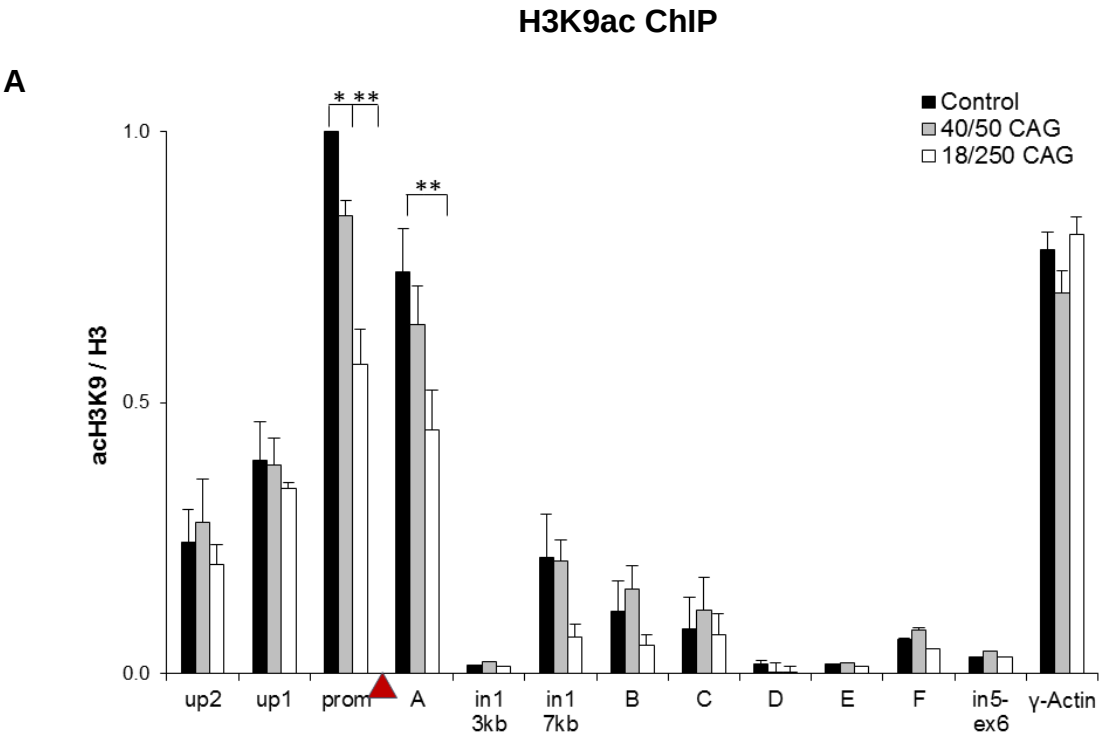
3.2 Transcriptionally impaired *HTT* correlates with increased repressive histone marks

Gene expression is regulated by epigenetic changes, including acetylation and methylation of histones, which were shown to be altered in transgenic mouse models of HD (Sadri-Vakili and Cha, 2006; Bowles *et al.*, 2012; McFarland *et al.*, 2012). It has been demonstrated that decreased acetylated histone H3 is linked to global downregulated gene expression (McFarland *et al.*, 2012). More generally, acetylation of lysine residues is associated with transcriptionally active genes, whereas methylation of lysine residues leads to transcriptional repression and is a marker of silenced genes. It was therefore important to investigate if the presence of expanded CAG repeats may alter the epigenetic landscape of the *HTT* gene. Histone modifications modify the accessibility of chromatin to the transcriptional machinery, therefore this study investigated their distribution across the *HTT* gene in control and HD lymphoblast cells.

3.2.1 Histone acetylation is altered at gene loci in HD

It is known that the acetylated histone H3 at lysine 9 (H3K9ac) is associated with active promoters and is considered to be a major hallmark of active transcription (Karmodiya *et al.*, 2012). H3K9ac ChIP analysis showed that this modification is strongly increased at the promoter region of *HTT* gene (amplicons prom and A), spanning both up- and downstream of the CAG repeats, similarly to the positive control, the highly expressed *γ-Actin* gene (Figure 3.7A). Moreover, the H3K9ac profile highly correlated

with the Pol II distribution, being enriched at the 5' end of the *HTT* gene, aiding gene activation, and was significantly reduced in HD cells (Figure 3.7A). Notably, these results were not due to differences in total histone H3 levels between control and HD lymphoblast cells, as shown by H3 total ChIP (Figure 3.7B). Moreover, this analysis revealed reduced levels of total histone H3 at the prom amplicon in control and HD cells, indicating a nucleosome-depleted region, which is frequently found associated with active promoters (Jonkers and Lis, 2015).



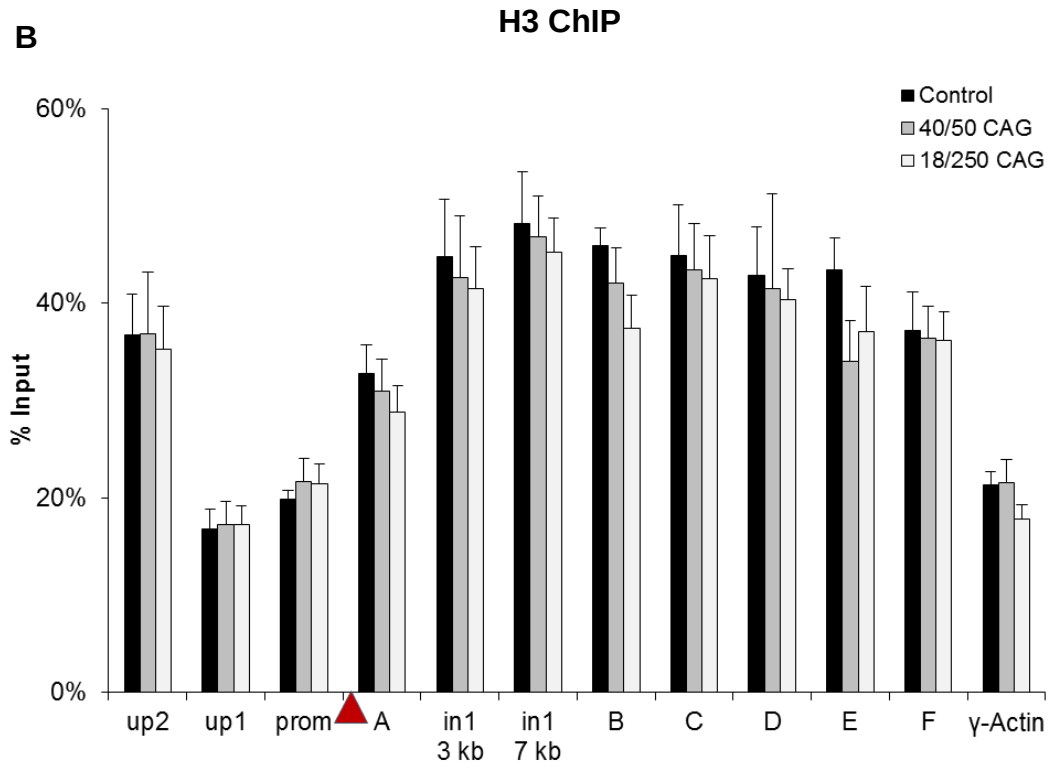


Figure 3.7 H3K9ac active histone mark is reduced in HD patient cells.

A. ChIP analysis show that the active histone mark H3K9ac is reduced on *γ-Actin* and *HTT* genes in control and HD lymphoblasts with short (40/50) and long (18/250) CAG tracts (n=3). H3K9ac levels were normalised to the total H3 levels, as determined by ChIP. *γ-Actin* is used as positive control.

B. ChIP analysis show that total histone H3 levels are similar on *HTT* gene in control and HD lymphoblast cells with short (40/50) and long (18/250) CAG tracts (n=5).

%Input is calculated based on the ratio (IP – No IP) / Input. IP is the immunoprecipitated material, No IP is the non-immunoprecipitated, similar to the IP sample but without antibody, and Input is the total starting material.

Asterisks (*) indicate statistical significance (* $p<0.05$; ** $p<0.01$; *** $p<0.001$), based on an unpaired, two-tailed distribution Student's t-test.

3.2.2 Repressive histone marks are involved in the *HTT* gene regulation

3.2.2.1 H3K9me2 histone mark

On the other hand, the distribution of the major histone modification associated with heterochromatin, the dimethylated histone H3 at lysine 9 (H3K9me2), was investigated across the *HTT* gene in control and HD cells. Interestingly, ChIP analysis

showed an inverse correlation between H3K9ac profile and repressive histone mark H3K9me2, implicated in heterochromatin formation and gene silencing (Figure 3.8A). H3K9me2 ChIP analysis in control lymphoblasts cells showed low levels of this repressive mark over the 5' end of *HTT* gene (amplicons up1, prom and A, Figure 3.8A). H3K9me2 levels are elevated around expanded CAG repeats in 18/250 HD cells (amplicons up1, prom and A), associated with transcriptional repression. The effects are not dramatic between control and patient cells possibly because ChIP despite being a robust method cannot distinguish between normal and mutant gene alleles in HD cells.

Surprisingly, H3K9me2 was greatly increased in the *HTT* intron 1 and spanned to the *HTT* exon 4 region in control cells with higher levels in HD cells (amplicons in1 3kb - D, Figure 3.8A). Moreover, H3K9me2 is significantly higher than over the highly expressed *γ-Actin* gene. The enrichment of the repressive mark in HD did not spread across the entire *HTT* gene (the region downstream of amplicon D is unaffected, Figure 3.8A), as well as levels of *γ-Actin* gene were similar between control and HD cells, thus confirming the specific and localised increase of this heterochromatin mark in the *HTT* gene. Currently, the Biological function of such high H3K9me2 repressive mark in the *HTT* gene is not clear and it requires further study. This analysis suggests that heterochromatin formation in Huntington's disease is due to expanded CAG repeats and may contribute to *HTT* gene repression in human cells.

3.2.2.2 H3K9me3 histone mark

Epigenetic-chromatin deregulation is becoming a predominant feature in HD. The abnormal epigenetic landscape can contribute to dysregulation of gene expression and neuronal dysfunction in HD (Lee, Hwang, Kim, *et al.*, 2013). H3K9me3 is a histone mark known to be associated with decreased transcriptional activity in gene expression

profiling. H3K9me3 levels and the methyltransferase SETDB1 were shown to be increased in HD transgenic animals (Ryu *et al.*, 2006), and the histone mark has been associated with the formation of constitutive heterochromatin domains and so gene silencing (Lee, Hwang, Kim, *et al.*, 2013; Lee, Hwang, Shin, *et al.*, 2013). Thus, we investigated the profile of nucleosomes containing H3K9me3 on the *HTT* gene using the ChIP method (Figure 3.8B). In contrast to H3K9me2, we observed low levels of H3K9me3 marks across the *HTT* gene in control lymphoblast cells. Interestingly, the levels of H3K9me3 were increased by more than 2 fold in cells with long CAG repeats, compared to control cells (amplicons prom - D, Figure 3.8B). These observations suggested that transcriptional dysfunction is associated with chromatin remodelling in a CAG expansion-dependent manner.

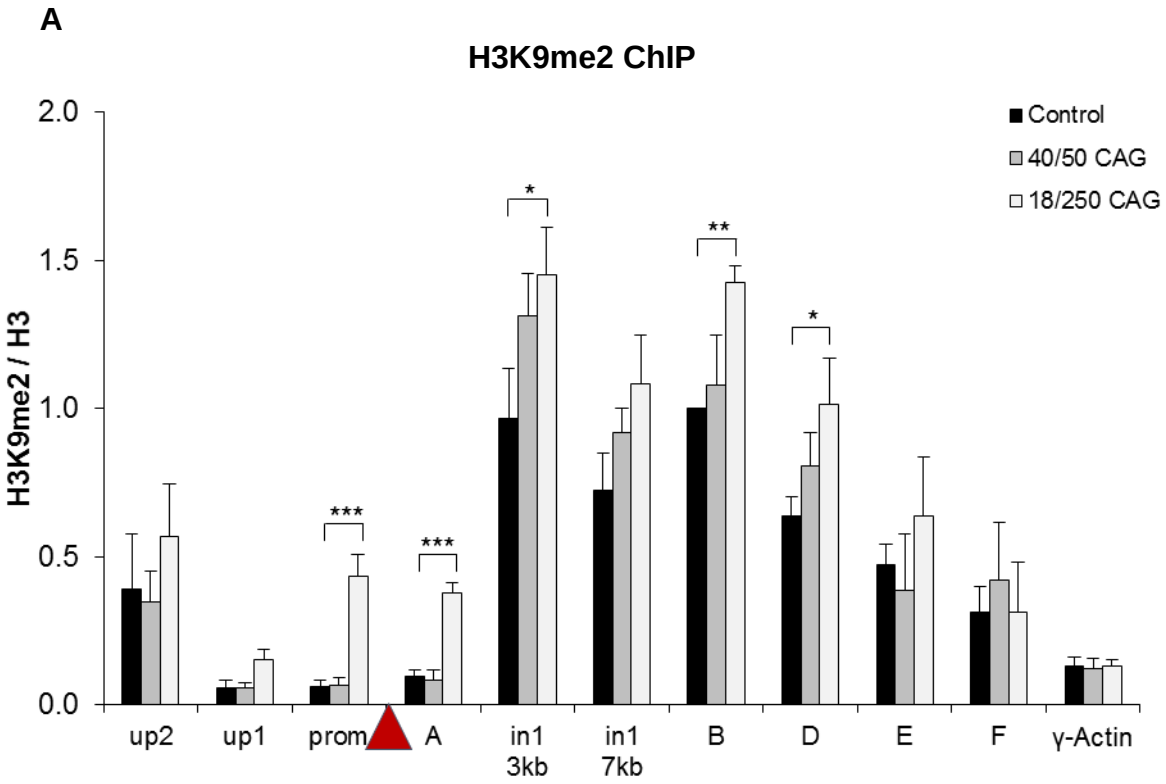
3.2.2.3 H3K27me3 histone mark

Overall, it is known that methylated H3K9 and H3K27 are related to transcriptionally inactive regions (Biagioli *et al.*, 2015). Trimethylated lysine 27 on histone H3, H3K27me3, a repressive histone mark present in promoters and enhancers, has been proposed to be modulated by CAG-repeat size in both embryonic stem cells (ESCs) and neuronal progenitor cells (NPCs) (Seong *et al.*, 2010; Biagioli *et al.*, 2015). Genome-wide studies also revealed that *HTT* can have different roles on H3K27me3 regulation depending on cell differentiation stage (Biagioli *et al.*, 2015).

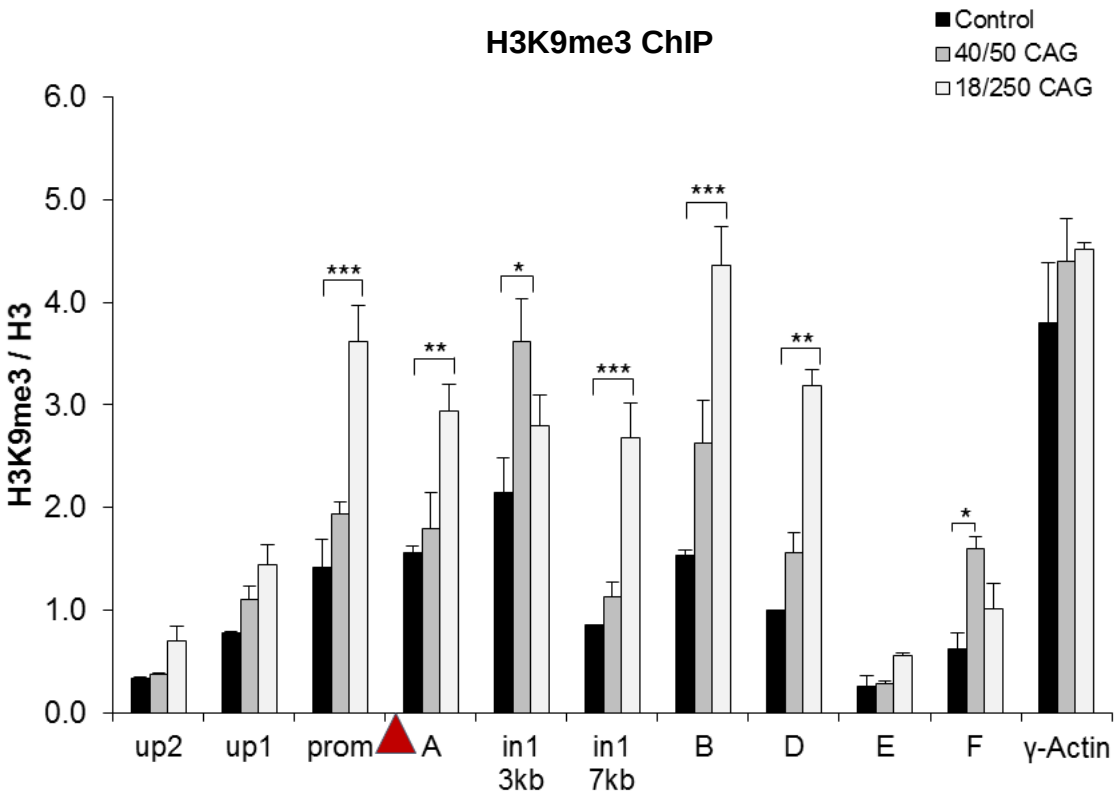
H3K27me3 ChIP analysis of the *HTT* gene in control human cells showed a peak within *HTT* intron1 at 3 kb. The levels are 5-fold higher compared with the region upstream and downstream of the *HTT* intron1 and the repressive mark is also significantly higher than the highly expressed *γ-Actin* gene. Interestingly, the levels of H3K27me3 were increased in the cells with the CAG expansion, compared to controls (amplicons

prom, A and in1 3kb, Figure 3.8C). These observations strongly suggest the importance of heterochromatin formation in the transcriptional repression of *HTT* gene and supports the idea that epigenetic mechanisms are affected by the CAG expansion.

Overall, the epigenetic-chromatin deregulation shown by the spread of repressive histone marks and loss of active histone modifications can modulate the *HTT* gene and transcription initiation efficiency.



B


H3K9me3 ChIP
Control
40/50 CAG
18/250 CAG
H3K9me3 / H3
6.0
5.0
4.0
3.0
2.0
1.0
0.0
***
**
*
***
***
**
*
up2
up1
prom
A
in1 3kb
in1 7kb
B
D
E
F
γ-Actin

C

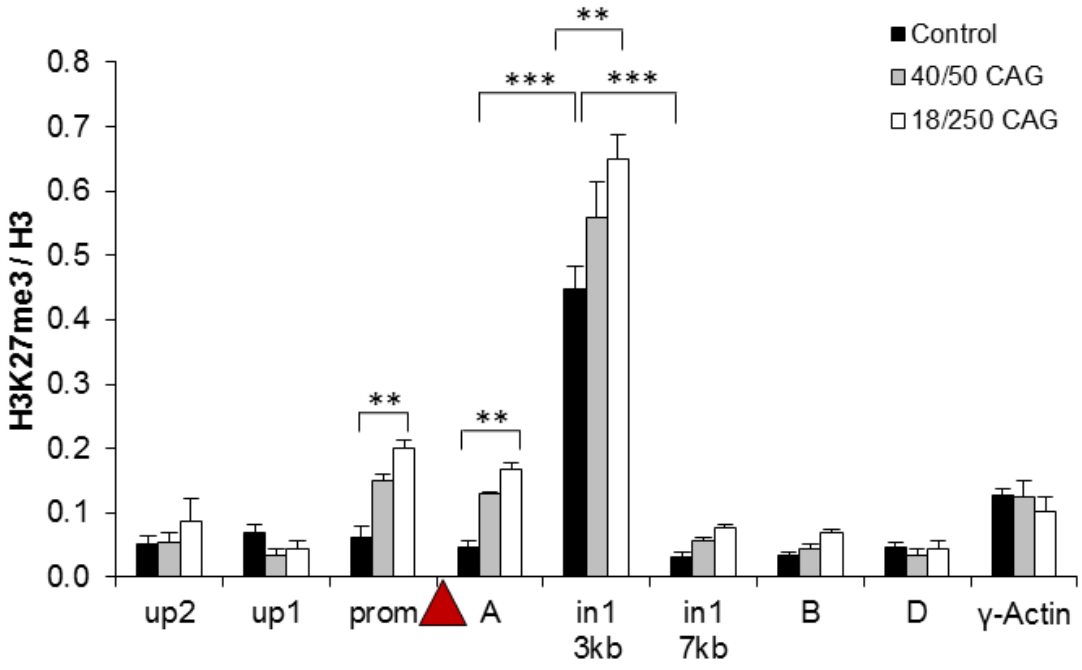

H3K27me3 ChIP
Control
40/50 CAG
18/250 CAG
H3K27me3 / H3
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
**
***
***
**
**
up2
up1
prom
A
in1 3kb
in1 7kb
B
D
γ-Actin

Figure 3.8 The repressive histone modifications are associated with the transcriptional regulation of the *HTT* gene.

A. ChIP analysis show that the heterochromatin modification H3K9me2 on the *HTT* gene is increased in HD lymphoblast cells compared to control cells, while the levels are similar on *γ-Actin* (n=3). H3K9me2 levels are normalised to the total H3 levels, as determined by ChIP. *γ-Actin* is used as background control gene.

B. ChIP analysis show that the repressive H3K9me3 mark on the *HTT* gene is increased in HD lymphoblasts compared to control cells, while the levels are similar on *γ-Actin* (n=3). H3K9me3 levels are normalised to the total H3 levels, as determined by ChIP.

C. ChIP analysis show that the repressive H3K27me3 mark on the *HTT* gene is increased in HD lymphoblast cells in the region surrounding the CAG repeats, while the levels are similar on *γ-Actin* (n=3). H3K27me3 levels are normalised to the total H3 levels, as determined by ChIP.

Asterisks (*) indicate statistical significance (* $p<0.05$; ** $p<0.01$; *** $p<0.001$), based on an unpaired, two-tailed distribution Student's t-test.

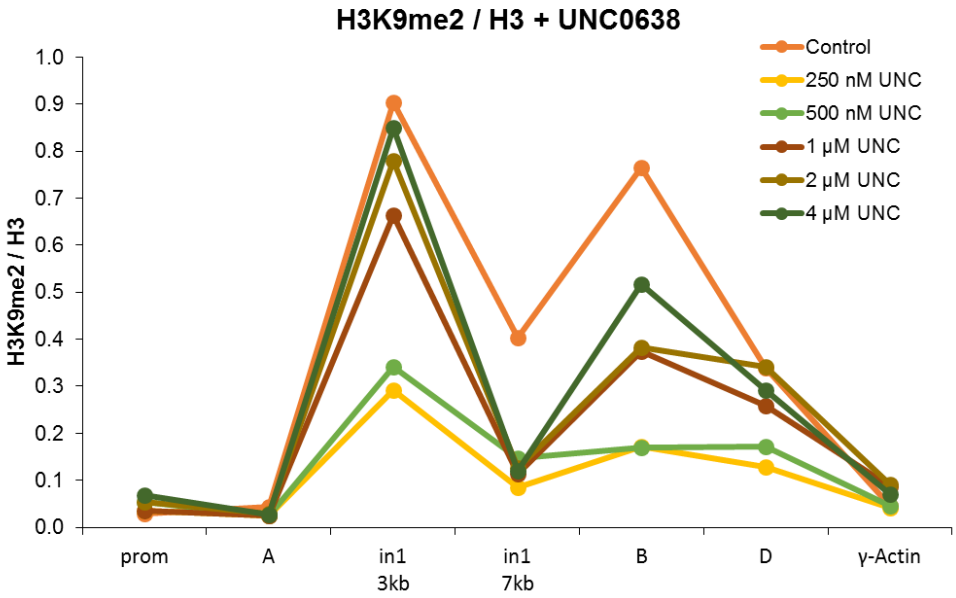
3.2.3 *HTT* transcript is not affected by methyltransferase inhibitor UNC0638 in HD

To further investigate the impact of the histone modification H3K9me2 formed on the *HTT* gene, we used the G9a histone methyltransferase inhibitor UNC0638. UNC0638 has previously shown to inhibit the enzymatic function of G9a, which mediates transcriptional repression by demethylation of histone H3 lysine-9, reducing the levels of H3K9me2 (Vedadi *et al.*, 2011; Ling *et al.*, 2012).

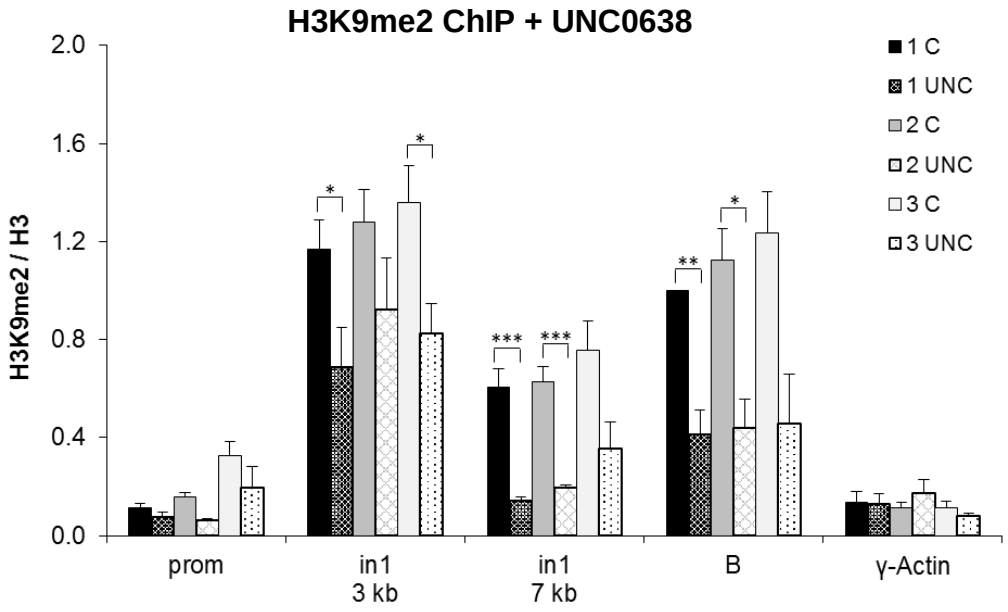
Given that H3K9me2 has a half-life of approximately 1 day (Zee *et al.*, 2010), and an exposure for longer than 48h results in greater reduction of the mark, a titration with different doses of UNC0638 was performed for 72 hours. Here, increasing doses of 250 nM, 500 nM, 1 µM, 2 µM and 4 µM were tested and H3K9me2 levels were analysed 72 hours post-treatment in lymphoblast cells (Figure 3.9A). The efficiency of the treatment with 1 µM or 2 µM UNC0638 and with 250 nM or 500 nM UNC0638 was similar, and H3K9me2 levels progressively increase when using higher doses of the inhibitor. The minimum dose of 250 nM was most effective as illustrated by the strongest H3K9me2 reduction, with an almost 3-fold decrease compared to the control. Thus, the concentration of UNC0638 used in the subsequent experiments was 250 nM.

Following treatment of control and HD lymphoblast cells with 250 nM UNC0638 for 72 hours, levels of H3K9me2 were significantly reduced (amplicons in1 3kb – B, Figure 3.9B). Knowing that UNC0638 can reduce the levels of the repressive mark and unravel the chromatin, the next step was to analyse the influence of H3K9me2 on the *HTT* RNA levels. Following UNC0638 treatment and subsequent reduction of the H3K9me2 levels, the *HTT* transcript was increased in all the amplicons in control lymphoblast cells (Figure 3.9C). Remarkably, RNA levels in the HD patients remained unchanged, suggesting that there are other repressive marks specifically playing a role in Huntington's disease, such as H3K9me3 or H3K27m3. This data suggest that it is not only the H3K9me2 chromatin modification affecting *HTT* transcripts but there is a highly dynamic and regulated process that occurs through interaction between various chromatin marks, DNA and RNA.

A



B


H3K9me2 ChIP + UNC0638
H3K9me2 / H3
2.0
1.6
1.2
0.8
0.4
0.0
1 C
1 UNC
2 C
2 UNC
3 C
3 UNC
prom
in1 3 kb
in1 7 kb
B
γ-Actin

C

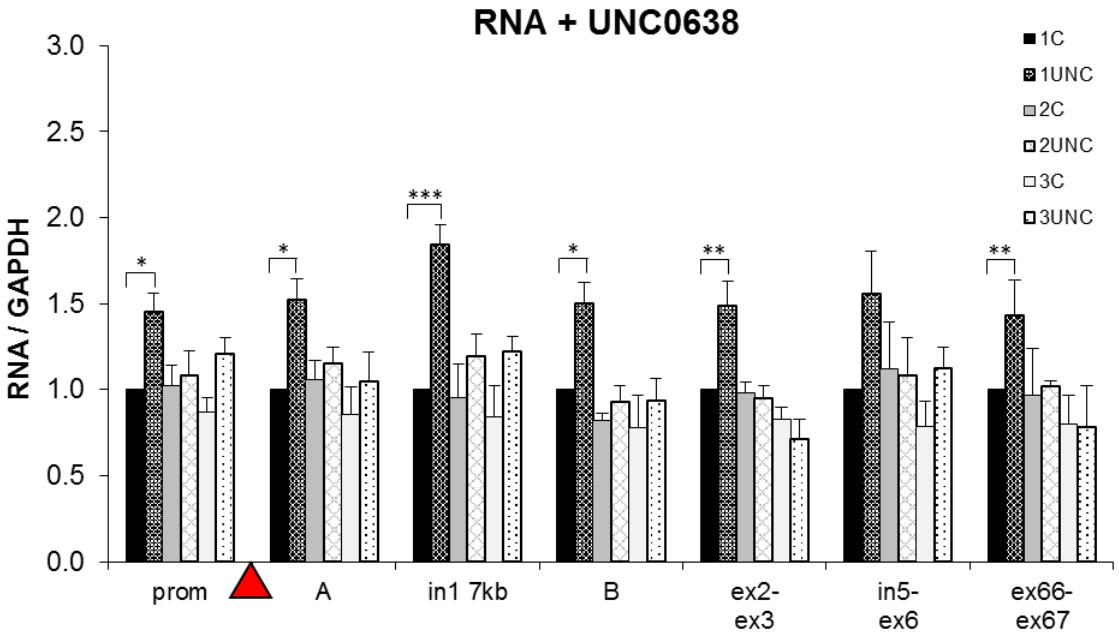

RNA + UNC0638
RNA / GAPDH
3.0
2.5
2.0
1.5
1.0
0.5
0.0
1C
1UNC
2C
2UNC
3C
3UNC
prom
A
in1 7kb
B
ex2-ex3
in5-ex6
ex66-ex67

Figure 3.9 H3K9me2 represses *HTT* transcript as revealed by UNC0638 treatment.

A. H3K9me2 ChIP analysis using a titration with different doses - 250 nM, 500 nM, 1 µM, 2 µM, 4 µM, of the inhibitor of H3K9me2 methyltransferase G9a, UNC0638 (n=2).

B. H3K9me2 ChIP analysis in control and HD lymphoblast cells show that treatment with 250 nM UNC0638 for 72 hours reduces the levels of the H3K9me2 repressive mark (n=4). H3K9me2 levels were normalised to the total H3 levels and relative to amplicon B.

C. RT-qPCR analysis show that *HTT* RNA in control lymphoblast cells increased after treatment with 250 nM UNC0638 for 72 hours, while *HTT* RNA levels remained similar in HD cells after treatment (n=5). Values are relative to untreated control cells and normalised to GAPDH RNA.

1C is Control lymphoblast cell line, 2C is HD lymphoblast cell line with 40/50 CAG repeats and 3C is HD lymphoblast cell line with 18/250 CAG repeats; UNC stands for UNC0638 treatment 250 nM.

Asterisks (*) indicate statistical significance (* $p<0.05$; ** $p<0.01$; *** $p<0.001$), based on an unpaired, two-tailed distribution Student's t-test.

3.3.4 SRSF6 is enriched at expanded CAG repeats in HD cells

Lastly, ChIP analysis on the serine/arginine-rich splicing factor 6 (SRSF6) that has been shown to bind to CAG repeats (Sathasivam *et al.*, 2013) demonstrated that this splicing factor is recruited at the 5' end of the *HTT* gene and is significantly enriched in HD patient cells (amplicons prom and A, Figure 3.10). SRSF6 regulates splicing and enables translation of spliced transcripts. Therefore, these results are consistent with Bates work, suggesting that expanded CAG repeats can affect the *HTT* gene splicing, through the binding of SRSF6, leading to production of small toxic polyglutamine proteins (Sathasivam *et al.*, 2013).

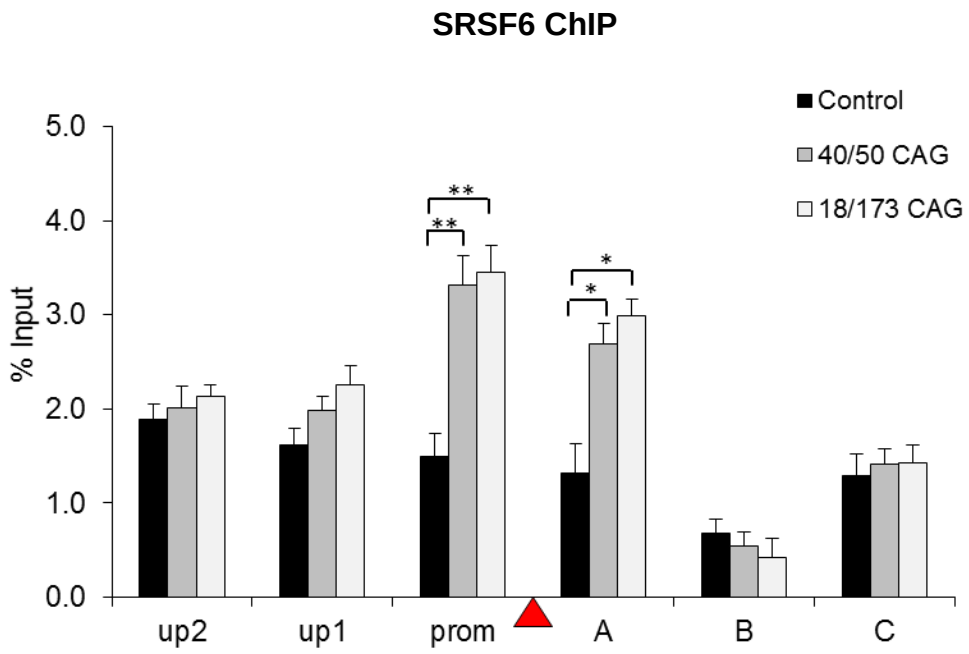


Figure 3.10 SRSF6 splicing factor is enriched in HD cells over CAG repeats.
ChIP analysis show that SRSF6 levels are enriched in the HTT gene in HD lymphoblast cells in the region surrounding the CAG repeats and were normalised to *γ-Actin* gene (n=3). %Input is calculated based on the ratio (IP – No IP) / Input. IP is the immunoprecipitated material, No IP is the non-immunoprecipitated, similar to the IP sample but without antibody, and Input is the total starting material.
Asterisks (*) indicate statistical significance (* $p<0.05$; ** $p<0.01$; *** $p<0.001$), based on an unpaired, two-tailed distribution Student's t-test.

3.3 Discussion of Chapter 3

Huntington's disease (HD) is an autosomal dominant neurodegenerative disorder caused by a CAG trinucleotide expansion within exon 1 of the huntingtin gene (*HTT*). There are no disease-modifying treatments and the diseases caused by CAG-expanded repeats in coding sequences have only recently been examined for possible RNA-related mechanisms of pathology (Budworth and McMurray, 2013). The aim of this chapter is to investigate the pathogenic molecular events caused by the expanded CAG repeats, including transcription and gene regulation of the *HTT* gene.

To study the transcriptional regulation of the *HTT* gene in HD, we used fibroblasts and lymphoblasts from healthy individuals and HD patients with different CAG repeat sizes. The results presented suggest that the presence of the CAG expansion in the *HTT* gene can affect RNA polymerase transcription, leading to dysregulation of transcription-associated processes in Huntington's disease. Overall, nascent and spliced *HTT* RNA profiles and RNA polymerase II distribution support major transcription initiation and elongation defects triggered by expanded repeats in the *HTT* gene. Furthermore, the data suggest that decreased acetylated histone is associated with downregulation of the *HTT* gene expression in HD cells. Consistent with transcriptional repression, the levels of the repressive histone marks are increased in HD patient cells around the expanded repeats. The Table 3.1 shows a summary of the results described in this Chapter. Overall, the results obtained in this chapter demonstrate that the presence of expanded CAG repeats in the *HTT* gene leads to transcriptional repression and heterochromatin formation associated with the pathology of HD.

Table 3.1 Summary of the results described in the Chapter 3.
The levels of RNA, RNA Polymerase II, H3K9 acetylation and H3K9 dimethylation after UNC9638 treatment are reduced in Huntington's' disease cells when compared to control cells (in green). In opposite, the levels of the repressive marks H3K9 di- and trimethylation and H3K27 trimethylation and the SRSF6 are increased in HD patient cells in relation to the control cells (in red). H3 total levels are similar between control and HD cells (black).

	Huntington's disease
Nascent RNA	**reduced**
Spliced RNA	**reduced**
RNA Polymerase II	**reduced**
H3K9 acetylation	**reduced**
H3K9 dimethylation	**increased**
H3K9me2 + UNC0638	**reduced**
H3K9 trimethylation	**increased**
H3K27 trimethylation	**increased**
H3 total	**similar**
SRSF6	**increased**

The pathogenesis of HD has been associated with the mutant HTT protein, however recent studies, as discussed below, have revealed novel pathogenic mechanisms associated with an RNA toxic gain-of-function. It has been shown that *Drosophila melanogaster* (fruit fly) and *Caenorhabditis elegans* (nematode worm) containing long CAG repeats can create RNA foci, blocking splicing and ultimately leading to neurodegeneration without the protein being translated (Li *et al.*, 2008; Wang *et al.*, 2011). Still, there are no indications that this mechanism exists in human HD. Another study shows that an antisense transcript of *HTT*, *HTTAS*, containing the CAG repeat downregulates the expression of the *HTT* gene (Chung *et al.*, 2011). The *HTTAS* is regulated by a weak promoter, is expressed less than the sense transcript, and is present in brain and other tissues. The effect of the CAG expansion is possibly through the

reduction of *HTTAS* expression, and would thus be expected to increase the expression of the *HTT* transcript. Moreover, long CAG expansions in HD are predicted to fold into hairpins (de Mezer *et al.*, 2011) which can be present in RNA foci affecting the splicing of other genes (Mykowska *et al.*, 2011). The long CAG hairpins can originate small CAG repeat RNAs, that are CAG repeat length dependent and neurotoxic (Banez-Coronel *et al.*, 2012).

It is possible that expanded CAG repeats in the first exon of the *HTT* mRNA induce toxicity by mechanisms implicated in gene expression deregulation. RNA toxicity means that the mutant transcript is able to induce pathogenesis. The codon repeat diseases in which the repeat expansion is not translated are usually affected at the level of transcription, either dysregulating transcription as in Fragile X Syndrome or in Friedreich's ataxia (Nelson *et al.*, 2013) or by affecting splicing as in myotonic dystrophy type 1 (DM1) and myotonic dystrophy type 2 (DM2) (Lee and Cooper, 2009). However, the RNA toxicity is more difficult to evaluate the diseases caused by CAG-expansion translated to polyglutamine, where protein and RNA toxic effects coexist. Indeed, the knowledge about the *HTT* transcript and its dynamics needs to have more research, also with the view of developing new targets for therapeutic approaches to be able to treat these devastating diseases.

It has been proposed that trinucleotide repeat expansions can have different effects depending on the cellular localization of the RNA. Those functions include alternative splicing in the nuclear compartment and regulation of mRNA translation in the cytoplasm. Similar to studies published before in human post-mortem HD brain samples (Evers *et al.*, 2015) and in central nervous system cells (Zuccato *et al.*, 2001), the results presented here demonstrate a defect in transcription initiation, thus confirming the validity of this model system to study HD. The transcriptional regulation defect was

supported by RNA Pol II ChIP and RT-qPCR data suggesting transcription initiation and elongation defects due to CAG expanded repeats. The transcription elongation defect was only seen in human HD iPSC-derived neurons, suggesting that downstream effects of CAG expansion on different phases of transcription can occur independently of each other and can be cell-type specific. It has been reported that the levels of elongating RNA Pol II at the HD locus, but not transcription initiation, were tissue-specific and are associated with CAG instability levels (Goula *et al.*, 2012). In the same way, H3K36me3, a post-translational histone modification associated with elongating RNA Pol II, was increased at the HD locus in the striatum but not in the cerebellum (Goula *et al.*, 2012). Interestingly, chromatin structure can alter CAG repeat instability due to a more open open conformation being more accessible to the DNA repair factors (Goula *et al.*, 2012). In line with this, the data presented here also suggests that transcription can modulate CAG instability in a tissue-dependent manner. The results presented suggest that the effect of CAG expansion on transcription may not be similar in every cell type, and in order to understand more deeply the effects is to expand the investigation for chromatin structure at the HD locus.

Several factors can affect transcription and alter chromatin structure. The accessibility of chromatin to the transcriptional machinery is usually altered by epigenetic changes of the gene, including histone acetylation and methylation. Previous studies have demonstrated that HD is associated with chromatin remodelling affecting many genes globally (Steffan *et al.*, 2001). Expression of mutant HTT in cultured cells decreases the level of acetylated histones H3 and H4, leading to hypoacetylation, and this reduction can be reversed by inhibitors of histone deacetylases, ameliorating the mutant *HTT* toxicity in *Drosophila* (Steffan *et al.*, 2001). Although histone modifications associated with epigenetic-chromatin deregulation are potential targets for HD therapies, it remains

unclear how chromatin structure influences CAG repeat instability and how changes in histone modifications can lead to transcription abnormalities of the *HTT* gene itself.

Interestingly, in control lymphoblast cells the active histone mark H3K9ac is enriched over the CAG repeats and reduced in HD cells, similar to the Pol II ChIP profile. Indeed, it has previously been shown that decreased acetylated histone acH3 (K9, K14) is associated with global downregulation of gene expression in HD mouse models (Bowles *et al.*, 2012; McFarland *et al.*, 2012). The acetylation process includes the transfer of an acetyl coenzyme A to the NH^{3+} on a histone lysine residue. Acetylation of lysine (K) residues on histones reduces the interactions between histones and DNA, supporting a relaxed chromatin environment.

Overall, histone deacetylase inhibitors improve transcriptional activity by increasing histone acetylation levels and making the DNA accessible to the transcriptional machinery. These inhibitors were shown to ameliorate learning and memory in mice (Guan *et al.*, 2009) and to exhibit neuroprotection in cultured neurons (Fischer *et al.*, 2010). As a result, histone deacetylase inhibitors are promising therapies in HD models as they are able to derepress transcription.

Consistent with transcriptional repression, the levels of the histone mark H3K9me2 are increased in HD patient cells around the CAG expanded repeats. Lysine methylation is well known because of its stability and crosstalk with other modifications. Histone proteins are methylated on lysine (K) residues and can be mono-, di-, or trimethylated. Specifically, methylated H3K9 and H3K27 are linked to transcriptionally inactive heterochromatin regions. It is not known whether specific loci are more sensitive to increased H3K9 methylation in HD cells.

There are some hypotheses for the phenomenon observed within the first intron of the *HTT* gene, the repressive H3K9me2 peak seen in control and HD cells. It is possible

that the long *HTT* intron 1 is providing a kinetic window for additional factors to act on transcription and/or splicing. Moreover, this histone modification could be also associated with activation of an alternative downstream transcriptional start site required for the flexibility in the regulation of *HTT* gene expression.

Recently it was suggested that CAG repeats lead to alternative splicing defects of *HTT* mRNA in the R6/2 HD mouse models (Sathasivam *et al.*, 2013). In this study the Bates laboratory predicted nine polyadenylation signals in the human intron 1 of the *HTT* gene (Sathasivam *et al.*, 2013). The two signals with the highest predictive score at 2.7 kb and 7.3 kb result in a polyadenylated transcript within the *HTT* intron 1 (Sathasivam *et al.*, 2013; Gipson *et al.*, 2013). Detailed computational prediction of the binding sites identified the serine/arginine-rich splicing factor 6 (SRSF6) mapping to CAG repeats (Sathasivam *et al.*, 2013). SRSF6 plays a role in splicing and through RNA co-immunoprecipitation they observed preferential association of SRSF6 with *HTT* mutant as compared with wild-type *HTT* RNA. The splicing factor SRSF6 binds to expanded CAG repeats in *HTT* transcripts leading to an aberrantly spliced transcript in HD patient tissues (Sathasivam *et al.*, 2013). On translation, this produces an exon 1 protein fragment in HD because there is a stop codon at the start of the intron 1 and HTT exon 1 proteins are known to be highly neurotoxic (Landles *et al.*, 2010).

SRSF6 ChIP analysis shows that this splicing factor is recruited at 5' end of the *HTT* gene and is significantly more enriched in HD patient cells. In terms of order of events, it appears that SRSF6 splicing factor binds to the 5' end of the *HTT* gene with an expanded CAG repeat. Then, the SR proteins recruit U1 small nuclear ribonucleoprotein complexes (Labourier *et al.*, 1999), which typically protect polyA signals (Berg *et al.*, 2012). Finally, if the intronic polyA cleavage sites are recognized, as U1 protection was inhibited, then a short polyadenylated *HTT* transcript would be formed. Therefore, it is

possible that a greater association of SRSF6 with expanded CAG repeats results in Pol II slows down transcription to be able to recognise the polyA signals and consequently an H3K9me2 repressive peak is formed. Moreover, the polyA signals can instruct the Pol II to mitigate transcription on downstream DNA in a manner that parallels accurate transcription termination. These studies lead to other questions: can the polyA cleavage sites be protected to reduce the repressive H3K9me2 peak within the *HTT* gene? Does the local intron 1 open a kinetic window for additional factors to act on splicing or polyadenylation site activation? This aberrant splicing, instigated by the binding of the SRSF6 to the CAG repeats is repeat length dependent and is only seen in mutant *HTT* transcripts. Thus, if this is an important pathologic mechanism, it will need to be well-thought-out when designing therapies towards the toxic HTT protein.

In addition to the H3K9me2 histone modification, the levels of H3K9me3 were elevated throughout the *HTT* gene in HD cells, confirming that altered gene transcription in HD is associated with heterochromatin. The increase of H3K9me3 has been correlated with the formation of constitute heterochromatin, and is believed to cause gene silencing in both global and local repression of transcription (Lee, Hwang, Kim, *et al.*, 2013; Lee, Hwang, Shin, *et al.*, 2013). Interestingly, the H3K27me3 chromatin mark was enriched in HD patient cells in the region surrounding the CAG repeats with a prominent peak within 3 kb of *HTT* intron 1, possibly sharing the same features of the H3K9me2 repressive peak. Indeed, H3K27 is associated with transcriptionally inactive regions and there are studies on the histone modifying enzyme PRC2, which methylates H3K27 at silent genes (Labadorf *et al.*, 2015). Although is still unclear what is the function of the H3K27me3 peak, it coincides with the first polyA signal observed in the Bates study (Sathasivam *et al.*, 2013), suggesting a link between H3K27me3 and aberrant splicing in *HTT*. Moreover, it has been demonstrated that in HD-affected human brain there is loss

of neuronal PRC2-H3K27me3 sites and increase of some PRC2 target genes. Thus, this data suggests that gene expression controlled by epigenetic changes in the genome is dependent on the chromatin environment and cell type context.

It is apparent that densely packed nucleosomes create a physical barrier for the proficient recruitment and processivity of RNA Pol II. Thus, genes present in packed chromatin regions are less likely to involve RNA Pol II into functional transcription. Overall, the results supported by RNA Pol II ChIP and RT-qPCR data and by decreased levels of chromatin accessibility in HD suggest that the pathomechanism is based on transcriptional dysregulation. Here it is demonstrated that the formation of repressive heterochromatin at expanded CAG repeats is a key event associated with transcriptional repression of the *HTT* gene in HD.

The effects of the G9a histone methyltransferase inhibitor UNC0638 was tested, in order to investigate the impact of the histone modification H3K9me2 in more detailed. UNC0638 has previously been shown to inhibit the enzymatic function of G9a, which is mainly responsible for demethylation of H3K9 (Vedadi *et al.*, 2011; Ling *et al.*, 2012). Interestingly, as demonstrated by ChIP, H3K9me2 is reduced in control and HD patient lymphoblast cells after treatment for 72 hours with 250 nM UNC0638. Moreover, deregulation of H3K9me2 leads to partial reactivation of the *HTT* gene in HD cells: a 2-fold increase of the *HTT* RNA in control cells. Although H3K9me2 repressive mark is highly enriched in the *HTT* gene and is a clear mark of HD, reducing the levels of H3K9me2 does not result in re-activation of *HTT* RNA in HD patient cells. This is most likely because other repressive marks play a role in HD in an orchestrated and non-independent way. This data may also suggest that secondary RNA and/or DNA toxic structures formed in HD (discussed in Chapter 4) appear earlier than the repressive histone modifications, forming a structural block and impeding the full reactivation of

HTT alleles containing CAG repeats. In this scenario, histone methyltransferase inhibitors, such as UNC0638, may be unable to reactivate *HTT* RNA because the toxic nucleic acid structures are not affected by the treatment and sustain the repressed gene. The data shown here indicates a potential epigenetic signalling cascade including H3K9 demethylation by G9a methyltransferase, followed by H3K9 trimethylation and H3K27 trimethylation, all triggered by CAG repeats and perhaps by the recognition of *HTT* polyA sites.

Overall, these results suggest that histone-modifying enzymes and chromatin factors are not acting independently but are part of a molecular complex where transcription regulation is possible through the orchestrated action of active and repressive histones. In particular, *HTT* alleles containing CAG repeats show reduced levels of histone acetylation and high levels of repressive histone modifications including H3K9me2/3 and H3K27me3.

Yet, whether histone modifications play a direct role in controlling Pol II recruitment, or they are a consequence of transcription, is still unclear. Although CAG expansion induces repressive marks, it is still elusive which events can trigger the formation of heterochromatin in HD. It is possible that a direct recruitment of epigenetic repressors by the CAG expansion DNA and the formation of secondary structures such as double-stranded RNA hairpins and RNA/DNA hybrids exists.

Chapter 4 | Double-stranded RNA are formed at triplet repeat expansions and contribute to transcriptional dysregulation in HD

4.1 R-loops formed on *HTT* gene are reduced in HD cells and do not interfere with efficient gene expression

The high GC-content of *HTT* RNA transcribed from CAG repeats is susceptible to the formation of greatly stable structures, such as RNA/DNA hybrids (R-loops), as demonstrated *in vitro* (Reddy *et al.*, 2011). R-loops are three-stranded nucleic acid structures, which occur when RNA hybridises to a complementary DNA strand (Thomas *et al.*, 1976). R-loops were proposed to be involved in the pathology of nucleotide expansion disorders, since they are induced following *in vitro* transcription of trinucleotide sequences in bacteria and human cells, promoting repeat instability (Grabczyk *et al.*, 2007; Lin *et al.*, 2010). Moreover, it was shown that R-loops are formed over both GAA and CGG repeats in cells from Friedreich ataxia (FRDA) and Fragile X syndrome (FXS) patients (Groh, Lufino, *et al.*, 2014). R-loops have also been detected at disease-associated CTG·CAG tracts on reporter constructs affecting instability of DNA trinucleotide repeats (Lin *et al.*, 2010).

Based on evidence that R-loops are involved in the pathology of trinucleotide expansion disorders (Lin *et al.*, 2010; Groh, Lufino, *et al.*, 2014), we hypothesized that their formation would be sensitive to DNA supercoiling and repeat length in Huntington's disease. Extended stretches of RNA/DNA hybrids can form at GC-rich regions (Lin *et al.*, 2010) and *HTT* has a high GC-content, especially in the region flanking the CAG repeats. To test if expanded CAG repeats can trigger formation of R-loops, lymphoblast cells from healthy individuals, adult and juvenile HD patients were used as cellular models, and R-loop formation on *HTT* gene was investigated using DNA/RNA immunoprecipitation (DIP) method. DIP relies on the immunoprecipitation of non-cross-linked purified RNA/DNA hybrids on any genomic location in human cells (Skourti-

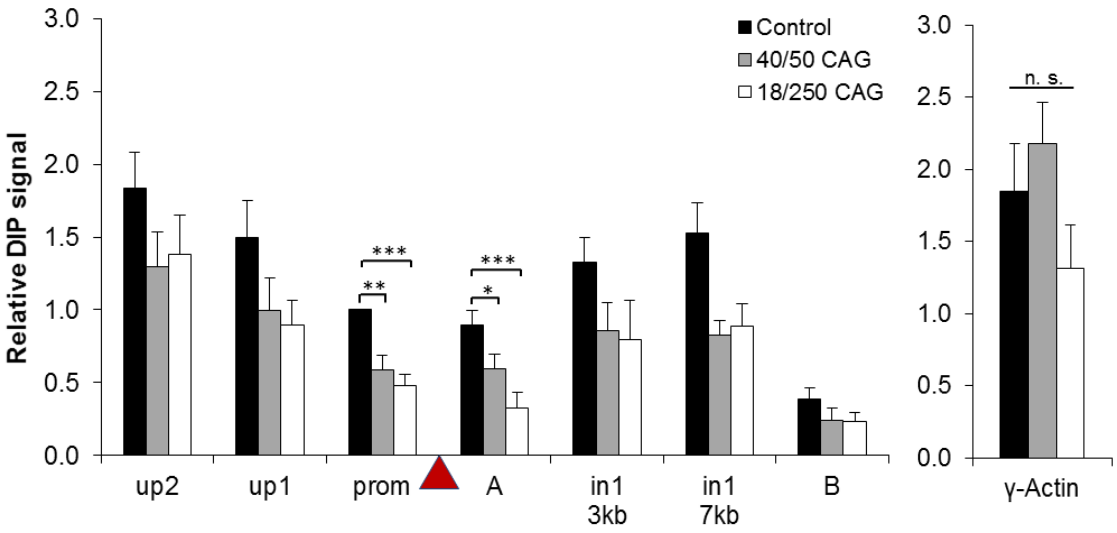
Stathaki *et al.*, 2011) using the monoclonal mouse S9.6 antibody, which efficiently detects R-loops (Boguslawski *et al.*, 1986).

As demonstrated in the Figure 4.1, R-loops were detected on *γ-Actin* and *HTT* genes in control cells. High levels of R-loops were seen upstream of the *HTT* transcription start site (TSS) (amplicons up2, up1) and this could be associated with *HTTAS* transcription. Surprisingly, the R-loop levels were significantly reduced in HD patient cells as compared to control cells in a CAG repeat length dependent manner (Figure 4.1A). The reduced levels of R-loops could be explained by the binding of various RNA-binding proteins, which interact with CAG repeats, preventing R-loop formation. Moreover, the packaging of *HTT* RNA by heterogeneous nuclear ribonucleoproteins (hnRNPs) may prevent a physical association with DNA during transcription, resulting in reduced R-loop levels. In addition, the preference of CAG repeats in forming alternative secondary structures, such as hairpins may prevent R-loop formation.

To explore the contribution of R-loops to *HTT* gene regulation, we used camptothecin (CPT), a drug which specifically inhibits DNA Topoisomerase I (Top1) (Pommier, 2006; Capranico *et al.*, 2010). Top1 relieves transcription-generated DNA supercoils and loss of Top1 induces R-loop formation (El Hage *et al.*, 2010; Wang, 2002; Tuduri *et al.*, 2009). Following CPT treatment, DIP analysis showed an increase of R-loop levels, particularly upstream of the CAG expansion in control cells, whereas the R-loop levels in HD cells in the *HTT* gene remained unchanged (Figure 4.1B). Our results demonstrated that CPT treatment is not specific to expanded repeats. The results in control cells are in line with a view that Top1 activates gene transcription and camptothecin has an inhibitory effect on Top1 activity (Capranico *et al.*, 2010). Taken together, these results showed that although R-loops are enriched and inducible in control

cells over expanded CAG repeat region, there is an alternative factor or mechanism that is preventing their formation and further induction in HD patient cells.

A



B

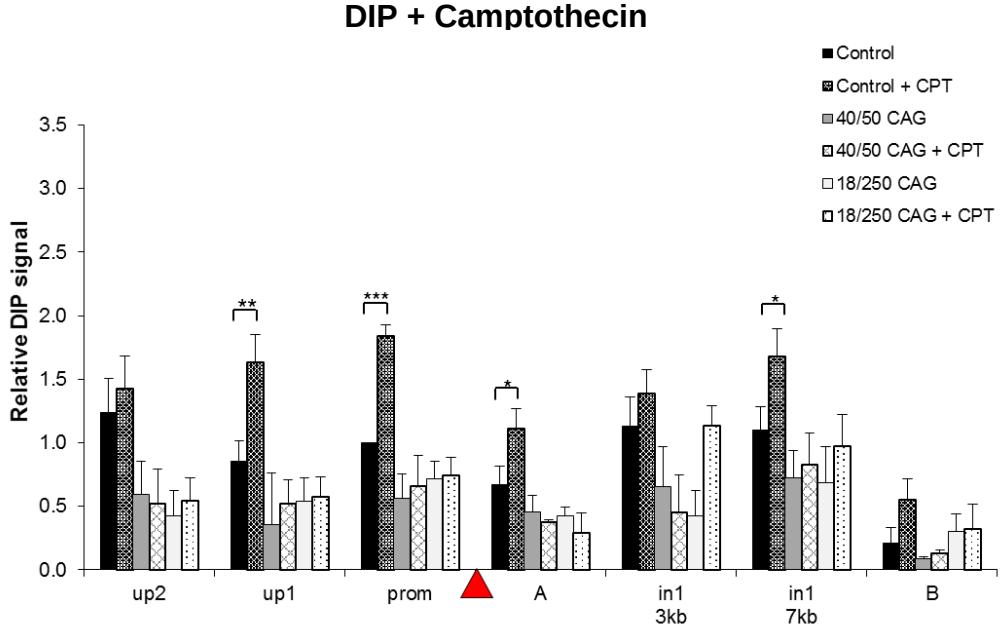


Figure 4.1 R-loops formed on the *HTT* gene are reduced in HD cells.

A. DNA/RNA hybrid immunoprecipitation (DIP) analysis show that the levels of DNA/RNA hybrids are reduced in the human *HTT* gene in HD lymphoblast cells compared to control lymphoblast cells (n=4). The values are represented relative to amplicon prom in control cells.

B. DNA/RNA hybrids formed on the *HTT* gene are restored in control lymphoblast cells treated with 10 μM camptothecin (CPT) for 6 hours but not in HD lymphoblast cells after treatment (n=6). The values are represented relative to amplicon prom in control cells.

Asterisks (*) indicate statistical significance (* $p<0.05$; ** $p<0.01$; *** $p<0.001$), based on an unpaired, two-tailed distribution Student's t-test.

4.2 dsRNA are enriched at transcribed expanded CAG repeats of *HTT* gene and correlate with expansion size in human HD cells

Previously it has been demonstrated that expanded CAG repeats may form hairpin structures *in vitro* (de Mezer et al., 2011). The expanded transcript is retained in the nucleus of human HD fibroblasts and is colocalized with the double-stranded RNA binding protein (dsRBP) MBNL1, suggesting that CAG repeats in the *HTT* mRNA adopt ds-like RNA conformations *in vivo* (de Mezer et al., 2011). The predisposition of CAG repeats to fold into long hairpin structures triggers toxicity of the pathogenic RNA due to unusual interactions with dsRBPs, leading then to abnormal function in RNA metabolism. Thus, the structural features of the CAG repeats on the *HTT* transcript, which can be recognized as damage signals by DNA repair systems, will need further attention.

In order to test the ability of CAG repeats to form hairpins *in vivo*, the dsRNA-specific J2 antibody, which binds RNA duplexes ≥40 base pairs (Schonborn *et al.*, 1991; Bonin *et al.*, 2000) was used to immunoprecipitate dsRNA (Figure 4.2A). The levels of dsRNA at the 5'end of the *HTT* gene were low in control cells. Interestingly, we observed an enrichment of dsRNA formed at the expanded repeats in HD patient cells (amplicons up1-A, Figure 4.2A), while the region downstream remained unchanged (amplicons in1

3kb-B). In line with *in vitro* work (de Mezer *et al.*, 2011), we detected an enrichment of dsRNA in HD cells, particularly at the 5'end of the *HTT* gene, at expanded CAG repeats (Figure 4.2A). Moreover, dsRNA enrichment correlates with expansion size, the larger the expansion, the more dsRNA is formed, independently of the cell type. The data suggests that CAG dsRNA hairpins may interfere with R-loops formed *in vivo*. It has been demonstrated that efficient folding of nascent RNA into secondary structures prevents RNA/DNA hybrid formation and genome instability in *S. cerevisiae* (Chen *et al.*, 2016). Therefore, accumulation of dsRNA can explain the reduced R-loop levels observed in HD.

To confirm the specificity of the dsRNA J2 antibody, dsRNA V1 RNase, which cleaves base-paired nucleotides in double-stranded regions of the RNA (Ehresmann *et al.*, 1987), and single-stranded-specific S1 nuclease treatments were carried out for 2 hours at 37°C after the dsRNA immunoprecipitation. Isolated RNA was analysed by RT-qPCR and positive signals were detected over prom and A regions of the *HTT* gene in HD cells (Figure 4.2B). J2-immunoprecipitated dsRNA was highly sensitive to dsRNA-specific V1 RNase as the signal was strongly reduced over the 5' region of the *HTT* gene but not strand-specific S1 nuclease-dependent, confirming dsRNA presence over the expanded CAG repeats (Figure 4.2).

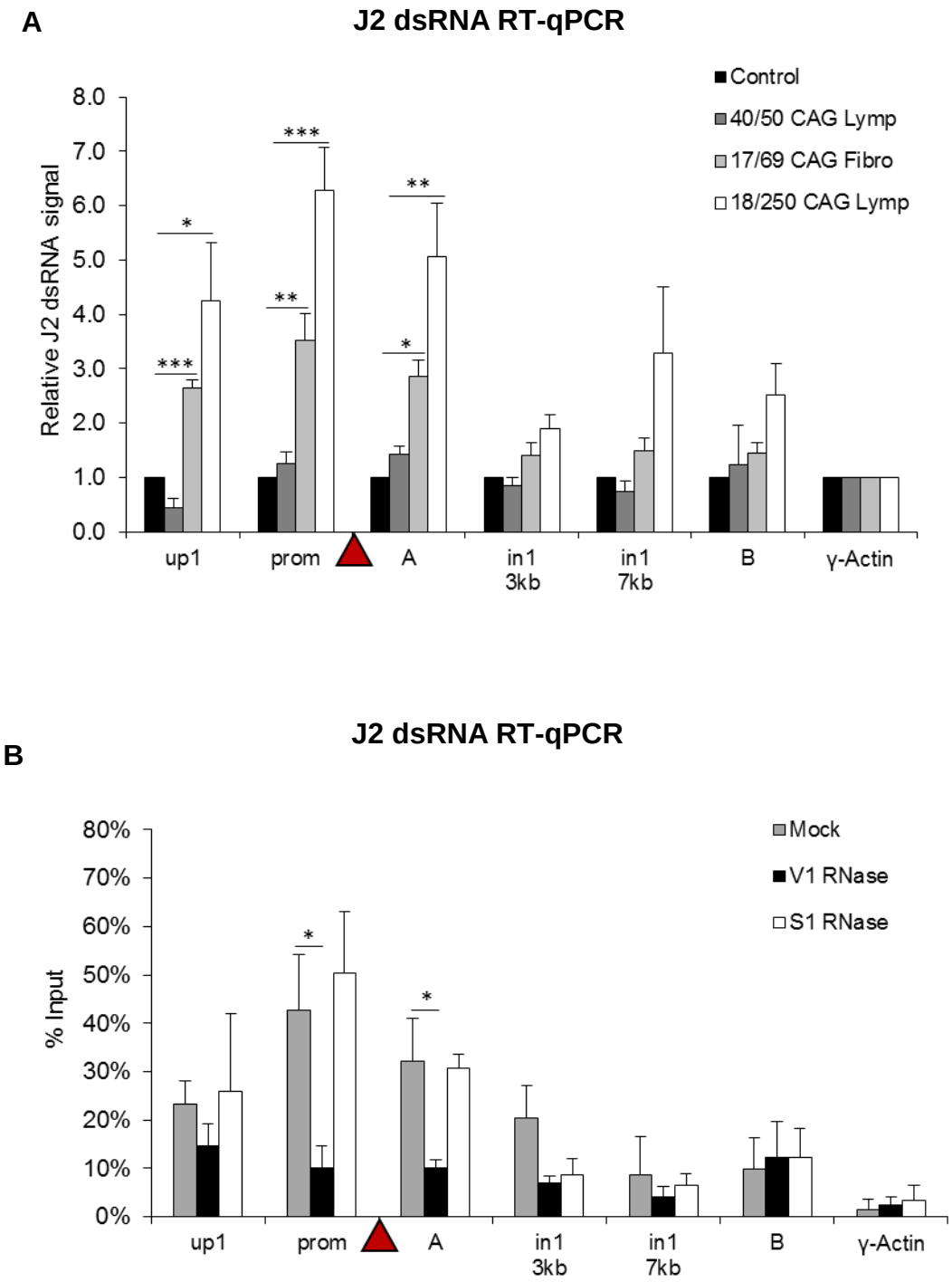


Figure 4.2 dsRNA are enriched at transcribed expanded CAG repeats of *HTT* gene.

A. qRT-PCR analysis from J2 pull-down, which recognizes dsRNA, show that dsRNA levels are increased on *HTT* gene over CAG repeats in HD lymphoblast and fibroblast cells (n=3 for fibroblast cells, n=4 for lymphoblast cells). Values are relative to untreated control cells, and normalized to *γ-Actin* probe.

B. dsRNA levels determined by RT–qPCR are reduced on *HTT* gene after V1 RNase treatment in HD fibroblast cells (17/69 CAG repeats) (n=3). Samples were untreated (Mock, grey bars), or treated with V1 RNase (black bars) that reduced dsRNA levels or S1 nuclease (white bars).

Asterisks (*) indicate statistical significance (* $p<0.05$; ** $p<0.01$; *** $p<0.001$), based on an unpaired, two-tailed distribution Student's t-test.

4.3 Investigation of LNA treatments for targeting of Huntingtin

4.3.1 RNA expression following treatments to reduce huntingtin expression

In order to investigate the contribution of the *HTT* dsRNA in the transcriptional repression of *HTT*, single stranded antisense oligonucleotides (ASOs) particularly locked nucleic acids (LNAs) were used. The LNAs used in this study target the CAG repeat expansion and were designed by Dr Miguel Varela (Wood laboratory) with the aim to reduce *HTT* expression as mixmers and gapmers.

In a mixmer, LNA and DNA nucleosides are intermixed throughout the sequence of the oligonucleotide and may inhibit translation by blocking the binding of the ribosomal subunits to the mRNA. Alternatively mixmers are able to bind close to the 5' end of the pre-mRNA preventing 5'-cap formation and thus inhibiting translation. A mixmer cannot induce cleavage of the target RNA but acts as a steric block to mediate the phenotype without destroying the RNA (Hagedorn *et al.*, 2017).

On the other hand, a gapmer contain LNA bases flanking a central DNA region and induce silencing via recruitment of RNase H to cleave the mRNA of a DNA-RNA duplex (Hagedorn et al., 2017). Generally, gapmers are more potent at inhibiting mRNA expression and protein, as the central DNA recruits the RNA-cleaving enzyme RNase H when the gapmer hybridizes to the mRNA (Braasch and Corey, 2001; Cerritelli and Crouch, 2009). Gapmers can target both exons and introns likewise efficiently and the activity occurs in both the cytoplasm and the nucleus, although mainly in the nucleus (Cerritelli and Crouch, 2009; Hagedorn *et al.*, 2017).

HD fibroblasts have been used in many studies and provide a validated model system to study the efficacy of ASOs in *HTT* gene expression (Fiszer *et al.*, 2011). The

methods to silence *HTT* were investigated previously in our lab by Dr Aisling O'Loughlin for their efficacy both at the message and protein level and also for the duration of silencing they can exert in HD fibroblasts (Figure 4.3A). The LNAs were transfected at doses of 12.5, 25, 50 and 100 nM into HD patient fibroblasts with Lipofectamine RNAiMAX and the *HTT* mRNA expression was measured after 48 hours (Figure 4.3A). To estimate the duration of silencing effects, LNAs were transfected at 100 nM and *HTT* mRNA expression was analysed after 24, 48, 96 and 186 hours (Figure 4.3B). The silencing of *HTT* mRNA expression was slightly more pronounced at the highest dose of 100 nM, with ~50% reduction. Furthermore, the efficacy of LNAs was comparable at all time points tested, until 96 hours, although there was some recovery at 186 hours (Figure 4.3B).

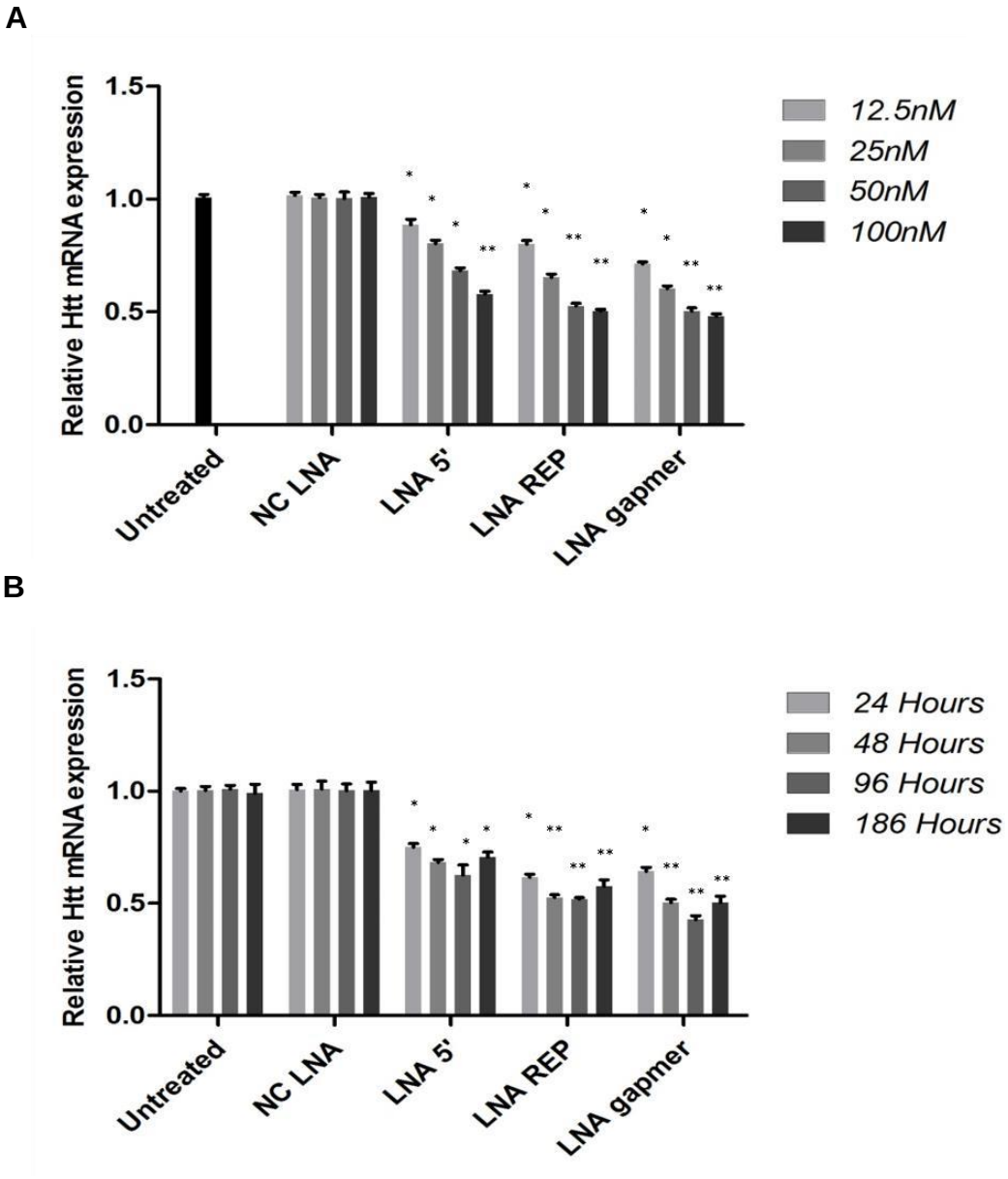


Figure 4.3 Investigation of antisense oligonucleotide treatments for *HTT* knockdown.

A. *HTT* mRNA levels assessed by RT-qPCR 48 hours after transfection with 12.5, 25, 50 and 100 nM LNA (n=4). HD patient-derived fibroblasts were left untreated, transfected with a negative control (NC) LNA or transfected with LNAs predicted to reduce *HTT* expression, one targeting an area upstream of the CAG triplet repeat (LNA-5'), and two targeting the CAG expansion (LNA-REP and gapmer). Values are normalized to housekeeping *β-Actin* gene. This experiment was performed by Dr Aisling O'Loughlin (Wood laboratory).

B. *HTT* mRNA levels assessed by RT-qPCR after transfection with 100 nM of LNA following different lengths of time post-transfection (n=4). HD Patient-derived fibroblasts were left untreated, transfected with a negative control (NC) LNA or transfected with LNAs predicted to reduce *HTT* expression, one targeting an area upstream of the CAG triplet repeat (LNA-5'), and two targeting the CAG expansion

(LNA-REP and gapmer). This experiment was performed by Dr Aisling O'Loughlin (Wood laboratory).

Asterisks (*) indicate statistical significance (* $p<0.05$; ** $p<0.01$; *** $p<0.001$), based on an unpaired, two-tailed distribution Student's t-test.

4.3.2 Protein expression following treatments to reduce huntingtin expression

In order to validate the previous results of *HTT* mRNA reduction following treatment with LNAs, the HTT protein levels of patient HD fibroblast cells, treated under the same conditions, were measured by ELISA. As before, different doses of LNAs were tested - 12.5, 25, 50 and 100 nM - and protein was harvested from cells 48 hours after transfection with Lipofectamine RNAiMAX. The LNA causing the most effective silencing was LNA-5', where HTT expression in cells treated with 100 nM was reduced to 41.9%. However, LNA-5' targets an area upstream of the CAG triplet repeat, and the aim of the ASO treatments in this context is to study *HTT* gene regulation and the effects of the CAG expansion. Thus, the HTT expression in cells treated with 100 nM of LNA mixmer and gapmer decreased to approximately 40%, revealing that the LNAs targeting the CAG repeats are the most promising candidates (Figure 4.4A). The results from the HTT ELISA assay are in line with the dose response studies measured by qPCR, confirming knockdown of HTT at a protein level as well at the mRNA level, confirming that both the LNAs mixmer and gapmer are allele selective and silence preferentially the mutant *HTT* allele.

Lastly, the toxicity of the LNAs was measured by a Cell Proliferation MTS assay (Figure 4.4B). The MTS is a colorimetric method for determining the toxicity of cellular treatments. The LNA treatments were slightly toxic at higher doses, while no measurable toxicity was detected at lower doses (Figure 4.4B). Overall, despite the slight toxicity, the LNA treatments used in this study are good candidates to further investigate *HTT* gene regulation in HD.

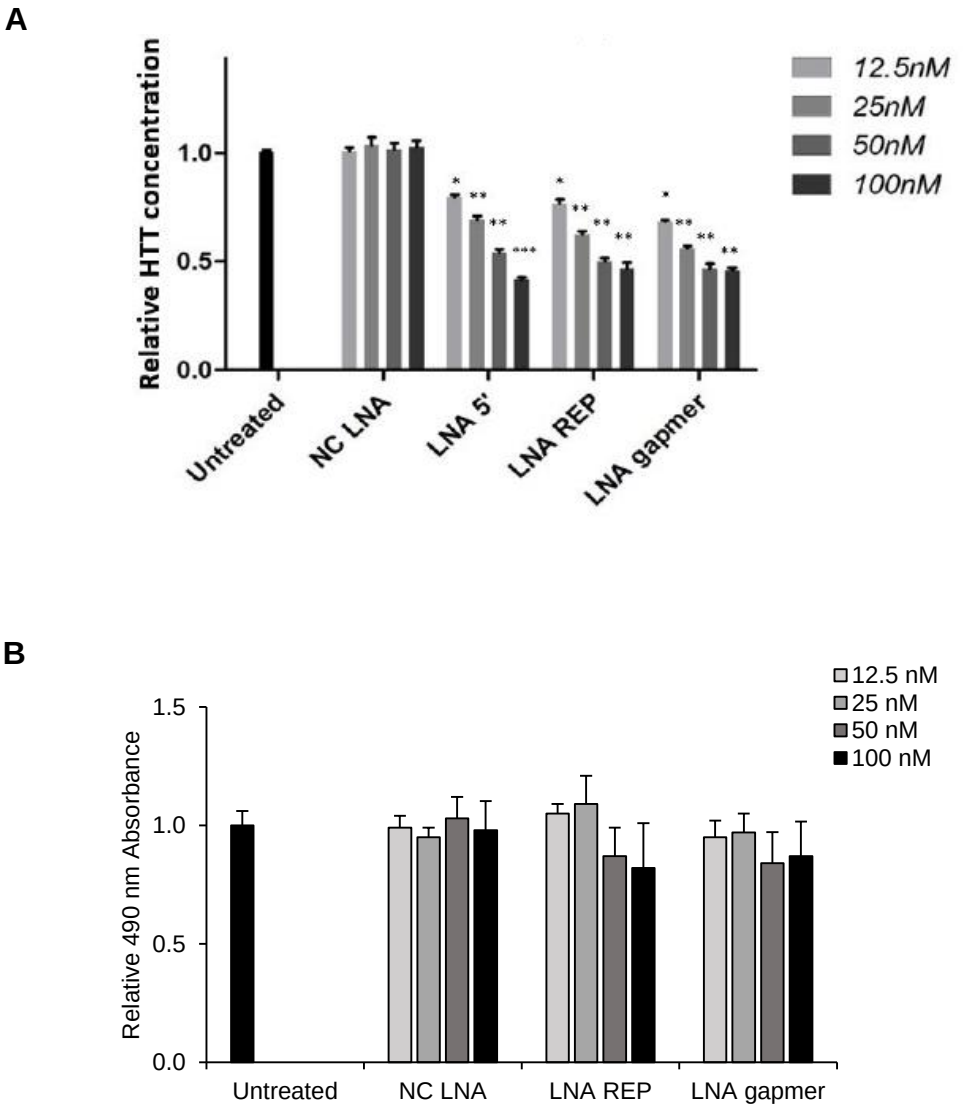


Figure 4.4 Quantification of HTT protein following LNA treatments for *HTT* knockdown.

A. HTT protein levels measured by Elisa 48 hours post-treatment (n=4). HD patient-derived fibroblasts were left untreated, transfected with a negative control (NC) or transfected with antisense oligonucleotides. This experiment was performed by Dr Aisling O'Loughlin (Wood laboratory).

B. Investigation of cell viability following LNA treatments to reduce HTT knockdown (n=3). HD fibroblasts were left untreated, transfected with the indicated doses of a negative control (NC) and with antisense oligonucleotides, predicted to reduce *HTT* expression in different doses. Cell viability was assessed using an MTS assay 48 hours post-treatment.

Asterisks (*) indicate statistical significance (* $p<0.05$; ** $p<0.01$; *** $p<0.001$), based on an unpaired, two-tailed distribution Student's t-test.

4.3.3 Effects of LNA treatments on *HTT* dsRNA expansion structures

To further investigate the effects of LNA treatments in silencing the dsRNA formed by CAG repeats, it was possible to take advantage of the LNAs designed by Dr Miguel Varela (Wood laboratory). The LNAs specifically target the CAG repeats and were tested previously for their toxicity and effectiveness in reducing the huntingtin expression. As in the previous tests the dose of 100 nM appeared to be the most effective in reducing the *HTT* gene expression, both LNAs gapmers and mixmers were transfected at 100 nM for 48 hours into HD fibroblasts with Lipofectamine RNAiMAX. Following dsRNA immunoprecipitation, *HTT* mRNA expression was analysed by RT-qPCR and normalised to *γ-Actin*. Notably, both LNAs gapmers and mixmers at 100 nM were able to reduce *HTT* dsRNA to ~50% in the region surrounding the repeats (amplicons prom – A) in HD cells (Figure 4.5).

The efficacy of LNAs was previously evaluated both at the mRNA and protein level and viability assays were performed to estimate a good safety profile. However, in this study, the antisense oligonucleotides will be used as a tool to investigate the involvement of dsRNA in transcriptional repression and heterochromatin formation.

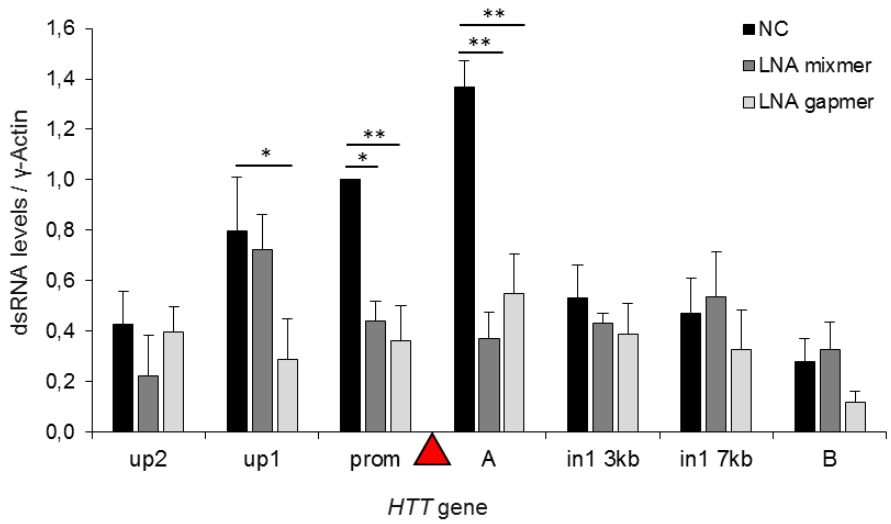


Figure 4.5 dsRNA levels are reduced in HD cells following LNA treatments.
HTT dsRNA levels measured by RT-qPCR 48 hours post-transfection with 100 nM LNA (n=3). HD fibroblasts show reduced levels of dsRNA after transfection with LNAs targeting the CAG expansion (LNA mixmer and LNA gapmer); NC is a negative control. Values are relative to amplicon prom probe, and normalized to housekeeping *γ-Actin* gene. Asterisks (*) indicate statistical significance (* $p<0.05$; ** $p<0.01$; *** $p<0.001$), based on an unpaired, two-tailed distribution Student's t-test.

4.4 Involvement of dsRNA in transcriptional repression and heterochromatin formation in HD

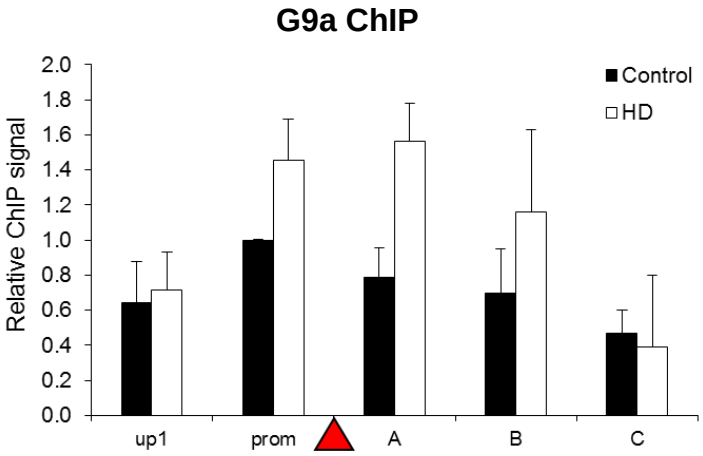
4.4.1 H3K9me2 is important for the formation of *HTT* dsRNA in HD

To further understand the dynamics of heterochromatin formation at expanded CAG repeats, the interaction between the histone methyltransferase G9a and *HTT* gene was investigated. G9a is the enzyme responsible for dimethylating lysine 9 of histone 3 (H3K9me2), a hallmark of gene silencing. H3K9me2 function is dependent on the recruitment of transcriptional repressors that impede transcriptional and gene activation (Scheer and Zaph, 2017). ChIP analysis demonstrated that G9a levels are increased in the

region surrounding the CAG expansion of the *HTT* gene in HD cells (Figure 4.6A). This result suggests that the CAG repeat expansion may recruit G9a enzyme, can be recruited by the dsRNA, leading to the formation of H3K9me2 histone modification.

To further investigate the role of dsRNA in heterochromatin formation, the G9a methyltransferase inhibitor UNC0638 was used. UNC0638 has previously shown to inhibit the enzymatic function of G9a, reducing the levels of H3K9me2 (Vedadi *et al.*, 2011; Ling *et al.*, 2012). After treatment of control and HD cells with UNC0638 for 72 hours, the levels of H3K9me2 were significantly reduced (Figure 3.9B). Control and HD cells, treated with UNC0638 methyltransferase inhibitor, were used for J2 dsRNA immunoprecipitation. Following UNC0638 treatment, and decreasing H3K9me2 repressive mark, the levels of dsRNA over CAG expansion significantly decreased, while no changes were detected in control cells (Figure 4.6B). This result suggests that H3K9me2 repressive mark is associated with dsRNA formation and reduced transcription in HD.

A



B

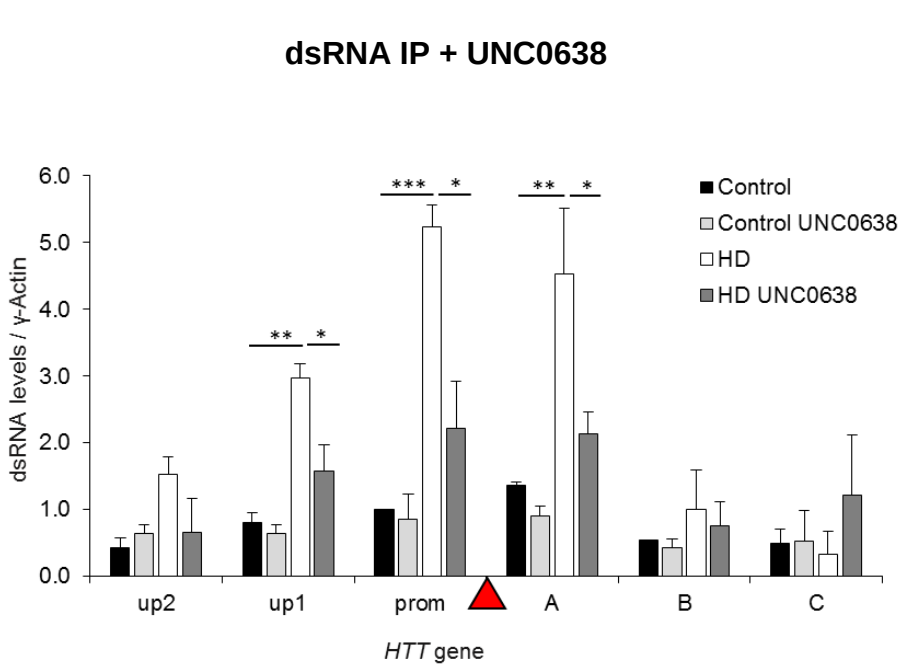


Figure 4.6 dsRNA are affected by changes in H3K9me2 repressive mark.

A. H3K9 methyltransferase G9a is enriched at expanded CAG repeats in HD fibroblast cells. G9a ChIP on *HTT* gene in control and HD cells. G9a levels are normalised relative to amplicon prom in control cells.

B. dsRNA levels as determined by qRT-PCR from J2 pull-down on *HTT* gene in control and HD fibroblast cells, after treatment with 250 nM UNC0638 for 72 hours (n=3). dsRNA levels are reduced in HD cells after UNC0638 treatment. Values are relative to prom amplicon and normalised to the *γ-Actin* probe.

Asterisks (*) indicate statistical significance (* $p<0.05$; ** $p<0.01$; *** $p<0.001$), based on an unpaired, two-tailed distribution Student's t-test.

4.4.2 Formation of the hairpin dsRNA correlates with H3K9me2 and *HTT* gene repression

Next, the correlation between dsRNA, RNA Polymerase II and H3K9me2 repressive marks was investigated. Interestingly, an increase in the level of *HTT* dsRNA coincide with the decreased levels of Pol II and with a significant increase of the H3K9me2 repressive mark over the 5' end of the *HTT* gene in HD cells (Figures 3.4, 3.8A and 4.2A, respectively). This suggests that *HTT* dsRNA formed over expanded CAG repeats can be functionally associated with the repressive heterochromatin and the RNA secondary structures may be involved in mediating transcriptional repression. When HD cells are treated with dsRNA-specific nuclease V, this leads to reduction of dsRNA levels in HD cells (Figure 4.2B). Furthermore, after RNase V treatment, the levels of Pol II in HD cells increase, indicating a re-activation of transcription, suggesting that dsRNA interfere with Pol II transcription of *HTT* gene, while no effect was seen in control cells (Figure 4.7A). Interestingly, after silencing of *HTT* dsRNA expression, the levels of H3K9me2 decrease in HD cells, which can be associated with active Pol II mediated transcription and derepression of the gene (Figure 4.7B). Overall, these results indicate that formation of dsRNA at expanded CAG repeats is associated with heterochromatin marks, suggesting that dsRNA can impair *HTT* gene expression. The crosstalk between dsRNA, heterochromatin formation and transcriptional repression of *HTT* gene can lead to a novel functional interplay and a role of dsRNA in *HTT* gene regulation in HD cells.

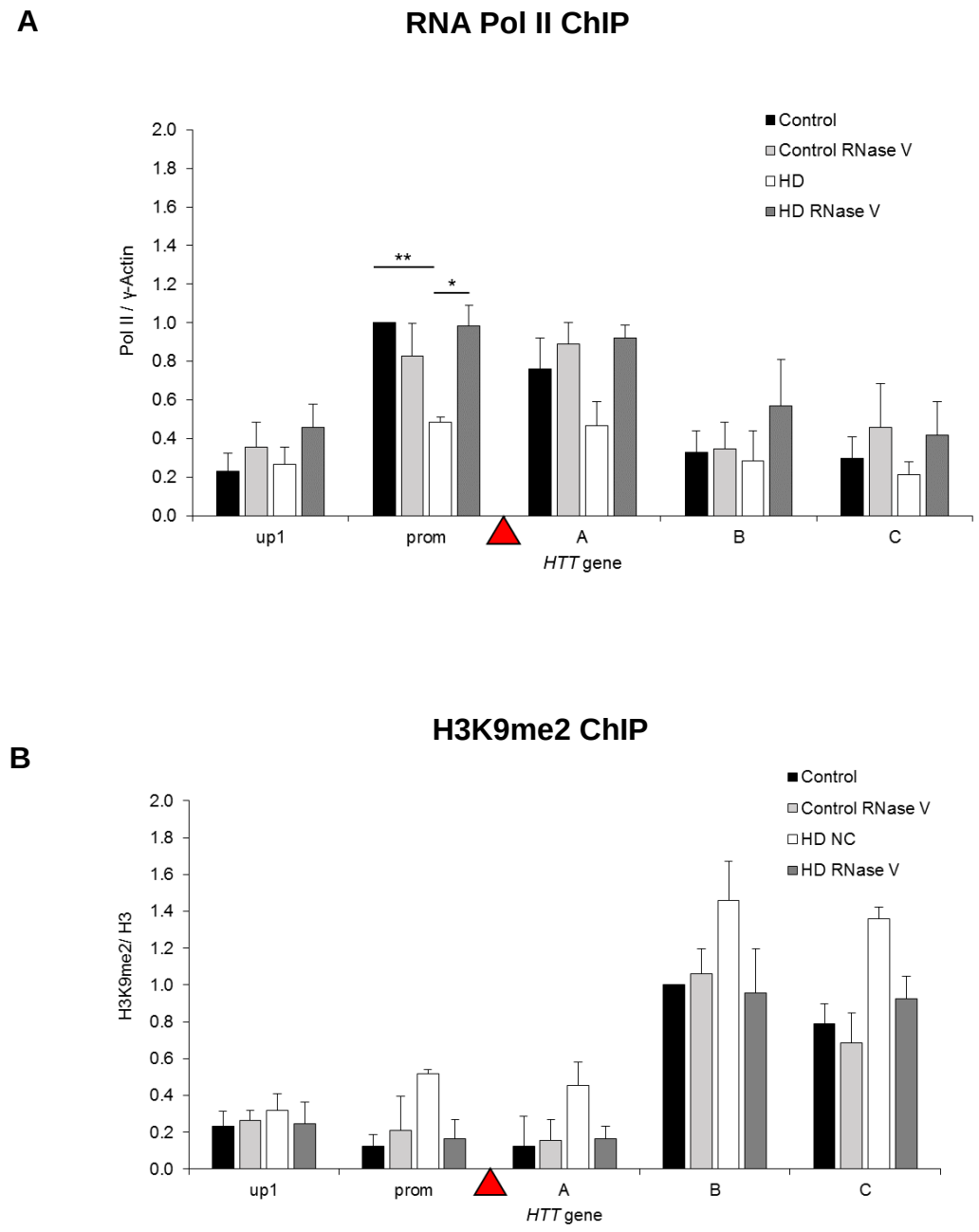


Figure 4.7 dsRNA can interfere with the transcriptional regulation of the *HTT* gene.

A. ChIP analysis show that RNA polymerase II, measured by RT-qPCR, is increased on the *HTT* gene in HD fibroblast cells, after treatment with V1 RNase, which is an enzyme that specifically degrades dsRNA (n=3). Control and HD fibroblasts are untreated, or treated with V1 RNase. Values are relative to amplicon prom probe, and normalized to housekeeping *γ-Actin* gene.

B. ChIP analysis of the H3K9me2 levels on *HTT* gene, measured by RT-qPCR (n=3). Control and HD fibroblasts were untreated, or treated with V1 RNase. Values are normalised to levels of total histone H3 and relative to amplicon B probe.

Asterisks (*) indicate statistical significance (* $p<0.05$; ** $p<0.01$; *** $p<0.001$), based on an unpaired, two-tailed distribution Student's t-test.

4.5 Summary

Overall, the data shown in this chapter demonstrated that dsRNA are formed at endogenous expanded CAG repeats in the *HTT* gene, leading to transcriptional repression of the *HTT* gene in HD cells.

The age of onset of HD is dependent on the CAG repeat within *HTT*, with longer expansions causing earlier onset (Langbehn *et al.*, 2010) and dsRNA levels are directly associated with the CAG-repeat length. Moreover, dsRNA can form independently of cell type, can be resolved by Ribonuclease V1 and reduced by antisense oligonucleotides (ASOs) designed to specifically target the CAG repeats.

Interestingly, the levels of the repressive histone mark H3K9me2 did affect dsRNA formation, as decreasing the levels of the H3K9me2 with UNC0638, reduced dsRNA formation in HD cells. Importantly, depletion of dsRNA led to decreased H3K9me2 levels and restored transcription of the *HTT* gene in HD cells, but not in control cells.

These results suggest that dsRNA formed by CAG repeats in the *HTT* gene promote heterochromatin formation and lead to *HTT* transcriptional repression. Although further investigations on the epigenetic landscape are required, this suggests a functional interplay between dsRNA and H3K9me2 with an impact on transcriptional repression.

Additionally, the identification of dsRNA-binding proteins would allow a better understanding on the pathomechanism of HD. It is possible that H3K9me2 function is dependent on the recruitment of epigenetic repressors by the *HTT* dsRNA that impede transcriptional and gene activation. Yet, whether histone modifications play a direct role in controlling Pol II or they are a consequence of transcription is still unclear. Thus, proteomic studies would be necessary to investigate the dsRNA interactors and their function in health and disease. In order to better elucidate on the pathomechanism of HD,

which can be extended to other trinucleotide expansion disorders, it will be essential to characterise the proteins that interact with dsRNA.

4.6 Investigation of the dsRNA-binding proteins in human cells

DsRNA is a key player in numerous Biological activities in human cells (Svoboda and Di Cara, 2006), and serve as a binding site for a variety of proteins (Wang *et al.*, 2015). The molecular mechanism underlying the interaction between dsRNA-binding proteins (dsRBPs) and dsRNA is still uncertain, particularly because dsRBP have not been well characterised in human cells (Wang *et al.*, 2015).

Many studies have identified factors involved in dsRNA Biology, mainly proteins regulating RNA metabolism, controlling protein synthesis and affecting RNA stability, such as the Argonaute proteins (Hutvagner and Simard, 2008). Other descriptions suggest that dsRBP are involved in anti-viral responses and regulation of translational initiation in response to stress, such as the PKR protein (Burgess and Mohr, 2015; Lemaire et al., 2008), which binds mutant *HTT* transcripts (Peel et al., 2001) and the interacting protein p74 (Coolidge and Patton, 2000). The presence of dsRNA is also linked to the loss of particular proteins, including Dicer (White et al., 2014), implicated in RNAi-mediated gene silencing (Saunders and Barber, 2003) and plays critical roles in genome regulation and surveillance (Johanson et al., 2013). Staufen and MBNL1 proteins function as disease-modifiers in the neuromuscular disorder DM1 and regulate alternative splicing events (Galka-Marciniak *et al.*, 2012; Bondy-Chorney, Parks, Ravel-Chapuis, Klinck, *et al.*, 2016).

Other studies investigating human dsRBPs, ADAD2, TRBP, Staufen 1 and ADAR1, have shown they play essential roles in development, translation, RNA editing and stability (Saunders and Barber, 2003). Subsequently, another study showed that

regardless of their conservative dsRNA binding domains, the dsRBPs exhibit different dynamic properties when in contact with different RNA substrates, which may affect their cellular function (Wang *et al.*, 2015).

These studies provide mechanistic insights and underscore the diversity of vital Biological functions performed by dsRBPs in the regulation of gene expression and RNA-related processes, however the majority of published literature is based on single-protein and single-gene studies and it is not clear whether the effects described are mediated by direct interaction of the protein factors with dsRNA.

To address these important questions towards a better understanding of the dsRNA Biology, the aim of the following subchapter was to identify proteins directly interacting with dsRNA in human cells.

4.6.1 The dsRNA IP methodology

Interactions between proteins and nucleic acids are complex and greatly regulated and with the helpful insights of Prof Shabaz Mohammed and Dr Alfredo Castello (Department of Biochemistry, University of Oxford), we developed a strategy to purify endogenous dsRNA and their associated proteins. The method to selectively recognise dsRNA-binding proteins relies on the usage of the dsRNA-specific J2 antibody (Schonborn *et al.*, 1991; Bonin *et al.*, 2000). The J2 antibody is well characterised and used by many independent groups to investigate dsRNA-related processes (Schonborn *et al.*, 1991; Mueller *et al.*, 2010; Bonin *et al.*, 2000; White *et al.*, 2014; Burgess and Mohr, 2015).

The dsRBP method was designed to isolate dsRNA and associated proteins from human fibroblast cells, and optimisation developments were focused on dsRBP immunoprecipitation, and reduced non-specific protein binding. As either formaldehyde

or UV-induced crosslinking prior to immunoprecipitation may artificially affect dsRNA levels or impede accessibility of J2 to endogenous dsRNA, dsRBP IP method was carried out using non-crosslinked cells. To do efficient extraction of dsRNA and proteins, a variety of buffers of different stringencies, from very mild to stringent conditions, were tested, however the fact that it only identified few dsRBP or reduced the dsRBP IP efficiency with enrichment of non-specific-binding, led us to choose an intermediate stringent 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 protease inhibitors. This buffer reduced non-specific binding and extracted dsRNA efficiently and was employed for all subsequent experiments.

To further reduce the pull-down of non-specific RNA-binding proteins, sonication of the lysates was maintained for 7 minutes, the sonication time previously used for dsRNA pull-down experiments. In addition, a benzonase treatment was employed after the dsRBP IP, which degrades all forms of nucleic acids without affecting proteins (Aygun *et al.*, 2008). Analysis of nucleic acids using agarose gel electrophoresis showed that benzonase treatment effectively disrupted RNA-protein interactions, removing the most abundant RNA species (28S and 18S rRNA) and genomic DNA starting at 3 U/ml (Figure 4.8A). For a more detailed protocol, see Chapter 2.

The optimised dsRBP method was carried out and immunoprecipitated samples were analysed by SDS-PAGE, silver staining gel, mass spectrometry and western blotting. Firstly, immunoprecipitated dsRNA-binding proteins were observed by SDS-PAGE and silver staining gel. As shown in Figure 4.8B, several dsRBP were detected in the IP sample, and the profile is different from the Control Input material, in the first lane. Importantly, the complex mixture in the IP profile of HD patient fibroblasts differs from the IP bands in the control lane. Moreover, these proteins were not present in the sample without the J2 dsRNA antibody (B). It seems that there is a unique pattern in the J2 IP

4.6.2 Mass spectrometry analysis of the dsRNA-binding proteins

After validation of the dsRBP Immunoprecipitation method, a proteomics strategy was developed in collaboration with Prof Shabaz Mohammed (Advanced Proteomics Facility, Department of Biochemistry, University of Oxford). This approach was performed on liquid chromatography tandem mass spectrometry (LC-MS/MS) with the goal to identify novel proteins interacting with dsRNA in HD human cells. SDS-containing protein samples were digested with trypsin according to the filter-aided sample preparation (FASP) procedure as described previously (Wisniewski, Zougman and Mann, 2009). The mass spectrometry-based proteomics was performed in both control and HD human samples after immunoprecipitation with J2 dsRNA antibody, for increased confidence of identified proteins. In total, were used three Biological replicates of the immunoprecipitated dsRBP from HD and control human samples.

The MS raw files were processed with MaxQuant software, version 1.5.0.0 (Cox and Mann, 2008) and analysed by Dr Ashling Holland (Wood laboratory) to calculate the relative enrichment for each identified protein in each Biological replicate. The raw files were searched against the human, *Homo sapiens*, UniProt KB FASTA database (20,219 proteins, August 2017) for MS/MS based peptide identification via Andromeda. Enzyme specificity was set to trypsin and Lys-C, with a minimum number of seven amino acids required for peptide identification. This method identified 123 proteins in total in all three independent Biological replicates of dsRBP IP. Perseus (version 1.5.5.3) was used for statistical and annotation enrichment analysis based on 3 independent Biological replicates, freely available open-source bioinformatics platforms were employed. Among the proteins identified, there was a strong correlation with the relative enrichment of dsRNA-binding proteins in Control and HD samples with a high reproducibility in the three Biological replicates (Pearson's correlation coefficients $r>0.7$) (Figure 4.9A).

Differentially expressed protein analysis was performed using the PANTHER (http://pantherdb.org/) to perform GO analysis. To determine which pathways play a potential role in our HD samples, all differentially expressed proteins were subjected to Kyoto Encyclopedia of Genes and Genomes (KEGG).

A

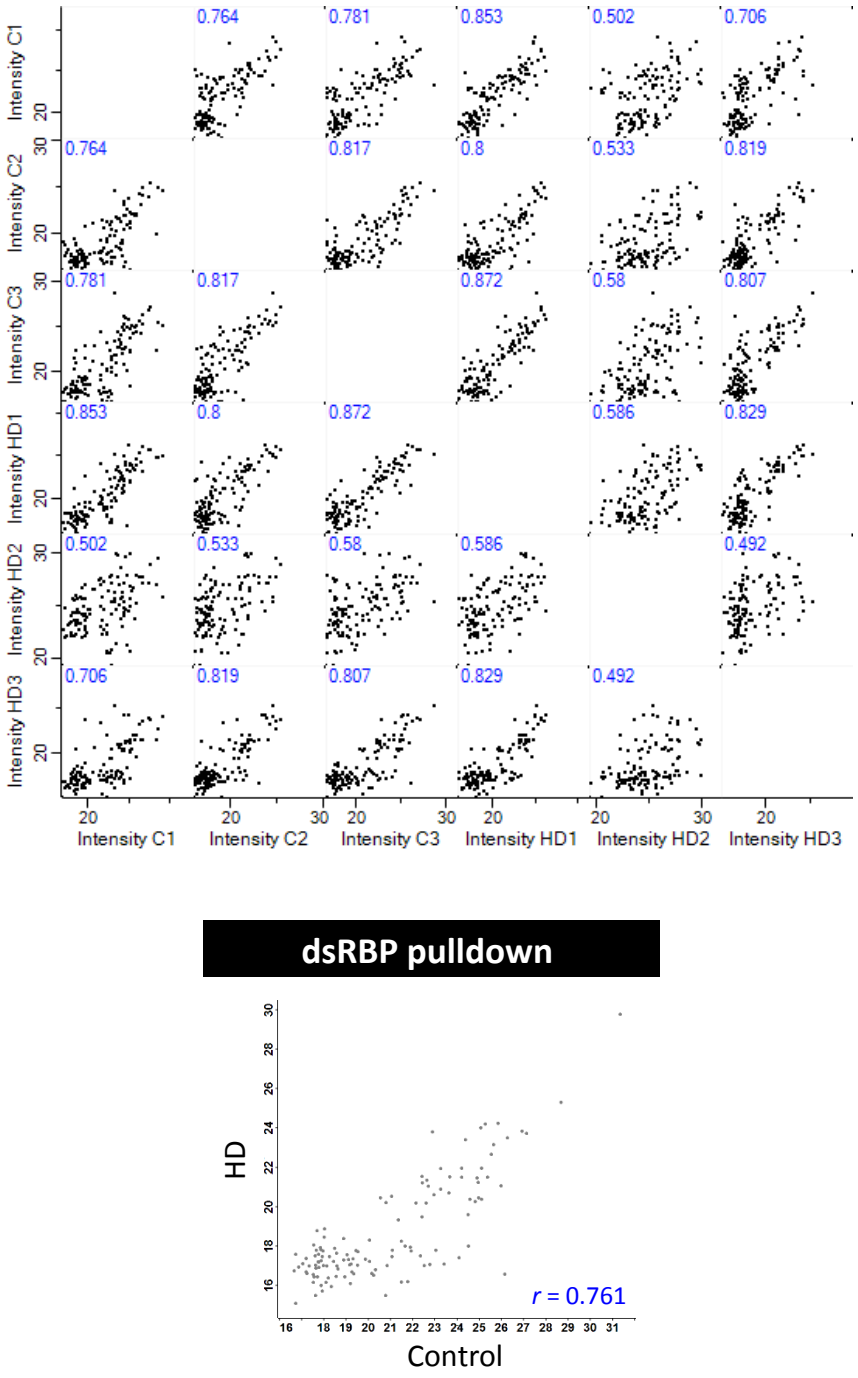
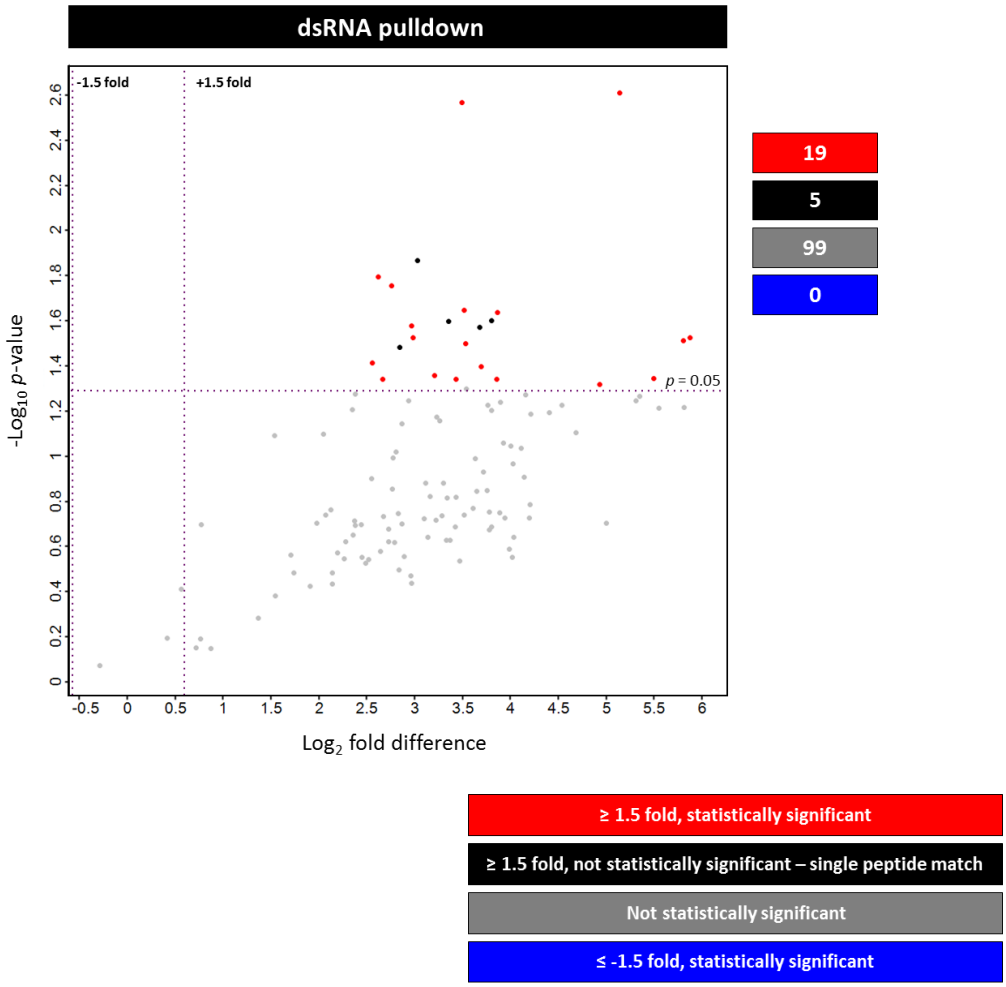


Figure 4.9 The dsRNA-binding proteins immunoprecipitation methodology shows a strong positive correlation.

Analysis of the Biological and technical reproducibility of the dsRBP IP method. The panel shows the correlation between log2 enrichment HD IP/Control IP of proteins quantified in three independent Biological replicates by mass spectrometry. *r* is Pearson's correlation coefficient.

For data handling, normalisation, statistics and annotation enrichment analysis, the freely available open-source bioinformatics platform Perseus (version 1.5.5.3) was employed. Proteins that were identified only by site modification, contaminants, found in the decoy database or quantified in only one of the replicates were excluded from protein analysis. For pairwise comparison of proteomes, a two-side *t*-test statistic was employed. All other statistical and bioinformatics data analyses were performed using the freely available software STRING (version 10.0), PANTHER, version 11.1 and Kyoto Encyclopedia of Genes and Genomes (KEGG), release 77.1. STRING (http:string-db.org/) was used to illustrate known and predicted protein-protein interactions. After employing the statistical data analysis, only 19 proteins were identified as significantly enriched in the dsRBP IP in HD (fold change≥1.5), with a *p*-value≤0.05 (Figure 4.10, red data dots). Data was displayed in confidence view using the high confidence (0.700) parameter for analysis. A representative list of proteins identified in the MS analysis and their fold change of enrichment is shown in Table 4.1 and a Heat map presenting differentially expressed proteins is shown in Figure 4.11.



	Altered Protein Expression		
	Increased	Decreased	Total DE
dsRNA pulldown	19	0	19

* ≥ 2 peptides
$p \leq 0.05$
fold change (≥ ± 1.5 fold)

Figure 4.10 Mass spectrometry results of three Biological replicates of dsRBP immunoprecipitation experiments using MaxQuant software.
Volcano plot shows mass spectrometry results of three Biological replicates of dsRBP IP experiments using MaxQuant software. Proteins significantly enriched in dsRBP IP with a corrected *p*-value ≥0.05 are shown in red, and represent the dsRNA interactors. Proteins identified with *p*-values>0.05 are depicted in grey. Proteins based on less than two peptides are represented by black dots.

Table 4.1 Representative list of proteins enriched in dsRBP in HD patients

The group of significantly enriched dsRBP ranked according to the fold change.

STATISTICALLY SIGNIFICANT

Protein name	Gene name	Accession number	Fold change	Unique peptides	Unique sequence coverage [%]	Mol. weight [kDa]	-Log Student's T-test p-value
Protein POF1B	POF1B	Q8WVV4-1	58.7	7	13.8	68.7	1.52631
Prelamin-A/C	LMNA	P02545	55.9	6	10.2	74.1	1.51265
Heat shock cognate 71 kDa protein	HSPA8	P11142	45.1	4	9.6	70.9	1.34544
Protein-glutamine gamma-glutamyltransferase K	TGM1	P22735	35.2	8	12.6	89.8	2.61043
Alpha-enolase	ENO1	P06733	30.6	12	31.8	47.2	1.31635
Protein S100-A16	S100A16	Q96FQ6	14.6	2	22.3	11.8	1.63641
Protein-glutamine gamma-glutamyltransferase E	TGM3	Q08188	14.5	13	24.0	76.6	1.34173
Peroxiredoxin-2	PRDX2	P32119	12.9	2	9.1	21.9	1.39694
Arginase-1	ARG1	P05089	11.6	6	22.4	34.7	1.49773
Polyubiquitin-B	UBB	P0CG47	11.5	5	61.6	25.8	1.64721
Plakophilin-1	PKP1	Q13835	11.3	5	7.0	82.9	2.56663
Neuroblast differentiation-associated protein AHNAK	AHNAK	Q09666	9.2	2	1.8	629.1	1.35553
Annexin A2	ANXA2	P07355	7.9	11	34.5	38.6	1.5239
Desmoplakin	DSP	P15924	7.8	50	19.4	331.8	1.57631
Serpin B12	SERPINB12	Q96P63	6.8	5	12.8	46.3	1.75561
Protein S100-A14	S100A14	Q9HCY8	6.4	2	21.2	11.7	1.33999
Desmoglein-1	DSG1	Q02413	6.1	16	20.1	113.8	1.79486
Desmocollin-1	DSC1	Q08554	5.9	7	10.0	100.0	1.41314

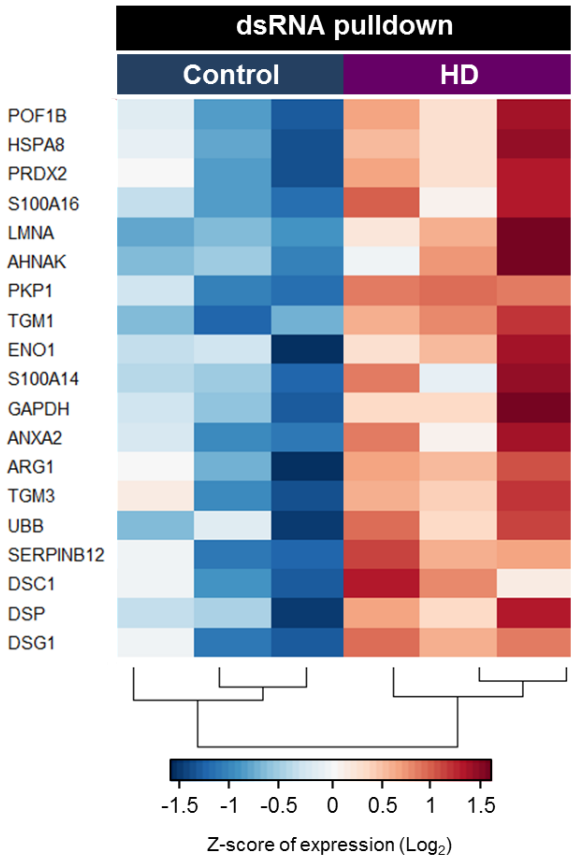


Figure 4.11 Heat map of differentially expressed proteins in control and HD.
Heat map displaying the expression of differentially expressed proteins identified from the proteomics analysis in control and HD cells. The colour scale illustrates the relative expression level of each protein across three Biological replicates; blue and red indicate lower and higher expression; the colour intensity indicates the degree of protein up or downregulation.

4.6.3 Characterisation and Biological validation of the dsRNA-binding proteins

To gain insights into the pathways that play Biological roles in HD cells, all differentially expressed proteins were subjected to the Kyoto Encyclopedia of Genes and Genomes (KEGG). Gene ontology analysis for cellular compartment indicated that the dsRNA-binding proteins are highly enriched in the nucleus (49.1%), but there are some proteins present in the cytoplasm (24.5%), in the nucleoplasm (13.2%) and in the nucleolus compartment (9.4%) (Figure 4.12). A small fraction of proteins in the dsRBP IP is present in other compartments, such as the cell junction.

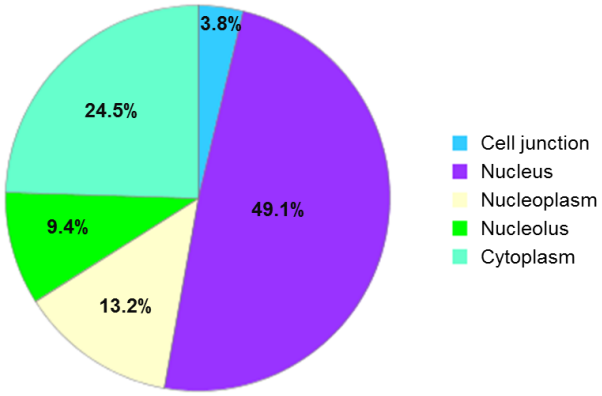


Figure 4.12 Analysis of cellular component enrichment of dsRNA interactors in HD. Cellular compartment analysis of dsRNA interactors in five cellular components, including cell junction, nucleus, nucleoplasm, nucleolus and cytoplasm.

A list of the significantly enriched proteins identified in the MS analysis and their Biological functions are shown in the Table 4.2. Expectantly, the dsRBP proteome might detect some new or previously identified strong candidates involved in dsRNA Biology. The dsRBP pull-down was enriched for proteins involved in transcriptional repression, such as the heat shock cognate 71 kDa protein (HSPA8) and alpha-enolase1 (ENO). HSPA8 is part of the HSP70 family, whose overexpression can protect cells with protein aggregates from apoptosis (Novoselova *et al.*, 2005) and ENO1 was previously found to be upregulated in human post-mortem brain tissue with HD, as a potential marker of

neuronal degeneration (Marangos and Paul, 1981). Interestingly, other proteins are involved in the DNA damage response such as polyubiquitin-B (UBB) and prelamin-A/C (LMNA). LMNA disrupts heterochromatin distribution and DNA damage repair pathways leading to genomic instability and senescence (Brayson and Shanahan, 2017). Notably, the neuroblast differentiation-associated protein AHNAK is enriched in dsRBP proteome. AHNAK is involved in DNA repair and regulation of RNA splicing and is also required for neuronal differentiation (Davis *et al.*, 2014). AHNAK and ANXA2 can bind to S100 protein bindings, which play a role in apoptosis. Namely, S100A16 and S100A14 are RNA-binding proteins important for gene expression regulation, together with PKP1. The S100A16 protein was previously found associated with CAG repeat transcripts (Marenholz and Heizmann, 2004). ENO1 and S100A16 proteins were also found aberrantly expressed in tumours (Marenholz and Heizmann, 2004; Perconti *et al.*, 2017). Figure 4.13 shows the most represented functions of the proteins upregulated in the dsRBP proteome, are transcription regulation, apoptotic processes, mRNA processing, oxidative stress and immune system activity.

Overall, the analysis indicates that the proteome is enriched for RNA-binding proteins or nucleic-acid binding proteins, as many proteins have dual properties and can bind both DNA and RNA (Hudson and Ortlund, 2014). Thus, the data suggests that dsRBP IP detects mainly direct interactions rather than secondary interactions facilitated by protein-protein binding.

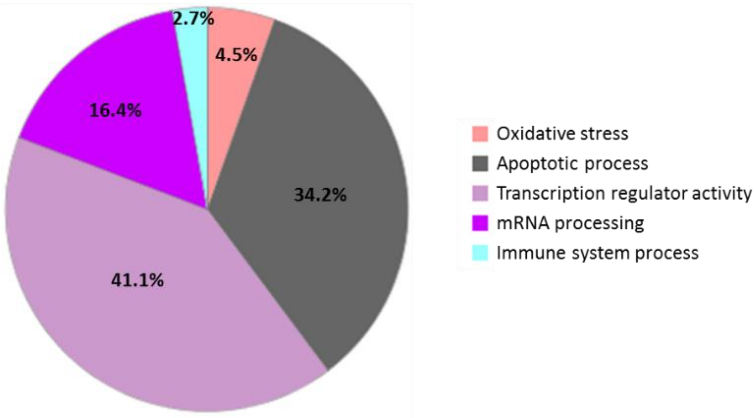


Figure 4.13 Molecular functions of the dsRNA interactors.
The most represented functions of the proteins upregulated in the dsRBP proteome are apoptotic processes, transcriptional regulation, mRNA processing, oxidative stress and immune system activity.

Table 4.2 Representative list of proteins enriched in dsRBP in HD patients
The group of significantly enriched dsRBP are involved in a wide range of Biological functions.

Protein	Description	Fold change
POF1B	Premature ovarian failure protein 1B	58.7
LMNA	Prelamin-A/C	55.9
HSPA8	Heat shock cognate 71 kDa protein	45.1
TGM1	Protein-glutamine gamma-glutamyltransferase K	35.2
ENO1	Alpha-enolase / C-myc promoter-binding protein	30.6
S100A16	S100 calcium-binding protein A16 / Aging-associated gene 13 protein	14.6
TGM3	Protein-glutamine gamma-glutamyltransferase E	14.5
PRDX2	Peroxiredoxin-2 / Natural killer cell-enhancing factor B	12.9
ARG1	Arginase-1	11.6
UBB	Polyubiquitin-B	11.5
PKP1	Plakophilin-1	11.3
AHNAK	Neuroblast differentiation-associated protein AHNAK	9.2
ANXA2	Annexin A2	7.9
DSP	Desmoplakin	7.8
SERPINB12	Serpin B12	6.8
S100A14	S100 calcium-binding protein A14	6.4
DSG1	Desmoglein-1	6.1
DSC1	Desmocollin-1 / Desmosomal glycoprotein 2/3	5.9

Next, western blotting of the highly enriched and most significant proteins was performed in order to validate the dsRBP proteome hits identified by mass spectrometry. In particular, the enrichment of four candidates was confirmed, including LMNA, HSPA8, ENO1 and S100A16, which are greatly enriched in the dsRBP proteome in HD (Figure 4.14). As in the Silver Staining gel the Control Input was shown to have reduced levels of the proteins compared to the Immunoprecipitated Control sample (Figure 4.8B), and as for some of the proteins analysed the material was not enough, thus only the Input HD was loaded in the Western Blot (Figure 4.14).

SDS-PAGE, silver staining gel, mass spectrometry and western blotting. Firstly, immunoprecipitated dsRNA-binding proteins were observed by SDS-PAGE and silver staining gel. As shown in Figure 4.8B, several dsRBP were detected in the IP sample, and the profile is different from the Input material, in the first lane.

The proteomics analysis had shown that all these proteins were overexpressed in the HD IP samples more than 8-fold compared to the control IP sample (Table 4.2). Several attempts were made to detect the neuroblast differentiation-associated protein (AHNAK), however due to its large size, 629 kDa, it was difficult to visualize it on the western blot. Although there are only few studies about proteins interacting with dsRNA in a disease context, it was expected that the dsRBP IP would detect some candidates that have previously been identified. The protein S100A16 was previously shown to be associated with CAG repeat transcripts (Marenholz and Heizmann, 2004) and alpha-enolase (ENO1) was found upregulated in human post-mortem brain tissue with HD (Marangos and Paul, 1981). Amongst proteins associated with gene repression, the ENO1 plays a role in transcriptional repressor activity and is a RNA Pol II core promoter sequence-specific binding and is a marker of neuronal oxidative processes (Ciancarelli *et*

4.6.4 The top candidate alpha-enolase1 (ENO1)

One important aspect of the dsRBP proteome is the discovery of novel Biological functions. Therefore, one of the most highly enriched proteins detected was chosen for preliminary functional assays. ENO1 has been previously shown to possess transcriptional repressor activity, acts as RNA polymerase II core promoter proximal region sequence-specific binding and can bind to both RNA and DNA (Ray and Miller, 1991). It has been demonstrated that ENO levels are higher in HD patients than in controls and expresses the neuronal oxidative processes involved in HD pathogenesis, acting as serum biomarker (Ciancarelli *et al.*, 2014). However, whether ENO1 also interacts with dsRNA in human HD fibroblasts has not been investigated. Therefore, western blotting was performed to validate the association between dsRNA and ENO1 proteins.

Firstly, HD samples were treated with benzonase which disrupted the ENO1-dsRNA complex. The next question was whether the dsRNA formed over the *HTT* gene could bind to ENO1. In this approach, after pull-down of the dsRNA, ENO1 was successfully immunoprecipitated and decreased in the presence of LNA gapmers and mixmers, which specifically target CAG repeats and which were shown to reduce *HTT* dsRNA. S100A16 was used as a control in the positive immunoprecipitation assay, because it was shown to interact with CAG repeat transcripts from the human brain (Marenholz and Heizmann, 2004). This result showed that ENO1 was efficiently immunoprecipitated from the dsRBP IP in HD cells and the LNAs had an effect on the ENO1-dsRNA interaction, suggesting that indeed ENO1 can bind and interacts to *HTT* dsRNA in HD (Figure 4.15A).

As dsRNA are naturally formed during transcription, it would be expected that inhibition of transcription would prevent co-immunoprecipitation of ENO1 in the dsRBP IP. Transcription was inhibited by treatment of cells with actinomycin D, which is the

4.7 Discussion of Chapter 4

The results shown in Chapter 4 show that formation of dsRNA over the CAG expanded repeats can trigger transcriptional repression of the *HTT* gene. The data provides additional understanding into the pathomechanism of human HD and potentially into the molecular pathology of other trinucleotide expansion disorders. In addition, this work describes the interaction with dsRNA and ENO1 protein involved in transcriptional regulation of the *HTT* gene in HD fibroblast cells.

dsRNA are formed at transcribed expanded CAG repeats

The data presented indicates that stable dsRNA structures are formed over expanded CAG repeats, while RNA/DNA hybrids are not detected in the *HTT* gene and do not interfere with efficient transcription in HD human cells. It is very important to observe that dsRNA levels are associated with CAG repeat length and this correlation is associated with the molecular pathology of HD.

The toxicity seen in HD may not just be affected by the CAG expansion, but may depend on the formation of an expanded CAG-hairpin RNA that is toxic. Long CAG trinucleotides form stable hairpin structures and is the stem of the RNA loop that has protein binding properties (Jasinska *et al.*, 2003; Galka-Marciniak *et al.*, 2012; Krzyzosiak *et al.*, 2012). It is known that these secondary RNA structures can be mislocalized in the cell, leading to abnormal interactions with harmful proteins dysregulating gene expression and alternative splicing (Marti, 2016). In terms of the nature of dsRNA, CNG repeats where N is A, C, G or U are more susceptible to form stable hairpins (Sobczak *et al.*, 2003) and this structure can interact with different RNA-binding proteins with a potential role in gene expression regulation.

The LNA mixmers used in this study target the CAG repeats and were designed to reduce HTT expression and the LNA gapmer targeting the CAG expansion was also included. Gapmers contain LNA bases flanking a central DNA region and induce silencing via recruitment of RNase H to cleave the mRNA of a DNA-RNA duplex. LNAs were shown to reduce expression of the HTT protein and mRNA *HTT* expression and in this study were used as a tool to investigate gene expression regulation.

In order to test the ability of CAG repeats to form hairpins *in vivo*, the dsRNA-specific J2 antibody, which binds RNA duplexes ≥40 base pairs (Schonborn *et al.*, 1991; Bonin *et al.*, 2000) was used. The J2 antibody has been widely used to detected endogenous dsRNA formation in mechanisms of RNAi and Pol II–associated Dicer (White *et al.*, 2014; Cruz and Houseley, 2014; Wu *et al.*, 2016) and in transcription processes in mammalian genes (Skourti-Stathaki *et al.*, 2014). To confirm the specificity of the dsRNA J2 antibody, dsRNA V1 RNase, which cleaves base-paired nucleotides in double-stranded regions of the RNA (Ehresmann *et al.*, 1987), and single-stranded-specific S1 nuclease treatments (Kertesz *et al.*, 2010) were carried out. RNase V preferentially cleaves phosphodiester bonds 3’ of double-stranded RNA, whereas S1 nuclease preferentially cleaves 3’ of single-stranded RNA (Kertesz *et al.*, 2010). The structures of RNA molecules are essential for their function and regulation, thus the data generated with these two complementary enzymes permit us to measure the degree to which each nucleotide was in a single- or double-stranded conformation. After RNase V1 and S1 nuclease treatment, it was possible to confirm by RT-qPCR the presence of endogenous dsRNA over the *HTT* gene and their enrichment specifically at 5’ end of the gene, where the CAG repeats are located. In addition, after transfection with LNAs, the dsRNA levels were reduced, indicating that indeed dsRNA are formed by CAG repeats.

Diseases such as DM1 and Fragile X syndrome are caused by a toxic RNA gain-of-function (Brook *et al.*, 1992; Orr, 2004). The toxicity of the harmful RNA is determined by its propensity to fold into hairpin structures, which initiates abnormal interaction with dsRNA-binding proteins, leading to loss of their normal function in RNA mechanism (de Mezer *et al.*, 2011). To gain insight into the structural features of complex triplet repeats of *HTT*, their structural conformations were studied *in vitro* and demonstrated the contribution of neighbouring sequence repeats to CAG repeat hairpin formation. HTT fragments are present within regions flanking polymorphic CAG repeats and other triplet repeats of neighbouring sequences. Particularly, in the case of *HTT* locus, the CAG tract is adjacent to polymorphic $(CCG)_n$, 12 bp downstream of the of $(CAG)_n$ expansion (de Mezer *et al.*, 2011). In the same study they demonstrated that the *HTT* transcript is retained in the nucleus of human HD fibroblasts and is colocalized with the MBNL1 protein (de Mezer *et al.*, 2011). It was observed that expanded transcripts can be formed of CAG interacting with CCG repeats or can be composed by the fold-back structure of pure CAG repeats (de Mezer *et al.*, 2011) and the hairpin structure is preserved in full-length mRNA (Michlewski and Krzyzosiak, 2004). These studies can possibly explain the enrichment of dsRNA detected by RT-qPCR in the *HTT* gene in the region downstream of the CAG expansion, due to the presence of a small CCG trinucleotide expansions (Figure 4.16).

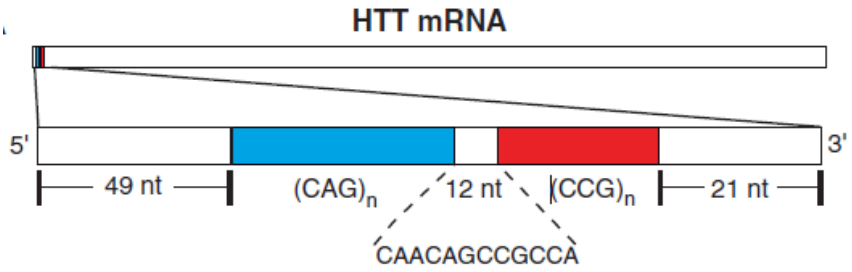


Figure 4.16 Schematic representation of HTT mRNA fragments.
White bars, specific flanking sequences; blue bars, polymorphic CAG repeats; red bar, polymorphic CCG repeat region (de Mezer *et al.*, 2011).

dsRNA and heterochromatin formation

In HD, the *HTT* dsRNA formed by CAG repeats is located at the 5'end of the *HTT* gene and is it known that it may have different roles depending on their genomic location and sequence conformation. The colocalisation of dsRNA with the repressive histone modification H3K9me2 indicates a functional interplay between dsRNA and *HTT* gene regulation in HD cells. It is possible that the dsRNA enrichment is linked to heterochromatin formation, facilitating the recruitment of repressive marks and mediating *HTT* repression in HD.

Treating the cells with the UNC0638 methyltransferase inhibitor reduces the levels of the repressive H3K9me2 chromatin mark and this leads to decreased levels of dsRNA on the expanded allele. Moreover, G9a profile colocalized with the dsRNA enrichment over the expanded repeats on the *HTT* gene. This provides an interesting possibility that *HTT* dsRNA may recruit G9a methyltransferase to promote deposition and spreading of the H3K9me2 histone mark. Notably, reducing dsRNA levels leads to a decrease in H3K9me2 and subsequent increase of RNA Pol II levels, alleviating the repression of the

HTT gene expression. These results indicate that dsRNA trigger H3K9me2 enrichment and heterochromatin formation leading to repression of the *HTT* gene in HD.

It is also apparent that dsRNA may be an early event during *HTT* transcriptional repression, appearing before the activation of the repressive histone modifications. This could explain the difficulty in reactivating HTT alleles using the histone methyltransferase inhibitor UNC0638, which suggests that dsRNA are not affected by these treatments and maintain the *HTT* gene in a repressed state.

Furthermore, it is possible that dsRNA interferes directly with RNA Pol II, because these structures may act as blockades for RNA polymerase. This structural block can directly interfere with transcription, resulting in reduced levels of RNA Pol II. It is known that during its journey, RNA Pol II has to go through many obstacles, including DNA-binding proteins, DNA damage and factors that are difficult to transcribe (Wilson *et al.*, 2013). In some cases, RNA Pol II may be ubiquitylated and degraded by the proteasome, after being irreversibly stalled or degraded at expanded loci (Wilson *et al.*, 2013). Interestingly, this mechanism of Pol II arrest may be a possible cause for reduced transcription in the region surrounding CAG repeats described here and could also be a feature of other trinucleotide expansion diseases. In addition, antisense transcripts have been detected in many triplet repeat-containing genes and may contribute to disease pathology. The consequences of antisense transcription on expanded repeats include regulation of sense transcription, production of toxic proteins and small repeat-containing RNAs that lead to post-transcriptional gene silencing (Groh, Silva, et al., 2014).

H3K9me2 serves also as a platform for recruitment of the heterochromatin protein 1 (HP1) that leads to stable repression via formation of a heterochromatic structure, allowing the binding of H3K9 trimethyltransferase such as SUV39H1 (Mozzetta *et al.*, 2015). Moreover, in Chapter 3 is shown that H3K9me3 repressive mark is also enriched

at expanded CAG repeats in HD. The chromatin reorganization is done by a cascade of several events that potentially include H3K9 dimethylation by G9a and H3K9 trimethylation by SUV39H1. All these events can be triggered by dsRNA present over CAG repeats. However, further work is required to understand these potential links.

Investigation of novel dsRNA interacting proteins in Huntington's disease

The knowledge of interactions between dsRNA and proteins in human diseases is still very limited. This work demonstrates a mass spectrometry approach to identify proteins binding to dsRNA in HD fibroblast cells, using the dsRNA-specific J2 antibody, and comparing their enrichment with the control fibroblast cells. This method identified 123 proteins differentially expressed in HD cells, and particularly 19 proteins that were highly enriched in HD, based on stringent statistical measures. The data suggest that binding of these proteins to dsRNA is transcription-dependent and dsRNA plays a role in multiple cellular mechanisms depending not only on their location, but on the binding of protein factors.

Proteomics analysis of dsRNA-binding proteins revealed proteins previously identified with a role in dsRNA Biology, including Staufen1 (Saunders and Barber, 2003), a splicing factor and a disease modifier in Myotonic dystrophy type I (Bondy-Chorney, Parks, Ravel-Chapuis, Jasmin, *et al.*, 2016) and novel dsRNA-binding candidates. The statistically significant dsRNA interactors are highly enriched and play a role in fundamental processes including transcriptional repression, regulation of RNA splicing and processing, apoptosis, DNA damage, chromatin organization and some proteins are markers of neuronal oxidative stress, and associated with diseases of the nervous system or found upregulated in tumours. The interactome is highly enriched for

proteins located in the nucleus, even though the majority of the proteins have dual DNA and RNA-binding roles.

This work suggests that in HD patient cells, upregulated proteins and RNA molecules can influence gene expression regulation. RNA is now involved in several roles due to activities caused by its secondary structures. RNA strands are able to fold back on themselves and generate stable conformations, affecting RNA transcription, splicing, translation and localization and all these processes are tightly and directly linked to dsRNA-binding proteins (Mortimer *et al.*, 2014).

The dsRNA interactor alpha-enolase, ENO1

Proteomics analysis identified alpha enolase1 (ENO1) as one of the most enriched proteins of the interactome. ENO1 has a fold change of 30.6 in HD compared to the control IP and was predicted under stringent features, such as the number of peptides (12), student's *t*-test (-log student *t*-test *p*-value = 1.31635) and intensity of each Biological sample. The interaction between ENO1 and dsRNA was confirmed using antisense oligonucleotides (LNAs gapmers) and the transcriptional inhibitor Actinomycin D. LNA gapmers specifically target CAG repeats and were shown to silence the *HTT* transcript and in their presence ENO1 from the dsRNA pull-down is reduced. As dsRNA are naturally formed during transcription, it was anticipated that inhibition of transcription would prevent co-immunoprecipitation of ENO1 in the dsRBP IP. Inhibition of transcription with Actinomycin D treatment results in reduced levels of dsRNA of the *HTT* gene and subsequent ENO1 reduction in the HD dsRBP pull-down. These results showed that ENO1 can bind to *HTT* dsRNA, and suggest that this interaction is Biologically relevant in HD, with ENO1 participating in dsRNA-mediated

processes. Notably, ENO1 may be involved in transcriptional repression of *HTT* gene regulation mediated by dsRNA in HD.

ENO1 has been implicated in RNA metabolism, it shows transcriptional repressor activity and can bind to both RNA and DNA (Ray and Miller, 1991) and it has been shown to be enriched in HD patients, acting as a serum biomarker for neuronal oxidative processes (Ciancarelli *et al.*, 2014). Transcriptional dysregulation plays an important role in Huntington's disease, but the causes of selective downregulation of specific genes, such as *HTT*, are unclear. In this work it is proposed that the mutant *HTT* dsRNA interacts with ENO1 in HD fibroblasts and, in addition to providing a novel mechanism underlying transcriptional dysregulation in HD, these results also show potential implications for dsRNA Biology. HTT-ENO1 interactions are enhanced in the presence of the expanded CAG repeats, are dependent on the dsRNA formation and contribute to transcriptional dysregulation in HD.

The repressive histone marks H3K9me2, H3K9me3 and H3K27me3 are enriched in HD cells and correlate with dsRNA over expanded CAG repeats in the *HTT* gene. It is possible that in addition to the role of dsRNA interactors in transcription, these factors could also affect chromatin organization, and function as epigenetic modifiers. Thus, it will be important to investigate the functions of dsRNA-binding proteins in RNA-mediated epigenetic regulation of gene expression.

The function of triplet repeats is still unclear, and this work together with other studies suggest that they can play a role in different Biological processes due to their interaction with binding proteins and due to their structural conformation (McLaughlin *et al.*, 1996; Peel *et al.*, 2001). The CAG repeats-containing dsRNA can be located in the nucleus, and involved in processes such as RNA splicing, maturation and transport. At the cytoplasm, they can also regulate mRNA stability and translation.

RNA structures may function differently depending on their genomic location, conformation and sequence composition as well as their triplet repeat expansion. The results described in this chapter propose a significant role for RNA structure in the pathology of HD. Currently there is evidence that *HTT* RNA contributes to the toxicity (Marti, 2016) and that RNA formed by CAG repeats are predisposed to aggregate into nuclear RNA foci (Echeverria and Cooper, 2012), which can be the reason for some proteins being trapped in the nucleus. Structural analysis of the *HTT* gene reveals that the CCG repeats downstream of CAG expansion provides stability to the hairpin structure (de Mezer *et al.*, 2011) and so it would be possible that some dsRNA hairpins are formed of CAG interacting with CCG repeats, yet LNA gapmer studies suggest that the hairpin structure is mostly formed of pure CAG repeats.

The dsRNA interactome can be an important resource to investigate dsRNA Biology in health and disease and it is apparent that RNA toxicity might be as complex as protein-induced toxicity causing RNA defects. The effects caused by the expanded RNA products can be a consequence of the interactions with RNA-binding proteins associated with the repeats (Budworth and McMurray, 2013).

However, extended Biological studies would be important to investigate deeper biochemical and functional roles in order to unravel their interplay with dsRNA and their contribution in health and disease. Indeed, differentiating between dsRNA that are 'harmful' from 'useful' is a relevant question in the dsRNA Biology and this could be assessed with the study of dsRBP.

Overall, this work suggests that dsRNA can promote transcriptional dysregulation in HD and elucidates a novel understanding of the molecular mechanisms underlying the HD pathology. The results suggest that dsRNA are capable of triggering transcriptional dysregulation in HD and these structures can be seen as potential targets to treat

trinucleotide expansion disorders. Besides elucidating the molecular mechanisms underlying HD pathology, this work also investigated interesting effects of dsRNA Biology, which may be a common feature of various disorders, which likely share Biological and pathogenic features.

Chapter 5 | Increased global transcription leads to RNA/DNA hybrid accumulation and DNA damage in cancer

The work described in this chapter has been published: Panagiotis Kotsantis, Lara Marques Silva, Sarah Irmscher, Rebecca M. Jones, Lisa Folkes, Natalia Gromak & Eva Petermann. **Increased global transcription activity as a mechanism of replication stress in cancer.** Nature Communications 7, 13087 doi: 10.1038/ncomms13087 (2016).

Introduction

Genome instability is a hallmark of most cancers (Negrini *et al.*, 2010), characterized by high mutation rates and genomic rearrangements (McGranahan *et al.*, 2012). Recent studies have shown that dysregulation of RNA/DNA hybrids (R-loops), the three-stranded nucleic acid structures that comprise nascent RNA hybridized with the DNA template (Skourti-Stathaki and Proudfoot, 2014), can affect genome integrity, chromosomal instability and recombination events (Hamperl and Cimprich, 2014). The link between R-loops and cancer has been demonstrated by studies on BRCA1 and BRCA2, which are mutated genes in breast and ovarian cancer that are required to prevent R-loop formation and genome instability (Bhatia *et al.*, 2014; Hatchi *et al.*, 2015). Importantly, spontaneous replication stress, arising when DNA replication fork progression slows or stalls, is seen as a central feature of cancer cells and an interesting phenotype feature to target for cancer therapy (Dobbelstein and Sorensen, 2015). Moreover, evidence suggests an involvement of R-loops as a source of replication stress in cancer Biology (Gan *et al.*, 2011). Yet it is still elusive whether R-loops are regulated differentially in tumour tissues and to what extent R-loops induce replication stress and subsequent genomic instability. To address this important question, the main aim of this work was to provide insights into the intricate crosstalk between transcription machinery, replication stress and cancer.

5.1 HRAS^{V12} leads to increased nascent RNA synthesis and transcription activity

In order to investigate the effects of oncogene activation on transcriptional dysregulation, RAS oncogene was overexpressed as this was previously shown to be sufficient to induce replication stress in cells (Di Micco *et al.*, 2006). Oncogenes promote cell proliferation, by interfering with growth factor signalling pathways. In the case of RAS and MYC oncogenes, they can induce cell growth by activating transcription and protein translation (Coleman *et al.*, 2004; White, 2005). Moreover, in various cancer types, the transcription machinery itself is dysregulated on a genome-wide scale (Hanahan and Weinberg, 2011). Oncogenic RAS regulates transcription through the extracellular signal-regulated kinase (ERK) pathway, which activates a number of transcription factors (White, 2005).

In this study used immortalized human fibroblasts transfected with pBabe-HRAS^{V12}-ER^{TAM} to express tamoxifen-inducible HRAS^{V12} (BJ-hTert HRASV12^{ER-TAM}) (De Vita *et al.*, 2005). Oncogenic HRAS^{V12} is induced by 4-hydroxytamoxifen (4OHT), which leads to increased mitogen-activated protein (MAP) kinases as shown by ERK1 and ERK2 phosphorylated proteins (Figure 5.1).

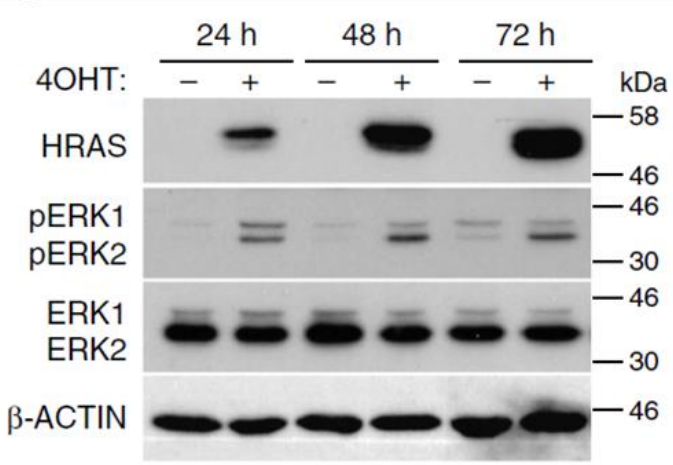
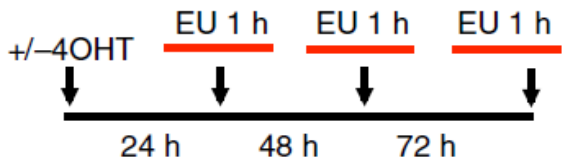


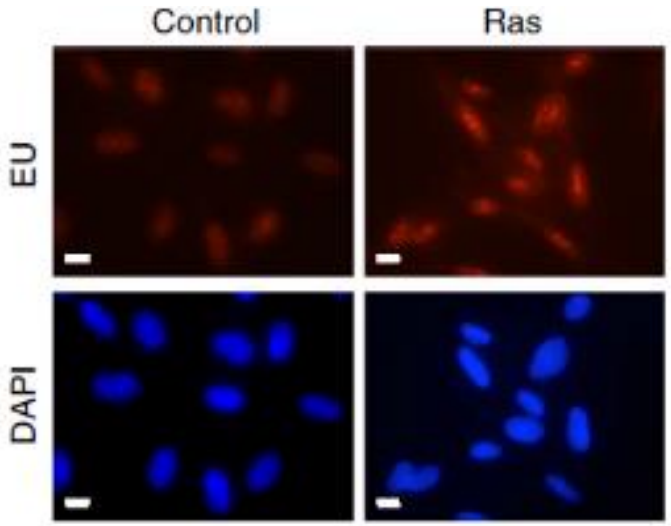
Figure 5.1 Protein levels of HRAS, pERK1/2, ERK1/2 and β-actin in BJ-HRAS^{V12} fibroblast cells after RAS induction for the times indicated.
This western blot was performed by Dr Panagiotis Kotsantis.

To characterise the effects of HRAS on the regulation of transcriptional activity, levels of nascent RNA synthesis were measured after RAS induction using nuclear incorporation of the modified precursor 5-ethynyluridine (EU) into newly transcribed RNA (Figure 5.2A, B). EU is a nucleoside analogue of uracil that is incorporated into RNA during active RNA synthesis, thus directly imaging spatially and temporally nascent global RNA transcription. To characterise the transcriptional dynamics, the nuclear EU intensity that represents RNA synthesis was measured 24h, 48h and 72h after HRAS^{V12} induction. As shown in Figure 5.2C, RNA levels increased approximately 2-fold after 48h and 72h HRAS^{V12} induction. After 24h, RNA synthesis in HRAS^{V12} induced cells was higher than the highest activity observed in control cells (Figure 5.2C). Importantly, 4OHT treatment did not change the levels of nascent RNA synthesis for the immortalized fibroblasts not expressing 4OHT-inducible HRAS^{V12} (Figure 5.2D). Therefore, the following experiments used 72h for the induction of HRAS^{V12}.

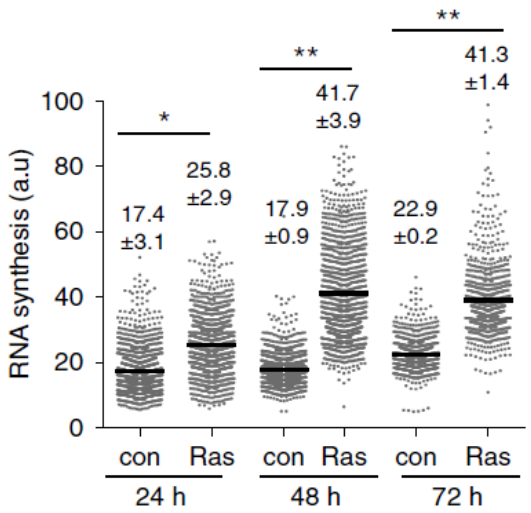
A



B


Control
Ras
EU
DAPI

C


RNA synthesis (a.u)
100
80
60
40
20
0
*
**
**
17.4
±3.1
25.8
±2.9
17.9
±0.9
41.7
±3.9
22.9
±0.2
41.3
±1.4
con
Ras
con
Ras
con
Ras
24 h
48 h
72 h

D

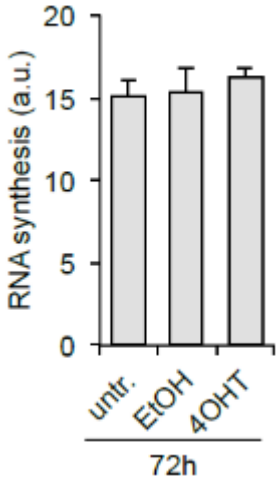

RNA synthesis (a.u.)
20
15
10
5
0
untr.
EtOH
4OHT
72h

Figure 5.2 HRAS^{V12} overexpression increases transcription activity.

A. Nascent RNA synthesis after 24 hours, 48 hours and 72 hours of RAS induction was measured by 1 hour of EU incorporation, as indicated.

B. Representative images of EU incorporation assay staining (in red) after 72 hours of RAS induction.

C. Quantification of nuclear EU intensity after RAS induction for 24 hours, 48 hours and 72 hours, n=4 (48 hours and 72 hours), n=5 (24 hours).

D. Nascent RNA synthesis measured by EU incorporation in BJ-hTert cells treated with EtOH, 4OHT or untreated for 72 hours (n=2).

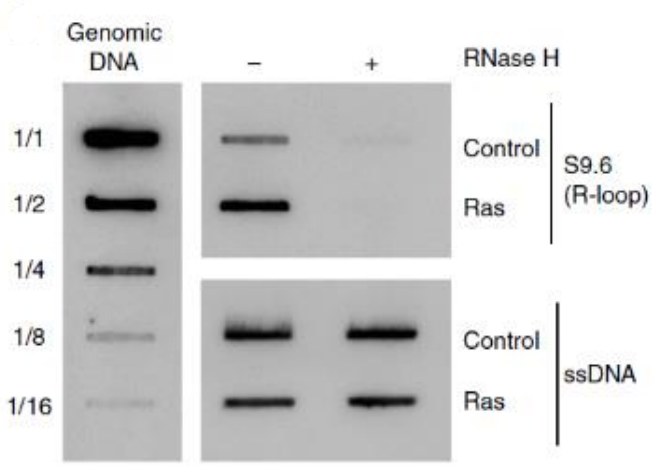
Asterisks (*) indicate statistical significance (* $p<0.05$; ** $p<0.01$; *** $p<0.001$), based on an unpaired, two-tailed distribution Student's t-test.

5.2 HRAS^{V12} leads to increased R-loop formation

It has been demonstrated that transcription, DNA replication and DNA repair are altered in cancer on a genome-wide scale (Hanahan and Weinberg, 2011). One important source of genome instability is RNA/DNA formation (Castellano-Pozo *et al.*, 2012), where the nascent RNA has re-annealed with the template. In line with this, the accumulation of RNA/DNA hybrids was investigated in HRAS^{V12} cells to test whether they are dysregulated and how this contributes to tumorigenesis.

RNA/DNA hybrids are detected using the monoclonal mouse antibody S9.6 (Boguslawski *et al.*, 1986). Employing the RNA/DNA hybrid slot blot to analyse isolated genomic DNA clearly showed a strong enrichment (3-fold increase) of immunoprecipitated RNA/DNA hybrids after 72h HRAS^{V12} induction (Figure 5.3A, B). S9.6 signal was removed by treatment with recombinant RNase H, which specifically degrades the RNA in RNA/DNA hybrids (Cerritelli and Crouch, 2009), suggesting that the signal was specific to the hybrids. The single-stranded DNA (ssDNA) antibody was used as a control and to confirm the specificity of the RNase H, degrading RNA/DNA hybrids. These results suggest that employing the RNA/DNA hybrid slot blot clearly showed a strong enrichment of endogenous RNA/DNA hybrids in HRAS^{V12}-induced cells.

A



B

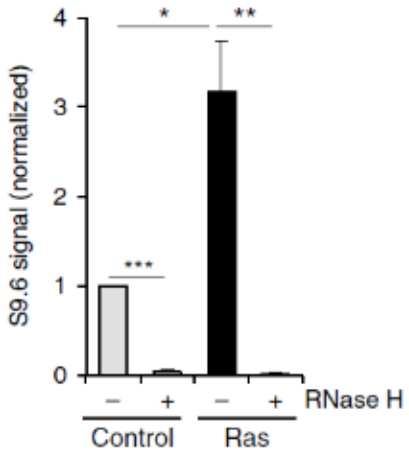


Figure 5.3 RNA/DNA hybrid slot blot of genomic DNA.

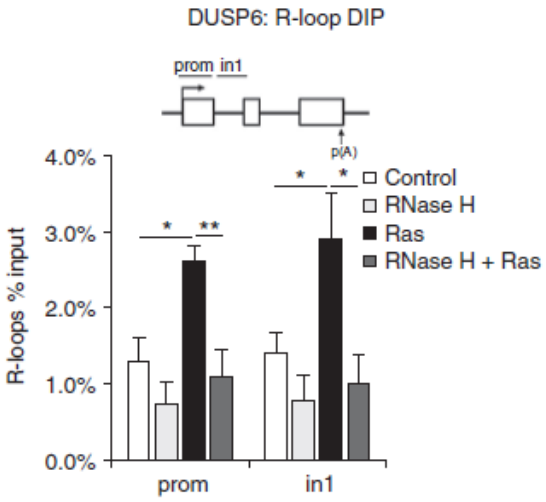
A. RNA/DNA hybrid slot blot of genomic DNA from $HRAS^{V12}$ cells after RAS induction for 72h, ± RNase H. S9.6 antibody was used to detect RNA/DNA hybrids (top panel) with single-strand DNA antibody (bottom panel) as a loading control. Serial dilutions of genomic DNA (1/1 = 4 µg) were probed with S9.6 antibody for standards (left panel).

B. RNA/DNA hybrid enrichment in $HRAS^{V12}$ compared with control.

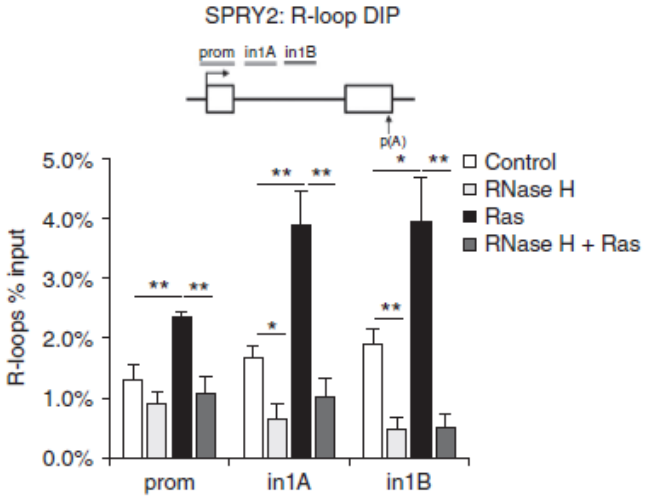
These analyses were performed by Dr. Panagiotis Kotsantis. Asterisks (*) indicate statistical significance (* $p<0.05$; ** $p<0.01$; *** $p<0.001$), based on an unpaired, two-tailed distribution Student's t-test.

The DNA/RNA immuno-precipitation (DIP) method permits quantitative detection of RNA/DNA hybrids on any endogenous genomic location in human cells (Skourti-Stathaki *et al*., 2011). In DIP, nuclei are extracted from non-crosslinked fibroblast cells, proteins removed and genomic RNA/DNA hybrids are isolated using the S9.6 antibody, which recognises RNA/DNA hybrids with a length of 6 bp with high-affinity (~0.5 nM) (Phillips *et al*., 2013). DIP is then followed by washes, DNA purification and qPCR analysis. DIP was used to investigate the distribution of RNA/DNA hybrid across RAS target versus non-RAS target control genes after 72h $HRAS^{V12}$ induction (Figures 5.4A – F). qPCR analysis on *DUSP6, SPRY2* and *c-FOS* genes, which are upregulated by activated RAS, indicated an enrichment of RNA/DNA hybrid formation on the promoter-proximal and intronic regions of all three target genes. DIP analysis showed that RNA/DNA formation is significantly increased over the transcribed regions of active genes (Figures 5.4A – C). On the non-RAS target control genes *GAPDH1*, *β-Actin* and *γ-Actin*, RNA/DNA hybrids are not formed (Figures 5.4D – F). To confirm the specificity of the DIP signal to detect RNA/DNA hybrids, DIP samples were treated before immunoprecipitation with RNase H, which specifically degrades the RNA in RNA/DNA hybrids (Cerritelli and Crouch, 2009). After RNase H treatment, the signal was strongly reduced for the control genes and target genes in both control and RAS induced cells, confirming that DIP specifically detected RNA/DNA hybrids. Overall, this data suggests that stimulation of transcription by $HRAS^{V12}$ results in increased R-loop formation.

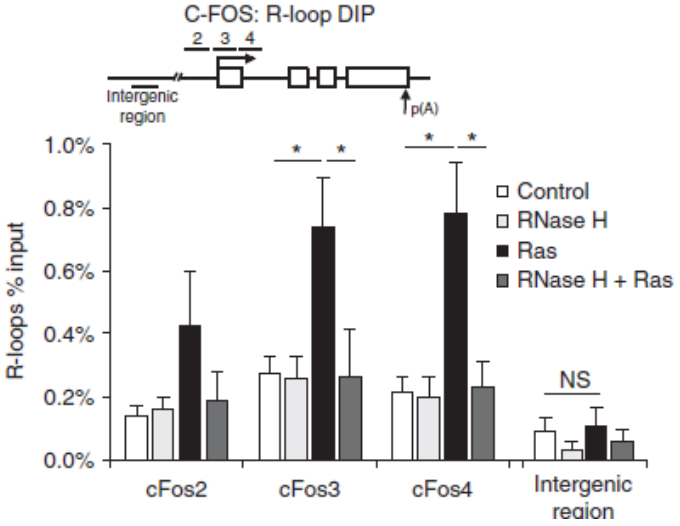
A


DUSP6: R-loop DIP
prom in1
p(A)
4.0%
3.0%
2.0%
1.0%
0.0%
R-loops % input
Control
RNase H
Ras
RNase H + Ras
prom
in1

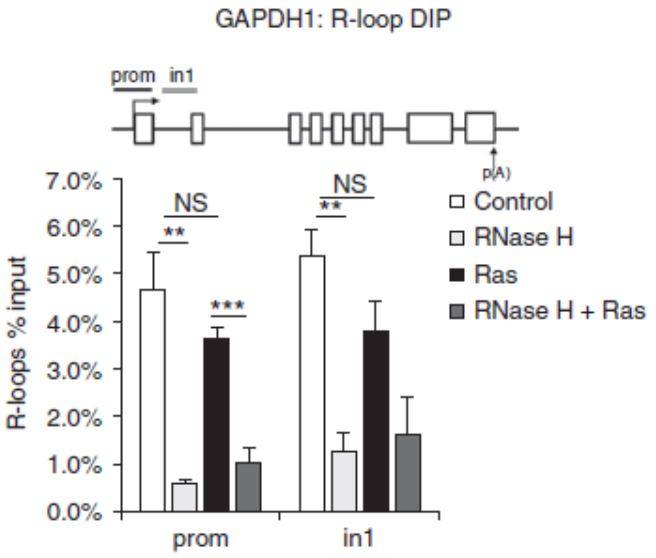
B


SPRY2: R-loop DIP
prom in1A in1B
p(A)
5.0%
4.0%
3.0%
2.0%
1.0%
0.0%
R-loops % input
Control
RNase H
Ras
RNase H + Ras
prom
in1A
in1B

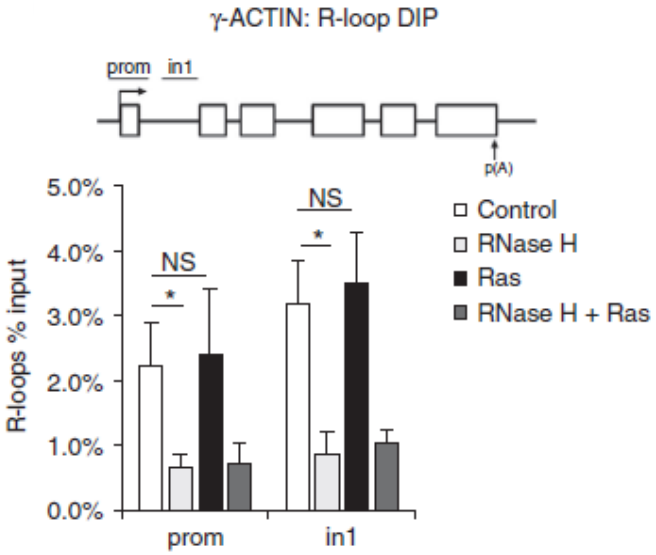
C


C-FOS: R-loop DIP
2 3 4
Intergenic region
p(A)
1.0%
0.8%
0.6%
0.4%
0.2%
0.0%
R-loops % input
Control
RNase H
Ras
RNase H + Ras
NS
cFos2
cFos3
cFos4
Intergenic region

D


GAPDH1: R-loop DIP
prom
in1
p(A)
R-loops % input
7.0%
6.0%
5.0%
4.0%
3.0%
2.0%
1.0%
0.0%
NS
**
NS
**
***
prom
in1
Control
RNase H
Ras
RNase H + Ras

E


γ-ACTIN: R-loop DIP
prom
in1
p(A)
R-loops % input
5.0%
4.0%
3.0%
2.0%
1.0%
0.0%
NS
*
NS
*
prom
in1
Control
RNase H
Ras
RNase H + Ras

F

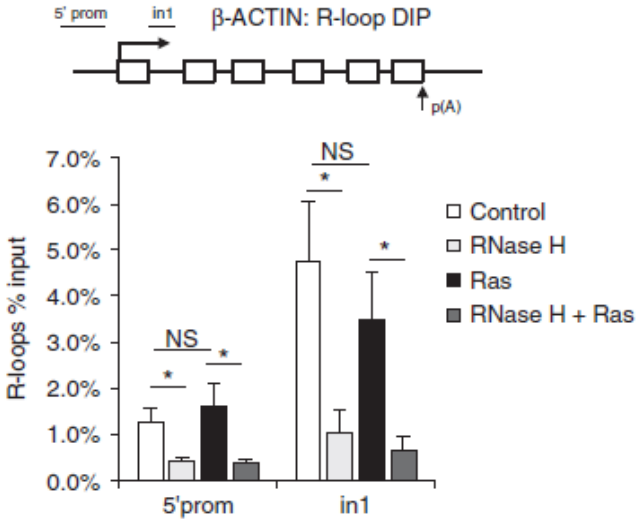

5' prom
in1
β-ACTIN: R-loop DIP
p(A)
R-loops % input
7.0%
6.0%
5.0%
4.0%
3.0%
2.0%
1.0%
0.0%
NS
*
NS
*
*
5'prom
in1
Control
RNase H
Ras
RNase H + Ras

Figure 5.4 HRAS^{V12} overexpression increases R-loop formation.
DIP analysis of R-loop enrichment on the *DUSP6* (A), *SPRY2* (B), *c-FOS* (C), *GAPDH* (D), *γ-Actin* (E), and *β-Actin* (F) genes in BJ-HRAS^{V12} after RAS induction for 72 hours (n=5). Intergenic region upstream of *c-FOS* promoter gene (C) was used as a background control. Values are illustrated as a percentage of input. Primer positions are shown in upper panels of A-F (white boxes represent exons).
Asterisks (*) indicate statistical significance (* p<0.05; ** p<0.01; *** p<0.001), based on an unpaired, two-tailed distribution Student's t-test.

5.3 Gene expression analysis and RNase A pre-treatment for R-loop DIP in HRAS^{V12} cells

Gene expression analysis showed increased RNA levels in RAS target genes *c-FOS*, *DUSP6*, *SPRY2* and *RNase H1* in HRAS^{V12} cells. In contrast, in the control genes *GAPDH*, *γ-Actin* and *β-Actin*, RNA levels remained constant between control and HRAS^{V12} cells (Figure 5.5).

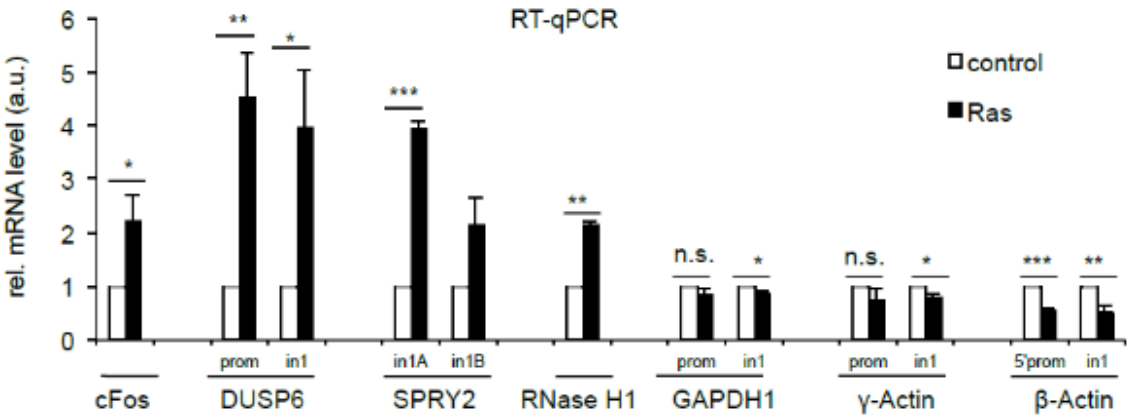
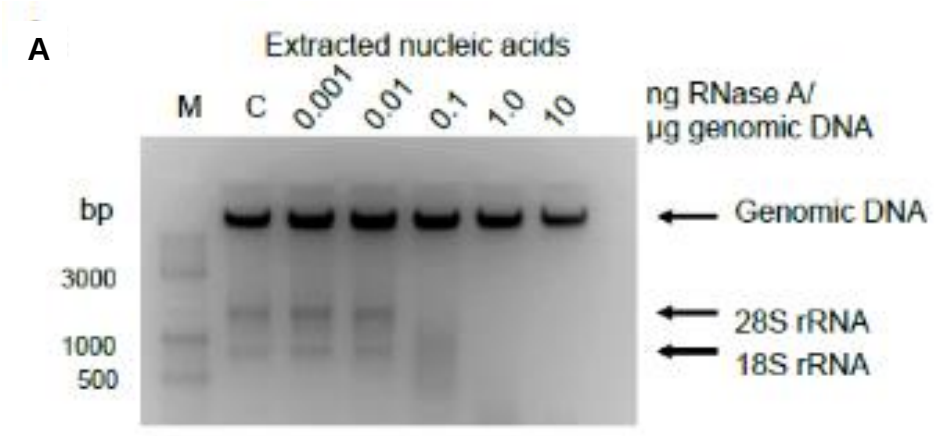
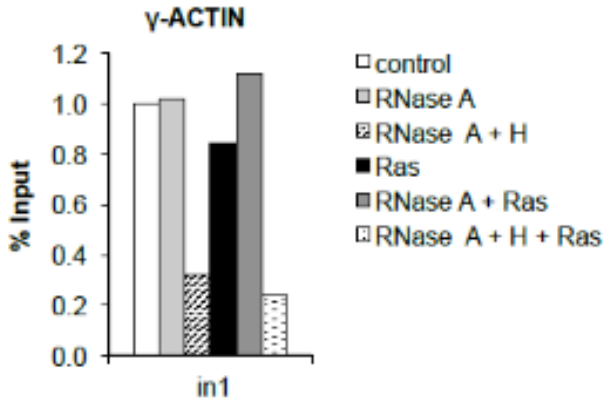


Figure 5.5 Gene expression analysis of HRAS^{V12} target and control genes.
RNA levels of *c-FOS*, *DUSP6*, *SPRY2*, *RNase H1*, *GAPDH*, *γ-Actin* and *β-Actin* genes in HRAS^{V12} cells, as determined by qRT-PCR (n=4). Values are relative to control cells for each amplicon.
Asterisks (*) indicate statistical significance (* p<0.05; ** p<0.01; *** p<0.001), based on an unpaired, two-tailed distribution Student's t-test.

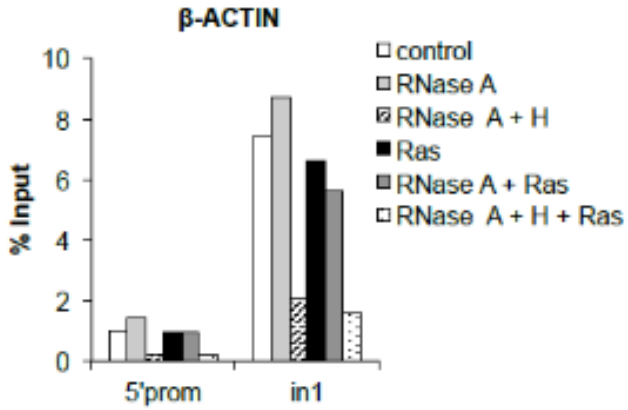
Previously it was shown that the S9.6 antibody could bind to AU-rich dsRNA, although with 5.6-fold lower affinity than for RNA/DNA hybrids (Phillips *et al.*, 2013). To investigate the ideal condition for RNA trimming, fibroblast cells were incubated with different concentrations of RNase A for 2 hours. In order to visualize the nucleic acids, an agarose gel electrophoresis was run and showed that RNase A considerably degraded the 28S and 18S rRNA species as low as 0.001 ng RNase A/µg genomic DNA (Figure 5.6A). It was possible to see a gradient of RNA degradation from 0.01 to 10 ng RNase A/µg genomic DNA. Also, at the highest RNase A concentration, it was possible to see a decrease of genomic DNA and associated RNA, the signal at the high molecular weight. Indeed, it seems that RNase A can also degrade genomic RNA/DNA hybrids at higher concentrations. Thereafter, in all subsequent experiments we used 1 ng RNase A/µg genomic DNA, where RNA/DNA hybrids were not affected (Figure 5.6A). In this way, in DIP experiments, RNase A treatment confirmed that the R-loop levels were not due to artificial hybridisation of free RNA species to DNA (or to other ssRNA forming dsRNA) during extraction and sample preparation or to S9.6 antibody recognizing double-stranded RNA (Figure 5.6B – D). In addition, the S9.6 signal could be removed by treatment with recombinant RNase H, supporting that it was specific to R-loops. These results confirm that stimulation of transcription by $HRAS^{V12}$ results in R-loop enrichment.



B



C



D

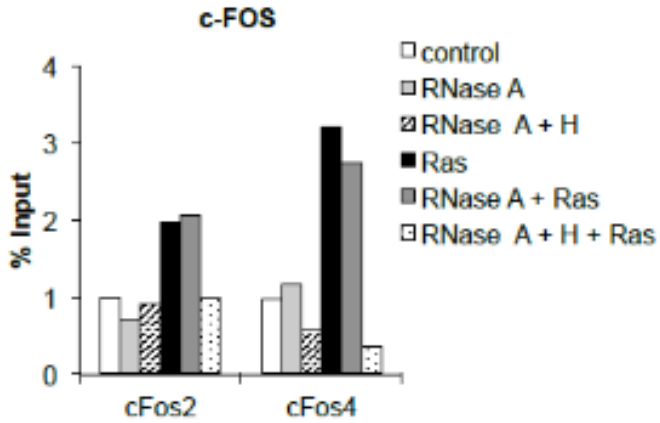


Figure 5.6 RNase A pre-treatment for R-loop DIP in HRAS^{V12} cells.

A. Agarose gel analysis of RNA trimming by RNase A during DIP. Purified genomic DNA (5 µg) was prepared according to the DIP protocol. RNase A was added to the indicated concentrations and incubated for 2 hours at 37°C. Control sample (C) was not treated with RNase A. Samples were treated with proteinase K and extracted

nucleic acids were separated on 1.2% agarose gel. DNA size markers are shown to the left of the gel (M). The migration of genomic DNA and ribosomal RNA (18S and 28S rRNA) is indicated on the right of the gel.

B. DIP analysis of R-loop induction on the *γ-Actin* gene after RAS induction for 72 hours (n=1).

C. DIP analysis of R-loop induction on the *β-Actin* gene after RAS induction for 72 hours (n=1).

D. DIP analysis of R-loop induction on the *c-FOS* gene after RAS induction for 72 hours (n=1).

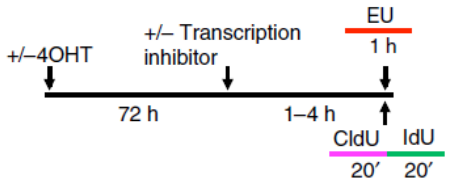
In (B), (C) and (D), genomic DNA samples were pre-treated with 1 ng/μg of RNase A and 1 ng/μg of RNase A + 10 U RNase H for 2 hours at 37°C before immunoprecipitation with S9.6 antibody. Values were normalised to the control cells in each amplicon.

5.4 Transcription promotes HRAS^{V12}-induced replication stress

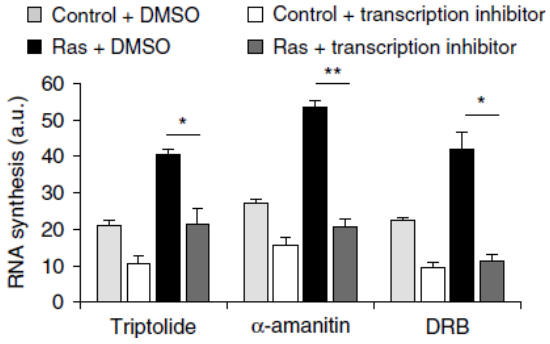
In order to test whether increased transcription causes replication stress in HRAS^{V12} cells, RNA synthesis was inhibited using small molecule inhibitors before replication fork progression was analysed. All the inhibitor compounds used here are able to inhibit eukaryotic transcription (Bensaude, 2011). Triptolide is fast and selective and able to completely arrest transcription by triggering rapid degradation of RNA Pol II, CDK9 inhibitors such as DRB are fast and reversible, with RNA Pol II pausing shortly after initiation, preventing a productive elongation of transcription, and α-amanitin is highly selective for RNA Pol II and RNA Pol III but its action is slow (Bensaude, 2011). To reduce effects on gene expression, the transcription inhibitors were incubated for short times at 100 min for triptolide and 5,6-dichloro-1-b-D-ribofuranosyl-1H-benzimidazole (DRB) and 4 hours for α-amanitin (Figure 5.7A). These treatments inhibited RNA synthesis, particularly in HRAS^{V12} cells (Figure 5.7B) but did not affect protein levels of HRAS^{V12} or levels of proteins with short half-lives such as Cyclin B1 and p53 (Figures 5.7C, D). Analysis of the distribution of replication fork speed demonstrated that all three transcription inhibitors increased replication fork speeds, particularly in HRAS^{V12} cells (Figure 5.7D). It is possible that such short incubations with transcription inhibitors may

not be sufficient to counteract the effects seen in transcription. Yet, these data suggest that RAS-induced replication stress is induced by active RNA synthesis. DRB and triptolide recovered replication more successfully than α-amanitin, suggesting that inhibiting transcription during early stages impedes more obstacles to replication forks than does inhibiting transcription during elongation (Figure 5.7E). Overall, our results indicate that deregulated transcription is a feature of cancer and promotes replication stress in $HRAS^{V12}$ cells.

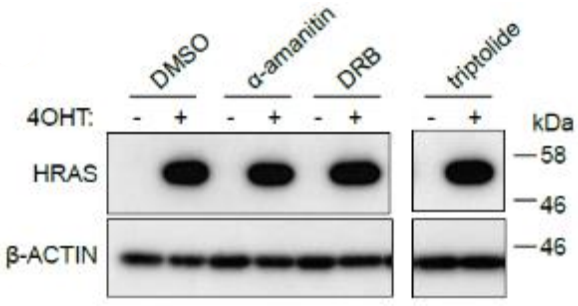
A



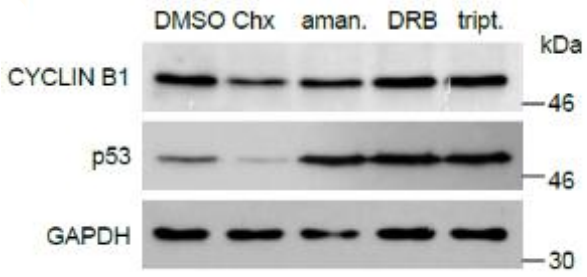
B



C



D



E

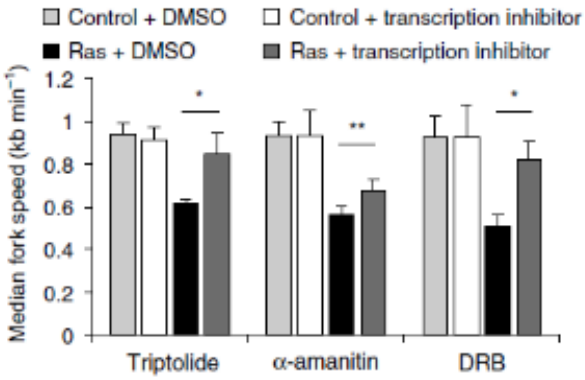


Figure 5.7 HRAS^{V12}-induced replication stress is promoted by ongoing transcription.

A. Seventy-two hours after RAS induction, BJ-HRAS^{V12} cells were incubated with transcription inhibitors triptolide, α-amanitin, DRB or dimethylsulfoxide (DMSO) for 100 min to 4 hours before and during EU.

B. Quantification of nascent RNA synthesis by EU incorporation in cells treated with transcription inhibitors 72 hours after RAS induction (n=3).

C. Protein levels of HRAS and β-ACTIN in cells treated with DMSO, DRB, or triptolide for 100 min or α-amanitin for 4 h, 72 hours after RAS induction.

D. Protein levels of CYCLIN B1, p53 and GAPDH (loading control) in BJHRAS^{V12} cells treated with 100 μg/ml cycloheximide or transcription inhibitors for 100 min.

E. Median replication fork speeds in cells treated with transcription inhibitors 72 hours after RAS induction.

Western blot and replication fork speed distribution analysis were performed by Dr. Panagiotis Kotsantis.

Asterisks (*) indicate statistical significance (* $p<0.05$; ** $p<0.01$; *** $p<0.001$), based on an unpaired, two-tailed distribution Student's t-test.

5.5 Transcription triggers R-loop-associated DNA damage

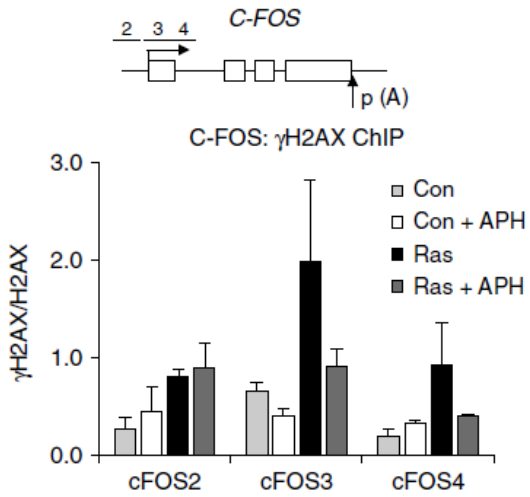
Genomic instability plays an essential role in the transition from normal to cancer cells, allowing them to growth and to metastasize, but not so much that it can be destructive to the cancer cells (Hanahan and Weinberg, 2011). Interestingly, another important source of replication stress is the collision of the replication machinery with R-loops. R-loops have protective roles and are implicated in many cellular processes, in a wide range of organisms, but still they can have deleterious effects on genome integrity (Santos-Pereira and Aguilera, 2015).

To investigate whether transcription and R-loop formation was associated with DNA damage in $HRAS^{V12}$ cells, γH2AX chromatin immunoprecipitation (ChIP) was performed. It has been demonstrated that DNA damage forms double stranded breaks (DSBs) and this event is generally followed by the phosphorylation of the histone variant H2AX, producing γH2AX (Kuo and Yang, 2008; Mah *et al.*, 2010). This phosphorylated protein plays a key role in recruiting DNA repair proteins at the sites possessing damaged chromatin, making γH2AX a sensitive molecular marker of DNA damage and repair (Podhorecka *et al*., 2010).

Therefore, γH2AX ChIP was carried out to test whether transcription and R-loop formation was associated with DNA damage in $HRAS^{V12}$ cells. In fact, γH2AX levels were increased in the promoter-proximal region of the *C-FOS* gene, associated with strong enrichment of R-loops, 72 hours after $HRAS^{V12}$ induction (Figure 5.8A). Moreover, $HRAS^{V12}$ ChIP samples were pre-treated with 0.5 mM Aphidicolin, a specific inhibitor of eukaryotic nuclear DNA replication (Spadari *et al.*, 1982), for 2 hours. This data showed that γH2AX levels decreased after Aphidicolin treatment, suggesting that the induction was replication dependent as it could be impeded by blocking replication (Figure 5.8A). In contrast, the levels of the DNA damage response, γH2AX, were similar

between control and HRAS^{V12}-induced cells before and after Aphidicolin treatment across the intron 1 region of the control *β-actin* gene (Figure 5.8B). Thus, DNA damage is seen in the genes previously shown to have increased R-loops and is replication-dependent. Cancer is a disease characterised by a deficient response to DNA damage, suggesting that R-loops may be the key players in dysregulating genome integrity. Overall, our results indicate that HRAS^{V12} triggers R-loop formation and DNA damage that also depends on replication.

A



B

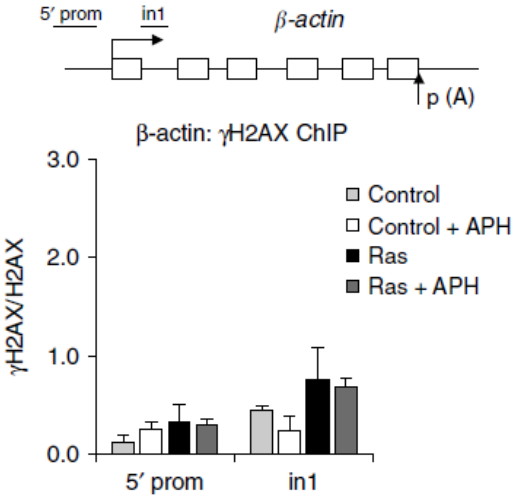


Figure 5.8 ChIP analysis of γH2AX after Aphidicolin treatment.

A. ChIP analysis of γH2AX versus total H2AX levels on the promoter of the *C-FOS* gene 72 hours after RAS induction (n=3). +APH samples were treated with 0.5 mM Aphidicolin for 2 hours before ChIP.

B. ChIP analysis of γH2AX versus total H2AX levels on the *β-actin* gene 72 hours after RAS induction as in A (n=3). qPCR primer positions are shown in upper panels.

Next, the role of R-loops in HRAS^{V12} cells was analysed in more detail. For this, the green fluorescent protein (GFP)-tagged human RNase H1 - an enzyme that degrades RNA/DNA hybrids when overexpressed - was transfected in HRAS^{V12} cells. Interestingly, the levels of protein and messenger RNA of endogenous RNase H1 were increased in cells overexpressing HRAS^{V12} (Figure 5.9A, B), confirming the enrichment of RNase H1 in the cells with increased R-loop levels.

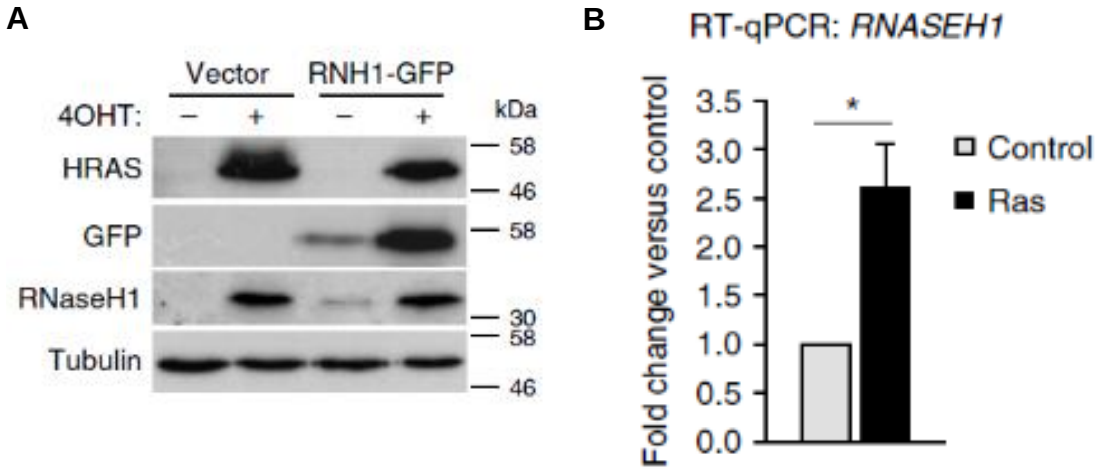


Figure 5.9 Endogenous protein and RNA levels of RNase H1.

A. Protein levels of HRAS, GFP, RNase H1 and Tubulin (loading control) 72 hours after RAS induction and 24 hours after transfection with RNase H1-GFP.

B. *RNase H1* mRNA quantification by quantitative reverse transcriptase–PCR after 72 hours RAS induction. *RNase H1* mRNA levels were normalized to *GAPDH* and control.

Western blot was performed by Dr. Panagiotis Kotsantis.

Asterisks (*) indicate statistical significance (* $p<0.05$; ** $p<0.01$; *** $p<0.001$), based on an unpaired, two-tailed distribution Student's t-test.

The work presented here together with data performed by members of Eva Petermann's group was published in the joint manuscript entitled ***Increased global transcription activity as a mechanism of replication stress in cancer***. *Nat. Commun. 7, 13087 (2016).*

$HRAS^{V12}$ oncogene promotes proliferation by increased transcription which leads to RNA synthesis. Increased RNA synthesis together with R-loop formation results in DNA damage. Members of Eva Petermann's group have also investigated in more detail the role of transcription factors in endogenous replication stress. The data suggests that in $HRAS^{V12}$ cells, the transcription factor TATA-box binding protein (TBP) is involved in the upregulation of transcription, which is associated with R-loop accumulation. Overall, this can be a mechanism of oncogene-induced DNA damage in tumorigenesis, providing an interesting connection between transcription machinery and genomic instability in cancer (Figure 5.10).

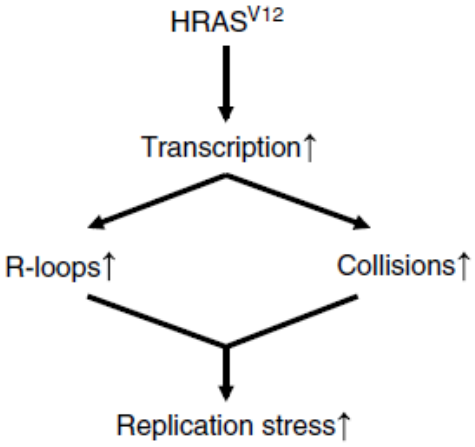


Figure 5.10 Model of how $HRAS^{V12}$ causes replication stress via increasing transcription.

Our data suggests that transcription contributes to $HRAS^{V12}$-induced replication stress primarily via direct conflicts with transcription machinery and R-loops.

5.6 Discussion of Chapter 5

Our data suggests that R-loop accumulation and increased transcription activity lead to replication stress and genomic instability in cancer. So far, the mechanisms involved in endogenous replication stress were deregulation of reactive oxygen species and cell cycle, namely premature or increased replication initiation, causing depletion of nucleotides or replication enzymes (Bester *et al.*, 2011). The transcription machinery is dysregulated in $HRAS^{V12}$ human cancer fibroblast cells and when upregulated has the potential to promote genomic instability. In this work, overexpression of oncogenic $HRAS^{V12}$ induces RNA synthesis and R-loop formation which can be a major cause for replication fork slowing and DNA breakage (Aguilera and Garcia-Muse, 2012). In the same way, it has been shown that loss of BRCA1 and BRCA2 tumour suppressor genes prevent accumulation of R-loops (Bhatia *et al.*, 2014; Hatchi *et al.*, 2015). Our results add a new view to this field, showing that deregulated transcription and increased R-loop levels is a feature of oncogenes, suggesting that they might be common in other cancer types.

Previous studies have demonstrated that R-loops can trigger genome instability, induce cancer and enhance the overall prognosis of the affected patient (Hamperl and Cimprich, 2014; Skourti-Stathaki and Proudfoot, 2014). Consequences of transcription-replication encounters have been proposed to be a major source of R-loop-associated DNA damage (Helmrich *et al.*, 2011). During replication fork, the genetic material is duplicated and transcription-replication complexes are interdependent and can lead to genomic instability in cancer (Aguilera and Garcia-Muse, 2012). Nonetheless, DNA damage is also present in non-replicating cells (Sordet *et al.*, 2009) most likely due to the exposure of single-stranded DNA of R-loops, which being more unstable is further

susceptible to lesions and to mutagenic molecules (Skourti-Stathaki and Proudfoot, 2014).

The importance of R-loops in replication stress mechanisms was previously demonstrated by a study showing that R-loops accumulate over the bacterial ColE1 origin of replication (Itoh and Tomizawa, 1980). Other studies have shown that R-loops unwind the DNA double helix, in order to make it more accessible for the initiation of replication (Baker and Kornberg, 1988). RNA/DNA hybrids also form at the origin of replication in mitochondria of *S. cerevisiae* and mammals (Xu and Clayton, 1996). The importance of R-loops in DNA replication stress has also been demonstrated in a genome-wide study in *E. coli* cells that have RNase H1-deficient, as they showed irregular replication fork progression (Maduike *et al.*, 2014). One recent study demonstrated that R-loops cause replication stress and DNA damage and can be resolved by factors of the Fanconi anemia (FA) pathway (Garcia-Rubio *et al.*, 2015). Thus, it is not surprising that genome instability associated with R-loops may contribute greatly to cancer of individuals with FA (Garcia-Rubio *et al.*, 2015).

Interestingly, cells overexpressing $HRAS^{V12}$ have increased levels of endogenous RNase H1. RNase H1 is located in both the nucleus and mitochondria (Cerritelli and Crouch, 2009; Suzuki *et al.*, 2010) where it plays a role in mitochondrial DNA replication and RNA processing, by removing the RNA/DNA hybrids during replication and transcription (Cerritelli and Crouch, 2009; Akman *et al.*, 2016). Overexpression of nuclear RNase H1 has been extensively used to remove R-loops (Cerritelli and Crouch, 2009) and is perhaps the best studied and effective way to reduce the cellular levels of R-loops. RNase H1 can counteract R-loop formation and the increase of these enzymes could be a response of cancer cells to increased R-loop levels. All these observations suggest that R-loops may contribute to tumorigenesis by providing proliferative

advantages and by promoting genome instability, which can lead to further mutations favourable to metastasis. These results support the idea that tumorigenesis is greatly associated with R-loop Biology.

R-loops are a double-edged sword (Skourti-Stathaki and Proudfoot, 2014), they form naturally during transcription and on the one hand they can have detrimental effects on genome integrity, on the other hand they are potential regulators of gene expression (Santos-Pereira and Aguilera, 2015; Skourti-Stathaki and Proudfoot, 2014). Additionally, they may be incorrectly recognised by DNA repair factors, leading to genome instability if not repaired (Sollier *et al.*, 2014). Recent years have produced growing evidence that dysregulation of R-loops can affect genome integrity leading to DNA damage, formation of double-strand breaks, chromosome breaks and recombination events (Hamperl and Cimprich, 2014). Therefore, it is expected that genome caretaker processes have evolved to protect R-loop levels (Aguilera and Garcia-Muse, 2012). There are specific proteins that act to maintain R-loops in balance and those have been identified mainly in tumours. For example, the E3 ubiquitin ligase BRE1 protects R-loop and its reduced expression leads to disruption of genome integrity (Chernikova *et al.*, 2012).

As mentioned above, the link between R-loops and cancer has been demonstrated in a study with the tumour suppressors BRCA1 and BRCA2, mutated in ovarian and breast cancer, where they were shown to function as "guardians" of R-loop and gene expression, preventing its accumulation (Bhatia *et al.*, 2014; Hatchi *et al.*, 2015). It seems that in cancer cells R-loops augment the probability of accumulating mutations which will favour the tumour growth and metastasis. It is very interesting to verify that some proteins have the potential to repress R-loops and others may induce their formation, thus leading to genome instability and so tumour growth. This has been shown in yeast models with the elongation factor Spt2 and Rad51 DNA repair protein (Sikdar *et al.*, 2008;

Wahba *et al.*, 2013). Overexpression of Spt2 results in increased R-loop formation and genome instability which can be reversed by RNase H1 overexpression (Sikdar *et al.*, 2008). However, it is still not clear how R-loops cause DNA damage. The present study suggests that one mechanism includes collisions between replication forks and R-loops as a major cause for DNA breakage.

Our findings show that transcription, R-loops and cancer are strongly related and demonstrate the potential to investigate additional factors playing a role in the mechanism. In the future, it would be interesting to investigate the role of R-loops in transcription as epigenetic modifiers. R-loops have recently been suggested to mediate heterochromatin formation in transcriptional termination of protein-coding human genes (Skourti-Stathaki *et al.*, 2014). It has been proposed that some tumour suppressors are silenced by DNA methylation (Karpf and Jones, 2002) and R-loop formation protects gene promoters from DNA methylation by the primary DNA methyltransferase DNMT3B in early development (Ginno *et al.*, 2012). Thus, it would be interesting to investigate R-loop function in cancer and its relation with DNA methylation. R-loops can add a new layer of gene expression regulation and all these studies can give a new perspective on the various R-loop cellular functions.

Changes in gene expression are another central aspect of cancer (Hanahan and Weinberg, 2011). In cells from healthy individuals, the expression of tumour suppressor genes impedes the uncontrolled proliferation and other features of tumorigenesis (Hanahan and Weinberg, 2011). R-loops can be regulated by different pathways depending on the cell type and indeed it has been proposed that they play a role in DNA damage and neurodegeneration (Sordet *et al.*, 2009), thereby leading to different kinetics.

It has been shown that oncogenes such as RAS and MYC do not only activate the cell cycle machinery but also induce cell growth through the activation of transcription

and protein translation (White, 2005). Indeed, overexpressed c-MYC can stimulate transcription by all three RNA polymerases (White, 2005; Lin *et al.*, 2012). One study showed that R-loop levels are kept low by the action of DNA topoisomerase IIIB to allow efficient transcription of the oncogene *c-MYC* (Yang *et al.*, 2014). This mechanism of c-MYC gene regulation can have an impact in tumour progression, which frequently shows overexpression of *c-MYC* (Yang *et al.*, 2014).

To test whether increased transcription causes replication stress in $HRAS^{V12}$ cells, small molecules were used to inhibit transcription activity. The most traditional inhibitors are: α-amanitin, DRB and triptolide (Bensaude, 2011). α-amanitin is highly selective for RNA Pol II and RNA Pol III but its action is slow, CDK9 inhibitors such as DRB are fast and reversible but many genes escape transcriptional inhibition. Lastly, triptolide is fast and selective and able to completely arrest transcription by causing rapid degradation of RNA Pol II (Bensaude, 2011). These treatments inhibited RNA synthesis but did not interfere with HRAS protein levels. The physiological effects of transcription inhibitors are known as the "transcriptional stress response" (Ljungman, 2007). All three transcription inhibitors increased replication fork speeds in the presence of $HRAS^{V12}$, suggesting that RAS-induced replication stress is triggered by active RNA synthesis. When transcription is inhibited by DRB and triptolide, replication is rescued more effectively than α-amanitin, suggesting that inhibition of early stages of transcription removes more obstacles to replication forks than inhibition during elongation.

To test whether transcription and R-loop are associated with DNA damage, γH2AX chromatin immunoprecipitation was performed. γH2AX, a sensitive molecular marker of DNA damage, was increased in HRAS cells and associated with strong enrichment of R-loops. H2AX expression can be used as a marker of DNA damage to detect genotoxic effects of different substances (Podhorecka *et al.*, 2010). In fact, evaluation of γH2AX

levels in cancer patient's samples may predict the efficiency of the anticancer treatment, inclusive the tumour cell sensitivity to DNA anticancer agents and toxicity of the treatment to the normal cells (Podhorecka *et al.*, 2010).

This study demonstrates that increased transcription activity in $HRAS^{V12}$ cells leads to replication stress and this mechanism is dependent on the increased expression of R-loop and of the transcription factor TATA-box binding protein (TBP) (shown by members of Petermann's laboratory). $HRAS^{V12}$-induced replication stress is facilitated by TBP, a general transcription factor that configures promoter DNA to the contours of the RNA Pol II surface (Kornberg, 2007). TBP binds to a DNA sequence called the TATA box, mediating the activation of eukaryotic genes transcribed by RNA Pol II. TATA box is frequently found 30 base pairs upstream of the transcription start site in some gene promoters (Kornberg, 2007). TBP is a component of the RNA polymerase II preinitiation complex (Lee and Young, 2000) and bends to the promoter DNA in a sequence-specific manner, aiding Pol II to adjust over the transcription start site to initiate transcription (Kornberg, 2007).

Components of the transcription machinery are frequently found mutated and differentially expressed in cancer (Bywater *et al.*, 2013) and TBP is one of the transcription factors involved in oncogene mechanisms (Johnson, Dubeau, Kawalek, *et al.*, 2003). TBP is differentially expressed in cancers and the *TBP* promoter has binding sites for oncogenic transcription factors. TBP mRNA and protein levels are increased in $HRAS^{V12}$ cells, however this increase does not affect RNA synthesis as much as $HRAS^{V12}$ overexpression, meaning that other transcription factors may play a role in this mechanism. Some cancer types exhibit mutations in TBP (Johnson, Dubeau, Kawalek, *et al.*, 2003; Johnson, Dubeau, White, *et al.*, 2003), yet it is still unclear whether such mutations are relevant for cancer pathogenesis. It has been shown that particular cellular

signalling cascades can induce TBP levels (Johnson, Dubeau, Kawalek, *et al.*, 2003). As shown previously, cells expressing oncogenic RAS have increased levels of TBP (Wang *et al.*, 1997) that can have pronounced effects on cellular transcription. Similarly, expression of one form of the hepatitis B protein was shown to be able to increase TBP levels through activation of RAS cellular signalling (Johnson *et al.*, 2000). TBP promoter is affected by RAS-mediated pathways through the activation of mitogen-activated protein kinase (MEK). Essentially, increased TBP can be a cause of MEK activation, a downstream target of RAS, and this could contribute to cellular transformation and tumorigenesis (Johnson *et al.*, 2000).

It has been demonstrated that cell cycle and transcription machinery are commonly dysregulated in cancer cells (Bywater *et al.*, 2013). As replication fork dynamics, transcription and the DNA damage are all interconnected in the cell, it would be important to understand these processes.

Our work elucidates the importance of R-loops in cancer, as their accumulation can be a major cause for DNA breakage and genomic instability. Additionally, increased RNA synthesis, transcription activity and increased R-loop levels could be correlated with replication stress in cancer. There are still multiple directions to explore the involvement of R-loops in cancer Biology. It would be advantageous to uncover whether R-loops have a direct effect on cancer progression. Lastly, further characterising the molecular mechanisms involved in RNA synthesis, increased R-loop levels and expression of RNase H1 associated with replication stress, may help to predict new biomarker candidates, and to unravel the role of R-loops and their relevance in cancer cells and tissues.

Chapter 6 Conclusions and outlook

6.1 Summary of results

6.1.1 Transcriptional and epigenetic regulation in Huntington's disease

The first part of this thesis aimed to investigate the pathogenic molecular events caused by the expanded CAG repeats in the *HTT* gene regulation. In this work, I have characterised the transcriptional and epigenetic regulation of the *HTT* gene in Huntington's disease. In particular, I have demonstrated that expanded CAG repeats adopt dsRNA structures that are involved in transcriptional repression and heterochromatin formation of the *HTT* gene in HD and potentially other trinucleotide expansion diseases. Furthermore, I have investigated the proteins interacting with dsRNA structures in HD fibroblasts to gain deeper understanding into dsRNA biology in health and disease. To this aim, I have combined dsRNA IP with quantitative mass spectrometry-based proteomic analysis to identify the dsRNA-binding proteins. I validated top candidates and demonstrated that a dsRNA interactor, ENO1, is involved in dsRNA-mediated transcriptional repression of the *HTT* gene (Figure 6.1).

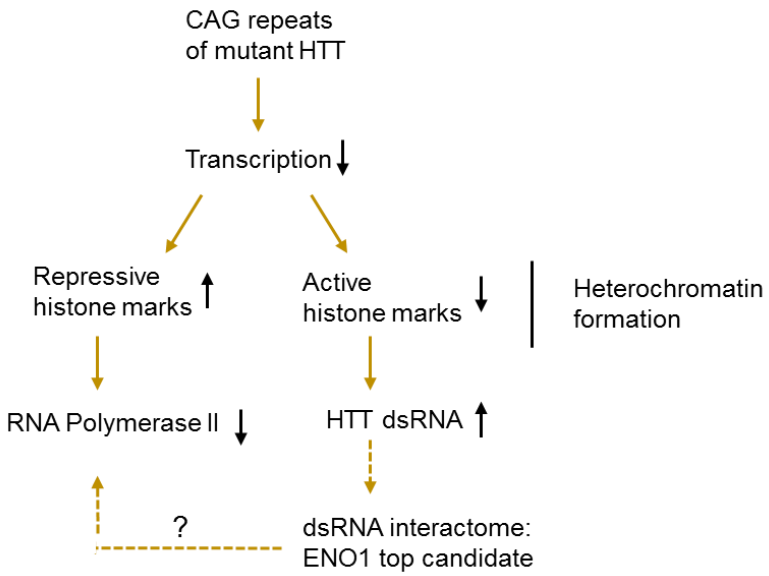


Figure 6.1 Model for the role of CAG expanded repeats of the mutant *HTT* in Huntington's disease. Transcribed CAG repeats induce transcriptional repression, leading to decreased levels of repressive histone marks and increased levels of active

histone marks (heterochromatin formation). This leads to downregulation of RNA Polymerase II (Pol II) expression and formation of HTT double-stranded RNA (dsRNA) in HD patient cells. ENO1 is a top HTT dsRNA interactor that may contribute to transcriptional repression and reduction of RNA Pol II.

Transcriptional alterations of HTT gene associated with CAG expansion in HD

Alterations in transcriptional regulation have been noted as a key mechanism in HD pathogenesis as a consequence of the expanded mutant *HTT* (Cha, 2007). Global gene expression studies have indicated that there was downregulation of transcripts in HD brain (Hodges *et al.*, 2006), and expression profiling in HD human brain (Hodges *et al.*, 2006) and mouse model brain (Luthi-Carter *et al.*, 2000; Zucker *et al.*, 2005; Kuhn *et al.*, 2007; Thomas *et al.*, 2011) also support this conclusion. Previously, most studies have investigated steady-state levels of *HTT* mRNA or relied on indirect read-outs of transcriptional activity, often compromised by the effects of cell loss in the brain (Luthi-Carter *et al.*, 2000; Zucker *et al.*, 2005; Hodges *et al.*, 2006). In addition, there has been little study of transcriptional regulation associated with the *HTT* gene itself *in cis*. The results from the work presented are in line with the previous studies, implicating transcription initiation and elongation defects as a feature of HD human cells. This work has expanded the characterisation of *HTT* gene regulation in health and disease, by analysing RNA Pol II recruitment and *HTT* pre-mRNA levels at multiple locations upstream and downstream of the CAG expansion in different healthy and HD patient-derived cells. It is possible that the pathology of diseases caused by CAG-expanded repeats translated to polyglutamine tracts also have a component dependent on RNA-related mechanisms. Furthermore, the distribution of the repressive histone modifications H3K9me2, H3K9me3, H3K27me3 are highly enriched around CAG repeats in HD cells, which are associated with transcriptionally inactive, heterochromatin regions. In contrast,

H3K9ac levels, linked with open, transcriptionally active genes in euchromatin, are lower in HD.

DsRNA are formed at triplet repeat expansions and contribute to transcriptional dysregulation in HD

Previous studies about the structural features of the CAG repeats suggested that the triplet repeats may form hairpin structures *in vitro* (de Mezer *et al.*, 2011). The present work demonstrated the formation of dsRNA at endogenous expanded CAG repeats that correlate with the expansion size, associated with more dramatic *HTT* repression, in HD human fibroblast and lymphoblast cells. Overall, these results indicate that accumulation of dsRNA at expanded CAG repeats are associated with increased levels of heterochromatin marks, suggesting that dsRNA can impair *HTT* gene expression. Importantly, depletion of dsRNA led to decreased H3K9me2 levels and restored transcription of the *HTT* gene in HD cells, but not in control cells. This suggests transcriptional repression is likely to be a consequence of high repressive mark levels, while other histone modifications may contribute to the dynamic nature of the epigenetic-chromatin regulation in HD. This should be taken into account in the design of future therapies, which would be more efficient if they target *HTT* dsRNA, as they are capable of triggering transcriptional silencing, instead of downstream epigenetic processes. The results suggest that dsRNA formed by CAG repeats in the *HTT* gene promote heterochromatin formation and lead to *HTT* transcriptional repression. The interplay between dsRNA, heterochromatin formation and transcriptional repression of the *HTT* gene suggest a novel functional role of dsRNA in *HTT* gene regulation in the HD pathology, but may also be a general feature of repeat expansion diseases.

Investigation of dsRNA-binding proteins in HD

In this study, a method was developed to characterise proteins interacting with dsRNA in healthy and HD fibroblast cells. Earlier RNA secondary structure prediction and *in vitro* immunoprecipitation studies provided evidence for functions of individual proteins in HD (Coolidge and Patton, 2000; Peel *et al.*, 2001; Sobczak *et al.*, 2003; de Mezer *et al.*, 2011). In this work, the proteome-wide identification of candidate dsRNA interactors may complement these functional studies with further elucidation on the dsRNA Biology in HD and provide new targets for future functional characterisations. Moreover, the dsRNA proteome was relevant as a discovery tool to identify top candidates, particularly for investigation of the top candidate, ENO1, which demonstrated that it interacts with dsRNA formed in the *HTT* gene, extending previous evidence of its function in the cerebral cortical tissue of HD as a potential marker of neuronal degeneration and as a neuronal oxidative biomarker in HD (Marangos and Paul, 1981; Ciancarelli *et al.*, 2014). There may be a potential cascade of epigenetic silencing including H3K9 dimethylation by G9a, followed by H3K9 trimethylation by SUV39H1 which can establish H3K27 trimethylation catalysed by the PRC2, all initially triggered by dsRNA formed at expanded repeats. This is in line with the view that histone modifying enzymes are part of a supermolecular complex where the regulation of transcription is dependent on the dynamics of the expanded CAG repeats. However, more work is required to uncover these potential links.

6.1.2 Investigation of transcriptional and genomic instability processes in cancer disease

The second major aim of this project was to investigate the effects of increased transcriptional activity on RNA/DNA hybrid formation and replication stress in cancer.

The research elucidates the importance of RNA/DNA hybrid in cancer as their accumulation is a major cause for DNA breakage and genomic instability. Together, our results suggest that increased RNA synthesis, transcription activity and RNA/DNA hybrid formation could be correlated with replication stress in cancer and results in replication fork slowing and DNA damage.

HRAS^{V12} leads to increased transcription activity and R-loop formation

Genome instability is a hallmark of most cancers (Negrini *et al.*, 2010), characterized by high mutation rates and genomic rearrangements (McGranahan *et al.*, 2012). Recent studies have shown that dysregulation of RNA/DNA hybrids (R-loops), the three-stranded nucleic acid structures that comprise nascent RNA hybridized with the DNA template (Skourti-Stathaki and Proudfoot, 2014), can affect genome integrity, chromosomal instability and recombination events (Hamperl and Cimprich, 2014). Therefore, it is expected that genome caretaker processes have evolved to protect R-loop levels (Aguilera and Garcia-Muse, 2012). In particular, the link between R-loops and cancer has been demonstrated by studies on BRCA1 and BRCA2, which are mutated genes in breast and ovarian cancer that are required to prevent R-loop formation and genome instability (Bhatia *et al.*, 2014; Hatchi *et al.*, 2015).

Even though it has been established that transcription promotes genomic instability, the molecular mechanisms involved remain poorly understood. In *E. coli* and *S. cerevisiae*, transcription was shown to hinder replication, leading to arrested replication forks (Prado and Aguilera, 2005; Mirkin and Mirkin, 2005).

Topoisomerase I (Top1) is a key enzyme in functioning at the boundary between DNA replication, transcription and mRNA maturation. It has been previously demonstrated that Top1 suppresses genomic instability in mammalian cells by preventing

a conflict between transcription and DNA replication (Tuduri *et al.*, 2009). In particular, abnormal post-transcription processes such as RNA splicing and export may result in formation of DNA breaks, described in yeast and mammalian cells depleted for Top1 (Huertas and Aguilera, 2003; Tuduri *et al.*, 2009). In higher eukaryotes, replication and transcription are synchronised processes and, unlike in bacteria, they occur within spatially and temporally separated domains (Wei *et al.*, 1998). When transcription occurs in S phase, it was suggested to be spatially separated from replication sites (Vieira *et al.*, 2004).

The formation and maintenance of an RNA/DNA hybrid has several potential consequences and is also influenced in terms of transcriptional regulatory elements. First evidence for an association between R-loops and transcription was seen 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 resolves 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-loop formation has been positively correlated with transcriptional regulation (Tracy and Lieber, 2000), possibly by the RNA function *in cis* with the DNA from which it was transcribed, whereas evidence also suggests that in yeast the Rad51 can help RNA in trans to form R-loops (Wahba *et al.*, 2013). In addition, mutations of RNase H, which digests the RNA in RNA/DNA hybrids, RNA helicases and Topoisomerases are associated with accumulation of R-loops, indicating that these enzymes function to remove these structures from the genome (Tuduri *et al.*, 2009; Wahba *et al.*, 2011). In particular, hydrolysis of RNA from RNA/DNA hybrid by the RNase H1 enzyme represses both replication stress and genomic instability (Tuduri *et al.*, 2009).

In this work, overexpression of oncogenic $HRAS^{V12}$ induces RNA synthesis and R-loop formation which can be a major cause for replication fork slowing and DNA breakage (Aguilera and Garcia-Muse, 2012). Our data may complement the previous studies, suggesting that deregulated transcription and R-loop-mediated genomic instability is a feature of oncogenes. Therefore, R-loops can add a new layer of gene expression regulation and together with other studies give a new perspective on the various R-loop cellular functions, suggesting that they might be common in other cancer types.

Increased transcription triggers R-loop associated DNA damage and replication stress

So far, the mechanisms involved in endogenous replication stress were deregulation of reactive oxygen species and cell cycle, namely premature or increased replication initiation, causing depletion of nucleotides or replication enzymes (Bester *et al*., 2011). Oncogene-induced replication stress is an important source of chromosomal instability in cancer cells, but the underlying molecular mechanisms are still poorly understood (Zeman and Cimprich, 2014; Gaillard *et al.*, 2015). This study provides the first molecular link between oncogene-induced and up-regulation of the transcription machinery and replication stress, presenting a new pathway of genomic instability in cancer.

Accumulation of R-loops can promote abnormal transcription of mRNA and rRNA genes in *E. coli* and *S. cerevisiae* (Tous and Aguilera, 2007; El Hage *et al.*, 2010). Later studies have associated the aberrant transcription phenotypes of cells lacking essential 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 encouraged future work investigating the close connection between transcription and genome integrity (Huertas and Aguilera, 2003). Later evidence also emphasised the potential for functional interplay between R-loops and transcription in different genome-wide studies, which have found a strong enrichment of R-loops at RNA Pol II gene promoters in human and murine cells (Ginno *et al.*, 2012; Chen *et al.*, 2015).

R-loops were shown to form around the bacterial ColE1 origin of replication *in vitro* (Itoh and Tomizawa, 1980). Moreover, it was suggested that R-loops help to unwind the DNA, making it more accessible for the initiation of replication (Baker and Kornberg, 1988) and can also form at the origin of replication in mitochondria of *S. cerevisiae* and mammals (Xu and Clayton, 1996; Brown *et al.*, 2008). 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 presented abnormal replication fork progression (Maduike *et al.*, 2014). However, it has not been investigated if genuine R-loops form during eukaryotic DNA replication.

During replication fork, the genetic material is duplicated and transcription-replication complexes are interdependent and meet along the template with RNA and DNA polymerases that can eventually result in a genetic clash, leading to genomic instability in cancer (Aguilera and Garcia-Muse, 2012). Nevertheless, DNA damage is also present in non-replicating cells (Sordet *et al.*, 2009) most likely due to the exposure of single-stranded DNA of R-loops, which being more unstable is further susceptible to lesions and to mutagenic molecules (Skourti-Stathaki and Proudfoot, 2014). Therefore, it is reasonable to consider that consequences of transcription-replication encounters can be a major source of R-loop-associated DNA damage (Helmrich *et al.*, 2011). As transcription machinery is often deregulated in cancer cells (Bywater *et al.*, 2013), our data suggests that the intricate crosstalk between transcription-associated replication

stress and accumulation of R-loops could be another pathway to promote genomic instability in cancer. There are still multiple directions to explore about the involvement of R-loops in cancer Biology. Characterising the molecular mechanisms involved in RNA synthesis, increased R-loop levels and expression of RNase H1 associated with replication stress, may help to predict and detect new candidates for markers in cancer cells and tissues.

6.2 Future directions

6.2.1 Interplay between dsRNA and heterochromatin

This study has provided a mechanistic link between the formation of dsRNA, enrichment of repressive heterochromatin and transcriptional defects in a human repeat expansion disease, Huntington's disease. However, several questions remain to be answered. It will be important to elucidate the role of histone deacetylase inhibitors (HDACis) in transcriptional reactivation of expanded alleles in HD. HDAC is that derepress transcription by increasing histone acetylation and making the DNA available to the transcriptional machinery are potential therapies in HD models, as noted earlier (Guan *et al.*, 2009; Fischer *et al.*, 2010). Promisingly, the first report of HDACi in phase I clinical trial was associated with sodium phenylbutyrate and led to correction of mRNA levels (Hogarth *et al.*, 2007), but the use of these broad-spectrum HDACis resulted in undesirable side effects, mainly related to the complexity of HDAC family and the uncertainty about the rate-limiting enzymes critical for the HD pathology. The search for newer and more effective HDACis is under development for the treatment of HD, but it is apparent that they preferentially hyperacetylate loci already transcriptionally active, instead of reactivate silent genomic locations (Lopez-Atalaya *et al.*, 2013). Inhibitors of HDAC exhibit neuroprotective and neuroregenerative properties, as such, they have

potential to become a promising therapeutic strategy, although some studies have shown that inhibition of HDACs can lead to only partial transcriptional reactivation of expanded alleles in polyglutamine expansion disorders (Coffee et al., 1999). Therefore, it is tempting to speculate that this may be associated with accumulation of dsRNA conformation during HDAC inhibition, which could reinforce the defect in transcription initiation and elongation. It would be interesting to treat lymphoblast and fibroblast cells with HDAC inhibitors, followed by J2 dsRNA IP and RT-qPCR analyses to test their effects on gene expression profile.

In addition to the classical histone modifications associated with transcriptional repression (H3K9me2), the levels of H3K9me3 were shown to be elevated at expanded CAG repeats, suggesting that this histone methylation may be important for *HTT* reduction. Although H3K9me3 levels and the corresponding SETDB1 methyltransferase were previously shown to be elevated in *post-mortem* brains of HD patients and in HD transgenic animals (Ryu *et al.*, 2006), future studies need to address how this histone mark is required for HD silencing. One possible mechanism *in trans* is that CBP acts as a transcriptional cofactor and supresses SETDB1 gene expression, conserving the normal level of H3K9me3 (Bannister and Kouzarides, 1996). Moreover, it seems that gene expression in HD can be directly related to H3K9me3-mediated chromatin remodelling and heterochromatin domains, leading to gene silencing (Lee, Hwang, Kim, *et al.*, 2013; Lee, Hwang, Shin, *et al.*, 2013). Further studies can investigate this mechanism by depleting H3K9me3 SETDB1 methyltransferase in combination with J2 dsRNA IP and RT-qPCR analysis of *HTT* gene expression. Furthermore, the potential recruitment of SETDB1 by dsRNA formed at expanded CAG may be evaluated by modifying dsRNA levels using RNase V1 and LNA treatments. SETDB1 recruitment could potentially also contribute to *HTT* repression in HD. Initial steps would consider if subunits of SETDB1

are enriched at expanded CAG repeats, with experiments investigating the interplay with dsRNA and other histone modifications such as histone acetylation.

6.2.2 Expansion of the CAG locus associated with chromatin and transcription

Intergenerational instability of the CAG repeat is a common feature in HD patients (>80% cases) (Duyao *et al.*, 1993), although in mice this instability increased with larger CAG repeat length (Wheeler *et al.*, 1999). Instability of the CAG expansion can lead to either its expansion or contraction. The factors that confer CAG repeat instability are still unclear. The formation of dsRNA at trinucleotide repeats in HD may have implications for our understanding of the CAG instability in HD.

It has been demonstrated that hairpin DNA structures formed during replication could underlie intergenerational CAG repeat instability (Liu and Leffak, 2012). DNA damage is seen in HD, most likely partly due to increased free radical generation and oxidative stress, and requires repair in differentiated tissues (Johri and Beal, 2012). Additionally, mHTT can interfere with DNA repair proteins, impairing their function of repairing double-strand breaks, increasing DNA damage in the neuron (Enokido *et al.*, 2010). In fact, a genome-wide association study demonstrated that the knockout of Msh3, involved in the DNA mismatch repair pathway, reduces somatic expansion and instability in HD mouse models and could be a measure of disease progression (Hensman Moss *et al.*, 2017). Importantly, it is possible that the DNA damage is resistant to repair due to secondary structures formed by expanded CAG repeats (Goula *et al.*, 2009), further aggravating the problem.

In addition, chromatin structure can affect CAG repeat instability because a more relaxed conformation is more accessible to the DNA repair factors (Goula *et al.*, 2012) and specific HDACs in human cells can enhance the expression of trinucleotide repeats

(Lahue and Frizzell, 2012). However, the interplay between trinucleotide repeats and DNA repair factors may be more complex and dependent on cellular environment, since mismatch repair factors have been shown to protect against intergenerational instability in mice (Ezzatizadeh *et al.*, 2012). The identification of proteins involved in DNA damage repair in the dsRNA interactome may disclose new factors with functions in repeat instability. Therefore, it will be important to comprehensively investigate the formation of dsRNA and their interaction with repair factors in different cell types. Transcription elongation can also promote CAG repeat instability and mediate tissue-specific somatic expansion (Goula *et al.*, 2012). A possible study could involve simultaneous proteomics-based quantitation of repair factor protein expression and analysis of repeat instability at different time points during differentiation of HD iPSCs to mature neurons.

This will be a major contribution to comprehensively investigate the role of dsRNA structures in trinucleotide repeat instability. Since increased CAG repeat length affects the pathology in HD, it is essential to explore their contribution in the production of transcripts and proteins. Intriguingly, when the expansions increase to more than 200 CAGs, toxicity is abolished, which might be related with reduced *HTT* RNA and protein levels (Dragatsis *et al.*, 2009; Morton *et al.*, 2009). It is possible that the presence of super-long CAG repeat expansions may have a paradoxical "protective" effect, since by reducing transcription of the mutant DNA, they may reduce the amount of toxic mutant protein, and delay the onset of disease. Future studies will need to address whether dsRNA contributes to alterations in transcription in the cell type specificity and extent of repeat instability *in vivo*.

6.2.3 Dynamics of the dsRNA interactome

In relation to the dsRNA interactors, it will be important to comprehensively describe the composition of dsRNA interactome in different cell types and tissues. Additional optimisation of the dsRNA IP method used in this work may extend to the proteome, perhaps disclosing low abundant proteins that were not identified due to technical limitations. As protein factors may play an important role in transcriptional silencing and repeat instability in HD, the investigation of dsRNA proteomes in clinically relevant tissues may help to clarify the molecular mechanisms of repeat expansion diseases. In particular, it will be interesting to investigate the HTT dsRNA interactors during development from embryonic stem cells to mature neurons, which could offer new insights into the association and Biology of repair factors or repressive chromatin complexes with dsRNA. However, since HD is monogenic, it may be crucial to improve the sensitivity of the dsRNA IP method or to develop a novel approach to particularly purify dsRNA and associated proteins from a single gene locus.

Gene reactivation by restoring histone acetylation in HD may be only partial due to the accumulation of dsRNA which could recruit alternative repressive factors to the genes. Thus, it will be important to target the primary trigger of transcriptional repression, including secondary structures such as dsRNA. Therefore, the present work may have implications for the future development of therapies for repeat expansion disorders.

6.2.4 R-loop as potential therapies to cancer disease

Our data demonstrated that cancer is a disease associated with genomic instability and R-loops. The DNA-RNA interaction, particularly R-loop conformation offers a unique target for small molecules, making them potential therapeutic structures (Shaw and Arya, 2008). It has been shown that various ligands can target R-loops, including

ethidium bromide, the polyamides distamycin and netropsin and the aminoglycosides neomycin and paramomycin (Shaw and Arya, 2008). These molecules show high binding affinities to R-loops, although some of them also recognise dsDNA and RNA and are mutagenic, limiting their potential Biological applications (Shaw and Arya, 2008). Still, evidence suggests that combining the properties of these ligands can achieve sub-nanomolar affinity for RNA/DNA hybrids (Shaw and Arya, 2008). In particular, complexes linking aminoglycosides to derivatives of ethidium bromide, provide a possible strategy for the development of specific RNA/DNA hybrid ligands in prospect drug design approaches. It has been shown that efficient folding of nascent RNA into secondary structures reduces R-loop formation and mutagenesis in *S. cerevisiae* (Chen *et al.*, 2016). As a result, designing inhibitors that stabilise the folding of repeat-associated RNA may be a method to decrease R-loop levels with less target effects.

As discussed previously for dsRNA-binding proteins, for R-loop therapies an alternate strategy would be to modify functions of proteins implicated in R-loop Biology. Inhibition of Top1 has been commonly used as a tool to induce R-loops in various experimental systems (Sordet *et al.*, 2009). In particular, R-loop formation is associated with widespread DNA damage in these studies, indicating that other strategies targeting R-loops may be important.

The RNA/DNA hybrid-interacting proteins can also be modulated to affect R-loop levels in diseases with a prevalent dysregulation of R-loops. The design of therapies for proteins requires control for off target effects as they may have several Biological functions. As the role of R-loops in human pathologies is still not fully elucidated, it is reasonable that this approach may be used in cancer. In cancer, transcription, DNA replication and DNA repair are dysregulated on a genome-wide scale (Hanahan and Weinberg, 2011). In line with this, it has been demonstrated that SETX/Sen1 can resolve

R-loops and is able to maintain genome integrity by coordinating transcription, DNA replication and the DNA damage response (Alzu *et al.*, 2012). SETX degrades the 3'-5' RNA degradation complex, the exosome, to sites of transcription-induced DNA damage (Richard *et al.*, 2013). The enrichment of helicases in human cancers suggests that these proteins may contribute to tumorigenesis and are associated with RNA/DNA hybrid accumulation. Another enzyme, Top1, suppresses genomic instability in mammalian cells by preventing a conflict between transcription and DNA replication (Tuduri *et al.*, 2009). Indeed, the Top1 inhibitor camptothecin is used as chemotherapeutic drugs (Capranico *et al.*, 2010) and its cellular toxicity is dependent on increased R-loop-mediated DNA damage (Sordet *et al.*, 2009; Tuduri *et al.*, 2009). Thus, it will be interesting to investigate the effects of R-loops in tumorigenesis, in different cancer cells, as they are a major source of replication stress and DNA damage and may potentially contribute to the increased cancer predisposition of individuals with cancer (Garcia-Rubio *et al*., 2015; Schwab *et al*., 2015).

Concluding remarks

Aggregation of the mHTT and its toxic functions has been shown to be the main molecular cause of HD. However, currently there is evidence that *HTT* RNA contributes to the toxicity and this can be extended to other types of polyglutamine diseases, which likely share Biological and pathogenic features. The presence of the CAG expansion can lead to the formation of dsRNA, resulting in dysregulation of transcription-associated processes in disease pathology. Here, this research revealed an unanticipated link between dsRNA and heterochromatin formation at expanded CAG repeats in mammalian *HTT* gene *in cis*. It is apparent that dsRNA recruits ENO1 transcriptional repressor, promoting the formation of repressive H3K9me2, resulting in transcriptional dysregulation over the CAG repeats. In particular, an H3K9me2 repressive mark is

formed and heterochromatin protein G9a is recruited, which reinforces Pol II pausing before efficient transcriptional activity, providing a molecular link between dsRNA and the pathology of repeat expansion disorders. This work will contribute to our understanding of how expansion mutations cause pathogenic molecular events on the *HTT* gene regulation. Future studies investigating the *HTT* gene are required and will need to be considered in the design of future therapies.

This work also elucidates the importance of R-loops in cancer, as their accumulation can be a major cause for DNA breakage and genomic instability. Increased RNA synthesis, transcription activity and accumulation of R-loop levels could be correlated with replication stress in cancer. There are still multiple directions to explore with respect to the involvement of R-loops in cancer Biology. It would be interesting to know whether R-loops have a direct effect on cancer progression. Moreover, further characterising the molecular mechanisms triggered by increased R-loop levels and expression of RNase H1 associated with replication stress, may help to predict and detect new candidates for markers in cancer cells and tissues. Nascent RNA synthesis can only be measured in live cells but increased expression of transcription factors and RNA polymerase subunits, as well as accumulation of R-loops and expression of RNases H1 or H2, might be promising candidates for markers of transcription-associated replication stress.

Overall, the role of dsRNA or RNA/DNA hybrids in gene expression regulation, provides a molecular link evolving the landscape of nucleic acid structures as transcriptional and epigenetic regulators in health and disease.

References

Adelman, K.,J. T. Lis. 2012. Promoter-proximal pausing of RNA polymerase II: emerging roles in metazoans. *Nature Reviews Genetics*, 13, 720-731.

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

Akman, G., R. Desai, L. J. Bailey, T. Yasukawa, I. Dalla Rosa, R. Durigon, J. B. Holmes, C. F. Moss, M. Mennuni, H. Houlden, R. J. Crouch, M. G. Hanna, R. D. Pitceathly, A. Spinazzola,I. J. Holt. 2016. Pathological ribonuclease H1 causes R-loop depletion and aberrant DNA segregation in mitochondria. *Proceedings of the National Academy of Sciences*, 113, E4276-4285.

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

Andresen, J. M., J. Gayan, L. Djousse, S. Roberts, D. Brocklebank, S. S. Cherny, L. R. Cardon, J. F. Gusella, M. E. MacDonald, R. H. Myers, D. E. Housman, N. S. Wexler,US-Venezuela Collaborative Res Grp. 2007. The relationship between CAG repeat length and age of onset differs for Huntington's disease patients with juvenile onset or adult onset. *Annals of Human Genetics*, 71, 295-301.

Andrew, S. E., Y. P. Goldberg, B. Kremer, F. Squitieri, J. Theilmann, J. Zeisler, H. Telenius, S. Adam, E. Almquist, M. Anvret, G. Lucotte, A. J. Stoess, G. Campanella,M. R. Hayden. 1994. Huntington Disease without Cag Expansion - Phenocopies or Errors in Assignment. *American Journal of Human Genetics*, 54, 852-863.

Andrew, S. E., Y. P. Goldberg, B. Kremer, H. Telenius, J. Theilmann, S. Adam, E. Starr, F. Squitieri, B. Y. Lin, M. A. Kalchman, R. K. Graham,M. R. Hayden. 1993. The Relationship between Trinucleotide (Cag) Repeat Length and Clinical-Features of Huntingtons-Disease. *Nature Genetics*, 4, 398-403.

Arber, C., S. V. Precious, S. Cambray, J. R. Risner-Janiczek, C. Kelly, Z. Noakes, M. Fjodorova, A. Heuer, M. A. Ungless, T. A. Rodriguez, A. E. Rosser, S. B. Dunnett,M. Li. 2015. Activin A directs striatal projection neuron differentiation of human pluripotent stem cells. *Development*, 142, 1375-1386.

Arrasate, M., S. Mitra, E. S. Schweitzer, M. R. Segal,S. Finkbeiner. 2004. Inclusion body formation reduces levels of mutant huntingtin and the risk of neuronal death. *Nature*, 431, 805-810.

Auerbach, W., M. S. Hurlbert, P. Hilditch-Maguire, Y. Z. Wadghiri, V. C. Wheeler, S. I. Cohen, A. L. Joyner, M. E. MacDonald,D. H. Turnbull. 2001. The HD mutation causes progressive lethal neurological disease in mice expressing reduced levels of huntingtin. *Human Molecular Genetics*, 10, 2515-2523.

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

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

Balachandran, S., P. C. Roberts, L. E. Brown, H. Truong, A. K. Pattnaik, D. R. Archer,G. N. Barber. 2000. Essential role for the dsRNA-dependent protein kinase PKR in innate immunity to viral infection. *Immunity*, 13, 129-141.

Bando, Y., R. Onuki, T. Katayama, T. Manabe, T. Kudo, K. Taira,M. Tohyama. 2005. Double-strand RNA dependent protein kinase (PKR) is involved in the extrastriatal degeneration in Parkinson's disease and Huntington's disease. *Neurochemistry International*, 46, 11-18.

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

Bannister, A. J.,T. Kouzarides. 1996. The CBP co-activator is a histone acetyltransferase. *Nature*, 384, 641-643.

Baralle, M., T. Pastor, E. Bussani,F. Pagani. 2008. Influence of Friedreich ataxia GAA noncoding repeat expansions on pre-mRNA processing. *American Journal of Human Genetics*, 83, 77-88.

Bassett, A., S. Cooper, C. Wu,A. Travers. 2009. The folding and unfolding of eukaryotic chromatin. *Current Opinion in Genetics & Development*, 19, 159-165.

Bates, G. P., R. Dorsey, J. F. Gusella, M. R. Hayden, C. Kay, B. R. Leavitt, M. Nance, C. A. Ross, R. I. Scahill, R. Wetzel, E. J. Wild,S. J. Tabrizi. 2015. Huntington disease. *Nature Reviews Disease Primers*, 1, 15005.

Begeti, F., L. C. Schwab, S. L. Mason,R. A. Barker. 2016. Hippocampal dysfunction defines disease onset in Huntington's disease. *Journal of Neurology, Neurosurgery, and Psychiatry*, 87, 975-981.

Behrens, S. E., L. Tomei,R. De Francesco. 1996. Identification and properties of the RNA-dependent RNA polymerase of hepatitis C virus. *EMBO JOURNALournal*, 15, 12-22.

Benn, C. L., T. Sun, G. Sadri-Vakili, K. N. McFarland, D. P. DiRocco, G. J. Yohrling, T. W. Clark, B. Bouzou,J. H. Cha. 2008. Huntingtin modulates transcription, occupies gene promoters in vivo, and binds directly to DNA in a polyglutamine-dependent manner. *Journal of Neuroscience*, 28, 10720-10733.

Bennett, C. F.,E. E. Swayze. 2010. RNA Targeting Therapeutics: Molecular Mechanisms of Antisense Oligonucleotides as a Therapeutic Platform. *Annual Review of Pharmacology and Toxicology*, 50, 259-293.

Bensaude, O. 2011. Inhibiting eukaryotic transcription: Which compound to choose? How to evaluate its activity? *Transcription*, 2, 103-108.

Berg, M. G., L. N. Singh, I. Younis, Q. Liu, A. M. Pinto, D. Kaida, Z. Zhang, S. Cho, S. Sherrill-Mix, L. Wan,G. Dreyfuss. 2012. U1 snRNP determines mRNA length and regulates isoform expression. *Cell*, 150, 53-64.

Bernstein, E., A. A. Caudy, S. M. Hammond,G. J. Hannon. 2001. Role for a bidentate ribonuclease in the initiation step of RNA interference. *Nature*, 409, 363-366.

Bester, A. C., M. Roniger, Y. S. Oren, M. M. Im, D. Sarni, M. Chaoat, A. Bensimon, G. Zamir, D. S. Shewach,B. Kerem. 2011. Nucleotide deficiency promotes genomic instability in early stages of cancer development. *Cell*, 145, 435-446.

Bhatia, V., S. I. Barroso, M. L. Garcia-Rubio, E. Tumini, E. Herrera-Moyano,A. Aguilera. 2014. BRCA2 prevents R-loop accumulation and associates with TREX-2 mRNA export factor PCID2. *Nature*, 511, 362-365.

Biagioli, M., F. Ferrari, E. M. Mendenhall, Y. Zhang, S. Erdin, R. Vijayvargia, S. M. Vallabh, N. Solomos, P. Manavalan, A. Ragavendran, F. Ozsolak, J. M. Lee, M. E. Talkowski, J. F. Gusella, M. E. Macdonald, P. J. Park,I. S. Seong. 2015. Htt CAG repeat expansion confers pleiotropic gains of mutant huntingtin function in chromatin regulation. *Human Molecular Genetics*, 24, 2442-2457.

Bidichandani, S. I., T. Ashizawa,P. I. Patel. 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.

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

Bondy-Chorney, E., T. E. C. Parks, A. Ravel-Chapuis, B. J. Jasmin,J. Cote. 2016. Staufen1s role as a splicing factor and a disease modifier in Myotonic Dystrophy Type I. *Rare Diseases*, 4.

Bondy-Chorney, E., T. E. C. Parks, A. Ravel-Chapuis, R. Klinck, L. Rocheleau, M. Pelchat, B. Chabot, B. J. Jasmin,J. Cote. 2016. Staufen1 Regulates Multiple Alternative Splicing Events either Positively or Negatively in DM1 Indicating Its Role as a Disease Modifier. *Plos Genetics*, 12.

Bonin, M., J. Oberstrass, N. Lukacs, K. Ewert, E. Oesterschulze, R. Kassing,W. Nellen. 2000. Determination of preferential binding sites for anti-dsRNA antibodies on double-stranded RNA by scanning force microscopy. *RNA*, 6, 563-570.

Boule, J. B.,V. A. Zakian. 2007. The yeast Pif1p DNA helicase preferentially unwinds RNA DNA substrates. *Nucleic Acids Research*, 35, 5809-5818.

Bowles, K. R., S. P. Brooks, S. B. Dunnett,L. Jones. 2012. Gene expression and behaviour in mouse models of HD. *Brain Research Bulletin*, 88, 276-284.

Braasch, D. A.,D. R. Corey. 2001. Locked nucleic acid (LNA): fine-tuning the recognition of DNA and RNA. *Chemistry and Biology*, 8, 1-7.

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

Bradley, D. W., Q. P. Dou, J. L. Fridovich-Keil,A. B. Pardee. 1990. Transformed and nontransformed cells differ in stability and cell cycle regulation of a binding activity to the murine thymidine kinase promoter. *Proceedings of the National Academy of Sciences*, 87, 9310-9314.

Brayson, D.,C. M. Shanahan. 2017. Current insights into LMNA cardiomyopathies: Existing models and missing LINCs. *Nucleus*, 8, 17-33.

Brook, J. D., M. E. McCurrach, H. G. Harley, A. J. Buckler, D. Church, H. Aburatani, K. Hunter, V. P. Stanton, J. P. Thirion, T. Hudson,et al. 1992. Molecular basis of myotonic dystrophy: expansion of a trinucleotide (CTG) repeat at the 3' end of a transcript encoding a protein kinase family member. *Cell*, 68, 799-808.

Brouwer, J. R., A. Huguet, A. Nicole, A. Munnich,G. Gourdon. 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., A. N. Tkachuk,D. A. Clayton. 2008. Native R-loops persist throughout the mouse mitochondrial DNA genome. *Journal of Biological Chemistry*, 283, 36743-36751.

Budworth, H.,C. T. McMurray. 2013. A brief history of triplet repeat diseases. *Methods Molecular Biology*, 1010, 3-17.

Burgess, H. M.,I. Mohr. 2015. Cellular 5'-3' mRNA exonuclease Xrn1 controls double-stranded RNA accumulation and anti-viral responses. *Cell Host Microbe*, 17, 332-344.

Bywater, M. J., R. B. Pearson, G. A. McArthur,R. D. Hannan. 2013. Dysregulation of the basal RNA polymerase transcription apparatus in cancer. *Nature Reviews Cancer*, 13, 299-314.

Campuzano, V., L. Montermini, M. D. Molto, L. Pianese, M. Cossee, F. Cavalcanti, E. Monros, F. Rodius, F. Duclos, A. Monticelli, F. Zara, J. Canizares, H. Koutnikova, S. I. Bidichandani, C. Gellera, A. Brice, P. Trouillas, G. De Michele, A. Filla, R. De Frutos, F. Palau, P. I. Patel, S. Di Donato, J. L. Mandel, S. Cocozza, M. Koenig,M. Pandolfo. 1996. Friedreich's ataxia: autosomal recessive disease caused by an intronic GAA triplet repeat expansion. *Science*, 271, 1423-1427.

Capranico, G., J. Marinello,L. Baranello. 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.

Carroll, J. B., S. C. Warby, A. L. Southwell, C. N. Doty, S. Greenlee, N. Skotte, G. Hung, C. F. Bennett, S. M. Freier,M. R. Hayden. 2011. Potent and selective antisense oligonucleotides targeting single-nucleotide polymorphisms in the Huntington disease gene / allele-specific silencing of mutant huntingtin. *Molecular Therapy*, 19, 2178-2185.

Casanova, E., A. Alonso-Llamazares, D. Zamanillo, C. Garate, P. Calvo,M. A. Chinchetru. 1996. Identification of a long huntingtin mRNA transcript in mouse brain. *Brain Research*, 743, 320-323.

Castel, A. L., J. D. Cleary,C. E. Pearson. 2010. Repeat instability as the basis for human diseases and as a potential target for therapy. *Nature Reviews Molecular Cell Biology*, 11, 165-170.

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

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

Cerritelli, S. M.,R. J. Crouch. 2009. Ribonuclease H: the enzymes in eukaryotes. *FEBS Journal*, 276, 1494-1505.

Cha, J. H. 2007. Transcriptional signatures in Huntington's disease. *Progress in NeuroBiology*, 83, 228-248.

Chen, J. L., D. M. VanEtten, M. A. Fountain, I. Yildirim,M. D. Disney. 2017. Structure and Dynamics of RNA Repeat Expansions That Cause Huntington's Disease and Myotonic Dystrophy Type 1. *Biochemistry*, 56, 3463-3474.

Chen, P. B., H. V. Chen, D. Acharya, O. J. Rando,T. G. Fazzio. 2015. R loops regulate promoter-proximal chromatin architecture and cellular differentiation. *Nature Structural and Molecular Biology*, 22, 999-1007.

Chen, X., J. R. Yang,J. Zhang. 2016. Nascent RNA folding mitigates transcription-associated mutagenesis. *Genome Research*, 26, 50-59.

Chernikova, S. B., O. V. Razorenova, J. P. Higgins, B. J. Sishc, M. Nicolau, J. A. Dorth, D. A. Chernikova, S. Kwok, J. D. Brooks, S. M. Bailey, J. C. Game,J. M. Brown. 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., C. P. Thienes, S. E. Mahoney, E. Analau, G. N. Filippova,S. J. Tapscott. 2005. Antisense transcription and heterochromatin at the DM1 CTG repeats are constrained by CTCF. *Molecular Cell*, 20, 483-489.

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

Ciancarelli, I., D. De Amicis, C. Di Massimo, C. Di Scanno, C. Pistarini, N. D'Orazio,M. G. T. Ciancarelli. 2014. Peripheral biomarkers of oxidative stress and their limited potential in evaluation of clinical features of Huntington's patients. *Biomarkers*, 19, 452-456.

Cicchetti, F., S. Lacroix, G. Cisbani, N. Vallieres, M. Saint-Pierre, I. St-Amour, R. Tolouei, J. N. Skepper, R. A. Hauser, D. Mantovani, R. A. Barker,T. B. Freeman. 2014. Mutant Huntingtin is Present in Neuronal Grafts in Huntington Disease Patients. *Annals of Neurology*, 76, 31-42.

Clemens, M. J.,A. Elia. 1997. The double-stranded RNA-dependent protein kinase PKR: structure and function. *Journal Interferon Cytokine Research*, 17, 503-524.

Coffee, B., F. Zhang, S. Ceman, S. T. Warren,D. Reines. 2002. Histone modifications depict an aberrantly heterochromatinized FMR1 gene in fragile x syndrome. *Am Journal of Human Genetics*, 71, 923-932.

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

Coleman, M. L., C. J. Marshall,M. F. Olson. 2004. RAS and RHO GTPases in G1-phase cell-cycle regulation. *Nature Reviews Molecular Cell Biology*, 5, 355-366.

Consortium, HD iPSC. 2012. Induced Pluripotent Stem Cells from Patients with Huntington's Disease Show CAG-Repeat-Expansion-Associated Phenotypes. *Cell Stem Cell*, 11, 264-278.

Coolidge, C. J.,J. G. Patton. 2000. A new double-stranded RNA-binding protein that interacts with PKR. *Nucleic Acids Research*, 28, 1407-1417.

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

Cruz, C.,J. Houseley. 2014. Endogenous RNA interference is driven by copy number. *Elife*, 3.

Davies, S. W., M. Turmaine, B. A. Cozens, M. DiFiglia, A. H. Sharp, C. A. Ross, E. Scherzinger, E. E. Wanker, L. Mangiarini,G. P. Bates. 1997. Formation of neuronal intranuclear inclusions underlies the neurological dysfunction in mice transgenic for the HD mutation. *Cell*, 90, 537-548.

Davis, T. A., B. Loos,A. M. Engelbrecht. 2014. AHNAK: The giant jack of all trades. *Cellular Signalling*, 26, 2683-2693.

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

de Mezer, M., M. Wojciechowska, M. Napierala, K. Sobczak,W. J. Krzyzosiak. 2011. Mutant CAG repeats of Huntingtin transcript fold into hairpins, form nuclear foci and are targets for RNA interference. *Nucleic Acids Research*, 39, 3852-3863.

De Vita, G., L. Bauer, V. M. da Costa, M. De Felice, M. G. Baratta, M. De Menna,R. Di Lauro. 2005. Dose-dependent inhibition of thyroid differentiation by RAS oncogenes. *Molecular Endocrinology*, 19, 76-89.

DeJesus-Hernandez, M., I. R. Mackenzie, B. F. Boeve, A. L. Boxer, M. Baker, N. J. Rutherford, A. M. Nicholson, N. A. Finch, H. Flynn, J. Adamson, N. Kouri, A. Wojtas, P. Sengdy, G. Y. Hsiung, A. Karydas, W. W. Seeley, K. A. Josephs, G. Coppola, D. H. Geschwind, Z. K. Wszolek, H. Feldman, D. S. Knopman, R. C. Petersen, B. L. Miller, D. W. Dickson, K. B. Boylan, N. R. Graff-Radford,R. Rademakers. 2011. Expanded GGGGCC hexanucleotide repeat in noncoding region of C9ORF72 causes chromosome 9p-linked FTD and ALS. *Neuron*, 72, 245-256.

Dereli-Oz, A., G. Versini,T. D. Halazonetis. 2011. Studies of genomic copy number changes in human cancers reveal signatures of DNA replication stress. *Molecular Oncology*, 5, 308-314.

Di Micco, R., M. Fumagalli, A. Cicalese, S. Piccinin, P. Gasparini, C. Luise, C. Schurra, M. Garre, P. G. Nuciforo, A. Bensimon, R. Maestro, P. G. Pelicci,F. d'Adda di Fagagna. 2006. Oncogene-induced senescence is a DNA damage response triggered by DNA hyper-replication. *Nature*, 444, 638-642.

Di Pardo, A., E. Amico, A. Basit, A. Armirotti, P. Joshi, D. M. Neely, R. Vuono, S. Castaldo, A. F. Digilio, F. Scalabri, G. Pepe, F. Elifani, M. Madonna, S. K. Jeong, B. M. Park, M. D'Esposito, A. B. Bowman, R. A. Barker,V. Maglione. 2017. Defective Sphingosine-1-phosphate metabolism is a druggable target in Huntington's disease. *Science Reports*, 7, 5280.

DiFiglia, M., E. Sapp, K. O. Chase, S. W. Davies, G. P. Bates, J. P. Vonsattel,N. Aronin. 1997. Aggregation of huntingtin in neuronal intranuclear inclusions and dystrophic neurites in brain. *Science*, 277, 1990-1993.

Dinger, M. E., K. C. Pang, T. R. Mercer, M. L. Crowe, S. M. Grimmond,J. S. Mattick. 2009. NRED: a database of long noncoding RNA expression. *Nucleic Acids Research*, 37, D122-126.

Dion, V.,J. H. Wilson. 2009. Instability and chromatin structure of expanded trinucleotide repeats. *Trends Genetics*, 25, 288-297.

Dobbelstein, M.,C. S. Sorensen. 2015. Exploiting replicative stress to treat cancer. *Nature Reviews Drug Discovery*, 14, 405-423.

Dominguez-Sola, D., C. Y. Ying, C. Grandori, L. Ruggiero, B. Chen, M. Li, D. A. Galloway, W. Gu, J. Gautier,R. Dalla-Favera. 2007. Non-transcriptional control of DNA replication by c-Myc. *Nature*, 448, 445-451.

Dorsey, E. R.,Huntington Study Grp COHORT Invest. 2012. Characterization of a Large Group of Individuals with Huntington Disease and Their Relatives Enrolled in the COHORT Study. *PLoS One*, 7.

Dragatsis, I., D. Goldowitz, N. Del Mar, Y. P. Deng, C. A. Meade, L. Liu, Z. Sun, P. Dietrich, J. Yue,A. Reiner. 2009. CAG repeat lengths > or =335 attenuate the phenotype in the R6/2 Huntington's disease transgenic mouse. *NeuroBiology of Disease*, 33, 315-330.

Dragatsis, I., M. S. Levine,S. Zeitlin. 2000. Inactivation of Hdh in the brain and testis results in progressive neurodegeneration and sterility in mice. *Nature Genetics*, 26, 300-306.

Drolet, M., P. Phoenix, R. Menzel, E. Masse, L. F. Liu,R. J. Crouch. 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.

Drouin-Ouellet, J., S. J. Sawiak, G. Cisbani, M. Lagace, W. L. Kuan, M. Saint-Pierre, R. J. Dury, W. Alata, I. St-Amour, S. L. Mason, F. Calon, S. Lacroix, P. A. Gowland, S. T. Francis, R. A. Barker,F. Cicchetti. 2015. Cerebrovascular and blood-brain barrier impairments in Huntington's disease: Potential implications for its pathophysiology. *Annals of Neurology*, 78, 160-177.

Dunah, A. W., H. Jeong, A. Griffin, Y. M. Kim, D. G. Standaert, S. M. Hersch, M. M. Mouradian, A. B. Young, N. Tanese,D. Krainc. 2002. Sp1 and TAFII130 transcriptional activity disrupted in early Huntington's disease. *Science*, 296, 2238-2243.

Durkin, S. G.,T. W. Glover. 2007. Chromosome fragile sites. *Annual Review of Genetics*, 41, 169-192.

Duyao, M., C. Ambrose, R. Myers, A. Novelletto, F. Persichetti, M. Frontali, S. Folstein, C. Ross, M. Franz, M. Abbott,et al. 1993. Trinucleotide repeat length instability and age of onset in Huntington's disease. *Nature Genetics*, 4, 387-392.

Duyao, M. P., A. B. Auerbach, A. Ryan, F. Persichetti, G. T. Barnes, S. M. McNeil, P. Ge, J. P. Vonsattel, J. F. Gusella, A. L. Joyner,et al. 1995. Inactivation of the mouse Huntington's disease gene homolog Hdh. *Science*, 269, 407-410.

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

Ehresmann, C., F. Baudin, M. Mougel, P. Romby, J. P. Ebel,B. Ehresmann. 1987. Probing the structure of RNAs in solution. *Nucleic Acids Research*, 15, 9109-9128.

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

Enokido, Y., T. Tamura, H. Ito, A. Arumughan, A. Komuro, H. Shiwaku, M. Sone, R. Foulle, H. Sawada, H. Ishiguro, T. Ono, M. Murata, I. Kanazawa, N. Tomilin, K. Tagawa, E. E. Wanker,H. Okazawa. 2010. Mutant huntingtin impairs Ku70-mediated DNA repair. *Journal of Cell Biology*, 189, 425-443.

Evans, S. J., I. Douglas, M. D. Rawlins, N. S. Wexler, S. J. Tabrizi,L. Smeeth. 2013. Prevalence of adult Huntington's disease in the UK based on diagnoses recorded in general practice records. *Journal of Neurology, Neurosurgery, and Psychiatry* 84, 1156-1160.

Evers, M. M., M. H. Schut, B. A. Pepers, M. Atalar, M. J. van Belzen, R. L. Faull, R. A. Roos,W. M. van Roon-Mom. 2015. Making (anti-) sense out of huntingtin levels in Huntington disease. *Molecular Neurodegeneration*, 10, 21.

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

Fischer, A., F. Sananbenesi, A. Mungenast,L. H. Tsai. 2010. Targeting the correct HDAC(s) to treat cognitive disorders. *Trends in Pharmacological Sciences*, 31, 605-617.

Fisher, E. R.,M. R. Hayden. 2014. Multisource Ascertainment of Huntington Disease in Canada: Prevalence and Population at Risk. *Movement Disorders*, 29, 105-114.

Fiszer, A., A. Mykowska,W. J. Krzyzosiak. 2011. Inhibition of mutant huntingtin expression by RNA duplex targeting expanded CAG repeats. *Nucleic Acids Research*, 39, 5578-5585.

French, S. 1992. Consequences of Replication Fork Movement through Transcription Units Invivo. *Science*, 258, 1362-1365.

Gagnon, K. T., H. M. Pendergraff, G. F. Deleavey, E. E. Swayze, P. Potier, J. Randolph, E. B. Roesch, J. Chattopadhyaya, M. J. Damha, C. F. Bennett, C. Montaillier, M. Lemaitre,D. R. Corey. 2010. Allele-selective inhibition of mutant huntingtin expression with antisense oligonucleotides targeting the expanded CAG repeat. *Biochemistry*, 49, 10166-10178.

Gaillard, H., T. Garcia-Muse,A. Aguilera. 2015. Replication stress and cancer. *Nature Reviews Cancer*, 15, 276-289.

Galka-Marciniak, P., M. O. Urbanek,W. J. Krzyzosiak. 2012. Triplet repeats in transcripts: structural insights into RNA toxicity. *Journal of Biological Chemistry*, 393, 1299-1315.

Gan, W., Z. Guan, J. Liu, T. Gui, K. Shen, J. L. Manley,X. Li. 2011. R-loop-mediated genomic instability is caused by impairment of replication fork progression. *Genes Development*, 25, 2041-2056.

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

Ginno, P. A., P. L. Lott, H. C. Christensen, I. Korf,F. Chedin. 2012. R-Loop Formation Is a Distinctive Characteristic of Unmethylated Human CpG Island Promoters. *Molecular Cell*, 45, 814-825.

Gipson, T. A., A. Neueder, N. S. Wexler, G. P. Bates,D. Housman. 2013. Aberrantly spliced HTT, a new player in Huntington's disease pathogenesis. *RNA Biology*, 10, 1647-1652.

Glajch, K. E.,G. Sadri-Vakili. 2015. Epigenetic Mechanisms Involved in Huntington's Disease Pathogenesis. *Journal of Huntingtons Disease*, 4, 1-15.

Gonitel, R., H. Moffitt, K. Sathasivam, B. Woodman, P. J. Detloff, R. L. Faull,G. P. Bates. 2008. DNA instability in postmitotic neurons. *Proceedings of the National Academy of Sciences*, 105, 3467-3472.

Gorgoulis, V. G., L. V. Vassiliou, P. Karakaidos, P. Zacharatos, A. Kotsinas, T. Liloglou, M. Venere, R. A. Ditullio, Jr., N. G. Kastrinakis, B. Levy, D. Kletsas, A. Yoneta, M. Herlyn, C. Kittas,T. D. Halazonetis. 2005. Activation of the DNA damage checkpoint and genomic instability in human precancerous lesions. *Nature*, 434, 907-913.

Goula, A. V., B. R. Berquist, D. M. Wilson, 3rd, V. C. Wheeler, Y. Trottier,K. Merienne. 2009. Stoichiometry of base excision repair proteins correlates with increased somatic CAG instability in striatum over cerebellum in Huntington's disease transgenic mice. *PLoS Genetics*, 5, e1000749.

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

Grabczyk, E., M. Mancuso,M. C. Sammarco. 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.,K. Usdin. 2000. 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 Research*, 28, 2815-2822.

Groh, M., M. M. Lufino, R. Wade-Martins,N. Gromak. 2014. R-loops associated with triplet repeat expansions promote gene silencing in Friedreich ataxia and fragile X syndrome. *PLoS Genetics*, 10, e1004318.

Groh, M., L. M. Silva,N. Gromak. 2014. Mechanisms of transcriptional dysregulation in repeat expansion disorders. *Biochemical Society Transactions*, 42, 1123-1128.

Guan, J. S., S. J. Haggarty, E. Giacometti, J. H. Dannenberg, N. Joseph, J. Gao, T. J. Nieland, Y. Zhou, X. Wang, R. Mazitschek, J. E. Bradner, R. A. DePinho, R. Jaenisch,L. H. Tsai. 2009. HDAC2 negatively regulates memory formation and synaptic plasticity. *Nature*, 459, 55-60.

Gusella, J. F., M. E. MacDonald,J. M. Lee. 2014. Genetic modifiers of Huntington's disease. *Movement Disorders*, 29, 1359-1365.

Haass, C.,D. J. Selkoe. 2007. Soluble protein oligomers in neurodegeneration: lessons from the Alzheimer's amyloid beta-peptide. *Nature Reviews Molecular Cell Biology*, 8, 101-112.

Hadzi, T. C., A. E. Hendricks, J. C. Latourelle, K. L. Lunetta, L. A. Cupples, T. Gillis, J. S. Mysore, J. F. Gusella, M. E. MacDonald, R. H. Myers,J. P. Vonsattel. 2012. Assessment of cortical and striatal involvement in 523 Huntington disease brains. *Neurology*, 79, 1708-1715.

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

Hagedorn, P. H., R. Persson, E. D. Funder, N. Albaek, S. L. Diemer, D. J. Hansen, M. R. Moller, N. Papargyri, H. Christiansen, B. R. Hansen, H. F. Hansen, M. A. Jensen,T. Koch. 2017. Locked nucleic acid: modality, diversity, and drug discovery. *Drug Discovery Today*.

Halazonetis, T., O. Zgheib, V. Gorgoulis,J. Bartek. 2008. An oncogene-induced DNA replication stress model for human cancer development. *Ejc Supplements*, 6, 63-64.

Hamperl, S.,K. A. Cimprich. 2014. The contribution of co-transcriptional RNA:DNA hybrid structures to DNA damage and genome instability. *DNA Repair (Amst)*, 19, 84-94.

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

Hatchi, E., K. Skourti-Stathaki, S. Ventz, L. Pinello, A. Yen, K. Kamieniarz-Gdula, S. Dimitrov, S. Pathania, K. M. McKinney, M. L. Eaton, M. Kellis, S. J. Hill, G. Parmigiani, N. J. Proudfoot,D. M. Livingston. 2015. BRCA1 recruitment to transcriptional pause sites is required for R-loop-driven DNA damage repair. *Molecular Cell*, 57, 636-647.

Hegde, M. V.,A. A. Saraph. 2011. Unstable genes unstable mind: beyond the central dogma of molecular Biology. *Medical Hypotheses*, 77, 165-170.

Helmlinger, D., S. Hardy, S. Sasorith, F. Klein, F. Robert, C. Weber, L. Miguet, N. Potier, A. Van-Dorsselaer, J. M. Wurtz, J. L. Mandel, L. Tora,D. Devys. 2004. Ataxin-7 is a subunit of GCN5 histone acetyltransferase-containing complexes. *Human Molecular Genetics*, 13, 1257-1265.

Helmrich, A., M. Ballarino,L. Tora. 2011. Collisions between Replication and Transcription Complexes Cause Common Fragile Site Instability at the Longest Human Genes. *Molecular Cell*, 44, 966-977.

Hensman Moss, D. J., A. F. Pardinas, D. Langbehn, K. Lo, B. R. Leavitt, R. Roos, A. Durr, S. Mead, Track-Hd investigators, Registry investigators, P. Holmans, L. Jones,S. J. Tabrizi. 2017. Identification of genetic variants associated with Huntington's disease progression: a genome-wide association study. *Lancet Neurology*, 16, 701-711.

Herrera-Moyano, E., X. Mergui, M. L. Garcia-Rubio, S. Barroso,A. Aguilera. 2014. The yeast and human FACT chromatin-reorganizing complexes solve R-loop-mediated transcription-replication conflicts. *Genes Development*, 28, 735-748.

Hodges, A., A. D. Strand, A. K. Aragaki, A. Kuhn, T. Sengstag, G. Hughes, L. A. Elliston, C. Hartog, D. R. Goldstein, D. Thu, Z. R. Hollingsworth, F. Collin, B. Synek, P. A. Holmans, A. B. Young, N. S. Wexler, M. Delorenzi, C. Kooperberg, S. J. Augood, R. L. Faull, J. M. Olson, L. Jones,R. Luthi-Carter. 2006. Regional and cellular gene expression changes in human Huntington's disease brain. *Human Molecular Genetics*, 15, 965-977.

Hogarth, P., L. Lovrecic,D. Krainc. 2007. Sodium phenylbutyrate in Huntington's disease: a dose-finding study. *Movement Disorders*, 22, 1962-1964.

Holley, R. W., J. H. Baldwin, J. A. Kiernan,T. O. Messmer. 1976. Control of growth of benzo(a)pyrene-transformed 3T3 cells. *Proceedings of the National Academy of Sciences*, 73, 3229-3232.

Hornstra, I. K., D. L. Nelson, S. T. Warren,T. P. Yang. 1993. High resolution methylation analysis of the FMR1 gene trinucleotide repeat region in fragile X syndrome. *Human Molecular Genetics*, 2, 1659-1665.

Hu, J., M. Matsui, K. T. Gagnon, J. C. Schwartz, S. Gabillet, K. Arar, J. Wu, I. Bezprozvanny,D. R. Corey. 2009. Allele-specific silencing of mutant huntingtin and ataxin-3 genes by targeting expanded CAG repeats in mRNAs. *Nature Biotechnology*, 27, 478-484.

Huang, C. C., P. W. Faber, F. Persichetti, V. Mittal, J. P. Vonsattel, M. E. MacDonald,J. F. Gusella. 1998. Amyloid formation by mutant huntingtin: threshold, progressivity and recruitment of normal polyglutamine proteins. *Somatic Cell and Molecular Genetics*, 24, 217-233.

Hudson, W. H.,E. A. Ortlund. 2014. The structure, function and evolution of proteins that bind DNA and RNA. *Nature Reviews Molecular Cell Biology*, 15, 749-760.

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

Huntington, G. 2003. On chorea. George Huntington, M.D. *The Journal of Neuropsychiatry & Clinical Neurosciences*, 15, 109-112.

Hutvagner, G.,M. J. Simard. 2008. Argonaute proteins: key players in RNA silencing. *Nature Reviews Molecular Cell Biology*, 9, 22-32.

Hynes, N. E.,T. Stoelzle. 2009. Key signalling nodes in mammary gland development and cancer: Myc. *Breast Cancer Research*, 11, 210.

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

Jasinska, A., G. Michlewski, M. de Mezer, K. Sobczak, P. Kozlowski, M. Napierala,W. J. Krzyzosiak. 2003. Structures of trinucleotide repeats in human transcripts and their functional implications. *Nucleic Acids Research*, 31, 5463-5468.

Jeffreys, A. J., J. K. Holloway, L. Kauppi, C. A. May, R. Neumann, M. T. Slingsby,A. J. Webb. 2004. Meiotic recombination hot spots and human DNA diversity. *Philosophical Transactions of the Royal Society of London*, 359, 141-152.

Jia, H., C. D. Morris, R. M. Williams, J. F. Loring,E. A. Thomas. 2015. HDAC inhibition imparts beneficial transgenerational effects in Huntington's disease mice via altered DNA and histone methylation. *Proceedings of the National Academy of Sciences*, 112, E56-64.

Jiang, H., A. Mankodi, M. S. Swanson, R. T. Moxley,C. A. Thornton. 2004. Myotonic dystrophy type 1 is associated with nuclear foci of mutant RNA, sequestration of muscleblind proteins and deregulated alternative splicing in neurons. *Human Molecular Genetics*, 13, 3079-3088.

Jimenez-Sanchez, M., F. Thomson, E. Zavodszky,D. C. Rubinsztein. 2012. Autophagy and polyglutamine diseases. *Progress in NeuroBiology*, 97, 67-82.

Jin, Y. N.,G. V. W. Johnson. 2010. The interrelationship between mitochondrial dysfunction and transcriptional dysregulation in Huntington disease. *Journal of Bioenergetics and Biomembranes*, 42, 199-205.

Johanson, T. M., A. M. Lew,M. M. Chong. 2013. MicroRNA-independent roles of the RNase III enzymes Drosha and Dicer. *The Open Biology Journal*, 3, 130144.

Johnson, S. A., N. Mandavia, H. D. Wang,D. L. Johnson. 2000. Transcriptional regulation of the TATA-binding protein by Ras cellular signaling. *Molecular Cell Biology*, 20, 5000-5009.

Johnson, S. A. S., L. Dubeau, M. Kawalek, A. Dervan, A. H. Schonthal, C. V. Dang,D. L. Johnson. 2003. Increased expression of TATA-binding protein, the central transcription factor, can contribute to oncogenesis. *Molecular and Cellular Biology*, 23, 3043-3051.

Johnson, S. A. S., L. Dubeau, R. J. White,D. L. Johnson. 2003. The TATA-Binding Protein as a Regulator of Cellular Transformation. *Cell Cycle*, 2, 442-444.

Johnstone, O.,P. Lasko. 2001. Translational regulation and RNA localization in Drosophila oocytes and embryos. *Annual Review of Genetics*, 35, 365-406.

Johri, A.,M. F. Beal. 2012. Antioxidants in Huntington's disease. *Biochimica et Biophysica Acta*, 1822, 664-674.

Jones, R. M., O. Mortusewicz, I. Afzal, M. Lorvellec, P. Garcia, T. Helleday,E. Petermann. 2013. Increased replication initiation and conflicts with transcription underlie Cyclin E-induced replication stress. *Oncogene*, 32, 3744-3753.

Jonkers, I.,J. T. Lis. 2015. Getting up to speed with transcription elongation by RNA polymerase II. *Nature Reviews Molecular Cell Biology*, 16, 167-177.

Joshi, M., A. Kulkarni,J. K. Pal. 2013. Small molecule modulators of eukaryotic initiation factor 2 alpha kinases, the key regulators of protein synthesis. *Biochimie*, 95, 1980-1990.

Jou, Y. S.,R. M. Myers. 1995. Evidence from Antibody Studies That the Cag Repeat in the Huntington Disease Gene Is Expressed in the Protein. *Human Molecular Genetics*, 4, 465-469.

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

Karmodiya, K., A. R. Krebs, M. Oulad-Abdelghani, H. Kimura,L. Tora. 2012. H3K9 and H3K14 acetylation co-occur at many gene regulatory elements, while H3K14ac marks a subset of inactive inducible promoters in mouse embryonic stem cells. *BMC Genomics*, 13, 424.

Karpf, A. R.,D. A. Jones. 2002. Reactivating the expression of methylation silenced genes in human cancer. *Oncogene*, 21, 5496-5503.

Kerschbamer, E.,M. Biagioli. 2015. Huntington's Disease as Neurodevelopmental Disorder: Altered Chromatin Regulation, Coding, and Non-Coding RNA Transcription. *Frontiers in Neuroscience*, 9, 509.

Kertesz, M., Y. Wan, E. Mazor, J. L. Rinn, R. C. Nutter, H. Y. Chang,E. Segal. 2010. Genome-wide measurement of RNA secondary structure in yeast. *Nature*, 467, 103-107.

Kim, E., M. Napierala,S. Y. R. Dent. 2011. Hyperexpansion of GAA repeats affects post-initiation steps of FXN transcription in Friedreich's ataxia. *Nucleic Acids Research*, 39, 8366-8377.

Kita, H., J. Carmichael, J. Swartz, S. Muro, A. Wyttenbach, K. Matsubara, D. C. Rubinsztein,K. Kato. 2002. Modulation of polyglutamine-induced cell death by genes identified by expression profiling. *Human Molecular Genetics*, 11, 2279-2287.

Kornberg, R. 2007. The molecular basis of eukaryotic transcription (Nobel Lecture). *Angewandte Chemie International Edition*, 46, 6956-6965.

Kotsantis, P., R. M. Jones, M. R. Higgs,E. Petermann. 2015. Cancer therapy and replication stress: forks on the road to perdition. *Advances in Clinical Chemistry*, 69, 91-138.

Kozlowski, P., M. de Mezer,W. J. Krzyzosiak. 2010. Trinucleotide repeats in human genome and exome. *Nucleic Acids Research*, 38, 4027-4039.

Krasilnikova, M. M., M. L. Kireeva, V. Petrovic, N. Knijnikova, M. Kashlev,S. M. Mirkin. 2007. Effects of Friedreich's ataxia (GAA)n*(TTC)n repeats on RNA synthesis and stability. *Nucleic Acids Research*, 35, 1075-1084.

Kremer, E. J., M. Pritchard, M. Lynch, S. Yu, K. Holman, E. Baker, S. T. Warren, D. Schlessinger, G. R. Sutherland,R. I. Richards. 1991. Mapping of DNA Instability at the Fragile-X to a Trinucleotide Repeat Sequence P(Ccg)N. *Science*, 252, 1711-1714.

Krol, J., A. Fiszer, A. Mykowska, K. Sobczak, M. de Mezer,W. J. Krzyzosiak. 2007. Ribonuclease dicer cleaves triplet repeat hairpins into shorter repeats that silence specific targets. *Molecular Cell*, 25, 575-586.

Krzyzosiak, W. J., K. Sobczak, M. Wojciechowska, A. Fiszer, A. Mykowska,P. Kozlowski. 2012. Triplet repeat RNA structure and its role as pathogenic agent and therapeutic target. *Nucleic Acids Research*, 40, 11-26.

Kuhn, A., D. R. Goldstein, A. Hodges, A. D. Strand, T. Sengstag, C. Kooperberg, K. Becanovic, M. A. Pouladi, K. Sathasivam, J. H. J. Cha, A. J. Hannan, M. R. Hayden, B. R. Leavitt, S. B. Dunnett, R. J. Ferrante, R. Albin, P. Shelbourne, M. Delorenzi, S. J. Augood, R. L. M. Faull, J. M. Olson, G. P. Bates, L. Jones,R. Luthi-Carter. 2007. Mutant huntingtin's effects on striatal gene expression in mice recapitulate changes observed in human Huntington's disease brain and do not differ with mutant huntingtin length or wild-type huntingtin dosage. *Human Molecular Genetics*, 16, 1845-1861.

Kulis, M.,M. Esteller. 2010. DNA methylation and cancer. *Advances in Genetics*, 70, 27-56.

Kumari, D., R. E. Biacsi,K. Usdin. 2011. Repeat Expansion Affects Both Transcription Initiation and Elongation in Friedreich Ataxia Cells. *Journal of Biological Chemistry*, 286, 4209-4215.

Kumari, D.,K. Usdin. 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.

Kuo, L. J.,L. X. Yang. 2008. Gamma-H2AX - a novel biomarker for DNA double-strand breaks. *In Vivo*, 22, 305-309.

Kurreck, J., E. Wyszko, C. Gillen,V. A. Erdmann. 2002. Design of antisense oligonucleotides stabilized by locked nucleic acids. *Nucleic Acids Research*, 30, 1911-1918.

La Spada, A. R.,J. P. Taylor. 2010. Repeat expansion disease: progress and puzzles in disease pathogenesis. *Nature Reviews Genetics*, 11, 247-258.

La Spada, A. R., E. M. Wilson, D. B. Lubahn, A. E. Harding,K. H. Fischbeck. 1991. Androgen receptor gene mutations in X-linked spinal and bulbar muscular atrophy. *Nature*, 352, 77-79.

Labadorf, A., A. G. Hoss, V. Lagomarsino, J. C. Latourelle, T. C. Hadzi, J. Bregu, M. E. MacDonald, J. F. Gusella, J. F. Chen, S. Akbarian, Z. Weng,R. H. Myers. 2015. RNA Sequence Analysis of Human Huntington Disease Brain Reveals an Extensive Increase in Inflammatory and Developmental Gene Expression. *PLoS One*, 10, e0143563.

Labbadia, J., H. Cunliffe, A. Weiss, E. Katsyuba, K. Sathasivam, T. Seredenina, B. Woodman, S. Moussaoui, S. Frentzel, R. Luthi-Carter, P. Paganetti,G. P. Bates. 2011. Altered chromatin architecture underlies progressive impairment of the heat shock response in mouse models of Huntington disease. *Journal of Clinical Investigation*, 121, 3306-3319.

Labourier, E., H. M. Bourbon, I. Gallouzi, M. Fostier, E. Allemand,J. Tazi. 1999. Antagonism between RSF1 and SR proteins for both splice-site recognition in vitro and Drosophila development. *Genes & Development*, 13, 740-753.

Ladd, P. D., L. E. Smith, N. A. Rabaia, J. M. Moore, S. A. Georges, R. S. Hansen, R. J. Hagerman, F. Tassone, S. J. Tapscott,G. N. Filippova. 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.

Lahue, R. S.,A. Frizzell. 2012. Histone deacetylase complexes as caretakers of genome stability. *Epigenetics*, 7, 806-810.

Landles, C.,G. P. Bates. 2004. Huntingtin and the molecular pathogenesis of Huntington's disease. Fourth in molecular medicine review series. *EMBO Reports*, 5, 958-963.

Landles, C., K. Sathasivam, A. Weiss, B. Woodman, H. Moffitt, S. Finkbeiner, B. Sun, J. Gafni, L. M. Ellerby, Y. Trottier, W. G. Richards, A. Osmand, P. Paganetti,G. P. Bates. 2010. Proteolysis of mutant huntingtin produces an exon 1 fragment that accumulates as an aggregated protein in neuronal nuclei in Huntington disease. *Journal Biological Chemistry*, 285, 8808-8823.

Langbehn, D. R., M. R. Hayden, J. S. Paulsen,Predict- H. D. Investigators of the Huntington Study Group and the. 2010. CAG-repeat length and the age of onset in Huntington disease (HD): a review and validation study of statistical approaches. *American Journal of Medical Genetics Part B: Neuropsychiatric Genetics*, 153B, 397-408.

Lee, J. E.,T. A. Cooper. 2009. Pathogenic mechanisms of myotonic dystrophy. *Biochemical Society Transactions*, 37, 1281-1286.

Lee, J., Y. J. Hwang, K. Y. Kim, N. W. Kowall,H. Ryu. 2013. Epigenetic mechanisms of neurodegeneration in Huntington's disease. *Neurotherapeutics*, 10, 664-676.

Lee, J., Y. J. Hwang, J. Y. Shin, W. C. Lee, J. Wie, K. Y. Kim, M. Y. Lee, D. Hwang, R. R. Ratan, A. N. Pae, N. W. Kowall, I. So, J. I. Kim,H. Ryu. 2013. Epigenetic regulation of cholinergic receptor M1 (CHRM1) by histone H3K9me3 impairs Ca(2+) signaling in Huntington's disease. *Acta Neuropathologica*, 125, 727-739.

Lee, J. M., E. M. Ramos, J. H. Lee, T. Gillis, J. S. Mysore, M. R. Hayden, S. C. Warby, P. Morrison, M. Nance, C. A. Ross, R. L. Margolis, F. Squitieri, S. Orobello, S. Di Donato, E. Gomez-Tortosa, C. Ayuso, O. Suchowersky, R. J. A. Trent, E. McCusker, A. Novelletto, M. Frontali, R. Jones, T. Ashizawa, S. Frank, M. H. Saint-Hilaire, S. M. Hersch, H. D. Rosas, D. Lucente, M. B. Harrison, A. Zanko, R. K. Abramson, K. Marder, J. Sequeiros, J. S. Paulsen, G. B. Landwehrmeyer, R. H. Myers, M. E. MacDonald, J. F. Gusella, PREDICT-HD Study Huntington, REGISTRY Study European, HD-MAPS Study Grp,COHORT Study HSG. 2012. CAG repeat expansion in Huntington disease determines age at onset in a fully dominant fashion. *Neurology*, 78, 690-695.

Lee, T. I.,R. A. Young. 2000. Transcription of eukaryotic protein-coding genes. *Annual Review of Genetics*, 34, 77-137.

Lee, Y., C. Ahn, J. Han, H. Choi, J. Kim, J. Yim, J. Lee, P. Provost, O. Radmark, S. Kim,V. N. Kim. 2003. The nuclear RNase III Drosha initiates microRNA processing. *Nature*, 425, 415-419.

Lemaire, P. A., E. Anderson, J. Lary,J. L. Cole. 2008. Mechanism of PKR Activation by dsRNA. *Journal of Molecular Biology*, 381, 351-360.

Lesnik, E. A.,S. M. Freier. 1995. Relative thermodynamic stability of DNA, RNA, and DNA:RNA hybrid duplexes: relationship with base composition and structure. *Biochemistry*, 34, 10807-10815.

Leuschner, P. J., S. L. Ameres, S. Kueng,J. Martinez. 2006. Cleavage of the siRNA passenger strand during RISC assembly in human cells. *EMBO Reports*, 7, 314-320.

Li, B., M. Carey,J. L. Workman. 2007. The role of chromatin during transcription. *Cell*, 128, 707-719.

Li, H., S. H. Li, Z. X. Yu, P. Shebourne,X. J. Li. 2001. Huntingtin aggregate-associated axonal degeneration is an early pathological event in Huntington's disease mice. *Journal of Neuroscience*, 21, 8473-8481.

Li, L. B.,N. M. Bonini. 2010. Roles of trinucleotide-repeat RNA in neurological disease and degeneration. *Trends Neuroscience*, 33, 292-298.

Li, L. B., Z. M. Yu, X. Y. Teng,N. M. Bonini. 2008. RNA toxicity is a component of ataxin-3 degeneration in Drosophila. *Nature*, 453, 1107-U1109.

Li, L. C., S. T. Okino, H. Zhao, D. Pookot, R. F. Place, S. Urakami, H. Enokida,R. Dahiya. 2006. Small dsRNAs induce transcriptional activation in human cells. *Proceedings of the National Academy of Sciences*, 103, 17337-17342.

Li, S. H., G. Schilling, W. S. Young, 3rd, X. J. Li, R. L. Margolis, O. C. Stine, M. V. Wagster, M. H. Abbott, M. L. Franz, N. G. Ranen,et al. 1993. Huntington's disease gene (IT15) is widely expressed in human and rat tissues. *Neuron*, 11, 985-993.

Lin, B., J. M. Rommens, R. K. Graham, M. Kalchman, H. MacDonald, J. Nasir, A. Delaney, Y. P. Goldberg,M. R. Hayden. 1993. Differential 3' polyadenylation of the Huntington disease gene results in two mRNA species with variable tissue expression. *Hum Molecular Genetics*, 2, 1541-1545.

Lin, C. Y., J. Loven, P. B. Rahl, R. M. Paranal, C. B. Burge, J. E. Bradner, T. I. Lee,R. A. Young. 2012. Transcriptional amplification in tumor cells with elevated c-Myc. *Cell*, 151, 56-67.

Lin, M. T.,M. F. Beal. 2006. Mitochondrial dysfunction and oxidative stress in neurodegenerative diseases. *Nature*, 443, 787-795.

Lin, Y., S. Y. Dent, J. H. Wilson, R. D. Wells,M. Napierala. 2010. R loops stimulate genetic instability of CTG.CAG repeats. *Proceedings of the National Academy of Sciences*, 107, 692-697.

Ling, B. M., S. Gopinadhan, W. K. Kok, S. R. Shankar, P. Gopal, N. Bharathy, Y. Wang,R. Taneja. 2012. G9a mediates Sharp-1-dependent inhibition of skeletal muscle differentiation. *Molecular Biology of the Cell*, 23, 4778-4785.

Liu, C. R., C. R. Chang, Y. Chern, T. H. Wang, W. C. Hsieh, W. C. Shen, C. Y. Chang, I. C. Chu, N. Deng, S. N. Cohen,T. H. Cheng. 2012. Spt4 Is Selectively Required for Transcription of Extended Trinucleotide Repeats. *Cell*, 148, 690-701.

Liu, G. Q.,M. Leffak. 2012. Instability of (CTG)(n)center dot(CAG)(n) trinucleotide repeats and DNA synthesis. *Cell and Bioscience*, 2,

Ljungman, M. 2007. The transcription stress response. *Cell Cycle*, 6, 2252-2257.

Loesch, D. Z., D. E. Godler, A. Evans, Q. M. Bui, F. Gehling, K. E. Kotschet, N. Trost, E. Storey, P. Stimpson, G. Kinsella, D. Francis, D. R. Thorburn, A. Venn, H. R. Slater,M. Horne. 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*, 13, 392-399.

Lopez-Atalaya, J. P., S. Ito, L. M. Valor, E. Benito,A. Barco. 2013. Genomic targets, and histone acetylation and gene expression profiling of neural HDAC inhibition. *Nucleic Acids Reearchs*, 41, 8072-8084.

Lu, J., G. Getz, E. A. Miska, E. Alvarez-Saavedra, J. Lamb, D. Peck, A. Sweet-Cordero, B. L. Ebert, R. H. Mak, A. A. Ferrando, J. R. Downing, T. Jacks, H. R. Horvitz,T. R. Golub. 2005. MicroRNA expression profiles classify human cancers. *Nature*, 435, 834-838.

Luthi-Carter, R., A. Strand, N. L. Peters, S. M. Solano, Z. R. Hollingsworth, A. S. Menon, A. S. Frey, B. S. Spektor, E. B. Penney, G. Schilling, C. A. Ross, D. R. Borchelt, S. J. Tapscott, A. B. Young, J. H. Cha,J. M. Olson. 2000. Decreased expression of striatal signaling genes in a mouse model of Huntington's disease. *Human Molecular Genetics*, 9, 1259-1271.

Maduike, N. Z., A. K. Tehranchi, J. D. Wang,K. N. Kreuzer. 2014. Replication of the Escherichia coli chromosome in RNase HI-deficient cells: multiple initiation regions and fork dynamics. *Molecular MicroBiology*, 91, 39-56.

Mah, L. J., A. El-Osta,T. C. Karagiannis. 2010. gammaH2AX: a sensitive molecular marker of DNA damage and repair. *Leukemia*, 24, 679-686.

Mangiarini, L., K. Sathasivam, M. Seller, B. Cozens, A. Harper, C. Hetherington, M. Lawton, Y. Trottier, H. Lehrach, S. W. Davies,G. P. Bates. 1996. Exon 1 of the HD gene with an expanded CAG repeat is sufficient to cause a progressive neurological phenotype in transgenic mice. *Cell*, 87, 493-506.

Marangos, P. J.,S. M. Paul. 1981. Brain levels of neuron-specific and nonneuronal enolase in Huntington's disease. *Journal of Neurochemistry*, 37, 1338-1340.

Marenholz, I.,C. W. Heizmann. 2004. S100A16, a ubiquitously expressed EF-hand protein which is up-regulated in tumors. *Biochemical and Biophysical Research Communications*, 313, 237-244.

Marti, E. 2016. RNA toxicity induced by expanded CAG repeats in Huntington's disease. *Brain Pathology*, 26, 779-786.

Martin, J. B.,J. F. Gusella. 1986. Huntington's disease. Pathogenesis and management. *The New England Journal of Medicine*, 315, 1267-1276.

Marzluff, W. F. 2005. Metazoan replication-dependent histone mRNAs: a distinct set of RNA polymerase II transcripts. *Current Opinion in Cell Biology*, 17, 274-280.

Masliah, G., P. Barraud,F. H. Allain. 2013. RNA recognition by double-stranded RNA binding domains: a matter of shape and sequence. *Cellular and Molecular Life Sciences*, 70, 1875-1895.

McFarland, K. N., S. Das, T. T. Sun, D. Leyfer, E. Xia, G. R. Sangrey, A. Kuhn, R. Luthi-Carter, T. W. Clark, G. Sadri-Vakili,J. H. Cha. 2012. Genome-wide histone acetylation is altered in a transgenic mouse model of Huntington's disease. *PLoS One*, 7, e41423.

McGranahan, N., R. A. Burrell, D. Endesfelder, M. R. Novelli,C. Swanton. 2012. Cancer chromosomal instability: therapeutic and diagnostic challenges. *EMBO Reports*, 13, 528-538.

McLaughlin, B. A., C. Spencer,J. Eberwine. 1996. CAG trinucleotide RNA repeats interact with RNA-binding proteins. *American Journal of Human Genetics*, 59, 561-569.

McMahon, S. J., M. G. Pray-Grant, D. Schieltz, J. R. Yates, 3rd,P. A. Grant. 2005. Polyglutamine-expanded spinocerebellar ataxia-7 protein disrupts normal SAGA and SLIK histone acetyltransferase activity. *Proceedings of the National Academy of Sciences*, 102, 8478-8482.

McMurray, C. T. 2001. Huntington's disease: new hope for therapeutics. *Trends Neuroscience*, 24, S32-38.

McMurray, C. T. 2010. Mechanisms of trinucleotide repeat instability during human development. *Nature Reviews Genetics*, 11, 786-799.

Meister, G., M. Landthaler, A. Patkaniowska, Y. Dorsett, G. Teng,T. Tuschl. 2004. Human Argonaute2 mediates RNA cleavage targeted by miRNAs and siRNAs. , 15, 185-197.

Michlewski, G.,W. J. Krzyzosiak. 2004. Molecular architecture of CAG repeats in human disease related transcripts. *Journal of Molecular Biology*, 340, 665-679.

Mirkin, E. V.,S. M. Mirkin. 2005. Mechanisms of transcription-replication collisions in bacteria. *Biology*, 25, 888-895.

Mirkin, S. M. 2007. Expandable DNA repeats and human disease. *Nature*, 447, 932-940.

Morrison, P. J., S. Harding-Lester,A. Bradley. 2011. Uptake of Huntington disease predictive testing in a complete population. *Clinical Genetics*, 80, 281-286.

Mortimer, S. A., M. A. Kidwell,J. A. Doudna. 2014. Insights into RNA structure and function from genome-wide studies. *Nature Reviews Genetics*, 15, 469-479.

Morton, A. J., D. Glynn, W. Leavens, Z. Zheng, R. L. Faull, J. N. Skepper,J. M. Wight. 2009. Paradoxical delay in the onset of disease caused by super-long CAG repeat expansions in R6/2 mice. *NeuroBiology Disease*, 33, 331-341.

Moseley, M. L., T. Zu, Y. Ikeda, W. Gao, A. K. Mosemiller, R. S. Daughters, G. Chen, M. R. Weatherspoon, H. B. Clark, T. J. Ebner, J. W. Day,L. P. Ranum. 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., E. Boyarchuk, J. Pontis,S. Ait-Si-Ali. 2015. Sound of silence: the properties and functions of repressive Lys methyltransferases. *Nature Reviews Biology*, 16, 499-513.

Mueller, S., V. Gausson, N. Vodovar, S. Deddouche, L. Troxler, J. Perot, S. Pfeffer, J. A. Hoffmann, M. C. Saleh,J. L. Imler. 2010. RNAi-mediated immunity provides strong protection against the negative-strand RNA vesicular stomatitis virus in Drosophila. *Proceedings of the National Academy of Sciences*, 107, 19390-19395.

Mykowska, A., K. Sobczak, M. Wojciechowska, P. Kozlowski,W. J. Krzyzosiak. 2011. CAG repeats mimic CUG repeats in the misregulation of alternative splicing. *Nucleic Acids Research*, 39, 8938-8951.

Nakama, M., K. Kawakami, T. Kajitani, T. Urano,Y. Murakami. 2012. DNA-RNA hybrid formation mediates RNAi-directed heterochromatin formation. *Genes Cells*, 17, 218-233.

Nakamori, M., C. E. Pearson,C. A. Thornton. 2011. Bidirectional transcription stimulates expansion and contraction of expanded (CTG)*(CAG) repeats. *Human Molecular Genetics*, 20, 580-588.

Nasir, J., S. B. Floresco, J. R. O'Kusky, V. M. Diewert, J. M. Richman, J. Zeisler, A. Borowski, J. D. Marth, A. G. Phillips,M. R. Hayden. 1995. Targeted disruption of the Huntington's disease gene results in embryonic lethality and behavioral and morphological changes in heterozygotes. *Cell*, 81, 811-823.

Neelsen, K. J., I. M. Zanini, R. Herrador,M. Lopes. 2013. Oncogenes induce genotoxic stress by mitotic processing of unusual replication intermediates. *Journal Cell Biology*, 200, 699-708.

Negrini, S., V. G. Gorgoulis,T. D. Halazonetis. 2010. Genomic instability--an evolving hallmark of cancer. *Nature Revies Biology*, 11, 220-228.

Nelson, D. L., H. T. Orr,S. T. Warren. 2013. The unstable repeats--three evolving faces of neurological disease. *Neuron*, 77, 825-843.

Nie, Z. Q., G. Q. Hu, G. Wei, K. R. Cui, A. Yamane, W. Resch, R. N. Wang, D. R. Green, L. Tessarollo, R. Casellas, K. J. Zhao,D. Levens. 2012. c-Myc Is a Universal Amplifier of Expressed Genes in Lymphocytes and Embryonic Stem Cells. *Cell*, 151, 68-79.

Novoselova, T. V., B. A. Margulis, S. S. Novoselov, A. M. Sapozhnikov, J. van der Spuy, M. E. Cheetham,I. V. Guzhova. 2005. Treatment with extracellular HSP70/HSC70 protein can reduce polyglutamine toxicity and aggregation. *Journal of Neurochemistry*, 94, 597-606.

Ohshima, K., L. Montermini, R. D. Wells,M. Pandolfo. 1998. Inhibitory effects of expanded GAA.TTC triplet repeats from intron I of the Friedreich ataxia gene on transcription and replication in vivo. *Journal Biological Chemistry*, 273, 14588-14595.

Orr, H. T. 2004. RNA gains a new function: a mediator of neurodegeneration. *Trends Neuroscience*, 27, 233-234.

Orr, H. T.,H. Y. Zoghbi. 2007. Trinucleotide repeat disorders. *Annual Review of Neuroscience*, 30, 575-621.

Palhan, V. B., S. Chen, G. H. Peng, A. Tjernberg, A. M. Gamper, Y. Fan, B. T. Chait, A. R. La Spada,R. G. Roeder. 2005. Polyglutamine-expanded ataxin-7 inhibits STAGA histone acetyltransferase activity to produce retinal degeneration. *Proceedings of the National Academy of Sciences*, 102, 8472-8477.

Patel, R. C.,G. C. Sen. 1998. PACT, a protein activator of the interferon-induced protein kinase, PKR. *EMBO JOURNAL*, 17, 4379-4390.

Paulsen, J. S., D. R. Langbehn, J. C. Stout, E. Aylward, C. A. Ross, M. Nance, M. Guttman, S. Johnson, M. MacDonald, L. J. Beglinger, K. Duff, E. Kayson, K. Biglan, I. Shoulson, D. Oakes, M. Hayden,Predict HD Investigators & Coordin. 2008. Detection of Huntington's disease decades before diagnosis: the Predict-HD study. *Journal of Neurology Neurosurgery and Psychiatry*, 79, 874-880.

Peel, A. L., R. V. Rao, B. A. Cottrell, M. R. Hayden, L. M. Ellerby,D. E. Bredesen. 2001. Double-stranded RNA-dependent protein kinase, PKR, binds preferentially to Huntington's disease (HD) transcripts and is activated in HD tissue. *Human Molecular Genetics*, 10, 1531-1538.

Pelechano, V.,L. M. Steinmetz. 2013. Gene regulation by antisense transcription. *Nature Reviews Genetics*, 14, 880-893.

Perconti, G., C. Maranto, D. P. Romancino, P. Rubino, S. Feo, A. Bongiovanni,A. Giallongo. 2017. Pro-invasive stimuli and the interacting protein Hsp70 favour the route of alpha-enolase to the cell surface. *Science Reports*, 7, 3841.

Pfister, E. L., L. Kennington, J. Straubhaar, S. Wagh, W. Z. Liu, M. DiFiglia, B. Landwehrmeyer, J. P. Vonsattel, P. D. Zamore,N. Aronin. 2009. Five siRNAs Targeting Three SNPs May Provide Therapy for Three-Quarters of Huntington's Disease Patients. *Current Biology*, 19, 774-778.

Phillips, D. D., D. N. Garboczi, K. Singh, Z. Hu, S. H. Leppla,C. E. Leysath. 2013. The sub-nanomolar binding of DNA-RNA hybrids by the single-chain Fv fragment of antibody S9.6. *Journal of Molecular Recognition*, 26, 376-381.

Pietrobono, R., E. Tabolacci, F. Zalfa, I. Zito, A. Terracciano, U. Moscato, C. Bagni, B. Oostra, P. Chiurazzi,G. Neri. 2005. Molecular dissection of the events leading to inactivation of the FMR1 gene. *Human Molecular Genetics*, 14, 267-277.

Pillai, J. A., L. A. Hansen, E. Masliah, J. L. Goldstein, S. D. Edland,J. Corey-Bloom. 2012. Clinical severity of Huntington's disease does not always correlate with neuropathologic stage. *Movement Disorders*, 27, 1099-1103.

Pires-daSilva, A., K. Nayernia, W. Engel, M. Torres, A. Stoykova, K. Chowdhury,P. Gruss. 2001. Mice deficient for spermatid perinuclear RNA-binding protein show neurologic, spermatogenic, and sperm morphological abnormalities. *Developmental Biology*, 233, 319-328.

Podhorecka, M., A. Skladanowski,P. Bozko. 2010. H2AX Phosphorylation: Its Role in DNA Damage Response and Cancer Therapy. *Journal Nucleic Acids*, 2010,

Pommier, Y. 2006. Topoisomerase I inhibitors: camptothecins and beyond. *Nature Reviews Cancer*, 6, 789-802.

Portal, M. M., V. Pavet, C. Erb,H. Gronemeyer. 2015. Human cells contain natural double-stranded RNAs with potential regulatory functions. *Nature Structural & Molecular Biology*, 22, 89-97.

Prado, F.,A. Aguilera. 2005. Impairment of replication fork progression mediates RNA polII transcription-associated recombination. *EMBO Journal*, 24, 1267-1276.

Ray, R.,D. M. Miller. 1991. Cloning and characterization of a human c-myc promoter-binding protein. *Biology*, 11, 2154-2161.

Reddy, K., M. Tam, R. P. Bowater, M. Barber, M. Tomlinson, K. Nichol Edamura, Y. H. Wang,C. E. Pearson. 2011. Determinants of R-loop formation at convergent bidirectionally transcribed trinucleotide repeats. *Nucleic Acids Research*, 39, 1749-1762.

Renton, A. E., E. Majounie, A. Waite, J. Simon-Sanchez, S. Rollinson, J. R. Gibbs, J. C. Schymick, H. Laaksovirta, J. C. van Swieten, L. Myllykangas, H. Kalimo, A. Paetau, Y. Abramzon, A. M. Remes, A. Kaganovich, S. W. Scholz, J. Duckworth, J. H. Ding, D. W. Harmer, D. G. Hernandez, J. O. Johnson, K. Mok, M. Ryten, D. Trabzuni, R. J. Guerreiro, R. W. Orrell, J. Neal, A. Murray, J. Pearson, I. E. Jansen, D. Sondervan, H. Seelaar, D. Blake, K. Young, N. Halliwell, J. B. Callister, G. Toulson, A. Richardson, A. Gerhard, J. Snowden, D. Mann, D. Neary, M. A. Nalls, T. Peuralinna, L. Jansson, V. M. Isoviita, A. L. Kaivorinne, M. Holtta-Vuori, E. Ikonen, R. Sulkava, M. Benatar, J. Wuu, A. Chio, G. Researchtagno, G. Borghero, M. Sabatelli, D. Heckerman, E. Rogaeva, L. Zinman, J. D. Rothstein, M. Sendtner, C. Drepper, E. E. Eichler, C. Alkan, Z. Abdullaev, S. D. Pack, A. Dutra, E. Pak, J. Hardy, A. Singleton, N. M. Williams, P. Heutink, S. Pickering-Brown, H. R. Morris, P. J. Tienari, B. J. Traynor,ITALSGEN Consortium. 2011. A Hexanucleotide Repeat Expansion in C9ORF72 Is the Cause of Chromosome 9p21-Linked ALS-FTD. *Neuron*, 72, 257-268.

Richard, P., S. Feng,J. L. Manley. 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.

Ryu, H., J. Lee, S. W. Hagerty, B. Y. Soh, S. E. McAlpin, K. A. Cormier, K. M. Smith,R. J. Ferrante. 2006. ESET/SETDB1 gene expResearchsion and histone H3 (K9) trimethylation in Huntington's disease. *Proceedings of the National Academy of Sciences*, 103, 19176-19181.

Sadri-Vakili, G., B. Bouzou, C. L. Benn, M. O. Kim, P. Chawla, R. P. Overland, K. E. Glajch, E. Xia, Z. Qiu, S. M. Hersch, T. W. Clark, G. J. Yohrling,J. H. Cha. 2007. Histones associated with downregulated genes are hypo-acetylated in Huntington's disease models. *Human Molecular Genetics*, 16, 1293-1306.

Sadri-Vakili, G.,J. H. Cha. 2006. Mechanisms of disease: Histone modifications in Huntington's disease. *Nature Clinical Practice Neurology*, 2, 330-338.

Sah, D. W.,N. Aronin. 2011. Oligonucleotide therapeutic approaches for Huntington disease. *Journal Clinical Investigation*, 121, 500-507.

Santos-Pereira, J. M.,A. Aguilera. 2015. R loops: new modulators of genome dynamics and function. *Nature Reviews Genetics*, 16, 583-597.

Santos-Pereira, J. M., A. B. Herrero, M. L. Garcia-Rubio, A. Marin, S. Moreno,A. Aguilera. 2013. The Npl3 hnRNP prevents R-loop-mediated transcription-replication conflicts and genome instability. *Genes Development*, 27, 2445-2458.

Sathasivam, K., A. Neueder, T. A. Gipson, C. Landles, A. C. Benjamin, M. K. Bondulich, D. L. Smith, R. L. Faull, R. A. Roos, D. Howland, P. J. Detloff, D. E. Housman,G.

P. Bates. 2013. Aberrant splicing of HTT generates the pathogenic exon 1 protein in Huntington disease. *Proceedings of the National Academy of Sciences*, 110, 2366-2370.

Saunders, L. R.,G. N. Barber. 2003. The dsRNA binding protein family: critical roles, diverse cellular functions. *FASEB Journal*, 17, 961-983.

Scheer, S.,C. Zaph. 2017. The Lysine Methyltransferase G9a in Immune Cell Differentiation and Function. *Frontiers in Immunology*, 8, 429.

Schonborn, J., J. Oberstrass, E. Breyel, J. Tittgen, J. Schumacher,N. Lukacs. 1991. Monoclonal-Antibodies to Double-Stranded-Rna as Probes of Rna Structure in Crude Nucleic-Acid Extracts. *Nucleic Acids Researchearch*, 19, 2993-3000.

Schulte, J.,J. T. Littleton. 2011. The Biological function of the Huntingtin protein and its relevance to Huntington's Disease pathology. *Current Trends in Neurology*, 5, 65-78.

Schumacher, J. M., K. Lee, S. Edelhoff,R. E. Braun. 1995a. Distribution of Tenr, an Rna-Binding Protein, in a Lattice-Like Network within the Spermatid Nucleus in the Mouse. *Biology of Reproduction*, 52, 1274-1283.

Schumacher, J. M., K. Lee, S. Edelhoff,R. E. Braun. 1995b. Spnr, a Murine Rna-Binding Protein That Is Localized to Cytoplasmic Microtubules. *Journal of Cell Biology*, 129, 1023-1032.

Schwab, L. C., K. Richetin, R. A. Barker,N. Deglon. 2017. Formation of hippocampal mHTT aggregates leads to impaired spatial memory, hippocampal activation and adult neurogenesis. *NeuroBiology of Disease*, 102, 105-112.

Schwab, R. A., J. Nieminuszczy, F. Shah, J. Langton, D. Lopez Martinez, C. C. Liang, M. A. Cohn, R. J. Gibbons, A. J. Deans,W. Niedzwiedz. 2015. The Fanconi Anemia Pathway Maintains Genome Stability by Coordinating Replication and Transcription. *Molecular Cell*, 60, 351-361.

Seok, H., H. Lee, E. S. Jang,S. W. Chi. 2017. Evaluation and control of miRNA-like off-target repResearchsion for RNA interference. *Cell Molecular Life Science*,

Seong, I. S., J. M. Woda, J. J. Song, A. Lloret, P. D. Abeyrathne, C. J. Woo, G. Gregory, J. M. Lee, V. C. Wheeler, T. Walz, R. E. Kingston, J. F. Gusella, R. A. Conlon,M. E. MacDonald. 2010. Huntingtin facilitates polycomb repResearchsive complex 2. *Human Molecular Genetics*, 19, 573-583.

Shaw, N. N.,D. P. Arya. 2008. Recognition of the unique structure of DNA:RNA hybrids. *Biochimie*, 90, 1026-1039.

Shin, J. H.,Z. Kelman. 2006. The replicative helicases of bacteria, archaea, and eukarya can unwind RNA-DNA hybrid substrates. *Journal of Biological Chemistry*, 281, 26914-26921.

Shishkin, A. A., I. Voineagu, R. Matera, N. Cherng, B. T. Chernet, M. M. Krasilnikova, V. Narayanan, K. S. Lobachev,S. M. Mirkin. 2009. Large-scale expansions of Friedreich's ataxia GAA repeats in yeast. *Molecular Cell*, 35, 82-92.

Sicot, G., G. Gourdon,M. Gomes-Pereira. 2011. Myotonic dystrophy, when simple repeats reveal complex pathogenic entities: new findings and future challenges. *Human Molecular Genetics*, 20, R116-R123.

Sikdar, N., S. Banerjee, H. Zhang, S. Smith,K. Myung. 2008. Spt2p defines a new transcription-dependent gross chromosomal rearrangement pathway. *PLoS Genetics*, 4, e1000290.

Sipione, S., D. Rigamonti, M. Valenza, C. Zuccato, L. Conti, J. Pritchard, C. Kooperberg, J. M. Olson,E. Cattaneo. 2002. Early transcriptional profiles in huntingtin-inducible striatal cells by microarray analyses. *Human Molecular Genetics*, 11, 1953-1965.

Skourti-Stathaki, K., K. Kamieniarz-Gdula,N. J. Proudfoot. 2014. R-loops induce repResearchsive chromatin marks over mammalian gene terminators. *Nature*, 516, 436-439.

Skourti-Stathaki, K.,N. J. Proudfoot. 2014. A double-edged sword: R loops as threats to genome integrity and powerful regulators of gene expResearchsion. *Genes Development*, 28, 1384-1396.

Skourti-Stathaki, K., N. J. Proudfoot,N. Gromak. 2011. Human senataxin Researcholves RNA/DNA hybrids formed at transcriptional pause sites to promote Xrn2-dependent termination. , 42, 794-805.

Snell, R. G., J. C. Macmillan, J. P. Cheadle, I. Fenton, L. P. Lazarou, P. Davies, M. E. Macdonald, J. F. Gusella, P. S. Harper,D. J. Shaw. 1993. Relationship between Trinucleotide Repeat Expansion and Phenotypic Variation in Huntingtons-Disease. *Nature Genetics*, 4, 393-397.

Sobczak, K., M. de Mezer, G. Michlewski, J. Krol,W. J. Krzyzosiak. 2003. RNA structure of trinucleotide repeats associated with human neurological diseases. *Nucleic Acids Research*, 31, 5469-5482.

Sollier, J., C. T. Stork, M. L. Garcia-Rubio, R. D. Paulsen, A. Aguilera,K. A. Cimprich. 2014. Transcription-coupled nucleotide excision repair factors promote R-loop-induced genome instability. , 56, 777-785.

Sopher, B. L., P. D. Ladd, V. V. Pineda, R. T. Libby, S. M. Sunkin, J. B. Hurley, C. P. Thienes, T. Gaasterland, G. N. Filippova,A. R. La Spada. 2011. CTCF Regulates Ataxin-7 ExpResearchsion through Promotion of a Convergently Transcribed, Antisense Noncoding RNA. *Neuron*, 70, 1071-1084.

Sordet, O., C. E. Redon, J. Guirouilh-Barbat, S. Smith, S. Solier, C. Douarre, C. Conti, A. J. Nakamura, B. B. Das, E. Nicolas, K. W. Kohn, W. M. Bonner,Y. Pommier. 2009. Ataxia telangiectasia mutated activation by transcription- and topoisomerase I-induced DNA double-strand breaks. *EMBO Reports*, 10, 887-893.

Spadari, S., F. Sala,G. Pedralinoy. 1982. Aphidicolin - a Specific Inhibitor of Nuclear-DNA Replication in Eukaryotes. *Trends in Biochemical Sciences*, 7, 29-32.

St Johnston, D., N. H. Brown, J. G. Gall,M. Jantsch. 1992. A conserved double-stranded RNA-binding domain. *Proceedings of the National Academy of Sciences*, 89, 10979-10983.

Steffan, J. S., L. Bodai, J. Pallos, M. Poelman, A. McCampbell, B. L. Apostol, A. Kazantsev, E. Schmidt, Y. Z. Zhu, M. Greenwald, R. Kurokawa, D. E. Housman, G. R. Jackson, J. L. Marsh,L. M. Thompson. 2001. Histone deacetylase inhibitors arResearcht polyglutamine-dependent neurodegeneration in Drosophila. *Nature*, 413, 739-743.

Steffan, J. S., A. Kazantsev, O. Spasic-Boskovic, M. Greenwald, Y. Z. Zhu, H. Gohler, E. E. Wanker, G. P. Bates, D. E. Housman,L. M. Thompson. 2000. The Huntington's disease protein interacts with p53 and CREB-binding protein and repResearchses transcription. *Proceedings of the National Academy of Sciences*, 97, 6763-6768.

Steinbach, P., D. Glaser, W. Vogel, M. Wolf,S. Schwemmle. 1998. The DMPK gene of severely affected myotonic dystrophy patients is hypermethylated proximal to the largely expanded CTG repeat. *American Journal of Human Genetics*, 62, 278-285.

Sutcliffe, J. S., D. L. Nelson, F. Zhang, M. Pieretti, C. T. Caskey, D. Saxe,S. T. Warren. 1992. DNA methylation repResearchses FMR-1 transcription in fragile X syndrome. *Human Molecular Genetics*, 1, 397-400.

Suzuki, Y., J. B. Holmes, S. M. Cerritelli, K. Sakhuja, M. Minczuk, I. J. Holt,R. J. Crouch. 2010. An upstream open reading frame and the context of the two AUG codons affect the abundance of mitochondrial and nuclear RNase H1. *Biology*, 30, 5123-5134.

Svoboda, P.,A. Di Cara. 2006. Hairpin RNA: a secondary structure of primary importance. *Cell Molecular Life Science*, 63, 901-908.

Thomas, E. A., G. Coppola, B. Tang, A. Kuhn, S. Kim, D. H. Geschwind, T. B. Brown, R. Luthi-Carter,M. E. Ehrlich. 2011. In vivo cell-autonomous transcriptional abnormalities revealed in mice expResearchsing mutant huntingtin in striatal but not cortical neurons. *Human Molecular Genetics*, 20, 1049-1060.

Thomas, M., R. L. White,R. W. Davis. 1976. Hybridization of RNA to double-stranded DNA: formation of R-loops. *Proceedings of the National Academy of Sciences*, 73, 2294-2298.

Toth, G., Z. Gaspari,J. Jurka. 2000. Microsatellites in different eukaryotic genomes: survey and analysis. *Genome Research*, 10, 967-981.

Tous, C.,A. Aguilera. 2007. Impairment of transcription elongation by R-loops in vitro. *Biochemical and Biophysical Research Communications*, 360, 428-432.

Tracy, R. B.,M. R. Lieber. 2000. Transcription-dependent R-loop formation at mammalian class switch sequences. *EMBO Journal*, 19, 1055-1067.

Tsantoulis, P. K., A. Kotsinas, P. P. Sfikakis, K. Evangelou, M. Sideridou, B. Levy, L. Mo, C. Kittas, X. R. Wu, A. G. Papavassiliou,V. G. Gorgoulis. 2008. Oncogene-induced replication stResearchs preferentially targets common fragile sites in preneoplastic lesions. A genome-wide study. *Oncogene*, 27, 3256-3264.

Tuduri, S., L. Crabbe, C. Conti, H. Tourriere, H. Holtgreve-Grez, A. Jauch, V. Pantesco, J. De Vos, A. Thomas, C. Theillet, Y. Pommier, J. Tazi, A. Coquelle,P. Pasero. 2009. Topoisomerase I suppResearchses genomic instability by preventing interference between replication and transcription. *Nature Cell Biology*, 11, 1315-U1125.

Vedadi, M., D. Barsyte-Lovejoy, F. Liu, S. Rival-Gervier, A. Allali-Hassani, V. Labrie, T. J. Wigle, P. A. Dimaggio, G. A. Wasney, A. Siarheyeva, A. Dong, W. Tempel, S. C. Wang, X. Chen, I. Chau, T. J. Mangano, X. P. Huang, C. D. Simpson, S. G. Pattenden, J. L. Norris, D. B. Kireev, A. Tripathy, A. Edwards, B. L. Roth, W. P. Janzen, B. A. Garcia, A. Petronis, J. Ellis, P. J. Brown, S. V. Frye, C. H. Arrowsmith,J. Jin. 2011. A chemical probe selectively inhibits G9a and GLP methyltransferase activity in cells. *Nature Chem Biology*, 7, 566-574.

Verkerk, A. J., M. Pieretti, J. S. Sutcliffe, Y. H. Fu, D. P. Kuhl, A. Pizzuti, O. Reiner, S. Richards, M. F. Victoria, F. P. Zhang,et al. 1991. Identification of a gene (FMR-1) containing a CGG repeat coincident with a breakpoint cluster region exhibiting length variation in fragile X syndrome. *Cell*, 65, 905-914.

Vieira, K. F., P. P. Levings, M. A. Hill, V. J. Crusselle, S. H. L. Kang, J. D. Engel,J. R. Bungert. 2004. Recruitment of transcription complexes to the beta-globin gene locus in vivo and in vitro. *Journal of Biological Chemistry*, 279, 50350-50357.

Vonsattel, J. P.,M. DiFiglia. 1998. Huntington disease. *J Neuropathol Exp Neurol*, 57, 369-384.

Voorhoeve, P. M., C. le Sage, M. Schrier, A. J. Gillis, H. Stoop, R. Nagel, Y. P. Liu, J. van Duijse, J. Drost, A. Griekspoor, E. Zlotorynski, N. Yabuta, G. De Vita, H. Nojima, L. H. Looijenga,R. Agami. 2006. A Geneticsic screen implicates miRNA-372 and miRNA-373 as oncogenes in testicular germ cell tumors. *Cell*, 124, 1169-1181.

Wahba, L., J. D. Amon, D. Koshland,M. Vuica-Ross. 2011. RNase H and multiple RNA biogenesis factors cooperate to prevent RNA:DNA hybrids from generating genome instability. , 44, 978-988.

Wahba, L., S. K. Gore,D. Koshland. 2013. The hoMolecularogous recombiNatureion machinery modulates the formation of RNA-DNA hybrids and associated chromosome instability. *Elife*, 2, e00505.

Walker, F. O. 2007. Huntington's disease. *Lancet*, 369, 218-228.

Wang, H. D., A. Trivedi,D. L. Johnson. 1997. Hepatitis B virus X protein induces RNA polymerase III-dependent gene transcription and increases cellular TATA-binding protein by activating the Ras signaling pathway. *Molecular and Cellular Biology*, 17, 6838-6846.

Wang, J. C. 2002. Cellular roles of DNA topoisomerases: a Molecular perspective. *Nature Rev Biology*, 3, 430-440.

Wang, L. C., K. Y. Chen, H. Pan, C. C. Wu, P. H. Chen, Y. T. Liao, C. Li, M. L. Huang,K. M. Hsiao. 2011. Muscleblind participates in RNA toxicity of expanded CAG and CUG repeats in Caenorhabditis elegans. *Cell Molecular Life Science*, 68, 1255-1267.

Wang, Q., J. Khillan, P. Gadue,K. Nishikura. 2000. Requirement of the RNA editing deaminase ADAR1 gene for embryonic erythropoiesis. *Scienceence*, 290, 1765-1768.

Wang, X., L. Vukovic, H. R. Koh, K. Schulten,S. Myong. 2015. Dynamic profiling of double-stranded RNA binding proteins. *Nucleic Acids Research*, 43, 7566-7576.

Watts, J. M., K. K. Dang, R. J. Gorelick, C. W. Leonard, J. W. Bess, Jr., R. Swanstrom, C. L. Burch,K. M. Weeks. 2009. Architecture and secondary structure of an entire HIV-1 RNA genome. *Nature*, 460, 711-716.

Wei, X. Y., J. Samarabandu, R. S. Developmentdhar, A. J. Siegel, R. Acharya,R. Berezney. 1998. Segregation of transcription and replication sites into higher order domains. *Science*, 281, 1502-1505.

Wellington, C. L., R. Singaraja, L. Ellerby, J. Savill, S. Roy, B. Leavitt, E. Cattaneo, A. Hackam, A. Sharp, N. Thornberry, D. W. Nicholson, D. E. Bredesen,M. R. Hayden. 2000. Inhibiting caspase cleavage of huntingtin reduces toxicity and aggregate formation in neuronal and nonneuronal cells. *Journal Biological Chemistry*, 275, 19831-19838.

West, S., N. Gromak,N. J. Proudfoot. 2004. Human 5' --> 3' exonuclease Xrn2 promotes transcription termiNatureion at co-transcriptional cleavage sites. *Nature*, 432, 522-525.

Wheeler, V. C., W. Auerbach, J. K. White, J. Srinidhi, A. Auerbach, A. Ryan, M. P. Duyao, V. Vrbanac, M. Weaver, J. F. Gusella, A. L. Joyner,M. E. MacDonald. 1999. Length-dependent gametic CAG repeat instability in the Huntington's disease knock-in mouse. *Human Molecular Genetics*, 8, 115-122.

Wheeler, V. C., F. Persichetti, S. M. McNeil, J. S. Mysore, S. S. Mysore, M. E. MacDonald, R. H. Myers, J. F. Gusella, N. S. Wexler,U. S-Venezuela Collaborative Researchearch Group. 2007. Factors associated with HD CAG repeat instability in Huntington disease. *Journal of Medical Genetics*, 44, 695-701.

White, E., M. Schlackow, K. Kamieniarz-Gdula, N. J. Proudfoot,M. Gullerova. 2014. Human nuclear Dicer Researchtricts the deleterious accumulation of endogenous double-stranded RNA. *Nature Structural & Molecular Biology*, 21, 552-559.

White, J. K., W. Auerbach, M. P. Duyao, J. P. Vonsattel, J. F. Gusella, A. L. Joyner,M. E. MacDonald. 1997. Huntingtin is required for neurogenesis and is not impaired by the Huntington's disease CAG expansion. *Nature Genetics*, 17, 404-410.

White, R. J. 2005. RNA polymerases I and III, growth control and cancer. *Nature Reviews Biology*, 6, 69-78.

Wilburn, B., D. D. Rudnicki, J. Zhao, T. M. Weitz, Y. Cheng, X. F. Gu, E. Greiner, C. S. Park, N. Wang, B. L. Sopher, A. R. La Spada, A. Osmand, R. L. Margolis, Y. E. Sun,X. W. Yang. 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., M. Harreman,J. Q. Svejstrup. 2013. Ubiquitylation and degradation of elongating RNA polymerase II: the last Researchort. *Biochimica et Biophysica Acta*, 1829, 151-157.

Wisniewski, J. R., A. Zougman,M. Mann. 2009. CombiNatureion of FASP and StageTip-based fractioNatureion allows in-depth analysis of the hippocampal membrane proteome. *Journal Proteome Research*, 8, 5674-5678.

Wisniewski, J. R., A. Zougman, N. Nagaraj,M. Mann. 2009. Universal sample preparation method for proteome analysis. *Nature Methods*, 6, 359-362.

Wojciechowska, M.,W. J. Krzyzosiak. 2011. Cellular toxicity of expanded RNA repeats: focus on RNA foci. *Human Molecular Genetics*, 20, 3811-3821.

Wu, K. X., P. Phuektes, P. Kumar, G. Y. Goh, D. Moreau, V. T. Chow, F. Bard,J. J. Chu. 2016. Human genome-wide RNAi screen reveals host factors required for enterovirus 71 replication. *Nature Communications*, 7, 13150.

Xu, B. J.,D. A. Clayton. 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. *EMBO Journal*, 15, 3135-3143.

Yang, Y., K. M. McBride, S. Hensley, Y. Lu, F. Chedin,M. T. Bedford. 2014. Arginine methylation facilitates the recruitment of TOP3B to chromatin to prevent R loop accumulation. , 53, 484-497.

Zee, B. M., R. S. Levin, B. Xu, G. LeRoy, N. S. Wingreen,B. A. Garcia. 2010. In vivo Researchidue-specific histone methylation dynamics. *J Biology Chem*, 285, 3341-3350.

Zeman, M. K.,K. A. Cimprich. 2014. Causes and consequences of replication stResearchs. *Nature Cell Biology*, 16, 2-9.

Zhang, Y., A. A. Shishkin, Y. Nishida, D. Marcinkowski-Desmond, N. Saini, K. V. Volkov, S. M. Mirkin,K. S. Lobachev. 2012. Genome-wide Screen Identifies Pathways that Govern GAA/TTC Repeat Fragility and Expansions in Dividing and Nondividing Yeast Cells. *Molecular Cell*, 48, 254-265.

Zheng, X.,P. C. Bevilacqua. 2004. Activation of the protein kinase PKR by short double-stranded RNAs with single-stranded tails. *RNA*, 10, 1934-1945.

Zu, T., B. Gibbens, N. S. Doty, M. Gomes-Pereira, A. Huguet, M. D. Stone, J. Margolis, M. Peterson, T. W. Markowski, M. A. Ingram, Z. Nan, C. Forster, W. C. Low, B. Schoser, N. V. Somia, H. B. Clark, S. Schmechel, P. B. Bitterman, G. Gourdon, M. S. Swanson, M. Moseley,L. P. Ranum. 2011. Non-ATG-initiated translation directed by microsatellite expansions. *Proceedings of the Natureional Academy of Sciences*, 108, 260-265.

Zuccato, C., A. CiamMoleculara, D. Rigamonti, B. R. Leavitt, D. Goffredo, L. Conti, M. E. MacDonald, R. M. Friedlander, V. Silani, M. R. Hayden, T. Timmusk, S.

Sipione,E. Cattaneo. 2001. Loss of huntingtin-mediated BDNF gene transcription in Huntington's disease. *Science*, 293, 493-498.

Zuccato, C., M. Tartari, A. Crotti, D. Goffredo, M. Valenza, L. Conti, T. Cataudella, B. R. Leavitt, M. R. Hayden, T. Timmusk, D. Rigamonti,E. Cattaneo. 2003. Huntingtin interacts with RESEARCHT/NRSF to modulate the transcription of NRSE-controlled neuronal genes. *Nature Genetics*, 35, 76-83.

Zucker, B., R. Luthi-Carter, J. A. Kama, A. W. Dunah, E. A. Stern, J. H. Fox, D. G. Standaert, A. B. Young,S. J. Augood. 2005. Transcriptional dysregulation in striatal projection- and interneurons in a mouse model of Huntington's disease: neuronal selectivity and potential neuroprotective role of HAP1. *Human Molecular Genetics*, 14, 179-189.

Appendix

nature COMMUNICATIONS

ARTICLE

Received 7 Mar 2016 | Accepted 31 Aug 2016 | Published 11 Oct 2016

DOI: 10.1038/ncomms13087 OPEN

Increased global transcription activity as a mechanism of replication stress in cancer

Panagiotis Kotsantis[1], Lara Marques Silva[2], Sarah Irmscher[2], Rebecca M. Jones[1], Lisa Folkes[3], Natalia Gromak[2] & Eva Petermann[1]

Cancer is a disease associated with genomic instability that often results from oncogene activation. This in turn leads to hyperproliferation and replication stress. However, the molecular mechanisms that underlie oncogene-induced replication stress are still poorly understood. Oncogenes such as $HRAS^{V12}$ promote proliferation by upregulating general transcription factors to stimulate RNA synthesis. Here we investigate whether this increase in transcription underlies oncogene-induced replication stress. We show that in cells overexpressing $HRAS^{V12}$, elevated expression of the general transcription factor TATA-box binding protein (TBP) leads to increased RNA synthesis, which together with R-loop accumulation results in replication fork slowing and DNA damage. Furthermore, overexpression of TBP alone causes the hallmarks of oncogene-induced replication stress, including replication fork slowing, DNA damage and senescence. Consequently, we reveal that increased transcription can be a mechanism of oncogene-induced DNA damage, providing a molecular link between upregulation of the transcription machinery and genomic instability in cancer.

[1] Institute of Cancer and Genomic Sciences, University of Birmingham, Edgbaston, Birmingham B15 2TT, UK. [2] Sir William Dunn School of Pathology, University of Oxford, South Parks Road, Oxford OX1 3RE, UK. [3] Department of Oncology, CRUK/MRC Oxford Institute for Radiation Oncology, University of Oxford, Old Road Campus Research Building, Roosevelt Drive, Oxford OX3 7DQ, UK. Correspondence and requests for materials should be addressed to N.G. (email: natalia.gromak@path.ox.ac.uk) or to E.P. (email: e.petermann@bham.ac.uk).

Cancer is a disease of genomic instability, characterized by high mutation rates and genomic rearrangements that ultimately drive aggressiveness and resistance to therapy[1]. One of the mechanisms proposed to cause genomic instability in cancer is replication stress, which occurs when DNA replication fork progression in S phase slows or stalls. This leads to collapse of forks into DNA double-strand breaks (DSBs), as well as incomplete sister chromatid separation in the following mitosis[2]. Markers of spontaneous replication stress are found in tumour samples and cells expressing active oncogenes, and replication stress promotes chromosomal instability, the most common form of genomic instability in sporadic cancers[3–6]. Spontaneous replication stress is therefore increasingly regarded as a central feature of cancer cells and there is much interest in specifically targeting this phenotype for cancer therapy[7]. However, progress in this field is hindered, because the molecular mechanisms underlying spontaneous replication stress in cells are still largely unknown. This impairs our ability to investigate replication stress *in vitro* and *in vivo*, and to identify potential biomarkers or therapeutic targets.

How can mechanisms of spontaneous replication stress be identified? The overexpression of oncogenes such as RAS, MOS, MYC, CDC25A or CYCLIN E is sufficient to induce replication stress in cultured cells[4,5,8–11]. These oncogenes all act in the growth factor signalling pathways that stimulate proliferation by promoting cell growth and division. Molecular changes associated with increased proliferation are therefore prime candidates for causing replication stress. Indeed, we and others have reported that CYCLIN E-induced replication stress results from accelerated S-phase entry and increased replication initiation during S phase[9,12]. So far, however, little attention has been paid to the fact that oncogenes such as RAS and MYC do not only activate the cell cycle machinery but also promote cell growth through the activation of transcription and protein translation[13–16]. Overexpressed c-MYC acts as a 'universal amplifier', stimulating transcription by all three RNA polymerases[14,17,18]. Oncogenic RAS promotes transcription through the mitogen-activated protein kinase extracellular signal-regulated kinase (ERK), which activates transcription factors such as TIFIA (RRN3), UBTF, TIFIIIB (BRF1) and TATA-box binding protein (TBP)[14,15,19], and can also promote transcription through other factors such as TERT[20]. Moreover, components of the general transcription machinery itself are found mutated, differentially expressed and deregulated across a variety of cancers[14,21].

Upregulation of transcription in cancer cells has the potential to be a direct cause of replication stress as interference between transcription and replication leads to replication fork slowing and genomic instability[22,23]. This can result from direct collisions or torsional stress between the two active protein complexes[24]. Another important source of replication stress is the collision of the replication machinery with RNA–DNA hybrids (R-loops) where the nascent RNA has re-annealed with the template[25]. Deregulated transcription is a common feature of cancers, but its importance for replication stress and consequent genomic instability has not been examined.

Here we use $HRAS^{V12}$ overexpression[5,26] to investigate whether the oncogene-induced increase in transcription is a mechanism of endogenous replication stress. We report that increased transcription activity in cells expressing $HRAS^{V12}$ causes replication stress, as high levels of RNA synthesis and transcription intermediates interfere with replication fork progression. Replication stress in these cells depends on R-loop accumulation and on increased expression of the general transcription factor TBP. Importantly, overexpression of TBP alone induces replication stress and genomic instability. Our data suggest that increased transcription activity is a mechanism contributing to replication stress in cancer.

Results

$HRAS^{V12}$ increases nascent RNA synthesis and R-loop formation. To investigate whether increased transcription activity contributes to oncogene-induced replication stress, we used immortalized human fibroblasts that have been stably transfected with pBabe-$HRAS^{V12}$-ER^{TAM}, to express tamoxifen-inducible $HRAS^{V12}$ (BJ-hTert $HRASV12^{ER\text{-}TAM}$)[27]. $HRAS^{V12}$-ER^{TAM} has been well characterized as a system for inducing oncogenic RAS[26,28,29]. Addition of 4-hydroxytamoxifen (4OHT) led to $HRAS^{V12}$ accumulation, which activated mitogen-activated protein kinase signalling as evidenced by ERK1/2 phosphorylation (Fig. 1a). Cells proliferated slightly faster for up to 6 days followed by growth arrest, which was previously shown to result from DNA damage-induced apoptosis and senescence[4,5] (Supplementary Fig. 1a–e). When using immortalized BJ fibroblasts that were not expressing 4OHT-inducible $HRAS^{V12}$, 4OHT treatment did not affect any of the phenotypes investigated in this study (Supplementary Fig. 1f–i).

First, we investigated the effect of oncogenic HRAS on global transcription activity. For this, we quantified nascent RNA synthesis using nuclear incorporation of the modified RNA precursor 5-ethynyluridine (EU) for 1 h (Fig. 1b,c). RNA synthesis increased rapidly after $HRAS^{V12}$ induction and was elevated more than twofold after 48 and 72 h $HRAS^{V12}$ induction (Fig. 1d). RNA synthesis after $HRAS^{V12}$ induction was higher than the highest activity observed in control cells (Fig. 1d), arguing against changes in cell cycle distribution as the sole explanation for the increase. Seventy-two hours of $HRAS^{V12}$ induction was used for most subsequent experiments.

We next used the S9.6 antibody that detects RNA/DNA hybrids[30,31], to test whether R-loop formation was increased in $HRAS^{V12}$-overexpressing cells. First, we performed slot blot analysis of isolated genomic DNA[32], which revealed a threefold increase in R-loops after 72 h $HRAS^{V12}$ induction (Fig. 1e,f). The S9.6 signal could be removed by treatment with recombinant RNase H, supporting that it was specific to R-loops. We then used S9.6 immunoprecipitation of RNA/DNA hybrids from cells (DNA immunoprecipitation (DIP))[33] after 72 h $HRAS^{V12}$ induction, to investigate the distribution of R-loops across RAS target versus control genes (Fig. 2a–f). We used quantitative PCR (qPCR) to quantify R-loop distribution on *DUSP6*, *SPRY2* and *C-FOS*, genes that are upregulated by activated RAS (Fig. 2a–c and Supplementary Fig. 2a). We observed an increased R-loop formation on the promoter-proximal and selected intron regions of all three genes. DIP analysis of *C-FOS* showed that in line with previously described R-loop accumulation in actively transcribed genes[33], R-loops were significantly increased over the transcribed regions of the gene (Fig. 2c and also see Supplementary Table 1 for PCR primer sequences). We also quantified R-loop formation on non-RAS target control genes *GAPDH1*, *ACTB* (β-ACTIN) and *ACTG* (γ-ACTIN). We observed no increase in R-loops across any of these genes (Fig. 2d–f). RNase H treatment confirmed that DIP specifically detected R-loops (Fig. 2a–f). RNase A treatment confirmed that DIP signal was not due to annealing of free RNA species to DNA during sample preparation or to S9.6 antibody recognizing double-stranded RNA (Supplementary Fig. 2b–e). These data support that stimulation of transcription by $HRAS^{V12}$ results in increased R-loop formation.

$HRAS^{V12}$ causes replication stress and G1 53BP1 foci. $HRAS^{V12}$ overexpression caused replication fork slowing after

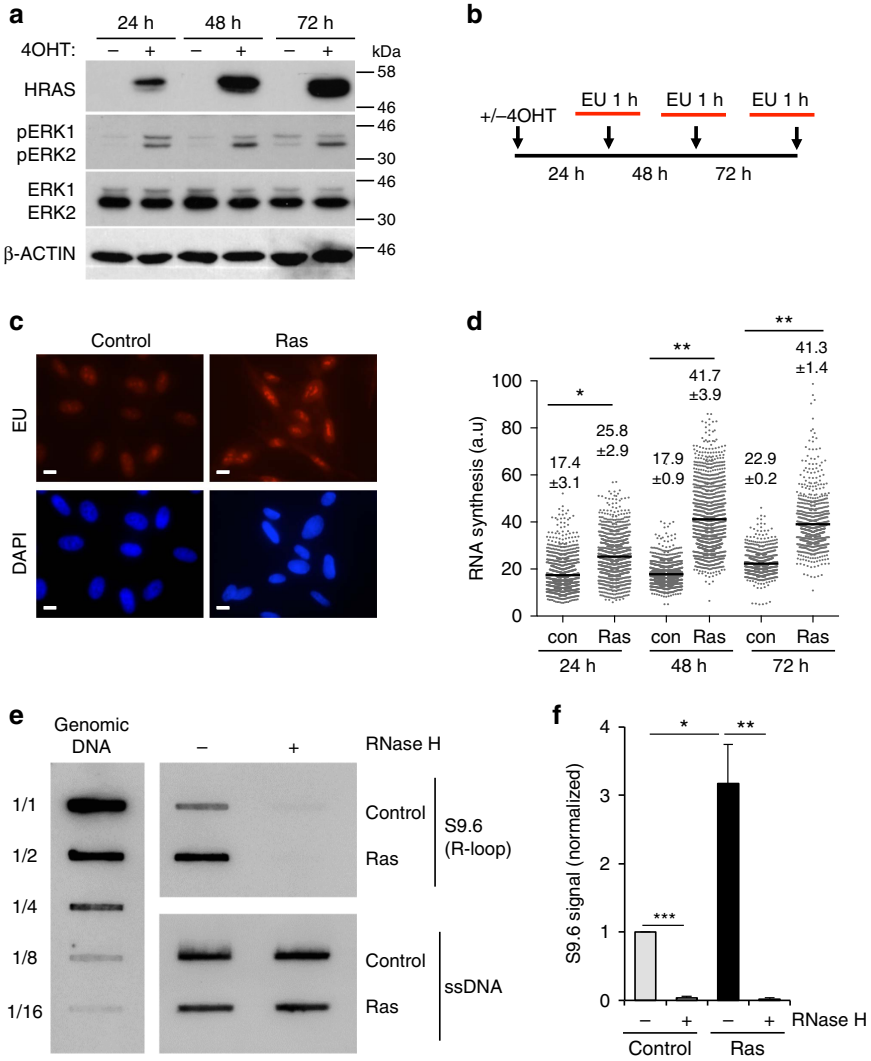


Figure 1 | HRAS^{V12} overexpression increases transcription activity. (**a**) Protein levels of HRAS, pERK1/2, ERK1/2 and β-ACTIN in BJ-HRASV12 cells after RAS induction for the times indicated. (**b**) EU incorporation (1 h) was used to measure nascent RNA synthesis after RAS induction for the times indicated. (**c**) Representative images of EU staining (red) after RAS induction for 72 h. (**d**) Quantification of nuclear EU intensity after RAS induction for the times indicated. $N=4$ (48 and 72 h), $N=5$ (24 h). (**e**) RNA/DNA hybrid slot blot of genomic DNA from BJ-HRAS^{V12} cells after RAS induction for 72 h, ± RNase H. S9.6 antibody was used to detect RNA/DNA hybrids (top panel) with single-strand DNA antibody (bottom panel) as a loading control. Serial dilutions of genomic DNA (1/1 = 4 μg) were probed with S9.6 antibody for standards (left panel). (**f**) Fold enrichment in RNA/DNA hybrids compared with control. $N=3$. Means ± s.e.m. (bars) are shown. Student's *t*-test, $^{*}P<0.05$, $^{**}P<0.01$ and $^{***}P<0.001$. Scale bars, 10 μm.

48 h, indicative of replication stress (Fig. 3a–c) and induced phosphorylation of replication stress response factors RPA32 (serine 33) and CHK1 (serine 345) (Fig. 3d,e). These modifications were already evident at 24 h HRAS^{V12} induction, which may also reflect the proposed role for RPA and ATR in the transcription stress response[34]. As expected, fork slowing was associated with nuclear foci formation of DNA damage markers γH2AX and 53BP1 (Fig. 3f,g). One consequence of replication fork slowing is mitotic entry with under-replicated DNA, which leads to micronuclei formation and appearance of 53BP1 bodies in the following G1 phase[35,36]. Accordingly, 53BP1 foci increased after 4 days HRAS^{V12} induction and were mostly found in G1 cells (Fig. 3h). Similarly, micronuclei formation peaked at 4 days HRAS^{V12} induction (Fig. 3i). This suggests that HRAS^{V12}-induced DNA damage requires mitotic progression, as was previously reported for CYCLIN E[10]. Four days of HRAS^{V12} induction was used for subsequent investigation of 53BP1 foci formation.

Transcription promotes HRAS^{V12}-induced replication stress. To test whether increased transcription causes replication stress in cells harbouring HRAS^{V12}, we transiently inhibited RNA synthesis using small molecule inhibitors before measuring replication fork progression. To minimize effects on gene expression, incubations were kept short at 100 min for triptolide and 5,6-dichloro-1-β-D-ribofuranosyl-1H-benzimidazole (DRB) and 4 h for α-amanitin (Fig. 4a). These treatments inhibited ongoing RNA synthesis (Fig. 4b) but did not affect protein levels of HRAS^{V12} or levels of proteins with short half lives such as CYCLIN B1 and p53 (Supplementary Fig. 3a,b). All three transcription inhibitors increased replication fork speeds specifically in the presence of HRAS^{V12} (Fig. 4c–f). Although such short incubations with transcription inhibitors may not be sufficient to reverse all effects of transcription, these data suggest that RAS-induced replication stress is promoted by active RNA synthesis. DRB and triptolide rescued replication more effectively than α-amanitin, suggesting that effective inhibition of early

stages of transcription removes more obstacles to replication forks than does inhibiting transcription during elongation (Fig. 4f).

We previously observed that increased CDK activity and new origin firing underlies replication stress in cells overexpressing CYCLIN E[12]. In contrast, inhibiting new origin firing using CDK inhibitor roscovitine could not rescue HRAS^{V12}-induced replication fork slowing (Supplementary Fig. 3c–e). Together, our data suggest that replication stress in cells expressing HRAS^{V12} is promoted by transcription but not by CDK activity.

To test whether HRAS^{V12}-induced DNA damage was transcription dependent, we first incubated cells with DRB for 100 min and stained for 53BP1 24 h later. DRB reduced 53BP1 foci formation; however, its impact was limited by the short incubation time (Supplementary Fig. 3f,g). We therefore used γH2AX chromatin immunoprecipitation (ChIP) to test whether transcription and R-loop formation was associated with DNA damage in cells harbouring HRAS^{V12}. Indeed, we observed an increase in γH2AX signal over the transcribed promoter-proximal region of the *C-FOS* gene, correlating with strong induction of R-loops, 72 h after HRAS^{V12} induction (Figs 2c and 4g). This γH2AX induction was replication dependent, as it could be prevented by blocking replication with Aphidicolin (Fig. 4g and Supplementary Fig. 1d). In contrast, we did not detect an increase in replication-dependent γH2AX levels across the intron 1 region of the *β-ACTIN* gene (Fig. 4h). This suggests that HRAS^{V12} triggers R-loop-associated DNA damage that also depends on replication.

R-loops promote HRAS^{V12}-induced replication stress. We next decided to further investigate the role of R-loops in HRAS^{V12}-induced replication stress. We used transient transfection to express green fluorescent protein (GFP)-tagged human RNaseH1, an enzyme that degrades RNA/DNA hybrids on overexpression[37] (Fig. 5a,b). Interestingly, we observed that protein and messenger RNA levels of endogenous RNaseH1 were elevated in cells overexpressing HRAS^{V12}, suggesting an increased requirement for R-loop processing activities (Fig. 5b,c). The specificity of RNaseH1 antibody was verified using small interfering RNA (siRNA) depletion of RNaseH1 (Supplementary Fig. 4a). Overexpression of GFP-RNaseH1 reduced R-loop levels in the nucleus, as indicated by S9.6 immunostaining (Fig. 5d,e and also see Supplementary Fig. 5 for validation of immunostaining method). As the expression construct contains the RNaseH1 mitochondrial targeting sequence, mitochondrial R-loops were also reduced (Fig. 5d). RNaseH1 overexpression effectively improved replication fork progression in cells harbouring HRAS^{V12} (Fig. 5f,g). GFP-RNaseH1 also decreased HRAS^{V12} induction of 53BP1 foci (Fig. 5h,i) and DSB induction measured by pulse-field gel electrophoresis (Supplementary Fig. 4b,c). GFP-RNaseH1 overexpression did not affect hydroxyurea-induced 53BP1 foci formation or the cell cycle profile (Supplementary Fig. 4d,e). These data support that HRAS^{V12}-induced replication stress is promoted by the presence of R-loops.

We also tested whether some of the replication stress caused by HRAS^{V12} might be due to ribonucleotide (NTP) depletion as a result of increased RNA synthesis. Supplementing growth medium with ribonucleosides improved fork speeds and reduced 53BP1 foci formation in the presence of HRAS^{V12} (Supplementary Fig. 6a–c). However, HPLC quantification showed no reduction in NTP levels or dNTP levels as detectable (Supplementary Fig. 6d,e). This is in line with previous reports demonstrating that nucleoside supplementation can rescue fork progression even in the absence of detectable nucleotide depletion[38]. Our data thus suggest that transcription contributes to HRAS^{V12}-induced replication stress predominantly via direct conflicts with transcription machinery and R-loops (Fig. 5j).

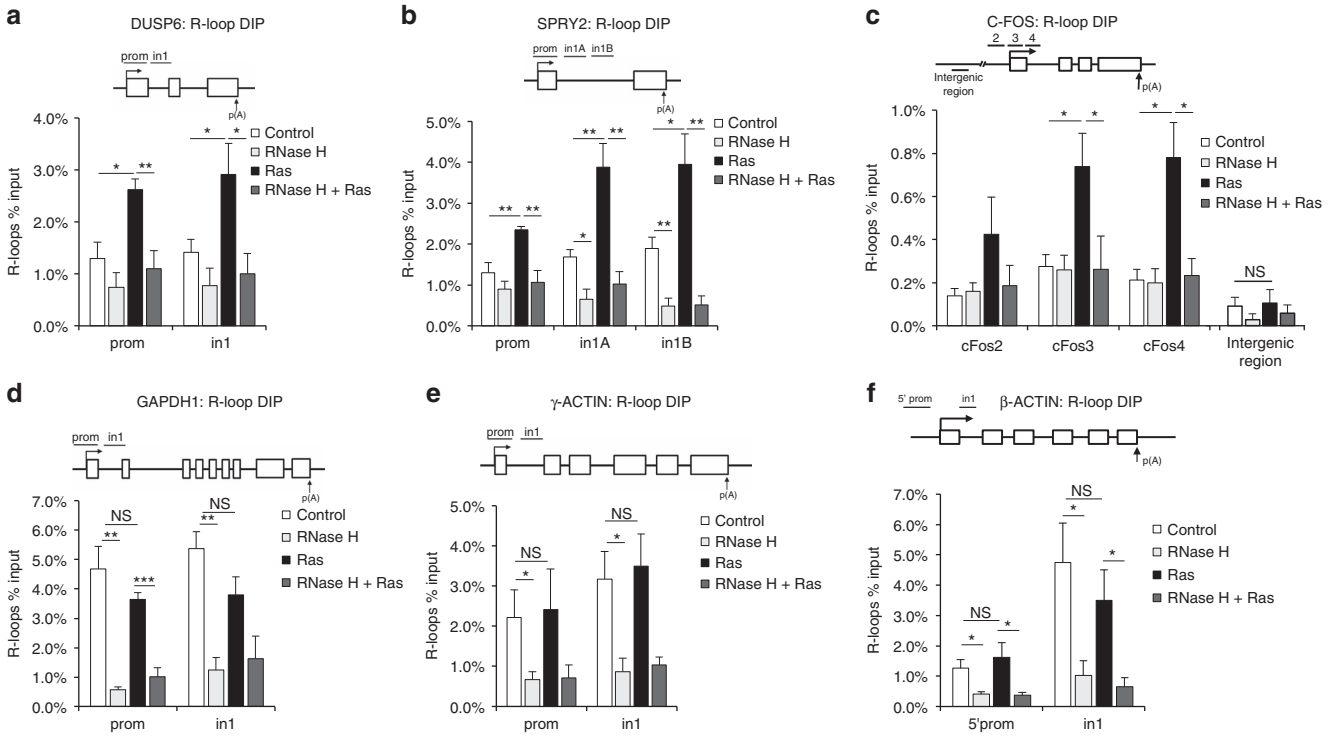


Figure 2 | HRAS^{V12} overexpression increases R-loop formation. DIP analysis of R-loop induction on the *DUSP6* (**a**), *SPRY2* (**b**), *C-FOS* (**c**), *GAPDH* (**d**), *γ-ACTIN* (**e**) and *β-ACTIN* (**f**) genes in BJ-HRAS^{V12} cells after RAS induction for 72 h. Intergenic region upstream of *C-FOS* gene (**c**) served as a background control. Values are percentage of input. $N=5$. Primer positions are shown in upper panels of **a**–**f** (white boxes represent exons). Means ± s.e.m. (bars) are shown. Student's *t*-test, $^{*}P<0.05$, $^{**}P<0.01$ and $^{***}P<0.001$. NS, nonsignificant.

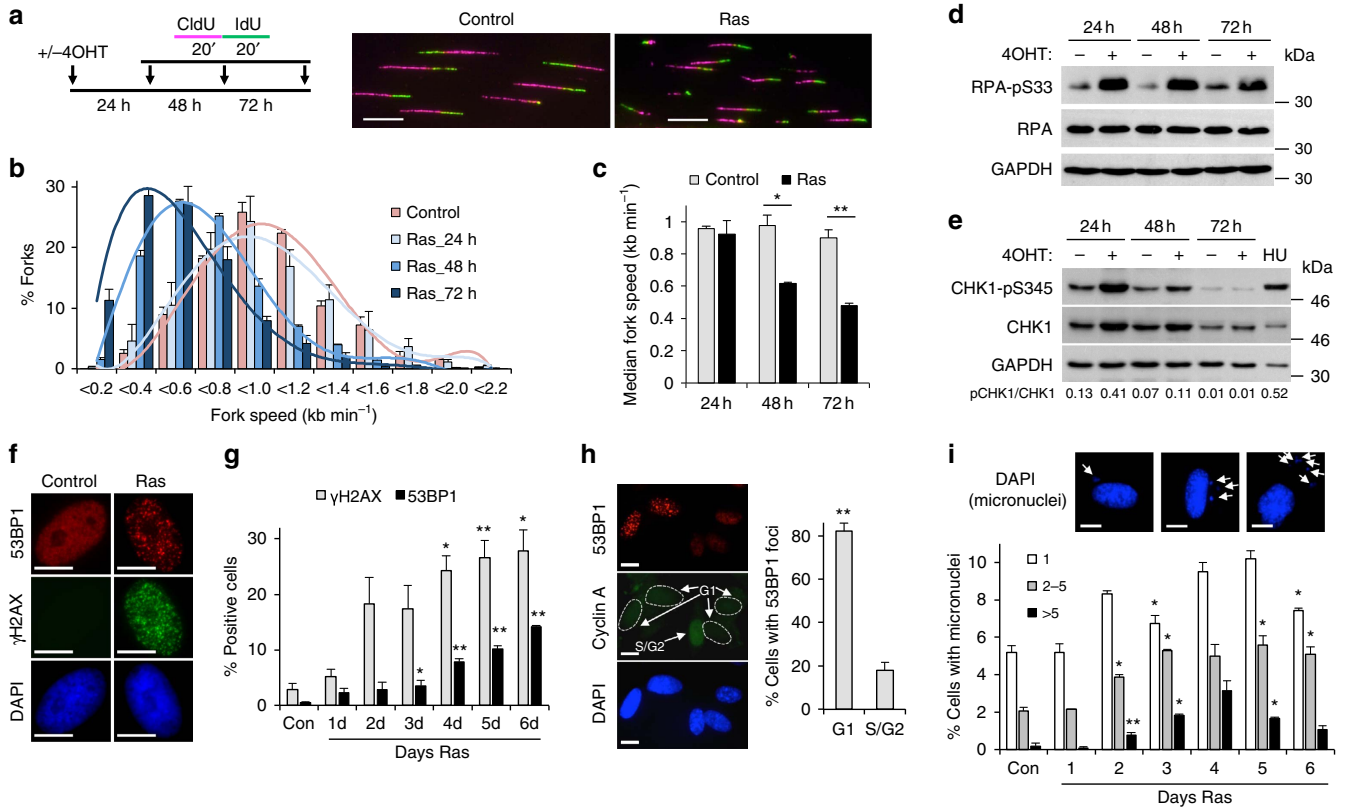


Figure 3 | HRAS^{V12} causes replication stress and genomic instability. (**a**) Top: DNA fibre labelling was performed in BJ-HRAS^{V12} cells after RAS induction for 24, 48 and 72 h. Bottom: representative images of DNA fibres after 72 h RAS induction. (**b**) Distributions of replication fork speeds after RAS induction. $N=3$. (**c**) Median replication fork speeds after RAS induction. $N=3$. (**d**) Protein levels of phospho-S33 RPA32, RPA32 and GAPDH after RAS induction. (**e**) Protein levels of phospho-S345 CHK1, CHK1 and GAPDH after RAS induction. Hydroxyurea (HU, 2 mM for 24 h) was used as a positive control. (**f**) Representative images of γH2AX and 53BP1 foci after RAS induction for 96 h. (**g**) Percentages of cells containing more than 8 γH2AX or 53BP1 foci after RAS induction for 96 h. $N=2$. Asterisks compare with control. (**h**) Cell cycle distribution of 53BP1-positive cells after 96 h RAS induction as determined by co-staining with Cyclin A. Right panel: representative images; left panel: quantification. $N=3$. (**i**) Representative images of cells with micronuclei and percentages of cells containing micronuclei after RAS induction for 1–6 days. $N=2$. Asterisks compare with control. Means ± s.e.m. (bars) are shown. Student's *t*-test, $^{*}P<0.05$ and $^{**}P<0.01$. Scale bars, 10 μm.

HRAS^{V12}-induced replication stress depends on TBP. We next turned our attention to transcription factors that may promote RNA synthesis and therefore replication stress downstream of HRAS^{V12}. TBP expression is induced by RAS signalling in a number of human, murine and *Drosophila* cell types[19,39,40], and TBP mRNA and protein levels were accordingly increased after HRAS^{V12} induction (Fig. 6a–c). We used siRNA depletion to test whether TBP promotes HRAS^{V12}-induced replication stress (Fig. 6d,e). TBP depletion decreased nascent RNA synthesis in HRAS^{V12}-expressing cells (Fig. 6f). Importantly, TBP depletion also prevented HRAS^{V12}-induced fork slowing (Fig. 6g,h) and TBP-depleted cells displayed fewer HRAS^{V12}-induced 53BP1 foci (Fig. 6i). Similar results were obtained using a different siRNA sequence targeting TBP (Supplementary Fig. 7a–e). We noticed that the alternative TBP siRNA also reduced RNA synthesis and replication fork speeds in control cells, suggesting that generally low RNA synthesis may affect replication (Supplementary Fig. 7c,d). TBP depletion did not prevent hydroxyurea-induced 53BP1 foci (Supplementary Fig. 6f) and did not affect the cell cycle profile of 53BP1 foci-positive cells, suggesting that the reduction in foci was not due to a G2 arrest (Supplementary Fig. 7g).

To test whether TBP acts in the same pathway as ongoing RNA synthesis, we combined TBP siRNA with DRB treatment. Combining both treatments rescued replication fork progression and 53BP1 foci formation to a similar extent as TBP depletion alone (Fig. 6j,k and Supplementary Fig. 8a). To further test whether ribonucleoside addition was affecting replication stress via a different pathway to transcription[41], we combined TBP siRNA and DRB with nucleoside supplementation. Compared with transcription inhibition alone, exogenous nucleosides had no additional effect on either the rescue of replication fork progression or the reduction in 53BP1 foci (Supplementary Fig. 8b–e).

These data suggest that the main pathway of HRAS^{V12}-induced replication stress is via upregulation of transcription, and that TBP is involved in this upregulation of transcription and the replication fork slowing and genomic instability that result from it (Fig. 6l).

TBP overexpression alone causes replication stress. We reasoned that if TBP was a downstream effector in HRAS^{V12}-induced replication stress, then overexpression of TBP alone should cause replication stress. We therefore generated BJ-hTert fibroblasts for doxycycline-inducible overexpression of TBP (BJ-TBPind; Fig. 7a). Inducing TBP overexpression over several days led to a steady increase in RNA synthesis activity as measured by EU incorporation (Fig. 7b,c). Importantly, the increase in transcription activity during TBP overexpression was accompanied by replication fork slowing, consistent with replication stress (Fig. 7d and Supplementary Fig. 9a).

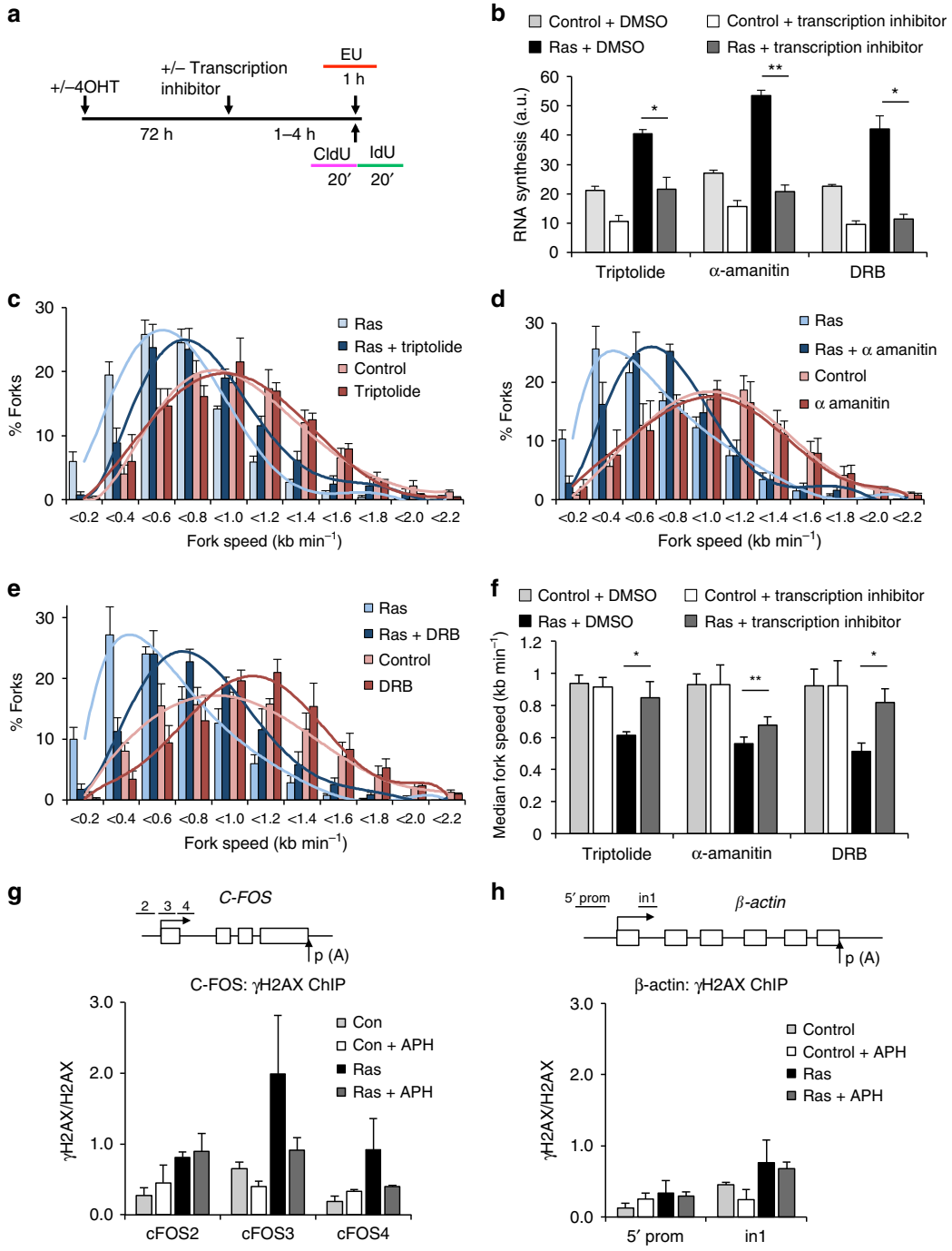


Figure 4 | HRAS^{V12}-induced replication stress is promoted by ongoing transcription. (**a**) Seventy-two hours after RAS induction, BJ-HRAS^{V12} cells were incubated with transcription inhibitors triptolide, α-amanitin, DRB or dimethylsulfoxide (DMSO) for 100 min to 4 h before and during EU or DNA fibre labelling. (**b**) Quantification of nascent RNA synthesis by EU incorporation in cells treated with transcription inhibitors 72 h after RAS induction. $N=2$ (con + triptolide, con + DRB), $N=3$ (all other samples). (**c**–**e**) Distributions of replication fork speeds in cells treated with transcription inhibitors 72 h after RAS induction. $N=3$ (DRB, control + α-amanitin, control + triptolide), $N=5$ (other samples). (**f**) Median replication fork speeds in cells treated with transcription inhibitors 72 h after RAS induction. (**g**) ChIP analysis of γH2AX versus total H2AX levels on the promoter of the *C-FOS* gene 72 h after RAS induction. + APH samples were treated with 0.5 μM Aphidicolin for 2 h before ChIP. $N=3$. (**h**) ChIP analysis of γH2AX versus total H2AX levels on the *β-ACTIN* gene 72 h after RAS induction as in g. $N=3$. qPCR primer positions are shown in upper panels. Means ± s.e.m. (bars) are shown. Student's *t*-test, $^{*}P<0.05$ and $^{**}P<0.01$.

Similar results were observed in human MRC5 fibroblasts overexpressing TBP (Supplementary Fig. 9b–d). To test whether TBP-induced fork slowing depended on ongoing transcription, we treated TBP-overexpressing cells with DRB for 100 min, to inhibit nascent RNA synthesis (Fig. 7e). DRB treatment rescued the TBP-induced fork slowing (Fig. 7f and Supplementary Fig. 9e), supporting that TBP overexpression causes replication stress through nascent RNA synthesis.

We next characterized the effect of TBP overexpression on DNA damage, genomic instability and proliferation. Similar to

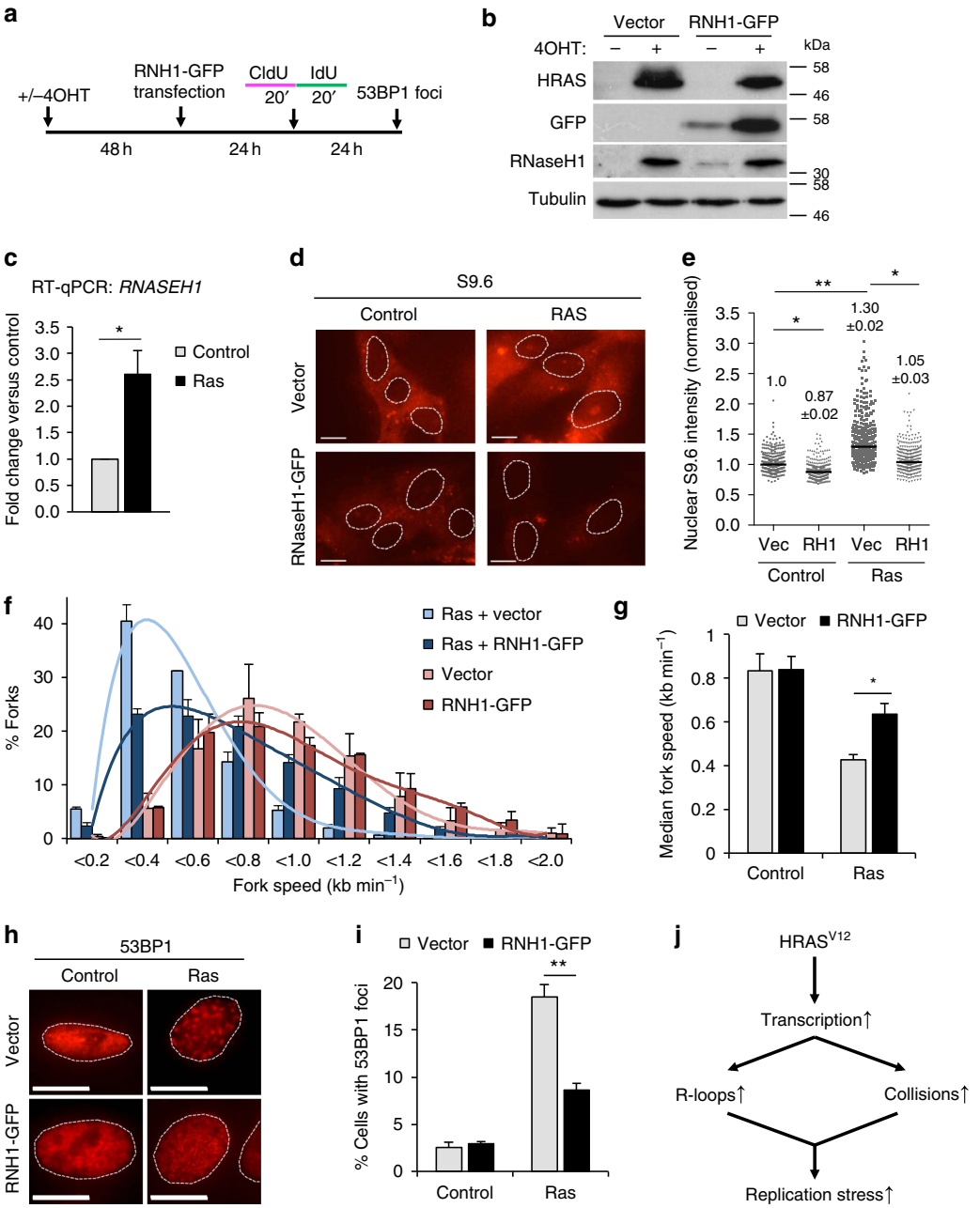


Figure 5 | HRAS^{V12} induces replication stress through R-loop formation. (**a**) After RAS induction for 48 h, BJ-HRAS^{V12} cells were transfected with pCMV6-AC-RNaseH1-GFP or empty vector and processed for DNA fibre analysis or 53BP1 foci staining after 24 or 48 h. (**b**) Protein levels of HRAS, GFP, RNaseH1 and TUBULIN (loading control) 72 h after RAS induction and 24 h after transfection with pCMV6-AC-RNaseH1-GFP. (**c**) *RNASEH1* mRNA quantification by quantitative reverse transcriptase–PCR after 72 h RAS induction. *RNASEH1* mRNA levels were normalized to *GAPDH* and control. (**d**) Representative images of S9.6 immunostaining ± pCMV6-AC-RNaseH1-GFP 72 h after RAS induction. (**e**) Quantification of nuclear S9.6 intensity ± pCMV6-AC-RNaseH1-GFP 72 h after RAS induction. $N=3$. (**f**) Distribution of fork speeds ± pCMV6-AC-RNaseH1-GFP 72 h after RAS induction. $N=3$. (**g**) Median replication fork speeds ± pCMV6-AC-RNaseH1-GFP 72 h after RAS induction. $N=3$. (**h**) Representative images of cells with 53BP1 foci, ± pCMV6-AC-RNaseH1-GFP 96 h after RAS induction. (**i**) Percentage of cells displaying more than 8 53BP1 foci, ± pCMV6-AC-RNaseH1-GFP 96 h after RAS induction. $N=3$. (**j**) Model of how HRAS^{V12} causes replication stress via increasing transcription. Means ± s.e.m. (bars) are shown. Student's *t*-test, $^{*}P<0.05$ and $^{**}P<0.01$. Scale bars, 10 μm.

HRAS^{V12}-expressing cells, cells overexpressing TBP displayed increased formation of micronuclei and 53BP1 foci (Fig. 7g), and 53BP1 foci were mostly found in G1 cells (Fig. 7h). Finally, TBP overexpression eventually led to growth arrest and senescence, similar to that caused by HRAS^{V12} (Fig. 7i and Supplementary Fig. 9f,g). In agreement with this, p53 levels were increased after TBP induction (Fig. 7a and also see Supplementary Fig. 10 for original images of western blottings).

Thus, our data show that TBP alone is able to increase transcription activity and cause replication stress with features that resemble oncogene-induced replication stress.

TBP expression and replication stress in cancer. Finally, we investigated the relationship of *TBP* and *RNASEH1* mRNA expression with oncogenes and replication stress markers in tumour samples using The Cancer Genome Atlas (TCGA) data

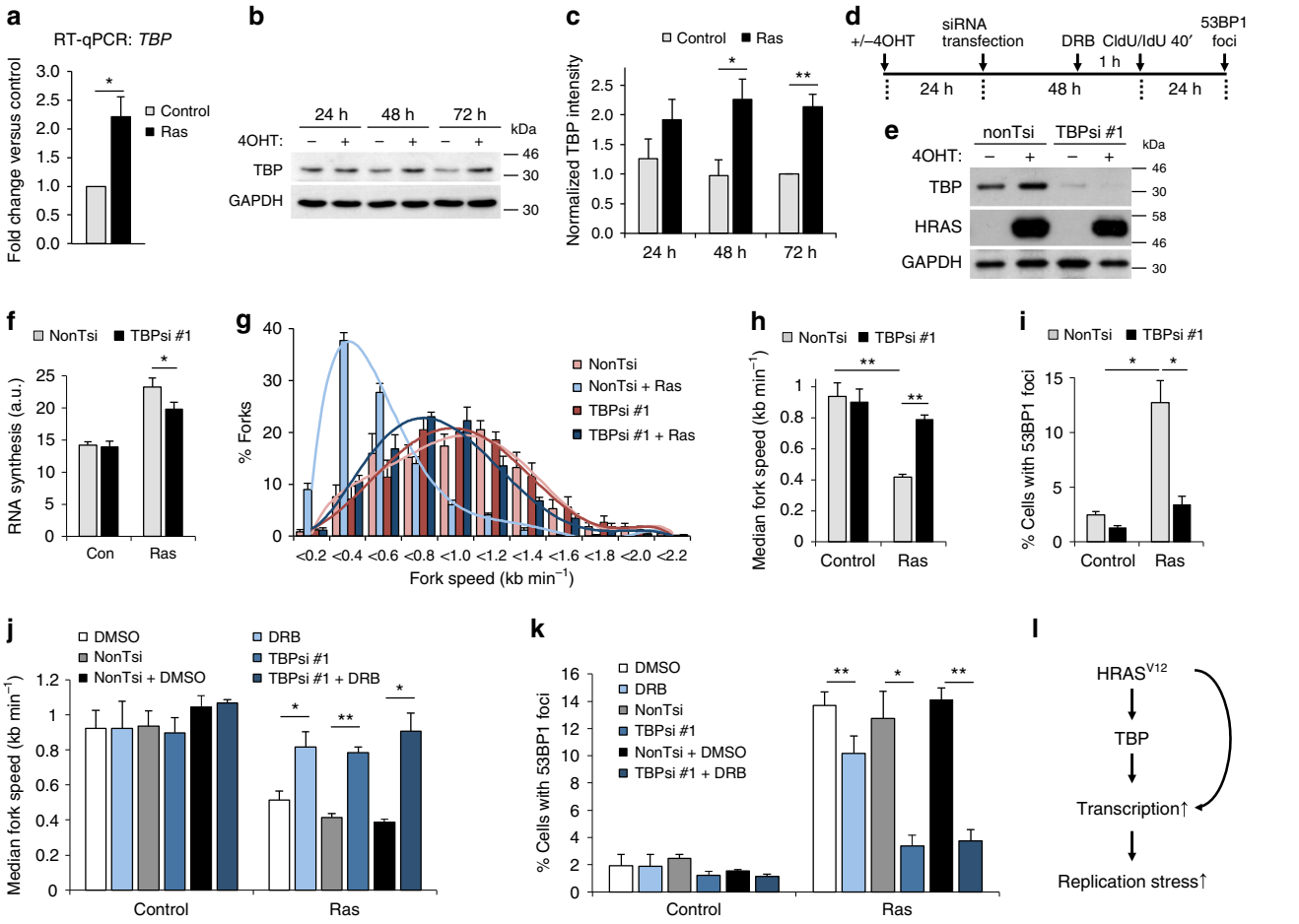


Figure 6 | TBP is an effector in HRAS^{V12}-induced replication stress. (**a**) *TBP* mRNA quantification by quantitative reverse transcriptase-PCR in BJ-HRAS^{V12} cells 72 h after RAS induction. *TBP* mRNA levels were normalized to *GAPDH* and control. (**b**) Protein levels of TBP and β-ACTIN after RAS induction for the times indicated. (**c**) Densitometry quantification of TBP levels based on western blotting as in **b** after RAS induction for the times indicated. Values were normalized to 72 h control. $N=3$ (24 and 48 h), $N=6$ (72 h). (**d**) Twenty-four hours after RAS induction, cells were transfected with TBP siRNA (TBPsi #1) or control siRNA (nonTsi). Cells were processed for DNA fibre analysis or western blotting 48 h later and for 53BP1 staining 24 h later. (**e**) Protein levels of TBP, HRAS and GAPDH (loading control) 72 h after RAS induction and 48 h after siRNA transfection. (**f**) Quantification of nascent RNA synthesis by EU incorporation ± TBPsi #1 72 h after RAS induction. $N=3$. (**g**) Distribution of replication fork speeds ± TBPsi #1 72 h after RAS induction. $N=3$. (**h**) Median replication fork speeds ± TBPsi #1 72 h after RAS induction. $N=3$. (**i**) Percentages of cells containing more than eight 53BP1 foci, ± TBPsi #1 96 h after RAS induction. $N=3$. (**j**) Median replication fork speeds in cells treated with TBPsi #1 and DRB 72 h after RAS induction, compared with TBPsi #1 or DRB alone. $N=3$. (**k**) Percentages of cells treated with TBPsi #1 and DRB containing more than eight 53BP1 foci after 96 h after RAS induction, compared with TBPsi #1 or DRB alone. $N=3$. (**l**) Model for the role of TBP in HRAS^{V12}-induced replication stress. Means ± s.e.m. (bars) are shown. Student's *t*-test, $^{*}P<0.05$, $^{**}P<0.01$ and $^{***}P<0.001$.

sets[42–50] (Table 1). In a number of cancers, *TBP* or *RNASEH1* expression correlated positively with expression of *RAS* or *MYC* oncogenes and the *CHEK1* or *CHEK2* checkpoint kinases, which are activated by oncogene-induced replication stress[4,5]. These data support that the relationship between oncogenes, TBP, R-loops and replication stress could be present in cancer tissues.

Discussion

We report that increased transcription is a new mechanism of oncogene-induced replication stress. Overexpression of oncogenic HRAS^{V12} increases RNA synthesis and R-loop formation, and this directly contributes to replication stress induced by HRAS^{V12}. HRAS^{V12}-induced replication stress is mediated by the general transcription factor TBP, which is a downstream target of RAS signalling. Accordingly, overexpression of TBP itself causes replication stress and genomic instability. The transcription machinery is frequently deregulated in cancer cells[21] and our data suggest that transcription-associated replication stress could be an important mechanism promoting genomic instability in cancer (Fig. 7j).

Markers of replication stress are observed in early tumours and cancers and other conditions of high proliferation, such as stem cell reprogramming or viral infection[51,52]. Recent studies have provided first candidates for the mechanisms causing such endogenous replication stress. So far, all of these have involved deregulation of the cell cycle, including re-replication[53] and premature or increased replication initiation, resulting in depletion of nucleotides or replication enzymes[9,10,12]. In addition, reactive oxygen species play important roles in DNA damage in cancer, but whether they can cause replication stress is still uncertain[54,55]. Our data support that increased transcription is another pathway causing endogenous replication stress, which may act in parallel or independently of cell cycle deregulation. HRAS^{V12} overexpression did increase the density of active replication origins, which could be reversed by CDK inhibition.

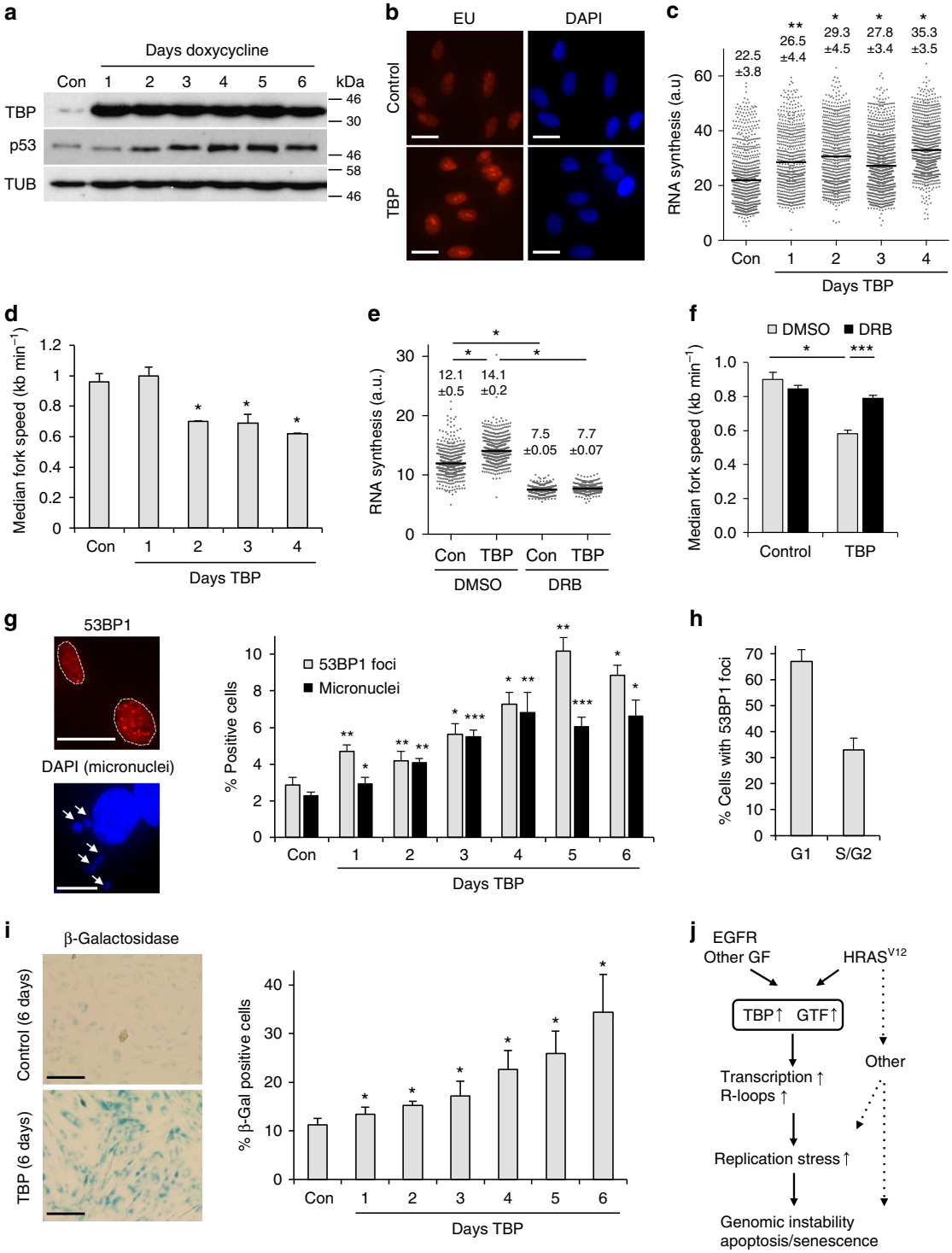


Figure 7 | TBP overexpression causes replication stress and senescence. (**a**) Protein levels of TBP, p53 and TUBULIN (loading control) in BJ-TBPind cells treated with doxycycline for 1–6 days, to induce TBP overexpression. No doxycycline was used as a control. (**b**) Nascent RNA synthesis as measured by EU incorporation after TBP induction for 3 days. Scale bars, 10 μm. (**c**). Quantification of nuclear EU intensity after TBP induction for 1–4 days. $N=4$ (con, day 1, 2 and 4), $N=5$ (day 3). (**d**) Median replication fork speeds in after TBP induction. $N=2$ (day 1, 2 and 4), $N=4$ (con, day 3). Asterisks compare with control. (**e**) Nascent RNA synthesis as measured by EU incorporation in cells treated with DRB or dimethylsulfoxide (DMSO) (control) for 100 min, 72 h after TBP induction. $N=3$. (**f**) Median replication fork speeds in BJ-TBPind cells treated with DRB 72 h after TBP induction. $N=3$. (**g**). Percentage of cells displaying more than eight 53BP1 foci or micronuclei after TBP induction. Right panel: representative images of cells with 53BP1 foci and micronuclei. Asterisks compare with control. $N=2$–7. (**h**) Cell cycle distribution of 53BP1-positive cells 96 h after TBP induction as determined by co-staining with Cyclin A. $N=2$. (**i**) Representative images and percentages of β-galactosidase staining after TBP induction for 1–6 days. Scale bars, 100 μm. (**j**) Model of how HRASV^{12} and other growth factor oncogenes such as epidermal growth factor receptor (EGFR) induce replication stress by increasing transcription through TBP and other transcription factors. Additional mechanisms, such as reactive oxygen species, may also contribute to HRASV^{12}-induced DNA damage. Means ± s.e.m. (bars) are shown. Student's *t*-test, $^{*}P<0.05$, $^{**}P<0.01$ and $^{***}P<0.001$.

Table 1 | TCGA data sets showing correlation between mRNA expression of *TBP* or *RNASEH1* with oncogenes and replication stress markers.

TCGA data set	*MYC*	*RAS*	*CHEK1*	*CHEK2*
TBP mRNA expression correlation with:				
Colorectal adenocarcinoma		0.41 (5%)**	0.54 (1%)**	0.42 (2%)**
Glioblastoma multiforme	0.41 (5%)**	0.33 (10%)*	0.47 (1%)**	
Stomach adenocarcinoma		0.34 (10%)**	0.42 (5%)**	0.45 (1%)**
Lung adenocarcinoma		0.39 (1%)**	0.37 (1%)**	
Lung squamous cell carcinoma				0.32 (5%)*
Bladder urothelial carcinoma	0.46 (1%)**	0.34 (5%)*		
Acute myeloid leukemia	0.31 (5%)*			0.30 (10%)*
Skin cutaneous melanoma		0.31 (1%)**		
Pheochromocytoma and Paraganglioma	0.31 (5%)*	−0.36 (5%)*		
Liver hepatocellular carcinoma			0.45 (1%)**	
RNASEH1 mRNA expression correlation with:				
Colon adenocarcinoma			0.48 (1%)**	
Glioblastoma multiforme		0.71 (1%)**		
Lung adenocarcinoma		0.5 (1%)**	0.48 (1%)**	0.38 (5%)**
Lung squamous cell carcinoma		0.31 (5%)*	0.44 (1%)**	0.36 (1%)*
Breast invasive carcinoma		0.37 (10%)**	0.57 (1%)**	0.4 (5%)**
Pheochromocytoma and Paraganglioma		0.38 (10%)**		
Prostate adenocarcinoma		0.3 (10%)**		
Pancreatic adenocarcinoma			0.48 (1%)**	
Thyroid carcinoma		0.51 (5%)**		
Brain lower grade glioma		−0.37 (10%)**		

Blank cells, no correlation; *MYC, CMYC/NMYC; RAS, NRAS/HRAS/KRAS;* TBP, TATA-box binding protein.
Values: Pearson's coefficient *r* (0.3–0.5: positive correlation, 0.5–1: strong positive correlation, −0.3–−0.5: negative correlation, empty cell: no correlation). Values in brackets: top % of all genes; *$P<0.00005$ and **$P<0.000001$, Student's *t*-test.

However, CDK inhibition was unable to relieve $HRAS^{V12}$-induced replication fork slowing, suggesting that increased replication initiation is not a cause but a consequence of $HRAS^{V12}$-induced replication stress (Supplementary Fig. 3d,e). Our data suggest that $HRAS^{V12}$ causes replication stress by a mechanism that is different from oncogenes such as CYCLIN E and CDC25A[10,12].

Our data support that $HRAS^{V12}$ overexpression promotes accumulation of R-loops, which can be a major cause for replication fork slowing and DNA breakage[23]. An elegant study recently reported that loss of tumour suppressors BRCA1 or BRCA2 increases R-loop levels, because these proteins act to prevent R-loop formation[56]. Our findings add an important new angle to this observation, showing that oncogenes can induce R-loops, which suggests that increased R-loop levels might also be common in cancer cells that are proficient in BRCA1 or BRCA2. Interestingly, we observed increased protein levels of endogenous RNaseH1 in cells overexpressing $HRAS^{V12}$ (Fig. 5b,c). RNaseH1 and RNaseH2 are the main RNase activities counteracting R-loop formation and upregulation of these enzymes may be a response of cancer cells to increased R-loop levels[23]. Although the biology of increased RNaseH1 expression in response to $HRAS^{V12}$ requires further investigation, it supports the idea that oncogenic replication stress is strongly connected with R-loop metabolism.

One exciting implication of our findings is the potential to discover new factors involved in promoting replication stress. Components of the transcription machineries are found overexpressed in cancer[21,57]. TBP is one of several general transcription factors that have been implicated in oncogene-induced cell transformation[39,58]. The *TBP* promoter contains binding sites for oncogenic transcription factors and TBP is differentially expressed in cancers (Table 1)[59]. RAS has been reported to upregulate TBP expression through RAF-MEK and RALGDS signalling[19], and TBP levels are also increased by other growth factor signalling pathways such as epidermal growth factor/epidermal growth factor receptor[59,60] (Fig. 7j). We report here that increased TBP levels cause replication stress and markers of genomic instability. TBP expression also correlates with the expression of MYC, RAS and checkpoint kinases in a number of cancers (Table 1). As TBP overexpression did not increase RNA synthesis as strongly as $HRAS^{V12}$ overexpression and TBP depletion did not completely rescue $HRAS^{V12}$-induced replication stress, other transcription factors such as UBTF may also be involved downstream of $HRAS^{V12}$ (Fig. 7j).

A previous report showed that TBP is important for $HRAS^{V12}$ transforming function, and that TBP overexpression alone promotes anchorage-independent growth and tumour growth *in vivo*[49]. Our data presented here support the idea that TBP promotes the oncogenic phenotype by increasing transcription activity and replication stress. There is however no evidence that TBP itself is an oncogene, as TBP overexpression promoted tumour growth but not tumourigenesis itself[49]. Some cancer types show recurrent mutations in TBP that affect the length of the amino-terminal glutamine-rich motif (for example, Q72dup in diffuse large B-cell lymphoma and adrenocortical carcinoma)[61]. It is yet unknown whether such mutations have pathogenic relevance in cancer and whether they could be oncogenic.

Finally, increased RNA synthesis and elevated levels of R-loops could be correlated with replication stress in cancer. Identifying the molecular mechanisms that directly cause spontaneous replication stress should therefore help to predict and detect replication stress more accurately in cells and tissues. Nascent RNA synthesis can only be measured in live cells but increased expression of transcription factors and RNA polymerase subunits[21], as well as increased R-loops and expression of RNases H1 or H2, might be promising candidates for markers of transcription-associated replication stress.

Methods

Cell lines and reagents. Human BJ-hTert HRASV12$^{ER\text{-}TAM}$ (Agami and de Vita labs[26,27]) BJ-hTert and MRC5 fibroblasts (ATCC) were authenticated using 8-locus short tandem repeat (STR) profiling (LGC Standards). Cells were confirmed to be free of Mycoplasma infection and were grown in DMEM medium (Sigma) with 10% fetal bovine serum (Sigma) supplemented with L-glutamine (Gibco) in a humidified atmosphere containing 5% CO_2. For BJ-TBPind cells, human TBP complementary DNA (Origene, SC118124) was inserted into a pInducer20 lentivirus construct to generate TBP-pInducer20 and infected BJ-hTert cells were selected with 500 μg ml^{-1} G418 (Gibco). HRAS^{V12} was induced with 333 nM 4OHT (Sigma) and TBP expression was induced with 2 μg ml^{-1} doxycycline (Sigma). 4OHT or doxycycline remained present during all experiments.

DRB (100 μM) was from Sigma, and triptolide (1 μM) and α-amanitin (10 μg ml^{-1}) were from Tocris Bioscience. Ribonucleosides (adenosine, guanosine, uridine and cytidine, 10 μM) were obtained from Sigma. Roscovitine (25 μM) was from Sigma. Hydroxyurea (2 mM) was from Acros Organics, bleomycin (10 μg ml^{-1}) was from Sigma and cycloheximide (100 μg ml^{-1}) was from Calbiochem.

DNA fibre analysis. Cells were pulse labelled with 25 μM CldU and 250 μM IdU for 20 min and harvested. DNA fibre spreads were prepared by spotting 2 μl of cells (5×10^5 cells per ml in PBS) onto microscope slides followed by lysis with 7 μl of 0.5% SDS, 200 mM Tris-HCl pH 7.4 and 50 mM EDTA. Slides were tilted and DNA spreads fixed in methanol/acetic acid (3:1). HCl-treated fibre spreads were incubated with rat anti-bromodeoxyuridine (detects CldU, BU1/75, AbD Serotec, 1:1,000) and mouse anti-bromodeoxyuridine (detects IdU, B44, Becton Dickinson, 1:500) for 1 h, fixed with 4% paraformaldehyde (PFA) to increase staining intensity and incubated with anti-rat IgG AlexaFluor 555 and anti-mouse IgG AlexaFluor 488 (Molecular Probes) for 1.5 h. Images were acquired on an Nikon E600 microscope using a Nikon Plan Apo ×60 (1.3 numerical aperture) oil lens, a Hamamatsu digital camera (C4742-95) and the Volocity acquisition software (Perkin Elmer). Images were analysed using ImageJ (http://rsbweb.nih.gov/ij/). In each independent experiment, at least 300 fibres were measured per condition.

Immunofluorescence. Cells were washed and fixed as follows. γH2AX and 53BP1: CSK buffer (10 mM PIPES, 300 mM sucrose, 100 mM NaCl and 3 mM $MgCl_2$) for 1 min, 0.5% Triton X-100 in CSK buffer for 1 min and 4% PFA for 10 min at room temperature; Cyclin A and CldU: 4% PFA for 10 min and 0.3% Triton X-100 in PBS for 5 min at room temperature; S9.6: methanol for 10 min on ice and 0.5% Triton X-100 in PBS for 5 min at room temperature. RNase H (New England Biolabs) was used at 0.05 U μl^{-1} for 36 h at 37 °C. RNase A (Invitrogen) was used at 0.05 ng μl^{-1}–2 μg μl^{-1} for 2 h at 37 °C. Samples were blocked with 3% BSA/10% fetal bovine serum. Primary antibodies were mouse anti-phospho-Histone H2AX (Ser139) (JBW301, Millipore 05-636, 1:1,000), rabbit anti-53BP1 (Bethyl A300-272A, 1:30,000), mouse anti-Cyclin A (6E6, Thermo Scientific MS1061, 1:50), rat anti-CldU (BU1/75, AbD Serotec OBT0030G, 1:250) and mouse anti-RNA/DNA hybrid (S9.6, gift from Professor Richard Gibbons, hybridoma supernatant 1:100). Secondary antibodies were anti-mouse IgG AlexaFluor 488 and anti-rabbit IgG AlexaFluor 555 (Molecular Probes). DNA was counterstained with 4,6-diamidino-2-phenylindole (DAPI) and images acquired as above. For quantification of nuclear S9.6 intensity, ImageJ was used to generate nuclear masks based on DAPI staining and mean S9.6 fluorescence intensities per pixel were quantified per nucleus.

EU incorporation assay. EU incorporation assays were performed using the Click-iT RNA Alexa Fluor 594 Imaging Kit (Invitrogen) according to the manufacturer's instructions. Cells were incubated with 1 mM EU for 1 h, fixed with 4% PFA for 15 min at room temperature, permeabilized with 0.5% Triton X-100 for 15 min and Click-iT reaction was performed. DNA was counterstained with DAPI and images were acquired as above. ImageJ was used to generate nuclear masks based on DAPI staining and mean AlexaFluor 594 fluorescence intensities per pixel were quantified per nucleus.

siRNA and DNA transfection. siRNAs against TBP (TBPsi #1: sense 5′-GAA UCUUGGUUGUAAACUU-3′ and TBPsi #2: sense 5′-GGAUAAGAGAGCC ACGAAC-3′) were purchased from Ambion and Dharmacon. siRNA against RNaseH1 (s48357) was purchased from Ambion. 'Allstars negative control siRNA' was purchased from Qiagen. Cells were transfected with 50 nM siRNA using Dharmafect 1 reagent (GE Dharmacon). For RNaseH1 overexpression, cells were transfected with 2.5 μg pCMV6-AC-RNase H1-GFP (Origene) using TransIT-2020 (Mirus Bio). Empty pcDNA 3.1 (+) vector was purchased from Invitrogen.

Western blotting. Cell extracts were prepared in UTB buffer (50 mM Tris-HCl pH 7.5, 150 mM β-mercaptoethanol and 8 M urea) and sonicated to release DNA-bound proteins. Primary antibodies used were mouse anti-HRAS (Santa Cruz sc-29, 1:500), mouse anti-TBP (1TBP18, Abcam ab818, 1:2,000), rabbit anti-RNaseH1 (Abcam ab83179, 1:3,000), rabbit anti-phospho-ERK1/2 (Cell Signaling 9101, 1:1,000), rabbit anti-ERK1/2 (Cell Signaling 9102, 1:1,000), rabbit anti-γH2AX (Bethyl A300-081A, 1:1,000), rabbit anti-phospho-S345 CHK1 (Cell Signaling 2341, 1:1,000), rabbit anti-CHK1 (Cell Signaling 2345S, 1:1,000), rabbit anti-phospho-S33 RPA32 (Bethyl A300-426A, 1:1,000), mouse anti-RPA32 (9H8, Abcam ab2175, 1:2,000), rabbit anti-Turbo-GFP (Evrogen AB513, 1:10,000), mouse anti-p53 (DO-1, gift from Professor David Lane, 1:100), mouse anti-CyclinB1 (V152, Abcam ab72, 1:1,000), mouse anti-αTUBULIN (B512, Sigma T6074, 1:10,000), rabbit anti-βACTIN (Cell Signaling A967S, 1:5,000) and mouse anti-GAPDH (6C5, Abcam ab8245, 1:10,000).

ChIP analysis. ChIP was performed using 5 μg of anti-γH2AX (Merck Millipore, 07-164) and H2AX (Merck Millipore, 07-627) antibodies as previously described[62,63]. Where indicated, samples were treated with 0.5 μM Aphidicolin (Sigma) for 2 h before ChIP. The immunoprecipitated (γH2AX or H2AX antibody in immunoprecipitation (IP) reaction), control (beads only) and input DNAs were used as templates for qPCR, containing QuantiTect SYBR Green PCR Master Mix (QIAGEN) and gene-specific primers (see Supplementary Table 1).

RNA/DNA hybrid immunoprecipitation. DIP analysis was performed with RNA/DNA hybrid antibody (0.3 μg μl^{-1} per IP reaction) purified from S9.6 hybridoma cell lines as previously described[33]. The immunoprecipitated (S9.6 antibody in IP reaction), control (beads only) and input DNAs were used as templates for qPCR. DIP RNase H-sensitivity analysis was carried out before IP step with the addition of 25 U RNase H (NEB, M0297S). One hundred microlitres of nuclease digestion reaction contained 1 × reaction buffer and it was performed for 3 h at 37 °C.

Slot-blot experiments. Slot-blot experiments were carried out as described[32]. Genomic DNA (1.2 μg) were treated with 2 U of RNase H per μg of DNA (NEB, M0297S) for 2 h at 37 °C before loading on the slot blot. Half of the DNA sample was probed with S9.6 antibody (1:1,000) and the other half with anti-ssDNA antibody (MAB3031, Millipore, 1:25,000) as described[32]. Secondary antibody was goat anti-mouse horseradish peroxidase (1:10,000). Images were acquired with LAS-4000 (Fujifilm) and quantified using Image Studio Lite software (Li-COR Biosciences).

Cell proliferation and β-galactosidase assays. For proliferation assays, 1×10^5 or 2×10^5 cells were seeded in 12-well plates and incubated with 10 mg ml^{-1} resazurin for 2 h at the indicated time points. Resorufin fluorescence at 590 nm was measured using a BMG Labtech PHERAstar FS microplate reader. Cell numbers were determined by multiplying the number of initially seeded cells with (FR_n/FR_0) (FR_n: fluorescence reading at time point, FR_0: initial fluorescence reading). β-Galactosidase staining was performed using the Senescence β-Galactosidase Staining Kit (Cell Signaling) according to the manufacturer's instructions.

Nucleotide quantification. Cells were harvested and nucleotides extracted with 70% ice-cold methanol. Precipitated proteins were removed by centrifugation and supernatants stored at −80 °C. Supernatants were dried in a heated vacuum centrifuge and reconstituted in HPLC starting eluent. Samples were analysed by HPLC (Waters 2695, Watford, UK) with a photodiode array detector (Waters 2996). Separation was achieved using an Ace C18 (3 μm, 3 × 125 mm, Hichrom, UK) column maintained at 35 °C with eluent A: 10 mM potassium dihydrogen phosphate and 10 mM tetrabutylammonium hydrogen sulfate, 10% methanol pH 6.9; eluent B: 50 mM potassium dihydrogen phosphate, 6 mM tetrabutylammonium hydrogen sulfate and 30% methanol pH 7, using a flow rate of 0.6 ml min^{-1} and a gradient of 25–80% B over 25 min, with a run time of 30 min. Nucleotides were identified by comparing with absorbance spectra and retention times of commercially available standards.

Quantitative real-time PCR. Total RNA was harvested using TRIZOL reagent (Invitrogen) followed by DNase I treatment (Roche), 1.5 μg of total RNA was reverse transcribed using SuperScript Reverse Transcriptase III (Invitrogen) with random hexamers (Invitrogen), following manufacturer's instructions. The qPCR primers for amplification are listed in Supplementary Table 1. For quantitative real-time PCR, 2 μl of cDNA was analysed using a Rotor-Gene RG-3000 real-time PCR machine (Corbett Research) with QuantiTect SYBR green (Qiagen). Cycling parameters were 95 °C for 15 min, followed by 45 cycles of 94 °C for 20 s, 58 °C–62 °C for 20 s and 72 °C for 20 s. Fluorescence intensities were plotted against the number of cycles by using an algorithm provided by the manufacturer.

Pulsed field gel electrophoresis. Cells (2×10^6) per sample were treated as indicated, harvested and melted into 1.0% InCert-Agarose (Lonza) inserts. Inserts were digested in 0.5 M EDTA-1% *N*-laurylsarcosyl-proteinase K (1 mg ml^{-1}) at room temperature for 48 h and washed three times in TE buffer. Inserts were loaded onto a separation gel (1% chromosomal-grade agarose, Bio-Rad). Separation was performed using a CHEF DR III (BioRad; 120 field angle, 240 s switch time, 4 V cm^{-1}, 14 °C) for 20 h. Images of ethidium bromide-stained gels were acquired using a Syngene G:BOX gel imaging system. DSBs (chromosome

fragments >2 Mb) were quantified by densitometry using ImageJ. Intensity of DNA entering the gel was normalized to total DNA and untreated control to obtain final values.

Statistical analysis. Unless stated otherwise, all values are means ±1 s.e.m. of results from independent biological repeats. Scatter plots show pooled data, but numerical values displayed on plots represent the means ±1 s.e.m. of the results from independent repeats. Numbers of repeats N are indicated in the figure legends. Statistical tests were performed using the one-tailed Student's t-test. Coexpression data were obtained using CBioPortal[61,64] and statistical analyses were performed using the Student's t-test based on $t=(r^*\sqrt{n-2})/(\sqrt{-r^2})$ where r is Pearson's coefficient and n is number of samples in analysis.

Data availability. The authors declare that all the data supporting the findings of this study are available within the article and its Supplementary Information files and from the corresponding authors upon reasonable request.

References

1. McGranahan, N., Burrell, R. A., Endesfelder, D., Novelli, M. R. & Swanton, C. Cancer chromosomal instability: therapeutic and diagnostic challenges. *EMBO Rep.* **13,** 528–538 (2012).
2. Kotsantis, P., Jones, R. M., Higgs, M. R. & Petermann, E. Cancer therapy and replication stress: forks on the road to perdition. *Adv. Clin. Chem.* **69,** 91–138 (2015).
3. Burrell, R. A. *et al.* Replication stress links structural and numerical cancer chromosomal instability. *Nature* **494,** 492–496 (2013).
4. Bartkova, J. *et al.* Oncogene-induced senescence is part of the tumorigenesis barrier imposed by DNA damage checkpoints. *Nature* **444,** 633–637 (2006).
5. Di Micco, R. *et al.* Oncogene-induced senescence is a DNA damage response triggered by DNA hyper-replication. *Nature* **444,** 638–642 (2006).
6. Dereli-Oz, A., Versini, G. & Halazonetis, T. D. Studies of genomic copy number changes in human cancers reveal signatures of DNA replication stress. *Mol. Oncol.* **5,** 308–314 (2011).
7. Dobbelstein, M. & Sorensen, C. S. Exploiting replicative stress to treat cancer. *Nat. Rev. Drug Discov.* **14,** 405–423 (2015).
8. Dominguez-Sola, D. *et al.* Non-transcriptional control of DNA replication by c-Myc. *Nature* **448,** 445–451 (2007).
9. Bester, A. C. *et al.* Nucleotide deficiency promotes genomic instability in early stages of cancer development. *Cell* **145,** 435–446 (2011).
10. Neelsen, K. J., Zanini, I. M., Herrador, R. & Lopes, M. Oncogenes induce genotoxic stress by mitotic processing of unusual replication intermediates. *J. Cell Biol.* **200,** 699–708 (2013).
11. Carlos, A. R. *et al.* ARF triggers senescence in Brca2-deficient cells by altering the spectrum of p53 transcriptional targets. *Nat. Commun.* **4,** 2697 (2013).
12. Jones, R. M. *et al.* Increased replication initiation and conflicts with transcription underlie Cyclin E-induced replication stress. *Oncogene* **32,** 3744–3753 (2013).
13. Beier, R. *et al.* Induction of cyclin E-cdk2 kinase activity, E2F-dependent transcription and cell growth by Myc are genetically separable events. *EMBO J.* **19,** 5813–5823 (2000).
14. White, R. J. RNA polymerases I and III, growth control and cancer. *Nat. Rev. Mol. Cell Biol.* **6,** 69–78 (2005).
15. Zhao, J., Yuan, X., Frodin, M. & Grummt, I. ERK-dependent phosphorylation of the transcription initiation factor TIF-IA is required for RNA polymerase I transcription and cell growth. *Mol. Cell* **11,** 405–413 (2003).
16. Coleman, M. L., Marshall, C. J. & Olson, M. F. RAS and RHO GTPases in G1-phase cell-cycle regulation. *Nat. Rev. Mol. Cell Biol.* **5,** 355–366 (2004).
17. Lin, C. Y. *et al.* Transcriptional amplification in tumor cells with elevated c-Myc. *Cell* **151,** 56–67 (2012).
18. Nie, Z. *et al.* c-Myc is a universal amplifier of expressed genes in lymphocytes and embryonic stem cells. *Cell* **151,** 68–79 (2012).
19. Johnson, S. A., Mandavia, N., Wang, H. D. & Johnson, D. L. Transcriptional regulation of the TATA-binding protein by Ras cellular signaling. *Mol. Cell Biol.* **20,** 5000–5009 (2000).
20. Gonzalez, O. G. *et al.* Telomerase stimulates ribosomal DNA transcription under hyperproliferative conditions. *Nat. Commun.* **5,** 4599 (2014).
21. Bywater, M. J., Pearson, R. B., McArthur, G. A. & Hannan, R. D. Dysregulation of the basal RNA polymerase transcription apparatus in cancer. *Nat. Rev. Cancer* **13,** 299–314 (2013).
22. French, S. Consequences of replication fork movement through transcription units *in vivo*. *Science* **258,** 1362–1365 (1992).
23. Aguilera, A. & Garcia-Muse, T. R loops: from transcription byproducts to threats to genome stability. *Mol. Cell* **46,** 115–124 (2012).
24. Mirkin, E. V. & Mirkin, S. M. Replication fork stalling at natural impediments. *Microbiol. Mol. Biol. Rev.* **71,** 13–35 (2007).
25. Gan, W. *et al.* R-loop-mediated genomic instability is caused by impairment of replication fork progression. *Genes Dev.* **25,** 2041–2056 (2011).
26. Voorhoeve, P. M. *et al.* A genetic screen implicates miRNA-372 and miRNA-373 as oncogenes in testicular germ cell tumors. *Cell* **124,** 1169–1181 (2006).
27. De Vita, G. *et al.* Dose-dependent inhibition of thyroid differentiation by RAS oncogenes. *Mol. Endocrinol.* **19,** 76–89 (2005).
28. Drost, J. *et al.* BRD7 is a candidate tumour suppressor gene required for p53 function. *Nat. Cell Biol.* **12,** 380–389 (2010).
29. Herranz, H., Hong, X., Hung, N. T., Voorhoeve, P. M. & Cohen, S. M. Oncogenic cooperation between SOCS family proteins and EGFR identified using a *Drosophila* epithelial transformation model. *Genes Dev.* **26,** 1602–1611 (2012).
30. Boguslawski, S. J. *et al.* Characterization of monoclonal antibody to DNA.RNA and its application to immunodetection of hybrids. *J. Immunol. Methods* **89,** 123–130 (1986).
31. El Hage, A., French, S. L., Beyer, A. L. & Tollervey, D. Loss of topoisomerase I leads to R-loop-mediated transcriptional blocks during ribosomal RNA synthesis. *Genes Dev.* **24,** 1546–1558 (2010).
32. Sollier, J. *et al.* Transcription-coupled nucleotide excision repair factors promote R-loop-induced genome instability. *Mol. Cell* **56,** 777–785 (2014).
33. Skourti-Stathaki, K., Proudfoot, N. J. & Gromak, N. Human senataxin resolves RNA/DNA hybrids formed at transcriptional pause sites to promote Xrn2-dependent termination. *Mol. Cell* **42,** 794–805 (2011).
34. Derheimer, F. A. *et al.* RPA and ATR link transcriptional stress to p53. *Proc. Natl Acad. Sci. USA* **104,** 12778–12783 (2007).
35. Chan, K. L., Palmai-Pallag, T., Ying, S. & Hickson, I. D. Replication stress induces sister-chromatid bridging at fragile site loci in mitosis. *Nat. Cell Biol.* **11,** 753–760 (2009).
36. Lukas, C. *et al.* 53BP1 nuclear bodies form around DNA lesions generated by mitotic transmission of chromosomes under replication stress. *Nat. Cell Biol.* **13,** 243–253 (2011).
37. Groh, M., Lufino, M. M., Wade-Martins, R. & Gromak, N. R-loops associated with triplet repeat expansions promote gene silencing in Friedreich ataxia and fragile X syndrome. *PLoS Genet.* **10,** e1004318 (2014).
38. Techer, H. *et al.* Signaling from Mus81-Eme2-dependent DNA damage elicited by Chk1 deficiency modulates replication fork speed and origin usage. *Cell Rep.* **14,** 1114–1127 (2016).
39. Johnson, S. A. *et al.* Increased expression of TATA-binding protein, the central transcription factor, can contribute to oncogenesis. *Mol. Cell Biol.* **23,** 3043–3051 (2003).
40. Wang, H. D., Trivedi, A. & Johnson, D. L. Hepatitis B virus X protein induces RNA polymerase III-dependent gene transcription and increases cellular TATA-binding protein by activating the Ras signaling pathway. *Mol. Cell Biol.* **17,** 6838–6846 (1997).
41. Aird, K. M. *et al.* Suppression of nucleotide metabolism underlies the establishment and maintenance of oncogene-induced senescence. *Cell Rep.* **3,** 1252–1265 (2013).
42. Cancer Genome Atlas Network. Comprehensive molecular characterization of human colon and rectal cancer. *Nature* **487,** 330–337 (2012).
43. Brennan, C. W. *et al.* The somatic genomic landscape of glioblastoma. *Cell* **155,** 462–477 (2013).
44. Cancer Genome Atlas Research Network. Comprehensive molecular characterization of gastric adenocarcinoma. *Nature* **513,** 202–209 (2014).
45. Cancer Genome Atlas Research Network. Comprehensive molecular profiling of lung adenocarcinoma. *Nature* **511,** 543–550 (2014).
46. Cancer Genome Atlas Research Network. Comprehensive genomic characterization of squamous cell lung cancers. *Nature* **489,** 519–525 (2012).
47. Cancer Genome Atlas Research Network. Genomic and epigenomic landscapes of adult *de novo* acute myeloid leukemia. *N. Engl. J. Med.* **368,** 2059–2074 (2013).
48. Cancer Genome Atlas Research Network. Comprehensive molecular characterization of urothelial bladder carcinoma. *Nature* **507,** 315–322 (2014).
49. Cancer Genome Atlas Network. Comprehensive molecular portraits of human breast tumours. *Nature* **490,** 61–70 (2012).
50. Cancer Genome Atlas Research Network. The molecular taxonomy of primary prostate cancer. *Cell* **163,** 1011–1025 (2015).
51. Ruiz, S. *et al.* Limiting replication stress during somatic cell reprogramming reduces genomic instability in induced pluripotent stem cells. *Nat. Commun.* **6,** 8036 (2015).
52. Zimmerman, E. S. *et al.* Human immunodeficiency virus type 1 Vpr induces DNA replication stress *in vitro* and *in vivo*. *J. Virol.* **80,** 10407–10418 (2006).
53. Davidson, I. F., Li, A. & Blow, J. J. Deregulated replication licensing causes DNA fragmentation consistent with head-to-tail fork collision. *Mol. Cell* **24,** 433–443 (2006).
54. Ogrunc, M. *et al.* Oncogene-induced reactive oxygen species fuel hyperproliferation and DNA damage response activation. *Cell Death Differ.* **21,** 998–1012 (2014).
55. Rai, P. *et al.* Continuous elimination of oxidized nucleotides is necessary to prevent rapid onset of cellular senescence. *Proc. Natl Acad. Sci. USA* **106,** 169–174 (2009).

56. Bhatia, V. *et al.* BRCA2 prevents R-loop accumulation and associates with TREX-2 mRNA export factor PCID2. *Nature* **511,** 362–365 (2014).
57. Ribeiro, J. R., Lovasco, L. A., Vanderhyden, B. C. & Freiman, R. N. Targeting TBP-associated factors in ovarian cancer. *Front. Oncol.* **4,** 45 (2014).
58. Bywater, M. J. *et al.* Inhibition of RNA polymerase I as a therapeutic strategy to promote cancer-specific activation of p53. *Cancer Cell* **22,** 51–65 (2012).
59. Fromm, J. A., Johnson, S. A. & Johnson, D. L. Epidermal growth factor receptor 1 (EGFR1) and its variant EGFRvIII regulate TATA-binding protein expression through distinct pathways. *Mol. Cell Biol.* **28,** 6483–6495 (2008).
60. Zhang, Q., Zhong, Q., Evans, A. G., Levy, D. & Zhong, S. Phosphorylation of histone H3 serine 28 modulates RNA polymerase III-dependent transcription. *Oncogene* **30,** 3943–3952 (2011).
61. Cerami, E. *et al.* The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. *Cancer Discov.* **2,** 401–404 (2012).
62. Kadener, S., Fededa, J. P., Rosbash, M. & Kornblihtt, A. R. Regulation of alternative splicing by a transcriptional enhancer through RNA pol II elongation. *Proc. Natl Acad. Sci. USA* **99,** 8185–8190 (2002).
63. West, S., Gromak, N. & Proudfoot, N. J. Human 5′ --> 3′ exonuclease Xrn2 promotes transcription termination at co-transcriptional cleavage sites. *Nature* **432,** 522–525 (2004).
64. Gao, J. *et al.* Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Sci. Signal.* **6,** pl1 (2013).

Acknowledgements

P.K. and R.M.J. are supported by a Worldwide Cancer Research grant (13-1048) and a Medical Research Council New Investigator Research Grant (MR/J007595/1) to E.P. N.G. is supported by a Royal Society University Research Fellowship and a Medical Research Council New Investigator Research Grant (MR/J007870/1). L.M.S. is supported by a fellowship from the Portuguese Foundation for Science and Technology. The results shown here are in part based on data generated by the TCGA Research Network: http://cancergenome.nih.gov/. We thank Profs Gabriella de Vita, Grant S. Stewart, David Lane, Richard Gibbons and Dr Alexandra Milona for providing materials.

Author contributions

P.K. designed and performed most of the experiments, analysed data and contributed to writing the paper. L.M.S. performed DIP, ChIP, reverse transcriptase–quantitative PCR (RT–qPCR) and slot-blot experiments. S.I. performed ChIP and DIP experiments. R.M.J performed roscovitine and CldU staining experiments. L.F. designed, performed and analysed nucleotide quantification. N.G. designed, performed and analysed DIP, ChIP, RT–qPCR and slot-blot experiments, and contributed to writing the paper. E.P. conceived and coordinated the project, designed experiments, analysed data and wrote the paper.

Additional information

Supplementary Information accompanies this paper at http://www.nature.com/naturecommunications

Competing financial interests: The authors declare no competing financial interests.

Reprints and permission information is available online at http://npg.nature.com/reprintsandpermissions/

How to cite this article: Kotsantis, P. *et al.* Increased global transcription activity as a mechanism of replication stress in cancer. *Nat. Commun.* **7,** 13087 doi: 10.1038/ncomms13087 (2016).

This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/

© The Author(s) 2016

ResearchGate

See discussions, stats, and author profiles for this publication at: https://www.researchgate.net/publication/264627993

Mechanisms of transcriptional dysregulation in repeat expansion disorders

Article *in* Biochemical Society Transactions · August 2014

DOI: 10.1042/BST20140049 · Source: PubMed

CITATIONS
7

READS
142

3 authors, including:



Matthias Groh
University of Oxford
4 PUBLICATIONS **121** CITATIONS
SEE PROFILE



Natalia Gromak
University of Oxford
25 PUBLICATIONS **1,665** CITATIONS
SEE PROFILE

All content following this page was uploaded by Natalia Gromak on 14 May 2015.

The user has requested enhancement of the downloaded file.

Mechanisms of transcriptional dysregulation in repeat expansion disorders

Matthias Groh[1], Lara Marques Silva[1] and Natalia Gromak[1]*

[1]Sir William Dunn School of Pathology, University of Oxford, South Parks Road, OX1 3RE, UK
*- communicating author (natalia.gromak@path.ox.ac.uk)

Key words: RNA/DNA hybrids, R-loops, transcription, repeat expansion disease, Friedreich ataxia; Fragile X syndrome

Abbreviations: Pol II, RNA polymerase II; mRNA, messenger RNA; FXN, frataxin; FRDA, Friedreich ataxia; FXS, Fragile X syndrome; SCA, spinocerebellar ataxia; DM1, Myotonic Dystrophy type 1; HD, Huntington's disease; HTT, Huntingtin.

Abstract

Around forty human diseases are associated with expansion of repeat sequences. These expansions can reside within coding or non-coding parts of the genes, affecting the host gene function. The presence of such expansions results in production of toxic RNA and/or protein or causes transcriptional repression and silencing of the host gene. Although the molecular mechanisms of expansion diseases are not well understood, mounting evidence suggests that transcription through expanded repeats plays an essential role in disease pathology. The presence of an expansion can affect RNA Polymerase transcription, leading to dysregulation of transcription-associated processes, such as RNA splicing, formation of RNA/DNA hybrids (R-loops), production of antisense, short non-coding and bidirectional RNA transcripts. Here we summarise current advances in this field and discuss possible roles of transcriptional defects in disease pathology.

Transcriptional dysregulation in expansion diseases

Around forty human diseases are associated with expanded repeat sequences found in coding and non-coding gene regions [1]. 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 [2]. Iinterestingly, the size of expansion also increases in postmitotic neuronal cells in the absence of DNA replication [3]. This observation raises the importance of RNA Polymerase II (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 RNA/DNA hybrids (R-loops), production of antisense, short non-coding and bidirectional RNA transcripts (Figure 1).

Expanded GAA repeats are known to directly inhibit Pol II transcription *in vitro* [4, 5] and on reporter constructs in bacteria and mammalian cells [6, 7]. In line with the results from these model systems, the GAA expansion in the first intron of the frataxin (*FXN*) gene leads to reduction of FXN mRNA and protein, causing Friedreich ataxia (FRDA) [8-11]. On a molecular level, it has been demonstrated that expanded GAA repeats affect both Pol II transcription initiation and elongation on the *FXN* gene, correlating with increased levels of repressive histone modifications and abnormal DNA methylation [9, 11, 12]. Other diseases, including myotonic dystrophy 1 (DM1), characterised by CTG expansion in the *DMPK* gene, and amyotrophic lateral sclerosis (ALS) and frontotemporal dementia (FTD), triggered by the GGGGCC expansion in *C9orf72* gene, have been also associated with reduced Pol II transcription [13, 14].

The rate of Pol II transcriptional elongation through expanded repeats is an important factor contributing to disease pathology. CAG expansion in *HTT* gene in Huntington's disease (HD), results in production of toxic huntingtin protein (*mHTT)*. Depletion of Spt4/Supt4, a transcription elongation factor that aids Pol II processivity by reducing dissociation of Pol II from the DNA template, reduced *mHTT* protein in neuronal cells and decreased its aggregation and toxicity [15]. Moreover, the process of transcription has recently emerged as an essential player in modulating repeat instability in yeast and mammalian cells [2, 16]. Interestingly, GAA expansions can act as independent promoters and initiate transcription in yeast, potentially resulting in production of antisense and/or bidirectionally-transcribed RNAs, which could have pathological functions [2].

Expanded repeats can also mediate regulation of transcription *in trans*. In particular, in HD *mHTT* protein can bind a number of transcription factors and promote transcriptional down-regulation of many genes, associated with decrease in histone acetylation, an effect which can be reversed by treatment with HDAC inhibitors [17].

R-loops in expansion-associated pathology

Unusual DNA structures (triplex DNA and RNA/DNA hybrids) have been proposed to play a role in the molecular mechanisms of several expansion diseases, contributing to transcriptional gene silencing [18-20]. RNA/DNA hybrids (R-loops) are formed during the process of transcription, when nascent RNA hybridizes to DNA template behind the elongating RNA Polymerase (Figure 2). R-loops are detected in organisms from bacteria to humans and are implicated in many processes [21]. In mammals, R-loops were originally discovered in the immunoglobulin class switch regions, essential for generating the antibody diversity in activated B cells [22]. R-loops have been shown to impede transcription elongation [23, 24] and accumulate in cells depleted of the key splicing factor SRSF1, resulting in genome instability and appearance of double-strand breaks [25]. Recent studies demonstrated that R-loops are enriched over CpG promoters and may be involved in protection of these regions from DNA methylation [26]. We recently showed that R-loops formed downstream of the polyA signal in human genes, promoting RNA Pol II transcriptional termination [27].

RNA/DNA hybrids are induced at GAA repeats following *in vitro* transcription and in bacteria [28, 29]. R-loops formed on plasmids containing CTG repeats in *E.coli* and mini-gene constructs in human cells promote repeat instability, consistent with their role in disease pathology [30]. Recently we demonstrated that R-loops are formed over both GAA and CGG expanded alleles in cells from FRDA and FXS patients, employing the DNA-immuno-precipitation (DIP) approach with an antibody recognizing RNA/DNA hybrids [31]. These stable R-loops correlate with the size of the expansion, co-localise with repressive chromatin marks, characteristic of these diseases, and impede *FXN* gene transcription in patients' cells (Figure 2). R-loops can trigger the formation of repressive chromatin and cause transcriptional silencing of the *FXN* gene, providing a molecular link between R-loops and the pathology of expansion diseases [31]. In line with R-loops being formed on expanded CGG repeats of FXS patients [31], recently it has been demonstrated that promoter-bound trinucleotide repeat mRNA drives epigenetic silencing in Fragile X syndrome [32]. Moreover, R-loops and G-quadruplexes associated with expanded (GGGGCC) repeats in C9orf72 initiate the molecular events leading to ALS/FTD neurodegenerative diseases [13].

There are a number of possible ways by which R-loops can trigger human diseases. Similar to R-loops at the 3'ends of human genes [27], expansion-associated R-loops may form a structural block, directly interfering with Pol II transcription elongation. In addition, R-loops may trigger the formation of repressive chromatin over the expansion region, analogous to heterochromatin formation at centromeres in *S. pombe* [33]. R-loops may also directly block the progression of the DNA replication fork, leading to fork collapse, formation of double-strand breaks, and further stimulation of repeat instability [19].

The ability of R-loops to trigger transcriptional repression in expansion diseases *in vivo* is important for understanding the basic molecular mechanism underlying their pathology. The effects of R-loops on Pol II transcription, genome stability and gene silencing may be a cause of many human diseases related to expansion of tandem repeat sequences. Hence, R-loops may represent potentially attractive therapeutic targets in such diseases in the future.

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 insulin receptor, chloride channel 1 and troponin T type 2, implicated in the clinical symptoms of DM1 [34]. Similar to CUG repeats, other expanded sequences, associated with SCA3, SCA8 and FXTAS were also shown to affect splicing *in trans* [34, 35].

It has been demonstrated that the speed of RNA Pol II transcriptional elongation can trigger changes in alternative splicing [36]. 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 [37]. In particular, expanded GAA repeats inserted in reporter

constructs trigger aberrant splicing in a position- and context-dependent manner [38]. 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 [38]. Similarly, expanded CAG repeats in the *HTT* gene affect its splicing, possibly through aberrant binding of splicing factor SFSF6, leading to production of short polyadenylated CAG-repeat containing mRNAs, which are translated into small toxic poly-glutamine proteins [39].

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 [40]. 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 [41]. Similarly, in the *ATXN7* gene a CAG-repeat spanning antisense SCAANT1 transcript negatively regulates *ATXN7* sense transcript levels in *cis* [42].

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 [43]. Conversely, transcription of the opposite strand produces long non-coding CUG-containing *ATXN8OS* RNA, leading to RNA-mediated toxicity, similar to DM1 [34]. 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 [44].

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 above). However, more recently it has been shown that the CTG repeats in *DMPK* are also transcribed in the antisense direction [45]. 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 [45]. 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 [46]. 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 [35, 47-49].

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 [45, 48]. 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* [49, 50].

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 [30, 51]. 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 [52]. Although the function of *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 [53].

Conclusions

Transcription plays a major role in driving the pathology of trinucleotide expansion diseases, starting from changes in Pol II transcription initiation and elongation and resulting in formation of R-loops, production of toxic, small, antisense and bidirectional RNAs and mis-regulation of RNA splicing (Figure 2). This plethora of gene expression events will have diverse consequences on disease pathology. Future research, focusing on systematic characterisation of the features of transcriptional mis-regulation in expansion repeat diseases, combined with development of new cellular disease models, will allow further insights into specific molecular mechanisms underlying these disorders.

Acknowledgements

NG is supported by a Royal Society University Research Fellowship and MRC NIRG (MR/J007870/1). MG is supported by a PhD studentship from Ataxia UK/Motor Neuron Diseases Association

(Gromak/Jun11/6278). LMS is supported by the fellowship from Portuguese Foundation for Science and Technology. We thank Chris Norbury for critically reading the manuscript.

Figure legends

Figure 1. Role of RNA transcription in pathology of repeat expansion diseases.

A generic gene containing expanded sequences within 5' and 3'UTRs, exons and non-coding sequences is shown. Diseases associated with specific expansions discussed in this review are presented below the gene. The expansion location is denoted by a triangle. TSS denotes transcriptional start site. DNA is shown as double helix and RNA is represented as the 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.

Figure 2. R-loops mediate gene silencing in Friedreich ataxia and Fragile X syndrome.

Model for the role of R-loops in gene silencing, mediated by expanded trinucleotide repeats in the *FXN* and *FMR1* genes. Low R-loop level on the wild type allele allows efficient transcriptional elongation and gene expression, associated with active chromatin marks. Transcribed (GAA)n and (CGG)n expanded repeats form R-loops, resulting in decreased initiation and elongation of RNA Pol II. This leads to down-regulation of *FXN* and *FMR1* expression, associated with formation of repressive DNA and chromatin marks. Figure adapted from [31].

References

1. 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. Nat Rev Mol Cell Biol. **11**, 165-70
2. 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**, 254-65
3. 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**, 3467-72
4. 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. Am J Hum Genet. **62**, 111-21
5. Grabczyk, E. and Usdin, K. (2000) 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**, 2815-22

6. 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**, 1075-84
7. 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. J Biol Chem. **273**, 14588-95
8. Campuzano, V., Montermini, L., Molto, M.D., Pianese, L., Cossee, M., Cavalcanti, F., Monros, E., Rodius, F., Duclos, F., Monticelli, A., Zara, F., Canizares, J., Koutnikova, H., Bidichandani, S.I., Gellera, C., Brice, A., Trouillas, P., De Michele, G., Filla, A., De Frutos, R., Palau, F., Patel, P.I., Di Donato, S., Mandel, J.L., Cocozza, S., Koenig, M., and Pandolfo, M. (1996) Friedreich's ataxia: autosomal recessive disease caused by an intronic GAA triplet repeat expansion. Science. **271**, 1423-7
9. Kumari, D., Biacsi, R.E., and Usdin, K. (2011) Repeat expansion affects both transcription initiation and elongation in friedreich ataxia cells. J Biol Chem. **286**, 4209-15
10. 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. Nat Chem Biol. **2**, 551-8
11. 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 Res. **39**, 8366-77
12. 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 Res. **35**, 3383-90
13. 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., Rothstein, J.D., and Wang, J. (2014) C9orf72 nucleotide repeat structures initiate molecular cascades of disease. Nature
14. 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. J Nucleic Acids. **2013**, 567435
15. 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., and Cheng, T.H. (2012) Spt4 is selectively required for transcription of extended trinucleotide repeats. Cell. **148**, 690-701
16. 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
17. 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., Clark, T.W., Yohrling, G.J., and Cha, J.H. (2007) Histones associated with downregulated genes are hypo-acetylated in Huntington's disease models. Hum Mol Genet. **16**, 1293-306
18. Wells, R.D. (2008) DNA triplexes and Friedreich ataxia. FASEB J. **22**, 1625-34
19. McIvor, E.I., Polak, U., and Napierala, M. (2010) New insights into repeat instability: role of RNA*DNA hybrids. RNA Biol. **7**, 551-8
20. 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. Mol Cell. **3**, 465-75
21. Aguilera, A. and Garcia-Muse, T. (2012) R loops: from transcription byproducts to threats to genome stability. Mol Cell. **46**, 115-24
22. 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. Nat Immunol. **4**, 442-51

23. Huertas, P. and Aguilera, A. (2003) Cotranscriptionally formed DNA:RNA hybrids mediate transcription elongation impairment and transcription-associated recombination. Mol Cell. **12**, 711-21
24. Tous, C. and Aguilera, A. (2007) Impairment of transcription elongation by R-loops in vitro. Biochem Biophys Res Commun. **360**, 428-32
25. Li, X. and Manley, J.L. (2005) Inactivation of the SR protein splicing factor ASF/SF2 results in genomic instability. Cell. **122**, 365-78
26. 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. Mol Cell. **45**, 814-25
27. 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. Mol Cell. **42**, 794-805
28. 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 Res. **35**, 5351-9
29. 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 Res. **39**, 1749-62
30. Lin, Y., Dent, S.Y., Wilson, J.H., Wells, R.D., and Napierala, M. (2010) R loops stimulate genetic instability of CTG.CAG repeats. Proc Natl Acad Sci U S A. **107**, 692-7
31. Groh, M., Lufino, M., Wade-Martins, R., and Gromak, N. (2014) R-loops associated with triplet repeat expansions promote gene silencing in Friedreich ataxia and Fragile X syndrome. PLoS Genet, in press
32. 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**, 1002-5
33. Nakama, M., Kawakami, K., Kajitani, T., Urano, T., and Murakami, Y. (2012) DNA-RNA hybrid formation mediates RNAi-directed heterochromatin formation. Genes Cells. **17**, 218-33
34. Echeverria, G.V. and Cooper, T.A. (2012) RNA-binding proteins in microsatellite expansion disorders: mediators of RNA toxicity. Brain Res. **1462**, 100-11
35. 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 Res. **40**, 11-26
36. 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. Mol Cell. **12**, 525-32
37. 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. Mol Cell. **35**, 82-92
38. Baralle, M., Pastor, T., Bussani, E., and Pagani, F. (2008) Influence of Friedreich ataxia GAA noncoding repeat expansions on pre-mRNA processing. Am J Hum Genet. **83**, 77-88
39. 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., Detloff, P.J., Housman, D.E., and Bates, G.P. (2013) Aberrant splicing of HTT generates the pathogenic exon 1 protein in Huntington disease. Proc Natl Acad Sci U S A. **110**, 2366-70
40. Pelechano, V. and Steinmetz, L.M. (2013) Gene regulation by antisense transcription. Nat Rev Genet. **14**, 880-93
41. 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. Hum Mol Genet. **20**, 3467-77

42. 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**, 1071-84
43. 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., Day, J.W., and Ranum, L.P. (2006) Bidirectional expression of CUG and CAG expansion transcripts and intranuclear polyglutamine inclusions in spinocerebellar ataxia type 8. Nat Genet. **38**, 758-69
44. Wilburn, B., Rudnicki, D.D., Zhao, J., Weitz, T.M., Cheng, Y., Gu, X., Greiner, E., Park, C.S., Wang, N., Sopher, B.L., La Spada, A.R., Osmand, A., Margolis, R.L., Sun, Y.E., and Yang, X.W. (2011) An antisense CAG repeat transcript at JPH3 locus mediates expanded polyglutamine protein toxicity in Huntington's disease-like 2 mice. Neuron. **70**, 427-40
45. 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. Mol Cell. **20**, 483-9
46. 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. Hum Mol Genet. **20**, 3757-68
47. 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 Res. **31**, 6243-8
48. 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. Mol Cell. **25**, 575-86
49. 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 Genet. **8**, e1002481
50. 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 Genet. **7**, e1001340
51. 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**, 580-8
52. 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. Hum Mol Genet. **16**, 3174-87
53. Loesch, D.Z., Godler, D.E., Evans, A., Bui, Q.M., Gehling, F., Kotschet, K.E., Trost, N., Storey, E., Stimpson, P., Kinsella, G., Francis, D., Thorburn, D.R., Venn, A., Slater, H.R., and Horne, M. (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. Genet Med. **13**, 392-9

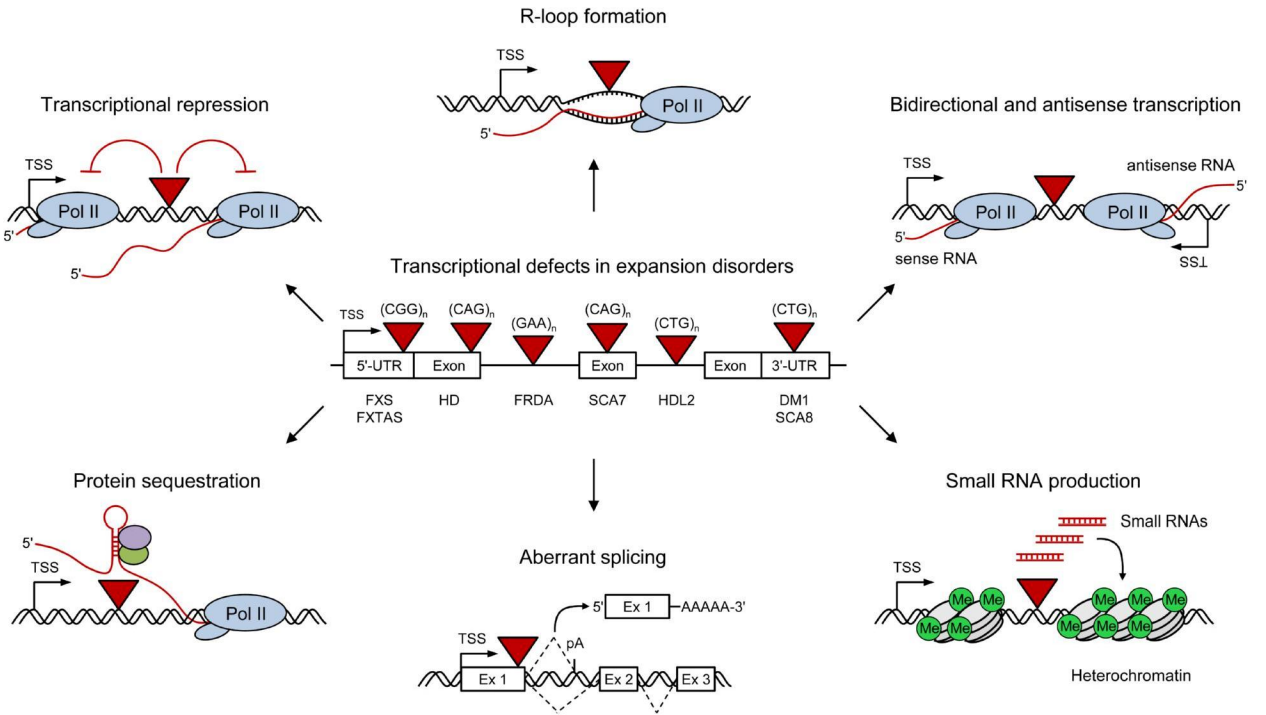
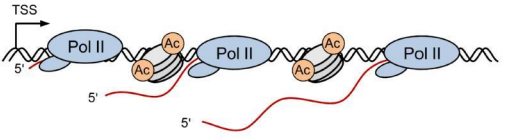

R-loop formation
TSS
Pol II
5'
Transcriptional repression
TSS
Pol II
Pol II
5'
5'
Bidirectional and antisense transcription
TSS
antisense RNA
5'
Pol II
Pol II
5'
sense RNA
SS⊥
Transcriptional defects in expansion disorders
TSS
(CGG)n
(CAG)n
(GAA)n
(CAG)n
(CTG)n
(CTG)n
5'-UTR
Exon
Exon
Exon
3'-UTR
FXS
FXTAS
HD
FRDA
SCA7
HDL2
DM1
SCA8
Protein sequestration
5'
TSS
Pol II
Aberrant splicing
5'
Ex 1
AAAAA-3'
TSS
pA
Ex 1
Ex 2
Ex 3
Small RNA production
Small RNAs
TSS
Me
Heterochromatin

Figure 1

Normal allele



Expanded trinucleotide allele

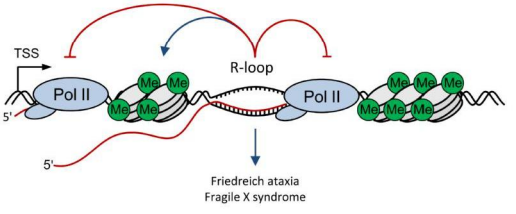


Figure 2