The role of transcription and RNA/DNA structures in the pathology of Huntington's Disease

Chiara Beghè

Sir William Dunn School of Pathology

St John's College

University of Oxford



Supervisor: Prof. Natalia Gromak

Second Supervisor: Prof. Shona Murphy

A thesis submitted for the degree of

Doctor of Philosophy in Molecular Cell Biology in Health and Disease

Trinity Term, 2023

To my family and Francesco,

who supported me unconditionally.

Abstract

Huntington's Disease (HD) is a repeat expansion disease where a CAG-triplet expansion in the first exon 1 of the huntingtin (*HTT)* gene leads to the production of a mutant HTT (mHTT) protein. mHTT forms toxic aggregates that alter mitochondrial function, impair protein degradation and lead to neuronal cell death, causing degeneration of medium spiny neurons (MSNs) in the brains of HD patients. Despite the well-defined genetic cause, the molecular mechanisms underlying HD pathology are still not fully understood.

Little is known about the transcriptional regulation of the *HTT* gene itself. To address this, I carried out transcriptional and chromatin analysis in cells from Control and HD patients. I observed transcriptional down-regulation of the *HTT* gene which correlates with its epigenetic signature in HD. The evidence from multiple repeat expansion diseases suggests that unusual nucleic acid structures are involved in their pathogenesis. Non-canonical RNA/DNA structures can interfere with transcription and have been implicated in repeat instability; therefore, I investigated their formation in HD cells. Interestingly, I observed an enrichment of double-stranded RNAs over expanded CAG repeat *HTT* transcripts, implicating them in HD pathogenesis.

The current understanding of the role of HTT protein in transcription and genome maintenance is still limited. Here I show that the HTT protein binds to active forms of RNA Polymerase II and several factors important for transcription and R-loop resolution. R-loops are non-canonical structures composed of an RNA/DNA hybrid and a single strand of DNA formed during transcription. Recently, R-loops have emerged as key regulators of important biological processes including transcription, DNA replication, generation of antibody diversity and epigenetics. However, if not properly regulated, R-loops can become pathogenic and accumulate in neurodegenerative disorders and cancer. Until now, the contribution of R-loops to HD pathology has not been investigated but here I show that the HTT protein is in complex with R-loop structures and may be important for their resolution during transcription. Consistent with this role of HTT protein, I detected an accumulation of pathological R-loops in several HD cellular models, including MSNs, the most affected cells in HD. In line with this, global transcription down-regulation was observed in HD cells. R-loops can be a source of DNA damage and inflammation, hallmarks of HD. Therefore, an investigation of the mechanistic link relating these phenotypes was also addressed.

This work provides novel insight into the role of unusual RNA/DNA structures in HD and their contribution to early neurotoxicity in HD, potentially uncovering new targets for therapeutic interventions.

Acknowledgments

I would like to thank my supervisor, Prof. Natalia Gromak, who gave me the opportunity to work on this fantastic project and supported me at every stage of this journey. She has not only offered me scientific support, but she has also thought me the importance of perseverance and dedication in academia. Her inspiration has motivated me to always strive for excellence. I would also like to thank my second supervisor, Prof. Shona Murphy, for her invaluable support, encouragement and engaging Italian chats.

I would like to gratefully acknowledge the Royal Society and the Sir William Dunn School of Pathology for generously funding my PhD studies.

This work would not have been possible without the members of the Gromak and Murphy labs, both past and present, who were there in the happy and sad days providing support, friendship and cakes! Thank you to the sweetest labmates Clelia Accalai, Flavia Constantinescu, Agnese Cristini and Joana Guiro. Thanks to my students, Rob Torrance and Elena Becker, for their exceptional contributions. Special thanks to Maria Sole Giordano for her support during my last year of PhD and for taking time out of her busy days to help with my thesis and experiments. Thanks also to Michael Tellier for his invaluable scientific advices and for all his help with bioinformatical analysis.

I would like to sincerely thank my thesis committee - Prof. Nicholas Proudfoot and Prof. Jernej Ule - for their insightful discussion, which has enriched my project.

These years would not have been the same without my friends from Oxford, Parma and London, whose friendship has made these years unforgettable. Thank you for being a part of this!

I am especially grateful to my family for always being there despite the distance. Thank you for giving me unconditional support and the freedom to choose my own path and chase my dreams. Thanks to my mum, the sweetest one in the world. To my dad, whose quiet support and countless trips to and from the airport mean more than words can say. To my sister and my little niece Zoe, you both have my heart. To my grandparents, who played a great role in shaping who I am today.

Last but not least, a special thanks to my partner Francesco, who has been a rock for me throughout all these years and particularly during my PhD. Your love and support have been my driving force, and I am immensely grateful to have you always by my side.

Table of Contents

Abstract i
Acknowledgments ii
Table of Contents iii
List of Figures vii
List of Tables ix
Abbreviations x
Chapter 1 Introduction 1
1.1 Repeat expansion diseases 2
1.1.1 Coding and Non-Coding RE diseases 2
1.1.2 Protein-induced toxicity in RE diseases 5
1.1.3 RNA-induced toxicity in RE diseases 7
1.1.4 DNA-mediated mechanisms associated with RE diseases 9
1.1.4.1 Unusual RNA/DNA structures formed by expanded repeats 9
1.1.4.2 Epigenetic dysregulation in RE diseases 10
1.2 Huntington's Disease 13
1.2.1 The pathology of HD 13
1.2.2 The Huntingtin gene 15
1.2.3 Instability of expanded CAG repeats 19
1.2.4 The Huntingtin protein 20
1.2.4.1 The structure of the HTT protein 20
1.2.4.2 HTT proteolysis and post-translational modifications 22
1.2.5 The functions of the HTT protein 23
1.2.5.1 HTT, intracellular trafficking and cell division 23
1.2.5.2 HTT and autophagy 25
1.2.5.3 HTT and mitochondria 25
1.2.5.4 HTT and apoptosis 26
1.2.5.5 HTT and transcription regulation 26
1.2.5.6 HTT, chromatin and epigenetics 29
1.2.5.7 HTT and DNA damage 30
1.2.6 RNA toxicity induced by expanded *HTT* mRNA 32
1.2.7 Immune dysfunction in HD 34
1.2.8 HD Therapeutics 35
1.3 R-loops in health and disease 39
1.3.1 Structure and detection of R-loops 39
1.3.2 Physiological roles of R-loops 40
1.3.3 R-loop interacting proteins 42
1.3.4 R-loops and human diseases 46

1.4 Aims of this work....51
1.5 Publications....53
Chapter 2 Materials and Methods....54
2.1 Mammalian cell culture techniques....55
2.1.1 Cell lines....55
2.1.2 iPSC lines....55
2.1.3 Cell Growth Curve....56
2.1.4 Pharmacological treatments....56
2.1.5 siRNA-mediated knock-down....56
2.1.6 Medium Spiny Neurons (MSNs) differentiation protocol....57
2.2 RNA analysis methods....58
2.2.1 Purification of total RNA....58
2.2.2 Reverse transcription qPCR (RT-qPCR)....59
2.2.3 5'-EU labelling....59
2.2.4 dsRIP analysis....59
2.3 DNA analysis methods....60
2.3.1 Agarose gel electrophoresis....60
2.3.2 Real-time quantitative PCR (qPCR)....60
2.4 Protein analysis methods....61
2.4.1 Preparation of whole cell extracts....61
2.4.2 Cellular fractionation....61
2.4.3 Co-Immuno-precipitation (co-IP)....62
2.4.4 Protein silver staining....63
2.4.5 Western Blot analysis....63
2.4.6 Immuno-fluorescence (IF)....64
2.4.7 Chromatin immuno-precipitation (ChIP)....65
2.4.8 R-loop Immuno-precipitation (R-loop IP)....66
2.5 R-loop analysis methods....67
2.5.1 RNA/DNA hybrid immuno-precipitation (DRIP)....67
2.5.2 RNA/DNA hybrid Slot Blot....69
2.6 Bioinformatic analysis....69
2.6.1 Statistical analysis....69
2.6.2 Gene annotation....70
2.6.3 DRIP-seq data processing....70
2.6.4 ChIP-seq data processing....71
2.6.5 Meta profiles....71
2.7 Tables....72
Chapter 3 Transcriptional regulation and epigenetic signature if the *HTT* gene....77
3.1 The *HTT* gene transcriptional regulation in health and disease....78
3.1.1 *HTT* expression varies in different cell lines....78

3.1.1.1 The *HTT* gene expression in HeLa and SH-SY5Y cells....78
3.1.1.2 Generation of Medium Spiny Neurons (MSNs) from iPS cells....78
3.1.2 Sense and antisense transcription of the HTT gene....84
3.1.3 Transcriptional profile of the HTT gene in HeLa cells....84
3.2 HD mutation affects *HTT* gene expression....89
3.2.1 Nascent and mRNA *HTT* levels are reduced in HD....89
3.2.2 Pol II levels are decreased at the *HTT* promoter in HD....93
3.3 HD affects *HTT* gene epigenetic regulation....93
3.3.1 Active chromatin marks are decreased over *HTT* promoter in HD....93
3.3.2 Repressive chromatin marks are increased over *HTT* in HD....95
3.4 Analysis of available transcriptomic data in HD mouse models....97
3.4.1 Mouse transcriptomic data shows transcription down-regulation in HD....97
3.5 Investigation of HTT protein levels in HD....100
3.5.1 HTT protein levels are decreased in HD cells....100
3.6 Summary....102
Chapter 4 Structural features of CAG triplet repeats....103
4.1 Non-canonical RNA/DNA structures are involved in HD pathology....104
4.2 Non-canonical RNA/DNA structures are formed over CAG repeats of the *HTT* gene....104
4.2.1 R-loops are formed over the *HTT* gene but not enriched at expanded repeats....104
4.2.2 Double-stranded RNAs are enriched on expanded *HTT* transcripts....109
4.2.3 dsRNAs on expanded *HTT* transcripts are sensitive to RNase III digestion....109
4.3 Formation of non-canonical RNA/DNA structures in HD cells....113
4.3.1 Genome-wide R-loop levels in HD cells....113
4.3.1.1 R-loop Slot Blot shows global increase of RNA/DNA hybrids in HD cells....113
4.3.1.2 DRIP-seq analysis in Control and HD cells....118
4.3.2 Genome-wide dsRNA levels in HD cells....123
4.3.2.1 dsRNA Slot Blot shows global decrease of dsRNA in HD cells....123
4.3.2.2 Establishment of the dsRNA immuno-precipitation method....123
4.4 Summary....130
Chapter 5 The HTT protein is involved in transcription and R-loop biology....131
5.1 Introduction....132
5.2 The role of the HTT protein in the nucleus....132
5.2.1 Full-length HTT protein is present in the nucleus and localises to gene promoters....132
5.2.2 HTT interacts with transcriptional and R-loop binding factors....136
5.2.3 HTT forms a complex with RNA/DNA hybrids *in vivo*....138
5.3 Global transcription is impaired in HD cells....141
5.3.1 Global nascent transcription is down-regulated in HD fibroblasts....141
5.4 HTT is involved in the DNA damage response....144
5.4.1 Depletion or mutation of HTT protein leads to DNA damage accumulation....144
5.4.2 CPT-induced DNA damage and R-loop accumulation in siHTT-depleted HeLa cells....146

5.4.3 HTT and XPG proteins are part of the same multi-protein complex 148
5.5 Inflammation phenotypes are observed in HD cells 151
5.5.1 Depletion or mutation of HTT protein leads to enhanced inflammatory response 151
5.5.2 Co-depletion of HTT and XPG proteins leads to accumulation of DNA damage and inflammation 153
5.6 Summary 156
Chapter 6 Discussion and Conclusions 158
6.1 Key findings and novelty 159
6.2 Transcriptional defects at the *HTT* gene in HD 161
6.3 Formation of non-canonical RNA/DNA structures across the *HTT* gene 163
6.4 Formation of non-canonical RNA/DNA structures in HD cells 168
6.5 Global nascent transcription is impaired in HD 169
6.6 HTT and R-loop homeostasis 172
6.7 HTT and DNA damage 173
6.8 HTT and inflammation 175
6.9 Conclusions and future perspectives 178
References 180

List of Figures

Chapter 1 ..... **1**

Figure 1.1. Location of unstable repeat expansions causing human diseases. 3

Figure 1.2. Molecular mechanisms of RE diseases. 6

Figure 1.3. Schematic of the molecular pathology of Huntington's Disease (HD). 14

Figure 1.4. The molecular basis of HD. 16

Figure 1.5. Schematic diagram of the huntingtin (HTT) protein. 21

Figure 1.6. Schematic representing known cellular processes which are affected in HD. 24

Figure 1.7. Structure and biological functions of R-loops. 45

Chapter 3 ..... **77**

Figure 3.1. Analysis of *HTT* gene expression in different cells. 79

Figure 3.2. Generation and characterization of an MSN model for HD. 82

Figure 3.3. GRO-cap across the *HTT* gene in lymphoblast cells. 85

Figure 3.4. Distribution of Pol II and active chromatin marks across the *HTT* gene in HeLa cells 87

Figure 3.5. Distribution of repressive chromatin marks across the *HTT* gene in HeLa cells. 88

Figure 3.6. *HTT* gene expression is impaired in HD lymphoblasts. 90

Figure 3.7. *HTT* gene expression is impaired in HD fibroblasts and MSNs. 92

Figure 3.8. Pol II and active H3K9ac mark are decreased over the promoter-proximal regions of the *HTT* gene in HD lymphoblasts. 94

Figure 3.9. Increased heterochromatin formation over the *HTT* gene in HD lymphoblasts. 96

Figure 3.10 Mouse transcriptomic data shows transcriptional down-regulation in HD. 99

Figure 3.11. Decreased HTT protein levels in HD cells. 101

Chapter 4 ..... **103**

Figure 4.1. RNA/DNA hybrids are formed across the *HTT* gene but not enriched over expanded repeats in HD lymphoblast cells. 105

Figure 4.2. RNA/DNA hybrids are formed across the *HTT* gene but not enriched over expanded repeats in HD neuronal cells. 108

Figure 4.3. Double-stranded RNAs are enriched on expanded *HTT* transcripts. 110

Figure 4.4. Double-stranded RNAs on expanded *HTT* transcripts are sensitive to RNase III treatment. 111

Figure 4.5. R-loop Slot Blot shows an overall increase of RNA/DNA hybrids in HD lymphoblast and fibroblast cells. 114

Figure 4.6. Accumulation of R-loops in HD cells is dependent on transcription. 115

Figure 4.7. S9.6 Slot Blot shows an overall increase of RNA/DNA hybrids in MSN HD cells. 117

Figure 4.8. DRIP-seq and DRIP-qPCR analysis in HD lymphoblast cells. 120

Figure 4.9. DRIP-seq analysis in HD lymphoblast cells. 121

Figure 4.10. DRIP-qPCR analysis in HD MSN cells. 122

Figure 4.11. J2 Slot Blot analysis shows a decrease of dsRNA in HD lymphoblast cells. 124

Figure 4.12. Establishment of the dsRNA-IP method. 126
Figure 4.13. Optimization of the dsRNA-IP method. 127
Figure 4.14. DHX9 levels are decreased over the HTT gene in HD lymphoblast cells. 129
Chapter 5 131
Figure 5.1. The HTT protein is localized in the nucleus. 133
Figure 5.2. The HTT protein binds to gene promoters. 135
Figure 5.3. Endogenous HTT co-IP optimization. 137
Figure 5.4. The HTT protein is in complex with R-loop and transcriptional factors. 139
Figure 5.5. The HTT protein binds to R-loops in vivo. 140
Figure 5.6. Global transcription is downregulated in HD fibroblast cells. 142
Figure 5.7. Accumulation of DNA damage in different HD models. 145
Figure 5.8. siHTT-depleted HeLa cells are susceptible to CPT-induced DNA damage. 147
Figure 5.9. siHTT-depleted HeLa cells show R-loop accumulation upon CTP treatment. 149
Figure 5.10. Depletion of HTT leads to upregulation of XPG protein levels. 150
Figure 5.11. Enhanced inflammatory response in siHTT-depleted HeLa and HD cells. 152
Figure 5.12. Enhanced inflammatory response in cells co-depleted for HTT and XPG proteins. 155
Chapter 6 158
Figure 6.1. Major molecular mechanisms contributing to HD pathogenesis described in this study. 160
Figure 6.2. The *HTT* gene in health and disease 166
Figure 6.3. Interplay between R-loops, DNA damage and inflammation in HD. 177

List of Tables

Chapter 1 ..... **1**
Table 1.1 R-loops and links to human diseases 50
Chapter 2 ..... **54**
Table 2.1 List of cell lines 72
Table 2.2 List of DNA oligonucleotides 73
Table 2.3 List of RNA oligonucleotides 75
Table 2.4 List of antibodies 76

Abbreviations

Listed below are abbreviations used in this thesis.

ABBREVIATION	FULL NAME
3'UTR	3'-untranslated region
5'UTR	5'-untranslated region
AAV	Adeno-associated viruses
Ab	Antibody
ActD	Actinomycin d
ADAR1	Adenosine deaminase, RNA-specific 1
AGS	Aicardi-Goutières syndrome
AID	Activation-induced deaminase
ALS4	Amyotrophic lateral sclerosis type 4
ALT	Alternative lengthening of telomers
AOA2	Ataxia with oculomotor apraxia type 2
AS	Angelman syndrome
ASO	Antisense oligonucleotide
AT	Ataxia telangiectasia
bp	Base pair
BCA	Bicinchoninic acid
BDNF	Brain-derived neurotrophic factor
BSA	Bovine serum albumin
CBP	CREB binding protein
CDNA	Complementary DNA
cGAS	Cyclin GMP-AMP Synthase
Chip	Chromatin immune-precipitation
Co-IP	Co-immuno-precipitation
CNS	Central nervous system
CPT	Camptothecin
CRISPR	Clustered regularly interspaced short palindromic repeats
CSF	Cerebrospinal fluid
CRYO-EM	Cryo-Electron Microscopy
CSR	Class-switch recombination
DDR	dna damage response
D. melalogaster	Drosophila melanogaster
DMEM	Dulbecco's modified eagle's medium
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
DNAse	Deoxyribonuclease
DRIP	DNA/RNA immunoprecipitation
DSB	Double-strand break
ds	Double-stranded
DTT	Dithiothreitol
E.coli	Escherichia coli
EBV	Epstein-Barr virus
EDTA	Ethylenediaminetetraacetic acid
EM	Electron Microscopy
ER	Endoplasmic Reticulum
ESCs	Embryonic stem cells
EtOH	Ethanol
FA	Fanconi Anemia
FMR1	Fragile x mental retardation 1 gene
FRDA	Friedreich Ataxia
FXN	Frataxin gene
FXS	Fragile x syndrome
FXTAS	Fragile x-associated tremor/ataxia syndrome
g	g-force
GAPDH	Glyceraldehyde 3-phosphate dehydrogenase gene

gDNA	Genomic DNA
GWAS	Genome-wide association study
h(s)	Hour(s)
H3	Histone H3
H3K9	Lysine 9 of Histone H3
H3K27	Lysine 27 of Histone H3
H3K36	Lysine 36 of Histone H3
HD	Huntington's Disease
HDAC	Histone Deacetylase
HEAT	Huntingtin, elongation factor-3, protein phosphatase 2A and TOR1
hESC(s)	Human embryonic stem cell(s)
HTT	Huntingtin protein
IFN	Interferon
Ig	Immunoglobulin
IL	Interleukin
IP	Immunoprecipitation
iPSC(s)	Induced pluripotent stem cell(s)
ISG	Interferon stimulated gene
Kb	Kilobase
KDA	Kilodalton
KO	Knock-out
MeOH	Methanol
Min	Minute(s)
mL(s)	Millilitre(s)
mRNA	Messenger RNA
MSN	Medium Spiny Neuron
mtDNA	Mitochondrial DNA
mTOR	Mammalian target of Rapamycin
ncRNA	Non-coding RNA
NCOR	Nuclear receptor co-repressor 2
NF-kB	Nuclear Factor kappa-light-chain-enhancer of activated B cells
NES	Nuclear Export Signal
NLC	Nucleolin
NLS	Nuclear localisation signal
OCT4	Octamer-binding transcription factor 4
ORF	Open Reading Frame
PBS	Phosphate buffered saline
PCR	Polymerase chain reaction
PMSF	Phenylmethylsulfonyl fluoride
Pol II	RNA polymerase II
PolyQ	Polyglutamine
PTM	Post-translational modification
PWS	Prader-willi syndrome
Q	Glutamine (CAG codon)
qPCR	Quantitative polymerase chain reaction
rDNA	Ribosomal DNA
RA	Retinoic acid
RAN	Repeat-associated non-AUG
RBP	RNA-binding protein
RE	Repeat Expansion
RISC	RNA-induced silencing complex
RNA	Ribonucleic acid
RNAi	RNA interference
RNase	Ribonuclease
ROS	Reactive oxidative species
RT-qPCR	Reverse transcription-quantitative polymerase chain reaction
SCAN1	Spinocerebellar ataxia with axonal neuropathy 1
SDS	Sodium dodecylsulfate
SDS-PAGE	Sodium dodecylsulfate polyacrylamide gel electrophoresis
SEM	Standard error of the mean

siRNA	Small interfering RNA
SMRF-Seq	Single-molecule R-loop footprinting
SNP	Single-nucleotide polimorphism
SRSF6	Serine/arginine-rich splicing factor 6
SSB	Single-strand break
ss	Single-stranded
STING	Stimulator of Interferon Genes
TBK1	TANK binding kinase 1
TERRA	Telomeric repeat-containing RNA
TCR	Transcription-coupled nucleotide excision repair
TOP1	Topoisomerase I
TOP3B	Topoisomerase III b
TNFα	Tumor necrosis factor-α
TSS	Transcription start site
TTS	Transcription termination site
UBE3A	Ubiquitin protein ligase E3A
WB	Western Blot
WT	Wild-type
XPF	Xeroderma pigmentosum group F
XPG	Xeroderma pigmentosum group G
ZFN	Zing finger nuclease

Chapter 1

Introduction

1.1 Repeat expansion diseases

Around forty percent of the human genome is composed of repetitive sequences (Gemmell, 2021). Tandem repeats (TRs) are regions comprising parallel repeating units and are usually classified according to the size of the repeating unit as satellites (>60 base pairs), minisatellites (10–60 base pairs), or microsatellites/short tandem repeats (STRs) (<9 base pairs) (Gymrek *et al.*, 2016), with STRs making up ~ 1-3% of the human genome. Although the biological roles of STRs are largely unknown, recently, a growing body of research has implicated them in the regulation of gene expression (Fotsing *et al.*, 2019; Wright & Todd, 2023). However, these repeats are often unstable and hypervariable. Changes in their lengths can drastically alter phenotypes and cause more than 50 debilitating neurodegenerative, neurological, and neuromuscular human genetic disorders, collectively known as repeat expansion (RE) disorders (Depienne & Mandel, 2021; Malik *et al.*, 2021) (Fig. 1.1). These disorders are associated with genetic anticipation, earlier disease onset and increased severity in successive generations. Furthermore, for many of these diseases, it has been demonstrated that the length of the repeat expansion is inversely correlated with the age of the onset of the disease (Paulson, 2018). These repeat expansion-associated diseases can be classified into two types depending on their occurrence within coding or non-coding gene regions and further classified on the basis of different pathogenic mechanisms (Malik *et al.*, 2021) (Fig. 1.1). When the expansion is located in the coding region of a gene, it causes the production of a toxic mutant protein, such as in Huntington's Disease (HD). In contrast, expansions in non-coding gene regions can alter transcription and protein expression, or may influence splicing and aspects of antisense RNA regulation, as observed in Fragile X syndrome (FXS) and Friedreich's Ataxia (FRDA) (Malik *et al.*, 2021).

1.1.1 Coding and Non-Coding RE diseases

Massive expansions can occur in transcribed but non-coding regions of genes, such as 5′ untranslated regions (UTRs), introns, or 3′UTRs, and can therefore be present in the pre-messenger (pre-m) RNA and/or the mRNA but not translated into protein (Fig. 1.1). Long expanded repeats such as the CGG repeat in FXS, the GAA repeat in FRDA, and the CTG repeat in myotonic dystrophy

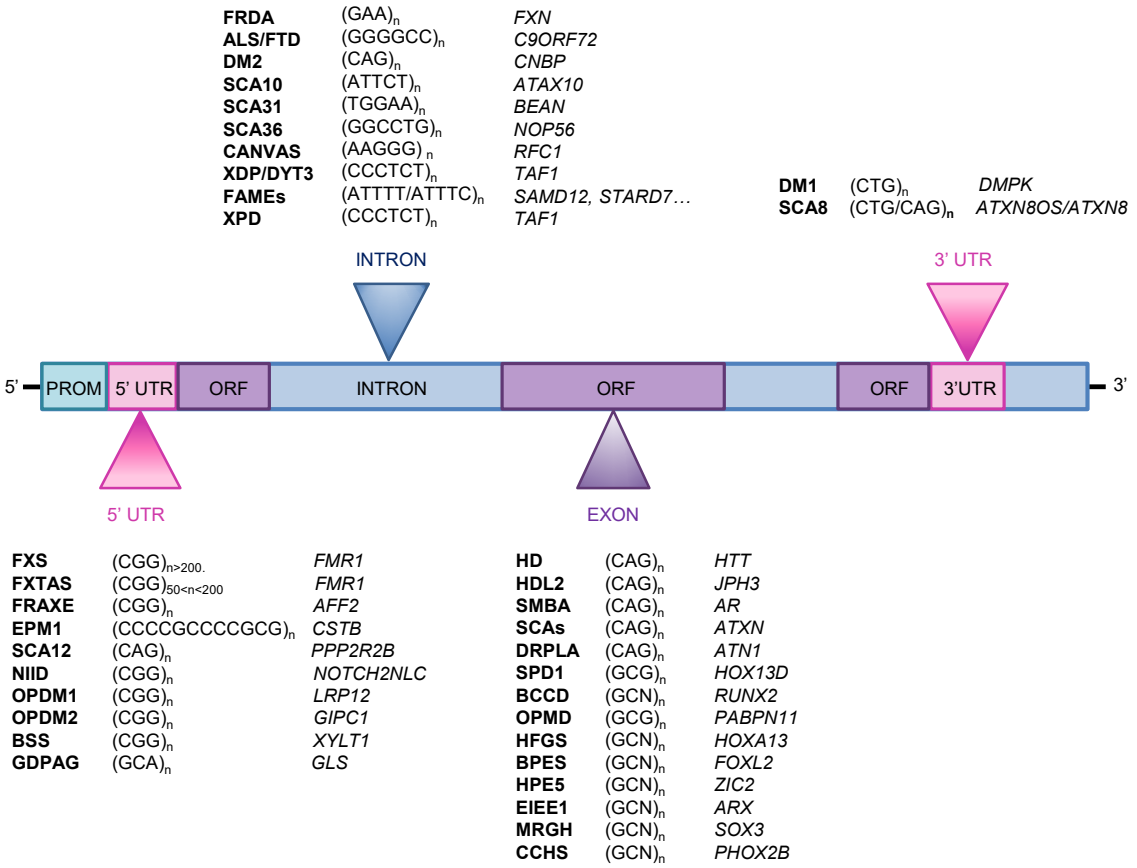


Figure 1.1. Location of unstable repeat expansions causing human diseases.
Diagram of a generic gene indicating the location of disease-causing repeats. Repeat expansions can occur within the 5'UTRs, exons, introns or 3'UTRs of a gene. Depending on their genomic location expanded repeats can trigger different disorders and have a variety of effects at DNA, RNA and protein levels. The triangles above and below the gene indicate the location of the repeats. Details on the sequences of the expanded repeats and the affected genes are listed for each associated human disease. The pathogenic mechanisms for each repeat expansion disorder are determined by a variety of factors and can broadly be split into toxic gain-of-function (GOF) or deleterious loss of gene function (LOF). Both GOF and LOF mechanisms typically contribute to the pathology of the disease.
ALS (amyotrophic lateral sclerosis); BCCD, brachydactyly and cleidocranial dysplasia; BPES, Blepharophimosis Syndrome; BSS, Baratela-Scott syndrome; CANVAS, cerebellar ataxia, neuropathy and vestibular areflexia syndrome; CCHS, congenital central hypoventilation syndrome; DM1, myotonic dystrophy type 1; DM2, myotonic dystrophy type 2; DRPLA, dentatorubral-pallidoluysian atrophy; EIEE1, early infantile epileptic encephalopathy type 1; EPM1, progressive myoclonus epilepsy type 1 (Unverricht-Lundborg disease); FAME, familial adult myoclonic epilepsy; FRAXE, fragile XE syndrome; FRDA, Friedreich's ataxia; FTD/ALS, frontotemporal dementia / amyotrophic lateral sclerosis; FXS, fragile X syndrome; FXTAS, fragile X-associated tremor ataxia syndrome; GDPAG, global developmental delay, progressive ataxia, and elevated glutamine; HD, Huntington's Disease; HDL2, Huntington's Disease-like 2; HFGS, hand-foot-genital syndrome; HPE5, holoprosencephaly type 5; MRGH, mental retardation with isolated growth hormone deficiency; OPMD, oculopharyngeal muscular dystrophy; NIID, neuronal intranuclear inclusion disease; SBMA, spinal and bulbar muscular atrophy; SPD1, synpolydactyly type 1; SCA, spinocerebellar ataxia; XDP, X-linked dystonia parkinsonism; UTR (untranslated region). Adapted from (Depienne *et al.,* 2021).

(DM) are unstable and they can expand by hundreds to more than one thousand repeats during transmission from one generation to the next (Depienne & Mandel, 2021). FRDA is the most common hereditary ataxia, and it is primarily caused by a homozygous GAA trinucleotide repeat expansion mutation within intron 1 of the frataxin *(FXN)* gene and results in the deficiency of the essential mitochondrial protein frataxin (Delatycki *et al.*, 2000). This leads to progressive mitochondrial dysfunction, oxidative stress, and cell death (Pandolfo & Pastore, 2009). Several models have been proposed to explain how repeats in non-coding regions could lead to the down-regulation of affected genes. For example, it has been suggested that repeat expansions can cause the formation of unusual DNA structures such as hairpins, R-loops, quadruplexes, and DNA triplexes, which could interfere with transcription (Khristich & Mirkin, 2020). Moreover, transcription through triplet repeats could trigger instability by exposing single-stranded RNA as the RNA Polymerase II (Pol II) moves through the repeats, allowing the repeats to form hairpins and slipped strand structures (Salinas-Rios *et al.*, 2011).

In contrast, even moderate expansions of roughly 15-100 repeats are sufficient to cause disease when they occur in the coding regions of genes (Fig. 1.1). Repeat expansions in coding regions are mainly represented by trinucleotide CAG repeats and mostly result in the formation of polyglutamine (polyQ) tracts in the encoded protein, which can gain a toxic function and/or form pathogenic aggregates (Paulson, 2018). These include many hereditary conditions such as HD, several forms of spinocerebellar ataxias (SCAs), and spinobulbar muscular atrophy (SBMA) (Paulson, 2018).

Collectively, the molecular processes that contribute to the development of repeat expansion disorders are quite diverse. This is in part due to differences in the repeat sequence, size, and location of the affected gene, as well as whether or not the repeated sequence is translated into a protein. The mechanisms that directly contribute to the pathology of repeat expansion diseases can be broadly classified into three interrelated categories: toxic gain-of-function of proteins, RNA-induced toxicity and DNA-mediated mechanisms. Gain-of-function of toxic protein can arise either from canonically-translated proteins harbouring expanded repeats, which are prone to form protein aggregates, or from the non-canonical repeat-associated non-AUG (RAN) translation of toxic repeat peptides (Depienne

& Mandel, 2021). RNA-induced toxicity involves bi-directional transcription, formation of toxic RNA foci and sequestration of RNA-binding proteins. Finally, DNA-based mechanisms include repeat-induced transcriptional gene silencing and the formation of non-canonical RNA/DNA structures followed by epigenetic dysregulation (Fig. 1.2). In the following Sections, I discuss each of the main molecular mechanisms of pathogenesis in RE diseases and how synergy between mechanisms produces the unique and complex pathology of each disease.

1.1.2 Protein-induced toxicity in RE diseases

Previously, a significant amount of research on RE disorders was focused on examining the mechanisms through which certain proteins implicated in these disorders undergo changes in their functional capabilities, either through loss- or gain-of-functions (Fig. 1.2). For example, FRDA and FXS, two non-coding RE disorders, are caused by a loss-of-function of the FXN protein and Fragile X mental retardation protein (FMRP), respectively (Campuzano *et al.*, 1996; Santoro *et al.*, 2012). In contrast, one of the most extensively studied hallmarks among coding RE disorders is the presence of abnormally folded pathogenic proteins that can gain novel toxic functions, form aggregates, and ultimately cause neuronal death (Sweeney *et al.*, 2017). Repeat-containing peptides may be produced both via canonical translation of the repeat in a coding region or via repeat-associated non-AUG (RAN) translation. At least nine disorders, including HD, several SCAs and SBMA, are thought to be caused by CAG repeat expansions in the coding sequence that result in the expression of polyQ proteins (Lieberman *et al.*, 2019). For example, HD may be primarily caused by a gain-of-function of the mutant huntingtin (mHTT) protein as expression of the short HTT exon 1 protein containing expanded repeat or insertion of a polyQ-encoding CAG repeat in the unrelated *Hprt* gene is sufficient to cause neurotoxicity in mice (Schilling *et al.*, 1999; Ordway *et al.* 1997). RAN translation instead can initiate in the absence of an AUG start codon, in multiple reading frames, through an expanded repeat to produce homopolymeric or dipeptide-repeat-containing proteins (Kearse & Todd, 2014). For example, in FXS, despite CCG repeats being localized in putative non-coding regions, RAN translation through these repeats leads to the production of toxic peptides, which have been shown

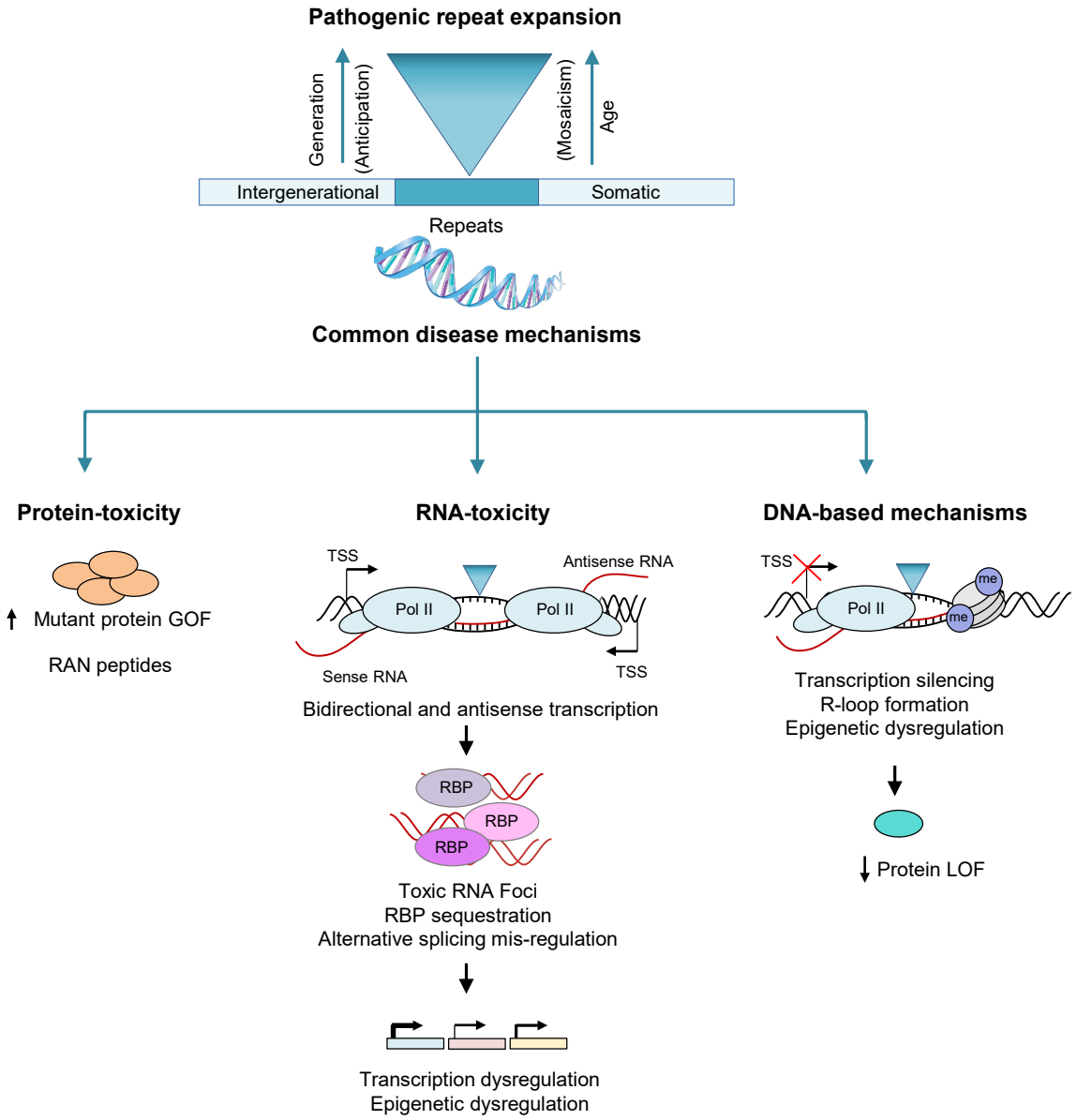


Figure 1.2. Molecular mechanisms of RE diseases.
Repeat expansions in coding sequences result in the production of mutant proteins, which are prone to form toxic aggregates. RAN peptides can also contribute to toxicity. Bidirectional and antisense transcription of expanded repeats favour the formation of repeat-including RNAs, which can fold into complex structures, including hairpins and G-quadruplexes, and can aberrantly sequester RNA-binding proteins (RBPs) into RNA foci. Sequestration of RBPs involved in splicing could result in splicing mis-regulation and global transcription dysregulation and epigenetic changes, common mechanisms of RE disorders. In contrast, repeat-driven transcriptional silencing could result from hypermethylation of DNA and/or the formation of unusual RNA/DNA structures, including R-loops, at the gene promoter of the expanded gene. This could result in partial or complete loss of the native protein encoded by the gene harbouring the repeats. Both R-loops forming over expanded repeats and RNA foci can activate DNA damage and inflammatory responses (not shown). Finally, for all of these mechanisms, the length of the expansion is inversely proportional to the age at onset and correlates positively with the severity of the disease. Intergenerational and somatic instability of the repeats amplify their potential for toxicity and often affect the age of disease onset. GOF, gain-of-function; LOF, loss-of-function; Pol II, RNA polymerase II; TSS, Transcriptional Start Site; RAN, Repeat-associated non-AUG. Diagram was adapted from Malik *et al*. 2021.

to interfere with a variety of cellular functions in several disease models (Sellier *et al.*, 2017; Ma *et al.*, 2019).

Repeats in proteins can affect protein folding and in HD, the structure of the mHTT protein containing the polyQ-tract differs from that of HTT. The polyglutamine tract has been shown to act as a flexible domain, and this flexibility is impaired with the expanded polyglutamine tracts of the mHTT (Caron *et al.*, 2013). SBMA is caused by a CAG expansion trinucleotide repeats in the Androgen Receptor (*AR*) gene. Interestingly, the polyQ-containing AR proteins become toxic only after activation by the natural androgenic ligands, possibly due to aberrant androgen-induced conformational changes of the polyQ AR, which generate misfolded species. Toxic polyQ AR proteins form nuclear aggregates which trigger cytotoxicity and cell death (Jochum *et al.*, 2012).

Protein-coding transcripts from repeat-containing loci were originally considered to be the only molecular players in disease manifestation, whereas very little was known about the role of RNA in these processes. Recently, a toxic role for mutant RNA transcripts has also been proposed. Thus, a much more complex pathogenic picture is emerging (Malik *et al.*, 2021) (Fig. 1.2).

1.1.3 RNA-induced toxicity in RE diseases

Genome-wide studies have revealed that only ~2% of transcripts are translated into functional proteins. The majority of transcripts are therefore “non-coding” RNAs, which can play crucial roles tightly regulating gene expression (Statello *et al.*, 2021). Interestingly, recent research has led to the proposal that RNA toxicity and aberrant RNA processing may contribute to many repeat disorders (Liu *et al.*, 2017). For example, RNA transcripts can accumulate in toxic RNA foci and cause the aberrant sequestration of RNA-binding proteins (RBPs). In addition, transcription through expanded repeats can lead to the production of antisense, short non-coding, bi-directional transcripts. Finally, expanded repeats can disrupt various transcription and RNA-related processes, including RNA splicing (Fig 1.2).

Toxic RNA foci

RNA expressed from expanded repeats can accumulate in cells in the form of complexes commonly referred to as RNA foci (Fig. 1.2). Myotonic dystrophy type 1 (DM1) results from an

expanded CTG repeat in the 3′UTR of the *DMPK* gene which encodes myotonic dystrophy protein kinase. DM1 is the best-characterized example of neuromuscular disorder with complex RNA-related aetiology linked to RNA gain-of-function (Hunter *et al.*, 1992). The expanded *DM1* transcripts have been shown to form toxic RNA foci in the cell nucleus and sequester RBPs (Taneja *et al.*, 1995; Jiang *et al.*, 2004). The sequestration of RBPs disrupts the ability of these factors to regulate splicing and leads to mis-splicing of their target genes. In particular, *DMPK* mRNA foci recruit members of the muscle blind-like family of RNA splicing proteins, MBNL1, 2 and 3, and disrupt their function and normal nuclear distribution (Jiang *et al.*, 2004). Furthermore, RNA foci have been established as a common hallmark of many RE diseases, including DM2, amyotrophic lateral sclerosis/frontotemporal dementia (ALS/FTD) and many SCAs (Zhang & Ashizawa, 2017; Malik *et al.*, 2021)

Bi-directional transcription of expanded repeats

Recently, a role for antisense transcripts in the pathogenesis of RE disorders has emerged (Fig. 1.2). Antisense transcription is a widespread phenomenon in the human genome that leads to the production of the so-called natural antisense transcripts (NATs), which may play important roles in RNA stability and transcriptional regulation (Wight & Werner, 2013). NATs can regulate gene expression by recruiting chromatin-remodelling complexes, resulting in epigenetic changes and transcriptional deregulation of target genes (Wanowska *et al.*, 2018). In the context of DM1, in addition to already mentioned sense transcription of CUG repeat-containing mRNA, a new potentially toxic mechanism involving antisense transcription of *DM1* (*DM1-AS*) has been described. *DM1-AS* RNA foci have been detected in the heart, muscle and brain of human and mouse DM1 embryos and fetuses (Michel *et al.*, 2015). Although the role of *DM1-AS* transcripts remains unknown, these transcripts have been recently proposed to belong to the class of long non-coding RNAs (Gudde *et al.*, 2017). The mechanisms underlying DM1 pathology may therefore involve toxic expanded sense and antisense *DMPK* transcripts that form numerous nuclear foci at early stages during development. In FRDA, an antisense transcript *FAST-1* (*FXN* Antisense Transcript-1) has been detected and shown to influence *FXN* transcription regulation (Mikaeili *et al.*, 2018).

Alternative splicing in RE disorders

Multiple studies have linked mis-splicing events with human diseases. Alternative splicing is the process by which different coding regions of the pre-mRNA are included in the mature mRNA (Marasco & Kornblihtt, 2022) (Fig. 1.2). In HD, splicing of the *HTT* mRNA exon 1 to exon 2 is affected by the length of the CAG repeat, resulting in the expression of a small, polyadenylated mRNA that encodes a highly pathogenic exon 1 huntingtin protein (Sathasivam *et al.*, 2013). Given the close relationship between splicing and transcriptional speed, it has been demonstrated that transcription slows down after transcribing through the repeat and speeds up again towards the end of intron 1, favouring the production of the incompletely spliced *HTT* exon 1 transcript (Neueder *et al.*, 2018). Transcription deceleration at GAA repeats results in aberrant alternative splicing and generation of an early terminated *FXN* transcript (*FXN*-ett). The levels of *FXN*-ett correlate with the length of GAA repeats and do not contribute to the synthesis of functional frataxin (Li *et al.*, 2022). In the case of DM1, toxic RNA foci formed by *DMPK* transcripts sequester splicing factors, such as MBNLs, resulting in altered RNA splicing events in the genome that lead to DM1 symptoms (Jiang *et al.*, 2004; Wang *et al.*, 2012a).

1.1.4 DNA-mediated mechanisms associated with RE diseases

1.1.4.1 Unusual RNA/DNA structures formed by expanded repeats

The sequence of nucleotides plays an important role in the structures adopted by DNA and RNA. Indeed, certain sequence characteristics favour the formation of alternative secondary structures. Notably, repeats can form multiple non-canonical structures, including RNA/DNA hybrids, slipped strand DNA structures, hairpins, triplexes, and G-quadruplexes, due to inter- and intramolecular interactions within/between the repetitive elements. The formation of these structures can increase gene instability and affect gene expression processes including DNA replication, repair, and transcription (Fig. 1.2). In DM1, the CUG repeats of the *DMPK* transcript have been shown to form “slippery” hairpins, which associate anomalously with RBPs (see Toxic RNA foci Section 1.1.3). Furthermore, the stability of these hairpins has been demonstrated to increase with the repeat length (Napierala & Krzyzosiak, 1997; van Cruchten *et al.*, 2019). In SCA1, imperfect CAG repeats

in mutant *SCA1* transcripts have been demonstrated to fold into undistorted hairpin structures (Sobczak & Krzyzosiak, 2004). In ALS/FTD, the GGGGCC expansion in the non-coding region of the *C9orf72* gene was found to generate both DNA and RNA polymorphic structures, such as G-quadruplexes and RNA/DNA hybrids, which alter transcription and promote the sequestration of unique RBPs (Haeusler *et al.*, 2014; Abu Diab *et al.*, 2018). Moreover, CAG repeats have been also shown to form extensive RNA secondary structures *in vitro* (de Mezer *et al.*, 2011; Wojciechowska & Krzyzosiak, 2011). In addition, R-loops, composed of RNA/DNA hybrids and a displaced single-strand DNA, are implicated in the pathogenesis of several expanded repeat disorders affecting transcription, DNA replication and genomic stability (Groh & Gromak, 2014; Freudenreich, 2018). Although the impact of R-loop formation at repeat expansions is still unclear, R-loop enrichment has been demonstrated in ALS/FTD, FRDA, FXTAS and FXS diseases *in vivo*, suggesting that their formation may be a general feature of RE diseases (Colak *et al.*, 2014; Groh *et al.*, 2014a; Loomis *et al.*, 2014). Our lab has demonstrated the pathological formation of R-loops on expanded repeats of *FXN* and *FMR1* genes in patient cells, associated with FRDA and FXS, respectively (Groh *et al.*, 2014). In particular, it was demonstrated that pathological R-loops promote the deposition of repressive chromatin marks and cause transcription silencing of the host-expanded *FXN* gene in FRDA (Groh *et al.*, 2014).

1.1.4.2 Epigenetic dysregulation in RE diseases

It is evident that epigenetic gene regulation is integral to the pathogenesis of several RE disorders (Nageshwaran & Festenstein, 2015) (Fig. 1.2). Repeat expansions can alter local genome architecture and the expression of neighbouring genes. Studies have shown a statistical enrichment of specific epigenetic marks around expansion-prone repeats. The genomic coordinates of the repeat tracts were determined at base pair resolution and the epigenetic profiles around repeats were computationally established (Essebier *et al.*, 2016). Interestingly, these studies showed that chromatin and DNA changes around expanding repeat loci have distinct epigenetic features that differ from those found in a random set of repeats (Essebier *et al.,* 2016). Specific histone modifications have been associated with gene silencing surrounding the GAA repeat expansion in

the *FXN* intron 1 in FRDA. In detail, together with increased heterochromatin formation (Greene *et al.*, 2007; Groh *et al.*, 2014a), a decrease in active chromatin marks, including acetylation of the histone H3 and H4 lysine residues, has been demonstrated to occur in regions surrounding the GAA expansion in multiple FRDA cellular models (Soragni *et al.*, 2008; Chutake *et al.*, 2015). To reverse FXN silencing, histone deacetylase inhibitors (HDACi) have been shown to increase frataxin expression (Herman *et al.,* 2006; Polak *et al.,* 2016). It was also shown that heterochromatinization induced by GAA-repeat hyper-expansion in FRDA can be reduced upon HDAC inhibition by vitamin B3 (nicotinamide) in several FRDA models and patients (Chan *et al.,* 2013). More recently, HDCAis have been demonstrated to partially restore cellular pathways related to the neuronal function that was affected by *FXN* loss in FRDA neuronal cells (Lai *et al.,* 2019). In DM1, the expanded CTG repeat is also thought to modify chromatin surrounding *DMPK* and cause altered expression of the neighbouring gene *SIX5* (Cho *et al.*, 2005; Lopez Castel *et al.*, 2011). These local chromatin modifications may be initiated by the presence of bidirectional transcripts through an RNA interference (RNAi) mechanism (Cho *et al.,* 2005). DNA methylation of *DMPK* around repeats is altered and inhibition of the DNA methyltransferase (DNMT) by 5-aza-deoxycytidine promotes an increase in the number of repeats in DM1 cells (Gorbunova *et al.,* 2004). This destabilization of repeat numbers mediated by DNMT inhibitors provides strong evidence that epigenetic processes play a role in regulating the length of expansion repeats and modulating disease phenotype.

Repeat expansions also cause global epigenetic dysregulation. In HD, epigenetic dysregulation contributes to the alteration of gene expression and neuronal dysfunction (Bassi *et al.,* 2017). Although a precise mechanism remains still unknown, impairment of acetylation and methylation homeostasis is thought contribute to HD pathogenesis. Hypoacetylation of histones H3 and H4 has been shown in many HD cellular and animal models (Sadri-Vakili *et al.,* 2007). The importance of altered acetylation in the context of HD has been also supported by the effect of several HDACis in several HD models. For example, neuronal degeneration is decreased in *D. melanogaster* models of HD following treatment with HDACis (Steffan *et al.,* 2001). In FXS, CGG repeat expansions have been shown to cause severe disruption to chromatin boundaries (Sun *et al.,* 2018).

The ones presented here are just a few of the myriad ways in which epigenetic processes underlie aspects of RE disorder pathogenesis. Understanding the mechanisms of epigenetic regulation and the link between epigenetic disruption and disease phenotype will bring us closer to developing effective and selective epigenetic therapies for these diseases.

1.2 Huntington's Disease

1.2.1 The pathology of HD

Huntington's Disease (HD) is a fatal hereditary neurodegenerative disorder that is characterized by a loss of motor control and cognitive ability and, ultimately, death. HD affects 1/10,000 individuals and is characterized by neurodegeneration of specific brain areas, with the striatum being the most severely affected, followed by progressive neuronal loss and shrinkage of the affected brain regions (Novak & Tabrizi, 2010) (Fig 1.3). The degeneration of the striatum in the patients' brain results in a loss of inhibitory control over motor movements, leading to chorea (involuntary jerky movements) and dystonia (sustained muscle contractions). Moreover, selective degeneration of the medium spiny neurons (MSNs), which are involved in the regulation of cognitive and motor functions, leads to cognitive impairment, including deficits in attention, working memory, and executive function (Novak & Tabrizi, 2010).

HD is an autosomal dominant disease caused by an expanded CAG repeat stretch in exon 1 of the huntingtin *(HTT)* gene that translates into a polyQ tract in the huntingtin (HTT) protein (MacDonald *et al.,* 1993) (Fig. 1.3-1.4). In humans, the normal range for CAG repeats is between 10-35 and individuals with 36 or more CAG repeats are at risk for developing the disease. Individuals with 36-39 CAG repeats may or may not develop the disease, but they are at risk of passing on the expansion to their offspring. This range is considered the "intermediate" or "reduced penetrance" range, as not all individuals with these repeat lengths will develop the disease. Individuals with 40 or more CAG repeats have a high likelihood of developing HD, with earlier onset and more severe symptoms associated with longer repeat lengths (Juvenile HD) (Fig. 1.4A) (Saudou & Humbert, 2016). Despite the well-known genetic origin of HD, the molecular mechanisms that drive disease pathogenesis are still not fully understood. As for many neurodegenerative diseases, the typical hallmark of HD is the presence of toxic aggregates in HD patients' brains (Hoffner *et al.*, 2007), mitochondrial dysfunction (Carmo *et al.*, 2018) and disruption of the neuronal circuitry (Miller *et al.*, 2011). Although the role of RNA toxicity in HD has not yet been fully characterized, new pathogenic roles of expanded CAG repeat-containing RNA are emerging (see Section 1.2.6).

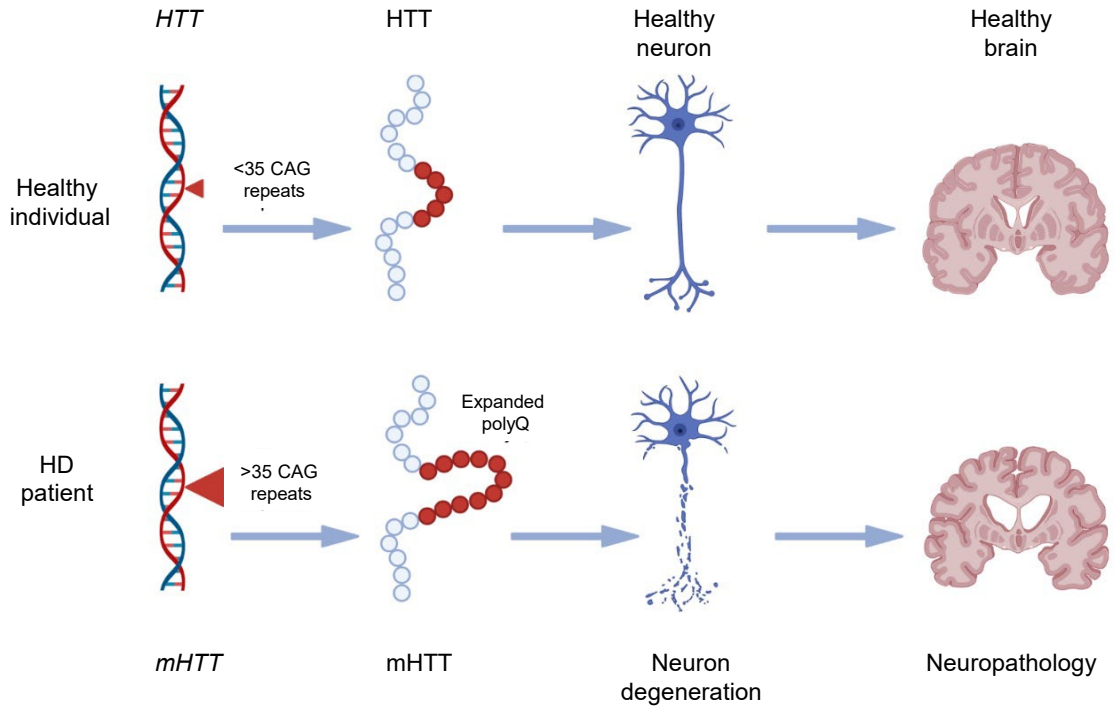


Figure 1.3. Schematic of the molecular pathology of Huntington's Disease (HD).
In healthy individuals, the huntingtin (*HTT*) gene contains up to 35 CAG repeats (red triangle) and encodes the HTT protein. The HTT protein is involved in several molecular mechanisms and contributes to neuronal survival. In HD patients, the huntingtin (*HTT*) gene carries expansions of 36 or more CAG repeats. The expanded CAG tract in exon 1 of the *HTT* gene translates into a polyglutamine tract (polyQ) in the mutant huntingtin (mHTT) protein. mHTT is cleaved by proteases generating toxic N-terminal polyQ fragments which form aggregates and/or translocate into the nucleus where they interfere with transcription, causing neuronal death. The neuropathology of HD is characterized by a general shrinkage of the patient's brain and specific neurodegeneration of medium spiny neurons (MSNs) in the striatum. Additionally, shrinkage of the cerebral cortex and other brain regions is evident, especially in more advanced cases.

1.2.2 The Huntingtin gene

The *HTT* gene was the first disease-associated gene mapped to a human chromosome (Gusella *et al.*, 1983). Only in 1993, scientists identified the DNA sequence and determined the precise nature of the HD-associated mutation in the *HTT* gene (MacDonald *et al*., 1993). Since then, most of the studies focused on the study of the role of the mHTT protein whereas relatively few studies have been carried out to characterize the regulation of *HTT* expression in health and disease. The *HTT* gene is a complex gene spanning approximately 180 kilobases of genomic DNA on the short arm of chromosome 4 (Fig 1.4). It contains 67 exons, which undergo alternative splicing to generate multiple isoforms of the HTT protein. Moreover, the expanded CAG repeats in the *HTT* gene result in the transcription of unusually long and unstable RNA molecules, that can form secondary structures and interact with other cellular components, as well as in the production of mutant proteins (mHTT) that have a high propensity to form aggregates. These aggregates are a defining characteristic of HD (Fig. 1.4 B and C) (Saudou & Humbert, 2016).

The *HTT* gene is highly conserved and found in many species, including humans, primates, rodents and even some invertebrates. The conservation of *HTT* across different species underscores its important biological function in normal development (Lin *et al.*, 1995). The *HTT* gene is expressed ubiquitously in both rodents and humans, with the highest levels in the brain (Kauffman *et al.*, 2003). The *Htt* mouse gene is on chromosome 5 and shows 86% homology to the human *HTT* gene. One main difference between *HTT* and *Htt* is that only 6 CAG physiological repeats are present in mice whereas the number of CAG repeats in unaffected human individuals is 20-22 (Barnes *et al.*, 1994). The human *HTT* promoter has high GC content and lacks TATA and CCAAT regulatory elements. Furthermore, unique features of the human *HTT* promoter include two 20 bp perfect repeats, two 17 bp direct repeats, one full *Alu* element and one *Alu* element 3' fragment, both of which are in reverse orientation to the direction of *HTT* transcription. These regulatory elements may serve as unique binding sites for trans-acting factors that may either direct transcriptional initiation or enhance expression of the *HTT* gene (Lin *et al.*, 1995). Gene expression is a highly-controlled

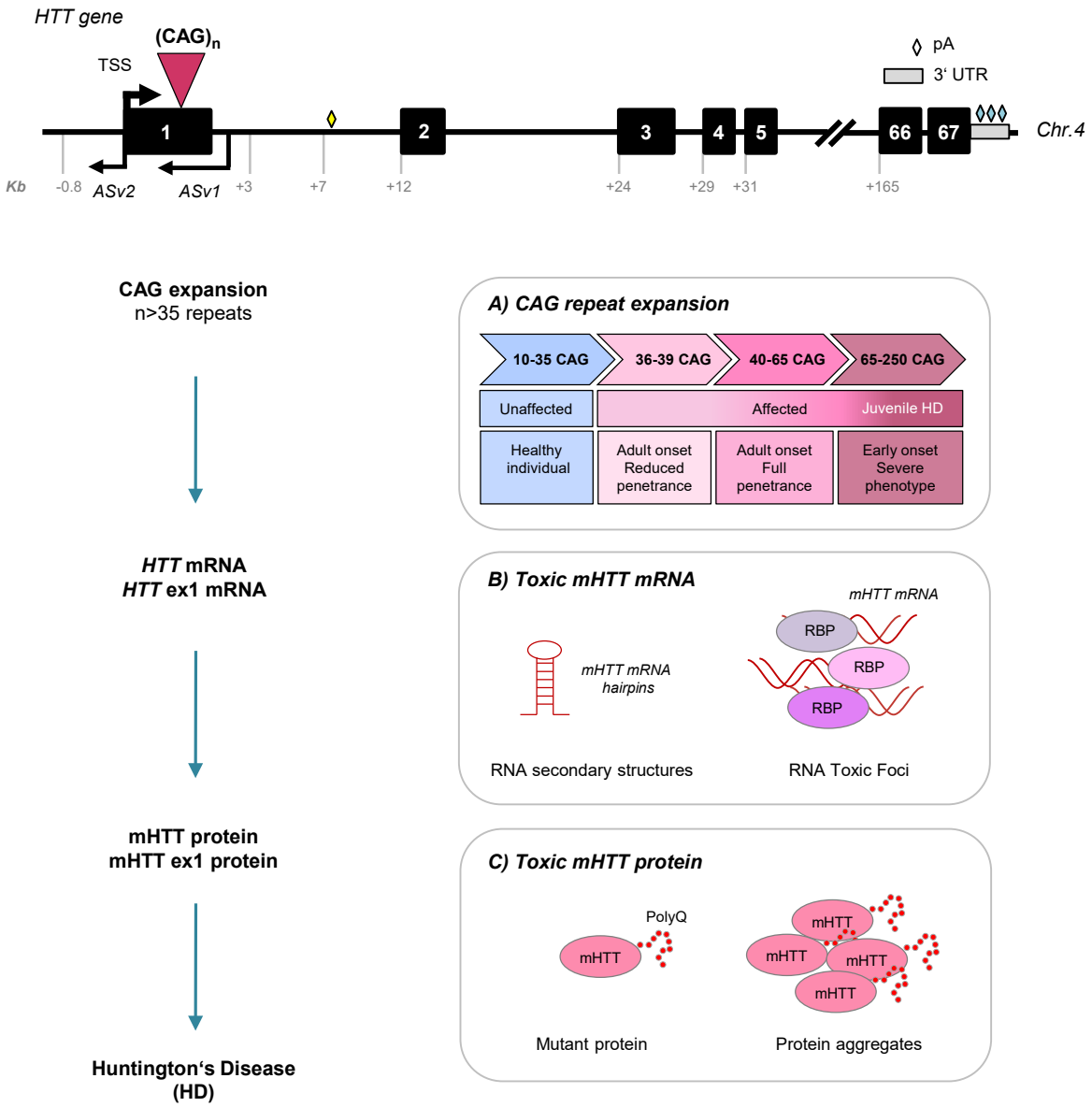


Figure 1.4. The molecular basis of HD.
The diagram depicts the huntingtin (*HTT*) gene. Black rectangles are exons, the small grey rectangle is the 3'UTR, lines are introns, and the red triangle represents the $(CAG)_n$ expansion. TSS is the transcriptional start site. The small diamonds represent poly A sites: a cryptic poly A site in the first intron of the *HTT* gene (yellow, Sathasivam *et al*., 2013) and three poly A sites at the 3'UTR (blue, Romo *et al.,* 2017). The small numbers in grey indicate the distances to TSS in kilobases. *ASv1* and *ASv2* indicate *HTT* antisense transcripts (Chung *et al,*. 2011). **A)** In humans, the normal range for CAG repeats is between 10 and 35. Individuals with 36-39 CAG repeats may or may not develop the disease (reduced penetrance), whereas individuals with 40 or more CAG repeats have a high likelihood of developing HD (full penetrance) with earlier onset and more severe symptoms associated with longer repeat lengths (Juvenile HD). The expansion of the CAG repeats above 35 leads to the production of mutant *HTT* (*mHTT*) mRNAs, which forms RNA foci and sequester RNA-binding proteins (RBPs) (shown in **B**), and mHTT proteins, which are prone to aggregate (shown in **C**). Moreover, expanded repeats lead to premature transcription termination at the cryptic poly A site in the first intron of the *HTT* gene with the production of *HTT* ex1 mRNAs and highly toxic mHTT ex1 mutant proteins. The presence of m*HTT* mRNAs and toxic mHTT proteins in patient brains causes HD.

process that involves the coordinated action of transcription factors (TFs) that bind to cis-regulatory elements to orchestrate transcription. Indeed, the human *HTT* promoter contains predicted binding sites for several TFs such as NF-κB, AP2, Sp1, AP4 and AML1 (Holzmann *et al.*, 2001; Wang *et al.*, 2012b; Becanovic *et al.*, 2015). Moreover, the identification of a binding site for the signal transducer and activator of transcription 1 (STAT1) transcription factor, in the *HTT* intron 5, showed that *HTT* transcription can be also modulated by regulatory sites within the *HTT* gene body (De Souza *et al*, 2018). Furthermore, numerous non-coding single-nucleotide polymorphisms (SNPs) have been detected in the *HTT* gene (Becanovic *et al.*, 2015). SNPs have been identified as potential modifiers of disease progression, altering the binding affinity of TFs to their binding sites, ultimately changing gene transcriptional activity. Interestingly, several of these SNPs are associated only with the mutant *HTT* allele, and their presence could be used for selective silencing of the mutant allele.

Alternative cleavage and polyadenylation events in the 3'UTR of the *HTT* nascent RNA produce three *HTT* mRNA isoforms, which vary in abundance in patient and control neural stem cells, fibroblasts, motor cortex, and cerebellum (Lin *et al.*, 1993; Romo *et al.*, 2017; Romo *et al.*, 2018) (Fig. 1.4). These *HTT* 3'UTR isoforms have different localizations, half-lives, polyA tail lengths, microRNA sites and RNA-binding protein sites. However, their function is yet not clear. Additionally, research has shown that the CAG repeat-dependent aberrant splicing of *HTT* exon 1 leads to the production of a short polyadenylated mRNA, which translates into a toxic exon 1 HTT protein (Sathasivam *et al.*, 2013; Neueder *et al.*, 2017; Neueder *et al.*, 2018; Franich *et al.*, 2019) (Fig. 1.4). Initial studies have shown that manipulation of the expression levels of the Serine/arginine-rich splicing factor 6 (SRSF6), which is predicted to bind CAG repeats, could affect the extent of this incomplete splicing event (Sathasivam *et al.*, 2013; Neueder *et al.*, 2018). However, it was recently demonstrated that the silencing of SRSF6 did not modulate incomplete splicing of the *HTT* gene in multiple HD mouse models (Mason *et al.*, 2020). Therefore, the factors that bind to *HTT* mRNA and influence either the efficiency of splicing and/or the probability of premature termination remain unknown.

Antisense transcription has been shown to influence *HTT* gene expression. An *HTT* antisense transcript containing the CAG repeat tract expansion was identified in brain frontal cortex tissues

from healthy and HD patients (Chung *et al.*, 2011). This antisense transcript is 5' capped, polyA tailed, contains three exons, and can be alternatively spliced into two smaller transcripts (*HTTASv1* and *HTTASv2*), one of which contains the CAG expanded tract (Fig. 1.4). However, contrasting results about its potential role are reported in the literature. It was shown that the expression of the *HTTAS* negatively correlates with the number of CAG repeats in *HTT* and that its expression negatively regulates *HTT* expression. Indeed, overexpression of *HTTAS* in HeLa and SH-SY5Y cells specifically decreased the levels of the endogenous *HTT* transcript, while an increase of *HTT* was observed upon siRNA-mediated knock-down of *HTTAS*, supporting the role of antisense transcription in the regulation of *HTT* expression (Chung *et al.*, 2011). In contrast, Evers *et al.* did not observe any variation in *HTT* mRNA levels after altered transcription of *HTTAS* in fibroblast and post-mortem brain tissue from HD patients (Evers *et al.*, 2015).

Although the regulation of the *HTT* gene remains poorly understood, its expression patterns have been extensively studied (Li *et al.*, 1993; Dixon *et al.*, 2004). Despite being associated with a primarily neurological disorder, *HTT* is expressed ubiquitously, with highest expression observed in the brain and testes (Li *et al.*, 1993; Dixon *et al.*, 2004; Van Raamsdonk *et al.*, 2007). However, *HTT* expression patterns within the brain do not correlate with HD pathophysiology. Expression is highest in the cortex, cerebellum and hippocampus and detectable in the striatum. Regions of reduced *HTT* expression within the striatum are positively correlated with regions of neuronal loss in the HD brain, indicating that reduced *HTT* in these areas is the result of neuronal depletion rather than altered *HTT* expression (Li *et al.*, 1993; Strong *et al.*, 1993). Moreover, it has been reported that in neuronally-differentiated cells *HTT* mRNA is predominantly localized in the cell nucleus (Didiot *et al.*, 2018). This suggests that *HTT* mRNA subcellular localization may play a role in neuronal regulatory mechanisms. Its depletion could result in deleterious effects, explaining the differential vulnerability that causes specific neurodegeneration of neurons in HD.

In summary, understanding the regulatory mechanisms that control *HTT* gene expression at the transcriptional level is critical to understand HTT function in health and disease and for the development of therapeutic strategies aimed at reducing mHTT levels.

1.2.3 Instability of expanded CAG repeats

The CAG repeats of the *HTT* gene are somatically and meiotically unstable, progressively lengthening throughout life and tending to expand between generations (Kennedy *et al.*, 2003). In studies that analysed samples of blood and post-mortem cortex from individuals with HD, greater CAG expansion was associated with an earlier age of disease onset (Swami *et al.*, 2009; GeM-HD, 2019). Furthermore, individuals with pure CAG repeat tracts have an earlier age at onset than individuals with CAA interruptions at the 3′ end of the repeat tract (GeM-HD, 2019). While such interruptions do not change the number of glutamines in the polyQ tract, they result in a reduction in somatic expansion. This suggests that somatic instability of the CAG repeat has a role in pathogenesis. The degree of somatic instability varies between tissues, with expansion particularly prominent in neurons from brain regions that show marked pathology such as the striatum and cortex (Aronin *et al.*, 1995). However, what drives somatic instability remains to be established. All current models involve slip-outs that form when complementary repeat-containing DNA strands anneal out of register (Axford *et al.*, 2013). Such structures could form when DNA is unwound, for example during DNA replication, transcription or repair. It has been demonstrated that transcription modulates somatic CAG instability *in vivo* and it has been suggested that transcriptional dynamics across CAG expansion may be impaired, particularly in the striatum of an HD mouse model (Goula *et al.*, 2012). In detail, it was proposed that high levels of elongating Pol II and high transcriptional activity at the HD locus in HD mice could increase the propensity of repeat instability (Goula *et al.*, 2012). Additionally, depletion of the transcription elongation factor SPT4 in yeast and mammalian cells specifically decreases the transcription of CAG-expanded genes, thereby reducing the amount of toxic gene products (Liu *et al.*, 2012). Using cellular models, it was also shown that transcription across CAG/CTG repeats induces R-loop formation, which promotes CAG/CTG instability (Lin *et al.*, 2010). Furthermore, the formation of unusual DNA structures at repeat expansions can predispose the tract to errors in mismatch repair. Also, variants in several DNA repair genes are associated with somatic instability in both animal models and HD patients (Flower *et al.*, 2019; GeM-HD, 2019) (see Section 1.2.5.7).

1.2.4 The Huntingtin protein

1.2.4.1 The structure of the HTT protein

The *HTT* gene encodes a ~348 kDa highly conserved HTT protein which is ubiquitously expressed and plays important roles in a range of cellular processes including autophagy, vesicular transport, and regulation of transcription (Saudou & Humbert, 2016; Tabrizi *et al.*, 2020). The lack of clear enzymatic activity in HTT, and few conserved functional motifs poses a challenge in understanding the role of HTT protein in different biological processes. The HTT protein is known to interact with a large number of proteins, suggesting its role as a scaffold in coordinating cellular processes (Shirasaki *et al.*, 2012; Saudou & Humbert, 2016; Guo *et al.*, 2018).

An obvious feature of the HTT protein is the polyQ stretch localized at its N-terminus. Despite being 3,144 amino acids long (with a polyQ repeat of 23 glutamines), much of the attention in the literature has focused on the first exon of the HTT protein where the polyQ stretch can be found (Saudou & Humbert, 2016) (Fig. 1.5). In 1994, the polyQ was shown to form a polar zipper structure and, for the first time, it was suggested that its physiological role was to bind transcription factors that also contain a polyQ region (Perutz *et al.*, 1994). More recently, secondary structure analysis of this region of HTT has shown that the polyQ stretch is a key regulator of HTT binding to its partners, and it can adopt several conformations, including α-helix, random coil and extended loop, similar to those found in several transcriptional regulators (Caron *et al.*, 2013; Gemayel *et al.*, 2015; Guo *et al.*, 2018). The polyQ stretch is immediately preceded by an amphipathic α-helix that is important for endoplasmic reticulum (ER) retention, nuclear export, and for modulating HTT aggregation (Atwal *et al.*, 2007). In mammals, a proline-rich domain involved in protein-protein interactions follows the polyQ stretch, consistent with a role for HTT as a protein-interaction hub (Kim *et al.*, 2009) (Fig. 1.5).

Downstream of the polyQ, the so-called HEAT (Huntingtin, elongation factor-3, protein phosphatase 2A and TOR1) repeats are found (Fig. 1.5). The HEAT repeats are sequences of ~40 amino acids thought to be important for mediating inter- and intramolecular interactions within protein complexes (Andrade & Bork, 1995; Yoshimura & Hirano, 2016). A recent cryo-EM structure

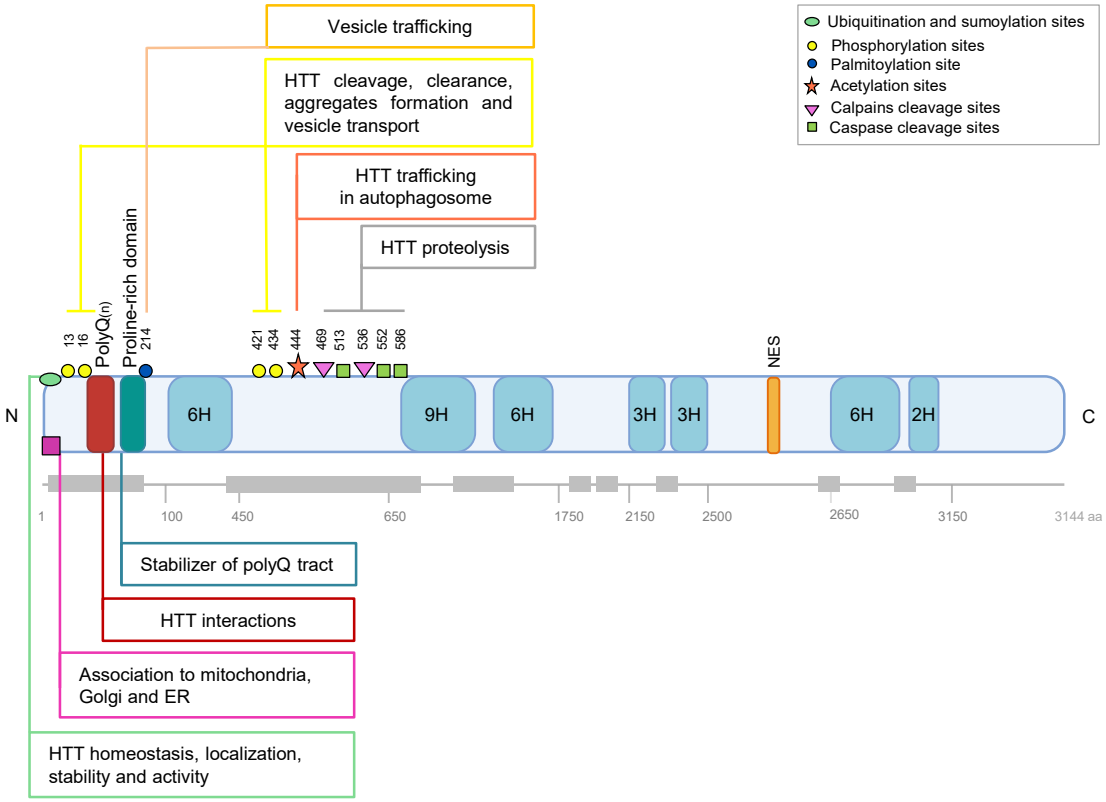


Figure 1.5. Schematic diagram of the huntingtin (HTT) protein.
$PolyQ_{(n)}$ indicates the polyglutamine tract, followed by the proline-rich domain; the blue rectangles indicate the HEAT repeats (H, where H indicates the number of predicted HEAT repeats organized in larger domains). Amino acid positions in black indicate sites of post-translational modifications. The yellow circles indicate the phosphorylation sites, while the pink triangles and green squares indicate the caspase and calpain cleavage sites, respectively. The orange star represents the acetylation site. At the N-terminal, the pink square and the green oval represent the sequence responsible for the association to mitochondria, Golgi and endoplasmic reticulum (ER) and the ubiquitination and/or sumoylation sites, respectively. NES is the nuclear export signal. Numbers in grey, below the protein structure, correspond to the limits of the indicated H domains and grey boxes to the highly disordered regions.

of HTT at 4Å resolution showed that the protein is largely α-helical and contains multiple HEAT repeats arranged in a solenoid fashion (Guo *et al.*, 2018), supporting the hypothesis that HTT acts as a molecular scaffold and confirming prior biophysical, computational, and mass spectrometry studies (Andrade & Bork, 1995; Takano & Gusella, 2002; Tabrizi *et al.*, 2020) (Fig. 1.5). The HEAT repeats are separated by large, disordered regions that are known to be post-translationally modified by phosphorylation, glycosylation, and proteolytic cleavage (Bates *et al.*, 2015) (Fig 1.5).

HTT is a mainly cytoplasmic protein, but nuclear localization of the full-length protein and of the N-terminal fragments has been reported (Hoogeveen *et al.*, 1993; Kegel *et al.*, 2002; Zheng *et al.*, 2013). The movement of HTT across the nuclear membrane is dynamic and seems to be facilitated by the HEAT repeats and regulated by a nuclear import signal in the amino terminus and by a COOH-terminal nuclear export signal (Xia *et al.*, 2003; Atwal *et al.*, 2007) (Fig. 1.5). The 17 amino acids preceding the polyQ region of the HTT protein interact with the translocated promoter region (TPR), a nuclear basket protein, which exports proteins from the nucleus. Removal of amino acids 3-13 causes HTT to accumulate in the nucleus (Cornett *et al.*, 2005).

1.2.4.2 HTT proteolysis and post-translational modifications

HTT is susceptible to proteolysis and HTT fragments arise from proteolysis by caspases (Graham *et al.*, 2006), calpains and other proteases (Bizat *et al.*, 2003; Gafni *et al.*, 2004) (Fig. 1.5). Proteolytic sites are present on both wild-type HTT (wtHTT) and mHTT proteins. The physiological consequences of the cleavage of wtHTT remain elusive. Proteolysis of mHTT leads to the generation of small N-terminal fragments that contain the polyQ stretch and translocate into the nucleus, where they have been shown to be toxic (Saudou *et al.*, 1998; Benn *et al.*, 2005; Graham *et al.*, 2006). Furthermore, wtHTT protein interacts with many proteins involved in post-translational modifications (PTMs) and protein degradation. HTT itself is also subjected to several PTMs such as phosphorylation, acetylation, sumoylation, palmitoylation, and others (Fig 1.5). PTMs are postulated to influence HTT aggregation, function, and subcellular localization (Steffan *et al.*, 2004; Yanai *et al.*, 2006). In addition, PTMs have been found to be significantly altered in the mHTT protein, resulting in phenotypic changes in animal models of HD (Graham *et al.*, 2006; Gu *et al.*, 2009).

1.2.5 The functions of the HTT protein

Despite the mutant HTT protein being toxic and triggering neurodegeneration, growing evidence suggests that the loss of wtHTT function also contributes to HD (Zuccato & Cattaneo, 2009). Indeed, down-regulation of wtHTT levels induces HD-related changes including progressive neurodegeneration and motor dysfunction or worsens these changes in HD models (Dragatsis *et al.*, 2000; Molero *et al.*, 2009). HTT has been shown to play an important role in development and neurogenesis. This is supported by the fact that depletion of HTT results in severe neural developmental defects and embryonic lethality (Duyao *et al.*, 1995; Nasir *et al.*, 1995), whereas conditional knock-out of HTT in mice at embryonic day 14.5 causes progressive neurodegeneration that mimics HD phenotypes (Dragatsis *et al.*, 2000). Furthermore, it was shown that mice expressing less than 50% of normal HTT levels bypass embryonic lethality but display aberrant brain development (Nasir *et al.*, 1995; White *et al.*, 1997; O'Kusky *et al.*, 1999).

The fact that wtHTT interacts with several proteins supports the idea that HTT is implicated in a broad range of cellular pathways from transcriptional regulation, nucleo-cytoplasmic shuttling, mitochondrial dynamics, vesicle trafficking, synaptic function, endocytosis, ciliogenesis and anti-apoptotic activity (Saudou & Humbert, 2016) (Fig. 1.6). Therefore, it is now becoming clear that some HD pathological events may be caused by the loss or modifications of the normal functions of wtHTT. The next Section focuses on some of the best-studied functions of wtHTT at the molecular level, which are also disrupted by the presence of the abnormal polyQ expansion (Fig. 1.6).

1.2.5.1 HTT, intracellular trafficking and cell division

The precise role of HTT as an integrator of vesicular trafficking in neurons is still unclear. However, it has been proposed to interact with the molecular motor machinery, either directly with dynein or indirectly through proteins such as the Huntingtin-associated protein 1 (HAP1), and coordinate the transport of organelles in neurons in both the anterograde and retrograde directions (Colin *et al.*, 2008; Caviston & Holzbaur, 2009). Indeed, silencing of HTT decreases vesicular transport in *D. melanogaster* (Gunawardena *et al.*, 2003), with specific defects in brain-derived neurotrophic factor (BDNF)-containing vesicles and autophagosomes, amongst others (Wong &

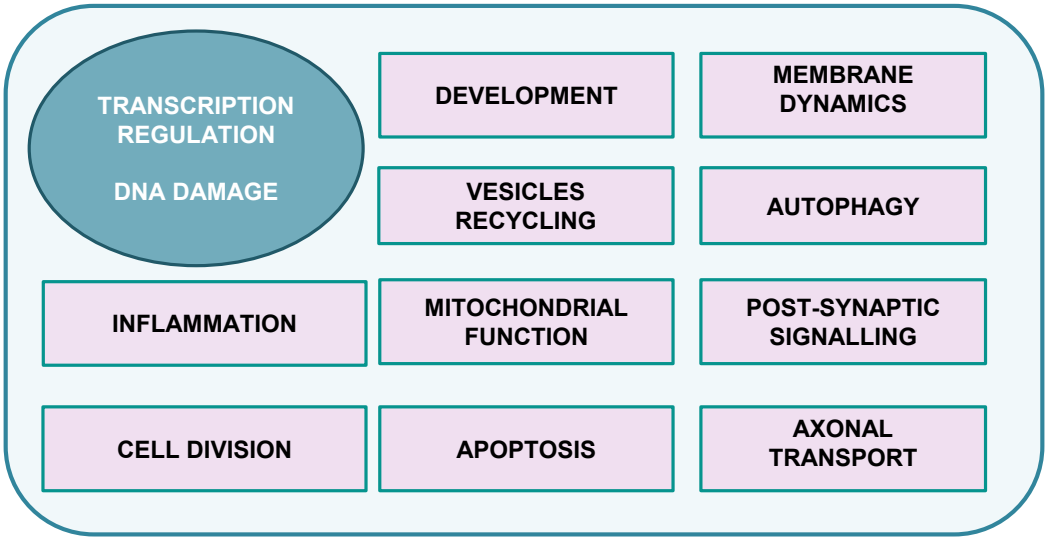


Figure 1.6. Schematic representing known cellular processes which are affected in HD.
The HTT protein is widely expressed and plays essential roles in the early stages of development. It is mainly localized in the cell cytoplasm and involved in various cellular processes, including vesicular trafficking, autophagy, membrane dynamics, axonal transport, post-synaptic signalling, apoptosis, mitochondrial homeostasis and cell division. Although mainly cytoplasmatic, HTT can be found also in the nucleus where it has been shown to be involved in transcription regulation and DNA damage repair. In HD, the presence of an abnormal polyQ expansion disrupts all the biological functions of the HTT protein and causes inflammation. Each of these pathways is described in detail in the main text. The blue square and the green oval represent the cell cytoplasm and nucleus, respectively.

Holzbaur, 2014). In contrast, the expression of a fragment of the fly HTT rescues transport in mammalian neurons silenced for HTT, showing that this function has been conserved during evolution (Zala *et al.*, 2013). Moreover, HTT has been shown to play a role in coordinate cell division (Elias *et al.*, 2014). Particularly during mitosis, HTT locates at the poles of the spindle, together with dynein, and promotes the accumulation of essential proteins involved in mitosis, including the nuclear mitotic apparatus protein (NUMA) and the leucine-glycine-asparagine protein (LGN) (Godin *et al.*, 2010). Hence, loss of HTT function during mitosis leads to spindle mis-orientation.

1.2.5.2 HTT and autophagy

Evidence for the involvement of HTT in autophagy, an important biological process that is essential for the removal of damaged organelles and toxic or aggregated proteins, has emerged following a series of studies highlighting autophagy defects in HD (Martin *et al.*, 2014). Recent studies in *D. melanogaster* and mice have demonstrated that HTT plays a significant role in selective autophagy. In particular, loss of the polyQ stretch, even in wtHTT, appeared to enhance autophagic capacity in neurons (Zheng *et al.*, 2010). HTT was found to function as a scaffold for selective autophagy by promoting cargo recognition and autophagy initiation in both fly and mammalian cells, mediating the binding between p62 and the autophagy-initiating kinase, ULK1 (Rui *et al.*, 2015). Moreover, mHTT contributes to the abnormal activation of the autophagy pathway in various HD models by sequestering and inactivating the mammalian target of rapamycin (mTOR) kinase, leading to an induction of autophagy (Kegel *et al.*, 2000; Ravikumar *et al.*, 2004). However, although the formation of autophagosomes is more abundant in HD, a defect in the loading of the autophagosomes is observed, leading to a reduced capacity of cells to degrade aggregated proteins and organelles (Martin *et al.*, 2014).

1.2.5.3 HTT and mitochondria

Mitochondrial dysregulation has also been associated with HD (Zuccato *et al.*, 2010). mHTT has been shown to affect mitochondria dynamics resulting in defects in ATP production, Ca++ buffering capacity, and apoptosis, which are associated with neurodegeneration in HD (Sawa *et al.*, 1999; Panov *et al.*, 2002). In particular, mHTT has been shown to interact with the outer

mitochondrial membrane, resulting in mitochondrial calcium abnormalities (Panov *et al.*, 2002; Choo *et al.*, 2004). mHTT also interferes with normal organellar axonal transport and can therefore reduce the transport of mitochondria to synapses, as well as ATP production (Yano *et al.*, 2014). Finally, mHTT has been shown to inhibit the expression of mitochondrial genes, including the nuclear coactivator PGC-1α (Cui *et al.*, 2006). A consequence of mitochondrial dysfunction is the aberrant production of reactive oxygen species (ROS), which in turn causes more damage to mitochondria. Interestingly, highly metabolically active cells such as Medium Spiny Neurons (MSNs), severely affected in HD, are extremely susceptible to excessive oxidative damage, age-dependent mitochondrial deficiency and ROS accumulation (Pickrell *et al.*, 2011; Nissanka *et al.*, 2018). Thus, mitochondrial dysfunction can explain the increased vulnerability of MSNs in HD.

1.2.5.4 HTT and apoptosis

One of the earliest assigned roles for wtHTT was its involvement in apoptosis (Saudou *et al.*, 1998). wtHTT protects the central nervous system (CNS) from a range of pro-apoptotic insults by exerting anti-apoptotic functions (Leavitt *et al.*, 2001; Rigamonti *et al.*, 2001). Specifically, wtHTT blocks the formation of the apoptosome complex and consequent activation of caspase-3 and caspase-9, thus preventing apoptosis (Rigamonti *et al.*, 2000). Further studies showed that the anti-apoptotic activity of HTT reflects its ability to bind and sequester pro-apoptotic molecules, including the huntingtin-associated protein 1 (HAP1) (Gervais *et al.*, 2002). Furthermore, transcriptional effects of wtHTT on tumour suppressor protein 53 (p53)-regulated genes, which are involved in apoptosis, may also play an important role (Steffan *et al.*, 2000). Conclusively, loss-of-function of the wtHTT anti-apoptotic role may contribute to neuronal cell death in HD.

1.2.5.5 HTT and transcription regulation

Transcriptional dysregulation is an early event in HD pathology and its presence across multiple HD models suggests that it plays a major role in disease pathogenesis. Expression profiles of several genes were found to be profoundly altered in HD in DNA microarray studies (Sipione *et al.*, 2002), and, more recently, in genome-wide transcriptional profile analysis (GeM-HD, 2019; Al-Dalahmah *et al.*, 2020; Seefelder & Kochanek, 2021). Transcription dysregulation has been also

demonstrated in the post-mortem brains of HD patients and in various tissue specimens of transgenic HD mice (La Spada *et al.*, 2011). Interestingly, these transcriptional changes are particularly pronounced in the striatum, the most affected brain region in HD (Myers *et al.*, 1985; Tabrizi *et al.*, 2013). Specifically, the down-regulation of genes involved in signalling pathways, neurotransmitter receptors, ion channel functions, neuropeptides, synaptic transmission and homeostasis, has been demonstrated in many human and transgenic mouse models of HD. Conversely, up-regulation of genes related to RNA metabolism, protein folding, and stress markers has been observed (Hodges *et al.*, 2006; Langfelder *et al.*, 2016; Ament *et al.*, 2017). The underlying mechanisms responsible for the selective up-regulation and down-regulation of genes are still unknown. In addition to the previously mentioned roles of HTT in autophagy, vesicular transport, neurotransmission, and mitochondrial function, recent evidence suggests that HTT also plays an important role in transcription. The loss of a transcriptional function of the HTT protein may be one of the potential mechanisms that could account for this altered gene expression. HTT nuclear localization and the presence of a polyQ tract, which is found in many transcription factors and can act as a transcriptional regulatory domain, suggest that HTT has the potential to regulate transcription (Benn *et al.*, 2008). In line with this, HTT protein was shown to directly associate with Pol II (Gao *et al.*, 2019).

HTT binds to transcription factors

HTT has been shown to interact with other proteins belonging to the transcriptional machinery, strengthening the notion that it functions as a gene expression regulator in mammalian cells. Several transcription factors are known to interact with HTT protein, including the tumour suppressor protein 53 (p53), the cAMP response element (CREB)-binding protein (CBP) (Steffan *et al.*, 2000), TATA-box binding protein associated factor 4 (TAF4) (Gao *et al.*, 2019) and the nuclear factor kappa-light-chain-enhancer of activated B cells (NF-kB) (Takano & Gusella, 2002). HTT also interacts with transcriptional activators and repressors such as the nuclear co-repressor (NCOR), the repressor element-1 transcription factor/neuron restrictive silencer factor (REST/NRSF) and the transcriptional corepressor C-terminal-binding protein (CtBP) (Kegel *et al.*, 2002; Yohrling *et al.*, 2003). Through these interactions, HTT can influence many cell functions at the level of

transcriptional networks. For instance, HTT can potentiate transcription factors and inhibit repressors (and vice-versa), thereby promoting and repressing gene transcription with a wide range of cellular outcomes. The best-known target of wtHTT-mediated transcriptional regulation is the brain-derived neurotrophic factor (*BDNF*) gene (Zuccato *et al.*, 2003). Indeed, it has been demonstrated that wtHTT can regulate the activity of genes that contain neuron-restrictive silencer elements (NRSEs) by modulating the cytoplasmic recruitment of NRSE-binding transcription factors to the nucleus. Particularly, wtHTT has been shown to promote *BDNF* transcription in the cortical neurons by sequestration of the REST/NRSF available in the cytoplasm, thereby preventing its negative activity on *BDNF* transcription (Zuccato *et al.*, 2003). In addition, the interaction of HTT with transcriptional regulatory proteins can be altered by the presence of the polyglutamine expansion (Benn *et al.*, 2008). For example, mHTT interacts with the transcriptional factor specificity protein 1 (Sp1) with higher affinity than wtHTT. This causes a polyQ length-dependent disruption of Sp1 function via inhibiting its association with TAFII130/TAF4 (Shimohata *et al.*, 2000; Li *et al.*, 2002). Consequently, a decrease in the expression of a number of Sp1-regulated genes has been shown in HD (Luthi-Carter *et al.*, 2000). Furthermore, mHTT can repress transcription by sequestration of several transcription regulators, including CBP, within the mHTT aggregates (Nucifora *et al.*, 2001).

HTT binds to nuclear receptors

HTT binds nuclear receptors, including the cholesterol metabolism regulators liver X receptors (LXR) α and β, the peroxisome proliferator-activated receptor-γ (PPARγ), the vitamin D receptor (VDR), the thyroid hormone receptor-α1 (TRα1) (Futter *et al.*, 2009), and the peroxisome proliferator-activating receptor-y coactivator-α1 (PGC-1α) (Cui *et al.*, 2006). PGC-1α is a transcriptional coactivator which coordinates the transcriptional profiles responsible for mitochondrial biogenesis. Therefore, abnormal interaction between mHTT and PGC-1α may contribute to mitochondrial dysfunction observed in HD (Section 1.2.5.3). LXRs are critical for brain function and their knock-out leads to cholesterol dysregulation and subsequent neurodegeneration (Wang *et al.*, 2002). *In vitro* studies have shown that wtHTT acts as a co-factor for LXR-mediated gene transcription and polyQ expansion reduces this activity (Futter *et al.*, 2009).

1.2.5.6 HTT, chromatin and epigenetics

It has been demonstrated that nuclear fragments of mHTT are sufficient to cause transcriptional dysregulation *in vivo* (Benn *et al.,* 2005), suggesting that mHTT can interfere with transcription at the promoter level by directly binding to DNA. This direct interaction can alter the conformation of DNA or sterically obstruct the access of other transcription factors, therefore potentially leading to transcriptional dysregulation in HD (Benn *et al.*, 2008). Although the direct relationship between the HTT protein and chromatin is poorly understood, a recent study showed that HTT occupies specific genomic locations in the mouse genome and HTT binding is correlated with epigenomic, and transcriptional changes observed in HD (Pearl *et al.*, 2020). Thus, it is possible that altered huntingtin-chromatin interactions directly contribute to the transcriptional and epigenetic dysregulation seen in HD.

Epigenetic mis-regulation, both in the regions surrounding repeat expansions and at the whole genome scale, has become a prevailing feature contributing to the altered expression, stability, and mutational repair rates of various tandem repeats, resulting in neuronal dysfunction (see Section 1.1.4.2) (Evans-Galea *et al.*, 2013). In the case of HD, progressive changes in gene expression have been observed in premanifest HD stages, and multiple epigenetic modifications have been associated with HD (Alcala-Vida *et al.*, 2021; Hyeon *et al.*, 2021). The full-length recombinant HTT protein has been shown to promote the histone H3K27 tri-methylase activity of the multi-subunit polycomb repressive complex 2 (PRC2) *in vitro*, and this stimulatory effect is enhanced by expanded polyQ repeats (Seong *et al.*, 2010). Moreover, expression of mHTT in embryonic stem cells causes a genome-wide change in the H3K27me3 pattern (Biagioli *et al.*, 2015) and induced PCR2-deficiency in MSNs affects transcription, leading to progressive neurodegeneration (von Schimmelmann *et al.*, 2016). In support of this, recent ChIP-seq experiments revealed genotype-specific co-localization of HTT with H3K27me3 marks in striatal tissue from a knock-in HD mouse model (Pearl *et al.*, 2020).

Increased levels of the repressive mark H3K9 trimethylation (H3K9me3) have been detected in HD patients and HD mouse models and up-regulation of these marks was accompanied by the up-regulation of its specific methyltransferase known as SET domain bifurcated histone lysine methyltransferase 1 (SETDB1) (Ryu *et al.*, 2006). Recently, wtHTT has been also found to bind the

activating transcription factor 7-interacting protein 1 (ATF7IP), a regulator of the histone H3 methyltransferase SETB1 (Irmak *et al.*, 2018). Specifically, HTT was shown to inhibit the interaction of ATF7IP-SETDB1 complex together with other heterochromatin regulators and transcriptional repressors, maintaining low levels of H3K9me3 in human embryonic stem cells (hESCs) (Irmak *et al.*, 2018). Hence, loss of HTT and/or mHTT can promote aberrant enrichment of H3K9me3 marks at distinct genes, including transcriptional regulators of neuronal differentiation.

In addition to changed histone marks, extensive changes in DNA methylation signatures have been reported in HD patients and transgenic mice, including changes in 5-methylcytosine (5mC) and 5-hydroxymethylcytosine (5-hmC) patterns (Ng *et al.*, 2013). Recently, a large study characterized the levels of DNA methylation in six different tissues from *HTT* knock-in mice, transgenic *HTT* sheep and humans (Lu *et al.*, 2020). Epigenome-wide association studies (EWAS) revealed that the most significant HD-associated methylation sites were found in the *HTT/Htt* locus and in close proximity to the repeat expansion mutation itself, suggesting a role for these methylation sites as disease modifiers (Lu *et al.*, 2020).

1.2.5.7 HTT and DNA damage

The length of uninterrupted CAG repeats has been shown to be the main determinant of the course of HD and accounts for ~60 % of the variation in age at onset (GeM-HD, 2019). However, the remaining variability is also heritable and results from differences elsewhere in the genome (Wexler *et al.*, 2004).

Somatic instability is thought to play a role in HD pathogenesis, particularly in the striatum (Swami *et al.*, 2009). DNA damage response (DDR) and mismatch repair (MMR) pathways may directly influence the repeat sequences themselves, affecting the somatic instability of the CAG tract and leading to incremental expansions in CAG length (Axford *et al.*, 2013; Tabrizi *et al.*, 2020). Indeed, recent genome-wide association studies (GWAS) conducted in HD have identified genetic loci associated with disease onset, and highlighted many genes that are involved in DDR and MMR pathways as disease severity and symptom onset modifiers (GeM-HD, 2015; Moss *et al.*, 2017; GeM-HD, 2019). The strongest signal came from genetic variations in the DNA repair gene *FAN1*,

a nuclease of the Fanconi anemia (FA) pathway that cuts DNA during the repair of crosslinks between DNA strands (GeM-HD, 2019). Increased expression of FAN1 was significantly associated with delayed age at onset and slower progression of HD, suggesting a protective role of FAN1 in the context of expanded *HTT* CAG repeats (Goold *et al.*, 2019). Indeed, FAN1 overexpression reduces CAG repeat expansion in different HD cellular models, whereas FAN1 knockdown increases CAG repeat expansion (Goold *et al.*, 2019). However, how FAN1 protects against CAG repeat instability remains unclear. FAN1 was shown to stabilize CAG repeats by binding the DNA mismatch repair protein MutL protein homolog 1 (MLH1), and restrict its recruitment by MutS Homolog 3 (MSH3), thus inhibiting the assembly of a functional MMR complex that would otherwise promote CAG repeat expansion. Furthermore, FAN1 was shown to promote accurate repair of loopouts that form at CAG repeats via its nuclease activity (Goold *et al.*, 2021; Porro *et al.*, 2021; McAllister *et al.*, 2022).

GWAS highlighted other prominent modifications in members of the MMR pathway, including *MLH1*, *MSH3, PMS1, PMS2* and *LIG1*. However, the mechanisms underlying the association of these modifiers with HD disease remain unclear. Increased expression of the mismatch repair gene *MSH3* was associated with faster HD progression, implying a detrimental role for MSH3 in HD pathogenesis (Flower *et al.*, 2019).

Alternatively, it has been suggested that the HTT protein directly participates in the DDR as a component of DNA damage repair complexes (Massey & Jones, 2018; Gao *et al.*, 2019). Specifically, mHTT aggregates have been shown to interfere with DDR mechanisms, resulting in chronic activation of the ataxia telangiectasia-mutated (ATM) and p53 signalling pathways (Lu *et al.*, 2014). The N-terminal methionine of HTT protein has been shown to act as a direct sensor of oxidative stress, which, when oxidised, triggers a cytoplasm-to-nucleus translocation of HTT (DiGiovanni *et al.*, 2016). In the nucleus, HTT is recruited to sites of DNA damage where it may act as a scaffold in the ATM oxidative DNA damage response complex (Lu *et al.*, 2014; Maiuri *et al.*, 2017). Additionally, HTT protein has been shown to be part of a transcription-coupled repair (TCR) complex (Gao *et al.*, 2019). The HTT-assembled TCR complex has been proposed to be involved in repairing DNA lesions during active transcription, while mHTT was shown to impair the complex

functionality, resulting in impaired DNA damage repair and neurotoxicity. Persistent DNA lesions within the actively-transcribed genes or promoter regions may contribute to the transcriptional dysregulation observed in HD (Section 1.2.5.5) (Gao *et al.*, 2019).

Recently, it was discovered that mHTT also causes deficient recruitment of the p53-binding protein 1 (53BP1), a critical component of DNA double-strand break (DSB) signalling and repair, to the nucleus following Topoisomerase I (Top-1)-induced DNA breaks (Palminha *et al.*, 2022). Furthermore, HD fibroblasts repair Top-1-induced DNA damage at a slower rate, indicating that HD cells exhibit defective repair of chromosomal DNA breaks, contributing to premature neuronal death (Palminha *et al.*, 2022).

1.2.6 RNA toxicity induced by expanded *HTT* mRNA

RNA-induced toxicity has not been extensively addressed in HD, where mHTT protein has been traditionally thought to play a major role in the pathogenesis. However, recently, a pathogenic role of expanded repeat-containing RNA has been proposed. Indeed, nuclear RNA clusters are a common HD pathological feature and expanded *HTT* mRNAs have been shown to induce toxicity and interfere with cell viability in a neuronal cell model (Banez-Coronel *et al.*, 2012). Moreover, recent GWAS studies showed that the length of uninterrupted CAG repeats, but not the poly Q tract, defines the age of onset, highlighting an under-appreciated role of mutant DNA/RNA in HD (GeM-HD, 2015; GeM-HD, 2019).

Multiple pieces of evidence indicate that CAG repeats can form extensive RNA secondary structures. Mutant CAG transcripts have been shown to fold into secondary hairpin structures *in vitro* and form nuclear foci *in vivo* (de Mezer *et al.*, 2011). In addition, it has been reported that mutant *HTT* mRNA is retained in the nucleus, and forms large clusters driven by expanded repeats in striatal and cortical neurons of HD mouse models and patient brains, supporting the role of toxic RNA in HD pathogenesis (Ly *et al.*, 2022). The co-localization of expanded transcripts with the dsRNA-binding protein MBNL1 suggests that *HTT* mRNA may adopt double-strand (ds)-like RNA conformations *in vivo* (de Mezer *et al.*, 2011; Sun *et al.*, 2015). For many expanded repeat disease loci, transcription is found on both strands, raising the possibility that ds expanded repeat RNA could

be a culprit in these diseases (Malik *et al.*, 2021). Studies in the *D. melanogaster* model reported that neuronal toxicity of the repeat-containing dsRNA, generated by bidirectional transcription, is dependent upon the activity of Dicer (Lawlor *et al.*, 2011). Furthermore, expanded CAG.CUG repeat dsRNAs have been shown to cause inflammation-mediated pathology when expressed in *D. melanogaster* (van Eyk *et al.*, 2019).

The ability of expanded CAG-RNA to induce RNA interference and trigger downstream silencing effects of transcripts containing CTG repeats in the coding region or in the 3'UTR has been also described (Krol *et al.*, 2007; Banez-Coronel *et al.*, 2012). However, experiments performed in SHSY-5Y neuroblastoma cells detected only a ~10% decrease in the expression of these transcripts following transfection with *HTT*-RNA containing expanded CAG repeats (Banez-Coronel *et al.*, 2012). Another recent study claimed that small interfering RNAs based on huntingtin trinucleotide repeats are highly toxic to cancer cells *in vitro* and *in vivo* (Murmann *et al.*, 2018). This may explain the low incidence of most cancer types in individuals with HD or other polyglutamine disorders (Sorensen *et al.*, 1999; Ji *et al.*, 2012). Moreover, one of the major mechanisms of RNA-induced toxicity in HD is dysregulated splicing (Schilling *et al.*, 2019). Indeed, it has been shown that mutant *HTT* RNA sequesters proteins that belong to the spliceosome pathway in a CAG repeat length-dependent manner, inducing mis-splicing in an HD cellular model (Schilling *et al.*, 2019).

The nucleolus is a major site for expanded CAG RNA toxicity in polyQ diseases. *HTT*-RNA containing expanded CAG repeats is able to directly interact with Nucleolin (NLC), a nucleolar protein that regulates ribosomal RNA (rRNA) transcription (Tsoi *et al.*, 2012). Such RNA-protein interaction causes DNA hypermethylation in the rRNA promoter and results in a reduction in rRNA transcription. Down-regulation of rRNA transcription induces nucleolar stress followed by apoptosis (Tsoi *et al.*, 2012). Therefore, a contribution of disrupted rDNA transcription to the aforementioned genome-wide transcriptional dysregulation cannot be excluded.

In conclusion, it is important to note that while each pathogenic mechanism related to toxic protein and RNA in HD has been discussed as a separate process, disruptions in one pathway likely contribute, at least in part, to HD pathogenesis by worsening the dysfunction of other critical events, ultimately resulting in the complex HD phenotype.

1.2.7 Immune dysfunction in HD

Inflammatory changes in the CNS, commonly referred to as neuroinflammation, are crucial components in the pathogenesis of several neurodegenerative diseases (Bjorkqvist *et al.*, 2008). Yet the exact underlying molecular mechanisms and the impact of the innate and adaptive immune systems in HD pathology are still not fully understood. Over the past few years, efforts have been made to understand whether mHTT expression can trigger immune activation in the brain and, hence, play a role in HD pathophysiology (Rocha *et al.*, 2016).

Microglia are the resident immune cells of the CNS and constantly monitor the microenvironment of the neurons and glial cells, being able to switch to different activation states through various signalling pathways (Borst *et al.*, 2021). The presence of mHTT has been shown to induce the activation of microglia and astrocytes leading to the production and release of inflammatory cytokines, such as Interleukin (IL-) 6, IL-8, tumor necrosis factor-α (TNFα), chemokines monocyte chemoattractant protein-1 (MCP-1)/CCL2, RANTES/CCL5 and matrix metalloproteinase-9 (MMP-9) (Sapp *et al.*, 2001; Bradford *et al.*, 2009; Silvestroni *et al.*, 2009). Indeed, mHTT has been shown to positively regulate the NF-kB signalling pathway via direct interaction with the inhibitor of nuclear factor-κB (IκB) kinase (IKK), inducing the release of IL-6 and IL-8 (Khoshnan *et al.*, 2004). Consequently, higher levels of IL-6, IL-8 and TNFα were found in the striatum of HD patients, whereas increased expression of IL-6, IL-8 and MMP-9 were observed in the cortex and cerebellum of HD patients (Silvestroni *et al.*, 2009). In addition, microglia generated from HD human pluripotent stem cells have been shown to be hyper-reactive to stimulation in the production of pro-inflammatory cytokines, such as IL-6 and TNFα (O'Regan *et al.*, 2021). Furthermore, microglia have been shown to release elevated levels of reactive oxygen species (ROS) that have neurotoxic potential (O'Regan *et al.*, 2021). Whether this inflammatory response is triggered by degenerating neurons or results from mHTT expression in microglial cells and leads to the activation of the immune response through various pathways, is still a matter of debate (Rocha *et al.*, 2021); however, the finding of microglial activation in the cortex and even in the cerebellum, a region generally spared by HD pathology, favours the latter hypothesis.

Immune-related pathology in HD is not merely restricted to CNS and a wide range of peripheral inflammatory changes have been also reported. Investigation of blood samples from HD patients revealed increased levels of inflammatory cytokines, particularly IL-6, IL-8, TNFα and different chemokines (Bjorkqvist *et al.*, 2008). Mitochondrial dysfunction observed in HD may also be associated with inflammatory responses. As mentioned in Section 1.2.5.3, mitochondria impairment leads to the over-production of reactive oxidate species (ROS) which can play important functions in regulating inflammatory responses. Moreover, it has been demonstrated that mHTT induces the release of mitochondrial RNAs which are recognized by the dsRNA sensor protein kinase R (PKR) and trigger the activation of innate immune signalling in HD neurons (Lee *et al.*, 2020). Finally, the cyclic GMP-AMP synthase (cGAS) protein, a key mediator of inflammatory responses, was shown to be upregulated in HD cells and its increase was causally linked to inflammatory and autophagy responses that may contribute to HD pathology (Sharma *et al.*, 2020). Specifically, the cGAS-stimulator of interferon genes (STING) pathway activates the TANK binding kinase 1 (TBK1). TBK1 phosphorylates the interferon regulatory factor 3 and 7 (IRF3/IRF7), transcription factors which dimerize and move into the nucleus, where they induce transcription of genes encoding various cytokines, including CCL5 and Cxcl10. Interestingly, CCL5 and Cxcl10 have been reported to be upregulated in several HD models and in HD patients (Dong & Cong, 2018; Sharma *et al.*, 2020).

1.2.8 HD Therapeutics

Currently, there are no available therapeutics capable of stopping or slowing the progression of HD (Novak & Tabrizi, 2010; Tabrizi *et al.*, 2022). The pathological events that occur downstream from mHTT are intricate, and it is challenging to target individual pathways pharmacologically. Moreover, such targeting may be insufficient to modify the course of the disease. Indeed, so far, several HD-specific treatments focused on the relief of symptoms have failed efficacy trials. Nevertheless, there are several medications used to manage the symptoms of HD, including antipsychotics, antidepressants, and mood stabilizers. Tetrabenzine, the first drug specifically licensed for the treatment of HD-related chorea (Frank, 2009), is typically used in combination with

drugs that are useful in managing non-motor symptoms, such as selective serotonin re-uptake inhibitors for depression (McLauchlan *et al.*, 2022).

Due to the difficulties in targeting individual downstream pathways, novel therapeutic strategies are now focusing on specific pathogenic mechanisms, such as targeting the production of mutant *HTT*, either at the RNA or DNA level, with the aim of ameliorating the relevant cellular deficit. This can be achieved using antisense oligonucleotides (ASOs) and RNAi, which bind and degrade *HTT* mRNA transcripts (Wild & Tabrizi, 2017). The first ASO administered to HD patients was tominersen. Tominersen targets exon 36 of human *HTT* mRNA and degrades both mutant and wild-type transcripts by RNAse H (Tabrizi *et al.*, 2022). Initial clinical trial showed a promising reduction in mHTT in patients cerebrospinal fluids (CSF). However, tominersen dosing was halted in March 2021 due to adverse events observed in treated patients. *In vivo* studies of recombinant adeno-associated viral (AAV)-mediated RNAi therapies have been shown to be efficient in lowering huntingtin in several HD models, including transgenic mice, mini-pigs and non-human primates (Grondin *et al.*, 2012; Evers *et al.*, 2018). Additionally, small molecules, like PTC518 and Branapalm, orally available splicing modulators, have been developed to reduce *HTT* expression through mRNA splicing modulation (Krach *et al.*, 2022). Moreover, agents that directly interact with *HTT* DNA, such as zinc finger transcriptional repressors and CRISPR-associated system (Cas) genome editing constructs (CRISPR-Cas9) have also shown potential (Malankhanova *et al.*, 2017). However, a key challenge in targeting repeat-expanded alleles is to distinguish between the mutant and healthy allele, as wtHTT has important cellular roles. Therefore, the ideal therapeutic strategy would selectively lower mHTT while preserving the wtHTT. While most HD patients are heterozygous for the HD gene mutation, targeting only mHTT through allele-selective *HTT* lowering therapies is challenging. One approach involves using polymorphisms, such as expanded CAG repeats and single-nucleotide polymorphisms (SNPs), to develop SNP-allele-specific therapies. Several SNPs have been identified in the *HTT* gene and their variants have been utilized in developing such therapies (van Bilsen *et al.*, 2008; Pfister *et al.*, 2009; Ostergaard *et al.*, 2013; Conroy *et al.*, 2022).

Somatic instability has been demonstrated to act as a driver of disease age-of-onset, progression and severity. Therefore, compounds that can modulate instability have been exploited therapeutically. For example, Nakamori *et al*. showed that the small molecule naphthyridine-azaquinolone (NA) specifically binds expanded CAG-repeat tracts, thereby preventing further expansion and instead encouraging repeat contraction, in cells and mice (Nakamori *et al.*, 2020). Furthermore, DNA repair pathways identified as disease modifiers from GWAS, as discussed in Section 1.2.5.7, can influence somatic instability of the CAG tract either driving or preventing repeat expansion (GeM-HD, 2015; GeM-HD, 2019). This knowledge is currently being translated into the development of new therapeutics. For example, it has been shown that loss of the MMR component MSH3 protein has no phenotype and does not affect the lifespan or cause cancer in mouse models (Iyer & Pluciennik, 2021). Therefore, inhibition of MSH3, by knockdown with MSH3 ASO or by small molecule MSH3 inhibitors is expected to slow repeat expansion and be well tolerated. Conversely, the use of promotors of FAN1 activity, which has been shown to have a protective role in the context of CAG expanded repeats, could be exploited for therapeutics.

Small-inhibitor molecules, targeting the DNA damage pathway, have been also identified as potential therapeutics for HD (Toledo-Sherman *et al.*, 2019). These molecules inhibit the ATM protein, a kinase that is hyperactive in HD. However, developing these small molecules is challenging due to the risk of off-target binding and non-specific effects. Therefore, there is ongoing research to modify these compounds and improve their selectivity (Toledo-Sherman *et al.*, 2019; Van de Poel *et al.*, 2021). Finally, the transplantation of healthy cells into the brain to replace damaged or dead cells, known as stem cell therapy, has shown promising results in HD preclinical studies (Islam *et al.*, 2021; Park *et al.*, 2022). However, before stem cell therapy can be used as a viable treatment for HD, more research is needed to optimize the cell types used, the delivery methods, and the dosage required. Additionally, the safety and long-term effects of stem cell therapy must be thoroughly evaluated in clinical trials before it can be approved for widespread use.

In order to develop a novel alternative therapeutic approach, it is crucial to establish both the effectiveness and safety of the treatment in the appropriate preclinical models of HD. Regrettably, some promising therapies, including mHTT-lowering strategies and the splicing modulator

Branapalman, have been discontinued due to adverse reactions in patients. As a result, the number of HD drugs currently undergoing advanced testing is limited. Consequently, pursuing targets that are specific to HD and address the underlying cause of the disease is likely to become the new focus of research in this field.

1.3 R-loops in health and disease

As described in section 1.1.4.1, non-canonical RNA/DNA structures, such as R-loops, have been implicated in several repeat expanded disorders (Groh & Gromak, 2014). However, the role of R-loops in HD has not been explored. The following section provides a general review of R-loops in health and disease giving a background for my study on HD.

1.3.1 Structure and detection of R-loops

R-loops were initially characterized *in vitro* by electron microscopy (EM) over 40 years ago (Thomas *et al.*, 1976). However, it was not until 20 years later that it was suggested that endogenous R-loops serve a biological purpose in living organisms (Drolet *et al.*, 1995). R-loops are three-stranded structures that are formed co-transcriptionally when nascent RNA re-anneals to the template DNA. This results in the formation of an RNA/DNA hybrid and a displaced non-template ssDNA (Garcia-Muse & Aguilera, 2019) (Fig 1.7). In mammals, R-loops are abundant and occupy a significant part of the genome, ranging from 200 to 500 base pairs in length (Stork *et al.*, 2016; Kouzine *et al.*, 2017). They are transient and reversible structures that naturally arise during transcription, occurring in various regions of eukaryotic genomes, including regions transcribed by Pol I, II and III (Tran *et al.*, 2017). Typically, R-loops form over unmethylated CpG island promoters, but their formation is observed also at the 3' end of a number of human genes (Ginno *et al.*, 2013). Furthermore, R-loop formation may be a critical step in some gene transcriptional termination (Skourti-Stathaki *et al.*, 2011), suggesting that R-loops may possess more general physiological roles. Indeed, multiple studies across various organisms have demonstrated that R-loops play a critical role in several biological processes, including immunoglobulin class-switch recombination, gene expression, epigenetics, DNA replication and DNA repair (Niehrs & Luke, 2020) (Fig. 1.7). Mis-regulation of R-loops can result in increased DNA damage and genome instability, and have been recently associated with neurodegenerative diseases and cancer (Groh & Gromak, 2014; Santos-Pereira & Aguilera, 2015; Crossley *et al.*, 2019) (Fig. 1.7). Therefore, an understanding of the molecular mechanisms underlying R-loop regulation is crucial to uncovering their cellular functions.

One of the major challenges in R-loop biology is detecting these structures *in vivo*. Currently, there are two main methods used to study R-loop distribution. The first method involves the use of the S9.6 antibody, which recognizes RNA/DNA hybrids with a 0.6 mM affinity and no sequence specificity (Boguslawski *et al.*, 1986; Phillips *et al.*, 2013; Bou-Nader *et al.*, 2022). This allows for RNA/DNA immuno-precipitation followed by sequencing (DRIP-seq) to map R-loop distribution and abundance (Ginno *et al.*, 2012; Sanz & Chédin, 2019; Crossley *et al.*, 2020). The second method involves the expression of a tagged, catalytically inactive form of RNAse H1 (dRNAse H1), which recognizes but does not process RNA/DNA hybrids. *In vivo* dRNAse H1 binding sites are mapped using chromatin immunoprecipitation (R-ChIP) and CUT&RUN-based approaches (Chen *et al.*, 2017; Yan *et al.*, 2019). Indirectly, R-loops can also be mapped at the single-molecule level using the single-molecule R-loop footprinting (SMRF-seq) strategy, which utilizes the chemical reactivity of the displaced single-stranded DNA to non-denaturing sodium bisulfite, followed by single-molecule long-read sequencing (Malig *et al.*, 2020). R-loops can also be visualized by electron EM (Kaback *et al.*, 1979) that can measure R-loop frequency and size at the single-molecule level (Stoy *et al.*, 2023), or more commonly by immuno-fluorescence (IF) microscopy. However, recent evidence suggests that S9.6-based imaging approaches may be limited by significant but lower affinity of S9.6 for dsRNA species (Chedin *et al.*, 2021; Smolka *et al.*, 2021). Future research to visualize these structures in living cells and to study their kinetics by single molecule approaches will provide valuable insight into their role in health and disease.

1.3.2 Physiological roles of R-loops

R-loops were initially considered to be by-products of transcription, but their potential role in gene regulation was only recently discovered. In 1995, R-loop formation was shown to occur in *E.coli* as a result of the transcription process (Drolet *et al.*, 1995). The presence of R-loops within gene promoters suggested that they may play an important role in gene regulation (Skourti-Stathaki & Proudfoot, 2014) (Fig. 1.7). Indeed, at promoters, R-loops have been linked to maintaining the unmethylated state of CpG islands, ultimately facilitating transcription (Ginno *et al.*, 2012; Grunseich *et al.*, 2018). In addition, R-loop-dependent recruitment of GADD45A and active de-

methylation by ten-eleven translocation (TET) enzymes was proposed as a mechanism for loss of CpG island methylation (Arab *et al.*, 2019). R-loops at promoters can also guide the binding of chromatin remodelers and transcriptional regulators (Chen *et al.*, 2015; Cloutier *et al.*, 2016). Conversely, R-loops at the 3' end of some mammalian genes are proposed to facilitate Pol II pausing before efficient termination (Skourti-Stathaki *et al.*, 2011). The helicase senataxin (SETX) is involved in the resolution of these R-loops, allowing access for the terminator factors XRN2, leading to efficient termination (Skourti-Stathaki *et al.*, 2011; Morales *et al.*, 2016). R-loops are prevalent in highly transcribed genes and also accumulate in regions of DNA repeats such as centromeres, telomeres, and retrotransposons (Graf *et al.*, 2017).

Apart from their role in transcription and gene regulation, R-loops also facilitate immunoglobulin (Ig) class-switch recombination, DNA replication of mitochondrial DNA, and CRISPR-Cas9 activity (Yu *et al.*, 2003; Horvath & Barrangou, 2010) (Fig. 1.7). The first example of endogenous regulatory R-loops in eukaryotes came from studies on Ig class-switch recombination (Yu *et al.*, 2003). R-loops that form within the G-rich switch regions have been shown to promote recombination-based deletions which contribute to antibody isotype diversity in activated B cells (Yu *et al.*, 2003). During mitochondrial replication, R-loops are cleaved by the mitochondrial RNase MRP to produce RNA primers for DNA polymerase (Lee & Clayton, 1997). Furthermore, R-loops are associated with the popular CRISPR-Cas9 system. CRISPR-Cas9 system provides adaptive immunity for bacteria, recognizing and destroying foreign nucleic acids of viral origin (Horvath & Barrangou, 2010). The viral DNA, integrated into a CRISPR array in the bacterial genome, is transcribed and associates with the Cas9 nuclease. The Cas9-RNA complex invades the viral DNA to form an R-loop with the matching sequence while displacing the opposing non-target strand. This generates a DNA break, which eventually results in the destruction of the viral DNA. This system is now widely used for genome editing applications, using guide RNAs to target Cas9 to a specific sequence in the genome, a process which involves R-loop formation *in trans* (Pacesa *et al.*, 2022). Additionally, R-loops are involved in maintaining the integrity of the genome by promoting DNA double-strand break (DBS) repair, repair of short telomeres, and preserving DNA topology (Aguilera & García-Muse, 2012) (Fig. 1.7).

1.3.3 R-loop interacting proteins

Despite the multiple fundamental functions of R-loops in cellular processes and human disease, our current knowledge about R-loop regulation and the factors involved in these processes is still lagging behind. Recently, several protocols have been optimised to detect R-loop-interacting factors. To identify R-loop-interacting proteins in human cells in a comprehensive way, we recently employed an unbiased affinity purification approach with the S9.6 antibody, followed by Western Blot (WB) or mass spectrometry (MS) (Cristini *et al.*, 2018; Beghè & Gromak, 2022). This protocol has been validated to allow the purification of RNA/DNA hybrid-binding factors from the chromatin fraction of non-cross-linked cells. MS analysis identified an "R-loop interactome" dataset comprising 469 proteins which has been successfully used to reveal novel R-loop-binding factors and their R-loop-associated functions (Cristini *et al.*, 2018; Beghè & Gromak, 2022). Wang *et al*. analysed the interaction of proteins with synthetic over-expressed RNA/DNA hybrids from *BAMBI* and *DPP9* genes in human B cells, resulting in the detection of ~800 R-loop-binding factors (Wang *et al.*, 2018). However, this protocol, as well as the one developed by our lab, may be limited in the detection of transient/weak interactions with RNA/DNA hybrids. Therefore, the identification of transient interactions has been facilitated in recent years by several proximity-based labelling approaches (Beck *et al.*, 2014). In these technologies, the common theme is the transfer of a biotin label from the target to proximate proteins (and RNA in the case of IPL and APEX), that can be purified by streptavidin affinity and identified by MS. RNA-DNA Proximity Proteomics (RDProx) was developed to map the R-loop proximal proteome of human cells using MS (Mosler *et al.*, 2021). This method relies on the use of the fusion protein of the hybrid-binding domain (HBD) of RNase H1 and an engineered variant of ascorbate peroxidase (APEX2) (Lam *et al.*, 2015). In addition, Yan *et al.* (2022) used TurboID, a biotin ligase which catalyses proximity labelling (Branon *et al.*, 2018), to uncover the RNase H proximal proteome.

Using these methods, over the last few years, many proteins associated with R-loops have been identified, reflecting the diversity of R-loop processes. These proteins can regulate cellular R-loop levels in cells, either directly or indirectly, mostly by preventing the nascent RNA from hybridizing with DNA, thus reducing excessive R-loop accumulation. Among them, there are

proteins required for efficient transcriptional elongation, termination, polyadenylation, RNA splicing, packaging, and export (Huertas & Aguilera, 2003; Li & Manley, 2005; Skourti-Stathaki *et al.*, 2011; Santos-Pereira *et al.*, 2013; Bhatia *et al.*, 2014). Cells have developed several mechanisms to prevent R-loop accumulation and avoid the potential threat to transcription and genome integrity. For example, R-loop levels can be directly modulated by dedicated enzymes, such as members of the RNase H family that specifically degrade the RNA within the RNA/DNA hybrids (Cerritelli & Crouch, 2009). Furthermore, RNA/DNA helicases, including senataxin (SETX) (Skourti-Stathaki *et al.*, 2011), Aquarius (AQR) (Sollier *et al.*, 2014), and DHX9 (Cristini *et al.*, 2018), can unwind RNA/DNA hybrids, while topoisomerase I (Top-1) can relax negative supercoils by cutting one of the DNA strands, preventing R-loop formation (Tuduri *et al.*, 2009). A novel role for the tumour suppressor DEAD-Box Helicase 41 (DDX41) in R-loop biology was also recently identified (Mosler *et al.*, 2021). Loss of DDX41 was shown to trigger R-loop accumulation accompanied by replication stress and a genomic instability-associated inflammatory response. R-loops resolution may involve factors that normally resolve other forms of transcriptional stress. Transcription-blocking, bulky DNA lesions are removed by TCR, which has been shown to process R-loops, leading to DSBs (Sollier *et al.*, 2014). In particular, R-loops have been shown to be processed by the Xeroderma Pigmentosum complementation group F (XPF) and group G (XPG) nucleases, alongside with the flap endonuclease 1 (FEN1) (Sollier *et al.*, 2014). In non-replicating cells, R-loops are cleaved at both 3′- and 5′-extremities to induce DSBs, and this dual incision is mediated by XPF/XPG or XPF/FEN1 (Cristini *et al.*, 2019). However, it is still unclear what determines the choice between these specific combinations of nucleases. While R-loop processing causes DNA damage, it may prove an effective way to resolve R-loop mediated transcriptional stress. The breast cancer susceptibility factors BRCA1 and BRCA2 are key factors alleviating R-loop-mediated transcriptional stress both by promoting elongation and resolving R-loops that form after Pol II pausing (Hatchi *et al.*, 2015; Shivji *et al.*, 2018). In addition, the TurboID approach developed by Yan *et al.* (2022) identified homeodomain and zinc finger-containing proteins as highly enriched in the proximity of R-loops. Particularly, the activity-dependent neuroprotector homeodomain protein (ADNP), one of the most frequently mutated in patients with syndromic autism, was identified and

it was shown to directly regulate R-loop structures. Mutations in the *ADNP* gene, causing Helsmoortel-VanDerAa syndrome, lead to R-loop accumulation at ADNP targets, revealing a novel molecular function for ADNP in R-loop suppression (Yan *et al.*, 2022).

The future discovery and characterization of new biological reagents to detect R-loop-interacting factors in cells and advances in proteomics are likely to improve the methods currently available, thereby enhancing our understanding of R-loop regulation.

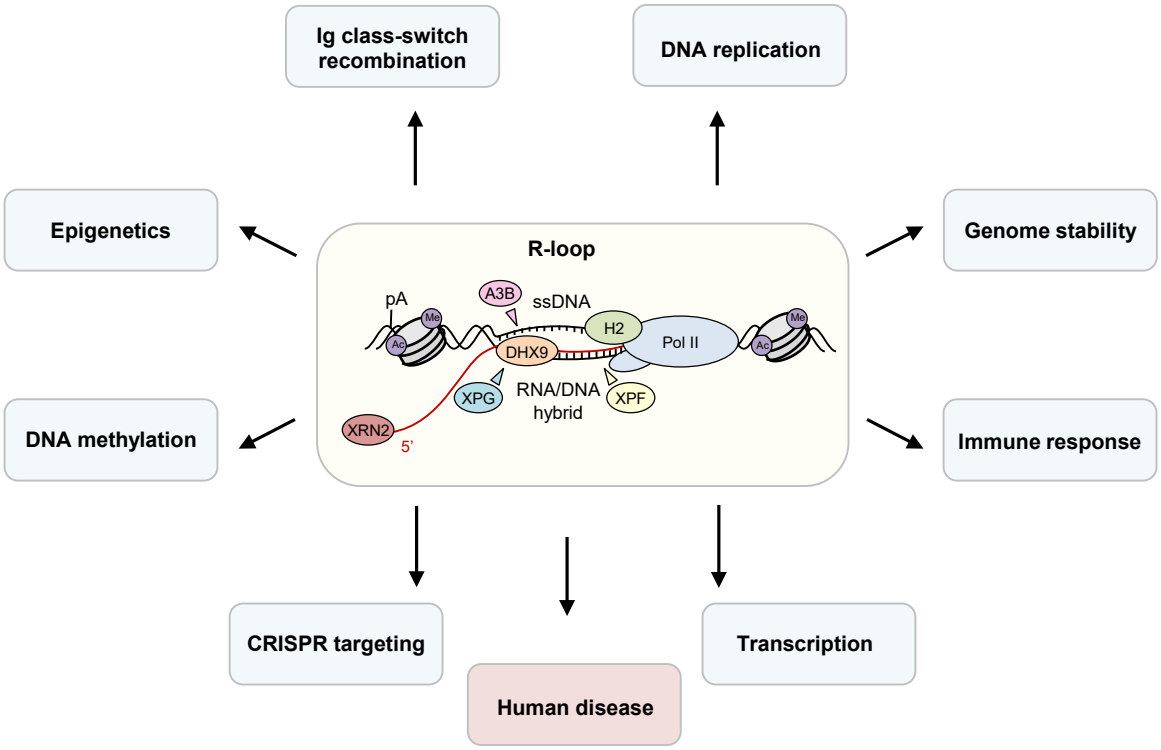


Figure 1.7. Structure and biological functions of R-loops.
Schematic diagram representing the structure of co-transcriptionally-formed R-loops (yellow box). Nascent RNA (red) exiting RNA polymerase II (Pol II) can re-anneal with the single-stranded DNA template strand (ssDNA, black), forming an RNA/DNA hybrid and displacing the non-template DNA strand. Furthermore, transcripts produced by Pol I and III can also form R-loops (not shown). Different R-loop-binding proteins can associate with the ssDNA, the RNA/DNA hybrid and the nascent RNA of the R-loop and some representative examples are shown in the diagram. Further details for each of these R-loop-binding factors can be found in the text. Black arrows indicate R-loop-associated functions and processes (blue boxes, physiological processes; red boxes, disease). Ac and Me indicate methylation and acetylation marks, respectively. A3B, APOBEC3B; DHX9, DExH-Box Helicase 9; H2, RNase H2; XPF, Xeroderma Pigmentosum complementation group F endonuclease; XPG, Xeroderma pigmentosum group G endonuclease; XRN2, 5′-3′-exoribonuclease 2. pA indicates poly A site.

1.3.4 R-loops and human diseases

Although R-loops form naturally during transcription and play an important role in several cellular processes, aberrant accumulation of R-loops is linked to DNA damage and genome instability, which can underlie the pathology of human diseases, including cancer and neurodegeneration (Sollier *et al.*, 2014; Richard & Manley, 2017) (Table. 1.1).

Cancer and senescence

R-loop involvement in cancer is not surprising since R-loop accumulation leads to genome instability and replicational stress, which are major hallmarks of tumour cells (Wells *et al.*, 2019). There are several mechanisms in place to maintain R-loop levels in balance, and alterations in genome caretaker processes can affect R-loop levels and genome stability (Aguilera & García-Muse, 2012; Wells *et al.*, 2019). In breast cancer cells subjected to high levels of estrogen signalling, R-loops were found to accumulate and drive DNA damage at estrogen-induced genes (Stork *et al.*, 2016). Similarly, oncogenic mutations in *HRAS*, which belong to the Ras family of oncogenes, cause R-loop accumulation and subsequent DNA damage and replication stress (Kotsantis *et al.*, 2016). Furthermore, several factors linked to cancer have also been implicated in R-loop regulation. Among other roles, BRCA1 and BRCA2 are major regulators of genome maintenance, and they play a critical role in DNA repair. In some cancers, including breast and ovarian cancer, mutations in BRCA1/2 have been identified, leading to increased R-loop levels and altered expression of oncogenes (Bhatia *et al.*, 2014; Hill *et al.*, 2014; Savage *et al.*, 2014; Hatchi *et al.*, 2015; Shivji *et al.*, 2018; San Martin Alonso & Noordermeer, 2021). In Ewing's sarcoma patient cells, R-loops induced by the oncogenic protein EWS-FLI, a fusion of EWS RNA binding protein 1 (EWSR1) and Friend leukaemia integration 1 transcription factor (FL11), sequester the available BRCA1, rendering these cancers functionally haplo-insufficient for BRCA1, which impedes DNA repair (Gorthi *et al.*, 2018). In addition, mutations in the splicing factors U2AF1 and SRSF2, associated with the myelodysplastic syndrome, have also been linked to an increase in R-loops (Chen *et al.*, 2018). Another example of the interconnection between R-loops and cancer is related to the AID-mediated mutagenesis during immunoglobulin class switch. The single-stranded DNA (ssDNA) in R-loops is a substrate for

cytidine deamination by activation-induced cytidine deaminase (AID), leading to mutagenesis (McCann *et al.*, 2021). These mutations can cause DSB formation, resulting in pathological chromosomal translocations between the Ig loci and other active genes, leading to oncogenic gene expression observed in B cell malignancies (Casellas *et al.*, 2016). In the past, overexpression of the APOBEC family of AID-related enzymes in breast cancer has been linked to genomic mutations, suggesting a role for these antiviral proteins in cancer (Burns *et al.*, 2013; Law *et al.*, 2020). Our lab recently uncovered a novel role for APOBEC3B (A3B) in R-loop homeostasis and demonstrated its contribution to R-loop mutagenesis in cancer (McCann *et al.*, 2021). In detail, it was shown that A3B can specifically bind to R-loops and interact with R-loop-associated proteins (McCann *et al.*, 2021; Zhang *et al.*, 2022). Moreover, A3B was shown to regulate nuclear levels of R-loops through a deamination-dependent mechanism, which affected R-loop distribution genome-wide. These findings suggest that A3B may bind to the exposed nascent ssRNA and/or ssDNA in R-loop structures and deaminate the exposed ssDNA cytosines resulting in mutagenesis (McCann *et al.*, 2021). Furthermore, bioinformatic studies have also shown higher APOBEC3 mutation densities on the non-transcribed strand of actively-expressed genes in multiple cancer types (Chervova *et al.*, 2021). These findings highlight the importance of A3B in R-loop biology and its potential contribution to mutagenesis and cancer development.

R-loops that form at telomeres have been implicated in cellular senescence and the telomeric repeat-containing RNA (TERRA) is a critical component of these R-loops. While TERRA-mediated R-loops are essential for telomere maintenance, they can also lead to telomere fragility and genomic instability (Arora *et al.*, 2014; Feretzaki *et al.*, 2020). The regulation of TERRA and telomeric R-loops is often disrupted in human cancers. Loss of the chromatin remodeler ATRX, which acts at sub-telomeres, is associated with increased telomeric R-loops in cancer cells exhibiting alternative lengthening of telomeres (ALT) activity. BRCA1 and the helicase FANCM suppress telomeric R-loops, thereby keeping telomere replication stress and ALT activity within acceptable limits (Silva *et al.*, 2021; Vohhodina *et al.*, 2021). In addition, adenosine deaminase acting on RNA (ADAR1) can edit A-C mismatches to I-C matched pairs, which facilitates the resolution of telomeric R-loops by RNase H2 and ensures genome stability at telomeres in cancer cells (Shiromoto *et al.*, 2021).

ADAR1-dependent control of telomeric R-loops is critical for the proliferation of non-ALT or telomerase-activated cancer cells and identifies ADAR1 as a promising therapeutic target of telomerase-positive cancers (Shiromoto *et al.*, 2021).

Neurological diseases

Several neurodegenerative disorders have been associated with R-loop accumulation (Table 1.1). Neurodegenerative disorders affect non-dividing neuronal cells, where excessive R-loop accumulation could lead to transcription-associated DNA damage and genome instability, independently of DNA replication. Cerebellar ataxias are linked to mutations in DNA repair genes, including *ATM*, which result in increased DNA damage in neurons (Katyal *et al.*, 2014; Becherel *et al.*, 2015). Mutations in *SETX* are associated with ataxia with oculomotor apraxia 2 (AOA2) and ALS4 (Becherel *et al.*, 2015; Grunseich *et al.*, 2018). Prader-Willi (PWS) and Angelman (AS) syndromes, two autism-spectrum disorders caused by large deletions or mutation of the maternal allele of the ubiquitin protein ligase E3A (*UBE3A*), involve epigenetic silencing of the paternal allele of UBE3A through the expression of an antisense Ube3a transcript (*Ub3a-ATS)* (Cassidy *et al.,* 2000). Interestingly, R-loops were shown to regulate the neuronal expression of the paternal *Ub3a-ATS* transcript*,* and the stabilization of R-loops with topotecan, a Top1 inhibitor, result in chromatin decondensation, and inhibition of *Ub3a-ATS* transcription, which also increases *UBE3A* expression (Powell *et al.*, 2013). Furthermore, mutations in RNA-binding proteins, such as the TAR DNA-binding protein 43 (TDP-43) implicated in ALS/FTD, have been also linked to R-loop accumulation (Giannini *et al.*, 2020; Wood *et al.*, 2020).

As mentioned in Section 1.1.4.1, R-loop formation at expanded trinucleotide DNA repeats has been associated with heterochromatin formation and transcriptional repression of expanded host genes linked to FRDA, FXS and ALS/FTD (Huertas & Aguilera, 2003; Groh *et al.*, 2014a; Loomis *et al.*, 2014). Mutations in RNase H2, an enzyme that degrades RNA in RNA/DNA hybrids, cause Aicardi Goutières syndrome (AGS), a severe neuroinflammatory encephalopathy (Crow *et al.*, 2006). AGS is also caused by mutations in other nucleic acid-processing enzymes, including TREX1, ADAR1 and SAMHD1, which result in the accumulation of endogenous nuclei acid with activation

of the cGAS-STING pathway and interferon responses (Mackenzie *et al.*, 2016; Park *et al.*, 2021). Recent studies suggest that R-loops may contribute to AGS pathogenesis as a source of immunogenic nucleic acids that trigger DNA damage and inflammation (Cristini *et al.*, 2022; Giordano *et al.*, 2022). In particular, the absence of RNase H2 was shown to result in the accumulation of co-transcriptional R-loops and impaired transcription. Accumulated R-loops are then processed by XPG and XPF endonucleases, leading to DNA damage and activation of the immune-response, features associated with AGS (Cristini *et al.*, 2022).

Table 1.1 R-loops and links to human diseases

Disease	R-loop Factors/ Locus	References
Cancer		
Myelodysplastic syndromes (MDS), acute myeloid leukemia (AML), uveal melanoma, bladder, pancreatic and lung cancer	SRSF2, U2AF1, SF3B1, ERCC6L2 U2AF2 and ZRSR2	Chen *et al*., 2018; Tummala *et al*., 2018;
Breast/Ovarian cancer	BRCA1/2, Estrogen	Stork *et al.,* 2016; Bathia *et al.,* 2014; Shivij *et al.,* 2018;
HRAS induced-cancer	HRAS	Kotsantis *et al.,* 2016;
Gastric cancer	RNAses	Bauer *et al.*, 2020;
Ewing's sarcoma	EWS-FLI, BRCA1	Gorthi *et al.,* 2018;
Multiple myeloma (MM) and Burkhitt's lymphoma	TRD3-TOP3B	Yang *et al.*, 2014;
Alternative lengthening of telomeres (ALT)-dependent cancers	Telomeric repeat- containing RNA (TERRA), ATRX (alpha thalassemia/mental retardation X-linked), ADAR1	Arora *et al.,* 2014; Shiromoto *et al*., 2021;
AIDS-associated malignancies	APOBEC3B, TREX complex	Jackson *et al.,* 2014; McCann *et al*., 2021;
Fanconi anemia (FA)	FANCA, FANCD2, FANACI, FANCJ, and FANCM	Schwab *et al.*, 2015; Garcia-Rubio *et al.*, 2015;
Neurological disorders		
Ataxia with oculomotor apraxia type 2 (AOA2)	SETX	Becherel *et al*., 2015;
Juvenile amyotrophic lateral sclerosis-4 (ALS4)	SETX	Grunseich *et al*., 2018;
Prader-Willi Syndrome (PWS) and Angelman Syndrome (AS)	SNORD116	Powell *et al.,* 2013;
Amyotrophic lateral sclerosis (ALS) and frontotemporal dementia (FTD)	C9ORF72, TDP-43/FUS	Haeusler *et al.,* 2014; Giannini *et al.,* 2020; Wood *et al.,* 2020;
Friedreich ataxia (FRDA); Fragile X syndrome (FXS); FXTAS (fragile X tremor ataxia syndrome) and FRAXE (fragile X syndrome type E)	FXN, FMR	Groh *et al.,* 2014; Loomis *et al.,* 2014; Colak *et al.,* 2014;
Aicardi-Goutières syndrome (AGS)	RNaseH2, SAMHD1, TREX1	Cristini *et al.,* 2022; Giordano *et al.,* 2022.

1.4 Aims of this work

Huntington's Disease is a fatal neurodegenerative disorder that profoundly impacts the lives of patients and their families. Although the genetic basis of the disease is well established and progress has been made in identifying the function of the mutant HTT protein, the mechanisms that link the expansion of CAG repeats in the *HTT* gene to HD phenotypes are not fully understood. It has been shown that transcription is defective in HD, and it is widely accepted that HTT protein plays a role in transcription. However, a thorough understanding of the mechanisms underlying the transcriptional dysregulation seen in HD cells is lacking. In particular, the transcriptional regulation of the *HTT* gene itself remains poorly understood, and the factors that influence it are unknown. Interestingly, the formation of non-canonical RNA/DNA structures, such as R-loops and dsRNAs, has been shown to be associated with the pathogenesis of several repeat expansion disorders and neurodegenerative diseases. These pathological structures can interfere with transcription and can affect repeat stability. However, whether these unusual RNA/DNA structures contribute to HD pathology has never been explored. Additionally, it is not yet clear how the loss-of-function of wtHTT and the gain-of-function of mHTT protein contribute to transcriptional dysregulation in HD. Therefore, the three main aims of this project were to study the transcriptional regulation of the *HTT* gene, to investigate the role of non-canonical RNA/DNA structures in HD pathogenesis and to understand the role of HTT protein in transcription and R-loop biology.

Aim 1: The transcriptional regulation of the *HTT* gene (Chapter 3)

The first main aim of this project was to investigate the transcriptional regulation of the *HTT* gene in Control and HD cells and study the impact of the expanded repeats on this process. Furthermore, I also investigated the *HTT* chromatin structure and its epigenetic signature. I showed that the phenotypes observed are conserved in different HD cellular models, including Medium Spiny Neurons differentiated from iPS cells (provided by Prof. Tabrizi) and an HD mouse model.

Aim 2: The role of non-canonical RNA/DNA structures in HD (Chapter 4)

The formation of non-canonical RNA/DNA structures, such as R-loops and double-stranded RNAs (dsRNAs), over the expanded CAG repeats *in vivo,* has not been fully addressed in the context

of HD. I have therefore investigated dsRNA and R-loop formation in HD. Initially, I analysed and characterized their formation over expanded *HTT* repeats in human cells from Control and HD patients. I detected and further characterized dsRNA structures forming over expanded *HTT* RNA. I then investigated R-loop formation genome-wide and observed R-loop enrichment in HD patient cells, validating these results on specific genes by DRIP-seq.

Aim 3: The role of the HTT protein in the nucleus (Chapter 5)

The aim of the experiments detailed in Chapter 5 was to explore the role of HTT protein in the nucleus. In detail, firstly I analysed HTT binding to genomic DNA. I then investigated the interaction of the HTT protein with different transcriptional and R-loop factors and its association with R-loops *in vivo*. Since R-loops have been recently shown to be a source of DNA damage and inflammation, I investigated these phenotypes in Control and HD cells. I observed that DNA damage and inflammation correlate with R-loop accumulation in HD cells. Finally, I have addressed the potential molecular mechanism underlying the role of HTT protein in maintaining R-loop homeostasis.

1.5 Publications

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

R-Loop Immunoprecipitation: A Method to Detect R-Loop Interacting Factors.
Beghè C, Gromak N.
Methods Mol Biol. 2022;2528:215-237. Doi: 10.1007/978-1-0716-2477-7_14

DNA damage contributes to neurotoxic inflammation in Aicardi-Goutières syndrome astrocytes.
Giordano AMS, Luciani M, Gatto F, Abou Alezz M, **Beghè C**, Della Volpe L, Migliara A, Valsoni S, Genua M, Dzieciatkowska M, Frati G, Tahraoui-Bories J, Giliani SC, Orcesi S, Fazzi E, Ostuni R, D'Alessandro A, Di Micco R, Merelli I, Lombardo A, Reijns MAM, Gromak N, Gritti A, Kajaste-Rudnitski A.
J Exp Med. 2022 Apr 4;219(4):e20211121. Doi: 10.1084/jem.20211121

R-loop homeostasis and cancer mutagenesis promoted by the DNA cytosine deaminase APOBEC3B. McCann, J Cristini A, Law E, Lee SY, Tellier M, Carpenter M, **Beghè C**, Kim J, Jarvis M, Stefanovska B, Temiz N, Bergstrom E, Salamango D, Brown M, Murphy S, Alexandrov LB, Miller K, Gromak N*, Harris R*
In revision in Nature Genetics, bioRXiv https://doi.org/10.1101/2021.08.30.458235

Chapter 2

Materials and Methods

2.1 Mammalian cell culture techniques

2.1.1 Cell lines

Epstein-Barr virus (EBV)-transformed lymphoblast cell lines from healthy (GM04844, GM15851, GM6895, GM14926, GM04780, GM04205) and HD (GM04856, GM14044, GM04868, GM04768, GM05538) patients were obtained from Coriell Institute for Medical Research (Table 2.1). Lymphoblast cell lines were grown in Roswell Park Memorial Institute 1640 medium (RPMI 1640, Sigma-Aldrich) supplemented with 15% fetal bovine serum (FBS, Gibco) and 100 μg/mL streptomycin (Gibco, Thermo Fisher Scientific) at 37°C in 5% CO_2.

Human dermal fibroblast cell lines from healthy (GM04548) and HD patients (GM04857, GM09197) were obtained from Coriell Institute for Medical Research (Table 2.1). Fibroblast cell lines were grown in Dulbecco's modified Eagle's medium (DMEM, Sigma-Aldrich) supplemented with 10% FBS (Gibco), 1% Non-Essential Amino Acids (NEEA, Lonza) and 100 μg/mL streptomycin (Gibco) at 37°C in 5% CO_2. For both lymphoblast and fibroblast cell lines, CAG repeat sizes are indicated in Table 2.1.

HeLa cells were obtained from Dr N. Gromak (University of Oxford). SH-SY5Y cells were kindly provided by Prof. Peter Cook (University of Oxford). HeLa and SH-SY5Y cells were grown in DMEM supplemented with 10% FBS (Gibco) and 100 μg/mL streptomycin (Gibco) at 37°C in 5% CO_2. All experiments were performed with early passages of the cell lines and cell lines were checked for mycoplasma contamination.

To prepare frozen stock of HeLa, lymphoblast and fibroblast cells, $8x10^6$ cells were resuspended in 1.5 mL of freezing medium (90% v/v FBS, 10% DMSO). All frozen stocks were stored at -80°C. Frozen cells were thawed gradually warming up at room temperature, diluted in fresh media followed by centrifugation and resuspension again in medium with antibiotics.

2.1.2 iPSC lines

Feeder-free iPSCs from healthy individuals were kindly provided by Prof. Richard Wade-Martins (DPAG, Universtity of Oxford) and grown in mTeSr medium (Life Technologies). On the day of cell seeding, 6-well plates were pre-coated by adding 1 mL of Matrigel (Corning) previously

diluted in DMEM-KO medium (Thermo Fisher Scientific) as per manufacturer instructions. The plate was gently shaken to cover all the surface area with the coating solution and transferred to a humidified incubator at 37°C in 5% CO_2 for at least 1 h. The Matrigel was gently removed from the plate right before seeding. After mTeSr medium was removed, cells were gently washed with 2 mL of PBS (Gibco), and then 1 mL of PBS (Gibco) supplied with 0.05 mM EDTA (PBS-EDTA 0.02 %) was added. The cells were incubated at 37°C in 5% CO_2 for about 3 min. PBS-EDTA 0.02 % was then aspirated and cells were resuspended in fresh mTeSr medium supplied with 10 µM Rock inhibitor (Y-27632 dihydrochloride, Tocris) at a ratio from 1:3 to 1:6, seeded on coated wells and fed daily thereafter. To prepare frozen stock of iPSCs, one 90 % confluent well of a 6-well plate was resuspended in 1 mL of freezing medium (90% v/v Knockout Serum Replacement (KSR, Invitrogen), 10% DMSO). All frozen stocks were stored in liquid nitrogen. Frozen cells were thawed gradually at room temperature, diluted in fresh media followed by centrifugation and resuspension again in fresh medium.

2.1.3 Cell Growth Curve

At day 0, $5x10^3$ fibroblast cells were seeded onto a 96-well plate. Cells were harvested by trypsinization at the indicated time points (Fig. 5.6D) and resuspended in 200 µl of PBS (Gibco). Growth curves were generated by manual cell counting using the Neubauer Chamber.

2.1.4 Pharmacological treatments

10 µM Camptothecin (CPT, Sigma-Aldrich) was added to HeLa cells treated with siLuc and siHTT RNAs 1 h prior Immuno-fluorescence (IF) experiments. Actinomycin D (ActD, Sigma-Aldrich) was added to a final concentration of 5 µg/mL to the media of Control and HD lymphoblast cell lines 18 h prior Slot Blot analysis.

2.1.5 siRNA-mediated knock-down

The siRNA-mediated knock-downs were carried out as described previously (Cristini *et al.*, 2022). siRNA transfections were performed for 72 h in HeLa cells using Lipofectamine 2000 (Thermo Fisher Scientific) according to the manufacturer's instructions. HeLa cell transfections were performed in 6-well plates 24 h after seeding with 22 nM siRNA and the medium was exchanged 6 h

after transfection. The day after, transfection was repeated, and the cells were re-seeded 24 h before the experiment. siRNAs were purchased from Horizon Discovery and their sequences can be found in Table 2.3.

2.1.6 Medium Spiny Neurons (MSNs) differentiation protocol

MSNs differentiation was carried out as previously described with few adjustments (Fjodorova & Li, 2018). For MSNs differentiation, iPSCs colonies were detached with PBS (Gibco) supplied with 0.05 mM EDTA (PBS-EDTA 0.02%), suspended by pipetting as single cells and plated on Matrigel (Corning)-coated dishes (80,000 cells/cm^2) with mTeSr medium in the presence of 10 μM of the Rock inhibitor Y-27632 (Sigma). The complete N2B27 differentiation culture medium was prepared as follows. N2B27 medium included 100 mL of DMEM /F-12 medium (Thermo Fisher Scientific), 50 mL of Neurobasal medium (Gibco), 1.5 mL of 200 mM L-glutamine (Lonza), 1.5 mL of 10,000 U/mL penicillin/streptomycin (Gibco), 1 mL of 100 X N-2 supplement (Thermo Fisher Scientific) supplemented 1:10 with 7.5% bovine serum albumin (Thermo Fisher Scientific), 1 mL of 50X B-27 supplement (with/without vitamin A) (Gibco), and 150 μL of 50 mM 2-mercaptoethanol solution (Gibco). When the cell culture reached ~90% confluence (usually 2 days after plating), the cells were supplied with complete N2B27 medium without vitamin A, 100 nM LDN-193189 (Tocris), 200 nM dorsomorphin dihydrochloride (DM) (Abcam), 10 μM SB 431542 (R&D Systems) (Neural induction medium A). The Neural induction medium A was changed daily for the next 4 days. At day 5, the cells were supplied with complete N2B27 medium without vitamin A, 100 nM LDN-193189 (Tocris), 200 nM DM (Abcam) (Neural induction medium B). The Neural induction medium B was changed daily for the next 4 days. At day 9, cells were incubated with N2B27 supplied with 10 μM Rock inhibitor (Y-27632 dihydrochloride, Tocris) and 25ng/mL Activin A (BioLegend) (1:8.000). Medium was collected in 50 mL tubes and the culture was washed with PBS once. 1 mL of PBS-EDTA 0,02% was added and cells were incubated for 2 min at 37°C until they visibly detached under the microscope. PBS-EDTA 0,02% was manually aspirated and the previously collected medium was supplied to the cells. Cells were gently scratched with a 10 mL pipette, collected, diluted in Neural induction medium B and replated 1:10 ratio on pre-coated wells. The day

after (day 10), the cells were supplied with complete N2B27 medium without vitamin A, 25 ng/mL recombinant human Activin A (BioLegend) (LGE patterning medium). Thereafter, LGE patterning medium was changed daily or every other day depending on cell growth for the next 9 days. At day 20, the culture was washed once with PBS (Gibco), and then incubated with 1 mL of PBS-EDTA 0,02% for 2 min at 37°C until the cells were visibly detached under the microscope. The PSB-EDTA 0,02% was removed and 1 mL of LGE patterning medium was added. The cells were gently scratched with a p1000 mL pipette, collected, and diluted in fresh LGE patterning medium. The cells were then counted, and 10.000-20.000 cells/24 well-plate were plated on pre-coated wells. The day after, the LGE patterning medium was switched to fresh complete N2B27 medium with vitamin A, 10 ng/mL recombinant human brain-derived neurotrophic factor (BDNF) (PeproTech), 10 ng/mL recombinant human glial cell line-derived neurotrophic factor (GDNF) (PeproTech) and 100 nM of Retinoic Acid (RA) (Sigma-Aldrich) (Terminal differentiation medium), thereafter changed every 4 days until the end of the protocol (day 35).

2.2 RNA analysis methods

2.2.1 Purification of total RNA

Total RNA was extracted using TRIzol reagent (Invitrogen) following manufacturer's instructions. Tissue culture media was removed, and cells were lysed by the addition of 1 ml TRIzol (for up to $5x10^6$ cells) and vigorous pipetting. 200 µl of chloroform were added and samples were incubated for 5 min at room temperature followed by 20 min centrifugation at 16,000 g at 4°C. The top transparent phase was carefully transferred to a new tube and RNA was precipitated by the addition of an equal volume of isopropanol, 20 µg of glycogen (20 mg/ml, Roche) and incubated overnight at -20°C. RNA was then precipitated by 25 min centrifugation at 16,000 g at 4°C. Pellets were washed twice with 1 ml 75% EtOH and centrifuged for 10 min at 16,000 g at 4°C and air-dried at room temperature until the pellet became transparent. Pellets were dissolved in RNase-free ddH_2O followed by 2 h DNase I treatment (Roche) to remove contaminating genomic DNA, followed by heat inactivation for 15 min at 75°C.

2.2.2 Reverse transcription qPCR (RT-qPCR)

1-2 µg of total RNA was reverse transcribed using SuperScript Reverse Transcriptase III (Invitrogen) and 100 ng of random primers (Cat. N. 48190011, ThermoFisher), following manufacturer's instruction. For quantitative real-time PCR, 2 µl of cDNA were analysed using Rotor-Gene® Q (QIAGEN) with QuantiTect SYBR green (QIAGEN) using primers indicated in Table 2.2, according to manufacturer's instruction.

2.2.3 5'-EU labelling

Fibroblast cells were seeded on coverslips and incubated with 1 mM 5-ethynyl uridine (EU) for 1 h to label newly synthesized RNA. Global nascent RNA transcription was detected using Click-iT RNA Alexa Fluor 488 Imaging Kit (ThermoFisher, C10329) according to the manufacturer's instructions. Coverslips were mounted using Fluoromount-G (SouthernBiotech, 0100-01) containing DAPI and slides were visualized with an inverted confocal microscope (Olympus Fluoview FV1200). Fluorescence intensity was analysed with ImageJ (version 1.52p) as previously described (Cristini *et al.*, 2022).

2.2.4 dsRIP analysis

dsRIP was performed essentially as described previously (Aktas *et al.*, 2017). One day before the experiment, 2 flasks of $4x10^5$ lymphoblast cells per experimental condition were seeded. Flask was processed in parallel according to the following protocol. Cells were pelleted by 5 min centrifugation at 500 g at 4°C, washed in cold PBS and incubated in Cell Lysis Buffer (85 mM KCl, 5 mM PIPES pH 8.0, 0.5% NP-40 0.5 mM PMSF, 0.8 µg/µl pepstatin A, 1 µg/µl leupeptin) for 10 min on ice. Nuclei were collected by 5 min centrifugation at 500 g at 4°C and then incubated in RIP buffer (25mM HEPES pH 7.2, 75 mM NaCl, 5 mM $MgCl_2$, 0.1% NP-40) for 10 min in ice. Extracts were then sonicated for 5 min at medium power (cycles of 30 seconds "ON" and 30 seconds "OFF", Diagenode Bioruptor) and centrifugated for 10 min at 16,000 g at 4°C. Lysate of the same condition were pulled together. Total RNA from 10% of the lysate (Input) was resuspended in TRIzol (Invitrogen). The remaining lysate was incubated with 3 µg dsRNA antibody J2 (Sciscon) at 4°C for 2 h. Lysates were then incubated with 25 µl of Dynabeads Protein A (ThermoFisher) for 1.5 h at 4°C.

The beads were washed 4 times with RIP buffer and bound RNA was purified using TRIzol following manufacturer's instructions. The obtained RNA was finally resuspended in 25 µl of ddH_2O and 1 h of DNase I treatment (Roche) was performed to remove contaminating genomic DNA, followed by heat inactivation for 15 min at 75°C. 3µl of total eluted RNA was transcribed with Superscript III using random primers (ThermoFisher). cDNA was used as a template for quantitative real-time PCR. dsRIP specificity was tested by digestion with recombinant RNase III (2U, NEB), 45 min at 37°C in the presence of 1X RNase reaction buffer (NEB) after immuno-precipitation. Attempts to perform the treatment before immuno-precipitation were not successful due to the composition of the RIP Buffer, used for nuclear lysis, which is not compatible with the RNase III enzyme activity.

2.3 DNA analysis methods

2.3.1 Agarose gel electrophoresis

Agarose gel electrophoresis was used to measure the size of PCR products and sonicated DNA fragments for DRIP, R-loop IP and ChIP experiments. DNA was mixed with either 5X Gel Pilot Loading Dye (QIAGEN) or 6X Gel Loading Dye Purple (New England BioLabs) and separated at 100 V using 1-2.5% agarose gels containing 1X TE and 0.005% ethidium bromide. DNA sizes were then estimated by comparison with a 100 bp DNA ladder (Gene Ruler, Thermo Fisher Scientific) loaded alongside the samples. DNA bands were visualized under UV light.

2.3.2 Real-time quantitative PCR (qPCR)

Real-time quantitative PCR (qPCR) was employed to amplify and simultaneously quantify specific DNA sequences. The fluorescent dye SYBR Green (QIAGEN), which intercalates into DNA, is used to monitor DNA amplification in qPCR. For each 10 µl reaction, 2 µl of purified DNA or cDNA were mixed with 5 µl of 2x QuantiTect PCR SYBR Green master mix (QIAGEN), 0.5 µl of forward primer (10 µM), 0.5 µl of reverse primer (10 µM) and 2 µl of ddH_2O (Table 2.2). All qPCR reactions were performed in technical triplicates (or duplicates) to minimize technical error. In addition, a non-template control and a dilution series of purified genomic DNA (from HeLa or HEK293 cells) were amplified alongside samples for each primer pair. Genomic standard dilutions

(1/10-1/400) for qPCR had the following concentrations: 20 ng/ µl (1/1), 4 ng/ µl (1/5), 0.8 ng/ µl (1/25), 0.2 ng/ µl (1/100), 0.05 ng/ µl (1/400).

2.4 Protein analysis methods

2.4.1 Preparation of whole cell extracts

Whole-cell extracts were prepared by washing cells twice in ice-cold PBS (Gibco). After centrifugation, cell pellets were resuspended either in RIPA lysis buffer (25 mM TRI-HCl pH 7.6, 150 mM NaCl, 1% NP-40, 1% Deoxycholate, 0.1% SDS) or Tris-SDS lysis buffer (10mM Tris HCl pH 7.5, 1% SDS) supplemented with 1X Complete EDTA-free protease inhibitors (Sigma-Aldrich) and phosphatase inhibitors (ThermoFisher Scientific). After incubation on ice for 20 min, samples were sonicated (2-5 min sonication for RIPA extracts or 15 min sonication for Tris-SDS extracts) at medium power to reduce the viscosity (cycles of 30 seconds "ON" and 30 seconds "OFF", Diagenode Bioruptor). Extracts were then centrifuged for 15 min at 16,000 g at 4°C. Protein concentration of the supernatant was determined using either Bradford assay (BioRad Laboratories) or BCA assay (ThermoFisher Scientific) for RIPA and Tris-SDS extracts, respectively, and stored at -80°C until further use.

2.4.2 Cellular fractionation

Whole-cell extracts were prepared in RIPA Buffer as reported in Section 2.4.1. Subcellular fractionation was performed as previously described with minor changes (Senichkin *et al.*, 2021). Cells were washed in ice-cold PBS (Gibco) and centrifuged for 5 min at 500 g at 4°C. Pelleted cells were then resuspended in 200 µl of ice-cold Hypotonic Buffer (20 mM Tris-HCl pH 7.4, 10 mM KCl, 2 mM $MgCl_2$, 1 mM EGTA, 0.5 mM DTT, 0.5 mM PMSF) and incubated on ice for 3 min. Additional NP-40 was added to a final concentration of 0.1% and samples were incubated for another 3 min on ice. The cells were then centrifuged for 5 min at 1,000 g at 4°C to separate the nuclei (pellet) and the cytoplasm (supernatant). The supernatant was centrifuged for 2 min at 16,000 g at 4°C to pellet cell debris (cytoplasmic fraction, C). Nuclei were resuspended in 200 µl Isotonic Buffer (20 mM Tris-HCl pH 7.4, 150 mM KCl, 2 mM $MgCl_2$, 1 mM EGTA, 0.5 mM DTT, 0.5 mM PMSF) containing 0.1% NP-40 and incubated on ice for 10 min followed by centrifugation for 3 min at 1,000

g at 4°C. Nuclei were then resuspended in 100 µl ice-cold RIPA Buffer (25 mM Tris-HCl pH 7.6, 150 mM NaCl, 1% NP-40, 1% Deoxycholate, 0.1% SDS) for 20-30 min on ice and then centrifugated for 3 min at 2,000 g at 4°C. Supernatant represented nucleoplasmic fraction (NP), whereas the pellet represented the chromatin fraction (Chr). The chromatin fraction was then incubated with 40 U of benzonase (E1014, Sigma-Aldrich) for 15 min at 37°C on a shaker at 500 rpm to degrade nucleic acid species. Protein concentrations of the different fractions obtained were quantified with Bradford assay (BioRad Laboratories) and stored at -80°C until further use.

2.4.3 Co-Immuno-precipitation (co-IP)

Non-cross-linked HeLa cells were plated on 150 mm plates one day prior to the experiment to achieve 70-80% confluency the following day. BSA-blocked protein A Dynabeads (Invitrogen) were incubated with HTT antibody (3 µg, MAB2170, Millipore) and matched isotype mouse-IgG antibody (M5409, Sigma) in 1 mL IP buffer (25 mM Tris-HCl pH 8.0, 150 mM NaCl, 0.5% NP-40%, 10% glycerol, 2.5 mM $MgCl_2$) supplemented with 1X Complete EDTA-free protease inhibitors (Sigma-Aldrich) for 2 h at 16 rpm on a rotating wheel at 4 °C. Pelleted cells were washed twice in PBS (Gibco) containing protease inhibitors (Roche) and then centrifuged for 5 min at 500 g at 4°C. Cell pellets were resuspended and incubated in 800 µl of Lysis Buffer (50 mM Tris-HCl pH 8.0, 150 mM NaCl, 1% NP-40, 10% glycerol, 2.5 mM $MgCl_2$), supplemented with 0.5 mM PMSF, 1X Complete EDTA-free protease (Sigma-Aldrich) and 1X PhosSTOP phosphatase (Roche) inhibitors in the presence of 40 U benzonase (E1014, Sigma-Aldrich) for 30 min at 16 rpm on a rotating wheel at 4 °C. Following lysis, insoluble material was removed by centrifugation at 16,000 g for 15 min at 4°C and supernatant was diluted 1:2 in Dilution Buffer (150 mM NaCl, 10% glycerol, 2.5 mM $MgCl_2$) supplemented with 0.5 mM PMSF, 1X Complete EDTA-free protease (Sigma-Aldrich) and 1X PhosSTOP phosphatase (Roche) inhibitors.

Proteins were quantified by Bradford assay. At least 100 µl of whole cell lysate was stored at -20°C as 'protein Input' for Western Blot analysis. 1 mg of proteins were subjected to immuno-precipitation with antibody-conjugated to protein A Dynabeads for 2 h at 16 rpm on a rotating wheel at 4°C. Immuno-complexes were washed three times with 1 mL of IP buffer and three times with 1

mL of IP buffer without NP-40, both supplemented with 1X Complete EDTA-free protease inhibitors (Sigma-Aldrich). Immuno-precipitated complexes were eluted in 40 µl of 1X LDS (Invitrogen) and 100 mM DTT for 10 min at 70 °C before analysis via Western Blot. Proteins were resolved on 4-12% Bis-Tris or 3-8% Tris-Acetate (Invitrogen) by SDS-PAGE and immunoblotted with indicated antibodies (Table 2.4) (see Section 2.4.5).

2.4.4 Protein silver staining

Silver staining was performed as reported previously (Cristini *et al.*, 2018). Samples were prepared and run on a SDS-PAGE gel. Gels were rinsed in ddH_2O for 3 min and fixed in Fixative Solution (50% EtOH v/v, 10% Glacial Acetic Acid, 1.85% formaldehyde) for 1 h at room temperature. Fixed gels were washed at least 6 times for 10 min in large volumes of ddH_2O at room temperature using a horizontal shaker. Gels were then incubated in ddH_2O containing 2.5 µM of 1M DTT for 1 h at room temperature. Further rinsing with ddH_2O (3x brief washes) was followed by staining in Silver Nitrate Solution (0.1% w/v $AgNO_3$) for 1 h at room temperature, rinsing in ddH_2O once more and developing in Silver Developing Solution (3% w/v Na_2CO_3, 1.85% formaldehyde). Gels were shaken gently for 3-15 min in Silver Developing Solution (3% w/v Na_2CO_3, 1.85% formaldehyde) and the development was stopped by incubation in Silver Stopping Solution (50% EtOH, 10% acetic acid) for 30 min at room temperature under gentle horizontal shaking. The fixed gels were then washed extensively with ddH_2O for 30 min and dried.

2.4.5 Western Blot analysis

Protein samples were prepared by denaturing in 1X LDS (Invitrogen) and 100 mM DTT by boiling at 70 °C for 10 min. Samples were then separated using SDS-PAGE. Commercial gels, used in this study, were either NuPAGE 3-8% Tris-Acetate or NuPAGE 4-12% Bis-Tris gels (Invitrogen) using either Tris-Acetate or MOPS NuPAGE running buffer (Invitrogen), respectively. Homemade polyacrylamide gels were cast following Immunoblotting protocol (Sigma-Aldrich) and run in homemade running buffer (25 mM Tris, 192 mM glycine, 1 % w/v SDS).

Both commercial and homemade gels were equilibrated in either 2X NuPaGE Transfer Buffer (Invitrogen) or homemade transfer buffer (25mM Tris, 192 M glycine) containing 10% w/v

MeOH for 15 min at room temperature. Amersham Hybond-P PVDF membranes (GE Healthcare) were activated by incubation for 1 min in 100% MeOH, followed by washes in ddH_2O for 5 min and equilibration in transfer buffer for 15 min. Proteins were transferred onto PVDF membrane using a Trans-Blot Semi-Dry transfer cell (Bio-Rad laboratories) at 15 V for 1.15-2.15 h using either 2X NuPAGE Transfer Buffer or homemade transfer buffer (see composition above). Membranes were then blocked in either 5% BSA Tris-buffered saline with 0.1% Tween (0.1% TBST) (to preserve phosphorylation status) or 10% Milk 0.1% TBST. All washes were done in 0.2% TBST. Primary antibodies were diluted in either 2% Milk 0.1% TBST or 5% BSA 0.1% TBST and membranes were incubated overnight under gentle horizontal shaking at 4°C. After 3 x washes in 0.2% TBST for 5 min, membranes were incubated with secondary antibodies diluted in either 1% Milk 0.1% TBST or 5% BSA 0.1% TBST for 1 h under gentle horizontal shaking at room temperature. Membranes were then extensively washed in 0.2% TBST. Images were acquired by chemiluminescence using autoradiography.

2.4.6 Immuno-fluorescence (IF)

Cells were fixed with 4% paraformaldehyde (in 1X PBS, Gibco) for 15 min at room temperature and permeabilized with 0.1% Triton X-100 for 20 min at room temperature. For blocking non-specific sites, cells were incubated for 30 min in 1X PBS (Gibco) supplemented with 2% BSA (B8894, Sigma-Aldrich) and stained overnight at 4°C with primary antibodies (refer to Table 2.4 for the complete list of antibodies and dilutions). For S9.6 IF, cells were fixed with ice-cold methanol for 10 min at -20°C, washed with PBS and permeabilized with 0.1% Triton X-100 for 10 min at room temperature. After PBS washes, slides were treated with 150 U/ml of RNase H (NEB, M0297) or left untreated for 4 h at 37°C. After PBS washes, slides were blocked with 8% BSA before incubating with purified S9.6 antibody (Kerafast) overnight at 4°C. The next day, after three washes with 1X PBS (Gibco) of 15 min, slides were incubated with donkey anti-rabbit IgG, Alexa Fluor 488 (1:500 dilution; A-21206; Thermo Fisher Scientific) or donkey anti-mouse IgG, Alexa Fluor 555 (1:500 dilution; A-31570; Thermo Fisher Scientific) for 1.5 h at room temperature. Coverslips were mounted using Fluoromount-G (SouthernBiotech, 0100-01) containing DAPI and slides were

visualized with an inverted confocal microscope (Olympus Fluoview FV1200). Fluorescence intensity was analysed with ImageJ (version 1.52p) as described in Cristini *et al.,* (2022). For γH2AX IF, image analysis was performed using Image J and Cell Profiler software as previously described (Giordano *et al.*, 2022). The macro for the analysis of EU intensity has been created by Dr. Alain Wainman, Light Microscopy Facility Manager (Dunn School of Pathology, Oxford University).

2.4.7 Chromatin immuno-precipitation (ChIP)

Chromatin immuno-precipitation was performed as described previously with minor modifications (Cristini *et al.*, 2018). One day before the experiment, cells were plated to reach 80% confluence on the day of the experiment. After 24 h, cells were crosslinked with 1% formaldehyde at 37°C for 15 min on a horizontal shaker and the reaction was quenched by addition of 0.125 M glycine at 37°C for 5 min. All subsequent steps were carried out at 4°C or on ice. Cells were washed twice with PBS (Gibco) containing protease inhibitors (0.5 mM PMSF, 1X Complete EDTA-free protease inhibitors, Sigma-Aldrich). Nuclei were isolated by lysing cells with 300 µl of Cell Lysis Buffer (5 mM PIPES pH 8.0, 85 mM KCl, 0.5% NP-40 and 1X Complete EDTA-free protease inhibitors, Sigma-Aldrich). After incubation on ice for 10 min, nuclei were centrifuged for 5 min at 500 g at 4°C and lysed in 400 µl Nuclear Lysis Buffer (50 mM Tris-HCl pH 8.0, 0.5 mM EDTA, 1% SDS, 0.5 mM PMSF, 1X Complete EDTA-free protease inhibitors, Sigma-Aldrich). Nuclear extracts were then incubated 10 min on ice before they were sonicated for 10-15 min at medium power (cycles of 30 seconds "ON" and 30 seconds "OFF", Diagenode Bioruptor). Insoluble material was removed by centrifugation at 16,000 g for 10 min at 4°C and samples were diluted in 2.6 mL of ChIP IP buffer (16.7 mM Tris-HCl pH 8.0, 1.2 mM EDTA pH 8.0, 167 mM NaCl, 0.01% SDS, 1.1% Triton-X-100, 0.5 mM PMSF, 1X Complete EDTA-free protease inhibitors, Sigma-Aldrich). Samples were precleared by incubation with 20 µl of protein A dynabeads (Invitrogen) blocked with BSA (B8894, Sigma-Aldrich) for 1 h at 16 rpm on a rotating wheel at 4°C. Beads were removed by incubation on the magnet for 3 min and 100 µl were taken as 'Input' samples and stored at - 20°C. Pre-cleared chromatin was then incubated overnight at 4°C rotating at 16 rpm with selected antibody or without antibody addition 'No Antibody' (for control sample). Following overnight incubation, samples were

incubated with 25 µl BSA-blocked protein A Dynabeads to collect the immuno-complexes for 2 h rotating at 16 rpm at 4°C. Beads were washed with 1.6 mL of the following buffers: once with buffer A (20 mM Tris-HCl pH 8.0, 2 mM EDTA, 0.1% SDS, 1% Triton X-100 and 0.15 M NaCl), once with buffer B (20 mM Tris-HCl pH 8.0, 2 mM EDTA, 0.1% SDS, 1% Triton X-100 and 0.5 M NaCl), once with buffer C (10 mM Tris-HCl pH 8.0, 1 mM EDTA, 1% NP-40, 1% Sodium Deoxycholate and 0.25 M LiCl) and twice with buffer D (10 mM Tris-HCl pH 8.0 and 1 mM EDTA). Immuno-precipitated complexes were eluted twice by addition of 250 µl of Elution Buffer (1% SDS and 0.1 M $NaHCO_3$) and rotation at 16 rpm at room temperature. Samples, alongside Inputs, were de-crosslinked by incubation at 65°C for at least 4 h in the presence of 0.016 µg/µl RNase A (PureLink, Invitrogen) and 0.3 M NaCl and then digested with 0.04 µg/µl proteinase K (Sigma Aldrich) for 2 h at 45°C.

DNA was purified with QIAquick PCR purification kit (QIAGEN) following manufacturer's instructions. DNA was then eluted with 60-100 µl of ddH_2O and used as a template for analysis by qPCR with Rotor-Gene® Q (QIAGEN) using QuantiTect SYBR green. The amount of immuno-precipitated material at a particular gene region was calculated as the percentage of Input after subtracting the background signal (no antibody control). The complete list of primers used in this study is shown in Table 2.2. 2-5 µg of the following antibodies were used for ChIP immunoprecipitation: Pol II antibody (Santa Cruz, H-224), H3 (Abcam Plc, ab1791), acH3K9 (Abcam Plc, ab4441), H3K9me2 (Abcam Plc, ab1220), H3K9me3 (Abcam Plc, ab8898), H3K27me3 (Abcam Plc, ab6002), H3K36me3 (Abcam Plc, ab9050), HTT (MAB2170) and DHX9 (Abcam Plc, ab26271). The complete list of antibodies used in this study is shown inTable 2.4.

2.4.8 R-loop Immuno-precipitation (R-loop IP)

R-loop immuno-precipitation (R-loop IP) was performed as described previously (Cristini *et al.*, 2018; Beghe & Gromak, 2022). HeLa cells were split one day before the experiment. Non-crosslinked HeLa cells at 85% confluency were harvested and washed twice with ice-cold PBS (Gibco). Cells were lysed in 600 µl of Cell Lysis Buffer (85 mM KCl, 5 mM PIPES pH 8.0, 0.5% NP-40, 0.5 mM PMSF, 0.8 µg/µl pepstatin A, 1 µg/µl leupeptin) for 10 min on ice, followed by

centrifugation at 500 g for 5 min to collect the nuclei. Nuclear lysis in 600 µl of ChIP Cell Lysis Buffer was performed twice to ensure better nuclei isolation. The nuclei were then resuspended in 300 µl of RSB+DSST buffer (10 mM Tris pH 7.5, 200 mM NaCl, 2.5mM $MgCl_2$, 0.5% Triton X-100, 0.2% NaDOC, 0.1% SDS, 0.05% NaSarkosyl). Following 10 min incubation on ice, extracts were sonicated for 10 min at medium power (cycles of 30 seconds "ON" and 30 seconds "OFF", Diagenode Bioruptor). Extracts were then diluted 1:4 in RSB+T buffer (10 mM Tris pH 7.5, 200 mM NaCl, 2.5 mM $MgCl_2$, 0.5% Triton X-100). 100 µl fraction of the diluted extracts was removed and stored as 'Protein Input'. The remaining diluted extracts were subjected to immuno-precipitation using 100 µl (~2 µg) of S9.6 antibody (Gromak Lab, home-made), bound to protein A Dynabeads (Invitrogen), pre-blocked with 0.5% BSA in PBS for 2 h on a rotating wheel at 4°C. As a control, 2 µg of mouse-IgG antibody (M5409, Sigma) was used instead. RNase A (PureLink, Invitrogen) was added during immuno-precipitation at 0.1 ng RNase A/µg genomic DNA. Samples were incubated at 4°C for 2 h at 16 rpm on a rotating wheel. Following incubation for 2 h, RNA/DNA hybrids and associated proteins were collected on the magnet and washed four times with RSB+T buffer and twice with RSB buffer (10 mM TRIS pH 7.5, 200 mM NaCl, 2.5mM $MgCl_2$) for 5 min at 4°C at 16 rpm on a rotating wheel. Washed immuno-complexes were analysed by SDS-PAGE and RNA/DNA hybrid Slot Blot. Samples were eluted using 40 µl of 2x LDS (Invitrogen) and 100 mM of 1M DTT by boiling for 10 min at 70°C (for SDS-PAGE) or using 500 µl of 1% SDS and 0.1 M $NaHCO_3$ for 30 min at room temperature at 16 rpm on a rotating wheel (for RNA/DNA hybrid Slot Blot).

2.5 R-loop analysis methods

2.5.1 RNA/DNA hybrid immuno-precipitation (DRIP)

RNA/DNA hybrids immunoprecipitation was performed as described previously (Cristini *et al.*, 2022), using the S9.6 antibody (Boguslawski *et al.*, 1986). Non-crosslinked cells were harvested and washed twice in ice-cold PBS (Gibco), resuspended in 800 µl of Cell Lysis Buffer (5 mM PIPES pH 8.0, 85 mM KCl, 0.5% NP-40) and incubated on ice for 10 min. Nuclei were collected by centrifugation for 5 min at 500 g at 4°C and lysed in Nuclear Lysis Buffer (50 mM Tris-HCl pH 8.0, 5 mM EDTA, 1% SDS). After incubation on ice for 10 min, samples were subjected to Proteinase K

treatment (Sigma Aldrich) for a total of 3 h at 55ºC. The initial addition of 5 µl Proteinase K was followed by the addition of 3 µl and 2 µl of Proteinase K after 1 h and 2 h of incubation, respectively. After digestion, 200 µl of 5 M KAc were added and protein and cell debris were removed by centrifugation for 5 min at 16,000 g at 4°C. Genomic nucleic acids were precipitated with 800 µl of isopropanol and incubated at -20°C overnight. DNA was then pelleted for 10 min at 16,000 g at 4°C. DNA pellets were washed in 75% ethanol, air-dried and resuspended in 100 µl of RNA-DNase-free water. 20 µg of DNA was diluted in 400 µl of IP dilution buffer (16.7 mM Tris-HCl pH 8.0, 1.2 mM EDTA, 167 mM NaCl, 0.01% SDS, 1.1% Triton X-100) and sonicated for 5 min at medium power (cycles of 30 seconds "ON" and 30 seconds "OFF", Diagenode Bioruptor) to an average length of 500 bp. After sonication, samples were centrifuged for 10 min at 16,000 g at 4°C and diluted in 2.6 mL of IP dilution buffer with addition of protease inhibitors (0.5 mM PMSF, 0.8 µg/ml pepstatin A, 1 µg/ml leupeptin) and pre-cleared with protein A Dynabeads (Invitrogen), blocked with BSA (B8894, Sigma Aldrich) rotating at 16 rpm for 1 h at 4°C. Beads were then removed by incubation on the magnet for 3 min and 100 µl were taken as 'Input' samples and stored at - 20°C. 15 µg of precleared genomic DNA was subjected to overnight immuno-precipitation at 4°C with 14,44 µl of S9.6 (Gromak Lab, home-made) or without antibody addition 'No Antibody' (control sample). Incubation with beads, washes and elution were performed as described for ChIP. Finally, samples were incubated with 2 µl of Proteinase K (Sigma Aldrich) at 45°C for 2 h.

For qPCR analysis, DNA was purified with QIAquick PCR purification kit (QIAGEN) and analysed by qPCR with Rotor-Gene® Q and QuantiTect SYBR green (QIAGEN). The amount of immuno-precipitated material at a particular gene region was calculated as the percentage of input after subtracting the background signal (no antibody control). Where stated, the values were normalized to the indicated probes. The primers used for DRIP are listed in Table 2.2.

For DRIP-sequencing analysis, multiple S9.6 IPs were pooled. DNA was purified with MinElute® PCR purification Kit (QIAGEN) and subjected to library preparation and sequencing on a NovaSeq 6000 with 150 bp paired-end reads at Oxford Genomics Centre (WTCHG, University of Oxford).

2.5.2 RNA/DNA hybrid Slot Blot

RNA/DNA hybrid Slot Blot was performed as previously described (Giordano *et al.*, 2022) Briefly, R-loops were purified as described in the DRIP protocol (Section 2.5.1), 1 µg of purified genomic DNA containing R-loops was diluted in 500 µl of 1X TE buffer (10 mM Tris-HCl pH 8.0, 1 mM EDTA pH 8.0) and spotted on Hybond-N+ membrane (Amersham Biosciences) using a Bio-Dot SF apparatus (Bio-Rad Laboratories). RNase H sensitivity was carried out by incubation with 1.7 U of RNase H (NEB, M0297) per µg of genomic DNA for 2.5 h shaking at 500 rpm at 37°C (prolonged to 4.5 h in an attempt to completely lose S9.6 signal in MSN cells). The membrane was cross-linked 2.5 min under UV light (Uvitec) and blocked in PBS (Gibco) containing 2% dried skimmed milk for 1 h at room temperature. The membrane was then probed with S9.6 antibody diluted 1:1000 in PBS (Gibco) containing 2% dried skimmed milk. For loading control, 250 ng of genomic DNA was heated at 95°C for 10 min and loaded on the Slot Blot. The membrane was denatured in 0.5 M NaOH and 1.5 M NaCl for 5 min and neutralized for 2 min in 0.5 M Tris-HCl, pH 7.2, and 1.5 M NaCl. The membrane was probed with anti-ssDNA (MAB3034; Millipore) after UV crosslinking and blocking in PBS (Gibco) containing 2% dried skimmed milk for 1 h at room temperature. Images were acquired with LAS-4000 (Fujifilm) or by chemiluminescence using autoradiography. S9.6 and ssDNA signals were quantified using Image Studio Lite software (Ver. 5.2, Li-COR Biosciences). Slot Blot analysis was also used to detect dsRNA with J2 antibody (Sciscon), following the same protocol described above after RNA extraction as described in Section 2.2.1.

2.6 Bioinformatic analysis

2.6.1 Statistical analysis

Unless otherwise stated, experimental differences for Western Blot analysis, DRIP, ChIP and RT-qPCR were tested for significance with a two-tailed unpaired t-test. One-way ANOVA with correction for multiple comparisons (Tukey test) was used to calculate statistical significance in IF experiments. For sequencing data, significance was calculated by using the Wilcoxon rank sum test, Kruskal-Wallis test and paired Wilcoxon signed rank test. Statistical tests were performed using

GraphPad Prism 9.4 software. Information on biological replicates (n) is indicated in the figure legends. Values are expressed as mean ± SEM. Image Studio Lite (Ver. 5.2, Li-Cor Bioscience) was used to quantify band density. Asterisks indicate statistical significance (* $p<0.05$, ** $p<0.01$, *** $p<0.001$, **** $p<0.001$).

2.6.2 Gene annotation

The GENCODE V41 annotation, based on the hg38 version of the human genome, and the GENCODE M30 annotation, based on the GRCm39 version of the mouse genome, were used to extract the list of protein-coding genes.

2.6.3 DRIP-seq data processing

Adapters were trimmed with Cutadapt version 3.5 in paired-end mode with the following parameters: -q 15, 10 -minimum-length 10 -A AGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT -a AGATCGGAAGAGCACACGTCTGAACTCCAGTCA (Martin, 2011). The rRNA reads were removed by mapping the trimmed reads to the Human ribosomal DNA complete repeating unit (GenBank: U13369.1) with STAR version 2.7.8a and the parameters --runThreadN 16 –readFilesCommand gunzip -c –k --outReadsUnmapped Fastx --limitBAMsortRAM 20000000000 --outSAMtype BAM SortedByCoordinate (Dobin *et al.*, 2013). The unmapped reads were mapped to the human hg38 reference genome with STAR and the parameters: --runThreadN 16 --readFilesCommand gunzip -c –k --alignIntronMax 1 --limitBAMsortRAM 20000000000 --outSAMtype BAM SortedByCoordinate. SAMtools version 1.14 was used to keep properly paired and mapped reads and to remove PCR duplicates (Li *et al.*, 2009). Reads mapping to the DAC Exclusion List Regions (accession: ENCSR636HFF) were removed with BEDtools version 2.30.0 (Quinlan & Hall, 2010). FPKM-normalized bigwig files were created with deepTools version 3.5.1 bamCoverage tool with the parameters -bs 10 -p max –normalizeUsing RPKM (Ramirez *et al.*, 2016).

DRIP-seq peaks were called with MACS2 version 2.2.7.1 and the parameters: callpeak -f BAMPE -g 2.91e9 -B -q 0.01 --call-summits (Zhang *et al.*, 2008). The total number of DRIP peaks for Control (150,304 peaks), 40/50 CAG (102,025 peaks), and 18/250 CAG (133,086 peaks) were

obtained by comparing each IP to its respective input. The 'unchanged' (40/50: 53,600 peaks, 18/250: 66,576 peaks), 'increased/gained' (40/50: 1,984 peaks, 18/250: 7,254 peaks), and 'decreased/lost' (40/50: 614 peaks, 18/250: 1,891 peaks) DRIP-seq peaks were obtained with the MACS2 bdgdiff tool by comparing the disease cell lines versus the control cell line. The difference in the number of peaks compared to the total number of DRIP peaks for Control and HD cell lines is due to the difference in the peak calling algorithm used (MACS2 callpeak vs MACS2 bdgdiff) and the breakdown of the initial peaks (IP vs Input) into smaller peaks when one or more parts of an initial peak are enriched differently in Control or HD cell lines.

2.6.4 ChIP-seq data processing

RNA polymerase II, H3K27ac, H3K9me3, H3K27me3, and their respective Input ChIP-seq data in six months male wild-type or HTT Q140 mice were obtained from GSE144684 (Alcala-Vida *et al.*, 2021). Adapters were trimmed with Cutadapt in single-end mode with the following parameters: -q 15, 10 –minimum-length 10 -a AGATCGGAAGAGCACACGTCTGAACTCCAGTCA. The reads were mapped to the mouse GRCm39 reference genome with STAR and the parameters: --runThreadN 16 --readFilesCommand gunzip -c –k --alignIntronMax 1 --limitBAMsortRAM 20000000000 --outSAMtype BAM SortedByCoordinate. SAMtools was used to keep properly mapped reads and to remove PCR duplicates. Reads mapping to the mouse DAC Exclusion List Regions (accession: ENCSR636HFF, the DAC mm10 version has been converted for the GRCm39 genome with the UCSC liftOver tool) were removed with BEDtools. FPKM-normalized bigwig files were created with deepTools bamCoverage tool with the parameters -bs 10 -p max –normalizeUsing RPKM -e 150.

2.6.5 Meta profiles

Metagene profiles were generated from FPKM-normalized bigwig files with deepTools compute. Matrix tool with a bin size of 10 bp and the plotting data obtained with plotProfile –outFileNameData tool. Graphs were then created with GraphPad Prism 9.4.

2.7 Tables

Table 2.1 List of cell lines

The table shows the cell lines used in this work. Cell lines indicated by * and by ** are originated from the same patient. Cell lines indicated by # have been generated by the Tabrizi lab. N.D., not determined.

Cell line	Genotype	Expansion	Gender
Lymphoblasts			
GM04844 (Control 1)*	WT	16/20	Female
GM06895 (Control 2)	WT	N.D.	Male
GM14926 (Control 3)	WT	N.D.	Male
GM15851 (Control 4)	WT	N.D.	Male
GM04780 (Control 5)	WT	N.D.	Male
GM04205 (Control 6)	WT	18/17	Male
GM121878 (Control 7)	WT	N.D.	Female
GM04856 (HD 1) **	Homozygous	40/50	Female
GM14044 (HD 2)	Heterozygous	18/250	Male
GM04868 (HD 3)	Heterozygous	13/42	Female
GM04798 (HD 4)	Heterozygous	N.D./47	Male
GM05538 (HD 5)	Heterozygous	N.D./68	Male
Fibroblasts			
GM04845 (Control) *	WT	16/20	Female
GM04857 **	Homozygous	40/50	Female
GM09197	Heterozygous	18/173	Male
Medium Spiny Neurons[#]			
Control	WT	N.D.	
125Q	Heterozygous	N.D./130	
Astrocytes[#]			
Control	WT	N.D.	
56Q	Heterozygous	N.D./56	
125Q	Heterozygous	N.D./145	
iPSCs	WT	N.D.	
HeLa	WT	N.D.	Female
SH-SY5Y	WT	N.D.	Female

Table 2.2 List of DNA oligonucleotides

The table indicates the sequences of DNA oligonucleotides used in this research. Sequences are given in 5' to 3' direction.

Name	Sequence (5' to 3')
Gene Expression, ChIP and DRIP	
ACTB	
prom (F)	CCGAAAGTTGCCTTTTATGGC
prom (R)	CAAAGGCGAGGCTCTGTGC
in1 (F)	CGGGGTCTTTGTCTGAGC
in1 (R)	CAGTTAGCGCCCAAAGGAC
in5 (F)	GGAGCTGTCACATCCAGGGTC
in5 (R)	TGCTGATCCACATCTGCTGG
5'pause (F)	TTACCCAGAGTGCAGGTGTG
5'pause (R)	CCCCAATAAGCAGGAACAGA
γActin	
prom (F)	GGAAAGATCGCCATATATGGAC
prom (R)	TCACCGGCAGAGAAACGCGAC
in1 (F)	CCGCAGTGCAGACTTCCGAG
in1 (R)	CGGGCGCGTCTGTAACACGG
DDX1	
up (F)	AATTTGGGGATCCTGGCAGA
up (R)	TTGAGTGTTCGTCAGCCTCT
prom (F)	GGCCTTTCCTTCCATCGGA
prom (R)	CACAAACGCACCGGAGAAG
***FXN** (F)*	AAACTGACCCGACCTTTATTCCA
***FXN** (R)*	GGAATCCCCCAAGGTCACA
***GAPDH** (F)*	ACATCAAGAAGGTGGTGAAG
***GAPDH** (R)*	GGGTCTTACTCCTTGGAGGC
HTT	
up2 (F)	GGCTGTCCGGGTGAGTATGG
up2 (R)	CCATGCGGAAAGGATCCCCC
up1 (F)	TCGCCTCACCCCATTACAGTC
up1 (R	GGTTCTGCCTCACACAGCAAG
prom (F)	GCCGCTCAGGTTCTGCTTTT
prom (R)	TCAGCTTTTCCAGGGTCGCC
ex1-in1 (F)	CCGGCTGTGGCTGAGGAG
ex1-in1 (R)	TCAGCTTTTCCAGGGTCGCC
in13kb (F)	AGTCTCACATTGTCACCCAGGC
in13kb (R)	ATTAGCTGGACATGGTGGT
in17kb (F)	GCATTTGGGAGGAAGACCTGTG
in17kb (R)	ACAGAAATGCTCAACGGCCC
B (F)	TTGGTGAACTCGTGCCCTGT
B (R)	TCATTCAGTGAGTGTTTCTCGACT
D (F)	AGTGGTTTGCTGGGTTGAGG
D (R)	GGCTCTCACCACCTTGCATC
E (F)	GGCTCCTCTTGACAGTTTCTCTTC
E (R)	CTGCGTCATCACTGCACAGC
F (F)	GGGATGTGTCTTCCATAAATCTC
F (R)	GCTCGAGCTGTAACCTTGGAA
ex1-ex2 (F)	CCGGCTGTGGCTGAGGAG
ex-ex2 (R)	ACACGGTCTTTCTTGGTAGCTGA
ex66-ex67 (F)	GGTCATGCTGTCCCTCTCC
ex66-ex67 (R)	AGCTTGCCCATCCTGCTG

***PTB** (F)*	GCCGTTGGTACAAAGGTAGG
***PTB** (R)*	GCCCCTTAGGAATGGAAAAG
***Gemin7** (F)*	TCTTCTTCCACCTGGACCAC
***Gemin7 (**R)*	GGGACAGAGAGAGTGCCTTG
***TFF1** prom (F)*	TATGAATCACTTCTGCAGTGAG
***TFF1** prom (R)*	GAGCGTTAGATAACATTTGCC
Immunity genes	
ISG15 (F)	CGCAGATCACCCAGAAGATCG
ISG15 (R)	TTCGTCGCATTTGTCCACCA
ISG20 (F)	TAGCCGCTCATGTCCTCTTT
ISG20 (R)	TGAGGGAGAGATCACCGATT
IFIT (F)	GCCTTGCTGAAGTGTGGAGGAA
IFIT (R)	ATCCAGGCGATAGGCAGAGATC
IFNGR1 (F)	GCCACAGGTCCCTGTTTTTA
IFNGR1 (R)	TCCAACCCTGGCTTTAACTC
NFkB (F)	TAGCCACAGAGATGGAGGAG
NFkB (R)	CCGAGTCGCTATCAGAGGTA
OAS1 (F)	CGCCTAGTCAAGCACTGGTA
OAS1 (R)	CAGGAGCTCCAGGGCATA
p21 (F)	CAGCATGACAGATTTCTACCACTC
p21 (R)	CTCGCGCTTCCAGGACTG
STING (F)	CATGGGCTGGCATGGTCATA
STING (R)	ATATACAGCCGGCTGGCTCAC
TNFα (F)	CCTCTCTCTAATCAGCCCTCTG
TNFα (R)	GAGGACCTGGGAGTAGATGAG
CCL5 (F)	ACCACTCCCTGCTGCTTTGC
CCL5 (R)	CACACTTGGCGGTTCCTTGG
DRIP-qPCR validation	
SFSF2 (F)	CTTTCACGACAAGCGCGAC
SFSF2 (R)	GTGGTGTGAGTCCGGGG
HIST1H1B (F)	TTCGCCTTCTTCGGAGTCTT
HIST1H1B (R)	AAACTCAACAAGAAGGCGGC
RNU2 (F)	GCCGGCTTGACTTAACTGCT
RNU2 (R)	TAGCCACGGTGGAAACGAAA
DOCK (F)	GGTGGTGGTGGTAATACTGC
DOCK (R)	GACCCCACCCCACGTAAAAA
FGF3 (F)	CATCCACCCACCCATCTGAC
FGF3 (R)	GGGAGTCCTCAGAAGCCATG
NRXN3 (F)	TTCTCTGCCAACCTCCGTTC
NRXN3 (R)	CACCACCACCACCAGGAAT
MSNs differentiation	
CTIP2 (F)	CTCCGAGCTCAGGAAAGTGTC
CTIP2 (R)	TCATCTTTACCTGCAATGTTCTCC
DRD2 (F)	CTGAGGGCTCCACTAAAGGAG
DRD2 (R)	CATTCTTCTCTGGTTTGGCG
Neurogenin (F)	AGGAAGAGGACGTGTTAGTGC
Neurogenin (R)	GCAATCGTGTACCAGACCCAG
Oct4 (F)	GTGTTCAGCCAAAAGACCATCT
Oct4 (R)	GGCCTGCATGAGGGTTTCT
Pax6 (F)	TGGGCAGGTATTACGAGGACTG
Pax6 (R)	ACTCCCGCTTATACTGGGCTA

Table 2.3 List of RNA oligonucleotides

The table indicates RNA oligonucleotides (Horizon) used for siRNA experiments in HeLa cells.

Target gene	**Sequence (5' to 3')**
siLuciferase (siLuc)	CTTACGCTGAGTACTTCGA
HTT	AACAUAGUGGCACAGUCUGUC
XPG	GAAAGAAGAUGCUAAACGU

Table 2.4 List of antibodies

The table indicates the targets of all antibodies employed in this work. Dilutions for Immunofluorescence, Western Blot and Slot Blot analysis are given in column 'IF/WB/SB dilution'. The amount of antibody used per sample for ChIP and IP experiments is also indicated in the respective columns.

Target	Catalogue n°	ChIP	IP	IF/WB/SB dilution	Supplier
ADAR1	sc271854			1:200	Santa Cruz
β-tubulin III	RB-435P			1:2,000 (IF)	Babco
CTIP2	25B6			1:200	Abcam Plc.
DHX9	ab26271			1:2,000	Abcam Plc.
Histone H3, total	ab1791	5 µg		1:5,000	Abcam Plc.
Histone H3K9me2	ab1220	5 µg		-	Abcam Plc.
Histone H3K9me3	ab8898	5 µg		-	Abcam Plc.
Histone acH3K9	ab4441	5 µg		-	Abcam Plc.
Histone H3K27me3	ab6002	5 µg		-	Abcam Plc.
Histone H3K36me3	ab9050	5 µg		-	Abcam Plc.
HTT	MAB2166			1:3,333	Sigma-Aldrich
HTT	MAB2170		3 µg		Sigma-Aldrich
IgG	M5409		2/3 µg		Sigma-Aldrich
Lamin B1	ab16048			1:2,000	Abcam Plc.
J2	10010200		3 µg	1:2,000	Sciscon
Nucleolin	ab22758			1:1,000	Abcam Plc.
PCNA	ab152112			1:5,000	Abcam Plc.
PKR	sc6282			1:200	Santa Cruz
Pol II clone 8WG16	ab817			1:1,000	Abcam Plc.
Pol II-S5P	ab5131			1:2,000	Abcam Plc.
Pol II-S2P	ab5095			1:2,000	Abcam Plc.
Pol II	NBP2-32080	1 µg			Novus
RNA/DNA hybrids	S9.6 (hybridoma line)		2 µg	1:1,000	Gromak Lab
RNA/DNA hybrids	S9.6			1:200 (IF)	Kerafast
Top 1	ab109374			1:500	Abcam Plc.
Tubulin	ab4074			1:5,000	Abcam Plc.
XPF	Ab76948			1:500	Abcam Plc.
XPG	11331-1-AP			1:500	Proteintech
γH2AX	NB100-384			1:1,000	Novus
γH2AX	05-636			1:200 (IF)	Sigma-Aldrich
αssDNA	MAB3034			1:5,000	Millipore

Chapter 3

Transcriptional regulation and epigenetic signature of the *HTT* gene

3.1 The *HTT* gene transcriptional regulation in health and disease

The CAG repeat expansion in the *HTT* gene was identified as the genetic origin of HD in 1993 (MacDonald *et al*., 1993). However, the mechanisms underlying the transcriptional regulation of the *HTT* gene itself are still not fully understood. Therefore, the first objective of the present work was to investigate the transcriptional regulation of the *HTT* gene initially in healthy cells, and then in cells derived from Control and HD patients (Table 2.1). The characterisation of the factors that influence *HTT* gene expression will contribute to the understanding of the HTT protein's function in physiological conditions and in HD pathogenesis, and provide insights into the potential mechanisms affected by expanded CAG repeats.

3.1.1 *HTT* expression varies in different cell lines

3.1.1.1 The *HTT* gene expression in HeLa and SH-SY5Y cells

The investigation of *HTT* expression in various cell lines is crucial to identify a cellular model suitable to study the molecular mechanisms involved in HD disease. The gain-of-function of mutant HTT protein is a major driver of HD. However, growing evidence suggests that the loss of wild-type HTT protein also contributes to HD (Zuccato & Cattaneo, 2009). To study the potential contribution of loss-of-function of the wild-type *HTT* RNA and protein (see Chapter 5), I decided to take advantage of HeLa and SH-SY5Y cellular models. HeLa is an immortalized cell line derived from cervical cancer cells, whereas SH-SY5Y is a neuroblastoma cell line often used as a neuronal model for the study of neurodegenerative disorders. To investigate the *HTT* expression levels in HeLa and SH-SY5Y cells, quantitative reverse transcription-PCR (RT-qPCR) was performed with primers designed to detect both nascent (amplicons 'ex1-in1' and 'in2-ex3') and spliced (amplicons 'ex1-ex2' and 'ex66-ex67') *HTT* RNA levels (Fig. 3.1A). Notably, higher levels of both nascent and spliced *HTT* RNA were detected in SH-SY5Y cells compared to HeLa cells (Fig. 3.1B), suggesting that the levels of *HTT* RNA vary in cell lines of different origin.

3.1.1.2 Generation of Medium Spiny Neurons (MSNs) from iPS cells

Despite HTT protein being ubiquitously expressed throughout the body, the disease is restricted to specific neuronal populations within the brain of HD patients. Specifically, at the onset

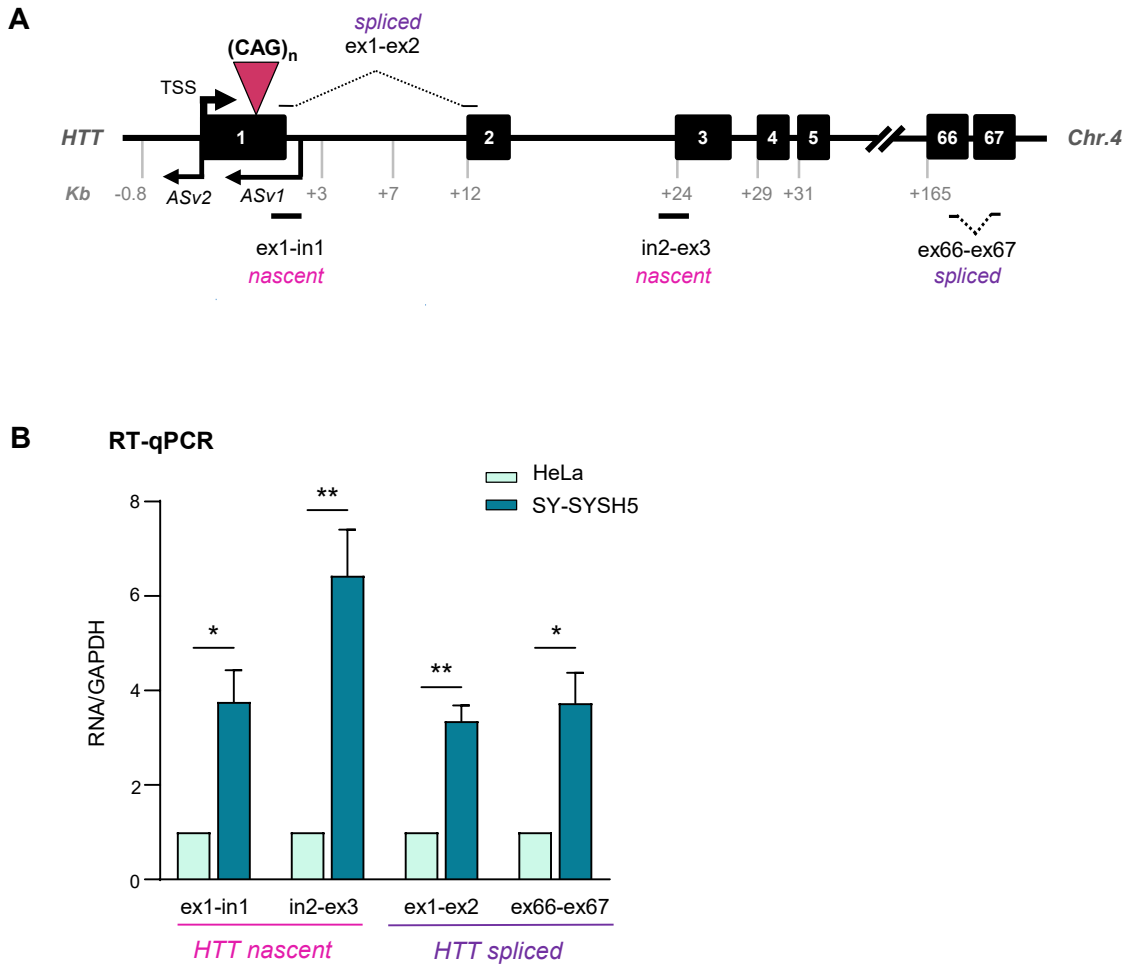


Figure 3.1. Analysis of *HTT* gene expression in different cells.

A. Diagram of the *HTT* gene. Black boxes are exons, lines are introns, and the red triangle is the $(CAG)_n$ expansion. Amplicons for RT-qPCR analysis are shown below and above the gene diagram. Numbers indicate sizes in kilobase (kb) from the Transcriptional Start Site (TSS). *ASv1* and *ASv2* indicate *HTT* antisense transcripts.

B. RT-qPCR analysis for *HTT* gene in HeLa and SH-SY5Y. Values are relative to HeLa cells and normalized to mRNA *GAPDH* (n=3; means ± SEM, * $p > 0.05$, ** $p > 0.01$ by two-tailed unpaired t-test).

of HD, there is a selective and gradual degeneration of medium spiny neurons (MSNs) in the striatum of HD patients. As the disease progresses to later stages, the neurodegeneration extends to other regions of the brain, including the cerebral cortex (Novak & Tabrizi, 2010). Induced Pluripotent Stem Cells (iPSCs) can be differentiated in specific neuronal lineages for disease modelling (Kaye *et al.*, 2022). Therefore, to study the physiologically relevant neurological context of HD, a protocol was optimized to differentiate iPSCs into MSNs, the cells preferentially lost in HD patients. Differentiation of lateral ganglionic eminence (LGE) progenitors and MSNs from iPSCs was optimized using a recently-published protocol (Fjodorova & Li, 2018). The experimental scheme of the differentiation timeline can be found in Fig. 3.2A (top panel). First, iPSCs were differentiated into forebrain progenitors with a cocktail of SMAD inhibitors (Day 0-9) and then patterned toward an LGE fate with Transforming Growth Factor- β (TGF- β) family protein Activin A (Day 9-21). Using this protocol, LGE progenitors can be replated and terminally differentiated into MSNs (Day 22-35). Changes in cell morphology have been monitored during each stage of the protocol and representative bright field images of iPSCs at Day 0 and MSNs at Day 35 are shown in Fig. 3.2A (bottom panel). Moreover, the neuronal identity of the differentiated cells was investigated via RT-qPCR analysis and immuno-fluorescence (IF). Neuronal induction resulted in down-regulation of the pluripotency gene *Oct4* and up-regulation of the neural markers *Pax6* and *Neurogenin*, observed by gene expression analysis (Fig. 3.2B). Furthermore, mature MSN commitment was confirmed by the expression of the chicken ovalbumin upstream promoter transcription factor (COUP-TF)-interacting protein 2 (*CTIP2*) and the dopamine-receptor 2 (*DRD2*), crucial for MSN and striatal development (Fig. 3.2B). In addition, IF experiments showed that LGEs at Day 21 expressed CTIP2 (Fig. 3.2C).

HTT gene expression levels were also evaluated in these cells. iPSCs exhibited higher expression of *HTT* compared to their neuronal counterparts (LGEs on Day 9 and Day 21 and MSNs on Day 35) at messenger RNA (mRNA) level (Fig. 3.2D). This is in line with recent studies showing that human embryonic stem cells (hESCs) exhibit higher *HTT* mRNA levels than their differentiated neuronal counterparts (Irmak *et al.*, 2018). Neuronal cells may therefore be susceptible to neurodegeneration in Huntington's disease due to naturally lower levels of wild-type *HTT* mRNA

compared to other cells. This relatively low expression of *HTT* may be further impacted by the presence of *HTT* mutations, resulting in specific HD phenotypes in neuronal cells.

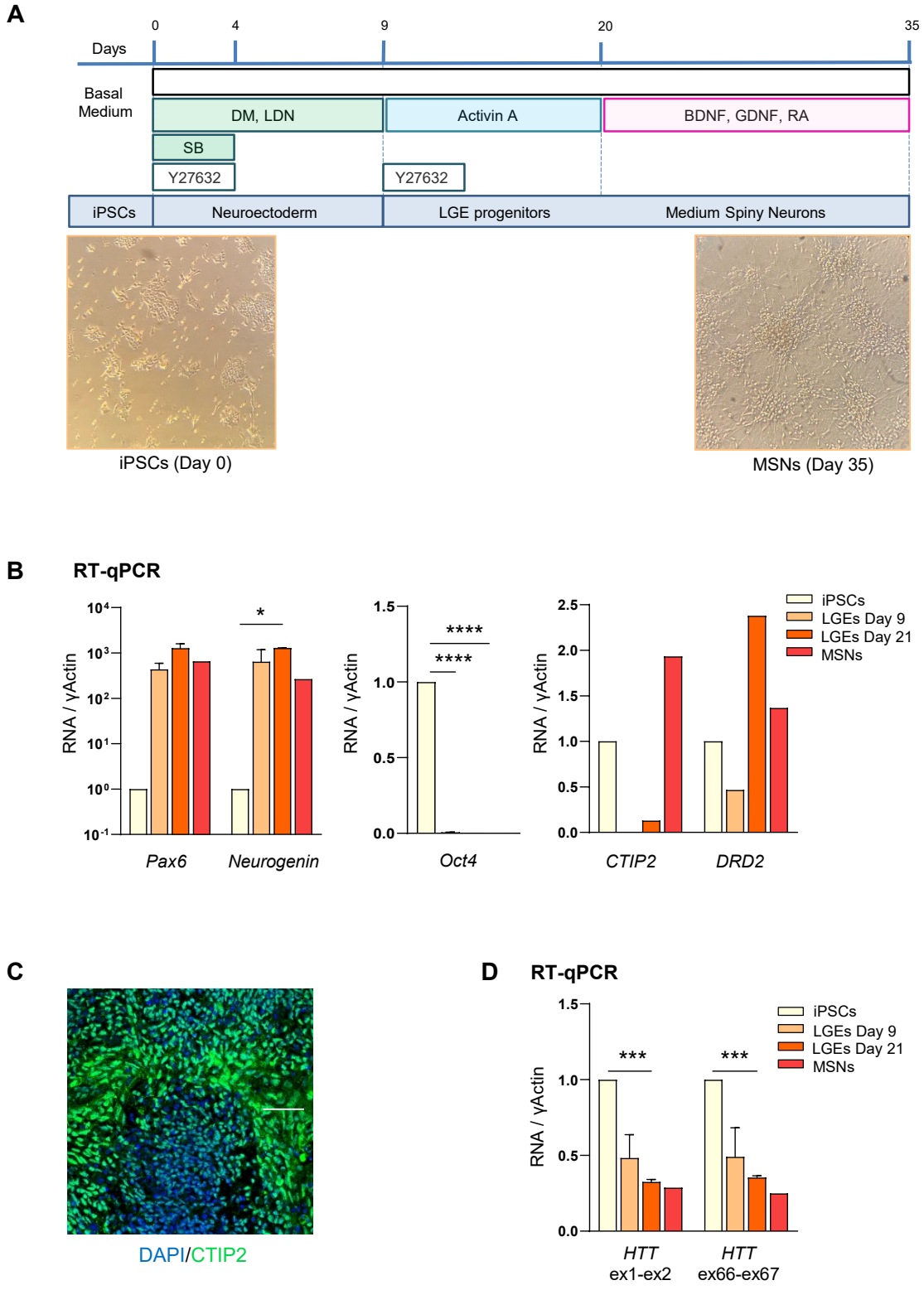


Figure 3.2. Generation and characterization of an MSN model for HD.
The caption of the figure can be found below.

Figure 3.2 (continued). Generation and characterization of an MSN model for HD.

A. **Top panel.** Overview of the experimental procedure of MSN differentiation protocol from iPSCs (methodology details are in Section 2.1.6). **Bottom panel.** Representative bright field images showing the cell morphology for iPSCs (Day 0) and MSNs (Day 35).

B. RT-qPCR analysis for *Pax6*, *Neurogenin, Oct4, CTIP2* and *DRD2* genes in iPSC, LGE and MSN cells. Values are relative to iPSCs and normalised to *γActin* (n=2; means ± SEM, * $p>0.05$, **** $p>0.0001$ by two-tailed unpaired t-test; MSNs Day 35, n=1).

C. Immunostaining for CTIP2 (green) and DAPI (blue) of LGEs derived from iPSCs at Day 21. Scale bar: 50 μm. Representative images from n = 2 biologically independent experiments. Image generated with the help of Dr. M.S. Giordano.

D. RT-qPCR analysis for the *HTT* gene in iPSC, LGE and MSN cells. Values are relative to iPSCs and normalised to *γActin* (n=2; means ± SEM, *** $p>0.001$ by two-tailed unpaired t-test; MSNs Day 35, n=1).

3.1.2 Sense and antisense transcription of the *HTT* gene

To date, there has been no comprehensive assessment of the regulation of *HTT* gene expression, and minimal attention has been given to the genomic features of the HD locus. To explore the *HTT* transcriptional profile, available global nuclear run-on (GRO-cap) genome-wide data, in lymphoblast cells from healthy individuals, was analysed (Fig. 3.3) (Core *et al.*, 2014). GRO-cap is a variation of the GRO-seq method that enriches for 5′-capped (m7G) RNAs (GRO-cap), allowing to capture transcriptional start sites (TSSs) for both stable and unstable transcripts (Core *et al.*, 2014). Multiple TSS peaks were detected at the beginning of the *HTT* gene in lymphoblast cells, most likely corresponding to different TSS/capping sites (Fig. 3.3). Interestingly, peaks corresponding to two additional TSSs in the first intron of the *HTT* gene, ~ 1 kb and ~ 3 kb downstream of the 5'-TSS, were also detected (Fig 3.3).

Antisense transcription is an important mechanism of gene expression regulation (Wight & Werner, 2013). A natural antisense transcript (*HTTAS*) was previously identified at the *HTT* locus (Chung *et al.*, 2011). This antisense transcript was shown to be alternatively spliced into two smaller transcripts (*HTTASv1* and *v2*). GRO-cap in lymphoblast cells detects the *HTTASv1* antisense transcript on the minus strand, in line with previous data (Chung *et al.*, 2011) (Fig 3.3). Furthermore, this analysis reveals the use of an alternative TSS for *HTTAS* (Fig 3.3, small peak upstream of the *HTT* promoter).

3.1.3 Transcriptional profile of the *HTT* gene in HeLa cells

To further investigate the transcriptional profile of the *HTT* gene, chromatin immunoprecipitation (ChIP) analysis using an anti-RNA Polymerase II (Pol II) antibody, raised against total Pol II, was performed in HeLa cells (Fig 3.4A-B). To analyse Pol II levels across the *HTT* gene, qPCR primers spanning the whole *HTT* gene body were designed (Fig. 3.4A and Table 2.2). Pol II ChIP-qPCR analysis showed that Pol II is enriched over the promoter-proximal region in HeLa cells (amplicon 'prom', Fig. 3.4B). This Pol II peak is conserved in most human genes and represents an early promoter-proximal Pol II pausing event (Jonkers & Lis, 2015). This peak

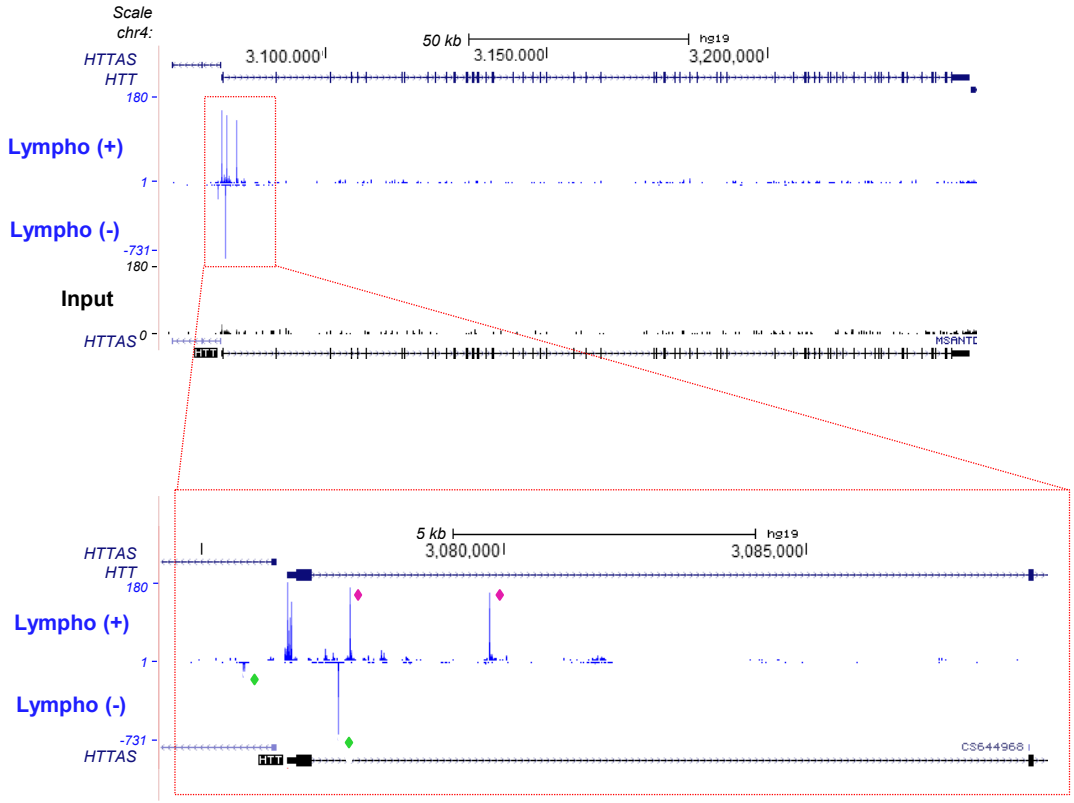


Figure 3.3. GRO-cap across the *HTT* gene in lymphoblast cells.
Top panel. Snapshot of GRO-cap read distribution across the *HTT* gene in wild-type lymphoblast cells (Control 7 cell line, Table 2.1). **Bottom panel.** Zoom on the exon1-intron1 region of the *HTT* gene. (+) and (-) represent sense and antisense RNA strands, respectively. Green diamonds indicate TSSs identified on the (-) strand. Pink diamonds indicate additional TSSs in the intron 1 of the *HTT* gene. Data are from Core *et al.*, 2014. Figure generated with the help of Dr. M. Tellier.

corresponds to the main TSS of the *HTT* gene and coincides with the location of the TSS GRO-cap peaks in lymphoblast cells (Fig. 3.3). Furthermore, the Pol II signal detected in the body of the *HTT* gene was low, a feature seen previously in other human genes (Core *et al.*, 2008). The high peak detected in the amplicon 'up2' may correspond to *HTTAS* (Fig 3.4B). As a positive control, the promoter-proximal amplicon of the *γActin* gene, which is highly transcribed, showed high levels of Pol II (Fig 3.4B).

To map active chromatin marks across the *HTT* gene, ChIP-qPCR analysis of H3K9ac and H3K36me3 marks were also performed (Fig. 3.4C-D). H3K9ac mark co-localizes with active gene promoters and it is associated with active transcription (Zhang *et al.*, 2015). In contrast, the H3K36me3 mark is usually associated with active transcription within gene bodies (Yoh *et al.*, 2008; Barski *et al.,* 2007). Consistent with the peak of Pol II profile (Fig. 3.4B), the *HTT* gene exhibited an enrichment of the H3K9ac mark over the *HTT* promoter (amplicon 'prom'), whereas low levels of this mark were detected in the gene body (Fig. 3.4C). As a positive control, the promoter-proximal amplicon of the *γActin* gene showed high level of H3K9ac mark. Enrichment of the H3K36me3 mark was detected further downstream of the *HTT* promoter in amplicons 'in1-3kb', 'in1-7kb', 'B', 'D' and 'F', as expected (Fig. 3.4D).

To investigate heterochromatin formation over the *HTT* gene, H3K9me2/me3 and H3K27me3 ChIP-qPCR were performed (Fig. 3.5). Generally, methylation of H3K9 is associated with constitutive heterochromatin formation and transcriptional silencing (Miller & Grant, 2013). Consistent with the fact that *HTT* is expressed in HeLa cells, a low level of H3K9me2/me3 repressive marks was detected over its promoter region (Fig. 3.5A-B). However, a region of higher heterochromatin was found over amplicons 'in1-3kb', 'in1-7kb' and 'B', especially for the H3K9me2 repressive mark (Fig. 3.5A-B). Moreover, H3K9me2 levels over these amplicons were significantly higher than over the promoter-proximal amplicon of the highly expressed *γActin* gene, thus confirming the specific increase of this heterochromatin mark in *HTT* gene (Fig. 3.5A). Furthermore, this heterochromatin patch coincides with a recently mapped poly A site in the Intron 1 of the *HTT* gene (Sathasivam *et al.*, 2013).

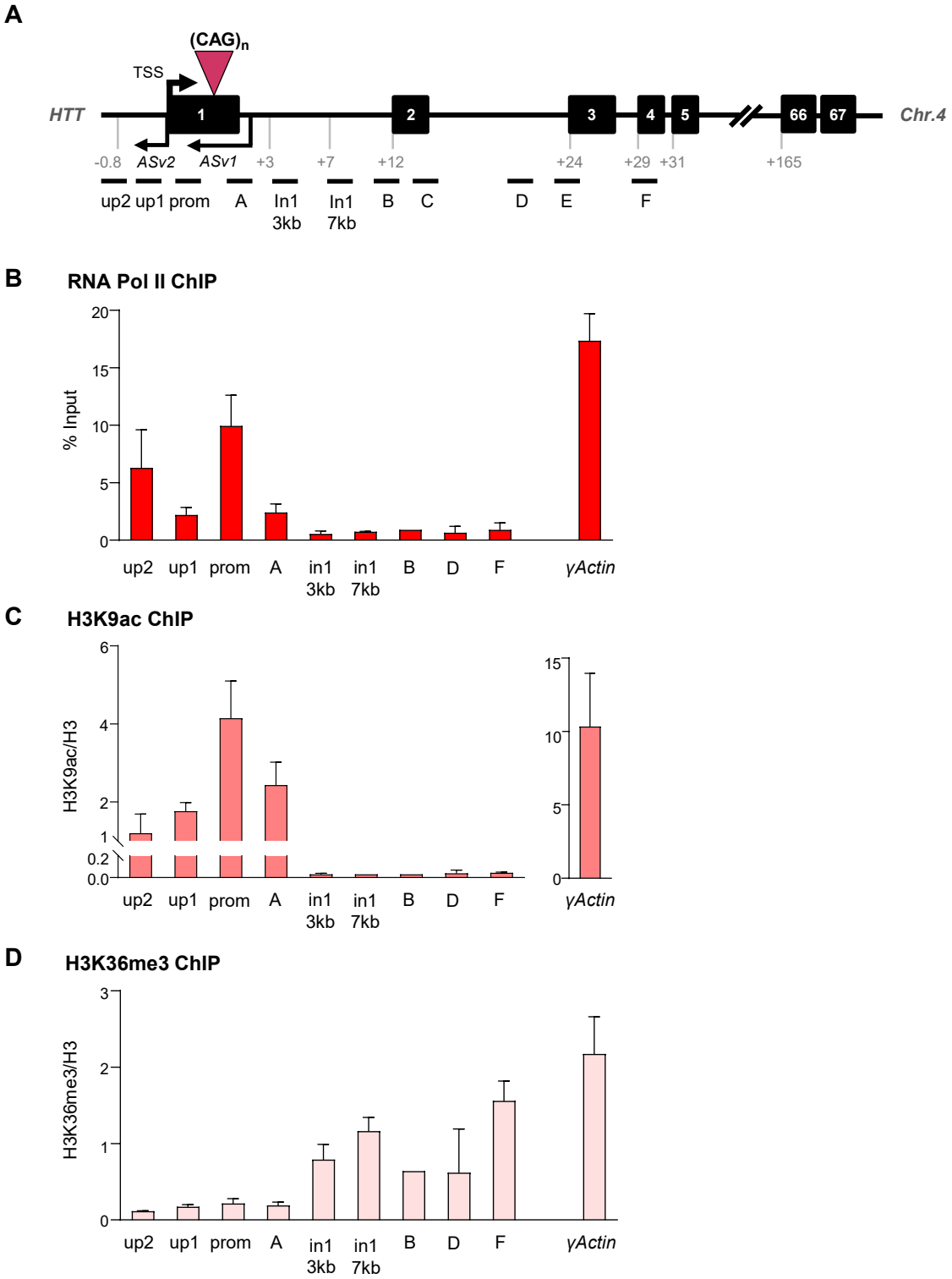


Figure 3.4. Distribution of Pol II and active chromatin marks across the *HTT* gene in HeLa cells. **(A)** Diagram of the *HTT* gene. Amplicons for ChIP-qPCR are shown below the gene diagram. Numbers indicate sizes in kilobase (kb) from the Transcriptional Start Site (TSS). *ASv1* and *ASv2* indicate *HTT* antisense transcripts. ChIP-qPCR for total Pol II **(B)**, H3K9ac **(C)** and H3K36me3 **(D)** across the *HTT* gene and the promoter-proximal amplicon of *γActin* gene in HeLa cells. For Pol II ChIP-qPCR, the values are expressed as % of Input (n=3; means ± SEM). For histone marks ChIP-qPCR, the values are normalised to the levels of total histone H3, as determined by ChIP-qPCR (n=3; means ± SEM).

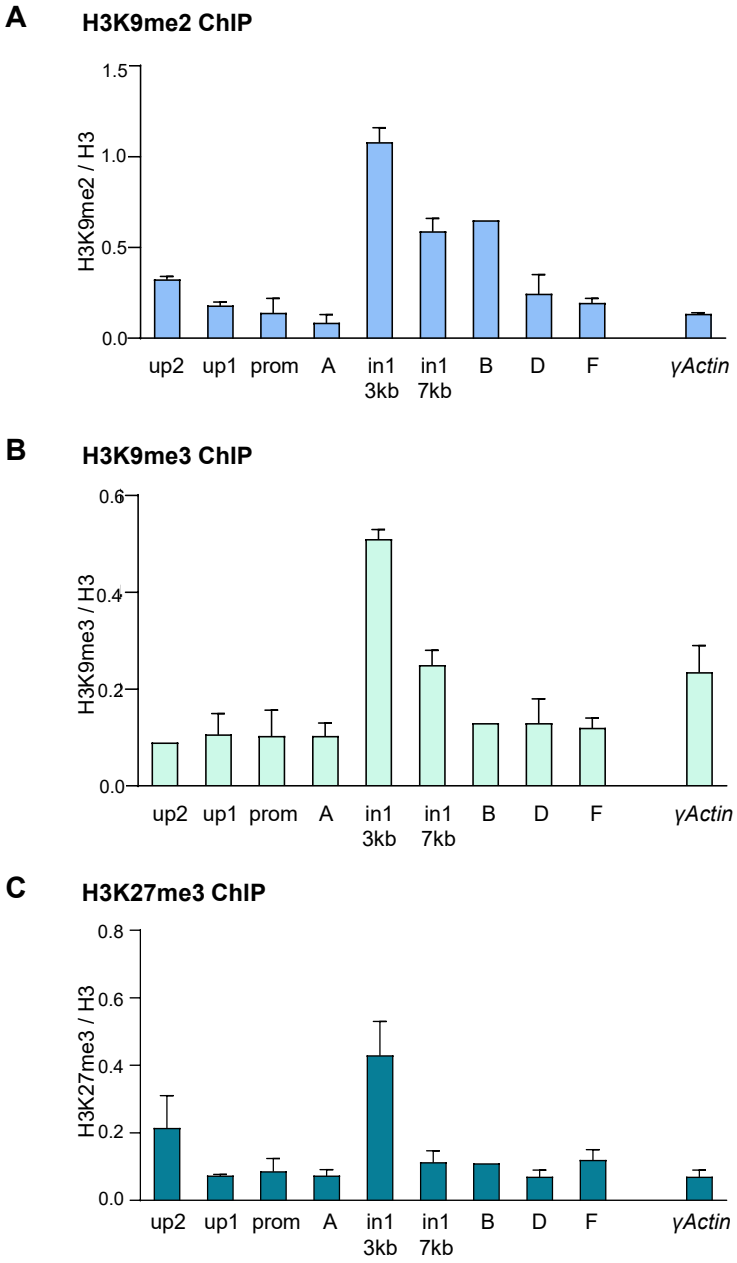


Figure 3.5. Distribution of repressive chromatin marks across the *HTT* gene in HeLa cells. ChIP-qPCR for H3K9me2 **(A)**, H3K9me3 **(B)** and H3K27me3 **(C)** across the *HTT* gene and the promoter-proximal amplicon of *γActin* gene in HeLa cells. Histone marks are normalised to the levels of total histone H3, as determined by ChIP-qPCR (n=3; means ± SEM).

In mammals, H3K27me3 is a repressive chromatin mark catalysed by proteins of the polycomb group (PcG), a group of transcriptional gene repressors that are conserved throughout evolution (Bogliotti & Ross, 2012). These PcG proteins form multiprotein complexes, including polycomb repressive complex 1 and 2 (PRC1 and PRC2), which are responsible for the establishment and maintenance of the H3K27me3 mark in mammals (Bogliotti & Ross, 2012). ChIP-qPCR analysis for H3K27me3 showed an overall low level of this mark across the *HTT* gene (Fig. 3.5C). However, the H3K27me3 mark was particularly enriched over the amplicon 'in1-3kb'. In addition, over this amplicon, the levels of the H3K27me3 mark were significantly higher than over the promoter-proximal amplicon of the *γActin* gene, thus confirming the specific increase of this heterochromatin mark in *HTT* gene (Fig. 3.5C). Interestingly, the location of this mark corresponds to the position of H3K9me2/me3 peaks described above.

3.2 HD mutation affects *HTT* gene expression

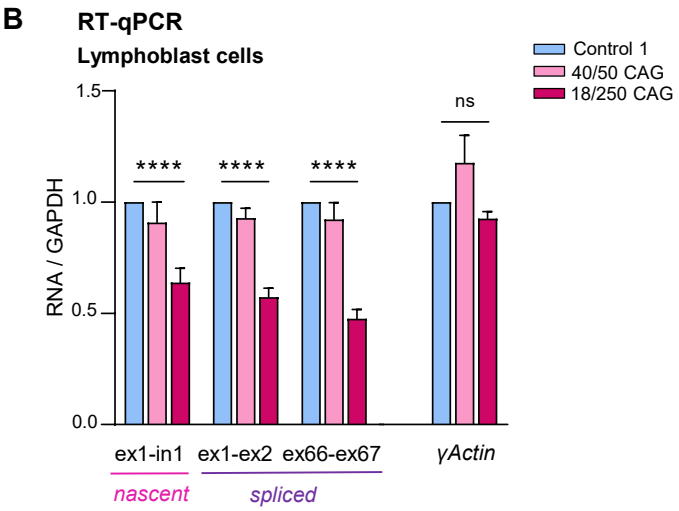
3.2.1 Nascent and mRNA *HTT* levels are reduced in HD

It has been demonstrated that transcription through expanded repeats can play an essential role in disease pathology (Groh *et al.*, 2014b). Indeed, transcription through expanded repeats can lead to dysregulation of transcription-associated processes, including RNA splicing, production of antisense, short non-coding and bidirectional RNA transcripts and cause the formation of non-canonical RNA/DNA structures, such as RNA/DNA hybrids (R-loops). Therefore, the transcription of the *HTT* gene has been analysed in Control and HD cells carrying expanded CAG repeats of different lengths.

To test if expanded CAG repeats at the *HTT* gene locus affect the transcriptional regulation of the *HTT* gene, Epstein-Barr virus (EBV)-immortalised lymphoblast and primary fibroblast cells derived from either healthy individuals or HD patients were purchased from the Coriell Institute from Medical Research (Camdem, USA) (Fig. 3.6A). Furthermore, an HD neuronal model, generated by Tabrizi lab, was analysed. In particular, medium spiny neurons (MSNs) were differentiated from Control and HD iPSCs (Fig. 3.6A). HD lines carried either homozygous or heterozygous HD

A

Cell line	Genotype	Expansion
Lymphoblasts		
GM04844 (Control 1)*	WT	16/20
GM06895 (Control 2)	WT	N.D.
GM14926 (Control 3)	WT	N.D.
GM04856 **	Homozygous	40/50
GM14044	Heterozygous	18/250
Fibroblasts		
GM04845 (Control 1) *	WT	16/20
GM04857 **	Homozygous	40/50
GM09197	Heterozygous	18/173
Medium Spiny Neurons		
Control	WT	N.D./20
125Q	Heterozygous	N.D./130



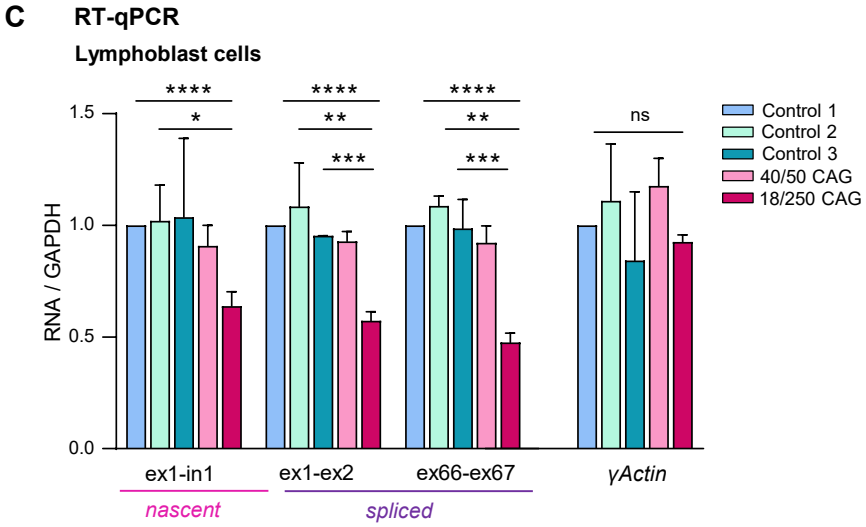


Figure 3.6. *HTT* gene expression is impaired in HD lymphoblasts.

A. Patient cell lines used in this Chapter. The repeat sizes are indicated. Cell lines indicated by * and ** are originated from the same patient. N.D., not determined.

B. RT-qPCR analysis for *HTT* and *γActin* genes in Control and HD lymphoblast cells. Values are relative to the Control 1 cell line and normalised to *GAPDH* (n=5; means ± SEM, **** $p>0.0001$ by two-tailed unpaired t-test).

C. RT-qPCR analysis for *HTT* and *γActin* genes in additional Control and HD lymphoblast cells. Values are relative to Control 1 cell line and normalized to *GAPDH* (n>3; means ± SEM, * $p>0.05$, ** $p>0.01$, ***$p>0.001$, **** $p>0.0001$ by two-tailed unpaired t-test).

mutations (Fig. 3.6A). To investigate *HTT* expression levels in these cells, RNA was extracted, and RT-qPCR analysis was performed. qPCR primers were used to detect either nascent or mRNA *HTT*levels (Fig. 3.1A). Using RT-qPCR analysis, a reduction in nascent and mRNA *HTT* levels was detected in the 18/250 CAG HD lymphoblasts compared to Control cells (Fig. 3.6B). A similar reduction in 18/250 CAG HD cells was detected when *HTT* expression was compared to two additional expression was compared to two additional independent Control lymphoblast lines, thus eliminating the possibility of patient-patient variation (Fig. 3.6C). Furthermore, a strong transcriptional decrease was observed in both 40/50 CAG and 18/250 CAG HD primary fibroblasts compared to the Control cells, suggesting that primary cells may show a stronger phenotype compared to immortalized cell lines (Fig. 3.7A). Interestingly, an even stronger reduction in nascent and mRNA *HTT* levels was observed in iPSC-derived HD MSNs, compared to healthy cells (Fig. 3.7B). In these experiments, mRNA levels of several housekeeping genes including *γActin* (promoter-proximal amplicon), *7SK* and *GAPDH* were comparable between healthy and HD cells and therefore used as controls.

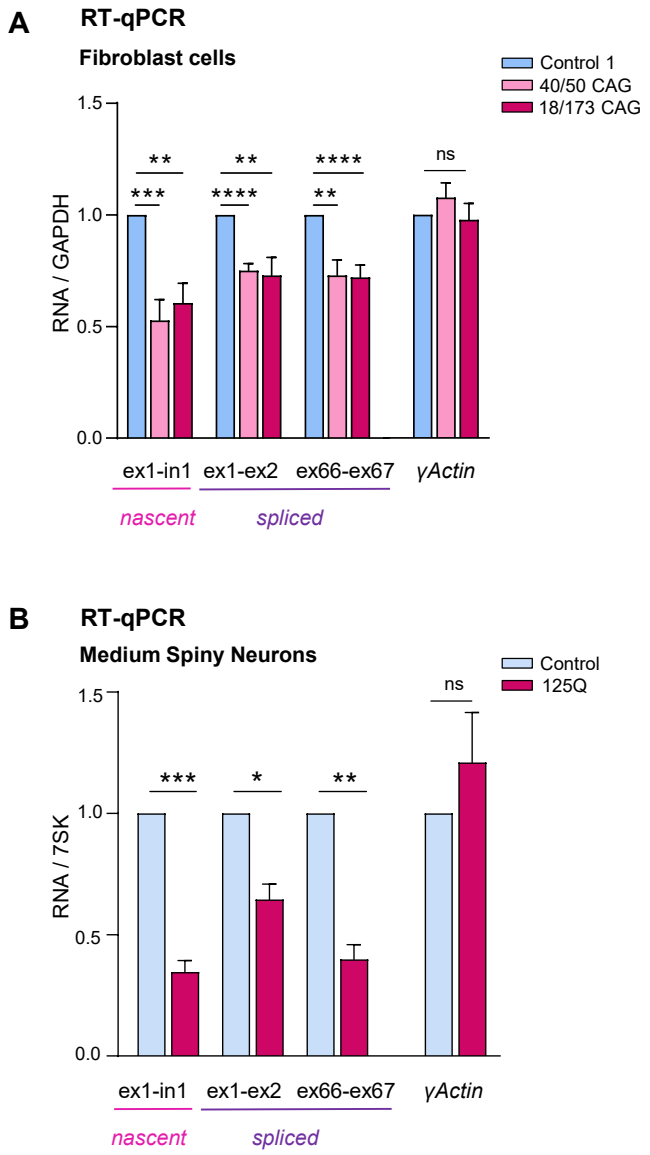


Figure 3.7. *HTT* gene expression is impaired in HD fibroblasts and MSNs.

A. RT-qPCR analysis for *HTT* and *γActin* genes in Control and HD fibroblast cells. Values are relative to the Control 1 cell line and normalised to *GAPDH* (n=4 means ± SEM, ** p>0.01, *** p>0.001, **** p>0.0001 by two-tailed unpaired t-test).

B. RT-qPCR analysis for *HTT* and *γActin* genes in Control and 125Q HD MSNs. Values are relative to the Control cell line and normalised to *7SK* (n>2; means ± SEM, * p> 0.05, ** p>0.01, *** p>0.001 by two-tailed unpaired t-test).

3.2.2 Pol II levels are decreased at the *HTT* promoter in HD

To further investigate the transcriptional profile of the *HTT* gene in HD cells, ChIP-qPCR analysis using an antibody recognising total Pol II was performed in lymphoblast cells. A set of qPCR primers spanning the whole *HTT* gene body were used, with some amplicons flanking the CAG repeats (amplicons 'prom' and 'A') and amplicons further in the gene body (amplicons 'E' and 'F') (Fig. 3.4A and Table 2.2). ChIP-qPCR showed that Pol II was enriched over the promoter-proximal region in lymphoblasts (Fig. 3.8A, blue bars), similar to HeLa cells (Fig. 3.4B). Interestingly, Pol II level was reduced at amplicon 'A' in HD cells compared to Control cells (Fig. 3.8A, pink and red bars). Although not statistically significant, Pol II levels were also reduced in all promoter-proximal amplicons in HD cells (amplicons 'up1' and 'prom') compared to Control cells. Furthermore, levels of Pol II over the *γActin* gene were similar in Control and HD cells, indicating that decreased Pol II levels were specific to the promoter-proximal regions of the *HTT* gene. This result, together with RT-qPCR data (Fig. 3.6-3.7), further confirms transcriptional down-regulation of the *HTT* gene in HD cell lines.

3.3 HD affects *HTT* gene epigenetic regulation

Studies have shown that alterations in the epigenetic status surrounding repeat expansions play a crucial role in regulating the stability of tandem repeats (Evans-Galea *et al.*, 2013). Furthermore, these changes may also impact the transcriptional rate of the expanded gene, ultimately resulting in changes in gene expression (Evans-Galea *et al.*, 2013). Our lab has previously demonstrated that heterochromatin formation at expanded GAA repeats contributes to frataxin (*FXN*) gene repression in Friedreich Ataxia (Groh *et al.*, 2014a). Since *HTT* gene chromatin structure and its epigenetic signature have not been yet characterized, I analysed the presence of both active and repressive heterochromatin marks across the *HTT* gene in Control and HD cells.

3.3.1 Active chromatin marks are decreased over *HTT* promoter in HD

To analyse active transcription at the *HTT* gene, ChIP-qPCR for H3K9ac mark was performed. ChIP-qPCR analysis across *HTT* showed that the H3K9ac active mark is enriched over the *HTT* promoter and reduced in HD cells (Fig. 3.8B). This result further supports Pol II ChIP-qPCR

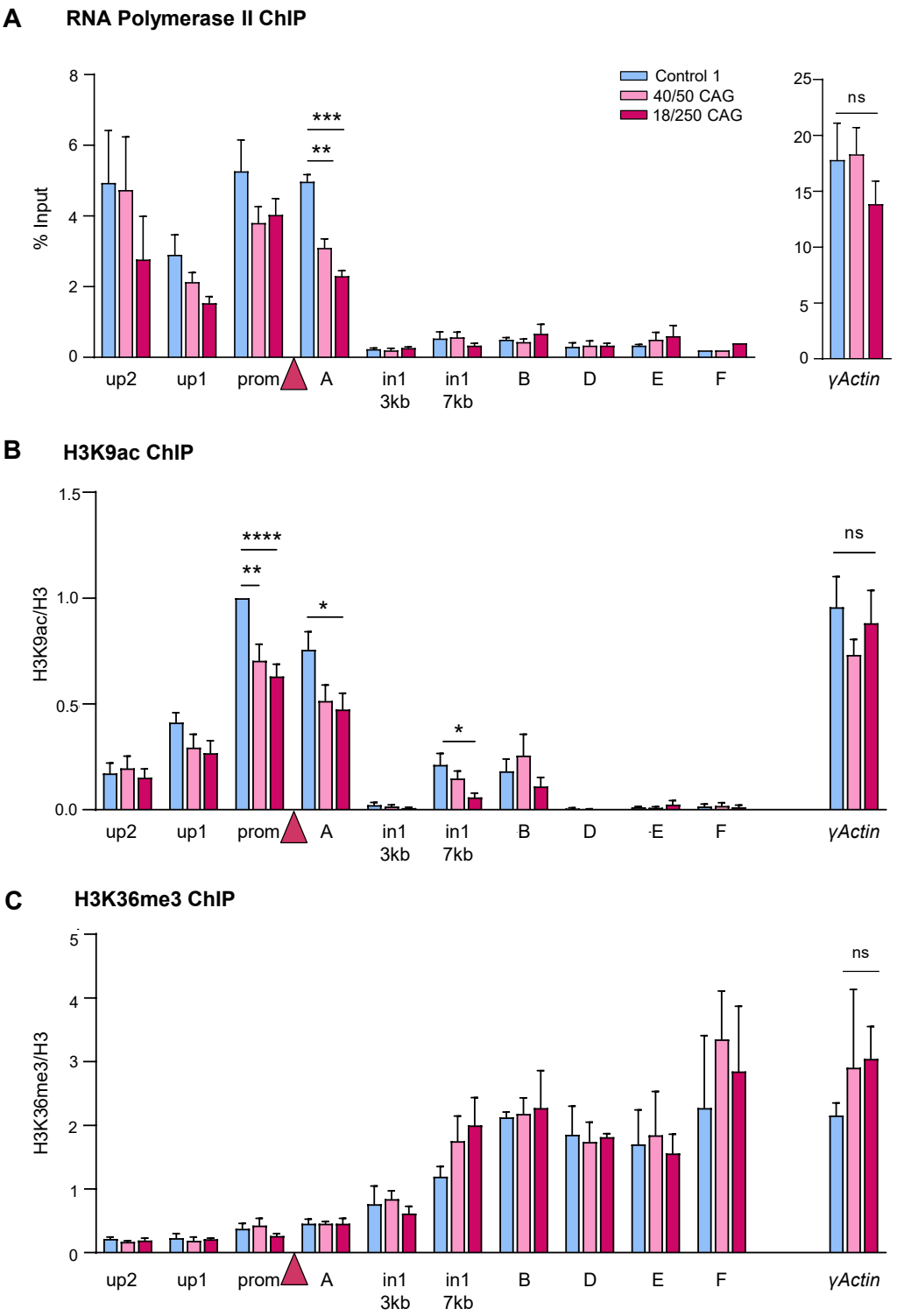


Figure 3.8. Pol II and active H3K9ac mark are decreased over the promoter-proximal regions of the *HTT* gene in HD lymphoblasts.

ChIP-qPCR for Pol II **(A)**, H3K9ac **(B)**, H3K36me3 **(C)** across *HTT* and *γActin* genes in Control and HD lymphoblast cells. For Pol II ChIP-qPCR, the values are expressed as % of Input (n=3; means ± SEM, ** p>0.01, *** p>0.001 by two-tailed unpaired t-test). For histone marks ChIP-qPCR, the values are normalised to the levels of total histone H3, as determined by ChIP-qPCR (n=5 (B), n=3 (C); means ± SEM, * p> 0.05, ** p>0.01, **** p>0.0001 by two-tailed unpaired t-test). The red triangle represents the $(CAG)_n$ expansion.

data (Fig. 3.8A) and the transcription down-regulation observed in HD cells by RT-qPCR (Fig. 3.6-3.7). Furthermore, levels of H3K9ac over the *γActin* gene were similar in Control and HD cells indicating that decreased H3K9ac levels were specific for the promoter-proximal regions of the *HTT* gene (Fig. 3.8B). In addition, an increase of the H3K9ac mark downstream of the *HTT* promoter, in the region corresponding to the amplicon 'in1-7kb', was detected in Control cells. This peak of H3K9ac mark could correspond to a second transcriptional starting site (TSS) in the *HTT* body, although this was not detected by GRO-cap. Interestingly, this H3K9ac peak correlates with the presence of a polyA signal at 7327 bp in *HTT* intron 1 as previously reported (Sathasivam *et al.*, 2013). Furthermore, H3K9ac ChIP-qPCR analysis in HeLa cells did not detect this peak in *HTT* intron 1 (Fig. 3.4C), suggesting that it may be cell-type-specific.

To investigate if the decrease in active chromatin marks was localized only at the promoter-proximal regions in HD cells or if it was also detected in the *HTT* gene body, H3K36me3 ChIP-qPCR analysis was performed. Similar to HeLa cells (Fig 3.4D), the distribution of the active H3K36me3 mark reflects a typical profile for a protein-coding gene, with enrichment of the signal in the *HTT* gene body in all three cell lines (Fig 3.8C). This further supports efficient transcription over the body of the *HTT* gene despite low Pol II and H3K9ac levels over this region (Fig 3.8A-B). However, no significant differences in its distribution were detected in the Control and HD cells across either *HTT* or *γActin* genes (Fig. 3.8C).

3.3.2 Repressive chromatin marks are increased over *HTT* in HD

Previous studies have demonstrated that expanded repeats are associated with increased levels of repressive heterochromatin marks (Groh *et al.*, 2014a). To investigate the relationship between transcriptional down-regulation and repressive heterochromatin marks formed over repeats, the distribution of H3K9me2/me3 and H3K27me3 histone modifications, associated with facultative heterochromatin, was investigated over the *HTT* gene in Control and HD cells. ChIP-qPCR for the repressive H3K9me2 mark showed that it was increased in the promoter-proximal region of the *HTT* gene in 18/250 CAG HD cells (amplicons 'up1', 'prom', 'A') and its levels were significantly higher compared to the highly expressed *γActin* gene (Fig. 3.9A). Interestingly, a patch of heterochromatin

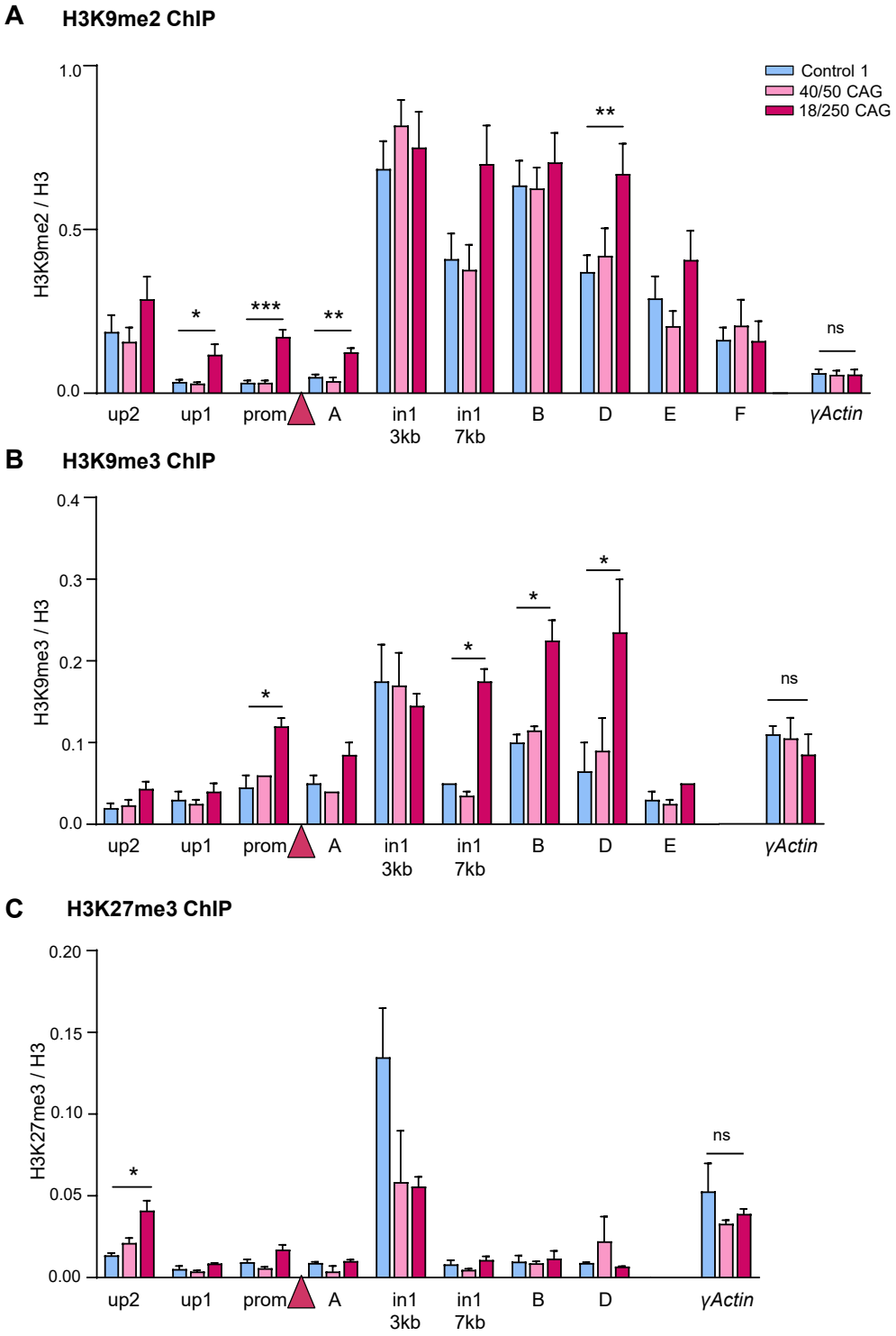


Figure 3.9. Increased heterochromatin formation over the *HTT* gene in HD lymphoblasts.
ChIP-qPCR for **(A)** H3K9me2, **(B)** H3K9me3, **(C)** H3K27me3 across *HTT* and *γActin* genes in Control and HD lymphoblast cells. Histone marks are normalised to the levels of total histone H3, as determined by ChIP-qPCR (n=4 (A), n=3 (B), n=2 (C); means ± SEM, * p> 0.05, ** p>0.01, *** p>0.001 by two-tailed unpaired t-test). The red triangle represents the $(CAG)_n$ expansion.

marked by H3K9me2 was also observed over the *HTT* gene body in all three cell lines, downstream of amplicon 'in1-3kb' (Fig. 3.9A), similar to the one detected in HeLa cells (Fig. 3.5A). However, its role in *HTT* regulation is currently unknown. In contrast, H3K9me3 ChIP-qPCR analysis revealed a different distribution of this repressive mark compared to H3K9me2. H3K9me3 was increased over the *HTT* gene body in 18/250 CAG HD cells (amplicons 'prom', 'in1-7kb', 'B', 'D' and 'E') (Fig. 3.9B). Overall, the epigenetic signature of the *HTT* gene in HD cells, showing an increase in repressive heterochromatin marks, correlates with its transcriptional down-regulation. The fact that the epigenetic silencing was observed in the HD cell line with longer repeats (18/250 CAG) could suggest that longer expansions may have a higher propensity to induce heterochromatinization. Further allele-specific analyses are needed to understand if these changes are specific to the mutant allele.

The distribution of the repressive H3K27me3 mark was also analysed. As found in HeLa cells (Fig. 3.5C), ChIP-qPCR for the H3K27me3 mark showed a peak of H3K27me3 signal in the middle of the *HTT* gene body, corresponding to the amplicon 'in1-3kb'. This peak was particularly high in Control cells and enriched compared to the *γActin* gene (Fig. 3.9C). Interestingly, this peak correlates both with low levels of Pol II and high levels of H3K9me2 in this region (Fig. 3.8A and 3.9A).

3.4 Analysis of available transcriptomic data in HD mouse models

3.4.1 Mouse transcriptomic data shows transcription down-regulation in HD

Epigenetic and transcriptional regulation are altered in HD brain tissues, particularly in the striatum of HD patients and mice (Bassi *et al.*, 2017). To investigate if HD mutations also locally affect *HTT* gene transcription and epigenetic regulation in mouse models of disease, available H3K27ac, H3K27me3, H3K3me3 and Pol II ChIP-seq data from the whole striatum of WT and HD Q140 heterozygous mice were analysed (Alcala-Vida *et al.*, 2021). In detail, this knock-in mouse model of HD was generated by inserting a chimeric mouse/human exon 1 containing 140 CAG repeats in the murine *HTT* gene (Menalled *et al.*, 2003). This mouse model displays early behavioural and neuropathological anomalies, including specific protein aggregates that first appear in the

striatum, providing a powerful model to study the mechanisms involved in the neuropathology of HD (Menalled *et al.*, 2003).

Pol II ChIP-seq profiles across the *HTT* gene showed that Pol II is enriched over the promoter-proximal region of the mouse *HTT* gene and decreased in HD Q140 compared to WT mice by ~35% (Fig. 3.10A, top panel). H3K27ac is found at both proximal and distal regions of the TSS (Heintzman *et al.*, 2009). H3K27ac ChIP-seq profiles across the mouse *HTT* gene showed that this active mark is enriched over the *HTT* promoter and reduced in HD Q140 mice by ~35%, in line with Pol II levels (Fig. 3.10A, bottom panel). These results are in line with the promoter-proximal decrease of Pol II and H3K9ac marks observed in human HD cells (Fig. 3.8A-B). Overall, this is consistent with the transcriptional down-regulation of the *HTT* gene observed in HD patient cells (Fig. 3.6-10). Furthermore, genome-wide analysis demonstrated a ~20% decrease in Pol II levels on promoter regions in HD Q140 mice compared to WT (Fig. 3.10B), in line with already reported transcription dysregulation observed in HD patient cells and mouse models (Hodges *et al.*, 2006; Langfelder *et al.*, 2016; Ament *et al.*, 2017). Moreover, meta profiles for H3K3me3 and H3K27me3 demonstrated a global increase of these repressive marks in HD Q140 compared to WT mice, despite these marks being present at a relatively low level in the genome (IP / Input values < 1) (Fig. 3.10C). The epigenetic dysregulation and transcription down-regulation found in different HD models suggest that loss of *HTT* expression could contribute to disease.

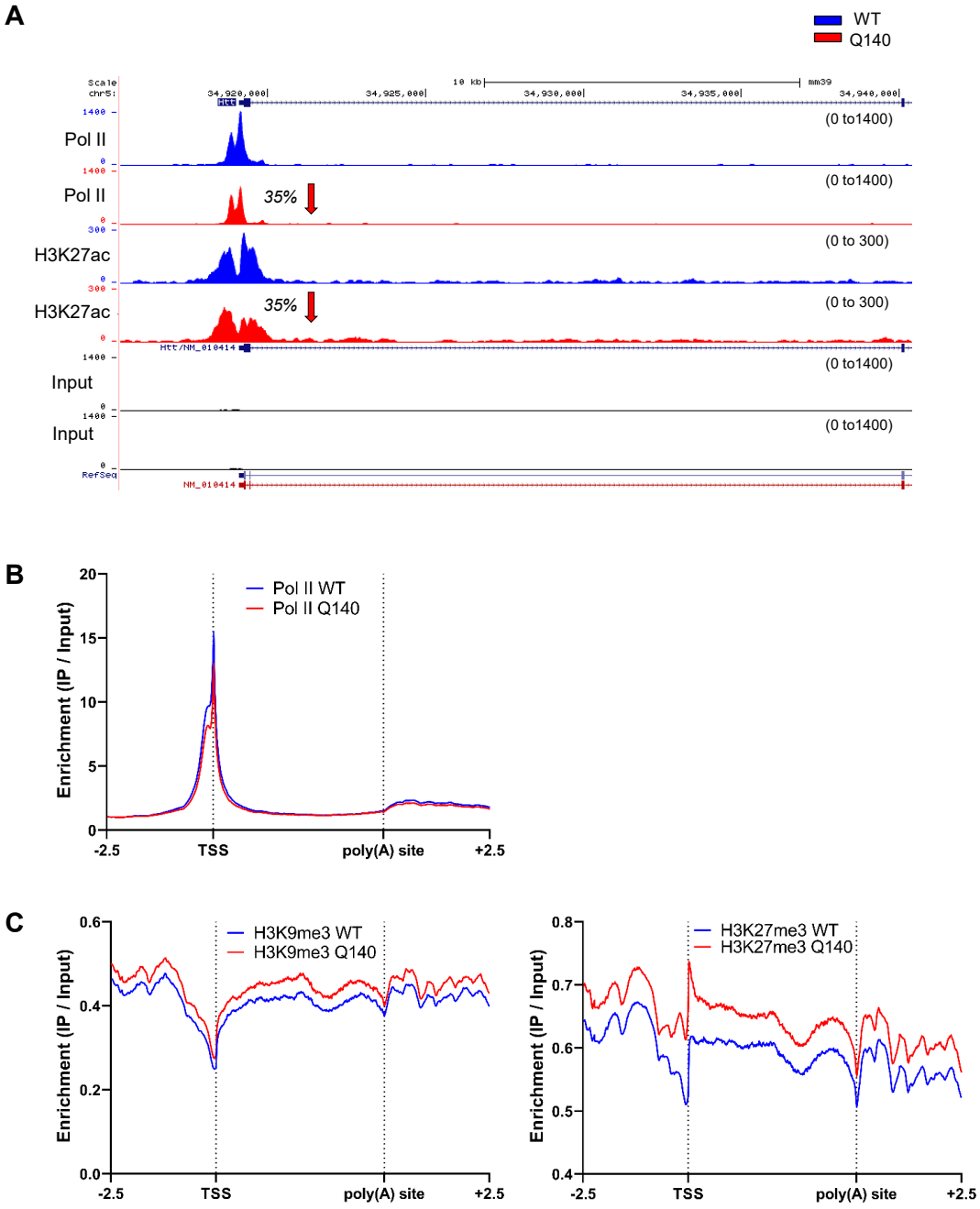


Figure 3.10 Mouse transcriptomic data shows transcriptional down-regulation in HD.

A. Screenshot of Pol II and H3K27ac ChIP-seq profiles of the *HTT* gene in WT (blue) and HD Q140 (red) mice. Numbers in brackets indicate the viewing range (reads per kilobase of transcript per million mapped reads, RPKM). Data are from Alcala-vida *et al*., 2021.

B. Meta-analysis of read enrichment, defined as RPKM-normalised IP signal ratioed to RPKM-normalised Input signal, for Pol II ChIP-seq in HD WT (blue) and HD Q140 (red) mice. Distance from Transcriptional Starting Site (TSS) or poly (A) site in kb. Data from Alcala-vida *et al*., 2021.

C. Meta-analysis of read enrichment for H3K9me3 and H3K27me3 ChIP-seq WT (blue) and HD Q140 (red) mice. Distance from TSS or poly (A) site is shown in kb. Data are from Alcala-vida *et al*., 2021.

Figure generated by Dr. M. Tellier.

3.5 Investigation of HTT protein levels in HD

3.5.1 HTT protein levels are decreased in HD cells

The central role of the mutant HTT protein in HD pathogenesis is clear, however, there is still an incomplete understanding of the molecular mechanisms that drive pathology. Evaluating the expression of the wild-type and mutant form of HTT is crucial for designing therapeutic approaches to lower mutant *HTT* levels. Previous studies have already shown lower levels of mHTT in striatal and cortical mouse tissues (Franich *et al.*, 2018) and in juvenile HD brain samples (Evers *et al.*, 2015).

Having observed lower levels of *HTT* mRNA in HD cell lines, the total HTT protein levels were also evaluated. Western Blot analysis of several Control and HD lymphoblast cells showed a reduction in HTT protein levels in diseased cells (Fig. 3.11A). A similar trend of HTT protein reduction was observed in HD fibroblasts compared to Control (Fig. 3.11B). This reduction may correspond to the lower levels of mHTT, observed in multiple HD models. Further use of specific antibodies able to distinguish wild-type and mutant HTT would enhance this analysis.

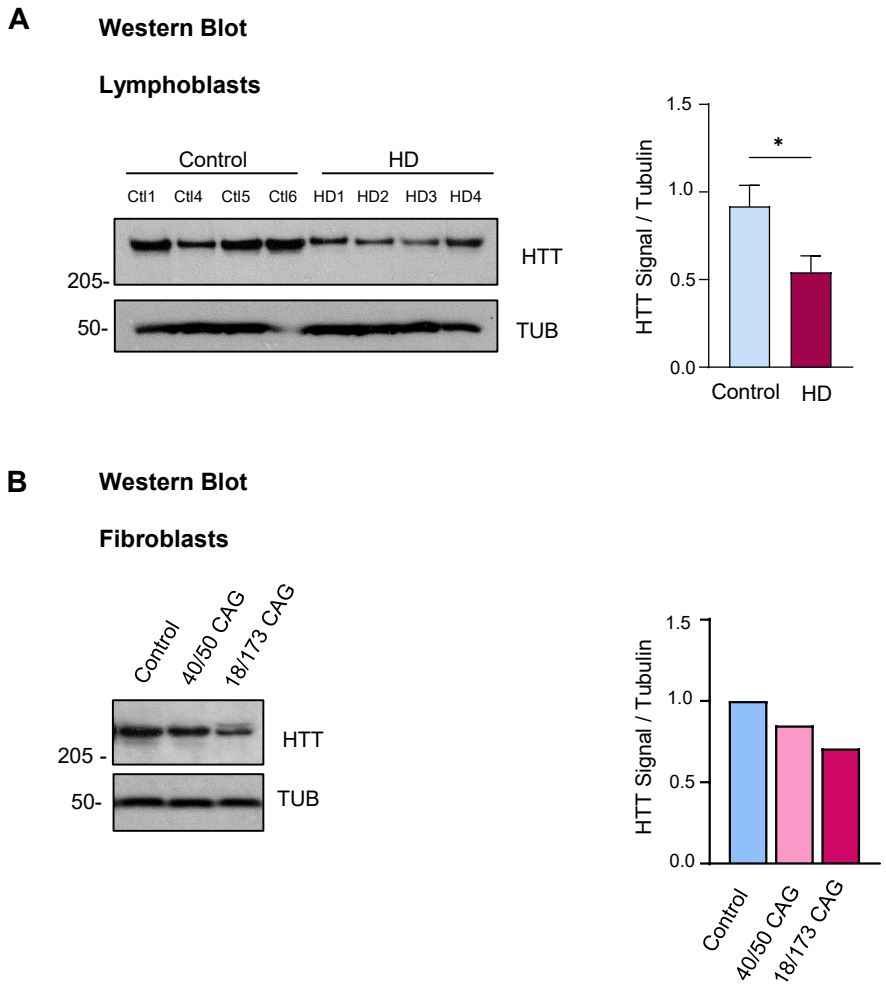


Figure 3.11. Decreased HTT protein levels in HD cells.

A. **Left panel.** Western Blot analysis of 20 µg of whole cell extract (WCE) obtained from Control and HD lymphoblast cell lines (Table 2.1). Representative image of Western Blot, probed with MAB2166 HTT antibody. Tubulin is a loading control. **Right panel.** Quantitation of protein levels. HTT protein levels were normalized to tubulin (n=2, means ± SEM; * $p > 0.05$ by two-tailed unpaired t-test).

B. **Left panel.** Western Blot analysis of 20 µg of whole cell extract (WCE) obtained from Control and HD fibroblast cell lines (Table 2.1). Representative image of Western Blot, probed with MAB2166 HTT antibody. Tubulin was a loading control. **Right panel.** Quantitation of protein levels. HTT protein levels were normalized to tubulin (n=1).

3.6 Summary

In this Chapter, the transcriptional and epigenetic signature of the *HTT* gene has been analysed in different cellular and animal models. It was shown that *HTT* mRNA expression varies in different cell lines (Fig. 3.1-3.2). Furthermore, specific genomic features of the *HTT* gene were initially investigated in healthy cells (Fig. 3.3-3.5) and then in cells derived from Control and HD patients. Overall, the results in this Chapter demonstrate that the presence of expanded CAG repeats in the *HTT* gene, in either HD patient cell lines or HD mouse models, somewhat impairs transcription-associated processes (Fig. 3.6-3.10). Specifically, down-regulation of *HTT* transcription was observed in the HD context, together with an increase in heterochromatin formation and a decrease in active chromatin marks over the *HTT* gene promoter (Fig. 3.8-3.10). This is in line with previous studies that showed a decrease in mutant *HTT* mRNA levels in adult-onset and juvenile HD patients compared to the controls (Evers *et al.*, 2015). However, Evers *et al.* reported that reduced mutant *HTT* mRNA expression did result in lower mutant HTT protein levels only in juvenile HD patients and not in adult-onset HD patients compared to controls. Here I show that the levels of HTT protein were decreased in both adult onset and juvenile HD cells (Fig. 3.11). All these features may have critical implications for the molecular pathology of HD, highlighting the importance of the loss-of-function of the wild-type HTT protein in the context of the disease.

In Chapter 4, the molecular mechanisms that may trigger such transcriptional down-regulation, including the presence of unusual RNA/DNA structures over the *HTT* gene, are investigated.

Chapter 4

Structural features of CAG triplet repeats

4.1 Non-canonical RNA/DNA structures are involved in HD pathology

Different non-canonical RNA/DNA structures, including double-stranded RNAs (dsRNAs) and R-loops, are known to be implicated in the pathogenesis of several expanded repeat disorders (Groh & Gromak, 2014; Freudenreich, 2018). These structures can affect transcription, DNA replication and genome stability and their formation has been proposed to be a general feature of RE disorders (Krzyzosiak et *al.,* 2012; Paulson, 2018). Moreover, the formation of unusual RNA/DNA structures over repeats has been recently demonstrated to facilitate repeat expansions and instability (Khristich & Mirkin, 2020; Tabrizi *et al.*, 2020). In the context of HD, CAG repeats have been shown to adopt RNA hairpin structures *in vitro* (de Mezer *et al.*, 2011; Wojciechowska & Krzyzosiak, 2011) and DNA-containing CAG repeats have been shown to form R-loops *in vitro* (Lin *et al.*, 2010). However, investigation of their formation *in vivo* is still missing. Genome-wide R-loop imbalance contributes to the pathology of several diseases (Table 1.1). Furthermore, recent studies have demonstrated that genome-wide accumulation of R-loops can impair several cellular functions and compromise cell survival, causing DNA damage, inflammation, and cell death (Cristini *et al.*, 2022; Giordano *et al.*, 2022).

The experiments described in this Chapter were aimed at investigating the formation of non-canonical RNA/DNA structures over expanded CAG repeats and genome-wide in HD cells.

4.2 Non-canonical RNA/DNA structures are formed over CAG repeats of the *HTT* gene

4.2.1 R-loops are formed over the *HTT* gene but not enriched at expanded repeats

Previous research conducted in our lab demonstrated that R-loop formation occurs over expanded GAA repeats and contributes to the *FXN* and *FMR1* gene silencing observed in FRDA and FXS, respectively (Groh *et al.*, 2014b).

To investigate whether R-loops form over the *HTT* gene, RNA/DNA hybrid immuno-precipitation (DRIP) was performed. In this technique, nuclear DNA and associated RNA, for example, R-loops, are extracted from non-crosslinked cells and the R-loops are then immuno-precipitated using the S9.6 antibody, which is specific for RNA/DNA hybrids. Immuno-precipitation

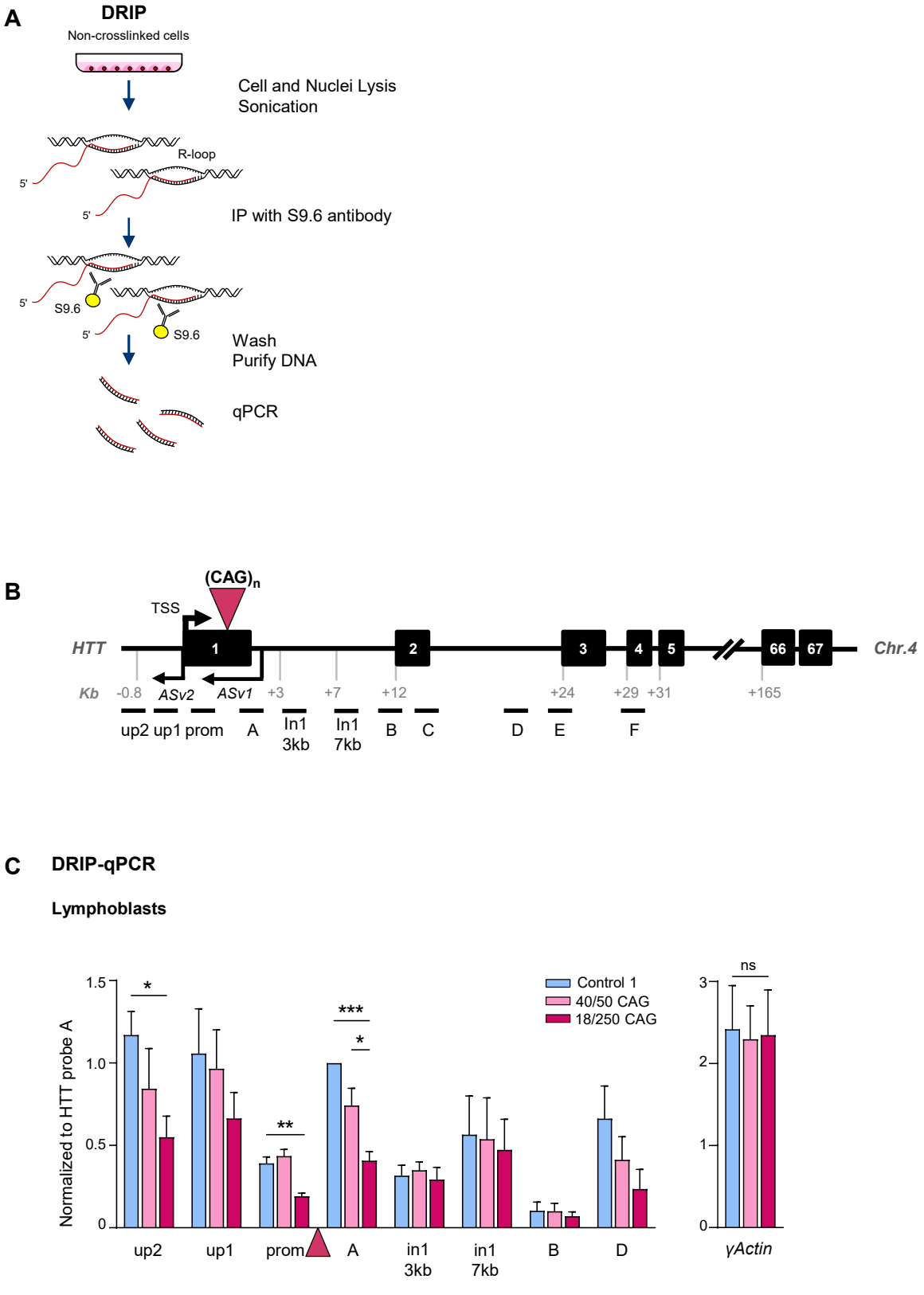


Figure 4.1. RNA/DNA hybrids are formed across the *HTT* gene but not enriched over expanded repeats in HD lymphoblast cells.
The caption of the figure can be found below.

Figure 4.1 (continued). RNA/DNA hybrids are formed across the *HTT* gene but not enriched over expanded repeats in HD lymphoblast cells.

A. Overview of the DRIP method. Initially, nuclei are extracted from non-crosslinked cells and proteins are removed by proteinase K treatment. Genomic DNA containing RNA/DNA hybrids (R-loops) is then precipitated and sonicated. RNA/DNA hybrids are immuno-precipitated using the S9.6 antibody which is specific for RNA/DNA hybrids. Immuno-precipitation is followed by extensive washes, DNA purification and qPCR analysis.

B. Diagram of the *HTT* gene. Black boxes are exons, lines are introns, and the red triangle represents the $(CAG)_n$ expansion. Amplicons for qPCR analysis are shown below the gene diagram. Numbers indicate sizes in kilobase (kb) from the Transcriptional Start Site (TSS). *ASv1* and *ASv2* indicate *HTT* antisense transcripts.

C. DRIP-qPCR on endogenous *HTT* gene in Control 1 and HD lymphoblast cells. The first intron of the *γActin* gene is used as a positive control. The red triangle represents the $(CAG)_n$ expansion. The values are normalized to *HTT* amplicon A in Control cells (n=4; mean ± SEM, * $p < 0.05$, ** $p < 0.01$ and *** $p < 0.001$ by two-tailed unpaired t-test).

is followed by extensive washes, DNA purification and qPCR analysis of specific genomic regions (Fig. 4.1A) (Skourti-Stathaki *et al.*, 2011; Cristini *et al.*, 2018). A set of primers was designed to investigate RNA/DNA hybrid distribution across the *HTT* gene in Control and HD lymphoblast cells (Fig. 4.1B-C and Table 2.2). The intron 1 region of the *γActin* gene, where a high level of transcriptional R-loops has previously been detected, was used as a positive control (Groh *et al.*, 2014a). DRIP-qPCR showed that RNA/DNA hybrids are formed over the *HTT* gene and peak in the region upstream of the *HTT* promoter (amplicons 'up2' and 'up1', corresponding to *HTTAS*) (Fig. 4.1C). Another peak was observed downstream of the *HTT* promoter after the CAG repeats (amplicon 'A') in Control and HD cells (Fig 4.1C). However, no RNA/DNA hybrid enrichment was observed over expanded CAG repeats compared to Control (Fig. 4.1C). The slight decrease in RNA/DNA hybrid levels over the expanded repeats of the *HTT* gene in HD cells was in line with the decrease of transcription observed in 40/50 CAG and 18/250 CAG HD lymphoblast cells by Pol II ChIP and RT-qPCR analysis (Fig. 3.6B-C and 3.8A). R-loops levels were unchanged in the *γActin* control gene, indicating that decreased levels of R-loops are specific for the expanded *HTT* gene (Fig. 4.1C).

To investigate whether the RNA/DNA hybrid formation over the expanded repeats in the *HTT* gene takes place in committed neuronal cells, where disease phenotypes are most profound, iPSC-derived astrocytes and MSNs from Control and HD patients were also analysed (all cells were obtained from the Tabrizi lab). In these neuronal cells, a slight decrease in RNA/DNA hybrid levels was detected in HD cells compared to Control cells, in line with the results obtained using lymphoblast cells (Fig 4.2).

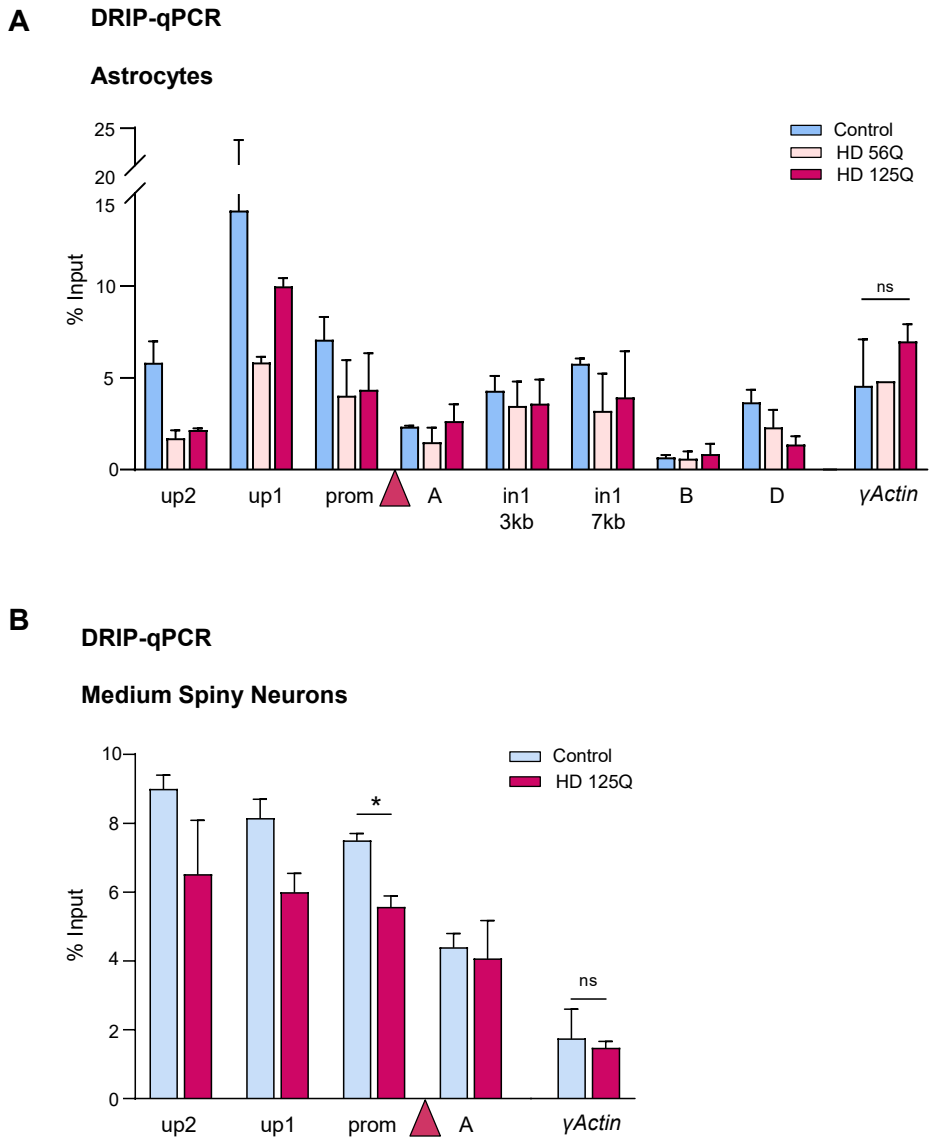


Figure 4.2. RNA/DNA hybrids are formed across the *HTT* gene but not enriched over expanded repeats in HD neuronal cells.

A. DRIP-qPCR on endogenous *HTT* gene in Control and HD iPSC-derived astrocytes. The first intron of the *γActin* gene is used as a positive control. The red triangle represents the $(CAG)_n$ expansion. The values are expressed as % of Input (n=2; mean ± SEM).

B. DRIP-qPCR on endogenous *HTT* gene in Control and HD iPSC-derived MSNs. The first intron of the *γActin* gene is used as a positive control. The red triangle represents $(CAG)_n$ expansion. The values are expressed as % of Input (n≥2; mean ± SEM, * $p < 0.05$ by two-tailed unpaired t-test).

4.2.2 Double-stranded RNAs are enriched on expanded *HTT* transcripts

To investigate if dsRNAs are formed on expanded *HTT* transcripts, dsRNA immuno-precipitation (dsRIP) using the J2 antibody (Sciscon), which allows specific detection of dsRNA, was employed (Fig. 4.3A). This protocol was adapted from a recent publication with minor modifications (Aktas *et al.*, 2017). Specifically, anti-dsRNA antibody J2 recognises dsRNA, independently of the sequence of the antigen, provided that the dsRNA length of the helix is more than 40 bp (Schonborn *et al*, 1991; Bonin *et al*, 2000). In detail, in the dsRIP technique, cell lysis was followed by nuclei lysis with RIP buffer and nuclear extracts were subjected to J2 antibody immuno-precipitation (Fig. 4.3A). The immuno-precipitation of dsRNA was followed by extensive washes and RNA extraction. Immuno-precipitated RNA was then reversed transcribed and analysed by qPCR with primers across the *HTT* gene (Fig. 4.3A). An enrichment of the dsRNA signal was detected on expanded *HTT* transcripts in the region surrounding the CAG repeats (amplicons 'up1', 'prom', 'A') in both HD cell lines, with longer repeat regions giving greater enrichment (Fig. 4.3B, dark pink bars). dsRNAs were less enriched in regions further downstream of the CAG repeats (amplicon 'B'). dsRNA enrichment was absent from *FXN* and *GAPDH* control genes, indicating that increased dsRNA levels are specific for CAG repeats of the *HTT* gene and that they do not represent a global feature of multiple genes in HD cells (Fig. 4.3B). Comparison of the dsRIP-qPCR data in Control and HD cell lines suggested that the level of dsRNAs correlates with expansion length. Moreover, the actual enrichment of dsRNA could be higher considering an overall reduction in nascent and mRNA *HTT* levels detected in the HD cells (Fig. 3.6B-C).

4.2.3 dsRNAs on expanded *HTT* transcripts are sensitive to RNase III digestion

To confirm the specificity of the dsRNA signal, dsRIP was performed with the addition of recombinant *E.coli* RNase III (NEB), which specifically digests dsRNA, following immuno-precipitation. Following RNase III digestion, the dsRNA signal was strongly reduced on all *HTT* amplicons and on the *FXN* and *GAPDH* genes in both Control and HD cells, suggesting that genuine dsRNAs are detected in this experiment (Fig. 4.4). RNase III treatment in the HD cell lines with 18/250 CAG repeats was less efficient compared to Control and 40/50 CAG cell lines, potentially

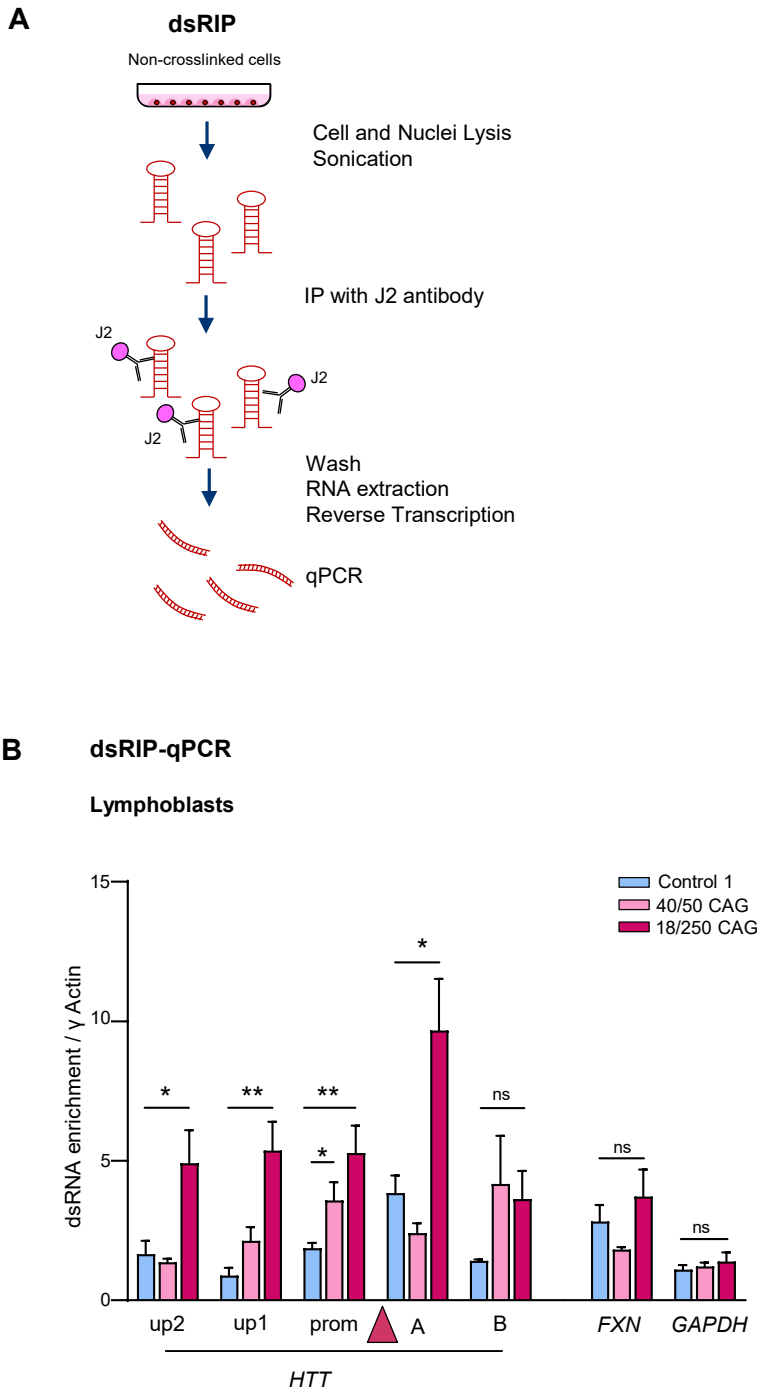


Figure 4.3. Double-stranded RNAs are enriched on expanded *HTT* transcripts.

A. Overview of the dsRIP method. Initially, cell lysis is followed by nuclei lysis. dsRNAs are immuno-precipitated using J2 antibody, specific for dsRNA. Immuno-precipitation is followed by extensive washes, RNA extraction, reverse transcription, and qPCR analysis.

B. dsRIP-qPCR on endogenous *HTT* gene in Control 1 and HD lymphoblast cells. Amplicons in *FXN* and *GAPDH* genes are used as negative controls. The red triangle represents the $(CAG)_n$ expansion. The values are normalized to *γActin* amplicon for each cell line respectively (n=4; mean ± SEM, * $p < 0.05$ and ** $p < 0.01$ by two-tailed unpaired t-test).

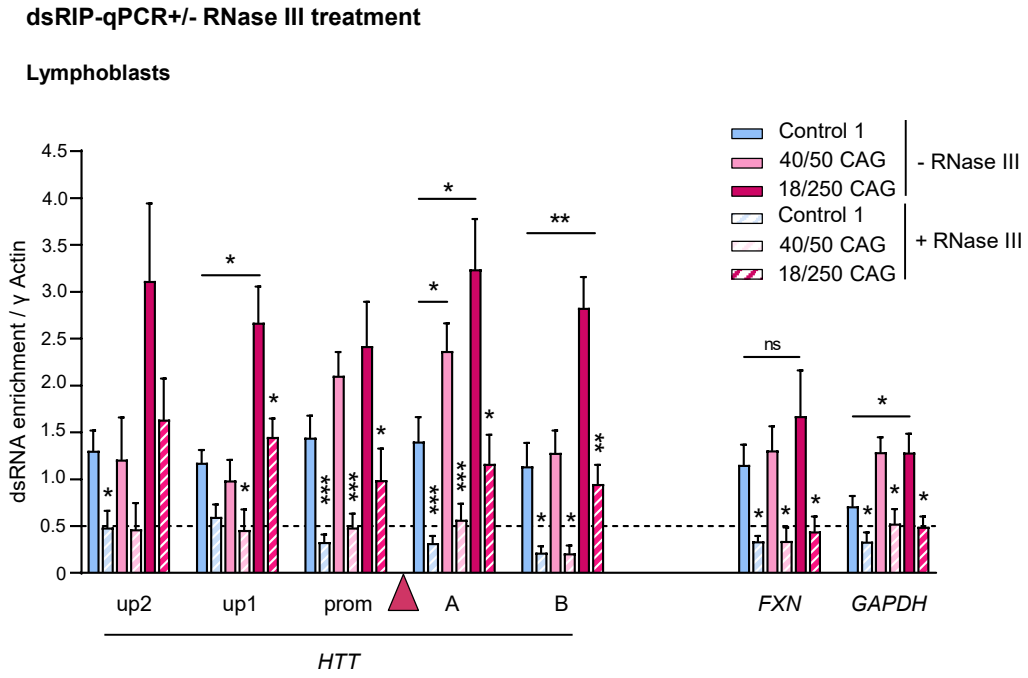


Figure 4.4. Double-stranded RNAs on expanded *HTT* transcripts are sensitive to RNase III treatment.
dsRIP samples were treated with recombinant *E.coli* RNAse III (NEB, 2U) for 45 min at 37ºC following immuno-precipitation with J2 antibody. Amplicons in *FXN* and *GAPDH* genes are used as negative controls. The dashed line represents the background signal. The red triangle represents the $(CAG)_n$ expansion. The values are normalized to *γActin* amplicon for each cell line respectively (n=5; mean ± SEM, * $p < 0.05$, ** $p < 0.01$ and *** $p < 0.001$ by two-tailed unpaired t-test).

due to the higher enrichment and stability of dsRNA in these cells. Nevertheless, the RNase III treatment confirmed the specificity of the dsRNA signal suggesting that genuine dsRNAs are formed over the expanded *HTT* transcripts (Fig. 4.4).

4.3 Formation of non-canonical RNA/DNA structures in HD cells

4.3.1 Genome-wide R-loop levels in HD cells

4.3.1.1 R-loop Slot Blot shows global increase of RNA/DNA hybrids in HD cells

To investigate if genome-wide R-loop accumulation could be implicated in HD pathogenesis, R-loop Slot Blot analysis was performed in Control and HD lymphoblast cells. Genomic DNA, containing R-loops, was extracted from these cells and loaded on a Slot Blot probed with the S9.6 antibody. A small but significant increase in the global level of nuclear RNA/DNA hybrids was observed in both HD cell lines (40/50 CAG and 18/250 CAG) compared to the Control (Fig. 4.5A). Importantly, the S9.6 signal was abolished by RNase H treatment, indicating that it is specific for RNA/DNA hybrids. In these experiments, ssDNA was used as a loading control. Given the modest effect observed, a panel of three additional Control and HD cell lines was analysed (Fig. 4.5B). This extended analysis showed a trend of increased RNA/DNA hybrids in HD cells compared to Controls, confirming that R-loop accumulation may be a feature of HD pathology in lymphoblast cells (Fig. 4.5A-B). Moreover, a similar trend was also observed in fibroblast HD cells (Fig. 4.5C). Because the majority of cellular R-loops are a consequence of active transcription, we reasoned that a general transcription inhibitor should cause a decrease in the R-loop signal. Therefore, Control and 18/250 CAG HD lymphoblast cells were treated with the transcriptional inhibitor actinomycin D (Perry & Kelley, 1970). Following the treatment of cells with actinomycin D for 18 h, R-loops were reduced in both cell lines, confirming that they are dependent on transcription (Fig. 4.6).

To investigate the formation of R-loops in neuronal cells in the context of HD, global RNA/DNA hybrids were investigated in MSNs, differentiated from iPSCs by Tabrizi lab, with the Control cell line containing 20 CAG repeats and the HD 125Q MSNs containing ~130 CAG repeats. In this experiment, MSNs derived from iPSCs of different passages (p20, p26, p36) were used (Fig 4.7). Interestingly, there was an ~5-fold increase in R-loops levels in HD cell lines compared to Control (Fig. 4.7A). When samples were exposed to RNase H digestion (+/- RNase H conditions), the R-loop increase was still detected, although to a lesser extent (~2-fold), most probably due to non-specific R-loop degradation during incubation at 37ºC (Fig. 4.7B). Furthermore, the S9.6 signal

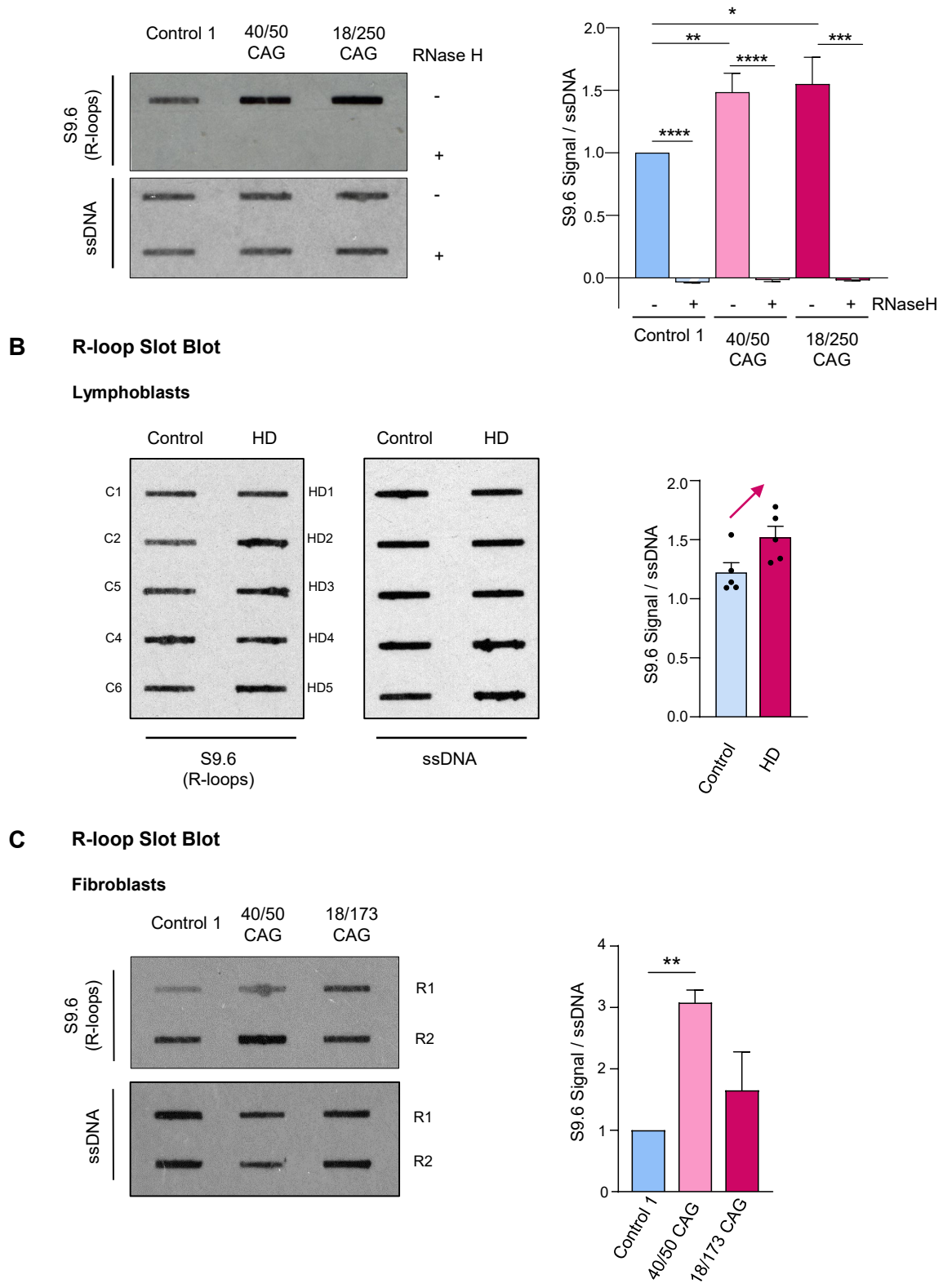


Figure 4.5. R-loop Slot Blot shows an overall increase of RNA/DNA hybrids in HD lymphoblast and fibroblast cells.
The caption of the figure can be found below.

Figure 4.5 (continued). R-loop Slot Blot shows an overall increase of RNA/DNA hybrids in HD lymphoblast and fibroblast cells.

A. **Left panel.** Slot Blot analysis with S9.6 antibody in Control 1 and HD lymphoblast cells treated with RNase H enzyme. Nuclear DNA was extracted from cells, treated -/+ RNase H (1.7 U, NEB) for 2.5 h at 37ºC and analysed by Slot Blot, probed with S9.6 antibody. RNAse H treatment led to a complete loss of R-loop signal confirming the specificity of the antibody. Slot Blot probed with the antibody recognizing ssDNA was used as a loading control. **Right panel.** Quantitation of the S9.6 Slot Blot signal. Values shown are normalized to the Control 1 cell line and to the corresponding ssDNA signal (n=5; mean ± SEM, * $p < 0.05$, ** $p < 0.01$ and **** $p < 0.001$ by two-tailed unpaired t-test).

B. **Left panel**. Slot Blot analysis with S9.6 antibody in additional Control and HD lymphoblast cells (see Table 2.1 for cell lines description). Nuclear DNA was extracted from cells and analysed by Slot Blot, probed with S9.6 antibody. Slot Blot probed with the antibody recognizing ssDNA was used as a loading control. **Right panel.** Quantitation of the S9.6 Slot Blot signal. Values shown are represented as the average of n≥3 biological replicates (mean ± SEM, ns $p = 0.1434$ by two-tailed unpaired t-test), each dot representing different cell lines from independent individuals. Values shown are normalized to the relative ssDNA signal.

C. **Left panel.** Slot Blot analysis with S9.6 antibody in Control 1 and HD fibroblast cells. R1 and R2 indicate two independent biological repeats. Nuclear DNA was extracted from cells and analysed by Slot Blot, probed with S9.6 antibody. Slot Blot probed with the antibody recognizing ssDNA was used as a loading control. **Right panel.** Quantitation of the S9.6 Slot Blot signal. Values shown are normalized to the Control 1 cell line and to the corresponding ssDNA signal (n=2; mean ± SEM, ** $p < 0.01$ by two-tailed unpaired t-test).

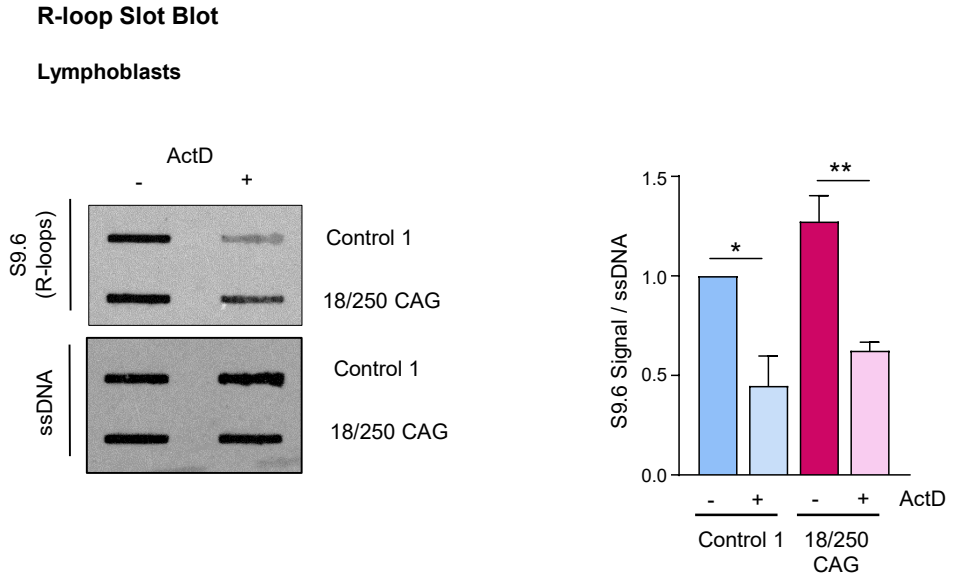


Figure 4.6. Accumulation of R-loops in HD cells is dependent on transcription.
Left panel. Slot Blot analysis with S9.6 antibody in Control and HD lymphoblast cells treated with 5 µg/mL of actinomycin D (+ ActD) or ethanol (- ActD) for 18 h. Nuclear DNA was then extracted from cells and analysed by Slot Blot, probed with S9.6 antibody. Slot Blot probed with the antibody recognizing ssDNA was used as a loading control. **Right panel.** Quantitation of the S9.6 Slot Blot signal. Values shown are normalized to the Control 1 cell line and to the relative ssDNA signal (n=2; mean ± SEM, * $p < 0.05$ and ** $p < 0.01$ by two-tailed unpaired t-test).

was only partially abolished by RNase H treatment in HD MSNs, despite the exposure of the samples to RNase H treatment for 4.5 h at 37ºC. The resistance of these R-loops to RNase H degradation maybe be due to their increased stability, as previously demonstrated for other pathological R-loops (Crossley *et al.*, 2023). Furthermore, the analysed samples may contain traces of dsRNA contamination, which can be also detected by S9.6 antibody (Phillips *et al.*, 2013). Additional RNase III treatment would be necessary to investigate this further.

Interestingly, HD MSNs differentiated from iPSCs with higher number of cell passages showed a greater R-loop enrichment compared to HD MSNs differentiated from younger iPSCs (Fig. 4.7). It is known that age can influence CAG repeat stability favouring repeat expansion, and finally worsening disease phenotypes (Kacher *et al.*, 2021). MSNs differentiated from older iPSCs may therefore have greater repeat instability and undergo expansion.

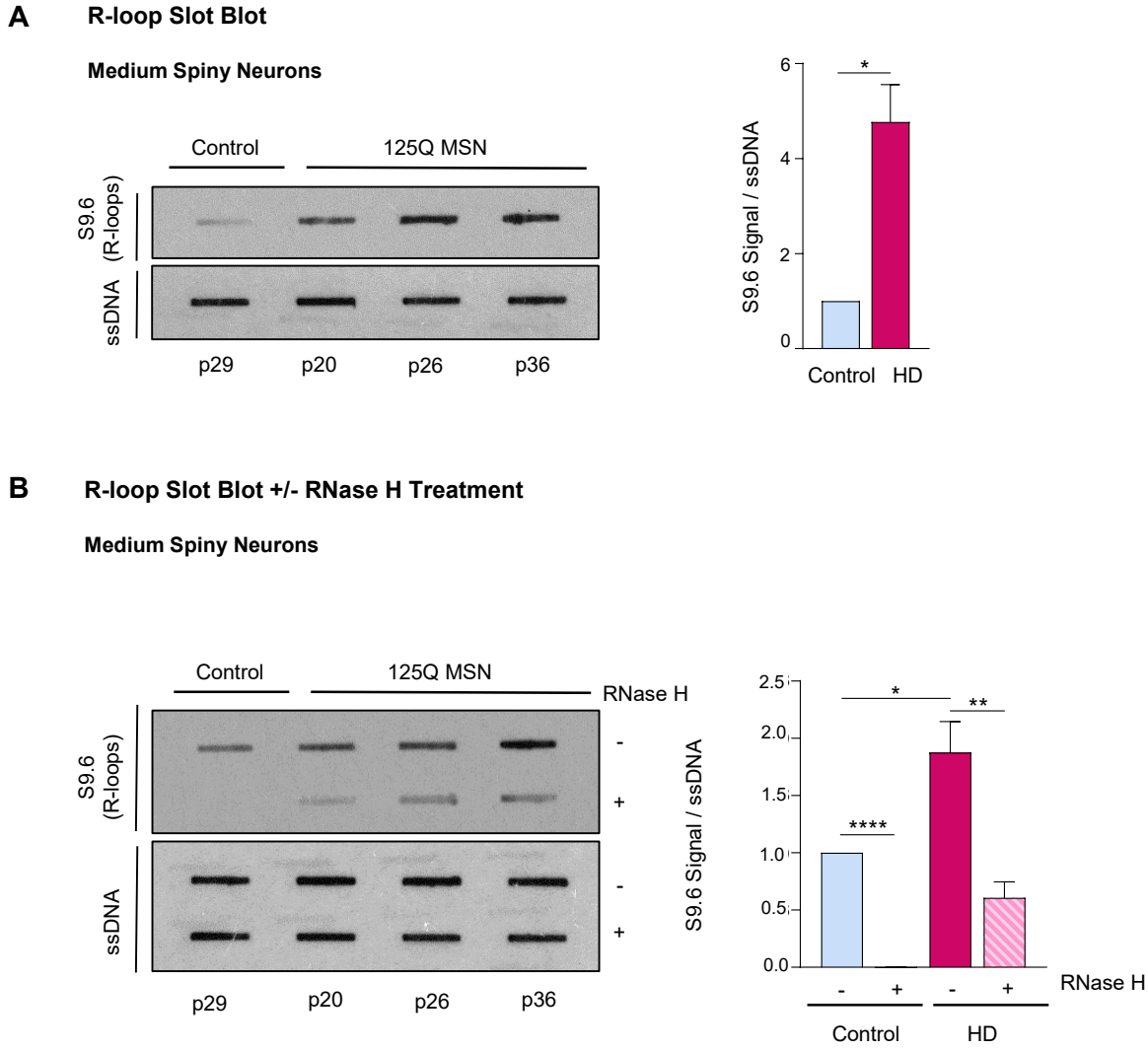


Figure 4.7. S9.6 Slot Blot shows an overall increase of RNA/DNA hybrids in MSN HD cells.

A. **Left panel.** Slot Blot analysis with S9.6 antibody in Control and HD MSNs differentiated from iPSCs (Table 2.1). Nuclear DNA was extracted from cells and analysed by Slot Blot, probed with S9.6 antibody. Slot blot probed with the antibody recognizing ssDNA was used as a loading control. The passage number of iPSCs used for MSN differentiation are indicated below the blot. **Right panel.** Quantitation of the S9.6 Slot Blot signal. Values shown are normalized to the Control cell line and to the corresponding ssDNA signal (n≥3; mean ± SEM, * $p < 0.05$ by two-tailed unpaired t-test).

B. **Left panel.** Slot Blot analysis with S9.6 antibody in Control and HD MSN cells treated with RNase H enzyme. Nuclear DNA was extracted from cells, treated -/+ RNase H (1.7 U, NEB) for 2.5 h at 37ºC (prolonged to 4.5 h in an attempt to completely lose S9.6 signal in HD MSN cells, not shown) and analysed by Slot Blot, probed with S9.6 antibody. RNAse H treatment led to a complete loss of R-loop signal in Control cells confirming the specificity of the antibody, although partial R-loop signal is still detectable in + RNase H treated HD MSN cells. Slot blot probed with the antibody recognizing ssDNA was used as a loading control. The passage number of iPSCs used for MSN differentiation are indicated below the blot. **Right Panel.** Quantitation of the S9.6 Slot Blot signal. Values shown are normalized to the Control cell line and to the corresponding ssDNA signal (n≥3; mean ± SEM, * $p < 0.05$ **, $p < 0.01$ and **** $p < 0.001$ by two-tailed unpaired t-test).

4.3.1.2 DRIP-seq analysis in Control and HD cells

To investigate whether the global increase in R-loop levels observed by Slot Blot was also observed at the level of specific genes, RNA/DNA hybrid immunoprecipitation (DRIP)-seq with the S9.6 antibody was employed on nuclear DNA containing R-loops, extracted from Control (Control 1) and HD lymphoblast cells (40/50 CAG and 18/250 CAG). Confirming the results observed by the S9.6 Slot Blot analysis, the DRIP-seq detected more 'gained/enriched' DRIP peaks in both HD cell lines compared to Control 1 (Fig. 4.8A). Specifically, 1,984 and 7,254 'increased/gained' DRIP peaks were observed in 40/50 CAG HD and 18/250 CAG HD cells, respectively, compared to Control 1 (Fig. 4.8A). Only the changes within genic regions have been considered. A final list of ~500 'increased' peaks common in both HD cell lines compared to Control 1 was created. It must be taken into account that a lot of the increased DRIP peaks in HD cells, detected by DRIP-seq, were located in highly repetitive regions, complicating primer design. Finally, these results were validated by DRIP-qPCR analysis performed with an additional independent Control cell line. As an example, DRIP-seq profiles of the 'increased' *HIST1H1B* and *SFRF2* genes in Control (blue), and HD lymphoblast cells (pink and red) are shown in Fig. 4.8B (left panel). DRIP-qPCR validation with one additional control cell line (green) confirmed DRIP-seq results (Fig. 4.8B, right panel). Additional DRIP-seq profiles and qPCR validations for the 'increased' *DOCK* and *RNU2* genes are shown in Fig. 4.9A. Some genes, including *FGF3* and *NRX3N*, showed unchanged R-loop levels in Control and HD cells (Fig. 4.9B). Interestingly, the level of R-loop accumulation correlates with expansion length, with 18/250 CAG HD cells showing overall greater levels of R-loop accumulation at specific genes.

To understand whether R-loop accumulation at the level of specific genes could be detected in neuronal cells, DRIP-qPCR was performed in MSNs differentiated from iPSCs (Tabrizi lab), where a stronger global R-loop enrichment was observed by Slot Blot (Fig. 4.7). Representative individual gene results are shown for *HIST1H1B* and *RNU2,* which show increased R-loop levels in HD MSNs compared to Control (Fig. 4.10A). Other genes, like *γActin* and *ACTB*, showed unchanged R-loop levels in Control and HD cells (Fig. 4.10B).

Overall these results demonstrate R-loop accumulation at the level of specific genes in HD lymphoblasts and MSNs, suggesting a potential role for R-loop mis-regulation in HD pathology.

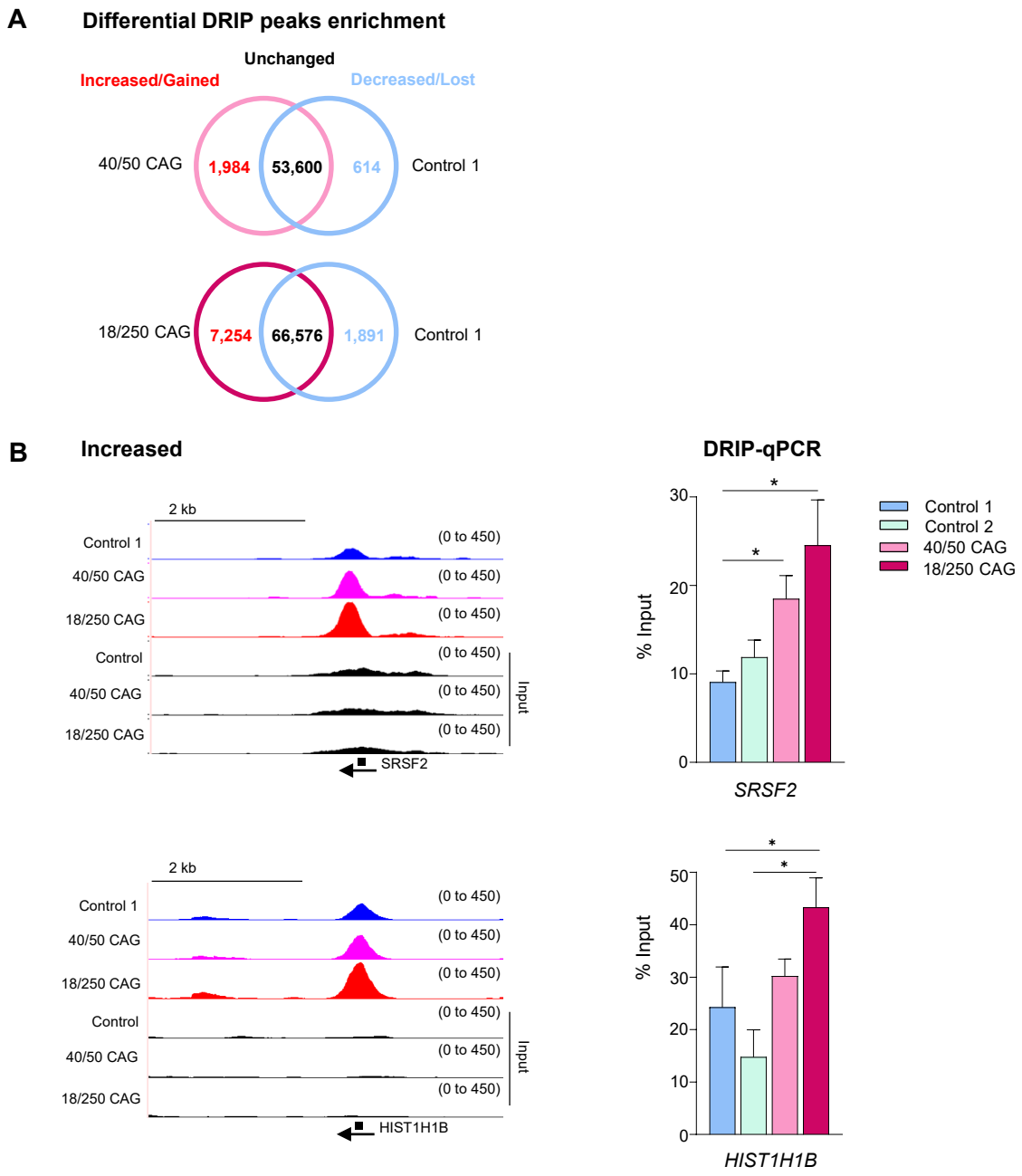


Figure 4.8. DRIP-seq and DRIP-qPCR analysis in HD lymphoblast cells.

A. Venn diagram showing the overlap between the differentially enriched DRIP peaks in Control 1 and HD cells following analysis of DRIP-seq (peak calling). Control 1 is shown in blue and HD cells (40/50 CAG and 18/250 CAG) are shown in pink and red, respectively. The number of DRIP peaks in each category is indicated in the Venn diagram. Unchanged peaks are represented in black, increased are shown in red and decreased in blue. A DRIP peak was considered as increased/decreased in Control or HD conditions when the change was statistically significant independently of the fold enrichment (>1.1 fold enrichment).

B. Screenshots of DRIP-seq profiles for 'increased' *SFRF2* (top) and *HIST1H1B* (bottom) genes in Control and HD cells. Independent quantitation by DRIP-qPCR is shown in the histograms to the right. The values are expressed as % of Input (n≥3; mean ± SEM, * $p < 0.05$ by two-tailed unpaired t-test). Numbers in brackets indicate the viewing range (FPKM).

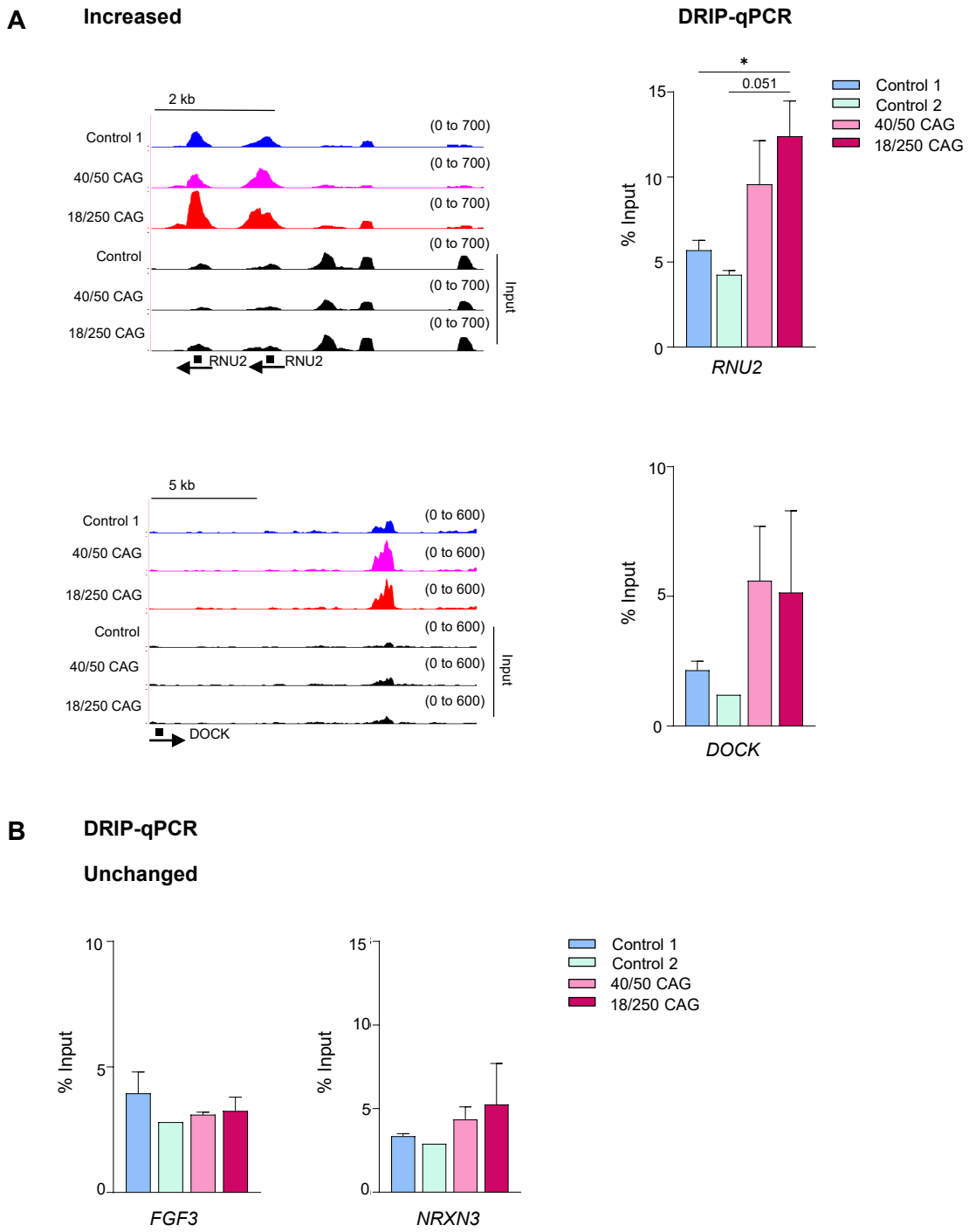


Figure 4.9. DRIP-seq analysis in HD lymphoblast cells.

A. Screenshots of DRIP-seq profiles for 'increased' *RNU2* and *DOCK* genes in Control 1 and HD cells (top and bottom, respectively). Independent quantitation by DRIP-qPCR is shown in the histograms to the right. The values are expressed as % of Input (n≥2; mean ± SEM, * $p < 0.05$ by two-tailed unpaired t-test). Numbers in brackets indicate the viewing range (FPKM).

B. Independent quantitation of unchanged *FGF3* and *NRXN3* genes by DRIP-qPCR. The values are expressed as % of Input (n≥2, apart for Control 2 n=1; mean ± SEM).

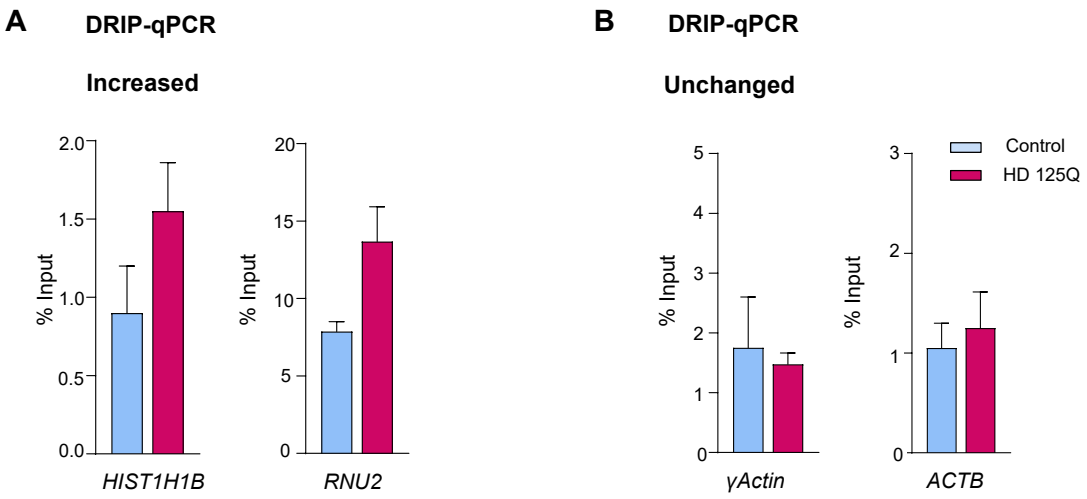


Figure 4.10. DRIP-qPCR analysis in HD MSN cells.

A. DRIP-qPCR in iPSC-derived MSNs for *HIST1H1B* and *RNU2* genes with increased R-loop levels. The values are expressed as % of Input (n≥3; mean ± SEM).

B. DRIP-qPCR in iPSC-derived MSNs for *γActin* and *ACTB* genes with unchanged R-loop levels. The values are expressed as % of Input (n≥3; mean ± SEM).

4.3.2 Genome-wide dsRNA levels in HD cells

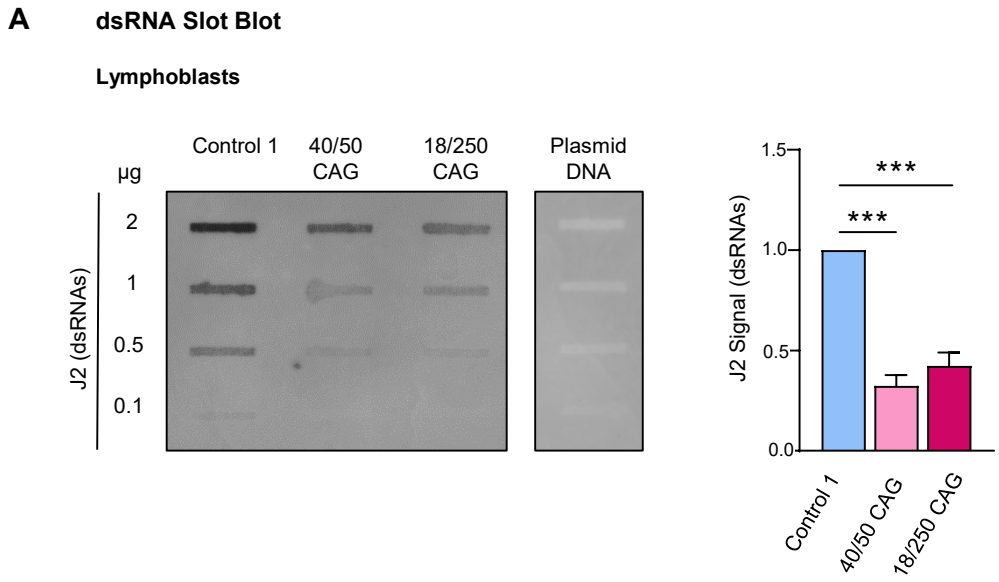
4.3.2.1 dsRNA Slot Blot shows global decrease of dsRNA in HD cells

As demonstrated in Fig. 4.3-4.4, expanded *HTT* transcripts can fold into dsRNA-like structures. Therefore, dsRNA enrichment was also investigated genome-wide in HD cells. Nuclear RNA from Control and HD lymphoblast cell lines was extracted and dsRNA global levels were analysed via Slot Blot assay, probed with the J2 antibody. As a negative control, plasmid DNA dilutions were loaded alongside the RNA samples. Interestingly, a ~60% reduction in dsRNA signal was detected in HD cells compared to Control (Fig. 4.11A). Importantly, the J2 signal was abolished by RNase III treatment, indicating that it was specific to dsRNAs (Fig. 4.11B). This reduction is in line with the transcription dysregulation observed in human and mouse HD models reported in this thesis (see Section 3.4 and 5.3) and in the literature (Hodges *et al.*, 2006; Ament *et al.*, 2017).

4.3.2.2 Establishment of the dsRNA immuno-precipitation method

As previously mentioned, dsRNA may play a role in HD pathogenesis (Chapters 1 and 4). More importantly, proteins can also regulate the cellular fate of dsRNAs and impact gene expression. Previously, a methodology to capture the mRNA interactome was developed to generate a comprehensive atlas of RNA-binding proteins (Castello *et al.*, 2012). However, besides some well-characterized dsRNA binders, a method to detect the dsRNA interactome in an unbiased way was still missing. Having observed that expanded *HTT* transcripts can fold into dsRNA-like structures *in vivo* (Fig. 4.3-4.4), the identification of dsRNA-interacting proteins could potentially shed a light on additional RNA-related pathogenic mechanisms in HD. Therefore, an affinity purification method was designed to purify dsRNA-binding proteins with the final goal of investigating their functions in health and disease.

To characterize the dsRNA interactome in human cells, lymphoblast cells from a healthy donor (Table 2.1, Control 1 line) were employed. Immuno-precipitation of dsRNA duplexes using the anti-dsRNA J2 antibody (Sciscon) was adapted and optimized from a previously published protocol, as described in Section 4.2.2 (Aktas *et al.*, 2017) (Fig. 4.12A).



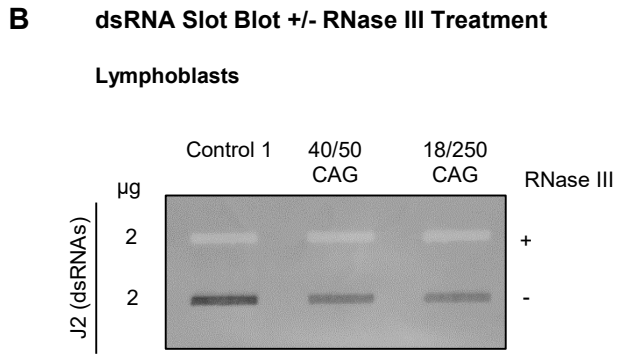


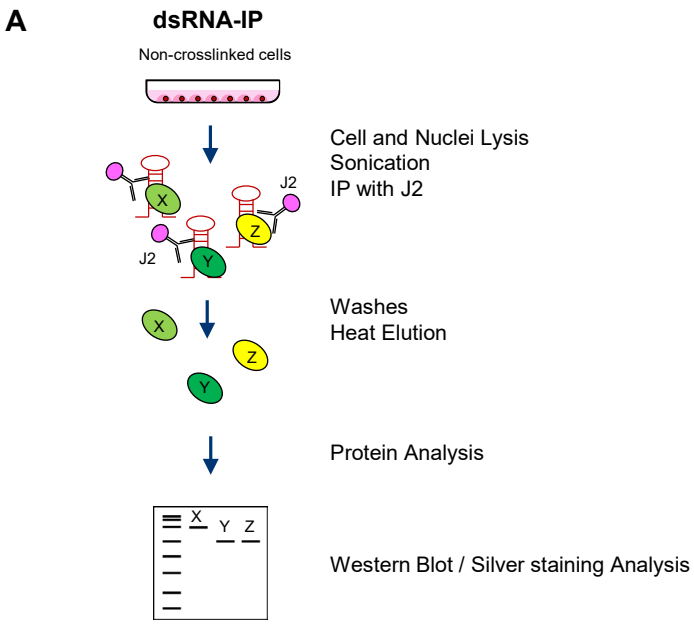
Figure 4.11. J2 Slot Blot analysis shows a decrease of dsRNAs in HD lymphoblast cells.

A. **Left panel.** Slot Blot analysis with J2 antibody in Control 1 and HD lymphoblast cells. Nuclear RNA was extracted from cells and analysed by Slot Blot, probed with J2 antibody. Serial dilutions of nuclear RNA from Control 1 and HD cells were blotted on the membrane (on the left) together with serial dilutions of plasmid DNA (negative control, on the right). **Right panel.** Quantitation of the J2 Slot Blot signal. Values for 1µg of nuclear RNA were normalized to the Control 1 (n=4; mean ± SEM, *** $p < 0.001$ by two-tailed unpaired t-test).

B. Slot Blot analysis with J2 antibody in Control 1 and HD lymphoblast cells treated with RNase III enzyme. 2µg of nuclear RNA was treated -/+ RNase III (NEB, 2U) for 1 h at 37ºC. RNAse III treatment led to a complete loss of dsRNA signal confirming the specificity of the antibody (n=1).

Silver staining of immuno-precipitated dsRNA-interacting proteins revealed a complex mixture of proteins, which differ significantly from the nuclear input material (Fig. 4.12B). Importantly, these proteins were not immuno-precipitated when the J2 antibody was omitted from the purification procedure, although some fainter bands could be detected in the silver staining. Similarly, when matched-mouse isotype control IgG antibody was used instead of J2 antibody, very few bands and much weaker staining intensity was observed (Fig. 4.12B). To optimize the procedure and to minimize the amount of unspecific binding observed in the 'IgG' and the 'No antibody' samples, the immuno-precipitation has been carried out with increasing salt concentration (75 mM and 150 mM NaCl). The increase in salt concentration to 150 mM NaCl did not improve the specificity of the method (Fig. 4.13A). However, the introduction of a pre-clearing step prior to immuno-precipitation did improve the specificity, leading to a modest decrease in non-specific binding in the control 'IgG' and 'No antibody' samples (Fig. 4.13B). Collectively, these results suggested that most of the proteins co-immuno-precipitated by J2 antibody were not due to unspecific binding to the beads or to the constant region of the antibody. Moreover, the presence of a different pattern of bands compared to the nuclear input suggested that the dsRNA-IP enriched for a unique set of interacting factors (Fig. 4.12-4.13A-B).

Although the knowledge of dsRNA-binding proteins *in vivo* in human cells is still limited, it was reasoned that dsRNA-IP would detect some candidates, previously implicated in dsRNA biology experimentally. Therefore, these candidates were tested using Western Blot analysis. For example, the antiviral dsRNA-binding protein kinase R (PKR), which has been demonstrated to bind *HTT* transcripts and be activated in HD tissues (Peel *et al.*, 2001), was immuno-precipitated (Fig. 4.13C). Moreover, DHX9 and RNA-specific adenosine deaminase 1 (ADAR1), well-known dsRNA binding proteins, were also identified (Fig. 4.13C). Overall, this data suggests that the RNA-IP approach is a powerful tool for identifying the dsRNA interactome.



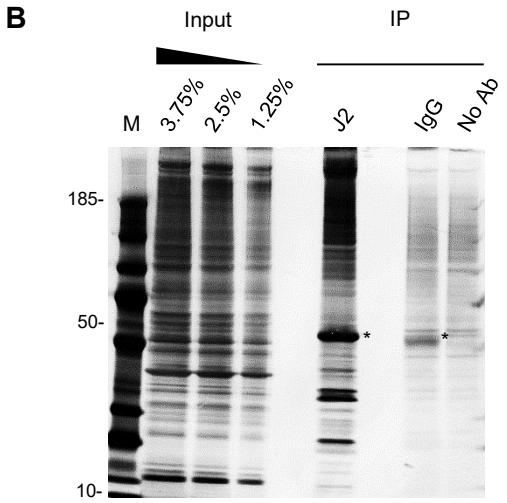


Figure 4.12. Establishment of the dsRNA-IP method.

A. Overview of the dsRNA-IP method. Initially, cell lysis is followed by nuclei lysis. dsRNAs are immuno-precipitated using J2 antibody, specific for dsRNA. Immuno-precipitation is followed by extensive washes and immuno-precipitated dsRNA-interacting proteins are eluted and analysed by silver staining and by Western Blot analysis.

B. Silver staining of protein samples from dsRNA-IP, following SDS-PAGE. 'No antibody' and 'IgG antibody' were used as controls. * indicates the heavy chain from the J2 and IgG antibodies. ½ of the IP has been loaded on the gel. The percentage of Inputs loaded on the gel is indicated at the top, and the marker (M) size is on the left.

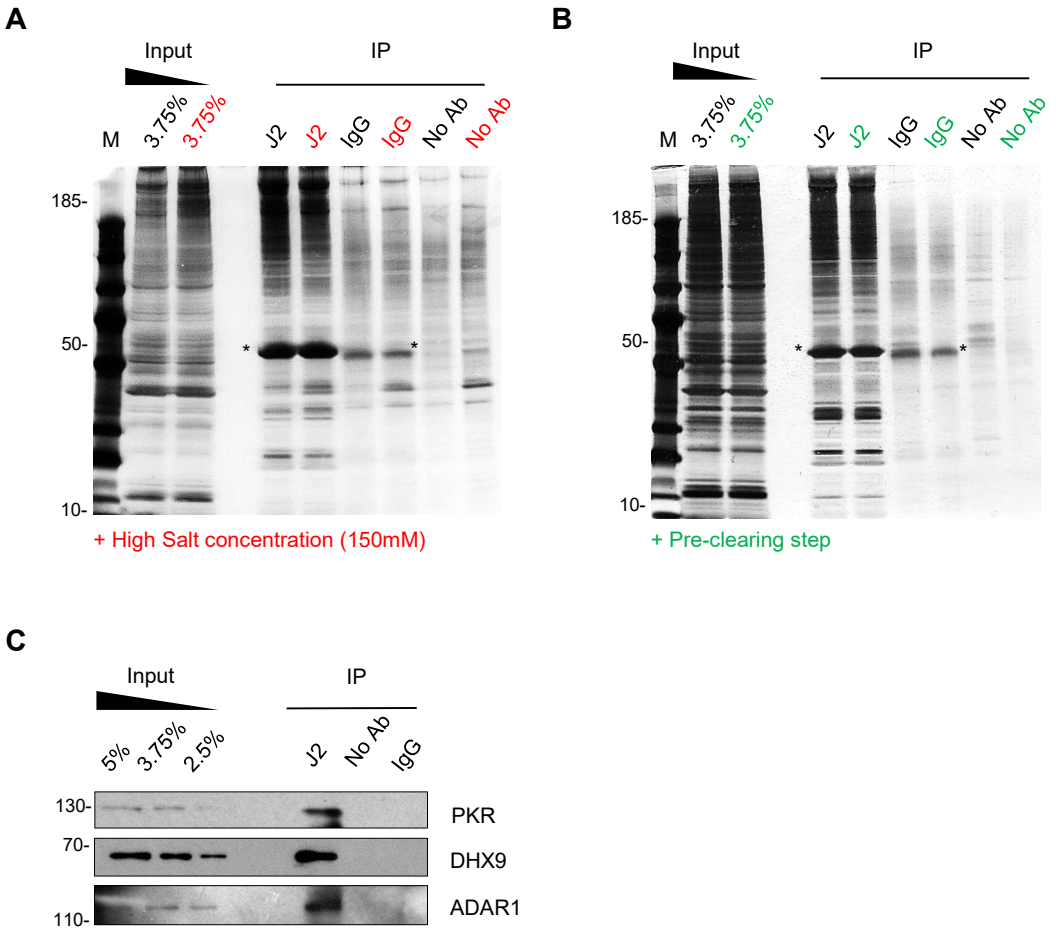


Figure 4.13. Optimization of the dsRNA-IP method.

A. Optimization of the dsRNA-IP method with increased salt concentrations (red, 150mM NaCl). Silver staining of protein samples from dsRNA-IP, following SDS-PAGE. Lanes 'No antibody' and 'IgG antibody' were used as controls. * indicates the heavy chain of the J2 and IgG antibodies. ½ of the IP has been loaded on the gel. The percentage of Inputs loaded on the gel is indicated at the top, and the marker (M) size is on the left.

B. Optimization of the dsRNA-IP method with the addition of the pre-clearing step (green). Silver staining of protein samples from dsRNA-IP, following SDS-PAGE. Lanes 'No antibody' and 'IgG antibody' were used as controls. * indicates the heavy chain of the J2 and IgG2a antibodies. ½ of the IP has been loaded on the gel. The percentage of Inputs is indicated at the top, and the marker (M) size is on the left.

C. Western Blot analysis of protein samples from dsRNA-IP using PKR, DHX9 and ADAR1 antibodies. ½ of the IP has been loaded on the gel. The percentage of Inputs is indicated at the top, and the marker (M) size is on the left.

To test if candidate dsRNA-binding proteins can bind the dsRNA formed over expanded *HTT* transcripts in HD cells (Fig. 4.3-4.4), ChIP-qPCR for DHX9 was performed using primers across the *HTT* gene (Fig. 4.14 and Table 2.2). DHX9 ChIP-qPCR did not show increased binding of DHX9 protein to expanded *HTT* compared to Control (Fig. 4.14). Instead, the levels of DHX9 protein were decreased across the expanded *HTT* gene in HD cells (Fig. 4.14). Furthermore, DHX9 levels were unchanged over the control *γActin* gene, suggesting that decreased levels of DHX9 were specific to the *HTT* gene (Fig. 4.14). This may suggest that DHX9 does not preferentially bind expanded CAG repeats of the *HTT* gene. This result correlates with the *HTT* gene transcriptional down-regulation observed in HD lymphoblast cells described in Chapter 3 and could be linked to the role that DHX9 plays in transcription regulation (Lee & Pelletier, 2016).

Additional efforts aimed at the characterization of the dsRNA interactome and the understanding of its contribution to HD will provide a link between dsRNA biology and disease pathogenesis.

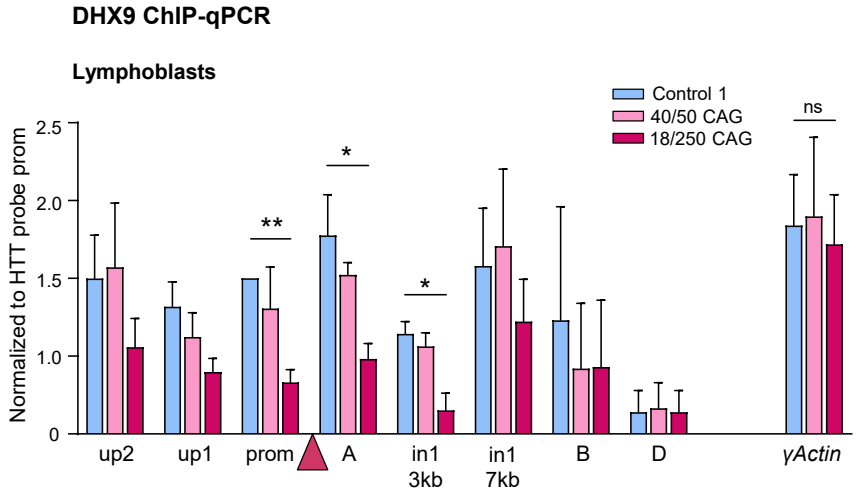


Figure 4.14. DHX9 levels are decreased over the *HTT* gene in HD lymphoblast cells.
DHX9 ChIP-qPCR on endogenous *HTT* gene in Control 1 and HD lymphoblast cells. The first intron of the *γActin* gene is used as a positive control. The values are normalized to *HTT* amplicon 'prom' in Control 1 cells (n=3; mean ± SEM, * $p < 0.05$ and ** $p < 0.01$ by two-tailed unpaired t-test).

4.4 Summary

The experiments described in this Chapter aimed to investigate the involvement of non-canonical RNA/DNA structures in HD pathology. The results demonstrate that R-loops are formed on the *HTT* gene, but not enriched over expanded repeats in lymphoblast and neuronal cells (Fig. 4.1-4.2). In contrast, dsRNAs are enriched over the expanded *HTT* transcripts, implicating them in HD pathogenesis (Fig. 4.3-4.4). In addition, a global accumulation of R-loops was detected in HD lymphoblast, fibroblast and MSN cells (Fig. 4.5 to 4.7). Interestingly, MSN HD cells, differentiated from iPSCs with higher passages, showed a stronger R-loop enrichment, raising a possibility of a potential influence of age-related genomic instability on the observed phenotype. Further analysis via DRIP-seq and qPCR revealed a subset of specific genes with increased R-loop levels in HD lymphoblast and MSN cells (Fig. 4.8 to 4.10).

As proteins can regulate the cellular fate of dsRNAs, an affinity purification method was designed to purify dsRNA-binding proteins with a final goal of investigating their function in health and disease (Fig. 4.12-4.13).

Further investigation of the role of the HTT protein in the nucleus and its potential involvement in R-loop regulation and other HD phenotypes, including DNA damage and inflammation, is described in Chapter 5.

Chapter 5

The HTT protein is involved in transcription and R-loop biology

5.1 Introduction

Wild-type HTT (wtHTT) protein is predominantly found in the cytoplasm. However, it has the ability to shuttle between the cytoplasm and the nucleus (Schulte & Littleton, 2011). Nonetheless, the function of HTT protein in the nucleus remains largely unknown. Interestingly, it has been shown that mis-localization of mutant HTT (mHTT) protein to the nucleus, alongside transcriptional and epigenetic changes, are among the first phenotypes observed in several HD models (Hodges *et al.*, 2006; Langfelder *et al.*, 2016; Bassi *et al.*, 2017).

A combination of reduced HTT protein levels (as seen in Chapter 3, Fig 3.11), and the presence of mHTT in the nucleus could contribute to the transcriptional dysregulation and the R-loop accumulation observed in different cellular HD models, as described in Chapters 3 and 4. Therefore, in this Chapter, I have investigated the role of HTT protein in the nucleus and its potential involvement in R-loop biology. Cells transiently depleted for HTT protein using siRNA-mediated knock-downs were used in combination with cells from HD patients in order to study the role of wtHTT and mHTT proteins. R-loops have been shown to compromise genome stability by interfering with both transcription and DNA damage repair processes (Crossley *et al.*, 2019). Therefore transcription, DNA damage and activation of immune responses have been analysed in HTT-depleted HeLa cells and HD cells from patients to dissect the potential interconnections between these regulatory mechanisms.

5.2 The role of the HTT protein in the nucleus

5.2.1 Full-length HTT protein is present in the nucleus and localises to gene promoters

Uncovering the function of wild-type and mutant huntingtin protein in the nucleus is important for understanding HD pathogenesis. Therefore, the localization of HTT protein was investigated in HeLa and SH-SY5Y cells by cellular fractionation, followed by Western Blot analysis. A fractionation protocol, which enabled the efficient separation of cytoplasm (C), nucleoplams (NP) and chromatin (Chr) fractions, was optimized and employed in HeLa and SH-SY5Y cells (Senichkin *et al.*, 2021) (Fig 5.1A). The purity of the resulting cellular fractions was

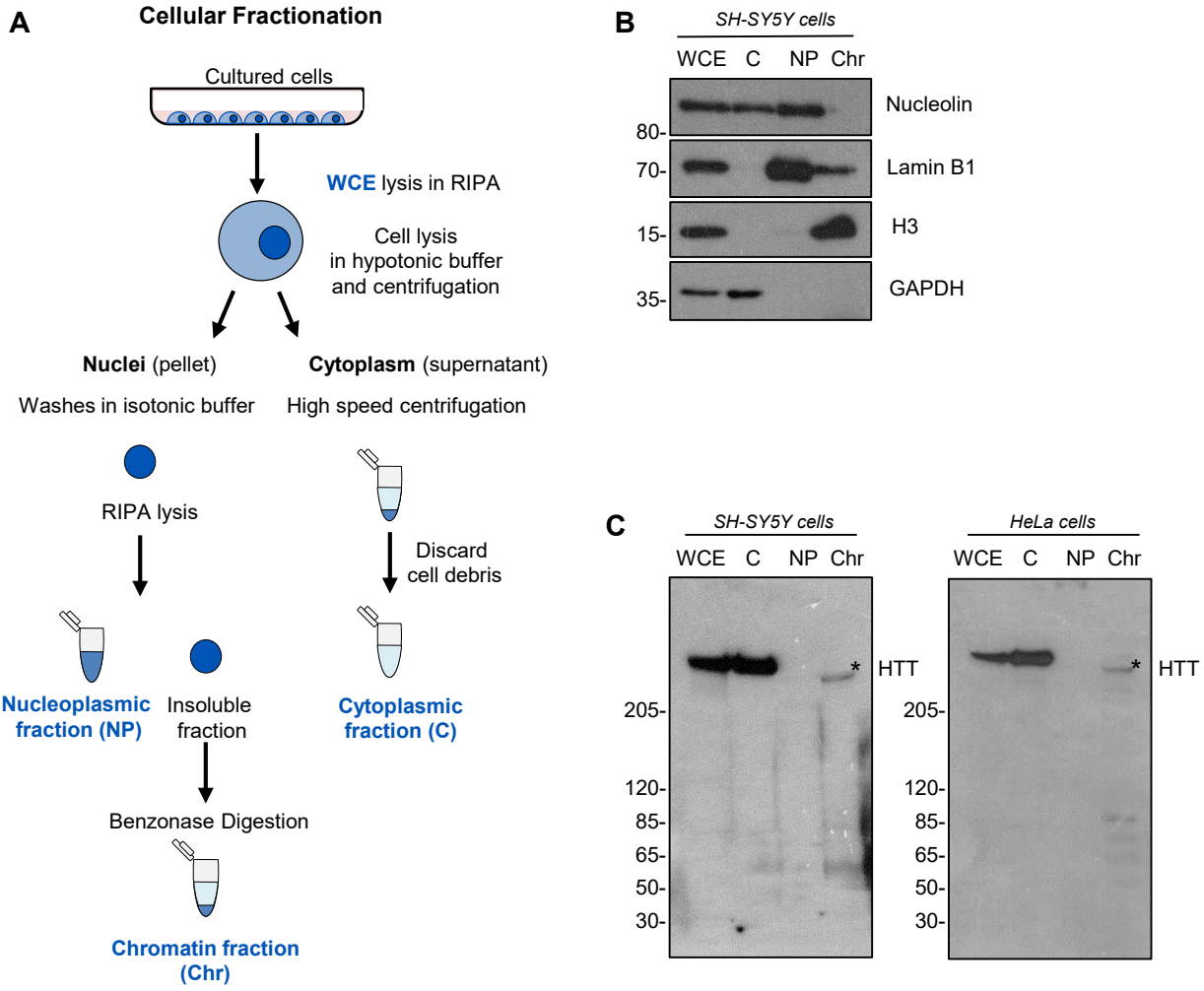


Figure 5.1. The HTT protein is localized in the nucleus.

A. Graphical representation of the cellular fractionation method. WCE is extracted with RIPA buffer. For subcellular fractionation, cells are harvested and washed with PBS, centrifuged and resuspended in hypotonic buffer. The lysate is centrifuged to separate nuclei (pellet) and cytoplasm (supernatant). Then, the cytoplasm (C) fraction is centrifuged at high speed to remove cell debris. Nuclei are then washed in isotonic buffer and centrifuged again. Nuclei are then incubated in RIPA buffer for 20 min and centrifuged. The supernatant corresponds to nucleoplasmic sub-fraction (NP), whereas the pellet is incubated with benzonase for 10 min at 37ºC to degrade all nucleic acid species, resulting in the chromatin fraction (Chr).

B. Validation of the cellular fractionation protocol for SH-SY5Y cells. Western Blot analysis of 20 µg of protein extracts obtained from cytoplasmic and nuclear fractions. The purity of the resulting fractions was assessed by staining for the markers of the cytoplasm (GAPDH), nuclear envelope (Lamin B1), nucleoplasm (Nucleolin) and chromatin (H3). Fractions abbreviations are: Whole Cell Extract (WCE), Cytoplasm (C), Nucleoplasm (NP), and Chromatin (Chr). Marker size is found on the left and antibodies are indicated on the right.

C. Western Blot analysis of 20 µg of protein extracts obtained from cytoplasmic and nuclear fractions in SH-SY5Y (left) and HeLa (right) cells and probed with MAB2166 HTT antibody. Full-length huntingtin was detected in WCE and C, but also detectable in Chr fraction (indicated with *). Smaller HTT fragments were also detected in the nuclear fraction upon higher exposure (only some are visible on the present gel, ~100 kDa and ~50 kDa).

assessed by detection of markers of cytoplasm (GAPDH), nuclear envelope (Lamin B1), chromatin (H3) and nucleoplasm (nucleolin) using Western Blot analysis (Fig 5.1B).

The majority of the HTT protein was localized in the cytoplasm of SH-SY5Y and HeLa cells (Fig 5.1C, fraction C). The signal for full-length wtHTT protein was significantly lower in the nuclear fraction, and detectable as a faint band in the chromatin fraction (Fig 5.1C, fraction Chr), in line with previous studies (Kegel *et al.*, 2002; Atwal *et al.*, 2007). The molecular weight of the HTT band in the Chr fraction is slightly lower than that of the WCE and cytoplasm fractions (Fig 5.1C). This could be due to post-translational modifications of HTT protein in the nucleus or to an artefact of the gel. Furthermore, both SH-SY5Y and HeLa cells had several smaller HTT fragments in the nucleus (Fig 5.1C, fraction Chr), in agreement with previous work (Kegel *et al.*, 2002).

To date, very few studies have attempted to map the genomic occupancy of the HTT protein (Benn *et al.*, 2008; Pearl *et al.*, 2020). Therefore, to determine whether HTT binds human genes, ChIP-qPCR assay using MAB2170 HTT antibody was employed in HeLa cells. To test the specificity of this antibody, the experiments were carried out in siLuc or siHTT-treated HeLa cells. ChIP-qPCR confirmed HTT binding over the promoters of actively transcribed genes including *HTT*, *ACTB*, *γActin*, and *DDX1* (Fig 5.2A). In contrast, *TFF1*, which is not expressed in HeLa cells, displayed only background levels of HTT binding (Fig 5.2A). In addition, HTT binding was decreased following siHTT-mediated depletion, indicating the specificity of the ChIP signal (Fig 5.2A, purple bars). The knock-down efficiency in these cells was estimated to be ~70% on protein levels, based on Western Blot analysis (Fig. 5.2B), and ~50% on RNA levels, based on RT-qPCR analysis (Fig. 5.2C). Since HTT is a ~360 kDa protein and it is relatively stable, its full depletion is technically challenging. Moreover, the efficiency of knock-down has been evaluated only in the WCE and it is likely to be less efficient in the nuclear fraction. However, the observed decrease in the HTT ChIP-qPCR signal in siHTT-treated cells indicates HTT depletion in the nucleus (Fig 5.2A, purple bars). Overall these results suggest that HTT protein can bind gene promoters *in vivo*. Hence, perturbation of HTT-chromatin interactions either due to the reduction of wtHTT protein in the nucleus or due to HD mutations could contribute to the transcriptional dysregulation observed in HD.

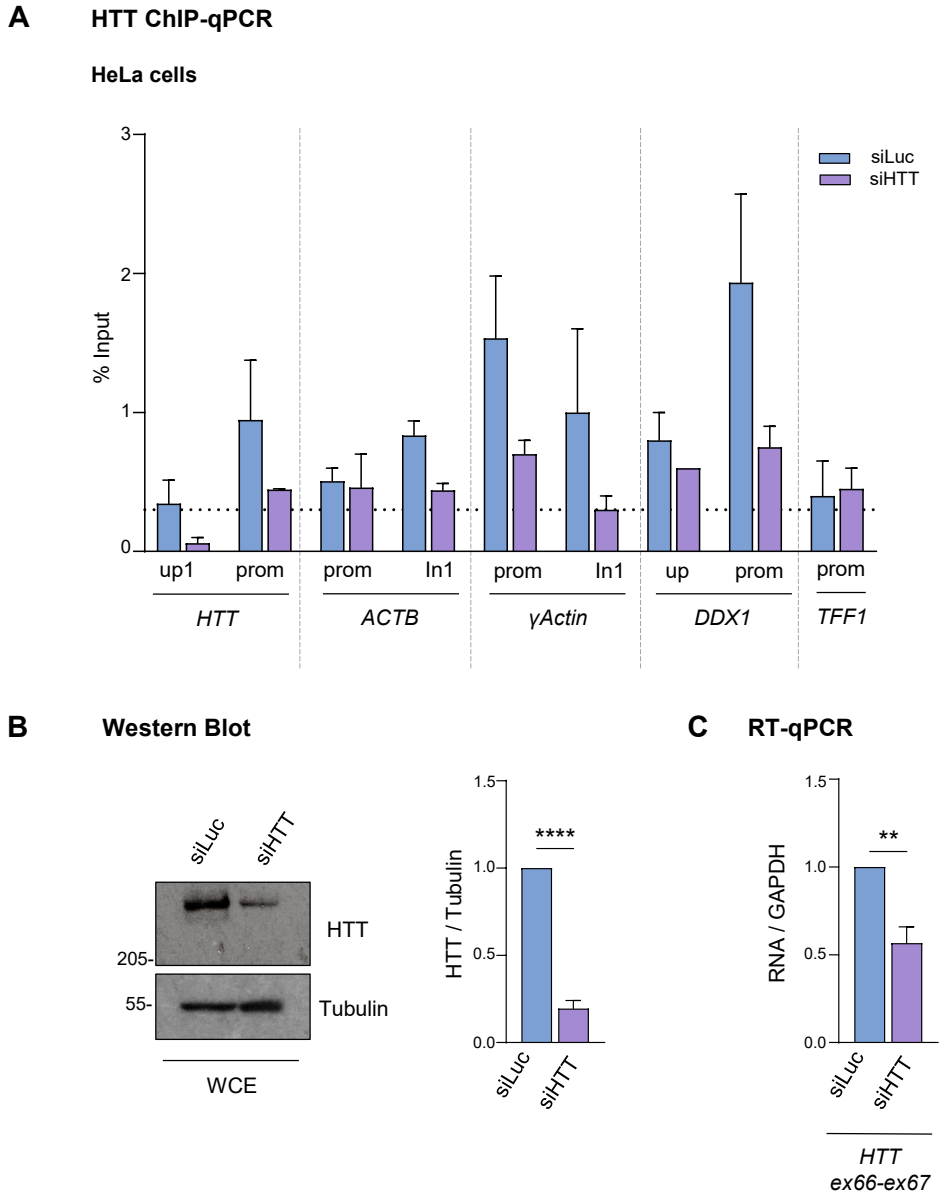


Figure 5.2. The HTT protein binds to gene promoters.

A. HTT ChIP-qPCR analysis for *HTT*, *ACTB, γActin* and *DDX1* genes in HeLa cells transfected with siLuc (blue bars) and siHTT RNAs (purple bars). The non-expressed *TFF1* gene is used as a negative control. Horizontal dotted line indicates background signal. The values are expressed as % of Input (n>2; means± SEM).

B. **Left panel.** Western Blot analysis of 20 µg of protein extracts obtained from the whole cell extract (WCE) of HeLa cells transfected with siLuc or siHTT RNAs and probed with MAB2166 anti-HTT antibody. Tubulin is a loading control. **Right panel.** Quantitation of protein levels. Values are normalized to tubulin (n=7; means± SEM, **** $p < 0.0001$ by two-tailed unpaired t-test).

C. RT-qPCR analysis of *HTT* mRNA levels (amplicon 'ex66-ex67', see gene scheme in 3.1A). Values are normalized to *GAPDH* (n=5; means± SEM, ** $p < 0.01$ by two-tailed unpaired t-test).

5.2.2 HTT interacts with transcriptional and R-loop binding factors

In order to investigate the role of HTT protein in transcriptional processes and its potential involvement in R-loop biology, co-IP experiments of endogenous HTT protein were performed from the WCE of HeLa cells. Two HTT-specific antibodies were purchased (MAB2166 and MAB2170, Sigma-Aldrich) and employed in co-IP experiments (Fig 5.3A). Initially, immuno-precipitated HTT-interacting proteins were analysed by SDS-PAGE followed by silver staining. Using MAB1266, which recognises 181-810 aa fragment, near the N-terminal end of the HTT protein, very few proteins were pulled-down (Fig. 5.3B), despite ~750 known interactors of HTT (Shirasaki *et al.*, 2012). In contrast, MAB2170, which recognizes 1247-1646 aa within HTT, led to an efficient pull-down of a complex mixture of HTT-interacting proteins (Fig. 5.3B). The low efficiency of pull-down using MAB2166 may be due to the location of its epitope on HTT protein (Fig 5.3A). This N-terminal part of the HTT protein is important for the binding of HTT interactors and forms intramolecular interactions which determine the overall conformation of HTT (Saudou & Humbert, 2008; Guo *et al.*, 2018). Hence, MAB2166 antibody could directly obscure binding sites for HTT interactors, or alternatively, affect binding by disrupting the HTT structure. Therefore, co-IP experiments were carried out using MAB2170 antibody, whereas MAB2166 was employed for Western Blot analysis (Fig 5.3A). The use of different antibodies for IP and Western Blot experiments increases the likelihood that the species detected represent HTT protein rather than non-specific bands. Co-IP experiments were performed in the presence of benzonase which degrades all nucleic acids but still retains protein-mediated interactions.

Using co-IP, ~360 kDa full-length HTT protein was efficiently pulled down (Fig. 5.3C). Alongside full-length HTT protein, a smaller band at ~100 kDa and two other bands at ~55 kDa and ~25 kDa were detected. It is likely that the bands at ~55 and ~25 kDa are due to background cross-reactivity between the secondary antibody and the heavy and light chains of the MAB2170 antibody, respectively (Fig 5.3C). Western Blots of the immuno-complexes revealed novel interactions between HTT and several factors with known roles in transcription and R-loop biology (Fig 5.4A). The interaction between HTT and Pol II, already reported in the literature (Gao *et al.*, 2019), was observed and used as a positive control. Interestingly, an enrichment for transcription-associated

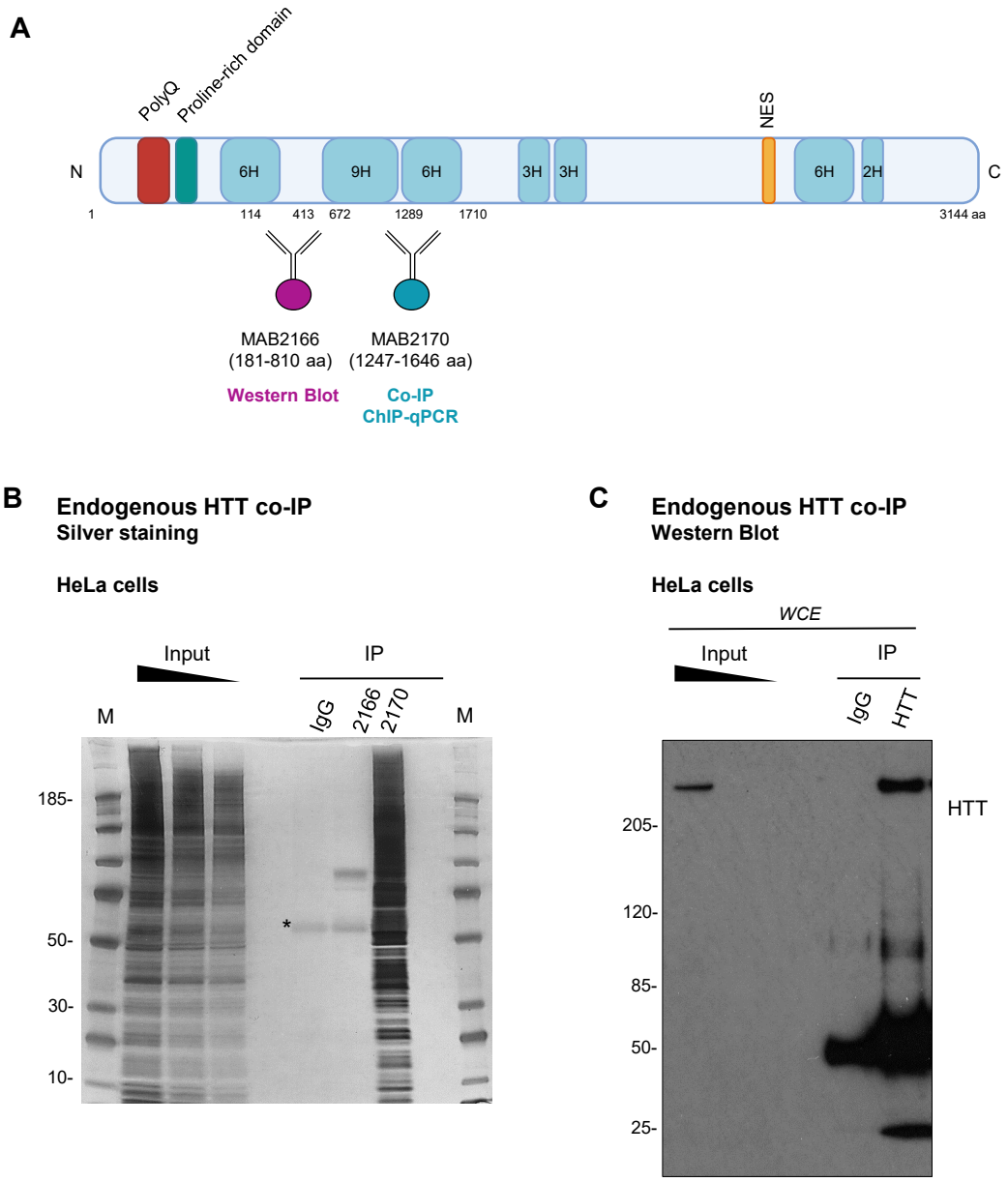


Figure 5.3. Endogenous HTT co-IP optimization.

A. Schematic of the HTT protein and HTT antibodies used in this study. The HTT protein is 3,144 amino acids (aa) long and contains several HEAT repeats (H), important for protein-protein interactions. HTT epitopes recognised by MAB2166 and MAB2170 monoclonal antibodies are highlighted.

B. Silver staining of protein samples from the endogenous HTT co-IP with MAB2166 and MAB2170 antibodies, following SDS-PAGE. Isotope-matched IgG antibody is used as a negative control. * indicates the heavy chain from the HTT and IgG antibodies. ½ of the IP has been loaded on the gel. Marker (M) size is on the left.

C. Western Blot analysis of endogenous HTT co-IP using MAB2170 HTT antibody from HeLa WCE probed with MAB2166 HTT antibody. Full-length gel showing the efficiency of the pull-down. The full-length HTT protein is detected in the HTT-IP and not in the IgG IP sample (negative control). Additional bands are detected at ~100 , ~55 and ~25 kDa. Representative blot from n>4 biologically independent experiments. ½ of the IP has been loaded on the gel. Marker size is on the left.

Figure generated with the help of Rob Torrance.

forms of Pol II, such as initiating (Ser5P) and elongating (Ser2P), was also observed (Fig 5.4A). Moreover, interactions between HTT and several known R-loop factors, particularly those involved in R-loop resolution, were also detected. These included the endonuclease and DNA damage repair factor XPG, the 5'-3' exoribonuclease XRN2, the RNA/DNA helicase DHX9, and DNA topoisomerase I (Top1) (Fig 5.4A). In contrast, Proliferating Cell Nuclear Antigen (PCNA) was not detected, suggesting a specificity of observed interactions and therefore it was used as a negative control (Fig 5.4A). Furthermore, some of these interactions were also detected in fibroblast cells (Fig 5.4B).

These data suggest that these proteins together with HTT can form a multi-protein complex that can act as a scaffold for R-loop resolution. However, it should be noted that the co-IP experiments have been performed in WCE and further experiments in nuclear extracts are needed to confirm that these interactions occur also in the nucleus.

5.2.3 HTT forms a complex with RNA/DNA hybrids *in vivo*

To examine the potential role of HTT protein in R-loop biology, the next step was to determine whether HTT binds R-loops *in vivo*. Recently, our lab has designed an R-loop IP method which allows to detect the interaction between proteins of interest and R-loops (Cristini *et al.*, 2018; Beghe & Gromak, 2022). This technique relies on the pull-down of R-loops, with R-loop interacting proteins, from cell nuclear extracts using S9.6 antibody, followed by Western Blot or Mass Spectrometry. Initial attempts to detect full-length HTT protein in HeLa cells by R-loop IP were unsuccessful most probably due to low levels of HTT in the nucleus of HeLa cells (Fig. 5.1C). Indeed, several papers have shown that HTT has a low abundance in the nuclei of non-neuronal cells (De Rooij *et al*, 1996; Kegel *et al.*, 2002; Didiot *et al.*, 2018). Consequently, in order to enrich for nuclear HTT protein, a reverse R-loop IP was performed. Here, HTT protein was initially enriched by IP with the MAB2170 HTT antibody, and then HTT-bound nucleic acids were extracted and analysed via Slot Blot probed with S9.6 antibody (Fig 5.5A). Using this approach, R-loops were immuno-precipitated with HTT protein both in the nuclear (NE) and whole cell extracts (WCE) of HeLa cells (Fig 5.5B, HTT-IP samples). R-loops were not detected in the control IgG-IP

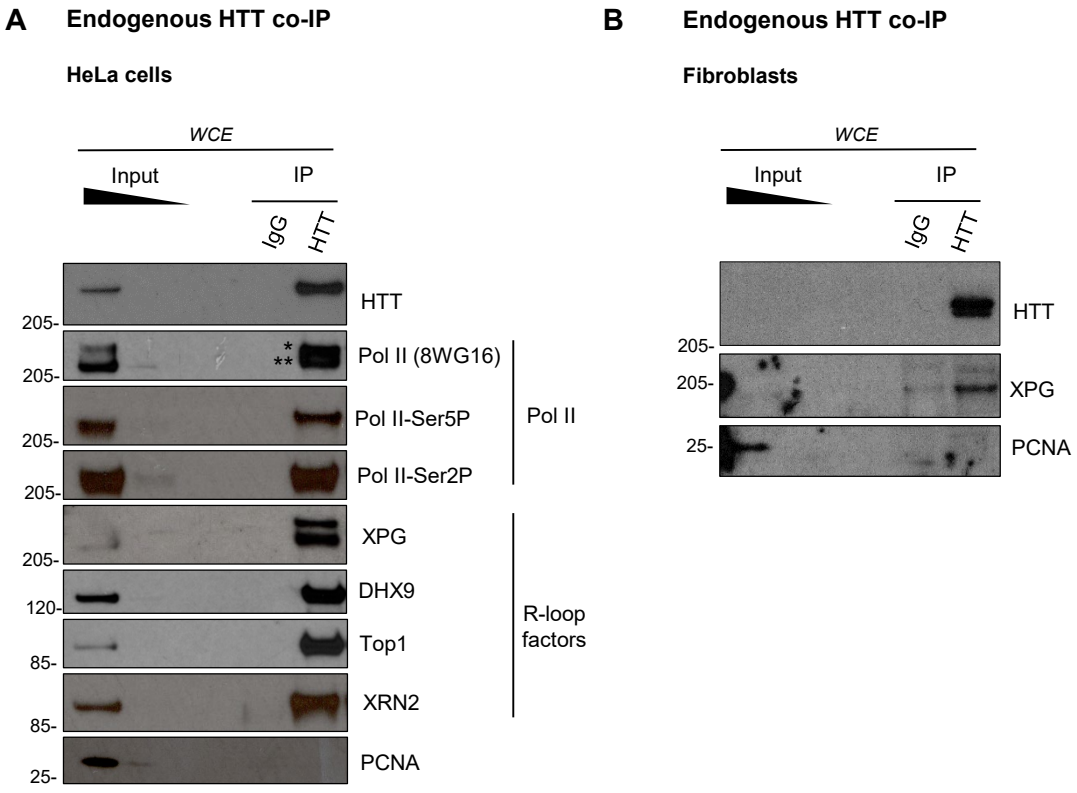


Figure 5.4. The HTT protein is in complex with R-loop and transcriptional factors.

A. Western Blot analysis of endogenous HTT co-IP from HeLa WCE probed with indicated antibodies. Asterisks indicate hypo (IIa)** and hyper-phosphorylated (IIo)* Pol II. IgG and PCNA are used as negative controls. ½ of the IP has been loaded on the gel. Marker size is on the left. Representative blot from n>4 biologically independent experiments. Figure generated with the help of Rob Torrance.

B. Western Blot analysis of endogenous HTT co-IP from fibroblast WCE probed with indicated antibodies. IgG and PCNA are used as negative controls. ½ of the IP has been loaded on the gel. Marker size is on the left (n=1).

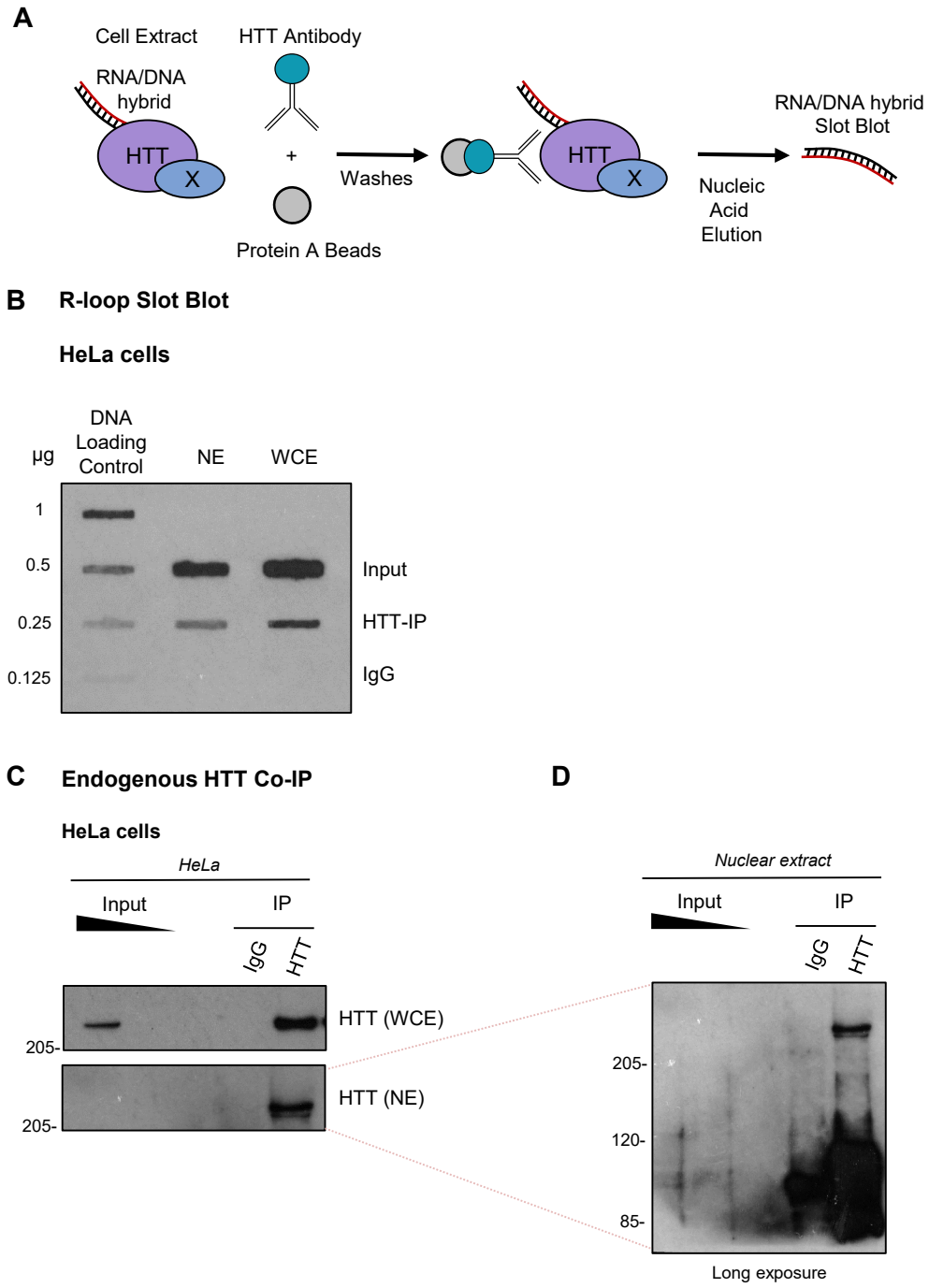


Figure 5.5. The HTT protein binds to R-loops *in vivo*.

A. Schematic of the reverse RNA/DNA hybrid IP procedure. Following HTT co-IP from nuclear extract (NE) or WCE, nucleic acids bound to HTT are extracted and loaded on a Slot Blot probed with S9.6 antibody, for R-loop detection.

B. S9.6 Slot Blot for detection of R-loops in NE and WCE following HTT co-IP in HeLa cells. Genomic DNA dilutions (1-0.125µg) were used as a loading control. IgG is a negative control.

C. Western Blot analysis of endogenous HTT co-IP from HeLa WCE and NE probed with anti-HTT antibody. The full-length HTT is pulled down from WCE and NE. IgG is anegative control.

D. Longer exposure and full view of the Western Blot analysis in NE in panel C. A strong additional band in the HTT-IP sample is present at ~120 kDa.

supporting the specificity of the HTT-R-loop interaction (Fig. 5.5B). Western Blot analysis of the HTT-IP samples in nuclear extracts demonstrated an enrichment of the full-length HTT protein (Fig 5.5C, NE). However, a high level of background signal was detected at ~120 kDa in these nuclear extract samples, which may correspond to nuclear HTT fragments (Fig 5.5D). Indeed, HTT has been shown to undergo proteolysis in the nucleus and HTT fragments have been associated with HD pathology (Saudou & Humbert, 2008). However, further work is required to determine the exact nature of these ~120 kDa nuclear species. Collectively, these data show that HTT interacts with R-loops *in vivo* in HeLa cells.

5.3 Global transcription is impaired in HD cells

5.3.1 Global nascent transcription is down-regulated in HD fibroblasts

HTT protein was found to be in complex with different transcriptional and R-loops factors (Fig 5.4). Therefore, it was addressed whether transcription was affected in HD cells. Here, nascent RNA synthesis was quantified using 5-ethynyl uridine (EU) labelling in Control and HD fibroblast cells. This analysis revealed a significant decrease in nascent global transcription in HD fibroblast cells compared to the Control cells (Fig 5.6A-B). Further analysis revealed transcriptional down-regulation both in the nucleoplasm, which mainly reflects nascent Pol II transcription, and nucleolus, associated with rDNA transcription (Fig 5.6C). These observations are consistent with HTT interacting with Pol II as demonstrated in HTT co-IP experiments (Fig 5.4) and as shown in previous studies (Gao *et al.*, 2019). Furthermore, these data are in line with a previously described role of the HTT protein in nucleolar transcription (Lee *et al.*, 2011; Tsoi *et al.*, 2012).

Finally, the transcription down-regulation observed by EU analysis in HD fibroblast cells is in line with the mouse transcriptomic data presented in Section 3.4, showing ~20% decrease in Pol II levels in promoter regions in HD Q140 mice compared to WT (Fig 3.10B). In addition, this result is consistent with the global reduction in nuclear dsRNA levels observed in HD lymphoblast cells (Fig. 4.11). Overall, these results confirm global transcriptional down-regulation in HD, supporting this conclusion in different cellular and animal models.

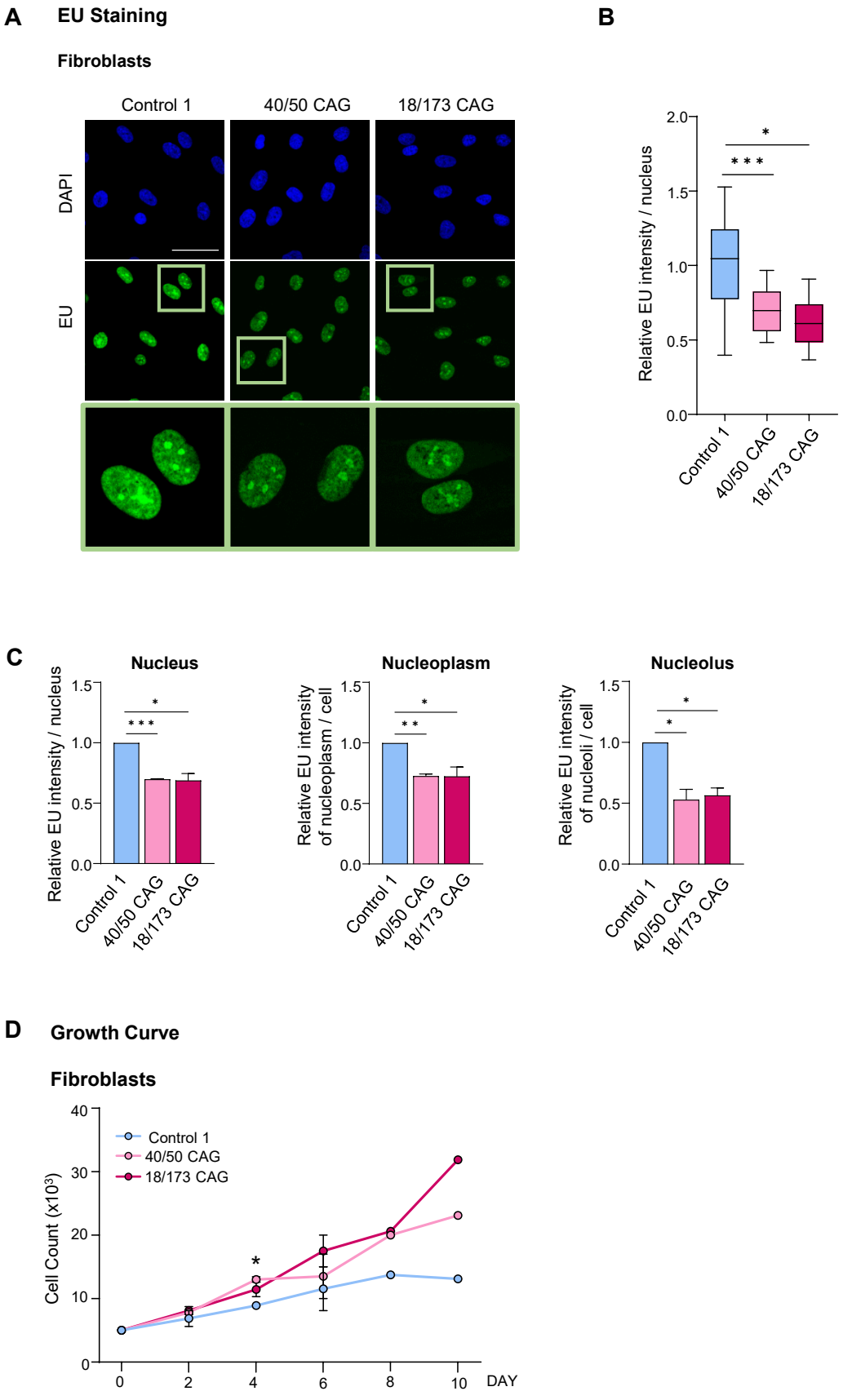


Figure 5.6. Global transcription is down-regulated in HD fibroblast cells.
The caption of the figure can be found below.

Figure 5.6 (continued). Global transcription is down-regulated in HD fibroblast cells.

A. IF analysis of 5'-ethynyl uridine (EU) incorporation in Control 1 and HD fibroblasts. EU (green) and DAPI (blue). Scale bar: 50 μm. Representative images from n=2 biologically independent experiments. The three green squares represent enlarged views.

B. Quantification of EU intensity per nucleus in Control 1 and HD fibroblasts. EU is normalized to the average intensity in the Control 1 condition. Boxplot settings are: box: 25-75 percentile range; whiskers: 10-90 percentile range; horizontal bars: median. >50 nuclei were quantified per condition per experiment (n=2; means± SEM, *$p < 0.05$ and *** $p < 0.001$ by two-tailed unpaired t-test).

C. Quantification of EU intensity per nucleus (left panel), nucleoplasm/cell (nucleus without nucleolus; middle panel) or nucleoli/cell (right panel) in Control 1 and HD fibroblasts. EU intensity is normalized to the average intensity in the Control condition. >50 nuclei were quantified per condition per experiment (n=2; means± SEM, * $p < 0.05$ and ** $p < 0.01$ *** $p < 0.001$ by two-tailed unpaired t-test).

D. Growth curve of Control 1 and HD fibroblast cells. The cell count was measured every 48 h. Days are indicated on the x-axis and the cell count is on the y-axis (n=2, n=1 for day 8 and 10; means± SEM, * $p < 0.05$ by two-tailed unpaired t-test).

Transcription can affect DNA replication progression, inducing genome instability (Saponaro, 2022). The accumulation of R-loops, transcriptional stress and DNA damage can in turn also impact DNA replication. Therefore, cell proliferation of Control and HD fibroblast cells was monitored over a 10-day period. Interestingly, the growth curve showed that HD fibroblasts grew faster following 4 days in culture, while the Control cells grew slower and reached a relative plateau after 6 days (Fig 5.6D). These observations are in line with previous studies (Hung *et al.*, 2018) and consistent with reports demonstrating aberrant cell-cycle regulation and DNA damage repair in HD cells (Hoffner *et al.*, 2002; Maiuri *et al.*, 2017).

5.4 HTT is involved in the DNA damage response

The accumulation of DNA damage has been described in HD cells and has typically been attributed to defects in DNA damage repair (Massey & Jones, 2018; Gao *et al.*, 2019; GeM-HD, 2019). Indeed, HTT has been suggested to play a critical role in the transcription-coupled DNA repair (TCR) process (Gao *et al.*, 2019; Pradhan *et al.*, 2022). However, it is possible that DNA damage in HD cells may also arise due to R-loop accumulation and transcriptional impairment described in this Thesis (Chapter 4 and Fig. 5.6A-B-C). Indeed, R-loop accumulation has previously been linked to increased DNA damage in cells, in part due to the presence of unstable single-stranded DNA (ssDNA) in the R-loops (Skourti-Stathaki & Proudfoot, 2014; Sollier *et al.*, 2014). Moreover, transcription-associated breaks can be particularly harmful as they occur in transcribed regions, and they can drive genome instability (Cristini *et al.*, 2021). However, a mechanistic explanation linking impaired transcription and R-loop accumulation to DNA damage in HD is still lacking. Therefore, an investigation of DNA damage phenotypes and the possible link with R-loop accumulation, as observed in HD cells, was further explored below.

5.4.1 Depletion or mutation of HTT protein leads to DNA damage accumulation

The DNA damage phenotypes were analysed in siHTT-depleted HeLa cells and HD cells from patients. Western Blot analysis of the phosphorylated histone variant H2AX (γH2AX), a marker of DNA damage, revealed an increase in DNA damage in siHTT-depleted HeLa cells relative to the

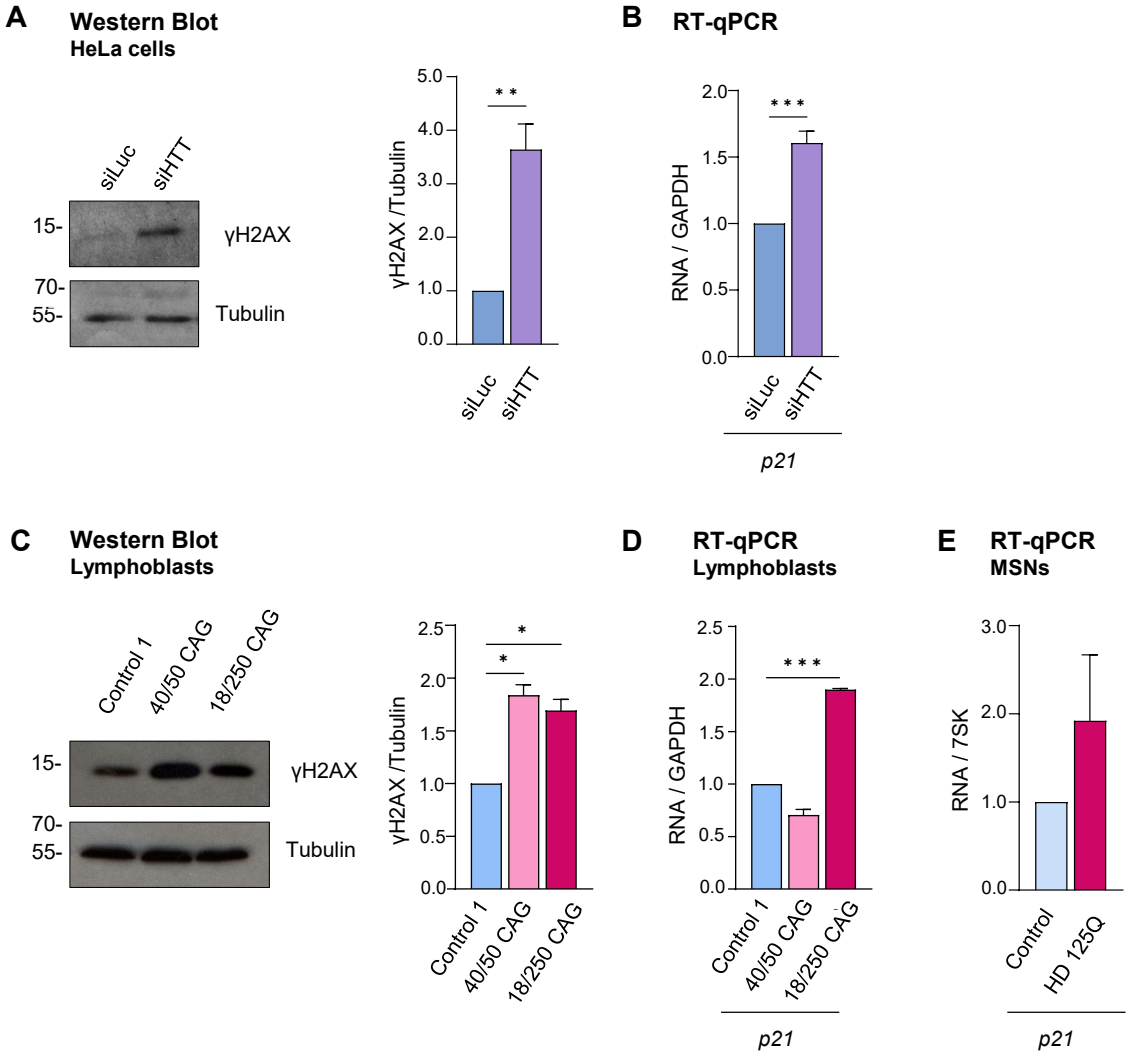


Figure 5.7. Accumulation of DNA damage in different HD models.

A. **Left panel**. Western Blot analysis of γH2AX performed with 20 µg of protein extracts obtained from HeLa cells transfected with siLuc (blue bars) or siHTT RNAs (purple bars). Tubulin is a loading control. **Right panel.** Quantitation of γH2AX protein levels. Values are normalized to tubulin (n=3; means± SEM, ** $p < 0.01$ by two-tailed unpaired t-test). Figure generated with the help of Rob Torrance.

B. RT-qPCR analysis of *p21* mRNA levels in HeLa cells transfected with siLuc (blue bars) or siHTT RNAs (purple bars). Values are normalized to *GAPDH* (n=3; means± SEM, *** $p < 0.001$ by two-tailed unpaired t-test).

C. **Left panel.** Western Blot analysis of γH2AX performed with 20 µg of protein extracts obtained from Control 1 and HD lymphoblast cells. Tubulin is a loading control. **Right panel.** Quantitation of γH2AX protein levels. Values are normalized to tubulin (n=2; means± SEM, * $p < 0.05$ by two-tailed unpaired t-test).

D. RT-qPCR analysis of *p21* mRNA levels in Control 1 (blue bars) and HD lymphoblast cells (pink and dark pink bars). Values are normalized to *GAPDH* (n=2; means± SEM, *** $p < 0.001$ by two-tailed unpaired t-test).

E. RT-qPCR analysis of *p21* mRNA levels in Control (blue bars) and HD MSNs (dark pink bars). Values are normalized to *7SK* (n>2; means± SEM).

siLuc-treated cells (Fig 5.7A). In addition, an increased mRNA expression of the *p21* gene, a cyclin kinase inhibitor involved in the DNA damage response, was detected in siHTT-depleted cells compared to siLuc-treated cells by RT-qPCR analysis (Fig 5.7B). This indicates that HTT protein may be important for either preventing or repairing DNA damage, in line with previous studies (Gao *et al.*, 2019).

Consistently, an analysis of γH2AX levels in Control and HD lymphoblast cells by Western Blot analysis showed an accumulation of DNA damage in HD cells (Fig 5.7C), again pointing to a global effect of mHTT. Moreover, an increase in *p21* mRNA expression levels was detected by RT-qPCR analysis in 18/250 CAG HD lymphoblasts and HD MSNs compared to Control cells (Fig 5.7D-E).

5.4.2 CPT-induced DNA damage and R-loop accumulation in siHTT-depleted HeLa cells

To further investigate the potential interplay between R-loops and DNA damage, it was possible to take an advantage of a highly specific inhibitor for Top1, camptothecin (CPT) (Capranico *et al*, 2010). CPT is known to promote R-loop accumulation and replication- and transcription-associated DNA damage caused by unresolved DNA supercoiling (Sordet *et al.*, 2009; Marinello *et al*, 2016). Furthermore, it has been recently shown that HD cells display a slower repair of DNA damage induced by CPT (Palminha *et al.*, 2022). Therefore, siLuc and siHTT-treated HeLa cells were incubated with either DMSO or 10 μM CPT for 1 h, followed by γH2AX staining and immuno-fluorescence (IF). As expected, CPT-treatment induced DNA damage accumulation, as indicated by γH2AX levels measured by IF in both siLuc- and siHTT-treated cells compared to the untreated and DMSO conditions (Fig 5.8). Interestingly, siHTT-depleted cells showed a significant increase in γH2AX signal following CPT treatment compared to siLuc-treated cells (Fig 5.8).

Subsequently, we examined whether increased γH2AX levels in HTT-depleted cells upon CPT treatment were accompanied by an accumulation of R-loops, as measured by S9.6 IF. It is known that CPT perturbs R-loop levels at early treatment time points (Cristini *et al.*, 2019).

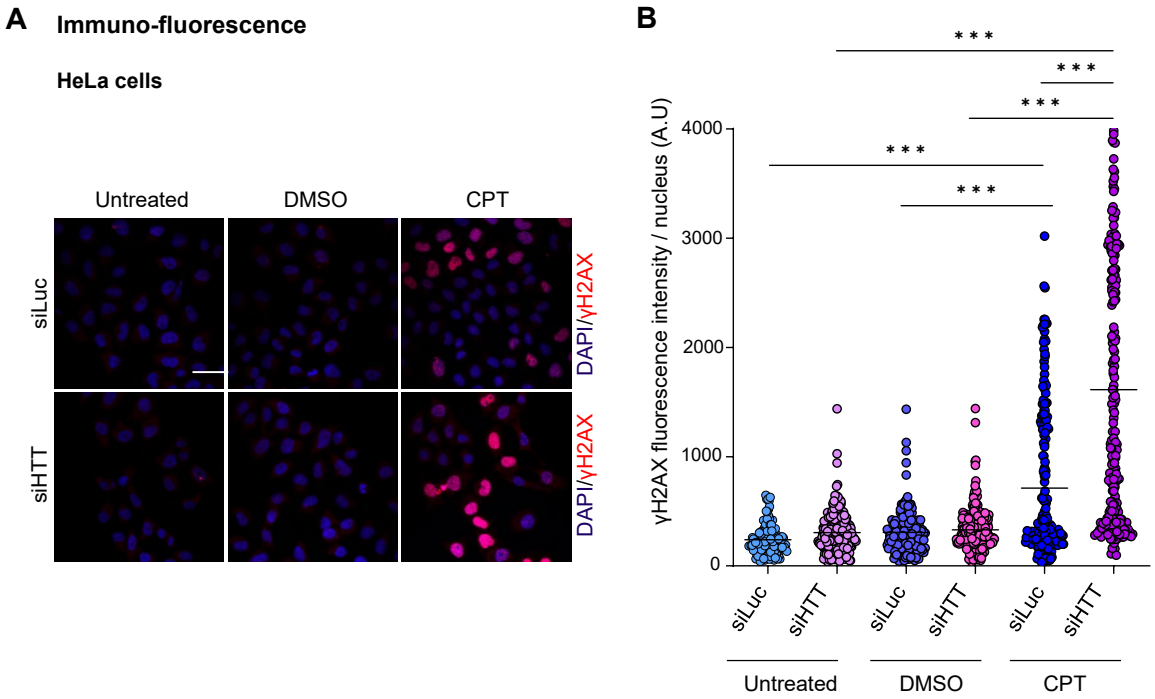


Figure 5.8. siHTT-depleted HeLa cells are susceptible to CPT-induced DNA damage.

A. IF analysis of HeLa cells transfected with siLuc or siHTT RNAs and treated with either DMSO or 10 µM CPT for 1h. Cells were stained for γH2AX (red) and DAPI (blue). Scale bar: 50µm. Representative images from n=1 experiment.

B. Quantification of γH2AX fluorescence intensity per nucleus from a representative experiment (≥200 nuclei were analysed per condition). The horizontal black bars represent the means, and each dot is one nucleus (*** $p < 0.01$ by one-way ANOVA Tukey's multiple comparisons test).

Therefore, siLuc and siHTT-treated cells were incubated with either DMSO or 10 μM CPT for 1 h, followed by S9.6 staining and IF. In line with the increased DNA damage accumulation (Fig 5.8), HTT-depleted cells showed a significant increase in S9.6 signal following CPT treatment compared to the siLuc-treated cells (Fig 5.9). Furthermore, treatment with recombinant RNase H1 *in vitro* had led to a modest but significant reduction in CTP-induced S9.6 signal, demonstrating the specificity of the antibody (Fig 5.9, +/- RNH).

Overall these data suggest that HTT-depleted HeLa cells are particularly sensitive to CPT treatment. Interestingly, HTT-depleted HeLa cells treated with CPT accumulate DNA damage and R-loops, suggesting that HTT protein may be involved in DNA damage and R-loop resolution.

5.4.3 HTT and XPG proteins are part of the same multi-protein complex

HTT co-IP experiments revealed novel interactions of HTT with several transcriptional and R-loops factors (Section 5.2.2, Fig. 5.4). Among these, the interaction of HTT with the endonuclease and DNA damage repair factor XPG was further characterized. XPG protein, also known as ERCC5, has been initially identified as the causative gene for Xeroderma Pigmentosum complementation group G (Zafeiriou *et al.*, 2001). Its function in TCR, together with the ERCC1-XPF endonuclease (XPF), has been widely described (Tsutakawa *et al.*, 2020). Additionally, it has become clear that XPG exerts several other important roles in genome maintenance, including its role in R-loop resolution (Sollier *et al.*, 2014). Interestingly, Western Blot analysis showed that XPG protein levels were up-regulated upon depletion of HTT in HeLa cells (Fig 5.10A). This result was also confirmed by the analysis of the XPG expression by IF in siHTT-depleted HeLa cells (Fig 5.10B). These data suggest that the expression of HTT and XPG proteins may be co-regulated. Finally, this further supports the idea that HTT and XPG proteins may be part of the same multi-protein complex involved in DNA damage repair and R-loop resolution.

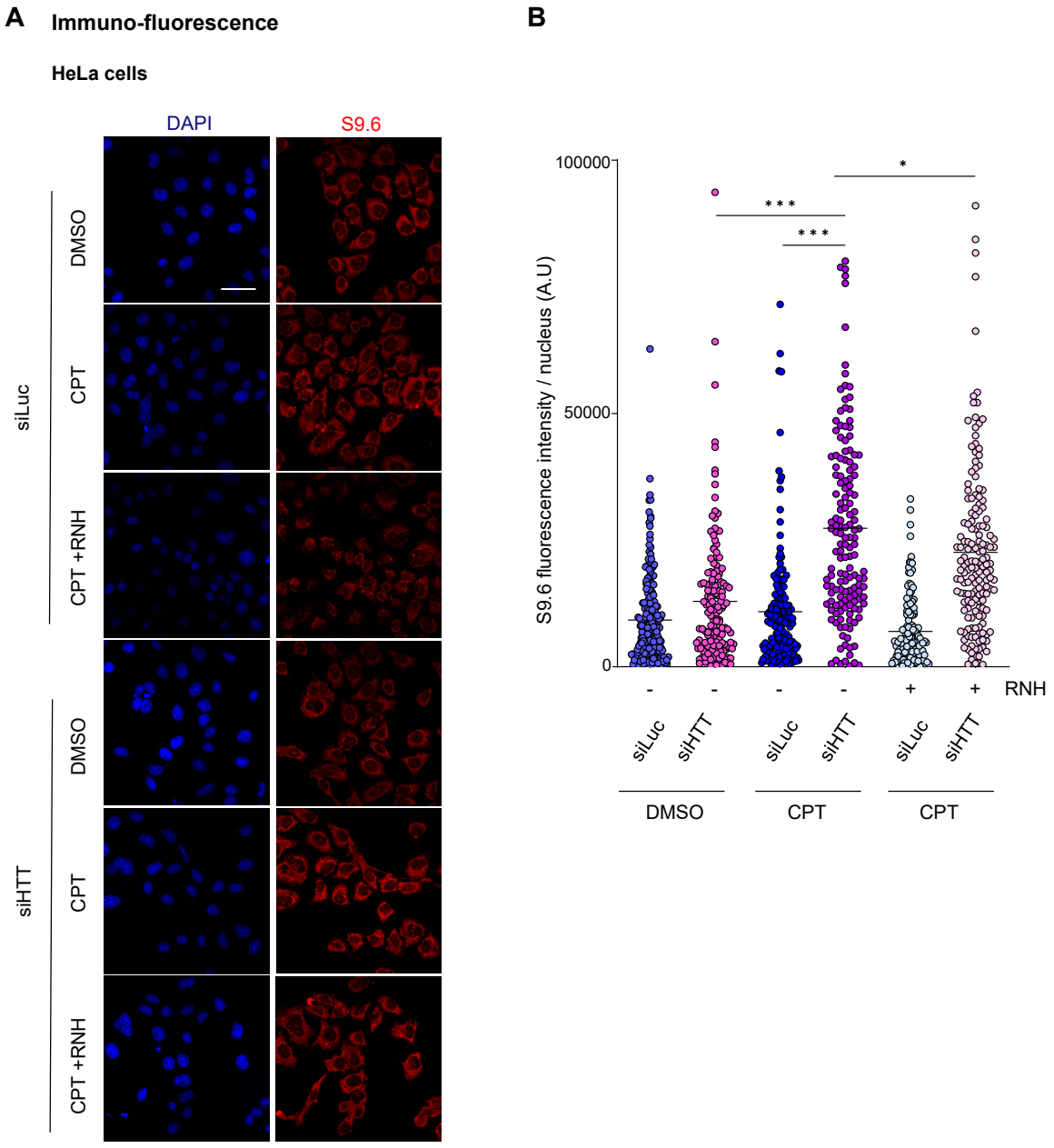


Figure 5.9. siHTT-depleted HeLa cells show R-loop accumulation upon CTP treatment.

A. IF analysis of HeLa cells transfected with siLuc or siHTT RNAs and treated with either DMSO or 10 µM CPT for 1 h. Cells were stained for S9.6 (red) and DAPI (blue). Condition CPT +RNH was treated with recombinant RNase H1 *in vitro* (after fixation) for 4 h at 37ºC, whereas the other conditions were incubated for the same amount of time with buffer. Scale bar: 50µm. Representative images from n=1 experiment.

B. Quantification of nuclear S9.6 fluorescence intensity per nucleus from a representative experiment (≥200 nuclei were analysed per condition). The horizontal black bars represent the means, and each dot is one nucleus (*** $p < 0.01$ by one-way ANOVA Tukey's multiple comparisons test).

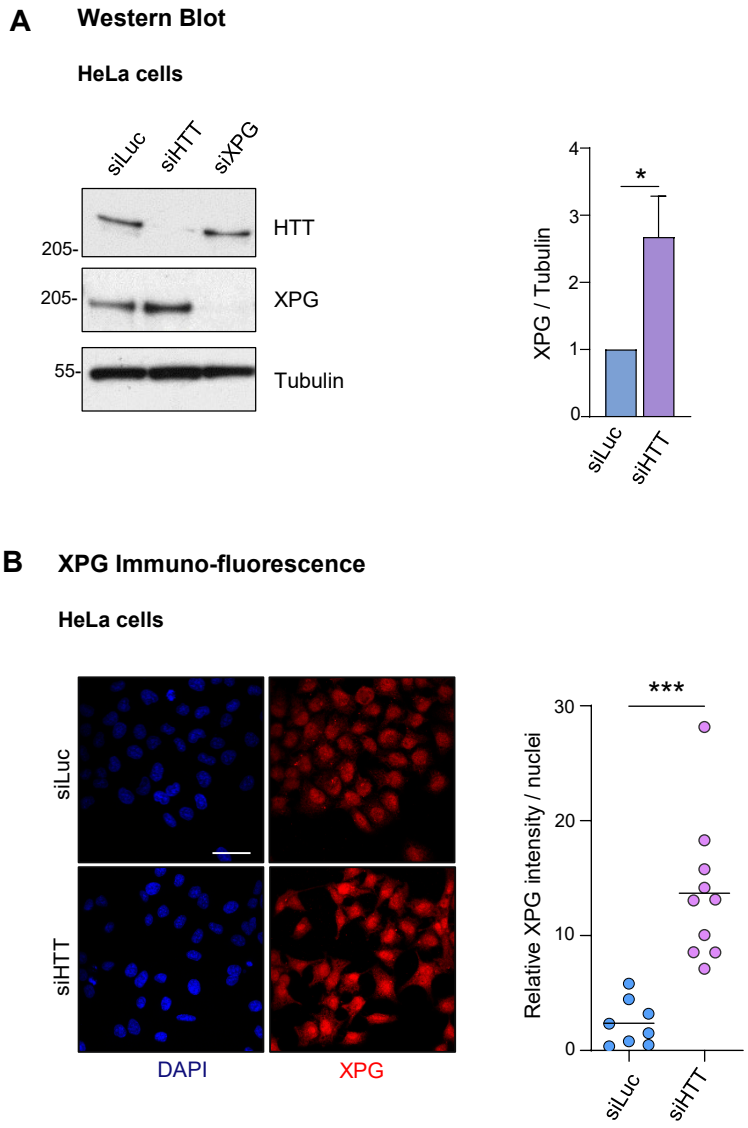


Figure 5.10. Depletion of HTT leads to upregulation of XPG protein levels.

A. **Left panel.** Western Blot analysis of HTT and XPG performed with 20 µg of protein extracts obtained from HeLa cells transfected with siLuc, siHTT or siXPG RNAs. Tubulin is a loading control. **Right panel.** Quantitation of XPG protein levels in siLuc and siHTT-depleted HeLa cells. Values are normalized to tubulin (n=6; means± SEM, * $p < 0.05$ by two-tailed unpaired t-test).

B. **Left panel.** IF analysis of HeLa cells transfected with siLuc or siHTT RNAs and stained for XPG (red) and DAPI (blue). Scale bar: 50µm. Representative images from n=1 experiment. **Right panel.** Quantification of XPG fluorescence intensity per nucleus (≥400 nuclei were analysed per condition). Each dot represents the XPG intensity/nucleus per image (*** $p < 0.01$ by one-way ANOVA Tukey's multiple comparisons test).

5.5 Inflammation phenotypes are observed in HD cells

Extensive inflammatory changes have been implicated in HD pathogenesis (Rocha *et al.*, 2016). However, the drivers of inflammation in HD are not known. Interestingly, DNA damage and R-loops have been shown to trigger inflammatory signalling (Cristini *et al.*, 2022; Giordano *et al.*, 2022; Crossley *et al.*, 2023). Therefore, to investigate this in the context of HD, inflammatory phenotypes have been measured in siHTT-depleted HeLa cells and HD cells from patients. Furthermore, the potential mechanistic link between inflammation, R-loop accumulation and DNA damage was also explored.

5.5.1 Depletion or mutation of HTT protein leads to enhanced inflammatory response

To characterize the immune response in siHTT-depleted HeLa and HD cells from patients, I examined Type 1 INF and other pro-inflammatory markers by RT-qPCR analysis. Depletion of HTT protein in HeLa cells modestly induced a panel of inflammation and immune-related mRNAs, including *ISG20, IFNGR1* and *TNFα* (Fig 5.11A). Furthermore, inflammatory responses were analysed in HD lymphoblast cells from patients. I observed an upregulation of interferon-stimulated genes (ISGs) including *ISG20, ISG15, OAS1, STING*, *IFIT1* and *NFkB*, in HD cells compared to Control cells (Fig 5.11B). Interestingly, the inflammatory phenotypes were particularly high in the 18/250 CAG HD cell line containing longer repeats, which also showed higher levels of R-loop and DNA damage accumulation (Fig 4.5 and 5.7).

Inflammation was also measured in MSNs. Upregulation of pro-inflammatory factors (*TNFα*) and ISGs (*ISG15, OAS1* and *IFNGR1)* was observed in HD MSNs (Fig 5.11C). Moreover, a high up-regulation of the pro-inflammatory chemokine *CCL5* and its downstream pathway factor *NFkB* was observed, in line with studies performed in the striatum, cortex and cerebellum of HD patients (Bradford *et al.*, 2009; Silvestroni *et al.*, 2009). Interestingly, HD MSNs differentiated from iPSCs with higher cell passage showed higher levels of inflammation, in line the R-loop accumulation phenotype (Fig 4.7). Overall, these data demonstrate that siHTT-depleted cells and HD cells from patients show enhanced inflammatory response.

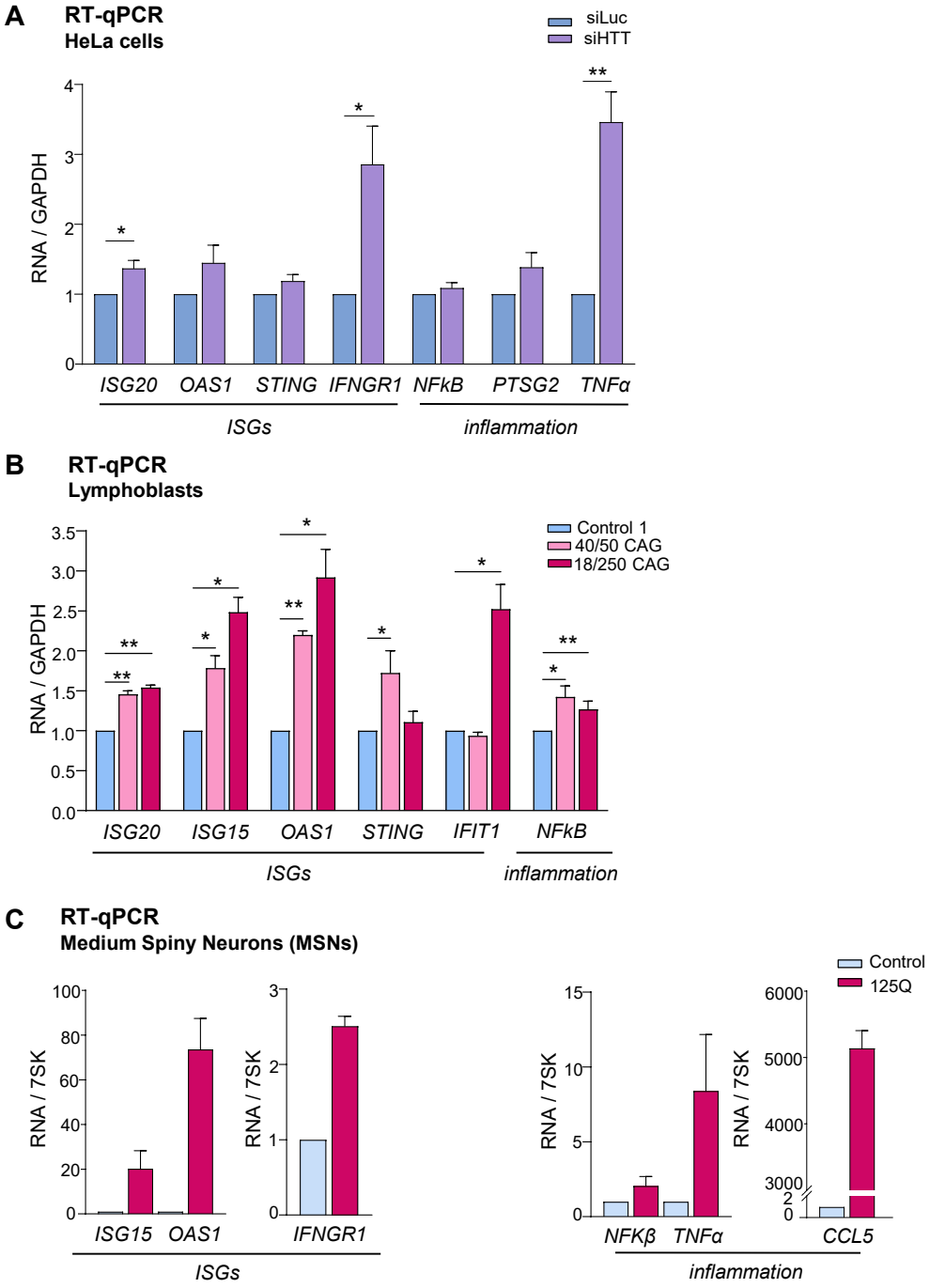


Figure 5.11. Enhanced inflammatory response in siHTT-depleted HeLa and HD cells.

A. RT-qPCR analysis of indicated mRNAs in HeLa cells transfected with siLuc (blue bars) or siHTT RNAs (purple bars). Values are relative to siLuc-treated cells and normalized to *GAPDH (*n>3; means± SEM, *p <0.05, ** p < 0.01 by two-tailed unpaired t-test).

B. RT-qPCR analysis of indicated mRNAs in Control (blue) and HD (pink and dark pink) lymphoblast cells. Values are relative to Control cells and normalized to *GAPDH* (n>2; means± SEM, *p <0.05, ** p < 0.01 by two-tailed unpaired t-test). Figures A and B generated with the help of Elena Becker.

C. RT-qPCR analysis of indicated mRNAs in Control (blue bars) and HD (pink bars) MSNs. Values are relative to Control cells and normalized to *7SK* (n=2 biologically independent experiments for Control, n=3 for HD cells differentiated from iPSCs of different passages (p26, p41, p45); means± SEM).

Furthermore, the levels of immune response and DNA damage observed in HD lymphoblast and MSNs correlate with the levels of R-loops. This raises the possibility that these processes may be connected on a molecular level.

5.5.2 Co-depletion of HTT and XPG proteins leads to accumulation of DNA damage and inflammation

As previously mentioned, several studies have shown that DNA damage and R-loops can promote inflammation (Cristini *et al.*, 2022; Giordano *et al.*, 2022; Crossley *et al.*, 2023). In the context of HD, it has been proposed that the accumulation of endogenous nucleic acids in the cytoplasm of HD cells could trigger immune activation (Sharma *et al.*, 2020). However, the nature of these endogenous nucleic acids has not been described. Therefore, it was of particular interest to investigate whether the R-loop accumulation observed in the HD cells could contribute to the activation of inflammatory responses and DNA damage, characteristic of HD.

It has been shown that both single-strand and double-strand breaks (SSBs and DSBs) can arise from the processing of pathological co-transcriptional R-loops by structure-specific endonucleases, including XPG and XPF, and trigger inflammation (Sollier *et al.*, 2014; Cristini *et al.*, 2019). Therefore, experiments in HeLa cells co-depleted for HTT and XPG proteins were performed. Analysis by RT-qPCR detected an efficient depletion of XPG and HTT on mRNA and protein levels (Fig 5.12A-B). Moreover, the depletion of HTT has led to an increase in XPG mRNA and protein levels (Fig 5.12A-B), in line with the data presented in Section 5.4.3. The independent depletion of HTT and XPG resulted in a modest induction of immune-related and DNA damage-related genes without affecting the expression of the control housekeeping gene *TUBB* (Fig 5.12C, purple and pink bars). Interestingly, the co-depletion of HTT and XPG proteins increased the activation of inflammatory and DNA damage responses compared to their individual depletion (Fig 5.12C, green bars). These findings further support the idea that HTT and XPG proteins may be a part of a complex involved in R-loop resolution and/or DNA damage repair pathways, mis-regulation of which drives inflammation. Taken together these results suggest that HTT functions to either prevent

or resolve DNA damage and R-loop accumulation and that this function may be impaired in HD, resulting in the up-regulation of immune response.

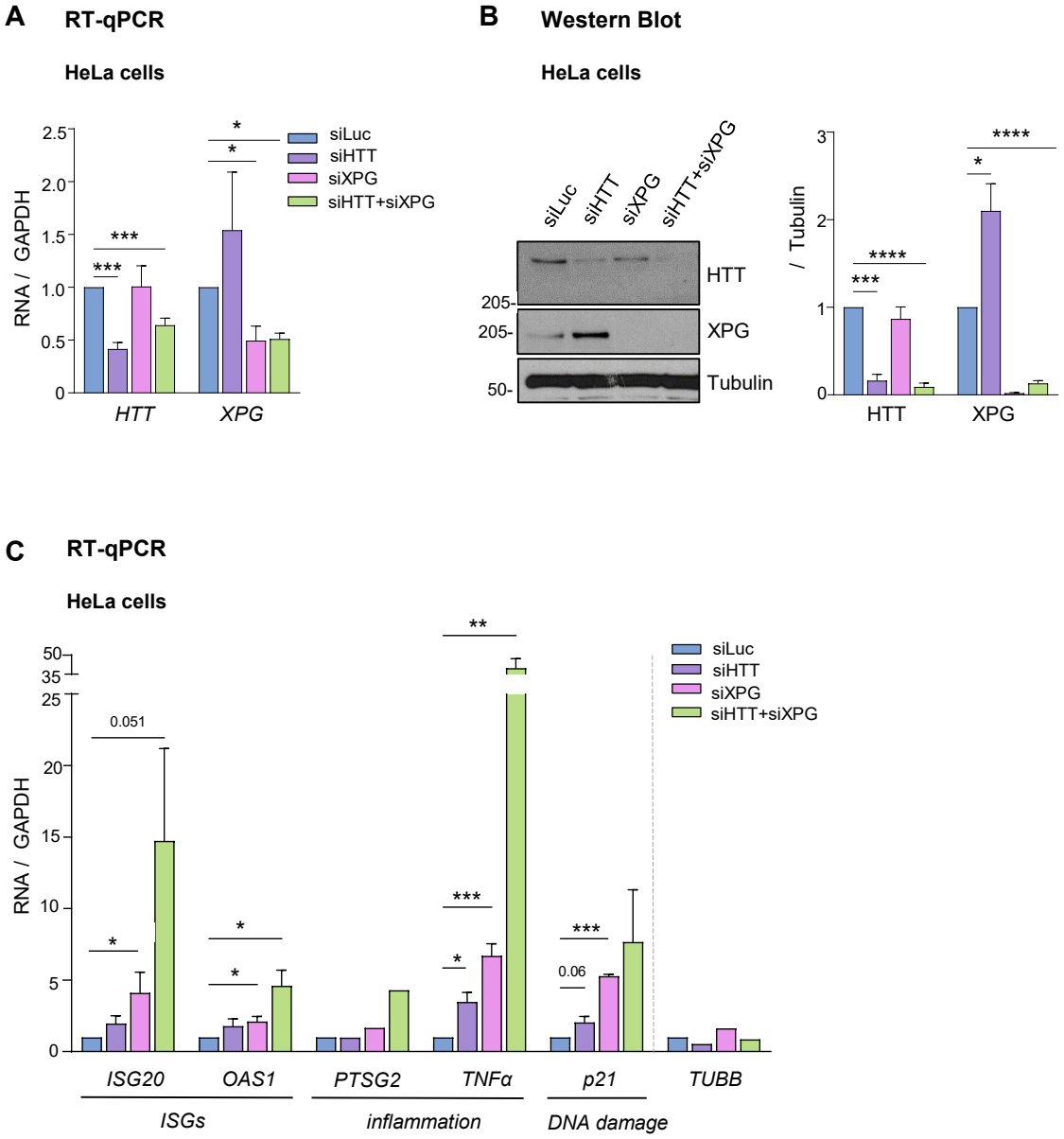


Figure 5.12. Enhanced inflammatory response in cells co-depleted for HTT and XPG proteins.

A. RT-qPCR analysis of *HTT* and *XPG* mRNAs in HeLa cells transfected with siLuc (blue bars), siHTT (purple bars), siXPG (pink bars) or siHTT+siXPG (green bars) RNAs. Values are relative to siLuc-treated cells and normalized for *GAPDH* (n>3; means± SEM, **p* <0.05, *** *p* < 0.001 by two-tailed unpaired t-test).

B. **Left panel.** Western Blot analysis of HTT and XPG performed with 20 µg of protein extracts obtained from HeLa cells transfected with siLuc, siHTT, siXPG or siHTT+XPG RNAs. Tubulin is a loading control. **Right panel.** Quantitation of HTT and XPG protein levels in siLuc, siHTT, siXPG and siHTT+XPG-depleted HeLa cells. Values are normalized to tubulin (n=3; means± SEM, **p* <0.05, *** *p* < 0.001 and *****p* < 0.0001 by two-tailed unpaired t-test).

C. RT-qPCR analysis of indicated mRNAs in HeLa cells transfected with siLuc (blue bars), siHTT (purple bars), siXPG (pink bars) or siHTT+siXPG (green bars) RNAs. Values are relative to Control cells and normalized to *GAPDH* (n>3 apart n=1 for *PTSG2* and *TUBB;* means± SEM, **p* <0.05, ** *p* < 0.01, *** *p* < 0.001 by two-tailed unpaired t-test).

5.6 Summary

The data presented in this Chapter provide novel insight into the role of HTT protein in transcription and R-loop biology with potential implications for understanding HD pathogenesis. In particular, HTT protein was found to localize in the nucleus and bind promoters of transcribed genes in HeLa cells (Fig. 5.1-5.2). Moreover, HTT co-IP experiments revealed novel interactions of HTT with key proteins involved in transcription (Fig. 5.3-5.4). Specifically, HTT protein interacts with phosphorylated Pol II, further implying its role in active transcription (Fig. 5.4). Additionally, HTT binds R-loops *in vivo* and interacts with factors involved in R-loop regulation, including Top1, XPG, DHX9 and XRN2 (Fig. 5.4-5.5). Finally, nascent transcription is impaired in HD cellular and mouse models (Fig. 5.6 and Fig. 3.10B). The observation that HTT protein interacts with active Pol II during transcription and several R-loop factors may imply that HTT acts as a scaffold for R-loop resolution during transcription. As such, dysregulation in R-loop resolution could account for the R-loop accumulation and transcriptional dysregulation observed in various HD cellular models (Chapter 4 and Fig. 5.6). Furthermore, DNA damage and activation of inflammatory response have been reported in HD, but the drivers of these processes are currently unknown. Since R-loop accumulation has been shown to induce activation of DNA damage and inflammatory response, the potential link between these phenotypes has been addressed in this Chapter. siHTT-depleted HeLa cells were sensitive to CPT-treatment, showing an increased DNA damage and R-loop accumulation (Fig. 5.8-5-9). Interestingly, the levels of immune response and DNA damage, observed in HD cells correlated, with the levels of R-loops (Fig. 5.7; Fig. 5.11 and Chapter 4). This raises the possibility that these processes are linked on molecular levels. Importantly, further characterization of novel interactions between HTT and XPG, a endonuclease that targets R-loops and plays a role in TCR, revealed that the expression of these two proteins may be co-regulated (Fig. 5.10). Furthermore, a strong induction of immune response was observed following co-depletion of HTT and XPG proteins (Fig. 5.12). Collectively, these data support the idea that HTT and XPG proteins may be a part of the same multi-protein complex involved in the DNA damage and R-loop resolution. However, the mechanistic

explanation for the exact role of HTT in R-loop regulation, inflammation and DNA damage pathways and their relevance to HD pathogenesis needs to be further explored.

Chapter 6

Discussion and Conclusions

6.1 Key findings and novelty

The data presented here provide novel insights into the molecular mechanisms underlying Huntington's Disease, a neurodegenerative disorder with no available cure. The main aims of this project were, firstly, to study the transcriptional regulation of the *HTT* gene; secondly, to investigate the role of non-canonical RNA/DNA structures in HD pathogenesis and finally, to investigate the role of HTT protein in the nucleus. My key findings are represented in Fig. 6.1 and summarized below.

The transcription of the *HTT* gene was shown to be impaired in HD. In details, transcription of the *HTT* gene is mis-regulated in HD, with increased levels of heterochromatin and the formation of dsRNA structures over expanded repeats. Impaired transcription of the *HTT* gene in HD cells leads to lower expression of the HTT protein, which may contribute to pathology. Heterochromatin and dsRNAs formed over expanded repeats could also play a role in somatic repeat instability in HD.

The HTT protein was previously shown to be involved in transcriptional processes. My results also implicate the HTT protein in transcriptional regulation and uncover a novel role for this protein in R-loop biology. I show here that the HTT protein forms a complex with R-loop structures *in vivo* and binds to several R-loop factors, suggesting that the HTT protein plays a role in R-loop homeostasis. In line with this, I found increased levels of pathological R-loops in HD cells, both genome-wide and on specific genes. This could result either from lower levels of the wild-type HTT protein and loss of its physiological function in R-loop biology or from the gain-of-function of the mHTT protein. I also show that R-loop accumulation in HD cells is accompanied by an increase in inflammation and DNA damage, raising the possibility that pathological R-loops could contribute to these HD phenotypes. These results were also observed in MSNs, the most affected cells in HD, supporting their crucial role in HD pathogenesis. Interestingly, co-IP experiments revealed a novel interaction of the HTT protein with XPG, an endonuclease involved in transcription-coupled DNA repair (TCR) pathway and R-loop resolution. Furthermore, the data show that the expression of HTT and XPG proteins is co-regulated and that the co-depletion of HTT and XPG proteins in HeLa cells results in inflammatory responses. This suggests that these proteins may form a complex that can be

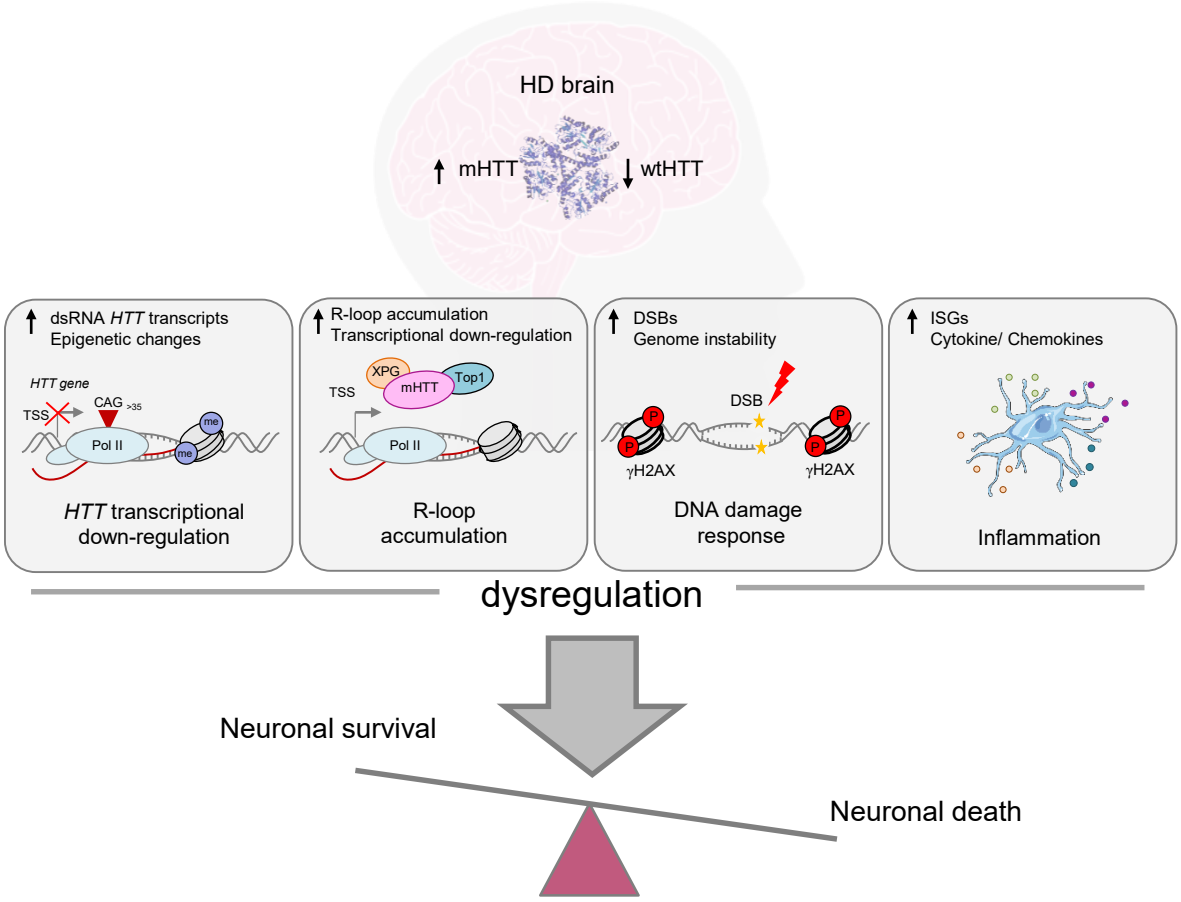


Figure 6.1. Major molecular mechanisms contributing to HD pathogenesis described in this study.
Several cellular processes are disturbed by the presence of mutant huntingtin (mHTT) or by the loss of wild-type huntingtin (wtHTT). The figure shows a simplified overview of the pathogenic pathways contributing to HD as discussed in this study. Specifically: **(i)** transcriptional down-regulation of the *HTT* gene and its epigenetic silencing was observed in HD cells. Furthermore, dsRNAs form over expanded *HTT* transcripts. Impairment in the transcription of the *HTT* gene resulted in lower levels of HTT protein in HD cells. **(ii)** accumulation of R-loops and transcription down-regulation were detected in neuronal and non-neuronal HD cells, together with **(iii)** accumulation of DNA damage and **(iv)** activation of inflammatory response. All these effects could contribute to genome instability, finally leading to neuronal cell death in HD. Although each pathway is represented as an independent process, in reality, disruptions in one pathway probably influence HD pathogenesis by exacerbating the dysfunction of other key events**.**

involved in the DNA damage and R-loop resolution. Therefore, HTT mutations or deletion could disrupt the functional integrity of the TCR complex, leading to R-loop accumulation. Collectively, the data presented here provide novel insights towards understanding the contribution of unusual RNA/DNA structures to HD pathogenesis, potentially uncovering new targets for therapeutic interventions. In addition, using HD as a disease model, this work provides new insight into the mechanisms underlying molecular pathology of human repeat expansion diseases in general.

6.2 Transcriptional defects at the *HTT* gene in HD

Thus far, no previous study has provided a comprehensive evaluation of the regulation of *HTT* expression. The aim of this study was to understand the mechanisms underlying the transcriptional regulation of the *HTT* gene and provide insights into its epigenetic signature in Control and HD cells. My findings indicate that Pol II transcription is down-regulated at the *HTT* gene in HD cells. Furthermore, the transcriptional down-regulation correlates with an increased repressive epigenetic signature of *HTT* in HD. Specifically, the down-regulation of *HTT* expression is associated with the reduction of active chromatin marks, including H3K9ac, and the formation of repressive H3K9me2/me3 chromatin marks over the promoter-proximal region of the *HTT* gene in HD cells. These findings are consistent with previous studies in other repeat expansion disorders demonstrating decreased acetylation and increased histone methylation at expanded repeat loci (Greene *et al.*, 2007; Groh *et al.*, 2014a; Chutake *et al.*, 2015). A recent study showed high levels of DNA methylation at the *HTT/Htt* locus, located in the vicinity of the CAG expansion in exon1 in humans, mice, and sheep HD models (Lu *et al.*, 2020). I speculate that the DNA and histone methylation, including H3K9me2/me3 and H3K27me3 marks, observed at the HD locus, particularly in the first intron of the gene (Fig. 3.9), play a role in the regulation of *HTT* gene expression. However, to understand their contribution, further investigation using drugs, targeting specific methyltransferases, is necessary. This will provide a more detailed understanding of the interplay between epigenetics and transcription of the *HTT* gene.

It is worth noting that highly pathogenic exon 1 HTT protein with an expanded polyQ stretch is sufficient to model numerous HD-related phenotypes (Barbaro *et al.*, 2015). The exon 1 HTT

protein is produced by mis-splicing of *HTT* exon 1 to exon 2 that generates a small transcript, comprising the 5′UTR, exon 1 and terminating at cryptic polyA sites in intron 1 (Sathasivam *et al.*, 2013). The production of this transcript is dependent on the length of the CAG repeat. However, the factors that influence the efficiency of splicing and/or the premature termination of *HTT* remain unknown. It is possible that increased heterochromatinization of *HTT* intron 1 causes premature termination of *HTT.* Therefore, investigating the potential interplay between epigenetics and the regulation of cryptic poly A sites in the *HTT* gene could provide insight into the production of this truncated exon 1 HTT protein.

Intriguingly, the extent of the transcription impairment and epigenetic silencing correlates with the length of the CAG repeats in the *HTT* gene of HD cells. In fact, the genes with longer CAG repeats (18/250 CAG) showed greater transcriptional impairment and stronger heterochromatinization. Recent genome-wide association studies (GWAS) have revealed that variants in MMR enzymes (e.g., *MLH1*, *MSH2* and *MSH3*), which are known to modify the germline and somatic instability of expanded *HTT* CAG repeats in HD mouse models (Pinto *et al.*, 2013), act as modifiers of the age of motor symptom onset in HD (GeM-HD, 2019). Therefore, it is important to investigate if impaired transcription and formation of heterochromatin across the *HTT* gene affects the somatic instability of the CAG repeat in HD. For example, longer CAG repeats could act as a "roadblock" to Pol II, impeding efficient transcriptional elongation. It was proposed that chromatin structure and/or transcriptional activity at the HD locus may contribute to somatic CAG instability in HD mouse models (Goula *et al.*, 2012). Since the largest expansions in CAG tract length have been observed in the HD striatum and the cerebral cortex, where neurons can harbour repeat tract expansions in the hundreds (Gonitel *et al.*, 2008; Mouro Pinto *et al.*, 2020), it would be interesting to characterize the transcriptional and epigenetic profile of the *HTT* gene in these neuronal cells.

Finally, *HTT* gene transcriptional down-regulation correlates with the lower levels of HTT protein in HD cell lines and this could also potentially contribute to the disease phenotype (Fig. 3.11). Although the expression of mHTT plays a major role in the disease, a potential contribution of loss of normal protein function has been recently proposed as a possible modifying factor of the disease. Overexpression of HTT has been shown to have a mild neuroprotective effect in a variety of model

systems (Rigamonti *et al.*, 2000; Leavitt *et al.*, 2006). However, HTT antibody used in this study did not allow the distinction between wild-type and mutant HTT. Therefore, the use of specific antibodies for wtHTT and mHTT would enhance this analysis. In addition, lower levels of wild-type *HTT* mRNA were detected in MSNs compared to the parental iPSC cells (Fig. 3.2). Further analysis of the HTT protein levels in these cells is needed to confirm whether lower levels of *HTT* RNA correspond to lower levels of HTT protein. However, it can be speculated that neuronal cells are susceptible to neurodegeneration in HD due to their naturally lower levels of wild-type *HTT* mRNA compared to other cells, such as the iPSCs reported in this study. Comparing *HTT* mRNA and HTT protein levels in different tissues could be of interest to evaluate if the CNS has the lowest amount of HTT in physiological conditions, which may be further impacted by the presence of HTT mutations, resulting in more pronounced HD phenotypes in neuronal cells.

6.3 Formation of non-canonical RNA/DNA structures across the *HTT* gene

Previous studies have shown that heterochromatin and transcriptional silencing of genes associated with human repeat expansion diseases, including FRDA and FXS, co-localize with the formation of pathological R-loops (Groh *et al.*, 2014a). The formation of stable R-loops over CAG repeat tracts has been demonstrated *in vitro* (Lin *et al.*, 2010). However, R-loop formation over the *HTT* gene has not been previously investigated in the context of HD. Therefore, I investigated the formation of R-loops across the *HTT* gene using lymphoblast and neuronal cells derived from HD patients. Notably, R-loops were detected over the *HTT* gene and their levels were comparable to those detected over the highly expressed *γActin* gene (Fig. 4.1-4.2). These findings are consistent with a recent DRIP-seq study showing that R-loops are formed at the *HTT* gene in neuronal progenitor cells (Thongthip *et al.*, 2022). Surprisingly, R-loops are not enriched over expanded CAG repeats at the *HTT* gene in HD cells. Instead, R-loop formation over the *HTT* gene was slightly decreased in a transcriptional and CAG repeat-dependent manner in HD cells. Further evaluation of the stability of R-loops on expanded repeats *in vivo* could help to highlight the differences between structures formed over physiological and expanded CAG repeats and provide insights into their role in HD.

Transcription of genes containing long repeats can generate repeat-containing RNAs that might be directly involved in disease toxicity (Liu *et al.*, 2017). However, the contribution of expanded *HTT* RNA to neuronal toxicity in the context of HD remains poorly understood. To investigate this, I analysed the presence of dsRNAs over expanded CAG repeats in HD cells. I observed an enrichment of dsRNA structures over expanded *HTT* transcripts, indicating their potential involvement in HD pathology (Fig. 4.3-4.4). Interestingly, I observed a higher enrichment of dsRNAs in the cell lines with longer repeats (18/250 CAG). This may be due to the increased stability of dsRNA structures formed by longer CAG repeats. In support of this, *in silico* structural modelling of CAG repeat-containing *HTT* mRNAs suggest that hairpin structures, comprising the CAG repeat region, become longer and more stable as the CAG repeat length increases (Sobczak & Krzyzosiak, 2004). Furthermore, the J2 antibody, which specifically recognizes dsRNA structures, more easily recognizes longer hairpins (Bonin *et al.*, 2000). Thus, it is also possible that the formation of dsRNA in cells with shorter repeats (40/50 CAG) may have been underestimated. Finally, it can also be speculated that the stronger dsRNA enrichment in the cell line with longer repeat expansions could result in a more severe phenotype. It would be interesting to determine whether these dsRNA structures are formed on nascent *HTT* transcripts during transcription. Given that repeat instability likely arises from the formation of non-canonical RNA/DNA structures that predispose the tract to errors in mismatch repair (MMR), the presence of dsRNAs at CAG repeats, together with epigenetic silencing of the *HTT* gene, could promote somatic instability. In support of this, *in vitro* data showed that the MMR MutSbeta complex, which includes MSH2 and MSH3, is very efficient at repairing short slip-outs, but fails to repair long or clustered slip-outs, favouring expansions (Panigrahi *et al.*, 2010). A small molecule, naphthyridine-azaquinolone, was shown to specifically bind expanded CAG repeats and induce their contraction in cells from affected individuals and brains of mice (Nakamori *et al.*, 2020). Therefore, specific compounds that can target dsRNA at the expanded *HTT* locus could be exploited therapeutically.

Another intriguing possibility is that the observed dsRNAs may be generated via bidirectional transcription of the expanded repeats. In support of this, an antisense *HTTAS* transcript containing the expanded repeat tract has been previously identified in brain frontal cortex tissues

from healthy and HD patients (Chung *et al.*, 2011). Furthermore, repeat-containing dsRNAs generated by bidirectional transcription have been shown to cause inflammation-mediated pathology in *D. melanogaster* (Lawlor *et al.*, 2011; van Eyk *et al.*, 2019).

Different repeat lengths could favour the formation of different RNA structures, which could influence the binding of dsRNA-binding factors. *HTT* repeat-containing RNAs have been shown to co-localize with MBLN splicing factors (de Mezer *et al.*, 2011; Sun *et al.*, 2015), but knowledge of the proteins that can bind dsRNAs directly is still incomplete. Therefore, an RNA-IP method to immuno-precipitate dsRNA and detect dsRNA-binding protein was optimized in Chapter 4 (Fig. 4.12-4.13). This approach has been shown to be a valuable discovery tool, as evidenced by a recent publication that utilized a similar approach to investigate dsRNA-binding proteins in Sindbis virus (SINV)-infected human cells (Girardi *et al.*, 2023). The RNA-IP method proved to be efficient in immuno-precipitating well-known dsRNA-binding proteins including DHX9, ADAR1 and PKR, underlining its potential for the investigation of dsRNA-binding proteins in health and disease. However, additional optimisation steps, including the use of nucleic acid competitors, and treatments with RNase III and benzonase, could further improve the specificity of the technique. Once the RNA-IP method is validated, proteomic analysis will be employed to identify proteins associated with dsRNA structures in health and disease. The characterization of the dsRNA interactome and the understanding of its contribution to HD will provide a critical link between dsRNA biology and disease pathogenesis. To test if candidate dsRNA-binding proteins identified by the RNA-IP could bind the dsRNA formed over expanded *HTT* transcripts in HD cells, ChIP-qPCR for DHX9 was performed using primers across the *HTT* gene (Fig. 4.14). The results suggested that DHX9 is not enriched, but rather it is decreased over expanded CAG repeats of the *HTT* gene, in line with the role that DHX9 plays in transcriptional regulation (Lee & Pelletier, 2016). Therefore, it will be interesting to test other candidates which, for example, can promote dsRNA-mediated transcriptional repression of the *HTT* gene and contribute to pathology. Characterizing the proteins that bind the dsRNAs formed over expanded CAG repeats could have therapeutic implications for HD.

In summary, my study reveals different transcriptional dynamics and chromatin environment for the *HTT* gene in health and disease (Fig 6.2). In healthy cells, repeats between 10 and 35 are

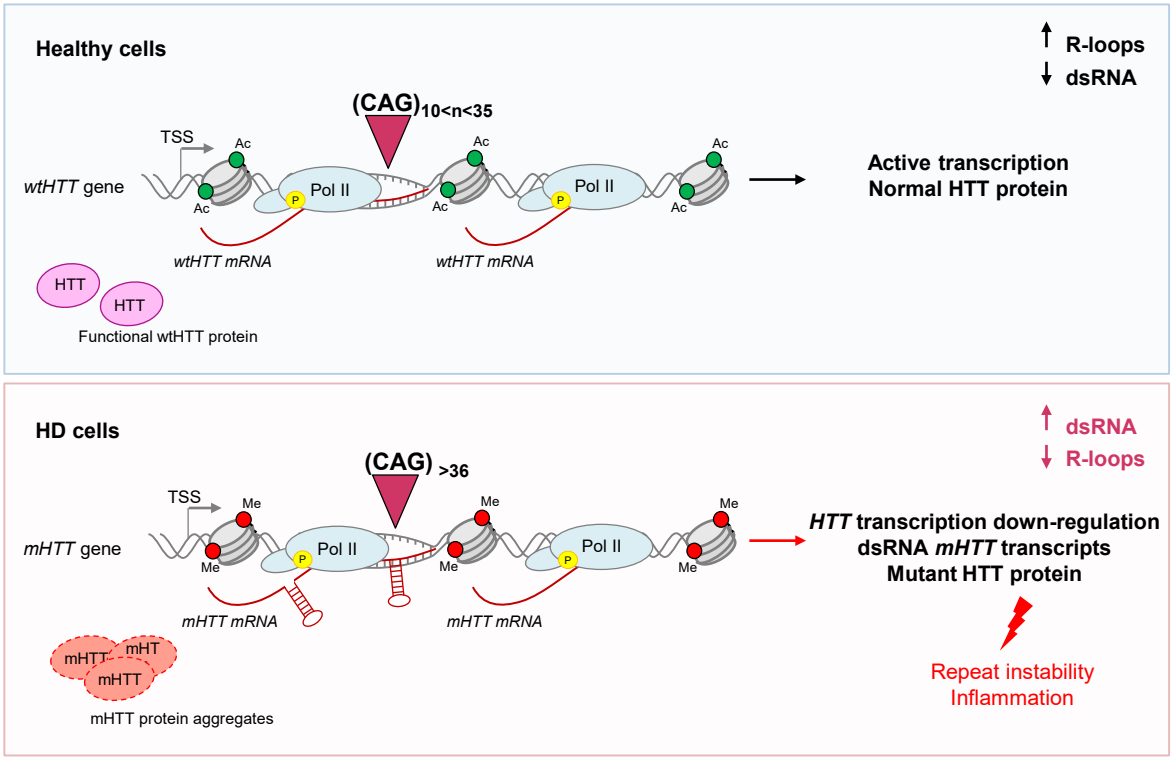


Figure 6.2. The *HTT* gene in health and disease.
Top panel. In healthy cells, the *HTT* gene contains between 10 and 35 CAG repeats. Transcription of the *HTT* gene leads to the production of a functional wild-type HTT (wtHTT) protein. **Bottom panel.** In HD patients, the *HTT* gene carries more than 36 CAG repeats and encodes for the toxic mutant HTT (mHTT) protein. The transcription of the expanded *HTT* gene is down-regulated and correlates with heterochromatin formation over expanded repeats in HD cells. The presence of R-loops and the accumulation of dsRNA structures over expanded transcripts could directly affect the *HTT* gene transcription. Furthermore, the presence of heterochromatin and non-canonical RNA/DNA structures could also play a role in somatic repeat instability in HD. dsRNA m*HTT* transcripts could also contribute to the inflammation phenotype observed in HD. Ac and Me indicate methylation and acetylation marks, respectively. Pol II, RNA Polymerase II; TSS, Transcriptional Start Site.

present on the *HTT* allele and transcription is active, leading to the production of a wild-type HTT protein. However, above 36 repeats HD pathology is triggered. I observed that transcription down-regulation of the expanded *HTT* gene correlates with its epigenetic signature through an increase in heterochromatin marks. The transcription of the *HTT* gene could be directly affected by the presence of R-loops and the accumulation of dsRNA structures over expanded transcripts (Fig. 6.2). Considering that both dsRNA and R-loops are found over the *HTT* gene in health and disease, an intriguing hypothesis could be that there is competition between R-loops and dsRNA at the *HTT* gene. For example, in HD, the mutant mRNA could be sequestered into the dsRNA-like structures and this would prevent the formation of stable R-loops over the gene. In support of this, it has been recently shown that RNA G4 structures can be converted into R-loops by the RNA helicase DDX1 to promote *IgH* Class Switch Recombination, demonstrating that the formation of these non-canonical structures is tightly regulated (Ribeiro de Almeida *et al.,* 2018). Therefore, disrupting the balance of these structures may have important consequences for disease pathology. Furthermore, heterochromatin formation and the presence of non-canonical RNA/DNA structures could also play a role in somatic repeat instability in HD (Fig. 6.2). Finally, expanded *HTT* dsRNA could act as a trigger for inflammation. Therefore, understanding the interplay between the formation of non-canonical secondary structures, transcription and somatic instability at CAG expanded repeats will shed light on fundamental molecular mechanisms that underlie HD pathogenesis.

6.4 Formation of non-canonical RNA/DNA structures in HD cells

Dysregulation of R-loops has been associated with several neurodegenerative diseases and cancers (see Table 1.1). However, the formation of R-loops in the context of HD has not been previously investigated. My results (Chapter 4) indicate that R-loops accumulate in HD cells both globally and at specific genomic locations, suggesting a potential role for these non-canonical structures in HD pathogenesis (Fig 4.5-4-7). The R-loop accumulation in HD MSN cells indicates that the R-loop phenotype is not replication-dependent and that pathological R-loops in HD cells are mainly formed co-transcriptionally. Furthermore, the R-loop phenotype in these cells correlates with the passage number of the iPSCs, used for MSNs differentiation. The DNA sizing conducted in Tabrizi lab previously showed that HD MSNs derived from iPSCs with higher passages carried more repeats (~140-145 CAG) than the ones originated from iPSCs with lower passages (~130 CAG) (pers. comm.). Thus, repeat instability and expansion of the CAG repeats may result in more severe phenotypes in HD cells, including higher levels of R-loop accumulation. To address this question more precisely, the DNA samples used for DRIP analysis will be analysed for CAG repeat size.

All Slot Blot experiments in this work were performed in the presence of a control sample treated with recombinant RNase H, which specifically degrades RNA in the RNA/DNA hybrid (Cerritelli & Crouch, 2009). The RNase H treatment substantially reduced the S9.6 signal in lymphoblast cell samples (Fig. 4.5). However, some signal remained in HD MSN samples after RNase H treatment (Fig 4.7). Interestingly, recent studies have described that pathological R-loops are partially resistant to RNase H digestion (Crossley *et al.*, 2020; Crossley *et al.*, 2023). In these studies, RNase H-resistant R-loops were found to have distinct sequence characteristics, including a high GC-skew and the formation of G-quadruplexes (Crossley *et al.*, 2020). Therefore, it is possible that pathological R-loops that form in HD MSN cells exhibit these characteristics, resulting in decreased sensitivity to RNase H. The cellular threat associated with R-loop accumulation to long-lived, non-dividing, and transcriptionally "high maintenance" cells such as neurons, could stem from a progressive accumulation of R-loops leading to DNA damage accumulation. Thus, in neurons, R-loop accumulation may cause genomic instability, neurodegeneration and cell death more readily

compared to other cell lines. Understanding the origin and the characteristics of these pathological R-loops will inform the design of therapeutic strategies targeted to these structures.

The DRIP-seq analysis performed in this study has certain limitations. The addition of RNase H digestion would provide further evidence for the specificity of the R-loops detected. Moreover, the comparisons between the samples were made after the hybrid signal was normalized to the total mapped reads from each sample. The addition of a 'spike-in' control during sample preparation will provide better internal normalization between different samples (Crossley *et al.*, 2020).

Most of the research in the HD field is carried out using mouse models. Although mouse models are useful to identify and test new therapeutics, they fail to fully recapitulate all the human phenotypes and lack genetic context. With this rationale, we decided to focus our study on human cells from HD patients. However, it can be argued that the use of cells from patients could result in misleading conclusions due to patient-patient genetic variability. To overcome this limitation, we used a panel of Control and HD cell lines (Table 2.1). In addition, the most striking phenotypes observed in cells derived from HD patients, including the reduction of *HTT* mRNA levels and the R-loop accumulation, were supported by the data from the HD-relevant neuronal MSN cells. However, isogenic cells would further allow the identification of the molecular mechanisms caused by the disease mutation. For this reason, we are currently optimizing a MSN differentiation protocol from iPSCs (Chapter 3) to generate an HD MSN isogenic model where key findings of this work will be validated.

6.5 Global nascent transcription is impaired in HD

In Chapter 5, I demonstrated that transcription is globally down-regulated in HD fibroblast cells, consistent with genome-wide studies in an HD mouse model (Fig.5.6) (Alcala-Vida *et al.*, 2021). A recent study in hESCs revealed that HTT binds the activating transcription factor 7-interacting protein 1 (ATF7IP), a regulator of SETDB1 (Irmak *et al.*, 2018). The binding of HTT to ATF7IP inhibits the interaction of the ATF7IP-SETDB1 complex with other heterochromatin regulators and transcriptional repressors, thereby maintaining low levels of H3K9me3 in hESCs. Loss of HTT leads to a global increase in H3K9me3 levels (Irmak *et al.*, 2018). Consistent with this,

increased levels of H3K9me3 were also found in HD patient brain cells (Lee *et al.*, 2013). This could explain how loss or mutation of HTT protein may promote the transcriptional down-regulation observed in HD cells. Interestingly, loss or mutation of HTT could also directly promote the heterochromatinization of its own locus, resulting in the increased H3K9me3 levels across the *HTT* gene observed in Chapter 3 (Fig. 3.9).

Another mechanism that could explain the global impairment of transcription in HD cells is linked to the methyltransferase SET domain–containing 2 (SETD2) protein. SETD2 was originally identified as an interactor of HTT protein in yeast (Faber *et al.*, 1998). SETD2 also interacts with the Ser2P C-terminal domain of Pol II and deposits H3K36me3 mark during transcriptional elongation (Yoh *et al.*, 2008), which increases along gene bodies peaking at the 3' end (Barski *et al.*, 2007). Although its activity as a chromatin remodeler responsible for H3K36me3 is well characterized, surprisingly, no function linked to its role as an HTT interactor has been yet identified. Notably, SETD2 associates with N-terminal HTT in disease-associated brain deposits (Passani *et al.*, 2000), and it is suggested that mHTT-SETD2 interaction alters SETD2 activity via sequestration (Gao *et al.*, 2014). This may lead to transcriptional dysregulation. Therefore, investigating the role of SETD2 and its interaction with HTT could contribute to a better understanding of transcription dysregulation in HD.

The HTT protein is also shown to bind to the PRC2 complex in a polyQ-length sensitive manner (Seong *et al.*, 2010). Ablation of PRC2 in MSNs reproduces several cellular phenotypes in HD, including aberrant de-repression of transcripts that code for developmentally regulated transcription factors, loss of gene expression associated with MSN identity, and prolonged cell death (von Schimmelmann *et al.*, 2016). Therefore, the aberrant interaction of the mutant HTT and PRC2 may also contribute to the transcriptional dysregulation observed in HD.

Pearl *et al.,* (2020) recently performed ChIP-seq analysis of histone modifications and HTT protein in a mouse HD model. ChIP-seq analysis revealed a positive correlation between HTT protein occupancy and H3K27me3 mark and a negative correlation with the H3K4me3 mark, a modification that is associated with transcriptionally active/poised chromatin (Pearl *et al.*, 2020). These findings support the idea that HTT protein may play a direct role in mediating heterochromatin formation,

finally resulting in gene repression. Interestingly, I observed a peak of the H3K27me3 mark in the first intron of the *HTT* gene, which was shown to be de-regulated in HD cells, however, its regulatory role remains unclear. Therefore, it could be interesting to analyse HTT occupancy at this peak and understand if it correlates with the levels of this mark in HD cells. In support of a direct role of HTT in transcriptional regulation, Pearl *et al.* also revealed that HTT protein, although preferentially enriched in intergenic regions and co-localized with facultative and constitutive heterochromatin, can also bind TSS-proximal regions at actively transcribed genes (Pearl *et al.*, 2020). Therefore, HTT ChIP-qPCR analysis and HTT co-IP experiments were performed to investigate whether transcriptional dysregulation in HD is occurring due to direct interactions between HTT and chromatin/transcriptional factors. I found that HTT binds to promoters *in vivo* (Fig. 5.2). This correlates with transcription dysregulation at promoter-proximal regions observed in the 140Q HD mice (Fig. 3.10). Therefore, altered huntingtin-chromatin interactions at gene promoters may play a direct role in driving transcriptional dysregulation in HD. I also observed that the HTT protein was shown to bind its own promoter (Fig. 5.2), suggesting a *cis* role of the HTT protein in regulating its own expression.

IP experiments of endogenous HTT revealed novel interactions of the HTT protein with key proteins involved in transcription (Fig. 5.4). For example, HTT interacts with both PoI II-Ser5P and Pol II-Ser2P, which are highest during transcription initiation and elongation, respectively (Bowman & Kelly, 2014) (Fig 5.4). The role of HTT protein in transcription elongation could also be related to its interaction with SETD2, as mentioned above. Thus, the data presented here suggest that the HTT protein plays a role in transcriptional initiation, and it may also travel with Pol II during transcription, pointing towards its possible function during transcription elongation and termination. Taken together, the transcription down-regulation observed in HD cells could arise from the loss-of-function of HTT protein in transcriptional regulation and epigenetic homeostasis.

6.6 HTT and R-loop homeostasis

The HTT protein interacts with several factors involved in R-loop resolution, including DHX9, XPG, Top1 and XRN2 (Fig. 5.4). R-loops are known to form during transcription and are enriched at gene promoters and terminators (Skourti-Stathaki *et al.*, 2011; Ginno *et al.*, 2012). DHX9 and XRN2 are important in resolving R-loops during transcription termination (Skourti-Stathaki *et al.*, 2011; Cristini *et al.*, 2018), while XPG and Top1 act in R-loop clearance pathways in both replicating and non-replicating cells (Sollier *et al.*, 2014; Cristini *et al.*, 2019; Promonet *et al.*, 2020). Since HTT interacts with Pol II and several R-loop resolution factors, it is possible that HTT protein act as a scaffold for R-loop resolution during transcription. Additional co-IP experiments with these factors are required to validate their mutual interaction with HTT. Furthermore, co-IP experiments were performed in whole cell extracts and additional endogenous HTT co-IP immuno-precipitations from nuclear extracts are important to confirm these interactions in the nucleus.

It is possible that the role of HTT in R-loop resolution is impaired in HD, leading to the observed R-loop accumulation in HD cells. If HTT is a scaffold for R-loop resolution and R-loop are dysregulated in HD, one would expect the protein to bind R-loops. Indeed, endogenous HTT co-IP experiments showed that HTT interacts with R-loops *in vivo*, further supporting its role in transcription and R-loop biology (Fig. 5.5). However, these results do not necessarily indicate a direct interaction between HTT and R-loops, as other factors, such as Pol II, may mediate this interaction.

Finally, it is important to note that when performing R-loop IP experiments, full-length HTT was not detected in the IP due to the low abundance of nuclear HTT protein in HeLa cells. However, a 120 kDa band was detected in the IP lane which may represent a fragment of wtHTT of physiological importance (Fig. 5.5). Further work is required to understand the nature of this species and its potential function, if any, in R-loop biology.

Among other R-loop and transcriptional factors, HTT co-IP experiments revealed a novel interaction of the HTT protein with XPG, an endonuclease involved in DNA damage repair and R-loop resolution (Sollier *et al.*, 2014). XPG, originally identified as the causative gene for Xeroderma Pigmentosum (XP) complementation group G (Zafeiriou *et al.*, 2001), is a component of the

nucleotide excision repair (NER) pathway, playing a role in both global genomic NER (GG-NER) and transcription-coupled NER (TCR). Mutations in the NER pathway lead to various forms of XP disorders (XP-A to XP-G and XP-V), which are rare autosomal recessive neurodegenerative diseases (Fassihi *et al.*, 2016). Interestingly, two families of XP disorders, XP-G, involving mutations of the *XP-G/ERCC5* gene, and XP-B, caused by mutations in the *XP-B/ERCC3* gene, display neurological features that closely resemble those of HD, including chorea and cognitive impairment (Garcia-Moreno *et al.*, 2018). This suggests that DNA repair disorders may mimic HD and that the NER pathway, and XPG protein in particular, may play a significant role in HD pathogenesis. In support of this, the expression of HTT and XPG proteins was shown to be co-regulated (Fig. 5.10). Furthermore, the co-depletion of XPG and HTT proteins in HeLa cells resulted in inflammatory responses, implying that these two proteins are involved in the same pathway (Fig. 5.12). Further investigation is needed to define the nature of HTT-XPG interaction and the mechanism by which these proteins participate in DNA damage and R-loop resolution.

6.7 HTT and DNA damage

Dysregulation of R-loops has previously been associated with increased DNA damage, which can be attributed to the presence of displaced ssDNA in the R-loop structure, as well as to various R-loop clearance pathways that generate double-strand breaks (DSBs) (Sollier *et al.*, 2014; Richard & Manley, 2017). DNA damage accumulation in HD cells has been described here (Fig. 5.7-5.8), and in the literature, and is typically attributed to defects in DNA repair (Massey & Jones, 2018; Gao *et al.*, 2019). However, it is possible that DNA damage in HD cells also arises due to accumulation of R-loops. To investigate the potential interplay between R-loops and DNA damage, I took an advantage of CPT, a highly specific inhibitor for Topoisomerase I (Top1), which traps Top1 on DNA and forms Top1 cleavage complexes (Top1ccs), that can lead to DNA DSBs if not repaired properly (Capranico *et al.*, 2010). CPT is cytotoxic to neuronal cells as it causes transcription-blocking DSBs and R-loop accumulation (Cristini *et al.*, 2019). Interestingly, HTT-depleted HeLa cells treated with CPT have more DNA damage and R-loops than CPT-treated Control cells (Fig. 5.8-5.9). In line with this, a recent study has shown that HD fibroblasts have a defect in the repair of

CPT-induced DNA damage and accumulate Top1ccs which are then turned into cytotoxic DSBs (Pahlmina *et al.,* 2022). Since Top1ccs stabilization and DSB formation at actively transcribed regions promote R-loop formation (Aguilera & García-Muse, 2012), an intriguing possibility is that the accumulation of R-loops, observed in siHTT-treated and in HD cells, could originate from these Top1cc lesions. Further to this point, XPG protein is involved in the resolution of Top1cc lesions. In details, the XPG protein, together with XPF and FEN1 nucleases, mediates the dual incision of the R-loops and generates SSBs (Cristini *et al.*, 2019). To gather a further understanding of the role of R-loops in HD pathology, it would be interesting to map the genomic location of Top1cc by ChIP-seq as well as to localize DSBs by γH2AX ChIP-seq. Furthermore, since I observed the interaction between wtHTT and both Top 1 and XPG (Fig. 5.4), it would be also interesting to map the genomic localization of HTT, XPG and Top1 proteins in HD cells. These experiments will help to determine if R-loops co-localise with DSBs and if the HTT protein is present at these lesion sites and plays a direct role in DSBs and R-loop resolution alongside XPG protein.

Notably, non-replicating cells, such as neurons, are particularly affected by defects in the Top1cc removal pathway and/or R-loop resolution (Cristini *et al.*, 2020). Indeed, the high metabolic activity of neurons may promote reactive oxygen species (ROS)-dependent Top1cc trapping followed by DNA damage and R-loop accumulation. Consequently, defects in the Top1cc removal pathway and/or in R-loop resolution are features of several neurodegenerative disorders, including ALS, SCA1 and others (Cristini *et al.*, 2020). Therefore, both Top1cc lesions and R-loop accumulation may also contribute to the HD neurodegenerative phenotype.

It has been recently proposed that transcription-blocking DNA damage during normal ageing can cause genome-wide Pol II stalling. This, in turn, results in reduced basal transcription and dysregulation of several ageing pathways (Gyenis *et al.,* 2023). Particularly, *Xpg*$^{-/-}$ mice showed a correlation between transcriptional decline, DNA repair deficiency and pathology. On the basis of the data showed in this thesis, it can thus be speculated that in HD cells Pol II stalling at DNA lesions could trigger R-loop formation and activate the DNA damage checkpoint ATM. This might contribute to cell death and senescence, thereby establishing a molecular link between transcription dysregulation, R-loops and DNA damage in HD.

6.8 HTT and inflammation

Inflammation in the brain and peripheral tissues is a well-documented hallmark of HD, which often precedes pathological symptoms (Rocha *et al.,* 2016). However, the specific trigger for this inflammatory signalling in HD cells has yet to be determined. RNA/DNA hybrids have been shown to activate several innate immune sensors, including the DNA damage sensor cGAS, resulting in inflammatory responses through activation of the cGAS-STING pathway (Weinreb *et al.*, 2021; Cristini *et al.*, 2022; Giordano *et al.*, 2022). Moreover, harmful R-loops can also stimulate an immune response by promoting micronuclei formation (De Magis *et al*, 2019). Additionally, DNA damage-driven R-loops have been shown to trigger the release of cytoplasmic ssDNAs, leading to chronic inflammation (Chatzidoukaki *et al.,* 2021). Interestingly, cGAS has been shown to be upregulated in HD cells and it has been proposed that activation of the cGAS-STING pathway may contribute to inflammatory responses observed in HD (Sharma *et al.*, 2020). The activation of the cGAS-STING pathway can lead to the downstream expression of several ISGs and various chemokines and cytokines, including Ccl5 and Cxcl10, which were shown to be upregulated in HD in this work and in previous studies (Dong & Cong, 2018; Sharma *et al.*, 2020). Furthermore, cytoplasmic micronuclei formation has been shown in HD (Sharma *et al.*, 2020). Therefore, I propose that the R-loop accumulation observed in HD cells could be a trigger for inflammation in HD. Indeed, it is noteworthy that the levels of R-loops in HD cells, both lymphoblasts and MSNs, correlate with the activation of several inflammatory responses (Fig. 5.11). In the context of AGS, it was shown that persistent R-loops accumulate in RNase H2-deficient cells and they are converted into DNA breaks through XPF- and XPG-mediated cleavage, triggering DNA damage and inflammatory response (Cristini *et al.,* 2022). It was also shown that depletion of XPG or XPF decreased the DNA damage and inflammation observed in RNase H2A-deficient cells, indicating that DNA breaks and inflammation may result from XPG and/or XPF-mediated R-loop cleavage. However, the data in this study shows that DNA damage and inflammation are upregulated in cells co-depleted for XPG and HTT. It is known that R-loop cleavage needs a functional TCR complex (Sollier *et al.*, 2014). Furthermore, it was recently demonstrated that HTT forms a TCR complex with Pol II and its

function is abrogated in HD (Gao *et al.,* 2019). Particularly, mutant HTT disrupts TCR by interfering with the activity of components of the TCR complex, including the DNA repair enzyme polynucleotide-kinase-3' phosphatase PNKP, ataxin-3 and the cAMP response element (CREB)-binding protein (CPB), resulting in the accumulation of DSBs in actively transcribed regions (Gao *et al.,* 2019). In line with the roles of XPG and HTT protein in TCR, mentioned above, it can be speculated that loss and/or mutation of HTT disrupts the functional integrity of the TCR complex, resulting in a dramatic increase in stalled transcription, DNA damage and R-loop formation which cannot be properly resolved by XPG. Therefore, this unveils a novel mechanism for R-loop-triggered inflammation in disease (Fig 6.3). It will be interesting to further investigate potential additional players required to process these pathological R-loops, triggering DNA damage and inflammation in HD.

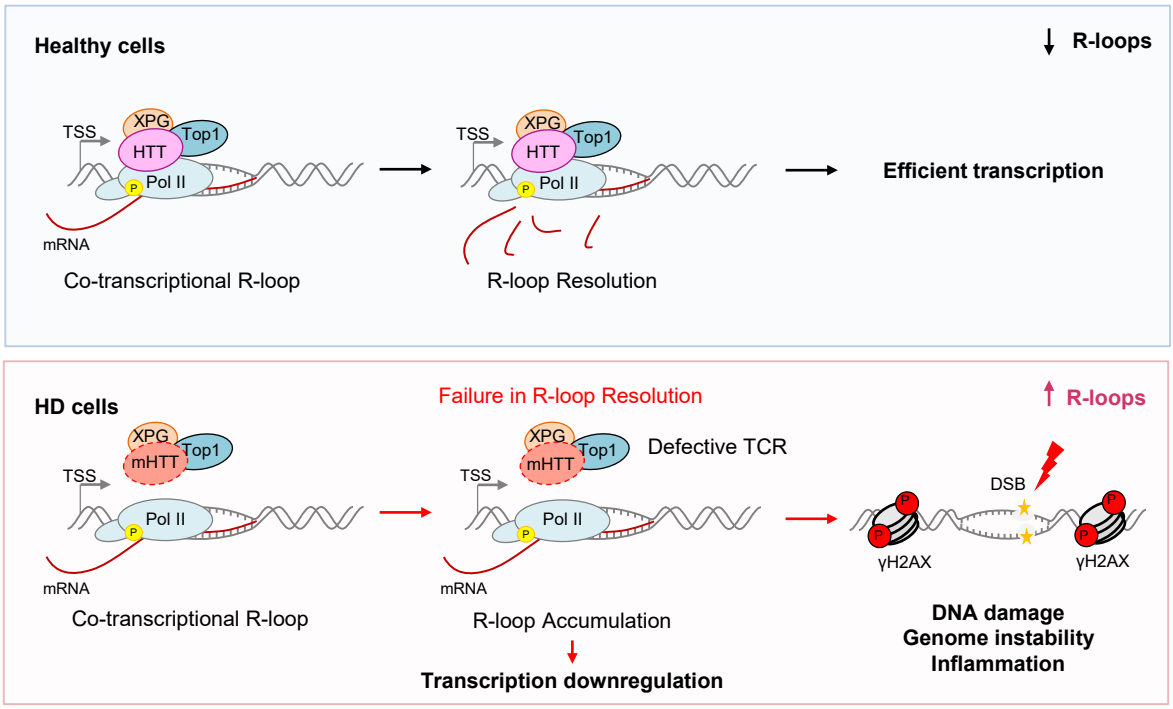


Figure 6.3. Interplay between R-loops, DNA damage and inflammation in HD.
Top panel. In healthy cells, HTT interacts with Pol II during transcription and R-loop factors. This complex resolves co-transcriptional R-loops exceeding the physiological levels, allowing efficient transcription. **Bottom panel.** Absence or mutation of HTT results in defective TCR, accumulation of R-loops, especially at genomic repetitive regions, which could affect genome stability. Furthermore, R-loop accumulation could contribute to transcription down-regulation, DNA damage and inflammation, characteristic of HD pathology.; XPG, Xeroderma pigmentosum group G endonuclease; Pol Il, RNA Polymerase II, DSB, double-strand break; TCR, Transcription coupled repair; Topi I, Topoisomerase I; TSS, Transcriptional Start Site; γH2AX, phosphorylation of the Ser-139 residue of the histone variant H2AX. P, phosphorylation.

6.9 Conclusions and future perspectives

Both loss-of-function of wtHTT and gain-of-function of mHTT protein contribute to HD. However, the contribution of wtHTT and mHTT to the mechanisms described in this work remains unclear. Since HTT protein is involved in transcriptional processes, I opted for transient RNAi-mediated HTT depletion to lower the level of HTT protein. The ChIP-qPCR for HTT protein in HTT-depleted HeLa cells showed a partial nuclear decrease of HTT (Fig. 5.2). However, it was recently shown that siRNAs can silence nuclear *HTT* mRNA with low efficiency, probably due to the higher stability or retention of nuclear *HTT* mRNA compared to cytoplasmic *HTT* mRNA (Didiot *et al.*, 2018). Therefore, degron-based approaches will offer a rapid and specific degradation of the HTT protein, minimizing perturbations to the cellular system. Additionally, it would be interesting to overexpress the mHTT protein, either the full-length and/or the toxic exon 1 HTT fragment, and study the changes in transcription and R-loop homeostasis. Alternatively, over-expression of wild-type HTT could have a protective effect.

The present study has provided novel insights into the role of R-loops in HD (Fig. 6.3). However, the correlation between R-loop accumulation, increase in DSBs and inflammatory signalling described in Chapter 5 requires further insights. In order to determine whether R-loops are the cause of increased DNA damage and inflammatory signalling in HD cells, the levels of R-loops could be modulated in diseased/knock-down cells and the effects on DNA damage and inflammation monitored. For example, R-loop levels can be decreased by over-expression of RNase H *in vivo*, allowing the specific depletion of R-loops. Moreover, modulation of R-loops could be achieved via the over-expression of different R-loop resolution factors, resulting in changes in R-loop levels. These approaches were successfully used to assess the role of R-loop homeostasis in the context of AGS (Cristini *et al.,* 2022). Alternatively, the DNA damage response could be modulated by ATM inhibitors, while monitoring R-loop levels.

The HTT protein is involved in many different cellular processes. However, its role in regulating genome integrity remains unclear. Here, I argue that the HTT protein can act as a scaffold for resolving R-loops, thereby directly playing a role in the regulation of the genome integrity (Fig.

6.3). Interestingly, our DRIP-seq revealed that many of the mis-regulated R-loops reside in highly repetitive genomic regions, suggesting that these structures may occur at specific places and not randomly in the genome, resulting in genome instability. Further investigation into the occurrence of R-loops accumulation in repetitive genomic regions is needed to understand the interplay between HTT protein, R-loop homeostasis and genome stability. This is particularly relevant considering that it was recently shown that ~50% reduction in mHTT protein reduces age-dependent CAG somatic instability at the *HTT* locus and at *Atxn2* CAG-containing locus in knock-in HD mice and human cells, suggesting a broader role for the HTT protein in regulating expanded repeat loci in the genome (Coffey *et al.*, 2020). Therefore, the proposed role of HTT protein as a scaffold for R-loop resolution may be in line with the role of the HTT protein in the regulation of somatic instability of long CAG tracts and of other repetitive sequences in general.

Overall, the work presented here provides novel insights into the molecular mechanisms driving HD pathogenesis by successfully characterizing the transcriptional regulation of the *HTT* gene and its epigenetic signature in human cells derived from HD patients. Transcription is down-regulated at the *HTT* gene in HD cells and the transcriptional impairment correlates with epigenetic silencing of the gene. Furthermore, dsRNA structures form over expanded CAG repeats, thus implicating them in repeat instability and HD pathogenesis. Finally, novel insights into the role of the HTT protein in the nucleus and the contribution of R-loops to HD pathogenesis, described here, open a new avenue of research in the field. In the future, this work may lead to new therapies targeting R-loops and dsRNAs in HD, particularly if they are shown to be causative of genome instability and neuroinflammation in HD neurons.

References

Abu Diab M, Mor-Shaked H, Cohen E, Cohen-Hadad Y, Ram O, Epsztejn-Litman S, Eiges R (2018) The G-rich Repeats in FMR1 and C9orf72 Loci Are Hotspots for Local Unpairing of DNA. *Genetics* 210: 1239-1252

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

Aktas T, Avsar Ilik I, Maticzka D, Bhardwaj V, Pessoa Rodrigues C, Mittler G, Manke T, Backofen R, Akhtar A (2017) DHX9 suppresses RNA processing defects originating from the Alu invasion of the human genome. *Nature* 544: 115-119

Al-Dalahmah O, Sosunov AA, Shaik A, Ofori K, Liu Y, Vonsattel JP, Adorjan I, Menon V, Goldman JE (2020) Single-nucleus RNA-seq identifies Huntington disease astrocyte states. *Acta Neuropathol Commun* 8: 19

Alcala-Vida R, Seguin J, Lotz C, Molitor AM, Irastorza-Azcarate I, Awada A, Karasu N, Bombardier A, Cosquer B, Skarmeta JLG *et al* (2021) Age-related and disease locus-specific mechanisms contribute to early remodelling of chromatin structure in Huntington's disease mice. *Nat Commun* 12: 364

Ament SA, Pearl JR, Grindeland A, St Claire J, Earls JC, Kovalenko M, Gillis T, Mysore J, Gusella JF, Lee JM *et al* (2017) High resolution time-course mapping of early transcriptomic, molecular and cellular phenotypes in Huntington's disease CAG knock-in mice across multiple genetic backgrounds. *Hum Mol Genet* 26: 913-922

Andrade MA, Bork P (1995) HEAT repeats in the Huntington's disease protein. *Nat Genet* 11: 115-116

Arab K, Karaulanov E, Musheev M, Trnka P, Schafer A, Grummt I, Niehrs C (2019) GADD45A binds R-loops and recruits TET1 to CpG island promoters. *Nat Genet* 51: 217-223

Aronin N, Chase K, Young C, Sapp E, Schwarz C, Matta N, Kornreich R, Landwehrmeyer B, Bird E, Beal MF *et al* (1995) CAG expansion affects the expression of mutant Huntingtin in the Huntington's disease brain. *Neuron* 15: 1193-1201

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

Atwal RS, Xia J, Pinchev D, Taylor J, Epand RM, Truant R (2007) Huntingtin has a membrane association signal that can modulate huntingtin aggregation, nuclear entry and toxicity. *Hum Mol Genet* 16: 2600-2615

Axford MM, Wang YH, Nakamori M, Zannis-Hadjopoulos M, Thornton CA, Pearson CE (2013) Detection of slipped-DNAs at the trinucleotide repeats of the myotonic dystrophy type I disease locus in patient tissues. *PLoS Genet* 9: e1003866

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

Barbaro BA, Lukacsovich T, Agrawal N, Burke J, Bornemann DJ, Purcell JM, Worthge SA, Caricasole A, Weiss A, Song W *et al* (2015) Comparative study of naturally occurring huntingtin fragments in Drosophila points to exon 1 as the most pathogenic species in Huntington's disease. *Hum Mol Genet* 24: 913-925

Barnes GT, Duyao MP, Ambrose CM, McNeil S, Persichetti F, Srinidhi J, Gusella JF, MacDonald ME (1994) Mouse Huntington's disease gene homolog (Hdh). *Somat Cell Mol Genet* 20: 87-97

Barski A, Cuddapah S, Cui K, Roh TY, Schones DE, Wang Z, Wei G, Chepelev I, Zhao K (2007) High-resolution profiling of histone methylations in the human genome. *Cell* 129: 823-837

Bassi S, Tripathi T, Monziani A, Di Leva F, Biagioli M (2017) Epigenetics of Huntington's Disease. *Adv Exp Med Biol* 978: 277-299

Bates GP, Dorsey R, Gusella JF, Hayden MR, Kay C, Leavitt BR, Nance M, Ross CA, Scahill RI, Wetzel R *et al* (2015) Huntington disease. *Nat Rev Dis Primers* 1: 15005

Bauer M, Nascakova Z, Mihai AI, Cheng PF, Levesque MP, Lampart S, Hurwitz R, Pfannkuch L, Dobrovolna J, Jacobs M *et al* (2020) The ALPK1/TIFA/NF-kappaB axis links a bacterial carcinogen to R-loop-induced replication stress. *Nat Commun* 11: 5117

Becanovic K, Norremolle A, Neal SJ, Kay C, Collins JA, Arenillas D, Lilja T, Gaudenzi G, Manoharan S, Doty CN *et al* (2015) A SNP in the HTT promoter alters NF-kappaB binding and is a bidirectional genetic modifier of Huntington disease. *Nat Neurosci* 18: 807-816

Becherel OJ, Sun J, Yeo AJ, Nayler S, Fogel BL, Gao F, Coppola G, Criscuolo C, De Michele G, Wolvetang E (2015) A new model to study neurodegeneration in ataxia oculomotor apraxia type 2. *Human Molecular Genetics* 24: 5759-5774

Beck DB, Narendra V, Drury WJ, 3rd, Casey R, Jansen PW, Yuan ZF, Garcia BA, Vermeulen M, Bonasio R (2014) In vivo proximity labeling for the detection of protein-protein and protein-RNA interactions. *J Proteome Res* 13: 6135-6143

Beghe C, Gromak N (2022) R-Loop Immunoprecipitation: A Method to Detect R-Loop Interacting Factors. *Methods Mol Biol* 2528: 215-237

Benn CL, Landles C, Li H, Strand AD, Woodman B, Sathasivam K, Li SH, Ghazi-Noori S, Hockly E, Faruque SM *et al* (2005) Contribution of nuclear and extranuclear polyQ to neurological phenotypes in mouse models of Huntington's disease. *Hum Mol Genet* 14: 3065-3078

Benn CL, Sun T, Sadri-Vakili G, McFarland KN, DiRocco DP, Yohrling GJ, Clark TW, Bouzou B, Cha JH (2008) Huntingtin modulates transcription, occupies gene promoters in vivo, and binds directly to DNA in a polyglutamine-dependent manner. *J Neurosci* 28: 10720-10733

Bhatia V, Barroso SI, Garcia-Rubio ML, Tumini E, Herrera-Moyano E, Aguilera A (2014) BRCA2 prevents R-loop accumulation and associates with TREX-2 mRNA export factor PCID2. *Nature* 511: 362-365

Biagioli M, Ferrari F, Mendenhall EM, Zhang Y, Erdin S, Vijayvargia R, Vallabh SM, Solomos N, Manavalan P, Ragavendran A *et al* (2015) Htt CAG repeat expansion confers pleiotropic gains of mutant huntingtin function in chromatin regulation. *Hum Mol Genet* 24: 2442-2457

Bizat N, Hermel JM, Boyer F, Jacquard C, Creminon C, Ouary S, Escartin C, Hantraye P, Kajewski S, Brouillet E (2003) Calpain is a major cell death effector in selective striatal degeneration induced in vivo by 3-nitropropionate: implications for Huntington's disease. *J Neurosci* 23: 5020-5030

Bjorkqvist M, Wild EJ, Thiele J, Silvestroni A, Andre R, Lahiri N, Raibon E, Lee RV, Benn CL, Soulet D *et al* (2008) A novel pathogenic pathway of immune activation detectable before clinical onset in Huntington's disease. *J Exp Med* 205: 1869-1877

Bogliotti YS, Ross PJ (2012) Mechanisms of histone H3 lysine 27 trimethylation remodeling during early mammalian development. *Epigenetics* 7: 976-981

Boguslawski SJ, Smith DE, Michalak MA, Mickelson KE, Yehle CO, Patterson WL, Carrico RJ (1986) Characterization of monoclonal antibody to DNA.RNA and its application to immunodetection of hybrids. *J Immunol Methods* 89: 123-130

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

Borst K, Dumas AA, Prinz M (2021) Microglia: Immune and non-immune functions. *Immunity* 54: 2194-2208

Bou-Nader C, Bothra A, Garboczi DN, Leppla SH, Zhang J (2022) Structural basis of R-loop recognition by the S9. 6 monoclonal antibody. *Nature Communications* 13: 1641

Bowman EA, Kelly WG (2014) RNA polymerase II transcription elongation and Pol II CTD Ser2 phosphorylation: A tail of two kinases. *Nucleus* 5: 224-236

Bradford J, Shin JY, Roberts M, Wang CE, Li XJ, Li S (2009) Expression of mutant huntingtin in mouse brain astrocytes causes age-dependent neurological symptoms. *Proc Natl Acad Sci U S A* 106: 22480-22485

Branon TC, Bosch JA, Sanchez AD, Udeshi ND, Svinkina T, Carr SA, Feldman JL, Perrimon N, Ting AY (2018) Efficient proximity labeling in living cells and organisms with TurboID. *Nat Biotechnol* 36: 880-887

Burns MB, Temiz NA, Harris RS (2013) Evidence for APOBEC3B mutagenesis in multiple human cancers. *Nature genetics* 45: 977-983

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

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

Carmo C, Naia L, Lopes C, Rego AC (2018) Mitochondrial Dysfunction in Huntington's Disease. *Adv Exp Med Biol* 1049: 59-83

Caron NS, Desmond CR, Xia J, Truant R (2013) Polyglutamine domain flexibility mediates the proximity between flanking sequences in huntingtin. *Proc Natl Acad Sci U S A* 110: 14610-14615

Casellas R, Basu U, Yewdell WT, Chaudhuri J, Robbiani DF, Di Noia JM (2016) Mutations, kataegis and translocations in B cells: understanding AID promiscuous activity. *Nat Rev Immunol* 16: 164-176

Cassidy SB, Dykens E, Williams CA (2000) Prader-Willi and Angelman syndromes: Sister imprinted disorders. *American journal of medical genetics* 97: 136-146

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

Caviston JP, Holzbaur EL (2009) Huntingtin as an essential integrator of intracellular vesicular trafficking. *Trends Cell Biol* 19: 147-155

Cerritelli SM, Crouch RJ (2009) Ribonuclease H: the enzymes in eukaryotes. *FEBS J* 276: 1494-1505

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

Chatzidoukaki O, Stratigi K, Goulielmaki E, Niotis G, Akalestou-Clocher A, Gkirtzimanaki K, Zafeiropoulos A, Altmüller J, Topalis P, Garinis GA (2021) R-loops trigger the release of cytoplasmic ssDNAs leading to chronic inflammation upon DNA damage. *Science Advances* 19;7(47):eabj5769.

Chedin F, Hartono SR, Sanz LA, Vanoosthuyse V (2021) Best practices for the visualization, mapping, and manipulation of R-loops. *EMBO J* 40: e106394

Chen L, Chen J-Y, Huang Y-J, Gu Y, Qiu J, Qian H, Shao C, Zhang X, Hu J, Li H (2018) The augmented R-loop is a unifying mechanism for myelodysplastic syndromes induced by high-risk splicing factor mutations. *Molecular cell* 69: 412-425. e416

Chen L, Chen J-Y, Zhang X, Gu Y, Xiao R, Shao C, Tang P, Qian H, Luo D, Li H (2017) R-ChIP using inactive RNase H reveals dynamic coupling of R-loops with transcriptional pausing at gene promoters. *Molecular cell* 68: 745-757. e745

Chen PB, Chen HV, Acharya D, Rando OJ, Fazzio TG (2015) R loops regulate promoter-proximal chromatin architecture and cellular differentiation. *Nature structural & molecular biology* 22: 999-1007

Chervova A, Fatykhov B, Koblov A, Shvarov E, Preobrazhenskaya J, Vinogradov D, Ponomarev GV, Gelfand MS, Kazanov MD (2021) Analysis of gene expression and mutation data points on contribution of transcription to the mutagenesis by APOBEC enzymes. *NAR cancer* 3: zcab025

Cho DH, Thienes CP, Mahoney SE, Analau E, Filippova GN, Tapscott SJ (2005) Antisense transcription and heterochromatin at the DM1 CTG repeats are constrained by CTCF. *Mol Cell* 20: 483-489

Choo YS, Johnson GV, MacDonald M, Detloff PJ, Lesort M (2004) Mutant huntingtin directly increases susceptibility of mitochondria to the calcium-induced permeability transition and cytochrome c release. *Hum Mol Genet* 13: 1407-1420

Chung DW, Rudnicki DD, Yu L, Margolis RL (2011) A natural antisense transcript at the Huntington's disease repeat locus regulates HTT expression. *Hum Mol Genet* 20: 3467-3477

Chutake YK, Costello WN, Lam CC, Parikh AC, Hughes TT, Michalopulos MG, Pook MA, Bidichandani SI (2015) FXN Promoter Silencing in the Humanized Mouse Model of Friedreich Ataxia. *PLoS One* 10: e0138437

Cloutier SC, Wang S, Ma WK, Al Husini N, Dhoondia Z, Ansari A, Pascuzzi PE, Tran EJ (2016) Regulated formation of lncRNA-DNA hybrids enables faster transcriptional induction and environmental adaptation. *Molecular cell* 61: 393-404

Coffey SR, Andrew M, Ging H, Hamilton J, Flower M, Kovalenko M, Bragg RM, Cantle JP, McHugh CA, Carrillo JM *et al* (2020) Huntingtin lowering reduces somatic instability at CAG-expanded loci. *bioRxiv*: 2020.2007.2023.218347

Colak D, Zaninovic N, Cohen MS, Rosenwaks Z, Yang WY, Gerhardt J, Disney MD, Jaffrey SR (2014) Promoter-bound trinucleotide repeat mRNA drives epigenetic silencing in fragile X syndrome. *Science* 343: 1002-1005

Colin E, Zala D, Liot G, Rangone H, Borrell-Pages M, Li XJ, Saudou F, Humbert S (2008) Huntingtin phosphorylation acts as a molecular switch for anterograde/retrograde transport in neurons. *EMBO J* 27: 2124-2134

Conroy F, Miller R, Alterman JF, Hassler MR, Echeverria D, Godinho B, Knox EG, Sapp E, Sousa J, Yamada K *et al* (2022) Chemical engineering of therapeutic siRNAs for allele-specific gene silencing in Huntington's disease models. *Nat Commun* 13: 5802

Core LJ, Martins AL, Danko CG, Waters CT, Siepel A, Lis JT (2014) Analysis of nascent RNA identifies a unified architecture of initiation regions at mammalian promoters and enhancers. *Nat Genet* 46: 1311-1320

Core LJ, Waterfall JJ, Lis JT (2008) Nascent RNA sequencing reveals widespread pausing and divergent initiation at human promoters. *Science* 322: 1845-1848

Cornett J, Cao F, Wang CE, Ross CA, Bates GP, Li SH, Li XJ (2005) Polyglutamine expansion of huntingtin impairs its nuclear export. *Nat Genet* 37: 198-204

Cristini A, Geraud M, Sordet O (2021) Transcription-associated DNA breaks and cancer: A matter of DNA topology. *Int Rev Cell Mol Biol* 364: 195-240

Cristini A, Groh M, Kristiansen MS, Gromak N (2018) RNA/DNA Hybrid Interactome Identifies DXH9 as a Molecular Player in Transcriptional Termination and R-Loop-Associated DNA Damage. *Cell Rep* 23: 1891-1905

Cristini A, Gromak N, Sordet O (2020) Transcription-dependent DNA double-strand breaks and human disease. *Mol Cell Oncol* 7: 1691905

Cristini A, Ricci G, Britton S, Salimbeni S, Huang SN, Marinello J, Calsou P, Pommier Y, Favre G, Capranico G *et al* (2019) Dual Processing of R-Loops and Topoisomerase I Induces Transcription-Dependent DNA Double-Strand Breaks. *Cell Rep* 28: 3167-3181 e3166

Cristini A, Tellier M, Constantinescu F, Accalai C, Albulescu LO, Heiringhoff R, Bery N, Sordet O, Murphy S, Gromak N (2022) RNase H2, mutated in Aicardi-Goutieres syndrome, resolves co-transcriptional R-loops to prevent DNA breaks and inflammation. *Nat Commun* 13: 2961

Crossley MP, Bocek M, Cimprich KA (2019) R-Loops as Cellular Regulators and Genomic Threats. *Mol Cell* 73: 398-411

Crossley MP, Bocek MJ, Hamperl S, Swigut T, Cimprich KA (2020) qDRIP: a method to quantitatively assess RNA–DNA hybrid formation genome-wide. *Nucleic acids research* 48: e84-e84

Crossley MP, Song C, Bocek MJ, Choi JH, Kousorous J, Sathirachinda A, Lin C, Brickner JR, Bai G, Lans H *et al* (2023) R-loop-derived cytoplasmic RNA-DNA hybrids activate an immune response. *Nature* 613: 187-194

Crow YJ, Hayward BE, Parmar R, Robins P, Leitch A, Ali M, Black DN, Van Bokhoven H, Brunner HG, Hamel BC (2006) Mutations in the gene encoding the 3′-5′ DNA exonuclease TREX1 cause Aicardi-Goutieres syndrome at the AGS1 locus. *Nature genetics* 38: 917-920

Cui L, Jeong H, Borovecki F, Parkhurst CN, Tanese N, Krainc D (2006) Transcriptional repression of PGC-1alpha by mutant huntingtin leads to mitochondrial dysfunction and neurodegeneration. *Cell* 127: 59-69

De Magis A, Manzo SG, Russo M, Marinello J, Morigi R, Sordet O, Capranico G (2019) DNA damage and genome instability by G-quadruplex ligands are mediated by R loops in human cancer cells. *Proceedings of the National Academy of Sciences* 116: 816-825

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

De Rooij KE, Dorsman JC, Smoor MA, Den Dunnen JT, Van Ommen GJ (1996) Subcellular localization of the Huntington's disease gene product in cell lines by immunofluorescence and biochemical subcellular fractionation. *Hum Mol Genet* 5: 1093-1099

De Souza RAG, Kosior N, Thomson SB, Mathelier A, Zhang AW, Becanovic K, Wasserman WW, Leavitt BR (2018) Computational Analysis of Transcriptional Regulation Sites at the HTT Gene Locus. *J Huntingtons Dis* 7: 223-237

Delatycki MB, Williamson R, Forrest SM (2000) Friedreich ataxia: an overview. *J Med Genet* 37: 1-8

Depienne C, Mandel JL (2021) 30 years of repeat expansion disorders: What have we learned and what are the remaining challenges? *Am J Hum Genet* 108: 764-785

Didiot MC, Ferguson CM, Ly S, Coles AH, Smith AO, Bicknell AA, Hall LM, Sapp E, Echeverria D, Pai AA *et al* (2018) Nuclear Localization of Huntingtin mRNA Is Specific to Cells of Neuronal Origin. *Cell Rep* 24: 2553-2560 e2555

DiGiovanni LF, Mocle AJ, Xia J, Truant R (2016) Huntingtin N17 domain is a reactive oxygen species sensor regulating huntingtin phosphorylation and localization. *Hum Mol Genet* 25: 3937-3945

Dixon KT, Cearley JA, Hunter JM, Detloff PJ (2004) Mouse Huntington's disease homolog mRNA levels: variation and allele effects. *Gene Expr* 11: 221-231

Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR (2013) STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 29: 15-21

Dong X, Cong S (2018) Identification of differentially expressed genes and regulatory relationships in Huntington's disease by bioinformatics analysis. *Mol Med Rep* 17: 4317-4326

Dragatsis I, Levine MS, Zeitlin S (2000) Inactivation of Hdh in the brain and testis results in progressive neurodegeneration and sterility in mice. *Nat Genet* 26: 300-306

Drolet M, Phoenix P, Menzel R, Masse E, Liu LF, Crouch RJ (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. *Proc Natl Acad Sci U S A* 92: 3526-3530

Duyao MP, Auerbach AB, Ryan A, Persichetti F, Barnes GT, McNeil SM, Ge P, Vonsattel JP, Gusella JF, Joyner AL *et al* (1995) Inactivation of the mouse Huntington's disease gene homolog Hdh. *Science* 269: 407-410

Elias S, Thion MS, Yu H, Sousa CM, Lasgi C, Morin X, Humbert S (2014) Huntingtin regulates mammary stem cell division and differentiation. *Stem Cell Reports* 2: 491-506

Essebier A, Vera Wolf P, Cao MD, Carroll BJ, Balasubramanian S, Boden M (2016) Statistical Enrichment of Epigenetic States Around Triplet Repeats that Can Undergo Expansions. *Front Neurosci* 10: 92

Evans-Galea MV, Hannan AJ, Carrodus N, Delatycki MB, Saffery R (2013) Epigenetic modifications in trinucleotide repeat diseases. *Trends Mol Med* 19: 655-663

Evers MM, Miniarikova J, Juhas S, Valles A, Bohuslavova B, Juhasova J, Skalnikova HK, Vodicka P, Valekova I, Brouwers C *et al* (2018) AAV5-miHTT Gene Therapy Demonstrates Broad Distribution and Strong Human Mutant Huntingtin Lowering in a Huntington's Disease Minipig Model. *Mol Ther* 26: 2163-2177

Evers MM, Schut MH, Pepers BA, Atalar M, van Belzen MJ, Faull RL, Roos RA, van Roon-Mom WM (2015) Making (anti-) sense out of huntingtin levels in Huntington disease. *Mol Neurodegener* 10: 21

Faber PW, Barnes GT, Srinidhi J, Chen J, Gusella JF, MacDonald ME (1998) Huntingtin interacts with a family of WW domain proteins. *Hum Mol Genet* 7: 1463-1474

Fassihi H, Sethi M, Fawcett H, Wing J, Chandler N, Mohammed S, Craythorne E, Morley AM, Lim R, Turner S *et al* (2016) Deep phenotyping of 89 xeroderma pigmentosum patients reveals unexpected heterogeneity dependent on the precise molecular defect. *Proc Natl Acad Sci U S A* 113: E1236-1245

Feretzaki M, Pospisilova M, Valador Fernandes R, Lunardi T, Krejci L, Lingner J (2020) RAD51-dependent recruitment of TERRA lncRNA to telomeres through R-loops. *Nature* 587: 303-308

Fjodorova M, Li M (2018) Robust Induction of DARPP32-Expressing GABAergic Striatal Neurons from Human Pluripotent Stem Cells. *Methods Mol Biol* 1780: 585-605

Flower M, Lomeikaite V, Ciosi M, Cumming S, Morales F, Lo K, Hensman Moss D, Jones L, Holmans P, Investigators T-H *et al* (2019) MSH3 modifies somatic instability and disease severity in Huntington's and myotonic dystrophy type 1. *Brain* 142: 1876-1886

Fotsing SF, Margoliash J, Wang C, Saini S, Yanicky R, Shleizer-Burko S, Goren A, Gymrek M (2019) The impact of short tandem repeat variation on gene expression. *Nat Genet* 51: 1652-1659

Franich NR, Basso M, Andre EA, Ochaba J, Kumar A, Thein S, Fote G, Kachemov M, Lau AL, Yeung SY *et al* (2018) Striatal Mutant Huntingtin Protein Levels Decline with Age in Homozygous Huntington's Disease Knock-In Mouse Models. *J Huntingtons Dis* 7: 137-150

Franich NR, Hickey MA, Zhu C, Osborne GF, Ali N, Chu T, Bove NH, Lemesre V, Lerner RP, Zeitlin SO *et al* (2019) Phenotype onset in Huntington's disease knock-in mice is correlated with the incomplete splicing of the mutant huntingtin gene. *J Neurosci Res* 97: 1590-1605

Frank S (2009) Tetrabenazine as anti-chorea therapy in Huntington disease: an open-label continuation study. Huntington Study Group/TETRA-HD Investigators. *BMC Neurol* 9: 62

Freudenreich CH (2018) R-loops: targets for nuclease cleavage and repeat instability. *Curr Genet* 64: 789-794

Futter M, Diekmann H, Schoenmakers E, Sadiq O, Chatterjee K, Rubinsztein DC (2009) Wild-type but not mutant huntingtin modulates the transcriptional activity of liver X receptors. *J Med Genet* 46: 438-446

Gafni J, Hermel E, Young JE, Wellington CL, Hayden MR, Ellerby LM (2004) Inhibition of calpain cleavage of huntingtin reduces toxicity: accumulation of calpain/caspase fragments in the nucleus. *J Biol Chem* 279: 20211-20220

Gao R, Chakraborty A, Geater C, Pradhan S, Gordon KL, Snowden J, Yuan S, Dickey AS, Choudhary S, Ashizawa T *et al* (2019) Mutant huntingtin impairs PNKP and ATXN3, disrupting DNA repair and transcription. *Elife* 8

Gao YG, Yang H, Zhao J, Jiang YJ, Hu HY (2014) Autoinhibitory structure of the WW domain of HYPB/SETD2 regulates its interaction with the proline-rich region of huntingtin. *Structure* 22: 378-386

Garcia-Moreno H, Fassihi H, Sarkany RPE, Phukan J, Warner T, Lehmann AR, Giunti P (2018) Xeroderma pigmentosum is a definite cause of Huntington's disease-like syndrome. *Ann Clin Transl Neurol* 5: 102-108

Garcia-Muse T, Aguilera A (2019) R Loops: From Physiological to Pathological Roles. *Cell* 179: 604-618

Garcia-Rubio ML, Perez-Calero C, Barroso SI, Tumini E, Herrera-Moyano E, Rosado IV, Aguilera A (2015) The Fanconi Anemia Pathway Protects Genome Integrity from R-loops. *PLoS Genet* 11: e1005674

Gemayel R, Chavali S, Pougach K, Legendre M, Zhu B, Boeynaems S, van der Zande E, Gevaert K, Rousseau F, Schymkowitz J *et al* (2015) Variable Glutamine-Rich Repeats Modulate Transcription Factor Activity. *Mol Cell* 59: 615-627

Gemmell NJ (2021) Repetitive DNA: genomic dark matter matters. *Nat Rev Genet* 22: 342

GeM-HD GMoHsDG-HC (2015) Identification of genetic factors that modify clinical onset of Huntington's disease. *Cell* 162.3: 516–26.

GeM-HD (2019) CAG Repeat Not Polyglutamine Length Determines Timing of Huntington's Disease Onset . *Cell* 178.4: 887-900.

Gervais FG, Singaraja R, Xanthoudakis S, Gutekunst CA, Leavitt BR, Metzler M, Hackam AS, Tam J, Vaillancourt JP, Houtzager V *et al* (2002) Recruitment and activation of caspase-8 by the Huntingtin-interacting protein Hip-1 and a novel partner Hippi. *Nat Cell Biol* 4: 95-105

Giannini M, Bayona-Feliu A, Sproviero D, Barroso SI, Cereda C, Aguilera A (2020) TDP-43 mutations link Amyotrophic Lateral Sclerosis with R-loop homeostasis and R loop-mediated DNA damage. *PLoS Genet* 16: e1009260

Ginno PA, Lim YW, Lott PL, Korf I, Chedin F (2013) GC skew at the 5' and 3' ends of human genes links R-loop formation to epigenetic regulation and transcription termination. *Genome Res* 23: 1590-1600

Ginno PA, Lott PL, Christensen HC, Korf I, Chedin F (2012) R-loop formation is a distinctive characteristic of unmethylated human CpG island promoters. *Mol Cell* 45: 814-825

Giordano AMS, Luciani M, Gatto F, Abou Alezz M, Beghe C, Della Volpe L, Migliara A, Valsoni S, Genua M, Dzieciatkowska M *et al* (2022) DNA damage contributes to neurotoxic inflammation in Aicardi-Goutieres syndrome astrocytes. *J Exp Med* 219

Girardi E, Messmer M, Lopez P, Fender A, Chicher J, Chane-Woon-Ming B, Hammann P, Pfeffer S (2023) Proteomics-based determination of double-stranded RNA interactome reveals known and new factors involved in Sindbis virus infection. *RNA* 29: 361-375

Godin JD, Colombo K, Molina-Calavita M, Keryer G, Zala D, Charrin BC, Dietrich P, Volvert ML, Guillemot F, Dragatsis I *et al* (2010) Huntingtin is required for mitotic spindle orientation and mammalian neurogenesis. *Neuron* 67: 392-406

Gonitel R, Moffitt H, Sathasivam K, Woodman B, Detloff PJ, Faull RL, Bates GP (2008) DNA instability in postmitotic neurons. *Proc Natl Acad Sci U S A* 105: 3467-3472

Goold R, Flower M, Moss DH, Medway C, Wood-Kaczmar A, Andre R, Farshim P, Bates GP, Holmans P, Jones L *et al* (2019) FAN1 modifies Huntington's disease progression by stabilizing the expanded HTT CAG repeat. *Hum Mol Genet* 28: 650-661

Goold R, Hamilton J, Menneteau T, Flower M, Bunting EL, Aldous SG, Porro A, Vicente JR, Allen ND, Wilkinson H *et al* (2021) FAN1 controls mismatch repair complex assembly via MLH1 retention to stabilize CAG repeat expansion in Huntington's disease. *Cell Rep* 36: 109649

Gorbunova V, Seluanov A, Mittelman D, Wilson JH (2004) Genome-wide demethylation destabilizes CTG.CAG trinucleotide repeats in mammalian cells. *Hum Mol Genet* 13: 2979-2989

Gorthi A, Romero JC, Loranc E, Cao L, Lawrence LA, Goodale E, Iniguez AB, Bernard X, Masamsetti VP, Roston S (2018) EWS–FLI1 increases transcription to cause R-loops and block BRCA1 repair in Ewing sarcoma. *Nature* 555: 387-391

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

Graf M, Bonetti D, Lockhart A, Serhal K, Kellner V, Maicher A, Jolivet P, Teixeira MT, Luke B (2017) Telomere Length Determines TERRA and R-Loop Regulation through the Cell Cycle. *Cell* 170: 72-85 e14

Graham RK, Deng Y, Slow EJ, Haigh B, Bissada N, Lu G, Pearson J, Shehadeh J, Bertram L, Murphy Z *et al* (2006) Cleavage at the caspase-6 site is required for neuronal dysfunction and degeneration due to mutant huntingtin. *Cell* 125: 1179-1191

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

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

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

Groh M, Silva LM, Gromak N (2014b) Mechanisms of transcriptional dysregulation in repeat expansion disorders. *Biochem Soc Trans* 42: 1123-1128

Grondin R, Kaytor MD, Ai Y, Nelson PT, Thakker DR, Heisel J, Weatherspoon MR, Blum JL, Burright EN, Zhang Z *et al* (2012) Six-month partial suppression of Huntingtin is well tolerated in the adult rhesus striatum. *Brain* 135: 1197-1209

Grunseich C, Wang IX, Watts JA, Burdick JT, Guber RD, Zhu Z, Bruzel A, Lanman T, Chen K, Schindler AB (2018) Senataxin mutation reveals how R-loops promote transcription by blocking DNA methylation at gene promoters. *Molecular cell* 69: 426-437. e427

Gu X, Greiner ER, Mishra R, Kodali R, Osmand A, Finkbeiner S, Steffan JS, Thompson LM, Wetzel R, Yang XW (2009) Serines 13 and 16 are critical determinants of full-length human mutant huntingtin induced disease pathogenesis in HD mice. *Neuron* 64: 828-840

Gudde A, van Heeringen SJ, de Oude AI, van Kessel IDG, Estabrook J, Wang ET, Wieringa B, Wansink DG (2017) Antisense transcription of the myotonic dystrophy locus yields low-abundant RNAs with and without (CAG)n repeat. *RNA Biol* 14: 1374-1388

Gunawardena S, Her LS, Brusch RG, Laymon RA, Niesman IR, Gordesky-Gold B, Sintasath L, Bonini NM, Goldstein LS (2003) Disruption of axonal transport by loss of huntingtin or expression of pathogenic polyQ proteins in Drosophila. *Neuron* 40: 25-40

Guo Q, Bin H, Cheng J, Seefelder M, Engler T, Pfeifer G, Oeckl P, Otto M, Moser F, Maurer M *et al* (2018) The cryo-electron microscopy structure of huntingtin. *Nature* 555: 117-120

Gusella JF, Wexler NS, Conneally PM, Naylor SL, Anderson MA, Tanzi RE, Watkins PC, Ottina K, Wallace MR, Sakaguchi AY *et al* (1983) A polymorphic DNA marker genetically linked to Huntington's disease. *Nature* 306: 234-238

Gyenis A, Chang J, Demmers JJPG *et al* (2023) Genome-wide RNA polymerase stalling shapes the transcriptome during aging. *Nat Genet* 55, 268–279

Gymrek M, Willems T, Guilmatre A, Zeng H, Markus B, Georgiev S, Daly MJ, Price AL, Pritchard JK, Sharp AJ *et al* (2016) Abundant contribution of short tandem repeats to gene expression variation in humans. *Nat Genet* 48: 22-29

Haeusler AR, Donnelly CJ, Periz G, Simko EA, Shaw PG, Kim MS, Maragakis NJ, Troncoso JC, Pandey A, Sattler R *et al* (2014) C9orf72 nucleotide repeat structures initiate molecular cascades of disease. *Nature* 507: 195-200

Hatchi E, Skourti-Stathaki K, Ventz S, Pinello L, Yen A, Kamieniarz-Gdula K, Dimitrov S, Pathania S, McKinney KM, Eaton ML *et al* (2015) BRCA1 recruitment to transcriptional pause sites is required for R-loop-driven DNA damage repair. *Mol Cell* 57: 636-647

Heintzman ND, Hon GC, Hawkins RD, Kheradpour P, Stark A, Harp LF, Ye Z, Lee LK, Stuart RK, Ching CW *et al* (2009) Histone modifications at human enhancers reflect global cell-type-specific gene expression. *Nature* 459: 108-112

Herman D, Jenssen K, Burnett R, Soragni E, Perlman SL, Gottesfeld JM (2006) Histone deacetylase inhibitors reverse gene silencing in Friedreich's ataxia. *Nat Chem Biol* 2: 551-558

Hill SJ, Clark AP, Silver DP, Livingston DM (2014) BRCA1 pathway function in basal-like breast cancer cells. *Molecular and cellular biology* 34: 3828-3842

Hodges A, Strand AD, Aragaki AK, Kuhn A, Sengstag T, Hughes G, Elliston LA, Hartog C, Goldstein DR, Thu D *et al* (2006) Regional and cellular gene expression changes in human Huntington's disease brain. *Hum Mol Genet* 15: 965-977

Hoffner G, Kahlem P, Djian P (2002) Perinuclear localization of huntingtin as a consequence of its binding to microtubules through an interaction with beta-tubulin: relevance to Huntington's disease. *J Cell Sci* 115: 941-948

Hoffner G, Soues S, Djian P (2007) Aggregation of expanded huntingtin in the brains of patients with Huntington disease. *Prion* 1: 26-31

Holzmann C, Schmidt T, Thiel G, Epplen JT, Riess O (2001) Functional characterization of the human Huntington's disease gene promoter. *Brain Res Mol Brain Res* 92: 85-97

Hoogeveen AT, Willemsen R, Meyer N, de Rooij KE, Roos RA, van Ommen GJ, Galjaard H (1993) Characterization and localization of the Huntington disease gene product. *Hum Mol Genet* 2: 2069-2073

Horvath P, Barrangou R (2010) CRISPR/Cas, the immune system of bacteria and archaea. *Science* 327: 167-170

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

Hung CL, Maiuri T, Bowie LE, Gotesman R, Son S, Falcone M, Giordano JV, Gillis T, Mattis V, Lau T *et al* (2018) A patient-derived cellular model for Huntington's disease reveals phenotypes at clinically relevant CAG lengths. *Mol Biol Cell* 29: 2809-2820

Hunter A, Tsilfidis C, Mettler G, Jacob P, Mahadevan M, Surh L, Korneluk R (1992) The correlation of age of onset with CTG trinucleotide repeat amplification in myotonic dystrophy. *J Med Genet* 29: 774-779

Hyeon JW, Kim AH, Yano H (2021) Epigenetic regulation in Huntington's disease. *Neurochem Int* 148: 105074

Irmak D, Fatima A, Gutierrez-Garcia R, Rinschen MM, Wagle P, Altmuller J, Arrigoni L, Hummel B, Klein C, Frese CK *et al* (2018) Mechanism suppressing H3K9 trimethylation in pluripotent stem cells and its demise by polyQ-expanded huntingtin mutations. *Hum Mol Genet* 27: 4117-4134

Islam J, So KH, Kc E, Moon HC, Kim A, Hyun SH, Kim S, Park YS (2021) Transplantation of human embryonic stem cells alleviates motor dysfunction in AAV2-Htt171-82Q transfected rat model of Huntington's disease. *Stem Cell Res Ther* 12: 585

Iyer RR, Pluciennik A (2021) DNA Mismatch Repair and its Role in Huntington's Disease. *J Huntingtons Dis* 10: 75-94

Ji J, Sundquist K, Sundquist J (2012) Cancer incidence in patients with polyglutamine diseases: a population-based study in Sweden. *Lancet Oncol* 13: 642-648

Jiang H, Mankodi A, Swanson MS, Moxley RT, Thornton CA (2004) Myotonic dystrophy type 1 is associated with nuclear foci of mutant RNA, sequestration of muscleblind proteins and deregulated alternative splicing in neurons. *Hum Mol Genet* 13: 3079-3088

Jochum T, Ritz ME, Schuster C, Funderburk SF, Jehle K, Schmitz K, Brinkmann F, Hirtz M, Moss D, Cato AC (2012) Toxic and non-toxic aggregates from the SBMA and normal forms of androgen receptor have distinct oligomeric structures. *Biochim Biophys Acta* 1822: 1070-1078

Jonkers I, Lis JT (2015) Getting up to speed with transcription elongation by RNA polymerase II. *Nat Rev Mol Cell Biol* 16: 167-177

Kaback DB, Angerer LM, Davidson N (1979) Improved methods for the formation and stabilization of R-loops. *Nucleic acids research* 6: 2499-2517

Kacher R, Lejeune FX, Noel S, Cazeneuve C, Brice A, Humbert S, Durr A (2021) Propensity for somatic expansion increases over the course of life in Huntington disease. *Elife* 10

Katyal S, Lee Y, Nitiss KC, Downing SM, Li Y, Shimada M, Zhao J, Russell HR, Petrini JH, Nitiss JL (2014) Aberrant topoisomerase-1 DNA lesions are pathogenic in neurodegenerative genome instability syndromes. *Nature neuroscience* 17: 813-821

Kauffman JS, Zinovyeva A, Yagi K, Makabe KW, Raff RA (2003) Neural expression of the Huntington's disease gene as a chordate evolutionary novelty. *J Exp Zool B Mol Dev Evol* 297: 57-64

Kaye J, Reisine T, Finkbeiner S (2022) Huntington's disease iPSC models-using human patient cells to understand the pathology caused by expanded CAG repeats. *Fac Rev* 11: 16

Kearse MG, Todd PK (2014) Repeat-associated non-AUG translation and its impact in neurodegenerative disease. *Neurotherapeutics* 11: 721-731

Kegel KB, Kim M, Sapp E, McIntyre C, Castano JG, Aronin N, DiFiglia M (2000) Huntingtin expression stimulates endosomal-lysosomal activity, endosome tubulation, and autophagy. *J Neurosci* 20: 7268-7278

Kegel KB, Meloni AR, Yi Y, Kim YJ, Doyle E, Cuiffo BG, Sapp E, Wang Y, Qin ZH, Chen JD *et al* (2002) Huntingtin is present in the nucleus, interacts with the transcriptional corepressor C-terminal binding protein, and represses transcription. *J Biol Chem* 277: 7466-7476

Kennedy L, Evans E, Chen CM, Craven L, Detloff PJ, Ennis M, Shelbourne PF (2003) Dramatic tissue-specific mutation length increases are an early molecular event in Huntington disease pathogenesis. *Hum Mol Genet* 12: 3359-3367

Khoshnan A, Ko J, Watkin EE, Paige LA, Reinhart PH, Patterson PH (2004) Activation of the IkappaB kinase complex and nuclear factor-kappaB contributes to mutant huntingtin neurotoxicity. *J Neurosci* 24: 7999-8008

Khristich AN, Mirkin SM (2020) On the wrong DNA track: Molecular mechanisms of repeat-mediated genome instability. *J Biol Chem* 295: 4134-4170

Kim MW, Chelliah Y, Kim SW, Otwinowski Z, Bezprozvanny I (2009) Secondary structure of Huntingtin amino-terminal region. *Structure* 17: 1205-1212

Kotsantis P, Silva LM, Irmscher S, Jones RM, Folkes L, Gromak N, Petermann E (2016) Increased global transcription activity as a mechanism of replication stress in cancer. *Nat Commun* 7: 13087

Kouzine F, Wojtowicz D, Baranello L, Yamane A, Nelson S, Resch W, Kieffer-Kwon KR, Benham CJ, Casellas R, Przytycka TM *et al* (2017) Permanganate/S1 Nuclease Footprinting Reveals Non-B DNA Structures with Regulatory Potential across a Mammalian Genome. *Cell Syst* 4: 344-356 e347

Krach F, Stemick J, Boerstler T, Weiss A, Lingos I, Reischl S, Meixner H, Ploetz S, Farrell M, Hehr U *et al* (2022) An alternative splicing modulator decreases mutant HTT and improves the molecular fingerprint in Huntington's disease patient neurons. *Nat Commun* 13: 6797

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

Krzyzosiak WJ, Sobczak K, Wojciechowska M, Fiszer A, Mykowska A, Kozlowski P (2012) Triplet repeat RNA structure and its role as pathogenic agent and therapeutic target. *Nucleic acids research* 40: 11-26

La Spada AR, Weydt P, Pineda VV (2011) Huntington's Disease Pathogenesis: Mechanisms and Pathways. In: *Neurobiology of Huntington's Disease: Applications to Drug Discovery*, Lo D.C., Hughes R.E. (eds.)Boca Raton (FL)

Lai JI, Nachun D, Petrosyan L, Throesch B, Campau E, Gao F, Baldwin KK, Coppola G, Gottesfeld JM, Soragni E (2019) Transcriptional profiling of isogenic Friedreich ataxia neurons and effect of an HDAC inhibitor on disease signatures. *J Biol Chem* 294: 1846-1859

Lam SS, Martell JD, Kamer KJ, Deerinck TJ, Ellisman MH, Mootha VK, Ting AY (2015) Directed evolution of APEX2 for electron microscopy and proximity labeling. *Nat Methods* 12: 51-54

Langfelder P, Cantle JP, Chatzopoulou D, Wang N, Gao F, Al-Ramahi I, Lu XH, Ramos EM, El-Zein K, Zhao Y *et al* (2016) Integrated genomics and proteomics define huntingtin CAG length-dependent networks in mice. *Nat Neurosci* 19: 623-633

Law EK, Levin-Klein R, Jarvis MC, Kim H, Argyris PP, Carpenter MA, Starrett GJ, Temiz NA, Larson LK, Durfee C (2020) APOBEC3A catalyzes mutation and drives carcinogenesis in vivo. *Journal of Experimental Medicine* 217

Lawlor KT, O'Keefe LV, Samaraweera SE, van Eyk CL, McLeod CJ, Maloney CA, Dang TH, Suter CM, Richards RI (2011) Double-stranded RNA is pathogenic in Drosophila models of expanded repeat neurodegenerative diseases. *Hum Mol Genet* 20: 3757-3768

Leavitt BR, Guttman JA, Hodgson JG, Kimel GH, Singaraja R, Vogl AW, Hayden MR (2001) Wild-type huntingtin reduces the cellular toxicity of mutant huntingtin in vivo. *Am J Hum Genet* 68: 313-324

Leavitt BR, van Raamsdonk JM, Shehadeh J, Fernandes H, Murphy Z, Graham RK, Wellington CL, Raymond LA, Hayden MR (2006) Wild-type huntingtin protects neurons from excitotoxicity. *J Neurochem* 96: 1121-1129

Lee H, Fenster RJ, Pineda SS, Gibbs WS, Mohammadi S, Davila-Velderrain J, Garcia FJ, Therrien M, Novis HS, Gao F *et al* (2020) Cell Type-Specific Transcriptomics Reveals that Mutant Huntingtin Leads to Mitochondrial RNA Release and Neuronal Innate Immune Activation. *Neuron* 107: 891-908 e898

Lee J, Hwang YJ, Boo JH, Han D, Kwon OK, Todorova K, Kowall NW, Kim Y, Ryu H (2011) Dysregulation of upstream binding factor-1 acetylation at K352 is linked to impaired ribosomal DNA transcription in Huntington's disease. *Cell Death Differ* 18: 1726-1735

Lee J, Hwang YJ, Shin JY, Lee WC, Wie J, Kim KY, Lee MY, Hwang D, Ratan RR, Pae AN *et al* (2013) Epigenetic regulation of cholinergic receptor M1 (CHRM1) by histone H3K9me3 impairs Ca(2+) signaling in Huntington's disease. *Acta Neuropathol* 125: 727-739

Lee T, Pelletier J (2016) The biology of DHX9 and its potential as a therapeutic target. *Oncotarget* 7: 42716-42739

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, Genome Project Data Processing S (2009) The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25: 2078-2079

Li SH, Cheng AL, Zhou H, Lam S, Rao M, Li H, Li XJ (2002) Interaction of Huntington disease protein with transcriptional activator Sp1. *Mol Cell Biol* 22: 1277-1287

Li SH, Schilling G, Young WS, 3rd, Li XJ, Margolis RL, Stine OC, Wagster MV, Abbott MH, Franz ML, Ranen NG *et al* (1993) Huntington's disease gene (IT15) is widely expressed in human and rat tissues. *Neuron* 11: 985-993

Li X, Manley JL (2005) Inactivation of the SR protein splicing factor ASF/SF2 results in genomic instability. *Cell* 122: 365-378

Li Y, Li J, Wang J, Zhang S, Giles K, Prakash TP, Rigo F, Napierala JS, Napierala M (2022) Premature transcription termination at the expanded GAA repeats and aberrant alternative polyadenylation contributes to the Frataxin transcriptional deficit in Friedreich's ataxia. *Hum Mol Genet* 31: 3539-3557

Lieberman AP, Shakkottai VG, Albin RL (2019) Polyglutamine repeats in neurodegenerative diseases. *Annual Review of Pathology: Mechanisms of Disease* 14: 1-27

Lin B, Nasir J, Kalchman MA, McDonald H, Zeisler J, Goldberg YP, Hayden MR (1995) Structural analysis of the 5' region of mouse and human Huntington disease genes reveals conservation of putative promoter region and di- and trinucleotide polymorphisms. *Genomics* 25: 707-715

Lin B, Rommens JM, Graham RK, Kalchman M, MacDonald H, Nasir J, Delaney A, Goldberg YP, Hayden MR (1993) Differential 3' polyadenylation of the Huntington disease gene results in two mRNA species with variable tissue expression. *Hum Mol Genet* 2: 1541-1545

Lin Y, Dent SY, Wilson JH, Wells RD, Napierala M (2010) R loops stimulate genetic instability of CTG.CAG repeats. *Proc Natl Acad Sci U S A* 107: 692-697

Liu CR, Chang CR, Chern Y, Wang TH, Hsieh WC, Shen WC, Chang CY, Chu IC, Deng N, Cohen SN *et al* (2012) Spt4 is selectively required for transcription of extended trinucleotide repeats. *Cell* 148: 690-701

Liu EY, Cali CP, Lee EB (2017) RNA metabolism in neurodegenerative disease. *Dis Model Mech* 10: 509-518

Loomis EW, Sanz LA, Chedin F, Hagerman PJ (2014) Transcription-associated R-loop formation across the human FMR1 CGG-repeat region. *PLoS Genet* 10: e1004294

Lopez Castel A, Nakamori M, Tome S, Chitayat D, Gourdon G, Thornton CA, Pearson CE (2011) Expanded CTG repeat demarcates a boundary for abnormal CpG methylation in myotonic dystrophy patient tissues. *Hum Mol Genet* 20: 1-15

Lu AT, Narayan P, Grant MJ, Langfelder P, Wang N, Kwak S, Wilkinson H, Chen RZ, Chen J, Simon Bawden C *et al* (2020) DNA methylation study of Huntington's disease and motor progression in patients and in animal models. *Nat Commun* 11: 4529

Lu XH, Mattis VB, Wang N, Al-Ramahi I, van den Berg N, Fratantoni SA, Waldvogel H, Greiner E, Osmand A, Elzein K *et al* (2014) Targeting ATM ameliorates mutant Huntingtin toxicity in cell and animal models of Huntington's disease. *Sci Transl Med* 6: 268ra178

Luthi-Carter R, Strand A, Peters NL, Solano SM, Hollingsworth ZR, Menon AS, Frey AS, Spektor BS, Penney EB, Schilling G *et al* (2000) Decreased expression of striatal signaling genes in a mouse model of Huntington's disease. *Hum Mol Genet* 9: 1259-1271

Ly S, Didiot MC, Ferguson CM, Coles AH, Miller R, Chase K, Echeverria D, Wang F, Sadri-Vakili G, Aronin N *et al* (2022) Mutant huntingtin messenger RNA forms neuronal nuclear clusters in rodent and human brains. *Brain Commun* 4: fcac248

Ma L, Herren AW, Espinal G, Randol J, McLaughlin B, Martinez-Cerdeno V, Pessah IN, Hagerman RJ, Hagerman PJ (2019) Composition of the Intranuclear Inclusions of Fragile X-associated Tremor/Ataxia Syndrome. *Acta Neuropathol Commun* 7: 143

MacDonald ME, Ambrose CM, Duyao MP, Myers RH, Lin C, Srinidhi L, Barnes G, Taylor SA, James M, Groot N, MacFarlane H (1993) A novel gene containing a trinucleotide repeat that is expanded and unstable on Huntington's disease chromosomes. *Cell* 72: 971-983

Mackenzie KJ, Carroll P, Lettice L, Tarnauskaitė Ž, Reddy K, Dix F, Revuelta A, Abbondati E, Rigby RE, Rabe B (2016) Ribonuclease H2 mutations induce a cGAS/STING-dependent innate immune response. *The EMBO journal* 35: 831-844

Maiuri T, Mocle AJ, Hung CL, Xia J, van Roon-Mom WM, Truant R (2017) Huntingtin is a scaffolding protein in the ATM oxidative DNA damage response complex. *Hum Mol Genet* 26: 395-406

Malankhanova TB, Malakhova AA, Medvedev SP, Zakian SM (2017) Modern Genome Editing Technologies in Huntington's Disease Research. *J Huntingtons Dis* 6: 19-31

Malig M, Hartono SR, Giafaglione JM, Sanz LA, Chedin F (2020) Ultra-deep Coverage Single-molecule R-loop Footprinting Reveals Principles of R-loop Formation. *J Mol Biol* 432: 2271-2288

Malik I, Kelley CP, Wang ET, Todd PK (2021) Molecular mechanisms underlying nucleotide repeat expansion disorders. *Nat Rev Mol Cell Biol* 22: 589-607

Marasco LE, Kornblihtt AR (2022) The physiology of alternative splicing. *Nat Rev Mol Cell Biol*

Marinello J, Bertoncini S, Aloisi I, Cristini A, Malagoli Tagliazucchi G, Forcato M, Sordet O, Capranico G (2016) Dynamic Effects of Topoisomerase I Inhibition on R-Loops and Short Transcripts at Active Promoters. *PLoS One* 11: e0147053

Martin DD, Heit RJ, Yap MC, Davidson MW, Hayden MR, Berthiaume LG (2014) Identification of a post-translationally myristoylated autophagy-inducing domain released by caspase cleavage of huntingtin. *Hum Mol Genet* 23: 3166-3179

Martin M (2011) Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet journal* 17: 10-12

Mason MA, Gomez-Paredes C, Sathasivam K, Neueder A, Papadopoulou AS, Bates GP (2020) Silencing Srsf6 does not modulate incomplete splicing of the huntingtin gene in Huntington's disease models. *Sci Rep* 10: 14057

Massey TH, Jones L (2018) The central role of DNA damage and repair in CAG repeat diseases. *Dis Model Mech* 11

McAllister B, Donaldson J, Binda CS, Powell S, Chughtai U, Edwards G, Stone J, Lobanov S, Elliston L, Schuhmacher LN *et al* (2022) Exome sequencing of individuals with Huntington's disease implicates FAN1 nuclease activity in slowing CAG expansion and disease onset. *Nat Neurosci* 25: 446-457

McCann JL, Cristini A, Law EK, Lee SY, Tellier M, Carpenter MA, Beghè C, Kim JJ, Jarvis MC, Stefanovska B *et al* (2021) R-loop homeostasis and cancer mutagenesis promoted by the DNA cytosine deaminase APOBEC3B. *bioRxiv*: 2021.2008.2030.458235

McLauchlan DJ, Lancaster T, Craufurd D, Linden DEJ, Rosser AE (2022) Different depression: motivational anhedonia governs antidepressant efficacy in Huntington's disease. *Brain Commun* 4: fcac278

Menalled LB, Sison JD, Dragatsis I, Zeitlin S, Chesselet MF (2003) Time course of early motor and neuropathological anomalies in a knock-in mouse model of Huntington's disease with 140 CAG repeats. *J Comp Neurol* 465: 11-26

Michel L, Huguet-Lachon A, Gourdon G (2015) Sense and Antisense DMPK RNA Foci Accumulate in DM1 Tissues during Development. *PLoS One* 10: e0137620

Mikaeili H, Sandi M, Bayot A, Al-Mahdawi S, Pook MA (2018) FAST-1 antisense RNA epigenetically alters FXN expression. *Sci Rep* 8: 17217

Miller BR, Walker AG, Barton SJ, Rebec GV (2011) Dysregulated Neuronal Activity Patterns Implicate Corticostriatal Circuit Dysfunction in Multiple Rodent Models of Huntington's Disease. *Front Syst Neurosci* 5: 26

Miller JL, Grant PA (2013) The role of DNA methylation and histone modifications in transcriptional regulation in humans. *Subcell Biochem* 61: 289-317

Molero AE, Gokhan S, Gonzalez S, Feig JL, Alexandre LC, Mehler MF (2009) Impairment of developmental stem cell-mediated striatal neurogenesis and pluripotency genes in a knock-in model of Huntington's disease. *Proc Natl Acad Sci U S A* 106: 21900-21905

Morales JC, Richard P, Patidar PL, Motea EA, Dang TT, Manley JL, Boothman DA (2016) XRN2 links transcription termination to DNA damage and replication stress. *PLoS genetics* 12: e1006107

Mosler T, Conte F, Longo GMC, Mikicic I, Kreim N, Mockel MM, Petrosino G, Flach J, Barau J, Luke B *et al* (2021) R-loop proximity proteomics identifies a role of DDX41 in transcription-associated genomic instability. *Nat Commun* 12: 7314

Moss DJH, Pardinas AF, Langbehn D, Lo K, Leavitt BR, Roos R, Durr A, Mead S, investigators T-H, investigators R *et al* (2017) Identification of genetic variants associated with Huntington's disease progression: a genome-wide association study. *Lancet Neurol* 16: 701-711

Mouro Pinto R, Arning L, Giordano JV, Razghandi P, Andrew MA, Gillis T, Correia K, Mysore JS, Grote Urtubey DM, Parwez CR *et al* (2020) Patterns of CAG repeat instability in the central nervous system and periphery in Huntington's disease and in spinocerebellar ataxia type 1. *Hum Mol Genet* 29: 2551-2567

Murmann AE, Gao QQ, Putzbach WE, Patel M, Bartom ET, Law CY, Bridgeman B, Chen S, McMahon KM, Thaxton CS *et al* (2018) Small interfering RNAs based on huntingtin trinucleotide repeats are highly toxic to cancer cells. *EMBO Rep* 19

Myers RH, Sax DS, Schoenfeld M, Bird ED, Wolf PA, Vonsattel JP, White RF, Martin JB (1985) Late onset of Huntington's disease. *J Neurol Neurosurg Psychiatry* 48: 530-534

Nageshwaran S, Festenstein R (2015) Epigenetics and Triplet-Repeat Neurological Diseases. *Front Neurol* 6: 262

Nakamori M, Panigrahi GB, Lanni S, Gall-Duncan T, Hayakawa H, Tanaka H, Luo J, Otabe T, Li J, Sakata A *et al* (2020) A slipped-CAG DNA-binding small molecule induces trinucleotide-repeat contractions in vivo. *Nat Genet* 52: 146-159

Napierala M, Krzyzosiak WJ (1997) CUG repeats present in myotonin kinase RNA form metastable "slippery" hairpins. *J Biol Chem* 272: 31079-31085

Nasir J, Floresco SB, O'Kusky JR, Diewert VM, Richman JM, Zeisler J, Borowski A, Marth JD, Phillips AG,

Hayden MR (1995) Targeted disruption of the Huntington's disease gene results in embryonic lethality and behavioral and morphological changes in heterozygotes. *Cell* 81: 811-823

Neueder A, Dumas AA, Benjamin AC, Bates GP (2018) Regulatory mechanisms of incomplete huntingtin mRNA splicing. *Nat Commun* 9: 3955

Neueder A, Landles C, Ghosh R, Howland D, Myers RH, Faull RLM, Tabrizi SJ, Bates GP (2017) The pathogenic exon 1 HTT protein is produced by incomplete splicing in Huntington's disease patients. *Sci Rep* 7: 1307

Ng CW, Yildirim F, Yap YS, Dalin S, Matthews BJ, Velez PJ, Labadorf A, Housman DE, Fraenkel E (2013) Extensive changes in DNA methylation are associated with expression of mutant huntingtin. *Proc Natl Acad Sci U S A* 110: 2354-2359

Niehrs C, Luke B (2020) Regulatory R-loops as facilitators of gene expression and genome stability. *Nat Rev Mol Cell Biol* 21: 167-178

Nissanka N, Bacman SR, Plastini MJ, Moraes CT (2018) The mitochondrial DNA polymerase gamma degrades linear DNA fragments precluding the formation of deletions. *Nat Commun* 9: 2491

Novak MJ, Tabrizi SJ (2010) Huntington's disease. *BMJ* 340: c3109

Nucifora FC, Jr., Sasaki M, Peters MF, Huang H, Cooper JK, Yamada M, Takahashi H, Tsuji S, Troncoso J, Dawson VL *et al* (2001) Interference by huntingtin and atrophin-1 with cbp-mediated transcription leading to cellular toxicity. *Science* 291: 2423-2428

O'Kusky JR, Nasir J, Cicchetti F, Parent A, Hayden MR (1999) Neuronal degeneration in the basal ganglia and loss of pallido-subthalamic synapses in mice with targeted disruption of the Huntington's disease gene. *Brain Res* 818: 468-479

Ordway JM., Tallaksen-Greene S, Gutekunst CA, Bernstein EM., Cearley JA., Wiener HW, Dure LS, Lindsey R., Hersch SM., Jope RS., Albin RL, Detloff PJ (1997) Ectopically expressed CAG repeats cause intranuclear inclusions and a progressive late onset neurological phenotype in the mouse. *Cell, 91(6), 753–763*.

O'Regan GC, Farag SH, Casey CS, Wood-Kaczmar A, Pocock JM, Tabrizi SJ, Andre R (2021) Human Huntington's disease pluripotent stem cell-derived microglia develop normally but are abnormally hyper-reactive and release elevated levels of reactive oxygen species. *J Neuroinflammation* 18: 94

Ostergaard ME, Southwell AL, Kordasiewicz H, Watt AT, Skotte NH, Doty CN, Vaid K, Villanueva EB, Swayze EE, Bennett CF *et al* (2013) Rational design of antisense oligonucleotides targeting single nucleotide polymorphisms for potent and allele selective suppression of mutant Huntingtin in the CNS. *Nucleic Acids Res* 41: 9634-9650

Pacesa M, Loeff L, Querques I, Muckenfuss LM, Sawicka M, Jinek M (2022) R-loop formation and conformational activation mechanisms of Cas9. *Nature* 609: 191-196

Palminha NM, Dos Santos Souza C, Griffin J, Liao C, Ferraiuolo L, El-Khamisy SF (2022) Defective repair of topoisomerase I induced chromosomal damage in Huntington's disease. *Cell Mol Life Sci* 79: 160

Pandolfo M, Pastore A (2009) The pathogenesis of Friedreich ataxia and the structure and function of frataxin. *J Neurol* 256 Suppl 1: 9-17

Panigrahi GB, Slean MM, Simard JP, Gileadi O, Pearson CE (2010) Isolated short CTG/CAG DNA slip-outs are repaired efficiently by hMutSbeta, but clustered slip-outs are poorly repaired. *Proc Natl Acad Sci U S A* 107: 12593-12598

Panov AV, Gutekunst CA, Leavitt BR, Hayden MR, Burke JR, Strittmatter WJ, Greenamyre JT (2002) Early mitochondrial calcium defects in Huntington's disease are a direct effect of polyglutamines. *Nat Neurosci* 5: 731-736

Park HJ, Han A, Kim JY, Choi J, Bae HS, Cho GB, Shin H, Shin EJ, Lee KI, Kim S *et al* (2022) SUPT4H1-edited stem cell therapy rescues neuronal dysfunction in a mouse model for Huntington's disease. *NPJ Regen Med* 7: 8

Park K, Ryoo J, Jeong H, Kim M, Lee S, Hwang S-Y, Ahn J, Kim D, Moon HC, Baek D (2021) Aicardi-Goutières syndrome-associated gene SAMHD1 preserves genome integrity by preventing R-loop formation at transcription–replication conflict regions. *PLoS genetics* 17: e1009523

Passani LA, Bedford MT, Faber PW, McGinnis KM, Sharp AH, Gusella JF, Vonsattel JP, MacDonald ME (2000) Huntingtin's WW domain partners in Huntington's disease post-mortem brain fulfill genetic criteria for direct involvement in Huntington's disease pathogenesis. *Hum Mol Genet* 9: 2175-2182

Paulson H (2018) Repeat expansion diseases. *Handb Clin Neurol* 147: 105-123

Pearl JR, Shetty AC, Cantle JP, Bergey DE, Bragg RM, Coffey SR, Kordasiewicz HB, Hood LE, Price ND,

Ament SA *et al* (2020) Altered Huntingtin-Chromatin Interactions Predict Transcriptional and Epigenetic Changes in Huntington's Disease. *bioRxiv*: 2020.2006.2004.132571

Peel AL, Rao RV, Cottrell BA, Hayden MR, Ellerby LM, Bredesen DE (2001) Double-stranded RNA-dependent protein kinase, PKR, binds preferentially to Huntington's disease (HD) transcripts and is activated in HD tissue. *Hum Mol Genet* 10: 1531-1538

Perry RP, Kelley DE (1970) Inhibition of RNA synthesis by actinomycin D: characteristic dose-response of different RNA species. *J Cell Physiol* 76: 127-139

Perutz MF, Johnson T, Suzuki M, Finch JT (1994) Glutamine repeats as polar zippers: their possible role in inherited neurodegenerative diseases. *Proc Natl Acad Sci U S A* 91: 5355-5358

Pfister EL, Kennington L, Straubhaar J, Wagh S, Liu W, DiFiglia M, Landwehrmeyer B, Vonsattel JP, Zamore PD, Aronin N (2009) Five siRNAs targeting three SNPs may provide therapy for three-quarters of Huntington's disease patients. *Curr Biol* 19: 774-778

Phillips DD, Garboczi DN, Singh K, Hu Z, Leppla SH, Leysath CE (2013) The sub-nanomolar binding of DNA-RNA hybrids by the single-chain Fv fragment of antibody S9.6. *J Mol Recognit* 26: 376-381

Pickrell AM, Fukui H, Wang X, Pinto M, Moraes CT (2011) The striatum is highly susceptible to mitochondrial oxidative phosphorylation dysfunctions. *J Neurosci* 31: 9895-9904

Pinto RM, Dragileva E, Kirby A, Lloret A, Lopez E, St Claire J, Panigrahi GB, Hou C, Holloway K, Gillis T *et al* (2013) Mismatch repair genes Mlh1 and Mlh3 modify CAG instability in Huntington's disease mice: genome-wide and candidate approaches. *PLoS Genet* 9: e1003930

Polak U, Li Y, Butler JS, Napierala M (2016) Alleviating GAA Repeat Induced Transcriptional Silencing of the Friedreich's Ataxia Gene During Somatic Cell Reprogramming. *Stem Cells Dev* 25: 1788-1800

Porro A, Mohiuddin M, Zurfluh C, Spegg V, Dai J, Iehl F, Ropars V, Collotta G, Fishwick KM, Mozaffari NL *et al* (2021) FAN1-MLH1 interaction affects repair of DNA interstrand cross-links and slipped-CAG/CTG repeats. *Sci Adv* 7

Powell WT, Coulson RL, Gonzales ML, Crary FK, Wong SS, Adams S, Ach RA, Tsang P, Yamada NA, Yasui DH (2013) R-loop formation at Snord116 mediates topotecan inhibition of Ube3a-antisense and allele-specific chromatin decondensation. *Proceedings of the National Academy of Sciences* 110: 13938-13943

Pradhan S, Gao R, Bush K, Zhang N, Wairkar YP, Sarkar PS (2022) Polyglutamine Expansion in Huntingtin and Mechanism of DNA Damage Repair Defects in Huntington's Disease. *Front Cell Neurosci* 16: 837576

Promonet A, Padioleau I, Liu Y, Sanz L, Biernacka A, Schmitz AL, Skrzypczak M, Sarrazin A, Mettling C, Rowicka M *et al* (2020) Topoisomerase 1 prevents replication stress at R-loop-enriched transcription termination sites. *Nat Commun* 11: 3940

Quinlan AR, Hall IM (2010) BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* 26: 841-842

Ramirez F, Ryan DP, Gruning B, Bhardwaj V, Kilpert F, Richter AS, Heyne S, Dundar F, Manke T (2016) deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Res* 44: W160-165

Ravikumar B, Vacher C, Berger Z, Davies JE, Luo S, Oroz LG, Scaravilli F, Easton DF, Duden R, O'Kane CJ *et al* (2004) Inhibition of mTOR induces autophagy and reduces toxicity of polyglutamine expansions in fly and mouse models of Huntington disease. *Nat Genet* 36: 585-595

Ribeiro de Almeida C, Dhir S, Dhir A, Moghaddam AE, Sattentau Q, Meinhart A, Proudfoot NJ (2018) RNA Helicase DDX1 Converts RNA G-Quadruplex Structures into R-Loops to Promote IgH Class Switch Recombination. *Mol Cell* 70(4):650-662.e8

Richard P, Manley JL (2017) R loops and links to human disease. *Journal of molecular biology* 429: 3168-3180

Rigamonti D, Bauer JH, De-Fraja C, Conti L, Sipione S, Sciorati C, Clementi E, Hackam A, Hayden MR, Li Y *et al* (2000) Wild-type huntingtin protects from apoptosis upstream of caspase-3. *J Neurosci* 20: 3705-3713

Rigamonti D, Sipione S, Goffredo D, Zuccato C, Fossale E, Cattaneo E (2001) Huntingtin's neuroprotective activity occurs via inhibition of procaspase-9 processing. *J Biol Chem* 276: 14545-14548

Rocha NP, Charron O, Latham LB, Colpo GD, Zanotti-Fregonara P, Yu M, Freeman L, Furr Stimming E, Teixeira AL (2021) Microglia Activation in Basal Ganglia Is a Late Event in Huntington Disease Pathophysiology. *Neurol Neuroimmunol Neuroinflamm* 8

Rocha NP, Ribeiro FM, Furr-Stimming E, Teixeira AL (2016) Neuroimmunology of Huntington's Disease: Revisiting Evidence from Human Studies. *Mediators Inflamm* 2016: 8653132

Romo L, Ashar-Patel A, Pfister E, Aronin N (2017) Alterations in mRNA 3' UTR Isoform Abundance Accompany Gene Expression Changes in Human Huntington's Disease Brains. *Cell Rep* 20: 3057-3070

Romo L, Mohn ES, Aronin N (2018) A Fresh Look at Huntingtin mRNA Processing in Huntington's Disease. *J Huntingtons Dis* 7: 101-108

Rui YN, Xu Z, Patel B, Chen Z, Chen D, Tito A, David G, Sun Y, Stimming EF, Bellen HJ *et al* (2015) Huntingtin functions as a scaffold for selective macroautophagy. *Nat Cell Biol* 17: 262-275

Ryu H, Lee J, Hagerty SW, Soh BY, McAlpin SE, Cormier KA, Smith KM, Ferrante RJ (2006) ESET/SETDB1 gene expression and histone H3 (K9) trimethylation in Huntington's disease. *Proc Natl Acad Sci U S A* 103: 19176-19181

Sadri-Vakili G, Bouzou B, Benn CL, Kim MO, Chawla P, Overland RP, Glajch KE, Xia E, Qiu Z, Hersch SM *et al* (2007) Histones associated with downregulated genes are hypo-acetylated in Huntington's disease models. *Hum Mol Genet* 16: 1293-1306

Salinas-Rios V, Belotserkovskii BP, Hanawalt PC (2011) DNA slip-outs cause RNA polymerase II arrest in vitro: potential implications for genetic instability. *Nucleic Acids Res* 39: 7444-7454

San Martin Alonso M, Noordermeer SM (2021) Untangling the crosstalk between BRCA1 and R-loops during DNA repair. *Nucleic Acids Research* 49: 4848-4863

Santoro MR, Bray SM, Warren ST (2012) Molecular mechanisms of fragile X syndrome: a twenty-year perspective. *Annu Rev Pathol* 7: 219-245

Santos-Pereira JM, Aguilera A (2015) R loops: new modulators of genome dynamics and function. *Nat Rev Genet* 16: 583-597

Santos-Pereira JM, Herrero AB, Garcia-Rubio ML, Marin A, Moreno S, Aguilera A (2013) The Npl3 hnRNP prevents R-loop-mediated transcription-replication conflicts and genome instability. *Genes Dev* 27: 2445-2458

Sanz LA, Chédin F (2019) High-resolution, strand-specific R-loop mapping via S9. 6-based DNA–RNA immunoprecipitation and high-throughput sequencing. *Nature protocols* 14: 1734-1755

Saponaro M (2022) Transcription-Replication Coordination. *Life (Basel)* 12

Sapp E, Kegel KB, Aronin N, Hashikawa T, Uchiyama Y, Tohyama K, Bhide PG, Vonsattel JP, DiFiglia M (2001) Early and progressive accumulation of reactive microglia in the Huntington disease brain. *J Neuropathol Exp Neurol* 60: 161-172

Sathasivam K, Neueder A, Gipson TA, Landles C, Benjamin AC, Bondulich MK, Smith DL, Faull RL, Roos RA, Howland D *et al* (2013) Aberrant splicing of HTT generates the pathogenic exon 1 protein in Huntington disease. *Proc Natl Acad Sci U S A* 110: 2366-2370

Saudou F, Finkbeiner S, Devys D, Greenberg ME (1998) Huntingtin acts in the nucleus to induce apoptosis but death does not correlate with the formation of intranuclear inclusions. *Cell* 95: 55-66

Saudou F, Humbert S (2008) The biology of Huntington's disease. *Handb Clin Neurol* 89: 619-629

Saudou F, Humbert S (2016) The Biology of Huntingtin. *Neuron* 89: 910-926

Savage KI, Matchett KB, Barros EM, Cooper KM, Irwin GW, Gorski JJ, Orr KS, Vohhodina J, Kavanagh JN, Madden AF (2014) BRCA1 Deficiency Exacerbates Estrogen-Induced DNA Damage and Genomic InstabilityEstrogen-Induced Genomic Instability in BRCA1-Deficient Cells. *Cancer research* 74: 2773-2784

Sawa A, Wiegand GW, Cooper J, Margolis RL, Sharp AH, Lawler JF, Jr., Greenamyre JT, Snyder SH, Ross CA (1999) Increased apoptosis of Huntington disease lymphoblasts associated with repeat length-dependent mitochondrial depolarization. *Nat Med* 5: 1194-1198

Schilling G, Becher MW, Sharp AH, Jinnah HA, Duan K, Kotzuk JA, Slunt HH, Ratovitski T, Cooper JK, Jenkins NA *et al* (1999) Intranuclear inclusions and neuritic aggregates in transgenic mice expressing a mutant N-terminal fragment of huntingtin. *Hum Mol Genet* 8: 397-407

Schilling J, Broemer M, Atanassov I, Duernberger Y, Vorberg I, Dieterich C, Dagane A, Dittmar G, Wanker E, van Roon-Mom W *et al* (2019) Deregulated Splicing Is a Major Mechanism of RNA-Induced Toxicity in Huntington's Disease. *J Mol Biol* 431: 1869-1877

Schonborn J, Oberstrass J, Breyel E, Tittgen J, Schumacher J, Lukacs N (1991) Monoclonal antibodies to double-stranded RNA as probes of RNA structure in crude nucleic acid extracts. *Nucleic Acids Res* 19: 2993-3000

Schulte J, Littleton JT (2011) The biological function of the Huntingtin protein and its relevance to Huntington's Disease pathology. *Curr Trends Neurol* 5: 65-78

Schwab RA, Nieminuszczy J, Shah F, Langton J, Lopez Martinez D, Liang CC, Cohn MA, Gibbons RJ, Deans AJ, Niedzwiedz W (2015) The Fanconi Anemia Pathway Maintains Genome Stability by Coordinating Replication and Transcription. *Mol Cell* 60: 351-361

Seefelder M, Kochanek S (2021) A meta-analysis of transcriptomic profiles of Huntington's disease patients. *PLoS One* 16: e0253037

Sellier C, Buijsen RAM, He F, Natla S, Jung L, Tropel P, Gaucherot A, Jacobs H, Meziane H, Vincent A *et al* (2017) Translation of Expanded CGG Repeats into FMRpolyG Is Pathogenic and May Contribute to Fragile X Tremor Ataxia Syndrome. *Neuron* 93: 331-347

Senichkin VV, Prokhorova EA, Zhivotovsky B, Kopeina GS (2021) Simple and Efficient Protocol for Subcellular Fractionation of Normal and Apoptotic Cells. *Cells* 10

Seong IS, Woda JM, Song JJ, Lloret A, Abeyrathne PD, Woo CJ, Gregory G, Lee JM, Wheeler VC, Walz T *et al* (2010) Huntingtin facilitates polycomb repressive complex 2. *Hum Mol Genet* 19: 573-583

Sharma M, Rajendrarao S, Shahani N, Ramirez-Jarquin UN, Subramaniam S (2020) Cyclic GMP-AMP synthase promotes the inflammatory and autophagy responses in Huntington disease. *Proc Natl Acad Sci U S A* 117: 15989-15999

Shimohata T, Nakajima T, Yamada M, Uchida C, Onodera O, Naruse S, Kimura T, Koide R, Nozaki K, Sano Y *et al* (2000) Expanded polyglutamine stretches interact with TAFII130, interfering with CREB-dependent transcription. *Nat Genet* 26: 29-36

Shirasaki DI, Greiner ER, Al-Ramahi I, Gray M, Boontheung P, Geschwind DH, Botas J, Coppola G, Horvath S, Loo JA *et al* (2012) Network organization of the huntingtin proteomic interactome in mammalian brain. *Neuron* 75: 41-57

Shiromoto Y, Sakurai M, Minakuchi M, Ariyoshi K, Nishikura K (2021) ADAR1 RNA editing enzyme regulates R-loop formation and genome stability at telomeres in cancer cells. *Nat Commun* 12: 1654

Shivji MK, Renaudin X, Williams ÇH, Venkitaraman AR (2018) BRCA2 regulates transcription elongation by RNA polymerase II to prevent R-loop accumulation. *Cell reports* 22: 1031-1039

Silva B, Arora R, Bione S, Azzalin CM (2021) TERRA transcription destabilizes telomere integrity to initiate break-induced replication in human ALT cells. *Nature Communications* 12: 3760

Silvestroni A, Faull RL, Strand AD, Moller T (2009) Distinct neuroinflammatory profile in post-mortem human Huntington's disease. *Neuroreport* 20: 1098-1103

Sipione S, Rigamonti D, Valenza M, Zuccato C, Conti L, Pritchard J, Kooperberg C, Olson JM, Cattaneo E (2002) Early transcriptional profiles in huntingtin-inducible striatal cells by microarray analyses. *Hum Mol Genet* 11: 1953-1965

Skourti-Stathaki K, Proudfoot NJ (2014) A double-edged sword: R loops as threats to genome integrity and powerful regulators of gene expression. *Genes Dev* 28: 1384-1396

Skourti-Stathaki K, Proudfoot NJ, Gromak N (2011) Human senataxin resolves RNA/DNA hybrids formed at transcriptional pause sites to promote Xrn2-dependent termination. *Mol Cell* 42: 794-805

Smolka JA, Sanz LA, Hartono SR, Chédin F (2021) Recognition of RNA by the S9. 6 antibody creates pervasive artifacts when imaging RNA: DNA hybrids. *Journal of Cell Biology* 220: e202004079

Sobczak K, Krzyzosiak WJ (2004) Imperfect CAG repeats form diverse structures in SCA1 transcripts. *J Biol Chem* 279: 41563-41572

Sollier J, Stork CT, Garcia-Rubio ML, Paulsen RD, Aguilera A, Cimprich KA (2014) Transcription-coupled nucleotide excision repair factors promote R-loop-induced genome instability. *Mol Cell* 56: 777-785

Soragni E, Herman D, Dent SY, Gottesfeld JM, Wells RD, Napierala M (2008) Long intronic GAA*TTC repeats induce epigenetic changes and reporter gene silencing in a molecular model of Friedreich ataxia. *Nucleic Acids Res* 36: 6056-6065

Sordet O, Redon CE, Guirouilh-Barbat J, Smith S, Solier S, Douarre C, Conti C, Nakamura AJ, Das BB, Nicolas E *et al* (2009) Ataxia telangiectasia mutated activation by transcription- and topoisomerase I-induced DNA double-strand breaks. *EMBO Rep* 10: 887-893

Sorensen SA, Fenger K, Olsen JH (1999) Significantly lower incidence of cancer among patients with Huntington disease: An apoptotic effect of an expanded polyglutamine tract? *Cancer* 86: 1342-1346

Statello L, Guo CJ, Chen LL, Huarte M (2021) Gene regulation by long non-coding RNAs and its biological functions. *Nat Rev Mol Cell Biol* 22: 96-118

Steffan JS, Agrawal N, Pallos J, Rockabrand E, Trotman LC, Slepko N, Illes K, Lukacsovich T, Zhu YZ, Cattaneo E *et al* (2004) SUMO modification of Huntingtin and Huntington's disease pathology. *Science* 304: 100-104

Steffan JS, Bodai L, Pallos J, Poelman M, McCampbell A, Apostol BL, Kazantsev A, Schmidt E, Zhu YZ, Greenwald M *et al* (2001) Histone deacetylase inhibitors arrest polyglutamine-dependent neurodegeneration in Drosophila. *Nature* 413: 739-743

Steffan JS, Kazantsev A, Spasic-Boskovic O, Greenwald M, Zhu YZ, Gohler H, Wanker EE, Bates GP, Housman DE, Thompson LM (2000) The Huntington's disease protein interacts with p53 and CREB-binding protein and represses transcription. *Proc Natl Acad Sci U S A* 97: 6763-6768

Stork CT, Bocek M, Crossley MP, Sollier J, Sanz LA, Chedin F, Swigut T, Cimprich KA (2016) Co-transcriptional R-loops are the main cause of estrogen-induced DNA damage. *Elife* 5

Stoy H, Zwicky K, Kuster D, Lang KS, Krietsch J, Crossley MP, Schmid JA, Cimprich KA, Merrikh H, Lopes M (2023) Direct visualization of transcription-replication conflicts reveals post-replicative DNA:RNA hybrids. *Nat Struct Mol Biol* 30: 348-359

Strong TV, Tagle DA, Valdes JM, Elmer LW, Boehm K, Swaroop M, Kaatz KW, Collins FS, Albin RL (1993) Widespread expression of the human and rat Huntington's disease gene in brain and nonneural tissues. *Nat Genet* 5: 259-265

Sun JH, Zhou L, Emerson DJ, Phyo SA, Titus KR, Gong W, Gilgenast TG, Beagan JA, Davidson BL, Tassone F *et al* (2018) Disease-Associated Short Tandem Repeats Co-localize with Chromatin Domain Boundaries. *Cell* 175: 224-238 e215

Sun X, Li PP, Zhu S, Cohen R, Marque LO, Ross CA, Pulst SM, Chan HY, Margolis RL, Rudnicki DD (2015) Nuclear retention of full-length HTT RNA is mediated by splicing factors MBNL1 and U2AF65. *Sci Rep* 5: 12521

Swami M, Hendricks AE, Gillis T, Massood T, Mysore J, Myers RH, Wheeler VC (2009) Somatic expansion of the Huntington's disease CAG repeat in the brain is associated with an earlier age of disease onset. *Hum Mol Genet* 18: 3039-3047

Sweeney P, Park H, Baumann M, Dunlop J, Frydman J, Kopito R, McCampbell A, Leblanc G, Venkateswaran A, Nurmi A *et al* (2017) Protein misfolding in neurodegenerative diseases: implications and strategies. *Transl Neurodegener* 6: 6

Tabrizi SJ, Estevez-Fraga C, van Roon-Mom WMC, Flower MD, Scahill RI, Wild EJ, Munoz-Sanjuan I, Sampaio C, Rosser AE, Leavitt BR (2022) Potential disease-modifying therapies for Huntington's disease: lessons learned and future opportunities. *Lancet Neurol* 21: 645-658

Tabrizi SJ, Flower MD, Ross CA, Wild EJ (2020) Huntington disease: new insights into molecular pathogenesis and therapeutic opportunities. *Nat Rev Neurol* 16: 529-546

Tabrizi SJ, Scahill RI, Owen G, Durr A, Leavitt BR, Roos RA, Borowsky B, Landwehrmeyer B, Frost C, Johnson H *et al* (2013) Predictors of phenotypic progression and disease onset in premanifest and early-stage Huntington's disease in the TRACK-HD study: analysis of 36-month observational data. *Lancet Neurol* 12: 637-649

Takano H, Gusella JF (2002) The predominantly HEAT-like motif structure of huntingtin and its association and coincident nuclear entry with dorsal, an NF-kB/Rel/dorsal family transcription factor. *BMC Neurosci* 3: 15

Taneja KL, McCurrach M, Schalling M, Housman D, Singer RH (1995) Foci of trinucleotide repeat transcripts in nuclei of myotonic dystrophy cells and tissues. *J Cell Biol* 128: 995-1002

Thomas M, White RL, Davis RW (1976) Hybridization of RNA to double-stranded DNA: formation of R-loops. *Proc Natl Acad Sci U S A* 73: 2294-2298

Thongthip S, Carlson A, Crossley MP, Schwer B (2022) Relationships between genome-wide R-loop distribution and classes of recurrent DNA breaks in neural stem/progenitor cells. *Sci Rep* 12: 13373

Toledo-Sherman L, Breccia P, Cachope R, Bate JR, Angulo-Herrera I, Wishart G, Matthews KL, Martin SL, Cox HC, McAllister G *et al* (2019) Optimization of Potent and Selective Ataxia Telangiectasia-Mutated Inhibitors Suitable for a Proof-of-Concept Study in Huntington's Disease Models. *J Med Chem* 62: 2988-3008

Tran PLT, Pohl TJ, Chen CF, Chan A, Pott S, Zakian VA (2017) PIF1 family DNA helicases suppress R-loop mediated genome instability at tRNA genes. *Nat Commun* 8: 15025

Tsoi H, Lau TC, Tsang SY, Lau KF, Chan HY (2012) CAG expansion induces nucleolar stress in polyglutamine diseases. *Proc Natl Acad Sci U S A* 109: 13428-13433

Tsutakawa SE, Sarker AH, Ng C, Arvai AS, Shin DS, Shih B, Jiang S, Thwin AC, Tsai MS, Willcox A *et al* (2020) Human XPG nuclease structure, assembly, and activities with insights for neurodegeneration and cancer from pathogenic mutations. *Proc Natl Acad Sci U S A* 117: 14127-14138

Tuduri S, Crabbe L, Conti C, Tourriere H, Holtgreve-Grez H, Jauch A, Pantesco V, De Vos J, Thomas A, Theillet C *et al* (2009) Topoisomerase I suppresses genomic instability by preventing interference between replication and transcription. *Nat Cell Biol* 11: 1315-1324

Tummala H, Dokal AD, Walne A, Ellison A, Cardoso S, Amirthasigamanipillai S, Kirwan M, Browne I, Sidhu JK, Rajeeve V *et al* (2018) Genome instability is a consequence of transcription deficiency in patients with bone marrow failure harboring biallelic ERCC6L2 variants. *Proc Natl Acad Sci U S A* 115: 7777-7782

van Bilsen PH, Jaspers L, Lombardi MS, Odekerken JC, Burright EN, Kaemmerer WF (2008) Identification and allele-specific silencing of the mutant huntingtin allele in Huntington's disease patient-derived fibroblasts. *Hum Gene Ther* 19: 710-719

van Cruchten RTP, Wieringa B, Wansink DG (2019) Expanded CUG repeats in DMPK transcripts adopt diverse hairpin conformations without influencing the structure of the flanking sequences. *RNA* 25: 481-495

Van de Poel A, Toledo-Sherman L, Breccia P, Cachope R, Bate JR, Angulo-Herrera I, Wishart G, Matthews KL, Martin SL, Peacock M *et al* (2021) Structure-Based Exploration of Selectivity for ATM Inhibitors in Huntington's Disease. *J Med Chem* 64: 5018-5036

van Eyk CL, Samaraweera SE, Scott A, Webber DL, Harvey DP, Mecinger O, O'Keefe LV, Cropley JE, Young P, Ho J *et al* (2019) Non-self mutation: double-stranded RNA elicits antiviral pathogenic response in a Drosophila model of expanded CAG repeat neurodegenerative diseases. *Hum Mol Genet* 28: 3000-3012

Van Raamsdonk JM, Murphy Z, Selva DM, Hamidizadeh R, Pearson J, Petersen A, Bjorkqvist M, Muir C, Mackenzie IR, Hammond GL *et al* (2007) Testicular degeneration in Huntington disease. *Neurobiol Dis* 26: 512-520

Vohhodina J, Goehring LJ, Liu B, Kong Q, Botchkarev Jr VV, Huynh M, Liu Z, Abderazzaq FO, Clark AP, Ficarro SB (2021) BRCA1 binds TERRA RNA and suppresses R-Loop-based telomeric DNA damage. *Nature communications* 12: 3542

von Schimmelmann M, Feinberg PA, Sullivan JM, Ku SM, Badimon A, Duff MK, Wang Z, Lachmann A, Dewell S, Ma'ayan A *et al* (2016) Polycomb repressive complex 2 (PRC2) silences genes responsible for neurodegeneration. *Nat Neurosci* 19: 1321-1330

Wang ET, Cody NA, Jog S, Biancolella M, Wang TT, Treacy DJ, Luo S, Schroth GP, Housman DE, Reddy S *et al* (2012a) Transcriptome-wide regulation of pre-mRNA splicing and mRNA localization by muscleblind proteins. *Cell* 150: 710-724

Wang IX, Grunseich C, Fox J, Burdick J, Zhu Z, Ravazian N, Hafner M, Cheung VG (2018) Human proteins that interact with RNA/DNA hybrids. *Genome Res* 28: 1405-1414

Wang L, Schuster GU, Hultenby K, Zhang Q, Andersson S, Gustafsson JA (2002) Liver X receptors in the central nervous system: from lipid homeostasis to neuronal degeneration. *Proc Natl Acad Sci U S A* 99: 13878-13883

Wang R, Luo Y, Ly PT, Cai F, Zhou W, Zou H, Song W (2012b) Sp1 regulates human huntingtin gene expression. *J Mol Neurosci* 47: 311-321

Wanowska E, Kubiak MR, Rosikiewicz W, Makalowska I, Szczesniak MW (2018) Natural antisense transcripts in diseases: From modes of action to targeted therapies. *Wiley Interdiscip Rev RNA* 9

Weinreb JT, Ghazale N, Pradhan K, Gupta V, Potts KS, Tricomi B, Daniels NJ, Padgett RA, De Oliveira S, Verma A *et al* (2021) Excessive R-loops trigger an inflammatory cascade leading to increased HSPC production. *Dev Cell* 56: 627-640 e625

Wells JP, White J, Stirling PC (2019) R loops and their composite cancer connections. *Trends in Cancer* 5: 619-631

Wexler NS, Lorimer J, Porter J, Gomez F, Moskowitz C, Shackell E, Marder K, Penchaszadeh G, Roberts SA, Gayan J *et al* (2004) Venezuelan kindreds reveal that genetic and environmental factors modulate Huntington's disease age of onset. *Proc Natl Acad Sci U S A* 101: 3498-3503

White JK, Auerbach W, Duyao MP, Vonsattel JP, Gusella JF, Joyner AL, MacDonald ME (1997) Huntingtin is required for neurogenesis and is not impaired by the Huntington's disease CAG expansion. *Nat Genet* 17: 404-410

Wight M, Werner A (2013) The functions of natural antisense transcripts. *Essays Biochem* 54: 91-101

Wild EJ, Tabrizi SJ (2017) Therapies targeting DNA and RNA in Huntington's disease. *Lancet Neurol* 16: 837-847

Wojciechowska M, Krzyzosiak WJ (2011) Cellular toxicity of expanded RNA repeats: focus on RNA foci. *Hum Mol Genet* 20: 3811-3821

Wong YC, Holzbaur EL (2014) The regulation of autophagosome dynamics by huntingtin and HAP1 is disrupted by expression of mutant huntingtin, leading to defective cargo degradation. *J Neurosci* 34: 1293-1305

Wood M, Quinet A, Lin Y-L, Davis AA, Pasero P, Ayala YM, Vindigni A (2020) TDP-43 dysfunction results in R-loop accumulation and DNA replication defects. *Journal of cell science* 133: jcs244129

Wright SE, Todd PK (2023) Native functions of short tandem repeats. *Elife* 12

Xia J, Lee DH, Taylor J, Vandelft M, Truant R (2003) Huntingtin contains a highly conserved nuclear export signal. *Hum Mol Genet* 12: 1393-1403

Yan Q, Shields EJ, Bonasio R, Sarma K (2019) Mapping native R-loops genome-wide using a targeted nuclease approach. *Cell reports* 29: 1369-1380. e1365

Yan Q, Wulfridge P, Doherty J, Fernandez-Luna JL, Real PJ, Tang H-Y, Sarma K (2022) Proximity labeling identifies a repertoire of site-specific R-loop modulators. *Nature Communications* 13: 53

Yanai A, Huang K, Kang R, Singaraja RR, Arstikaitis P, Gan L, Orban PC, Mullard A, Cowan CM, Raymond LA *et al* (2006) Palmitoylation of huntingtin by HIP14 is essential for its trafficking and function. *Nat Neurosci* 9: 824-831

Yang Y, McBride KM, Hensley S, Lu Y, Chedin F, Bedford MT (2014) Arginine methylation facilitates the recruitment of TOP3B to chromatin to prevent R loop accumulation. *Mol Cell* 53: 484-497

Yano H, Baranov SV, Baranova OV, Kim J, Pan Y, Yablonska S, Carlisle DL, Ferrante RJ, Kim AH, Friedlander RM (2014) Inhibition of mitochondrial protein import by mutant huntingtin. *Nat Neurosci* 17: 822-831

Yoh SM, Lucas JS, Jones KA (2008) The Iws1:Spt6:CTD complex controls cotranscriptional mRNA biosynthesis and HYPB/Setd2-mediated histone H3K36 methylation. *Genes Dev* 22: 3422-3434

Yohrling GJ, Farrell LA, Hollenberg AN, Cha JH (2003) Mutant huntingtin increases nuclear corepressor function and enhances ligand-dependent nuclear hormone receptor activation. *Mol Cell Neurosci* 23: 28-38

Yoshimura SH, Hirano T (2016) HEAT repeats - versatile arrays of amphiphilic helices working in crowded environments? *J Cell Sci* 129: 3963-3970

Yu K, Chedin F, Hsieh CL, Wilson TE, Lieber MR (2003) R-loops at immunoglobulin class switch regions in the chromosomes of stimulated B cells. *Nat Immunol* 4: 442-451

Zafeiriou DI, Thorel F, Andreou A, Kleijer WJ, Raams A, Garritsen VH, Gombakis N, Jaspers NG, Clarkson SG (2001) Xeroderma pigmentosum group G with severe neurological involvement and features of Cockayne syndrome in infancy. *Pediatr Res* 49: 407-412

Zala D, Hinckelmann MV, Yu H, Lyra da Cunha MM, Liot G, Cordelieres FP, Marco S, Saudou F (2013) Vesicular glycolysis provides on-board energy for fast axonal transport. *Cell* 152: 479-491

Zhang C, Lu Y-j, Chen B, Bai Z, Hervieu A, Licciardello MP, Wang M, Mitsopoulos C, Al-Lazikani B, Totorici M *et al* (2022) R-loop editing by DNA cytosine deaminase APOBEC3B determines the activity of estrogen receptor enhancers. *bioRxiv*: 2022.2010.2021.513235

Zhang N, Ashizawa T (2017) RNA toxicity and foci formation in microsatellite expansion diseases. *Current opinion in genetics & development* 44: 17-29

Zhang T, Cooper S, Brockdorff N (2015) The interplay of histone modifications - writers that read. *EMBO Rep* 16: 1467-1481

Zhang Y, Liu T, Meyer CA, Eeckhoute J, Johnson DS, Bernstein BE, Nusbaum C, Myers RM, Brown M, Li W *et al* (2008) Model-based analysis of ChIP-Seq (MACS). *Genome Biol* 9: R137

Zheng S, Clabough EB, Sarkar S, Futter M, Rubinsztein DC, Zeitlin SO (2010) Deletion of the huntingtin polyglutamine stretch enhances neuronal autophagy and longevity in mice. *PLoS Genet* 6: e1000838

Zheng Z, Li A, Holmes BB, Marasa JC, Diamond MI (2013) An N-terminal nuclear export signal regulates trafficking and aggregation of Huntingtin (Htt) protein exon 1. *J Biol Chem* 288: 6063-6071

Zuccato C, Cattaneo E (2009) Brain-derived neurotrophic factor in neurodegenerative diseases. *Nat Rev Neurol* 5: 311-322

Zuccato C, Tartari M, Crotti A, Goffredo D, Valenza M, Conti L, Cataudella T, Leavitt BR, Hayden MR, Timmusk T *et al* (2003) Huntingtin interacts with REST/NRSF to modulate the transcription of NRSE-controlled neuronal genes. *Nat Genet* 35: 76-83

Zuccato C, Valenza M, Cattaneo E (2010) Molecular mechanisms and potential therapeutical targets in Huntington's disease. *Physiol Rev* 90: 905-981