Huntingtin CAG repeat expansions in induced pluripotent stem cell models of Huntington's Disease

Jasmine Jo Donaldson

A thesis submitted to Cardiff University for the degree of
Doctor of Philosophy

September 2019

APPENDIX 1 - STATEMENTS AND DECLARATIONS TO BE SIGNED BY THE CANDIDATE AND INCLUDED IN THE THESIS

STATEMENT 1

This thesis is being submitted in partial fulfilment of the requirements for the degree of ... (*insert PhD, MD, MPhil, etc., as appropriate*)

Signed _____ Date _____

STATEMENT 2

This work has not been submitted in substance for any other degree or award at this or any other university or place of learning, nor is it being submitted concurrently for any other degree or award (outside of any formal collaboration agreement between the University and a partner organisation)

Signed _____ Date _____

STATEMENT 3

I hereby give consent for my thesis, if accepted, to be available in the University's Open Access repository (or, where approved, to be available in the University's library and for inter-library loan), and for the title and summary to be made available to outside organisations, subject to the expiry of a University-approved bar on access if applicable.

Signed _____ Date _____

DECLARATION

This thesis is the result of my own independent work, except where otherwise stated, and the views expressed are my own. Other sources are acknowledged by explicit references. The thesis has not been edited by a third party beyond what is permitted by Cardiff University's Use of Third Party Editors by Research Degree Students Procedure.

Signed _____ Date _____

WORD COUNT 44,787 words

(Excluding summary, acknowledgements, declarations, contents pages, appendices, tables, diagrams and figures, references, bibliography, footnotes and endnotes)

Acknowledgements

I would like to sincerely thank each of my supervisors; Lesley Jones, Nicholas Allen and Thomas Massey. Lesley, for your support, encouragement and all of the opportunities you have given me over the course of my PhD. Nick, for your endless hours teaching me tissue culture, unwavering support and your never-ending ideas. And Tom, for your excellent guidance throughout this project and for being more like a friend.

I would also like to thank the Wellcome Trust for funding my PhD studentship, and all those on the Wellcome Trust course who have made my time in Cardiff so enjoyable. Particularly Frankie, Tommy, Matt, Tom, Hannah, Matt.

Thanks goes to all the people in the Allen lab that made the long nights and weekend work hilarious, particularly Kim, Emma, Jincy, Elisa and Aurelien. There's no one else I'd rather fight over a tissue culture hood with.

I would particularly like to thank Lyn for all the time you've given me over the past 3 years, I am in awe of your kind nature and ability to teach. Special thanks also go to all the present and past members of the Jones/ Massey lab; Laura, Freja, Bran, Laura, Ant, Sergey and Joe, who have helped me over the years and made lab meetings fun! And Adele, who knew learning CRISPR could be so fun. You make the lab the best place to be, thanks for your help with everything.

Thanks also to Emily, Aine, Caer, Jess and Rhys. Cardiff wouldn't feel like a home without you all.

A huge thank you goes to Sinead, who I was lucky enough to meet during my first year in Cardiff and who has been a huge part of my life ever since. You made this PhD journey infinitely brighter and I'm so thankful we were able to experience and finish it together.

Thank you to all my family and friends back home, who have offered a welcome relief from work when I return. A special thanks to my parents, who continue to inspire me daily in all that they do. Thank you for always believing in me.

Finally, thank you to Jamie. For your constant love and support and for making me smile at the end of every day.

Thesis summary

Huntington's disease (HD) is an autosomal dominant neurodegenerative disorder caused by the expansion of a CAG repeat in the huntingtin gene, *HTT*. The length of the CAG repeat is inversely correlated with age at motor onset, but other factors influence onset including genetic variation elsewhere in the genome. Recent genome-wide association studies have identified genetic variants in the vicinity of *FAN1*, a nuclease involved in DNA interstrand cross-link repair, as modifiers of age at onset of HD. It is thought that FAN1 might modify age at onset through direct modulation of the expanded *HTT* CAG repeat. Further expansions of the CAG repeat in post-mitotic neurons are thought to accelerate disease-onset and therefore factors that license or inhibit expansions represent therapeutic targets.

This thesis demonstrates that HD-iPSC lines with expanded CAG repeat tracts of >100 CAG repeats (Q109) exhibit repeat instability in culture, with the repeat tract undergoing further expansions in pluripotent cells and upon neuronal differentiation. These cell lines therefore represent a cellular model of repeat expansion which can be utilised to characterise how DNA repair genes affect cells harbouring expanded CAG repeats. Employing CRISPR-Cas9 and a piggyBac transposon-based homologous recombination approach the expanded *HTT* CAG repeat tract was genetically corrected to a wild-type repeat length of 22 *HTT* CAGs (Q22). Corrected HD-iPSCs retained pluripotency and differentiation potential. Additionally, metabolic deficits in Q109 neural progenitor cells were rescued in Q22 neural progenitor cells.

This thesis also aimed to establish the effect of FAN1 on *HTT* CAG repeat expansions. Using CRISPR-Cas9, *FAN1* was knocked-out of an isogenic pair of HD-iPSCs containing either 22 or >100 *HTT* CAG repeats. FAN1$^{-/-}$ clones were significantly more susceptible to interstrand cross-linking agent MMC, indicative of defective interstrand cross-link repair. Increased CAG repeat expansion was observed in FAN1$^{-/-}$ iPSCs and iPSC-derived neurons, suggestive of a protective role of FAN1 against CAG repeat expansions.

These novel model systems provide a platform for investigating the cellular phenotypes associated with expanded CAG repeats, and the effects of DNA-repair associated genetic modifiers.

List of Abbreviations

AAG	Alkyladenine glycosylase
ADF	Advanced DMEM/F-12
ADP	Adenosine Diphosphate
APE1	Apurinic/ apyrimidinic endonuclease
ATM	Ataxia telangiectasia mutated
ATP	Adenosine Triphosphate
ATR	Ataxia telangiectasia and Rad3 related protein
ATXN	Ataxin
BER	Base excision repair
BCA	Bicinchoninic acid assay
BSA	Bovine serum albumin
BDNF	Brain-derived neurotrophic factor
BWA	Burrows-Wheeler Aligner
CAG	Cytosine-Adenine-Guanosine Trinucleotide
CBP	CREB (cAMP response element-binding protein)- binding protein
CDK7	Cyclin-dependent kinase 7
CNV	Copy Number Variant
CREB	cAMP response element-binding protein
CRISPR	Clustered Regularly Interspaced Short Palindromic Repeats
CSB	Cockayne syndrome protein B
CTIP2	COUP TF1-interacting protein 2
DARPP-32	Dopamine- and cAMP-regulated neuronal phosphoprotein
DDR	DNA damage response
DM1	Myotonic Dystrophy Type 1
DM2	Myotonic Dystrophy Type 2

DMPK	Myotonic dystrophy protein kinase
DNA	Deoxyribonucleic acid
DNA-PK	DNA-dependent protein kinase
DRD1	Dopamine 1 receptor
DRD2	Dopamine 2 receptor
DRP1	Dynamin-related protein 1
DRPLA	Dentatorubral- pallidoluysian atrophy
eQTL	Expression quantitative trait loci
EXO1	Exonuclease 1
FA	Fanconi anaemia
FACS	Fluorescence-Activated Cell Sorting
FAN1	FANCD2/FANCI-associated nuclease 1
FANCC	Fanconi anemia, complementation group C
FANCJ	Fanconi anemia, complementation group J
FMR1	Fragile X mental retardation 1
FOXG1	Forkhead Box G1
FXD	Fragile X-related disorders
γH2AX	Phosphorylated histone family member X
GABA	Gamma-Aminobutyric Acid
GATK	Genome Analysis Toolkit
gRNA	Guide RNA
GWAS	Genome-Wide Association Studies
HCS	High-Content Screening
HD	Huntington's Disease
HEAT	Huntingtin, elongation factor 3, protein phosphatase 2A and TOR1
hESCs	Human Embryonic Stem Cells
hPSCs	Human Pluripotent Stem Cells
HR	Homologous recombination

HTT	Huntingtin Protein
ICL	Interstrand cross-links
iPSCs	Induced pluripotent stem cells
JHD	Juvenile HD
KO	Knock-out
LB	Luria broth
LIG1	DNA ligase 1
mHTT	Mutant huntingtin protein
MLH1	MutL homolog 1
MLH3	MutL homolog 3
MMC	Mitomycin C
MMR	Mismatch repair
MPT	Mitochondrial permeability transition pore
mRNA	Messenger RNA
MSH2	MutS homolog 2
MSH3	MutS homolog 3
MSH6	MutS homolog 6
MSN	Medium Spiny Neuron
MTRM10	Myotubularin related protein 10
MutSα	Heterodimer of MSH2 and MSH6
MutSβ	Heterodimer of MSH2 and MSH3
MutLα	Heterodimer of MLH1 and PMS2
MutLγ	Heterodimer of MLH1 and MLH3
N-17	First 17 amino acids in exon 1 of HTT
NAD^+	Nicotinamide adenine dinucleotide
NADH	Nicotinamide adenine dinucleotide + hydrogen
NCoR	Nuclear receptor co-repressor

NEIL1	nei endonuclease VIII like 1
NHEJ	Non-homologous end joining
NPCs	Neural progenitor Cells
NTH1	Endonuclease III-like protein 1
OGG1	8-oxoG DNA glycosylase
p53	Tumour protein 53
PAR	Poly-ADP-ribose
PARP	Poly-(ADP-Ribose) Polymerase
PAX6	Paired Box 6
PBS	Phosphate buffered saline
PCR	Polymerase Chain Reaction
PFA	Paraformaldehyde
PGC-1α	Peroxisome proliferator-activated receptor gamma coactivator-1α
PMS1	Postmeiotic Segregation Increased 1
PMS2	Postmeiotic Segregation Increased 2
PNKP	Polynucleotide kinase- phosphatase
POLβ	DNA polymerase beta
POLRA	DNA-directed RNA polymerase II subunit RPB1
QC	Quality control
qRT-PCR	Quantitative reverse-transcription PCR
RE1	Repressor element 1
REST	Repressor element 1-silencing transcription factor
RIPA	Radio-immunoprecipitation assay buffer
RLT	RNA lysis buffer
RNA	Ribonucleic acid
ROS	Reactive Oxygen Species
RPA	Replication protein A3

RRM2B	Ribonucleotide-diphosphate reductase subunit M2 B
SCA	Spinocerebellar ataxia
shRNA	Short-hairpin RNA
SMBA	Spinal and bulbar muscular atrophy
SNP	Single nucleotide polymorphism
U2OS	Osteosarcoma epithelial cell line
UBR5	Ubiquitin-protein ligase E3 component N-recognin 5
UBZ	Ubiquitin-binding zinc finger
VCF	Variant calling file
VRR-NUC	Virus-type replication-repair nuclease domain
WT	Wild Type
XPC	Complementation group C
ZFN	Zinc-finger nuclease

Table of Contents

List of Figures

List of Tables

Chapter 1: General Introduction

1.1 Background to Huntington's Disease

1.1.1 Historical Background

The first definite descriptions of Huntington's disease (HD) date back to 1842, when they were reported in a letter written by Charles Oscar Waters, which was published in *Practise of Medicine.* Waters described the hereditary nature of the disease as well as some of the symptoms that the patients experienced; involuntary movements and cognitive decline leading to a dementia-like phenotype. In 1846 Dr. Charles Gorman described in his thesis that the disease tended to localise geographically, and in 1860, Johan Christopher Lund also reported on the progressive nature of the disease. Lund referred to the disease as 'twitches' or 'the inherited disease' and commented on the fact that these changes were not usually observed until the affected person had reached middle age.

However, the most complete description of the disease came in 1872 when the American physician George Huntington, building on two generations of observations before him, published his description of "chorea" (Huntington 1872). The comprehensive and graphic description of the disease meant that it subsequently came to bear his name. Huntington noted that the disease was of a hereditary nature, there was a tendency towards 'insanity' and suicide and primarily it manifested in adult life. Charles Davenport confidently described the inheritance of HD as autosomal dominant in 1911 (Davenport 1912) though it wasn't until 1983 when studies confirmed this (Hayden *et al.* 1983). Shortly thereafter, the disease-causing gene was mapped to the short arm of human chromosome 4 (Gusella *et al.* 1983). Over the next decade, with the help from families affected by HD and a large Venezuelan HD kindred, considerable efforts were made to identify the gene using linkage and linkage disequilibrium mapping. In 1993, the disease-causing mutation was identified as a polymorphic $(CAG)_n$ trinucleotide repeat at the 5' end of *IT15 (*interesting transcript 15) which subsequently became the huntingtin gene, *HTT* (MacDonald *et al.* 1993). This was a time of great excitement in the field as it meant that pre-manifest diagnoses could be made. The identification of the disease-causing mutation marked the start of a new research chapter for the HD community.

1.1.2 Epidemiology

Though HD occurs in human populations across the world, it demonstrates patterns of ethnic variation. The worldwide prevalence of HD is estimated to be 3 in 100,000, whereas the prevalence

in Caucasian populations is estimated to be around 1/6-8,000 individuals. It is thought that the variability in global prevalence rates are ancestry-specific, likely arising from genetic differences at the *HTT* locus. Populations with a high prevalence of HD tend to have alleles in the intermediate disease length range: the average *HTT* CAG repeat length is longer in individuals of European descent than those of East Asian or African descent, perhaps owing to a higher rate of *de novo* mutations (Semaka *et al.* 2006). Prevalence rates may also be influenced by the stigma associated with such disease, health care provisions available and the number of people presenting to medical professionals (Wexler 2010). The rates of predictive, prenatal and pre-implantation tests remain as low as 2-20% in most countries (Paulsen *et al.* 2006; Baig *et al.* 2016). Despite this, diagnosis of HD in the UK has doubled in the last 20 years, likely driven by medical advances, availability of diagnostic screening, an ageing population and longer HD survival (Evans *et al.* 2013).

1.1.3 Clinical phenotypes

HD is a progressive neurodegenerative disorder, clinically characterised by involuntary choreatic movements, psychiatric disturbances and cognitive decline (Huntington 1872). Clinical diagnosis of HD is based on a confirmed family history of the disease, or a positive genetic test (Warner *et al.* 1993). Disease onset is preceded by a variable prodromal phase, where subtle motor, cognitive and behavioural changes manifest up to 15 years before clinical diagnosis (Ross *et al.* 2014). Once disease symptoms begin however, the disease progresses inexorably, and death usually occurs 15-20 years following clinical onset. There is currently no cure for HD, but therapies aim to manage symptoms to give patients a better quality of life.

1.1.3.1 Motor symptoms

The slow and progressive nature of the disease means that motor symptoms gradually develop, and in cases when the individual or family is not familiar with the disease, symptoms are often not obvious until they start to impact on daily life. The motor symptoms exhibited by HD patients include chorea, hypokinesia, dystonia and tics (McColgan and Tabrizi 2018). Choreic movements are the most common symptom of HD, presenting early on as unpredictable, involuntary jerking movements, often exacerbated by anxiety and stress, as well as excitement. As the disease progresses, choreatic movements reduce and patients' movements tend to slow, showing: hypokinesia, bradykinesia and akinesia (Sánchez-Pernaute *et al.* 2000). With the decline of motor coordination, gait is significantly impaired and patients often stumble and fall. Involuntary motor control also effects the muscles involving speech; dysarthria is a common symptom reported in HD (Skodda *et al.* 2014). Dysarthria refers to difficulties in respiration, phonation, resonance and articulation, and individuals with HD often demonstrate alterations of speech rate and irregularity.

In the later stages of the disease, individuals exhibit problems with swallowing, which increases the risk of choking.

1.1.3.2 Cognitive symptoms

Subtle cognitive impairments were found to precede the onset of motor symptoms in HD mutation carriers. By the time of clinical diagnosis, cognition has often declined quite drastically, though reports vary considerably. Emotion recognition, executive functioning, visuospatial functioning and processing speeds are all thought to be impaired in HD, as measured by a range of clinical tests (Ross and Tabrizi 2011; Papoutsi *et al.* 2014; Eddy *et al.* 2016). Cognition declines over many years and most cases of HD result in dementia (Vitanova *et al.* 2019). Language is relatively spared, though as alluded to, patients do have difficulties with their speech.

1.1.3.3 Psychiatric symptoms

The onset and severity of neuropsychiatric symptoms varies greatly in individuals with HD. Dysphoria, agitation, irritability, apathy or anxiety occur fairly frequently, with estimates between 33-76%, whereas psychotic symptoms including delusions and hallucinations were comparatively less common, occurring in 3-11% of cases (Paulsen *et al.* 2001; Van Duijn *et al.* 2007; Orth *et al.* 2010). Psychosis tends to occur in the later stages of the disease, and often accompanies cognitive decline. Depression was noted early on in descriptions of HD (Huntington 1872) and lifetime prevalence estimates of depression range from 9% to 63% (Abbott *et al.* 1983; Morris 1991; Pflanz *et al.* 1991; Shiwach 1994; Paulsen *et al.* 2001). Obsessive-compulsive symptoms also tend to occur and lead to irritability and aggression (Van Duijn *et al.* 2007). The presentation of such symptoms varies considerably and is it often difficult to distinguish when they begin. Individuals with HD have often witnessed a parent suffer the same fate and thus it is difficult to disentangle whether symptoms are caused organically by the disease or psychologically due to the impact of the disease on their lives. Indeed, 50% of adults at risk for HD reported adverse events associated with the diagnosis of HD in one of their parents (van der Meer *et al.* 2012). After pneumonia, suicide is the most frequent cause of death, and occurs more frequently in early-symptomatic and premanifest gene carriers, with estimates between 10% to 20% (Paulsen *et al.* 2001; Orth *et al.* 2010).

1.1.4 Onset of Huntington's Disease

The disease typically manifests in the third and fourth decades of life, however, disease onset before the age of 20 occurs in 10% of HD cases, and these cases are defined as Juvenile HD (Hayden 1983). As in adult-onset HD, Juvenile HD patients demonstrate a triad of symptoms, however they tend to be much more severe and have a faster rate of progression (Telenius *et al.* 1993; Nahhas *et al.* 2005; Quarrell *et al.* 2013). Bradykinesia, rigidity and dystonia are more common than chorea,

and juveniles exhibit severe cognitive impairments. Disease onset above the age of 60 is known as late-onset HD (Kremer *et al.* 1993). Late-onset HD tends to have a slower progression and individuals exhibit mild chorea and subtle cognitive disturbances. The age of disease onset spans 2-85 years and is primarily determined by the *HTT* CAG repeat length (Duyao *et al.* 1993), but also influenced by the presence of genetic and environmental modifiers (Wexler 2004; GeM-HD Consortium 2015; GeM-HD Consortium 2019). There is considerable heterogeneity in the clinical phenotypes of individuals with HD and thus disease onset is often diagnosed with the onset of motor symptoms. It is commonly agreed however, that subtle symptoms are present up to a decade before motor onset, in a period termed the prodromal phase. Pre-symptomatic HD individuals show poorer performance in cognitive (Paulsen *et al.* 2006), motor and oculomotor tests (Gordon *et al.* 2000; Kirkwood *et al.* 2000). Neuropathological changes have also been reported up to ten years before the emergence of symptoms, including reduced striatal volume (Aylward *et al.* 2004), cortical thinning (Rosas *et al.* 2005) and the loss of grey and white matter (Thieben 2002).

1.2 The Neuropathology of Huntington's Disease

1.2.1 Cortico-striatal circuitry

HD is a neurodegenerative disorder predominantly affecting the striatum, the primary input to the basal ganglia (Graybiel 1998). The basal ganglia comprises a set of forebrain structures involved with motor control, cognition, mood and learning. Striatal atrophy leads to some of the clinical symptoms observed in HD (Alexander 1994; Kassubek *et al.* 2004). The striatum, which includes the caudate nucleus and the putamen, is primarily composed of medium spiny neurons (MSNs), comprising at least 95% of the total striatal cells, and cholinergic aspiny interneurons, with local connections, whose role is to modulate the output of the MSNs (Kemp and Powell 1971). MSNs projections are inhibitory, and they release the neurotransmitter gamma-ammino butyric acid (GABA). The striatum receives glutamatergic input from the thalamus and dopaminergic input from the substantia nigra pars compacta (Bolam *et al.* 2000). Projections from the striatum can occur via two pathways; an excitatory direct pathway or an inhibitory indirect pathway (Figure 1.1) (Albin *et al.* 1989; DeLong 1990). MSNs projecting via the direct pathway contain substance P and primarily express dopamine 1 (DRD1) receptors and those projecting via the indirect pathway contain enkephalin and primarily express dopamine 2 (DRD2) receptors (Keefe and Gerfen 1995). The striatum receives excitatory glutamatergic input from the cortex. In the direct pathway, this excitation increases GABAergic inhibitory output from the MSNs to the internal division of the globus pallidus and the substantia nigra pars reticular (Alexander 1994). Both the internal division of the globus pallidus and the substantia nigra pars reticular tonically inhibit the thalamus, thus

increased inhibition to these nuclei results in disinhibition of the thalamocortical and brainstem motor circuits. In the indirect pathway, the excitatory input from the cortex results in increased inhibitory output of the MSNs to the external division of the globus pallidus, which in turn tonically inhibits the subthalamic nucleus. The subthalamic nucleus is an excitatory nucleus with projections to the external division of the globus pallidus, substantia nigra pars compacta, internal division of the globus pallidus and the substantia nigra pars reticular. The excitation of the internal division of the globus pallidus and substantia nigra pars reticular results in increased inhibition of the thalamus and dampens downstream motor cortex neuron activity. Additionally, the substantia nigra pars compacta projects dopaminergic output to the striatum, modulating both direct and indirect pathways. The functional antagonism of these pathways works to closely regulate the functions of the striatum.

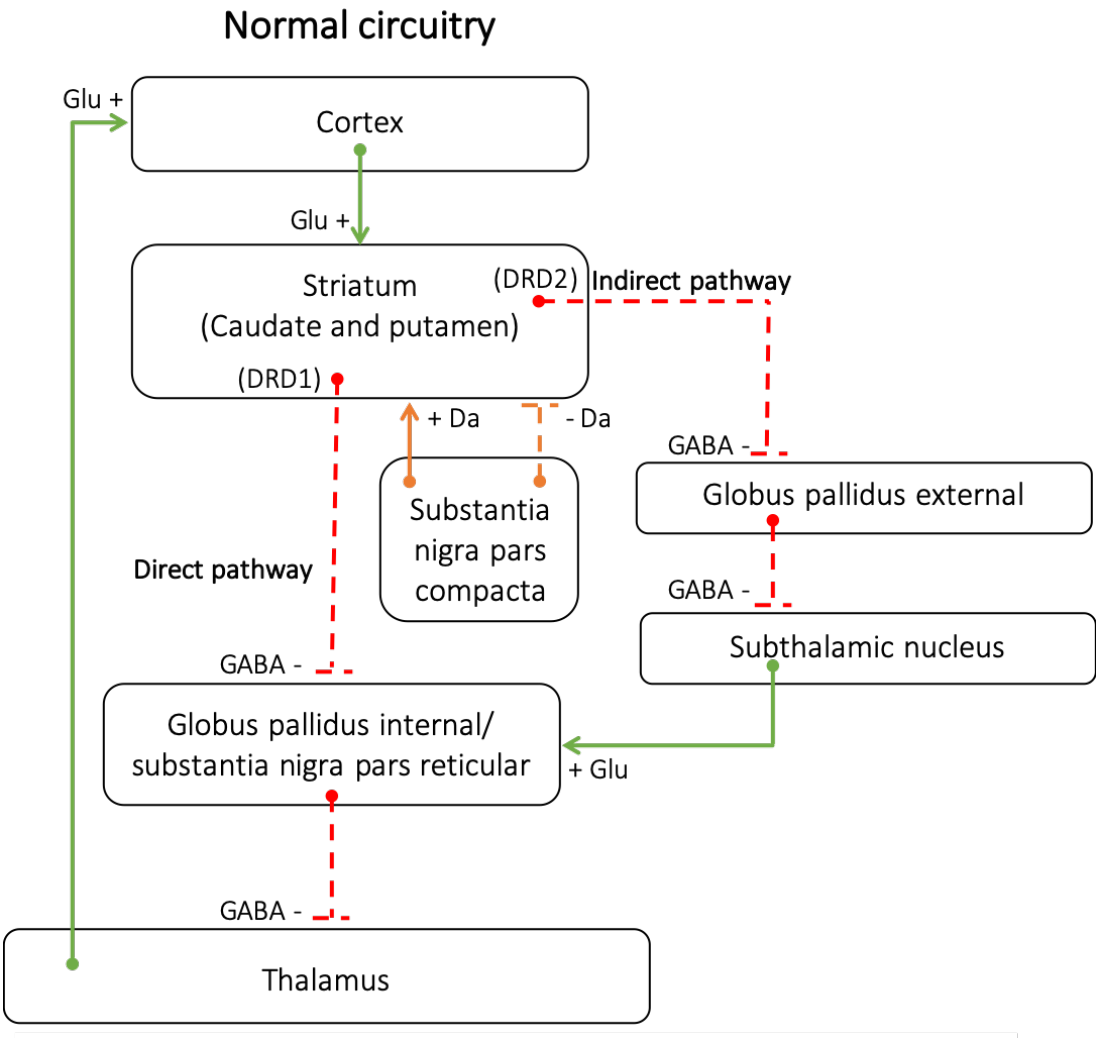


Figure 1.1 Cortico-striatal circuitry. Excitatory glutamatergic signals from the cortex stimulate the excitatory direct pathway, or the inhibitory indirect pathway from the striatum to the thalamocortical and brainstem motor circuits. (Abbreviations: DRD1: dopamine receptor 1; DRD2: dopamine receptor 2; Da: dopamine; Glu: glutamate).

1.2.2 Cortico-striatal circuitry in HD

In HD the GABAergic MSNs are particularly vulnerable to degeneration, whilst the aspiny interneurons are largely spared (Graveland *et al.* 1985). Disruption of these striatal pathways leads to the development of motor dysfunction (Figure 1.2). In the earlier stages of the disease, the enkephalin containing MSNs projecting to the external division of the globus pallidus are more susceptible to cell death than the substance P containing MSNs (Sapp *et al.* 1995). The selective loss of these neurons in the indirect pathway results in the loss of inhibition to the thalamus and motor circuits, resulting in chorea (Albin *et al.* 1990). As the disease progresses however, the loss of substance P containing MSNs results in the loss of excitatory input to the thalamus and upstream motor circuits, explaining the onset of bradykinetic movements.

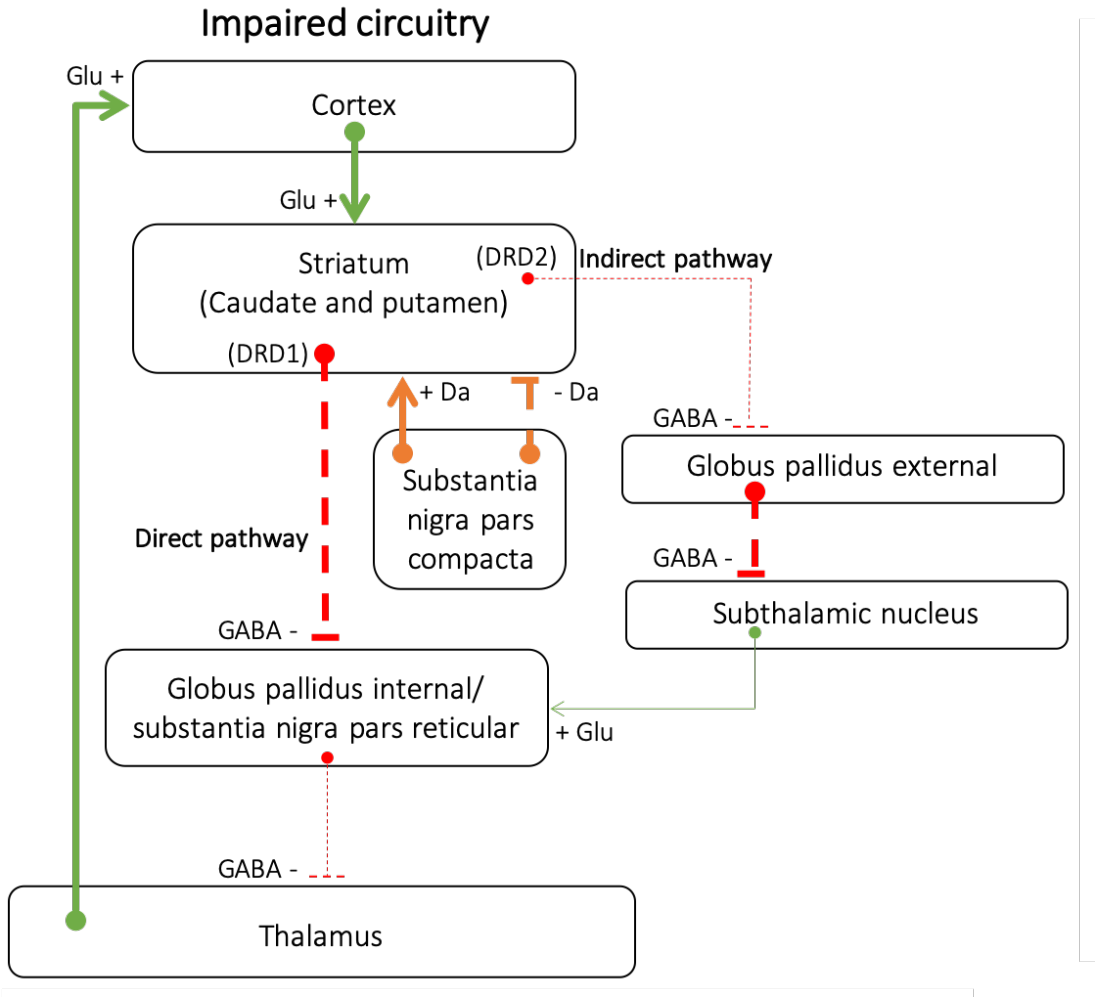


Figure 1.2 Impaired cortico-striatal circuitry in HD. Selective loss of MSNs in the indirect pathway results in disinhibition of the external division of the globus pallidus, and resultant increased inhibition of the subthalamic nucleus by the external division of the globus pallidus. This results in decreased glutamatergic output to the internal division of the globus pallidus and substantia nigra pars reticular and disinhibition of the thalamus, increasing excitatory output from the thalamus to the thalamocortical and brainstem motor circuits. (Abbreviations: DRD1: dopamine receptor 1; DRD2: dopamine receptor 2; Da: dopamine; Glu: glutamate).

1.2.3 Gross Neuropathology of Huntington's Disease

HD neuropathology follows a distinct pattern and is primarily marked by the selective loss of neuronal populations, particularly in the caudate nucleus and putamen in the striatum. Brain regions receiving striatal input, the globus pallidus and substantia nigra (Albin *et al.* 1989), are also affected. In later stages of the disease the cerebral cortex shows considerable degeneration, resulting from the loss of pyramidal neurons in layers III, V and VI (Vonsattel *et al.* 1985). Loss of cortical neurons likely results in the critical cognitive and personality-related changes observed in HD patients (Reiner *et al.* 2011). Much like clinical symptom presentation, there is considerable variation in the patterns of pathology demonstrated by cases of HD, but the degeneration is often reflective of clinical presentations of the individual.

A classification system was described to grade neuropathological severity in the brain, based on the pattern of striatal degeneration, and these grades correlate closely with clinical symptoms (Vonsattel *et al.* 1985). Such assessments are limited however in their ability to account for the heterogeneity of pathology which HD cases show (Hadzi *et al.* 2012). As well as striatal atrophy, cortical atrophy, cortical thinning (Rosas *et al.* 2002; Rosas *et al.* 2005), white matter atrophy (Dunlap 1927) and grey matter atrophy (Squitieri *et al.* 2009) are also evident in HD brains, in fact HD brains demonstrate a 20% reduction in global brain weight compared to controls (Hadzi *et al.* 2012). Structural and functional imaging studies such as functional magnetic resonance imaging (fMRI) and diffusion tensor imaging (DTI) allow *in vivo* imaging which provides an insight into the correlation of neuropathology with disease progression and symptom severity. Imaging studies reveal that neuronal atrophy can precede clinical onset by up to 12 years, suggesting that brain imaging measures could be used as markers of progression (Tabrizi *et al.* 2009; Tabrizi *et al.* 2011; Tabrizi *et al.* 2013). In adult onset HD, the cerebellum is particularly spared and shows minimal atrophy. However, cases of juvenile HD often demonstrate cerebellar atrophy, in particular a loss of Purkinje cells. Juvenile HD patients are prone to seizures and so it is speculated that cerebellar neuronal loss might be a secondary consequence of hypoxic-ischemic events.

Intracellular aggregates and nuclear inclusions of mutant HTT (mHTT) have been demonstrated in the brains of individuals with HD (Vonsattel *et al.* 1985). Inclusions are seen at a higher frequency in the cortex of juvenile HD patients than individuals with adult-onset HD, suggesting they may depend on the length of the polymorphic *HTT* CAG region. Aggregates are often reported prior to the onset of disease symptoms, leaving many to conclude that the formation of intracellular aggregates is toxic and contributes to disease pathogenesis (Li *et al.* 2001; Lee *et al.* 2004). It has been postulated that intracellular aggregates might interfere with normal cellular functioning through disruption of axonal transport, synaptic transmission or the sequestration of essential

proteins into the inclusions (Cummings *et al.* 1998; McCampbell 2000; Steffan *et al.* 2000). Conversely, others suggest that the formation of such aggregates is neuroprotective, decreasing the amounts of the soluble toxic protein (Guidetti *et al.* 2001; Zhou *et al.* 2001; Arrasate *et al.* 2004). Supporting these findings, one mouse model of HD demonstrated widespread neuronal inclusion formation, yet no signs of degeneration or neuronal dysfunction were observed (Slow *et al.* 2005). Additionally, proteins promoting inclusion formation in cellular models of HD were neuroprotective and prevented mHTT toxicity (Bodner *et al.* 2006).

1.3 The Genetic Cause of Huntington's Disease

1.3.1 The Huntingtin Gene (*HTT*) and mutation

The genetic cause of HD was mapped to chromosome 4p16.3 in 1983 using linkage analysis from human polymorphisms (Gusella *et al.* 1983). A decade later, in 1993, the disease-causing mutation was identified as a trinucleotide repeat expansion of cytosine, adenine, guanine (CAG) within the open read frame of exon 1 of the huntingtin gene (*HTT*) (MacDonald *et al.* 1993). *HTT* spans a region of ~210 kb, contains 67 exons and encodes the 348 kDa protein, huntingtin (HTT). The CAG trinucleotide repeat encodes glutamine, and the mutation therefore results in an expanded glutamine tract in HTT, which is thought to be responsible for the disease pathogenesis. The *HTT* CAG repeat is followed by a polymorphic (7-12) CCG repeat and a polymorphic (2-3) CCT repeat, both of which encode proline. The polyproline repeat tract is not thought to effect disease pathogenesis (Andrew *et al.* 1994; Panegyres *et al.* 2006; Ciosi *et al.* 2019).

Two major *HTT* isoforms have been described; a 10.3 kb transcript, and a 13.7 kb transcript containing an extended 3' UTR, which is particularly enriched in the brain (Lin *et al.* 1993). However, a number of additional isoforms have also been identified through the alternative splicing of *HTT* (Hughes *et al.* 2014; Ruzo *et al.* 2015; Neueder *et al.* 2017).

1.3.2 *HTT* CAG Repeat Length and Age at Onset

The length of the trinucleotide repeat inversely correlates with the age at which an individual develops HD; longer *HTT* CAG repeat lengths are associated with an earlier age of onset and faster disease progression (MacDonald *et al.* 1993). The *HTT* CAG repeat is polymorphic and unaffected individuals demonstrate *HTT* CAG repeat lengths between 6-26 CAGs. *HTT* CAG repeat lengths between 36-39 show reduced disease penetrance whilst repeat lengths greater than 40 CAGs are fully penetrant and invariably cause the disease. It is estimated that around 90% of HD individuals have a *HTT* CAG repeat length between 40-50 CAGs (Wexler 2004) and these repeat lengths are associated with adult-onset HD. Repeat lengths greater than 50 CAGs generally cause juvenile HD,

and repeat lengths greater than 80 *HTT* CAGs are associated with childhood onset HD, where the age of onset is before 10 years of age (Telenius *et al.* 1993).

1.3.3 Homozygous mutation carriers

Whilst most individuals with HD are heterozygous for the disease mutation, a number of homozygous mutation carries have been identified (Wexler *et al.* 1987; Squitieri *et al.* 2003). These individuals have been beneficial in establishing whether disease phenotypes are alleviated by the presence of a normal gene product, or whether the aberrant gene product is sufficient to cause complete phenotypic dominance. Wexler *et al.* (1987) reported no effect on phenotype or progression of HD homozygotes compared with HD heterozygotes (Wexler *et al.* 1987), and these findings have been echoed in further studies, indicating a toxic gain-of-function (Durr *et al.* 1999; Lee *et al.* 2012). A larger study however, reported no change in age at onset of the disease but an accelerated clinical phenotype and quicker rate of progression (Squitieri *et al.* 2003). These data suggest that whilst there may be a small difference in clinical presentation between heterozygotes and homozygotes, these differences are not great.

1.3.4 Trinucleotide repeat disorders

Trinucleotide repeat expansions in several genes are responsible for a number of other inherited neurogenerative disorders (Table 1.1). Trinucleotide repeats are one of many types of microsatellites that are defined as short sequences of DNA typically repeated between 5-50 times, but can range into the thousands (Tóth *et al.* 2000). Microsatellite instability contributes to a number of cancers (Yamamoto and Imai 2015), but trinucleotide repeats are of particular interest as their expansion, regardless of gene location, often results in neuronal or neuromuscular dysfunction. Trinucleotide repeat disorders share common characteristics; a dynamic repeat that is prone to expand intergenerationally and somatically, a correlation between the length of the expanded repeat and the age at which disease symptoms develop and increased severity between generations. Expansions in trinucleotide repeats accounts for over 30 disorders, and due to their mutation convergence, research findings can often be translated across diseases.

Trinucleotide repeat disorders result from the expansion of a number of triplet repeats: (CAG)•(CTG), (CGG)•(CCG), (GAA)•(TTC) and (GCN)•(NGC). Expansion of a trinucleotide repeat can occur in various regions of a gene; the coding region, 5' untranslated region (5' UTR), 3' UTR, intron and promoter region. To data, ten neurodegenerative diseases have been identified which are caused by the expansion of a trinucleotide CAG repeat within a coding region of a gene, seven of which are spinocerebellar ataxias (SCA 1, 2, 3, 6, 7, 12, 17). Fragile X-related disorders (FXD) result from the expansion of CGG trinucleotide repeat in the 5' UTR of *FMR1.* The severity of the disorder

increases as the length of the CGG repeat tract increases. Alleles with repeats in *FRM1* in the range of 55-200 CGGs are described as premutation alleles, and they give rise to fragile X-associated primary ovarian insufficiency and fragile-X associated tremor/ataxia syndrome, ovarian failure disorder and a neurodegenerative disorder, respectively. *FMR1* CGG repeats greater than 200 leads to fragile X syndrome, the most common single genetic cause of intellectual disability and autism. Myotonic dystrophy type 1 (DM1) is an autosomal dominant neuromuscular disease caused by the expansion of a CTG trinucleotide repeat in the 3' UTR of the *DMPK* gene (Brook *et al.* 1992). CTG repeats over a threshold of 50 are sufficient to cause DM1. Interestingly, myotonic dystrophy type 2 (DM2), characterised by myotonia and muscle dysfunction, is caused by a CCTG tetranucleotide repeat expansion in intron 1 of the *ZNF9* gene (Liquori *et al.* 2001).

In each of the diseases, wild-type repeat tracts are polymorphic and tend to range from ~4-50 repeats. Repeat lengths above a certain threshold invariably lead to the disease, but once this threshold is reached, further expansions correlate with earlier disease onset and increased disease severity. It is likely that the gene in which the expansion resides restricts the pattern of neuronal dysfunction and degeneration, as despite their mutational commonality, the neuropathology and phenotypes demonstrated in each disorder differ considerably. Spinocerebellar ataxias demonstrate cerebellar neuronal loss, particularly of the Purkinje cells (Chong *et al.* 1995; David *et al.* 1998), whilst in HD, striatal and cortical atrophy are prominent, with cerebellar neurons largely spared (Vonsattel *et al.* 1985). As in HD, longer repeat tracts are associated with an earlier age of onset, faster progression and more severe symptoms. However, considerable variation in age of onset is seen in each of the disorders, suggestive of environmental and genetic modifiers which are thought to influence disease pathogenesis. Sharing an underlying genetic cause, however, has provided us with invaluable knowledge about other disorders, and research is often translational between trinucleotide repeat disorders. While this thesis will focus on the *HTT* CAG repeat, research from other trinucleotide repeat disorders will be mentioned.

Repeat	Disease	Gene	Genomic locus	Repeat location	Repeat Length		References
					Wild-type	Pathogenic	
CAG	Dentatorubral-pallidoluysian atrophy (DRPLA)	*ATN1*	12p13.31	exon	6-35	>49	(Koide *et al.* 1994; Ueno *et al.* 1995)
	Huntington's Disease (HD)	*HTT*	4p16.3	exon	6-26	>35	(MacDonald *et al.* 1993)
	Spinal and bulbar muscular atrophy (SMBA)	*AR*	Xq11-12	exon	13-31	>40	(Spada *et al.* 1991)
	Spinocerebellar ataxia 1 (SCA1)	*ATXN1*	6p23	exon	6-39	>41	(Chong *et al.* 1995)
	Spinocerebellar ataxia 2 (SCA2)	*ATXN2*	12q24.1	exon	13-31	>32	(Imbert *et al.* 1996)
	Spinocerebellar ataxia 3 (SCA3)	*ATXN3*	14q32.1	exon	11-44	>60	(Dürr *et al.* 1996)
	Spinocerebellar ataxia 6 (SCA6)	*CACNA1A*	19p13.1	exon	4-18	>18	(Jodice *et al.* 1997; Riess *et al.* 1997)
	Spinocerebellar ataxia 7 (SCA7)	*ATXN7*	3p12-13	exon	4-19	>33	(David *et al.* 1998)
	Spinocerebellar ataxia 12 (SCA12)	*PPP2R2B*	5q32	5' UTR	4-32	>42	(Holmes *et al.* 1999; Fujigasaki *et al.* 2001)

	Spinocerebellar ataxia 17 (SCA17)	*TBP*	6q27	exon	25-40	>40	(Nakamura 2001)
CTG	Fuchs endothelial corneal dystrophy (FECD)	*TCF4*	18q21.2	intron	10-37	>50	(Wieben *et al.* 2012)
	Huntington's Disease-like 2 (HDL2)	*JPH3*	16q24.2	3' UTR	6-28	40-59	(Holmes *et al.* 2001)
	Myotonic dystrophy 1 (DM1)	*DMPK*	19q13.3	exon	5-37	>50	(Brook *et al.* 1992)
	Spinocerebellar ataxia 8 (SCA8)	*ATXN8*	13q21.3	3' UTR	15-37	>74	(Moseley *et al.* 2000; Ikeda *et al.* 2004)
GAA	Friedreich ataxia (FRDA)	*FXN1*	9q21.11	exon	5-30	>70	(Campuzano *et al.* 1996)
CGG	FRA7A	ZFN713	7p11.2	intron	5-22	>68	(Metsu *et al.* 2014)
	Fragile X syndrome (FXS)	*FMR1*	Xp27.3	5' UTR	6-55	>200	(Yu *et al.* 1991)

Table 1.1 Trinucleotide Repeat Diseases

1.4 Genetic modifiers of Huntington's Disease

Soon after the identification of the disease-causing mutation in HD, the inverse relationship between the length of the *HTT* CAG repeat and the age at onset was also discovered (Andrew *et al.* 1993; Kremer *et al.* 1993; MacDonald *et al.* 1993). In fact, the length of the *HTT* CAG repeat is the primary determinant for the age at which an individual presents motor symptoms. Despite this relationship, there is considerable variation in motor onset that cannot be explained by CAG repeat length alone. The length of the expanded *HTT* CAG repeat is thought to account for around 56% of the variance in age at onset in individuals with 40-55 *HTT* CAG repeats, suggesting the remaining variance could be influenced by genetic, environmental or stochastic effects (Craufurd and Dodge 1993; Wexler 2004). Evidence from family and sibling studies revealed that the unexplained variation in age of onset is largely heritable (Squitieri *et al.* 2000; Li *et al.* 2003; Djoussé *et al.* 2004). Corroborating these findings are those from studies with the largest HD kindred in the world; the Venezuelan kindred (Wexler 2004). Using heritability studies with the Venezuelan kindred, it was estimated that around 40% of the variation that could not be explained by the length of the CAG repeat was attributable to genetic variation elsewhere in the genome (Wexler 2004).

1.4.1 Genome Wide Association Studies

Capitalising on advancements in technology, unbiased, genome-wide association studies (GWAS) can be performed with these samples to identify novel genetic variants that might modify disease onset and progression. These disease-modifying variants therefore represent genes or pathways that could be targeted therapeutically to prevent or delay disease onset, or slow disease progression. GWAS studies have successfully been used to identify risk variants in a range of disorders including coronary heart disease, type 1 diabetes, Alzheimer's disease and Schizophrenia (Visscher *et al.* 2017).

Global efforts have since enabled the collection of DNA and clinical information from a large number of HD subjects. Individuals were ascertained from a number of studies: HSG PHAROS (Shoulson 2006), COHORT (Dorsey 2012), TREND-HD (Huntington Study Group TREND-HD 2009), PREDICT-HD (Paulsen *et al.* 2008), Registry (Orth *et al.* 2010) and Enroll-HD (Landwehrmeyer *et al.* 2017). The initial HD-GWAS comprised individuals with 40-53 *HTT* CAG repeats. For each individual a residual value was calculated based on the age at motor onset predicted by their *HTT* CAG repeat length and their actual age at motor onset. This residual value therefore represented the difference in years between the persons age at clinical onset and predicted age at onset and was used as a phenotype in the GWA analysis (GeM-HD Consortium 2015). The primary GWAS conducted in HD consisted of three data sets (GWA 1-3) and comprised 4,082 individuals (GeM-HD Consortium

2015), whilst the latest GWAS consisted of 9037 subjects (GeM-HD Consortium 2019). The first GWAS identified two independent loci at a chromosome 15 locus containing Fanconi Anaemia-associated nuclease 1 (*FAN1)* and myotubularin related protein 10 (*MTMR10)* and a single effect at a chromosome 8 locus containing ribonucleotide-diphosphate reductase subunit M2 B (*RRM2B*) and ubiquitin-protein ligase E3 component N-recognin 5 (*UBR5)* which modified the age of onset of HD. The signals at the chromosome 15 locus were replicated in an independent cohort of 1,462 individuals with HD (Bettencourt *et al.* 2016). Pathway analysis revealed DNA repair pathways were also significantly associated with age of onset (GeM-HD Consortium 2015) and follow-up studies identified a genome-wide significant variant on chromosome 3 containing Mut L homolog 1 (*MLH1*) (Lee *et al.* 2017). These findings were striking as DNA repair machinery had been implicated in disease pathogenesis in a number of mouse models of HD, and *MLH1*, a mismatch repair gene, had been identified as a modifier of somatic instability and disease-phenotypes in HD knock-in mice (Kovtun and McMurray 2001; McMurray 2010; Pinto *et al.* 2013). The latest GWAS not only replicated the genome-wide significant signals at chromosome 15 and chromosome 8, but also identified several other loci that modify the age of onset in HD (GeM-HD Consortium 2019). Many of these loci included DNA repair genes; Mut S homolog 3 (*MSH3), MLH1*, postmeiotic segregation increased 1 (*PMS1),* postmeiotic segregation increased 2 (*PMS2)* and DNA ligase 1 (*LIG1).* The *HTT* gene itself was associated with an altered age at onset, with the uninterrupted length of the CAG tract, rather than the length of the polyglutamine tract driving onset (GeM-HD Consortium 2019).

Variants in *FAN1* and *PMS2* were also significantly associated with an altered age at onset of a number of SCAs, suggesting that a common pathogenic mechanism at the level of the DNA might influence disease onset and progression (Bettencourt *et al.* 2016). Excitingly, the modifiers identified from the human genetic studies largely aligned with those previously identified in model systems of HD. DNA repair pathways have been implicated in HD pathogenesis for over two decades, and it is thought that they might modify disease onset through somatic expansion of the *HTT* CAG repeat.

1.4.2 FAN1

To date, four independent modifier single nucleotide polymorphisms (SNPs) have been identified at the chromosome 15 locus containing *FAN1* and *MTMR10*. Protein-changing and expression-associated variation suggest that *FAN1* is the likely driver of these signals. The modifier SNPs at the chromosome 15 locus are bi-directional. Two of the SNPs are associated with an earlier age at onset. These are coding SNPs for putative damaging variants in *FAN1* and likely result in a loss of function of FAN1. The most significant SNP, rs150393409 ($p= 1.8 \times 10^{-28}$) is in linkage disequilibrium (r^2 =0.72) with a SNP encoding an amino acid change from arginine to histidine at codon 507

(R507H) at the C terminal end of FAN1 DNA binding domain and is associated with 5.2 years earlier motor onset. Conditional analysis suggests R507H is the pathogenic variant. A second coding genome wide significant SNP on chromosome 15, rs151322829 ($p= 1.4 \times 10^{-8}$) tags a SNP encoding an arginine to tryptophan missense variant at codon 377 (R377W) in FAN1 DNA binding domain. Conversely two SNPs are associated with a delayed age at onset and these correspond with cis-eQTL (expression quantitative trait loci) SNPs for increased FAN1 expression in cortical tissue. These data indicate that increased FAN1 expression might be protective.

FAN1 is a 5' endonuclease and a 5'-3' exonuclease involved in interstrand cross-link repair and replication fork metabolism and as such maintains chromosomal stability (Kratz *et al.* 2010; MacKay *et al.* 2010; Chaudhury *et al.* 2014; Lachaud *et al.* 2016). FAN1 was initially identified as a protein interactor of mismatch repair proteins MLH1 and PMS2 (MutLα) (Cannavo *et al.* 2007), though the function of this interaction is still not known. FAN1 was later identified as a member of the Fanconi anaemia pathway with a role in interstrand cross-link repair (Liu *et al.* 2010; MacKay *et al.* 2010), though unlike other genes in the pathway, its loss does not cause Fanconi anaemia. Loss of FAN1 causes a recessive renal syndrome, karyomegalic interstitial nephritis, in humans and mice, and is characterised by renal fibrosis, tubular dilation and degeneration and polyploidy (Zhou *et al.* 2012; Airik *et al.* 2016). The functions of FAN1 are discussed further in section 1.6.2.4.

1.5 CAG repeat instability

The expanded *HTT* CAG repeat is unstable, undergoing progressive length increases over time in both germ cells and somatic cells. Variations in trinucleotide repeat lengths have been described in human post-mortem brain tissue and mouse models of various triplet repeat disorders; HD, SCAs, DM and FXD (Nagafuchi *et al.* 1994; Kennedy *et al.* 2003; De Temmerman *et al.* 2004; De Biase *et al.* 2007; Gonitel *et al.* 2008). These expansion events are thought to be of great importance, as repeat length variability affects phenotype, and it is therefore likely that changes in repeat length might contribute to pathogenesis.

1.5.1 Germline Expansions

Many of the trinucleotide repeat diseases demonstrate anticipation, a phenomenon in which the age of disease onset becomes progressively earlier, and disease symptoms become more severe in successive generations. Anticipation is influenced by the sex of the affected parent, and in HD, anticipation commonly occurs through paternally transmitted alleles, whilst the offspring off affected mothers tend to demonstrate a similar age at onset (Ridley *et al.* 1988; Andrew *et al.* 1994; Koshy and Zoghbi 1997). With the discovery of the disease-causing mutation, it was evident that a

change in *HTT* CAG repeat length between generations was responsible for anticipation (Andrew *et al.* 1993; MacDonald *et al.* 1993). Intergenerational instability of the *HTT* CAG repeat refers to expansions or contractions of the repeat upon transmission. Intergenerational instability of the *HTT* CAG repeat is not uncommon, occurring in 70-80% of cases, but is usually only altered by one or two *HTT* CAGs, often showing contractions if maternally transmitted, and expansions if paternally transmitted (Duyao *et al.* 1993). However, anticipation is most evident in juvenile HD cases, 80% of which are paternally transmitted, where substantial expansions result in a considerably earlier age at onset than the transmitting parent (Telenius *et al.* 1994). Large intergenerational expansions have been documented in humans, the largest being an expansion of 160 *HTT* CAGs from a father with 54 *HTT* CAGs to 214 *HTT* CAGs in the offspring (Seneca *et al.* 2004). Maternally transmitted expansions are also documented, occurring for up to 20% of JHD cases (Nahhas *et al.* 2005). Though some maternally transmitted expansions are large (> 60 *HTT* CAGs) (Nahhas *et al.* 2005), these tend to be rare and expansions seldom exceed 20 *HTT* CAGs (Sánchez *et al.* 1997). Whilst *de novo* expansion of the *HTT* CAG repeat is rare, sporadic cases of HD result from the expansion of intermediate alleles with *HTT* CAG repeat lengths in the range of 30-37 (Myers *et al.* 1993; Telenius *et al.* 1994). Intergenerational instability is also documented in non-human primates (Putkhao *et al.* 2013) and mouse models of HD, although it occurs at a much lower frequency in mouse-models (2%) when *HTT* CAG repeat lengths are in the common pathogenic range (Wheeler *et al.* 1999). Intergenerational instability in mouse models of HD is increased in a length-dependent manner, where instability occurs at a higher frequency in mice with *HTT* CAG repeats of 90 and 109 than those with 72 (Wheeler *et al.* 1999). Interestingly, in mouse models, female transmissions showed higher levels of instability than male transmissions, but these were largely contractions, whilst male transmissions were largely expansions (Wheeler *et al.* 1999).

Intergenerational instability of the *HTT* repeat length upon paternal transmission is largely reflective of the variability in CAG repeat length in the sperm of males with HD. The length of the *HTT* CAG repeat tract in sperm demonstrated considerable variation in the range of *HTT* CAG repeat lengths, and on average, the repeat length was higher than that in blood (Duyao *et al.* 1993). Single sperm analysis revealed expansion events in 92-99% of sperm cells with an expanded allele. Interestingly, expansions were also observed on 42% of intermediate alleles, yet contractions were observed on alleles in the 15-18 *HTT* CAG repeat range (Leeflang *et al.* 1995). Longer repeat tracts showed a higher propensity for repeat expansions, and demonstrated greater mosaicism of repeat lengths, suggesting that constitutive repeat length is a determinant of repeat instability in sperm (Wheeler *et al.* 2007). The findings that sperm from male HD carriers demonstrated *HTT* CAG repeat variability suggests that repeat instability occurs during spermatogenesis and paternal

transmission. Examination of offspring pairs in a large HD kindred revealed significant sibling correlation, suggesting that instability is likely governed by genetic factors (Wheeler *et al.* 2007).

1.5.2 Somatic Expansion

The *HTT* CAG repeat is also somatically unstable. Post-mortem tissue from individuals affected with HD shows considerable expansion and repeat mosaicism, particularly in the brain (Telenius *et al.* 1994; De Rooij *et al.* 1995; Kennedy *et al.* 2003; Shelbourne *et al.* 2007). Mosaicism is most obvious in the regions of the brain most susceptible to disease pathogenesis, such as the basal ganglia and cerebellar cortex, perhaps contributing to their selective degeneration. Whilst expansions increase with age (Dragileva *et al.* 2009), the length of the repeat tract appears to be the primary driver, as individuals with juvenile onset HD demonstrate significantly higher levels of mosaicism in the basal ganglia than those with adult-onset HD (Telenius *et al.* 1993). While most areas of the brain exhibit varying degrees of mosaicism, the cerebellum is largely spared, mirroring HD neuropathology (Telenius *et al.* 1994). Striatal neurons from brain tissue of late-stage HD cases show little somatic instability, potentially owing to the extensive neuronal loss in the striatum at death. Indeed, if somatic instability precedes and drives neuronal degeneration, it is likely that neurons exhibiting the greatest instability are the most susceptible to cell death. Larger studies have determined that increased levels of somatic instability in cortical tissue from HD patients correlate with an earlier age at onset than that predicted by the length of their *HTT* CAG repeat tract (Swami *et al.* 2009). Establishing cause and effect of somatic expansions are difficult, as findings are largely based on post-mortem brain tissue, which is only available after death, and usually demonstrates considerable atrophy. Post-mortem brain tissue from individuals who passed away from an unrelated cause, can therefore shed light on these findings. Small-pool PCR analysis of post-mortem HD brain tissue with no atrophy revealed expansions greater than 1000 *HTT* CAG in striatal and cortical samples, suggesting that somatic expansion might precede disease onset (Kennedy *et al.* 2003). Whilst these results are promising, the availability of such tissue is rare, and these findings are based on limited samples.

Similar findings have been described in mouse models of HD. Transgenic mice containing exon 1 of the human *HTT* gene with $(CAG)_{115}$-$(CAG)_{155}$, R6/1, R6/2 and R6/5, demonstrate intergenerational and somatic repeat instability (Mangiarini *et al.* 1996; Mangiarini *et al.* 1997). Somatic instability has also been documented in knock-in mouse models, where an expanded repeat is inserted into the endogenous *HTT* locus: Hdh^{Q72}, Hdh^{Q80} (Kennedy 2000), Hdh^{92} and Hdh^{111} (Wheeler *et al.* 1999), Hdh^{Q150} (Gonitel *et al.* 2008). Expansions have been documented in a number of tissues including

the brain, heart, spleen, kidneys and liver, though somatic expansions in mouse models also show tissue-specificity, and the greatest expansions are evident in the in the striatum, cortex and liver (Mangiarini *et al.* 1997; Manley, Pugh, *et al.* 1999; Lee *et al.* 2011). Tissue-specificity is not confined to HD; in DM1 patients, considerably larger CTG repeat expansions were demonstrated in skeletal muscle, cardiac muscle, brain, liver and testes compared with peripheral blood lymphocytes (Peterlin *et al.* 1996). Additionally, post-mortem brains of individuals with DRPLA had longer repeat lengths in the Purkinje cells and glia than granule cells (Ueno *et al.* 1995; Watanabe *et al.* 2000). Further evidence for the involvement of somatic expansions in disease pathogenesis comes from mouse models, when suppression of repeat expansion alleviates disease symptoms (Budworth *et al.* 2015). These tissue-specific differences are therefore thought to contribute to disease pathogenesis, so understanding the factors governing repeat expansions is of great interest.

1.5.3 CAG Repeat Structure

The propensity for repeats to expand and contract is thought to occur due to the non-helical structures that expanded repeats adopt. Indeed, it has been demonstrated that repetitive sequences can adopt several different conformations including hairpin loops, unpaired loops, slipped-DNAs, G-quadruplexes and sticky DNA (Mirkin and Frank-Kamenetskii 1994). The structures formed and their stability largely depend on the sequence of the repeat and their length, but both cis- and trans- modifiers influence structure (Usdin and Grabczyk 2000; McGinty and Mirkin 2018). All repetitive sequences have the potential to form slipped, mispaired DNA which enables the formation of an unpaired or hairpin loop on either strand. The formation of these structures usually occurs transiently following unwinding of the DNA double helix during DNA replication, repair, recombination or transcription. Single stranded CAG repeats show a propensity to form imperfect hairpins comprised of Watson-Crick base pairing (C=G) stabilising the structure, with bulky mismatched nucleotides (A●A/ T●T) along the stem (Figure 1.3 DNA structures that may form as a result of an expanded CAG repeat).

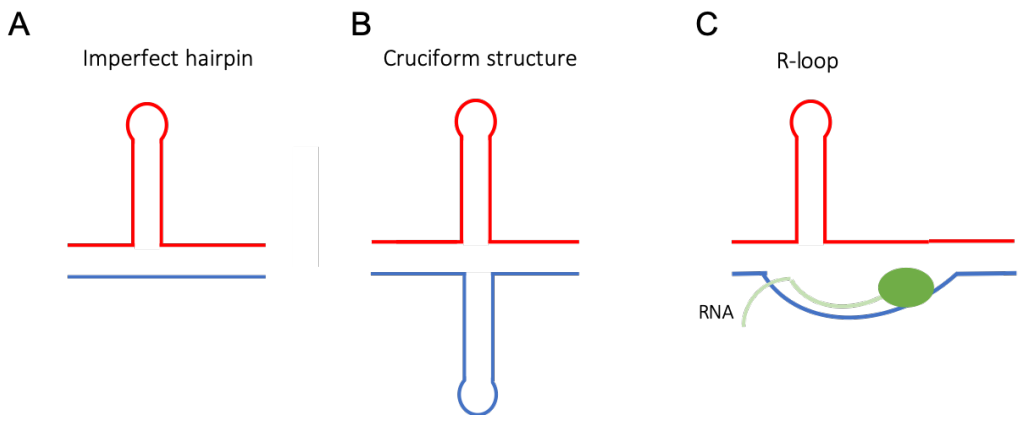


Figure 1.3 DNA structures that may form as a result of an expanded CAG repeat. (A) An imperfect hairpin in which Watson-Crick G-C base pairing stabilises the hairpin, but intervening bases are not paired. (B) A cruciform structure consisting of a hairpin structure on both strands (C) R-loop formation during transcription exposes a single strand and predisposes it to hairpin formation.

Slipped DNA structures are the intermediates for expansion and contraction. In DNA replication, if the single-stranded looped structures occur on the leading strand, contractions are thought to occur, conversely, if they occur on the lagging strand, the looped structures are thought to be incorporated into the DNA and result in expansions (Mirkin 2007; McMurray 2010; McGinty and Mirkin 2018). These structures are more likely to occur on the newly synthesised lagging strand, as the strand remains single-stranded until Okazaki fragments are synthesised. During transcription, R-loops may transiently form; stretches of RNA-DNA hybrids, which temporarily result in a single-stranded non-template DNA strand. CAG repeats are guanine-rich and show a propensity to form R-loops (Sugimoto *et al.* 1995; Nadel *et al.* 2015). The persistence of R-loops is thought to increase repeat instability as they expose single-stranded DNA, rendering it more susceptible to DNA damage.

Longer repeat tracts demonstrate increased rates of expansion, likely owing to the increased energetic stability of hairpins formed from longer repeats (Marquis Gacy *et al.* 1995). Interruptions in the repeat tract disturb the stability of these extrahelical structures, and hence fewer repeat expansions are seen. These findings are consistent with those in human studies, whereby individuals with interruptions tend to have a later onset than predicted by their CAG repeat length alone (GeM-HD Consortium 2019; Wright *et al.* 2019).

1.6 Mechanisms of Repeat Instability

In theory repeat expansions can occur at any point that the DNA becomes unpaired, thereby implicating replication, repair and transcription.

1.6.1 Replication mediated repeat instability

Trinucleotide repeat instability was first thought to occur during replication. Studies conducted in simple organisms; bacteria and yeast, revealed instability occurred due to replication slippage on either the leading or lagging strand (Schweitzer and Livingston 1999; Usdin and Grabczyk 2000). During replication, due to the similarity of the trinucleotide repeat, nucleotides can misalign, generating an extra-helical loop which may then be incorporated into the duplex DNA by a DNA polymerase (Schweitzer and Livingston 1997; Schweitzer and Livingston 1999; Kovtun and McMurray 2006). If the extrahelical loop occurs on the template strand, contractions are thought to occur and if it occurs on the daughter strand then expansions may occur. Interestingly, yeast and bacteria show a propensity towards repeat contractions, whereas mammalian cells show a propensity towards expansions. DNA replication may underlie intergenerational repeat expansions (McMurray 2010).

Although DNA replication might contribute to repeat instability, it's likely limited to early development or cell populations still able to proliferate, glial cells for example (Watanabe *et al.* 2000; Claassen and Lahue 2007; Shelbourne *et al.* 2007). However, repeat expansions occur in neurons, which are post-mitotic and do not replicate their DNA. Somatic expansions in neurons correlate well with disease onset, and suppression of repeat expansions alleviates disease phenotypes, suggesting that repeat expansions in non-dividing cells are likely to be relevant to the disease (Swami *et al.* 2009; Budworth *et al.* 2015). These data suggest that the mechanisms underlying repeat expansions are not solely division dependent.

1.6.2 DNA repair mediated repeat instability

DNA repair has long been implicated in repeat instability in HD and other trinucleotide repeat diseases, and there are a number of DNA repair proteins that appear essential for repeat expansions (Kovtun and McMurray 2006; McMurray 2010; McGinty and Mirkin 2018). The DNA damage response is crucial for the maintenance of genomic stability and as such we have evolved intricate DNA repair pathways. If left unrepaired, DNA damage contributes to cancers, ageing and neurodegenerative diseases (Roos *et al.* 2016). The DNA damage response comprises a number of different DNA repair pathways, responsible for resolving a range of DNA lesions. The main repair pathways in the DNA damage response are mismatch repair, base excision repair, translesion DNA synthesis, nucleotide excision repair, Fanconi anaemia, non-homologous end joining and homologous recombination. Many elements of the DDR are highly conserved from prokaryotes to eukaryotes and many of the pathways demonstrate functional redundancy.

1.6.2.1 Mismatch Repair

The mismatch repair (MMR) pathway was first implicated in repeat instability in bacteria (Jaworski *et al.* 1995; Bichara *et al.* 2000) and yeast (Schweitzer and Livingston 1997), with insight from findings of microsatellite instability and tumour development in individuals with MMR defects (Umar *et al.* 1994; de Wind *et al.* 1995; Kobayashi *et al.* 1996; Reitmair *et al.* 1997). In yeast and bacteria, the loss of MMR proteins increases repeat instability, whereas in mouse and human cell models, repeat expansions are dependent on proteins in the MMR pathway (Wheeler *et al.* 1999; Wheeler *et al.* 2003; Pinto *et al.* 2013). Functional research looking at the role of MMR in trinucleotide repeat disease spans over two decades, and their identification as potential genetic modifiers of HD pathogenesis highlights the importance of their contribution to disease. The MMR pathway is an evolutionarily highly conserved pathway, maintaining genomic integrity by detecting and repairing mis-paired bases and small (15 bp) insertion/deletion loops (Fishel and Wilson 1997). Functional MMR is a multi-step process involving the detection of mispaired DNA, excision along the erroneous nascent strand, exonuclease nucleotide removal, gap filling by polymerase and nick ligation (Hsieh and Yamane 2008).

In eukaryotes, two complexes are involved in mismatch recognition: MutSα and MutSβ. MutSα is a heterodimer comprised of MutS homolog 2 (MSH2) and MutS homolog 6 (MSH6), and MutSβ is another heterodimer comprised of MSH2 and MutS homolog 3 (MSH3). Whilst the binding specificity of the two complexes overlaps, it is not identical. MutSα recognises single base mismatches and small insertions or deletions whilst MutSβ tends recognise larger insertions and deletions (Fishel and Wilson 1997; Gradia *et al.* 1999). *In vivo*, repeat expansions appear to require MutSβ; knock-out of Msh2 or Msh3 attenuates both germline and somatic *HTT* CAG repeat expansions (Manley, Shirley, *et al.* 1999; Wheeler *et al.* 1999; Wheeler *et al.* 2003; Owen *et al.* 2005; Dragileva *et al.* 2009; Tomé *et al.* 2013). The effects of MSH6 are variable in HD models, with reports of a protective effect (Dragileva *et al.* 2009), or a small, tissue-specific role in expansions (Owen *et al.* 2005). Knockdown of MSH6 in mouse models of induced pluripotent stem cells from FRDA patients potentiates expansions however (Du *et al.* 2012). MutSβ binds to CAG hairpin loops, and upon their binding certain mismatch recognition properties change (Owen *et al.* 2005). Additionally, it is thought that CAG hairpin loops demonstrate a unique conformational population which, when bound to MutSβ, prevents its dissociation from the repeat (Lang *et al.* 2011).

Two MMR complexes that act downstream of MutS; MutLα and MutLγ, have also been implicated in *HTT* CAG repeat expansions (Pinto *et al.* 2013; Gazy *et al.* 2019). MutLα, a heterodimer of MutL homolog 1 (MLH1) and postmeiotic segregation increased 2 (PMS2) and MutLγ, a heterodimer of MLH1 and MutL homolog 3 (MLH3) are recruited by MutS and are primarily involved in the

activation of endonucleases, and meiotic recombination, respectively (Jiricny 2006a; Jiricny 2006b). Loss of Mlh1 and Mlh3 abolishes *HTT* somatic expansions, however no effect on intergenerational expansions could be deciphered as Mlh1/ Mlh3 mice are sterile (Pinto *et al.* 2013). These data suggest that the mechanism governing CAG repeat expansions involves both MutS complexes and MutL complexes. Consistent with this, in a mouse model of DM1, disruption of the ATPase function of MSH2, which is required for recruitment of MutL complexes, reduced expansions (Tomé *et al.* 2009). Exonuclease 1 (EXO1) is an endonuclease which acts downstream of MutL proteins, whose loss exacerbated both somatic and intergenerational repeat expansions in a mouse model of fragile X (Zhao *et al.* 2018).

1.6.2.2 Base Excision Repair

The base excision repair pathway is responsible for the removal of chemically damaged bases, including oxidised bases. Oxidative damage accumulates in mouse models of HD in a CAG-repeat-length dependent manner (Bogdanov *et al.* 2001; Goula *et al.* 2012). Somatic expansions increase in both a length- and age-dependent manner, and evidence suggests that expansions may occur through the removal of oxidised bases (Kovtun *et al.* 2007). 8-oxoG DNA glycosylase (OGG1) is the primary glycosylase involved in removing the oxidised base via base excision repair. The loss of OGG1 reduced the rate of *HTT* CAG repeat expansions in a mouse model of HD (Kovtun *et al.* 2007) and knock-out of OGG1 delayed HD onset and progression (Budworth *et al.* 2015). Though it should be noted that the effect size was small and required over 1000 mice to achieve a significant result. Additionally, exposure to oxidising agents such as potassium bromate ($KBrO_3$) induced *Fmr1* CGG repeat expansions in a mouse model of Fragile X (Entezam *et al.* 2010) and exposure of HD fibroblasts to hydrogen peroxide also potentiates *HTT* CAG repeat expansion (Kovtun *et al.* 2007). The loss of nei endonuclease VIII like 1 (NEIL1) also reduced somatic expansions in a mouse model of HD but did not have any effect on intergenerational expansions (Møllersen *et al.* 2012). It appears the effects of base excision repair proteins on somatic expansion are gene-specific, as loss of alkyladenine glycosylase (AAG) or human endonuclease III-like protein 1 (NTH1) had no effect on *HTT* CAG somatic expansions (Kovtun *et al.* 2007). Notably there is no genetic signal of *OGG1* or *NEIL1* in the HD GWAS.

Due to the abundance of guanine residues in the repeat tract (i.e. CAG, CGG, CTG), guanines are preferentially oxidised to 7,8-dihyrdro-8-oxoguanine (Jarem *et al.* 2009). It is thought that base excision repair might mediate repeat expansions in non-dividing cells through the removal of an oxidised base which creates a single strand break, followed by endonuclease processing and subsequent polymerase gap filling (Parsons and Edmonds 2015). Gap-filling may occur by long patch or short patch base excision repair mechanisms. Long patch base excision repair might

incorporate the single strand loops into the repair patch leading to an expansion (McMurray 2010). Consistent with this, heterozygous loss of DNA polymerase (pol) β, a DNA polymerase in base excision repair, reduced rates of somatic expansion in the brain and tail and reduced the number of expanded alleles detected in sperm, in a mouse model of fragile X disorder (Lokanga *et al.* 2015).

1.6.2.3 Nucleotide Excision Repair

The nucleotide excision repair pathway has also been implicated in CAG repeat expansions, but there is limited research in the context of HD. Nucleotide excision repair comprises two pathways; global genome repair and transcription coupled repair (Hanawalt *et al.* 2000). Loss of global genome repair protein, complementation group C (Xpc) from a mouse model of HD had no effect on intergenerational or somatic instability of the *HTT* CAG repeat (Dragileva *et al.* 2009). Variants in transcription-coupled repair genes, Cockayne syndrome protein B (*CSB*), replication protein A3 (*RPA*) and cyclin-dependent kinase 7 (*CDK7*) were associated with an intergenerational expansion bias in SCA3 (Martins *et al.* 2014). CSB also promotes both intergenerational and somatic expansions in a mouse model of Fragile-X disorder (Zhao and Usdin 2014). CSB might influence repeat expansions indirectly through the base-excision repair pathway; CSB expression influences that of OGG1 (Javeri *et al.* 2011), and CSB also stimulates the repair activities of both OGG1 (Tuo *et al.* 2002), NEIL1 (Muftuoglu *et al.* 2009; Aamann *et al.* 2014) and APE1 (Wong *et al.* 2007).

1.6.2.4 Fanconi Anaemia pathway

Interstrand cross-links occur when two nucleotides on complementary strands become covalently linked, and are particularly toxic as they prevent unwinding of the DNA and therefore interfere with replication and transcription. Failure to resolve interstrand cross-links results in replication fork collapse and increased double strand breaks and genomic instability (Noll *et al.* 2006). The importance of FAN1 in interstrand cross-link repair has been demonstrated in a number of studies as loss of FAN1 increases the vulnerability of human cells and mice to interstrand cross-linking agents such as mitomycin C (Kratz *et al.* 2010; Zhou *et al.* 2012; Thongthip *et al.* 2016). Additionally, mice lacking FAN1 demonstrate karyomegalic nuclei in the kidney and liver at 6 months of age (Thongthip *et al.* 2016). These data demonstrate the importance of FAN1 for the survival of animals exposed to cross-linking damage.

An important event in the resolution of an interstrand cross-links is the ubiquitylation of the FANCI-FAND2 (ID) complex that recruits and coordinates the relevant Fanconi anaemia proteins to the site of DNA damage (Smogorzewska *et al.* 2010). However, it is now believed that FAN1 is able to resolve cross-links in a pathway independently of other Fanconi anaemia proteins (Jin and Cho 2017). FAN1 is a highly conserved nuclease with four functional domains: an N-terminal ubiquitin-

binding zinc finger (UBZ) domain that facilitates protein-protein interactions; a SAF-A/B, Acinus and PIAS (SAP) domain involved in DNA binding of nuclear proteins; a tetratricopeptide repeat (TPR) thought to be involved with dimerization and a C-terminal virus-type replication repair nuclease (VRR_NUC) domain which contains a highly conserved PD-D/EXK motif that confers endo- and exonuclease properties to FAN1 (Kratz *et al.* 2010; Liu *et al.* 2010; MacKay *et al.* 2010). Several crystal structures of FAN1 bound to DNA have been reported. Despite these data the mechanism of FAN1 in ICL repair is not completely clear. The nuclease activity of FAN1 suggests that it could participate in many steps of the interstrand cross-link repair pathway including interstrand crosslink unhooking or incision (Jin and Cho 2017).

Though DNA repair machinery had been implicated in HD pathogenesis previously, neither the Fanconi anaemia pathway nor FAN1 had been associated with HD or trinucleotide repeat disorders. Recently it was demonstrated that increased expression of FAN1 is associated with a later age of onset of HD and delayed progression (Goold *et al.* 2019), and it was postulated that FAN1 might directly modulate the length of the expanded repeat. FAN1 is structure-, rather than sequence-specific, and recognises branched structures (Ceccaldi *et al.* 2016; Jin and Cho 2017). Expanded CAG repeats adopt non-helical structures in DNA, and FAN1 could be recruited to such structures to participate in their processing. One study to date has looked at the effect of FAN1 on *HTT* CAG repeat expansions (Goold *et al.* 2019). Human bone osteosarcoma epithelial (U2OS) cells were stably transduced with HTT exon 1 containing 30, 70, 97 or 118 *HTT* CAGs. Knock-down of FAN1 in these cells significantly increased *HTT* CAG repeat expansion in U2OS with 97 and 118 *HTT* CAG repeats but had no effect on repeat lengths of 30 or 70 *HTT* CAGs. Overexpression of FAN1 in the U2OS cell line with 118 *HTT* CAG repeats further reduced repeat expansions. To establish whether the nuclease domain or the coding SNP identified in the HD-GWAS would influence repeat instability, FAN1 variants were introduced into the U2OS cell line with 118 *HTT* CAG repeats. Neither FAN1 nuclease inactivation nor the R507H variant had any effect on *HTT* CAG repeat expansions in the U2OS cells, suggesting that FAN1 exerts its protective effects in a nuclease-independent manner (Goold *et al.* 2019). In the same study, short-hairpin RNA (shRNA) was used to knock-down FAN1 in induced pluripotent stem cells derived from an individual with 109 *HTT* CAG repeats. Knock-down of FAN1 significantly increased *HTT* CAG repeat expansions in both dividing and post-mitotic MSNs, further suggesting a protective role of FAN1 in HD.

In a mouse model of Fragile X Syndrome, FAN1 deficiency significantly increased the rate of somatic repeat expansion compared to FAN1 wild-type mice, suggesting a protective effect of FAN1 against repeat expansions (Zhao and Usdin 2018). The loss of FAN1 was most notable in the kidney and the brain but had no effect on intergenerational expansions. Interestingly, in this study the expression

of FAN1 did not correlate well with the protective effect, since FAN1 expression was lower in the liver and relatively high in the testes.

Pathway	Gene	Model	Effect on expansion		Reference
			Somatic	Germline	
MMR	Msh2	$Hdh^{Q111/+}$: $Msh2^{-/-}$, R6/1:$Msh2^{-/-}$, HD lymphoblasts 45/47 *HTT* CAG	Ablation	Ablation	(Manley, Shirley, *et al.* 1999; Wheeler *et al.* 1999; Wheeler *et al.* 2003; Kovalenko *et al.* 2012; Lai *et al.* 2016)
	Msh3	$Hdh^{Q111/+}$: $Msh2^{-/-}$, R6/1:$Msh3^{-/-}$	Ablation	Reduced	(Dragileva *et al.* 2009; Tomé *et al.* 2013)
	Msh6	$Hdh^{Q111/+}$: $Msh6^{-/-}$	No effect	Variably increased contractions	(Dragileva *et al.* 2009)
	MLH1	$Hdh^{Q111/+}$: $Mlh2^{-/-}$	Ablation	Homozygous null mice are sterile	(Pinto *et al.* 2013)
	MLH3	$Hdh^{Q111/+}$: $Mlh3^{-/-}$	Ablation		(Pinto *et al.* 2013)
BER	OGG1	R6/1:$Ogg1^{-/-}$, $Hdh^{Q150/Q150}$:$ogg1^{-/-}$	Reduction	No effect	(Kovtun *et al.* 2007; Budworth *et al.* 2015)
	NEIL1	R6/1:$Neil1^{-/-}$	Reduction	No effect	(Møllersen *et al.* 2012)
	FEN1	R6/1:$Fen1^{+/-}$	No effect	Variably increased	(Spiro and McMurray 2003)

NER	XPC	$Hdh^{Q111/+}$:$Xpc^{-/-}$	No effect	No effect	(Dragileva *et al.* 2009)
FA	FAN1	HD iPSCs Q109:FAN1 knock-down U2OS 97, 118 *HTT* CAG: FAN1 knock-down	Increased		(Goold *et al.* 2019)

Table 1.2 Effect of DNA repair enzymes on CAG repeat expansions. Mismatch Repair (MMR), base excision repair (BER), nucleotide excision repair (NER), Fanconi anaemia (FA) pathway.

1.6.3 Transcription mediated repeat instability

Expansion in model systems is facilitated by transcription. In bacteria, inducible transcription of expanded repeat sequences demonstrated that repeat instability was increased following transcription induction (Bowater *et al.* 1997; Mochmann and Wells 2004). Instability depended on the length, orientation and sequence of the repeat sequences, perhaps suggesting that transcription facilitates the formation of secondary structures that act as expansion intermediates. Transcriptional studies in bacteria are difficult to interpret, as they are proliferating and therefore repeat instability might reflect both transcription and replication. Bacterial studies showed that repeat instability resulting from mutations in NER and MMR genes was greatly enhanced following the induction of transcription (Wierdl *et al.* 1996; Oussatcheva *et al.* 2001). In human cells it was demonstrated that CAG instability could be mediated by transcription, and that instability required components of MMR and NER (Lin *et al.* 2006). Furthermore, bidirectional transcription across repeat sequences potentiated repeat instability (Nakamori *et al.* 2011). CAG repeat stability was improved in human cells following exposure to RNaseH, an endonuclease involved in degrading the RNA in an R-loop, and thus removing R-loops (Lin *et al.* 2010). It is thus likely that transcription facilitates repeat instability through the formation of R-loops and the exposure of single-stranded DNA, allowing the formation of secondary structures and aberrant processing by DNA repair machinery.

Interestingly, there does not appear to be a correlation between mRNA levels of *HTT* and levels of somatic instability seen in the brain regions most affected (Goula *et al.* 2012). Transcriptional elongation might influence repeat instability however, as transcriptional elongation rates are higher in the striatum than in the cerebellum of mouse models of HD, and the striatum demonstrates higher levels of *HTT* CAG repeat instability (Goula *et al.* 2012).

1.7 Huntington's Disease Pathogenesis

1.7.1 The Huntingtin Protein (HTT)

HTT is ubiquitously expressed and is essential for human development and normal brain function, yet its role in cellular functioning is not clearly defined. The importance of HTT in embryonic development was established through the findings that complete loss of HTT is embryonic lethal, with death ensuing between 8.5 to 10 days of gestation (Duyao *et al.* 1995; Nasir *et al.* 1995). HTT is involved with a wide range of cellular processes including transcription, autophagy, endocytosis, vesicular transport, cell division, apoptosis, cytoskeletal structure/ function and the DNA damage response (Saudou and Humbert 2016; Maiuri *et al.* 2017). Intracellularly, HTT associates with a

number of organelles, including the nucleus, endoplasmic reticulum, mitochondria and the Golgi apparatus (Velier *et al.* 1998; Kegel *et al.* 2002), and in neurons it is found in neurites and synapses and has associations with vesicles and microtubules (DiFiglia *et al.* 1995; Hoffner *et al.* 2002). HTT contains 37 putative huntingtin, elongation factor 3, protein phosphatase 2A and TOR1 (HEAT) repeats, that are largely conserved in vertebrates (Andrade and Bork 1995; Takano and Gusella 2002). HEAT repeats indicate protein-protein interactions, and their frequent occurrence imply a large range of binding partners and suggesting HTT might act as a molecular scaffold to coordinate cellular processes (Harjes and Wanker 2003; Goehler *et al.* 2004; Li and Li 2004). The polymorphic polyglutamine tract is a key regulator of interactor binding (Harjes and Wanker 2003) and it is therefore presumed that HTT might adopt specific conformations depending on its binding partner. Proteins containing a polyglutamine region are often involved in transcriptional regulation (Whan *et al.* 2010), therefore the polyglutamine region might regulate HTT function through interaction with proteins involved in transcription (Zuccato *et al.* 2010).

The amino (N)-terminus of HTT has been studied extensively as it contains the expandable CAG repeat and therefore the polyglutamine tract. The first 17 amino acids in exon 1 of HTT (N-17) are thought to assume an alpha-helical structure with vesicular and endoplasmic reticulum binding capacities (Atwal *et al.* 2007; M. W. Kim *et al.* 2009). Highly conserved nuclear export signals have been found at both N-17 and at the carboxy (C)-terminus of HTT, indicating a role of HTT in protein shuttling to and from the nucleus (Xia 2003; Zheng *et al.* 2013). Disruption of N-17 or an expanded polyglutamine tract increases nuclear accumulation of HTT and subsequent mHTT toxicity (Cornett *et al.* 2005). N-17 is subject to a range of post-translational modifications including acetylation, sumoylation, ubiquitination, palmitoylation and phosphorylation (Steffan *et al.* 2004; Aiken *et al.* 2009; Thompson *et al.* 2009). These modifications are thought to regulate HTT localisation, function and interactions.

1.7.2 The mutant Huntingtin Protein (mHTT)

Due to the large interactome of HTT, resolving its function and therefore how mHTT leads to pathogenic effects is challenging, but it is thought that the expandable polyglutamine stretch at the N-terminal of the protein might alter the protein configuration or its interactions with other cellular proteins (Jou and Myers 1995). The effects of mHTT are complex and multifaceted, and a number of pathogenic mechanisms have been proposed for neuronal death in HD.

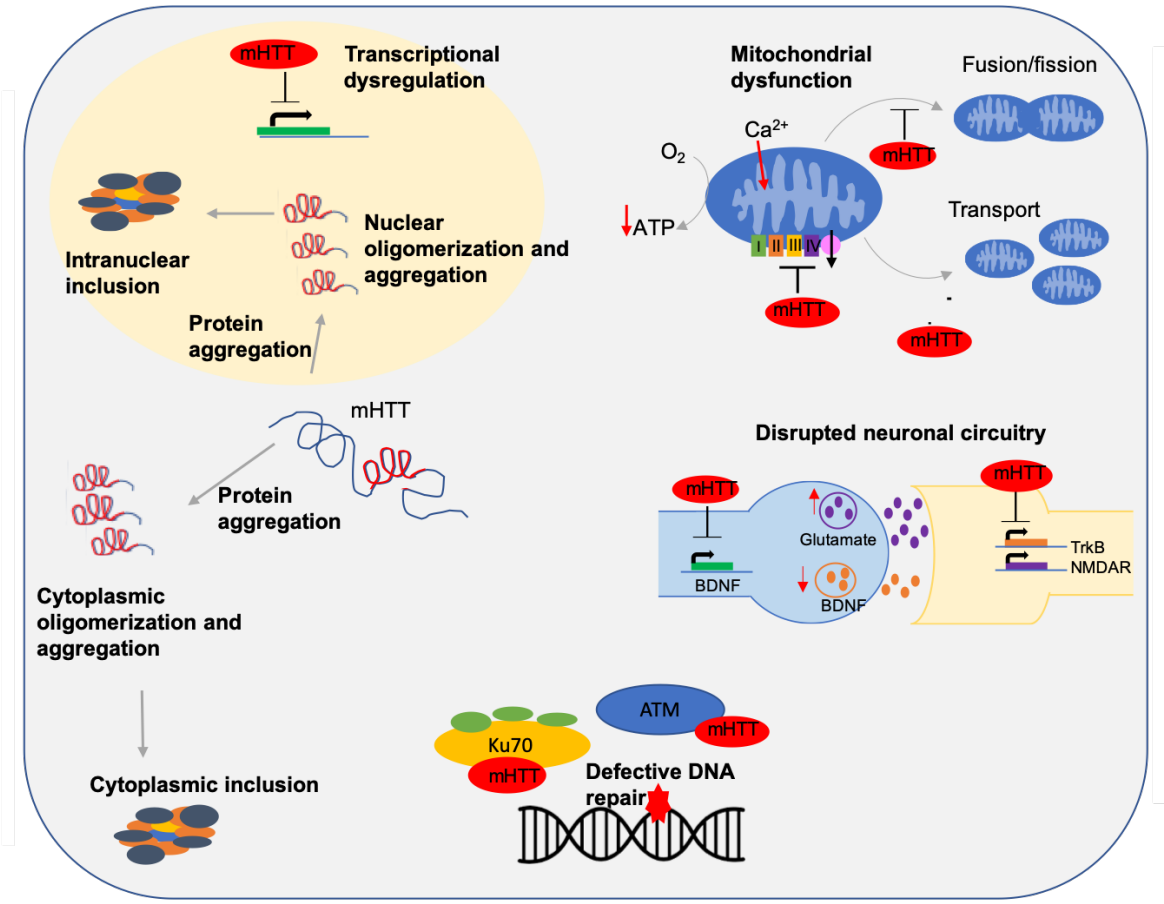


Figure 1.4 Pathogenic effects of mHTT.

1.7.2.1 mHTT fragments and inclusions

Proteolytic cleavage of HTT has been implicated in disease pathogenesis. HTT has a number of cleavage sites and several proteases including caspases, calpain and cathepsins have been demonstrated to cleave HTT (Wellington *et al.* 1998; Gafni and Ellerby 2002; Bizat *et al.* 2003; Gafni *et al.* 2004; Graham *et al.* 2006). Cleavage of mHTT results in fragments of both HTT and mHTT, though N-terminal fragments of mHTT promote neuronal toxicity (Cooper *et al.* 1998). HD brains show an upregulation of proteases which increases the generation of mHTT N-terminal fragments which tend to accumulate in the cytoplasm and nucleus (Sanchez Mejia and Friedlander 2001; Friedlander 2003). Studies using mouse models have demonstrated a correlation between N-terminal mHTT fragments and an increase in aggregates and disease-phenotype (Wang *et al.* 2008; Marcellin *et al.* 2012). Expression of mHTT N-terminus fragments in mouse models appear as toxic as full length mHTT and these mice demonstrate many of the disease-phenotypes HD mouse models expressing full-length mHTT show. As well as proteolytic cleavage of mHTT, N-terminal fragments of mHTT have been generated through the incomplete splicing of *HTT* RNA (Sathasivam *et al.* 2013). The incomplete splicing of *HTT* results in transcripts of exon 1 of *HTT* and these

transcripts are enriched in patient samples with juvenile onset HD compared with patient samples with adult-onset HD (Neueder *et al.* 2017). The translocation of mHTT fragments to the nucleus is thought to be cytotoxic, perhaps through their interference with transcription (Cooper *et al.* 1998; Ross and Tabrizi 2011). Toxicity of the expanded polyglutamine tract can be rescued in mouse models of HD following caspase-6 inhibitors (Saudou *et al.* 1998; Graham *et al.* 2006).

N-terminal fragments of mHTT are prone to misfolding and assemble into oligomers which may form insoluble aggregates (Scherzinger *et al.* 1997). Soluble oligomers of N-terminal mHTT have been shown to induce apoptosis, whilst insoluble fibrils inhibit apoptosis and promote necrosis (Ramdzan *et al.* 2017). Aggregates were thought to confer toxicity to the cell via disruption of transport machinery due to their large size (Li *et al.* 2001; Lee *et al.* 2004; Woerner *et al.* 2016), or through the sequestration of proteins critical to cell survival (Schaffar *et al.* 2004; Park *et al.* 2013). However, there is evidence to suggest that these aggregates are in fact neuroprotective and reduce the risk of neuronal death (Arrasate *et al.* 2004; Bodner *et al.* 2006). Oligomers appear particularly cytotoxic and so it's assumed that the formation of aggregates might be protective as they reduce levels of the toxic soluble mHTT fragments (Takahashi *et al.* 2008; Lajoie and Snapp 2010; Leitman *et al.* 2013).

1.7.2.2 mHTT is a transcriptional repressor

mHTT has also been implicated in abnormal transcription (Krainc 2006; Cha 2007; Park et al. 2018; Faragó et al. 2019). Global gene expression studies have revealed considerable downregulation of transcription in the HD brain (Hodges et al. 2006) and cellular models of HD (An et al. 2012; The HD iPSC Consortium 2012; Xu et al. 2017). Transcriptional changes are most obvious in the striatum (Tabrizi et al. 2013), and many of the dysregulated genes include those relating to synaptic function in MSNs and inflammation (Seredenina and Luthi-Carter 2012; Bragg et al. 2017; Peng et al. 2018). HTT interacts with many transcriptionally active proteins including: nuclear receptor co-repressor (NCoR) (Boutell et al. 1999), CREB (cAMP response element-binding protein)-binding protein (CBP) (Nucifora et al. 2001), tumour protein 53 (p53), and repressor element 1-silencing transcription factor (REST) (Zuccato et al. 2003; Chen et al. 2017). The interaction of mHTT with p53 inhibits the transcription of genes regulated by p53, thereby preventing or altering an appropriate stress-response (Steffan et al. 2000). Additionally, mHTT binds to the promoter of peroxisome proliferator-activated receptor gamma coactivator-1α (PGC-1α), a transcriptional coactivator involved in metabolic processes including mitochondrial biogenesis and mitochondrial respiration (Cui *et al.* 2006; Finck and Kelly 2006). Reduced expression of PGC-1α is thought to contribute to mitochondrial dysfunction. Expression of PGC-1α is drastically reduced in the caudate, specifically

MSNs, relative to other neuronal populations, suggesting that it may contribute to their selective vulnerability (Cui *et al.* 2006; Weydt *et al.* 2006; Chaturvedi *et al.* 2010; McMeekin *et al.* 2018).

mHTT is also implicated in dysregulation of brain-derived neurotrophic factor (BDNF). HTT promotes transcription and transport of BDNF, a neurotrophin produced by cortical neurons that is critical for striatal neuron survival (Baydyuk and Xu 2014). BDNF mediates glutamate release at the cortico-striatal synapse and, BDNF promotes the differentiation and survival of striatal neurons in vitro. The promoter region of BDNF contains repressor element 1 (RE1) sites, the binding of REST to such sites prevents transcription. HTT has been found to sequester REST in the cytoplasm, thereby preventing its recruitment to RE1 sites in the nucleus and increasing transcription of *BDNF* (Zuccato *et al.* 2001; Zuccato *et al.* 2003). In HD, mHTT permits nuclear entry of REST where it accumulates and prevents the transcription of *BDNF* and a number of other genes. BDNF levels are thought to be further reduced at the cortico-striatal synapse in HD through disruption of axonal transport and vesicular trafficking (Baydyuk and Xu 2014).

1.7.2.3 mHTT promotes apoptosis

mHTT is considered to lead to a dominant, gain of function phenotype, as the presence of the HD mutation is necessary and sufficient to cause disease. However, there is growing evidence that the loss of normal HTT function might also contribute to disease pathogenesis. Heterozygous loss of HTT induces neurodegeneration and behavioural and cognitive deficits, suggesting that the loss of one copy of HTT is detrimental (Nasir *et al.* 1995; Zeitlin *et al.* 1995; O'Kusky *et al.* 1999). Primary neuronal HD cultures and patient derived neurons are particularly vulnerable to glutamate toxicity and growth factor withdrawal (The HD iPSC Consortium 2012), however, overexpression of wild-type HTT has a neuroprotective effect in these models (Leavitt *et al.* 2001; Leavitt *et al.* 2006), whilst further knock-down of HTT exacerbates cell death (Zhang *et al.* 2006). HTT is therefore thought to have antiapoptotic properties. Indeed, HTT inhibits the activation of caspase-3 and caspase-9 (Rigamonti *et al.* 2001; Zhang *et al.* 2003), and as discussed, promotes the transcription and transport of BDNF to cortico-striatal synapses which is critical for striatal MSN survival.

1.7.2.4 mHTT and mitochondrial dysfunction

Mitochondria are essential to cell survival, as they play pivotal roles in energy metabolism, apoptosis pathways and calcium homeostasis (Chan 2006). Mitochondrial defects and oxidative stress appear to underlie a number of neurodegenerative disorders, including Parkinson's Disease, Alzheimer's Disease and amyotrophic lateral sclerosis (Mattson *et al.* 2008; Arun *et al.* 2015). The striatum is particularly vulnerable to impairments in mitochondrial functioning; systemic treatment of rodents and non-human primates with the mitochondrial toxin 3-nitropropionic acid (3-NPA)

induced choreiform movements, dystonia and striatal degeneration, with preferential loss of the medium spiny neurons and in man it induced dystonia and striatal atrophy (Ludolph *et al.* 1991; He *et al.* 1995; Liot *et al.* 2009). Additionally, mutations in mitochondrial genes, such as a mutation in complex III, ubiquinol-cytochrome *c* reductase can lead to basal ganglia dysfunction and striatal atrophy (Sarzi *et al.* 2007; Spruijt *et al.* 2007; Barel *et al.* 2008). These findings suggest that regional patterns of neurodegeneration and tissue-specificity might result from deficits in mitochondrial function.

Enzymatic activity of a number of mitochondrial enzymes is reduced or altered in HD brains. Post-mortem brain tissue revealed reduced activity of complexes II, III and IV in the striatum (Vonsattel *et al.* 1985; Browne *et al.* 1997), and altered mitochondrial morphology (Tellez-Nagel *et al.* 1974; Goebel *et al.* 1978). Reduced complex II expression has also been reported in the striatum of HD patients (Benchoua *et al.* 2006) and overexpression of complex II in primary striatal neurons has a protective effect (Damiano *et al.* 2013). Mitochondrial dysfunction appears to parallel region-specific degeneration in the brain (Mazziotta *et al.* 1987; Kuwert *et al.* 1990). Interestingly, no changes in mitochondrial enzymatic activity were seen in HD individuals who had been clinically diagnosed but were not yet demonstrating disease symptoms, indicating no obvious prodromal mitochondrial dysfunction.

Concordant with metabolic dysfunction, patients with HD often display weight loss (Djoussé *et al.* 2002) and muscle wasting (Sathasivam *et al.* 1999). Proton nuclear magnetic resonance (NMR) spectroscopy has revealed increases in lactate concentrations in the cortex and the striatum of HD patients (Jenkins *et al.* 1993; Jenkins *et al.* 1998) and reduced ATP production in muscle (Lodi *et al.* 2000).

To better predict the influence of mitochondrial toxicity, peripheral tissues from living patients are advantageous. Lymphoblasts from HD patients with varying lengths of expanded repeats show reduced complex I enzyme activity, and reduced basal respiration, ATP coupling, spare respiratory capacity (Mejia *et al.* 2016). Reduced mitochondrial membrane potential and calcium dyshomeostasis are often reported in HD models (Cooke *et al.* 2003). Following stress, lymphoblasts derived from HD patients demonstrated increased cell-death and increased mitochondrial membrane depolarisation (Sawa *et al.* 1999), both of which correlated with the length of the expanded *HTT* CAG repeat. Consistent with this, mitochondria from HD lymphoblast cells demonstrated a lower membrane potential and depolarised at lower calcium loads (Panov *et al.* 2002). Additionally, mitochondria from transgenic HD rats demonstrated Ca^{2+} sensitivity; decreased Ca^{2+} threshold for permeability transition, reduced Ca^{2+} capacities and reduced membrane potential stability in response to Ca^{2+} (Choo *et al.* 2004; Gellerich *et al.* 2008).

There is debate as to whether these mitochondrial deficits are caused directly through the interaction of HTT and mitochondria, or through transcriptional changes induced by mHTT. Indeed, HTT interacts with the outer mitochondrial membrane and mHTT reduces the Ca^{2+} threshold required for mitochondrial permeability transition (MPT) pore opening (Choo *et al.* 2004). MPT pore opening induces cytochrome c release and apoptosis, and is a key component in the induction of cell death (Crompton 1999; Krieger and Duchen 2002; Halestrap 2006; Rasola *et al.* 2010). mHTT aggregates have been demonstrated to impair mitochondrial movement along neuronal processes, and as a result mitochondria accumulate at the aggregates (Trushina *et al.* 2004; Chang *et al.* 2006). Soluble N-terminal mHTT fragments also appear toxic as they associate with mitochondria in an age-dependent manner and correlate with disease progression (Orr et al. 2008). N-terminal fragments prevent the association of trafficking proteins with the mitochondria, thereby reducing the distribution and rate of mitochondrial movement (Orr et al. 2008).

Mitochondria are dynamic organelles, able to undergo both fusion and fission to maintain metabolic homeostasis within a cell. Fusion and fission are important for mitochondrial remodelling and rearrangements, critical for maintaining mitochondrial integrity as they enable component sharing across the tubular network. Dysregulation of mitochondrial dynamics is seen in a number of diseases, including HD, and results in the accumulation of damaged mitochondria and an increase in reactive oxygen species (Costa *et al.* 2010; Shirendeb *et al.* 2011; Song *et al.* 2011). mHTT interacts with mitochondrial fission GTPase dynamin-related protein 1 (DRP1), enhancing its enzymatic activity and increasing mitochondrial fission and fragmentation, resulting in a disruption of anterograde mitochondrial movement and synaptic connectivity (Song *et al.* 2011; Reddy and Shirendeb 2012; Shirendeb *et al.* 2012). Treatment of HD fibroblasts and iPSCs with a mitochondrial fission inhibitor reduced excessive mitochondrial fragmentation induced by mHTT, and had beneficial downstream effects on cell survival, and mitochondrial functioning (Guo *et al.* 2013). p53 translocates to mitochondria in a DRP1 dependent manner and its plausible that the recruitment of p53 initiates cell death (Matoba *et al.* 2006; Guo *et al.* 2013). Indeed, knock-down of p53 improved mitochondrial functioning and cell viability in a mouse model of HD (Bae *et al.* 2005).

Increased levels of reactive oxygen species (ROS) are also reported in HD, and compelling evidence supports a critical role of ROS in disease pathogenesis (Wyttenbach 2002; Pérez-Severiano *et al.* 2004; Firdaus *et al.* 2006; Ooi *et al.* 2019). ROS accumulation leads to increased DNA, protein and lipid damage (Sies 1991; Yu 1994). Mitochondrial dysfunction, induced by mHTT in the mechanisms discussed above, leads to increased ROS production, and this in turn can lead to bioenergetic failure. mHTT aggregates have been shown to directly increase ROS production themselves, and the increase of ROS preceded dysfunction and cell death in cells expressing exon 1 of mHTT (Hands

et al. 2011). Though oxidative stress induces DNA damage during normal ageing, oxidative lesions such as 8-Oxo-2'-deoxyguanosine (8-oxo-dG) are increased in models of HD, particularly in the caudate and putamen (Browne *et al.* 1997; Siddiqui *et al.* 2012). Mitochondrial DNA itself is particularly damaged by increasing levels of ROS, and a number of studies have reported mitochondrial DNA depletion in HD cells compared to wild-type cells, and the mitochondrial DNA levels showed an inverse correlation with the length of the *HTT* CAG repeat (Liu *et al.* 2008; Napoli *et al.* 2013; Hering *et al.* 2015). While it appears the accumulation of ROS might follow mitochondrial dysfunction, their generation likely contributes to a toxic oxidation cycle whereby increased ROS levels lead to an increase cellular dysfunction and downstream cell death and degeneration. Consistent with this, anti-oxidants have a neuroprotective effect in models of HD (Stack *et al.* 2010; Johri and Beal 2012; Liu *et al.* 2017).

1.7.2.5 mHTT and the DNA damage response

Along with the HD pathomechanisms described above, the DNA damage response has emerged as a central part of neuronal dysfunction in HD pathogenesis. Accumulation of DNA damage, particularly following oxidative stress in mouse models of HD potentiates repeat expansions, likely leading to increased RNA or protein toxicity (Kovtun *et al.* 2007; Entezam *et al.* 2010). It is likely however, that HTT itself plays a role in the DNA damage response. Cells and mouse models expressing mHTT demonstrate chronic activation of the ATM-p53 pro-apoptotic signalling cascade (Giuliano *et al.* 2003; Bae *et al.* 2005; Maiuri *et al.* 2017). HTT binds at cofilin-actin rods in the nucleus following cellular stress and is involved in nuclear actin reorganisation (Munsie *et al.* 2011). Additionally, HTT has a range of binding partners involved in the DNA damage response. mHTT, but not wild-type HTT, interacts with Ku70, a double strand break repair protein involved in non-homologous end joining (Enokido *et al.* 2010). Ku70 is involved with the recruitment of DNA-dependent protein kinase (DNA-PK) complex to the DNA lesion, which activates enzymatic activity for non-homologous end-joining activities. Soluble mHTT reduced the activity of the DNA-PK complex, likely by preventing the interaction of Ku70 and DNA-PK, leading to an accumulation of double-strand breaks (Qi *et al.* 2007; Enokido *et al.* 2010).

In response to oxidative damage, phosphorylation of N-17 of HTT occurs at serine residues 13 and 16, promoting nuclear entry of HTT where it is thought to act as a molecular scaffold in the DNA damage response (Maiuri *et al.* 2013). Both HTT and mHTT localise at sites of DNA damage in an ataxia telangiectasia mutated kinase (ATM)-dependent manner, where they interact with a number of base excision repair proteins including XRCC1, FEN1, APE1 and HMGB1 (Maiuri *et al.* 2017). Though mHTT is still recruited to sites of DNA damage, it exhibits slowed chromatin dissociation which is thought to slow or impair DNA repair.

It has recently been demonstrated that HTT is part of a transcription-coupled repair complex which is involved with the recognition of DNA damage and facilitation of DNA repair during transcriptional elongation (Gao *et al.* 2019). HTT complexes with RNA polymerase II subunit A (POLR2A), ATXN 3, polynucleotide-kinase-3'-phosphate (PNKP), and CBP. PNKP has 3' phosphatase activity and plays a role in DNA strand break repair (Jilani *et al.* 1999; Karimi-Busheri *et al.* 1999). The presence of this complex has been demonstrated in both the nucleus and mitochondria, suggesting it is involved with repair of nuclear and mitochondrial DNA. Whilst both HTT and mHTT are capable of binding to the catalytic domain of PNKP, mHTT reduced its 3' phosphatase activity by 70-80% (Gao *et al.* 2019). The interaction between HTT and the transcription-coupled repair complex appears to mediate DNA strand breaks during transcription, as mice with an expanded *HTT* CAG repeat demonstrated increased DNA damage in actively transcribing genes (Gao *et al.* 2019). PNKP overexpression drastically reduced caspase-3 activation and reduced cell toxicity.

1.8 Induced Pluripotent stem cells

Mouse models of HD have yielded impressive data and have helped to understand the role of mHTT and genetic modifiers on disease pathogenesis. However, rodents are fundamentally different species to humans, and additionally, animal studies are costly, time-consuming and sampling is usually confined to limited time-points. Reprogramming somatic cells such as fibroblasts, blood or keratinocytes into induced pluripotent stem cells (iPSCs) has invaluable potential for disease modelling, drug discovery and compound screening (Hu *et al.* 2016). Reprogramming of somatic cells is achieved by expression of just four transcription factors; *OCT4, SOX2, c-MYC* and *KLF4* (Takahashi and Yamanaka 2006; Takahashi *et al.* 2007). Expression of such reprogramming factors has been achieved via a wide range of methods including retroviral, lentiviral and adenoviral transduction (Takahashi *et al.* 2007; Hockemeyer *et al.* 2009; The HD iPSC Consortium 2012), delivery of recombinant proteins (D. Kim *et al.* 2009; Zhou *et al.* 2009), mRNA or miRNA transfection (Warren *et al.* 2010; Miyoshi *et al.* 2011) and excisable reprogramming cassettes (Woltjen *et al.* 2009). iPSCs are considered indistinguishable from human embryonic stem cells (hESCs) in terms of morphology, proliferation, gene expression and teratoma formation (Takahashi and Yamanaka 2006), yet don't face availability or ethical concerns that are associated with hESCs. Since reprogramming technologies were first described, hundreds of patient-specific stem cell lines have been generated and characterised. iPSCs derived from patients provide an invaluable renewable resource of cells that can be differentiated into disease-relevant cell types and many demonstrate disease-associated phenotypes. Indeed, iPSCs have been generated from patients with Parkinson's

disease (Soldner *et al.* 2009; Qing *et al.* 2017), Alzheimer's disease, Amyotrophic Lateral Sclerosis (Dimos *et al.* 2008) and Huntington's disease (The HD iPSC Consortium 2012).

1.8.1 The use of iPSCs to study Huntington's Disease

Induced pluripotent stem cells (iPSCs) derived from patients with HD provide a unique opportunity for modelling HD pathogenesis. Moreover, HD-iPSCs and neural cells derived from them demonstrate similar phenotypes to those reported in human tissue and rodent models (Table 1.3). MSNs in the striatum are particularly vulnerable to cell death in HD, and many protocols differentiate HD-iPSCs to striatal MSNs to identify the pathogenic processes that might contribute to their selective degeneration. In 2012 the HD-iPSC consortium described the generation and characterisation of 14 iPSC lines derived from HD patients and both related and unrelated controls (The HD iPSC Consortium 2012). The use of HD-iPSCs was validated by demonstrating a host of well-established HD phenotypes previously described in mouse models of HD, including; transcriptional changes, impaired mitochondria bioenergetics, cell-adhesion and cytoskeletal alterations, altered electrophysiological properties and increased cell death following growth factor withdrawal (The HD iPSC Consortium 2012). Subsequent studies reporting the generation of HD-iPSC lines with varying *HTT* CAG repeat lengths have recapitulated many of these phenotypes and also identified a several novel phenotypes (Table 1.3). Transcriptomic analyses of HD-iPSC lines have revealed changes in pathways relating to cell signalling, cell cycle, axonal guidance, apoptosis and neurodevelopment (An *et al.* 2012; The HD iPSC Consortium 2012; Ring *et al.* 2015; Szlachcic *et al.* 2015; The HD iPSC Consortium 2017). Proteomic techniques have also highlighted alterations in mitochondrial respiration, DNA damage-mediated apoptosis and oxidative stress (Chae *et al.* 2012; McQuade *et al.* 2014). The identification of these pathways in HD-iPSCs parallels those implicated in post-mortem brain tissue from humans and rodents, demonstrating the potential of iPSCs.

Some phenotypes are only seen after exposing the HD-iPSCs or HD-iPSC derived neurons to exogenous stressors. Elevated levels of caspase-3/7 activity are seen in MSNs following BDNF withdrawal (Zhang *et al.* 2010; The HD iPSC Consortium 2012; Mattis *et al.* 2014), oxidative stress (Chiu *et al.* 2015; Nekrasov *et al.* 2016) or glutamate pulsing (The HD iPSC Consortium 2012; The HD iPSC Consortium 2017). iPSC-derived neurons are generally considered immature, and formation of mHTT aggregates is rare. However, mHTT aggregates were reported in iPSC-derived neurons following exposure to the proteasome inhibitor MG132 (Jeon *et al.* 2012), and one paper reported HTT-positive aggregates in 6-month old neurons without the need for cellular stressing (Nekrasov *et al.* 2016). The disease phenotypes are often more pronounced in iPSC lines generated from patients with longer repeat lengths (The HD iPSC Consortium 2012; The HD iPSC Consortium 2017). iPSCs have also been generated from homozygous HD mutation carriers, though there was

no effect of two copies of mHTT on reprogramming or differentiation potential (Camnasio *et al.* 2012). Similarly, there was no distinction in phenotype of repeats between HD-iPSCs with one or two copies of the expanded *HTT* CAG repeat (Camnasio *et al.* 2012).

Despite the variability and large range of *HTT* CAG repeat lengths in HD-iPSCs, most reports noted that the length of the *HTT* CAG repeat length remained fairly stable during reprogramming, iPSC culture and neuronal differentiation (Zhang *et al.* 2010; Camnasio *et al.* 2012). Somatic expansion was reported in one HD-iPSC line however, HD-109, derived from an individual with 109 *HTT* CAG repeats (The HD iPSC Consortium 2012). HD-109 demonstrated *HTT* CAG repeat expansion in culture; expansion of the expanded allele was seen from 110 to 118 over a number of passages (The HD iPSC Consortium 2012). This cell line represents an ideal tool for directly linking HD phenotypes seen in vitro, to *HTT* CAG repeat expansion reported in HD post-mortem brain tissue and mouse models of HD.

Cell Line	Expanded *HTT* CAG repeat length	Donor source tissue	Reprogramming method	Phenotype	Reference
HD-iPS4, HD-iPS11	72	Fibroblast	Retroviral	Enhanced caspase-3/7 activity following growth factor withdrawal; altered mitochondrial bioenergetics, susceptibility to cell death and reduced expression of cadherin, BDNF and TGF-β; reduced neural rosette-formation and poor differentiation efficiencies; differential expression of cytoskeletal, apoptotic and oxidative-stress related proteins; mHTT aggregate formation following proteasome inhibition with MG132	(Park *et al.* 2008; Zhang *et al.* 2010; An *et al.* 2012; Chae *et al.* 2012; Jeon *et al.* 2012)
HD509 HD832 HD1657	44/42 43/39 45	Fibroblast	Lentivirus	Increase in lysosomal activity in HD-iPSCs and iPSC-derived neurons	(Camnasio *et al.* 2012)
HD-60	60	Fibroblast	Lentivirus	Altered energetics, cell adhesion and cytoskeleton; increased caspase 3/7 activation following BDNF withdrawal; calcium dyshomeostasis and increased cell death following glutamate pulsing	(The HD iPSC Consortium 2012)
HD-109	109	Fibroblast	Lentivirus	Increased nuclear condensation following BDNF withdrawal; *HTT* CAG repeat instability	(The HD iPSC Consortium 2012)
HD-180	180	Fibroblast	Lentivirus	Altered energetics, cell adhesion and cytoskeleton; increased cell death; increased caspase 3/7 activation and condensed nuclei following BDNF withdrawal; calcium dyshomeostasis and increased cell death following glutamate pulsing; altered electrophysiology; inhibition of proteasome activity, particularly UBR5 increases mHTT aggregation	(The HD iPSC Consortium 2012; Xu *et al.* 2017; Koyuncu *et al.* 2018)

CS21iHD60nXX	60	Fibroblast	Episomal	No significant effect of BDNF withdrawal; differential expression of neurodevelopmental genes; dysregulated calcium signalling, glutamate and GABA receptor signalling and axonal guidance pathways; altered maturation; impaired neural rosette formation; disrupted cytoarchitecture in cortical organoids	(Mattis *et al.* 2014; The HD iPSC Consortium 2017; Conforti *et al.* 2018)
CS09iHD109nXX CS97iHD180nXX	109 180	Fibroblast	Episomal	Reduced numbers of neural progenitor cells; vulnerability to BDNF withdrawal due to glutamate toxicity; altered transcriptomics; reduction in cortical neurite outgrowth and axonal fasciculation; down-regulation of voltage gated calcium channels; phenotype severity was length-dependent; CS09iHD109nXX also demonstrated differential expression of neurodevelopmental genes; dysregulated calcium signalling, glutamate and GABA receptor signalling and axonal guidance pathways; altered maturation; neuroectodermal acquisition defects; reduction in PNKP activity and accumulation of DNA strand breaks	(Mattis *et al.* 2014; The HD iPSC Consortium 2017; Conforti *et al.* 2018; Mehta *et al.* 2018; Gao *et al.* 2019; Goold *et al.* 2019)
HD71 HD109	71 109	Fibroblast	Episomal	Increased expression of oxidative-stress-related genes, *SOD1;* dysregulation of pathways involving p53, ERK1/2 and Wnt; dysregulated expression levels in HD109 of genes involved in DNA damage response and apoptosis; HD71 lines show dysregulation of several transcriptional regulator genes	(Szlachcic *et al.* 2015; Świtońska *et al.* 2019)
F-HD-iPSC D-HD-iPSC	50 109	Fibroblast	Retroviral	Cytoplasmic vacuoles in astrocytes- increased levels in D-HD-iPSC	(Juopperi *et al.* 2012)
iPSHD11 IPSHD334 IPSHD22	40 42 47	Fibroblast	Episomal	Aggregates of mHTT in 6 month old neurons; elevated levels of lysosomes and autophagosomes; nuclear indentations; neuronal death during ageing; enhanced calcium entry	(Nekrasov *et al.* 2016)

Table 1.3 Human iPSC-based models of HD

1.8.2 Genome-editing of iPSCs

Genome editing of iPSCs using Clusters of Regularly Interspaced Short Palindromic Repeats (CRISPR)/Cas9 technologies has led to a new era in the field of genome editing (Shui *et al.* 2016; Yumlu *et al.* 2019). CRISPR-Cas9 is an increasingly popular technique to study the function of a gene(s) of interest in the context of disease-phenotypes. CRISPR-Cas systems exist in nature as a bacterial defence mechanism against foreign DNA (Barrangou *et al.* 2007), yet can be manipulated to target the genome at virtually any genomic sequence (Charpentier and Doudna 2013; Jiang and Doudna 2017). Cas9 can be directed to cleave at a desired location via its assembly with a guide RNA. The guide RNA consists of a CRISPR-RNA (crRNA) which confers target specificity, and a trans-activating RNA (tracrRNA) which is involved with Cas9 recruitment (Jinek *et al.* 2012). Target specificity of the crRNA also depends on the presence of a short conserved sequence motif known as the protospacer adjacent motif (PAM) (Bolotin *et al.* 2005; Deveau *et al.* 2008). CRISPR-Cas9 is a type II CRISPR system in which the Cas9 protein alone is sufficient to bind the crRNA and cleave the target DNA (Nishimasu and Nureki 2017). Cas9 cleavage results in a double strand break which may be repaired by either non-homologous end joining or homology-directed repair. Non-homologous end joining is error-prone and therefore likely to introduce small insertions or deletions (indels) into the targeted DNA sequence (Perez *et al.* 2008). Non-homologous end-joining is thus useful for mediating gene knock-outs, as an out-of-frame indel can result in premature stop codons (Cong and Zhang 2014). Conversely, the homology-directed repair pathway may be leveraged to introduce a desired gene edit such as a coding mutation or the correction of a disease causing mutation (Chen *et al.* 2011). In this respect a repair template is used, which is homologous to the target site, but contains a defined target locus modification (Saleh-Gohari and Helleday 2004).

1.8.3 Isogenic HD-iPSCs

The first wave of disease modelling studies in iPSCs focused on confirming previously reported phenotypes seen in animal models or immortalised cell lines, and this relied heavily on comparing the phenotypes seen in iPSCs derived from healthy donors with iPSCs derived from donors with a disease. Whilst this approach continues to give us a vast amount of information regarding general disease phenotypes, it tells us less about subtle phenotypes that might present in disease state or what phenotypes can be attributed solely to the disease-causing mutation(s). Here, the desirable ability to study disease in the genetic background of the patient perhaps poses the greatest confounder, as genetic background contributes to high heterogeneity in the properties of the iPSCs. In HD, a highly penetrant, monogenic disease, substantial evidence suggests that genetic variation elsewhere in the genome can significantly influence the age at onset and progression of the disease

(GeM-HD Consortium 2015; Moss *et al.* 2017; GeM-HD Consortium 2019). Genetic and epigenetic variability in iPSC lines can cause functional variability in the differentiated somatic cells (Sandoe and Eggan 2013), and combined with variability introduced in stem cell reprogramming and in vitro culture, it is necessary to implement relevant controls.

An increasingly popular approach is to generate isogenic iPSC lines differing only at the site of the HD mutation, allowing cleaner phenotypic comparison. A number of isogenic HD-iPSC/ ESC lines have been generated and previously reported that mutation-related phenotypes have been rescued following genetic correction of the HD-mutation to a non-pathogenic repeat length; mitochondrial dysfunction, impaired neural rosette formation, vulnerability to growth factor withdrawal, neuronal viability, DNA damage and gene expression changes (An *et al.* 2012; Xu *et al.* 2017; Ruzo *et al.* 2018; Ooi *et al.* 2019). Gene expression changes are particularly sensitive to differences in genetic background, and Xu *et al.* found that after controlling for genetic background in an isogenic system, there was a significant reduction in differential gene expression levels (Xu *et al.* 2017). In the case of genetic modifiers, isogenic iPSCs enable the isolation of the genetic modifier's effect in the context of the expanded repeat.

1.9 Project Aims

Using human genetic data, DNA repair proteins have been identified as determinants of the age at onset in HD. Previous work implicated DNA repair proteins in potentiating *HTT* CAG repeat expansions in striatal neurons and thereby accelerating onset of HD by increasing the toxicity of the expanded protein. *HTT* CAG repeat expansions in mouse models of HD are well described, yet physiologically relevant cellular models of repeat expansion are limited. iPSCs derived from a patient with HD provide an ideal tool for studying disease phenotypes and the factors driving them. With the establishment of a reproducible model of *HTT* CAG repeat expansion, genome-editing can be utilised to manipulate genes of interest and investigate the effect on repeat expansions and associated phenotypes.

Primary objectives:

1) Establish a robust cellular model of *HTT* CAG repeat expansion that can subsequently be used to determine the role of genetic and drug modifiers on repeat expansions.
2) Generate and characterise isogenic HD-iPSCs which differ only in the disease-causing mutation which can be used to dissect phenotypes associated with an expanded *HTT* CAG repeat.
3) Knock-out FAN1 from our cellular assay to investigate the effects of FAN1 on *HTT* CAG repeat expansions in stem cells and neurons.

Chapter 2: Materials and Methods

2.1 Cell culture

2.1.1 Cell culture reagents

Cell Culture Reagent	Supplier	Catalogue Number
Essential 8 Flex Medium Kit	Life Technologies	A2858501
Advanced DMEM/F-12	Life Technologies	12634028
Knockout DMEM/F-12	Life Technologies	12660012
BrainPhys Neuronal Medium	Stem Cell Technologies	05790
Neurobasal MED SFM	Life Technologies	21103049
RHVTN-N (RUO)	Life Technologies	A14700
Corning Matrigel- Growth Factor Reduced	BD Biosciences	354230
Poly-D-Lysine hyrobromide	Sigma-Aldrich	P6407
Glutamax	Thermo Fisher	35050-038
Penicillin/ Streptomycin (5000U/5000 µg)	Gibco	15070063
CryoStor CS10	Stem Cell Technologies	07931
ReLeSR	Stem Cell Technologies	05873
StemPro Accutase Cell Dissociation Reagent	Life Technologies	A1110501
D-PBS	Life Technologies	14190250
Y-27632 dihydrochloride (ROCK inhibitor)	Tocris	1254
Ganciclovir	Merck	345700
G418 (Geneticin)	Life Technologies	10131035
SCR7 pyrazine	Sigma-Aldrich	14892-97-8
MACS NeuroBrew-21	Miltenyi	130-093-566
MACS NeuroBrew-21 (w/o Vitamin A)	Miltenyi	130-097-263
StemMACS™ IWR-1-endo	Miltenyi	130-110-491
Human BDNF, research grade	Miltenyi	130-096-286
StemMACS™ SB431542	Miltenyi	130-105-336
LDN -193189	StemGent	04-0019
Recombinant Human/Murine/Rat Activin A (Insect derived)	PeproTech	120-14
CHIR 99021	Bio-Techne	4423
PD0332991	Bio-Techne	4786

DAPT	Bio-Techne	2634
LM22A4	Bio-Techne	4607
Forskolin	Bio-Techne	1099
Ascorbic Acid	Sigma-Aldrich	A4544
Calicium Chloride	Sigma-Aldrich	499609-1G
Trypan Blue	Gibco	15250061
Countess Cell Counting Chamber Slides	Thermo Fisher	C10312

Table 2.1 List of reagents used for stem cell culture and differentiation

Neuronal differentiation media	Composition
SLI	Advanced DMEM/F12 with 1% Glutamax, 1% Penicillin/ Streptomycin, 1% MACS neurobrew without Vitamin A, 10uM SB (SB431542), 1uM LDN (LDN193189) and 1.5uM IWR1 (IWR-1-endo)
LIA	Advanced DMEM/F12 (ADF) with 1% Glutamax, 1% Penicillin/ Streptomycin, 1% MACS neurobrew without Vitamin A, 1uM LDN (LDN193189), 1.5uM IWR1 (IWR-1-endo) and 25 ng/ml Activin A
SJA	ADF with 1% Glutamax, 1% penicillin/streptomycin, 2% MACS neurobrew with Vitamin A, 2µM PD0332991, 10µM DAPT, 10ng/ml BDNF, 0.5µM LM22A4, 10µM Forskolin, 3µM CHIR 99021, 0.3mM GABA, 1.8mM $CaCl_2$ and 0.2mM Ascorbic acid
SJB	Equal amounts of ADF and Neurobasal A with 1% Glutamax, 1% penicillin/streptomycin, 2% MACS neurobrew with Vitamin A, 2µM PD0332991, 10ng/ml BDNF, 0.5µM LM22A4, 3 µM CHIR 99021, 1.8mM $CaCl_2$ and 0.2mM Ascorbic acid
BrainPhys Terminal	BrainPhys neuronal medium with 1% Glutamax, 1% penicillin/streptomycin, 2% MACS neurobrew with Vitamin A, 10ng/ml BDNF and 0.5µM LM22A4

Table 2.2 Media composition for neuronal differentiation of iPSCs

2.1.2 iPSC maintenance

Human Q109 and Q60 induced pluripotent stem cells (iPSCs) were generated by the HD-iPSC consortium from a human fibroblast line with an expanded CAG *HTT* allele of 109 repeats or 60 repeats, respectively (The HD iPSC Consortium 2012). Three clonal lines were generated from the Q109 line; Q109 N1, N4 and N5 and three generated from the Q60 line; Q60 N4, N5 and N8. The iPSCs were cultured on vitronectin-coated plates (0.5 µg/cm^2) (Life Technologies) in Essential 8 Flex medium (Life Technologies) under standard culturing conditions (37˚C, 5% CO_2).

Cells were passaged every 3-4 days, when reaching confluency of ~ 70%. For passaging, cells were incubated with ReLeSR (Stem Cell Technologies) for 1 minute at 37˚C. After aspirating the ReLeSR, cells were dissociated into small clumps in fresh warmed medium and were seeded into a new plate at a density of 1:12.

For freezing, cells were dissociated with ReLeSR as described above, centrifuged at 1000 rpm for 3 minutes and resuspended in CryoStor CS10 (Stem Cell Technologies) with approximately 1x10^6 cells/ 0.5 mL CryoStor CS10. Cryovials containing the cell suspension were then transferred to a CoolCell Freezing Container (Corning) and placed at -80°C where cells were frozen at a rate of -1°C/minute.

For thawing iPSCs, cryovials were warmed to 37˚C in a water bath for 1-2 minutes until partially thawed. 1 mL of warm Advanced DMEM/F-12 (ADF) (Life Technologies) was then added to the cells dropwise. The cell suspension was then transferred to an Eppendorf, centrifuged at 1000 rpm for 3 minutes, and resuspended in warmed E8 Flex Medium containing 10µM Y-27632 dihydrochloride (Rock Inhibitor).

2.1.3 Monolayer differentiation of iPSCs to striatal neurons

iPSCs were pre-treated for 1 hour with 10 µM Y-27632 and then dissociated into a single cell suspension using Accutase (Life Technologies). iPSCs were plated into 12 well plates coated with Growth Factor Reduced Matrigel (0.5 ng/mL) (BD Biosciences) and cultured in iPSC medium until the cells reached ~80% confluency. The differentiation protocols used are adapted versions of previously published protocols (Arber *et al.* 2015; Kemp *et al.* 2016). An outline of the differentiation protocol is shown in Figure 2.1 with media compositions described in Table 2.2.

NPC differentiation: iPSCs were induced into NPCs through monolayer differentiation. For the first 8 days, cells were grown in SLI media (Table 2.2) which was changed daily. On day 8, cells were passaged at a 1:1.5 ratio using Accutase in the presence of 10 µM the ROCK inhibitor Y-27632. Dissociated cells were plated onto Matrigel coated 12 well plates and grown in LIA media (Table 2.2), which was changed daily. At day 16, samples of these cultures were taken for RNA, DNA, cryopreservation or passaged for LGE neural differentiation.

For neural differentiation, tissue culture plates and glass coverslips were coated with Poly-D-Lysine (Sigma-Aldrich). Briefly, vessels were incubated for 1 hour at room temperature with 10 µg/ml of Poly-D-Lysine, followed by 3 washes in water and allowed to air dry overnight. The following day vessels were coated with Matrigel (0.5 ng/ml) for 1 hour at 37°C. A single cell suspension of NPCs was obtained by Accutase treatment in the presence of 10 µM Y-27632, cells were counted (see 2.1.4) and then the volume of cell solution adjusted to $1X10^6$ cells/ml. 80µl of this cell suspension was then plated onto a 13 mm glass coverslips coated with PDL and Matrigel to achieve a final cell number of 50,000/coverslip. After 1 hr incubation at 37 °C and 5% CO_2 in a humidified incubator, each well was gently flooded with SJA medium (Table 2.2). For RNA analysis at day 37, cells were plated directly onto PDL and Matrigel coated 12 well plates at a density of 500,000 cells/well. Half media changes were performed every other day until day 23 when the cells had a complete media change into SJB medium (Table 2.2). Cells were then maintained in this medium with half volume media changes every other day until day 37. At day 37 if samples were not fixed for immunocytochemistry or taken for RNA or DNA, media was changed to BrainPhys Terminal media (Table 2.2) and half volume media changes were performed every 2 days.

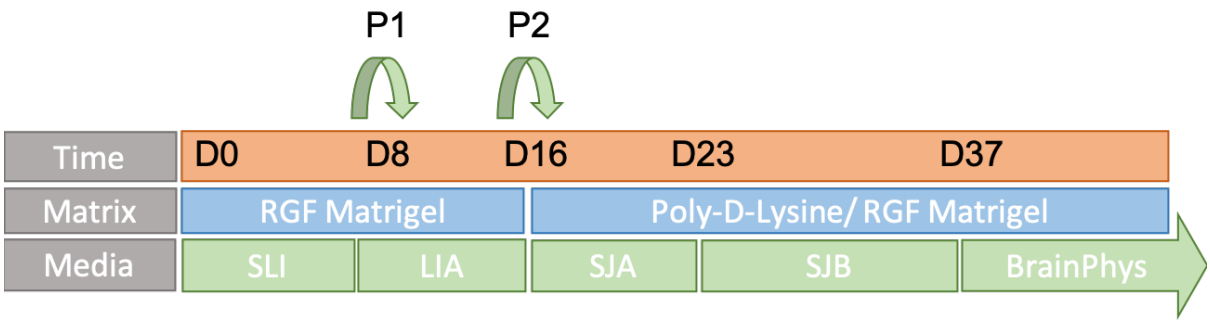


Figure 2.1 Differentiation of iPSCs into neural precursor cells and striatal neurons. Schematic depicts time in the number of days (orange), cellular matrix used (blue) and media used at each differentiation stage (green).

2.1.4 Cell counting

Cells were counted using the Countess II FL Automated Cell Counter (Thermo Fisher). Cells were dissociated to a single cell suspension using Accutase and mixed in equal proportions with Trypan Blue. Cells were mixed thoroughly and 10 µL of the cell/Trypan blue suspension was pipetted into each side of a Countess Cell Counting Chamber Slides (Thermo Fisher) for duplicate measurements. Slides were inserted into the instrument and information regarding total cell count, live cell count and average cell size were recorded.

2.2 Nucleic Acid Extraction

2.2.1 DNA extraction

For sensitive applications, DNA was routinely extracted using the QIAamp DNA Mini Kit (QIAGEN) following manufacturer's instructions. DNA was eluted in 100 µL of nuclease-free water and kept at -20°C.

2.2.2 DNA extraction for genotyping

For genomic DNA extraction from a 96-well plate, cells were washed with DPBS, and 50µL of DNA lysis buffer (1 M Tris pH 8.0, 0.5 M EDTA, 10% SDS, 5 M NaCl) supplemented with 0.5 mg/mL proteinase K (Sigma-Aldrich) was added to each well. The plate was gently vortexed to ensure thorough mixing and then incubated at 55°C overnight in a humidified chamber. The following day, 10 µL of the DNA sample was diluted in 90 µL 10 mM Tris pH 8.0 and the mixture was incubated at 95°C for 30 minutes to inactivate the proteinase K.

2.2.3 RNA extraction

RNA was extracted using RNAeasy mini Kit (QIAGEN) and RNase-free DNase kit (QIAGEN) according to manufacturer's instructions. Briefly, cells were washed once with DPBS and lysed in 350 µL RLT buffer containing 10 µL/mL ß-mercaptoethanol. At this point lysates could be stored at -80°C for up to 6 months. Frozen lysates were thawed gently on ice and then vortexed for 1 minute to ensure homogenisation of cells. Ethanol was then added at a 1:1 ratio and the sample was mixed thoroughly before being added to a RNeasy spin column and centrifuged to bind the RNA to the column membrane. To ensure complete DNA removal from the sample, the RNase-free DNase kit was used. Using the buffers provided, the column membrane was washed several times and RNA was eluted into 30 µL RNase-free water. RNA was stored at -80°C.

2.2.4 Nucleic acid quantitation

The concentration and quality of both DNA and RNA was determined using Nanodrop ND1000 spectrophotometer (Labtech international, Uckfield, UK). The 260/ 230 and 260/280 ratios were used to determine quality of the DNA and RNA, respectively.

2.3 Quantitative Real-time polymerase chain reaction (qRT-PCR)

Both SYBR® Green and TaqMan® qRT-PCR technologies were used in this project.

2.3.1 cDNA synthesis

RNA was reverse transcribed to cDNA using qScript cDNA Synthesis Kit (Quantabio) according to the manufacturers protocol. The same RNA concentration was used in every experiment to ensure consistency. Reactions were kept on ice and performed in 0.2 mL PCR tubes:

Reagent	Volume (µL)
RNA (10 pg- 1 µg total RNA)	variable
qScript Reaction Mix (5X)	4
qScript RT	1
Nuclease-free water	Up to 20

The reaction was vortexed gently followed by a brief centrifugation. Reactions were then placed in a T100 Thermal Cycler (Biorad) on the following programme:

Cycles	Temperature	Time
1	22°C	5 minutes
1	42°C	30 minutes
1	85°C	5 minutes
	4°C	Forever

2.3.2 SYBR Green qRT-PCR

Quantification of target gene transcripts was performed on the LightCycler 96 Instrument (Roche) using Fast Start Essential DNA Green Master Mix (Roche). Primers used were designed and validated previously, with established efficiencies. 10 µM stocks for each primer pair that was to be used (Table 2.3) were prepared and kept at -20°C. A mastermix was made and 8 µL was pipetted into each well together with 2 µL containing 50ng of cDNA:

Reagent	Volume (µL)
Fast Start Essential DNA Green Master (2X)	5
F + R primers (10 µM each)	0.5
Nuclease-free water	2.5
cDNA (50ng)	2.0
Total	10

The LightCycler PCR plates (Star Labs) were sealed with MicroAmp Optical Adhesive Film (AppliedBiosystems), vortexed and centrifuged briefly. The qRT-PCR was run on the Lightcycler 96 Instrument following a reaction protocol of:

Cycles	Temperature	Time
1	95°C	1 minutes
40	95°C	10 seconds
	60°C	30 seconds

Gene name	Forward (F)/ Reverse (R)	Sequence (5'- 3')
OCT4	F	CTCACCCTGGGGGTTCTATT
	R	CTCCAGGTTGCCTCTCACTC
SOX2	F	TACCTCTTCCTCCCACTCCA
	R	GGTAGTGCTGGGACATGTGA
NANOG	F	GATGGGAGGAGGGGAGAGGA
	R	TTTGGAAGCTGCTGGGGAAG
PAX6	F	AGGCCAGCAACACACCTAG
	R	AGCCAGATGTGAAGGAGGAA
GSX2	F	ATGTCGCGCTCCTTCTATGT
	R	ATGCCAAGCGGGATGAAGAAA
FOXG1	F	AGGAGGGCGAGAAGAAGA
	R	ACTCGTAGATGCCGTTGAGC
MAP2	F	AAAGCTGATGAGGGCAAGAA
	R	GGCCCCTGAATAAATTCCAT
CTIP2	F	TGGGTGCCTGCTATGACAAG
	R	GATGCCTTTCGTGGGTGAGA
DARPP-32	F	CCTGAAGGTCATCAGGCAGT
	R	GGTCTTCCACTTGGTCCTCA
GAPDH	F	TGCACCACCAACTGCTTAGC
	R	GGCATGGACTGTGGTCATGAG
ß-Actin	F	CCCAGCACAATGAAGATCA
	R	ACATCTGCTGGAAGGTGGAC

Table 2.3 List of Syber Green qRT-PCR primers and their sequences

2.3.3 TaqMan qRT-PCR

Quantification of target gene transcripts was also performed on the StepOne Plus Instrument (Applied Biosystems) using TaqMan Fast advanced Master mix and TaqMan gene expression assays. TaqMan gene expression assay probes were pre-designed by Life Technologies (Table 2.4).

Gene name	TaqMan Assay ID
GAPDH	Hs02758991_g1
ß-ACTIN	Hs01060665_g1 ACTB
FAN1	Hs00429686_m1
FAN1	Hs00429687_m1 FAN1

Table 2.4 List of TaqMan gene expression assays used and their IDs

A master mix was made and 8 µL was pipetted into each well together with 2 µL containing 75ng of cDNA.

Reagent	Volume (µL)
TaqMan Fast Advanced Master Mix	5
TaqMan Assay (10x)	0.5
Nuclease-free water	2.5
cDNA (75 ng)	2.0
Total	10

Reactions were pipetted into MicroAmp® Fast Optical 96-well reaction plates (Applied Biosystems) and sealed with MicroAmp® Optical adhesive film (Applied Biosystems). Plates were vortexed and centrifuged before being loaded into the StepOne Plus Machine. The thermal cycler protocol was as follows:

Cycles	Temperature	Time
1	50°C	2 minutes
1	95°C	2 minutes
40	95°C	1 second
	60°C	20 seconds

Data were analysed in Microsoft Excel using the delta-delta CT method to determine relative gene expression levels. Data was based on three biological replicates and samples were run in triplicate.

2.4 Molecular cloning of piggyBac transposon

A 5' homology arm, a 1.7-kb region upstream of the HTT locus and a 1.6-kb 3' homology arm containing human HTT exon 1 with 22 CAG repeats were synthesised using GeneArt (Thermo Fisher). The two homology arms were cloned into a piggyBac Multivector SGK-004 (Transposagen) using a two-step cloning process (see 3.3.2). To clone the homology arm into the targeting vector, both vector and insert DNA were treated with restriction enzymes, causing double strand breaks that produce complementary overhanging single stranded nucleotides. 500 ng of the 5' homology arm and 500ng of the piggyBac vector were digested for 1 hour at 37°C with NsiI-HF (NEB) and AscI (NEB). For the 3' homology arm, piggyBac vector and insert DNA were digested with NotI (NEB) for 1 hour at 37°C followed by the addition of BsiWI (NEB) and incubation at 55°C for a further hour. Restriction digest products were then electrophoresed on a 2% agarose gel. Correctly sized bands were excised from the gel and purified using QIAquick Gel Extraction Kit (QIAGEN) according to the manufacturer's instructions. The piggyBac vector was then combined with 5' homology arm or the 3' homology arm (performed after confirmation of successful 5' homology arm insertion) for the ligation of the two. Ligation was performed at room temperature for 5 minutes using Quick Ligase (NEB) according to manufacturer's instructions. For transformation of the vector, NEB Stable Competent *Escherichia coli* cells were incubated on ice for 2 minutes with 50 ng of plasmid DNA. After 2 minutes on ice, cells were heat shocked at 42°C for 30 seconds, followed by another incubation on ice for 2 minutes. Transformed bacteria were streaked onto Luria Broth (LB) agar plates containing ampicillin and incubated at 37°C overnight. The following day, single colonies were selected from the streaked LB agar plate and dipped into liquid LB with ampicillin. The bacterial cultures were incubated overnight in a shaking incubator, ensuring constant agitation. Plasmid DNA was then extracted from half of the bacterial culture for confirmation using QIAprep Spin Miniprep Kit (QIAGEN) and digested with NsI-HF and AscI for the 5' homology arm, or NotI and BsiWI for the 3' homology arm to confirm the presence of the insert. Polymerase chain reactions (PCRs) were performed from four primer pairs which amplified ligation junctions of the homology arms (Table 2.5) to confirm homology arm insertion.

PCRs were performed using MyTaq™ DNA Polymerase (Bioline). PCR reactions included:

Reagent	Volume (µL)
MyTaq DNA Polymerase	0.2
MyTaq Reaction Buffer (5X)	2.0
Forward primer (10 µM)	0.5
Reverse primer (10 µM)	0.5
Nuclease-free water	5.8
DNA (100ng)	1.0
	10.0

Reactions were vortexed gently and centrifuged before being placed in a T100 Thermal-Cycler under the following PCR conditions:

Cycles	Temperature	Time
1	95°C	1 minute
35	95°C	15 seconds
	Variable*	30 seconds
	72°C	30 seconds
1	72°C	5 minutes
	4°C	Forever

Name	*Forward (F)/ Reverse (R)*	*Sequence (5'- 3')*	**Annealing Temperature*
3' homology arm-NsiI	F	CTAAATGCACAGCGACGGAT	54°C
	R	CTGAGGAAGCTGAGGAGGC	
3' homology arm- AscI	F	GATGTGTCCCAGATGGCATT	58°C
	R	CCATGATTACGCCAAGCTCG	
5' homology arm- NotI	F	GTTGTAAAACGACGGCCAGT	55°C
	R	GGGGCGTTTCTTTATGGGAG	
5' homology arm- BsiWI	F	CCATTCATTGCCCCGGTG	57°C
	R	TCCAGACTGCCTTGGGAAAA	

Table 2.5 List of primers used to confirm presence and orientation of homology arms cloned into piggBac targeting vector

Orientation of the insert was confirmed by Sanger sequencing (see 2.7) of PCR products. Once both homology arms had successfully been cloned into the vector, it was important to yield endotoxin-free plasmid DNA which would be suitable for transfection of iPSCs. Liquid LB and ampicillin were inoculated with bacteria transformed with the complete targeting plasmid, and plasmid DNA was extracted using an Endotoxin-Free Maxi Plasmid Purification Kit (QIAGEN) according to manufacturer's instructions. For long term storage of the bacteria, glycerol stocks were made by combining equal quantities of the overnight bacterial culture with 50% glycerol and freezing at -80°C.

2.5 CRISPR-Cas9 targeting

2.5.1 Genetic correction of the HD mutation

2.5.1.1 gRNA design and synthesis

To edit exon 1 of the *HTT* gene, we selected a guide RNA (gRNA) target site of known high efficiency (An *et al.* 2012) (Table 2.6).

gRNA	Location	Sequence (5'- 3')
Isogenic-HTT	Exon 1 HTT	GCCTCCGGGGACTGCCGTGCGG

Table 2.6 gRNA sequence for targeting exon 1 of HTT. Shown is the 20 nucleotide long gRNA sequence (black) and protospacer adjacent motif (green).

To generate a functional gRNA duplex, a universal *trans*-activating CRISPR RNA (tracrRNA) (IDT) was annealed to the CRISPR-Cas9 crRNA (IDT). The crRNA contains the target specific sequence for guiding the Cas9 protein to a genomic location. gRNA oligos were synthesised (IDT) and resuspended in IDTE (1X TE solution) to a concentration of 200 µM. The tracrRNA and crRNA were annealed at an equimolar ratio by incubating for 2 minutes at 95°C, with IDT duplex buffer (30 mM HEPES, 100 mM potassium acetate), to generate a gRNA duplex.

	Volume (µL)
crRNA	1.0
tracrRNA	1.0
Duplex buffer	2.5
Total	4.5

Cas9 protein was then diluted in storage buffer to a final concentration of 4 µg/µL.

Cas9 storage buffer	Concentration
Tris-HCl, pH 7.4	10 mM
NaCl	300 mM
EDTA	0.1 mM
DTT	1 mM

To form the RNP complex, Cas9 protein was mixed with the gRNA duplex and incubated at room temperature for 10-20 minutes.

	Volume (µL)
Cas9 protein (4 µg/µl)	1.0
crRNA:tracrRNA complex	1.0
Total	2.0

2.5.1.2 Nucleofection of iPSCs

Nucleofection was used in the generation of the isogenic HD-iPSCs (see also Chapter 3) and the FAN1 knock-out iPSCs (see also Chapter 4). Cells were transfected using the Lonza 4D-Nucleofector and the P3 Primary Cell 4D-Nucleofector X Kit according to the manufacturers' instructions. For one nucleofection, cells were dissociated to a single cell suspension with Accutase and 2 x 10^5 cells were resuspended in 20 µL of complete P3 buffer (16.4 µL of P3 buffer and 3.6 µL of P3 supplement) and 2 µL per RNP complex was added. Final amounts per nucleofection: 2 x 10^5 cells, 4 µg Cas9 (IDT), 4 µg gRNA and 1 µg of plasmid DNA. Cells were nucleofected using program CA-137 and immediately after transfection cells were replated into warmed E8 medium, supplemented with 10 µM Y-27632 and 50 µM SCR7, a non-homologous end joining inhibitor.

2.5.1.3 Selection with Geneticin (G418)

48 hours after transfection, cells were washed and E8 media was supplemented with 100 µg/mL Geneticin (G418) (Sigma-Aldrich). Cells were fed with E8 + G418 every other day until day 10, when daily cell death was high and so cells were washed and fed daily. Individual colonies were picked between days 13-16 after selection was started into wells of a 96 well plate containing fresh E8 and G418. G418 selection was stopped 48 hours after picking. When colonies were ready to be split, cells were split into a duplicate 96 well plate for clonal expansion and the remaining cells were using for screening.

2.5.1.4 Genotyping using polymerase chain reaction (PCR)

To confirm successful targeting at the *HTT* locus, PCRs using primers amplifying exon 1 of the *HTT* locus (*HTT*-exon 1 and LKH-*HTT*) were performed (Table 2.7). Genomic DNA was extracted from the candidate clones (2.2.2). PCRs over the *HTT* CAG repeat were performed using LA Taq DNA polymerase with GC buffers (TaKaRa), ideal for amplification of GC-rich regions that are often challenging to PCR with high fidelity. PCR reactions included:

	Volume (µL)
TaKaRa LA Taq (5 U / µL)	0.1
GC Buffer II (2X)	5.0
dNTP Mixture (2.5 mM each)	1.6
Forward primer (10 µM)	0.5
Reverse primer (10 µM)	0.5
Nuclease-free water	1.3
DNA (100ng)	1.0
Total	10

Reactions were vortexed gently and centrifuged before being placed in a T100 Thermal-Cycler under the following PCR conditions:

Cycles	Temperature	Time
1	94°C	3 minutes
35	94°C	30 seconds
	65°C	30 seconds
	72°C	1 minute 30
1	72°C	5 minutes
	4°C	Forever

2.5.1.5 Agarose gel electrophoresis

PCR products were run on 2% agarose gels made with 0.5% TBE. 2 µL of DNA Gel Loading Dye (6X) (Thermo Scientific) was added to each PCR product. 3.5 µL of HyperLadder™ 100bp (Bioline) was loaded into the first lane of each gel, and 12 µL of PCR product and loading dye was loaded into each lane. Gels were electrophoresed at 90 V for 1 hour 30.

Primer name	*Forward (F)/ Reverse (R)*	*Sequence (5'- 3')*	**Annealing Temperature*
HTT-exon 1	F	ATGAAGGCCTTCGAGTCCCTCAAGTCCTTC	65°C
	R	GGCGGCTGAGGAAGCTGAGGA	
FAM-huHTT-exon 1	F	FAM-ATGAAGGCCTTCGAGTCCCTCAAGTCCTTC	65°C
	R	GGCGGCTGAGGAAGCTGAGGA	
LKH-HTT	F	CCCATTCATTGCCCCGGTGCTG	65°C
	R	CTGGGTCACTCTGTCTCTG	
HTT-sequencing	F	ATTCATTGCCCCGGTGC	60°C
	R	CTGGGTCACTCTGTCTCTG	
TK-NEO-cassette	F	TCGCCTTCTTGACGAGTTCT	58°C
	R	GCAGATACCGCACCGTATTG	
LHA-junction	F	TACAGGCATGCACCACTACA	59°C
	R	TACAGGCATGCACCACTACA	
RHA-junction	F	TACCCGAGCCGATGACTTAC	59°C
	R	ATCAAGTGTCTGAAGCAGCC	

Table 2.7 List of primers used for *HTT* genotyping

2.5.1.6 piggyBac selection cassette excision

To remove the selection cassette, 2 x 10^5 HD-iPSCs were dissociated to single cells using Accutase in the presence of 10 µM Y-276321 (see 3.3.2.3). Cells were transfected (see 2.5.1.2) with 1 µg of the piggyback excision-only transposase vector and then seeded into a 10 cm dish containing warmed E8 supplemented with 10 µM Y-276321. 48 hours after transfection cells were placed on negative selection with 1 µM ganciclovir and media was changed daily. Colonies were picked after 5-7 days into fresh media containing 10 µM Y-276321 and 1 µM ganciclovir.

As above, when colonies were ready to be split, cells were split into a duplicate well for clonal expansion, whilst the remaining cells were harvested for genomic DNA extraction. Genomic DNA was

extracted using DNA lysis buffer (see 2.2.2) and several PCRs were performed to confirm successful targeting and removal of the piggyBac selection cassette. Junction-based PCRs were performed using primers amplifying each of the homology arm junctions (LHA-junction, RHA-junction) (Table 2.7). Primers amplifying a region of the piggyBac selection cassette (TK-NEO-cassette) were used to confirm the absence of the cassette from the whole genome (Table 2.7). LKH-HTT primers were used to amplify exon 1 of *HTT* and *HTT*-sequencing primers (Table 2.7) were then used for Sanger Sequencing (see 2.7) the amplicon in each direction.

2.5.2 *FAN1* knock-out

2.5.2.1 gRNA design and synthesis

Two guide RNAs (gRNAs) targeting the second exon of FAN1 were designed using DESKGEN Cloud (www.deskgen.com) (Table 2.8) based on their predicted high efficiency and low probability for off-target effects.

gRNA	Location	Sequence (5'- 3')
FAN1 KO 1	Exon 2 R14.3 FAN1	CTGATTGATAAGCTTCTACGAGG
FAN1 KO 2	Exon 2 V46.2 FAN1	GCACCATTTTACTGCAAACGGGG

Table 2.8 gRNA sequence for targeting exon 2 of FAN1. Shown is the 20 nucleotide long gRNA sequence (black) and protospacer adjacent motif (green).

RNA oligos were synthesised (IDT) and resuspended in IDTE (1X TE solution) to a concentration of 200 µM. A tracrRNA fluorescently labelled with ATTO-550 (IDT) and crRNA were annealed at an equimolar ratio by incubating for 2 minutes at 95°C, to generate a gRNA duplex. Cas9 protein was then diluted in storage buffer to a final concentration of 4 µg/µL. To form the RNP complex, Cas9 protein was mixed with the gRNA duplex and incubated at room temperature for 10-20 minutes. For each gRNA to be used, 2 µL of RNP was produced for nucleofection.

2.5.2.2 Flow cytometry analysis of nucleofected iPSCs

Cells were transfected as in 2.5.1.2. FACs was used to sort transfected cells based on fluorescence after nucleofection with the fluorescently labelled RNP complex. 24 hours after transfection, cells were dissociated with accutase to a single cell suspension, centrifuged and resuspended in 250µL E8 + Penicillin/ Streptomycin (1:1000) + 10 µM Y-27632 which was then kept on ice. Cells were sorted on the FACS ARIA Fusion (BD Biosciences) to select for the top 10% of cells emitting the highest levels of fluorescence, increasing efficiency. Cells were replated onto a 10 cm tissue culture dish containing E8 + 10 µM Y-27632 and media was changed every other day before colonies were picked.

2.5.2.3 Genotyping of targeted *FAN1* clones

5-7 days after transfection, single colonies were picked into 96-well plates containing E8 + 10 µM Y-27632. As above, when required cell colonies were split into a duplicate 96-well plate. Cells grown on one plate were used for clonal expansion, whilst genomic DNA extraction and PCR screening was performed with cells in the other.

Genomic DNA was extracted as above using DNA lysis buffer (2.2.2). To detect successful deletion of the target region, a PCR using primers amplifying exon 2 of FAN1 was performed (Table 2.9).

Primer name	Forward (F)/ Reverse (R)	Sequence (5'-3')	* Annealing Temperature
FAN1-KO	F	CGCTTTTCCCCTTGCTGAAT	58°C
	R	CTGCACCAATTCCAGGCTTA	
FAN1-T7	F	TCAGAGTTCGCTTTTCCCCT	58°C
	R	GATGCTAGGCTTCCCAAACA	

Table 2.9 List of primers used for FAN1 genotyping

PCRs were performed using MyTaq™ DNA Polymerase (Bioline). PCR reactions included:

	Volume (µL)
MyTaq DNA Polymerase	0.2
MyTaq Reaction Buffer (5X)	2.0
Forward primer (10 µM)	0.5
Reverse primer (10 µM)	0.5
Nuclease-free water	5.8
DNA (100ng)	1.0
Total	10

Reactions were vortexed gently and centrifuged before being placed in a T100 Thermal-Cycler under the following PCR conditions:

Cycles	Temperature	Time
1	95°C	1 minute
35	95°C	15 seconds
	Variable*	30 seconds
	72°C	30 seconds
1	72°C	5 minutes
	4°C	Forever

2.6 CAG repeat sizing

Genomic DNA was isolated from cells using the QIAamp DNA mini kit (Qiagen). Fluorescently labelled primers amplifying the *HTT* CAG repeat (Table 2.7) and the resulting PCR products were analysed using capillary electrophoresis (GA3130*x*L Genetic Analyser, Applied Biosystems). For capillary electrophoresis, PCR products were mixed with Hi-Di™ Formamide (Applied Biosystems) and a GeneScan™-600 LIZ size standard (Applied Biosystems):

	Volume (µL)
PCR product	0.5
Hi-Di formamide	9.1
GeneScan-600 LIZ size standard	0.4
Total	10

Reactions were centrifuged and then subject to a denaturation step:

Cycles	Temperature	Time
1	95°C	4 minutes
1	4°C	3 minutes

Samples were added to the GA3130*x*L Genetic Anayser (Applied Biosystems) and a run module was created according to manufacturer's recommendations. Peak scanner v2.0 and GeneMapper software were used to view *HTT* CAG repeat size distribution traces and determine *HTT* CAG repeat length.

CAG repeat length was determined from the size of the modal (highest) peak in the trace:

HTT CAG repeat length = (size of modal peak – 80)/ 3

To quantify repeat instability, instability and expansion indexes were calculated using previously described methods (Lee *et al.* 2011).

2.7 Genotyping methods

2.7.1 Sanger sequencing

For Sanger sequencing, genomic DNA was extracted using QIAamp DNA Mini Kit (QIAGEN). PCRs were performed with the relevant primers (Table 2.5, Table 2.7 and Table 2.9) and 50 ng of genomic DNA. PCR products were purified using the QIAquick PCR Purification Kit (QIAGEN). Purified PCR product was eluted in 30 µL of nuclease-free water. 5 µL of purified PCR products (20-80 ng/µL) were added to an

Eppendorf with 5 µL of either the forward or reverse primer (5 pmol/µL) that was used in the PCR. 10 µL of PCR product and primer were then sent to GATC for their LightRun sequencing.

2.7.2 Exome Sequencing

Exome sequencing was kindly carried out by Branduff McAllister in Cardiff University. Exome libraries were produced from 75 ng (10 µL at 7.5 ng/ µL) of genomic DNA using the TruSeq Rapid Exome Library Prep Kit (Illumina), according to manufacturer's instructions. DNA clean-up was performed using SPRI beads (Beckman Coulter). Exomes were sequenced in-house using the HiSeq 4000 to produce raw FASTQ files. To call and annotate the variants, three pipelines were used: a standard GATK pipeline, quality control (QC) pipeline and an annotation pipeline. Sequenced reads were aligned to the human reference genome [Human GRCh37 (hg19) build] using Burrows-Wheeler Aligner (BWA). Base recalibration and de-duping were performed as standard QC measures, followed by quality control following the Picard pipeline (www.gatk.co.uk) to produce variant-ready BAM files. Variants are called using the Genome Analysis Toolkit (GATK) (www.gatk.co.uk), creating variant calling files (VCFs). Additional QC measures were employed, considering variants >=10 reads and genotyping quality >=30. Variant annotation used a custom hail pipeline and annotated variants with gnomad and dbNSFP.

2.7.3 SNP array

SNP array genotyping experiments were kindly carried out by Alexandra Evans in Cardiff Universities Medical Research Centre (MRC) department, Hadyn Ellis Building. Genomic DNA was extracted using QIAamp DNA Mini Kit and 200 ng (50 ng/ µL) was required for genotyping. Samples were genotyped on the Infinium PsychArray-24 Kit (Illumina) or the Infinium Global Screening Array-24 (Illumina) and scanned using the iScan System (Illumina). Data were exported from Genome Studio and analysed using PennCNV (Fang and Wang 2018). Sample level quality control was applied based on the standard deviation of Log R ratio set at 0.3, minimum SNP number of 10 and minimum region size of 100,000bp.

2.8 Protein Extraction and Western Blotting

Cells were washed once in DPBS and 150 µL of radio-immunoprecipitation assay buffer (RIPA) (Sigma-Aldrich) with protease inhibitors (Sigma-Aldrich) and phosphatase inhibitors (Sigma-Aldrich) was directly added to the tissue culture plate. Cells were scraped and the lysate was vortexed for 30 seconds and incubated on ice for 30 minutes, vortexing every 5 minutes. Samples were then centrifuged at 14,000 rpm for 15 minutes at 4°C. The supernatant was transferred to another tube for quantitation and storage at -80°C.

For protein quantification, the Pierce bicinchoninic acid assay (BCA) protein assay kit was used (Thermo Fisher). Protein concentrations were determined relative to dilutions of a standard protein bovine

serum albumin (BSA) ranging from 25 µg/mL to 200 µg/mL. A BCA working reagent was prepared by mixing 50 parts of BCA reagent A with 1 part of BCA reagent B. 10 µL of the sample and standard was pipetted into the microplate in triplicate. 200 µL of the working reagent was then added to each well and the plate was mixed thoroughly on a plate shaker for 30 seconds. The plate was then wrapped in tin foil and incubated at 37°C for 30 minutes. Plate absorbance was read at 562nm on the CLARIOstar spectrophotometer (BMG Labtech). Protein concentration was extrapolated from the standard curve of the BSA standards.

1X samples buffers were prepared by adding 2.5 µL of NuPAGE LDS Sample Buffer (4x) to 30 µg of protein sample and deionised water up to 10 µL. The samples were then denatured at 70°C for 10 minutes. To prepare the running buffer, 50 mL of 20X NuPAGE Tris-Acetate SDS Running Buffer was added to 950 mL of deionised water. The running buffer was added to the Mini Gel Tank (Thermo Fisher) and samples were separated on NuPAGE 3-8% protein gel (Fisher Scientific). Samples were loaded into the appropriate wells, alongside 5 µL of protein standard (NEB) and run for 1 hour at 150 V. PVDF membranes (Thermo Fisher) were activated in 100% methanol for 2 minutes, before transferring the protein to the membrane using a Mini Blot Module (Thermo Fisher). Protein was transferred in NuPAGE Transfer Buffer (Thermo Fisher) for 45 minutes at 120 V. The membrane was blocked in 5% milk in Tris Buffered Saline containing 0.1% Tween (TBS-T) for one hour at room temperature. Primary antibodies (Table 2.10) were also diluted in 5% milk in TBS-T in a 50 mL falcon to which the membrane was transferred. The membrane and antibody solution were incubated at 4°C overnight on a rotating tube mixer. The following day membranes were washed in TBST 3 times for 20 minutes. Secondary antibodies used were conjugated to infra-red dyes (LI-COR). Secondary antibodies were also diluted in 5% milk 1:15,000 and the membrane was incubated in the secondaries for 1 hour at room temperature. The membrane was washed with TBST a further 3 times for 20 minutes before being visualised on the LI-COR Odessay CLx system.

Target	Species	Dilution	Supplier	Catalogue Number
Beta-tubulin	Mouse	1:1000	AbCam	Ab6046
FAN1	Sheep	1:1000	CHDI	N/A

Table 2.10 Antibodies used in Western Blotting

2.9 Cell Imaging

2.9.1 Immunocytochemistry

Cells plated on glass coverslips had culture media removed and were washed once in DPBS (Life Technologies) before incubation with 4% paraformaldehyde (PFA) (Sigma-Aldrich) for 15 min at room temperature. Post-incubation, cells were washed 3 times in DPBS. Fixed cells were stored in DPBS at 4°C until staining. Dilutions and conditions for each antibody can be seen in (Table 2.11). Following fixation, depending on the antibody used, cells were permeabilised using either 0.1% triton-X (Sigma-Aldrich) in PBS for 20 minutes or ice-cold absolute ethanol for 2 minutes. If the target was a cell surface antigen, no permeabilization step was required. Cells were then washed 3 times in DPBS. To prevent non-specific binding, cells were blocked in a blocking buffer of 1-3% BSA (Sigma-Aldrich) with 3% normal goat serum (NGS) (Invitrogen) unless otherwise stated, in PBS for one hour at room temperature. Following blocking, the cells were incubated for one hour at 37°C in a humidified chamber with the primary antibody at an appropriate dilution in blocking media. For iPSCs and NPCs the droplet method was employed; coverslips were placed on a parafilm covered plate lid on top of a bed of moist tissue paper, and 50 µL of primary antibody dilution was gently pipetted onto the coverslip, with care to ensure full coverage of the solution. For neurons, primary antibody incubation was performed in tissue culture plates. Cells were then washed in 3 x 20 minute washes in DPBS. Cells were then incubated for one hour at 37°C in a humidified chamber with the appropriate secondary antibody (Table 2.12). Following secondary incubation, cells were washed again in 3 x 20 minute washes in DPBS. Coverslips were then mounted using vector shield containing DAPI nuclear stain onto a glass microscope slide.

1° antibody	Species	Supplier (catalogue number)	Dilution	Permeabilisation	Blocking
Beta III Tubulin	Rabbit polyclonal	Abcam	1:1000	/	3% NGS + 3% BSA
CTIP2	Rat monoclonal	Abcam (ab18465)	1:200	0.1% Triton-X in PBS, 20 minutes, RT	3% NGS + 3% BSA + 0.1% Triton-X in PBS
DARPP-32	Rabbit monoclonal	Abcam (ab40801)	1:200	0.1% Triton-X in PBS, 20 minutes, RT	3% NGS + 3% BSA + 0.1% Triton-X in PBS
GABA	Rabbit polyclonal	Sigma-Aldrich (A2052)	1:100	/	3% NGS + 3% BSA
γH2A.X	Mouse monoclonal	Abcam (ab11174)	1:250	0.1% Triton-X in PBS, 20 minutes, RT	1% BSA
MAP2	Mouse monoclonal	Sigma-Aldrich (m1406)	1:500	Ice-cold absolute ethanol, 2 minutes	3% NGS + 3% BSA
NESTIN	Mouse monoclonal	Millipore (MAB5326)	1:500	0.1% Triton-X in PBS, 20 minutes, RT	3% NGS + 3% BSA + 0.1% Triton-X in PBS
OCT4	Mouse monoclonal	Santa Cruz (sc5279)	1:100	Ice-cold absolute ethanol, 2 minutes	3% NGS + 3% BSA + 0.1% Triton-X in PBS
PAX6	Mouse	Developmental Studies Hybridoma Bank	1:10	0.1% Triton-X in PBS, 20 minutes, RT	3% NGS + 3% BSA + 0.1% Triton-X in PBS
Poly ADP-Ribose	Rabbit	Sigma (11835238001)	1:200	0.1% Triton-X in PBS, 20 minutes, RT	1% BSA
SSEA-4	Mouse	Developmental Studies Hybridoma Bank	1:10	/	3% NGS + 3% BSA
ZO-I	Mouse	BD Transduction Laboratories (610966)	1:100	3% NGS + 3% BSA + 0.1% Triton-X in PBS	3% NGS + 3% BSA + 0.1% Triton-X in PBS

Table 2.11 Primary antibodies used for immunocytochemistry

2° antibody	Dilution	Supplier	Catalogue Number
Alexa Fluor goat anti-mouse IgG 488	1:400	Invitrogen	A11001
Alexa Fluor goat anti-mouse IgG 568	1:800	Invitrogen	A11004
Alexa Fluor goat anti-rabbit IgG 488	1:800	Invitrogen	A27034
Alexa Fluor goat anti-rabbit IgG 568	1:800	Invitrogen	A11011

Table 2.12 Secondary antibodies used for immunocytochemistry

2.9.2 Imaging and analysis

Cells were imaged using a Lecia DM6000B inverted fluorescence microscope. For a more detailed examination of nuclear foci, cells were imaged using the Zeiss LSM710 Confocal Microscope. Confocal images were analysed using Volocity Software (Quorum Technologies). Volocity software parameters were set up to automatically count cells using DAPI. Volocity software was also used to count γH2A.X foci. Foci were excluded if they were <0.05 µm or >1 µm. Data analysis and representation were carried out using GraphPad Prism 8.

2.9.3 IncuCyte live cell imaging

For Incucyte measurements, iPSCs were dissociated into a single cell suspension and plated onto vitronectin coated 6-well plates at a density of 25,000 or 50,000 cells/well in E8 media. 24 hours after plating, cells were washed to remove any dead cells and the tissue culture plates were transferred to the IncuCyte S3 live-cell imaging system (Essen BioScience), housed inside a 37°C/ 5% CO_2 incubator. The plates were imaged every 2 hours for 96 hours with a 10x objective lens. Culture media was changed after 48 hours. Samples were run in triplicate wells and 16 phase images were taken per well. Analysis was performed using the IncuCyte S3 software. Phase contrast masking was applied, with an area filter of 50 µm to exclude dead cells.

2.9.4 Seahorse XF Cell Mito Stress Test

The Agilent Seahorse XF Mito Stress test was used to measure mitochondrial function through the direct measurement of oxygen consumption rate. Neural precursor cells were dissociated to a single cell suspension using Accutase and plated into XF Cell Culture Microplates at a density of 80,000 cells/well in LIA medium (Table 2.2).

The day prior to the experiment, the Seahorse XFe Analyser was switched on to allow the instrument to stabilise and the temperature increase to 37°C. To hydrate the sensor cartridge, 200 µL of Seahorse XF calibrant was added to each well of the Seahorse XF96 sensor cartridge and incubated overnight at 37°C in a non-CO_2 incubator. On the day of the assay, assay medium was prepared by supplementing Seahorse XF DMEM medium with 1 mM pyruvate, 2 mM glutamine and 10 mM glucose. Assay compounds were then diluted to 10x stock concentrations in warmed assay media: 20 µM oligomycin, 20 µM carbonilcyanide p-triflouromethoxyphenylhydrazone (FCCP) and 5 µM rotenone and antimycin A. The hydrated sensor cartridge was then loaded with 20 µL oligomycin in port A, 22 µL FCCP in port B and 25 µL rotenone and antimycin A in port C to give a final compound concentration of 2 µM, 2 µM and 0.5 µM respectively. The cell culture microplate was removed from the 5% CO_2/ 37°C incubator, growth media removed, cells washed once in assay medium and then incubated with 175 µL of assay

medium for 45 minutes at 37°C in a non-CO_2 incubator. The hydrated sensor cartridge was then inserted into the Seahorse XFe96 instrument and using the Wave software, a Seahorse XF Cell Mito Stress test protocol was set up. After a calibration step with the sensor cartridge, the cell culture microplate was added to the instrument and the Seahorse XF Cell Mito Stress test protocol began.

2.10 DNA damage assays

To examine the DNA damage response to exogenous stressors, iPSCs and neurons were treated with Mitomycin C (MMC) and potassium bromate ($KBrO_3$).

2.10.1 Mitomycin C

For MMC cell survival assays, cells were seeded into wells of a 6 well plate at a density of 50,000 cells/ well. 24 hours after plating, cells were treated in duplicate with increasing concentrations of MMC from 10-60 ng/mL, for 16 hours. After 24 hours, media was removed, cells were washed once in DPBS and cells were fed with warmed E8 media. After 72 hours, cells were dissociated using Accutase and counted using the Countess II Automated Cell Counter (2.1.4). Total and live cell counts were obtained for each well in triplicate.

To establish the DNA damage response of iPSCs and neurons to MMC treatment, cells were seeded onto coverslips in 12 well tissue culture plates. iPSCs were seeded 3 days prior to MMC treatment, whilst neurons were seeded 14 days prior to MMC treatment. iPSCs were treated with 25 ng/mL or 50 ng/mL MMC for 24 hours prior to fixing, and neurons were treated with 25 ng/mL MMC for 24 hours prior to fixing. Cells were fixed as above (2.9.1) and stained with Poly-ADP-Ribose and γH2A.X.

2.10.2 Potassium Bromate

Neurons were seeded 14 days prior to $KBrO_3$ exposure onto coverslips in a 12 well tissue culture plate. Neurons were treated with 100 μM $KBrO_3$ (Sigma-Aldrich) for 30 minutes. Following 30 minute exposure to $KBrO_3$ neurons were washed with PBS and either fixed immediately, or fed with fresh E8 media and fixed after a recovery window of 24 hours. Cells were fixed as above (2.9.1) and stained with Poly-ADP-Ribose and γH2A.X. 3 coverslips were stained for each genotype at each time point, and 5 representative images were taken of each coverslip.

Chapter 3: The development and validation of isogenic HD-iPSC lines

3.1 Introduction

HD is caused by an expanded CAG repeat in the huntingtin gene, *HTT.* The expanded *HTT* CAG repeat tract is known to expand both intergenerationally and somatically (Ranen *et al.* 1995; Gusella and MacDonald 1996), giving rise to a variation of *HTT* CAG repeat lengths in different regions of the brain and other organ tissues (Macdonald et al 1993, Zuhlke 1993). In particular, marked *HTT* CAG expansions are seen in striatal medium spiny projection neurons, the cells most vulnerable to degeneration, perhaps contributing to their selective degeneration (Telenius *et al.* 1994). In post-mortem human brains, considerable *HTT* CAG repeat expansions (> 100) are seen compared to inherited *HTT* CAG repeat lengths reported in the blood. This somatic instability correlates with an earlier age of motor onset than expected from the CAG repeat length alone (Shelbourne *et al.* 2007; Swami *et al.* 2009), suggesting *HTT* CAG repeat expansions might drive disease pathogenesis. DNA repair pathways have been implicated in somatic expansion of the *HTT* CAG repeat in mouse models of HD (Wheeler *et al.* 2003; Pinto *et al.* 2013; Iyer *et al.* 2015) and in altering age at onset of disease by genetics (GeM-HD Consortium 2015; GeM-HD Consortium 2019). Factors promoting CAG repeat expansions therefore represent novel therapeutic targets; if we can block this process, we could delay disease onset or slow disease progression, so understanding the mechanisms behind the process are imperative.

HTT CAG repeat expansion in mouse models of HD is well reported (Wheeler *et al.* 2003; Lee *et al.* 2010; Pinto *et al.* 2013), however immortalised cell lines taken from these animals seldom show ongoing expansion and those that do are highly replicative cells (Manley, Pugh, *et al.* 1999). In addition, animal models are costly, time-consuming and usually confined to sampling at limited time-points; therefore, having a model of CAG repeat expansion in a physiologically relevant cell system which is renewable, reproducible, and allows for high throughput screening is desirable.

An iPSC line with an expanded CAG repeat tract of 109 CAGs in *HTT* was reported to expand further in culture over a number of passages (The HD iPSC Consortium 2012), representing a physiologically-relevant in vitro cell model which could be used as a surrogate of repeat expansions *in vivo* to assay environmental or genetic factors influencing expansion. In this context HD-iPSCs represent an advantageous alternative to animal models as they provide a renewable resource of cells that can be differentiated into disease-relevant cell types and demonstrate CAG-repeat-length associated phenotypes (An *et al.* 2012; The HD iPSC Consortium 2012; Mattis and Svendsen 2017). Advances in gene-editing technologies have meant that genetic variants can be easily introduced into iPSC models and the influence on disease pathogenesis can be investigated at a cellular and molecular level.

However, whilst studying disease in the context of the patient's genetic background has been one of the biggest advantages of iPSC disease-modelling, it also poses the greatest challenge, as genetic background contributes considerably to heterogeneity in the properties of the iPSCs. Interpretation of phenotypes from multiple individual patient stem cell lines therefore could result in inappropriate conclusions being drawn. Indeed, although HD is monogenic, substantial evidence suggests that genetic variation elsewhere in the genome influences the age at which motor symptoms present in HD patients (GeM-HD Consortium 2015; GeM-HD Consortium 2019). Hence, ideal comparisons should evaluate cell lines with an identical genetic background, which differ only in the mutation under investigation, i.e., through the generation and analysis of isogenic cell lines.

A number of isogenic HD-iPSC/ ESC lines have been generated and previously reported that mutation-related phenotypes have been rescued following genetic correction of the HD-mutation to a non-pathogenic repeat length. These include phenotypes related to mitochondrial dysfunction, neural rosette formation, disrupted neurogenesis, sensitivity to stressors, neuronal viability, DNA damage and gene expression changes (An *et al.* 2012; Xu *et al.* 2017; Ruzo *et al.* 2018; Ooi *et al.* 2019). Isogenic lines have been generated for a number of diseases using numerous gene-targeting methods including homologous recombination (An *et al.* 2012), Cre-LoxP recombination (Qing *et al.* 2017), Zinc Finger Nuclease (Hockemeyer *et al.* 2009) and TALEN genome editing (Ooi *et al.* 2019) many of which leave 'genomic scars' and have reportedly low efficiencies. An alternative to these techniques is the combined use of a site-specific nuclease and a homologous recombination repair template. Here we have used CRISPR-Cas9 assisted homologous recombination to integrate a piggyBac transposon vector, to generate foot-print free isogenic HD-iPSCs. piggyBac is a DNA transposon from the cabbage looper moth which can efficiently transpose into mammalian systems (Ding *et al.* 2005), but may also be used as a homologous recombination repair template to facilitate genome-editing of large, or challenging manipulations (Xie *et al.* 2014; Wang *et al.* 2017). Whilst homologous recombination in hESCs/iPSCs is rare, CRISPR-Cas9 enhances the efficiency and specificity of homologous recombination, and the addition of selectable markers in the piggyBac transposon increases the identification of these rare recombination events.

3.2 Chapter aims

This chapter aims to establish a model of CAG repeat expansion in HD-iPSCs with which we can assess the contribution of genetic modifiers. To ensure the phenotypes we see are not due to variances in genetic background, we also aim to generate control isogenic HD-iPSC lines with WT CAG repeat lengths. The isogenic HD-iPSCs enable the isolation of the effect of the genetic modifier in the context of the expanded *HTT* CAG repeat. It will aim to characterise the model and understand differences in

their genetic makeup. It will examine whether mutation corrected iPSCs are able to differentiate into different neuronal lineages- in particular forebrain neurons. It will then look at HD-like phenotypes previously reported in HD-iPSCs and examine whether these phenotypes are rescued with the genetic correction. This chapter aims to establish a line that can be used to help delineate the effects of genetic modifiers on CAG repeats of wild-type and pathogenic repeat lengths and the phenotypes associated with both.

3.3 Results

3.3.1 HD-iPSCs exhibit CAG repeat instability at the *HTT* locus

Expansion of the *HTT* CAG repeat in a HD-iPSC line derived from a juvenile HD patient with a 109 CAG repeat (Q109) was reported in iPSCs and iPSC-derived neural progenitor cells (NPCs) (The HD iPSC Consortium 2012) and we sought to establish an assay for this expansion that could be manipulated to identify factors promoting or inhibiting the process. Another HD-iPSC line derived from a patient with 60 CAG repeats (Q60) was used to ascertain whether repeat expansions also occurred in a cell line with a CAG repeat length closer to the common disease-associated range. Three HD-iPSC clonal lines derived from each patient were cultured; Q109N1, Q109N4, Q109N5, and Q60N4, Q60N5, Q60N8. These parent populations demonstrated heterogeneity with respect to *HTT* CAG repeat length and so subcloning was performed to achieve subclones with a single length *HTT* CAG repeat. Parent lines were therefore plated as single cells and single colonies picked, giving a number of subclones (Q109N1, n=13; Q109N4, n=16; Q109N5, n=7; Q60N4, n=7; Q60N5, n=8; Q60N8, n=11). A PCR amplifying the CAG repeat region was performed and resultant CAG repeat length analysed. CAG repeat length was fairly consistent across the Q60 subclones, with all of the subclone repeat lengths in the range of 56 to 59 CAGs (Figure 3.1A), although none of the subclones had a CAG repeat length of 60 as previously reported. In contrast, considerable variation was present in the repeat lengths of the Q109 subclones, with repeat lengths ranging from 92 to 122 (Figure 3.1B), indicative of a mosaic population. None of the subclones retained the original repeat length of 109 CAGs. To get an indication of repeat expansion over time in the HD-iPSCs, individual subclones were cultured for 4 weeks and DNA samples were taken at each passage. Q109 subclones exhibited varying degrees of instability of the mutant HD allele; most of the repeat length changes were expansions, although contractions were also observed in two of the subclones (Figure 3.1C, Figure 3.1D). CAG repeat lengths for the WT allele were also determined and showed no change. The Q60 cell line displayed mild instability of the CAG repeat, with an expansion of two CAG repeats in one subclone (Figure 3.1C), suggesting the line is fairly stable, or expansions occur at a slower rate. Because the effect of modifiers is likely to be greater in a cell line that exhibits expansions at a higher rate, we only used the Q109 lines for subsequent experiments.

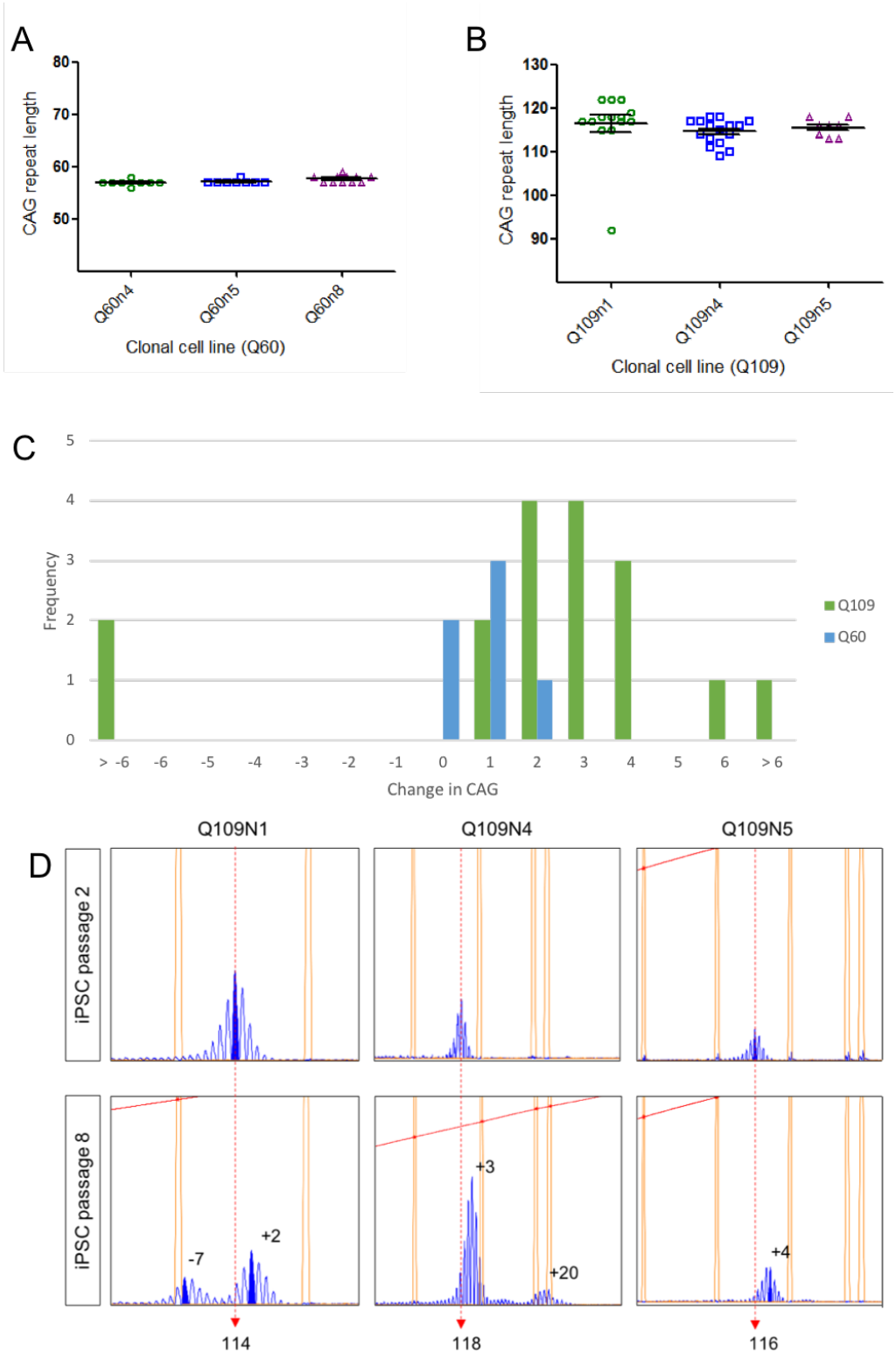


Figure 3.1 HD-iPSCs exhibit repeat instability in culture (A) CAG repeat lengths in the mutant *HTT* allele from multiple subclones from three clonal lines of Q60 (Q60N4, n=7; Q60N5, n=8; Q60N8, n=11) and (B) three clonal lines of Q109 (Q109N1, n=13; Q109N4, n=16; Q109N5, n=7) lines represent mean and standard error of mean. (C) Change in CAG repeat length across multiple passages in a number of HD-iPSC Q109 subclones (Q109N1, n=5; Q109N4, n=10; Q109N5, n=2) and HD-iPSC Q60 subclones (Q60N4, n=2; Q60N5, n=4). (D) Representative electropherograms of fluorescent PCR and capillary electrophoresis of the CAG repeat in HD-iPSC clones derived from Q109 cells.

3.3.1 HD-iPSC derived neurons exhibit CAG repeat instability at the *HTT* locus

Four Q109 subclones, with similar starting *HTT* CAG repeat lengths, were differentiated into NPCs and further directed towards a striatal fate (2.1.3). Cell samples for DNA were taken at each NPC passage and weekly upon neuronal differentiation. HD-NPCs and neurons demonstrated *HTT* CAG expansions (Figure 3.2) but changes in CAG repeat length occurred at a lower frequency than those occurring in iPSCs. Expansion rates showed variation between subclones: Q109N1 SC11 demonstrated the highest *HTT* CAG expansion rates with an average expansion rate of 1 CAG repeat increase per 18.8 days, compared with an average rate of 1 CAG repeat increase per 80 days in Q109N1 SC5, and an average increase of 1 CAG repeat per 31.5 days in Q109N4 SC1 and Q109N5 SC10.

In a number of subclones in dividing cells (iPSC/ NPC), we observed a bimodal distribution of peaks where an additional peak is visible in the trace at an expanded or contracted length (Figure 3.1D, Figure 3.2B). In this case, contractions are more common. This likely represents two populations in the culture and raises the question whether shorter repeats have a selective advantage in vitro.

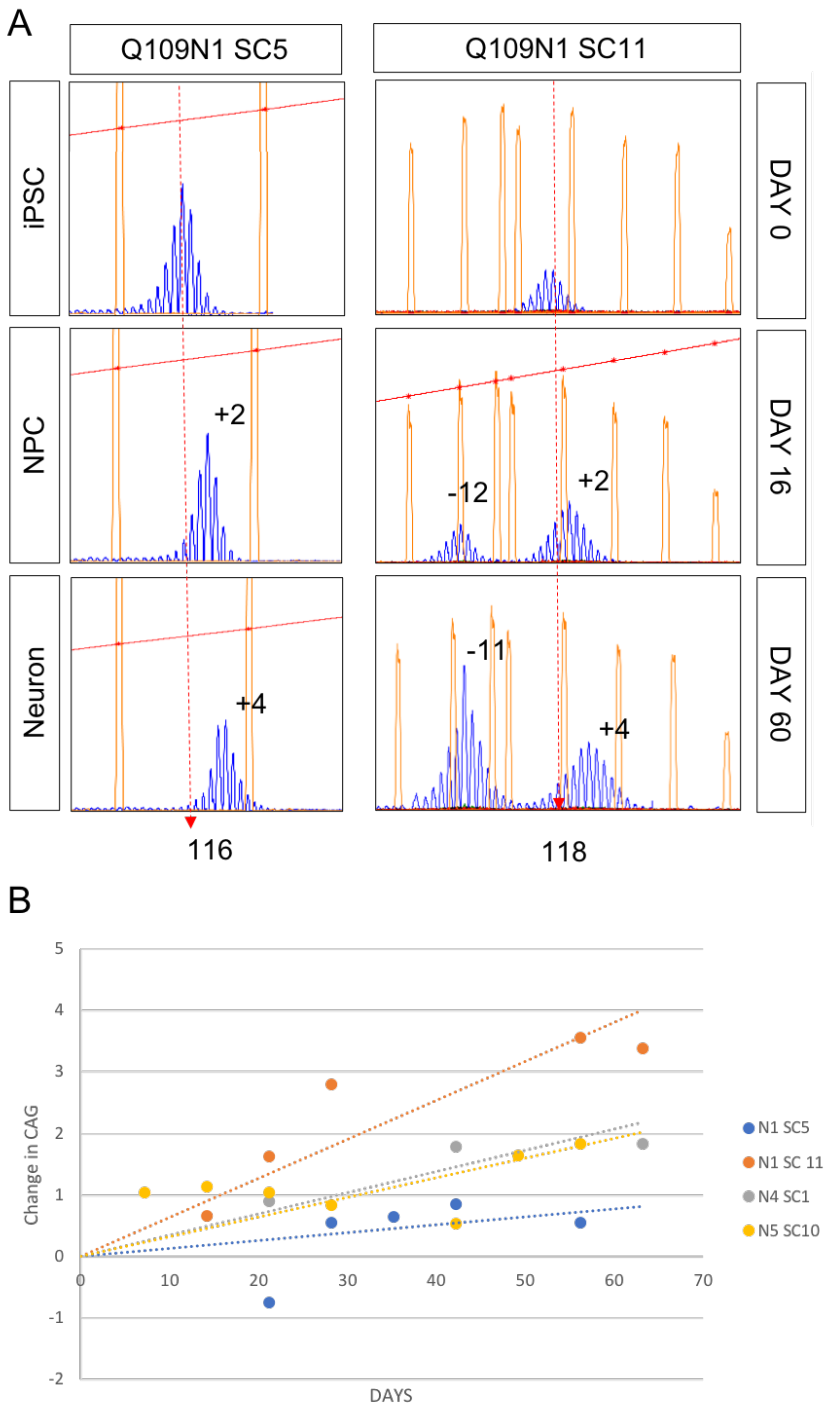


Figure 3.2 Neurons derived from Q109-iPSCs exhibit expansion in culture (A) Representative electropherograms of fluorescent PCR and capillary electrophoresis of the CAG repeat in HD-iPSC derived NPCs and neurons from Q109 cells. Q109N1 SC5 and Q109N1 SC11 represent two subclones of the parent clone Q109N1. (B) Graph showing change in modal *HTT* CAG repeat length in post-mitotic neurons generated from 4 subclones of Q109 (Q109N1, n=2; Q109N4, n=1; Q109N5, n=1). Day 0 represents day 0 of neuronal plating.

3.3.2 Genetic correction of HD-iPSCs

3.3.2.1 Generation of piggyBac transposon

To establish isogenic controls for the Q109 HD- iPSCs, we used the CRISPR/Cas9 nuclease system together with a piggyBac transposon selection cassette-based homologous recombination donor (Figure 3.3), enabling correction of the expanded CAG repeat tract to a wild-type length of 22. To edit exon 1 of the *HTT* gene, we selected a guide RNA (gRNA) target site of known high efficiency (An *et al.* 2012) (Figure 3.3) close to the CAG repeat tract, and chose a TCAA tetra-nucleotide 61 bp away from the double strand break (DSB) site for introduction of a silent mutation (to TTAA) necessary for insertion or release of the piggyBac selection cassette. So as not to cleave the donor construct upon transfection, the PAM site was mutated (CGG > CCG) creating a silent mutation. The piggyBac vector contains both a neomycin-resistance gene for positive selection of vector integration and the herpes simplex virus thymidine kinase gene for negative selection for subsequent piggyBac vector sequence removal (Figure 3.4A).

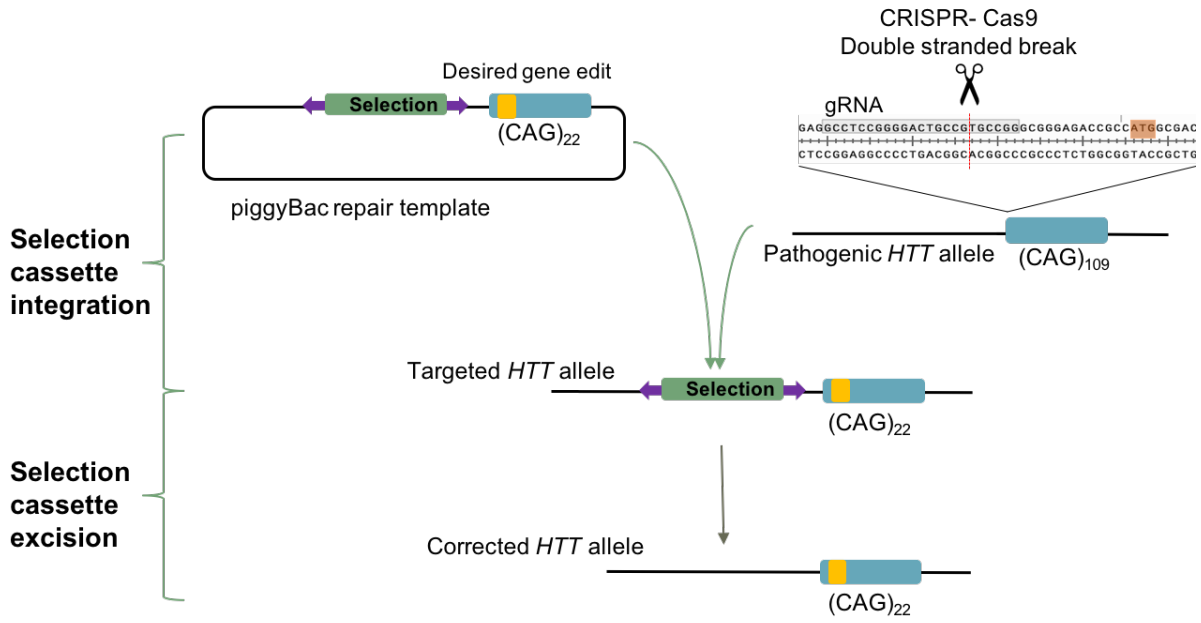


Figure 3.3 Schematic depiction of the 2-step footprint-free editing strategy using the CRISPR-Cas9 and piggyBac systems. Schematic shows gRNA sequence and predicted Cas9 cut site in *HTT* locus marked by the red line, which lies 19 bp away from HTT start codon.

The donor vector was created with a two-stage cloning process (Figure 3.4). The 5' homology arm containing a sequence homologous to a 1.7-kb region upstream of the *HTT* locus was amplified by PCR and ligated into piggyBac Multivector SGK-004 (Transposagen) following digestion with NotI and BsiWI (Figure 3.4) (see 2.4). A 1.6-kb 3' homology arm containing human *HTT* exon 1 with 22 CAG repeats was amplified and then ligated into the vector following digestion with AscI and NsiI (Figure 3.4C), creating a 9.8-kb donor vector (Figure 3.4A). Correct homology arm insertion and orientation was confirmed by Sanger sequencing.

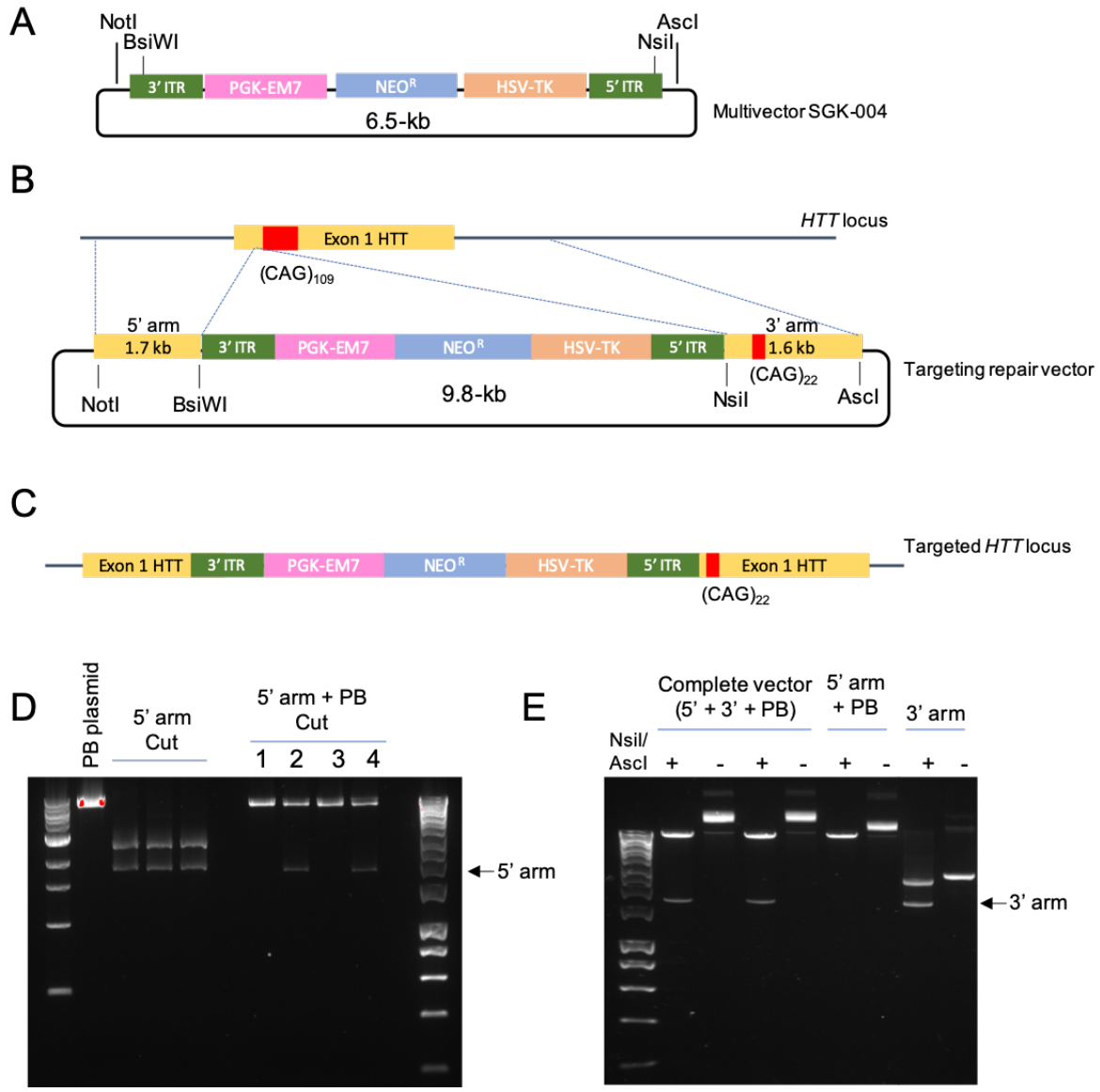


Figure 3.4 piggyBac targeting vector construction. (A) Map of piggyBac Multivector SGK-004 plasmid. (B) Schematic depicting the HTT locus and the homology PB targeting repair vector design. (C) Schematic showing the targeted allele after cassette integration. (D) DNA electrophoresis gel showing digestion of the PB plasmid and 5' homology arm vector with NotI and BsiWI. Following ligation of 5' homology arm and the PB plasmid, digestion with NotI and BsiWI confirms correct ligation in clones 2 and 4. (E) DNA electrophoresis gel showing digestion of the 3' homology arm vector, PB plasmid with 5' insertion with AscI and NsiI. Ligation of the 3' homology arm into the vector was then confirmed by digestion with AscI and NsiI.

3.3.2.2 Targeted correction at the *HTT* locus in Q109 HD-iPSCs

200,000 Q109N1 HD-iPSC cells were transfected with the donor vector together with the Cas9 RNP complex and allowed to recover in 10 µM Y-27632 for 24 hours, followed by selection in 100 µg/mL G418 for 13-16 days. After selection we obtained 288 viable colonies which were expanded and screened using PCR (Figure 3.5). Clones were determined to be targeted candidates if the PCR showed amplification of the normal allele in the absence of the expanded allele product, as the selection cassette insertion disrupted the primer binding sequence (Figure 3.5). This screen does not exclude false positives such as cassette insertion without homologous recombination occurring, but it does reduce the screening pool. Of the 288 colonies screened, targeting of the expanded allele occurred in five clones (1.7%) and targeting of the wild-type allele occurred in two clones (0.7%) (Figure 3.5). Of these, three where the expanded allele was targeted were taken forward for further characterisation; 22-C1, 22-C2, 22-C3, and one clone in which the WT allele was targeted was taken forward; 109-C4 (Figure 3.6C). This provided us with a cell line with 109 CAG repeats that had also gone through the targeting process, ensuring that the gene-editing process itself did not influence any readout.

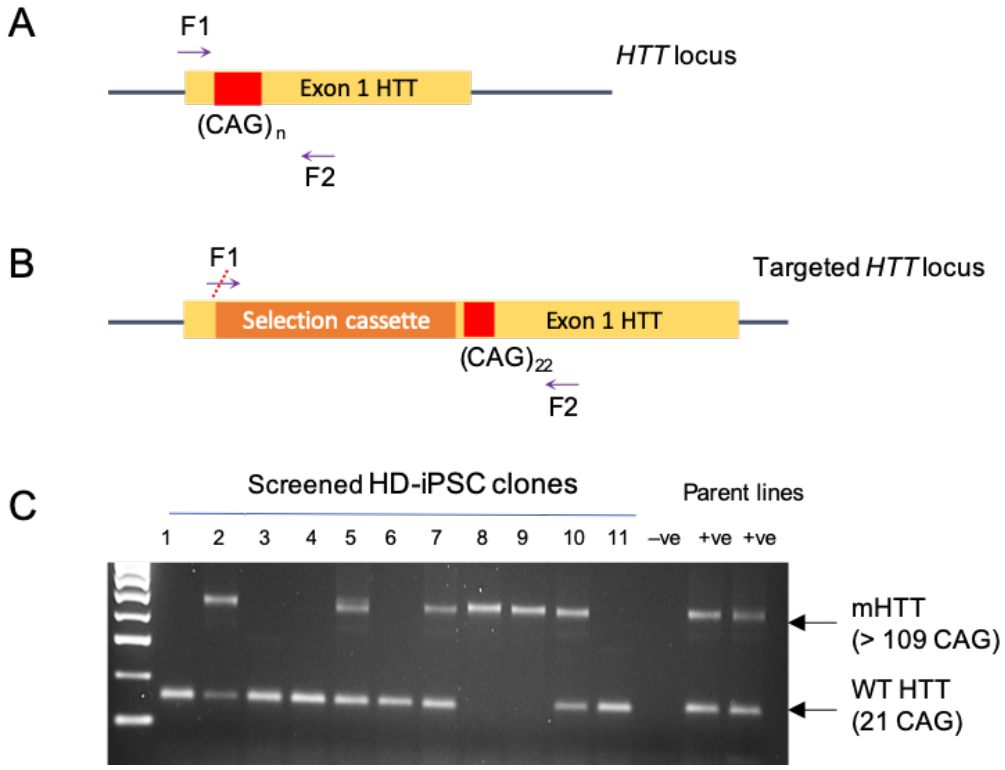


Figure 3.5 Targeted correction at the *HTT* locus in Q109 HD-iPSCs (A)Primers amplifying exon 1 of *HTT* iPSCs were used to screen for targeted clones. In the Q109 iPSCs, PCR products are generated from amplification of both the WT allele and the mutant allele. (B) Insertion of the targeting repair vector disrupts the forward primer binding site, and thus no PCR amplification is detected in the alleles that have been successfully targeted. (C) Diagnostic PCR screen for CAG repeat length identifies clones in which the WT allele or the mutant allele have been targeted.

3.3.2.3 piggyBac selection cassette removal

To remove the selection cassette, targeted HD-iPSCs were transfected with the piggyBac excision-only transposase vector, followed by negative selection with 1 μM ganciclovir (Figure 3.6A). All cells still harbouring the selection cassette containing the herpes simplex virus thymidine kinase gene would convert ganciclovir to its toxic by-product (Barese *et al.* 2012) and as a result the cell colonies would die, however it does not rule out mutations occurring inside the selection cassette making them immune to the ganciclovir. Resistant colonies were screened by PCR using primers amplifying the selection cassette, and clones with no PCR amplification were determined as free of the PB selection cassette (Figure 3.6B). This screen ensured that the targeting construct was removed from both the *HTT* locus anywhere else in the genome where random integration might have occurred.

Of note, on removal of the selection cassette from one of the targeted clones, the mutant *HTT* allele was present in the PCR screen over the *HTT* CAG repeat, suggesting that integration without homologous recombination had occurred in this clone. This clone was not included in the three isogenic clones we carried forward. There was difficulty removing the vector from another clone in which the mutant allele was targeted; iPSCs were transfected with the excision vector twice yet screening of each of the picked colonies revealed that the selection cassette was still present. As we already had three successfully targeted clones, we chose not to take this clone further.

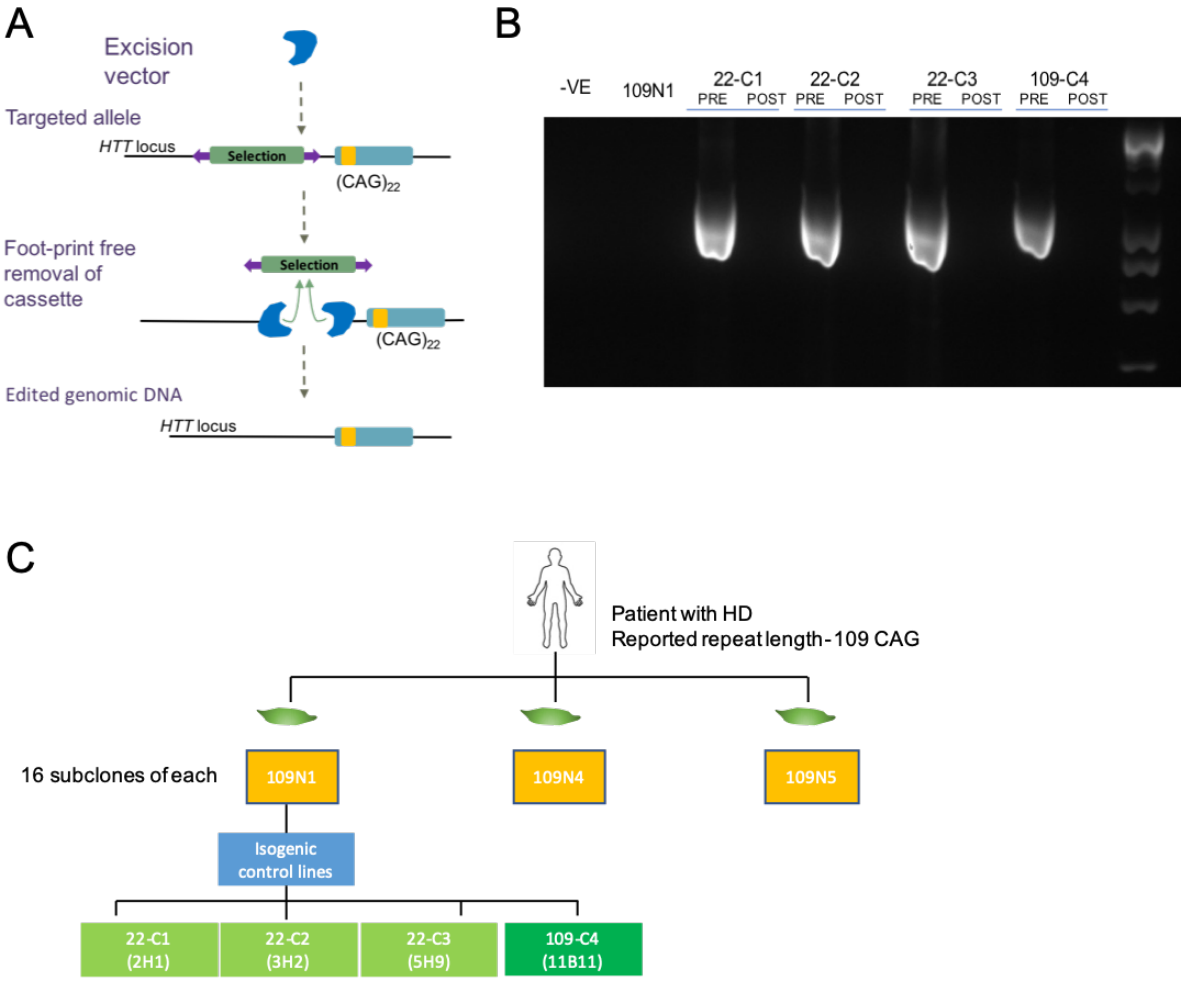


Figure 3.6 Excision of piggyBac piggyBac targeting vector from targeted iPSC clones. (A) Method of excision. (B) PCR screen amplifying selection cassette illustrates loss of selection cassette in resultant clones after excision. (C) Schematic depicting relation of gene-edited clones to parent 109N1 line. Yellow boxes indicate parent iPSC lines generated from three fibroblast cells derived from a patient diagnose at the age of 3 with Juvenile HD. The pale green boxes indicate corrected clones where mHTT was replaced with WT HTT, whilst the dark green box indicates a clone in which the WT allele was targeted and therefore retains the mHTT allele.

3.3.2.4 Footprint-free excision of the piggyBac selection cassette

Capillary electrophoresis and Sanger sequencing were performed to confirm the absence of the expanded CAG repeat tract in the corrected HD-iPSC clones (Figure 3.7). Resultant colonies were PCR amplified using another set of primers to the *HTT* locus, run on the GAGI30x and analysed by peak scanner. Representative traces from the software show the loss of the mutant HD allele in the corrected iPSCs and the return of the WT HD allele after cassette excision in the clones which had the WT allele targeted (Figure 3.7A).

Furthermore, Sanger sequencing analysis of the TTAA sites that flank the inverted terminal repeat sequences of the PB selection cassette confirmed effective excision of the selection cassette (Figure 3.7B).

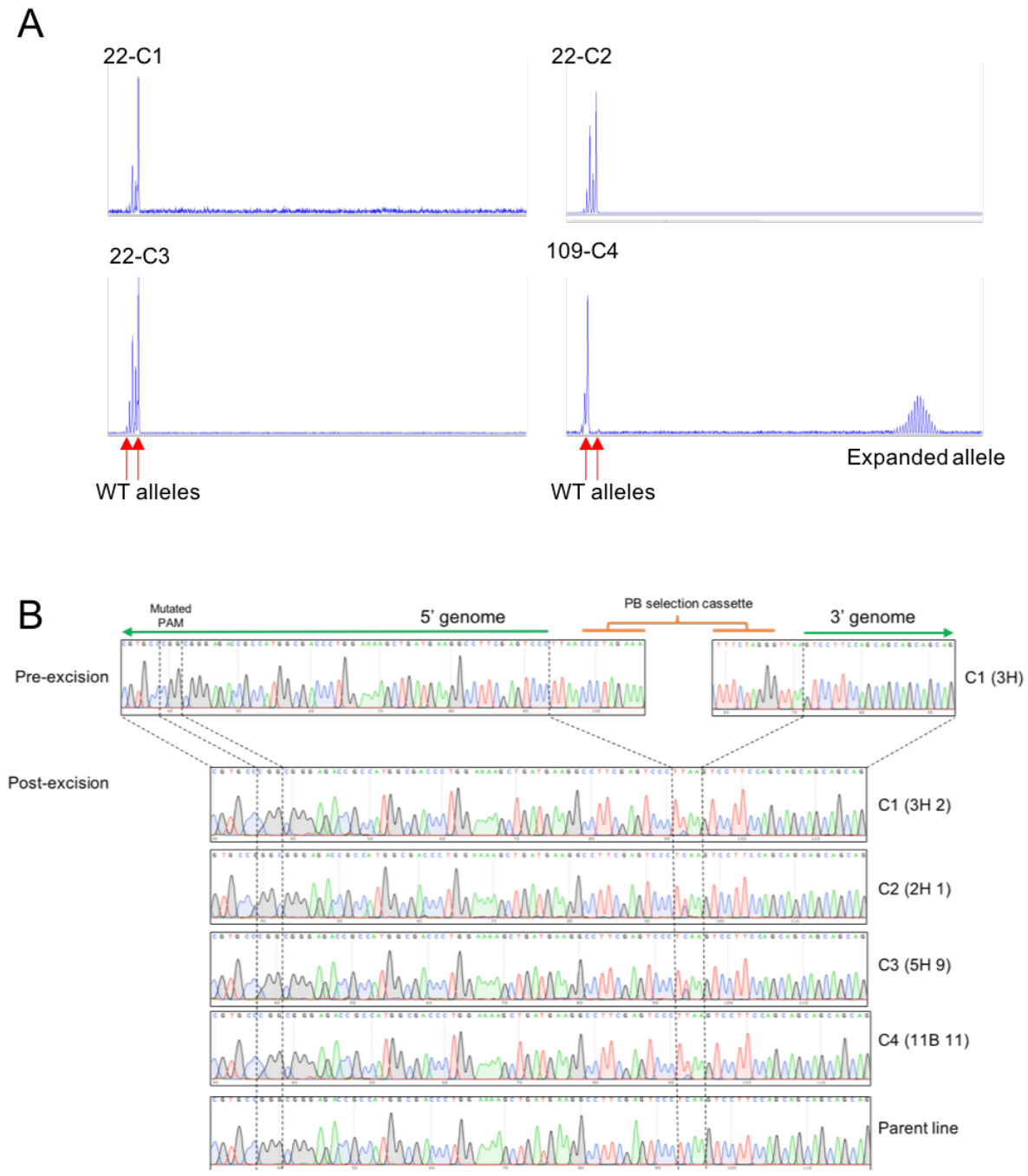


Figure 3.7 Generation of footprint-free edited isogenic HD-iPS cell lines. (A) Electropherograms of fluorescent PCR across CAG repeat show the WT allele present in all targeted HD-iPSC clones but show loss of expanded allele in 22-C1, C2, C3 clones. Clone 109-C4 retains the expanded repeat and the WT allele is re-inserted with a CAG repeat length +1 of original WT allele. (B) Sanger sequencing chromatograms confirm the specific PB cassette integration (upper), seamless excision around the TTAA integration site and GGG>CCG PAM site mutation introduction after excision (lower).

3.3.3 Whole exome sequencing

Whole exome sequencing was performed on genomic DNA from Q109N1, Q109N5 and all isogenic clones (Branduff McAllister), to check for any off-target effects of CRISPR-Cas9 editing process and whether single nucleotide variants had accrued through general cell culture maintenance. Exome libraries were sequenced using the HiSeq 400 and sequence reads were aligned to the human reference genome [Human GRCh37 (hg19) build] using Burrows-Wheeler Aligner (BWA). Variants were called using Genome Analysis Toolkit (GATK).

Whole exome sequencing identified a total of 106 variants that differed in at least one of the samples, and 34 of these were coding changes. Of these, 11 calls passed the GATK variant calling filter (Table 3.1). The genotypes are shown by either 0/0 (homozygote reference), 0/1 (heterozygote reference) or 1/1 (homozygote alternative allele). 6 non-synonymous coding variants were detected in 109N1 and 7 were detected in 109N5. 1 novel non-synonymous coding variant was identified in each of the gene edited clones 22-C1, 22-C2, 22-C3 and 109-C4, indicative of de novo mutations acquired through normal stem cell culture (Table 3.1). None of the variants identified were in genes known to be a modifier of HD phenotype, but perhaps unsurprisingly many appeared to be involved in signalling or growth pathways. It is well established that prolonged culture of iPSCs, especially under selective pressures, can lead to adaption such as increased growth rate or reduced apoptosis (Laurent *et al.* 2011).

Although the incidence of CRISPR/Cas9-induced off-target indels is low in human iPSCs (Suzuki *et al.* 2014; Veres *et al.* 2014) we performed unbiased detection of potential off-target sites predicted by the DESKGEN (www.deskgen.com), and the CRISPR Design tool from the Massachusetts Institute of Technology (www.crispr.mit.edu). Each off-target site was examined, and results confirmed the absence of any off-target indels.

N1	N5	22-C1	22-C2	22-C3	109-C4	VARIANT	GENE
0/1	0/1	0/1	0/1	0/1	0/1	6:24437445:G:A	GPLD1
0/1	0/1	0/1	0/1	0/1	0/1	6:24437458:T:C	GPLD1
0/1	0/1	0/1	0/1	1/1	0/1	6:160505207:A:G	IGF2R
0/1	0/1	0/0	0/1	0/1	0/1	7:100552371:C:G	MUC3A
0/0	0/0	0/0	0/0	0/0	0/1	11:9462016:G:A	IPO7
0/1	0/1	1/1	0/1	0/1	0/0	11:45241245:G:A	PRDM11
0/0	0/1	0/0	0/0	0/0	0/0	11:108357082:C:A	KDELC2
0/0	0/0	0/0	0/0	0/0	0/1	12:109998845:C:G	MMAB
0/0	0/0	0/0	0/0	0/0	0/1	18:48331573:C:T	MRO
0/0	0/0	0/0	0/1	0/0	0/0	19:56539724:G:T	NLRP5
0/1	0/1	0/1	0/1	0/1	0/0	21:47855916:C:T	PCNT

Table 3.1 Single nucleotide variants identified by whole exome sequencing. Single nucleotide variants are compared with the parent line Q109N1, and another clonal line Q109N5. The genotypes are shown by either 0/0 (homozygote reference), 0/1 (heterozygote reference) or 1/1 (homozygote alternative allele).

3.3.4 Virtual karyotyping of Q109 and corrected HD-iPSCs

All resultant clones were genotyped using Infinium PsychArray-24 chip which has a coverage of 593,260 single nucleotide polymorphisms (SNPs). This array produced a 'virtual karyotype' allowing us to detect gross chromosomal duplications, deletions or rearrangements and provided information about any copy number variants (CNVs) that had accrued throughout the targeting process. CNVs can be detected from a Log R ratio and B allele frequency. The Log R ratio uses a reference panel to determine the normal intensity expected at each probe (Salomón-Torres *et al.* 2015). A Log R ratio of zero indicates that the sample has the same copy number as the reference. An L R ratio above zero indicates a copy number gain, and an L R ratio below zero indicates a copy number loss. The B allele frequency represents the intensity of one allele compared with the other (Attiyeh *et al.* 2009). Homozygous SNPs have a B allele frequency of 0 (AA) or 1 (BB). Heterozygosity for the SNP (AB) has a B allele frequency of 0.5. A duplication will result in three copies of the SNP which has four potential genotypes (AAA, BBB, AAB, BBA) and so has a B allele frequency of 0, 0.33, 0.67, and 1 (Attiyeh *et al.* 2009).

After quality control (see 2.7.3) we identified two CNVs that were present in each of the samples (Table 3.2). Deletions were seen at 2q22.1 (Figure 3.8A) and 14q24.3 (Figure 3.8B), loci containing LDL receptor related protein 1B (*LRP1B) and* neurexin 3 (*NRXN3*), respectively. CNVs are considered a major source for human genome variability (Soldner *et al.* 2011). However, as these deletions were also present in both Q109N1 and Q109N5, it suggests these CNVs were present in the patient and are not due to the gene-editing process itself.

At later passages we report a duplication of chromosome 1 in the Q109N1 parent line (Figure 3.9D), and the corrected-iPSC daughter clones. This finding suggests a pre-disposition to the duplication of chromosome 1, which might have functional consequences for phenotypic assessment.

Total CNVs	CNV Locus	1p36.33-q44	2q22.1	14q24.3
	Approx Position	chr1: full chromosome	chr2:141448312-141932472	chr14:77984540-78108110
	Status	Duplication	Deletion	Deletion
	Length	248,511,674	484,161	123,571
22-C1	Passage 3			
	Passage 13			
22-C2	Passage 3			
	Passage 7			
	Passage 19			
22-C3	Passage 3			
	Passage 12			
	Passage 20			
109-C1	Passage 3			
	Passage 20			
	Passage 28			
	Passage 36			
109N1	Passage 17			
	Passage 39			
109N5	Passage 18			

Table 3.2 CNV analysis performed after genotyping targeted clones on PsychArray. Final CNV analysis reveals deletions at 2q22.1 and 14q24.3 in all samples, and duplication of chromosome 1 in various samples at later passages. Yellow highlighted squares indicate that the sample has the CNV highlighted.

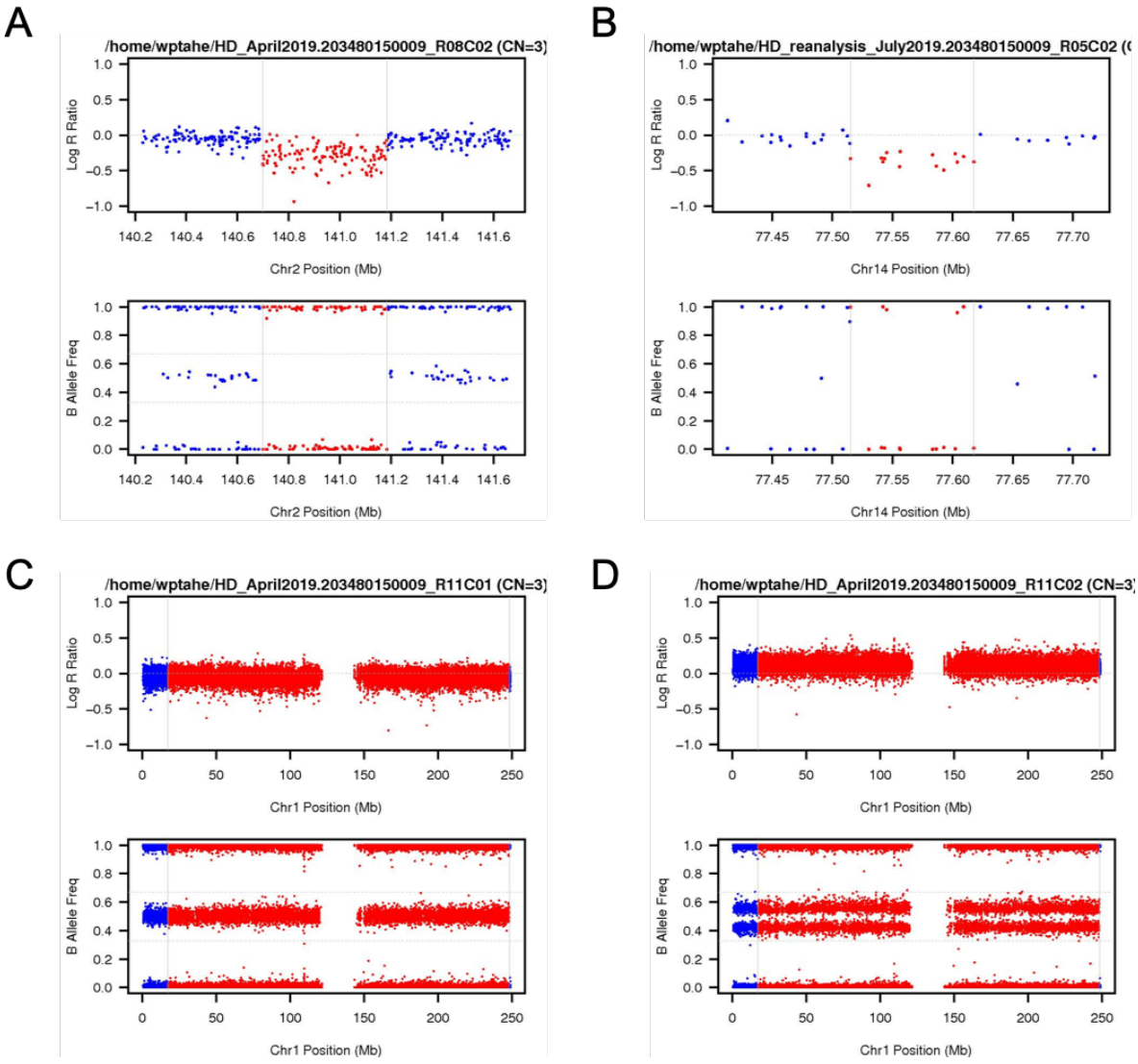


Figure 3.8 SNP array plots of corrected and HD-iPSCs. (A) Deletion at 2q22.1 in all samples, parent and daughter clones, characterised by a decrease of the Log R ratio below zero and an absence of SNPs around 0.5 in the B allele frequency plot (B) Deletion at 14q24.3 in all samples. (C) karyotypically normal plot of chromosome 1 characterised by a Log R Ratio of 0 and a B allele frequency of 0.5, indicating two copies of the chromosome. (D) Duplication of chromosome 1 characterised by an increase of the Log R Ratio above 0 and a split in the B allele frequency to give two bands around 0.4 and 0.6, indicating three copies of the chromosome

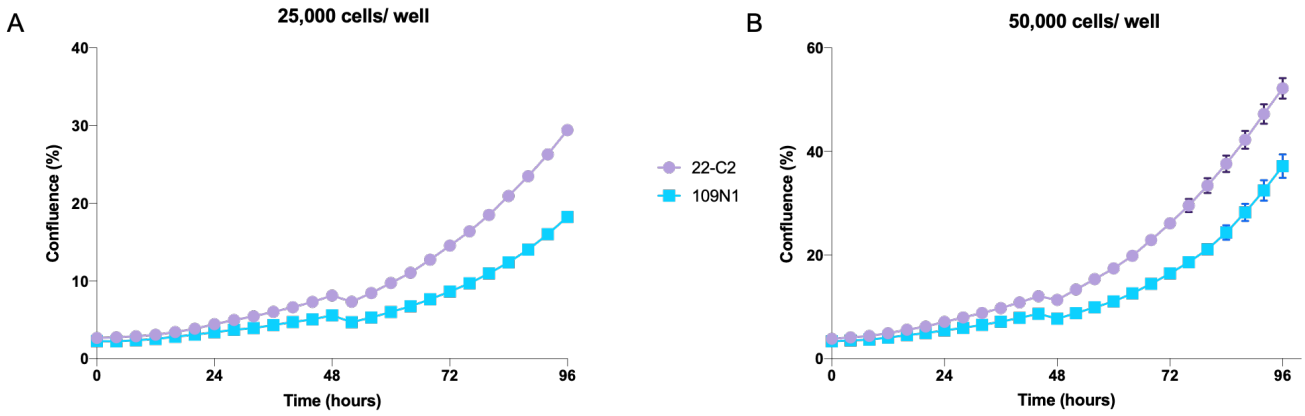
3.3.5 Proliferation rates in Q109 and corrected-iPSCs

Reports of the effect of *HTT* CAG repeat lengths on proliferation rates in HD-iPSCs have been varied, with some reports noting no effect of CAG repeat length on replication rates and others reporting differences. To investigate whether genetic correction of the CAG repeat affects proliferative properties of the iPSCs we used the IncuCyte-Zoom cell confluency assay (see 2.9.3). iPSCs were dissociated to a single cell suspension using Accutase and cells were seeded into warm media with 10 µM Y-27632 at a density of 25,000 or 50,000 in each well of a 12 well plate. Two densities were chosen to examine whether proliferation is affected by cell density as well as genotype. Cell confluency (% of the surface area of cells) was measured every 2 hours for 96 hours.

There was a significant effect of genotype on proliferation rates in cells plated at both densities. Cells plated at a density of 25,000 cells/well had doubling rates of 28.5 hours and 34.2 hours for 22-C1 and 109N1, respectively (Figure 3.9A). This was significant after Two-way ANOVA followed by Sidak's post-hoc testing ($p<0.0001$). Cells plated at a density of 50,000 cells/well had doubling rates of 25.7 hours and 29.1 hours for 22-C1 and 109N1, respectively (Figure 3.9B). Cell doubling rates were lower in cells plated at a density of 50,000/well. Cell media was changed after 48 hours giving rise to the small dip that is evident on the graph.

To ensure that the confluence mask applied was accurately measuring proliferation, and not being skewed by cell size, average cell size was determined. Cells were dissociated to a single cell population using accutase, centrifuged, and resuspended in warmed ADF media. 10 µl of the single cell suspension was mixed in equal amounts with Trypan Blue and added to a Countess II FL Automated Cell Counter slide. Average cell size was automatically determined by the Countess for any cells that did not stain positive for Trypan blue. Average cell sizes for 22-C1 and 109N1 were 15.43 µm and 15.61 µm, respectively, suggesting that the confluence mask applied would not be affected by cell size. However, it should be considered that cell spreading might differ between the lines, even if the cell volumes in flow are similar.

These data are preliminary results from one experiment with one cell line from each genotype. This experiment should therefore be repeated in the remaining isogenic and HD-iPSCs to accurately determine the effect of genotype on proliferation.



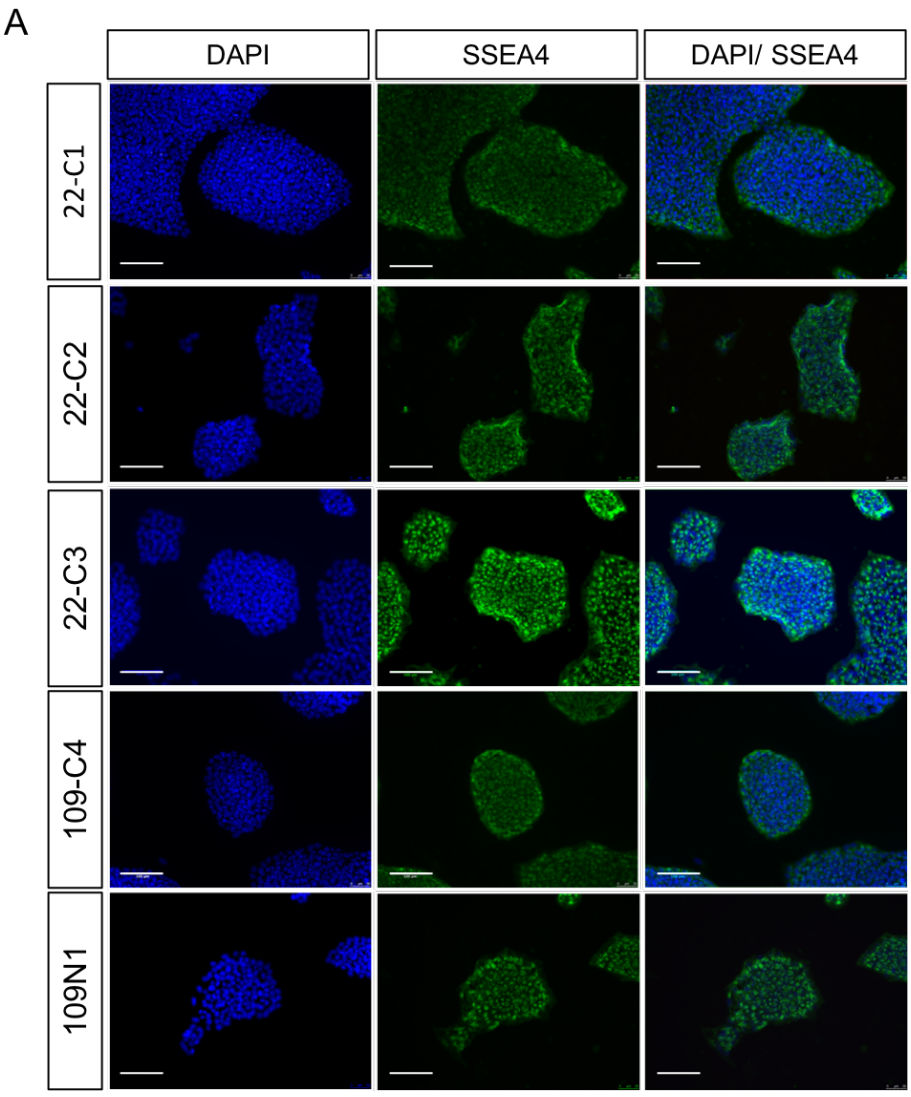
C

Doubling rates	25,000 cells	50,000 cells
22-C2	28.5 hrs	25.7 hrs
109N1	34.2 hrs	29.1 hrs

Figure 3.9 Proliferation rates in gene corrected iPSCs. Cell confluency (%) was calculated using Incucyte Zoom Software by phase-contrast images. Each data point represents triplicate wells of cells plated at (A) 25,000 cells/well, $p<0.0001$ and (B) 50,000 cells/ well, $p<0.0001$. p-values determined by Two Way ANOVA followed by Sidak's multiple comparisons test. (C) Doubling times of iPSCs plated at either 25,000 or 50,000 per well. Values expressed as Means ± SEM with 3 technical replicates. n= 1 independent experiment.

3.3.6 Characterisation of pluripotency in Q109 and corrected-iPSCs

We then examined whether the pluripotent characteristics of the parental HD-iPSC line, previously shown to express pluripotency markers, were maintained in the corrected iPSC lines. Indeed, all corrected iPSC lines stained positive for pluripotency markers OCT4 and SSEA4 (Figure 3.10) and had similar mRNA levels of the pluripotency genes *OCT4* and *NANOG*, whilst expression levels of *SOX2* were significantly higher in one clone, 22-C1 (Figure 3.10C).



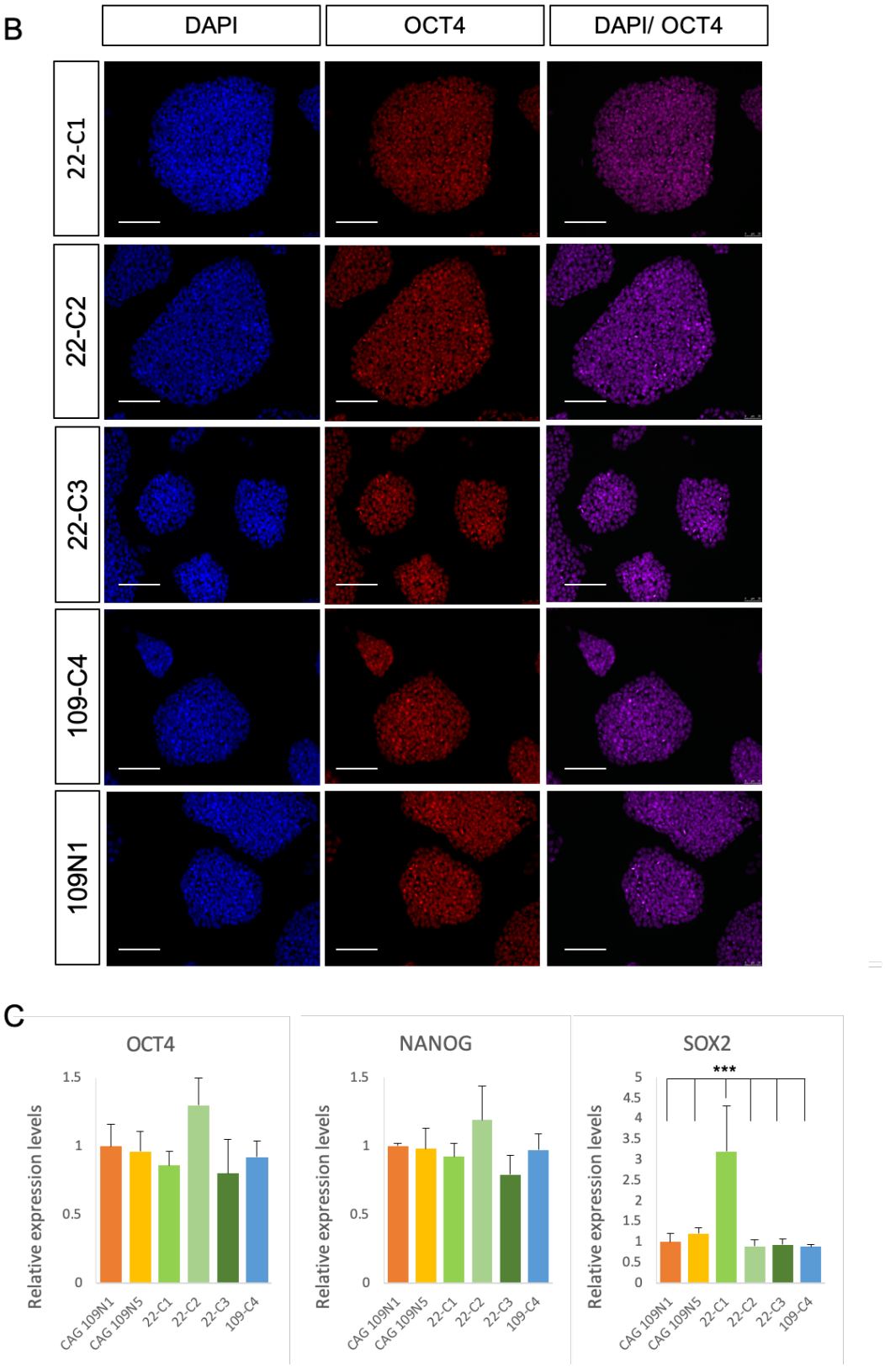


Figure 3.10 Q22 iPSCs retain pluripotency following genetic correction. Undifferentiated iPSCs stained positive for the pluripotency markers (A) SSEA4 and (B) OCT4. Scale bars = 100 µm. (C) The expression levels of pluripotency genes *OCT4*, *NANOG* and *SOX2* are comparative across parent lines and targeted lines. Data are normalised to *GAPDH* and *β-actin*. Expression values are relative to the parent line Q109N1-iPSC RNA. Asterisks indicate significant differences from Q22-C1, ***p< 0.001. determined by One-Way ANOVA followed by Tukey HSD test. n=3 per clone, values for independent technical replicates shown as mean ± SEM.

3.3.7 Neural rosette formation

Neural rosettes are radially organised cellular structures with a central lumen resembling the developing neural tube (Wilson and Stice 2006; Elkabetz *et al.* 2008). NPCs organise into neural rosettes both *in vivo* and *in vitro.* The self-organisation of NPCs into neural rosettes *in vitro* is thought to mimic neural tube formation during *in vivo* development (Prajumwongs *et al.* 2016). HTT has been implicated in the regulation of this morphogenetic self-organisation during neural development, as evidenced by the irregular organisation of rosette structures formed by $HTT^{-/-}$ lines in contrast to the highly regular organisation that WT HTT lines display (Ruzo *et al.* 2018). Consistent with this, previous reports have indicated that NPCs generated from HD cell lines demonstrate reduced cell adhesion and form impaired or fewer neural rosettes (Jeon *et al.* 2012; The HD iPSC Consortium 2012; Xu *et al.* 2017). Despite these findings, the failure of NPCs to form neural rosettes does not appear to effect the differentiation potential of HD-iPSCs/ ESCs carrying an expanded CAG repeat (Jeon *et al.* 2012; Xu *et al.* 2017).

I early study reported no change in neural rosette formation in NPCs derived from HD-ESCs with an expanded *HTT* CAG repeat (Niclis *et al.* 2009). In our hands both 22-C1 and 109-C4 demonstrate the ability to form neural rosettes from day 12 onwards (Figure 3.11) when following a monolayer differentiation protocol (see 2.1.3). Rosette structures stained positive for beta-tubulin, a neuronal microtubular element marker and ZO-I a marker for tight junction proteins. We do however observe irregularity in the spatial distribution of neural rosettes in NPCs with an expanded *HTT* CAG repeat compared with those with a WT length repeat. Smaller lumen sizes have been reported in rosettes formed from HD cells (Ruzo *et al.* 2018), though we find a larger lumen size. Due to time restraints we did not take phenotypic assessment of neural rosette formation further. To assess whether phenotypic differences exist between genotypes, a more extensive characterisation of neural rosettes should be performed, assessing their organisation and spatial distribution, the number of rosettes and the lumen size.

It is worth noting that variations occurred between differentiations, one possible reason for which could be due to cell density when neural induction was started. This was evidenced with two differentiations held in parallel, some wells were split at a 2:3 ratio at day 8 whilst other wells were not split and carried through to day 16 in the same plate. Cells split at day 8 did not show rosette formation in either cell line, but cells not split at day 8 did develop rosettes that appeared between days 12 and 16, irrespective of genotype.

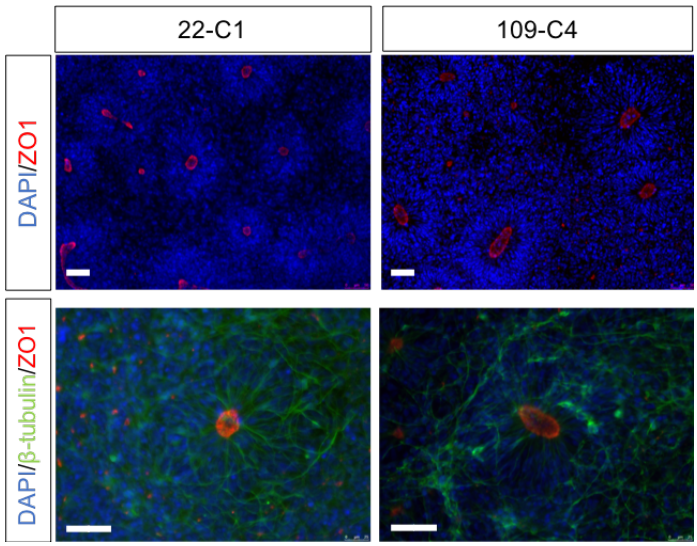


Figure 3.11 Assessment of neural rosette formation following neural induction. Immunostaining of day 12 cultures showed uniform expression of neuroepithelial markers in all clones irrespective of CAG repeat length. Staining with the microtubular element Beta-tubulin (green), and tight junction protein ZO-1 (red) identifying neural tube-like lumens, reveals neural rosette structure. All nuclei are counterstained with DAPI (blue). Scale bars = 50µm.

3.3.8 Differentiation of HD and corrected-iPSCs into neurons

Studying MSNs derived from HD-iPSCs may shed light on pathogenic mechanisms contributing to HD. Each clonal line grew well as a monolayer and following 16 days of neural induction and patterning each stained positive for nestin, an intermediate filament expressed by NPCs (Lendahl *et al.* 1990) (Figure 3.12A). Forkhead Box G1 (*FOXG1*) and GS Homeobox 2 (*GSX2*) were also highly expressed at day 16, markers of forebrain progenitors and lateral ganglionic eminence (LGE) markers respectively (Figure 3.12C) (Delli Carri *et al.* 2013; Greig *et al.* 2013).

NPCs were further directed towards a striatal fate as shown by immunostaining with mature neuronal marker microtubule-associated protein 2 (MAP2) and neurotransmitter gamma-aminobutryic acid (GABA) (Figure 3.12B), and expression of MSN markers COUP-TF-interacting protein (*CTIP2*) and dopamine- and cAMP- regulated neuronal phosphoprotein (*DARPP-32*) (Figure 3.12D). Striatal gene expression gradually increased over time in the differentiating cultures, however significance could not be established as only one sample of each clonal line was ascertained. Reliably establishing the differentiation potential of Q109 and Q22 iPSCs into MSNs is essential, immunocytochemistry should be used to confirm proportion of DARPP-32 and CTIP2 positive neurons, but this was not performed in this thesis.

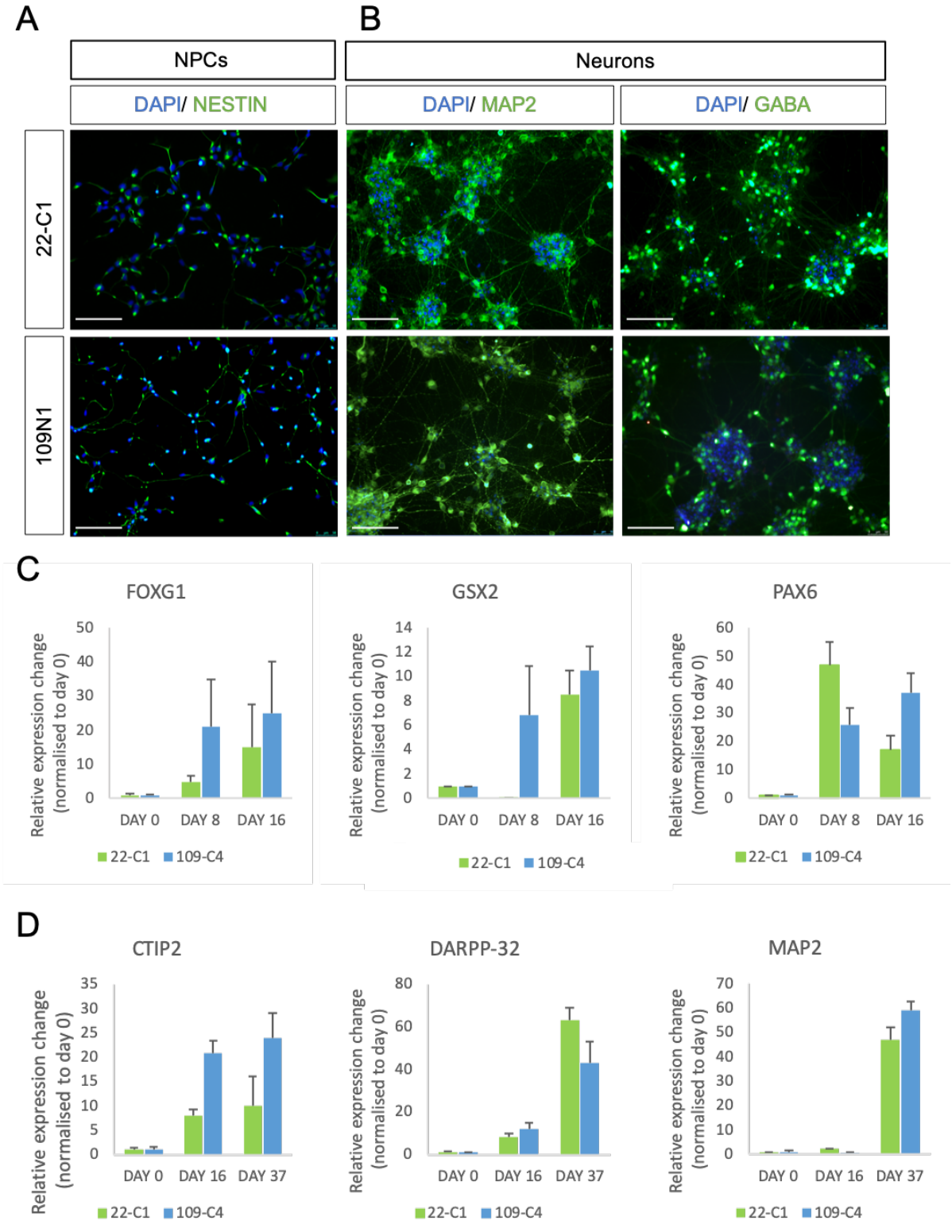


Figure 3.12 Differentiation of Q109 and Q22 iPSCs into NPCs and neurons. (A) Immunostaining at day 16 with NPC marker nestin (green). (B) Further differentiation of NPCs to day 37 results in neurons staining positive for neuronal marker MAP2 (green) and inhibitory neuron marker GABA (green). All nuclei are counterstained with DAPI (blue). Scale bars = 100µm. (C) NPC markers *FOXG1*, *GSX2* and *PAX6* relative expression levels at days 8 and 16 (D) Neuronal marker *CTIP*, *DARPP-32* and *MAP2* relative expression levels at days 16 and 37 (n= 1 per clone, values for independent technical replicates shown as mean ± SEM). All graphs are relative to iPS cells and plotted as 2^-ΔΔCT method.

3.3.9 Reduced mitochondrial bioenergetics are rescued in corrected-NPCs

Another commonly reported phenotype in HD is altered mitochondrial energetics (Seong *et al.* 2005). Early reports noted significant defects in the enzymatic activity of complexes in the electron transport chain in post-mortem striatal tissue of HD patients (Brennan *et al.* 1985; Browne *et al.* 1997). Reduced mitochondrial membrane potential and reduced mitochondrial calcium buffering capacity are also well documented in HD mouse models and lymphoblastoid cell lines (Sawa *et al.* 1999; Panov *et al.* 2002; Milakovic *et al.* 2006). Mitochondrial impairments are also seen in HD-iPSC and ESC derived neural cells, with reports of lower basal respiration rates and reduced ATP production (An *et al.* 2012; Xu *et al.* 2017; Ooi *et al.* 2019). Using the Seahorse XF96 extracellular flux analyser to measure oxygen consumption rate (OCR) and extracellular acidification rate (ECAR) we can get an idea of mitochondrial respiration in the NPCs. The integrated drug delivery system allows various complexes of the electron transport chain to be probed through the addition of the mitochondrial stressors oligomycin, FCCP, rotenone and antimycin A and the resultant changes in OCR examined (see 2.9.4). Mitochondrial basal respiration, maximal respiration and ATP production can then be determined from the changes in OCR (Figure 3.13).

Oligomycin is an ATP synthase (complex V) inhibitor, the addition of which prevents the flow of electrons through the electron transport chain, thereby reducing ATP-linked respiration (Divakaruni *et al.* 2014). FCCP is an uncoupling agent which activates proton conductance through the plasma membrane (To *et al.* 2010). Increased proton conductance causes the proton gradient to collapse and disrupts the mitochondrial membrane potential. As a result, electrons are still transferred through the electron transport chain and oxygen is consumed by complex IV at a maximal rate, yet no ATP is produced. The change in OCR following FCCP addition can therefore give an indication of the spare respiratory capacity and the ability of the cell to respond to an increased energy demand under stress (Divakaruni *et al.* 2014). Rotenone is a complex I inhibitor, preventing the transfer of electrons from the iron-sulphur centres of complex I to ubiquinone (Palmer *et al.* 1968; Heinz *et al.* 2017), whilst antimycin A is a complex III inhibitor, inhibiting oxidation of ubiquinol which prevents electron transfer between cytochrome b and c (Alexandre and Lehninger 1984; Maguire *et al.* 1992). Rotenone and antimycin A delivery together inhibits mitochondrial respiration and allows the determination of maximal respiration and non-mitochondrial respiration.

NPCs were plated into the Seahorse XF96 plates at a density of 80,000 cells/ well (see 2.9.4). Basal respiration rates were elevated in corrected NPCs compared with Q109 NPCs (Figure 3.13B). Both ATP production (Figure 3.13C) and maximal respiration (Figure 3.13D) were increased in NPCs with a CAG repeat length of 22 compared to those with a repeat length > 109 CAG. Spare respiratory capacity was also increased in corrected NPCs (Figure 3.13E) suggesting that cells with a WT repeat length might cope with increased energy demands of the cell more appropriately than cells with an expanded repeat.

These data suggest neural cells derived from HD-iPSCs demonstrate mitochondrial impairment, but this can be rescued following genetic correction of the expanded *HTT* CAG repeat.

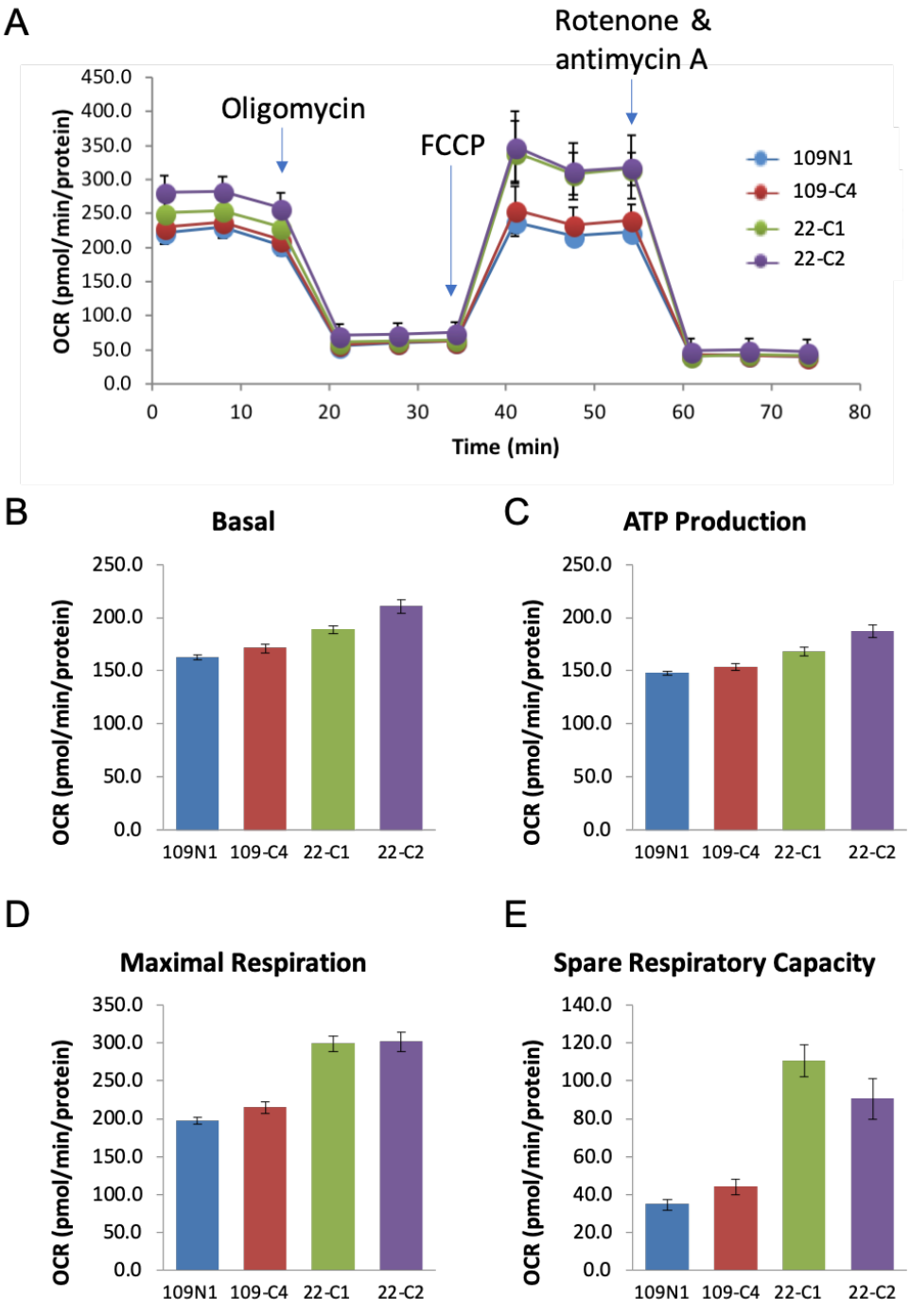


Figure 3.13 Mitochondrial energetics in corrected and HD-NPCs. (A) Representative Seahorse Bioanalyser traces showing oxygen consumption rates (OCR) in corrected and HD-NPCs (80,000 per well). Quantitative data showing (B) Basal respiration rates (C) ATP production (D) Maximal respiration rates and (E) Spare respiratory capacity in NPCs derived from corrected-iPSCs or HD-NPCs. n=1 for each cell line, with at least 14 technical replicates. Values expressed as Means ± SEM. OCR measurements normalized to absorbance at 560 nm.

3.4 Discussion

This chapter describes the development of a HD-iPSC model of CAG repeat expansion with which to assay the effects of DNA repair genes on *HTT* CAG expansion and downstream phenotypes. To establish isogenic models to accurately examine the effects of these modifiers on cells harbouring an expanded CAG repeat, we describe the seamless genetic correction of HD-iPSCs from a *HTT* CAG repeat length of >109 to a WT repeat length of 22 using a CRISPR-Cas9 and piggyBac transposon-based homologous recombination approach. Corrected HD-iPSCs maintained a normal 'virtual' karyotype, pluripotency and were able to differentiate into the neuronal subtypes most susceptible to degeneration in the disease. Furthermore, mitochondrial dysfunction, a phenotype previously reported in the HD-iPSCs, was rescued following genetic correction of the expanded *HTT* CAG repeat.

HD-iPSCs maintain the expanded CAG repeat and in combination with their directed differentiation into multiple cell types, including disease relevant neuronal cell types, provide a unique opportunity to illuminate the underlying cellular pathology. Q109 iPSC lines show expansion of the *HTT* CAG repeat across multiple passages, in multiple subclones, and upon neuronal differentiation, supporting the findings from the HD stem cell consortium who also observed expansion in the Q109N1 line across multiple passages of NPCs (The HD iPSC Consortium 2012). The cell system therefore represents a suitable model for assaying CAG repeat expansion which we hope will provide us with relevant tools to dissect the effects of genetic modifiers on cells harbouring an expanded *HTT* CAG repeat. The Q60 line displayed little mosaicism and only one subclone demonstrated an expansion > 1 *HTT* CAG repeat when cultured, possibly reflecting the length-dependent bias of CAG repeat expansions. Indeed, cells harbouring longer CAG repeats are more likely to undergo further expansions in culture. Somatic instability of the *HTT* CAG repeat was found to be more pronounced in the post-mortem brains of juvenile cases of HD than adult onset cases of HD (Telenius et al. 1994; Aronin et al. 1995). This is echoed in other triplet repeat diseases such as spinocerebellar ataxia 1 (Chong et al. 1995) and spinal and bulbar muscular atrophy (Tanaka *et al.* 2002) where there is increasing mosaicism with longer CAG repeat lengths. In iPSCs derived from Myotonic dystrophy type 1 (DM1) patients it was proposed that the interval between 57 and 126 CAG repeats was an important length threshold where triplet repeat expansion rates dramatically increased (Du *et al.* 2013). What we do not know is whether there are different mechanisms involved in the expansions of long and short repeats, or whether the same pathways for expansion are used under all conditions, and in all cell types.

Alternatively, the Q60 cell line may exhibit fewer expansions than the Q109 line due to inherent variability in genetic background between the two patient cell lines, or differential expression levels of genes. Whilst our approach for assaying genetic modifiers of HD is candidate driven, it would be

interesting to perform RNA sequencing and exome sequencing on subclones from both lines as this could potentially reveal novel targets.

Although the exact mechanisms behind repeat expansions are still not known, recent human genetic data suggests that DNA repair pathways play a central role in disease pathogenesis and a plausible mechanism is that this occurs through somatic expansions (GeM-HD Consortium 2015; Bettencourt *et al.* 2016; GeM-HD Consortium 2019). Repetitive DNA sequences can form unusual secondary structures to which mismatch repair proteins (MMR) bind and, in the process of repair, cause somatic instability of the CAG repeat (Usdin and Grabczyk 2000; Mirkin 2007). Indeed, knock-out of several MMR proteins attenuates somatic CAG expansion (Wheeler *et al.* 2003; Dragileva *et al.* 2009). Later chapters will discuss the role of DNA damage and repair in the context of the *HTT* CAG repeat further.

The expansions observed in the Q109 lines are reproducible across a number of subclones and occur in iPSCs at higher frequencies than previously reported in cell models (Manley, Pugh, *et al.* 1999; Cannella *et al.* 2009; Niclis *et al.* 2009). In this study, substantial expansions were seen over 4 weeks in culture making them an attractive system compared with mouse models, due to the considerably shorter time course that an experiment could conceivably run for. The main interests of our group are to examine the role that genetic modifiers of HD, identified by human genetic studies, might have in potentiating or preventing CAG repeat expansions, therefore this time-course provides a window with which the effects of modifiers in CAG repeat expansion can be assayed. Combined with this, iPSCs are physiologically relevant, easier to genetically manipulate and higher throughput than animal models, making HD-iPSCs a desirable cell line to work with when assaying genetic variants.

Despite these obvious advantages, we do observe variability in *HTT* CAG expansion rates amongst subclones derived from the same clonal line. Some subclones showed considerable CAG repeat expansion, whilst others showed little to no CAG repeat expansion under the same culture conditions. These differences in expansion rates suggest a stochastic mechanism perhaps driven by genomic integrity or epigenetic regulators. Single nucleotide variants, copy number variants and chromosomal rearrangements represent sources of genetic variation commonly reported in iPSC studies. It is therefore imperative to use multiple subclones for each phenotype under investigation.

As human brain tissue is only available post-mortem, it is difficult to link *HTT* CAG somatic expansions directly with HD progression (Swami *et al.* 2009). Whilst mouse models also face the same problem, multiple time points may be taken and the *HTT* CAG expansion observed can be correlated to behavioural phenotype. But unlike mouse models, iPSCs (and the neurons derived from them) provide an unlimited renewable resource of cells that can be sampled at different stages of differentiation and maturity. This provides us with a detailed look at the rate at which expansions occur and how these

affect the HD phenotypes seen in iPSCs, theoretically allowing us to accurately detect phenotypes resulting from different CAG repeat lengths.

Perhaps unsurprisingly, the rate of expansions observed is higher in iPSCs than neurons, though overall are still high. Proliferating stem cells are subject to DNA damage induced by replication mishaps and stress responses imposed when manually passaging. This likely contributes to the higher *HTT* CAG expansion rates seen in dividing cells (iPSCs) compared with post-mitotic neurons. However, in this cell system we do observe neuronal *HTT* CAG expansions, suggesting that the expansions are not solely division-dependent. Whether there are distinct pathways responsible for driving expansions in mitotic and post-mitotic cells is still not known. Whilst neurons are exempt from replication-related DNA repair, they are subject to DNA damage from endogenous metabolic processes. For example, the brain has high oxygen demands and consequently high levels of reactive oxygen species (ROS) are produced through normal mitochondrial respiration (Cooke et al. 2003). ROS are damaging to both mitochondrial and nuclear DNA (Yu 1994; Arun *et al.* 2015; Liu *et al.* 2017).

Despite the ubiquitous expression of HTT, neuronal populations are most susceptible to disease degeneration. Here we show CAG repeat expansions and mitochondrial dysfunction in neurons and NPCs differentiated following a striatal protocol. The reasons for differentiating HD-iPSCs towards a neuronal fate was to examine the expansion properties in a disease-relevant cell type. It should be noted that extensive characterisation of neuronal subtype was not performed, and so we cannot conclude the proportion of MSNs we were sampling. It is of interest to direct the iPSCs into other lineages to look at expansions and associated phenotypes. A recent study looked at the tissue-specific effects of CAG repeat length on somatic expansion in an allelic HD-hESC series (Q30, Q45, Q65, Q81) differentiated to myotubes, neurons and hepatocytes (Ooi et al. 2019). No expansion was seen in hepatocytes or myotubes at any repeat length, but mild instability was seen in NPCs and neurons derived from the Q65-hESC line. Whilst this is not congruous with the high expansion rates seen in the livers of transgenic mice (Lee et al. 2011), it provides some support for the selective vulnerability of neurons to the disease. The differences seen between HD-ESCs and mouse models might reflect the finding that the unstable repeats are highly enriched in polyploid hepatocytes, suggesting that the unique population of polyploid cells reflect a state where repeat expansions are permissible (Lee et al. 2011). hESCs represent a very 'clean' population of cells, with little heterogeneity. Cardiovascular symptoms often present in early stage HD patients, however differentiation of HD-hESC lines to cardiomyocytes revealed no effect on CAG expansion length (Jacquet *et al.* 2015). Maturity of the iPSC derived cardiomyocytes and hepatocytes might affect this.

In proliferating cells, after a number of passages, we observed an intra-allele bimodal distribution, indicating that a large (> 10 CAG) contraction or expansion occurred between time-points. A bimodal

peak distribution following an expansion might reflect a proportion of cells that are expansion-prone whilst others may not be. It may be that a single cell undergoes a single large expansion or contraction, and with multiple rounds of division, a small subpopulation of cells with a shorter/longer repeat length is then detectable by bulk PCR. What is driving expansions or contractions in a subset of cells is still not clear. Whether these repeat length changes confer a growth advantage in some way would be interesting to investigate. Developing single-cell sequencing methods to determine CAG length in iPSCs, NPCs and neurons would provide information about cell population diversity.

There are contradictory reports as to whether *HTT* CAG repeat length affects cell proliferation rates. Our preliminary data suggest that there is a *HTT* CAG-repeat-length dependent effect on iPSC proliferation rates, with the Q22-C2 cell line demonstrating significantly higher growth rates than the Q109N1 cell line. This is consistent with a recent paper that reported a length-dependent decrease in proliferation rates in an allelic HD-ESC series with *HTT* CAG repeats ranging from 35 to 85, suggesting that shorter repeat lengths do convey a growth advantage under certain conditions or cell types (Ooi *et al.* 2019). Conversely, proliferation rates were not affected when the HD mutation was genetically corrected from a *HTT* CAG repeat length of 72 to 20 in HD-iPSCs (An *et al.* 2012). Whilst our data does hint towards an effect of *HTT* CAG repeat length on replication rates, the experiment should be repeated further with the remaining isogenic and HD-iPSCs to accurately determine effect. Whilst the use of the Incucyte to measure cell proliferation was simple, it was likely not entirely accurate as it relied on the use of a confluence mask to establish proliferation rates. Although we determined similar cell sizes between Q22 and Q109 when the cells were in flow, these may not be representative of cell spreading when the cells are attached. We are optimising more accurate assays to determine cell proliferation in our cell lines, including 5-ethynyl-2'-deoxyuridine (EdU) kits. Edu is a nucleoside analogue to thymidine is incorporated into DNA during DNA synthesis, and this can produce a fluorescent read out for more sensitive determination of cell proliferation (Chehrehasa *et al.* 2009).

If our hypothesis is correct, we expect the cells harbouring the longest expansions to die first, thus the frequency of large expansions in post-mortem brain samples is likely an underestimation. Whilst in our cell system we are able to routinely measure CAG repeat lengths, if CAG expansion does reduce cell viability, we might miss these events in a neuronal culture. Collecting media at each passage to analyse the CAG repeat lengths in cells that have died seems logical but is time-consuming and difficult with small yields. Whilst bulk-PCR is extremely advantageous for high-throughput repeat length analysis, small-pool PCR would inform us of any rare but large expansions that might be missed with bulk-PCR. Using small-pool PCR we would get an idea of the repeat expansion dynamics in culture over time i.e. how mosaicism changes at each passage and the relative frequency by which expansions/ contractions

occur in a given population. Small-pool PCR would also allow you to follow these rarer large expansions over time to understand whether they are more prone to cell death.

Another point worth noting is that when we measure CAG repeat length in a neuronal culture, we are measuring that in a mixed population of neurons. Neuronal cultures are mixed populations of neuronal subtypes and it's estimated only ~20-40% of the culture will be MSNs. This raises the question as to whether we are diluting the phenotype we see with other neuronal subtypes in the population. Whilst the cortex also shows high levels of somatic instability (Swami *et al.* 2009), it could be that we are selectively losing the neurons undergoing the largest expansion changes (which may be the MSNs) and reading the CAG repeat lengths of different neurons. This is a limitation many face when attributing phenotypes specifically to one neuronal subtype, particularly when taking electrophysiological recordings from neurons that are not necessarily integrated into relevant physiological neural networks. Though this assumes higher MSN vulnerability is intrinsic, when it may relate to the interaction with other neurons, such as dopamine and glutamate innervation and excitotoxicity.

As mentioned above, the need for appropriate controls is imperative so that interpretations of phenotypes can be made accurately. Isogenic HD-iPSCs and ESCs have been generated through a number of different methods, CRISPR-Cas9/ piggyBac being one of them. The piggyBac transposon was chosen for its ability to integrate large amounts of cargo and seamlessly excise, leaving no or minimal 'scars' which have been found to influence expression of the gene it integrates into.

Virtual karyotyping using a SNP array is a relatively inexpensive method with high sensitivity for detecting CNVs as small as 100 kb, a resolution 50-fold better than karyotyping (Wapner *et al.* 2013). This technique allows a high throughput karyotyping pipeline, where iPSCs can be assessed through culture to ensure they are maintaining a normal CNV/ chromosome profile and should be run routinely. To what extent the SNP arrays can inform us of somatic CNVs in subpopulations of cells in an iPSC line is still not known, and therefore traditional karyotyping should still be performed on a reduced number of clones. Additionally, traditional karyotyping methods would provide information on the visual appearance of the chromosomes and chromosomal abnormalities, which would not be detected on an array. All samples contained a deletion at 2q23.1, which contains the gene *NRXN3,* variants in which have been associated with AD, autism, addiction and schizophrenia. NRXN3 is a presynaptic adhesion molecule involved in regulating neurotransmitter release (Hishimoto *et al.* 2019). A deletion was also seen on chromosome 2 in the *LRP1B* gene. LRP1B is a putative tumour suppressor, mutations in which have been associated with a number of cancers (Langbein *et al.* 2002; Prazeres *et al.* 2011; Ni *et al.* 2013). Both of the CNVs identified were present in the parent line, Q109 N1 and Q109 N5, suggesting that these CNVs are not an artefact of cell culture, but likely existed in the patient. Whilst we do not

see CNVs in any of the genes highlighted as genetic modifiers of age at onset, it is important to bear CNVs in mind when assessing phenotypes.

At later passages, the parent Q109N1 line and the Q22 cell lines show a duplication of chromosome 1. It is well established that prolonged culture of iPSCs, especially under the strong selective pressures of maintaining pluripotency, can lead to adaption such as increased growth rate, reduced apoptosis and the acquisition of chromosomal abnormalities including copy number variations or trisomy (CNVs) (Laurent *et al.* 2011). Of all the trisomies, trisomy 1 is the rarest (Banzai *et al.* 2004). In mice, trisomy 1 resulted in hypotelorism; a reduced distance between organs or body parts, a general reduction in size of the mice, and developmental retardation (v. Domarus 1983). In human pluripotent stem cells trisomies of chromosome 1, 12, 7 and 20 are common (Baker *et al.* 2016), indicating that genes within these chromosomes confer selective advantage in culture (Blum *et al.* 2009; Avery *et al.* 2013). Such genetic aberrations can alter behaviour of the cells and effect downstream phenotypes, so it is crucial to be aware of them (Enver *et al.* 2005; Yang *et al.* 2008; Werbowetski-Ogilvie *et al.* 2009).

It is often reported that HD-iPSCs display impaired neural rosette formation, a phenotype thought to be indicative of a developmental impairment (Ruzo *et al.* 2018). However, reports are inconsistent. A recent paper reports a HD-ESC allelic series in which all HD lines form neural rosettes but rosette organisation is impaired in HD and *HTT*$^{-/-}$ lines (Ruzo *et al.* 2018), whilst other studies report that HD-iPSCs are not capable of forming neural rosette structures at all (The HD iPSC Consortium 2012; Xu *et al.* 2017). In contrast, a study looking at HD-embryonic stem cells (ESCs) with 37 and 51 CAG repeats reports normal neural rosette structure formation (Niclis *et al.* 2009). Consistent with this, we find that isogenic clones of both WT and expanded repeat lengths differentiate well and demonstrate characteristic neural rosette structures. It has been speculated that these differences in neural rosette formation among the HD-iPSC lines is due to different CAG repeat lengths; those with shorter CAG repeat lengths appear to form rosette structures (Niclis *et al.* 2009) whereas those with impairments seem to harbour longer CAG repeat lengths; 72(Jeon *et al.* 2012) and 180(Xu *et al.* 2017), however our data are not consistent with this theory.

These isogenic lines are now being used routinely by others investigating a number of phenotypic outcomes including transcriptomics, autophagy and lysosomal storage. To examine whether previously reported phenotypes were rescued following genetic correction of the CAG repeat we first looked at mitochondrial respiration which has shown to have deficits in numerous HD mouse models (Oliveira 2010) and HD cell lines (Seong *et al.* 2005; An *et al.* 2012). Consistent with previous studies we observed lower basal and maximal respiration rates in HD lines compared to WT or corrected cell lines. This assay should thus be utilised for further phenotype analysis, examining whether variants of interest might have any effect on metabolic activity, either directly or indirectly through expansion of the CAG repeat.

Robust phenotypes like those mentioned above are fundamentally important in the screening of genetic and environmental disease modifiers and to aid in the discovery of new drug targets.

This novel isogenic model can be used to examine the effects of genetic modifiers on CAG expansion and associated phenotypes in cells with an expanded CAG repeat and wild-type length repeat.

3.5 Chapter Summary

- iPSCs derived from an individual with 109 *HTT* CAG repeats demonstrate reproducible repeat expansions in iPSCs and post-mitotic neurons.
- Using piggyBac technologies and CRISPR-Cas9 the expanded *HTT* CAG repeat was genetically corrected to a wild-type repeat length of 22 *HTT* CAGs.
- Q22 iPSCs retain pluripotency and differentiation potential.
- Genetic correction of HD-iPSCs 'rescues' metabolic dysfunction.

Chapter 4: Functional knock-out of FAN1 in an isogenic iPSC model of Huntington's Disease

4.1 Introduction

The length of the expanded CAG repeat tract is inversely correlated with the age at which an individual develops HD and it accounts for up to 68% of the variance in age at onset (Djoussé *et al.* 2003). However, genetic and environmental factors account for the remaining variance and are thought to modify disease onset and progression (Wexler 2004). Such genetic modifiers are expected to highlight genes and pathways capable of altering the course of HD. This highlights new, human-validated targets for therapeutic development, with the goal of developing rational treatments to delay or prevent onset of HD symptoms.

Recent genome-wide association studies (GeM-HD GWAS) have identified several genomic loci significantly associated with deviation in the age at onset of HD from that predicted by CAG repeat length alone. Many of these are linked to DNA repair processes. Four independent genome-wide significant signals have been found on chromosome 15 at a genomic locus containing *FANCD2/FANCI-Associated Nuclease 1* (*FAN1)*(GeM-HD Consortium 2015; GeM-HD Consortium 2019). FAN1 is an 5' endonuclease and 5'-3' exonuclease (Kratz *et al.* 2010; MacKay *et al.* 2010). It was identified as a member of the Fanconi anaemia pathway with a role in interstrand cross-link repair (Liu *et al.* 2010; MacKay *et al.* 2010). Independently of this pathway it is involved in replication fork metabolism and chromosomal stability (Chaudhury *et al.* 2014; Lachaud *et al.* 2016). Biallelic loss of *FAN1* causes a rare kidney disease, karyomegalic interstitial nephritis in humans and mice (Zhou *et al.* 2012; Airik *et al.* 2016), whilst heterozygous nonsense mutations in FAN1 have been associated with several familial colorectal cancers (Seguí *et al.* 2015), possibly through increased predisposition to somatic mutations in other genes. Unlike the other genes in the FA pathway, mutations in *FAN1* do not cause FA.

Additional genome-wide significant signals have been identified at the genomic loci of genes associated with DNA maintenance processes; *MLH1, PMS1, MSH3, PMS2 and LIG1* (GeM-HD Consortium 2019). Pathway analysis revealed the significant association of genes in the mismatch repair pathways as modifiers of age at onset. These findings likely extend beyond HD and into other triplet repeat diseases, as age at onset in other polyglutamine repeat disorders also appears to be modulated by genetic variation in DNA repair pathways, suggestive of a common pathogenic mechanism acting at the level of the DNA (Bettencourt *et al.* 2016).

One mechanism by which DNA repair proteins might act in HD is through somatic instability of the *HTT* CAG repeat to drive disease pathogenesis. Repetitive DNA sequences form unusual secondary

structures (Marquis Gacy *et al.* 1995; Mirkin 2007) and it is thought that these might be aberrantly processed by DNA repair machinery, possibly being detected by mismatch repair proteins and then cleaved by nucleases like FAN1 (McMurray 2010). HD mouse models have yielded data that is consistent with this; knock-out of MMR genes *Mlh1*, *Mlh3*, *Msh2* and *Msh3* alleviated disease phenotypes and reduced somatic instability (Dragileva *et al.* 2009; Pinto *et al.* 2013; Tomé *et al.* 2013). FAN1 interacts with MLH1 and PMS2, so it is plausible that FAN1 and MMR proteins act through a concerted mechanism to modulate age at onset.

The modifier SNPs at the *FAN1* locus are bi-directional. Two of the SNPs are associated with an earlier age at onset. These are coding SNPs for putative damaging variants in *FAN1* and likely result in a loss of function of FAN1. The most significant SNP, rs150393409 ($p= 1.8 \times 10^{-28}$) is in linkage disequilibrium ($r^2 =0.72$) with a SNP encoding an amino acid change (R507H) at the C terminal end of FAN1 DNA binding domain and is associated with 5.2 years earlier motor onset. A second coding genome wide significant SNP on chromosome 15, rs151322829 ($p= 1.4 \times 10^{-8}$) tags a SNP encoding a missense variant (R377W). Conversely two SNPs are associated with a delayed age at onset and these correspond with cis-eQTL SNPs for increased FAN1 expression in cortical tissue. These data indicate that increased FAN1 expression might be protective and in support of this, knock-down of FAN1 in an iPSC model of HD increases the rate of *HTT* CAG expansions seen in iPSCs and post-mitotic neurons (Goold *et al.* 2019). Additionally, Fan1 was found to be protective against somatic CGG expansions in the *Fmr1* gene in a mouse model of Fragile X (X.-N. N. Zhao and Usdin 2018).

With mounting evidence that FAN1 is involved in the modulation of trinucleotide repeat expansions, we sought to establish a model that could be used to assay the effects of FAN1 on CAG expansions and associated phenotypes.

4.2 Chapter Aims

This chapter aims to generate and validate the knock-out of FAN1 from isogenic iPSCs harbouring *HTT* CAG repeat lengths of 22 CAG or >109 CAG. It will examine whether FAN1 knock-out iPSCs retain their pluripotency and are able to differentiate into neuronal lineages. It will also assess the effect of FAN1 deficiency on growth rates and susceptibility to interstrand cross-linking agents.

4.3 Results

4.3.1 GeM-HD Risk Alleles in HD-iPSCs

The GeM-HD GWAS highlighted numerous SNPs in the genomic region of *FAN1*. The most significant SNPs from each of the signals with $p < 1 \times 10^{-6}$ (Table 4.1) are the best candidates for linkage to true genetic modifiers and we therefore genotyped the iPSCs at these positions, enabling cleaner genetic manipulation for repeat expansion assays. The Q109 iPSCs contained the relatively common SNP rs35811129 located in the intron of *MTMR10*, downstream of *FAN1*. This SNP is likely either a regulatory SNP involved in modulating FAN1 expression or linked to such a SNP.

Chr	SNP	Major Allele	Minor Allele	Minor AF /%	Effect size/ years/ minor allele	Q109 cell line alleles	Notes
15	rs150393409	C	A	1.4	-5.2	C/C	Linkage disequilibrium with R507H in FAN1
15	rs35811129	G	A	27.5	+1.3	G/A	3' FAN1 UTR Intron MTMR10
15	rs151322829	C	T	0.7	-3.8	C/C	Missense variant R377W in FAN1
15	rs34017474	T	C	38.2	+0.8	/	Intron variant FAN1/ MTMR10

Table 4.1 Genotypes of the genome wide significant SNPs identified in the GeM-GWAS on chromosome 15 in Q109 iPSCs.

4.3.2 CRISPR-Cas9 mediated knock-Out of FAN1

Two guide RNAs (gRNAs) targeting the second exon of *FAN1* were designed using DESKGEN Cloud (www.deskgen.com) based on their predicted high efficiency and low probability for off-target effects (Figure 4.1A). Exon 2 is the first coding exon of *FAN1* and therefore a premature stop codon is likely to result in nonsense mediated decay. gRNAs were synthesized and duplexed with a universal ATTO 550 labeled Alt-R Cas9 tracrRNA and then incubated with Alt-R S.p.Cas9 Nuclease 3NLS to form a ribonucleotide protein (RNP) complex. Q109N1 and 22-C2 (gene corrected) iPSCs were electroporated

with the fluorescently labelled RNP complex. The use of a fluorescently labelled tracrRNA allowed the sorting of cells using FACS to enrich the cells that contained the ATTO 550 fluorescent marker, therefore reducing the screening pool of colonies to be picked. After 24 hours transfected cells and non-transfected control cells were dissociated and sorted on a BD FACS Aria Fusion flow cytometer. FACS analysis showed a clear population shift in fluorescence after transfection with the fluorescently labelled RNP. Through FACS, the top 10% of cells emitting the highest levels of ATTO 550 fluorescence were sorted into warm media (Figure 4.1B). These cells were then plated into a 10 cm dish and grown for 4-6 days before picking clones into a 96 well plate for screening and subsequent expansion.

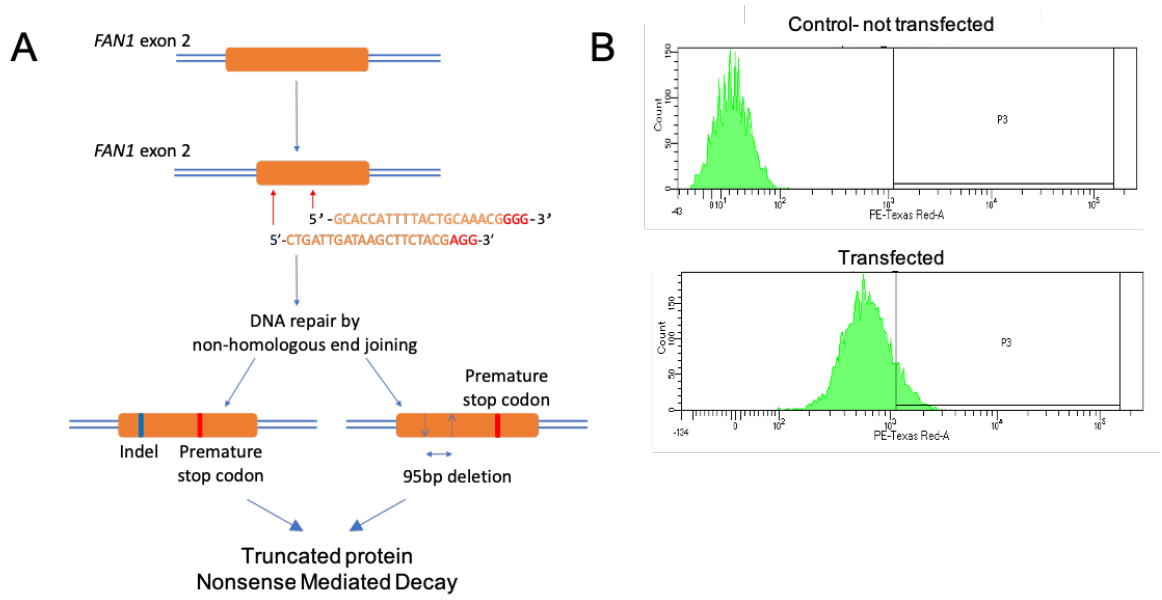
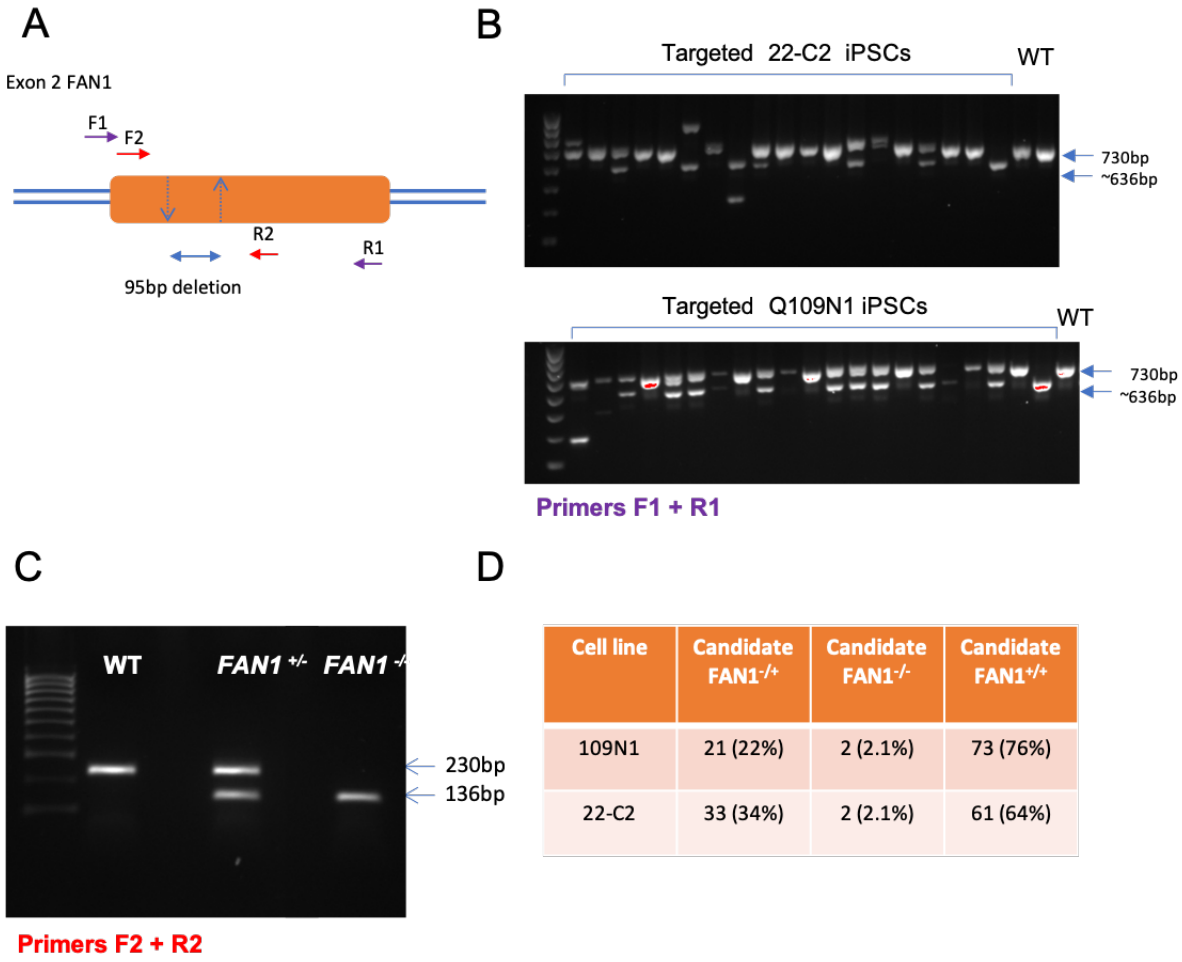


Figure 4.1 Generation of FAN1$^{-/-}$ iPSC lines. (A) Schematic showing CRISPR-Cas9 targeting of exon 2 of *FAN1* using two gRNAs. gRNA sequences are shown in 5'-3' orientation (orange) followed by the PAM sequence (red). (B) FACS analysis of transfected iPSCs revealed a population shift in fluorescence after FACS (bottom) compared to control (top). The top 10% of fluorescently labelled cells (P3) were sorted and replated for subsequent picking and expansion.

4.3.3 Screening of targeted clones

Genomic DNA was extracted from resultant colonies and a PCR screen amplifying exon 2 of *FAN1* was performed (Figure 4.2A). PCR products were gel electrophoresed (Figure 4.2B). Candidate clones were determined to be those which produced a PCR product at a lower molecular weight (136 bp) than the predicted wild-type PCR product (230 bp). Two primer pairs were used to ensure the banding pattern was consistent across PCRs (Figure 4.2B, Figure 4.2C) and to ensure any larger insertions or deletions would be detected.

96 colonies from each transfection in Q109N1 and 22-C2 cells were picked. In the Q109N1 line, 21 of the 96 picked colonies (22%) had a fragment deletion evident by two bands visible after gel electrophoresis, indicative of potential 109-*FAN1*$^{+/-}$ clones and 2 clones were candidate 109-*FAN1*$^{-/-}$ lines (2%), evident by the absence of a band at the WT molecular weight (230 bp) and the presence of only 1 band at a lower molecular weight (136 bp) (Figure 4.2B, Figure 4.2D). The targeting efficiencies were comparable for the 22-C1 iPSC transfection, where 33 clones presented with two bands (34%) and 2 clones had one band (2%) at the lower molecular weight (Figure 4.2D). It should be noted that many of the clones with a fragment deletion on one allele might have insertions or deletions on the other allele, rendering the clone *FAN1*$^{-/-}$, and so the targeting efficiencies calculated here are likely an underestimation.



Cell line	Candidate FAN1$^{-/+}$	Candidate FAN1$^{-/-}$	Candidate FAN1$^{+/+}$
109N1	21 (22%)	2 (2.1%)	73 (76%)
22-C2	33 (34%)	2 (2.1%)	61 (64%)

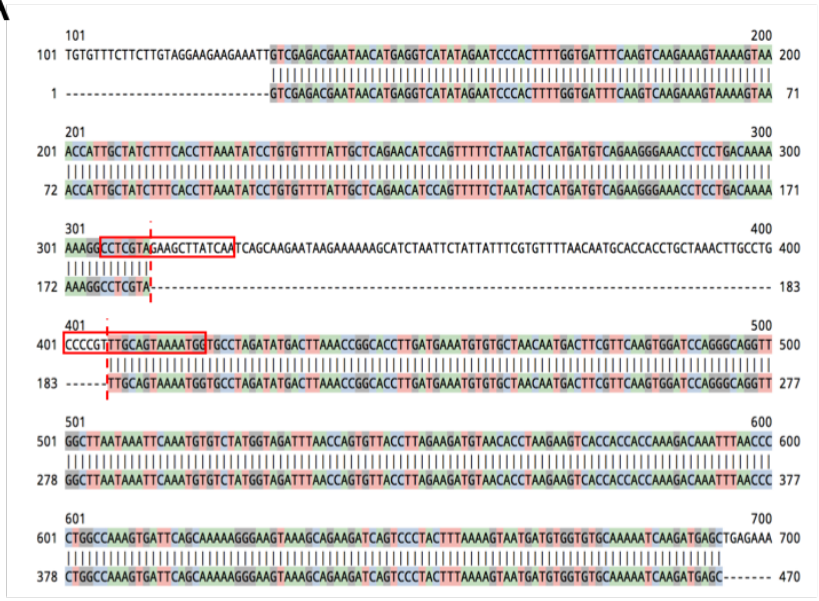
Figure 4.2 Screening of *FAN1* targeted iPSCs (A) Schematic of PCR screen over exon 2. Primer pairs used are indicated. F1 + R1 (purple) generates a 730 bp PCR product in WT clones, F2 + R2 (red) generates a 230bp PCR product in WT clones. (B) Diagnostic PCR screen over exon 2 of *FAN1* in 22-C2(top) and Q109N1 (bottom) transfected clones shows multiple banding patterns as a result of the transfection, highlighting variability in CRISPR-Cas9 targeting. (C) Diagnostic PCR screen using a second set of primers, F2 + R2, reveals similar banding pattern and highlights WT, *FAN1*$^{-/-}$ and *FAN1*$^{+/+}$ candidate clones. Representative banding patterns of WT alleles (230bp), *FAN1*$^{-/+}$ (230/136 bp) and *FAN1*$^{-/-}$ (136/136 bp). (D) CRISPR-Cas9 targeting efficiencies for each Q109N1 and 22-C2 line, calculated based on the presence of a fragment deletion.

Due to high targeting efficiency, not all of the targeted clones were Sanger sequenced, however all candidate clones were frozen down. Of those that were Sanger sequenced, the most common edit was a 94bp deletion, created by the ligation of the two fragmented ends. This edit was predicted based on the gRNA sequence and predicted Cas9 cut site (Figure 4.3A). The variability of the CRISPR-Cas9 targeting was evident by both gel electrophoresis (Figure 4.2B) and Sanger sequencing. Other deletions included the loss of 92bp; 95bp; 104bp or indels around the cut sites (Figure 4.3B).

We then assessed whether the 94bp fragment deletion altered *FAN1* expression. There was a significant decrease in *FAN1* expression, relative to the parent line, measured by RT-qPCR (Figure 4.3C). FAN1 is usually encoded by 1017 amino acids with an expected molecular weight of 114 kDa, however, the loss of 94bp results in a stop codon after 23 amino acids, truncating the protein. The loss of FAN1 is confirmed by Western Blot (Figure 4.3D).

Two FAN1$^{-/-}$ lines derived from 22-C2 (22-FAN1$^{-/-}$-C1 and 22-FAN1$^{-/-}$-C2) and two lines derived from Q109N1 (109-FAN1$^{-/-}$-C1 and 109-FAN1$^{-/-}$-C2) were taken forward for further characterisation (Figure 4.4). Due to time constraints FAN1$^{+/-}$ clones were not used in subsequent experiments.

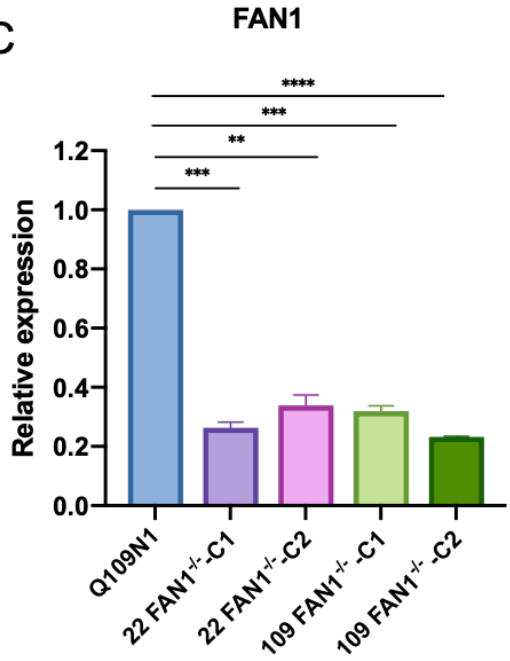
A



B

Clone	*FAN1* locus genotype Allele 1	Allele 2
#1	1bp insertion/ 1 bp deletion	WT
#2, #6, #7, #8	94bp deletion	94 bp deletion
#3	95bp deletion	WT
#4	102bp deletion (in frame)	WT
#5	Silent mutation/ 92bp deletion	Multiple sequences
#9	94bp deletion	WT

C



D

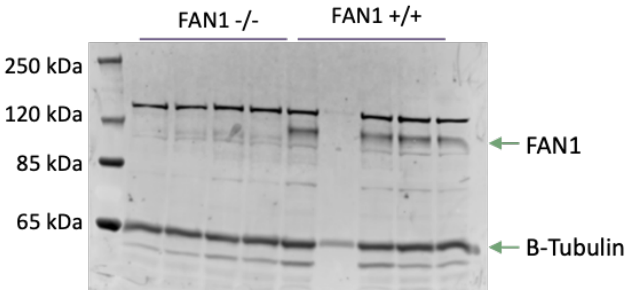


Figure 4.3 Validation of CRISPR-Cas9 mediated FAN1 knock-out in iPSCs. (A) Sanger sequencing of PCR products demonstrates the targeted deletion as expected. Red boxes indicate gRNA sequences and dotted line indicates predicted Cas9 cut-site. (B) Sanger sequencing of multiple candidate clones reveals a number of different *FAN1* locus genotypes. Genotypes of each allele are shown, representing both *FAN1*$^{-/-}$ and *FAN1*$^{+/-}$ clones. (C) Relative FAN1 expression levels were determined by RT-qPCR. Data was normalised to Q109N1. (D) Western blot with polyclonal antibody raised in sheep against FAN1 confirms FAN1 knock-out from targeted iPSCs.

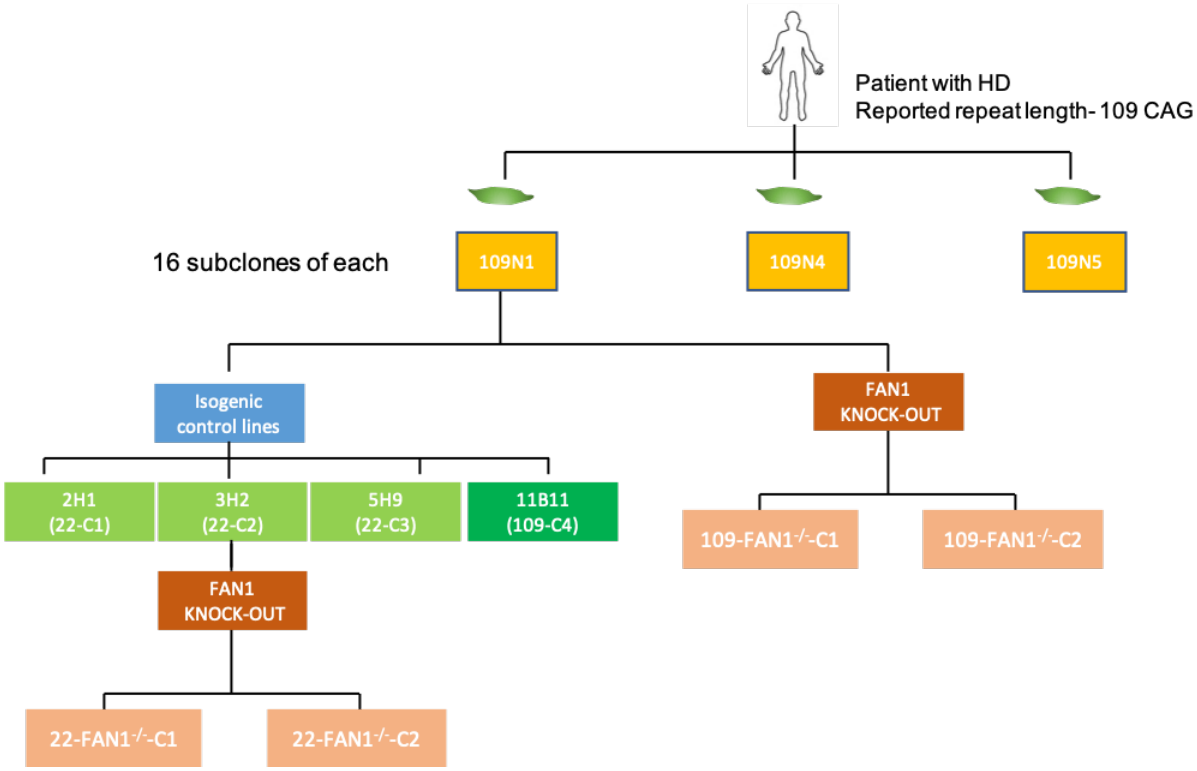


Figure 4.4 HD-iPSC family tree. Diagram depicting iPSCs derived from a patient diagnosed at 3 years old with Juvenile HD. Diagram shows the iPSCs that we have in house and those that have been gene-edited during this project.

4.3.4 Virtual karyotyping of FAN1$^{-/-}$ and corrected HD-iPSCs

All resultant clones were genotyped using either the PsychArray-24 chip which has a coverage of 593,260 markers or the InfiniumTM Global Screening Array-24 v2.0 which has a coverage of 665,608 markers. As described in Chapter 3, these arrays allow high-throughput 'virtual karyotyping' and therefore are ideal for looking at chromosomal duplications or deletions in a number of CRISPR clones at multiple time-points.

As in Chapter 3, all samples, including the parent Q109N1 line, contained a similar deletion at 2q22.1 and at 14q24.3 (Table 4.2). Additionally, all samples demonstrated loss of heterozygosity at 3q26.1 and 12q14.2 (Figure 4.5A and Figure 4.5B). The locus of 3q26.1 contains a pseudogene and 12q.14.2 contains *SRGAP1* which encodes a GTPase activator. Loss of heterozygosity may occur through mitotic recombination, where heterozygosity is lost upon segregation of the chromosomes. The cell still contains two copies of each gene; however, they are both from either the maternal or paternal allele. Although loss of heterozygosity was not detected in chapter 3, this was likely due to different coverage of markers between the two chips, as all samples re-run on the InfiniumTM Global Screening Array-24 demonstrate a loss of heterozygosity in these regions.

An additional deletion was called at 22q11.1. 22q11.1 is centromeric, and genotyping over this region provides unreliable results, so this CNV is likely a false-positive. To ensure CNVs are accurately called, plots must be analysed by eye.

As discussed in Chapter 3, a similar phenomenon was observed with respect to chromosome 1 in the FAN1 targeted clones and parent line. Immediately after transfection targeted clones were diploid. At some point between passage 13 and 18 however, three out of the four targeted clones underwent a duplication of chromosome 1 (Table 4.2, Figure 4.5B). One line, 109 FAN1$^{-/-}$-C2, did not undergo a duplication of chromosome 1, and this appeared to be stable over >40 passages (Figure 4.5A). There was no apparent effect of the loss of FAN1 *or HTT CAG length* on karyotype.

Total CNVs	CNV Locus	1p36.33-q44	2q22.1	3q26.1	12q14.2	14q24.3
	Approx Position	chr1: full chromosome	chr2:141448312-141932472	chr3:166000000-166400000	chr12:64011062-64116159	chr14:77984540-78108110
	Status	Duplication	Deletion	Loss of Heterozygosity	Loss of Heterozygosity	Deletion
	Length	248,511,674	484,161	400,000	105,098	123,571
	Gene		*LRP1B*		*SRGAP1*	*NRXN3*
22-FAN1$^{-/-}$C1	Passage 9					
	Passage 12					
	Psssage 23					
22-FAN1$^{-/-}$C2	Passage 5					
	Passage 11					
	Passage 20					
109-FAN1$^{-/-}$C1	Passage 4					
	Passage 13					
	Passage 29					
	Passage 40					
109-FAN1$^{-/-}$C2	Passage 9					
	Passage 13					
	Passage 19					
	Passage 32					
109N1	Passage 28					
	Passage 39					

Table 4.2 CNV analysis performed after genotyping targeted clones on PsychArray and Global Screening Array Final CNV analysis reveals deletions at 2q22.1 and 14q24.3 in all samples, and duplication of chromosome 1 in various samples at later passages. Loss of heterozygosity was observed at 3q26.1 and 12q14.1. Yellow highlighted squares indicate that the sample has the CNV highlighted. Green highlighted squares indicate loss of heterozygosity.

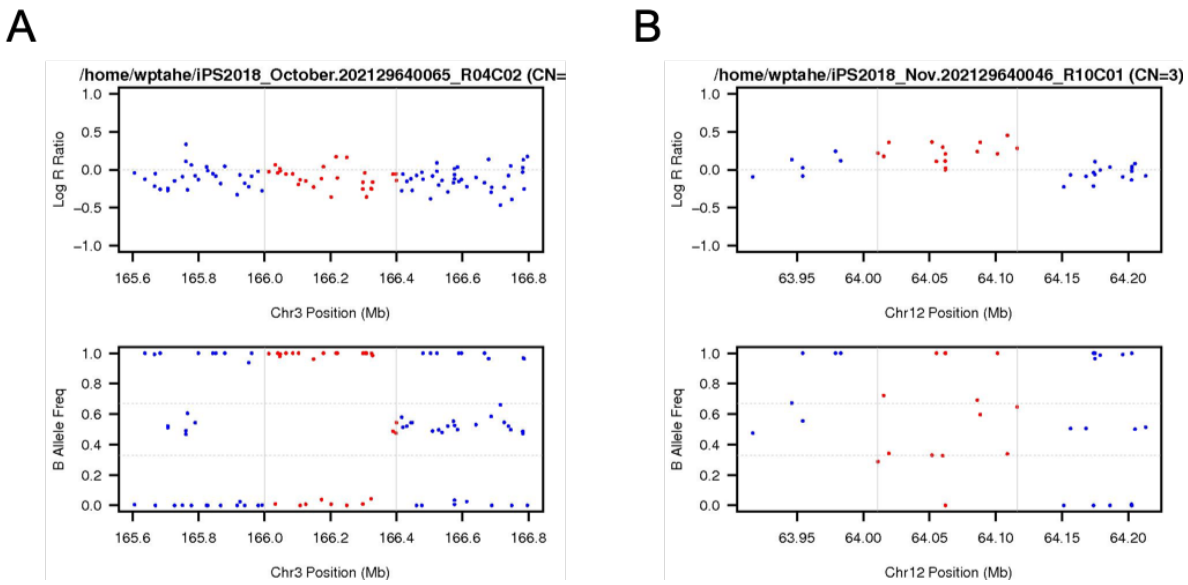


Figure 4.5 SNP array plots of FAN1$^{-/-}$ clones. (A) Loss of heterozygosity on chromosome 3, characterised by a Log R Ratio of 0, but an absence of SNPs around 0.5 in the B allele Frequency Plot (B) Loss of heterozygosity on chromosome 12.

4.3.6 FAN1$^{-/-}$ iPSCs show increased vulnerability to mitomycin C

To investigate whether the knock-out of FAN1 results in the cells being hypersensitive to treatment with DNA crosslinking agents, we looked at cell viability of HD-iPSCs treated mitomycin C (MMC). MMC is a cross-linking agent which covalently binds nucleotides, generating inter- and intrastrand cross-links (Roh et al. 2008). Cells from patients with Fanconi Anemia show increased vulnerability to MMC (Claassen et al. 1986). Exposure of cells to MMC can therefore be considered an indirect assay of FAN1 function. FAN1$^{-/-}$ iPSCs were significantly more sensitive to MMC at 10 and 20 ng/mL when compared with FAN1$^{+/+}$ iPSCs (Figure 4.6). These data are consistent with a functional knockout of FAN1 in the engineered FAN1$^{-/-}$ HD-iPSC lines, as expected.

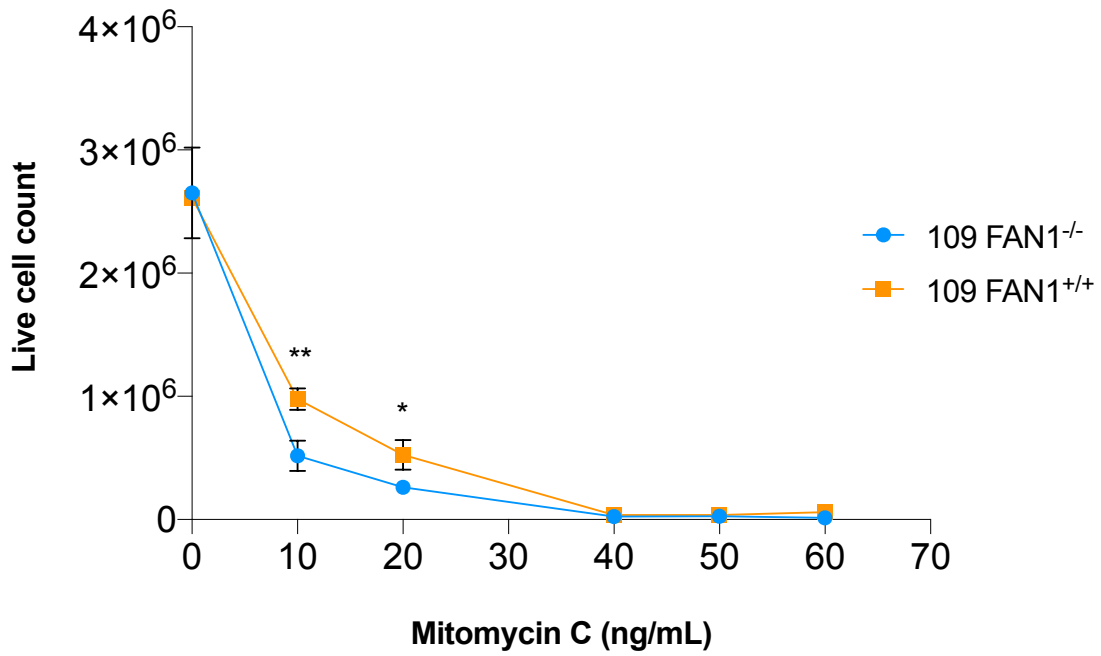
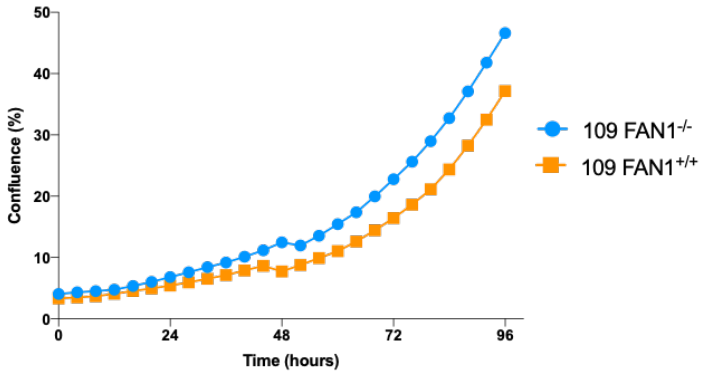


Figure 4.6 Survival of HD-iPSCs after treatment with indicated doses of mitomycin C (MMC). iPSCs were treated in duplicate with increasing concentrations of MMC. After 4 days, live cell numbers were determined and plotted. 109-FAN1$^{-/-}$ iPSCs (n=2) were significantly more vulnerable to MMC treatment at 10 ng/ml and 20 ng/ml than 109-FAN1$^{+/+}$ iPSCs (n=2). Asterisks indicate significant differences; *p<0.05, **p<0.01 determined by unpaired t-test. Values expressed as mean ± SEM.

4.3.7 Proliferation rates in Q109 and FAN1$^{-/-}$ iPSCs

FAN1 is involved in repairing interstrand cross-links, if these go unresolved then replication may become stalled or delayed. We therefore examined whether there were differences in proliferation between iPSCs with and without FAN1. As in Chapter 3 we used the Incucyte to analyse confluence (%) by applying a confluence mask to an occupied area. We were then able to use confluence as a surrogate for proliferation. Confluence rates (%) grew at a significantly higher rate in FAN1$^{-/-}$ iPSCs compared with FAN1$^{+/+}$ lines (Figure 4.7A). Doubling rates were determined to be 29.1 hours and 27.9 hours for FAN1$^{+/+}$ and FAN1$^{-/-}$ iPSCs, respectively (Figure 4.7B). Cell media was changed after 48 hours, giving rise to the dip in the curves. As in chapter 3, to ensure the confluence mask represented cell number rather than cell size, cell size was determined using the Countess II FL Automated Cell Counter. Average cell sizes for 109-FAN1$^{+/+}$ and 109-FAN1$^{-/-}$ were 18.98 µm and 19.5 µm, respectively. Whilst these cell sizes are similar, this difference in cell size is likely amplified, driving a difference in confluence mask and therefore skewing the doubling rates ascertained. More accurate measures of proliferation are therefore required.

A



B

Genotype	Doubling Rate (hours)
109 FAN1$^{+/+}$	29.1 hrs
109 FAN1$^{-/-}$	27.9 hrs

Figure 4.7 Surrogate proliferation rates in FAN1$^{-/-}$ and FAN1$^{+/+}$ iPSCs. Cell confluency (%) was calculated using Incucyte Zoom Software by phase-contrast images. Cells were plated at 50,000 cells/well. $p<0.0001$ determined by Two Way ANOVA followed by Sidak's multiple comparisons test. Values expressed as Means ± SEM with 3 technical replicates. n= 1 independent experiment.

4.3.8 Characterisation of pluripotency in *FAN1*$^{-/-}$ iPSCs

We then examined whether the pluripotent characteristics of the parental HD-iPSC line, previously shown to express pluripotency markers, were maintained in the gene-edited iPSC lines. All targeted clones retained a normal iPSC morphology (Figure 4.8A). FAN1 targeted clones had variable mRNA levels of the pluripotency genes *OCT4*, *NANOG*, *SOX2* (Figure 4.8B) compared to Q109N1. All but one of the clones had an upregulation of *OCT4* and all of the clones had an upregulation of *NANOG*. This might be due to considerable heterogeneity in the parent Q109 line where some iPSCs may be skewed towards spontaneous differentiation. *SOX2* was also variable amongst subclones. Immunostaining of the targeted clones revealed OCT4 staining in the nuclei of all cells (Figure 4.8C). These data therefore confirm pluripotency in all FAN1$^{-/-}$ clones, but care should be taken with sampling for RT-qPCR experiments.

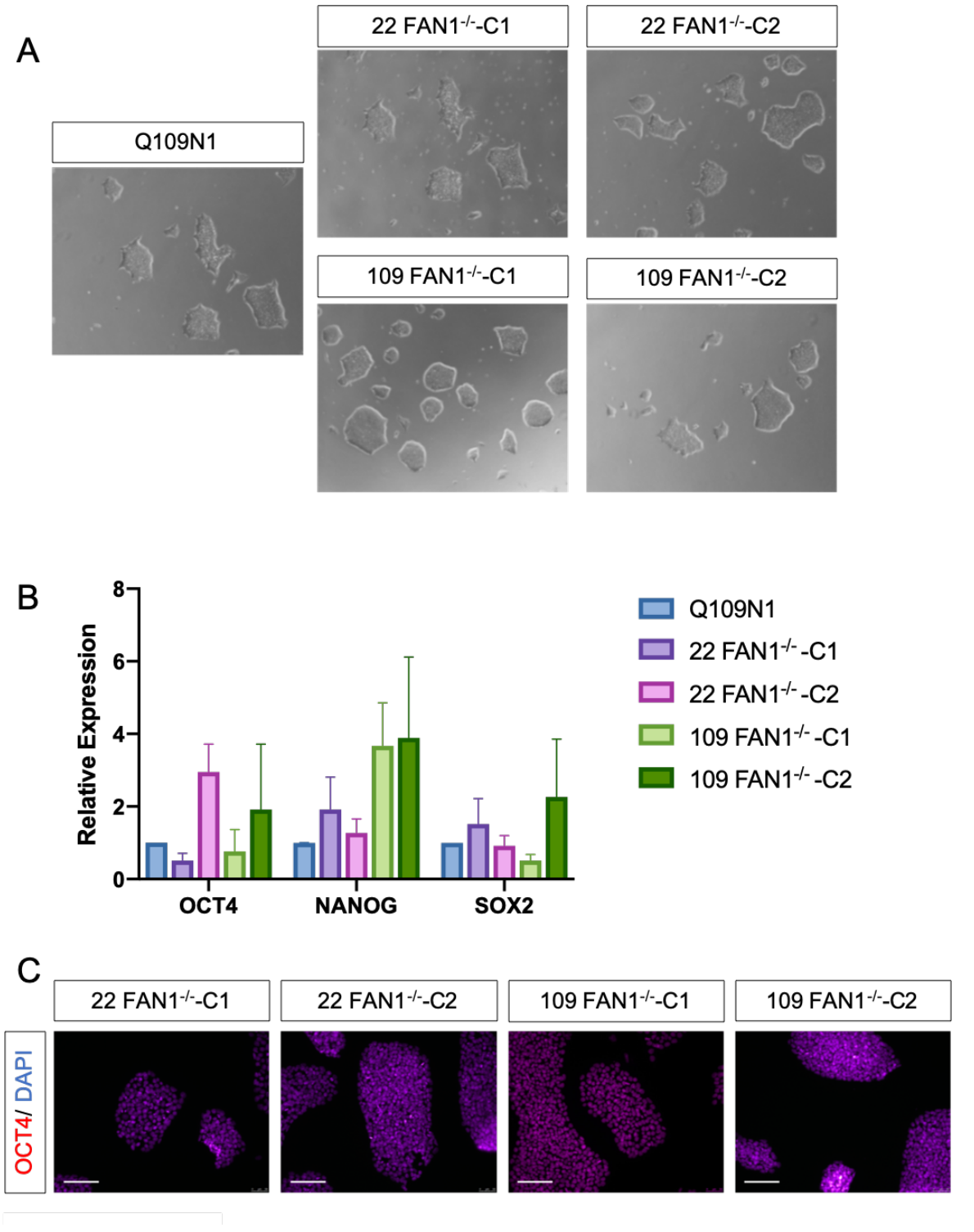


Figure 4.8 Assessment of pluripotency in FAN1 targeted iPSCs. (A) Brightfield images of FAN1$^{-/-}$ iPSCs. All targeted iPSCs retain normal iPSC morphology. (B) The expression levels of pluripotency genes *OCT4, NANOG* and *SOX2* are comparative across the parent line Q109N1 and targeted lines. Data are normalised to *GAPDH* and *β-actin*. Expression values are relative to the parent line Q109N1-iPSC RNA. (n= 3 per clone, values for independent technical replicates shown as mean ± SEM (C) Undifferentiated iPSCs stained positive for the pluripotency marker OCT4 (red). All nuclei are counterstained with DAPI (blue). Scale bars = 100 µm.

4.3.9 Directed neural differentiation of *FAN1*$^{-/-}$ iPSCs

FAN1$^{+/+}$ and FAN1$^{-/-}$ iPSCs were differentiated to NPCs following the protocol described in section 2.1.3. Each clonal line grew well as a monolayer, and neural rosettes were visible from day 10. Following 16 days of neural induction forebrain progenitor markers PAX6, FOXG1 and CTIP2 (Arlotta et al. 2008; Greig et al. 2013) were upregulated (Figure 4.9A). Each NPC line stained positive for nestin (Figure 4.9B).

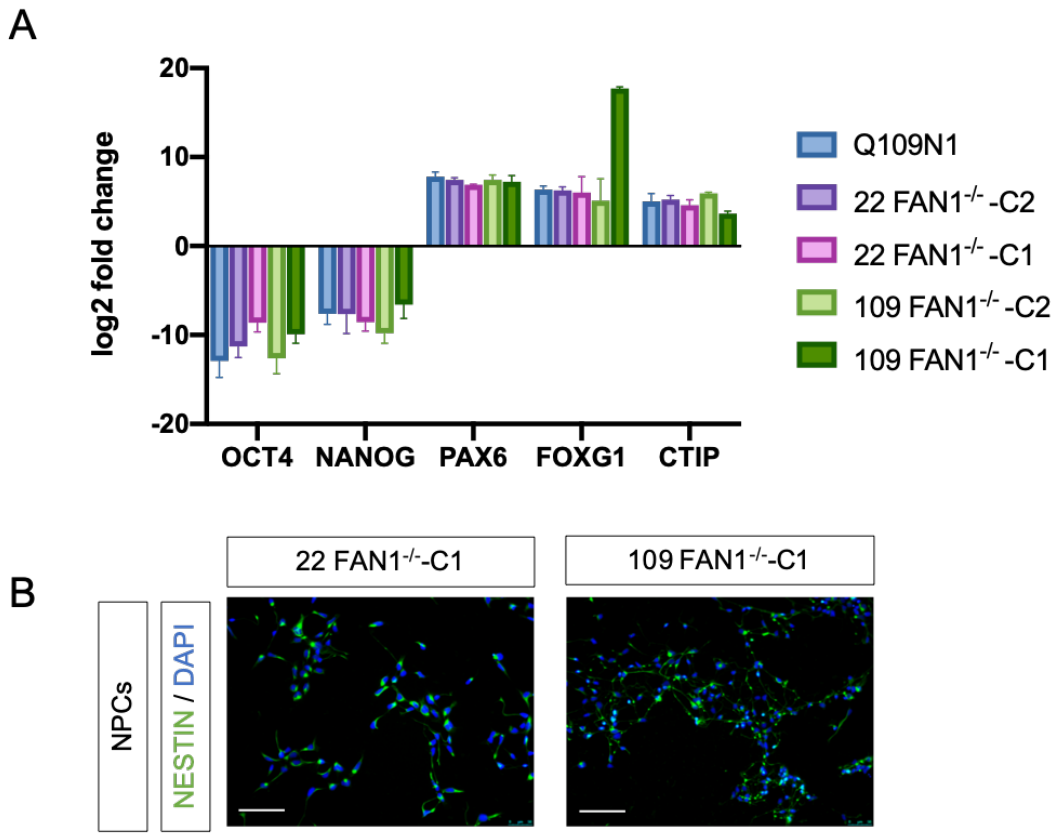


Figure 4.9 Differentiation of targeted HD-iPSCs into NPCs and neurons. (A) Neural induction results in the decreased expression of pluripotency markers *OCT4* and *NANOG* and increased expression of n neural precursor markers *FOXG1*, *CTIP* and *PAX6* at day 16. Data are normalised to *GAPDH* and *β-actin*. Expression values are relative to the RNA from the respective line at iPSC stage. (n= 3 per clone, values for independent technical replicates shown as mean ± SEM) (B) NPCs stain positive for NPC marker nestin at day 16 of neural induction. Scale bars = 100 µm.

4.4 Discussion

This chapter describes the use of CRISPR-Cas9 to knock FAN1 out of an isogenic system. The knock-out of FAN1 from iPSC lines harbouring a CAG repeat of 22 and 109 will allow us to examine whether there are differential phenotypic effects of FAN1 in a disease or 'healthy' cellular model. Until now much research has focused on the phenotype in the context of disease, however with advancements in gene-editing technologies we are now able to look at the phenotype in an isogenic system. Our research group is interested in the role of FAN1 in the context of HD, but to control for genetic background, metabolism and cell type, the isogenic lines are critically important.

Within the past few years CRISPR-Cas9 editing efficiencies have greatly improved and this chapter reports editing rates in concordance with that, with editing efficiencies of 30-35% (Liang *et al.* 2017). There are a number of factors that can improve gene-editing efficiency; guide RNAs chosen, guide RNA number, target sequence, concentration of components and FAC sorting after transfection. Each of these were considered when designing the experiment. To generate a FAN1 knock-out we employed the use of two guide RNAs with predicted Cas9 cut sites 95bp apart with the aim of introducing a large fragment deletion in *FAN1.* An out-of-frame deletion was predicted by the gRNA cut sites, with the subsequent expectation that a premature stop codon should be introduced. This targeting method allowed for a single PCR across exon 2 of FAN1 to detect the fragment deletion, obviating the use of a restriction digest or T7 surveyor assay - both of which are time consuming. It should also be noted that we only calculated the efficiencies of the large fragment deletions as this was our primary screen. There is no doubt that many of the other clones that were picked would be $FAN1^{-/-}$ or $FAN1^{-/+}$, due to indels arising through non-homologous end-joining repair. As we already had a number of homozygous and heterozygous knock-outs, however, no additional clones were sequenced. These do however represent a stock of multiple candidate clones that could be used in additional experiments.

In chapter 3 we discussed the use of CRISPR-Cas9 to genetically correct the expanded *HTT* CAG repeat length to a wild-type repeat length. Although we would expect editing efficiencies to be lower due to the use of a homologous recombination repair template, they were notably lower than previous reports. In the generation of the $FAN1^{-/-}$ iPSCs we utilised a fluorescently labelled universal tracrRNA. This was labelled with ATTO-550, which allowed for the detection of any cells that had taken up the RNP complex upon transfection. We chose to sort the top 10% of cells emitting the highest fluorescence, but FACS parameters could be made more stringent to reduce the screening pool even further. Whilst it does not necessarily mean that the cells have the edit, or that a double strand break has even been induced at the desired locus, it reduced the initial cell pool from ~200,000 cells to ~8,000. Almost a third of screened clones had an edit detectable by distinct bands in gel electrophoresis. Cell

survival after FACS is often reported to be low, however, we found no indication of that. We consistently had survival of an ample number of colonies for screening. We would therefore suggest that using the fluorescently labelled tracrRNA and FACS sorting 24 hours later should be done in all gene-editing applications, as it substantially reduces the time spent picking and screening clones and is more cost-effective.

An interesting observation was the sequence variability in the edited clones. This was evidenced by the multiple banding patterns observed when the *FAN1* exon 2 PCR products were gel electrophoresed. Whilst you would expect slightly different combinations of banding patterns, there were bands ~200bp larger and smaller than expected. Due to time restraints these clones were not investigated further, though it would be interesting to Sanger sequence the clones to examine how the DNA was repaired in each cell following CRISPR-Cas9. Sanger sequencing the spurious bands in an additional experiment revealed a 93 bp insert of DNA between the two primer pairs which endogenously resides ~ 1 kb away from the CRISPR-Cas9 cut sites. This likely indicates homologous recombination with DNA that was topographically near.

Successful targeting of FAN1 was confirmed by a reduction in *FAN1* mRNA in RT-qPCR and the absence of FAN1 in a western blot. Whilst FAN1 expression was reduced, we still see some FAN1 expression, possibly representing *FAN1* transcripts that escaped nonsense-mediated decay. Although we report a reduction in *FAN1* expression in the FAN1$^{-/-}$ cell lines, the CT values required to reach the threshold in the parent line were still relatively high, suggesting that FAN1 might be expressed endogenously at low levels. This in itself might be due to the presence of SNP rs35811129 in the 3' UTR of *FAN1* which was identified as a genetic modifier of age at onset (GeM-HD Consortium 2015; GeM-HD Consortium 2019). The presence of this SNP is likely associated with differential expression of FAN1 and an altered age at onset than predicted by *HTT* CAG repeat length. Whether the SNP is the functional SNP or tagging the actual variant that might alter FAN1 expression is not known. However, endogenously low levels of FAN1 might partly explain the *HTT* CAG repeat instability that this cell line exhibits. To establish relative FAN1 expression levels in the Q109 line, RT-qPCRs should be performed with RNA from several iPSC lines from different subjects, to determine whether FAN1 is expressed at similar levels.

At the protein level, western blots with FAN1 antibodies reveal that FAN1 at the expected molecular weight was not present in the knock-out lines. Bands were also present at other molecular weights that most likely represent non-specific products as they are seen in both FAN1 positive and FAN1 knock-out cells. The FAN1 antibody was raised in sheep against full-length GST-FAN1 (MacKay *et al.* 2010). Non-specific binding is a common issue in western blotting and in our hands, commercial antibodies gave little to no detection of the FAN1 protein, indicating a need for better antibodies. However, our data are consistent with the loss of FAN1 and this is supported by the functional results from MMC assays.

As in Chapter 3, duplication of chromosome 1 is seen at later passages in the parent line and three out of the four FAN1$^{-/-}$ lines. Unfortunately, this does reduce the utility of the lines beyond a certain passage as we cannot be sure of the effect that the duplication of chromosome 1 might have on oncogenes, tumour suppressor genes or possible genetic modifiers of HD. As was seen in Chapter 3, the Q109N1 line and daughter clones derived from it show a propensity for a duplication of chromosome 1. Interestingly, one FAN1$^{-/-}$ clone did not undergo a duplication of chromosome 1 over 40 passages, perhaps indicating that it is unlikely to. Each clone likely represents a mixed population where a proportion of cells contain the duplication, and a proportion do not. It therefore raises the possibility of subcloning each line and selecting those without the duplication. Whether this would then introduce a selection bias is unknown. Reprogramming and cell culture are often noted as the biggest drivers of karyotypic abnormalities (Yoshihara *et al.* 2017), however, with all subclones derived from the same parent line and cultured under identical conditions these explanations do not provide a complete answer. It is likely that genetic stability changes in number of subclones due to adaptation of culture conditions (Okita and Yamanaka 2011). For subsequent experiments cells at the lowest possible passage were used to ensure phenotypes observed are from cells that are diploid.

Interestingly over ~30-40 passages, none of the cell lines accrued any additional CNVs that were not present in the parent line. It is well known that late passage PSCs are more likely to have genomic changes than cells at an earlier passage. We would expect there to be genomic variation present in the parental Q109N1 line, and upon clonal isolation and selection we might assume that we would amplify this genetic variation. It is therefore fairly striking that no variation was observed at the CNV level between the lines, particularly 22-FAN1$^{-/-}$ as these are two generations away from the parent Q109N1 line and have been through two rounds of gene-editing. Gene-editing introduces strong selective pressures that favour the most adaptable cells, meaning that cells that have variation in genes that confer a growth or survival advantage are likely to be selected for in culture (Laurent *et al.* 2011). Exome sequencing of the samples might reveal SNVs that are not detectable on a SNP array and it is likely that, as in chapter 3, variation might exist in genes involved in growth and survival.

As reported in chapter 3, all samples contained a deletion at 2q23.1, which contains the gene *LRP1B*, a putative tumour suppressor (Langbein *et al.* 2002; Prazeres *et al.* 2011; Ni *et al.* 2013). A deletion was also identified at 14q24.3, which contains the gene *NRXN3,* a presynaptic adhesion molecule involved in regulating neurotransmitter release (Hishimoto *et al.* 2019). Whilst we do not see CNVs in any of the genes highlighted as a genetic modifier of age at onset, the identification of such CNVs is important as they may influence phenotypic outcomes. However, the presence of these CNVs in each of the lines suggests that changes in phenotype observed can be attributed to the loss of FAN1 rather than by the presence of different CNVs in different clones.

The SNP array identified a loss of heterozygosity at two loci. 3q26.1 contains a pseudogene and 12q.14.2 contains *SRGAP1* which encodes a GTPase activator. Loss of heterozygosity (LOH) is commonly reported in many cancer types when the loss of a wild-type allele unmasks a recessive mutation in a tumour suppressor gene (Ryland *et al.* 2015). LOH can occur as a result of a DSB repaired through homologous recombination, resolved through crossing over where the maternal and paternal chromosome arms are exchanged. This results in everything distal to the initial DSB being homozygous with maternal or paternal origin (Zheng *et al.* 2005). A study looking at culture-induced effects on LOH in ESCs revealed a correlation between passage number and LOH sites detected with the detection of 1.3 new LOH sites per passage (Zheng *et al.* 2005). Whilst we do not see LOH rates as high as this, it is important to consider this in the clinical implications of iPSCs. Several mechanisms likely contribute to genomic stability in the iPSCs. HD-ESCs have reports of deficits in DNA repair (Ooi *et al.* 2019), and manipulating DNA repair genes in the iPSCs is likely to influence genomic stability, so care should be taken. However, in this instance, loss of heterozygosity was also present in the parental line, suggesting that it has not resulted from prolonged cell culture.

FAN1 is required for the processing of replication forks and protection against chromosome instability. Cells with defects in a number of the Fanconi anaemia genes show significantly more spontaneous chromosomal aberrations compared to wild-type cells and these are exacerbated following treatment with interstrand cross-linking agents (Schlacher *et al.* 2012). Following treatment with MMC, Fan1$^{-/-}$ mouse embryonic fibroblasts show increased levels of chromosome breaks and radial chromosomes (Thongthip *et al.* 2016). Therefore, we might expect to see more chromosomal instability in our FAN1$^{-/-}$ lines, however, there appears to be no effect of FAN1 knock-out and CNVs in the cells. Similarly, in a Chicken B cell (DT40) line, FAN1 knock-out had little effect on spontaneous aberrations, but upon treatment with MMC there were elevated levels of chromosomal aberrations compared with FAN1$^{+/+}$ cells (Yoshikiyo *et al.* 2010).

These data show little effect of FAN1 deficiency on proliferation rates in the iPSCs. This is consistent with data showing FAN1 knock-out of DT40 cells had no effect on doubling times or cell-cycle profiles (Yoshikiyo *et al.* 2010). This may be due to functional redundancy as Fanconi anaemia complementation group J knock-out cells (FANCJ$^{-/-}$) cells have decreased proliferation rates, and these are decreased further with the additional knock-out of FAN1. Conversely, the reduction in proliferation rates in Fanconi anaemia complementation group C knock-out cells (FANCC$^{-/-}$) cells was not decreased any further in FAN1$^{-/-}$/FANCC$^{-/-}$ cells.

109-FAN1$^{-/-}$ cells showed decreased survival after treatment with the interstrand cross-linking agent MMC compared to 109-FAN1$^{+/+}$. This is consistent with numerous studies where cells with FAN1 deficiency show increased sensitivity to a number of cross linking agents such as MMC, cisplatin,

acetaldehyde and formaldehyde (Kratz *et al.* 2010; Yoshikiyo *et al.* 2010; Zhou *et al.* 2012; Thongthip *et al.* 2016). Interestingly, FAN1$^{+/-}$ MEFs did not show a cell death phenotype following MMC or cisplatin treatment, suggesting FAN1 is not haplo-insufficient for interstrand cross-link repair (Thongthip *et al.* 2016) and this is consistent with the recessive nature of KIN in humans. Whether there would be a differential effect of vulnerability to mitomycin C in the 22 FAN1$^{-/-}$ iPSCs compared to the 109 FAN1$^{-/-}$ iPSCs should be investigated in future experiments.

Our data show variability of the expression of pluripotency genes determined by RT-qPCR. Expression levels were variable but slightly higher than the parent Q109N1 line. This might be due to a number of factors. Samples for RNA were often taken when iPSCs were ready to be passaged. In hindsight, taking samples when cells were still growing at an exponential rate would be preferable, so that expression levels of any genes involved in growth and division would not be altered. Additionally, spontaneous differentiation was present in differing levels in a number of subclones. After cloning, spontaneous differentiation is often present due to the stress introduced in the cloning procedure. Spontaneous differentiation in culture would render the expression levels of pluripotency genes relatively lower when compared to the expression levels of the housekeeping genes, due to the presence of a number of cells which are likely no longer pluripotent. In future care should be taken to take RNA when cells do not have spontaneous differentiations in the culture.

Ideally more data would have been obtained for the differentiation potential of the FAN1$^{-/-}$ iPSCs, both immunocytochemistry and qPCR, however multiple challenges were faced with these. Multiple methods were attempted to prevent the neurons detaching following fixation, but few worked consistently, and they yielded variable results. As neurons tended to detach immediately post-fixation, the protocol was adapted slightly to either fix the cells with PFA for a slightly shorter time period or add the PFA directly to the neuronal media at a 1:1 ratio. Both of these techniques were used with limited success, as the neurons were still too fragile, and the majority detached in the first three washes with PBS. The method that yielded the best results was to supplement the neuronal media with laminin a week before fixation. The laminin meant that the neurons withstood the fixation procedure, but by the final wash many of the neurons had still detached, leaving only a few to visualise. Moving forward, it would be beneficial to optimise the cellular matrix again for these neurons, as it is evident that it is not optimal for these neurons.

This chapter describes the generation and characterisation of iPSC lines deficient in functional FAN1 in an isogenic system. These cells are thus an excellent model system for investigating the effects of FAN1 on HD pathogenesis. They also provide an excellent tool for use in a battery of experiments looking at DNA damage and repair, gene expression studies and chromosomal aberrations.

4.5 Chapter Summary

- Using CRISPR-Cas9, FAN1 was successfully knocked-out of cell lines with an expanded (>109Q) *HTT* CAG repeat and a wild-type (22Q) *HTT* CAG repeat.
- FAN1 knock-out was confirmed by reduced *FAN1* expression in a qRT-PCR and the loss of FAN1 in a Western Blot.
- The loss of FAN had no discernible effect on chromosomal stability in iPSCs over >30 passages.
- FAN1$^{-/-}$ iPSCs showed increased vulnerability to cross-linking agent MMC, indicating the absence of functional FAN1.
- FAN1$^{-/-}$ iPSCs maintained pluripotency and differentiation potential.

Chapter 5: Investigating the effect of FAN1 knock-out on CAG repeat expansions and associated phenotypes

5.1 Introduction

Since the identification of FAN1 as a potential modifier of age at onset in HD in 2015 (GeM-HD Consortium 2019), a number of papers have sought to determine whether the signal in the GWAS is underlain by FAN1, and what role FAN1 might play in HD pathogenesis. FAN1 was an attractive candidate due to its role in DNA repair, both in the resolution of interstrand-cross links (Liu *et al.* 2010), replication fork metabolism (Lachaud *et al.* 2016) and homologous recombination (Kratz *et al.* 2010), and its interactions with a number of mismatch repair proteins (MacKay *et al.* 2010). It has long been reported that DNA repair proteins are involved with HD pathogenesis, and it's thought that they might act through somatic expansion of the CAG repeat. Indeed, a number of mismatch repair (MMR) proteins; Msh2, Msh3, Mlh1 and Pms2 and base excision repair (BER) proteins; Ogg1 and Neil1 are required for expansion of the repeat in mouse models of HD (Wheeler *et al.* 2003; Kovtun *et al.* 2007; Møllersen *et al.* 2012; Pinto *et al.* 2013; Lai *et al.* 2016). The modulation of repeat expansions appears gene-specific, as the loss of MMR protein Msh6 (Dragileva *et al.* 2009), or the loss of the glycosylase NTH1 (Kovtun *et al.* 2007), another BER protein, bares little effect on repeat instability.

The research into the links between FAN1 and HD is still in its infancy. Whilst genome wide significant signals have been identified at the genomic locus of *FAN1* in HD (GeM-HD Consortium 2015; GeM-HD Consortium 2019) and other polyglutamine disorders (Bettencourt *et al.* 2016), functional research is still exploratory, and to date only one published paper has identified a functional link between HD and FAN1 (Goold *et al.* 2019). FAN1 is highly expressed in the brain (Ardlie *et al.* 2015) and transcriptome-wide association study results revealed a significant association between increased *FAN1* expression and a delayed age at onset of HD and slower disease progression (Goold *et al.* 2019). Furthermore, knock-down of *FAN1* increased *HTT* CAG repeat expansions in U2OS cells and HD-iPSCs. Chromatin immunoprecipitation assays have confirmed the interaction of FAN1 with both a wild-type and expanded *HTT* CAG repeat in U2OS cells and lymphoblastoid cell lines from HD patients (Goold *et al.* 2019). Interestingly, the authors reported that the nuclease activity of FAN1 was not required for protecting against *HTT* CAG repeat expansions in U2OS cells, suggesting it might act in a nuclease-independent manner to modulate repeat expansions. The protective effects of FAN1 appear to extend beyond HD to other repeat expansion disorders; homozygous loss of Fan1 in a mouse model of Fragile-X resulted in increased CGG repeat expansions in somatic tissues, the brain and liver (X.-N. N. Zhao and Usdin 2018), but had no effect on intergenerational repeat expansions. These papers provide support

for the proposition of a mechanism at the level of the DNA that might underpin disease pathogenesis in a number of trinucleotide repeat disorders.

Repetitive sequences, including CAG repeats have been shown to form non-helical structures such as hairpins, G-quadruplexes and R-loops (Mirkin 2007; McMurray 2010; Iyer *et al.* 2015). Theoretically, these structures could form any time single stranded DNA is exposed and therefore may form transiently during replication, repair, recombination or transcription (McGinty and Mirkin 2018). Structure stability is largely governed by the length of the repeat tract, longer repeats result in more energetically stable structures and thus they occur at a higher frequency (Marquis Gacy *et al.* 1995). If these single-stranded looped structures occur on a newly synthesised DNA strand, they may be incorporated into the duplexed DNA and cause expansions, conversely if they occur on the leading strand, contractions may occur. Somatic expansion of the *HTT* CAG repeat likely contributes to disease phenotype through increased toxicity of RNA or mHTT protein. Understanding factors governing repeat expansions are therefore of great importance.

Interestingly, the accumulation of oxidative lesions through normal ageing has been shown to enhance repeat expansions, likely through the generation of a single strand break during base excision repair (Kovtun *et al.* 2007). Exposure to oxidising agents such as potassium bromate ($KBrO_3$) induced *Fmr1* CGG repeat expansions in a mouse model of Fragile X (Entezam *et al.* 2010), and exposure of HD fibroblasts to hydrogen peroxide induced *HTT* CAG repeat expansion (Kovtun *et al.* 2007). Repetitive regions represent hotspots for oxidative damage due to the abundance of guanine residues in the expanded repeat tract (i.e. CAG, CGG, CTG) which are preferentially oxidised to 7,8-dihyrdro-8-oxoguanine (8-oxoG) (Jarem *et al.* 2009). In HD, increased levels of ROS are generated, and often accumulate as a result of mitochondrial dysfunction (Wyttenbach 2002; Pérez-Severiano *et al.* 2004; Firdaus *et al.* 2006). These findings suggest that a toxic oxidation loop likely exists, whereby further expansions are potentiated by increased endogenous levels of ROS. Examining these links and the role that FAN1 might play is therefore imperative in our understanding of factors governing repeat expansions and phenotypes associated with an expanded *HTT* CAG repeat.

5.2 Chapter Aims

Utilising cell lines generated and characterised in Chapters 3 and 4, this chapter primarily aims to look at the effect of FAN1 knock-out on *HTT* CAG repeat expansions in iPSCs and post-mitotic neurons. It will then explore the association between the length of the *HTT* CAG repeat tract and mitochondrial functioning and response to exogenous stressors, and the effect that the loss of FAN1 might have on each of those parameters.

5.3 Results

5.3.1 *HTT* CAG repeat Instability in FAN1$^{-/-}$ iPSCs

To ascertain whether FAN1 has any effect on *HTT* CAG repeat length in proliferating cells, cells were cultured over a number of passages and genomic DNA was extracted at each passage to allow us to follow changes in repeat length closely (see 2.2.1). *HTT* exon 1 PCR amplicons were capillary electrophoresed using the GA3130*x*L Genetic Analyser (see 2.6). Each peak in the electropherogram represents one *HTT* CAG unit. *HTT* CAG repeat lengths were determined from the size of the highest peak in the electropherogram, which represents the modal *HTT* CAG repeat length (Figure 5.1A). Traces roughly follow a Gaussian distribution, with peaks immediately to the left of the modal peak often considered as PCR stutter, whilst peaks to the right of the modal peak represent mosaicism of the *HTT* CAG repeat (Lee *et al.* 2011). As *HTT* CAG repeat instability occurs and the population becomes more mosaic the peak distribution often broadens. Cells were initially cultured for 11 passages over ~40 days, as in Goold *et al.* (2019). A population shift in the electropherogram traces is observed between samples at early and late passage and this is most marked in the 109-FAN1$^{-/-}$ iPSCs, where modal *HTT* CAG expansions were greatest (Figure 5.1A).

The loss of FAN1 from the HD-iPSCs resulted in a significant increase in CAG repeat expansions compared with those seen in 109-FAN1$^{+/+}$ iPSCs (Figure 5.1B). This is consistent with previous reports suggesting that FAN1 is protective against triplet repeat expansions (X.-N. N. Zhao and Usdin 2018; Goold *et al.* 2019). In the 109-FAN1$^{+/+}$ iPSCs the *HTT* CAG repeat expanded at an average rate of 1 CAG repeat increase per 23.5 days, compared with an average rate of 1 CAG repeat increase per 9.9 days in 109-FAN1$^{-/-}$ iPSCs. As only two 109-FAN1$^{-/-}$ lines were characterised, expansion data in the iPSCs represents two biological replicates of these two lines (N=4, n=2). Data from the 109-FAN1$^{+/+}$ lines, however, represents the parent Q109N1 line and four Q109N1 subclones (n=5). Clones of each genotype are consistent in their rates of expansion. These data suggest that the effects of FAN1 knock-out are detectable in iPSCs over short periods of time in culture and could present a window of time for assaying genetic modifiers of HD. It should be noted that whilst baseline *HTT* CAG repeat lengths differ in each of the lines, there does not appear to be a length-dependent effect on expansions seen.

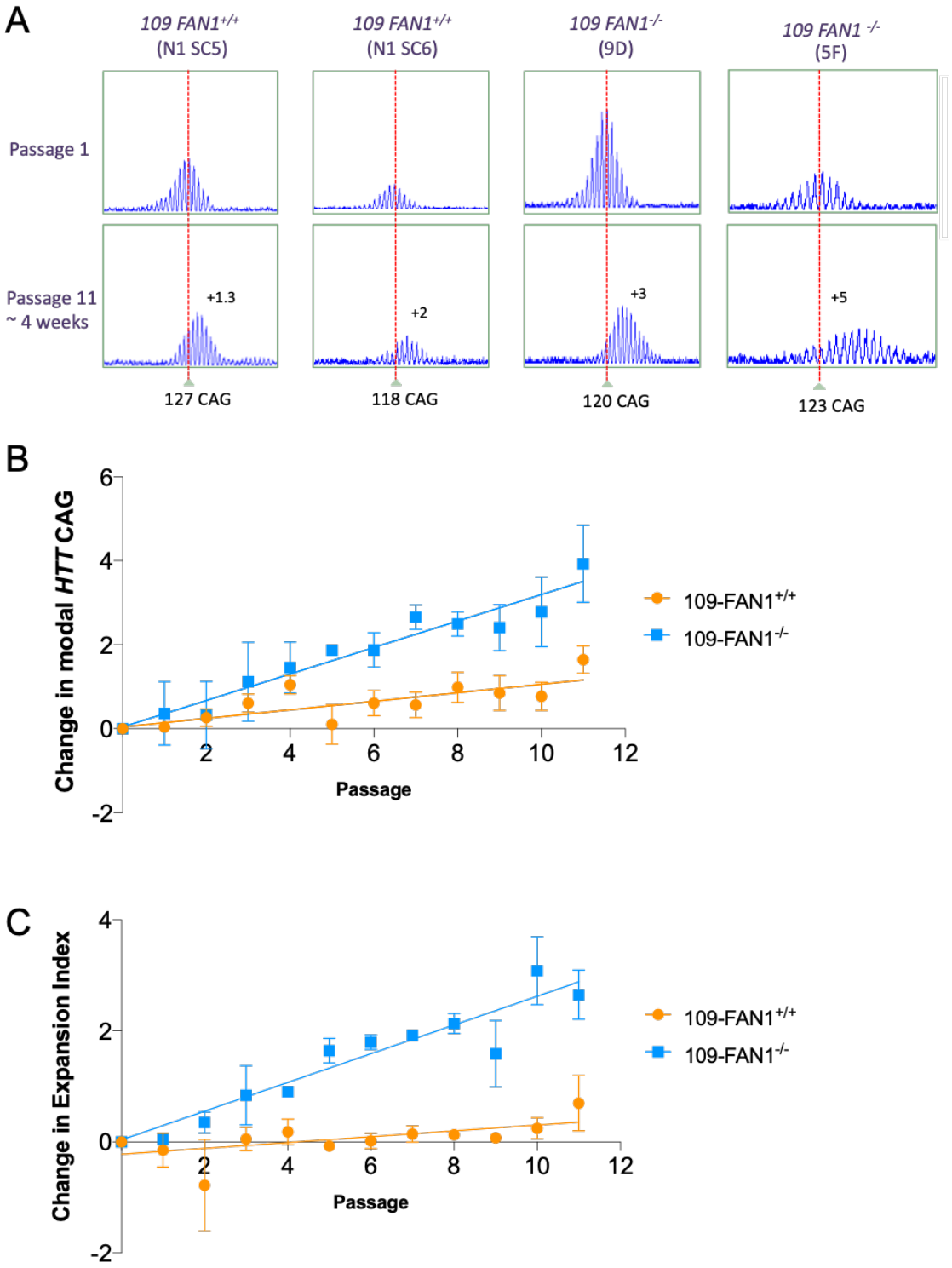


Figure 5.1 109-FAN1$^{-/-}$ iPSCs exhibit increased *HTT* CAG repeat expansion in culture. (A) Representative electropherograms of fluorescent PCR and capillary electrophoresis of the *HTT* CAG repeat in 109-FAN1$^{-/-}$ iPSCs compared to 109-FAN1$^{+/+}$ iPSCs. The red dotted line indicates baseline *HTT* CAG repeat length. (B) Change in modal *HTT* CAG repeat lengths in 109-FAN1$^{-/-}$(N=4, n=2) iPSCs compared to 109-FAN1$^{+/+}$ iPSCs (n=5). $p<0.05$ as determined by a mixed-effects analysis followed by Sidak's multiple comparisons test. (C) Change in expansion index of 109-FAN1$^{-/-}$ iPSCs and 109-FAN1$^{+/+}$ iPSCs. $p<0.001$ as determined by a mixed-effects analysis followed by Sidak's multiple comparisons test. Samples run in duplicate. Values expressed as mean ± SEM.

The expansion index can be determined using previously published methods (Lee *et al.* 2010). This measure takes both change in *HTT* CAG repeat length and number of peaks larger and smaller than the modal peak into account. Broadening of the electropherogram trace is often seen in tissue samples showing high repeat instability, and therefore, quantification of this can provide a useful measure of *HTT* CAG repeat instability. As peaks to the left of the modal peak are often considered PCR stutter, the expansion index is determined using peaks to the right of the modal baseline peak. There was a significantly greater change in expansion index in the 109-FAN1$^{-/-}$ iPSCs compared with 109-FAN1$^{+/+}$ iPSCs (Figure 5.1C).

To examine *HTT* CAG expansion over an extended period of time in culture (30 passages, ~105 days), two clones of each genotype; 22-FAN1$^{-/-}$, 22-FAN1$^{+/+}$, 109-FAN1$^{-/-}$ and 109-FAN1$^{+/+}$ were cultured and genomic DNA was extracted routinely. CAG expansion in repeats below a threshold (~27 CAG) is seldom seen, but it is likely that it occurs through *de novo* mutations (Myers *et al.* 1993). We therefore tested whether a wild-type repeat existing in the genetic background of an individual who had HD would expand, and whether loss of FAN1 would affect expansion. Expansion of the *HTT* CAG repeat was length dependent; the *HTT* CAG repeat length remained stable in both 22-FAN1$^{-/-}$ and 22-FAN1$^{+/+}$ iPSCs (Figure 5.2), indicating that there was no effect of the loss of FAN1 on cells with a wild-type repeat length. Minor variation was seen around the repeat length, but this can be attributed to variation in the capillary electrophoresis analysis. All samples were therefore analysed by capillary electrophoresis in duplicate, and for each experiment samples were analysed on the same run.

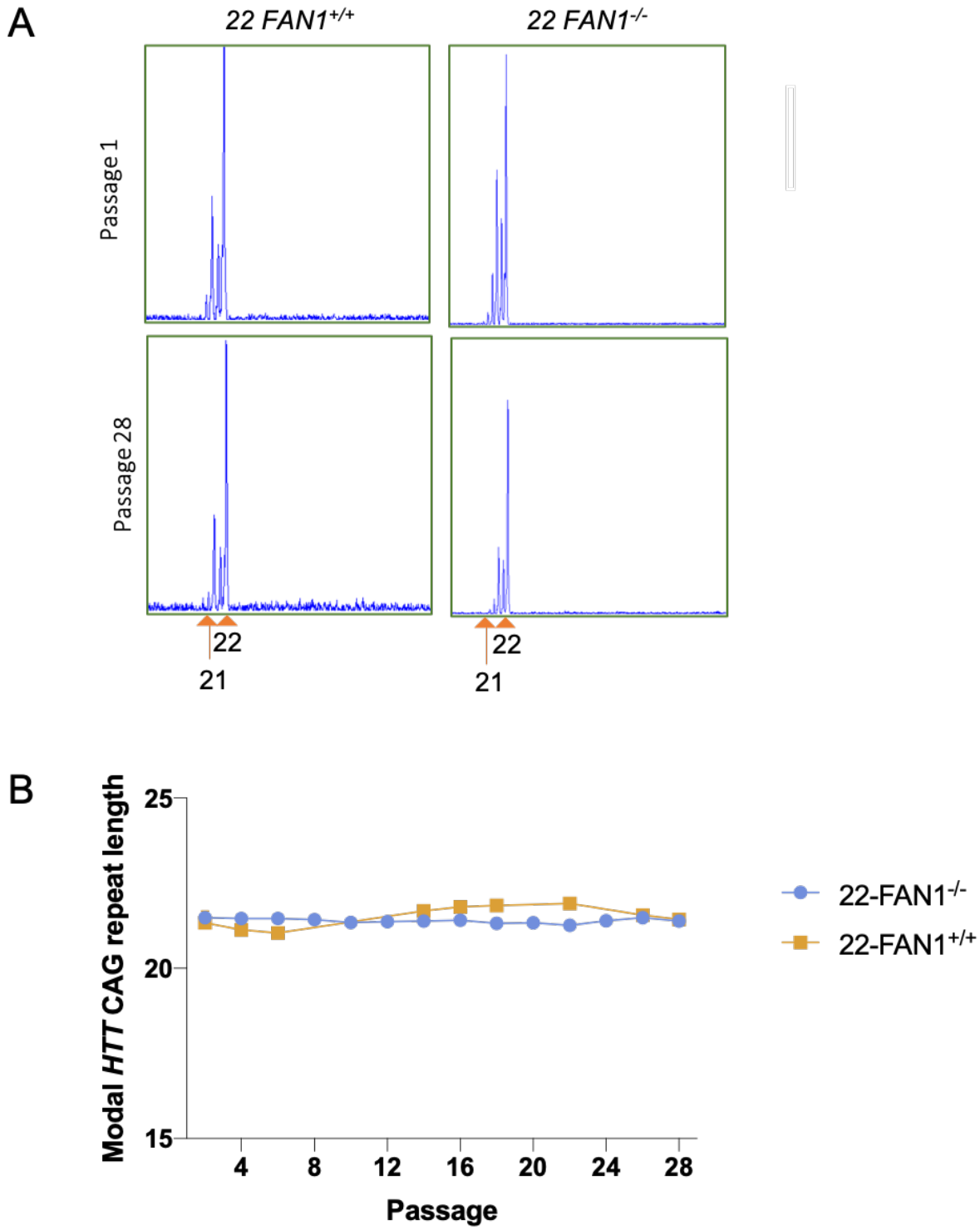


Figure 5.2 22-FAN1$^{-/-}$ and FAN1$^{+/+}$ iPSCs do not demonstrate *HTT* CAG repeat expansions in culture (A) Representative electropherograms of fluorescent PCR and capillary electrophoresis of the *HTT* CAG repeat in 22-FAN1$^{-/-}$ iPSCs (n=2) compared to 22-FAN1$^{+/+}$ iPSCs (n=2). Both wild-type alleles are seen with *HTT* CAG repeat lengths of 21 and 22. No expansion was observed. (B) Plots of modal CAG repeat in both 22-FAN1$^{+/+}$ and 22-FAN$^{-/-}$ iPSCs shows no change in *HTT* CAG repeat length over 28 passages. Samples run in duplicate. Values expressed as mean ± SEM.

Expansion of the *HTT* CAG repeat was seen in each of the lines with an expanded *HTT* CAG repeat in prolonged iPSC culture. *HTT* CAG expansion rates were significantly higher in 109-FAN1$^{-/-}$-C2 (Figure 5.3B) than both 109-FAN1$^{+/+}$-C1 (Figure 5.3C) and 109-FAN1$^{+/+}$-C2 (Figure 5.3D) with an increase of 1 *HTT* CAG every 10.75 days compared to an increase of 1 *HTT* CAG every 16.5 days and 17.8 days, respectively. In one of the lines, 109-FAN1$^{-/-}$-C1, a cell-death phenotype was observed between passages 13 and 14, where embryoid bodies formed from iPSC colonies, followed by considerable cell detachment for 24-48 hours. Cells that remained adhered demonstrated slow growth and initially passaged poorly. To ensure the cellular matrix was not responsible for cell detachment, detached cells were replated onto a tissue culture plate containing fresh matrix and no attachment of the cells was observed, suggesting they had detached due to cell death. This phenotype was observed three times in the same line at a similar passage. Despite their initial slow growth, the cells that remained adhered grew well as colonies and had a similar morphology to iPSCs in other clonal lines (Figure 5.4D), suggesting that the cells that survived likely did so because of an adaptive survival advantage. *HTT* CAG repeat analysis revealed a sudden decrease (3 CAGs) in *HTT* CAG repeat length in this clone following recovery (Figure 5.3A). This is illustrated in the CAG peak distribution changes observed in electropherograms of these samples (Figure 5.4). At passage 12 the electropherograms showed a broad peak, indicative of a population with a range of *HTT* CAG repeat lengths, which was reduced following the observed cell death (Figure 5.4). Whilst the *HTT* CAG repeat length 109-FAN1$^{-/-}$-C1 continued to expand, it was at a much slower rate; an average increase of 1 CAG every 17.5 days compared to the expansion rate seen prior to passage 14 of an average increase of 1 CAG every 10.7 days. There were no significant differences in the expansion rate of 109-FAN1$^{-/-}$-C1 compared with 22-FAN1$^{-/-}$-C1 or 22-FAN1$^{-/-}$-C2.

Expansion events appear to be stochastic, with each subclone exhibiting a different expansion profile (Figure 5.3). Overlaying the expansion traces illustrates the comparative expansion trace for each of the subclones (Figure 5.3E). Quantifying the expansion index produces a similar graph (Figure 5.3F), where expansion indices of all subclones increase with time. At the later passages, there does not appear to be an effect of the loss of FAN1 on expansion indices, as that of 109-FAN1$^{-/-}$-C1 and 109-FAN1$^{+/+}$-C1 are comparable, as are 109-FAN1$^{-/-}$-C2 and 109-FAN1$^{+/+}$-C2.

A steady increase in *HTT* CAG repeat length over time was observed in 109 FAN1$^{-/-}$ lines (Figure 5.4A,B), whereas multiple distinct peaks were often observed in the 109-FAN1$^{+/+}$ lines (Figure 5.4C). The presence of two main peaks within an electropherogram trace is known as bimodality and is illustrative of a population of cells which have exhibited an expansion or a contraction of the *HTT* CAG repeat. Bimodality in the FAN1$^{+/+}$ iPSCs often coincided with spontaneous differentiation in culture (Figure 5.4D). Although this observation might be coincidental, heterogeneity in tissues has been shown to give

rise to bimodal peak distribution, and this could be reflected in these data. It's not clear however, whether clonality might drive repeat instability, or if repeat instability increases the propensity for spontaneous differentiation. Indeed, whether these two phenomena are linked, or in fact a result of a duplication of chromosome 1, should be investigated further.

Whilst these data support a role of FAN1 in modulation of the *HTT* CAG repeat, particularly in the first 40 days, the steady expansion rates in both $FAN1^{+/+}$ and $FAN1^{-/-}$ imply that multiple mechanisms affect repeat instability, and suitable controls should be in place that prevent inappropriate conclusions being drawn.

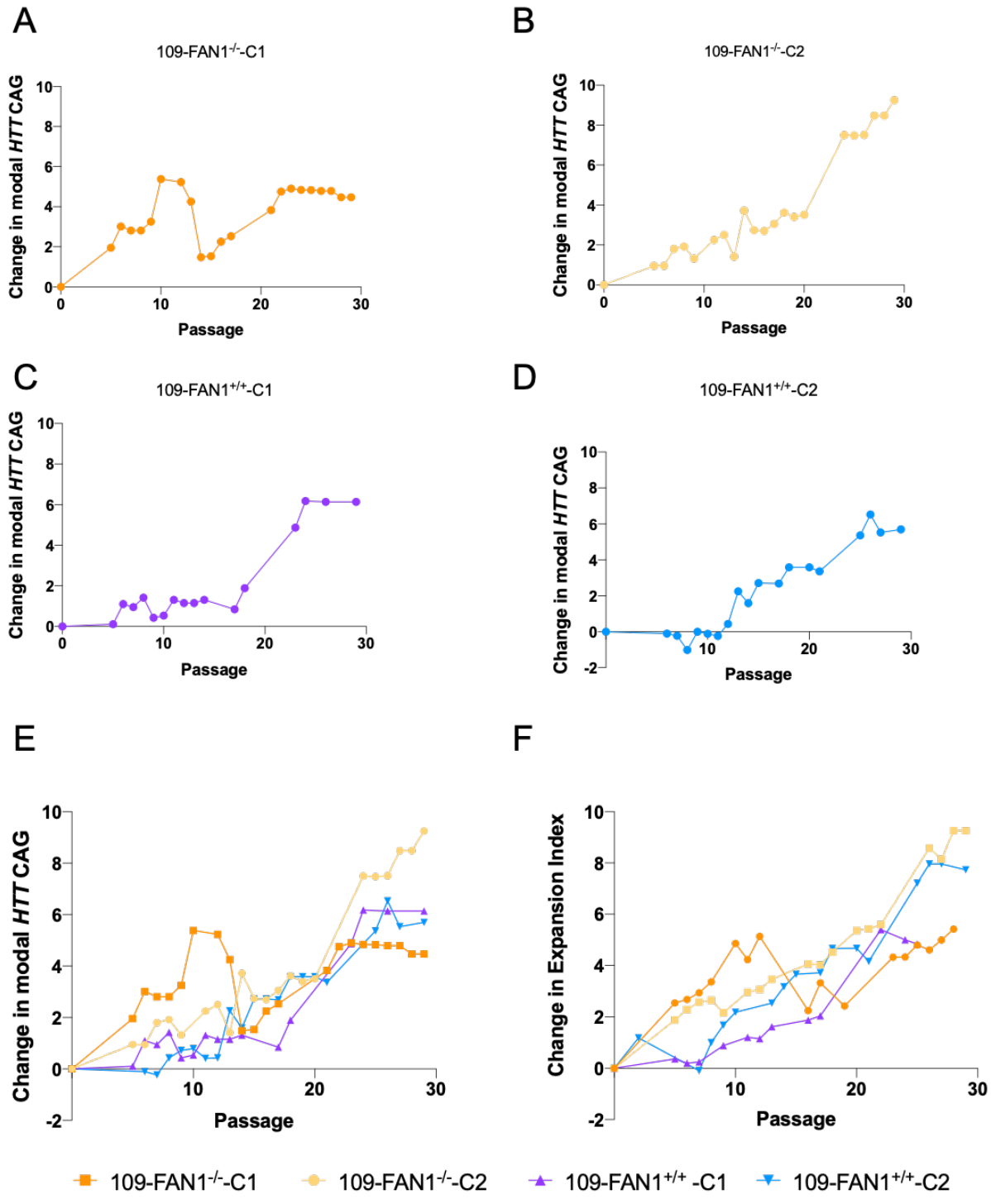


Figure 5.3 *HTT* CAG expansion in 109-FAN1$^{-/-}$ and 109-FAN1$^{+/+}$ iPSCs cultured over an extended period of time. Change in modal *HTT* CAG repeat length over 30 passages in (A) 109-FAN1$^{-/-}$-C1, (B) 109-FAN1$^{-/-}$-C2, (C) 109-FAN$^{+/+}$-C1, (D) 109-FAN$^{+/+}$-C2, (E) all four iPSC lines and (F) Change in the expansion index over 30 passages in each of the iPSC lines.

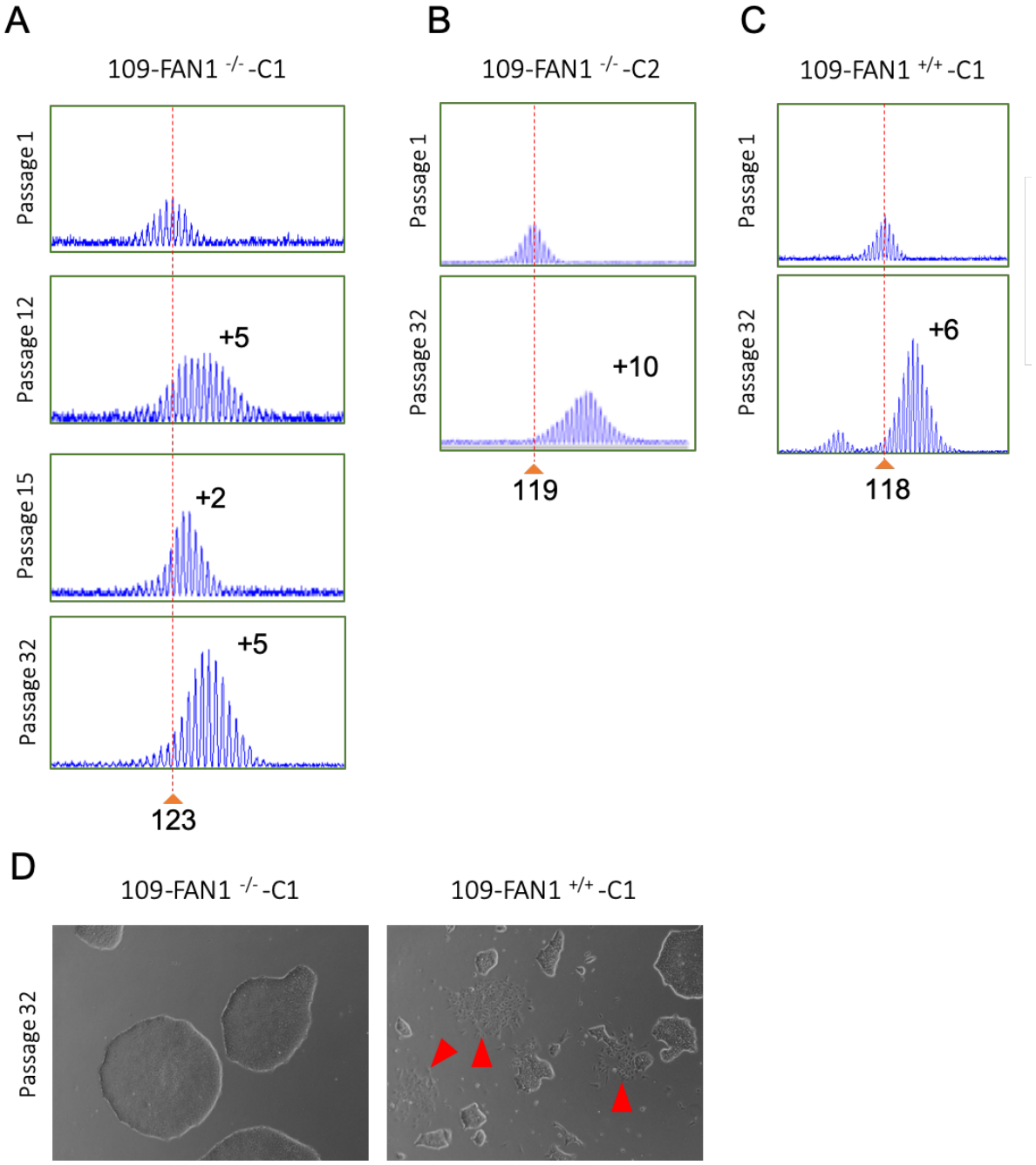


Figure 5.4 *HTT* CAG expansion and spontaneous differentiation in prolonged culture of iPSCs. Representative electropherograms of fluorescent PCR and capillary electrophoresis of the *HTT* CAG repeat in (A) 109-FAN1$^{-/-}$-C1 (B) 109-FAN1$^{-/-}$-C2 and (C) 109-FAN$^{+/+}$-C1. The red dotted line indicates baseline *HTT* CAG repeat length. (D) Brightfield images of iPSCs showing pluripotent colonies at passage 32. Red arrows highlight spontaneous differentiation.

5.3.2 *HTT* CAG repeat Instability in FAN1$^{-/-}$ iPSC-derived neurons

To disentangle the effects of division, spontaneous differentiation and the stress of manual passaging, iPSCs were differentiated into NPCs and then further differentiated as post-mitotic neurons (see 4.3.9). Day 0 represents the day NPCs were plated for terminal neuronal differentiation and thus are post-mitotic. There was a significantly higher rate of expansion in the 109-FAN1$^{-/-}$ neurons (N=4, n=2) than the 109-FAN1$^{+/+}$ neurons (n= 5) (Figure 5.5A). This difference was evident in all subclones. The 109-FAN1$^{-/-}$ neurons expanded at an average rate of 1 *HTT* CAG repeat every 21.2 days, whilst 109-FAN1$^{+/+}$ neurons expanded at an average rate of 1 *HTT* CAG repeat every 105.6 days. These differences were reflected in the expansion indices, where expansion indices were significantly higher in 109-FAN1$^{-/-}$ lines compared with the 109-FAN1$^{+/+}$ lines in the latter weeks of the differentiation (Figure 5.5B). The expansion rates demonstrated in neurons were lower than those observed in iPSCs, supporting the notion that a myriad of factors influence *HTT* CAG expansion rates. As in the iPSCs, no expansion was seen in the 22-FAN1$^{+/+}$ or 22-FAN$^{-/-}$ neurons (Figure 5.5C).

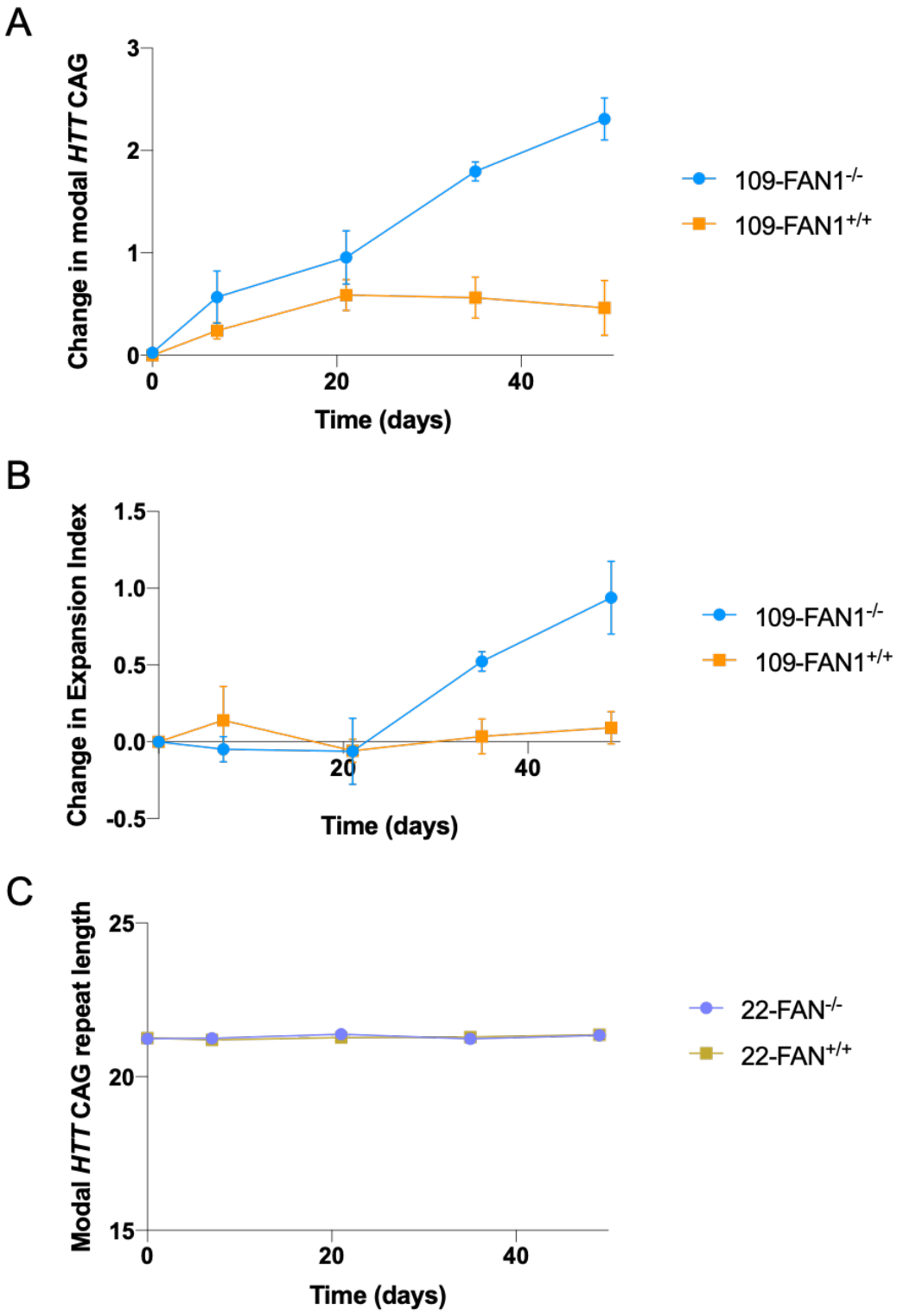


Figure 5.5 *HTT CAG repeat expansion in neurons derived from 109-FAN1$^{-/-}$ and 22-FAN1$^{-/-}$ iPSCs. (A) Change in modal HTT CAG repeat length in* 109-FAN1$^{-/-}$ neurons (N=4, n=2) compared to 109-FAN1$^{+/+}$ neurons (n=5). $p<0.05$ as determined by a mixed-effects analysis followed by Sidak's multiple comparisons test. Day 0 represents plating NPCs for terminal neuronal differentiation (B) change in expansion index of 109-FAN1$^{-/-}$ neurons compared to 109-FAN1$^{+/+}$ neurons. $p<0.001$ as determined by a mixed-effects analysis followed by Sidak's multiple comparisons test. (C) Modal *HTT* CAG repeat length in 22-FAN1$^{-/-}$ (n=2) and 22-FAN1$^{+/+}$ (n=2). For each time point two samples of each genotype were harvested. Samples run in duplicate. Values expressed as mean ± SEM.

5.3.3 DNA damage assays

5.3.3.1 Endogenous DNA damage in HD-iPSCs

Although many avenues point to a role of DNA repair machinery in HD, few papers have looked at the interplay between the *HTT* CAG repeat and DNA damage in an isogenic context. Using an isogenic system, we were able to look at DNA maintenance systems in the context of repeat length and examine the effect of genetic modifiers on the system. To investigate whether differences exist in DNA damage we first looked at levels of poly- ADP-ribose (PAR) staining between iPSCs and neurons with either 22 CAG repeats or >109 CAG repeats. ADP-ribosyltransferases, such as poly(ADP-ribose) polymerase 1 (PARP-1), 2 (PARP-2) and 3 (PARP-3) are rapidly recruited to sites of DNA damage following both single strand and double strand breaks (Boehler *et al.* 2011). PARPs are also involved in the stabilisation of replication forks and chromatin remodelling (Bryant *et al.* 2009; Ray Chaudhuri and Nussenzweig 2017) and resolution of interstrand cross-links (Gunn *et al.* 2016). PARPs use nicotinamide adenine dinucleotide (NAD^+) as a substrate to generate and transfer PAR molecules to target proteins, a process known as ADP ribosylation, to aid in DNA repair (Gupte *et al.* 2017). Following DNA damage, the activation of PARPs leads to a decrease in NAD^+ and an increase in PAR; staining of which can reveal endogenous levels of DNA damage. Preliminary data demonstrated a trend towards a significant difference in PAR immunostaining between cells with an expanded *HTT* CAG repeat and cells with a wild-type length repeat in iPSCs and neurons (Figure 5.6).

To determine relative levels of double strand breaks *in vitro*, cells can be stained with γ-H2A.X (Figure 5.7A). When a double strand break occurs, the first step in recruitment of additional DNA repair proteins is the phosphorylation of histone H2A.X (Kuo and Yang 2008). γ-H2A.X presents as discrete foci which can be counted and quantified to ascertain relative levels of double strand breaks in a system (Figure 5.7B). It was recently reported in an isogenic series of HD-ESCs that an ESC line with 81 *HTT* CAG repeats demonstrated significantly more γ-H2A.X foci than it's isogenic counterpart with 30 *HTT* CAG repeats (Ooi *et al.* 2019). We report a trend towards significance between the numbers of γ-H2A.X foci per cell in iPSCs with an expanded *HTT* CAG repeat compared to iPSCs with a wild-type repeat (Figure 5.7C). These findings hint at deficits in DNA repair, or accumulation of endogenous DNA damage at baseline, but experiments require repeating to determine significance.

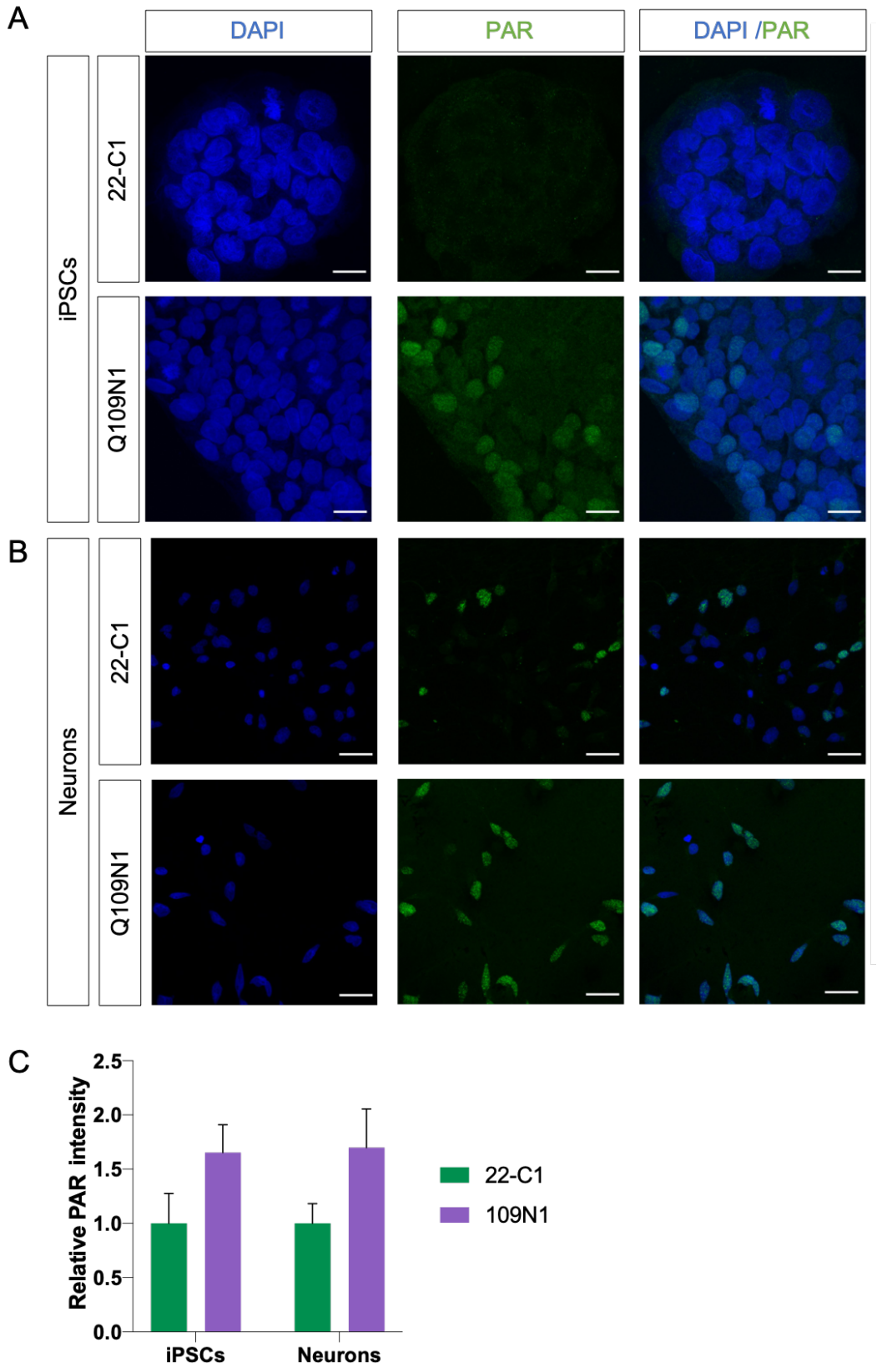


Figure 5.6 PAR staining in iPSCs and neurons harbouring an expanded *HTT* CAG repeat. PAR staining in 22-C1 (n=2) and 109N1 (n=2) (A) iPSCs and (B) neurons. Scale bars represent 20 μm. (C) Relative PAR intensities in 22-C1 and 109N1 iPSCs and neurons. Intensities were determined by Volocity software and are relative to the intensities of 22-C1. Student's unpaired t-test revealed no significant difference in iPSCs (p=0.1462) or neurons (p=0.0926)

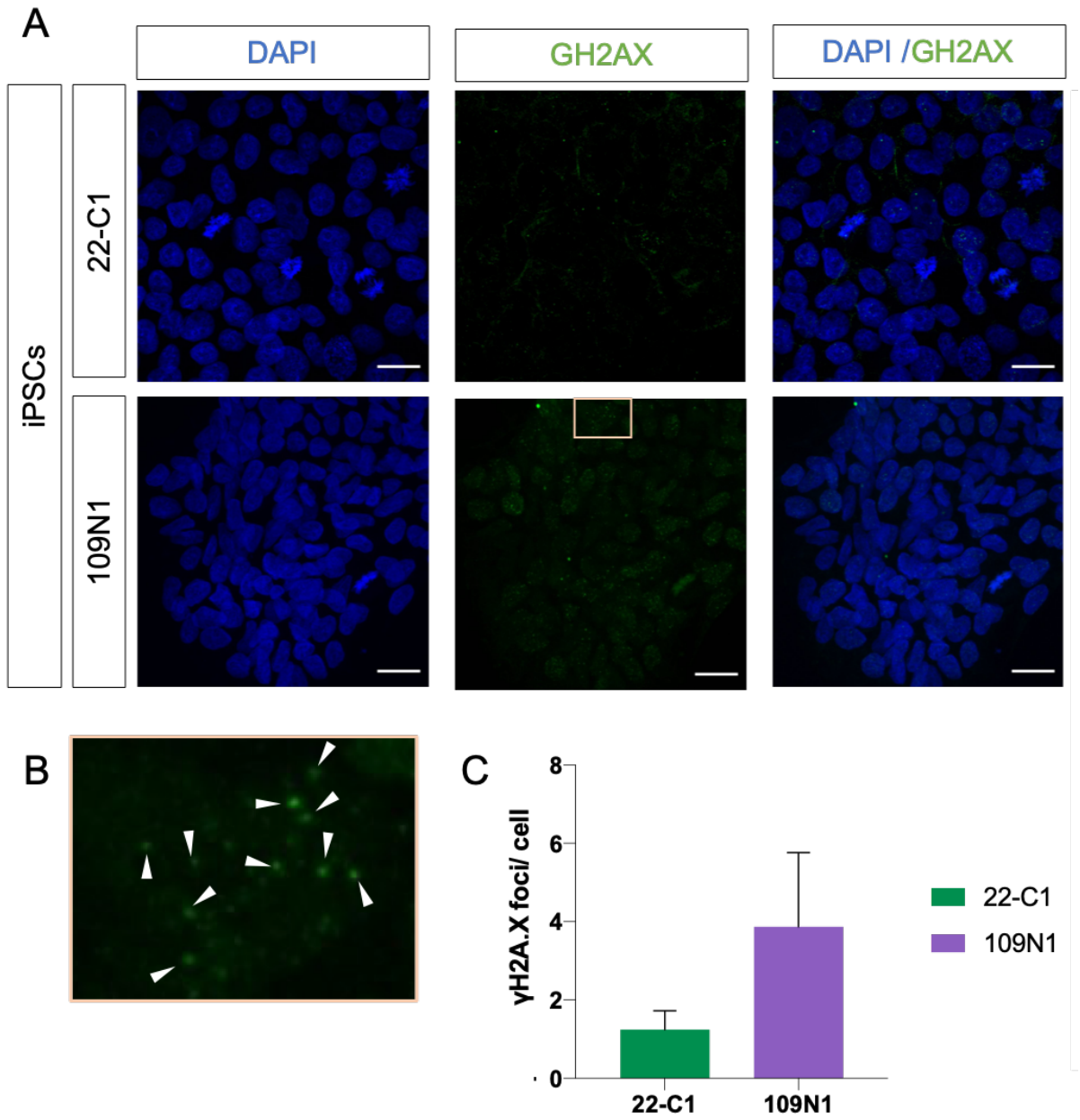
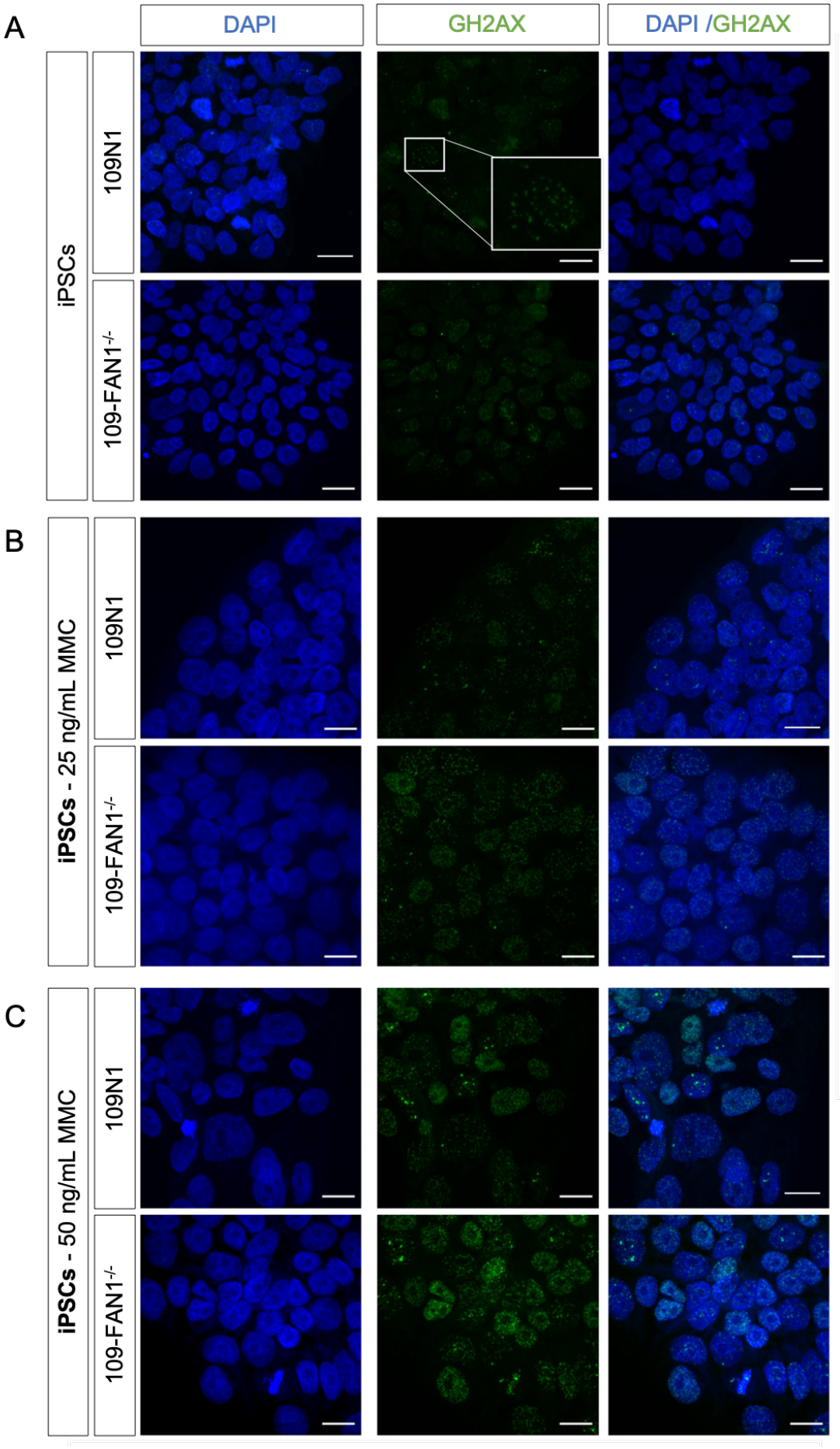


Figure 5.7 γ-H2A.X foci staining in iPSCs with an expanded HTT CAG repeat. (A) γ-H2AX staining in 22-C1 and 109N1 iPSCs. (B) Magnified section of 109N1 γ-H2AX staining showing individual γ-H2AX foci. Scale bars represent 20 μm. (C) Quantitation of the average number of γ-H2A.X foci per cell. Foci counts were determined using Volocity software. Values are expressed as mean ± SEM. Student's unpaired t-test revealed no significant difference (p=0.0841).

5.3.3.2 HD-FAN1$^{-/-}$ iPSCs demonstrate increased DNA damage following treatment with MMC

To assay the effects of FAN1 in our cell system, we treated iPSCs and neurons with mitomycin C (MMC). In Chapter 4 we confirmed lack of a functional FAN1 protein due to the increased vulnerability of 109-FAN1$^{-/-}$ lines to MMC at concentrations 10-20 ng/mL. At higher MMC concentrations however, iPSCs demonstrated equal susceptibility to MMC exposure, irrespective of FAN1 status, owing to its cytotoxic nature in culture. Cells were therefore exposed to MMC at concentrations of 25 ng/mL or 50 ng/mL for 24 hours to investigate whether this vulnerability was mirrored in the levels of the double-strand break marker, γ-H2A.X. At baseline, there were no significant differences in the levels of γ-H2AX between 109N1 and 109-FAN1$^{-/-}$-C1 (Figure 5.8A). Exposure of cells to 25 ng/mL MMC resulted in significantly more γ -H2A.X foci in the 109-FAN1$^{-/-}$-C1 iPSCs than 109N1 at 25 ng/mL MMC (Figure 5.8D), whereas after exposure to 50 ng/mL MMC, γ-H2A.X foci levels were comparable in both lines (Figure 5.8D). These data indicate a deficiency of interstrand cross-link repair in the FAN1 knock-out lines at lower concentrations of MMC, but suggest that at higher concentrations of MMC, the repair systems responsible for resolving cross-links are saturated in both cell lines, demonstrated by the high levels of γ-H2A.X foci observed in both cell lines.

Recessive loss of function mutations in FAN1 cause karyomegalic interstitial nephritis, a rare kidney disease where enlarged nuclei are observed in the kidneys and other tissues (Airik *et al.* 2016). This phenotype of karyomegalic nuclei is recapitulated in the kidneys of Fan1$^{-/-}$ mice treated with cisplatin, which like MMC is another cross-linking agent. We therefore sought to establish whether there were any observable differences in nuclear size in our cell system. As described in Chapter 4, 109-FAN1$^{-/-}$ iPSCs were on average 0.6 μm larger than 109N1 iPSCs (see 4.3.7). In this experiment, cell nucleus volume was determined using Volocity software (see 2.9.2), and revealed no significant difference in nuclear volume in 109-FAN1$^{-/-}$ iPSCs compared with 109N1 iPSCs. Though the cell nuclei are observably larger upon treatment with MMC in both genotypes, there was not a significant effect of MMC treatment on nuclear size (Figure 5.8B and Figure 5.8C).


A
DAPI
GH2AX
DAPI /GH2AX
iPSCs
109N1
109-FAN1-/-
B
iPSCs - 25 ng/mL MMC
109N1
109-FAN1-/-
C
iPSCs - 50 ng/mL MMC
109N1
109-FAN1-/-

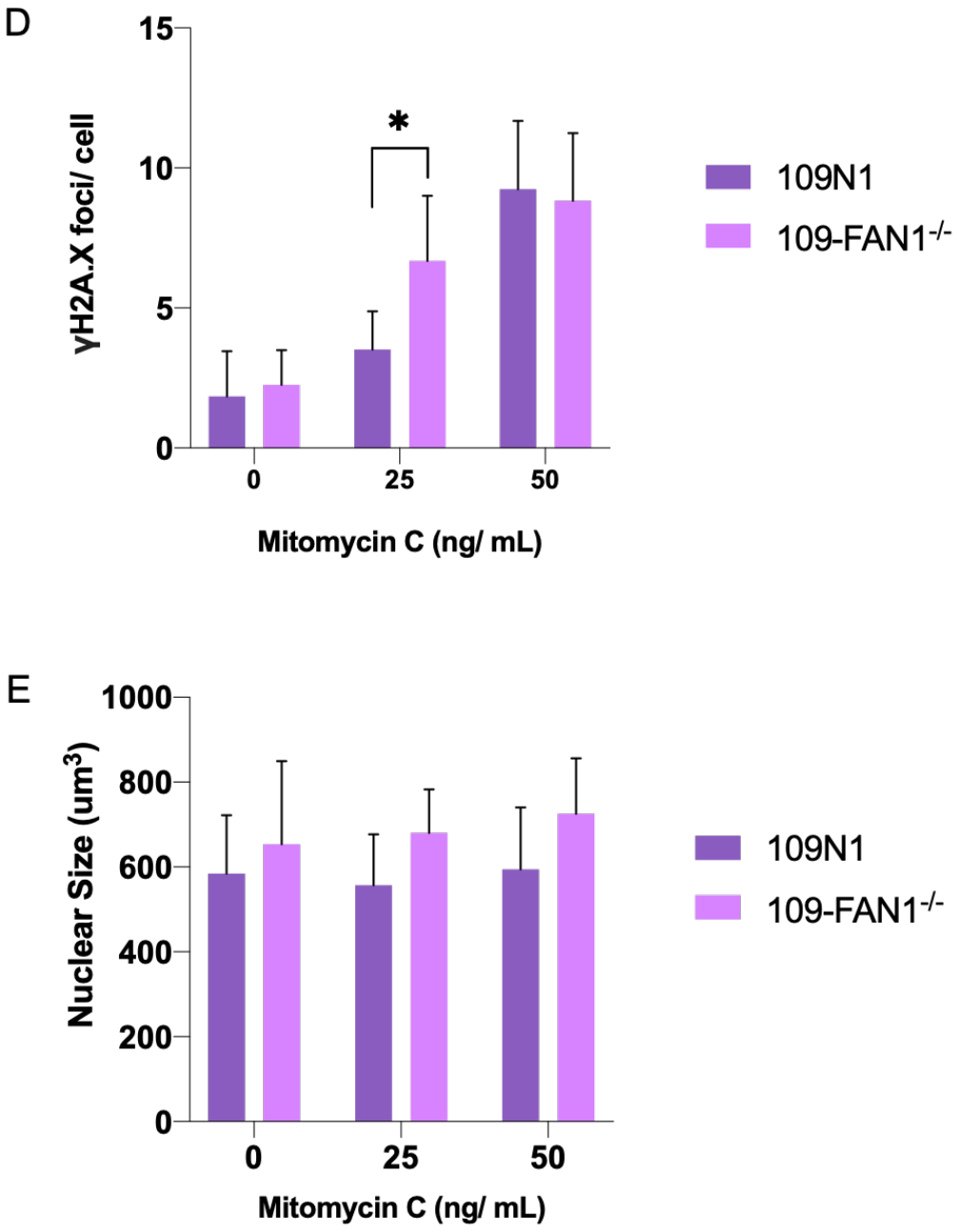
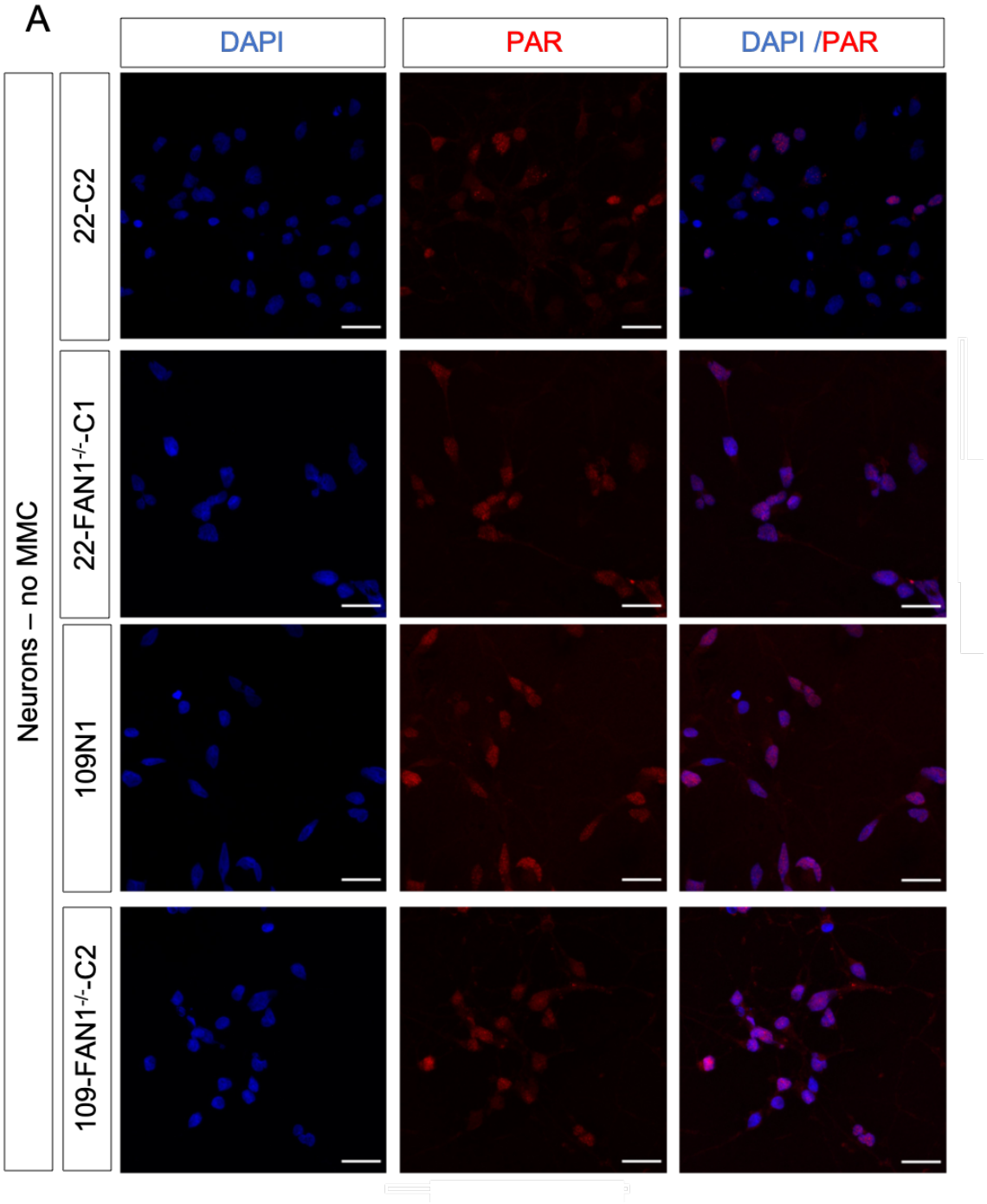


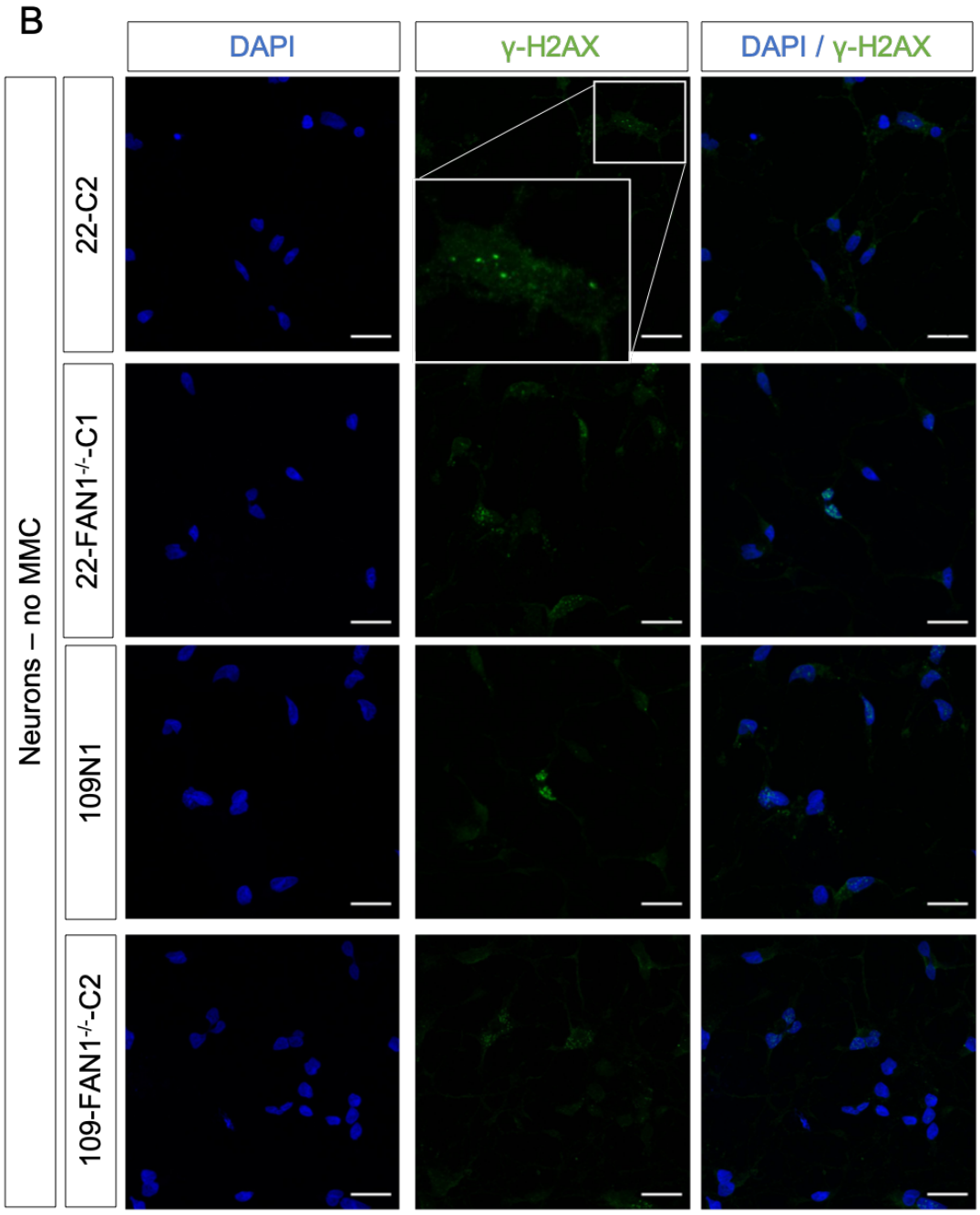
Figure 5.8 γ-H2AX staining in iPSCs following exposure to MMC. (A) γ-H2A.X foci in 109N1 and 109-FAN1$^{-/-}$ iPSCs at baseline. White box shows magnified cell nuclei containing γ-H2A.X foci for clearer representation. (B) γ-H2A.X foci in 109N1 and 109-FAN1$^{-/-}$ iPSCs at 25 ng/mL MMC or (C) 50 ng/mL MMC. Scale bars represent 20 μm (D) Quantitation of the average number of γ-H2A.X foci per cell. Foci counts were determined using Volocity software. *p<0.05, determined by a mixed-model ANOVA followed by Tukey's HSD test. Values are expressed as mean ± SEM. (E) Average nuclear size in 109N1 and 109-FAN1$^{-/-}$ iPSCs treated with increasing concentrations of MMC. Size determined by Volocity software, selecting nuclei > 450 μm^3 and ensuring correct nuclear fragmentation. n=1, 3 technical replicates. Values are expressed as mean ± SEM.

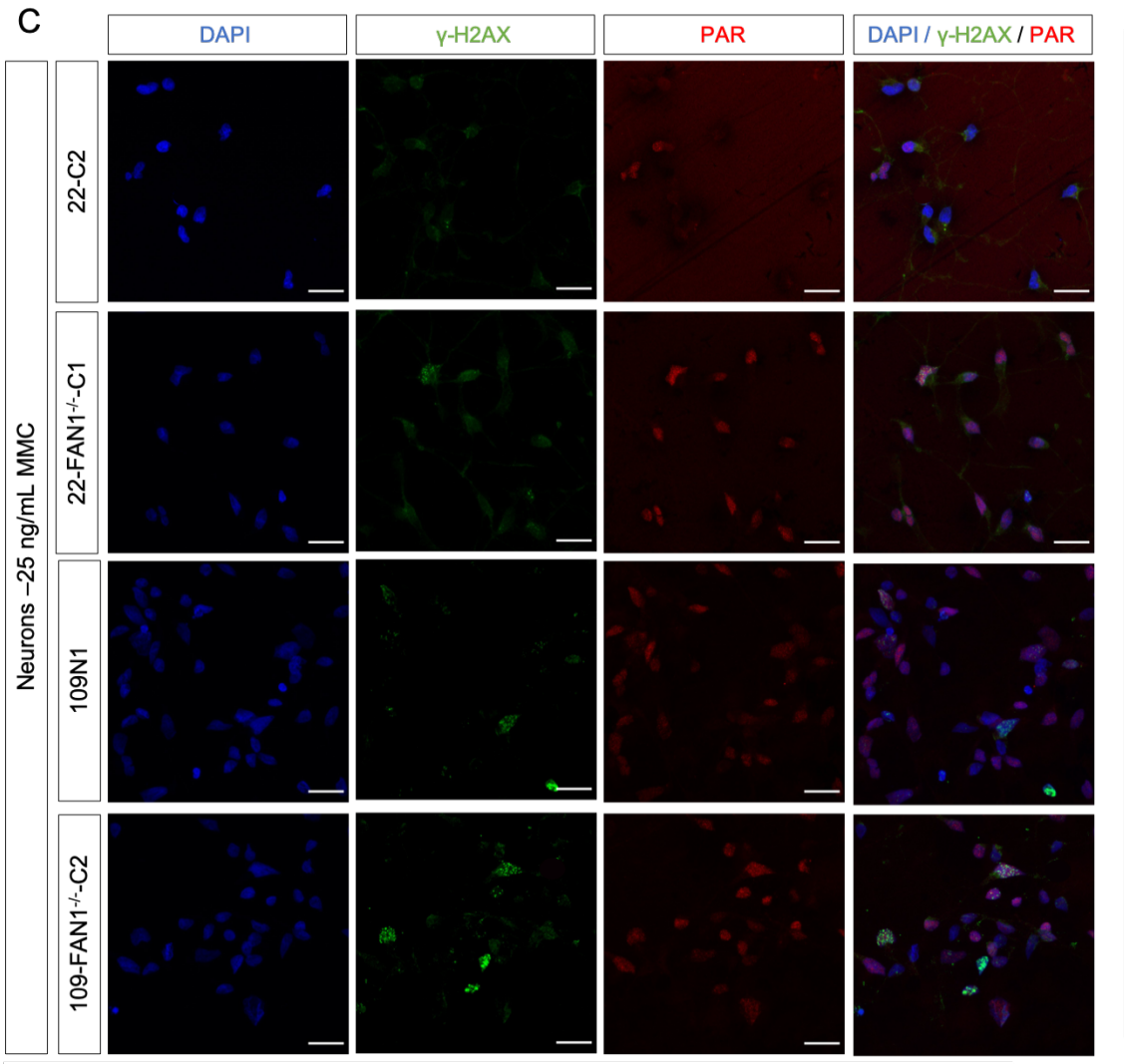
5.3.3.3 Neuronal exposure to MMC

We then sought to establish whether the response to MMC was similar in neurons, and whether the length of the *HTT* CAG repeat tract affected the DNA damage response. We stained for both single strand and double strand breaks by immunostaining with PAR and γ-H2A.X, respectively (Figure 5.9). Loss of FAN1 resulted in significantly elevated levels of PAR staining in 22-FAN1$^{-/-}$-C1 relative to 22-C2 at baseline, though there was no significant differences between any of the other genotypes (Figure 5.9A). Following exposure to 25 ng/mL MMC for 24 hours, there were no significant differences in PAR staining in any of the lines (Figure 5.9C). MMC has a half-life of only 40 minutes *in vivo*, and so these findings might be reflective of the rapid repair rates of single strand breaks in cells.

Comparing average numbers of γ-H2A.X foci per cell for each genotype at baseline revealed no significant difference between 22-FAN1$^{-/-}$, 109N1 and 109-FAN1$^{-/-}$ or 22-C2. Following exposure of each neuronal cell line to 25 ng/mL of MMC, the average number of γ-H2A.X foci per cell observably increased in each line, though findings were not significant. This is due to variability and lack of repetition. The data have large error bars, indicative of the considerable differences in foci counts observed between images of the same sample on different coverslips and the experiments bear repeating to establish an effect.


A
DAPI
PAR
DAPI /PAR
Neurons – no MMC
22-C2
22-FAN1-/--C1
109N1
109-FAN1-/--C2


B
DAPI
γ-H2AX
DAPI / γ-H2AX
Neurons – no MMC
22-C2
22-FAN1-/--C1
109N1
109-FAN1-/--C2


C
DAPI
γ-H2AX
PAR
DAPI / γ-H2AX / PAR
Neurons –25 ng/mL MMC
22-C2
22-FAN1-/--C1
109N1
109-FAN1-/--C2

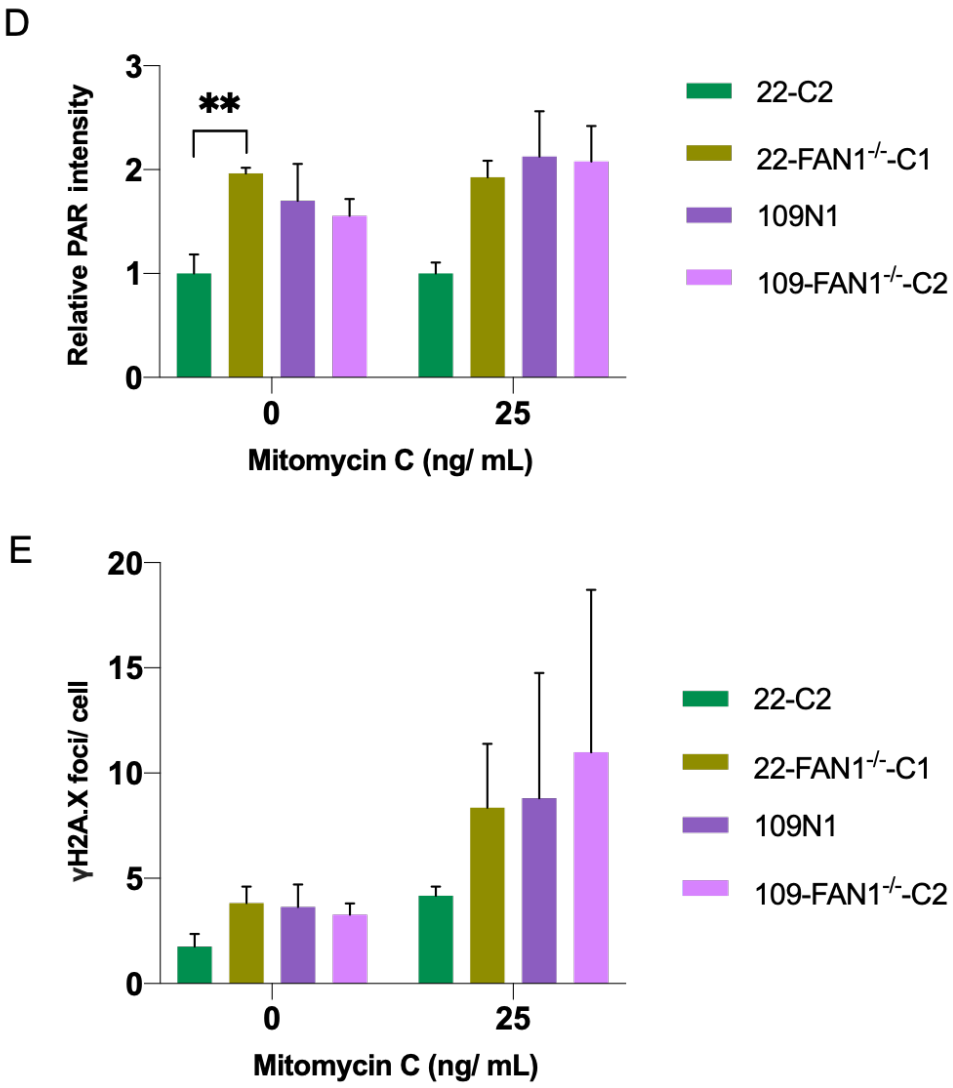


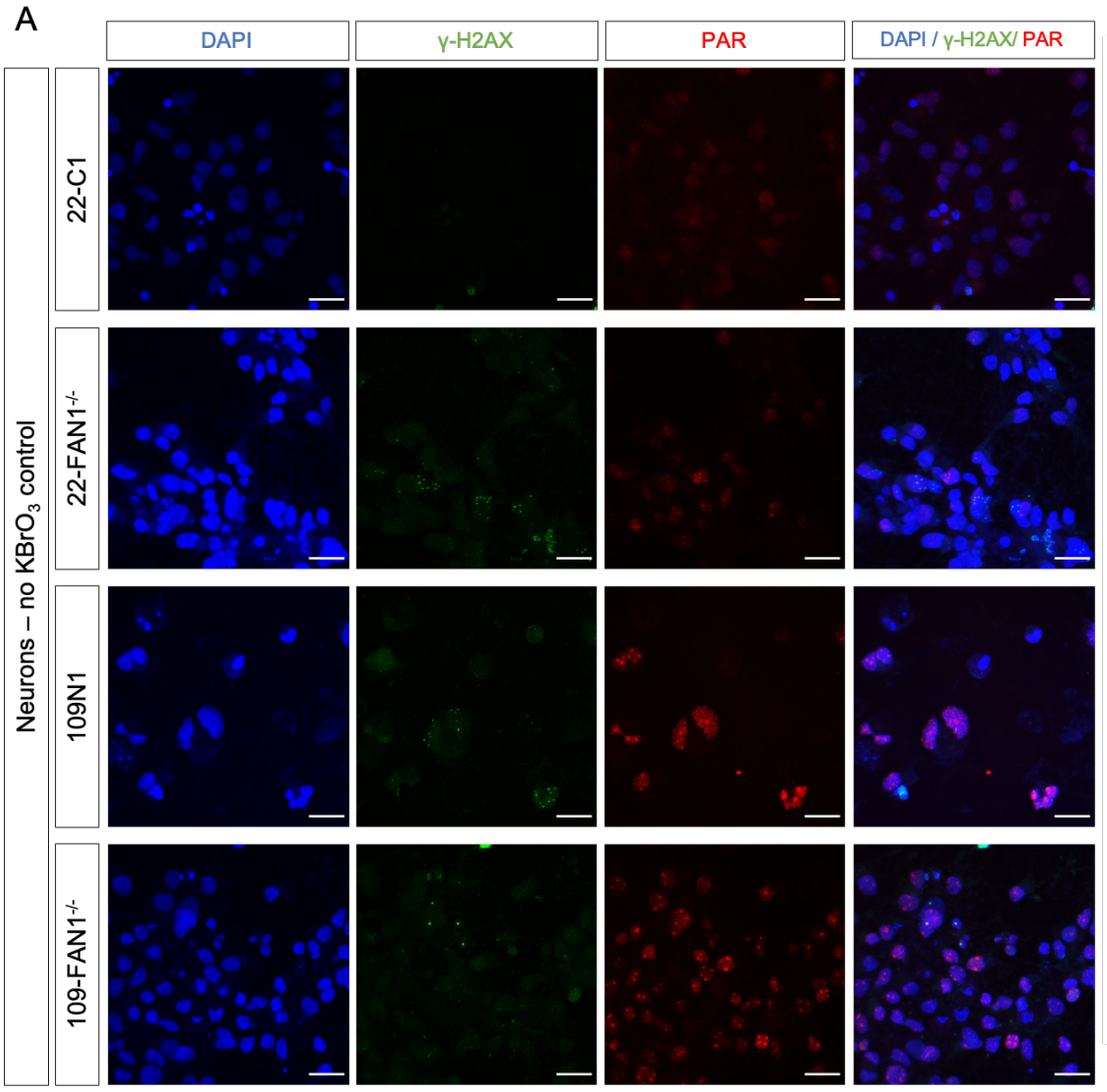
Figure 5.9 PAR and γ-H2AX staining in neurons following exposure to MMC. Staining with (A) PAR, (B) γ-H2A.X and (C) both PAR and γ-H2A.X in 22-C2, 22-FAN1$^{-/-}$-C1, 109N1 and 109-FAN1$^{-/-}$-C2 following exposure with 25 ng/mL MMC. Scale bars represent 20 µm. (D) Relative PAR intensities in 22-C2, 22-FAN1$^{-/-}$-C1, 109N1 and 109N1 at baseline and following exposure to 25 ng/mL MMC. Intensities were determined by Volocity software and are relative to the intensities of 22-C2 at baseline. **p<0.01, determined by a mixed-model ANOVA followed by Tukey's multiple comparisons test (p=0.0098). (E) Quantitation of the average number of γ-H2A.X foci per cell. Foci counts were determined using Volocity software. n=1, 3 technical replicates of each. Values are expressed as mean ± SEM.

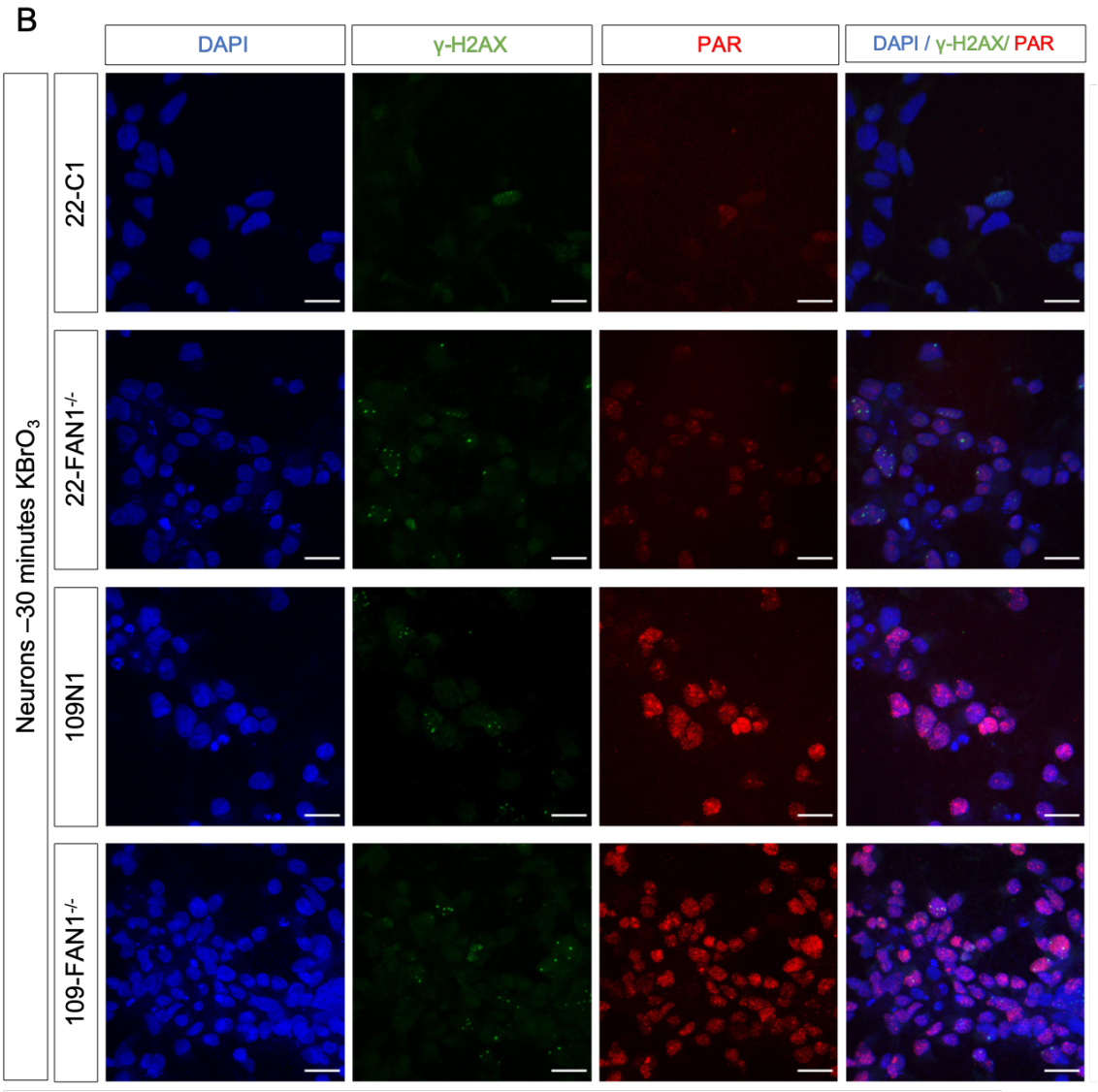
5.3.3.4 Neuronal exposure to $KBrO_3$

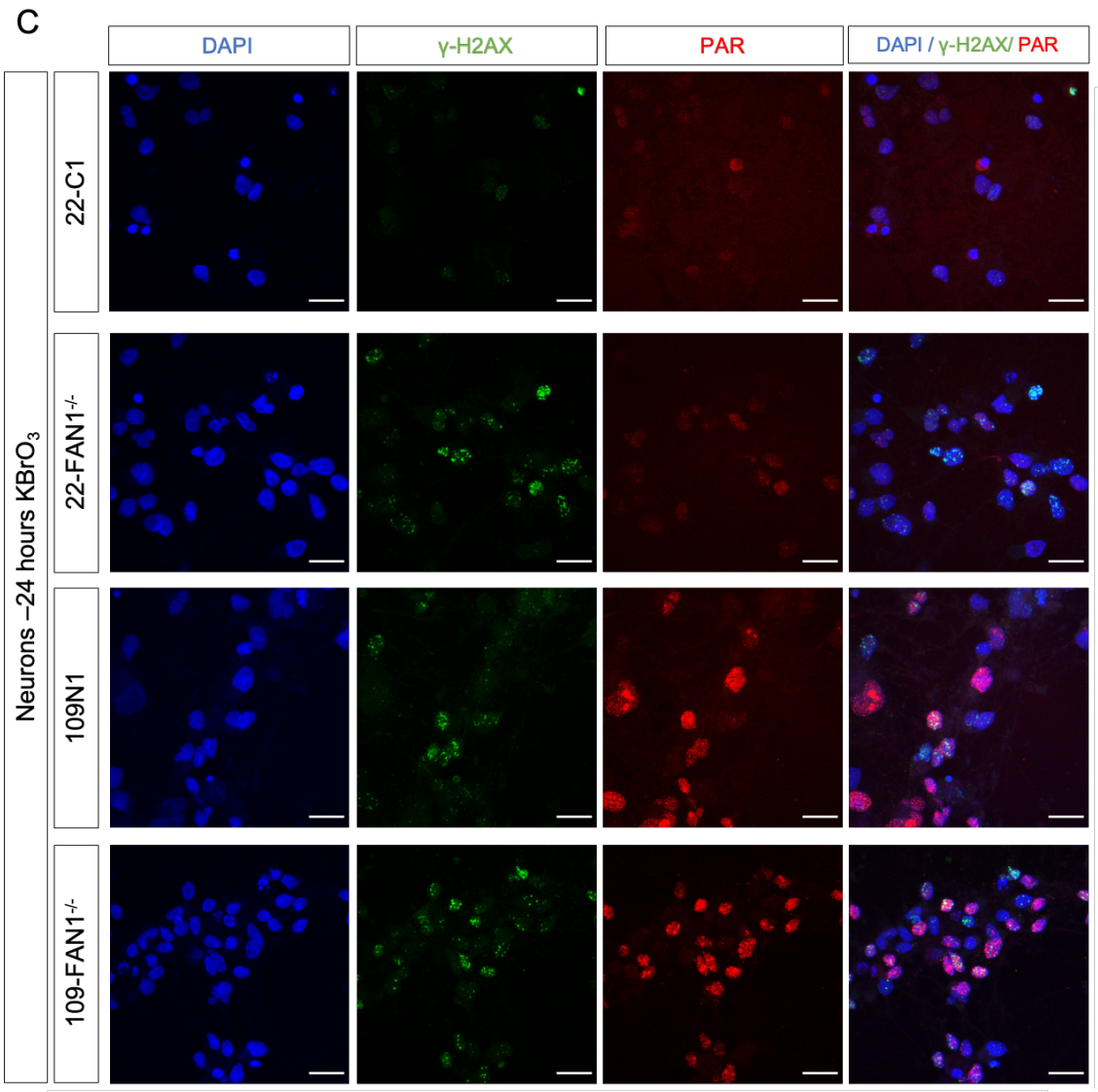
$KBrO_3$ is a potent oxidising agent, whose actions predominantly result in the generation of 7,8-dihydro-8-oxo-guanine (8-oxoG) (Ballmaier and Epe 2006). Removal of this lesion prevents the formation of 8-oxoG: A mispairs, which are poorly repaired (Entezam *et al.* 2010). To examine the response to each of the cell lines to oxidative damage, neurons were exposed to 100 µM $KBrO_3$ for 30 minutes, and fixed and stained with PAR and γ-H2A.X immediately after exposure or following a 24 hour recovery period. These time points allowed us to examine the initial response to $KBrO_3$ and to investigate deficits in recovery following oxidative stress. At baseline, PAR levels were higher in neuronal cells derived from iPSCs with an expanded CAG repeat length compared to those derived from iPSCs with a wild-type repeat length, irrespective of FAN1 status (Figure 5.10A). Whilst this largely supports our previous findings in iPSCs and neurons, it does contradict those of 22-FAN1$^{-/-}$, whose levels of PAR staining at baseline appears variable (Figure 5.9A and Figure 5.10A). Upon treatment with $KBrO_3$ the response was rapid, and PAR staining was elevated at 30 minutes (Figure 5.10B). Surprisingly, there was very little response of 22-C1 to $KBrO_3$ in the initial 30 minutes. PAR staining moderately increased in 22-FAN1$^{-/-}$ and 109N1 after 30 minutes treatment with $KBrO_3$ although findings were not significant. There was, however, a significant increase in PAR in 109-FAN1$^{-/-}$ following treatment with $KBrO_3$. Both 109N1 and 109-FAN1$^{-/-}$ PAR levels were significantly higher than 22-C1, and 109-FAN1$^{-/-}$ was significantly higher than 22-FAN1$^{-/-}$ (Figure 5.10D). At 24 hours, PAR staining was significantly reduced in both 109N1 and 109-FAN1$^{-/-}$ suggesting that whilst cells with an expanded repeat are susceptible to initial DNA damage, they are able to repair the oxidative damage (Figure 5.10C). However, although PAR levels decreased in both 109N1 and 109-FAN1$^{-/-}$, they were still elevated compared to baseline, suggesting incomplete recovery. Continuing this experiment for a longer duration would establish whether PAR levels in the neurons stressed with $KBrO_3$ would return to baseline. In this context, *HTT* CAG repeat length appeared to be the primary factor governing PAR expression in response to DNA damage.

γ-H2A.X staining at baseline demonstrated no significant differences between any of the cell lines (Figure 5.10E). Following exposure to $KBrO_3$ for 30 minutes, there was a marginal increase in the average number of γ -H2A.X foci per cell in 22-C1, 22-FAN1$^{-/-}$ and 109N1 neurons. However, the average number of γ-H2A.X foci had significantly increased after 30 minutes exposure to KBr03 in the 109-FAN1$^{-/-}$-C1, suggesting immediate deficits in DNA repair in this cell line, mediated in part by the expanded *HTT* CAG repeat. Following a recovery window of 24 hours, each line demonstrated an increase in γ-H2A.X foci. There was a significant difference in the average number of γ-H2A.X foci between 22-FAN1$^{-/-}$ and 22-C1. 22-FAN1$^{-/-}$ also showed significantly more γ-H2A.X foci than 109-FAN1$^{-/-}$ levels, suggesting that after longer incubation with $KBrO_3$, the loss of FAN1 had a greater influence on double strand breaks than *HTT* CAG repeat length.

These data demonstrate that in response to oxidative damage, single strand breaks precede double strand breaks, determined via staining with PAR and γ-H2A.X, respectively. Cells with long *HTT* CAG repeats have higher levels of PAR staining, endogenously and following oxidative stress with $KBrO_3$. Neurons lacking FAN1 demonstrate increased PAR staining compared to $FAN1^{+/+}$ neurons with the same *HTT* CAG repeat lengths following exposure to oxidative stress, although these differences are small. FAN1 status appears important in mediating levels of double strand breaks, as $FAN1^{-/-}$ neurons demonstrate higher levels of γ-H2A.X foci irrespective of *HTT* CAG repeat length.




B
DAPI
γ-H2AX
PAR
DAPI / γ-H2AX/ PAR
Neurons –30 minutes $KBrO_3$
22-C1
22-FAN1-/-
109N1
109-FAN1-/-


C
DAPI
γ-H2AX
PAR
DAPI / γ-H2AX/ PAR
Neurons –24 hours KBrO3
22-C1
22-FAN1-/-
109N1
109-FAN1-/-

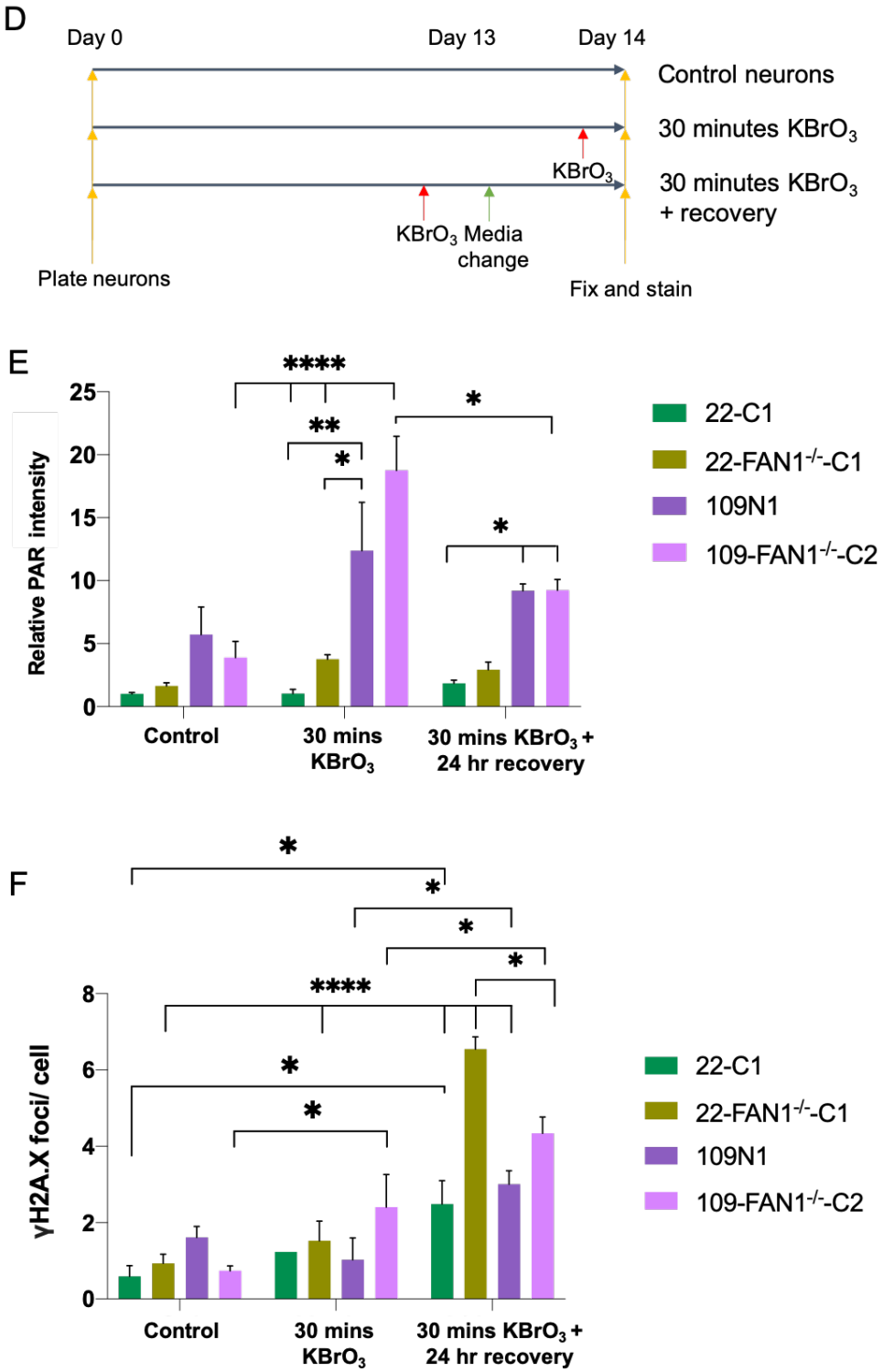


Figure 5.10 PAR and γ-H2AX staining in neurons following exposure to potassium bromate ($KBrO_3$). Staining with PAR and γ-H2AX at (A) baseline- no $KBrO_3$ exposure, (B) 30 minutes 100 µM $KBrO_3$ exposure or (C) 30 minutes 100 µM $KBrO_3$ exposure followed by 24 hours recovery period in fresh media. Scale bars = 20 µm. (D) Schematic of experimental outline. (E) Quantitation of PAR staining in neurons exposed to $KBrO_3$ for 30 minutes or 30 minutes followed by a 24 hour recovery period. Intensities were determined by Volocity software and are relative to the intensities of 22-C1 at baseline. (F) Quantitation of the average number of γ-H2A.X foci per cell in each condition. Foci counts were determined using Volocity software. *p<0.05, **p<0.01, ***p<0.001, ****p<0.0001, determined by One-Way ANOVA followed by Tukey HSD test. n=1, 3 technical replicates of each. Values are expressed as mean ± SEM.

5.3.4 Mitochondrial function in FAN1$^{-/-}$ NPCs

Mitochondrial deficits have been previously reported in post-mortem brain tissue from HD patients, mouse models of HD and HD cell systems (Brennan *et al.* 1985; Sawa *et al.* 1999; Oliveira 2010). We demonstrated previously that corrected HD-iPSCs exhibit significantly higher basal respiration rates, higher maximal respiration rates and higher ATP production (section 3.14). There have also been links to mitochondrial dysfunction and DNA damage (1.7.2.4). We therefore sought to establish whether the loss of FAN1 has an effect on mitochondrial function, and whether the effects differ between an NPC line with an expanded *HTT* CAG repeat or shorter *HTT* CAG repeat length.

For these experiments, NPCs derived from 22 and 109-FAN1$^{+/+}$ and FAN1$^{-/-}$ iPSCs were used. As we only had two clonal lines of 109-FAN1$^{-/-}$, additional controls were used for comparison; the parental line, 109N1, and a number of subclones derived from Q109N1 and Q109N4. NPCs were thawed and recovered in a T25 tissue culture flask in LIA medium (see section 2.1). Following recovery NPCs were plated into Seahorse XF96 microculture plates at a density of 80,000 cells per well. Oxygen consumption rates between experiments and cell lines were variable, and clustering by genotype did not seem representative of the population, so data points for each clone are presented separately.

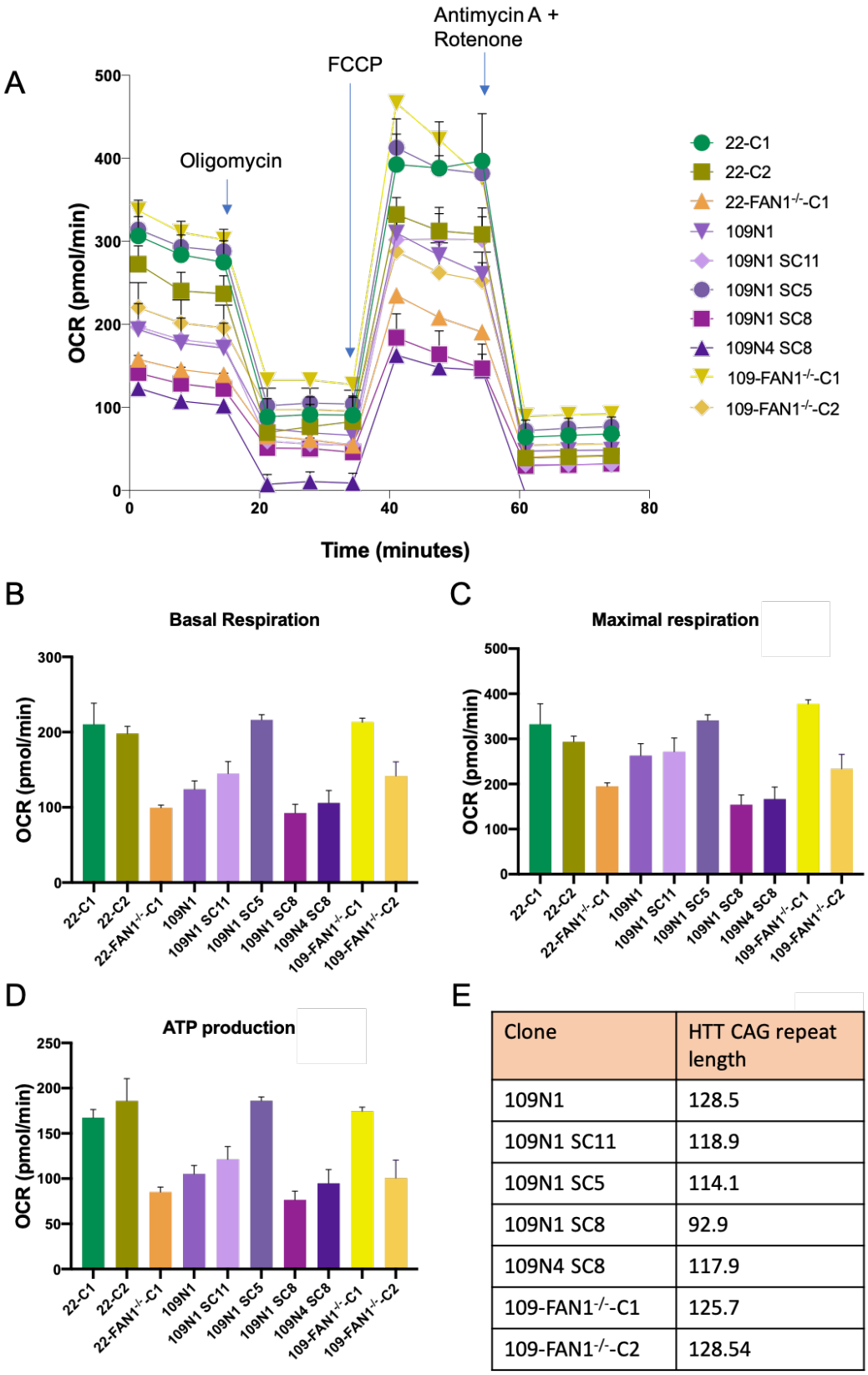
Data from cells in the first experiment were consistent with findings we reported in Chapter 3. Oxygen consumption rates were comparable for both 22-C1 and 22-C2, and basal respiration rates were significantly higher in those lines compared with three out of the five clones with an expanded repeat; 109N1, 109N1 SC8 and 109N4 SC8 (Figure 5.11B and Table 5.1). Maximal respiration was significantly higher in corrected iPSCs when compared with 109N1 SC8 and 109N4 SC8 (Figure 5.11C). ATP production was also significantly higher in the corrected iPSCs when compared with four out of the five clonal lines with an expanded repeat (Figure 5.11D, Table 5.1), suggesting that whilst there was little difference in maximal respiration in two of the 109 subclones, they demonstrated an ATP production deficit. Conversely one subclone with an expanded CAG repeat demonstrated rates of basal respiration (Figure 5.11B), maximal respiration (Figure 5.11C) and ATP production (Figure 5.11D) that were similar to that of corrected iPSCs, possible reflecting an adaptive advantage of this subclone.

In the corrected iPSCs, the knock-out of FAN1 significantly reduced basal respiration (Figure 5.11B), maximal respiration (Figure 5.11C) and ATP production (Figure 5.11D), indicating a link, independent of the *HTT* CAG repeat tract, between FAN1 and mitochondrial respiration. Although the data are underpowered and need repeating, they provide the first hints of an interaction between FAN1 deficiency and mitochondrial functioning. Conversely, knock-out of FAN1 from cells harbouring an expanded *HTT* CAG repeat had the opposite effect. 109-FAN1$^{-/-}$-C1 and 109-FAN1$^{-/-}$-C2 demonstrate

similar basal respiration rates to the parental 109N1 line. Maximal respiration rates and ATP production were significantly higher 109-FAN1$^{-/-}$-C1 NPCs compared with 109N1 NPCs.

It is important to consider the *HTT* CAG repeat length when analysing such data, as we and others have found a significant effect of CAG repeat length on phenotype (An *et al.* 2012; Xu *et al.* 2017; Ooi *et al.* 2019). A recent paper demonstrated deficits in basal respiration rates, ATP production and maximal respiration which worsened in a length-dependent manner in isogenic ESCs with *HTT* CAG repeat lengths of 30, 45, 65 and 81 (Ooi *et al.* 2019). As CAG length variability exists in our lines, we assessed whether there were associations between oxygen consumption rates and CAG repeat length measured in the NPCs (Figure 5.11E). No apparent difference in oxygen consumption rates based on the length of the expanded *HTT* CAG repeat was observed. For example, the measured *HTT* CAG repeat length of 109N1 SC8 was 92 *HTT* CAGs but this line demonstrated the lowest oxygen consumption rates and associated measures when compared with other NPC lines. Our data therefore suggest that metabolic deficits result from the presence of mutant HTT, rather than worsen in a length-dependent manner, though it could be that a length-dependent effect occurs up to a certain threshold. The *HTT* CAG repeats reported here are long, and perhaps all exceed a threshold, above which repeat length might bare little effect.

In the second experiment, basal respiration (Figure 5.12B), maximal respiration (Figure 5.12C), and ATP production (Figure 5.12D), were significantly higher in 22-C1 than each of the other clones, including 22-C2. 22-C1 demonstrated a higher basal respiration rate than literature might suggest is usual for these cell lines, although reports vary (An *et al.* 2012; Xu *et al.* 2017; Ooi *et al.* 2019). Interestingly, basal respiration (Figure 5.12B), maximal respiration (Figure 5.12C), and ATP production (Figure 5.12D) are comparable in the remaining four lines, irrespective of genotype. The loss of FAN1 from 22-C2 or 109N1 had no significant effect on metabolic rates. These findings contradict those from the previous experiment, where loss of FAN1 from 109N1 increased basal respiration rate and maximal respiration, and loss of FAN1 from 22-C2 reduced metabolic rates. These findings demonstrate variability between experiments and highlight the sensitivity of this assay.



E

Clone	HTT CAG repeat length
109N1	128.5
109N1 SC11	118.9
109N1 SC5	114.1
109N1 SC8	92.9
109N4 SC8	117.9
109-FAN1$^{-/-}$-C1	125.7
109-FAN1$^{-/-}$-C2	128.54

Figure 5.11 Mitochondrial energetics in FAN1$^{+/+}$ and FAN1$^{-/-}$ corrected and HD-NPCs (A) Representative Seahorse Bioanalyser traces showing oxygen consumption rates (OCR) in NPCs of multiple genotypes. Cells were plated at 80,000 cells/ well. (B) Quantitative data showing basal respiration (C) Maximal respiration (D) ATP production (E) *HTT* CAG repeat lengths of each of the clones. Significant results can be seen in table below. Values expressed as Means ± SEM with 8 technical replicates of each. OCR measurements normalised to absorbance at 560 nm.

Multiple Comparisons	Basal respiration	Maximal respiration	ATP production	Spare respiratory capacity
22-C1 vs. 22-FAN1$^{-/-}$-C1	p<0.0001	p=0.0087	p=0.0016	
22-C1 vs. 109N1	p=0.0029		p=0.0443	
22-C1 vs. 109N1 SC8	p<0.0001	p=0.0002	p=0.0003	
22-C1 vs. 109N4 SC8	p=0.0001	p=0.0006	p=0.0090	
22-C1 vs. 109-FAN1$^{-/-}$-C1				p=0.0037
22-C1 vs. 109-FAN1$^{-/-}$-C2	p=0.0398		p=0.0223	
22-C2 vs. 22-FAN1$^{-/-}$-C1	0.0004		p<0.0001	
22-C2 vs. 109N1	0.0198		p=0.0020	
22-C2 vs. 109N1 SC8	p=0.0001	p=0.0076	p<0.0001	p=0.0183
22-C2 vs. 109N1 SC11			p=0.0313	
22-C2 vs. 109N4 SC8	p=0.0011	p=0.0220	p=0.0003	p=0.0153
22-C2 vs. 109-FAN1$^{-/-}$-C2			p=0.0009	
22-FAN1$^{-/-}$-C1 vs. 109N1 SC5	p<0.0001	p=0.0041	p<0.0001	
22-FAN1$^{-/-}$-C1 vs. 109-FAN1$^{-/-}$-C1	p<0.0001	p=0.0001	p=0.0004	p=0.0038
109N1 vs. 109N1 SC5	p=0.0011		p=0.0020	
109N1 vs. 109N1 SC8				p=0.0007
109N1 vs. 109N4 SC8				p=0.0006
109N1 vs. 109-FAN1$^{-/-}$-C1	p=0.0019		p=0.0146	
109N1 SC11 vs. 109N1 SC5	p=0.0280		p=0.0303	
109N1 SC11 vs. 109N1 SC8		p=0.0465		p=0.0078
109N1 SC11 vs. 109N4 SC8				p=0.0064
109N1 SC11 vs. 109-FAN1$^{-/-}$-C1	p=0.0436			
109N1 SC5 vs. 109N1 SC8	p<0.0001	p<0.0001	p<0.0001	p=0.0117
109N1 SC5 vs. 109N4 SC8	p<0.0001	p<0.0003	p=0.0003	p=0.0097
109N1 SC5 vs. 109-FAN1$^{-/-}$-C2	p=0.0178		p=0.0008	
109N1 SC8 vs. 109-FAN1$^{-/-}$-C1	p<0.0001	p<0.0001	<p0.0001	p<0.0001
109N4 SC8 vs. 109-FAN1$^{-/-}$-C1	p<0.0001		p=0.0026	p<0.0001
109-FAN1$^{-/-}$-C1 vs. 109-FAN1$^{-/-}$-C2	p=0.0282	p=0.0050	p=0.0069	p=0.0018

Table 5.1 Table of significance for the multiple comparisons performed on the Seahorse data presented in Figure 5.11, determined by One-Way ANOVA followed by Tukey HSD test.

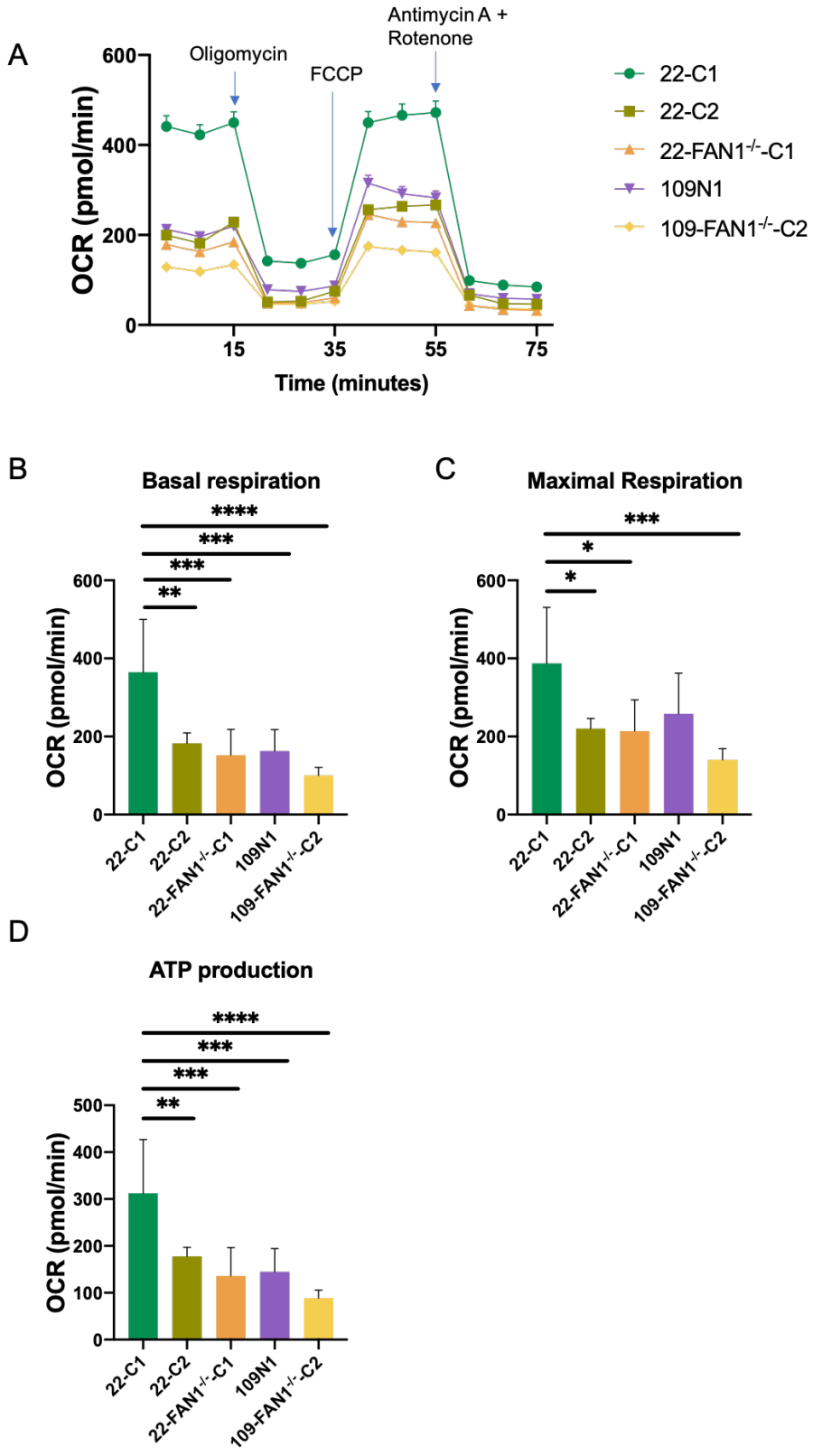


Figure 5.12 Mitochondrial energetics in FAN1$^{+/+}$ and FAN1$^{-/-}$ corrected and HD-NPCs. (A) Representative Seahorse Bioanalyser traces showing oxygen consumption rates (OCR) in NPCs of multiple genotypes. Cells were plated at 80,000 cells/ well. (B) Quantitative data showing basal respiration (C) maximal respiration (D) ATP production. *p<0.05, **p<0.01, ***p<0.001, ****p<0.0001, determined by One-Way ANOVA followed by Tukey HSD test. Values expressed as Mean ± SEM with 14 technical replicates of each. OCR measurements normalised to absorbance at 560 nm.

5.4 Discussion

HD pathogenesis is complex and a number of molecular interactions and signalling pathways have been implicated, mitochondrial dysfunction, DNA damage accumulation and DNA repair deficits (Damiano *et al.* 2010; Usdin *et al.* 2015; GeM-HD Consortium 2019). How these networks lead to neuronal dysfunction and degeneration is still unknown. This chapter aimed to explore these networks in the context of the length of the *HTT* CAG repeat and investigate the role that FAN1, as the most significant genetic modifier of age at onset of HD, might play in contributing to disease phenotypes.

As described in chapter 3, the HD iPSCs demonstrate robust and reproducible *HTT* CAG repeat expansions. Knock-out of FAN1 in iPSCs with an expanded CAG repeat (> 100 *HTT* CAG) increased the frequency and rate of expansion in dividing iPSCs and post-mitotic neurons. A recent transcriptome wide association study demonstrated that increased *FAN1* expression is associated with a later age at onset and a slower disease progression (Goold *et al.* 2019), implying that FAN1 protects against CAG repeat expansion and may modulate HD pathogenesis via this mechanism. Somatic expansion in HD post-mortem brain tissue correlates with disease onset (Swami *et al.* 2009). Additionally, suppression of somatic expansion in mouse models alleviates motor symptoms (Pinto *et al.* 2013), delays onset and slows progression of the disease (Budworth *et al.* 2015), highlighting the importance of identifying factors involved in preventing or promoting *HTT* CAG repeat expansions.

Fragile X related disorders are another class of trinucleotide repeat disorders caused by CGG-repeat expansions in the *FMR1* gene. The loss of *Fan1* from a mouse model of Fragile X also potentiated CGG-repeat expansions in the *Fmr1* gene (X.-N. N. Zhao and Usdin 2018), suggesting that the protective effects of FAN1 extend beyond HD, and likely have a wider clinical impact in a number of trinucleotide repeat disorders. The accumulating evidence from human genetic studies (GeM-HD Consortium 2015; Bettencourt *et al.* 2016; Moss *et al.* 2017; GeM-HD Consortium 2019) and the recent functional findings in both HD (Goold *et al.* 2019) and Fragile X disorder (X.-N. N. Zhao and Usdin 2018) support the notion that a common pathogenic mechanism acting at the level of the DNA might underlie disease pathogenesis in a number of trinucleotide repeat disorders.

Multiple pathways have been implicated in trinucleotide repeat expansion. Manipulation of proteins in the MMR pathways show that Msh2, Msh3, Mlh1 and Pms2 are required for *HTT* CAG repeat expansions and knock-out of *Mlh1, Msh2* and *Msh3* alleviates disease symptoms (Wheeler *et al.* 2003; Pinto *et al.* 2013). Additionally, knock-out of members of the base excision repair machinery, Ogg1 (Kovtun *et al.* 2007), Neil1 (Møllersen *et al.* 2012) and Polβ (Lokanga *et al.* 2015) suppresses somatic expansions. Loss of nucleotide excision repair protein, XPA (Hubert *et al.* 2011), but not XPC (Dragileva *et al.* 2009), also reduces trinucleotide expansions. Exo1, a 5'-3' exonuclease nuclease is also protective

against repeat expansions in a Fragile X mouse model (X.-N. N. Zhao and Usdin 2018). EXO1 is an effector in multiple DDR pathways and may be recruited to mispaired bases by MutLα (MLH1-PMS2) or MutLγ (MLH1-MLH3), where it is involved in the removal of the nicked strand to allow for resynthesis (Desai and Gerson 2014). To maintain genomic integrity, there is often functional redundancy within DNA repair pathways and it has been demonstrated that FAN1 nuclease activity in mismatch repair is able to compensate for the loss of EXO1 nuclease activity (Desai and Gerson 2014). Given that FAN1 also interacts with MLH1 and PMS2 (MacKay *et al.* 2010), FAN1 and EXO1 might act via a similar mechanism to faithfully repair mismatched DNA and prevent repeat expansions (X.-N. N. Zhao and Usdin 2018). It has been suggested that FAN1 might sequester MLH1 and/or PMS2, thereby preventing the formation of the MMR complex which is involved in *HTT* CAG repeat expansions (Goold *et al.* 2019). This could also explain why repeat expansions do still occur when cells or animals have functional FAN1.

The DNA damage response is increasingly recognised as a key modulator of disease pathogenesis in HD and other triplet repeat disorders (GeM-HD Consortium 2015; Bettencourt *et al.* 2016; GeM-HD Consortium 2019). It is not known however, if the expanded *HTT* CAG repeat tract acts as a substrate for DNA maintenance processes, or whether HTT itself plays a role in in the DNA damage response, which is then impaired by the presence of an expanded polyglutamine tract. Elevated levels of DNA damage markers have been demonstrated in models of HD (Giuliano *et al.* 2003; Gao *et al.* 2015; Gao *et al.* 2019; Ooi *et al.* 2019). At baseline we observed a trend towards a significance in levels of DNA damage markers PAR and γ-H2A.X, though it is critical to repeat these experiments to establish an effect. These findings could suggest either increased endogenous levels of DNA damage due to the accumulation of ROS, or deficits in DNA repair and subsequent accrual of DNA damage. HD fibroblasts with an expanded polyglutamine tract showed deficits in double strand break repair following irradiation and reduced rates of ataxia telangiectasia mutated kinase (ATM) shuttling into the nucleus (Ferlazzo *et al.* 2014). Additionally, HTT actively participates in oxidative DNA damage repair; it has been found to localise at BER sites where it interacts and scaffolds a number of BER proteins including XRCC1, FEN1, APE1 and HMGB1 (Maiuri *et al.* 2017). mHTT is still recruited to chromatin and therefore able to localise at sites of DNA damage, though remains bound to chromatin for longer than HTT, suggesting mHTT might slow or impair the repair processes. These findings suggest that HTT plays a critical role in DNA repair, which is then impaired upon polyglutamine expansion. Consistent with this, we find neurons with an expanded *HTT* CAG repeat demonstrate increased levels of DNA damage, as marked by PAR staining, compared to those with a wild-type *HTT* repeat, following 30 minutes exposure to the potent oxidising agent, $KBrO_3$. After a 24 hour recovery period, a significant reduction in single strand breaks was observed in Q109 lines, although levels did not return to baseline, suggesting incomplete or delayed repair of single strand breaks. However, there was an increase in the levels of double strand

breaks in each line after 24 hours, which have been demonstrated to occur following oxidative damage and may persist over 24 hours (Fortini *et al.* 2012). To examine whether this DNA damage would be sustained, repaired, or might induce neuronal death, future experiments should examine DNA damage markers and apoptosis at least 72 hours after $KBrO_3$ exposure. Interestingly, cells lacking FAN1 demonstrated elevated levels of DNA damage following 30 minutes exposure to $KBrO_3$. This increase in DNA damage was observed in neurons with both an expanded CAG repeat and a wild-type CAG repeat, suggesting that cells lacking FAN1 are more susceptible to DNA damage following oxidative stress.

Mitochondrial deficits in HD-iPSCs are well documented (An *et al.* 2012; Xu *et al.* 2017; Ooi *et al.* 2019) and in chapter 3 we demonstrated deficits in mitochondrial functioning in HD-iPSCs with an expanded repeat tract. These deficits were alleviated upon genetic correction of the disease-causing mutation and increases in basal respiration, maximal respiration and ATP production were observed. Data from this chapter largely support these findings. In most cases, oxygen consumption rates were increased in NPCs with a wild-type *HTT* CAG repeat tract compared with those with an expanded *HTT* CAG repeat, further supporting the notion that mHTT leads to deficits in mitochondrial functioning. Interestingly we found no correlation between metabolic rates and *HTT* CAG repeat length. For example, clones with the largest repeat lengths (125/128 *HTT* CAG) had higher metabolic rates than clones with shorter repeat lengths (115/92 *HTT* CAG), although rates were still comparable. These data are contradictory to those reported recently in isogenic hESCs, where they found increasing metabolic deficits as the CAG repeat length increased from 30 to 81 *HTT* CAG. One explanation for this might be that the lengths of the repeat tracts in each of our iPSC lines greatly exceeds those used in Ooi *et al.* (2019). CAG repeat lengths > 80 are commonly associated with childhood onset of HD, and it might be that CAG repeat lengths over a certain threshold have the same deleterious effect on mitochondrial functioning. Conversely, these findings might reflect adaptation of certain cells in culture.

To examine the interplay between DNA damage and mitochondrial functioning further, we assessed the effect of a loss of FAN1 on mitochondrial parameters. In one experiment, the loss of FAN1 from a cell line with 22 *HTT* CAGs significantly reduced oxygen consumption rates, yet conversely the loss of FAN1 from a cell line with an expanded *HTT* CAG repeat increased oxygen consumption rates. In a second experiment, there was no effect of genotype on oxygen consumption rates, leaving results inconclusive. These findings highlight a clear need for repetition of experiments, particularly when the experiments are sensitive to subtle changes.

The effect of FAN1 on *HTT* CAG expansions was most pronounced in the first 40 days, perhaps why this was chosen as the primary assay period in the recent paper by Goold *et al.* (2019). In this short window, a marked effect of FAN1 knock-out was observed compared to wild-type clones and therefore it represents an optimal period in which potential genetic modifiers of HD may be assayed. Through

extended time in culture however, there appear to be multiple factors that influence *HTT* CAG instability. One clone, 109-FAN1$^{-/-}$ demonstrated a reproducible cell death phenotype which corresponded to a reduction in *HTT* CAG repeat length. This appeared to be an adaptive tendency of the cells, as those that survived grew well in culture and exhibited *HTT* CAG repeat expansions at a much slower frequency than prior to the observed cell death. The issues surrounding this are complex, it may in part be that the relatively rapid expansion of the expanded repeat drives cell death or the repeat expansions may be a secondary consequence to the stress a particular subset of cells encountered. It is likely that a population of cells accrued variants in genes that encouraged their survival through the loss of FAN1. There is functional redundancy in many DNA repair pathways, to ensure that DNA damage can be repaired efficiently. Transcriptional changes in proteins involved in these pathways might therefore increase cell survival and reduce *HTT* CAG expansions. Interestingly, this was the only cell line that did not undergo a duplication of chromosome 1 over 40 passages. These findings might suggest that the cells that survived had an adaptive advantage in culture that was perhaps more favourable than that of a duplication of chromosome 1. No discernible cell death phenotype was observed in neurons based on FAN1 status or the length of the *HTT* CAG repeat length.

Another challenge faced after prolonged iPSC culture was spontaneous differentiation. This spontaneous differentiation often coincided with multiple clonality of the *HTT* CAG repeats, i.e. large contractions or expansions in a subset of cells. The fact that the bimodal peak distribution occurs in a population of cells that are still dividing suggests that a single rare event, a contraction or expansion, could be responsible for the emergence of distinct peaks through multiple rounds of division. However, in germline and somatic tissues, multiple CAG repeat lengths are largely reflective of a heterogenous cell population; where a subset of cells in an organ are prone to instability, while others are not. Large repeat expansions are largely confined to the spermatogonia or primary spermatocytes in the testes (X. N. Zhao and Usdin 2018), the hepatocytes of the liver (Lee *et al.* 2011; Gazy *et al.* 2019), and the striatal neurons, and to a lesser extent cortical neurons in the brain (Swami *et al.* 2009; Lokanga *et al.* 2013). The reasons for these findings are not completely understood, although it has been speculated that it might result from differences in the relative expression levels of proteins involved in contraction and expansion events (Zhao *et al.* 2019). It raises the question as to whether the heterogeneity observed in the iPSC population might be driving bimodality of the CAG repeat. To determine whether spontaneously differentiated cells are the source of the multiple *HTT* CAG repeat populations, cells could be sorted based on the expression of a pluripotency cell surface antigen such as SSEA-4 and *HTT* CAG repeat lengths determined from the different populations.

The loss of FAN1 from iPSC-derived neurons also significantly potentiated neuronal *HTT* CAG expansion, suggesting a protective effect of FAN1 in both dividing and post-mitotic cells. Although expansions are

not solely dependent on division, *HTT* CAG repeat expansion in neurons occurred at a significantly lower frequency than those occurring in iPSCs which are still dividing. It is possible there are distinct mechanisms governing expansions in post-mitotic and dividing cells. However, research suggests that replication, DNA repair and recombination all influence trinucleotide repeat instability. During replication, due to the similarity of the trinucleotide repeat, nucleotides can misalign, generating an extra-helical loop which may then be incorporated into the duplex DNA by a DNA polymerase (Schweitzer and Livingston 1997; Schweitzer and Livingston 1999). It's also likely that replication facilitates the formation of secondary structures that act as expansion intermediates as DNA is unwound and exposed.

It is widely accepted that expansion of a trinucleotide repeat is length dependent; trinucleotide repeats below a certain threshold seldom show expansion in mouse models or cell systems (Manley, Pugh, *et al.* 1999). *De novo* mutations of a trinucleotide repeat mainly occur through germline expansion. It might be speculated however, that a wild-type repeat length in genetic background of a patient with HD would have the appropriate environment to promote expansions of the repeat. Furthermore, the effect that knock-out of FAN1 might have on repeat expansion of a wild-type allele was not known. We report no effect of FAN1 knock-out on *HTT* CAG repeat expansion of a wild-type repeat length; over prolonged culture of iPSCs or in post-mitotic neurons, though these time courses are comparatively short compared with a patient's lifetime. These data support the notion that trinucleotide repeats over a certain length form secondary and tertiary structures that act as a substrate for repeat expansion or drive DNA damage themselves.

If our hypothesis is true, and somatic instability drives degeneration, then it might be expected that any rarer large expansions that occur are quickly excluded from the population. These expansion events are too rare and occur at too low a frequency to be detected by bulk PCR methods. Electrophoretic traces generated by bulk-PCR methods greatly underestimate the true variability in CAG repeat lengths that exist in a population, particularly around the expanded allele (Monckton *et al.* 1999). However, small-pool PCR or single-cell sequencing techniques would offer a detailed look at the expansion profile of each cell over time (Swami *et al.* 2009). Genomic DNA was harvested at each time point to allow for small-pool PCR to be performed at a later date. This technique would allow us to follow rare expansions or contractions and understand the foundation of bimodal peak distributions. Advancements in long read sequencing technologies also enables us to sequence individual repeat tracts in a high-throughput manner (Ciosi *et al.* 2018).

Immunocytochemistry was variable, we observed differences between the staining on coverslips of the same condition performed at the same time. To bolster our findings, it would be sensible to perform additional techniques to assess the levels of DNA damage markers in response to DNA damage.

Extracting RNA and protein from samples treated with DNA damaging agents such as MMC or $KBrO_3$, would enable us to look at expression changes via qPCR and western blot. These techniques would allow a complete look at expression changes in the total population of cells treated, rather than those imaged in the fields chosen.

In conclusion, we provide further support for a role of FAN1 in protecting against *HTT* CAG repeat expansions in HD, both in iPSCs and post-mitotic neurons. FAN1 is likely involved with the complex DNA repair network involved in the governance of repeat instability. Q109 neurons demonstrated increased levels of DNA damage following exposure to oxidising agent $KBrO_3$, indicative of defective or delayed repair. The loss of FAN1 renders cells more susceptible to cross-linking agents such as mitomycin C, and oxidising agents such as $KBrO_3$. However, preliminary data suggests that there is not a differential effect of the loss of FAN1 on cells with a wild-type repeat compared with cells with an expanded repeat.

5.5 Chapter Summary

- Knock-out of FAN1 potentiates further *HTT* CAG repeat expansions in iPSCs and post-mitotic neurons with an expanded *HTT* CAG repeat
- Cells with FAN1 knock-out are more susceptible to DNA damage following cross-linking agent MMC
- Knock-out of FAN1 and/or the presence of an expanded *HTT* CAG repeat renders neurons more susceptible to DNA damage following exposure to oxidising agent $KBrO_3$.
- No discernible effect of FAN1 knock-out on mitochondrial functioning

Chapter 6: General Discussion

6.1 Summary of Findings

The aims of this thesis were to establish a cellular model of *HTT* CAG repeat expansion which could be used as a surrogate of *in vivo* repeat expansions to assay genetic and drug modifiers. We demonstrate that an iPSC line derived from a patient with 109 *HTT* CAG (Q109) repeats exhibits reproducible repeat expansions in iPSCs and post-mitotic neurons, and thus represents a suitable model to investigate mechanisms governing repeat expansions *in vitro*. We also report the genetic correction using CRISPR-Cas9 and piggyBac technologies, of an iPSC line with an expanded *HTT* CAG repeat (Q109), to a wild-type repeat length of 22 *HTT* CAGs (Q22). Q22 clones retained pluripotency and were successfully differentiated into forebrain progenitors and post-mitotic neurons. Upon correction of the HD-mutation we observe a reversal of a well-established HD-phenotype; Q22 NPCs exhibited increased basal respiration, maximal respiration and ATP production compared with Q109 NPCs. Additionally, Q109 neurons demonstrated increased levels of DNA damage upon exposure to exogenous stressors. These data suggest that the expanded *HTT* CAG repeat, and/or resulting expanded polyglutamine tract confer detrimental effects.

Recent human genetic studies have identified genes and pathways relating to DNA maintenance machinery as modifiers of the age at onset of HD and it is thought they might act through somatic expansion of the *HTT* CAG repeat. FAN1 is a DNA repair protein canonically associated with interstrand cross-link repair (MacKay *et al.* 2010; Jin and Cho 2017) and has been identified as a significant modifier of age at onset in HD and several spinocerebellar ataxias (GeM-HD Consortium 2015; Bettencourt *et al.* 2016; GeM-HD Consortium 2019). Using CRISPR-Cas9 we successfully knocked-out *FAN1* in Q22 and Q109 iPSCs. iPSCs lacking FAN1 were more susceptible to interstrand cross-linking agent MMC and demonstrated increased levels of DNA damage following exposure to $KBrO_3$. The loss of FAN1 from Q109 iPSCs potentiated repeat expansion of the expanded *HTT* CAG repeat in iPSCs and iPSC-derived neurons, suggesting that FAN1 might play a protective role in preventing repeat expansions in HD. These novel model systems provide a platform for investigating the cellular phenotypes associated with expanded CAG repeats, and the effects of DNA-repair associated genetic modifiers.

6.2 Mechanisms underlying Huntington's Disease pathogenesis

6.2.1 Mitochondrial dysfunction in Huntington's Disease

Mitochondrial deficits and altered bioenergetics are a commonly reported phenotype in HD (Seong *et al.* 2005; Damiano *et al.* 2010). Indeed, weight loss is common in HD patients, independent of increased locomotor activity and their calorific intake (van der Burg *et al.* 2009). Patients with HD exhibit respiratory chain complex deficiencies, reduced ATP levels (Lodi *et al.* 2000), and a lower anaerobic respiration threshold following exercise (Ciammola *et al.* 2011). A number of papers have reported the reversal of mitochondrial dysfunction upon correction of the HD mutation in HD-iPSCs/ESCs (An *et al.* 2012; Xu *et al.* 2017; Ooi *et al.* 2019). Mitochondrial functioning was therefore a robust phenotype to validate the isogenic lines generated in this study. Consistent with previous reports, we find that mitochondrial dysfunction in Q109 NPCs was rescued upon genetic correction of the expanded repeat.

HTT has a wide range of cellular functions. Effects of mHTT are widespread and effect multiple pathways. Mitochondrial dysfunction likely arises through multiple pathogenic mechanisms; suppression of genes critical for mitochondrial functioning such as PGC-1α (Cui *et al.* 2006; Che *et al.* 2011); enhanced enzymatic activity of DRP1, increased mitochondrial fission and fragmentation (Song *et al.* 2011; Cherubini *et al.* 2015; Rodinova *et al.* 2019); calcium dyshomeostasis (De Mario *et al.* 2016); interference of trafficking proteins and subsequent reduced mitochondrial mobility (Trushina *et al.* 2004). Indeed, such effects have been reported in humans, mini-pigs and rodents (Arun *et al.* 2015; Carmo *et al.* 2018). A toxic oxidation cycle likely exists, whereby mHTT induces mitochondrial dysfunction which in turn generates increased levels of ROS (Figure 6.1). ROS have damaging effects via a number of mechanisms and are particularly harmful to both mitochondrial and nuclear DNA (Liu *et al.* 2017). Mitochondrial DNA damage causes mitochondrial DNA depletion, which in turn reduces mitochondrial bioenergetics; contributing to a cycle of increased mitochondrial dysfunction and activation of apoptotic pathways (Browne *et al.* 2006; Damiano *et al.* 2010). Bioenergetic failure, and subsequent degeneration are also thought to result from accumulation of nuclear DNA damage (Ying *et al.* 2003; Cardinale *et al.* 2015; Maiuri *et al.* 2017). Increased DNA damage may result in increased *HTT* CAG repeat expansion, thereby increasing the length of the toxic polyglutamine tract and starting the cycle of toxicity again. Indeed, oxidative damage is a driver of repeat expansions in HD and Fragile X disorder (Kovtun *et al.* 2007; Entezam *et al.* 2010).

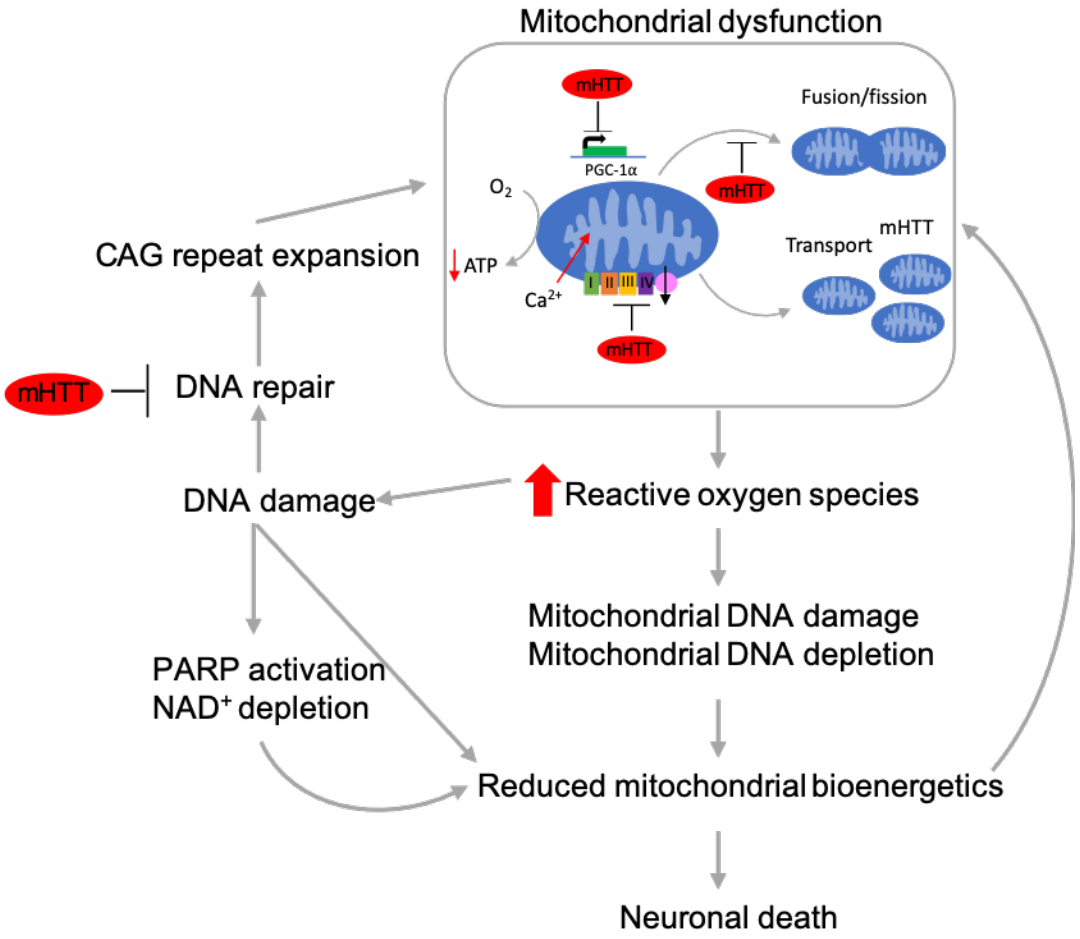


Figure 6.1 Proposed toxic oxidation cycle *in HD.*

6.2.2 DNA damage in Huntington's Disease

DNA repair proteins have long been implicated in HD pathogenesis, yet research has seldom looked at DNA damage in HD in an isogenic context. The interest in DNA damage resulting from the human genetic data, that highlighted DNA handling pathways as important modifiers of the disease, renewed the focus on the role of DNA damage and repair in the context of disease pathogenesis (GeM-HD Consortium 2015; Moss *et al.* 2017; GeM-HD Consortium 2019). Elevated levels of DNA damage are reported in models of HD, both at baseline and upon exposure to oxidative damage (Gao *et al.* 2019; Ooi *et al.* 2019). We observe a trend towards a significant difference in endogenous levels of DNA damage between Q22 and Q109 and, upon exposure to we report a significant increase in DNA damage markers, PAR and γH2A.X, in neurons with an expanded *HTT* CAG repeat. These findings suggest that repair systems in cells with mHTT have defective or delayed DNA repair. Indeed, HTT itself plays a role in in the DNA damage response, which may then be impaired by the presence of an expanded polyglutamine tract (Giuliano *et al.* 2003; Enokido *et al.* 2010; Gao *et al.* 2015; Maiuri *et al.* 2017). A number of proteins containing a polyglutamine tract

including HTT, androgen receptor (AR), ataxin (ATXN)1, ATXN2 and ATXN3, have demonstrated roles in DNA repair (Giuliano *et al.* 2003; Chatterjee *et al.* 2015; Gao *et al.* 2015).

Following oxidative stress, the N-terminus of HTT is phosphorylated and this post-translational modification allows the translocation of HTT into the nucleus (DiGiovanni *et al.* 2016). HTT is then recruited to sites of damage in the nucleus, in an ATM-dependent manner, where it acts as a molecular scaffold to proteins of the DNA damage response pathway (Maiuri *et al.* 2017). mHTT is still recruited to sites of DNA damage, but appears hypo-phosphorylated and therefore might prevent appropriate scaffolding (Atwal *et al.* 2011). Interestingly, the loss of ATM from mammalian cells and Drosophila was protective against mHTT fragments and in a mouse model of HD, heterozygous loss of Atm improved behavioural phenotypes and neuropathology (Lu *et al.* 2014). Additionally, upon DNA damage HTT may be phosphorylated by cyclin-dependent kinase 5 (Cdk5), which functions to mediate p53-dependent neuronal death (Anne *et al.* 2007). Hypo-phosphorylation, or decreased expression of *CDK5*, as has been shown to occur through ageing (Lu *et al.* 2004), could therefore contribute to neuronal death. These data are consistent with a model that supports a role of HTT in DNA damage repair which is impaired upon elongation of the polyglutamine tract.

Increased DNA damage resulting from ROS may also indirectly affect mitochondrial bioenergetics. DNA damage is recognised by PARP proteins, and using NAD+, PARP generates PAR chains to aid in DNA repair (Kameshita et al. 1984). NAD+ plays vital roles in energy metabolism in eukaryotes, where it acts as an oxidising agent to transfer electrons from one molecule to another to fuel oxidative phosphorylation (Xia et al. 2009). NAD+ levels are limited however, and PARP-1 competes with a number of other enzymes that require NAD+ (Cantó et al. 2015); sirtuins 1 (SIRT1) and 3 (SIRT3) (Haigis and Guarente 2006). SIRT3 is mainly located in the mitochondria where it has a role in regulation of the mitochondrial permeability transition pore, prevents cell death and has neuroprotective properties (Fried 2007; Hafner *et al.* 2010; Kim *et al.* 2011). SIRT1 functionally interacts with PGC-1α, which as mentioned, has a key role in mitochondrial biogenesis (Kong *et al.* 2010). Inefficient DNA repair, or sustained DNA damage, results in hyper-PARylation where PARP-1 metabolises NAD+ at such high rates that it impedes metabolic functioning. Additionally, excessive PARP activation has been demonstrated to regulate neuronal cell death in Alzheimer's (Hensley et al. 1996) and Parkinson's Disease (Mandir et al. 2002), and depletion of NAD+ and mitochondrial permeability transition have directly been shown to mediate PARP-1 induced cell death (Ying et al. 2003; Alano et al. 2004). Supporting these findings, PARP-1 inhibitors demonstrated neuroprotective effects in mouse models of HD (Cardinale et al. 2015; Paldino et al.

2017). These findings suggest that mHTT can indirectly elicit mitochondrial dysfunction via increased DNA damage levels.

Thus, while there is no doubt that increased ROS generated from mitochondrial dysfunction can induce DNA damage, it is likely that mHTT itself also impairs DNA repair, resulting in the accumulation of DNA damage later in life.

6.2.3 Somatic expansion and disease

Research looking at HD age at onset and disease progression has identified a correlation between levels of somatic instability in the brain and age at onset, suggesting that somatic instability could drive disease pathogenesis (Swami *et al.* 2009). Consistent with this, evidence from mouse models suggests that suppression of somatic instability delays age at onset, alleviates disease phenotypes and slows disease progression (Wheeler *et al.* 2003; Dragileva *et al.* 2009; Budworth *et al.* 2015). However, much of the research looking at the factors driving repeat instability have not looked at the phenotypes associated with increasing *HTT* CAG repeat lengths, and those that have are often of small effect and require a large number of animals to see a difference (Budworth *et al.* 2015; Budworth and McMurray 2016). With many of these models, it is difficult to definitely establish cause and effect, i.e. whether somatic instability contributes to disease pathogenesis, or whether somatic instability is a secondary consequence. Indeed, a number of mouse models do not demonstrate somatic instability, but still demonstrate a disease phenotype, implying somatic expansion is not a primary determinant of disease (Gray *et al.* 2008). Somatic expansion of the *HTT* CAG repeat likely contributes to disease phenotype through a diverse range of functions. The genetic expansion of the repeat can influence pathology through increased toxicity of RNA or protein. mHTT has not only been implicated in DNA repair and mitochondrial functioning, but also; pro-apoptotic pathways, autophagy, disrupted cell signalling; N-terminal fragment oligomers and subsequent aggregation (Saudou and Humbert 2016). As discussed, the toxic effects of the protein are complex, and enhanced toxicity caused by an expanding protein may contribute to disease via any one of these mechanisms or multiple mechanisms.

6.2.4 Mechanism of repeat expansion

DNA repair mechanisms are central to repeat expansions, and each of the repair pathways in the DNA damage response has been investigated with respect to a role they might play in mediating disease-associated repeat expansions (Kovtun and McMurray 2008; Jarem *et al.* 2009; McMurray 2010; McGinty and Mirkin 2018). DNA repair proteins in both base excision repair and nucleotide excision repair pathways have been implicated in repeat expansions, yet the strongest driver of repeat expansions is the mismatch repair pathway, with strong evidence in favour of its

requirement for repeat expansions to occur (McMurray 2008; Iyer *et al.* 2015; Schmidt and Pearson 2016). Despite these findings, the mechanism appears complex as cross-talk between pathways clearly occurs. Recently, another DNA repair protein was implicated in the puzzle. FAN1 was identified as a modifier of age at onset of HD and a number of SCAs, and recent functional findings support the involvement of FAN1 as a disease modifier through modulation of the expanded repeat tract (GeM-HD Consortium 2015; Bettencourt *et al.* 2016; GeM-HD Consortium 2019). The loss of FAN1 in a mouse model of Fragile X disorder potentiated CGG repeat expansions, and similarly, the knock-down of FAN1 increased CAG repeat expansions in an iPSC model of HD (X.-N. N. Zhao and Usdin 2018; Goold *et al.* 2019). Consistent with these findings we report increased *HTT* CAG repeat expansions in iPSCs and post-mitotic neurons following the loss of FAN1. It is likely that FAN1 plays an important role in the complex network of DNA maintenance machinery governing repeat expansions.

Repetitive DNA sequences show a propensity to adopt unusual secondary structures such as slipped strands, hairpin loops and R-loops (Mirkin and Frank-Kamenetskii 1994; Mirkin 2007; McGinty and Mirkin 2018). These may form transiently when DNA is unpaired during cellular processes such as replication, repair and transcription. It is thought that these structures may be aberrantly processed by DNA repair pathways in an attempt to resolve the non-helical structures. MutSβ (MSH2-MSH3) and subsequent recruitment of MutLα (MLH1-PMS2) or MutLγ (MLH1-MLH3) are required for repeat expansions to occur (Manley, Shirley, *et al.* 1999; Dragileva *et al.* 2009; Pinto *et al.* 2013). The MutL complexes may cleave either strand, and it is thought that if the strand opposite the hairpin is nicked, gap filling occurs and the looped bases are incorporate into the repaired strands, causing expansion.

There are a number of mechanisms by which FAN1 might confer protection against repeat expansions (Figure 6.2). FAN1 interacts with both MLH1 and PMS2 (Cannavo *et al.* 2007) and although the function of this interaction is unknown, it is plausible that their interaction mediates repeat instability. FAN1 might sequester MutL complexes thereby preventing or reducing their recruitment to MutSβ and preventing further recruitment of MMR machinery. FAN1 is a structure-specific nuclease capable of recognising branched structures (Pennell *et al.* 2014; Ceccaldi *et al.* 2016). FAN1 might recognise the non-helical structures adopted by expanded CAG repeats during DNA repair or transcription and bind to the DNA. Indeed, there is some evidence suggesting that FAN1 does bind to the CAG repeat, though there did not appear to be a differential effect of binding between expanded and wild-type repeat lengths (Goold *et al.* 2019). FAN1 might competitively bind to the expanded *HTT* CAG repeat, thereby preventing the binding of MutSβ and subsequent recruitment of mismatch repair machinery; directly inhibiting repeat expansions. One might then

expect that increasing expression of FAN1 would increase the sequestration of MLH1-PMS2, or the binding of FAN1 to the CAG repeat, and reduce repeat expansions further. Indeed, overexpression of FAN1 in U2OS cells with 118 *HTT* CAG repeats was found to reduce repeat expansions below that of the wild-type cell lines (Goold *et al.* 2019).

Alternatively, FAN1 might bind to the unusual secondary structures formed by the expanded CAG repeat in a MutL-dependent or independent manner to promote accurate DNA repair. There are a number of ways FAN1 might promote error-free repair of the repeat. In interstrand cross-link repair, FAN1 is thought to introduce incisions in the DNA necessary for unhooking of the cross-link (Kratz *et al.* 2010; MacKay *et al.* 2010; Shereda *et al.* 2010; Smogorzewska *et al.* 2010; Yoshikiyo *et al.* 2010). Three crystal structures of FAN1 indicate distinct DNA-binding properties; substrate scanning, substrate-latching, and substrate unwinding (Zhao *et al.* 2014). FAN1 may therefore promote local unwinding, incision and removal of the secondary structure formed by the expanded CAG repeat. Interestingly, sensitive to nitrogen mustard 1A (SNM1A), another interstrand cross-link repair protein, functions as a structure-specific, hairpin opening endonuclease (Tiefenbach and Junop 2012). However, FAN1 has weak cleavage at un-nicked Holliday junctions, which have a similar structure to hairpins, and instead favours simple branched structures such as 5' flaps (Pennell *et al.* 2014).

FAN1 might alternatively act on 5' flaps generated by another endonuclease. EXO1 is also a 5'-3' exonuclease and 5' flap endonuclease which acts downstream in the mismatch repair pathway to process the nick generated by MutL. Recent evidence demonstrates that EXO1, like FAN1, is a suppressor of repeat expansions, primarily reducing intergenerational expansions in a mouse model of fragile X disorder (Zhao *et al.* 2018). EXO1 also interacts with the MutL complexes suggesting FAN1 and EXO1 might be recruited to the unusual secondary structures in a similar manner. Additionally, EXO1 participates in mismatch repair recognition and processing of interstrand-cross links in post-mitotic cells, suggesting that as FAN1 and EXO1 have similar substrate specificities, they might have similar functions in repeat expansions (Pizzolato *et al.* 2015; Kato *et al.* 2017; Zhao *et al.* 2018). Moreover, evidence suggests that there is functional redundancy of EXO1 and FAN1 (Desai and Gerson 2014). There appears to be overlap between the functions of mismatch repair and interstrand cross-link repair (Kato *et al.* 2017), and it is plausible that FAN1 acts via a mismatch repair-like process to protect against repeat expansions. FAN1 preferentially cleaves 5' flaps, and both EXO1 and FAN1 could cleave 5' flaps generated by strand-displacement following repair synthesis by DNA polymerase beta (POLβ).

Goold *et al.* (2018), recently demonstrated that the nuclease activity of FAN1 was not required for protection against repeat expansions, though these findings should be repeated in a post-mitotic

cell system. FAN1 might act in a nuclease-independent manner as a molecular scaffold for the assembly of proteins involved in the error-free repair of the abnormal secondary structure.

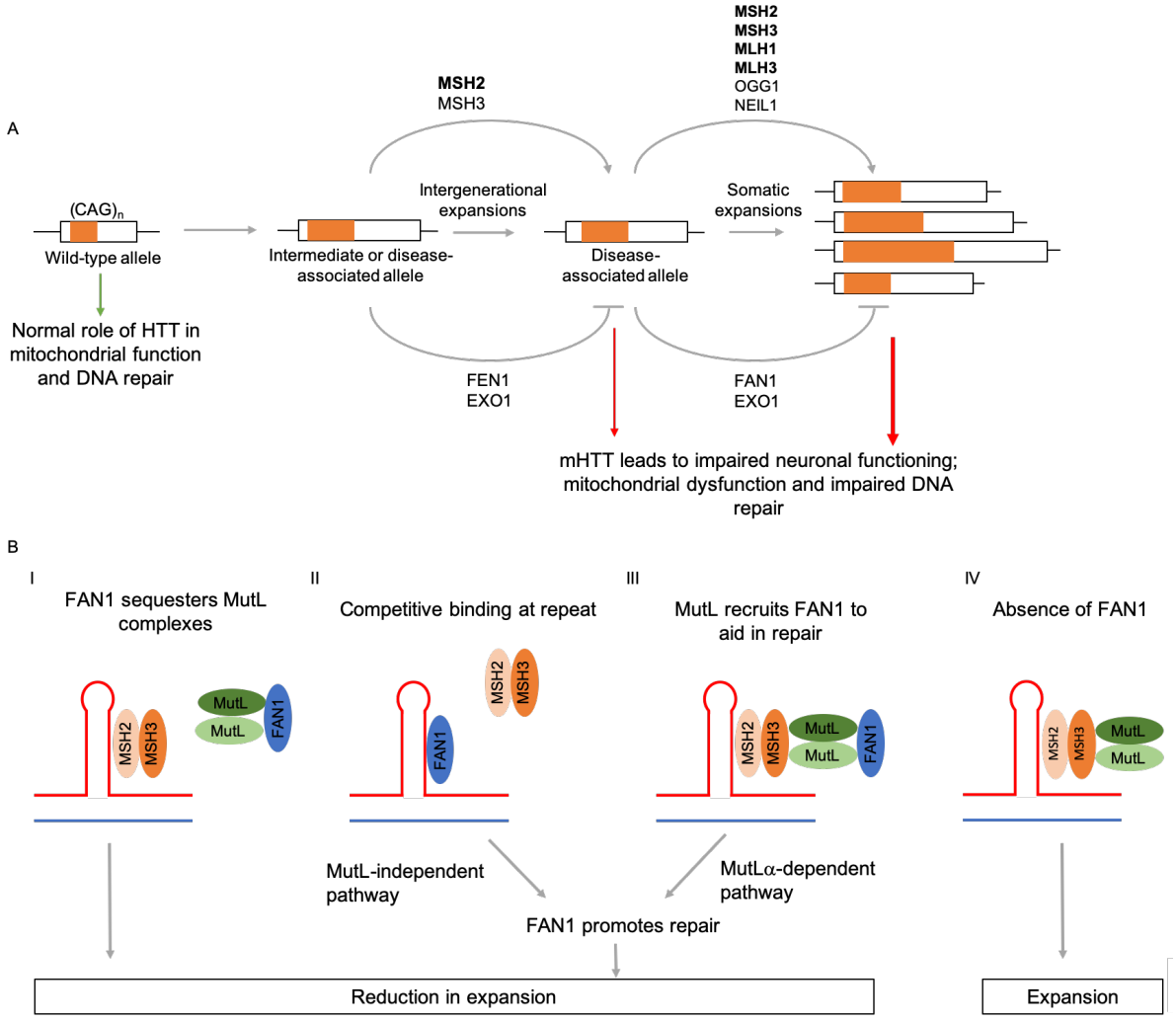


Figure 6.2 (A) DNA repair proteins involved with intergenerational or somatic *HTT CAG repeat expansions.* (B) Possible mechanisms by which FAN1 protects against repeat expansions.

6.3 iPSCs as a model of HTT CAG repeat instability

Somatic instability of the *HTT* CAG repeat is length-dependent and tissue-specific, occurring primarily in the striatal and cortical neurons in the brain, populations selectively lost in HD pathogenesis (Swami *et al.* 2009). Post-mortem HD brain tissue and mouse models of HD have aided our understanding of somatic instability considerably, yet their use is not without caveats. Post-mortem brain tissue from mouse or humans only allows the ascertainment of somatic

instability at one time point, and in post-mortem human brain tissue considerable atrophy has usually already occurred. Additionally, the use of animals to investigate a gene of interest is costly, time-consuming and often involves artificial constructs to study disease-phenotypes. The expression of endogenous mHTT in the genetic background of an individual who lived with HD offers an ideal model system for studying disease-relevant phenotypes, hence iPSCs derived from an individual with HD represent a physiologically relevant system. iPSCs are easily manipulated by CRISPR-Cas9, self-renewing, may be differentiated into disease-relevant cell-types and sampling time points are flexible. Whilst the potential of iPSCs is undoubtable, they are not without issue.

6.3.1 Clonal variability

We observe variability amongst subclones derived from the same clonal line, which made repeat expansion analysis challenging. Some subclones demonstrated considerable repeat expansion, whilst other subclones demonstrated no repeat expansion. We also observed variability in expansions rates between experiments; using the same subclone, at the same passage, but up to 12 months apart. These differences highlight the stochastic mechanism of repeat expansions, perhaps driven by genomic integrity or epigenetic regulators. Genetic and epigenetic variations are thought to contribute to functional variability of clones in culture (Cahan and Daley 2013). Exome sequencing was performed with a number of subclones, and interestingly revealed very few single nucleotide variants between the samples. To increase confidence in results, I would argue that more than 3 subclones of each genotype should be utilised when assessing repeat expansion characteristics. In the case of FAN1, a seemingly protective modifier of HD, we and others report an increase in repeat expansions following its loss. However, measuring an increase in repeat expansions was more difficult than initially expected, as a number of factors clearly govern repeat expansions in iPSCs. Culturing iPSCs with and without FAN1 over 30 passages revealed no significant difference between genotypes, yet up to passage 12 and in post-mitotic neurons, there were significant differences.

6.3.2 Juvenile Huntington's Disease

The Q109 iPSC line was derived from an individual with juvenile HD. The individual developed HD at 3 years old, and the fibroblast sample was obtained when the individual was at 9 years old (The HD iPSC Consortium 2012). The juvenile form of HD typically arises in patients with >60 *HTT* CAG repeats but is clinically characterised by HD onset before the age of 20. Individuals with juvenile HD present with neurodegeneration and clinical phenotypes that may be distinct from adult-onset HD, notably motor deficits tend to manifest as bradykinesia and dystonia, rather than chorea (Quarrell *et al.* 2013; Cronin *et al.* 2019). Much of the literature looking at HD phenotypes involves cell lines

derived from juvenile HD individuals, or mouse models with expanded *HTT* CAG repeats > 60. These models are commonly used as they demonstrate observable phenotypes that often worsen as the repeat length increases. Like many mouse models, the phenotypes seen in Q109 iPSC line are more severe than those seen in cell lines with shorter CAG repeat lengths. This raises the question as to whether the phenotypes observed in these models are accurately reflecting phenotypes of individuals with adult-onset HD, who make up 90% of the HD population and usually have repeat lengths in the range of 40-50 *HTT* CAGs. It is generally accepted that the utility of model systems with an expanded repeat length above the common pathogenic range enables the observation of phenotypes, which should be relevant to adult-onset HD in the later stages of disease, in a more controlled manner, without the need for long experiments to mimic ageing. Understanding the pathogenic mechanisms underlying disease-phenotypes should therefore still be relevant. The Q109 iPSCs and iPSC-derived neurons demonstrate high rates of repeat expansions that have yet been not observed in iPSCs with shorter repeat lengths, or even those with longer repeat lengths (Manley, Pugh, et al. 1999; The HD iPSC Consortium 2012; Xu et al. 2017), yet occur naturally in HD patients. Therefore, it is likely that the use of a juvenile model of HD accelerates the disease phenotypes and allows detection on a smaller time-scale.

6.3.3 Duplication of chromosome 1

The discovery of the duplication of chromosome 1 at later passages in both the parent Q109N1 line, and the resulting daughter clones does have consequences. Chromosomal duplications have been associated with altered metabolic activity, growth rates and cell survival and therefore are likely to effect cellular phenotypes. Excluding the longitudinal *HTT* CAG expansion assays in chapter 5, the work presented in this thesis focused on iPSCs and iPSC-derived neurons that were genotyped as 'karyotypically normal', i.e. had no gross chromosomal duplications or deletions. Phenotypes characterised in this thesis should therefore reflect the length of the *HTT* CAG repeat tract, or the status of FAN1. It should be considered however, what factors might predispose the cell lines to a chromosomal duplication, and whether these factors are driving phenotypic differences preceding chromosomal duplication. External karyotyping of Q109N1, Q109N4 and Q109N5 reveals a duplication of chromosome 1 at later passages of Q109N1 but reports a stable karyotype in Q109N4 and Q109N5 beyond 40 passages. Whether these clones would undergo a duplication of chromosome 1 at later passages is not known. It is plausible that the propensity for a duplication of chromosome 1 was present in the original fibroblast cell used to derive Q109N1, or that there were variants in this fibroblast line which then benefited through duplication of chromosome 1. Due to the repeat instability that these clonal lines exhibit, the application of these cell lines is undeniably beneficial, but the confounds that a duplication of chromosome 1 presents

cannot be ignored. The sensitivity of the assays we perform, particularly looking at metabolic functioning and DNA damage response, means we must implement stringent controls. To identify what effect the duplication of chromosome 1 might have, it would be interesting to perform each assay with subclones prior to and after the duplication of chromosome 1 in Q109N1, alongside Q109N4 and Q109N5, which do not appear to show the same fate. Moving forward, as is already done, extensive characterisation of each clone and its derived daughter clones will be performed at numerous passages.

6.4 Wider Implications

DNA repair is essential for genomic stability and mutations in a number of repair pathways are mutagenic. Loss-of-function variants in a number of mismatch repair genes including *MLH1, MSH2, MSH6* and PMS2 cause Lynch syndrome, or hereditary non-polyposis colorectal cancer, which accounts for 2-4% of colorectal cancers (Kupfer and Ellis 2016; Kanth *et al.* 2017). Though MSH3 is implicated in repeat instability, it appears that its loss is well tolerated, and is not thought to increase the risk of cancer (De Wind *et al.* 1999; Adam *et al.* 2016). There is a lot of work underway to explore therapeutic avenues relating to the mismatch repair system. FAN1 appears to be protective, and transcriptome wide association analyses suggest that increased FAN1 expression is associated with a later age at onset and delayed progression of HD (Goold *et al.* 2019). Overexpression of a protein is challenging therapeutically, and therefore inhibition of a transcriptional repressor of FAN1 might have therapeutic potential.

These cells are well characterised and represent an ideal model for high-throughput CRISPR or drug screens. Repeat expansions are reproducible, and changes in repeat expansion may be detected within a month. These cells thus represent suitable building blocks to dissect the whole network of proteins that may be involved in repeat expansions, and their targets.

6.5 Future Directions

The generation of these cell lines represents a novel tool which can be used to dissect the role of *HTT* repeat expansion on phenotype. The lines are well characterised and may now be utilised to investigate a wide range of phenotypes that have been implicated in HD. In this thesis, we assessed mitochondrial functioning using the Seahorse Bioanalyser. To explore this area further we intend to use a range of mitochondrial assays, and we are particularly interested in cell-by-cell assays. Several assays are of interest in our system including mitochondrial trafficking assays (Shahen *et al.* 2018) and ADP:ATP ratios (Berg *et al.* 2009). Additionally, we could look at the NADH-NAD^+ redox state and interrogate it in cells with an expanded repeat or genetic modifiers of interest (Hung *et*

al. 2011; Bilan *et al.* 2014). We aim to establish a number of assays in our cell systems, and once suitable assays are established we can assess the effect of an expanding CAG repeat on phenotype. Papers often report increasing severity of phenotype in cell lines with a wide range of *HTT* CAG repeat lengths, however, we are interested in the direct effect of incremental repeat expansions over time on phenotype.

Due to time restraints, some experiments performed, and data collected need further replication. There is inherent variability in the assays, likely due to their sensitivity. In addition, stem cells are notoriously variable and also sensitive to environmental stressors or even batch variations in growth medium. Therefore, the number of assays performed, and clonal lines utilised could be expanded to account for technical difficulty and ensure replicability of results.

The tissue-specific nature of repeat expansions is well documented in humans and mice, but less so *in vitro*. It is not known why neuronal populations exhibit the highest rates of somatic instability, though is likely linked to the high oxidative demands of the brain. iPSCs can be differentiated into a range of cell types, and disease-associated phenotypes can be assessed. A recent study used isogenic hESCs of varying repeat lengths (Q35-Q81) and differentiated them to various tissue subtypes; hepatocytes, cardiomyocytes, myocytes and neurons (Ooi *et al.* 2019). They reported mild (±1-2 *HTT* CAG) instability of the *HTT* CAG repeat of Q81-hESCs in striatal neurons, though the *HTT* CAG repeat in hESCs lines was generally stable. Utilising an iPSC line with an expanded repeat, which exhibits expansion in culture would allow a detailed look at differences in expansion rates and the factors driving them between tissues.

To establish how FAN1 might affect repeat modulation it is crucial to know which domains of FAN1 are required to prevent *HTT* repeat expansion. It was recently shown in U2OS cells that the nuclease activity of FAN1 was not required to protect against repeat expansions, as its loss yielded no effect on expansion rates (Goold *et al.* 2019). These findings are perhaps surprising and bear repetition in a more physiological cell line. U2OS cells are replicative cells and so the effect of the loss of the nuclease domain could not be assessed in the context of post-mitotic cells. As we demonstrated, a number of factors govern repeat instability in dividing cells, and thus to reliably establish the role that FAN1 might play *in vivo* efforts should be made to recapitulate the disease system as closely as possible. Colleagues in our research group have recently generated D690A and R507H variants in the Q109 cell lines. These cell lines will allow us to establish whether coding variants in FAN1, or the nuclease activity of FAN1, are required for the expansions we see in culture. As discussed, several hypotheses exist for a mechanism by which FAN1 might protect against repeat expansions. Probing different domains of FAN1 allows us to identify which activities of FAN1 are dependent on repeat modulation. Understanding these factors are critical for establishing whether FAN1 is

actively involved with DNA repair and prevention of aberrant repair; whether FAN1 is responsible for the recruitment of additional DDR proteins; or conversely sequestration of MutLα or MutLγ.

Of particular interest is the effect that knock-out of FAN1 might have on a cellular model with an *HTT* CAG repeat length in the common pathogenic range; 40-50 CAGs, or one that does not show expansion in culture. Utilising a cell line with a shorter *HTT* CAG repeat, i.e. the Q60 iPSC line briefly characterised in chapter 3, would allow us to examine whether the loss of FAN1 would induce expansions of an otherwise stable repeat. We might expect less of an effect, as expansion events are more frequent in cells harbouring longer repeat tracts. Indeed, loss of FAN1 from U2OS cells with a CAG repeat length of 70 remained stable in culture over 8 weeks, whilst the loss of FAN1 had a significant effect on expansion rates in U2OS cells with 97 CAG or 118 CAG repeats (Goold *et al.* 2019). It should be noted however, that a CAA interruption was present in the *HTT* CAG repeat tract of the 70Q U2OS cell line, likely contributing to its stability in culture. To determine if the effects of FAN1 are limiting, overexpression experiments should be performed in the 109N1 line.

It is also necessary to understand the role of FAN1 in the context of the *HTT* CAG repeat, both wild-type and expanded, and therefore we plan to carry out an in-depth RNA-sequencing analysis. Analysis of differential gene expression will also be performed in the following conditions: iPSCs, NPCs and neurons from Q22, Q109 and Q22-FAN1$^{-/-}$ and Q109-FAN1$^{-/-}$ will be analysed. These findings will allow us to see general gene expression changes, but also investigate the effect that mHTT might have on the expression of genes involved in DNA damage and mitochondrial functioning. A general suppression of DNA repair capacity has been demonstrated in HD PBMCs (Askeland *et al.* 2018). Following this it would also be interesting to assess acute transcriptional changes following oxidative damage in each of the lines.

6.6 Concluding remarks

Although HD is monogenic, substantial evidence now suggests that genetic variation elsewhere in the genome may modify the age at which symptoms develop and disease progression. Understanding what these modifiers are, and how they might alter disease pathogenesis provides novel targets for therapeutics. In HD, recent GWAS findings have identified a number of DNA repair genes and DNA repair pathways as modifiers of the age at onset of HD. These findings bring research spanning two decades to the forefront of the HD research field, linking findings from human genetics to DNA repair genes implicated in functional studies from mouse models. DNA repair processes are thought to act on the expanded repeat tract and either prevent or promote further repeat expansions. Using HD iPSCs and iPSC-derived neurons we show that FAN1 protects against repeat expansions *in vitro*. These findings likely extend beyond those of HD and may have

translational relevance in a range of disorders caused by a trinucleotide repeat expansion. FAN1 therefore represents a therapeutic target that might have therapeutic potential in a range of trinucleotide repeat disorders. The novel model systems generated and characterised in this thesis provide a platform for investigating the cellular phenotypes associated with expanded CAG repeats, and the effects of DNA-repair associated genetic modifiers.

References

Aamann, M.D., Hvitby, C., Popuri, V., Muftuoglu, M., Lemminger, L., Skeby, C.K., … Stevnsner, T. (2014). Cockayne Syndrome group B protein stimulates NEIL2 DNA glycosylase activity. *Mechanisms of Ageing and Development*. doi: https://doi.org/10.1016/j.mad.2013.12.008.

Abbott, M.H., Chase, G.A., Jensen, B.A. and Folstein, M.F. (1983). The Association of Affective Disorder with Huntington's Disease in a Case Series and in Families. *Psychological Medicine*. doi: https://doi.org/10.1017/S0033291700047966.

Adam, R., Spier, I., Zhao, B., Kloth, M., Marquez, J., Hinrichsen, I., … Aretz, S. (2016). Exome Sequencing Identifies Biallelic MSH3 Germline Mutations as a Recessive Subtype of Colorectal Adenomatous Polyposis. *American Journal of Human Genetics*. doi: https://doi.org/10.1016/j.ajhg.2016.06.015.

Aiken, C.T., Steffan, J.S., Guerrero, C.M., Khashwji, H., Lukacsovich, T., Simmons, D., … Marsh, J.L. (2009). Phosphorylation of threonine 3: Implications for huntingtin aggregation and neurotoxicity. *Journal of Biological Chemistry*. doi: https://doi.org/10.1074/jbc.M109.013193.

Airik, R., Schueler, M., Airik, M., Cho, J., Porath, J.D., Mukherjee, E., … Schueler, M. (2016). A FANCD2/FANCI-Associated Nuclease 1-Knockout Model Develops Karyomegalic Interstitial Nephritis. *Journal of the American Society of Nephrology* **27**(12):3552–3559. doi: https://doi.org/10.1681/asn.2015101108.

Alano, C.C., Ying, W. and Swanson, R.A. (2004). Poly(ADP-ribose) Polymerase-1-mediated Cell Death in Astrocytes Requires NAD+ Depletion and Mitochondrial Permeability Transition. *Journal of Biological Chemistry*. doi: https://doi.org/10.1074/jbc.M313329200.

Albin, R.L., Reiner, A., Anderson, K.D., Penney, J.B. and Young, A.B. (1990). Striatal and nigral neuron subpopulations in rigid Huntington's disease: Implications for the functional anatomy of chorea and rigidity-akinesia. *Annals of Neurology*. doi: https://doi.org/10.1002/ana.410270403.

Albin, R.L., Young, A.B. and Penney, J.B. (1989). The functional anatomy of basal ganglia disorders. *Trends in Neurosciences*. doi: https://doi.org/10.1016/0166-2236(89)90074-X.

Alexander, G.E. (1994). Basal ganglia-thalamocortical circuits: Their role in control of movements. *Journal of Clinical Neurophysiology*. doi: https://doi.org/10.1097/00004691-199407000-00004.

Alexandre, A. and Lehninger, A.L. (1984). Bypasses of the antimycin A block of mitochondrial electron transport in relation to ubisemiquinone function. *BBA - Bioenergetics*. doi: https://doi.org/10.1016/0005-2728(84)90086-0.

An, M.C., Zhang, N., Scott, G., Montoro, D., Wittkop, T., Mooney, S., … Ellerby, L.M. (2012). Genetic correction of huntington's disease phenotypes in induced pluripotent stem cells. *Cell Stem Cell* **11**(2):253–263. doi: https://doi.org/10.1016/j.stem.2012.04.026.

Andrade, M.A. and Bork, P. (1995). HEAT repeats in the Huntington's disease protein. *Nature Genetics*.

Andrew, S.E., Goldberg, Y.P., Kremer, B., Telenius, H., Theilmann, J., Adam, S., … Hayden, M.R. (1993). The relationship between trinucleotide (CAG) repeat length and clinical features of Huntington's disease. *Nature Genetics*. doi: https://doi.org/10.1038/ng0893-398.

Andrew, S.E., Paul Goldberg, Y., Kremer, B., Squitieri, F., Theilmann, J., Zeisler, J., … Hayden, M.R. (1994). Huntington disease without CAG expansion: Phenocopies or errors in assignment? *American Journal of Human Genetics*.

Anne, S.L., Saudou, F. and Humbert, S. (2007). Phosphorylation of huntingtin by cyclin-dependent kinase 5 is induced by DNA damage and regulates wild-type and mutant huntingtin toxicity in neurons. *Journal of Neuroscience*. doi: https://doi.org/10.1523/JNEUROSCI.1831-07.2007.

Arber, C., Precious, S. V., Cambray, S., Risner-Janiczek, J.R., Kelly, C., Noakes, Z., … Li, M. (2015). Activin A directs striatal projection neuron differentiation of human pluripotent stem cells. *Development*. doi: https://doi.org/10.1242/dev.117093.

Ardlie, K.G., DeLuca, D.S., Segrè, A. V., Sullivan, T.J., Young, T.R., Gelfand, E.T., … Lockhart (2015). The Genotype-Tissue Expression (GTEx) pilot analysis: Multitissue gene regulation in humans. *Science*. doi: https://doi.org/10.1126/science.1262110.

Arlotta, P., Molyneaux, B.J., Jabaudon, D., Yoshida, Y. and Macklis, J.D. (2008). Ctip2 Controls the Differentiation of Medium Spiny Neurons and the Establishment of the Cellular Architecture of the Striatum. *Journal of Neuroscience*. doi: https://doi.org/10.1523/JNEUROSCI.2986-07.2008.

Aronin, N., Chase, K., Young, C., Sapp, E., Schwarz, C., Matta, N., … DiFiglia, M. (1995). CAG expansion affects the expression of mutant huntingtin in the Huntington's disease brain. *Neuron*. doi: https://doi.org/10.1016/0896-6273(95)90106-X.

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

Arun, S., Liu, L. and Donmez, G. (2015). Mitochondrial Biology and Neurological Diseases. *Current Neuropharmacology*. doi: https://doi.org/10.2174/1570159x13666150703154541.

Askeland, G., Dosoudilova, Z., Rodinova, M., Klempir, J., Liskova, I., Kuśnierczyk, A., … Eide, L. (2018).

Increased nuclear DNA damage precedes mitochondrial dysfunction in peripheral blood mononuclear cells from Huntington's disease patients. *Scientific Reports*. doi: https://doi.org/10.1038/s41598-018-27985-y.

Attiyeh, E.F., Diskin, S.J., Attiyeh, M.A., Mossé, Y.P., Hou, C., Jackson, E.M., … Maris, J.M. (2009). Genomic copy number determination in cancer cells from single nucleotide polymorphism microarrays based on quantitative genotyping corrected for aneuploidy. *Genome Research*. doi: https://doi.org/10.1101/gr.075671.107.

Atwal, R.S., Desmond, C.R., Caron, N., Maiuri, T., Xia, J., Sipione, S. and Truant, R. (2011). Kinase inhibitors modulate huntingtin cell localization and toxicity. *Nature Chemical Biology*. doi: https://doi.org/10.1038/nchembio.582.

Atwal, R.S., Xia, J., Pinchev, D., Taylor, J., Epand, R.M. and Truant, R. (2007). Huntingtin has a membrane association signal that can modulate huntingtin aggregation, nuclear entry and toxicity. *Human Molecular Genetics*. doi: https://doi.org/10.1093/hmg/ddm217.

Avery, S., Hirst, A.J., Baker, D., Lim, C.Y., Alagaratnam, S., Skotheim, R.I., … Knowles, B.B. (2013). BCL-XL mediates the strong selective advantage of a 20q11.21 amplification commonly found in human embryonic stem cell cultures. *Stem Cell Reports*. doi: https://doi.org/10.1016/j.stemcr.2013.10.005.

Aylward, E.H., Sparks, B.F., Field, K.M., Yallapragada, V., Shpritz, B.D., Rosenblatt, A., … Ross, C.A. (2004). Onset and rate of striatal atrophy in preclinical Huntington disease. *Neurology*. doi: https://doi.org/10.1212/01.WNL.0000132965.14653.D1.

Bae, B. Il, Xu, H., Igarashi, S., Fujimuro, M., Agrawal, N., Taya, Y., … Sawa, A. (2005). p53 mediates cellular dysfunction and behavioral abnormalities in Huntington's disease. *Neuron*. doi: https://doi.org/10.1016/j.neuron.2005.06.005.

Baig, S.S., Strong, M., Rosser, E., Taverner, N. V., Glew, R., Miedzybrodzka, Z., … Quarrell, O.W. (2016). 22 Years of predictive testing for Huntington's disease: The experience of the UK Huntington's Prediction Consortium. *European Journal of Human Genetics*. doi: https://doi.org/10.1038/ejhg.2016.36.

Baker, D., Hirst, A.J., Gokhale, P.J., Juarez, M.A., Williams, S., Wheeler, M., … Barbaric, I. (2016). Detecting Genetic Mosaicism in Cultures of Human Pluripotent Stem Cells. *Stem Cell Reports*. doi: https://doi.org/10.1016/j.stemcr.2016.10.003.

Ballmaier, D. and Epe, B. (2006). DNA damage by bromate: Mechanism and consequences. *Toxicology*. doi: https://doi.org/10.1016/j.tox.2006.01.009.

Banzai, M., Sato, S., Matsuda, H. and Kanasugi, H. (2004). Trisomy 1 in a case of a missed abortion.

Journal of Human Genetics. doi: https://doi.org/10.1007/s10038-004-0164-1.

Barel, O., Shorer, Z., Flusser, H., Ofir, R., Narkis, G., Finer, G., … Birk, O.S. (2008). Mitochondrial Complex III Deficiency Associated with a Homozygous Mutation in UQCRQ. *American Journal of Human Genetics*. doi: https://doi.org/10.1016/j.ajhg.2008.03.020.

Barese, C.N., Krouse, A.E., Metzger, M.E., King, C.A., Traversari, C., Marini, F.C., … Dunbar, C.E. (2012). Thymidine kinase suicide gene-mediated ganciclovir ablation of autologous gene-modified rhesus hematopoiesis. *Molecular Therapy*. doi: https://doi.org/10.1038/mt.2012.166.

Barrangou, R., Fremaux, C., Deveau, H., Richards, M., Boyaval, P., Moineau, S., … Horvath, P. (2007). CRISPR provides acquired resistance against viruses in prokaryotes. *Science*. doi: https://doi.org/10.1126/science.1138140.

Baydyuk, M. and Xu, B. (2014). BDNF signaling and survival of striatal neurons. *Frontiers in Cellular Neuroscience*.

Benchoua, A., Trioulier, Y., Zala, D., Gaillard, M.C., Lefort, N., Dufour, N., … Brouillet, E. (2006). Involvement of mitochondrial complex II defects in neuronal death produced by N-terminus fragment of mutated huntingtin. *Molecular Biology of the Cell*. doi: https://doi.org/10.1091/mbc.E05-07-0607.

Berg, J., Hung, Y.P. and Yellen, G. (2009). A genetically encoded fluorescent reporter of ATP:ADP ratio. *Nature Methods*. doi: https://doi.org/10.1038/nmeth.1288.

Bettencourt, C., Hensman-Moss, D., Flower, M., Wiethoff, S., Brice, A., Goizet, C., … Jones, L. (2016). DNA repair pathways underlie a common genetic mechanism modulating onset in polyglutamine diseases. *Annals of Neurology* **79**(6):983–990. doi: https://doi.org/10.1002/ana.24656.

De Biase, I., Rasmussen, A., Endres, D., Al-Mahdawi, S., Monticelli, A., Cocozza, S., … Bidichandani, S.I. (2007). Progressive GAA expansions in dorsal root ganglia of Friedreich's ataxia patients. *Annals of Neurology*. doi: https://doi.org/10.1002/ana.21052.

Bichara, M., Pinet, I., Schumacher, S. and Fuchs, R.P.P. (2000). Mechanisms of dinucleotide repeat instability in Escherichia coli. *Genetics*.

Bilan, D.S., Matlashov, M.E., Gorokhovatsky, A.Y., Schultz, C., Enikolopov, G. and Belousov, V. V. (2014). Genetically encoded fluorescent indicator for imaging NAD + /NADH ratio changes in different cellular compartments. *Biochimica et Biophysica Acta - General Subjects*. doi: https://doi.org/10.1016/j.bbagen.2013.11.018.

Bizat, N., Hermel, J.M., Boyer, F., Jacquard, C., Créminon, C., Ouary, 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. *Journal of Neuroscience*.

Blum, B., Bar-Nur, O., Golan-Lev, T. and Benvenisty, N. (2009). The anti-apoptotic gene survivin contributes to teratoma formation by human embryonic stem cells. *Nature Biotechnology*. doi: https://doi.org/10.1038/nbt.1527.

Bodner, R.A., Outeiro, T.F., Altmann, S., Maxwell, M.M., Cho, S.H., Hyman, B.T., … Kazantsev, A.G. (2006). Pharmacological promotion of inclusion formation: A therapeutic approach for Huntington's and Parkinson's diseases. *Proceedings of the National Academy of Sciences of the United States of America*. doi: https://doi.org/10.1073/pnas.0511256103.

Boehler, C., Gauthier, L.R., Mortusewicz, O., Biard, D.S., Saliou, J.M., Bresson, A., … Dantzer, F. (2011). Poly(ADP-ribose) polymerase 3 (PARP3), a newcomer in cellular response to DNA damage and mitotic progression. *Proceedings of the National Academy of Sciences of the United States of America*. doi: https://doi.org/10.1073/pnas.1016574108.

Bogdanov, M.B., Andreassen, O.A., Dedeoglu, A., Ferrante, R.J. and Beal, M.F. (2001). Increased oxidative damage to DNA in a transgenic mouse model of Huntington's disease. *Journal of Neurochemistry*. doi: https://doi.org/10.1046/j.1471-4159.2001.00689.x.

Bolam, J.P., Hanley, J.J., Booth, P.A.C. and Bevan, M.D. (2000). Synaptic organisation of the basal ganglia. *Journal of Anatomy*.

Bolotin, A., Quinquis, B., Sorokin, A. and Dusko Ehrlich, S. (2005). Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin. *Microbiology*. doi: https://doi.org/10.1099/mic.0.28048-0.

Boutell, J.M., Thomas, P., Neal, J.W., Weston, V.J., Duce, J., Harper, P.S. and Jones, A.L. (1999). Aberrant interactions of transcriptional repressor proteins with the Huntington's disease gene product, huntingtin. *Human Molecular Genetics*. doi: https://doi.org/10.1093/hmg/8.9.1647.

Bowater, R.P., Jaworski, A., Larson, J.E., Parniewski, P. and Wells, R.D. (1997). Transcription increases the deletion frequency of long CTG.CAG triplet repeats from plasmids in Escherichia coli. *Nucleic Acids Research*. doi: https://doi.org/10.1093/nar/25.14.2861.

Bragg, R.M., Coffey, S.R., Weston, R.M., Ament, S.A., Cantle, J.P., Minnig, S., … Carroll, J.B. (2017). Corrigendum: Motivational, proteostatic and transcriptional deficits precede synapse loss, gliosis and neurodegeneration in the B6.Htt Q111/+ model of Huntington's disease. *Scientific reports*.

Brennan, W.A., Bird, E.D. and Aprille, J.R. (1985). Regional Mitochondrial Respiratory Activity in Huntington's Disease Brain. *Journal of Neurochemistry*. doi: https://doi.org/10.1111/j.1471-4159.1985.tb07192.x.

Brook, J.D., McCurrach, M.E., Harley, H.G., Buckler, A.J., Church, D., Aburatani, H., … Housman, D.E. (1992). Molecular basis of myotonic dystrophy: Expansion of a trinucleotide (CTG) repeat at the 3′ end of a transcript encoding a protein kinase family member. *Cell*. doi: https://doi.org/10.1016/0092-8674(92)90154-5.

Browne, S.E., Bowling, A.C., MacGarvey, U., Baik, M.J., Berger, S.C., Muqit, M.M.K., … Beal, M.F. (1997). Oxidative damage and metabolic dysfunction in huntington's disease: Selective vulnerability of the basal ganglia. *Annals of Neurology*. doi: https://doi.org/10.1002/ana.410410514.

Browne, S.E., Bowling, A.C., MacGarvey, U., Baik, M.J., Berger, S.C., Muqit, M.M.K., … Beal, M.F. (2006). Oxidative damage in Huntington's disease pathogenesis. *Annals of Neurology*. doi: https://doi.org/10.1089/ars.2006.8.2061.

Bryant, H.E., Petermann, E., Schultz, N., Jemth, A.S., Loseva, O., Issaeva, N., … Helleday, T. (2009). PARP is activated at stalled forks to mediate Mre11-dependent replication restart and recombination. *EMBO Journal*. doi: https://doi.org/10.1038/emboj.2009.206.

Budworth, H., Harris, F.R., Williams, P., Lee, D.Y., Holt, A., Pahnke, J., … McMurray, C.T. (2015). Suppression of Somatic Expansion Delays the Onset of Pathophysiology in a Mouse Model of Huntington's Disease. *PLoS Genetics*. doi: https://doi.org/10.1371/journal.pgen.1005267.

Budworth, H. and McMurray, C.T. (2016). Problems and solutions for the analysis of somatic CAG repeat expansion and their relationship to Huntington's disease toxicity. *Rare Diseases*. doi: https://doi.org/10.1080/21675511.2015.1131885.

van der Burg, J.M., Björkqvist, M. and Brundin, P. (2009). Beyond the brain: widespread pathology in Huntington's disease. *The Lancet Neurology*.

Cahan, P. and Daley, G.Q. (2013). Origins and implications of pluripotent stem cell variability and heterogeneity. *Nature Reviews Molecular Cell Biology*.

Camnasio, S., Carri, A.D., Lombardo, A., Grad, I., Mariotti, C., Castucci, A., … Cattaneo, E. (2012). The first reported generation of several induced pluripotent stem cell lines from homozygous and heterozygous Huntington's disease patients demonstrates mutation related enhanced lysosomal activity. *Neurobiology of Disease*. doi: https://doi.org/10.1016/j.nbd.2011.12.042.

Campuzano, V., Montermini, L., Moltò, M.D., Pianese, L., Cossée, M., Cavalcanti, F., … Pandolfo, M. (1996). Friedreich's ataxia: Autosomal recessive disease caused by an intronic GAA triplet repeat expansion. *Science*. doi: https://doi.org/10.1126/science.271.5254.1423.

Cannavo, E., Gerrits, B., Marra, G., Schlapbach, R. and Jiricny, J. (2007). Characterization of the Interactome of the Human MutL Homologues MLH1, PMS1, and PMS2. *Journal of Biological Chemistry*

282(5):2976–2986. doi: https://doi.org/10.1074/jbc.M609989200.

Cannella, M., Maglione, V., Martino, T., Ragona, G., Frati, L., Li, G.M. and Squitieri, F. (2009). DNA instability in replicating Huntington's disease lymphoblasts. *BMC Medical Genetics*. doi: https://doi.org/10.1186/1471-2350-10-11.

Cantó, C., Menzies, K.J. and Auwerx, J. (2015). NAD+ Metabolism and the Control of Energy Homeostasis: A Balancing Act between Mitochondria and the Nucleus. *Cell Metabolism*.

Cardinale, A., Paldino, E., Giampà, C., Bernardi, G. and Fusco, F.R. (2015). PARP-1 inhibition is neuroprotective in the R6/2 mouse model of huntington's disease. *PLoS ONE*. doi: https://doi.org/10.1371/journal.pone.0134482.

Carmo, C., Naia, L., Lopes, C. and Rego, A.C. (2018). Mitochondrial dysfunction in huntington's disease. *Advances in Experimental Medicine and Biology*.

Ceccaldi, R., Sarangi, P. and D'Andrea, A.D. (2016). The Fanconi anaemia pathway: New players and new functions. *Nature Reviews Molecular Cell Biology*.

Cha, J.H.J. (2007). Transcriptional signatures in Huntington's disease. *Progress in Neurobiology*.

Chae, J. Il, Kim, D.W., Lee, N., Jeon, Y.J., Jeon, I., Kwon, J., … Song, J. (2012). Quantitative proteomic analysis of induced pluripotent stem cells derived from a human Huntington's disease patient. *Biochemical Journal*. doi: https://doi.org/10.1042/BJ20111495.

Chan, D.C. (2006). Mitochondria: Dynamic Organelles in Disease, Aging, and Development. *Cell*.

Chang, D.T.W., Rintoul, G.L., Pandipati, S. and Reynolds, I.J. (2006). Mutant huntingtin aggregates impair mitochondrial movement and trafficking in cortical neurons. *Neurobiology of Disease*. doi: https://doi.org/10.1016/j.nbd.2005.12.007.

Charpentier, E. and Doudna, J.A. (2013). Biotechnology: Rewriting a genome. *Nature*. doi: https://doi.org/10.1038/495050a.

Chatterjee, A., Saha, S., Chakraborty, A., Silva-Fernandes, A., Mandal, S.M., Neves-Carvalho, A., … Hazra, T.K. (2015). The Role of the Mammalian DNA End-processing Enzyme Polynucleotide Kinase 3'-Phosphatase in Spinocerebellar Ataxia Type 3 Pathogenesis. *PLoS Genetics*. doi: https://doi.org/10.1371/journal.pgen.1004749.

Chaturvedi, R.K., Calingasan, N.Y., Yang, L., Hennessey, T., Johri, A. and Beal, M.F. (2010). Impairment of PGC-1alpha expression, neuropathology and hepatic steatosis in a transgenic mouse model of Huntington's disease following chronic energy deprivation. *Human Molecular Genetics*. doi: https://doi.org/10.1093/hmg/ddq229.

Chaudhury, I., Stroik, D.R. and Sobeck, A. (2014). FANCD2-Controlled Chromatin Access of the Fanconi-Associated Nuclease FAN1 Is Crucial for the Recovery of Stalled Replication Forks. *Molecular and Cellular Biology*. doi: https://doi.org/10.1128/mcb.00457-14.

Che, H.V.B., Metzger, S., Portal, E., Deyle, C., Riess, O. and Nguyen, H.P. (2011). Localization of sequence variations in PGC-1 influence their modifying effect in Huntington disease. *Molecular Neurodegeneration*. doi: https://doi.org/10.1186/1750-1326-6-1.

Chehrehasa, F., Meedeniya, A.C.B., Dwyer, P., Abrahamsen, G. and Mackay-Sim, A. (2009). EdU, a new thymidine analogue for labelling proliferating cells in the nervous system. *Journal of Neuroscience Methods*. doi: https://doi.org/10.1016/j.jneumeth.2008.10.006.

Chen, F., Pruett-Miller, S.M., Huang, Y., Gjoka, M., Duda, K., Taunton, J., … Davis, G.D. (2011). High-frequency genome editing using ssDNA oligonucleotides with zinc-finger nucleases. *Nature Methods*. doi: https://doi.org/10.1038/nmeth.1653.

Chen, G.L., Ma, Q., Goswami, D., Shang, J. and Miller, G.M. (2017). Modulation of nuclear REST by alternative splicing: a potential therapeutic target for Huntington's disease. *Journal of Cellular and Molecular Medicine*. doi: https://doi.org/10.1111/jcmm.13209.

Cherubini, M., Puigdellívol, M., Alberch, J. and Ginés, S. (2015). Cdk5-mediated mitochondrial fission: A key player in dopaminergic toxicity in Huntington's disease. *Biochimica et Biophysica Acta - Molecular Basis of Disease*. doi: https://doi.org/10.1016/j.bbadis.2015.06.025.

Chiu, F.L., Lin, J.T., Chuang, C.Y., Chien, T., Chen, C.M., Chen, K.H., … Kuo, H.C. (2015). Elucidating the role of the A 2A adenosine receptor in neurodegeneration using neurons derived from Huntington's disease iPSCs. *Human Molecular Genetics*. doi: https://doi.org/10.1093/hmg/ddv318.

Chong, S.S., McCall, A.E., Cota, J., Subramony, S.H., Orr, H.T., Hughes, M.R. and Zoghbi, H.Y. (1995). Gametic and somatic tissue–specific heterogeneity of the expanded SCA1 CAG repeat in spinocerebellar ataxia type 1. *Nature Genetics*. doi: https://doi.org/10.1038/ng0795-344.

Choo, Y.S., Johnson, G.V.W., MacDonald, M., Detloff, P.J. and Lesort, M. (2004). Mutant huntingtin directly increases susceptibility of mitochondria to the calcium-induced permeability transition and cytochrome c release. *Human Molecular Genetics*. doi: https://doi.org/10.1093/hmg/ddh162.

Ciammola, A., Sassone, J., Sciacco, M., Mencacci, N.E., Ripolone, M., Bizzi, C., … Malfatto, G. (2011). Low anaerobic threshold and increased skeletal muscle lactate production in subjects with Huntington's disease. *Movement Disorders*. doi: https://doi.org/10.1002/mds.23258.

Ciosi, M., Ciosi, M., Cumming, S.A., Mubarak, A., Symeonidi, E., Herzyk, P., … Monckton, D.G. (2018). Library preparation and MiSeq sequencing for the genotyping-by-sequencing of the Huntington

disease HTT exon one trinucleotide repeat and the quantification of somatic mosaicism. *Protocol Exchange*. doi: https://doi.org/10.1038/protex.2018.089.

Ciosi, M., Maxwell, A., Cumming, S.A., Moss, D.J.H., Alshammari, A.M., Flower, M.D., … Monckton, D.G. (2019). A genetic association study of glutamine-encoding DNA sequence structures, somatic CAG expansion, and DNA repair gene variants, with Huntington disease clinical outcomes. *EBioMedicine* **48**:568–580. doi: https://doi.org/https://doi.org/10.1016/j.ebiom.2019.09.020.

Claassen, D.A. and Lahue, R.S. (2007). Expansions of CAG•CTG repeats in immortalized human astrocytes. *Human Molecular Genetics*. doi: https://doi.org/10.1093/hmg/ddm270.

Claassen, E., Kortbeek, H. and Arwert, F. (1986). Effects of mitomycin C on the rate of DNA synthesis in normal and Faconi anaemia cells. *Mutation Research DNA Repair Reports*. doi: https://doi.org/10.1016/0167-8817(86)90004-0.

Conforti, P., Besusso, D., Bocchi, V.D., Faedo, A., Cesana, E., Rossetti, G., … Cattaneo, E. (2018). Faulty neuronal determination and cell polarization are reverted by modulating HD early phenotypes. *Proceedings of the National Academy of Sciences of the United States of America*. doi: https://doi.org/10.1073/pnas.1715865115.

Cong, L. and Zhang, F. (2014). Genome engineering using crispr-cas9 system. *Chromosomal Mutagenesis: Second Edition*.

Cooke, M.S., Evans, M.D., Dizdaroglu, M. and Lunec, J. (2003). Oxidative DNA damage: mechanisms, mutation, and disease. *The FASEB Journal*. doi: https://doi.org/10.1096/fj.02-0752rev.

Cooper, J.K., Schilling, G., Peters, M.F., Herring, W.J., Sharp, A.H., Kaminsky, Z., … Ross, C.A. (1998). Truncated N-terminal fragments of huntingtin with expanded glutamine repeats form nuclear and cytoplasmic aggregates in cell culture. *Human Molecular Genetics*. doi: https://doi.org/10.1093/hmg/7.5.783.

Cornett, J., Cao, F., Wang, C.E., Ross, C.A., Bates, G.P., Li, S.H. and Li, X.J. (2005). Polyglutamine expansion of huntingtin impairs its nuclear export. *Nature Genetics*. doi: https://doi.org/10.1038/ng1503.

Costa, V., Giacomello, M., Hudec, R., Lopreiato, R., Ermak, G., Lim, D., … Scorrano, L. (2010). Mitochondrial fission and cristae disruption increase the response of cell models of Huntington's disease to apoptotic stimuli. *EMBO Molecular Medicine*. doi: https://doi.org/10.1002/emmm.201000102.

Craufurd, D. and Dodge, A. (1993). Mutation size and age at onset in Huntington's disease. *Journal of Medical Genetics*.

Crompton, M. (1999). The mitochondrial permeabitity transition pore and its role in cell death. *Biochemical Journal*.

Cronin, T., Rosser, A. and Massey, T. (2019). Clinical Presentation and Features of Juvenile-Onset Huntington's Disease: A Systematic Review. *Journal of Huntington's Disease*.

Cui, L., Jeong, H., Borovecki, F., Parkhurst, C.N., Tanese, N. and Krainc, D. (2006). Transcriptional Repression of PGC-1α by Mutant Huntingtin Leads to Mitochondrial Dysfunction and Neurodegeneration. *Cell*. doi: https://doi.org/10.1016/j.cell.2006.09.015.

Cummings, C.J., Mancini, M.A., Antalffy, B., DeFranco, D.B., Orr, H.T. and Zoghbi, H.Y. (1998). Chaperone suppression of aggregation and altered subcellular proteasome localization imply protein misfolding in SCA1. *Nature Genetics*. doi: https://doi.org/10.1038/502.

Damiano, M., Diguet, E., Malgorn, C., D'Aurelio, M., Galvan, L., Petit, F., … Brouillet, E. (2013). A role of mitochondrial complex II defects in genetic models of Huntington's disease expressing N-terminal fragments of mutant huntingtin. *Human Molecular Genetics*. doi: https://doi.org/10.1093/hmg/ddt242.

Damiano, M., Galvan, L., Déglon, N. and Brouillet, E. (2010). Mitochondria in Huntington's disease. *Biochimica et Biophysica Acta - Molecular Basis of Disease*.

Davenport, C.B. (1912). Heredity in Relation to Eugenics. *The ANNALS of the American Academy of Political and Social Science*. doi: https://doi.org/10.1177/000271621204200148.

David, G., Dürr, A., Stevanin, G., Cancel, G., Abbas, N., Benomar, A., … Brice, A. (1998). Molecular and clinical correlations in autosomal dominant cerebellar ataxia with progressive macular dystrophy (SCA7). *Human Molecular Genetics*. doi: https://doi.org/10.1093/hmg/7.2.165.

Delli Carri, A., Onorati, M., Castiglioni, V., Faedo, A., Camnasio, S., Toselli, M., … Cattaneo, E. (2013). Human Pluripotent Stem Cell Differentiation into Authentic Striatal Projection Neurons. *Stem Cell Reviews and Reports*. doi: https://doi.org/10.1007/s12015-013-9441-8.

DeLong, M.R. (1990). Primate models of movement disorders of basal ganglia origin. *Trends in Neurosciences*.

Desai, A. and Gerson, S. (2014). Exo1 independent DNA mismatch repair involves multiple compensatory nucleases. *DNA Repair*. doi: https://doi.org/10.1016/j.dnarep.2014.06.005.

Deveau, H., Barrangou, R., Garneau, J.E., Labonté, J., Fremaux, C., Boyaval, P., … Moineau, S. (2008). Phage response to CRISPR-encoded resistance in Streptococcus thermophilus. *Journal of Bacteriology*. doi: https://doi.org/10.1128/JB.01412-07.

DiFiglia, M., Sapp, E., Chase, K., Schwarz, C., Meloni, A., Young, C., … Aronin, N. (1995). Huntingtin is a cytoplasmic protein associated with vesicles in human and rat brain neurons. *Neuron*. doi: https://doi.org/10.1016/0896-6273(95)90346-1.

DiGiovanni, L.F., Mocle, A.J., Xia, J. and Truant, R. (2016). Huntingtin N17 domain is a reactive oxygen species sensor regulating huntingtin phosphorylation and localization. *Human Molecular Genetics*. doi: https://doi.org/10.1093/hmg/ddw234.

Dimos, J.T., Rodolfa, K.T., Niakan, K.K., Weisenthal, L.M., Mitsumoto, H., Chung, W., … Eggan, K. (2008). Induced pluripotent stem cells generated from patients with ALS can be differentiated into motor neurons. *Science*. doi: https://doi.org/10.1126/science.1158799.

Ding, S., Wu, X., Li, G., Han, M., Zhuang, Y. and Xu, T. (2005). Efficient transposition of the piggyBac (PB) transposon in mammalian cells and mice. *Cell*. doi: https://doi.org/10.1016/j.cell.2005.07.013.

Divakaruni, A.S., Paradyse, A., Ferrick, D.A., Murphy, A.N. and Jastroch, M. (2014). Analysis and interpretation of microplate-based oxygen consumption and pH data. *Methods in Enzymology*.

Djoussé, L., Knowlton, B., Cupples, L.A., Marder, K., Shoulson, I. and Myers, R.H. (2002). Weight loss in early stage of Huntington's disease. *Neurology*. doi: https://doi.org/10.1212/01.WNL.0000031791.10922.CF.

Djoussé, L., Knowlton, B., Hayden, M., Almqvist, E.W., Brinkman, R., Ross, C., … Myers, R.H. (2003). Interaction of normal and expanded CAG repeat sizes influences age at onset of Huntington disease. *American Journal of Medical Genetics Part A* **119A**(3):279–282. doi: https://doi.org/10.1002/ajmg.a.20190.

Djoussé, L., Knowlton, B., Hayden, M.R., Almqvist, E.W., Brinkman, R.R., Ross, C.A., … Myers, R.H. (2004). Evidence for a modifier of onset age in Huntington disease linked to the HD gene in 4p16. *Neurogenetics*. doi: https://doi.org/10.1007/s10048-004-0175-2.

v. Domarus, H. (1983). Hypotelorism in trisomy 1-producing mice. *The Anatomical Record*. doi: https://doi.org/10.1002/ar.1092060309.

Dorsey, E.R. (2012). Characterization of a large group of individuals with huntington disease and their relatives enrolled in the COHORT study. *PLoS ONE*. doi: https://doi.org/10.1371/journal.pone.0029522.

Dragileva, E., Hendricks, A., Teed, A., Gillis, T., Lopez, E.T., Friedberg, E.C., … Wheeler, V.C. (2009). Intergenerational and striatal CAG repeat instability in Huntington's disease knock-in mice involve different DNA repair genes. *Neurobiology of disease* **33**(1):37–47. doi: https://doi.org/10.1016/j.nbd.2008.09.014.

Du, J., Campau, E., Soragni, E., Jespersen, C. and Gottesfeld, J.M. (2013). Length-dependent CTG.CAG triplet-repeat expansion in myotonic dystrophy patient-derived induced pluripotent stem cells. *Human Molecular Genetics*. doi: https://doi.org/10.1093/hmg/ddt386.

Du, J., Campau, E., Soragni, E., Ku, S., Puckett, J.W., Dervan, P.B. and Gottesfeld, J.M. (2012). Role of mismatch repair enzymes in GAA·TTC triplet-repeat expansion in friedreich ataxia induced pluripotent stem cells. *Journal of Biological Chemistry*. doi: https://doi.org/10.1074/jbc.M112.391961.

Van Duijn, E., Kingma, E.M. and Van Der Mast, R.C. (2007). Psychopathology in verified Huntington's disease gene carriers. *Journal of Neuropsychiatry and Clinical Neurosciences*. doi: https://doi.org/10.1176/jnp.2007.19.4.441.

Dunlap, C. (1927). Pathological Changes in Huntington's Chorea. *Journal of Mental Science*. doi: https://doi.org/10.1192/bjp.74.305.335.

Durr, A., Hahn-Barma, V., Brice, A., Pecheux, C., Dode, C. and Feingold, J. (1999). Homozygosity in Huntington's disease [2]. *Journal of Medical Genetics*.

Dürr, A., Stevanin, G., Cancel, G., Duyckaerts, C., Abbas, N., Didierjean, O., … Brice, A. (1996). Spinocerebellar ataxia 3 and Machado-Joseph disease: Clinical, molecular, and neuropathological features. *Annals of Neurology*. doi: https://doi.org/10.1002/ana.410390411.

Duyao, M., Ambrose, C., Myers, R., Novelletto, A., Persichetti, F., Frontali, M., … Macdonald, M. (1993). Trinucleotide repeat length instability and age of onset in Huntington's disease. *Nature Genetics*. doi: https://doi.org/10.1038/ng0893-387.

Duyao, M.P., Auerbach, A.B., Ryan, A., Persichetti, F., Barnes, G.T., McNeil, S.M., … MacDonald, M.E. (1995). Inactivation of the mouse huntington's disease gene homolog Hdh. *Science*. doi: https://doi.org/10.1126/science.7618107.

Eddy, C.M., Parkinson, E.G. and Rickards, H.E. (2016). Changes in mental state and behaviour in Huntington's disease. *The Lancet Psychiatry*.

Elkabetz, Y., Panagiotakos, G., Al Shamy, G., Socci, N.D., Tabar, V. and Studer, L. (2008). Human ES cell-derived neural rosettes reveal a functionally distinct early neural stem cell stage. *Genes and Development*. doi: https://doi.org/10.1101/gad.1616208.

Enokido, Y., Tamura, T., Ito, H., Arumughan, A., Komuro, A., Shiwaku, H., … Okazawa, H. (2010). Mutant huntingtin impairs Ku70-mediated DNA repair. *Journal of Cell Biology*. doi: https://doi.org/10.1083/jcb.200905138.

Entezam, A., Lokanga, A.R., Le, W., Hoffman, G. and Usdin, K. (2010). Potassium bromate, a potent

DNA oxidizing agent, exacerbates germline repeat expansion in a fragile X premutation mouse model. *Human Mutation*. doi: https://doi.org/10.1002/humu.21237.

Enver, T., Soneji, S., Joshi, C., Brown, J., Iborra, F., Orntoft, T., … Andrews, P.W. (2005). Cellular differentiation hierarchies in normal and culture-adapted human embryonic stem cells. *Human Molecular Genetics*. doi: https://doi.org/10.1093/hmg/ddi345.

Evans, S.J.W., Douglas, I., Rawlins, M.D., Wexler, N.S., Tabrizi, S.J. and Smeeth, L. (2013). Prevalence of adult Huntington's disease in the UK based on diagnoses recorded in general practice records. *Journal of Neurology, Neurosurgery and Psychiatry*. doi: https://doi.org/10.1136/jnnp-2012-304636.

Fang, L. and Wang, K. (2018). Identification of copy number variants from SNP arrays using PennCNV. *Methods in Molecular Biology*.

Faragó, A., Zsindely, N. and Bodai, L. (2019). Mutant huntingtin disturbs circadian clock gene expression and sleep patterns in Drosophila. *Scientific Reports*. doi: https://doi.org/10.1038/s41598-019-43612-w.

Ferlazzo, M.L., Sonzogni, L., Granzotto, A., Bodgi, L., Lartin, O., Devic, C., … Foray, N. (2014). Mutations of the Huntington's disease protein impact on the ATM-dependent signaling and repair pathways of the radiation-induced DNA double-strand breaks: Corrective effect of statins and bisphosphonates. *Molecular Neurobiology*. doi: https://doi.org/10.1007/s12035-013-8591-7.

Finck, B.N. and Kelly, D.P. (2006). PGC-1 coactivators: Inducible regulators of energy metabolism in health and disease. *Journal of Clinical Investigation*.

Firdaus, W.J.J., Wyttenbach, A., Giuliano, P., Kretz-Remy, C., Currie, R.W. and Arrigo, A.P. (2006). Huntingtin inclusion bodies are iron-dependent centers of oxidative events. *FEBS Journal*. doi: https://doi.org/10.1111/j.1742-4658.2006.05537.x.

Fishel, R. and Wilson, T. (1997). MutS homologs in mammalian cells. *Current Opinion in Genetics and Development*. doi: https://doi.org/10.1016/S0959-437X(97)80117-7.

Fortini, P., Ferretti, C., Pascucci, B., Narciso, L., Pajalunga, D., Puggioni, E.M.R., … Dogliotti, E. (2012). DNA damage response by single-strand breaks in terminally differentiated muscle cells and the control of muscle integrity. *Cell Death and Differentiation*. doi: https://doi.org/10.1038/cdd.2012.53.

Fried, H.M. (2007). Nutrient-sensitive mitochondrial NAD+ levels dictate cell survival. *Chemtracts*.

Friedlander, R.M. (2003). Apoptosis and caspases in neurodegenerative diseases. *New England Journal of Medicine*.

Fujigasaki, H., Verma, I.C., Camuzat, A., Margolis, R.L., Zander, C., Lebre, A.S., … Brice, A. (2001).

SCA12 is a rare locus for autosomal dominant cerebellar ataxia: A study of an Indian family. *Annals of Neurology*. doi: https://doi.org/10.1002/1531-8249(200101)49:1<117::AID-ANA19>3.0.CO;2-G.

Gafni, J. and Ellerby, L.M. (2002). Calpain Activation in Huntington's Disease. *Journal of Neuroscience*.

Gafni, J., Hermel, E., Young, J.E., Wellington, C.L., Hayden, M.R. and Ellerby, L.M. (2004). Inhibition of calpain cleavage of Huntingtin reduces toxicity: Accumulation of calpain/caspase fragments in the nucleus. *Journal of Biological Chemistry*. doi: https://doi.org/10.1074/jbc.M401267200.

Gao, R., Chakraborty, A., Geater, C., Pradhan, S., Gordon, K.L., Snowden, J., … Sarkar, P.S. (2019). Mutant huntingtin impairs PNKP and ATXN3, disrupting DNA repair and transcription. *eLife*. doi: https://doi.org/10.7554/eLife.42988.

Gao, R., Liu, Y., Silva-Fernandes, A., Fang, X., Paulucci-Holthauzen, A., Chatterjee, A., … Sarkar, P.S. (2015). Inactivation of PNKP by Mutant ATXN3 Triggers Apoptosis by Activating the DNA Damage-Response Pathway in SCA3. *PLoS Genetics*. doi: https://doi.org/10.1371/journal.pgen.1004834.

Gazy, I., Hayward, B., Potapova, S., Zhao, X. and Usdin, K. (2019). Double-strand break repair plays a role in repeat instability in a fragile X mouse model. *DNA Repair*. doi: https://doi.org/10.1016/j.dnarep.2018.12.004.

Gellerich, F.N., Gizatullina, Z., Nguyen, H.P., Trumbeckaite, S., Vielhaber, S., Seppet, E., … Striggow, F. (2008). Impaired regulation of brain mitochondria by extramitochondrial Ca 2+ in transgenic huntington disease rats. *Journal of Biological Chemistry*. doi: https://doi.org/10.1074/jbc.M709555200.

GeM-HD Consortium, G.M. of H.D. (2015). Identification of Genetic Factors that Modify Clinical Onset of Huntington's Disease. *Cell* 162(3):516–526. doi: https://doi.org/10.1016/j.cell.2015.07.003.

GeM-HD Consortium, G.M. of H.D. (2019). Huntington's disease onset is determined by length of uninterrupted CAG, not encoded polyglutamine, and is modified by DNA maintenance mechanisms. *Cell*. doi: https://doi.org/10.1101/529768.

Giuliano, P., de Cristofaro, T., Affaitati, A., Pizzulo, G.M., Feliciello, A., Criscuolo, C., … Varrone, S. (2003). DNA damage induced by polyglutamine-expanded proteins. *Human Molecular Genetics*.

Goebel, H.H., Heipertz, R., Scholz, W., Iqbal, K. and Tellez-Nagel, I. (1978). Juvenile huntington chorea: Clinical, ultrastructural, and biochemical studies. *Neurology*.

Goehler, H., Lalowski, M., Stelzl, U., Waelter, S., Stroedicke, M., Worm, U., … Wanker, E.E. (2004). A protein interaction network links GIT1, an enhancer of huntingtin aggregation, to Huntington's disease. *Molecular Cell*. doi: https://doi.org/10.1016/j.molcel.2004.09.016.

Gonitel, R., Moffitt, H., Sathasivam, K., Woodman, B., Detloff, P.J., Faull, R.L.M. and Bates, G.P. (2008). DNA instability in postmitotic neurons. *Proceedings of the National Academy of Sciences of the United States of America*. doi: https://doi.org/10.1073/pnas.0800048105.

Goold, R., Flower, M., Moss, D.H., Medway, C., Wood-Kaczmar, A., Andre, R., … Tabrizi, S.J. (2019). FAN1 modifies Huntington's disease progression by stabilizing the expanded HTT CAG repeat. *Human Molecular Genetics* **28**(4):650–661. doi: https://doi.org/10.1093/hmg/ddy375.

Gordon, A.M., Quinn, L., Reilmann, R. and Marder, K. (2000). Coordination of prehensile forces during precision grip in Huntington's disease. *Experimental Neurology*. doi: https://doi.org/10.1006/exnr.2000.7348.

Goula, A.V., Stys, A., Chan, J.P.K., Trottier, Y., Festenstein, R. and Merienne, K. (2012). Transcription Elongation and Tissue-Specific Somatic CAG Instability. *PLoS Genetics*. doi: https://doi.org/10.1371/journal.pgen.1003051.

Gradia, S., Subramanian, D., Wilson, T., Acharya, S., Makhov, A., Griffith, J. and Fishel, R. (1999). hMSH2-hMSH6 forms a hydrolysis-independent sliding clamp on mismatched DNA. *Molecular Cell*. doi: https://doi.org/10.1016/S1097-2765(00)80316-0.

Graham, R.K., Deng, Y., Slow, E.J., Haigh, B., Bissada, N., Lu, G., … Hayden, M.R. (2006). Cleavage at the Caspase-6 Site Is Required for Neuronal Dysfunction and Degeneration Due to Mutant Huntingtin. *Cell*. doi: https://doi.org/10.1016/j.cell.2006.04.026.

Graveland, G.A., Williams, R.S. and Difiglia, M. (1985). Evidence for degenerative and regenerative changes in neostriatal spiny neurons in Huntington's disease. *Science*. doi: https://doi.org/10.1126/science.3155875.

Gray, M., Shirasaki, D.I., Cepeda, C., André, V.M., Wilburn, B., Lu, X.H., … Yang, X.W. (2008). Full-length human mutant huntingtin with a stable polyglutamine repeat can elicit progressive and selective neuropathogenesis in BACHD mice. *Journal of Neuroscience*. doi: https://doi.org/10.1523/JNEUROSCI.0857-08.2008.

Graybiel, A.M. (1998). The basal ganglia and chunking of action repertoires. *Neurobiology of Learning and Memory*.

Greig, L.C., Woodworth, M.B., Galazo, M.J., Padmanabhan, H. and Macklis, J.D. (2013). Molecular logic of neocortical projection neuron specification, development and diversity. *Nature Reviews Neuroscience*.

Guidetti, P., Charles, V., Chen, E.Y., Reddy, P.H., Kordower, J.H., Whetsell, W.O., … Tagle, D.A. (2001). Early degenerative changes in transgenic mice expressing mutant huntingtin involve dendritic

abnormalities but no impairment of mitochondrial energy production. *Experimental Neurology*. doi: https://doi.org/10.1006/exnr.2000.7626.

Gunn, A.R., Banos-Pinero, B., Paschke, P., Sanchez-Pulido, L., Ariza, A., Day, J., … Lakin, N.D. (2016). The role of ADP-ribosylation in regulating DNA interstrand crosslink repair. *Journal of Cell Science*. doi: https://doi.org/10.1242/jcs.193375.

Guo, X., Disatnik, M.H., Monbureau, M., Shamloo, M., Mochly-Rosen, D. and Qi, X. (2013). Inhibition of mitochondrial fragmentation diminishes Huntington's disease-associated neurodegeneration. *Journal of Clinical Investigation*. doi: https://doi.org/10.1172/JCI70911.

Gupte, R., Liu, Z. and Kraus, W.L. (2017). Parps and adp-ribosylation: Recent advances linking molecular functions to biological outcomes. *Genes and Development*.

Gusella, J.F. and MacDonald, M.E. (1996). TRINUCLEOTIDE INSTABILITY: A Repeating Theme in Human Inherited Disorders. *Annual Review of Medicine*. doi: https://doi.org/10.1146/annurev.med.47.1.201.

Gusella, J.F., Wexler, N.S., Conneally, P.M., Naylor, S.L., Anderson, M.A., Tanzi, R.E., … Martin, J.B. (1983). A polymorphic DNA marker genetically linked to Huntington's disease. *Nature*. doi: https://doi.org/10.1038/306234a0.

Hadzi, T.C., Hendricks, A.E., Latourelle, J.C., Lunetta, K.L., Cupples, L.A., Gillis, T., … Vonsattel, J.P. (2012). Assessment of cortical and striatal involvement in 523 Huntington disease brains. *Neurology*. doi: https://doi.org/10.1212/WNL.0b013e31826e9a5d.

Hafner, A. V., Dai, J., Gomes, A.P., Xiao, C.Y., Palmeira, K.C.M., Rosenzweig, A. and Sinclair, D.A. (2010). Regulation of the mPTP by SIRT3-mediated deacetylation of CypD at lysine 166 suppresses age-related cardiac hypertrophy. *Aging*. doi: https://doi.org/10.18632/aging.100252.

Haigis, M.C. and Guarente, L.P. (2006). Mammalian sirtuins - Emerging roles in physiology, aging, and calorie restriction. *Genes and Development*.

Halestrap, A.P. (2006). Calcium, mitochondria and reperfusion injury: A pore way to die. *Biochemical Society Transactions*.

Hanawalt, P.C., Crowley, D.J., Ford, J.M., Ganesan, A.K., Lloyd, D.R., Nouspikel, T., … Tornaletti, S. (2000). Regulation of nucleotide excision repair in bacteria and mammalian cells. *Cold Spring Harbor Symposia on Quantitative Biology*.

Hands, S., Sajjad, M.U., Newton, M.J. and Wyttenbach, A. (2011). In vitro and in vivo aggregation of a fragment of huntingtin protein directly causes free radical production. *Journal of Biological Chemistry*. doi: https://doi.org/10.1074/jbc.M111.307587.

Harjes, P. and Wanker, E.E. (2003). The hunt for huntingtin function: Interaction partners tell many different stories. *Trends in Biochemical Sciences*.

Hayden, M.R. (1983). Huntington's Chorea. *Springer-Verlag*.

He, F., Zhang, S., Qian, F. and Zhang, C. (1995). Delayed dystonia with striatal ct lucencies induced by a mycotoxin (3-nitropropionic acid). *Neurology*. doi: https://doi.org/10.1212/WNL.45.12.2178.

Heinz, S., Freyberger, A., Lawrenz, B., Schladt, L., Schmuck, G. and Ellinger-Ziegelbauer, H. (2017). Mechanistic Investigations of the Mitochondrial Complex i Inhibitor Rotenone in the Context of Pharmacological and Safety Evaluation. *Scientific Reports*. doi: https://doi.org/10.1038/srep45465.

Hensley, K., Butterfield, D.A., Hall, N., Cole, P., Subramaniam, R., Mark, R., … Carney, J.M. (1996). Reactive oxygen species as causal agents in the neurotoxicity of the Alzheimer's disease-associated amyloid beta peptide. *Annals of the New York Academy of Sciences*.

Hering, T., Birth, N., Taanman, J.W. and Orth, M. (2015). Selective striatal mtDNA depletion in end-stage Huntington's disease R6/2 mice. *Experimental Neurology*. doi: https://doi.org/10.1016/j.expneurol.2015.02.004.

Hishimoto, A., Pletnikova, O., Lang, D.L., Troncoso, J.C., Egan, J.M. and Liu, Q.R. (2019). Neurexin 3 transmembrane and soluble isoform expression and splicing haplotype are associated with neuron inflammasome and Alzheimer's disease. *Alzheimer's Research and Therapy*. doi: https://doi.org/10.1186/s13195-019-0475-2.

Hockemeyer, D., Soldner, F., Beard, C., Gao, Q., Mitalipova, M., Dekelver, R.C., … Jaenisch, R. (2009). Efficient targeting of expressed and silent genes in human ESCs and iPSCs using zinc-finger nucleases. *Nature Biotechnology*. doi: https://doi.org/10.1038/nbt.1562.

Hodges, A., Strand, A.D., Aragaki, A.K., Kuhn, A., Sengstag, T., Hughes, G., … Luthi-Carter, R. (2006). Regional and cellular gene expression changes in human Huntington's disease brain. *Human Molecular Genetics*. doi: https://doi.org/10.1093/hmg/ddl013.

Hoffner, G., Kahlem, P. and Djian, P. (2002). Perinuclear localization of huntingtin as a consequence of its binding to microtubules through an interaction with β-tubulin: Relevance to Huntington's disease. *Journal of Cell Science*.

Holmes, S.E., O'hearn, E., Rosenblatt, A., Callahan, C., Hwang, H.S., Ingersoll-Ashworth, R.G., … Margolis, R.L. (2001). A repeat expansion in the gene encoding junctophilin-3 is associated with Huntington disease-like 2. *Nature Genetics*. doi: https://doi.org/10.1038/ng760.

Holmes, S.E., O'Hearn, E.E., McInnis, M.G., Gorelick-Feldman, D.A., Kleiderlein, J.J., Callahan, C., …

Margolis, R.L. (1999). Expansion of a novel CAG trinucleotide repeat in the 5' region of PPP2R2B is associated with SCA12. *Nature Genetics*. doi: https://doi.org/10.1038/70493.

Hsieh, P. and Yamane, K. (2008). DNA mismatch repair: Molecular mechanism, cancer, and ageing. *Mechanisms of Ageing and Development*. doi: https://doi.org/10.1016/j.mad.2008.02.012.

Hu, S., Zhao, M.-T., Jahanbani, F., Shao, N.-Y., Lee, W.H., Chen, H., … Wu, J.C. (2016). Effects of cellular origin on differentiation of human induced pluripotent stem cell–derived endothelial cells. *JCI Insight*. doi: https://doi.org/10.1172/jci.insight.85558.

Hubert, L., Lin, Y., Dion, V. and Wilson, J.H. (2011). Xpa deficiency reduces CAG trinucleotide repeat instability in neuronal tissues in a mouse model of SCA1. *Human Molecular Genetics*. doi: https://doi.org/10.1093/hmg/ddr421.

Hughes, A.C., Mort, M., Elliston, L., Thomas, R.M., Brooks, S.P., Dunnett, S.B. and Jones, L. (2014). Identification of novel alternative splicing events in the huntingtin gene and assessment of the functional consequences using structural protein homology modelling. *Journal of Molecular Biology*. doi: https://doi.org/10.1016/j.jmb.2013.12.028.

Hung, Y.P., Albeck, J.G., Tantama, M. and Yellen, G. (2011). Imaging cytosolic NADH-NAD + redox state with a genetically encoded fluorescent biosensor. *Cell Metabolism*. doi: https://doi.org/10.1016/j.cmet.2011.08.012.

Huntington, G. (1872). On chorea. *Medical and Surgical Reporter*. doi: https://doi.org/10.1001/archneur.1967.00470270110014.

Huntington Study Group TREND-HD, I. (2009). Randomized controlled trial of ethyl-eicosapentaenoic acid in Huntington disease: The TREND-HD study (Archives of Neurology (2008) 65, 12 (1582-1589)). *Archives of Neurology*.

Ikeda, Y., Dalton, J.C., Moseley, M.L., Gardner, K.L., Bird, T.D., Ashizawa, T., … Ranum, L.P.W. (2004). Spinocerebellar ataxia type 8: Molecular genetic comparisons and haplotype analysis of 37 families with ataxia. *American Journal of Human Genetics*. doi: https://doi.org/10.1086/422014.

Imbert, G., Saudou, F., Yvert, G., Devys, D., Trottier, Y., Garnier, J.M., … Brice, A. (1996). Cloning of the gene for spinooerebellar ataxia 2 reveals a locus with high sensitivity to expanded CAG/glutamine repeats. *Nature Genetics*. doi: https://doi.org/10.1038/ng1196-285.

Iyer, R.R., Pluciennik, A., Napierala, M. and Wells, R.D. (2015). DNA Triplet Repeat Expansion and Mismatch Repair. *Annual Review of Biochemistry*. doi: https://doi.org/10.1146/annurev-biochem-060614-034010.

Jacquet, L., Neueder, A., Földes, G., Karagiannis, P., Hobbs, C., Jolinon, N., … Ilic, D. (2015). Three Huntington's disease specific mutation-carrying human embryonic stem cell lines have stable number of CAG repeats upon In Vitro differentiation into cardiomyocytes. *PLoS ONE*. doi: https://doi.org/10.1371/journal.pone.0126860.

Jarem, D.A., Wilson, N.R. and Delaney, S. (2009). Structure-dependent DNA damage and repair in a trinucleotide repeat sequence. *Biochemistry*. doi: https://doi.org/10.1021/bi9007403.

Javeri, A., Guy Lyons, J., Huang, X.X. and Halliday, G.M. (2011). Downregulation of Cockayne syndrome B protein reduces human 8-oxoguanine DNA glycosylase-1 expression and repair of UV radiation-induced 8-oxo-7,8-dihydro-2'-deoxyguanine. *Cancer Science*. doi: https://doi.org/10.1111/j.1349-7006.2011.02005.x.

Jaworski, A., Rosche, W.A., Gellibolian, R., Kang, S., Shimizu, M., Bowater, R.P., … Wells, R.D. (1995). Mismatch repair in Escherichia coli enhances instability of (CTG)(n) triplet repeats from human hereditary diseases. *Proceedings of the National Academy of Sciences of the United States of America*. doi: https://doi.org/10.1073/pnas.92.24.11019.

Jenkins, B.G., Koroshetz, W.J., Beal, M.F. and Rosen, B.R. (1993). Evidence for irnnairment of energy metabofism in vivo in huntington's disease using localized 1h nmr spectroscopy. *Neurology*.

Jenkins, B.G., Rosas, H.D., Chen, Y.C.I., Makabe, T., Myers, R., MacDonald, M., … Koroshetz, W.J. (1998). 1H NMR spectroscopy studies of Huntington's disease. Correlations with CAG repeat numbers. *Neurology*. doi: https://doi.org/10.1212/WNL.50.5.1357.

Jeon, I., Lee, N., Li, J.Y., Park, I.H., Park, K.S., Moon, J., … Song, J. (2012). Neuronal properties, in vivo effects, and pathology of a Huntington's disease patient-derived induced pluripotent stem cells. *Stem Cells* **30**(9):2054–2062. doi: https://doi.org/10.1002/stem.1135.

Jiang, F. and Doudna, J.A. (2017). CRISPR–Cas9 Structures and Mechanisms. *Annual Review of Biophysics*. doi: https://doi.org/10.1146/annurev-biophys-062215-010822.

Jilani, A., Ramotar, D., Slack, C., Ong, C., Yang, X.M., Scherer, S.W. and Lasko, D.D. (1999). Molecular cloning of the human gene, PNKP, encoding a polynucleotide kinase 3'-phosphatase and evidence for its role in repair of DNA strand breaks caused by oxidative damage. *Journal of Biological Chemistry*. doi: https://doi.org/10.1074/jbc.274.34.24176.

Jin, H. and Cho, Y. (2017). Structural and functional relationships of FAN1. *DNA Repair*.

Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J.A. and Charpentier, E. (2012). A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science*. doi: https://doi.org/10.1126/science.1225829.

Jiricny, J. (2006a). MutLα: At the Cutting Edge of Mismatch Repair. *Cell*.

Jiricny, J. (2006b). The multifaceted mismatch-repair system. *Nature Reviews Molecular Cell Biology*.

Jodice, C., Mantuano, E., Veneziano, L., Trettel, F., Sabbadini, G., Calandriello, L., … Frontali, M. (1997). Episodic ataxia type 2 (EA2) and spinocerebellar ataxia type 6 (SCA6) due to CAG repeat expansion in the CACNA1A gene on chromosome 19p. *Human Molecular Genetics*. doi: https://doi.org/10.1093/hmg/6.11.1973.

Johri, A. and Beal, M.F. (2012). Antioxidants in Huntington's disease. *Biochimica et Biophysica Acta - Molecular Basis of Disease*.

Jou, Y.S. and Myers, R.M. (1995). Evidence from antibody studies that the CAG repeat in the huntington disease gene is expressed in the protein. *Human Molecular Genetics*. doi: https://doi.org/10.1093/hmg/4.3.465.

Juopperi, T.A., Kim, W.R., Chiang, C.H., Yu, H., Margolis, R.L., Ross, C.A., … Song, H. (2012). Astrocytes generated from patient induced pluripotent stem cells recapitulate features of Huntingtons disease patient cells. *Molecular Brain*. doi: https://doi.org/10.1186/1756-6606-5-17.

Kameshita, I., Matsuda, Z., Taniguchi, T. and Shizuta, Y. (1984). Poly(ADP-ribose) synthetase. Separation and identification of three proteolytic fragments as the substrate-binding domain, the DNA-binding domain, and the automodification domain. *Journal of Biological Chemistry*.

Kanth, P., Grimmett, J., Champine, M., Burt, R. and Samadder, N.J. (2017). Hereditary Colorectal Polyposis and Cancer Syndromes: A Primer on Diagnosis and Management. *American Journal of Gastroenterology*.

Karimi-Busheri, F., Daly, G., Robins, P., Canas, B., Pappin, D.J.C., Sgouros, J., … Weinfeld, M. (1999). Molecular characterization of a human DNA kinase. *Journal of Biological Chemistry*. doi: https://doi.org/10.1074/jbc.274.34.24187.

Kassubek, J., Juengling, F.D., Kioschies, T., Henkel, K., Karitzky, J., Kramer, B., … Landwehrmeyer, G.B. (2004). Topography of cerebral atrophy in early Huntington's disease: A voxel based morphometric MRI study. *Journal of Neurology, Neurosurgery and Psychiatry*.

Kato, N., Kawasoe, Y., Williams, H., Coates, E., Roy, U., Shi, Y., … Gautier, J. (2017). Sensing and Processing of DNA Interstrand Crosslinks by the Mismatch Repair Pathway. *Cell Reports*. doi: https://doi.org/10.1016/j.celrep.2017.10.032.

Keefe, K.A. and Gerfen, C.R. (1995). D1-D2 dopamine receptor synergy in striatum: effects of intrastriatal infusions of dopamine agonists and antagonists on immediate early gene expression.

Neuroscience. doi: https://doi.org/10.1016/0306-4522(95)00024-D.

Kegel, K.B., Meloni, A.R., Yi, Y., Kim, Y.J., Doyle, E., Cuiffo, B.G., … Difiglia, M. (2002). Huntingtin is present in the nucleus, interacts with the transcriptional corepressor C-terminal binding protein, and represses transcription. *Journal of Biological Chemistry*. doi: https://doi.org/10.1074/jbc.M103946200.

Kemp, J.M. and Powell, T.P. (1971). The structure of the caudate nucleus of the cat: light and electron microscopy. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*. doi: https://doi.org/10.1098/rstb.1971.0102.

Kemp, P.J., Rushton, D.J., Yarova, P.L., Schnell, C., Geater, C., Hancock, J.M., … Telezhkin, V. (2016). Improving and accelerating the differentiation and functional maturation of human stem cell-derived neurons: role of extracellular calcium and GABA. *Journal of Physiology*. doi: https://doi.org/10.1113/JP270655.

Kennedy, L. (2000). Dramatic mutation instability in HD mouse striatum: does polyglutamine load contribute to cell-specific vulnerability in Huntington's disease? *Human Molecular Genetics*. doi: https://doi.org/10.1093/hmg/9.17.2539.

Kennedy, L., Evans, E., Chen, C.M., Craven, L., Detloff, P.J., Ennis, M. and Shelbourne, P.F. (2003). Dramatic tissue-specific mutation length increases are an early molecular event in Huntington disease pathogenesis. *Human Molecular Genetics* **12**(24):3359–3367. doi: https://doi.org/10.1093/hmg/ddg352.

Kim, D., Kim, C.H., Moon, J. Il, Chung, Y.G., Chang, M.Y., Han, B.S., … Kim, K.S. (2009). Generation of Human Induced Pluripotent Stem Cells by Direct Delivery of Reprogramming Proteins. *Cell Stem Cell*.

Kim, M.W., Chelliah, Y., Kim, S.W., Otwinowski, Z. and Bezprozvanny, I. (2009). Secondary Structure of Huntingtin Amino-Terminal Region. *Structure*. doi: https://doi.org/10.1016/j.str.2009.08.002.

Kim, S.H., Lu, H.F. and Alano, C.C. (2011). Neuronal sirt3 protects against excitotoxic injury in mouse cortical neuron culture. *PLoS ONE*. doi: https://doi.org/10.1371/journal.pone.0014731.

Kirkwood, S.C., Siemers, E., Hodes, M.E., Conneally, P.M., Christian, J.C. and Foroud, T. (2000). Subtle changes among presymptomatic carriers of the Huntington's disease gene. *Journal of Neurology Neurosurgery and Psychiatry*. doi: https://doi.org/10.1136/jnnp.69.6.773.

Kobayashi, K., Matsushima, M., Koi, S., Saito, H., Sagae, S., Kudo, R. and Nakamura, Y. (1996). Mutational analysis of mismatch repair genes, hMLH1 and hMSH2, in sporadic endometrial carcinomas with microsatellite instability. *Japanese Journal of Cancer Research*. doi: https://doi.org/10.1111/j.1349-7006.1996.tb03151.x.

Koide, R., Ikeuchi, T., Onodera, O., Tanaka, H., Igarashi, S., Endo, K., … Tsuji, S. (1994). Unstable expansion of CAG repeat in hereditary dentatorubral–pallidoluysian atrophy (DRPLA). *Nature Genetics*. doi: https://doi.org/10.1038/ng0194-9.

Kong, X., Wang, R., Xue, Y., Liu, X., Zhang, H., Chen, Y., … Chang, Y. (2010). Sirtuin 3, a new target of PGC-1α, plays an important role in the suppression of ROS and mitochondrial biogenesis. *PLoS ONE*. doi: https://doi.org/10.1371/journal.pone.0011707.

Koshy, B.T. and Zoghbi, H.Y. (1997). The CAG/polyglutamine tract diseases: Gene products and molecular pathogenesis. *Brain Pathology*.

Kovalenko, M., Dragileva, E., St. Claire, J., Gillis, T., Guide, J.R., New, J., … Wheeler, V.C. (2012). Msh2 Acts in Medium-Spiny Striatal Neurons as an Enhancer of CAG Instability and Mutant Huntingtin Phenotypes in Huntington's Disease Knock-In Mice. *PLoS ONE*. doi: https://doi.org/10.1371/journal.pone.0044273.

Kovtun, I. V., Liu, Y., Bjoras, M., Klungland, A., Wilson, S.H. and McMurray, C.T. (2007). OGG1 initiates age-dependent CAG trinucleotide expansion in somatic cells. *Nature*. doi: https://doi.org/10.1038/nature05778.

Kovtun, I. V. and McMurray, C.T. (2001). Trinucleotide expansion in haploid germ cells by gap repair. *Nature Genetics*. doi: https://doi.org/10.1038/86906.

Kovtun, I. V. and McMurray, C.T. (2006). Models of repair underlying trinucleotide DNA expansion. *Genetic Instabilities and Neurological Diseases, Second Edition*.

Kovtun, I. V. and McMurray, C.T. (2008). Features of trinucleotide repeat instability in vivo. *Cell Research*.

Koyuncu, S., Saez, I., Lee, H.J., Gutierrez-Garcia, R., Pokrzywa, W., Fatima, A., … Vilchez, D. (2018). The ubiquitin ligase UBR5 suppresses proteostasis collapse in pluripotent stem cells from Huntington's disease patients. *Nature Communications*. doi: https://doi.org/10.1038/s41467-018-05320-3.

Krainc, D. (2006). Transcriptional Abnormalities in Huntington's Disease. *Transcription Factors in the Nervous System: Development, Brain Function, and Diseases*.

Kratz, K., Schöpf, B., Kaden, S., Sendoel, A., Eberhard, R., Lademann, C., … Jiricny, J. (2010). Deficiency of FANCD2-Associated Nuclease KIAA1018/FAN1 Sensitizes Cells to Interstrand Crosslinking Agents. *Cell*. doi: https://doi.org/10.1016/j.cell.2010.06.022.

Kremer, B., Squitieri, F., Telenius, H., Andrew, S.E., Theilmann, J., Spence, N., … Hayden, M.R. (1993). Molecular analysis of late onset Huntington's disease. *Journal of Medical Genetics*.

Krieger, C. and Duchen, M.R. (2002). Mitochondria, Ca2+ and neurodegenerative disease. *European Journal of Pharmacology*. doi: https://doi.org/10.1016/S0014-2999(02)01842-3.

Kuo, L.J. and Yang, L.-X.X. (2008). γ-H2AX- A novel biomaker for DNA double-strand breaks. *In Vivo*. doi: https://doi.org/0258-851X/2008.

Kupfer, S.S. and Ellis, N.A. (2016). Hereditary colorectal cancer. *The Molecular Basis of Human Cancer*.

Kuwert, T., Lange, H.W., Langen, K.J., Herzog, H., Aulich, A. and Feinendegen, L.E. (1990). Cortical and subcortical glucose consumption measured by pet in patients with huntington's disease. *Brain*. doi: https://doi.org/10.1093/brain/113.5.1405.

Lachaud, C., Moreno, A., Marchesi, F., Toth, R., Blow, J.J. and Rouse, J. (2016). Ubiquitinated Fancd2 recruits Fan1 to stalled replication forks to prevent genome instability. *Science*. doi: https://doi.org/10.1126/science.aad5634.

Lai, Y., Budworth, H., Beaver, J.M., Chan, N.L.S., Zhang, Z., McMurray, C.T. and Liu, Y. (2016). Crosstalk between MSH2-MSH3 and polβ promotes trinucleotide repeat expansion during base excision repair. *Nature Communications*. doi: https://doi.org/10.1038/ncomms12465.

Lajoie, P. and Snapp, E.L. (2010). Formation and toxicity of soluble polyglutamine oligomers in living cells. *PLoS ONE*. doi: https://doi.org/10.1371/journal.pone.0015245.

Landwehrmeyer, G.B., Fitzer-Attas, C.J., Giuliano, J.D., Gonçalves, N., Anderson, K.E., Cardoso, F., … Sampaio, C. (2017). Data Analytics from Enroll-HD, a Global Clinical Research Platform for Huntington's Disease. *Movement Disorders Clinical Practice*. doi: https://doi.org/10.1002/mdc3.12388.

Lang, W.H., Coats, J.E., Majka, J., Hura, G.L., Lin, Y., Rasnik, I. and McMurray, C.T. (2011). Conformational trapping of mismatch recognition complex MSH2/MSH3 on repair-resistant DNA loops. *Proceedings of the National Academy of Sciences of the United States of America*. doi: https://doi.org/10.1073/pnas.1105461108.

Langbein, S., Szakacs, O., Wilhelm, M., Sukosd, F., Weber, S., Jauch, A., … Kovacs, G. (2002). Alteration of the LRP1B gene region is associated with high grade of urothelial cancer. *Laboratory Investigation*. doi: https://doi.org/10.1038/labinvest.3780458.

Laurent, L.C., Ulitsky, I., Slavin, I., Tran, H., Schork, A., Morey, R., … Loring, J.F. (2011). Dynamic changes in the copy number of pluripotency and cell proliferation genes in human ESCs and iPSCs during reprogramming and time in culture. *Cell Stem Cell*. doi: https://doi.org/10.1016/j.stem.2010.12.003.

Leavitt, B.R., Guttman, J.A., Hodgson, J.G., Kimel, G.H., Singaraja, R., Vogl, A.W. and Hayden, M.R. (2001). Wild-type Huntingtin reduces the cellular toxicity of mutant Huntingtin in vivo. *American Journal of Human Genetics*. doi: https://doi.org/10.1086/318207.

Leavitt, B.R., Van Raamsdonk, J.M., Shehadeh, J., Fernandes, H., Murphy, Z., Graham, R.K., … Hayden, M.R. (2006). Wild-type huntingtin protects neurons from excitotoxicity. *Journal of Neurochemistry*. doi: https://doi.org/10.1111/j.1471-4159.2005.03605.x.

Lee, J.M., Chao, M.J., Harold, D., Elneel, K.A., Gillis, T., Holmans, P., … Gusella, J.F. (2017). A modifier of Huntington's disease onset at the MLH1 locus. *Human Molecular Genetics*. doi: https://doi.org/10.1093/hmg/ddx286.

Lee, J.M., Pinto, R.M., Gillis, T., St. Claire, J.C. and Wheeler, V.C. (2011). Quantification of age-dependent somatic CAG repeat instability in Hdh CAG knock-in mice reveals different expansion dynamics in striatum and liver. *PLoS ONE*. doi: https://doi.org/10.1371/journal.pone.0023647.

Lee, J.M., Ramos, E.M., Lee, J.H., Gillis, T., Mysore, J.S., Hayden, M.R., … Gusella, J.F. (2012). CAG repeat expansion in Huntington disease determines age at onset in a fully dominant fashion. *Neurology* **78**(10):690–695. doi: https://doi.org/10.1212/WNL.0b013e318249f683.

Lee, J.M., Zhang, J., Su, A.I., Walker, J.R., Wiltshire, T., Kang, K., … Wheeler, V.C. (2010). HA novel approach to investigate tissue-specific trinucleotide repeat instability. *BMC Systems Biology*. doi: https://doi.org/10.1186/1752-0509-4-29.

Lee, W.C.M., Yoshihara, M. and Littleton, J.T. (2004). Cytoplasmic aggregates trap polyglutamine-containing proteins and block axonal transport in a Drosophila model of Huntington's disease. *Proceedings of the National Academy of Sciences of the United States of America*. doi: https://doi.org/10.1073/pnas.0400243101.

Leeflang, E., Zhang, L., Tavare, S., Hubert, R., Srinidhi, J., E.Macdonald, M., … Arnheim, N. (1995). Single sperm analysis of the trinucleotide repeats in the huntington's disease gene: Quantification of the mutation frequency spectrum. *Human Molecular Genetics*. doi: https://doi.org/10.1093/hmg/4.9.1519.

Leitman, J., Ulrich Hartl, F. and Lederkremer, G.Z. (2013). Soluble forms of polyQ-expanded huntingtin rather than large aggregates cause endoplasmic reticulum stress. *Nature Communications*. doi: https://doi.org/10.1038/ncomms3753.

Lendahl, U., Zimmerman, L.B. and McKay, R.D.G. (1990). CNS stem cells express a new class of intermediate filament protein. *Cell*. doi: https://doi.org/10.1016/0092-8674(90)90662-X.

Li, H., Li, S.H., Yu, Z.X., Shelbourne, P. and Li, X.J. (2001). Huntingtin aggregate-associated axonal

degeneration is an early pathological event in Huntington's disease mice. *Journal of Neuroscience*.

Li, J.L., Hayden, M.R., Almqvist, E.W., Brinkman, R.R., Durr, A., Dodé, C., … Myers, R.H. (2003). A genome scan for modifiers of age at onset in Huntington disease: The HD MAPS study. *American Journal of Human Genetics*. doi: https://doi.org/10.1086/378133.

Li, S.H. and Li, X.J. (2004). Huntingtin-protein interactions and the pathogenesis of Huntington's disease. *Trends in Genetics*.

Liang, X., Potter, J., Kumar, S., Ravinder, N. and Chesnut, J.D. (2017). Enhanced CRISPR/Cas9-mediated precise genome editing by improved design and delivery of gRNA, Cas9 nuclease, and donor DNA. *Journal of Biotechnology*. doi: https://doi.org/10.1016/j.jbiotec.2016.11.011.

Lin, B., Rommens, J.M., Graham, R.K., Kalchman, M., Macdonald, H., Nasir, J., … Hayden, M.R. (1993). Differential 3' polyadenylation of the huntington disease gene results in two mRNA species with variable tissue expression. *Human Molecular Genetics*. doi: https://doi.org/10.1093/hmg/2.10.1541.

Lin, Y., Dent, S.Y.R., Wilson, J.H., Wells, R.D. and Napierala, M. (2010). R loops stimulate genetic instability of CTG·CAG repeats. *Proceedings of the National Academy of Sciences of the United States of America*. doi: https://doi.org/10.1073/pnas.0909740107.

Lin, Y., Dion, V. and Wilson, J.H. (2006). Transcription promotes contraction of CAG repeat tracts in human cells. *Nature Structural and Molecular Biology*. doi: https://doi.org/10.1038/nsmb1042.

Liot, G., Bossy, B., Lubitz, S., Kushnareva, Y., Sejbuk, N. and Bossy-Wetzel, E. (2009). Complex II inhibition by 3-NP causes mitochondrial fragmentation and neuronal cell death via an NMDA- and ROS-dependent pathway. *Cell Death and Differentiation*. doi: https://doi.org/10.1038/cdd.2009.22.

Liquori, C.L., Ricker, K., Moseley, M.L., Jacobsen, J.F., Kress, W., Naylor, S.L., … Ranum, L.P.W. (2001). Myotonic dystrophy type 2 caused by a CCTG expansion in intron I of ZNF9. *Science*. doi: https://doi.org/10.1126/science.1062125.

Liu, C.S., Cheng, W.L., Kuo, S.J., Li, J.Y., Soong, B.W. and Wei, Y.H. (2008). Depletion of mitochondrial DNA in leukocytes of patients with poly-Q diseases. *Journal of the Neurological Sciences*. doi: https://doi.org/10.1016/j.jns.2007.07.016.

Liu, T., Ghosal, G., Yuan, J., Chen, J. and Huang, J. (2010). FAN1 acts with FANCI-FANCD2 to promote DNA interstrand cross-link repair. *Science* 329(5992):693–696. doi: https://doi.org/10.1126/science.1192656.

Liu, Z., Zhou, T., Ziegler, A.C., Dimitrion, P. and Zuo, L. (2017). Oxidative Stress in Neurodegenerative Diseases: From Molecular Mechanisms to Clinical Applications. *Oxidative Medicine and Cellular*

Longevity.

Lodi, R., Schapira, A.H.V., Manners, D., Styles, P., Wood, N.W., Taylor, D.J. and Warner, T.T. (2000). Abnormal in vivo skeletal muscle energy metabolism in Huntington's disease and dentatorubropallidoluysian atrophy. *Annals of Neurology*. doi: https://doi.org/10.1002/1531-8249(200007)48:1<72::AID-ANA11>3.0.CO;2-I.

Lokanga, R.A., Entezam, A., Kumari, D., Yudkin, D., Qin, M., Smith, C.B. and Usdin, K. (2013). Somatic Expansion in Mouse and Human Carriers of Fragile X Premutation Alleles. *Human Mutation*. doi: https://doi.org/10.1002/humu.22177.

Lokanga, R.A., Senejani, A.G., Sweasy, J.B. and Usdin, K. (2015). Heterozygosity for a Hypomorphic Polβ Mutation Reduces the Expansion Frequency in a Mouse Model of the Fragile X-Related Disorders. *PLoS Genetics*. doi: https://doi.org/10.1371/journal.pgen.1005181.

Lu, T., Pan, Y., Kao, S.Y., Li, C., Kohane, I., Chan, J. and Yankner, B.A. (2004). Gene regulation and DNA damage in the ageing human brain. *Nature*. doi: https://doi.org/10.1038/nature02661.

Lu, X.H., Mattis, V.B., Wang, N., Al-Ramahi, I., Van Den Berg, N., Fratantoni, S.A., … Yang, X.W. (2014). Targeting ATM ameliorates mutant Huntingtin toxicity in cell and animal models of Huntington's disease. *Science Translational Medicine*. doi: https://doi.org/10.1126/scitranslmed.3010523.

Ludolph, A.C., he, F., Spencer, P.S., Hammerstad, J. and Sabri, M. (1991). 3-Nitropropionic Acid - Exogenous Animal Neurotoxin and Possible Human Striatal Toxin. *Canadian Journal of Neurological Sciences / Journal Canadien des Sciences Neurologiques*.

MacDonald, M.E., Ambrose, C.M., Duyao, M.P., Myers, R.H., Lin, C., Srinidhi, L., … Harper, P.S. (1993). A novel gene containing a trinucleotide repeat that is expanded and unstable on Huntington's disease chromosomes. *Cell* **72**(6):971–983. doi: https://doi.org/10.1016/0092-8674(93)90585-E.

MacKay, C., Déclais, A.C., Lundin, C., Agostinho, A., Deans, A.J., MacArtney, T.J., … Rouse, J. (2010). Identification of KIAA1018/FAN1, a DNA Repair Nuclease Recruited to DNA Damage by Monoubiquitinated FANCD2. *Cell*. doi: https://doi.org/10.1016/j.cell.2010.06.021.

Maguire, J.J., Kagan, V.E. and Packer, L. (1992). Electron transport between cytochrome c and alpha tocopherol. *Biochemical and Biophysical Research Communications*. doi: https://doi.org/10.1016/0006-291X(92)92368-8.

Maiuri, T., Mocle, A.J., Hung, C.L., Xia, J., van Roon-Mom, W.M.C. and Truant, R. (2017). Huntingtin is a scaffolding protein in the ATM oxidative DNA damage response complex. *Human Molecular Genetics*. doi: https://doi.org/10.1093/hmg/ddw395.

Maiuri, T., Woloshansky, T., Xia, J. and Truant, R. (2013). The huntingtin N17 domain is a multifunctional CRM1 and ran-dependent nuclear and cilial export signal. *Human Molecular Genetics*. doi: https://doi.org/10.1093/hmg/dds554.

Mandir, A.S., Simbulan-Rosenthal, C.M., Poitras, M.F., Lumpkin, J.R., Dawson, V.L., Smulson, M.E. and Dawson, T.M. (2002). A novel in vivo post-translational modification of p53 by PARP-1 in MPTP-induced parkinsonism. *Journal of Neurochemistry*. doi: https://doi.org/10.1046/j.1471-4159.2002.01144.x.

Mangiarini, L., Sathasivam, K., Mahal, A., Mott, R., Seller, M. and Bates, G.P. (1997). Instability of highly expanded CAG repeats in mice transgenic for the Huntington's disease mutation. *Nature Genetics*. doi: https://doi.org/10.1038/ng0297-197.

Mangiarini, L., Sathasivam, K., Seller, M., Cozens, B., Harper, A., Hetherington, C., … Bates, G.P. (1996). Exon I of the HD gene with an expanded CAG repeat is sufficient to cause a progressive neurological phenotype in transgenic mice. *Cell*. doi: https://doi.org/10.1016/S0092-8674(00)81369-0.

Manley, K., Pugh, J. and Messer, A. (1999). Instability of the CAG repeat in immortalized fibroblast cell cultures from Huntington's disease transgenic mice. *Brain Research*.

Manley, K., Shirley, T.L., Flaherty, L. and Messer, A. (1999). Msh2 deficiency prevents in vivo somatic instability of the CAG repeat in Huntington disease transgenic mice. *Nature Genetics*. doi: https://doi.org/10.1038/70598.

Marcellin, D., Abramowski, D., Young, D., Richter, J., Weiss, A., Marcel, A., … Lotz, G.P. (2012). Fragments of HdhQ150 Mutant Huntingtin Form a Soluble Oligomer Pool That Declines with Aggregate Deposition upon Aging. *PLoS ONE*. doi: https://doi.org/10.1371/journal.pone.0044457.

De Mario, A., Scarlatti, C., Costiniti, V., Primerano, S., Lopreiato, R., Calì1, T., … Carafoli, E. (2016). Calcium handling by endoplasmic reticulum and mitochondria in a cell model of huntington's disease. *PLoS Currents*. doi: https://doi.org/10.1371/currents.hd.37fcb1c9a27503dc845594ee4a7316c3.

Marquis Gacy, A., Goellner, G., Juranić, N., Macura, S. and McMurray, C.T. (1995). Trinucleotide repeats that expand in human disease form hairpin structures in vitro. *Cell*. doi: https://doi.org/10.1016/0092-8674(95)90074-8.

Martins, S., Pearson, C.E., Coutinho, P., Provost, S., Amorim, A., Dubé, M.P., … Rouleau, G.A. (2014). Modifiers of (CAG)(n) instability in Machado-Joseph disease (MJD/SCA3) transmissions: an association study with DNA replication, repair and recombination genes. *Human genetics*. doi: https://doi.org/10.1007/s00439-014-1467-8.

Matoba, S., Kang, J.G., Patino, W.D., Wragg, A., Boehm, M., Gavrilova, O., … Hwang, P.M. (2006). p53

regulates mitochondrial respiration. *Science*. doi: https://doi.org/10.1126/science.1126863.

Mattis, V.B. and Svendsen, C.N. (2017). Modeling Huntington's disease with patient-derived neurons. *Brain Research* 1656:76–87.

Mattis, V.B., Tom, C., Akimov, S., Saeedian, J., Østergaard, M.E., Southwell, A.L., … Svendsen, C.N. (2014). HD iPSC-derived neural progenitors accumulate in culture and are susceptible to BDNF withdrawal due to glutamate toxicity. *Human Molecular Genetics*. doi: https://doi.org/10.1093/hmg/ddv080.

Mattson, M.P., Gleichmann, M. and Cheng, A. (2008). Mitochondria in Neuroplasticity and Neurological Disorders. *Neuron*.

Mazziotta, J.C., Phelps, M.E., Pahl, J.J., Huang, S.C., Baxter, L.R., Riege, W.H., … Markham, C.H. (1987). Reduced Cerebral Glucose Metabolism in Asymptomatic Subjects at Risk for Huntington's Disease. *New England Journal of Medicine*. doi: https://doi.org/10.1056/NEJM198702123160701.

McCampbell, A. (2000). CREB-binding protein sequestration by expanded polyglutamine. *Human Molecular Genetics*. doi: https://doi.org/10.1093/hmg/9.14.2197.

McColgan, P. and Tabrizi, S.J. (2018). Huntington's disease: a clinical review. *European Journal of Neurology*.

McGinty, R.J. and Mirkin, S.M. (2018). Cis- and Trans-Modifiers of Repeat Expansions: Blending Model Systems with Human Genetics. *Trends in Genetics*.

McMeekin, L.J., Li, Ye, Fox, S.N., Rowe, G.C., Crossman, D.K., Day, J.J., … Cowell, Rita M. (2018). Cell-specific deletion of PGC-1α from medium spiny neurons causes transcriptional alterations and age-related motor impairment. *Journal of Neuroscience*. doi: https://doi.org/10.1523/JNEUROSCI.0848-17.2018.

McMurray, C.T. (2008). Hijacking of the mismatch repair system to cause CAG expansion and cell death in neurodegenerative disease. *DNA Repair*. doi: https://doi.org/10.1016/j.dnarep.2008.03.013.

McMurray, C.T. (2010). Mechanisms of trinucleotide repeat instability during human development. *Nat Rev Genet* 11(11):786–799.

McQuade, L.R., Balachandran, A., Scott, H.A., Khaira, S., Baker, M.S. and Schmidt, U. (2014). Proteomics of Huntington's disease-affected human embryonic stem cells reveals an evolving pathology involving mitochondrial dysfunction and metabolic disturbances. *Journal of Proteome Research*. doi: https://doi.org/10.1021/pr500649m.

van der Meer, L.B., van Duijn, E., Wolterbeek, R. and Tibben, A. (2012). Adverse childhood

experiences of persons at risk for Huntington's disease or BRCA1/2 hereditary breast/ovarian cancer. *Clinical Genetics*. doi: https://doi.org/10.1111/j.1399-0004.2011.01778.x.

Mehta, S.R., Tom, C.M., Wang, Y., Bresee, C., Rushton, D., Mathkar, P.P., … Mattis, V.B. (2018). Human Huntington's Disease iPSC-Derived Cortical Neurons Display Altered Transcriptomics, Morphology, and Maturation. *Cell Reports*. doi: https://doi.org/10.1016/j.celrep.2018.09.076.

Mejia, E.M., Chau, S., Sparagna, G.C., Sipione, S. and Hatch, G.M. (2016). Reduced Mitochondrial Function in Human Huntington Disease Lymphoblasts is Not Due to Alterations in Cardiolipin Metabolism or Mitochondrial Supercomplex Assembly. *Lipids*. doi: https://doi.org/10.1007/s11745-015-4110-0.

Metsu, S., Rainger, J.K., Debacker, K., Bernhard, B., Rooms, L., Grafodatskaya, D., … Fitzpatrick, D.R. (2014). A CGG-repeat expansion mutation in ZNF713 causes FRA7A: Association with autistic spectrum disorder in two families. *Human Mutation*. doi: https://doi.org/10.1002/humu.22683.

Milakovic, T., Quintanilla, R.A. and Johnson, G.V.W. (2006). Mutant Huntingtin expression induces mitochondrial calcium handling defects in clonal striatal cells: Functional consequences. *Journal of Biological Chemistry*. doi: https://doi.org/10.1074/jbc.M603845200.

Mirkin, S.M. (2007). Expandable DNA repeats and human disease. *Nature*.

Mirkin, S.M. and Frank-Kamenetskii, M.D. (1994). H-DNA and related structures. *Annual Review of Biophysics and Biomolecular Structure*.

Miyoshi, N., Ishii, H., Nagano, H., Haraguchi, N., Dewi, D.L., Kano, Y., … Mori, M. (2011). Reprogramming of mouse and human cells to pluripotency using mature microRNAs. *Cell Stem Cell*. doi: https://doi.org/10.1016/j.stem.2011.05.001.

Mochmann, L.H. and Wells, R.D. (2004). Transcription influences the types of deletion and expansion products in an orientation-dependent manner from GAC·GTC repeats. *Nucleic Acids Research*. doi: https://doi.org/10.1093/nar/gkh787.

Møllersen, L., Rowe, A.D., Illuzzi, J.L., Hildrestrand, G.A., Gerhold, K.J., Tveterås, L., … Klungland, A. (2012). Neil1 is a genetic modifier of somatic and germline CAG trinucleotide repeat instability in R6/1 mice. *Human Molecular Genetics*. doi: https://doi.org/10.1093/hmg/dds337.

Monckton, D.G., Cayuela, M.L., Gould, F.K., Brock, G.J.R., De Silva, R. and Ashizawa, T. (1999). Very large (CAG)(n) DNA repeat expansions in the sperm of two spinocerebellar ataxia type 7 males. *Human Molecular Genetics*. doi: https://doi.org/10.1093/hmg/8.13.2473.

Morris, M. (1991). Psychiatric aspects of Huntington's disease. *Huntington's Disease* **22**:pp 81-126.

Moseley, M.L., Schut, L.J., Bird, T.D., Day, J.W. and Ranum, L.P.W. (2000). Large, expanded repeats in sca8 are not confined to patients with cerebellar ataxia: Reply. *Nature Genetics*. doi: https://doi.org/10.1038/73415.

Moss, D.J.H., Tabrizi, S.J., Mead, S., Lo, K., Pardiñas, A.F., Holmans, P., … Tan, L. (2017). Identification of genetic variants associated with Huntington's disease progression: a genome-wide association study. *The Lancet Neurology*. doi: https://doi.org/10.1016/S1474-4422(17)30161-8.

Muftuoglu, M., de Souza-Pinto, N.C., Dogan, A., Aamann, M., Stevnsner, T., Rybanska, I., … Bohr, V.A. (2009). Cockayne syndrome group B protein stimulates repair of formamidopyrimidines by NEIL1 DNA glycosylase. *Journal of Biological Chemistry*. doi: https://doi.org/10.1074/jbc.M807006200.

Munsie, L., Caron, N., Atwal, R.S., Marsden, I., Wild, E.J., Bamburg, J.R., … Truant, R. (2011). Mutant huntingtin causes defective actin remodeling during stress: Defining a new role for transglutaminase 2 in neurodegenerative disease. *Human Molecular Genetics*. doi: https://doi.org/10.1093/hmg/ddr075.

Myers, R.H., Macdonald, M.E., Koroshetz, W.J., Duyao, M.P., Ambrose, C.M., Taylor, S.A.M., … Gusella, J.F. (1993). De novo expansion of a (CAG)n repeat in sporadic Huntington's disease. *Nature Genetics*. doi: https://doi.org/10.1038/ng1093-168.

Nadel, J., Athanasiadou, R., Lemetre, C., Wijetunga, N.A., Ó Broin, P., Sato, H., … Greally, J.M. (2015). RNA:DNA hybrids in the human genome have distinctive nucleotide characteristics, chromatin composition, and transcriptional relationships. *Epigenetics and Chromatin*. doi: https://doi.org/10.1186/s13072-015-0040-6.

Nagafuchi, S., Yanagisawa, H., Sato, K., Shirayama, T., Ohsaki, E., Bundo, M., … Yamada, M. (1994). Dentatorubral and pallidoluysian atrophy expansion of an unstable CAG trinucleotide on chromosome 12p. *Nature Genetics*. doi: https://doi.org/10.1038/ng0194-14.

Nahhas, F.A., Garbern, J., Krajewski, K.M., Roa, B.B. and Feldman, G.L. (2005). Juvenile onset Huntington disease resulting from a very large maternal expansion. *American Journal of Medical Genetics*. doi: https://doi.org/10.1002/ajmg.a.30891.

Nakamori, M., Pearson, C.E. and Thornton, C.A. (2011). Bidirectional transcription stimulates expansion and contraction of expanded (CTG)•(CAG) repeats. *Human Molecular Genetics*. doi: https://doi.org/10.1093/hmg/ddq501.

Nakamura, K. (2001). SCA17, a novel autosomal dominant cerebellar ataxia caused by an expanded polyglutamine in TATA-binding protein. *Human Molecular Genetics*. doi: https://doi.org/10.1093/hmg/10.14.1441.

Napoli, E., Wong, S., Hung, C., Ross-Inta, C., Bomdica, P. and Giulivi, C. (2013). Defective mitochondrial

disulfide relay system, altered mitochondrial morphology and function in Huntington's disease. *Human Molecular Genetics*. doi: https://doi.org/10.1093/hmg/dds503.

Nasir, J., Floresco, S.B., O'Kusky, J.R., Diewert, V.M., Richman, J.M., Zeisler, J., … Hayden, M.R. (1995). Targeted disruption of the Huntington's disease gene results in embryonic lethality and behavioral and morphological changes in heterozygotes. *Cell*. doi: https://doi.org/10.1016/0092-8674(95)90542-1.

Nekrasov, E.D., Vigont, V.A., Klyushnikov, S.A., Lebedeva, O.S., Vassina, E.M., Bogomazova, A.N., … Kiselev, S.L. (2016). Manifestation of Huntington's disease pathology in human induced pluripotent stem cell-derived neurons. *Molecular Neurodegeneration*. doi: https://doi.org/10.1186/s13024-016-0092-5.

Neueder, A., Landles, C., Ghosh, R., Howland, D., Myers, R.H., Faull, R.L.M., … Bates, G.P. (2017). The pathogenic exon 1 HTT protein is produced by incomplete splicing in Huntington's disease patients. *Scientific Reports*. doi: https://doi.org/10.1038/s41598-017-01510-z.

Ni, S., Hu, J., Duan, Y., Shi, S., Li, R., Wu, H., … Li, Y. (2013). Down expression of LRP1B promotes cell migration via RhoA/Cdc42 pathway and actin cytoskeleton remodeling in renal cell cancer. *Cancer Science*. doi: https://doi.org/10.1111/cas.12157.

Niclis, J.C., Ellisdon, A.M., Cram, D.S., Trounson, A.O., Dottori, M., Bottomley, S.P. and Verlinsky, Y. (2009). Human embryonic stem cell models of Huntington disease. *Reproductive biomedicine online* **19**(1):106–13. doi: https://doi.org/10.1016/S1472-6483(10)60053-3.

Nishimasu, H. and Nureki, O. (2017). Structures and mechanisms of CRISPR RNA-guided effector nucleases. *Current Opinion in Structural Biology*.

Noll, D.M., McGregor Mason, T. and Miller, P.S. (2006). Formation and repair of interstrand cross-links in DNA. *Chemical Reviews*.

Nucifora, J., Sasaki, M., Peters, M.F., Huang, H., Cooper, J.K., Yamada, M., … Ross, C.A. (2001). Interference by huntingtin and atrophin-1 with CBP-mediated transcription leading to cellular toxicity. *Science*. doi: https://doi.org/10.1126/science.1056784.

O'Kusky, J.R., Nasir, J., Cicchetti, F., Parent, A. and Hayden, M.R. (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 Research*. doi: https://doi.org/10.1016/S0006-8993(98)01312-2.

Okita, K. and Yamanaka, S. (2011). Induced pluripotent stem cells: Opportunities and challenges. *Philosophical Transactions of the Royal Society B: Biological Sciences*.

Oliveira, J.M.A. (2010). Nature and cause of mitochondrial dysfunction in Huntington's disease: Focusing on huntingtin and the striatum. *Journal of Neurochemistry*.

Ooi, J., Langley, S.R., Xu, X., Utami, K.H., Sim, B., Huang, Y., … Pouladi, M.A. (2019). Unbiased Profiling of Isogenic Huntington Disease hPSC-Derived CNS and Peripheral Cells Reveals Strong Cell-Type Specificity of CAG Length Effects. *Cell Reports*. doi: https://doi.org/10.1016/j.celrep.2019.02.008.

Orr, A.L., Li, S., Wang, C.E., Li, H., Wang, J., Rong, J., … Li, X.J. (2008). N-terminal mutant huntingtin associates with mitochondria and impairs mitochondrial trafficking. *Journal of Neuroscience*. doi: https://doi.org/10.1523/JNEUROSCI.0106-08.2008.

Orth, M., Handley, O.J., Schwenke, C., Dunnett, S.B., Craufurd, D., Ho, A.K., … Landwehrmeyer, G.B. (2010). Observing Huntington's disease: The European Huntington's disease network's registry. *PLoS Currents*. doi: https://doi.org/10.1371/currents.RRN1184.

Oussatcheva, E.A., Hashem, V.I., Zou, Y., Sinden, R.R. and Potaman, V.N. (2001). Involvement of the Nucleotide Excision Repair Protein UvrA in Instability of CAG·CTG Repeat Sequences in Escherichia coli. *Journal of Biological Chemistry*. doi: https://doi.org/10.1074/jbc.M104697200.

Owen, B.A.L., Yang, Z., Lai, M., Gajek, M., Badger, J.D., Hayes, J.J., … McMurray, C.T. (2005). (CAG)n-hairpin DNA binds to Msh2-Msh3 and changes properties of mismatch recognition. *Nature Structural and Molecular Biology*. doi: https://doi.org/10.1038/nsmb965.

Paldino, E., Cardinale, A., D'Angelo, V., Sauve, I., Giampà, C. and Fusco, F.R. (2017). Selective sparing of striatal interneurons after poly (ADP-ribose) polymerase 1 inhibition in the R6/2 mouse model of huntington's disease. *Frontiers in Neuroanatomy*. doi: https://doi.org/10.3389/fnana.2017.00061.

Palmer, G., Horgan, D.J., Tisdale, H., Singer, T.P. and Beinert, H. (1968). Studies on the respiratory chain-linked reduced nicotinamide adenine dinucleotide dehydrogenase. XIV. Location of the sites of inhibition of rotenone, barbiturates, and piericidin by means of electron paramagnetic resonance spectroscopy. *Journal of Biological Chemistry*.

Panegyres, P.K., Beilby, J., Bulsara, M., Toufexis, K. and Wong, C. (2006). A study of potential interactive genetic factors in Huntington's disease. *European Neurology*. doi: https://doi.org/10.1159/000093867.

Panov, A. V., Gutekunst, C.A., Leavitt, B.R., Hayden, M.R., Burke, J.R., Strittmatter, W.J. and Greenamyre, J.T. (2002). Early mitochondrial calcium defects in Huntington's disease are a direct effect of polyglutamines. *Nature Neuroscience*. doi: https://doi.org/10.1038/nn884.

Papoutsi, M., Labuschagne, I., Tabrizi, S.J. and Stout, J.C. (2014). The cognitive burden in Huntington's disease: Pathology, phenotype, and mechanisms of compensation. *Movement Disorders*.

Park, H., Miyazaki, H., Yamanaka, T. and Nukina, N. (2018). Non-coding RNA Neat1 and Abhd11os expressions are dysregulated in medium spiny neurons of Huntington disease model mice. *Neuroscience Research*.

Park, I.H., Arora, N., Huo, H., Maherali, N., Ahfeldt, T., Shimamura, A., … Daley, G.Q. (2008). Disease-Specific Induced Pluripotent Stem Cells. *Cell*. doi: https://doi.org/10.1016/j.cell.2008.07.041.

Park, S.H., Kukushkin, Y., Gupta, R., Chen, T., Konagai, A., Hipp, M.S., … Hartl, F.U. (2013). PolyQ proteins interfere with nuclear degradation of cytosolic proteins by sequestering the Sis1p chaperone. *Cell*. doi: https://doi.org/10.1016/j.cell.2013.06.003.

Parsons, J.L. and Edmonds, M.J. (2015). The Base Excision Repair Pathway. *Encyclopedia of Cell Biology*.

Paulsen, J.S., Hayden, M., Stout, J.C., Langbehn, D.R., Aylward, E., Ross, C.A., … Penziner, E. (2006). Preparing for preventive clinical trials: The predict-HD study. *Archives of Neurology*. doi: https://doi.org/10.1001/archneur.63.6.883.

Paulsen, J.S., Langbehn, D.R., Stout, J.C., Aylward, E., Ross, C.A., Nance, M., … Hayden, M. (2008). Detection of Huntington's disease decades before diagnosis: The Predict-HD study. *Journal of Neurology, Neurosurgery and Psychiatry*. doi: https://doi.org/10.1136/jnnp.2007.128728.

Paulsen, J.S., Ready, R.E., Hamilton, J.M., Mega, M.S. and Cummings, J.L. (2001). Neuropsychiatric aspects of Huntington's disease. *Journal of Neurology Neurosurgery and Psychiatry*. doi: https://doi.org/10.1136/jnnp.71.3.310.

Peng, C., Zhu, G., Liu, X. and Li, H. (2018). Mutant Huntingtin Causes a Selective Decrease in the Expression of Synaptic Vesicle Protein 2C. *Neuroscience Bulletin*. doi: https://doi.org/10.1007/s12264-018-0230-x.

Pennell, S., Déclais, A.-C.C., Li, J., Haire, L.F., Berg, W., Saldanha, J.W., … Smerdon, S.J. (2014). FAN1 activity on asymmetric repair intermediates is mediated by an atypical monomeric virus-type replication-repair nuclease domain. *Cell Reports* **8**(1):84–93. doi: https://doi.org/10.1016/j.celrep.2014.06.001.

Pérez-Severiano, F., Santamaría, A., Pedraza-Chaverri, J., Medina-Campos, O.N., Ríos, C. and Segovia, J. (2004). Increased Formation of Reactive Oxygen Species, but No Changes in Glutathione Peroxidase Activity, in Striata of Mice Transgenic for the Huntington's Disease Mutation. *Neurochemical Research*. doi: https://doi.org/10.1023/B:NERE.0000018843.83770.4b.

Perez, E.E., Wang, J., Miller, J.C., Jouvenot, Y., Kim, K.A., Liu, O., … June, C.H. (2008). Establishment of HIV-1 resistance in CD4+ T cells by genome editing using zinc-finger nucleases. *Nature Biotechnology*.

doi: https://doi.org/10.1038/nbt1410.

Peterlin, B., Logar, N. and Zidar, J. (1996). CTG repeat analysis in lymphocytes, muscles and fibroblasts in patients with myotonic dystrophy. *Pflügers Archiv European Journal of Physiology*. doi: https://doi.org/10.1007/BF02346337.

Pflanz, S., Besson, J.A.O., Ebmeier, K.P. and Simpson, S. (1991). The clinical manifestation of mental disorder in Huntington's disease: a retrospective case record study of disease progression. *Acta Psychiatrica Scandinavica*. doi: https://doi.org/10.1111/j.1600-0447.1991.tb05511.x.

Pinto, R.M., Dragileva, E., Kirby, A., Lloret, A., Lopez, E., St. Claire, J., … Wheeler, V.C. (2013). Mismatch Repair Genes Mlh1 and Mlh3 Modify CAG Instability in Huntington's Disease Mice: Genome-Wide and Candidate Approaches. *PLoS Genetics*. doi: https://doi.org/10.1371/journal.pgen.1003930.

Pizzolato, J., Mukherjee, S., Schärer, O.D. and Jiricny, J. (2015). FANCD2-associated nuclease 1, but not exonuclease 1 or flap endonuclease 1, is able to unhook DNA interstrand cross-links in vitro. *Journal of Biological Chemistry*. doi: https://doi.org/10.1074/jbc.M115.663666.

Prajumwongs, P., Weeranantanapan, O., Jaroonwitchawan, T. and Noisa, P. (2016). Human Embryonic Stem Cells: A Model for the Study of Neural Development and Neurological Diseases. *Stem Cells International*. doi: https://doi.org/10.1155/2016/2958210.

Prazeres, H., Torres, J., Rodrigues, F., Pinto, M., Pastoriza, M.C., Gomes, D., … Soares, P. (2011). Chromosomal, epigenetic and microRNA-mediated inactivation of LRP1B, a modulator of the extracellular environment of thyroid cancer cells. *Oncogene*. doi: https://doi.org/10.1038/onc.2010.512.

Putkhao, K., Kocerha, J., Cho, I.K., Yang, J., Parnpai, R. and Chan, A.W.S. (2013). Pathogenic cellular phenotypes are germline transmissible in a transgenic primate model of Huntington's disease. *Stem Cells and Development*. doi: https://doi.org/10.1089/scd.2012.0469.

Qi, M.L., Tagawa, K., Enokido, Y., Yoshimura, N., Wada, Y. ichi, Watase, K., … Okazawa, H. (2007). Proteome analysis of soluble nuclear proteins reveals that HMGB1/2 suppress genotoxic stress in polyglutamine diseases. *Nature Cell Biology*. doi: https://doi.org/10.1038/ncb1553.

Qing, X., Walter, J., Jarazo, J., Arias-Fuenzalida, J., Hillje, A.-L. and Schwamborn, J.C. (2017). CRISPR/Cas9 and piggyBac-mediated footprint-free LRRK2-G2019S knock-in reveals neuronal complexity phenotypes and α-Synuclein modulation in dopaminergic neurons. *Stem Cell Research* **24**:44–50. doi: https://doi.org/10.1016/J.SCR.2017.08.013.

Quarrell, O.W., Nance, M.A., Nopoulos, P., Paulsen, J.S., Smith, J.A. and Squitieri, F. (2013). Managing juvenile Huntington's disease. *Neurodegenerative Disease Management*. doi:

https://doi.org/10.2217/nmt.13.18.

Ramdzan, Y.M., Trubetskov, M.M., Ormsby, A.R., Newcombe, E.A., Sui, X., Tobin, M.J., … Hatters, D.M. (2017). Huntingtin Inclusions Trigger Cellular Quiescence, Deactivate Apoptosis, and Lead to Delayed Necrosis. *Cell Reports*. doi: https://doi.org/10.1016/j.celrep.2017.04.029.

Ranen, N.G., Stine, O.C., Abbott, M.H., Sherr, M., Codori, A.M., Franz, M.L., … Ross, C.A. (1995). Anticipation and instability of IT-15 (CAG)(N) repeats in parent-offspring pairs with Huntington disease. *American Journal of Human Genetics*.

Rasola, A., Sciacovelli, M., Pantic, B. and Bernardi, P. (2010). Signal transduction to the permeability transition pore. *FEBS Letters*.

Ray Chaudhuri, A. and Nussenzweig, A. (2017). The multifaceted roles of PARP1 in DNA repair and chromatin remodelling. *Nature Reviews Molecular Cell Biology*.

Reddy, P.H. and Shirendeb, U.P. (2012). Mutant huntingtin, abnormal mitochondrial dynamics, defective axonal transport of mitochondria, and selective synaptic degeneration in Huntington's disease. *Biochimica et Biophysica Acta - Molecular Basis of Disease*.

Reiner, A., Dragatsis, I. and Dietrich, P. (2011). Genetics and neuropathology of huntington's disease. *International Review of Neurobiology*.

Reitmair, A.H., Risley, R., Bristow, R.G., Wilson, T., Ganesh, A., Jang, A., … Meuth, M. (1997). Mutator phenotype in Msh2-deficient murine embryonic fibroblasts. *Cancer Research*.

Ridley, R.M., Frith, C.D., Crow, T.J. and Conneally, P.M. (1988). Anticipation in Huntington's disease is inherited through the male line but may originate in the female. *Journal of Medical Genetics*.

Riess, O., Schöls, L., Böttger, H., Nolte, D., Menezes Vieira-Saecker, A.M., Schimming, C., … Laccone, F.A. (1997). SCA6 is caused by moderate CAG expansion in the α(1A)-voltage-dependent calcium channel gene. *Human Molecular Genetics*. doi: https://doi.org/10.1093/hmg/6.8.1289.

Rigamonti, D., Sipione, S., Goffredo, D., Zuccato, C., Fossale, E. and Cattaneo, E. (2001). Huntingtin's Neuroprotective Activity Occurs via Inhibition of Procaspase-9 Processing. *Journal of Biological Chemistry*. doi: https://doi.org/10.1074/jbc.C100044200.

Ring, K.L., An, M.C., Zhang, N., O'Brien, R.N., Ramos, E.M., Gao, F., … Ellerby, L.M. (2015). Genomic Analysis Reveals Disruption of Striatal Neuronal Development and Therapeutic Targets in Human Huntington's Disease Neural Stem Cells. *Stem Cell Reports*. doi: https://doi.org/10.1016/j.stemcr.2015.11.005.

Rodinova, M., Krizova, J., Stufkova, H., Bohuslavova, B., Askeland, G., Dosoudilova, Z., … Hansikova, H.

(2019). Deterioration of mitochondrial bioenergetics and ultrastructure impairment in skeletal muscle of a transgenic minipig model in the early stages of Huntington's disease. *Disease Models & Mechanisms*. doi: https://doi.org/10.1242/dmm.038737.

Roh, D.S., Cook, A.L., Rhee, S.S., Joshi, A., Kowalski, R., Dhaliwal, D.K. and Funderburgh, J.L. (2008). DNA cross-linking, double-strand breaks, and apoptosis in corneal endothelial cells after a single exposure to mitomycin C. *Investigative Ophthalmology and Visual Science*. doi: https://doi.org/10.1167/iovs.08-2115.

De Rooij, K.E., De Koning Gans, P.A.M., Roos, R.A.C., Van Ommen, G.J.B. and Den Dunnen, J.T. (1995). Somatic expansion of the (CAG)n repeat in Huntington disease brains. *Human Genetics*. doi: https://doi.org/10.1007/BF00225192.

Roos, W.P., Thomas, A.D. and Kaina, B. (2016). DNA damage and the balance between survival and death in cancer biology. *Nature Reviews Cancer*.

Rosas, H.D., Hevelone, N.D., Zaleta, A.K., Greve, D.N., Salat, D.H. and Fischl, B. (2005). Regional cortical thinning in preclinical Huntington disease and its relationship to cognition. *Neurology*. doi: https://doi.org/10.1212/01.wnl.0000174432.87383.87.

Rosas, H.D., Liu, A.K., Hersch, S., Glessner, M., Ferrante, R.J., Salat, D.H., … Fischl, B. (2002). Regional and progressive thinning of the cortical ribbon in Huntington's disease. *Neurology*. doi: https://doi.org/10.1212/WNL.58.5.695.

Ross, C.A., Aylward, E.H., Wild, E.J., Langbehn, D.R., Long, J.D., Warner, J.H., … Tabrizi, S.J. (2014). Huntington disease: Natural history, biomarkers and prospects for therapeutics. *Nature Reviews Neurology*.

Ross, C.A. and Tabrizi, S.J. (2011). Huntington's disease: From molecular pathogenesis to clinical treatment. *The Lancet Neurology*.

Ruzo, A., Croft, G.F., Metzger, J.J., Galgoczi, S., Gerber, L.J., Pellegrini, C., … Brivanlou, A.H. (2018). Chromosomal instability during neurogenesis in Huntington's disease. *Development (Cambridge)* **145**(2):dev156844. doi: https://doi.org/10.1242/dev.156844.

Ruzo, A., Ismailoglu, I., Popowski, M., Haremaki, T., Croft, G.F., Deglincerti, A. and Brivanlou, A.H. (2015). Discovery of novel isoforms of Huntingtin reveals a new hominid-specific exon. *PLoS ONE*. doi: https://doi.org/10.1371/journal.pone.0127687.

Ryland, G.L., Doyle, M.A., Goode, D., Boyle, S.E., Choong, D.Y.H., Rowley, S.M., … Gorringe, K.L. (2015). Loss of heterozygosity: What is it good for? *BMC Medical Genomics*. doi: https://doi.org/10.1186/s12920-015-0123-z.

Saleh-Gohari, N. and Helleday, T. (2004). Conservative homologous recombination preferentially repairs DNA double-strand breaks in the S phase of the cell cycle un human cells. *Nucleic Acids Research*. doi: https://doi.org/10.1093/nar/gkh703.

Salomón-Torres, R., González-Vizcarra, V.M., Medina-Basulto, G.E., Montaño-Gómez, M.F., Mahadevan, P., Yaurima-Basaldúa, V.H., … Villa-Angulo, R. (2015). Genome-wide identification of copy number variations in holstein cattle from Baja California, Mexico, using high-density SNP genotyping arrays. *Genetics and Molecular Research*. doi: https://doi.org/10.4238/2015.October.2.18.

Sánchez-Pernaute, R., Künig, G., Del Barrio Alba, A., De Yébenes, J.G., Vontobel, P. and Leenders, K.L. (2000). Bradykinesia in early Huntington's disease. *Neurology*.

Sánchez, A., Milà, M., Castellví-Bel, S., Rosich, M., Jiménez, D., Badenas, C. and Estivill, X. (1997). Maternal transmission in sporadic Huntington's disease. *Journal of Neurology Neurosurgery and Psychiatry*. doi: https://doi.org/10.1136/jnnp.62.5.535.

Sanchez Mejia, R.O. and Friedlander, R.M. (2001). Caspases in Huntington's disease. *Neuroscientist*.

Sandoe, J. and Eggan, K. (2013). Opportunities and challenges of pluripotent stem cell neurodegenerative disease models. *Nature Neuroscience*.

Sapp, E., Ge, P., Aizawa, H., Bird, E., Penney, J., Young, A.B., … Difiglia, M. (1995). Evidence for a preferential loss of enkephalin immunoreactivity in the external globus pallidus in low grade Huntington's disease using high resolution image analysis. *Neuroscience*. doi: https://doi.org/10.1016/0306-4522(94)00427-7.

Sarzi, E., Brown, M.D., Lebon, S., Chretien, D., Munnich, A., Rotig, A. and Procaccio, V. (2007). A novel recurrent mitochondrial DNA mutation in ND3 gene is associated with isolated complex I deficiency causing leigh syndrome and dystonia. *American Journal of Medical Genetics, Part A*. doi: https://doi.org/10.1002/ajmg.a.31565.

Sathasivam, K., Hobbs, C., Turmaine, M., Mangiarini, L., Mahal, A., Bertaux, F., … Bates, G.P. (1999). Formation of polyglutamine inclusions in non-CNS tissue. *Human Molecular Genetics*. doi: https://doi.org/10.1093/hmg/8.5.813.

Sathasivam, K., Neueder, A., Gipson, T.A., Landles, C., Benjamin, A.C., Bondulich, M.K., … Bates, G.P. (2013). Aberrant splicing of HTT generates the pathogenic exon 1 protein in Huntington disease. *Proceedings of the National Academy of Sciences of the United States of America*. doi: https://doi.org/10.1073/pnas.1221891110.

Saudou, F., Finkbeiner, S., Devys, D. and Greenberg, M.E. (1998). Huntingtin acts in the nucleus to induce apoptosis but death does not correlate with the formation of intranuclear inclusions. *Cell*. doi:

https://doi.org/10.1016/S0092-8674(00)81782-1.

Saudou, F. and Humbert, S. (2016). The Biology of Huntingtin. *Neuron*.

Sawa, A., Wiegand, G.W., Cooper, J., Margolis, R.L., Sharp, A.H., Lawler, J.F., … Ross, C.A. (1999). Increased apoptosis of Huntington disease lymphoblasts associated with repeat length-dependent mitochondrial depolarization. *Nature Medicine*. doi: https://doi.org/10.1038/13518.

Schaffar, G., Breuer, P., Boteva, R., Behrends, C., Tzvetkov, N., Strippel, N., … Hartl, F.U. (2004). Cellular toxicity of polyglutamine expansion proteins: Mechanism of transcription factor deactivation. *Molecular Cell*. doi: https://doi.org/10.1016/j.molcel.2004.06.029.

Scherzinger, E., Lurz, R., Turmaine, M., Mangiarini, L., Hollenbach, B., Hasenbank, R., … Wanker, E.E. (1997). Huntingtin-encoded polyglutamine expansions form amyloid-like protein aggregates in vitro and in vivo. *Cell*. doi: https://doi.org/10.1016/S0092-8674(00)80514-0.

Schlacher, K., Wu, H. and Jasin, M. (2012). A Distinct Replication Fork Protection Pathway Connects Fanconi Anemia Tumor Suppressors to RAD51-BRCA1/2. *Cancer Cell*. doi: https://doi.org/10.1016/j.ccr.2012.05.015.

Schmidt, M.H.M. and Pearson, C.E. (2016). Disease-associated repeat instability and mismatch repair. *DNA Repair*.

Schweitzer, J.K. and Livingston, D.M. (1997). Destabilization of CAG trinucleotide repeat tracts by mismatch repair mutations in yeast. *Human Molecular Genetics*. doi: https://doi.org/10.1093/hmg/6.3.349.

Schweitzer, J.K. and Livingston, D.M. (1999). The effect of DNA replication mutations on CAG tract stability in yeast. *Genetics*.

Seguí, N., Mina, L.B., Lázaro, C., Sanz-Pamplona, R., Pons, T., Navarro, M., … Valle, L. (2015). Germline Mutations in FAN1 Cause Hereditary Colorectal Cancer by Impairing DNA Repair. *Gastroenterology*. doi: https://doi.org/10.1053/j.gastro.2015.05.056.

Semaka, A., Creighton, S., Warby, S. and Hayden, M.A. (2006). Predictive testing for Huntington disease: Interpretation and significance of intermediate alleles. *Clinical Genetics*.

Seneca, S., Fagnart, D., Keymolen, K., Lissens, W., Hasaerts, D., Debulpaep, S., … De Meirleir, L. (2004). Early onset Huntington disease: A neuronal degeneration syndrome. *European Journal of Pediatrics*. doi: https://doi.org/10.1007/s00431-004-1537-3.

Seong, I.S., Ivanova, E., Lee, J.M., Choo, Y.S., Fossale, E., Anderson, M.A., … MacDonald, M.E. (2005). HD CAG repeat implicates a dominant property of huntingtin in mitochondrial energy metabolism.

Human Molecular Genetics. doi: https://doi.org/10.1093/hmg/ddi319.

Seredenina, T. and Luthi-Carter, R. (2012). What have we learned from gene expression profiles in Huntington's disease? *Neurobiology of Disease*.

Shahen, V.A., Cantrill, L.C., Sangani, N.B., Christodoulou, J. and Gold, W.A. (2018). A simple and efficient toolset for analysing mitochondrial trafficking in neuronal cells. *Acta Histochemica*. doi: https://doi.org/10.1016/j.acthis.2018.09.001.

Shelbourne, P.F., Keller-McGandy, C., Bi, W.L., Yoon, S.R., Dubeau, L., Veitch, N.J., … Augood, S.J. (2007). Triplet repeat mutation length gains correlate with cell-type specific vulnerability in Huntington disease brain. *Human Molecular Genetics* 16(10):1133–1142. doi: https://doi.org/10.1093/hmg/ddm054.

Shereda, R.D., Machida, Y.J.Y. and Machida, Y.J.Y. (2010). Human KIAA1018/FAN1 localizes to stalled replication forks via its ubiquitin-binding domain. *Cell Cycle*. doi: https://doi.org/10.4161/cc.9.19.13207.

Shirendeb, U., Reddy, A.P., Manczak, M., Calkins, M.J., Mao, P., Tagle, D.A. and Reddy, P.H. (2011). Abnormal mitochondrial dynamics, mitochondrial loss and mutant huntingtin oligomers in Huntington's disease: Implications for selective neuronal damage. *Human Molecular Genetics*. doi: https://doi.org/10.1093/hmg/ddr024.

Shirendeb, U.P., Calkins, M.J., Manczak, M., Anekonda, V., Dufour, B., McBride, J.L., … Reddy, P.H. (2012). Mutant Huntingtin's interaction with mitochondrial protein Drp1 impairs mitochondrial biogenesis and causes defective axonal transport and synaptic degeneration in Huntington's disease. *Human Molecular Genetics*. doi: https://doi.org/10.1093/hmg/ddr475.

Shiwach, R. (1994). Psychopathology in Huntington's disease patients. *Acta Psychiatrica Scandinavica*. doi: https://doi.org/10.1111/j.1600-0447.1994.tb01587.x.

Shoulson, I. (2006). At risk for Huntington disease: The PHAROS (Prospective Huntington At Risk Observational Study) cohort enrolled. *Archives of Neurology*. doi: https://doi.org/10.1001/archneur.63.7.991.

Shui, B., Hernandez Matias, L., Guo, Y. and Peng, Y. (2016). The Rise of CRISPR/Cas for Genome Editing in Stem Cells. *Stem Cells International*.

Siddiqui, A., Rivera-Sánchez, S., Castro, M.D.R., Acevedo-Torres, K., Rane, A., Torres-Ramos, C.A., … Ayala-Torres, S. (2012). Mitochondrial DNA damage Is associated with reduced mitochondrial bioenergetics in Huntington's disease. *Free Radical Biology and Medicine* 53(7):1478–1488. doi: https://doi.org/10.1016/j.freeradbiomed.2012.06.008.

Sies, H. (1991). Oxidative stress: From basic research to clinical application. *American Journal of Medicine*.

Skodda, S., Schlegel, U., Hoffmann, R. and Saft, C. (2014). Impaired motor speech performance in Huntington's disease. *Journal of Neural Transmission*. doi: https://doi.org/10.1007/s00702-013-1115-9.

Slow, E.J., Graham, R.K., Osmand, A.P., Devon, R.S., Lu, G., Deng, Y., … Hayden, M.R. (2005). Absence of behavioral abnormalities and neurodegeneration in vivo despite widespread neuronal huntingtin inclusions. *Proceedings of the National Academy of Sciences of the United States of America*. doi: https://doi.org/10.1073/pnas.0503634102.

Smogorzewska, A., Desetty, R., Saito, T.T., Schlabach, M., Lach, F.P., Sowa, M.E., … Elledge, S.J. (2010). A Genetic Screen Identifies FAN1, a Fanconi Anemia-Associated Nuclease Necessary for DNA Interstrand Crosslink Repair. *Molecular Cell* **39**(1):36–47. doi: https://doi.org/10.1016/j.molcel.2010.06.023.

Soldner, F., Hockemeyer, D., Beard, C., Gao, Q., Bell, G.W., Cook, E.G., … Jaenisch, R. (2009). Parkinson's Disease Patient-Derived Induced Pluripotent Stem Cells Free of Viral Reprogramming Factors. *Cell*. doi: https://doi.org/10.1016/j.cell.2009.02.013.

Soldner, F., Laganière, J., Cheng, A.W., Hockemeyer, D., Gao, Q., Alagappan, R., … Jaenisch, R. (2011). Generation of isogenic pluripotent stem cells differing exclusively at two early onset parkinson point mutations. *Cell*. doi: https://doi.org/10.1016/j.cell.2011.06.019.

Song, W., Chen, J., Petrilli, A., Liot, G., Klinglmayr, E., Zhou, Y., … Bossy-Wetzel, E. (2011). Mutant huntingtin binds the mitochondrial fission GTPase dynamin-related protein-1 and increases its enzymatic activity. *Nature Medicine*. doi: https://doi.org/10.1038/nm.2313.

Spada, A.R.L., Wilson, E.M., Lubahn, D.B., Harding, A.E. and Fischbeck, K.H. (1991). Androgen receptor gene mutations in X-linked spinal and bulbar muscular atrophy. *Nature*. doi: https://doi.org/10.1038/352077a0.

Spiro, C. and McMurray, C.T. (2003). Nuclease-Deficient FEN-1 Blocks Rad51/BRCA1-Mediated Repair and Causes Trinucleotide Repeat Instability. *Molecular and Cellular Biology*. doi: https://doi.org/10.1128/mcb.23.17.6063-6074.2003.

Spruijt, L., Smeets, H.J., Hendrickx, A., Bettink-Remeijer, M.W., Maat-Kievit, A., Schoonderwoerd, K.C., … Hintzen, R.Q. (2007). A MELAS-associated ND1 mutation causing Leber hereditary optic neuropathy and spastic dystonia. *Archives of Neurology*. doi: https://doi.org/10.1001/archneur.64.6.890.

Squitieri, F., Cannella, M., Simonelli, M., Sassone, J., Martino, T., Venditti, E., … Ciarmiello, A. (2009).

Distinct brain volume changes correlating with clinical stage, disease progression rate, mutation size, and age at onset prediction as early biomarkers of brain atrophy in Huntington's disease. *CNS Neuroscience and Therapeutics*. doi: https://doi.org/10.1111/j.1755-5949.2008.00068.x.

Squitieri, F., Gellera, C., Cannella, M., Mariotti, C., Cislaghi, G., Rubinsztein, D.C., … Di Donato, S. (2003). Homozygosity for CAG mutation in Huntington disease is associated with a more severe clinical course. *Brain*. doi: https://doi.org/10.1093/brain/awg077.

Squitieri, F., Sabbadini, G., Mandich, P., Gellera, C., Di Maria, E., Bellone, E., … Novelletto, A. (2000). Family and molecular data for a fine analysis of age at onset in Huntington disease. *American Journal of Medical Genetics*. doi: https://doi.org/10.1002/1096-8628(20001211)95:4<366::AID-AJMG13>3.0.CO;2-2.

Stack, C., Ho, D., Wille, E., Calingasan, N.Y., Williams, C., Liby, K., … Beal, M.F. (2010). Triterpenoids CDDO-ethyl amide and CDDO-trifluoroethyl amide improve the behavioral phenotype and brain pathology in a transgenic mouse model of Huntington's disease. *Free Radical Biology and Medicine*. doi: https://doi.org/10.1016/j.freeradbiomed.2010.03.017.

Steffan, J.S., Agrawal, N., Pallos, J., Rockabrand, E., Trotman, L.C., Slepko, N., … Marsh, J.L. (2004). SUMO Modification of Huntingtin and Huntington's Disease Pathology. *Science*. doi: https://doi.org/10.1126/science.1092194.

Steffan, J.S., Kazantsev, A., Spasic-Boskovic, O., Greenwald, M., Zhu, Y.Z., Gohler, H., … Thompson, L.M. (2000). The Huntington's disease protein interacts with p53 and CREB-binding protein and represses transcription. *Proceedings of the National Academy of Sciences of the United States of America*. doi: https://doi.org/10.1073/pnas.100110097.

Sugimoto, N., Nakano, S. ichi, Katoh, M., Matsumura, A., Nakamuta, H., Ohmichi, T., … Sasaki, M. (1995). Thermodynamic Parameters To Predict Stability of RNA/DNA Hybrid Duplexes. *Biochemistry*. doi: https://doi.org/10.1021/bi00035a029.

Suzuki, K., Yu, C., Qu, J., Li, M., Yao, X., Yuan, T., … Izpisua Belmonte, J.C. (2014). Targeted gene correction minimally impacts whole-genome mutational load in human-disease-specific induced pluripotent stem cell clones. *Cell Stem Cell*. doi: https://doi.org/10.1016/j.stem.2014.06.016.

Swami, M., Hendricks, A.E., Gillis, T., Massood, T., Mysore, J., Myers, R.H. and Wheeler, V.C. (2009). Somatic expansion of the Huntington's disease CAG repeat in the brain is associated with an earlier age of disease onset. *Human Molecular Genetics* 18(16):3039–3047. doi: https://doi.org/10.1093/hmg/ddp242.

Świtońska, K., Szlachcic, W.J., Handschuh, L., Wojciechowski, P., Marczak, Ł., Stelmaszczuk, M., …

Figiel, M. (2019). Identification of altered developmental pathways in human juvenile HD iPSC with 71Q and 109Q using transcriptome profiling. *Frontiers in Cellular Neuroscience*. doi: https://doi.org/10.3389/fncel.2018.00528.

Szlachcic, W.J., Switonski, P.M., Krzyzosiak, W.J., Figlerowicz, M. and Figiel, M. (2015). Huntington disease iPSCs show early molecular changes in intracellular signaling, the expression of oxidative stress proteins and the p53 pathway. *DMM Disease Models and Mechanisms*. doi: https://doi.org/10.1242/dmm.019406.

Tabrizi, S.J., Langbehn, D.R., Leavitt, B.R., Roos, R.A., Durr, A., Craufurd, D., … Stout, J.C. (2009). Biological and clinical manifestations of Huntington's disease in the longitudinal TRACK-HD study: cross-sectional analysis of baseline data. *The Lancet Neurology*. doi: https://doi.org/10.1016/S1474-4422(09)70170-X.

Tabrizi, S.J., Scahill, R.I., Durr, A., Roos, R.A.C., Leavitt, B.R., Jones, R., … Stout, J.C. (2011). Biological and clinical changes in premanifest and early stage Huntington's disease in the TRACK-HD study: The 12-month longitudinal analysis. *The Lancet Neurology*. doi: https://doi.org/10.1016/S1474-4422(10)70276-3.

Tabrizi, S.J., Scahill, R.I., Owen, G., Durr, A., Leavitt, B.R., Roos, R.A., … Langbehn, D.R. (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. *The Lancet Neurology*. doi: https://doi.org/10.1016/S1474-4422(13)70088-7.

Takahashi, K., Tanabe, K., Ohnuki, M., Narita, M., Ichisaka, T., Tomoda, K. and Yamanaka, S. (2007). Induction of Pluripotent Stem Cells from Adult Human Fibroblasts by Defined Factors. *Cell*. doi: https://doi.org/10.1016/j.cell.2007.11.019.

Takahashi, K. and Yamanaka, S. (2006). Induction of Pluripotent Stem Cells from Mouse Embryonic and Adult Fibroblast Cultures by Defined Factors. *Cell*. doi: https://doi.org/10.1016/j.cell.2006.07.024.

Takahashi, T., Kikuchi, S., Katada, S., Nagai, Y., Nishizawa, M. and Onodera, O. (2008). Soluble polyglutamine oligomers formed prior to inclusion body formation are cytotoxic. *Human Molecular Genetics*. doi: https://doi.org/10.1093/hmg/ddm311.

Takano, H. and Gusella, J.F. (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 Neuroscience*. doi: https://doi.org/10.1186/1471-2202-3-15.

Tanaka, F., Reeves, M.F., Ito, Y., Matsumoto, M., Li, M., Miwa, S., … Sobue, G. (2002). Tissue-Specific Somatic Mosaicism in Spinal and Bulbar Muscular Atrophy Is Dependent on CAG-Repeat Length and

Androgen Receptor–Gene Expression Level. *The American Journal of Human Genetics*. doi: https://doi.org/10.1086/302578.

Telenius, H., Kremer, B., Goldberg, Y.P., Theilmann, J., Andrew, S.E., Zeisler, J., … Hayden, M.R. (1994). Somatic and gonadal mosaicism of the Huntington disease gene CAG repeat in brain and sperm. *Nature Genetics*. doi: https://doi.org/10.1038/ng0494-409.

Telenius, H., Kremer, H.P.H., Thellmann, J., Andrew, S.E., Almqvist, E., Anvret, M., … Hayden, M.R. (1993). Molecular analysis of juvenile huntington disease: The major influence on (CAG)n repeat length is the sex of the affected parent. *Human Molecular Genetics*. doi: https://doi.org/10.1093/hmg/2.10.1535.

Tellez-Nagel, I., Johnson, A.B. and Terry, R.D. (1974). Studies on brain biopsies of patients with huntington's chorea. *Journal of Neuropathology and Experimental Neurology*. doi: https://doi.org/10.1097/00005072-197404000-00008.

De Temmerman, N., Sermon, K., Seneca, S., De Rycke, M., Hilven, P., Lissens, W., … Liebaers, I. (2004). Intergenerational instability of the expanded CTG repeat in the DMPK gene: Studies in human gametes and preimplantation embryos. *American Journal of Human Genetics*. doi: https://doi.org/10.1086/422762.

The HD iPSC Consortium (2012). Induced Pluripotent Stem Cells from Patients with Huntington's Disease Show CAG-Repeat-Expansion-Associated Phenotypes. *Cell Stem Cell* **11**(2):264–278. doi: https://doi.org/10.1016/j.stem.2012.04.027.

The HD iPSC Consortium (2017). Developmental alterations in Huntington's disease neural cells and pharmacological rescue in cells and mice. *Nature Neuroscience*. doi: https://doi.org/10.1038/nn.4532.

Thieben, M.J. (2002). The distribution of structural neuropathology in pre-clinical Huntington's disease. *Brain*. doi: https://doi.org/10.1093/brain/awf179.

Thompson, L.M., Aiken, C.T., Kaltenbach, L.S., Agrawal, N., Illes, K., Khoshnan, A., … Steffan, J.S. (2009). IKK phosphorylates Huntingtin and targets it for degradation by the proteasome and lysosome. *Journal of Cell Biology*. doi: https://doi.org/10.1083/jcb.200909067.

Thongthip, S., Bellani, M., Gregg, S.Q., Sridhar, S., Conti, B.A., Chen, Y., … Smogorzewska, A. (2016). Fan1 deficiency results in DNA interstrand cross-link repair defects, enhanced tissue karyomegaly, and organ dysfunction. *Genes and Development*. doi: https://doi.org/10.1101/gad.276261.115.

Tiefenbach, T. and Junop, M. (2012). Pso2 (SNM1) is a DNA structure-specific endonuclease. *Nucleic Acids Research*. doi: https://doi.org/10.1093/nar/gkr1059.

To, M.S., Aromataris, E.C., Castro, J., Roberts, M.L., Barritt, G.J. and Rychkov, G.Y. (2010). Mitochondrial uncoupler FCCP activates proton conductance but does not block store-operated Ca2+ current in liver cells. *Archives of Biochemistry and Biophysics*. doi: https://doi.org/10.1016/j.abb.2010.01.004.

Tomé, S., Holt, I., Edelmann, W., Morris, G.E., Munnich, A., Pearson, C.E. and Gourdon, G. (2009). MSH2 ATPase domain mutation affects CTG•CAG repeat instability in transgenic mice. *PLoS Genetics*. doi: https://doi.org/10.1371/journal.pgen.1000482.

Tomé, S., Manley, K., Simard, J.P., Clark, G.W., Slean, M.M., Swami, M., … Pearson, C.E. (2013). MSH3 Polymorphisms and Protein Levels Affect CAG Repeat Instability in Huntington's Disease Mice. *PLoS Genetics*. doi: https://doi.org/10.1371/journal.pgen.1003280.

Tóth, G., Gáspári, Z. and Jurka, J. (2000). Microsatellites in different eukaryotic genomes: Surveys and analysis. *Genome Research*. doi: https://doi.org/10.1101/gr.10.7.967.

Trushina, E., Dyer, R.B., Badger, J.D., Ure, D., Eide, L., Tran, D.D., … McMurray, C.T. (2004). Mutant Huntingtin Impairs Axonal Trafficking in Mammalian Neurons In Vivo and In Vitro. *Molecular and Cellular Biology*. doi: https://doi.org/10.1128/mcb.24.18.8195-8209.2004.

Tuo, J., Chen, C., Zeng, X., Christiansen, M. and Bohr, V.A. (2002). Functional crosstalk between hOgg1 and the helicase domain of Cockayne syndrome group B protein. *DNA Repair*. doi: https://doi.org/10.1016/S1568-7864(02)00116-7.

Ueno, S. ichi, Kondoh, K., Komure, Y., Komure, O., Kuno, S., Kawai, J., … Sano, A. (1995). Somatic mosaicism of CAG repeat in dentatorubral-pallidoluysian atrophy (DRPLA). *Human Molecular Genetics*. doi: https://doi.org/10.1093/hmg/4.4.663.

Umar, A., Boyer, J.C., Thomas, D.C., Nguyen, D.C., Risinger, J.I., Boyd, J., … Kunkel, T.A. (1994). Defective mismatch repair in extracts of colorectal and endometrial cancer cell lines exhibiting microsatellite instability. *Journal of Biological Chemistry*.

Usdin, K. and Grabczyk, E. (2000). DNA repeat expansions and human disease. *Cellular and Molecular Life Sciences*.

Usdin, K., House, N.C.M. and Freudenreich, C.H. (2015). Repeat instability during DNA repair: Insights from model systems. *Critical Reviews in Biochemistry and Molecular Biology*.

Velier, J., Kim, M., Schwarz, C., Kim, T.W., Sapp, E., Chase, K., … DiFiglia, M. (1998). Wild-type and mutant huntingtins function in vesicle trafficking in the secretory and endocytic pathways. *Experimental Neurology*. doi: https://doi.org/10.1006/exnr.1998.6832.

Veres, A., Gosis, B.S., Ding, Q., Collins, R., Ragavendran, A., Brand, H., … Musunuru, K. (2014). Low incidence of Off-target mutations in individual CRISPR-Cas9 and TALEN targeted human stem cell clones detected by whole-genome sequencing. *Cell Stem Cell*. doi: https://doi.org/10.1016/j.stem.2014.04.020.

Visscher, P.M., Wray, N.R., Zhang, Q., Sklar, P., McCarthy, M.I., Brown, M.A. and Yang, J. (2017). 10 Years of GWAS Discovery: Biology, Function, and Translation. *American Journal of Human Genetics*.

Vitanova, K.S., Stringer, K.M., Benitez, D.P., Brenton, J. and Cummings, D.M. (2019). Dementia associated with disorders of the basal ganglia. *Journal of Neuroscience Research* 0(0). doi: https://doi.org/10.1002/jnr.24508.

Vonsattel, J.P., Myers, R.H., Stevens, T.J., Ferrante, R.J., Bird, E.D. and Richardson, E.P. (1985). Neuropathological classification of huntington's disease. *Journal of Neuropathology and Experimental Neurology*. doi: https://doi.org/10.1097/00005072-198511000-00003.

Wang, C.E., Tydlacka, S., Orr, A.L., Yang, S.H., Graham, R.K., Hayden, M.R., … Li, X.J. (2008). Accumulation of N-terminal mutant huntingtin in mouse and monkey models implicated as a pathogenic mechanism in Huntington's disease. *Human Molecular Genetics*. doi: https://doi.org/10.1093/hmg/ddn175.

Wang, G., Yang, L., Grishin, D., Rios, X., Ye, L.Y., Hu, Y., … Pu, W.T. (2017). Efficient, footprint-free human iPSC genome editing by consolidation of Cas9/CRISPR and piggyBac technologies. *Nature Protocols*. doi: https://doi.org/10.1038/nprot.2016.152.

Wapner, R.J., Martin, C.L., Levy, B., Ballif, B.C., Eng, C.M., Zachary, J.M., … Jackson, L. (2013). Chromosomal microarray versus karyotyping for prenatal diagnosis. *Obstetrical and Gynecological Survey*.

Warner, J.P., Barron, L.H. and Brock, D.J.H. (1993). A new polymerase chain reaction (pcr) assay for the trinucleotide repeat that is unstable and expanded on huntington's disease chromosomes. *Molecular and Cellular Probes*. doi: https://doi.org/10.1006/mcpr.1993.1034.

Warren, L., Manos, P.D., Ahfeldt, T., Loh, Y.H., Li, H., Lau, F., … Rossi, D.J. (2010). Highly efficient reprogramming to pluripotency and directed differentiation of human cells with synthetic modified mRNA. *Cell Stem Cell*. doi: https://doi.org/10.1016/j.stem.2010.08.012.

Watanabe, H., Tanaka, F., Doyu, M., Riku, S., Yoshida, M., Hashizume, Y. and Sobue, G. (2000). Differential somatic CAG repeat instability in variable brain cell lineage in dentatorubral pallidoluysian atrophy (DRPLA): A laser-captured microdissection (LCM)-based analysis. *Human Genetics*. doi: https://doi.org/10.1007/s004390000400.

Wellington, C.L., Ellerby, L.M., Hackam, A.S., Margolis, R.L., Trifiro, M.A., Singaraja, R., … Hayden, M.R. (1998). Caspase cleavage of gene products associated with triplet expansion disorders generates truncated fragments containing the polyglutamine tract. *Journal of Biological Chemistry*. doi: https://doi.org/10.1074/jbc.273.15.9158.

Werbowetski-Ogilvie, T.E., Bossé, M., Stewart, M., Schnerch, A., Ramos-Mejia, V., Rouleau, A., … Bhatia, M. (2009). Characterization of human embryonic stem cells with features of neoplastic progression. *Nature Biotechnology*. doi: https://doi.org/10.1038/nbt.1516.

Wexler, A. (2010). Stigma, history, and Huntington's disease. *The Lancet*. doi: https://doi.org/10.1016/s0140-6736(10)60957-9.

Wexler, N.S. (2004). Venezuelan kindreds reveal that genetic and environmental factors modulate Huntington's disease age of onset. *Proceedings of the National Academy of Sciences of the United States of America* 101(10):3498–3503. doi: https://doi.org/10.1073/pnas.0308679101.

Wexler, N.S., Young, A.B., Tanzi, R.E., Travers, H., Starosta-Rubinstein, S., Penney, J.B., … Gusella, J.F. (1987). Homozygotes for huntington's disease. *Nature*. doi: https://doi.org/10.1038/326194a0.

Weydt, P., Pineda, V. V., Torrence, A.E., Libby, R.T., Satterfield, T.F., Lazarowski, E.R.R., … La Spada, A.R. (2006). Thermoregulatory and metabolic defects in Huntington's disease transgenic mice implicate PGC-1α in Huntington's disease neurodegeneration. *Cell Metabolism*. doi: https://doi.org/10.1016/j.cmet.2006.10.004.

Whan, V., Hobbs, M., McWilliam, S., Lynn, D.J., Lutzow, Y.S., Khatkar, M., … Tellam, R.L. (2010). Bovine proteins containing poly-glutamine repeats are often polymorphic and enriched for components of transcriptional regulatory complexes. *BMC Genomics*. doi: https://doi.org/10.1186/1471-2164-11-654.

Wheeler, V.C., Auerbach, W., White, J.K., Srinidhi, J., Auerbach, A., Ryan, A., … MacDonald, M.E. (1999). Length-dependent gametic CAG repeat instability in the Huntington's disease knock-in mouse. *Human Molecular Genetics* 8(1):115–122. doi: https://doi.org/10.1093/hmg/8.1.115.

Wheeler, V.C., Lebel, L.-A., Vrbanac, V., Teed, A., te Riele, H. and MacDonald, M.E. (2003). Mismatch repair gene Msh2 modifies the timing of early disease in Hdh(Q111) striatum. *Human molecular genetics* 12(3):273–281.

Wheeler, V.C., Persichetti, F., McNeil, S.M., Mysore, J.S., Mysore, S.S., MacDonald, M.E., … Wexler, N.S. (2007). Factors associated with HD CAG repeat instability in Huntington disease. *Journal of Medical Genetics*. doi: https://doi.org/10.1136/jmg.2007.050930.

Wieben, E.D., Aleff, R.A., Tosakulwong, N., Butz, M.L., Highsmith, W.E., Edwards, A.O. and Baratz, K.H.

(2012). A Common Trinucleotide Repeat Expansion within the Transcription Factor 4 (TCF4, E2-2) Gene Predicts Fuchs Corneal Dystrophy. *PLoS ONE*. doi: https://doi.org/10.1371/journal.pone.0049083.

Wierdl, M., Greene, C.N., Datta, A., Jinks-Robertson, S. and Petes, T.D. (1996). Destabilization of simple repetitive DNA sequences by transcription in yeast. *Genetics*.

Wilson, P.G. and Stice, S.S. (2006). Development and differentiation of neural rosettes derived from human embryonic stem cells. *Stem Cell Reviews*.

de Wind, N., Dekker, M., Berns, A., Radman, M. and te Riele, H. (1995). Inactivation of the mouse Msh2 gene results in mismatch repair deficiency, methylation tolerance, hyperrecombination, and predisposition to cancer. *Cell*. doi: https://doi.org/10.1016/0092-8674(95)90319-4.

De Wind, N., Dekker, M., Claij, N., Jansen, L., Van Klink, Y., Radman, M., … Riele, H. Te (1999). HNPCC-like cancer predisposition in mice through simultaneous loss of Msh3 and Msh6 mismatch-repair protein functions. *Nature Genetics*. doi: https://doi.org/10.1038/15544.

Woerner, A.C., Frottin, F., Hornburg, D., Feng, L.R., Meissner, F., Patra, M., … Hipp, M.S. (2016). Cytoplasmic protein aggregates interfere with nucleocytoplasmic transport of protein and RNA. *Science*. doi: https://doi.org/10.1126/science.aad2033.

Woltjen, K., Michael, I.P., Mohseni, P., Desai, R., Mileikovsky, M., Hämäläinen, R., … Nagy, A. (2009). PiggyBac transposition reprograms fibroblasts to induced pluripotent stem cells. *Nature*. doi: https://doi.org/10.1038/nature07863.

Wong, H.K., Muftuoglu, M., Beck, G., Imam, S.Z., Bohr, V.A. and Wilson, D.M. (2007). Cockayne syndrome B protein stimulates apurinic endonuclease 1 activity and protects against agents that introduce base excision repair intermediates. *Nucleic Acids Research*. doi: https://doi.org/10.1093/nar/gkm404.

Wright, G.E.B., Collins, J.A., Kay, C., McDonald, C., Dolzhenko, E., Xia, Q., … Hayden, M.R. (2019). Length of Uninterrupted CAG, Independent of Polyglutamine Size, Results in Increased Somatic Instability, Hastening Onset of Huntington Disease. *American Journal of Human Genetics*. doi: https://doi.org/10.1016/j.ajhg.2019.04.007.

Wyttenbach, A. (2002). Heat shock protein 27 prevents cellular polyglutamine toxicity and suppresses the increase of reactive oxygen species caused by huntingtin. *Human Molecular Genetics*. doi: https://doi.org/10.1093/hmg/11.9.1137.

Xia, J. (2003). Huntingtin contains a highly conserved nuclear export signal. *Human Molecular Genetics*. doi: https://doi.org/10.1093/hmg/ddg156.

Xia, W., Wang, Z., Wang, Q., Han, J., Zhao, C., Hong, Y., … Ying, W. (2009). Roles of NAD / NADH and NADP+ / NADPH in Cell Death. *Current Pharmaceutical Design*. doi: https://doi.org/10.2174/138161209787185832.

Xie, F., Ye, L., Chang, J.C., Beyer, A.I., Wang, J., Muench, M.O. and Kan, Y.W. (2014). Seamless gene correction of β-thalassemia mutations in patient-specific iPSCs using CRISPR/Cas9 and piggyBac. *Genome Research*. doi: https://doi.org/10.1101/gr.173427.114.

Xu, X., Tay, Y., Sim, B., Yoon, S.-I.I., Huang, Y., Ooi, J., … Pouladi, M.A. (2017). Reversal of Phenotypic Abnormalities by CRISPR/Cas9-Mediated Gene Correction in Huntington Disease Patient-Derived Induced Pluripotent Stem Cells. *Stem Cell Reports* **8**(3):619–633. doi: https://doi.org/10.1016/j.stemcr.2017.01.022.

Yamamoto, H. and Imai, K. (2015). Microsatellite instability: an update. *Archives of Toxicology*.

Yang, S., Lin, G., Tan, Y.Q., Zhou, D., Deng, L.Y., Cheng, D.H., … Lu, G.X. (2008). Tumor progression of culture-adapted human embryonic stem cells during long-term culture. *Genes Chromosomes and Cancer*. doi: https://doi.org/10.1002/gcc.20574.

Ying, W., Garnier, P. and Swanson, R.A. (2003). NAD+ repletion prevents PARP-1-induced glycolytic blockade and cell death in cultured mouse astrocytes. *Biochemical and Biophysical Research Communications*. doi: https://doi.org/10.1016/S0006-291X(03)01483-9.

Yoshihara, M., Hayashizaki, Y. and Murakawa, Y. (2017). Genomic Instability of iPSCs: Challenges Towards Their Clinical Applications. *Stem Cell Reviews and Reports*.

Yoshikiyo, K., Kratz, K., Hirota, K., Nishihara, K., Takata, M., Kurumizaka, H., … Jiricny, J. (2010). KIAA1018/FAN1 nuclease protects cells against genomic instability induced by interstrand cross-linking agents. *Proceedings of the National Academy of Sciences* **107**(50):21553–21557. doi: https://doi.org/10.1073/pnas.1011081107.

Yu, B.P. (1994). Cellular defenses against damage from reactive oxygen species. *Physiological Reviews*.

Yu, S., Pritchard, M., Kremer, E., Lynch, M., Nancarrow, J., Baker, E., … Richards, R.I. (1991). Fragile X genotype characterized by an unstable region of DNA. *Science*. doi: https://doi.org/10.1126/science.252.5009.1179.

Yumlu, S., Bashir, S., Stumm, J. and Kühn, R. (2019). Efficient Gene Editing of Human Induced Pluripotent Stem Cells Using CRISPR/Cas9. In: Luo, Y. (ed.). *CRISPR Gene Editing: Methods and Protocols*. New York, NY: Springer New York, pp. 137–151.

Zeitlin, S., Liu, J.P., Chapman, D.L., Papaioannou, V.E. and Efstratiadis, A. (1995). Increased apoptosis

and early embryonic lethality in mice nullizygous for the Huntington's disease gene homologue. *Nature Genetics*. doi: https://doi.org/10.1038/ng1095-155.

Zhang, N., An, M.C., Montoro, D. and Ellerby, L.M. (2010). Characterization of human Huntington's disease cell model from induced pluripotent stem cells. *PLoS Currents*. doi: https://doi.org/10.1371/currents.RRN1193.

Zhang, Y., Leavitt, B.R., Van Raamsdonk, J.M., Dragatsis, I., Goldowitz, D., MacDonald, M.E., … Friedlander, R.M. (2006). Huntingtin inhibits caspase-3 activation. *EMBO Journal*. doi: https://doi.org/10.1038/sj.emboj.7601445.

Zhang, Y., Li, M., Drozda, M., Chen, M., Ren, S., Sanchez, R.O.M., … Friedlander, R.M. (2003). Depletion of wild-type huntingtin in mouse models of neurologic diseases. *Journal of Neurochemistry*. doi: https://doi.org/10.1046/j.1471-4159.2003.01980.x.

Zhao, Q., Xue, X., Longerich, S., Sung, P. and Xiong, Y. (2014). Structural insights into 5′ flap DNA unwinding and incision by the human FAN1 dimer. *Nature Communications* **5**(1):5726. doi: https://doi.org/10.1038/ncomms6726.

Zhao, X.-N.N. and Usdin, K. (2018). FAN1 protects against repeat expansions in a Fragile X mouse model. *DNA Repair* **69**:1–5. doi: https://doi.org/10.1016/j.dnarep.2018.07.001.

Zhao, X., Gazy, I., Hayward, B., Pintado, E., Hwang, Y.H., Tassone, F. and Usdin, K. (2019). Repeat instability in the fragile x-related disorders: Lessons from a mouse model. *Brain Sciences*. doi: https://doi.org/10.3390/brainsci9030052.

Zhao, X., Zhang, Y., Wilkins, K., Edelmann, W. and Usdin, K. (2018). MutLγ promotes repeat expansion in a Fragile X mouse model while EXO1 is protective. *PLoS Genetics*. doi: https://doi.org/10.1371/journal.pgen.1007719.

Zhao, X.N. and Usdin, K. (2014). Gender and cell-type-specific effects of the transcription-coupled repair protein, ERCC6/CSB, on repeat expansion in a mouse model of the fragile X-related disorders. *Human Mutation*. doi: https://doi.org/10.1002/humu.22495.

Zhao, X.N. and Usdin, K. (2018). Timing of expansion of fragile X premutation alleles during intergenerational transmission in a mouse model of the fragile X-related disorders. *Frontiers in Genetics*. doi: https://doi.org/10.3389/fgene.2018.00314.

Zheng, H.T., Peng, Z.H., Li, S. and He, L. (2005). Loss of heterozygosity analyzed by single nucleoside polymorphism array in cancer. *World Journal of Gastroenterology*.

Zheng, Z., Li, A., Holmes, B.B., Marasa, J.C. and Diamond, M.I. (2013). An N-terminal nuclear export

signal regulates trafficking and aggregation of huntingtin (Htt) protein exon 1. *Journal of Biological Chemistry*. doi: https://doi.org/10.1074/jbc.M112.413575.

Zhou, H., Li, S.H. and Li, X.J. (2001). Chaperone Suppression of Cellular Toxicity of Huntingtin Is Independent of Polyglutamine Aggregation. *Journal of Biological Chemistry*. doi: https://doi.org/10.1074/jbc.M104140200.

Zhou, H., Wu, S., Joo, J.Y., Zhu, S., Han, D.W., Lin, T., … Ding, S. (2009). Generation of Induced Pluripotent Stem Cells Using Recombinant Proteins. *Cell Stem Cell*.

Zhou, W., Otto, E.A., Cluckey, A., Airik, R., Hurd, T.W., Chaki, M., … Hildebrandt, F. (2012). FAN1 mutations cause karyomegalic interstitial nephritis, linking chronic kidney failure to defective DNA damage repair. *Nature Genetics* **44**(8):910–915. doi: https://doi.org/10.1038/ng.2347.

Zuccato, C., Ciammola, A., Rigamonti, D., Leavitt, B.R., Goffredo, D., Conti, L., … Cattaneo, E. (2001). Loss of huntingtin-mediated BDNF gene transcription in Huntington's disease. *Science*. doi: https://doi.org/10.1126/science.1059581.

Zuccato, C., Tartari, M., Crotti, A., Goffredo, D., Valenza, M., Conti, L., … Cattaneo, E. (2003). Huntingtin interacts with REST/NRSF to modulate the transcription of NRSE-controlled neuronal genes. *Nature Genetics*. doi: https://doi.org/10.1038/ng1219.

Zuccato, C., Valenza, M. and Cattaneo, E. (2010). Molecular mechanisms and potential therapeutical targets in Huntington's disease. *Physiological Reviews*.