The impact of FAN1 on a CRISPR-Cas9 gene editing approach for Huntington's Disease

Laura Heraty

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

December 2023

Acknowledgments

Firstly, I would like to thank my supervisors; Vincent Dion, Tom Massey and Lesley Jones. Vincent; thank you for your constant support and guidance throughout the PhD, you have been like a cheerleader encouraging me along the way and I could not have asked for a more positive role model throughout the last few years. Tom; thank you for the helpful discussions and providing different perspectives, and for helping proof read this thesis. Lesley; you were such a positive role model for woman in science and I really appreciated all your support during the first couple years of my PhD.

Thank you to Andrew for being the best lab manager; providing any tips or advice and also for just being a general source of positivity.

Next I would like to thank all members of the Dion and HD lab, past and present, I could not have asked for a better group of people to have worked with during my PhD. Special thanks to Emma and Chris; you have both been invaluable in helping with various technical aspects throughout my PhD and I could not have managed without you. Work aside you are both amongst the funniest people I have ever met and it has been a pleasure to work with you. Alys; thank you for your constant support and friendship, and for proof-reading this thesis, I am sure you had more enjoyable things to be doing! Alvaro; your positive and jokey attitude always makes any work day better, so thank you! Aeverie; providing all the best baked-goods has been amazing, especially during the write up period. Thank you to Jasmine, Nina and Joe; you have been invaluable concerning all things stem cell and are always positive and happy to help. Finally, Antoine; I could not imagine a better fellow Dion lab PhD student to share this experience with. I have been so fortunate to work with not only an exceptional colleague but also an exceptional friend.

Thank you to all the DRI PhD students, you have been such a positive tight-knit group, it has been wonderful to be part of. I feel incredibly lucky to have made such good friends within the DRI, including former PhD students; Marieta, Katy and Dina. I am the last of us to finish! You three have provided such an incredible support network and I feel privileged to call you my friends.

Finally, I would like to thank Nicholas and my family. You have been a constant source of love and support and I could not have done this without you.

Contributions

Chris Smith generated and validated the *FAN1* knock-out lines described in Chapter 3 (section 3.2.4) and assisted in the clonal expansion and validation of *FAN1* knock-out lines in HD-iPSC iNeuron lines, described in Chapter 5.

Emma Randal assisted with the generation of PacBio SMRTbell sequencing libraries described in Chapter 3 and 4.

Joseph Stone generated and validated the HD-iPSC iNeuron lines described in Chapter 5.

Thesis Summary

Fifteen neurodegenerative or neuromuscular diseases are caused by expanded CAG/CTG repeats at multiple loci, all of which are without disease-modifying treatment. One example is Huntington's disease (HD), an autosomal dominant disease caused by an expanded CAG repeat within exon one of the *HTT* gene. In HD the length of the CAG repeat tract is inversely correlated with the age at onset, however, recent genome-wide association studies have identified several genetic modifiers of disease. One such modifier was mapped to the region containing *FAN1*. FAN1, a DNA repair nuclease, is thought to provide protection against somatic expansion with loss-of-function variants associated with an earlier age of HD onset.

Our laboratory has previously described a potential therapeutic approach for CAG/CTG diseases. This involves a CRISPR-Cas9 D10A nickase gene editing system, that is capable of contracting expanded repeats in human cell-lines, to non-pathogenic lengths. Given the central role of FAN1 in protecting against somatic expansion, the aim of this thesis was to characterise whether FAN1 modulates CRISPR-Cas9-induced contractions. This is important for not only understanding which proteins are involved in generating or preventing nickase-induced contractions but also for patient stratification to identify patients which would benefit from this therapy.

This thesis demonstrates that in a HEK293 reporter cell line knocked out for *FAN1*, there is an increase in Cas9 nickase-induced contractions, relative to $FAN1^{+/+}$ cells. This indicates that FAN1 acts to protect against the induction of contractions. Furthermore, this protective role for FAN1 requires FAN1 binding at CAG/CTG repeats, as a DNA-binding mutant identified was not capable of rescuing contractions to $FAN1^{+/+}$ levels. Additional functional domains of FAN1, including the nuclease and UBZ-binding domains are also implicated. Follow-up validation of HD-patient derived iPSC lines has been described in an effort to characterise whether this protective phenotype is recapitulated iPSC-derived neurons.

Abbreviations

AAV	Adeno-associated virus
AD	Alzheimer's disease
APC	Astrocyte precursor
ASO	Antisense oligonucleotide
BDNF	Brain-derived neurotrophic factor
BFP	Blue fluorescent protein
CAG	Cytosine-Adenine-Guanosine Trinucleotide
CCS	Circular consensus sequence
CMV	Cytomegalovirus
CNS	Central nervous system
CRISPR	Clustered Regularly Interspaced Short Palindromic Repeats
crRNAs	CRISPR-derived RNAs
CSF	Cerebrospinal fluid
CTG	Cytosine-Thymine-Guanosine Trinucleotide
DM1	Myotonic dystrophy 1
DNA	Deoxyribonucleic acid
DRPLA	Dentatorubral-pallidoluysian atrophy
DSB	Double-stranded break
EB	Embryoid body
EF1-α	Elongation factor alpha 1
EGF	Epidermal growth factor
ESCs	Embryonic stem cells
eQTL	expression Quantitative trait loci
FA	Fanconi anemia
FACS	Fluorescent activated cell sorting
FAN1	FANCD2 and FANCI associated nuclease 1

FANCC	Fanconi anemia group C
FANCD2	Fanconi anemia group D2
FANCI	Fanconi anemia group I
FECD	Fuchs endothelial corneal dystrophy
FRT	Flippase recognition target site
FXS	Fragile X syndrome
GABA	γ-aminobutyric acid
GD	Glutaminase deficiency
GFP	Green fluorescent protein
GWA	Genome wide association
HD	Huntington's Disease
HDL2	Huntington's disease-like 2
HEK293	Human embryonic kidney immortalised cell line
HR	Homologous recombination
HTT	Huntingtin gene
HTT	Huntingtin protein
ICL	Interstrand crosslink
iCRISPR	Inducible CRISPR
iNeurons	Inducible neurons
iPSCs	Induced pluripotent stem cells
IRES	Internal ribosome entry site
KIN	Karyomegalic interstitial nephritis
LIF	Leukaemia inhibitory factor
LIG1	DNA ligase 1
LS	Lynch syndrome
mESC	Mouse embryonic stem cell
mRNA	Messenger RNA
miRNA	Micro RNA
mHTT	Mutant Huntingtin protein

MLH1	MutL homolog 1
MLH3	MutL homolog 3
MMC	Mitomycin C
MMR	Mismatch repair
MOI	Multiplicity of infection
MRI	Magnetic resonance imaging
MSH2	MutS homolog 2
MSH3	MutS homolog 3
MSH6	MutS homolog 6
MSNs	Medium spiny neurons
MTMR10	Myotubularin protein 10
NGN2	Neurogenin-2
NPC	Neural precursor
PARP	Poly (ADP-ribose) polymerase
PacBio	Pacific Biosciences
PAM	Protospacer adjacent motif
PD	Parkinson’s disease
PFA	Paraformaldehyde
PMS1	Post mitotic segregation increased 1
PMS2	Post mitotic segregation increased 2
PNS	Peripheral nervous system
PolyQ	Polyglutamine tract
RED	Repeat expansion diseases
RNA	Ribonucleic acid
RNP	Ribonucleoprotein complex
RT-qPCR	Reverse transcription -
SBMA	Spinal and bulbar muscular dystrophy
SCA	Spinocerebellar ataxia
siRNA	Small interfering RNA

sgCTG	Single guide CTG
sgRNA	Single guide RNA
shRNA	Short hairpin RNA
SMRT	Real time (PacBio sequencing platform)
SNP	Single nucleotide polymorphism
spPCR	Small-pool polymerase chain reaction
SSB	Single-stranded break
TNR	Trinucleotide repeat
TPR	Tetratricopeptide repeat
TS	Translesion synthesis
U2OS	Osteosarcoma epithelial cell line
UBZ	Ubiquitin-binding zinc finger
UTR	Untranslated region
VRR-NUC	Virus-type replication repair nuclease
XRCC1	X-Ray Repair Cross-Complementing Protein 1
ZMV	Zero-magnetic wave guide

Table of Contents

List of Figures

List of Tables

1 General Introduction

1.1.1 CAG/CTG repeat expansion diseases

Trinucleotide repeat (TNR) disorders are a subset of repeat expansion diseases (REDs) caused by the repetition of short tandem repeats, 3 bases in length, within exonic or intronic regions of DNA (Den Dunnen 2018). The majority of TNR diseases are caused by CAG/CTG expansions. To date, there are 16 known loci where the expansion of CAG/CTG repeats causes disease (Table 1-1). These can be further subdivided into two categories. The first are as a result of CAG expansions in the coding region, leading to the formation of a functional protein with an expanded polyglutamine (polyQ) tract. This category accounts for 12 diseases, including; Huntington's disease (HD), multiple spinocerebellar ataxias (1, 2, 3, 6, 7, 8, 12 and 17) (SCAs), spinal and bulbar muscular atrophy (SBMA), dentatorubral-pallidoluysian atrophy (DRPLA) and a recently identified SCA (Tan et al. 2023). The second category are diseases caused by CAG/CTG expansions within the untranslated (UTR) or intronic regions of DNA, including; Glutaminase deficiency (GD), Myotonic dystrophy 1 (DM1), Huntington's disease-like 2, Fuchs endothelial corneal dystrophy (FECD) and SCA 8 (Khristich and Mirkin 2020; Tan et al. 2023).

CAG/CTG repeat disorders share a number of characteristic traits; the majority are marked by neuronal or neuromuscular degeneration, are inherited in an autosomal dominant manner and are currently without disease-modifying treatment. In all these diseases the length of the wild-type CAG/CTG tract is polymorphic, with repeat lengths ranging anywhere from 5-50 repeats (Khristich and Mirkin 2020; Donaldson et al. 2021; Bunting et al. 2022). Repeat lengths above a specific threshold are sufficient to cause disease. Importantly, in the majority of these diseases the CAG/CTG repeat length is inversely correlated with disease onset, and often disease progression. Furthermore, with the exception of a few, these diseases demonstrate somatic CAG/CTG expansion over time which has been linked to increasing disease severity (Castel et al. 2010). Whilst there are many overlapping features within this subclass, there is also a great deal of heterogeneity with regards to specific

cell-types affected, disease pathology and age at onset. This is likely due to that fact that the particular gene affected dictates, at least in part, cell-type specific sensitivity and degeneration (Den Dunnen 2018).

Table 1-1: Current known diseases caused by expanded CAG/CTG repeats. *Adapted from (Khristich and Mirkin 2020).*

Repeat Unit	Disease	Gene	Repeat location	Pathogenic length (full penetrance)	References
CAG	Dentatorubral–pallidoluysian atrophy(DRPLA)	*ATN1*	exon	49-93	(Koide et al. 1994)
CAG	Huntington disease (HD)	*HTT*	exon	40-250	(MacDonald et al. 1993)
CAG	Spinal and bulbar muscular atrophy (SBMA)	*AR*	exon	38-70	(Spada et al. 1991)
CAG	Spinocerebellar ataxia 1 (SCA1)	*ATXN1*	exon	45-81	(Orr et al. 1993)
CAG	Spinocerebellar ataxia 2 (SCA2)	*ATXN2*	exon	37-270	(Pulst et al. 1996)
CAG	Spinocerebellar ataxia 3 (SCA3)	*ATXN3*	exon	60-87	(Kawaguchi et al. 1994)
CAG	Spinocerebellar ataxia (SCA6)	*CACNA1A*	exon	20-33	(Zhuchenko et al. 1997)

CAG	Spinocerebellar ataxia (SCA7)	*ATXN7*	exon	37-460	(David et al. 1997)
CAG	Spinocerebellar ataxia (SCA12)	*PPP2R2B*	5′ UTR	51-78	(Holmes et al. 1999)
CAG	Spinocerebellar ataxia (SCA17)	*TBP*	exon	49-66	(Nakamura et al. 2001)
CAG	Glutaminase deficiency (GD)	*GLS*	5′ UTR	400-1500	(van Kuilenburg et al. 2019)
CAG	Spinocerebellar ataxia	*THAP11*	exon	45-100	(Tan et al. 2023)
CTG	Fuchs endothelial corneal dystrophy (FECD)	*TCF4*	intron	>50	(Wieben et al. 2012)
CTG	Huntington disease-like 2 (HDL2)	*JPH3*	3′ UTR	>41	(Holmes et al. 2001)
CTG	Myotonic dystrophy 1 (DM1)	*DMPK*	3′ UTR	50-5000	(Harley et al. 1992)
CTG	Spinocerebellar ataxia 8 (SCA8)	*ATXN8*	intron	90-250	(Koob et al. 1999)

1.2 An Introduction to Huntington's Disease (HD)

1.2.1 Historical background

Huntington's Disease (HD), previously referred to as 'Huntington's chorea', was first described in detail by American physician George Huntington in 1872 (Huntington, 1872). However, prior to this, several other physicians had noted a type of chorea that was likely an early description of HD. In 1832 English physician Elliotson described a form of chorea that occurred in adult life and was frequently associated with 'idiotism' (Vale and Cardoso 2015). A decade later Charles Oscar Waters presented an account of a disease likely to be HD in the *Practice of Medicine*. This report detailed the hereditary nature of the disease as well as noting that patients suffered from both motor and cognitive decline over time (Waters, 1842). As well as a description of disease symptoms both Elliotson and Oscar Waters noted that this form of chorea did not resolve with treatment.

Other reports by Charles Foreman and Johan Lund in following years, 1846 and 1860, supported this previous work (Vale and Cardoso 2015), but it was not until 1872 when Huntington published his essay 'On Chorea' in the *Medical and Surgical Reporter of Philadelphia* that neurologists began to take notice. Huntington drew the attention of the medical community to a form of 'hereditary chorea' which he observed was confined to a few families in Long Island, New York. Huntington noted three major points that characterised the disease; it tended to manifest only in adult life, patients suffered from 'insanity' and suicide and importantly it was hereditary in nature requiring only one affected parent (Huntington, 1872).

1.2.2 Genetics of HD

Despite a long history of documenting HD cases, the causative HD gene was not mapped to chromosome 4 until 1983 (Gusella et al. 1983). This study involved a community of interrelated families in Venezuela carrying the HD gene with a pedigree detailing over 3000 individuals, where the gene had been

inherited from a common ancestor. A decade later a new gene within 4p16.3, *IT15 (interesting transcript 15)*, containing an expanded polymorphic CAG trinucleotide was identified as the causative HD gene. This was renamed the *Huntingtin (HTT)* gene (MacDonald et al. 1993). Almost all cases of HD occur in an autosomal dominant manner, with only 3% of cases reported as *de novo* (Paine 2015).

HD is caused specifically by an expanded CAG repeat tract within the first exon of the *HTT* gene (MacDonald et al. 1993). The length of the expanded repeat tract can be divided into 3 subtypes; non-pathogenic, intermediate, and pathogenic. Pathogenic repeats sufficient to cause disease are 36 CAG's or above, however 40 repeats are required for full disease penetrance (Bates et al. 2015). Similarly to other TNR disorders, an intermediate CAG threshold has been described for HD, where lengths between 25-35 repeats are not-pathogenic but are considered prone to further expansion in the germline, which may lead to disease in subsequent generations (Goldberg et al. 1993).

1.2.3 Epidemiology

The prevalence of HD worldwide is roughly 3 in 100,000, however this varies considerably upon geographical location (Pringsheim et al. 2012). For example, the prevalence in western populations is estimated to be 10.6-13.7 per 100,000, compared with Japan, Taiwan and Hong Kong with reported rates between 1-7 per million (McColgan and Tabrizi 2018). The rates of HD in the United Kingdom are reported to be 12.3 per 100,000 (Evans et al. 2013). Differences between frequency of HD based on geographic location relates to differences at the *HTT* locus between ethnic groups. These populations with high HD rates tend to have longer *HTT* CAG repeat lengths compared with individuals of African and Asian descent. This is likely due to a higher prevalence of intermediate alleles in the population, which are prone to *de novo* mutations (Semaka et al. 2006; Rawlins et al. 2016). It has been suggested that European migration has contributed significantly to patterns of HD around the world, which is supported by increased incident rates in Caucasians or individuals of north-west European descent (Kay et al. 2017).

1.2.4 Clinical progression

HD is considered a disease of mid-life with typical age at onset between 30-50 years, though there are also rarer juvenile cases or those that appear later in life (Roos 2010). The presence of chorea, or involuntary motor movements, is the characteristic symptom of disease and often, when it becomes progressively worse, is the symptom that prompts investigation and subsequent diagnosis. However, subtle motor and cognitive symptoms do occur up to 10-15 years prior to diagnosis in the 'prodromal' disease stage (Ross et al. 2014; Bates et al. 2015). In particular, cognitive and psychiatric disturbances are common prior to formal diagnosis in pre-manifest patients with reports of impaired emotional recognition, depression and irritability, to name a few (McColgan and Tabrizi 2018). After the onset of chorea, the average length of disease is 17-20 years. During this time patients become severely disabled and increasingly dependent on others for care, experience worsening chorea as well as further impairment in cognitive function (Walker 2007; Roos 2010).

1.2.5 Neuropathology of HD

Neuropathological findings have demonstrated that HD is considered a multisystem neurodegenerative disease, affecting multiple brain structures (Rüb et al. 2016). In fact, HD patient brains have been reported to weigh between 20-30% less than healthy controls, dependent upon disease severity (Vonsattel, Jean-Paul; DiFiglia 1998). The area of the brain most severely affected by pathology is the striatum, consisting of the caudate nuclease and putamen, which demonstrate selective bilateral atrophy (De La Monte et al. 1988; Aylward et al. 1997; Vonsattel et al. 2008). The striatum is part of the basal ganglia, along with the globus pallidus, subthalamic nucleus and substantia nigra (Smith et al. 1998). This set of subcortical nuclei that form the basal ganglia are important in the regulation of movement, mood, learning and memory (Graybiel 1998). The neurons that make up the striatum fall into two classifications; the medium spiny neurons (MSNs) which account for 80-85% of the neuronal population in humans and primates (Johnston et al. 1990;

Gerfen 1992). MSNs are projection neurons expressing the inhibitory neurotransmitter γ-aminobutyric acid (GABA), extending to the globus pallidus and substantia nigra (Ehrlich 2012). These are the neurons most vulnerable to selective degeneration in HD (Reiner et al. 1988; Plotkin and Surmeier 2015). The remaining neuronal cell types are a mix of heterogeneous interneurons which can modulate the function of MSNs, and are relatively spared from HD pathology (Graveland et al. 1985; Carmichael and Lockhart 2012).

A pathological grading system has been developed based on the pattern of striatal neurodegeneration observed in HD cases, which closely correlates with disease symptoms (Vonsattel et al. 1985). This consists of 5 grades, from 0-4, with each increasing grade demonstrating more severe pathology. By grade 4 authors observed 95% or more MSN loss (Vonsattel et al. 1985). However, given the heterogenous and multisystem nature of HD, marking only striatal pathology can be limited and indeed pathology is observed in other brain regions including; the globus pallidus (Lange et al. 1976; Douaud et al. 2006), sub-thalamic nucleus (Guo et al. 2012), the cerebral cortex (De La Monte et al. 1988; Cudkowicz and Kowall 1990; Macdonald and Halliday 2002; Hadzi et al. 2012), the white matter (De La Monte et al. 1988; Aylward et al. 1998) and cerebellum (Rüb et al. 2013). More recently, advances in magnetic resonance imaging (MRI) have allowed for the monitoring of pathological changes *in vivo* (McColgan and Tabrizi 2018). One such study, TRACK-HD, involved a 12-month longitudinal investigation of pre-manifest individuals with the *HTT* mutation or early symptomatic HD patients. Strikingly, imaging revealed increased rates of whole brain atrophy in the pre-manifest cohort, relative to age-matched controls. Additionally, imaging in early-HD patients also revealed increased atrophy relative to both controls and pre-manifest individuals, significantly in the caudate. These findings are in line with other MRI studies monitoring brain atrophy in early-HD patients (Henley et al. 2009; Hobbs et al. 2010).

A key pathological hallmark of HD at the cellular level, is nuclear and cytoplasmic mutant HTT (mHTT) inclusion bodies (Davies et al. 1997; DiFiglia et al. 1997; Gutekunst et al. 1999). These have been observed in

immunohistochemical studies of HD-patient brains (Gutekunst et al. 1999; Herndon et al. 2009) and HD transgenic and knock-in mouse models (Hodgson et al. 1999; Landles et al. 2010). Whilst mHTT inclusion bodies were initially thought to be the toxic species in HD, leading to selective cell death (Davies et al. 1997; Davies et al. 1999), this theory is under debate. Indeed, some studies indicate that the presence of mHTT inclusion bodies are protective (Saudou et al. 1998) and correlate with increased neuronal survival (Arrasate et al. 2004). Interestingly, these structures are more frequently observed in the cortex and spared interneurons of the striatum, compared with MSNs (Gutekunst et al. 1999; Kuemmerle et al. 1999). Despite current debate the mHTT protein clearly plays a role in HD pathogenesis and mHTT inclusions are a hallmark of disease pathology.

1.2.6 The Huntingtin (HTT) protein and mutant HTT (mHTT)

The non-pathogenic HTT protein is roughly 3,144 amino acids with a molecular weight of 384 kDa (Bates et al. 2015). *HTT* is widely expressed in different tissues with the highest expression in the brain and testes (Li et al. 1993; Paine 2015). There are two main *HTT* isoforms with comparable protein coding sequences and protein size, but these differ in size and sequence at the 3′ UTR. The longer 13.7 Kb transcript is the predominant transcript in the human brain (Lin et al. 1993). The normal function of the HTT protein is still under investigation. However, HTT has been demonstrated to interact with a number intracellular organelles, including; the endoplasmic reticulum and Golgi apparatus (DiFiglia et al. 1995; Hilditch-Maguire et al. 2000). Roles for the function of HTT include; transcriptional regulation (Thompson 2003; Zuccato et al. 2003), autophagy (Wong and Holzbaur 2014), vesicle transport (Caviston et al. 2011), and protein trafficking (Huang et al. 2004), to name a few (see Saudou and Humbert 2016 for review).

Additional roles for HTT in normal cellular function indicate that it is important in providing neuroprotection against apoptotic stimuli and excitotoxicity (Rigamonti et al. 2000; Leavitt et al. 2006). Wild-type HTT function has also

been linked to the upregulation of transcription of cortico-striatal brain-derived neurotrophic factor (BDNF) (Zuccato et al. 2001). BDNF is important in the function of cortico-striatal synapses and acts as a pro-survival factor with striatal neurons particularly dependent upon this for development and proper functioning (Baydyuk and Xu 2014). Furthermore, wild-type HTT appears to be required for BDNF vesicle trafficking, with mHTT showing deficits in trafficking, resulting in neuronal toxicity (Gauthier et al. 2004).

Interestingly, HTT is also critical for embryonic development in mice, with homozygous inactivation of homologous *Hdh* gene proving embryonic lethal (Duyao et al. 1995; Zeitlin et al. 1995). Comparatively, mice heterozygous for inactivation of *Hdh* demonstrated no developmental deficits compared with controls, indicating only one functional copy of *Hdh* is required for development. Furthermore, rare instances of HD patients who are homozygous for the HD mutation are born with no developmental defects, indicating that even the mHTT protein can perform the normal developmental function of HTT (Wexler et al. 1987).

The HTT protein is highly organised, with a series of linear domains, separated by disordered regions. The ordered domains are clusters of α-helical HEAT (Huntingtin, elongation factor 3, protein phosphatise 2A and TORI) repeats (Andrade and Bork 1995). Within the disordered regions are a number of post-translation modification sites which can be proteolytically cleaved by a number of proteases including; caspases (Goldberg et al. 1996; Kim et al. 2001; Hermel et al. 2004; Ratovitski et al. 2009), cathepsins (Lunkes et al. 2002), calpain (Gafni and Ellerby 2002) and the metalloproteinase MMP10 (Tebbenkamp et al. 2012). These cleavage sites are also present in mHTT (Saudou and Humbert 2016). Importantly, mHTT has been demonstrated to be more prone to nucleolytic cleavage than wild-type HTT (Goldberg et al. 1996). A variety of N-terminal mHTT fragments, generated by proteolytic cleavage, have been described and are implicated in HD pathogenesis. One such fragment includes an N-terminal fragment consisting of 100 amino acids, referred to as HTT exon 1 (Bates et al. 2015). This demonstrates a propensity to self-aggregate and is implicated in HD pathogenesis (Jayaraman et al. 2012; Trepte et al. 2014; Wagner et al. 2018). As an array of mHTT structures

have been described, including; monomeric, oligomeric and fibrillar aggregates and inclusion bodies, there is still debate regarding how these each contribute to disease pathology (Wanker et al. 2019).

1.2.7 mHTT and cellular pathology

The expression of mHTT has been associated with a range of cellular pathology. Transcriptional dysregulation in HD is a major hallmark of disease (Malla et al. 2021), with altered transcription demonstrated in HD patient brains relative to controls. Interestingly the greatest number of differentially expressed genes were in the caudate of the striatum, in agreement with disease pathology (Hodges et al. 2006). Studies have demonstrated that the mHTT can interact with regulators of the transcription machinery, amongst which are the cAMP response element-binding protein (CREB) and p53, which can repress transcription and lead to neuronal dysfunction and death (Steffan et al. 2000; Dunah et al. 2002; Schaffar et al. 2004; Zhai et al. 2005; Labbadia and Morimoto 2013). Aside from transcriptional dysregulation, the aggregation of mHTT in the cytoplasm has been proposed to cause wide-spread alterations in proteostasis (Margulis and Finkbeiner 2014). This is thought to occur due to the sequestration of proteins essential in maintaining proteostasis by mHTT aggregates (Park et al. 2013; Gasset-Rosa et al. 2017). Furthermore, cellular chaperones important for the disaggregation of misfolded proteins, have also been shown to be sequestered by mHTT aggregates (Park et al. 2013). It has been proposed that sequestration of chaperones by mHTT, reducing chaperon binding to other misfolded proteins causes their diversion to the proteosome for degradation. This becomes overwhelmed leading to proteostasis collapse and cellular dysfunction (Soares et al. 2019). Additionally, mHTT is associated with increasing mitochondrial dysfunction (Song et al. 2011; Reddy and Shirendeb 2012), synaptic dysfunction (Nithianantharajah and Hannan 2013), and axonal transport (Her and Goldstein 2008; Shirendeb et al. 2012) (Figure 1-1).

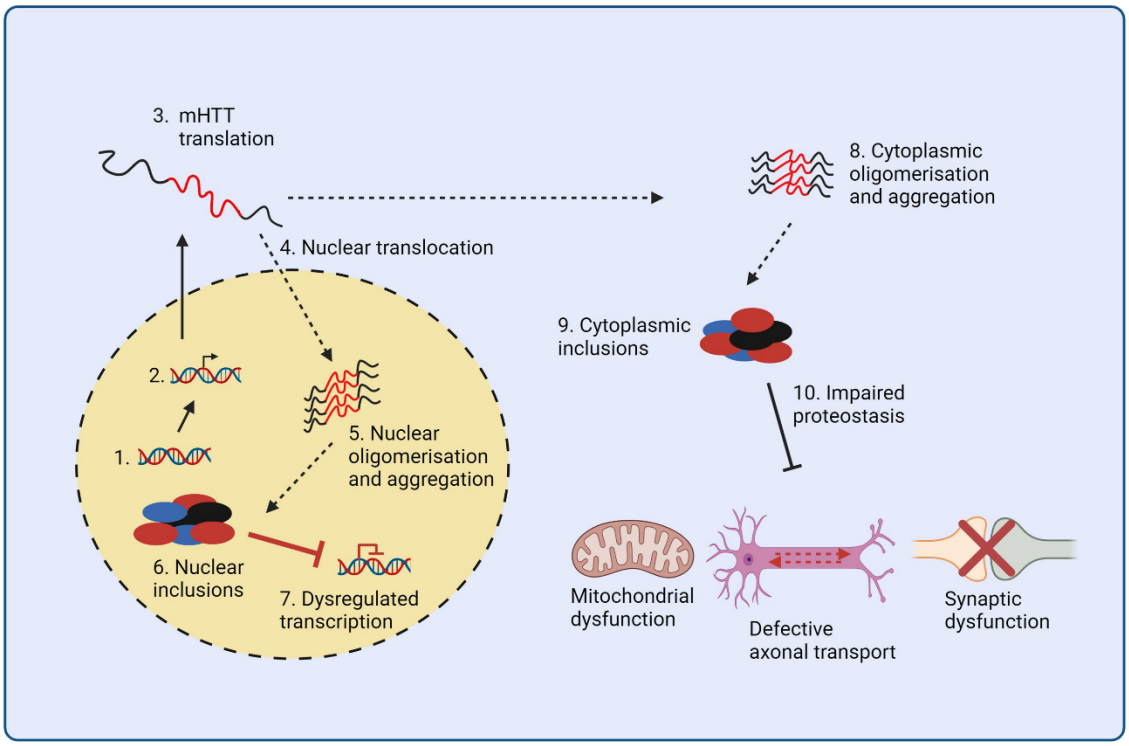


Figure 1-1: Molecular pathology of Huntington's Disease. *A number of cellular processes are dysregulated due to the expression of mHTT. mHTT undergoes transcription and then mRNA is translocated to the cytoplasm for translation to a full length mHTT protein, including the expanded polyQ tract (1-3). mHTT can form aggregates which can translocate to the nuclease which can further self-associate to form nuclear inclusions (4-6). Aggregation of mHTT in the nuclease can lead to transcriptional dysregulation (7). Cytoplasmic mHTT can also form aggregates and inclusion bodies which can act to sequester proteostasis machinery, leading to impaired proteostasis (8-10). Abberant mHTT can also lead to global cellular impairments, including mitochondrial dysfunction, defective axonal transport and synaptic dysfunction. Figure generated in BioRender.*

1.3 Genetic Modifiers of HD

In HD the length of the CAG tract is inversely correlated with age at onset, with increased repeat lengths associated with earlier onset (Duyao et al. 1993; Kremer et al. 1993; Snell et al. 1993; Trottier et al. 1994). Whilst the CAG tract length is the primary determinant for HD age at onset it only accounts for ~60% variance at motor onset, with the remaining variability accounted for by genetic and environmental factors (Wexler et al. 2004; McAllister et al. 2021). Recent efforts to collect data from large cohorts of individuals with HD has led to the identification of genetic modifiers through genome-wide association (GWA) studies (Lee et al. 2015; Lee et al. 2019).

GWA studies allow for unbiased analysis and the detection of variation in a patient's genome that may give insight into genes that can alter HD pathogenesis (Holmans et al. 2017). An initial GWA study was conducted to determine factors that affect the residual age of motor onset of HD patients. This value represents the difference between the observed age of onset and expected age based upon the *HTT* CAG length of each individual. This study, comprised of 4,082 HD individuals (GWA1-3) examining a CAG repeat size range of 40-53 repeats and identified two independent effects at the chromosome 15 locus containing *FANCD2 and FANCI associated nuclease 1* (*FAN1*) and *myotubularin protein 10* (*MTMR10*) (Lee et al. 2015). One signal was associated with an accelerated onset by 6.1 years and the other delayed onset by 1.4 years. An additional genome-wide significant signal was identified at chromosome 8, which hastened disease onset by 1.6 years. A follow-up study comprised of 9,064 individuals (GWA12345) supported these previous findings and also identified an additional set of significant associations at chromosome's 3, 5, 7, 19, and 11. Furthermore, an additional 2 modifiers of opposing effects were identified at the chromosome 15 locus (Lee et al. 2017; Lee et al. 2019).

Interestingly, many of the loci that reached genome-wide significance contain genes associated with DNA repair, indicating a direct role of DNA repair in modulating HD age of onset, perhaps through modulation of somatic instability

(see 1.7). For example, signals at chromosome 3, 2, 5, 7 and 19 have been linked to *MutL homolog 1* (*MLH1*), *Post mitotic segregation increased 1* (*PMS1*), *MutS homolog 3* (*MSH3*), *Post mitotic segregation increased 2* (*PMS2*), and *DNA ligase 1* (*LIG1*), respectively. All of these factors are associated with the DNA Mismatch repair (MMR) pathway (Iyer and Pluciennik 2021). Furthermore, the four independent signals at chromosome 15 have been linked to *FAN1*, with two tagged SNP alleles that cause missense variants and are predicted to be deleterious for protein function (Lee et al. 2019; McAllister et al. 2022). The other two protective variants correspond to *cis*-eQTLs for increased expression of *FAN1* in the cortex (Lee et al. 2019).

1.4 FAN1

FAN1, previously a hypothetical protein (KIAA1018), was initially identified as part of a cDNA screen to determine coding sequences of unidentified human genes expressed across central nervous system (CNS) tissues (Nagase et al. 1999). Subsequent secondary structure prediction analysis based on its consensus sequences indicated that KIAA1018 was likely to function in DNA repair or maintenance of genome stability (Kinch et al. 2005). This was supported in 2007 when Cannavo *et al* identified KIAA1018 in the interactome of MMR proteins MLH1 and PMS2. Recent studies have shed light on the nature of this interaction with MLH1, but any direct interaction with PMS2 has yet to be identified. After the discovery of KIAA1018 in the interactome of MMR proteins several independent groups sought to identify the action of this novel protein.

1.4.1 Structure of FAN1

Human FAN1 consists of four key functional domains; an N-terminal ubiquitin-binding zinc finger (UBZ) domain, a SAF-A.B, Acinus and PIAS (SAP) DNA binding domain, a tetratricopeptide repeat (TPR) domain and a C-terminal virus-type replication repair nuclease (VRR_NUC) domain (MacKay et al. 2010; Smogorzewska et al. 2010). Interestingly, all domains, with the

exception of the UBZ domain, are highly conserved in bacteria and unicellular eukaryotes (Jin and Cho 2017).

1.4.2 FAN1 nuclease activity

FAN1 has both 5′ structure-specific endonuclease activity and 5′-3′ exonuclease activity. Although FAN1 nuclease activity demonstrates a preference for 5′ flap DNA structures, it is capable of cleaving other structures, including; 3′ flaps, replication forks, duplexed DNA, nicked DNA, crosslinked DNA, gapped DNA substrates and DNA loop-out structures (Kratz et al. 2010; Liu et al. 2010; MacKay et al. 2010; Smogorzewska et al. 2010; Yoshikiyo et al. 2010; Wang et al. 2014; Zhao et al. 2014; Kim et al. 2020; Deshmukh et al. 2021). The ability of FAN1 to cleave multiple DNA intermediates reflects its versatility as a DNA repair protein, capable of acting in multiple repair pathways.

1.4.3 A role for FAN1 in DNA repair

In 2010 immunoprecipitation data established direct protein interactions between FAN1 and Fanconi anemia group D2 (FANCD2), a member of the Fanconi Anemia (FA) DNA repair pathway, in addition to corroborating previous findings that FAN1 interacts with MLH1 and PMS2 (Liu et al. 2010; MacKay et al. 2010; Smogorzewska et al. 2010).

The FA pathway is a specialized DNA repair pathway responsible for the resolution of interstrand crosslinks (ICLs). ICLs are highly toxic lesions resulting in covalent bond formation between opposite DNA strands, which can stall replication and transcription machineries (Thongthip et al. 2016). This pathway comprises 19 known core proteins, as well as many other associated proteins. Germline inactivation of any of these core genes is sufficient to cause FA, a genetic disease where patients are predisposed to bone marrow failure and cancer (Ceccaldi et al. 2016). Interestingly, inactivation of *FAN1* does not cause FA, instead it causes karyomegalic interstitial nephritis (KIN), a rare chronic kidney disease, suggesting that whilst FAN1 can associate with members of the FA pathway it is not essential (Zhou et al. 2012). Modelling in cell systems demonstrate that the absence or depletion of FAN1 in the

presence of ICL-inducing agents, mitomycin C (MMC) and cisplatin, sensitise cells to these lesions (Kratz et al. 2010; Liu et al. 2010; MacKay et al. 2010; Smogorzewska et al. 2010; Thongthip et al. 2016). Furthermore, work in *Caenorhabditis elegans* indicated this was an evolutionarily conserved function (MacKay et al. 2010; Smogorzewska et al. 2010).

Repair of ICLs by FA repair occurs via an initial lesion recognition step facilitated by DNA-binding protein Fanconi anemia complementation group M (FANCM). FANCM's ability to detect ICLs is dependent on its phosphorylation by ataxia telangiectasia (ATR) and interactions with other associated proteins (Castella et al. 2015). Upon binding, FANCM recruits the FA core complex, comprised of 14 key proteins, which acts as a ubiquitin ligase for FANCD2 and Fanconi anemia complementation group I (FANCI) (Meetei et al. 2005; Kim et al. 2008). Subsequent activation of FANCD2-I is required for nucleolytic incision and 'unhooking' of ICLs in a downstream process. After the ICLs are 'unhooked', translesion synthesis (TS) and homologous recombination (HR) steps occur to complete repair (for full review see Ceccaldi et al. 2016).

It is proposed that FAN1 is associated with the FA pathway through is recruitment to ICLs by monoubiquitinated FANCD2, via an interaction with FAN1's UBZ domain. This is supported by evidence demonstrating that mutations within the UBZ domain results in a failure of FAN1 to co-localise with FANCD2 at sites of DNA damage (Liu et al. 2010; Smogorzewska et al. 2010; Kratz et al. 2010). FAN1 is thought to resolve ICLs at potentially multiple stages; including ICL 'unhooking', resection of unhooked nucleotides and D-loop incision during the process of HR (Jin and Cho 2017). Whilst FAN1 can act in association with the FA pathway, evidence has demonstrated that FAN1 is able to repair ICLs independently of this. Two studies have shown FAN1 alone is able to cleave ICL substrates in biochemical-based assays (Wang et al. 2014; Zhao et al. 2014). Furthermore, double knock-out of *FAN1* in conjunction with *Fanconi anemia complementation group C* (*FANCC)* demonstrate increased sensitivity to cisplatin, relative to *FAN1*$^{-/-}$ cells (Yoshikiyo et al. 2010). These data are supported by findings that rescuing *FAN1*$^{-/-}$ lines with a C44/C47A mutation, which abolishes the FANCD2-FAN1 interaction, is sufficient to rescue MMC-induced sensitivity (Zhou et al. 2012;

Goold et al. 2021). Taken together this indicates that FAN1 is also able to localise and resolve ICLs independently of FANCD2.

Interestingly, the interaction between FAN1 and FANCD2 has been implicated in replication fork restart during replication stress, preventing fork collapse (Chaudhury et al. 2014). In this case, FAN1 nuclease activity appears to be tightly controlled by FANCD2. In the presence of FANCD2, in concert with DNA repair factors Mre11 and BLM helicase, FAN1 is recruited to sites of stalled forks, promoting replication fork restart via its nuclease activity (Chaudhury et al. 2014). However, in the absence of FANCD2, FAN1 is still able to localise to stalled replication forks but is not regulated and as such, this unrestrained access can result in fork degradation. Additional studies have also demonstrated that FAN1 nuclease activity is required for restraining replication fork progression and maintaining genome integrity during replication stress and replication fork stalling (Lachaud et al. 2016; Porro et al. 2017). How FAN1 can both promote replication fork restart and limit its progression remains to be seen, but it is clear that FAN1 is a diverse nuclease capable of interacting within multiple DNA repair mechanisms.

1.5 Mismatch Repair

In humans, deficiency of certain MMR proteins leads to Lynch Syndrome (LS). LS is the most commonly inherited cancer syndrome associated with increased risk of colonic cancer and other forms, including; endometrial, ovarian and stomach cancer (Tiwari et al. 2016; Liu et al. 2017a). The MMR pathway is highly conserved from *Escherichia coli* to *Saccharomyces cerevisiae* to mammals (Wildenberg and Meselson 1975; Wagner and Meselson 1976; Kramer et al. 1989; Reenan and Kolodner 1992; Prolla et al. 1994; Iyer et al. 2006). The canonical role of MMR involves recognition and subsequent repair of small DNA mismatches and larger insertion/deletion loop structures, which occur during replication due to errors in DNA synthesis (Modrich and Lahue 1996; Fishel and Wilson 1997; Modrich 2006; Jiricny 2013). MMR is a multi-step process involving; mismatch strand recognition,

recruitment of repair proteins and formation of a MMR protein complex, strand nicking, resection and subsequent gap filling and ligation (Usdin et al. 2015).

Eukaryotic MMR processing relies on the MutS and MutL complexes, which are involved in mismatch recognition and subsequent repair. There are 2 eukaryotic MutS complexes; MutSα (MSH2-MSH6) and MutSβ (MSH2-MSH3) (Modrich 2006). MutSα recognises single base-base mismatches or single-nucleotide loops, but demonstrates reduced affinity for larger insertion/deletion loops. In comparison, MutSβ is capable of recognising larger insertion/deletion loops between 2 to 10 bp (Drummond et al. 1995; Fishel and Wilson 1997; Palombo et al. 1996; Genschel et al. 1998). After mismatch recognition, by MutSα or MutSβ, a second complex, the MutL complex, is recruited to the mismatch. There are 3 eukaryotic MutL complexes; MutLα (MLH1-PMS2), MutLβ (MLH1-PMS1) and MutLγ (MLH1-MLH3). MutLα represents ~90% of MLH1 in human cells and is the primary MutL complex in MMR (Modrich 2006). Indeed, whilst MutLγ is more lowly expressed it has been demonstrated to support mismatch repair of base-base mismatches *in vitro* and can partially restore MMR function in the absence of MutLα (Flores-Rozas and Kolodner 1998; Cannavo et al. 2005; Chen et al. 2005). Comparatively, a role for MutLβ in canonical MMR has not been demonstrated (Räschle et al. 1999). Both MutLα and MutLγ possess endonuclease activity and once recruited to the site of the mismatch, catalyse strand incision which directs downstream repair of the mismatch (Iyer and Pluciennik 2021).

1.6 CAG/CTG repeat instability

A huge body of evidence regarding CAG/CTG disorders is concerned with the inherent instability of CAG/CTG repeats and the potential role this plays in disease pathogenesis (Wheeler and Dion 2021).

1.6.1 Formation of non-canonical DNA structures

A key aspect of expanded CAG/CTG repeat tracts is their tendency to adopt stable non-canonical structures, which are thought to be essential intermediates involved in driving repeat instability (Marquis Gacy et al. 1995;

Pearson and Sinden 1996; Pearson and Sinden 1998; Pearson et al. 2002). These structures may form transiently throughout DNA repair, replication and transcription, where DNA is unwound and exposed as single-stranded DNA (ssDNA) (Wheeler and Dion 2021). CAG/CTG repeat structures are capable of forming imperfect hairpins, where slipped-out ssDNA self-associates to form Watson-Crick base pairs as well as base-base mismatches (Marquis Gacy et al. 1995; Liu et al. 2010a; Khristich and Mirkin 2020). Indeed studies propose that the terminal end of hairpins may result in potential 5′-CAG-3′ tri-loop or 5′-AGCA-3′ tetraloop structures. The formation of tri-loop or tetra-loop structures have been shown to have variable stability, with tetra-loop structures demonstrating increased stability, and reduced DNA slippage relative to tri-loops (Pan et al. 2017; Xu et al. 2020). Furthermore, tetra-loop structures containing mismatched A-A or T-T pairing at the terminal ends of hairpins has been suggested to be important in the recognition by DNA repair proteins and subsequent resolution and repair of these structures (Deshmukh et al. 2021b). Interestingly, sequence context within the repeat appears to contribute to structure formation. For instance, the presence of CAA repeat interruptions in pure CAG repeat tracts has been proposed to reduce the propensity of CAG repeats to adopt hairpin structures (Xu et al. 2020). Aside from hairpin structures CAG repeats may also form CAG loop-out structures, where the slipped-out CAG repeat region does not self-associate with itself, forming a loop rather than a DNA hairpin (Pearson et al. 2002). These structures have been proposed to be intermediates for repeat expansion or contraction, which can occur as a result of; DNA replication, transcription and DNA repair (Mirkin 2007; McMurray 2010; McGinty and Mirkin 2018).

1.6.2 Germline instability

CAG/CTG repeat instability can be broadly characterised into two main categories; germline/intergenerational and somatic instability. Germline instability describes the phenomenon commonly seen in TNR disorders where mutations occur during parent-to-offspring transmissions, leading to contraction or expansion events (Pearson et al. 2005). Germline expansion

events explains the high level of anticipation observed in these diseases, where the disease manifests earlier through successive generations, often with more severe phenotypes (Semaka et al. 2006). Importantly, not all repeat diseases demonstrate germline instability, but HD does (Pearson 2003). One early study examining germline expansion in a large HD kindred demonstrated in 62 individuals, where the length of the *HTT* CAG repeat could be examined on the parental chromosomes, that 80% of meiotic transmissions demonstrated changes in CAG repeat length (Duyao et al. 1993).

An important factor which determines the direction of intergenerational instability, is the parental gender. In HD, on average, paternal transmission demonstrates increased expansions rates with no change or small contractions observed in female transmission (Duyao et al. 1993; Trottier et al. 1994; Kremer et al. 1995; Ranen et al. 1995). Though rare expansion events arising from maternal transmission have been documented in a juvenile-HD onset patient. In this case the germline *HTT* tract underwent an expansion from 70 CAG's in the mother, to ~130 in her daughter (Nahhas et al. 2005). Interestingly, when comparing paternal versus maternal transmission, the average length of CAG size changes was larger in paternal germlines (Duyao et al. 1993). Furthermore, longer CAG repeat lengths are more prone to germline instability, with a longer CAG repeat more prone to larger expansions between paternal transmissions (Telenius et al. 1994; Kremer et al. 1995; Nørremølle et al. 1995; Ranen et al. 1995). Whilst the average CAG expansion between transmission has been reported to vary from ~2-9 repeats on average, depending on reports, much larger *HTT* CAG expansions have also been documented. The largest documented occurring through paternal transmission from a father with 54 CAGs to a daughter with juvenile HD-onset with a repeat length of 214 CAGs (Seneca et al. 2004).

1.6.3 Somatic instability

Whilst a person's inherited CAG/CTG repeat length determines whether or not they will develop disease, an increasing body of evidence has highlighted somatic instability as a key player in disease onset. One current school-of-

thought proposes a two-step disease model for HD; the first step involves inheritance of an expanded CAG repeat, which critically is at a length where it is prone to further somatic expansion. Secondly, throughout the lifetime of the individual somatic expansion occurs, with repeat lengths likely far surpassing that of the inherited length. This somatic expansion occurs until a critical threshold is reached in vulnerable cells, upon which cellular dysfunction and death occur (Donaldson et al. 2021; Hong et al. 2021) (Figure 1-2).

In HD, evidence of somatic instability has been described in human post-mortem tissue in areas associated with pathology, including the striatum, cerebral cortex and basal ganglia (Telenius et al. 1994; De Rooij et al. 1995; Kennedy et al. 2003; Shelbourne et al. 2007; Gonitel et al. 2008). Though the MSNs of the striatum are most vulnerable and demonstrate earliest cell death in HD (Reiner et al. 1988; Plotkin and Surmeier 2015), one study documented little somatic expansion in in the striatum of an end-stage HD patient (Kennedy et al. 2003). This finding is likely explained by the fact that by end-stage disease, there is extensive neuronal loss in the striatum, therefore those cells with the highest degree of somatic expansion have likely died. The same study presented results that support this idea by examining CAG repeat instability in two pre-symptomatic HD patients with no striatal atrophy, who had died prior to HD-onset. Both patients demonstrated a dramatic increase in CAG repeat length in striatal cells, in some instances with CAG repeats greater than 1000 units, compared with an inherited CAG size of 41 or 51 repeats (Kennedy et al. 2003). Further work by Swami and colleagues supports a model for somatic instability driving disease onset. This study examined somatic instability in the cortex of individuals with early and late-onset HD, where their age of onset deviated significantly from that predicted by their constitutive *HTT* CAG tract. Findings revealed increased levels of somatic instability in the cohort with earlier disease onset, indicating that somatic expansions impacts disease onset (Swami et al. 2009).

Due to limited tissue availability and HD-post mortem brains often demonstrating late-end disease, it is important to investigate somatic expansion in other models. Indeed, *in vivo* models supports findings in post-mortem tissue. The earliest description of somatic instability in a HD mouse

model was described in a series of transgenic mice; R6/1, R6/2 and R6/5 (Mangiarini et al. 1997; Møllersen et al. 2010). These mice contain a human 1.9 Kb genomic *HTT* fragment including the human promoter sequence and expanded CAG repeat in exon 1 (Mangiarini et al. 1996). In all lines there was tissue-specific somatic expansion in brain regions associated with human pathology, including the striatum and cerebral cortex (Mangiarini et al. 1997). Knock-in mouse models of HD have also reported similar tissue-specific somatic expansion. In these lines exon 1 of the mouse *Hdh* gene is replaced with a chimeric *Hdh*:*HTT* exon 1 constructed from a *HTT* gene containing expanded CAG repeats (White et al. 1997). Somatic expansion has been reported in multiple brain regions, including striatum and cortex, with a range of CAG repeats; Hdh^{Q72} (Kennedy and Shelbourne 2000), Hdh^{Q80} (Ishiguro et al. 2001), Hdh^{Q92}, Hdh^{Q111} (Wheeler et al. 1999; Lee et al. 2011) and Hdh^{Q150} (Gonitel et al. 2008). These results are encouraging as they demonstrate *in vivo* models are capable of mimicking certain aspects associated with human disease.

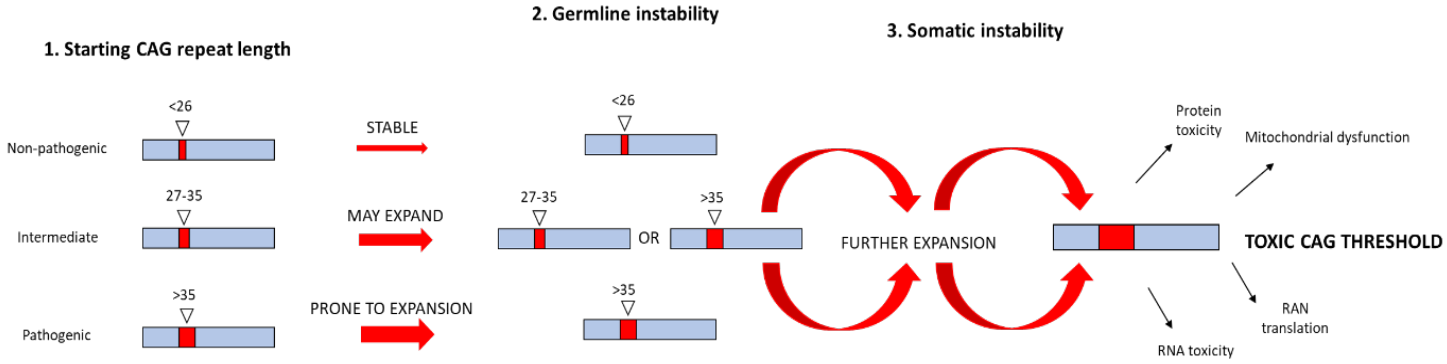


Figure 1-2: A two-step model of disease pathogenesis. *(1) An individual may inherit a non-pathogenic, intermediate or pathogenic HTT allele. Both intermediate and pathogenic alleles demonstrate the ability to expand in the germline, with longer repeats particularly prone to expansion (2). The inherited CAG repeat length continues to expand throughout the lifetime of an individual with increasing somatic expansion occurring until repeat units reach a toxic CAG threshold, at which point cellular dysfunction and death occur (3).*

1.7 DNA Repair and CAG/CTG instability

A range of DNA repair processes have been implicated in CAG/CTG repeat instability, both germline and somatic instability. Whilst some DNA repair factors demonstrate a protective effect, a range of repair proteins have also been shown to drive CAG/CTG instability. This section details DNA proteins or pathways implicated in instability, with a focus on HD.

1.7.1 Mismatch Repair

In mammalian systems MMR is a driver of CAG/CTG instability with a predominant bias towards expansion events (Schmidt and Pearson 2016). *In vivo* HD data, including transgenic and knock-in models, have demonstrated that MSH2 (MutS homolog 2) is required for somatic *Htt* CAG expansion, where loss of Msh2 ablates expansion in striatum of mice (Manley et al. 1999; Wheeler et al. 2003; Kovalenko et al. 2012). Furthermore, studies indicate Msh2 contributes to germline instability, with a loss of Msh2 in HD mice abolishing paternal germline expansion (Kovtun and McMurray 2001; Dragileva et al. 2009). MSH3 (MutS homolog 3), similarly to MSH2, is implicated in somatic instability, and to a lesser extent germline instability. Loss of Msh3 in HD mice completely ablates somatic expansion (Owen et al. 2005; Dragileva et al. 2009; Tomé et al. 2013). Comparatively, in a HD knock-in mouse model, loss of *Msh3* had only a moderate impact on germline instability, compared with *Msh2*$^{-/-}$ mice (Dragileva et al. 2009). Data regarding MSH6, (MutS homolog 6), is much less clear with *Msh6-/-* mice demonstrating tissue specific reduction of somatic *Htt* CAG expansion in one study (Owen et al. 2005). Contrastingly, another study demonstrated no effect of Msh6 on *Htt* CAG somatic expansion in the striatum, but a potential role for preventing contractions in the paternal germline (Dragileva et al. 2009). Taken together these data indicate a clear role for MutSβ in somatic and germline expansion, with the role for MutSα being less clear.

Regarding the role of MutL complexes, both *Mlh1* and *Mlh3* are required for somatic expansion in the striatum in HD knock-in mice (Pinto et al. 2013). However, no effect on germline expansion could be determined as both Mlh1 and Mlh3 deficient mice are sterile (Edelmann et al. 1996; Lipkin et al. 2002).

These data indicate MutLγ is required for somatic expansion in HD. This is interesting given that MutLγ is lowly expressed in human cells (Modrich 2006). However, this data is supported by findings in mouse embryonic stem cells (mESCs) modelling Fragile X syndrome (FXS) (Hayward et al. 2020). FXS is the most common inherited form of mental retardation and is caused by CGG expansions in the 5′ UTR region of the *FMR1* gene (Garber et al. 2008). Data from Hayward et al. 2020 demonstrated that the generation of a point mutation, D1185N, within the endonuclease domain of MLH3 is sufficient to eliminate somatic expansion in culture. This indicates that the endonuclease activity of MLH3 is required for somatic expansion. Furthermore data from a DM1 mouse model, another CAG/CTG disorder, have demonstrated that somatic expansion in *Pms2*$^{-/-}$ mice is reduced by ~50% compared with *Pms2*$^{+/+}$ mice (Gomes-Pereira et al. 2004). These data suggest that both MutLγ and MutLα, together with MutSβ, act to promote somatic expansion, to varying extents.

1.7.2 FAN1

Whilst MMR seems to be the main driver of somatic instability, FAN1 displays the opposite effect, protecting against instability and expansion events. This phenotype has been demonstrated in multiple systems, where loss of FAN1 increases somatic expansions in cellular and mouse models (Zhao and Usdin 2018; Goold et al. 2019; Kim et al. 2020; Loupe et al. 2020; Goold et al. 2021; Zhao et al. 2021; McAllister et al. 2022). This protective effect of FAN1 was first described in a mouse model of FXS (Zhao and Usdin 2018). *Fan1*$^{-/-}$ mice demonstrated increase somatic expansions over 3 and 6 months in the liver and brain, relative to their *Fan1*$^{+/+}$ litter mates. Interestingly, no effect on germline instability was reported in these mice (Zhao and Usdin 2018).

Further studies investigating FAN1's role in somatic instability at CAG/CTG repeats has been explored in both replicating and non-replicating cells models. In human osteosarcoma epithelial cells (U2OS) stably transduced with *HTT* exon 1 constructs containing 118 CAG repeats, *FAN1*$^{-/-}$ cells demonstrated a significant increase in somatic expansion over 42 days compared with a *FAN1*$^{+/+}$ line. Subsequent overexpression of wild-type FAN1 reduced somatic expansions to *FAN1*$^{+/+}$ levels (Goold et al. 2019). Furthermore, the same study

demonstrated that short-hairpin RNA (shRNA) knockdown of *FAN1* in HD-patient induced pluripotent stem cells (iPSCs), and iPSC-derived MSNs, increased somatic expansion over 70-80 days. Taken together, these data support a role for FAN1 in protecting against somatic instability. This has been corroborated in other studies where *FAN1*$^{-/-}$ HD patient iPSC-derived neural precursors (NPCs) and iPSC-derived MSNs demonstrate significantly increased expansion rates relative to *FAN1*$^{+/+}$ lines (Kim et al. 2020; McAllister et al. 2022). These data are supported by *in vivo* findings in a HD knock-in mouse model where *Fan1*$^{-/-}$ mice demonstrate increased somatic expansion in the striatum, relative to *Fan1*$^{+/+}$ mice (Loupe et al. 2020). Furthermore, biochemical-based assays have demonstrated that FAN1 is able to bind CAG/CTG and CGG/CCG loop out structures and act via endo- and exo-nucleolytic processing, as detected by gel-shift assays (Deshmukh et al. 2021b). This finding demonstrating that FAN1 can process such secondary structures associated with repetitive DNA regions indicates an active role for FAN1 in preventing expansions, through DNA-binding or nucleolytic processing.

Indeed, the role of FAN1 in preventing somatic instability seems to be multi-faceted and further investigation is required to elucidate these mechanisms. One way FAN1 is proposed to protect against somatic instability is through the modulation of the MMR machinery. In 2021, studies described the presence of two novel MLH1 interaction motifs, within FAN1 (Goold et al. 2021; Porro et al. 2021). Whilst a direct interaction between FAN1 and MLH1 had previously been reported (Cannavo et al. 2007), the nature of this interaction remained unclear. Although, in 2020 reports in a HD knock-in mouse model demonstrated that whilst *Fan1*$^{-/-}$ mice demonstrate increased somatic expansion in the striatum, loss of both *Fan1* and *Mlh1* completely ablates expansion (Loupe et al. 2020). This indicates that *MLH1* is epistatic to *FAN1*, indicating FAN1 may act to protect against MMR-mediated repeat expansion *in vivo*. Indeed, Goold and colleagues demonstrate that one interaction domain, termed the SPYF domain, is important in mediating FAN1 binding of MLH1. In this system, with U2OS cells transduced with an *HTT* exon 1 construct and a FAN1 SPYF mutant, somatic expansion rates failed to return

to *FAN1*$^{+/+}$ levels, indicating this domain is important in conferring protection against somatic expansion (Goold et al. 2021). These data together indicate that FAN1 binding to MLH1 may play a role in FAN1's ability to protect against somatic expansions. It is feasible that FAN1 acts to sequester MLH1 binding, preventing the downstream binding of the MutLα or MutLγ complexes (Gomes-Pereira et al. 2004; Pinto et al. 2013) at sights of CAG/CTG repeats, therefore protecting against expansions which can arise as a result of MMR.

Another way FAN1 may act to modulate somatic expansion is through DNA-binding to CAG/CTG repeats and processing of CAG/CTG loop-out structures. Whilst an initial report indicated the nuclease activity of FAN1 was not required for FAN1's protective function, as an overexpression of a nuclease-dead (D960A) FAN1 variant in U2OS cells was sufficient to rescue increase expansion rates (Goold et al. 2019), this has recently been challenged. Indeed, a second study by the same group concluded that this reported lack of requirement was likely due to an overexpression in the D960A variant construct. Where it is likely that although FAN1 nuclease activity was ablated, the significant FAN1 overexpression was likely sufficient to sequester MLH1, thus preventing expansions in this manner through modulation of the MMR machinery. Indeed a biochemical-based gel shift assay has demonstrated that the D960A variant is unable to cleave CAG/CTG or CGG/CGG loop outs, *in vitro*, suggesting a functional deficit (Deshmukh et al. 2021b). This is supported by findings in HD-patient iPSCs and NPCs where there is a dose-dependent effect of the D960A mutant, where *FAN1*$^{-/-}$ and *FAN1*$^{D960A/D960A}$ demonstrate comparable somatic expansion rates (McAllister et al. 2022). Furthermore, FAN1 has been shown to compete with MutSβ binding at CAG loop-out structures (Phadte et al. 2023). These data convincingly demonstrate that the MMR machinery and FAN1 have opposite effects, where MMR acts to promote somatic expansion whilst FAN1 acts to prevent this, by both competing with MMR for binding to repeat regions and through sequestration of MLH1 (Figure 1-3).

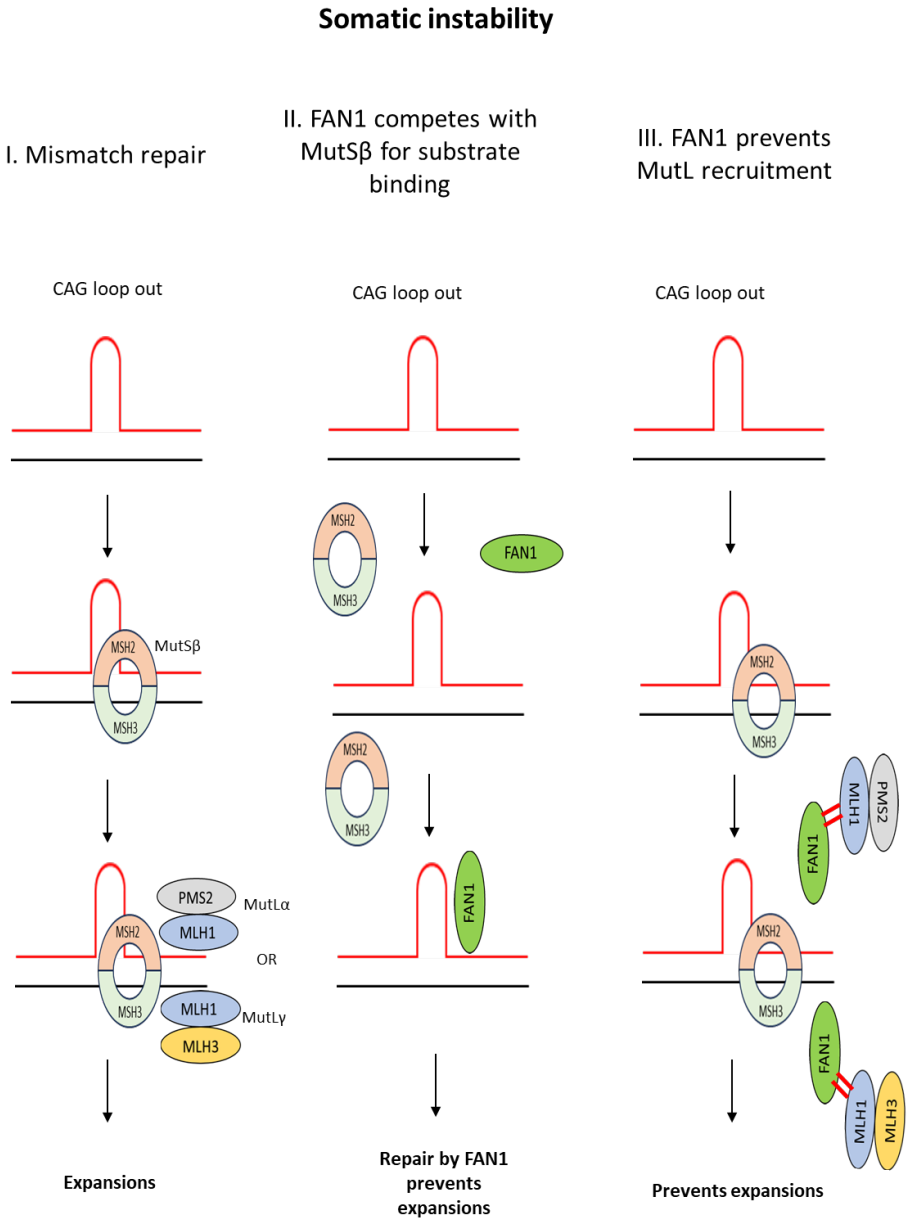


Figure 1-3: FAN1 protects against somatic instability. ***(I)*** *MMR acts to promote somatic expansion at sites of CAG/CTG loop outs, where MutSβ binds to looped-out structures and subsequent recruitment of MutL complexes results in error-prone repair and somatic expansion.* ***(II)*** *FAN1 and MutSβ compete for occupancy at sites of loop-outs. FAN1 binding prevents MutSβ binding and repair by FAN1 acts to prevent somatic expansions.* ***(III)*** *FAN1 may act to prevent somatic expansion through binding to MLH1, preventing processing of structures by MutL complexes.*

1.8 CRISPR-Cas9

The clustered, regularly interspaced, short palindromic repeats (CRISPR) system was originally identified as a primitive immune system of prokaryotes. This study described the presence of unusual sequences of a series of 29-repeats, separated by a 32 bp ‘spacer element’ (Ishino et al. 1987). The repetitive, palindromic nature of these regions led to the nomenclature ‘CRISPR’ (Jansen et al. 2002). This adaptive immune system in prokaryotes acts in response to viral or phage infection. After infection, with a previously unencountered pathogen, bacteria and archaea integrate foreign DNA into designated sites of the genome, flanked on either side by a CRISPR site (Mojica et al. 2005; Wiedenheft et al. 2012). This can be referred to as the ‘CRISPR locus’ and this process is repeated each time cells encounter a new pathogen. These CRISPR loci are transcribed and the primary transcript packaged into smaller CRISPR-derived RNAs (crRNAs), which are complementary to a specific pathogen. This cRNA, acts in complex with a Cas nuclease and trans-activating RNA (tracrRNA) to form an immune surveillance complex mediating the destruction and degradation of foreign nucleic acid (Deltcheva et al. 2011).

Many CRISPR- associated Cas proteins have been identified, including Cas9 (Brouns et al. 2008). Cas9 is a large protein comprised of two nuclease domains, a RuvC nuclease domain and HNH nuclease domain (Sapranauskas et al. 2011). In 2013 Cong and colleagues demonstrated that this Cas9 protein can be manipulated to induce precise genome editing in human and mouse cells. The classic Cas9 endonuclease can be targeted to a site of interest within the genome by a single-guide RNA (sgRNA), comprised of a tracrRNA and crRNA containing a target DNA sequence. Upon recruitment of Cas9 to the target site this induces a blunt-ended double-strand break (DSB), facilitating genome editing (Cong et al. 2013; Jiang et al. 2013; Ran et al. 2013). Since this discovery CRISPR-Cas9 has become a well-used and adapted gene editing tool, with techniques including; generating functional gene knock-outs, generating gene knock-ins via insertion of donor DNA and CRISPR-mediated gene regulation by CRISPR interference (Adli 2018; Akram et al. 2023).

1.9 Therapeutic approaches to CAG/CTG disorders

To date, there are no disease-modifying treatments for any of the 15 neurodegenerative or neuromuscular disorders caused by CAG/CTG expansions, only treatments aimed at managing symptoms. This section aims to highlight current therapeutic approaches, with a focus on HD.

1.9.1 Huntington's Disease

1.9.1.1 Antisense Oligonucleotides (ASOs)

The most well-developed potential therapeutic approach for HD, which has progressed to human clinical trials, relies on the use of gapmer ASOs to lower huntingtin protein levels. Gapmer ASOs contain modified nucleotides on two ends separated by a central DNA region specific to targeted RNA. These form a DNA/RNA heteroduplex that is recognised by RNase H causing cleavage and subsequent degradation of transcripts (Yasuhara et al. 2022). An ASO targeting *HTT* exon 36, Tominersen, results in RNAse H degradation of both wild-type and mutant *HTT* transcripts (Kordasiewicz et al. 2012). Despite promising results obtained in a phase 1-2a trial demonstrating a dose-dependent reduction in m*HTT* present in cerebrospinal fluid (CSF) (Tabrizi et al. 2019), trials were halted during the phase 3 trial. Preliminary data reported an increase adverse effects in dosing groups compared to placebo and a worsening clinical outcome score in patients in an 8-weekly dosing arm (Kingwell 2021; Tabrizi et al. 2022). A second gapmer ASO therapy, with selective m*HTT* allele targeting (Pfister et al. 2009), reached a phase 1/2 clinical trial. However, this trial was halted due to a lack of target engagement, with treatment failing to lower *mHTT* CSF levels (Kingwell 2021).

Despite these setbacks there are a series of ASOs in development targeting *mHTT*, which have demonstrated efficacy *in vitro* and *in vivo,* providing potential allele-specific therapy. One such ASO targeting SNPs associated with *mHTT* has collectively demonstrated specific mHTT lowering in HD fibroblasts and primary neurons, as well as a reduction in a transgenic rodent model and non-human primates (Skotte et al. 2014; Southwell et al. 2018).

Furthermore, the reduction of mHTT was associated with an improvement in cognitive and behavioural impairments in mice (Southwell et al. 2018). Another ASO, which targets the CAG repeat itself, has demonstrated mHTT lowering HD-patient derived fibroblasts and lymphoblasts (Evers et al. 2011). A follow-up study demonstrated sustained reduction *in vivo* which correlated with an improved performance in motor tasks (Datson et al. 2017).

These findings suggests there is still promise for the use of ASOs as a potential therapy, but there are potential drawbacks. Firstly, non-specific ASOs targeting *HTT* can lower both wild-type HTT and mHTT levels (Kordasiewicz et al. 2012). The long-term effect of lowering wild-type HTT is unknown but may have accounted for the adverse effects seen in the Tominersen trials (Tabrizi et al. 2022). Whilst targeting of *mHTT* using SNPs associated with *mHTT* alleviates this concern, this would not be applicable to all patients. For example, one SNP accounts for only 30% of European HD patients (Fang et al. 2023). Potentially, an ASO targeting the CAG site specifically, as described above, has the most promise, as it is selective for *mHTT* and has potential applications for other CAG repeat disorders. However, this is still limited by problems associated with ASO delivery. ASOs are unable to cross the blood-brain-barrier, therefore the mode of delivery is intrathecal, which is invasive, can be painful, and may not be feasible for all patients. In addition, intrathecal administration does not distribute ASOs evenly to all brain regions, with reduced delivery to subcortical structures, such as the striatum, which is primarily affected in HD (Lieberman et al. 2019). Furthermore, though mutant huntingtin concentrations may decrease upon administration of the drug, this is not a long-term solution as levels would revert over time. This approach, therefore, necessitates regular visits to hospital for intrathecal injections, every 3 to 4 months for the lifetime of the individuals.

1.9.1.2 RNAi based approaches

RNAi naturally occurs post-transcriptionally in cells through the use of endogenous microRNAs (miRNAs) that target and downregulate expression of transcripts (Carthew and Sontheimer 2009). This concept has been adapted for therapeutic use with short interfering RNAs (siRNAs) or miRNAs in development for HD, which tag target mRNA for degradation via an RNAi-silencing complex (Bobbin and Rossi 2016). Currently, one such miRNA therapy targeting *HTT* has demonstrated reduction of both wild-type and mutant HTT protein in HD-patient derived neurons and astrocytes (Keskin et al. 2019) and rodent models (Miniarikova et al. 2017; Thomson et al. 2023). Furthermore, intracranial injection of AAV5-miHTT in a HD minipig model demonstrated reduced *HTT* expression for up to one year post-treatment (Evers et al. 2018). These data, combined with proof of concept tolerability studies in a non-human primate model have led to support for the first AAV trial in humans for HD (Spronck et al. 2021). Whilst these data are suggest promise for RNAi-based therapeutics a major concern may be unwanted off-target affects and silencing of genes with partial sequence similarity (Gavrilov and Saltzman 2012; Redhwan et al. 2023).

1.9.1.3 DNA targeting approaches

Currently, all other therapeutic approaches described above indirectly target downstream processes, without correcting the cause of the disease itself; the expanded *HTT* CAG tract. Targeting the underlying cause of the disease should theoretically eliminate all disease pathogenesis (Tabrizi et al. 2022). The next few paragraphs details current gene editing tools available.

1.9.1.3.1 Zinc Finger Nucleases or Repressors

Zinc finger nucleases (ZFNs) are naturally occurring structural motifs, initially identified in the type IIS endonuclease, *Fok*I (Li et al. 1992). Importantly, these endonucleases have separate DNA-binding and DNA-cleavage domains. The DNA-binding domain is comprised of three zinc finger peptides which have

been demonstrated to be customisable, therefore allowing for the generation of synthetic zinc fingers and targeting to a specific region in the genome (Kim et al. 1996). However, the feasibility of ZFNs as a targeted gene editing approach for HD faces one major caveat, off-target cutting (Cathomen and Keith Joung 2008). Importantly, ZFNs generate double-stranded breaks (DSBs) at the target site. As these are highly mutagenic lesions, the risk of off-targets, coupled with a lack of predictability means this approach is not well-suited to gene-editing of post mitotic neurons (Wild and Tabrizi 2017; Phan et al. 2023).

Despite the major drawback of this approach, with respect to gene-editing, a therapeutic approach with particular promise for HD involves the use of a zinc finger transcriptional repressor. In this case, zinc finger peptides can be fused to a transcriptional repressor. This allows for specific DNA-binding and subsequent regulation of gene expression at the site of interest (Klug 2010). Two such studies have demonstrated the successful use of zinc finger repressors to lower *mHTT* expression (Garriga-Canut et al. 2012; Zeitler et al. 2019). Both groups reported selective lowering of *mHTT* in HD-patient derived cell lines with CAG repeats ranging from 38-67, with little to no effect on wild-type *HTT*. Further work *in vivo* demonstrated selective *mHTT* lowering in the striatum and an improvement of motor deficits (Garriga-Canut et al. 2012; Zeitler et al. 2019).

1.9.1.3.2 TALENs

Another example of engineered nucleases that can be adapted for gene-editing purposes are transcription activator-like effector nucleases (TALENs). These are similar to ZFNs, comprising a *Fok*I nuclease domain which is fused to an adaptable DNA-binding domain (Joung and Sander 2013). The DNA-binding domain contains highly conserved repeats which are derived from transcription-activator-like effectors (TALEs), originally identified in *Xanthomonas* bacteria (Boch and Bonas 2010). This approach has generated interest as studies have demonstrated TALENs can cleave DNA with a similar efficiency to ZFNs when targeting the same gene (Joung and Sander 2013).

Furthermore, compared with ZFNs, TALENs are easier to construct and have a higher editing success rate (Beumer et al. 2013).

One study has demonstrated that a TALEN construct targeting CAG trinucleotide repeats can induce targeted DSBs, leading to a high frequency of CAG contractions in yeast (Richard et al. 2014). Yeast, both heterozygous and homozygous for an expanded CAG tract, demonstrated near 100% contraction efficiency to 3-13 CAG repeats. To date, only one study, has examined the efficacy of TALENs in HD patient-derived fibroblasts (Fink et al. 2016). In this case TALENs were designed targeting SNPs associated with *mHTT*, allowing for allele specificity. TALENs were designed to induce deletions of CAG repeats above 15 units. Utilising this approach authors reported allele specific lowering of *mHTT*. Whilst this result is promising, as with ZFNs, TALENs induce DSBs at the target site, which may have unintended mutagenic consequences. However, like ZFNs, TALEs can be fused with transcriptional repressor domains, allowing for selective downregulation at a gene of interest (Becker and Boch 2021). Using this technology Fink and colleagues designed TALEs targeting SNPs associated with *mHTT* packaged into a vector backbone containing the KRAB transcriptional repressor. They observed *mHTT* specific lowering, compared with empty vector treatments. Whilst this provides a proof of principle for the use of TALEs, both with nuclease or repressor domains, as a potential therapy for HD further work to validate these findings *in vivo* has yet to be demonstrated.

1.9.1.3.3 CRISPR-Cas9 editing

Since its identification CRISPR-Cas9 editing has become an adaptable tool used by many in favour of ZFNs and TALENs. Current CRISPR-Cas9 editing approaches for HD have demonstrated the ability to inactivate mutant *HTT.* One study describes a dual sgRNA approach to selectively delete a ~44 Kb DNA region of mutant *HTT* containing the promoter, transcription start site and mutant CAG tract, in HD-patient derived iPSCs and NPCs (Shin et al. 2016). This was sufficient to prevent *mHTT* RNA and protein expression. Another

study published in 2017, similarly describes a dual sgRNA approach for targeting *mHTT,* demonstrating reduced *mHTT* RNA expression in a BacHD transgenic mouse model (Monteys et al. 2017). This approach differed from that reported by Shin and colleagues, instead of a large deletion, a smaller targeted deletion was achieved. This involved two sgRNA's which were located 5′ of *HTT* exon 1, in the *HTT* promoter, and within intron 1. In this way a smaller deletion occurs, encompassing the 5′ UTR and *HTT* exon 1. Importantly, both studies designed sgRNAs based on SNPs associated with *mHTT*, as such, this approach is potentially limited to a subset of patients (Shin et al. 2016; Monteys et al. 2017).

Additional non-allele specific Cas9 targeting approaches have been described in HD transgenic R6/2 (Ekman et al. 2019) and knock-in HD140Q mice (Yang et al. 2017). Both studies demonstrated a reduction in mHTT expression, a reduction in mHTT neuronal inclusions and recovery of certain motor deficits (Yang et al. 2017; Ekman et al. 2019). The work by Yang and colleagues has since been expanded, and excitingly, their recent study has demonstrated the efficacy of CRISPR-Cas9 as a potential therapy for HD in a knock-in pig model (Yan et al. 2023). This involves treatment with Cas9 and two sgRNAs flanking *HTT* exon 1. Cas9 and sgRNAs were packaged into two separate adeno-associated viruses (AAVs) and injected into the striatum of HD and control animals. Importantly, the AAV containing the sgRNAs also contains a human *HTT* donor template, containing a non-pathogenic 20 CAG repeat tract. This method then relies on homologue recombination and replacement of the pathogenic CAG repeat, with the non-pathogenic donor. This study has successfully demonstrated a reduction in mHTT levels, reduction in mHTT inclusions and an improvement in gait performance (Yan et al. 2023).

Whilst these studies demonstrate the potential for CRISPR-Cas9 editing as a therapeutic approach to HD, there are limitations. For example, two approaches rely on Cas9-targeting of SNPs associated with *mHTT*, and are therefore limited to a reduced percentage of HD patients (Shin et al. 2016; Monteys et al. 2017). The remainder describe the use of sgRNAs that target both *HTT* and *mHTT*, and are therefore not allele-selective (Yang et al. 2017; Ekman et al. 2019; Yan et al. 2023). Furthermore, all describe the use of a

Cas9 endonuclease, which induces DSBs. These are mutagenic DNA lesions which present an increased likelihood of off-target mutations (Cho et al. 2014).

1.10 Methods to monitor CAG/CTG repeats

Repetitive DNA elements have been notoriously difficult to study, given their propensity to form unusual DNA structures which present a particular challenges for DNA sequencing and characterization of DNA architecture (Massey et al. 2019). To date, a number of methods have been developed to assay for changes in expanded CAG/CTG regions. This section will describe those used in this thesis (Figure 1-4).

1.10.1 Small-pool polymerase chain reaction (spPCR)

Small-pool polymerase chain reaction (spPCR), once considered the gold-standard technique for monitoring changes in TNR regions, involves the sequential dilution of input DNA to a few genomic equivalents. These diluted DNA samples undergo PCR amplification and are resolved via gel electrophoresis. Following this, a southern-blot is performed with subsequent treatment with a radiolabelled probe, targeting the CAG/CTG repeat (Monckton et al. 1995; Gomes-Pereira, Mario Bidichandani and Monckton 2004). This method was adapted from previously used southern blotting techniques, which involved using large µg concentration of starting DNA which were subsequently digested and probed. The benefit of spPCR over this method is dilution of DNA prior to PCR allows for a broader range of allelic repeat lengths to be determined within a sample, detecting both the common and rarer alleles which are missed by a bulk approach. Furthermore, this technique by-passes the inherent amplification advantage of smaller alleles (Monckton et al. 1995).

1.10.2 High-throughput sequencing

The advent of high-throughput sequencing has allowed for increased accuracy of CAG/CTG repeat sizing. Many studies have described the targeted sequencing of expanded repeats with; Illumina MiSeq (Ciosi et al. 2019; Ciosi

et al. 2021; McAllister et al. 2022), Oxford Nanopore technology (Ebbert et al. 2018; Giesselmann et al. 2019; Rasmussen et al. 2022; Stevanovski et al. 2022) and Pacific Biosciences (PacBio) Real Time (SMRT) sequencing (Loomis et al. 2013; Cumming et al. 2018; Höijer et al. 2018; Ciosi et al. 2021; Mangin et al. 2021; Tsai et al. 2022). Specifically, Oxford Nanopore and PacBio sequencing technology is best suited for sequencing of expanded CAG/CTG repeats beyond ~80, due to MiSeq's restricted read-length (Ciosi et al. 2021). Whilst both techniques are able to sequence repeats beyond this, Oxford Nanopore technology has a higher error rate (Taylor et al. 2022). Therefore the long-read sequencing approach undertaken in this thesis was PacBio SMRT sequencing.

PacBio SMRT sequencing is a real-time sequencing approach which does not require a pause step between reads (Eid et al. 2009). After PCR amplification of the targeted region, a CAG/CTG repeat tract, for example, SMRTbell libraries are prepared by the purification of PCR amplicons and subsequent annealing of barcoded adapters to form a closed, circular SMRTbell template. This template is then loaded onto a SMRT cell chip where the SMRTbell template diffuse into a sequencing unit, the zero-magnetic wave guide (ZMV). In each individual ZMV a single polymerase is held which incorporates bases into the read strand. As this SMRTbell template is a closed circle, after base incorporation of one-strand the polymerase is then able to act on the other strand. The polymerase can then pass back through both strands multiple times, collecting multiple reads. At the end of the sequencing protocol these reads are assembled to form a circular consensus sequence (CCS) (Rhoads and Au 2015). In this thesis after the generation of a CCS all reads were then passed through the Repeat Detector algorithm (Taylor et al. 2022). This is an algorithm that is able to count tandem repeats in targeted sequencing data. It utilises sequencing files to determine a weighted profile to calculate repeat lengths in each sample, and generates histogram files as an output detailing the number of reads per repeat length. It has previously been demonstrated to accurately determine repeat sizes from SMRTbell sequencing data (Taylor et al. 2022).

1.10.3 GFP-based cell reporter cell line

In 2014 Santillan and colleagues described a green fluorescent protein (GFP) based reporter cell line, that was able to monitor CAG/CTG contractions *in vitro*. In this system a GFP mini-gene, separated by an intronic region containing 89 CAG repeats, was inserted into a Flp-In HEK293 T-Rex cell line. These cells contain a flippase recognition target (FRT) site at a transcriptionally active locus. This allows for the homogenous expression of a protein of interest, in this case GFP. In this system GFP expression serves as a proxy for CAG repeat length. The presence of expanded CAG repeats in the intronic region interferes with functional mRNA splicing, in a length dependent manner (Gorbunova et al. 2003). Therefore, longer CAG repeats demonstrate increased mRNA mis-splicing, and a dimmer GFP intensity. Comparatively, short repeats demonstrate reduced levels of mRNA mis-splicing, with higher GFP expression (Santillan et al. 2014; Cinesi et al. 2016). This cell system provides a rapid and sensitive tool for monitoring changes in CAG/CTG repeat tracts via flow cytometry. In 2016, data from our laboratory replicated these findings and confirmed that this cell system could not only monitor changes in CAG contractions, but also CAG expansions (Cinesi et al. 2016).

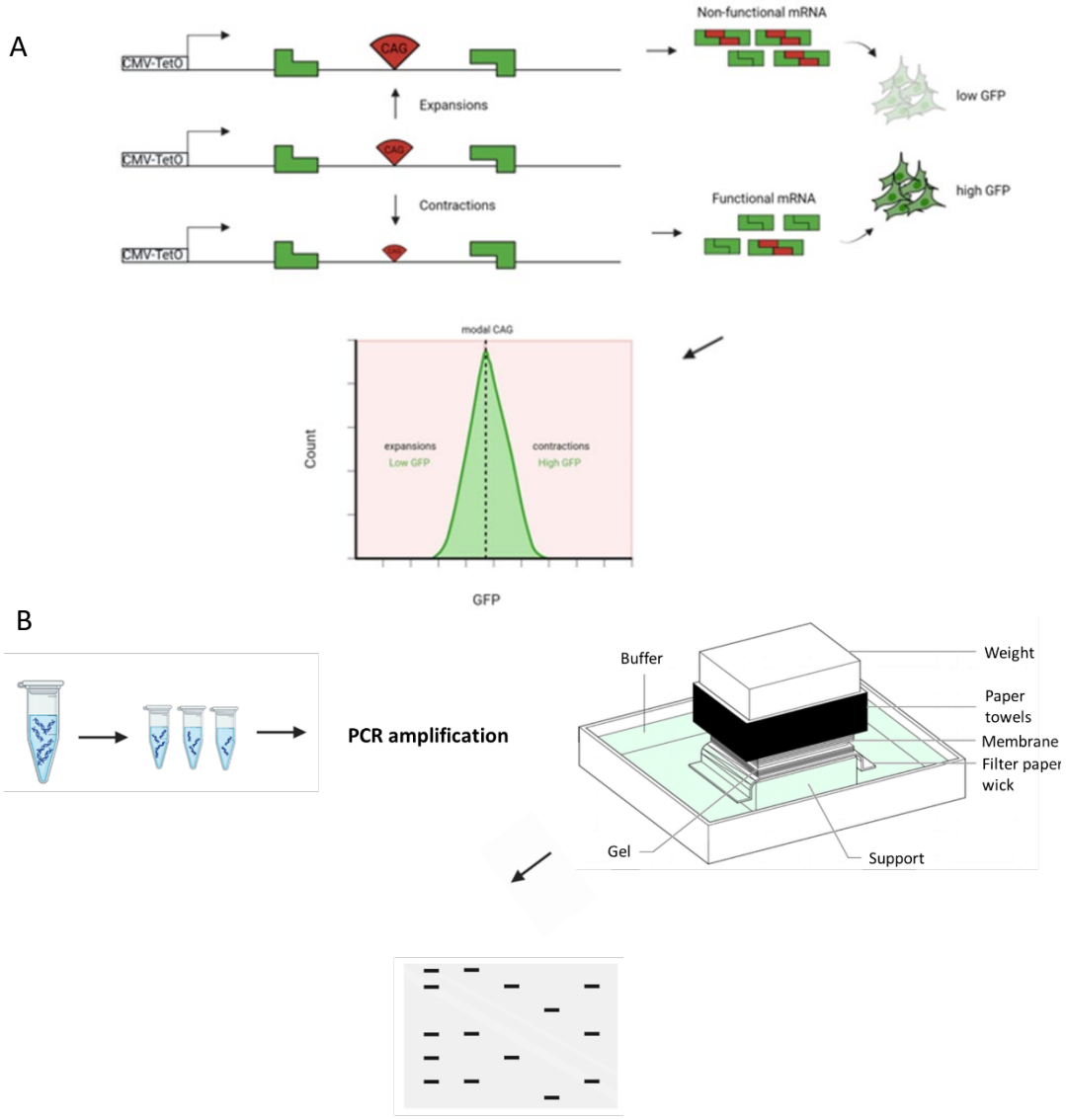

A
CMV-TetO
CAG
Expansions
Contractions
Non-functional mRNA
low GFP
Functional mRNA
high GFP
modal CAG
Count
expansions
Low GFP
contractions
High GFP
GFP
B
PCR amplification
Buffer
Weight
Paper towels
Membrane
Filter paper wick
Gel
Support
C
PCR amplification
SMRTbell template
1 Input
2 Repeat Detector
3 Output
Sequence
Circular Profile
Restrictive Profile
Permissive Profile
6 CAG
9 CAG

Figure 1-4: Comparison of methods used in this thesis to monitor changes in CAG/CTG repeats. ***(A)*** *Representation of the ectopic GFP-reporter based assay. A CAG repeat tract resides within an intron separating two exons of a GFP-mini gene. This interferes with mRNA splicing in a length-dependent manner, allowing GFP intensity to serve as a proxy for changes in CAG length.* ***(B)*** *DNA is collected for spPCR and diluted to a few genomic equivalents prior to PCR amplification. After this PCR products are run via gel electrophoresis and DNA transferred to a nylon membrane via southern blotting. The membrane is then probed with a radiolabelled probe to visualise the spectrum of alleles.* ***(C)*** *DNA samples are collected for SMRTbell sequencing and PCR amplification of the target region performed. SMRTbell libraries are then prepped via the addition of adapters to either end the double-stranded PCR molecule, which creates a closed circular template. Upon loading into the sequencing chip a DNA polymerase reads through the template generating a circular consensus sequence. Repeat Detector then acts to count reads and generate histogram profiles of repeat distributions.*

1.11 Contracting CAG/CTG repeats using CRISPR-Cas9 nickase

As previously discussed, directly targeting the genetic cause of HD, the expanded *HTT* CAG repeat, offers an exciting potential therapy. In theory, this genetic correction of an expanded repeat to a non-pathogenic length would ameliorate all disease pathogenesis, and could even be preventative, dependent upon treatment administration (Tabrizi et al. 2022). Our laboratory has recently designed the first editing approach that can successfully induce CAG/CTG contractions in expanded repeats, whilst avoiding concomitant expansions in human cells. This involves the use of a *Streptococcus pyogenes* Cas9 nickase with a sgRNA targeting an expanded CAG/CTG tract (Cinesi et al. 2016). Importantly, in addition to a sgRNA directing the Cas9 to a target site in the genome, the Cas9 requires a specific sequence to be present on the target strand, immediately adjacent to the target sequence. This is referred to as the protospacer adjacent motif (PAM) site. The PAM for the *S. pyogenes* Cas9 is 5′-NGG-3′, but it is also capable of binding 5′-NAG-3′, making it suitable for targeting CAG/CTG repeat tracts (Zhang et al. 2014). The Cas9 nickase differs from the traditional Cas9 endonuclease as it induces single-stranded breaks (SSBs) instead of DSBs. The formation of SSBs are made possible by engineering a D10A point mutation in the RuvC nuclease domain,

rendering it inactive (Cong et al. 2013). This leaves only the HNH nuclease domain active, which will create SSBs on the non-target DNA strand (Jinek et al. 2012).

Interestingly the ability of the nickase to induce contractions occurs in a length dependent manner, whereby repeat sizes within the non-pathogenic range remain unedited (Cinesi et al. 2016). This presents a potential allele-selective targeting approach applicable to not only HD, but other CAG/CTG diseases, as the expanded repeat is specifically targeted by the sgCTG. Currently, the DNA repair mechanism which induces Cas9 nickase-induced CAG/CTG contractions is not known. Previous data indicates that nickase-induced contractions are not likely to occur via single-strand break repair (SSBR). Detection of DNA nicks, or SSBs, relies on both X-Ray Repair Cross-Complementing Protein 1 (XRCC1) and Poly (ADP-ribose) polymerase (PARP) which together act to sense DNA nicks and mediate repair to downstream SSBR pathways (Caldecott 2003). In the GFP-based reporter cell line, described above, knockdown of XRCC1 and inhibition of PARP did not affect contraction rates, as detected by GFP expression via flow cytometry (Cinesi et al. 2016). Furthermore, knockdown of MMR factor, MSH2, the key component of the MutS complexes did not affect nickase-induced contraction rates, indicating that MMR may not be involved in generating nickase-induced contractions. Comparatively, ataxia telangiectasia mutated (ATM), a DNA damage repair kinase, seems to promote nickase-induced contractions, as cells treated with an ATM inhibitor demonstrated a significant reduction in contractions. Data from these findings propose a model for Cas9 nickase-induced contractions where Cas9 induces multiple nicks across the expanded CAG/CTG repeat. These nicks lead to the formation of a DNA gap structure which are converted to contractions in an ATM-dependent mechanism (Cinesi et al. 2016). However, further work is needed to fully understand how the Cas9 nickase acts to generate contractions in this system.

1.12 Thesis aims

Both genetic data and biochemical, cellular and *in vivo* modelling have indicated that the DNA repair protein, FAN1, protects against somatic expansion, not only at CAG/CTG repeats, but also at CGG/CCG repeats. Furthermore, data from our lab indicates that DNA repair machinery is involved in mediating Cas9-nickase induced contractions of CAG/CTG repeats in this potential CRISPR-Cas9 gene editing strategy. However, the specific DNA repair pathway and proteins that mitigate these contractions remains to be fully elucidated. Further understanding of what proteins modulate nickase-induced contractions would be useful for; understanding the mechanism of contractions, modulating contraction efficiency and potential screening to ensure patients with rare variants, for example in FAN1, may still benefit from this approach. Therefore, this thesis set out to explore three primary objectives:

1) Whether loss of FAN1 in GFP HEK293 reporter cell lines impacts Cas9 nickase-induced contraction rates.
2) Whether a loss of FAN1 impacts Cas9-nickase induced contractions in HD-patient iPSC and iPSC-derived neurons.
3) What functional domains of FAN1 may be involved in modulating contraction rates.

2 Materials and Methods

2.1 Cell culture

2.1.1 Cell culture reagents

Cell Culture Reagent	Supplier	Catalogue Number
Gibco DMEM (1X) + GlutaMAX	Thermo Fisher Scientific	31966-021
Dialysed Foetal Bovine Serum	Merck	F0392
Penicillin-Streptomycin (10000U/ml)	Thermo Fisher Scientific	11548876
Hygromycin B	Thermo Fisher Scientific	10697010
Blasticidin S HCl	Fisher Scientific Ltd	12172530
G418 Disulfate	VWR International	A6798.0050
Puromycin	Fisher Scientific	10296974
Phosphate Buffered Saline	Thermo Fisher Scientific	10010056
Trypsin-EDTA	Thermo Fisher Scientific	25300054
Doxycycline hyclate	Merck	D9891
Lipofectamine 2000	Fisher Scientific Ltd	11668500
Opti-MEM Reduced Serum Medium	Fisher Scientific Ltd	31985047
Polybrene	VectorBuilder	N/A
Dimethyl Sulfoxide (DMSO)	Merck	D2650
NC-Side 8 NucleoCounter® slides	Chemometec	942-0003
Solution 18 AO-DAPI	Chemometec	910-3018

Table 2-1: Table of HEK293 cell culture reagents.

Cell Culture Reagent	Supplier	Catalogue Number
Essential 8 Flex Medium Kit	Life Technologies	A2858501
DMEM/F-12 GlutaMAX	Thermo Fisher Scientific	31331028
Advanced DMEM/F-12	Thermo Fisher Scientific	12634028
Neurobasal™-A Medium	Thermo Fisher Scientific	10888022
Matrigel hESC Matrix	Scientific Laboratory Supplies	354277
Poly-D-Lysine	Thermo Fisher Scientific	A3890401
Penicillin-Streptomycin (10000U/ml)	Thermo Fisher Scientific	11548876
Y-27632 dihydrochloride (ROCK inhibitor)	Bio-Techne	1254
Phosphate Buffered Saline	Thermo Fisher Scientific	10010056

CryoStor CS10	Stem Cell Technologies	07930
ReLeSR	Stem Cell Technologies	5872
Accutase	Fisher Scientific Ltd	12780000
MACs Neurobrew-21	Miltenyi Biotec	130-093-566
MACs Neurobrew-21 w/o Vitamin A	Miltenyi Biotec	130-097-263
StemMACs™ SB431542	Miltenyi Biotec	130-105-336
StemMACs™ LDN-193189	Miltenyi Biotec	130-106-540
StemMACs™ IWR-1-endo	Miltenyi Biotec	130-110-491
PD0332991	Merck	PZ0199
DAPT	Bio-Techne	2634
Human BDNF research grade	Miltenyi Biotec	130-093-811
Forskolin	Bio-Techne	1099
CHIR99021	Bio-Teche	4423
Ascorbic Acid	Sigma-Aldrich	A4544
Calcium Chloride	Sigma-Aldrich	499609
Human FGF-2	Miltenyi Biotec	130-104-921
Plurisin-1	Stem Cell Technologies	72822
GlutaMAX	Thermo Fisher Scientific	35050038
Recombinant human NT-3	Peprotech EC Ltd	450-03
Blasticidin S HCl	Fisher Scientific Ltd	12172530
Doxycycline hyclate	Merck	D9891
Human EGF premium grade	Miltenyi Biotec	130-097-749
Human LIF research grade	Miltenyi Biotec	130-108-159
NC-Slide 8 NucleoCounter® slides	Chemometec	942-0003
Solution 18 AO-DAPI	Chemometec	910-3018

Table 2-2: Table of HD-iPSC cell culture reagents.

2.1.2 HEK293 cell maintenance

The HEK (human embryonic kidney) cell lines utilised in this thesis are Flp-In-T-Rex 293 cells (ThermoFisher Scientific). All HEK293 cell lines were maintained at 37°C with 5% CO_2 in Dulbecco's essential media (DMEM) with GlutaMAX and 10% dialysed fetal bovine serum (FBS). Cells were routinely treated with selection media supplemented with a combination of hygromycin (150 µg/mL), blasticidin (15 µg/mL), G418 (200 µg/mL) and puromycin (1 µg/mL), dependent on individual lines (see 2.1.2.1). Cells were passaged every 3-4 days when a confluency between 80-90% was reached. For

passaging, cells were washed briefly with phosphate-buffered saline (PBS) and detached using Trypsin/EDTA. Cells were then seeded on a new plate with a density of 1:5. Cells were used for experiments up to passage 40.

For cryopreservation, cells were washed with PBS, detached from the flask/plate using Trypsin/EDTA and subsequently centrifuged at 300 rcf for 5 min before being resuspended in Freezing media (50% DMEM with GlutaMAX, 30% dialysed FBS and 20% dimethyl sulfoxide (DMSO)) in 1 mL cryovials, with approximately $1x10^6$ cells/mL. Cryovials containing cells were transferred to -80°C in Cool Cells and frozen at a rate of -1°C/min and after at least 24 hr cells were transferred to liquid nitrogen.

For thawing, cryovials were removed from liquid nitrogen and warmed at 37°C in a water bath for 2 min until partially thawed. Cells were then resuspended dropwise in non-selective media. Selective media was subsequently added once cells had reached confluence.

2.1.2.1 Cell line specifics

All HEK293 cell lines had been previously engineered to stably express a GFP-Pem1 insert and doxycycline inducible promoter (CMV/$TetO_2$) (Santillan et al. 2014). Two cell lines were routinely used in Chapter 3 and 4, referred to as GFPSNick120 and GFP(CAG101), respectively.

GFPNickS120 cells contain the stably integrated GFP-Pem1 insert and doxycycline inducible promoter. Additionally, they contain a stably integrated *Streptococcus pyogenes* (*S.pyogenes*) Cas9 D10A nickase and sgCTG targeting CAG/CTG repeat tract containing a target sequence of CTGCTGCTGCTGCTGCTG. These cells were maintained in selection media containing hygromycin (150 µg/mL), blasticidin (15 µg/mL), G418 (200 µg/mL) and puromycin (1 µg/mL), ensuring continued expression of all stable sites. These lines were used to monitor for changes in Cas9 nickase-induced contractions.

GFP(CAG101) cells also contain the stably integrated GFP-Pem1 insert and doxycycline inducible promoter, but do not contain the Cas9 D10A nickase of sgCTG. These cells were maintained in media containing hygromycin (150

µg/mL) and blasticidin (15 µg/mL) and were used to monitor CAG/CTG somatic instability.

2.1.3 iPSC maintenance

Human HD-patient derived induced pluripotent stem cells (iPSCs) used in this thesis were generated by the HD-iPSC consortium from a juvenile HD-patient fibroblasts with an initial expanded CAG *HTT* allele of 109 repeats (Mattis et al. 2012). The two clonal lines, N1 and N5, used in this thesis, including *FAN1*$^{+/+}$ and *FAN1*$^{-/-}$ lines had been previously validated, as described in McAllister et al. 2022.

iPSCs were cultured on Matrigel-coated plates (50 µg/mL) in Essential 8 Flex media at 37°C and 5% CO_2 conditions. A complete media change was performed every 2 days. When cells reached ~70% confluence iPSCs were passaged using ReLeSR. Briefly, old media was removed and cells were washed with PBS and incubated with ReLeSR for 5 min at 37°C. ReLeSR was removed and cells resuspended in fresh E8 media and replated onto Matrigel-coated wells.

For cryopreservation, cells were detached from the plate using ReLeSR, as described above. Cells were centrifuged at 300 rcf for 3 min and resuspended in CryoStor CS10, with approximately 1x10^6 cells/mL. Cryovials containing cells were transferred to -80°C in Cool Cells where cells were frozen at a rate of -1°C/min and after at least 24 hr cells were transferred to liquid nitrogen.

For thawing, cryovials were removed from liquid nitrogen and warmed at in a bead bath at 37°C for 2 min, until partially thawed. Cells were then resuspended drop wise into Matrigel-coated (50 µg/mL) plates in E8 media containing 10 µM ROCK inhibitor. After 24 hr, a full media change was made to remove ROCK inhibitor.

Monolayer differentiation of iPSCs to cortical neurons

The differentiation protocol used in this thesis was adapted from a previously published protocol, with minor modifications (Telezhkin et al. 2016). All differentiations were carried out in a 6-well plate. At day 1 iPSCs (N1 subclone) at 60-70% confluency were pre-treated for 1 hr with fresh E8 media containing 10 μM ROCK inhibitor. Cells were then dissociated into a single-cell suspension using Accutase for 5 min at 37°C. Following this, Accutase was removed and cells were centrifuged for 3 min at 300 rcf and resuspended in E8 media with 10 μM ROCK inhibitor and left to adhere to on Matrigel-coated plates overnight, to reach confluency. The following day media was removed and cells were washed 3 times with PBS. Cells were then maintained in SLI media for 8-12 days, with medium changes every day. On day 8-12, cells were treated with 10 μM ROCK inhibitor for 1 hr. Cells were then dissociated using Accutase for 5 min at 37°C to generate a single-cell suspension. Cells were spun down at 300 rcf for 5 min and split in a 1:4 ratio and re-plated in NB media containing 10 μM ROCK inhibitor. The next day, ROCK inhibitor was removed and media changed every day until day 16. Between days 14-16 plurisin was added to remove any remaining iPSCs from the neural precursor (NPC) culture. From day 16 onwards NPCs could be frozen, as described above, or expanded for up to 7-days to increase NPC yield.

For terminal differentiation, NPCs were plated on 6-well plates that had been previously coated with Poly-L-lysine (12 hr) and Matrigel-coated (100 μg/mL) (1 hr) after dissociation with Accutase, as described above. NPCs were pre-treated with 10 μM ROCK inhibitor prior to dissociation. Cells were re-plated in SJA media containing 10 μM ROCK inhibitor, and left to adhere overnight. The following day, media was changed to SJA media with half-media changes every 2 days, for 7 days. After 7 days, medium was changed to SJB medium. For the following 14 days half media changes of SJB were performed every 2 days, after which media changes were reduced to twice weekly.

Neuronal differentiation media	Composition
SLI	Advanced DMEM-F12 with 1% Glutamax, 1% Penicillin/Streptomycin, 1% MACs Neurobrew (w/o) Vitamin A, 10 µm SB431542, 200 nM LDN-193189 and 1.5 µM IWR-1-endo
NB	Advanced DMEM-F12 with 1% Glutamax, 1% Penicillin/Streptomycin and 2% MACs Neurobrew (w/o) Vitamin A
NF	Advanced DMEM-F12 with 1% Glutamax, 1% Penicillin/Streptomycin, 2% MACs Neurobrew (w/o) Vitamin A and FGF-2
SJA	Advanced DMEM-F12 with 1% Glutamax, 1% Penicillin/Streptomycin, 2% MACs Neurobrew with Vitamin A, 2 µM PD0332991, 10µ M DAPT, 10 ng/mL BDNF, 10 µM Forskolin, 3 µM CHIR99021, 200 µM Ascorbic Acid, 0.8 mM $CaCl_2$, 300 µM GABA
SJB	Advanced DMEM-F12 with 1% Glutamax, 1% Penicillin/Streptomycin, 2% MACs Neurobrew with Vitamin A, 2 µM PD0332991, 200 µM Ascorbic Acid, 0.4 mM $CaCl_2$

Table 2-3: Media composition for iPSC monolayer differentiation to cortical neurons.

2.1.4 Directed differentiation of NPCs to astrocyte precursors

NPCs were differentiated to astrocyte precursors (APCs) using a protocol adapted from Serio et al. 2013. iPSCs were differentiated to NPCs, as described above. NPCs were proliferated until they reached confluence. Cells were then expanded for 3-6 passages in NEL media in T25 flasks, with cells

split 1:3 with Accutase, as described previously, when confluent. After expansion, cells were sorted for APC marker CD44 using FACS. Cells were dissociated using Accutase for 10 min at 37°C and centrifuged at 300 rcf for 3 min. $1x10^6$ were counted and resuspended in 98 µL of NEF media. 2 µL of the CD44 antibody or the isotype control were then added and cells were incubated on ice in the dark for 10 min. Cells were then washed once in NEF media and resuspended in 1 mL of PBS with 0.1% bovine serum albumin (BSA). The top 50% of CD44 positive cells were then collected and replated onto Matrigel-coated (100 µg/mL) plates. Cells were then proliferated in NEF media and upon reaching confluency frozen down as described in section 2.1.3.

Astrocyte differentiation media	Composition
NEL	Advanced DMEM-F12 with 1% Glutamax, 1% Penicillin/Streptomycin, 2% Neurobrew with Vitamin A, 20 ng/mL EGF, 20ng /mL LIF
NEF	Advanced DMEM-F12 with 1% Glutamax, 1% Penicillin/Streptomycin, 2% Neurobrew with Vitamin A, 20 ng/mL EGF, 20ng/mL FGF-2

Table 2-4: Media composition for differentiation of NPCs to APCs.

2.1.5 Generation of iNeuron iPSCs

'Inducible Neuron' (iNeuron) iPSCs using the N5 HD109 subclone were generated by Joseph Stone (Massey Lab, Cardiff University). Lines were generated following guidelines in Fernandopulle et al. 2018. iPSCs were cultured, as previously described in section 2.1.3, and upon reaching 70-80% confluence were dissociated using Accutase at 37°C for 5 min. After which cells were centrifuged at 300 rcf for 3 mins and $2x10^5$ cell were re-plated in a

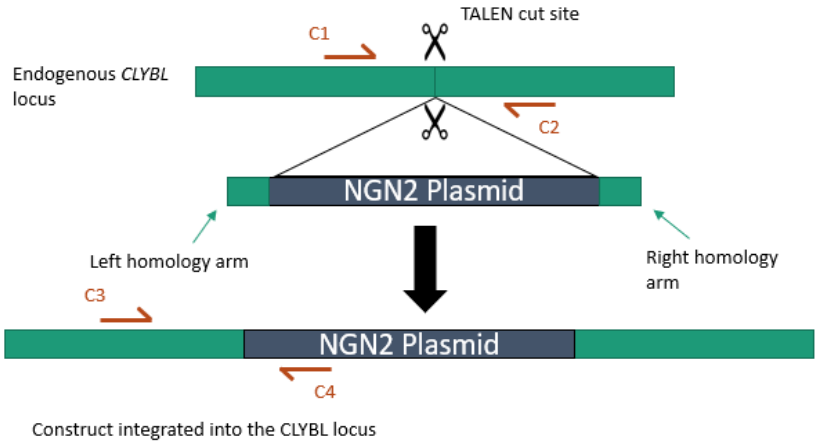
12-well tissue culture dish, pre-treated with Matrigel (50 µg/mL) in 1 mL of E8 media supplemented with 10 µM ROCK inhibitor. Cells were left in the incubator overnight and the next day transfected with plasmids containing; the left and right TALEN arms (Addgene:#62196 and #62197) and the human Neurogenin-2 (NGN2) donor (Addgene:#124229). Transfection was achieved by mixing two reaction tubes containing either Opti-MEM with all three plasmids or Opti-MEM and 5 µL of Lipofectamine Stem reagent (Thermo Fisher Scientific), as described below:

Component	Volume (µL)
OptiMEM	50
Lipofectamine	5

Component	Volume (µL)
OptiMEM	50
Addgene #62196 (0.375 µg)	variable
Addgene #62197 (0.375 µg)	variable
Addgene #124229 (0.75 µg)	variable

The contents of each tube were left to incubate separately at room temperature for 5 min, before mixing and then the total volume was incubated for a further 10 min. This 100 µL transfection solution was added dropwise to cells. Cells were then returned to 37°C and left for 24 hr, after which a full media change with E8 media was carried out. Once the transfected cells had reached confluency cells were split using ReLeSR, as previously described, into 2-wells of a 6-well plate. After 24 hr cells were treated with 5 µg/mL of Blasticidin, to select for wells containing the transgene. Cells which survived treatment were then plated as single cells and clonal expansion undertaken to screen for transgene expression.

A



B

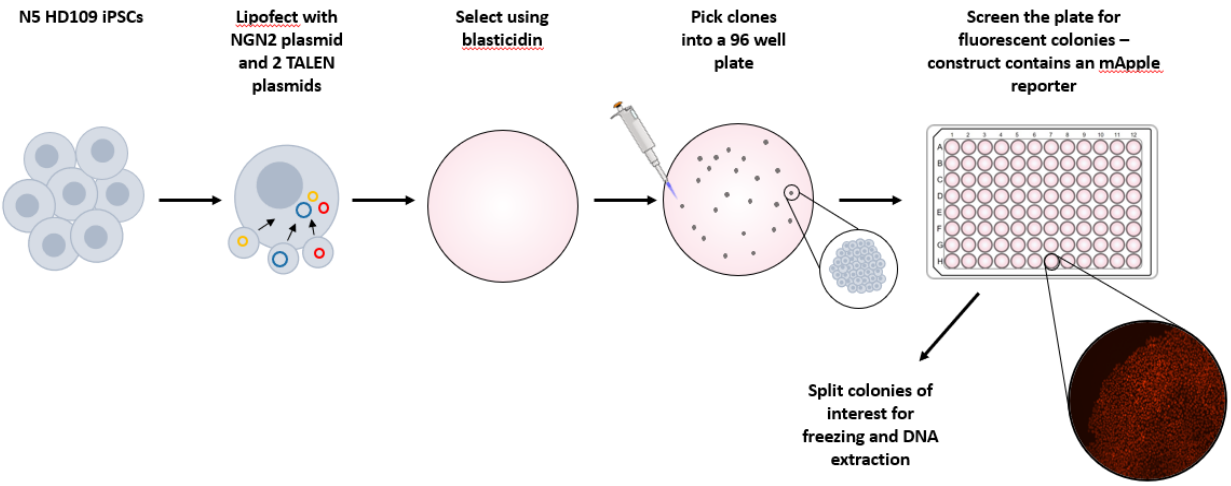


Figure 2-1: Depiction of TALEN-mediated generation of inducible Neuron iPSC lines. ***(A)*** *Graphic depicting TALEN cut site within the endogenous CLYBL locus. The left and right-homology arms of the two TALEN's direct the NGN2 plasmid to the correct integration site.* ***(B)*** *Schematic depicting the approach used to generate HD-iPSCs expression the NGN2 transgene. Figure by Joseph Stone.*

2.1.5.1 Doxycycline induced differentiation to cortical neurons

For terminal differentiation of iPSCs expressing the NGN2 transgene to cortical iNeurons, on day 0 iPSCs at 60-70% confluency were pre-treated with 10 µM ROCK inhibitor for 1 hr, during which time 6-well plates were coated with Matrigel (50 µg/mL). After 1 hr cells were dissociated into a single cell suspension with Accutase for 5 min at 37°C. Cells were then centrifuged at 300 rcf for 3 min and $1x10^6$ cells re-plated in a 6-well in induction media (IM) with 10 µM ROCK inhibitor. After 24 hr, a full media change was undertaken to remove ROCK inhibitor. Media was changed every day with fresh Induction media, until day 3. On day 3 cells were dissociated using Accutase, as previously described in section 2.1.4, and resuspended in culture media (CM). Cells were counted and $1x10^6$ re-plated for final terminal differentiation. Cells were plated on pre-coated 6-well plates that had been coated for 12 hr with Poly-L-lysine and for 1 hr subsequently with Matrigel (100 µg/mL). 24 hr after plating a full media change was performed. Subsequent media half media changes were undertaken twice weekly.

iNeuron differentiation media	Composition
IM	Advanced DMEM-F12 with 1% Glutamax, 1% Penicillin/Streptomycin, 1% Neurobrew (w/o) Vitamin A, 2 µg/mL doxycycline
CM	Advanced DMEM-F12 with 1% Glutamax, 1% Penicillin/Streptomycin, 1% Neurobrew with Vitamin A, 10 ng/mL BDNF, 10 ng/mL NT-3, 200 µM Ascorbic acid

Table 2-5: Media composition for doxycycline-induced differentiation of iNeurons.

2.1.6 Cell counting

Cells were counted using the chemometic Nucleocounter® NC-250™. Cells were washed with PBS dissociated into a single cell suspension using Trypsin/EDTA (HEK293) or Accutase (iPSCs and NPCs). A 20µl sample was taken and 1µL AO-DAPI (acridine orange and 4′, 6-diamidino-2-phenylindole) stain added. Samples were added to a chamber within NC-Slide A8 and cell count and viability recorded.

2.2 CRISPR-Cas9 targeting to generate knock-out cell lines

2.2.1 Generating *FAN1* knock-outs in GFPNickS120 cells

2.2.1.1 sgRNA design

For the generation of *FAN1* knock-outs in the GFPNickS120, cells were transfected with plasmids containing a single guide RNA (sgRNA) scaffold in a pUC57 backbone. As these cells already contain a stably integrated CRISPR Cas9 D10A nickase and sgCTG we could not deliver an exogenous Cas9 endonuclease RNP, as described in 2.3.2, as there was a risk of the Cas9 endonuclease binding to the stably expressed sgCTG in these cells, causing unwanted double-strand breaks at CAG/CTG repeats. The delivery of plasmids containing paired sgRNAs with *FAN1*-targeting sequence was sufficient to induced *FAN1* knock-out clones and relies on the generation of paired DNA nicks by the stably expressed CRISPR Cas9 D10A nickase. sgRNAs were designed targeting both exon 2 and exon 4 of *FAN1*.

Pair	sgRNA	Location	Sequence (5′ - 3′)
1	bVIN 567	Exon 2	CTACCTTCGGAGTTTCCTTG
	bVIN 568	Exon 2	GGATGACCGGTTGTTTGACC
2	bVIN 569	Exon 2	ACTTCGTTCAAGTGGATCCA
	bVIN 570	Exon 2	CACACATTTCATCAAGGTGC
3	bVIN 571	Exon 4	TTCCACTTGGTGAATCCCAA
	bVIN 572	Exon 4	CAAGTGGAAGGTCTTGGCTA
4	bVIN 573	Exon 4	GGTGGACGCCTTTCTCAAAT
	bVIN 574	Exon 4	GTCCATTGGGATTCACCAAG

2.2.1.2 Lipofectamine- mediated transfection of GFPSNickS120 cells

The day before transfection of cells with paired sgRNAs, $4x10^5$ cells were seeded in a 12-well tissue culture plate in DMEM non-selective media. The following day cells were transfected with paired sgRNAs using Lipofectamine 2000 (Thermo Fisher Scientific). Two tubes containing either; Opti-MEM media and lipofectamine or Opti-MEM media and the desired sgRNA pairs with a blue fluorescent protein (BFP) expressing plasmid, were prepared. The contents of each tube was mixed and incubated at room temperature for 5 min.

Component	Volume (µL)
Opti-MEM	50
Lipofectamine 2000	4

Component	Volume (µL)
OptiMEM	50
sgRNA (500 ng)	variable
BFP plasmid (500 ng)	variable

After incubation the contents of both tubes were combined and incubated for a further 10 min. During incubation 900 µL of fresh media was added to cells.

After 10 min the contents of the tube were added to the cells, left for 6 hr after which a complete media change was performed with DMEM non-selective media.

2.2.1.3 Fluorescent activated cell sorting (FACS) analysis of transfected cells

48 hr after transfection cells were dissociated to a single cell suspension using Trypsin/EDTA and FACS was used to sort cells based on those expressing BFP, which serves as a proxy for successful transfection. Cells were sorted on the FACS Aria Fusion (BD Biosciences) selecting for the top 10% of cells expressing BFP and single cells were plated in 96-well plates. Single cells were then expanded for subsequent genotyping to screen for *FAN1* knock-outs.

2.2.1.4 Genotyping of FAN1-targeted clones

After clonal expansion DNA was extracted, as described in section 2.5.1 . After extraction of genomic DNA from all clones PCRs amplifying either exon 2 or exon 4 of *FAN1* around targeted regions were carried out using the MangoTaq™ Polymerase (Bioline).

Primer name	Target region	Forward(F) / Reverse(R)	Sequence (5′ - 3′)	Annealing temperature (°C)
oVIN 2875	Exon 2	F	ACTCATGATGTCAGAAGGGAAAC	53
oVIN 2884	Exon 2	R	ACCGGCCAAATTCTTCATCCT	53
oVIN 2881	Exon 2	F	ACGTGAGGCATGTCATTGTG	53
oVIN 2887	Exon 2	R	CAGCTGCTTCCAATGTAACCA	53

oVIN 2889	Exon 4	F	CAGGTGTTGAGATGCCCAAA	53
oVIN 2876	Exon 4	R	TGGTCCTGGAAATTAGTACGGT	53

PCR reactions were as follows:

Component	Volume (µL)
Nuclease-free H_2O	14.25
Buffer	5.25
F primer (10 µM)	1.25
R primer (10 µM)	1.25
MgCl2	0.5
dNTPs	0.5
DMSO	0.75
Polymerase	0.5
DNA (50-100 ng)	1
Total	**25**

Reactions were vortexed and placed in the Mastercycler nexus GSX1 (Eppendorf) under the following conditions:

Cycles	Temperature (°C)	Time
1	95	5 min
	95	30 s
35	53	30 s
	72	1 min
1	72	10 min

PCR products were then sent for sanger sequencing, as described in 2.6

2.2.2 Generating stable FAN1 variant lines in GFPSNick120 cells

Stable FAN1 variant lines were generated in the GFPSNick120 *FAN1*$^{-/-}$ clone #7 line, as described in section 3.2.2. In order to generate lines expressing FAN1 variant constructs cells were transduced with the specific FAN1 construct, which had been packaged into a lentiviral vector. Plasmids were designed using VectorBuilder services and lentiviral constructs were packaged by Vectorbuilder (www.vectorbuilder.com). All plasmid constructs contained a mCherry tag under the control of a cytomegalovirus (CMV) promoter, which allowed for screening of successfully transduced cells.

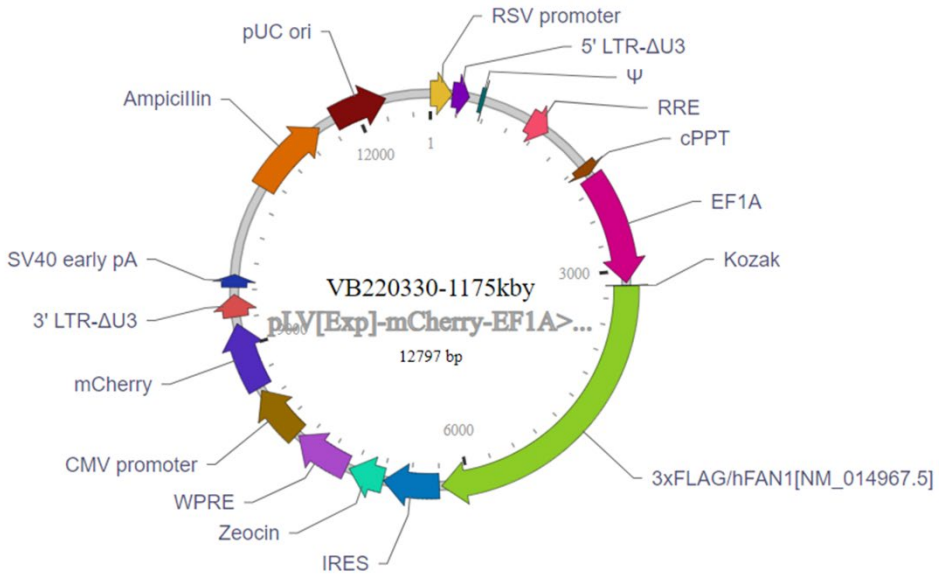


Figure 2-2: Demonstrative FAN1 variant construct map.

On the day before transduction, cells were split as described in 2.1.2, and $4x10^5$ cells were re-plated onto a 6-well culture dish. Cells were then left to adhere overnight. The following day a full media change was carried out and 1 mL of non-selective DMEM media with 5 µg/mL polybrene was added to each well. Cells were then transduced with the lentiviral constructs at a multiplicity of infection (MOI) of 1. Cells were left for 48 hr. Following which lentivirus was removed with a complete media change with 2 mL of DMEM-non selective media was performed. Cells were then left to reach 70-80% confluence. After cells reached confluence they were split as previously described and dissociated into a single-cell suspension. Cells were then sorted by FACS on the FACS Aria Fusion (BD Biosciences). Three populations of cells were sorted into separate FACS tubes; no mCherry, low mCherry (bottom 50%) or high mCherry (top 50%) (see 4.3.1.3). These populations were then re-plated and expanded. Upon expansion cells were frozen, as described in 2.1.2, to make stable stocks. Additionally, DNA and protein samples were taken to confirm integration of constructs via PCR (see below) and for western blotting to assess FAN1 protein levels.

A PCR amplifying the flanking regions of the FAN1 insert was carried out to screen for variant integration. PCR products were subsequently sent for Sanger sequencing, as described in 2.6. PCR reactions were as follows:

Component	Volume (µl)
Nuclease-free H^2O	14.25
Buffer	5.25
F primer (10µM)	1.25
R primer (10µM)	1.25
MgCl2	0.5
dNTPs	0.5
DMSO	0.75
Polymerase	0.5
DNA (50-100ng)	1
Total	**25**

Reactions were vortexed and placed in the Mastercycler nexus GSX1 (Eppendorf) under the following conditions:

Cycles	Temperature (°C)	Time
1	95	5 min
	95	30 s
35	53	30 s
	72	1 min
1	72	10 min

Primer name	Forward (F)/ Reverse (R)	Sequence (5′ - 3′)
oVIN 3562	F	CGGCTTTTGGAGTACGTCG
oVIN 3563	R	TTGCATTCCTTTGGCGAGAG

2.2.3 Generating a *FAN1/MLH1* knock-out in GFPSNick120 cells

2.2.3.1 sgRNA design

To generate a double *FAN1* and *MLH1* knock-out in the GFPSNick120 cells a validated *FAN1* knock-out clone, clone #7 was used (as described in section 3.2.2). An available plasmid in our lab containing the *Staphylococcus aureus* Cas9 and sgRNA backbone was utilised (Addgene:#96920). This was modified through the use of golden-gate cloning to contain an *MLH1*-targeting sequence, targeting exon 2 of MLH1.

sgRNA	Location	Sequence (5′ - 3′)
MLH1	Exon 2	CAAGACAATGGCACCGGGAT

Cells were subsequently transfected and sorted via FACs, and genotyped as described in 2.2.1.2 and 2.2.1.4. PCR primers are detailed below:

Primer name	Forward (F)/ Reverse (R)	Sequence (5′ - 3′)
oVIN 3525	F	CCGTCTCTTCCCTCTCTCTC
oVIN 3526	R	ACCCATTTTGTCTCCCACCA

2.3 Transformation of plasmids in DH5α cells

All plasmids were transformed into competent DH5α cells as follows; bacterial cells were removed from -80°C storage and thawed on ice, during which time plasmid DNA was diluted to 50 ng. Plasmid DNA was added to 80 µL of bacteria and incubated on ice for 30 min, following which, cells were heat-shocked at 42°C for 45 s. Tubes were then immediately placed on ice for a further 2 min. 200 µL of warm LB medium was added to each sample and left to agitate at 37°C for 1 hr. Upon completion of incubation, bacteria was coated onto LB agar plates containing 100 µg/mL ampicillin.

Following overnight incubation at 37°C, single colonies were selected and grown up in 3 mL of LB broth containing ampicillin (100 ug/mL), after which cells were centrifuged and pellets collected for plasmid DNA extraction. If increased yields were required, an additional incubation of bacteria in 20 mL LB broth and ampicillin overnight was first performed.

2.3.1 Isolation of plasmid DNA

Plasmid DNA was extracted utilising the GeneJET Plasmid Miniprep Kit (ThermoFisher Scientific) or the NucleoBond™ Xtra Midi Kit (Machery-Nigel™) following manufacturers guidelines.

2.3.2 Generating knock-outs in GFP(CAG101) cells

2.3.2.1 sgRNA design and synthesis

For the generation of *FAN1 or MLH1* knock-outs in GFP(CAG101) cells, two CRISPR-Cas9 RNAs (crRNAs) targeting exon 2 of *FAN1* or exon 1 of *MLH1* were synthesized. *FAN1* sgRNAs have been described previously in McAllister et al. 2022 and *MLH1* sgRNAs were designed using DESKGEN Cloud based on their predicted high on-target efficiency and low off-target scores.

sgRNA	Location	Sequence (5′ - 3′)
FAN1 target 1	Exon 2	CTGATTGATAAGCTTCTACG
FAN1 target 2	Exon 2	GCACCATTTTACTGCAAACG
MLH1 target 1	Exon 1	TGGCGCAAAATGTCGTTCG
MLH1 target 2	Exon 1	CGGCGGGGGAAGTTATCCAG

To create functional sgRNAs a universal *trans*-activating RNA (tracrRNA) was duplexed with the desired target crRNAs. sgRNA oligos were synthesised and resuspended IDTE (1XTE solution) to a final concentration of 200 µM. crRNAs were then duplexed with the tracrRNA, tagged with a fluorescent ATTO550 (IDT) marker, to create a functional sgRNA. The crRNA and tracrRNA were annealed in equimolar concentrations at 95°C for 2 min with IDT duplex buffer (30 mM HEPES, 100 mM potassium acetate).

Component	Volume (µl)
crRNA	1.7
tracrRNA	1.7
Duplex Buffer	1.7
Total	**5.1**

After sgRNA synthesis the Cas9 protein was diluted from an initial stock of 10 µg/µL to a final stock of 6.2 µg/µL in PBS. To form a functional ribonucleoprotein (RNP) complex containing the Cas9 and sgRNA both

components were mixed together and incubated at room temperature for 20 min.

Component	Volume (µL)
Cas9	2.5
sgRNA	2.5
Total	**5**

After each individual sgRNA was incubated with Cas9 to form a RNP complex, the complexes were mixed, to create a final 10 µl solution containing either both *FAN1* sgRNAs or both *MLH1* sgRNAs. Treatment of cells with two sgRNAs for each target allowed for screening homozygous knock-outs by targeted deletion (as described in 3.2.4.2).

2.3.2.2 Nucleofection of HEK293 cells

Cells were nucleofected to generate knock-out lines. Cells were transfected using the Lonza 4D-Nucleofector and P3 Primary Cell 4D-Nucleofector X kit, following manufacture's guidance. For transfection cells were dissociated using Trypsin/EDTA and $1x10^6$ cells resuspended in complete P3 buffer containing; Lonza P3 buffer, P3 supplement and the two separate RNP complexes containing either both *FAN1* sgRNAs or both *MLH1* sgRNAs. Cells were nucleofected using programme CA-137 and immediately after re-plated in pre-warmed non-selective DMEM media.

Component	Volume (µL)
Lonza P3 buffer	78
P3 supplement	22
RNP complex	10
Total	**110**

2.3.2.3 FACS analysis of nucleofected cells

24 hr after nucleofection, FACS was used to sort transfected cells based on ATTO550 fluorescence intensity, to select for cells successfully transfected with the RNP complex. Cells were dissociated to a single cell suspension using Trypsin/EDTA, centrifuged and resuspended in 500 µL DMEM media with Penicillin/Streptomycin (1:1000) and transferred to ice. Cells were sorted on the FACS Aria Fusion (BD Biosciences) selecting for the top 10% of cells expressing ATTO550. Cells were then re-plated at low density onto a 10 cm tissue culture dish containing DMEM non-selective media, to allow for colony growth and screening for edited cells.

2.3.2.4 Genotyping of FAN1 or MLH1-targeted clones

7-9 days after transfection, single colonies were picked into 96-well plates containing DMEM non-selective media. When cells reached confluency, all clones were split into duplicate 96-well plates. Cells grown on one plate were used for clonal expansion, whilst the other was used for genomic DNA extraction and PCR screening of desired edits.

DNA was extracted for genotyping using the QuickExtract™ DNA Extraction Solution. Briefly, 50 uL of QuickExtract buffer was added to each 96-well and plate was placed on a shaker at room temperature for 10 min. Contents were then mixed and transferred for 96-well PCR plate. Plates were placed in the thermocycler, under the following programme:

Temperature (°C)	Time (min)
65	6
95	2
Total	**8**

After DNA extraction to detect the successful deletion of the target region, PCR primers amplifying either *FAN1* exon 2 or *MLH1* exon 1 were used.

Primer name	Forward(F)/ Reverse (R)	Sequence (5′ - 3′)	Annealing temperature (°C)
FAN1 1	F	ACTCATGATGTCAGAAGGGAAAC	60
FAN1 2	R	TGTCTTTGGTGGTGGTGACT	
MLH1 1	F	ATCCTTCTAGGTAGCGGGCA	59
MLH1 2	R	CATACGTCTTGTACGCTTCAA	

PCRs were performed using MangoTaq™ Polymerase (Bioline). PCR reactions were as follows:

Component	Volume (µL)
Nuclease-free H_2O	14.25
Buffer	5.25
F primer (10 µM)	1.25
R primer (10 µM)	1.25
MgCl2	0.5
dNTPs	0.5
DMSO	0.75
Polymerase	0.5
DNA (50-100 ng)	1
Total	**25**

Reactions were vortexed and placed in the Mastercycler nexus GSX1 (Eppendorf) under the following conditions:

Cycles	Temperature (°C)	Time
1	95	5 min
	95	30 s
35	*variable	30 s
	72	1 min
1	72	10 min

2.3.3 Generating *FAN1* knock-outs in HD-iPSC iNeurons

2.3.3.1 sgRNA design and synthesis

FAN1 knock-outs were generated in HD-iPSC iNeurons using the *FAN1* sgRNAs described in 2.3.2.1. All components of sgRNA synthesis and formation of Cas9 RNP complex were the same, except that that tracrRNA had an ATTO647 fluorescent tag. This was due to these iPSCs expressing mApple fluorescence due to the integration of NGN2 transgene.

2.3.3.2 Nucleofection of iPSCs

As previously described, cells were transfected using the Lonza 4D-Nucleofector and P3 Primary Cell 4D-Nucleofector X kit, following manufacture's guidance. Prior to nucleofection cells were pre-treated with 10 μM ROCK inhibitor. After 1 hr cells were dissociated into a single cell suspension using Accutase for 5 min at 37°C, as previously described. Cells were counted and $1x10^6$ cells were resuspended in complete P3 buffer containing; Lonza P3 buffer, P3 supplement and the two separate RNP complexes containing both *FAN1* sgRNAs. Cells were nucleofected using programme CA-137 and immediately after re-plated in pre-warmed E8 flex media with 10 μM ROCK inhibitor.

2.3.3.3 FACS analysis of nucleofected iPSCs

24 hr after nucleofection FACS was used to sort transfected cells based on ATTO647 fluorescence intensity, to select for cells successfully transfected with the RNP complex. Cells were dissociated to a single cell suspension using Accutase, centrifuged at 300 rcf for 3 min and resuspended in 500 μL E8 flex media with Penicillin/Streptomycin (1:1000) and transferred to ice. Cells were sorted on the FACS Aria Fusion (BD Biosciences) selecting for the top 10% of cells expressing ATTO647. Cells were then re-plated at low density onto a 10 cm^2 tissue culture dish containing E8 flex media to allow for colony growth and screening for edited cells.

2.3.3.4 Genotyping of FAN1-targeted cells

5-7 days after transfection, single colonies were picked into 96-well plates containing E8 flex media supplemented with 10 μM ROCK inhibitor. When cells reached confluency all clones were split into duplicate 96-well plates. Cells grown on one plate for used for clonal expansion, whilst the other was used for genomic DNA extraction and PCR screening of desired edits. Genomic DNA extraction and genotyping by PCR was carried out using FAN1 primers as described in 2.3.2.4

2.4 GFP-based reporter assay

In order to monitor nickase-induced contractions in GFPNickS120 cells (described in 2.1.2.1) one technique used was to monitor changes in GFP expression which served as a proxy for changes in CAG repeat tract length (described in 1.10.3 and 3.2.1). Cells were cultured in DMEM media supplemented with doxycycline (1 μg/mL) for 7, 14 and 21 days, to induce transcription through the GFP-Pem1 construct. Cells were split as previously described and media was supplemented 3 times weekly.

2.4.1 Flow cytometry

After cells had been cultured for the correct length of time, cells were dissociated using Trypsin/EDTA and centrifuged at 300 rcf for 5 min. Cell pellets were then washed gently with PBS, and re-centrifuged. Cells were then resuspended in PBS containing 1 μg/ml of DAPI. This was important as it allowed for the removal of dead cells from downstream analysis. Appropriate controls were also applied; including an unstained control (no DAPI or GFP), a GFP only sample and a DAPI only sample. These samples allowed for compensation gating to be set in the downstream analysis.

2.4.2 Analysis of flow cytometry data

All analysis was performed using version 10.1 of FlowJo™. The gating strategy was as follows (Figure 2-3); gating of the HEK293 cell population was performed and a compensation matrix was applied to all samples. After selection of cell populations, cell doublets were removed from the analysis by

plotting FSC-A versus FSC-H. After doublet exclusion cell populations were divided into quadrants based on GFP and DAPI expression. The unstained control was first used to discriminate between background GFP levels, and served to set quadrant gates. Quadrant gating resulted in four populations; (Q1) DAPI+/GFP-, (Q2) DAPI+GFP+, (Q3) GFP+/DAPI- and (Q4) GFP-/DAPI. The Q3 population, representing those cells that express GFP, but importantly not DAPI, were selected and a histogram generated to view GFP expression distribution. For analysis two methods were used; firstly the mean GFP expression for all samples were calculated by FlowJo™ and mean GFP expression at day 7 and day 21 were compared and the fold-change calculated. The second mode of analysis involved gating the top 10% of GFP-expressing cells. The fold-change was then calculated between day 7 and day 21.

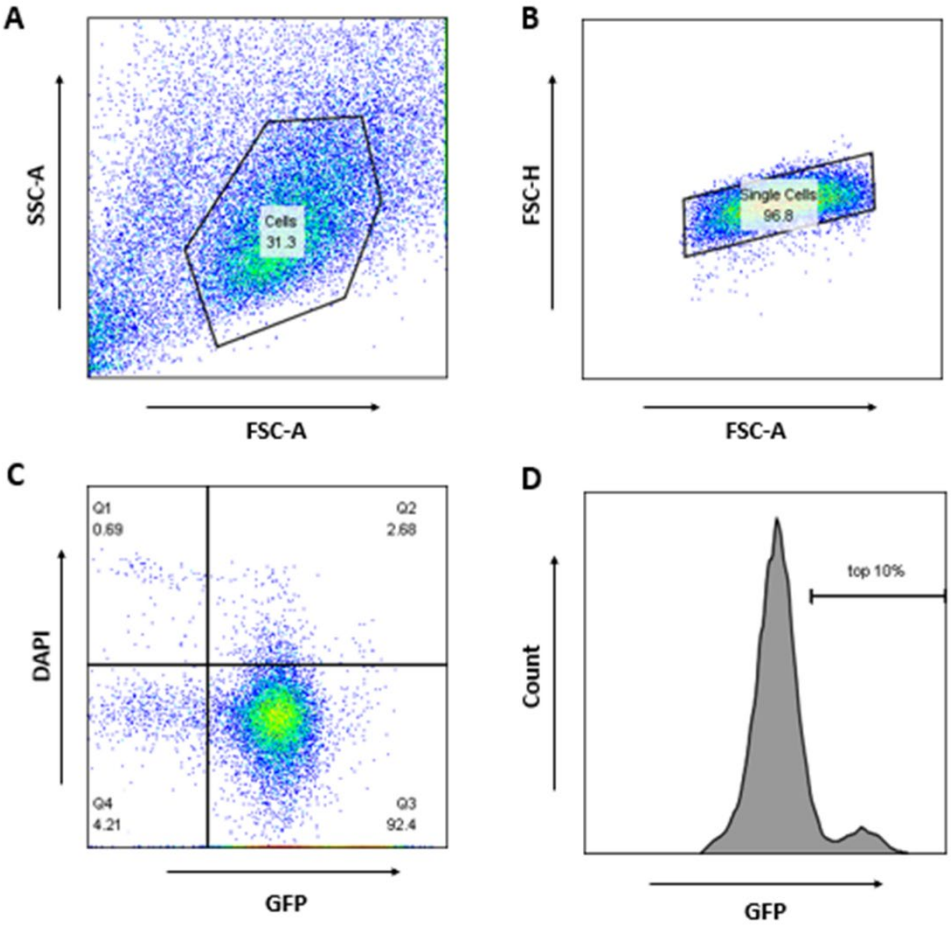


Figure 2-3: FlowJo™ gating strategy. ***(A)*** *HEK cell populations were selected based on forward scatter (FSC) and side scatter (SSC) profiles.* ***(B)*** *Doublets were excluded by plotting FSC-A versus FSC-H.* ***(C)*** *Cell profiles were split into quadrants based on GFP and DAPI intensity. Population Q3, represents cells of interest with GFP expression, but no DAPI expression. Unstained control cells were used to set quadrant gates.* ***(D)*** *GFP expressing cells were gated for the second set of analysis based on the top 10% of GFP-positive cells.*

2.5 Nucleic Acid Extraction

2.5.1 DNA Extraction

DNA extraction was performed using the NucleoSpin Tissue Kit® (Macherey-Nagel) following manufacturer's instructions. In brief, cell pellets previously stored at -80°C were resuspended in T1 buffer and 25 µL proteinase K was added. To ensure the complete removal of RNA from the sample 20 µL RNase A was added, following which all samples were incubated for 5 min at 37°C. Samples were then incubated for a further 15 min at 70°C after the addition of an additional buffer. Ethanol (96-100%) was then added to the sample and vortexed vigorously. All samples were added to the extraction column and a series of washes and centrifugation steps were performed. DNA was eluted into 30-100 µL nuclease-free water/BE buffer. DNA was subsequently stored at -20°C.

2.5.2 RNA Extraction

RNA was extracted using the RNeasy mini kit (QIAGEN) following manufacturer's instructions. Cell pellets previously stored at -80°C were lysed in 350 µL of BLT buffer which contained 10 µL/mL β-mercaptoethanol. Equal parts of 70% ethanol was added to the lysate and mixed well before being added to the RNeasy spin column and centrifuged. A DNA digestion step was next carried out to ensure the removal of DNA from the sample. The column was further treated with a series of buffers and centrifugation steps were performed in order to wash the column. RNA was eluted into RNAse-free water and stored at -80°C.

2.5.2.1 Nucleic acid quantification

RNA concentrations were determined using the Qubit™ RNA BR assay. For routine applications DNA concentration was determined by the DeNovix DS-11 Series spectrophotometer/fluorometer. The quality of both RNA and DNA was assessed by calculating the 260/230 and 260/280 ratios.

For long-read PacBio sequencing (see 2.10) DNA concentrations were recorded using the Qubit™ DNA HS assay, following manufacturer's guidelines.

2.6 Sanger sequencing

For Sanger sequencing, genomic DNA was extracted as described in 2.5.1. PCRs were performed using relevant primers with 50-100 ng of genomic DNA. PCR products were purified using the NucleoSpin Gel and PCR Clean-up kit (Macherey-Nagel) following manufacturer's guidelines. Purified PCR products were eluted 30 µL in nuclease-free water/NE buffer. 5 µL of purified PCR product were added to an Eppendorf with either 5 uL of the forward (5 µM) or reverse primer (5 µM). Samples were then sent to Genewiz for sequencing.

2.7 Agarose gel electrophoresis

PCR products were loaded on 1.5% agarose gels made with 1% Tris-acetate-EDTA (TAE). 8 µL of the Quick-Load® 1 Kb Plus DNA ladder (New England Biolabs) was loaded in the first lane, followed by 10 µL of each PCR reaction and a water only control. Gels were run at ~70-100 V for 1 hr, or longer depending on size of PCR product. Gels were imaged using a G:BOX Chemi XX9 imager (Syngene).

2.8 Quantitative Real-Time Polymerase Chain Reaction (qRT-PCR)

2.8.1 cDNA synthesis

Reverse transcription of RNA to cDNA was carried out using the qScript cDNA Synthesis Kit (Quantabio). Reactions were kept on ice and performed in 0.2 mL PCR tubes. In each case 1 µL of the reverse transcriptase and 4 µL of 5X reaction mix was added to a 15 µL solution containing the desired ratio of RNA and nuclease-free water.

Once all components were added the reaction mixture was vortexed gently and centrifuged. Reactions were then placed in the Mastercycler nexus GSX1 (Eppendorf) on the following programme:

Cycles	Temperature	Time (min)
1	22°C	5
1	42°C	30
1	85°C	5

2.8.2 TaqMan RT-qPCR

Quantification of target gene transcripts was performed on the StepOne plus instrument (Applied Biosystems) using TaqMan Standard Gene Expression Master Mix and gene specific probes were pre-designed by Life Technologies.

Gene name	TaqMan Assay ID
FAN1	Hs00429686_m1 FAN1
GAPDH	Hs02758991_g1 GAPDH

A reaction mix of 10 µL was made up containing:

Reagent	Volume (µL)
TaqMan Gene Expression Master Mix	5
TaqMan Assay	0.5
Nuclease free H_2O	0.5
cDNA (100 ng)	4
Total	**10**

Reactions were performed in MicroAmp® Fast Optical 96-well reaction plates (Applied Biosystems) and plates sealed with the MicroAmp® Optical adhesive film (Applied Biosystems). Plates were then briefly vortexed and centrifuged before being placed in the StepOne plus instrument (Applied Biosystems). The thermocycler reaction was as follows:

Cycles	Temperature (°C)	Time
1	50	2 min
1	95	10 min
40	95	15 s
	60	1 min

Data was analysed in Microsoft Excel utilising the delta-delta CT method to determine relative gene expression. All samples were run in triplicate with three biological replicates.

2.8.3 SYBR RT-qPCR

Quantification of target gene transcripts was performed on the StepOne plus instrument (Applied Biosystems) using the FastStart Universal SYBR Green Master Mix (Rox) (Sigma). Primers used had been designed and validated previously, with established efficiencies. All primers used were diluted to a 10 µM concentration. A master mix was made up and 8 µL pipetted into each well and 2 uL cDNA (50 ng) added.

Reagent	Volume (µL)
FastStart Universal SYBR Green Master Mix	5
Forward + Reverse primers (10 µM)	0.5
Nuclease-free water	2.5
cDNA (50 ng)	2
Total	**10**

Reactions were performed in MicroAmp® Fast Optical 96-well reaction plates (Applied Biosystems) and plates sealed with the MicroAmp® Optical adhesive film (Applied Biosystems). Plates were then briefly vortexed and centrifuged before being placed in the StepOne plus instrument (Applied Biosystems). The thermocycler reaction was as follows:

Cycles	Temperature (°C)	Time
1	95	1 min
	95	10 s
40	60	30 s
	72	30 s

Target	Forward (F)/ Reverse (R)	Sequence (5′ - 3′)
OCT4	F	CTCACCCTGGGGGTTCTATT
	R	CTCCAGGTTGCCTCTCACTC
SOX2	F	TACCTCTTCCTCCCACTCCA
	R	GGTAGTGCTGGGACATGTGA
NANOG	F	GGTAGTGCTGGGACATGTGA
	R	GATGGGAGGAGGGGAGAGGA
B- ACTIN	F	CCCAGCACAATGAAGATCA
	R	ACATCTGCTGGAAGGTGGAC

Table 2-6: List of primers used for SYBR RT-qPCR.

2.9 Small-pool PCR in GFPNickS120 cells

2.9.1 PCR protocol

Genomic DNA was extracted from samples as previously described in 2.5.1. A 10-fold dilution series was undertaken to determine optimal DNA concentrations at which to best visualise repeat size alleles. All PCR reactions were performed using the MangoTaq™ DNA polymerase (Bioline). PCR reactions were all 10 µL and performed in 0.2 mL PCR tubes as follows:

Reagent	Volume (µL)
5x Mango Buffer	2
$MgCl_2$	0.2
dNTPs (10 mM)	0.2
Forward primer (10 µM – oVIN459)	0.5
Reverse primer (10 µM – oVIN460)	0.5
DMSO	0.3
MangoTaq DNA Polymerase	0.2
H_20	6.1
DNA (concentration variable)	1
Total	**10**

Primer name	Target	Sequence (5′ - 3′)
oVIN 459	GFP Pem1 intron	AAGAGCTTCCCTTACACAACG
oVIN 460	GFP Pem1 intron	TCTGCAAATTCAGTGATGC

Reactions were vortexed and centrifuged and placed in the Mastercycler nexus GSX1 (Eppendorf) under the following conditions:

Cycles	Temperature (°C)	Time
1	95	5 min
	95	20 s
5	52	20 s
	72	4 min
	95	30 s
25	55	30 s
	72	1 min 30 s
1	72	10 min

2.9.2 Agarose gel electrophoresis

PCR products were loaded on 1.5% agarose gels made with 1% TAE. 8 µL of the Quick-Load® 1 Kb Plus DNA ladder (New England Biolabs) was loaded

into the first and last lane of each gel and 10 µL of all samples were subsequently loaded. Gels were run overnight at 140 V for 60 min and for 35 V for ~16 hr. Gels were imaged using a G:BOX Chemi XX9 imager (Syngene).

2.9.3 Southern blotting

Upon the completion of gel electrophoresis step all gels were imaged. Importantly gels were imaged with fluorescent rulers placed at either side of the gels which was essential for downstream analysis. Gels were then cut in function of PCR product size and washed twice in alkaline transfer buffer (0.4 M NaOH, 1 M NaCl) for 20 min. Gels were then measured, and two pieces of filter paper and one piece of nylon membrane was cut to size. DNA was then transferred onto the nylon membrane overnight by capillary action (Dion et al. 2008).

2.9.4 γ-32^{P}-dATP probe preparation

Prior to hybridisation of the membranes the radioactive probe was prepared as below:

Reagent	Volume (µL)
Buffer T4 PNK	2.5
oVIN100 (10 µM)	5
H_20	11.5
γ-32^{P}-dATP (10 mM)	5
T4 PNK	1

Reactions were then placed in thermocycler under the following conditions:

Cycles	Temperature	Time
1	37°C	30 min
1	65°C	20 min

Primer name	Primer sequence
oVIN-100	AGCAGCAGCAGCAGCAGCAGCAGCAGC

2.9.5 Membrane hybridisation

Following successful transfer, membranes were washed in neutralisation buffer (1.5 M NaCl, 0.5 M Tris base, pH 7.4) for 5 min. Membranes were incubated in ULTRAHyb (ThermoFisher Scientific) buffer and salmon sperm DNA for 1 hr at 48°C. Following this the probe was added to the membrane for 2 hr at 48°C. After complete hybridisation, membranes were washed in 15 mL of pre-heated (48°C) wash buffer (0.5X SSC, 0.1% SDS) and incubated for 30 min. This step was performed twice. After the above steps were performed membranes were exposed to a phosphoscreen at room temperature for 2 to 48 hr and images revealed with Bio-Rad's Personal Molecular Imager (PMI) and Quantity One software.

2.9.6 spPCR allele counts

After annotation of spPCR membranes using Adobe Photoshop and Illustrator, samples were blinded. A Poisson distribution was used to calculate the predicted number of total alleles amplified per blot (described in Gomes-Pereira, Mario Bidichandani and Monckton 2004).

2.10 Long-read PacBio sequencing in GFPNickS120 cells

2.10.1 PCR amplification

Genomic DNA was extracted as described in 2.5.1. Prior to PCR amplification DNA quality was assessed on the DeNovix DS-11 Series spectrophotometer/fluorometer. The quality of DNA was assessed by calculating the 260/230 and 260/280 ratios. Samples with scores within range were then taken forward for DNA quantification using the Qubit™ DNA HS assay, following manufacturer's guidelines.

PCR amplification of the GFP-Pem1 insert was carried out using MangoTaq™ DNA polymerase (Bioline). PCR reactions were all 50 µL and carried out in 0.2 mL PCR tubes, with a typical input of between 100-150 ng of DNA. PCR reactions were carried out as below:

Reagent	Volume (µL)
5x Mango Buffer	10
$MgCl_2$	1.5
dNTPs (10 mM)	0.5
Forward primer (10 µM – oVIN 459)	2.5
Reverse primer (10 µM – oVIN 2514)	2.5
DMSO	1.6
MangoTaq DNA polymerase	1
DNA (100-150 ng)	variable
Nuclease-free water	Up to 50 µL

Reactions were vortexed briefly and then placed in the Mastercycler nexus GSX1 (Eppendorf) under the following conditions:

Cycles	Temperature (°C)	Time
1	95	5 min
	95	30 s
33	62	30 s
	72	1 min 30 s
1	72	10 min

After PCR amplification 5 µL of each PCR reaction was taken forward for visual assessment via gel electrophoresis to confirm amplification, as in section 2.7. The remainder was taken forward for PCR-clean up and PacBio SMRTbell library preparation.

2.10.2 PCR-clean up and quantification

For each sample, 3 individual PCR reactions were carried out and pooled. The pooled PCR's were then cleaned-up following PacBio instructions for the generation of SMRTbell® libraries using PacBio® Barcoded Overhang Adapters for Multiplexing Amplicons. PCR clean-ups were carried out using AMPure® PB beads (PacBio) with 0.6X beads per PCR volume, as recommended by manufacturer. Briefly, PCR products were bound to beads

and left to incubate at room temperature for 5 min. After 5 min reaction mixtures were bound to DynaMag™ beads and washed twice with 80% ethanol. The bead pellet was left to dry and then cleaned-up PCR reactions were eluted off the beads in elution buffer after a 15 min incubation at 37°C (PacBio).

After PCR amplicons were cleaned-up 1 µL was taken forward and diluted in 9 µL of elution buffer. This was used for the quantification of PCR amplicons using the Qubit™ DNA HS assay, following manufacturer's guidelines.

2.10.3 SMRTbell library construction and quantification

After PCR amplicons were cleaned-up, as described above, library preparation was carried out. Briefly, each PCR amplicon was subject to; DNA Damage repair, End-repair/A-tailing and subsequent ligation with a barcoded overhang adaptor using the SMRTbell® Express Template Prep Kit 2.0. Following barcoding all samples were multiplexed following manufacturer's instructions for total DNA (ng) mass requirements. The final pooled library was then subjected to two sequential clean-up steps, as described in 2.10.2. The final library was eluted in 50 µL elution buffer (PacBio) with 1 µL taken for quantification using the Qubit™ DNA HS assay, following manufacturer's guidelines.

2.10.4 Fragment analysis

Prior to sequencing the purified library was run on the 5200 Fragment Analyser (Agilent) to determine the quality of the library preparation. Libraries were diluted to 0.5 ng/µL based on Qubit™ readings and 1 µL used for fragment analysis.

2.10.5 Sequencing

Libraries were diluted to the required concentration and sequenced according to manufacturer's guidance on the Sequel® II system. Analysis was then performed using Repeat Detector (Taylor et al. 2022).

2.11 Protein extraction and western blotting

2.11.1 Protein extraction and quantification

Radio-immunoprecipitation assay buffer (RIPA) (Thermo Fisher) containing an EDTA-free protease inhibitor (Merck) was added to cell pellets in variable amounts depending on sample size. Pellets were resuspended and placed on a shaker to incubate at 4°C for 30 min. Samples were then centrifuged at 12,000 rpm for 20 min at 4°C. Following this, the protein supernatant was transferred to a pre-cooled 1.5 mL Eppendorf and stored at -80°C.

Protein samples were quantified using the Pierce bicinchoninic acid assay (BCA) protein assay kit (Thermo Fisher). Concentrations were determined relative to dilutions of a standard protein bovine serum albumin (BSA) ranging from 0-2000 µg/mL. A BCA working reagent was prepared by mixing reagent A and B in a 50:1 ratio. 25 µL of each standard and protein sample was pipetted into a 96-well plate in duplicate. 200 µL of the working reagent was then added to each well and mixed on a plate shaker for 30 s. Samples were then incubated at 37°C for 30 min. Plate absorbance was read at 562 nm. Protein concentration was then extrapolated based on a standard curve of absorbance measures from protein standards.

2.11.2 SDS-page and western blotting

Protein samples were thawed on ice and a 1X sample solution was made by adding 5 µL of 4X LDS (Fisher Scientific) sample buffer to a 20 µL solution containing the protein sample at the desired concentration and deionised H_2O. Samples were denatured at 70°C for 10 min. A running buffer was prepared by diluting 50 mL of MES SDS running buffer (20X) (Fisher Scientific) in 950 mL of deionised H_2O. 500 µL of Bolt™ antioxidant was added and the buffer thoroughly mixed. The buffer was added to the Bolt Mini Gel Tank and samples were loaded alongside a protein standard (Precision Plus Protein™ Kaleidoscope™, Bio-Rad) and separated on a 4-12% Bis-Tris gel (Thermo Fisher). Samples were then transferred to a nitrocellulose membrane in a 1X prepared transfer buffer for 1 hr at 10 V. The membrane was blocked in a 5% milk in phosphate buffered saline containing 1% Tween (Merck) (PBS-T) for 1

hr at room temperature. Membranes were then incubated in the relevant primary antibody in 5% milk in PBS-T overnight at 4°C on a rotating tube mixer. The following day, membranes were washed in PBS-T for 3 x 10 min intervals and Alexa Fluor secondary antibodies diluted in 5% milk in PBS-T were added for 1 hr at room temperature. Membranes were washed in PBS-T for 3 x 10 min intervals before visualisation on the Licor Odyssey Clx.

Target	Species	Dilution	Supplier	Catalogue Number
B-actin	Mouse	1:2000	Sigma	A5441
FAN1	Sheep	1:500	CHDI	N/A
Cas9	Mouse	1:5000	Diagenode	C15200216
MLH1	Mouse	1:500	BD Biosciences	554073

Table 2-7: List of antibodies used for western blotting.

2.12 Immunocytochemistry (ICC)

IPSCs and neurons were plated for ICC in 96-well µClear (Greiner) plates. On the day of staining media was removed and cells were washed gently once with PBS. Cells were then fixed with 4% paraformaldehyde (PFA) (Scientific Laboratory Supplies) for 15 min at room temperature. Fixed cells were then washed 3 times with PBS before staining. If staining did not take place immediately, cells were stored in PBS at 4°C. Prior to staining with primary antibodies, cells were permeabilised and blocked for 1 hr at room temperature in a blocking solution containing 0.1% Triton-X (Sigma-Aldrich), 1% BSA and either 3% normal goat serum (NGS) or normal donkey serum (NDS) (Merck), dependent upon the secondary antibody used (Table 2-9). After 1 hr, blocking solution was removed and cells were incubated in primary antibodies overnight at 4°C. Following incubation antibodies were removed and cells were washed 3 times with PBS. Cells were then incubated with the appropriate secondary antibody for 1 hr at room temperature in the dark. During this step the secondary antibody solution also contained the nuclear stain DAPI. After

incubation cells were washed 3 times in PBS and were then imaged on Opera Phenix (Perkin Elmer) imaging platform.

1° antibody	Species	Supplier	Dilution	Blocking
OCT4	Rabbit	Cell signalling technologies (4903)	1:100	0.1% Triton-X, 1% BSA, 3% NGS
SOX2	Rabbit	Cell signalling technologies (2840)	1:100	0.1% Triton-X, 1% BSA, 3% NGS
NANOG	Rabbit	Cell signalling technologies (3579)	1:000	0.1% Triton-X, 1% BSA, 3% NGS
MAP2	Chicken	Abcam (ab92434)	1:500	0.1% Triton-X, 1% BSA, 3% NDS
Ki67	Rabbit	Abcam (ab15580)	1:500	0.1% Triton-X, 1% BSA, 3% NGS

Table 2-8: List of primary antibodies used for ICC.

2° antibody	Supplier	Catalogue number	Dilution
Alexa Fluor 488 goat anti-rabbit	Thermo Fisher Scientific	A11008	1:1000
Alexa Fluor 568 goat anti-rabbit	Thermo Fisher Scientific	A11011	1:1000
Alexa Fluor 680 goat anti-rabbit	Thermo Fisher Scientific	A21109	1:1000
Alexa Fluor 647 donkey anti-chicken	Thermo Fisher Scientific	A78952	1:1000

Table 2-9: List of secondary antibodies used for ICC.

2.12.1 Statistical analysis

All statistical analysis in this thesis was carried out in GraphPad Prism version 9.4.1. The appropriate statistical tests were selected dependent upon the experiment (see figure legends for details). For all assays a p value >0.05 was non-significant (ns) and p values <0.05 were the threshold for significance. The annotations in all figure legends are as follows; $p<0.05$ (*), $p<0.01$ (**), $p<0.001$ (***) and $p<0.0001$ (****).

For clarification the term 'biological replicates' in this thesis, relating to all cell culture experiments carried out in Chapters 3-5, refers to individual experiments carried out on either a different day or from a different batch of cells (either from a new cryovial or an independent starting culture population).

3 Exploring whether loss of FAN1 impacts CRISPR-Cas9 nickase contractions

3.1 Introduction

The age at onset for Huntington's disease (HD) is inversely correlated with the length of the *HTT* CAG tract. The CAG tract accounts for ~60% variance in the age at onset, with the remaining variability explained by genetic or environmental factors (Djoussé et al. 2003; McAllister et al. 2021). A recent boom in genome wide association (GWA) studies has identified a number of significant loci, which contain genes associated with DNA repair proteins. These genome-wide significant signals have been mapped to chromosomes 3, 2, 5, 7 and 19, with many of the genes thought to be responsible for these signals coding for proteins within the Mismatch repair (MMR) pathway (Lee et al. 2015; Lee et al. 2019).

The biggest genome-wide significant hit was mapped to the long arm of chromosome 15, implicating *FAN1.* FAN1 is a 5′ to 3′ exonuclease and a structure-specific endonuclease that has previously been implicated in ICL repair (Kratz et al. 2010; Liu et al. 2010b; Smogorzewska et al. 2010) and the recovery of stalled replication forks (Chaudhury et al. 2014; Porro et al. 2017). Recessive *FAN1* mutations cause a rare chronic kidney disease, KIN, which is characterized by progressive renal impairment (Zhou et al. 2012).

To date, four modifier SNPs at chromosome 15 have been identified through GWA studies, two of which, *rs150393409* and *rs151322829*, code for amino acid changes R507H and R377W in FAN1 (Lee et al. 2019). These variants have also been identified through exome sequencing of HD individuals heterozygous for these mutations (McAllister et al. 2022). Exactly how these variants contribute to HD pathogenesis remains to be seen. It has been suggested that these variants may result in loss of function, are hypomorphic or have dominant-negative or gain-of-function effects (Bastarache et al. 2018; Lee et al. 2019; McAllister et al. 2022). The other two SNPs are associated with a delayed age at onset and correspond with cis-expression quantitative

trait loci (cis-eQTL) for increased expression of FAN1 in the cortex (Lee et al. 2015; Lee et al. 2019).

One way genetic modifiers, such as *FAN1*, likely alter HD age at onset is by modulating somatic instability of the CAG repeat tract (Kennedy and Shelbourne 2000; Kovalenko et al. 2012). This is supported by a plethora of studies, pre-dating the GWA studies, in both transgenic and knock-in mouse models which demonstrate MMR proteins promote somatic expansion of CAG repeats. Indeed, loss of both Msh2 and Msh3 is sufficient to ablate somatic expansion in the striatum *in vivo* (Manley et al. 1999; Wheeler et al. 2003; Owen et al. 2005; Dragileva et al. 2009; Kovalenko et al. 2012; Tomé et al. 2013). Similarly, loss of Mlh1 or Mlh3, ablates somatic expansion in HD knock-in mice (Pinto et al. 2013; Roy et al. 2021).

Contrastingly to the MMR machinery, FAN1 is proposed to be important in protecting against somatic CAG expansion. The generation of *FAN1*$^{-/-}$ lines in immortalised and HD-patient derived iPSCs and neurons have demonstrated increased CAG repeat expansion over time, relative to *FAN1*$^{+/+}$ lines (Goold et al. 2019; McAllister et al. 2022). This is supported by *in vivo* data where *Fan1*$^{-/-}$ mice exhibit increased somatic expansion rates in the striatum compared with controls. Interestingly, this observed increase in somatic expansions was completely ablated in dual *Fan1* and *Mlh1* knock-out mice (Loupe et al. 2020). Additionally, a role for FAN1 in protecting against somatic expansion has also been reported in a Fragile-X syndrome mouse model, where loss of FAN1 results in increased somatic CGG expansion in the liver and brain tissue (Zhao and Usdin 2018; Zhao et al. 2021). Taken together, these findings indicate a broad role for FAN1 protecting against somatic expansion, that is applicable to multiple diseases.

Our laboratory has recently designed a modified CRISPR-Cas9 system that can induce contractions in expanded CAG/CTG tracts, whilst avoiding concomitant expansions in human cells. This involves the use of a Cas9 nickase with a sgRNA targeting a CAG tract (Cinesi et al. 2016). This system differs from the traditional Cas9 endonuclease as it induces single-stranded breaks (SSBs) instead of double-stranded breaks (DSBs). The formation of

SSBs are made possible by engineering a D10A point mutation in the RuvC nuclease domain, rendering it inactive (Cong et al. 2013). Interestingly, the ability of the nickase to induce contractions occurs in a length dependent manner, with repeat sizes within the non-pathogenic range remaining unedited (Cinesi et al. 2016). Furthermore, data from our lab indicate the induction of nickase-induced contractions is transcription dependent (Larin, et al; unpublished). This presents an exciting potential therapeutic approach where only the pathogenic allele is corrected. Importantly, this approach has the potential to be adapted for treatment of multiple disorders as the target for the nickase is the CAG repeat itself.

Data from Cinesi et al (2016) indicate that the DNA repair machinery is involved in nickase-induced contractions of CAG repeat tracts, but currently the specific proteins involved are unknown. PARP inhibition and knockdown of XRCC1 failed to yield any change in nickase-induced contraction rates, indicating that canonical SSB repair is an unlikely mechanism for these contraction events. Given the evidence supporting FAN1's role in somatic instability, I sought to determine whether it plays a role in CRISPR-Cas9 nickase-induced contractions. This is important, not only to improve our understanding of the mechanism of contractions, but also to establish whether patients that have rare FAN1 variants may still benefit from this therapy.

3.1.1 Chapter aims:

The aim of this chapter was to generate and validate the knock-out of *FAN1* in HEK293 cells containing a stably integrated CRISPR-Cas9 nickase and stably integrated GFP-reporter system containing an artificial CAG repeat construct (GFPNickS120). This will examine whether a loss of FAN1 impacts the efficiency of nickase-induced contractions. Furthermore, a second cell line, GFP(CAG101), which does not contain the Cas9 treatment allowed us confirm whether the phenotype of *FAN1*$^{-/-}$ in GFPNickS120 cells is specific to the action of the nickase.

3.2 Results

3.2.1 HEK293 GFP-based reporter system

The work detailed in this chapter involves HEK293 cells, which have been stably integrated with a green fluorescent protein (GFP) based reporter assay. This can be been used as a screening method to monitor effects on nickase-induced contractions (Cinesi et al. 2016). The GFP assay is under the control of a doxycycline inducible promoter and contains two exons of a GFP mini gene, separated by an intronic region containing a CAG repeat tract (Santillan et al. 2014). In this system the CAG tract interferes with the effective splicing of GFP exons in a length-dependent manner. Therefore, the longer the repeat tract the lower the GFP expression. This occurs as the expanded tract is not efficiently spliced out, but included as an exon within the mRNA, thus causing a frame shift mutation preventing the synthesis of a functional GFP transcript (Gorbunova et al. 2003; Santillan et al. 2014). This system enables GFP intensity to serve as a proxy for repeat size. In this case, cells that express a similar repeat size will form a narrow GFP distribution. If instability events occur, there will be a widening of the GFP distribution, indicative of contraction and expansion events (Figure 3-1).

In this chapter two cell lines containing the GFP reporter were used; the first were GFPNickS120 cells which contain the GFP reporter as well as a constitutively expressed Cas9 D10A nickase and sgCTG targeting the CAG/CTG repeat tract. These cells allow for the monitoring of Cas9-nickase induced contractions upon the induction of transcription by doxycycline. The second line, GFP(CAG101) cells contain the GFP reporter, but not the Cas9 nickase of sgCTG treatment. These can be used to monitor somatic expansions in these lines (described in 2.1.2.1).

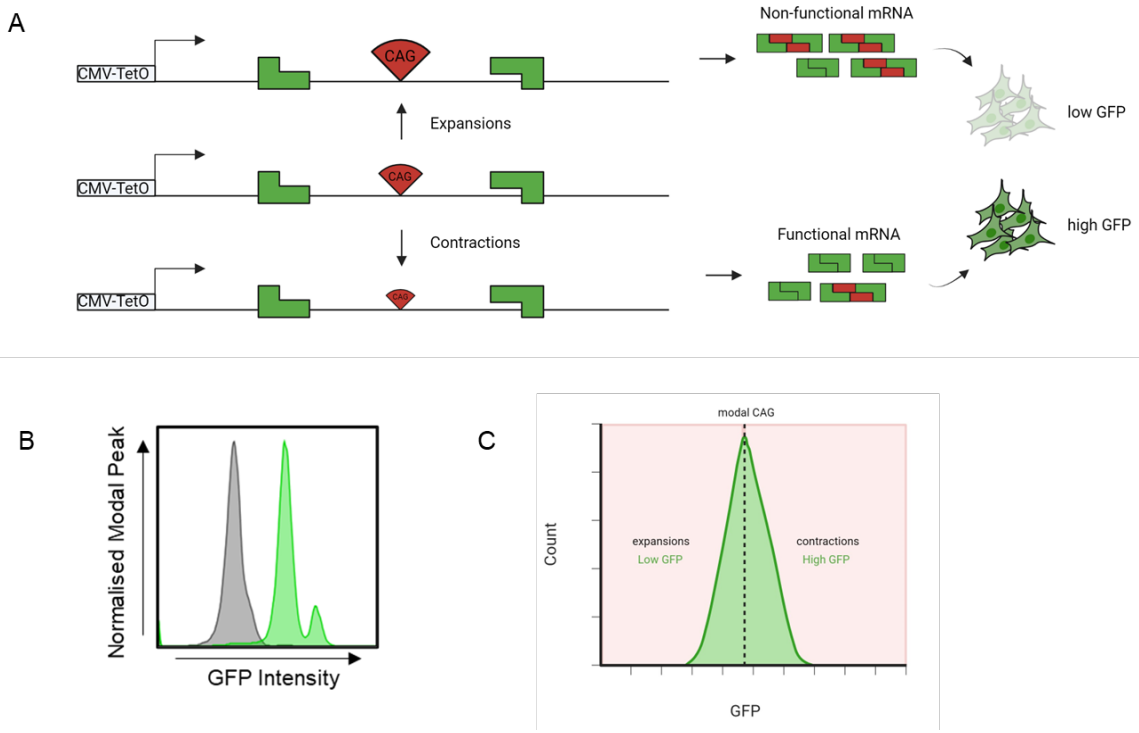


Figure 3-1: A GFP-based reporter for repeat size**.** ***(A)*** *Representation of the ectopic GFP-reporter based assay. A CAG repeat tract resides within an intron separating two exons of a GFP-mini gene. This interferes with mRNA splicing in a length-dependent manner, allowing GFP intensity to serve as a proxy for changes in CAG length.* ***(B)*** *Demonstrative histogram plot where the X axis represents GFP intensity. Cells that are untreated (grey) represent no/little GFP, as transcription has not been induced. Cells treated with doxycycline (green) undergo transcription to induce expression of GFP, demonstrating by a higher GFP profile.* ***(C)*** *Representation of the GFP distribution serving as a proxy for CAG length. Shifts to the right (higher GFP) indicate contraction events and shifts to the left (lower GFP) indicate expansions.*

3.2.2 Generation and validation of *FAN1* knock-out lines in cells containing the CRISPR-Cas9 nickase (GFPNickS120)

3.2.2.1 FAN1 targeting approach

To determine whether FAN1 affects CRISPR-Cas9 nickase-induced contractions of CAG repeats, I sought to knockout *FAN1* in our GFPNickS120 cell line, as described in 2.2.1. Importantly, these cell lines already contain a constitutively expressed *S.pyogenes* Cas9 D10A nickase and a sgCTG targeting the ectopic CAG tract. The challenge of knocking out *FAN1* in these cells is that they already contain the Cas9 nickase. Classic approaches for generating a knock-out cell line, involve the exogenous delivery of a ribonucleoprotein complex (RNP) comprised of a Cas9 endonuclease and functional sgRNA (Bloomer et al. 2022). However, in these cells providing an endogenous RNP containing the Cas9 endonuclease could result in binding with the constitutively expressed sgCTG, inducing DSBs at ~140 sites in the genome. To avoid this, I designed pairs of sgRNAs and utilised the Cas9 nickase already expressed in this cell line. This was necessary as the Cas9 nickase variant induces SSBs when guided to a target site in the genome by a sgRNA. These paired sgRNAs targeting both the sense and anti-sense strands should lead to a local DSB, which can result in a frameshift mutation in the DNA.

gRNAs were designed using Benchling design tools (https://www.benchling.com/crispr). The sgRNAs were selected based on their high editing efficiency and off-target prediction scores. Exon 2 was targeted as this is the first coding exon of *FAN1* and previous work has shown sgRNAs targeting this region result in the successful knock-out of *FAN1* (McAllister et al. 2022). To increase the likelihood of success, a second exon, exon 4, was targeted and all cells were transfected with two pairs of sgRNAs, targeting a combination of regions in exon 2 and/or 4 (Figure 3-2).

Transfections were performed using the transfection agent lipofectamine 2000. To separate the transfected cells from the untransfected ones, cells were co-transfected with a plasmid expressing a blue fluorescent protein (BFP). 48 hrs post-transfection both transfected and non-transfected control

cells were sorted using the BD FACS Aria Fusion flow cytometer. Cells that were in the brightest 10% of BFP expressing cells were sorted into 96 well plates, with one cell per well. Single cell clones were expanded and screened for knock-outs.

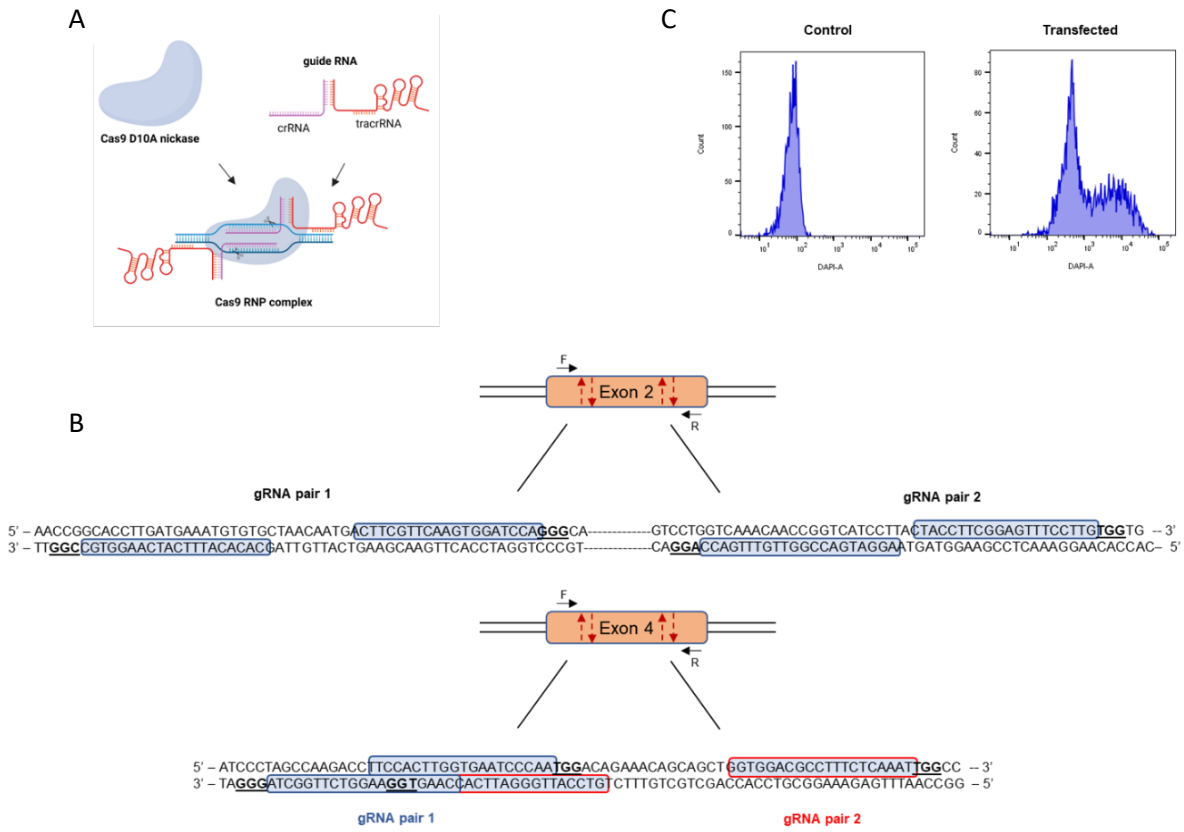


Figure 3-2 CRISPR-Cas9 editing strategy in GFPNickS120 cells. ***(A)*** *Schematic demonstrating sgRNA binding and orientation of paired guides, which should generate a staggered DSB.* ***(B)*** *Paired gRNAs used in this CRISPR and relative position and sequence in exon 2 and 4.* ***(C)*** *Control versus cells co-transfected with a BFP plasmid and sorted via FACS.*

3.2.2.2 Screening of targeted clones

Genomic DNA was harvested from individual clones and a two-pronged PCR-based screening approach was carried out. A PCR amplifying the CAG repeat region of interest was performed to identify clones with a comparable repeat size of ~120 CAG units or more (Figure 3-3). This led to the elimination of clones #1, #3 and #5. In addition, clone #6 failed to grow to confluence during this time and was removed from subsequent analysis.

Additionally, PCR amplicons spanning the targeted regions of exon 2 and 4 of *FAN1* were performed. PCR products were resolved via gel electrophoresis. Of the 7 clones screened, there were visible changes in the molecular weight of the PCR products for some clones (#4 and #6) compared with the parental control sample (Figure 3-3). However, we also expected small changes leading to frameshifts to occur that would not be detectable on an agarose gel. Thus, DNA from all candidate clones was subsequently sent for Sanger sequencing.

Sanger sequencing analysis indicated that clones #2 and #7 were potential *FAN1* knock-outs. Both of these had been transfected with the same gRNAs targeting exon 2 and exon 4 and sequencing was performed across these regions. Both clones #2 and #7 appear to have more than two copies of *FAN1,* with 3 and 4 copies, respectively, indicating they were polyploidy for *FAN1* (Figure 3-3). Analysis of targeted sites within these clones revealed 100% editing efficiency for gRNAs targeting exon 2, where all edits lead to the induction of a premature stop codon. Comparatively, the gRNAs targeting exon 4 had an 85.7% editing efficiency, with 1 allele of clone #7 left unedited. However, due to edits in exon 2, we predicted this allele to code for non-functional protein. Additionally, subsequent sequencing analysis of the *FAN1*$^{+/+}$ line and clones #2 and #7 for CAG repeat length indicated these lines have 125, 140 and 126 CAG repeats, respectively.

.

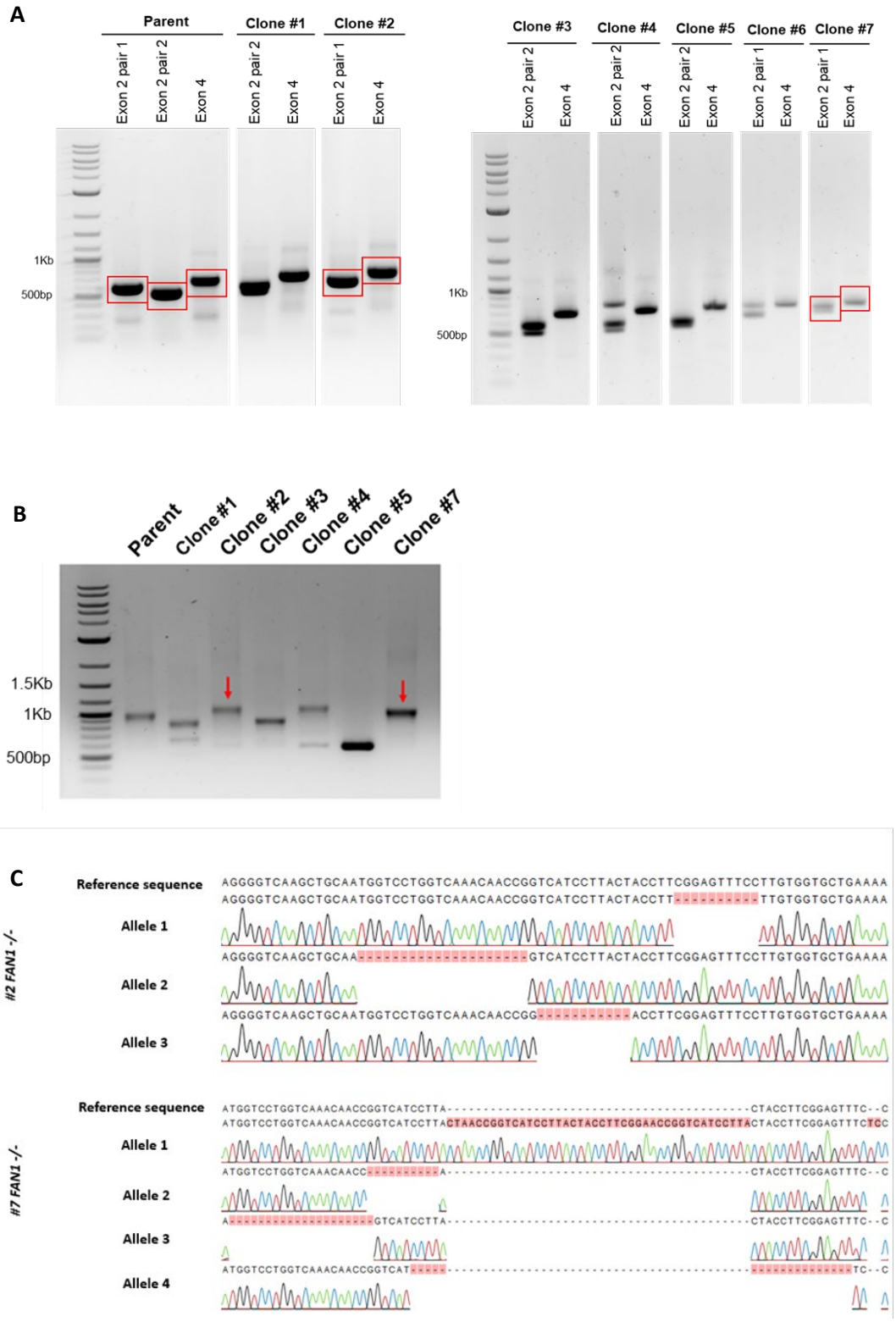


Figure 3-3: Screening of targeted clones. ***(A)*** *PCR spanning sgRNA cut sites.* ***(B)*** *PCR comparing CAG repeat lengths in all clones.* ***(C)*** *Sanger sequencing of clone #2 and clone #7 at exon 2 of FAN1* highlighting deletions generation by CRISPR-Cas9 editing.

3.2.2.3 Confirmation of FAN1 knock-out status and Cas9 expression

To confirm the *FAN1* knockout status genotype predicted by sanger sequencing, RT-qPCR and western blot analysis were performed. There was a significant decrease in *FAN1* mRNA levels in both clone #2 and #7 (*P=<0.0001* and *P=0.0002*), relative to the parent *FAN1*$^{+/+}$ cell line. A loss of functional FAN1 protein was confirmed via western blotting. A final screen was performed to determine whether both *FAN1*$^{-/-}$ clones express the Cas9 nickase. Western blot analysis confirmed this, with comparable expression observed between all lines (Figure 3-4). Taken together this data confirmed the generation of two *FAN1* knockout clones .

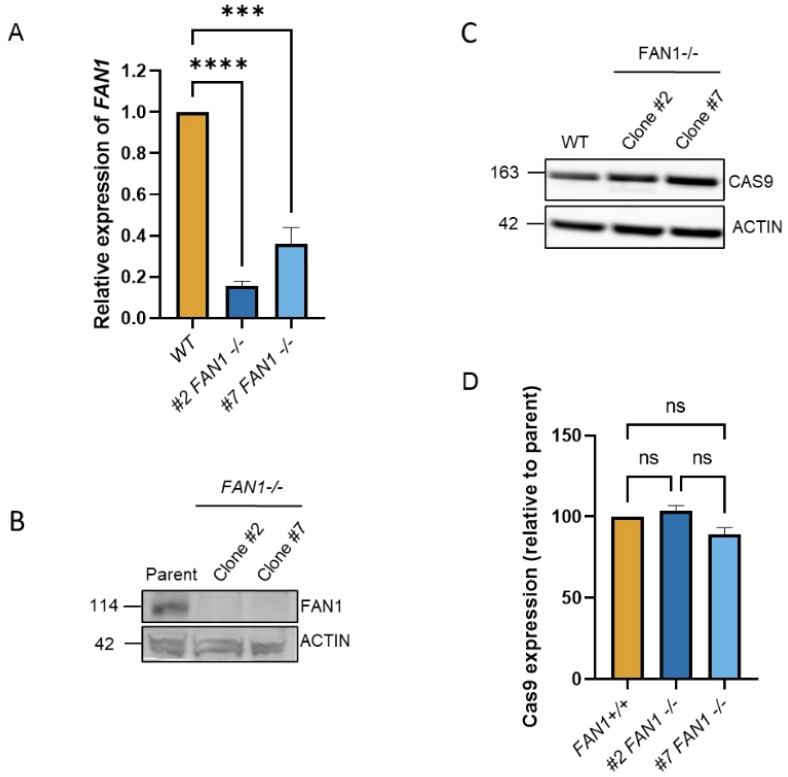


Figure 3-4: Validation of FAN1 knock-out status. *(A) RT-qPCR against FAN1 demonstrates significant reduction in mRNA expression in both FAN1*$^{-/-}$ *lines as determined by a One-Way-ANOVA followed by Tukey's multiple comparison testing. ***=P<0.001, **** =P<0.0001. Error represents ± SD (n=3).* ***(B)*** *Loss of functional FAN1 protein is confirmed via absence of detectable band on western blot. Molecular weight of FAN1 (114) and ACTIN (42)* in KDa. ***(C)*** *Expression of the Cas9 D10A nickase was confirmed by western blot. Molecular weight of CAS9 (163) and ACTIN (42) in KDa.* ***(D)*** *There was no significant difference in Cas9 expression levels between the FAN1 knock-out clones and parent line (n=3). ns = P>0.05, determined by One-Way-ANOVA followed by Tukey's multiple comparison testing Error represents ± SD.*

3.2.2.4 FAN1$^{-/-}$ clones are vulnerable to Mitomycin C

To determine whether *FAN1* knockout lines behaved as expected in the HEK293 background, I performed survival assays to a DNA damaging agent. FAN1 plays an important role in the repair of ICLs and a loss of FAN1 sensitises cells to multiple crosslinking agents (Kratz et al. 2010; MacKay et al. 2010; Smogorzewska et al. 2010). Mitomycin C (MMC) is a DNA cross-linking agent that results in the formation of covalent interstrand cross-links (ICLs) (Verweij 1990). Therefore, cell sensitivity to MMC can act as an indirect assay for FAN1 function.

A clonogenic assay was used to determine the effects of MMC on *FAN1*$^{-/-}$ and *FAN1*$^{+/+}$ clones. Cells were plated at a low density of 500 cells per 10cm^2 plate and left overnight to adhere. Cells were subsequently treated with increasing concentrations of MMC (5-50 ng/mL) for 24 hr, after which cells were allowed to recover and left to form colonies for a 10-day period. Following this, cells were fixed and stained with a crystal violet solution and colonies exceeding 50 cells were counted. There was an observed decrease in colony forming units as the concentration of MMC increased in both *FAN1*$^{-/-}$ clones compared with *FAN1*$^{+/+}$ cells, as expected (Figure 3-5).

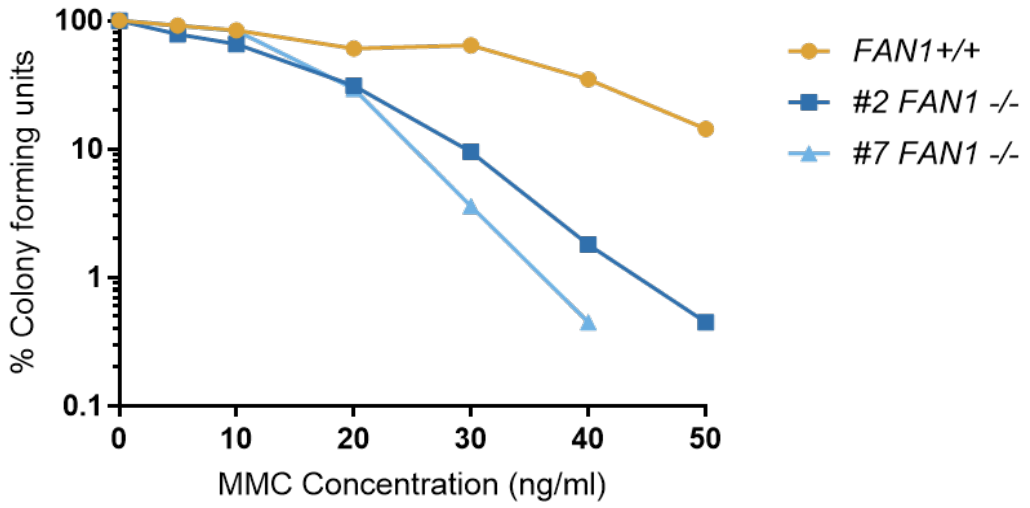


Figure 3-5: Mitomycin C assay to determine cell sensitivity to ICLs. *$FAN1^{-/-}$ cells demonstrate increased sensitivity to MMC compared with $FAN1^{+/+}$ over concentrations exceeding 20 ng/mL (n=1).*

3.2.2.5 FAN1 knock-out clones do not show proliferation deficits

FAN1 is involved in the repair of interstrand crosslinks (ICLs) which, if not resolved can stall the transcription and replication machineries (Thongthip et al. 2016). Independently of its role in ICL repair, FAN1 is also implicated in the recovery of stalled replication forks (Chaudhury et al. 2014; Lachaud et al. 2016). Although there has been no report suggesting a loss of FAN1 will impact the proliferation rates of HEK293, or other immortalised cell lines, it is plausible, given its importance for replication, that a loss may impact growth rates. To determine whether this was the case, a proliferation assay was carried out to assess if there were any differences between wild-type and knock-out cell lines.

$FAN1^{-/-}$ and *$FAN1^{+/+}$* cells were cultured for 12 days, and proliferation rates monitored. Cells were seeded at a starting density of $2x10^6$ cells in a 10 cm^2 plate. Every 3 days cells were recovered using trypsin and resuspended in pre-warmed non-selective DMEM + GlutaMAX (10% FBS) media and a

sample was taken to measure cell count. The cells were subsequently re-plated at the starting cell density. Over the time course a cumulative cell count was recorded. No significant difference was observed in growth rates between both *FAN1*$^{-/-}$ clones and the parent *FAN1*$^{+/+}$ line (Figure 3-6).

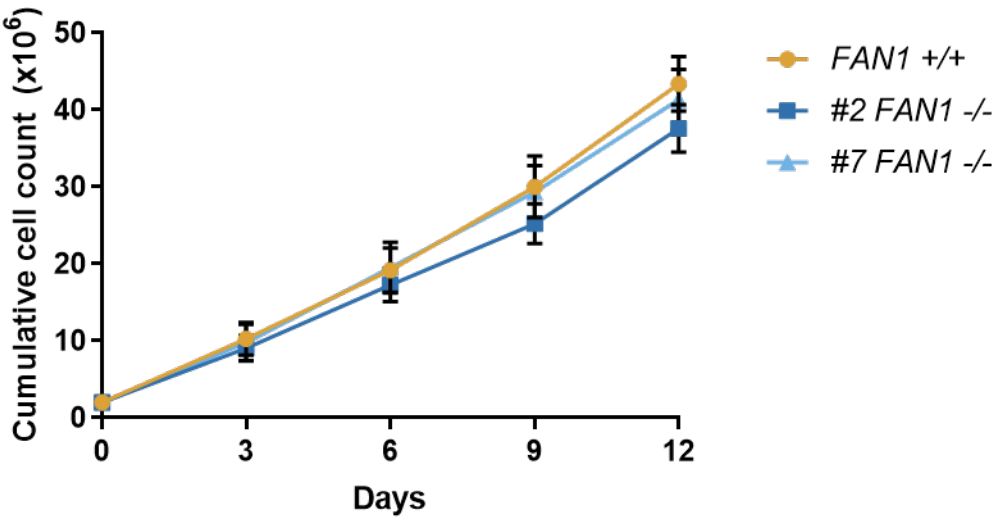


Figure 3-6: Monitoring proliferation rates in FAN1 knock-out lines and the parent line. *Cells were seeded at a density of 2x10^6 in 10 cm^2 dishes and passaged every 3 days with a count being taken before re-seeding. No difference in growth rates was determined by a Two-Way ANOVA followed by a Tukey's multiple comparison test. Error represents ± SD (n=4).*

3.2.3 Investigating nickase-induced contraction rates in *FAN1*$^{+/+}$ and *FAN1*$^{-/-}$ GFPNickS120 lines

3.2.3.1 FAN1$^{-/-}$ *clones demonstrate increased GFP expression over time*

To determine whether a loss of FAN1 impacts nickase-induced contractions, cells were grown in culture for a period of 21-days. During this time cells were treated with doxycycline to induce transcription through the CAG repeat tract. To determine how the rate of contractions changes over time, cells were treated in parallel with doxycycline for 7, 14 or 21-days. At the end of the time-course cells were collected for analysis via flow cytometry. In this system, changes in GFP intensity serves as a proxy for changes in repeat length, with an increase in GFP indicative of contraction events (Santillan et al. 2014; Cinesi et al. 2016).

Histograms were generated by plotting GFP intensity on the x-axis and the modal GFP data on the y-axis. The histograms demonstrate a clear difference between *FAN1*$^{+/+}$ and *FAN1*$^{-/-}$ clones throughout the 21-day time course, where both *FAN1*$^{-/-}$ clones appear to have increased rates of contractions over time compared with the *FAN1*$^{+/+}$ cell line. Data from at least five experimental replicates was collected for each line (Figure 3-7).

Two modes of analysis were used when comparing the effects of FAN1 loss on GFP expression. Firstly, I compared the difference between mean GFP expression at 7-days and 21-day post-treatment. The comparison between these time points was selected as previous data has demonstrated that at least 3 days of doxycycline expression in these lines is required for contraction events detectable via flow cytometry (Cinesi et al. 2016). This revealed a significant increase in GFP expression, and therefore contractions, in both *FAN1*$^{-/-}$ clones compared with the *FAN1*$^{+/+}$ cell line (*P=0.0032 and P=0.0245*). Secondly, the top 10% of cells expressing GFP were selected. These represent the top 10% of cells to have undergone larger contraction events from the modal CAG repeat. I then compared the change in this cell population from 7 to 21-days. In this analysis only *FAN1*$^{-/-}$ clone #2 demonstrated a significant increase in GFP expression, compared with the parent line (*P=0.0263*), but it worth noting in this analysis there was a greater degree of

variation in the biological replicates, particularly in the parent line and *FAN1* knock-out clone #7 (Figure 3-7).

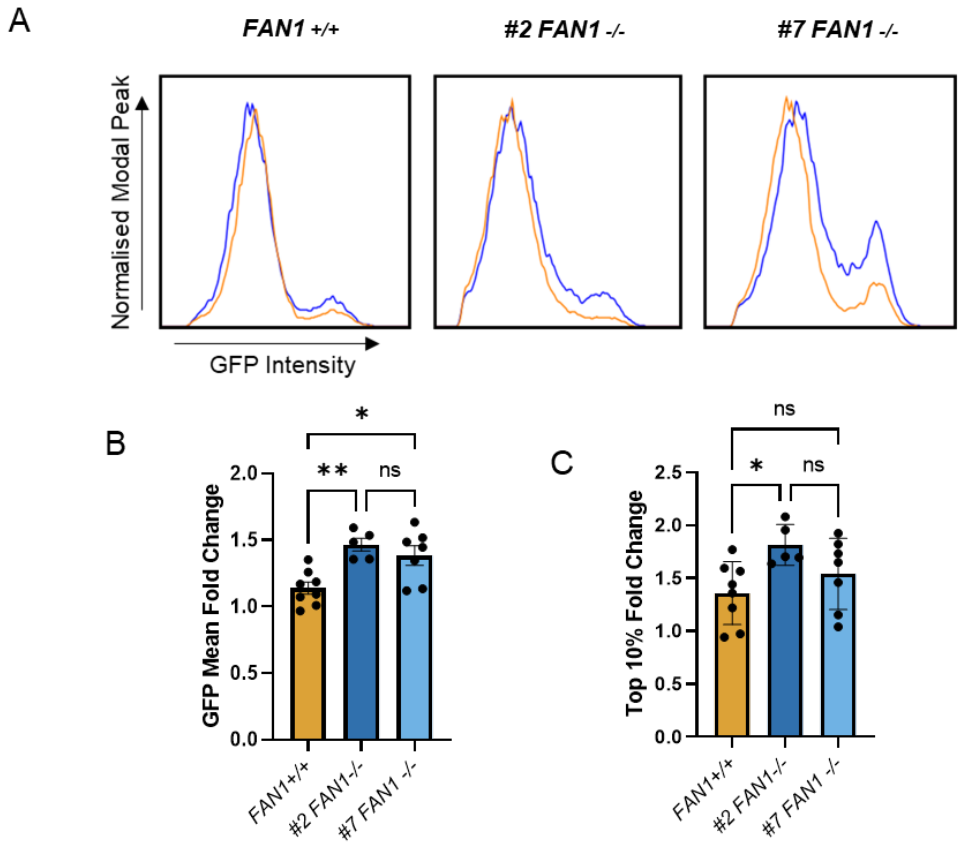


Figure 3-7: Monitoring changes in GFP expression in FAN1$^{+/+}$ and FAN1$^{-/-}$ lines. ***(A)*** *Demonstrative histogram profiles for all cell lines, where the x-axis indicated GFP expression and the y-axis GFP intensity normalised to the modal peak. Orange histograms are cells after 7-days of doxycycline treatment, and blue are after 21-days.* ***(B)*** *Bar chart comparing the change in GFP mean expression between 7 and 21-days post-doxycycline treatment.* ***(C)*** *Bar chart comparing the change in the top 10% GFP-expressing cells, representing the most contracted cells, at 7 and 21-days post-doxycycline treatment. *=P<0.05, **=P<0.01 as determined by a One-Way ANOVA followed by Tukey's multiple comparisons testing. Error represents ± SD (n=5-7).*

3.2.3.2 *FAN1$^{-/-}$ clones demonstrate a significant change in repeat length over time compared with FAN1$^{+/+}$ cells as detected by small-pool PCR*

Small-pool PCR (spPCR) was once considered the gold-standard method for analysis of trinucleotide repeat (TNRs) lengths. This technique involves the sequential dilution of an initial DNA sample to only a few genomic equivalents, preventing the bias against longer alleles that occurs using bulk-PCR methods and therefore enabling the detection of a full spectrum of allelic lengths (Gomes-pereira et al. 2004).

DNA samples were collected from both *FAN1$^{+/+}$* and *FAN1$^{-/-}$* lines at day 0 and day 21, post-doxycycline treatment. Serial dilutions of DNA were first carried out to obtain an optimal concentration of amplifiable alleles. Concentrations ranging from 7-15 genomic equivalents were used for subsequent PCR amplification and analysis. For each sample, blinded allele counts were performed and between 100-200 alleles were counted from at least 2 membranes per sample (Figure 3-8).

To investigate how the distribution of the repeat tract changed over the 21-day time course, alleles were binned into categories based on molecular weight markers using a 1 kb DNA ladder. Comparisons between day 0 and day 21 were performed independently for each cell line. A Mann-Whitney U test comparing distributions revealed no significant difference between the day 0 and day 21 samples in the *FAN1$^{+/+}$* line, but a significant difference between the two timepoints was observed in both *FAN1$^{-/-}$* clones (Figure 3-9). These results suggest that in both *FAN1* knockout lines there is more variation in CAG repeat length after 21 days, compared with the parent *FAN1$^{+/+}$* line.

A second analysis was carried out to quantify the percentage of contraction events after 21-days, relative to day 0. In this analysis the most common allele representing the modal CAG length was selected and alleles below this region were counted as contractions. Importantly, the distributions of allele frequencies at day 0 were comparable between all lines ($P > 0.05$), allowing for a fair comparison after 21-days of treatment. There was a significant

difference between the two lines at the day 21 timepoint, with both *FAN1*$^{-/-}$ lines demonstrating a substantial increase in the number of contractions compared to *FAN1*$^{+/+}$ cells (*P=<0.0001*) (Figure 3-9) . Taken together, these data suggest that a loss of FAN1 significantly increases contraction frequency, relative to *FAN1*$^{+/+}$ lines.

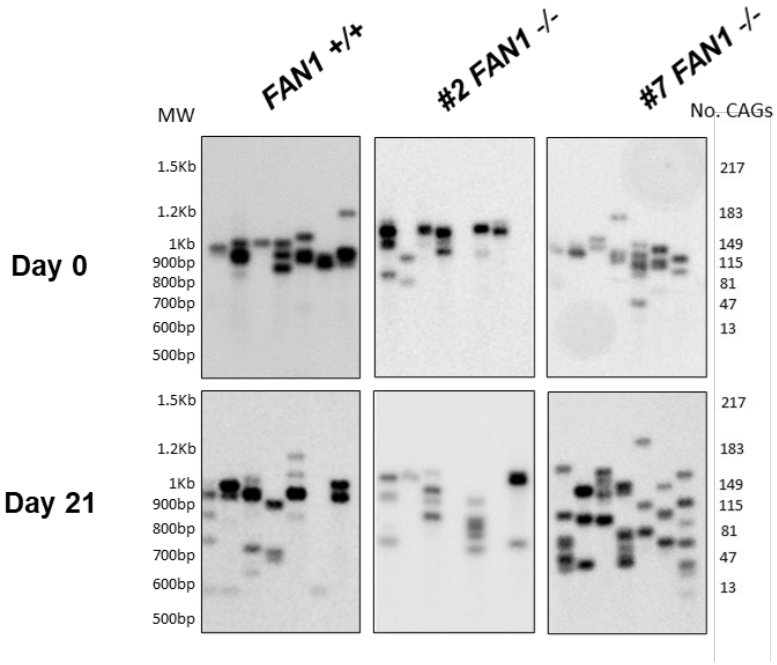
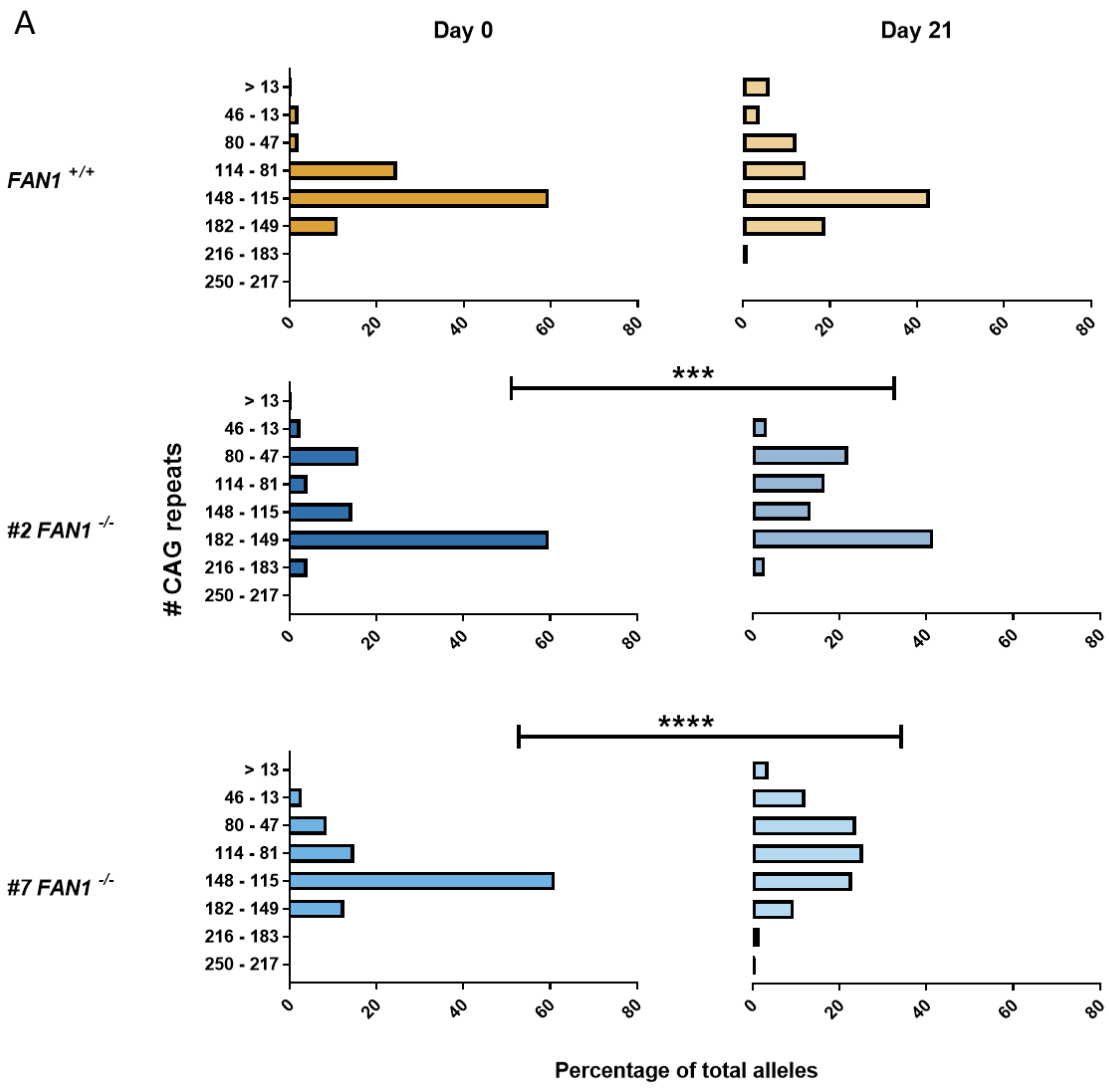


***Figure 3-8: Demonstrative spPCR blots**. Example spPCR blots for FAN1*$^{+/+}$ *and FAN1*$^{-/-}$ *lines at day 0 and day 21.*


A
Day 0
Day 21
FAN1 +/+
#2 FAN1 -/-
#7 FAN1 -/-
CAG repeats
> 13
46 - 13
80 - 47
114 - 81
148 - 115
182 - 149
216 - 183
250 - 217
0
20
40
60
80
***
****
Percentage of total alleles

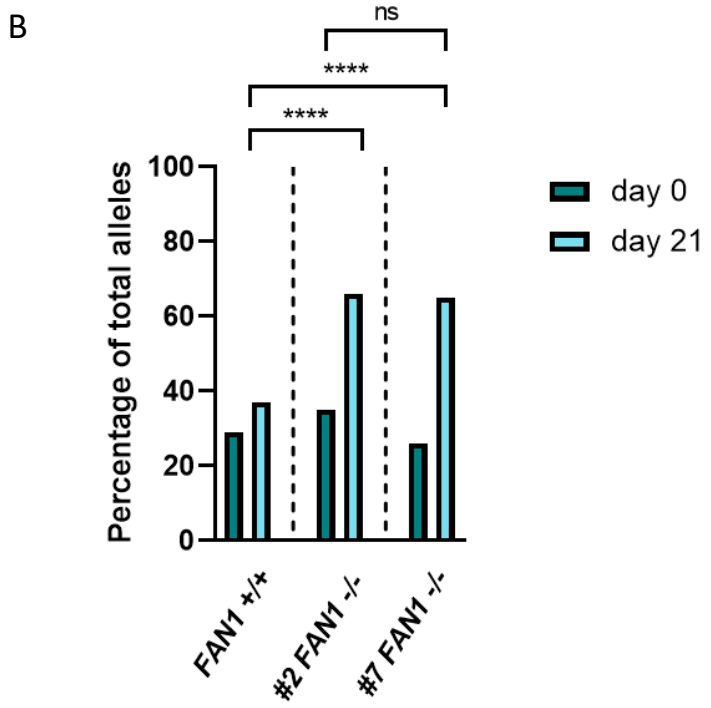

B
ns
****
****
Percentage of total alleles
100
80
60
40
20
0
day 0
day 21
FAN1 +/+
#2 FAN1 -/-
#7 FAN1 -/-

Figure 3-9: SpPCR analysis of FAN1$^{+/+}$ and FAN1$^{-/-}$ lines to monitor Cas9 nickase-induced contractions. ***(A)*** *Graphic representation of spPCR analysis monitoring distribution of CAG repeat size. ***P=<0.001, ****P=<0.0001 as determined by Mann-Whitney U comparing day 0 and day 21 distributions.* ***(B)*** *Both FAN1$^{-/-}$ clones demonstrate a significant increase in Cas9 nickase-induced contractions after 21-days relative to the FAN1$^{+/+}$ parent line. ****P=<0.0001 as determined by two-sided Fisher's exact testing.*

3.2.3.3 FAN1$^{-/-}$ lines reveal increased modal contraction over 21-days as detected via amplicon PacBio sequencing

The advent of high-throughput sequencing, as described in 1.10.2, has allowed for increased accuracy of CAG repeat sizing. To date, targeted sequencing of CAG repeats has been achieved with Illumina MiSeq, Oxford Nanopore technology and long-read PacBio sequencing. Oxford Nanopore and PacBio sequencing technology are best suited for sequencing of expanded repeats over ~80 repeats due to MiSeq's restricted read-length. In this thesis long-read PacBio SMRT sequencing was used to compare changes in CAG repeat lengths, as this has reduced error rates compared with the current Nanopore technology (Taylor et al 2022). Our lab has developed a tool that allows for accurate sizing of CAG repeat tracts from amplicon long-read PacBio sequencing data. Repeat Detector detects and counts tandem repeats in targeted sequencing data (described in 1.10.2) (Taylor et al. 2022).

DNA samples were collected from both *FAN1$^{+/+}$* and *FAN1$^{-/-}$* lines at day 0 and day 21 post-doxycycline treatment. Repeat detector generated histogram reads for all samples and for all lines I compared average CAG repeat size from day 0 to 21-days post treatment. For all samples five biological replicates were analysed. In the wild-type line we saw an average reduction of 7 CAG repeats over the time-course. Comparatively, both knock-out lines demonstrated a significant increase in CAG contraction over time, relative to the wild-type (*P=0.049 and P=0.0001*). In *FAN1$^{-/-}$* clone #2 there was an average contraction of 24 repeats compared with 41 repeats for clone #7 over 21-days (Figure 3-10). This difference between the knock-out clones may be attributed to the difference in starting CAG repeat length in these lines, where

clone #2 has an average starting length of 140 CAG's, compared with 126 in clone #7.

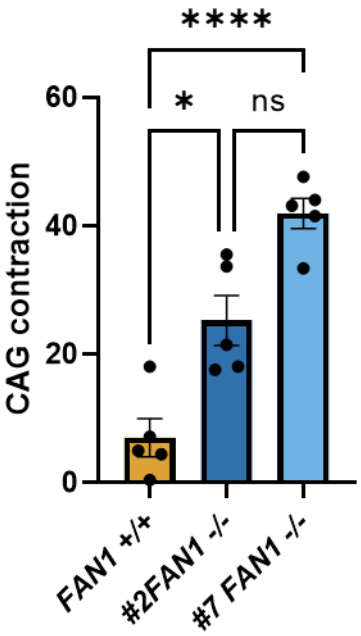


Figure 3-10: CAG contractions in FAN1+/+ and FAN1-/- lines as detected by long-read PacBio sequencing. *Both FAN1$^{-/-}$ lines demonstrate a significant increase in CAG contractions relative to the parent FAN1$^{+/+}$ line. *=P<0.05, ****P<0.0001 as calculated by a One-way ANOVA followed by Tukey's multiple comparison testing. Error represents ± SD, (n=5).*

3.2.4 Confirming somatic instability in HEK293 cells in the absence of the CRISPR-Cas9 nickase (GFP(CAG101) cells)

3.2.4.1 Generation of a FAN1 knock-out via CRISPR-Cas9 genome editing

Canonically a loss FAN1 results in increased somatic instability, with a clear bias towards expansion events in mammalian systems. This has been replicated in both immortalised cell lines with exogenous *HTT* CAG repeats and HD-patient derived iPSCs and neurons. In these cells, repeat lengths above 71 were sufficient to cause an increase in expansions over prolonged culture times, relative to wild-type lines (Goold et al. 2019; Kim et al. 2020; Goold et al. 2021; McAllister et al. 2022). Given that loss of FAN1 in GFPNickS120 cells increase Cas9-nickase induced contractions and analysis by sensitive long-read PacBio sequencing did not indicate increases in expansion events, I set out to confirm that in the absence of Cas9 *FAN1*$^{-/-}$ HEK293 cells demonstrate somatic expansion, in line with previous literature. This would demonstrate this contraction phenotype observed is specific to the action of the Cas9 nickase, and not a phenotype associated with HEK293 cells.

To determine whether a loss of FAN1 induces somatic instability in our GFP(CAG101) reporter cells, in the absence of the nickase, a CRISPR-Cas9 genome editing approach was used. In these lines as there is no constitutively expressed Cas9 nickase, a knock-out was generated using a well-established Cas9 RNP approach (Bloomer et al. 2022). In this case two gRNAs targeting exon 2 of *FAN1* were used to generate a 95 bp deletion in this region (McAllister et al. 2022). Deletions of this length have been shown to yield a high number of edited clones. If successful, this deletion results in a frameshift mutation, the introduction of a premature stop codon and therefore loss of FAN1 protein expression.

Briefly, *FAN1*-specific targeting cRNAs were synthesized and duplexed with a tracrRNA scaffold containing an ATTO 550 fluorescent tag. The functional sgRNA was then incubated with an Alt-R *S.p.* Cas9 Nuclease V3 to form a stable RNP complex. Cells were electroporated with this fluorescently labelled complex and after 24 hr cells were sorted based on fluorescence intensity, with

only the top 10% of cells selected. Single cell clones were then screened for the presence of the correct deletion product (Figure 3-11).

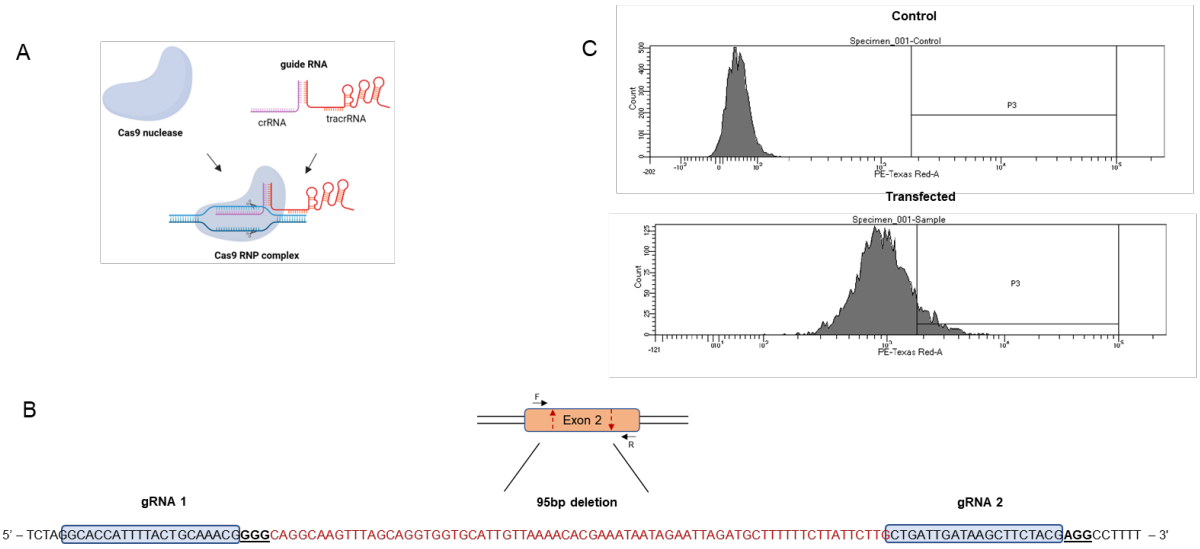


Figure 3-11: FAN1 targeting approach in GFP(CAG101) cells. (A) *CRISPR-Cas9 facilitates gene editing as a Cas9 nuclease is targeted to the desired region in the genome by a gRNA. Together the nuclease and gRNA form a complex leading to the formation of a DSB at the targeted editing site.* ***(B)*** *Two gRNAs were designed targeting exon 2 of FAN1 to create a 94p deletion.* ***(C)*** *The top 10% of fluorescently labelled cells (P3) were sorted and plated for characterization and expansion.*

3.2.4.2 Screening of targeted clones

After electroporation with Cas9 RNP complexes single-cell colonies were expanded. Genomic DNA was then extracted and clones screened via PCR amplification for the presence of the targeted 95 bp deletion in exon 2 of *FAN1*. PCR products were visualised via gel electrophoresis (Figure 3-12). Candidate knock-out clones were products with a molecular weight at 231 bp, compared with the predicted molecular weight of the wild-type *FAN1* (325 bp). These clones were then taken forward and selected for sanger sequencing to confirm the presence of deletion. For both clones sequenced there was the presence of a 94bp deletion, which whilst not the predicted 95bp deletion still results in a premature stop codon (Figure 3-12). Based on this initial PCR data there appeared to be two knock-out clones, however, upon further expansion one clone was found to be a mixed population. As such only one *FAN1* knock-out

clone was taken forward. Unfortunately, due to issues with the FAN1 antibody, western blot analysis of FAN1 protein expression could not be performed. However, based on sanger sequencing data demonstrating a frame-shift mutation in the DNA and an introduction of a premature stop codon, I was confident this clone is a *FAN1* knock-out.

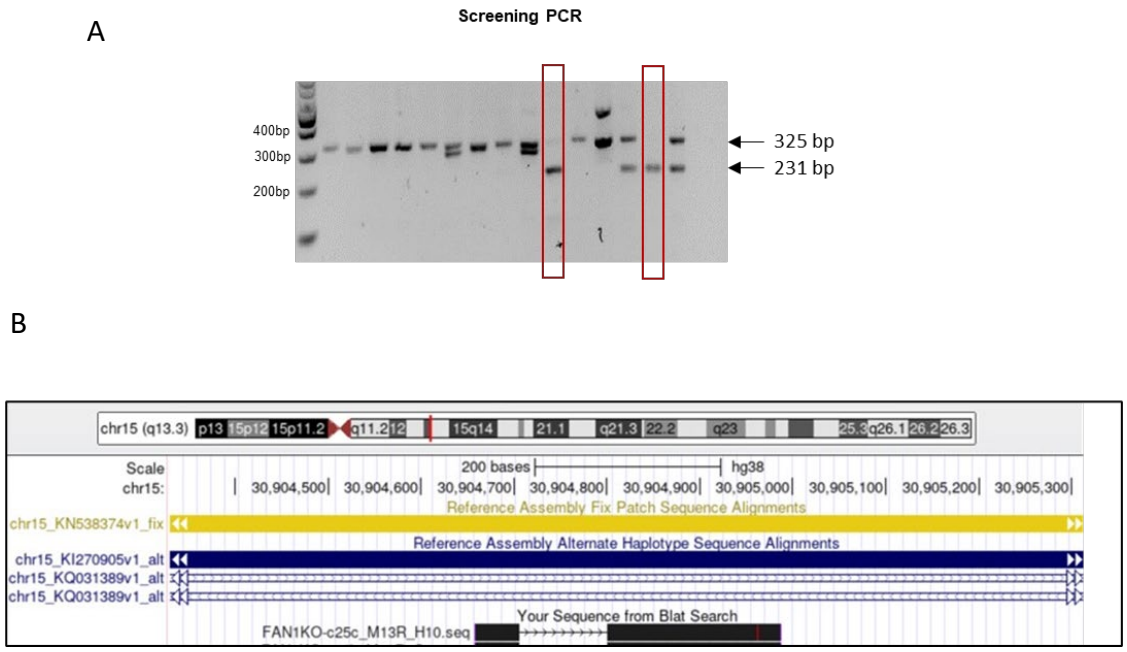


Figure 3-12: Validation of FAN1 knock-out in the GFP(CAG101) line. ***(A)*** *Screening PCR to confirm deletion product.* ***(B)*** *Sanger sequencing targeted 94 bp deletion in exon 2 of FAN1 (in black).*

3.2.4.3 A *FAN1*$^{-/-}$ line shows increased somatic expansion in culture compared with the *FAN1*$^{+/+}$ parent line

Repeat Detector was used to monitor somatic instability in *FAN1*$^{+/+}$ and *FAN1*$^{-/-}$ cells (Taylor et al. 2022). For each cell line, cell pellets were collected at day 0, day 21 and day 42 post-doxycycline treatment. DNA was extracted and amplicon PacBio SMRTbell sequencing libraries prepared. After demultiplexing, samples were run through Repeat Detector software and histograms were produced. For our analysis, the modal CAG repeat for each line at day 0 was calculated and compared with day 21 and day 42. For each sample at least three biological replicates were sequenced. Modal CAG values were plotted and a linear regression performed for each genotype. In the *FAN1*$^{-/-}$ line there was a significant increase (*P=<0.0001*) in CAG repeat number over 42 days, with an average increase of 6.7 repeats. Surprisingly, in *FAN1*$^{+/+}$ cells there was no significant change in repeat length over the time-course.

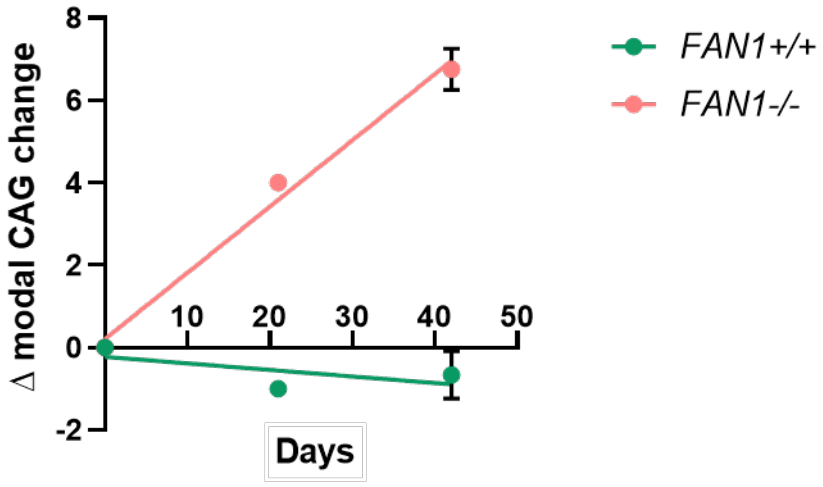


Figure 3-13: PacBio long-read sequencing to monitor somatic expansion in GFP(CAG101) cells. *The FAN1*$^{-/-}$ *demonstrates a significant increase in CAG repeat expansion over the 42-days (P=<0.0001), whilst the FAN1*$^{+/+}$ *line does not (P=0.1269) As determined by a linear regression calculating deviation from zero (n=4).*

3.3 Discussion

3.3.1 Summary of findings

This chapter describes the use of CRISPR-Cas9 gene editing to generate *FAN1* knock-outs in a GFPNickS120 reporter cell system, containing a stably expressed Cas9 D10A nickase and sgCTG targeting the CAG/CTG repeat tract. The characterisation of two *FAN1*-knockout clones in cells expressing the CRISPR-Cas9 nickase system allowed us to determine whether FAN1 impacts the efficiency of nickase-induced contractions, the mechanism of which is yet to be fully understood. Furthermore, to confirm these results were due to the action of the nickase, and not as a result of instability of CAG/CTG repeats in HEK293 cells a second line was utilised. These cells, GFP(CAG101), do not contain the Cas9-nickase treatment and allowed for the monitoring of somatic changes across the CAG/CTG repeat in $FAN1^{+/+}$ and $FAN1^{-/-}$ lines. This allowed me to determine whether loss of FAN1 causes somatic CAG expansion over culture time, which has been previously reported in other cellular models (Goold et al. 2019; Goold et al. 2021; McAllister et al. 2022).

3.3.2 FAN1 protects against Cas9 nickase-induced contractions in GFPNickS120 cells

To investigate whether FAN1 impacts nickase-induced contractions I generated two $FAN1^{-/-}$ lines through CRISPR-Cas9 editing. This gene-editing approach relied on the constitutive expression of the D10A Cas9 nickase in our cells and transfection of cells with paired sgRNAs targeting exon 2 and exon 4, respectively. The use of paired sgRNAs has been shown to result in a nick on both strands, generating a DSB with 5′ overhangs (Ran et al. 2013). Indeed, studies suggest that gene editing using a Cas9 nickase reduces potential off target effects, as the likelihood of another location in the genome containing the same two target sequences is reduced (Cho et al. 2014; Shen et al. 2014). When designing paired sgRNAs an optimal gap of 0-20

nucleotides between each sgRNA is considered best for gene editing, this was taken into account when designing our CRISPR-editing strategy (Ran et al. 2013).

Due to the requirement for paired sgRNAs this reduced the potential number of sgRNAs that could be used. As the majority of edits that occur via CRISPR-editing are a result of NHEJ which creates small insertion/deletion events, I was unable to resolve small changes in sequence length via gel electrophoresis, so all clones were sequenced. Sanger sequencing analysis confirmed clones #2 and #7 as likely *FAN1* knock-outs, based on frameshift mutations within both targeted exons. Interestingly, for both clones there was a range of edits across both exons with deletion events from 10-20 bp in length and larger insertion events (35 and 42 bp). Furthermore, for clones that were not predicted to be a knock-out, based on sanger sequencing, all demonstrated editing of one allele when targeted with either exon 2 or exon 4 sgRNAs (data not shown). This indicates that this method is associated with a high level of gene editing. The high level of editing observed using this approach may be attributed to the method, which involved FACs sorting of the top 10% of cells expressing a BFP plasmid that was co-transfected with sgRNAs. Although this is not a direct measure for uptake of a sgRNA complex this still serves as a tool for reducing the initial cell population based on those that have undergone successful transfection. Additionally, I observed a high degree of cell death following FACS. This reduced the number of clones available for screening, with only 7 screened. Therefore, it is possible repeating this CRISPR with more clones screened may reduce calculated editing efficiency.

Sanger sequencing revealed that both *FAN1*$^{-/-}$ clones had more than two alleles, indicating that both clones are polyploid for chromosome 15. This polyploid phenotype was also observed in sequencing data obtained from other clones (data not shown) and has also been observed at other chromosomal locations (Dion lab, unpublished). This is not a surprising finding given that HEK293 cells are immortalised. HEK293 cells have been described as having a hypotriploid karyotype with a modal chromosome number of 64 occurring in 30% of cells (American Type Culture Collection). However, one

study utilising HEK293 cells sourced from the American Type Culture Collection demonstrated a hypertriploid phenotype and significant changes in chromosome diversity and copy number than previously described (Lin et al. 2014). Furthermore, it is clear that chromosomal number and variants are distinct in HEK293 cells depending upon the source, with studies reporting chromosomal numbers ranging from anywhere between 56-78 chromosomes per cell (Stepanenko and Dmitrenko 2015).

Despite the presence of multiple copies of *FAN1,* sequencing data still indicated a *FAN1*$^{-/-}$ genotype. RT-qPCR investigating *FAN1* mRNA levels demonstrated a reduction in mRNA, with *FAN1*$^{-/-}$ clones #2 and #7 demonstrating an average of roughly 15% and 35% expression compared to *FAN1*$^{+/+}$, respectively. The presence of remaining mRNA could be explained by the presence of transcripts that have escaped nonsense-mediated decay, or may be indicative of the steady-state between transcription and degradation. Despite the presence of a low level of mRNA, the absence of protein via western blot confirms both clones are *FAN1* knock-outs . The FAN1 antibody used to confirm the presence of FAN1 is a polyclonal antibody raised in sheep against full-length GST-FAN1 (MacKay et al. 2010). Although there is a clear band present at 114 kDa, the molecular weight of FAN1, there are other contaminating bands present on the blot for all samples, which were not removed by an extended blocking time. Additionally, this antibody was not reliable between experiments, where endogenous FAN1 levels could often not be detected (as highlighted in 3.2.4.2). This highlights the need for better commercial monoclonal antibodies. Importantly, *FAN1*$^{-/-}$ clones both demonstrated sensitivity to cross-linking agent MMC, consistent with previous reports (Kratz et al. 2010; MacKay et al. 2010). Aside from a role in ICL repair, FAN1 is important for the maintenance of genome integrity and prevention of replication fork collapse (Chaudhury et al. 2014; Porro et al. 2017). Given this, we might expect a loss of FAN1 to impact cell proliferation rates. However, this was not the case and both knock-out clones demonstrated comparable cell growth rates over a 12-day period. This is consistent with previous reports in a *FAN1*$^{-/-}$ DT40 cell line, where no effect was reported on doubling times (Yoshikiyo et al. 2010). These data taken together data conclusively

demonstrate a loss of functional FAN1 in two clones, which serve as a tool to monitor the effect of FAN1 loss on CRISPR-Cas9 nickase-induced contractions.

After validation of both *FAN1*$^{-/-}$ clones I sought to investigate the effect of FAN1 loss on CRISPR-Cas9 nickase induced contractions. Determining the size of repetitive DNA elements, such as TNR regions, has historically proven particularly challenging. Multiple methods have been developed in an attempt to quantify these regions, with the majority replying upon PCR amplification through the repeat. However, these methods can lead to an inherent PCR bias towards smaller allele lengths meaning rarer expansion events may be overlooked. Non-amplification methods are also used, such as southern blotting, but these are time consuming and require a higher input of DNA (Massey et al. 2019). The presence of the stably integrated GFP reporter assay containing an expanded CAG tract in our HEK293 cell system presents a sensitive and rapid tool that can assess both contraction and expansion events (Cinesi et al. 2016). As such, this was the first tool used to screen for whether a loss of FAN1 impacts nickase-induced contractions.

Two methods of analysis were carried out utilising flow cytometry data, where GFP intensity serves as proxy for CAG repeat length. When comparing the change in mean GFP intensity from day 7 to day 21, both *FAN1*$^{-/-}$ lines demonstrated a significant increase, compared with the wild-type line. When focusing on only the top 10% of GFP positive cells, those cells with the largest contraction events, we only observed a significant increase in clone #2. This may be attributed to the larger spread of data points in this analysis, where more replicates may yield more informative results. One way to reduce variation between biological replicates may be to run all samples through the flow cytometer on the same day, accounting for any effect of run variation. Overall, using GFP intensity as a proxy for changes in CAG repeat length allowed for a quick assay to assess whether a loss of FAN1 impacts nickase-induced contractions, which can be complimented by downstream techniques.

Importantly, analysis of spPCR data supports these findings, demonstrating increased contractions in *FAN1*$^{-/-}$ lines. I consistently observed more

contractions in the *FAN1*$^{-/-}$ clones than in the wild-type parent by spPCR and when comparing the spread of allelic distribution no increase in expansion events was observed. Together, these results suggest that FAN1 inhibits but does not completely prevent nickase-induced contractions, as loss of FAN1 greatly improves the frequency of contractions. Some expansion events were present in spPCR which can be visualised on demonstrative blot images. These rare expansion events were seen in both *FAN1*$^{+/+}$ and *FAN1*$^{-/-}$ lines. This is to be expected as spPCR analysis reveals a more detailed allelic spectrum with less bias towards longer CAG repeats, and will therefore reveal rarer expansion events missed by our GFP-reporter based assay, which is less sensitive (Gomes-pereira et al. 2004).

Whilst spPCR is a useful tool which allows a visual assessment of allelic changes in CAG/CTG repeat lengths, detection of smaller expansions or contraction events is limited. As such, contraction rates reported by spPCR are likely an underestimation as subtle changes cannot be identified. High-throughput sequencing is useful, therefore, as it allows us to collect accurate sequencing information regarding CAG repeat tract lengths. Using long-read PacBio sequencing, we were able to assess changes in the average CAG repeat tract length over 21-days. In both *FAN1*$^{-/-}$ lines there was a significant average increase in CAG contractions over the time-course, compared to the wild-type. These data supports the findings of the GFP and spPCR assays, which serves to validate the use of this sequencing platform in conjunction with Repeat Detector for monitoring changes in CAG repeat tract lengths. This is important as this long-read sequencing approach allows for a more sensitive and accurate measure of CAG repeat size compared with the GFP-reporter, where fluorescence serves as a proxy for CAG size, or small-pool PCR which is unable to distinguish between small changes in repeat length. Interestingly, clone #7 underwent larger CAG contraction events on average compared to clone #2. Whilst this was not significant, this may be attributed to differences between the starting length of both repeats, with clone #2 having a longer starting expanded repeat, which may influence the degree of contraction.

Given FAN1's well characterised role in protection against somatic instability (Zhao and Usdin 2018; Goold et al. 2019; Kim et al. 2020; Goold et al. 2021; McAllister et al. 2022), with a loss of FAN1 increasing expansion events, it was perhaps surprising that a loss of FAN1 in our system results in an increase in contraction events, not expansion events. Studies demonstrate that FAN1 binds to CAG/CTG loop out structures (Kim et al. 2020; Deshmukh et al. 2021b), this FAN1 binding and subsequent nucleolytic processing of these structures likely leads to faithful DNA repair, maintaining stability at these sights. Therefore, whilst the direction of instability in the absence of FAN1 in Cas9-treated nickase cells is biased to contraction events, this still indicates that FAN1 acts to reduce instability in both directions at an expanded repeat.

3.3.3 Loss of FAN1 increases somatic expansions in GFP(CAG101) cells

In order to confirm the increase in contractions observed in GFPNickS120 lines are due to the action of the Cas9 nickase, and not an inherent phenotype of HEK293 cells a *FAN1*$^{-/-}$ clone in GFP(CAG101) cells lacking the Cas9-nickase was utilised. To generate a *FAN1* knock-out line and assay for changes in somatic CAG expansion, a CRISPR editing approach was carried out using a previously optimised protocol (McAllister et al. 2022). This approach involves the use of 2 sgRNAs in close proximity to generate an 95 bp deletion. The benefit of this approach is that it allows for the screening of the targeted deletion via PCR, without the need for a T7 surveyor assay, which is more time-consuming and has been reported to underestimate editing efficiency (Sentmanat et al. 2018). Single cell colonies were isolated and screened for successful editing. Due to the previously reported high editing efficiency of this targeting approach, not all clones were screened, but simply frozen down for potential pools of stocks to be utilised if needed. Of the clones screened we report 15% of cells with the desired edit, but this is likely an overestimation due to a lower screening pool. The overall editing efficiency was higher at 36% of clones screened, when taking into account differences in PCR banding observed via gel electrophoresis and the presence of two distinct bands at either the wild-type or desired 95 bp deletion, indicating a likely clone heterozygous for *FAN1* knockout. This editing efficiency is similar

to previously published reports (Liang et al. 2017). However, this may be an underestimation as small insertions/deletions are not visible by electrophoresis, but for our purposes these lines were not characterised further.

Cell systems modelling CAG repeat expansion in culture have demonstrated that starting repeat lengths of 97 or more are sufficient to show expansion phenotypes over a 40 day culture period (Goold et al. 2019; Goold et al. 2021; McAllister et al. 2022). To date, a U2OS cell line stably transduced with human *HTT* containing 118 CAG repeats demonstrates an average increase of ~1.5 CAGs over 40 days in culture. Furthermore, studies report an increase in CAG expansions in HD-patient derived iPSCs of ~1.5-2.5 CAGs over the same time course. Interestingly, reports also demonstrate that differentiated medium spiny neurons tend to expand more slowly than their iPSC derived counterpart (Goold et al. 2019; McAllister et al. 2022). In GFP(CAG101) *FAN1*$^{+/+}$ cells, I did not observe any increase in CAG expansion over a time course of 42 days. However, I report a striking average increase of 6.7 CAGs over the time course in a *FAN1*$^{-/-}$ line.

The lack of CAG expansion in *FAN1*$^{+/+}$ cells may be attributed to higher basal FAN1 expression levels in our cell system, compared with other cell types reported in the literature. A transcription-wide association study (TWAS) has indicated that increased FAN1 expression is significantly associated with delayed age at onset of HD, likely due to its protective effect against somatic expansions (Goold et al. 2019). FAN1 expression levels vary depending upon tissue and cell type. For example, expression of FAN1 from skin fibroblasts, which are the source for HD-patient derived iPSC lines, is relatively low compared to expression levels from kidney cells (Human Protein Atlas). However, expression of FAN1 will also vary depending on patient sample variation. Therefore it would be useful to look at endogenous expression levels of FAN1, and indeed other factors associated with somatic instability, such as the MMR proteins, which promote somatic expansion. For example, as HEK293 cells demonstrate a varied karyotype, it is possible in our lines gene dosage of MMR factors is important. Comparisons between basal FAN1 and MMR levels, and indeed comparison between different cell models would likely

be helpful in understanding differences between basal somatic expansion in culture. Furthermore, it is possible that in our cell line an extended culture time of our *FAN1*$^{+/+}$ cells would result in an expansion phenotype. However, the striking increase in somatic expansions in the absence of FAN1 demonstrates that FAN1 in HEK293 cells acts to protect somatic instability in line with current literature (Goold et al. 2019; Kim et al. 2020; Loupe et al. 2020; Goold et al. 2021; McAllister et al. 2022) . These data demonstrate that these HEK293 GFP reporter cell lines are able to monitor somatic instability *in vitro* and that the observed increase in contraction events seen *FAN1*$^{-/-}$ GFPNickS120 cells can be attributed specifically to the action of the Cas9 nickase.

3.3.4 Chapter Summary

- Loss of FAN1 in GFPNickS120 cells results in an increase in Cas9 nickase-induced contractions, relative to *FAN1*$^{-/-}$ cells
- Loss of FAN1 in GFP(CAG101) cells increases somatic expansions relative to *FAN1*$^{+/+}$ cells

4 Investigating the role of FAN1 variants and MLH1 on CRISPR-Cas9 nickase induced contractions

4.1 Introduction

As *FAN1* deletion leads to an enhanced frequency of nickase-induced contractions (Chapter 3), I then sought to determine which functions of FAN1 are relevant in this case. Since its identification, FAN1 has been linked to a range of DNA repair processes (Cannavo et al. 2007). FAN1 is comprised of an N-terminal UBZ-type ubiquitin binding domain, a SAP DNA binding domain, a TPR protein-protein interaction domain and a C-terminal nuclease domain (Figure 4-1). The protein domain organisation of FAN1 is highly conserved across most eukaryotes and prokaryotes, highlighting its fundamental role in maintenance of genomic integrity (MacKay et al. 2010; Yoshikiyo et al. 2010). Interestingly, FAN1 is the only known eukaryotic protein to contain a VRR_nuc domain, which are more commonly present in prokaryotes (Pennell et al. 2014).

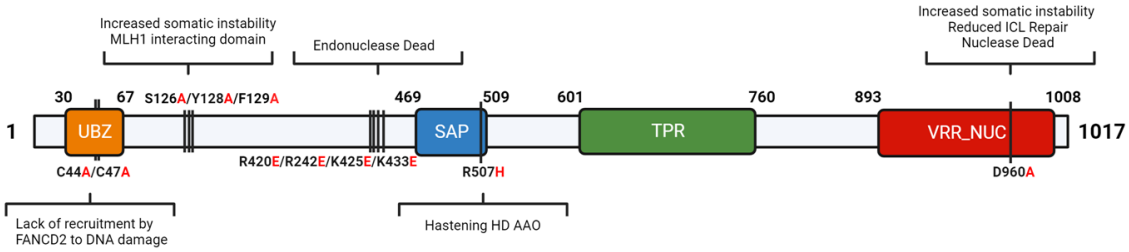


Figure 4-1: Schematic depicting the different FAN1 variants under investigation.

The first described role for FAN1 was its involvement in ICL repair, a function thought to be mediated by its UBZ domain. This domain facilitates an interaction with monoubiquitinated FANCD2, a member of the FA pathway (Kratz et al. 2010), where a C44A/C47A ablates co-localisation of FAN1 and FANCD2 to sites of DNA damage (Kratz et al. 2010; Liu et al. 2010b; Smogorzewska et al. 2010). The FA pathway, made up of 19 core proteins, is a specialized DNA repair responsible for the resolution of interstrand crosslinks

(ICLs) (Ceccaldi et al. 2016). These are highly toxic lesions resulting in covalent bond formation between opposite DNA strands, capable of stalling replicated and transcription machineries (Thongthip et al. 2016). Whilst FAN1's role in ICL repair was initially thought to be confined only to its association with the FA pathway, recent evidence has suggested that FAN1 can also repair ICLs independently of this pathway (Thongthip et al. 2016; Goold et al. 2021; Lachaud et al. 2016). How FAN1 functions to repair ICLs is not fully known, though all models rely on the nuclease activity of FAN1, which may act to; directly unhook ICLs, resect unhooked nucleotides or act via downstream repair on D-loop structures (Jin and Cho 2017).

FAN can act as both an endo- and exonuclease. The endonuclease activity of FAN1 is structure specific with a preference for 5′ flap structures, but is capable of cleaving other structures to a lesser extent, including looped DNA intermediates, which form in repetitive genomic regions. The exonuclease activity of FAN1 acts 5′ to 3′ requiring gapped, nicked or resected DNA as an entry point (Kratz et al. 2010; Smogorzewska et al. 2010; Wang et al. 2014). The ability of FAN1 to cleave multiple DNA intermediates reflects its versatility as a DNA repair protein, capable of acting in multiple repair pathways. Importantly, recent efforts to elucidate FAN1's role in somatic instability has implicated the nuclease domain of FAN1 as important in protecting against instability (Deshmukh et al. 2021b; Goold et al. 2021; McAllister et al. 2022).

In a U2OS model cell system data indicates that FAN1 nuclease activity is important in protecting against somatic expansions, as the rescue of *FAN1*$^{-/-}$ cells with a construct expressing nuclease-dead (D960A) FAN1 failed to reduce CAG expansion rates to *FAN1*$^{+/+}$ levels (Goold et al. 2021). This data has been supported by a recent publication utilising HD patient-derived iPSCs. A D960A mutation was introduced via CRISPR-Cas9 homology directed repair (McAllister et al. 2022). In this instance cells exhibited increased CAG expansion rates which were not significantly different from *FAN1*$^{-/-}$ lines, indicating the nuclease activity of FAN1 is required at repeat regions. Furthermore, in a Fragile-X mouse model mice with a homozygous D963A mutation showed increased CGG expansion rates in striatal and cortical tissue. This was similar to the levels seen in *Fan1*$^{-/-}$ mice (Zhao et al. 2021).

Taken together these data convincingly show that the nuclease action of FAN1 is required for protection against CAG and CGG repeat expansions.

The way in which FAN1's nuclease activity acts at sites of expanded repeats, preventing somatic instability, is not fully known. However, data from biochemical-based assays provide an insight to which structures FAN1 may cleave in this context. Expanded CAG/CTG repeats are prone to stable secondary structure formation, including imperfect CAG/CTG loop outs (Pan et al. 2017; Xu et al. 2020). These structures can stall replication and transcription, therefore they must be resolved by DNA repair machinery. Biochemical-based assays have demonstrated that both nuclease actions of FAN1, the endo- and exonuclease, can act CAG/CTG loop out structures, where both functions are capable of cleaving/processing loop-outs. This study demonstrated that CAG/CTG loop outs within duplex DNA adjacent to a 5′ flap or on a 5′ flap can be cleaved by FAN1 (Deshmukh et al. 2021b). The action of FAN1 at such structures, likely acts to prevent somatic instability by promoting faithful repair at these sights. This DNA-binding and nucleolytic processing may also limit access to such structures from other DNA repair factors, such as the MMR machinery, which are error-prone and drive somatic expansions (Loupe et al. 2020; Goold et al. 2021; Phadte et al. 2023).

The DNA-binding domain of FAN1 has recently been implicated in somatic instability through GWA and exome-sequencing studies, where multiple variants are associated with earlier age at onset of HD. Two key variants, R377W and R507H, reach genome-wide significance, both leading to an earlier age of onset of HD, with the R507H variant associated with ~6 years earlier onset, on average (Lee et al. 2015; Lee et al. 2019; McAllister et al. 2022). The R507H variant has been predicted to impact FAN1 function (Bastarache et al. 2018). Follow-up investigations regarding this mutant have yielded conflicting reports, with some studies indicating functional deficits in nucleolytic processing, likely as a result of reduced DNA-binding (Kim et al. 2020; McAllister et al. 2022). Contrastingly, other studies demonstrates no deficiency in DNA-binding or nucleolytic processing of CAG/CTG loop-out structures (Deshmukh et al. 2021) or no effect on somatic expansion (Goold et al. 2019). Despite these reports the strength of the genetic data suggests

there is some functional consequence of this mutant, hopefully further studies will aid our understanding.

FAN1 is also thought to regulate somatic expansion through its interaction with MLH1. Insights from a HD knock-in mouse model has demonstrated that knocking-out Fan1 increases somatic expansion in mice after 3 months, but the simultaneous loss of Mlh1 ablates this phenotype (Loupe et al. 2020). This indicated that functional MLH1 is required for somatic expansion at the CAG repeat, with increasing expansions in the absence of FAN1 expression. This implies that FAN1 limits the activity of the MMR repair machinery, thereby preventing somatic expansions. This is supported in recent publications that identified a novel MLH1 interaction motif, ^{126}SPYF^{129} , in FAN1 (Goold et al. 2021; Porro et al. 2021). Data from immortalized U2OS and HD patient-derived iPSC lines suggest this motif is essential in mediating the binding of FAN1-MLH1, with mutations in this domain (FAN1^{F129A}) increasing somatic expansion rates compared with FAN1^{WT}. Interestingly, this increase was not as striking as in *FAN1* knockout lines. However, in lines expressing both the SPYF-mutant and nuclease-dead mutant, FAN1$^{F129A/D960A}$, somatic expansion rates were comparable to knockout lines, indicating the nuclease activity in combination with FAN1's MLH1 binding ability are important in protecting against somatic expansions in this specific cellular system (Goold et al. 2021).

The role of FAN1 in somatic instability, as well as other DNA repair pathways, is rapidly being uncovered. In Chapter 3 I demonstrated that a loss of FAN1 increases CRISPR-Cas9 nickase induced contractions in a HEK293 cell reporter line. This could suggest that the mechanisms for somatic expansion may differ to that of nickase-induced contractions and furthermore, that patients with heterozygous FAN1 reduced or loss-of-function variants may still benefit from this therapeutic approach, or indeed benefit more. Therefore, it is important to determine what functional domains of FAN1 play in preventing CRISPR-Cas9 nickase contractions.

4.2 Chapter Aims:

This chapter aims to rescue *FAN1*$^{-/-}$ GFPNickS120 cells, which contain our stably integrated Cas9 nickase and sgCTG, with FAN1 variant constructs to further elucidate our understanding of how FAN1 acts to prevent nickase-induced contractions. The variants investigated in this chapter are; a wild-type FAN1 construct (*FAN1-/-*^{WT}*),* a C44A/C47A mutation in the UBZ domain (*FAN1*$^{-/-UBZ}$), a nuclease-dead construct (*FAN1*$^{-/-D960A}$), an endonuclease-dead construct (*FAN1*$^{-/-Endo}$) and a construct containing the DNA binding variant R507H (*FAN1*$^{-/-R507H}$). These variants target a range of FAN1 functional domains and roles in DNA repair, and examining the effects of these in nickase-induced contractions will aid our understanding of how FAN1 protects against contractions.

A secondary aim of this chapter was explore whether the MMR machinery is responsible for the increased nickase-induced contractions seen in *FAN1*$^{-/-}$ GFPNickS120 cells. This is interesting as given FAN1's role in modulating MLH1 activity, we could speculate that in the absence of FAN1 there is increased MMR at sites of Cas9 nicks, which results in error-prone repair resulting in a bias towards CAG/CTG contractions events. This will aid our understanding of whether FAN1's role in preventing nickase-induced contractions is similar or different to its role in somatic instability.

4.3 Results

4.3.1 Generating FAN1 rescue lines in GFPNickS120 *FAN1*$^{-/-}$ clone #7

4.3.1.1 FAN1 variant construct design

All variants were designed and packaged for lentiviral production using Vector Builder (www.vectorbuilder.com) services. Due to the presence of other resistance selection markers within the HEK293 *FAN1*$^{-/-}$ line, a zeocin selection cassette was the only available marker and was tagged to the 3' end of the FAN1 ORF via an IRES site. A second mCherry marker under a CMV promoter was also included in each construct to allow for a quick read-out of successful transduction (Figure 4-2).

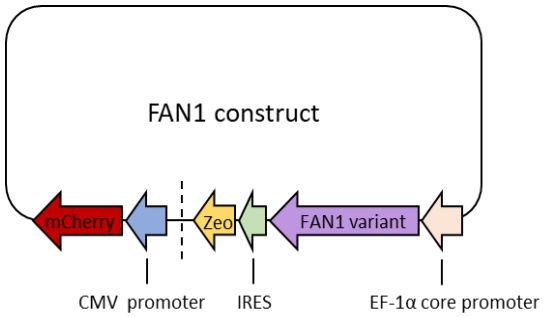


Figure 4-2: FAN1 variant construct design. *FAN variants were under the control of an EF-1α core promoter linked to a Zeocin resistance cassette through an IRES site. An mCherry tag under a CMV promoter allowed for screening of cells containing the desired vector.*

4.3.1.2 Selection with Zeocin stalls replication in FAN1$^{-/-}$ lines

Prior to transduction with desired FAN1 variant plasmids, a control mCherry expression plasmid was tested and different MOI's were screened in HEK293 cells to determine efficiency of transduction. MOI's of 2, 4 and 8 were tested and mCherry expression assessed 72 hours post transduction. Transduction efficiency was ~30, 50 and 90% for MOI's of 2, 4 and 8, respectively (Figure 4-3).

Previous literature has noted that significantly overexpressing a FAN1 nuclease-dead variant in *FAN1*$^{-/-}$ iPSCs decreased expansion rates. This argued that the nuclease activity was not essential for repeat expansion. However, it was later found that FAN1 overexpression likely led to sequestration of MMR protein MLH1, which is required to drive somatic expansion (Goold et al. 2019; Goold et al. 2021; Porro et al. 2021). Given this, I sought to keep FAN1 levels as close to the endogenous levels as possible and opted to use the lowest MOI of 2 which led to successful transduction.

The *FAN1*$^{-/-}$ clone #7 line, described in Chapter 3, was selected to perform FAN1 rescue experiments as its starting CAG repeat length was most comparable to the *FAN1*$^{+/+}$ parent line. As outlined in the above figure, FAN1 variant plasmids were designed with an zeocin selection cassette allowing for selection of cells which had been positively transduced. Therefore, the experimental design was to seed cells, transduce after 18 hours and after a 72 hour transduction period switch to selective media for 14 days (Figure 4-3). Previous kill curve experiments in HEK293 cells had determined a concentration of 100µg/ml was sufficient to kill any cells that had not integrated the resistance cassette over a 14 day selection period (data not shown).

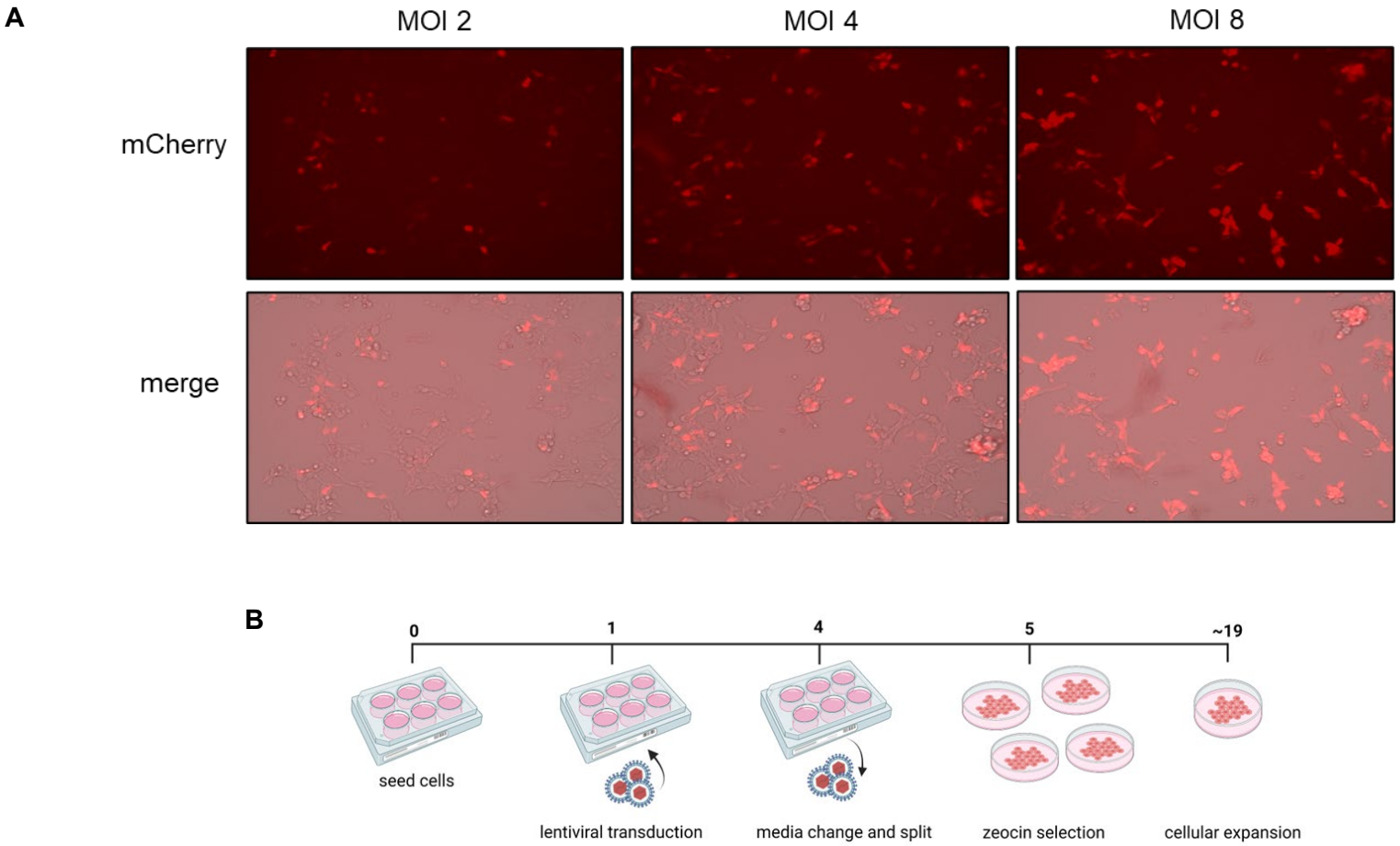


Figure 4-3: Initial experimental optimisation. ***(A)*** *Testing MOI conditions of control mCherry plasmid to screen for transduction efficiency in HEK293 cells. Merge = transmission light and mCherry.* ***(B)*** *Schematic outlining experimental plan. 18 hours after seeding cells were treated with lentiviral particles containing the desired FAN1 variant construct. 72 hours post-transduction media was changed to zeocin conditioned media, and zeocin selected was to be carried out for 14 days.*

Cells were transduced with an MOI of 2 and following lentiviral removal cells were maintained in zeocin selective media. Unexpectedly, I observed complete cell death 2 days after transduction. MOI's of 4 and 8 also led to complete cell death after 4 days, despite the appearance of mCherry positive cells. After using an MOI of 10, much higher than originally intended, I observed that cells which were mCherry positive survived the selection However, these cells still demonstrated limited growth rates compared to untreated cells and caused many cells to form rounded bodies that remained attached the surface of the plate and did not replicate.

The mCherry expression was under the control of the CMV promoter, whilst FAN1 and zeocin expression was under an EF1-α promoter. Therefore, a western blot was carried out to determine whether the EF1-α promoter was acting effectively. In cells with an MOI of 10 there was a vast overexpression of all FAN1 variants in *FAN1* $^{-/-}$ *#7* rescued lines, compared with endogenous protein levels (Figure 4-4). Indicating correct functioning of the EF1-α promoter. However, an overexpression to such an extent would likely mask any subtle phenotypes and was not fit for the purposes of the experiment.

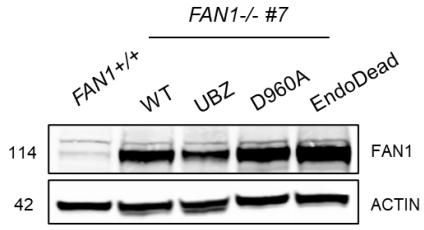


Figure 4-4: Western blot for variants at MOI of 10. *Comparative expression for all transduced variants compared with FAN1*$^{+/+}$ *lines. Molecular weight of FAN1 (114) and ACTIN (42) in KDa.*

4.3.1.3 FACs sorting of mCherry positive cells allows for creation of stable FAN1 variant lines

Due to the toxic effects of zeocin in the FAN1 knock-out cells an alternative method to generate stable lines was utilised. The mCherry marker acts as a useful tool for screening successfully transduced cells allowing for FACS to isolate cells expressing mCherry. As the mCherry expression is independent of FAN1 expression a new series of MOIs were screened, lower than previously, at 1, 2.5 and 5. For each MOI transduced, cells were sorted into three categories; no, low or high mCherry expression. Low and high indicating the bottom or top 50% cells within the mCherry positive population. This allowed for the more effective screening of a range of FAN1 expression levels in each condition, to reduce the risk of a large overexpression compared with endogenous levels (Figure 4-5).

After sorting cells into three populations, they were left to expand and samples were taken for protein analysis via western blot. Western blot analysis revealed no difference in FAN1 expression between the MOIs in both low or high expression mCherry populations (Figure 4-5). For all MOI's the population of cells in the low mCherry expression yielded FAN1 protein levels most comparable to endogenous levels. An MOI of 1 was taken forward based upon protein levels and calculated probability of copy number integration per cell being lower.

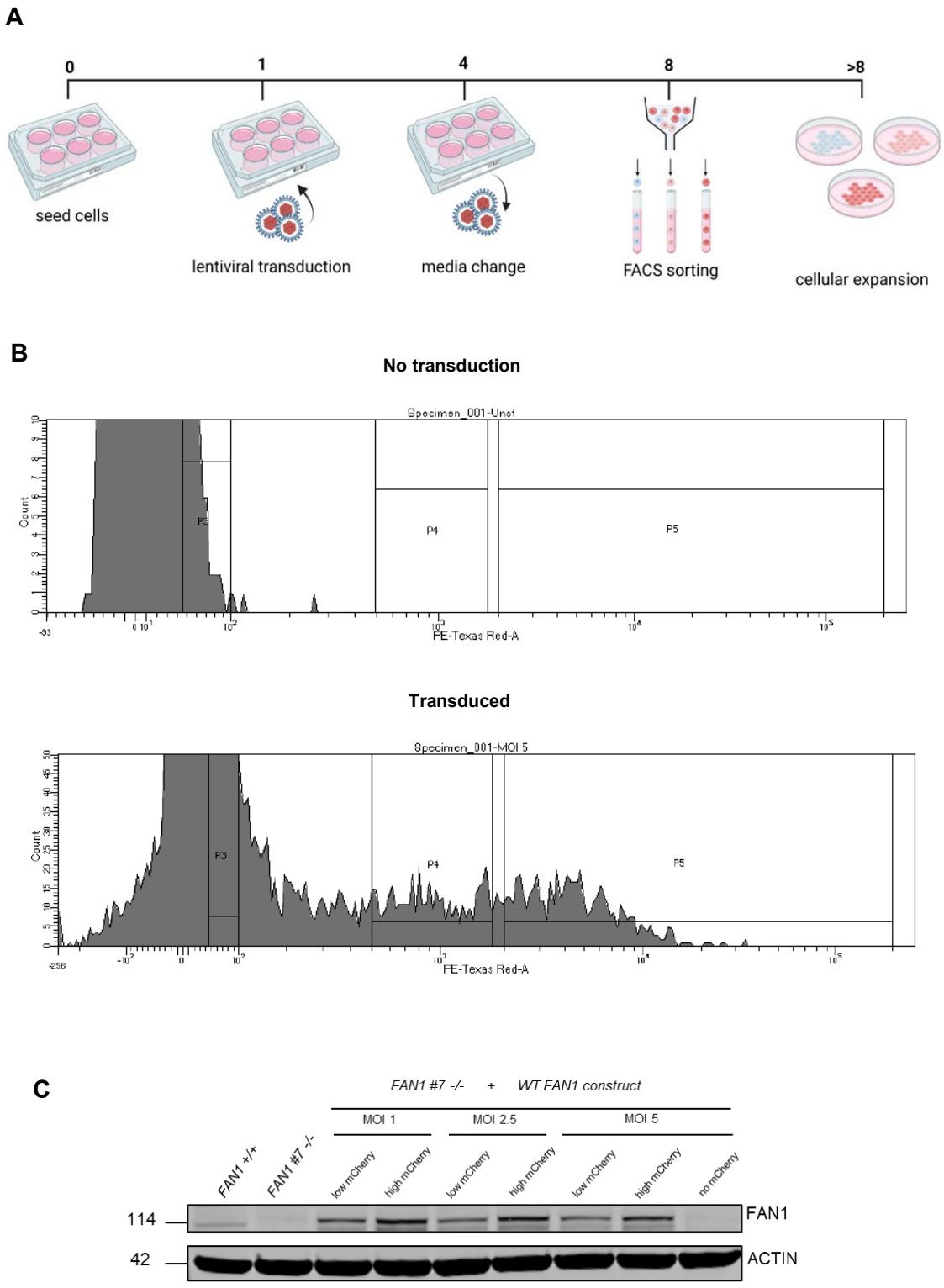


Figure 4-5: mCherry FACs sorting to allow for creation of stable variant lines. ***(A)*** *Experimental outline for the creation of stable variants. After transduction mCherry positive cells were FAC sorted into three populations; no, low and high mCherry expression. All populations were expanded and frozen for validation.* ***(B)*** *Demonstrative mCherry density plots comparing non-transduced and transduced cells. P3 indicates no mCherry expression and P4 and P5 indicate low or high, respectively.* ***(C)*** *Western blot demonstrating overexpression of FAN1 variants compared with endogenous FAN1$^{+/+}$ levels. Molecular weight of FAN1 (114) and ACTIN (42) in KDa.*

4.3.1.4 Validation of FAN1 variant lines

After determining the optimum transduction conditions the remaining variant lines were generated following this protocol (Figure 4-5). After sorting and cellular expansion, all lines were screened via PCR for integration of desired plasmid. Due to the presence of endogenous FAN1 DNA, PCR primers were designed specifically flanking the FAN1 insert region of our plasmid to generate a 3.5Kb amplicon. Samples were then sent for Sanger sequencing to confirm correct variant integration. After generation of cell lines, pellets were taken to monitor variant expression throughout the experimental timeline (Figure 4-6).

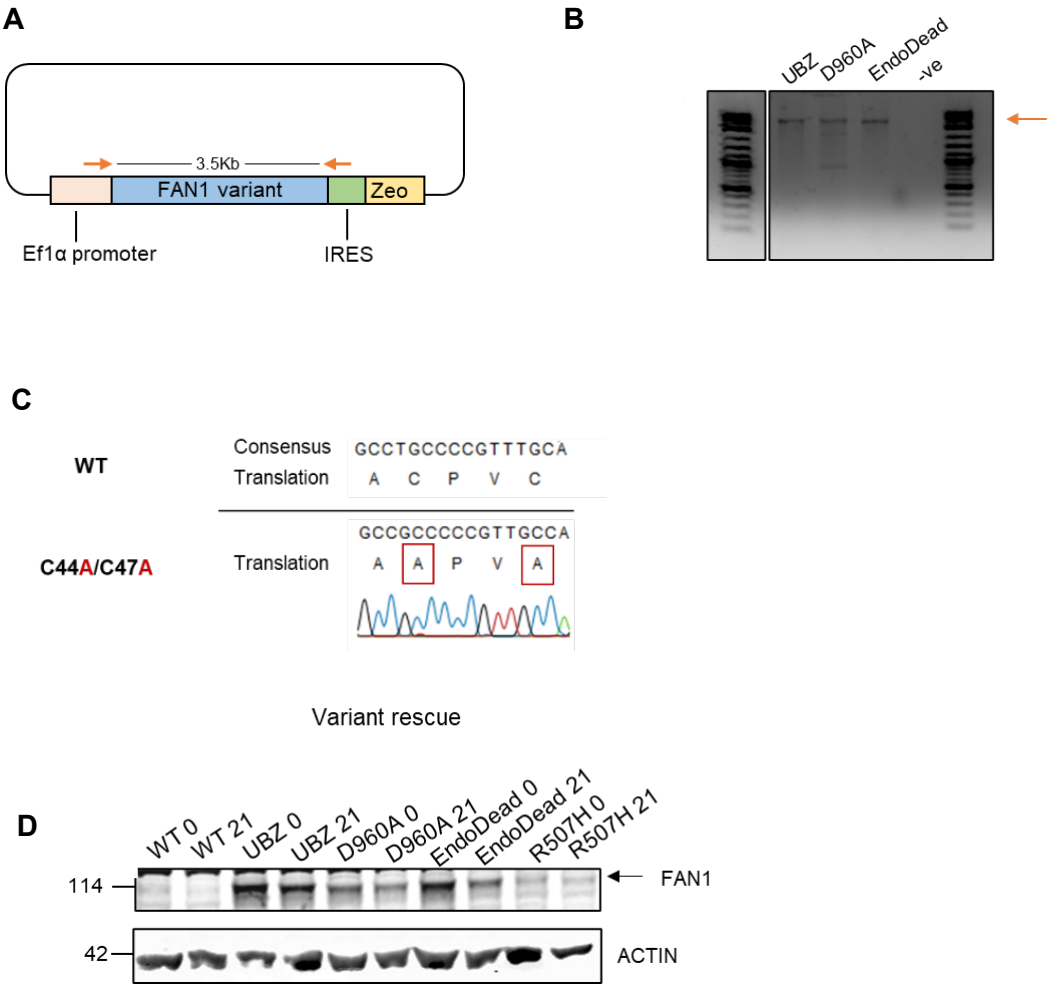


Figure 4-6: Validation of FAN1 variant lines. ***(A)*** *Schematic depicting primer pair to confirm integration of desired FAN1 variant plasmid. Amplicon size is 3.5Kb.* ***(B)*** *Demonstrative gel electrophoresis image confirming visible band at the desired fragment size for the rescued FAN1$^{-/-}$ lines.* ***(C)*** *Demonstrative Sanger sequence trace confirming variant integration.* ***(D)*** *Demonstrative western blot for FAN1$^{-/-}$ lines rescued with FAN1 variants at day 0 and day 21 of experimental timeline. Molecular weight of FAN1 (114) and ACTIN (42) in KDa.*

4.3.2 The effect of FAN1 variants on the GFP expression

To determine which FAN1 domains may play a role in nickase-induced contractions I made use of the integrated GFP mini gene within this cell type, whereby GFP intensity can serve as a proxy for changes in CAG repeat length (Santillan et al. 2014; Cinesi et al. 2016a). As previously detailed in Chapter 3, cells were grown in culture for 21 days in the presence of doxycycline to induce transcription of the GFP mini gene containing an intronic CAG repeat. To determine how rates of contractions may change over time, cells were treated in parallel for 7 or 21-days. At the end of the time course cells were collected for analysis via flow cytometry.

As previously described two modes of analysis were utilised to assay for changes in GFP expression. Firstly, data from all experimental replicates were collated and the fold-change between mean GFP expression at 7 and 21-days post-doxycycline treatment was analysed. As discussed in Chapter 3 the *FAN1*$^{-/-}$ *#7* line demonstrates an increase in GFP expression over the time course, compared to the *FAN1*$^{+/+}$ parent line. Therefore, I expected to see some reversal of this phenotype when complementing the *FAN1*$^{-/-}$ line with various FAN1 constructs.

As shown in Figure 4-7 only cells complemented with the R507H variant (*FAN1*$^{-/-R507H}$) demonstrated a significant rescue in mean GFP expression, similar to *FAN1*$^{+/+}$ cells (*P=0.0080*). However, when comparing the spread of data points there does appear to be a trend towards recovery of the phenotype in the knockout lines that have been rescued with the WT (*FAN1*$^{-/-WT}$), D960A (*FAN1*$^{-/-D960A}$) or EndoDead (*FAN1*$^{-/-Endo}$). The lack of significance may be due to variability in the *FAN1*$^{+/+}$ and *FAN1*$^{-/-}$ data sets, where each biological replicate was run through the flow cytometer on different days, which can introduce variation. In comparison, for all biological replicates carried out on the variant lines flow cytometry was run on the same day, which seems to decrease variability in the data set.

As previously described the top 10% of GFP-positive cells, representing larger contraction events, were also compared between day 7 and day 21 of doxycycline treatment and fold change calculated. In this case there was no

significant rescue with any FAN1 variants, however as previously discussed in Chapter 3 there does seem to be a higher degree of variation in this mode of analysis (Figure 4-7). As with the first analysis there is a clear trend towards phenotypic recovery in *FAN1*$^{-/-WT}$, *FAN1*$^{-/-D960A}$ and *FAN1*$^{-/-Endo}$ cells. Interestingly, in this analysis cells transduced with the C44A/C47A UBZ domain mutant (*FAN1*$^{-/-UBZ}$) and *FAN1*$^{-/-R507H}$ showed little/reduced recovery.

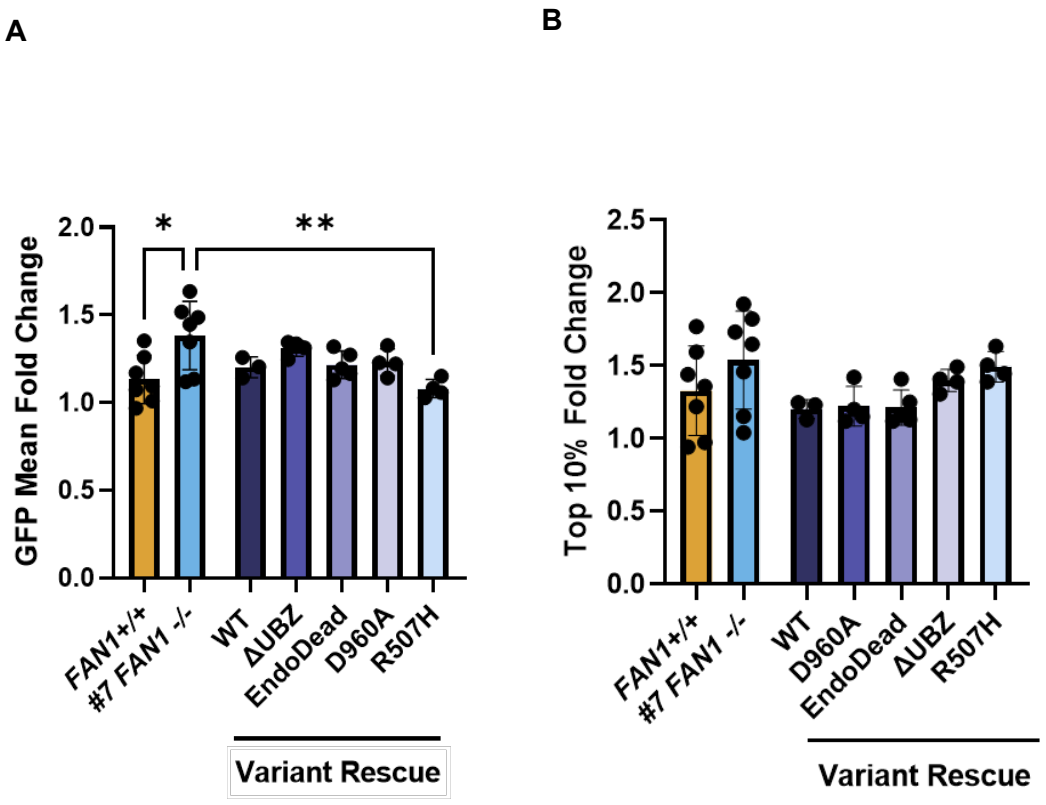


Figure 4-7: The effects of FAN1 variant rescue on GFP expression. ***(A)*** *Comparison of changes in GFP mean expression from day 7 to day 21 for all lines.* ***(B)*** *Fold change of contraction frequency in all lines from day 7 to day 21 (* P=<0.05, ** P=<0.01 One-way ANOVA with Tukey's multiple comparisons test). Error represents ± SD, (n=3-7). Biological replicates for FAN1*$^{+/+}$ *and #7 FAN1*$^{-/-}$ *lines are the same as those presented in Figure 3-7.*

4.3.3 PacBio long-read sequencing confirms changes in CAG contraction rates in rescue lines

As previously described, PacBio is a high-throughput sequencing approach allowing for long-read sequencing of repetitive DNA regions (Taylor et al. 2022). For all FAN1 variant lines, DNA samples were collected on day 0 and day 21 post-doxycycline treatment and prepared for amplicon SMRTbell sequencing. Repeat detector generated histogram reads for all samples and the mean CAG repeat was recorded and average change in CAG length calculated. The graph below indicates the size of the CAG contraction from day 0 to day 21. For all lines at least 4 biological replicates were analysed.

In this analysis only *FAN1*$^{-/-WT}$ cells rescued nickase-induced contractions to *FAN1*$^{+/+}$ levels (*P=0.3061*). All other variants demonstrated no rescue, or partial rescue, with profiles similar to the top 10% GFP expression analysis, indicating this mode of GFP analysis may be more accurate (Figure 4-8). *FAN1-/-*^{R507H} cells demonstrated no rescue when compared to *FAN1+/+* (*P=<0.0001)* and *FAN1-/-*^{WT} lines (*P=0.0003). FAN1-/-*^{UBZ} , *FAN1-/-*^{D960A} and *FAN1-/-*^{Endo} lines demonstrated a partial rescue compared with *FAN1*$^{+/+}$ lines, (*P=0.0002*, *P=0.0024* and *P=0.0031*). These lines demonstrate a partial recovery as whilst there is a significant difference in CAG contractions relative to *FAN1*$^{+/+}$ lines, these lines were not significantly different from *FAN1*$^{-/-}$WT lines. Together, these results argue that DNA-binding is essential for FAN1's protective effect against Cas9-nickase induced contractions, with the UBZ domain and nuclease functions of FAN1 also implicated.

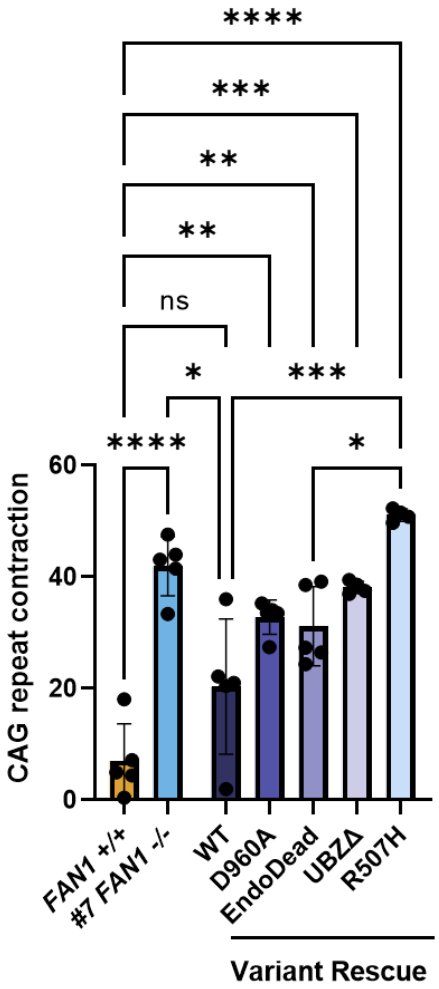


Figure 4-8: PacBio long-read sequencing investigating FAN1 variants. *Bar chart depicting the change in CAG repeat size from day 0 to day 21 of treatment with Cas9 nickase in all lines.(*P=<0.05, **P=<0.01, ***P=<0.001, ****P=<0.0001. One-way ANOVA with Tukey's multiple comparison test). Error represents ± SD, (n=4-5). Biological replicates for FAN1$^{+/+}$ and #7 FAN1$^{-/-}$ lines are those presented in Figure 3-10.*

4.3.4 Investigating the role of MLH1 in nickase-induced contractions

4.3.4.1 Generation of FAN1 and MLH1 double knock-out lines in GFP(CAG101) cells without Cas9 nickase expression

In mammalian cells, somatic expansion is dependent upon MMR machinery, and FAN1 can act to protect against MMR-induced expansion by competing with MutSβ for binding of CAG/CTG loop-outs or by modulating MLH1 recruitment (Goold et al. 2021; Phadte et al. 2023). In a HD knock-in mouse, loss of Fan1 increases somatic expansion relative to wild-type. Furthermore, a dual knock-out of *Fan1* and *Mlh1* ablates somatic expansion (Loupe et al. 2020). These data demonstrate that MLH1 is required for the increase expansion observed in FAN1 loss. Therefore, I hypothesized that the increased nickase-induced contractions seen in *FAN1*$^{-/-}$ cells may be mediated by MLH1, where FAN1 is required to reduce instability at the CAG repeat in both the absence and presence of the Cas9 nickase and MLH1 acts to increase instability.

Prior to investigating this effect on nickase-induced contractions I created dual *FAN1* and *MLH1* knockout lines the GFP(CAG101) cells, which do not contain the Cas9 nickase. This allowed me to determine whether a loss of MLH1, in *FAN1* knock-out cells, reduces somatic expansions to wild-type levels, in line with the literature (Loupe et al. 2020). This would further validate that the HEK293 cells are able to model aspects of somatic instability reported in other models.

The GFP(CAG101) *FAN1*$^{-/-}$ line was transfected with a dual sgRNA Cas9 RNP that generates a 68bp deletion in exon 1 of *MLH1,* leading to a frameshift mutation and presumably a null allele. sgRNAs were designed using DESKGEN Cloud (www.deskgen.com) based on their high on-target editing and low off-target scores. *MLH1*-specific targeting cRNAs were synthesized and duplexed with a tracrRNA scaffold containing an ATTO 550 fluorescent tag. The functional sgRNA was then incubated with an Alt-R *S.p.* Cas9 Nuclease V3 to form a stable RNP complex. Cells were electroporated with this fluorescently labelled complex and, after 24 hours, sorted based on

fluorescence intensity, with only the top 10% of cells selected (Figure 4-9). Single cell clones were then screened for the presence of the correct deletion product.

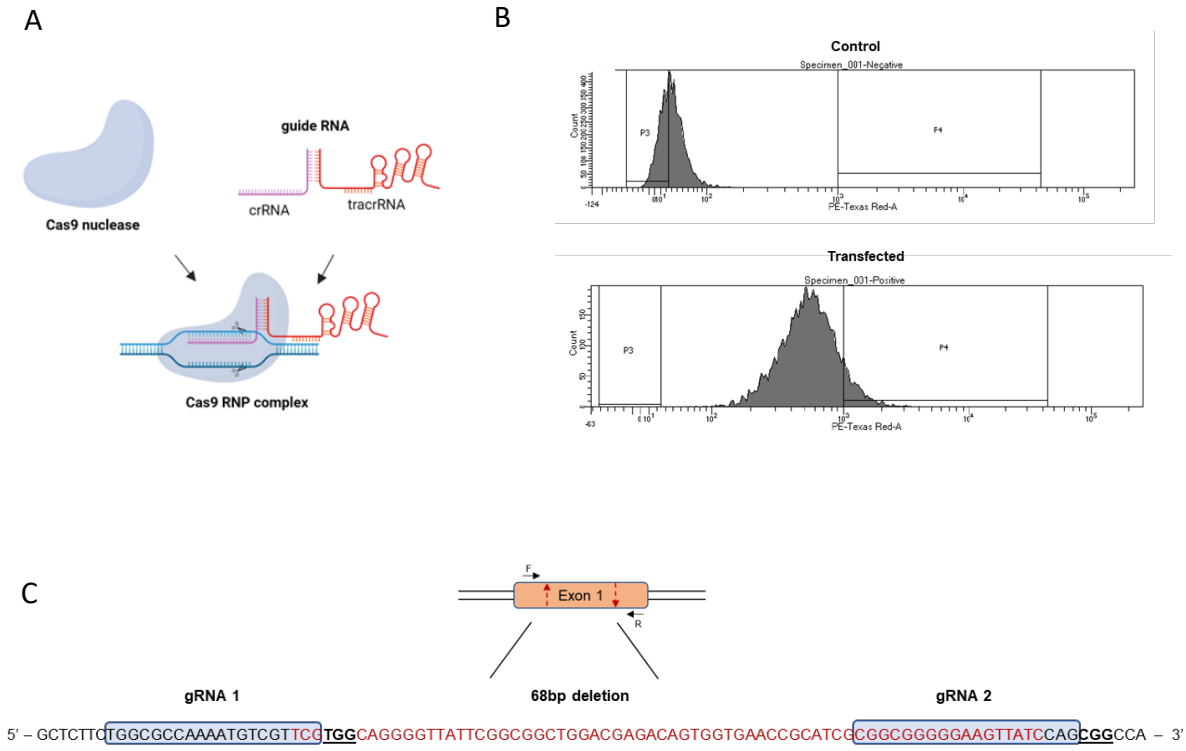


Figure 4-9: CRISPR-Cas9 MLH1 targeting strategy. ***(A)*** *CRISPR-Cas9 facilitates gene editing as a Cas9 nuclease is targeted to the desired in the genome by a gRNA. Together the nuclease and gRNA form a complex leading to the formation of a DSB at editing site.* ***(B)*** *Two gRNAs were designed targeting exon 1 of MLH1 to create a 68p deletion.* ***(C)*** *The top 10% of fluorescently labelled cells (P4) were sorted and plated for characterization and expansion.*

4.3.4.1.1 Screening of targeted clones

Genomic DNA was harvested from individual clones and PCR primers amplifying the predicted target sites were used and PCR-products run out via gel electrophoresis. Candidate clones with a lower molecular weight (302 bp) and presence of one clear band with the desired deletion were expanded and screened further. Sanger sequencing of edited clones was carried out to assess CAG repeat lengths comparable to the parent *FAN1*$^{-/-}$ line. Clones B5, C2 and C3 were carried forward for experiments as they had CAG repeat lengths closest to the parent line. An additional PCR across exon 2 of *FAN1* was carried out to re-confirm *FAN1* knock-out status, this confirmed the presence of a 94 bp deletion visible by gel electrophoresis (as described in 3.2.4.2). Sanger sequencing analysis of all three clones confirmed the presence of a 68bp deletion within exon 1 of *MLH1* and a loss of detectable protein via western blot confirmed knockout status (Figure 4-10).

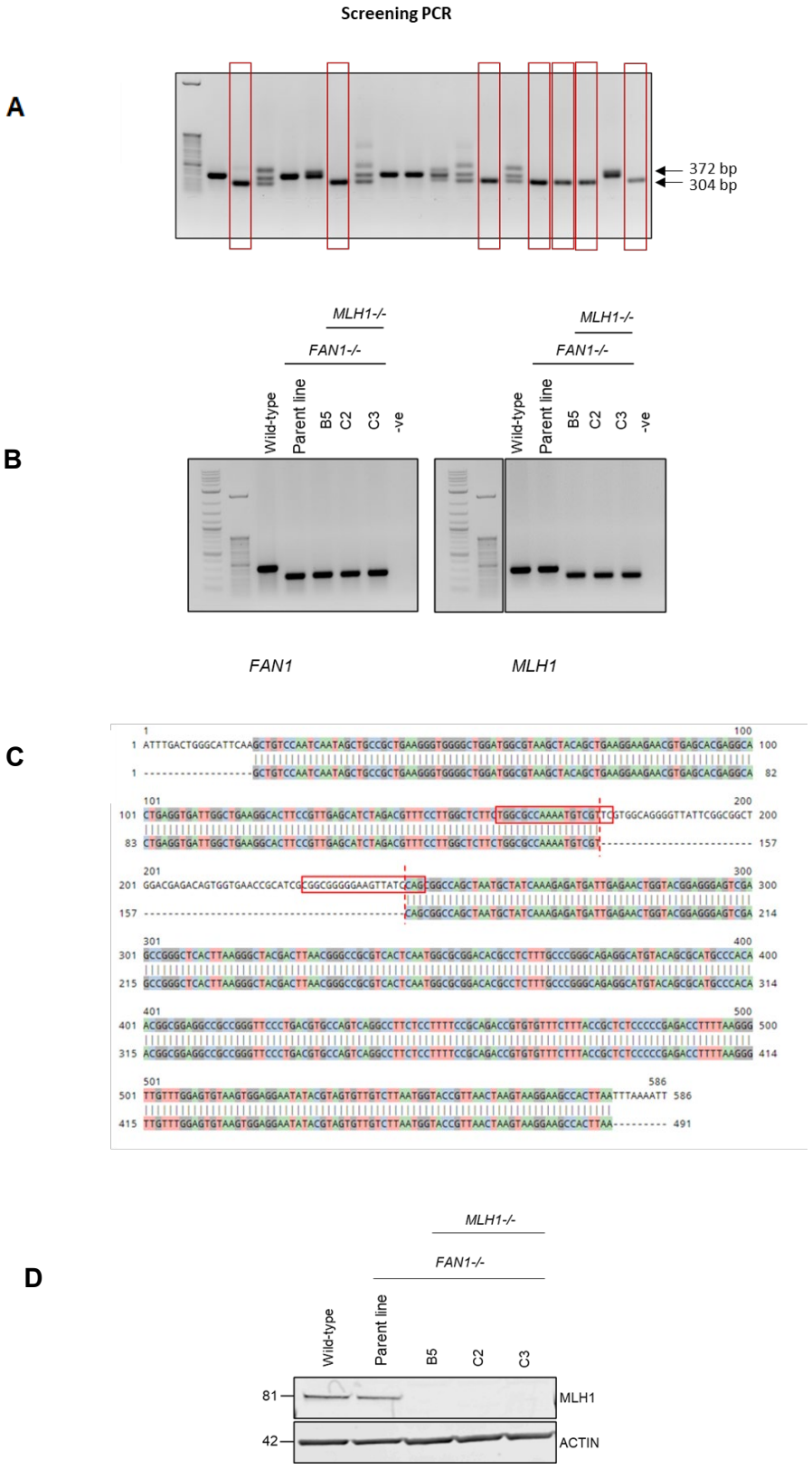


Figure 4-10: Screening CRISPR strategy. ***(A)*** *Screening gel electrophoresis results after transfection with Cas RNP complex targeting MLH1.* ***(B)*** *PCR confirmation in clones selected for dual FAN1 and MLH1 knock-out status.* ***(C)*** *Sanger sequencing of PCR products confirming 68bp deletion. Red box indicates sgRNA sequence and dotted line the predicted cut site.* ***(D)*** *Western blot confirming knockout of MLH1 in FAN1 and MLH1 double knock-out lines. Molecular weight of MLH1 (81) and ACTIN (42) in KDa.*

4.3.4.2 Loss of MLH1 ablates somatic expansion in FAN1$^{-/-}$ GFP(CAG101) cells

To determine whether a loss of MLH1, in conjunction with FAN1, in these double knock-out lines reduces somatic expansion to wild-type levels lines were cultured for a period of 42-days. Pellets were collected at day 0 and day 21 and 42 post-doxycycline treatment. DNA was extracted and amplicon PacBio SMRTbell sequencing libraries prepared. For this analysis the modal CAG repeat for each line at day 0 was calculated and compared with day 21 and day 42. Modal CAG values were plotted and a linear regression performed for each genotype. In all *FAN1/MLH1* double knock-out lines there was a complete loss of somatic expansion, comparable with wild-type levels (Figure 4-11).

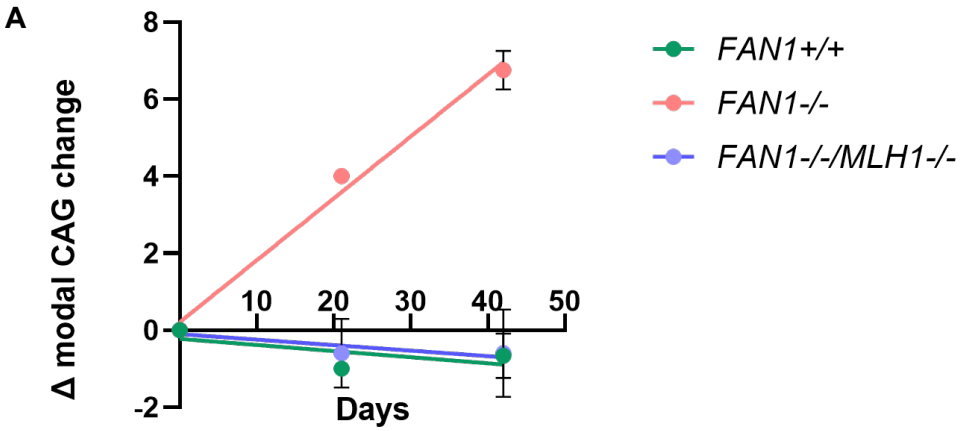


Figure 4-11: Investigating somatic expansion rates in FAN1$^{-/-}$/MLH1$^{-/-}$ clones. ***(A)*** *Demonstrative PacBio histogram plots.* ***(B)*** *Change in modal CAG repeat lengths over 42 days. The three FAN1/MLH1$^{-/-}$ lines were pooled for comparison on one graph. FAN1/MLH1$^{-/-}$ and FAN1$^{+/+}$ lines do not demonstrate a significant increase in CAG expansion (P=0.2670 and P=0.1269), compared with the FAN1$^{-/-}$ line which does (P<0.0001). As determined by a linear regression calculating deviation from zero (n=2-4) Error represents ± SD. Biological replicates for FAN1$^{+/+}$ and FAN1$^{-/-}$ lines are those presented Figure 3-13.*

4.3.4.3 Generation of a FAN1 and MLH1 double knock-out in GFPNickS120 cells

4.3.4.3.1 CRISPR-Cas9 mediated targeting of MLH1

After confirming that a double knock-out of *FAN1* and *MLH1* ablates somatic expansion in our HEK293-reporter cells, as expected, I sought to explore whether this interaction between FAN1 and MLH1 is also relevant in the context of nickase-induced contractions. I attempted to generate double knockout clones in the *FAN1*$^{-/-}$ clone #7 and single *MLH1*$^{-/-}$ lines in cells which express the CRISPR-Cas9 nickase treatment.

The challenge in knocking out genes in these cells is that they already contain a stably integrated *S. pyogenes* Cas9 nickase and sgCTG targeting the CAG repeat tract. If I provide the *S. pyogenes* Cas9 nuclease exogenously along with sgRNAs that target *MLH1*, there is a risk the Cas9 will bind the sgCTG in these cells and induce a double-strand break at ~140 sites in the genome that contain CAG repeats. To avoid this, I made use of a plasmid containing a functional *Staphylococcus aureus* Cas9 endonuclease and sgRNA scaffold (Addgene: #96920). The *S. aureus* Cas9 and *S. pyogenes* Cas9 RNAs have different scaffold structures, as such the *S. aureus* Cas9 will be unable to bind the *S. pyogenes* sgRNA preventing editing a CAG regions (Ran et al. 2015).

A sgRNA targeting exon 2 of MLH1 was designed using the sgRNA design tool CRISPick (CRISPick (broadinstitute.org)). After sgRNA sequence design, it was synthesised as a double stranded oligonucleotide with flanking BsmbI sites and appropriate overhang sequences for Golden Gate cloning into the *S. aureus* expression plasmid. The successful integration of the gRNA was confirmed via sanger sequencing (Figure 4-12).

As previously described in Chapter 3 transfections in were performed using the transfection agent lipofectamine 2000. Both *FAN1*$^{+/+}$ and *FAN1*$^{-/-}$ cells were transfected. To separate the transfected cells from the untransfected ones, cells were co-transfected with a plasmid expressing BFP. 48 hours post-

transfection both transfected and non-transfected control cells were sorted using the BD FACS Aria Fusion flow cytometer.

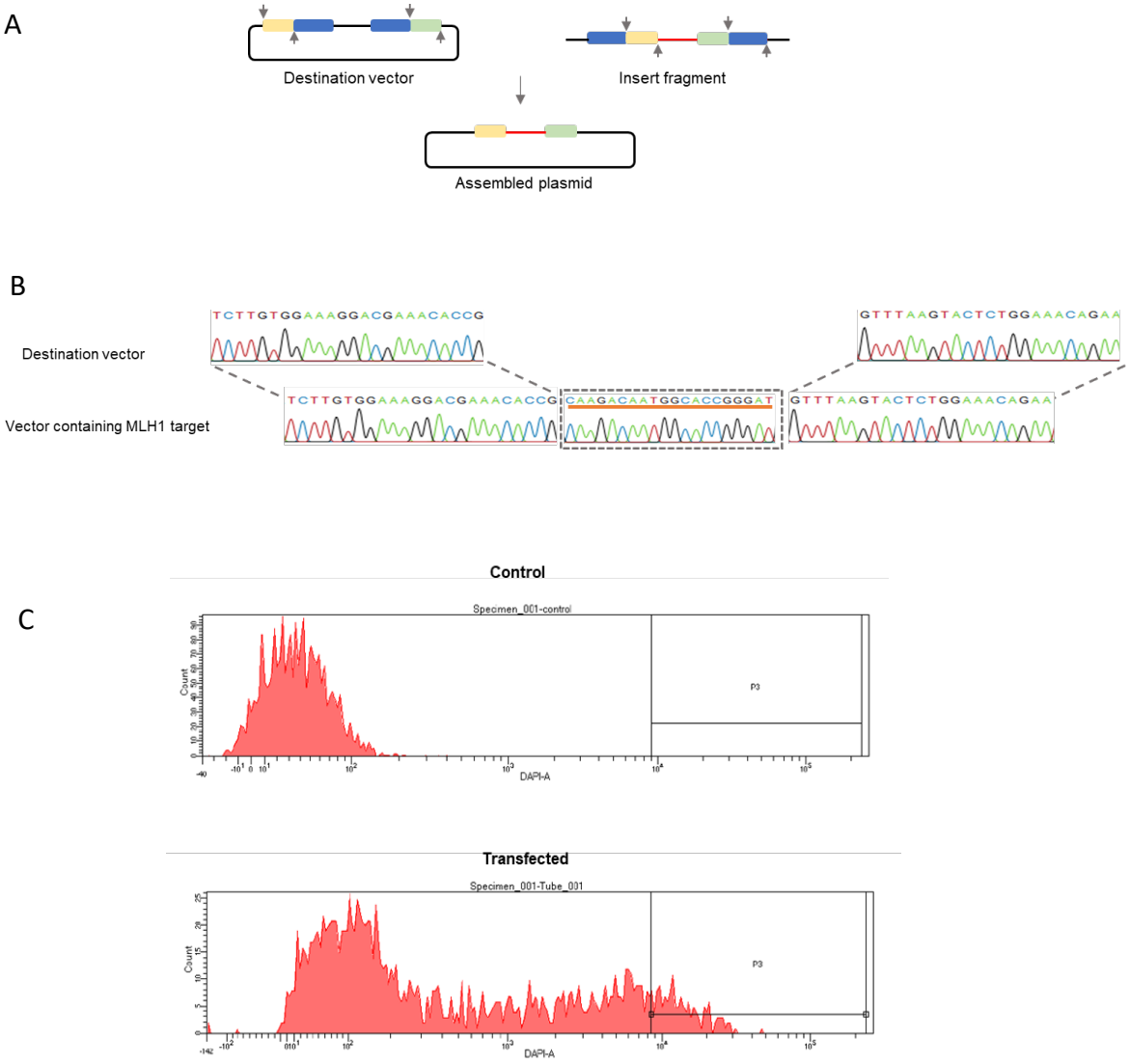


Figure 4-12: MLH1 CRISPR targeting strategy in GFPNickS120 cells. ***(A)*** *Schematic depicting golden gate cloning where Bsm1 sites in the destination vector and insert fragment direct targeted cutting to form products with complementary overhangs, allowing the assembly of desired plasmid with MLH1 target insert and sgRNA scaffold and S.aureus Cas9.* ***(B)*** *Sanger sequencing confirming successful integration of the MLH1 target sequence into the destination vector.* ***(C)*** *FACS analysis depicting un-transfected verses cells transfected with BFP. Cells expressing top 10% were gated for single cell clones.*

4.3.4.3.2 Screening of targeted clones

A two-pronged screening approach was employed to identify candidate clones. Firstly, genomic DNA was harvested from all clones and an initial PCR-based screening approach was carried out to identify clones with similar CAG repeat lengths. PCR products were visualised by gel electrophoresis and those containing an expanded CAG repeat comparable with the *FAN1*$^{-/-}$ parent line were then screened for MLH1 expression. Due to the availability of an efficient commercial antibody against MLH1, all clones with an expanded CAG repeat comparable to *FAN1*$^{+/+}$ and *FAN1*$^{-/-}$ lines were screened for a loss of protein expression by western blot. Unfortunately, this CRISPR-editing approach was not very efficient and to date I have only obtained one *FAN1* and *MLH1* double knock-out clone, and no single *MLH1* knock-outs. This double knock-out clone was characterised as described above, and western blotting was carried out to establish that Cas9 expression was comparable with the parent line (Figure 4-13).

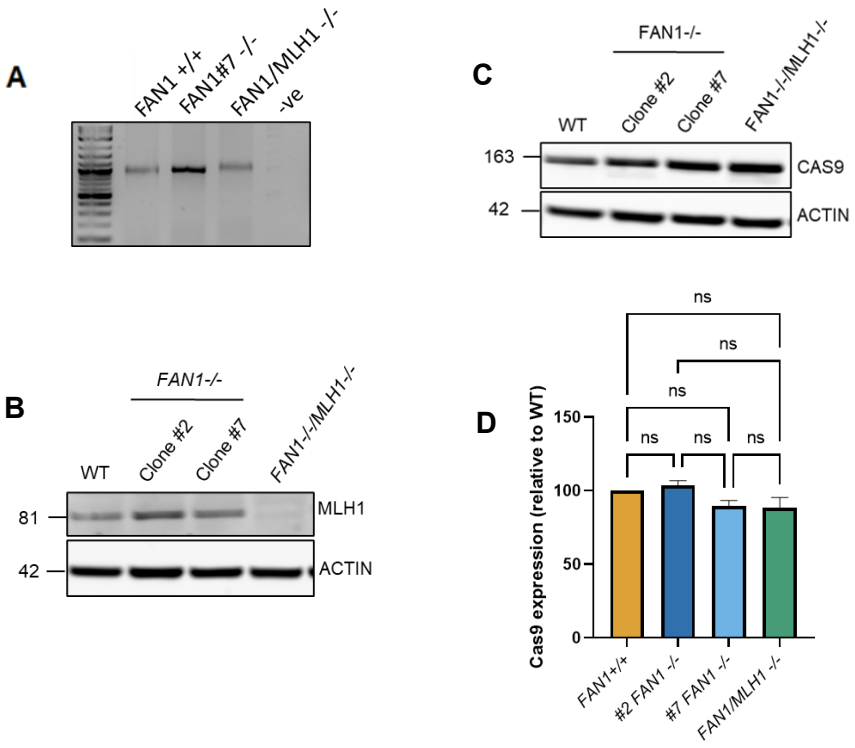


Figure 4-13: Characterization of FAN1$^{-/-}$/MLH1$^{-/-}$ clone. ***(A)*** *PCR comparing CAG repeat size between double knock-out and parent FAN1$^{+/+}$ and FAN1$^{-/-}$ lines.* ***(B)*** *Western blot confirming loss of functional MLH1 protein in double knock-out clone. Molecular weight of MLH1 (81) and ACTIN (42) in KDa.* ***(C)*** *Western blot confirming Cas9 expression in double knock-out line. Molecular weight of CAS9 (163) and ACTIN (42) in KDa.* ***(D)*** *Quantification of Cas9 expression between wild type, FAN1$^{-/-}$ and FAN1$^{-/-}$/MLH1$^{-/-}$ clones. (ns = P>0.05, One-way ANOVA and Tukey's multiple comparisons). Error represents ± SD, (n=3). Replicates for FAN1$^{+/+}$ and #2 and #7 FAN1$^{-/-}$ lines are those presented in Figure 3-4.*

4.3.4.3.3 FAN1 and MLH1 double knock-out GFPNickS120 line has no detectable proliferation deficits

We have previously shown that a loss of FAN1 in our HEK293 cells does not alter proliferation rates, which is consistent with previous literature (Yoshikiyo et al. 2010). However, it has not been established whether a combined loss of FAN1 and MLH1 would impact growth rates. To determine this, all lines were cultured for 12 days, and proliferation rates monitored. Cells were seeded at a starting density of $2x10^6$ cells in a 10 cm^2 plate. Every 3 days cells were recovered using trypsin and resuspended in pre-warmed DMEM + GlutaMAX (10% FBS) media. Cells were dissociated and a sample was taken to measure cell count. The cells were subsequently re-plated at the starting cell density. Over the time course a cumulative cell count was recorded (Figure 4-14).

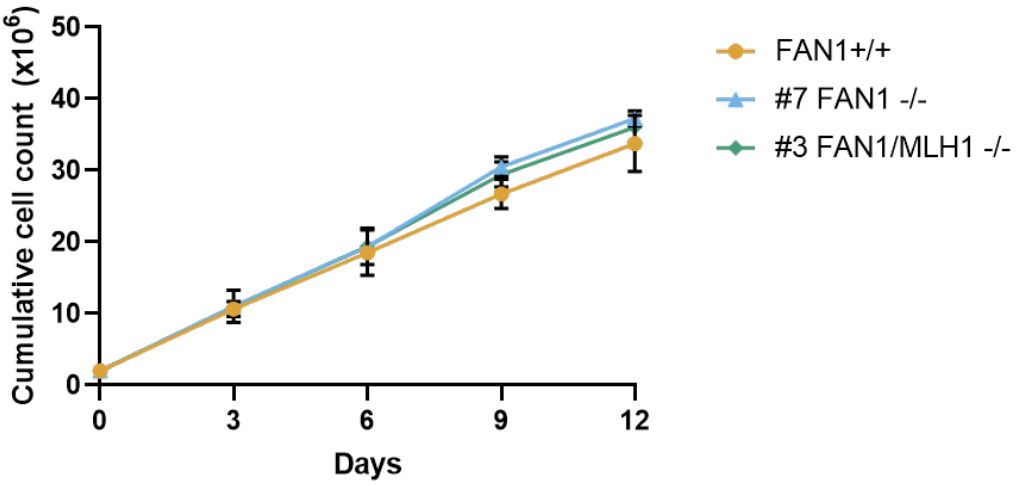


Figure 4-14: Comparison of growth rates between FAN1$^{-/-}$/MLH1$^{-/-}$ line and parent FAN1$^{+/+}$ and FAN1$^{-/-}$ lines. *No significant difference in growth rates was seen between all lines, as determined by a Two-way ANOVA followed by Tukey's multiple comparison test (n=4). Error represents ± SD. Biological replicates for FAN1$^{+/+}$ and #7 FAN1$^{-/-}$ lines are those presented in Figure 3-6.*

4.3.4.4 The *FAN1*$^{-/-}$/*MLH1*$^{-/-}$ GFPNickS120 line demonstrates increased GFP expression over time compared with *FAN1*$^{-/-}$ or *FAN1*$^{+/+}$ parent lines

To determine whether the dual loss of both FAN1 and MLH1 altered nickase-induced contraction rates compared with *FAN1* knockout or wild-type lines, cells were grown in culture for a period of 21-days, as previously described. During this time cells were treated with doxycycline to induce transcription through the CAG repeat tract. To determine how the rate of contractions may change over time, cells were treated in parallel with doxycycline for 7 or 21-days. At the end of the time-course cells were collected for analysis via flow cytometry and the fold change in GFP expression between the two time points was compared.

As previously described two modes of analysis were utilised to assay changes in GFP expression; firstly data from all experimental replicates was taken and the fold-change between mean GFP expression at 7 and 21-days post doxycycline treatment was analysed. *FAN1/MLH1*$^{-/-}$ cells demonstrated a significant increase in GFP expression relative to *FAN1*$^{+/+}$ cells (*P=<0.0001)* and a further increase relative to *FAN1*$^{-/-}$ levels (*P=0.0100)*. The second mode of analysis comparing changes in the top 10% of GFP-positive cells followed the same trend. There was a significant increase in GFP expression in the *FAN1/MLH1*$^{-/-}$ cells relative to *FAN1*$^{+/+}$ (*P=<0.0001*) and *FAN1*$^{-/-}$ lines (*P=<0.0001*) (Figure 4-15).

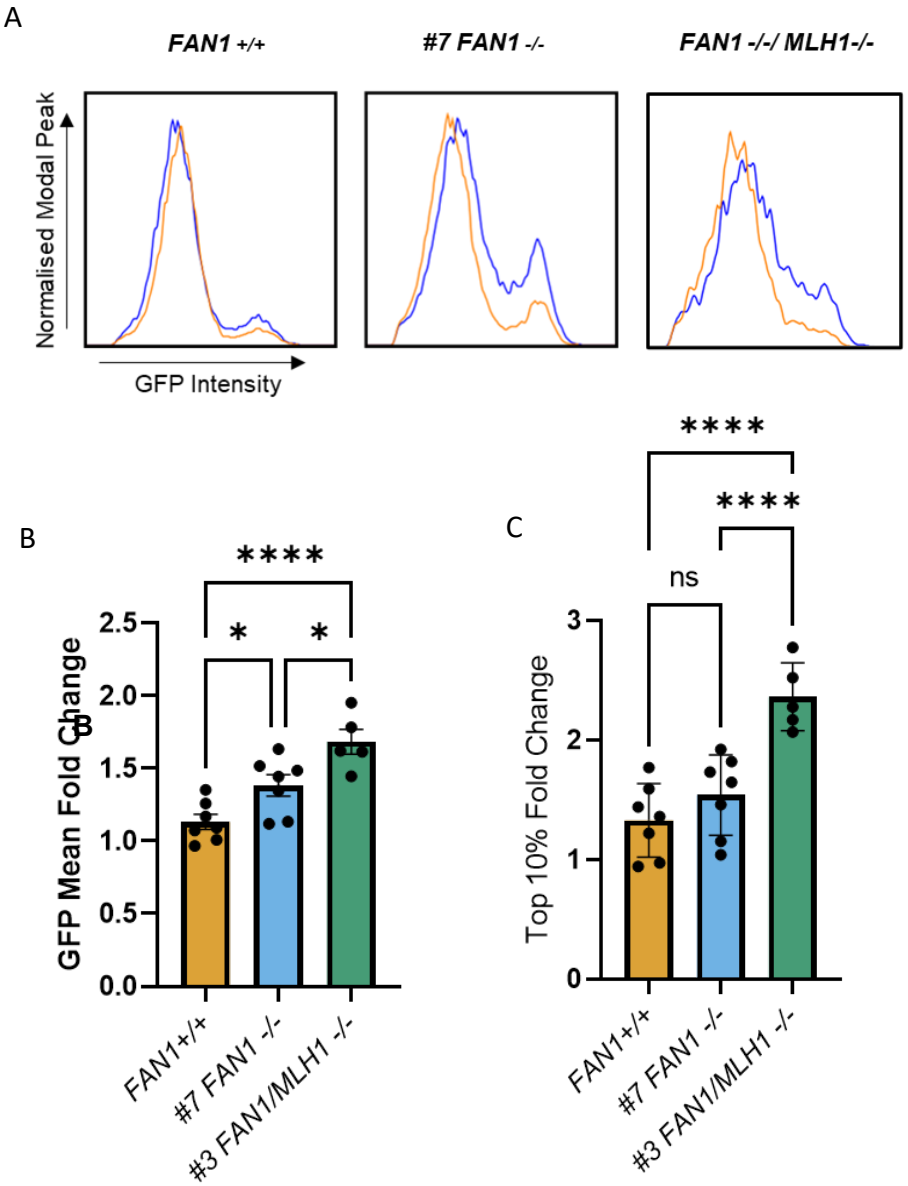


Figure 4-15: Investigating changes in GFP expression in FAN1-/-/MLH1-/- line and parent FAN1+/+ and FAN1-/- lines. ***(A)*** *Comparative GFP histogram profiles monitoring GFP expression over 7 (orange) and 21 (blue) days.* ***(B)*** *Fold change in mean GFP expression over 21 days.* ***(C)*** *Fold change in top 10% of GFP-positive cells over 21 days. (*P<0.05, ****P<0.0001. One-way ANOVA followed by Tukey's multiple comparisons). Error represents ± SD. Biological replicates for FAN1$^{+/+}$ and #7 FAN1$^{-/-}$ lines are those presented in Figure 3-7 and Figure 4-7.*

4.3.4.5 *FAN1$^{-/-}$/MLH1$^{-/-}$ GFPNickS120 line demonstrates increased nickase-induced contractions as demonstrated by sp-PCR*

As described in Chapter 3 spPCR was once considered the gold-standard method for analysis of trinucleotide repeats. The sequential dilution of an initial DNA sample concentration to only a few genomic equivalents, prevents bias against longer alleles that occur using bulk-PCR methods enabling the detection of a full spectrum of allelic lengths (Gomes-pereira et al. 2004).

DNA samples were collected from the *FAN1* and *MLH1* double knock-out line at day 0 and day 21 post-doxycycline treatment. As before, serial dilutions of DNA were first carried out to obtain an optimal concentration of amplifiable alleles. Concentrations ranging from 7-15 genomic equivalents were used for data analysis. For each sample, blinded allele counts were performed and at least 100 alleles counted from at least 2 membranes (Figure 4-16).

To investigate how the distribution of the repeat tract changed over the 21-day time course alleles were binned into categories based on molecular weight markers using a 1Kb DNA ladder. Comparisons between day 0 and day 21 was performed and revealed a significant difference in CAG repeat distribution after 21 days (*P=<0.0001*). A second analysis was carried out to quantify the fraction of contracted alleles after 21-days, compared to day 0. Importantly, the distributions of allele frequencies at day 0 were comparable between all lines (*P > 0.05*), however it is worth noting that in the double knock-out lines there appears to be slightly less heterogeneity at day 0. In this second analysis the most common allele representing the modal CAG length was selected, and alleles below this region were counted as contractions, respectively. Interestingly, there was a significant difference in contraction rates at day 21 between not only the *FAN1$^{-/-}$/MLH1$^{-/-}$* line relative to *FAN1$^{+/+}$* (*P=<0.0001*), but also the *FAN1$^{-/-}$* (*P=0.003*) parent line (Figure 4-17).

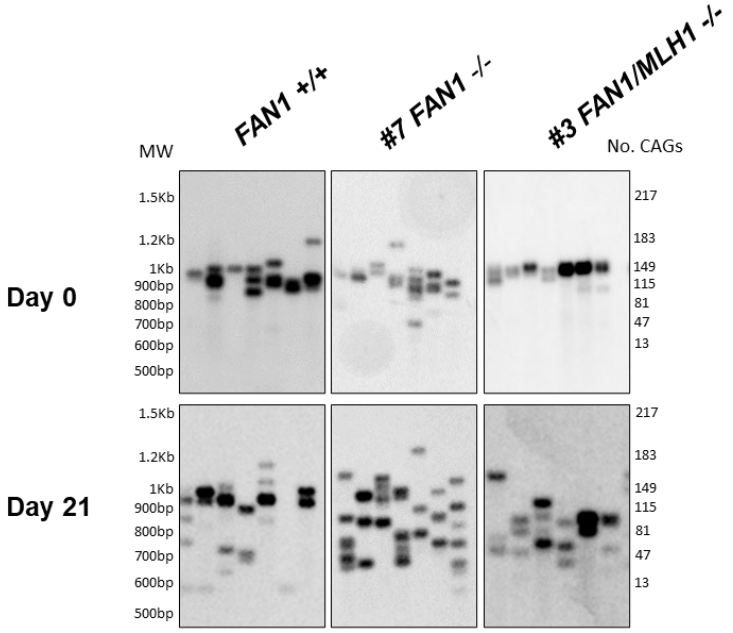
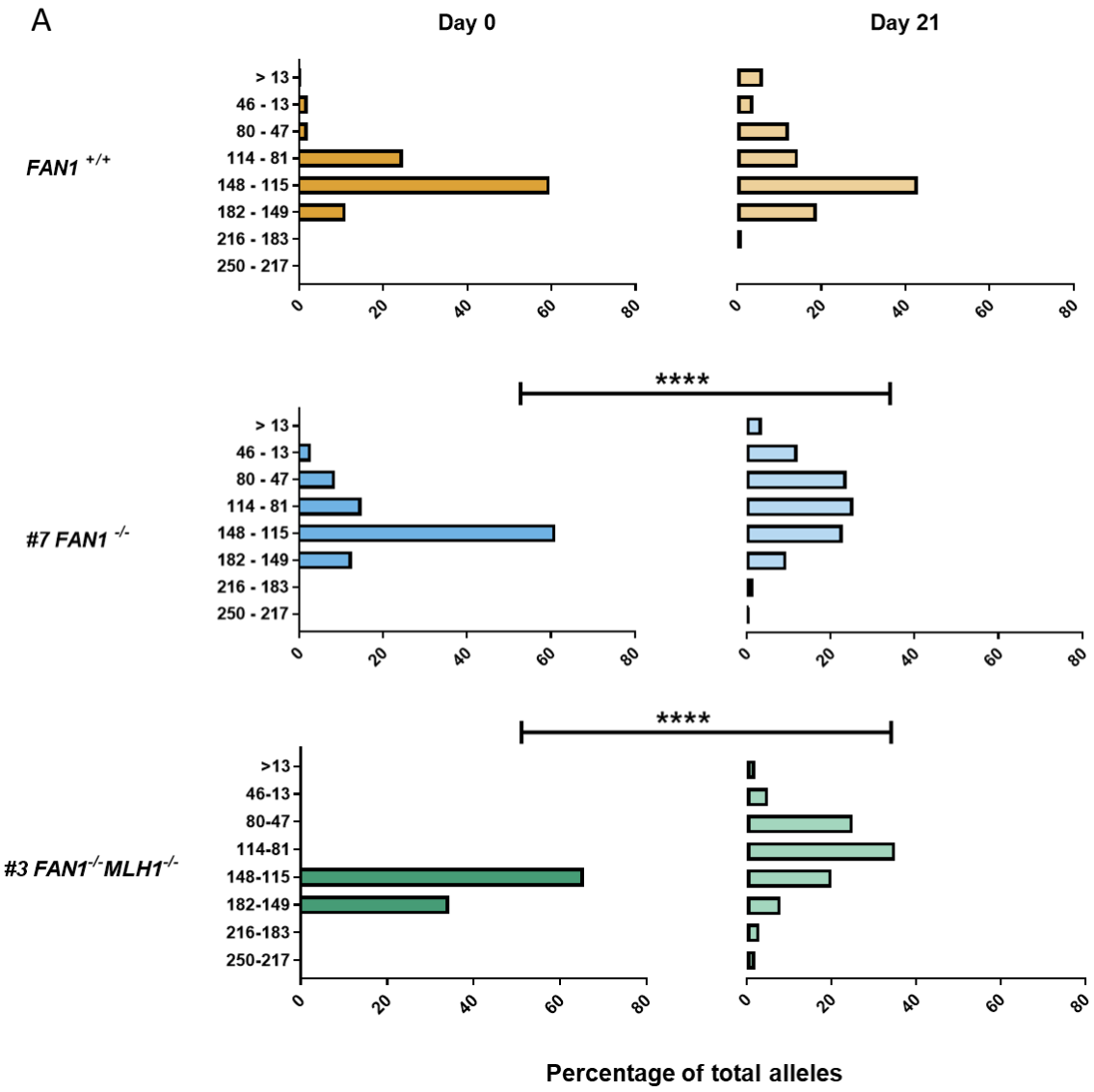


Figure 4-16: Demonstrative spPCR blot. *Example spPCR blots for FAN1$^{+/+}$, FAN1$^{-/-}$ and FAN1$^{-/-}$/MLH1$^{-/-}$ lines at day 0 and day 21. Images for FAN1$^{+/+}$ and #7 FAN1$^{-/-}$ are those in Figure 3-8.*


A
Day 0
Day 21
FAN1 +/+
#7 FAN1 -/-
#3 FAN1-/-MLH1-/-
> 13
46 - 13
80 - 47
114 - 81
148 - 115
182 - 149
216 - 183
250 - 217
>13
46-13
80-47
114-81
148-115
182-149
216-183
250-217
****
0
20
40
60
80
Percentage of total alleles

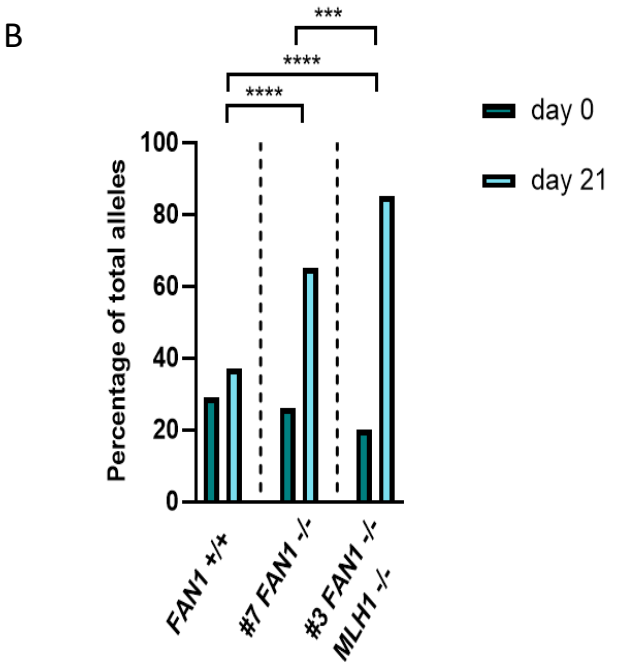

B
***
****
****
day 0
day 21
Percentage of total alleles
100
80
60
40
20
0
FAN1 +/+
#7 FAN1 -/-
#3 FAN1 -/- MLH1 -/-

Figure 4-17: SpPCR analysis of FAN1+/+, FAN1-/- and FAN1-/-/MLH1-/- lines to monitor Cas9 nickase-induced contractions. ***(A)*** *Graphic representation of spPCR analysis monitoring changes in CAG repeat distribution Comparative CAG repeat length profiles at day 0 and day 21 for FAN1$^{+/+}$, FAN1$^{-/-}$ and FAN1/MLH1$^{-/-}$ lines. ***P=<0.001, ****P=<0.0001 as determined by Mann-Whitney U comparing day 0 and day 21 distributions.* ***(B)*** *The FAN1$^{-/-}$/MLH1$^{-/-}$ clone demonstrate a significant increase in Cas9 nickase-induced contractions after 21-days relative to the FAN1$^{+/+}$ and FAN1-/- parent lines.***P=<0.001 ****P=<0.0001 as determined by two-sided Fisher's exact testing. Data from FAN1$^{+/+}$ and #7 FAN1$^{-/-}$ lines are those presented in Figure 3-9.*

4.3.4.6 FAN1$^{-/-}$/MLH1$^{-/-}$ GFPNickS120 cells show a comparable average CAG contraction compared with FAN1$^{-/-}$ parent line

The final mode of analysis used to compare whether a dual loss of FAN1 and MLH1 affects nickase-induced contractions rates, was long-read PacBio sequencing, as previously described. DNA samples were collected on day 0 and day 21 post-doxycycline treatment from the *FAN1/MLH1$^{-/-}$* line and compared with data from the *FAN1$^{+/+}$* and *FAN1$^{-/-}$* parent lines.

Repeat detector generated histogram reads for all samples and the mean CAG repeat was recorded and average change in CAG length calculated. The graph below indicates the size of the CAG contraction from day 0 to day 21. For all lines at least 4 biological replicates were analysed. There was a significant increase in CAG contractions between *FAN1$^{+/+}$* and the *FAN1/MLH1$^{-/-}$* line (*P=0.001*). However, there was no significant difference between the *FAN1$^{-/-}$* and *FAN1/MLH1$^{-/-}$* parent line. This was unexpected given the increase in GFP expression and contractions detected by spPCR, but taken together these data all indicate that MLH1 is not required to mediate the enhance nickase-induced contractions observed in *FAN1$^{-/-}$* lines.

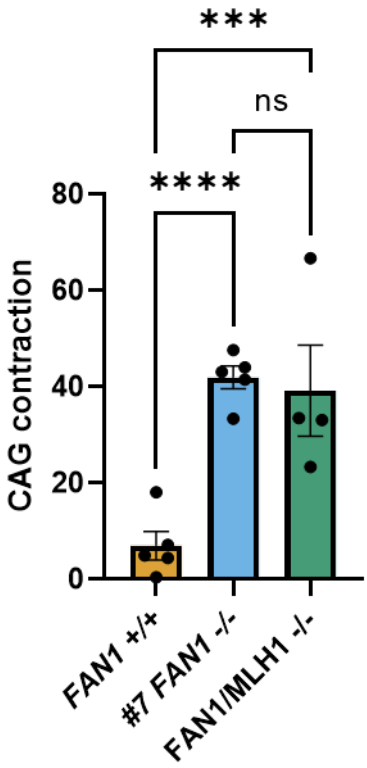


Figure 4-18: PacBio sequencing of GFPSNick120 cells to investigate whether FAN1/MLH1-/- GFPNickS120 cells demonstrate changes in nickase-induced contractions. *Bar chart depicting the average CAG contraction size from day 0 to day 21 days post Cas9 treatment. (*** P=<0.001, **** P=<0.0001, One-way ANOVA followed by Tukey's multiple comparison test). Error represents ± SD. Biological replicates for FAN1$^{+/+}$ and FAN1$^{-/-}$ are as presented in Figure 3-10 and Figure 4-8, (n=4-5).*

4.4 Discussion

This thesis aims to characterise the effect of FAN1 on CRISPR-Cas9 nickase-induced contractions. So far, literature investigating the role of FAN1 in the context of somatic CAG/CTG instability at expanded repeat tracts has implicated various functional domains of import. FAN1 protects against CAG/CTG expansion with genetic and functional data highlighting a range of roles for FAN1. The current school of thought is that FAN1 processes CAG/CTG loop out structures through either endo or exonucleolytic processing, leading to repair (Kim et al. 2020; Deshmukh et al. 2021b; McAllister et al. 2022; Phadte et al. 2023). However, the interaction between FAN1 and MLH1 is also key, where FAN1 competes with MMR for processing of CAG/CTG loop-out structures or additionally may sequester MLH1, preventing binding. This acts to protect against MMR-mediated repair of loop-out structures, which is error prone and leads to somatic expansions (Goold et al. 2019; Deshmukh et al. 2021a; Goold et al. 2021; Phadte et al. 2023). Whether or not FAN1 may impact our Cas9 nickase therapy has not been investigated.

4.4.1 Investigating the functional domains important for FAN1's protection against Cas9-nickase induced contractions

The first section of this chapter describes the generation of *FAN1*$^{-/-}$ GFPNicksS120 lines stably transduced with *FAN1* variant constructs, with the aim to explore how FAN1 may act to prevent nickase-induced contractions. The initial plan to generate these lines involved stable integration of a lentiviral construct followed by zeocin selection. However, zeocin selection resulted in cell cycle arrest and death in all lines with low MOI's. Only when transduced with an MOI of 10 did cells avoid cell death, and these lines demonstrated a significant increase of FAN1 protein level relative to endogenous expression and as such was not fit for these experiments. Through the use of an mCherry tag on a separate promoter positively transduced cells were able to be sorted for successful transduction and expanded to create stable lines. One limitation of this method meant that were no means of controlling for maintenance of construct expression over the time course. This did mean that over the period

there is likely a reduction in FAN1 construct expression to some degree in all lines, which can be detected via western blotting. Furthermore, western blotting did reveal variable protein expression between the lines, where ideally these would have been comparable. However, as demonstrated by downstream PacBio analysis *FAN1-/-^{WT}* reduced contraction rates similar to *FAN1$^{+/+}$* levels, demonstrating that FAN1 variant protein expression was sufficient for the purposes of these experiments. Though it is worth noting that too much FAN1 expression may lead to potential sequestration of MLH1, which could confound these data. However, given the data obtained from a dual *FAN1* and *MLH1* knock-out clone in GFPSNick120 cells, indicating MLH1 is not involved in the increased contractions observed in *FAN1$^{-/-}$* cells, this may not be an issue. It may be useful to further characterise these variant lines using qPCR targeting *FAN1* in genomic DNA, which will serve to inform us regarding the relative copy number of FAN1 variants within each cell line.

Once stable lines had been generated and validated, I sought to investigate which variants, if any, rescued the observed increased contraction phenotype in the *FAN1$^{-/-}$* clone #7 line. For all variant lines when comparing the top 10% of contracted cells there was no significant rescue to wild-type levels. However, this is likely due to the spread of data points in *FAN1$^{+/+}$* and *FAN1$^{-/-}$* lines. This highlights an important technical consideration for future experiments where run-to-run variation may impact findings, as such ideally separate biological replicates should be run on the same day to limit this as much as possible . Nevertheless, despite a lack of statistical significance, there was a trend towards recovery for all biological replicates in *FAN1$^{-/-WT}$*, *FAN1$^{-/-D960A}$* and *FAN1$^{-/-Endo}$* lines which demonstrated a reduction of GFP expression. Comparatively, *FAN1$^{-/-UBZ}$* and *FAN1$^{-/-R507H}$* demonstrated less recovery on average. However, when comparing GFP mean fold change as a measure, the *FAN1$^{-/-R507H}$* demonstrated a significant rescue to *FAN1$^{+/+}$* levels, with all other variant lines showing no significant change. This difference between the two modes of analysis was surprising, but could be explained by variation between cells counted for each sample, which may skew data. As such it was important to validate these findings using an additional assay.

As PacBio long-read sequencing presents a more robust and sensitive approach to measure CAG/CTG repeats, this was used to validate the GFP expression data (see summary in Table 4-1). For all variants, Repeat Detector generated histogram reads and the mean CAG repeat was recorded and the average CAG contraction from day 21 to day 0 was calculated. After 21 days only *FAN1*$^{-/-WT}$ cells were rescued to wild-type levels. This was encouraging, as when comparing protein expression between all variant lines, these cells had the lowest expression most comparable to endogenous levels. This suggests that the effect of each FAN1 variant in nickase-induced contractions is likely not due to differences in FAN1 protein expression, more likely the action of the variant itself. However, as the *FAN1*$^{-/-WT}$ lines used for these data do demonstrate a loss of FAN1 expression during the 21-day time period, due to our inability to select for the variant, it would be beneficial to utilise another batch of cells with an initial higher expression more comparable to other variant lines. Furthermore, stocks of all variant lines were frozen down and present a pool of cells where single-cell isolates can be generated in the future to screen for more comparable expression of variants.

The most significant finding from the PacBio dataset was that *FAN1*$^{-/-R507H}$ cells failed to reduce nickase-induced contractions to *FAN1*$^{+/+}$ and *FAN1*$^{-/-WT}$ levels. To date, genome, exome-wide studies and phenome association studies have implicated this variant in autism, schizophrenia, KIN and as a HD-age of onset modifier (Ionita-Laza et al. 2014; Lee et al. 2015; Bastarache et al. 2019; Lee et al. 2019). However, the functional role this variant may play in CAG repeat instability, or indeed in FAN1's role as a DNA damage repair protein, is unclear. Whilst the genetic data indicate that this amino acid change may be deleterious for FAN1 function, likely through reduced DNA binding, *in vitro* reports have been conflicting. One report utilising nuclear extracts obtained from *FAN1*$^{-/-}$ HEK293T cells overexpressing this variant demonstrated reduced binding to CAG loop-out structures (Kim et al. 2020). A separate study examined the effect of partially purified FAN1 proteins on a 5' DNA flap structure, which can be processed by FAN1 through its endo- or exonuclease activity (see Deshmukh et al. 2021 for review). Interestingly they noted a lack of nucleolytic processing, comparable with a nuclease-dead FAN1. However,

this is likely attributed to a lack of DNA binding, rather than a defect in nucleolytic processing (McAllister et al. 2022).

A recent publication has examined the effect of purified R507H FAN1 protein on a range of DNA substrates; including canonical 5' flap structures and CAG, CTG, CGG and CCG loop out structures (Deshmukh et al. 2021b). In contrast to the previous two studies they report no effect of this variant on DNA binding, as detected by an electrophoretic shift assay. Furthermore, they observed no significant difference in endo- or exonuclease processing on any structures, indicating that, in this system, the R507H variant does not result in a loss of function (Deshmukh et al. 2021b). The differences in these observations may be attributed to technical differences. Whilst the use of biochemical based assays to investigate the roles of DNA repair proteins *in vitro* is valuable, this 'stripped-back' approach is limited by the lack of proteins and other endogenous factors that likely influence FAN1 function. It may be that a highly purified prep, where no competition for DNA binding exists, masks any phenotypic differences. This could explain the presence of functional deficits reported in the other studies. For example, Kim *et al.* 2020 used an indirect pull-down assay when assessing FAN1 DNA-binding activity, it is therefore plausible that other FAN1-interacting protein partners may form part of this prep, thus interfering with R507H binding to DNA substrates. It is clear that a further exploration into the functional consequences of this variant is required, ideally using a whole-cell system approach. On such study in a mammalian cell system does support the findings of Deshmukh and colleagues (Goold et al. 2019). In this system *FAN1*$^{-/-}$ U2OS cells reconstituted with a GFP-tagged R507H FAN1 variant, reduced somatic CAG expansions to *FAN1*$^{+/+}$ levels in an ectopic *HTT* exon 1 construct, perhaps suggesting no functional deficits. Indeed, this study demonstrates cells expressing this variant do not confer resistance to MMC. However, in these cells this variant was vastly overexpressed, relative to endogenous FAN1 expression levels. Therefore the reduction of somatic expansion may be explained by increased FAN1 protein levels resulting in increased MLH1 sequestration, limiting the MMR machinery.

On the role of the R507H variant in our CRISPR-Cas9 system, a failure to rescue nickase-induced contractions to *FAN1*$^{+/+}$ and *FAN1*$^{-/-}$ ^{WT} levels,

suggests a functional deficit in this variant. An interpretation of these findings may be, firstly; the R507H variant results in decreased or no DNA-binding to DNA substrates present at sites of Cas9-induced DNA nicks. Secondly; this lack of DNA binding likely results in reduced or no nucleolytic processing by FAN1 at these sites, leaving processing of these structures available to other DNA repair proteins. The partial recovery I see in both *FAN1*$^{-/-\ ENDO}$ and *FAN1*$^{-/-\ D960A}$ lines support this hypothesis, as whilst catalytically impaired, these variants are still able to bind DNA substrates (Kratz et al. 2010; Zhao et al. 2014). The presence of bound FAN1 protein to these structures may reduce binding by other proteins, protecting against differential processing that may lead to contraction events, thus I observe a partial rescue. To further support these findings it would be useful to confirm functional deficits in DNA repair processes in these lines. Cells expressing mutant either D960A and R507H FAN1 demonstrate increased sensitivity to cross-linking agent MMC (Liu et al. 2010b; Kim et al. 2020; McAllister et al. 2022). Therefore, performing a clonogenic assay similar to that described in Chapter 3 would be useful to further validate these lines behave as expected.

The final variant investigated in this chapter concerned a C44A/C47A mutation in the UBZ domain of FAN1. Reports have demonstrated that this mutation results in a failure of FAN1 to co-localise with FANCD2 to DNA-damage induced foci (Kratz et al. 2010; MacKay et al. 2010; Smogorzewska et al. 2010). This interaction between FAN1 and monoubiquitinated FANCD2 allows for the recruitment of FAN1 to ICL-induced DNA damage where it can function in association with members of the FA pathway as a means of ICL repair (Jin and Cho 2017). To date, there is no literature investigating the importance of this domain in the context of CAG/CTG instability. One report has demonstrated that a loss of FANCD2 has no effect on CGG repeat expansion in a mouse model of Fragile-X-syndrome, suggesting the interaction between FAN1 and FANCD2 is not required for FAN1 recruitment to CCG/CGG secondary structures (Zhao et al. 2021).

Surprisingly, in our system, *FAN1*$^{-/-UBZ}$ cells failed to reduce nickase-induced contraction rates to *FAN1*$^{+/+}$ levels, however this may be considered a partial rescue, similarly with *FAN1*$^{-/-\ ENDO}$ and *FAN1*$^{-/-\ D960A}$ lines, as contractions rates

were not significantly different from *FAN1*$^{-/-\ WT}$ lines. Furthermore, when comparing relative expression of FAN1 protein levels between variants, this mutant consistently demonstrated higher expression levels compared with other variants, indicating that a lack of rescue was not due to insufficient expression levels. As this mutant has not been implicated in DNA repair at CAG/CTG repeats, I can only speculate as to a potential role for this domain in preventing nickase-induced contractions. Ubiquitin-mediated interactions are essential for the cellular DNA damage response, as such many DNA-damage repair proteins contain UBZ domains (Rizzo et al. 2014). Whilst one study has demonstrated in a FXD mouse model that *Fancd2*$^{-/-}$ do not demonstrate different somatic expansion rates in the brain compared with *Fancd2*$^{+/+}$, it is plausible that other interactions between FAN1 and ubiquitinated proteins may be important in localising FAN1 to CAG/CTG secondary structures. However, it is also possible that the DNA structures that form as a result of Cas9-induced DNA nicks are distinct from those normally present in un-nicked CAG/CTG repeats, therefore I cannot rule out entirely whether this interaction between FAN1 and FANCD2 is important in our system. This poses an interesting question as to whether the process of DNA repair that results in nickase-induced contractions is distinct from that of expansions or contractions that occur as a result of increased somatic instability.

Genotype	Cas9 Nickase	Somatic Instability
FAN1$^{+/+}$	*contractions	***expansions
FAN1$^{-/-}$	contractions	expansions
FAN1$^{-/-}$/*MLH1*$^{-/-}$	contractions	stable
FAN1 variant	**Cas9 Nickase**	**Somatic Instability**
Wild type	contractions	expansions
UBZ mutant	**contractions	No known impact
Endonuclease dead	**contractions	Not investigated
Nuclease dead	**Nuclease dead	expansions
R057H	R507H	****Increased HD AAO

Table 4-1: Table summarizing the amplicon PacBio data monitoring Cas9 nickase-induced contractions and current known implications of these genotypes or variants in somatic instability. **contractions still present but less than in FAN1$^{-/-}$ lines **partial reduction in contractions, but not to FAN1$^{+/+}$ levels . ***somatic expansions present, but reduced compared with FAN1$^{-/-}$ lines. **** R507H variant shows early age of onset as predicted by GWA studies, suggesting this variant may not protect against somatic expansion but data from functional studies is mixed. Green arrows = low, yellow arrows = moderate and red arrows = high levels of either contractions or expansions.*

4.4.2 Exploring whether MLH1 contributes to increased Cas9 nickase - induced contractions

Next, I sought to determine whether the enhanced frequencies of nickase-induced contractions seen in the *FAN1*$^{-/-}$ cells depend on MLH1. This was motivated by the work in somatic expansions where deleting *Mlh1* in a *Fan1* knock-out background abolished expansions. The prediction was that the double knock-out would be more stable that the *FAN1* knock-out, with contraction rates reduced to *FAN1*$^{+/+}$ levels. Firstly, to confirm somatic instability in HEK293 cells mirrors what has been demonstrated in previous literature, I generated three *FAN1* and *MLH1* double knock-out lines in GFP(CAG101) cells that do not express the Cas9 nickase. As expected in all three double knock-out clones I saw a complete loss of somatic expansion

over the time-course, detected by long-read PacBio sequencing. This was in agreement with the current literature demonstrating that a loss of MLH1 in a *FAN1*$^{-/-}$ background ablates somatic expansion (Loupe et al. 2020). This confirms that our HEK293-reporter cell line acts as a reliable model to monitor changes in somatic CAG/CTG expansion, where MMR is essential for somatic expansion.

By contrast, in GFPNickS120 *FAN1*$^{-/-}$*MLH1*$^{-/-}$ double knock-outs, the frequency of nickase-induced contractions was similar to that of the single *FAN1*$^{-/-}$ line, as detected by long-read PacBio sequencing. These results suggest that MLH1 is not required for nickase-induced contractions. This is consistent with a previous publication demonstrating that knockdown of MSH2 did not affect the levels of nickase-induced contractions (Cinesi et al. 2016). However, to date, only one double knock-out clone has been generated and no single *MLH1* knock-out clones, ideally a future aim of this project would be to confirm any phenotypic findings in additional generated clones. Moreover, initial screening of nickase-induced contractions, using GFP expression as a proxy indicated that a loss of *FAN1* and *MLH1* may increase contraction rates. This was observed in both means of analysis where the *FAN1*$^{-/-}$*MLH1*$^{-/-}$ line had significant increased GFP expression over time relative to both *FAN1*$^{+/+}$ and *FAN1*$^{-/-}$ parent lines. This apparent increase in contractions in the double knock-out clone was further supported by small-pool PCR analysis. Although the small pool PCR showed a higher level of instability in the double knockout, this was a single biological replicate that was also high in the PacBio dataset. The other advantage of the long-read sequencing approach is that it is high-throughput and more samples can be analysed, compared with small-pool PCR which is more laborious. Therefore, within the PacBio analysis it is apparent that this single biological replicate demonstrated higher contraction rates than the others. Overall, the *FAN1/MLH1* double knock-out has at least as many contractions as the *FAN1* single knock-out, based on the more sensitive long-read sequencing data. Future work to produce additional clones will act to confirm these findings.

The data gathered so far in Chapter 3 and supported by findings in this chapter indicate that FAN1 is important in protecting against CRISPR-Cas9 nickase

induced contractions. Specifically, the results presented in this chapter indicate that the DNA-binding and UBZ domains of FAN1 are important for this mechanism, with the nuclease activity only partially required. Furthermore, I demonstrate that this increase in contraction rates in *FAN1*$^{-/-}$ cells is not dependent on MLH1 activity, suggesting that FAN1's protective role in nickase-induced contractions is distinct from that of somatic expansion.

4.4.3 Chapter Summary

- Rescue of *FAN1*$^{-/-}$ cells with *FAN1* R507H variant in the DNA-binding domain fails to reduce nickase-induced contractions to *FAN1*$^{+/+}$ and *FAN1*$^{-/-WT}$ levels
- Rescue of *FAN1*$^{-/-}$ cells with *FAN1* C44A/C47A UBZ variant, a nuclease-dead or endonuclease dead variants partially reduce nickase-induced contraction rates to *FAN1*$^{+/+}$ *levels*
- Somatic expansion in a HEK293-derived reporter cell line requires MLH1 and MMR machinery
- Increased nickase-induced contraction rates in *FAN1*$^{-/-}$ lines is not dependent on MLH1

5 Investigating the role of the CRISPR-Cas9 nickase in HD-patient iPSCs and iPSC-derived neurons

5.1 Introduction

Stem cells are a unique set of unspecialised cells capable of continuous self-renewal and the ability to differentiate to any cell type (Robinton and Daley 2012). As early as the 1980s embryonic stem cells (ESCs) from mouse embryos were isolated (Evans, M J and Kaufmant 1981; Martin 1981), though isolation of human ESCs (hESCs) was not successful until much later (Thomson et al. 1998). hESCs are capable of differentiation into all three embryonic germ layers; endoderm, mesoderm and ectoderm. This discovery opened up a new avenue for modelling human diseases with early models demonstrating these cells can recapitulate disease phenotypes (Urbach et al. 2004). However, ethical concerns and the relative limited availability of these cells can present challenges to their use as a model system.

In 2006, the discovery that mouse embryonic and adult fibroblast cells are capable of transformation to induced pluripotent stem cells (iPSCs), through reprogramming with four transcription factors, presented iPSCs as an alternative model to ESCs (Takahashi and Yamanaka 2006). iPSCs, like ESCs, can be expanded indefinitely and are similar to ESCs in morphology, gene expression and epigenetic status of pluripotent cell-specific genes. Importantly, iPSCs retain the ability to differentiate into any cell type from the three germ cell layers, allowing for modelling of phenotypes in disease-specific cells (Takahashi and Yamanaka 2006; Takahashi et al. 2007; Yu et al. 2007). Furthermore, a key advantage of iPSCs is the ability to collect cells directly from patients which retain the genetic information of the patient, allowing for the investigation of how genetic factors may influence disease (Gan 2014).

To date, iPSCs derived from patients have been utilised to model many diseases, including Alzheimer's disease (AD) (Yagi et al. 2011; Nieweg et al. 2015; Liao et al. 2016), Parkinson’s disease (PD) (Ambasudhan et al. 2013; Brazdis et al. 2020; Avazzadeh et al. 2021) and HD (see Kaye et al., 2022 for

review). In these models, as neurons are often the cell-type of interest, many studies have been published detailing the programmed differentiation of iPSCs to neuronal cultures. These include differentiation to cortical neurons (Shi et al., 2012), dopaminergic (Hartfield et al. 2014) and GABAergic neurons (Sun et al. 2016), motor neurons (Nizzardo et al. 2010) and medium spiny neurons (MSNs) (Arber et al. 2015; Telezhkin et al. 2016), to name a few. A common protocol involves the step-wise addition of exogenous small molecules and growth factors, aimed to recapitulate the sequential process of neurogenesis and synaptogenesis (Hulme et al., 2022). In such cases, an initial culture media containing small molecules aimed at inhibiting pathways essential for iPSC maintenance, such as the SMAD signalling pathway, are used to direct iPSCs to a neural precursor (NPC) fate (Arber et al., 2015; Chambers et al., 2009, 2012; Chavali et al., 2020; Shi et al., 2012). Following NPC production treatment of cells with a second media supplemented with alternative growth factors and small molecules acts to drive neuronal direction and maturation to a particular sub-type (Hulme et al., 2022).

iPSCs reprogrammed from HD-patient fibroblasts provide a useful tool to model HD pathogenesis. To date, a number of cell lines have been generated from HD patients, in an effort to model disease and screen for potential therapeutic targets (Table 5-1). Reports from a number of these studies have demonstrated HD-disease phenotypes compared with controls, including; altered transcriptomics, altered neuronal morphology, defects in nuclear pore transport and altered bioenergetics (see Kaye et al., 2022 for review).

iPSC line	CAG repeat size	Reference
CS97iHD-180nX	180	(Mattis et al., 2012)
CS109iHD-109nX	109	(Mattis et al., 2012)
CS77iHD-77nX	77	(Mehta et al., 2018)
HD76	76	(Vigont et al., 2021)
HD70	70	(Tidball et al., 2016)
CS21iHD-60nX	60	(Mattis et al., 2012)
CS039HD-53nX	53	(Lim et al., 2017)
iPSHD11,iPSHD22, iPSHD34	40, 47,42	(Nekrasov et al., 2016)
CS04iHD-46nX	46	(Lim et al., 2017)
HD1, HD2	44, 42	(Mollica et al., 2018)

Table 5-1: HD-patient iPSC models. *(Adapted from Kaye et al., 2022).*

This chapter aims to investigate whether the phenotype of increased nickase-induced contractions observed *FAN1*$^{-/-}$ cells in a HEK293-reporter cell line is recapitulated in HD-patient iPSCs and iPSC-derived neurons. Importantly, although the HEK cells are a useful model the CAG tract investigated is under an inducible ectopic system, and crucially these cells replicate, therefore it is important to validate these findings in non-dividing neurons (Santillan et al. 2014; Cinesi et al. 2016). To do so subcloned CS09iHD-109nX (HD-iPSCs) iPSCs, initially generated and validated by the HD-iPSC Consortium, were used (Mattis et al. 2012). This study described the generation and validation of 14 iPSC lines derived from HD patient and healthy control fibroblasts. Importantly, the CS09iHD-109n1 subclone was reported to be able to successfully differentiate to MSNs, the major cell type affected in HD. Furthermore, many studies have not reported or do not demonstrate *HTT* somatic CAG expansion of these lines over time, this line does (Mattis et al. 2012; McAllister et al. 2022). The ability of this cell line to undergo somatic expansion presents a relevant *in vitro* model to compare effects of somatic expansion and effects of nickase-induced contractions.

5.1.1 Chapter aims

The aims of this chapter were to generate *FAN1*$^{+/+}$ and *FAN1*$^{-/-}$ HD-iPSCs stably expressing the Cas9 nickase treatment, through the use of lentiviral transduction with a vector expressing Cas9 with a blasticidin selectable marker. After the generation of such lines, downstream aims were to establish whether loss of FAN1 increased nickase-induced contraction rates in iPSCs and iPSC-derived neurons, in agreement with the data shown in Chapter 3.

5.2 Results

5.2.1 Attempting to generate HD-iPSCs stably expressing Cas9 nickase

5.2.1.1 Confirmation of HTT CAG repeat sizing in HD109 iPSCs

HD-iPSCs (N1 subclone), as described in McAllister et al., 2022, were obtained and *FAN1*$^{+/+}$ and *FAN1*$^{-/-}$ lines were screened for *HTT* CAG repeat size prior to downstream experiments. A previously optimised PCR was utilised amplifying exon 1 of *HTT* and the resultant PCR product was analysed using capillary electrophoresis with a fluorescent ladder to accurately size repeats. Both lines had an initial CAG repeat length of 125 CAGs and when grown in culture demonstrated typical iPSC morphology (Figure 5-1).

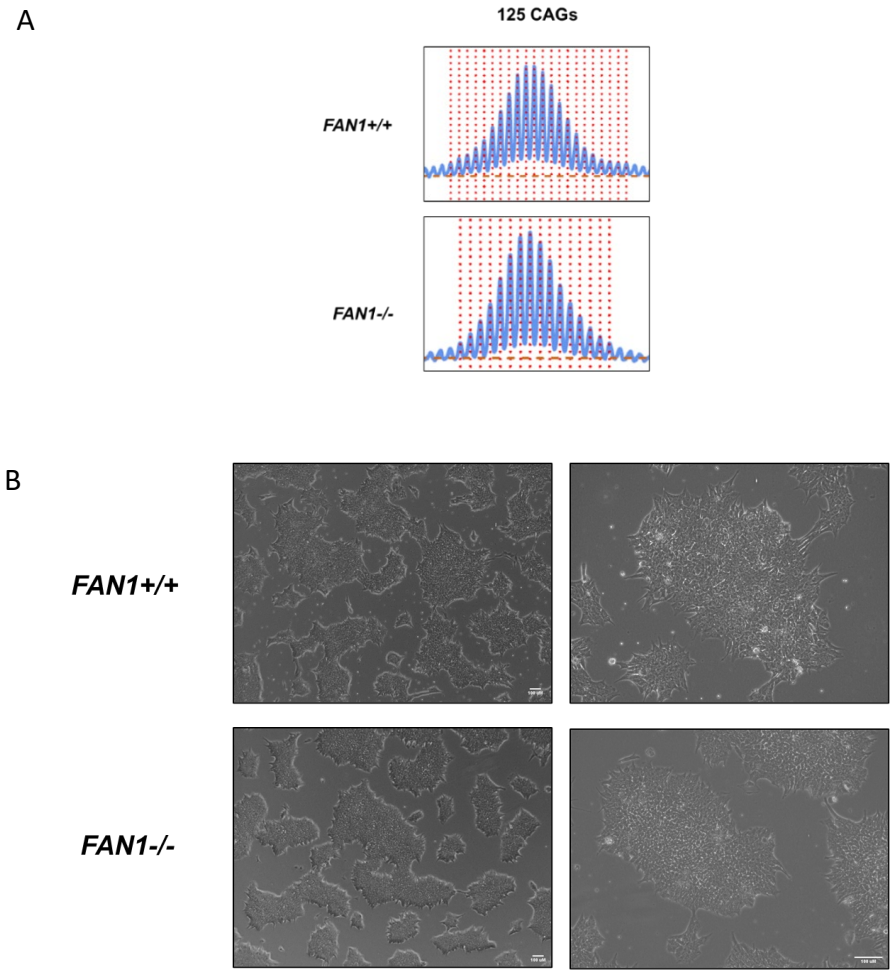
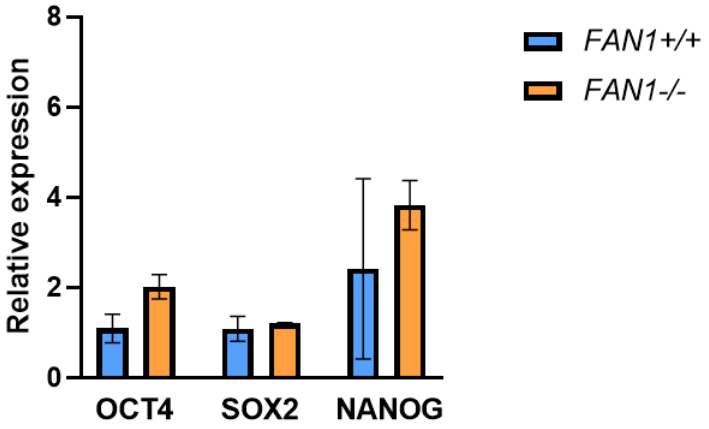


Figure 5-1 Characterisation of HD-IPSC. ***(A)*** *Representative electropherogram plots representing HTT CAG repeat sizing in FAN1*$^{+/+}$ *and FAN1*$^{-/-}$ *cells.* ***(B)*** *Brightfield images of iPSCs colonies in culture at 4x (left) and 10x (right) magnification. Scale = 100 um.*

5.2.1.2 Confirmation of pluripotency in HD FAN1$^{+/+}$ and FAN1$^{-/-}$ iPSCs

After CAG repeat sizing, cells were expanded and examined for expression of pluripotency markers. These lines have been characterised previously and demonstrated expression of pluripotency markers. It was important to reestablish this pluripotent potential prior to downstream experiments, as variability in expression of these markers could impact downstream differentiation to neuronal cell types of interest. Cell pellets were taken routinely over culture time and RNA extracted for RT-qPCR. Both lines expressed *OCT4*, *SOX2* and *NANOG* as expected, there was no significant in difference in expression between the two lines (Figure 5-2). Furthermore, both lines demonstrated specific nuclear immunostaining of all three pluripotency markers (Figure 5-2 and Figure 5-3).

A



B

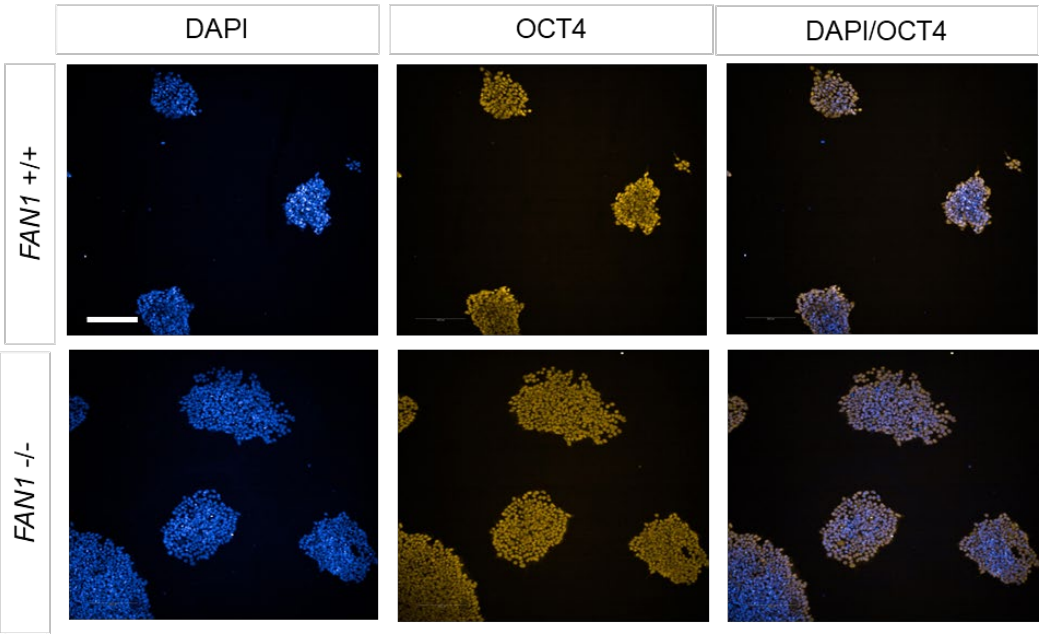


***Figure 5-2: Pluripotency marker expression in HD-iPSCs 1**: **(A)** mRNA expression levels of OCT, SOX2 and NANOG are comparable between FAN1$^{+/+}$ and FAN1$^{-/-}$ lines (n=3). One-way ANOVA followed by Tukey's multiple comparison testing. **(B)** Undifferentiated iPSCs demonstrate nuclear staining of OCT4. Scale bar =200 µm.*

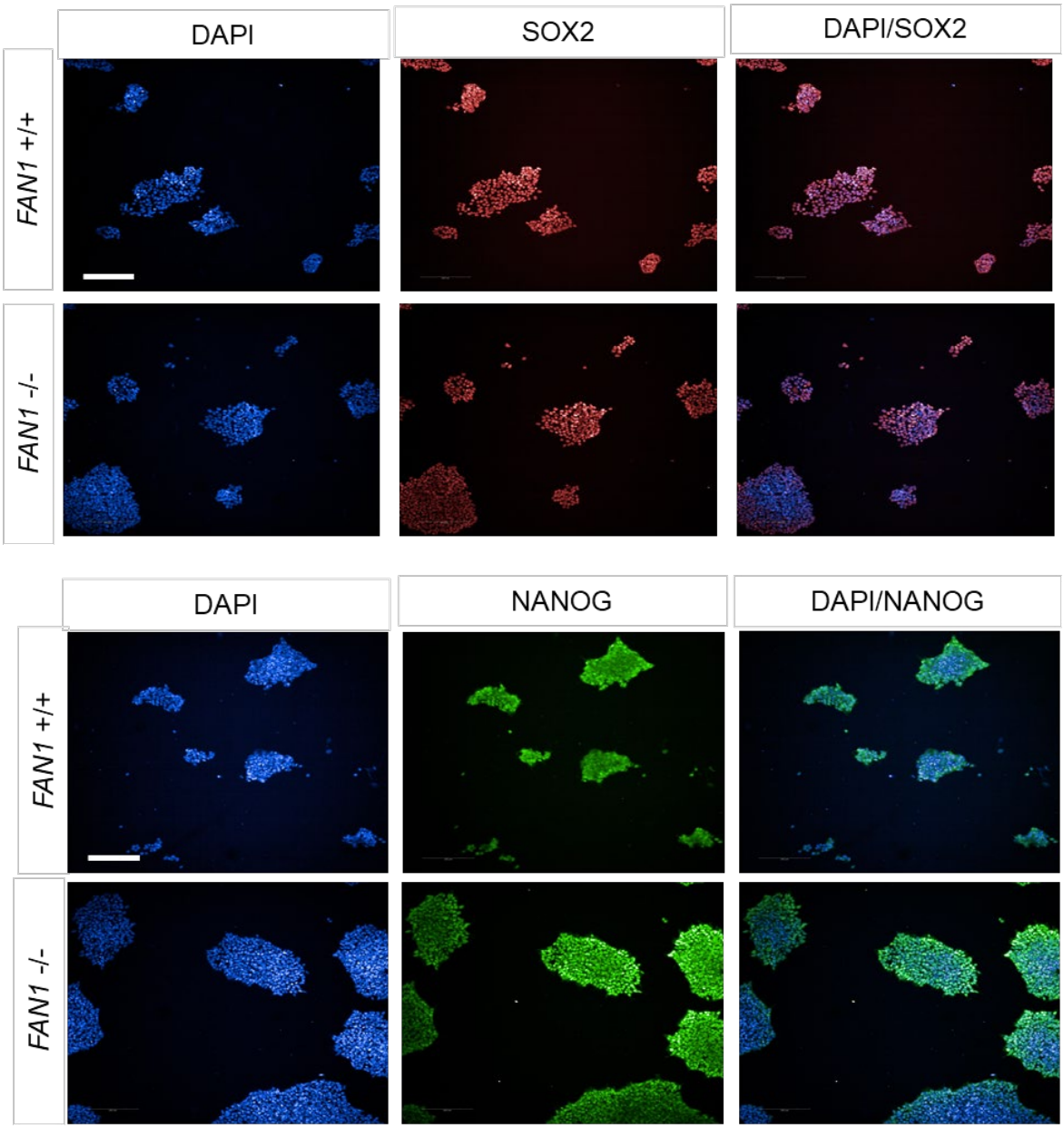


Figure 5-3: Pluripotency marker expression in HD-iPSCs 2: *Undifferentiated iPSCs demonstrate nuclear staining of SOX2 and NANOG in both FAN1$^{+/+}$ and FAN1$^{-/-}$ lines. Scale bar = 200 µm.*

5.2.1.3 EF1-α promoter demonstrates best expression of lentiviral construct relative to CMV in HD109 iPSCs

Prior to transduction of iPSCs with a lentiviral construct containing our Cas9 nickase treatment, it was important to establish whether HD-patient iPSCs would express our construct as desired. This was important as differences in promoter efficiencies have been reported for ESCs and iPSCs (Minoguchi and Iba 2008; Hoffmann et al. 2017).

I employed the use of a control vector obtained from VectorBuilder containing a mCherry fluorescent tag under control a EF1-α promoter and a GFP fluorescent tag under control of a CMV promoter (Figure 5-4). The use of fluorescent tags allowed for screening of successful transduction over time and for comparison of best expression between the two promoters. Prior to transduction 400,000 cells were plated in E8 Flex media containing 10 μM ROCK inhibitor and left to adhere overnight. The following morning cells were transduced with the control vector construct in media containing 5 μg/mL polybrene, which has been demonstrated to improve transduction efficiency (Li and Lu 2009), in some cell types and conditions. MOI's of 5 and 10 were used for transduction. After 24 hours media was changed and cells were then treated with 1 μg/mL of puromycin to select cells transduced with the lentiviral construct.

Cells were maintained in puromycin selection for 14 days and mCherry or GFP expression monitored routinely (Figure 5-5 and Figure 5-6). Whilst there was clear expression of both fluorescent markers the expression of mCherry was higher with reduced background. This suggested that EF1-α was the best promoter for transduction with these iPSCs for downstream treatment with the Cas9 nickase.

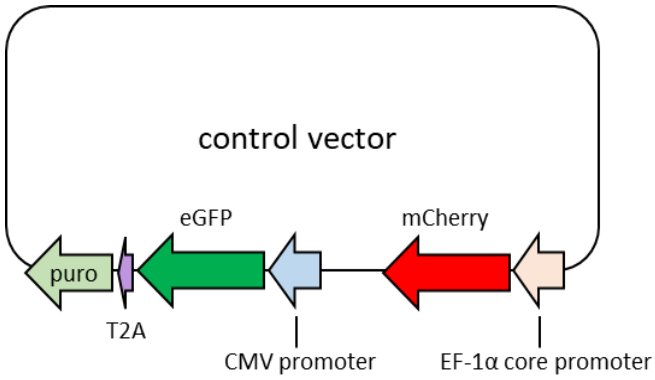


Figure 5-4: Control vector containing mCherry and eGFP markers. *Illustration of control lentiviral construct with both mCherry and eGFP tags under the control of EF1-α and CMV promoters.*

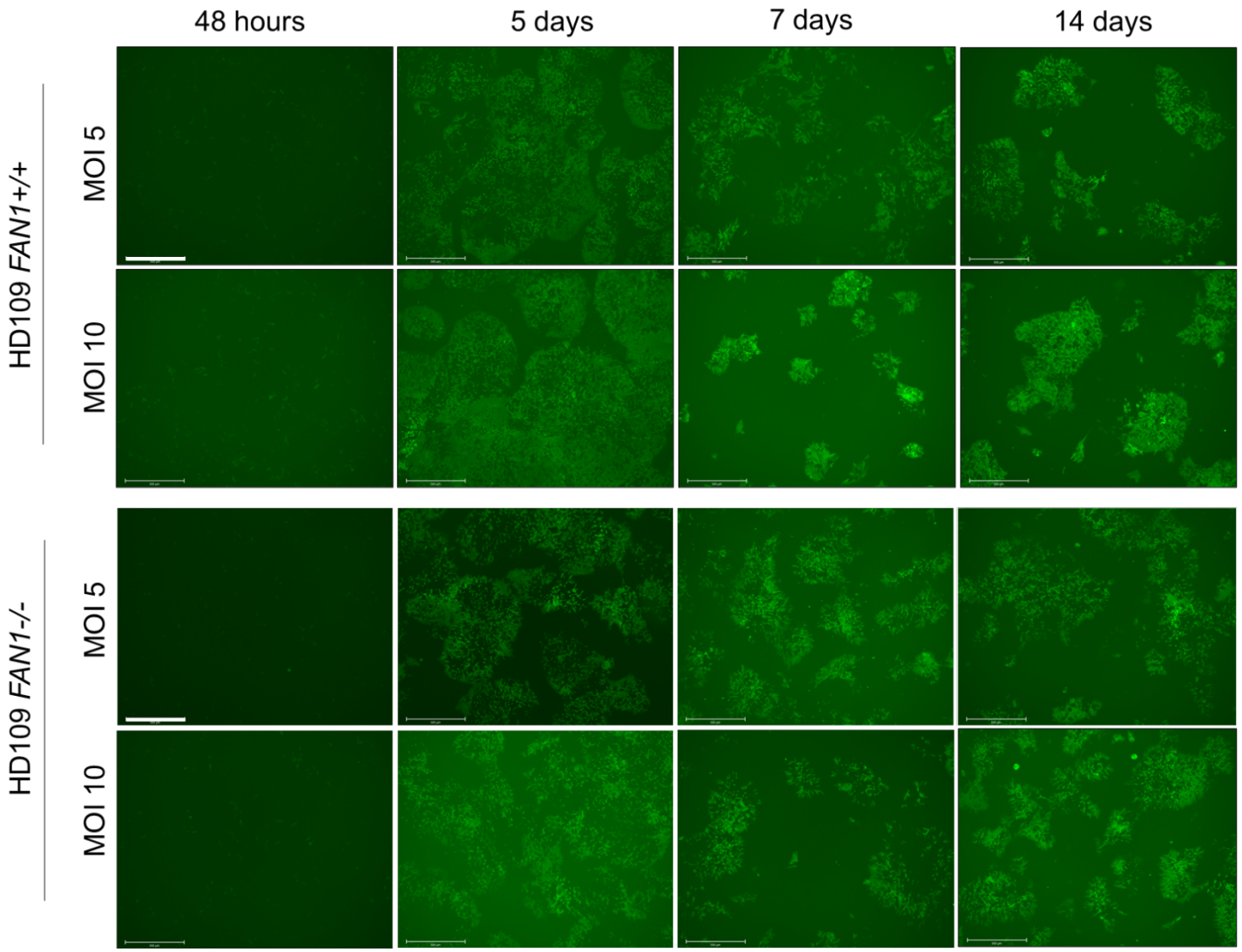


Figure 5-5: GFP expression in transduced HD109 FAN1$^{+/+}$ and FAN1$^{-/-}$ iPSCs over 14 days.
Scale bar = 500 um.

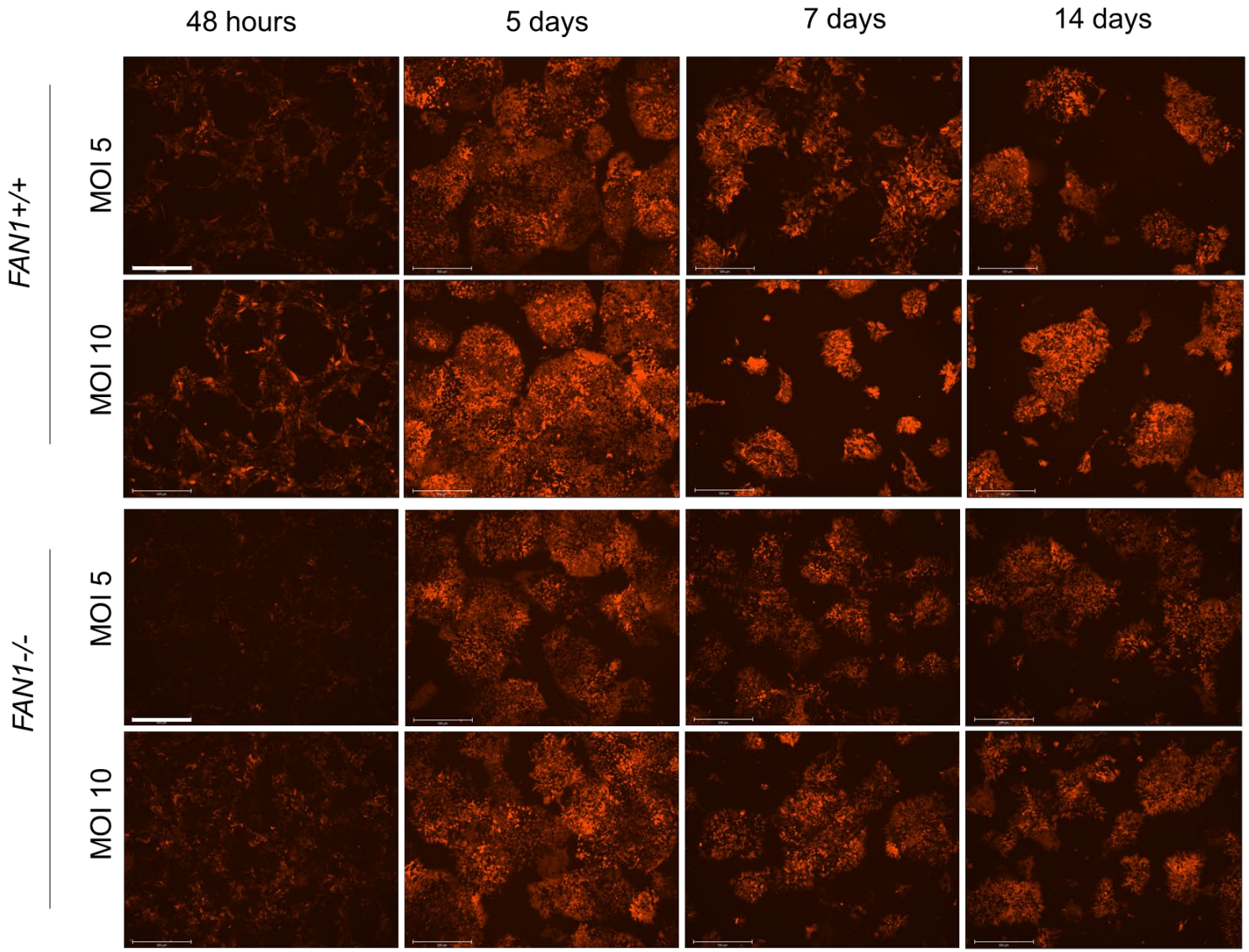


Figure 5-6: mCherry expression in transduced HD109 FAN1$^{+/+}$ and FAN1$^{-/-}$ iPSCs over 14 days. *Scale bar = 500 um.*

5.2.2 Cas9 is selectively silenced in HD-iPSCs

After confirming that the EF1-α promoter is expressed in HD-patient iPSCs, cells were transduced with a lentiviral vector expressing a human-codon optimised *S. pyogenes* D10A Cas9 nickase (Addgene #63593) (Figure 5-7). Transduction conditions were as described in section 5.2.1.3. 24 hours after transduction lentiviral media was removed and cells were placed under selection in 5 µg/mL blasticidin for 5 to 7 days. After selection, colonies were obtained from single cells with the aim to expand and screen clones for Cas9 integration and expression (Figure 5-7). During the expansion phase a bulk population of cells from both *FAN1*$^{+/+}$ and *FAN1*$^{-/-}$ lines were frozen down and cell pellets collected to confirm Cas9 integration.

PCR primers designed amplify a 120 bp region of the Cas9 vector, confirmed successful integration of DNA in both wild-type and knock-out iPSCs lines for both MOI 5 and 10 transduction conditions (Figure 5-8). A qPCR was carried out to quantify relative copy number integration relative to β- actin, in both conditions copy number was lower in *FAN1*$^{-/-}$ cells relative to wild-type (Figure 5-8). Based on these results MOI 5 appeared to be the best transduction condition, with most comparable expression in bulk populations, which helped to inform the number of single cells plated downstream for clonal expansion. After confirming successful integration bulk samples were screened for Cas9 expression via western blotting. Interestingly, despite confirmation of successful vector integration and blasticidin resistance cells failed to express Cas9 1 or 2 weeks post-transduction (Figure 5-8). To confirm this was not due to a problem with the lentiviral batch control HEK293 cells that do not express Cas9 were transduced with the same lentivirus at MOI's of 5 and 10, placed under blasticidin selection and a cell pellet taken at 1 week post transduction. At both MOI's HEK293 cells demonstrated clear expression of Cas9, indicating that this was not a lentiviral batch issue and in HD-patient iPSCs Cas9 is selectively silenced (Figure 5-8).

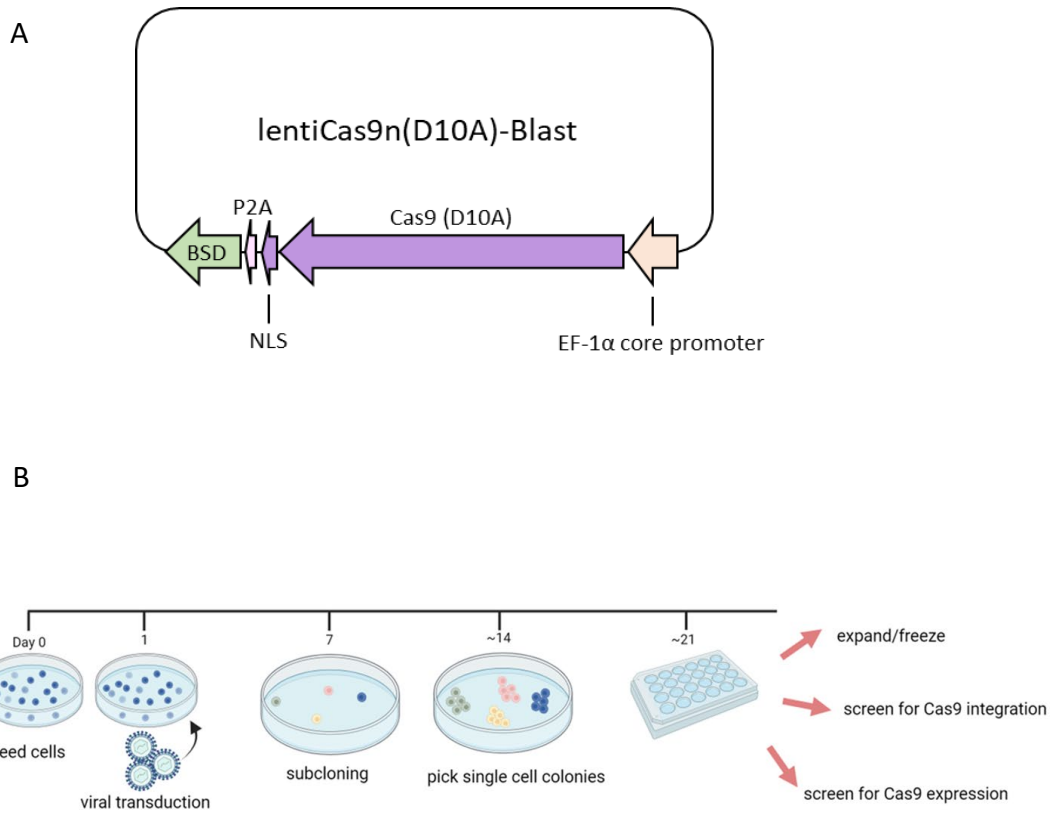


Figure 5-7: Experimental plan to generate HD-iPSCs expressing Cas9 nickase. ***(A)*** *Schematic depicting lentiviral construct containing Cas9 nickase under control of EF1-α promoter with a blasticidin selection marker.* ***(B)*** *Experimental outline for the generation of Cas9 nickase expressing iPSCs.*

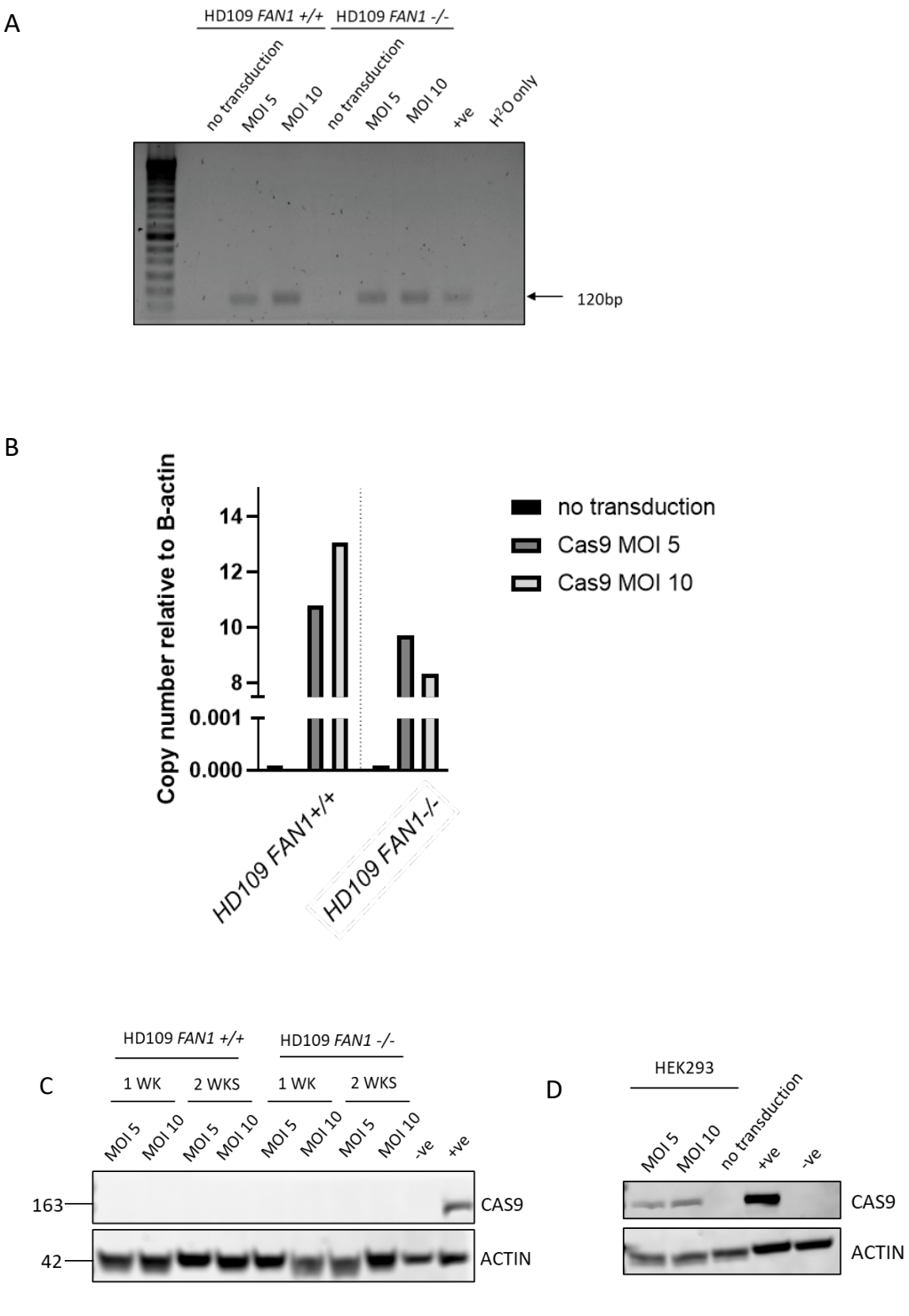


Figure 5-8: Cas9 is selectively silenced by HD-iPSCs. (A) *PCR amplifying 120 bp region demonstrating successful integration of Cas9 lentiviral vector in FAN1$^{+/+}$ and FAN1$^{-/-}$ cells.* ***(B)*** *qPCR confirming successful integration of Cas9 and copy number relative to β-actin.* ***(C)*** *Western blot demonstrating that after 1 and 2 weeks HD-iPSCs silence Cas9.* ***(D)*** *HEK293 cells express Cas9 1 week post-lentiviral treatment, confirming iPSC silencing is not due to poor lentiviral batch. Molecular weight of CAS9 (163) and ACTIN (42) in KDa.*

5.2.3 HD-iPSCs are capable of differentiation to iPSC-derived cortical neurons

Due to the selective silencing of Cas9 in both iPSC lines, I sought to differentiate both *FAN1*$^{+/+}$ and *FAN1*$^{-/-}$ lines to neurons, with the aim of transducing these with our Cas9 nickase therapy and monitoring contractions over time. Previous data from our lab has demonstrated that both neurons and astrocytes are able to undergo successful transduction with Cas9 and are capable of maintaining expression 6 weeks after transduction (Larin et al, unpublished; Murillo et al, unpublished).

There are a range of neuronal differentiation protocols depending on the specific cell type of interest. In HD the initial and most severely affected cell type are the medium spiny neurons (MSNs) of the striatum, however differentiations of iPSCs to MSNs are notoriously difficult with actual MSN yield as low as 10-15% (Li and Lu 2009; Lim et al. 2017). Comparatively, differentiation of iPSCs to cortical neurons tend to be more successful with higher reported rates of cortical neurons (Shi et al. 2012). I therefore opted to differentiate the cells to cortical neurons.

Initial attempts to generate cortical neurons utilising one well characterised protocol (Shi et al., 2012) failed to yield clean neuronal cultures, with a high degree of non-specific dividing cell types contaminating NPC and neuronal cultures. Furthermore, I often observed failure of NPCs to mature to neurons and exit the cell cycle, as demonstrated by ongoing proliferation in culture (data not shown). This occurred in both *FAN1*$^{+/+}$ and *FAN1*$^{-/-}$ lines, suggesting that the issue was either the cell line, the protocol, or both.

A second attempt to differentiate iPSCs was made, this time following the protocol from Telezhkin et al., 2016, with minor modifications. This relies on three separate differentiation media to enable neurogenesis and synaptogenesis, yielding mature cortical iPSC-derived neurons. iPSCs were cultured in E8 flex media and on day -1 split and replated in E8 flex media with 10 μM ROCK inhibitor. On day 0, if cells were confluent, they were differentiated in SLI media inhibiting SMAD and WNT signalling for 8-12 days and growth factors were then removed to induce NPC formation at ~day 16,

which can be expanded as NPCs with forebrain potential. NPCs are then further directed towards a cortical fate and are considered mature 3 weeks post-plating in sequential SJA and SJB media (Figure 5-9). One benefit of this protocol is that additional factors in SJA and SJB medium have demonstrated reduced proliferation rates of NPCs during the neuronal maturation and acts to synchronise NPC cell cycle exit (Telezhkin et al. 2016). Using this protocol, I was able to generate mature cortical neurons which stained positive for neuronal marker microtubule associate protein 2 (MAP2) at day 37 (Figure 5-10).

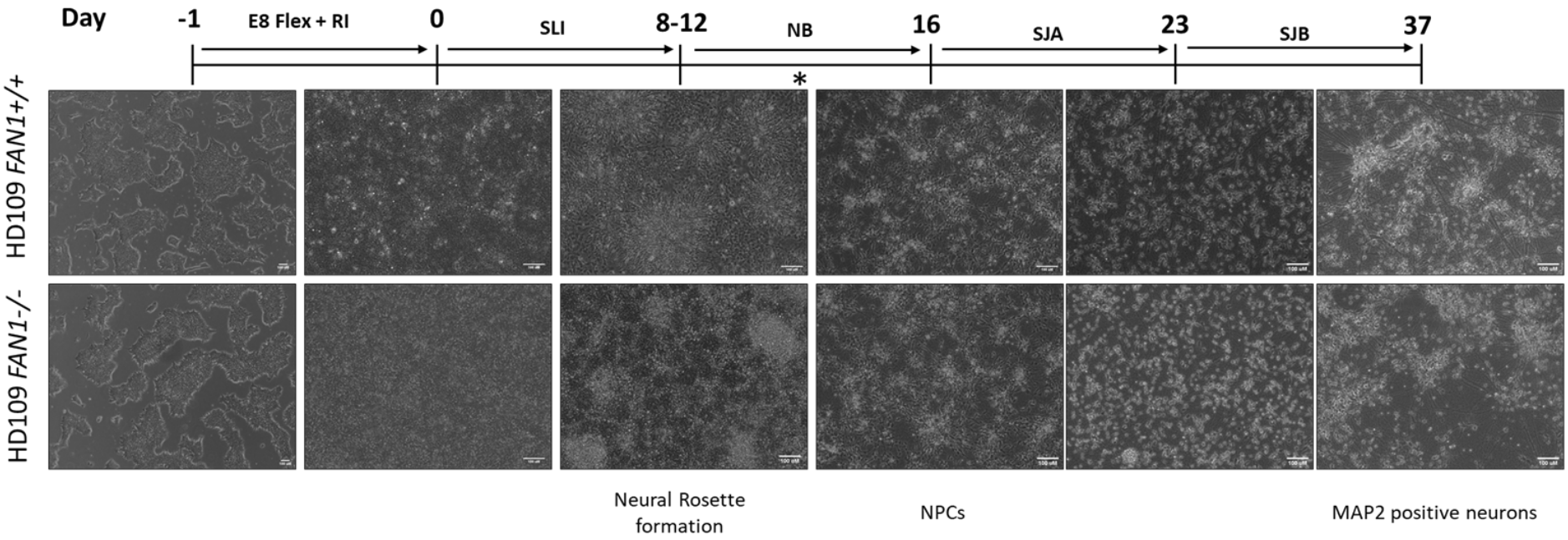


***Figure 5-9: Directed differentiation of HD-iPSCs to cortical neurons**. iPSCs were differentiated to NPCs through initial SLI media containing SMAD and WNT signaling inhibitors to induce neural rosette formation. After day 8-12 and upon appearance of rosettes cells were split and grown in NB media with the removal of inhibitors allowing for growth of NPCs. NPCs were then terminally differentiated to mature cortical neurons within 3 weeks after growth in SJA and SJB media. *NB media can be used to proliferate NPCs. Scale bar = 100 µm.*

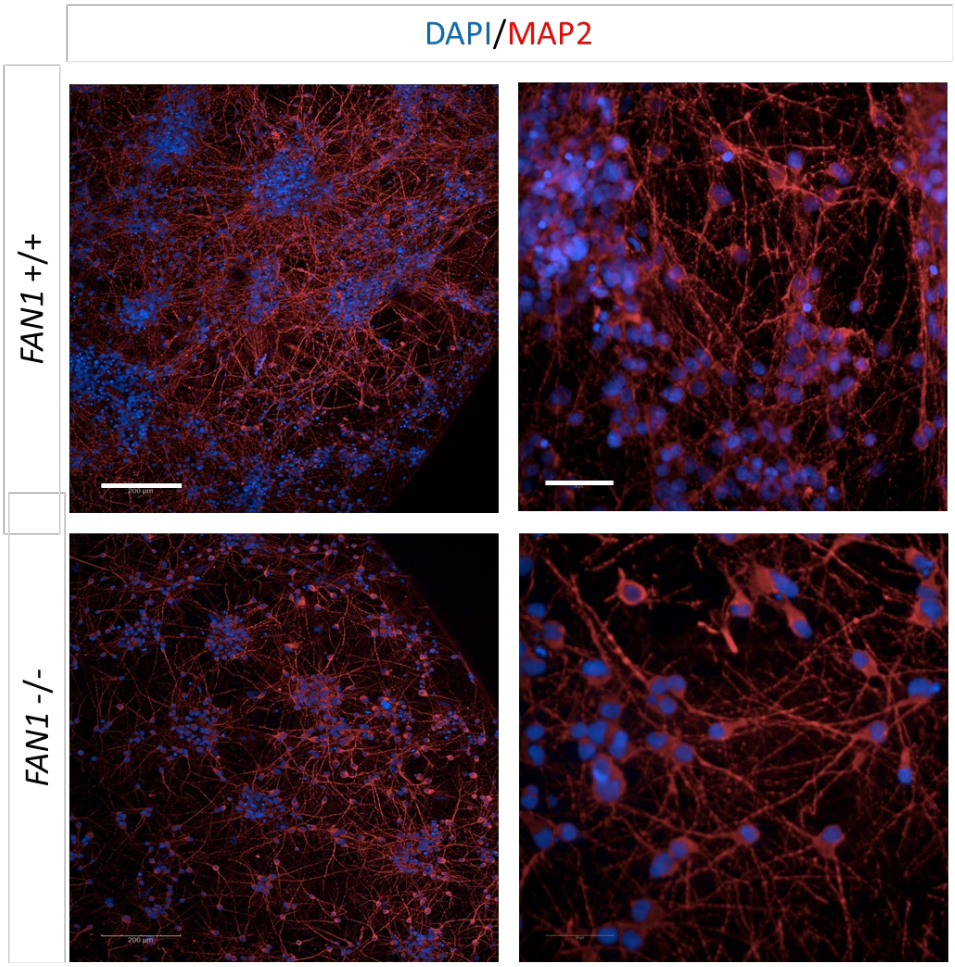


Figure 5-10: FAN1$^{+/+}$ and FAN1$^{-/-}$ cells were stain positive for neuronal specific MAP2. *Cells were fixed and stained at day 37, after a 3 week maturation period. Magnification at 4x (left) and 20x (right). Scale bars = 200 µm (left) and 50 µm (right).*

5.2.4 HD-iPSCs demonstrate variation between differentiations and presence of non-specific differentiated cell types

After confirming *FAN1*$^{+/+}$ and *FAN1*$^{-/-}$ lines were both capable of differentiating to MAP2 positive cortical neurons, I sought to reproduce these results with at least three separate differentiations per genotype, and transduce cells with our Cas9 nickase lentiviral vector alongside a sgCTG targeting the repeat tract.

Unfortunately, I was unable to replicate the differentiations previously described in both *FAN1*$^{+/+}$ and *FAN1*$^{-/-}$ lines. Whilst *FAN1*$^{-/-}$ lines were able to differentiate somewhat reliably, the quality of neurons produced was variable with some batches without contaminating cells and others demonstrating heterogenous cell types. Comparatively, in the *FAN1*$^{+/+}$ line, I was unable to generate mature neurons without contaminating non-specific cell types. Furthermore, I often observed problems early in the differentiation protocol, where holes began to form in the monolayer before the formation of neural rosettes, which often rendered these differentiations unviable (Figure 5-11). Those differentiations that continued beyond the initial neuronal induction period, yielded NPCs with a large amount of uncharacterised dividing cells. Upon plating these cultures for terminal neuron differentiation, these cells took over the culture (Figure 5-11). As these iPSCs are a subclone of the original HD109 line, I attempted to differentiate a separate subclone (N5), described in McAllister et al., 2022. However, in both *FAN1*$^{+/+}$ and *FAN1*$^{-/-}$ lines these did not progress past the neuronal induction stages due to the formation of holes in the monolayer (data not shown).

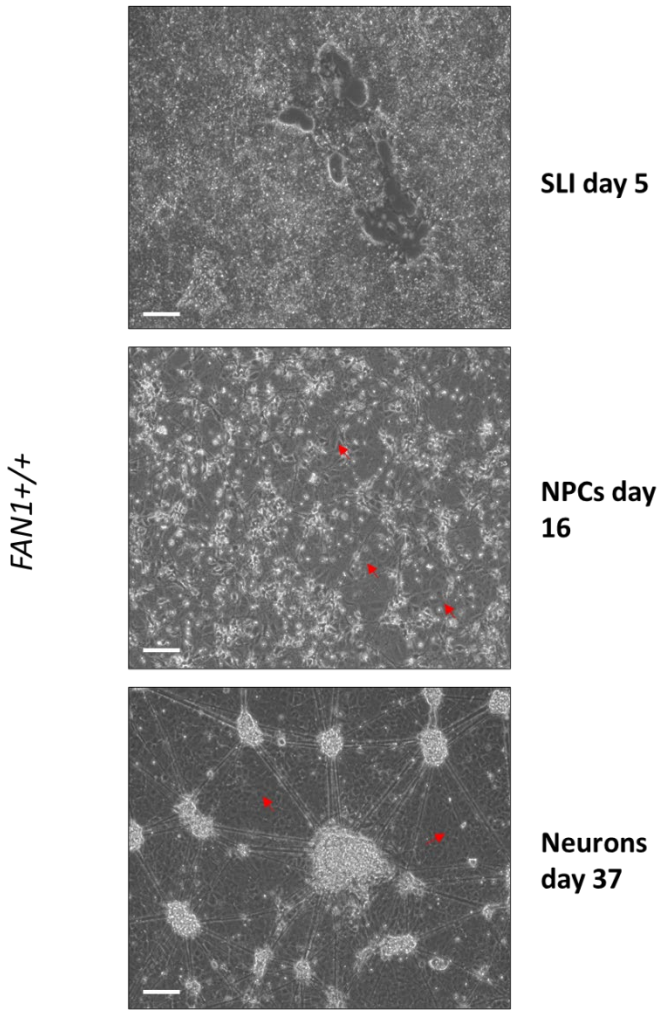


Figure 5-11: Demonstrative figures highlighting issues differentiating FAN1$^{+/+}$ iPSCs to cortical neurons. *Hole formation in neural induction monolayer (top) often led to unviable differentiations. The presence of non-specific dividing cell types (red arrows) within NPC populations (middle) lead to the outcompeting and further proliferation of these cells in neuronal cultures (bottom). Scale bar = 100 µm.*

5.2.5 Alternative models to investigate nickase-induced contractions in non-dividing HD-patient *FAN1*$^{+/+}$ and *FAN1*$^{-/-}$ cells

Due to inconsistencies between differentiations and contaminating cell types I was unable to use these iPSC-derived neurons to monitor nickase-induced contractions in *FAN1*$^{+/+}$ and *FAN1*$^{-/-}$ lines. However, I began work setting up potential model cell lines that may be used as an alternative for future work, which would allow us to investigate whether the phenotypes observed in the HEK293-reporter cell lines are recapitulated in non-dividing iPSC-derived cell types.

5.2.5.1 HD-109 iPSCs are able to differentiate to astrocyte precursors

Whilst I was unable to reliably induce differentiation of HD-iPSCs to cortical neurons, I was able to generate astrocyte precursors (APCs) using a protocol adapted from Serio et al., 2013. During development, neurogenesis precedes the process of astrogenesis and in this protocol existing NPCs can be utilised for subsequent differentiation to APCs. These APCs are capable of maturation to astrocytes, which, importantly, do not replicate (Serio et al. 2013). As such, a stock of APCs for both *FAN1*$^{+/+}$ and *FAN1*$^{-/-}$ lines present a source of cells capable of assessing nickase-induced contractions in non-dividing cells.

In this protocol NPCs from both *FAN1*$^{+/+}$ and *FAN1*$^{-/-}$ line were differentiated to APCs with epidermal growth factor (EGF) and leukaemia inhibitory factor (LIF) (Chambers et al. 2009; Serio et al. 2013; Telezhkin et al. 2016). After 6 weeks of treatment cells were stained with APC specific marker surface marker CD44 and sorted by FACS (Figure 5-12). This enabled collection and expansion of pure APC populations capable of further maturation to astrocytes.

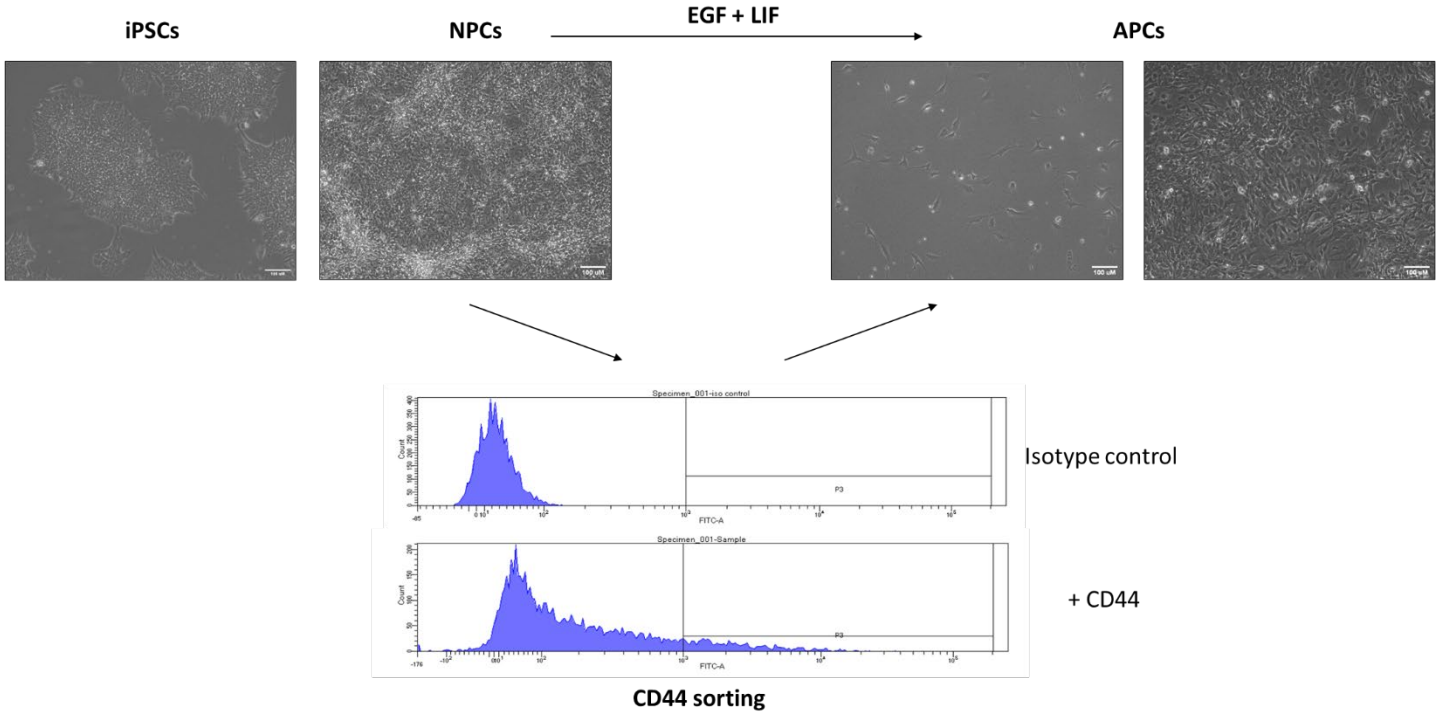


Figure 5-12: Generation of APCs. *After the generation of NPCs cells were treated with epidermal growth factor (EGF) and leaukaemia inhibitory factor (LIF) for up to 6 weeks, at which point cells were sorted for those positive for APC marker CD44. APCs were subsequently expanded for cryopreservation. Scale bar = 100 µm.*

5.2.5.2 Generation of inducible neurons (iNeurons) from patient HD-iPSCs

Due to limitations with small molecule inhibitor approaches for directed differentiation of iPSCs to neurons, recent advances in the field have led to an alternative approach relying on the overexpression of transcription factors to differentiate iPSCs to cell-type specific neurons. This method leads to rapid and direct differentiation to neuronal subtypes, termed 'inducible neurons' (iNeurons), bypassing the neural progenitor state (Chen et al., 2020; Zhang et al., 2013). Neurogenin-2 (NGN2) is a master transcription factor for robust expression of mature glutamatergic cortical neurons, with differentiation rates as high as 90% reported when using lentiviral vectors (Busskamp et al., 2014; Zhang et al., 2013). Recent advances in gene-editing technology has described the generation of iNeuron lines where a NGN2 transgene is stably integrated into the CLYBL safe harbour site in the genome under a doxycycline inducible promoter (Cerbini et al., 2015; Fernandopulle et al., 2018). This is mediated by the use of two TALEN arms designed with homology flanking sites within the CLYBL locus, allowing for the insertion of an inducible transgene via homologous recombination (Fernandopulle et al., 2018; Wang et al., 2017). Differentiation rates after single clone isolation is reported to be 100% (Fernandopulle et al., 2018).

Given the advantages presented by iNeurons these cells present a useful alternative to model CRISPR-Cas9 nickase induced contractions in HD patient iPSC-derived neurons. HD-iPSCs (N5 subclone) containing a stably integrated NGN2 transgene were generated by Joseph Stone (Massey Lab, Cardiff). Cells were transfected with three plasmids; two containing the TALEN homology arms and the third containing the NGN2 expression plasmid under a doxycycline-inducible promoter, expressing an mApple fluorescent protein. Cells expressing mApple were selected for further characterisation via PCR and sanger sequencing to confirm transgene insertion (see 2.1.5).

After confirmation of transgene insertion I set out to differentiate cells from iPSCs into cortical neurons following protocol by Fernandopulle et al., 2018 (Figure 5-13) . At day 0 iPSCs were dissociated and $1.5x10^6$ cells re-plated in a 6-well culture dish in an initial induction media. This media was

supplemented with 10 μM ROCK inhibitor for the first 24 hours, after which it was removed. Importantly, in addition to factors essential for cell maintenance, the induction media contained 2 μg/mL of doxycycline to induce transcription of NGN2 and blasticidin to maintain positive selection of the transgene. By day 3 iPSCs demonstrate clear neuronal morphology and were subsequently dissociated and $1x10^6$ cells re-plated in culture media. By day 11 iNeurons are considered mature and cells were fixed for staining with MAP2 and proliferation marker Ki67. INeurons stained positive for MAP2 and there were small numbers of Ki67 positive cells in some wells (Figure 5-13). These dividing cells can be removed with plurisin treatment, which selectively targets proliferating iPSCs (Zhang et al. 2014).

After confirming *FAN1*$^{+/+}$ cells were able to rapidly differentiate to MAP2-positive neurons after 11 days, I set out to generate *FAN1*$^{-/-}$ lines, derived from these HD-iPSC iNeurons. To do so the two *FAN1*-targeting cRNAs described in Chapter 3 (3.2.4.1) were utilised, with one modification. As these cells express mApple I was unable to use the fluorescent tracrRNA Atto550 scaffold, therefore an Atto647 was used instead. *FAN1*-specific targeting cRNAs were synthesized and duplexed with the Atto647-tagged tracrRNA. The functional sgRNA was then incubated with an Alt-R *S.p.* Cas9 Nuclease V3 to form a stable RNP complex. Cells were electroporated with this fluorescently labelled complex and after 24 hr cells were sorted based on fluorescence intensity, with only the top 10% of cells selected. Single cell clones were then screened for the presence of the correct deletion product. To date a total of 96 colonies have been screened, with the aid of Emma Randall and Christopher Smith (Dion lab) and present a pool of cells for further characterisation (Figure 5-14).

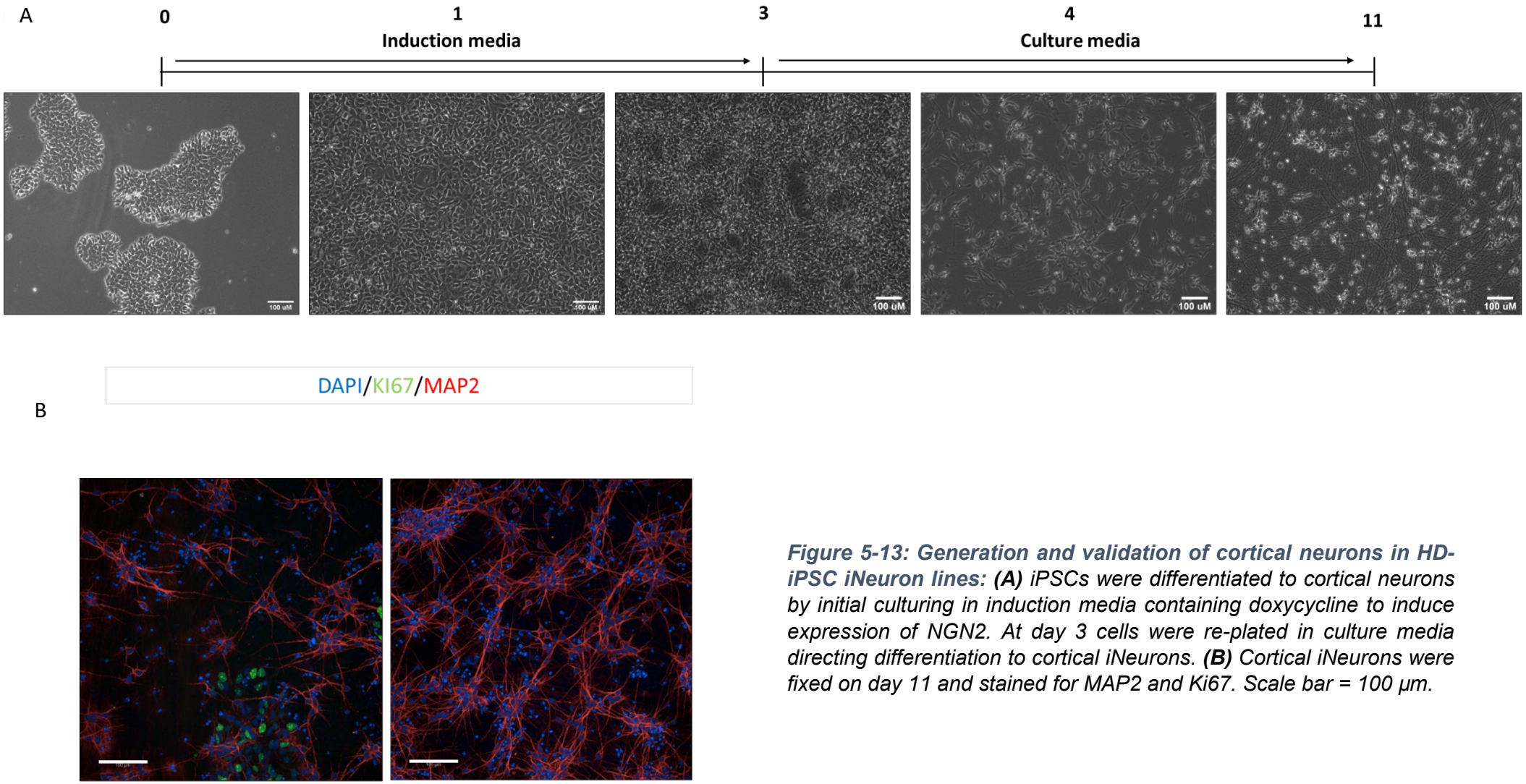


Figure 5-13: Generation and validation of cortical neurons in HD-iPSC iNeuron lines: ***(A)*** *iPSCs were differentiated to cortical neurons by initial culturing in induction media containing doxycycline to induce expression of NGN2. At day 3 cells were re-plated in culture media directing differentiation to cortical iNeurons.* ***(B)*** *Cortical iNeurons were fixed on day 11 and stained for MAP2 and Ki67. Scale bar = 100 µm.*

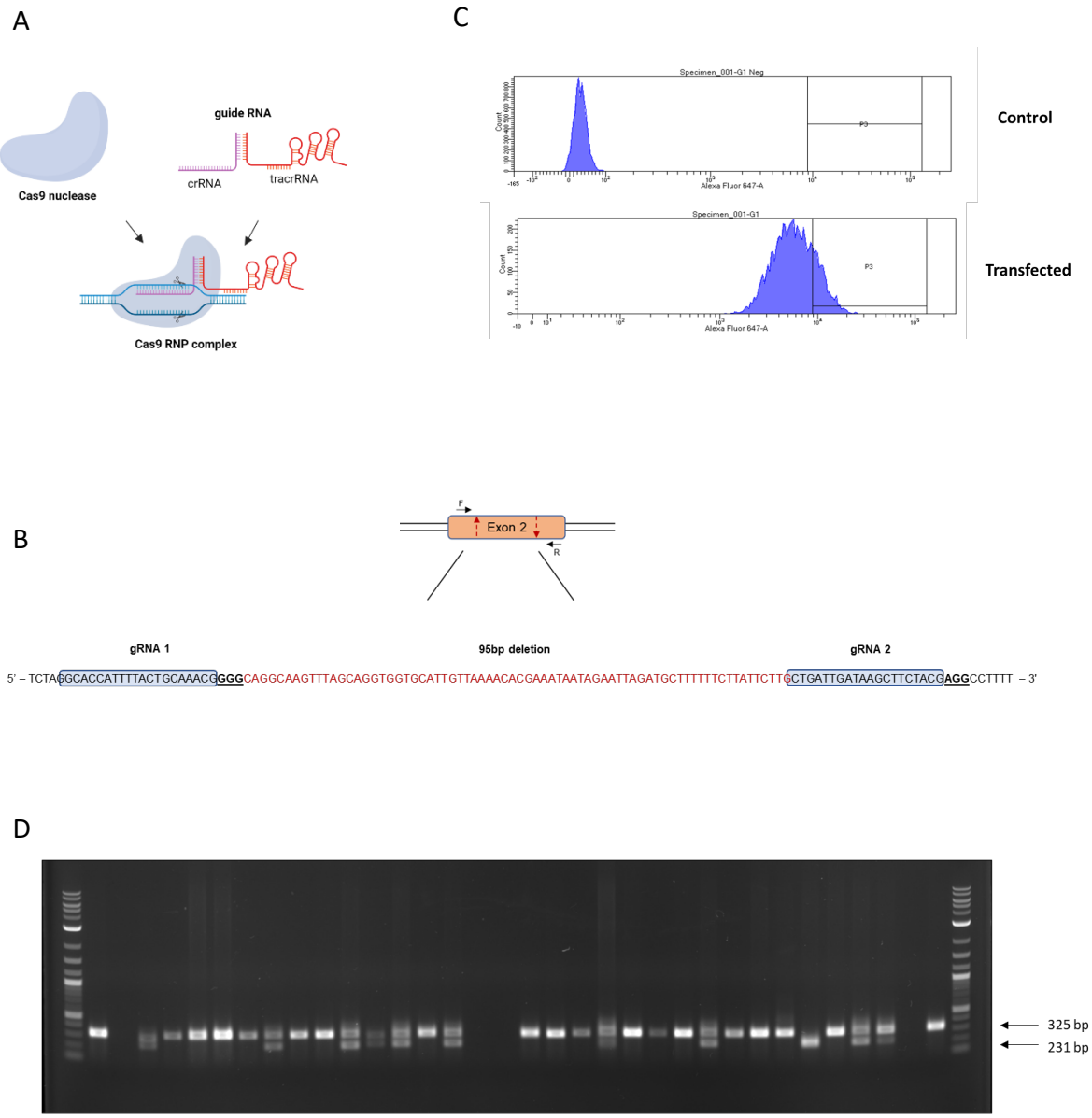


Figure 5-14: FAN1-targeting via CRISPR-Cas9 in HD-iPSC iNeuron lines. ***(A)*** *CRISPR-Cas9 facilitates gene editing as a Cas9 nuclease is targeted to the desired region in the genome by a sgRNA. The Cas9 nuclease and gRNA form a RNP complex leading to the formation of a DSB at the targeted editing site. (**B**) Two gRNAs were designed targeting FAN1 to create a 94p deletion within exon 2. (**C**) The top 10% of fluorescently labelled cells (P3) were sorted and plated for characterization and expansion. **(D)** A screening PCR amplifying exon 2 of FAN1 was carried out to identify potential homozygous FAN1 knock-outs with a molecular weight of 231 bp.*

5.3 Discussion

5.3.1 Selective silencing of Cas9 by HD-iPSCs

The initial aim of this chapter was to generate HD-iPSC lines that stably express the CRISPR-Cas9 nickase and sgCTG. HD-iPSCs provide a useful tool where CRISPR-Cas9 nickase-induced contraction rates can be monitored in patient-derived cell lines. Furthermore, these lines would allow for a direct comparison of nickase-induced contractions in replicating iPSCs and non-replicating neurons. However, I observed that HD-iPSCs silenced Cas9 expression after stable transduction with a *S. pyogenes* D10A Cas9 lentiviral construct. This was surprising as both lines demonstrated resistance to blasticidin after a 5-7 day selection period. In this lentiviral construct, both Cas9 and blasticidin were under the control of a Ef1-α promoter, separated by a P2A site. Prior to transduction with Cas9 I confirmed that Ef1-α promoter is efficiently expressed in HD-iPSCs through the use of a control vector containing an mCherry tag, with mCherry expression still high 2 weeks post-transduction. This was essential as studies have reported variable efficiency of promoter expression in ESCs and iPSCs, which are particularly prone to promoter silencing via methylation (Minoguchi and Iba 2008; Hoffmann et al. 2017). The absence of detectable expression by western blot cannot be attributed to batch effects of the lentivirus as the same stock was used to transduce HEK293 cells which demonstrated clear expression 1 week post-transduction. P2A is a protein splicing signal (Kim et al. 2011), therefore these observations suggest that the Cas9 protein is produced and translated, spliced off from Blasticidin S deaminase, then selectively degraded. To confirm this a pellet taken from these cells can be used for RT-qPCR, to assess Cas9 mRNA expression. This hypothesis suggests that treatment of cells with a proteasome inhibitor may ameliorate this effect.

The selective silencing of Cas9 was interesting as previous studies have demonstrated that iPSCs are capable of expressing Cas9 in culture. iPSCs transduced with Cas9 in a lentiviral construct have been described (Liao et al., 2022). Similarly to the method described in this thesis, cells were transduced with a vector that contained the Cas9 endonuclease and a blasticidin

resistance marker. However, in this case protein expression of Cas9 was not assayed by western blot, but Cas9 activity inferred utilising a BFP-GFP reporter-based assay. In this system cells were transduced with a gRNA targeting GFP, with depletion of GFP indicating functional Cas9 activity (Liao et al. 2022). This study does not report how routinely cells were assayed for Cas9 activity and, to my knowledge, long-term protein expression of Cas9 has not been reported in iPSCs. Additionally, recent progress in gene-targeting technology have detailed the generation inducible CRISPR (iCRISPR) iPSCs, where TALEN mediated targeting enables insertion of an iCRISPR cassette in a genomic safe harbour site (González et al. 2014). In this system expression of Cas9 is under the control of a doxycycline inducible promoter, and expression of Cas9, as detected by RT-qPCR, was stable for a 14-day time period in the presence of doxycycline. This study did not report detection of Cas9 via western blotting. Additionally, as these lines are used to generate gene knock-outs, Cas9 activity is inferred by presence of editing at the DNA level. Therefore, it is possible that Cas9 expression is downregulated over a longer induction period. Nevertheless, the generation of iCRISPR lines from HD-iPSCs may prove a useful tool in monitoring nickase-induced contractions in iPSCs. However, one drawback to this is that the initial HD-iPSC *FAN1*$^{+/+}$ lines would have to first undergo stable Cas9 insertion and validation, following which CRISPR targeting *FAN1* would need to be carried out, as the original *FAN1*$^{-/-}$ would no longer be isogenic to the new *FAN1*$^{+/+}$ iCRISPR line. Given that HD-iNeuron lines had already been generated, this option was not explored further but could be considered in the future.

Interestingly, a patent filed in 2022 has reported Cas9 silencing during iPSC differentiation to inducible neurons, where cells contain the NGN2 iNeuron system (WO 2023/105212 A1). In this case iPSCs were able to express Cas9 initially but upon differentiation, Cas9 underwent rapid silencing within 10 days. Furthermore, Cas9 mRNA was still present in the absence of detectable Cas9 protein, which also suggests that silencing is post-transcriptional. This study demonstrated that codon optimisation, where the Cas9 sequence is altered to reflect cell-type preferential codon usage, reduced this Cas9 silencing effect, with Cas9 still detectable at day 14 of differentiation.

Importantly, this study highlights the role of codon usage in successful expression of exogenous sequences. In this study the initial Cas9 was a commercially available Cas9, human-codon optimised to current gold standards, similar to the Cas9 nickase in our lentiviral construct. Whereas, this study demonstrate that cell-type specific codon optimisation may be required for successful expression of Cas9 protein in iPSCs and neurons. Therefore, it is possible that in our HD-iPSC lines codon usage may have a role in Cas9 silencing.

5.3.2 Alternative approaches to study nickase-induced contractions in HD-patient lines

Although many studies have reported successful differentiation of iPSCs to neuronal sub-types using dual SMAD inhibition and extrinsic factors, these have been noted to be limited due to low reproducibility, yield, and the propensity for undifferentiated proliferating cells to outcompete post-mitotic cells in culture over an extended period (Fernandopulle et al., 2018; Hulme et al., 2022). Whilst the HD109 line has been reported to be capable of differentiation to multiple brain cell types, including neurons, microglia and astrocytes (Mattis et al. 2012; Lim et al. 2017; Mehta et al. 2018; McAllister et al. 2022; Stöberl et al. 2023), I was unable to produce consistent neuronal batches. As an alternative to this approach many studies have reported the use of transcription factor overexpression as a means of reliable differentiation of iPSCs to neuronal subtypes. The advancement in gene-editing approaches has allowed for the TALEN-mediated generation of iNeuron iPSCs (Fernandopulle et al., 2018). These iNeurons provide a potentially valuable tool to determine whether FAN1 impacts nickase-induced contractions in non-replicating neurons, where mature neurons can be generated after only 11 days, with reduced batch-to-batch variation and presence of a heterogeneous population.

After generation of lines in a *FAN1*$^{+/+}$ background by Joseph Stone, I set out to validate the differentiation protocol outlined by Fernandopulle et al., 2018. After only 11 days of induction and maturation cells demonstrated neuronal morphology and were positively stained for neuronal marker MAP2. Co-

staining with Ki67, a marker of cell proliferation, did reveal some evidence of small patches of replicating cells, though this was infrequent. However, further optimisation of the differentiation protocol following this, including 48-hour plurisin treatment can effectively reduce the presence of these cells. Plurisin acts to inhibit stearoyl-CoA desaturase, a key enzyme in stem cell lipid metabolism (Zhang et al. 2014). This enables selective targeting of any iPSCs that may be present in the neuronal culture. Following confirmation of successful neuronal staining, a CRISPR targeting *FAN1* was carried out to generate isogenic *FAN1*$^{-/-}$ iPSCs with NGN2 transgene expression. Due to time limitations, these cells have not been fully characterised but are a tool for future investigation to monitor nickase-induced contractions in both *FAN1*$^{+/+}$ and *FAN1*$^{-/-}$ backgrounds.

Whilst iNeurons present a rapid and more reliable method of generating neuronal cultures, there are some limitations. Whilst these may not be applicable for the purposes of this thesis, which purely aim to investigate nickase-induced contractions in cell-types relevant to HD comparing *FAN1*$^{+/+}$ and *FAN1*$^{-/-}$ contraction rates, they are worth considering for any downstream applications. One study compared electrophysiological properties of iNeurons at 2-3 weeks and Embryoid body (EB)-derived dual SMAD inhibition neurons at 3-6 months maturation. EB-derived neurons were considered more mature based on action potential amplitude and firing, similar to human adult neurons. Comparatively, iNeurons more closely resembled second trimester human brain tissue (Rosa et al. 2020). Furthermore, there is some question regarding the degree of purity of iNeuron cultures, with regards to regional-specific neuronal cell types. Indeed, one study suggests that whilst overexpression with NGN2 leads to ~95% of cells staining positive for MAP2, the remaining neuronal population stained positive for PERIPHERIN, a marker of peripheral nervous system (PNS) neurons (Chen et al. 2020). This study demonstrated that a short CNS patterning step, using SMAD and WNT signally inhibitors prior to NGN induction, reduced this population. Finally, it may be important to consider co-culturing iNeurons with mature astrocytes. Rosa and colleagues demonstrate that co-culturing of iNeurons with astrocytes demonstrate a more mature neuronal profile, closer to EB-derived neurons. Indeed, the SMAD

differentiation approach to generate neurons naturally leads to low yields of astrocytes within the culture, which can increase the viability of neurons over long-term culture (Tanaka et al. 1999; Rosa et al. 2020). The generation of APCs, described in 5.2.5.1, therefore represent a useful stock of cells for potential co-culturing with iNeurons to improve their viability in culture. As these APCs are capable of maturation to, non-replicating, astrocytes and have both *FAN1*$^{+/+}$ and *FAN1*$^{-/-}$ genotypes with comparable expanded CAG repeat lengths, these would not confound experiments aimed to monitor nickase-induced contractions.

5.3.3 Chapter Summary:

- HD-iPSCs selectively silence a *S. pyogenes* Cas9 D10A nickase delivered by lentiviral transduction
- Due to the unreliable differentiation of *FAN1*$^{+/+}$ and *FAN1*$^{-/-}$ via dual SMAD inhibition, monitoring of Cas9 nickase-induced contractions was not successful in neuronal cells
- iPSC-derived APCs and iNeurons have been validated for future use

6 General Discussion

6.1 Summary of findings

In Chapter 3, *FAN1*$^{-/-}$ lines were generated in a HEK293 GFP reporter cell line with expanded CAG/CTG repeats. These cells also contain a stably integrated CRISPR-Cas9 D10A nickase and sgCTG targeting the repeat tract. As the expression of the GFP reporter is under the control of a doxycycline inducible promoter, this allowed me to investigate the effects of the Cas9 nickase on an expanded CAG/CTG repeat tract after the induction of transcription. In this chapter, three separate approaches demonstrated that a loss of FAN1 leads to an increase in Cas9 nickase-induced contractions. Importantly, this chapter was able to validate the use of the PacBio long-read sequencing, which mirrored the findings of the previously established GFP-reporter assay and small-pool PCR. This is essential as this sequencing approach demonstrates increased sensitivity allowing for the assessment of more subtle changes in CAG repeat tract lengths. Taken together, these data indicates FAN1 is protective against Cas9-induced contractions at CAG/CTG repeats. To confirm that this phenotype was specific to the action of the nickase, a *FAN1*$^{-/-}$ clone in GFP(CAG101) cells, without the Cas9 nickase treatment was generated. In line with current literature, loss of FAN1 increased somatic expansion over 42 days (Goold et al. 2019; McAllister et al. 2022). Taken together, these data indicate that FAN1 is important for protection against Cas9 nickase-induced contractions and our HEK293 reporter cell line is able to model phenotypes of somatic instability *in vitro*.

In Chapter 4, I sought to determine which functional domains of FAN1 were important for preventing Cas9 nickase-induced contractions. The *FAN1*$^{-/-}$ clone #7 line was stably transduced with FAN1 variants. Analysis via the more sensitive long-read PacBio sequencing demonstrated that only the cells expressing wild-type FAN1 were able to reduce nickase-induced contractions rates to *FAN1*$^{+/+}$ levels. Cells expressing constructs with mutations affecting nuclease activity, nuclease-dead and endonuclease-dead, demonstrated partial rescue to *FAN1*$^{+/+}$ levels. This phenotype was also seen in cells rescued with a FAN1 construct with a UBZ domain mutant. In cells rescued with a DNA-

binding FAN1 mutant, originally identified as a modifier of HD age at onset (Lee et al. 2015), there was a no recovery. Indicating DNA-binding is essential for FAN1's role in protecting against nickase-induced contractions. Finally, I have shown that knocking out MLH1 in the *FAN1*$^{-/-}$ background did not abolish the increased contractions observed in *FAN1*$^{-/-}$ lines, indicating the MMR machinery does not drive these contraction rates. This result emphasizes that the genetic requirements for nickase-induced contractions are different from those for somatic instability.

In Chapter 5, I sought to investigate whether the protective effect of FAN1 against nickase-induced contractions demonstrated in Chapters 3 and 4 were recapitulated in HD patient-derived iPSCs. Interestingly, iPSCs silenced Cas9 expression, despite confirmation of successful vector integration. Further attempts to model how FAN1 expression impacts contractions were attempted in *FAN1*$^{+/+}$ and *FAN1*$^{-/-}$ lines by generating iPSC-derived cortical neurons. Unfortunately, the process of generating these cells, via dual SMAD inhibition, demonstrated high variability with many neuronal cultures contaminated with non-differentiated cell types. However, tools to for future investigation have been established. These include astrocyte precursors in *FAN1*$^{+/+}$ and *FAN1*$^{-/-}$ lines, which are capable of further differentiation to mature, non-replicating astrocytes. Additionally, the generation of iNeurons in *FAN1*$^{+/+}$ and *FAN1*$^{-/-}$ lines have been described, which provide a valuable tool to investigate nickase-induced contractions in cortical neurons, with reduced culture variability.

6.2 FAN1 and somatic expansion

FAN1 is a structure specific endonuclease and 5′-3′ exonuclease implicated in multiple DNA repair pathways; including; ICL repair and recovery of stalled replication forks (Kratz et al. 2010; Liu et al. 2010b; MacKay et al. 2010; Smogorzewska et al. 2010; Chaudhury et al. 2014; Lachaud et al. 2016). Recent GWA studies have identified FAN1 as a genetic modifier of HD age at onset; with modifier signals in both coding and non-coding regions. Importantly, some genome-wide significant SNPs associated with earlier age at onset have been predicted to be deleterious for FAN1 function (Bastarache

et al. 2018; Lee et al. 2019). Multiple studies have demonstrated that FAN1 acts to prevent somatic expansions at expanded CAG/CTG and CGG/CCG repeats (Zhao and Usdin 2018; Goold et al. 2019; Kim et al. 2020; Goold et al. 2021a; Zhao et al. 2021; McAllister et al. 2022).

FAN1 is thought to protect against somatic expansions in two ways. The first is through binding to and processing of secondary structures at expanded repeats, such as CAG/CTG loop outs (Kim et al. 2020; Deshmukh et al. 2021b; Goold et al. 2021; McAllister et al. 2022; Phadte et al. 2023). These structures can act as blocks to replication or, importantly in non-replicating neurons, transcription and therefore need to be resolved. It has been proposed that FAN1 acts on these structures utilising both its endo- and exonuclease activity (Deshmukh et al. 2021b). A biochemical based assay modelling nucleolytic processing by FAN1 on small CAG/CTG extrusions proposes that resolution of such structures involves short-patch DNA excision and a DNA resynthesis mechanism, that may involve other nucleases, DNA polymerase and ligases (Phadte et al. 2023). This study also demonstrated that FAN1 competes with the MutSβ complex, from the MMR machinery, for DNA binding at these structures. Importantly, increasing concentrations of MutSβ inhibited FAN1 nuclease activity, suggesting that at sites of CAG/CTG loop outs FAN1 and MutSβ compete for occupancy (Phadte et al. 2023). In this model, repair by FAN1 may limit the opportunity for strand-slippage or error-associated repair which occurs via MMR, thus protecting against somatic expansions.

Aside from directly processing CAG/CTG loop out structures, FAN1 is also thought to protect against somatic expansions through binding to and sequestering MLH1. MLH1 is the main component of all three MutL complexes, which act downstream in the MMR repair pathway to resolve DNA mismatches (Schmidt and Pearson 2016). Importantly, there is evidence that MLH1 is epistatic to FAN1 as in a HD knock-in mouse model an increase in somatic expansions observed in *Fan1*$^{-/-}$ mice was ablated in *Fan1*$^{-/-}$/*Mlh1*$^{-/-}$ mice (Loupe et al. 2020). This indicates MLH1 is required for the increase in somatic expansions seen in the absence of FAN1. The two MutL complexes primarily implicated in somatic expansion are MutLα and Mutlγ (Gomes-Pereira et al. 2004; Pinto et al. 2013). In 2021, two studies described two novel

MLH1-interaction sites within FAN1 (Goold et al. 2021; Porro et al. 2021). Data from these studies, in U2OS cells, suggests that the interaction between FAN1 and MLH1 may be important in mediating efficient repair. In this scenario, FAN1 may act to sequester MLH1, preventing DNA binding and repair by MutL complexes, therefore limiting error-prone repair by MMR which can lead to somatic expansions (Figure 6-1). The importance of this interaction remains to be validated in other model systems.

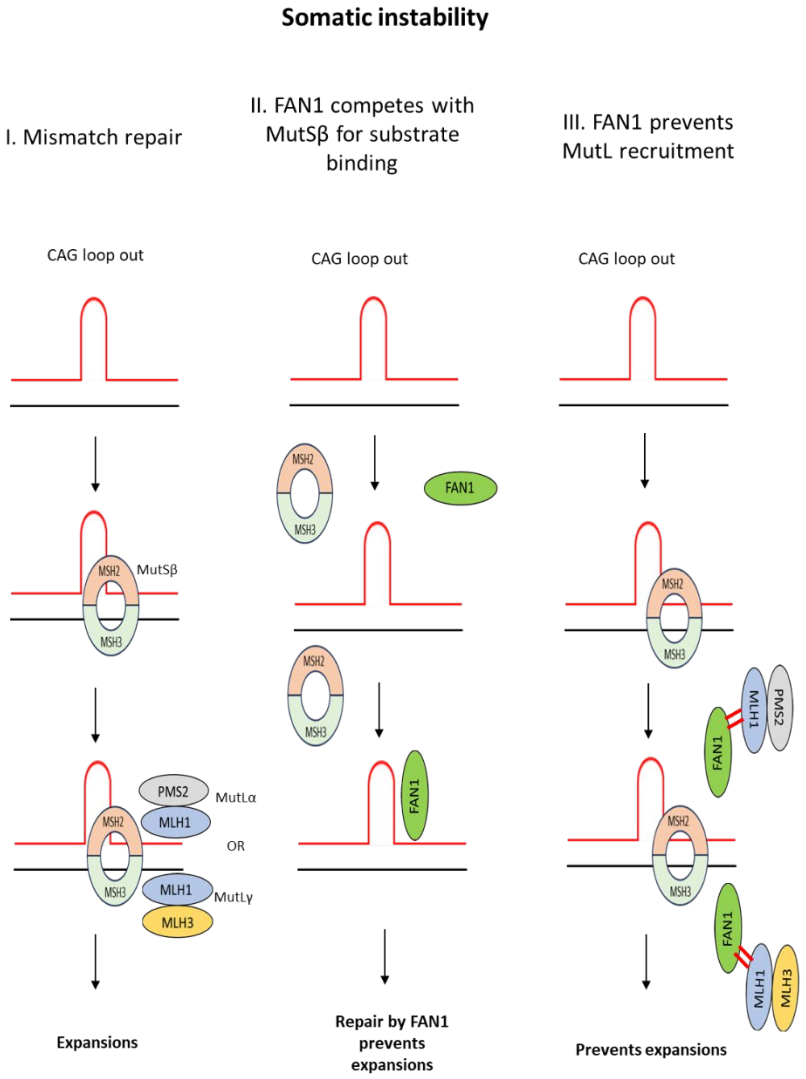


Figure 6-1: FAN1's protects against somatic instability at CAG/CTG loop-outs. ***(I)*** *MMR acts to promote somatic expansion and sites of CAG/CTG loop outs, where MutSβ binds to looped-out structures and subsequent recruitment of MutL complexes results in error-prone repair and somatic expansion.* ***(II)*** *FAN1 and MutSβ compete for occupancy at sites of loop-outs. FAN1 binding prevents MutSβ binding and repair by FAN1 acts to prevent somatic expansions.* ***(III)*** *FAN1 may act to prevent somatic expansion through binding to MLH1, preventing processing of structures by MutL complexes.*

6.3 FAN1 and CRISPR-Cas9 nickase-induced contractions

How DNA nicks are repaired is not fully understood, with multiple DNA repair pathways implicated, including; single-strand break repair (SSBR) (Caldecott 2008) or nick-induced homology directed repair (HDR) (Metzger et al. 2011; Davis and Maizels 2014; Davis and Maizels 2016). Given the complexity of DNA repair pathways involved, it is hard to determine exactly how FAN1 acts to prevent CRISPR-Cas9 nickase induced contractions, or indeed, how nicks induced by the Cas9 nickase give rise to contraction events. Understanding how contraction events occur, or which factors promote/inhibit contractions, could allow for a more efficient targeting approach which could assist in the development of a potential therapy.

In SSBR, DNA nicks are recognized by X-Ray Repair Cross-Complementing Protein 1 (XRCC1) and Poly (ADP-ribose) polymerase (PARP). This XRCC1-PARP complex acts to detect nicks and aids in mediation of repair SSBR pathways (Caldecott 2003). Previous data from our lab indicates that SSBR is not involved in the generation of nickase-induced contractions, where knockdown of XRCC1 and inhibition of PARP did not affect contraction rates, as detected by GFP expression via flow cytometry (Cinesi et al. 2016). Interestingly, studies have shown that DNA nicks/SSBs can be repaired by nick-induced HDR. In this pathway RAD51 promotes strand invasion and repair using a homologous dsDNA donor as a template (Davis and Maizels 2014). In this model a homologous donor, specifically a sister chromatid, is used and after strand invasion RAD51 facilitates the search for homology and formation of a connection between the invading DNA substrate and duplexed sister DNA, as in canonical HDR (Krejci et al. 2012). This leads to the generation of a D-loop structure and subsequent DNA synthesis occurs on the nicked-strand, using the homologous DNA as a template, promoting faithful repair.

Therefore, one way FAN1 could act to prevent nickase-induced contractions in our HEK293 cell line is through a role in nick-induced HDR (Figure 6-2). Whilst FAN1 has not been directly implicated in this pathway, it has been demonstrated that FAN1 is able to cleave D-loop structures *in vitro* (Kratz et

al. 2010). These D-loop structures can form during the process of HDR. These data, alongside one study demonstrating that siRNA knockdown of *FAN1* causes a delay in depletion of RAD51 foci (MacKay et al. 2010), indicate that FAN1 may play a role in the latter stages of HDR, which has been proposed as a downstream mechanism during ICL repair (Jin and Cho 2017; Deshmukh et al. 2021a).

In this model (Figure 6-2), if a DNA nick were to occur, either during replication or in the G2 phase of the cell cycle, this could lead to the induction of HDR. In this scenario, the generation of a DNA nick is followed by DNA end resection at the break site, providing a 3′-OH single-stranded DNA tail, as in canonical HDR (Li and Heyer 2008). Subsequent loading of RAD51 by BRCA2 to the exposed 3′ single-stranded DNA may occur, facilitating the strand invasion of a homologous sister chromatid and the generation of a D-loop structure. In this context, after DNA synthesis within the invading strand, FAN1 could act to resolve the D-loop, disengaging the invading strand before the formation of a double-holiday junction (Figure 6-2). This would likely lead to the formation of DNA flaps, which may also be processed by FAN1, likely through its endonuclease activity. Therefore, in this model FAN1 acts within the latter stages of HDR and aids in a more faithful repair of Cas9 nickase-induced contractions, thus preventing contraction events. Given this hypothesis it would be interesting for future experiments to investigate the functional consequence of loss of RAD51 or BRCA2. For example, if knockdown of these two proteins leads to an increase in Cas9 nickase-induced contractions, comparable with the FAN1 knock-out lines, we may conclude that HDR acts to prevent contraction events.

One caveat to this is that data indicating a role for FAN1 in HDR is from biochemical-based assays and replicating cells. Whilst FAN1 may also be important for HDR in our HEK293 system it is unclear whether this model is relevant for non-dividing neurons. HDR is generally thought to be confined to the late S or G2 phase of the cell-cycle (Zhao et al. 2017). However, some evidence has indicated terminal neurons may be capable of utilizing HDR pathways. One study aimed to monitor DSB repair in rod photoreceptor neurons in mice and demonstrated that ~15% of DSB were repaired by a sub-

pathway of HDR, the single-strand annealing pathway (Chan et al. 2011). Furthermore, a second study demonstrated that precise genome editing via HDR is possible in mature post-mitotic neurons in mice, when a homologous donor template is delivered via adeno-associated viral delivery (Nishiyama et al. 2017). Therefore, whilst the exact mechanisms of DNA repair utilized in neurons/non-dividing cells are unknown, there is evidence that HDR and FAN1 may be relevant in these cell types.

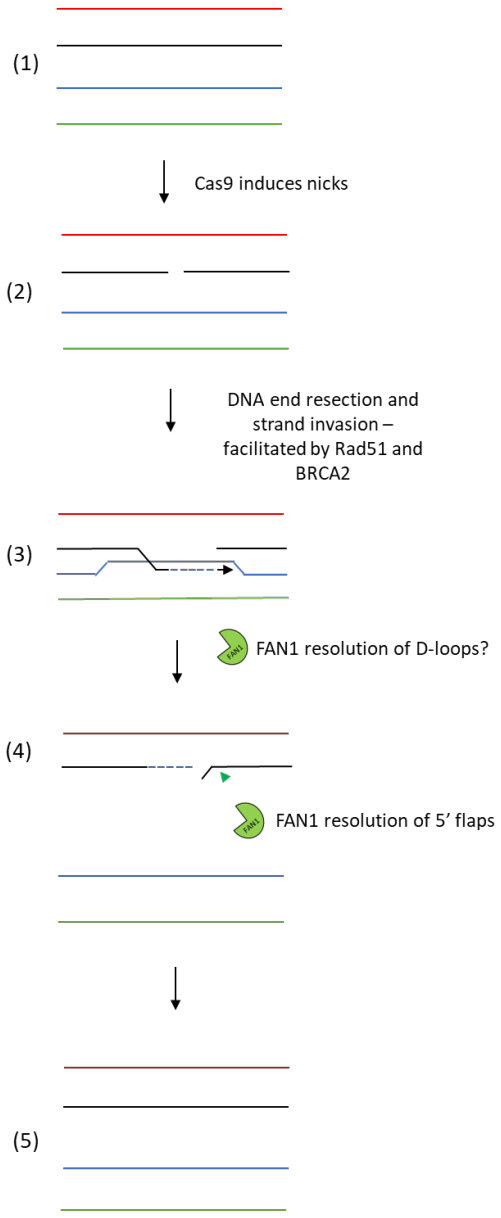


Figure 6-2: FAN1 may protect against nickase-induced contractions through a role in nick-induced HDR. *FAN1 may act in nick-induced HDR as a means of protecting against Cas9 nickase-induced contractions. In this pathway DNA nicks may be generated during late S phase or G2 (2), subsequent DNA end resection and strand invasion may occur, using a sister chromatid as a template for DNA synthesis (3). FAN1 could act downstream of this action to resolve D-loop structures or alternatively may also be involved in resolution of flap DNA (4), prior to DNA ligation.*

MMR is a driver of somatic instability, however in 2016 Cinesi et al suggested that MMR is not a primary driver of Cas9 nickase-induced contractions. This was based upon observations that siRNA knockdown of MSH2, the core factor of both MutSα and MutSβ, did not significantly reduce nickase-induced contractions. FAN1 is proposed to canonically protect against somatic expansion at CAG/CTG repeats through competition with MutSβ for binding to CAG/CTG loop outs and subsequent processing of such structures via it's nuclease activity (Phadte et al. 2023). Or additionally, through sequestration/binding of MLH1 (Goold et al. 2021; Porro et al. 2021). As siRNA-mediated knockdowns are never 100% efficient, and given the important role of FAN1 in modulating MMR machinery, I speculated that it was still possible that in the absence of FAN1, an increase in nickase-induced contractions may occur through an error prone MMR-mediated repair mechanism. Therefore, we would expect in a *FAN1* and *MLH1* double knock-out cell line, that the increase in nickase-induced contractions seen in *FAN1*$^{-/-}$ cells is rescued, and contraction rates are reduced to *FAN1*$^{+/+}$ levels or less. However, this was not the case. When examining the PacBio long-read sequencing data a *FAN1* and *MLH1* double knock-out clone failed to reduce contraction rates to *FAN1*$^{+/+}$ levels.

It is important to mention that this data is the result from only one clone. Thus, a suggestion for future work would be to generate more clones for validation, or given the availability of short-hairpin RNAs (shRNAs) packaged into a lentivirus in our lab, stable knockdown of *MLH1* in both *FAN1*$^{+/+}$ and *FAN1*$^{-/-}$ lines is also an option. This would negate any issues with results obtained from a single clone. It should be noted that I have attempted to make a second clone multiple times, as well as single *MLH1* knock-out clones, without success. This may be due to issues with the sgRNA targeting *MLH1*, as whilst on-target prediction scores aim to highlight the editing efficiency of sgRNAs, these are not always indicative of successful editing. Aside from shRNA knockdown of *MLH1* or re-designing CRISPR sgRNAs, a useful future experiment would be to rescue *FAN1*$^{-/-}$ cells with the SPYF mutant, which reduces FAN1-MLH1 binding (Goold et al. 2021). In this case, if complementation with this mutant fails to reduce contraction rates to *FAN1*$^{+/+}$

levels, this would support findings that binding of MLH1 by FAN1 is not required for FAN1s activity in preventing nickase-induced contractions.

Despite data indicating no active role for MMR in nickase-induced contractions, it is still possible that the structures which are likely to form at sites of Cas9-induced nicks within CAG/CTG repeats are important in FAN1's ability to prevent nickase-induced contractions. FAN1 is a diverse DNA repair protein, with the ability to bind and resolve an array of structures including; 5′ flaps (Kratz et al. 2010; Liu et al. 2010b; MacKay et al. 2010; Smogorzewska et al. 2010; Yoshikiyo et al. 2010; McAllister et al. 2022), DNA nicks, (Kratz et al. 2010; Smogorzewska et al. 2010) DNA gaps (Kratz et al. 2010; Pennell et al. 2014) and CAG/CTG loop outs (Kim et al. 2020; Deshmukh et al. 2021). Given these diverse functions of FAN1, it may be that FAN1 protects against nickase-induced contractions through binding and subsequent nucleolytic processing of such structures. Binding of FAN1 may reduce the ability of other more error-prone nucleases to bind to these sites and initiate an error-prone repair, resulting in nickase-induced contractions. After binding, nucleolytic processing by FAN1 may promote faithful resolution of these structures and subsequent repair.

This theory is supported by data from Chapter 4, indicating that only a rescue with a wild-type FAN1 construct is sufficient to reduce Cas9 nickase-induced contraction rates to *FAN1*$^{+/+}$ levels. Presumably, this wild-type variant would be able to bind an array of DNA structures that may arise after the formation of a Cas9-induced nick, including 5′ flap DNA, DNA nicks or CAG/CTG loop outs (see Figure 6-3 (2)). Downstream of FAN1 binding there is likely processing by either FAN1's endo- or exonuclease activity with subsequent processing of DNA by other repair factors (DNA polymerases, ligases) to promote faithful repair (see Figure 6-3 (3) and (4)). Further data from Chapter 4 supports this model. A mutant in the DNA-binding domain of FAN1, R507H, significantly fails to reduce nickase-induced contractions to wild-type levels. DNA binding to substrates likely reduces competitive binding of other nucleases/repair factors, in a similar manner to how FAN1 competes for binding with MutSβ (Phadte et al. 2023). Following binding, nucleolytic processing of such structures by FAN1 would occur and could promote an error-free downstream repair pathway (as

detailed in Figure 6-3). Therefore, the effect of the R507H mutant has a secondary effect. Not only does it impact DNA binding, but also would presumably have a knock-on effect on FAN1's nuclease activity, which would be dampened due to reduced binding.

This may to some extent explain findings which show a partial recovery when *FAN1*$^{-/-}$ cells are rescued with nuclease-dead (D960A) or endonuclease dead mutants. Whilst these mutants are limited in their nuclease abilities, they can still bind DNA substrates (Kratz et al. 2010; Zhao et al. 2014). This would provide partial protection against nickase-induced contractions by preventing binding of other DNA factors. However, a question remains then, if FAN1 is unable to process DNA structures using its nuclease activity, would another repair pathway take over? In this case, we would predict a nuclease-dead FAN1 may not confer any protection against nickase-induced contractions, as presumably FAN1 would dissociate from DNA substrates allowing for binding of other more error-prone nucleases that could cause contraction events. In this case the UBZ domain may be an important factor. Indeed, the PacBio long-read sequencing data demonstrates that a UBZ mutant also results in a partial rescue, similar to both the nuclease-dead (D960A) and endonuclease dead variants.

The role of the UBZ domain in this protective mechanism is unclear. It has been demonstrated that this FAN1 domain is important for its recruitment to sites of DNA damage by monoubiquitinated proteins, including FANCD2 and PCNA (Kratz et al. 2010; Liu et al. 2010b; Smogorzewska et al. 2010; Porro et al. 2017). Therefore, the partial reduction observed in *FAN1*$^{-/-}$ cells rescued with a C44A/C47A UBZ domain mutant may be explained by a reduction of FAN1 recruitment to sites of DNA damage. The partial rescue may derive from the ability of FAN1 to localize to sites of DNA damage independently of UBZ-domain-mediated recruitment (Yoshikiyo et al. 2010; Zhou et al. 2012; Goold et al. 2021). Alternatively, the UBZ domain may be important downstream in repair of nickase-induced contractions, mediating recruitment of a faithful polymerase or ligase for final repair. A potential recruitment phenotype by the UBZ domain, particularly of other DNA repair factors, could explain the partial rescue observed in nuclease-dead (D960A) and endonuclease dead variants.

In this case, whilst FAN1's nuclease ability is limited, it is possible that the UBZ domain may independently recruit other DNA repair factors. Faithful repair and prevention of Cas9-nickase induced contractions may still be promoted, even with limited nuclease activity by FAN1. It is therefore clear that further work is needed to assess the importance of the UBZ domain of FAN1 in the context of CAG/CTG repeat structures.

Taken together, the results gathered in Chapter 4 suggest that multiple functions of FAN1 are important in preventing nickase-induced contractions and the requirements are different from those involved in somatic expansion. This is relevant for not only HD, but other CAG/CTG diseases. Understanding how FAN1 processes DNA intermediates could be helpful in guiding future mechanistic studies investigating how FAN1 is recruited to, or processes, CAG/CTG loop-out structures.

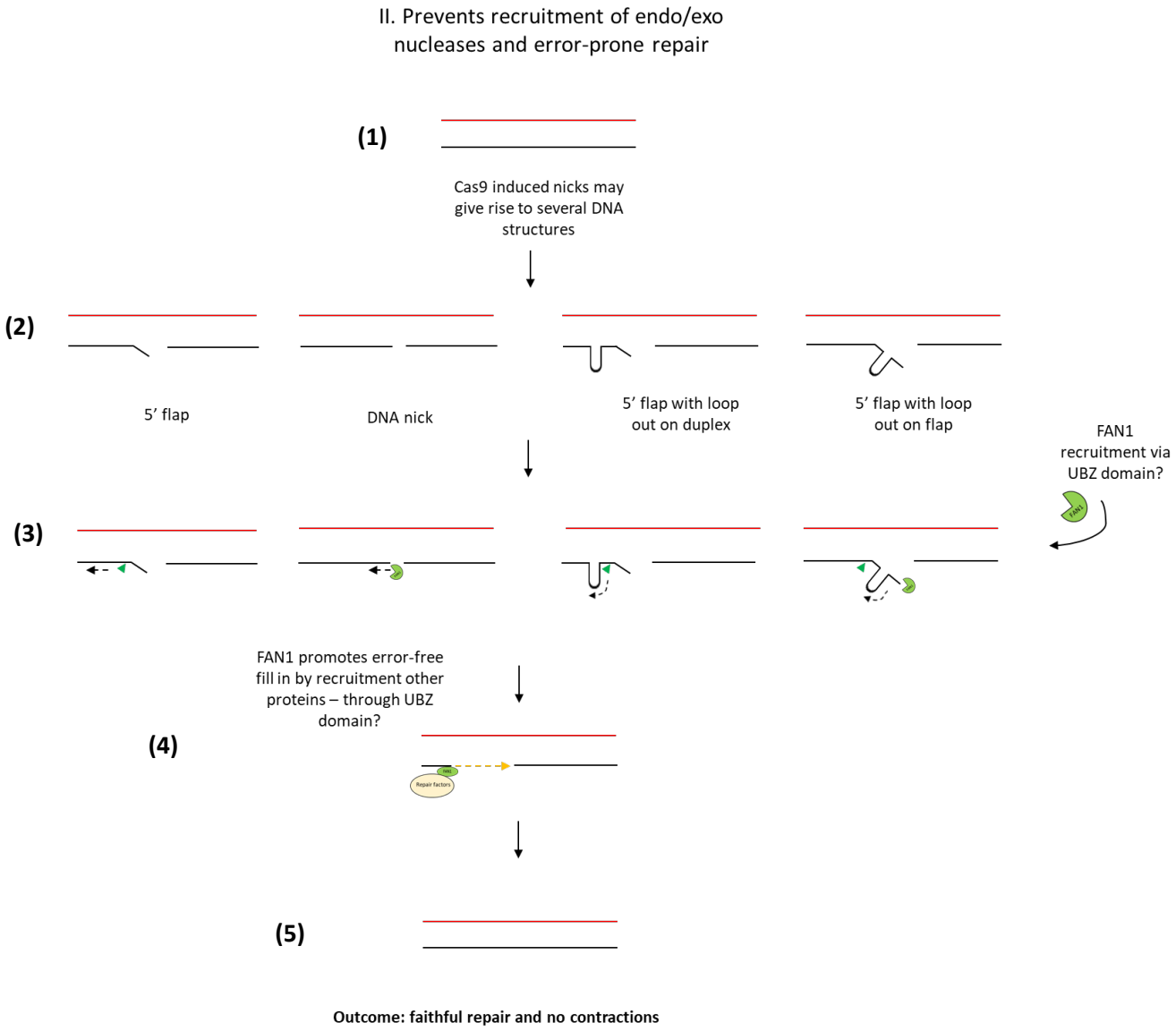


Figure 6-3: FAN1 faithfully repairs DNA substrates to prevent Cas9 nickase-induced contractions. *After induction of a DNA nick by Cas9 several DNA structures may form, including; a 5'flap, DNA nick or gap, and CAG/CTG loop-out structures. FAN1 may be recruited to these structures via association with its UBZ domain to other protein partners, or may localise to such structures independently. Binding of FAN1 to such structures likely occludes binding by other error-prone nucleases which can act at these sites. Endonuclease (green arrow) or exonuclease (black, dashed arrow) activity of FAN1 is likely required for processing of DNA structures. Downstream of this processing FAN1 may function to promoter error free repair by recruitment of faithful DNA polymerases. The result is faithful repair of Cas9-induced nicks and prevention of nickase-induced contractions.*

6.4 Wider Implications

To date, there is no disease-modifying treatment for any neurological CAG/CTG disorder, including HD. Whilst potential therapies have reached clinical trial stages, including ASOs targeting *HTT* and *mHTT*, these were halted due to adverse effects or a lack of target engagement (Kingwell 2021; Tabrizi et al. 2022). Furthermore, other proposed therapies for HD, or other CAG/CTG disorders, including siRNAs and CRISPR-Cas9 targeting of *HTT* are either; not allele-selective, target SNPs applicable to only a subset of patients or involve use of a Cas9 endonuclease targeting flanking sequences resulting in complete removal of all CAG/CTG repeats (Shin et al. 2016; Monteys et al. 2017; Yang et al. 2017; Ekman et al. 2019). The Cas9 nickase presents an allele-selective therapy which can contract only expanded repeats, leaving short non-pathogenic CAG/CTG repeats un-edited (Cinesi et al. 2016). As the distinct DNA repair mechanism causing Cas9-nickase induced contractions is unknown, it is important to understand how contractions are generated. This is important, not only in identifying the DNA repair mechanism involved but also can allow for screening of what factors may improve CAG/CTG contraction efficiency, without causing unwanted expansion events. The finding that loss of FAN1 actually promotes increased nickase-induced contraction events is significant, with implications not only for the mechanism of contractions but also suggests that patients with loss-of-function FAN1 variants will still benefit from this potential therapy.

Currently the therapeutic approach for this Cas9 nickase treatment would involve an adeno-associated virus (AAV) mediated delivery approach, via striatal injection. These provide benefits over lentiviral and retroviral delivery approaches as they do not integrate into the patient genome and are associated with a lower immune response (Mingozzi and High 2013). The AAV9 has been reported to demonstrate robust expression in striatal tissue (Marco et al. 2018). However, one current issue is that due to the limited packaging sizes of AAV's, the Cas9 nickase and sgCTG need to be packaged in separate AAVs, which will reduce the efficiency of the therapy (Wu et al. 2010). Therefore, further insight into how contractions are generated and what proteins are involved is useful. The finding that loss of FAN1 promotes

nickase-induced contractions may have implications for increasing the efficiency of this therapy. Such information may be useful as a co-therapeutic whereby a small molecule inhibitor may be delivered acutely at the same time as a AAV-based therapy to boost efficiency. If, in non-dividing neurons, the relatively rapid contractions observed in the HEK293 cells are recapitulated, it is likely that given the slow timescale of expansions compared to rapid contractions, this temporary inhibition may not have the long-term effects that chronic exposure to DNA repair inhibitors might have.

6.5 Limitations and Future work

One limitation of this thesis involves the FAN1 variant work investigated in Chapter 4. The method of complementing *FAN1*$^{-/-}$ cells with FAN1 variants, involved lentiviral transduction and FACS to identify cells that were stably transduced. Initially, the plan was to select cells with zeocin contained on the constructs to maintain expression. However, selection with zeocin was toxic to *FAN1*$^{-/-}$ cells, as such, stable selection of FAN1 variants could not be performed. Therefore, I did observe some loss of FAN1 expression in all lines as detected by western blot. Despite this, and some variability in protein expression between variant lines the data demonstrating that only the FAN1 wild-type variant, with the lowest protein expression, reduces nickase-induced contractions to *FAN1*$^{+/+}$ levels suggests insufficient protein expression was not the reason other variants failed to rescue cells. An alternative approach to this experiment may have been to use a TALEN-mediated gene-editing approach to create stable FAN1 variant lines. This approach would involve using TALEN homology arms, which target a safe-harbour site in the genome, such as the AASV1 or CLYBL site (González et al. 2014; Fernandopulle et al. 2018). In this system FAN1 variant constructs would be inserted into the same genomic region, and the copy number of variants would be the same, therefore presumably so too would protein expression. Furthermore, using a doxycycline inducible promoter, which has been demonstrated in iNeurons, controls for the timed expression of each variant.

As discussed, an initial aim of this thesis was to determine whether the phenotype demonstrated in HEK293 cells was recapitulated in non-dividing neurons. Due to difficulties in differentiation protocols this was not achieved. Therefore it is essential to determine whether the role of FAN1 in preventing nickase-induced contractions are relevant to neurons. If a loss of FAN1 in iPSC-derived neurons also yields an increase in nickase-induced contraction rates, we may conclude that similar protective mechanisms are at play in these cells. The work presented in Chapter 5 presents a foundation from which further work can be built upon. The generation of inducible iNeurons with both *FAN1*$^{+/+}$ and *FAN1*$^{-/-}$ lines present an invaluable tool for future work. To date, I have produced a stock of *FAN1* targeted CRISPR-edited cells, and have identified 4 homozygous knock-out clones, I do not expect that any major hurdles remain to bring this to fruition. If an increase in contractions is seen in *FAN1*$^{-/-}$ neurons, a logical next step would be to validate the variants investigated in Chapter 4. In Chapter 4 *FAN1*$^{-/-}$ cells were rescued with FAN1 variant constructs via lentiviral transduction. Whilst these results demonstrated that rescue with the wild-type variant reduced nickase-contraction rates, as expected, I did observe variation in protein expression.

The use of CRISPR-Cas9 presents a useful tool to investigate the role of FAN1 in preventing against nickase-induced contractions in iPSC-derived neurons. CRISPR-Cas9 editing with a HDR donor arm can allow for the insertion of a desired SNP, for example R507H, which appears critical in protecting against Cas9 nickase-induced contractions (Bloomer et al. 2022). This would be a useful experiment as the clones with this SNP under the endogenous *FAN1* promoter would negate the limitations of variable protein expression. These set of experiments would provide not only insight into repair mechanisms in non-dividing neurons, but also serve to further validate our HEK293 reporter cell line as a model system from which to monitor nickase-induced contractions and somatic instability.

Whilst the presence of CAG/CTG contractions in iPSC-derived neurons would be an essential proof-of-principle demonstrating the applicability of the Cas9 nickase as a therapy, it would be interesting to explore whether contractions in targeted cells lead to phenotypic recovery. To date, publications using the

HD109 iPSC lines to generate iPSC-derived neurons have demonstrated phenotypic abnormalities in these lines, compared with controls. These abnormalities include; bioenergetic deficits (Kedaigle et al. 2020), disruption of nucleocytoplasmic transport (Grima et al. 2017) and differences in neuronal resting membrane potentials indicating a delay in functional neuronal maturation (Mehta et al. 2018). These assays may be useful in comparing neurons with and without Cas9 nickase treatment. The creation of an isogenic control line for the HD109 iPSCs has been described (Stöberl et al. 2023). In these cells the expanded CAG repeats in *HTT* exon 1 were genetically corrected to a non-pathogenic length of 22 repeats. These provide a valuable tool allowing for comparison between disease and corrected control cells, from the same patient, where variation in genetics would not confound the data. Many studies have reported changes in transcriptomic profiling in HD iPSCs and iPSC-derived neurons (Kaye et al. 2022). Given the availability of an isogenic control line, an experiment that would be interesting would be to compare iPSC-derived neurons with or without Cas9 nickase treatment to isogenic controls using single-cell RNA sequencing. This would allow us to establish if treatment reverts disease-associated transcriptional phenotypes.

One phenotype that is hard to recapitulate in HD iPSC-derived neurons is the formation of mHTT inclusion bodies, a hallmark of disease pathology *in vivo*. Some studies have reported that HD-patient iPSC lines can form mHTT aggregates, but require treatment with proteasome and autophagy inhibitors or menadione to induce oxidative stress (Jeon et al. 2012; Liu et al. 2017b; Koyuncu et al. 2018). The inability to form mHTT aggregates under endogenous conditions has been reported to be due to the process of re-programming of HD-patient fibroblast to iPSCs, where re-programming alters epigenetic age-associated marks (Victor et al. 2018). One way to overcome this is direct conversion of HD-fibroblasts to neurons, therefore bypassing the pluripotency state, so cells retain their age. One study describes the directed differentiation of HD-fibroblasts to MSNs through micro-RNA based direct neural conversion (Victor et al. 2018). In these cells mHTT inclusion bodies were detected, without the requirement of exogenous stressors. These cells

may present a useful tool where we can monitor whether treatment with the Cas9 nickase prevents formation of mHTT inclusion.

6.6 Concluding remarks

Results from this thesis have demonstrated that, surprisingly, FAN1 acts to prevent CRISPR-Cas9 nickase induced contractions. Furthermore, it has led to insights regarding which FAN1 functions are required for this phenotype, indicating that FAN1 DNA-binding at the CAG/CTG repeat is essential for this protective effect. Both the nuclease function of FAN1 and the UBZ domain are also implicated. These data have advanced our understanding of how these contractions may be generated and provide a basis for future modelling of this phenotype in iPSC-derived cortical neurons. Furthermore, establishing that a loss of FAN1 enhances contractions while not promoting expansions suggests that patients with FAN1-damaging variants may benefit most from this therapy.

7 List of References

Adli, M. 2018. The CRISPR tool kit for genome editing and beyond. *Nature Communications* 9(1).

Akram, F. et al. 2023. An Insight into Modern Targeted Genome-Editing Technologies with a Special Focus on CRISPR/Cas9 and its Applications. *Molecular Biotechnology* 65(2), pp. 227–242.

Ambasudhan, R. et al. 2013. Isogenic Human iPSC Parkinson's Model Shows Nitrosative Stress-Induced Dysfunction in MEF2-PGC1α Transcription. *Cell* 155(6), pp. 1351–1364.

Andrade, M.A. and Bork, P. 1995. HEAT repeats in the Huntington's disease protein. *Nature Genetics 1995 11:2* 11(2), pp. 115–116.

Arber, C. et al. 2015. Activin a directs striatal projection neuron differentiation of human pluripotent stem cells. *Development (Cambridge)* 142(7), pp. 1375–1386.

Arrasate, M. et al. 2004. Inclusion body formation reduces levels of mutant huntingtin and the risk of neuronal death. *Nature* 431(7010), pp. 805–810.

Avazzadeh, S. et al. 2021. Modelling Parkinson's Disease: iPSCs towards Better Understanding of Human Pathology. *Brain Sciences 2021, Vol. 11, Page 373* 11(3), p. 373.

Aylward, E.H. et al. 1997. Longitudinal change in basal ganglia volume in patients with Huntington's disease. *Neurology* 48(2), pp. 394–399.

Aylward, E.H. et al. 1998. Frontal lobe volume in patients with Huntington's disease. *Neurology* 50(1), pp. 252–258. doi: 10.1212/WNL.50.1.252.

Bastarache, L. et al. 2018. Phenotype risk scores identify patients with unrecognized mendelian disease patterns. *Science* 359(6381), pp. 1233–1239.

Bastarache, L. et al. 2019. Mendelian disease patterns. 359(6381), pp. 1233–1239.

Bates, G.P. et al. 2015. Huntington disease. (April), pp. 1–21. doi: 10.1038/nrdp.2015.5.

Baydyuk, M. and Xu, B. 2014. BDNF signaling and survival of striatal neurons. *Frontiers in Cellular Neuroscience* 8(AUG), pp. 1–10.

Becker, S. and Boch, J. 2021. TALE and TALEN genome editing technologies. *Gene and Genome Editing* 2, p. 100007.

Beumer, K.J. et al. 2013. Comparing Zinc Finger Nucleases and Transcription Activator-Like Effector Nucleases for Gene Targeting in Drosophila. *G3: Genes, Genomes, Genetics* 3(9), pp. 1717–1725.

Bloomer, H. et al. 2022. CRISPR/Cas9 ribonucleoprotein-mediated genome and epigenome editing in mammalian cells. *Advanced Drug Delivery Reviews* 181, p. 114087.

Bobbin, M.L. and Rossi, J.J. 2016. RNA Interference (RNAi)-Based Therapeutics: Delivering on the Promise? *Annual Review of Pharmacology and Toxicology* 56, pp. 103–122.

Boch, J. and Bonas, U. 2010. Xanthomonas AvrBs3 family-type III effectors: Discovery and function. *Annual Review of Phytopathology* 48(Volume 48, 2010), pp. 419–436.

Brazdis, R.M. et al. 2020. Demonstration of brain region-specific neuronal vulnerability in human iPSC-based model of familial Parkinson's disease. *Human Molecular Genetics* 29(7),

pp. 1180–1191.

Van Den Broek, W.J.A.A. et al. 2002. Somatic expansion behaviour of the (CTG)n repeat in myotonic dystrophy knock-in mice is differentially affected by Msh3 and Msh6 mismatch–repair proteins. *Human Molecular Genetics* 11(2), pp. 191–198.

Brouns, S.J.J. et al. 2008. Small CRISPR RNAs guide antiviral defense in prokaryotes. *Science* 321(5891), pp. 960–964.

Bunting, E.L. et al. 2022. Polyglutamine diseases. *Current Opinion in Neurobiology* 72, pp. 39–47.

Busskamp, V. et al. 2014. Rapid neurogenesis through transcriptional activation in human stem cells. *Molecular Systems Biology* 10(11), pp. 1–21.

Caldecott, K.W. 2003. XRCC1 and DNA strand break repair. *DNA Repair* 2(9), pp. 955–969.

Caldecott, K.W. 2008. Single-strand break repair and genetic disease. *Nature Reviews Genetics* 9(8), pp. 619–631.

Cannavo, E. et al. 2005. Expression of the MutL Homologue hMLH3 in Human Cells and its Role in DNA Mismatch Repair. *Cancer Research* 65(23), pp. 10759–10766.

Cannavo, E. et al. 2007. Characterization of the interactome of the human MutL homologues MLH1, PMS1, and PMS2. *Journal of Biological Chemistry* 282(5), pp. 2976–2986.

Carmichael, O. and Lockhart, S. 2012. The Neuropathology of Huntington's Disease. *Brain Imaging in Behavioral Neuroscience* (November 2011), pp. 289–320.

Carthew, R.W. and Sontheimer, E.J. 2009. Origins and Mechanisms of miRNAs and siRNAs. *Cell* 136(4), pp. 642–655.

Castel, A.L. et al. 2010. Repeat instability as the basis for human diseases and as a potential target for therapy. *Nature Reviews Molecular Cell Biology 2010 11:3* 11(3), pp. 165–170.

Castella, M. et al. 2015. FANCI Regulates Recruitment of the FA Core Complex at Sites of DNA Damage Independently of FANCD2. *PLoS Genetics* 11(10), p. 1005563.

Cathomen, T. and Keith Joung, J. 2008. Zinc-finger nucleases: The next generation emerges. *Molecular Therapy* 16(7), pp. 1200–1207.

Caviston, J.P. et al. 2011. Huntingtin coordinates the dynein-mediated dynamic positioning of endosomes and lysosomes. *Molecular Biology of the Cell* 22(4), pp. 478–492.

Ceccaldi, R. et al. 2016. The Fanconi anaemia pathway: New players and new functions. *Nature Reviews Molecular Cell Biology* 17(6), pp. 337–349.

Cerbini, T. et al. 2015. Transcription Activator-Like Effector Nuclease (TALEN)-Mediated CLYBL Targeting Enables Enhanced Transgene Expression and One-Step Generation of Dual Reporter Human Induced Pluripotent Stem Cell (iPSC) and Neural Stem Cell (NSC) Lines. *PLOS ONE* 10(1), p. e0116032.

Chambers, S.M. et al. 2009. Highly efficient neural conversion of human ES and iPS cells by dual inhibition of SMAD signaling. *Nature Biotechnology* 27(3), pp. 275–280.

Chambers, S.M. et al. 2012. Combined small-molecule inhibition accelerates developmental timing and converts human pluripotent stem cells into nociceptors. *Nature Biotechnology* 30(7), pp. 715–720.

Chan, F. et al. 2011. Efficient mutagenesis of the rhodopsin gene in rod photoreceptor neurons in mice. *Nucleic Acids Research* 39(14), pp. 5955–5966.

Chaudhury, I. et al. 2014. FANCD2-Controlled Chromatin Access of the Fanconi-Associated Nuclease FAN1 Is Crucial for the Recovery of Stalled Replication Forks. *Molecular and Cellular Biology* 34(21), pp. 3939–3954.

Chavali, V.R.M. et al. 2020. Dual SMAD inhibition and Wnt inhibition enable efficient and reproducible differentiations of induced pluripotent stem cells into retinal ganglion cells. *Scientific Reports 2020 10:1* 10(1), pp. 1–14.

Chen, M. et al. 2020. Rapid generation of regionally specified CNS neurons by sequential patterning and conversion of human induced pluripotent stem cells. *Stem Cell Research* 48, p. 101945.

Chen, P.C. et al. 2005. Contributions by MutL homologues Mlh3 and Pms2 to DNA mismatch repair and tumor suppression in the mouse. *Cancer Research* 65(19), pp. 8662–8670.

Cho, S.W. et al. 2014. Analysis of off-target effects of CRISPR/Cas-derived RNA-guided endonucleases and nickases. *Genome Research* 24(1), pp. 132–141.

Cinesi, C. et al. 2016. Contracting CAG/CTG repeats using the CRISPR-Cas9 nickase. *Nature Communications* 7.

Ciosi, M. et al. 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, pp. 568–580.

Ciosi, M. et al. 2021. Approaches to Sequence the HTT CAG Repeat Expansion and Quantify Repeat Length Variation. Jones, L. et al. eds. *Journal of Huntington's Disease* 10(1), pp. 53–74.

Cong, L. et al. 2013. Multiplex genome engineering using CRISPR/Cas systems. *Science* 339(6121), pp. 819–823.

Cudkowicz, M. and Kowall, N.W. 1990. Degeneration of pyramidal projection neurons in Huntington's disease cortex. *Annals of Neurology* 27(2), pp. 200–204.

Cumming, S.A. et al. 2018. De novo repeat interruptions are associated with reduced somatic instability and mild or absent clinical features in myotonic dystrophy type 1. *European Journal of Human Genetics 2018 26:11* 26(11), pp. 1635–1647.

Datson, N.A. et al. 2017. The expanded CAG repeat in the huntingtin gene as target for therapeutic RNA modulation throughout the HD mouse brain. *PLOS ONE* 12(2), p. e0171127.

David, G. et al. 1997. Cloning of the SCA7 gene reveals a highly unstable CAG repeat expansion. *Nature Genetics 1997 17:1* 17(1), pp. 65–70.

Davies, S.W. et al. 1997. Formation of neuronal intranuclear inclusions underlies the neurological dysfunction in mice transgenic for the HD mutation. *Cell* 90(3), pp. 537–548.

Davies, S.W. et al. 1999. From neuronal inclusions to neurodegeneration: Neuropathological investigation of a transgenic mouse model of Huntington's disease. *Philosophical Transactions of the Royal Society B: Biological Sciences* 354(1386), pp. 971–979.

Davis, L. and Maizels, N. 2014. Homology-directed repair of DNA nicks via pathways distinct from canonical double-strand break repair. *Proceedings of the National Academy of Sciences of the United States of America* 111(10).

Davis, L. and Maizels, N. 2016. Two Distinct Pathways Support Gene Correction by Single-Stranded Donors at DNA Nicks. *Cell Reports* 17(7), pp. 1872–1881.

Deltcheva, E. et al. 2011. CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. *Nature* 471(7340), pp. 602–607.

Deshmukh, A.L. et al. 2021a. FAN1, a DNA Repair Nuclease, as a Modifier of Repeat Expansion Disorders. *Journal of Huntington's Disease* 10(1), pp. 95–122.

Deshmukh, A.L. et al. 2021b. FAN1 exo- not endo-nuclease pausing on disease-associated slipped-DNA repeats: A mechanism of repeat instability. *Cell Reports* 37(10).

DiFiglia, M. et al. 1995. Huntingtin is a cytoplasmic protein associated with vesicles in human and rat brain neurons. *Neuron* 14(5), pp. 1075–1081.

DiFiglia, M. et al. 1997. Aggregation of Huntingtin in Neuronal Intranuclear Inclusions and Dystrophic Neurites in Brain. *Science* 277(5334), pp. 1990–1993.

Dion, V. et al. 2008. Dnmt1 deficiency promotes CAG repeat expansion in the mouse germline. *Human Molecular Genetics* 17(9), pp. 1306–1317.

Djoussé, L. et al. 2003. Interaction of normal and expanded CAG repeat sizes influences age at onset of Huntington disease. *American Journal of Medical Genetics* 119 A(3), pp. 279–282.

Donaldson, J. et al. 2021. What is the Pathogenic CAG Expansion Length in Huntington's Disease. *Journal of Huntington's Disease* 10(1), pp. 175–202.

Douaud, G. et al. 2006. Distribution of grey matter atrophy in Huntington's disease patients: A combined ROI-based and voxel-based morphometric study. *NeuroImage* 32(4), pp. 1562–1575.

Dragileva, E. et al. 2009. Intergenerational and striatal CAG repeat instability in Huntington's disease knock-in mice involve different DNA repair genes. *Neurobiology of Disease* 33(1), pp. 37–47.

Drummond, J.T. et al. 1995. Isolation of an hMSH2-p160 Heterodimer That Restores DNA Mismatch Repair to Tumor Cells. *Science* 268(5219), pp. 1909–1912.

Dunah, A.W. et al. 2002. Sp1 and TAFII130 transcriptional activity disrupted in early Huntington's disease. *Science* 296(5576), pp. 2238–2243.

Den Dunnen, W.F.A. 2018. Trinucleotide repeat disorders. In: *Handbook of Clinical Neurology*. Elsevier B.V., pp. 383–391.

Duyao, M. et al. 1993. Trinucleotide repeat length instability and age of onset in Huntington's disease. *Nature Genetics* 4(4), pp. 387–392.

Duyao, M.P. et al. 1995. Inactivation of the Mouse Huntington's Disease Gene Homolog Hdh. *Science* 269(5222), pp. 407–410.

Ebbert, M.T.W. et al. 2018. Long-read sequencing across the C9orf72 'GGGGCC' repeat expansion: Implications for clinical use and genetic discovery efforts in human disease. *Molecular Neurodegeneration* 13(1), pp. 1–17.

Edelmann, W. et al. 1996. Meiotic Pachytene Arrest in MLH1-Deficient Mice Recent studies have shown that much of HNPCC is due to. *Cell* 85, pp. 1125–1134.

Ehrlich, M.E. 2012. Huntington's Disease and the Striatal Medium Spiny Neuron: Cell-Autonomous and Non-Cell-Autonomous Mechanisms of Disease. *Neurotherapeutics 2012 9:2* 9(2), pp. 270–284.

Eid, J. et al. 2009. Real-time DNA sequencing from single polymerase molecules. *Science* 323(5910), pp. 133–138.

Ekman, F.K. et al. 2019. CRISPR-Cas9-Mediated Genome Editing Increases Lifespan and Improves Motor Deficits in a Huntington's Disease Mouse Model. *Molecular Therapy - Nucleic Acids* 17, pp. 829–839.

Evans, M J and Kaufmant, M.H. 1981. Establishment in culture of pluripotential cells from mouse embryos. *Nature* 292(July), pp. 154–156.

Evans, S.J.W. et al. 2013. Prevalence of adult Huntington's disease in the UK based on diagnoses recorded in general practice records. *Journal of Neurology, Neurosurgery and Psychiatry* 84(10), pp. 1156–1160.

Evers, M.M. et al. 2011. Targeting several CAG expansion diseases by a single antisense oligonucleotide. *PLoS ONE* 6(9).

Evers, M.M. et al. 2018. AAV5-miHTT Gene Therapy Demonstrates Broad Distribution and Strong Human Mutant Huntingtin Lowering in a Huntington's Disease Minipig Model. *Molecular Therapy* 26(9), pp. 2163–2177.

Fang, L. et al. 2023. Haplotyping SNPs for allele-specific gene editing of the expanded huntingtin allele using long-read sequencing. *Human Genetics and Genomics Advances* 4(1).

Fernandopulle, M.S. et al. 2018. Transcription Factor–Mediated Differentiation of Human iPSCs into Neurons. *Current Protocols in Cell Biology* 79(1), p. e51.

Fink, K.D. et al. 2016. Allele-specific reduction of the mutant huntingtin allele using transcription activator-like effectors in human huntington's disease fibroblasts. *Cell Transplantation* 25(4), pp. 677–686.

Fishel, R. and Wilson, T. 1997. MutS homologs in mammalian cells. *Current Opinion in Genetics and Development* 7(1), pp. 105–113.

Flores-Rozas, H. and Kolodner, R.D. 1998. The Saccharomyces cerevisiae MLH3 gene functions in MSH3-dependent suppression of frameshift mutations. *Proceedings of the National Academy of Sciences of the United States of America* 95(21), p. 12404.

Gafni, J. and Ellerby, L.M. 2002. Calpain Activation in Huntington's Disease. *Journal of Neuroscience* 22(12), pp. 4842–4849.

Gan, R. 2014. Comparing ESC and iPSC — Based Models for Human. *Journal of Clinical Medicine* 3, pp. 1146–1162.

Garber, K.B. et al. 2008. Fragile X syndrome. *European Journal of Human Genetics* 16(6), pp. 666–672.

Garriga-Canut, M. et al. 2012. Synthetic zinc finger repressors reduce mutant huntingtin expression in the brain of R6/2 mice. *Proceedings of the National Academy of Sciences of the United States of America* 109(45), pp. E3136–E3145.

Gasset-Rosa, F. et al. 2017. Polyglutamine-expanded huntingtin exacerbates age-related disruption of nuclear integrity and nucleocytoplasmic transport. *Neuron* 94(1), p. 48.

Gauthier, L.R. et al. 2004. Huntingtin controls neurotrophic support and survival of neurons by enhancing BDNF vesicular transport along microtubules. *Cell* 118(1), pp. 127–138.

Gavrilov, K. and Saltzman, W.M. 2012. Therapeutic siRNA: Principles, challenges, and strategies. *Yale Journal of Biology and Medicine* 85(2), pp. 187–200.

Genschel, J. et al. 1998. Isolation of MutSβ from human cells and comparison of the mismatch repair specificities of MutSβ and MutSα. *Journal of Biological Chemistry* 273(31), pp. 19895–19901.

Gerfen, C.R. 1992. The neostriatal mosaic: Multiple levels of compartmental organization in the basal ganglia. *Annual Review of Neuroscience* 15(Volume 15, 1992), pp. 285–320.

Giesselmann, P. et al. 2019. Analysis of short tandem repeat expansions and their methylation state with nanopore sequencing. *Nature Biotechnology 2019 37:12* 37(12), pp. 1478–1481.

Goldberg, Y.P. et al. 1993. Molecular analysis of new mutations for Huntington's disease: intermediate alleles and sex of origin effects. *Nature Genetics 1993 5:2* 5(2), pp. 174–179.

Goldberg, Y.P. et al. 1996. Cleavage of huntingtin by apopain, a proapoptotic cysteine protease, is modulated by the polyglutamine tract. *Nature Genetics 1996 13:4* 13(4), pp. 442–449.

Gomes-Pereira, Mario Bidichandani, S. and Monckton, D.G. 2004. Analysis of Unstable Triplet Repeats Using Small-Pool Polymerase Chain Reaction. *Methods in Molecular Biology, Trinucleotide Repeat Protocols* , pp. 61–76.

Gomes-pereira, M. et al. 2004. Small-Pool Polymerase Chain Reaction. 277, pp. 61–76.

Gomes-Pereira, M. et al. 2004. Pms2 is a genetic enhancer of trinucleotide CAG·CTG repeat somatic mosaicism: implications for the mechanism of triplet repeat expansion. *Human Molecular Genetics* 13(16), pp. 1815–1825.

Gonitel, R. et al. 2008. DNA instability in postmitotic neurons. *Proceedings of the National Academy of Sciences of the United States of America* 105(9), pp. 3467–3472.

González, F. et al. 2014. An iCRISPR platform for rapid, multiplexable, and inducible genome editing in human pluripotent stem cells. *Cell Stem Cell* 15(2), pp. 215–226.

Goold, R. et al. 2019. FAN1 modifies Huntington's disease progression by stabilizing the expanded HTT CAG repeat. *Human Molecular Genetics* 28(4), pp. 650–661.

Goold, R. et al. 2021. FAN1 controls mismatch repair complex assembly via MLH1 retention to stabilize CAG repeat expansion in Huntington's disease. *Cell Reports* 36(9), p. 109649.

Gorbunova, V. et al. 2003. Selectable System for Monitoring the Instability of CTG/CAG Triplet Repeats in Mammalian Cells. *Molecular and Cellular Biology* 23(13), pp. 4485–4493.

Graveland, G.A. et al. 1985. Evidence for degenerative and regenerative changes in neostriatal spiny neurons in Huntington's disease. *Science* 227(4688), pp. 770–773.

Graybiel, A.M. 1998. The Basal Ganglia and Chunking of Action Repertoires. *Neurobiology of Learning and Memory* 70(1–2), pp. 119–136.

Grima, J.C. et al. 2017. Mutant Huntingtin Disrupts the Nuclear Pore Complex. *Neuron*

94(1), pp. 93-107.e6.

Guo, Z. et al. 2012. Striatal neuronal loss correlates with clinical motor impairment in Huntington's disease. *Movement Disorders* 27(11), pp. 1379–1386.

Gusella, J.F. et al. 1983. A polymorphic DNA marker genetically linked to Huntington's disease. *Nature* 306(5940), pp. 234–238.

Gutekunst, C.A. et al. 1999. Nuclear and Neuropil Aggregates in Huntington's Disease: Relationship to Neuropathology. *The Journal of Neuroscience* 19(7), p. 2522.

Hadzi, T.C. et al. 2012. Assessment of cortical and striatal involvement in 523 Huntington disease brains. *Neurology* 79(16), p. 1708.

Harley, H.G. et al. 1992. Unstable DNA sequence in myotonic dystrophy. *The Lancet* 339(8802), pp. 1125–1128.

Hartfield, E.M. et al. 2014. Physiological Characterisation of Human iPS-Derived Dopaminergic Neurons. *PLOS ONE* 9(2), p. e87388.

Hayward, B.E. et al. 2020. A point mutation in the nuclease domain of MLH3 eliminates repeat expansions in a mouse stem cell model of the Fragile X-related disorders. *Nucleic Acids Research* 48(14), pp. 7856–7863.

Henley, S.M.D. et al. 2009. Whole-brain atrophy as a measure of progression in premanifest and early Huntington's disease. *Movement Disorders* 24(6), pp. 932–936.

Her, L.S. and Goldstein, L.S.B. 2008. Enhanced Sensitivity of Striatal Neurons to Axonal Transport Defects Induced by Mutant Huntingtin. *Journal of Neuroscience* 28(50), pp. 13662–13672.

Hermel, E. et al. 2004. Specific caspase interactions and amplification are involved in selective neuronal vulnerability in Huntington's disease. *Cell Death & Differentiation 2004 11:4* 11(4), pp. 424–438.

Herndon, E.S. et al. 2009. Neuroanatomic profile of polyglutamine immunoreactivity inhuntington disease brains. *Journal of Neuropathology and Experimental Neurology* 68(3), pp. 250–261.

Hilditch-Maguire, P. et al. 2000. Huntingtin: an iron-regulated protein essential for normal nuclear and perinuclear organelles. *Human Molecular Genetics* 9(19), pp. 2789–2797.

Hobbs, N.Z. et al. 2010. Onset and progression of pathologic atrophy in Huntington disease: A longitudinal MR imaging study. *American Journal of Neuroradiology* 31(6), pp. 1036–1041.

Hodges, A. et al. 2006. Regional and cellular gene expression changes in human Huntington's disease brain. *Human Molecular Genetics* 15(6), pp. 965–977.

Hodgson, J.G. et al. 1999. A YAC mouse model for Huntington's disease with full-length mutant huntingtin, cytoplasmic toxicity, and selective striatal neurodegeneration. *Neuron* 23(1), pp. 181–192.

Hoffmann, D. et al. 2017. Detailed comparison of retroviral vectors and promoter configurations for stable and high transgene expression in human induced pluripotent stem cells. *Gene Therapy* 24(5), pp. 298–307.

Höijer, I. et al. 2018. Detailed analysis of HTT repeat elements in human blood using

targeted amplification-free long-read sequencing. *Human Mutation* 39(9), pp. 1262–1272.

Holmans, P.A. et al. 2017. Genetic modifiers of Mendelian disease: Huntington's disease and the trinucleotide repeat disorders. *Human Molecular Genetics* 26(R2), pp. R83–R90.

Holmes, S.E. et al. 1999. Expansion of a novel CAG trinucleotide repeat in the 5' region of PPP2R2B is associated with SCA12. *Nature Genetics* 23(4), pp. 391–392.

Holmes, S.E. et al. 2001. A repeat expansion in the gene encoding junctophilin-3 is associated with Huntington disease–like 2. *Nature Genetics 2001 29:4* 29(4), pp. 377–378.

Hong, E.P. et al. 2021. Huntington's Disease Pathogenesis: Two Sequential Components. *Journal of Huntington's Disease* 10(1), pp. 35–51. doi: 10.3233/JHD-200427.

Huang, K. et al. 2004. Huntingtin-interacting protein HIP14 is a palmitoyl transferase involved in palmitoylation and trafficking of multiple neuronal proteins. *Neuron* 44(6), pp. 977–986.

Hulme, A.J. et al. 2022. Making neurons, made easy: The use of Neurogenin-2 in neuronal differentiation. *Stem Cell Reports* 17(1), p. 14.

Huntington, G. 1967. On chorea. *Archives of Neurology* 17(3), pp. 332–333.

Ionita-Laza, I. et al. 2014. Scan statistic-based analysis of exome sequencing data identifies FAN1 at 15q13.3 as a susceptibility gene for schizophrenia and autism. *Proceedings of the National Academy of Sciences of the United States of America* 111(1), pp. 343–348.

Ishiguro, H. et al. 2001. Age-dependent and tissue-specific CAG repeat instability occurs in mouse knock-in for a mutant Huntington's disease gene. *Journal of Neuroscience Research* 65(4), pp. 289–297.

Ishino, Y. et al. 1987. Nucleotide sequence of the iap gene, responsible for alkaline phosphatase isozyme conversion in Escherichia coli, and identification of the gene product. *Journal of Bacteriology* 169(12), pp. 5429–5433.

Iyer, R.R. et al. 2006. DNA mismatch repair: Functions and Mechanisms. *Chem ReV* 106, pp. 302–323.

Iyer, R.R. and Pluciennik, A. 2021. DNA Mismatch Repair and its Role in Huntington's Disease. *Journal of Huntington's Disease* 10(1), p. 75.

Jansen, R. et al. 2002. Identification of genes that are associated with DNA repeats in prokaryotes. *Molecular Microbiology* 43(6), pp. 1565–1575.

Jayaraman, M. et al. 2012. Slow Amyloid Nucleation via α-Helix-Rich Oligomeric Intermediates in Short Polyglutamine-Containing Huntingtin Fragments. *Journal of Molecular Biology* 415(5), pp. 881–899.

Jeon, I. et al. 2012. Neuronal Properties, In Vivo Effects, and Pathology of a Huntington's Disease Patient-Derived Induced Pluripotent Stem Cells. *Stem Cells* 30(9), pp. 2054–2062.

Jiang, W. et al. 2013. RNA-guided editing of bacterial genomes using CRISPR-Cas systems. *Nature Biotechnology* 31(3), pp. 233–239.

Jin, H. and Cho, Y. 2017. Structural and functional relationships of FAN1. *DNA Repair* 56, pp. 135–143.

Jinek, M. et al. 2012. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* 337(6096), pp. 816–821.

Jiricny, J. 2013. Postreplicative mismatch repair. *Cold Spring Harbor Perspectives in Biology* 5(4), pp. 1–23.

Johnston, J.G. et al. 1990. Mechanisms of striatal pattern formation: conservation of mammalian compartmentalization. *Developmental Brain Research* 57(1), pp. 93–102.

Joung, J.K. and Sander, J.D. 2013. TALENs: A widely applicable technology for targeted genome editing. *Nature Reviews Molecular Cell Biology* 14(1), pp. 49–55.

Kawaguchi, Y. et al. 1994. CAG expansions in a novel gene for Machado-Joseph disease at chromosome 14q32.1. *Nature Genetics 1994 8:3* 8(3), pp. 221–228.

Kay, C. et al. 2017. Epidemiology of Huntington disease. In: *Handbook of Clinical Neurology*. Elsevier B.V., pp. 31–46.

Kaye, J. et al. 2022. Huntington's disease iPSC models—using human patient cells to understand the pathology caused by expanded CAG repeats. *Faculty Reviews* 11.

Kedaigle, A.J. et al. 2020. Bioenergetic deficits in Huntington's disease iPSC-derived neural cells and rescue with glycolytic metabolites. *Human Molecular Genetics* 29(11), p. 1757.

Kennedy, L. et al. 2003. Dramatic tissue-specific mutation length increases are an early molecular event in Huntington disease pathogenesis. *Human Molecular Genetics* 12(24), pp. 3359–3367.

Kennedy, L. and Shelbourne, P.F. 2000. *Dramatic mutation instability in HD mouse striatum: does polyglutamine load contribute to cell-specific vulnerability in Huntington's disease?*

Keskin, S. et al. 2019. AAV5-miHTT Lowers Huntingtin mRNA and Protein without Off-Target Effects in Patient-Derived Neuronal Cultures and Astrocytes. *Molecular Therapy - Methods and Clinical Development* 15(December), pp. 275–284.

Khristich, A.N. and Mirkin, S.M. 2020. On the wrong DNA track: Molecular mechanisms of repeat-mediated genome instability. *Journal of Biological Chemistry* 295(13), pp. 4134–4170.

Kim, J.H. et al. 2011. High Cleavage Efficiency of a 2A Peptide Derived from Porcine Teschovirus-1 in Human Cell Lines, Zebrafish and Mice. *PLOS ONE* 6(4), p. e18556.

Kim, J.M. et al. 2008. Cell cycle–dependent chromatin loading of the Fanconi anemia core complex by FANCM/FAAP24. *Blood* 111(10), p. 5215.

Kim, K.H. et al. 2020. Genetic and Functional Analyses Point to FAN1 as the Source of Multiple Huntington Disease Modifier Effects. *American Journal of Human Genetics* 107(1), pp. 96–110.

Kim, Y.G. et al. 1996. Hybrid restriction enzymes: Zinc finger fusions to Fok I cleavage domain. *Proceedings of the National Academy of Sciences of the United States of America* 93(3), pp. 1156–1160.

Kim, Y.J. et al. 2001. Caspase 3-cleaved N-terminal fragments of wild-type and mutant huntingtin are present in normal and Huntington's disease brains, associate with membranes, and undergo calpaindependent proteolysis. *Proceedings of the National Academy of Sciences of the United States of America* 98(22), pp. 12784–12789.

Kinch, L.N. et al. 2005. Identification of novel restriction endonuclease-like fold families among hypothetical proteins. *Nucleic Acids Research* 33(11), pp. 3598–3605.

Kingwell, K. 2021. Double setback for ASO trials in Huntington disease. *Nature reviews. Drug discovery* 20(6), pp. 412–413.

Klug, A. 2010. The discovery of zinc fingers and their development for practical applications in gene regulation and genome manipulation. *Quarterly Reviews of Biophysics* 43(1), pp. 1–21.

Koide, R. et al. 1994. Unstable expansion of CAG repeat in hereditary dentatorubral–pallidoluysian atrophy (DRPLA). *Nature Genetics 1994 6:1* 6(1), pp. 9–13.

Koob, M.D. et al. 1999. An untranslated CTG expansion causes a novel form of spinocerebellar ataxia (SCA8). *Nature Genetics 1999 21:4* 21(4), pp. 379–384.

Kordasiewicz, H.B. et al. 2012. Sustained Therapeutic Reversal of Huntington's Disease by Transient Repression of Huntingtin Synthesis. *Neuron* 74(6), pp. 1031–1044.

Kovalenko, M. et al. 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* 7(9), p. e44273.

Kovtun, I. V. and McMurray, C.T. 2001. Trinucleotide expansion in haploid germ cells by gap repair. *Nature Genetics* 27(4), pp. 407–411.

Koyuncu, S. et al. 2018. The ubiquitin ligase UBR5 suppresses proteostasis collapse in pluripotent stem cells from Huntington's disease patients. *Nature Communications* 9(1), pp. 1–22.

Kramer, W. et al. 1989. Cloning and nucleotide sequence of DNA mismatch repair gene PMS1 from Saccharomyces cerevisiae: homology of PMS1 to procaryotic MutL and HexB. *Journal of Bacteriology* 171(10), p. 5339.

Kratz, K. et al. 2010. Deficiency of FANCD2-Associated Nuclease KIAA1018/FAN1 Sensitizes Cells to Interstrand Crosslinking Agents. *Cell* 142(1), pp. 77–88.

Krejci, L. et al. 2012. Homologous recombination and its regulation. *Nucleic Acids Research* 40(13), pp. 5795–5818. doi: 10.1093/nar/gks270.

Kremer, B. et al. 1993. Molecular analysis of late onset Huntington's disease. *Journal of Medical Genetics* 30(12), pp. 991–995.

Kremer, B. et al. 1995. Sex-dependent mechanisms for expansions and contractions of the CAG repeat on affected Huntington disease chromosomes. *American Journal of Human Genetics* 57(2), pp. 343–350.

Kuemmerle, S. et al. 1999. Huntingtin aggregates may not predict neuronal death in Huntington's disease. *Annals of Neurology* 46(6), pp. 842–849.

van Kuilenburg, A.B.P. et al. 2019. Glutaminase Deficiency Caused by Short Tandem Repeat Expansion in GLS . *New England Journal of Medicine* 380(15), pp. 1433–1441.

De La Monte, S.M. et al. 1988. Morphometric Demonstration of Atrophic Changes in the Cerebral Cortex, White Matter, and Neostriatum in Huntington's Disease. *Journal of Neuropathology & Experimental Neurology* 47(5), pp. 516–525.

Labbadia, J. and Morimoto, R.I. 2013. Huntington's disease: Underlying molecular mechanisms and emerging concepts. *Trends in Biochemical Sciences* 38(8), pp. 378–385.

Lachaud, C. et al. 2016. Ubiquitinated Fancd2 recruits Fan1 to stalled replication forks to

prevent genome instability. *Science* 351(6275), pp. 846–849.

Landles, C. et al. 2010. Proteolysis of mutant huntingtin produces an exon 1 fragment that accumulates as an aggregated protein in neuronal nuclei in huntington disease. *Journal of Biological Chemistry* 285(12), pp. 8808–8823.

Lange, H. et al. 1976. Morphometric studies of the neuropathological changes in choreatic diseases. *Journal of the Neurological Sciences* 28(4), pp. 401–425.

Leavitt, B.R. et al. 2006. Wild-type huntingtin protects neurons from excitotoxicity. *Journal of Neurochemistry* 96(4), pp. 1121–1129.

Lee, J.M. et al. 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* 6(8).

Lee, J.M. et al. 2015. Identification of Genetic Factors that Modify Clinical Onset of Huntington's Disease. *Cell* 162(3), pp. 516–526.

Lee, J.M. et al. 2017. A modifier of Huntington's disease onset at the MLH1 locus. *Human Molecular Genetics* 26(19), pp. 3859–3867.

Lee, J.M. et al. 2019. CAG Repeat Not Polyglutamine Length Determines Timing of Huntington's Disease Onset. *Cell* 178(4), pp. 887-900.e14.

Li, G.B. and Lu, G.X. 2009. Gene Delivery Efficiency in Bone Marrow-derived Dendritic Cells: Comparison of Four Methods and Optimization for Lentivirus Transduction. *Molecular Biotechnology* 43(3), pp. 250–256.

Li, L. et al. 1992. Functional domains in Fok I restriction endonuclease. *Proceedings of the National Academy of Sciences of the United States of America* 89(10), p. 4275.

Li, S.H. et al. 1993. Huntington's disease gene (IT15) is widely expressed in human and rat tissues. *Neuron* 11(5), pp. 985–993.

Li, X. and Heyer, W.D. 2008. Homologous recombination in DNA repair and DNA damage tolerance. *Cell Research 2008 18:1* 18(1), pp. 99–113

Liang, X. et al. 2017. Enhanced CRISPR/Cas9-mediated precise genome editing by improved design and delivery of gRNA, Cas9 nuclease, and donor DNA. *Journal of Biotechnology* 241, pp. 136–146.

Liao, J.Q. et al. 2022. Generation of Monoclonal iPSC Lines with Stable Cas9 Expression and High Cas9 Activity. In: *Methods in Molecular Biology*. Humana Press Inc., pp. 575–588.

Liao, M.C. et al. 2016. Single-cell detection of secreted Aβ and sAPPα from human IPSC-derived neurons and astrocytes. *Journal of Neuroscience* 36(5), pp. 1730–1746.

Lieberman, A.P. et al. 2019. Polyglutamine Repeats in Neurodegenerative Diseases. *Annual Review of Pathology: Mechanisms of Disease* 14(1), pp. 1–27.

Lim, R.G. et al. 2017. Developmental alterations in Huntington's disease neural cells and pharmacological rescue in cells and mice. *Nature Neuroscience 2017 20:5* 20(5), pp. 648–660.

Lin, B. et al. 1993. Differential 3' polyadenylation of the Huntington disease gene results in two mRNA species with variable tissue expression. *Human Molecular Genetics* 2(10), pp. 1541–1545.

Lin, Y.C. et al. 2014. Genome dynamics of the human embryonic kidney 293 lineage in response to cell biology manipulations. *Nature Communications* 5(1), p. 12.

Lipkin, S.M. et al. 2002. Meiotic arrest and aneuploidy in MLH3-deficient mice. *Nature Genetics* 31(4), pp. 385–390.

Liu, D. et al. 2017a. DNA mismatch repair and its many roles in eukaryotic cells. *Mutation Research - Reviews in Mutation Research* 773, pp. 174–187.

Liu, G. et al. 2010a. Replication-dependent instability at (CTG)·(CAG) repeat hairpins in human cells. *Nature Chemical Biology* 6(9), pp. 652–659.

Liu, T. et al. 2010b. FAN1 acts with FANCI-FANCD2 to promote DNA interstrand cross-link repair. *Science* 329(5992), pp. 693–696.

Liu, Y. et al. 2017b. FOXOs modulate proteasome activity in human-induced pluripotent stem cells of Huntington's disease and their derived neural cells. *Human Molecular Genetics* 26(22), pp. 4416–4428.

Loomis, E.W. et al. 2013. Sequencing the unsequenceable: Expanded CGG-repeat alleles of the fragile X gene. *Genome Research* 23(1), pp. 121–128

Loupe, J.M. et al. 2020. Promotion of somatic CAG repeat expansion by Fan1 knock-out in Huntington's disease knock-in mice is blocked by Mlh1 knock-out. *Human Molecular Genetics* 00(00), pp. 1–10.

Lunkes, A. et al. 2002. Proteases acting on mutant huntingtin generate cleaved products that differentially build up cytoplasmic and nuclear inclusions. *Molecular Cell* 10(2), pp. 259–269.

MacDonald, M.E. et al. 1993. A novel gene containing a trinucleotide repeat that is expanded and unstable on Huntington's disease chromosomes. *Cell* 72(6), pp. 971–983.

Macdonald, V. and Halliday, G. 2002. Pyramidal Cell Loss in Motor Cortices in Huntington's Disease. *Neurobiology of Disease* 10(3), pp. 378–386.

MacKay, C. et al. 2010. Identification of KIAA1018/FAN1, a DNA Repair Nuclease Recruited to DNA Damage by Monoubiquitinated FANCD2. *Cell* 142(1), pp. 65–76.

Malla, B. et al. 2021. A Systematic Review of Transcriptional Dysregulation in Huntington's Disease Studied by RNA Sequencing. *Frontiers in Genetics* 12(October), pp. 1–22.

Mangiarini, L. et al. 1996. Exon I of the HD gene with an expanded CAG repeat is sufficient to cause a progressive neurological phenotype in transgenic mice. *Cell* 87(3), pp. 493–506.

Mangiarini, L. et al. 1997. Instability of highly expanded CAG repeats in mice transgenic for the Huntington's disease mutation. *Nature Genetics* 15(2), pp. 197–200.

Mangin, A. et al. 2021. Robust detection of somatic mosaicism and repeat interruptions by long-read targeted sequencing in myotonic dystrophy type 1. *International Journal of Molecular Sciences* 22(5), pp. 1–24.

Manley, K. et al. 1999. Msh2 deficiency prevents in vivo somatic instability of the CAG repeat in Huntington disease transgenic mice. *Nature Genetics* 23(4), pp. 471–473.

Marco, S. et al. 2018. RNAi-Based GluN3A Silencing Prevents and Reverses Disease Phenotypes Induced by Mutant huntingtin. *Molecular Therapy* 26(8), p. 1965.

Margulis, J. and Finkbeiner, S. 2014. Proteostasis in striatal cells and selective

neurodegeneration in Huntington's disease. *Frontiers in Cellular Neuroscience* 8(AUG), p. 93058.

Marquis Gacy, A. et al. 1995. Trinucleotide repeats that expand in human disease form hairpin structures in vitro. *Cell* 81(4), pp. 533–540.

Martin, G.R. 1981. Isolation of a pluripotent cell line from early mouse embryos cultured in medium conditioned by teratocarcinoma stem cells Developmental Biology : *Proc Natl Acad Sci U S A* 78(12), pp. 7634–7638.

Massey, T. et al. 2019. *Methods for Assessing DNA Repair and Repeat Expansion in Huntington's Disease*.

Mattis, V.B. et al. 2012. Induced pluripotent stem cells from patients with huntington's disease show CAG repeat expansion associated phenotypes. *Cell Stem Cell* 11(2), pp. 264–278.

McAllister, B. et al. 2021. Timing and Impact of Psychiatric, Cognitive, and Motor Abnormalities in Huntington Disease. *Neurology* 96(19), p. e2395.

McAllister, B. et al. 2022. Exome sequencing of individuals with Huntington's disease implicates FAN1 nuclease activity in slowing CAG expansion and disease onset. *Nature Neuroscience* 25(4), pp. 446–457.

McColgan, P. and Tabrizi, S.J. 2018. Huntington's disease: a clinical review. *European Journal of Neurology* 25(1), pp. 24–34.

McGinty, R.J. and Mirkin, S.M. 2018. Cis- and Trans-Modifiers of Repeat Expansions: Blending Model Systems with Human Genetics. *Trends in Genetics* 34(6), pp. 448–465.

McMurray, C.T. 2010. Mechanisms of trinucleotide repeat instability during human development. *Nature Reviews Genetics* 11(11), pp. 786–799.

Meetei, A.R. et al. 2005. A human ortholog of archaeal DNA repair protein Hef is defective in Fanconi anemia complementation group M. *Nature Genetics* 37(9), pp. 958–963.

Mehta, S.R. et al. 2018. Human Huntington's Disease iPSC-Derived Cortical Neurons Display Altered Transcriptomics, Morphology, and Maturation. *Cell Reports* 25(4), pp. 1081-1096.e6.

Metzger, M.J. et al. 2011. Single-strand nicks induce homologous recombination with less toxicity than double-strand breaks using an AAV vector template. *Nucleic Acids Research* 39(3), pp. 926–935.

Mingozzi, F. and High, K.A. 2013. Immune responses to AAV vectors: overcoming barriers to successful gene therapy. *Blood* 122(1), p. 23.

Miniarikova, J. et al. 2017. AAV5-miHTT gene therapy demonstrates suppression of mutant huntingtin aggregation and neuronal dysfunction in a rat model of Huntington's disease. *Gene Therapy* 24(10), p. 630.

Minoguchi, S. and Iba, H. 2008. Instability of Retroviral DNA Methylation in Embryonic Stem Cells. *Stem Cells* 26(5), pp. 1166–1173. doi: 10.1634/stemcells.2007-1106.

Mirkin, S.M. 2007. Expandable DNA repeats and human disease. *Nature* 447(7147), pp. 932–940.

Modrich, P. 2006. Mechanisms in eukaryotic mismatch repair. *Journal of Biological*

Chemistry 281(41), pp. 30305–30309.

Modrich, P. and Lahue, R. 1996. Mismatch repair in replication fidelity, genetic recombination, and cancer biology. *Annual Review of Biochemistry* 65, pp. 101–133.

Mojica, F.J.M. et al. 2005. Intervening sequences of regularly spaced prokaryotic repeats derive from foreign genetic elements. *Journal of Molecular Evolution* 60(2), pp. 174–182.

Møllersen, L. et al. 2010. Continuous and periodic expansion of CAG repeats in huntington's disease R6/1 mice. *PLoS Genetics* 6(12), pp. 1–11.

Monckton, D.G. et al. 1995. Somatic mosaicism, germline expansions, germline reversions and intergenerational reductions in myotonic dystrophy males: small pool PCR analyses. *Human Molecular Genetics* 4(1), pp. 1–8.

Monteys, A.M. et al. 2017. CRISPR/Cas9 Editing of the Mutant Huntingtin Allele In Vitro and In Vivo. *Molecular Therapy* 25(1), pp. 12–23.

Nagase, T. et al. 1999. Prediction of the coding sequences of unidentified human genes. XIII. The complete sequences of 100 new cDNA clones from brain which code for large proteins in vitro. *DNA Research* 6, pp. 63–70.

Nahhas, F.A. et al. 2005. Juvenile onset Huntington disease resulting from a very large maternal expansion. *American Journal of Medical Genetics Part A* 137A(3), pp. 328–331.

Nakamura, K. et al. 2001. SCA17, a novel autosomal dominant cerebellar ataxia caused by an expanded polyglutamine in TATA-binding protein. *Human Molecular Genetics* 10(14), pp. 1441–1448.

Nieweg, K. et al. 2015. Alzheimer's disease-related amyloid-β induces synaptotoxicity in human iPS cell-derived neurons. *Cell Death & Disease* 6(4), p. e1709.

Nishiyama, J. et al. 2017. Virus-Mediated Genome Editing via Homology-Directed Repair in Mitotic and Postmitotic Cells in Mammalian Brain. *Neuron* 96(4), pp. 755-768.e5.

Nithianantharajah, J. and Hannan, A.J. 2013. Dysregulation of synaptic proteins, dendritic spine abnormalities and pathological plasticity of synapses as experience-dependent mediators of cognitive and psychiatric symptoms in Huntington's disease. *Neuroscience* 251, pp. 66–74.

Nizzardo, M. et al. 2010. Human motor neuron generation from embryonic stem cells and induced pluripotent stem cells. *Cellular and Molecular Life Sciences* 67(22), pp. 3837–3847.

Nørremølle, A. et al. 1995. Correlation between magnitude of CAG repeat length alterations and length of the paternal repeat in paternally inherited Huntington's disease. *Clinical Genetics* 47(3), pp. 113–117.

Orr, H.T. et al. 1993. Expansion of an unstable trinucleotide CAG repeat in spinocerebellar ataxia type 1. *Nature Genetics 1993 4:3* 4(3), pp. 221–226.

Owen, B.A.L. et al. 2005. (CAG)n-hairpin DNA binds to Msh2-Msh3 and changes properties of mismatch recognition. *Nature Structural and Molecular Biology* 12(8), pp. 663–670.

Paine, H. 2015. Does loss of the normal protein function contribute to the pathogenesis of Huntington's disease? *Bioscience Horizons* 8, pp. 1–9.

Palombo, F. et al. 1996. hMutSβ, a heterodimer of hMSH2 and hMSH3, binds to insertion/deletion loops in DNA. *Current Biology* 6(9), pp. 1181–1184.

Pan, F. et al. 2017. Structure and Dynamics of DNA and RNA Double Helices of CAG and GAC Trinucleotide Repeats. *Biophysical Journal* 113(1), pp. 19–36.

Park, S.H. et al. 2013. PolyQ proteins interfere with nuclear degradation of cytosolic proteins by sequestering the Sis1p chaperone. *Cell* 154(1), pp. 134–145.

Pearson, C.E. et al. 2002. Slipped-strand DNAs formed by long (CAG)·(CTG) repeats: Slipped-out repeats and slip-out junctions. *Nucleic Acids Research* 30(20), pp. 4534–4547.

Pearson, C.E. 2003. Slipping while sleeping? Trinucleotide repeat expansions in germ cells. *Trends in Molecular Medicine* 9(11), pp. 490–495.

Pearson, C.E. et al. 2005. Repeat instability: Mechanisms of dynamic mutations. *Nature Reviews Genetics* 6(10), pp. 729–742.

Pearson, C.E. and Sinden, R.R. 1996. Alternative structures in duplex DNA formed within the trinucleotide repeats of the myotonic dystrophy and fragile X loci. *Biochemistry* 35(15), pp. 5041–5053.

Pearson, C.E. and Sinden, R.R. 1998. Trinucleotide repeat DNA structures: Dynamic mutations from dynamic DNA. *Current Opinion in Structural Biology* 8(3), pp. 321–330.

Pennell, S. et al. 2014. FAN1 activity on asymmetric repair intermediates is mediated by an atypical monomeric virus-type replication-repair nuclease domain. *Cell Reports* 8(1), pp. 84–93.

Pfister, E.L. et al. 2009. Five siRNAs Targeting Three SNPs May Provide Therapy for Three-Quarters of Huntington's Disease Patients. *Current Biology* 19(9), pp. 774–778.

Phadte, A.S. et al. 2023. FAN1 removes triplet repeat extrusions via a PCNA- and RFC-dependent mechanism. *Proceedings of the National Academy of Sciences of the United States of America* 120(33), p. e2302103120.

Phan, H.T.L. et al. 2023. Progress in and Prospects of Genome Editing Tools for Human Disease Model Development and Therapeutic Applications. *Genes 2023, Vol. 14, Page 483* 14(2), p. 483.

Pinto, R.M. et al. 2013. Mismatch Repair Genes Mlh1 and Mlh3 Modify CAG Instability in Huntington's Disease Mice: Genome-Wide and Candidate Approaches. *PLoS Genetics* 9(10).

Plotkin, J.L. and Surmeier, D.J. 2015. Corticostriatal synaptic adaptations in Huntington's disease. *Current Opinion in Neurobiology* 33, pp. 53–62.

Porro, A. et al. 2017. FAN1 interaction with ubiquitylated PCNA alleviates replication stress and preserves genomic integrity independently of BRCA2. *Nature Communications 2017 8:1* 8(1), pp. 1–14.

Porro, A. et al. 2021. FAN1-MLH1 interaction affects repair of DNA interstrand cross-links and slipped-CAG/CTG repeats. *Science Advances* 7(31).

Pringsheim, T. et al. 2012. The incidence and prevalence of Huntington's disease: A systematic review and meta-analysis. *Movement Disorders* 27(9), pp. 1083–1091.

Prolla, T.A. et al. 1994. Dual requirement in yeast DNA mismatch repair for MLH1 and PMS1, two homologs of the bacterial mutL gene. *Molecular and Cellular Biology* 14(1), p. 407.

Pulst, S.M. et al. 1996. Moderate expansion of a normally biallelic trinucleotide repeat in

spinocerebellar ataxia type 2. *Nature Genetics 1996 14:3* 14(3), pp. 269–276.

Ran, F.A. et al. 2013. Genome engineering using the CRISPR-Cas9 system. *Nature Protocols* 8(11), pp. 2281–2308.

Ran, F.A. et al. 2015. In vivo genome editing using Staphylococcus aureus Cas9. *Nature* 520(7546), pp. 186–191.

Ranen, N.G. et al. 1995. Anticipation and instability of IT-15 (CAG)(N) repeats in parent-offspring pairs with Huntington disease. *American Journal of Human Genetics* 57(3), pp. 593–602.

Räschle, M. et al. 1999. Identification of hMutLβ, a heterodimer of hMLH1 and hPMS1. *Journal of Biological Chemistry* 274(45), pp. 32368–32375.

Rasmussen, A. et al. 2022. High Resolution Analysis of DMPK Hypermethylation and Repeat Interruptions in Myotonic Dystrophy Type 1. *Genes* 13(6), p. 970.

Ratovitski, T. et al. 2009. Mutant huntingtin N-terminal fragments of specific size mediate aggregation and toxicity in neuronal cells. *Journal of Biological Chemistry* 284(16), pp. 10855–10867.

Rawlins, M.D. et al. 2016. The prevalence of huntington's disease. *Neuroepidemiology* 46(2), pp. 144–153.

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 (BBA) - Molecular Basis of Disease* 1822(2), pp. 101–110.

Redhwan, M.A.M. et al. 2023. Small interference (RNAi) technique: Exploring its clinical applications, benefits and limitations. *European Journal of Clinical Investigation* 53(10), p. e14039.

Reenan, R.A.G. and Kolodner, R.D. 1992. Isolation and Characterization of Two Saccharomyces Cerevisiae Genes Encoding Homologs of the Bacterial Hexa and Muts Mismatch Repair Proteins. *Genetics* 132(4), p. 963.

Reiner, A. et al. 1988. Differential loss of striatal projection neurons in Huntington disease. *Proceedings of the National Academy of Sciences* 85(15), pp. 5733–5737.

Rhoads, A. and Au, K.F. 2015. PacBio Sequencing and Its Applications. *Genomics, Proteomics and Bioinformatics* 13(5), pp. 278–289.

Richard, G.F. et al. 2014. Highly specific contractions of a single CAG/CTG trinucleotide repeat by TALEN in yeast. *PLoS ONE* 9(4), p. e95611.

Rigamonti, D. et al. 2000. Wild-type huntingtin protects from apoptosis upstream of caspase-3. *Journal of Neuroscience* 20(10), pp. 3705–3713.

Rizzo, A.A. et al. 2014. NMR Structure of the Human Rad18 Zinc Finger in Complex with Ubiquitin Defines a Class of UBZ Domains in Proteins Linked to the DNA Damage Response. *Biochemistry* 53(37), pp. 5895–5906.

Robinton, D.A. and Daley, G.Q. 2012. The promise of induced pluripotent stem cells in research and therapy. *Nature* 481(7381), pp. 295–305.

De Rooij, K.E. et al. 1995. Somatic expansion of the (CAG)n repeat in Huntington disease

brains. *Human Genetics* 95(3), pp. 270–274.

Roos, R.A. 2010. Huntington's disease: A clinical review. *Orphanet Journal of Rare Diseases* 5(1), p. 40.

Rosa, F. et al. 2020. In Vitro Differentiated Human Stem Cell-Derived Neurons Reproduce Synaptic Synchronicity Arising during Neurodevelopment. *Stem Cell Reports* 15(1), pp. 22–37.

Ross, C.A. et al. 2014. Huntington disease: Natural history, biomarkers and prospects for therapeutics. *Nature Reviews Neurology* 10(4), pp. 204–216.

Roy, J.C.L. et al. 2021. Somatic CAG expansion in Huntington's disease is dependent on the MLH3 endonuclease domain, which can be excluded via splice redirection. *Nucleic Acids Research* 49(7), pp. 3907–3918.

Rüb, U. et al. 2013. Degeneration of the cerebellum in huntingtons disease (HD): Possible relevance for the clinical picture and potential gateway to pathological mechanisms of the disease process. *Brain Pathology* 23(2), pp. 165–177.

Rüb, U. et al. 2016. Huntington's disease (HD): the neuropathology of a multisystem neurodegenerative disorder of the human brain. *Brain Pathology* 26(6), pp. 726–740.

Santillan, B.A. et al. 2014. GFP-based fluorescence assay for CAG repeat instability in cultured human cells. *PLoS ONE* 9(11).

Sapranauskas, R. et al. 2011. The Streptococcus thermophilus CRISPR/Cas system provides immunity in Escherichia coli. *Nucleic Acids Research* 39(21), pp. 9275–9282.

Saudou, F. et al. 1998. Huntingtin acts in the nucleus to induce apoptosis but death does not correlate with the formation of intranuclear inclusions. *Cell* 95(1), pp. 55–66.

Saudou, F. and Humbert, S. 2016. The Biology of Huntingtin. *Neuron* 89(5), pp. 910–926.

Schaffar, G. et al. 2004. Cellular toxicity of polyglutamine expansion proteins: Mechanism of transcription factor deactivation. *Molecular Cell* 15(1), pp. 95–105.

Schmidt, M.H.M. and Pearson, C.E. 2016. Disease-associated repeat instability and mismatch repair. *DNA Repair* 38, pp. 117–126.

Semaka, A. et al. 2006. Predictive testing for Huntington disease: Interpretation and significance of intermediate alleles. *Clinical Genetics* 70(4), pp. 283–294.

Seneca, S. et al. 2004. Early onset Huntington disease: A neuronal degeneration syndrome. *European Journal of Pediatrics* 163(12), pp. 717–721.

Sentmanat, M.F. et al. 2018. A Survey of Validation Strategies for CRISPR-Cas9 Editing. *Scientific Reports* 8(1), pp. 1–8.

Serio, A. et al. 2013. Astrocyte pathology and the absence of non-cell autonomy in an induced pluripotent stem cell model of TDP-43 proteinopathy. *PNAS* 110(12).

Shelbourne, P.F. et al. 2007. Triplet repeat mutation length gains correlate with cell-type specific vulnerability in Huntington disease brain. *Human Molecular Genetics* 16(10), pp. 1133–1142.

Shen, B. et al. 2014. Efficient genome modification by CRISPR-Cas9 nickase with minimal off-target effects. *Nature Methods* 11(4), pp. 399–402.

Shi, Y. et al. 2012a. Directed differentiation of human pluripotent stem cells to cerebral cortex neurons and neural networks. *Nature Protocols* 7(10), pp. 1836–1846.

Shi, Y. et al. 2012b. Human cerebral cortex development from pluripotent stem cells to functional excitatory synapses. *Nature Neuroscience* 15(3), pp. 477–486.

Shin, J.W. et al. 2016. Permanent inactivation of Huntington's disease mutation by personalized allele-specific CRISPR/Cas9. *Human Molecular Genetics* 25(20), pp. 4566–4576.

Shirendeb, U.P. et al. 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* 21(2), pp. 406–420.

Skotte, N.H. et al. 2014. Allele-specific suppression of mutant huntingtin using antisense oligonucleotides: Providing a therapeutic option for all Huntington disease patients. *PLoS ONE* 9(9).

Smith, Y. et al. 1998. Microcircuitry of the direct and indirect pathways of the basal ganglia. *Neuroscience* 86(2), pp. 353–387.

Smogorzewska, A. et al. 2010. A Genetic Screen Identifies FAN1, a Fanconi Anemia-Associated Nuclease Necessary for DNA Interstrand Crosslink Repair. *Molecular Cell* 39(1), pp. 36–47.

Snell, R.G. et al. 1993. Relationship between trinucleotide repeat expansion and phenotypic variation in Huntington's disease. *Nature Genetics* 4(4), pp. 393–397.

Soares, T.R. et al. 2019. Targeting the proteostasis network in Huntington's disease. *Ageing Research Reviews* 49(July 2018), pp. 92–103.

Song, W. et al. 2011. Mutant huntingtin binds the mitochondrial fission GTPase dynamin-related protein-1 and increases its enzymatic activity. *Nature Medicine 2011 17:3* 17(3), pp. 377–382.

Southwell, A.L. et al. 2018. Huntingtin suppression restores cognitive function in a mouse model of Huntington's disease. *Science Translational Medicine* 10(461), pp. 1–13.

Spada, A.R.L. et al. 1991. Androgen receptor gene mutations in X-linked spinal and bulbar muscular atrophy. *Nature* 352(6330), pp. 77–79.

Spronck, E.A. et al. 2021. Intrastriatal administration of AAV5-MIHTT in non-human primates and rats is well tolerated and results in MIHTT transgene expression in key areas of huntington disease pathology. *Brain Sciences* 11(2), pp. 1–18.

Steffan, J.S. et al. 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* 97(12), pp. 6763–6768.

Stepanenko, A.A. and Dmitrenko, V. V. 2015. HEK293 in cell biology and cancer research: Phenotype, karyotype, tumorigenicity, and stress-induced genome-phenotype evolution. *Gene* 569(2), pp. 182–190.

Stevanovski, I. et al. 2022. Comprehensive genetic diagnosis of tandem repeat expansion disorders with programmable targeted nanopore sequencing. *Science Advances* 8(9), p. 17.

Stöberl, N. et al. 2023. Mutant huntingtin confers cell-autonomous phenotypes on Huntington's disease iPSC-derived microglia. *Scientific Reports* 13(1), pp. 1–15.

Sun, A.X. et al. 2016. Direct Induction and Functional Maturation of Forebrain GABAergic Neurons from Human Pluripotent Stem Cells. *Cell Reports* 16(7), pp. 1942–1953.

Swami, M. et al. 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), pp. 3039–3047.

Tabrizi, S.J. et al. 2019. Targeting huntingtin expression in patients with Huntington's disease. *New England Journal of Medicine* 380(24), pp. 2307–2316.

Tabrizi, S.J. et al. 2022. Potential disease-modifying therapies for Huntington's disease: lessons learned and future opportunities. *The Lancet Neurology* 21(7), pp. 645–658.

Takahashi, K. et al. 2007. Induction of Pluripotent Stem Cells from Adult Human Fibroblasts by Defined Factors. *Cell* 131(5), pp. 861–872.

Takahashi, K. and Yamanaka, S. 2006. Induction of Pluripotent Stem Cells from Mouse Embryonic and Adult Fibroblast Cultures by Defined Factors. *Cell* 126(4), pp. 663–676.

Tan, D. et al. 2023. CAG Repeat Expansion in THAP11 Is Associated with a Novel Spinocerebellar Ataxia. *Movement Disorders* 38(7), pp. 1282–1293.

Tanaka, J. et al. 1999. Astrocytes prevent neuronal death induced by reactive oxygen and nitrogen species. *GLIA* 28(2), pp. 85–96.

Taylor, A.S. et al. 2022. Repeat Detector: versatile sizing of expanded tandem repeats and identification of interrupted alleles from targeted DNA sequencing. *NAR Genomics and Bioinformatics* 4(4).

Tebbenkamp, A.T.N. et al. 2012. Analysis of Proteolytic Processes and Enzymatic Activities in the Generation of Huntingtin N-Terminal Fragments in an HEK293 Cell Model. *PLOS ONE* 7(12), p. e50750.

Telenius, H. et al. 1994. Somatic and gonadal mosaicism of the Huntington disease gene CAG repeat in brain and sperm. *Nature Genetics* 6(4), pp. 409–414.

Telezhkin, V. et al. 2016. Forced cell cycle exit and modulation of GABAA, CREB, and GSK3β signaling promote functional maturation of induced pluripotent stem cell-derived neurons. *American Journal of Physiology - Cell Physiology* 310(7), pp. C520–C541.

Thompson, L.M. 2003. An expanded role for wild-type huntingtin in neuronal transcription. *Nature Genetics* 35(1), pp. 13–14.

Thomson, J.A. et al. 1998. Embryonic Stem Cell Lines Derived from Human Blastocysts. *Science* 282(November), pp. 1145–1148.

Thomson, S.B. et al. 2023. AAV5-miHTT-mediated huntingtin lowering improves brain health in a Huntington's disease mouse model. *Brain* 146(6), pp. 2298–2315.

Thongthip, S. et al. 2016. Fan1 deficiency results in DNA interstrand cross-link repair defects, enhanced tissue karyomegaly, and organ dysfunction. *Genes and Development* 30(6), pp. 645–659.

Tiwari, A.K. et al. 2016. Lynch syndrome in the 21st century: clinical perspectives. *QJM: An International Journal of Medicine* 109(3), pp. 151–158.

Tomé, S. et al. 2013. MSH3 Polymorphisms and Protein Levels Affect CAG Repeat Instability in Huntington's Disease Mice. *PLoS Genetics* 9(2), p. e1003280.

Trepte, P. et al. 2014. Spontaneous self-assembly of pathogenic huntingtin exon 1 protein into amyloid structures. *Essays in Biochemistry* 56(1), pp. 167–180.

Trottier, Y. et al. 1994. Instability of CAG repeats in Huntington's disease: Relation to parental transmission and age of onset. *Journal of Medical Genetics* 31(5), pp. 377–382.

Tsai, Y.C. et al. 2022. Identification of a CCG-Enriched Expanded Allele in Patients with Myotonic Dystrophy Type 1 Using Amplification-Free Long-Read Sequencing. *Journal of Molecular Diagnostics* 24(11), pp. 1143–1154.

Urbach, A. et al. 2004. Modeling for Lesch-Nyhan Disease by Gene Targeting in Human Embryonic Stem Cells. *Stem Cells* 22, pp. 635–641.

Usdin, K. et al. 2015. Repeat instability during DNA repair: Insights from model systems. *Critical Reviews in Biochemistry and Molecular Biology* 50(2), pp. 142–167.

Vale, T.C. and Cardoso, F. 2015. Chorea: A Journey through History. *Tremor and Other Hyperkinetic Movements* 5, pp. 1–6.

Verweij, J. and H.M.P. 1990. Mitomycin C: mechanism of action, usefulness and l imitations. *Rapid Communications of Oxford* , pp. 5–13.

Victor, M.B. et al. 2018. Striatal neurons directly converted from Huntington's disease patient fibroblasts recapitulate age-associated disease phenotypes. *Nature Neuroscience* 21(3), pp. 341–352.

Vonsattel, Jean-Paul; DiFiglia, M. 1998. Huntington Disease. *Journal of Neuropathology and Experimental Neurology* 57(5), pp. 369–384.

Vonsattel, J.-P. et al. 1985. *Neuropathological Classification of Huntinqton's Disease*.

Vonsattel, J.P.G. et al. 2008. Neuropathology of Huntington's Disease. *Handbook of Clinical Neurology* 89, pp. 599–618.

Wagner, A.S. et al. 2018. Self-assembly of Mutant Huntingtin Exon-1 Fragments into Large Complex Fibrillar Structures Involves Nucleated Branching. *Journal of Molecular Biology* 430(12), pp. 1725–1744.

Wagner, R. and Meselson, M. 1976. Repair tracts in mismatched DNA heteroduplexes. *Proceedings of the National Academy of Sciences* 73(11), pp. 4135–4139.

Walker, F.O. 2007. Huntington's disease. *Lancet* 369(9557), pp. 218–228.

Wang, C. et al. 2017. Scalable Production of iPSC-Derived Human Neurons to Identify Tau-Lowering Compounds by High-Content Screening. *Stem Cell Reports* 9(4), pp. 1221–1233.

Wang, R. et al. 2014. Mechanism of DNA interstrand cross-link processing by repair nuclease FAN1. *Science* 346(6213), pp. 1127–1130.

Wanker, E.E. et al. 2019. The pathobiology of perturbed mutant huntingtin protein–protein interactions in Huntington's disease. *Journal of Neurochemistry* 151(4), pp. 507–519.

Wexler, N.S. et al. 1987. Homozygotes for huntington's disease. *Nature* 326(6109), pp. 194–197.

Wexler, N.S. et al. 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), pp. 3498–503.

Wheeler, V.C. et al. 1999. *Length-dependent gametic CAG repeat instability in the Huntington's disease knock-in mouse*.

Wheeler, V.C. et al. 2003. Mismatch repair gene Msh2 modifies the timing of early disease in HdhQ111 striatum. *Human Molecular Genetics* 12(3), pp. 273–281.

Wheeler, V.C. and Dion, V. 2021. Modifiers of CAG/CTG Repeat Instability: Insights from Mammalian Models. *Journal of Huntington's Disease* 10(1), pp. 123–148.

White, J.K. et al. 1997. Huntingtin is required for neurogenesis and is not impaired by the Huntington's disease CAG expansion. *Nature Genetics* 17(december), pp. 404–410.

Wieben, E.D. et al. 2012. A Common Trinucleotide Repeat Expansion within the Transcription Factor 4 (TCF4, E2-2) Gene Predicts Fuchs Corneal Dystrophy. *PLOS ONE* 7(11), p. e49083.

Wiedenheft, B. et al. 2012. RNA-guided genetic silencing systems in bacteria and archaea. *Nature* 482(7385), pp. 331–338.

Wild, E.J. and Tabrizi, S.J. 2017. Therapies targeting DNA and RNA in Huntington's disease. *The Lancet Neurology* 16(10), pp. 837–847.

Wildenberg, J. and Meselson, M. 1975. Mismatch repair in heteroduplex DNA. *Proceedings of the National Academy of Sciences of the United States of America* 72(6), pp. 2202–2206.

Wong, Y.C. and Holzbaur, E.L.F. 2014. The Regulation of Autophagosome Dynamics by Huntingtin and HAP1 Is Disrupted by Expression of Mutant Huntingtin, Leading to Defective Cargo Degradation. *The Journal of Neuroscience* 34(4), p. 1293.

Wu, Z. et al. 2010. Effect of genome size on AAV vector packaging. *Molecular Therapy* 18(1), pp. 80–86.

Xu, P. et al. 2020. Dynamics of strand slippage in DNA hairpins formed by CAG repeats: Roles of sequence parity and trinucleotide interrupts. *Nucleic Acids Research* 48(5), pp. 2232–2245.

Yagi, T. et al. 2011. Modeling familial Alzheimer's disease with induced pluripotent stem cells. *Human Molecular Genetics* 20(23), pp. 4530–4539.

Yan, S. et al. 2023. Cas9-mediated replacement of expanded CAG repeats in a pig model of Huntington's disease. *Nature Biomedical Engineering* 7(5), pp. 629–646.

Yang, S. et al. 2017. CRISPR/Cas9-mediated gene editing ameliorates neurotoxicity in mouse model of Huntington's disease. *The Journal of Clinical Investigation* 127(7), p. 2719.

Yasuhara, H. et al. 2022. Reduction of Off-Target Effects of Gapmer Antisense Oligonucleotides by Oligonucleotide Extension. *Molecular Diagnosis and Therapy* 26(1), pp. 117–127.

Yoshikiyo, K. et al. 2010. KIAA1018/FAN1 nuclease protects cells against genomic instability induced by interstrand cross-linking agents. *Proceedings of the National Academy of Sciences of the United States of America* 107(50), pp. 21553–21557.

Yu, J. et al. 2007. Induced pluripotent stem cell lines derived from human somatic cells. *Science* 318(5858), pp. 1917–1920.

Zeitler, B. et al. 2019. Allele-selective transcriptional repression of mutant HTT for the treatment of Huntington's disease. *Nature Medicine* 25(7), pp. 1131–1142.

Zeitlin, S. et al. 1995. Increased apoptosis and early embryonic lethality in mice nullizygous for the Huntington's disease gene homologue. *Nature Genetics* 11(october), pp. 155–163.

Zhai, W. et al. 2005. In vitro analysis of huntingtin-mediated transcriptional repression reveals multiple transcription factor targets. *Cell* 123(7), pp. 1241–1253.

Zhang, L. et al. 2014a. Inhibition of stearoyl-coA desaturase selectively eliminates tumorigenic Nanog-positive cells: Improving the safety of iPS cell transplantation to myocardium. *Cell Cycle* 13(5), pp. 762–771.

Zhang, Y. et al. 2013. Rapid single-step induction of functional neurons from human pluripotent stem cells. *Neuron* 78(5), pp. 785–798.

Zhang, Y. et al. 2014. Comparison of non-canonical PAMs for CRISPR/Cas9-mediated DNA cleavage in human cells. *Scientific Reports* 4, pp. 1–5.

Zhao, Q. et al. 2014. Structural insights into 5' flap DNA unwinding and incision by the human FAN1 dimer. *Nature Communications* 5(1–9).

Zhao, X. et al. 2017. Cell cycle-dependent control of homologous recombination. *Acta Biochimica et Biophysica Sinica* 49(8), pp. 655–668.

Zhao, X. et al. 2021. FAN1's protection against CGG repeat expansion requires its nuclease activity and is FANCD2-independent. *Nucleic Acids Research* 49(20), pp. 11643–11652.

Zhao, X.N. and Usdin, K. 2018. FAN1 protects against repeat expansions in a Fragile X mouse model. *DNA Repair* 69, pp. 1–5.

Zhou, W. et al. 2012. FAN1 mutations cause karyomegalic interstitial nephritis, linking chronic kidney failure to defective DNA damage repair. *Nature Genetics* 44(8), pp. 910–915.

Zhuchenko, O. et al. 1997. Autosomal dominant cerebellar ataxia (SCA6) associated with small polyglutamine expansions in the α1A-voltage-dependent calcium channel. *Nature Genetics 1997 15:1* 15(1), pp. 62–69.

Zuccato, C. et al. 2001. Loss of huntingtin-mediated BDNF gene transcription in Huntington's disease. *Science* 293(5529), pp. 493–498.

Zuccato, C. et al. 2003. Huntingtin interacts with REST/NRSF to modulate the transcription of NRSE-controlled neuronal genes. *Nature Genetics* 35(1), pp. 76–83.