THE UNIVERSITY of ADELAIDE

THE ROLE OF *UPF3A* AND *UPF3B* IN EARLY DEVELOPMENT AND NEURAL DIFFERENTIATION

Debrah Sadie Sebolai

IBBL (Hons)

Neurogenetics Laboratory

The University of Adelaide

Thesis submitted for the degree of

Doctor of Philosophy in

Discipline of Paediatrics

Adelaide School of Medicine

The University of Adelaide

November 2019

Table of Contents

ABSTRACT

The nonsense mediated mRNA decay (NMD) pathway degrades transcripts with premature termination codons (PTCs) to prevent the production of C-terminally truncated proteins that might have dominant negative properties. Loss of function mutations and copy number variation in genes required for NMD have also been implicated in neurodevelopmental disorders (NDDs) such as autism spectrum disorders (ASD), intellectual disability (ID), childhood onset schizophrenia and attention deficit hyperactivity disorder (ADHD). One such gene, *UPF3B*, located on the X-chromosome, is the only NMD factor that has a gene paralog *UPF3A*, located on chromosome 13. Both are implicated in NDDs. Loss of function mutations in *UPF3B* lead to UPF3A protein stabilisation, a phenomena proposed to compensate for loss of *UPF3B* in order to maintain residual NMD activity and during critical periods of development. The impact of loss of *UPF3B* function in neurodevelopment has been investigated in *UPF3B* patient lymphoblastoid cell lines and mice and leads to a deregulation of mRNA important for brain function and development. The role of *UPF3A* has not yet been elucidated. *UPF3A* was initially shown to act as a weak NMD activator until recently when it was shown to primarily act to inhibit NMD of a majority, but not all, NMD-targeted mRNAs. In mouse loss of *UPF3A* has been shown to be embryonic lethal. Thus far, the role of NMD and *UPF3A* and *UPF3B* in particular, have yet to be determined in human cells of the developing brain. To elucidate the roles of *UPF3A* and *UPF3B* in human cells in a neurodevelopmental model, *UPF3A* and *UPF3B* knockout (KO) human Embryonic Stem Cells (hESCs) were generated using the CRISPR/Cas9 genome editing technology. Whole genome sequencing (WGS) of the gene-edited clones revealed that none of the small nucleotide variants (SNVs) detected in each gene-edited clone overlapped with the predicted off-target sites, however, notable genetic variation in each clone was attributed to extended cell culture as part of the CRISPR/Cas9 editing process, presence of

mosaicism in the parental cell line and normal passaging. Larger deletions and duplications (structural variants (SVs)) were detected in each gene-edited clones that overlapped with the predicted off-target sites. The hESCs were subsequently differentiated into neural stem cells (NSCs) and transcriptome wide mRNA sequencing analysis was performed both on the *UPF3A* and *UPF3B* KO hESCs and NSCs. Loss of *UPF3A* in hESCs did not have any significant effect on the transcriptome of hESCs but lead to deregulation of 3.03% of transcripts in NSCs, the majority of which (2.96%) were downregulated. Genes deregulated in *UPF3A* KO NSCs were involved in intracellular signalling pathways regulated by the TGF-β superfamily pathway such as extracellular matrix remodelling, blood vessel development and morphogenesis and regulation of cell adhesion. Mutations in genes that encode components of the extracellular matrix and the vascular system lead to embryonic lethality in mice. Loss of *UPF3B* had an impact at both the hESC and NSC stage. Loss of *UPF3B* in hESCs lead to a deregulation of 0.62% (of which 0.21% were upregulated and 0.41% were downregulated) while loss in NSCs resulted in a 0.29% deregulation (0.07% upregulated and 0.22% downregulated). Loss of *UPF3B* in hESCs lead to a deregulation in genes that are important in cell-cell adhesion, calcium ion binding, synapse assembly and regulation of signalling receptor activity and post synaptic membrane potential. In NSCs, loss of *UPF3B* lead to deregulation of genes important for neurodevelopment such as of *ARX*, an ID gene and neural function such as *ROBO2* which is important for axonal guidance. Our results suggest that *UPF3A* could act as an NMD inhibitor and is important in regulating the TGF-β superfamily pathway, while *UPF3B* is important for normal neurodevelopment and function. The deregulated genes in *UPF3A* and *UPF3B* KO NSCs only had two genes that overlapped suggesting that these paralogs have separate roles in NMD. In conclusion the CRISPR/Cas9 genome editing tool and hESCs were efficient and faithful tools to model NDDs caused by mutations in *UPF3A* and *UPF3B*. They provided resources to investigate the loss of

UPF3A and *UPF3B* in a suitable cell type that would otherwise be impossible to acquire or only able to acquire post mortem.

DECLARATION

I certify that this work contains no material which has been accepted for the award of any other degree or diploma in my name, in any university or other tertiary institution and, to the best of my knowledge and belief, contains no material previously published or written by another person, except where due reference has been made in the text. In addition, I certify that no part of this work will, in the future, be used in a submission in my name, for any degree or diploma in any university or other tertiary institution without the prior approval of the University of Adelaide and where applicable, any partner institution responsible for the joint-award of this degree.

I give permission for the digital version of my thesis to be made available on the web, via the University's digital research repository, the Library Search and also through web search engines, unless permission has been granted by the University to restrict access for a period of time.

I acknowledge the support I have received for my research through the provision of an Australian Government Research Training Program Scholarship.

Bachelor of Laboratory Medicine (Honours)

Student number: a1643807

Signed

DATE: 01 November 2019

ACKNOWLEDGEMENTS

There are a number of people to whom I am greatly indebted and who without their contributions and influences this dissertation may not have been written. I would like to thank my supervisors, Professor Jozef Gecz and Dr Lachlan Jolly for their intellectual input and guidance and more especially Jozef for giving me such a great opportunity to work in his laboratory. I would like to also thank Dr Claire Homan who taught me how to culture hESCs and neural stem cells and thank Dr Lachlan Moldenhauer for performing flow cytometry for my cell cycle analysis. I would also like to acknowledge Rhonda Hutchinson and Mario Nicola from SA Pathology for performing karyotype analysis on the generated CRISPR/Cas9 clones. I would like to thank Dr Mark Corbett and Dr Atma Ivancevic for analysing my whole genome sequence data and Dr Stephen Pederson and Urwah Nawaz for analysing my RNA Seq data.

To my loving late parents Francis and Patricia Renders who have nurtured me to be the dedicated and resilient person I am and to my parents in-law Moses and Dr Boingotlo Sebolai who have been my pillar of strength during this period and my husband for being supportive and encouraging.

ABBREVIATIONS

Standard Terms

%	Percentage
°C	Degrees Celsius
aa	amino acid
bp	Base pair
Cat	Catalogue
cDNA	Complementary DNA
DNA	Deoxyribonucleic Acid
g	Gravitational force
gDNA	Genomic DNA
hrs	Hours
Kb	Kilobase
KO	Knockout
L	Litre
M	Molar
Min	Minutes
mL	Millilitre
mM	Millimolar
mRNA	Messenger RNA
ng	Nanograms
nt	Nucleotide
RNA	Ribonucleic Acid
WT	Wildtype
µg	Microgram
µL	Microlitre
µM	Micromolar

Materials and Methods

BSA	Bovine serum albumin
DAPI	4’6-diamidino-2-phenylindole
DMEM	Dulbecco’s modified eagle medium
DMEM/F12	Dulbecco’s modified eagle medium: nutrient F-12
DMSO	Dimethyl sulfoxide
dNTP	Deoxyribonucleotide
DPBS	Dulbecco’s phosphate buffered saline
ECM	Extracellular matrix
EDTA	Ethylenediaminetetraacetic acid
EGF	Epidermal growth factor
ESC	Embryonic stem cell
FACS	Fluorescence-activated cell sorting
FBS	Foetal bovine serum
FCS	Foetal calf serum
FGF	Fibroblast growth factor
GFP	Green fluorescent protein
hESC	Human embryonic stem cell
IF	Immunofluorescence
iPSC	Induced pluripotent stem cell
LB	Luria-broth
iMEF	Irradiated mouse embryonic fibroblasts
NEAA	Non-essential amino acids
NSC	Neural stem cell
PBS	Phosphate buffered saline
PBST	Phosphate buffered saline Tween20
PCR	Polymerase chain reaction
Pen/Strep	Penicillin-Streptomycin
PFA	Paraformaldehyde
RT qPCR	Quantitative real time PCR
SMC4	Small molecule inhibitor complex 4

Non-Standard Terms

ADHD	Attention deficient hyperactivity disorder
ASD	Autism spectrum disorder
BND	Breakend
CASC3	Cancer susceptibility candidate 3
CBP	Cap binding protein
CNV	Copy number variant
CRISPR	Clustered regularly interspaced short palindromic repeats
DBS	Double stranded breaks
DEG	Differently expressed gene
DEL	Deletions
DHX34	DEAH (Asp-Glu-Ala-His) box polypeptide 34
DUP	Duplications
EIF4A3	Eukaryotic translation initiation factor 4A3
EJC	Exon junction complex
ERF	Eukaryotic release factor
gRNA	Guide RNA
HD	Homology directed
INDELS	Insertions/deletions
INS	Insertions
INV	Inversions
ID	Intellectual disability
KO	Knockout
LCL	Lymphoblastoid cell line
MAGOH	Mago-nashi homolog, proliferation-associated
NBAS	Neuroblastoma amplified sequence
NDDS	Neurodevelopmental disorders
NHEJ	Non-homologous end joining
NMD	Nonsense mediated mRNA decay pathway
NT	Non transfected
ORF	Open reading frame

PAM	Protospacer adjacent motif
PTC	Premature termination codon
RBM8A	RNA binding motif protein 8A
RNA Seq	RNA sequencing
RNPS1	RNA binding protein S1, serine-rich domain 1
SMG	Suppressor of morphological defects on genitalia
SNP	Small nucleotide polymorphism
SNV	Small nucleotide variation
STD	Standard deviation
SURF	SMG1, UPF1 and ERF1-ERF3 complex
SV	Structural variants
uORF	Upstream open reading frame
UTR	Untranslated region
UPF	Up frameshift factor
WIGB	Within bgen homolog (Drosophila)
WGS	Whole genome sequencing
D. melanogaster	*Drosophila melanogaster*
E.coli	*Escherichia coli*
S.cerevisiae	*Saccharomyces cerevisiae*

LIST OF TABLES

LIST OF FIGURES

Chapter 1 - Introduction

1.0 Introduction

Neurodevelopmental disorders (NDDs) are a large group of heterogeneous disorders caused by genetic, environmental or epigenetic factors that lead to aberrations in nervous system development. The clinical presentations seen in NDDs are postulated to be caused by alterations in neurodevelopmental processes such as neurogenesis, cell migration and neural connectivity (Ehninger et al. 2008). NDDs include a broad range of disorders such as attention deficit hyperactivity disorder (ADHD), epilepsy, autism spectrum disorders (ASDs), cerebral palsy and intellectual disability (ID) (Szpir 2006). These disorders result in abnormalities that affect an individual's behaviour, memory and learning (Szpir 2006) and the common feature relating them is that the disease onset occurs during periods of development (Ehninger et al. 2008).

Genetic factors are the main drivers of these disorders (Ehninger et al. 2008) with approximately 12% of genetically inherited disorders caused by premature termination codons (PTCs) (Mort, M. et al. 2008). PTCs in mRNAs can arise due to germline or somatic mutations in DNA and inaccurate pre-mRNA splicing (Keeling, Du & Bedwell 2013). PTCs normally cause a significant reduction in mRNA levels due to mRNA degradation via the nonsense-mediated mRNA decay (NMD) pathway, and lead to complete loss of protein function (Keeling, Du & Bedwell 2013).

1.1 Nonsense mediated mRNA decay pathway (NMD)

The NMD pathway exists and is conserved within all eukaryotes (Schweingruber et al. 2013). It was discovered in 1979 in β° thalassemia patients that had nonsense mutations in their *β-globin* gene and had reduced *β-globin* mRNA (Chang & Kan 1979). Proteins including the upframe-shift factors (UPFs), **s**uppressor with **m**orphogenetic effect on **g**enitalia (SMG) proteins and components of the exon junction complex (EJC) form a complex that is crucial in initiating the identification and decay of mRNA with PTCs (Buchwald et al. 2010). Whilst the degradation of transcripts with PTCs protects against the production of truncated proteins with deleterious dominant negative or gain of function effects, in some instances the truncated protein might retain some residual activity which would lessen the clinical severity incurred by the resulting haploinsufficiency. Thus NMD is recognised as a major modifier of clinical outcomes of genetic diseases caused by PTC type mutations (Khajavi, Inoue & Lupski 2006).

1.2 Classical NMD Mechanism(s)

The relay of genetic information from DNA to protein requires accurate transmission. There are several checkpoints or pathways that are responsible for eliminating any transcripts that may lack functionality or fidelity (Smith & Baker 2015). Before genetic information in DNA is relayed into protein, DNA is first transcribed into RNA. The RNA is processed and non-coding regions (introns) removed before it is translated into polypeptides. The NMD pathway is responsible at this stage to ensure that transcripts with PTCs in their open reading frames (ORFs) are degraded to prevent such transcripts to be translated into proteins.

When introns are removed from RNA (splicing), EJCs are deposited between 20-24 nucleotides (nts) upstream of the exon-exon boundary (Le Hir et al. 2000). The EJC has an inner core complex that contains the EJC factors, MAGOH/Y14 (RBM8A), BTZ and eIF4AIII. The inner core complex is tightly bound to the mRNA by the eIF4AIII protein (Ballut et al. 2005). The EJC also associates with peripheral proteins in its outer shell which contains splicing factors such as ACINUS, PININ, RNPS1, SAP18 and the mRNA export factor REF/Aly (Tange et al. 2005; Wang et al. 2018).

In the classical NMD mechanism, the EJC aids discrimination of physiological stop codons from PTCs. As the vast majority of physiological stop codons are encoded in the last exon, mRNA with PTCs have a unique mRNA landscape in which an EJC is located further downstream of a stop codon. A stop codon needs to be located >50-55 nts upstream of an EJC to be recognised by the NMD machinery (**Figure 1.1**) (Nagy & Maquat 1998). Thus the EJC aids the NMD mechanism to identify and subsequently degrade transcripts that contain a stop codon followed by an EJC (Nagy & Maquat 1998).

NMD occurs during mRNA translation. While it has been reported that NMD only occurs during the pioneering round of translation (Ishigaki et al. 2001) some studies reveal that mRNAs have the ability to be degraded during each round of translation (Durand & Lykke-Andersen 2013; Hoek et al. 2019; Rufener & Muhlemann 2013). After the assembly of the EJC on the mRNA, the NMD factor UPF3B binds to the EJC using its C-terminal domain (Kadlec, Izaurralde & Cusack 2004) to interact with the EJC component Y14 (Le Hir et al. 2001). When the mRNA is exported from the nucleus into the cytoplasm for translation, a second NMD factor UPF2 binds to the EJC through interactions with the N-terminal of UPF3B (Gehring et al. 2003) (**Figure 1.2**).

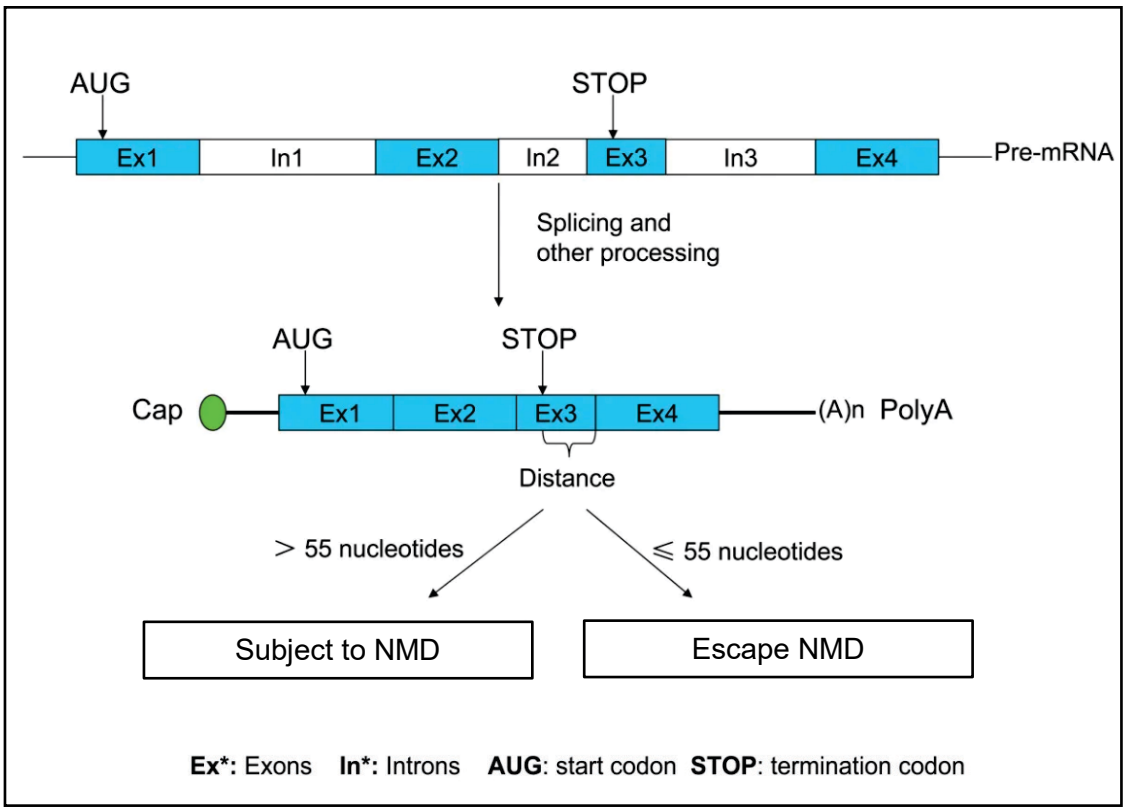


Figure 1.1: The 55 nucleotide NMD rule.

After pre-mRNA is transcribed it is processed to produce mRNA. These processing steps include splicing, capping and polyadenylation. During splicing after introns are removed exons are ligated together creating exon-exon boundaries. If a transcript contains a PTC located more than 55 nts upstream from an exon-exon boundary it is targeted and degraded by the NMD machinery, while PTCs located 55 nts or less normally escape NMD with exceptions to this rule. PTCs less than 55 nts from the exon-exon boundary do not elicit NMD as the ribosomes would have already displaced the EJCs by the time the scanning ribosomes recognise the PTC as EJCs are located 24 nts upstream of the exon-exon boundary and the ribosome footprint is approximately 20 nts on either side (Kataoka et al. 2000; Le Hir et al. 2000). Modified figure taken from (Zhang et al. 2009).

During translation, when the ribosome encounters an authentic stop codon, the ribosome-bound eukaryotic release factor 1 (eRF1) and eRF3 bind poly(A) binding protein cytoplasmic 1 (PABPC1) and terminate translation (Zahdeh & Carmel 2016). In the instance when ribosomes encounters a PTC, termination is inefficient and the ribosome stalls due to the absence of the termination stimulating factors between PABPC1 and the eRFs (Amrani, Sachs & Jacobson 2006). Instead of binding to PABPC1, eRF3 recruits UPF1 and SMG1 to the EJC complex and triggers the formation of a downstream complex called SURF (SMG1, UPF1, eRF1 and eRF3), together with SMG8 and SMG9 which collectively activates the first step in the NMD pathway (Kashima et al. 2006; Longman et al. 2013; Yamashita 2013).

The SURF complex allows formation of the decay inducing complex (DECID), which results from the association of SURF, ribosomes, UPF2 and the EJC. In this DECID complex UPF2 binds with the C-terminal of SMG1 inducing the dissociation of SMG8-SMG9, activating the kinase activity of SMG1. SMG1 then phosphorylates the C-terminal serine-glutamine motifs of UPF1, which is a key event in activating mRNA decay machinery (Arias-Palomo et al. 2011). The phosphorylation of UPF1 leads to the recruitment of SMG6 and SMG5-7 (Okada-Katsuhata et al. 2012). The PilT N-terminus domain of SMG6, initiates mRNA decay through its endonuclease activity (Franks, Singh & Lykke-Andersen 2010) which creates unprotected free 5’ and 3’ mRNA ends (Popp & Maquat 2014). SMG5-SMG7 complexes recruit decapping and deadenylases enzymes that remove the terminal 5’ and 3’ mRNA modifications, allowing access to 5’-3’ end and 3’-5’ RNA degradation enzymes respectively (Fukuhara et al. 2005).

UPF2 also binds to the regulatory cysteine-histidine rich domain of UPF1, causing a conformational change that promotes UPF1 helicase activity (Chakrabarti et al. 2011). This conformational change switches the helicase domain of UPF1 from a RNA-clamping function to a RNA-unwinding function (Chamieh et al. 2008). This is needed for the presentation of the 3'-cleavage product generated by SMG6 to the 5'- to -3' exonuclease XRN1 (Franks, Singh & Lykke-Andersen 2010) and to recycle essential protein factors (Popp & Maquat 2014).

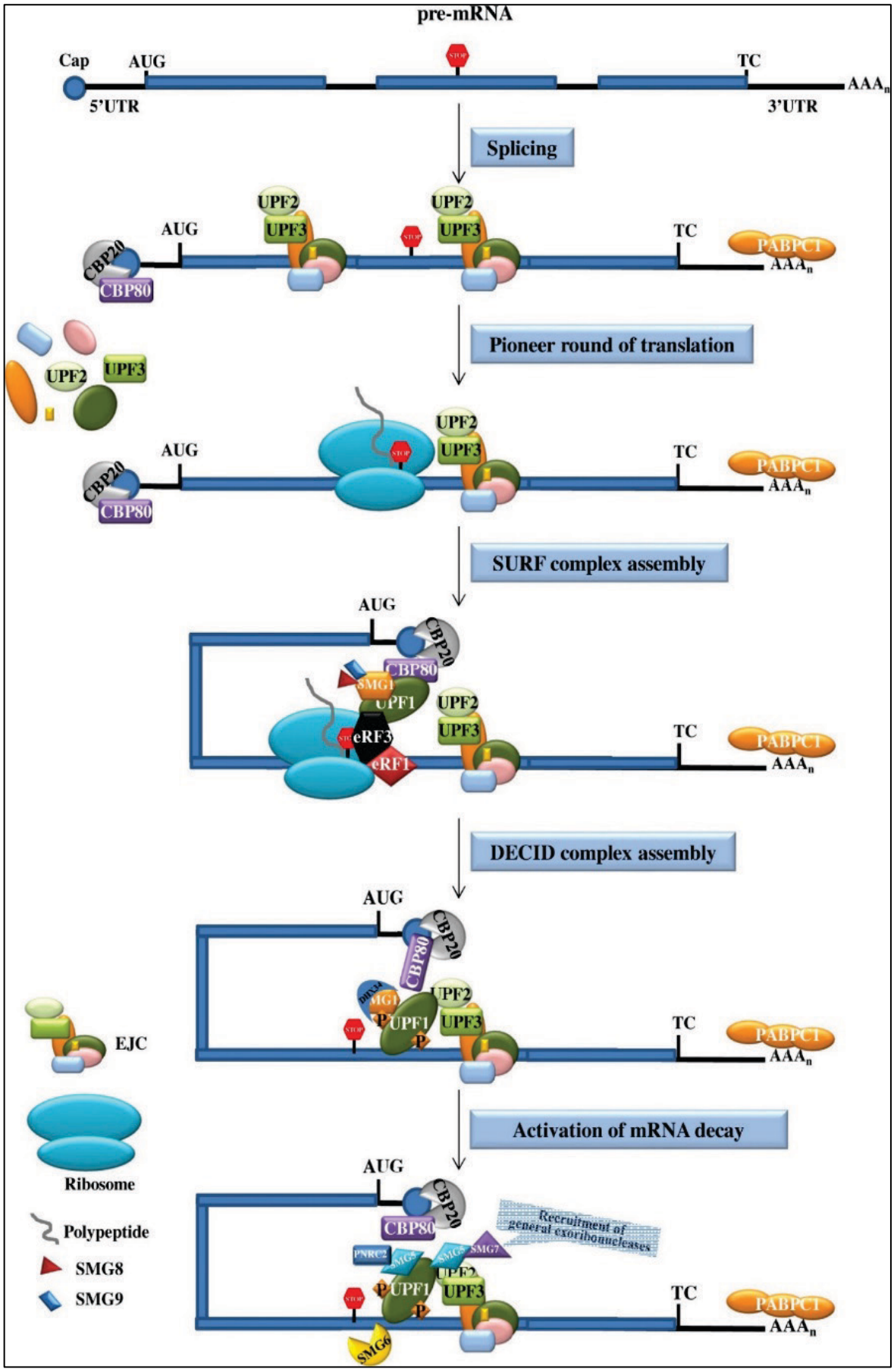

pre-mRNA
Cap
AUG
TC
AAAn
5'UTR
3'UTR
Splicing
UPF2
UPF3
CBP20
CBP80
PABPC1
Pioneer round of translation
SURF complex assembly
SMG1
UPF1
eRF3
eRF1
DECID complex assembly
P
Activation of mRNA decay
Recruitment of general exoribonucleases
PNRC2
SMG5
SMG6
SMG7
EJC
Ribosome
Polypeptide
SMG8
SMG9

Figure 1.2: NMD mechanism

In the nucleus during mRNA splicing, introns are removed and exons are ligated. 20-24 nt upstream from the exon-exon boundary EJCs are loaded on the transcript. UPF3B, an NMD factor also binds the EJC. In the cytoplasm UPF2 attaches to the EJC through binding to UPF3B. During translation, EJCs are removed by ribosomes. However if there is a PTC, the ribosomes stall, leading to the recruitment of the NMD factors and the eRFs which form the SURF complex. Association of the SURF complex, ribosomes, UPF2, and the EJC form the decay inducing complex and leads to degradation of the mutated transcript. Figure from (da Costa, Menezes & Romao 2017).

1.2.1 Alternative NMD pathway(s)

The classical NMD pathway model suggests that the NMD pathway occurs in a linear manner with respect to the involvement of the core NMD factors UPF1, UPF2 and UPF3B: UPF3B first binds to the EJC in the nucleus and after the transcript is exported into the cytoplasm, UPF2 binds UPF3B and finally UPF1 to form a bridge between the EJC and UPF1, resulting in NMD pathway activation (Chakrabarti et al. 2011; Gehring et al. 2003). However alternative NMD pathways exists that act independently of some of these core NMD factors (Chan et al. 2007; Gehring et al. 2005).

A prime example of this stems from observations in patients with *UPF3B* mutations, in which the PTC containing mutant mRNA of *UPF3B* is itself NMD degraded despite the absence of the UPF3B protein (Nguyen et al. 2012; Tarpey et al. 2007). There is also a UPF3A/B independent branch of NMD that does not require these gene paralogs that is shown to occur in T-cell receptors (TCR). The TCR locus undergoes programmed rearrangements that generate PTCs which are degraded by the UPF3A/B independent branch of NMD (Chan et al. 2007).

Gehring et al., 2005 identified two different branches of the NMD pathway that require different EJC factors. The UPF2 independent pathway that requires eIF4A3, Y14/Magoh and BTZ (**Figure 1.3a**). In support of the UPF2 independent pathway, there is prevailing evidence that UPF3B can bind directly to UPF1 independent of UPF2 (Gehring et al. 2005; Kunz et al. 2006; Neu-Yilik et al. 2017; Shum et al. 2016). The UPF2 dependent pathway requires the EJC factor RNPS1 (**Figure 1.3b**). These two pathways converge at a point where they both require UPF1 and UPF3B to activate NMD (Gehring et al. 2005).

The reason why different branches of the same pathway exist is unknown however it may be due to the availability of the NMD/EJC factors in that specific cell type (Huang & Wilkinson 2012). The exact mechanisms and complexity of the different NMD pathways and how these function in specific cell types across different developmental stages is still to be fully elucidated.

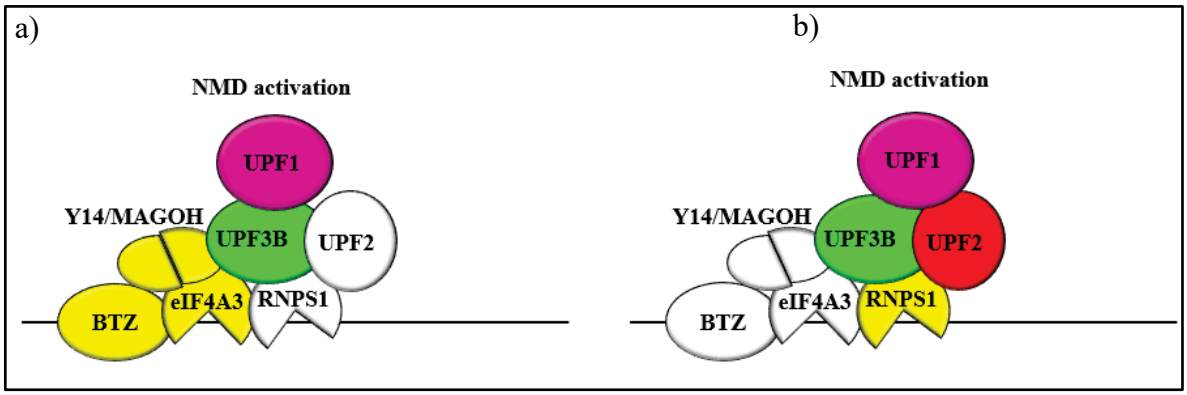


Figure 1.3: Alternative NMD pathway.

The two different branches of NMD that require different EJC factors. a) The UPF2 independent branch does not require the EJC factor RNPS1. b) The UPF2 dependent pathway requires the EJC factor RNPS1. These two pathways converge at a point where they both require UPF1 and UPF3B to activate NMD. White coloured shapes indicate absence of the protein. Picture modified from (Gehring et al. 2005).

1.2.2 The UPF3A dependent NMD pathway

In vertebrates, *UPF3B* is the only NMD factor that has an autosomal paralog gene, *UPF3A* (Jones & Wilkinson 2017). UPF3B is well established as a known strong NMD activator, whilst only limited studies on UPF3A have been conducted, which suggest UPF3A is a weak NMD activator (Kunz et al. 2006). *UPF3A* and *UPF3B* are expressed in the same tissue, although *UPF3B* expression is usually much higher than that of *UPF3A* except in the testis (Shum et al. 2016; Tarpey et al. 2007). Both UPF3A and UPF3B have a UPF2 and Y14 binding domain (Lykke-Andersen, Shu & Steitz 2000). The peptide sequence at the UPF2 binding domain is highly similar, but more divergent in the EJC binding domain (88% and 56% respectively) (Shum et al. 2016). The Y14 binding domain of these proteins is important in NMD activation and specifically, an arginine residue (R419) in UPF3B's Y14 binding domain is responsible for UPF3B's strong NMD activity. In UPF3A, this residue is replaced by alanine (A432) which weakens its interaction with Y14, and substitution of this alanine with arginine converts UPF3A into a strong NMD activator like UPF3B (Kunz et al. 2006; Shum et al. 2016).

When both UPF3A and UPF3B are present, UPF3B binds to UPF2. This is because UPF3B has a higher binding affinity for UPF2 compared to UPF3A. Consequently only a small amount of UPF3A is bond to UPF2 while most is degraded (**Figure 1.4a**). When UPF3B is absent, as in the case of patients with loss of function mutations in *UPF3B*, or in cells manipulated to deplete or delete *UPF3B*, there is stabilisation of the UPF3A protein (**Figure 1.4b**) (Chan et al. 2009; Jolly et al. 2013; Nguyen et al. 2012). The stabilisation of UPF3A in the absence of UPF3B has been postulated to be due to binding of UPF3A to UPF2, preventing its degradation. Aligned with evidence that UPF3A is a weak NMD activator, (Kunz et al. 2006), its stabilisation is proposed to partially rescue NMD function. Furthermore, the level of UPF3A stabilisation appears to be an important

determinant in the severity of clinical neurological phenotypes in patients with loss of function *UPF3B* mutations as patients with more UPF3A protein have a milder clinical phenotype and vice versa (Nguyen et al. 2012).

UPF3A came into existence during a *UPF3B* gene duplication event which occurred approximately 400 million years ago (Shum et al. 2016). As mentioned above, UPF3B is a strong NMD activator while UPF3A is a weak NMD activator (Kunz et al. 2006) and compensates for UPF3B when it is unavailable. However, the persistence of these gene paralogs suggests that they may have different functions (Innan & Kondrashov 2010). In support of this, recently *UPF3A* has been shown to act as an NMD inhibitor (Shum et al. 2016).

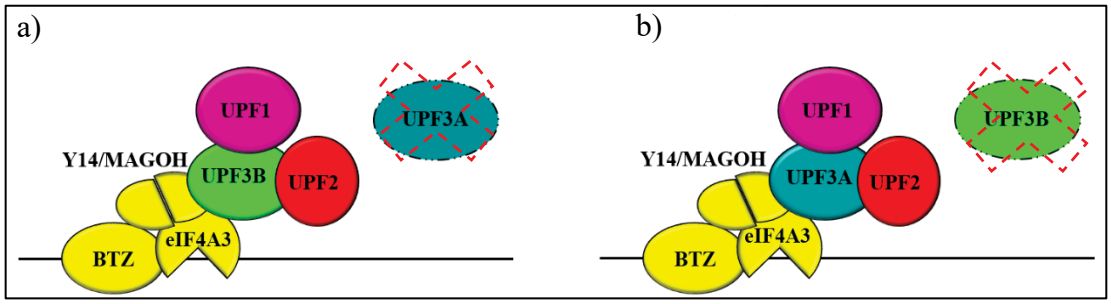


Figure 1.4: UPF3A dependent NMD pathway.

a) When UPF3B protein is present in the cell it binds to UPF2 and free UPF3A protein is degraded. b) When UPF3B protein is absent, the UPF3A protein is stabilised through bind to UPF2. The red cross indicates the absence of the protein. Picture modified from (Gehring et al. 2005).

1.2.3 NMD substrates

In addition to aberrant transcripts that contain PTCs, normal transcripts can be targeted for degradation by NMD due to the presence of 'NMD inducing features'. NMD targets normal mRNA transcripts that have an upstream open reading frame (uORF) (Karam et al. 2013) or alternative spliced mRNA, as alternative spliced exons may have stop codons in their uORFs (Rebbapragada & Lykke-Andersen 2009). Other NMD inducing features include leaky translation due to differential usage of an internal ribosomal entry site (Welch & Jacobson 1999), the presence of one or more introns in the 3'-untranslated region (UTR), a long 3'-UTR(>1.5 Kbp) (Schweingruber et al. 2013) and poly (A) site mutations or alternative polyadenylation sites (Martins et al. 2012). In all these different classes of NMD substrates, translation termination occurs at an unusual position which is distant from the poly(A) tail or with an EJC located between the stop codon and the poly(A) tail (Schweingruber et al. 2013).

1.2.4 Auto regulation of NMD

NMD regulates 3-10% of the transcriptome across the phylogenic tree, and therefore a defective NMD pathway can significantly alter the transcript profiles in cells and as such biological functions (He et al. 2003; Lelivelt & Culbertson 1999; Nguyen et al. 2012; Rehwinkel et al. 2005; Rodriguez-Gabriel et al. 2006; Wittmann, Hol & Jack 2006). The NMD mechanism therefore has a buffering mechanism that confers robustness of the NMD pathway, as such that when NMD is inhibited there is an upregulation of NMD factors triggered by a feedback regulatory loop (Chan et al. 2007; Huang, L et al. 2011; Mendell et al. 2004; Singh, Rebbapragada & Lykke-Andersen 2008). The mRNA species encoding the NMD factors that are upregulated during the feedback mechanism all contain NMD inducing features, and are direct NMD targets (**Table 1-1**).

NMD factors	3'-UTR length (nt)	uORF (amino acids)	3'-UTR intron
UPF1	1712	No	No[a]
UPF2	1275	Yes: 32, 35, 36 aa	No
UPF3A	889	No	No
UPF3B	852	No	No
SMG1	4715	Yes: 6 aa	No
SMG5	1361	Yes: 30 aa	No
SMG6	1653	No	No
SMG7	2259	Yes: 18 aa	No

[a]Intron located 3 nt downstream from stop codon.

Table 1-1: NMD factors involved in the NMD feedback regulatory network. (Yepiskoposyan et al. 2011)

1.3 Role of NMD in human embryonic stem cells

NMD factors are highly expressed in human pluripotent stem cells (hESCs) and their levels decrease upon loss of pluripotency (Alrahbeni et al. 2015; Bruno et al. 2011; Cho et al. 2012; Gong et al. 2009; Lou et al. 2016). NMD is important during early embryogenesis as loss of function mutations in *Upf1*, *Upf2*, *Upf3a*, *Smg1* and *Smg6* result in early embryonic lethality in mice (Li et al. 2015; McIlwain et al. 2010; Medghalchi et al. 2001; Shum et al. 2016; Weischenfeldt et al. 2008). Perturbation of the NMD pathway through knock down of *UPF1* and *UPF3B* in H9 hESCs and P19 cells leads to a reduction in the stem cell markers, NANOG and OCT3/4 and triggers the initial stages of endodermal differentiation (Lou et al. 2016; Lou et al. 2014).

In addition to having a role in maintaining stem cell pluripotency, NMD also has a role in controlling cell cycle progression as a perturbed NMD pathway via loss of *UPF1* and

UPF3B leads to an enrichment of cells in the G1 phase of different cell types including hESCs (Lou et al. 2016; Lou et al. 2014). In *Drosophila* deletion of NMD factors impaired cell proliferation and led to cell cycle arrest at the G2/M phase (Rehwinkel et al. 2005) while in HeLa cells it led to an early S phase arrest (Azzalin & Lingner 2006).

1.4 Role of NMD in neural development and function in animal models

The discovery that mutations in NMD and EJC factor genes are strongly associated with NDDs suggests that NMD is critical for neural development and function (Jaffrey & Wilkinson 2018; Nguyen et al. 2013). *UPF3B* null mice present with fear-conditioned learning and defects in neurogenesis such as failure of dendritic spine maturation and neural differentiation defects (Huang et al. 2018; Jolly et al. 2013). Expression of missense mutant *UPF3B* in rat neurons and mice neurons depleted of *UPF3B* have reduced branching of neurites (Alrahbeni et al. 2015; Jolly et al. 2013). Mice NSCs depleted of *UPF3B* are also hyper-proliferative and have a higher capacity for self-renewal (Huang et al. 2018; Jolly et al. 2013).

Mutations in other NMD and EJC factors in other species also have effects on neurogenesis and behaviour. For example haploinsufficiency of the EJC component *Magoh* in mice causes microcephaly due to intermediate neural progenitor depletion (Silver et al. 2010). Overexpression of Y14, another EJC component has been implicated in anxiety and abnormal social interaction in mice (Alachkar et al. 2013). In zebrafish, *upf1* morphants had brain patterning defects, predominantly at the midbrain-hindbrain boundary (Wittkopp et al. 2009). Heterozygous mutations in *upf2*, *smg1* and *smg6* in *Drosophila* causes disruption in the formation of neuromuscular junction synapse

structure, delayed neurotransmission responses and impaired synaptic vesicle cycling (Long et al. 2010). Commissural neurons from *Upf2* conditional knockout mice exhibited aberrant axonal trajectories (Colak et al. 2013).

Polypyrimidine tract binding proteins 1 (PTBP1) and PTBP2 are regulators of the neuron-specific alternative splicing program that are regulated by NMD (Boutz et al. 2007). They are important in the development of the nervous system. Mice deficient of *PTBP2* show neuronal progenitor defects and premature neurogenesis and die after birth (Licatalosi et al. 2012). PTBPs bind to regulatory sequences within or adjacent to alternative exons to repress or activate splicing or cause intron retention (Kafasla et al. 2012). PTBP1 is known to regulate nervous system-specific genes (Yap & Makeyev 2013) such as *Gabbr1* which encodes the $GABA_{B1}$ receptor mostly expressed at post synaptic terminals of inhibitory synapses (Pinard, Seddik & Bettler 2010) and *Dlg4* which encodes for the PSD-95 protein which is abundant in post synaptic neurons and regulates structure and function of excitatory synapses (Sheng & Hoogenraad 2007). Mice without the PSD-95 proteins show spatial learning defects, enhanced glutamate receptor-dependent long-term potentiation and reduced long term depression (Migaud et al. 1998). These function of NMD on regulating such important neuronal genes demonstrates the importance a well-functioning NMD pathway is for normal neurodevelopment and function and may explain some of the NDDs that present with a defective NMD.

1.5 Mutations in NMD factors cause genetic disease

Copy number variants and mutations in NMD and EJC genes have been reported to be associated with NDDs (Alachkar et al. 2013; Brunetti-Pierri, N. et al. 2008; Favaro et al. 2014; Favaro et al. 2011; Gulsuner et al. 2013; Nguyen et al. 2013). UPF3B, an NMD

factor whose loss of function causes NDDs is expressed in neurons and dendritic spines, which are essential structures, needed for proper neurotransmission, learning and memory processes (Laumonnier et al. 2010). Loss of function mutations in *UPF3B* cause a broad spectrum of NDDs that include autism, ADHD, childhood onset schizophrenia and ID (Addington et al. 2011; Lynch et al. 2012; Szyszka et al. 2012; Tarpey et al. 2007; Xu et al. 2013). Transcriptome profiling lymphoblastoid cell lines (LCLs) of patients with *UPF3B* mutations detected deregulated genes that have important neural functions such as neurogenesis and synaptic plasticity (Nguyen et al. 2012). Approximately 23% of transcripts are co-expressed between LCLs and brain tissue (Rollins et al. 2010). Heterozygous copy number losses of *UPF3A* has been associated with neural tube defects (Luo et al. 2000).

A heterozygous deletion of *UPF2* is associated with NDDs (Nguyen et al. 2013) and *de novo* missense *UPF2* mutations have been identified in patients with schizophrenia (Gulsuner et al. 2013). Transcriptome profiling of patients LCLs with heterozygous deletions of *UPF2* had thirty eight percent of deregulated genes overlapping with patients with hemizygous *UPF3B* mutations. Ninety five percent of these genes shared the same trend of deregulation, suggesting that the NDDs stemming from alteration of different genes encoding NMD factors may arise from deregulation of a common underlying set of transcripts. These genes also have important functions in neurodevelopment (Nguyen et al. 2013).

Thrombocytopenia absent radius (TAR) syndrome is caused by a heterozygous deletion and a non-coding single nucleotide polymorphism on the unaffected allele of *RBM8A* (Albers et al. 2012), which codes for the Y14 protein, a subunit of the EJC complex that

binds to UPF3B. About 7% of patients with a defective *RBM8A* present with ID (Alachkar et al. 2013). Micro deletions and duplication near the1q21.1 region which contains the *RBM8A* gene has been associated with ID, brain defects and congenital alterations without TAR syndrome (Rosenfeld et al. 2012). Some of these patients with abnormal *RBM8A* dosage have schizophrenia while others have schizophrenia with microcephaly and facial dysmorphic features (Brunetti-Pierri, Nicola et al. 2008).

A noncoding expansion in the EJC factor, *eIF4AIII* is also known to cause a craniofacial disorder associated with limb defects (Favaro et al. 2014) such as a midline cleft mandible, laryngeal defects and radial and tibial deficiencies associated with clubfeet called Richieri-Costa-Pereira syndrome (Favaro et al. 2011). The abnormal development of the pharyngeal arches are due to the reduction of *eIF4AIII* transcript levels (Favaro et al. 2014). Approximately 50% of affected individuals have learning and language disabilities (Favaro et al. 2011).

A homozygous amino acid substitution mutation in the *NBAS* gene located in a highly conserved position amongst species is known to cause a hereditary short stature (SOPH) syndrome. This syndrome also presents with optic nerve atrophy, postnatal growth retardation, facial dysmorphism and Pelger-Huet anomaly. Even though the mutation does not seem to affect the expression it affects its activity and it is assumed that this may therefore compromise its function in NMD (Maksimova et al. 2010). (Longman et al. 2013)

Patients with NDDs have also shown to have both copy number losses and duplication in NMD and EJC factors. Compared to a control cohort, a significant enrichment of copy number gains in *UPF2*, *SMG6*, *RMB8A*, *eIF4AIII* and *RNPS1* and copy number losses in *UPF3A* and *RBM8A* (Nguyen et al. 2013) were found in patients with NDDs.

1.6 Disease modelling tools

NDDs affect a considerable number of the world's population and creates a significant social, behavioural and economic burden to the patients, their families and the society overall. Currently available medications, such as anti-epileptic drugs (AEDs) are partially effective and have major side effects on nervous system function or development (Al-Harbi 2012). The search for better therapeutics for these disorders has been challenging, and only a few novel treatments have been developed in the last few decades. It appears the major obstacle along this drug development pipeline is the lack of suitable preclinical models that can be used for studying disease mechanisms, identifying therapeutic targets and effectively testing a larger number of potential drugs (Wen 2017).

It is desirable to study the molecular basis of a disease using the best available and appropriate models which provides the context of the affected cellular and tissue pathology as molecular pathways are shaped by cell type specific gene expression (Handley et al. 2015). Relevant human tissue or cell samples are however difficult to obtain and typically require surgery or are only available post-mortem. Isolated cells also cannot be maintained or expanded with conventional culture conditions and generally require immortalisation for long-term use (Kumar, Blangero & Curran 2018). In the absence of primary tissue or cells of interest, animal models (typical mice and rats) have been instrumental tools for modelling human diseases (Phillips et al. 2014).

1.6.1 The limitations of mice in disease modelling

While rodents continue to be tremendously valuable models for biomedical research, rodents do not always accurately model human disease or biological responses (Seok et al. 2013). The evolutionary distance between rodents and humans (diverged approximately 87 million years ago) (Springer et al. 2003) may in part explain the significant differences in biological function between these species which can prevent the recapitulation of human disease in mice (Avior, Sagi & Benvenisty 2016).

Mice and humans differ in many aspects such as embryonic development, genetic, physiological background and neocortical development including cortical expansion and the duration of human brain development. In addition to having a large brain size, the human cortex also has diverse types of interneurons compared to other mammals (Clancy, Darlington & Finlay 2001; Hamlin & Altschuld 2011; Hansen et al. 2013; Molnar & Clowry 2012; Strachan, Lindsay & Wilson 1997; Zhao & Bhattacharyya 2018). These differences may account for variable impacts that a particular mutation may have between mice and humans. For example monosomy X is viable in mice but lethal in humans (Saenger 1996), while mutations in *BLM* in humans cause Bloom syndrome but is fatal in mice (Chester et al. 1998). Lesch-Nyhan syndrome (Eiges et al. 2007; Urbach, Schuldiner & Benvenisty 2004) and Turner syndrome (Urbach & Benvenisty 2009) are some examples of diseases that also cannot be studied in mice due to species-specific differences.

1.6.2 iPSCs in disease modelling

The need for suitable model systems to recapitulate diseases caused by variants of unknown function has led to an advance in human disease research over the last decade. One critical development was that of 'somatic cell pluripotent reprogramming (cellular reprogramming)' technologies, where human somatic cells such as skin fibroblasts or blood cells are reprogramed into induced pluripotent stem cells (iPSCs) (**Figure 1.5**) (Kunisato et al. 2010; Takahashi et al. 2007; Takahashi & Yamanaka 2006; Yu et al. 2007). The directed differentiation of iPSCs into disease relevant cell types enables experiments to be performed on cell types otherwise difficult or impractical to obtain from patients. iPSCs are a genetically faithful model as they are derived from the patient and carry the patient's genetic identity. They therefore provide the opportunity to study cellular and molecular mechanisms in the same genetic environment for a given disease (Bellin et al. 2012; Wen et al. 2016). iPSCs provide a new platform for disease modelling and drug development (Wen 2017). For example in disease modelling patient's iPSCs were generated and differentiated into neurons to study PCDH19 Girls Clustering Epilepsy. In drug development, RG7800 was tested on motor neurons derived from iPSCs from patients with spinal Muscular Atrophy. RG7800 increased the survival of motor neurons and has been through clinical trials (Homan et al. 2018; Kletzl et al. 2019; Ratni et al. 2018).

There are several issues to however consider when using iPSCs based models. First and foremost, it requires access to patient derived cell lines, which is not always available. For instance, in mutations that lead to very early embryonic lethality such as *KHDC3L* and *TLE6* (Alazami et al. 2015; Zhang et al. 2019). Second, iPSCs can acquire new genetic alterations during reprogramming and subsequent culture (Puri & Nagy 2012).

Third the current practice in the field of iPSC disease modelling requires gene editing to correct the disease-causing mutations to provide the study of isogenic control cell lines. It is imperative to generate isogenic control cell lines as they have the same genetic background as the iPSCs. This involves further manipulation and selection, potentially affecting cellular phenotypes. Altogether, iPSCs reprogramming and gene editing is specialised and laborious technique that requires handling of large number of clonal lines to ensure selection of clonally pluripotent cells, precise gene-editing and preservation of karyotype. Even though hESCs and iPSCs have been shown to be very similar, there are important differences to note between the two cell types such as the persistence of transcriptional and epigenetic memory from the somatic cells tissue of origin (Bar-Nur et al. 2011; Chin et al. 2009; Doi et al. 2009; Kim, K et al. 2011; Ohi et al. 2011), differential DNA methylation signatures (Lister et al. 2011; Ruiz et al. 2012) and a higher degree of genetic aberrations acquisition in iPSCs (Gore et al. 2011; Yung et al. 2013). However, extended culture of iPSCs in culture conditions defined to support pluripotency has been shown to erase somatic cell memory at the transcriptional, epigenetic and genome topology levels (Krijger et al. 2016).

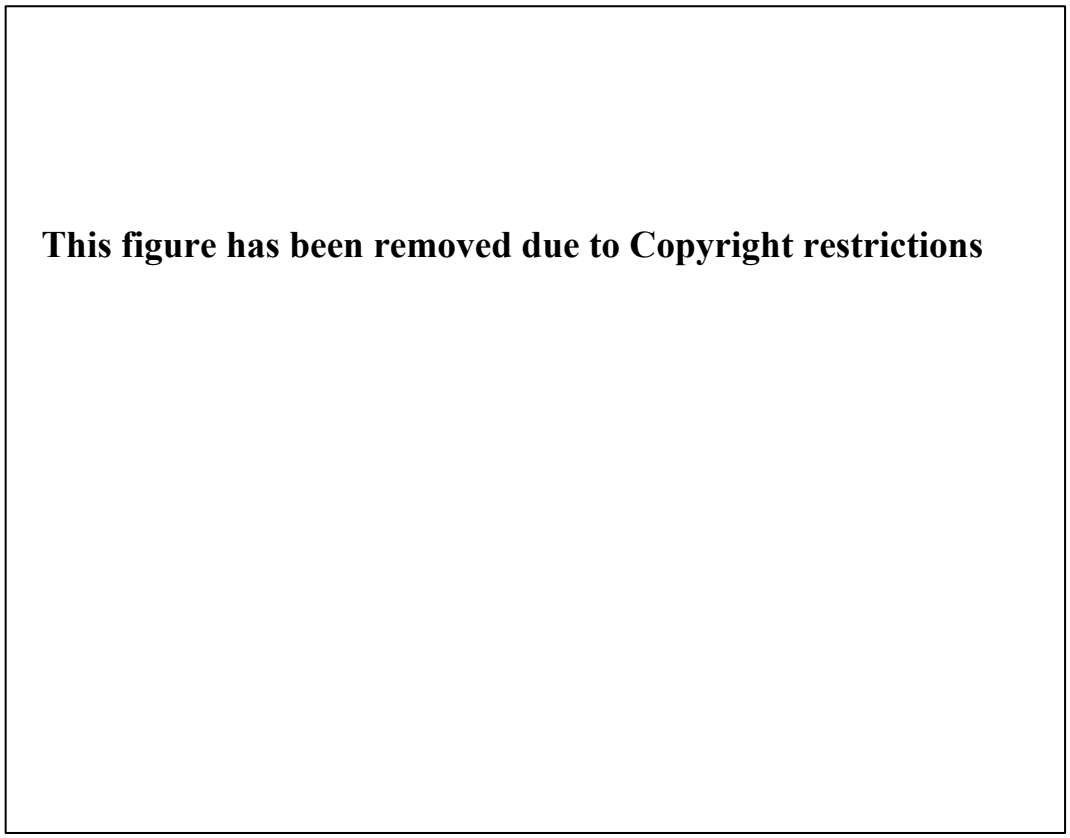


Figure 1.5: The use of iPSCs for regenerative medicine, disease modelling and drug screening.

Patient-derived somatic cells are reprogrammed to generate iPSCs carrying a disease specific genetic mutation. These cells have the ability to self-renew and are pluripotent and can be differentiated into the disease affected cell type for studying mechanisms of disease pathology. The generated iPSCs can be used for high-throughput screening of drugs. Figure was taken from (Ebert, Liang & Wu 2012).

1.6.3 The application of hESCs and gene editing technologies to model diseases

The field of genome engineering is rapidly evolving due to new technological developments. The ability to combine human embryonic stem cells (hESC)-based technology with advanced gene editing technology is an attractive approach to generate new *in vitro* disease models and resources for basic and applied biology research into genetic disease (Hendriks, Warren & Cowan 2016). It enables systematic interrogation of causative genetic variation across multiple diverse cell types and tissues of disease relevance (Ding, Qiurong et al. 2013).

1.6.3.1 Human embryonic stem cells (hESCs)

hESCs were first isolated in 1998 and are derived from the inner cell mass of the blastocyst (Thomson et al. 1998). hESC lines attracted attention as they have the ability to proliferate indefinitely *in-vitro* and maintain the potential to differentiate into any cell type of the body (Löser et al. 2010). The use of hESCs is important in the absence of patient-derived somatic cells for example in cases of genetic aberrations and aneuploidies that lead to very early embryonic lethality (As mentioned in section 1.6.2) (Urbach & Benvenisty 2009).

1.6.3.2 Gene editing technologies

Gene editing technologies employ the same concept. They use a specific sequence complementary to the targeted region, tethered to an endonuclease. The complementary sequence homes the endonuclease to the target gDNA where it cuts the genome generating double stranded breaks (DSBs). Gene editing technologies also exploit the endogenous DNA repair pathway that introduces mutations by deleting or inserting

nucleotides during the repair. They differ in the DNA recognition mechanism and the specific endonuclease involved. The first custom-engineered, site specific endonucleases used for genome editing in hESCs were the Zinc-Finger Nucleases (ZFNs) (Hockemeyer et al. 2009). These are fusion proteins composed of several tandem Zinc-finger DNA binding domains coupled to the *FokI* endonuclease catalytic domain. The DNA binding domain of ZFNs consists of three to six zinc finger DNA-binding domains (ZFDBD) assembled in an array. This array construction of the ZFN allows for specific targeting of genetic loci, as each ZFDBD binds to a specific nucleotide triplet. FokI endonuclease is only active when dimerised (**Figure 1.6a**) (Bibikova et al. 2003; Urnov et al. 2005). ZFNs are difficult to engineer and their design and construction remains technically challenging (Hendriks, Warren & Cowan 2016).

Another different custom-engineered endonuclease is the Transcription Activation-Like Effector Nuclease (TALEN) (**Figure 1.6b**). Like ZFNs, TALENs consist of a customised TALE DNA binding domain fused to a non-specific *FokI* nuclease domain. The TALE DNA binding domain comprises arrays of 33-35 amino acids (Joung & Sander 2013) where the amino acids in position 12 and 13 of each array determine nucleotide binding specificity (Bogdanove, Schornack & Lahaye 2010). As with the design of ZFNs, each DNA target sequence requires re-engineering of the TALE DNA binding domain (Hendriks, Warren & Cowan 2016).

Similar to the ZFNs and TALENs, the CRISPR/Cas9 technology stimulates DSBs at the target genomic site (Hsu & Zhang 2012). However, unlike the ZFNs and TALENs which use a method of tethering the *FokI* endonuclease catalytic domains to modular DNA binding proteins that dimerise to cause DSBs (Ran et al. 2013; Wood et al. 2011), the

CRISPR/Cas9 technology utilises the Cas9 endonuclease homed by a 20 nt guide RNA (gRNA) through Watson-Crick base pairing to a target DNA (**Figure 1.6c**) (Mali, Esvelt & Church 2013). Unlike the ZFNs and TALENs, the CRISPR/Cas9 genome editing technology is easy and efficient (Ran et al. 2013).

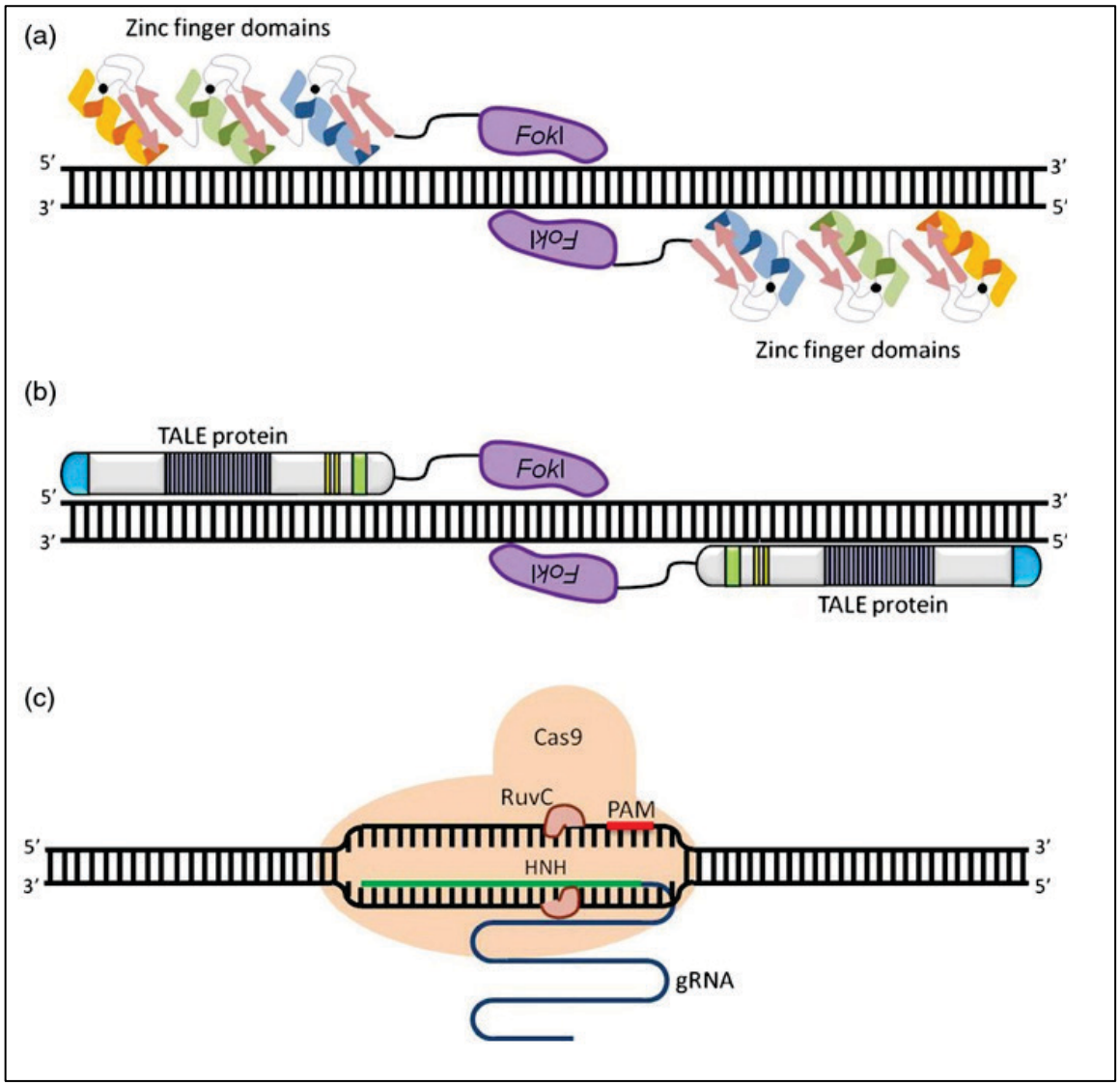


Figure 1.6: Gene editing technologies.

a,b) ZNFs and TALENs require a pair of fusion proteins with DNA binding domains fused to the *FokI* endonuclease on opposite DNA strands. These two fusion proteins are separated by a small fragment called a spacer sequence. In the nucleus, the *FokI* domains located at the C-termini of the chimeric protein dimerise to cause a double stranded break in the spacer sequence. c) The CRISPR/Cas9 technology generates DSBs by utilising the Cas9 endonuclease under the guidance of the gRNA to a complementary sequence in the genome. Figure taken from (Mahfouz, Piatek & Stewart 2014).

1.7 Hypothesis and aims

A defective NMD mechanism due to knockout/knockdown of NMD factors is known to cause a deregulation in approximately 3-10% of the whole transcriptome across multiple species (He et al. 2003; Lelivelt & Culbertson 1999; Nguyen et al. 2012; Rehwinkel et al. 2005; Rodriguez-Gabriel et al. 2006; Wittmann, Hol & Jack 2006). Genes that are deregulated as a consequence of malfunctioning NMD can explain the NDD phenotypes of patients but have not yet been systematically identified in a human brain context. Most of the research work on the NMD pathway has so far been performed using transformed cell lines, mouse/rat neural progenitor cells and mouse/rat stem cells derived neurons and other species such as the *Drosophila melanogaster* (Alrahbeni et al. 2015; Jolly et al. 2013; Lou et al. 2014; Nguyen et al. 2012; Rehwinkel et al. 2005).

Given that NMD regulates approximately 3-10% of the transcriptome and that at least a proportion of NMD targets can be species specific (Jolly et al. 2013; Rehwinkel et al. 2005) it is important to study the role of NMD in neurodevelopment in a relevant human brain disease model. There is paucity of suitable human cell-based disease models to investigate how a defective NMD pathway leads to NDDs, and in particular NDDs stemming from mutations in *UPF3B*, and its gene paralog *UPF3A*. Whilst the role of UPF3B in promoting NMD is well established, the role of UPF3A is enigmatic, with both NMD-promoting and NMD-inhibiting activity being reported in different experimental contexts (Shum et al. 2016). The main goal of this research is to generate *UPF3A* and *UPF3B* KO hESCs and differentiate them into neural stem cells (NSCS) and identify NMD targeted or regulated transcripts in *UPF3A* and *UPF3B* KO hESCs and hESCs derived neural stem cells. The significance of identifying these transcripts is that we will identify genes associated with phenotypes in *UPF3B* patients. It will also help in understanding the pathological mechanisms underpinning heterozygous copy losses of

UPF3A and the role of UPF3A in NMD (activator vs inhibitor). Having an understanding of the role of *UPF3A* and *UPF3B* in a brain cell model will help better understand the disease pathology and will help in drug design and discovery to improve patient's lives.

Hypothesis 1

UPF3A and *UPF3B* KO hESCs would be viable and not drastically affect pluripotency.

Aim 1

To use CRISPR/Cas9 genome editing technology to generate *UPF3A* and *UPF3B* KO hESCs clones.

Aim 2

Assess pluripotency in CRISPR/Cas9 edited hESCs.

Hypothesis 2

Transcriptome profiling of *UPF3A* and *UPF3B* KO hESCs and NSCs would elucidate the role of *UPF3A* and *UPF3B* in neurodevelopment and the role of *UPF3A* in NMD.

Aim 1

To differentiate *UPF3A* and *UPF3B* KO hESCs clones into NSCs.

Aim 2

To perform RNASeq on hESCs and hESC-derived neural stem cells in *UPF3A* and *UPF3B* KO clones.

Chapter 2 - Materials and Methods

2.1 Cell Culture Methods

2.1.1 List of equipment, reagents and medium

2.1.1.1 Cell culture equipment

- 37°C water bath
- Cell culture centrifuge with swing arm bucket rotor
- Cell culture humidified incubator, 5% CO_2, 37°C
- Biological safety cabinet class II (VWR, PA, USA)
- Dissecting microscope (Olympus SZX16 stereo microscope, Tokyo, Japan)
- Inverted microscope
- CoolCell®LX, freezing container (Bioscision Cat# BCS-405, CA, USA)
- Cell counter (Countess II FL, Life Technologies, CA, USA)
- Liquid nitrogen cell storage tank

2.1.1.2 Cell culture materials

- Six well plates (Corning® Costar® Cat# CLS3516, NJ, USA)
- 96 well plates (NUNC Cat# 167008, RKE, Denmark)
- Six well plate (NUNC Cat# 140675)
- 35 mm dishes (NUNC Cat# 150318)
- Cryovials (NUNC Cat# 377224)
- 10 ml Polypropylene tube graduated screw capped sterile (Technoplas Cat# P10316SU)
- Cell scrappers (TTP Cat# 99003)
- 1 mL disposable syringe (BD Cat# 302100)
- PricisionGlide® 21G needles (BD Cat# 302017)

2.1.1.3 Reagents

- Fetal Bovine Serum, ESC-qualified US origin (hESC FBS) (Gibco, Thermo Fisher Scientific Cat# 10439024, MA, USA).
- Fetal Bovine Serum, qualified, Australia (FBS) (Gibco, Thermo Fisher Scientific Cat# 10099141).
- Dulbecco's Modified Eagle Medium (DMEM, high glucose) (Gibco, Thermo Fisher Scientific Cat# 11995081).
- Dulbecco's Modified Eagle Medium/Nutrient Mixture F-12/GlutaMAX^{TM} supplement (DMEM/F12 GlutaMAX) (Gibco, Thermo Fisher Scientific Cat# 10565042).
- Dulbecco's phosphate-buffered sale (DPBS) (Gibco, Thermo Fisher Scientific Cat# 14190-144).

- Knockout Serum replacement (KSR) (Gibco, Thermo Fisher Scientific Cat# 10828028).
- β-Mercaptoethanol (55 mM) (Life Technologies Cat# 21985-023).
- Gelatin from bovine skin (Sigma-Aldrich Cat# G9391, NSW, Australia).
- Basic Fibroblast Growth Factor (AA 10-1055: bFGF) Human recombinant protein (Invitrogen, Thermo Fisher Scientific Cat# PHG0023).
- Human recombinant Epidermal Growth Factor (EGF) (StemCell Technologies, Cat# 78006.1, VIC, Australia).
- Y-27632 dihydrochloride (ROCK-Inhibitor) (Tocris Bioscience Cat# 1254, Bristol, UK).
- Neurobasal® medium (Gibco, Thermo Fisher Scientific Cat# 21103049).
- N-2 supplement (100X) (Gibco, Thermo Fisher Scientific Cat# 17502048).
- B-27 supplement, serum free (50X) (Gibco, Thermo Fisher Scientific Cat# 17504044).
- Trypan blue stain (0.4%) (Gibco, Thermo Fisher Scientific Cat# 15250016)
- Extra cellular matrix (ECM) from Engelbreth-Holm-Swarm murine sarcoma (Sigma-Aldrich Cat# E1270).
- Accutase (StemCell Technologies Cat# 07920).
- Insulin solution human (Sigma-Aldrich Cat# I9278).
- Dispase II (Gibco, Thermo Fisher Scientific Cat# 17105041).
- L-Glutamine (200 mM) (Gibco, Thermo Fisher Scientific Cat# 25030081).
- MEM Non-essential amino acids solution (NEAA) (Gibco, Thermo Fisher Scientific Cat# 11140050).
- Penicillin-Streptomycin (10,000 units) (P/S) (Gibco, Thermo Fisher Scientific Cat# 15140122).
- SB431542 (Tocris Bioscience Cat# 1614).
- Dorsomorphin (Sigma-Aldrich Cat# P5499).
- Poly-L-Ornithine solution (Sigma-Aldrich Cat# P4957).
- Laminin (Sigma-Aldrich Cat# L2020).
- Vitronectin XF™ (StemCell Technologies Cat# 07180).
- TeSR™ –E8™ (StemCell Technologies Cat# 05940).
- mTeSR-1 medium without phenol red (StemCell Technologies Cat# 85850).
- Gentle cell dissociation reagent (StemCell Technologies Cat# 07174).
- CellAdhere™ dilution buffer (StemCell Technologies Cat# 07183).
- Dimethylsulfoxide (DMSO) (Sigma-Aldrich Cat# D2438)
- Trypsin (Gibco, Thermo Fisher Scientific Cat# 2520056)

2.1.1.4 Medium/solutions for cell culture

Medium/solution	**Reagents**
iMEF medium	10% hESC FBS, 100U/mL P/S, 1X NEAA, DMEM (high glucose). Store at 4°C for 1 month.
Incomplete hESC medium	DMEM/F12 GlutaMAX, 20% KSR, 1X NEAA, 50 U/mL P/S. Store at 4°C for up to 1 months.
Complete hESC medium	Incomplete hESC medium, 0.1 mM β-Mercaptoethanol, 8ng/mL bFGF. Use immediately.
Complete hESC conditioned medium	Incomplete hESC medium conditioned with MEFs for 24hrs, 0.1 mM β-Mercaptoethanol, 8 ng/mL bFGF. Use immediately.
B-27 medium	Neurobasal medium, 1X B-27 supplement, 1 mM L-Glutamine, 100 U/mM P/S. Store at 4°C for up to 2 weeks.
N-2	DMEM/F12 GlutaMAX, 1X N-2 supplement, 5 µg/mL Insulin, 1 mM L-Glutamine, 1X NEAA, 0.1 mM β-Mercaptoethanol, 100 U/mL P/S. Store for up to 2 weeks.
Neural maintenance medium	1:1 N-2: B-27 medium. Store at 4°C for up to 2 weeks.
Neural induction medium #1	Neural maintenance medium, 10 µM SB431542, 1 µM Dorsomorphin. Store at 4°C and use within 5 days.
Neural induction medium #2	Neural maintenance medium, 1 µM Dorsomorphin. Store at 4°C and use within 5 days.
Dispase solution	Make 10 mg/mL stock by resuspending powder in DPBS without magnesium and calcium. Filter sterile through a 0.22 µM filter. Store at 4°C. Stable for 24 months.
Gelatin solution	Make 0.1% W/V gelatin solution in MilliQ water. Autoclave and store at 4°C. Filter through a 0.22 µM filter prior to use.
ECM	1:2 dilution of ECM and cold DMEM/F12 GlutaMax medium
Neural progenitor freezing medium	Neural maintenance medium with 10% DMSO and 20 ng/µL FGF and 20 ng/µL EGF.
hESC freezing medium	90% hESC FBS, 10% DMSO

2.1.2 Culturing hESCs

hESCs were cultured as either feeder depended or feeder free cultures. They were routinely cultured in feeder depended conditions and transferred onto feeder free conditions before experiments.

2.1.2.1 Standard hESCs culture on irradiated Mouse Embryonic Fibroblast (iMEF) feeder layer.

2.1.2.1.1 Preparation of iMEF Feeder cultures

Six-well plates were coated with 1 mL/well of 0.1% gelatin for a minimum of 2 hrs to overnight at 37°C before plating iMEFs (Stemcore, Ltd, Queensland, Australia). Gelatin was aspirated and left in the biological safety hood for 1 hr to dry before plating iMEFs. Cryopreserved iMEFs were thawed quickly by the addition of pre-warmed iMEF medium, to the cryovial of cells and transferring to a 10 mL warmed iMEF medium tube using a transfer pipette. The cells were centrifuged at 120 g for 5 mins. The supernatant was aspirated and the cells resuspended in iMEF medium. A 20 µL aliquot was removed and diluted 1:1 with trypan blue to perform a viable cell count using a haemocytometer. The cell suspension was diluted with iMEF medium to give $1.9x10^5$ cells/mL. The cells were mixed well by inversion and 1 mL of cell suspension added to each well of a gelatin coated 6-well plate, giving a density of 20, 000 MEFs/cm^2. An additional 1 mL of iMEF medium was added to each well of the plate. The cells were mixed to allow even spreading over the surface of the plate and incubated overnight at 37°C in 5% CO_2.

2.1.2.1.2 Culturing hESCs on iMEF feeder cultures

The H1 hESC line obtained from WiCell (Madison, WI, USA) were cultured on an iMEF feeder layer (Section 2.1.2.1.1) as described previously (Briggs et al. 2013) with an bFGF concentration of 8 ng/mL (Rosler et al. 2004) . The day after plating iMEFs, the medium was aspirated and the wells washed with Dulbecco's Phosphate Buffered Saline (DPBS) to remove any remaining FBS. 2 mL (section 2.1.2.3.1) or 1 mL (section 2.1.2.3.2.1) of complete hESC medium was added to each well and equilibrated for 1 hr in the incubator (37°C, 5% CO_2). hESCs were then seeded into each well (Section 2.1.2.3.1, 2.1.2.3.2.1). Cells were incubated at 37°C in 5% CO_2. Complete hESC medium was changed daily (and made fresh daily) with the morphology of hESCs colonies monitored closely by bright field analysis on an inverted microscope (Olympus CKX31, Tokyo, Japan) to determine when the cells required passaging. This was determined by the size of colonies and the presence of differentiation in the culture.

2.1.2.2 Standard hESC feeder-free culture conditions

2.1.2.2.1 iMEF conditioned medium

iMEF conditioned medium (CM) was prepared to harvest growth factors secreted by iMEFs and supply them to hESCs grown on feeder-free cultures. CM was prepared by plating $1x10^7$ iMEFs in a T125 flask with iMEF medium. A day after plating, the cells were washed once in DPBS. Incomplete hESCs medium (90 mL) was added and collected 24 hrs later. Medium was repeatedly replenished and collected every 24 hrs over a period of 6 days. The daily collected aliquots were filtered every second day as a batch with a 0.2 μM filter and stored at -80°C for up to 12 months.

2.1.2.2.2 Feeder-free hESCs culture using extracellular matrix (ECM)

The H1 clonal hESC line were cultured as described previously by (Xu et al. 2001). Six-well culture plates were coated with the ECM gel diluted 1:50 from stock in cold DMEM/F12 GlutaMax medium for between 2 hrs to overnight at 37°C. Before use, ECM was aspirated and the wells washed once with DPBS. 2 mL (section 2.1.2.3.1) or 1 mL (section 2.1.2.3.2.1, 2.1.2.3.2.3) of complete iMEF CM was added in each well and placed in the incubator to equilibrate for 1 hour. hESCs were seeded into each well (section 2.1.2.3.1 or 2.1.2.3.2.1 or 2.1.2.3.2.3). Cells were incubated at 37°C in 5% CO_2. Complete iMEF CM was changed daily (and made fresh daily) with the morphology of hESCs colonies monitored closely by bright/Phase contrast analysis on an inverted microscope to determine when the cells required passaging. This was determined by size of colonies and the presence of differentiating cells in culture.

2.1.2.2.3 Feeder-free hESCs culture using chemically defined TeSR™ –E8™ medium

H1 hESCs were cultured as described in the 'Maintenance of Human pluripotent stem cells in TeSR™ –E8™ technical manual' (StemCell Technologies). Six-well plates were coated with 1 mL/well of 10 μg/mL Vitronectin XF™ diluted with the cellAdhere™ dilution buffer for between 1 hr minimum to overnight at 37°C. Vitronectin XF™ was aspirated before use and the wells rinsed once with DPBS. 1 mL of TeSR™ –E8™ was added into each well and left in the incubator to equilibrate for 1 hr. hESCs were seeded into each well (section 2.1.2.3.2.2). Cells were incubated at 37°C in 5% CO_2. TeSR™ –E8™ medium was changed daily with the morphology of hESC colonies monitored closely by bright/Phase contrast analysis on an inverted microscope to determine when the cells required passaging. This was determined by size of colonies and the presence of differentiating cells in culture.

2.1.2.3 Passaging hESCs

hESCs were passaged every 5-7 days as required. Prior to passaging, plates were prepared as described (Section 2.1.2.1.1, 2.1.2.2.2 or 2.1.2.2.3) and returned to the incubator to equilibrate for approximately 1 hr with appropriate growth medium added.

2.1.2.3.1 Mechanical passaging of hESCs

Mechanical passaging was performed in a biological safety cabinet class II (VWR) using a dissecting microscope (Olympus SZX16 stereo microscope, Tokyo, Japan). hESC colonies were selected based on morphology at 40X magnification. Those that appeared pluripotent were dissected into smaller square shaped pieces using a 21-guage needle. Using a 200 µL pipette and barrier tip, dissected colonies were carefully lifted off the plate and transferred to the newly prepared plate. Care was taken to avoid cutting and transferring differentiating regions of hESC colonies. Pieces of the same hESC colony were transferred to different wells/plates to avoid subculture of colonies. Approximately 50-60 dissected squares were transferred to each well of a 6-well plate. Cells were incubated at 37°C in 5% CO_2 with daily medium changes.

2.1.2.3.2 Enzymatic passaging of hESCs

Prior to enzymatic passaging of hESCs any differentiating cells/colonies were removed from the culture by mechanically removing selected differentiating regions of colonies under the dissecting microscope. These regions were lifted off the plate using a 200 µL pipette and a barrier tip. Medium containing the differentiating cells was aspirated and 1 mL of PBS was added and aspirated to ensure removal of all differentiating cells from the culture. Enzymatic passaging was then immediately performed as described below.

2.1.2.3.2.1 Dispase passaging of hESCs

800 µL of 10 mg/mL Dispase pre-warmed to 37°C was added to each well of a 6-well plate. The plates were returned to the incubator for 2 mins or after the edges of colonies started to curl upwards. Dispase was removed by aspiration and 1 mL of incomplete hESC medium was added to the well to wash off any remaining Dispase. The medium was again aspirated and an additional 1 mL of incomplete hESC medium added. Using a cell scrapper hESC colonies were lifted off the surface of the plate and transferred to a 10 mL tube containing 6 mL of 37°C pre-warmed incomplete hESC medium using a 5 mL serological pipette. An additional 1 mL of medium was added to the well and any remaining hESC colonies collected and transferred to the tube. hESC colonies were centrifuged at 80 g for 3 mins with low break. The supernatant was removed by aspiration and 1 mL of complete hESC medium (or complete iMEF CM) added to the cell pellet. Using a 1000 µL pipette, the hESC colonies were gently resuspend and triturated by pipetting up and down 2-3 times until the colonies were mechanically dissociated into smaller pieces. Additional hESC medium was then added to ensure an adequate split ratio (approximately 1:4 to 1:5 split) and 1 mL of cell suspension transferred to pre-prepared hESC plates using a 5 mL serological pipette to avoid complete dissociation to single cells. The volume of medium was adjusted to 2 mL/well of a 6-well plate. Cells were incubated at 37°C in 5% CO_2 with daily medium changes.

2.1.2.3.2.2 Passaging cells grown using TeSR™-E™ medium

1 mL of gentle cell dissociation reagent was added for 10 mins at RT. The reagent was aspirated and 1 mL of DMEM/F12 GlutaMAX medium was added. The cells were gently detached using a cell scrapper. The colonies were transferred into a 10 mL tube containing 6 mL of warmed DMEM/F12 GlutaMAX medium. The wells were rinsed

once with 1 mL DMEM/F12 GlutaMAX medium and collected in the 10 mL tube. The cells were centrifuged at 300 g for 5 mins with low break. Medium was aspirated off and 1 mL of TeSR^{TM} –E8^{TM} medium was added to the cells and pipetted up and down with a P1000 pipette twice to triturate the colonies. A 1:6 – 1:10 dilution ratio (depending on the cell density prior to passaging) was performed and the cells re-plated. The cells were then transferred into the incubator at 37°C with 5% CO_2 in a humidified environment, re-feed daily and passaged every 6-7 days.

2.1.2.3.2.3 Accutase single cell passaging of hESCs

Fresh complete iMEF CM containing 10 μM of ROCK inhibitor (Y-27632 dihydrochloride) was added to cells for 2 hrs before single cell passaging. After 2 hrs, the medium was aspirated and cells rinsed once in DPBS and 800 μL of Accutase^{TM} cell detachment solution pre-warmed at 37°C was added to each well of a 6-well plate. The plate was returned to the incubator for 10 mins until hESC colonies could be seen to dissociate into single cells under the inverted microscope. 1 mL of incomplete hESC medium was added to the well and hESC removed from the well and transferred into a 10 mL tube with 6 mL of pre-warmed at 37°C incomplete hESC medium. An additional 1 mL of incomplete hESC medium was added to the well and any remaining hESCs removed from well and into the 10 mL tube. Cells were then collected by centrifugation at 120 g for 3 mins with low break. The supernatant was aspirated and the hESCs resuspended in 1 mL of complete iMEF CM containing 10 μM ROCK inhibitor using trituration with a P1000 μL pipette to generate a single cell suspension. Complete iMEF CM supplemented with 10 μM ROCK inhibitor was added to dilute cells appropriately and a 20 μL aliquot was removed and diluted 1:1 with trypan blue to perform a viable cell count using a haemocytometer. hESCs were plated at a high density by plating $1.68x10^5$ cells/cm^2 in pre-prepared plates (section 2.1.2.2.2) with adjustment of medium

to 2 mL/well of a 6-well plate. Cells were incubated at 37°C in 5% CO_2 with medium changed the following day to complete iMEF CM without ROCK inhibitor and subsequent daily medium changes.

2.1.2.4 Cryopreservation of hESCs

hESCs were expanded into a T25 cm^2 tissue culture flasks (ECM/iMEF) or 6-well plates (Vitronectin). hESC were passaged using Dispase (ECM/iMEFs) or Gentle cell dissociation reagent (Vitronectin) as described above. Cells harvested from T25 sized flask were frozen into a single cryogenic vial, while cells harvested from one well of a 6-well plate were frozen into a single cryogenic vial. After centrifugation cells were resuspended in 1 mL of hESC freezing medium ensuring minimum disruption to hESCs colony size. Resuspended cells were transferred into a cryogenic vial and placed into an isopropanol filled freezing containers and stored at -80°C. After 24 hrs the vials were transferred to liquid nitrogen for long term storage.

2.1.2.5 Thawing hESCs

To thaw hESCs, plates with iMEF feeder layers, ECM or Vitronectin were prepared 1 day in advance. The plates were rinsed once with DPBS and 1 mL of complete hESC medium, iMEF CM or TeSR-E8 medium with 10 μM ROCK inhibitor was added and placed in the incubator to equilibrate for 1 hr. 9 mL of incomplete hESCs medium (iMEFs, ECM based cultures) or DMEM/F12 GlutaMAX (Vitronectin based cultures) was warmed at 37°C in a water bath in a 10 mL tube. A cryogenic vial of the cells were quickly thawed by adding the warmed medium in the 10 mL tube to cells in a dropwise manner and collecting them in the same 10 mL tube with warmed medium. This was repeated until all cells in the cryogenic vial were collected. The cells were centrifuged at

80 g for 3 mins (iMEFs, ECM) or 300 g for 5 mins (Vitronectin) with low break. The supernatant was removed and cells resuspended in complete hESC medium, iMEF CM or TeSR-E8 with 10 μM ROCK inhibitor, mildly triturated by pipetting once or twice, and transferred to the plates with the equilibrated medium. Cells thawed from the harvest of T25 flask (i.e. ECM/iMEF based cultures) were distributed across two wells of a 6-well plates, while cells harvested from one well of the 6-well plate (i.e. Vitronectin based cultures) were distributed across six wells of a 6-well plate. Cells were then incubated at 37°C at 5% CO_2 in a humidified atmosphere. After 24 hrs the medium was changed to medium with no ROCK inhibitor. Medium changes were performed daily until cells were ready to be passaged.

2.1.3 CRISPR/Cas9 Genome Editing

The CRISPR/Cas9 gRNAs were designed to introduce gene disrupting deletions via exploitation of the NHEJ pathway in target genes *UPF1*, *UPF2*, *UPF3A* and *UPF3B* (see section 2.2). Validation of gRNA gene targeting was performed initially in HEK293T cells before use in hESCs. The validation protocol was based on Ran et al., 2013. The CRISPR/Cas9-gRNA delivery method employs the PX459 V2 plasmid (Addgene plasmid #62988, USA) that expresses the Cas9 protein, gRNA and the *PAC* gene encoding puromycin-N-acetyltransferase that confers resistance to the antibiotic puromycin.

2.1.3.1 Human Embryonic Kidney 293T cell line (HEK293T) culture

HEK293T cells were cultured in DMEM supplemented with 10% FBS and 100U Penicillin/Streptomycin (PenStrep). Cells were incubated in 5% CO_2 at 37°C in a humidified environment. Cells were passaged when 80% confluent by detaching the cells

with Trypsin as follows; medium was aspirated and cells rinsed once with DPBS. The cells were incubated in pre-warmed 37°C 0.25% Trypsin for 5 mins at 37°C. Complete DMEM was then added to neutralise the Trypsin, the dissociated cells were harvested by centrifugation at 120 g for 5 mins. The supernatant was aspirated and cells resuspended in the appropriate amount of complete DMEM. A cell count was performed and cells were replated at $1x10^6$ cell/well in a 6-well plate for transfection in 2 mL of medium.

2.1.3.2 Transfecting CRISPR/Cas9 plasmids in HEK293T cells

A day prior to transfections, HEK293T cells were seeded at a density of $1x10^6$ in a 6-well plate. After 24 hrs post seeding, transfections were performed using Lipofectamine 2000 (Cat# 11668019, Invitrogen, Thermo Fisher Scientific) as per manufacturer's instructions. Briefly 2 µg of plasmid DNA (PX459 V2 plasmid control or with inserted gRNAs) was added to 150 µL of Opti-MEM (Cat# 3198062, Gibco, Thermo Fisher Scientific) at RT. In a second tube, 8 µL of Lipofectamine 2000 was added to 150 µL of Opti-MEM at RT. The DNA: Opti-MEM solution was then added to the Opti-MEM: - Lipofectamine solution and incubated for 5 mins at RT. This mixture was added dropwise to cells while swirling the plate. After 24 hrs, puromycin selection was performed by adding 2 µg/mL of puromycin to cells for 48 hrs. Cells were expanded and cell pellets collected for RNA, gDNA and protein analysis.

2.1.3.3 Nucleofecting gRNA CRISPR/Cas9 plasmids in hESCs

Nucleofection was performed using the Amaxa P3 Primary Cell 4D nucleofector™ X kit L (Cat# V4XP-3024, Lonza, Mount Waverley, VIC). 2 hrs before nucleofection, complete iMEF CM was replaced with fresh iMEF CM containing 10 µM ROCK inhibitor. Complete hESC medium containing 10 µM ROCK inhibitor was also added to freshly prepared 100 mm iMEF coated culture dishes to equilibrate at 37°C in the

incubator. Single cell suspensions were collected by detaching cells with Accutase. A cell count was performed and the required cells ($1x10^6$ cells/nucleofection) were centrifuged at 115 g for 3 mins with low break. Medium was aspirated and cells were resuspended carefully in 100 µL of nucleofector solution (82 µL of nucleofector solution and 18 µL of supplement) at RT as per manufactures instruction. The cell solution was then added to 2 µg of plasmid DNA. The following plasmids were nucleofected: 1. p.GFPmax (negative control); 2. PX459 V2 (positive control); 3. PX459 V2 -G1-3 (where G1-3 represents one of the three gRNAs used to target each gene (*UPF3A* and *UPF3B*). Nucleofection was performed using program CB-150 with the Lonza 4D Nucleofector™ unit. Nucleofected cells were resuspended in 500 µL of equilibrated complete hESC medium supplemented with the ROCK inhibitor, pipetted up and down three times gently and placed in 100 mm dishes containing iMEFs. The cells were evenly distributed and placed in the incubator.

2.1.3.4 Selection and expansion of CRISPR/Cas9 edited clones.

See optimised protocol in **Appendix 1**.

2.1.4 Neural differentiation

2.1.4.1 Neural induction of hESCs

Neural differentiation was performed on hESC colonies (Lukovic et al. 2017) using the dual SMAD inhibition method in a 2D adhesion culture system (Shi, Kirwan & Livesey 2012). High quality pluripotent hESC cultures (≥ 90% pluripotent, <10% differentiated) were passaged onto a 6-well Nunc plate coated with 10 µg/mL of Vitronectin. After 5 days, cells were washed once in DPBS and 2 mL of Neural Induction Medium 1 (NIM1)

was added to cells. This event was called Day 0 of neural differentiation. Medium was changed every second day.

2.1.4.2 Passaging of Neuroepithelial sheet

At day 6 of neural differentiation, formation of a 'neuroepithelial sheet' was observed in which cells displayed a uniform characteristic compact morphology with a small prominent nuclei. At day 7 of neural differentiation, medium was changed to Neural Induction Medium 2 (NIM2). Medium was changed every second day until day 10 where the neuroepithelial sheet was denser and thickened at which point it was passaged (Neural Differentiation Passage 1). Prior to passaging, Nunc dishes (35 mm) were coated with Poly-L-ornithine (PLO) for 4 hrs then aspirated and coated with 20 µg/mL of Laminin (diluted in cold DPBS) overnight. The dishes were rinsed once with DPBS before use and 1 mL medium added and equilibrated for 1 hr.

The neuroepithelial sheet was passaged by adding 800 µL/well of pre-warmed 37°C Dispase and incubating at 37°C for 5 mins. The medium and Dispase were aspirated and cells rinsed with 1 mL Neural Maintenance (NM) medium. The medium was aspirated and 1 mL of NM medium was added. A cell scraper was used to carefully detach the neuroepithelial sheet from the plate. The cells were transferred to a 10 mL tube containing 6 mL of pre-warmed NM medium with a 5 mL serological pipette. Remaining cells in the wells were collected. The harvested cells were centrifuged at 80 g for 3 mins with low break. The medium was aspirated and cells were resuspended in NIM2 supplemented with 10 µM ROCK inhibitor and 20 ng/mL bFGF. The tube was inverted to mix cells and distributed across 2 Nunc dishes (i.e. equivalent to a 1:2 split). The next day medium was changed to NM medium supplemented with 20 ng/mL bFGF and refed every second day.

2.1.4.3 Passaging of Neural Rosettes

By day 13 of neural differentiation, the presence of 'neural rosettes' started appearing in the cultures. Neural rosettes represent neural stem cells (NSCs) adopting an elongation morphology and arranged in a 'pin-wheel' architecture dictated by an overt acquisition of apical-basal polarisation (Wilson & Stice 2006). At day 15 neural rosettes structures were manually passaged (Neural Differentiation Passage 2); Neural rosettes were manually dissected (as per manual passaging of hESCs), avoiding the edges (as they contain contaminating neural crest cells) and transferred into a 35 mm dish pre-coated with PLO/laminin and pre-equilibrated with NM medium supplemented with 20 ng/mL of bFGF.

Rosettes from two 35 mm dishes were transferred into one 35 mm dish to generate a high density culture. The next day medium was changed to neural maintenance medium without bFGF. Medium was changed every second day. Five days after manual selection of neural rosettes (i.e. day 20 of neural differentiation) , cells were passaged again (Neural differentiation Passage 3) by adding 800 µL of pre-warmed 37°C Dispase directly into the medium and incubated for 5 mins at 37°C. Cells were collected and centrifuged at 80 g for 3 mins with low break. The medium was aspirated and 1 mL NM medium added to cells and using a P1000 pipette, pipetted once to break large neural rosettes and plated into a 35 mm PLO/laminin coated dish containing equilibrated NM medium. NM medium and changed every second day.

2.1.4.4 Dissociating Neural Rosettes to Isolate Neural Stem cells

At 5 days after Dispase passaging the neural rosettes (i.e. day 25 of neural differentiation), neurons were observed at the periphery of the rosettes. At this stage the cultures were single cell passaged using Accutase (Neural Differentiation Passage 4). Medium was

aspirated from the cells, and the cells washed once with DPBS. Pre-warmed 37°C Accutase (700 µL) was added to the cells and incubated at 37°C for 10 mins. One mL of NM medium was added to the dish and the cells pipetted up and down twice using a P1000 pipette. Cells were collected and centrifuged at 120 g for 3 mins with low break. The medium was aspirated and 1 mL of NM medium supplemented with 20 ng/mL of bFGF and 20 ng/mL of EGF was added to cells and pipetted up and down four times to generate a single cell suspension. Cells were plated at a density of $1x10^6$ cells/well of a 6-well Nunc coated plate with ECM. Medium was changed every two days and cells passaged when they reached 90% confluency. From this point the cells in these cultures were referred to as NSCs.

2.1.4.5 Maintaining neural stem cells

NSCs cultures were passaged by Accutase dissociation upon reaching 80-90% confluency (NSC Passage). Six-well plates were pre-coated with ECM and subsequently rinsed with DPBS. 1 mL NM medium supplemented with 20 ng/mL of bFGF and 20 ng/mL EGF was added to the wells and equilibrated in the incubator for 1 hr. NSCs were rinsed once with DPBS and 800 µL of pre-warmed 37°C Accutase added and incubated until the NSCs started to lift off the surface. 1 mL of NM medium was added to each well and NSCs removed from the well and transferred into a 10 mL tube with 5 mL of NM medium pre-warmed at 37°C. An additional 1 mL of NM medium was added to the well and any remaining NSCs collected and added to the 10 mL tube. Cells were centrifuged at 120 g for 5 mins with low break. The supernatant was aspirated and the NSCs resuspended in 1 mL of NM medium supplemented with 20 ng/µL of bFGF and 20 ng/µL of EGF and using a P1000 pipette to generate a single cell suspension. A viable cell count was performed using a haemocytometer. NSCs were plated at a density by of $1x10^6$

cells/well of a 6-well plate. Medium was changed every second day with freshly added 20 ng/mL of bFGF and 20 ng/mL of EGF. NSCs were re-passaged once 80-90% confluent.

2.1.4.6 Cryopreservation of NSCs

At Passage 4 NSCs were cryopreserved. After Accutase dissociation, $2x10^6$ cells were resuspended into NSC freezing medium and transferred to a cryovial. The cryovial was placed in a coolCell®LX freezing container and stored at -80°C overnight and then transferred into liquid nitrogen for long term storage.

2.1.5 Cell proliferation assay

Cell proliferation assays were performed using the CellTiter 96® AQ_{ueous} non-radioactive cell proliferation assay (Cat# G5421, Promega) according to manufacturer's instructions. $3x10^4$ hESCs were plated in quadruplicate on a Vitronectin coated 96-well plate with 100 µL of mTeSR-1 medium without phenol red (Cat# 85850, StemCell Technologies). 20 µL of MTS/PMS was added to the medium and incubated for 1 hr before reading. Plate readings were taken at day 0, day 2, day 4 and day 6. Absorbance was measured at 490 nm using a FLUOstar Omega photospectromic plate reader (BMG LabTech, VIC, Australia).

2.1.6 Cell Cycle analysis by DNA content

Single cell passaging was performed on hESCs. A cell count was performed and $2x10^6$ cells were collected from each clonal hESC line and centrifuged at 120 g for 5 mins with low break. Cells were washed with DPBS and centrifuged at 200 g for 5 mins at 4°C. The

cells were then resuspended in 1 mL ice cold DPBS. The cells were added dropwise while vortexing gently into 9 mL of cold 70% ethanol in a 15 mL polypropylene centrifuge tube. Cells were stored at 4°C until they were ready for analysis. For analysis cells were centrifuged at 200 g for 10 mins at 4°C and resuspend in 3 mL cold DPBS. The cells were again centrifuged at 200 g for 10 mins at 4°C and resuspended in 500 µL of PI/Triton X-100 staining solution containing 50 µg/mL of Propidium iodide (Cat# P4864, Sigma-Aldrich) and 100 µg/mL of RNase A (Cat# 1007885, Qiagen) in 0.1% Triton X-100 (v/v) (Cat# 8787, Sigma-Aldrich) in 1X PBS (137 mM NaCl, 2.7 mM KCI, 10 mM Na_2HPO_4, and 2 mM KH_2PO_4, pH 7.4). The cells were incubated at 37°C for 15 mins and analysed by flow cytometry using the BD FACS Canto1 (NJ, USA) and analysis performed using the BD FACS Diva software. Live cells were plotted using the FSC/SSC (representing the distribution of cells in the light scatter based on size and intracellular composition).

2.1.7 Microscopy

Fluorescence was observed using the Zeiss AxioImager M2 fluorescence microscope (Carl Zeiss, Germany). Fluorescent images were acquired with the AxioCam MRm high resolution camera and the Axio Vision software Vs4.9.1.0 (Car Zeiss, Germany). Phase contrast images were observed on Zeiss Vert.A1 inverted microscope and images captured using the AxioCam MRm high resolution camera and the Axio Vision software Vs4.9.1.0 Carl Zeiss, Germany).

2.1.8 Karyotype analysis

G-band chromosome analysis of hESCs clones was performed by SA Pathology (SA, Australia). A total of 20 metaphase chromosomes were analysed per cell line.

2.2 CRISPR/Cas9 Cloning Methods

2.2.1 gRNA design

gRNAs for *UPF1*, *UPF2*, *UPF3A* and *UPF3B* were designed using an online gRNA CRISPR design tool (http://tools.genome-engineering.org). Sequences at the beginning of the genes ORFs were used as input sequences. Three guides with the highest output score were selected (**Table 2-1**). The score is based on 100 as the percentage of the guide binding the specific DNA sequence minus the off-target score (possibility of it binding to off-target regions in the genome with high sequence similarity). To facilitate ligation of a dsDNA fragment encoding the gRNA into the *Bsb1* restriction site of PX459 V2, nucleotide (nt) overhang sequences were added: 5'-CACC-3' at the 5' end of the gRNA and 3'-CAAA-5' added at the 5'end of the complementary strand (**Figure 2.1**). A **G** nt was also added at the beginning of the gRNA sequence if not already present to ensure compatibility with transcription via a U6 promoter (Ran et al. 2013).

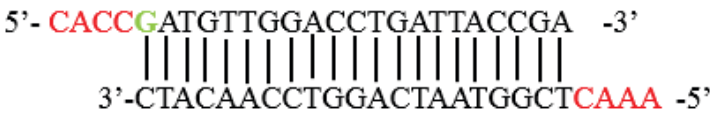


Figure 2.1: The nucleotide overhangs in the gRNA are important for ligation into the CRISPR plasmid.

After designing gRNAs the nt overhangs in red were added to the gRNA sequence. These overhangs are needed to ligate the gRNA into the *Bbs1* site in the CRISPR plasmid. The highlighted G (green) is important for U6 transcription and should be added if absent from the gRNA. Figure taken from (Ran et al. 2013).

NMD Factors	*Guide*	*Guide sequence*	*Scores*
UPF1	G1	AAAGTCGGTGAACTCGAACT	94
	G2	GGGGAGGTCCTTCGTGTAAT	94
	G3	AGAAGACACCTATTACACGA	90
UPF2	G1	GCTGGCTTTTTACGCTCAGC	89
	G2	GTGAGCTTGATATCGTCTTT	88
	G3	GCGTAAAAAGCCAGCAAGTA	74
UPF3A	G1	AACTTCGTCCTCCGGTTGCG	97
	G2	ACTTCGTCCTCCGGTTGCGG	96
	G3	CCGCAACCGGAGGACGAAGT	93
UPF3B	G1	GTAACCCTGTTAACCCCCGC	96
	G2	AACCCTGTTAACCCCCGCCG	95
	G3	TAACCCTGTTAACCCCCGCC	94

Table 2-1: gRNA sequences from the online CRISPR design tool.

2.2.2 Oligonucleotide annealing and phosphorylation

The oligonucleotides encoding the top and bottom strands of each gRNA were synthesized by Geneworks (Adelaide, Australia). Complementary fragments were phosphorylated and annealed by adding 1 µL (100 µM) of each DNA fragment, 1 µL 10X T4 ligation buffer (Cat# B0202S, New England BioLabs, MA, USA), 6.5 µL H_2O and 0.5 µL T4 PNK ligase (Cat# M0201S, New England BioLabs). The cycling conditions were; 37°C for 30 mins, 95°C for 5 mins and the temperature ramped down to 25°C at 5°C per min. The annealed oligonucleotides were diluted 1:250.

2.2.3 Plasmid digestion

Five µg of PX459 V2, 5 µL of FastDigest Bpil (Cat# FD1014, Thermo Fisher Scientific), 5 µL of FastAP (Cat# EF0654, Thermo Fisher Scientific) and 5 µL of 10X FastDigest buffer (Cat# FD1014, Thermo Fisher Scientific) were added together and made up to a total reaction mixture of 50 µL using H_2O and incubated overnight at 37°C with shaking. The digested DNA was visualised via DNA Agarose Gel Electrophoresis: (1% Agarose gel (Cat# HE121-0500, ProSciTech) and buffer consisting of 1X Tris Borate EDTA buffer (TBE):- 89 mM Tris Base, 89 mM Boric acid, 2 mM EDTA).

2.2.4 Gel Extraction

After running the digested plasmid on a 1% Agarose gel for 30 mins the DNA was excised from the gel and purified using the QIAquick gel extraction kit (Cat# 28704, Qiagen, Hilden, Germany) according to the manufacturer's instructions.

2.2.5 Plasmid ligation

A ligation reaction was set up by adding 50 ng of the *Bbsl* digested PX459 V2 plasmid, 1 µL of the diluted phosphorylated and annealed oligo duplexes, 2 µL of 10X T4 DNA ligase buffer (Cat# B0202S, New England BioLabs), 1 µL of T4 DNA ligase (Cat# M0202S, New England BioLabs) and made to a total reaction mixture of 10 µL with H_2O. The ligation reaction mixture was incubated at 16°C overnight in a thermocycler.

2.2.6 Transformation

Two µL of the ligated plasmid was added to 20 µL of ice cold chemically competent DHα5 cells (F-; φ80*lac*ZΔM15; Δ(*lac*ZYA-argF)U169; *deo*R; *rec*A1; *end*A1; *hsd*R17(r_k-m_k+); *pho*A; *sup*E44; *thi*-1; *gyr*A96; *rel*A1; γ-) (Cat # 18265017, Invitrogen) and incubated on ice for 10 mins. The cells were heat shocked at 42°C for 30s and

immediately put on ice for 2 mins. 100 µL of SOC medium (Cat # 15544034, Invitrogen) was added to the cells and plated on a Luria Broth (LB) agar plate containing 100 µg/mL ampicillin. The plates were incubated overnight at 37°C. Three colonies were selected per ligation and inoculated into a 3 mL of LB broth with 100 µg/mL ampicillin (Cat # A9518, Sigma). The culture was incubated while shaking at 37°C overnight. After 24 hrs, plasmid DNA was isolated using a QIAprep spin minikit (Cat# 27106, Qiagen) according to the manufacturer's instructions. Restriction DNA digest and Sanger sequencing reactions were performed on the extracted DNA to confirm correct ligation (See below).

2.2.7 DNA restriction digest

To confirm ligation, restriction digest reactions were performed as follows: 500 ng of plasmid, 2 µL of NEB buffer 2 (Cat# 7002S, New England BioLabs), 2 µL of BSA (10X) (Cat# 9001S, New England BioLabs), 0.5 µL of Kpnl (Cat# R0142s, New England BioLabs), 0.5 µL of EcoRV (Cat# R0195S, New England BioLabs) were combined and H_2O added to a final volume made up to 20 µL. Reactions were incubated overnight at 37°C with shaking and products separated by agarose gel electrophoresis as described above. Digested plasmids exhibiting the expected band sizes were subsequently subjected to Sanger sequencing to verify the fidelity of the inserted gRNA sequence using the primer: 5'- GAGGGCCTATTTCCCATGATTCC-3'

2.2.8 Plasmid DNA extraction for transfections and nucleofections

For preparation of transfection/ nucleofection competent plasmids, cultures of bacterial clones harbouring plasmids of interest were prepared and processed using the QIAfilter Plasmid Midi Kit (Cat# 12243, Qiagen) for transfections and Endofree Plasmid Maxi kit (Cat# 12362, Qiagen) for nucleofection's according to the manufacturer's instructions.

2.2.9 Glycerol stocks

Glycerol stocks of bacterial clones were prepared from the LB broth cultures by adding an equal volume of 80% glycerol (Cat# G5516, Sigma-Aldrich) to the bacterial culture. Glycerol stocks were immediately stored at -80°C.

2.2.10 pGEM-T Easy vector cloning

PCR amplified products were ligated using a 1:3 ratio with the pGEM-T Easy vector overnight at 4°C following manufactures instructions (Cat# A1360, Promega, Madison, WI, USA).

2.3 Biochemical Methods

2.3.1 RNA extraction with RNeasy isolation kit

RNA isolation was performed using Trizol (Cat# 15596018, Ambion) and the RNeasy mini kit (Cat# 74104, Qiagen) as per manufactures instructions. Briefly, Trizol (0.5 mL) was added to thawed cell pellets and incubated at RT for 5 mins. Chloroform (Cat# 1024451000, Amsure^{R}) (100 µL) was added to Trizol-cell solution, mixed vigorously by shaking for 1 min and left at RT for 3 mins. Samples were centrifuged at 200 g for 15 mins at 4°C. The upper aqueous phases containing RNA were transferred to a tube. 1 volume of 70% ethanol (Cat# 437433T, VWR) was added to the homogenised lysate and mixed well by pipetting and added to RNeasy mini-spin columns and processed as per manufacturer's instructions. RNA was separated using Gel Electrophoresis (1% Agarose gel) to check integrity. RNA yield was quantitated using the Nanodrop 1000 photo-spectrometer (Thermo Fischer Scientific).

2.3.2 RNA extraction using the Maxwell® RSC Instrument

RNA extractions were performed according to the Maxwell's RSC simple RNA extraction kit (Cat# AS1390, Promega) following manufacturer's instructions. 200 µL of chilled 1:50 1-Thioglycerol: Homogenization solution was added into each thawed cell pellet. The samples were vortexed to create an even cell suspension. 200 µL of lysis buffer was added to cells and vortexed vigorously for 15 secs. The samples were then processed on the Maxwell RSC instrument. RNA was quantitated using the Nanodrop and ran on a 1% Agarose gel to check for RNA integrity. Samples were stored at -80°C.

2.3.3 cDNA synthesis

Two µg of RNA was added to 1 µL of random heximers (100 ng) (RP-6-AMINO, Geneworks) and diluted with H_2O to a final volume of 12 µL. This was heated to 65°C for 5 mins and incubated on ice for 1 min. Subsequently, 4 µL of 5X first strand buffer (Cat# P/N y02321, Invitrogen), 2 µL of 0.1M DTT (Cat# P/N y00147, Invitrogen), 1 µL of dNTPs (Cat# P/N 55082-P/N 55085, Invitrogen) and 1 µL of superscript III RT (Cat# 56575, Invitrogen) were added and subjected to the following thermocycler protocol: 25°C for 5 min, 50°C for 50 min and inactivated by heating at 70°C for 15 mins. cDNA was diluted 1:3 and stored at -20°C.

2.3.4 Isolation of genomic DNA from cell

Genomic DNA (gDNA) was extracted from cells pellets using High Pure PCR template preparation kit (Cat# 11796828001, Roche) according to the manufacturer's instructions.

2.3.5 Isolation of gDNA from hESCs small colonies

hESC clones generated after CRISPR/Cas9 genome editing were genotyped without extensive clonal expansion. A small section of cells dissected from each clone at the time of clonal isolation was harvested and gDNA extracted using the QuickExtract DNA Extraction solution (Cat# QE0905T, Epicentre) according to manufacturer's instructions.

2.3.6 Preparation of gDNA for whole genome sequencing

DNA was extracted from cells pellets using the DNAeasy blood and Tissue Kit (Cat# 69504, Qiagen) according to manufacturer's instructions. The DNA quality was analysed by running on a 1% Agarose gel with 1X TBE buffer. For DNA quantification, DNA were prepared using the Qubit dsDNA BR assay kit (Q32853, Invitrogen) according to the manufacturer's instructing and analysed using the Qubit 3 Fluorometry.

2.3.7 Polymerase Chain Reaction (PCR) using Taq DNA polymerase

PCR reactions that used the Taq DNA polymerase (Cat# 11146173001, Roche, Basel, Switzerland) were performed according to the manufactures instruction. PCR reactions that used the MasterAmp^{TM} 2X PCR PreMix Buffer was prepared by adding 10 µL of the MasterAmp^{TM} 2X PCR PreMix Buffer, 1 µL of forward primer, 1 µL of reverse primer, 1 µL of DNA (100 ng/mL)/ 1 µL cDNA , 0.4 µL of Taq DNA polymerase and made up to 20 µL with water. The cycling conditions were the same as using Taq DNA polymerase, following manufactures instructions. Forward and reverse primers and the annealing temperatures are listed in **Table 2-2.**

2.3.8 Bacterial Colony PCR

Bacterial colony PCRs were performed on bacterial colonies selected after the transformation of the ligated cDNA PCR product into the pGEM-T Easy vector (section 2.2.10). This was performed by adding 10 µL of MasterAmp^{TM} 2X PCR PreMix Buffer J, 1 µL of cDNA Forward primer, 1 µL of cDNA Reverse primer, 0.4 µL of Taq DNA polymerase and made up to 20 µL with water. A pipette tip was then used to scrape off a small section of the bacterial colony and inoculated into the prepared PCR reaction. The

PCR reaction was conducted as per the Taq DNA polymerase manufacturer's instructions with an annealing temperature of 58°C.

2.3.9 PCR using KAPA HiFi DNA polymerase

DNA amplified using KAPA HiFi DNA polymerase (Cat# KP-KK2101, KAPA Biosystems) was performed according to the manufacturer's instructions. Forward and reverse primers and the annealing temperatures are listed in **Table 2-2.**

2.3.10 PCR using Herculase DNA polymerase

DNA amplified using Herculase II Fusion DNA polymerase (Cat# 600677, Agilent Technologies) was performed according to the manufacturer's instruction. Forward and reverse primers and the annealing temperatures are listed in **Table 2-2.**

2.3.11 PCR purification

PCR products were purified using the QIAQuick PCR purification kit (Cat# 28106, Qiagen) according to the manufacturer's instructions.

2.3.12 Sanger Sequencing

Purified DNA was sequenced using BigDye™ Terminator sequencing kit (Applied Biosystem™ BigDye Terminator v3.1 Cycle sequencing kit, Cat# 4337455, Thermo Fisher Scientific). Purified PCR DNA (10-40 ng), 1.5 µL of 5X buffer, 1 µL of Big dye, 0.5 µL of 10 µM primer were combined and made up to a total reaction of 10 µL with

H_2O. The cycling conditions for the reaction were as follows: 96°C for 5 min and 30 cycles of 96°C for 10 secs and 50°C for 5 secs with a final step at 60°C for 4 mins.

2.3.12.1 Sanger Sequencing reaction clean up

After sequencing, 80 µL of 75% isopropanol was added to each reaction. The tubes were vortexed gently and DNA precipitated by incubating in the dark for 10 mins. The tubes was centrifuged at 160 g for 20 mins. The supernatant was discarded, and the DNA pellet washed with 250 µL of 75% isopropanol with a gentle vortex. The tubes were centrifuged for 10 mins at 160 g and supernatant removed. The tubes were centrifuged for another 5 mins and the remaining isopropanol removed. The pellet was dried by heating at 95°C for 5 mins and sent for sequencing at SA Pathology (SA, Australia).

Gene	Use	Primers (5'-3')	Product size	Enzyme/ buffer	Annealing temperature
UPF1	Heteroduplex G1	F: ATTAGCAGTTGGGGCGGGAC R: GGAGGACTCGAGGCAAGTCTG	960	Herculase	60
UPF1	Heteroduplex: G2G3	F: GCCAGGTAGGTTTGCTAGAGAAG R: CAGATGCTCACCCAATGCCA	640	Herculase	60
UPF2	Heteroduplex	F: TGAATCTGAGTGATGGGTACG R: CTGCCTTACTTCATCTGAGCA	695	Herculase	60
UPF3A	Heteroduplex/ gDNA sequencing	F: GCTCTCGCGAGGTTTCGTCG R: TCGAAGTAGTCGTGTGCTGG	522	KAPA	65
UPF3A	cDNA sequencing	F: GCTCTCGCGAGGTTTCGT R: ATAAAGACTCAGGTCGGCGG	376	Buffer J (Cat# M07205J, Epicentre)	58
UPF3B	gDNA Sequencing	F: CACTTCAAGATGGCGGAATTG R: GGCGAAGAGATAGAAGGAGAAC	574	Buffer D (Cat# M07205D, Epicentre)	60
UPF3B	Heteroduplex	F: CTAGTGAGGCAAAATTTGGACG R: AACCTGAGAAAGAAAGGGGGC	758	Herculase	60
UPF3B	cDNA sequencing	F: ATTTGTTGGGGGCGGAGC R: AGCGGGATATTCCTGACCTTT	485	Buffer G (Cat# M07205G, Epicentre)	56
B-Actin	PCR	F-ATGGGTCAGAAGGATTCCTATGTG R- TGTTGAAGGTCTCAAACATGATCTGG	247	Taq	60
AHCYL1	WGS validation	F: AGCACTTGGATCGCATGAAA R: CAGCTGTGGTTAGGAAATAGTGTT	461	Taq	60
ANKRD13A	WGS validation	F: TCCAAGTCTCATGCACTTTCAGA R: GAGGGCAGTGGAATGATCCAA	513	Taq	60
C9orf3	WGS validation	F: TTTCATCGTGCATGTTGCTGT R: ATTCCACGTATACCACGGTCAC	498	KAPA	65
CSRNP3	WGS validation	F: GTAGTTCTATGGGCCCTGTCG R: CTCCGGGGTCATGGTGTAAG	530	Taq	60
DAO	WGS validation	F: CAAGAGAAGCTCAGGTGGCA R: TGCCAGACCCAGAGAAGGTA	487	Taq	60
DLGAP3	WGS validation	F: CCTGCGGGAGATGTGTGTAG R: CTCTGAGACTCCCAGCCTGA	524	Taq	60
GALNT6	WGS validation	F: TGGTTCTGGAGGATGAGGCT R: ACTTACTCAGGTGGTCGGGT	509	Taq	60
HEY1	WGS validation	F: GGCCTCACCCACGGAAC R: ACCTATCACAAAGTGTTGGCTT	573	Taq	60
HTRA1	WGS	F: GGCTGTTGCACAACCCATTA	495	Taq	60

	validation	R: GCTCCTTACAGAAGTTTTCCTGA			
NCOR2	WGS validation	F: GTGCCATCTCAGGGTTAGCAA R: TGGTCACGGGCGAACCT	490	Taq	60
NFATC4	WGS validation	F: AGGCTAAGCCAGCAGGAAAG R: AGTCTTCACCCTCCTCCCTC	557	KAPA	65
SLC6A11	WGS validation	F: CCTCGTCTGCATTAATGGGTG R: TAGCTGAGCACTTGCACATGAA	475	KAPA	65
STRN	WGS validation	F: GCTCTCATTGTGTGACTTGATGG R: AGCTTTTTCACAGCATCTGTATG	515	KAPA	65
TGM1	WGS validation	F: GGTGTCAATGCAGCTGGAGA R: GGAGGATGGGCTTGGGAAAA	577	KAPA	65
URB1	WGS validation	F: TATATTGTGTGGGTGTGGGTGGT R: ATTGCAGTTAGCAGAGGGTTAATAC	537	KAPA	65
ERAP2	WGS validation	F: CAGCAGTTTAAGTCACTGGTGC R: AGGATTAGTCGTAGAAGAGAAACCA	546	Taq	60
CLDN18	WGS validation	F: TCGGCTCAGCAAAGCTCTAAAA R: CTAAGACCATTTGGCCTCTCC	55	Taq	60
SEMA4A	WGS validation	F: CAAGAGTCCTCCCATCCTGC R: CCCAAACTAAGCCCCAGTCC	569	Taq	60
STIM1	WGS validation	F: GTAATGACTGTGGTTTCTATGGGC R: TTGTTGGTAAGAGTCCAGTGAC	500	Taq	60
TRAPPC8	WGS validation	F: CCTCCAATGCAGGTGTTCTT R: AAATATCCTGCATAATCTGGAACAG	581	KAPA	65
ZNF490	WGS validation	F: TCCACTCACTTTCGAATCCATGA R: TCACATATCTGCAATCCCGCA	513	Taq	60
KRT18	WGS validation	F: AAAGCCTGAGTCCTGTCCTTTCTC R: TGGAAATCCAGGTATGCCCAAC	590	KAPA	65
GZMH	WGS validation	F: AGAAGCCCTACTCTTCATGTCTT R: AAGAAAGTCCAGGTCAGGCAATG	510	Taq	60
NCOR2	WGS validation	F: TCATAGTGTTCCCGACTTGC R: CTAGAACTGGGCCCGTGTG	539	Taq	60

Table 2-2: Primers used in PCRs.

2.3.13 Quantitative Real Time PCR

Quantitative real time PCR (qPCR) was used to determine the expression of mRNA. 10 µL reactions were set up that contained 2 µL of cDNA, 1X Fast SYBR Green Real-Time PCR master mix (Cat# 4385612, Applied Biosystems), 0.5 µM of Forward primer, 0.5 µM of Reverse primer and made up to 10 µL with water. *Beta Actin* (*ACTB*) was used as a house keeping gene to normalise data. RT-qPCR primers are listed on **Table 2-3**. Each reaction was performed in triplicate. The plate was centrifuged for 30 secs at 234 g and ran on the CFX 384 real time system (Bio-Rad, CA, USA). The cycling conditions were a hot start at 95°C for 20 secs, and then 40 cycles of denaturation at 95°C for 3 secs and annealing at 60°C for 30 secs. This was followed by a melt curve set as follows: 95°C for 15 secs, 60°C for 1 min followed by a 0.3% continuous ramp from 60°C to 95°C. All primers used were validated by the ability to generate a single product by the analysis of a single melting temperature, and by validating the efficiency of the amplification. Efficiency of the amplification was determined by comparing the Cross Threshold (Ct) values acquired through serial dilutions of a control template. The Bio-Rad CFX manager software was used to calculate range percentage efficiency. Using this method all primers sets used were found to have an efficiency of 100 ± 10%. Results and normalised gene expression level were analysed using the Bio-Rad CFX manager software which uses the $2^{-\Delta\Delta CT}$ equation to calculate the normalised gene expression levels.

Gene	Primers	Product Size	Reference
UPF1	F-CATCATCCTGTCCTGTGTGC R-GACGCCATACCTTGCTCTG	103	
UPF2	F-GTTGTTACGGGCACTCTTCAT R-CCCCCTCAGCATGGAACAAA	130	(Tarpey et al. 2007)
UPF3A	F-CCATGGAGAGACTGGGAAG R-TGCCTTGCAGATCCTCCTG	158	
UPF3B	F- CTTCAGGGCAAAGAATAGAGAGA R- TTGACACAAGACTTACTCCTCTG	77	(Nguyen et al. 2012)
GADD45B	F-CAGCTACTGCGAAGAAAGC R-GTTTGTGGCAGCAACTCAAC	110	(Tarpey et al. 2007)
ATF4	F-TTAAGCCATGGCGCTTCTCA R-GGTCGAAGGGGGACATCAAG	168	
GAS5	F-CTTGCCTGGACCAGCTTAAT R-CAAGCCGACTCTCCATACCT	122	
SMG5	F-AAGGCAGCCAAAATGTACCAC R-GTAGGAGGCTTTGCAGATAC	124	(Tarpey et al. 2007)
PANK2	F-GGATTATTGGTCCAAGGG R-GTAATGATCACGGGATCTTC	107	(Tarpey et al. 2007)
c-JUN	F-GTGCCGAAAAAGGAAGCTGG R-CTGCGTTAGCATGCGTTGGC	175	
OCT3/4	F-GACAGGGGGAGGGGAGGAGCTAGG R-CTTCCCTCCAACCAGTTGCCCCAAAC	144	(Huang, HP et al. 2011)
NANOG	F-AGTCCCAAAGGCAAACAACCCACTTC R-TGCTGGAGGCTGAGGTATTTCTGTCTC	161	(Huang, HP et al. 2011)
SOX2	F-GGGAAATGGGAGGGGTGCAAAAGAGG R-TTGCGTGAGTGTGGATGGGATTGGTG	151	(Huang, HP et al. 2011)
B-ACTIN	F-ATGGGTCAGAAGGATTCCTATGTG R- TGTTGAAGGTCTCAAACATGATCTGG	247	(Tarpey et al. 2007)
NAT9	F: CGGGCATGCTTTCTAAACACA R: TGGGAGGCCTTAGGCAAGT	100	(Huang, L et al. 2011)
ATF3	F: GATGTCCTCTGCGCTGGAAT R: CTTCTTGTTTCGGCACTTTGC	151	(Huang, L et al. 2011)
SMG7	F: GCCCTCTTCGAGAGAAATTG R: CTGCTCGGTTTCATTGCTAA	143	(Huang, L et al. 2011)
HNMT NMD 1	F: GCTTGTTTTCTGACCACGGG R: ACAGCAATTCTGGTATCTTCCTA	134	Wilkinson, unpublished

HIST1H2AC	F: TTAGGCCGCTGGTTTTGGTG R: GCGTAGTTGCCTTTACGGAG	159	Wilkinson, unpublished
AUTS2	F: GCATTTCAGCCGAAGTTGACA R: AAATCTGCCAGGCGATGTGA	94	Wilkinson, unpublished

Table 2-3: Primers used for RT-qPCR analysis.

2.3.14 Taqman RT-qPCR

Pre-designed Taqman expression assays (Primer-probes) for RT-qPCR were purchased from Thermo Fisher Scientific (**Table 2-4**). At the 5'end they were conjugated to the FAM^{TM} label and on the 3'end a miner groove binder (MGB) and a non-fluorescence quencher (NFQ). Taqman amplifications were all performed as triplicate reactions using the housekeeper gene *ACTB* with a VIC® dye label. All 20 µL reactions contained: 2 µL cDNA, 1X Taqman gene expression master mix, 1X Taqman gene expression assay, 1X Taqman endogenous control and made up with water. PCR reactions were run on the CFX real time system (Bio-Rad) and data collected and analysed using the Bio-Rad CFX manager. All reactions were ran using the following parameters: 50°C for 2 mins, 95°C for 10 mins and 40 cycles of 95°C for 15 secs, 60°C for 60 secs.

Taqman Probes	**Cat #**
CDKN1A	Hs00355782_m1
CDKN2A	Hs00233365_m1
ATF3	Hs00231069_m1
Y14/RBM8A	Hs04234933_m1
ANXA1	Hs00167549_m1
OTX2	Hs0022228_m1
MESDC2	Hs00991868_m1
CLDN6	Hs04189215_g1
ATF6	Hs00232586_m1
ERN2	Hs01086607_m1
DDIT3	Hs00358796_g1
EIF2AK3	Hs00984003_m1
ROBO2	Hs00326067_m1
ARX	Hs00417686_m1
NETO1	Hs00371151_m1
LRRC4B	Hs00297475_s1
TGFB1	Hs00998133_m1
HAS2	Hs00193435_m1
ACTB	4326315E

Table 2-4: Primers used for Taqman RT-qPCR

2.3.15 Heteroduplex analysis of CRISPR/Cas9 Genome Edited cells

Heteroduplex analysis was performed by purifying PCR products derived from the amplification of genomic regions flanking gRNAs targeted sites in a pool of heterogeneous CRISPR edited cells. Reactions contained 18 µL of purified PCR product (20 ng/µL) and 2 µL of 10X Taq PCR buffer. Reaction products were denatured and slowly annealed with the cycling conditions in **Table 2-5**. The heteroduplex products were separated by polyacrylamide electrophoresis using either an 8% (Cat# EC6215, Invitrogen), 10% (Cat# EC62752, Invitrogen) or 10% (home-made) polyacrylamide gel with 1X TBE buffer for 30 mins.

Cycle number	Condition
1	95°C for 3 mins
2	95-85°C, -2°C s^{-1}
3	85°C for 1 min
4	85-75°C, -0.3°C s^{-1}
5	75°C for 1 min
6	75-65°C, -0.3°C s^{-1}
7	65°C for 1 min
8	65-55°C, -0.3°C s^{-1}
9	55°C for 1 min
10	55-45°C, -0.3°C s^{-1}
11	45°C for 1 min
12	45-35°C, -0.3°C s^{-1}
13	35°C for 1 min
14	35-25°C, -0.3°C s^{-1}
15	25°C for 1 min
16	25-4°C, -0.3°C s^{-1}
17	4°C hold

Table 2-5: Cycling conditions for heteroduplex analysis.

2.3.16 10% polyacrylamide gel

6 mL of 5X TBE buffer and 3.75 mL of 40% acrylamide/bisacrylamide (Cat# A9926, Sigma-Aldrich) were mixed and made up to 15 mL with water. 10 µL of TEMED (Cat# 17919, Thermo Fisher Scientific) was added to the above reaction and mixed. 37.5 µL of 20% APS (Cat# A3678-259, Sigma-Aldrich) was immediately added and mixed by inverting the tube. The acrylamide solution was poured into gel plates and allowed to polymerise for 30 mins before use.

2.3.17 Western Blot

2.3.17.1 Protein extraction

100 µL of RIPA buffer (65.3mM Tris, 150 mM NaCI, 1% (v/v) Nonidet P-40) was supplemented with 4% 25X Protease inhibitor cocktail (Cat # P8340, Sigma), 0.5% of Na_3VO_4 (200 mM), 0.5% of NaF (200 mM) and 0.5% of phenylmethylsulfonyl (PMSF) (200 mM)). The supplemented RIPA buffer was added to cells and pipetted to create uniformly resuspend cells. Cells were passed through a 21 G needle 10 times to lyse the cells. The lysate was centrifuged at 160 g for 30 mins at 4°C to clear debris. The supernatant was stored at -80°C.

2.3.17.2 Protein quantification

Aliquots of protein lysate samples were diluted 1:10 or 1:20 prior to Bradford quantification. 10 µL of the diluted samples were added to a 96 clear flat bottom well in triplicates along with standard samples of bovine serum albumin (BSA) standards (0, 0.2, 0.4, 0.6, 0.8 and 1.0 mg/mL) (Cat# A7030, Sigma-Aldrich). 200µL of 1X Bradford reagent (Cat# 500-0006, Biorad) was added to each sample and light absorbance at 620 nm was analysed by a photospectroscopic plate reader (FLUOstar Omega, LabTech). The protein concentration of the samples were quantified by using the standard curve method.

2.3.17.3 Separation of protein on an SDS-PAGE gel

40 µg of protein was prepared by adding, 1X loading buffer (Cat# NP0007, Invitrogen), 10% (v/v) DTT and made up to a total volume of 40 µL with water. Samples were denatured at 95°C for 5 mins and returned to ice. The XCell SureLock Mini-Cell system (Invitrogen, Thermo Fisher Scientific) was used to perform SDS-PAGE as per manufacturer's instructions. NuPAGE^{TM} 4-12% Bis-Tris gels (Cat # NP0322, Invitrogen) were used, and running buffer (Cat# NP0001, Invitrogen) prepared by adding 500 µL of anti-oxidant (Cat# NP0005, Invitrogen) to the inner chamber. The outer chamber was half filled with running buffer only. The denatured protein samples were separated by adding 10 µL of the sample into the gel wells and electrophoresis conducted at 150V for 150 mins.

2.3.17.4 Western transfer of protein to cellulose membrane

Western transfer was conducted using Xcell Blot II Module apparatus (Invitrogen, Thermo Fisher Scientific) as per manufacturer's instructions. Spongers, filter papers and the cellulose membrane were soaked in cold transfer buffer (1X Towbin (0.025 M Tris and 0.192 M Glycine), supplemented with 20% methanol (v/v)). Gels were soaked in transfer buffer for 10 mins. The transfer chamber was loaded with pre-soaked sponges, filter papers, gel, cellulose membrane, filter paper and spongers. The chamber was inserted into the tank and clamped. The inner chamber was filled with cold transfer buffer. Transfer was performed at 30V for 90 mins.

2.3.17.5 Immunoblotting

After protein transfer onto the cellulose membrane, ponceau staining was performed by incubating the cellulose membrane with the ponceau stain for 5 mins, and then rinsed with water for 5 mins, to determine successful transfer. The cellulose membrane was then washed with PBST (PBS with 0.2% Tween 20 (Cat# 9416, Sigma)) to remove the ponceau stain. The cellulose membrane was blocked with 5% skim milk (w/v), 5% BSA (w/v) and 5% Goat serum (GS) (Cat# 16210064, Gibco, Thermo Fisher Scientific) / Horse serum (HS) (Cat# 16050130, Gibco, Thermo Fisher Scientific) (v/v) depending on the secondary antibody used for 1 hr or overnight. The blocking solution was removed and replaced with a primary antibody solution in 1% blocking solution. Primary antibody concentration are listed in **Table 2-6**. Primary antibodies were incubated overnight in a humidified atmosphere at 4°C. The membrane was washed with PBS-T three times to remove the primary antibody and incubated with the secondary antibody (1:4000) in 1% blocking solution (**Table 2-6**). The membrane was washed with PBS-T three times to remove the secondary antibody. Bound secondary antibodies were detected by incubating the membrane with the Clarify™ western ECL substrate (Cat# 170-5061, Bio-Rad) as per manufacture's instruction. After 5 mins incubation, chemiluminescence was captured using the ChemiDoc™ MP Imaging system, Bio-Rad.

Protein	Primary antibody and dilution	Secondary antibody	Protein size
UPF1	RENT 1, Rabbit polyclonal Bethyl Laboratories A301-902 1:2000	HRP-conjugated goat anti-rabbit IgG DAKO P0448	150
UPF2	RENT2 (C-18), Goat polyclonal Santa Cruz Biotechnology SC-20227 1:500	HRP-conjugated donkey anti-goat/sheep IgG Millipore AB324P	148
UPF3A	UPF3A, Rabbit polyclonal Sigma-Aldrich HPA018325 1:1000	HRP-conjugated goat anti-rabbit IgG DAKO P0448	55
UPF3B	UPF3B, Rabbit polyconal Sigma-Aldrich HPA001800 1:250	HRP-conjugated goat anti-rabbit IgG DAKO P0448	58
β-ACTIN	B-ACTIN, mouse monoclonal Sigma-Aldrich A2228 1:1000	HRP-conjugated goat anti-mouse IgG DAKO P00447	42
Y14	Y14, Mouse monoclonal Santa Cruz Biotechnology SC-32312 1:400	HRP-conjugated goat anti-mouse IgG DAKO P00447	20
MAGOH	MAGOH, Mouse monoclonal Santa Cruz Biotechnology SC-56724	HRP-conjugated goat anti-mouse IgG DAKO P00447	17

Table 2-6: Antibodies used for western blot analysis.

2.3.18 Immunofluorescence analysis

Cells were fixed in 4% paraformaldehyde (PFA) (w/v) (Cat# P6148, Sigma-Aldrich) at RT for 15 mins. The fixed cells were washed three times with PBS and permeabilised in 0.5% Triton X-1000 in PBST for 2 mins for hESCs and 5 mins for neural rosettes. The cells were washed once with PBST and blocked in 5% HS in PBST for 1 hr at RT. The permeabilised cells were incubated with the primary antibody diluted (**Table 2-7**) in 0.5% HS-PBST overnight at 4°C. After the overnight incubation, cells were rinsed three times with PBST for 5 mins and incubated with a secondary antibody in 0.5% HS-PBST (1:1000 dilution) (**Table 2-7**) for 1 hr at RT. The secondary antibody was removed and cells washed three times with PBST and lastly rinsed with 1X PBS. Cells were mounted using ProLong^{TM} Diamond Antifade mountant with DAPI (Cat# P36962, Invitrogen, Thermo Fisher Scientific).

Protein	Primary antibody and dilutions	Secondary antibody
NANOG	NANOG Polyclonal Goat R&D Systems AF1997 1:500	Alexa Fluor Donkey anti-goat Invitrogen A21432
OCT3/4	OCT3/4 Monoclonal mouse Santa Cruz Biotechnology SC-5279 1:200	Alexa Fluor Donkey anti-mouse Invitrogen A21202
SOX2	SOX2 Polyclonal rabbit EMD Millipore AB5603 1:200	Alexa Fluor Donkey anti-rabbit Invitrogen A31573
PAX6	PAX6 Polyclonal rabbit BioLegend 901301 1:300	Alexa Fluor Donkey anti-rabbit Invitrogen A31573
N-CADHERIN	N-CADHERIN Polyclonal Mouse BD Transduction Laboratories 610921 1:300	Alexa Fluor Donkey anti-mouse Invitrogen A21202
NESTIN	NESTIN monoclonal Mouse Millipore MAB5326 1:300	Alexa Flour Donkey anti-mouse Invitrogen A21202

Table 2-7: Antibodies used in immunofluorescence analysis

2.4 Whole genome analysis of gene-edited clones

Whole genome sequence (WGS) analysis was performed by Dr Mark Corbett and Dr Atma Ivancevic (Neurogenetics Laboratory, University of Adelaide). DNA Samples were sequenced with 150 bp paired reads by Illumina HiSq X-10 at the Kinghorn Centre for Clinical Genomics (NSW, Australia). Published H1 WGS data were downloaded from the European Nucleotide Archive (H1 Published: SRR2070629) (Yazdi, P. G. et al. 2015). For SNV analysis Fastq formatted sequence files were mapped to the University of California Santa Cruz (UCSC) hg19 build of the human genome using Burrow-Wheeler Aligner (BWA) (Li & Durbin 2009). Duplicate reads were marked with Sambamba (Tarasov et al. 2015) and single nucleotide variants (SNVs) and indels were called using the genome analysis toolkit (GATK) V3.8 haplotype caller (Li & Durbin 2009; Van der Auwera et al. 2013). Samples were jointly genotyped with 16 additional unrelated WGS of DNA extracted from human blood using GATK to generate variant call format (VCF) files with non-reference calls for each samples. Bcftools were used to apply the following variant exclusion filters, where at least all variants had at least 10 reads per total (DP<10), a quarter of reads had alternate alleles (QD<20, FS>60.0, MQ<30.0, MQRankSum<-12.5, ReadPosRanlSum<-8.0), and reach the minimum qualities recommended by GATK (MIN(AD[1]/FMT/DP)<0.25) (Van der Auwera 2012) and the GATK variant quality score recalibration (VQSR) metrics to retain PASS and VQSRTrancehlINDEL99.00to99.90 variants. Comparisons between different samples were made with the *vcf-compare* command from the *bcftools* package. Variants loaded at potential off target sites were identified using *bfctools* view filtered with a custom bed file that had a 20 bp window flanking each predicted off target site. A list of off-targets was generated using the Cas-OFFinder program (www.rgenome.net/cas-offinder) (Bae, Park & Kim 2014) and the CRISPR design program (crispr.mit.edu). Relatedness between cell lines was determined using the *vcftools* relatedness2 algorithm. Variants

were annotated with ANNOVAR (Van der Auwera et al. 2013) and alignments at variants were viewed using the Integrative Genomics Viewer (IGV).

DELLY was used to call and genotype structural variants in all samples. Output vcf files were merged and re-genotyped using multi-sample cohort containing 160 samples as reference. *Bcftools* was used to merge the final vcfs into a multi-sample vcf file containing the eight samples of interest. *BedTools* was used to overlap each sample with the corresponding off-target regions, allowing for a 200 bp window around the predicted off-target regions. Resulting structural variants were de-duplicated (since the provided off-target regions have overlapping records) and filtered to remove low quality variants (set FILTER=PASS). Variant candidates were then extracted from the original vcf and run through **SVScore** to predict possible pathogenicity and look for gene overlaps.

2.5 RNA-Sequencing

RNA-Sequencing analysis was performed by Dr Stephen Pederson (Bioinformatics Hub, University of Adelaide) and Urwah Nawaz (Neurogenetics Laboratory, University of Adelaide). RNA-Seq samples were sequenced with 75 bp paired-end reads by Illumina TruSeq (SAHMRI Genomics Facility, SA, Australia). The raw fastq reads were checked for quality using FASTQC (Andrews 2010). Adapters and low quality bases were trimmed using TrimGalore (http://www.bioinformatics.babraham.ac.uk/projects/trim_galore/). Salmon v0.8.2 (Patro, Duggal & Kingsford 2015) was used for transcript quantification and the resulting transcript abundance were imported to R using tximport (Love, Soneson & Robinson 2017), converting transcript counts to gene counts. Gene level quantification was estimated by summing isoform counts using tximport. To filter out lowly expressed

genes, if 3 or less samples had a counts per million (CPM) < 1 for a particular gene, that gene was removed from the counts matrix. Differential expression analysis was carried out using DESeq2 (Love, Huber & Anders 2014). Genes were considered differentially expressed if they had an adjusted p-value of <0.05 and a log 2 fold change. The effects of loss of *UPF3A* and *UPF3B* were detected in hESCs and NSCs using the Wald test in a model with condition and collection or condition only respectively. Pairwise comparisons using DESeq2 was performed to compare gene expression between hESCs and NSCs in the main 3 genotypes, controls, *UPF3A* KO and *UPF3B* KO. Genes with an adjusted p-value of <0.05 and a log 2 fold change were used for gene ontology analysis using ClusterProfiler (Yu et al. 2012). All volcano plots were made using the EnhancedVolcano package for all DE analyses (Blighe, 2018). Heatmaps were generated by normalising the raw gene counts to log transformed CPM using the pheatmap package (pheatmap: Pretty Heatmaps v.1.0.12, 2019).

2.6 Statistical analysis

Statistical significance was calculated using a two-tailed students T test with significance set at $p<0.05$. Error bars represent standard deviation.

Chapter 3 - Generating *UPF3A* and *UPF3B* null hESC clones using CRISPR/Cas9 genome editing technology

3.1 Introduction

Complete loss of function mutations in *UPF3B* cause a spectrum of NDDs including autism, ID, ADHD and childhood onset schizophrenia (Addington et al. 2011; Lynch et al. 2012; Szyszka et al. 2012; Tarpey et al. 2007; Xu et al. 2013). *UPF3B*, is the only NMD factor gene that has a paralog *UPF3A*. *UPF3B* is located on the X chromosome and loss of function mutations of this gene affects males while female carriers are unaffected (Laumonnier et al. 2010). *UPF3A* is autosomal and located on chromosome 13 (Linder, Fischer & Gehring 2015). Heterozygous copy number losses of *UPF3A* is linked to NDDs (neural tube defects) (Luo et al. 2000). Intriguingly loss of function mutations in *UPF3B* lead to stabilisation of UPF3A protein (Chan et al. 2009; Jolly et al. 2013; Nguyen et al. 2012).

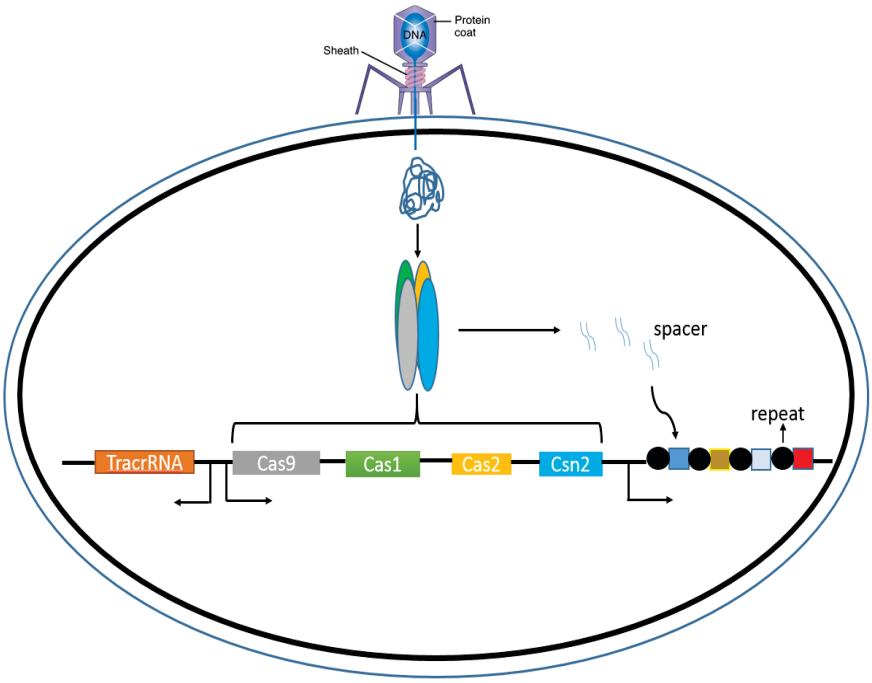
UPF3B is well established as an NMD activator and its role in neurodevelopment has mainly been derived from patient LCLs and other species such as mice. Initially UPF3A was shown to have weak NMD promoting activity when tethered artificially to an NMD reporter transgene (Kunz et al. 2006; Lykke-Andersen, Shu & Steitz 2000) but recent gain and loss function approaches aligned UPF3A function more as an NMD inhibitor upon assessment of changes in expression of endogenous NMD targeted mRNA (Shum et al. 2016). However, even this study provided evidence that *UPF3A* could still act on a minority of endogenous NMD transcripts as an NMD promoter. Whilst the role of UPF3A as an NMD factor itself is still unclear, its role in neurodevelopment is even less well known, having never been formally studied. Therefore there is an unmet need to study the role of both *UPF3A* and *UPF3B* in models of brain development to better understand the involvement of these factors in NMD in general and in human NDDs. The other reason for investigating the role of *UPF3A* parallel to *UPF3B* is because these two genes are paralogs that have been postulated to act redundantly. To investigate the role

of *UPF3A* and *UPF3B* in NDDs a cell model system that replicates neurodevelopment in patients with loss of function mutations in *UPF3A* and *UPF3B* was generated. The CRISPR/Cas9 genome editing technology and H1 hESCs (derived from a male embryo) were used to generate *UPF3A* and *UPF3B* KO hESCs. A male cell line was used, as deletions in *UPF3B* cause X-linked inherited NDDs. Directed differentiation of *UPF3A* and *UPF3B* KO hESCs into neural stem cells provides a model system of human brain development.

3.1.1 CRISPR/Cas9 in bacterial immunity

CRISPR/Cas9 is an adaptive immune system that protects bacteria and archaea from bacteriophages and plasmids (Hryhorowicz et al. 2016). It was first discovered in *E.coli* by Ishino in 1987 (Ishino et al. 1987). The CRISPR/Cas9 system develops 'immuno-memory' by inserting foreign viral DNA derived sequences called 'spacers' between repeat sequences within the bacterial *CRISPR* locus during the first viral exposure (Heler et al. 2015). The *CRISPR* locus contains all information about the cells (and its predecessor cells) encounters with foreign DNA (i.e. from phage's or plasmids) in the past (Rath et al. 2015). There are three known CRISPR systems (I to III) (Ran et al. 2013). The type II CRISPR-Cas system used for genome editing is from the bacteria *Streptococcus pyogenes* (Sternberg et al. 2016). The *type II CRISPR* locus contains a *trans-activating CRISPR* RNA (*tracrRNA*), the CRISPR associated (Cas) genes (*Cas9*, *Cas1*, *Cas2* and *Csn2*) (Cas complex) and a *long pre CRISPR* RNA (*pre-crRNA*) transcribed from the spacer repeat locus (Mali, Esvelt & Church 2013). A detailed mechanism of the CRISPR/Cas9 type II system is described in **Figure 3.1**.

a) Acquisition of new spacers



b) Target recognition and destruction

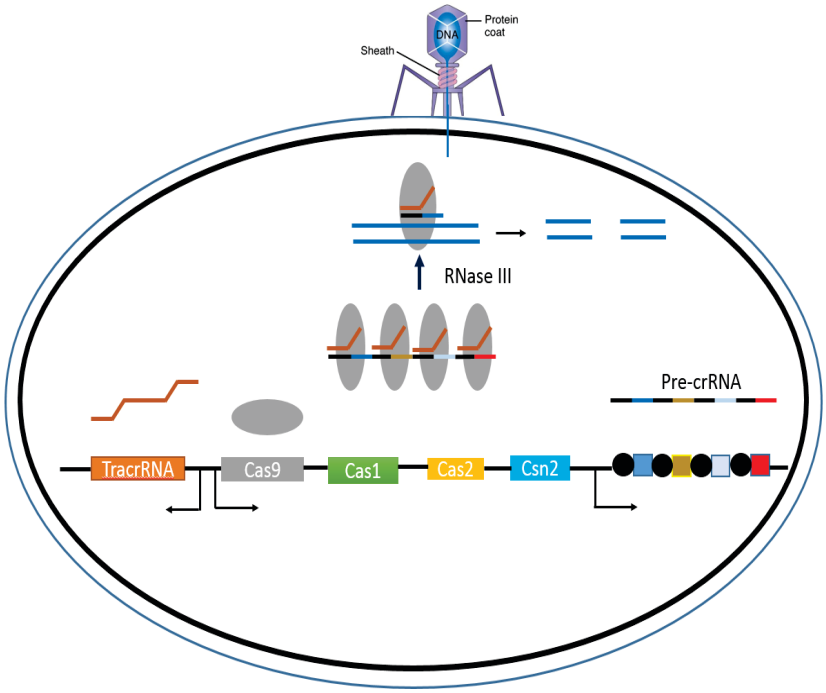


Figure 3.1: CRISPR/Cas9 immunity in bacteria.

a) In the CRISPR/Cas9 system, the Cas complex specifies the spacer sequence in the foreign DNA sequence next to a protospacer adjacent motif (PAM) site) cleaves the spacer sequence and inserts it between the repeats in the *CRISPR* locus (Heler et al. 2015). At this stage the DNA fingerprint of the infection is stored (Mali, Esvelt & Church 2013). b) In subsequent infections the *CRISPR* locus is transcribed to produce the CRISPR/Cas9 unit that guides the degradation of homologous sequences on the foreign DNA (Terns &

Terns 2011). The generation of the CRISPR/Cas9 unit occurs by *tracrRNA* binding to the long *pre crRNA* on the spacer sequence and the Cas9 endonuclease forming a stable ribonucleoprotein. RNase III cleaves the *tracrRNA: crRNA* (dsRNA) and the 5'end of each of the spacers is removed generating a CRISPR/Cas9 functional unit. The CRISPR/Cas9 unit hybridises with the complementary sequence upstream of the PAM sequence on the foreign DNA sequence and Cas9 cuts the DNA creating double stranded breaks (DSBs) (Mali, Esvelt & Church 2013) using its RuvC and HNH nuclease domains (Sander & Joung 2014). Figure modified from (Mali, Esvelt & Church 2013).

3.1.2 Exploiting CRISPR/Cas9 for genome editing

Genome editing technologies have been developed that both harnesses the CRISPR/Cas9 system 'genome-homing DSBs' mechanism and leverages the cells innate mechanisms that repairs DNA DSBs (e.g. that occur during exposure to damaging DNA radiation and chemicals, DNA replication and repair (Cannan & Pederson 2016) and meiosis I (de Massy 2013)). DNA DSBs threaten genome integrity and failure to detect and repair them are detrimental to the cell (Ceccaldi, Rondinelli & D'Andrea 2016). DNA DSB are repaired by either the non-homologous end joining (NJEJ) or homology directed (HD) repair, the latter using a repair donor DNA template. The HD repair pathway occurs during the S and G2 phase of the cell cycle where the two sister chromatids are in close proximity- a wildtype chromatid providing a donor template for repair of a homologous site in the damaged chromatid (Lieber 2010). In the absence of the donor template the NHEJ pathway is responsible for repairing the DSBs. Unlike the HD repair this pathway can occur in any phase of the cell cycle (Cannan & Pederson 2016) but is dominant in the GO/1 and G2 phase (Karanam et al. 2012; Lieber et al. 2003). Because the NHEJ pathway is error prone and generates insertion or deletions (indels) during the repair mechanism this often leads to frameshift mutations and subsequently PTCs when this occurs within protein-coding regions of the genome.

In this chapter, the CRISPR/Cas9 genome editing technology was used to create *UPF3A* and *UPF3B* KO hESCs. The methodology employed was based on a protocol by Ran et al., 2013 and leverages the NHEJ repair pathway. The protocol uses the PX459 V2 plasmid which contains three expression cassettes in *cis* that facilitate expression of the Cas9 protein, a specific gRNA and the *PAC* gene encoding puromycin N-acetyltransferase that confers resistance to the antibiotic puromycin (for antibiotic selection). The specific gRNAs cloned into PX459 V2 were first validated for their ability to facilitate editing at their targeted genetic loci in HEK293T cells. Next, these validated gRNAs were used to generate gene-edited hESC clones. Despite following a published method, problems with attaining clonal hESCs were encountered. This chapter outlines the validation of CRISPR/Cas9 gRNAs in HEK293T cells and the optimisation steps taken to generate hESC clones. The optimised method proved to be robust, reproducible and did not have any overt effects on pluripotency or karyotype of the gene-edited hESCs.

3.2 Results

3.2.1 CRISPR/Cas9 design

gRNAs that target the 5' coding regions (exon one or two) of *UPF1*, *UPF2*, *UPF3A* and *UPF3B* were designed using an online gRNA CRISPR design tool (http://tools.genome-engineering.org). 5' coding regions were targeted as frameshift events impacting the 5' end of the gene are most likely to cause a PTC and lead to complete loss of functional protein (Pelley 2007). The PX459 V2 plasmid is depicted in **Figure 3.2**. Three unique gRNAs were selected for each gene with the view that clonal lines derived from the use of several unique gRNAs targeted to the same gene will cater for any possible off-targets generated by any single gRNA. The effectiveness of the gRNAs in generating DSBs (indels) at the desired loci was performed in HEK293T cells before use in hESCs.

3.2.2 Validation of CRISPR guides in HEK293T cells

To determine the targeting efficiency of gRNAs, validation of the gRNAs were performed in HEK293T cells as these cells are human and easy to grow and transfect. Cells were transiently transfected with an empty PX459 V2 plasmid (PX2, negative control) or with PX459 V2 plasmid encoding the gRNAs that target either *UPF1*, *UPF2*, *UPF3A* or *UPF3B*. Non-transfected cells were included as an additional negative control. The expected outcome is that control experiments will have wildtype loci, whilst experiments that include gene specific gRNA will cause editing events at the respective loci. A key feature in this experiment is that transfected cells are analysed as a pool of cells (rather than clones) and as such, editing events at a loci will be highly variable (a range of indels) between each cell in the pool of cells analysed. This heterogeneity of editing events can however be resolved and would strongly support that the gRNA is effective at homing Cas9 to the desired loci and to induce DNA cleavage. The homing of

Cas9 endonuclease by the individual gRNAs was determined by three assays; 1. heteroduplex formation analysis, 2. Sanger sequencing, and 3. ability to reduce the translation of full-length protein from the target site in the heterogeneous edited cells.

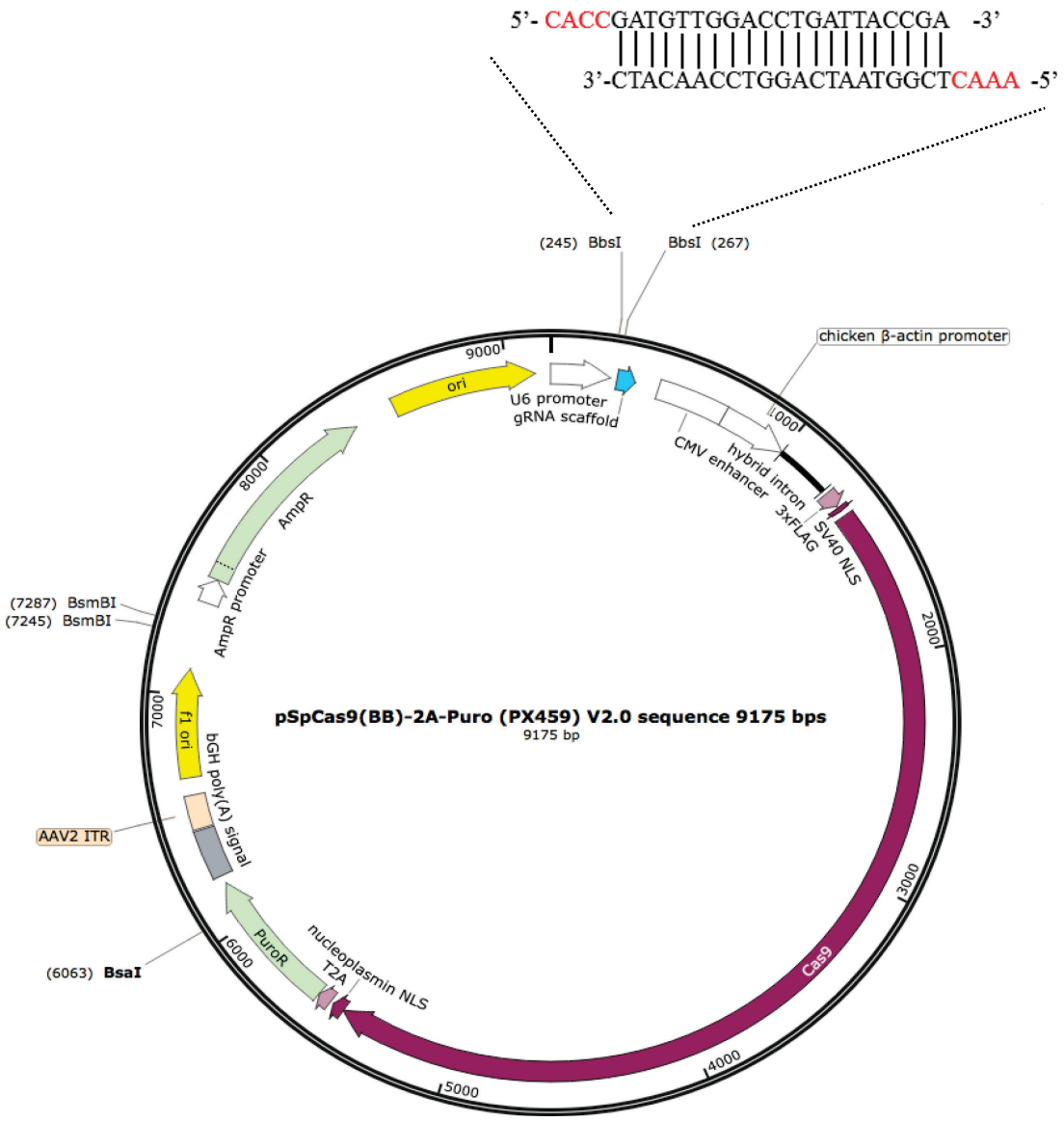


Figure 3.2: CRISPR/Cas9 genome editing PX459 V2 plasmid.

The PX459 V2 plasmid encodes the Cas9 endonuclease, a puromycin resistance cassette and the gRNA. The plasmid has a *Bbs1* site for insertion of the gRNA as shown.

3.2.2.1 Heteroduplex analysis

Heterogeneous CRISPR/Cas9 editing events at a gene loci in a pool of cells can be identified by heteroduplex assays. A heteroduplex is a dsDNA molecule with partial mismatched complementarity. Compared to homoduplex dsDNA molecules, heteroduplexes migrate slower during gel electrophoresis because of kinks and loops introduced by the mismatched pairs (Hestand et al. 2016). The sequence of PCR products derived from wildtype alleles are identical, and as such upon denaturing to ssDNA and reannealing, homoduplex products are formed. In contrast, PCR products derived from amplification of edited alleles from the transiently transfected pools of CRISPR/Cas9 edited cells should contain heterogeneous sequences caused by NHEJ-mediated indels and as such after denaturing and reannealing will form heteroduplexes. Thus, the heteroduplex formation of PCR products is diagnostic of CRISPR/Cas9 editing in the analysis of pooled cells (**Figure 3.3**).

gDNA was extracted from the above described transfection experiment (section 3.2.2) and Herculase II or KAPA HiFi, high fidelity PCR enzymes were used to generate PCR products across relevant loci. PCRs products from *UPF3A* loci were markedly smaller in size indicative of frequent large deletions within this locus (**Figure 3.4a**). PCR products across *UPF1*, *UPF2* and *UPF3B* loci generated single products by gel electrophoresis and were further analysed by the heteroduplex formation assay. PCR products were purified, denatured to produce ssDNA species and slowly annealed. The annealed products were separated using polyacrylamide gel electrophoresis which permits high resolution size discrimination. Annealed PCR products derived from CRISPR/Cas9 edited loci had extra bands on the gel compared to controls (**Figure 3.4b, c, d**). This revealed that the PCR template gDNA was heterogeneous, an outcome that reflects

CRISPR-Ca9 editing. The fastest migrating bands are homoduplexes that migrate easily and quicker and represent wild type loci products.

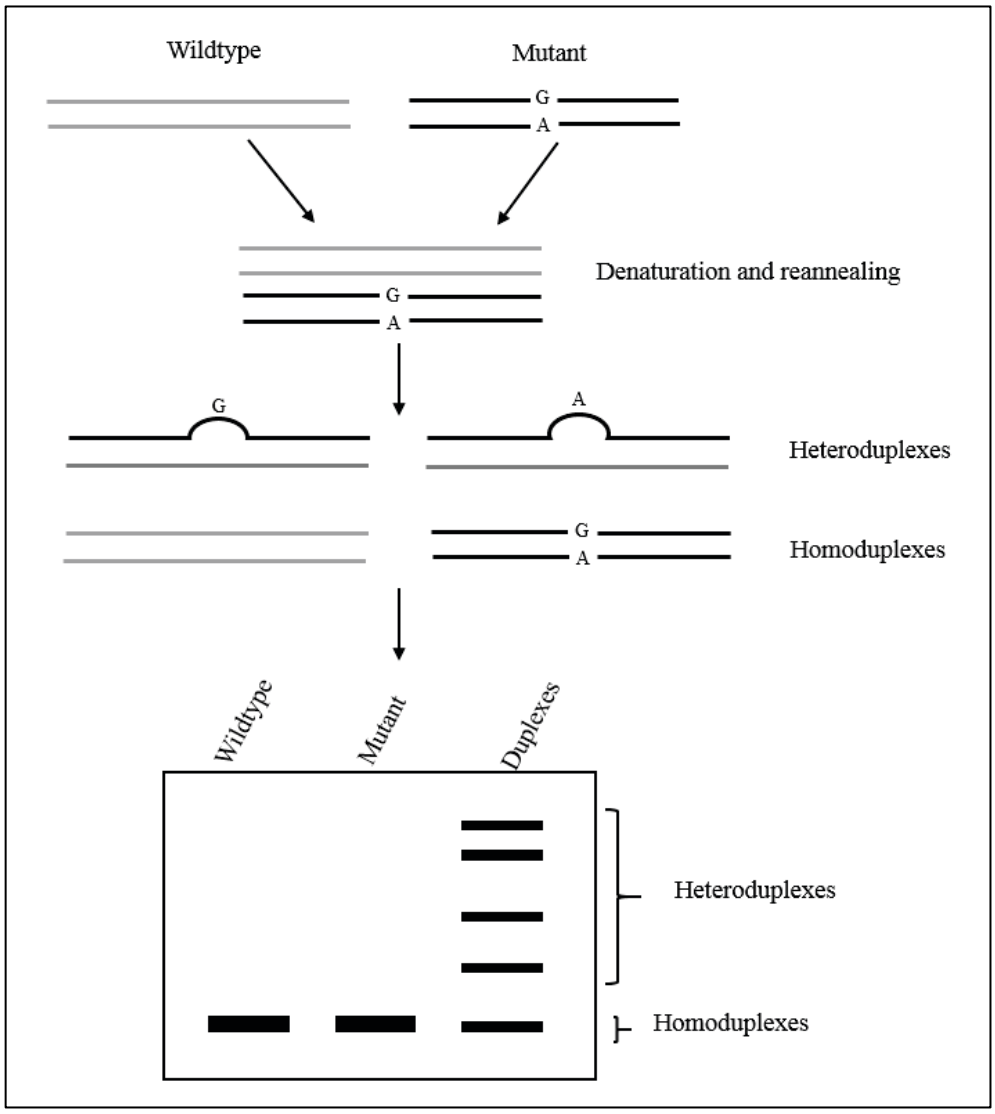


Figure 3.3: Analysis of heteroduplex formation in CRISPR edited heterogeneous cells.

Heteroduplex analysis on a pool of cells that contain heterogeneous genotypes. After denaturation and reannealing, alleles with either the wildtype or mutated genotypes will anneal randomly despite the mismatches. This will result in dsDNA that contain either the wildtype alleles, mutated alleles (homoduplexes) or either dsDNA strands that have mismatched sequences (heteroduplexes). The difference in homoduplexes and heteroduplexes is resolved by running the denatured and reannealed DNA on a polyacrylamide gel as these travel at different rates during electrophoresis. Figure modified from (Menounos & Patrinos 2010).

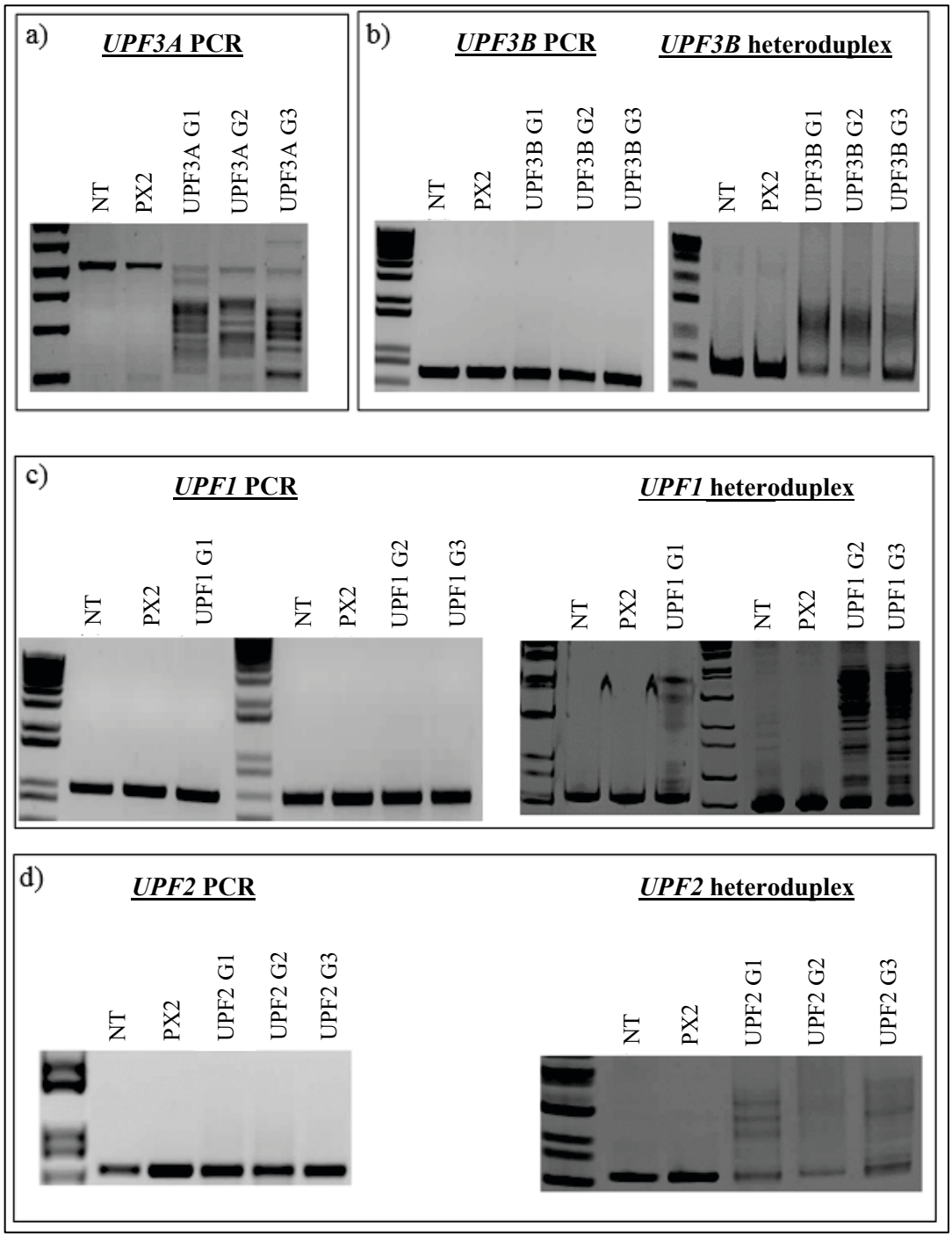


Figure 3.4: Efficient CRISPR/Cas9 targeting by gRNAs revealed by PCR and heteroduplex assay.

gRNAs targeted at (a) *UPF3A*, (b) *UPF3B*, (c) *UPF1* and (d) *UPF2* were tested for their ability to introduce indels in HEK293T cells. PX459 V2 empty vector (PX2) or one of the three gene specific gRNAs (G1, G2, G3) were transfected into HEK293T and gDNA extracted for use in PCR and subsequent heteroduplex assay. PCR and duplex products were separated on agarose and polyacrylamide gel respectively. NT: non-transfected. NB. Smaller size PCR products in (a) indicative of larger deletions.

3.2.2.2 Sequence traces of gDNA from CRISPR/Cas9 edited cells

A second measure used to determine if the gRNAs were effective at targeting/homing the Cas9 to the specific genomic location is use of Sanger sequencing of the PCR product generated above from the gDNA of transfected cells. Because each cell has a potentially unique array of indels at the targeted loci, effective Cas9 homing is visualised as a loss of a single sequence trace downstream of the gRNA sequence. The sequence chromatograms of the respective gene loci's (*UPF1, UPF2. UPF3A, UPF3B*) were clean in cells that were non-transfected (NT) or transfected with an empty PX459 V2 plasmid (PX2). In contrast, the sequence chromatograms derived from PCR products in HEK293T cell population targeted with PX459 V2 encoding loci specific gRNA had a mixed signal downstream of the respective gRNA sites (**Figure 3.5**) (sequence traces for *UPF2*, *UPF3A* and *UPF3B* not shown). This data aligns with effective homing of the Cas9 endonuclease by the gRNAs.

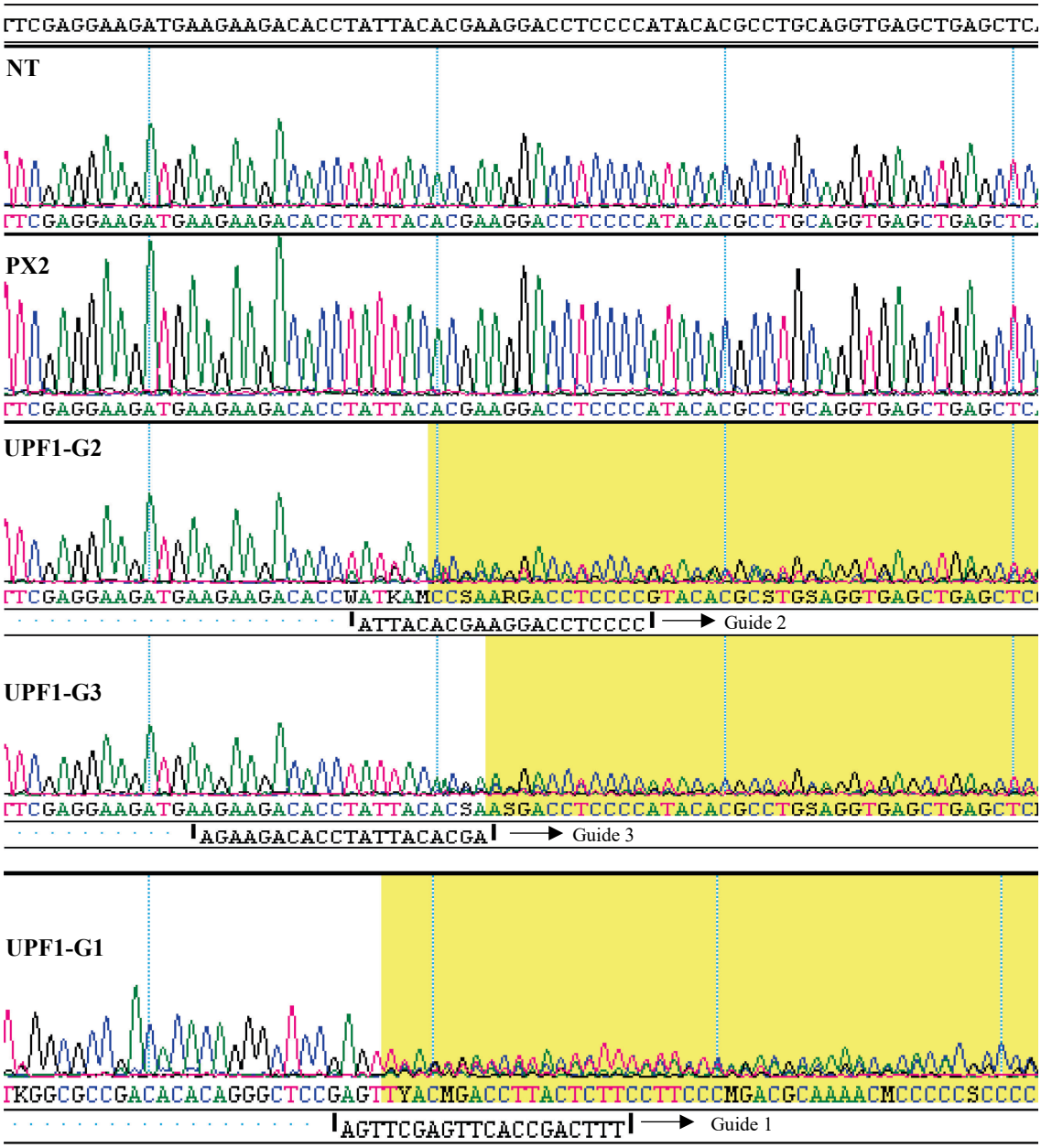


Figure 3.5: Aberrant sequence traces of *UPF1* edited heterogeneous cells.

gDNA was extracted from wildtype cells (NT), and cells either treated with an empty vector (PX2) or 3 different guides targeting *UPF1* loci in HEK293T cells. A PCR was performed flanking the targeted region and the PCR product sequenced using Sanger sequencing. NT: Non-transfected.

3.2.2.3 Protein analysis

The final and ultimate assessment of the gRNA design and functionality is that the targeted loci will not be able to produce functional protein due to creation of an early frameshift mutation. HEK293T cells were transfected as described above (section 3.2.2). Protein was isolated and subjected to western blot analysis. Because a range of indels were created including in-frame, a targeted loci would result in reduced rather than absent levels of the respective protein compared to controls (non-transfected and PX2). CRISPR/Cas9 targeted (G1, G2, G3) cells did indeed show obvious reduction in their protein levels in UPF1 and UPF2 and to a lesser extend UPF3B compared to controls (**Figure 3.6**). UPF3A protein is lowly detected in HEK293T cells, therefore no western blot was performed on the *UPF3A* edited cells.

In aggregate the three independent lines of validation including heteroduplex formation, Sanger sequencing and analysis of protein levels all aligned with the gRNAs being able to correctly target Cas9 to their respective gene loci and create indels of varying degrees that resulted in loss of protein expression.

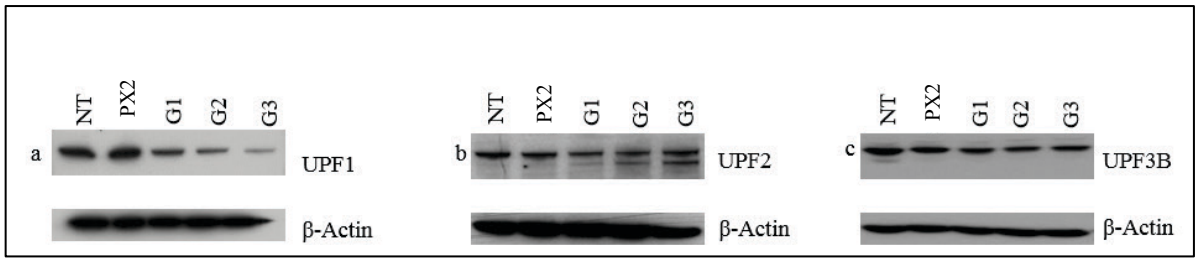


Figure 3.6: Western blot analysis of CRISPR edited HEK293T cells.

Protein was extracted from wildtype cells (NT), and cells either treated with an empty PX459 V2 vector or three different guides targeting (a) *UPF1* (b) *UPF2* (c) *UPF3B* loci in HEK293T cells. The protein lysates were probed for UPF1, UPF2 and UPF3B protein in the respective gene-edited cells. β-Actin (bottom panel) was used as a loading control. NT: Non-transfected.

3.2.3 Optimising CRISPR/Cas9 genome editing technology in hESCs.

Functional validation of the gRNAs in the HEK293T cells proved that the gRNAs were effective in homing Cas9 to the genomic loci to introduce DSBs. To cater for possible off-target effects three unique gRNAs were selected to generate *UPF3A* and *UPF3B* KO clones in hESCs. Generating *UPF3A* and *UPF3B* KO clones were prioritised due to their involvement in NDDs and relationship as paralogs. Due to time constraints, generation of *UPF1* and *UPF2* KO clones was not pursued further. The method used to generate loss of function mutations in *UPF3A* and *UPF3B* in hESCs was previously published (Ran et al. 2013). Briefly, in this method, clonal hESCs were generated by nucleofecting $2x10^5$ cells with 1 μg of the PX459 V2 plasmid and plating the cells in a 100 mm ECM coated dish. After nucleofection, cell culture medium was supplemented with 10 μM ROCK inhibitor for 24 hrs to promote cell survival. 24 hrs post nucleofection, ROCK inhibitor was removed and medium containing puromycin (0.5 μg ml^{-1}) was added for 2 additional days. During this time non-transfected cells die off whilst transfected cells survive and can be expanded as single colonies allowing isolation of clones. After following this protocol no surviving cells were observed after puromycin selection (data not shown). As such a series of optimising strategies were undertaken.

One protocol parameter that was postulated to influence cell survival was the post nucleofection cell density. hESCs typically grow best as colonies, and dissociating them into single cells and performing nucleofection leads to cell death. To promote healthier growth and increase cell survival post nucleofection the rationale was to increase cell density to allow cells to re-establish cell to cell contact (Byrne, Mali & Church 2014). The initial post-nucleofection plating density as published was considered low (at $2x10^5$ cells per 100 mm dish). Another reason contributing to the lack of survival post selection

may have also been a low nucleofection efficiency. Transfection efficiency was ~30% based on parallel nucleofection of a plasmid encoding GFP: pGFPmax control (**Figure 3.7a**). As such, a relatively low number of cells were likely to be expressing the puromycin resistance gene required to tolerate puromycin selection.

A second protocol with alternative plating densities was tested as described by Yang et al., (2014). In this method, $1x10^6$ cells and 1 μg of DNA was used and cells were plated at a very high density (in 1 well of a 48-well plate) (**Figure 3.7b**)). Under these conditions, however, some cells in the control (pGFPmax, no puromycin resistance gene) survived the puromycin selection process (data not shown). It has been reported that drug selection of cells cultured at a very high density can be inefficient (Moore et al. 2010). This method is therefore better suited for FACS based selection instead of drug selection.

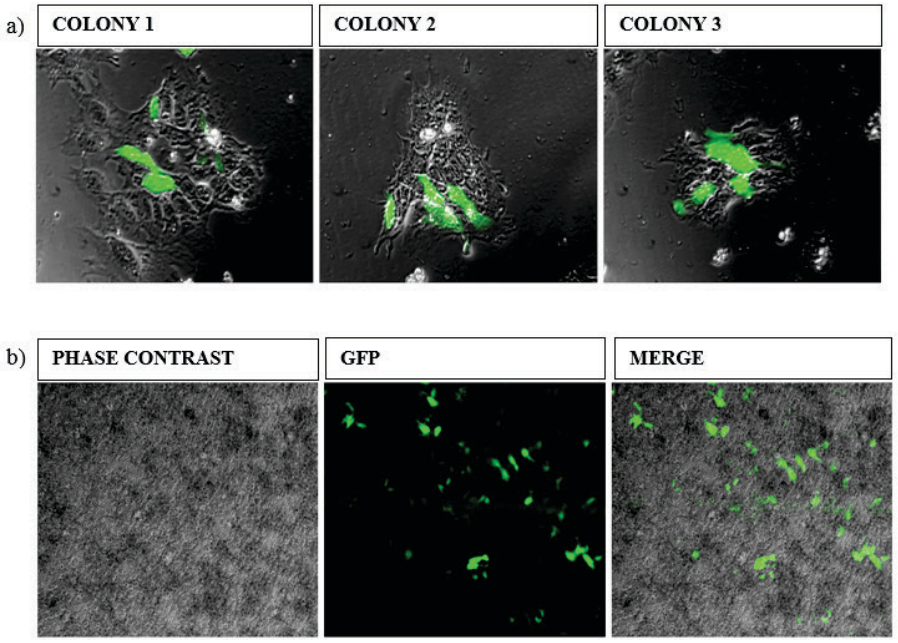


Figure 3.7: Different plating density of cells after nucleofection.

a) $1x10^6$ cells were nucleofected with 1 µg of p.GFPmax plasmid and plated at low density on ECM coated plates. Image taken at day 4 post nucleofection. b) $1x10^6$ cells were nucleofected with 1 µg of p.GFPmax and plated at high density on 48 well plates coated with ECM. Image taken 2 days post nucleofection. Images acquired at 40X magnification.

The Yang et al., (2014) method was repeated, but plating density reduced ($1x10^6$ cells in a well of a 6-well plate). In this instance, cells did survive post-selection and grew into colonies, however the colonies differentiated in culture over time (**Figure 3.8a**). The method by Yang et al., (2014) used medium supplemented with small molecule cocktail of four inhibitors (SMC4) after they performed drug selection. SMC4 contains SB431542, PD0325901, CHIR99021 and Thiazovivin which are inhibitors of the TGF-β, MEK, GSK and ROCK pathways, respectively. Combination of these molecules have been shown to increase the viability of hESCs as single cells and help in maintaining their pluripotency (Valamehr et al. 2012). The above experiment was repeated but modified to

include medium supplemented with SMC4 after puromycin selection. Colonies formed again, but the cells still differentiated and failed to maintain pluripotency (**Figure 3.8b**).

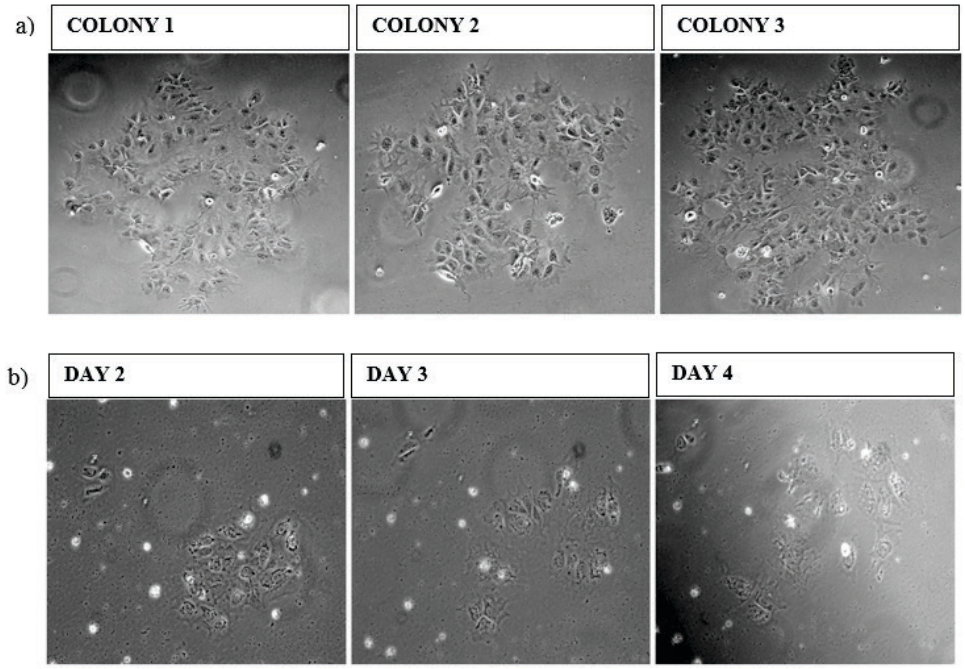


Figure 3.8: Cells plated on a feeder free culture (ECM) after nucleofection.

a) Differentiating colonies 10 days post selection. Cells fail to form colonies with well-defined borders. b) Cells were plated on ECM after nucleofection. After selection cells were cultured in media supplemented with SMC4. Imagen taken 4 days post selection. Images acquired at 40X magnification.

In another protocol by Byrne et al., 2014, SMC4 was used for a period of 8 days post selection, which in their hands greatly improved the viability of their cells. In another modification, instead of plating cells in a feeder independent culture (as I had done previously) they used a feeder dependent culture system. Given that my cultures were also feeder dependent prior to nucleofection and with rationale that growth on feeders can support hESCs growth even at low density, I assessed if puromycin selection could be performed on cells plated directly onto iMEF cultures. A previous study identified that

1 μg/mL of puromycin treatment over 24 hrs had no identifiable effect on iMEF viability (Taniguchi et al. 1998). In a pilot experiment, it was discovered that this regime was sufficient to kill wildtype hESCs with no overt effect on iMEF viability (data not shown).

Based on these findings I performed the following experiment; $1x10^6$ cells were nucleofected with 1 μg of PX459 V2 DNA. After nucleofection cells were plated on 100 mm iMEF prepared dishes in conditioned medium at low density to prevent resultant colonies from merging and cross contamination. 24 hrs later, puromycin selection (1 μg/mL for 24 hrs) was performed and medium changed to medium supplemented with SMC4 for 6 days at which point pluripotent colonies were identifiable by phase contrast microscopy. Long periods of SMC4 exposure has been reported to cause hESCs to differentiate (Yang, L. et al. 2014). In an attempt to further refine the protocol cells were treated with either 10 μM of ROCK inhibitor or SMC4 for only 1 day following selection (compared to 6 days as per above), however, the use of SMC4 for 6 days post selection was optimal in getting the highest numbers of pluripotent colonies (**Figure 3.9**). This strategy had no effect on pluripotency and genome stability of the generated clones (**Figure 3.10**).

In summary after several rounds of optimisation, including testing different cell densities, culture platforms, medium supplementation and puromycin selection criteria, an optimised protocol for the isolation of PX459 V2 transfected hESCs was developed. A step-by-step working protocol is detailed in (**Appendix 1**).

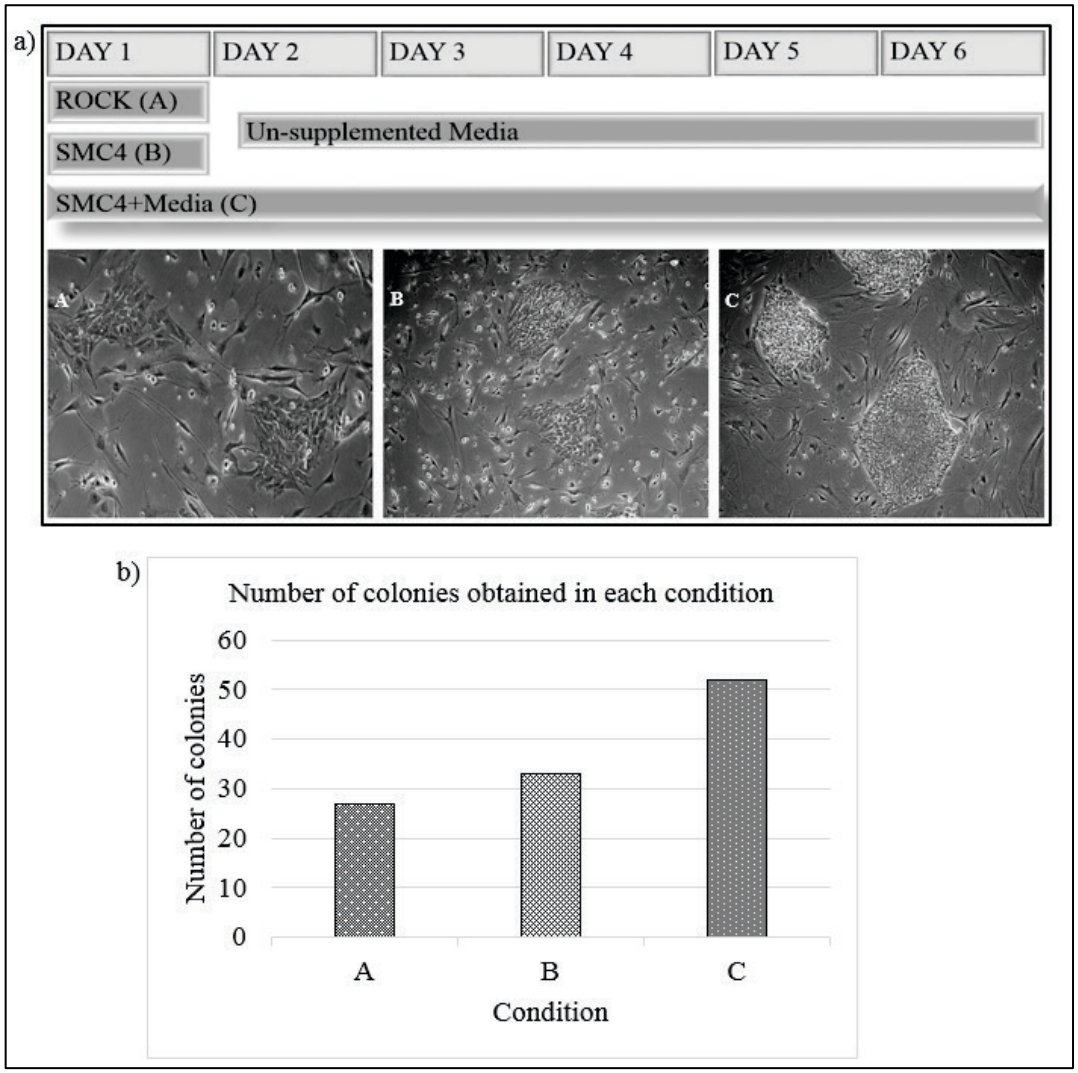


Figure 3.9: Cells plated on iMEFs and supplemented with ROCK or SMC4.

$1x10^6$ cells were nucleofected and plated on iMEF plates. After puromycin selection cells were either cultured with media supplemented with a) 10 µM ROCK for 1 day, b) SMC4 for 1 day or c) SMC4 for 6 days. b) Graph showing the number of clones that grew in each condition. Images acquired at 40X magnification.

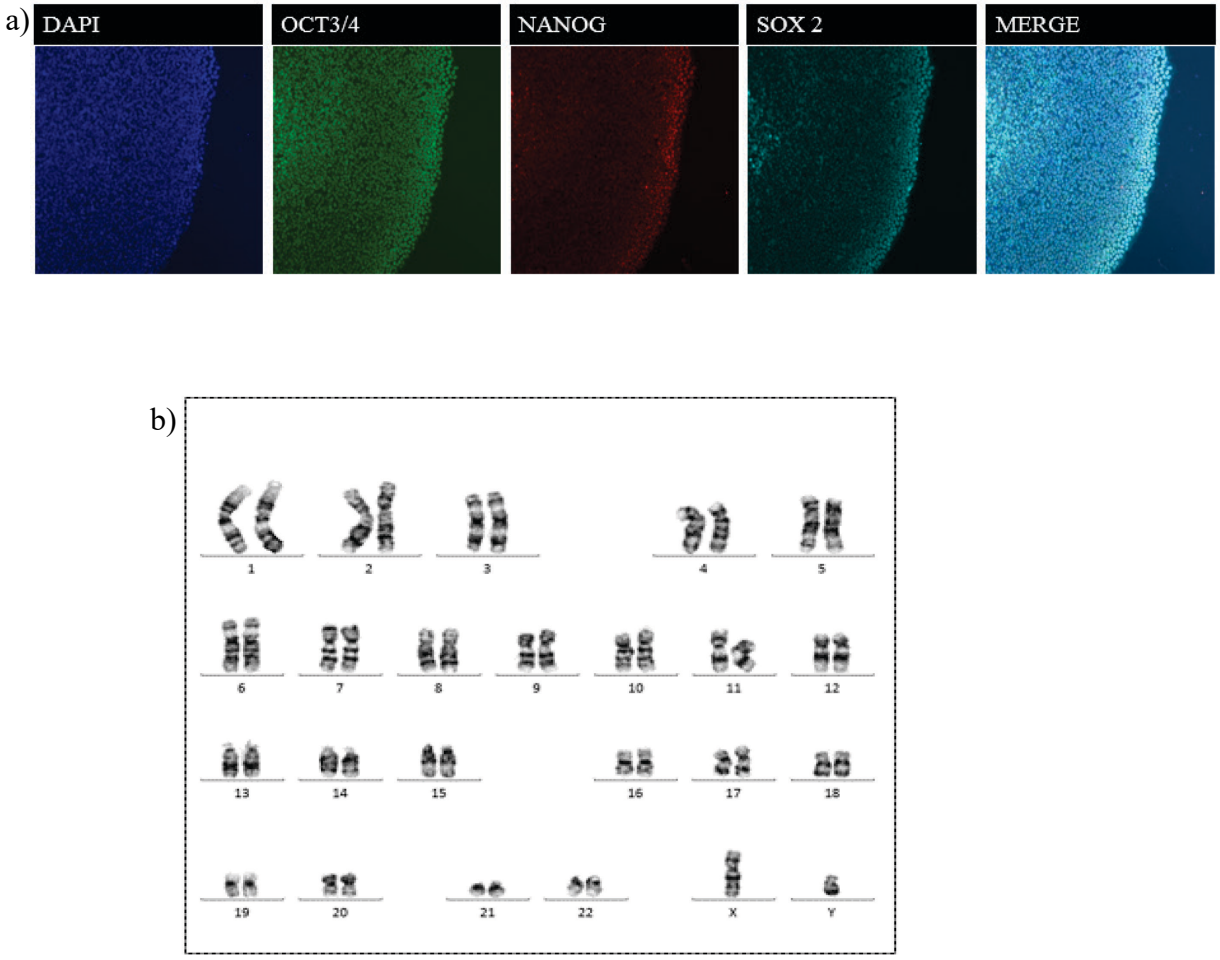


Figure 3.10: hESCs colonies isolated using optimised protocol maintain pluripotency and have a normal karyotype.

a) Representative immunofluorescence image of a hESC clone at day 10 post selection. Cells were fixed and stained with antibodies against pluripotent marker proteins; OCT3/4 (green), NANOG (red) and SOX2 (far red). Nuclei are counterstained with DAPI (blue). Images acquired at 40X magnification. b) Representative metaphase spread of chromosomes isolated from hESC colonies 10 days post selection and expanded in a T25 for one passage. Normal male karyotype is observed.

3.2.4 Generating clonal *UPF3A* and *UPF3B* KO human embryonic cells lines

UPF3A and *UPF3B* KO clones were generated using the optimised CRISPR/Cas9 gene-editing protocol (**Appendix 1**). Ten days post selection, 6 control colonies (nucleofected with an empty PX459 V2 plasmid; PX2) and all colonies nucleofected with the *UPF3A* and *UPF3B* gRNAs that survived puromycin selection were manually selected. hESC clones generated after CRISPR/Cas9 genome editing were genotyped without extensive clonal expansion. A small section of the clone was dissected from each clone at the time of clonal isolation and gDNA harvested. PCR was performed on the gDNA with primers flanking the targeted gene loci. The amplified gDNA was then subjected to Sanger sequencing. **Figure 3.11** and **Figure 3.12** show examples of sequence traces across the *UPF3A* and *UPF3B* targeted loci from a control and three *UPF3A* and three *UPF3B* gene-edited cells that were selected for further analysis.

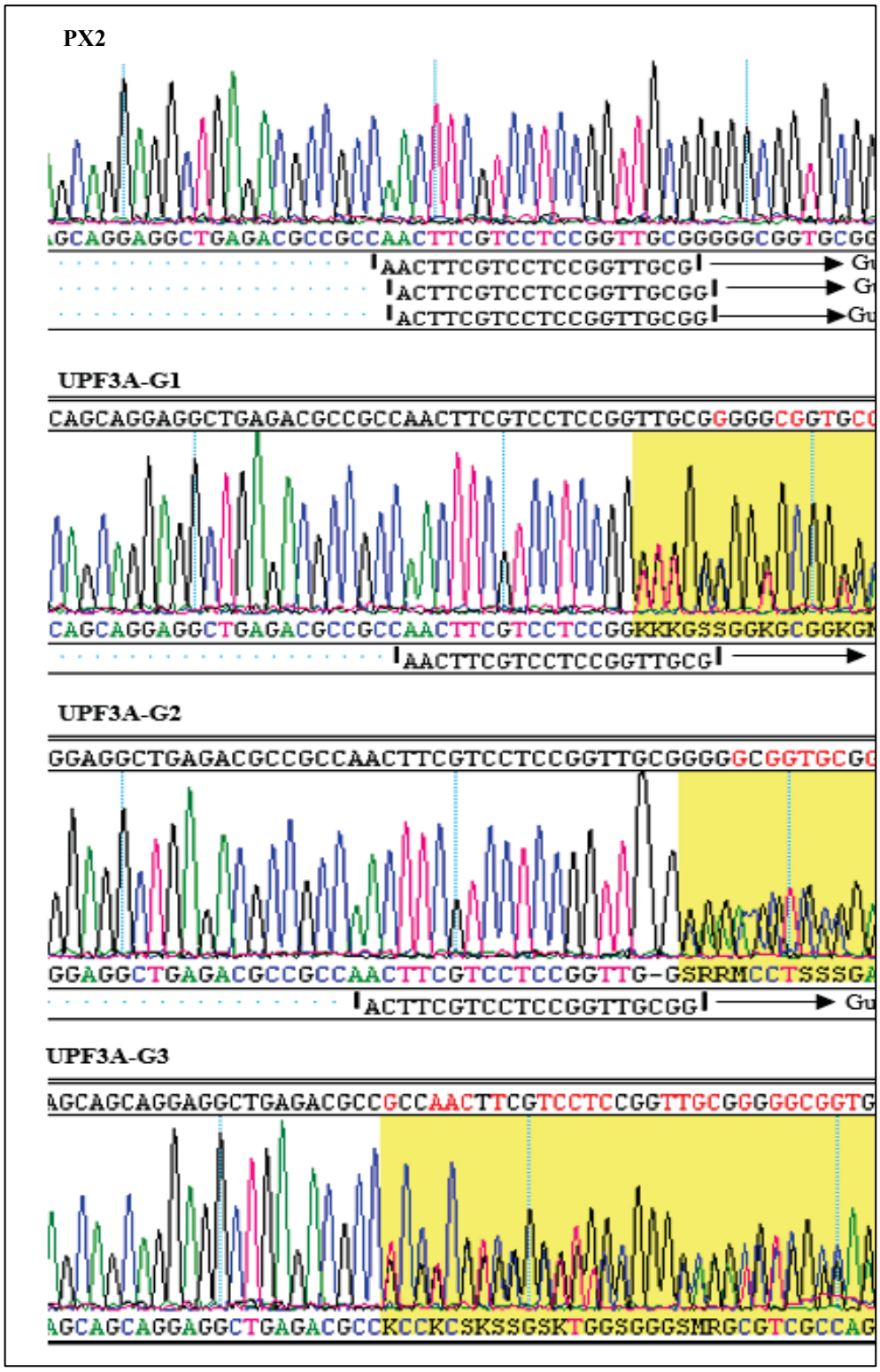


Figure 3.11: Sanger sequencing chromatographs showing *UPF3A* KO hESCs.

gDNA was extracted from a control (PX2) and three *UPF3A* KO clones. PCR amplification was performed with primers flanking the targeted *UPF3A* loci and Sanger sequencing performed on the PCR product. Sequence chromatograph of the sequenced clones. Guide 1 to 3 sequences on the sequence chromatograph show the location of where the gRNA target the *UPF3A* gene. *UPF3A-G3* has no guide aligned with the sequence chromatography as the gRNA sequence is included in the deleted section.

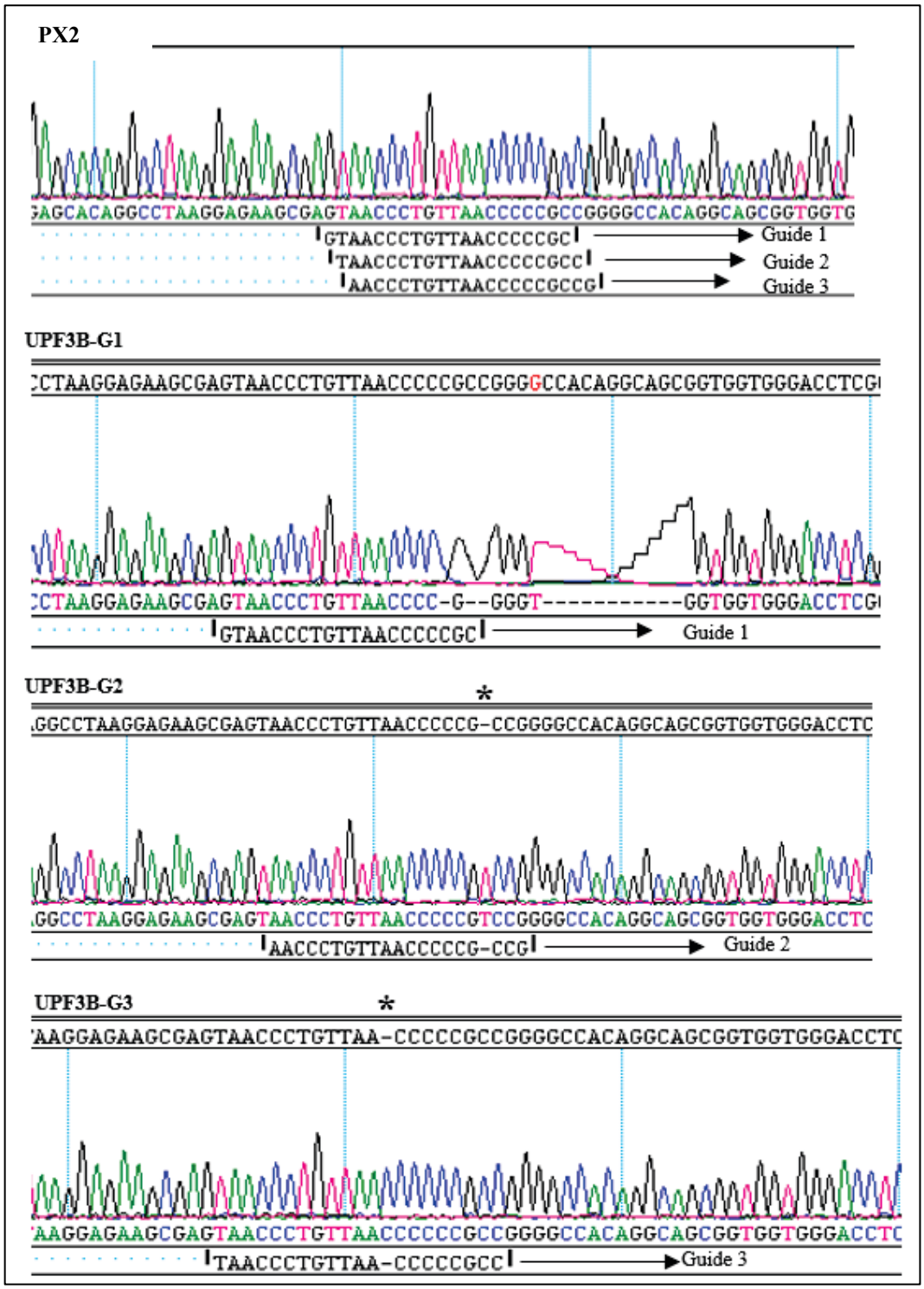
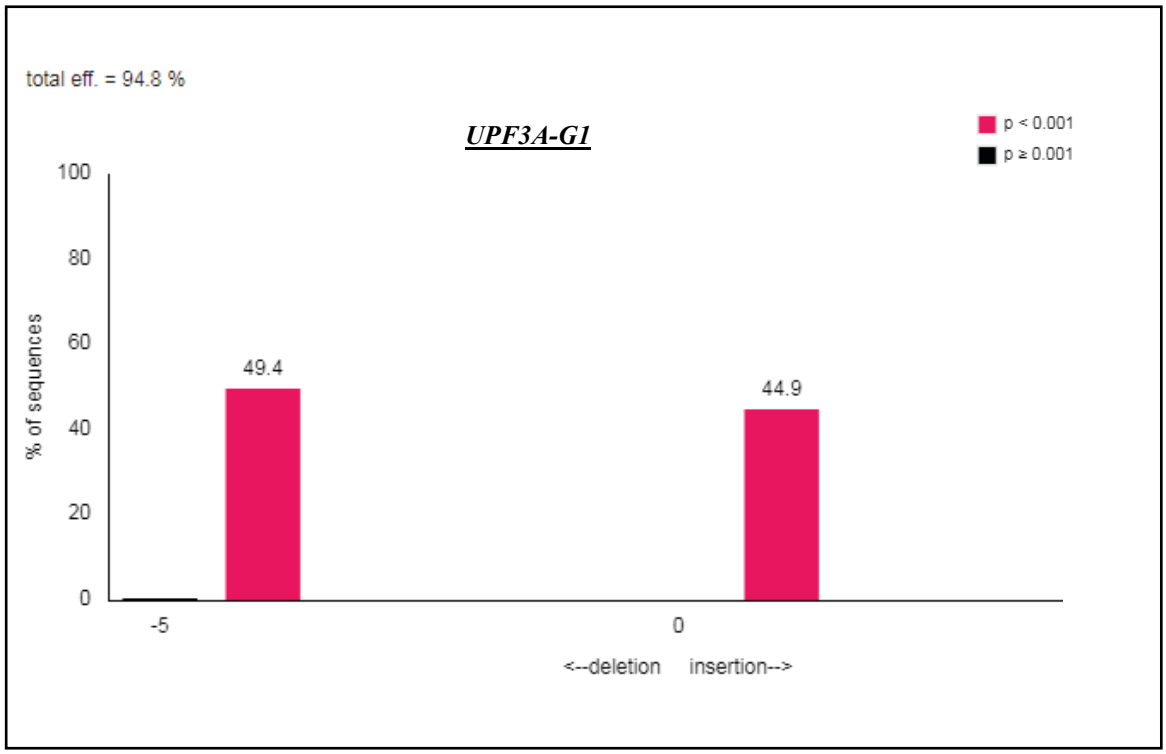


Figure 3.12: Sequence traces for *UPF3B* KO clones and controls selected for analysis.

gDNA was extracted from a control (PX2) and three *UPF3B* KO clones. PCR amplification was performed with primers flanking the targeted *UPF3B* loci and Sanger sequencing performed on the PCR product. Sequence chromatograph of the sequenced clones. Guide 1 to 3 sequences on the sequence chromatograph show the location where the gRNA target the *UPF3B* gene.

After Sanger sequencing, clones with wildtype, in-frame and possible frameshift mutations were identified by determining the number of indels generated. For *UPF3B*, this was done by looking at the sequence as only one sequence trace is expected as *UPF3B* is located on the X chromosome. However, *UPF3A* is autosomal and in an event where a single or both alleles have been edited, two sequence traces would be visible. To deconvolve the sequence and identify the number of indels on each *UPF3A* allele, a software known as TIDE was used. Sanger sequences from **Figure 3.11** were put through the TIDE software to determine the number of indels on each allele and results are shown in **Figure 3.13**. In addition to identifying the compound heterozygous indels at *UPF3A* loci, this program also detected non-clonal lines such as in *UPF3A* G3 CF and *UPF3A* G1 C1 clones (**Appendix 2**).



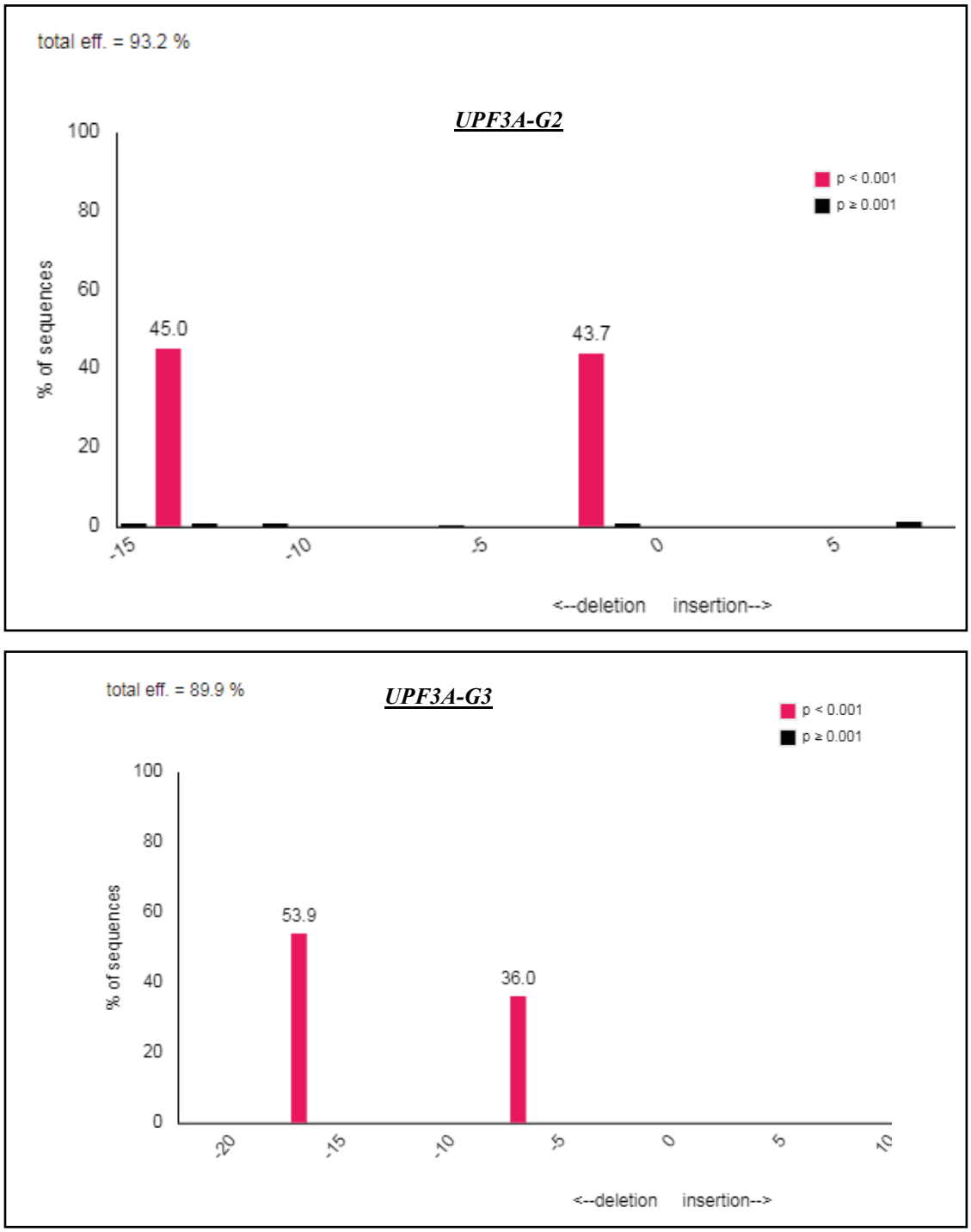


Figure 3.13: Detecting the number of indels in *UPF3A* KO hESCs using TIDE.

Output information from the TIDE program from three *UPF3A* clones from Figure 3.11. The x-axis shows the size of indels in each sequence trace, while the y-axis shows the percentage in abundance of each allele sequence in the sequence chromatography. To determine the number of indels on each allele, the Sanger sequencing trace from a control (empty PX459 V2 plasmid treated cells; PX2) and gene-edited clones (*UPF3A* KO clones) and the gRNA sequence were upload via the TIDE software. The program aligns the control sequence trace and the gRNA to determine the break region. The control and the gene-edited clone trace sequences are then aligned together to determine the differences in the two sequences. TIDE uses the heights of the peaks in each sequence trace to determine the abundance of the nucleotides (Brinkman et al., 2014).

The optimised CRISPR/Cas9 genome editing technology protocol was successfully used to generate gene-edited *UPF3A* and *UPF3B* KO hESC clones. **Table 3-1** shows a summary of all the clones selected from each guide and the number of clones that were targeted by Cas9 (which either generated a frameshift or an in-frame mutation). From the total number of clones that were selected a targeting efficiency (defined as the proportion of clones that had a disrupted target allele) of 34.9% was achieved with 16.5% having a mutation that generated a frameshift mutation. *UPF3A* gRNAs generated more clones with frameshift mutations (both single and bi-allelic) than *UPF3B* gRNAs (**Table 3-2**). Overall the 6 selected gRNAs had variable editing efficiencies.

Target gene and gRNA used	*Number of clones isolated*	*Clones with normal genotype*	*Clones with altered gene loci*	*Clones with frameshift loci*	*Percentage of clones with an altered gene loci*
UPF3A G1	9	4	5	3	
UPF3A G2	8	2	6	4	**48.8**
UPF3A G3	24	15	9	5	
UPF3B G1	27	20	7	3	
UPF3B G2	20	15	5	1	**26.5**
UPF3B G3	21	15	6	2	
Total clones	**109**	**71**	**38**	**18**	
Percentage		**65.1**	**34.9**	**16.5**	

Table 3-1: Targeting efficiency of the gRNAs.

Guides	Clones	Mutations	Type of mutation
UPF3A G1			
	UPF3A G1 C1	Normal, 5 bp del, 10 bp del, 12 bp del (mixed)	Discarded/mixed colony
	UPF3A G1 C3	1 bp ins, 4 bp del	Compound heterozygous
	UPF3A G1 C6	Normal, 1 bp ins	Heterozygous
	UPF3A G1 C8	1 bp ins	*
UPF3A G2			
	UPF3A G2 C1	5 bp del, 7 bp del	Compound heterozygous
	UPF3A G2 C4	Normal, 5 bp del	Heterozygous
	UPF3A G2 C5	2 bp del, 14 bp del	Compound heterozygous
	UPF3A G2 C7	Normal, 5 bp del	Heterozygous
UPF3A G3			
	UPF3A G3 C5	Normal, 1 bp del	Heterozygous
	UPF3A G3 C10	12 bp del, 14 bp del	Compound heterozygous
	UPF3A G3 C11	1 bp del, 1 bp insert	Compound heterozygous
	UPF3A G3 C14	7 bp del, 17 bp del	Compound heterozygous
	UPF3A G3 C15	Normal, 25 bp del	Heterozygous
	UPF3A G3 CF	Normal, 1 bp del, 14 bp del (mixed)	Discarded/mixed colony
UPFB G1			
	UPF3B G1 C2	14 bp del	
	UPF3B G1 C13	29 bp del	
	UPF3B G1 C14	4 bp del	
	UPF3B G1 C17	5 bp del, Normal (mixed)	Discarded/mixed colony
UPF3B G2			
	UPF3B G2 C8	1 bp ins	
UPF3B G3			
	UPF3B G3 C12	1 bp ins	
	UPF3B G3 C13	13 bp del	

Table 3-2: Indels generated in the selected *UPF3A* and *UPF3B* KO clones with frameshift mutations.

* This clone could either have a large deletion on one allele that is not being detected by the designed primers or have both alleles with a one bp insertion.

3.2.5 Karyotype analysis

Karyotype analysis was performed on controls and all clones that had a frameshift mutation after genome editing. Twenty metaphases were analysed from cultures that did not show any abnormalities when screened. However, if an abnormality was detected a larger number of metaphases were analysed. From the 24 clones that were karyotyped only one clone had an abnormal karyotype showing mosaicism for trisomy 12 (47, XY, +12/46, XY) (**Table 3-3** and **Figure 3.14**). For downstream analysis four controls and three *UPF3A* and *UPF3B* KO clones with a normal karyotypes were randomly selected (clones highlighted in bold). These clones are represented in **Figure 3.11** and **Figure 3.12**. The mutation annotations of these KO clones are presented in Chapter 5 (**Table 5-1**). The transcript level, protein expression, and the expected size of proteins from the mutant *UPF3A* and *UPF3B* alleles are also analysed in Chapter 5.

Guides	Clones	Passage number	Metaphases analysed	Karyotype
CONTROLS	**PX1**	**3**	**20**	**Normal**
	PX2	**3**	**20**	**Normal**
	PX3	**3**	**20**	**Normal**
	PX4	**3**	**20**	**Normal**
	PX5	3	20	Normal
	PX6	3	20	Normal
UPF3A G1				
	UPF3A G1 C3	**6**	**20**	**Normal**
	UPF3A G1 C6	8	20	Normal
	UPF3A G1 C8	5	20	Normal
UPF3A G2				
	UPF3A G2 C1	7	20	Normal
	UPF3A G2 C4	5	20	Normal
	UPF3A G2 C5	**5**	**20**	**Normal**
	UPF3A G2 C7	6	20	Normal
UPF3A G3				
	UPF3A G3 C5	5	20	Normal
	UPF3A G3 C10	5	20	Normal
	UPF3A G3 C11	7	22	(47, XY, +12/46, XY)
	UPF3A G3 C14	**6**	**20**	**Normal**
	UPF3A G3 C15	7	20	Normal
UPFB G1				
	UPF3B G1 C2	**6**	**20**	**Normal**
	UPF3B G1 C13	6	20	Normal
	UPF3B G1 C14	5	20	Normal
UPF3B G2				
	UPF3B G2 C8	**5**	**20**	**Normal**
UPF3B G3				
	UPF3B G3 C12	**5**	**20**	**Normal**
	UPF3B G3 C13	6	20	Normal

Table 3-3: Karyotyping gene-edited *UPF3A* and *UPF3B* hESC clones.

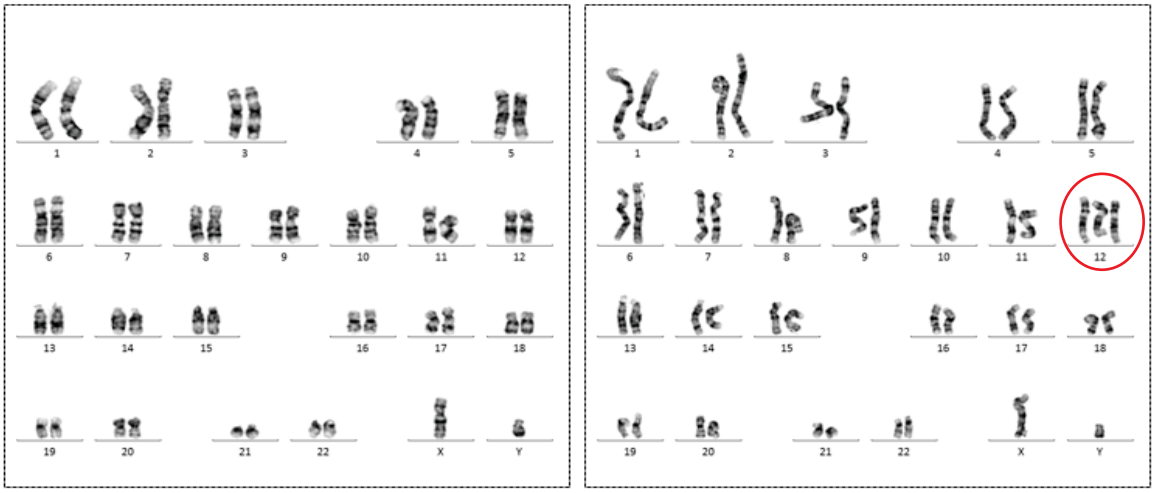


Figure 3.14: Karyotype analysis of selected clones.

Representative metaphase spread of chromosomes isolated from a hESCs that had an abnormal karyotype showing trisomy 12. A total of 22 metaphase spreads were analysed.

3.3 Discussion

CRISPR/Cas9 genome editing technology is an easy and efficient method to edit the genome. It has a stream-lined workflow requiring a single ligation step of the DNA encoding the gRNA when using the plasmid expressing the Cas9 and gRNA. This is particularly useful when editing more than one gene. Furthermore, expression of multiple gRNA sequences using a single CRISPR/Cas9 or multiple plasmids can enable simultaneous editing of several genomic sites within the mammalian genome (Cong, Le et al. 2013). The CRISPR/Cas9 editing technology enables studies to be performed for loss of function mutations by creating indels (utilising the NHEJ pathway) which may lead to frameshift mutations (Kuscu et al. 2017) or it can be used to generate specific knock-in mutations using a single-stranded oligo DNA nucleotide (DNA repair template) (HDR pathway) (Cong, L. et al. 2013). When combined with hESCs, these tools provide a powerful resource to investigate the effects of genetic variants across a broad range of developmental models and tissue types. In this Chapter, CRISPR/Cas9 gene editing technology was applied to hESCs to generate frameshift mutations (i.e. knockout mutations) in genes involved in the NMD pathway. Loss of function mutations in NMD genes result in a range of NDDs in humans.

The optimised method derived from the initial study from Ran et al., 2013 to generate clonal hESCs using CRISPR/Cas9 proved to be efficient and robust. It did not have any effects on the pluripotency and genome stability of the hESCs as shown by immunofluorescence staining of pluripotency markers (OCT3/4, NANOG and SOX2) and karyotype analysis (**Figure 3.10**). A targeting efficiency (defined as the proportion of clones that had a disrupted targeted allele) of between 26.5 and 48.8% (*UPF3B* and *UPF3A* respectively) was achieved (**Table 3-1**). This targeting efficiency was high compared to Mali et al., 2013 and Byrne et al, 2014 where they reported targeting

efficiencies between 2-4% and 1-10% respectively (Byrne, Mali & Church 2014; Mali et al. 2013). Our targeting efficiency was however lower compared to a study by Ding et al., 2013 that reported a targeting efficiency of between 51-79% (Ding, Q. et al. 2013). The method used to introduce the CRISPR/Cas9 components (Cas9 and gRNA) into cells has been shown to have an impact on the targeting efficiency. Lower targeting efficiency has been reported when the Cas9 protein and the gRNAs were delivered as crRNA and tracrRNA fusion transcripts (mRNA) (Byrne, Mali & Church 2014; Mali et al. 2013). In studies where a higher targeting efficiency was reported, the CRISPR/Cas9 components were delivered as DNA via two separate plasmids that expressed either the Cas9 endonuclease or gRNA (Ding, Q. et al. 2013). This was however, unexpected as it has been reported that Cas9 protein and gRNA mRNA delivery has a higher targeting efficiency compared to the CRISPR/Cas9 DNA delivery (Sahel, Mittal & Chitkara 2019). Therefore the difference in the targeting efficiencies could be due to the gRNAs used as the ability of a gRNA to efficiently create DSB in the target DNA can vary based on the guide RNA sequence and position in the targeted gene (Wang, T. et al. 2014). From the *UPF3A* KO hESCs clones that were selected, both heterozygous and compound heterozygous clones were obtained (41.7% and 50% respectively (**Table 3-2**)). No clone with homozygous mutations were detected, however Ding et al., 2013 detected 7-25% homozygous clones. One *UPF3A* KO clone had a single bp insertion detected using TIDE (a program that deconvolutes sequence traces and detects the number of indels in each sequence). This suggests that either this clone has a 1 bp insertion in each allele, or that, there is one 1 bp insertion on one allele and a large deletion in the other allele that is not being detected by the primers flanking the targeted genomic loci. However this was not investigated further. The CRISPR/Cas9 genome editing technology was able to generate a range of indels ranging from 1 bp deletions to as large as 55 bp (data not shown) in hESCs. In HEK293T cells CRISPR/Cas9-mediated editing of the *UPF3A* locus resulted

in some large deletions whereas editing of *UPF1*, *UPF2* and *UPF3B* only had small indels (**Figure 3.4**). The large deletions that were observed in HEK293T where not observed in *UPF3A* CRISPR/Cas9 edited hESCs, with the largest deletion detected being 25 bp in hESCs (**Table 3-2**). The difference in the size of deletions generated by the same gRNA in HEK293T cells and hESCs could be due to the different cell types.

Karyotype analysis was performed on control cell lines and all clones with frameshift mutations generated at the targeted loci. Karyotyping analysis revealed one abnormal clone with mosaicism for trisomy 12 (**Figure 3.14**). Karyotyping after CRISPR-Cas 9 is important as hESCs genomes are unstable (Henry et al. 2018). During the editing process, hESCs are passaged as single cells, nucleofected and selected with an antibiotic. These unfavourable environments impose a strong selective pressure on viability which may favour genetic alterations with positive influence (Peterson & Loring 2014). Gain of chromosome 12, 17, 20 and X have been frequently detected in hESCs cultures (Baker et al. 2007; Nguyen et al. 2014). These abnormalities may provide a selective advantage by promoting self-renewal or proliferation in culture for example through gain of NANOG which is located on chromosome 12 (Tosca et al. 2015).

In conclusion CRISPR/Cas9 genome editing technology was successfully used to generate a series of *UPF3A* and *UPF3B* frameshift mutant hESC lines that would enable the study of the impact of *UPF3A* and *UPF3B* loss-of-function mutations in disease. The use of iMEFs and SMC4 were important in the gene-editing method to attain pluripotent colonies. The process of editing also had no discernible effect on the pluripotency and genome stability at the level of analysis performed. However it has been shown that CRISPR/Cas9 editing events can be associated with generation of off-target mutations

caused by gRNA binding to mismatched DNA sequences. The experimental design of choosing three independent gRNAs to generate unique gene edited lines served to control against potential off-target effects. In the next chapter the genome wide interrogation of the impact of CRISPR/Cas9 editing was performed.

Chapter 4 : Whole genome sequencing of CRISPR/Cas9 gene-edited hESCs

4.1 Introduction

CRISPR/Cas9 gene editing technology is a highly effective gene editing tool used for modifying the genome for functional elucidation of genetic variants in cell culture (Wang, Tim et al. 2014) and animals models such as mice (Platt et al. 2014; Swiech et al. 2015). As described in chapter 3, CRISPR/Cas9 genome editing technology utilises a gRNA that binds a complementary sequence of DNA to home the Cas9 endonuclease to specific genome sites and generates DSBs (Pattanayak et al. 2013). However, the complementary binding of the gRNA to the DNA is not completely specific. It has been shown that the gRNAs can bind mismatched sequences of DNA at low frequency and as such Cas9 can be homed to undesired locations where it can potentially generate undesired 'off-target' mutations (Cradick et al. 2013; Fu et al. 2013; Hsu et al. 2013).

Binding of the gRNA to a mismatched DNA sequence has been shown to depend on several factors such as the number of nucleotide mismatches and the position of the mismatch (Fu et al. 2013). The 8-13 nucleotide 5' of the PAM sequence in the gRNA is known as the 'seed sequence' and has been shown to lead to the specificity of the gRNA binding to the complementary DNA sequence (Byrne, Mali & Church 2014; Cong, Le et al. 2013). The remaining nucleotides of the gRNA sequence have been shown to tolerate some degree of mismatch. In addition to binding mismatched DNA sequences, the gRNA can bind DNA sequences that contain insertions (DNA bulge) or deletions (RNA bulge) within the targeted region, further introducing opportunities for off-target homing of Cas9 (Lin et al. 2014).

For clinical applications such as gene therapy, off-target mutations are a major concern as these genomic changes can affect important genes and create chromosomal rearrangement (Yang, Luhan et al. 2014). Off-target mutations can also have an impact

on results of various functional studies when elucidating genetic variants if coding regions within a gene are affected. This indicates that comprehensive off-target analysis should be performed on CRISPR gene-edited cells to detect any off-target mutations.

A limited number of studies have employed whole genome sequencing (WGS) of CRISPR/Cas9 gene-edited hESCs and iPSCs clones, and have shown little or no off-target mutations respectively (Smith et al. 2014; Veres et al. 2014) while other human cell lines have low levels of off-targets mutations (Anderson et al. 2015). Even though no off-target mutations were identified in iPSCs exposed to CRISPR/Cas9 gene-editing systems, individual iPSCs clones had unique indels caused by clonal heterogeneity (Veres et al. 2014).

This chapter describes in detail the experiment and the analysis performed to detect off-target mutations in the CRISPR/Cas9 clones generated as part of this study using WGS. It includes a brief description of how the gene-edited hESCs were generated, a detailed description of the detection of clone specific *de novo* small nucleotide variants (SNVs) and structural variants (SVs) and validation of the variant calls. The results showed that none of the unique *de novo* SNVs detected were likely due to off-target mutations generated by CRISPR/Cas9 gene. However, SV analysis detected large deletions (21-578 bp, with the largest detected being ~13 megabases) and duplications (294-626 bp) that overlapped with predicted off-target sites and their validation is still in progress to determine if they are real. The analysis revealed that the detected *de novo* variants were more likely due to extended culturing of hESCs, the process of CRISPR/Cas9 editing, and presence of mosaicism in the parental cell line. Even though a relatively large number of unique variants were detected in each clone, a test of genetic relatedness between the

clones, parental cell line and an external H1 cell line (published H1 WGS data) revealed that they were highly similar with only minor variations.

4.2 Results

4.2.1 hESCs generated using the CRISPR/Cas9 gene editing technology

H1 hESCs were used to generate gene modified stem cells. H1 hESCs were expanded and single cell passaging was performed to generate a pool of cells ($7x10^6$ cells) to execute a series of nucleofection experiments required to generate all clonal lines in parallel. After nucleofection, puromycin selection was performed, colonies isolated and clonal cell lines expanded (**Figure 4.1a)**. Loci specific genotyping was performed, and the clones with indels generating frameshift mutations karyotyped. Four controls (nucleofected with the empty PX459 V2 vector (PX2)), and three *UPF3A* and three *UPF3B* KO clones (each generated with a different gRNA) were selected for further analysis. Each clone was expanded for approximately 20 passages before gDNA was collected and submitted for WGS. Samples submitted for WGS included the parental cell line at passage number 30 (P30) from the initial vial/ stock, one control (PX2) and six of the KO clones each at an approximate average passage number 86 (P86) (**Figure 4.1b**).

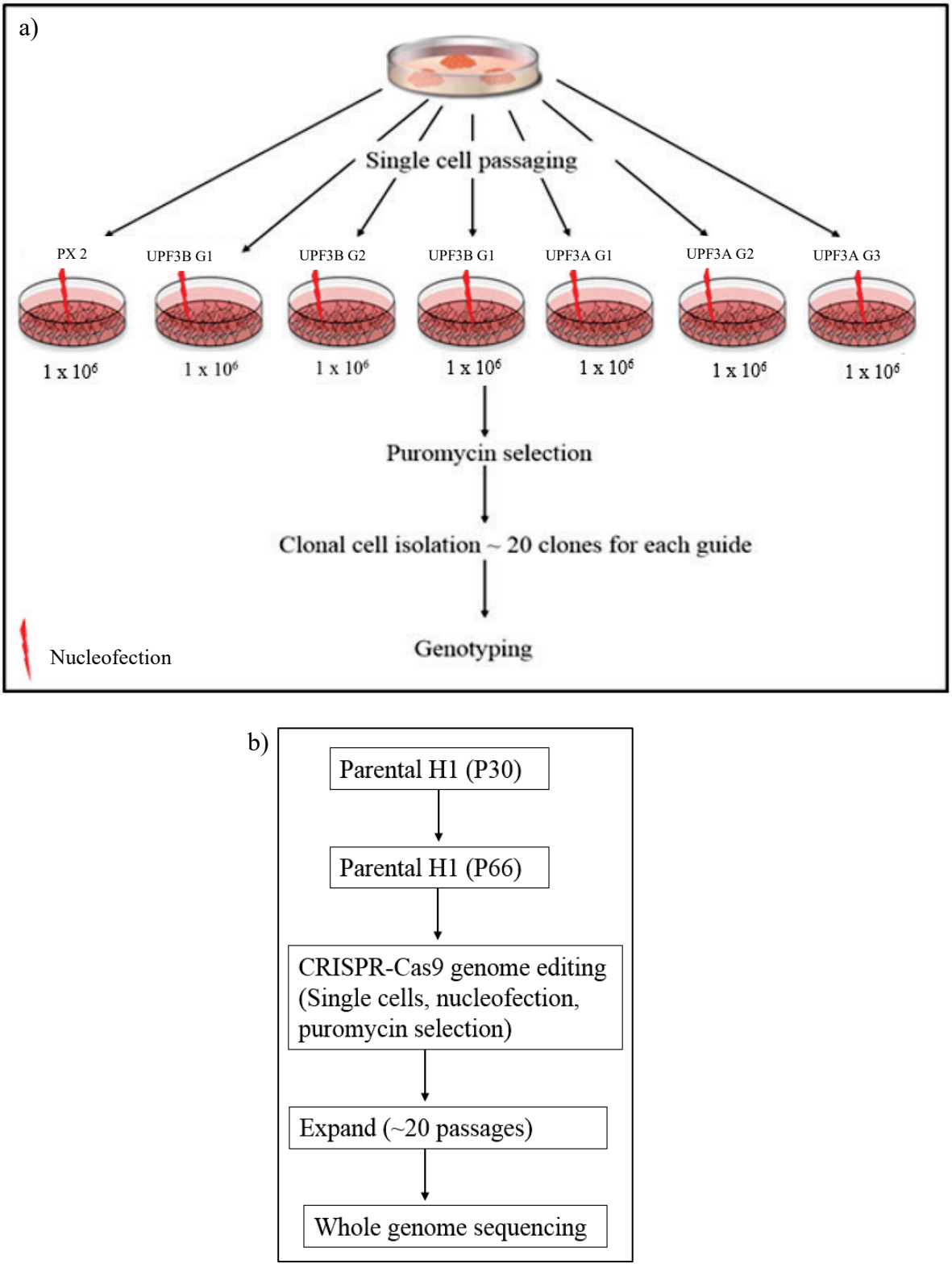


Figure 4.1: CRISPR/Cas9 gene-editing process and clones selected for WGS analysis.

a) *UPF3A* and *UPF3B* KO clones were generated from a single cell derived from a large pool of H1 hESCs that were nucleofected with the designed CRISPR/Cas9 plasmid and selected through drug selection and expanded. b) Samples submitted for WGS included the parental cell line (P30) and seven gene-edited clones (~P86).

4.2.2 Identification of unique *de novo* SNVs and SVs analysis in gene-edited hESCs.

DNA samples were sequenced using 150 bp paired reads by Illumina HiSq X-10. Fastq formatted sequence files were mapped to the UCSC hg19 build of the reference human genome and SNVs were called using GATK (Van der Auwera et al. 2013) and SVs called with DELLY (Rausch et al. 2012) (**Figure 4.2**). To comprehensively investigate off-target mutations generated by CRISPR/Cas9 technology, clonal specific *de novo* mutations were identified from the WGS data.

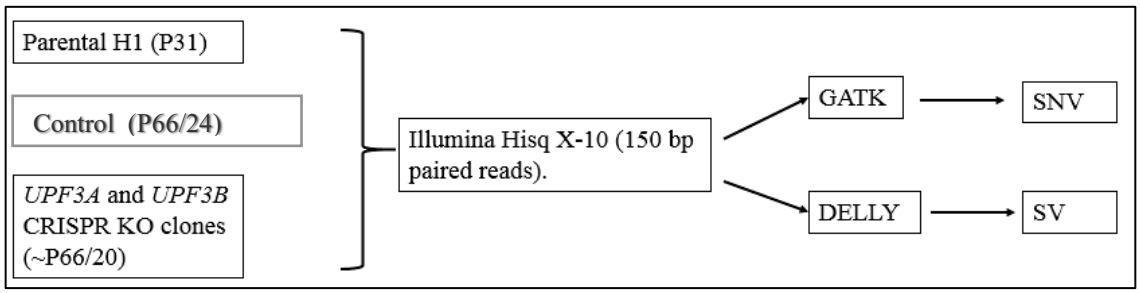


Figure 4.2: Analysis of SNVs and SVs in *UPF3A* and *UPF3B* KO hESCs clones.

WGS was performed with 150 bp paired reads using Illumina Hisq X-10 on the parental cell line, a control and three *UPF3A* and *UPF3B* KO clones. The sequencing results were analysed using GATK to detect SNVs and DELLY for SVs in the CRISPR edited clones.

SNV analysis detected approximately $4.4x10^6$ variants in the parental H1 and each of the analysed control and KO clones when compared to the human genome reference sequence (**Table 4-1a**).The identification of *de-novo* SNVs due to off-target mutagenesis was achieved by removing germ line and acquired variants present in the parental H1 line from the PX2 control and KO clones. Removal of the H1 variants revealed variants ranging between $16.7x10^4$ and $23.2x10^4$ per clonal cell line. The detection of approximately $20x10^4$ variants across clones after removal of germ line and

acquired variants in the parental was mainly due to the lower passage number of the parental cell line, ~56 passages lower than the control and KO clones (see section 4.2.1). This is because the parental clone had fewer variants due to a lower passage number. The longer the cells are passaged and stay in culture, the more variants they acquire. Therefore a higher passage number cells with have a higher number of variants compared to lower passage cells. The detected variants in the KO clones (each using 1 of 6 different gRNAs) were further filtered to remove all variants present in the control. After filtering, the number of SNVs in the KO clones was reduced to between $6.1x10^3$ to $8.4x10^3$ SNVs per cell line. The reason for this reduction is that the control clone is approximately the same passage number as the KO clones, and also went through the same experimental process leading to clonal cell isolation. Further filtering of the variants was performed to identify unique *de novo* variants by removing all variants that were detected in at least two clone. This further reduced the variants detected in each clone to approximately ~3083 variants. These unique *de novo* variants where then categorised into single nucleotide polymorphisms (SNPs) and insertions/deletions (indels). As these variants can affect results of functional studies when elucidating the role of *UPF3A* and *UPF3B* if coding regions within a gene are affected, variants located in exons or those with a role in splicing were analysed. From these variants, nonsynonymous, frameshift or variants with unknown function were selected. Genes that contained these variants were then identified (**Table 4-1a**). To explore if any of these unique *de novo* SNVs are due to likely off-targets mutations, a list of predicted off-targets was generated and overlapped with the SNVs for each clone (See section 4.2.3).

Cas9-induced DNA DSBs are predominantly repaired by NHEJ to produce small indels and SVs (Veres et al. 2014; Yang, Luhan et al. 2014). Single gRNAs induced deletions

up to 600 bp in mouse zygotes (Shin, Ha Youn et al. 2017). Large deletions and complex genomic rearrangements at CRISPR/Cas9 targeted loci in mouse ESCs and human differentiated cell line have also been reported (Kosicki, Tomberg & Bradley 2018). To determine if Cas9-induced DNA DSBs lead to SVs such as deletions, insertions, inversions or duplications, unique *de-novo* SVs were identified using DELLY. The SVs identified in this section are additional larger variants not detected using GAKT (independent of SNVs). The filtering criteria of the data to detect unique *de novo* SVs was performed stepwise as above (SNVs), where the Fastq formatted sequence files were first mapped to the reference human genome and the H1, control and any shared variants between two or more clones were removed from the KO clones systematically. The number of unique *de novo* variants were between 1524 and 2135. These variants were then categorised based on the type of SV in each clonal line. This revealed that from the SVs detected, deletions represented the largest proportion and inversions being the least frequent across all clones (**Table 4-1b**). To explore if any of these unique *de novo* SVs are due to likely off-target mutations, a list of predicted off-targets was generated and overlapped with the SVs for each clone (See section 4.2.3).

4.2.3 Off-target mutation analysis in gene-edited clones

To determine if any of the unique *de novo* SNVs and SVs in each KO clone were due to off-target mutations, a list of predicted gRNA off-targets sites were generated using the Cas-OFFinder program (www.rgenome.net/cas-offinder) (Bae, Park & Kim 2014) (**Off-targets-a**) and the CRISPR design program (crispr.mit.edu) (**Off-target-b**) (**Table 4-2**). SNVs located at the potential off-target sites were identified using *bfctools* viewer filtered with a custom bed file that had a 20 bp window flanking each predicted off-target site. From the total list of off-targets generated none overlapped with the high quality filtered

clonal *de novo* indels. The identified indels in each clone corresponded to the previously identified on-target indels generated in *UPF3A* and *UPF3B* (**Table 4-2b**). By reducing stringency to also include 'low quality' SNVs, re-analysis identified only one off-target indel in *LINC01865* in a single clone. Validation of this off-target variant was unable to be performed due to lack of suitable primers available for flanking regions.

To determine if any of the unique *de novo* SVs detected in KO clones overlapped with predicted off-targets sites, *BedTools* was used to overlap each sample with the corresponding off-target sites, allowing for a 200 bp window around the predicted off-target regions. Resulting SVs were de-duplicated/merged (since the provided off-target regions have overlapping records) and filtered to remove low quality variants (set FILTER=PASS). **Table 4-2b** shows the number of deletions and duplications of variants that overlapped with the off-target sites. The deletions were between 21-578 bp, with the largest detected being ~13 megabases and large duplications of between 294-626 bp (**Appendix 3**).

a)

	H1	PX2	UPF3A-G1	UPF3A-G2	UPF3A-G3	UPF3B-G1	UPF3B-G2	UPF3B-G3
PASS variants (hg19 reference)	4403252	4391192	4403564	4419210	4382615	4359304	4345768	4363856
H1 variants removed		20337	21539	23176	18927	18075	16705	17740
PX2 variants removed			7696	8423	7254	7343	6148	6841
Variants present in 1 clone or more			4534	4710	4612	3674	3408	4266
SNV-unique variants			3162	3713	2642	3669	2740	2575
SNP-unique variants			1580	1765	1381	2371	1640	1437
INDELS-unique variants			1582	1948	1261	1298	1100	1138
SNV located in exonic regions or important in splicing			24	29	28	24	31	24
Nonsynonymous/frameshift/unknown			12	15	17	17	19	14
number of unique genes			2	2	2	6	3	0

b)

	H1	PX2	UPF3A-G1	UPF3A-G2	UPF3A-G3	UPF3B-G1	UPF3B-G2	UPF3B-G3
PASS variants (hg19 reference)	59048	58036	57950	58416	58584	56818	59404	57253
H1 variants removed		5843	6360	7122	7176	6738	6330	6689
PX2 variants removed			3890	4720	4741	4455	3850	4277
Variants detected in 1 or more clones			2366	2607	2606	2414	2323	2430
Unique variants			1524	2113	2135	2041	1527	1847
DEL			1243	1753	1817	1747	1302	1594
INS			63	48	66	84	64	60
BND			73	134	85	32	45	46
INV			19	28	25	27	23	20
DUP			126	150	142	151	93	127

Table 4-1: Numbers of SNVs and SVs detected by WGS.

The effects of filtering strategies and breakdown of the types of unique *de novo* variants in each clonal line. a) SNVs were broken down into SNPs and indels. Number of variants located in exonic regions, involved in splicing and with unknown functions were analysed. The number of genes impacted by the detected variants were identified. b) SVs classified according to the type of the SVs identified.

a)

Predicted off-target mutations			
Samples	**Off-targets-A**	**Off-targets-B**	**Total**
UPF3A-G1	3256	21	3277
UPF3A-G2	3780	28	3808
UPF3A-G3	6148	31	6179
UPF3B-G1	3806	31	3837
UPF3B-G2	4064	32	4096
UPF3B-G3	4833	46	4879

b)

Variants overlapping with predicted off targets					
Sample	**DEL**	**DUP**	**INS**	**SNPs**	**INDELS**
UPF3A-G1	13	8	0	0	0
UPF3A-G2	8	6	0	0	0
UPF3A-G3	7	5	0	0	1
UPF3B-G1	10	7	0	0	2
UPF3B-G2	10	7	0	0	1
UPF3B-G3	6	7	0	0	1

Table 4-2: Off-target mutation analysis in CRISPR/Cas9 edited cells.

a) The total numbers of predicted off-targets and b) the number of SVs (del, dup, ins) and SNVs (SNPs and indels) detected in each gene-edited clone that overlapped with the predicted off-target sites.

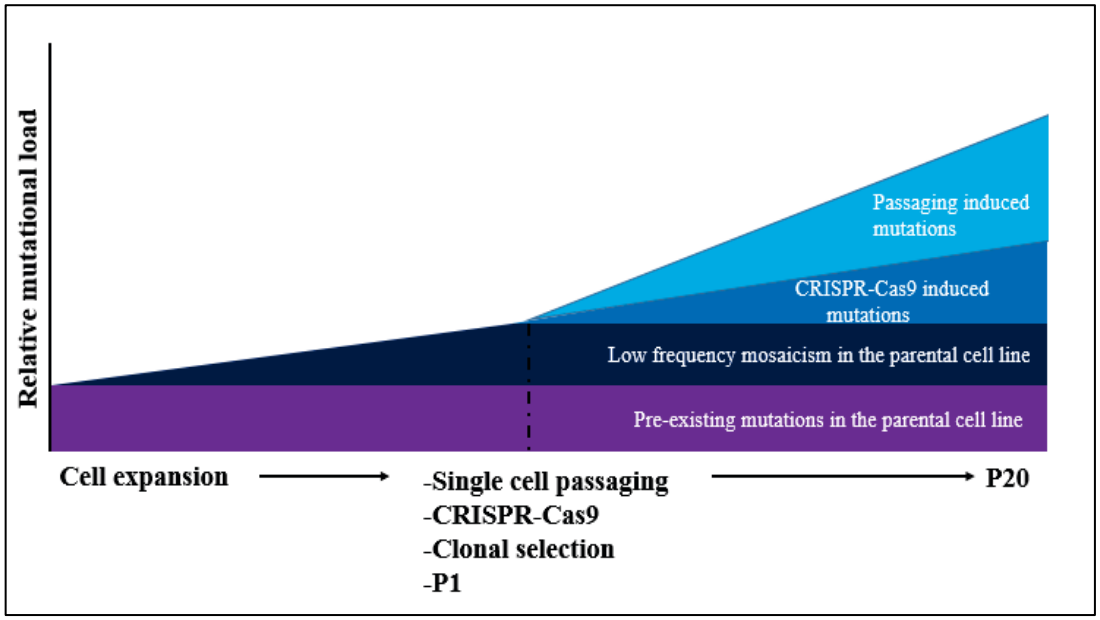
4.2.4 Mutational load in human embryonic stem cells

hESCs are subjected to *in-vitro* culture conditions which can provide a 'selective' pressure by selecting cells with growth acceleration which can be underpinned by chromosomal alterations (Draper et al. 2004). Accumulation of CNVs detected in high passage hESCs results from a series of individual mutation events during prolonged culture periods. High resolution SNP genotyping on hESCs, iPSCs, somatic stem cells, primary cells and tissues reveal hESCs have a higher frequency of genome aberrations (Laurent et al. 2011). The unique *de novo* SNV variants detected in the gene-edited cells did not overlap with any of the off-target sites. The presence of mosaicism in the parental cell line, the CRISPR/Cas9 editing process and cell passaging i.e. the number of cell

divisions, - could be contributing to the mutational load in these cells. Newly introduced mutations are often present in a subgroup of cells in a given culture leading to mosaic mutations (Martincorena & Campbell 2015). Isolation and expansion of single cell clones from a pool of cells with mosaicism could lead to derivation of clonal lines with mutations that are present at very low frequencies in the parental pool of cells. An example of this was evident during validation of one of the low quality variants, which displayed mosaicism (**Figure 4.5**). The CRISPR/Cas9 genome editing process in hESCs could also possibly also lead to accumulation of novel mutations as this method includes steps that can be stressful to the cells. During editing, hESCs are passaged as single cells, nucleofected and selected with an antibiotic and clonally expanded. A strong selective pressure may arise that favours the surviving cells which may also be at least partly attributed to genomic alterations in these cells (**Figure 4.3a**) (Peterson & Loring 2014).

The Cas9- induced DNA DSBs are predominantly repaired by NHEJ to produce small indels (Yang, Luhan et al. 2014). To focus on mutations generated during cell culture and not CRISPR/Cas9, SNVs (with only H1 variants filtered out, and shared variants shared between two clones or more removed) were separated into SNPs and indels as the CRISPR/Cas9 does not typically introduce missense changes. There was no overt difference in the number of SNP or indels variants found in the control and the *UPF3A* KO clones (**Figure 4.3b**). However, in *UPF3B* KO clones, SNPs (which are more cell culture related) were more abundant when compared to the number of indels (**Figure 4.3b**). It is noteworthy that *UPF3B KO* clones were more difficult to culture compared to the *UPF3A KO* clones and control clones. This difficulty could impact additional selective pressure and contribute to these cells acquiring more culture related mutations (SNPs).

a)



b)

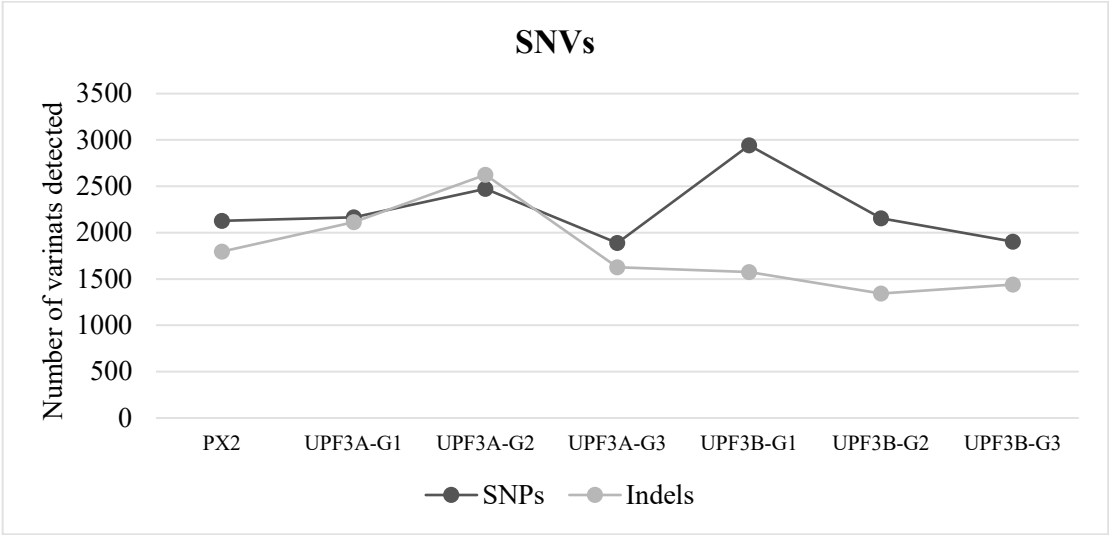


Figure 4.3: Passaging induced mutations and the CRISPR/Cas9 editing process are the main contributing factors to an increased mutational load in hESCs.

a) Factors that contribute to a high mutational load in hESCs in higher passage number hESCs. b) Graph showing the number of SNVs in each CRISPR clones. SNVs were divided into SNPs and indels. Figure adapted from (Ji et al. 2012).

4.2.5 Validation of variant calling confidence

To confirm high quality variant calls, a total of fifteen variants were selected. The variants were validated by genomic PCR followed by Sanger sequencing with 100% overall confirmation (**Figure 4.4** and **Appendix 4**). To verify the stringency in the filtering criteria and to ensure important variants were not filtered out, eight low quality variants were selected that were part of the variants that did not meet the PASS filtering criteria and were therefore removed from the analysis. Seven of the eight low quality variants did not validate (results not shown) as expected while one, *TRAPPC8* showed low frequency mosaicism (~10.5% of a 3 bp deletion in the cells population) in clone *UPF3B-G1* (**Figure 4.5**) indicating that the sensitivity of variant detection in the analysis is high.

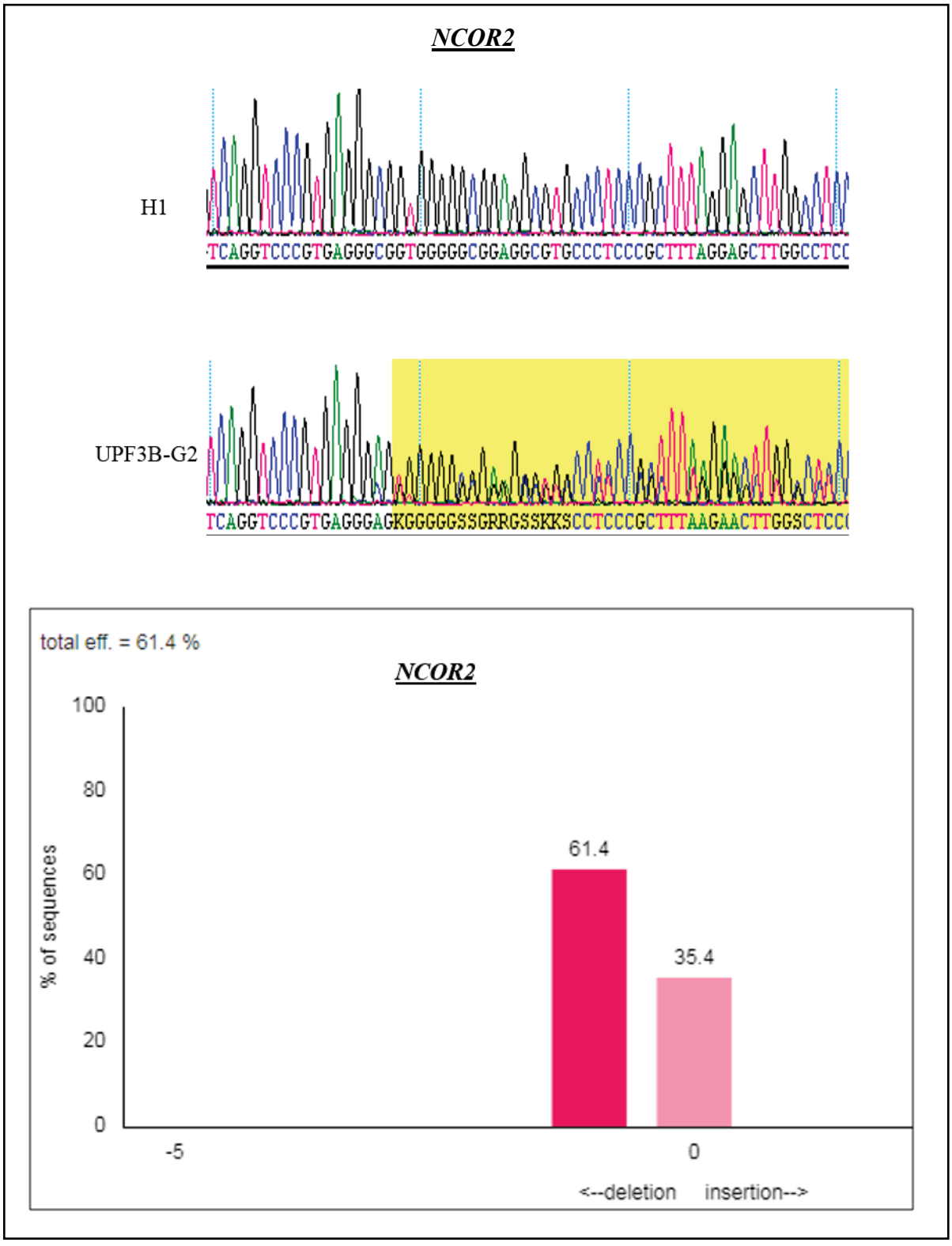
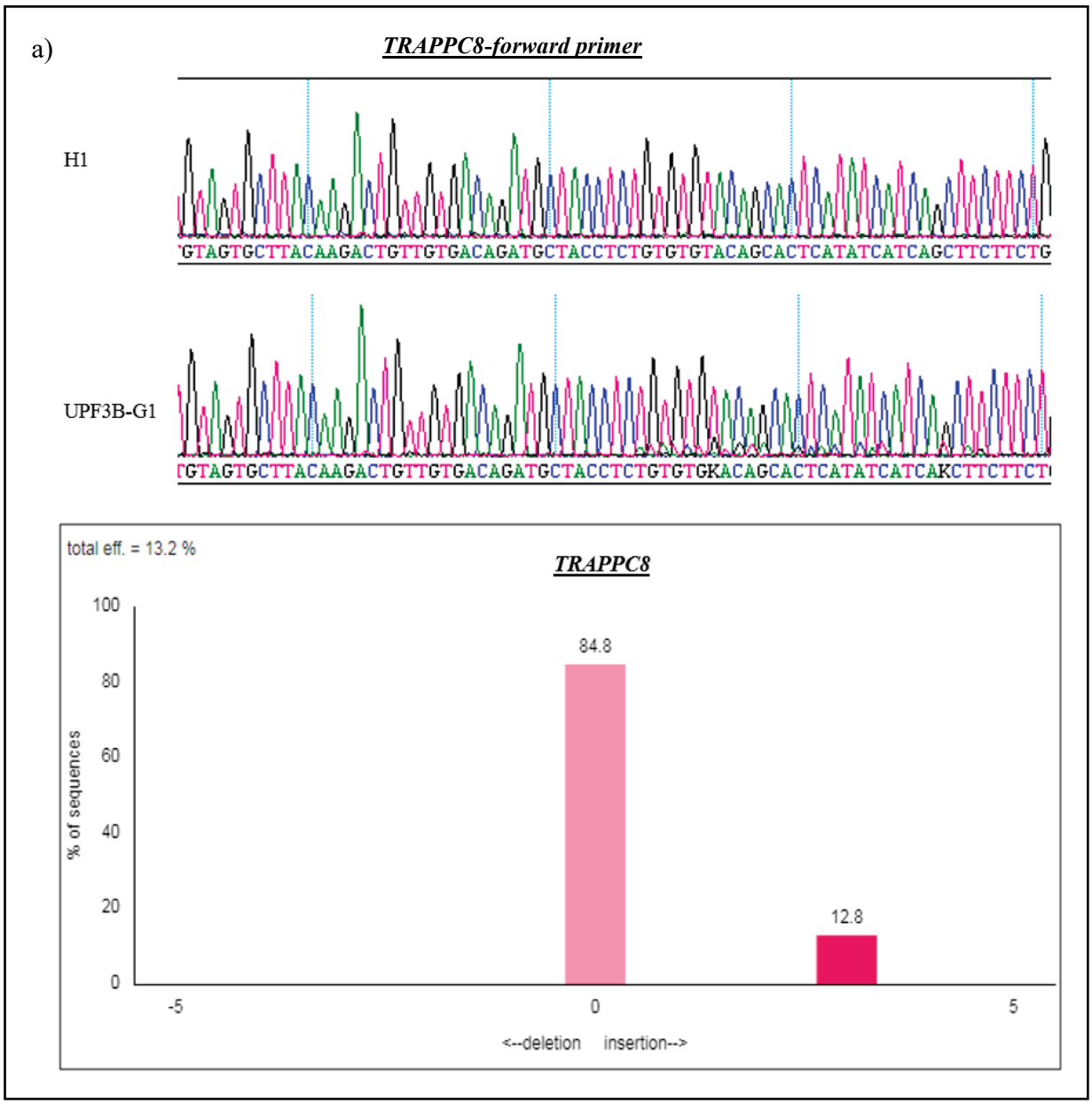


Figure 4.4: Validation of high quality variant in the WGS analysis.

High quality variants were selected and validated through Sanger sequencing. *UF3B-G2* had a heterozygous mutation of the *NCOR2* gene. TIDE revealed that the heterozygous mutation had a wild type allele and an allele with a 1 bp deletion.

a)


TRAPPC8-forward primer
H1
GTAGTGCTTACAAGACTGTTGTGACAGATGCTACCTCTGTGTGTACAGCACTCATATCATCAGCTTCTTCTG
UPF3B-G1
GTAGTGCTTACAAGACTGTTGTGACAGATGCTACCTCTGTGTGKACAGCACTCATATCATCAKCTTCTTCT
total eff. = 13.2 %
TRAPPC8
100
80
60
40
20
0
% of sequences
84.8
12.8
-5
0
5
<--deletion insertion-->

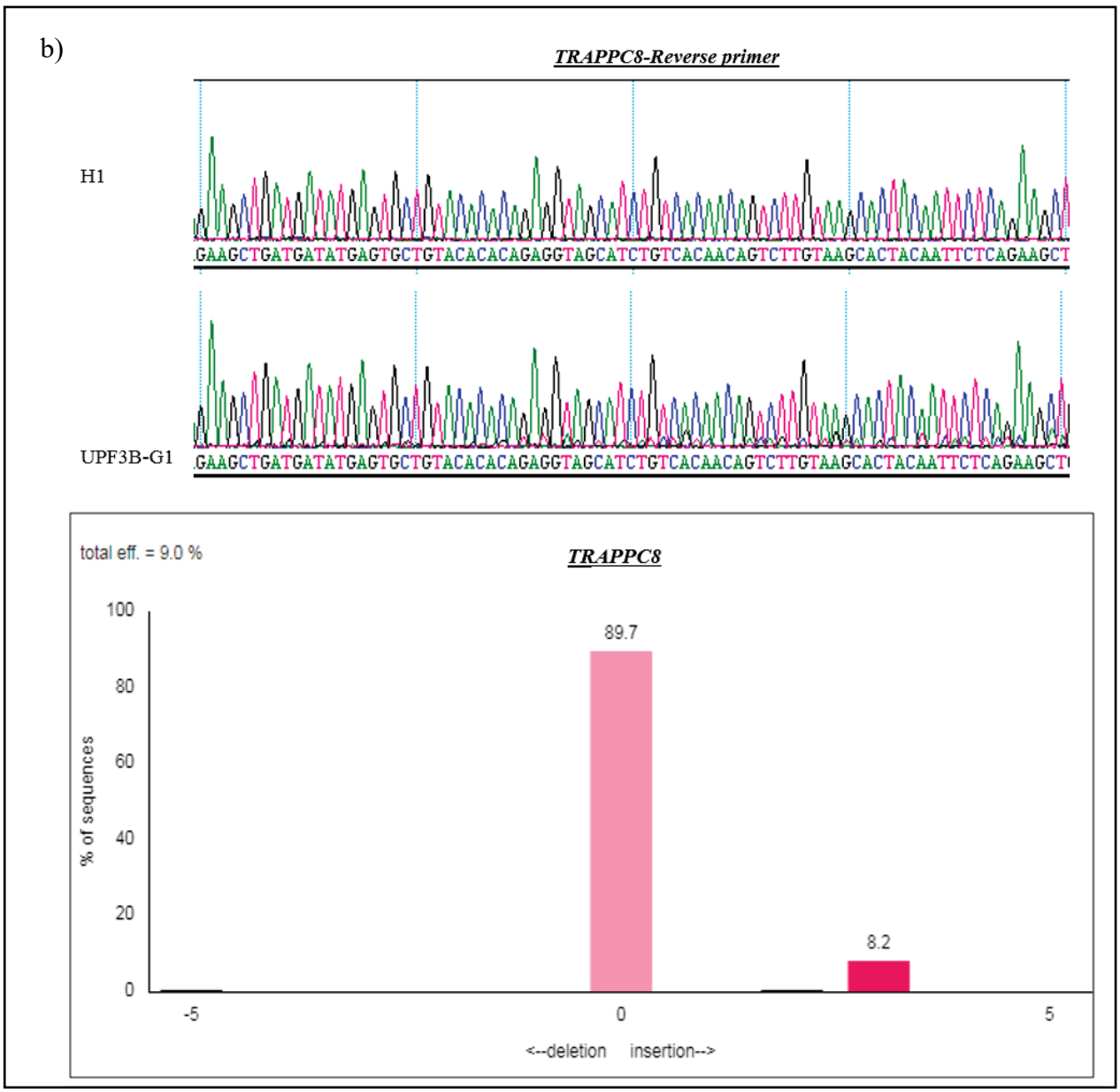


Figure 4.5: Validation of low quality variants in the WGS analysis.

Low quality variants were reanalysed. Mosaicism of *TRAPPC8* was discovered in one of the clones (*UPF3B-G1*) at a frequency of ~10.5% of the cells population. Sanger sequencing was performed using the a) forward and b) reverse primers. TIDE was used to determine the abundance of each mutation in the cell population.

4.2.6 Relatedness of the CRISPR gene-edited clones and the parental cell line

To ensure the KO clones had not genetically drifted apart from each other or the parental cell line due to the number of variants detected in each individual clone, an analysis was performed to compare how similar the gene-edited clones were to the parental cell line and an external published H1 clone (Yazdi, Puya G et al. 2015). The analysis used to determine the relatedness between the cell lines was the *vcftools* relatedness2 algorithm. The matrix (**Table 4-3**) shows how samples are related to each other and the closer to 0.5 the closer the samples are to being identical. The value in each CRISPR clone confirms that the gene-edited clones, the parental cell line and an external published H1 cell line are all similar with minor variation between them.

	H1 (In house)	PX2	UPF3A-G1	UPF3A-G2	UPF3A-G3	UPF3B-G1	UPF3B-G2	UPF3B-G3	H1 (External lab)
H1 (In house)	0.5	0.478767	0.479134	0.479316	0.478714	0.477299	0.477318	0.477871	0.466535
PX2	0.478767	0.5	0.478824	0.478883	0.47847	0.477387	0.477529	0.477837	0.466145
UPF3A-G1	0.479134	0.478824	0.5	0.479408	0.478776	0.477335	0.477375	0.477882	0.466858
UPF3A-G2	0.479316	0.478883	0.479408	0.5	0.478744	0.47662	0.476607	0.477379	0.467378
UPF3A-G3	0.478714	0.47847	0.478776	0.478744	0.5	0.477461	0.477424	0.478122	0.466981
UPF3B-G1	0.477299	0.477387	0.477335	0.47662	0.477461	0.5	0.477374	0.477554	0.464876
UPF3B-G2	0.477318	0.477529	0.477375	0.476607	0.477424	0.477374	0.5	0.477477	0.465015
UPF3B-G3	0.477871	0.477837	0.477882	0.477379	0.478122	0.477554	0.477477	0.5	0.465642
H1 (External lab)	0.466535	0.466145	0.466858	0.467378	0.466981	0.464876	0.465015	0.465642	0.5

Table 4-3: Relatedness between parental cell line and KO clones.

4.3 Discussion

The CRISPR/Cas9 genome editing technology holds great promise in the field of gene therapy such as disrupting genes that cause cancer (Wang & Sun 2017) and repairing mutated genes that cause genetic disease (Bengtsson et al. 2017). The use of *ex vivo* gene therapy, where cells are corrected before transplantation back into a patient (Crudele & Chamberlain 2018) or *in vivo* gene therapy are emerging therapeutic strategies (Gregory-Evans, Bashar & Tan 2012). However, a major concern with gene therapy and disease modelling using engineered nucleases is their potential to generate off-target mutations (Lockyer 2016; Veres et al. 2014). gRNAs tolerate up to five bp mismatches which may home Cas9 to undesired genomic locations and result in SVs such as insertions, deletions or translocations (Cho et al. 2014).

WGS analysis of the gene-edited cells generated in this study revealed that none of the unique *de novo* SNVs detected in each clone overlapped with the predicted off-target sites, suggesting that CRISPR/Cas9 editing technology did not introduce any detectable off-target mutations (**Table 4-2a**). The detected variants were mainly due to the accumulation of mutations during culturing the hESCs (passaging induced mutations) and/or the process of clonal selection including nucleofection and antibiotic selection which could create a strong selective pressure that may favour cells with genomic alterations that promote cell survival (Peterson & Loring 2014). Compared to a study by Veres et (2014) which detected an average of 193 variants in four gene-edited clones, the number of variants detected in this study were considerable high (detected an average of 3083 variants across six clones). Perhaps one explanation for the higher number of variants detected in this study could be that WGS was performed at a higher passage number (~P20) after CRISPR/Cas9 editing while Veres et al (2014) performed WGS at an earlier passage (P7). Another possible explanation could be that compared to our

analysis, Vere et al., WGS analysis included two TALEN edited clones and three controls (edited clones with wildtype alleles from both the TALEN and CRISPR experiment) and a parental cell line. The use of more samples compared to our studies (9 samples compared to 8) could have resulted in the lower variants they detected. The controls they used in their study are wildtype clones that were nucleofected with the same plasmid as the KO clones. This could have therefore reduced the variants further if the gRNA generated the same indels in the clones. Another study that performed WGS in CRISPR/Cas9-mediated gene-edited iPSCs, identified between 217 and 281 SNVs. This study also included TALEN edited iPSCs in their WGS analysis but did not include a control. WGS analysis was performed immediately after selection at passage 1 (P1) (Smith et al. 2014). The main pitfall of this study/analysis is that the parental cell line that was used (to filter out variants from the gene-edited clones) to analyse the WGS was 56 passages lower than the control and gene-edited cells, and is contributing to a large number of variants being detected in each clone. Another study that analysed the WGS of six CRISPR/Cas9-mediated gene-edited clones compared to the parental iPSCs detected an average of 24 932 SNVs in each clone which is high compared to the variants detected in this study. This could be because they first reprogrammed fibroblasts into iPSCs and used CRISPR/Cas9 to correct the mutation (Lyu et al. 2018). iPSCs are known to acquire genetic mutations during reprogramming (Puri & Nagy 2012). This study however did not mention the passage number when gDNA was extracted for WGS. Despite the number of variants detected in the gene-edited clones in this study, when compared to the parental cell line and an external H1 cell line they were all similar with minor variations as analysed using the *vcftools* relatedness2 algorithm (**Table 4-3**).

As indels and SVs cover almost all the mutations generated by CRISPR/Cas9 (Veres et al. 2014), unique *de novo* SVs from the WGS data were also analysed. Deletions (21-578

bp, with the largest being 13 megabases) and duplications (294-626 bp) detected in the gene-edited cells overlapped with the predicted off-target sites. The validation of these deletions and duplications are in progress to determine if they are real. Large deletions (up to 600 bp) were detected in CRISPR edited mice (Shin, H. Y. et al. 2017) and other cell types such as mouse ESCs, mouse hematopoietic progenitors and human differentiated cell lines revealed deletions extending over kilobases (Kosicki, Tomberg & Bradley 2018). The CRISPR/Cas9 had minimal impact on coding regions outside of the target sites. Most clones had two to three genes affected with *UPF3B-G1* clone having the most genes affected (six genes) and *UPF3B-G3* having no genes affected (**Table 4-1a**). In summary none of the unique *de novo* SNVs variants detected were likely due to off-target mutations generated by CRISPR/Cas9, but rather extended culturing of hESCs, the process of obtaining clonal populations of cells and mosaicism present in the parental cell line. *De novo* SVs (deletions and duplications) over lapping the off-target sites were identified and validation of these variants is in progress.

Chapter 5 - Characterisation of *UPF3A* and *UPF3B* dependent NMD in hESCs.

5.1 Introduction

NMD is a posttranscriptional mRNA quality control pathway that identifies, targets and degrades mRNA that contain PTCs within their coding regions (Smith & Baker 2015). Approximately 12% of genetic disorders are caused by mutations that generate PTCs in the respective genes (Mort, Matthew et al. 2008). NMD targets the majority, but not all of such PTC-containing mRNAs. In addition to the PTC containing transcripts, NMD also acts on a range of physiological targets with 'endogenous' PTC-like features introduced by means of alternative splicing, intron retention, uORF, intronic sequences within 3' UTR and lengthy 3'UTRs (>1.5 Kbp) among others. Human mutations and CNVs in NMD and EJC factors which compromise NMD activity leads to unfavourable neurodevelopmental outcomes, reaffirming the importance of the NMD pathway in the regulation of the physiological transcriptome and normal cellular function.

NMD factors, UPF3A and UPF3B both have a short and long isoform and share a highly similar protein sequence and are expressed in the same tissues but not at the same expression level (Tarpey et al. 2007). UPF3B is known to be a strong NMD activator and is important in neurodevelopment while the role of UPF3A has not yet been fully elucidated (**Figure 5.1**). On the path towards determining the molecular and cellular processes disrupted due to loss of *UPF3A* and *UPF3B* during neurodevelopment in a human brain model and the role of *UPF3A* in NMD, *UPF3A* and *UPF3B* KO hESCs were generated (Chapter 3). In this chapter the focus was to characterise *UPF3A* and *UPF3B* KO hESCs. The characterisation determined if the KO alleles produced any functional protein, and if the NMD pathway was disrupted in the hESCs KO clones by analysing NMD targets that are known to be deregulated when *UPF3B* is absent. As depletion of NMD factors are known to impact cell cycle progression and initial stages

of hESC differentiation, analysis of cell cycle kinetics, pluripotency and differentiation of KO hESCs were also assessed.

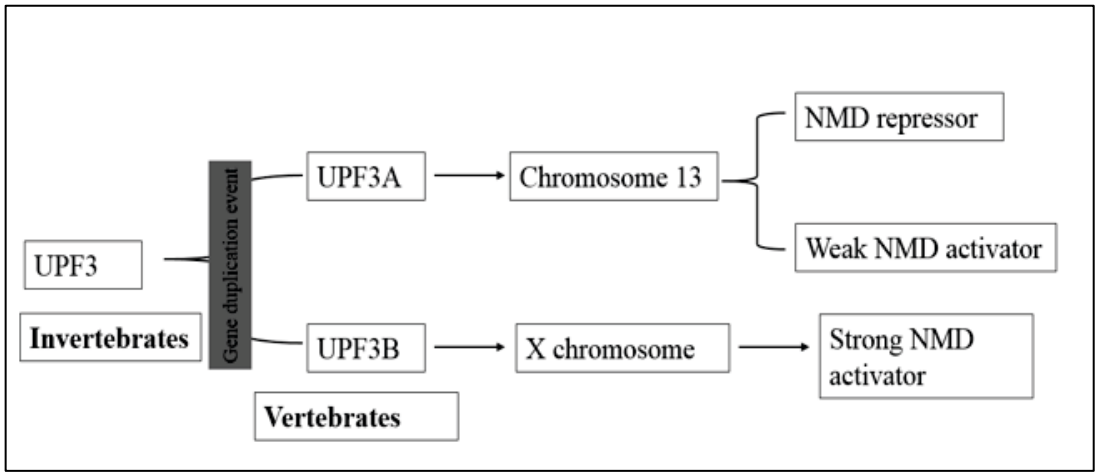


Figure 5.1: Role of UPF3A and UPF3B in NMD

During the divergence of the vertebrate and invertebrate lineage, a gene duplication event occurred that resulted in the emergence of the two gene paralogs *UPF3A* (duplicated) and *UPF3B* (ancestral). *UPF3A* is located on chromosome 13 and its role is uncertain, while *UPF3B* is located on the X chromosome and is a known strong NMD activator.

5.2 Results

5.2.1 Premature termination codons (PTC) were introduced early in the coding regions of *UPF3A* and *UPF3B* KO hESC clones.

The gRNAs used to generate *UPF3A* and *UPF3B* KO hESCs clones were designed to target either exon one or two depending on which gRNA had the highest score from the CRISPR design tool (Chapter 3). Frameshift type mutations in these 5'regions of the gene are most likely to prevent the production of functional proteins. **Table 5-1** shows the three *UPF3A* and three *UPF3B* KO hESCs clones that were selected for analysis (**Figure**

3.12 and **Figure 3.13**) and their mutation annotations. The mutations created in these clones all generated frameshift mutations and introduced PTCs early in their respective ORFs. **Figure 5.2** shows the expected size of proteins from the mutant *UPF3A* and *UPF3B* alleles (a peptide translated from the start codon to the first PTC). The hypothetical truncated proteins are small and predicted non-functional as they lack the essential UPF2 and Y14 binding sites.

Gene	**Clone**	**Mutation annotation**
UPF3A	UPF3A-G1	Allele 1: c.152_153insT
		Allele 2: c.153_156del
	UPF3A-G2	Allele 1: c.155_156del
		Allele 2: c.155_168del
	UPF3A-G3	Allele 1: c.135_141del
		Allele 2: c.138_154del
UPF3B	UPF3B-G1	c.57_75delinsGGGGT
	UPF3B-G2	c.158_159insT
	UPF3B-G3	c.52_53insC

Table 5-1: Mutation annotations of the generated *UPF3A* and *UPF3B* KO clones.

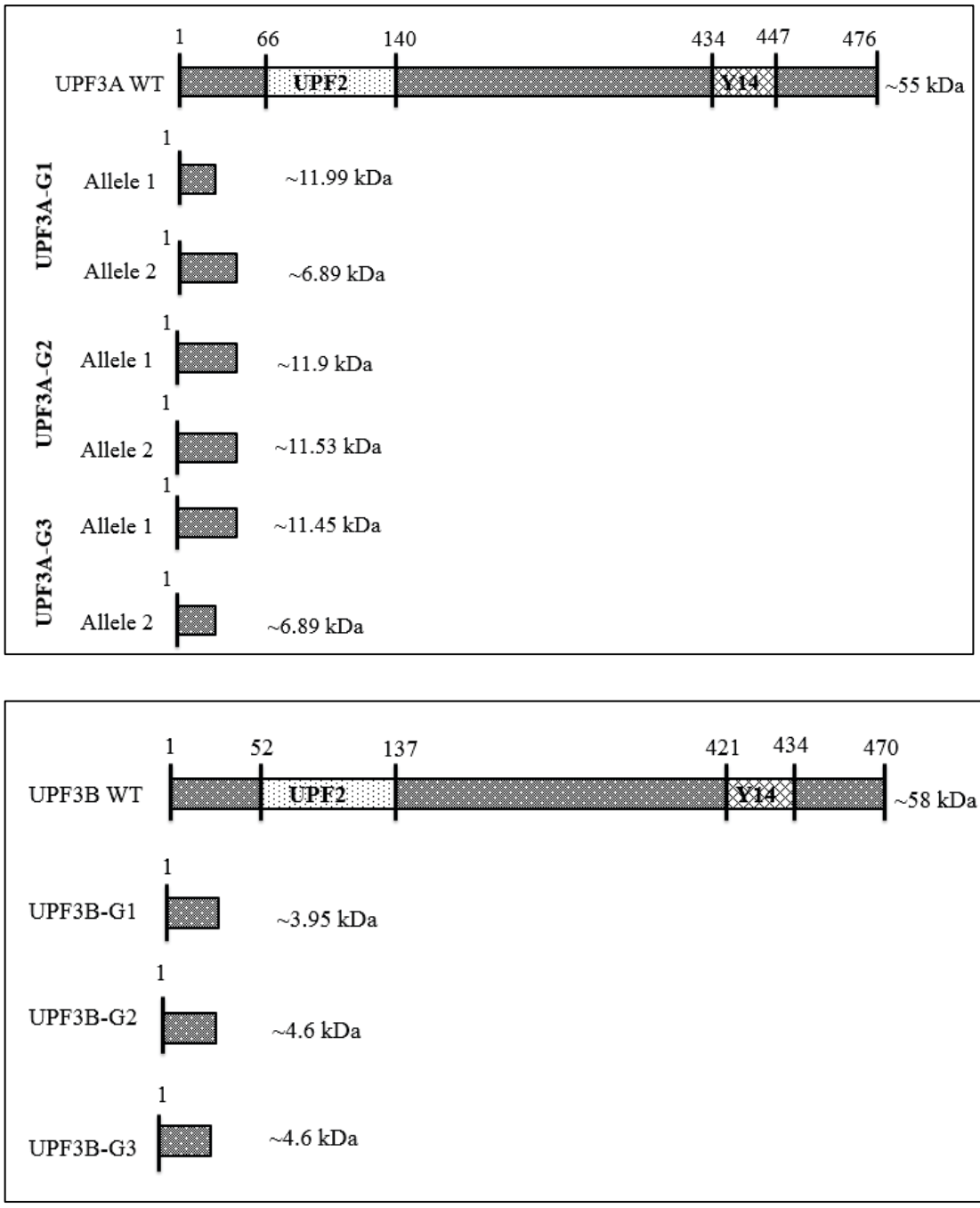


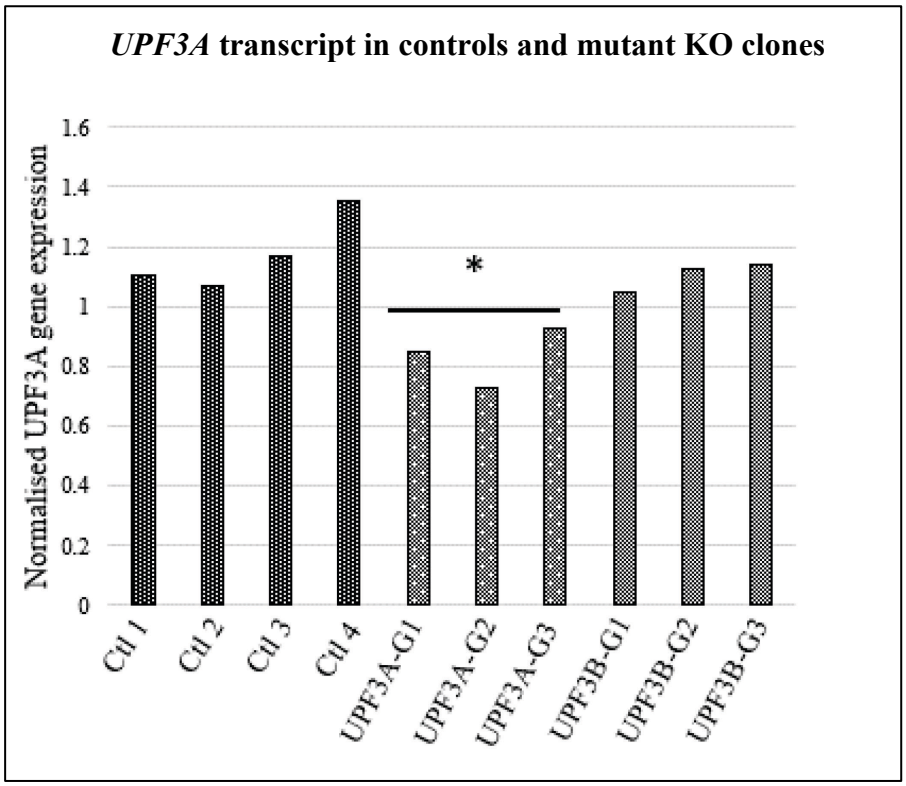
Figure 5.2: Schematic of wildtype and hypothetical UPF3A and UPF3B proteins produced from wildtype and mutant alleles.

UPF3A and UPF3B proteins have two similar binding domains located at the N- and C-terminus, a UPF2 and Y14 binding domain respectively. Numbers above the wildtype schematic represent amino acid residues. The hypothetical proteins produced from the alleles from the three *UPF3A* and three *UPF3B* KO hESC are represented.

5.2.2 *UPF3A* transcript levels are reduced in *UPF3A* KO clones while *UPF3B* transcripts levels are unchanged in *UPF3B* KO clones.

As the indels created in the *UPF3A* and *UPF3B* KO clones introduced PTCs in their respective transcripts, real time qPCR analysis was performed on mRNA extracted from the clones to determine *UPF3A* and *UPF3B* transcript levels. *UPF3A* KO clones had significantly reduced *UPF3A* transcript levels (29.1%) compared to the controls while *UPF3B* transcript levels in *UPF3B* KO clones were slightly upregulated, suggesting they escaped NMD (**Figure 5.3**). PTCs introduced in the *UPF3A* and *UPF3B* KO clones are located in exon 1 (near the initiation codon). These PTCs are predicted to be degraded by NMD pathway as they are located more than 50-55 nts upstream of the last EJC. The 50-55 nt rule states that PTCs located 50-55 nts upstream of the 3'most exon-exon junction are degraded by mRNA. However, they are exceptions to this rule as it has been shown that PTCs located close to the translation initiation codon escape being targeted by the NMD pathway despite being located 50-55 nts upstream of the EJC (Inacio et al. 2004; Perrin-Vidoz et al. 2002) which is observed in the generated *UPF3A* and *UPF3B* KO clones.

a)



b)

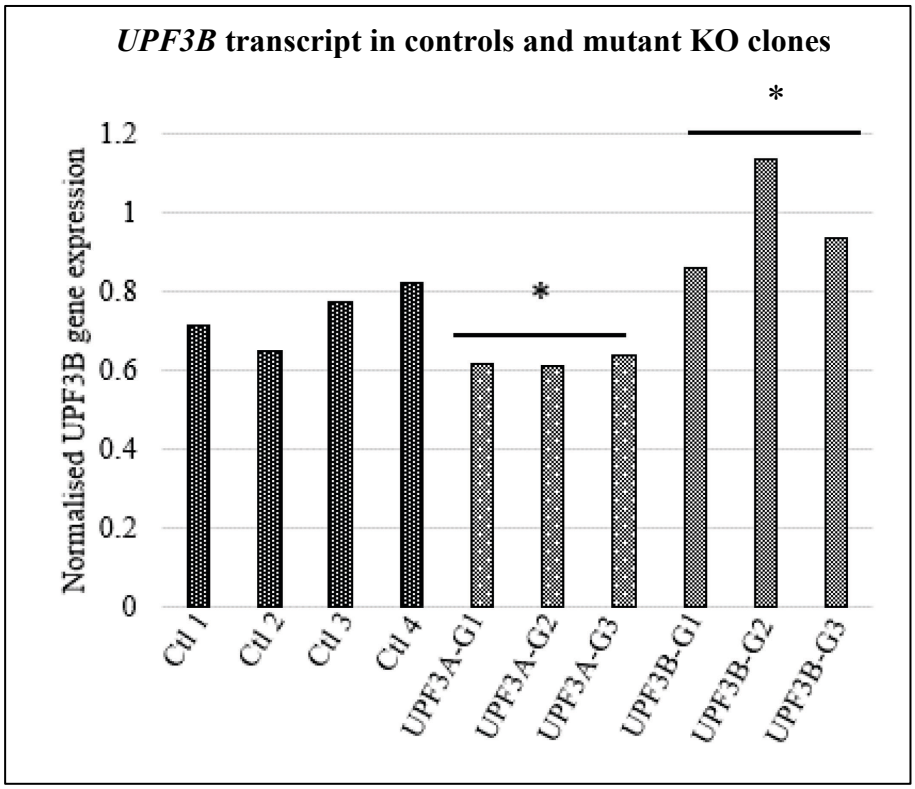


Figure 5.3: *UPF3A* and *UPF3B* transcript levels in *UPF3A* and *UPF3B* KO clones.

Real time qPCR was used to analyse the transcript levels of a) *UPF3A* and b) *UPF3B* in mRNA extracted from controls (n=4), *UPF3A* (n=3) and *UPF3B* (n=3) KO hESCs clones performed in triplicates. All data normalised to *ACTB* mRNA expression. Graphs represent mean value for each genotype in each triplicate. *$p<0.05$ significantly different to controls by Student T-test assuming equal variance.

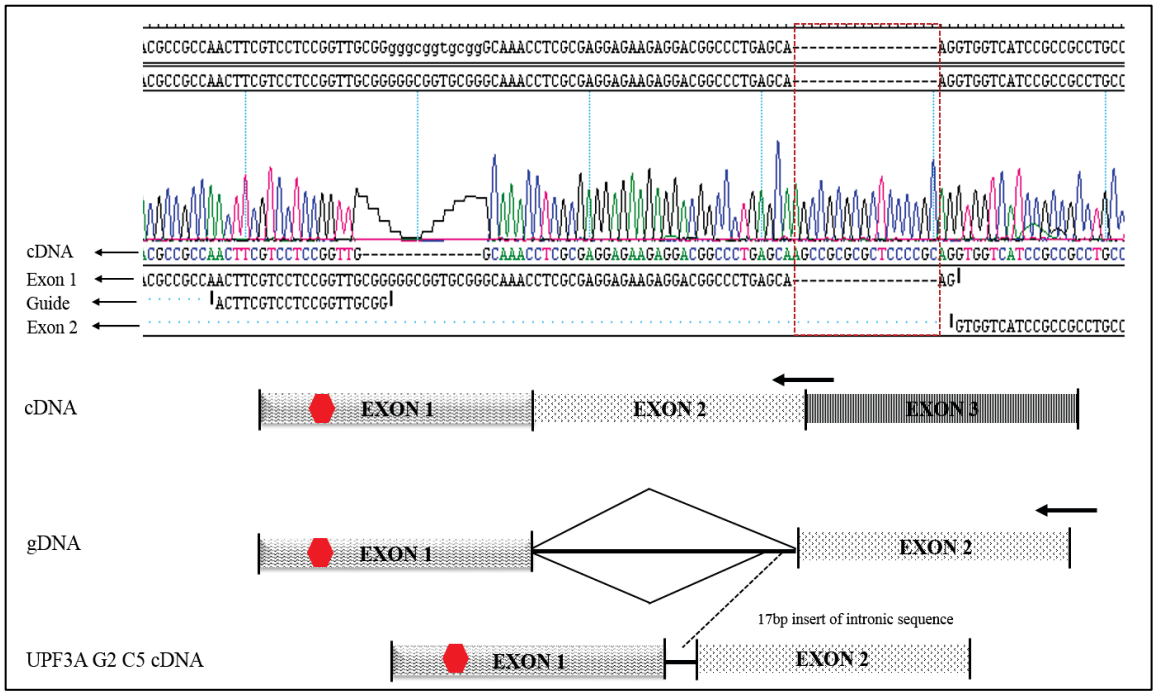
5.2.3 mRNA expressed from *UPF3A* KO alleles display usage of cryptic splicing sites.

UPF3A KO clones were selected based on the results from the TIDE software which identifies the number of indels generated using Sanger sequencing of regions spanning across the target gRNA site (**Figure 3.13**). However, this software does not give any information regarding the actual position of the mutation on each allele. To determine the position of the mutation on each allele in the *UPF3A* KO clones, a PCR was performed using cDNA isolated from each clone as a template. The PCR spanned the region transcribed from the genomic site targeted by the gRNA. The PCR products were then ligated into the pGEM-T Easy vector and transformed into bacteria.

Ten bacterial colonies were selected for each *UPF3A* KO clone. The plasmid DNA was isolated and the inserts sequenced. From each clone, half of the isolated plasmids had sequences that aligned with one allele of the hESC clone, and half from the other allele suggesting bi-allelic expression of the alleles in hESCs. The reason for observing an equal ratio of each allele could be due to only 29.1% of the transcript being degraded compared to controls. However, in addition to the two sequences detected (from each allele) in each clone, *UPF3A-G2* and *UPF3B-G3* KO clones had a third sequence detected, suggesting a third mRNA species was being produced (occurred in 1:10 of the selected bacterial clones) (**Figure 5.4a** and **Figure 5.4b**). Sequences of the third mRNA species was generated from one of the two alleles as allelic mutation was clearly present (**Table 5-1** and **Figure 5.4a-b**). This sequencing revealed that this mRNA species reflected an incorrectly spliced product, harbouring an additional 17 bp and 110 bp from intron one sequence in *UPF3A-G2* and *UPF3A-G3* clones respectively. The aberrant splicing in *UPF3A* KO clones (G2 and G3) created an in-frame mutation in addition to the frameshift

mutations introduced by the CRISPR/Cas9 technology. This in-frame mutation can potentially produce a near full length protein, but with a partial deletion and additional 'intronic sequence derived' amino acids.

a) UPF3A-G2 allele 2



b) UPF3A-G3 allele 1

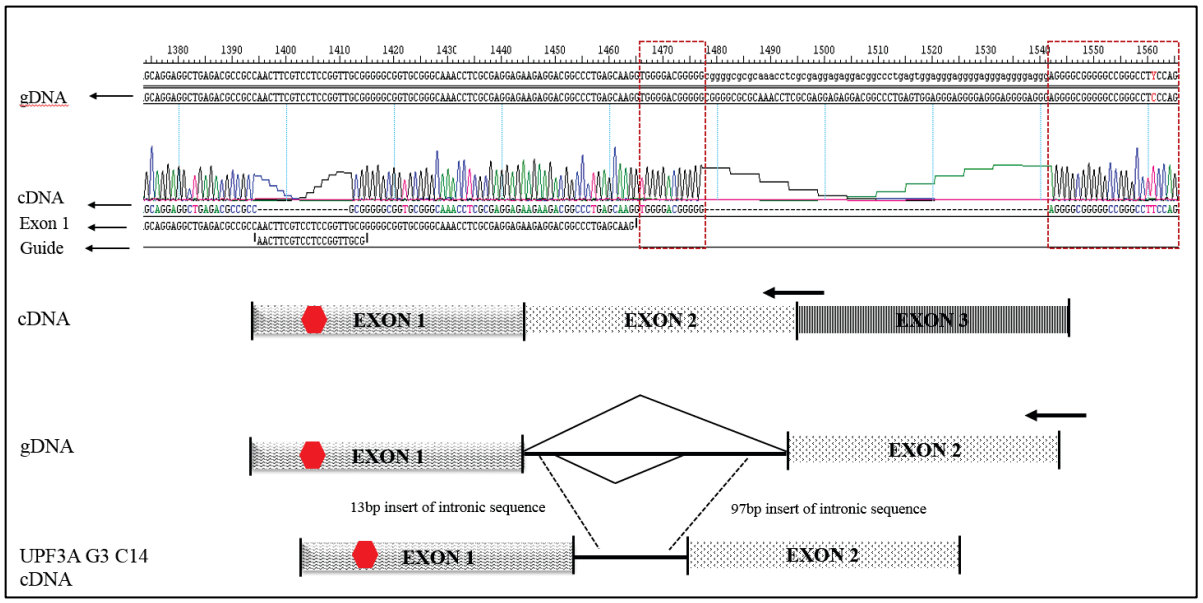


Figure 5.4: Aberrant splicing in *UPF3A* KO hESCs clones.

Sanger sequencing performed on bacterial cell colonies that were transformed with the PCR product from a) *UPF3A-G2* and b) *UPF3A-G3*. In *UPF3A-G2* and *UPF3A-G3* clones an additional sequence was detected that had a section from intron one sequence.

5.2.4 Expression from the *UPF3A* and *UPF3B* KO alleles did not produce protein.

The ultimate measure to determine if the mutant alleles truly represented null alleles and as such confirming the generation of KO hESCs clones was to determine if any protein was being expressed. hESCs were harvested and protein lysates extracted and subjected to western blot analysis. UPF3A and UPF3B protein were absent from lysates from *UPF3A* and *UPF3B* KO hESCs clones respectively (**Figure 5.5a** and **Figure 5.5b**). Western blot analysis, even at a high exposure of the membrane did not detect any UPF3A protein (**Figure 5.5a**). This data thus confirm that the mutant alleles generated by CRISPR/Cas9 targeted disruption of *UPF3A* and *UPF3B* are truly null alleles unable to produce functional protein, and as such confirming the generation of *UPF3A* and *UPF3B* KO hESC clones.

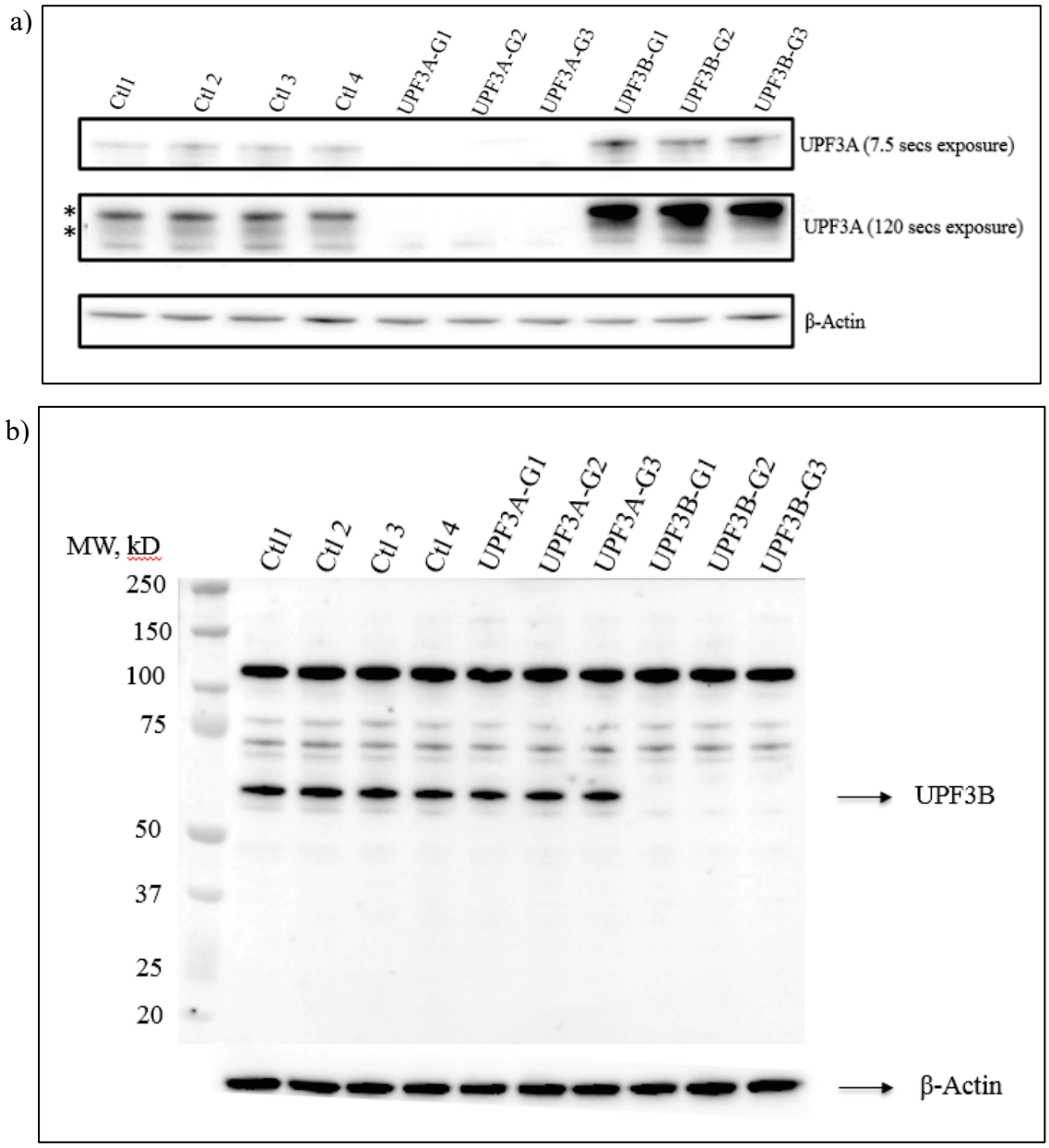


Figure 5.5: *UPF3A* and *UPF3B* KO clones had no protein expression.

Protein lysates from controls (n=4), *UPF3A* (n=3) and *UPF3B* (n=3) KO clones were harvested and subject to western blot analysis. Membranes were probed with the a) UPF3A antibody and b) UPF3B antibody. Asterisk marks on the UPF3A western blot indicates the two isoforms of UPF3A and the lower band which appears in all clones is non-specific. Bands that appear in all clones with the same intensity are non-specific in the UPF3B western blot.

5.2.5 UPF3A protein is stabilised in *UPF3B* KO clones

Western blot analysis of the UPF3A protein in *UPF3B* KO clones indicated that UPF3A protein was stabilised in the absence of the UPF3B protein (**Figure 5.6a** and **b**). This phenomena has been previously described as evidence of a compensatory NMD mechanism in other cell types (Chan et al. 2009; Jolly et al. 2013; Nguyen et al. 2012). UPF3A and UPF3B are highly similar proteins that bind to UPF2 using their N-terminal domain (Kunz et al. 2006; Lykke-Andersen, Shu & Steitz 2000). The UPF2 binding domain of UPF3A is important in stabilisation of this protein. UPF3A and UPF3B both compete to bind UPF2. When UPF3B is present, most of the available UPF2 protein is bound to UPF3B and unbound UPF3A is degraded. However when UPF3B protein is absent, UPF3A is able to bind to UPF2 and is stabilised (Chan et al. 2009).

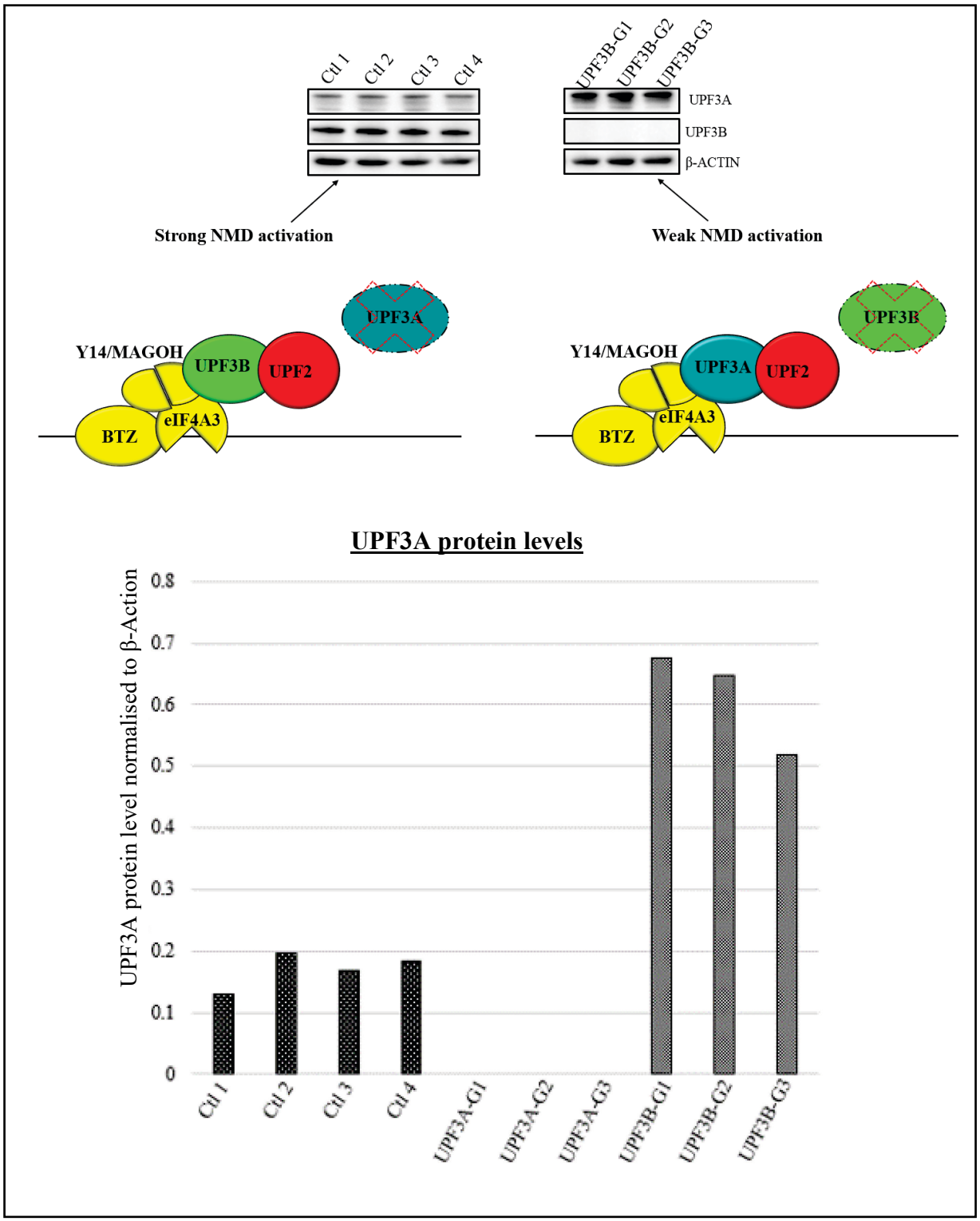


Figure 5.6: UPF3A protein is stabilised in *UPF3B* KO hESCs clones.

a) When UPF3B is present it binds to UPF2 and UPF3A protein is degraded and vice versa. Protein lysates were extracted from the controls (n=4) and *UPF3B* KO clones (n=3) and subject to western blot analysis. Membranes were probed with the UPF3A and UPF3B antibody. b) Densitometry was performed directly on image captured with low exposure time and normalised against the level of the β-Actin protein in the same sample. Red cross indicates absence of the protein.

5.2.6 Evidence of an NMD negative feedback regulatory network in hESCs

NMD has a dual role in regulating gene expression by degrading transcripts with PTCs generated from germline, somatic mutations, transcription errors and physiological transcripts from genes that contain NMD inducing features such as an uORF, long 3'-UTR (>1.5 Kbp) (Boehm et al. 2014; Nicholson et al. 2010). Some of these endogenous NMD features are also present in transcripts which encode NMD factors themselves (**Table 1-1**). As a result when NMD is perturbed, the NMD factors are known to be upregulated suggesting a self-buffering mechanism exists to maintain NMD magnitude (Huang, L et al. 2011).

To discover evidence of this buffering mechanism in *UPF3A* and *UPF3B* KO clones, real time qPCR was performed on mRNA isolated from the clones to test the expression of NMD and EJC genes. The *UPF1* and *UPF2* transcript levels were significantly upregulated in *UPF3B* KO clones and unaffected in *UPF3A* KO clones. *SMG7* and *MAGOH* expression was downregulated in *UPF3A* KO clones while *UPF3B* KO clones only showed a reduction in *MAGOH* (**Figure 5.7a)**. Western blot analysis showed there was no significant difference between *UPF3A*, *UPF3B* KO clones and controls for the probed NMD and EJC factor proteins (data not shown) (**Figure 5.7b**).

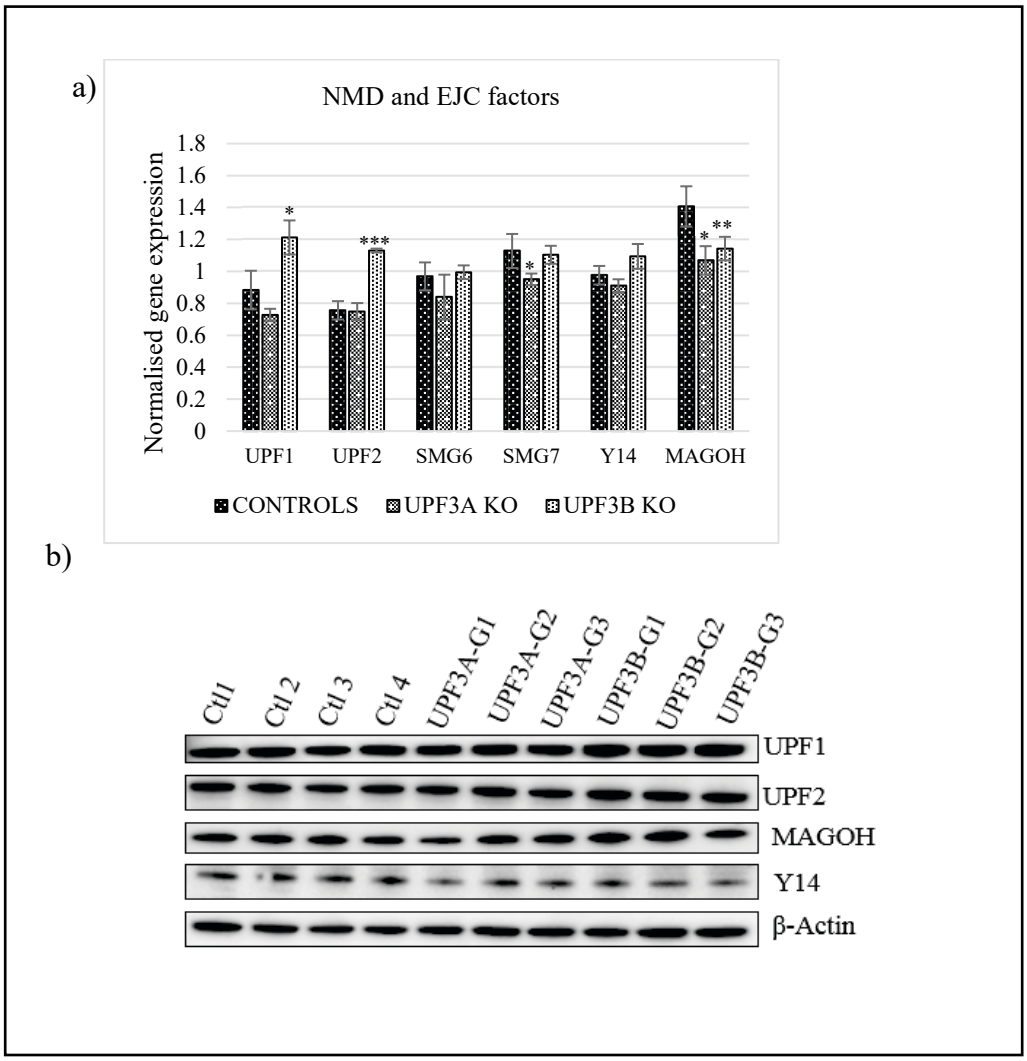


Figure 5.7: Loss of *UPF3B* and not *UPF3A* is involved in the buffering mechanism.

Analysis of mRNA and protein expression of genes encoding NMD and EJC factors. mRNA and protein lysates was isolated from n=4 controls and n=3 *UPF3A* and *UPF3B* KO clonal hESCs and analysed separately. a) Real time analysis of NMD and EJC factor mRNA. Real time analysis was performed in triplicates per sample. All data normalised to *ACTB* mRNA expression. Graphs represent as mean values of each genotype, error bars represent standard deviation. *p<0.05, **p<0.01, ***p<0.001 represent statistical difference compared to controls by Student-T test assuming equal variances. b) Western blot analysis of EJC and NMD factors on protein lysates. β-Actin serves as a loading control.

5.2.7 NMD 'targeted' transcripts are deregulated in response to *UPF3B* deletion.

Following the successful generation of *UPF3A* and *UPF3B* KO hESCs clones, evidence of altered NMD was sought, by testing the mRNA expression of a suite of endogenous NMD target mRNAs. mRNA was isolated from *UPF3A* and *UPF3B* KO clones and controls and the levels of NMD targeted mRNA sensitive to both the classical (*UPF3B* sensitive) and the alternate NMD (*UPF3B* in-sensitive) pathways were analysed. The mRNA expression of the classical NMD target genes *GAS5* and *ATF4* were upregulated in *UPF3B* KO clones, but not affected in *UPF3A* KO clones. Surprisingly, the mRNA of other NMD classical targets *cJUN*, *GADD45B* and *ATF3* were significantly downregulated in both the *UPF3A* and *UPF3B* KO hESCs (**Figure 5.8a**). Intriguingly these genes all encode stress response proteins (Inoue et al. 2018; Jovaisaite, Mouchiroud & Auwerx 2014; Liebermann & Hoffman 2008). Alternative NMD pathway targeted mRNAs encoded by *PANK2* and *SMG5* were not affected in either *UPF3A* and *UPF3B* KO clones (**Figure 5.8b**).

A second set of putative NMD-targeted mRNA for testing was selected from RNA Seq transcriptome data generated from an iPSC derived from a male individual with complete loss-of-function mutations in *UPF3B* (unpublished, provided by Prof. Miles Wilkinson, University of California, San Diego, CA, USA). Four deregulated transcripts (encoded by *HIST1H2AC*, *ANXA1*, *HNMT*, *AUTS2*) were selected and their expression analysed using real time qPCR performed on mRNA isolated from the *UPF3A* and *UPF3B* KO clones (**Figure 5.8c**). Three of the deregulated transcripts were validated in *UPF3B* KO clones (*HIST1H2AC*, *ANXA1*, *HNMT)* whilst *AUTS2* was unchanged. This analysis was not a robust approach because a small number of transcripts (12 transcripts) were selected to determine the effects of loss of *UPF3A* and *UPF3B* in hESCs. As such the

identification of transcriptome wide changes in these cells were subsequently performed by RNA seq (Chapter 6).

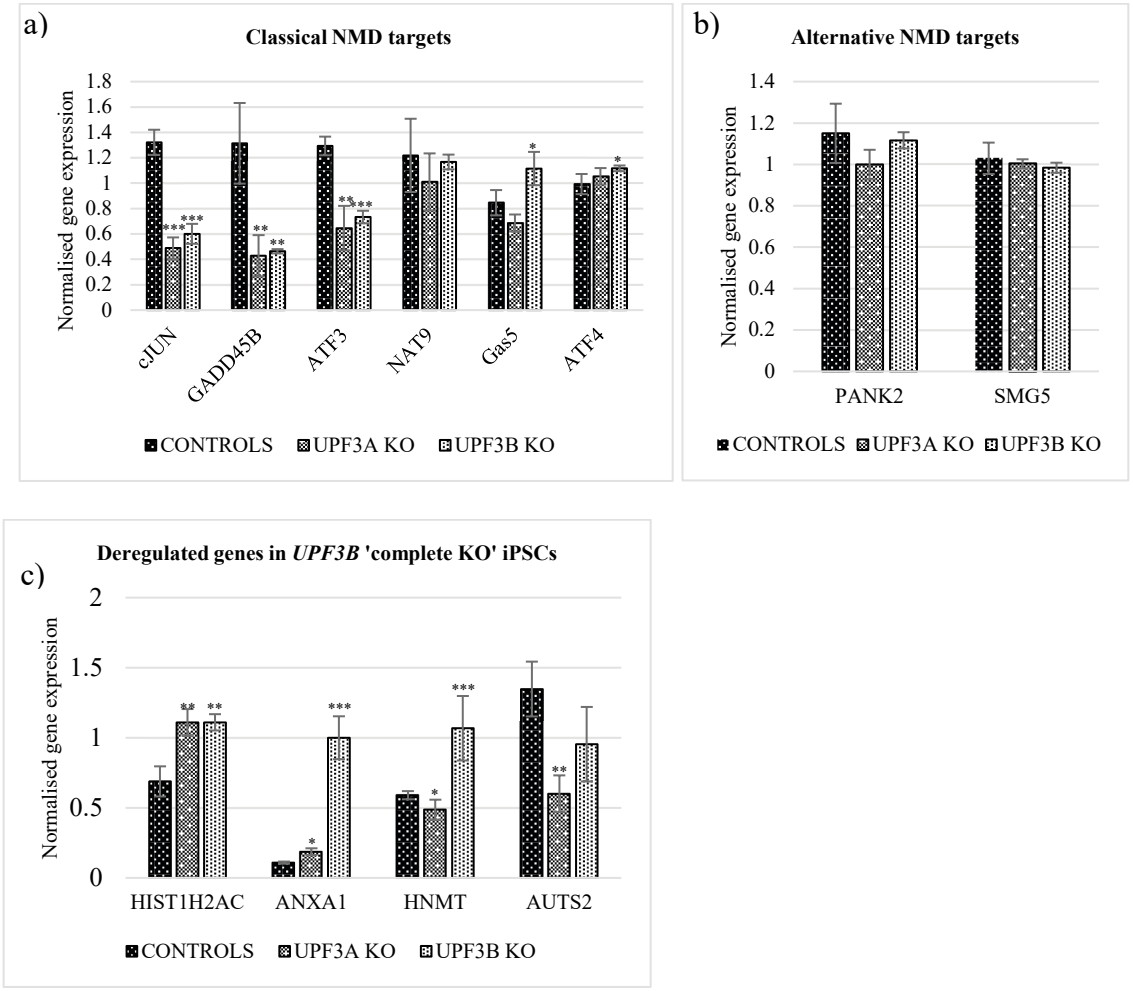


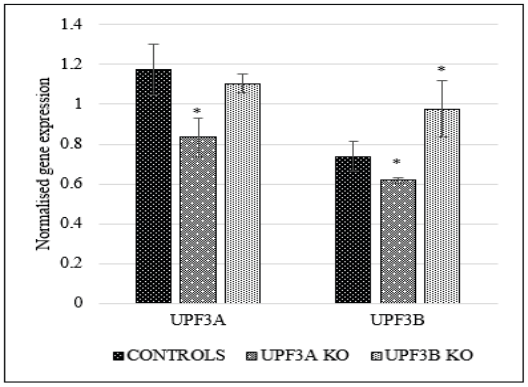
Figure 5.8: Deregulation of NMD targets in response to loss of *UPF3A* and *UPF3B*.

Real time qPCR analysis of mRNA known to be subjected to NMD in *UPF3A* and *UPF3B* KO hESCs compared to wildtype controls. a) Analysis of mRNAs targeted by the classical NMD pathway. b) Analysis of mRNAs targeted by the alternative NMD pathway. c) Analysis of mRNAs deregulated in *UPF3B* 'complete knockout' iPSCs. n=4 independent control clonal lines compared to n=3 independent *UPF3A* KO and n=3 *UPF3B* KO hESC lines done in triplicate. All data normalised to *ACTB* mRNA expression. Graphs represent mean value for each genotype, error bars represent standard deviation. *p<0.05, **p<0.001, ***p<0.0001 significantly different to controls by student T-test assuming equal variance.

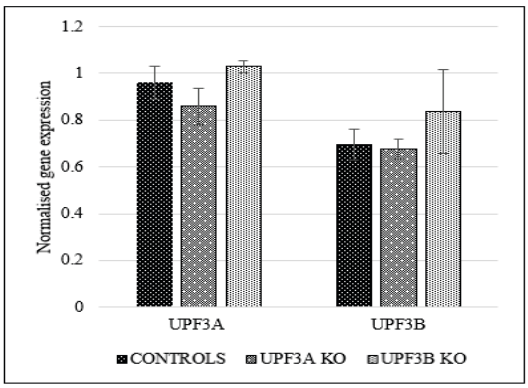
5.2.8 Cell culture environment /passage number have an impact on NMD activity variability.

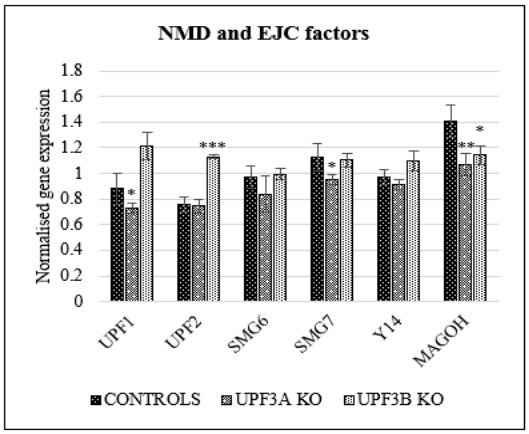
To investigate the unexpected results where the stress response transcripts were downregulated in both *UPF3A* and *UPF3B* KO hESC clones (section 5.2.7), cell pellets were collected at an earlier passage number, approximately a 10 passage difference between collections periods and mRNA extracted. Bona fide NMD targets and stress response transcripts were analysed. The bona fide NMD targets (*GAS5*, *ATF4*) and the NMD factors involved in the buffering feedback loop (*UPF1*, *UPF2*) were upregulated in both the high and low passage conditions suggesting that these are true and robust effects of a perturbed NMD pathway. However genes involved in the stress response (*cJUN, GADD45B, ATF3*) transcript levels changed between the low and high passage (**Figure 5.9**). In the earlier passage these transcripts were unaffected, suggesting that the use of NMD targeted genes that are also involved in the stress response pathways are not ideal indicators of NMD activity. For RNA Seq analysis and NSC differentiations cells that were at a lower passage where the stress response genes were not affected were used.

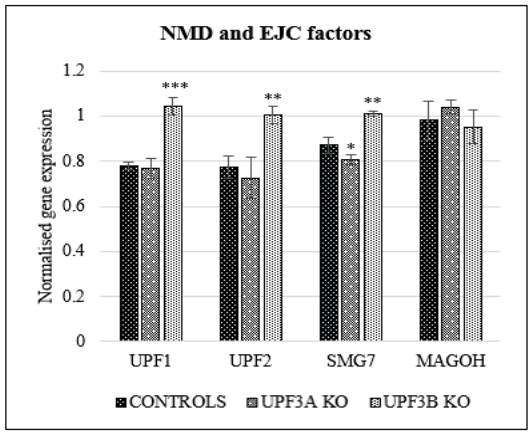
Late Passage

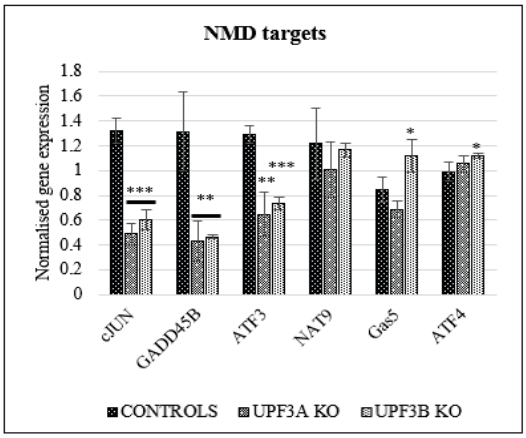


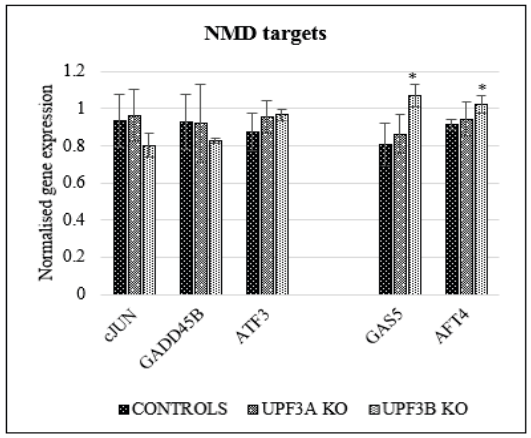
Early Passage

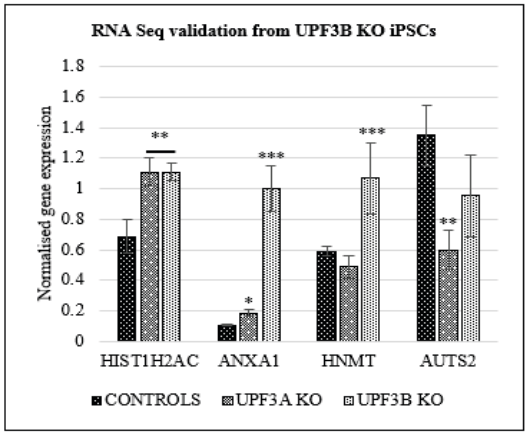












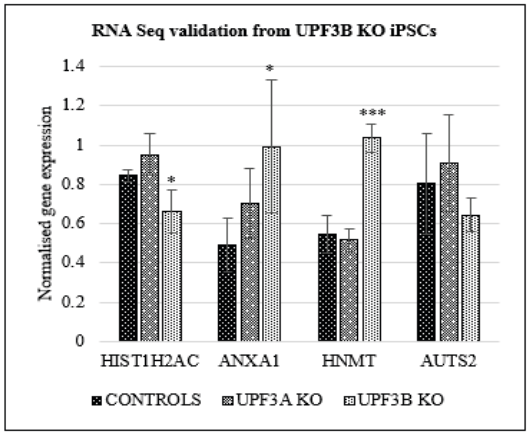


Figure 5.9: Impact of cell culture environment/passage number on NMD activity.

Real time qPCR was performed on mRNA's known to be regulated by NMD or are bona fide NMD targets in early and late passages of cell from controls (n=4), *UPF3A* KO hESCs clones (n=3) and *UPF3B* KO hESCs clones (n=3), performed in triplicates. All data normalised to *ACTB* mRNA expression. Graphs are represented as mean value for each genotype, error bars represent standard deviation. *p<0.05, **p<0.01, ***p<0.001 significantly different to controls by student T-test assuming equal variance.

5.2.9 NMD and the unfolded protein response (UPR) stress response pathway

As NMD is important in degrading transcripts with PTCs to prevent truncated proteins to be produced, when NMD is perturbed these proteins can be made and may lead to the production of misfolded proteins. This in turn can lead to an unfolded protein response, (UPR) which is a stress response pathway activated by endoplasmic reticulum (ER) stress. This pathway helps cells to adapt to the excess levels of misfolded proteins (**Figure 5.10**) (Goetz & Wilkinson 2017).

To investigate if loss of *UPF3A* and *UPF3B* lead to activation of the UPR stress response pathway, three different mRNAs which encode UPR 'sensors' (i.e. normally upregulated by activation of the UPR pathway) were analysed using real time qPCR. Compared to controls, *UPF3A* and *UPF3B* KO clones had a reduction in *EIF2AK3* and *ERN2* transcripts, however this was not significant. *UPF3B* KO clones had a slight insignificant increase in *ATF6* transcript levels. *DDIT3/CHOP* (a downstream target of ATF4) was significantly downregulated in *UPF3A* KO clones (**Figure 5.11**). These results suggest that the UPR pathway is not significantly activated by loss of *UPF3A* and *UPF3B* in hESCs.

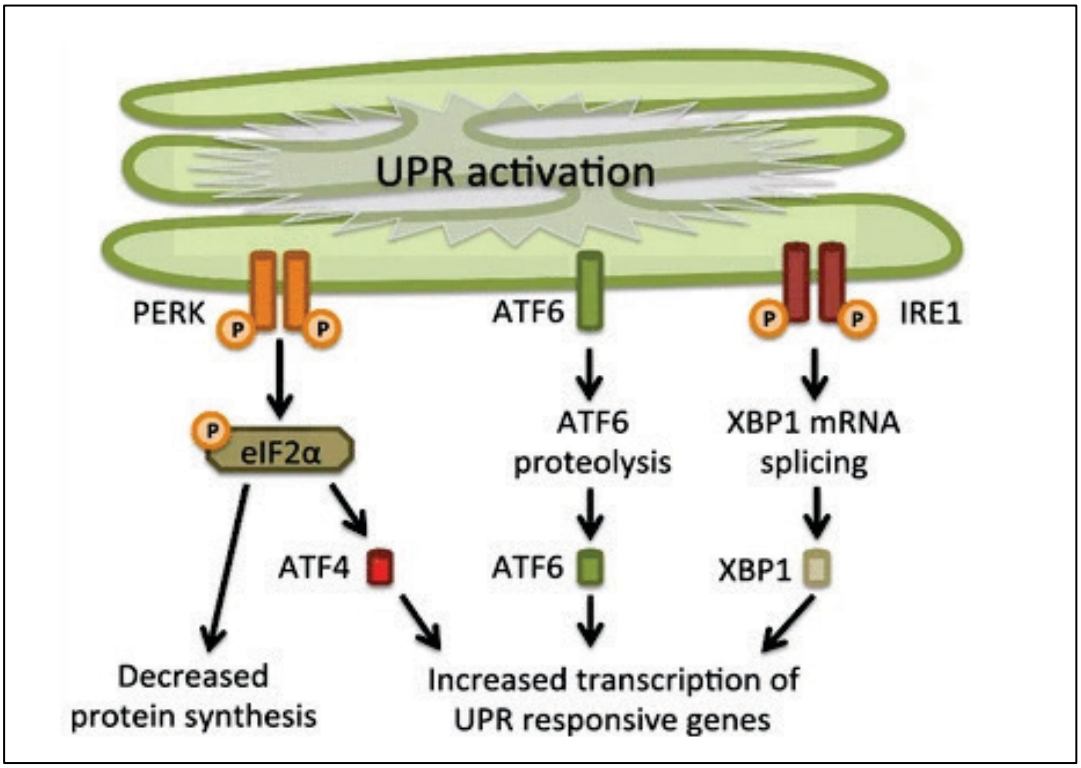


Figure 5.10: The UPR pathway

ER stress leads to the activation of the UPR pathway. This pathway has three different branches which have different stress sensors namely, PERK, ATF6 and IRE1. The activation of these different branches is important to allow cells to adapt to the high levels of unfolded proteins and return the cells to normal homeostasis. Picture taken from (Scheper & Hoozemans 2015).

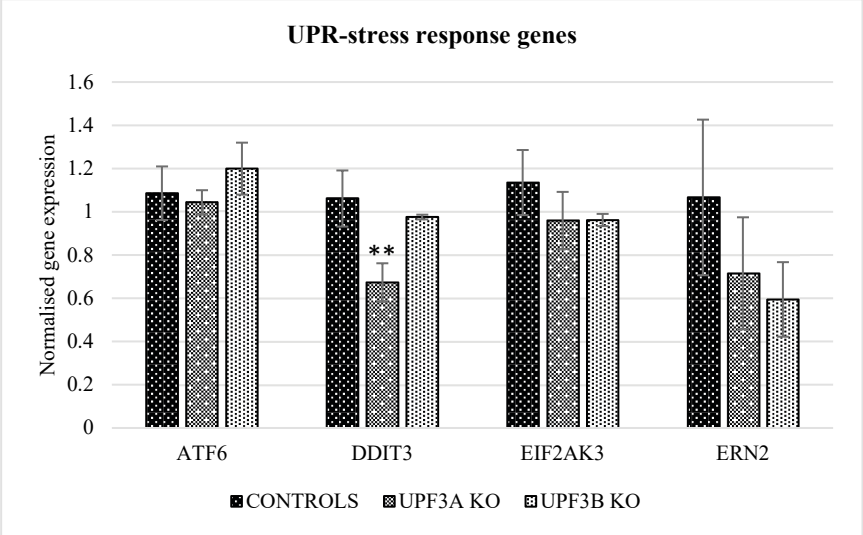


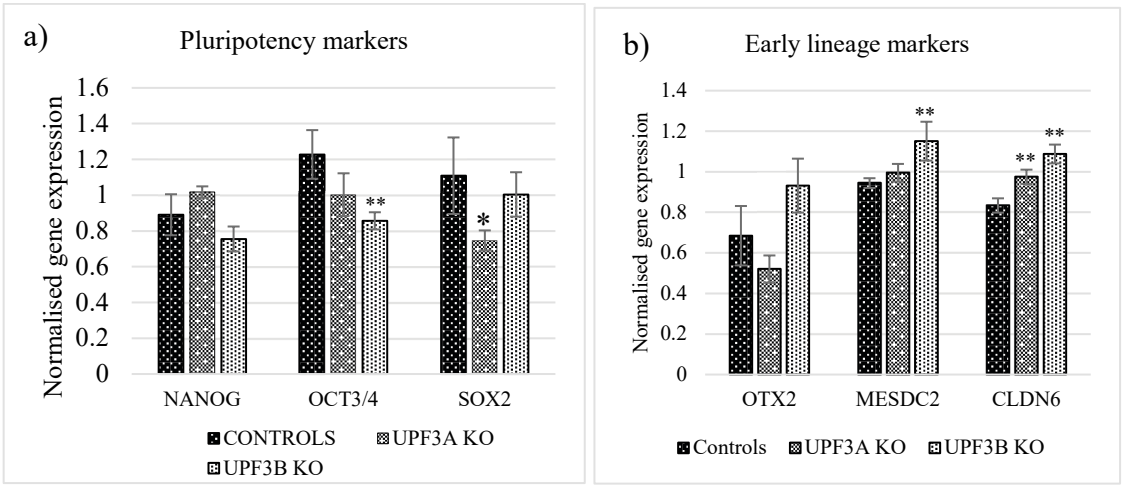
Figure 5.11: UPR-stress response genes.

Real time qPCR analysis of mRNA of the UPR transcripts sensors and targets in controls (n=4), *UPF3A* (n=3) and *UPF3B* (n=3) KO clones, performed in triplicate. All data is normalised to *ACTB* mRNA expression. Graphs are represented as mean and STDEV of number of clones. **p<0.01 significantly different to controls by student T-test assuming equal variance.

5.2.10 Deletion of *UPF3A* or *UPF3B* lead to the initial stages of endoderm and mesoderm differentiation.

NMD factors are highly expressed in human pluripotent stem cells and their levels decrease upon loss of pluripotency (Alrahbeni et al. 2015; Bruno et al. 2011; Cho et al. 2012; Gong et al. 2009; Lou et al. 2016). Inhibition of NMD via *UPF1* and *UPF3B* knockdown in P19 and H9 cells led to a reduction in the expression of the stem cell markers, - NANOG and OCT3/4, and triggered the initial stages of endodermal differentiation (Lou et al. 2016; Lou et al. 2014). To determine if loss of the NMD factor genes *UPF3A* and *UPF3B* also have a role in pluripotency maintenance in H1 hESCs, the expression of a suite of well characterised pluripotency marker genes *NANOG*, *OCT3/4* and *SOX2* were determined in cultures of *UPF3A* and *UPF3B* KO hESCs.

In the first instance, real time qPCR was employed to determine mRNA expression levels. Loss of *UPF3A* or *UPF3B* was associated with significant reductions in transcript levels of *SOX2* and *OCT3/4* respectively, suggesting that the loss of *UPF3A* and *UPF3B* in hESCs triggered initial stages of differentiation (**Figure 5.12a**). To extend this analysis, the expression of early differentiation markers were assessed. Indeed loss of *UPF3B* leads to an increase in transcript levels of *MESDC2*, an early mesodermal marker (Laco et al. 2018; Veltman et al. 2005) and *CLDN6*, an early endodermal marker (Anderson et al. 2008), while loss of *UPF3A* lead to an upregulation of *CLDN6* only (**Figure 5.12b**). Finally, to assess the expression of pluripotency markers with single cell resolution, immunofluorescence detection of their protein expression was employed. This analysis revealed broad expression of the markers in all cells in the culture, revealing that the reduction in the pluripotency markers is not due to overt differentiation of particular subsets of cells in culture (**Figure 5.12c**).



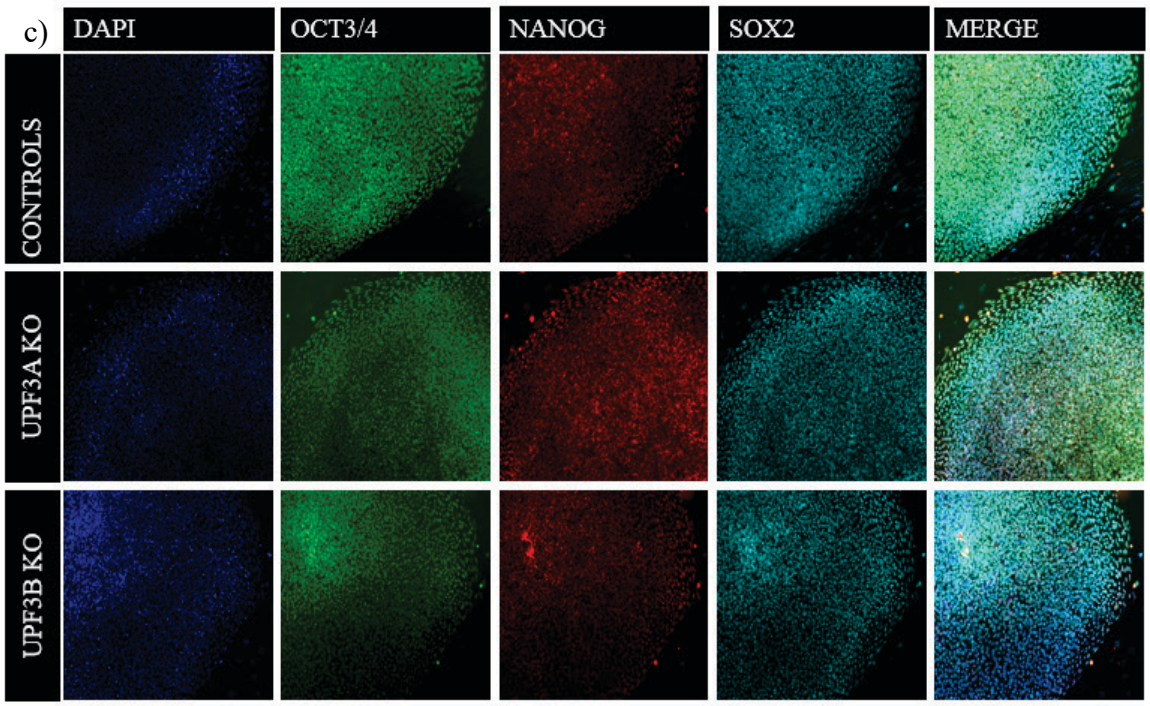


Figure 5.12: Loss of *UPF3A* and *UPF3B* impact the expression of pluripotency and early differentiation marker genes in hESCs.

a) Real time qPCR analysis of pluripotent markers, *NANOG*, *OCT3/4* and *SOX2* mRNA and b) early lineage markers mRNA in controls (n=4), *UPF3A* KO hESCs (n=3) and *UPF3B* KO hESCs (n=3) performed in triplicate. All data normalised to *ACTB* mRNA expression. Graph values represent mean value derived for each genotype, error bars represent standard deviation. *p<0.05, **p<0.01 significantly different to controls by student T-test assuming equal variances. c) hESCs were fixed and stained with pluripotency markers, OCT3/4 (green), NANOG (red), SOX2 (far red) and DAPI (blue). Images acquired at 40X magnification.

5.2.11 *UPF3A* and *UPF3B* are important for cell cycle progression and proliferation

In this section the cell cycle kinetics of *UPF3A* and *UPF3B* KO hESCs were analysed. hESCs have a short cell cycle, highlighted by short G phase's and a high proportion of cells in the S phase (White & Dalton 2005). Upon differentiation, this ratio changes leading to an enrichment of cells in the G1 phase (Barta et al. 2013). NMD is known to impact cell cycle and target mRNAs encoding regulators of the cell cycle (Lou et al. 2016; Lou et al. 2014). *UPF3B* KO (and to a lesser extend *UPF3A* KO) hESCs were shown to have reduced *OCT3/4* mRNA expression encoding a pluripotency cell marker which is important in cell cycle progression (**Figure 5.12a**). Downregulation of *OCT3/4* prevents proliferation by blocking the G1 cell progression (Lee et al. 2010). Collectively this suggests that loss of *UPF3A* and *UPF3B* and a compromised NMD in general may impact cell cycle progress in hESCs.

Single cells from controls, *UPF3A* and *UPF3B* KO hESCs were collected for cell cycle analysis just prior to a normal passaging event. Cell cycle analysis by DNA content using propidium iodide (PI) was used to determine cycle phase by FACS. Both loss of *UPF3A* and *UPF3B* had an impact on the cell cycle kinetics, however the differences were distinct. Loss of *UPF3A* led to a significant reduction in the percentage of cells in the G1 phase and an enrichment of cells in the G2/M phase. In contrast, loss of *UPF3B* had the reciprocal effect, with an enrichment of cells in the G1 phase (although not significant) and a significant reduction of cells in the G2/M phase **(Figure 5.13a**). Next, the mRNA expression genes encoding the cell cycle inhibitors *CDKN1A* and *CDKN2A* were analysed. *CDKN1A* and *CDKN2A* cause an increase in cells at the G1 phase by preventing cells to enter the S phase by inhibiting the phosphorylation of retinoblastoma proteins through inactivation of cyclin-dependent kinase (CDK) CDK2 and CDK4 (Harper et al.

1995; Zhao et al. 2016). These transcripts have been shown to be upregulated when *UPF1* and *UPF3B* is downregulated in hESCs (Lou et al. 2016). Both transcripts were deregulated (*CDKN1A* downregulated and *CDKN2A* upregulated) in *UPF3B* KO clones. The *CDKN1A* transcript level in *UPF3A* KO hESCs were also downregulated, however it was not significant **(Figure 5.13b**).

To determine if the defects in cell cycle progression had an impact on proliferation, the MTS proliferation assay was performed on the cells. Equivalent number of cells from controls, *UPF3A* and *UPF3B* KO hESCs were plated in quadruplicates and assay readings taken from day zero and every two days for six days. Both *UPF3A* and *UPF3B* KO clones showed reduced proliferation rates compared to the controls, but statistical significance was not reached **(Figure 5.13c**).

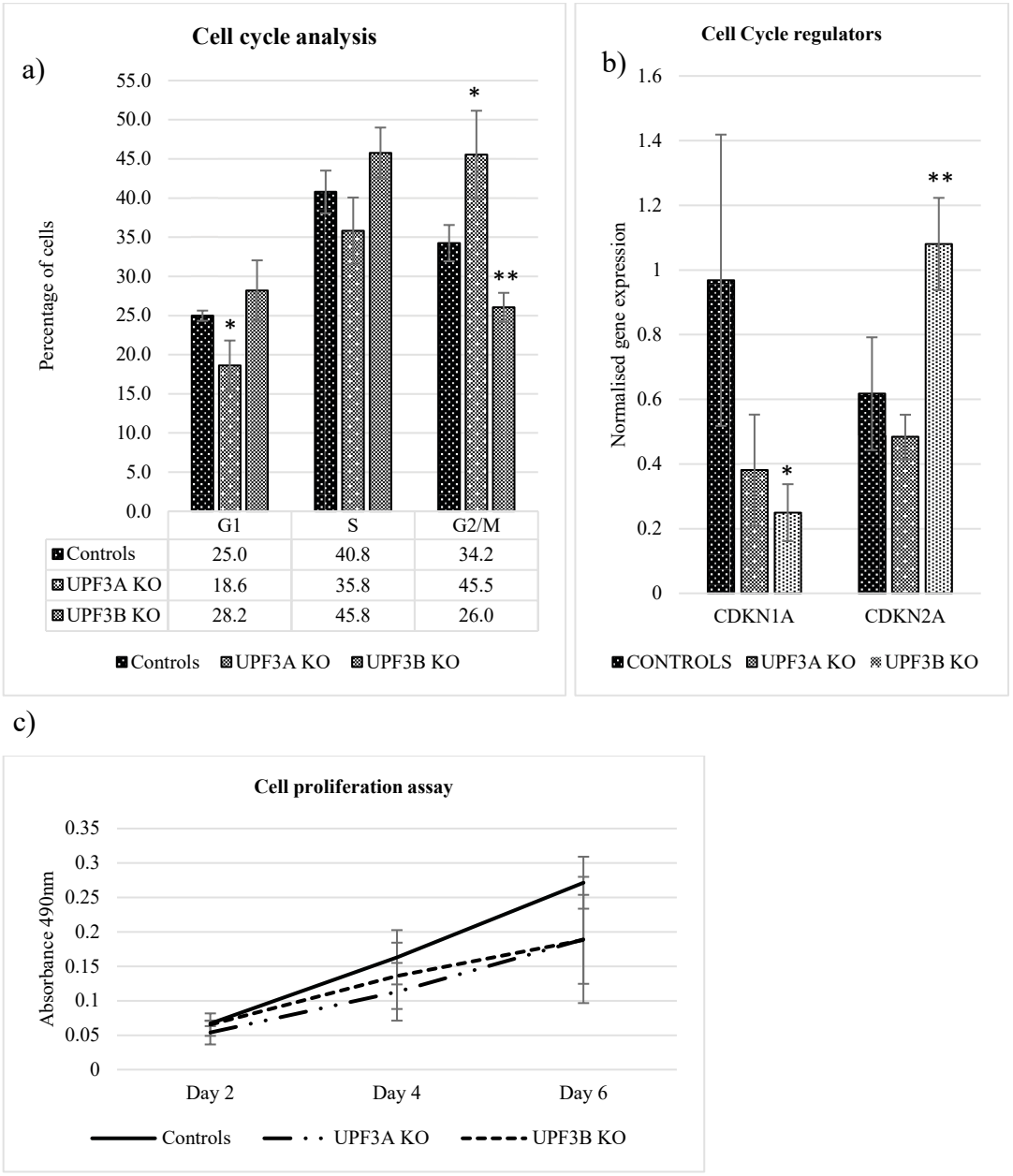


Figure 5.13: Loss of *UPF3A* and *UPF3B* leads to delayed cell cycle progression.

a) Cell cycle analysis of *UPF3A* and *UPF3B* KO hESCs. Single cells were collected prior to passaging. Cells were fixed and nuclear stained with PI and FACS analysis performed n=4 independent control clonal lines compared to n=3 independent *UPF3A* KO and n=3 *UPF3B* KO hESC lines. b) Real time qPCR analysis of mRNA cell cycle inhibitors. n=4 independent control clonal lines compared to n=3 independent *UPF3A* KO and n=3 *UPF3B* KO hESC lines performed in triplicate. All data normalised to *ACTB* mRNA expression. c) Cell proliferation assay was performed using the MTS proliferation kit. n=4 independent control clonal lines compared to n=3 independent *UPF3A* KO and n=3 *UPF3B* KO hESC lines. Experiment was performed in quadruplicate. Graphs represent mean value for each genotype, error bars represent standard deviation. *p<0.05, **p<0.001, ***p<0.0001significantlly different to controls by student T-test assuming equal variance.

5.3 Discussion

Null mutations in NMD factors *Upf1*, *Upf2*, *Upf3a*, *Smg1* and *Smg6* result in early embryonic lethality in mice. Death has been shown to occur after the immediate implantation stage to between embryonic stage (E) 9.5 (Li et al. 2015; McIlwain et al. 2010; Medghalchi et al. 2001; Shum et al. 2016; Weischenfeldt et al. 2008). This indicates that these NMD factors are not essential at the blastocyst or epiblast stage but are however required during gastrulation. This suggests that even though these NMD factors are embryonic lethal, it is possible to generate viable KO hESCs of these NMD factors.

To determine the role of *UPF3A* and *UPF3B* in neurodevelopment, and the role of *UPF3A* in NMD, *UPF3A* and *UPF3B* KO hESCs were generated using CRISPR/Cas9 genome editing technology (Chapter 3). The insertions and deletions (indels) created in the *UPF3A* and *UPF3B* KO hESC clones resulted in the introduction of PTCs in their respective ORFs. *UPF3A* KO clones had reduced *UPF3A* mRNA levels compared to the controls. *UPF3B* mRNA levels were slightly upregulated and not degraded in *UPF3B* KO clones, suggesting they were escaping NMD. Western blot analysis of protein lysates extracted from *UPF3A* and *UPF3B* KO hESCs indicated that the frameshift mutations did completely inhibit functional full-length protein production (**Figure 5.5**). UPF3A protein was stabilised in *UPF3B* KO clones validating the existence of the posttranscriptional regulatory switch originally discovered in non-hESCs (Chan et al. 2009) (**Figure 5.6**). This stabilisation is thought to compensate for loss of *UPF3B* as UPF3A is a 'weak NMD activator'. For example in patients with loss of function mutations in *UPF3B,* UPF3A is stabilised. The level of UPF3A stabilised was different between patients. Patients with more UPF3A stabilised generally presented with a less severe clinical phenotype (Nguyen et al. 2012).

NMD is self-regulated by a negative feedback regulatory network. When NMD is perturbed, the NMD factors are upregulated suggesting a buffering mechanism that regulates the NMD magnitude (Huang, L et al. 2011). *UPF1* and *UPF2* transcript levels were significantly upregulated only in *UPF3B* KO clones and unaffected in *UPF3A* KO clones (**Figure 5.7**). This implies that loss of *UPF3A* does not exert a strong reduction in NMD (as seen by it not having an impact on the classical NMD targets (*ATF4* and *GAS*)), or that it operates in a separate NMD pathway branch (e.g. alternative branch) or fashion that does not target *UPF1* and *UPF2* (Huang, L et al. 2011). The NMD factor transcripts that were upregulated have NMD inducing features (Huang & Wilkinson 2012; Yepiskoposyan et al. 2011).

The NMD substrate genes involved in the stress response pathway (*cJUN*, *GADD45B*, *ATF3*) were significantly downregulated in both the *UPF3A* and *UPF3B* KO clones while *GAS5* and *ATF4* were only upregulated in *UPF3B* KO clones. *GADD45B* and *ATF3* were surprisingly downregulated rather than upregulated (**Figure 5.8**). These two genes are known bona fide NMD targets. A reason for this could be attributed to culture conditions (i.e. presences of stress factors)/passage number as these genes are also involved in the stress response pathway. Cells at a lower passage number were collected and the transcripts levels of *cJUN*, *GADD45B*, *ATF3* were re-analysed. These transcripts were unaffected at the lower passage number. This suggest that either the passage number or culture conditions had an impact on the expression of these genes. Extended cell culture can also have a considerable influence on the gene expression profile and subsequent characterisation of hESCs (Allegrucci et al. 2007). DMEM based media used for culturing hESCs contain 500-1500 fold more of physiological concentrations of methionine/folate cycle components (Steele et al. 2005) an environment that would deviate the methylation of hESCs increasing with each passage (Allegrucci et al. 2005).

PANK2 and *SMG5* which are the alternative NMD substrates were not affected in both *UPF3A* and *UPF3B* KO clones indicating that *UPF3A* and *UPF3B* do not have a role in the alternative NMD pathway. From RNA Seq data from *UPF3B* iPSCs (Miles Wilkinson, personal communication), three (*HIST1H2AC*, *ANXA1*, *HNMT*) of the deregulated transcripts could be validated in the *UPF3B* KO clones. As this analysis was not robust enough to be able to effectively analyse the impact of *UPF3A* and *UPF3B* on the transcriptome in hESCs, RNA Seq analysis was performed which is presented in Chapter 6.

Loss of *UPF3A* and *UPF3B* NMD factor genes were predicted to have an impact on the UPR stress response pathway as protein coding mutations can increase the levels of misfolded proteins in the ER (Hetz et al. 2003; Valastyan & Lindquist 2014). NMD is a quality control pathway that degrades transcripts with PTCs to prevent the production of truncated proteins (Popp & Maquat 2014). A defective NMD pathway could therefore lead to accumulation of transcripts with PTCs due to mutations or aberrant/alternative splicing that could potentially lead to excess misfolded proteins in the ER. NMD is known to play a role in regulating NMD targets *ATF3* and *ATF4* (Mendell et al. 2004) which encode transcription factors involved in the UPR pathway (Goetz & Wilkinson 2017). *upf1* and *upf2* mutant yeast had protein aggregates that were similar to protein aggregates caused by misfolding (Jamar, Kritsiligkou & Grant 2018). The transcript levels of three different sensors that are normally upregulated by activation of the UPR pathway were analysed using real time qPCR. *UPF3A* and *UPF3B* KO clones had a reduction in *EIF2AK3* and *ERN2* transcripts, however this was not significant. *UPF3B* KO clones had a slight insignificant increase in *ATF6* transcript levels. *DDIT3* a downstream target of *ATF4* was significantly downregulated in *UPF3A* KO clones (**Figure 5.11**). These results suggest that loss of *UPF3A* and *UPF3B* does not activate any of the three branches of the

UPR pathway in hESCs. This suggests that the perturbation of the classical NMD pathway elicits no significant ER stress in hESCs.

Loss of either *UPF3A* or *UPF3B* in hESC cells was able to trigger the initial stages of differentiation by causing a reduction in transcript levels of the human embryonic stem cell markers *SOX2* and *OCT3/4* respectively. Loss of *UPF3B* led to an increase in transcript levels of *MESDC2*, an early mesodermal marker and *CLDN6* which is an early endodermal marker, while loss of *UPF3A* lead to an upregulation of *CLDN6* only. High NMD activity is maintained during hESCs differentiation into the mesodermal and ectodermal cell lineages, whilst NMD activity is downregulated during hESCs differentiation into the endoderm (Lou et al. 2016). Loss of both *UPF3A* and *UPF3B* triggered the initial stages of endoderm differentiation (**Figure 5.12**). The WNT pathway inhibits the endoderm lineage differentiation (Loh et al. 2014). Inhibition of NMD causes an upregulation of negative regulators of the WNT signalling pathway (Lou et al. 2016). *MESDC2*, an early mesodermal lineage marker (Holdener et al. 1994) is also known to inhibit WNT/β-catenin signalling (Lin et al. 2011) and therefore could be playing a role in driving the *UPF3B* KO clones towards the endoderm differentiation. The initial stages of differentiation that was caused by loss of *UPF3A* and *UPF3B* was supported by a significant impact on cell cycle progression and cell proliferation albeit insignificant. The cell cycle defects in both the clones could also be attributed to loss of pluripotency. In the *UPF3B* KO hESCs clones, the cell cycle was slightly but not significantly blocked at the G1 phase and downregulation of *OCT3/4* was also observed (**Figure 5.12a**, **Figure 5.13a**). *OCT3/4* a pluripotency cell marker is important in cell cycle progression and its downregulation prevents proliferation by blocking the G1 cell progression (Lee et al. 2010). In addition to causing an enrichment in the G1 phase, cells were reduced in the G2/M phase. Loss of *UPF3A* had an opposite effect, where cells were reduced in the G1

phase, and had an enrichment of cells in the G2/M phase (**Figure 5.13a**). The cell cycle inhibitors *CDKN1A* and *CDKN2A* were deregulated only in *UPF3B* KO hESCs clones and unaffected in *UPF3A* KO hESCs clones suggesting that the defect in cell cycle in *UPF3A* KO hESCs could be attributed to other factors other than deregulation in cell cycle inhibitors (**Figure 5.13b**).

In conclusion CRISPR/Cas9 genome editing technology was faithful in generating *UPF3A* and *UPF3B* KO hESCs with no protein expression. In *UPF3B* KO hESCs UPF3A protein was stabilised, a phenomena been previously described as evidence of a compensatory NMD mechanism in other cell types which occurs in *UPF3B* depleted cells (Chan et al. 2009; Jolly et al. 2013; Nguyen et al. 2012) Loss of *UPF3A* in hESCs had no impact on the buffering feedback loop mechanism and the classical NMD targets such as *ATF4* and *GAS5*, which were regulated by UPF3B. However, both the genes caused a defect in cell cycle progression although distinct and induced the initial stages of endoderm differentiation. The stress response genes which are bona fide NMD targets are not ideal indicators of NMD as cell culture conditions/passage number can impact their expression. As the main aim of this study was to study *UPF3A* and *UPF3B* in a human brain model, in the next chapter *UPF3A* and *UPF3B* KO hESCs were differentiated into NSC and RNA Seq analysis performed at both hESC and NSC stage to determine the transcriptome wide impact due to loss of these genes.

Chapter 6 – Transcriptome-wide impact of loss of *UPF3A* or *UPF3B* NMD as a disease model

6.1 Introduction

Neural differentiation of hESCs lacking *UPF3A* or *UPF3B in vitro* can provide an understanding on the role of NMD in human neurodevelopment. Loss of *UPF3B* causes a broad spectrum of NDDs in humans and leads to defects in learning and neurogenesis in mice while heterozygous copy number losses of *UPF3A* is associated with neural tube defect (Addington et al. 2011; Huang et al. 2018; Laumonnier et al. 2010; Luo et al. 2000; Nguyen et al. 2012; Tarpey et al. 2007). Mutations in other NMD and EJC factors also have effects on neurogenesis and behaviour (Alachkar et al. 2013; Colak et al. 2013; Long et al. 2010; Silver et al. 2010; Wittkopp et al. 2009).

To begin to determine the role of *UPF3A* and *UPF3B* in human neurodevelopment, in this chapter *UPF3A* and *UPF3B* KO hESCs were successfully differentiated into NSCs using the dual SMAD inhibition method (Shi, Kirwan & Livesey 2012). The generated NSCs exhibited reduced pluripotency markers and an increase in NSC markers. Transcriptome profiling was performed on mRNA extracted from *UPF3A* and *UPF3B* KO hESCs and NSCs. Validation of the RNA Seq analysis was performed on the same mRNA submitted for RNA seq from hESCs and NSCs and also on neural rosettes.

6.2 Results

6.2.1 Neural differentiation of *UPF3A* and *UPF3B*-deficient NMD hESCs into NSCs.

UPF3A and *UPF3B* KO hESCs were differentiated into NSC lines using the dual SMAD inhibition method (Shi, Kirwan & Livesey 2012) . The synergistic action of the SMAD inhibitors Noggin and SB431542 are efficient in driving the differentiation of hESCs to NSCs through inhibiting the Activin/Nodal/BMP kinase receptors. Inhibition of these kinase receptors inhibits downstream SMAD signalling and drives neuroectodermal differentiation. This method recapitulates temporal development of neural cell populations *in vivo* in a step wise manner and generates dorsal cortical excitatory neurons (Chambers et al. 2009; Shi, Kirwan & Livesey 2012; Shi et al. 2012).

The first observable neural cell population derived from hESCs after neural induction is reminiscent of the neuroepithelia of the neural plate and occurs approximately 8 to 12 days after neural induction. These NSCs maintain the expression of SOX2 and begin expressing neuroepithelial markers such as Nestin, SOX1, SOX3, PSA-NCAM and MUSASHI-1 (Dhara & Stice 2008). The neuroepithelial layer is subsequently replaced by a second population of NSCs, 'neural rosettes structures' and occurs between days 12-17 of neural differentiation (Shi, Kirwan & Livesey 2012). Neural rosette structures are NSCs that express markers of the neuroepithelial cells of the neural plate *in vivo* and arrange radially forming structures reminiscent of the *in vivo* structural formation of the developing neural tube (**Figure 6.1**) (O'Rahilly & Müller 2007).

NSCs are derived from single cell dissociation of neural rosettes and serial passaging in the presence of bFGF and EGF mitogens. The derived NSCs are highly homogenous and

have the ability to divide over extended periods in culture. They maintain the expression of marker genes that define NSCs rosettes (PAX6, Nestin) but loose the radial arrangement of rosette structure NSCs. Cells are instead found in random/lattice like arrangements. Upon withdrawal of mitogens, these NSCs have the capacity to differentiate into excitatory neurons of dorsal cortex identity (Dhara & Stice 2008).

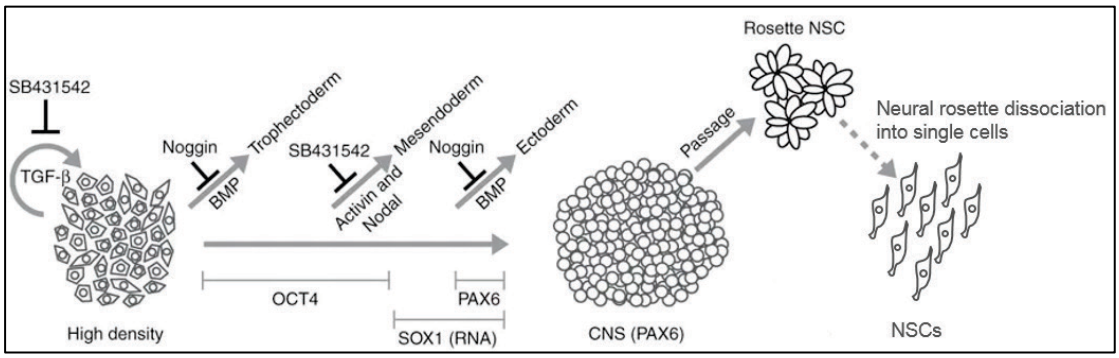


Figure 6.1: Neural induction using the dual SMAD inhibition method.

Addition of the SMAD inhibitors Noggin and SB431542 to high density hESC cultures prevents differentiation into the trophectoderm, mesendoderm and ectoderm leading to the neural default state. Efficient generation of neural cells from hESCs leads to neuropithelial cells and subsequently neural rosettes. As the differentiation progresses it leads to the reduction in pluripotency markers (OCT4) and upregulation of neural markers (SOX1, PAX6). Image modified from (Chambers et al. 2009).

Controls, *UPF3A* and *UPF3B* KO hESCs were differentiated by plating hESCs as colonies on Vitronectin Nunc coated 35 mm dishes. Approximately 5 to 6 days post passaging, cultures were differentiated using the above described dual SMAD inhibition method (Shi, Kirwan & Livesey 2012). After the appearance of neural rosettes at day 13 and their expansion (section 2.1.4.3), NSCs were generated by dissociating neural rosettes into single cells and culturing them in medium supplemented with 20 ng/mL

bFGF and 20 ng/mL EGF. Neural rosettes (day 15) and NSCs (P4) were collected and fixed for immunofluorescence and RNA collected from NSCs at both the neural rosette stage and from derived NSC lines for analysis. Only NSCs RNA was submitted for analysis. Immunofluorescence was used to visualise the expression of marker gene proteins which define the NSCs at the neural rosette and NSC stage. **Figure 6.2** shows representative images of neural rosettes and NSCs derived from hESC controls, *UPF3A* and *UPF3B* KO clones. The controls and KO clones all differentiated efficiently to NSCs with no overt differences morphologically.

To further analyse the neural differentiation experiment, RNA Seq data was used to analyse markers of pluripotency, endoderm, mesoderm and ectoderm lineages in hESCs and NSCs. The results indicated a reduction in the pluripotent markers and an increase in expression of neural markers in NSCs as expected. Most of the mesoderm and endoderm markers remained unchanged in both hESCs and NSCs or either downregulated in NSCs (*NODAL, FOXA2*) (**Figure 6.3**).

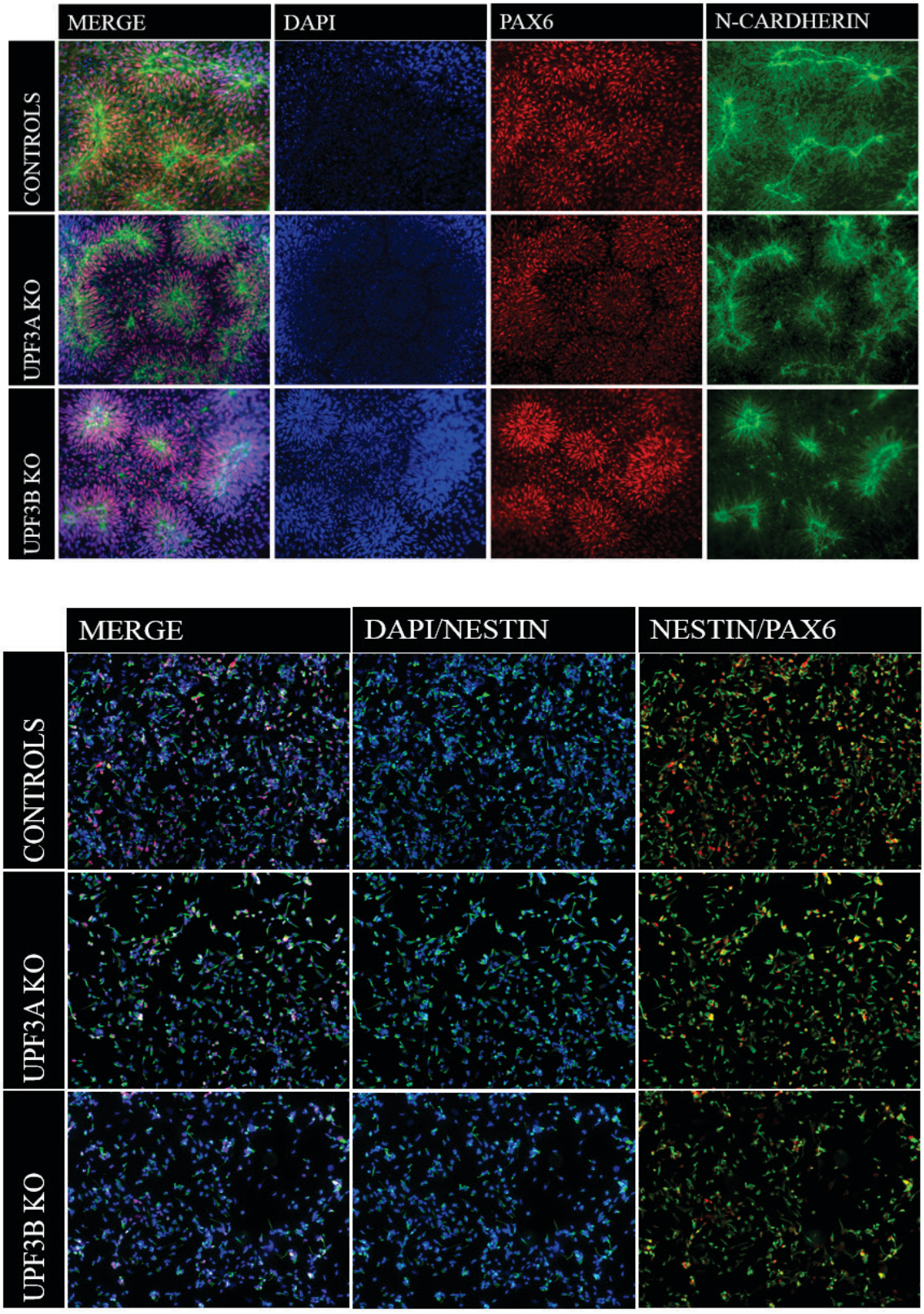


Figure 6.2*: UPF3A* and *UPF3B* KO neural rosettes and neural stem cells.

Neural rosettes and NSCs were generated from controls, *UPF3A* and *UPF3B* KO hESCs. hESC clones were directed to differentiate to NSCs via the 2D dual-SMAD inhibition culture system. After 15 days neural rosettes were fixed and stained with N-Cadherin to identify the neural rosette structures, and DAPI and PAX6 to identify the cell type. After the dissociation of neural rosettes to generate single cells (NSCs), the cells were fixed

and stained with neural cell markers Nestin and PAX6. a) Representative immunofluorescent images of neural rosettes stained with DAPI (blue), PAX6 (red) and N-Cadherin (green). b) Representative immunofluorescent images of NSCs stained with DAPI (blue), PAX6 (red) and Nestin (green). Images acquired at 40X magnification.

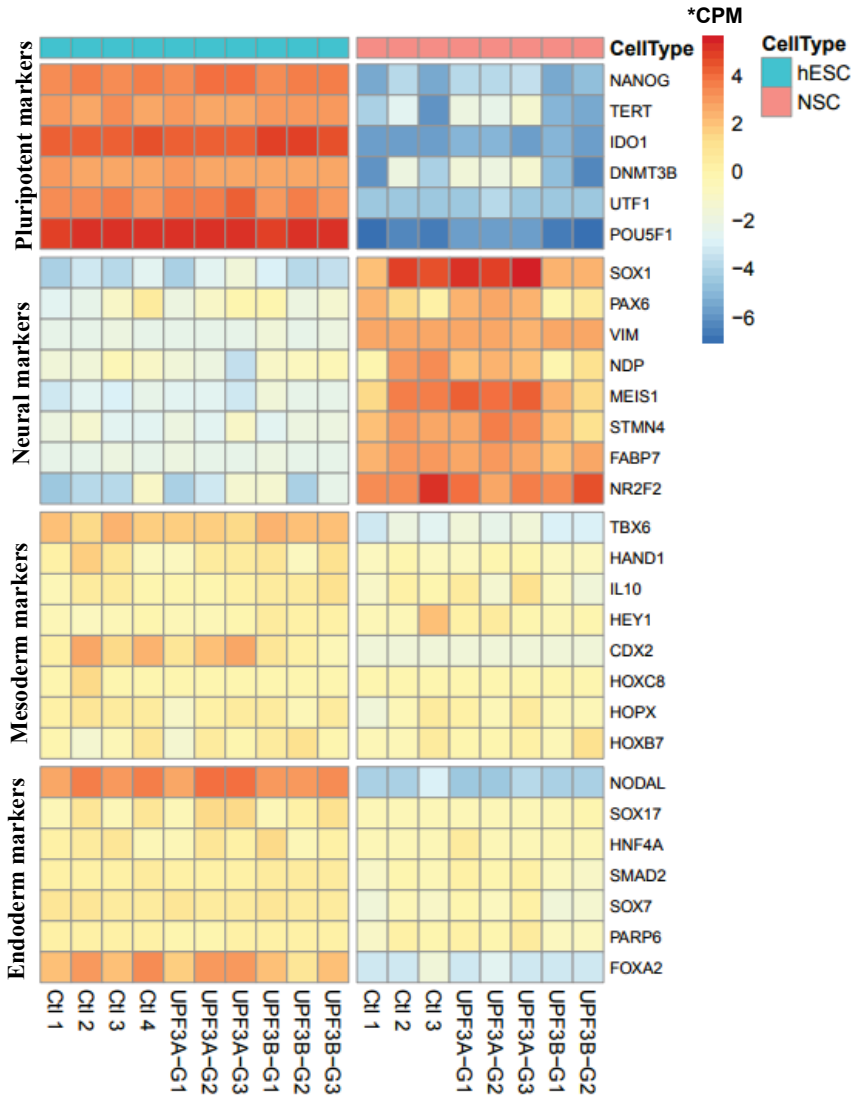


Figure 6.3: Genome-wide analysis of pluripotent and tri-lineage markers in *UPF3A* and *UPF3B* KO clones in hESCs and NSCs.

The mRNA expression level of the pluripotency, endoderm, mesoderm and ectoderm germ layers genes extracted from RNA Seq analysis of hESCs and NSCs clones and visualised using a heatmap. hESCs (controls n=4, *UPF3A* KO n=3, *UPF3B* KO n=3). NSCs (controls n=3, *UPF3A* KO n=3, *UPF3B* KO n=2). *CPM: counts per million.

Another way to determine the success of the neural differentiation in addition to staining the NSCs with neural markers was using the RNA-Seq data that was generated from analysing the transcriptome changes from hESCs to NSCs in the three genotypes (controls (wildtype), *UPF3A* and *UPF3B* KO) individually. The principal component analysis (PCA) of the global gene expression profiles of all data sets showed that the hESCs and NSCs clustered into two distinct cell populations (**Figure 6.4**).

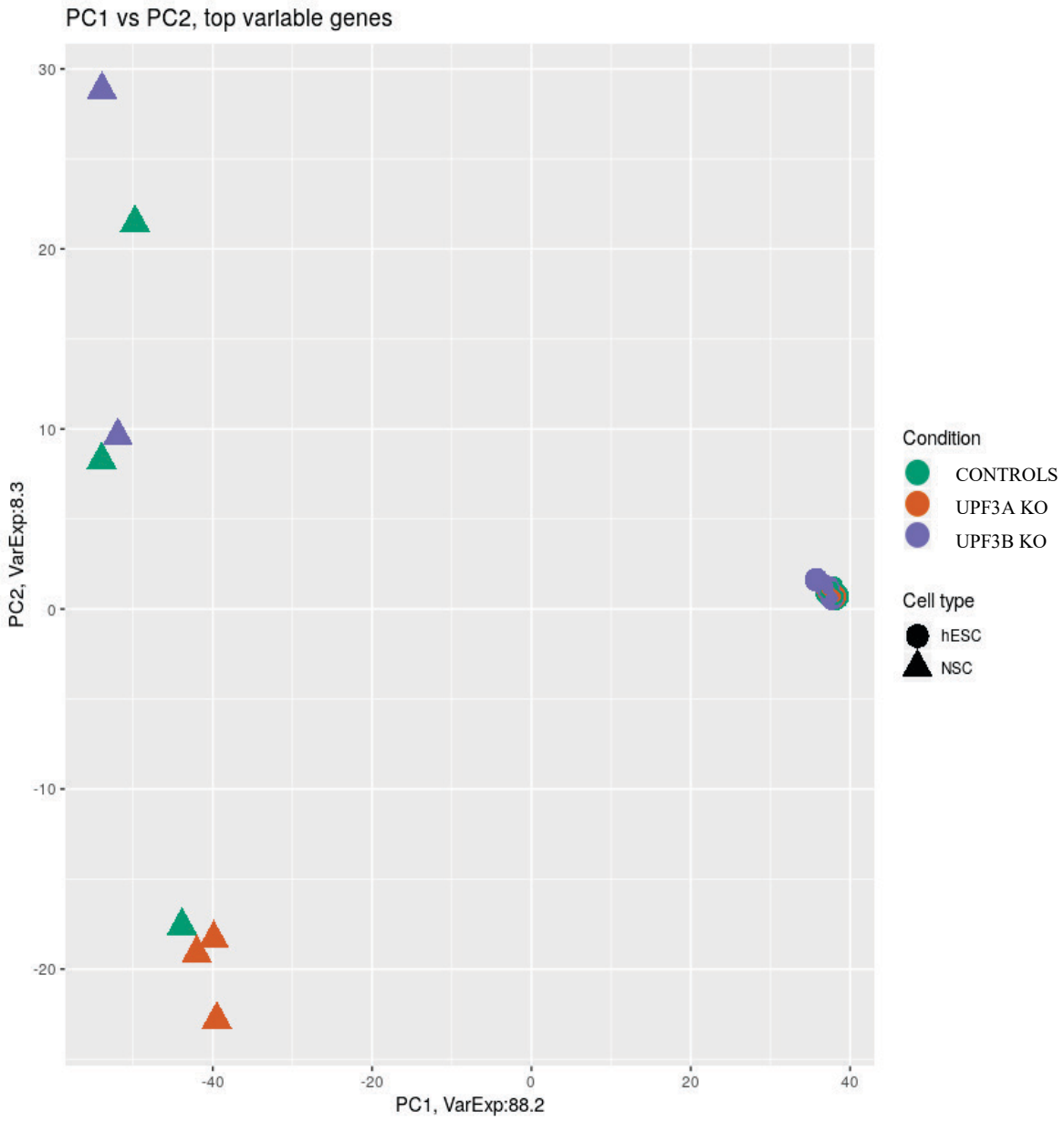
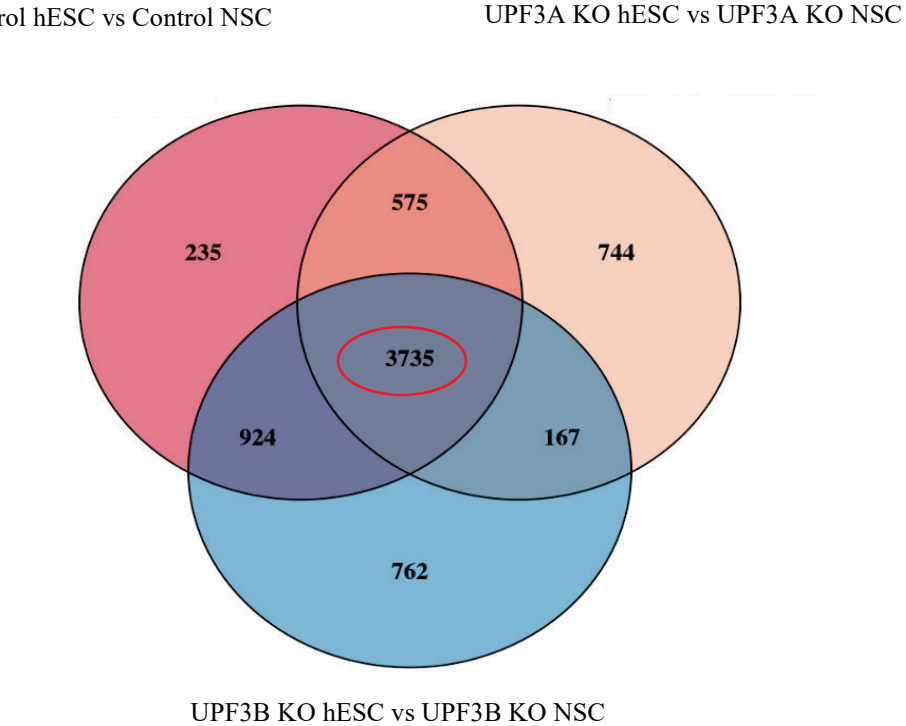


Figure 6.4: Gene-edited KO hESCs and NSCs have different global gene expression profiles

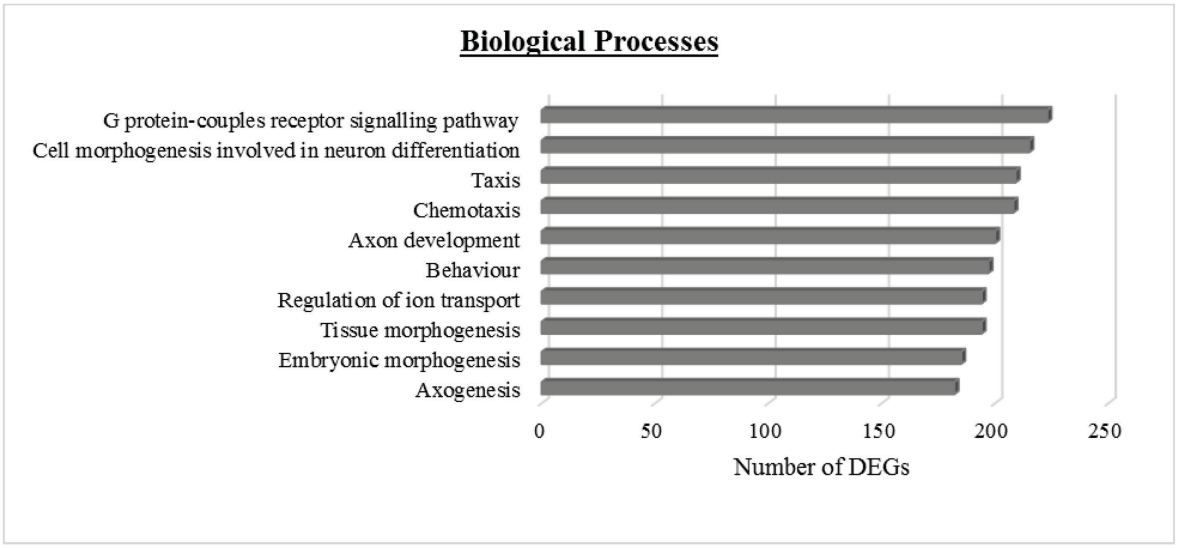
Principal component analysis (PCA) of gene expression profiles of hESCs and NSCs. RNA was extracted from control (n=4), *UPF3A* KO (n=3) and *UPF3B* KO (n=3) hESCs and controls (n=3), *UPF3A* KO (n=3) and *UPF3B* KO (n=2) NSCs. The RNA was subjected to RNA Seq. Samples are coloured based on their genotype and shapes represent the different cell types.

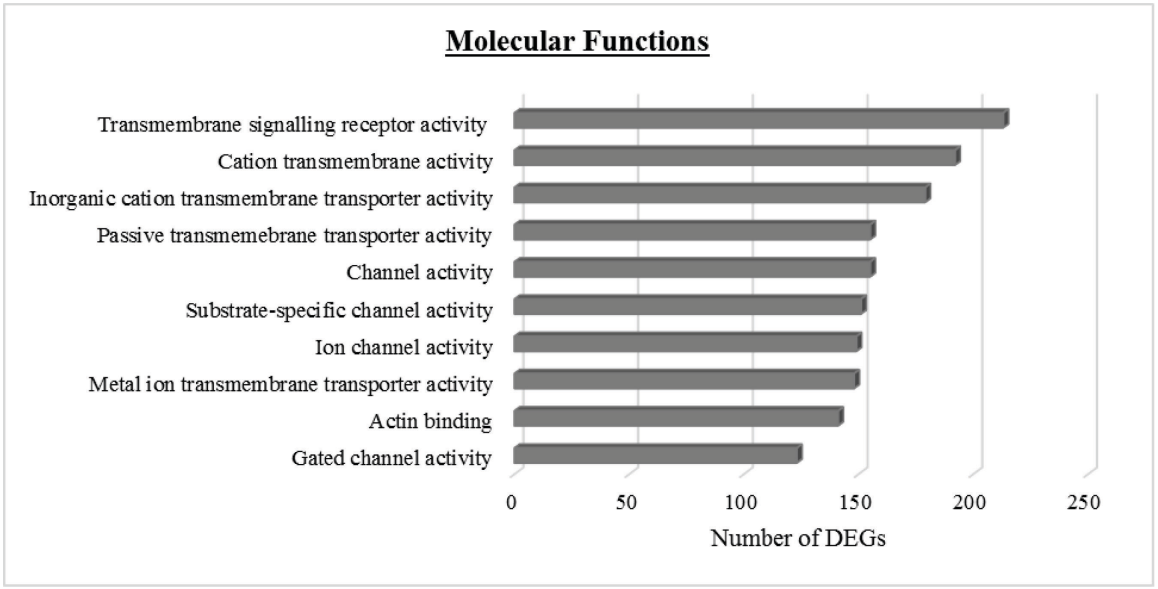
Changes in the transcriptome in hESCs and NSCs were compared between the three genotypes to determine shared genes. The differentially expressed genes that were shared with a log fold change of 2 and $p<0.05$ (3735 genes) were considered to represent genes involved in the differentiation process (**Figure 6.5a**). These genes were subjected to Gene ontology (GO) analysis (Yu et al. 2012). The highest top ten significant GO terms associated with these commonly differentially expressed genes between hESCs and NSCs identified pathways involved in neuron differentiation, axogenesis, neuronal cell body and neuron to neuron synapse, which aligns with successful neural differentiation of hESC clones (**Figure 6.5b** and **Appendix 11-13**).

a)



b) GO analysis on genes involved in differentiation





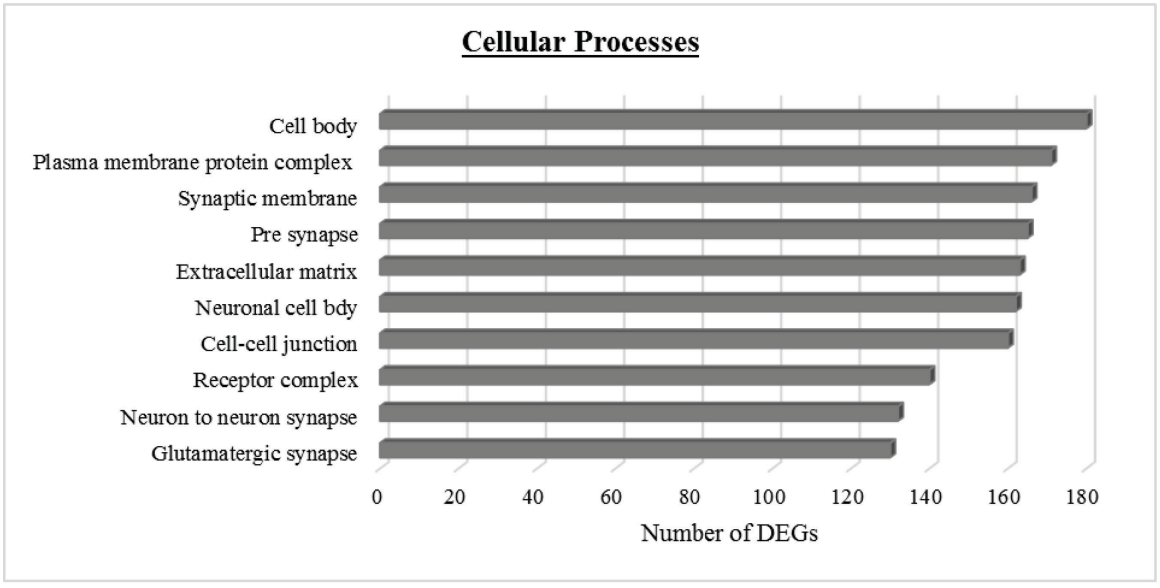


Figure 6.5: The common genes shared between the genotypes from the hESCs and NSCs stage had GO terms for neural cell function and development.

RNA was extracted from control (n=4), *UPF3A* KO (n=3) and *UPF3B* KO (n=3) hESCs and controls (n=3), *UPF3A* KO (n=3) and *UPF3B* KO (n=2) NSCs. The RNA was subjected to RNA Seq. The transcriptome changes in the three separate genotypes (controls, *UPF3A* KO and *UPF3B* KO) were analysed in hESCs and NSCs. Changes were compared between genotypes to identify shared genes. a) Venn diagram showing the shared genes between controls, *UPF3A* and *UPF3B* KO clones. b) Shared genes with a log fold of 2 and $p<0.05$ (3735) highlighted with a red circle in (a) were analysed using GO analysis to identified the pathways the differentially expressed genes are involved in.

Even though the neural differentiation was successful based on IF and transcriptome analysis, the principal component analysis (PCA) of the global expression of NSCs showed variation between the controls (**Figure 6.6**). Because the differentiation of NSCs from hESCs protocol requires approximately a month this may increase heterogeneity of the final neural cell population (Kim et al. 2014). Despite this variation, analysis was still performed with this data.

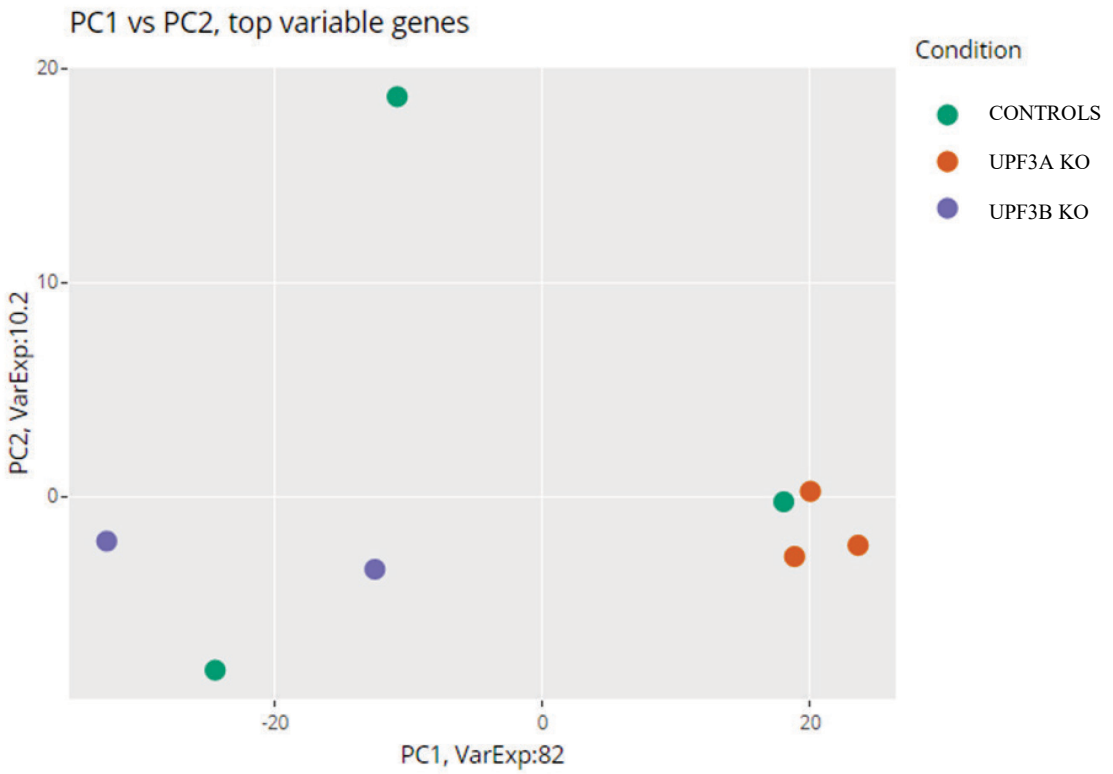


Figure 6.6: PCA plot showing controls, *UPF3A* and *UPF3B* KO NSCs.

Principal component analysis (PCA) of gene expression profiles of NSCs. Samples are coloured based on their genotype. Controls n=3, *UPF3A* KO n=3 and *UPF3B* NSCs n=2.

6.2.2 NMD factors are differentially expressed in hESCs and NSCs

UPF1, *UPF2* and *UPF3B* are highly expressed in hESCs compared to differentiated human cells and their level decreases upon loss of pluripotency (Lou et al. 2016). *UPF1* has been shown to be downregulated by microRNA-128 (miR-128), a neural expressed miRNA during neural differentiation (Karam & Wilkinson 2012). Downregulation of NMD also occurs to permit muscle (Gong et al. 2009), epidermal keratinocyte and human pancreatic progenitor cell differentiation (Lou et al. 2016). The expression of NMD and EJC factors were compared in hESCs and NSCs using the RNA Seq data. Indeed the

NMD factors, *UPF2* and *UPF3B* were reduced in NSCs compared to hESCs, while UPF1 levels were variable. However other NMD (*SMG1*, *SMG6*, *SMG7* and *SMG9*) and EJC factors (*NBAS* and *RMB8A*) were increased in NSCs (**Figure 6.7**). This difference in NMD and EJC factors across the hESCs and NSCs may possibly be due to the different NMD pathways operating in the different cell types. NMD efficiency has been shown to differ across cell and tissue type (Bateman et al. 2003; Huang, L et al. 2011; Zetoune et al. 2008).

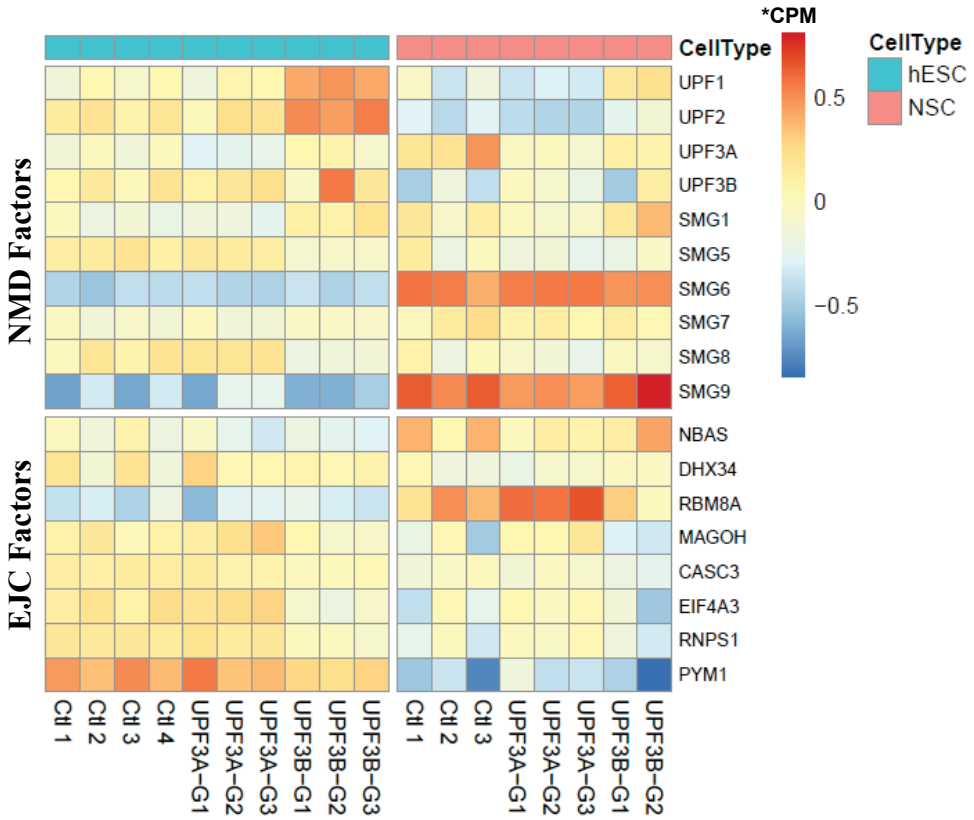


Figure 6.7 NMD factors are differential expressed in hESCs and NSCs.

The mRNA expression levels of NMD and EJC genes were extracted from the RNA Seq data of hESCs and NSCs clones and visualised using a heatmap. hESCs (controls n=4, *UPF3A* KO n=3, *UPF3B* KO n=3). NSCs (controls n=3, *UPF3A* KO n=3, *UPF3B* KO n=2). *CPM: counts per million.

6.2.3 *UPF3A* is important in regulating early embryonic development.

The function of *UPF3A* in the NMD mechanism has not yet been elucidated with data suggesting conflicting roles. Studies suggests it may act as a weak NMD activator (Kunz et al. 2006; Lykke-Andersen, Shu & Steitz 2000), while others suggest it is an NMD inhibitor with some residual NMD activity (Shum et al. 2016). Another study suggested it does not have a role in NMD as *UPF3A* knockdown in HeLa cells had little no effect on the transcripts that were affected by either knockdown of *UPF1* and *UPF3B* (Chan et al. 2007). To determine the role of *UPF3A* in NMD in hESCs and NSCs, differentially expressed genes were identified in *UPF3A* KO hESCs and NSCs compared to controls. Genes that were differentially expressed by at least two-fold with a $p<0.05$ were used for downstream analysis.

The principal component analysis (PCA) of the global gene expression profiles of hESCs showed that the three genotypes (controls, *UPF3A* KO and *UPF3B* KO) clustered into three distinct groups with collection dates having an impact on the clustering of *UPF3A* KO and control hESCs (**Figure 6.8**). In hESCs, loss of *UPF3A* had no impact on the transcriptome. This suggests that *UPF3A* is not important or has no role in the embryonic stem cell state. However, loss of *UPF3A* in NSCs led to a 3.03% deregulation of the transcriptome (425 of 14022 of genes expressed). Of the 425 genes that were differentially expressed, 2.96% (415 of the 14022 transcripts expressed) were downregulated, whereas only 0.07% transcripts (10 of the 14022) were upregulated (**Figure 6.9a**). The top 40 upregulated and downregulated genes are presented in **Appendix 5** and **6**. Some of the deregulated transcripts (*BMP2, BMP4, SMAD7, TGFB1, TGFB, TGFBR2*) are involved in the transforming growth factor β (TGF-β) pathway. GO analysis of the differentially expressed genes also had GO terms involved in intracellular

signalling pathways regulated by the TGF-β superfamily pathway. Genes important in cell cycle progression (*CDKN1A*, *CDKN2A, CDKN2B*) were also deregulated (**Figure 6.9b** and **Appendix 7-9**).

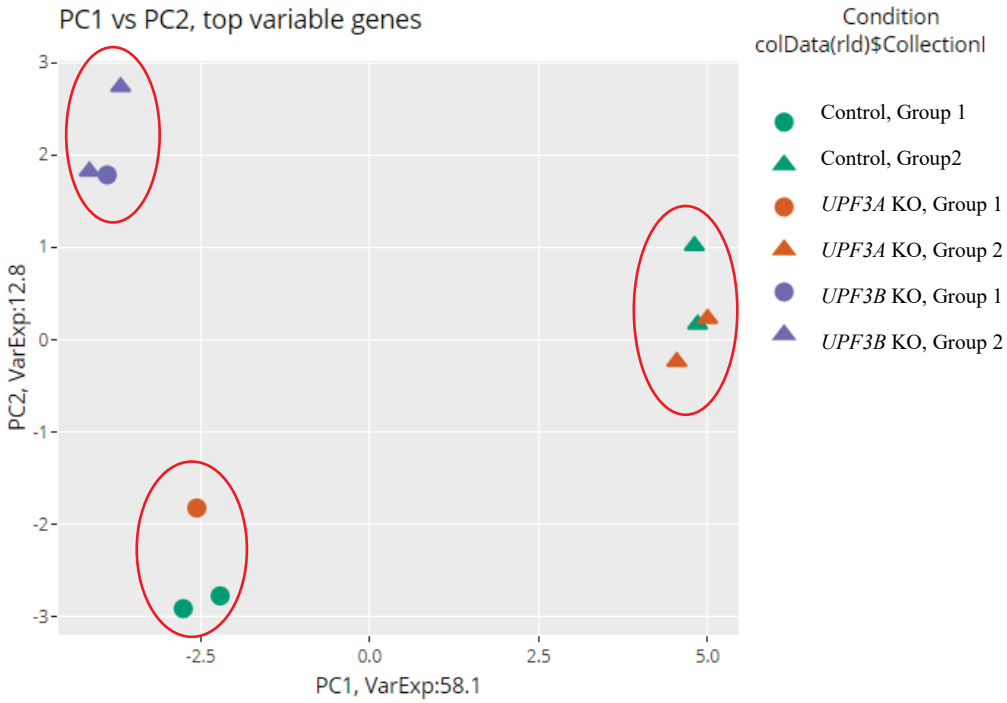


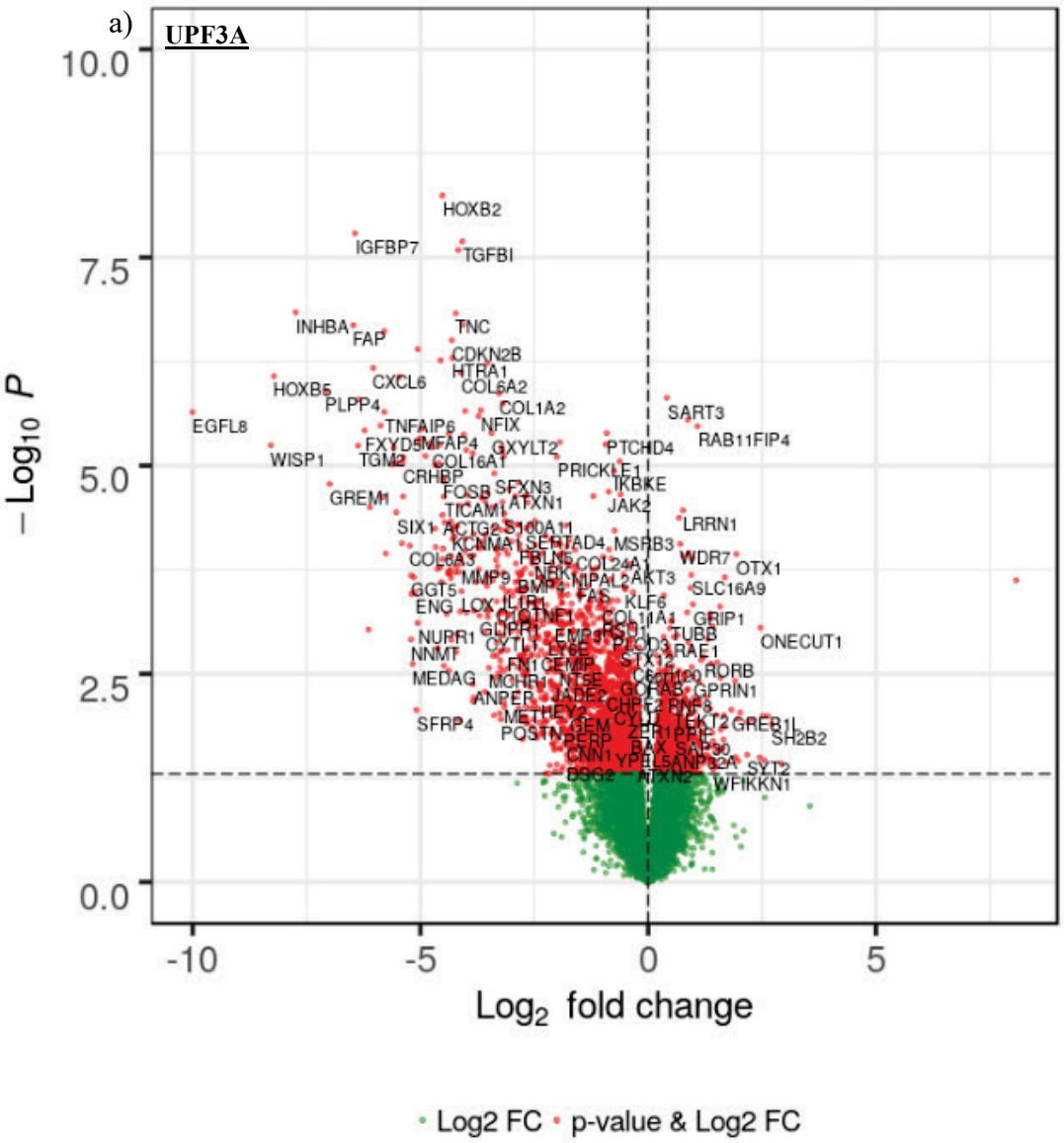
Figure 6.8: PCA plot showing controls, *UPF3A* and *UPF3B* KO hESCs.

Principal component analysis (PCA) of gene expression profiles of hESCs. Samples are coloured based on their genotype and shapes indicate the two different dates of cell pellet collection. Controls hESCs n=4, *UPF3A* KO hESCs n=3, *UPF3B* KO hESCs n=3.

The TGF-β superfamily is a regulator of embryonic development (Gordon & Blobe 2008). It is important to regulate embryonic development and cellular homeostasis such as regulation of proliferation, differentiation, apoptosis and extracellular matrix (ECM) remodelling in a cell (Battegay et al. 1990; De Caestecker 2004; Massague 2000; Massagué & Gomis 2006; Siegel & Massague 2003). A defective TGF-β pathway leads

to developmental disorders, vascular diseases and cancer (Blobe, Schiemann & Lodish 2000; Massagué, Blain & Lo 2000). The top 3 biological processes affected by loss of *UPF3A* in NSCs are ECM, extracellular structure organisation and blood vessel development. GO analysis performed on genes expressed on cortical neurons after differentiation revealed that approximately 45.3% of GO terms were enriched in categories involved in nervous system development and central nervous system development while 29.47% of GO terms were enriched in categories involved in embryo development, organ morphogenesis and circulatory system development (van de Leemput et al. 2014). Therefore it is not unexpected to see such GO terms in the *UPF3A* KO clones. ECM is generated in early life and mutations in genes that encode ECM components can lead to embryonic lethality (Bateman, Boot-Handford & Lamande 2009; Hynes 2009). Lack of TGF-β signalling due to TGF-β receptor mutations in mouse embryonic cells causes embryonic lethality at around E10.5 (Carvalho et al. 2007), while deletion of either the TGF-β 1 ligand in mice severely disturbs vasculogenesis in the yolk sac and leads to embryonic lethality at mid gestation (Arthur et al. 2000; Carvalho et al. 2004; Dickson et al. 1995; Larsson et al. 2001; Leveen et al. 2002). Loss of *UPF3A* in NSCs suggests that it may have a negative impact on TGF-β pathway signalling and loss of *UPF3A* leads to defects in early development. The deregulated genes in NSCs do not have an impact on neural development or function but are however involved in pathways important for embryogenesis and gastrulation (**Appendix 7-9**).

a)


UPF3A
10.0
7.5
5.0
2.5
0.0
−Log10 P
-10
-5
0
5
Log2 fold change
HOXB2
IGFBP7
TGFBI
INHBA
FAP
TNC
CDKN2B
HTRA1
CXCL6
COL6A2
HOXB5
PLPP4
COL1A2
SART3
EGFL8
TNFAIP6
NFIX
RAB11FIP4
FXYD5
MFAP4
GXYLT2
PTCHD4
WISP1
TGM2
COL16A1
PRICKLE1
CRHBP
IKBKE
GREM1
FOSB
SFXN3
ATXN1
JAK2
TICAM1
LRRN1
SIX1
ACTG2
S100A11
KCNMA1
SERTAD4
MSRB3
WDR7
COL6A3
FBLN5
COL24A1
OTX1
MMP9
NRL
MPAL2
AKT3
SLC16A9
GGT5
BMP4
FAS
KLF6
ENG
LOX
IL1R1
COL11A1
GRIP1
NUPR1
GLIPR1
EMP3
PSD1
TUBB
ONECUT1
NNMT
CYTL1
LY6E
PLOD3
RAE1
STX12
FN1
CEMIP
MEDAG
MCHR1
NT5E
RORB
ANPEP
JADE2
GORAB
GPRIN1
METTL7A
CHPF2
RNF8
SFRP4
POSTN
PERP
GEM
CYLD
TEKT2
GREB1L
ZFR1
PPIL
SH2B2
CNN1
BAX
SAP30
YPEL5
ANP32A
SYT2
PSG2
NTN2L
WFIKKN1
Log2 FC
p-value & Log2 FC

b) Biological processes

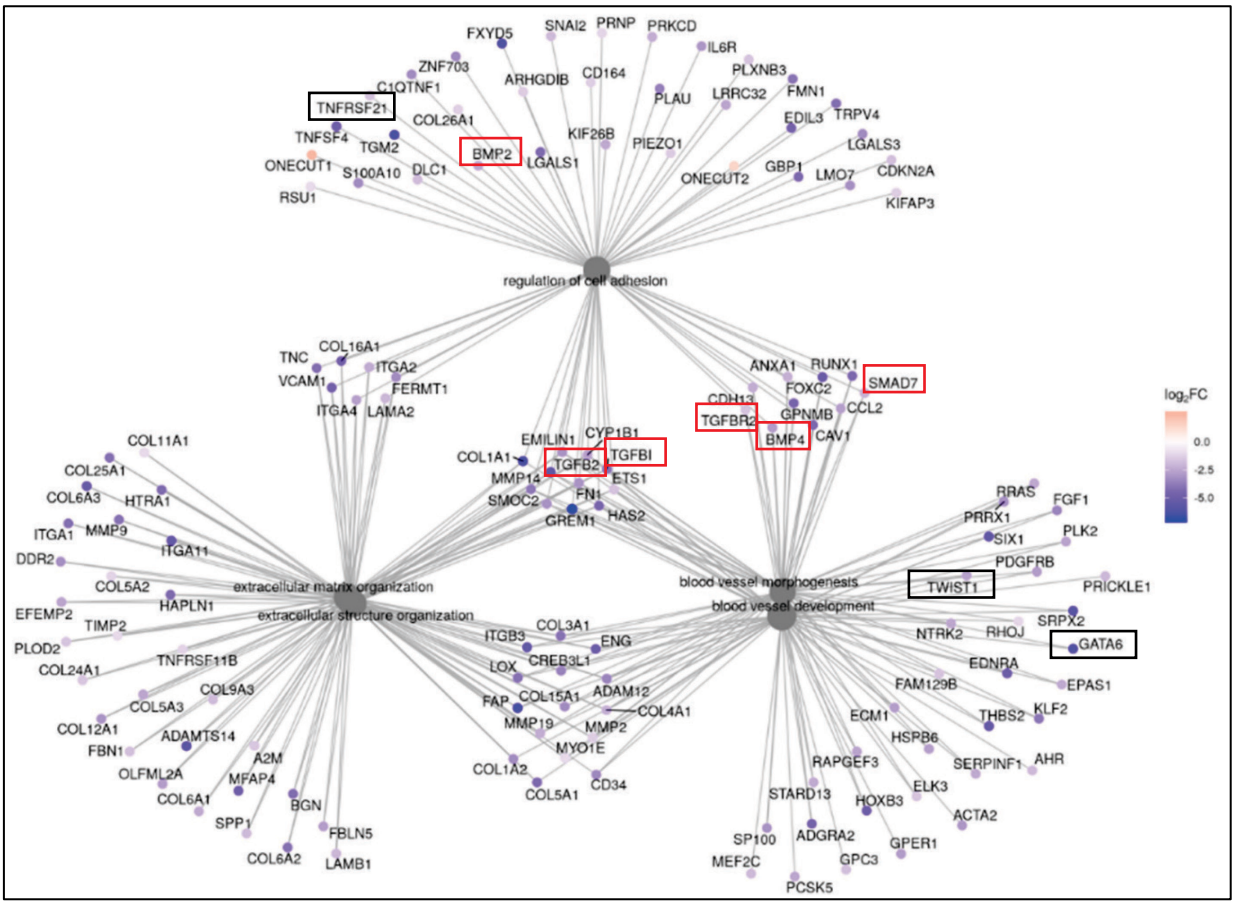


Figure 6.9: Loss of *UPF3A* has a negative impact of pathways that are regulated by the TGF-β pathway

RNA was extracted from NSCs from controls (n=3), *UPF3A* KO (n=3) and *UPF3B* KO (n=2) clones and subject to RNA Seq analysis. a) Volcano plot showing all deregulated genes with $p<0.05$. b) GO analysis was performed on differentially expressed genes with a log fold change of 2 and $p<0.05$ presented using ClusterProfiler to visualise the genes and pathways affected. Red boxes highlight genes that are deregulated that are involved in the TGF-β pathway and black boxes highlight genes that are deregulated in *UPF3A* KO NSCs and P19 cells with *UPF3A* knockdown.

6.2.4 Loss of *UPF3B* in hESCs and NSCs lead to a deregulation of pathways involved in neural function.

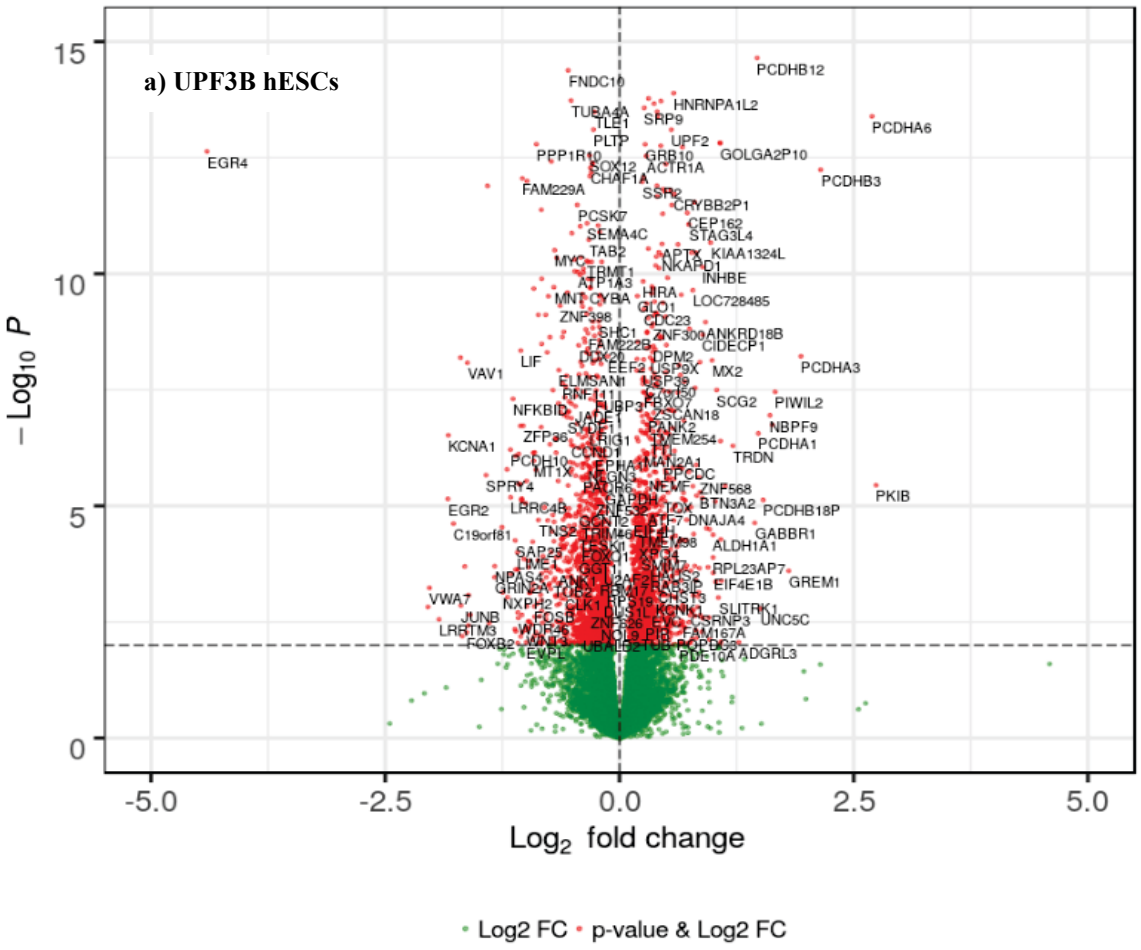
To determine the changes in the transcriptome due to loss of *UPF3B* in hESCs and NSC, RNA Seq analysis was performed on RNA extracted from the *UPF3B* KO hESCs and NSCs. Genes that were differentially expressed by at least two-fold with a $p<0.05$ were used for analysis. The deregulated genes in the *UPF3B* KO hESCs had roles in neural function and brain development. Loss of *UPF3B* in hESCs also resulted in deregulation of genes involved in cell cycle progression such as *CDKN1A*. *GREM1*, an inhibitor of the bone morphogenetic proteins (BMPs) (Walsh et al. 2010) has been shown to have a role in cell proliferation and growth was upregulated (Curran et al. 2012; Frank et al. 2006; Maciel, Melo & Campos 2009; Maciel et al. 2008; Sneddon et al. 2006) (**Figure 6.10** and **Appendix 10**).

Loss of *UPF3B* in hESCs resulted in a deregulation of 0.62% (90 of the 14569 expressed genes) of the transcriptome. From the deregulated genes, 0.21% (31 of 14569) were upregulated while 0.41% (59 of 14569) were downregulated. The top 40 down and upregulated genes are presented in **Appendix 5** and **6**. GO analysis of genes that were deregulated had GO terms that were involved in neural function and brain development such as regulation of post synaptic membrane potential, synapse assemble and calcium ion binding. (**Figure 6.10** and **Appendix 10**). Despite these genes involvement in neural function and development, they are normally expressed in hESCs such as *ADGL3* (*LPHN3*), *SHANK1*, *GABBR1*, *NETO1* and *PCDH10*. The top two significant biological GO terms were homophillic cell adhesion via plasma membrane adhesion molecules and cell-cell adhesion via plasma membrane adhesion molecules and the top deregulated pathway in molecular functions was calcium ion binding (**Figure 6.10**). The affected genes involved in these three pathways were the protocadherins (PCDH) genes.

Protocadherins are a subgroup of cadherins superfamily (Sano et al. 1993). Protocadherins are transmembrane proteins with β folds connected by a linker that binds 2-3 calcium ions. Cadherin-mediated interactions are calcium ion dependent and mediate calcium-dependent homophilic cell-cell recognition and adhesion (Chitaev & Troyanovsky 1998). The deregulated genes involved in homophillic cell adhesion via plasma membrane adhesion molecules and cell-cell adhesion via plasma membrane adhesion molecules are also involved in synapse assembly. Synapses are site of cell-cell adhesion and intercellular communication (Missler, Sudhof & Biederer 2012). Cadherin are also known synaptic cell adhesion molecules. (Pokutta & Weis 2007).

NETO1 is a gene required for synaptic plasticity and learning in mice (Ng et al. 2009), while *SHANK1* is involved in synapse assembly and neuronal signalling (Naisbitt et al. 1999). *LRRC4B*, amongst the deregulated gene is implicated in ASD (Jiang et al. 2013), schizophrenia (Kirov et al. 2012; Purcell et al. 2014) and ID (Sangu et al. 2017). LRRC4B is a member of the Netrin-G ligand 2 (NGL-2), which belongs to the NGL family of synaptic adhesion molecules (Woo, Kwon & Kim 2009). NGL-2 is detected at postsynaptic locations (Um et al. 2018) and is important in regulating the strength of synaptic transmission (DeNardo et al. 2012). Neurotransmitter level regulation at the site of release is essential for proper neuronal function (Anbalagan et al. 2019). *LPHN3* (*ADGRL3*) a gene that was also upregulated is associated with ADHD (Choudhry et al. 2012). *LPHN3* is a brain specific member of the LPHN subfamily of G-protein coupled receptors which are important in regulation of neurotransmitters exocytosis (Linets'ka, Storchak & Himmelreich 2002) and synaptic development (O'Sullivan et al. 2014). In addition to the involvement in cell to cell contacts and in synapse assembly, some of the deregulated genes were also implicated in NDDs for example *SCG2*, a schizophrenia gene (Sequeira, Martin & Vawter 2012), *PCDH10* (Schoch et al. 2017), *NETO1* (Krumm

et al. 2013; O'Donnell et al. 2010) and *SHANK1* (Gong & Wang 2015; Sato et al. 2012) which are ASD associated genes. Loss of *UPF3B* in hESCs resulted in deregulation of neural genes that are expressed in hESCs that are important in neuronal cell function and development.



b) GO analysis of differentially expressed genes in *UPF3B* KO clones

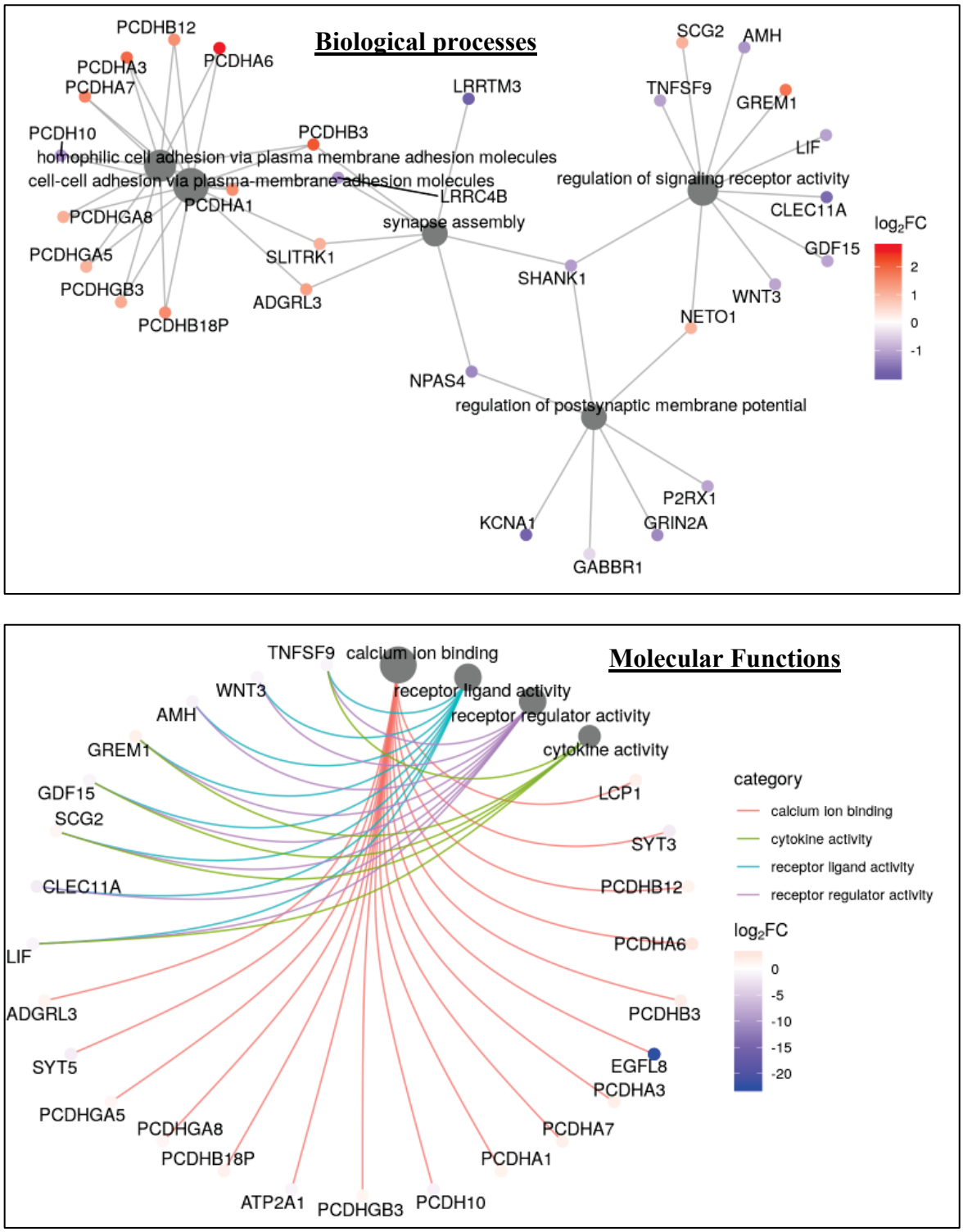


Figure 6.10 UPF3B is important in neurodevelopment and function.

RNA was extracted from hESCs from controls (n=4), *UPF3A* KO (n=3) and *UPF3B* KO (n=3) clones and subject to RNA Seq analysis. Deregulated genes with a $p<0.05$ were presented on a volcano plot while deregulated genes with a log fold change of 2 and $p<0.05$ were used for GO analysis using ClusterProfiler. a) Volcano plots showing the changes in the transcriptome in *UPF3B* KO hESCs. b) Gene ontology analysis using ClusterProfiler to visualise the genes and pathways affected.

Loss of *UPF3B* in NSCs lead to a deregulation of 0.29% (41 of 14022 expressed genes) of the transcriptome. From the deregulated genes, 0.07% (10 of 14022 expressed genes) were upregulated and 0.22% (31 of 14022 expressed genes) downregulated (**Figure 6.11**). This was unexpected as more genes would be expected to be upregulate. RNA seq analysis of the *UPF3B* deficient NSCs also revealed that some of the genes that were deregulated had important functions in neurodevelopment (*ROBO2*, *NRXN1*) and were implicated in NDDs such as *ARX*, *MAP6* and *RELN*. Neurexin 1 (NRXN1) is a presynaptic neuronal adhesion molecule that interacts with postsynaptic neuroligins in both glutamatergic and GABAergic synapses and is important in synaptic formation and function (Reissner et al. 2008) while ROBO2 promotes axon interaction and guidance (Hocking et al. 2010). However no GO terms were identified from this data. This could have been that they were no genes that were enriched for a particular pathway. The results from *UPF3B* KO hESCs and NSCs suggest *UPF3B* is important for normal neurodevelopment and loss of *UPF3B* results in deregulation of genes important for synaptic strength regulation and neural activity.

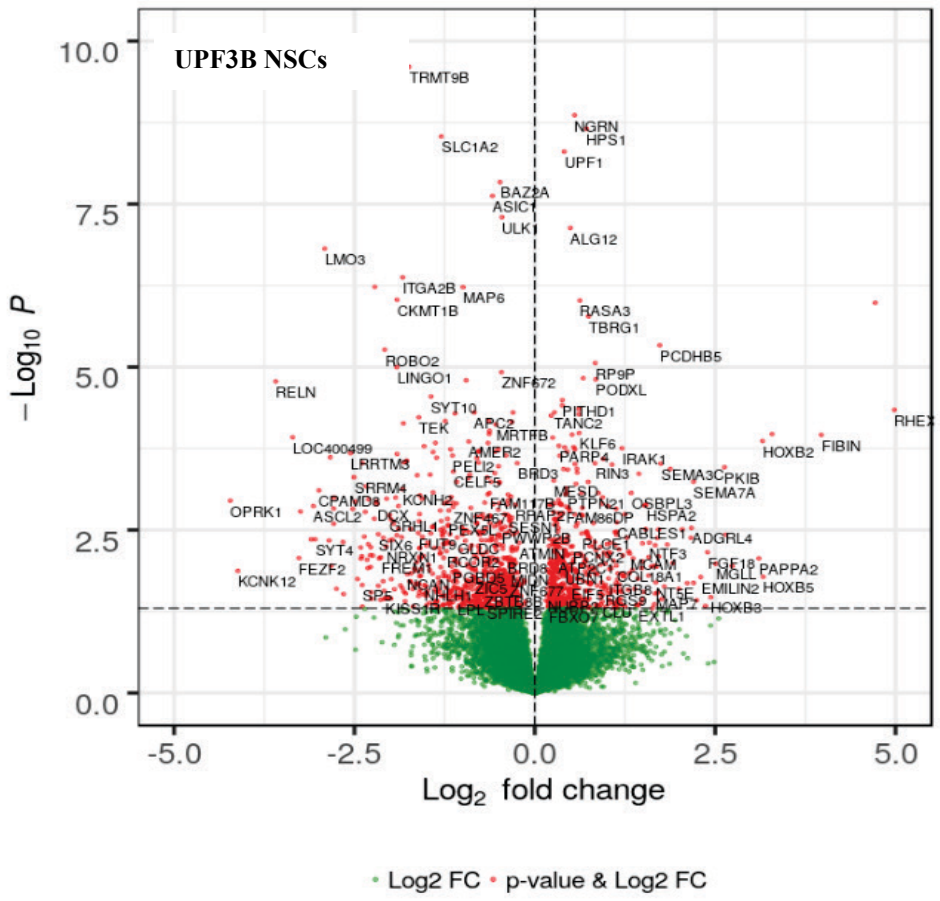


Figure 6.11 Volcano plot showing the UPF3B KO NSCs transcriptome.

RNA was extracted from NSCs from controls (n=3), *UPF3A* KO n=3 and *UPF3B* KO (n=2) clones and subject to RNA Seq analysis. Deregulated genes with a $p<0.05$ were presented on a volcano plot.

6.2.5 Impact of loss of *UPF3A* and *UPF3B* on neural differentiation

To determine how the changes in the transcriptome from hESCs to NSCs was affected by loss of either *UPF3A* or *UPF3B*, the three genotypes (control (wildtype), *UPF3A* KO and *UPF3B* KO) were compared from hESCs Vs NSCs. These changes were then compared between genotypes. The Venn diagram shows the number of unique and shared genes that were identified in this analysis (**Figure 6.5a**) in controls, *UPF3A* and *UPF3B* KO clones. The common 3735 genes in the three genotypes showed an enrichment of

genes specific for neuronal cells (**Figure 6.5b**). When GO analysis was performed using the unique genes in the *UPF3A* and *UPF3B* KO clones only five GO terms were identified in *UPF3A* KO clones and none in *UPF3B* KO clones. The *UPF3A* KO clones had GO terms enriched with genes involved in the inflammatory response (**Figure 6.12**).

The inflammatory response like the other intracellular signalling pathways regulated by the TGF-β superfamily pathway was affected by loss of *UPF3A* in NSCs. This pathway regulates the generation and effector functions of immune cell types (Flavell et al. 2010; Sanjabi, Oh & Li 2017). TGF-β is a pleiotropic cytokine involved in both suppressive and inflammatory immune responses (Sanjabi, Oh & Li 2017) and controls immune cell function (Kehrl et al. 1986). TGF-β is known to supress T cell proliferation and controls its effector functions. TGF-β also inhibits the expression of effector molecules by CTLs such as IFN-γ and perforin (Ahmadzadeh & Rosenberg 2005; Bonig et al. 1999; Ranges et al. 1987; Smyth et al. 1991).

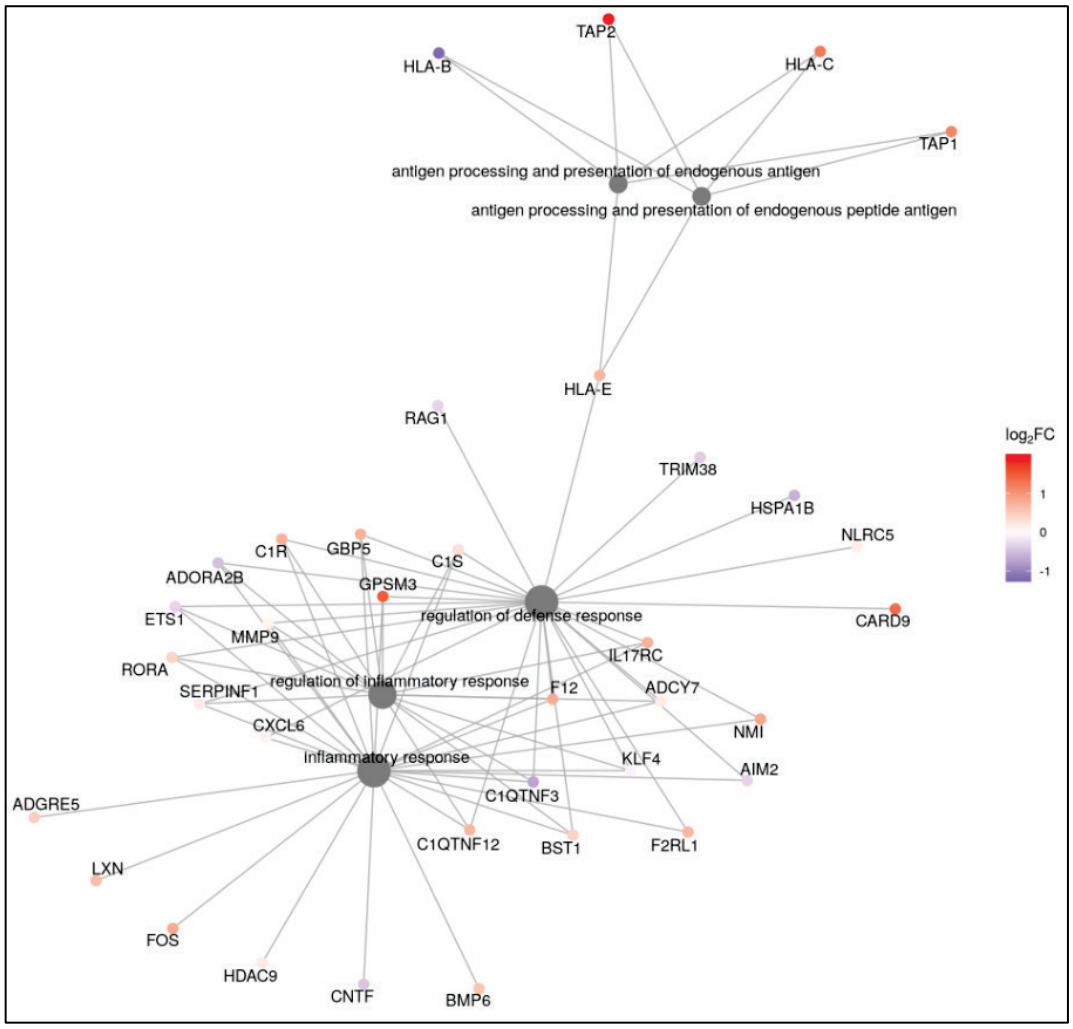


Figure 6.12 Loss of *UPF3A* affected the inflammatory response pathway during the differentiation of hESCs to NSCs.

RNA was extracted from control (n=4), *UPF3A* KO (n=3) and *UPF3B* KO (n=3) hESCs and controls (n=3), *UPF3A* KO (n=3) and *UPF3B* KO (n=2) NSCs. The RNA was subjected to RNA Seq. The transcriptome changes in the three genotypes (controls, *UPF3A* KO and *UPF3B* KO) were analysed in hESCs Vs. NSCs. Changes were compared in each genotype from hESCs to NSCs, the changes were then compared between the genotypes to identify shared and unique genes. ClusterProfiler to visualise the unique genes and pathways affected from *UPF3A* KO during the differentiation of hESCs to NSCs compared to control clones.

6.2.6 Validation of RNA Seq data

To validate the RNA Seq analysis data, 5 genes were selected from the analysis and Taqman RT-qPCR performed on the mRNA that was sent for RNA Seq (hESCs and NSCs) and the additional RNA samples collected from the neural rosette stage cultures. *NETO1* and *LRRC4B* were two genes that were deregulated in *UPF3B* KO hESCs that have a role in neurodevelopment and are associated with NDDs. *NETO1* was upregulated and *LRRC4B* downregulated in hESCs as shown in the RNA Seq data. *LLRC4B* was also significantly downregulated at the neural rosette stage, while *NETO1* was significantly upregulated in NSCs in addition to hESCs. *ARX* and *ROBO2* were selected in *UPF3B* KO NSCs, and were significantly downregulated in NSCs. *ARX* was not detected in hESCs and was significantly downregulated in neural rosettes and NSCs while *ROBO2* was downregulated only in NSCs. For *UFP3A* KO NSCs *HAS2*, an important enzyme for hyaluronic acid synthesis (Cai, Li & Na 2011) was selected for validation. *HAS2* is important for mesendoderm differentiation in human (Xu et al. 2018) and *Has2* null mice die mid gestation with a reduced body size and severe cardiac and vascular abnormalities (Camenisch et al. 2000). This gene was also significantly upregulated in *UPF3A* KO hESCs and in *UPF3B* KO neural rosettes. From the 5 genes that were selected all validated with 100% confirmation indicating accuracy in the analysis (**Figure 6.13**).

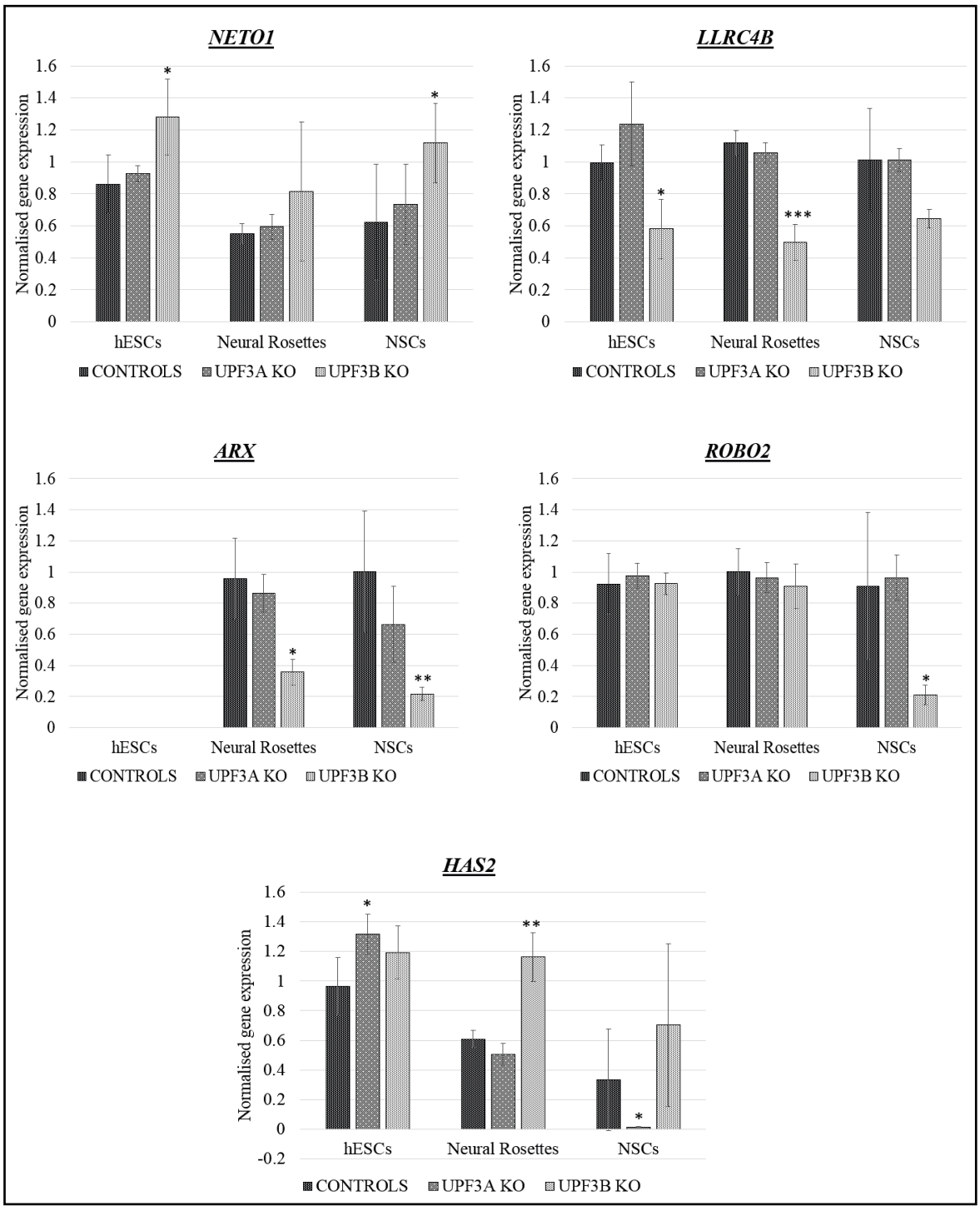


Figure 6.13: Validation of RNA Seq analysis

RNA expression levels of five genes selected to validate RNA Seq analysis using Taqman. Gene expression was analysed on mRNA extracted from hESCs (controls n=4, compared to *UPF3A* KO n=3, *UPF3B* KO n=3, experiment performed in triplicate), neural rosettes (controls n=4, compared to *UPF3A* KO n=3, *UPF3B* KO n=3, experiment performed in triplicate) and NSCs stage (controls n=3, compared to *UPF3A* KO n=3, *UPF3B* KO n=2, experiment performed in triplicate, done in duplicate. All data normalised to *ACTB* mRNA expression. Graphs represent mean value of each genotype, errors bars represent standard deviation. *p<0.05, **p<0.01 significantly different to controls by student T-test assuming equal variance.

6.3 Discussion

To determine the function of *UPF3A* and *UPF3B* in NMD in hESCs and in neurodevelopment, *UPF3A* and *UPF3B* KO hESCs were differentiated into NSCs and transcriptome profiling performed on mRNA extracted from *UPF3A* and *UPF3B* KO hESCs and NSCs. Transcriptome profiling of *UPF3A* KO hESCs revealed that *UPF3A* most likely, by these measures, does not have a role in hESCs, which are cells derived from the inner cell mass of the blastocyst (Guo et al. 2016). hESCs are similar to mouse post-implantation epiblast derived stem cells (Brons et al. 2007; Tesar et al. 2007). The reason why no genes were deregulated in *UPF3A* KO hESCs by RNA seq despite having an effect on pluripotency (lead to a reduction in *SOX2* mRNA and an increase in *CLDN6*, an early endoderm marker) could be speculative (**Figure 5.12**). This could have been that the mRNA expression of the pluripotent markers was performed on different cell pellets. However, even in mRNA extracted from *UPF3A* KO hESCs that was sent for RNA Seq analysis, *HAS2* a deregulated gene in *UPF3A* KO NSCs was significantly upregulated in *UPF3A* KO hESCs but not detected in the RNA seq analysis (**Figure 6.13**). *HAS2* is a gene important for mesendoderm differentiation in humans (Xu et al. 2018) and its upregulation in *UPF3A* KO hESCs is in support of triggering the initial stages of endoderm differentiation. The reason why this is occurring could be due to the sensitivities of the different methods.

Transcriptome profiling of *UPF3A* KO NSCs, however, revealed that loss of *UPF3A* caused a 3.03% deregulation in the transcriptome of which 2.96% were downregulated and only 0.07% were upregulated (**Figure 6.9**). These results suggest that *UPF3A* may act in inhibiting NMD. However, from these genes it was not possible to identify if they were NMD targets as the analysis was performed at the genes and not at the transcript

level as NMD only acts post transcriptionally on mRNAs that contain NMD inducing features such as an uORF, alternatively spliced mRNA and long 3'-UTR (>1.5 Kbp). The genes that were deregulated are involved in intracellular signalling pathways regulated by the TGF-β superfamily pathway (**Figure 6.9, Appendices 7-9**). In addition to the top 5 GO terms that were affected, the transmembrane receptor protein serine/threonine kinase signalling pathway, wound healing, embryonic morphogenesis and gastrulation pathways were also affected. From the genes that were differentially expressed, 3 genes were identified (*TNFRSF21*, *TWIST1*, *GATA6*) (**Figure 6.9b**) that were also deregulated in P19 cells were *UPF3A* was knocked down (Shum et al. 2016).

The suggested deregulation of the TGF-β superfamily pathway in these *UPF3A* KO NSCs may partly explain embryonic lethality in *UPF3A* null mice. *Twist1* null mice are embryonic lethal and die at E11.5 with failed fusion of the cranial neural folds (Chen & Behringer 1995) and *Gata6* null mice die before E7.5 due to failure of visceral endoderm differentiation (Xin et al. 2006). These results suggest that even though UPF3A is expressed as early as in the one cell stage it is only required later in embryonic development. *UPF3A* null mice death occurs around E4.5 to E8.5 (Shum et al. 2016). This may suggest that *UPF3A* has a critical role during the epiblast stage (when the primitive endoderm cells form) and gastrulation (E4.5-E7.5) (Nowotschin & Hadjantonakis 2010; Solnica-Krezel & Sepich 2012). The ECM and structure organisation were the top two GO terms. ECM is present in all cells and is generated early in embryonic life (Hynes 2009). It regulates cellular processes such as adhesion, migration, proliferation, differentiation and survival (Daley, Peters & Larsen 2008). Among the specific growth factors that control ECM synthesis, the TGF-β is the most important (Border & Noble 1994; Verrecchia & Mauviel 2007). ECM is essential and mutations in genes that encode components of the ECM can lead to embryonic lethality

(Bateman, Boot-Handford & Lamande 2009). Mutations in ECM protein also lead to defects in blood vessel development and affect endoderm differentiation (Alpy et al. 2005; Lohler, Timpl & Jaenisch 1984; Schnieke, Harbers & Jaenisch 1983; Smyth et al. 1999). Blood vessel development was the third statistically GO term with the most deregulated genes. The cardiovascular system is the first functional organ system to form and failure of the proper establishment of this network affects organ development and embryo viability (Coultas, Chawengsaksophak & Rossant 2005; Udan, Culver & Dickinson 2013). Deletion of *SMG1*, an NMD factor in mice causes embryonic lethality due to embryos failing to form a vascular system and die at E8.5 However, additional studies will need to be performed to confirm this hypothesis about the role of *UPF3A* in early embryonic development and the TGF-β superfamily signalling pathway.

Loss of *UPF3B* in hESCs and NSCs resulted in a deregulation of genes that are involved in neural function and brain development (**Figure 6.10** and **Appendix 10**). In hESCs, GO terms for cell-cell adhesion via plasma membrane adhesion molecules were observed. The cell-cell adhesion system is involved in many aspects of neuronal development including cell migration, axon-bundle formation, synapse formation and formation of complex glia networks which surround axons and synapses (Togashi, Sakisaka & Takai 2009). Defects in synapse formation has been shown to lead to NDDs (Verpelli & Sala 2012). From the genes that were identified in both *UPF3B* KO hESCs and *UPF3B* KO NSCs, deregulated genes were identified that are important in synapse formation and plasticity such as *SHANK1*, *NRXN1*, *RELN*, *LRRC4B*, *LPPHN3* and *NETO1* among others (**Figure 6.10, Appendix 5-6**).

The adhesion systems are important for brain morphology and highly coordinated brain functions such as memory and learning (Sanes & Yamagata 1999; Washbourne et al. 2004; Yamagata, Sanes & Weiner 2003). In early development of the nervous system, neurons migrate to their location and elongate their axons towards their target. Growing axons are guided by various attractive or repulse target derived cues (Tessier-Lavigne & Goodman 1996). After reaching the final destination axonal growth cones recognise their target cells and form synapses (Scheiffele 2003). *ROBO2*, which was downregulated in NSCs is the main receptor for directing axons within the dorsal tracts and repels neuron cell bodies from the floor plate (Kim, M et al. 2011). Loss of *UPF3B* in NSCs and neural rosettes also lead to a deregulation of genes associated with ID such as ARX. Patients with ARX have ID and epilepsy with or without structural defects in the brain such as lissencephaly, microcephaly and agenesis of the corpus callosum as well as abnormal genitalia (Kato et al. 2004; Katsarou, Moshe & Galanopoulou 2017; Shoubridge, Fullston & Gecz 2010).

Transcriptome profiling of *UPF3B* KO NSCs only led to a 0.29% deregulation in the transcriptome. This may be due to the sample size (controls n=3 and *UPF3B* KO NSCs n=2) or the variation in the control samples (**Figure 6.6**). Even though this would have had an impact on analysing the effects of loss of *UPF3B* in NSCs, there is confidence in the results obtained as for example *ARX* which is downregulated in NSCs is also downregulated in neural rosettes (controls n=4 and *UPF3B* KO neural rosettes n=3) and *NETO1* which is also deregulated in both *UPF3B* KO hESCs and NSCs (**Figure 6.13**).

NMD is known to impact cell cycle and target mRNAs encoding regulators of the cell cycle (Lou et al. 2016). Perturbation of NMD leads to defective cell cycle progression

(Azzalin & Lingner 2006; Lou et al. 2016; Lou et al. 2014; Rehwinkel et al. 2005). In this study loss of *UPF3B* in hESCs and *UPF3A* in NSCs deregulated genes that are important in cell cycle progression such as *CDKN1A*, *CDKN2A* and *CDKN2B* which encode for proteins that inhibitor cell cycle progression at the G1 phase (Harper et al. 1995; Zhao et al. 2016). *CDKN1A* was downregulated in *UPF3B* KO hESCs, while in *UPF3A* KO NSCs, downregulation of *CDKN1A*, *CDKN2A* and *CDKN2B* was observed.

These results suggest that *UPF3B* is important in neurodevelopment and the NDDs seen in patients with *UPF3B* mutations could be due to the deregulation of genes that are important in neural function and brain development. From the neural genes that were affected most of them were involved in cell adhesion and synapse assembly and were associated with NDDs. Loss of *UPF3A* in hESCs did not have any impact in hESCs, however, loss of *UPF3A* in NSCs lead to a 3.03% deregulation of the transcriptome, and 2.96% of that was downregulated implying that UPF3A may acting as an NMD inhibitor. Loss of *UPF3A* had an impact on the TGF-β signalling pathway. This pathway is important in early development and perturbation of this signalling pathway leads to embryonic lethality (Carvalho et al. 2007). These results together with the fact that loss of *UPF3A* in mice leads to embryonic lethality, suggests that *UPF3A* plays an important role in early intracellular pathways important in early embryonic development.

From the overall results *UP3FA* and *UPF3B* seem to have difference roles. *UPF3B* is important in early embryonic development (hESCs) and leads to NDDs by deregulating genes that are important in neurodevelopment and function, thus learning and memory. However, *UPF3A* seems to be important during early development and loss of *UPF3A* affects intracellular signalling pathways regulated by TGF-β superfamily pathway. In

support of these genes not acting redundantly, only two genes were shared between *UPF3A* KO NSCs and *UPF3B* KO hESCs. As UPF3A is postulated to act like an NMD inhibitor, its stabilisation in patients with *UPF3B* mutations is important for the downregulations of transcripts that are upregulated in response to loss of NMD activity, therefore acting like a rheostat.

Chapter 7 -Final Discussion

The role of *UPF3A* and *UPF3B* in neurodevelopment and the role of *UPF3A* in NMD was elucidated by generating *UPF3A* and *UPF3B* KO hESC clones using CRISPR/Cas9 genome editing technology. A neurodevelopmental model was developed by differentiating *UPF3A* and *UPF3B* KO hESCs into NSCs. Studies to determine the role of *UPF3B* in neurodevelopment has been performed on patient LCLs, mice, rat and *Drosophila* (Alrahbeni et al. 2015; Huang et al. 2018; Jolly et al. 2013; Nguyen et al. 2012; Nguyen et al. 2013; Rehwinkel et al. 2005; Tarpey et al. 2007). However, it is important to study the molecular basis of a disease using the most appropriate models reflecting the affected cells or tissue types as molecular pathways are shaped by cell type specific gene expression (Handley et al. 2015). NMD targets are also not conserved across species. For example *ARHGAP24* which is an NMD target in humans does not have an uORF in mice (Jolly et al. 2013).

CRISPR/Cas9 editing technology faithfully generated *UPF3A* and *UPF3B* null hESCs (Chapter 3). WGS analysis of the gene-edited clones detected an average of 3083 unique *de novo* SNVs in each *UPF3A* and *UPF3B* KO clone (**Table 4-1**). None of the detected SNVs were due to off-target mutations but could most likely be attributed to cell culture and the presence of mosaicism in the parental cell line as discussed in section 4.2.4. Another reason for detecting a high number of SNVs in these clones is that the parental cell line H1, which was used for filtering out variants in the gene-edited clones, was at a lower passage (56 passages lower). The identification of off-target mutations from unique *de novo* SVs detected large deletions and duplications that overlapped with the predicted off-target sites (**Table 4-2**, **Appendix 3**). Validation of these variants are still in progress to determine if they are real. Despite such a large number of unique *de novo* SNVs detected in each clone, the gene-edited clones, the parental cell line (H1) and an H1 cell culture (Published H1 WGS data) provided by an external source were all similar with

minor variation between them, when tested using the *vcftools* relatedness algorithm (**Table 4-3**).

To determine the impact of loss of *UPF3A* in neurodevelopment and NMD, transcriptome profiling was performed on *UPF3A* KO hESCs and NSCs. Loss of *UPF3A* causes embryonic lethality in mice and embryos die at around E4.5 and E8.5 (Shum et al. 2016). This suggests that *UPF3A* is important between the period of epiblast formation and during gastrulation and not at the earlier stages (Tam & Behringer 1997; Tam & Loebel 2007). Loss of *UPF3A* in hESCs triggered the initial stages of endoderm differentiation and led to defects in cycle cell progression (**Figure 5.12**, **Figure 5.13**). Loss of *UPF3A* did not have an impact on *UPF1* and *UPF2* which are involved in the NMD buffering mechanism and did not affect the classical NMD targets *GAS5* and *ATF4*, which were upregulated in response to loss of *UF3B* (**Figure 5.7**, **Figure 5.8**).

Transcriptome analysis of *UPF3A* KO NSCs, revealed that *UPF3A* led a 3.03% deregulation in the transcriptome of NSCs (**Figure 6.9**). From the genes that were deregulated, 97.6% of those were downregulated. This is similar to a study where *UPF3A* was knocked down in P19 cells and found that 83% of the deregulated genes were down regulated (Shum et al. 2016). Pluripotency analysis in *UPF3A* KO hESCs, revealed that loss of *UPF3A* triggered the initial stages of endodermal differentiation (**Figure 5.12**). There is evidence that during endoderm differentiation, NMD activity is reduced. Assuming that if indeed UPF3A is an NMD inhibitor, it could be acting as an inhibitor during endoderm differentiation. GO analysis showed an enrichment of genes involved intracellular signalling pathways regulated by TGF-β superfamily pathway (**Figure 6.9**, **Appendix 7-9**). These pathways have important functions in embryonic morphogenesis.

Knockdown of *UPF3A* in P19 cells had GO terms with enriched genes in chordate embryonic development, embryonic organ development and neural tube development (Shum et al. 2016).

UPF3B is important in hESCs and NSCs. Loss of *UPF3B* in hESCs resulted in upregulation of the classical NMD targets *GAS5* and *ATF4* and transcripts involved in the NMD feedback loop (*UPF1* and *UPF2*) (**Figure 5.7**, **Figure 5.8**). Loss of *UPF3B* also led to a reduction in OCT3/4 and upregulation of early markers of the mesoderm and endoderm, *MEDC2* and *CLDN6* respectively (**Figure 5.12**). This triggering of the initial stages of differentiation was supported by defects in cells cycle progression (**Figure 5.13**). Transcriptome profiling of *UPF3B* KO hESCs and NSCs showed that the deregulated genes in both the hESCs and NSCs had important roles in neural function (*LRRC4B*, *LPHN3*, *GABBR1, SHANK1*) and brain development (*ARX*, *ROBO2*) and are associated with NDDs. The genes that were deregulated in hESCs were enriched with GO terms related to cell-cell adhesion via plasma membrane adhesion molecules, regulation of signalling receptor activity, calcium ion binding, intrinsic component of postsynaptic membrane and synapse assembly (**Figure 6.10** and **Appendix 10**).

Future Directions

Loss of *UPF3B* in hESCs and NSCs lead to deregulated genes that are involved in neurodevelopment, neural function and are associated with NDDs. Genes deregulated in hESCs affected pathways involved in cell-cell adhesion and homophilic cell adhesion via plasma membrane adhesion molecule, regulation of signalling receptor activity, synapse assembly and signalling receptor activity. Differentiation of *UPF3B* KO hESCs into neurons would be important to pursue for investigating how the defects in this pathway

are affecting normal neurodevelopment. As synapse assembly was compromised due to defects in cell-cell adhesion and homophillic cell adhesion via plasma membrane adhesion molecules, performing electrophysiology and microelectrode arrays to study cultural networks would be essential. Differentiation of hESCs into neurons would also be important for analysing the morphology of neuron structures such as neuritic outgrowth, as there is compiling evidence that shows that, loss of *UPF3B* in neurons lead to defects in dendritic spine maturation (Huang et al. 2018) and neurite growth (Alrahbeni et al. 2015; Jolly et al. 2013).

Loss of *UPF3A* in NSCs had a much broader impact, affecting the intracellular pathways regulated by the TGF-β superfamily pathway which are important in early development. As loss of *UPF3A* is embryonically lethal, with death occurring around E4.5-E6.5 (period between epiblast formation and gastrulation), and mutations in proteins involved in the intracellular signalling regulated TGF-β superfamily leading to embryonic lethality at around the same period, it could be speculated that *UPF3A* is important during this period, when the epiblast starts to different into the three germ layers. This study revealed that UPF3A could be acting like an NMD inhibitor, as NMD activity is normally reduced during endoderm differentiation. This suggests that *UPF3A* might be important for this reduction in NMD activity during endoderm differentiation as the primitive endoderm occurs as early as E4.5. Differentiating *UPF3A* KO hESCs into the different lineages and performing transcriptome analysis would elucidate the role of *UPF3A* during differentiation and determine which cell lineage is affected by loss of *UPF3A*.

Conclusion

CRISPR/Cas9 genome editing technology is a suitable and easy method to generate mutations in hESCs to model disease. The generated *UPF3A* and *UPF3A* KO clones faithfully modelled 'some' disease pathology already shown to occur in these cells. For example, in *UPF3B* KO hESCs, stabilisation of UPF3A protein was detected, validating the existence of the posttranscriptional regulatory switch originally discovered in non-hESCs (Chan et al. 2009; Jolly et al. 2013; Nguyen et al. 2012) Our results suggest that *UPF3A* and *UPF3B* have different roles in NMD and do not act redundantly as loss of *UPF3A* had no impact on the transcripts involved in the NMD feedback loop (*UPF1*, *UPF2*) and the classical NMD targets (*GAS5*, *ATF4*) which were upregulated in *UPF3B* KO hESCs. Even though both loss of *UPF3A* and *UPF3B* affected pluripotency and cell cycle analysis, the effects were distinct, for example loss of *UPF3A* affected *SOX2*, and the early marker for endoderm differentiation, while loss of *UPF3B* caused a reduction in *OCT3/4* and upregulation of an early marker of the mesoderm and endoderm. In cell cycle progression, these paralogs had opposite effects. Loss of *UPF3A* led to a reduction in cells at the G1 phase and enrichment of cells in the G2/M phase, while loss of *UPF3B* led to an enrichment of cells in the G1 phase (although insignificant) and a reduction in the G2/M phase. Genes deregulated in *UPF3A* and *UPF3B* KO NSC clones also had two genes shared, further implying that these gene paralogs are acting on separate pathways. The *UPF3A* and *UPF3B* KO hESCs in this study are the first to be generated. Extensive analysis (WGS, pluripotency, karyotype analysis) has been performed on these CRISPR KO clones and will therefore provide useful resources for future experiments.

Appendices

Appendix 1: Optimised CRISPR/Cas9 editing technology in hESCs

Steps 1-3: Culturing hESCs for CRISPR/Cas9 editing

1. Culture hESCs as described in section 2.1.1 on an iMEF feeder layer.
2. Once hESCs are growing well on iMEFs with little differentiation transition to feeder free conditions as described in section 2.1.2
3. Expand cultures for CRISPR/Cas9 genome editing depending on the size of the experiment.

Steps 4-17: Preparation of hESCs for CRISPR/Cas9 editing

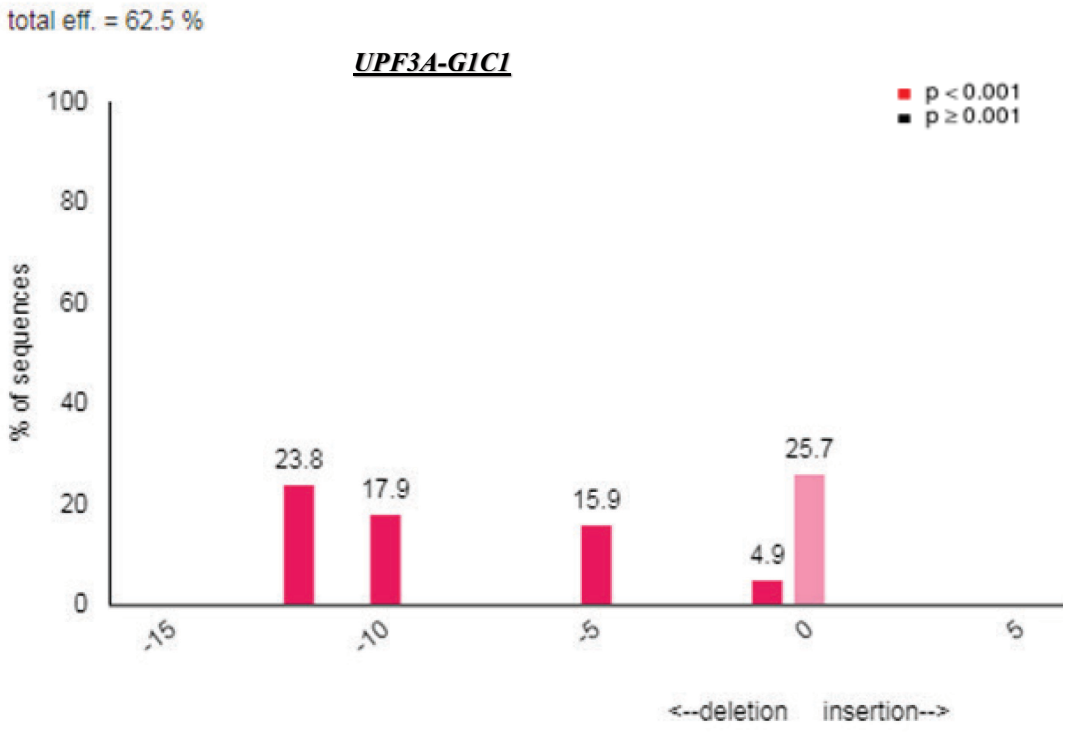
4. Coat 100 mm cell culture dishes with 0.1% gelatin for a minimum of 2 hrs or overnight at 37°C.
5. Aspirated gelatin and leave dishes in the hood to dry for an hr.
6. Thaw iMEFs and plate at a density of $1x10^6$ in a 100 mm culture dish with MEF medium.
7. The following day aspirate iMEF medium and rinse once with DPBS.
8. 11 mL of complete hESC medium with 10 μM ROCK inhibitor was added to each 100 mm dish and transfer into the incubator to equilibrate for 1 hr.
9. Inspect hESC cultures for differentiation under a dissecting microscope in a biological safety cabinet class II under sterile conditions. Differentiated colonies typically lose the defined edges and have large differentiated cells appearing at the border.
10. Remove any regions of differentiation by cutting the differentiating cells using a 21 G needle using a dissecting microscope in a biological safety cabinet class II under stile conditions and scrapping with a pipette tip.
11. Aspirate the medium with the differentiated cells and wash once with DPBS.

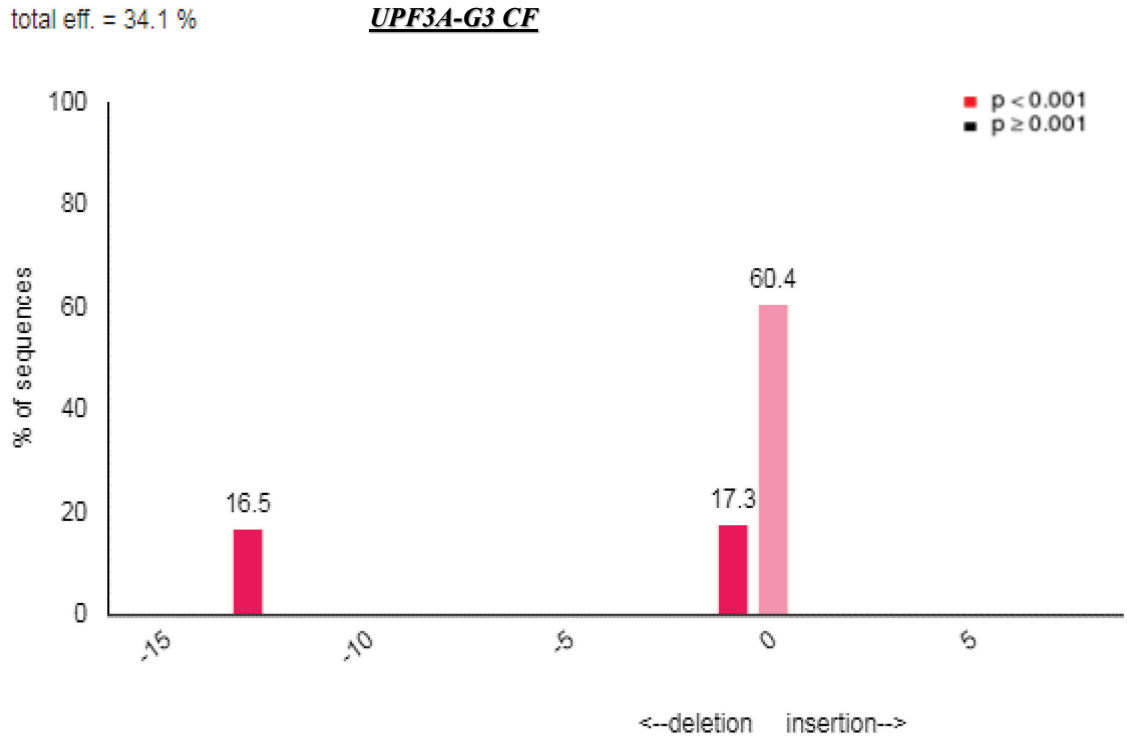
12. Add 2 mL of complete hESCs conditioned medium with 10 μM ROCK inhibitor and leave for 2 hrs before nucleofection.
13. After 2 hrs single cell passage hESCs using Accutase as described in section 2.1.2.3.
14. Perform a viable cell count with trypan blue using the automated cell counter.
15. Collect $1x10^6$ cells per nucleofection (pGFPmax (negative control), PX459 V2 (positive control), G1, G2 and G3 (for each gene)) and spin at 115 g for 3 mins break 1.
16. Aspirate and resuspended cells carefully in 100 μL room temperature nucleofection solution (82 μL of nucleofector solution and 18 μL of the supplement) and then in 2 μg of plasmid DNA.
17. Transfer the 100 μL cell suspension into a nucleocuvette vessel and gently tap to make sure the sample covers the bottom of the cuvette.
18. Place the nucleocuvette vessel with the lid closed into the retainer of the Lonza 4D nucleofector™ unit.
19. Start the nucleofection process using the CB-150 program.
20. After completion, resuspend cells immediately in 500 μL of pre-equilibrated hESC medium with 10 μM ROCK inhibitor from the 100 mm culture dishes.
21. Mix cells gently by pipetting up and down three times using the supplied Pasteur pipettes in the kit and place in 100 mm iMEF coated dishes. Evenly distributed cells in the wells and place in the incubator.

Step 22-33: Selecting and expansion of CRISPR/Cas9 edited clones.

22. After 24 hrs, aspirate medium and rinse once with DPBS.
23. Add 11 mL of complete iMEF CM supplemented with 1 μg/mL puromycin to the cells for 24 hrs.

24. After 24 hrs, aspirate medium and rinse once with DPBS and add complete hESC medium supplemented with SMC4 (Cat# FAL354537, BD Biosciences) for 6 days.
25. Perform daily medium changes with complete hESC medium supplemented with SMC4.
26. Approximately 10 days post nucleofection colonies are ready to be passaged.
27. Prepare 6-well iMEF plates ready for a manual passage.
28. Cut single colonies and place in individual wells of a 6-well plate. Care should be taken to collect all pieces that belong to a single colony to generate a clonal cell population.
29. 6-7 days after the passage a small section of the colony is scrapped off and gDNA extracted using the QuickExtract DNA Extraction solution.
30. Perform a gDNA PCR using primers that flank the region where the gRNA is targeted and sequence the PCR product.
31. Select clones with indels that generate a frameshift and expand further.
32. Perform karyotype analysis on the selected clones. Discard clones with an abnormal karyotype.
33. Freeze cells for future analysis and collect material for analysis.

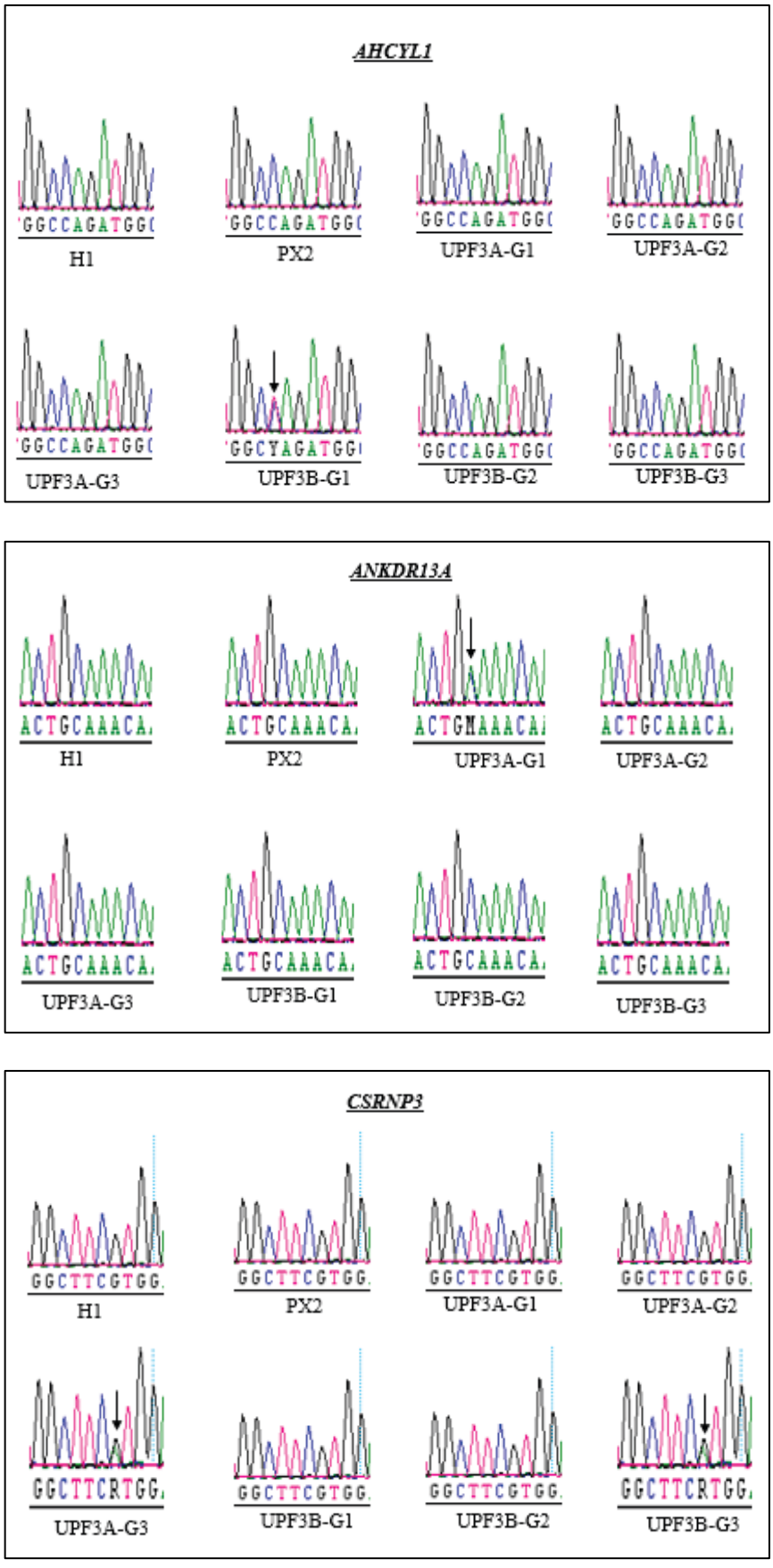


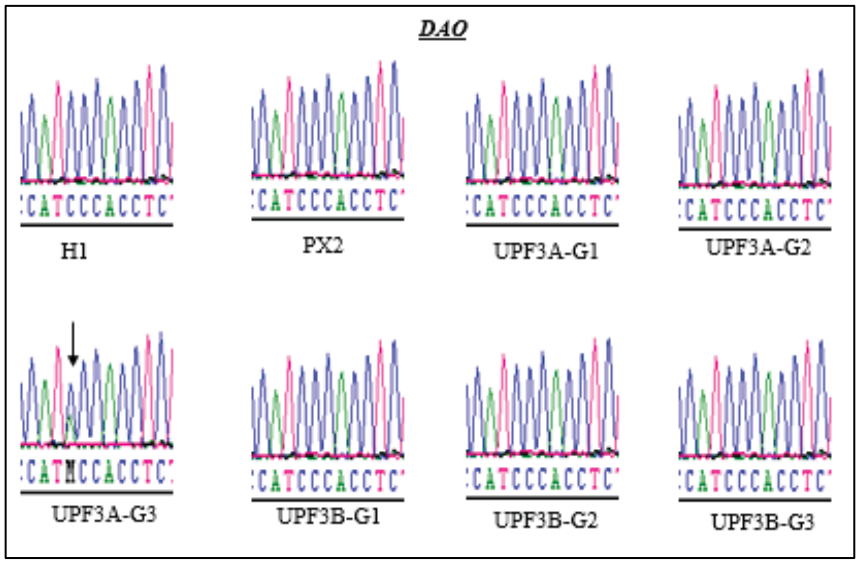


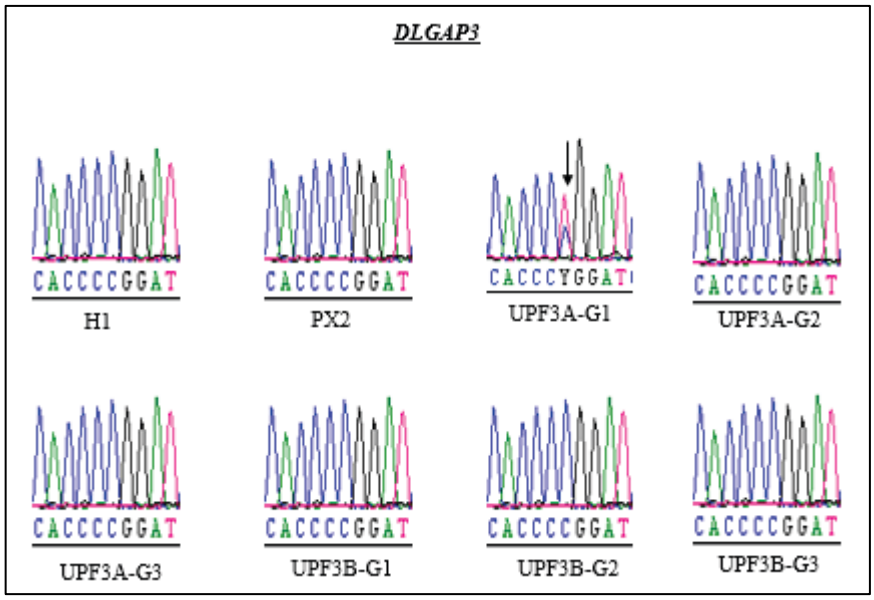
Appendix 2: Non-clonal *UPF3A* KO hESCs detected after sequencing.

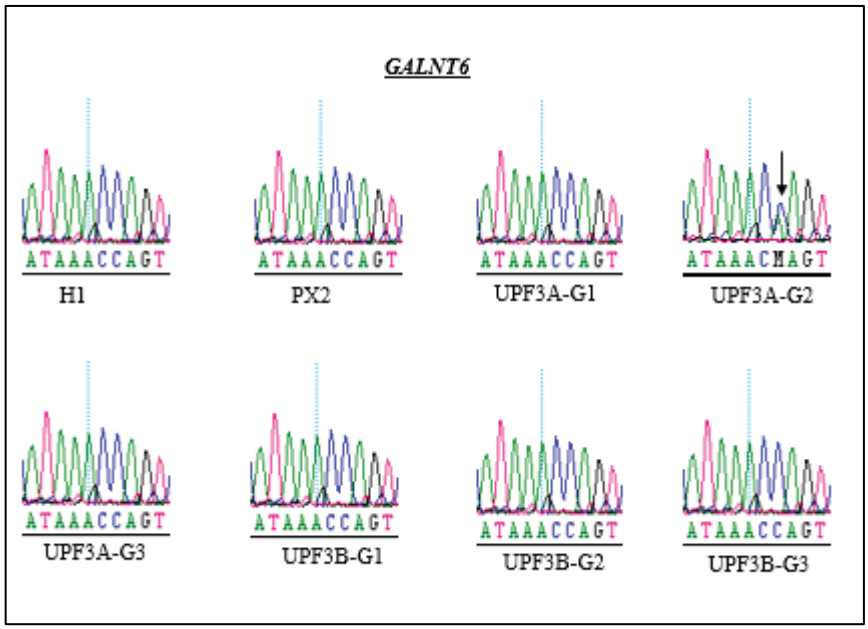
SV chrom	SV start	SV end	SV length	SV type	VARIANT ID
10	34293405	34293729	324	deletion	DEL00422905
12	53347832	66451463	13103631	duplication	DUP00519678
13	62303996	62304208	212	deletion	DEL00559774
13	115047251	115047269	18	deletion	DEL00571862
16	18043642	18043665	23	deletion	DEL00633493
16	18043659	18043680	21	deletion	DEL00633494
17	216825	217125	300	duplication	DUP00666849
17	216858	217259	401	duplication	DUP00666850
17	216925	217551	626	duplication	DUP00666853
17	76623469	76623880	411	deletion	DEL00699564
19	14869542	14869697	155	deletion	DEL00736899
2	2796485	2796952	467	duplication	DUP00052770
2	2796643	2796937	294	duplication	DUP00052772
20	44658049	44658065	16	deletion	DEL00781369
21	44456026	44456344	318	deletion	DEL00806515
21	44456027	44456180	153	deletion	DEL00806516
22	48647705	48647801	96	deletion	DEL00826081
22	48648234	48648695	461	deletion	DEL00826087
22	48648550	48648713	163	deletion	DEL00826091
22	48648627	48648695	68	deletion	DEL00826093
22	48648650	48648985	335	deletion	DEL00826094
3	75683342	75683479	137	deletion	DEL00146487
4	91488716	91489034	318	deletion	DEL00195393
4	185575620	185575772	152	deletion	DEL00210684
7	1944553	1945131	578	deletion	DEL00291822
7	79428569	79428614	45	deletion	DEL00320977
7	154467755	154468189	434	deletion	DEL00338865
8	29125786	29126102	316	deletion	DEL00350267

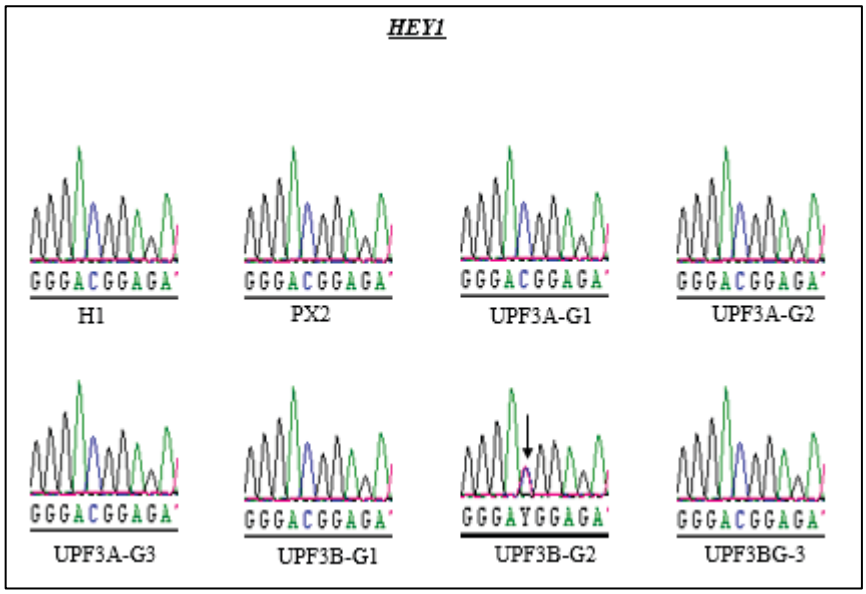
Appendix 3: Unique *de novo* SVs that overlap with predicted off-targets.

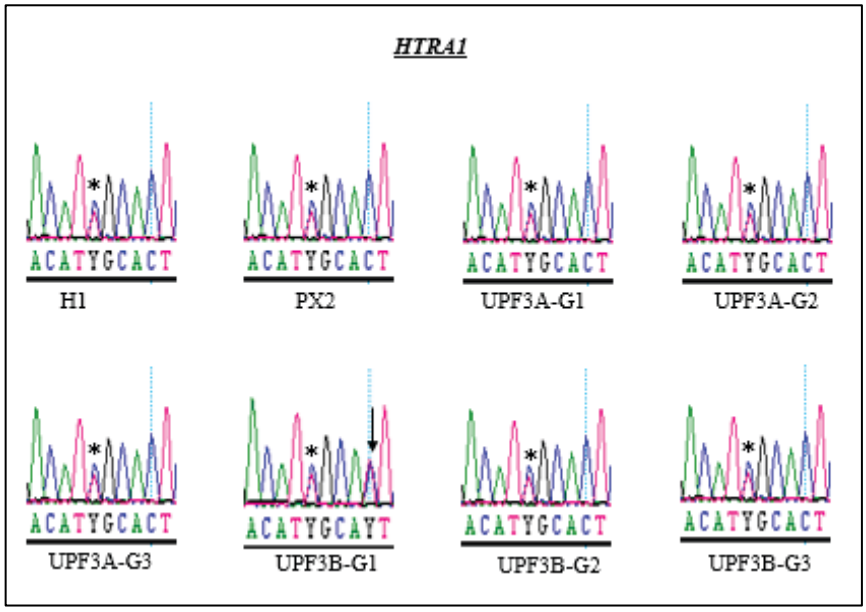

AHCYL1
H1
PX2
UPF3A-G1
UPF3A-G2
UPF3A-G3
UPF3B-G1
UPF3B-G2
UPF3B-G3
ANKDR13A
H1
PX2
UPF3A-G1
UPF3A-G2
UPF3A-G3
UPF3B-G1
UPF3B-G2
UPF3B-G3
CSRNP3
H1
PX2
UPF3A-G1
UPF3A-G2
UPF3A-G3
UPF3B-G1
UPF3B-G2
UPF3B-G3

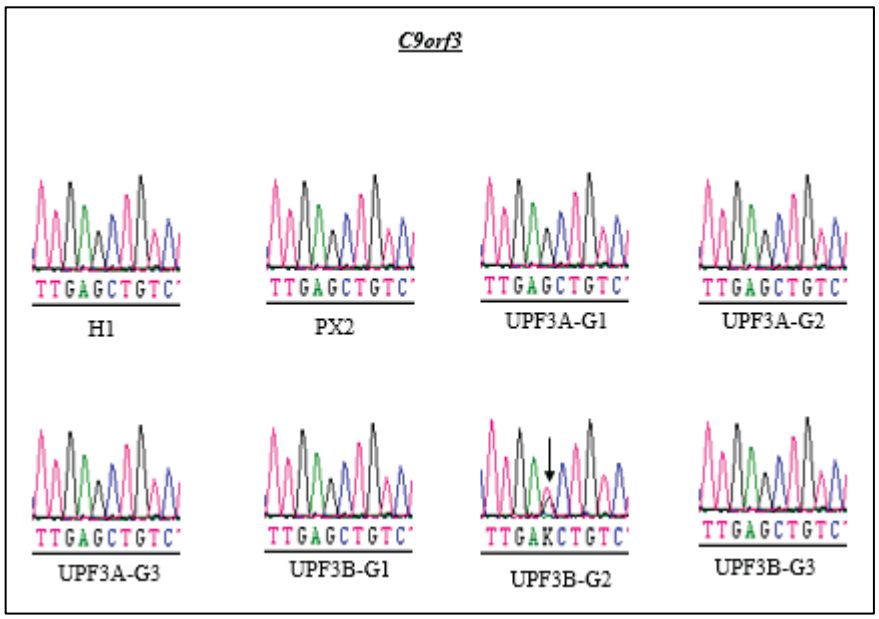

DAO
H1
PX2
UPF3A-G1
UPF3A-G2
UPF3A-G3
UPF3B-G1
UPF3B-G2
UPF3B-G3

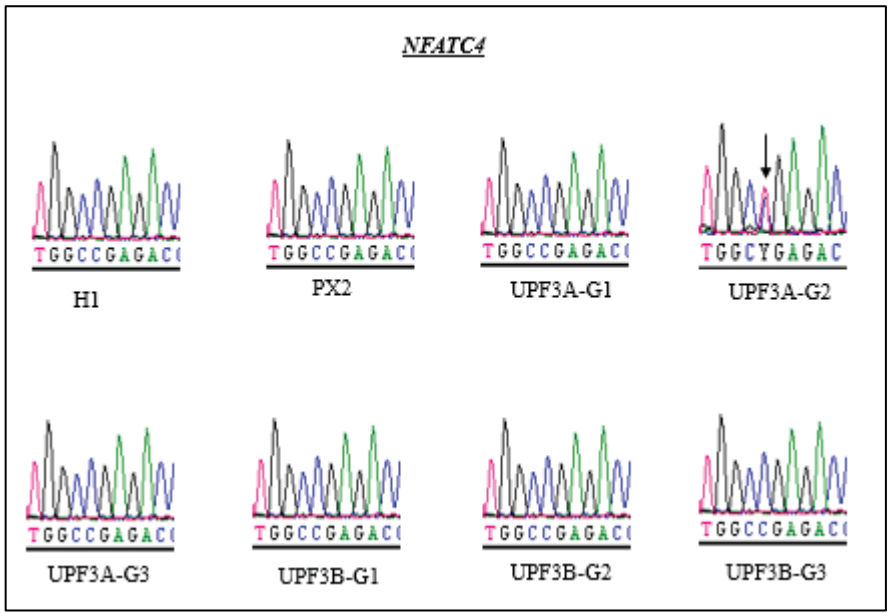

DLGAP3
H1
PX2
UPF3A-G1
UPF3A-G2
UPF3A-G3
UPF3B-G1
UPF3B-G2
UPF3B-G3

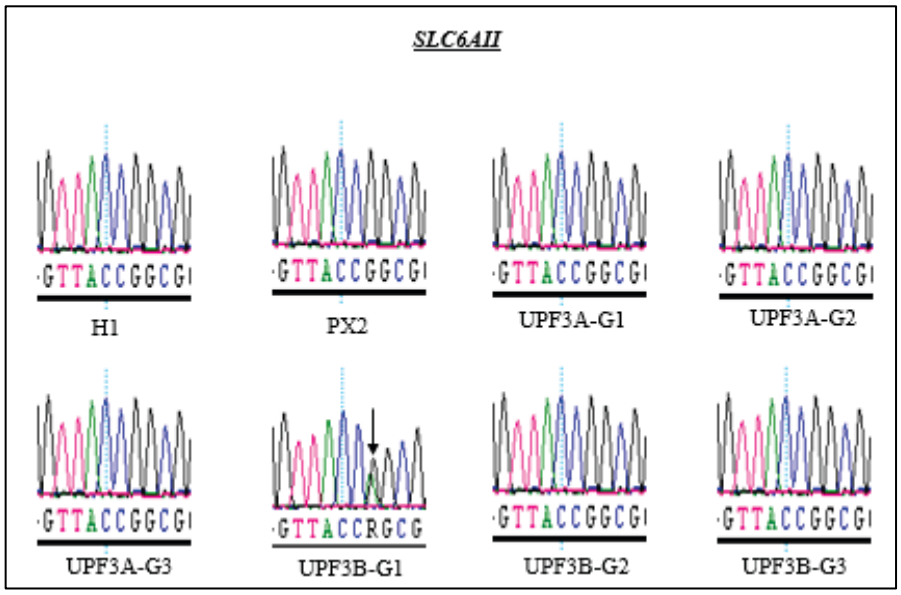

GALNT6
H1
PX2
UPF3A-G1
UPF3A-G2
UPF3A-G3
UPF3B-G1
UPF3B-G2
UPF3B-G3

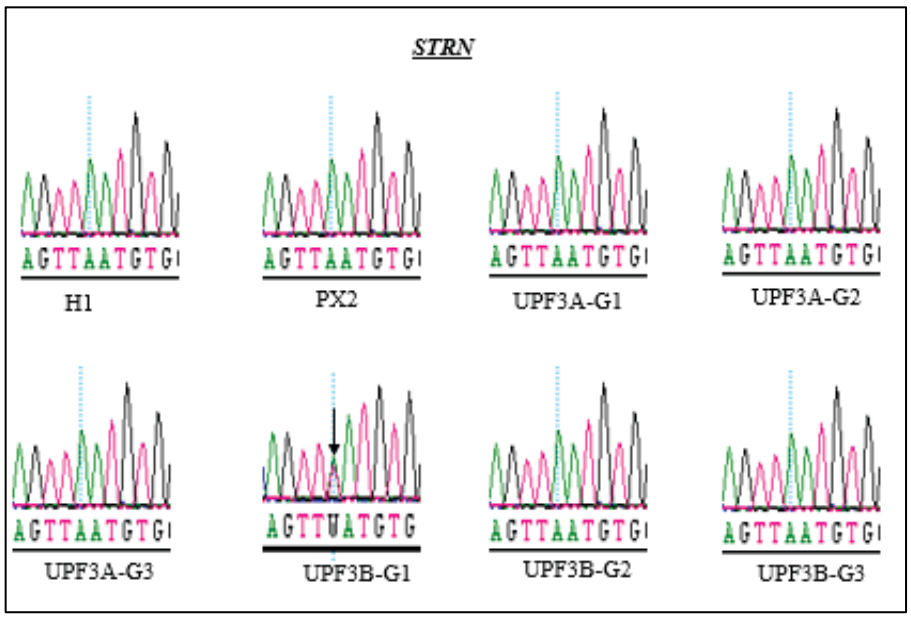

HEY1
GGGACGGAGA
H1
GGGACGGAGA
PX2
GGGACGGAGA
UPF3A-G1
GGGACGGAGA
UPF3A-G2
GGGACGGAGA
UPF3A-G3
GGGACGGAGA
UPF3B-G1
GGGATGGAGA
UPF3B-G2
GGGACGGAGA
UPF3BG-3

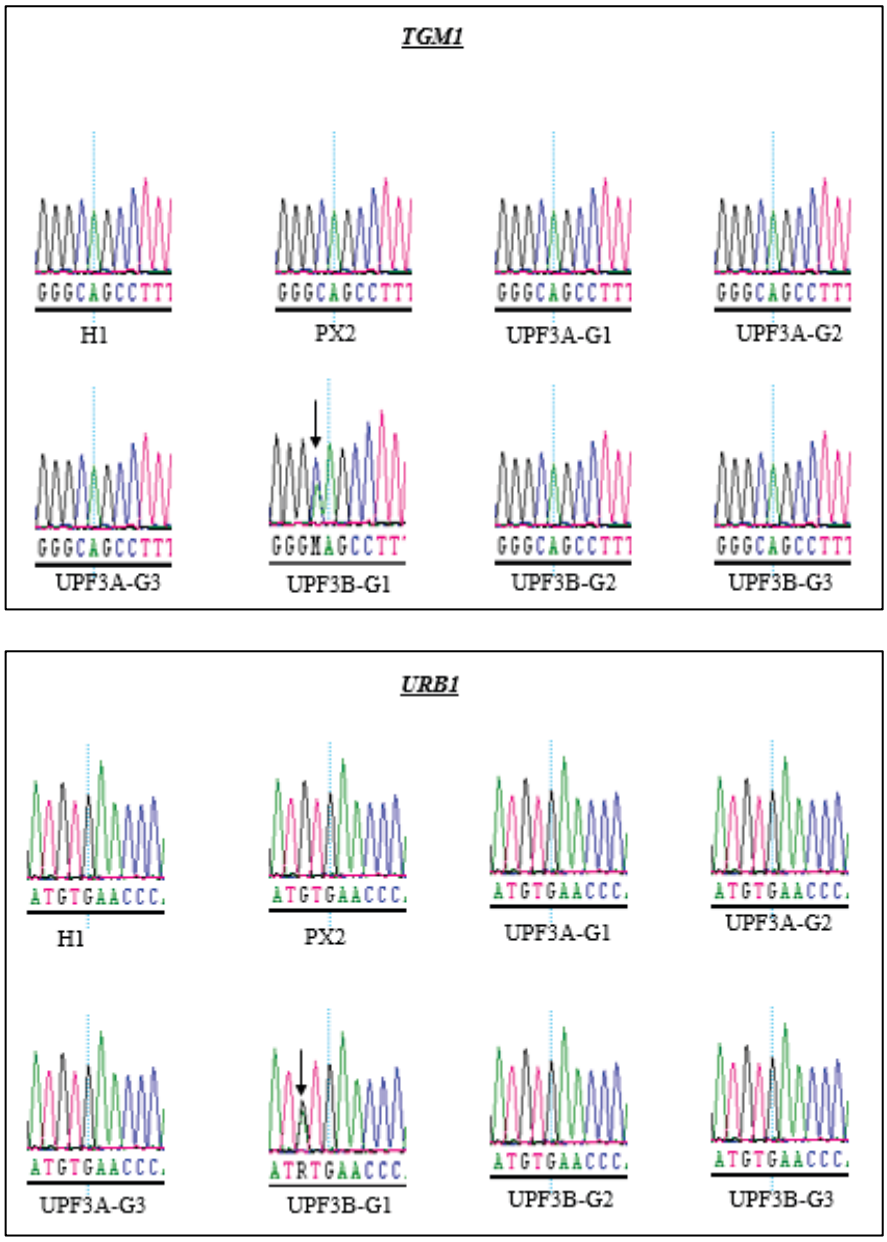

HTRA1
ACATYGCACT
H1
ACATYGCACT
PX2
ACATYGCACT
UPF3A-G1
ACATYGCACT
UPF3A-G2
ACATYGCACT
UPF3A-G3
ACATYGCAYT
UPF3B-G1
ACATYGCACT
UPF3B-G2
ACATYGCACT
UPF3B-G3


C9orf3
TTGAGCTGTC
H1
TTGAGCTGTC
PX2
TTGAGCTGTC
UPF3A-G1
TTGAGCTGTC
UPF3A-G2
TTGAGCTGTC
UPF3A-G3
TTGAGCTGTC
UPF3B-G1
TTGAKCTGTC
UPF3B-G2
TTGAGCTGTC
UPF3B-G3


NFATC4
H1
PX2
UPF3A-G1
UPF3A-G2
UPF3A-G3
UPF3B-G1
UPF3B-G2
UPF3B-G3


SLC6A11
H1
PX2
UPF3A-G1
UPF3A-G2
UPF3A-G3
UPF3B-G1
UPF3B-G2
UPF3B-G3


STRN
H1
PX2
UPF3A-G1
UPF3A-G2
UPF3A-G3
UPF3B-G1
UPF3B-G2
UPF3B-G3



Appendix 4: Whole genome sequencing validation.
H1: parental cell line. PX2: Control cell line

UPF3B KO hESCs	*UPF3A* KO NSCs	*UPF3B* KO NSCs
SYT3	COL1A1	MEGF6
EGR4	HOXB2	TRMT9B
FAM229A	TGFBI	APOM
DUSP8	IGFBP7	SLC1A2
EGFL8	ITGA1	LMO3
LIF	TNC	BLOC1S5-TXNDC5
CLEC11A	INHBA	ITGA2B
VAV1	FAP	CNR1
NFKBID	COL3A1	CKMT1B
ZFP36	ADAMTS14	ROBO2
FAM222A	CDKN2B	LINGO1
KCNA1	EDNRA	RELN
PCDH10	HTRA1	SYT10
CDKN1A	HAS2	ACACB
FOSL1	BNC2	TEK
TMEM132E	CXCL6	FCHO1
SPRY4	COL6A2	ELN
ISG20	HOXB5	LOC400499
CYP1B1	PAMR1	GADD45G
LRRC4B	PLPP4	ANO5
ST8SIA3	COL1A2	ACSBG1
EGR2	CFH	LRRTM3
ATP2A1	CDH11	DLL1
WNT2B	NFIX	PELI2
SLA	TNFAIP6	OLFM3
C19orf81	EGFL8	FBN3
C3	FMN1	ONECUT2
SAP25	MXRA5	ARX
LIME1	SRPX2	CELF3
HSPA12A	MFAP4	
GDF15	FXYD5	
RING1	RUNX1	
NPAS4	GXYLT2	
FOXD3	CEBPD	
RGS11	HOXB3	
SHANK1	COL16A1	
ZNF648	TGFB2	
GRIN2A	WISP1	
VWA7	TNFSF4	
NXPH2	GPNMB	

Appendix 5: Table showing downregulated genes in *UPF3A* KO NSCs and *UPF3B* KO hESCs and NSCs.

UPF3B KO hESCs	*UPF3A* KO NSCs	*UPF3B* KO NSCs
LCP1	RAB11FIP4	PCDHA6
TXNIP	OTX1	PCDHB5
POMZP3	ZNF813	RHEX
LOC645513	DLGAP3	FIBIN
PCDHB12	RIMS4	PLCG2
PCDHA6	ONECUT2	HOXB2
GOLGA2P10	ONECUT1	IRAK1
LOC730268		SEMA3C
PCDHB3		OGFRL1
PCDHA3		
SCG2		
PIWIL2		
NBPF9		
PCDHA7		
PCDHA1		
GADD45G		
TRDN		
PKIB		
PCDHGB3		
PCDHB18P		
NETO1		
GABBR1		
PCDHGA8		
GREM1		
EIF4E1B		
GSAP		
PCDHGA5		
SLITRK1		
UNC5C		
ZBED6CL		
ADGRL3		

Appendix 6: Table showing upregulated genes in *UPF3A* KO NSCs and *UPF3B* KO hESCs and NSCs.

ID	Description	DERatio	BGRatio	adjp
GO:0030198	extracellular matrix organization	61/395	240/11586	4.02E-33
GO:0043062	extracellular structure organization	62/395	273/11586	1.23E-30
GO:0001568	blood vessel development	68/395	477/11586	4.59E-21
GO:0030155	regulation of cell adhesion	61/395	449/11586	1.95E-17
GO:0048514	blood vessel morphogenesis	58/395	414/11586	4.39E-17
GO:0001525	angiogenesis	50/395	349/11586	9.92E-15
GO:0001501	skeletal system development	48/395	346/11586	1.89E-13
GO:0032963	collagen metabolic process	22/395	65/11586	3.49E-13
GO:0090287	regulation of cellular response to growth factor stimulus	35/395	201/11586	3.71E-12
GO:0048705	skeletal system morphogenesis	31/395	157/11586	4.55E-12
GO:0090092	regulation of transmembrane receptor protein serine/threonine kinase signaling pathway	31/395	167/11586	2.77E-11
GO:1901342	regulation of vasculature development	35/395	219/11586	5.60E-11
GO:0007178	transmembrane receptor protein serine/threonine kinase signaling pathway	37/395	245/11586	6.07E-11
GO:0008285	negative regulation of cell proliferation	52/395	474/11586	1.81E-10
GO:0032101	regulation of response to external stimulus	50/395	446/11586	2.51E-10
GO:0030199	collagen fibril organization	16/395	40/11586	2.76E-10
GO:0045765	regulation of angiogenesis	32/395	196/11586	4.31E-10
GO:0061448	connective tissue development	31/395	186/11586	5.96E-10
GO:0007162	negative regulation of cell adhesion	30/395	179/11586	1.28E-09
GO:0031589	cell-substrate adhesion	36/395	256/11586	1.29E-09
GO:0051272	positive regulation of cellular component movement	44/395	378/11586	2.47E-09
GO:2000147	positive regulation of cell motility	43/395	368/11586	4.04E-09
GO:0060349	bone morphogenesis	20/395	80/11586	5.67E-09
GO:0002062	chondrocyte differentiation	21/395	91/11586	8.40E-09
GO:0040017	positive regulation of locomotion	44/395	393/11586	9.54E-09
GO:0060348	bone development	26/395	147/11586	1.41E-08
GO:0048729	tissue morphogenesis	47/395	447/11586	1.57E-08
GO:0071559	response to transforming growth factor beta	29/395	186/11586	2.10E-08
GO:0007492	endoderm development	16/395	51/11586	2.23E-08
GO:0007160	cell-matrix adhesion	27/395	164/11586	3.12E-08
GO:0030335	positive regulation of cell migration	41/395	360/11586	3.25E-08
GO:0051216	cartilage development	25/395	143/11586	4.85E-08
GO:0035987	endodermal cell differentiation	13/395	32/11586	5.01E-08
GO:0001503	ossification	35/395	275/11586	5.22E-08
GO:0071560	cellular response to transforming growth factor beta stimulus	28/395	181/11586	5.98E-08
GO:0042060	wound healing	39/395	338/11586	7.27E-08

GO:0007179	transforming growth factor beta receptor signaling pathway	25/395	147/11586	9.09E-08
GO:0006954	inflammatory response	39/395	347/11586	1.63E-07
GO:0009611	response to wounding	43/395	412/11586	1.77E-07
GO:0001706	endoderm formation	13/395	35/11586	1.94E-07
GO:0060485	mesenchyme development	28/395	190/11586	1.96E-07
GO:0017015	regulation of transforming growth factor beta receptor signaling pathway	19/395	90/11586	5.02E-07
GO:0048598	embryonic morphogenesis	42/395	409/11586	5.06E-07
GO:0003416	endochondral bone growth	12/395	31/11586	6.16E-07
GO:0045766	positive regulation of angiogenesis	21/395	113/11586	6.69E-07
GO:0060350	endochondral bone morphogenesis	15/395	54/11586	7.24E-07
GO:1903844	regulation of cellular response to transforming growth factor beta stimulus	19/395	92/11586	7.51E-07
GO:0098868	bone growth	12/395	32/11586	9.56E-07
GO:1904018	positive regulation of vasculature development	22/395	127/11586	1.02E-06
GO:0061061	muscle structure development	43/395	443/11586	1.80E-06
GO:0045785	positive regulation of cell adhesion	31/395	252/11586	1.91E-06
GO:0048704	embryonic skeletal system morphogenesis	15/395	59/11586	2.86E-06
GO:0050900	leukocyte migration	28/395	213/11586	2.96E-06
GO:0048706	embryonic skeletal system development	17/395	79/11586	3.49E-06
GO:0022407	regulation of cell-cell adhesion	29/395	231/11586	4.38E-06
GO:0001655	urogenital system development	30/395	248/11586	5.53E-06
GO:0001667	ameboidal-type cell migration	33/395	294/11586	5.72E-06
GO:0032964	collagen biosynthetic process	11/395	30/11586	7.21E-06
GO:0060389	pathway-restricted SMAD protein phosphorylation	13/395	46/11586	9.50E-06
GO:0001704	formation of primary germ layer	17/395	84/11586	9.54E-06
GO:0002683	negative regulation of immune system process	30/395	254/11586	9.82E-06
GO:0001816	cytokine production	40/395	417/11586	1.08E-05
GO:0048762	mesenchymal cell differentiation	23/395	158/11586	1.38E-05
GO:0090288	negative regulation of cellular response to growth factor stimulus	19/395	109/11586	1.54E-05
GO:0060351	cartilage development involved in endochondral bone morphogenesis	11/395	32/11586	1.60E-05
GO:0050673	epithelial cell proliferation	30/395	261/11586	1.87E-05
GO:0050727	regulation of inflammatory response	24/395	174/11586	1.89E-05
GO:0010712	regulation of collagen metabolic process	10/395	26/11586	2.32E-05
GO:0033627	cell adhesion mediated by integrin	12/395	41/11586	2.53E-05
GO:1904888	cranial skeletal system development	13/395	50/11586	2.93E-05
GO:0090130	tissue migration	27/395	221/11586	3.03E-05

GO:0030855	epithelial cell differentiation	37/395	382/11586	3.38E-05
GO:0042476	odontogenesis	16/395	80/11586	3.39E-05
GO:0010718	positive regulation of epithelial to mesenchymal transition	12/395	42/11586	3.44E-05
GO:0010717	regulation of epithelial to mesenchymal transition	15/395	70/11586	3.67E-05
GO:0072001	renal system development	27/395	223/11586	3.68E-05
GO:0001837	epithelial to mesenchymal transition	18/395	104/11586	4.46E-05
GO:0060393	regulation of pathway-restricted SMAD protein phosphorylation	12/395	43/11586	4.62E-05
GO:0002063	chondrocyte development	11/395	35/11586	4.72E-05
GO:0010469	regulation of signaling receptor activity	27/395	230/11586	7.18E-05
GO:0048701	embryonic cranial skeleton morphogenesis	11/395	37/11586	9.10E-05
GO:2000027	regulation of animal organ morphogenesis	21/395	147/11586	9.35E-05
GO:0045596	negative regulation of cell differentiation	42/395	489/11586	0.00011
GO:0010810	regulation of cell-substrate adhesion	22/395	162/11586	0.0001112
GO:0070482	response to oxygen levels	29/395	266/11586	0.0001124
GO:0003418	growth plate cartilage chondrocyte differentiation	8/395	17/11586	0.0001186
GO:0003433	chondrocyte development involved in endochondral bone morphogenesis	8/395	17/11586	0.0001186
GO:0022408	negative regulation of cell-cell adhesion	17/395	101/11586	0.0001755
GO:0050678	regulation of epithelial cell proliferation	26/395	226/11586	0.0002006
GO:0030512	negative regulation of transforming growth factor beta receptor signaling pathway	13/395	58/11586	0.0002031
GO:0001822	kidney development	25/395	211/11586	0.0002032
GO:0097529	myeloid leukocyte migration	16/395	91/11586	0.000232
GO:0003417	growth plate cartilage development	9/395	25/11586	0.0002687
GO:0050865	regulation of cell activation	30/395	294/11586	0.0002896
GO:1903845	negative regulation of cellular response to transforming growth factor beta stimulus	13/395	60/11586	0.0003125
GO:0050866	negative regulation of cell activation	17/395	105/11586	0.0003175
GO:0007369	gastrulation	19/395	131/11586	0.0003413
GO:0060536	cartilage morphogenesis	8/395	19/11586	0.0003474
GO:0010631	epithelial cell migration	25/395	219/11586	0.0004239
GO:0090132	epithelium migration	25/395	219/11586	0.0004239
GO:0043065	positive regulation of apoptotic process	40/395	476/11586	0.0004348
GO:0043068	positive regulation of programmed cell death	40/395	478/11586	0.0004865

GO:0001818	negative regulation of cytokine production	20/395	150/11586	0.0006488
GO:0001817	regulation of cytokine production	34/395	376/11586	0.0007403
GO:0090101	negative regulation of transmembrane receptor protein serine/threonine kinase signaling pathway	15/395	87/11586	0.0007875
GO:0003413	chondrocyte differentiation involved in endochondral bone morphogenesis	8/395	21/11586	0.0008806
GO:0009612	response to mechanical stimulus	19/395	139/11586	0.0008901
GO:0006935	chemotaxis	35/395	400/11586	0.001078
GO:0042330	taxis	35/395	402/11586	0.001216
GO:0032965	regulation of collagen biosynthetic process	8/395	22/11586	0.001343
GO:0031347	regulation of defense response	34/395	386/11586	0.001376
GO:0090100	positive regulation of transmembrane receptor protein serine/threonine kinase signaling pathway	14/395	79/11586	0.001413
GO:0001666	response to hypoxia	25/395	235/11586	0.001653
GO:1903034	regulation of response to wounding	15/395	93/11586	0.001939
GO:0035137	hindlimb morphogenesis	8/395	23/11586	0.001998
GO:0010862	positive regulation of pathway-restricted SMAD protein phosphorylation	9/395	31/11586	0.00221
GO:0042481	regulation of odontogenesis	6/395	11/11586	0.00227
GO:0070663	regulation of leukocyte proliferation	16/395	107/11586	0.002345
GO:0042692	muscle cell differentiation	25/395	240/11586	0.002461
GO:0002526	acute inflammatory response	12/395	60/11586	0.002495
GO:0048771	tissue remodeling	15/395	95/11586	0.002574
GO:0001649	osteoblast differentiation	20/395	164/11586	0.002805
GO:0030574	collagen catabolic process	8/395	24/11586	0.002909
GO:0002009	morphogenesis of an epithelium	32/395	363/11586	0.003025
GO:0036293	response to decreased oxygen levels	25/395	246/11586	0.003906
GO:0061041	regulation of wound healing	13/395	74/11586	0.004104
GO:0003414	chondrocyte morphogenesis involved in endochondral bone morphogenesis	6/395	12/11586	0.00441
GO:0003429	growth plate cartilage chondrocyte morphogenesis	6/395	12/11586	0.00441
GO:0090171	chondrocyte morphogenesis	6/395	12/11586	0.00441
GO:0070661	leukocyte proliferation	18/395	140/11586	0.004638
GO:0031214	biomineral tissue development	14/395	87/11586	0.004793
GO:0048562	embryonic organ morphogenesis	21/395	185/11586	0.004892
GO:0043542	endothelial cell migration	19/395	155/11586	0.004912
GO:0007507	heart development	34/395	409/11586	0.00522
GO:0050670	regulation of lymphocyte proliferation	15/395	101/11586	0.005758

GO:0060317	cardiac epithelial to mesenchymal transition	8/395	26/11586	0.005819
GO:0010632	regulation of epithelial cell migration	20/395	172/11586	0.006
GO:0110110	positive regulation of animal organ morphogenesis	11/395	54/11586	0.006045
GO:0032944	regulation of mononuclear cell proliferation	15/395	102/11586	0.006544
GO:0072006	nephron development	15/395	102/11586	0.006544
GO:0051146	striated muscle cell differentiation	20/395	173/11586	0.006575
GO:0071407	cellular response to organic cyclic compound	33/395	395/11586	0.00673
GO:0060562	epithelial tube morphogenesis	24/395	237/11586	0.006861
GO:0046651	lymphocyte proliferation	17/395	130/11586	0.007229
GO:0060326	cell chemotaxis	18/395	145/11586	0.007774
GO:0071229	cellular response to acid chemical	18/395	145/11586	0.007774
GO:0003422	growth plate cartilage morphogenesis	6/395	13/11586	0.007955
GO:0048589	developmental growth	37/395	474/11586	0.007982
GO:0032943	mononuclear cell proliferation	17/395	131/11586	0.008052
GO:0048568	embryonic organ development	27/395	291/11586	0.008549
GO:0007517	muscle organ development	26/395	274/11586	0.00871
GO:0050729	positive regulation of inflammatory response	11/395	56/11586	0.008844
GO:0061333	renal tubule morphogenesis	11/395	56/11586	0.008844
GO:0010634	positive regulation of epithelial cell migration	16/395	118/11586	0.008938
GO:1903037	regulation of leukocyte cell-cell adhesion	19/395	163/11586	0.01055
GO:0072073	kidney epithelium development	15/395	106/11586	0.01074
GO:0072009	nephron epithelium development	13/395	82/11586	0.01366
GO:0072132	mesenchyme morphogenesis	9/395	38/11586	0.01446
GO:0022617	extracellular matrix disassembly	10/395	48/11586	0.01459
GO:0097530	granulocyte migration	10/395	48/11586	0.01459
GO:0007596	blood coagulation	21/395	198/11586	0.01468
GO:0034330	cell junction organization	23/395	231/11586	0.01504
GO:0006936	muscle contraction	21/395	200/11586	0.01723
GO:0050817	coagulation	21/395	200/11586	0.01723
GO:0035265	organ growth	16/395	124/11586	0.0173
GO:1905207	regulation of cardiocyte differentiation	9/395	39/11586	0.01824
GO:0061326	renal tubule development	12/395	72/11586	0.01893
GO:0007423	sensory organ development	31/395	377/11586	0.01939
GO:0007599	hemostasis	21/395	202/11586	0.02016
GO:0002685	regulation of leukocyte migration	14/395	98/11586	0.02056
GO:0003148	outflow tract septum morphogenesis	7/395	22/11586	0.02087
GO:1905209	positive regulation of cardiocyte differentiation	7/395	22/11586	0.02087
GO:0071772	response to BMP	15/395	112/11586	0.02158
GO:0071773	cellular response to BMP stimulus	15/395	112/11586	0.02158

GO:0060021	roof of mouth development	12/395	73/11586	0.02197
GO:0002694	regulation of leukocyte activation	25/395	271/11586	0.02259
GO:0050707	regulation of cytokine secretion	13/395	86/11586	0.02355
GO:0035107	appendage morphogenesis	15/395	113/11586	0.02412
GO:0035108	limb morphogenesis	15/395	113/11586	0.02412
GO:0050710	negative regulation of cytokine secretion	8/395	31/11586	0.02529
GO:2000826	regulation of heart morphogenesis	8/395	31/11586	0.02529
GO:0003007	heart morphogenesis	19/395	173/11586	0.02554
GO:0034329	cell junction assembly	20/395	189/11586	0.02584
GO:0061138	morphogenesis of a branching epithelium	16/395	129/11586	0.02904
GO:0035270	endocrine system development	12/395	75/11586	0.02935
GO:0030510	regulation of BMP signaling pathway	11/395	63/11586	0.02951
GO:1901654	response to ketone	16/395	130/11586	0.0321
GO:0010714	positive regulation of collagen metabolic process	6/395	16/11586	0.03402
GO:0032967	positive regulation of collagen biosynthetic process	6/395	16/11586	0.03402
GO:0030282	bone mineralization	11/395	64/11586	0.03457
GO:0048545	response to steroid hormone	25/395	278/11586	0.03534
GO:0038065	collagen-activated signaling pathway	5/395	10/11586	0.03691
GO:0002695	negative regulation of leukocyte activation	13/395	90/11586	0.03931
GO:0001656	metanephros development	11/395	65/11586	0.04035
GO:0032102	negative regulation of response to external stimulus	20/395	195/11586	0.04136
GO:0030509	BMP signaling pathway	14/395	104/11586	0.04161
GO:0071396	cellular response to lipid	31/395	392/11586	0.04266
GO:0002040	sprouting angiogenesis	13/395	91/11586	0.04446
GO:0007159	leukocyte cell-cell adhesion	19/395	180/11586	0.04545
GO:0030278	regulation of ossification	16/395	134/11586	0.04743

Appendix 7: ***UPF3A*** **KO NSCs Biological Process ontology.**

ID	Description	DERatio	BGRatio	adjp
GO:0005201	extracellular matrix structural constituent	39/395	99/11706	2.73E-29
GO:0030020	extracellular matrix structural constituent conferring tensile strength	17/395	30/11706	2.73E-15
GO:0019838	growth factor binding	27/395	101/11706	1.01E-14
GO:0005539	glycosaminoglycan binding	25/395	122/11706	1.24E-10
GO:0019955	cytokine binding	19/395	66/11706	1.43E-10
GO:0005178	integrin binding	19/395	88/11706	3.72E-08
GO:0005126	cytokine receptor binding	21/395	119/11706	2.03E-07
GO:0038023	signaling receptor activity	45/395	480/11706	2.06E-07
GO:0005125	cytokine activity	14/395	58/11706	2.59E-06
GO:0030545	receptor regulator activity	25/395	193/11706	3.63E-06
GO:0005160	transforming growth factor beta receptor binding	11/395	34/11706	4.01E-06
GO:0008201	heparin binding	17/395	92/11706	4.70E-06
GO:0048407	platelet-derived growth factor binding	7/395	11/11706	6.95E-06
GO:0050840	extracellular matrix binding	12/395	44/11706	7.34E-06
GO:0048018	receptor ligand activity	22/395	167/11706	2.17E-05
GO:0005518	collagen binding	12/395	50/11706	3.52E-05
GO:0004888	transmembrane signaling receptor activity	32/395	348/11706	0.000125
GO:0050839	cell adhesion molecule binding	35/395	406/11706	0.000162
GO:0050431	transforming growth factor beta binding	7/395	16/11706	0.000208
GO:1901681	sulfur compound binding	19/395	160/11706	0.000915
GO:0000978	RNA polymerase II proximal promoter sequence-specific DNA binding	30/395	367/11706	0.003395
GO:0000987	proximal promoter sequence-specific DNA binding	30/395	376/11706	0.00547
GO:0098631	cell adhesion mediator activity	9/395	47/11706	0.01133
GO:0000982	transcription factor activity, RNA polymerase II proximal promoter sequence-specific DNA binding	26/395	315/11706	0.01192
GO:0001228	DNA-binding transcription activator activity, RNA polymerase II-specific	24/395	286/11706	0.01882
GO:0002020	protease binding	12/395	89/11706	0.02095
GO:0004252	serine-type endopeptidase activity	10/395	63/11706	0.0221
GO:0001077	proximal promoter DNA-binding transcription activator activity, RNA polymerase II-specific	19/395	201/11706	0.02483
GO:0005509	calcium ion binding	31/395	439/11706	0.04125
GO:0043394	proteoglycan binding	6/395	23/11706	0.04411

Appendix 8: ***UPF3A*** **KO NSCs Molecular Function ontology.**

ID	Description	DERatio	BGRatio	adjp
GO:0031012	extracellular matrix	71/407	305/12152	6.85E-38
GO:0062023	collagen-containing extracellular matrix	56/407	215/12152	4.02E-32
GO:0005581	collagen trimer	22/407	51/12152	4.47E-17
GO:0005788	endoplasmic reticulum lumen	39/407	214/12152	7.90E-16
GO:0009986	cell surface	51/407	442/12152	2.09E-12
GO:0044420	extracellular matrix component	16/407	41/12152	3.10E-11
GO:0005912	adherens junction	47/407	481/12152	1.09E-08
GO:0070161	anchoring junction	47/407	491/12152	2.23E-08
GO:0005925	focal adhesion	37/407	372/12152	9.71E-07
GO:0005924	cell-substrate adherens junction	37/407	373/12152	1.05E-06
GO:0030055	cell-substrate junction	37/407	376/12152	1.30E-06
GO:0005604	basement membrane	15/407	71/12152	3.30E-06
GO:0005583	fibrillar collagen trimer	7/407	11/12152	4.57E-06
GO:0098643	banded collagen fibril	7/407	11/12152	4.57E-06
GO:0098644	complex of collagen trimers	8/407	17/12152	9.54E-06
GO:0043235	receptor complex	27/407	241/12152	1.26E-05
GO:0042383	sarcolemma	17/407	102/12152	1.47E-05
GO:0005901	caveola	11/407	57/12152	0.0008279
GO:0009897	external side of plasma membrane	16/407	130/12152	0.002382
GO:0044853	plasma membrane raft	11/407	73/12152	0.009868
GO:0098797	plasma membrane protein complex	27/407	343/12152	0.0118
GO:0008305	integrin complex	6/407	20/12152	0.01224
GO:0098636	protein complex involved in cell adhesion	6/407	20/12152	0.01224
GO:0005913	cell-cell adherens junction	12/407	91/12152	0.0169
GO:0045121	membrane raft	20/407	222/12152	0.01962
GO:0098857	membrane microdomain	20/407	223/12152	0.0209
GO:0098589	membrane region	20/407	230/12152	0.03215

Appendix 9: *UPF3A* KO NSC Cellular component ontology.

ID	Description	DERatio	BGRatio	adjp
GO:0099055	integral component of postsynaptic membrane	7/81	81/12598	0.00014
GO:0098936	intrinsic component of postsynaptic membrane	7/81	86/12598	0.00021
GO:0099061	integral component of postsynaptic density membrane	5/81	38/12598	0.00068
GO:0099699	integral component of synaptic membrane	7/81	106/12598	0.00085
GO:0099146	intrinsic component of postsynaptic density membrane	5/81	41/12598	0.00101
GO:0099240	intrinsic component of synaptic membrane	7/81	115/12598	0.00145
GO:0099060	integral component of postsynaptic specialization membrane	5/81	52/12598	0.0033
GO:0045211	postsynaptic membrane	1/9	243/12598	0.00412
GO:0098948	intrinsic component of postsynaptic specialization membrane	5/81	55/12598	0.00435
GO:0098839	postsynaptic density membrane	5/81	58/12598	0.00565
GO:0097060	synaptic membrane	10/81	326/12598	0.00717
GO:0042734	presynaptic membrane	2/27	115/12598	0.01567
GO:0099634	postsynaptic specialization membrane	5/81	74/12598	0.01833
GO:0098978	glutamatergic synapse	1/9	298/12598	0.02

Appendix 10: *UPF3B* KO hESCs cellular component ontology.

ID	Description	DERatio	BGRatio	adjp
GO:0042330	taxis	209/3770	440/12553	1.97E-11
GO:0006935	chemotaxis	208/3770	438/12553	2.37E-11
GO:0097485	neuron projection guidance	122/3770	225/12553	1.22E-10
GO:0007411	axon guidance	121/3770	224/12553	2.25E-10
GO:0061564	axon development	200/3770	432/12553	1.83E-09
GO:0007186	G protein-coupled receptor signaling pathway	223/3770	498/12553	5.43E-09
GO:0007409	axonogenesis	182/3770	394/12553	2.79E-08
GO:0007389	pattern specification process	148/3770	310/12553	1.42E-07
GO:0048667	cell morphogenesis involved in neuron differentiation	215/3770	493/12553	3.00E-07
GO:0007610	behavior	197/3770	449/12553	1.05E-06
GO:0007600	sensory perception	148/3770	317/12553	1.09E-06
GO:0003013	circulatory system process	152/3770	339/12553	2.17E-05
GO:0007423	sensory organ development	181/3770	419/12553	2.51E-05
GO:0030198	extracellular matrix organization	126/3770	270/12553	2.53E-05
GO:0003002	regionalization	113/3770	236/12553	2.60E-05
GO:0043062	extracellular structure organization	139/3770	306/12553	3.65E-05
GO:0008015	blood circulation	148/3770	332/12553	5.58E-05
GO:0043269	regulation of ion transport	194/3770	460/12553	6.86E-05
GO:0098742	cell-cell adhesion via plasma-membrane adhesion molecules	101/3770	210/12553	0.000127
GO:0001655	urogenital system development	121/3770	263/12553	0.000142
GO:0010469	regulation of signaling receptor activity	125/3770	275/12553	0.0002
GO:0001501	skeletal system development	158/3770	366/12553	0.000255
GO:0090596	sensory organ morphogenesis	94/3770	196/12553	0.000459
GO:0045165	cell fate commitment	80/3770	161/12553	0.000637
GO:0072001	renal system development	109/3770	237/12553	0.000706
GO:0048598	embryonic morphogenesis	185/3770	449/12553	0.001102
GO:0090066	regulation of anatomical structure size	165/3770	393/12553	0.00121
GO:0050808	synapse organization	145/3770	339/12553	0.001776
GO:0055123	digestive system development	56/3770	104/12553	0.001778
GO:0030900	forebrain development	133/3770	306/12553	0.001851
GO:0001822	kidney development	103/3770	225/12553	0.002057
GO:0050890	cognition	106/3770	234/12553	0.002643
GO:0099177	regulation of trans-synaptic signaling	148/3770	350/12553	0.003066
GO:0007156	homophilic cell adhesion via plasma membrane adhesion molecules	69/3770	138/12553	0.003451
GO:0050804	modulation of chemical synaptic transmission	147/3770	349/12553	0.004406
GO:0048729	tissue morphogenesis	194/3770	483/12553	0.004704
GO:0043010	camera-type eye development	107/3770	239/12553	0.004717
GO:0043588	skin development	87/3770	186/12553	0.005332
GO:0048592	eye morphogenesis	60/3770	117/12553	0.006226
GO:0048568	embryonic organ development	136/3770	320/12553	0.006285
GO:0048565	digestive tract development	51/3770	95/12553	0.006666
GO:0061448	connective tissue development	90/3770	195/12553	0.00715

GO:0007187	G protein-coupled receptor signaling pathway, coupled to cyclic nucleotide second messenger	63/3770	125/12553	0.007381
GO:0010721	negative regulation of cell development	116/3770	267/12553	0.01055
GO:0030855	epithelial cell differentiation	178/3770	442/12553	0.01094
GO:0032970	regulation of actin filament-based process	135/3770	321/12553	0.01357
GO:0051216	cartilage development	72/3770	150/12553	0.014
GO:0050673	epithelial cell proliferation	123/3770	288/12553	0.01504
GO:0048562	embryonic organ morphogenesis	93/3770	206/12553	0.01562
GO:0048880	sensory system development	121/3770	283/12553	0.01692
GO:0051961	negative regulation of nervous system development	110/3770	253/12553	0.0185
GO:0009952	anterior/posterior pattern specification	67/3770	138/12553	0.01881
GO:0001654	eye development	120/3770	281/12553	0.01993
GO:0150063	visual system development	120/3770	281/12553	0.01993
GO:0007611	learning or memory	93/3770	207/12553	0.02009
GO:0007015	actin filament organization	141/3770	341/12553	0.02465
GO:0042391	regulation of membrane potential	136/3770	327/12553	0.02526
GO:0035637	multicellular organismal signaling	63/3770	129/12553	0.02856
GO:0007416	synapse assembly	68/3770	142/12553	0.02916
GO:0007626	locomotory behavior	72/3770	153/12553	0.03423
GO:0032956	regulation of actin cytoskeleton organization	120/3770	284/12553	0.03702
GO:0032102	negative regulation of response to external stimulus	97/3770	221/12553	0.04261
GO:0001505	regulation of neurotransmitter levels	114/3770	268/12553	0.04311
GO:0042063	gliogenesis	95/3770	216/12553	0.04751

Appendix 11: **Shared genes (hESCs and NSCs): Biological process ontology.**

ID	Description	DERatio	BGRatio	adjp
GO:0004888	transmembrane signaling receptor activity	213/3776	437/12668	2.20E-14
GO:0022838	substrate-specific channel activity	151/3776	294/12668	4.45E-12
GO:0005216	ion channel activity	149/3776	289/12668	4.46E-12
GO:0015267	channel activity	155/3776	305/12668	6.32E-12
GO:0022803	passive transmembrane transporter activity	155/3776	305/12668	6.32E-12
GO:0022839	ion gated channel activity	120/3776	234/12668	3.39E-09
GO:0046873	metal ion transmembrane transporter activity	148/3776	306/12668	4.01E-09
GO:0022836	gated channel activity	123/3776	242/12668	4.02E-09
GO:0005261	cation channel activity	112/3776	220/12668	3.23E-08
GO:0015276	ligand-gated ion channel activity	57/3776	98/12668	4.64E-06
GO:0022834	ligand-gated channel activity	57/3776	98/12668	4.64E-06
GO:0008324	cation transmembrane transporter activity	192/3776	453/12668	4.93E-06
GO:0022890	inorganic cation transmembrane transporter activity	179/3776	418/12668	6.59E-06
GO:0022843	voltage-gated cation channel activity	54/3776	97/12668	9.20E-05
GO:0005267	potassium channel activity	47/3776	82/12668	0.000191
GO:0048018	receptor ligand activity	94/3776	203/12668	0.000436
GO:0030545	receptor regulator activity	104/3776	231/12668	0.000593
GO:0005249	voltage-gated potassium channel activity	36/3776	59/12668	0.000635
GO:0099094	ligand-gated cation channel activity	43/3776	76/12668	0.00102
GO:0005244	voltage-gated ion channel activity	66/3776	133/12668	0.001145
GO:0022832	voltage-gated channel activity	66/3776	133/12668	0.001145
GO:0015079	potassium ion transmembrane transporter activity	55/3776	107/12668	0.0022
GO:0008509	anion transmembrane transporter activity	96/3776	217/12668	0.004039
GO:0004930	G protein-coupled receptor activity	82/3776	180/12668	0.005131
GO:0015081	sodium ion transmembrane transporter activity	48/3776	92/12668	0.005616
GO:0015077	monovalent inorganic cation transmembrane transporter activity	111/3776	260/12668	0.005765
GO:0003779	actin binding	141/3776	346/12668	0.007225
GO:0005539	glycosaminoglycan binding	64/3776	135/12668	0.0114
GO:0015370	solute:sodium symporter activity	23/3776	35/12668	0.01189
GO:0019838	growth factor binding	53/3776	109/12668	0.0262
GO:0005125	cytokine activity	41/3776	79/12668	0.03123
GO:0030594	neurotransmitter receptor activity	31/3776	55/12668	0.03594
GO:0004112	cyclic-nucleotide phosphodiesterase activity	17/3776	24/12668	0.03945
GO:0098631	cell adhesion mediator activity	30/3776	53/12668	0.04291
GO:0015108	chloride transmembrane transporter activity	32/3776	58/12668	0.04803

Appendix 12: **Shared genes (hESCs and NSCs): Molecular Functions ontology.**

ID	Description	DERatio	BGRatio	adjp
GO:0031012	extracellular matrix	163/3963	339/13183	9.29E-10
GO:0043235	receptor complex	140/3963	281/13183	1.34E-09
GO:0099699	integral component of synaptic membrane	74/3963	123/13183	2.62E-09
GO:0045177	apical part of cell	128/3963	261/13183	4.66E-08
GO:0099240	intrinsic component of synaptic membrane	76/3963	133/13183	4.67E-08
GO:0097060	synaptic membrane	166/3963	362/13183	7.01E-08
GO:0016324	apical plasma membrane	107/3963	210/13183	1.01E-07
GO:0045211	postsynaptic membrane	129/3963	271/13183	4.70E-07
GO:0034702	ion channel complex	109/3963	220/13183	5.70E-07
GO:1902495	transmembrane transporter complex	115/3963	236/13183	7.02E-07
GO:1990351	transporter complex	116/3963	244/13183	3.71E-06
GO:0099055	integral component of postsynaptic membrane	55/3963	96/13183	1.74E-05
GO:0005911	cell-cell junction	160/3963	369/13183	2.02E-05
GO:0098797	plasma membrane protein complex	171/3963	403/13183	4.23E-05
GO:0062023	collagen-containing extracellular matrix	110/3963	238/13183	6.04E-05
GO:0098936	intrinsic component of postsynaptic membrane	56/3963	101/13183	6.17E-05
GO:0034703	cation channel complex	82/3963	166/13183	8.15E-05
GO:0099056	integral component of presynaptic membrane	35/3963	55/13183	0.000181
GO:0099634	postsynaptic specialization membrane	48/3963	85/13183	0.000242
GO:0098982	GABA-ergic synapse	36/3963	58/13183	0.000303
GO:0098793	presynapse	165/3963	399/13183	0.000523
GO:0043025	neuronal cell body	162/3963	391/13183	0.000573
GO:0098889	intrinsic component of presynaptic membrane	37/3963	62/13183	0.000847
GO:0016323	basolateral plasma membrane	76/3963	158/13183	0.000883
GO:0044297	cell body	180/3963	445/13183	0.000939
GO:0098839	postsynaptic density membrane	37/3963	65/13183	0.004067
GO:0031225	anchored component of membrane	51/3963	100/13183	0.00616
GO:0099572	postsynaptic specialization	130/3963	314/13183	0.007127
GO:0098984	neuron to neuron synapse	132/3963	320/13183	0.007595
GO:0099060	integral component of postsynaptic specialization membrane	35/3963	62/13183	0.009285
GO:0098978	glutamatergic synapse	130/3963	317/13183	0.01238
GO:0098948	intrinsic component of postsynaptic specialization membrane	36/3963	65/13183	0.01249
GO:0034705	potassium channel complex	40/3963	75/13183	0.01455
GO:0031253	cell projection membrane	108/3963	259/13183	0.02718
GO:0042734	presynaptic membrane	61/3963	132/13183	0.04316

Appendix 13: Shared genes (hESCs and NSCs): Cellular Processes ontology.

References

Addington, AM, Gauthier, J, Piton, A, Hamdan, FF, Raymond, A, Gogtay, N, Miller, R, Tossell, J, Bakalar, J, Inoff-Germain, G, Gochman, P, Long, R, Rapoport, JL & Rouleau, GA 2011, 'A novel frameshift mutation in UPF3B identified in brothers affected with childhood onset schizophrenia and autism spectrum disorders', *Mol Psychiatry*, vol. 16, no. 3, Mar, pp. 238-239.

Ahmadzadeh, M & Rosenberg, SA 2005, 'TGF-β1 attenuates the acquisition and expression of effector function by tumor antigen-specific human memory CD8 T cells', *The Journal of Immunology*, vol. 174, no. 9, pp. 5215-5223.

Al-Harbi, KS 2012, 'Treatment-resistant depression: therapeutic trends, challenges, and future directions', *Patient preference and adherence*, vol. 6, p. 369.

Alachkar, A, Jiang, D, Harrison, M, Zhou, Y, Chen, G & Mao, Y 2013, 'An EJC factor RBM8a regulates anxiety behaviors', *Curr Mol Med*, vol. 13, no. 6, Jul, pp. 887-899.

Alazami, AM, Awad, SM, Coskun, S, Al-Hassan, S, Hijazi, H, Abdulwahab, FM, Poizat, C & Alkuraya, FS 2015, 'TLE6 mutation causes the earliest known human embryonic lethality', *Genome Biol*, vol. 16, Nov 5, p. 240.

Albers, CA, Paul, DS, Schulze, H, Freson, K, Stephens, JC, Smethurst, PA, Jolley, JD, Cvejic, A, Kostadima, M & Bertone, P 2012, 'Compound inheritance of a low-frequency regulatory SNP and a rare null mutation in exon-junction complex subunit RBM8A causes TAR syndrome', *Nature Genetics*, vol. 44, no. 4, p. 435.

Allegrucci, C, Denning, CN, Burridge, P, Steele, W, Sinclair, KD & Young, LE 2005, 'Human embryonic stem cells as a model for nutritional programming: an evaluation', *Reproductive Toxicology*, vol. 20, no. 3, pp. 353-367.

Allegrucci, C, Wu, Y-Z, Thurston, A, Denning, CN, Priddle, H, Mummery, CL, Ward-van Oostwaard, D, Andrews, PW, Stojkovic, M & Smith, N 2007, 'Restriction landmark genome scanning identifies culture-induced DNA methylation instability in the human embryonic stem cell epigenome', *Human Molecular Genetics*, vol. 16, no. 10, pp. 1253-1268.

Alpy, F, Jivkov, I, Sorokin, L, Klein, A, Arnold, C, Huss, Y, Kedinger, M, Simon-Assmann, P & Lefebvre, O 2005, 'Generation of a conditionally null allele of the laminin alpha1 gene', *Genesis*, vol. 43, no. 2, Oct, pp. 59-70.

Alrahbeni, T, Sartor, F, Anderson, J, Miedzybrodzka, Z, McCaig, C & Muller, B 2015, 'Full UPF3B function is critical for neuronal differentiation of neural stem cells', *Mol Brain*, vol. 8, p. 33.

Amrani, N, Sachs, MS & Jacobson, A 2006, 'Early nonsense: mRNA decay solves a translational problem', *Nat Rev Mol Cell Biol*, vol. 7, no. 6, Jun, pp. 415-425.

Anbalagan, S, Blechman, J, Gliksberg, M, Gordon, L, Rotkopf, R, Dadosh, T, Shimoni, E & Levkowitz, G 2019, 'Robo2 regulates synaptic oxytocin content by affecting actin dynamics', *Elife*, vol. 8, Jun 10.

Anderson, EM, Haupt, A, Schiel, JA, Chou, E, Machado, HB, Strezoska, Z, Lenger, S, McClelland, S, Birmingham, A, Vermeulen, A & Smith, A 2015, 'Systematic analysis of CRISPR-Cas9 mismatch tolerance reveals low levels of off-target activity', *Journal of Biotechnology*, vol. 211, Oct 10, pp. 56-65.

Anderson, WJ, Zhou, Q, Alcalde, V, Kaneko, OF, Blank, LJ, Sherwood, RI, Guseh, JS, Rajagopal, J & Melton, DA 2008, 'Genetic targeting of the endoderm with claudin-6CreER', *Developmental Dynamics*, vol. 237, no. 2, pp. 504-512.

Andrews, S 2010, 'FastQC: a quality control tool for high throughput sequence data'.

Arias-Palomo, E, Yamashita, A, Fernandez, IS, Nunez-Ramirez, R, Bamba, Y, Izumi, N, Ohno, S & Llorca, O 2011, 'The nonsense-mediated mRNA decay SMG-1 kinase is regulated by large-scale conformational changes controlled by SMG-8', *Genes Dev*, vol. 25, no. 2, Jan 15, pp. 153-164.

Arthur, HM, Ure, J, Smith, AJ, Renforth, G, Wilson, DI, Torsney, E, Charlton, R, Parums, DV, Jowett, T, Marchuk, DA, Burn, J & Diamond, AG 2000, 'Endoglin, an ancillary TGFbeta receptor, is required for extraembryonic angiogenesis and plays a key role in heart development', *Dev Biol*, vol. 217, no. 1, Jan 1, pp. 42-53.

Avior, Y, Sagi, I & Benvenisty, N 2016, 'Pluripotent stem cells in disease modelling and drug discovery', *Nature reviews Molecular cell biology*, vol. 17, no. 3, p. 170.

Azzalin, CM & Lingner, J 2006, 'The human RNA surveillance factor UPF1 is required for S phase progression and genome stability', *Current Biology*, vol. 16, no. 4, pp. 433-439.

Bae, S, Park, J & Kim, J-S 2014, 'Cas-OFFinder: a fast and versatile algorithm that searches for potential off-target sites of Cas9 RNA-guided endonucleases', *Bioinformatics*, vol. 30, no. 10, pp. 1473-1475.

Baker, DE, Harrison, NJ, Maltby, E, Smith, K, Moore, HD, Shaw, PJ, Heath, PR, Holden, H & Andrews, PW 2007, 'Adaptation to culture of human embryonic stem cells and oncogenesis in vivo', *Nature Biotechnology*, vol. 25, no. 2, Feb, pp. 207-215.

Ballut, L, Marchadier, B, Baguet, A, Tomasetto, C, Séraphin, B & Le Hir, H 2005, 'The exon junction core complex is locked onto RNA by inhibition of eIF4AIII ATPase activity', *Nature Structural and Molecular Biology*, vol. 12, no. 10, p. 861.

Bar-Nur, O, Russ, HA, Efrat, S & Benvenisty, N 2011, 'Epigenetic memory and preferential lineage-specific differentiation in induced pluripotent stem cells derived from human pancreatic islet beta cells', *Cell Stem Cell*, vol. 9, no. 1, pp. 17-23.

Barta, T, Dolezalova, D, Holubcova, Z & Hampl, A 2013, 'Cell cycle regulation in human embryonic stem cells: links to adaptation to cell culture', *Exp Biol Med (Maywood)*, vol. 238, no. 3, Mar, pp. 271-275.

Bateman, JF, Boot-Handford, RP & Lamande, SR 2009, 'Genetic diseases of connective tissues: cellular and extracellular effects of ECM mutations', *Nat Rev Genet*, vol. 10, no. 3, Mar, pp. 173-183.

Bateman, JF, Freddi, S, Nattrass, G & Savarirayan, R 2003, 'Tissue-specific RNA surveillance? Nonsense-mediated mRNA decay causes collagen X haploinsufficiency in Schmid metaphyseal chondrodysplasia cartilage', *Hum Mol Genet*, vol. 12, no. 3, Feb 1, pp. 217-225.

Battegay, EJ, Raines, EW, Seifert, RA, Bowen-Pope, DF & Ross, R 1990, 'TGF-β induces bimodal proliferation of connective tissue cells via complex control of an autocrine PDGF loop', *Cell*, vol. 63, no. 3, pp. 515-524.

Bellin, M, Marchetto, MC, Gage, FH & Mummery, CL 2012, 'Induced pluripotent stem cells: the new patient?', *Nature reviews Molecular cell biology*, vol. 13, no. 11, p. 713.

Bengtsson, NE, Hall, JK, Odom, GL, Phelps, MP, Andrus, CR, Hawkins, RD, Hauschka, SD, Chamberlain, JR & Chamberlain, JS 2017, 'Muscle-specific CRISPR/Cas9 dystrophin gene editing ameliorates pathophysiology in a mouse model for Duchenne muscular dystrophy', *Nat Commun*, vol. 8, Feb 14, p. 14454.

Bibikova, M, Beumer, K, Trautman, JK & Carroll, D 2003, 'Enhancing gene targeting with designed zinc finger nucleases', *Science*, vol. 300, no. 5620, pp. 764-764.

Blobe, GC, Schiemann, WP & Lodish, HF 2000, 'Role of transforming growth factor β in human disease', *New England Journal of Medicine*, vol. 342, no. 18, pp. 1350-1358.

Boehm, V, Haberman, N, Ottens, F, Ule, J & Gehring, NH 2014, '3' UTR length and messenger ribonucleoprotein composition determine endocleavage efficiencies at termination codons', *Cell Rep*, vol. 9, no. 2, Oct 23, pp. 555-568.

Bogdanove, AJ, Schornack, S & Lahaye, T 2010, 'TAL effectors: finding plant genes for disease and defense', *Current Opinion in Plant Biology*, vol. 13, no. 4, pp. 394-401.

Bonig, H, Banning, U, Hannen, M, Kim, YM, Verheyen, J, Mauz-Korholz, C & Korholz, D 1999, 'Transforming growth factor-beta1 suppresses interleukin-15-mediated interferon-gamma production in human T lymphocytes', *Scandinavian Journal of Immunology*, vol. 50, no. 6, Dec, pp. 612-618.

Border, WA & Noble, NA 1994, 'Transforming growth factor beta in tissue fibrosis', *New England Journal of Medicine*, vol. 331, no. 19, Nov 10, pp. 1286-1292.

Boutz, PL, Stoilov, P, Li, Q, Lin, C-H, Chawla, G, Ostrow, K, Shiue, L, Ares, M & Black, DL 2007, 'A post-transcriptional regulatory switch in polypyrimidine tract-binding proteins reprograms alternative splicing in developing neurons', *Genes & Development*, vol. 21, no. 13, pp. 1636-1652.

Briggs, JA, Sun, J, Shepherd, J, Ovchinnikov, DA, Chung, TL, Nayler, SP, Kao, LP, Morrow, CA, Thakar, NY, Soo, SY, Peura, T, Grimmond, S & Wolvetang, EJ 2013, 'Integration-free induced pluripotent stem cells model genetic and neural developmental features of down syndrome etiology', *Stem Cells*, vol. 31, no. 3, Mar, pp. 467-478.

Brons, IG, Smithers, LE, Trotter, MW, Rugg-Gunn, P, Sun, B, Chuva de Sousa Lopes, SM, Howlett, SK, Clarkson, A, Ahrlund-Richter, L, Pedersen, RA & Vallier, L 2007, 'Derivation of pluripotent epiblast stem cells from mammalian embryos', *Nature*, vol. 448, no. 7150, Jul 12, pp. 191-195.

Brunetti-Pierri, N, Berg, JS, Scaglia, F, Belmont, J, Bacino, CA, Sahoo, T, Lalani, SR, Graham, B, Lee, B & Shinawi, M 2008, 'Recurrent reciprocal 1q21. 1 deletions and duplications associated with microcephaly or macrocephaly and developmental and behavioral abnormalities', *Nature Genetics*, vol. 40, no. 12, p. 1466.

Brunetti-Pierri, N, Berg, JS, Scaglia, F, Belmont, J, Bacino, CA, Sahoo, T, Lalani, SR, Graham, B, Lee, B, Shinawi, M, Shen, J, Kang, SH, Pursley, A, Lotze, T, Kennedy, G, Lansky-Shafer, S, Weaver, C, Roeder, ER, Grebe, TA, Arnold, GL, Hutchison, T, Reimschisel, T, Amato, S, Geragthy, MT, Innis, JW, Obersztyn, E, Nowakowska, B, Rosengren, SS, Bader, PI, Grange, DK, Naqvi, S, Garnica, AD, Bernes, SM, Fong, CT, Summers, A, Walters, WD, Lupski, JR, Stankiewicz, P, Cheung, SW & Patel, A 2008, 'Recurrent reciprocal 1q21.1 deletions and duplications associated with microcephaly or macrocephaly and developmental and behavioral abnormalities', *Nat Genet*, vol. 40, no. 12, Dec, pp. 1466-1471.

Bruno, IG, Karam, R, Huang, L, Bhardwaj, A, Lou, CH, Shum, EY, Song, HW, Corbett, MA, Gifford, WD, Gecz, J, Pfaff, SL & Wilkinson, MF 2011, 'Identification of a microRNA that activates gene expression by repressing nonsense-mediated RNA decay', *Mol Cell*, vol. 42, no. 4, May 20, pp. 500-510.

Buchwald, G, Ebert, J, Basquin, C, Sauliere, J, Jayachandran, U, Bono, F, Le Hir, H & Conti, E 2010, 'Insights into the recruitment of the NMD machinery from the crystal structure of a core EJC-UPF3b complex', *Proc Natl Acad Sci U S A*, vol. 107, no. 22, Jun 1, pp. 10050-10055.

Byrne, SM, Mali, P & Church, GM 2014, 'Genome editing in human stem cells', *Methods Enzymol*, vol. 546, pp. 119-138.

Cai, J-l, Li, M & Na, Y-q 2011, 'Correlation between hyaluronic acid, hyaluronic acid synthase and human renal clear cell carcinoma', *Chinese Journal of Cancer Research*, vol. 23, no. 1, pp. 59-63.

Camenisch, TD, Spicer, AP, Brehm-Gibson, T, Biesterfeldt, J, Augustine, ML, Calabro, A, Jr., Kubalak, S, Klewer, SE & McDonald, JA 2000, 'Disruption of hyaluronan synthase-2 abrogates normal cardiac morphogenesis and hyaluronan-mediated transformation of epithelium to mesenchyme', *J Clin Invest*, vol. 106, no. 3, Aug, pp. 349-360.

Cannan, WJ & Pederson, DS 2016, 'Mechanisms and consequences of double-strand DNA break formation in chromatin', *Journal of Cellular Physiology*, vol. 231, no. 1, pp. 3-14.

Carvalho, RL, Itoh, F, Goumans, M-J, Lebrin, F, Kato, M, Takahashi, S, Ema, M, Itoh, S, van Rooijen, M & Bertolino, P 2007, 'Compensatory signalling induced in the yolk sac vasculature by deletion of TGFβ receptors in mice', *Journal of Cell Science*, vol. 120, no. 24, pp. 4269-4277.

Carvalho, RL, Jonker, L, Goumans, MJ, Larsson, J, Bouwman, P, Karlsson, S, Dijke, PT, Arthur, HM & Mummery, CL 2004, 'Defective paracrine signalling by TGFbeta in yolk sac vasculature of endoglin mutant mice: a paradigm for hereditary haemorrhagic telangiectasia', *Development*, vol. 131, no. 24, Dec, pp. 6237-6247.

Ceccaldi, R, Rondinelli, B & D'Andrea, AD 2016, 'Repair pathway choices and consequences at the double-strand break', *Trends in Cell Biology*, vol. 26, no. 1, pp. 52-64.

Chakrabarti, S, Jayachandran, U, Bonneau, F, Fiorini, F, Basquin, C, Domcke, S, Le Hir, H & Conti, E 2011, 'Molecular mechanisms for the RNA-dependent ATPase activity of Upf1 and its regulation by Upf2', *Mol Cell*, vol. 41, no. 6, Mar 18, pp. 693-703.

Chambers, SM, Fasano, CA, Papapetrou, EP, Tomishima, M, Sadelain, M & Studer, L 2009, 'Highly efficient neural conversion of human ES and iPS cells by dual inhibition of SMAD signaling', *Nature Biotechnology*, vol. 27, no. 3, p. 275.

Chamieh, H, Ballut, L, Bonneau, F & Le Hir, H 2008, 'NMD factors UPF2 and UPF3 bridge UPF1 to the exon junction complex and stimulate its RNA helicase activity', *Nat Struct Mol Biol*, vol. 15, no. 1, Jan, pp. 85-93.

Chan, WK, Bhalla, AD, Le Hir, H, Nguyen, LS, Huang, L, Gecz, J & Wilkinson, MF 2009, 'A UPF3-mediated regulatory switch that maintains RNA surveillance', *Nat Struct Mol Biol*, vol. 16, no. 7, Jul, pp. 747-753.

Chan, WK, Huang, L, Gudikote, JP, Chang, YF, Imam, JS, MacLean, JA, 2nd & Wilkinson, MF 2007, 'An alternative branch of the nonsense-mediated decay pathway', *Embo j*, vol. 26, no. 7, Apr 4, pp. 1820-1830.

Chang, JC & Kan, YW 1979, 'beta 0 thalassemia, a nonsense mutation in man', *Proc Natl Acad Sci U S A*, vol. 76, no. 6, Jun, pp. 2886-2889.

Chen, ZF & Behringer, RR 1995, 'twist is required in head mesenchyme for cranial neural tube morphogenesis', *Genes Dev*, vol. 9, no. 6, Mar 15, pp. 686-699.

Chester, N, Kuo, F, Kozak, C, O'Hara, CD & Leder, P 1998, 'Stage-specific apoptosis, developmental delay, and embryonic lethality in mice homozygous for a targeted disruption in the murine Bloom's syndrome gene', *Genes & Development*, vol. 12, no. 21, pp. 3382-3393.

Chin, MH, Mason, MJ, Xie, W, Volinia, S, Singer, M, Peterson, C, Ambartsumyan, G, Aimiuwu, O, Richter, L & Zhang, J 2009, 'Induced pluripotent stem cells and embryonic stem cells are distinguished by gene expression signatures', *Cell Stem Cell*, vol. 5, no. 1, pp. 111-123.

Chitaev, NA & Troyanovsky, SM 1998, 'Adhesive but not lateral E-cadherin complexes require calcium and catenins for their formation', *J Cell Biol*, vol. 142, no. 3, Aug 10, pp. 837-846.

Cho, H, Kim, KM, Han, S, Choe, J, Park, SG, Choi, SS & Kim, YK 2012, 'Staufen1-mediated mRNA decay functions in adipogenesis', *Mol Cell*, vol. 46, no. 4, May 25, pp. 495-506.

Cho, SW, Kim, S, Kim, Y, Kweon, J, Kim, HS, Bae, S & Kim, J-S 2014, 'Analysis of off-target effects of CRISPR/Cas-derived RNA-guided endonucleases and nickases', *Genome Research*, vol. 24, no. 1, pp. 132-141.

Choudhry, Z, Sengupta, SM, Grizenko, N, Fortier, ME, Thakur, GA, Bellingham, J & Joober, R 2012, 'LPHN3 and attention-deficit/hyperactivity disorder: interaction with maternal stress during pregnancy', *Journal of Child Psychology and Psychiatry and Allied Disciplines*, vol. 53, no. 8, Aug, pp. 892-902.

Clancy, B, Darlington, RB & Finlay, BL 2001, 'Translating developmental time across mammalian species', *Neuroscience*, vol. 105, no. 1, pp. 7-17.

Colak, D, Ji, SJ, Porse, BT & Jaffrey, SR 2013, 'Regulation of axon guidance by compartmentalized nonsense-mediated mRNA decay', *Cell*, vol. 153, no. 6, Jun 6, pp. 1252-1265.

Cong, L, Ran, FA, Cox, D, Lin, S, Barretto, R, Habib, N, Hsu, PD, Wu, X, Jiang, W & Marraffini, LA 2013, 'Multiplex genome engineering using CRISPR/Cas systems', *Science*, vol. 339, no. 6121, pp. 819-823.

Cong, L, Ran, FA, Cox, D, Lin, S, Barretto, R, Habib, N, Hsu, PD, Wu, X, Jiang, W, Marraffini, LA & Zhang, F 2013, 'Multiplex genome engineering using CRISPR/Cas systems', *Science*, vol. 339, no. 6121, Feb 15, pp. 819-823.

Coultas, L, Chawengsaksophak, K & Rossant, J 2005, 'Endothelial cells and VEGF in vascular development', *Nature*, vol. 438, no. 7070, Dec 15, pp. 937-945.

Cradick, TJ, Fine, EJ, Antico, CJ & Bao, G 2013, 'CRISPR/Cas9 systems targeting beta-globin and CCR5 genes have substantial off-target activity', *Nucleic Acids Res*, vol. 41, no. 20, Nov, pp. 9584-9592.

Crudele, JM & Chamberlain, JS 2018, 'Cas9 immunity creates challenges for CRISPR gene editing therapies', *Nat Commun*, vol. 9, no. 1, Aug 29, p. 3497.

Curran, SP, Hickey, FB, Watson, A, Godson, C & Brazil, DP 2012, 'Deletion of Gremlin1 increases cell proliferation and migration responses in mouse embryonic fibroblasts', *Cellular Signalling*, vol. 24, no. 4, Apr, pp. 889-898.

da Costa, PJ, Menezes, J & Romao, L 2017, 'The role of alternative splicing coupled to nonsense-mediated mRNA decay in human disease', *International Journal of Biochemistry and Cell Biology*, vol. 91, no. Pt B, Oct, pp. 168-175.

Daley, WP, Peters, SB & Larsen, M 2008, 'Extracellular matrix dynamics in development and regenerative medicine', *J Cell Sci*, vol. 121, no. Pt 3, Feb 1, pp. 255-264.

De Caestecker, M 2004, 'The transforming growth factor-β superfamily of receptors', *Cytokine and Growth Factor Reviews*, vol. 15, no. 1, pp. 1-11.

de Massy, B 2013, 'Initiation of meiotic recombination: how and where? Conservation and specificities among eukaryotes', *Annual Review of Genetics*, vol. 47, pp. 563-599.

DeNardo, LA, de Wit, J, Otto-Hitt, S & Ghosh, A 2012, 'NGL-2 regulates input-specific synapse development in CA1 pyramidal neurons', *Neuron*, vol. 76, no. 4, Nov 21, pp. 762-775.

Dhara, SK & Stice, SL 2008, 'Neural differentiation of human embryonic stem cells', *Journal of Cellular Biochemistry*, vol. 105, no. 3, pp. 633-640.

Dickson, MC, Martin, JS, Cousins, FM, Kulkarni, AB, Karlsson, S & Akhurst, RJ 1995, 'Defective haematopoiesis and vasculogenesis in transforming growth factor-beta 1 knock out mice', *Development*, vol. 121, no. 6, Jun, pp. 1845-1854.

Ding, Q, Lee, Y-K, Schaefer, EA, Peters, DT, Veres, A, Kim, K, Kuperwasser, N, Motola, DL, Meissner, TB & Hendriks, WT 2013, 'A TALEN genome-editing system for generating human stem cell-based disease models', *Cell Stem Cell*, vol. 12, no. 2, pp. 238-251.

Ding, Q, Regan, SN, Xia, Y, Oostrom, LA, Cowan, CA & Musunuru, K 2013, 'Enhanced efficiency of human pluripotent stem cell genome editing through replacing TALENs with CRISPRs', *Cell Stem Cell*, vol. 12, no. 4, Apr 4, pp. 393-394.

Doi, A, Park, I-H, Wen, B, Murakami, P, Aryee, MJ, Irizarry, R, Herb, B, Ladd-Acosta, C, Rho, J & Loewer, S 2009, 'Differential methylation of tissue-and cancer-specific CpG island shores distinguishes human induced pluripotent stem cells, embryonic stem cells and fibroblasts', *Nature Genetics*, vol. 41, no. 12, p. 1350.

Draper, JS, Moore, HD, Ruban, LN, Gokhale, PJ & Andrews, PW 2004, 'Culture and characterization of human embryonic stem cells', *Stem cells and development*, vol. 13, no. 4, pp. 325-336.

Durand, S & Lykke-Andersen, J 2013, 'Nonsense-mediated mRNA decay occurs during eIF4F-dependent translation in human cells', *Nat Struct Mol Biol*, vol. 20, no. 6, Jun, pp. 702-709.

Ebert, AD, Liang, P & Wu, JC 2012, 'Induced pluripotent stem cells as a disease modeling and drug screening platform', *Journal of Cardiovascular Pharmacology*, vol. 60, no. 4, p. 408.

Ehninger, D, Li, W, Fox, K, Stryker, MP & Silva, AJ 2008, 'Reversing neurodevelopmental disorders in adults', *Neuron*, vol. 60, no. 6, Dec 26, pp. 950-960.

Eiges, R, Urbach, A, Malcov, M, Frumkin, T, Schwartz, T, Amit, A, Yaron, Y, Eden, A, Yanuka, O, Benvenisty, N & Ben-Yosef, D 2007, 'Developmental study of fragile X syndrome using human embryonic stem cells derived from preimplantation genetically diagnosed embryos', *Cell Stem Cell*, vol. 1, no. 5, Nov, pp. 568-577.

Favaro, FP, Alvizi, L, Zechi-Ceide, RM, Bertola, D, Felix, TM, de Souza, J, Raskin, S, Twigg, SR, Weiner, AM, Armas, P, Margarit, E, Calcaterra, NB, Andersen, GR, McGowan, SJ, Wilkie, AO, Richieri-Costa, A, de Almeida, ML & Passos-Bueno, MR 2014, 'A noncoding expansion in EIF4A3 causes Richieri-Costa-Pereira syndrome, a craniofacial disorder associated with limb defects', *Am J Hum Genet*, vol. 94, no. 1, Jan 2, pp. 120-128.

Favaro, FP, Zechi-Ceide, RM, Alvarez, CW, Maximino, LP, Antunes, LFB, Richieri-Costa, A & Guion-Almeida, ML 2011, 'Richieri-Costa–Pereira syndrome: a unique acrofacial dysostosis type. An overview of the Brazilian cases', *American Journal of Medical Genetics Part A*, vol. 155, no. 2, pp. 322-331.

Flavell, RA, Sanjabi, S, Wrzesinski, SH & Licona-Limon, P 2010, 'The polarization of immune cells in the tumour environment by TGFbeta', *Nature Reviews: Immunology*, vol. 10, no. 8, Aug, pp. 554-567.

Frank, NY, Kho, AT, Schatton, T, Murphy, GF, Molloy, MJ, Zhan, Q, Ramoni, MF, Frank, MH, Kohane, IS & Gussoni, E 2006, 'Regulation of myogenic progenitor proliferation in human fetal skeletal muscle by BMP4 and its antagonist Gremlin', *J Cell Biol*, vol. 175, no. 1, Oct 9, pp. 99-110.

Franks, TM, Singh, G & Lykke-Andersen, J 2010, 'Upf1 ATPase-dependent mRNP disassembly is required for completion of nonsense- mediated mRNA decay', *Cell*, vol. 143, no. 6, Dec 10, pp. 938-950.

Fu, Y, Foden, JA, Khayter, C, Maeder, ML, Reyon, D, Joung, JK & Sander, JD 2013, 'High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells', *Nature Biotechnology*, vol. 31, no. 9, Sep, pp. 822-826.

Fukuhara, N, Ebert, J, Unterholzner, L, Lindner, D, Izaurralde, E & Conti, E 2005, 'SMG7 is a 14-3-3-like adaptor in the nonsense-mediated mRNA decay pathway', *Mol Cell*, vol. 17, no. 4, Feb 18, pp. 537-547.

Gehring, NH, Kunz, JB, Neu-Yilik, G, Breit, S, Viegas, MH, Hentze, MW & Kulozik, AE 2005, 'Exon-junction complex components specify distinct routes of nonsense-mediated mRNA decay with differential cofactor requirements', *Mol Cell*, vol. 20, no. 1, Oct 7, pp. 65-75.

Gehring, NH, Neu-Yilik, G, Schell, T, Hentze, MW & Kulozik, AE 2003, 'Y14 and hUpf3b form an NMD-activating complex', *Mol Cell*, vol. 11, no. 4, Apr, pp. 939-949.

Goetz, AE & Wilkinson, M 2017, 'Stress and the nonsense-mediated RNA decay pathway', *Cellular and Molecular Life Sciences*, vol. 74, no. 19, pp. 3509-3531.

Gong, C, Kim, YK, Woeller, CF, Tang, Y & Maquat, LE 2009, 'SMD and NMD are competitive pathways that contribute to myogenesis: effects on PAX3 and myogenin mRNAs', *Genes Dev*, vol. 23, no. 1, Jan 1, pp. 54-66.

Gong, X & Wang, H 2015, 'SHANK1 and autism spectrum disorders', *Sci China Life Sci*, vol. 58, no. 10, Oct, pp. 985-990.

Gordon, KJ & Blobe, GC 2008, 'Role of transforming growth factor-β superfamily signaling pathways in human disease', *Biochimica et Biophysica Acta (BBA)-Molecular Basis of Disease*, vol. 1782, no. 4, pp. 197-228.

Gore, A, Li, Z, Fung, HL, Young, JE, Agarwal, S, Antosiewicz-Bourget, J, Canto, I, Giorgetti, A, Israel, MA, Kiskinis, E, Lee, JH, Loh, YH, Manos, PD, Montserrat, N, Panopoulos, AD, Ruiz, S, Wilbert, ML, Yu, J, Kirkness, EF, Izpisua Belmonte, JC, Rossi, DJ, Thomson, JA, Eggan, K, Daley, GQ, Goldstein, LS & Zhang, K 2011, 'Somatic coding mutations in human induced pluripotent stem cells', *Nature*, vol. 471, no. 7336, Mar 3, pp. 63-67.

Gregory-Evans, K, Bashar, AM & Tan, M 2012, 'Ex vivo gene therapy and vision', *Current Gene Therapy*, vol. 12, no. 2, Apr 1, pp. 103-115.

Gulsuner, S, Walsh, T, Watts, AC, Lee, MK, Thornton, AM, Casadei, S, Rippey, C, Shahin, H, Nimgaonkar, VL, Go, RC, Savage, RM, Swerdlow, NR, Gur, RE, Braff, DL, King, MC & McClellan, JM 2013, 'Spatial and temporal mapping of de novo mutations in schizophrenia to a fetal prefrontal cortical network', *Cell*, vol. 154, no. 3, Aug 1, pp. 518-529.

Guo, G, von Meyenn, F, Santos, F, Chen, Y, Reik, W, Bertone, P, Smith, A & Nichols, J 2016, 'Naive Pluripotent Stem Cells Derived Directly from Isolated Cells of the Human Inner Cell Mass', *Stem Cell Reports*, vol. 6, no. 4, Apr 12, pp. 437-446.

Hamlin, RL & Altschuld, RA 2011, 'Extrapolation from mouse to man', Am Heart Assoc.

Handley, A, Schauer, T, Ladurner, AG & Margulies, CE 2015, 'Designing cell-type-specific genome-wide experiments', *Molecular Cell*, vol. 58, no. 4, pp. 621-631.

Hansen, DV, Lui, JH, Flandin, P, Yoshikawa, K, Rubenstein, JL, Alvarez-Buylla, A & Kriegstein, AR 2013, 'Non-epithelial stem cells and cortical interneuron production in the human ganglionic eminences', *Nat Neurosci*, vol. 16, no. 11, Nov, pp. 1576-1587.

Harper, JW, Elledge, SJ, Keyomarsi, K, Dynlacht, B, Tsai, LH, Zhang, P, Dobrowolski, S, Bai, C, Connell-Crowley, L, Swindell, E & et al. 1995, 'Inhibition of cyclin-dependent kinases by p21', *Molecular Biology of the Cell*, vol. 6, no. 4, Apr, pp. 387-400.

He, F, Li, X, Spatrick, P, Casillo, R, Dong, S & Jacobson, A 2003, 'Genome-wide analysis of mRNAs regulated by the nonsense-mediated and 5' to 3' mRNA decay pathways in yeast', *Mol Cell*, vol. 12, no. 6, Dec, pp. 1439-1452.

Heler, R, Samai, P, Modell, JW, Weiner, C, Goldberg, GW, Bikard, D & Marraffini, LA 2015, 'Cas9 specifies functional viral targets during CRISPR-Cas adaptation', *Nature*, vol. 519, no. 7542, Mar 12, pp. 199-202.

Hendriks, WT, Warren, CR & Cowan, CA 2016, 'Genome editing in human pluripotent stem cells: approaches, pitfalls, and solutions', *Cell Stem Cell*, vol. 18, no. 1, pp. 53-65.

Henry, MP, Hawkins, JR, Boyle, J & Bridger, JM 2018, 'The Genomic Health of Human Pluripotent Stem Cells: Genomic Instability and the Consequences on Nuclear Organization', *Frontiers in genetics*, vol. 9.

Hestand, MS, Houdt, JV, Cristofoli, F & Vermeesch, JR 2016, 'Polymerase specific error rates and profiles identified by single molecule sequencing', *Mutat Res*, vol. 784-785, Feb-Mar, pp. 39-45.

Hetz, C, Russelakis-Carneiro, M, Maundrell, K, Castilla, J & Soto, C 2003, 'Caspase-12 and endoplasmic reticulum stress mediate neurotoxicity of pathological prion protein', *Embo j*, vol. 22, no. 20, pp. 5435-5445.

Hockemeyer, D, Soldner, F, Beard, C, Gao, Q, Mitalipova, M, DeKelver, RC, Katibah, GE, Amora, R, Boydston, EA & Zeitler, B 2009, 'Efficient targeting of expressed and silent genes in human ESCs and iPSCs using zinc-finger nucleases', *Nature Biotechnology*, vol. 27, no. 9, p. 851.

Hocking, JC, Hehr, CL, Bertolesi, GE, Wu, JY & McFarlane, S 2010, 'Distinct roles for Robo2 in the regulation of axon and dendrite growth by retinal ganglion cells', *Mechanisms of Development*, vol. 127, no. 1-2, Jan-Feb, pp. 36-48.

Hoek, TA, Khuperkar, D, Lindeboom, RGH, Sonneveld, S, Verhagen, BMP, Boersma, S, Vermeulen, M & Tanenbaum, ME 2019, 'Single-Molecule Imaging Uncovers Rules Governing Nonsense-Mediated mRNA Decay', *Mol Cell*, vol. 75, no. 2, Jul 25, pp. 324-339 e311.

Holdener, BC, Faust, C, Rosenthal, NS & Magnuson, T 1994, 'Msd is required for mesoderm induction in mice', *Development*, vol. 120, no. 5, pp. 1335-1346.

Homan, CC, Pederson, S, To, T-H, Tan, C, Piltz, S, Corbett, MA, Wolvetang, E, Thomas, PQ, Jolly, LA & Gecz, J 2018, 'PCDH19 regulation of neural progenitor cell differentiation suggests asynchrony of neurogenesis as a mechanism contributing to PCDH19 Girls Clustering Epilepsy', *Neurobiology of Disease*, vol. 116, pp. 106-119.

Hryhorowicz, M, Lipinski, D, Zeyland, J & Slomski, R 2016, 'CRISPR/Cas9 Immune System as a Tool for Genome Engineering', *Archivum Immunologiae et Therapiae Experimentalis*, Oct 3.

Hsu, PD, Scott, DA, Weinstein, JA, Ran, FA, Konermann, S, Agarwala, V, Li, Y, Fine, EJ, Wu, X, Shalem, O, Cradick, TJ, Marraffini, LA, Bao, G & Zhang, F 2013, 'DNA targeting specificity of RNA-guided Cas9 nucleases', *Nature Biotechnology*, vol. 31, no. 9, Sep, pp. 827-832.

Hsu, PD & Zhang, F 2012, 'Dissecting neural function using targeted genome engineering technologies', *ACS Chemical Neuroscience*, vol. 3, no. 8, pp. 603-610.

Huang, HP, Chen, PH, Yu, CY, Chuang, CY, Stone, L, Hsiao, WC, Li, CL, Tsai, SC, Chen, KY, Chen, HF, Ho, HN & Kuo, HC 2011, 'Epithelial cell adhesion molecule (EpCAM) complex proteins promote transcription factor-mediated pluripotency reprogramming', *J Biol Chem*, vol. 286, no. 38, Sep 23, pp. 33520-33532.

Huang, L, Lou, CH, Chan, W, Shum, EY, Shao, A, Stone, E, Karam, R, Song, HW & Wilkinson, MF 2011, 'RNA homeostasis governed by cell type-specific and branched feedback loops acting on NMD', *Mol Cell*, vol. 43, no. 6, Sep 16, pp. 950-961.

Huang, L, Shum, EY, Jones, SH, Lou, CH, Dumdie, J, Kim, H, Roberts, AJ, Jolly, LA, Espinoza, JL, Skarbrevik, DM, Phan, MH, Cook-Andersen, H, Swerdlow, NR, Gecz, J & Wilkinson, MF 2018, 'A Upf3b-mutant mouse model with behavioral and neurogenesis defects', *Mol Psychiatry*, vol. 23, no. 8, Aug, pp. 1773-1786.

Huang, L & Wilkinson, MF 2012, 'Regulation of nonsense-mediated mRNA decay', *Wiley Interdiscip Rev RNA*, vol. 3, no. 6, Nov-Dec, pp. 807-828.

Hynes, RO 2009, 'The extracellular matrix: not just pretty fibrils', *Science*, vol. 326, no. 5957, Nov 27, pp. 1216-1219.

Inacio, A, Silva, AL, Pinto, J, Ji, X, Morgado, A, Almeida, F, Faustino, P, Lavinha, J, Liebhaber, SA & Romao, L 2004, 'Nonsense mutations in close proximity to the initiation codon fail to trigger full nonsense-mediated mRNA decay', *J Biol Chem*, vol. 279, no. 31, Jul 30, pp. 32170-32180.

Innan, H & Kondrashov, F 2010, 'The evolution of gene duplications: classifying and distinguishing between models', *Nature Reviews Genetics*, vol. 11, no. 2, p. 97.

Inoue, M, Uchida, Y, Edagawa, M, Hirata, M, Mitamura, J, Miyamoto, D, Taketani, K, Sekine, S, Kawauchi, J & Kitajima, S 2018, 'The stress response gene ATF3 is a direct target of the Wnt/β-catenin pathway and inhibits the invasion and migration of HCT116 human colorectal cancer cells', *PLoS One*, vol. 13, no. 7, p. e0194160.

Ishigaki, Y, Li, X, Serin, G & Maquat, LE 2001, 'Evidence for a pioneer round of mRNA translation: mRNAs subject to nonsense-mediated decay in mammalian cells are bound by CBP80 and CBP20', *Cell*, vol. 106, no. 5, Sep 7, pp. 607-617.

Ishino, Y, Shinagawa, H, Makino, K, Amemura, M & Nakata, A 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*, vol. 169, no. 12, Dec, pp. 5429-5433.

Jaffrey, SR & Wilkinson, MF 2018, 'Nonsense-mediated RNA decay in the brain: emerging modulator of neural development and disease', *Nature Reviews: Neuroscience*, vol. 19, no. 12, Dec, pp. 715-728.

Jamar, NH, Kritsiligkou, P & Grant, CM 2018, 'Loss of mRNA surveillance pathways results in widespread protein aggregation', *Scientific Reports*, vol. 8, no. 1, p. 3894.

Ji, J, Ng, SH, Sharma, V, Neculai, D, Hussein, S, Sam, M, Trinh, Q, Church, GM, Mcpherson, JD & Nagy, A 2012, 'Elevated coding mutation rate during the reprogramming of human somatic cells into induced pluripotent stem cells', *Stem Cells*, vol. 30, no. 3, pp. 435-440.

Jiang, YH, Yuen, RK, Jin, X, Wang, M, Chen, N, Wu, X, Ju, J, Mei, J, Shi, Y, He, M, Wang, G, Liang, J, Wang, Z, Cao, D, Carter, MT, Chrysler, C, Drmic, IE, Howe, JL, Lau, L, Marshall, CR, Merico, D, Nalpathamkalam, T, Thiruvahindrapuram, B, Thompson, A, Uddin, M, Walker, S, Luo, J, Anagnostou, E, Zwaigenbaum, L, Ring, RH, Wang, J, Lajonchere, C, Wang, J, Shih, A, Szatmari, P, Yang, H, Dawson, G, Li, Y & Scherer, SW 2013, 'Detection of clinically relevant genetic variants in autism spectrum disorder by whole-genome sequencing', *Am J Hum Genet*, vol. 93, no. 2, Aug 8, pp. 249-263.

Jolly, LA, Homan, CC, Jacob, R, Barry, S & Gecz, J 2013, 'The UPF3B gene, implicated in intellectual disability, autism, ADHD and childhood onset schizophrenia regulates neural progenitor cell behaviour and neuronal outgrowth', *Hum Mol Genet*, vol. 22, no. 23, Dec 1, pp. 4673-4687.

Jones, SH & Wilkinson, M 2017, 'RNA decay, evolution, and the testis', *RNA Biol*, vol. 14, no. 2, Feb, pp. 146-155.

Joung, JK & Sander, JD 2013, 'TALENs: a widely applicable technology for targeted genome editing', *Nature reviews Molecular cell biology*, vol. 14, no. 1, p. 49.

Jovaisaite, V, Mouchiroud, L & Auwerx, J 2014, 'The mitochondrial unfolded protein response, a conserved stress response pathway with implications in health and disease', *Journal of Experimental Biology*, vol. 217, no. 1, pp. 137-143.

Kadlec, J, Izaurralde, E & Cusack, S 2004, 'The structural basis for the interaction between nonsense-mediated mRNA decay factors UPF2 and UPF3', *Nat Struct Mol Biol*, vol. 11, no. 4, Apr, pp. 330-337.

Kafasla, P, Mickleburgh, I, Llorian, M, Coelho, M, Gooding, C, Cherny, D, Joshi, A, Kotik-Kogan, O, Curry, S & Eperon, IC 2012, 'Defining the roles and interactions of PTB', Portland Press Limited.

Karam, R, Wengrod, J, Gardner, LB & Wilkinson, MF 2013, 'Regulation of nonsense-mediated mRNA decay: implications for physiology and disease', *Biochim Biophys Acta*, vol. 1829, no. 6-7, Jun-Jul, pp. 624-633.

Karam, R & Wilkinson, M 2012, 'A conserved microRNA/NMD regulatory circuit controls gene expression', *RNA Biol*, vol. 9, no. 1, Jan, pp. 22-26.

Karanam, K, Kafri, R, Loewer, A & Lahav, G 2012, 'Quantitative live cell imaging reveals a gradual shift between DNA repair mechanisms and a maximal use of HR in mid S phase', *Molecular Cell*, vol. 47, no. 2, pp. 320-329.

Kashima, I, Yamashita, A, Izumi, N, Kataoka, N, Morishita, R, Hoshino, S, Ohno, M, Dreyfuss, G & Ohno, S 2006, 'Binding of a novel SMG-1-Upf1-eRF1-eRF3 complex (SURF) to the exon junction complex triggers Upf1 phosphorylation and nonsense-mediated mRNA decay', *Genes Dev*, vol. 20, no. 3, Feb 1, pp. 355-367.

Kataoka, N, Yong, J, Kim, VN, Velazquez, F, Perkinson, RA, Wang, F & Dreyfuss, G 2000, 'Pre-mRNA splicing imprints mRNA in the nucleus with a novel RNA-binding protein that persists in the cytoplasm', *Molecular Cell*, vol. 6, no. 3, pp. 673-682.

Kato, M, Das, S, Petras, K, Kitamura, K, Morohashi, K, Abuelo, DN, Barr, M, Bonneau, D, Brady, AF, Carpenter, NJ, Cipero, KL, Frisone, F, Fukuda, T, Guerrini, R, Iida, E, Itoh, M, Lewanda, AF, Nanba, Y, Oka, A, Proud, VK, Saugier-Veber, P, Schelley, SL, Selicorni, A, Shaner, R, Silengo, M, Stewart, F, Sugiyama, N, Toyama, J, Toutain, A, Vargas, AL, Yanazawa, M, Zackai, EH & Dobyns, WB 2004, 'Mutations of ARX are associated with striking pleiotropy and consistent genotype-phenotype correlation', *Hum Mutat*, vol. 23, no. 2, Feb, pp. 147-159.

Katsarou, AM, Moshe, SL & Galanopoulou, AS 2017, 'INTERNEURONOPATHIES AND THEIR ROLE IN EARLY LIFE EPILEPSIES AND NEURODEVELOPMENTAL DISORDERS', *Epilepsia Open*, vol. 2, no. 3, Sep, pp. 284-306.

Keeling, KM, Du, M & Bedwell, DM 2013, 'Therapies of nonsense-associated diseases'.

Kehrl, JH, Roberts, AB, Wakefield, LM, Jakowlew, S, Sporn, MB & Fauci, AS 1986, 'Transforming growth factor beta is an important immunomodulatory protein for human B lymphocytes', *Journal of Immunology*, vol. 137, no. 12, Dec 15, pp. 3855-3860.

Khajavi, M, Inoue, K & Lupski, JR 2006, 'Nonsense-mediated mRNA decay modulates clinical outcome of genetic disease', *Eur J Hum Genet*, vol. 14, no. 10, Oct, pp. 1074-1081.

Kim, D-S, Ross, PJ, Zaslavsky, K & Ellis, J 2014, 'Optimizing neuronal differentiation from induced pluripotent stem cells to model ASD', *Frontiers in Cellular Neuroscience*, vol. 8, p. 109.

Kim, K, Zhao, R, Doi, A, Ng, K, Unternaehrer, J, Cahan, P, Hongguang, H, Loh, Y-H, Aryee, MJ & Lensch, MW 2011, 'Donor cell type can influence the epigenome and differentiation potential of human induced pluripotent stem cells', *Nature Biotechnology*, vol. 29, no. 12, p. 1117.

Kim, M, Roesener, AP, Mendonca, PR & Mastick, GS 2011, 'Robo1 and Robo2 have distinct roles in pioneer longitudinal axon guidance', *Dev Biol*, vol. 358, no. 1, Oct 1, pp. 181-188.

Kirov, G, Pocklington, AJ, Holmans, P, Ivanov, D, Ikeda, M, Ruderfer, D, Moran, J, Chambert, K, Toncheva, D, Georgieva, L, Grozeva, D, Fjodorova, M, Wollerton, R, Rees, E, Nikolov, I, van de Lagemaat, LN, Bayes, A, Fernandez, E, Olason, PI, Bottcher, Y, Komiyama, NH, Collins, MO, Choudhary, J, Stefansson, K, Stefansson, H, Grant, SG, Purcell, S, Sklar, P, O'Donovan, MC & Owen, MJ 2012, 'De novo CNV analysis implicates specific abnormalities of postsynaptic signalling complexes in the pathogenesis of schizophrenia', *Mol Psychiatry*, vol. 17, no. 2, Feb, pp. 142-153.

Kletzl, H, Marquet, A, Gunther, A, Tang, W, Heuberger, J, Groeneveld, GJ, Birkhoff, W, Mercuri, E, Lochmuller, H, Wood, C, Fischer, D, Gerlach, I, Heinig, K, Bugawan, T, Dziadek, S, Kinch, R, Czech, C & Khwaja, O 2019, 'The oral splicing modifier RG7800 increases full length survival of motor neuron 2 mRNA and survival of motor neuron protein: Results from trials in healthy adults and patients with spinal muscular atrophy', *Neuromuscul Disord*, vol. 29, no. 1, Jan, pp. 21-29.

Kosicki, M, Tomberg, K & Bradley, A 2018, 'Repair of double-strand breaks induced by CRISPR–Cas9 leads to large deletions and complex rearrangements', *Nature Biotechnology*, vol. 36, no. 8, p. 765.

Krijger, PHL, Di Stefano, B, de Wit, E, Limone, F, Van Oevelen, C, De Laat, W & Graf, T 2016, 'Cell-of-origin-specific 3D genome structure acquired during somatic cell reprogramming', *Cell Stem Cell*, vol. 18, no. 5, pp. 597-610.

Krumm, N, O'Roak, BJ, Karakoc, E, Mohajeri, K, Nelson, B, Vives, L, Jacquemont, S, Munson, J, Bernier, R & Eichler, EE 2013, 'Transmission disequilibrium of small CNVs in simplex autism', *Am J Hum Genet*, vol. 93, no. 4, Oct 3, pp. 595-606.

Kumar, S, Blangero, J & Curran, JE 2018, 'Induced Pluripotent Stem Cells in Disease Modeling and Gene Identification', in *Disease Gene Identification*, Springer, pp. 17-38.

Kunisato, A, Wakatsuki, M, Shinba, H, Ota, T, Ishida, I & Nagao, K 2010, 'Direct generation of induced pluripotent stem cells from human nonmobilized blood', *Stem cells and development*, vol. 20, no. 1, pp. 159-168.

Kunz, JB, Neu-Yilik, G, Hentze, MW, Kulozik, AE & Gehring, NH 2006, 'Functions of hUpf3a and hUpf3b in nonsense-mediated mRNA decay and translation', *Rna*, vol. 12, no. 6, Jun, pp. 1015-1022.

Kuscu, C, Parlak, M, Tufan, T, Yang, J, Szlachta, K, Wei, X, Mammadov, R & Adli, M 2017, 'CRISPR-STOP: gene silencing through base-editing-induced nonsense mutations', *Nat Methods*, vol. 14, no. 7, Jul, pp. 710-712.

Laco, F, Woo, TL, Zhong, Q, Szmyd, R, Ting, S, Khan, FJ, Chai, CL, Reuveny, S, Chen, A & Oh, S 2018, 'Unraveling the inconsistencies of cardiac differentiation efficiency induced by the GSK3β inhibitor CHIR99021 in human pluripotent stem cells', *Stem Cell Reports*, vol. 10, no. 6, pp. 1851-1866.

Larsson, J, Goumans, MJ, Sjostrand, LJ, van Rooijen, MA, Ward, D, Leveen, P, Xu, X, ten Dijke, P, Mummery, CL & Karlsson, S 2001, 'Abnormal angiogenesis but intact hematopoietic potential in TGF-beta type I receptor-deficient mice', *Embo j*, vol. 20, no. 7, Apr 2, pp. 1663-1673.

Laumonnier, F, Shoubridge, C, Antar, C, Nguyen, LS, Van Esch, H, Kleefstra, T, Briault, S, Fryns, JP, Hamel, B, Chelly, J, Ropers, HH, Ronce, N, Blesson, S, Moraine, C, Gecz, J & Raynaud, M 2010, 'Mutations of the UPF3B gene, which encodes a protein widely expressed in neurons, are associated with nonspecific mental retardation with or without autism', *Mol Psychiatry*, vol. 15, no. 7, Jul, pp. 767-776.

Laurent, LC, Ulitsky, I, Slavin, I, Tran, H, Schork, A, Morey, R, Lynch, C, Harness, JV, Lee, S & Barrero, MJ 2011, 'Dynamic changes in the copy number of pluripotency and cell proliferation genes in human ESCs and iPSCs during reprogramming and time in culture', *Cell Stem Cell*, vol. 8, no. 1, pp. 106-118.

Le Hir, H, Gatfield, D, Izaurralde, E & Moore, MJ 2001, 'The exon-exon junction complex provides a binding platform for factors involved in mRNA export and nonsense-mediated mRNA decay', *Embo j*, vol. 20, no. 17, Sep 3, pp. 4987-4997.

Le Hir, H, Izaurralde, E, Maquat, LE & Moore, MJ 2000, 'The spliceosome deposits multiple proteins 20-24 nucleotides upstream of mRNA exon-exon junctions', *Embo j*, vol. 19, no. 24, Dec 15, pp. 6860-6869.

Lee, J, Go, Y, Kang, I, Han, YM & Kim, J 2010, 'Oct-4 controls cell-cycle progression of embryonic stem cells', *Biochem J*, vol. 426, no. 2, Feb 9, pp. 171-181.

Lelivelt, MJ & Culbertson, MR 1999, 'Yeast Upf proteins required for RNA surveillance affect global expression of the yeast transcriptome', *Mol Cell Biol*, vol. 19, no. 10, Oct, pp. 6710-6719.

Leveen, P, Larsson, J, Ehinger, M, Cilio, CM, Sundler, M, Sjostrand, LJ, Holmdahl, R & Karlsson, S 2002, 'Induced disruption of the transforming growth factor beta type II receptor gene in mice causes a lethal inflammatory disorder that is transplantable', *Blood*, vol. 100, no. 2, Jul 15, pp. 560-568.

Li, H & Durbin, R 2009, 'Fast and accurate short read alignment with Burrows-Wheeler transform', *Bioinformatics*, vol. 25, no. 14, Jul 15, pp. 1754-1760.

Li, T, Shi, Y, Wang, P, Guachalla, LM, Sun, B, Joerss, T, Chen, YS, Groth, M, Krueger, A, Platzer, M, Yang, YG, Rudolph, KL & Wang, ZQ 2015, 'Smg6/Est1 licenses embryonic stem cell differentiation via nonsense-mediated mRNA decay', *Embo j*, Mar 14.

Licatalosi, DD, Yano, M, Fak, JJ, Mele, A, Grabinski, SE, Zhang, C & Darnell, RB 2012, 'Ptbp2 represses adult-specific splicing to regulate the generation of neuronal precursors in the embryonic brain', *Genes Dev*, vol. 26, no. 14, Jul 15, pp. 1626-1642.

Lieber, MR 2010, 'The mechanism of double-strand DNA break repair by the nonhomologous DNA end-joining pathway', *Annual Review of Biochemistry*, vol. 79, pp. 181-211.

Lieber, MR, Ma, Y, Pannicke, U & Schwarz, K 2003, 'Mechanism and regulation of human non-homologous DNA end-joining', *Nature reviews Molecular cell biology*, vol. 4, no. 9, p. 712.

Liebermann, DA & Hoffman, B 2008, 'Gadd45 in stress signaling', *Journal of molecular signaling*, vol. 3, no. 1, p. 15.

Lin, C, Lu, W, Zhai, L, Bethea, T, Berry, K, Qu, Z, Waud, WR & Li, Y 2011, 'Mesd is a general inhibitor of different Wnt ligands in Wnt/LRP signaling and inhibits PC-3 tumor growth in vivo', *FEBS Letters*, vol. 585, no. 19, pp. 3120-3125.

Lin, Y, Cradick, TJ, Brown, MT, Deshmukh, H, Ranjan, P, Sarode, N, Wile, BM, Vertino, PM, Stewart, FJ & Bao, G 2014, 'CRISPR/Cas9 systems have off-target activity with insertions or deletions between target DNA and guide RNA sequences', *Nucleic Acids Research*, vol. 42, no. 11, pp. 7473-7485.

Linder, B, Fischer, U & Gehring, NH 2015, 'mRNA metabolism and neuronal disease', *FEBS Letters*, vol. 589, no. 14, pp. 1598-1606.

Linets'ka, MV, Storchak, LH & Himmelreich, NH 2002, '[Effect of synaptosomal cytosolic [3H]GABA pool depletion on secretory ability of alpha-latrotoxin]', *Ukr Biokhim Zh (1999)*, vol. 74, no. 3, May-Jun, pp. 65-72.

Lister, R, Pelizzola, M, Kida, YS, Hawkins, RD, Nery, JR, Hon, G, Antosiewicz-Bourget, J, O'malley, R, Castanon, R & Klugman, S 2011, 'Hotspots of aberrant epigenomic reprogramming in human induced pluripotent stem cells', *Nature*, vol. 471, no. 7336, p. 68.

Lockyer, EJ 2016, 'The potential of CRISPR-Cas9 for treating genetic disorders', *Bioscience Horizons: The International Journal of Student Research*, vol. 9.

Loh, KM, Ang, LT, Zhang, J, Kumar, V, Ang, J, Auyeong, JQ, Lee, KL, Choo, SH, Lim, CY & Nichane, M 2014, 'Efficient endoderm induction from human pluripotent stem cells by logically directing signals controlling lineage bifurcations', *Cell Stem Cell*, vol. 14, no. 2, pp. 237-252.

Lohler, J, Timpl, R & Jaenisch, R 1984, 'Embryonic lethal mutation in mouse collagen I gene causes rupture of blood vessels and is associated with erythropoietic and mesenchymal cell death', *Cell*, vol. 38, no. 2, Sep, pp. 597-607.

Long, AA, Mahapatra, CT, Woodruff, EA, 3rd, Rohrbough, J, Leung, HT, Shino, S, An, L, Doerge, RW, Metzstein, MM, Pak, WL & Broadie, K 2010, 'The nonsense-mediated decay pathway maintains synapse architecture and synaptic vesicle cycle efficacy', *J Cell Sci*, vol. 123, no. Pt 19, Oct 1, pp. 3303-3315.

Longman, D, Hug, N, Keith, M, Anastasaki, C, Patton, EE, Grimes, G & Caceres, JF 2013, 'DHX34 and NBAS form part of an autoregulatory NMD circuit that regulates endogenous RNA targets in human cells, zebrafish and Caenorhabditis elegans', *Nucleic Acids Res*, vol. 41, no. 17, Sep, pp. 8319-8331.

Löser, P, Schirm, J, Guhr, A, Wobus, AM & Kurtz, A 2010, 'Human embryonic stem cell lines and their use in international research', *Stem Cells*, vol. 28, no. 2, pp. 240-246.

Lou, CH, Dumdie, J, Goetz, A, Shum, EY, Brafman, D, Liao, X, Mora-Castilla, S, Ramaiah, M, Cook-Andersen, H, Laurent, L & Wilkinson, MF 2016, 'Nonsense-

Mediated RNA Decay Influences Human Embryonic Stem Cell Fate', *Stem Cell Reports*, vol. 6, no. 6, Jun 14, pp. 844-857.

Lou, CH, Shao, A, Shum, EY, Espinoza, JL, Huang, L, Karam, R & Wilkinson, MF 2014, 'Posttranscriptional control of the stem cell and neurogenic programs by the nonsense-mediated RNA decay pathway', *Cell Rep*, vol. 6, no. 4, Feb 27, pp. 748-764.

Love, MI, Huber, W & Anders, S 2014, 'Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2', *Genome Biology*, vol. 15, no. 12, p. 550.

Love, MI, Soneson, C & Robinson, MD 2017, 'Importing transcript abundance datasets with tximport', *dim (txi. inf. rep $ infReps $ sample1)*, vol. 1, no. 178136, p. 5.

Lukovic, D, Diez Lloret, A, Stojkovic, P, Rodríguez-Martínez, D, Perez Arago, MA, Rodriguez-Jimenez, FJ, González-Rodríguez, P, López-Barneo, J, Sykova, E & Jendelova, P 2017, 'Highly efficient neural conversion of human pluripotent stem cells in adherent and animal-free conditions', *Stem Cells Translational Medicine*, vol. 6, no. 4, pp. 1217-1226.

Luo, J, Balkin, N, Stewart, JF, Sarwark, JF, Charrow, J & Nye, JS 2000, 'Neural tube defects and the 13q deletion syndrome: evidence for a critical region in 13q33-34', *American Journal of Medical Genetics*, vol. 91, no. 3, Mar 20, pp. 227-230.

Lykke-Andersen, J, Shu, M-D & Steitz, JA 2000, 'Human Upf proteins target an mRNA for nonsense-mediated decay when bound downstream of a termination codon', *Cell*, vol. 103, no. 7, pp. 1121-1131.

Lynch, SA, Nguyen, LS, Ng, LY, Waldron, M, McDonald, D & Gecz, J 2012, 'Broadening the phenotype associated with mutations in UPF3B: two further cases with renal dysplasia and variable developmental delay', *European Journal of Medical Genetics*, vol. 55, no. 8-9, Aug-Sep, pp. 476-479.

Lyu, C, Shen, J, Wang, R, Gu, H, Zhang, J, Xue, F, Liu, X, Liu, W, Fu, R, Zhang, L, Li, H, Zhang, X, Cheng, T, Yang, R & Zhang, L 2018, 'Targeted genome engineering in human induced pluripotent stem cells from patients with hemophilia B using the CRISPR-Cas9 system', *Stem Cell Research & Therapy*, vol. 9, no. 1, Apr 6, p. 92.

Maciel, TT, Melo, RS & Campos, AH 2009, 'The bone morphogenetic protein antagonist gremlin promotes vascular smooth muscle cell apoptosis', *Journal of Vascular Research*, vol. 46, no. 4, pp. 325-332.

Maciel, TT, Melo, RS, Schor, N & Campos, AH 2008, 'Gremlin promotes vascular smooth muscle cell proliferation and migration', *J Mol Cell Cardiol*, vol. 44, no. 2, Feb, pp. 370-379.

Mahfouz, MM, Piatek, A & Stewart, CN, Jr. 2014, 'Genome engineering via TALENs and CRISPR/Cas9 systems: challenges and perspectives', *Plant Biotechnol J*, vol. 12, no. 8, Oct, pp. 1006-1014.

Maksimova, N, Hara, K, Nikolaeva, I, Chun-Feng, T, Usui, T, Takagi, M, Nishihira, Y, Miyashita, A, Fujiwara, H, Oyama, T, Nogovicina, A, Sukhomyasova, A, Potapova, S, Kuwano, R, Takahashi, H, Nishizawa, M & Onodera, O 2010, 'Neuroblastoma amplified sequence gene is associated with a novel short stature syndrome characterised by optic nerve atrophy and Pelger-Huet anomaly', *J Med Genet*, vol. 47, no. 8, Aug, pp. 538-548.

Mali, P, Esvelt, KM & Church, GM 2013, 'Cas9 as a versatile tool for engineering biology', *Nat Methods*, vol. 10, no. 10, Oct, pp. 957-963.

Mali, P, Yang, L, Esvelt, KM, Aach, J, Guell, M, DiCarlo, JE, Norville, JE & Church, GM 2013, 'RNA-guided human genome engineering via Cas9', *Science*, vol. 339, no. 6121, Feb 15, pp. 823-826.

Martincorena, I & Campbell, PJ 2015, 'Somatic mutation in cancer and normal cells', *Science*, vol. 349, no. 6255, Sep 25, pp. 1483-1489.

Martins, R, Proenca, D, Silva, B, Barbosa, C, Silva, AL, Faustino, P & Romao, L 2012, 'Alternative polyadenylation and nonsense-mediated decay coordinately regulate the human HFE mRNA levels', *PLoS One*, vol. 7, no. 4, p. e35461.

Massague, J 2000, 'How cells read TGF-beta signals', *Nat Rev Mol Cell Biol*, vol. 1, no. 3, Dec, pp. 169-178.

Massagué, J, Blain, SW & Lo, RS 2000, 'TGFβ signaling in growth control, cancer, and heritable disorders', *Cell*, vol. 103, no. 2, pp. 295-309.

Massagué, J & Gomis, RR 2006, 'The logic of TGFβ signaling', *FEBS Letters*, vol. 580, no. 12, pp. 2811-2820.

McIlwain, DR, Pan, Q, Reilly, PT, Elia, AJ, McCracken, S, Wakeham, AC, Itie-Youten, A, Blencowe, BJ & Mak, TW 2010, 'Smg1 is required for embryogenesis and regulates diverse genes via alternative splicing coupled to nonsense-mediated mRNA decay', *Proc Natl Acad Sci U S A*, vol. 107, no. 27, Jul 6, pp. 12186-12191.

Medghalchi, SM, Frischmeyer, PA, Mendell, JT, Kelly, AG, Lawler, AM & Dietz, HC 2001, 'Rent1, a trans-effector of nonsense-mediated mRNA decay, is essential for mammalian embryonic viability', *Human Molecular Genetics*, vol. 10, no. 2, pp. 99-105.

Mendell, JT, Sharifi, NA, Meyers, JL, Martinez-Murillo, F & Dietz, HC 2004, 'Nonsense surveillance regulates expression of diverse classes of mammalian transcripts and mutes genomic noise', *Nature Genetics*, vol. 36, no. 10, p. 1073.

Menounos, PG & Patrinos, GP 2010, 'Mutation Detection by Single Strand Conformation Polymorphism and Heteroduplex Analysis', in *Molecular Diagnostics*, Elsevier, pp. 45-58.

Migaud, M, Charlesworth, P, Dempster, M, Webster, LC, Watabe, AM, Makhinson, M, He, Y, Ramsay, MF, Morris, RG & Morrison, JH 1998, 'Enhanced long-term potentiation and impaired learning in mice with mutant postsynaptic density-95 protein', *Nature*, vol. 396, no. 6710, p. 433.

Missler, M, Sudhof, TC & Biederer, T 2012, 'Synaptic cell adhesion', *Cold Spring Harbor Perspectives in Biology*, vol. 4, no. 4, Apr 1, p. a005694.

Molnar, Z & Clowry, G 2012, 'Cerebral cortical development in rodents and primates', in *Progress in Brain Research*, vol. 195, Elsevier, pp. 45-70.

Moore, JC, Atze, K, Yeung, PL, Toro-Ramos, AJ, Camarillo, C, Thompson, K, Ricupero, CL, Brenneman, MA, Cohen, RI & Hart, RP 2010, 'Efficient, high-throughput transfection of human embryonic stem cells', *Stem Cell Research & Therapy*, vol. 1, no. 3, Jul 26, p. 23.

Mort, M, Ivanov, D, Cooper, DN & Chuzhanova, NA 2008, 'A meta-analysis of nonsense mutations causing human genetic disease', *Hum Mutat*, vol. 29, no. 8, Aug, pp. 1037-1047.

Mort, M, Ivanov, D, Cooper, DN & Chuzhanova, NA 2008, 'A meta-analysis of nonsense mutations causing human genetic disease', *Human Mutation*, vol. 29, no. 8, pp. 1037-1047.

Nagy, E & Maquat, LE 1998, 'A rule for termination-codon position within intron-containing genes: when nonsense affects RNA abundance', *Trends Biochem Sci*, vol. 23, no. 6, Jun, pp. 198-199.

Naisbitt, S, Kim, E, Tu, JC, Xiao, B, Sala, C, Valtschanoff, J, Weinberg, RJ, Worley, PF & Sheng, M 1999, 'Shank, a novel family of postsynaptic density proteins that binds to the NMDA receptor/PSD-95/GKAP complex and cortactin', *Neuron*, vol. 23, no. 3, Jul, pp. 569-582.

Neu-Yilik, G, Raimondeau, E, Eliseev, B, Yeramala, L, Amthor, B, Deniaud, A, Huard, K, Kerschgens, K, Hentze, MW & Schaffitzel, C 2017, 'Dual function of UPF3B in early and late translation termination', *Embo j*, vol. 36, no. 20, pp. 2968-2986.

Ng, D, Pitcher, GM, Szilard, RK, Sertie, A, Kanisek, M, Clapcote, SJ, Lipina, T, Kalia, LV, Joo, D, McKerlie, C, Cortez, M, Roder, JC, Salter, MW & McInnes, RR 2009, 'Neto1 is a novel CUB-domain NMDA receptor-interacting protein required for synaptic plasticity and learning', *PLoS Biol*, vol. 7, no. 2, Feb 24, p. e41.

Nguyen, HT, Geens, M, Mertzanidou, A, Jacobs, K, Heirman, C, Breckpot, K & Spits, C 2014, 'Gain of 20q11.21 in human embryonic stem cells improves cell survival by increased expression of Bcl-xL', *Molecular Human Reproduction*, vol. 20, no. 2, Feb, pp. 168-177.

Nguyen, LS, Jolly, L, Shoubridge, C, Chan, WK, Huang, L, Laumonnier, F, Raynaud, M, Hackett, A, Field, M, Rodriguez, J, Srivastava, AK, Lee, Y, Long, R, Addington, AM, Rapoport, JL, Suren, S, Hahn, CN, Gamble, J, Wilkinson, MF, Corbett, MA & Gecz, J 2012, 'Transcriptome profiling of UPF3B/NMD-deficient lymphoblastoid cells from patients with various forms of intellectual disability', *Mol Psychiatry*, vol. 17, no. 11, Nov, pp. 1103-1115.

Nguyen, LS, Kim, HG, Rosenfeld, JA, Shen, Y, Gusella, JF, Lacassie, Y, Layman, LC, Shaffer, LG & Gecz, J 2013, 'Contribution of copy number variants involving nonsense-mediated mRNA decay pathway genes to neuro-developmental disorders', *Hum Mol Genet*, vol. 22, no. 9, May 1, pp. 1816-1825.

Nicholson, P, Yepiskoposyan, H, Metze, S, Zamudio Orozco, R, Kleinschmidt, N & Muhlemann, O 2010, 'Nonsense-mediated mRNA decay in human cells: mechanistic insights, functions beyond quality control and the double-life of NMD factors', *Cell Mol Life Sci*, vol. 67, no. 5, Mar, pp. 677-700.

Nowotschin, S & Hadjantonakis, AK 2010, 'Cellular dynamics in the early mouse embryo: from axis formation to gastrulation', *Curr Opin Genet Dev*, vol. 20, no. 4, Aug, pp. 420-427.

O'Donnell, L, Soileau, B, Heard, P, Carter, E, Sebold, C, Gelfond, J, Hale, DE & Cody, JD 2010, 'Genetic determinants of autism in individuals with deletions of 18q', *Hum Genet*, vol. 128, no. 2, Aug, pp. 155-164.

O'Rahilly, R & Müller, F 2007, 'Neurulation in the normal human embryo', in *Ciba Foundation Symposium 181-Neural Tube Defects: Neural Tube Defects: Ciba Foundation Symposium 181,* Wiley Online Library, pp. 70-89.

O'Sullivan, ML, Martini, F, von Daake, S, Comoletti, D & Ghosh, A 2014, 'LPHN3, a presynaptic adhesion-GPCR implicated in ADHD, regulates the strength of neocortical layer 2/3 synaptic input to layer 5', *Neural Dev*, vol. 9, Apr 17, p. 7.

Ohi, Y, Qin, H, Hong, C, Blouin, L, Polo, JM, Guo, T, Qi, Z, Downey, SL, Manos, PD & Rossi, DJ 2011, 'Incomplete DNA methylation underlies a transcriptional memory of somatic cells in human iPS cells', *Nature Cell Biology*, vol. 13, no. 5, p. 541.

Okada-Katsuhata, Y, Yamashita, A, Kutsuzawa, K, Izumi, N, Hirahara, F & Ohno, S 2012, 'N- and C-terminal Upf1 phosphorylations create binding platforms for SMG-6 and SMG-5:SMG-7 during NMD', *Nucleic Acids Res*, vol. 40, no. 3, Feb, pp. 1251-1266.

Patro, R, Duggal, G & Kingsford, C 2015, 'Salmon: accurate, versatile and ultrafast quantification from RNA-seq data using lightweight-alignment', *Biorxiv*, p. 021592.

Pattanayak, V, Lin, S, Guilinger, JP, Ma, E, Doudna, JA & Liu, DR 2013, 'High-throughput profiling of off-target DNA cleavage reveals RNA-programmed Cas9 nuclease specificity', *Nature Biotechnology*, vol. 31, no. 9, Sep, pp. 839-843.

Pelley, JW 2007, *Elsevier's Integrated Biochemistry*, Mosby.

Perrin-Vidoz, L, Sinilnikova, OM, Stoppa-Lyonnet, D, Lenoir, GM & Mazoyer, S 2002, 'The nonsense-mediated mRNA decay pathway triggers degradation of most BRCA1 mRNAs bearing premature termination codons', *Hum Mol Genet*, vol. 11, no. 23, Nov 1, pp. 2805-2814.

Peterson, SE & Loring, JF 2014, 'Genomic instability in pluripotent stem cells: implications for clinical applications', *J Biol Chem*, vol. 289, no. 8, Feb 21, pp. 4578-4584.

Phillips, KA, Bales, KL, Capitanio, JP, Conley, A, Czoty, PW, 't Hart, BA, Hopkins, WD, Hu, SL, Miller, LA & Nader, MA 2014, 'Why primate models matter', *American Journal of Primatology*, vol. 76, no. 9, pp. 801-827.

Pinard, A, Seddik, R & Bettler, B 2010, 'GABAB receptors: physiological functions and mechanisms of diversity', *Advances in Pharmacology*, vol. 58, pp. 231-255.

Platt, RJ, Chen, S, Zhou, Y, Yim, MJ, Swiech, L, Kempton, HR, Dahlman, JE, Parnas, O, Eisenhaure, TM & Jovanovic, M 2014, 'CRISPR-Cas9 knockin mice for genome editing and cancer modeling', *Cell*, vol. 159, no. 2, pp. 440-455.

Pokutta, S & Weis, WI 2007, 'Structure and mechanism of cadherins and catenins in cell-cell contacts', *Annual Review of Cell and Developmental Biology*, vol. 23, pp. 237-261.

Popp, MW & Maquat, LE 2014, 'The dharma of nonsense-mediated mRNA decay in mammalian cells', *Mol Cells*, vol. 37, no. 1, Jan, pp. 1-8.

Purcell, SM, Moran, JL, Fromer, M, Ruderfer, D, Solovieff, N, Roussos, P, O'Dushlaine, C, Chambert, K, Bergen, SE, Kahler, A, Duncan, L, Stahl, E, Genovese, G, Fernandez, E, Collins, MO, Komiyama, NH, Choudhary, JS, Magnusson, PK, Banks, E, Shakir, K, Garimella, K, Fennell, T, DePristo, M, Grant, SG, Haggarty, SJ, Gabriel, S, Scolnick, EM, Lander, ES, Hultman, CM, Sullivan, PF, McCarroll, SA & Sklar, P 2014, 'A polygenic burden of rare disruptive mutations in schizophrenia', *Nature*, vol. 506, no. 7487, Feb 13, pp. 185-190.

Puri, MC & Nagy, A 2012, 'Concise review: embryonic stem cells versus induced pluripotent stem cells: the game is on', *Stem Cells*, vol. 30, no. 1, pp. 10-14.

Ran, FA, Hsu, PD, Wright, J, Agarwala, V, Scott, DA & Zhang, F 2013, 'Genome engineering using the CRISPR-Cas9 system', *Nature Protocols*, vol. 8, no. 11, Nov, pp. 2281-2308.

Ranges, GE, Figari, IS, Espevik, T & Palladino, MA, Jr. 1987, 'Inhibition of cytotoxic T cell development by transforming growth factor beta and reversal by recombinant tumor necrosis factor alpha', *Journal of Experimental Medicine*, vol. 166, no. 4, Oct 1, pp. 991-998.

Rath, D, Amlinger, L, Rath, A & Lundgren, M 2015, 'The CRISPR-Cas immune system: biology, mechanisms and applications', *Biochimie*, vol. 117, pp. 119-128.

Ratni, H, Ebeling, M, Baird, J, Bendels, S, Bylund, J, Chen, KS, Denk, N, Feng, Z, Green, L, Guerard, M, Jablonski, P, Jacobsen, B, Khwaja, O, Kletzl, H, Ko, CP, Kustermann, S, Marquet, A, Metzger, F, Mueller, B, Naryshkin, NA, Paushkin, SV, Pinard, E, Poirier, A, Reutlinger, M, Weetall, M, Zeller, A, Zhao, X & Mueller, L 2018, 'Discovery of Risdiplam, a Selective Survival of Motor Neuron-2 (SMN2) Gene Splicing Modifier for the Treatment of Spinal Muscular Atrophy (SMA)', *Journal of Medicinal Chemistry*, vol. 61, no. 15, Aug 9, pp. 6501-6517.

Rausch, T, Zichner, T, Schlattl, A, Stutz, AM, Benes, V & Korbel, JO 2012, 'DELLY: structural variant discovery by integrated paired-end and split-read analysis', *Bioinformatics*, vol. 28, no. 18, Sep 15, pp. i333-i339.

Rebbapragada, I & Lykke-Andersen, J 2009, 'Execution of nonsense-mediated mRNA decay: what defines a substrate?', *Curr Opin Cell Biol*, vol. 21, no. 3, Jun, pp. 394-402.

Rehwinkel, J, Letunic, I, Raes, J, Bork, P & Izaurralde, E 2005, 'Nonsense-mediated mRNA decay factors act in concert to regulate common mRNA targets', *Rna*, vol. 11, no. 10, Oct, pp. 1530-1544.

Reissner, C, Klose, M, Fairless, R & Missler, M 2008, 'Mutational analysis of the neurexin/neuroligin complex reveals essential and regulatory components', *Proc Natl Acad Sci U S A*, vol. 105, no. 39, Sep 30, pp. 15124-15129.

Rodriguez-Gabriel, MA, Watt, S, Bahler, J & Russell, P 2006, 'Upf1, an RNA helicase required for nonsense-mediated mRNA decay, modulates the transcriptional response to oxidative stress in fission yeast', *Mol Cell Biol*, vol. 26, no. 17, Sep, pp. 6347-6356.

Rollins, B, Martin, MV, Morgan, L & Vawter, MP 2010, 'Analysis of whole genome biomarker expression in blood and brain', *Am J Med Genet B Neuropsychiatr Genet*, vol. 153B, no. 4, Jun 5, pp. 919-936.

Rosenfeld, JA, Traylor, RN, Schaefer, GB, McPherson, EW, Ballif, BC, Klopocki, E, Mundlos, S, Shaffer, LG & Aylsworth, AS 2012, 'Proximal microdeletions and microduplications of 1q21.1 contribute to variable abnormal phenotypes', *Eur J Hum Genet*, vol. 20, no. 7, Jul, pp. 754-761.

Rosler, ES, Fisk, GJ, Ares, X, Irving, J, Miura, T, Rao, MS & Carpenter, MK 2004, 'Long-term culture of human embryonic stem cells in feeder-free conditions', *Developmental Dynamics*, vol. 229, no. 2, Feb, pp. 259-274.

Rufener, SC & Muhlemann, O 2013, 'eIF4E-bound mRNPs are substrates for nonsense-mediated mRNA decay in mammalian cells', *Nat Struct Mol Biol*, vol. 20, no. 6, Jun, pp. 710-717.

Ruiz, S, Diep, D, Gore, A, Panopoulos, AD, Montserrat, N, Plongthongkum, N, Kumar, S, Fung, HL, Giorgetti, A, Bilic, J, Batchelder, EM, Zaehres, H, Kan, NG, Scholer, HR, Mercola, M, Zhang, K & Izpisua Belmonte, JC 2012, 'Identification of a specific reprogramming-associated epigenetic signature in human induced pluripotent stem cells', *Proc Natl Acad Sci U S A*, vol. 109, no. 40, Oct 2, pp. 16196-16201.

Saenger, P 1996, 'Turner's syndrome', *New England Journal of Medicine*, vol. 335, no. 23, pp. 1749-1754.

Sahel, DK, Mittal, A & Chitkara, D 2019, 'CRISPR/Cas System for Genome Editing: Progress and Prospects as a Therapeutic Tool', *Journal of Pharmacology and Experimental Therapeutics*, vol. 370, no. 3, Sep, pp. 725-735.

Sander, JD & Joung, JK 2014, 'CRISPR-Cas systems for editing, regulating and targeting genomes', *Nature Biotechnology*, vol. 32, no. 4, Apr, pp. 347-355.

Sanes, JR & Yamagata, M 1999, 'Formation of lamina-specific synaptic connections', *Current Opinion in Neurobiology*, vol. 9, no. 1, Feb, pp. 79-87.

Sangu, N, Shimojima, K, Takahashi, Y, Ohashi, T, Tohyama, J & Yamamoto, T 2017, 'A 7q31.33q32.1 microdeletion including LRRC4 and GRM8 is associated with severe intellectual disability and characteristics of autism', *Hum Genome Var*, vol. 4, p. 17001.

Sanjabi, S, Oh, SA & Li, MO 2017, 'Regulation of the Immune Response by TGF-beta: From Conception to Autoimmunity and Infection', *Cold Spring Harbor Perspectives in Biology*, vol. 9, no. 6, Jun 1.

Sano, K, Tanihara, H, Heimark, RL, Obata, S, Davidson, M, St John, T, Taketani, S & Suzuki, S 1993, 'Protocadherins: a large family of cadherin-related molecules in central nervous system', *Embo j*, vol. 12, no. 6, Jun, pp. 2249-2256.

Sato, D, Lionel, AC, Leblond, CS, Prasad, A, Pinto, D, Walker, S, O'Connor, I, Russell, C, Drmic, IE, Hamdan, FF, Michaud, JL, Endris, V, Roeth, R, Delorme, R, Huguet, G, Leboyer, M, Rastam, M, Gillberg, C, Lathrop, M, Stavropoulos, DJ, Anagnostou, E, Weksberg, R, Fombonne, E, Zwaigenbaum, L, Fernandez, BA, Roberts, W, Rappold, GA, Marshall, CR, Bourgeron, T, Szatmari, P & Scherer, SW 2012, 'SHANK1 Deletions in Males with Autism Spectrum Disorder', *Am J Hum Genet*, vol. 90, no. 5, May 4, pp. 879-887.

Scheiffele, P 2003, 'Cell-cell signaling during synapse formation in the CNS', *Annual Review of Neuroscience*, vol. 26, pp. 485-508.

Scheper, W & Hoozemans, JJ 2015, 'The unfolded protein response in neurodegenerative diseases: a neuropathological perspective', *Acta Neuropathologica*, vol. 130, no. 3, Sep, pp. 315-331.

Schnieke, A, Harbers, K & Jaenisch, R 1983, 'Embryonic lethal mutation in mice induced by retrovirus insertion into the alpha 1(I) collagen gene', *Nature*, vol. 304, no. 5924, Jul 28-Aug 3, pp. 315-320.

Schoch, H, Kreibich, AS, Ferri, SL, White, RS, Bohorquez, D, Banerjee, A, Port, RG, Dow, HC, Cordero, L, Pallathra, AA, Kim, H, Li, H, Bilker, WB, Hirano, S, Schultz, RT, Borgmann-Winter, K, Hahn, CG, Feldmeyer, D, Carlson, GC, Abel, T & Brodkin, ES 2017, 'Sociability Deficits and Altered Amygdala Circuits in Mice Lacking Pcdh10, an Autism Associated Gene', *Biological Psychiatry*, vol. 81, no. 3, Feb 1, pp. 193-202.

Schweingruber, C, Rufener, SC, Zund, D, Yamashita, A & Muhlemann, O 2013, 'Nonsense-mediated mRNA decay - mechanisms of substrate mRNA recognition and degradation in mammalian cells', *Biochim Biophys Acta*, vol. 1829, no. 6-7, Jun-Jul, pp. 612-623.

Seok, J, Warren, HS, Cuenca, AG, Mindrinos, MN, Baker, HV, Xu, W, Richards, DR, McDonald-Smith, GP, Gao, H & Hennessy, L 2013, 'Genomic responses in mouse models poorly mimic human inflammatory diseases', *Proceedings of the National Academy of Sciences*, vol. 110, no. 9, pp. 3507-3512.

Sequeira, PA, Martin, MV & Vawter, MP 2012, 'The first decade and beyond of transcriptional profiling in schizophrenia', *Neurobiol Dis*, vol. 45, no. 1, Jan, pp. 23-36.

Sheng, M & Hoogenraad, CC 2007, 'The postsynaptic architecture of excitatory synapses: a more quantitative view', *Annu Rev Biochem*, vol. 76, pp. 823-847.

Shi, Y, Kirwan, P & Livesey, FJ 2012, 'Directed differentiation of human pluripotent stem cells to cerebral cortex neurons and neural networks', *Nature Protocols*, vol. 7, no. 10, Oct, pp. 1836-1846.

Shi, Y, Kirwan, P, Smith, J, Robinson, HP & Livesey, FJ 2012, 'Human cerebral cortex development from pluripotent stem cells to functional excitatory synapses', *Nature Neuroscience*, vol. 15, no. 3, p. 477.

Shin, HY, Wang, C, Lee, HK, Yoo, KH, Zeng, X, Kuhns, T, Yang, CM, Mohr, T, Liu, C & Hennighausen, L 2017, 'CRISPR/Cas9 targeting events cause complex deletions and insertions at 17 sites in the mouse genome', *Nat Commun*, vol. 8, May 31, p. 15464.

Shin, HY, Wang, C, Lee, HK, Yoo, KH, Zeng, X, Kuhns, T, Yang, CM, Mohr, T, Liu, C & Hennighausen, L 2017, 'CRISPR/Cas9 targeting events cause complex deletions and insertions at 17 sites in the mouse genome', *Nat Commun*, vol. 8, p. 15464.

Shoubridge, C, Fullston, T & Gecz, J 2010, 'ARX spectrum disorders: making inroads into the molecular pathology', *Hum Mutat*, vol. 31, no. 8, Aug, pp. 889-900.

Shum, EY, Jones, SH, Shao, A, Dumdie, J, Krause, MD, Chan, WK, Lou, CH, Espinoza, JL, Song, HW, Phan, MH, Ramaiah, M, Huang, L, McCarrey, JR, Peterson, KJ, De Rooij, DG, Cook-Andersen, H & Wilkinson, MF 2016, 'The Antagonistic Gene Paralogs Upf3a and Upf3b Govern Nonsense-Mediated RNA Decay', *Cell*, vol. 165, no. 2, Apr 07, pp. 382-395.

Siegel, PM & Massague, J 2003, 'Cytostatic and apoptotic actions of TGF-beta in homeostasis and cancer', *Nature Reviews: Cancer*, vol. 3, no. 11, Nov, pp. 807-821.

Silver, DL, Watkins-Chow, DE, Schreck, KC, Pierfelice, TJ, Larson, DM, Burnetti, AJ, Liaw, HJ, Myung, K, Walsh, CA, Gaiano, N & Pavan, WJ 2010, 'The exon junction complex component Magoh controls brain size by regulating neural stem cell division', *Nat Neurosci*, vol. 13, no. 5, May, pp. 551-558.

Singh, G, Rebbapragada, I & Lykke-Andersen, J 2008, 'A competition between stimulators and antagonists of Upf complex recruitment governs human nonsense-mediated mRNA decay', *PLoS Biol*, vol. 6, no. 4, Apr 29, p. e111.

Smith, C, Gore, A, Yan, W, Abalde-Atristain, L, Li, Z, He, C, Wang, Y, Brodsky, RA, Zhang, K, Cheng, L & Ye, Z 2014, 'Whole-genome sequencing analysis reveals high specificity of CRISPR/Cas9 and TALEN-based genome editing in human iPSCs', *Cell Stem Cell*, vol. 15, no. 1, Jul 03, pp. 12-13.

Smith, JE & Baker, KE 2015, 'Nonsense-mediated RNA decay - a switch and dial for regulating gene expression', *Bioessays*, Mar 27.

Smyth, MJ, Strobl, SL, Young, HA, Ortaldo, JR & Ochoa, AC 1991, 'Regulation of lymphokine-activated killer activity and pore-forming protein gene expression in human peripheral blood CD8+ T lymphocytes. Inhibition by transforming growth factor-beta', *Journal of Immunology*, vol. 146, no. 10, May 15, pp. 3289-3297.

Smyth, N, Vatansever, HS, Murray, P, Meyer, M, Frie, C, Paulsson, M & Edgar, D 1999, 'Absence of basement membranes after targeting the LAMC1 gene results in embryonic lethality due to failure of endoderm differentiation', *J Cell Biol*, vol. 144, no. 1, Jan 11, pp. 151-160.

Sneddon, JB, Zhen, HH, Montgomery, K, van de Rijn, M, Tward, AD, West, R, Gladstone, H, Chang, HY, Morganroth, GS, Oro, AE & Brown, PO 2006, 'Bone morphogenetic protein antagonist gremlin 1 is widely expressed by cancer-associated stromal cells and can promote tumor cell proliferation', *Proc Natl Acad Sci U S A*, vol. 103, no. 40, Oct 3, pp. 14842-14847.

Solnica-Krezel, L & Sepich, DS 2012, 'Gastrulation: making and shaping germ layers', *Annual Review of Cell and Developmental Biology*, vol. 28, pp. 687-717.

Springer, MS, Murphy, WJ, Eizirik, E & O'Brien, SJ 2003, 'Placental mammal diversification and the Cretaceous–Tertiary boundary', *Proceedings of the National Academy of Sciences*, vol. 100, no. 3, pp. 1056-1061.

Steele, W, Allegrucci, C, Singh, R, Lucas, E, Priddle, H, Denning, C, Sinclair, K & Young, L 2005, 'Human embryonic stem cell methyl cycle enzyme expression: modelling epigenetic programming in assisted reproduction?', *Reproductive biomedicine online*, vol. 10, no. 6, pp. 755-766.

Sternberg, SH, Richter, H, Charpentier, E & Qimron, U 2016, 'Adaptation in CRISPR-Cas Systems', *Mol Cell*, vol. 61, no. 6, Mar 17, pp. 797-808.

Strachan, T, Lindsay, S & Wilson, DI 1997, *Molecular genetics of early human development*, Bios Scientific Pub Limited.

Swiech, L, Heidenreich, M, Banerjee, A, Habib, N, Li, Y, Trombetta, J, Sur, M & Zhang, F 2015, 'In vivo interrogation of gene function in the mammalian brain using CRISPR-Cas9', *Nature Biotechnology*, vol. 33, no. 1, p. 102.

Szpir, M 2006, 'New thinking on neurodevelopment', *Environmental Health Perspectives*, vol. 114, no. 2, Feb, pp. A100-107.

Szyszka, P, Sharp, SI, Dedman, A, Gurling, HM & McQuillin, A 2012, 'A nonconservative amino acid change in the UPF3B gene in a patient with schizophrenia', *Psychiatr Genet*, vol. 22, no. 3, Jun, pp. 150-151.

Takahashi, K, Tanabe, K, Ohnuki, M, Narita, M, Ichisaka, T, Tomoda, K & Yamanaka, S 2007, 'Induction of pluripotent stem cells from adult human fibroblasts by defined factors', *Cell*, vol. 131, no. 5, pp. 861-872.

Takahashi, K & Yamanaka, S 2006, 'Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors', *Cell*, vol. 126, no. 4, pp. 663-676.

Tam, PP & Behringer, RR 1997, 'Mouse gastrulation: the formation of a mammalian body plan', *Mechanisms of Development*, vol. 68, no. 1-2, Nov, pp. 3-25.

Tam, PP & Loebel, DA 2007, 'Gene function in mouse embryogenesis: get set for gastrulation', *Nat Rev Genet*, vol. 8, no. 5, May, pp. 368-381.

Tange, TO, Shibuya, T, Jurica, MS & Moore, MJ 2005, 'Biochemical analysis of the EJC reveals two new factors and a stable tetrameric protein core', *Rna*, vol. 11, no. 12, Dec, pp. 1869-1883.

Taniguchi, M, Sanbo, M, Watanabe, S, Naruse, I, Mishina, M & Yagi, T 1998, 'Efficient production of Cre-mediated site-directed recombinants through the utilization of the puromycin resistance gene, pac: a transient gene-integration marker for ES cells', *Nucleic Acids Res*, vol. 26, no. 2, Jan 15, pp. 679-680.

Tarasov, A, Vilella, AJ, Cuppen, E, Nijman, IJ & Prins, P 2015, 'Sambamba: fast processing of NGS alignment formats', *Bioinformatics*, vol. 31, no. 12, Jun 15, pp. 2032-2034.

Tarpey, PS, Raymond, FL, Nguyen, LS, Rodriguez, J, Hackett, A, Vandeleur, L, Smith, R, Shoubridge, C, Edkins, S, Stevens, C, O'Meara, S, Tofts, C, Barthorpe, S, Buck, G, Cole, J, Halliday, K, Hills, K, Jones, D, Mironenko, T, Perry, J, Varian, J, West, S, Widaa, S, Teague, J, Dicks, E, Butler, A, Menzies, A, Richardson, D, Jenkinson, A, Shepherd, R, Raine, K, Moon, J, Luo, Y, Parnau, J, Bhat, SS, Gardner, A, Corbett, M, Brooks, D, Thomas, P, Parkinson-Lawrence, E, Porteous, ME, Warner, JP, Sanderson, T, Pearson, P, Simensen, RJ, Skinner, C, Hoganson, G, Superneau, D, Wooster, R, Bobrow, M, Turner, G, Stevenson, RE, Schwartz, CE, Futreal, PA, Srivastava, AK, Stratton, MR & Gecz, J 2007, 'Mutations in UPF3B, a member of the nonsense-mediated mRNA decay complex, cause syndromic and nonsyndromic mental retardation', *Nat Genet*, vol. 39, no. 9, Sep, pp. 1127-1133.

Terns, MP & Terns, RM 2011, 'CRISPR-based adaptive immune systems', *Curr Opin Microbiol*, vol. 14, no. 3, Jun, pp. 321-327.

Tesar, PJ, Chenoweth, JG, Brook, FA, Davies, TJ, Evans, EP, Mack, DL, Gardner, RL & McKay, RD 2007, 'New cell lines from mouse epiblast share defining features with human embryonic stem cells', *Nature*, vol. 448, no. 7150, Jul 12, pp. 196-199.

Tessier-Lavigne, M & Goodman, CS 1996, 'The molecular biology of axon guidance', *Science*, vol. 274, no. 5290, Nov 15, pp. 1123-1133.

Thomson, JA, Itskovitz-Eldor, J, Shapiro, SS, Waknitz, MA, Swiergiel, JJ, Marshall, VS & Jones, JM 1998, 'Embryonic stem cell lines derived from human blastocysts', *Science*, vol. 282, no. 5391, Nov 6, pp. 1145-1147.

Togashi, H, Sakisaka, T & Takai, Y 2009, 'Cell adhesion molecules in the central nervous system', *Cell Adh Migr*, vol. 3, no. 1, Jan-Mar, pp. 29-35.

Tosca, L, Feraud, O, Magniez, A, Bas, C, Griscelli, F, Bennaceur-Griscelli, A & Tachdjian, G 2015, 'Genomic instability of human embryonic stem cell lines using different passaging culture methods', *Mol Cytogenet*, vol. 8, p. 30.

Udan, RS, Culver, JC & Dickinson, ME 2013, 'Understanding vascular development', *Wiley Interdiscip Rev Dev Biol*, vol. 2, no. 3, May-Jun, pp. 327-346.

Um, SM, Ha, S, Lee, H, Kim, J, Kim, K, Shin, W, Cho, YS, Roh, JD, Kang, J, Yoo, T, Noh, YW, Choi, Y, Bae, YC & Kim, E 2018, 'NGL-2 Deletion Leads to Autistic-like Behaviors Responsive to NMDAR Modulation', *Cell Rep*, vol. 23, no. 13, Jun 26, pp. 3839-3851.

Urbach, A & Benvenisty, N 2009, 'Studying early lethality of 45,XO (Turner's syndrome) embryos using human embryonic stem cells', *PLoS One*, vol. 4, no. 1, p. e4175.

Urbach, A, Schuldiner, M & Benvenisty, N 2004, 'Modeling for Lesch-Nyhan disease by gene targeting in human embryonic stem cells', *Stem Cells*, vol. 22, no. 4, pp. 635-641.

Urnov, FD, Miller, JC, Lee, Y-L, Beausejour, CM, Rock, JM, Augustus, S, Jamieson, AC, Porteus, MH, Gregory, PD & Holmes, MC 2005, 'Highly efficient endogenous human gene correction using designed zinc-finger nucleases', *Nature*, vol. 435, no. 7042, p. 646.

Valamehr, B, Abujarour, R, Robinson, M, Le, T, Robbins, D, Shoemaker, D & Flynn, P 2012, 'A novel platform to enable the high-throughput derivation and characterization of feeder-free human iPSCs', *Sci Rep*, vol. 2, p. 213.

Valastyan, JS & Lindquist, S 2014, 'Mechanisms of protein-folding diseases at a glance', *Disease Models & Mechanisms*, vol. 7, no. 1, pp. 9-14.

van de Leemput, J, Boles, NC, Kiehl, TR, Corneo, B, Lederman, P, Menon, V, Lee, C, Martinez, RA, Levi, BP, Thompson, CL, Yao, S, Kaykas, A, Temple, S & Fasano, CA 2014, 'CORTECON: a temporal transcriptome analysis of in vitro human cerebral cortex development from human embryonic stem cells', *Neuron*, vol. 83, no. 1, Jul 2, pp. 51-68.

Van der Auwera, G 2012, <https://gatkforums.broadinstitute.org/gatk/discussion/2806/howto-apply-hard-filters-to-a-call-set>.

Van der Auwera, GA, Carneiro, MO, Hartl, C, Poplin, R, Del Angel, G, Levy-Moonshine, A, Jordan, T, Shakir, K, Roazen, D, Thibault, J, Banks, E, Garimella, KV, Altshuler, D, Gabriel, S & DePristo, MA 2013, 'From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline', *Curr Protoc Bioinformatics*, vol. 43, pp. 11 10 11-33.

Veltman, IM, Vreede, LA, Cheng, J, Looijenga, LH, Janssen, B, Schoenmakers, EF, Yeh, ET & van Kessel, AG 2005, 'Fusion of the SUMO/Sentrin-specific protease 1 gene SENP1 and the embryonic polarity-related mesoderm development gene MESDC2 in a patient with an infantile teratoma and a constitutional t (12; 15)(q13; q25)', *Human Molecular Genetics*, vol. 14, no. 14, pp. 1955-1963.

Veres, A, Gosis, BS, Ding, Q, Collins, R, Ragavendran, A, Brand, H, Erdin, S, Cowan, CA, Talkowski, ME & Musunuru, K 2014, 'Low incidence of off-target mutations in individual CRISPR-Cas9 and TALEN targeted human stem cell clones detected by whole-genome sequencing', *Cell Stem Cell*, vol. 15, no. 1, Jul 03, pp. 27-30.

Verpelli, C & Sala, C 2012, 'Molecular and synaptic defects in intellectual disability syndromes', *Current Opinion in Neurobiology*, vol. 22, no. 3, Jun, pp. 530-536.

Verrecchia, F & Mauviel, A 2007, 'Transforming growth factor-beta and fibrosis', *World Journal of Gastroenterology*, vol. 13, no. 22, Jun 14, pp. 3056-3062.

Walsh, DW, Godson, C, Brazil, DP & Martin, F 2010, 'Extracellular BMP-antagonist regulation in development and disease: tied up in knots', *Trends Cell Biol*, vol. 20, no. 5, May, pp. 244-256.

Wang, H & Sun, W 2017, 'CRISPR-mediated targeting of HER2 inhibits cell proliferation through a dominant negative mutation', *Cancer Letters*, vol. 385, Jan 28, pp. 137-143.

Wang, T, Wei, JJ, Sabatini, DM & Lander, ES 2014, 'Genetic screens in human cells using the CRISPR-Cas9 system', *Science*, vol. 343, no. 6166, Jan 3, pp. 80-84.

Wang, T, Wei, JJ, Sabatini, DM & Lander, ES 2014, 'Genetic screens in human cells using the CRISPR-Cas9 system', *Science*, vol. 343, no. 6166, pp. 80-84.

Wang, Z, Ballut, L, Barbosa, I & Le Hir, H 2018, 'Exon Junction Complexes can have distinct functional flavours to regulate specific splicing events', *Scientific Reports*, vol. 8, no. 1, p. 9509.

Washbourne, P, Dityatev, A, Scheiffele, P, Biederer, T, Weiner, JA, Christopherson, KS & El-Husseini, A 2004, 'Cell adhesion molecules in synapse formation', *J Neurosci*, vol. 24, no. 42, Oct 20, pp. 9244-9249.

Weischenfeldt, J, Damgaard, I, Bryder, D, Theilgaard-Monch, K, Thoren, LA, Nielsen, FC, Jacobsen, SE, Nerlov, C & Porse, BT 2008, 'NMD is essential for hematopoietic stem and progenitor cells and for eliminating by-products of programmed DNA rearrangements', *Genes Dev*, vol. 22, no. 10, May 15, pp. 1381-1396.

Welch, EM & Jacobson, A 1999, 'An internal open reading frame triggers nonsense-mediated decay of the yeast SPT10 mRNA', *Embo j*, vol. 18, no. 21, Nov 1, pp. 6134-6145.

Wen, Z 2017, 'Modeling neurodevelopmental and psychiatric diseases with human iPSCs', *Journal of Neuroscience Research*, vol. 95, no. 5, pp. 1097-1109.

Wen, Z, Christian, KM, Song, H & Ming, G-l 2016, 'Modeling psychiatric disorders with patient-derived iPSCs', *Current Opinion in Neurobiology*, vol. 36, pp. 118-127.

White, J & Dalton, S 2005, 'Cell cycle control of embryonic stem cells', *Stem Cell Rev*, vol. 1, no. 2, pp. 131-138.

Wilson, PG & Stice, SS 2006, 'Development and differentiation of neural rosettes derived from human embryonic stem cells', *Stem Cell Rev*, vol. 2, no. 1, pp. 67-77.

Wittkopp, N, Huntzinger, E, Weiler, C, Sauliere, J, Schmidt, S, Sonawane, M & Izaurralde, E 2009, 'Nonsense-mediated mRNA decay effectors are essential for zebrafish embryonic development and survival', *Mol Cell Biol*, vol. 29, no. 13, Jul, pp. 3517-3528.

Wittmann, J, Hol, EM & Jack, HM 2006, 'hUPF2 silencing identifies physiologic substrates of mammalian nonsense-mediated mRNA decay', *Mol Cell Biol*, vol. 26, no. 4, Feb, pp. 1272-1287.

Woo, J, Kwon, SK & Kim, E 2009, 'The NGL family of leucine-rich repeat-containing synaptic adhesion molecules', *Mol Cell Neurosci*, vol. 42, no. 1, Sep, pp. 1-10.

Wood, AJ, Lo, T-W, Zeitler, B, Pickle, CS, Ralston, EJ, Lee, AH, Amora, R, Miller, JC, Leung, E & Meng, X 2011, 'Targeted genome editing across species using ZFNs and TALENs', *Science*, vol. 333, no. 6040, pp. 307-307.

Xin, M, Davis, CA, Molkentin, JD, Lien, CL, Duncan, SA, Richardson, JA & Olson, EN 2006, 'A threshold of GATA4 and GATA6 expression is required for cardiovascular development', *Proc Natl Acad Sci U S A*, vol. 103, no. 30, Jul 25, pp. 11189-11194.

Xu, C, Inokuma, MS, Denham, J, Golds, K, Kundu, P, Gold, JD & Carpenter, MK 2001, 'Feeder-free growth of undifferentiated human embryonic stem cells', *Nature Biotechnology*, vol. 19, no. 10, Oct, pp. 971-974.

Xu, X, Wang, L, Liu, B, Xie, W & Chen, Y-G 2018, 'Activin/Smad2 and Wnt/β-catenin up-regulate HAS2 and ALDH3A2 to facilitate mesendoderm differentiation of human embryonic stem cells', *Journal of Biological Chemistry*, vol. 293, no. 48, pp. 18444-18453.

Xu, X, Zhang, L, Tong, P, Xun, G, Su, W, Xiong, Z, Zhu, T, Zheng, Y, Luo, S, Pan, Y, Xia, K & Hu, Z 2013, 'Exome sequencing identifies UPF3B as the causative gene for a Chinese non-syndrome mental retardation pedigree', *Clin Genet*, vol. 83, no. 6, Jun, pp. 560-564.

Yamagata, M, Sanes, JR & Weiner, JA 2003, 'Synaptic adhesion molecules', *Curr Opin Cell Biol*, vol. 15, no. 5, Oct, pp. 621-632.

Yamashita, A 2013, 'Role of SMG-1-mediated Upf1 phosphorylation in mammalian nonsense-mediated mRNA decay', *Genes Cells*, vol. 18, no. 3, Mar, pp. 161-175.

Yang, L, Grishin, D, Wang, G, Aach, J, Zhang, C-Z, Chari, R, Homsy, J, Cai, X, Zhao, Y & Fan, J-B 2014, 'Targeted and genome-wide sequencing reveal single nucleotide variations impacting specificity of Cas9 in human stem cells', *Nat Commun*, vol. 5, p. 5507.

Yang, L, Yang, JL, Byrne, S, Pan, J & Church, GM 2014, 'CRISPR/Cas9-Directed Genome Editing of Cultured Cells', *Current Protocols in Molecular Biology*, vol. 107, Jul 01, pp. 31 31 31-17.

Yap, K & Makeyev, EV 2013, 'Regulation of gene expression in mammalian nervous system through alternative pre-mRNA splicing coupled with RNA quality control mechanisms', *Mol Cell Neurosci*, vol. 56, Sep, pp. 420-428.

Yazdi, PG, Pedersen, BA, Taylor, JF, Khattab, OS, Chen, Y-H, Chen, Y, Jacobsen, SE & Wang, PH 2015, 'Increasing nucleosome occupancy is correlated with an increasing mutation rate so long as DNA repair machinery is intact', *PLoS One*, vol. 10, no. 8, p. e0136574.

Yazdi, PG, Pedersen, BA, Taylor, JF, Khattab, OS, Chen, YH, Chen, Y, Jacobsen, SE & Wang, PH 2015, 'Increasing Nucleosome Occupancy Is Correlated with an Increasing

Mutation Rate so Long as DNA Repair Machinery Is Intact', *PLoS One*, vol. 10, no. 8, p. e0136574.

Yepiskoposyan, H, Aeschimann, F, Nilsson, D, Okoniewski, M & Muhlemann, O 2011, 'Autoregulation of the nonsense-mediated mRNA decay pathway in human cells', *Rna*, vol. 17, no. 12, Dec, pp. 2108-2118.

Yu, G, Wang, L-G, Han, Y & He, Q-Y 2012, 'clusterProfiler: an R package for comparing biological themes among gene clusters', *OMICS: A Journal of Integrative Biology*, vol. 16, no. 5, pp. 284-287.

Yu, J, Vodyanik, MA, Smuga-Otto, K, Antosiewicz-Bourget, J, Frane, JL, Tian, S, Nie, J, Jonsdottir, GA, Ruotti, V, Stewart, R, Slukvin, II & Thomson, JA 2007, 'Induced pluripotent stem cell lines derived from human somatic cells', *Science*, vol. 318, no. 5858, Dec 21, pp. 1917-1920.

Yung, SK, Tilgner, K, Ledran, MH, Habibollah, S, Neganova, I, Singhapol, C, Saretzki, G, Stojkovic, M, Armstrong, L, Przyborski, S & Lako, M 2013, 'Brief report: human pluripotent stem cell models of fanconi anemia deficiency reveal an important role for fanconi anemia proteins in cellular reprogramming and survival of hematopoietic progenitors', *Stem Cells*, vol. 31, no. 5, May, pp. 1022-1029.

Zahdeh, F & Carmel, L 2016, 'The role of nucleotide composition in premature termination codon recognition', *BMC Bioinformatics*, vol. 17, no. 1, Dec 7, p. 519.

Zetoune, AB, Fontanière, S, Magnin, D, Anczuków, O, Buisson, M, Zhang, CX & Mazoyer, S 2008, 'Comparison of nonsense-mediated mRNA decay efficiency in various murine tissues', *BMC Genetics*, vol. 9, no. 1, p. 83.

Zhang, W, Chen, Z, Zhang, D, Zhao, B, Liu, L, Xie, Z, Yao, Y & Zheng, P 2019, 'KHDC3L mutation causes recurrent pregnancy loss by inducing genomic instability of human early embryonic cells', *PLoS Biol*, vol. 17, no. 10, Oct, p. e3000468.

Zhang, Z, Xin, D, Wang, P, Zhou, L, Hu, L, Kong, X & Hurst, LD 2009, 'Noisy splicing, more than expression regulation, explains why some exons are subject to nonsense-mediated mRNA decay', *BMC Biol*, vol. 7, p. 23.

Zhao, R, Choi, BY, Lee, M-H, Bode, AM & Dong, Z 2016, 'Implications of genetic and epigenetic alterations of CDKN2A (p16INK4a) in cancer', *EBioMedicine*, vol. 8, pp. 30-39.

Zhao, X & Bhattacharyya, A 2018, 'Human Models Are Needed for Studying Human Neurodevelopmental Disorders', *Am J Hum Genet*, vol. 103, no. 6, Dec 6, pp. 829-857.