Generation of Stem Cell Models of ITPR1-Related Neurological Disorders Using a Novel Genome Engineering Approach

Ricardo Parolin Schnekenberg

Merton College

A thesis submitted for the degree of

Doctor of Philosophy

Trinity Term 2023

Acknowledgements

I would like to express my immense gratitude to my primary supervisor, Prof. Andrea Németh, who, nearly 12 years ago, invited a young medical student to 'help her with some research in cerebellar ataxias'. She supported me not only academically but also personally throughout the years, even in periods when I was living and working in Brazil and had all but given up pursuing an academic career. I also thank her for being incredibly understanding and encouraging during the COVID pandemic, when my attention mostly shifted to epidemiological research.

Similarly I am forever indebted to Dr. Philip Hublitz, my secondary supervisor, for the amazing dedication, effort and patience he has shown in helping me through the final years of this DPhil project. Despite leaving Oxford to return to his native Germany, he continued to be available for numerous cycles of thesis revisions and has been incredibly responsive to my many emails.

The incredibly talented Dr. Hannah Sleven, a postdoctoral research scientist in the Nemeth lab, taught me how to properly maintain stem cell cultures and is credited with having the original idea which eventually became this entire thesis. After Hannah left, being the only member of a lab was not easy but it was made easier by the help and support provided by members of Prof. Robert MacLaren's lab. Thank you Dr. Christina Martinez (who taught me how to run Western blots), Dr. Ahmad Salman (taught me immunocytochemistry and confocal microscopy), Dr. Michelle McClements (who kindly allowed me to share reagents and equipment), Dr. Caroline Peddle, and fellow Mertonians Dr. Laurel Chandler, Dr. Lewis Fry, Dr. Imran Yusuf and Dr. Harry Orlans.

I am also thankful to members of Prof. Stephanie Halford's lab (Dr. Solomon Merepa and Dr. Suzanne Broadgate), Prof. George Tofaris's lab (Dr. Benedict Tanudjojo), Prof. Kevin Talbot's lab (Dr. Ruxandra Dafinca, who taught me calcium imaging), Prof. David Bennett's lab (Dr. Lucy McDermott, Dr. Alex Davies, Dr. Teodora Trendafilova) and Prof. Esther Becker's lab. I am very grateful to Dr. Jussi-pekka Tolonen with whom I spent countless hours debating possible mechanisms of ITPR1 pathology and every single ITPR1 variant ever reported.

I have no doubt that being a member of the beautiful Merton College made my academic and social experience in Oxford much more pleasant and enriching. I thank Merton College for being incredibly understanding and for funding my rather long postgraduate degree. Similarly, I am immensely grateful to the Clarendon Fund for providing me with a prestigious Clarendon

scholarship, without which I would never have been able to pursue this DPhil.

I would like to thank my close friends and family for supporting me through this academic roller coaster. At the most difficult times of this project, my incredible girlfriend Dominique Doering suffered with me and because of me. She stood by my side through the entire duration of this DPhil and should be credited for me actually finishing it. In this foreign country, she was and is my support system. I thank her and also her family, who have taken me as one of their own and made me feel I can call three different countries home.

Finally, I would like to show my most sincere gratitude for my parents, Marcia Parolin and Paulo Schnekenberg, for their unwavering emotional, intellectual and financial support. Despite my extremely long absence from family life, very occasional phone calls and near complete lack of desire to engage in conversations about my work, my parents somehow remained firm in their support and motivated me to complete this DPhil. Thank you.

Table of Contents

Abstract

Mutations in the ITPR1 gene cause different forms of neurological disease in humans. Large heterozygous deletions result in adult-onset SCA15, heterozygous missense mutations cause infantile-onset SCA29 or Gillespie syndrome, and homozygous loss-of-function mutations cause autosomal recessive Gillespie syndrome. However, the mechanisms by which ITPR1 mutations lead to motor and cognitive symptoms are still poorly understood. Stem cell models hold promise for advancing mechanistic research and drug development, but their generation has been notoriously challenging despite the recent advances in genome engineering technology.

This thesis describes the development of a CRISPR/Cas9 system designed to introduce ITPR1 mutations into human stem cells using positive-negative antibiotic selection. The aim was to develop a system that could not only generate stem cell models of different ITPR1 mutations but also be easily adapted for any other genes or mutations of interest.

The Cassette Insertion-Replacement CRISPR Editing System (CIRCES) was designed as two stages of CRISPR/Cas9-induced homology-directed repair. Sequential application of positive and negative selection was planned to remove the overwhelming majority of wild-type cells, thereby increasing the frequency of correctly edited cells. Constructs were generated to introduce SCA29 and Gillespie syndrome mutations, with individual components of the system initially tested in HEK293T cells before further optimisation in H9 human embryonic stem cells.

Integration of the selection cassette was achieved in all three ITPR1 target exons. Analysis of clonally-isolated colonies revealed cassette insertion was always identified in combination with an *in trans* small insertion and/or deletion, suggestive of biallelic loss of function. While no heterozygous missense point mutations were generated with CIRCES, seven clones with biallelic loss-of-function mutations were confirmed to be ITPR1 knockouts, thereby reproducing the genetic defect found in autosomal recessive Gillespie syndrome. Loss of ITPR1 was not compensated for by upregulation of other IP_3 receptors and neuronal differentiation induced

significant downregulation of IP_3 receptors in both knockouts and controls. Despite ITPR1 knockout, intracellular calcium dynamics were unchanged.

The results presented in this thesis suggest total loss of ITPR1 is tolerated in human embryonic stem cells and during early neuronal development. The methods used were able to generate several stem cell lines that can be used as cellular models of Gillespie syndrome, which should be useful tools in the study of ITPR1-related diseases and calcium signalling. Adaptation of this approach to different targets will facilitate generation of cellular models for other human diseases.

List of Abbreviations

A-EJ ..... Alternative End Joining
AD ..... Autosomal Dominant
ADCA ..... Autosomal Dominant Cerebellar Ataxia
AMPAR ..... α-Amino-3-Hydroxy-5-Methyl-4-Isoxazolepropionic Acid Receptor
AR ..... Autosomal Recessive
Ara-C ..... Cytosine Arabinoside (Cytarabine)
ARM ..... Armadillo Repeat Motif
ASE ..... Antigen Signal Enhancement
ATP ..... Adenosine Triphosphate
BDNF ..... Brain-Derived Neurotrophic Factor
bGH ..... Bovine Growth Factor
BSD ..... Blasticidin-S-Deaminase
c-NHEJ ..... Canonical Non-Homologous End Joining
CA8 ..... Carbonic Anhydrase 8
CARP ..... Carbonic Anhydrase Related Protein
Cas9 ..... CRISPR-Associated Protein 9
CHO ..... Chinese Hamster Ovary cells
CIRCES ..... Cassette Insertion-Replacement CRISPR/Cas9 Editing System
CRISPR ..... Clustered Regularly Interspaced Short Palindromic Repeats
crRNA ..... CRISPR-encoded RNA
CTD ..... C-terminal Domain
DAG ..... Diacylglycerol
DAPI 4,6-diamidino-2-phenylindole
DD ..... Destabilisation Domain
DHPG ..... Dihydroxyphenylglycine
DKO ..... Double Knock-Out
DMEM ..... Dulbecco's Modified Eagle's Medium
DMSO ..... Dimethyl Sulfoxide
DPBS ..... Dulbecco's Phosphate Buffered Saline
DSB ..... Double Strand Break
dsDNA ..... Double-Stranded DNA
DSS ..... Drosophila-Specific Sequence
EDTA ..... Ethylenediaminetetraacetic Acid
EF1α ..... Elongation Factor 1α
EGFP ..... Enhanced Green Fluorescent Protein

ER ..... Endoplasmic Reticulum
ESC ..... Embryonic Stem Cell
ExoSAP ..... Exonuclease I / Shrimp Alkaline Phosphatase
FACS ..... Fluorescence-Activated Cell Sorting
FBS ..... Foetal Bovine Serum
FCS ..... Foetal Calf Serum
FDR ..... False-Discovery Rate
gDNA ..... Genomic DNA
GFR ..... Growth-Factor Reduced
GPCR ..... G Protein-Coupled Receptor
GS ..... Gillespie Syndrome
HBSS ..... Hanks' Balanced Salt Solution
HDR ..... Homology-Directed Repair
HEPES ..... 4-(2-Hydroxyethyl)-1-Piperazineethanesulfonic Acid
hESC ..... Human Embryonic Stem Cell
HR ..... Homologous Recombination
HSV ..... Herpes Simplex Virus
IBC ..... $IP_[3]$-Binding Core
IBD ..... $IP_[3]$-Binding Domain
ICC ..... Immunocytochemistry
ICE ..... Inference of CRISPR Edits
InsP3R1 ..... Inositol 1,4,5-Trisphosphate Receptor Type 1
IP3 ..... Inositol 1,4,5-Trisphosphate
IP3R1 ..... Inositol 1,4,5-Trisphosphate Receptor Type 1 (Protein)
IP3R2 ..... Inositol 1,4,5-Trisphosphate Receptor Type 2 (Protein)
IP3R3 ..... Inositol 1,4,5-Trisphosphate Receptor Type 3 (Protein)
iPSC ..... Induced Pluripotent Stem Cells
IPTG ..... Isopropyl β-D-1-Thiogalactopyranoside
ITPR1 ..... Inositol 1,4,5-Trisphosphate Receptor Type 1 (Gene)
ITPR2 ..... Inositol 1,4,5-Trisphosphate Receptor Type 2 (Gene)
ITPR3 ..... Inositol 1,4,5-Trisphosphate Receptor Type 3 (Gene)
KO ..... Knock-Out
LB ..... Lysogeny Broth
LEI ..... Left External-to-Internal
LOH ..... Loss of Heterozygosity
LTD ..... Long-Term Depression
MACS ..... Magnetic-Assisted Cell Sorting

mESC ..... Mouse Embryonic Stem CEll
mGluR ..... Metabotropic Glutamate Receptor
MMEJ ..... Microhomology-Mediated End Joining
MRI ..... Magnetic Resonance Imaging
Ngn2 ..... Neurogenin-2
NGS ..... Next-Generation Sequencing
NHEJ ..... Non-Homologous End Joining
NLS ..... Nuclear Localisation Signal
NMD ..... Nonsense-Mediated Decay
pA ..... Polyadenylation Signal
PAGE ..... Polyacrylamide Gel Electrophoresis
PAM ..... Protospacer-Adjacent Motif
PBS ..... Phosphate Buffered Saline
PBST ..... Phosphate Buffered Saline with Tween-20
PC ..... Purkinje Cell
PCH ..... Pontocerebellar Hypoplasia
PCR ..... Polymerase Chain Reaction
PFA ..... Paraformaldehyde
PKC ..... Protein Kinase C
PLC ..... Phospholipase C
PMSF ..... Phenylmethylsulfonyl Fluoride
PND ..... Post-natal Day
PTD ..... Post-transfection Day
PuroR ..... Puromycin N-Acetyl-Transferase Resistance (Gene)
PVDF ..... Polyvinylidene Fluoride
qPCR ..... Quantitative Polymerase Chain Reaction
REI ..... Right External-to-Internal
RFLP ..... Restriction Fragment Length Polymorphism
RNP ..... Ribonucleoprotein
ROCKi ..... Rho-Associated Coiled Kinase Inhibitor (Y-27632 Dihydrochloride)
RPM ..... Revolutions Per Minute
rSAP ..... Recombinant Shrimp Alkaline Phosphatase
RT ..... Room Temperature
RT-PCR ..... Reverse Transcription Polymerase Chain Reaction
SCA ..... Spinocerebellar Ataxia
SD ..... Supressor Domain
SDS ..... Sodium Dodecyl Sulfate

sgRNA ..... Single Guide RNA
SMAD ..... Suppressor of Mothers Against Decapentaplegic
SNP ..... Single Nucleotide Polymorphism
SpCas9 ..... *Streptococcus pyogenes* Cas9
ssDNA ..... Single-Stranded DNA
ssODN ..... Single-Stranded Oligodeoxynucleotide
TALEN ..... Transcription Activator-Like Effector Nuclease
TBS ..... Tris Buffered Saline
TBST ..... Tris Buffered Saline with 0.1% Tween-20
TIDE ..... Tracking of Indels by Decomposition
TK ..... Thymidine Kinase
TM ..... Transmembrane
TMD ..... Transmembrane Domain
tracrRNA ..... trans-activating crRNA
WIMM ..... Weatherall Institute of Molecular Medicine (Oxford)
WT ..... Wild-Type
X-Gal 5-Bromo-4-Chloro-3-Indolyl-Beta-D-Galacto-Pyranoside
ZFN ..... Zinc-Finger Nuclease

1 | Introduction

Hereditary disorders have long been known to affect the nervous system. The advent of Next-Generation Sequencing (NGS), over the past 15 years, has revolutionised clinical diagnostics and enabled researchers to identify hundreds of new genes associated with human disease. However, the rate of gene and mutation discovery has far outpaced the ability to study these newly-defined disorders. In particular, the functional consequences of mutations on the nervous system are often very unclear.

Functional analysis of gene variants relevant to neurological disease is especially challenging as the nervous system is relatively inaccessible in humans, and the relevance of animal models to human disease can be uncertain. A new era of cellular modelling, combining stem cell technology with novel genome engineering methods has emerged fairly recently to add another set of tools with which to understand neurological disease.

The aim of my thesis was to develop a novel genome engineering system to enable rapid and efficient introduction of mutations into human embryonic stem cells. As proof of principle, I knocked out ITPR1, a gene known to be involved in several neurogenetic disorders. I will ultimately also discuss how this system can be adapted to target other genes, enabling its application in the study of a wide variety of human disorders.

1.1 Human Disorders Associated with ITPR1 Mutations

1.1.1 Gillespie Syndrome

In 1965 Frederick Gillespie reported two siblings, born from healthy non-consanguineous parents, with bilateral partial aniridia, cerebellar ataxia and intellectual disability[1]. Gillespie syndrome, as it came to be known, was for decades assumed to be autosomal recessive, although at least one publication had reported direct transmission from an affected parent to their offspring[2].

Gillespie syndrome is normally suspected at birth or in the first few months of life in hypotonic

infants with fixed dilated pupils. Careful ophthalmologic examination shows bilateral partially-formed irises that commonly have a distinct scalloped appearance due to abnormal development of the iris sphincter and associated stroma[3]. Patients show developmental delay, ataxia and intellectual disability. Most reported patients do not walk unassisted until at least age 6, while many in their second decade of life have not yet reached this motor milestone[4, 5]. Intellectual disability is often reported as moderate or severe. Additionally, extra-neurological features such as pulmonary valve stenosis, gastroesophageal reflux, scoliosis and mild facial dysmorphism are often present[4–8].

Almost invariably, patients have cerebellar atrophy identified on MRI, but occasionally no atrophy is seen in patients younger than 2 years of age. The vermis, especially its superior portion, is mostly affected[5]. The radiological findings have been noted to be progressive, especially in the first 5 years of life, and can eventually also affect the cerebellar hemispheres, particularly the superior foliae[9].

1.1.2 SCA15/16

In 2001 two large multi-generational families showing a pattern of autosomal dominant adult-onset pure cerebellar ataxia were published[10, 11]. Initial linkage studies mapped the Australian family to chromosome 3p[12], whereas the phenotype in the Japanese family was initially mapped to chromosome 8q[11]. Although they were thought to represent different novel spinocerebellar ataxias (SCAs) — type 15 (SCA15) was assigned to the former and type 16 (SCA16) to the latter —, further assessments in the Japanese family suggested it too mapped to chromosome 3p[13]. Eventually both would be found to be caused by heterozygous deletions in ITPR1, therefore "SCA16 is really SCA15"[14]. Since then, 20 other families have been described, all with a remarkably similar phenotype marked by slow or very slow progression of gait and limb ataxia[15–23].

The disorder is for the most part a pure cerebellar ataxia, generally mild, with very slow progression that does not shorten lifespan, with most patients remaining ambulant even decades after disease onset. Reported ages of onset range anywhere between the ages of 10 to late

70s. Gait ataxia or dysarthria are commonly reported as first symptoms, but it may be heralded by years of unsteadiness, handwriting issues or clumsiness. Gaze-evoked nystagmus is found in most patients, although it may be transient in some. Head and/or truncal tremor has been inconstantly reported, and it may accompany upper extremity tremor[21].

Although the disorder was initially described as fitting a pure cerebellar ataxia consistent with ADCA type III in Anita Harding's classification, some families show additional features that might be associated with the phenotype. Mild dysphagia has been been reported in several families[20], usually after two decades of symptoms. Mild cognitive decline has been noticed in some patients after a long disease course[21], usually in their 70s or 80s, and at least one of these patients shows important bilateral frontoparietal atrophy on MRI[18]. Three families have been reported to show involuntary movements such as facial myokymias and buccolingual dyskynesias[18] or mild choreic movements of mouth and fingers[20].

Cerebellar atrophy is always identified on MRIs of SCA15 patients, with the vermis disproportionately affected, more so superiorly and dorsally, but without brainstem involvement[10, 11].

1.1.3 SCA29

Three years after the first SCA15 pedigrees were described, another autosomal dominant ataxia syndrome mapping to the same locus was described, coincidentally, in an Australian family[24]. Different from SCA15, however, patients showed poor balance and incoordination from a few months of age and the disorder was not progressive. And in contrast to Gillespie syndrome, patients had normally-formed irises. According to convention, this condition was called spinocerebellar ataxia type 29 (SCA29).

The original family spanned four generations and had at least 20 affected members with similar clinical histories, namely congenital ataxia, delayed walking and requirement of special education. Other specific signs of cerebellar dysfunction varied between patients, but deficits were generally mild and several family members reported slight improvements in coordination with age. Interestingly, in this family cognitive impairment was the most important disability. There

were no signs of pyramidal tract dysfunction or sensory changes, and lifespan was not affected by the condition.

While the 3-year-old ataxic proband had a normal MRI, the mother had atrophy of the cerebellar vermis. Of the five other affected family members who were imaged, three had normal scans, and there was no association between degree of ataxia or cognitive impairment and presence of radiological findings.

1.2 Genetics of ITPR1-Related Disorders

Over the past 15 years, Gillespie syndrome, SCA29 and SCA15/16 have all been found to be caused by mutations in ITPR1, a large gene that spans over 350kb of genomic sequence near the telomere of chromosome 3p. Its longest transcript has 62 exons, and the last 59 encode the 2758-residue protein IP3R1. Often also referred to as InsP3R1, the inositol-1,4,5-triphosphate receptor type 1 is a tetrameric ligand-gated calcium channel that is central in the IP_3 calcium signalling pathway. Several other neurological disorders, including many ataxias, have been directly or indirectly associated with dysfunctions in the IP_3 signalling pathway (Fig. 1.1).

1.2.1 SCA15 Genetics

ITPR1 was an early candidate gene for SCA15 but it was incorrectly excluded based on normal results on Sanger sequencing, Southern blotting and analysis of RNA transcripts[12]. Re-analysis of the Australian SCA15 family with high-density SNP genotyping years later identified a large genomic deletion encompassing genes SUMF1 and ITPR1 that segregated with the phenotype[15]. Additional families helped narrow the genetic defect down to ITPR1 exons 1-10 as a minimal deletion size known to cause SCA15, and at least two families are known to have whole ITPR1 gene deletions[16, 22]. A single family has been reported to have a heterozygous missense mutation (P1059L), but their phenotype included additional features such as mild peripheral neuropathy, low vitamin E and low apoprotein B levels that were not seen in other SCA15 patients[16]. Additionally, neither this specific mutation nor other missense mutations in ITPR1 have been reported in other adult-onset ataxia pedigrees[25] despite over

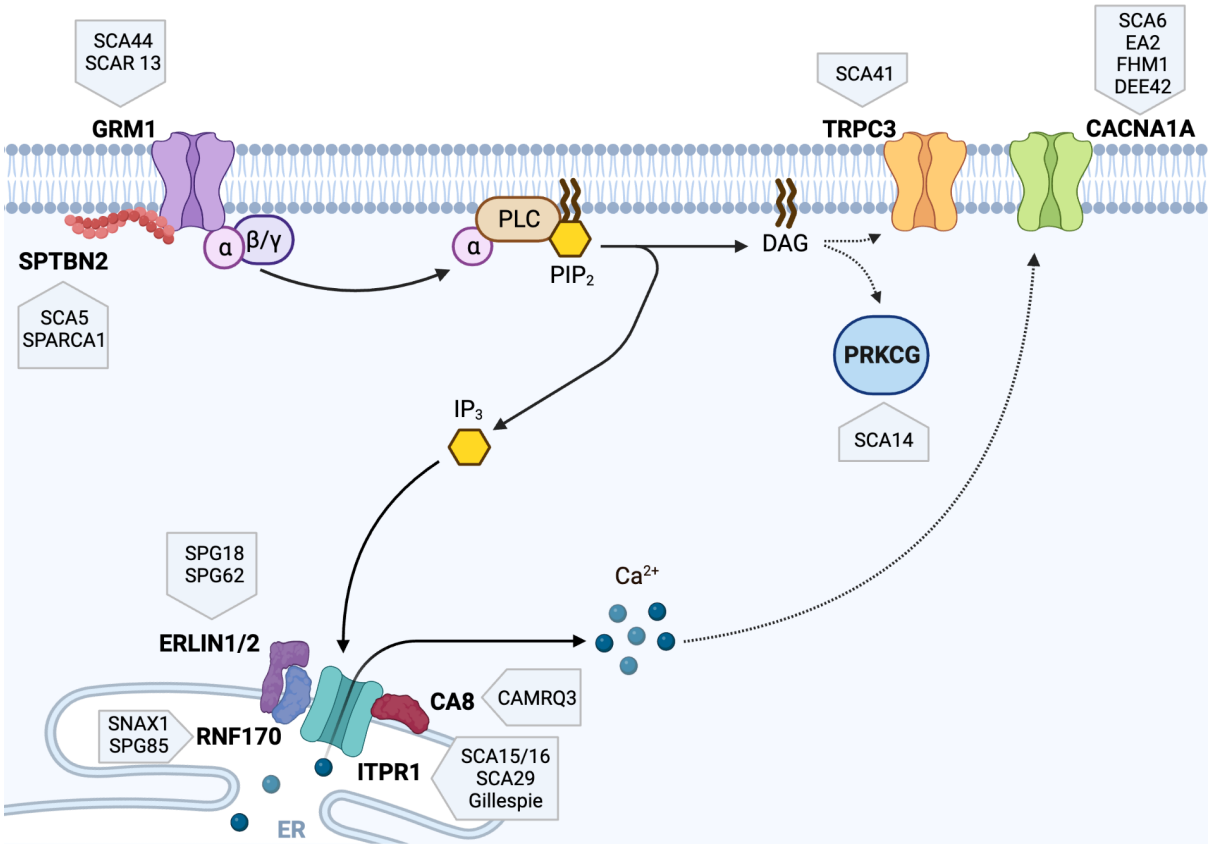

Figure 1.1: Many ataxias and other motor disorders are associated with mutations in genes coding for proteins that are directly or indirectly involved in the IP_3 signalling pathway. Shown in bold are the genes associated with human neurogenetic syndromes. Abbreviations for specific disorders associated with each gene are shown in the light blue arrows. The details of the IP_3 signalling pathway are explained in section 1.3.1.

a decade of next-generation sequencing being used as in the clinic as a diagnostic tool. Thus, it is generally accepted that SCA15 is an adult-onset ataxia caused by large genomic deletions affecting ITPR1 that is fully penetrant[21].

Interestingly, despite great availability of diagnostic methodologies that assess for copy-number changes at high genomic resolution, all SCA15-related deletions reported to date affect the 5' region of ITPR1. Notwithstanding its large genomic size, no intragenic or 3' deletions have been identified.

1.2.2 SCA29 Genetics

The advent of next-generation sequencing permitted the molecular cause of SCA29 to be established[26]. In the original family, whole-exome sequencing identified the missense ITPR1 mutation V1562M, which segregated with the phenotype. Since its original description, at least

one more family with the same mutation has been identified[27]. A three-generation Canadian family had a similar, albeit more severe, phenotype, which was associated with a heterozygous N602D mutation[26]. Prior to starting this project, as part of an exome sequencing project in the Nemeth lab, I reported another unrelated patient with a very similar phenotype who was also found to carry the same N602D mutation[28].

Following the initial identification of heterozygous point mutations in ITPR1 as causes of congenital ataxia, in general, any cases of congenital or infantile-onset ataxia associated with heterozygous missense mutations in ITPR1 were diagnosed as SCA29. Reported pedigrees became smaller and most cases were *de novo* rather than inherited[29–36]. Additionally, neonatal hypotonia, nystagmus and extra-neurological features such as strabismus were more commonly reported. The N602D and V1562M mutations have not been reported since, whereas newly reported cases in unrelated families showed clear clustering for exon 10, with few residues being repeatedly affected. For example, there have been 11 reported cases (in 6 publications) associated with mutations in T267, 9 cases (in 5 publications) associated with R269 and 4 cases (in 4 publications) linked to S277.

As of today, SCA29 is understood as a neurodevelopmental syndrome usually evident at birth or in the first few months of life. After uneventful pregnancies, these patients are typically born at term with normal height, weight and head circumference and the first manifestation is usually nystagmus or hypotonia[33]. As they age, delays in achieving early motor milestones such as head support and sitting raise the first red flags of a neurodevelopmental condition. Signs of cerebellar dysfunction become more clear with time, and intention and/or postural tremor may become apparent. Some patients might be initially diagnosed as ataxic cerebral palsy[28]. Most patients do not walk unassisted until 3-6 years of age, but for some it can take >10 years[33]. Patients show delayed speech and learning difficulties, and cognition is almost always impaired to some extent, with almost all cases being classified as mild intellectual disability. Although there are a few cases of normal intelligence reported, on psychometric evaluation these patients tend to show IQs in the low 70s. Seizures are not considered part of the phenotype, and there are usually no pyramidal or extrapyramidal signs. As mentioned, the disorder is not progressive and some families even notice minor improvements as children age[26–28]. MRIs can be normal in

the first couple years of life, but generally cerebellar atrophy, especially of the superior vermis, is seen[9].

1.2.3 Gillespie Syndrome Genetics

A molecular cause for Gillespie syndrome was not established until 2016 when two publications reported that affected patients had point mutations in ITPR1, and that the disorder had both autosomal dominant and recessive inheritance[4, 5]. Most sporadic cases were due to *de novo* missense mutations, whereas recessive cases were associated with biallelic loss-of-function mutations.

De novo mutations identified in Gillespie patients repeatedly affected only a few residues of the channel domain encoded by exons 58 and 59, such as E2109Q, G2554R, F2601L and K2611del[5]. Recessive mutations, on the other hand, were loss-of-function variants expected to cause protein truncation before the channel domain, such as H93fs, R728X and Q1567X[4].

1.3 Type 1 Inositol-1,4,5-Triphosphate Receptors

In response to certain extracellular neurotransmitter and hormones, mammalian cells increase their cytosolic calcium concentration even in the absence of extracellular calcium[37]. The cellular organelle that stores this pool of releasable calcium is the endoplasmic reticulum (ER) and the intracellular messenger that triggers its release is inositol-1,4,5-triphosphate (IP_3)[38]. Channels in the ER membrane that release calcium upon binding of IP_3 were therefore aptly named IP_3 Receptors. They were first purified from the cerebellum of rats[39], where IP_3 binding levels had been found to be several hundred times greater than in peripheral tissues[40]. This cerebellar IP_3 receptor, encoded by the ITPR1 gene, was labelled Type 1 inositol-1,4,5-triphosphate receptor, but is often referred to as IP3R1 or InsP3R1.

1.3.1 The Inositol Triphosphate Signalling Pathway

The IP_3 signalling pathway is one of the most well characterised ways by which cells transduce extracellular signals to initiate intracellular actions (Fig. 1.2). Certain neurotransmitters

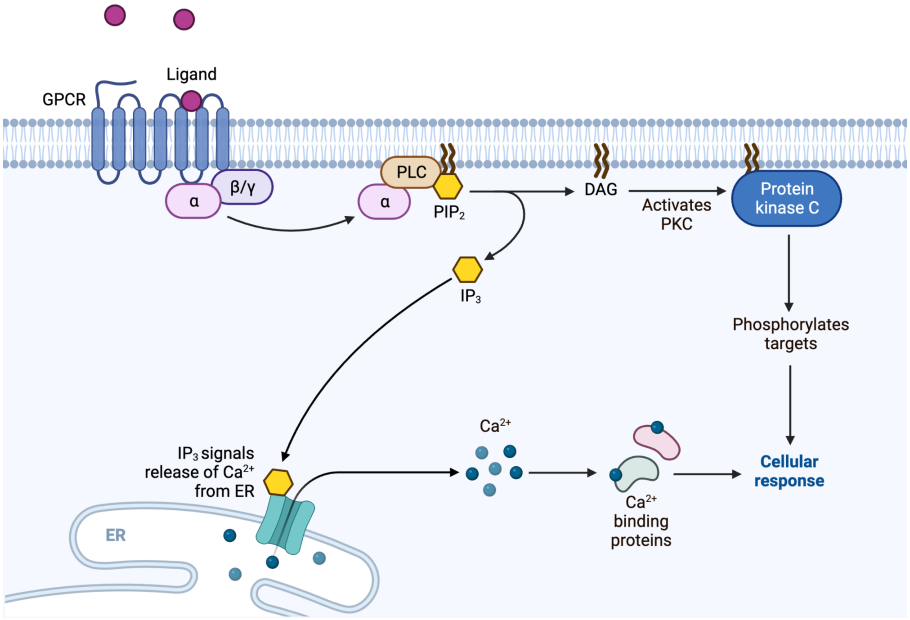

Figure 1.2: Schematics of the IP_3 signalling pathway. Binding of extracellular ligands to transmembrane G-protein coupled receptors (GPCRs) triggers the activation of phospholipase C (PLC), which catalyses the hydrolysis of membrane phospholipid PIP_2 into IP_3 and diacylglycerol (DAG). In response to IP_3, endoplasmic reticulum (ER) transmembrane IP_3 receptors allow the release of calcium to the cytoplasm which can start or amplify downstream signalling cascades.

(e.g. glutamate), growth factors and hormones, upon binding to the extracellular domain of transmembrane G Protein-Coupled Receptors (GPCRs), activate the enzyme Phospholipase C (PLC), which hydrolyses membrane phospholipid PIP_2 (phosphatidylinositol-4,5-biphosphate) into second-messengers IP_3 and diacylglycerol (DAG)[41]. While DAG can directly activate channels in the plasma membrane and initiate other pathways through activation of protein kinases such as PKC and PKD[42], the sole recognised function of IP_3 is to trigger the release of calcium from the ER through IP_3 Receptors. Temporal and spatial fluctuations in cytosolic calcium concentrations in turn trigger many different downstream effects, such as gene expression[43], apoptosis[44], neurogenesis[45] and synaptic plasticity[46].

1.3.2 Structure and Function of IP_3 Receptors

IP_3 Receptors exist as tetrameric mushroom-shaped channels with its stalk piercing the ER membrane through to the luminal side while the large cap sits on the cytoplasmatic side (Fig. 1.3)[47]. The central pore is formed through the assembly of the transmembrane domains of all four subunits[48], but each subunit is individually able to bind IP_3[49, 50]. Channel

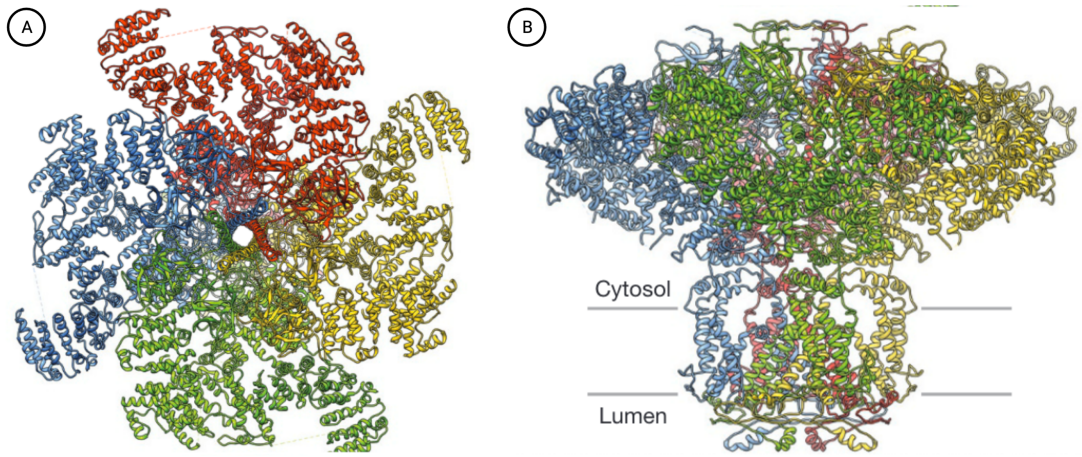

Figure 1.3: Tetrameric structure of IP_3 Receptor type 1. (A) View from the cytoplasmatic side, showing the four subunits (individually coloured) that assemble to create a transmembrane channel with a centrally-located pore. (B) Side view of the receptor. Images from ref. [47].

gating requires that all four IP_3-binding pockets be simultaneously occupied[51], effectively functioning as a coincidence detector for IP_3.

Each $\sim$310kDa subunit contains three major domains: a N-terminal IP_3-binding domain (IBD, $\sim$600aa), a central regulatory domain ($\sim$1650aa), and a C-terminal channel domain ($\sim$550aa) (Fig. 1.4). Both N and C termini are located in the cytoplasmatic side of the ER membrane (Fig. 1.4B). Mutations causing SCA29 cluster in the IP_3-binding domain, AD Gillespie mutations are exclusively located in or near the channel domain and AR Gillespie mutations cause premature protein termination mostly in the regulatory domain (Fig. 1.4A).

1.3.2.1 IP_3 Binding

The IBD is fully capable of binding IP_3. Mutagenesis studies revealed that small changes in the first 225 residues could lead to drastically reduced IP_3 binding. Similarly, it was found that IP_3 binding occurred with 10-fold higher affinity if all 225 N-terminal aminoacids were completely deleted[53, 54]. Therefore the first 225 N-terminal residues of IBD form a Suppressor Domain (SD)[53], while the remainder form the IP_3-Binding Core (IBC) (Fig. 1.4).

The core binding region is structurally formed by a β-trefoil fold (residues 224-436, hereby referred as IBC-β, yellow section on Fig. 1.4A) and an armadillo repeat foil (residues 437-604, referred to as IBC-α, red section on Fig. 1.4A). In the absence of an IP_3 molecule, the IBC-α and IBC-β are arranged in an open-clam configuration[55]. Both domains approximate

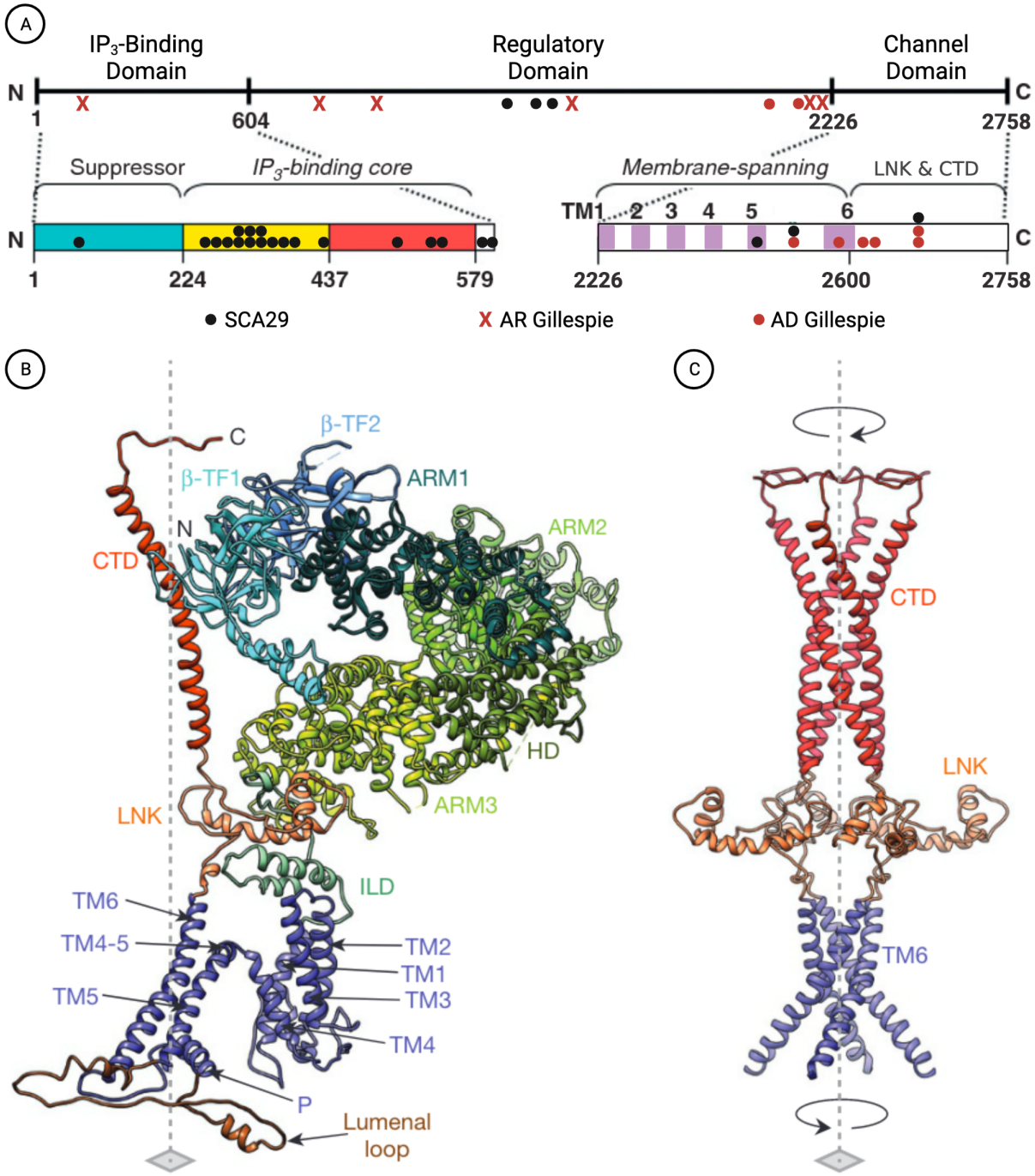

Figure 1.4: Structure of IP3R1 subunits. (A) Each subunit contains three major functional domains. The IP_3-Binding Domain (IBD) comprises the N-terminal 604 residues and is subdivided into a supressor domain (SD) and a IP_3-binding core (IBC). The channel domain is located in the C-terminus and is made of a membrane-spanning region and cytoplasmatic linker (LNK) and C-terminal (CTD) regions. Between the IBD and the CD is the long regulatory domain. Most SCA29 mutations (black dots) cluster in the IBD, whereas all heterozygous Gillespie syndrome-causing mutations (red dots) have so far been located in or near the CD. Recessive Gillespie mutations (red X's) are predicted to cause protein truncation, and up to now have only been described in the regulatory and supressor domains. Some residues in the CD, when mutated, can lead to either SCA29 or Gillespie syndrome. (B) Tertiary structure of a single IP3R1 subunit viwed from the side. The SD is shown as β-TF1 and the IBC is shown as β-TF2 (broadly equivalent to residues 224-436) and armadillo solenoid fold 1 (ARM1, residues 437-600). The regulatory domain forms the bulk of the cytoplasmatic side of each subunit and comprises ARM2, a helical domain (HD) and a intervening lateral domain (ILD), which is known to be involved in channel gating. The CD contains the six transmembrane (TM) domains, a linker region (LNK) and the CTD. In the lumenal side a loop is formed between TM5 and TM6 which contains a selectivity filter for the channel. (C) Central core structure of the assembled tetrameric receptor. Images adapted from refs. [47] and [52].

forming a 'closed-clam' in the presence of IP_3. The SD sits at the hinge, interacting with both IBC-β and IBC-α (Fig. 1.5A, SD shown was β-TF1, IBC-α as ARM1 and IBC-β as β-TF2). Highly basic residues in these two domains are critical for interactions with IP_3[53] (Fig. 1.5B).

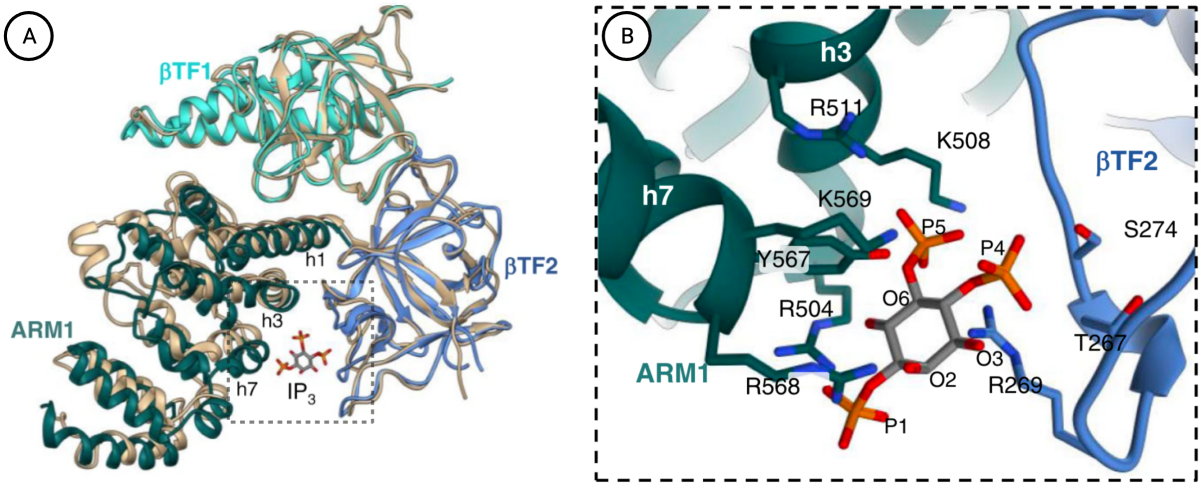

Figure 1.5: Structure of the IP_3-Binding Domain in the presence of an IP_3 molecule. (A) The suppressor domain is shown as β-TF1. The IP_3-Binding Core is shown as its two main components, β-TF2 and ARM1, in its open clam configuration. (B) Detailed view of the IP_3-binding pocket showing residues important for interaction with IP_3. Image adapted from ref. [56].

Most SCA29 patients have been found to have mutations in the IP_3-Binding Domain. Of these, mutations are heavily clustered around a few residues that form the IP_3-binding pocket (e.g. R269W/G, T267M/A, R568G, Y567C, Figs. 1.5B and 1.4A). Systematic validation of SCA29-causing mutations has revealed that most lead to drastically reduced IP_3 binding affinity, thereby abolishing calcium release[57].

At least one pathogenic mutation is known to exist in the suppressor domain (R36C), and it causes a SCA29-like phenotype with pinpoint pupils[58]. This variant was shown to increase IP_3 binding affinity and to affect the dynamics of calcium release[58].

Other mutations, such as N602D described in the Canadian SCA29 family, do not affect IP3 binding, but do impair calcium release despite being distant from the channel domain[57, 59]. It is assumed that it disrupts the signal transduction mechanism between the ligand-binding and channel domains that is required for channel gating[57].

1.3.2.2 Calcium Channel

The channel domain contains six transmembrane domains, and its expression alone is sufficient for IP3R1 to form homotetramers and localise to the ER membrane. While constructs lacking all six transmembrane domains (TMDs) are soluble monomers, those that express at least two assemble into multimers and localise to the ER membrane[60]. The pore of the channel is structurally formed by TMD5 and TMD6, but lined only by TMD6[48]. In lumenal side, between TMD5 and TMD6 lies a critical selectivity filter[61] (lumenal loop in fig. 1.4B), while in the cytosolic side TMD6 continues C-terminally to a linker region (LNK) and to a central coiled-coil[48] (CTD in figs. 1.4A and B). The LNK region is thought to be critical for channel gating[62]. Expression constructs containing the channel domain but lacking the regulatory and IP_3-binding domains produce constitutionally 'leaky' calcium channels in the ER membrane[63]. Leaky channels with calcium permeation and IP3-binding properties similar to wild-type channels are also produced if TMD1-4 are deleted from the full length construct[64].

AD Gillespie mutations described to date have been found to affect only a restricted repertoire of residues found in the lumenal loop (between TMD5 and 6), TMD6 and LNK regions. The two most common mutations have been functionally validated[4, 59], and both show abolished channel function (G2554R and K2611del).

Some patients with a SCA29-like phenotype have also been found to have heterozygous missense mutations in these critical channel regions. Surprisingly, there have been reports of patients with and without aniridia for the same mutation G2554R[5, 33]. For other residues, different mutations lead to different phenotypes. For instance, the glycine residue at position 2666 can lead to Gillespie syndrome when mutated to arginine[65], but to a SCA29-like condition when mutated to glutamic acid[66]. The same can be seen for residue 2591 (SCA29-like with N>T mutation, Gillespie syndrome with N>I) (Dr. Jussi Tolonen, personal communication, and ref. [67]). A *de novo* I2598N mutation has been associated with severe pontocerebellar hypoplasia in one patient[68] while I2598T has been linked to pontocerebellar atrophy in another[69]. Additionally, some patients with phenotypes that include micro or macrocephaly[70] and peripheral neuropathy (in addition to ataxia)[71] have been found to have *de novo* muta-

tions in critical residues in the IP3R1 channel domain.

1.3.2.3 Regulatory Domain

IP_3 receptors are unusual because the ligand-binding site is very distant from the channel. The large region between these two functional domains comprises the bulk of receptor on the cytoplasmatic side and is known as the regulatory domain. Segments of it are known to be important for channel gating in response to IP_3 binding, but the specifics of how this occurs is still a topic of intense debate. Due to its large size, it contains numerous sites for interaction with other proteins.

All but one[7] of the AR Gillespie mutations are predicted to lead to protein truncation in the regulatory domain[4, 6] (Fig. 1.4A).

Some AD Gillespie mutations have also been reported in the regulatory domain. A few patients have been reported with different mutations affecting the residue E2109, which is thought to be a critical calcium sensor for channel activation[72]. The other mutation, described in a single family[73], introduces an intronic splice site which leads to in-frame incorporation of five aminoacids and may also disrupt either channel gating or calcium sensing.

Some SCA29 mutations also affect residues in this region (Fig. 1.4A). The V1562M mutation identified in the original Australian family has been shown to neither prevent IP3 binding nor affect calcium release[57]. It is known however to block a known interactor, CA8 (also known as CARP), from binding to IP3R1. At least one other mutation, S1502D, is also known to disrupt the normal regulation of IP3R1 by CA8 (itself a known ataxia gene[74]), without other effects in IP_3 binding or calcium release[57].

1.4 Molecular Mechanisms of ITPR1-Related Diseases

Lymphocytes and fibroblasts of SCA15 patients express roughly half the levels of ITPR1 mRNA transcripts as unaffected controls[15, 16, 19]. Similarly, IP3R1 protein levels are markedly reduced, but still detectable, in cells of SCA15 patients, but shorter N-truncated proteins are not found[15, 19]. Therefore, ITPR1 haploinsufficiency is the mechanism assumed to

cause SCA15.

In contrast, SCA29 is thought to be caused by dominant-negative mechanisms because many of the missense mutations found in affected patients cause IP3R1 loss of function in functional studies[57]. Most identified mutations significantly impair IP_3 binding (e.g. T267M, R269W), while others were shown to either prevent channel gating (e.g. N602D or T594I) or block IP3R1's regulation by CA8 (e.g. V1562M or S1502D)[57, 59].

Gillespie syndrome is thought to be caused by complete or near complete loss of ITPR1 function. Autosomal recessive mutations lead to premature protein termination, with all reported mutations to date occurring before the channel domain[4]. Autosomal dominant (usually *de novo*) mutations affect exclusively the channel domain and are expected to severely disrupt channel function in a dominant-negative manner[4, 5, 59].

1.4.1 Challenges to Current Understanding

The currently proposed mechanisms for each of the types of mutations present a paradox. If SCA15, assumed to be a neurodegenerative condition with 100% penetrance[21], is caused by haploinsufficiency of ITPR1, then obligatory parental carriers of AR Gillespie loss-of-function mutations would be expected to develop SCA15. There is no indication that this is the case, however. AR Gillespie patients have unremarkable family histories and are born of healthy parents who, despite MRI investigation, are not found to have even subclinical cerebellar atrophy[4].

This unexplained paradox complicates clinical diagnostics. Although not supported by available literature, some adult-onset ataxia patients have been diagnosed with SCA15 because they carried a heterozygous truncating variant in ITPR1 (Prof. Andrea Nemeth, personal communication).

Additionally, it is possible that other neurogenetic syndromes are associated with ITPR1 mutations. Recently, two unrelated patients have been reported with a phenotype that resembled pontocerebellar hypoplasia (PCH), with severe developmental delay, ataxia and hypoplasia of the pons and cerebellum on imaging. Interestingly, both cases involved mutations in the same

residue (I2598N or I2598T)[68, 69].

1.4.2 Currently Available Disease Models

Existing ITPR1 disease models have not been able to significantly improve our understanding of pathogenetic mechanisms.

Animal models of ITPR1, predominantly mice, have not adequately recapitulated the phenotype seen in humans. While knockout of *itpr1* in mice leads to an ataxic phenotype, it is significantly more severe than that seen in Gillespie patients. Only 20% of knockout animals are born, and these develop rapidly progressive ataxia and dystonia which leads to premature lethality by the fourth week of life[75]. No aniridia is seen in these animals. Also in contrast to human SCA15 patients, heterozygous knockout of ITPR1 does not lead to an ataxic phenotype in mice, although mild incoordination can be detected[76]. Crucially, however, no animal models with dominant-negative missense mutations have been reported to date[5].

Current cellular models are often based on transfection of mutant ITPR1 constructs into immortalised cell lines to investigate changes in intracellular calcium in response to IP3R1 stimulation. These experiments generally require that all endogenous IP_3 receptors (ITPR1, ITPR2 and ITPR3) are knocked out for the mutant construct's effect to be detected. Upon transfection, constructs encode self-assembling tetramers which contain four mutant subunits. This approach, while useful, is not a good representation of what probably happens in patients' cells, where both wild-type and mutant subunits are expressed. Co-transfection experiments have been used to address this issue, but dominant-negative effects in channel function have only been detected when a 5 or 6-fold excess of mutant construct was used[4, 57]. In this set up, nearly half of all tetramers are expected to contain only mutant subunits, in contrast to 6.25% if the same amounts of mutant and wild-type subunits were produced.

1.4.3 Stem Cell Disease Models

Better models are required to progress the understanding of pathogenetic mechanism of ITPR1-related disorders. As ITPR1 pathology seems to be largely confined to the central nervous

system, one possible approach would be to derive induced pluripotent stem cells (iPSCs) from Gillespie and/or SCA29 patients and use neuronal differentiation protocols to generate mutant neuronal cultures which could then be studied for disease-relevant phenotypes. This approach, amongst other limitations, requires access to patients carrying particularly informative mutations. Additionally, reprogramming of patient-derived somatic cells is an expensive and lengthy step with variable results that can significantly impact downstream differentiation experiments (i.e. somatic memory)[77].

An alternative is to harness recent developments in genome engineering technology to directly edit the genome of human embryonic stem cells (hESCs) to introduce disease-relevant ITPR1 mutations. This approach is not dependent on access to patients and does not involve cellular reprogramming, therefore established and well characterised hESC lines with proven neuronal differentiation potential can be used.

1.5 Approaches for Editing the Genome of Human Stem Cells

Targeted editing of the genome of stem cells was first accomplished in mouse embryos in 1987[78]. Human embryonic stem cell lines, however, were not isolated until 1998[79] and not transfected until 2001[80]. Targeted editing of hESCs was reported for the first time in 2003[81], but by 2011 there were only 11 other reports of successful modification[82]. Similarly, despite iPSC technology being available since 2006[83], the first engineered iPSC lines were not described until 2011[84]. Therefore, targeted genomic manipulation of human stem cells is a rather recent possibility.

1.5.1 Homologous Recombination

Homologous Recombination (HR) was the method first used to precisely edit the genome of mouse ESCs (mESCs)[85, 86]. It relies on the cell's own enzymatic machinery to recombine foreign DNA with a specific region of its genome[87], as guided by the regions of homology between them. As it was an incredibly inefficient method on its own, it was combined with selection markers so that successfully edited cells could be more easily identified. Despite ini-

tially positive results in human embryonic stem cells[81], few groups were able to successfully replicate these results until other methods became available[88, 89].

1.5.2 Engineered Site-Specific Nucleases

The intrinsically low efficiency of homologous recombination was found to increase dramatically when double-strand breaks were introduced at the target locus[90, 91]. In mESCs, early experiments showed that when DSBs were induced, HR was increased by over 50-fold[92] while in other cell lines it was found to increase by over 1000-fold[91]. To exploit this fact, artificial restriction enzymes such as Zinc-Finger Nucleases[93] (ZFNs, Fig. 1.6A) and Transcription Activator-Like Effector Nucleases[94] (TALENs, Fig. 1.6B) were synthetically engineered to bind and introduce DSBs at specific DNA sequences. By co-transfecting a homologous donor template alongside the nucleases, hESCs and iPSCs were successfully edited using ZFNs for the first time in 2009[89, 95] and using TALENs in 2011[96]. Both ZFNs and TALENs contained the cleavage domain of the FokI endonuclease fused with DNA-binding domains of other proteins[97]. Hence, each time a genomic site is to be targeted, a new protein construct must be designed and tested, requiring significant cloning effort and time.

1.5.3 RNA-Guided Nucleases

Difficulties in the process of engineering and testing site-specific nucleases prevented its widespread adoption as molecular tools. It was therefore revolutionary when two landmark publications reported in 2012 that bacterial nucleases could be directed to introduce DSBs at specific DNA sequences using RNA-based guides[99, 100] (Fig. 1.6C). This system, based on Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) and CRISPR-Associated (Cas) Nucleases, was rapidly repurposed to edit virtually any DNA sequence.

Within a year of publication, genetically altered mice[101], zebrafish[102], hESCs[103] and iPSCs [104] were reported using this technique.

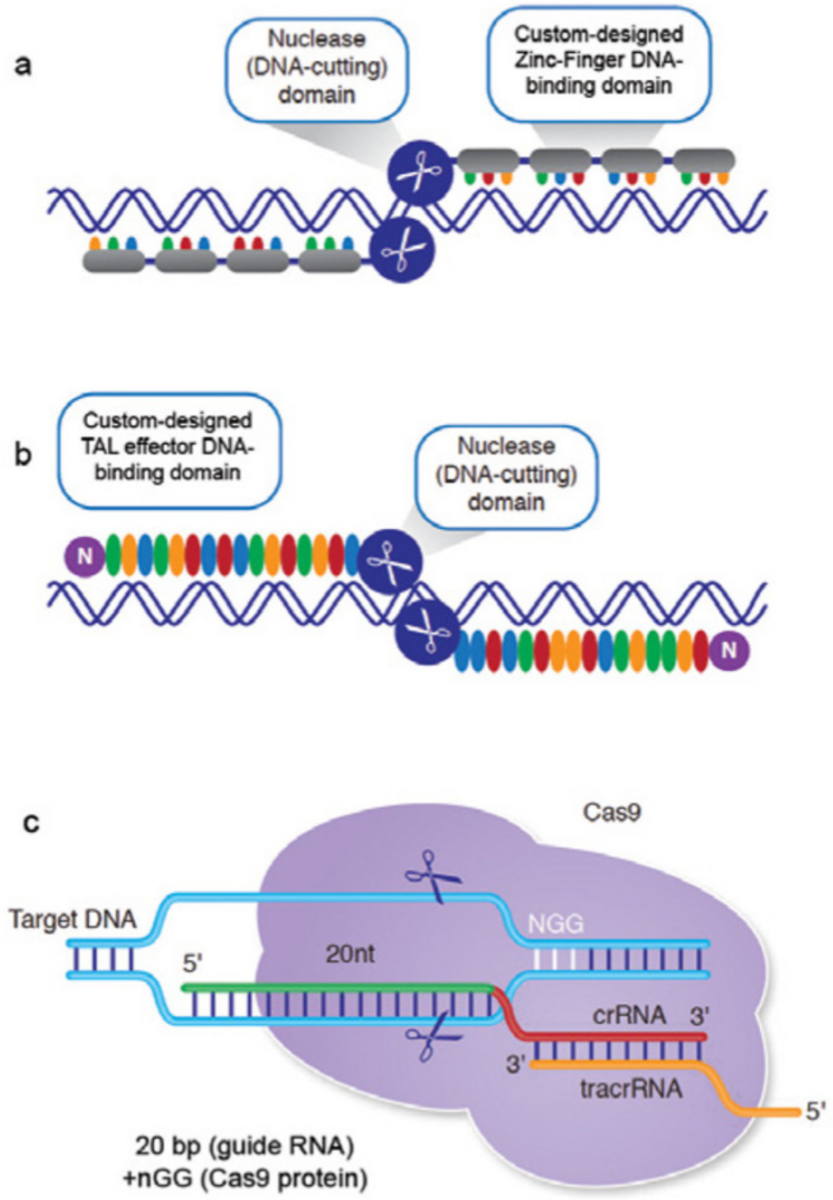

Figure 1.6: Genome editing with nucleases. (A) Zinc-finger nucleases (ZFNs) (B) TALENs and (C) CRISPR/Cas9. Different from ZFNs and TALENs, the CRISPR/Cas9 system uses an invariable nuclease (Cas9) which is guided to a precise genomic location by a short guide RNA (originally a combination of a target-specific crRNA and an activating tracrRNA). Image from ref. [98].

1.6 CRISPR/Cas9 Genome Engineering

Compared to previous methods, the CRISPR/Cas9 system proved incredibly more efficient and practical, and editing of human cell lines has now become commonplace. However, editing human pluripotent or embryonic stem cells is much more challenging, and very low rates of editing were commonly reported prior to the start of this project[105]. Although some publications mentioned editing rates of up to 6% in hESCs or iPSCs[106], these frequencies, calcu-

lated using high-throughput sequencing or mismatch cleavage assays in cell populations, were rarely found by researchers trying to isolate successfully edited stem cell lines. Previous experience in the Nemeth lab trying to edit the SPTBN2 gene in hESCs using CRISPR/Cas9 had achieved only two cell lines with biallelic loss-of-function mutations, and none carrying the desired missense mutation, after manually screening approximately 600 clones. In research fields such as ataxia or neurodevelopmental disorders, in which hundreds of new disease genes (and thousands of new causal mutations) have been discovered only recently, improvements in the approach used to edit human stem cells using CRISPR are necessary to accelerate investigation of mechanisms and potential therapies. Below, I summarise the key features of the CRISPR/Cas9 system relevant to this thesis.

1.6.1 Origins as a Prokaryotic Adaptive Immune System

Prokaryotes cleave the genetic material of invading bacteriophage viruses by expressing small RNAs which hybridise to complimentary sequences in the invading organism. Specialised prokaryotic nucleases detect and cleave regions where hybridisation has occurred[99, 100].

These small guiding RNAs are encoded in the host genome at regular intervals in the spacer sequences found between palindromic repetitive sequences, and are arranged in clusters encoding for several different targets[107].

The entire cluster of repeats and spacers are transcribed into long RNAs (pre-crRNAs) before they are processed into multiple smaller RNAs (crRNAs) each containing a repeat and a spacer[108].

Many conserved genes were regularly found adjacent to CRISPRs in bacterial and archeal genomes[109]. These CRISPR-Associated (Cas) genes were shown to be the effectors of many parts of prokaryotic immune system, but disruption of a single gene (Cas9) could shut down the immunity[110].

Cas9 is a nuclease with two active sites, RuvC- and HNH-like, that are located on different domains (Fig. 1.7A). Each active site acts independently on individual DNA strands to generate a double-stranded break (DSB) in the protospacer sequence [99]. However, complementarity

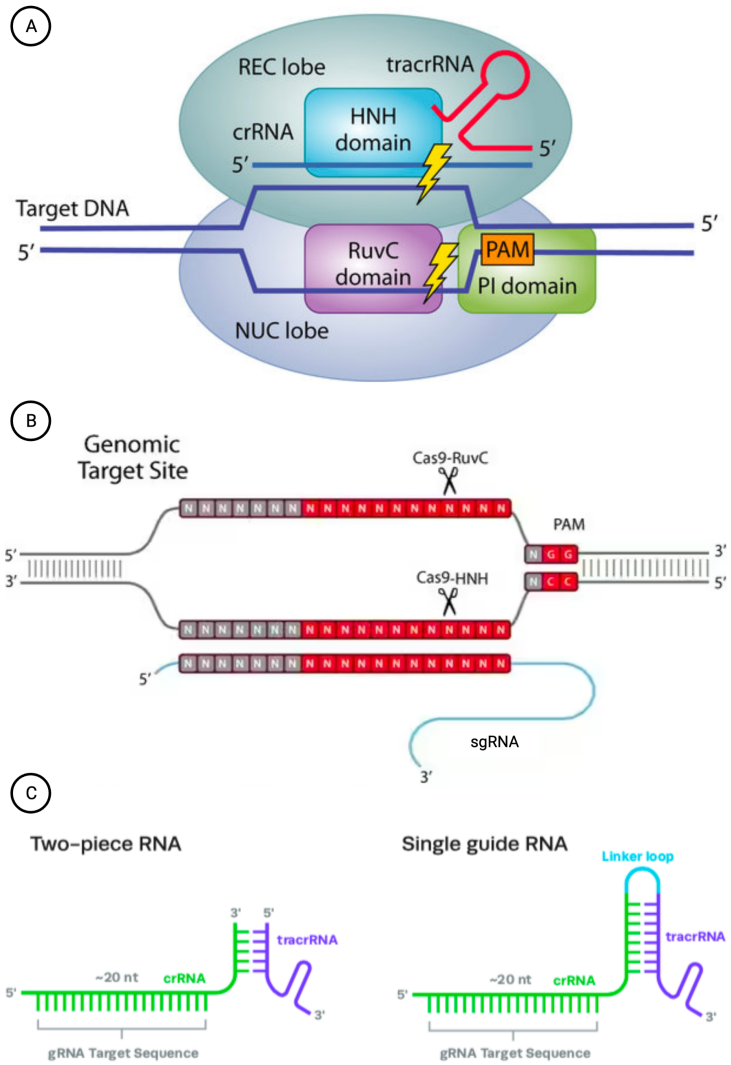

Figure 1.7: CRISPR/Cas9 System. (A) The Cas9 nuclease has separate domains, RuvC and HNH, to introduce double-strand breaks in each of the DNA strands. (B) For Cas9-mediated cleavage to occur, two conditions must be met. First, there must be complementarity between the sequence in the guide RNA and the target strand. Second, a protospacer-adjacent motif (PAM) must be present in the non-complimentary strand. (C) In prokaryotes the guiding RNA system is made of two separately transcribed RNA sequences. The crRNA contains a RNA sequence complimentary to the target strand, whereas the tracrRNA contains a sequence that pairs with the crRNA, forming a secondary RNA structure which is then recognised by the Cas9. In the adapted eukaryotic genome engineering system, the crRNA and tracrRNA have been fused so they can be transcribed as a single guide RNA (sgRNA).

between spacer-encoded crRNA (host genome) and protospacer (foreign genome) sequences is not sufficient for cleavage to occur, otherwise host auto-immunity would ensue. An additional requirement for cleavage is the presence of a conserved motif located immediately downstream of the protospacer sequence (Fig. 1.7A and 1.7B). If this Protospacer-Adjacent Motif (PAM), which is present only in the foreign sequence, is mutated or not present, cleavage does not occur[110].

1.6.2 Early Adaptation to a Genome Engineering System

Although CRISPR/Cas systems were found in nearly half of all known bacteria and most archeae[107], different variations existed. Earlier work that paved the way for the prokaryotic immune system to be adapted for use as a genomic editing system in eukaryotic cells used proteins and RNA sequences from the type II system found in *Streptococcus pyogenes*[99].

Its Cas9, referred to as SpCas9, requires a 20bp protospacer (i.e. the target sequence) followed by a PAM in the form of 5'-NGG-3', which is one of the least restrictive motifs amongst all known Cas nucleases[111]. This effectively means that, for cleavage to occur, the 20bp target sequence must be immediately followed by any nucleotide and two guanines. Cleavage by SpCas9 occurs predictably and reliably, on both strands, at the protospacer usually between the 3rd and 4th nucleotides 5' of the PAM[99, 100] (Fig. 1.7B).

Also particular to the type II system, processing of pre-crRNAs and Cas9-mediated cleavage requires presence of a trans-activating crRNA (tracrRNA), which contains a sequence that basepairs perfectly with the repeat found in crRNAs[112] (Fig. 1.7A and 1.7C). The tracrRNA serves as a scaffold to which the SpCas9 binds in association with the crRNA. However, it was found that by cleverly juxtaposing the sequences encoding the tracrRNA and crRNA, a fully functional single guide RNA (sgRNA) could be constructed[99, 104] (Fig. 1.7C).

Thus, a two-component (sgRNA and SpCa9) system can be used to introduce DSBs at precise locations with the only requirement that 20 nucleotides of the sgRNA be complementary to a target sequence immediately preceding a NGG motif. Following these rules, and with minor adaptations to allow efficient expression of sgRNA and SpCas9 in eukaryotic cells, human cell lines can be edited[104].

1.6.3 Relevant Cellular Mechanisms of DNA Repair

Successful genome editing using the CRISPR/Cas9 system is dependent on two separate processes. The first is generation of a DSB at the target locus, and it is entirely achieved by foreign machinery delivered to the cell's nucleus. The second is an imperfect correction of this DSB

carried out by the cell's own DNA repair machinery. While the first process is dictated by the user, the second is entirely controlled by the cell. By understanding how mammalian cells naturally repair DSBs, it is possible to bias the outcomes towards an expected result.

Once a DSB is detected, the cell has two primary strategies for correcting it: re-ligating the ends directly, known as Non-Homologous End Joining (NHEJ), or repairing the gap with the help of a template, known as Homology-Directed Repair (Fig. 1.8).

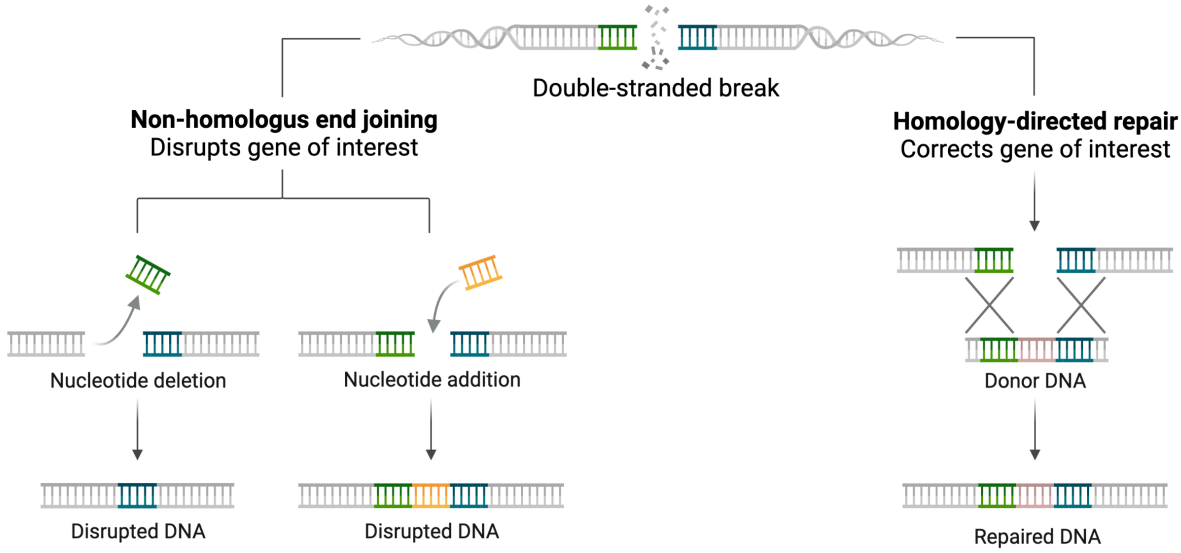

Figure 1.8: Basic mechanisms of DNA repair following a double-stranded break. Non-homologous end joining (NHEJ) is the mechanism by which direct re-ligation of DNA ends is attempted. It is considered an error-prone process as often small insertions or deletions (indels) are generated at the re-ligation site. Homology-directed repair (HDR) is a mechanism considered error-free as it uses information from a homologous template (e.g. sister chromatid) to repair the DSB with fidelity.

The process commonly described as NHEJ is now understood to include at least two different enzymatic pathways, Canonical Non-Homologous End Joining (c-NHEJ) and Microhomology-Mediated End Joining (MMEJ, also called Alternative End-Joining, A-EJ or alt-NHEJ)[113]. Notwithstanding the exact biochemical mechanisms involved, the outcomes are similar: either the DSB is perfectly repaired back to its original sequence (by c-NHEJ) or the gap is repaired with small insertions or deletions (indels, a consequence of MMEJ or c-NHEJ). The pattern of indels generated at a given locus are not random[114], and in some cases, dependent on knowledge of the local sequence context, it is possible to predict the types of mutations that will be generated[115]. However, despite recent efforts to try to harness MMEJ to achieve precise editing, template-less CRISPR/Cas9 editing is generally used with the expectation that

indels that disrupt the reading frame of coding genes can be generated.

Precise corrections are most commonly associated with the Homology-Directed Repair (HDR) pathway, which is the basis of original homologous recombination techniques. A DNA template containing regions of homology to one or both ends is required, as HDR is the high-fidelity process by which the cell uses the sister chromatid as a template to repair a DSB that could not be directly re-ligated[113]. The final result is a repaired double-stranded DNA in which a segment has been copied from a homologous region found in the template.

1.6.4 Delivery Strategies

Components of the CRISPR/Cas9 system must reach the nucleus of the cell to edit its genome. As mentioned, this adapted system requires a Cas9 nuclease and a single guide RNA (sgRNA). If precise editing is attempted, a third component required is the HDR template. Therefore for all CRISPR/Cas9 editing experiments, two or three components must reach the nucleus.

The Cas9 nuclease can be delivered as a DNA construct (e.g. as a circular dsDNA plasmid)[104], as mRNA[116] or as an assembled protein[117]. Similarly, the sgRNA can be delivered either as DNA (to be transcribed by the cell's own machinery) or as pre-transcribed (or synthetic) RNA. Finally, the HDR template is delivered as DNA, although it may be single or double-stranded, linear (e.g. PCR product or single-stranded oligonucleotide) or circular (e.g. plasmid) .

How individual components are delivered to cells is important not only because of the different techniques required, but also because it directly impacts two key metrics: efficiency and specificity.

1.6.5 Efficiency

Despite best efforts, not all cells are successfully edited in CRISPR/Cas9 experiments. This becomes extremely important when considering how to develop analysis strategies at scale. If efficiency is very low, isolating mutant clones for downstream analyses becomes very time-consuming and expensive, limiting the number of successful clones that can be generated.

Efficiency is often understood as the percentage of successfully edited cells. But as target cell lines are diploid or polyploid, the number of targets is generally greater than the number of cells. For practical purposes, the rate of successful editing is thus calculated based on the total number of available targets.

1.6.5.1 Assessment of Indel Formation

As generation of Cas9-induced DSBs cannot (under normal circumstances) be directly measured, a proxy measurement is taken by analysing the frequency of alternative sequences at the target locus. This is a combined measurement reflective of its individual components, namely Cas9 cleavage efficiency, sgRNA hybridisation efficiency, target susceptibility and DNA repair (in)efficiency.

Comparative assessment of the efficiency of two different sgRNAs, for example, requires side-by-side experiments to be undertaken where each guide is co-transfected with Cas9. The most efficient sgRNA will be the one that generates the most mutations (indels) at the target locus. In such comparisons, the indel frequency is often referred to as 'cutting efficiency' or cut rate, as what is truly being estimated is the ability of a given sgRNA to induce Cas9-mediated cutting at a specific site.

Indel formation can be quantified from populations of cells in different ways. The traditional method is to use mismatch cleavage assays, such as SURVEYOR or T7 Endonuclease assay[105, 118]. In these, PCR amplicons containing the target region are denatured and allowed to slowly reanneal, promoting the formation of heteroduplexes between strands of slightly different lengths differing only by the presence of small internal insertions and deletions (indels). Enzymatic digestion with structure-selective endonucleases cleaves heteroduplexes only, leaving homoduplexes intact. By running the products on a gel, it is possible to quantify the proportion of cut vs uncut amplicons, thereby estimating the proportion of indels at a given locus.

Another common method uses Sanger sequencing to estimate indel frequency. PCR products across the target region are amplified from both edited and unedited (control) cell populations. As indels disrupt the sequencing frame and generate secondary peaks in the chromatograms,

the ratio between primary and secondary peaks (and associated base qualities) can be used to quantify indel frequency[119]. Initially results were roughly estimated from visual inspection of chromatograms (e.g. no frameshifts indicated negligible frequency of indels, whereas complete disruption of the reading frame starting at or near the expected cut site indicated very high frequencies), but statistical software such as TIDE and ICE were soon developed to generate more accurate estimates with sensitivity to detect frequencies down to 1-2%[120, 121].

Amplicon-based next-generation sequencing (NGS) can also be used. Not only does it provide a very accurate estimation of indel frequency, it also sequences and quantifies the proportion of each individual read, mutated or not, represented in the amplicon. Though expensive, it is considered the gold standard for analysing indel formation.

An important limitation of all methods discussed is their reliance on short-range PCR amplification across the expected cut site, which biases results towards smaller insertions and deletions that allow for an amplicon to be produced[122]. Indeed, recent studies using long-range PCRs and long-read NGS have identified surprisingly high rates (>20% in one study[123]) of large deletions that would have been missed using the methods that rely on short-range PCR[122–127].

1.6.5.2 Factors Affecting CRISPR/Cas9 Efficiency

The rate at which SpCas9 induces the production of indels at a target locus is influenced by a large number of factors. Many are external to the CRISPR/Cas system per se, but nonetheless affect the experimental outcomes. For instance, transfection efficiency is often overlooked or assumed to be 100% when results are compared between studies, despite the fact that editing outcomes are only possible in cells that were successfully transfected. Cell type also immensely influences editing results. Not only because different cells command different transfection methods and efficiencies, but also because DNA repair mechanisms may be more or less active. Immortalised cell lines, such as HEK293 or those derived from tumours (e.g. HeLa cells) are very amenable to CRISPR/Cas9 editing, whereas human embryonic stem cells or iPSCs are considered very resistant. Targeting the same locus in different cell lines with the same SpCas9 and sgRNA, Mali et al. achieved 25% indel frequency in HEK293T cells but only 4% in a iPSC

line[104]. In contrast, mouse ES cells are easily editable, in keeping with their susceptibility to traditional homologous recombination techniques. Chromatin accessibility is also a factor, as more efficient editing is achieved in euchromatin targets compared to ones found in regions of heterochromatin[128]. DNA repair pathways after DSB are also influenced by chromatin context[129].

Cas9-related factors such as version, concentration, duration and delivery impact its efficiency. Changes in the Cas9 structure not surprisingly can alter its function and therefore its efficiency. Several different versions of SpCas9 have been produced. For instance, mutations introduced in its HNH- or RuvC-like domains can give rise to 'nickases' (SpCas9n) that generate single-stranded nicks[99, 103]. Other versions such as eSpCas9(1.1), SpCas9-HF1 and HypaCas9 have been engineered to be less tolerant of mismatched protospacer sequences, but high stringency often comes at a cost of reduced efficiency at some targets[130–132]. Concentration and duration are two factors that are positively correlated with efficiency: more Cas9 molecules per cell or prolonged persistence leads to higher editing frequencies[106, 117, 133]. These two factors largely explain why different delivery methods lead to different efficiencies. Plasmid-based methods of delivery lead to high concentrations, but often take a few days to reach maximal expression, whereas Cas9 protein is active immediately after transfection but degrades rapidly[117].

Finally, choices pertaining to sgRNAs exert massive influence on cutting efficiency. Similar to Cas9, concentration and duration are important. Higher levels of sgRNAs in general lead to higher cutting efficiency up to a plateau [133]. Similarly, expression and/or stability of sgRNAs in cells improves editing, as RNAs are inherently unstable molecules that are rapidly degraded by nucleases. Chemical modification of terminal nucleotides in sgRNAs can reduce their degradation, thereby prolonging persistence in cells and increasing cutting efficiency[133]. Mismatches between the guide and the target can dramatically reduce its efficiency, but some mismatches are more tolerated than others. Single mismatches in the PAM site can nearly abolish Cas9 activity[106], whereas multiple mismatches in the most distal protospacer positions from the PAM can be partially tolerated[106]. Indeed, for many targets even shorter 18bp sgRNAs can have similar indel efficiency as standard 20bp ones[134]. As a gen-

eral rule, the further away a mismatch is to the PAM, the more likely it is to be tolerated[135]. Other sequence-related factors have been noticed to affect the function of sgRNA, including GC content, repeated nucleotides and position-specific nucleotides[136]. Thus, bioinformatic tools and predictive scores have been developed to aid the design of sgRNAs against human genomic targets[136] (e.g. Doench score[137]). Yet, *in silico* tools are not perfect and often sgRNAs with high predictive scores fail for no apparent reason. Thus it is commonly suggested to design and experimentally test between 3-5 guides for each locus to be targeted.

1.6.6 Specificity

Although the CRISPR/Cas9 system can be used for precise genome editing, it often introduces DSBs at locations away from the intended target, potentially introducing unwanted genomic variation[138]. These off-target effects have various contributing factors, and extensive work has been dedicated to understand and minimise their occurrence.

Cas9's partial tolerance to mismatched nucleotides between sgRNA and target sequence means that highly homologous regions elsewhere in the genome might also be edited[106, 124, 130]. Most of these can be avoided or predicted when designing sgRNAs by using bioinformatic tools that identify potential off-target sites and their consequences (e.g. intergenic, intronic, exonic). However, genomic variation between individuals or cell lines can alter the landscape of potential off-target sites[139].

Several Cas9 versions have been developed as an attempt to reduce unwarranted cutting. Nickases (SpCas9n), which introduce only single-stranded nicks, can be used with a pair of closely spaced sgRNAs to create a staggered DSB with a >50-fold lower rate of off-target effects[103]. In a different strategy, high-fidelity version such as the enhanced SpCas9 (eSpCas9, fig. 1.9) have been rationally engineered for maximum stringency with sgRNA-protospacer hybridisation[130]. For most targets and sgRNAs they can drastically reduce off-target cutting while maintaing similar or only slightly reduced on-target efficiency[132].

Finally, several of the factors that are known to increase on-target efficiency will also increase off-target cutting, such as concentration or duration of Cas9 and sgRNAs in the cell[106].

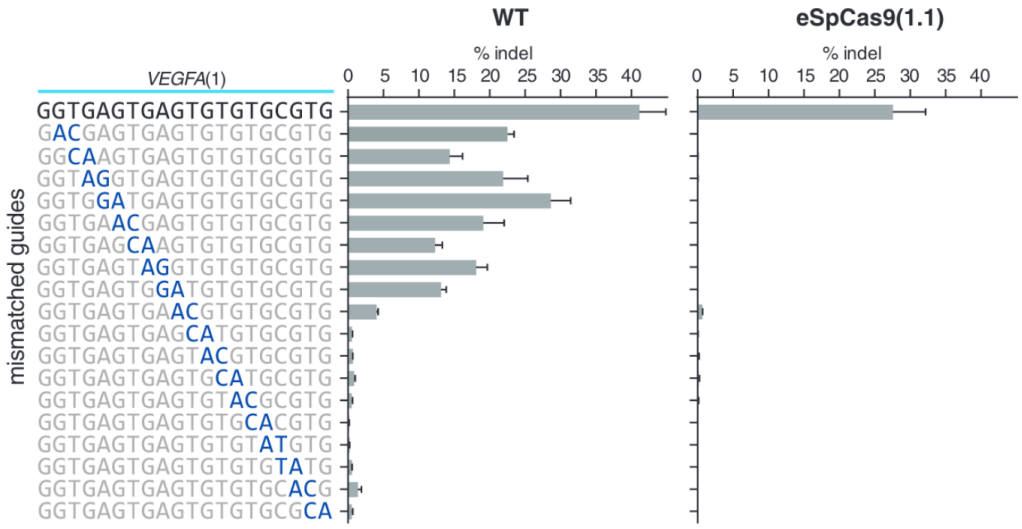

Figure 1.9: Comparative on- and off-target efficiency of wild-type SpCas9 and enhanced SpCas9. Some Cas9 versions can have markedly reduced rates of off-target effects. The rates of indel formation at the target VEGFA locus is shown, for both wild-type (WT) and enhanced SpCas9 (eSpCas9) using the fully matched sgRNA (top row) and also with 18 other guide sequences containing two nucleotide mismatches. Image from [130].

1.6.7 Harnessing HDR for Precise Genome Editing using CRISPR/Cas9

For the purposes of precise editing of the mammalian genome, the homology-directed repair (HDR) pathway can be exploited to repair the DSB using a foreign template. CRISPR/Cas9-mediated HDR uses the exact same principle behind the technique of homologous recombination (discussed in section 1.5.1), which has been used since the 1980s for generation of transgenic animals[78]. The difference is that a DSB is first introduced at the target region in order to trigger the DNA repair machinery, thereby massively increasing rates of successful homologous recombination using the foreign template. The sequence to be edited (e.g. added, removed, or changed) is placed between two longer arms of sequences that bear homology to the ends of the DSB (Fig. 1.8). Foreign HDR templates can be delivered as either single-stranded or double-stranded DNA, and dsDNA template can be either circular (e.g. a plasmid) or linear (e.g. a PCR product, a synthetic gene block or a linearised plasmid).

Several approaches have been developed, each with different practicalities, efficiencies and problems. Very few studies however have tried to compare different approaches while accounting for the multitude of variables that are known to influence final outcomes.

If the desired result is to introduce large genomic changes, such as the insertion of transgenes

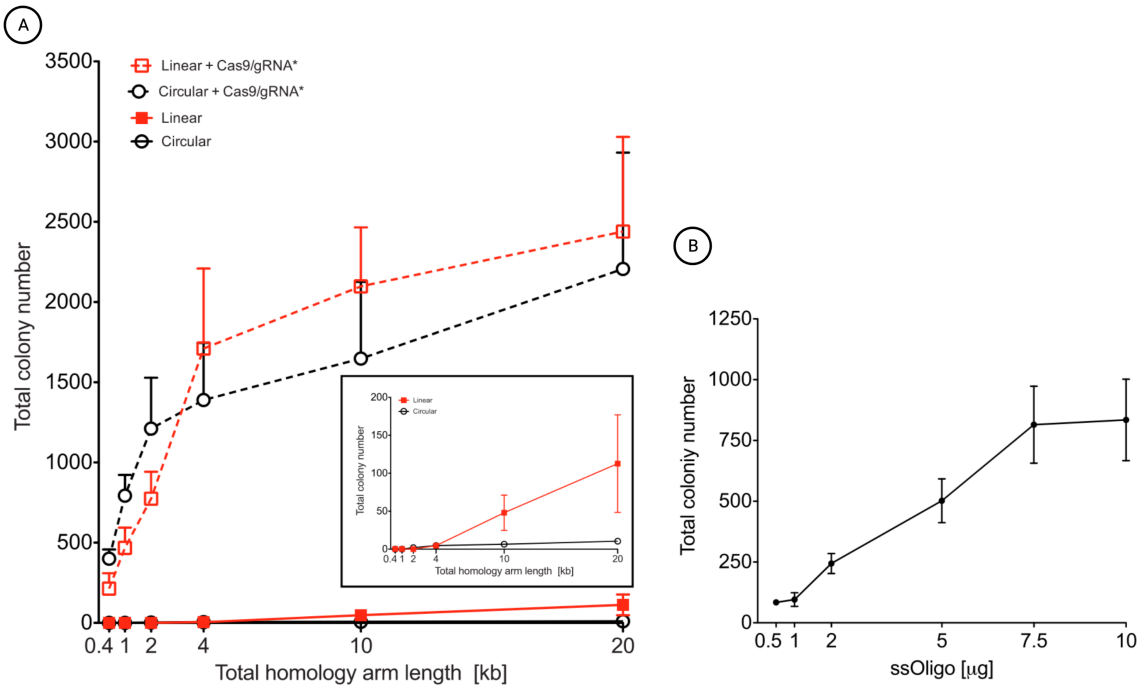

Figure 1.10: Donor template factors influencing HDR rates. (A) Effect of double-stranded homology arm length on the total number of colonies expressing a 4bp-corrected selective marker. Both linearised and circular constructs are shown. The inset magnifies the rate of homologous recombination in the absence of a DSB breaks (i.e. without a transfected Cas9/sgRNA). (B) Effect of total transfected mass of a 100nt single-stranded DNA donor on HDR rates. Adapted from [62].

(e.g. GFP or antibiotic resistance cassettes), dsDNA HDR templates are generally used. As a general rule, efficiency of integration drops as insert length increases[140, 141], with longer regions of homology flanking the insert (homology arms) improving integration rates[62] (Fig. 1.10A). Prior to the development of ZFNs, TALENs and CRISPR/Cas systems, homologous recombination experiments used linear dsDNA templates with very long homology arms in order to overcome the inherently low efficiency of spontaneous recombination, but with Cas9-assisted homologous recombination much shorter regions of homology can be used. For example, while the first successful generation of genomically engineered hESC lines used 12kB of total homology[81], recent studies in iPSCs have achieved similar or better results with less than 2kb[142]. CRISPR/Cas9 studies in mammalian cells (including hESCs and iPSCs) have consistently shown that integration efficiency rises with length of homology arms, but efficiency gains are marginal past the 2kb mark (i.e. roughly 1kb for each arm)[62, 141–143] (Fig. 1.10A). Templates can be delivered as linear dsDNA, such as a PCR product or a linearised plasmid, or can be delivered as part of circular dsDNA plasmid.

For smaller mutations, single-stranded oligonucleotides (ssODNs) are generally preferred. Commercially available and relatively inexpensive, ssODN do not require any preparatory steps normally associated with dsDNA donors such as cloning or PCRs. There has been much controversy in the literature regarding optimal design of ssODNs. Recent data suggests that, similar to dsDNA templates, efficiency reduces as insert length increases, but up to 1kb inserts can still be reliably inserted[140]. Interestingly, for ssODNs it is likely that longer homology arms are not necessarily better, and might even be detrimental[135, 140]. Maximum efficiency can be achieved with homology arms as short as 35bp[140, 144]. Distance between mutation and DSB is critical, as efficiency of editing drops rapidly as distance increases. Rates of successful editing are very low beyond 30bp from the DSB[135, 140, 144].

As previously mentioned, direct comparisons between single-stranded and double-stranded donor templates are difficult due to the large number of variables that influence the final outcome, including biases toward different DNA repair pathways in different cell lines and organisms. Results between the best and most carefully conducted studies are often contradictory. For example, while a study in HEK293T and mouse ESCs concluded that ssODNs were superior to dsDNA for introduction of a a 700bp transgene[140], another has found that circular plasmids with long homology arms are superior to ssODNs for introduction of a 4bp mutation[62] (compare Figs. 1.10 A and B).

Regardless of the approach taken, HDR is generally inefficient, as even when a foreign template is provided most DSBs are still repaired by c-NHEJ or MMEJ rather than HDR[105].

1.6.8 Isolation of Successfully Edited Clones

CRISPR/Cas9 experiments are performed on populations of cells, but not all will be successfully edited. Unless the successful edit leads to an immediate phenotype that can be used for cell selection (i.e. change in expression of a fluorescent marker or surface protein), isolation of single-cell-derived clones is required to derive cell lines with homogeneous genotypes. Invariably, however, a number of single-cell clones will have to be screened before a successfully edited cell is found. The process of establishing clonal populations can be achieved in different

ways.

Cell sorting uses specialised flow cytometry equipment to seed individual cells into separate wells of 96-well culture plates. Plates are cultured to allow for cell proliferation, giving rise to clonal populations. Cell sorting hardware is not found in most labs as it is expensive and requires significant training for its correct use. Commonly cell sorting is offered as a commercial service to researchers, which can be expensive and requires booking time slots in advance. Some cell types tolerate the sorting procedure better than others, but generally hESCs and iPSCs suffer from viability issues after cell sorting[145].

Limiting dilution is a simple technique that can be done in any lab. Cultures are dissociated to single-cells using standard reagents and diluted down to very low concentrations. Culture plates are seeded so that on average fewer than one cell is found per well. The downside of this method is that the number of clones is far less than the number of wells seeded, and most wells seeded will not actually have any cells. As identification of single-cells in 96-well plates under the microscope is challenging, time-consuming and error-prone, all seeded wells must be maintained for at least a few days. Therefore, if the goal is to screen between 90-100 clones, more than 2 full 96-well plates must be seeded and maintained. Logistically, this method entails significant costs with culture medium, plastics and hands-on labour.

Another common method for clonal isolation is colony picking. Dissociated single cells are seeded at very low density onto culture dishes and allowed to grow until large spatially-isolated colonies can be manually transferred to individual wells of a 96-well plate. The procedure traditionally requires an inverted microscope that can be placed, after disinfection, in the laminar flow hood. If such a microscope is not available, the procedure can still be if colonies are allowed to grow large enough to be seen without a microscope. The technique for colony picking is generally easy and requires minimal training if done with a microscope, but can be challenging without one.

1.6.8.1 Screening Clones for Successful Edits

Although preliminary phenotypic screening can enrich for successfully-edited cells, regardless of the method used for clonal isolation, clones eventually will have to be screened for the presence of desired edits. This entails extraction of genomic DNA and use of techniques such as PCR genotyping, restriction fragment length polymorphism (RFLP) and/or sequencing. Depending on the number of clones that are expected to be screened in order to isolate one or more edited cloned, the screening procedure can be costly, repetitive and frustrating.

Considering the large number of variables involved with CRISPR/Cas9 experiments, it is generally impossible to predict how effective editing is likely to be for any given target. Human stem cell genome engineering projects normally grow and screen a few hundred clones in the hope that at least one correctly-edited clone will be identified.

1.7 Development of Stem Cell Models of ITPR1 Mutations

At the time this project started, the general methodologies for editing human stem cells were not extensively developed. Although the discovery of the CRISPR/Cas9 system in 2012[99] has revolutionised genome editing, the process for human stem cells remained laborious, expensive and time consuming. Previous experience in our lab, for example, required screening of nearly 600 hESC colonies to obtain two biallelic knockout clones, a process that took almost a year.

The overall aim of this thesis was to develop a robust and efficient system that would facilitate the introduction of mutations into human stem cells while reducing the uncertainties, costs and hands-on labour required. Ideally, such a system would be easily adaptable to other genes and types of mutations while allowing for cellular differentiation assays.

A secondary aim was to generate stem cell models of ITPR1-related diseases, such as Gillespie syndrome, SCA29 and SCA15. Given the different types of mutations that cause these three conditions, a faster, cheaper, and highly efficient method to produce cellular models was required to investigate their mechanisms.

The third aim was to understand whether complete genetic knockout of ITPR1 impacted the

early stages of neuronal development.

Although the field has moved rapidly in recent years[146], the methodology developed in the course of this work is an important addition for those interested in generating stem cell models of human diseases. If successful in introducing ITPR1 mutations into hESCs, it should also be a valuable tool for other genes of interest.

2 | Materials and Methods

2.1 Molecular Methods

2.1.1 Polymerase Chain Reaction

Two main PCR kits were used, depending on the application.

For long amplicons (>1kB) requiring sequence fidelity (i.e. amplicons were to be cloned into vectors or used in other downstream experiments), NEB Q5 High Fidelity Polymerase Kit was used (unless stated otherwise). Reactions were set up on ice. Reactions were assembled into 1X Q5 Reaction Buffer containing 200μM dNTPs, 0.5μM forward primer, 0.5μM reverse primer, 0.02U/μL Q5 High-Fidelity DNA Polymerase and variable amounts of template. Samples were initially denatured at 98°C for 30 seconds, and then underwent 35 cycles of denaturation at 98°C for 10 seconds, annealing at 60°C (unless stated otherwise) for 10 seconds, and extension at 72°C for 30 seconds per kilobase of amplicon product, followed by a final extension at 72°C for 2 minutes. Samples were cooled down to 10°C and stored at -20°C or submitted to gel electrophoresis.

For short amplicons, Bioline MyTaq High-Sensitivity Red Mix was used unless stated otherwise. Reactions were set at room temperature, and contained 1X MyTaq HS Red Mix, 0.4μM forward primer, 0.4μM reverse primer, and variable amounts of template. Samples were initially denatured at 95°C for 1 minute, and then underwent 35 cycles of denaturation at 95°C for 15 seconds, annealing at 60°C (unless stated otherwise) for 15 seconds, and extension at 72°C for 15 seconds. A final extension at 72°C for 1 minute was applied, before samples were cooled down to 10°C before storage.

For human genomic templates, 25-100ng gDNA were used. For plasmid templates, 10-100pg were used. All primers are listed in the Appendix. Unless stated otherwise, all PCR experiments contained a no-template control reaction, where nuclease-free water was used instead of template.

2.1.2 Agarose Gel Electrophoresis

Agarose gels at 1% concentration were prepared by dissolving 1.5g of low-melting point Agarose (Geneflow, UK) in 150mL of 1x Tris-Acetate-EDTA (TAE) buffer (40mM Tris, 20mM acetic acid, 1mM EDTA) by heating the mixture in the microwave at full power for 3 minutes. Once dissolved, 7.5μL of 10mg/mL Ethidium Bromide solution (Sigma-Aldrich, UK) were added and the gel was cast on a plastic tray and allowed to cool down to room temperature in a fume hood.

Gels were immersed in a tank filled with 1X TAE buffer. 10μL of PCR product mixed with 2μL of 6X Loading Dye (Thermo Scientific, UK) were loaded to each well. Unless stated otherwise, at least one well in the same row was loaded with 5μL of GeneRuler DNA Ladder Mix (Thermo Scientific, UK).

Gels were run at 100V for 50 minutes, followed by visualisation and imaging in a Syngene U Genius Imaging System.

2.1.3 Gel Extraction

UltraSlim Blue LED transillumination was used to assess adequate band separation in gels. UV-based transillumination was not used to avoid DNA damage, and for that reason no recorded images of gels before band excision exist. When separation was deemed appropriate for safe band excision without contamination from nearby lanes or bands, desired bands were excised using a sterile disposable scalpel and carefully transferred to a 1.7mL Eppendorf tube. Unless specified otherwise, QIAquick Gel Extraction Kit (Qiagen) was used following the manufacturer recommended protocol.

2.1.4 DNA Precipitation

1/10 volume of 3M Sodium Acetate pH 5.0 and 2.5 volumes of -80C absolute ethanol were added to the sample. The reaction was vortexed for 10 seconds and incubated at -80°C for at least one hour. Precipitated DNA was recovered by >16,000g centrifugation at 4°C for 20 mins.

The supernatant was carefully discarded. The pellet was washed twice with room-temperature 70% ethanol and allowed to air dry for 15 minutes. The dried DNA pellet was resuspended in the desired volume of buffer or water, followed by measurement of DNA concentration using the NanoDrop 2000 Spectrophotometer (Thermo Scientific, UK).

2.1.5 PCR Purification

PCR reactions were purified using two different methods, depending on the planned downstream applications.

If Sanger sequencing was the only application planned, the ExoSap cleanup method was used. An ExoSap Mix was prepared by freshly mixing 1μL of ExoI nuclease (NEB), 2μL of rSAP (NEB) and 37μL of nuclease-free water. To 10μL of each PCR reaction to be purified, 5μL of ExoSap mix was added. Reactions were incubated at 37°C for 30 minutes, followed by heat-inactivation at 85°C for 15 minutes. This process ensures that that reaction primers are digested and dephosphorylated (by the exonuclease and rSAP, respectively) and cannot interfere with the downstream dideoxy sequencing reaction. Purified samples were diluted 1:10 in nuclease-free water before being sent for sequencing.

For all other applications, PCR reactions were purified using silica columns. Unless specified otherwise, QIAquick PCR Purification Kit (Qiagen) was used following the recommended manufacturer protocol was followed. DNA concentration was measured using the NanoDrop 2000 Spectrophotometer (Thermo Scientific, UK).

2.1.6 Annealing and Phosphorylation of Oligonucleotides

Partially complementary single-stranded oligodeoxynucleotides (ssODNs) were ordered from IDT. 100nmol of each ssODN were mixed in 10μL 1X DNA Ligase Buffer supplemented with 5 units of T4 PNK (NEB). The reaction was incubated at 37°C for 30 mins for optimal phosphorylation, followed by denaturation at 95°C for 5 mins. Annealing was achieved by gradually lowering the temperature by 0.7°C every 12 seconds for 99 cycles using a benchtop thermocycling machine.

2.1.7 Restriction Enzyme Digestion

Standard digestions were carried out by maintaining a ratio of 1μg of plasmid to 1-2 U of enzyme in a 10μL final volume of 1X recommended buffer. If a plasmid was simultaneously digested by two different enzymes (double digestion), the final reaction volume was doubled so that the final glycerol concentration never exceeded 5-10%. Digestions were carried out at 37°C for 1hr, unless specified otherwise. All restriction enzymes used were from New England Biolabs (NEB, UK).

2.1.8 DNA Ligation

DNA fragments to be ligated were mixed at a insert-to-backbone molar ratio of 3:1 up to 8:1 in 20μL of 1X T4 DNA Ligase Buffer (NEB). One microliter of T4 DNA Ligase (10,000 Weiss units; NEB) was added and the reaction was incubated at room temperature for 10 minutes, followed by heat-inactivation at 65°C for 10 minutes, unless stated otherwise.

2.2 Use of Plasmid Vectors

2.2.1 General Plasmid Handling

Plasmids received from Addgene arrived in stab agar culture bottles. A sterile plastic tip was used to pierce the bacteria-containing agar, and then spread carefully on fresh LB-agar plates with ampicillin. Plates were grown at 37°C for 16-20hrs. A single antibiotic-resistant colony was picked with a fresh sterile tip, inoculated into 5mL of Ampicillin-LB broth and incubated at 37°C for 16-18hrs with shaking at 250rpm.

Glycerol stocks for long-term storage were prepared by mixing 0.5mL of liquid culture with 0.5mL of autoclaved 50% glycerol, transferring to a cryovial, and storing at -80°C. Remaining liquid cultures were pelleted by centrifugation at 3,000g for 15 minutes, and supernatant was decanted. Pelleted samples were processed using QIAprep Miniprep Kit (Qiagen) following the manufacturer recommended protocol. Purified plasmids were eluted in 50μL of 10mM Tris-Cl pH 8.5 (Elution Buffer), concentration was assessed using NanoDrop, and stored at

-20°C.

2.2.2 Preparation of Bacterial Culture Agar Plates

For general bacterial work, LB-agar plates containing 100μg/mL ampicillin were prepared by dissolving 6g Agar (Invitrogen cat.30391023) and 10g LB (Invitrogen cat.12795084) into 400mL deionised water. The mixture was autoclaved, slowly cooled down to 50°C in a water bath, and 0.4mL Ampicillin 100mg/mL (Sigma cat.A9518) was added. Between 20-25mL were poured into sterile Petri dishes and allowed to set at room temperature. Covered plates were stored at 4°C for a maximum of 8 weeks. Plates were re-warmed to 37°C for 30 minutes before being used.

If blue/white colony screening was planned, IPTG/X-Gal LB-agar plates containing 100μM IPTG, 40μg/mL X-Gal and 100μg/mL ampicillin were prepared as described above, but once the autoclaved mixtured had cooled down to 50°C in a water bath, and 0.4mL Ampicillin 100mg/mL (Sigma), 0.4mL 0.1M IPTG (Bioline), and 0.8mL 20mg/mL X-Gal (Bioline) were added.

2.2.3 Transformation of Plasmids into Competent Bacteria

For each transformation reaction, 25μL of NEB10-beta competent bacteria were thawed on ice for 10 minutes. 1ng of plasmid or 2.5μL of ligation reactions added to the bacteria and incubated on ice for 30 minutes. After heat-shocking at 42°C in a water bath for 30 seconds, reactions were chilled on ice for 5 minutes. 500μL of pre-warmed SOC medium were added and the reaction was incubated at 37°C with shaking at 250rpm for 1 hour. Between 20-250μL were applied to LB-agar plates containing 100μg/mL ampicillin, unless stated otherwise.

2.2.4 Isolation and Purification of Plasmids from Bacteria

LB-agar plates were incubated at 37°C for 16-18hrs. Between 2-10 ampicillin-resistant colonies from each transformation reaction were manually picked with a sterile tip and inoculated into 5mL of LB broth containing 100μg/mL ampicillin. Liquid cultures were incubated at 37°C for

16-18hrs with shaking at 250rpm. Bacteria were pelleted by centrifugation at 3,000g for 15 minutes, followed by decantation of the supernatant.

Unless stated otherwise, pelleted samples were processed using the QIAprep Spin Miniprep Kit (Qiagen) following the manufacturer recommended protocol. Purified DNA was then eluted using a low-salt buffer.

2.2.5 Screening of Bacterial Colonies by Colony PCR

Some cloning experiments used colony PCR to screen bacterial colonies for correct ligation products before plasmid purification. Between 8-16 bacterial colonies in LB/Agar plates were individually picked using a P10 plastic tip, dipped briefly in a pre-prepared 20μL MyTaq HS Red Mix PCR (containing the adequate primer pair) and ejected in a mirror PCR tube containing 150μL LB medium (without ampicillin) and kept at RT. The PCR thermocycling protocol had an initial denaturation at 95°C for 1 minute, followed by 35 cycles of amplification (95°C for 15s, 55-60°C for 15s, 72°C x 20s), and a final extension at 72°C for 1 minute. After thermocycling, the reactions were electrophoresed on agarose gels and imaged. Colonies found to carry the correct insert were identified and retrieved from the mirror PCR tube into a 5mL LB-ampicillin liquid culture which was incubated at 37°C with 250 rpm shaking overnight. The following day the liquid cultures were processed as described previously.

2.3 Cell Culture

All cell culture work was done in dedicated tissue culture facilities. Immortalised HEK293T and H9 human embryonic stem cell cultures were carried out in separate facilities. Tissue culture hoods were sprayed and wiped twice with 70% Industrial Denatured Alcohol (IDA) before and after any cell culture work. Pipettes, tip boxes and any other equipment brought into the hoods were sprayed and wiped with IDA before and after any cell culture work. Aseptic technique was followed. All media being added to cell cultures were pre-warmed to 37°C unless otherwise specified.

2.3.1 Culture of Immortalised HEK293T Cells

The HEK293T cell line was a kind gift from Prof. Zam Cader (WIMM, Oxford) to Dr. Hannah Sleven, a previous post-doc in Prof. Andrea Nemeth's lab.

2.3.1.1 Maintenance Conditions

HEK293T cells were maintained in Dulbecco's Modified Eagle Medium (DMEM) (Gibco, cat. #11995065) supplemented with 10% Fetal Calf Serum (Gibco) and 1% Penicillin/Streptomycin (Gibco cat. 15140122).

2.3.1.2 Resuspension of Frozen Cells

Frozen cryovials containing cell stocks were retrieved from liquid nitrogen, rapidly thawed to 37°C in a water bath and transferred to a 15mL polypropylene tube containing 9mL of pre-warmed maintenance medium. Cells were collected by centrifugation at 500g for 5 mins, and residual freezing medium was removed by aspirating the supernatant. Cell pellets were resuspended in 15mL of pre-warmed maintenance medium, transferred to a new T75 culture flask and kept at 37°C in a 5% CO_2 incubator.

2.3.1.3 Cell Passaging

HEK293T cells were passaged whenever they covered more than 75% of the surface area of the culture flask, typically every 3-4 days. Growth medium was aspirated and discarded, and the cell layer was carefully washed with 5mL of room temperature PBS. Cells were dissociated by incubating with 1.5mL of 1X TrypLE Express (Gibco) for 5 minutes at 37°C. Dissociation reagent was inactivated by 10-fold dilution in serum-containing growth medium, and an aliquot was used to seed a new culture flask at the desired split ratio (1:10 to 1:30).

2.3.1.4 Transfection of Guide RNAs into HEK293T Cells

One day before transfection, HEK293T cultures at 70-90% confluency in T75 flasks were washed in 1X PBS and dissociated using 1mL of 1X TrypLE Express (Gibco). Following incubation at 37°C for 5 minutes, cells were resuspended in 10mL of DMEM/FCS medium

and transferred to a polypropylene tube. Cell count was established using a BioRad TC20 automated counter. Unless otherwise specified, cells were diluted to 2.8 x 10e5 cells/mL and 0.5mL thereof were seeded to each well of a 24-well culture plate. The plate was kept at 37°C in a 5% CO_2 incubator overnight.

Transfection was performed only if wells were between 70-95% confluent. For each transfection reaction, 500ng of plasmids to be transfected were diluted in 50μL of Opti-MEM (Gibco) in a sterile tube, mixed with 1.5μL of TransIT LT1 reagent (Mirus). Following incubation for 25 minutes at RT, the reaction was added to the cell culture medium in each well, and culture plates were returned to the incubator for at least 24 hrs before assessing transfection efficiency.

Transfection experiments usually included untransfected and EGFP-transfected control wells. Overall transfection efficiency was estimated by visually determining the percentage of fluorescent cells amongst all surviving GFP-transfected cells 24-48hrs post-transfection.

2.3.1.5 Freezing Cell Stocks

T75 flasks at 70-80% confluency had their medium aspirated and the cell monolayer was carefully washed once with 1X PBS ([-]Ca^{2+} [-]Mg^{2+} pH 7.4, Gibco cat. 10010). One millilitre of TrypLe Express (Gibco) was added to cells. After a 5 minute incubation at 37°C, dissociation reagent was inactivated by 10-fold dilution in serum-containing medium. Cells were transferred to a polypropylene tube and pelleted by centrifugation at 500g for 5 minutes. After the supernatant was discarded, the cell pellet was reconstituted in chilled freezing medium (90% FCS / 10% DMSO) and 1mL was transferred to each cryovial. Cryovials were placed in a Mr. Frosty freezing container (Thermo Scientific, UK) and stored overnight at -80°C. The following day, the cryovials were moved to liquid nitrogen vapour for long-term storage.

2.3.2 Culture of Human Embryonic Stem Cells

The H9 human embryonic stem cell line (WA09-FI) was purchased from WiCell (Agreement No. 15-W0175, passage number 24 on arrival) and importation into the United Kingdom was approved by the Steering Committee of the UK Stem Cell Bank. All H9 experiments described

use cells between passage numbers 28-40.

2.3.2.1 Maintenance Conditions

H9 hESCs were maintained in feeder-free conditions in mTeSR1 medium (StemCell). Complete medium was prepared by slowly thawing 100mL of 5X mTeSR1 Supplement overnight at 4°C and adding it to 400mL of mTeSR1 Basal medium. Freshly prepared bottles were split into 45mL aliquots and stored at -20°C. When needed, aliquots were thawed overnight at 4°C and kept in the fridge for a maximum of 2 weeks. Maintenance medium was changed daily, unless stated otherwise. As a general rule, the following volumes of pre-warmed complete mTeSR medium were used:

- 6-well plates: 2mL/well
- 24-well plates: 0.5mL/well
- 96-well plates: 100μL/well
- 60mm dishes: 3mL/dish

Unless specified, H9 hESCs were grown under antibiotic-free conditions.

2.3.2.2 Preparation of Culture Ware Prior to Stem Cell Culture

All tissue culture plates and dishes were first coated with a matrix to support cell attachment and expansion of H9 hESCs. Following manufacturer's recommended protocol, new vials of Matrigel hESC-qualified Basement Membrane Matrix (Corning #354277), hereby referred to as Matrigel, were aliquoted according to the lot-specific protein concentration (dilution factor) included on the Certificate of Analysis. When needed, aliquots were retrieved from -80°C storage and, working on ice and using chilled pipettes and tips, freshly diluted in 12.5mL of ice-cold KnockOut DMEM (Gibco #10829018). Plates and dishes were freshly coated when required by adding diluted Matrigel as follows:

- 6-well plates: 1mL/well
- 24-well plates: 150uL/well

- 96-well plates: 30uL/well
- 60mm dishes: 2mL/dish

Culture ware containing Matrigel were incubated at -37°C for at least 1 hour. Diluted mMtrigel was aspirated, leaving a thin gel coating, immediately before cells were added to each plate or dish.

2.3.2.3 Resuspension of Frozen Cell Stocks

Defrosting medium was prepared by supplementing complete mTeSR1 medium with 10μM ROCKi (Y-27632 Rho Kinase Inhibitor, Abcam ab120129) and 1% Penicillin/Streptomycin (Gibco cat. 15140122). Frozen cryovials containing cell stocks were retrieved from liquid nitrogen, rapidly thawed to 37°C in a water bath and transferred to a 15mL polypropylene tube containing 9 mL of pre-warmed KnockOut DMEM (Gibco). Cells were pelleted by centrifugation at 300g for 5 minutes, and residual freezing medium was removed by aspiration of supernatant. Cell pellets were resuspended in 6mL of pre-warmed defrosting medium, seeded to 3 Matrigel-coated wells of a 6-well-plate, and kept at 37°C in a 5% CO_2 incubator. The following day plates were checked for viability and defrosting medium was replaced with standard maintenance medium.

2.3.2.4 Cell Passaging

H9 hESCs were passaged at 70% confluency or when the majority of colonies were large, compact and had dense centres. The frequency at which cell passaging was required varied with seeding density, but on average cultures were passaged on a weekly basis. Generally one well at optimal density was split at ratios of 1/6 to 1/10. Unless specified otherwise, cells were passaged in aggregates.

Growth medium was aspirated and discarded, and the cell layer was washed twice with 1-2mL of room temperature 1X PBS ([-]Ca^{2+} [-]Mg^{2+} pH 7.4, Gibco cat. 10010). One millilitre of room-temperature ReLeSR Passaging Reagent (StemCell) was added to the well and aspirated after 30 seconds. Following incubation at 37°C for exactly 5 minutes, 1mL of maintenance

medium was added and the side of the plate was hand tapped firmly for 15-60 seconds to manually detach cells. Aggregate size was verified using an inverted microscope. Using a 5mL pipette, medium containing aggregates was transferred to a 15mL conical tube and, depending on the desired split ratio, mixed with variable volumes of additional maintenance medium. Matrigel-coated culture plates were seeded with diluted aggregate-containing medium at the desired split ratio, returned to the incubator and moved in several back-and-forth and side-to-side movements to evenly distribute the aggregates.

2.3.2.5 Single Cell Dissociation

If H9 hESC cultures had to be dissociated into single cells prior to transfection or clonal isolation, Accutase dissociation was performed. A pre-treatment with Rho Kinase Inhibitor (ROCKi) was applied between 1-4 hours prior to dissociation. Maintenance medium was replaced with complete mTeSR1 supplemented with 10μM ROCKi (Y-27632 Rho Kinase Inhibitor, Abcam ab120129) and the plate was returned to the 37°C incubator until dissociation.

For dissociation, medium from each well was aspirated and discarded, and the cell layer was washed twice with 1mL of room temperature 1X PBS ([-]Ca^{2+} [-]Mg^{2+} pH 7.4). One milliliter of room-temperature Accutase Cell Detachment Solution (BioLegend, cat. #423201) was added per well and the plate was incubated at 37°C for 8-15 minutes until complete detachment (visually confirmed by inverted microscopy). Detached cells were transferred to a 15mL conical tube containing 9mL of pre-warmed KnockOut DMEM. Remaining multi-cell clusters were dissociated by pipetting up and down 6-10 times with a 10mL serological pipette. Cells were pelleted by centrifugation at 300g for 5 minutes. After the supernatant was carefully aspirated, cells were resuspended in the 5mL of maintenance medium and a 10μL aliquot was used for cell counting. A volume containing the desired cell count was used for downstream experiments.

2.3.2.6 Freezing Cell Stocks

Each well of a 6-well plate at 70% confluency was frozen into 3 cryovials as follows. Medium from each well was aspirated and discarded, and the cell layer was washed twice with 1mL

of room temperature 1X PBS ([-]Ca^{2+} [-]Mg^{2+} pH 7.4). One mililiter of room-temperature ReLeSR Passaging Reagent (StemCell) was added to the well and aspirated after 30 seconds. Following incubation at 37°C for exactly 5 minutes, 1.5mL of cold maintenance medium was added per well and the side of the plate was tapped firmly by hand for 30-60 seconds to detach cells. Aggregate sizes were verified using an inverted microscope. After 1.5mL of cold 2X Freezing Mix (20% DMSO [Sigma], 60% ES-qualified Fetal Bovine Serum [Gibco], 20% KnockOut DMEM [StemCell]) was added to each well, 1mL was transferred to each pre-labelled 2mL cryovial. Cryovials were placed in a Mr. Frosty freezing container (Thermo Scientific, UK) and stored overnight at -80°C. The following day, cryovials were moved to liquid nitrogen vapour for long-term storage.

2.3.2.7 Reverse Transfection of H9 hESCs

Transfection of plasmids into H9 hESCs was done in a 24-well format. For each transfection reaction, a total 1μg of purified endotoxin-free plasmids was mixed with 3μL of TransIT LT1 Reagent (Mirus) in 100μL of Opti-MEM and incubated at RT for 20 minutes. Matrigel-coated transfection plates were prepared by adding 125μL of maintenance medium supplemented with 10μM ROCKi (Y-27632, Abcam) or 10X CloneR Cloning Supplement (StemCell) and stored in a 37°C incubator until needed. Healthy H9 cultures of 60-80% confluency were pre-treated with ROCKi and single-cell dissociated using Accutase. Dissociated cells were resuspended to 1.4 x 10^6 cells/mL in maintenance medium supplemented with 10μM ROCKi. After 250μL of resuspended cells were seeded to each well of the 24-well plate, the full transfection reaction was added ans plates were incubated undisturbed overnight. The following day medium containing transfection reagents was aspirated and fresh complete medium supplemented with 10μM ROCKi or 10X CloneR Supplement was added.

2.3.2.8 Clonal Isolation and Analysis

2.3.2.8.1 Preparation of Cloning Medium Cloning medium was prepared by mixing 1 part CloneR Cloning Supplement (Stemcell) with 9 parts of complete mTeSR1 medium. Medium was made fresh for single-cell seeding and kept at 4°C for up to one week.

2.3.2.8.2 Single-Cell Seeding Seeding and clonal isolation of H9 hESCs was performed in 60mm dishes. Healthy H9 cultures of 60-80% confluency were pre-treated with ROCKi and single-cell dissociated using Accutase as described previously. Following cell counting, dissociated cells were resuspended to 180 cells/mL in cloning medium and 3mL were used to seed each Matrigel-coated 60mm dish. Dish was returned to the incubator and incubated undisturbed. After 2 days, a full media change with 3mL of cloning medium was performed. The following day, 1mL of cloning medium was added to each dish. From the fourth day on, full media changes with 3mL of standard maintenance medium were performed daily until colony picking.

2.3.2.8.3 Colony Picking Single-cell-derived colonies were picked after at least 9 days in culture. Colonies were prioritised by morphology (round shape, defined edges, dense centres), size (larger preferred over smaller), and isolation (large gap from other colonies was preferred). 60mm dishes were carefully inspected under an inverted microscope, colonies were identified, prioritised and marked. Existing growth medium was replaced with standard maintenance medium not containing selective antibiotics or ROCKi, unless specified otherwise. Dishes were then incubated at 37°C in a 5% CO_2 incubator for >2 hours before picking. Colony picking was done under aseptic conditions, inside the cell culture hood, using the EVOS XL Imaging System (Invitrogen). Using a P200 pipette set to 75μL, individual colonies were isolated by scraping while slowly aspirating and transferred to Matrigel-coated 96-well plates. To break large aggregates, contents of each well were aspirated up and forcefully ejected five times before the next colony was picked. To avoid significant temperature and pH changes, picking was done in batches of eight colonies before the dish and its matching 96-well-plate were returned to the incubator for at least 30 minutes before the next attempt. After all colonies had been picked, the 96-well-plate was incubated undisturbed overnight.

2.3.2.8.4 Expansion of Picked Colonies The day after colony picking, all wells in 96-well plates were topped up with 50uL of maintenance medium and returned to the incubator. The following day and daily thereafter, growth medium was aspirated and 100uL of maintenance medium supplemented with blasticidin was added, unless specified otherwise. All wells were

carefully inspected daily for survival, density, and signs of contamination. Colonies were grown for 6-12 days.

2.3.2.8.5 Splitting Isolated Colonies for Expansion and DNA Extraction When any wells had reached >80% confluency, cells were passaged into a new expansion plate, and the remaining cells were used for gDNA extraction. For each well, after 1x PBS wash, 20uL of ReLeSR Passaging Reagent (StemCell) were added and quickly aspirated. Following 5 minutes incubation at 37°C, 70μL maintenance medium was used to rapidly flush and aspirate each well, transferring detached aggregates to a new Matrigel-coated 96-well expansion plate.

2.3.2.8.6 Harvesting Cells for DNA Analysis Remaining attached cells in the original 96-well plate were harvested for gDNA extraction by adding 25μL TrypLE Express (Gibco) to each well and incubating at 37°C for 30 minutes. Each well was flushed with 150μL of 1X PBS, and the entire well volume was transferred to each matching well of a 96-well PCR plate. Cells were pelleted by centrifugation at 4000rpm for 15 minutes followed by aspiration of supernatant.

2.3.2.8.7 DNA Extraction from 96-Well Plates Genomic DNA was enzymatically extracted using MyTaq Extract-PCR Kit (Bioline). An extraction mastermix was prepared by mixing 7 volumes of nuclease-free water, 2 volumes of Buffer A, and 1 volume of buffer B, and 10uL thereof was added to each cell pellet. The plate was vortexed at high speed for 30 secs and incubated at 37°C for 30 minutes. After heat-inactivation for 80°C for 10 minutes, 40μL of nuclease-free water was added to each well. The plate was vortexed once more, and any cell debris were pelleted by centrifugation at 4000rpm for 10 minutes.

3 | Development of a CRISPR/Cas9-Based Genome Editing System

3.1 Introduction

Dr. Hannah Sleven, then a post-doctoral research scientist in the Nemeth lab, had conceptualised a two-stage CRISPR/Cas9-based homology-directed repair (HDR) approach designed to generate heterozygous missense mutations while markedly reducing colony screening efforts. This system has been further elaborated, assembled and fine-tuned in this DPhil project, and is hereby presented as Cassette Insertion-Replacement CRISPR Editing System (CIRCES).

3.1.1 General Principle

CIRCES was designed to edit human stem cells in two sequential rounds of CRISPR/Cas9-mediated HDR editing. In the first, the target gene is effectively knocked out by insertion of a selection cassette, while the second round replaces the cassette with the desired final mutation. By harnessing the power of both positive and negative selection, the inherently low HDR efficiency in human stem cells should be bypassed.

In the "Insertion Stage", CRISPR/Cas9 is used to generate a double-strand break (DSB) at the genomic target locus (Figure 3.1A). While most DSBs will be repaired by the cell using the NHEJ pathway, by providing an HDR donor template that contains a selection cassette flanked by locus-specific homology arms (Figure 3.1B), some cells are expected to integrate ("insert") the selection cassette into the target region by HDR (Figure 3.1C). Since the insertion cassette contains an antibiotic resistance gene, application of positive antibiotic selection ensures that only cassette-expressing cells reach the end of this stage (Figure 3.1C). A caveat of this approach is that cells with randomly integrated copies of the cassette would similarly survive selection. However, as detailed below, these would be eradicated by negative selection in an efficient and high-throughput manner in the next stage.

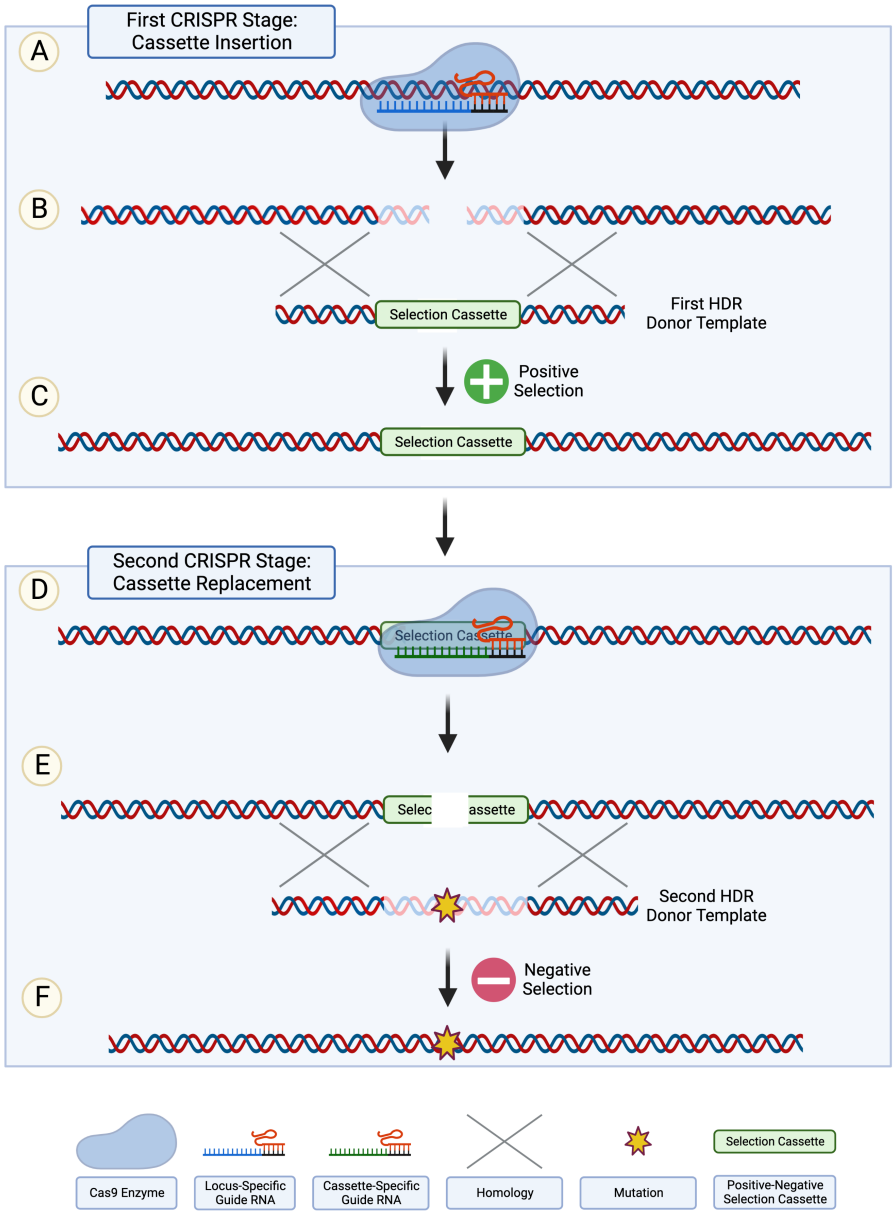

Figure 3.1: Cassette Insertion-Replacement CRISPR Editing System (CIRCES). (A) For the first CRISPR round, part of the CIRCES Insertion Stage, a sgRNA is used to direct Cas9 to generate a DSB at its desired target locus. (B) Foreign DNA sequences can be introduced at a DSB break-site by providing an HDR template. For the insertion stage of CIRCES, this template is a positive-negative selection cassette flanked by homology arms to the target locus. (C) Although the efficiency of HDR-mediated integration of the selection cassette is low, populations of successfully integrated cells can be selected and expanded via positive antibiotic selection. (D) Selected cells are subjected to a second round of genome editing where a new sgRNA directs Cas9 to generate a DSB in the (now integrated) selection cassette itself. (E) A second HDR template, containing any desired mutations flanked by the same previously used homology arms, is used, by some cells, as template for another round of homology-directed DNA repair. (F) The efficiency of HDR-mediated integration of the desired mutations is also expected to be low, but negative selection using a suicide gene approach is used to selectively remove cells which continue to express the integrated cassette. As a consequence, surviving cells after both CIRCES selection rounds should be greatly enriched for cells containing the desired mutation(s).

In "Replacement Stage", CRISPR/Cas9 targets the integrated cassette itself (Figure 3.1D). By providing a second HDR template, which contains the desired final mutation (Figure 3.1E),

some cells should replace the mutant template for the genomic cassette (Figure 3.1F). Using this two-step system, it should be possible to obtain clones that were doubly targeted at the same locus. Cells still expressing the cassette (either on-target or randomly inserted) are largely removed by application of negative antibiotic selection. At the end of the second stage, theory predicts that significant enrichment of doubly corrected (mutant) cells will have occurred, as only cells that underwent two rounds of CRISPR HDR should be able to survive (Figure 3.1F).

The main benefit of this system would be an extensive reduction in the number of colonies that need to be screened for a successfully edited clone to be identified, enabling rapid, efficient and cost-effective generation of mutant cell lines. As will be shown, the first stage, once refined, was extremely efficient at generating knock-out mutations, but the second stage had unexpected limitations that did not enable generation of "scarless" missense mutations. The concepts, successes and difficulties of using CIRCES are the subject of the following three chapters.

3.1.2 Rationale and Design of the Selection Cassette

CIRCES was a novel system, and as such the positive-negative selection cassette had to be designed and assembled before it could be used to target ITPR1.

The key feature in this concept is adequate expression of the genes critical for positive and negative selection. Several antibiotics are available for positive selection in mammalian cells. Blasticidin was chosen because of the small size of its resistance gene, Blasticidin-S-Deaminase (BSD, 396bp CDS). In preliminary experiments in H9 hESCs, blasticidin selection was found to be largely unaffected by cell density. For the subsequent negative selection, the choice of genes is limited. A 996bp truncated Herpes Simplex Virus Thymidine Kinase (HSV-ΔTK, from now on referred to as TK) was chosen as a "suicide gene"[147]. Cells expressing TK undergo apoptosis when exposed to ganciclovir, a commonly used antiviral drug[148]. Expression of the positive selection gene (BSD) must always be coupled to the negative marker (TK) to ensure that any cells surviving blasticidin selection will die upon treatment with ganciclovir unless the cassette ceases to be expressed. For this, a fusion protein approach was chosen. Applying this method to human stem cells requires the BSD-TK fusion protein to be under the

control of a strong and stably expressed mammalian promoter. The full length EF1α promoter (1182bp) was chosen due to its proven stability in H9 hESCs[149].

Assembling an ideal CRISPR/Cas9 target for the Replacement Stage required several considerations. The EF1α-BSD-TK sequence is the core of the double selection, which must remain functional unless removed from the genome. Despite the optimal central location in the cassette and its non-human genomic sequence, the BSD gene cannot serve as a CRISPR/Cas9 target for the second stage since NHEJ-mediated repair could lead to loss-of-function of the entire fusion protein due to reading frame disruption. As such, the second editing step needed to target a sequence located well outside the EF1α-BSD-TK block[122]. This offered the opportunity to use an artificial non-mammalian sequence that maximised cutting efficiency while minimising potential off-target effects. A 23bp sequence from the dCG6744 *Drosophila melanogaster* gene was chosen due to is minimal homology to the human genome, and because a guide RNA had been previously reported to target it efficiently (Dr. Joey Riepsaame, Genome Engineering Oxford at Sir William Dunn School of Pathology, personal communication). This Drosophila-specific sequence (DSS) was positioned 5' of the EF1α-BSD-TK block, leaving 50bp of padding sequence that could be disrupted by minor indels without impacting EF1α promoter activity. The final component required to ensure adequate cassette activity is the termination signal for protein expression in eukaryotic cells (bGH pA - bovine growth hormone polyadenylation signal). The 3050bp structure of the cassette is illustrated in Figure 3.2.

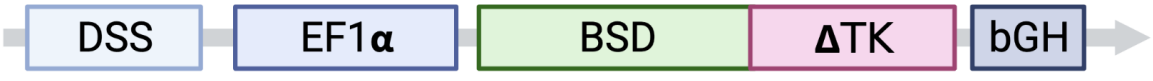

Figure 3.2: Structure of the CIRCES cassette. DSS - Drosophila-Specific Sequence, EF1α - EF1α Promoter, BSD - Blasticidin-S-Deaminase (antibiotic resistance gene), ΔTK - Truncated Herpes-Simplex Virus Thymidine Kinase ("suicide gene"), bGH - Bovine Growth Hormone Polyadenylation Signal.

3.1.3 Components

CRISPR/Cas9 components and HDR templates are to be delivered to cells as plasmid vectors. Both CRISPR rounds (Fig. 3.3A and B) require different plasmids to be simultaneously transfected. One plasmid expresses both the Cas9 nuclease and the targeting sgRNA (Fig. 3.3, left) while the other contains the HDR donor template (Fig. 3.3, right).

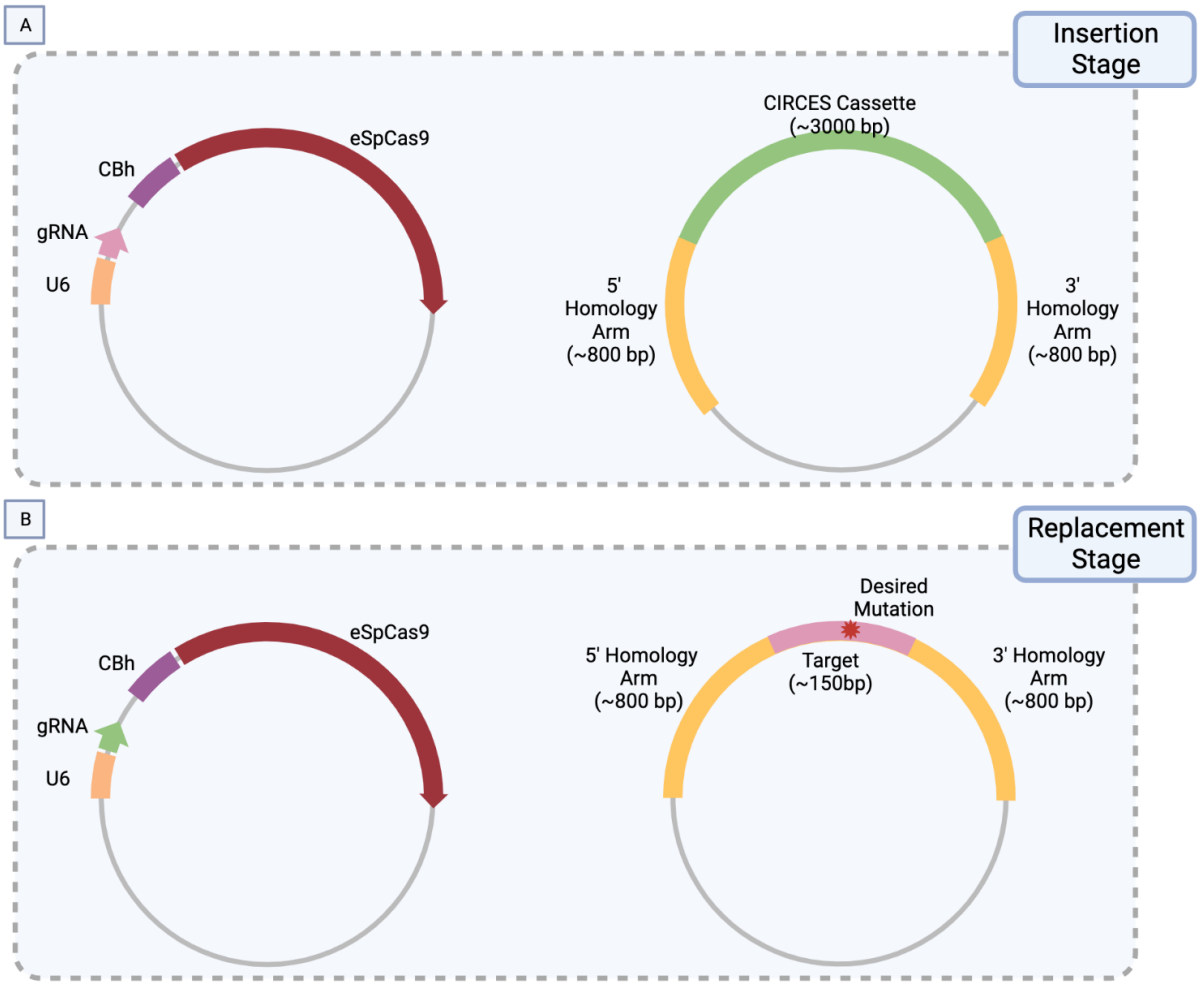

Figure 3.3: CIRCES Components. (A) To initiate the Insertion Stage of CIRCES, two plasmids are transfected to wild-type cells. The first is a pX330-based vector that expresses the target-specific sgRNA (under the control of the hU6 promoter) and the enhanced SpCas9(1.1) enzyme (under the control of the CBh promoter) [left-hand side]. The second is a pGEM-based plasmid as the Insertion Stage HDR donor template that contains the CIRCES cassette flanked by ∼800bp regions of homology to the genomic context of the target region [right-hand side]. (B) For the Replacement Stage of CIRCES, cells that have completed the insertion stage are transfected with two different plasmids. The first (left-hand side) is an eSpCas9 vector with the Drosophila sgRNA. When expressed in cells, this sgRNA will direct the eSpCas9 to cleave the integrated cassette at its DSS target site. The second transfected plasmid is the HDR donor template (right-hand side), which is flanked by ∼800bp homology arms to deliver any desired mutations to be introduced. Given that the first HDR template is cloned excluding ∼150bp of genomic context immediately adjacent to the target sgRNA site, if the goal is to achieve "seamless" genomic editing then these ∼150bp must be present between the homology arms of the second HDR template.

3.1.3.1 Homology-Directed Repair Donor Templates

Two different HDR donor templates are required for each target. The insertion stage template contains the selection cassette flanked by homology arms to the relevant target gene (Fig. 3.3A, right). The replacement stage template contains the same flanking homology arms, but encloses the mutation of choice to be introduced in a way that replaces the selection cassette introduced previously (Fig. 3.3B, right). The aim is to achieve "scarless" introduction of the mutation, leaving the remaining genomic context intact. To improve HDR-mediated integration and replacement of a large 3kb cassette, flanking homologous sequences of approximately 800bp were chosen.

3.1.3.2 Cas9 Enzyme Selection and Guide RNA Design

Given the two CRISPR rounds, it was essential to minimise potential off-target activity[138]. The 'enhanced' Cas9 version eSpCas9(1.1), which had recently been reported to have similar on-target efficiency as the original SpCas9 while being almost devoid of detectable off-target effects[130], was used as an expression vector (Fig. 3.3A, left) that allows simultaneous delivery of both the Cas9 and the corresponding sgRNA.

Unlike traditional SpCas9, eSpCas9 is an extremely specific enzyme which requires virtually perfect complementarity between the sgRNA and its target sequence[130, 132]. This stringency, however, comes at a cost. The hU6 promoter is thought to preferentially start transcription from a guanine (G) nucleotide[150]. For that reason, unless a 20bp sgRNA natively starts with a G, it has been common practice to append a 5' mismatched guanine (5'GN_{20}) to boost transcription of non-5'G-native guides[150]. Given eSpCas9's stringency, however, this practice is expected to significantly reduce cleavage efficiency. Consequently, use of eSpCas9 restricts the choice of optimal sgRNAs and it is imperative that their on-target efficiency be experimentally confirmed.

Validation of sgRNAs was therefore performed in HEK293T cells to assess for on-target indel formation. As transfection efficiency is never complete nor constant, antibiotic selection of HEK293T cells actively expressing eSpCas9 is preferred. As no antibiotic-selectable version

of eSpCas9 was available at the time, a Puromycin-resistant version of eSpCas9 was cloned, tested and later submitted to the Addgene repository (no.101039).

3.1.4 ITPR1 Target Regions and Mutations

Using CIRCES, the aim was to develop disease models recapitulating proposed genetic mechanisms of SCA29 and Gillespie Syndrome by creating human embryonic stem cell lines with the following mutations:

- Autosomal Recessive Gillespie: Biallelic Loss-of-Function (Exons 10, 58 and 59)
- Autosomal Dominant / De Novo Gillespie: G2554R (Exon 58), K2611del (Exon 59)
- Autosomal Dominant / De Novo SCA29: R269G (Exon 10)

Thus three different target regions — ITPR1 Exons 10, 58, and 59 — with three target mutations — R269G, G2554R, K2611del — were chosen.

As on-target cassette insertion effectively deletes an entire exon, it was expected to give rise to a loss-of-function allele and that some biallelic loss-of-function clones would be identified in each target by the end of the protocol.

The Insertion Stage of CIRCES requires that for *each target* region two target-specific components be cloned and transfected into cells:

- Expression vector containing a validated target-specific sgRNA and the eSpCas9
- HDR template containing the CIRCES cassette flanked by target-specific homology arms

In addition, for the Replacement Stage, another two components have to be delivered:

- Expression vector containing the Drosophila sgRNA and eSpCas9 (this vector is target-independent)
- Mutant HDR donor template (which must be prepared for each target mutation)

This chapter describes cloning and validation of all 10 plasmids (3 target-specific constructs per target exon and 1 Drosophila sgRNA-eSpCas9 vector) so that they could be used directly

in hESCs to achieve genomic editing of ITPR1 using CIRCES.

3.2 Materials and Methods

3.2.1 Assembly of the CIRCES Cassette

Puromycin selection is not used in the cassette, but the cloning approach chosen effectively created a fully functional EF1α-Puro-TK cassette before the Puromycin resistance gene was swapped for the Blasticidin gene. Using vectors already available in our lab, assembly of the CIRCES cassette progressed in three rounds of cloning (see Fig. 3.9 in the results section). The first round joined the EF1α promoter and a Puromycin-TK fusion gene in a pCAGEN vector (EF1α-Puro-TK plasmid). To this vector, the Drosophila-Specific Sequence (DSS) was inserted. Finally the Puromycin resistance gene was replaced with Blasticidin-S-Deaminase (BSD) while maintaining the correct reading frame.

3.2.1.1 First Round: Cloning the Intermediary EF1αPuro-TK Vector

pEF-GFP (Addgene #11154) was double digested with KpnI-HF and HindIII-HF, and the fragment containing the pCAGEN vector backbone and EF1α promoter was gel extracted.

pLCA.66/2272 (Addgene #22733) that contained the fusion Puromycin-N-Acetyltransferase-Thymidine-Kinase (Puro-TK) gene and the bGH polyadenylation signal was used as template in a Q5 High-Fidelity PCR reaction with primers containing KpnI and HindIII overhangs. The PCR product was digested with KpnI-HF and HindIII-HF and gel purified.

The obtained fragments were ligated and transformed into NEB10-beta competent bacteria. Clones were screened and verified by double KpnI/HindIII digestion followed by Sanger sequencing.

3.2.1.2 Second Round: Insertion of the Drosophila-Specific Sequence Upstream of the EF1α Promoter

The 23bp Drosophila-specific sequence (DSS) was inserted upstream of EF1α promoter in the EF1α-Puro-TK vector. The plasmid was linearised using an inverse PCR approach with opposite-facing side-by-side primers containing NotI and BsmBI overhangs. To generate ap-

propriate 'sticky ends', the amplicon was ethanol precipitated, reconstituted in CutSmart Buffer, double digested with NotI-HF and DpnI for 2 hours at 37°C, and heat-inactivated at 65°C for 20 minutes. After another round of ethanol precipitation, it was reconstituted in NEB Buffer 3.1 and simultaneously digested with BsmBI and dephosphorylated with rSAP for 2 hours at 55°C. After a final round of heat-inactivation at 80°C for 20 minutes, it was column purified.

The DSS was cloned as two partially complimentary single-stranded DNA oligonucleotides containing the 23bp Drosophila gDNA sequence flanked by NotI and HindIII overhangs. After phosphorylation and annealing, the annealed oligonucleotide was diluted 200-fold before it was ligated into 50ng of the purified EF1α-Puro-TK PCR product and transformed into NEB10-beta competent cells. After plasmid purification, correct insertion was verified by Sanger sequencing.

3.2.1.3 Final Round: Replacing Puromycin with Blasticidin resistance

To finalise the assembly of the cassette, the Puromycin-N-Acetyltranferase gene was replaced by BSD.

The BSD gene was amplified from the pLV-TRET-hNgn1-UBC-Bla (Addgene #61473) plasmid in a Q5 High-Fidelity PCR reaction using 20 cycles and primers containing 20bp overhangs complementary to the 3' of the EF1α promoter and the 5' end of the TK gene (primer sequences in Appendix A). To reduce the chance of template plasmid contaminating the PCR product, the amplified product was diluted 1:250 in nuclease-free water and 0.25μL was used in a second 18-cycle PCR reaction with the same set of primers.

The previously assembled backbone was inverse PCR-amplified as a single large amplicon and 10U of DpnI were added to the final product to digest remaining methylated plasmid template.

The two-round purified BSD gene and the TK-bGH-backbone-EF1α PCR product were combined in a NEBuilder HiFi DNA Assembly reaction following the standard NEB protocol and transformed into NEB10-beta competent bacteria. After plasmid isolation, clones were screened by double digestion with XbaI/NheI and colonies were further verified by Sanger sequencing.

3.2.2 Cloning of the HDR Donor Templates

Different HDR donor templates are required for Insertion and Replacement stages, and both are delivered as circular double-stranded plasmids. Given that the same homology arms are used in both templates, an optimised three-round cloning strategy was used to generate the transfection-ready donor templates (Fig. 3.10 in the results section).

3.2.2.1 First Round: Cloning Wild-Type ITPR1 Sequences into pGEM Vectors

Genomic DNA from wild-type H9 hESCs was extracted using the QIAmp DNA Mini kit (Qiagen) following the recommended protocol and used as template in Q5 High-Fidelity PCR reactions to generate 1500-2000bp amplicons encompassing each target ITPR1 exon and both homology arms (primer sequences in the Appendix A). Amplified fragments were gel purified and sequenced. Sequences were validated against the hg38 human genomic reference sequence to identify polymorphisms which might affect primer and/or sgRNA complementarity. Products were A-tailed, ligated into pGEM-T Easy vectors (Promega), and selected by blue/white colony screening. The ligation outcome was confirmed by bi-directional Sanger sequencing using M13 primers.

3.2.2.2 Second Round: Generating Insertion Stage Target-Specific ITPR1-Cassette HDR Template Plasmids

The wild-type ITPR1 pGEM constructs were inverse PCR amplified using opposing primers that contain 20bp overhangs homologous to the ends of the DSS-EF1α-BSD-TK Cassette. This PCR reaction amplifies the entire length of the intermediary vector, excluding only a 100-150bp region that contains the expected guide RNA target site and the nucleotides to be ultimately edited. This gap is the exact location where the cassette will be inserted. In parallel, the fully assembled CIRCES cassette (DSS-EF1α-BSD-TK) was PCR amplified from its pCAGEN vector. The products of all four PCR reactions (cassette, ITPR1 Exons 10, 58, and 59) were gel extracted using Macherey-Nagel's Nucleospin Gel Extraction kit using the manufacturer recommended protocol. To clone each of the Insertion Stage HDR Donor plasmids (Exon10-Cassette, Exon58-Cassette, and Exon59-Cassette), NEBuilder HiFi DNA Assembly reactions

were set up with two-fold excess of cassette amplicon relative to 40ng ITPR1 Exon pGEM amplicon, and transformed into NEB10-beta competent cells. After plating onto ampicillin-agar plates, picked colonies were screened by colony PCR. Correctly ligated colonies were purified using Zippy Endo-Free Miniprep kit (Zymo). Sequence integrity was confirmed by Sanger sequencing using multiple internal primers.

3.2.2.3 Third Round: Creating the Replacement Stage Mutant ITPR1 HDR Donor Plasmids

The replacement stage HDR donor templates that contain the mutations to be engineered are created by simple site-directed mutagenesis on the intermediary wild-type uninterrupted sequence pGEM plasmids. Mutagenesis primers (Appendix A) for the ITPR1 mutations R269G, G2554R, and K2611del were designed using NEBaseChanger software. Mutations were introduced into their respective previously cloned WT ITPR1-Exon pGEM constructs using Q5 Site-Directed Mutagenesis kit (NEB) following the recommended protocol. Mutated constructs were transformed into Zymo's Mix & Go JH5α competent cells, grown on ampicillin-agar plates, and purified using Zymo's Zippy Endo-Free Miniprep Kit. Sanger sequencing was used to confirm both correct mutagenesis and integrity of homologous regions.

3.2.3 Cloning of the eSpCas9-sgRNA Vectors

3.2.3.1 Cloning of eSpCas9-puro

As described previously, no antibiotic-selectable versions of eSpCas9 were available at the start of this project. To aid validation of sgRNAs in HEK293T cells, a puromycin-resistant version was made. The eSpCas9(1.1) plasmid (Addgene #71814) was split into two fragments by digestion with FseI and NotI-HF. A fragment containing the vector backbone, U6 promoter, sgRNA scaffold, CBh promoter, SV40 NLS, eSpCas9(1.1), and part of the nucleoplasmin NLS was gel purified. The NLS-T2A-Puromycin resistance gene coding sequence was extracted from the pSpCas9n(BB)-2A-Puro (PX462) V2.0 plasmid (Addgene #62987) by digestion with the same restriction enzymes. Fragments were gel purified and ligated. The ligation reaction was transformed into OneShot Stbl3 ultracompetent bacteria (Invitrogen) us-

ing the high-efficiency protocol recommended by the manufacturer and grown on ampicillin-agar plates. Colonies were screened using double NotI-HF and FseI digestion, and successful cloning was confirmed by Sanger sequencing.

3.2.3.2 Cloning sgRNAs into eSpCas9-puro

3.2.3.2.1 Digestion of eSpCas9-puro eSpCas9-puro plasmid was digested with BbsI, dephosphorylated with rSAP (NEB), and gel purified. Concentration was determined using Nanodrop and aliquots of 50ng were frozen and used when needed.

3.2.3.2.2 Cloning sgRNAs into eSpCas9-Puro Partially complimentary single-stranded oligonucleotides containing protospacer sequences were annealed as described in the main methods chapter. 50pmol of annealed oligonucleotides were ligated to 50ng of BbsI-digested eSpCas9-puro backbone and transformed into OneShot Stbl3 ultracompetent bacteria (Invitrogen) following the manufacturer recommended protocol. Plasmids were purified and correct insertion was confirmed by Sanger sequencing.

3.2.4 Assessing Cas9-Mediated Editing Efficiency

As described in the main methods chapter, HEK293T cells were transfected in 24-well-plates with 500ng of eSpCas9-puro plasmids containing each of the different guide RNAs.

3.2.4.1 Puromycin Selection of HEK293T Cells

Puromycin selection started 24hrs post transfection. Culture medium was carefully aspirated and slowly replaced by 0.5mL of medium containing 2.0μg/mL Puromycin. To avoid disturbing the fragile monolayer of cultured cells, no further media changes were performed until cells were imaged and harvested for DNA analysis. After 72 hours of selection a significant number of dead floating cells could already be seen, but for completeness selection was allowed to proceed for 96 hours.

3.2.4.2 Harvesting HEK293T Cells for DNA Analysis

After 4 days of selection, medium and debris were carefully aspirated and discarded. Cells were detached from wells by adding 0.5mL of ice-cold 1X PBS Ca^{++} and Mg^{++}-free. Following 5min RT incubation, cells were manually scraped using a 1mL pipette tip and transferred to a microcentrifuge tube. Cells were pelleted by centrifugation at 5,000g for 10 minutes and the supernatant was discarded. Pellets were flash frozen in dry ice and stored at -20°C until processed.

3.2.4.3 DNA Extraction from HEK293T Cells

Genomic DNA was extracted using Bioline's Extract-PCR Kit. An extraction mastermix was prepared by diluting 125μL of Buffer A and 62.5μL of Buffer B in 437.5μL of nuclease-free water. Each cell pellet was resuspended in 25μL of extraction mastermix. After vigorous vortexing, samples were incubated at 75°C for 30 minutes followed by heat-inactivation at 95°C for 10 minutes. Extracted gDNA samples were allowed to cool down to RT before 75μL of nuclease-free water was added to each sample to dilute out PCR inhibitors and samples were stored at -20°C.

3.2.4.4 Calculation of Cas9-Mediated Editing Efficiency

PCR amplification with exon specific screening primers was done across the expected cut site. Amplicons were enzymatically purified using the ExoSAP method (as described in the general methods section) and submitted for Sanger sequencing. Cutting efficiency (percentage of the pool represented by non-wild type sequences, also referred to as "indel percentage" or "editing efficiency") was calculated for each target-sgRNA combination using Synthego's ICE (Inference of CRISPR Edits) v.2 online tool[121] (see 1.6.5.1 in the main introduction chapter).

3.2.5 Validation of the Drosophila sgRNA in a Reporter System

3.2.5.1 Generation of a Stably Inserted CIRCES Reporter System

ITPR1 Exon 10 was arbitrarily chosen as the insertion site for the CIRCES Reporter system. HEK293T cells were co-transfected with 500ng eSpCas9-puro ITPR1 Exon 10 sgRNA #1 plasmid (see table 3.1) and 500ng ITPR1 Exon 10-Cassette pGEM HDR Template plasmid in 24-well plates. As a control, cells were transfected with either 1000ng of EF1α-EGFP (non-targeting control) or eSpCas9-puro EMX1 sgRNA #1 (positive control). The following day, cells were dissociated and passaged to T25 culture flasks. Blasticidin 5μg/mL selection was started on day two post-transfection. Cells were passaged to a T75 flask after an additional two days and subpassaged at 1:10 to 1:20 split ratios whenever they reached over 70% confluency. Fresh blasticidin was added at each passaging step. Blasticidin concentration was doubled to 10μg/mL from post-transfection day 17 to ensure highest stringency selection.

3.2.5.2 Targeting the CIRCES Cassette using the Drosophila sgRNA

Blasticidin-resistant cells by post-transfection day 27 were dissociated and seeded for transfection in 24-well plates. The following day, blasticidin-resistant cells were transfected with 500ng of Drosophila sgRNA eSpCas9-puro vector (see table 3.1). Control cells were transfected in duplicate with either 500ng EF1α-EGFP (transfection and puromycin selection control) or eSpCas9-puro EMX1 sgRNA #1 (positive selection and negative editing control). Puromycin selection was started at 2μg/mL one day post-transfection and maintained for 4 days.

3.2.5.3 Confirmation of Cassette Insertion and Editing of the DSS Site

To confirm on-target insertion of the cassette (ITPR1 Exon 10 locus), gDNA from surviving cells was subjected to PCR reactions amplifying the cassette in its target locus using an Exon 10 specific forward primer (not present in the homologous regions used in the HDR template; this approach is discussed in more detail in chapter 3) and a Blasticidin-gene specific reverse primer. Confirmation of insertion in the correct ITPR1 Exon 10 locus and calculation of Drosophila sgRNA editing efficiency were achieved by Sanger sequencing the PCR amplicon using internal

primers and chromatogram analysis using Synthego's ICE v.2 as described earlier.

3.3 Results

3.3.1 Construction and Validation of eSpCas9/sgRNA Vectors

3.3.1.1 Antibiotic-Selectable Version of eSpCas9

Antibiotic selection is used during single guide RNA (sgRNA) validation to select for cells effectively expressing the sgRNA and eSpCas9 nuclease. A puromycin-resistant version of eSpCas9 was made by cloning a T2A self-cleaving peptide and the Puromycin-N-Acetyltransferase resistance gene (Puro) to the N-terminus of eSpCas9(1.1) (Fig. 3.4).

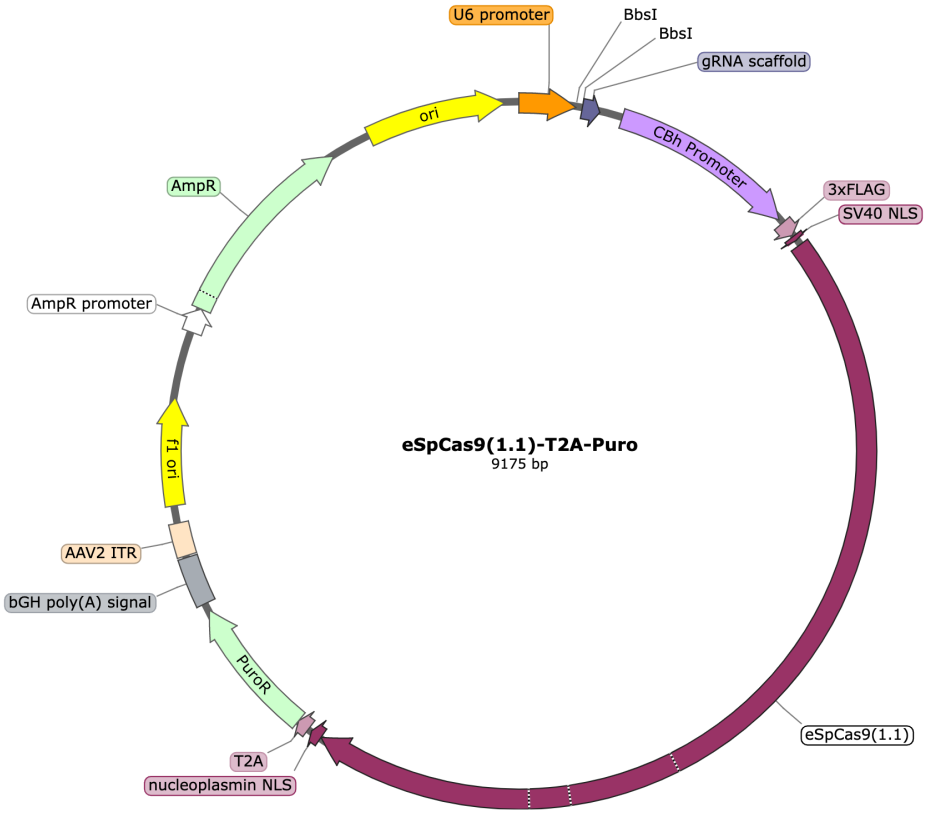

Figure 3.4: Map of the eSpCas9-T2A-puro plasmid (Addgene #101039). Protospacer sequences inserted between the BbsI sites are transcribed along with the tracrRNA scaffold as a single-guide RNA (sgRNA) under the control of the hU6 promoter. The CBh promoter drives the expression of the eSpCas9-T2A-Puro product.

It was tested against the benchmarking gene EMX1 using sgRNAs known to have high efficiency[106, 134, 150]. The sgRNAs were cloned into eSpCas9-puro empty vectors and ligation was confirmed by Sanger sequencing (Figs. 3.5B-C).

HEK293T cells were transfected with either EMX1-sgRNA-1/eSpCas9-puro or EMX1-sgRNA-

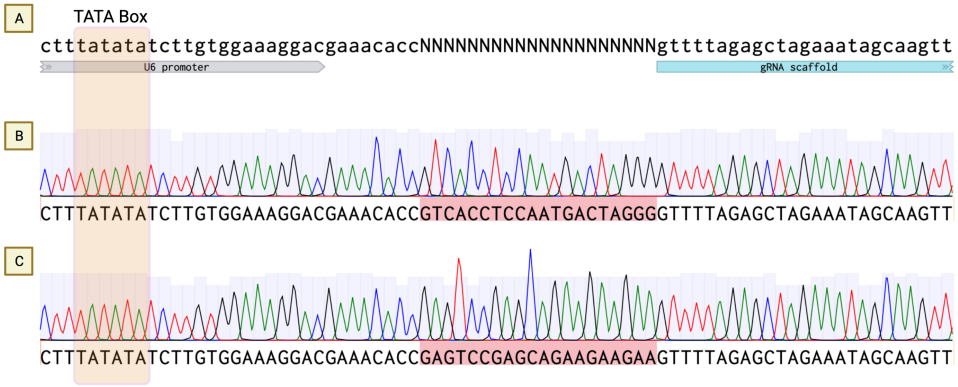

Figure 3.5: Confirmation of EMX1 sgRNA insertion into eSpCas9-Puro. (A) The reference sequence of a BbsI-digested eSpCas9-puro vector, highlighting the location of hU6 promoter sequence and the sgRNA scaffold which are part of the construct. Transcription from the hU6 promoter starts 24nt downstream from its TATA box (highlighted). Sanger sequencing chromatograms showing successfully cloned EMX1 sgRNA #1 (B) and EMX1 sgRNA #3 (C) in eSpCas9-puro.

3/eSpCas9-puro and selected with puromycin for 4 days. Untransfected cells or cells transfected only with an EGFP construct were completely killed, whereas cells transfected with eSpCas9-puro constructs were fully confluent (data not shown), confirming puromycin resistance. Nuclease activity was evaluated by assessing DNA from puromycin-resistant cells for indel formation at the expected EMX1 targets. Sanger sequencing revealed frameshifts starting at the expect cut sites for both sgRNAs (Figure 3.6), confirming that fusion with puromycin did not impact eSpCas9's nuclease activity.

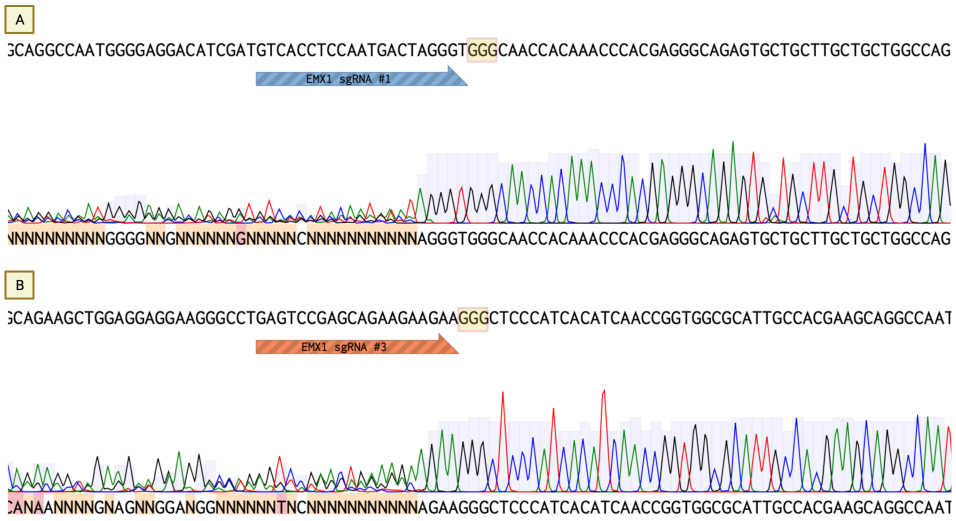

Figure 3.6: Sanger sequencing chromatograms of the target EMX1 sgRNA #1 (A) and #3 (B) loci in HEK293T cells transfected with eSpCas9-puro-gRNA constructs and puromycin selected for 4 days. Sequencing reads were generated using reverse sequencing primers. Abrupt reduction in peak height and appearance of multiple peaks coincide with the expected cut site of each guide, between the 3rd and 4th nucleotides 5' of the protospacer adjacent motif (PAM) sequence (highlighted GGG).

3.3.1.2 Design and Validation of eSpCas9-sgRNA Vectors Targeting ITPR1

To identify the sgRNAs most likely to be effective for each of the three ITPR1 targets, genomic sequences for ITPR1 Exons 10, 58, and 59 were submitted to Benchling's CRISPR Guide Design tool. A list of possible sgRNAs was ranked in decreasing order of optimised Doench's on-target score[137]. Between four and six guides (Table 3.1) for each target were prioritised for validation according to a subjective compromise between the following factors:

- High On-Target Optimised Doench Score[137]
- Low Combined Off-Target Score
- Guides preferably starting with a Guanine (1st choice) for optimal sgRNA transcription[103]
- Distance between cut site and target mutation site (the closer the better)[151]

In total, 16 prioritised sgRNAs were cloned into eSpCas9-T2A-puro (hereby referred to only as eSpCas9) and confirmed by Sanger sequencing (Table 3.1).

Target	Identification	Sequence (5' to 3')
Exon 10	ITPR1 ex10 sgRNA #1	GCACGTCTTCCTGAGAACCA
	ITPR1 ex10 sgRNA #5	TTCTCACCTGTGACGAACAC
	ITPR1 ex10 sgRNA #6	ACTGCCGGCCCGTGGTTCTC
	ITPR1 ex10 sgRNA #6opp	GAGAACCACGGGCCGGCAGT
	ITPR1 ex10 sgRNA #7	TGTGGCCGACTGCCGGCCCG
	ITPR1 ex10 sgRNA #8	TGGCAGATGTGGCCGACTGC
Exon 58	ITPR1 ex58 sgRNA #1	AGAGGTAAATTAATCCCGAG
	ITPR1 ex58 sgRNA #2	AAGAGGTAAATTAATCCCGA
	ITPR1 ex58 sgRNA #3orig	ACTCAGGAAGCCGTCCAAAG
	ITPR1 ex58 sgRNA #3fix	ACTCAGGAAACCGTCCAAAG
	ITPR1 ex58 sgRNA #4	AAAGAGGTAAATTAATCCCG
Exon 59	ITPR1 ex59 sgRNA #0	TTATCTGTGGTGAGTGTCGC
	ITPR1 ex59 sgRNA #1	CTCACCACAGATAAAGCACG
	ITPR1 ex59 sgRNA #2	ATAACTCTAGCAGCAAACAG
	ITPR1 ex59 sgRNA #3	AAGACCACGTGCTTTATCTG
	ITPR1 ex59 sgRNA #7	TTGACACTTTTGCTGACCTG
DSS (Cassette)	Drosophila sgRNA	AAGCTCCTCTGCGTAATGGC

Table 3.1: sgRNAs tested in HEK293T cells. These 17 sgRNAs were cloned into eSpCas9-puro and transfected into HEK293T cells for validation. Guides ITPR1 Ex10 sgRNA #6 and #6opp are the reverse complement of each other. ITPR1 Exon 58 sgRNA #3 spans the single nucleotide polymorphism rs901854, which though annotated as a 'G' on reference human genome GRCh38.p13, was found to be heterozygous A/G in HEK293T cells and homozygous A/A in H9 hESCs. Therefore a new mismatch-corrected 'A' version was prepared (ITPR1 Exon 58 sgRNA #3fix), while the original version was relabelled as #3orig.

HEK293T cells were transfected with one of 16 different ITPR1 sgRNA-eSpCas9-puro constructs. Transfection was efficient and no significant toxicity was detected (Figure 3.7B). After 72hrs under puromycin selection, all wells transfected with eSpCas9-puro ITPR1 sgRNA plasmids were fully confluent (Figure 3.7C). In contrast untransfected or eGFP-transfected wells had no surviving cells (Figure 3.7D).

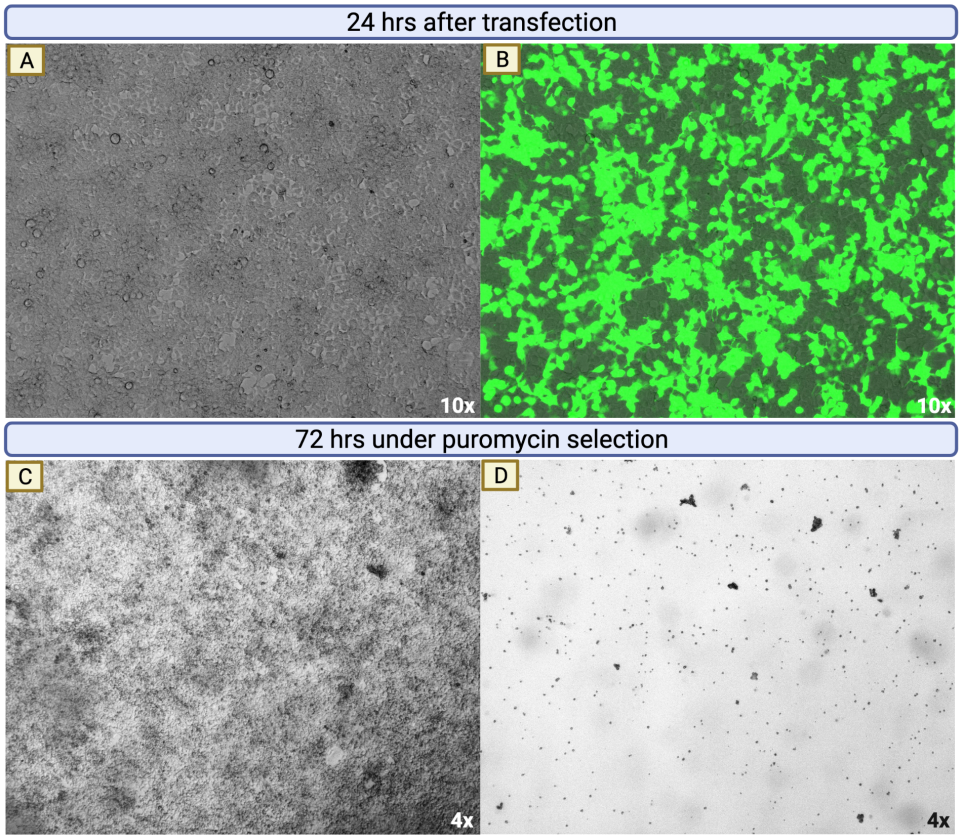

Figure 3.7: Validating ITPR1 sgRNAs using eSpCas9-puro in HEK293T cells. Control eGFP-transfected well 24 hrs transfection, immediately before the start of puromycin selection, shown under 10x magnification in bright-field (A) and fluorescence (B) microscopy showing >80% transfection efficiency. (C) Well transfected with eSpCas9-puro ITPR1 Exon 10 sgRNA #1 was fully confluent three days under puromycin selection. (D) Control eGFP-transfected well had no surviving cells after 3 days of puromycin selection.

Expected ITPR1 target sites were amplified from puromycin-resistant cells and Sanger-sequenced. Cutting efficiency results are shown in Figure 3.8. ITPR1 Exon 10 sgRNA #1 had the highest overall cutting efficiency, with indels detected in about 81% of the sequenced reads (Figure 3.8A). For Exons 58 and 59 the highest cutting efficiencies were between 53 and 55%, respectively, for ITPR1 Exon 59 sgRNA #2 (Fig. 3.8C) and Exon 58 sgRNA #3fix (Fig. 3.8B). Interestingly, despite a single base mismatch, Exon 58 sgRNA #3orig still resulted in 35% indel frequency, higher than the 11% achieved by the fully matched Exon 58 sgRNA #4 (see table

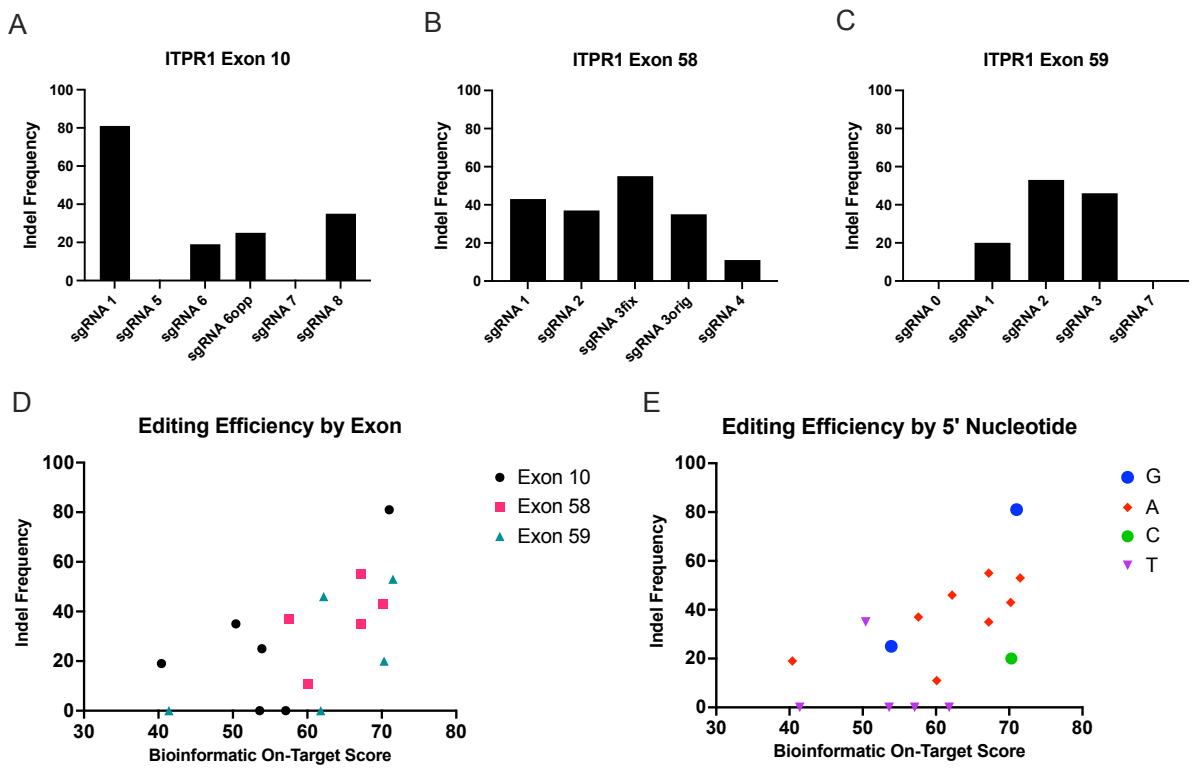

Figure 3.8: Editing efficiency of eSpCas9-puro ITPR1 sgRNA constructs in HEK293T cells. Indel frequency was calculated using Synthego's ICE software on Sanger sequencing chromatograms across the target site. All guides tested for ITPR1 Exon 10 (A), Exon 58 (B), and Exon 59 (C) are shown. Correlation between *in silico* predictions and experimental results are shown categorised by target exon (D) and 5' nucleotide of protospacer sequence (E). Bioinformatic on-target score was generated by Benchling's CRISPR Guide Design Tool using the optimised Doench score[137].

3.1 and fig. 3.8B).

Four sgRNAs (Exon 10 sgRNAs #5 and #7, and Exon 59 sgRNAs #0 and #7) did not lead to any detectable indel formation (<1%). A common denominator was that all four started with a thymidine (Fig. 3.8E), which is known to severely diminish transcription from the U6 promoter[152].

These results confirmed that guides ITPR1 Exon 10 sgRNA #1, Exon 58 sgRNA #3fix, and Exon 59 sgRNA #2 were the most effective at their intended targets. These three constructs were thus chosen to be used in subsequent experiments aiming at introducing ITPR1 mutations in human embryonic stem cells.

3.3.1.3 Construction of the Cassette-Targeting eSpCas9-sgRNA Vector

In CIRCES, the second round of CRISPR aims to generate a double-strand break in the integrated cassette itself. For that, a highly specific and effective Drosophila sgRNA is needed to target the Drosophila-Specific Sequence (DSS) site.

The Drosophila sgRNA (Table 3.1) was cloned into eSpCas9-puro and confirmed using Sanger sequencing. Validation of its cutting efficiency is described later in this chapter as it required components developed in the next section (see 3.3.3).

3.3.2 Construction of Homology-Directed Repair (HDR) Donor Templates

3.3.2.1 Assembly of the CIRCES Positive-Negative Selection Cassette

One of the goals of this system was to have an easily adaptable positive-negative selection cassette construct for which homology arms of any gene of interest could be readily added using standard and inexpensive molecular biology techniques. The assembly process involved three sequential rounds of cloning to achieve the final functional CIRCES cassette in a pCAGEN backbone (Figure 3.9).

3.3.2.2 Constructing Target and Mutation-Specific HDR Templates for Insertion and Replacement Stages

For each ITPR1 target region the flanking ∼800bp homology regions of both insertion and replacement stage HDR donor templates were identical, which greatly simplified the cloning process (Figure 3.10). A single intermediary plasmid containing wild-type target and flanking sequences was used to prepare both templates. H9 hESC gDNA sequences encompassing the target exons were cloned into pGEM vectors (Figure 3.10A-B). Sanger sequencing confirmed correct cloning and identified the strand, which provided valuable information for downstream experiments. These intermediary plasmids containing uninterrupted wild-type ITPR1 sequences (Figure 3.10B) were not transfected but were used in the cloning of both insertion (Figure 3.10D) and replacement (Figure 3.10E) HDR donor templates.

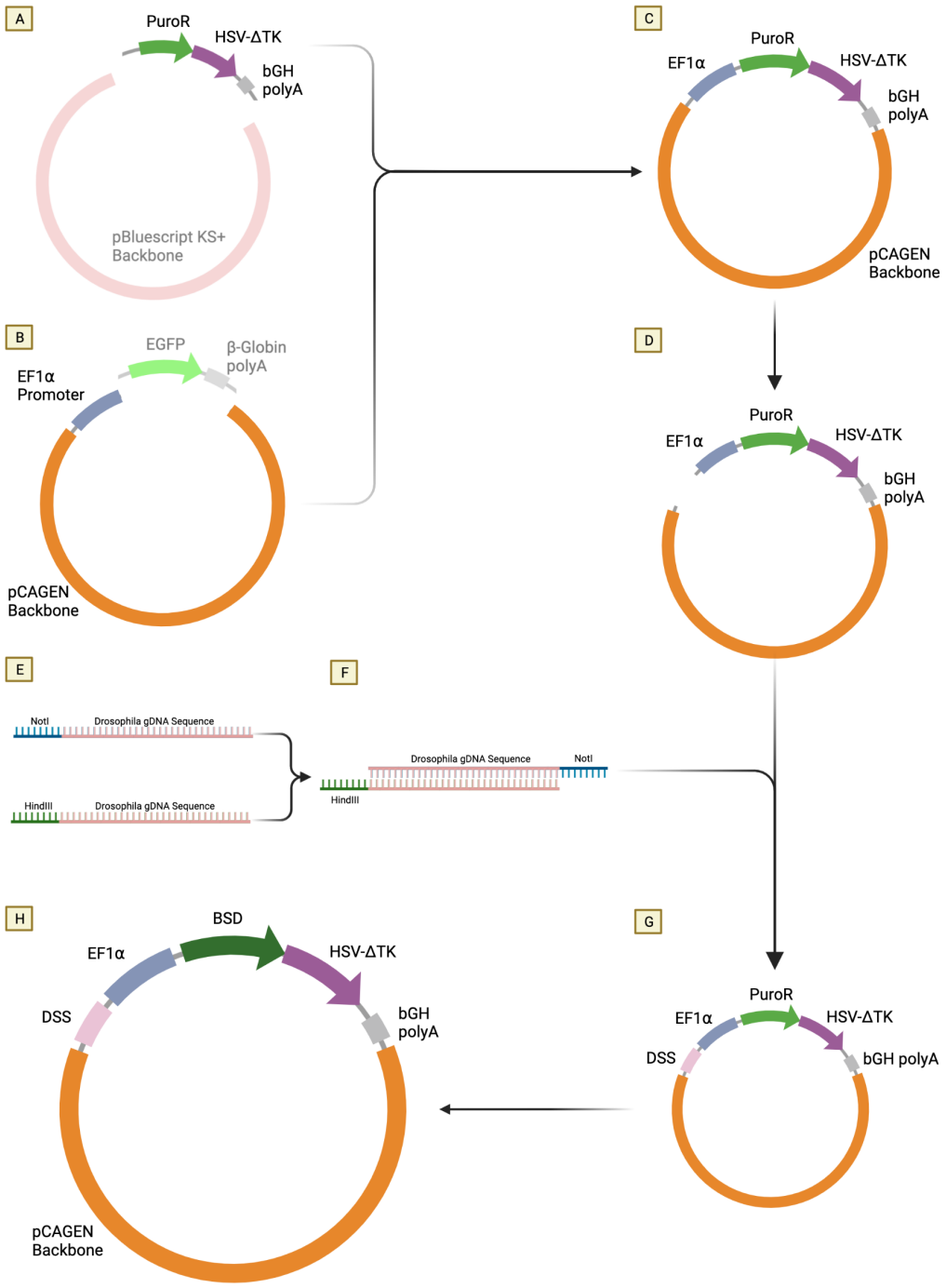

Figure 3.9: Assembly of the CIRCES Positive-Negative Selection Cassette. (A) The segment containing the PuroR-TK fusion protein and bGH polyadenylation signal was excised from the pLCA.66/2272 plasmid using restriction enzymes. It was ligated into (B) a linear PCR product containing the EF1α promoter and pCAGEN backbone from the pEF-EGFP plasmid. (C) The ligated EF1α-PuroR-TK plasmid was linearised (D) by inverse PCR using primers with overhangs compatible with NotI and HindIII. (E) Partially complimentary oligonucleotides containing the Drosophila genomic DNA target sequence (DSS) and NotI or HindIII overhangs were annealed (F) and then ligated into (D). (G) The DSS-EF1α-PuroR-TK plasmid was linearised by inverse PCR to exclude the PuroR gene, and ligated to another amplicon containing the BSD gene (from the pLV-TRET-hNgn1-UBC-Bla plasmid, not shown) to conclude the assembly of the CIRCES Cassette (H).

Figure 3.10: HDR Donor Template Cloning Strategy. (A) Using H9 hESC gDNA, a high-fidelity PCR reaction amplified ∼1800bp of uninterrupted genomic sequence encompassing the WT target site (ITPR1 Exon 10 is shown as an example). The forward and reverse primers in this reaction mark the external limits of the homology arms. (B) An intermediary plasmid was prepared by cloning the PCR product into a pGEM T-Easy vector. This was used as template in an inverse PCR reaction to generate a linear amplicon (inner orange semi-circle) containing the 5' homology arm (ITPR1 Intron 9 shown), the pGEM backbone, and the 3' homology arm (ITPR1 Intron 10 shown). This amplicon is ligated into the CIRCES cassette-containing amplicon, which was similarly PCR amplified (green inner semi-circle) from (C), the pCAGEN vector containing the assembled cassette. (D) The ligated amplicons form the first HDR Donor plasmid, which is used in the Insertion Stage. It is important to note that a small region encompassing the target site (here ITPR1 Exon 10, shown in red) is deliberately *excluded* from the first HDR template. This feature prevents Cas9 re-cutting after correct insertion. (E) The second HDR Donor plasmid, to be used in the Replacement Stage, contains the desired mutation to be introduced (ITPR1 Exon 10 mutation R269G shown as an example) and the immediately adjacent target context (here ITPR1 Exon 10, shown in red) that is not present in the first HDR donor plasmid. This replacement stage HDR donor template is cloned by simple site-directed mutagenesis to the intermediary wild-type pGEM plasmid (B).

3.3.2.2.1 Creating Insertion Stage HDR Templates Homology arms for each target were defined by PCR amplification in its respective intermediary pGEM plasmid using opposing primers, being careful to *exclude* the target region (e.g. ITPR1 Exon 10) from the resulting amplicon (Figure 3.10B, orange inner semi-circle). By ligating with an amplicon containing the CIRCES cassette (Figure 3.10C, green inner semi-circle), the three Insertion Stage HDR

template plasmids were cloned (ITPR1 Exon10-Cassette, ITPR1 Exon58-Cassette, and ITPR1 Exon59-Cassette, Figure 3.10D represents ITPR1 Exon 10-Cassette). Sequence integrity of the cassette and homology arms was confirmed by Sanger sequencing in all vectors.

3.3.2.2.2 Creating Replacement Stage HDR Templates Using each intermediary wild-type ITPR1 pGEM plasmid (e.g. ITPR1 Exon 10 WT pGEM, Figure 3.10B) as template, ITPR1 R269G, G2554R, and K2611del mutations were introduced using site-directed mutagenesis to create the Replacement Stage HDR template plasmids (Figure 3.10E represents ITPR1 Exon 10-R269G pGEM donor plasmid). Correct mutagenesis was confirmed using Sanger sequencing.

3.3.3 Validation of the Cassette-Targeting Cas9-gRNA Vector

Although the Drosophila sgRNA sequence had been reported to be effective at creating DSBs in Drosophila S2 cells (Dr. Joey Riepsaame, Genome Engineering Oxford, personal communication), it was important to validate its efficiency in the intended cassette and human genomic context, especially given the stringency requirements of eSpCas9. The target for this Drosophila sgRNA, the DSS, is not present in the human genome, therefore the CIRCES cassette had to first be inserted in genome of human cells to function as a reporter system before it could be used as target in a standard sgRNA validation experiment.

3.3.3.1 HEK293T CIRCES Reporter System

Creating a CIRCES cassette reporter system in HEK293T cells offered the first opportunity to experimentally test the Insertion Stage of CIRCES. ITPR1 Exon 10 was arbitrarily used as the target site for the insertion of the CIRCES cassette into HEK293T cells.

HEK293T cells were co-transfected with eSpCas9-puro ITPR1 Exon 10 sgRNA #1 plasmid (constructed and validated in section 3.3.1.2) and ITPR1 Exon 10-Cassette HDR template plasmid (constructed in section 3.3.2.2.1). A simultaneously performed control eGFP-only transfection indicated over 50% transfection efficiency. After 5 days under blasticidin selection, cells transfected only with a GFP construct died, whereas cells transfected with ITPR1 Exon

10-Cassette HDR template plasmid reached 80% confluency.

To identify at which point most transfected episomal plasmids had been lost or degraded, control cells had been co-transfected with a GFP construct and ITPR1 Exon 10-Cassette HDR template, but no sgRNA/eSpCas9 vectors. Post-transfection day 17, these cells, which displayed normal blasticidin resistance, had lost virtually all their GFP fluorescence, indicating loss of most episomal plasmids. From this point onward, surviving cells were assumed to be significantly enriched for on-target integration of the CIRCES cassette.

3.3.3.2 Targeting the CIRCES Cassette with a Drosophila-Specific sgRNA

After 27 days, surviving blasticidin-resistant cells were transfected with the Drosophila-sgRNA/eSpCas9-puro plasmid to assess for indel formation at the DSS in the CIRCES cassette. In parallel, control transfections were also carried out in blasticidin-resistant cells. For visual confirmation of transfection efficiency and functional confirmation of puromycin selection, two control wells were transfected only with an eGFP construct. Additionally, a third control well was transfected with a puromycin-resistant construct that does not target the DSS (eSpCas9-puro EMX1 sgRNA #1). Blasticidin selection was stopped on the day of transfection and puromycin selection started the following day. After 4 days, the puromycin-treated eGFP-transfected control well had virtually no surviving cells, while wells transfected with the Drosophila or EMX1 sgRNA plasmids were fully confluent, indicating successful puromycin selection.

To confirm the cassette had been correctly inserted into the ITPR1 Exon 10 locus, gDNA from surviving cells was used in a PCR reaction designed to amplify the cassette from its target locus using locus- and cassette-specific primers (Figure 3.11A).

Bands suggestive of on-target integration were identified in wells transfected with either EMX1 sgRNA (used as negative control for indels at DSS) or Drosophila sgRNA (Fig. 3.11C). Sanger sequencing revealed significant indel formation at the integrated DSS site in cells transfected with Drosophila sgRNA (40-46%) but not in cells transfected with EMX1 sgRNA (Figure 3.11B).

These results confirmed that the Drosophila sgRNA/eSpCas9 vector effectively targets the DSS

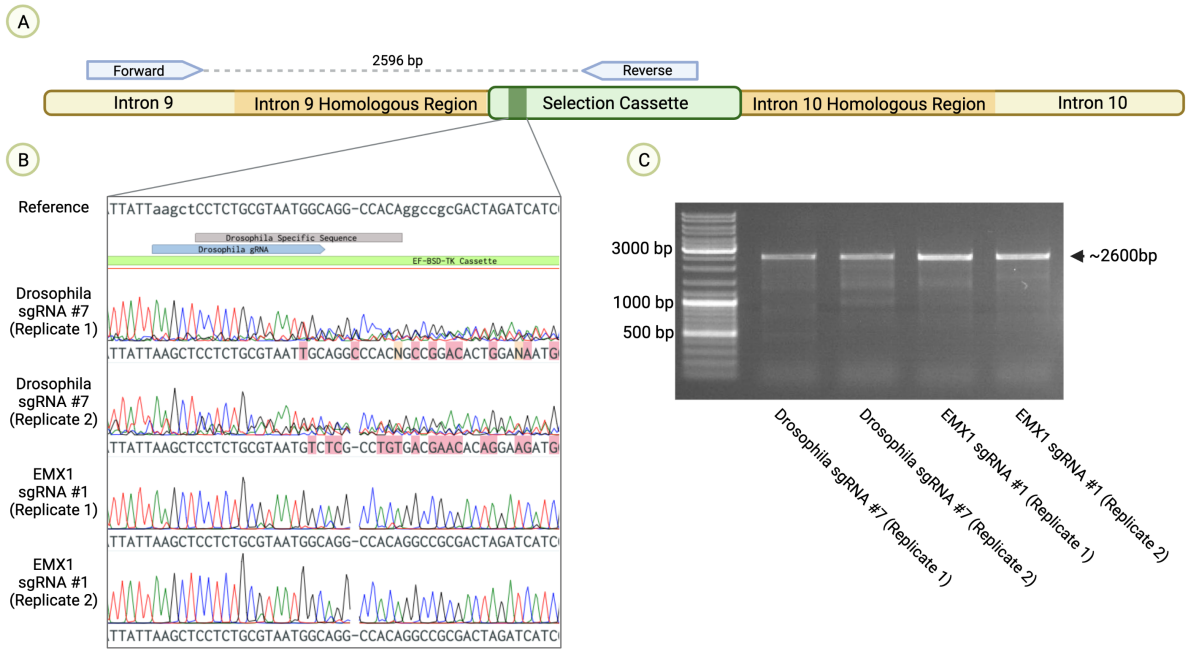

Figure 3.11: Assessment of Drosophila sgRNA Editing Efficiency in a HEK293T CIRCES Reporter System. (A) To avoid detecting randomly integrated copies of the cassette, PCR reactions used a forward primer external to the 5' homology arm and a reverse primer internal to the BSD gene in the cassette. The resulting 2596bp amplicon encompasses all of the 5' homology arm (intron 9), the DSS, the EF1α promoter, and most of the BSD gene. (B) Sanger sequencing chromatograms of the purified amplicons showed the expected unedited DSS sequence in control samples transfected with EMX1 sgRNA #1 eSpCas9 construct and significant indel formation in the samples transfected with Drosophila sgRNA (40% in Replicate 1 and 46% in Replicate 2). (C) PCR products showed that a $\sim$2600bp band could be clearly identified in all four samples, indicating correct Exon 10 insertion of the CIRCES cassette.

in its intended genomic context, thus validating its use for the second stage of CIRCES.

3.4 Discussion

The basic molecular tools required to edit the ITPR1 gene in human stem cells using the Cassette Insertion-Replacement CRISPR Editing System were generated and reported in this chapter. This system was designed so that some of these tools, such as the CIRCES cassette, the generic sgRNA/eSpCas9-puro plasmid and the Drosophila sgRNA/eSpCas9 plasmid, could also be used to target other genes with minimal cloning adaptations.

Each step in CIRCES requires two plasmids to be simultaneously transfected: one encoding both the Cas9 nuclease and the sgRNA and the other containing the HDR template.

The goal of the first step is to introduce a double-strand break at the desired genomic target while a custom repair template with the CIRCES cassette is available for HDR. An efficient sgRNA is required to direct the Cas9 nuclease to its target while minimising the risks of off-target effects. This was accomplished by using eSpCas9(1.1). Unlike traditional SpCas9, eSpCas9 is an extremely specific enzyme which requires virtually perfect complementarity between guide RNAs and their target sequences[130, 132]. Therefore it is reported to be intolerant of 21bp sgRNAs with an appended 5' mismatched guanine (5'GN_{20})[153], a common technique[150] used to improve U6 promoter-mediated transcription of 20bp guides that do not natively start with a guanine.

This extra stringency significantly reduces the number of possible optimal guide RNA designs (5'GN19) in a given target region. This was a concern, as especially for ITPR1 Exons 58 and 59 none of the prioritised sgRNAs natively started with a guanine (Table 3.1). However, while the validation experiments using eSpCas9-puro in HEK293T cells confirmed that little to no detectable cutting by 20bp guides starting with a thymine (5'TN_{19}), sgRNAs that started with an adenine were highly effective. These results corroborate the findings of a recent study analysing transcriptional activity of the U6 promoter that showed significant expression of RNAs with a 5' adenine[152].

Efficient and specific guides were identified for each ITPR1 target (Table 3.2) by experimentally validating 4-6 designs per target, a process that can be done in under three weeks from bioin-

formatic design of the guides to determination of cutting efficiency. While guide validation is ongoing, target-specific HDR donor templates can be prepared. Therefore, the preliminary cloning and validation steps prior to effectively transfecting cells can be completed in less than three weeks.

Target	Identification	Sequence (5'-3')	PAM	Editing Efficiency
ITPR1 Exon 10	ITPR1 ex10 sgRNA #1	GCACGTCTTCCTGAGAACCA	CGG	81%
ITPR1 Exon 58	ITPR1 ex58 sgRNA #3fix	ACTCAGGAAACCGTCCAAAG	AGG	55%
ITPR1 Exon 59	ITPR1 ex59 sgRNA #2	ATAACTCTAGCAGCAAACAG	GGG	53%
DSS	Drosophila sgRNA	AAGCTCCTCTGCGTAATGGC	AGG	40-46%

Table 3.2: Validated sgRNAs to be used in the following chapters.

Although the first assembly of the CIRCES cassette was a lengthy process that involved cloning parts from multiple different vectors, it does not need to be repeated. It now exists in a standard pCAGEN cloning vector that can be easily shared and used to prepare HDR templates for any target. To flank the cassette with target-specific homology arms, a simple and universal cloning strategy was developed and optimised (Fig. 3.10). From the outset CIRCES was planned to be easily adaptable to any target gene. Cloning multi-kilobase segments of DNA between plasmids can be very challenging for reasons that are not always readily apparent; therefore, multiple different cloning strategies were considered and tested. Although this thesis only reported on ITPR1, many other constructs were prepared, including SPTBN2, GRM1, EBF3, EMX1 and also other fluorescent-based cassette versions such as EGFP and mCherry. The most efficient way of achieving complete constructs was isothermal assembly with NEBuilder HiFi DNA Assembly kit. In almost all cases where it was used, it generated full constructs on first attempt, obviating the need for other cloning methods. It is thus recommended as the method of choice for anyone who wishes to adapt CIRCES to other target genes.

Using only these standard molecular biology techniques, all three double-stranded circular HDR donor templates were assembled in 3 weeks. Once the eSpCas9-puro/sgRNA plasmid is validated and the target-specific HDR donor plasmid is assembled, they can be transfected into wild-type cells for the first CRISPR round of CIRCES.

For the second CIRCES step, once all cells display stable Blasticidin resistance and have presumably integrated the cassette, the aim is to generate a double-strand break at the Drosophila-

specific sequence (DSS) within the cassette itself. Although the target genomic region may vary, the DSS remains the same. Thus, the same Drosophila eSpCas9/gRNA plasmid is used in the replacement stage regardless of the target. Therefore it was critical to have confirmation that this key component of CIRCES worked as intended. The validation of the Drosophila sgRNA offered the opportunity to show for the first time that the insertion stage worked as predicted by theory, as the non-human DSS had to be inserted into the human genome before it could be targeted using Drosophila sgRNA and eSpCas9 for the assessment of editing efficiency.

Although the goal of this thesis was to specifically target the ITPR1 gene in human stem cells, the CIRCES approach is designed to be used to target any gene in any cell line. Three essential components developed and tested in this chapter can be directly shared with other researchers: the 'empty' eSpCas9-T2A-puro vector (which has since been requested more than 58 times by the scientific community), the Drosophila sgRNA in eSpCas9-puro, and the CIRCES cassette in the pCAGEN vector backbone. Using simple molecular techniques and following the cloning strategies outlined in this chapter, it should be possible to go from target identification to transfection-ready CIRCES components in less than three weeks.

4 | Application of CIRCES in Human Stem Cells

4.1 Introduction

This chapter reports the application of CIRCES in human embryonic stem cell (hESCs). The aim was to target three different ITPR1 loci to generate cellular models that can be used to study SCA29 and Gillespie Syndrome.

For CIRCES to work as intended, components developed in the previous chapter had to be delivered and tested in hESCs. While delivering circular plasmids to immortalised cells such as HEK293T is not an issue, the same cannot be said about stem cells, which fall in the category of "very hard to transfect" cells[80]. The technology for expressing exogenous genetic material in human embryonic stem cells has greatly progressed since it was first achieved 20 years ago[80]. However, at the start of this project, it generally involved either viral vectors or electroporation/nucleofection[150]. These methods were considered not suitable for the repeated testing and optimisation of CIRCES. Nucleofection in particular requires expensive equipment and consumables. In our lab, it achieved highly variable, yet generally low, transfection efficiencies (data not shown). Perhaps more importantly, it significantly impacted cell viability. Thus, an easier, more effective, and less toxic method of simultaneously delivering multiple large plasmids into H9 cells was required. While a broader range of different methods and chemical reagents were considered, this chapter describes the establishment of a transfection procedure adapted to H9 hESCs using a polymeric reagent.

Although in the previous chapter Insertion Stage CIRCES components had been shown to work in HEK293T cells, practical issues remained about how best to apply effective blasticidin selection in H9 cells. Human stem cells are sensitive to transfected agents, so it was also important to show that delivery of CIRCES components was not toxic and could elicit blasticidin resistance. On a higher level, completeness of Insertion Stage positive selection relied on three assumptions: (1) effective blasticidin resistance is provided by overexpression of episomal blasticidin resistance gene (BSD) or by genomic integration of the BSD gene; (2) all transfected plasmids

are lost or degraded within a reasonable time frame; and (3) positive selection using blasticidin kills all cells not expressing a BSD resistance gene. A major uncertainty was the expected duration of transient blasticidin resistance due to episomal plasmid retention.

For the two-step CIRCES to work as intended, it was imperative that all cells undergoing the second CRISPR stage have at least one copy of the selection cassette integrated to their genome, either on- or off-target. Random off-target cassette integration events are expected[154], but ganciclovir-based negative selection[147, 148, 155] during Replacement Stage is designed to remove any cells that continue to express the cassette after the second CRISPR round. As it involves asymmetrically cleaving and replacing a large integrated cassette with a short $\sim$150bp target sequence containing the final mutation to be introduced, this second HDR step was expected to have very low efficiency. Thus a very high purity of cassette-integrated cells entering the Replacement Stage and high efficiency of negative selection were necessary factors to minimise the number of colonies that need to be screened at the end of the protocol. For example, if 100% of cells entering the Replacement Stage had integrated cassettes and the second CRISPR round was 100% effective at replacing the cassette with the target mutation, then no negative selection would be needed and every colony screened at the end of the protocol would be correctly edited, i.e. zero false discovery rate (FDR). On the other hand, even under 100% ganciclovir negative selection efficiency and a more realistic <0.01% second CRISPR HDR efficiency, if 1% of cells entering replacement stage did not have an integrated cassette (i.e. 99% positive selection efficiency), colony screening at the end of the protocol would identify 99 wild-type colonies for every correctly edited colony (99% FDR).

The design of the CIRCES two-step protocol does not require a clonal isolation step between Insertion and Replacement stages. If positive and negative selection are efficient, this intermediary genetic bottleneck is best avoided. However, during development of the protocol, it was necessary to rapidly identify and correct possible issues with cassette insertion efficiency, positive selection, promoter methylation, and — as this chapter will show — indels. The seemingly trivial, yet immensely important, question of when transfected plasmids are completely lost was difficult to answer. In addition to the tightly clustered way that hESCs grow and form colonies, the unknown frequency of random cassette integration events compounded the

difficulty in defining a specific time point at which blasticidin selection could be considered complete. For these reasons, as this chapter describes, individual cells were investigated at the end of Insertion Stage.

This required designing and testing a strategy for accurately identifying on-target cassette insertion and determining its zygosity. Random integration was expected to happen, albeit less frequently than on-target integration.

One of the main concerns at the start of this project was the frequency with which EF1α promoter would be silenced between CIRCES stages. Promoter methylation occurring between the end of Insertion Stage and the initiation of ganciclovir selection in Replacement Stage would make non-edited cells immune to negative selection, thus significantly increasing FDR. Progression to Replacement Stage was therefore conditional on experimental proof that Insertion Stage was able to generate a pool of cassette-integrated cells compatible with the overall goals and planned FDR of the proposed method.

This chapter describes how Insertion Stage was applied on H9 hESCs and was able to generate a large number of colonies carrying biallelic loss-of-function mutations that are useful for the study of ITPR1-related diseases. The following chapter will delve more deeply into the investigation of issues identified during Insertion Stage and methods proposed to correct them.

4.2 Materials and Methods

4.2.1 Chemical Reverse Transfection of H9 hESCs

The optimised Mirus TransIT LT1 transfection protocol is described in the main methods chapter. Table 4.1 lists the constructs and concentrations used for reverse chemical transfection in this chapter.

Well Label	eSpCas9-puro Plasmid	HDR Donor Template	Control Plasmid	Total DNA Transfected
Exon 10-Cassette	ITPR1 Exon 10 sgRNA #1	ITPR1 Exon 10-Cassette	-	1000ng
Exon 58-Cassette	ITPR1 Exon 58 sgRNA #3fix	ITPR1 Exon 58-Cassette	-	1000ng
Exon 59-Cassette	ITPR1 Exon 59 sgRNA #2	ITPR1 Exon 59-Cassette	-	1000ng
GFP	-	-	EF-GFP	1000ng
GFP/Cassette	-	ITPR1 Exon 10-Cassette	EF-GFP	1000ng

Table 4.1: Details of reverse transfections in H9 hESCs.

4.2.2 Blasticidin Selection of H9 hESCs

Blasticidin hydrochloride 10mg/mL stock solution (InvivoGen cat. anti-bl-05) was directly diluted into maintenance medium (mTeSR1) to achieve a final concentration of 5μg/mL unless stated otherwise. Blasticidin was replenished with every medium change, which was daily unless stated otherwise.

4.2.3 Strategy For Detection of On-Target Cassette Integration

Detection of on-target cassette integration relies on target-specific PCR primers. To avoid possible amplification of remaining contaminating or episomal plasmids, one of the primers binds to a flanking genomic region not present in (external to) the HDR donor plasmid, whilst the other anneals to an internal cassette-specific locus (the BSD gene). A correctly sized product generated by this 'External to Internal PCR' is suggestive of on-target integration of the cassette (Figure 4.1).

In itself, however, the 'External to Internal PCR' cannot distinguish between heterozygous and homozygous integration. For that, a second PCR specific to the non-integrated allele is performed with primers spanning the target exon. The results from this 'Screening PCR' are

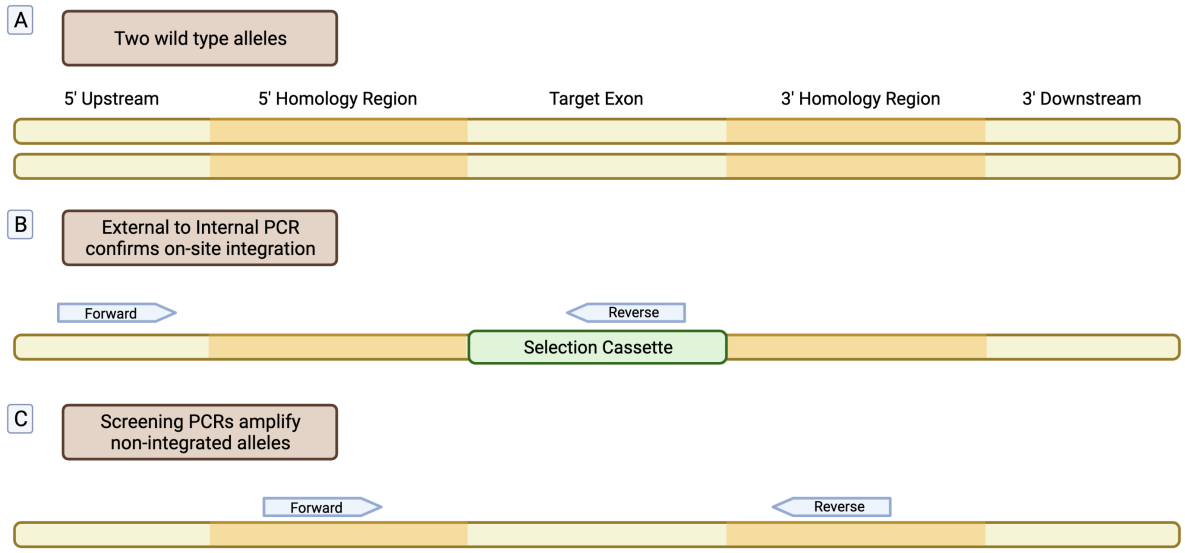

Figure 4.1: PCR strategy for detection of on-target cassette insertion. A) Representation of a target locus in a diploid genome such as H9 hESCs. B) External to Internal PCR reaction detects the presence of an on-target cassette by using a forward primer external to the region present in the 5' homology arm and a reverse primer in the Blasticidin-S-Deaminase (BSD) gene, which is internal to the cassette. This strategy prevents false-positive results due to amplification of cassette-containing HDR plasmids, which do not contain the binding site for the forward primer. Note this illustration is not to scale as the selection cassette is over 3kb in length. C) Screening PCR reaction detects alleles which do not have an inserted cassette. Its forward and reverse primers bind to the homologous regions adjacent to the target exon. Amplicon length can easily distinguish between the genomic non-integrated allele (500-600bp) and contaminating insertion stage HDR plasmid (>3000bp).

highly informative. If a wild-type-sized amplicon is generated, it indicates that at least one non-integrated allele is present. Therefore, if the 'External to Internal' reaction is also positive, the two results together indicate heterozygous (monoallelic) cassette integration. Alternatively, a large (>3kb) product in 'Screening PCR' would indicate either homozygous cassette-integration or presence of contaminating HDR plasmid. If no product is amplified at all, it could be due to insufficient quality/quantity of the gDNA sample or a too large product (e.g. genomic rearrangement, concatenations, etc.). Being a much smaller amplicon ($\sim$500bp from wild-type) and using a high-sensitivity polymerase kit, it is efficiently amplified even from poorly extracted, dirty, or dilute gDNA samples. Thus, if positive, it is a marker of sample adequacy that also aids in the interpretation of the 'External to Internal' PCR reaction.

4.2.4 Restriction Digest Screening of Exon 10 Colonies

gDNA was extracted from isolated colonies and used as template for Exon 10 Screening PCR reactions. 0.5μL of diluted extracted gDNA was used in a 10μL MyTaq HS Red Mix PCR re-

action. Following ExoSAP purification, a 2μL aliquot was diluted 10-fold and used for Sanger Sequencing (Azenta LifeSciences). The remaining were digested by directly adding 1μL of SspI 1:5 mix (4μL 10X SspI Buffer, 8μL SspI, 28μL nuclease-free water) and incubating at 37°C overnight. Digested reactions were run on agarose gels at 100V for one hour before being imaged.

4.2.5 Sequencing of Cassette Insertion Junctions

Left External to Internal PCRs were used to assess the 5' junction and Right External to Internal PCRs were used for the 3' junctions (Fig. 4.12A). PCR reactions were performed as described in the general methods section, electrophoresed on agarose gels and Sanger sequenced using internal primers.

4.3 Results

4.3.1 Delivery of CIRCES Components to hESCs

4.3.1.1 Optimisation of Chemical Transfection

The CIRCES protocol relies on efficient delivery of eSpCas9-sgRNA and HDR template plasmids to cells. Chemical transfection, although not commonly used in stem cell CRISPR literature, offers the possibility of significantly reducing transfection costs and hands-on time.

To investigate if chemical transfection could be used in H9 hESCs, the Mirus Trans-IT LT1 polymeric transfection reagent was used. H9 hESCs were transfected with an eGFP-expressing plasmid of similar size to CIRCES plasmid components.

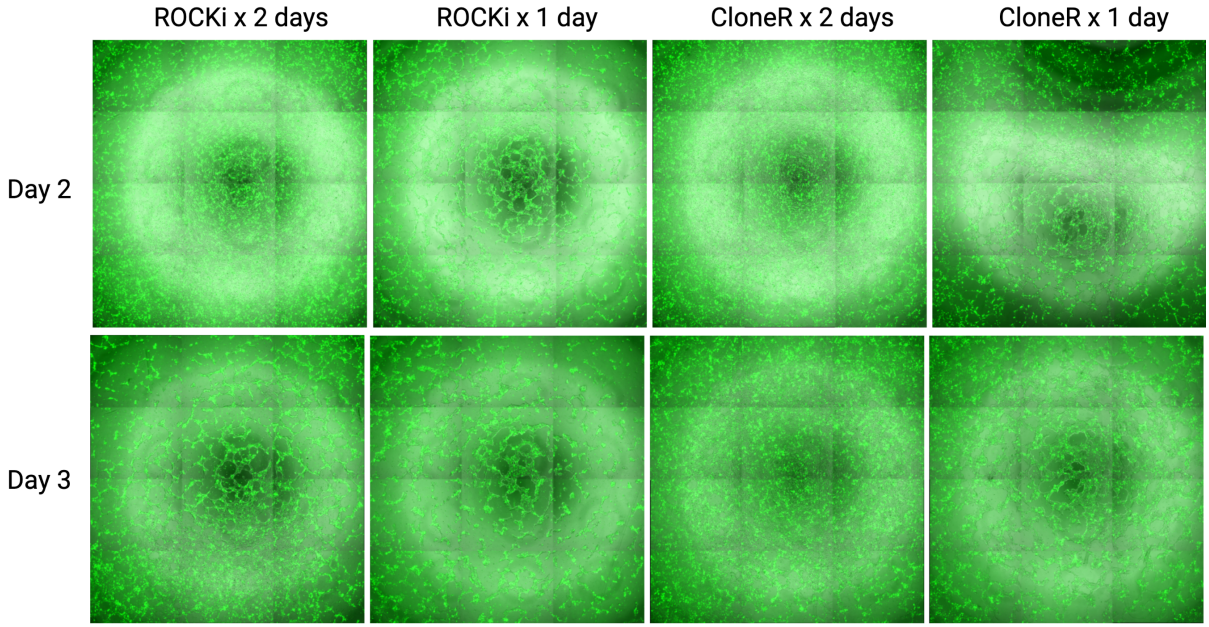

Figure 4.2: Chemical transfection of H9 hESCs. 3.5 x 10e5 H9 hESCs were chemically transfected with 1μg of a 7kb EF1α-eGFP construct per well of 24-well-plate and treated for 1 or 2 days with either 10μM ROCKi (Y-27632) or CloneR supplement (StemCell). Cultures were analysed for GFP fluorescence and cell density at 2 and 3 days post-transfection using fluorescence and bright-field microscopy. Experiment was performed in duplicate. Representative merged panorama bright-field and fluorescent images at 4x magnification are shown.

Three days post-transfection, most cells expressed GFP, indicating high transfection efficiency (Fig. 4.2). However, cell density was noted to decrease between the 1st and 3rd days post-transfection. As an attempt to improve post-transfection cell viability, transfected cells were cultured for either 1 (standard) or 2 days in the presence of Rho kinase inhibitors (ROCKi).

Alternatively, cells were instead cultured for 1 or 2 days in the presence of CloneR supplement (StemCell). Cell survival seemed to be improved by longer ROCKi treatment at post-transfection day 2 (PTD2) (Fig. 4.2 top), but at PTD3 the effects were indistinguishable from standard treatment. A 2-day treatment with CloneR, however, led to a significant improvement in viability at both PTD2 and PTD3 (Fig. 4.2, third column).

These results confirmed that circular dsDNA vectors can be reliably and effectively delivered to H9 hESCs using simple and inexpensive chemical transfection, obviating the need for nucleofection. It also showed that post-transfection viability can be improved by using the CloneR supplement.

4.3.1.2 CIRCES Insertion Stage Components Are Effective and Non-Toxic

The next step was to assess toxicity of Insertion Stage components. H9 hESCs were chemically co-transfected with the validated eSpCas9-sgRNAs and HDR template plasmids for each ITPR1 target (chapter 3).

At PTD2-3, no apparent differences in cell viability between control and experimental transfections were detected, suggesting that CIRCES components were not intrinsically toxic to H9 cells (data not shown). Blasticidin selection at 3μg/mL was started on PTD5. After 2 days, control cells (transfected only with a eGFP-expressing construct) had been killed, indicating effective blasticidin concentration and duration for positive selection. In contrast, cells transfected with Exon10-Cassette / eSpCas9-sgRNA, Exon58-Cassette / eSpCas9-sgRNA and Exon59-Cassette / eSpCas9-sgRNA were all at high densities (Figure 4.3). These results confirmed that chemical co-transfection of Insertion Stage CIRCES components is not toxic and can successfully elicit blasticidin resistance in H9 hESCs.

4.3.2 Stable Transformation by Prolonged Blasticidin Selection

Blasticidin-resistance in the first week post transfection is mostly due to overexpression from remaining episomal plasmids, and therefore expected to be transient. As plasmids are lost through replicative dilution and nuclease-mediated degradation[156], prolonged blasticidin se-

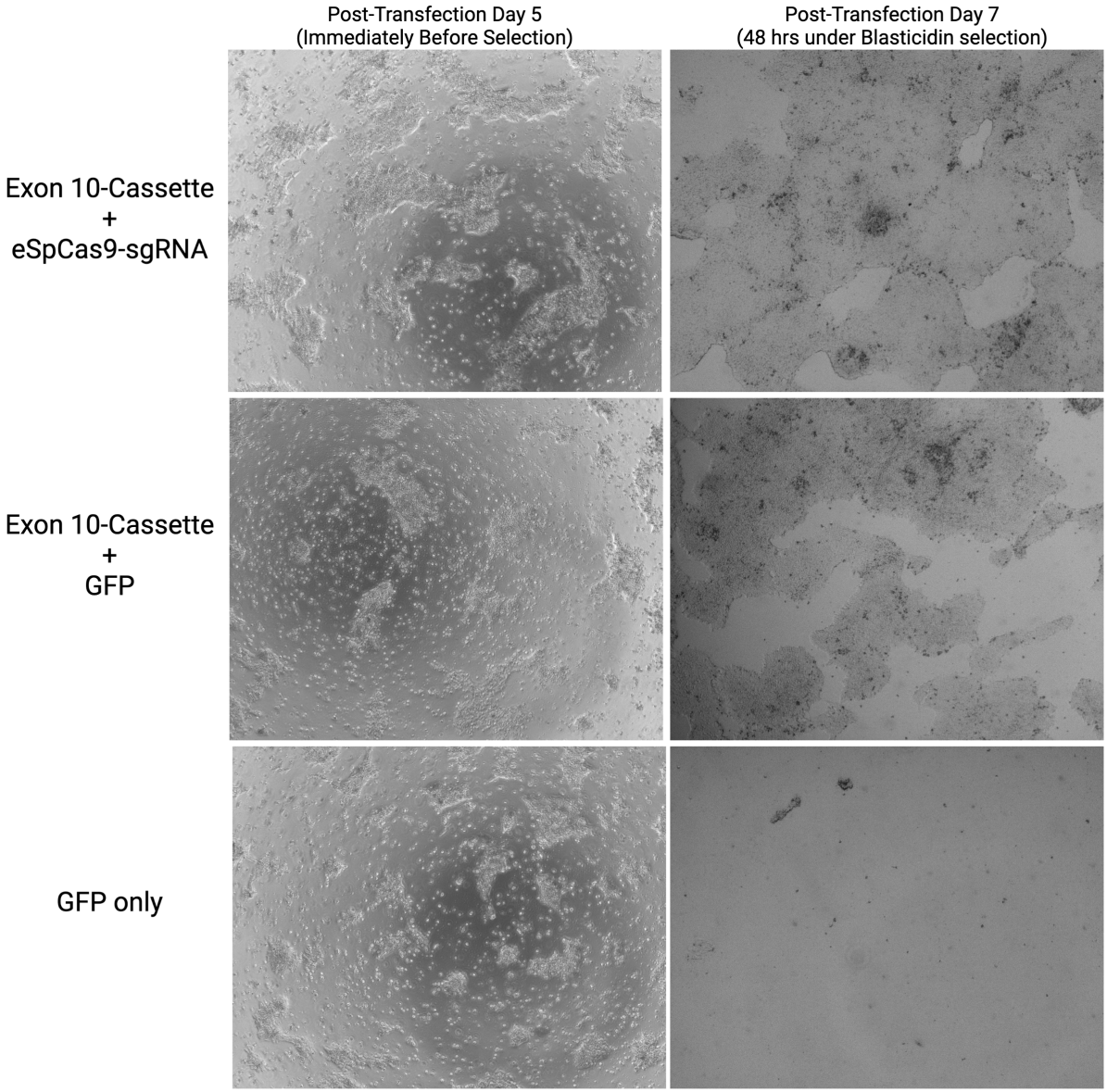

Figure 4.3: Blasticidin resistance in transfected hESCs. H9 cells transfected with exon-specific CIRCES insertion stage components survived two days of blasticidin 3μg/mL selection while all cells transfected only with a GFP plasmid were killed by blasticidin. Only images shown of Exon 10-cassette/eSpCas9-sgRNA are shown, but are representative of results with Exons 58 and 59. Bright-field images at 4X taken at PTD5 and PTD7.

lection and frequent passaging was thought to be necessary to adequately select for stable integration of the cassette.

On PTD7, surviving cells (in Exon 10, Exon 58 and Exon 59 targeting experiments) were single-cell dissociated and passaged to maintain the same number of cells over a 9-fold larger surface area while blasticidin selective pressure was maintained. Eleven days post-transfection, due to the COVID-19 lockdown, all experiments had to be stopped and cell cultures frozen. After several months, cells were re-thawed (counted as PTD12), and blasticidin selection was

resumed on PTD14.

At PTD20, cells no longer displayed signs of active selection and all wells had large coalescing colonies (Figure 4.4). As this point was beyond the expected duration of transient plasmid over-expression (7-10 days), surviving cells were considered to have undergone genomic integration of the cassette. This time-point was considered end of Insertion Stage.

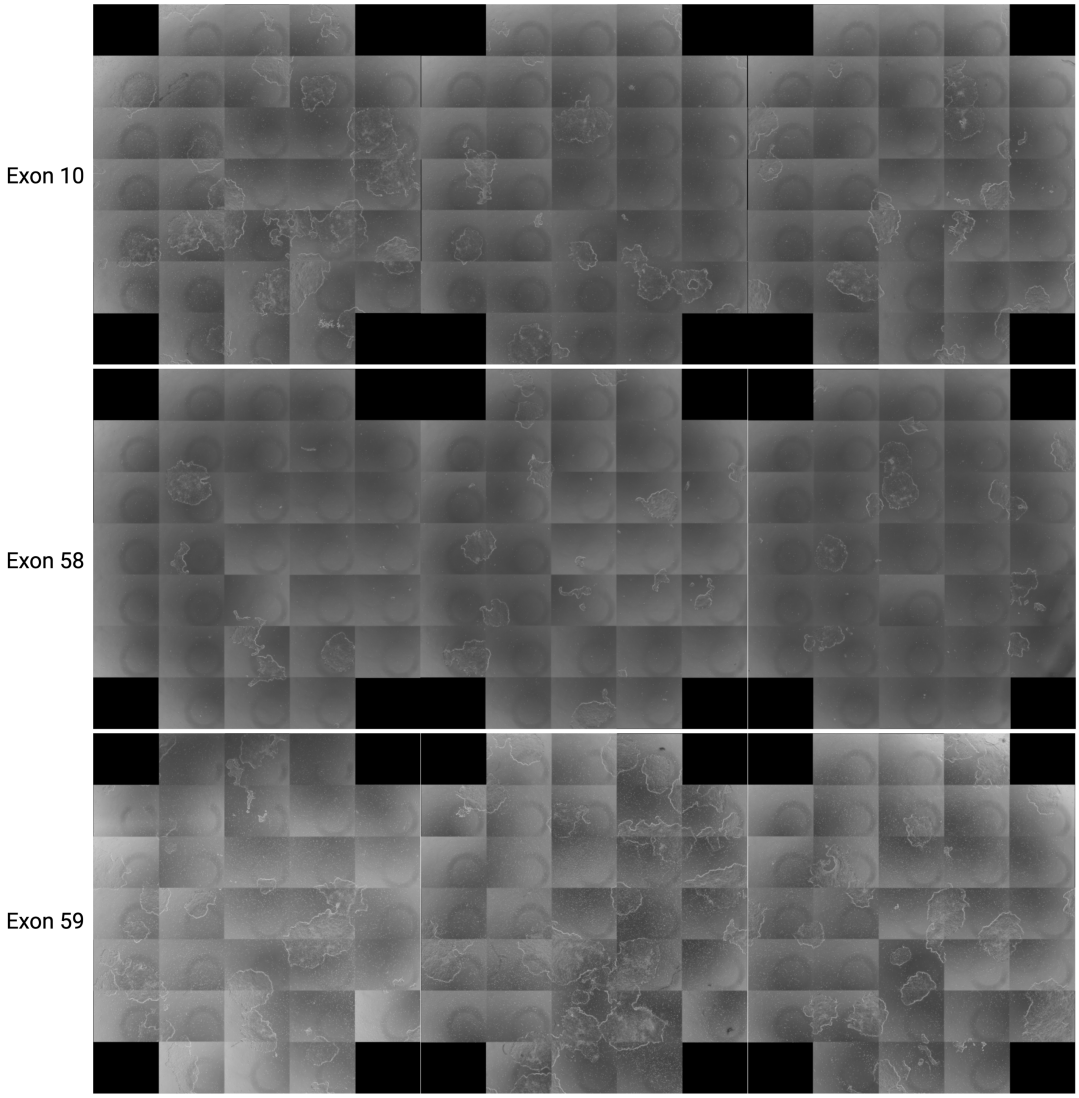

Figure 4.4: Coalescent clusters of blasticidin-selected H9 hESCs at PTD20. Each row is a composite image created from a panorama of 20% of the surface area imaged at 4x magnification.

4.3.3 Assessment of Insertion Stage Completeness

Although the CIRCES protocol was not designed to include an intermediary clonal isolation step between the insertion and replacement stages, it was important to confirm and assess rates

of on-target cassette integration at the end of the insertion stage. As previously discussed, for the two-stage CIRCES protocol to work as intended, virtually all cells advancing to Replacement Stage must contain an integrated cassette, either on- or off-target. Insertion stage completeness could be inferred by comparing clone survival and cassette integration rates under different selective conditions.

For that, insertion stage H9 hESCs were clonally isolated. At PTD24-26 (after 19-21 days under blasticidin selection) Exon 10, Exon 58 and Exon 59 cultures had all formed large clusters with dense centres. These wells were single-cell dissociated, re-seeded at clonal density and further characterised.

4.3.3.1 Blasticidin-Free Colony Screening

To assess rates of cassette insertion at the point of dissociation (representing the end of insertion stage), a set of dishes containing single cell-derived colonies was grown under blasticidin-free conditions. After 10-12 days (34-38 days after initial transfection), numerous large colonies could be identified in all dishes. For each target exon, 16 colonies were analysed using the Screening and External-to-Internal PCRs (see methods section).

Results suggestive of on-target cassette integration were identified in two Exon 59 colonies (Fig. 4.5F, colonies 03 and 08). In the 32 colonies isolated from Exon 10 and Exon 58 transfections, no insertion was identified (Figs. 4.5B and 4.5D). All colonies tested also generated WT-sized amplicons on Screening PCR (Figs. 4.5A,C,E), confirming presence of at least one non-integrated target allele. Analysed in conjunction, these PCR results were suggestive for monoallelic (heterozygous) Exon 59 cassette insertion events in two different clones. Preliminary Sanger sequencing of the PCR product in colony 03 (Exon 59) confirmed integration in the correct locus.

The detectable insertion rate was lower than expected (2/48 colonies) given all clones had survived a long course of positive selection. Nevertheless, confirmation of on-target integration of the cassette in H9 hESCs was important proof of concept, showing that the approach was able to insert a large 3kb cassette at a predetermined genomic locus in human embryonic stem

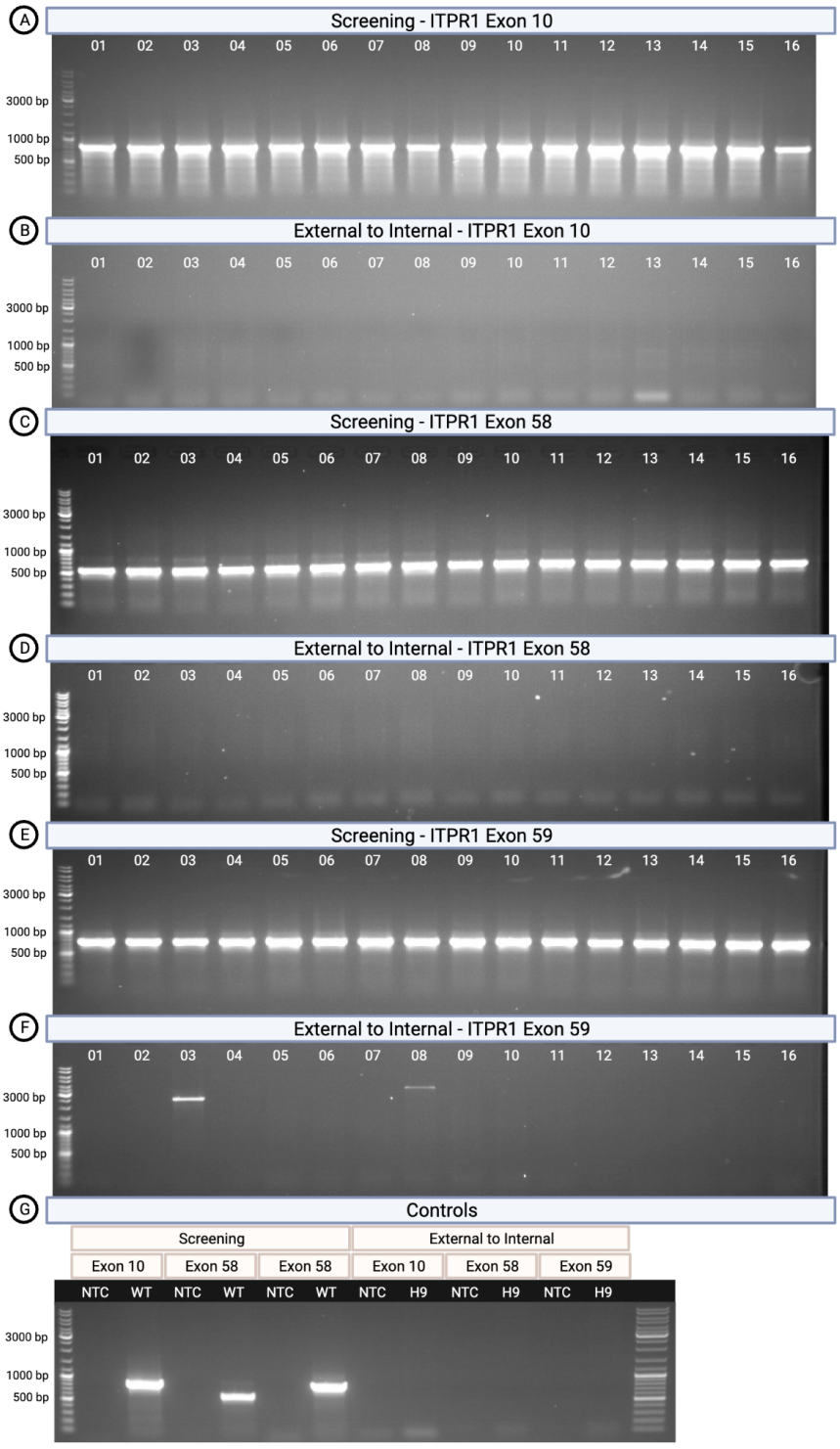

Figure 4.5: PCR-based investigation of single-cell clones established at the proposed end of insertion stage (PTD 24-26). To reflect genotype frequencies found at PTD24-26, positive selection with blasticidin was not applied during the clonal isolation procedure. For each target, 16 colonies were manually picked and analysed using two PCR reactions: screening PCR (can only amplify non-integrated alleles) and external-to-internal PCR (can only amplify integrated alleles). Using this two-PCR colony investigation approach, cassette integration was only detected in two Exon 59 colonies (cols. 03 and 08 in [F]). Assuming a diploid genome and single-cell origin of each colony, monoallelic vs biallelic cassette integration can be inferred by the presence or absence of a band in the screening reaction (positive band in cols. 03 and 08 in [E]) indicate presence of non-integrated allele, therefore both colonies are likely to have a monoallelic cassette insertion in Exon 59).

cells.

4.3.3.2 Blasticidin-Selective Colony Screening

Hypothesising that the low on-target insertion rate was partially due to incomplete blasticidin selection, a mirror set of dishes with PTD24-26 clonally isolated cells were grown under continued selection with blasticidin. After 10-12 days, despite identical seeding density and origin as the previous experiment, very few colonies survived. Of the few that did, not all survived passaging to 96-well-plates for expansion. These findings indicated greater than expected sensitivity to blasticidin selection after clonal isolation and suggested the blasticidin selection prior to clonal isolation had not been fully effective.

Results from Screening PCR and External-to-Internal PCR for all colonies are shown in figure 4.6. Out of eight colonies picked from Exon 58 dishes, seven survived passaging and six has results suggestive of on-target cassette integration. Exon 58 colony 02 showed multiple band sizes on both Screening and External-to-Internal PCR reactions, suggestive of multiple genotypes in the same colony (i.e. mixed colony). The other five cassette-integrated Exon 58 colonies likely represented monoallelic integrations, given the coincident detection of non-integrated alleles (presumably wild-type).

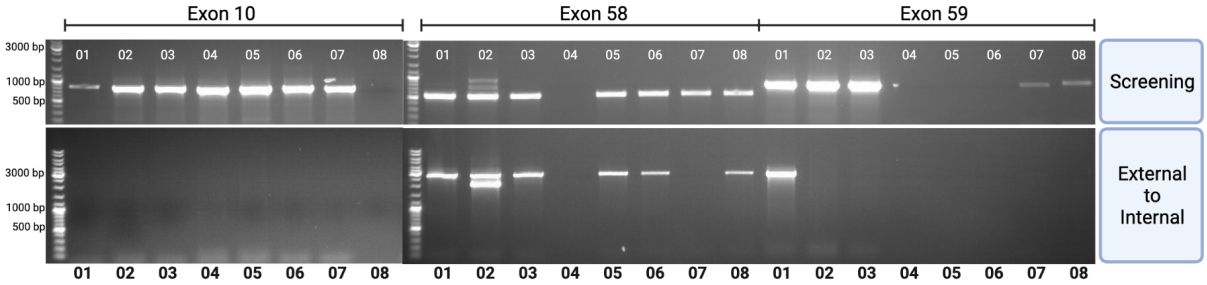

Figure 4.6: PCR investigation of colonies grown under blasticidin-selective conditions. Cells at the proposed end of insertion stage (PTD 24-26) were clonally isolated and grown under continued blasticidin selection. For each target, 8 colonies were manually picked and analysed using the two PCR screening approach. Not all colonies survived or generated sufficient gDNA for analysis (colonies lacking visible band on the screening PCR [top row], e.g. Exon 10 col. 08). Cassette insertion can be identified in Exon 58 cols. 01, 02, 03, 05, 06, 08 and in Exon 59 col. 01 and 07 (barely visible). Multiple bands in seen in Exon 58 col. 02 are indicative of it being a mixed clone.

For Exon 59, 6 out of 8 colonies picked survived passaging. Of these, 5 generated WT-sized bands on screening PCR, but two were quite faint, indicating poor sample quality/quantity. Initial analysis suggested on-target integration occurred in only one colony, but re-evaluation

of gel images revealed a faint band in a second colony (Figure 4.6, Exon 59, colonies 01 and 07, respectively).

Exon 10 proved more challenging. Although 7 out of 8 picked colonies survived passaging, all had bands consistent with non-integrated alleles (Figure 4.6A) and none had detectable on-target cassette integration (Figure 4.6B).

Overall, blasticidin-selective clonal isolation achieved a higher rate of detectable on-target cassette insertion and led to the identification of several cassette-integrated Exon 58 colonies, as well as other Exon 59 colonies. The lower overall survival but higher rate of cassette integration in colonies grown under continuing positive selection was indicative of a high number of presumably wild-type cells having unexpectedly reached the end of Insertion Stage. This issue would need to be addressed before cells could be progressed on to the replacement stage, otherwise the false-discovery rate (FDR) at the end of the protocol would be unacceptably high.

4.3.4 Detailed Investigation of Exon 10 Colonies

Analysis of clonally-isolated colonies showed that on-target cassette integration could be identified in two of the three ITPR1 target exons by the proposed end of Insertion Stage. In addition, continued blasticidin selection during clonal isolation markedly improved the rate of identification of cassette-integrated colonies. This raised the possibility that more stringent blasticidin selection could better remove non-integrated cells, thereby improving the cassette-integration rate. As no successfully cassette-integrated Exon 10 colonies had yet been identified, this hypothesis was tested by subjecting PTD12 blasticidin-selected (3.0μg/mL) cells to further selection at either medium dose (5.0μg/mL) or high dose (10.0μg/mL) blasticidin for 10 additional days before clonal isolation. Importantly, selection was maintained also during the clonal stage.

By PTD32 only eight colonies had survived under each selective condition (colonies 01-08 in figure 4.7A and colonies 01-08 in figure 4.7B). Eight additional colonies from each blasticidin concentration were isolated from backup dishes which had been seeded at 10 times clonal density (Figure 4.7, colonies 09-16). Despite 31/32 colonies having acceptable gDNA quality/quantity for analysis, none were found to have PCR results compatible with on-target

cassette insertion (Fig. 4.7A and 4.7B). These results suggested that more stringent blasticidin selection during Insertion Stage was not effective at isolating a clone with detectable cassette at the ITPR1 Exon 10 locus.

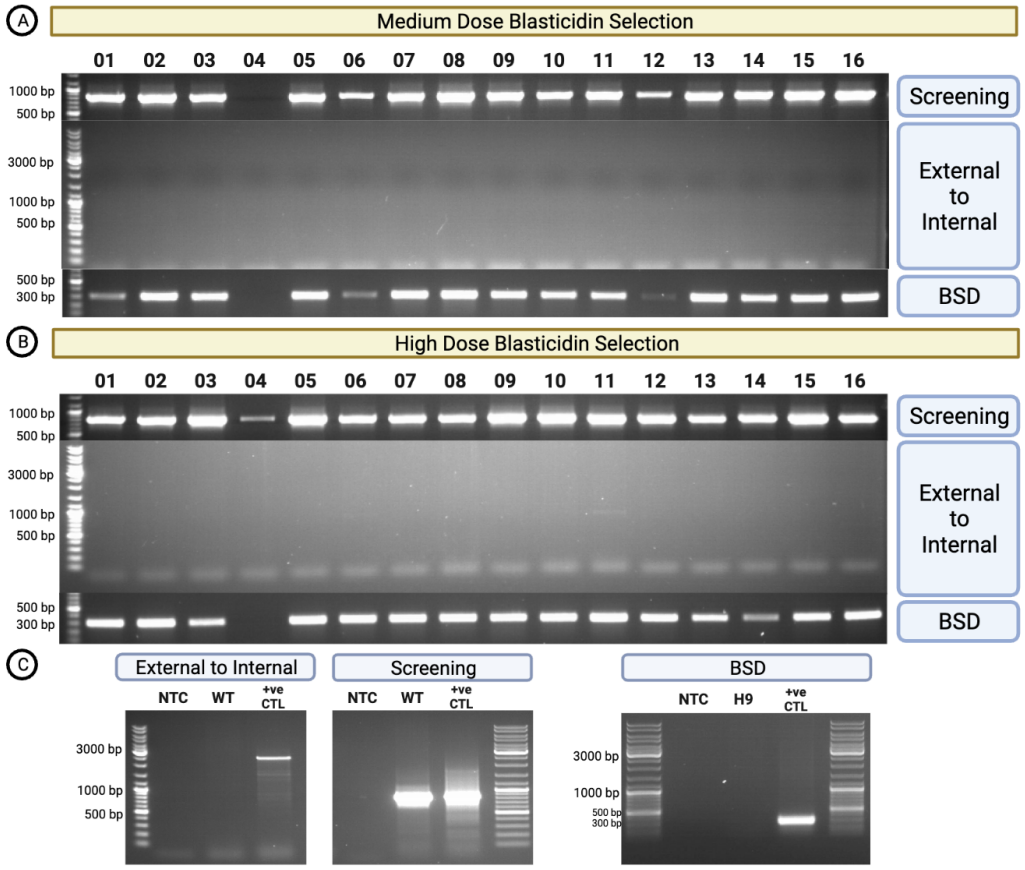

Figure 4.7: Investigation of Exon 10 colonies grown under different blasticidin concentrations. H9 hESCs transfected with Exon 10 sgRNA/eSpCas9 and Exon 10-Cassette HDR template were selected with a low standard Blasticidin concentration (3.0μg/mL) for 11 days. Concentration was then increased to either 5.0μg/mL (medium dose) or 10.0μg/mL (high dose) and cells were cultured for an additional 10 days prior to being single-cell dissociated and clonally isolated. Blasticidin selection was maintained during clonal isolation. At PTD32 16 colonies from each condition were analysed using the two-PCR screening strategy in addition to a PCR amplifying the BSD gene. Results for these three PCR reactions are shown for medium dose-selected colonies in (A) and for high dose-selected colonies in (B). Controls for all three PCR reactions are shown in (C). NTC = No template control. HEK293T CIRCES Reporter cells (exon 10) prepared in the previous chapter were used as positive control for Exon 10 External to Internal PCR. ITPR1 Exon 59 clone #8 were used as positive control for BSD PCR reaction.

4.3.4.1 Blasticidin-S-Deaminase Genes Are Found in Blasticidin-Resistant Colonies

For CIRCES to work as expected, it is required that at least a small subset of cells have inserted the cassette on target. While for Exon 57 and Exon 58 this was confirmed, for Exon 10 it was not. It was important therefore to investigate the causes for the apparent failure of on-target cassette insertion.

The previous experiments showed a remarkable ability of H9 hESCs to withstand high doses of blasticidin for prolonged periods despite not having detectable on-target integration of the CIRCES cassette. Possible causes included ineffective blasticidin selection, prolonged retention of episomal plasmid, high-rates of random cassette integration, or PCR issues.

To investigate the underlying mechanism of blasticidin resistance, a PCR was designed to amplify only the Blasticidin-S-Deaminase (BSD) gene present in the cassette. Given its small size (300bp), the BSD-specific PCR reaction should be highly efficient even in low quality DNA samples, unlike the large (>2.5kb) Exon 10 External-to-Internal reaction. A positive band in the BSD-specific PCR reaction cannot, on its own, distinguish between on-target vs. off-target integration vs. episomal plasmid. But given the significant time elapsed since transfection, a BSD+ result in conjunction with a negative External-to-Internal PCR would be highly suggestive of either off-target or undetectable on-target integration of the cassette. In contrast, the absence of a BSD band would be strongly suggestive of ineffective blasticidin selection.

All but two of the 32 colonies from the previous experiment were found to carry the BSD gene (Fig. 4.7C and 4.7F). One BSD^{-} colony had not generated a Screening amplicon while the other had a faint band on Screening PCR (colony 04 in figure 4.7A and colony 04 in figure 4.7B, respectively), pointing to less-than-ideal sample quality/quantity that could have affected accurate detection of single-copy integration.

These results indicated that almost all tested Exon 10 colonies carried at least one copy of the blasticidin resistance gene in their genomes. It was therefore unlikely that the low rate of cassette integration in Exon 10 colonies could be ascribed mainly to ineffective blasticidin selection in this experiment.

4.3.4.2 Some Blasticidin-Resistant Colonies Likely Have Undetected On-Target Cassette Integration

Although the BSD gene detection experiment suggested that off-target integration could be remarkably frequent, it was also technically possible that the Exon 10 External-to-Internal PCR reaction was inefficient despite having been validated previously (see appendix). Alternatively,

partial cassette duplications/concatenations or imprecise HDR accompanied by large deletions or more complex rearrangements could lead to false negative results, either due to excessive separation between primers or due to deletion of primer-binding sites[122, 123, 157–159].

To investigate these possibilities, all 32 Exon 10 colonies were assessed for possible loss of heterozygosity (LOH) in the Exon 10 locus. Wild-type H9 hESCs are heterozygous for the rs41289624 (A/T) SNP in ITPR1 intron 10, which is present in the 3' homology arm of the respective HDR template and is amplified as part of Exon 10 Screening PCR (Fig. 4.8A,B).

The restriction enzyme SspI can be used to selectively cleave only amplicons containing the T allele at rs41289624, while leaving the A allele intact (Fig. 4.8B,C,D). Conditional on the single-cell origin of colonies, by analysing gel electrophoresis results of SspI-digested Screening PCR amplicons, it is possible to infer whether the amplicon contains two non-integrated alleles (heterozygous A/T genotype) or only one (either A or T) (Fig. 4.8C). Identification of heterozygous A/T genotypes would strongly suggest absence of Exon 10 cassette integration.

Surprisingly, 8 colonies had results compatible with LOH at the ITPR1 locus (Fig. 4.8E,F). These findings were confirmed by Sanger sequencing of the undigested screening amplicons, which showed the 8 LOH colonies to be pseudo-homozygous for the A allele while the remainder were heterozygous (Fig. 4.8H).

Although these results were not proof of on-target cassette integration at the Exon 10 locus, they suggested that the PCR-based screening strategy might not be detecting all on-target integration events and that some on-target integration of the cassette may have occurred with additional large mutations in *cis*.

4.3.4.3 High Rate of Indels in Blasticidin-Resistant Exon 10 Colonies

Sanger sequencing reads used to confirm LOH also spanned the expected eSpCas9/sgRNA Exon 10 cut site. Unexpectedly it revealed that all eight LOH colonies also carried a pseudo-homozygous 16bp deletion (NM_001378452.1:c.789_804del p.Leu264Glyfs*50) affecting the PAM and sgRNA binding sites (Fig. 4.8G).

The remaining 24 Exon 10 colonies were subsequently investigated by Sanger sequencing. Re-

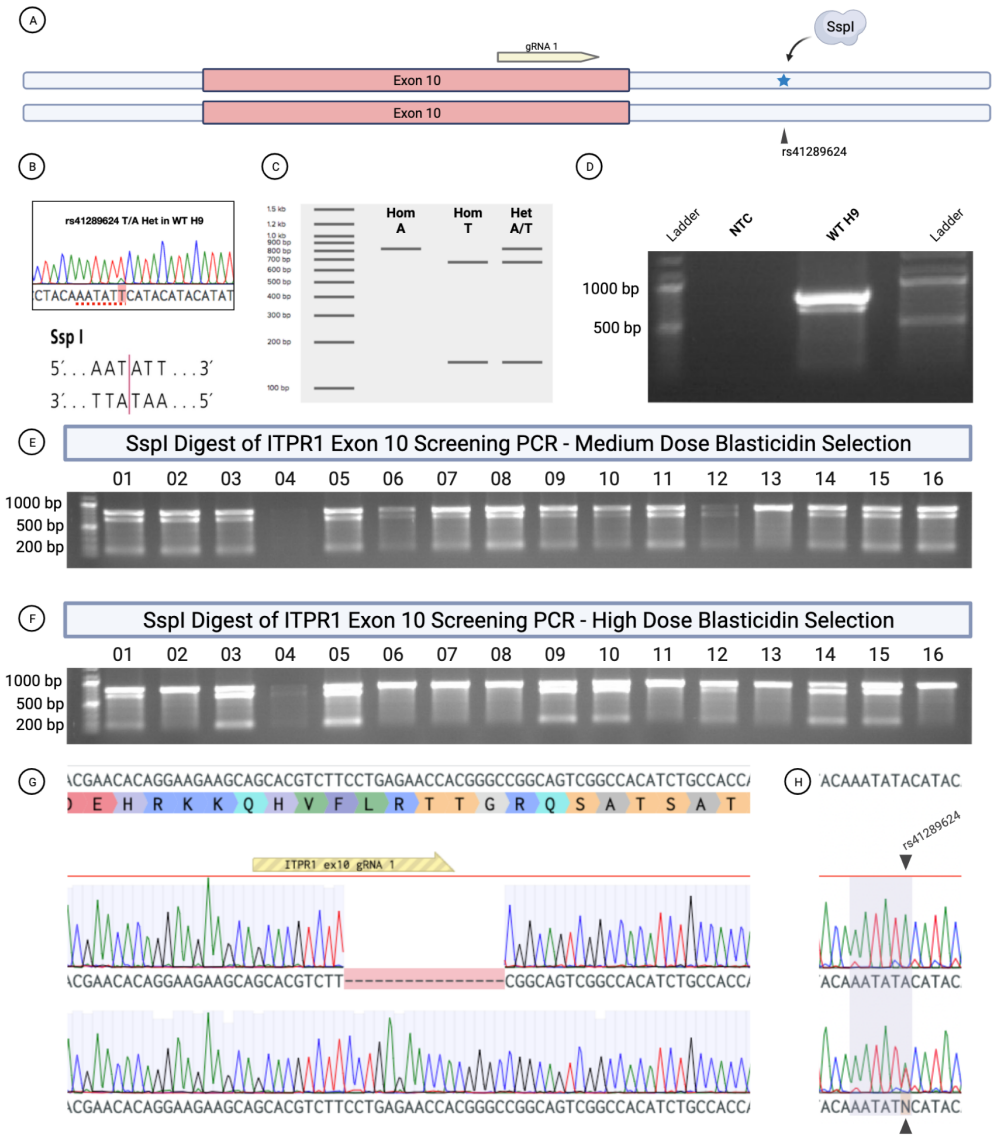

Figure 4.8: Investigation of Exon 10 colonies for loss of heterozygosity at ITPR1 Intron 10. (A) The SNP rs41289624 is located within intron 10 region of homology used in the construction of HDR templates. (B) WT H9 hESCs are heterozygous for rs41289624 with a genotype T/A (note small secondary peak in the chromatogram), which introduces a SspI recognition site in the T but not in the A allele, while the 3' homology arm in the Exon 10 HDR template contains only the A allele (not shown). (C) Different SspI digestion patterns are expected in PCR products containing either or both of the alleles. (D) Screening PCR amplicon from WT H9 hESCs digested with SspI showing the double banding pattern between 500-1000bp that is indicative of heterozygosity at rs41289624. (E) SspI digestion of 16 single cell-derived colonies selected at 5.0μg/mL (medium dose) blasticidin. Colony 04 did not produce a screening amplicon, likely due to insufficient or impure gDNA. Colony 13 shows a single high band between 500-1000bp, indicating all amplified products have the A allele at rs41289624. All other colonies show the double banding pattern indicative of heterozygosity. (F) SspI digestion of 16 single cell-derived colonies selected at 10.0μg/mL (high dose) blasticidin. Colonies 02, 06, 07, 08, 11, 13 and 16 produced a homozygous A/A digestion pattern. Colony 04 produced a very faint screening amplicon, but a double banding pattern can be seen. All other colonies had digestion patterns consistent with heterozygosity. (G) All colonies showing homozygous A/A alleles at rs41289624 also had a 16bp deletion spanning in the Exon 10 sgRNA #1 protospacer and PAM sequences (reference genomic sequence and aminoacid translation at the top, sgRNA protospacer sequence represented as a yellow arrow with its head directed to the 5'-CGG-3' PAM). Sanger sequencing chromatogram depicting the 16bp deletion in the middle row and WT control in the bottom row. (H) Reference genomic sequence (top) and sequencing chromatograms of colonies with homozygous A/A genotype at rs41289624 (centre) and heterozygous T/A (bottom). NTC = No template control.

markably at least 16 of the 24 colonies were heterozygous for the same 16bp deletion, while 5 colonies were homozygous wild-type (confirmed by a heterozygous rs41289624 allele). Two colonies showed multiple sequences, consistent with mixed origin, but also contained identifiable indel sequences.

Taken together, out of the 31 Exon 10 colonies which could be sequenced, 26 contained indels (table 4.2 contains the sequencing results of all 31 colonies). Since none of the Exon 10 colonies had produced PCR results compatible with on-target cassette insertion, they had been assumed to represent mostly wild-type genotypes. But these results showed that over 80% of colonies in fact carried NHEJ-mediated mutations in the target region.

Medium Dose Blasticidin Selected Colonies					**High Dose Blasticidin Selected Colonies**				
Colony	**Allele #1**	**Allele #2**	**rs41289624**	**Comments**	**Colony**	**Allele #1**	**Allele #2**	**rs41289624**	**Comments**
1	WT	c.789_804del	Het		1	WT	c.789_804del	Het	
2	WT	c.789_804del	Het		2	-	c.789_804del	Hom	LOH
3	WT	c.789_804del	Het		3	WT	WT	Het	
4	-	-	-	Failed sequencing	4	WT	c.789_804del	Het	
5	Indel	Indel	-	Mixed colony	5	WT	WT	Het	
6	WT	c.789_804del	Het		6	-	c.789_804del	Hom	LOH
7	WT	c.789_804del	Het		7	-	c.789_804del	Hom	LOH
8	WT	c.789_804del	Het		8	-	c.789_804del	Hom	LOH
9	WT	c.789_804del	Het		9	WT	WT	Het	
10	WT	c.789_804del	Het		10	WT	c.789_804del	Het	
11	WT	c.789_804del	Het		11	-	c.789_804del	Hom	LOH
12	WT	indel	-	Mixed colony	12	WT	c.789_804del	Het	
13	-	c.789_804del	Hom	LOH	13	-	c.789_804del	Hom	LOH
14	WT	c.789_804del	Het		14	WT	WT	Het	
15	WT	c.789_804del	Het		15	WT	WT	Het	
16	WT	c.789_804del	Het		16	-	c.789_804del	Hom	LOH

Table 4.2: Sanger sequencing results from 32 Exon 10 colonies selected at either medium or high blasticidin concentrations. The same screening PCR amplicon contained both the expected cut site for ITPR1 Exon 10 sgRNA #1 and intronic SNP rs41289624. Results for column rs41289624 are derived from Sanger sequencing. LOH = loss of heterozygosity as identified by SspI digestion (from figs. 4.8E and 4.8F). Het = heterozygous, Hom = homozygous, WT = wild-type.

4.3.5 Identification of Exon 10 On-Target Cassette Insertion

Repeated detection of the same indel (NM_001378452.1:c.789_804del p.Leu264Glyfs*50) in more than 80% of Exon 10 colonies raised the possibility that these were daughter cells from one or few dominant clones. If that was the case, it was possible that correct on-target insertions existed but had not been detected merely because too few colonies were analysed. Instead of simply screening more colonies, a new experiment was designed to identify distinct integration events by genotyping individual clusters of blasticidin-resistant cells earlier in the Insertion

Stage.

WT H9 hESCs were co-transfected with Exon 10 Insertion Stage components (eSpCas9-sgRNA and HDR template) and allowed to grow spatially separated over a large surface area under blasticidin selection. By PTD14 several clusters of surviving cells had formed and appeared to be expanding (similar to but more spatially separated than those shown in figure 4.4). Though these clusters did not have single-cell origin and were thus almost certainly of mixed origins, randomly testing some of them directly with a sensitive PCR method was expected to be able to detect Exon 10 on-target cassette integration events. If not identified, negligible rates of on-target integration would be assumed and it would not be worth pursuing clonal isolation and colony screening.

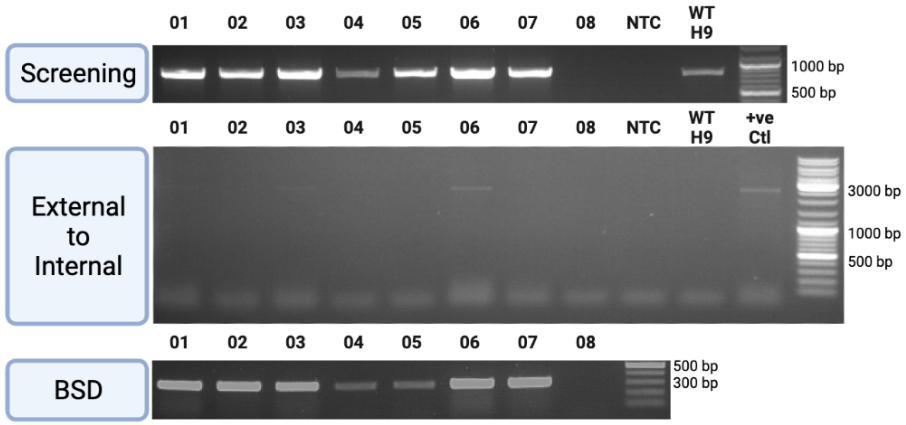

Figure 4.9: Analysis of Exon 10 Clusters. Two days post-transfection, a 70-80% confluent well of H9 hESCs transfected with Exon 10 sgRNA/eSpCas9 and Exon 10-Cassette HDR template was dissociated to single cells, passaged to a 21-fold larger surface area and allowed to grow as spatially separated clusters under 5.0ug/mL blasticidin selection. At PTD14 8 clusters were manually picked and separately expanded for further 10 days under blasticidin selection, at which point DNA was extracted for analysis. PCR results from these 8 clones are shown. Cluster 08 did not survive expansion. The other 7 clones all had detectable BSD genes. In three clusters (01,03 and 06) faint external-to-internal bands were identified, indicating small sub-populations of cells with on-target cassette integration. HEK293T CIRCES Reporter cells (exon 10) prepared in the previous chapter were used as positive control for External to Internal PCR. NTC = No template control.

Eight PTD18 clusters were manually isolated and seven survived expansion under blasticidin selection. Results suggestive of on-target insertion were identified in three clusters (Figure 4.9), indicating for the first time that on-target Exon 10 cassette insertion had indeed occurred and could be detected. Sanger sequencing of the Screening PCR amplicons revealed that all seven had heterozygous rs41289624 SNPs (data not shown), which discarded the possibility of single-cell origin. All seven clusters had detectable BSD genes, suggestive of at least a sub-

population of cells with either off-target or undetectable on-target genomic integration of the cassette (Figure 4.9).

Given that Exon 10 cassette insertion could be detected in clusters, frozen PTD14 cells were thawed and grown under blasticidin selection for four days before single-cell dissociation. To maximise chances of recovery, clonal isolation dishes were seeded at 3 different densities and selected with blasticidin. Surprisingly, however, blasticidin selection during clonal isolation led to rapid elimination of most nascent colonies and was stopped after 4 days. In total, only two colonies survived clonal isolation.

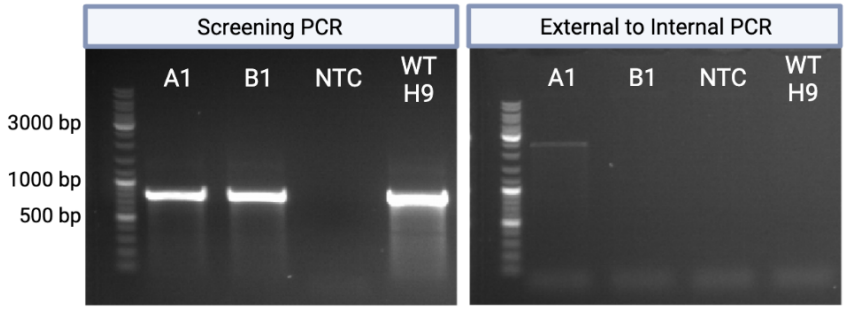

Figure 4.10: Analysis of two single-cell derived Exon 10 Colonies. Both A1 and B1 colonies had non-integrated alleles as shown by strong bands in the screening PCR, but only colony A1 generated a correctly-sized 2488bp amplicon in the External to Internal PCR.

One colony was found to have on-target monoallelic cassette insertion (colony A1, fig. 4.10). This was the first isolation of a clone carrying Exon 10 on-target cassette insertion. Investigation of the other colony (colony B1 in fig. 4.10) revealed it instead had biallelic indels. A frameshifting indel (c.797_798delCCinsG p.T266Rfs*53) was identified in trans with an in-frame indel (c.797_801delinsGTAAGGCC p.266_267delTTinsSKA). It likely has a randomly integrated cassette copy which allowed it to survive selection.

This experiment confirmed that the CIRCES strategy was able to successfully insert the selection cassette at the three target loci, and that colonies carrying these mutations could be clonally isolated.

4.3.6 Indels Always Accompanied Cassette Insertion

Across the three target exons, none of the 10 cassette-integrated colonies identified to this point had biallelic on-target cassette insertions. Thus it was important to investigate these individual colonies for the genotype of the presumably wild-type non-integrated ITPR1 target allele. To this end 'Screening PCR' reactions, which can only amplify alleles that do *not* contain a cassette (Fig. 4.1C), for all 10 cassette-integrated colonies (1 in Exon 10, 6 in Exon 58, 3 in Exon 59) were Sanger sequenced.

Target	Colony	Genotype of Non-Integrated Allele	Protein Prediction	Comments
Exon 10	A1	c.796dupA	p.(Thr266Asnfs*25)	
Exon 58	01	c.7698dupA + WT	p.(Glu2567Argfs*8) + WT	Mixed colony
Exon 58	02	c.7698dupA and c.7689_7701+5del	p.(Glu2567Argfs*8) and Abnormal Splicing	Mixed colony
Exon 58	03	c.7698dupA	p.(Glu2567Argfs*8)	
Exon 58	05	c.7698dupA	p.(Glu2567Argfs*8)	
Exon 58	06	c.7698dupA	p.(Glu2567Argfs*8)	
Exon 58	08	c.7698dupA	p.(Glu2567Argfs*8)	
Exon 59	01	c.7713dupT	p.(Ala2572Cysfs*3)	
Exon 59	03	c.7713dupT	p.(Ala2572Cysfs*3)	
Exon 59	08	c.7712_7713dupTT	p.(Ala2572Leufs*22)	

Table 4.3: Sanger sequencing results of non-integrated alleles from all 10 colonies isolated in this chapter and found to have heterozygous on-target cassette insertion.

All ten colonies were found to have frameshifting indels in their non-integrated alleles (Table 4.3). The single Exon 10 colony was found to carry a c.796dupA (p.T266Nfs*25) mutation (Fig. 4.11A). All six Exon 58 colonies had the same frameshifting single adenine insertion (c.7698dupA p.Glu2567Argfs*8, Fig. 4.11B), but two Exon 58 colonies generated multiple sequences which indicated mixed origins. For Exon 59, two colonies carried a single thymine insertion (NM_001378452.1c.7713dup p.Ala2572Cysfs*3, Figure 4.11C, top) and one carried a double thymine insertion (c.7712_7713dup p.Ala2572Leufs*22, Figure 4.11C, bottom).

Indels occurred at the expected cleavage site of transfected sgRNAs, confirming CRISPR/Cas9-mediated editing. The compound heterozygosity of cassette integration (which replaces an entire exon) and frameshifting indels meant that likely all identified colonies of single-cell origin would have most, if not all, of their ITPR1 expression affected by either non-sense mediated decay or protein truncation. Taken together, all 10 cassette-integrated colonies harboured mutations likely to lead to ITPR1 knockout.

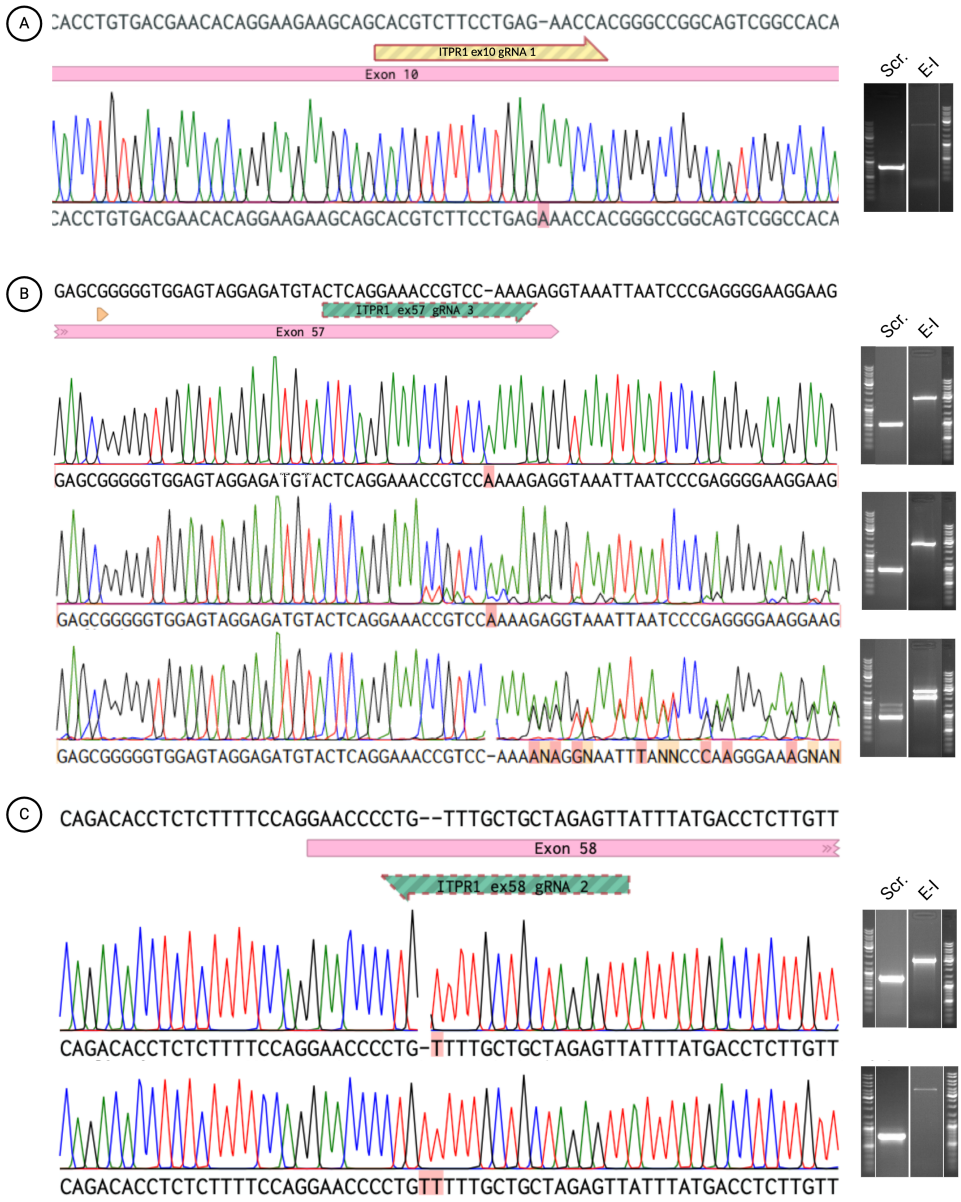

Figure 4.11: Representative Sanger sequencing chromatograms of indels found in non-integrated alleles of colonies with monoallelic on-target cassette insertion. A) Exon 10, colony A1. B) Exon 58, colony #5 (top, c.7698dupA), colony #2 (middle, primary sequence showing c.7698dupA and smaller peaks showing c.7689_7701+5del indicating this is a mixed sequence) and colony #1 (bottom, showing c.7698dupA and WT in the same chromatogram). C) Exon 59, colonies #3 (top, c.7712_7713dupTT) and #8 (bottom, c.7713dupT). Wild-type H9 reference sequences are shown at the top of each panel. sgRNA sequences used for each target are shown as arrows pointing towards the PAM site (NGG). Gel lanes for each colony's respective Screening (Scr) and External-to-Internal (E-I) PCRs are shown on the right. The three stronger ladder bands represent, from top to bottom, 3000bp, 1000bp and 500bp.

4.3.7 Identification of Mutations Near Cassette Insertion Junctions

Previous experiments identified 10 colonies with ITPR1 gene disruption due to CIRCES cassette insertion and all carried indels in the non-cassette allele. Six were thought to be of single-cell origin (i.e. not a mixed colony) and likely to represent independent cassette-integration

events (i.e. not daughter cells from one original event) based on amplicon sizes, sequencing results of non-cassette allele and different clonal isolation experiments.

Some of these clones generated 5' External to Internal PCR amplicons with sizes that were incompatible with precise HDR integration. While cassette insertion is naturally allele-disrupting as it removes most or all of the target exon, it was important to understand the genomic context immediately upstream and downstream from integrated cassettes. For that, attempts were made at amplifying and completely sequencing both 5' and 3' insertional junctions in these six clones.

Left (5') and Right (3') External to Internal PCR reactions (Figure 4.12A) were performed on each of the six clones and resulting amplicons were sequenced. A diagram of PCR and sequencing results is shown in Fig. 4.12E.

All clones produced amplicons on 5' ("left") External-to-Internal (LEI) reactions as expected (Fig. 4.12B), since this reaction had been used as a screening method for isolating these clones. LEI amplicons were in line with expected sizes for Exon 10 (2488bp, Fig. 4.12B) and Exon 58 (2718bp, Fig. 4.12B), but differed substantially between the three Exon 59 clones (#1, #3 and #8 in Fig. 4.12B). If cassette insertion had occurred exactly as intended, Exon 59 LEI should have generated a single 2596bp amplicon (similar to that seen in clone #3 in figure 4.12B). This suggested the presence of a deletion in the cassette-intron junction in clone #1 and an insertion in clone #8.

In contrast, right (3') External-to-Internal (REI) reactions generated visible bands in only half of the clones previously confirmed using LEI reactions. While Exon 59 clones #1 and #3 had identically sized bands matching the expected 2202bp amplicon size (Fig. 4.12B), Exon 58 clone #6 produced a 4-5kb band, which was much larger than the expected 2182bp (Fig. 4.12B). Clones A1 (Exon 10, expected 2398bp), #5 (Exon 58) and #8 (Exon 59) did not generate REI amplicons despite repeated attempts.

These results suggested significant variation in the genomic boundaries of inserted cassette alleles, with only one clone (Exon 59 #3) showing PCR results in line with expected sequences in the 5' and 3' cassette insertion junctions. Succesful amplicons were sequenced, the results

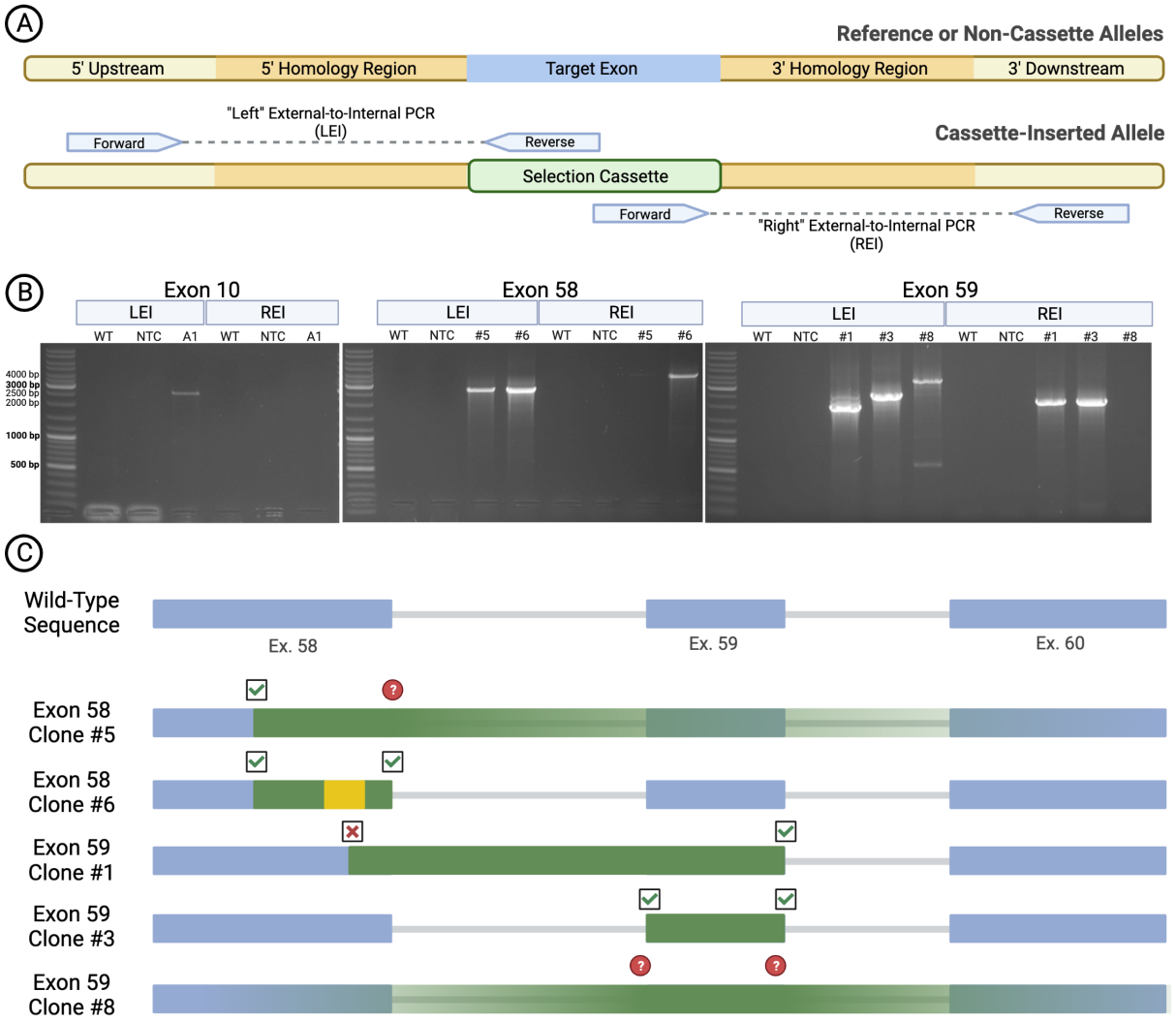

Figure 4.12: Identification of Mutations Associated with On-Target Cassette Insertion. A) Target locus-specific PCR reactions spanning 5' (left) and 3' (right) cassette insertion junctions were used to determine integrity of neighbouring sequences. B) Results of Left and Right External-to-Internal PCR Reaction on all six cassette-integrated clones. Colony A1 (Exon 10) and associated controls are shown in the left-hand side, colonies #5 and #6 (Exon 58) are shown in centre and colonies #1, #3 and #8 (Exon 59) are shown in right-hand side. WT = Wild-type, NTC = No template control, LEI = Left (5') External-to-Internal reaction, REI = Right (3') External-to-Internal reaction. C) Diagram of bidirectional Sanger sequencing results of all LEI and REI amplicons shown in [B]. Green rectangles represent the CIRCES Cassette. Yellow rectangle represents parts of the pGEM plasmid backbone. Green checkmarks indicate intact cassette-target junction, red question marks indicate that the integrity of the cassette-target junction could not be determined.

of which are shown as a diagram in figure 4.12C.

Overall, results indicated that out of the six cassette-integrated clones, only one (Exon 59 clone #3) could be confirmed as having undergone perfect HDR (i.e. did not have any associated deletions or insertions in cis) using its donor template. Another clone (Exon 58 #6) was confirmed to have intact target locus boundaries, but parts of the plasmid backbone were unexpectedly inserted within the cassette (data not shown). The remaining four clones had either proven or

suspected imperfect HDR. Three clones (Exon 10 A1, Exon 58 #5 and Exon 59 #8) did not generate 3' amplicons, a finding that is suggestive of complex genomic rearrangements that prevented PCR amplification.

4.4 Discussion

At the start of this project it was clear that a reliable and inexpensive method of delivering CIRCES constructs to hESCs was required. Although different methods were considered, reverse chemical transfection with an optimised protocol based on Mirus Trans-IT LT1 reagent and CloneR supplement led to high transfection efficiency with minimal toxicity. These results were surprising considering how difficult it has been, historically, to transfect hESCs[81, 160]. Although head-to-head comparisons were not performed, it is likely that efficiency achieved with this protocol is similar or higher to that achieved using nucleoporation, but much simpler and cheaper. The described approach requires minimal hands-on time and allows for easy set-up of parallel transfections; it should be broadly applicable to any researcher working with hESC and/or iPSC cultures.

Using this optimised method, CIRCES components were successfully delivered to H9 hESCs. Transfected HDR template plasmids drive expression of the cassette but are not capable of replicating in mammalian cells. Thus the initial transfected copy number is expected to continuously drop due to nuclease-mediated degradation[156] and replicative dilution[161]. Without genomic integration, blasticidin resistance should be transient, lasting until all transfected plasmids have been depleted. Earlier optimisation experiments showed that EF1α-eGFP constructs transfected into WT H9 hESCs generated strong visible fluorescence for 72-96 hrs, followed by a rapid decline, with virtually no GFP+ cells identified after 8-10 days (data not shown). This suggested that insertion stage duration had to be at least 10 days to allow for virtually all episomal plasmids to be lost before proceeding to replacement stage (fig. 3.1). Hence, the finding that many H9 cells were able to survive blasticidin selection for 14-20 days (Fig. 4.4) was very suggestive of genomic integration of the cassette, either on- or off-target.

Confirmation of on-target integration of a $\sim$3kb cassette required design and testing of several sets of locus-specific PCR reactions (see appendix). It was imperative to distinguish *bona fide* integration from false-positive results due to low level HDR donor plasmid contamination in the sample. Preliminary experiments had identified that trace amounts of plasmid contamination could affect PCR results despite sample pre-treatment with plasmid-specific restriction

enzymes such as DpnI (data not shown). Thus PCR reactions were designed to distinguish between human sequences within plasmids and true human chromosomal sequences. Two sets of PCR reactions (External-to-Internal and Screening) not only confirmed on-target integration, but also informed zygosity of integration and allowed for sequencing of each allele.

Although this genotyping strategy was efficient, inexpensive, and easy to perform, there were limitations. Given the large size of the External-to-Internal amplicon (>2kb), it was prone to yielding false-negative results on low-quality DNA samples. Large deletions[123, 159] and complex genomic integrations (e.g. multimeric insertions, inversions, duplications, plasmid insertions, etc)[123, 157] could also lead to false-negative results.

Indeed, even in clones where cassette insertion was identified, additional mutations were very common. Most cassette insertion events seem to have occurred through imperfect HDR, with unintended insertions and deletions occurring at one of both insertion junctions. While this is unlikely to be an issue when generating knockout models, it highlights that in genome editing the user controls only where a double-strand break is generated, but the cell controls the precision of its repair.

Additional mutations despite HDR is not a problem exclusive to CIRCES. As most original CRISPR studies used only short PCR-based genotyping to identify and isolate successfully edited clones, many unintended large genomic changes were missed. Data from mouse ESCs and immortalised cell lines suggested these to be extremely common in edited lines[122, 123, 158, 159]. A recent reanalysis of 'successfully edited'[151] iPSC models of Alzheimer's Disease revealed up to 40% of clones contained additional large mutations which affected results from functional experiments[126]. These had been originally missed using only short-range PCR and were suspected only when loss of heterozygosity of nearby SNPs was detected, using the same approach as reported in this chapter to investigate Exon 10 colonies[126].

It is therefore possible that many cassette insertions in ITPR1 targets were accompanied by additional mutations, many of which prevented detection using a single External to Internal PCR reaction. Had the colony screening procedure used two separate reactions covering both 5' and 3' insertion junctions, it is probable that many more clones with cassette integration

would have been identified. Thus, the true insertion stage on-target integration rate is almost certainly higher than the results presented in this chapter.

A large number blasticidin-resistant colonies, however, had two clearly non-integrated alleles yet had detectable copies of the blasticidin-S-deaminase (BSD) gene. Although it cannot be excluded that the BSD amplification was due to residual plasmid contamination, it would be extremely unlikely to have detectable levels of plasmid more than a 30 days after transfection. It is possible that many, if not all, of these colonies represent off-target integrations of the cassette.

Random integration was always expected to happen, and the HDR templates were delivered as circular plasmids because it is generally accepted that linear dsDNA randomly integrate at higher rates[62, 154]. Given how historically difficult it was to establish transgenic hESC lines and how poorly efficient HDR-mediated knock-ins were in these cells[103, 105, 106], it was surprising to find that a large cassette with long regions of homology to a specific double-strand broken target would still integrate elsewhere in the genome at higher frequencies, likely in a homology-independent manner. Determining precisely where cassette copies were ectopically integrated was considered beyond the scope of this project and therefore not attempted.

Clonal isolation experiments at the planned end of Insertion Stage were critical to understand the outcome of genomic editing in individual cells. Despite selection with high blasticidin concentrations (3-fold higher than the minimally effective concentration of 2.5μg/mL; kill curves in the appendix), no biallelic cassette insertions were identified. Analysis of heterozygous cassette-integrants revealed that all carried NHEJ-mediated indels in their non-integrated alleles. These findings have significant implications for the original CIRCES protocol. Given that only cassette-containing alleles can be replaced by target mutations during the Replacement Stage, any indels in non-integrated alleles would reach the end of the protocol unchanged. Thus, it would be impossible to generate the desired mutant/wild-type heterozygous genotypes. For that reason, it was not deemed worthwhile to progress to Replacement Stage unless biallelic cassette insertions could be achieved at much higher frequencies or indels could be prevented from being generated in the non-integrated alleles.

Although these findings certainly an were an impediment for the two-stage CIRCES protocol as planned, it significantly accelerated identification of potentially biallelic knockouts. In a single CRISPR step, single-cell derived colonies carrying biallelic loss-of-function mutations could be genotyped and clonally isolated within 40 days from transfection, with screening of only 8-16 colonies. In comparison, even after significant advances to genome engineering techniques since this project started, most CRISPR/Cas9 genome editing experiments today aiming to generate iPSC or hESC gene knockouts screen at least a full 96-well plate of isolated clones in order to maximise chances of isolating one or two biallelic knockouts. As the protocol was not optimised for the goal of isolating biallelic knockouts, as it currently stands it is unnecessarily long, and thus it is likely that similar or even better results could be achieved in an even shorter time frame. Simple modifications to the design of the cassette to reduce the total insert length could also significantly improve knock-in rates. This selection-based approach that was used to disrupt ITPR1 could also be easily adapted to target any other genes in order to rapidly generate knockouts stem cell models.

5 | Efficient Generation of Biallelic Loss-of-Function Mutations

5.1 Introduction

Clonal isolation of CIRCES intermediates revealed two main issues which had to be addressed before considering advancing CIRCES to Replacement Stage: indels in non-HDR alleles and inefficient positive selection.

5.1.1 Indels in Non-Cassette Alleles

It was predicted that Insertion Stage would predominantly generate monoallelic cassette insertions with a minor frequency of biallelic targeting. Despite screening dozens of clones, to this point only monoallelic integrations had been identified. Inefficient positive selection lowered the odds (discussed below) of identifying biallelic insertions, but analysis of monoallelic clones revealed that all carried indels in their non-cassette alleles. Identification of indels in the non-integrated allele had both positive and negative implications.

On the positive, it accelerates the process for developing biallelic knockout cell lines, and the protocol can be re-designed to be shorter and more efficient. On the negative, it significantly impacted the original goal of CIRCES for generation of heterozygous point mutations. Indels in the non-integrated allele would not be amenable to editing at the second CRISPR round. This meant that the full CIRCES protocol could not be performed as planned. Carried forward to Replacement Stage, any compound heterozygous cassette/indel genotypes, at the end of the protocol, would only be able to give rise to clones with either WT/indel or mutant/indel genotypes. For heterozygous point mutations to be successfully generated at the Replacement Stage, indels would have to be prevented from occurring during the Insertion Stage in at least a small fraction of cells or biallelic cassette insertion would have to be achieved in the majority of cells.

Upon CRISPR/Cas9 cleavage, the eukaryotic cell attempts DNA repair by one of two main pathways: non-homologous end joining (NHEJ) or homology-directed repair (HDR)[113, 162]. Cassette/indel compound heterozygosity indicates that each cleaved allele was repaired by a different mechanism: cassette-integrated alleles are a result of HDR, whilst indels were created by NHEJ. Multiple cycles of Cas9 cutting and error-free end joining might occur at the target locus before the cellular NHEJ repair mechanism eventually makes an error, introducing or removing nucleotides which prevent this cycle from continuing[151, 163]. If this hypothesis is correct and a high-expression long-persistence Cas9/sgRNA vector is used, then it might be nearly impossible to achieve cassette insertion in one allele while the other retains its wild-type sequence[143, 151].

Possible interventions in this process include attempts to decrease NHEJ (i.e. reduce the occurrence of errors in the repair process) or to increase HDR (i.e. increase the likelihood of biallelic cassette insertion). Several interventions have been studied to address this problem[164–168]. Nocodazole is an anti-mitotic agent that effectively stalls cultured cells in their G2/M cell cycle phase, and has been one of the most studied and effective HDR enhancers as it had been shown to increase HDR rates by nearly 40% in HEK293T and 50% in iPSCs cells[105, 143]. In H9 hESCs, however, results were disappointing[105]. Moreover, many of the small molecules that have been shown to increase HDR in model systems[169, 170] are toxic to hESCs. For instance, the nocodazole protocol used successfully in iPSCs[143] was extremely toxic to H9 hESCs even at lower doses (data not shown). Therefore, small molecule-based enhancement of HDR or inhibition of NHEJ was not considered likely to be sufficiently effective for the goals of this project.

As the CIRCES cassette is a large payload that is likely to be integrated through HDR with very low efficiency, I hypothesised that the incidence of indels in non-integrated alleles might be reduced by promoting HDR-mediated integration of an alternative short template containing a silent mutation in the protospacer/PAM sites[135, 151]. If the alternative blocker template occurred at much higher efficiency than CIRCES cassette insertion, then it would be expected that a subset of blasticidin-selected clones would have the inserted cassette in one allele and the silent blocking mutation in the other. However, since the silent blocking mutation would

not be removed from the second allele during replacement stage, the CIRCES protocol would no longer be 'scarless'. But if this strategy worked at high efficiency, it would be theoretically possible to design the alternative allele to not only contain the silent blocking mutation but also the desired final mutation. In this case, the expected genotype outcome after the Insertion Stage would be cassette/mutant, and replacement stage could then be modified to only replace the cassette back to its wild-type allele (or any other sequence).

It was unclear whether a particular alternative template design would be superior for this specific application. An approach using short single-stranded DNA oligonucleotide (ssODNs) with chemically modified terminal residues was eventually chosen with the aim to maximise stability in transfected cells[133, 163]. Conflicting data emerged from different publications regarding the optimal design strategy. While some groups published high rates of ssODN-mediated editing with a symmetrical design (i.e. both homology arms are roughly the same length relative to the expected DSB location)[105, 171], other data suggested that specific symmetry and polarity rules had to be followed[144, 151, 172]. Similarly, some studies suggested very short ssODNs with only 30-40nt of homology in each side[62], while others had better results with ssODNs with homology arms up to 90nt[105, 135, 171]. As effectively three different target loci were being tested, it was an opportunity to test different design alternatives[135, 163, 172]. Moreover ssODNs had to be chemically co-transfected with plasmids into H9 hESCs, something for which established protocols were not available. Finally, next-generation sequencing was used to detect whether ssODN-mediated editing could be detected at low levels in pooled cells and this was followed by clonal isolation in attempts to isolate successfully-edited clones.

5.1.2 Incomplete Blasticidin Selection

Success of the two-stage CIRCES protocol depends on near-complete elimination of non-cassette-expressing cells by the end of the Insertion Stage. As discussed in the previous chapter, even small numbers of non-cassette-expressing cells reaching the second CRISPR round would lead to deterioration of the overall protocol efficiency by significantly increasing the false-discovery rate. An assumption was that resistance to blasticidin could only be elicited in cells carrying at least one copy of the BSD gene, either genomically inserted (on or off-

target) or episomally (i.e. transfected foreign genetic material). The second assumption was that episomal DNA is eventually completely lost.

Blasticidin selection had to be sufficient in concentration and duration so that only cells with genomic integration of the cassette would be able to survive through the end of the insertion stage. However, a common finding was identification of presumably wild-type cells even after long periods of positive selection with blasticidin. Higher doses of blasticidin did not eliminate such cells, thus they were considered to display *bona fide* blasticidin resistance. Possible reasons for this phenomenon included both off-target (random) integration of the cassette and prolonged retention (and expression) of episomal cassette-containing HDR template plasmids. It was not straightforward nor necessary to confirm, exclude or differentiate these two possibilities with absolute confidence. As designed, the CIRCES protocol eliminates cells containing randomly integrated cassette by negative selection with ganciclovir during the replacement stage[148]. Likewise, any cells expressing episomal copies of the cassette would be removed by ganciclovir treatment. Final CIRCES efficiency would be affected, however, by loss of residual plasmid occurring during the short period between end of positive selection and initiation of negative selection. The second part of this chapter describes how these competing possibilities were investigated, and presents evidence that a previously unappreciated mechanism of blasticidin resistance interferes with positive selection during insertion stage.

In this chapter I troubleshoot these two critical issues in an attempt to prevent the generation of cassette/indel genotypes and improve the efficiency of positive-selection during the insertion stage.

5.2 Materials and Methods

5.2.1 Design of Single-Stranded HDR Templates

Single-stranded oligonucleotides (ssODNs) were ordered from IDT as chemically modified to contain two phosphorothioate bonds at each end to improve stability and reduce exonuclease-mediated degradation[133, 163]. All ssODNs were designed to be complimentary to the non-target strand as per references [135, 144, 172]. ssODNs were initially resuspended to 100μM concentration using 10mM Tris-Cl, pH 8.5 (Buffer EB, Qiagen).

For Exon 10 two 128nt ssODNs containing a silent c.801G>T PAM blocking mutation were designed to be asymmetric (-91/+37) relative to the expected Cas9-sgRNA cleavage, according to Jacob Corn's design algorithm[172]. A second donor (Ex10-128-R269G) with this design also contained the c.805C>G (p.R269G) SCA29 mutation in close proximity to the silent mutation (Figure 5.1A).

For Exon 58, two different approaches were tested (Figure 5.1B). A short 60nt sense ssODN[135] containing only the sgRNA-blocking c.7698A>G silent mutation 2nt from cutting site (Ex58-60-Silent). Also, a long 150nt[173] sense ssODN containing the same silent mutation in addition to the AD Gillespie-causing c.7660G>C (p.G2554R) mutation 37nt distant from the expected cleavage site (Ex58-60-G2554R, Figure 5.1B).

For Exon 59, given the long distance (120bp) between the expected cut site and the desired p.K2611del mutation site, only a 60nt antisense silent ssODN containing the c.7708C>T sgRNA-blocking mutation (3nt from cutting site) was designed[135] (Ex59-60-Silent, Figure 5.1C).

5.2.2 Chemical Co-Transfection of Plasmids and ssODNs

Optimisation experiments demonstrated that 100pmol of ssODNs led to best transfection efficiency and least toxicity (as measured by co-transfection with an EGFP construct, data not shown). As such, 100pmol of each ssODN were used on top of the established reverse transfection procedure without further changes.

5.2.3 Blasticidin Selection of Cell Clusters

At PTD2, wells were pre-treated for one hour with 10μM Rho-Kinase Inhibitor (ROCKi) before dissociation with ReLeSR (StemCell). Dissociated cells were resuspended in 4mL of growth medium supplemented with 10μM ROCKi, pipetted up and down several times in a 1000μL plastic tip to break larger aggregates and were equally distributed into two 60mm Matrigel-coated culture dishes. Twelve hours later, 1μL of blasticidin stock (10mg/mL) was directly added to each dish and swirled around to mix. After 8 hours, a full medium change was undertaken with fresh complete culture medium supplemented with 5μg/mL blasticidin. The following day, all dishes were topped up with 1mL fresh complete culture medium supplemented with 5μg/mL blasticidin. After this point, daily full medium changes supplemented with blasticidin 5μg/mL were performed.

5.2.4 Next-Generation Sequencing of Amplicons

Screening PCR amplicons were produced for Exon 58 and Exon 59 using Q5 High-Fidelity Polymerase and primers containing overhangs for Illumina NGS adaptors (listed in the appendix). PCR reactions were gel purified and submitted for Illumina paired-end 2x250 sequencing (Genewiz, Leipzig, Germany). Only one amplicon per transfection sample was sequenced. Fastq files were processed using CRISPResso2[174] with minimum read qualities set to 30. Only reads that could be aligned to the amplicon sequence after pre-processing were considered for calculations.

5.2.5 Negative Selection with Ganciclovir

Ganciclovir (Invivogen cat. sud-gcv) was reconstituted in distilled water pH 12, diluted to 10mg/mL and sterile filtered. Stocks of 5mg/mL (approximately 20mM) were frozen at -20C and defrosted when needed to prepare 2mM working stock.

H9 hESCs were negatively selected with 5μM ganciclovir concentration by adding 1:400 ganciclovir working stock to maintenance medium. Fresh ganciclovir was added with every daily media change.

5.3 Results

5.3.1 Using ssODNs to Prevent NHEJ-Mediated Indel Formation

Single-stranded HDR templates between 60-150nt were designed to introduce blocking mutations for Cas9 targeting in all ITPR1 target loci. Two ssODNs also contained SCA29 or Gillespie-causing mutations (Fig. 5.1). Mutant ssODNs were co-transfected into WT H9 hESCs with target-specific eSpCas9-sgRNA and cassette-containing HDR template plasmids. Parallel control transfections were also undertaken with eSpCas9-sgRNA and cassette HDR templates but without ssODN templates.

Although two different ssODNs of identical length were designed for Exon 10 (Fig. 5.1A), cells transfected with either of them developed significant toxicity in the first 48 hrs post-transfection. Toxicity was not seen in cells transfected only with Exon 10 sgRNA/eSpCas9 and cassette-containing HDR template. As toxicity led to near complete cell death, it was not possible to continue with experiments targeting Exon 10.

No significant toxicity was noticed in transfection conditions for Exons 58 and 59. Following the optimised protocol developed in the previous chapter, transfected cells were single-cell dissociated and passaged to large cell culture vessels on PTD2 to allow for growth of spatially separated clusters. Blasticidin selection at 5μg/mL was initiated on PTD3.

Control cells transfected only with an GFP construct were killed within 48hrs of blasticidin selection. At PTD14, across all conditions, clusters of blasticidin-resistant cells were obtained (data not shown) and analysed.

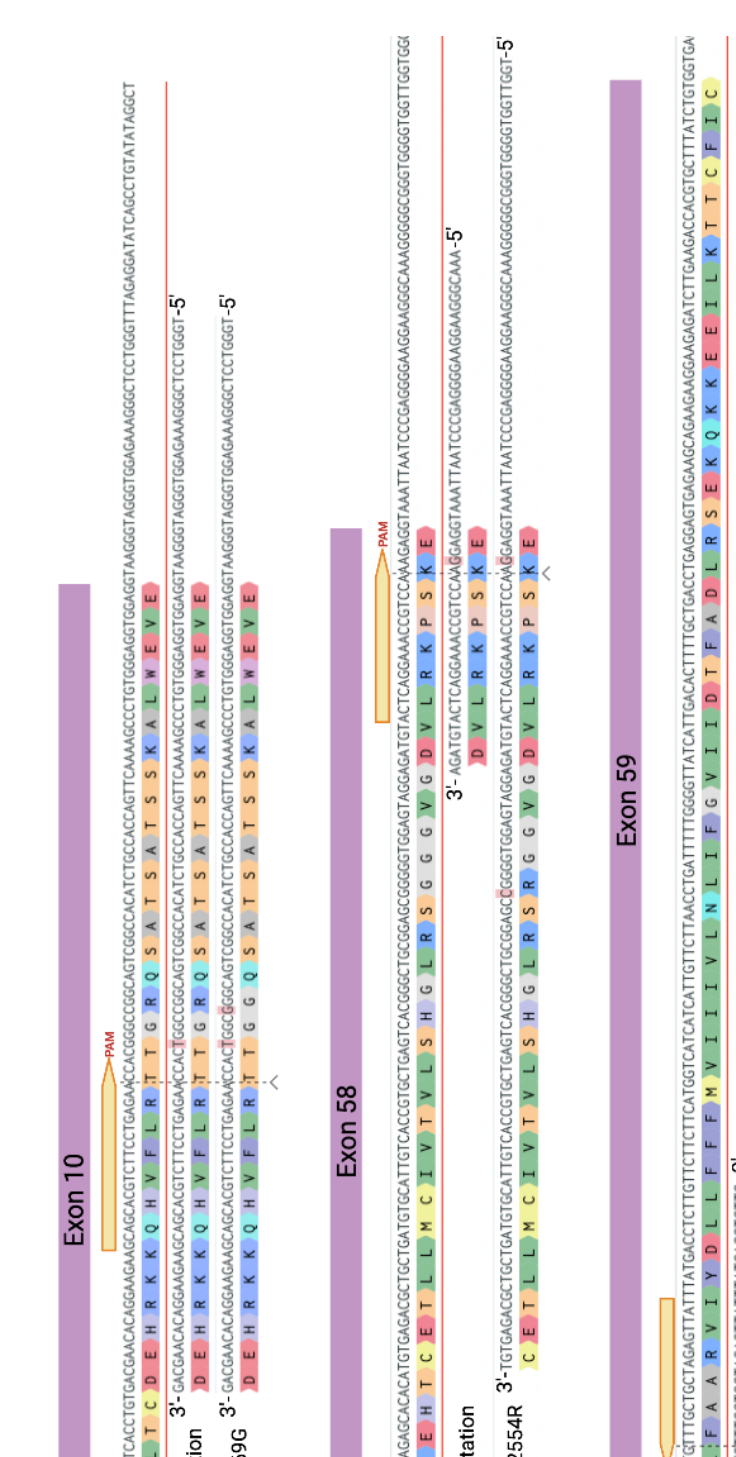

Figure 5.1: ssODN HDR templates used. (A) For exon 10, the asymmetric design proposed by Richardson et. al[172] was used. It is a 128nt ssODN with asymmetric homology arms (-91/+37) around the expected cut site of Exon 10 sgRNA #1. It includes a silent PAM mutation (NGG>NTG). Two versions were designed, one of which contains also a single-nucleotide change 8nt from the expected cut site which introduces a p.Arg269Gly SCA29-causing mutation. (B) For exon 58, two different ssODNs designs were used. The total length of the first is only 60nt, with 23nt of homology to the PAM-distal side of the expected cut of Exon 58 sgRNA #3fix and 37nt homology to the PAM-proximal side. The second is a 150nt ssODN with asymmetric homology arms of -91/+59nt. Both include a silent A>G mutation in the seed region of the protospacer at position -2 from PAM, as silent PAM mutations could not be generated due to risk of disrupting the exon 58 donor splice site. The longer ssODN also includes a Gillespie Syndrome-causing p.Gly2554Arg mutation in the PAM-distal side 36nt distant from the cut site. (C) For exon 59 a single perfectly symmetric (-30/+30) 60nt ssODN was designed to include a silent C>T mutation in the -1 position (seed region) of the protospacer recognised by Exon 58 sgRNA #2. Regardless of the length and symmetry, all ssODNs used were complementary to non-target strand as per ref. [172]. sgRNA protospacer sequences for each target are shown as yellow arrows. Expected cut sites (between -3 and -4 from PAM) are highlighted as grey dashed lines.

5.3.1.1 Analysis of Blasticidin-Resistant Clusters

To determine whether PTD14 clusters contained sub-populations of cells with on-target cassette insertions, eight randomly selected clusters from each condition were isolated, expanded under blasticidin selection and investigated for presence of on-target cassette and ssODN blocking mutation incorporation.

After 10 days, all expanded clusters were subjected to gDNA extraction, PCR and Sanger sequencing (Fig. 5.2). Sequencing of non-integrated alleles in surviving clusters was not able to identify incorporation of blocking mutations. The amplified non-integrated alleles mostly consisted of wild-type sequences, with generally low indel frequencies (Fig. 5.2).

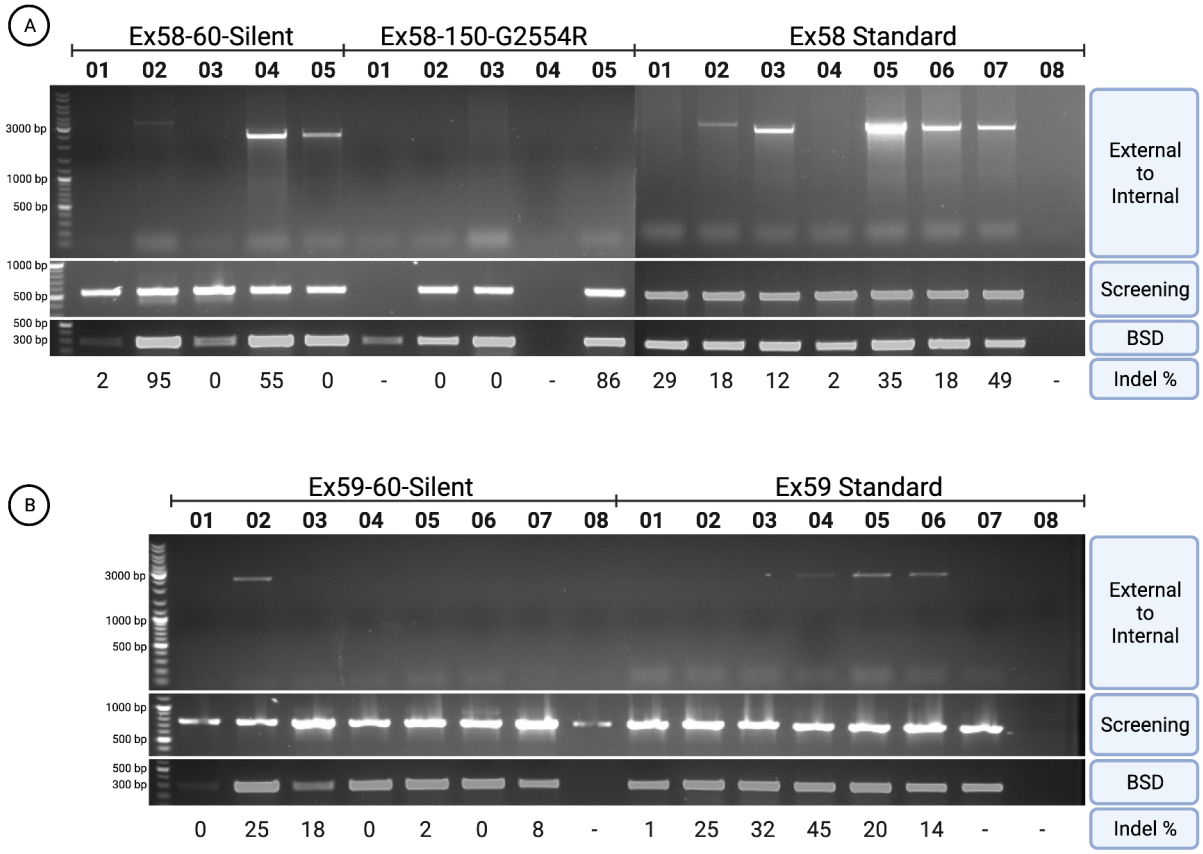

Figure 5.2: Analysis of PTD14 Blasticidin-resistant clusters. H9 hESCs were co-transfected with standard CIRCES insertion stage plasmids (target-specific sgRNA/eSpCas9 + cassette HDR template) with or without 100pmol ssODN donor template (fig. 5.1) and allowed to grow under 5μg/mL blasticidin selection to form isolated clusters. At PTD14, 8 clusters from each condition were manually isolated, expanded for 10 additional days under blasticidin selection and analysed using External-to-Internal, Screening and BSD PCR reactions. Screening PCR amplicons were sequenced and analysed using Synthego ICE to calculate indel frequency in each cluster. (A) Clusters from cells transfected with Ex58-60-Silent ssODN template (cols. 01-05 on left-hand side), with Ex58-150-G2554R (cols. 01-05 in centre) and without ssODNs (cols. 01-08 on right-hand side). (B) Clusters from cells transfected with Ex59-60-Silent on cols. 01-08 on left-hand side and from cells not transfected with ssODNs on cols. 01-08 on right-hand side.

Across the different transfections, cassette insertions were detected in slightly over a third of clusters that produced sufficient DNA for analysis (Fig. 5.2). All but one had detectable BSD genes, supporting on- or off-target cassette integration.

Overall, these results suggested that while on-target cassette insertion could be identified in many of the clusters, incorporation of ssODNs-introduced blocking mutations could not be identified using Sanger sequencing.

5.3.1.2 Next-Generation Sequencing of Populations of Blasticidin-Resistant Cells

To investigate whether ssODN-mediated editing had occurred at all, populations of blasticidin-resistant PTD17 cells were analysed using amplicon-based next-generation sequencing. The genomic region encompassing the expected sgRNA cut site (and any of the blocking mutations) was PCR amplified and sequenced to high depth to identify low frequency alleles. Crucially, this method is unable to detect presence of the cassette, as the sequenced amplicons only cover non-cassette alleles. Additionally, it cannot detect larger indels that cannot be amplified by a screening PCR reaction.

Total aligned reads varied between 11762 (Ex58-60-Silent) and 47647 (Ex59-60-Silent), while total aligned reads ranged between 87355 (Ex59 Standard) and 225910 (Ex58-60-Silent).

Using this approach, precise HDR-mediated incorporation of blocking mutations was identified in 4.4% of all processed reads for Ex58-60-Silent, 0.5% for Ex58-150-G2554R, and 3.13% for Ex59-60-Silent (Fig. 5.3A). In standard CIRCES transfections (e.g. no ssODNs used), indels were identified in 19.1% of Exon 58 and 20.3% of Exon 59 non-cassette alleles. In ssODN-transfected cells, indels were identified in 20.1% of Ex58-60-Silent, 27.5% of Ex58-150nt-G2554R and 16.5% of Ex59-60-Silent's non-cassette reads (Fig. 5.3B).

Across all conditions, wild-type sequences represented 70-80% of all non-cassette reads.

Ex59-60-Silent ssODN led to the best results, with nearly 17% of non-reference non-cassette alleles having the silent blocking mutation. For Ex58-60-Silent this number was approximately 11%, while for Ex58-150-G2554R it was less than 2%.

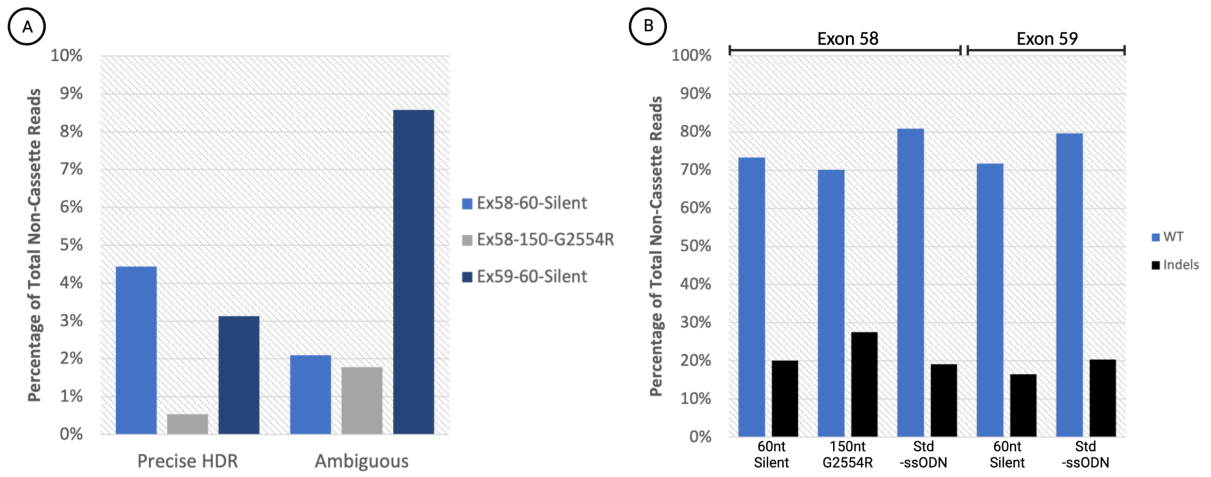

Figure 5.3: Deep Sequencing Results of Exon 58 and Exon 59 Screening PCR Amplicons. PTD17 H9 blasticidin-resistant clusters were dissociated to single cells and an aliquot was used for Exon 58 or Exon 59 Screening PCR using primers containing NGS adaptors. Amplicons were sequenced at high depth using Illumina paired-end sequencing as a commercial service by Genewiz (Leipzig, Germany). (A) Rates of precise HDR (i.e. only the desired mutation without any other mutations or indels detected in the amplicon) and ambiguous reads (i.e. cannot be classified due to identical scores in two or more categories). (B) Percentage of non-cassette reads that are wild-type or indels for each transfection. Std -ssODN denotes control standard insertion stage CIRCES transfections without ssODNs.

These results suggested that co-transfection preferably with short ssODNs might be a useful strategy to reduce indels in the non-cassette allele.

5.3.1.3 Analysis of Single-Cell Derived Colonies

Although further optimisations are required to improve the ratio of blocking mutation incorporation, the NGS results hinted at the possibility that for some conditions between 1:6 and 1:9 of cassette-integrated clones had integrated a silent mutation in their non-cassette alleles.

To test whether it was possible to isolate clones with on-target cassette insertion in trans with silent blocking mutations, remaining PTD14 blasticidin-resistant clusters were single-cell dissociated and clonally isolated under blasticidin selection. After 12 days, however, colonies could only be generated from cells transfected with either Ex58-150-G2554R or Ex59-60-Silent. Despite seeding more than 1000 cells, no single cell-derived colonies from the other conditions survived continued blasticidin selection.

All surviving colonies from these two transfections were manually isolated, expanded and analysed by PCR and Sanger sequencing (Fig. 5.4).

For cells transfected with Ex58-150-G2554R, 14/15 colonies produced Screening PCR ampli-

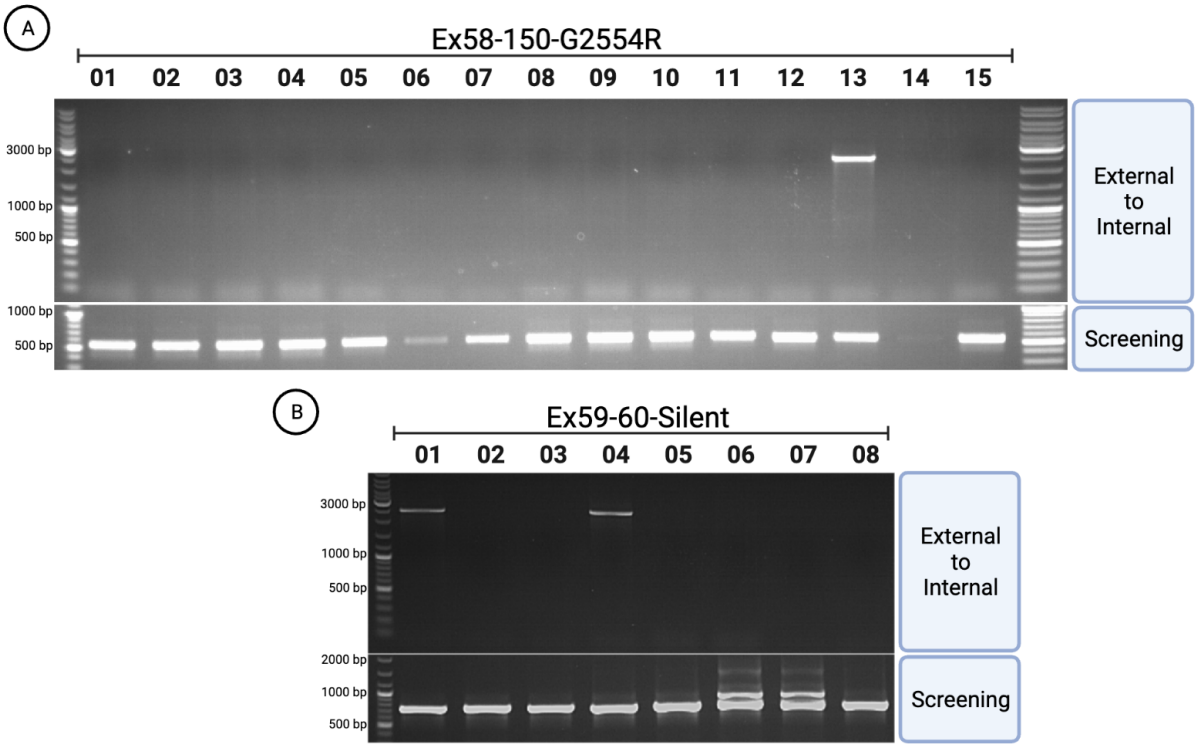

Figure 5.4: Analysis of ssODN-transfected Single Cell-Derived Clones. Blasticidin-resistant PTD14 clusters remaining after manual isolation represented in fig. 5.2 were dissociated to single cells, seeded at clonal density and allowed to grow under continued blasticidin selection for 12 days. Surviving colonies were only found in dishes seeded from Ex58-150-G2554R and Ex59-60-Silent transfections. All surviving colonies were picked, expanded for 10 more days under blasticidin selection and harvested for DNA extraction and PCR-based analysis using External-to-internal and Screening PCR. (A) 15 colonies isolated from Ex58-150-G2554R transfections. Colony 13 shows heterozygous exon 58 on-target cassette integration. Colony 14 did not generate enough DNA for accurate analysis. (B) 8 surviving Ex59-60-Silent colonies. Heterozygous exon 59 on-target cassette integration was identified in colonies 01 and 04. Colonies 06 and 07 were found to carry a 210bp insertion, which explains the two bands seen on Screening PCR.

cons, but, of these, only one colony had results compatible with on-target cassette insertion (Fig. 5.4A).

For Ex59-60nt-Silent transfections, 2 colonies had results suggestive of on-target cassette insertion while all 8 surviving colonies produced Screening PCR amplicons (Fig. 5.4B).

Sanger sequencing revealed that knock-in of blocking mutations had not happened in any of the 22 colonies that could be analysed. All three cassette-integrated colonies had indels affecting the protospacer and/or PAM regions (c.7698dupA for the Exon 58 colony and c.7702-8_7711del for the two Exon 59 colonies).

Sequencing also revealed an unexpectedly high rate of indels in the other 19 colonies. All Exon 58 colonies carried biallelic mutations. Remarkably, genotypes were identical across the

13 colonies that did not have a detectable cassette: a 21bp deletion encompassing the Exon 58 donor splice site (c.7682_7701+1del) and a 1bp frameshifting insertion (c.7698dupA). For the six Exon 59 colonies without a detectable cassette, two colonies were homozygous wild-type, two were had biallelic frameshifting indels (a 17bp deletion [c.7706_7722del] and a 210bp insertion [p.F2571Efs*3]), and the last two were heterozygous for different indels.

Repeating the entire clonal isolation procedure using PTD23 blasticidin-resistant cells led to essentially the same results (data not shown). Despite seeding a larger number of cells, very few colonies survived blasticidin selection. No ssODN HDR-mediated blocking mutations were identified and nearly all surviving colonies had biallelic indels. Out of 21 surviving colonies for Exon 58 (Ex58-150-G2554R), 19 had the same biallelic indel genotype (c.7682_7701+1del; c.7698dupA) and 2 were homozygous wild-type. None had detectable on-target cassette integration. Only a single Exon 59 (Ex59-60nt-Silent) survived blasticidin selection and it had a homozygous WT genotype, suggestive of random cassette insertion.

Taken together, these results showed an unexpectedly high rate of biallelic indels in blasticidin-resistant colonies that did not have on-target cassette insertion. Out of the 34 colonies with biallelic loss-of-function mutations (and likely to represent ITPR1 knockouts) isolated, two clones (Exon 58 clones A4 and B4, both carrying the c.7682_7701+1del/c.7698dupA genotype) were expanded and cryostored for further experiments.

Clonal isolation experiments did not identify any clones with blocking mutations, but the number of surviving colonies was much lower than expected. While 14 colonies were isolated for Ex58-150-G2554R transfections, NGS results suggested that for this particular transfection only 1 in every 50 cassette-integrated clones had incorporated the blocking mutation. Thus it was not likely that it would have been found by only screening 14 colonies. For cells transfected with Ex59-60-Silent, on the other hand, chances were much higher, but only 6 colonies survived, and only two had a detectable cassette.

5.3.1.4 Progression to CIRCES Replacement Stage

Next-generation sequencing had revealed that 70-80% of non-cassette ITPR1 alleles of PTD14 blasticidin-resistant cells were wild-type.

These results indicated that many cells are seemingly resistant to blasticidin but do not have detectable on-target cassette insertions at the end of the Insertion Stage. Two main reasons that could explain this phenomenon were random genomic integration of the cassette and/or prolonged episomal retention of the cassette-containing HDR template plasmid. In either case, negative selection with ganciclovir should remove all cassette-expressing cells during replacement stage. Given that the ssODN strategy had low, but detectable, levels of editing, continuation into the replacement stage could potentially allow for isolation of correctly edited clones.

To test this hypothesis, PTD17 blasticidin-resistant cells were single-cell dissociated and co-transfected with plasmids for CIRCES replacement stage, namely Drosophila sgRNA/eSpCas9 and target-specific ITPR1 HDR templates. Transfections were >50% efficient (based on eGFP test transfections, data not shown) and no significant toxicity was observed. Cells were re-plated at low density and selected with 5μM ganciclovir (Fig. 5.5A).

To assess effectiveness of the ganciclovir negative selection, control blasticidin-resistant cells had also been transfected with EF1α-GFP construct but not with Replacement Stage CIRCES components (Fig. 5.5A).

Despite initial signs of ganciclovir-mediated toxicity, after 6 days the number of surviving colonies was broadly similar between replacement stage transfections and control GFP-transfected cells. Given it was impossible for control GFP-transfected cells to replace their cassette copies due to lack of homology, this indicated either practical issues with ganciclovir selection (i.e. insufficient concentration or instability in the culture medium) or a large number of cells not properly expressing the HSV-TK gene.

To identify the reason behind the failure of ganciclovir selection, three low-density control GFP-transfected dishes that had been ganciclovir-selected were used for a side experiment (Fig. 5.5B). In one, ganciclovir was replaced with medium dose blasticidin, while combined

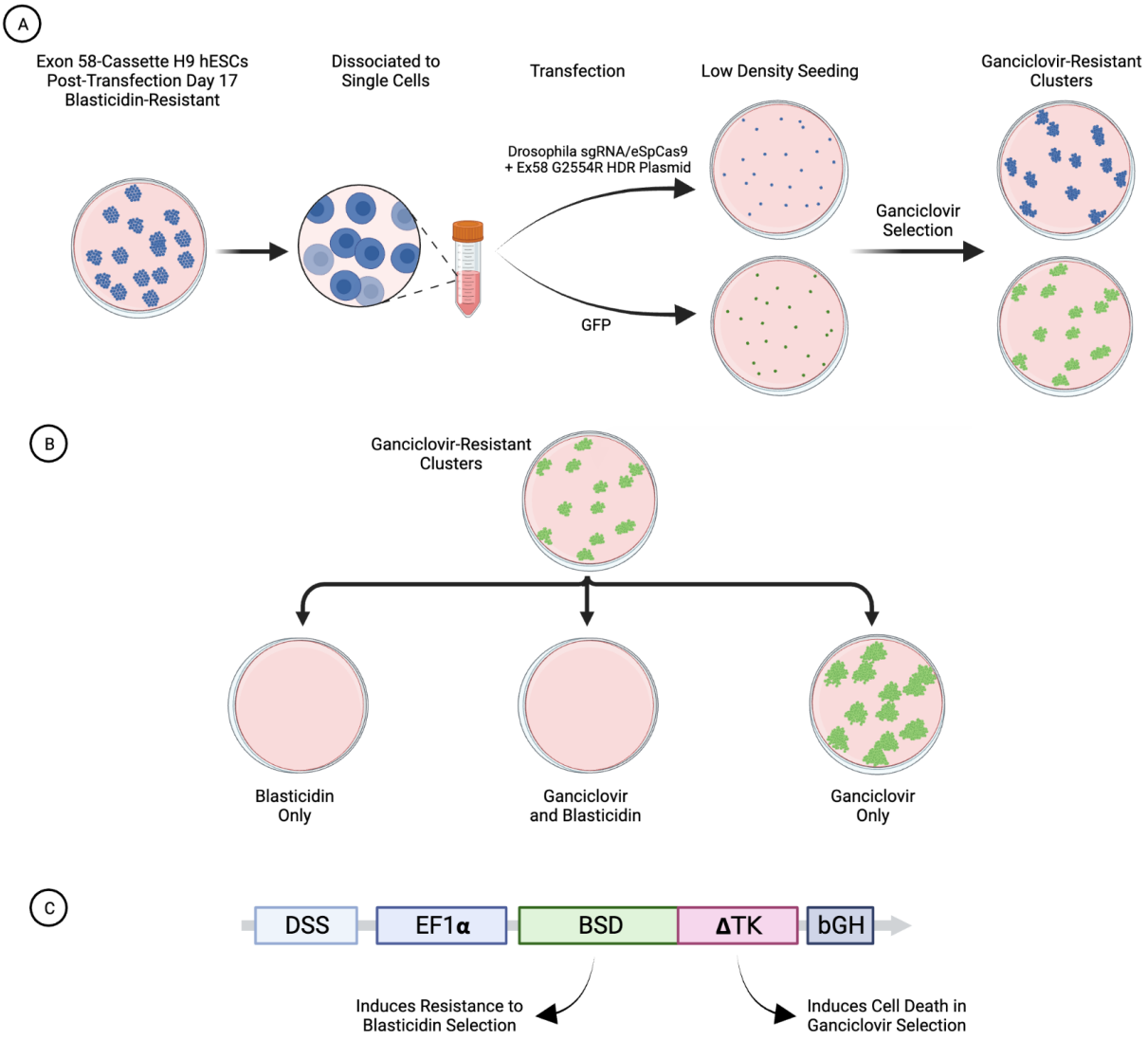

Figure 5.5: Replacement Stage Experiments. (A) H9 hESCs had been originally transfected with Insertion Stage + ssODNs and allowed to grow as spatially separated clusters under 5μg/mL blasticidin selection. At PTD17, blasticidin selection was discontinued and clusters were dissociated to single-cells and transfected with Drosophila sgRNA/eSpCas9 [validated in Chapter 3] and mutant ITPR1 HDR templates or with a GFP construct (control). Transfected cells were seeded at low density and 5μM ganciclovir selection was initiated. After 6 days, the number of ganciclovir-resistant clusters was similar between dishes transfected with replacement stage plasmids vs. control GFP constructs. The image depicts an attempt to generate ITPR1 G2554R (starting from Ex58-60-Silent transfections) clones, but parallel experiments were also performed for ITPR1 K2611del (starting from Ex59-60-Silent transfections) with similar results. (B) As multiple dishes had been seeded at low-density after transfections depicted in [A], it was possible to investigate the mechanisms of ganciclovir resistance by altering selection conditions. Starting 6 days after the second round of transfections, each dish with ganciclovir-resistant clusters was exposed to blasticidin 5μg/mL; combined ganciclovir 5μM and blasticidin 5μg/mL; or continued under ganciclovir 5μM selection. After 2 days, ganciclovir-resistant clusters continued to expand under ganciclovir only selection, but were completely eliminated in blasticidin-only or combined blasticidin-ganciclovir selection. (C) Diagram of CIRCES cassette. The BSD-TK fusion protein stably expressed by the cassette promotes survival in the presence of blasticidin and apoptosis in the presence of ganciclovir. The Drosophila sgRNA introduces a DSB at the Drosophila-Specific Sequence (DSS) site to trigger DNA repair mechanisms. If the cassette stops being expressed (e.g. replaced by a Replacement Stage HDR template), the cell would be rendered ganciclovir-resistant.

blasticidin and ganciclovir selection was applied in another. The third dish was maintained on ganciclovir selection only. Assuming adequate concentrations and stability of selection agents,

cells expressing a CIRCES cassette would survive blasticidin-only selection but be killed by combined blasticidin-ganciclovir selection. In contrast, cells not expressing a viable cassette would be able to survive in the ganciclovir-only selection, but not in any of the other two dishes due to blasticidin. However, if the ganciclovir dosing was insufficient, cassette-expressing cells would survive in all three conditions (Fig. 5.5B).

After 48 hrs, cells were killed in both blasticidin-only and combined blasticidin-ganciclovir dishes (schematics of the results shown in Fig. 5.5B and cassette structure in Fig. 5.5C). These results indicated that ganciclovir-selection failure was not caused by inadequate ganciclovir dosing, but rather by large numbers of cells not expressing the selection cassette. In the three day period between positive selection discontinuation and initiation of negative selection, it appeared that a large number of cells effectively ceased to express the CIRCES cassette (Fig. 5.5C).

Progress onto Replacement Stage is not possible unless cells not expressing the cassette can be adequately removed during the Insertion Stage. Given the already existent issues with occurrence of indels in most non-cassette alleles, the likelihood of identifying a precisely edited heterozygous mutant would be very low.

5.3.2 Passive Transfer of Blasticidin Resistance

Results from different experimental lines converged onto the realisation that many cells lost blasticidin-resistance once they were selected at clonal or very low density. Both epigenetic silencing and untimely loss of residual episomal plasmid were considered as possible, though unlikely, causes. As an alternative explanation, I hypothesised that blasticidin resistance in wild-type cells could also be mediated by proximity to (few) cassette-expressing cells.

To investigate if cells can transfer passive resistance, a mixing experiment was prepared. A previously isolated blasticidin-resistant clone of a known genotype (Exon 58 clone #6 - cassette/indel) was single-cell dissociated and mixed at a 1:9 ratio with wild-type H9 cells. Positive and negative control wells were seeded only with WT H9 hESCs or mutant cells respectively (Fig. 5.6). After five days under blasticidin selection, wells were imaged and total cell counts

were established.

Wells seeded only with wild-type cells (90,000 cells/well) were fully killed (Fig. 5.6, left column) after 5 days under blasticidin selection, indicating adequate blasticidin dosage and length of selection. In contrast, wells seeded with only cassette-integrated cells (10,000 cells/well) had numerous healthy colonies (mean of 33,000 cells/well), confirming true blasticidin resistance (Fig. 5.6, centre column).

Mixed wells cultured under blasticidin selection, surprisingly, were fully confluent (Fig. 5.6, right column). Despite being seeded only with 10,000 truly blasticidin-resistant cells at the start of the experiment, the final average cell count in mixed wells was 1.83 million cells/well. Thus the total cell counts were 54-fold higher in mixed wells compared to cassette-only wells, despite having the exact same number of resistant cells at the start.

This experiment has been repeated multiple times with similar results, strongly suggesting a 'bystander effect' for blasticidin in H9 hESCs. Although mechanisms are unclear, susceptible cells passively develop blasticidin resistance by being in close proximity to truly blasticidin-resistant cells. This unexpected finding is likely the cause of most issues with positive selection during Insertion Stage. Unless this is addressed, it will not be possible to progress to Replacement Stage because the False-Discovery Rate would be too high.

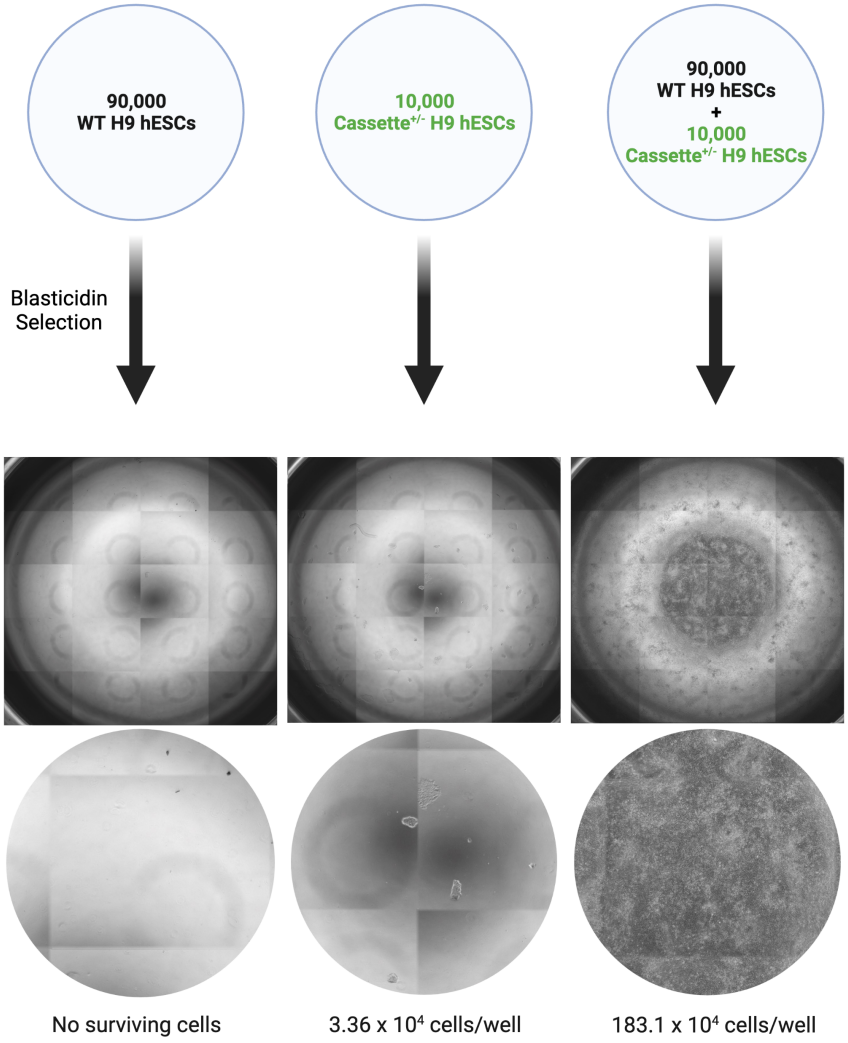

Figure 5.6: Bystander Blasticidin-Resistance Effect. Mixing experiments were carried out with WT H9 hESCs and ITPR1 Ex58 clone #6 (isolated in chapter 4, known to have on-target insertion of the cassette in one ITPR1 exon 58 locus and a frameshifting indel in the other). Both WT and blasticidin-resistant H9 cultures were dissociated to single cells prior to seeding. Wells were seeded with 90,000 WT H9 hESCs only; 10,000 cassette-expressing clone #6 hESCs only; or a mixture of 90,000 WT and 10,000 clone #6 cells. Blasticidin selection at 5μg/mL (medium dose) was initiated the following day. Medium with blasticidin was changed daily, and therefore cellular debris were removed. After 5 days under blasticidin selection, no cells remained in the wells seeded only with WT cells. In the wells seeded only with clone #6 cells, numerous small colonies could be seen. Wells seeded with both WT and blasticidin-resistant cells, in contrast, had achieved 100% confluency despite blasticidin selection. Experiment was run in quadruplicate. Wells were fully imaged at 4x magnification and panorama images of the entire well (shown in the middle row) created using an EVOS M7000 automated microscope (Thermo Scientific). Bottom row of the image shows a close-up view of approximately one field of view at 4x magnification. Cells in each well were dissociated to single cells, two wells were pooled and an aliquot was counted twice on an automated cell counter (BioRad TC20) without Trypan Blue. Average total cell counts for each well were calculated and are shown at the bottom of the image.

5.4 Discussion

The insertion stage of CIRCES was shown in the previous chapter to successfully insert a large selection cassette in three target ITPR1 loci. Detailed investigations revealed a surprisingly high number of colonies with biallelic loss-of-function mutations which could be used as cellular models of Gillespie syndrome. Generation of SCA29 models, however, required precise introduction of heterozygous missense point mutations, which can only be produced during the Replacement Stage.

Two main issues were identified that prevented progression to Replacement Stage: insufficient blasticidin selection and excessive indel formation.

The underlying assumption had been that only cells carrying genetic copies of the Blasticidin-S-Deaminase (BSD) gene would survive prolonged positive selection with Blasticidin. As the results of this chapter have shown, this assumption was unfortunately proven incorrect. Large quantities of cells lacking BSD copies do survive blasticidin selection when mixed with a small number of BSD^+ cells. This is an extremely interesting finding which, to the best of my knowledge, has not been described before, but unfortunately it imposes a large burden on the current CIRCES protocol.

It is known that susceptibility to ganciclovir (a negative selection agent) can be transferred to neighbouring cells in a process known as 'bystander effect'[175]. HSV-TK^- cells (i.e. normally resistant to ganciclovir) are killed by toxic ganciclovir metabolites produced by HSV-TK^+ cells. The most accepted mechanism is transfer of ganciclovir-phosphate through intercellular gap junctions[176].

Blasticidin resistance is mediated by expression of Blasticidin-S-Deaminase, an enzyme that inactivates blasticidin to deaminohydroxiblasticidin[177]. For blasticidin resistance to be transmitted to neighbouring cells, the 13.5kDa-enzyme blasticidin-S-deaminase would need to be transferred between cells or be secreted (and remain active) in their vicinity. One intriguing possibility is transfer through tunnelling nanotubes, which are recently described[178] membrane connections between proximal cells that have been shown to allow for transfer of entire

organelles[179] and also proteins such as prions (PrP)[180], Huntingtin (Htt)[181, 182], α-synuclein (α-syn)[183] and β-amyloid (APP)[184].

Detailed investigation of the mechanisms behind passive transfer of resistance was beyond the scope of this project, and at this time it is not known whether it is specific to H9 hESCs or also happens in other human cell types. However, its implication for the originally planned CIRCES protocol are significant. As previously discussed, even small numbers of cells lacking genomically integrated copies of the cassette would lead to substantial increases in false-discovery rates at the end of the Replacement Stage. To work around this limitation during Insertion Stage, cells to be selected would need replating at clonal density as soon as possible after transfection. Even in the most optimistic scenarios of CRISPR HDR efficiency, this would be impracticable. For example, under optimistic assumptions of 50% transfection efficiency, 0.1% cells undergoing HDR and 25% survival[185] after single-cell seeding, on average at least three 100mm dishes would need to be grown at clonal density (25 cells/cm^2) for a single cassette-inserted clone to be identified.

If passive transfer of antibiotic resistance is a blasticidin-specific phenomenon, the only realistic option would be to use a different combination of antibiotic and resistance gene. As reported in Chapter 3, a puromycin-resistant version of the CIRCES cassette was an intermediary step in the cloning of the final blasticidin-resistant version. While it would be relatively straightforward to adapt the existing CIRCES protocol, puromycin selection of H9 hESCs has proved challenging in preliminary experiments. Puromycin selection is highly density dependent and requires dynamic concentration adjustments as colonies grow or shrink. Accidental killing of all cells due to incorrect puromycin dosage for a given cell density happened often during preliminary testing (data not shown). Other potential antibiotic candidates included Hygromycin B, Geneticin/G418 or Zeocin. Cloning of the Hygromycin resistance gene into the CIRCES cassette was attempted, but despite significant time and resources it could not be achieved (data not shown). The other two antibiotic options were not tested due to time constraints.

While this project was ongoing, the concept of a two-stage CRISPR protocol with positive-negative selection was independently published by another group and used to edit iPSCs and

hESCs[186]. The main difference between the published protocol and CIRCES related to the use of antibiotic selection. In Ikeda et al.'s protocol[186], positive selection is achieved using Magnetic-Assisted Cell Sorting (MACS, Fig. 5.7A). In their method, which is shown in fig. 5.7, a selection cassette is inserted into the target locus via HDR and drives the expression of a truncated version of surface marker CD19 and a mCherry fluorescent marker. Using a anti-CD19 antibody conjugated to magnetic beads, cassette-expressing cells are enriched by MACS and clonally isolated. Fluorescent colonies are picked, genotyped and re-targeted in a second round of CRISPR using a cassette-specific sgRNA and a second HDR template containing the final mutation to be introduced. Negative selection occurs by enrichment for tCD19-negative cells through MACS followed by clonal isolation and picking of non-fluorescent colonies. Overall their method is very similar to CIRCES, using plasmid-based delivery of Cas9, sgRNA and dsDNA HDR templates containing a 3.3kb cassette flanked by 400bp homology arms. This method has been used to introduce precise homozygous and heterozygous modifications in at least 4 different genes in both iPSCs and hESCs, highlighting that the theory behind CIRCES is correct and that with optimisations it too can likely achieve precise introduction of patient-derived mutations into human stem cells.

It is likely that positive selection can eventually be optimised to achieve complete selection. However, the excessive formation of indels during the first CRISPR round is a much more difficult problem to solve. Puzzlingly, in Ikeda et al's publication, several clones with heterozygous on-target cassette integration and wild-type non-cassette alleles were identified. Indeed, very few indels were described. The rate of biallelic cassette integration was also significantly higher, although their fluorescent method of detection was used to select clones with multiple cassette copies. While the proposed ssODN indel blocking strategy has given promising results, it was unlikely that it alone could reduce indels to a low enough rate for CIRCES to be a viable protocol for generating heterozygous missense mutations such as those found in SCA29, AD Gillespie Syndrome and in many other ataxia-causing genes that the Nemeth lab currently studies, such as EBF3[187], SPTBN2[28], KCNC3[28], GRM1[188].

At its core, the "indel problem" is likely to stem from excessive Cas9 cutting. It is unclear, however, why the same issue was not reported by Ikeda et al., and it might well be locus-specific.

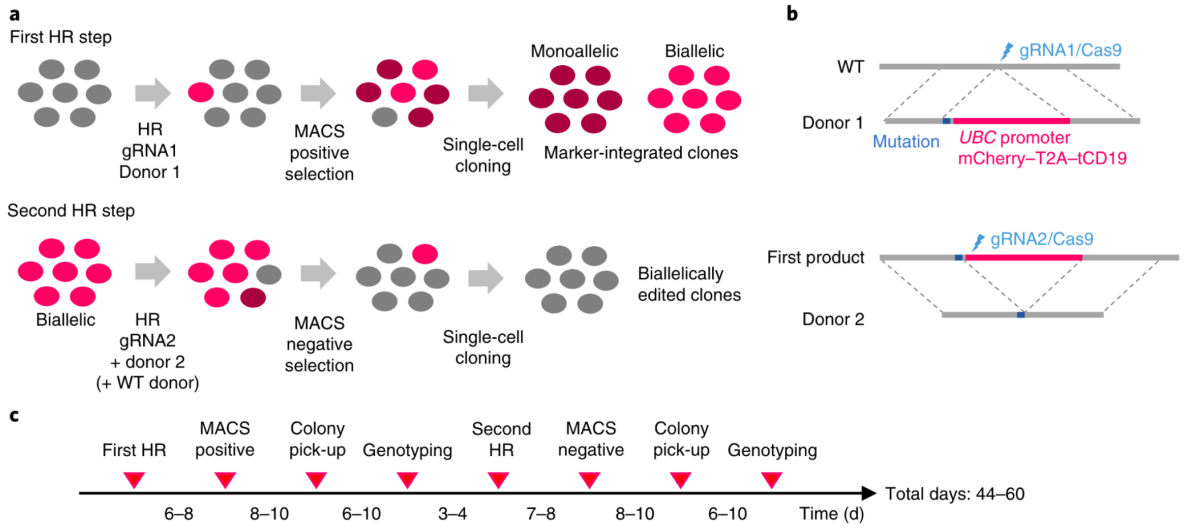

Figure 5.7: Two-step CRISPR HDR protocol as described by Ikeda et al. (A) The protocol uses double positive and double negative selection and includes a critical intermediary clonal isolation step. Cells are transfected with sgRNA/SpCas9 and HDR template plasmids. Their cassette uses a UBC promoter to drive bicistronic expression of expression of a mCherry fluorescent protein and tCD19. MACS is used to enrich for cells expressing the surface marker tCD19, which are then clonally isolated. Only colonies showing mCherry fluorescence are isolated, and their fluorescence intensity (dim or bright) can suggest monoallelic vs biallelic cassette integration. Genotyping is used at this stage to confirm correct sequences and optimal clones are expanded and used for a second CRISPR HDR step with sgRNA/SpCas9 targetting the cassette itself and HDR templates. Another round of MACS is used this time to enrich for tCD19-negative cells, followed by single-cell seeding and manual isolation only of non-fluorescent colonies. When biallelic cassette insertion was identified after the first CRISPR stage, heterozygous or homozygous precise edits could be achieved after the second CRISPR stage by using one or two simultaneous HDR templates, respectively. (B) Diagram showing the insertion and removal of the UBC-mCherry-T2A-tCD19 cassette. (C) Expected timeline of the Ikeda et al. protocol, highlighting that heterozygous mutations can be achieved within 44-60 days.

While the ssODN approach tried to block Cas9 cutting by means of altering its recognition sequences, another alternative would be to reduce the half-life of Cas9 within the cell. Theoretically it could be achieved by reducing the total amount of eSpCas9 plasmid transfected. More promising options would be to directly transfect either *in vitro* transcribed eSpCas9 mRNA[116] or Cas9 ribonucleoprotein[117]. Neither of these methods, however, would be compatible with the chemical transfection method already in use. Given these challenges and the time constraints associated with this project, a decision was made not to pursue Replacement Stage further.

Excessive indel formation, while a hindrance to the original aim of generating heterozygous missense mutations, turned out to be a significant advantage to the development of knockout cell lines. Biallelic mutations were identified in all previous experiments despite incomplete positive selection. As described, single-cell derived colonies carrying biallelic truncating mu-

tations could be isolated, expanded and genotyped within 4-5 weeks from initial transfection with minimal costs and hands-on time. Minor modifications to the protocol could make this time-frame even shorter. As no negative selection would have to be undertaken, the cassette can be made much smaller by removing the TK gene, the Drosophila-Specific Sequence and any unnecessary buffering sequences. These simple modifications would reduce total cassette size by 1152bp (38%). For an additional gain in efficiency, the full length EF1α promoter (1179bp) could be replaced by its short core promoter version (212bp) to achieve a cassette 70% shorter. Although untested at this point, it is possible that an alternative selection agent such as Zeocin[189] is not affected by this bystander resistance effect. If so, replacing the BSD gene for the BleoR gene that induces zeocin-resistance would lead to a total cassette length of 750bp, instead of 3kb. A short cassette does not require long homology arms[140], which could represent additional benefits during cloning and PCR-based screening. Indeed, at 750bp cargo it is possible to achieve high-rates of on-target insertion at least in HEK293T cells using only 30-40bp homology arms[140]. Although this remains to be tested in human stem cells, if confirmed it would greatly facilitate the process of assembling the HDR templates. For example, the entire HDR donor could be assembled using simple PCR techniques and delivered as a linear dsDNA or long ssDNA[190]. Alternatively, HDR templates could be synthesised commercially. Either way, it could make nearly 3 weeks of preliminary cloning work fully unnecessary.

Despite blasticidin's proximity protection, it might still be possible to accelerate clonal isolation of biallelic mutants even further. Replating transfected cells over double the surface area (42x) as the current protocol would be expected to reduce proximity between different independent clusters and thus reduce the number of contaminating non-cassette expressing cells. Since it was demonstrated that antibiotic selection must be kept during clonal isolation, blasticidin-resistant clusters could be dissociated, pooled, and seeded for clonal isolation a full week earlier. These modifications could shorten the total time between transfection and clonal isolation (including genotyping) to 3-4 weeks.

In summary, the CIRCES system, as designed, was not able to generate heterozygous missense mutations. Its Insertion Stage, however, proved unexpectedly efficient at generating biallelic

truncating mutations in human stem cells. In its current form, it can repeatedly identify several biallelic colonies with minimal colony screening within 4-5 weeks and using only standard lab equipment and inexpensive reagents. Small optimisations to cassette design and protocol are expected to lead to further improvements in efficiency. If pre-transfection cloning steps are considered, this method could be used to go from primer design to genotyped knockout stem cell lines in 8-10 weeks.

6 | Functional Effects of ITPR1 Knockout

6.1 Introduction

The Type 1 IP_3 receptor (IP3R1) is a key component of calcium signalling pathways of most cortical and cerebellar neurons. Whilst SCA15, SCA29 and Gillespie syndrome are directly caused by mutations in ITPR1, many other neurological conditions have been indirectly associated with its function (Fig. 1.1). For example, in SCA2, SCA3 and Huntington's Disease, mutant proteins were shown to bind IP3R1 and lead to increased sensitivity for IP_3[191–193]. Similarly, genetic reduction of IP3R1 in animal models of Familial Alzheimer's Disease was shown to normalise exaggerated calcium signalling and rescue biochemical, electrophysiological and behavioural phenotypes[194].

Given the importance of ITPR1 in the neurodevelopmental and neurodegenerative diseases and the current uncertainty over the mechanisms of SCA15, SCA29 and Gillespie syndrome, this project aimed to develop human cellular models that can be used to study ITPR1-related disorders. Although the main objective of this thesis was to develop the tools capable of generating such models, this chapter takes the first steps towards better understanding ITPR1 pathology using the cell lines engineered in the previous chapters.

Several H9 hESC lines with biallelic ITPR1 mutations were generated in CIRCES experiments. Of these, nine were chosen to be further characterised based on location and types of mutations (Table 6.1).

Two lines had biallelic mutations in exon 10, four lines had mutations in exon 59 and three lines had mutations in exon 59. The types of mutations varied. Most clones had one allele disrupted by the insertion of a cassette and the other allele compromised by an indel (hereby referred as cassette/indel, all were isolated in chapter 4). Two clones, both targeting Exon 58, had biallelic indels (e.g. indel/indel, both isolated in chapter 5). Indel/indel clones were included because they most closely mirror autosomal recessive Gillespie syndrome alleles[4] and, being more frequently generated than cassette/indel clones, they might be functionally equivalent to

cassette/indel clones. For most research applications where rapid gene knockouts are required, these biallelic indels might be sufficient, especially if several different clones are generated, thereby diluting potential phenotypic consequences of randomly integrated cassette copies.

ID	Target	Allele #1	Allele #2	Effect Predictions
A1	Ex10	c.796dupA	Cassette	p.(Thr266Asnfs*25) / Cassette Disruption
B1	Ex10	c.797_798delCCinsG	c.797_801delinsGTAAGGCC	p.(T266Rfs*53) / p.(266_267delTTinsSKA)
A4	Ex58	c.7698dupA	c.7682_7701+1del	p.(Glu2567Argfs*8) / Splicing Defect
B4	Ex58	c.7698dupA	c.7682_7701+1del	p.(Glu2567Argfs*8) / Splicing Defect
#5	Ex58	c.7698dupA	Cassette	p.(Glu2567Argfs*8) / Cassette Disruption
#6	Ex58	c.7698dupA	Cassette	p.(Glu2567Argfs*8) / Cassette Disruption
#1	Ex59	c.7713dupT	Cassette	p.(Ala2572Cysfs*3) / Cassette Disruption
#3	Ex59	c.7713dupT	Cassette	p.(Ala2572Cysfs*3) / Cassette Disruption
#8	Ex59	c.7712_7713dupTT	Cassette	p.(Ala2572Leufs*22) / Cassette Disruption

Table 6.1: Genotypes and effect predictions for the nine mutant ITPR1 H9 hESC clones that are characterised in this chapter. The term 'cassette disruption' was used to indicate that one of the alleles is expected to be disrupted by the cassette, although the exact consequence is difficult to predict.

Investigation of clones proceeded in three main phases.

First, undifferentiated H9 hESC clones were analysed to establish whether ITPR1 mutations effectively led to knockout of IP3R1. This was investigated at the mRNA level to determine whether mutant alleles were expressed or underwent nonsense-mediated decay. Possible alternative splicing was assessed and ITPR1 expression was quantified. At the protein level, loss of IP3R1 was assessed using subcellular fractionation followed by Western blotting to detect low level expression of membrane-associated proteins.

Given the existence of three different IP_3 receptors, the second phase investigated potential compensatory expression of other IP_3 receptor isoforms. It was possible that functional redundancy was obtained by expression of more than one receptor type in the same cell. As a recent study had identified expression of all three receptors in H9 hESCs[195], ITPR2 and ITPR3 transcripts were also quantified to assess whether loss of ITPR1 led to detectable compensatory changes.

The final phase was to differentiate mutant clones to neuronal cells and evaluate for functional consequences of ITPR1 knockout. The cerebellum is most clearly affected by mutations in ITPR1, as demonstrated by functional and animal studies and corroborated by the high expression of IP3R1 in Purkinje cells[196]. The ideal model to study ITPR1 mutations would

be knockout cell lines differentiated into cerebellar lineages (e.g. Purkinje cells) to recapitulate disease-relevant cellular phenotypes. However, reliable human cerebellar differentiation protocols were not available at the time this project commenced, and even today remain very challenging. Since ITPR1 is also expressed in the cerebral cortex[196], and intellectual disability is part of SCA29 and Gillespie syndrome[1, 24], experiments aimed to generate cells which have been found to be transcriptionally similar to excitatory neurons found in layers 2/3 of the cerebral cortex[197]. For that, the rapid neuronal differentiation protocol described by Zhang and colleagues[197] was used. It is based on the forced expression of the transcription factor neurogenin-2, which has been reported to generate reliably firing neurons approximately equivalent to the second trimester of brain development[198]. Since intellectual disability is a core feature of infantile-onset ITPR1-related syndromes, understanding whether loss of ITPR1 impacts early cortical neurons is critical. For that, differentiated cells were investigated for changes in the expression of IP_3 receptors, assessed for markers of neuronal identity and assessed for calcium responses to glutamatergic stimulation.

The results from these experiments raise further questions about current models of the molecular mechanisms of ITPR1 mutations and highlight some of the challenges inherent to these investigations. Based on the integration of these results with other data, a new theoretical model of ITPR1 pathogenesis is proposed, which will be testable with current cellular models and others that can be refined using the methodologies generated in previous chapters.

6.2 Materials and Methods

6.2.1 Analysis of ITPR1 Transcripts

Analysis of alternative splicing was performed on reverse transcribed RNA from H9 hESCs, with most steps described in subsection 6.2.2.

Primers were designed to span target mutation loci. For exon 10, primers binding in exons 7 and 11 were used. A single amplification reaction was used to investigate mutants in exons 58 and 59, with primers binding in exons 57 and 61. Multiple primer sets were initially designed, optimised and confirmed only to generate amplicons from cDNA, but not gDNA. PCR reactions were set up and imaged as described in general methods chapter.

A small aliquot of each amplicon was Sanger sequenced as described in the general methods chapter. For samples A4 and B4, which required band isolation prior to sequencing, a stab-PCR band purification approach was used. Under blue LED light, each band was carefully stabbed once with a 10μL plastic tip and used to stir 20μL of nuclease-free water in a polypropylene tube. One microliter was used as template in a repeat PCR reaction using the same amplification primers (exons 57-61) for 35 cycles. Purification and amplification were confirmed by agarose gel electrophoresis and an aliquot of each band was Sanger sequenced.

6.2.2 Quantification of RNA Transcripts

RNA was extracted from frozen pellets of H9 hESCs and induced day 23 neurons, harvested from matrigel-covered 6-well plates as described in 6.2.3.1. Each frozen pellet was kept on ice throughout, and 2X DNA/RNA Shield (Zymo) was added and and tubes were vortexed until thawed. RNA extraction was performed using Zymo Quick-RNA Miniprep kit according to the recommended protocol. Adult cerebellum total RNA was purchased from Takara Bio (cat. 636535).

RNA samples were quantified using a spectrophotometer (NanoDrop One, Thermo-Fisher) and 3μg of each sample were used immediately for cDNA synthesis using QuantiTect Reverse

Target Gene	Exons	Transcript	Catalog
ITPR1-DT	1-2	NR_108075.1	Hs04402651_m1
ITPR1 (5')	3-4	NM_001378452.1	Hs00181881_m1
ITPR1 (MID)	36-37	NM_001378452.1	Hs00976039_m1
ITPR1 (3')	61-62	NM_001378452.1	Hs00976063_m1
ITPR2	10-11	NM_002223.3	Hs00181916_m1
ITPR3	55-56	NM_002224.3	Hs01573555_m1
18S	1	X03205.1	Hs99999901_s1
GAPDH	2	NM_001289746.1	Hs99999905_m1

Table 6.2: Taqman hybridisation probes used for qPCR. All probes purchased from Applied Biosciences.

Transcription Kit (Qiagen) according to the recommended protocol, with the optional application of gDNA Wipeout buffer and with a 35 minute Reverse Transcription reaction incubation to favour production of longer cDNAs. Resulting cDNAs were stored at -20°C until needed.

Gene expression was assessed in cDNA samples by quantitative PCR (qPCR) using predesigned TaqMan hybridisation probes (Table 6.2), including two housekeeping genes (18S and GAPDH). 10μL Fast Advanced PCR Reactions (Applied Biosciences, cat. 4444556) were set up on ice and distributed to a 384-well plate using a digital multidispensing pipette, with three technical replicates for each combination of sample and hybridisation probe. Three no-template control (NTC) reactions were performed for each hybridisation probe using nuclease-free water instead of sample cDNA template. Reactions were run and quantified in a Roche Lightcycler 480 II real-time thermocycler.

Quantification cycle (Cq) values for each well were analysed using qRAT[199] (qPCR Relative expression Analysis Tool). Only technical replicates with Cq values between 5 and 35 and within 2 standard deviations for each combination of probe and sample were included in the analysis. Relative expressions were calculated according to the $\Delta\Delta$Cq method, calibrating to the geometric mean of two housekeeping genes (18S and GAPDH) and normalising to the level of expression found in wild-type samples. Fold change and standard deviation values generated by qRAT were plotted using custom scripts in Python and Matplotlib. Single comparison moderated T-test statistics were calculated for each sample against wild-type, and P-values were calculated and adjusted for multiple comparisons using the Benjamini & Hochberg method[200].

6.2.3 Detection of IP_3 Receptors

Multiple experimental conditions were used to optimise immunoblotting of IP_3Rs, which are high-molecular weight proteins of >300kDa and challenging to detect using Western blotting. The methods described below reflect the optimised conditions largely based on the details kindly shared by Dr. Lara Terry from Prof. David Yule's lab (University of Rochester Medical Center) and were used for the blots presented in this thesis.

6.2.3.1 Cell Harvesting

Cells for whole-cell protein extraction were grown on matrigel coated 6-well plates and included both H9 hESCs and day 23 Ngn2-differentiated neurons. H9 hESCs were harvested as described in the general methods section, centrifuged and the cell pellets were snap frozen and stored at -80°C until needed. Plates containing Ngn2-differentiated neurons were placed directly on ice, the medium was aspirated, cells were carefully rinsed once with ice-cold DPBS, and cells were manually scraped, transferred to pre-chilled tubes, and immediately snap frozen on dry ice.

6.2.3.2 Crude Subcellular Fractionation

H9 hESCs were subjected to a crude subcellular fractionation protocol[201] to enrich for microsomal proteins. Cells in 6-well plates were harvested with ReLeSR (StemCell) and transferred to pre-chilled polypropylene tubes on ice. Cells were pelleted by centrifugation at 300g for 5 minutes at 4°C and the supernatant was discarded. Cell pellets were resuspended in ice-cold Digitonin fractionation buffer and incubated on ice for 10 minutes on a shaker. The soluble cytosolic fraction was separated by centrifugation at 2000g for 10 minutes at 4°C and discarded. Remaining pellets, representing the combination of enriched microsomal, nuclear and mitochondrial fractions, were snap frozen on dry ice before being stored at -80°C until protein extraction.

The Digitonin fractionation buffer was always made fresh and contained 25μg/mL Digitonin (Chemcruz, cat. 11024-24-1), 150mM NaCl and 50mM HEPES pH 7.4. One tab of cOm-

plete Easy Protease Inhibitor Cocktail EDTA-free (Roche cat. 45-4693159001) and 250μL of 100μM PMSF were added to 25mL of fractionation buffer immediately before use.

6.2.3.3 Protein Extraction

Frozen cell pellets were thawed on ice for 15 minutes prior to the addition of ice-cold Tris-Triton extraction buffer (BioWorld, cat. 42020310-2) containing protease inhibitors. Samples were pipetted up and down 10 times to mix and incubated at 4°C in a rotatory shaker at 20rpm for 45 minutes. Insoluble debris were pelleted by centrifugation at 16,000g for 10 minutes. The cleared supernatants were transferred to prechilled new plastic tubes and kept on ice.

6.2.3.4 Establishing Protein Concentration

Protein concentration in each sample lysate was established using the Pierce BCA Protein Assay Kit (Thermo Scientific cat. 23227). Preliminary experiments had shown that the Tris-Triton Extraction Buffer interfered with the BCA chemistry up to 1:8 dilution. For that reason the BCA reaction was performed on aliquots of lysates diluted 1:10 in water and the BCA protein standards were also diluted using 1:10 dilutions of Tris-Triton Buffer (with Protease Inhibitors) in water. In total 11 different concentration standards were used and both samples and standards were analysed in triplicate in 96-well microplates using 8 parts BCA working reagent to 1 part sample. All H9 hESC and Ngn2-differentiated samples were analysed in the same plate. The plate was mixed at 800RPM for 1 minute on a plate mixer and incubated at 37°C for 30 minutes, before being allowed to cool down to room temperature for 5 minutes. Bubbles in wells were carefully popped using a 24G needle before the plate was read uncovered in a iMark Microplate Absorbance Reader (Bio-Rad) at 595nm using standard settings.

Absorbance results were analysed using Microplate Manager 6 software (Bio-Rad) to automatically substract readings from null standards and generate a 4-parameter logistic standard curve against which original sample concentrations could be inferred. After labelling and aliquotting, aliquots were stored at -80°C.

6.2.3.5 Protein Gel Electrophoresis

Protein lysates were diluted to the desired concentration in Tris-Triton Buffer (with protease inhibitors) and mixed with 5X Lithium Dodecyl Sulfate Protein Loading Buffer (National Diagnostics, cat. EC-887) supplemented with 10% beta-mercaptoethanol (2% final concentration). Samples were heated to 70°C for 10 minutes, followed by centrifugation at maximum speed in a 8-strip tabletop minicentrifuge for 3 minutes to pellet any insoluble debris.

Supernatants were loaded onto wells of Bio-Rad TGX Stain-Free Tris-Glycine 4-15% 1mm precast polyacrylamide gels in a Bio-Rad Criterion gel tank. At least one well was loaded with 5μL of BLUeye Prestained protein ladder. 1X Tris-Glycine-SDS Running buffer (National Diagnostics, cat. EC-870) was used and electrophoresis was started at 50V for 10 minutes followed by 150-200V for 40-60 minutes.

6.2.3.6 Electrophoretic Protein Transfer

The gel was washed in 1X Towbin transfer buffer (25mM Tris, 192mM glycine, 20% methanol, pH 8.3) for 15 minutes before it was assembled in a wet transfer stack cassette with a pre-activated (submerged in 100% methanol for 3 minutes) Immobilon LB-PVDF blotting membrane (Bio-Rad) sandwiched between permeable sponges. Ice-cold 1X Towbin transfer buffer (20% methanol) was used and an ice pack at -80°C was added to the tank to prevent overheating and a magnetic stirrer bar was added to aid with temperature and pH stabilisation. The tank was placed on a magnetic stand and stirring was maintained at the highest possible speed. Electroblotting was run at 90V for 90 minutes using a high-current powerpack (Bio-Rad).

6.2.3.7 Membrane Blocking and Antibody Incubations

The transfer stack was rinsed in TBST before being incubated in TBST for 15 minutes with gentle rocking to aid separation of gel and membrane. The gel was removed and the membrane was washed three times for 5 minutes each in fresh TBST before blocking in 5% skim milk in TBST for 1 hr at room temperature (RT) with gentle rocking. It was then incubated with primary antibody diluted in 5% milk TBST (table 6.3) for 18-24hrs at 4°C with gentle rocking.

The membrane was washed 4 times for 5 minutes in TBST with rocking before being incubated with secondary antibody diluted in 5% milk TBST (concentrations on table 6.3) for 1 hour at RT. Finally, it was washed 4 more times for 5 minutes each in TBST.

Primary Antibodies

Target Protein	Host Species	Clonality	Manufacturer	Catalog	Concentration
IP3R1 (C-terminus)	Mouse	Monoclonal	Neuromab	L24/18	1:1000
IP3R1 (N-terminus)	Rabbit	Monoclonal	CST	8568	1:1000
IP3R2	Mouse	Monoclonal	Santa Cruz	sc-398434	1:1000
IP3R-I/II/IIII	Mouse	Monoclonal	Santa Cruz	sc-377518	1:1000
Beta-Tubulin	Rabbit	Polyclonal	Abcam	ab6046	1:5000

Secondary Antibodies

Target Species	Host	Conjugate	Manufacturer	Catalog	Concentration
Mouse	Donkey	HRP	Abcam	ab205724	1:20,000
Rabbit	Donkey	HRP	Abcam	ab205722	1:10,000

Table 6.3: Antibodies used in Western Blotting.

6.2.3.8 Chemiluminescent Protein Detection

The rinsed membrane was incubated with Clarity ECL reagent (Bio-Rad cat. 1705060) for 5 minutes and imaged on a ChemiDoc MP Digital Imaging System (Bio-Rad) for up to 10 minute exposure. To capture the protein ladder markers, the membrane was immediately also imaged using fluorescent DyLight 680 settings, which capture all blue and green prestained markers.

6.2.3.9 Image Processing

Digital chemiluminescent and fluorescent images of each blot were merged on ImageJ by manually overlaying the protein ladder fluorescent image onto the corresponding region of the chemiluminescent image. Of note, this approach leaves visible edges between both images.

6.2.3.10 Multi-protein Detection

To reprobe the membrane using a different set of primary and secondary antibodies, the membrane was washed 4 times for 5 minutes in TBST with rocking before two incubations for 10 minutes in mild stripping buffer (15g/L glycine, 0.1% SDS, 1% Tween-20, pH 2.2) and then washed twice for 10 minutes in PBS followed by two washes for 5 minutes in TBST. At this point the membrane could be blocked and incubated with antibodies as described previously.

6.2.4 Lentivirus Production

Lentivirus were produced either by Dr. Hannah Sleven (Nuffield Department of Clinical Neurosciences) or by Dr. Ryan Beveridge (Weatherall Institute of Molecular Medicine, University of Oxford). HEK293T cells were plated at 7x10e5 cells/well in 6-well plates and co-transfected the following day with 888ng pLV-hEF1a-rtTA3 (Addgene #61472) or pLV-TRET-hNgn2-UBC-Puro (Addgene #61474), 666ng psPAX2 (Addgene #12260) and 444ng pMD2.G (Addgene #12259) using PEIpro (Polyplus cat. 101000017). Six hours post-transfection, culture medium was aspirated and replaced with fresh DMEM. Two days post-transfection the medium was aspirated and saved at 4°C (first lentiviral harvest) and fresh DMEM was added. The following day, the culture medium was collected (second lentiviral harvest) and combined with the saved first harvest. Harvested media were centrifuged at 2000g for 5 minutest at 4°C to pellet cellular debris and the supernatant was passed through a 0.45μm acetate cellulose filter, aliquoted and stored at -80°C.

6.2.5 Neuronal Differentiation

H9 hESCs were differentiated to neuronal cells by forced expression of neurogenin-2 (Ngn2). The protocol used was adapted from Zheng et al[197] with the help from Dr. Hannah Sleven, a post-doc in the Nemeth lab.

Dissociated wild-type or engineered H9 hESCs were transduced with pLV-hEF1a-rtTA3 (Addgene #61472) and pLV-TRET-hNgn2-UBC-Puro (Addgene #61474) lentiviral vectors in 6-well plates at 1.8x10e5 cells/well and expanded. On day -1, accutase-dissociated transduced cells in mTesR1 supplemented with ROCKi were plated onto matrigel-covered glass coverslips at either 2x10e5 or 6.7x10e5 cells/cm2 for Calcium Imaging/ICC or Protein/RNA extraction respectively.

On day 0, growth medium was replaced with the N2 Induction Medium (Table 6.4) and lentivirally-delivered Ngn2 expression was induced by doxycycline (4μg/mL). Medium was replaced on day 2 and puromycin (0.4 μg/mL) was added to eliminate non-Ngn2-expressing cells.

On day 3, medium was replaced with Neurobasal/B27 Medium (Table 6.5) containing doxycyclin and puromycin. The following day, puromycin was stopped and Cytosine β-D-arabinofuranoside (Ara-C 1.5uM) was started.

On day 7, doxycyline and Ara-C were stopped.

On day 10, 50% of Neurobasal/B27 medium was replaced with 1.5 volumes of BrainPhys/B27 medium (Table 6.6). This was repeated every 3-4 days until cells were analysed or harvested (days 23-25). Glial co-cultures were not used at any point in this protocol.

Reagent	Source	Catalog	Final Concentration
DMEM/F12 with GlutaMAX	Gibco	31331-028	1X
MEM NEAA	Gibco	11140050	1X
N2 Supplement	Gibco	17502-048	1X
Penicillin/Streptomycin	Gibco	15140-122	0.5X
Doxycyline Hyclate	Sigma	D9891-1G	4ug/mL
Y-27632 (ROCKi)	Abcam	ab120-129	10uM
Recombinant Human BDNF	Peprotech	450-02B	10ng/mL
Recombinant Human NT3	Peprotech	450-03B	10ng/mL
Puromycin	Invivogen	ant-pr-1	0.4ug/mL

Table 6.4: Composition of N2 Induction Media. Note that Puromycin is only added on day 2.

Reagent	Source	Catalog	Final Concentration
Neurobasal	Gibco	21103049	1X
Glutamax Supplement	Gibco	35050-061	0.5X
B27 Supplement, Serum Free	Gibco	17504044	1X
Penicillin/Streptomycin	Gibco	15140-122	0.5X
Doxycyline Hyclate	Sigma	D9891-1G	4ug/mL
Recombinant Human BDNF	Peprotech	450-02B	10ng/mL
Recombinant Human NT3	Peprotech	450-03B	10ng/mL
Puromycin	Invivogen	ant-pr-1	0.4ug/mL
Matrigel, Growth Factor Reduced (GFR)	Corning	356231	3.5uL/mL
Cytosine β-D-arabinofuranoside (Ara-C)	Sigma	C1768-100MG	1.5uM

Table 6.5: Composition of Neurobasal/B27 Medium. Note that puromycin is only used on day 3, doxycycline is only used between days 3 to 6, and Ara-C is only used between days 4 to 6.

6.2.6 Immunofluorescence

Cells were grown on matrigel-covered glass coverslips in 24-well plates. Fixation was achieved using 200uL of 4% paraformaldehyde (Thermo Pierce cat. 28908) in PBS for 10 minutes at RT. After fixation, PFA was diluted 1:5 with 800uL DPBS and plates were stored at 4°C.

Reagent	Source	Catalog	Final Concentration
BrainPhys Neuronal Medium	StemCell	05790	1X
B27 Supplement, Serum Free	Gibco	17504044	1X
Penicillin/Streptomycin	Gibco	15140-122	0.5X
Recombinant Human BDNF	Peprotech	450-02B	10ng/mL
Recombinant Human NT3	Peprotech	450-03B	10ng/mL
Matrigel, Growth Factor Reduced (GFR)	Corning	356231	3.5uL/mL

Table 6.6: Composition of BrainPhys/B27 Medium.

The Antigen Signal Enhancement (ASE) immunocytochemistry protocol[202] was used. Coverslips were washed 3 times in PBS, permeabilised in 0.5% Triton X-100 for 10 minutes at RT, blocked for 45 minutes at RT using ASE Blocking Solution (2% Donkey serum, 50mM Glycine, 0.05% Tween-20, 0.1% Triton X-100, 0.01% BSA in PBS) and incubated with primary antibodies (Table 6.7) for 18hrs at 4°C in ASE Dilution Buffer (10mM Glycine, 0.05% Tween-20, 0.1% Triton X-100, 0.1% Hydrogen Peroxide in PBS). After 3 washes with PBS-T (0.05% Tween-20 in PBS), coverslips were incubated with secondary antibodies (Table 6.8) in the dark for 1hr at RT in ASE Secondary Antibody solution (0.1% Tween-20 in PBS) and stained with 2μg/mL DAPI for 5 minutes at RT. Following 3 washes with PBS-T, coverslips were mounted onto glass slides using ProLong Diamond Antifade Mountant (Invitrogen cat. P36961) and allowed to cure at RT in the dark for 18-24hrs before being stored at 4°C until imaging. No-primary controls used wild-type coverslips prepared exactly as described, except that the ASE Dilution Buffer did not contain any primary antibodies.

Target Protein	Host Species	Clonality	Manufacturer	Catalog	Concentration
IP3R1	Rabbit	Polyclonal	Invitrogen	PA1-901	1:250
IP3R2	Mouse	Monoclonal	Santa Cruz	sc-398434	1:100
MAP2	Chicken	Polyclonal	Abcam	ab5392	1:1000
Tuj1/TUBB3	Mouse	Monoclonal	Santa Cruz	sc-80005	1:100
Synaptophysin	Rabbit	Polyclonal	Abcam	ab68851	1:500

Table 6.7: Primary antibodies used for immunofluorescence.

Target Species	Conjugate	Host Species	Manufacturer	Catalog	Concentration
Mouse IgG	Alexa 647	Donkey	Abcam	ab150111	1:500
Rabbit IgG	Alexa 555	Donkey	Abcam	A-31572	1:500
Chicken IgY	CF488A	Donkey	Biotium	20166-1	1:500

Table 6.8: Secondary antibodies used for immunofluorescence.

Confocal Z-stack images were acquired using an Olympus FV3000 confocal microscope. For

each combination of cell types, primary antibodies and objective, all imaging parameters were first optimised using wild-type cells and then checked for absence of signal in no-primary control coverslips.

6.2.7 Calcium Release Assays

Intracellular calcium release assays were performed on day 23-25 Ngn2-differentiated neurons using a ratiometric membrane-permeable calcium indicator and live-cell imaging on a confocal microscope set up for continuous perfusion-aspiration. Intracellular calcium release was triggered by exposing cells to DHPG, a selective group I metabotropic glutamate receptor[203].

Cells were loaded with 2μM Fura 2-AM (Invitrogen cat. F1221) in BrainPhys (without Phenol Red, StemCell cat. 05791) and incubated at 37°C for 45 minutes in a 5% CO2 incubator. The loading medium was aspirated and cells were rinsed once with BrainPhys before coverslips were carefully removed from culture plates and set up in the microscope. To allow complete de-esterification of intracellular dyes, coverslips were continuously perfused with Hank's Balanced Salt Solution for 20 minutes prior to imaging. Images were acquired with a 10x objective every 1 second at both 340nm and 380nm excitation. Cells were perfused sequentially with HBSS (for the first 60 seconds), DHPG 50μM in HBSS (for 120 seconds) and HBSS (for the next 300 seconds). All solutions contained calcium and magnesium unless stated otherwise.

6.3 Results

6.3.1 Impact of Mutations on Splicing

Since the aim of this work was to produce knockout cell lines, biallelic loss-of-function mutations should lead to significant disruption to RNA transcripts. Exons 10, 58 and 59, coding for critical regions of IP3R1, had been specifically chosen so that alternative splicing should not generate partially functional IP_3-gated calcium channels. Total RNA from mutants and WT H9 cells was analysed by RT-PCR with primers spanning target exons to investigate whether ITPR1 splicing was altered by either the presence of the cassette and/or indels.

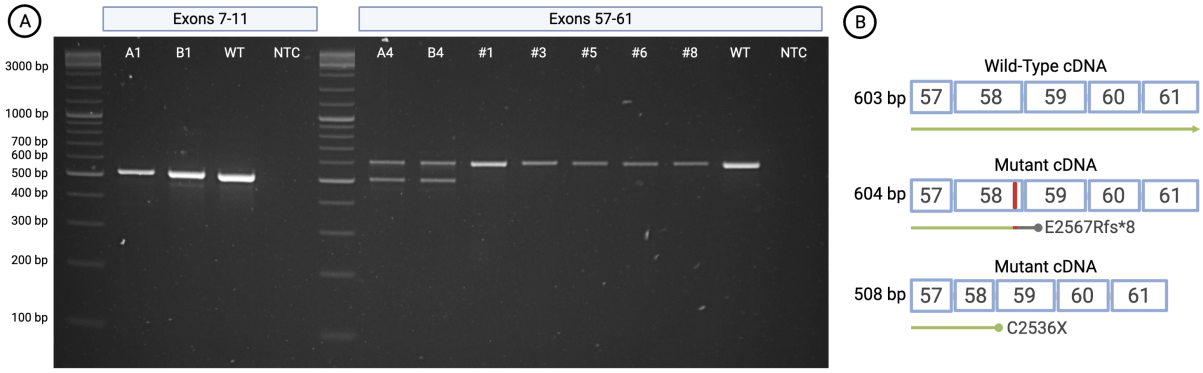

Figure 6.1: Analysis of ITPR1 transcripts produced from mutant clones. A) RT-PCR amplicons spanning the target exon for each mutant. B) Diagram of sequencing results from both amplicons identified in clones A4 and B4. The c.7698dupA allele produced a longer transcript with predicted functional consequence of p.Glu2567Argfs*8. The c.7682_7701+1del allele produced a shorter transcript using a cryptic donor splice site in exon 58 which immediately introduced a stop codon (p.Cys2536X). WT = Wild-type, NTC = No template control.

Abnormally spliced transcripts were identified in clones A4 and B4 (Fig. 6.1A), both which carry biallelic indels in exon 58 (c.7698dupA/c.7682_7701+1del) but do not have inserted cassettes. Sanger sequencing confirmed that transcripts produced from both alleles contained premature stop codons in clones A4 and B4. (Fig. 6.1B).

The alternatively spliced product identified in these clones was not surprising given that A4 and B4 clones carried a 21bp deletion that spanned the splice donor site for this exon. To check whether a near-full length protein product could be expected from these aberrant transcripts, both cDNA amplicons were individually purified and sequenced.

The longer sized amplicon (604bp), as expected, contained the single nucleotide insertion

(c.7698dupA) which is predicted to disrupt the reading frame and introduce a premature stop codon after 8 residues (Fig. 6.1B). The shorter amplicon (508bp) used a cryptic donor splice site internal to Exon 58, resulting in a 95bp-shorter transcript lacking most of Exon 58, leading to a frameshift and immediate incorporation of a stop codon (Fig. 6.1B). These results confirmed that clones A4 and B4 produced only transcripts expected to cause premature truncation of IP3R1.

cDNA amplicons in the other seven mutant lines revealed that all but two expressed only the frameshifted indel-containing allele expected to cause premature protein truncation. Alleles with an inserted cassette did not appear to be expressed and no occurrences of target exon skipping were identified.

Analysis of clone B1, which carries biallelic indels in Exon 10 (Table 6.1), detected transcription from both alleles (in-frame and out-of-frame indels). Clone #1, previously shown to have an inserted cassette and a frameshifting 1bp insertion in Exon 59, surprisingly produced a wild-type transcript, for unclear reasons.

Taken together, these results showed that except for lines #1 and B1, the other 7 mutants lines were possible knockouts for ITPR1.

6.3.2 Impact of Mutations on ITPR1 Expression

To assess whether mutant transcripts were still expressed at normal levels, ITPR1 expression in H9 hESCs was quantified by RT-qPCR using a TaqMan hybridisation probe to an upstream nonmutated region spanning the first two coding exons (Figure 6.2A).

Seven out of nine mutant clones expressed ITPR1 at significantly lower levels than WT H9 hESCs (Figure 6.2B, blue bars). Line #8 had the lowest expression at 28$\pm$6%, while most other lines had expression levels of only 30-40% ($P<0.01$). Lines #1 and #3 were exceptions, expressing ITPR1 at 74$\pm$8% and 78$\pm$12%, respectively (not statistically different from controls). Results were essentially identical when a different probe (hybridising to exons 36-37) was used (Figure 6.2B, orange bars).

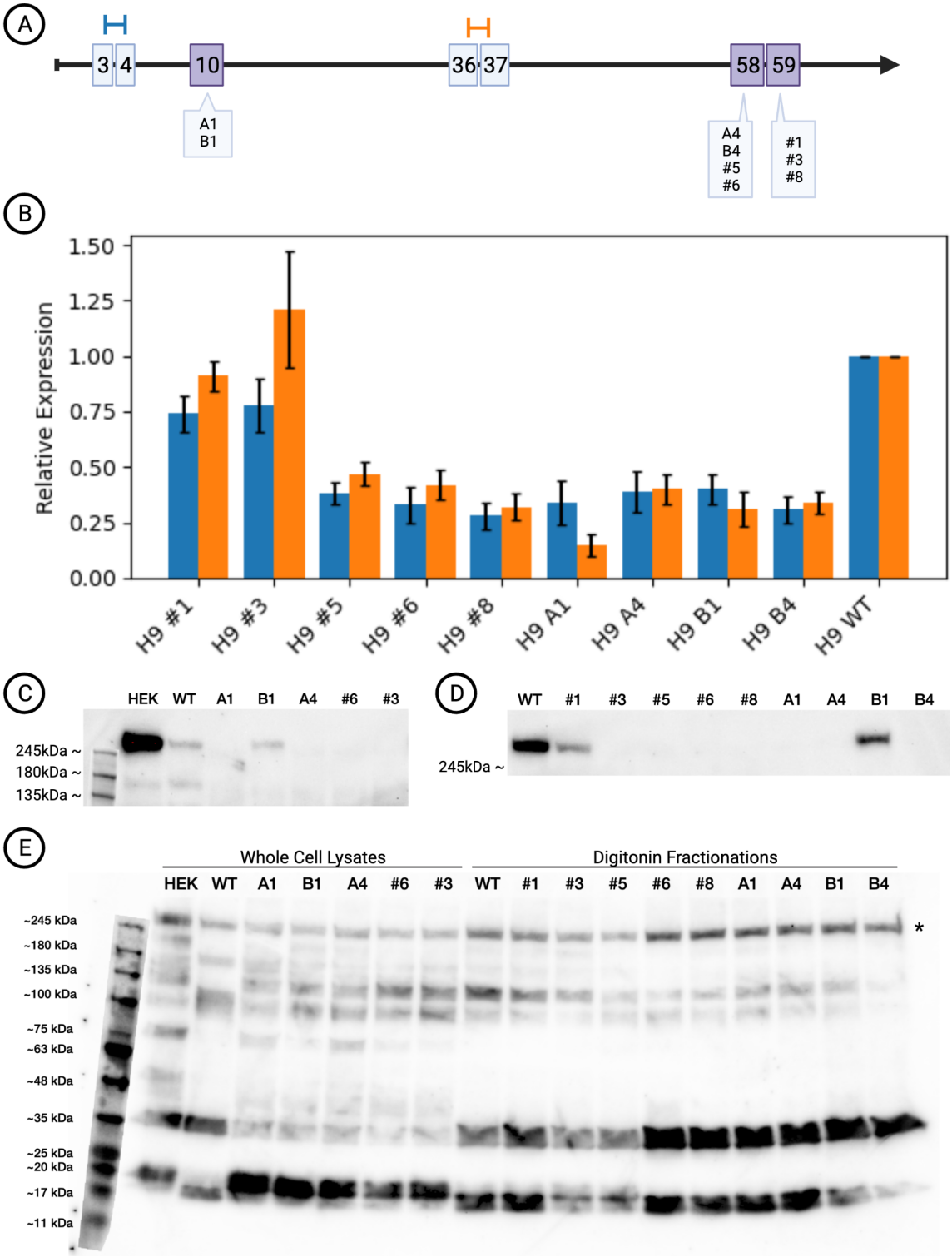

Figure 6.2: Expression of ITPR1/IP3R1 in mutant H9 hESCs. (A) Simplified exon structural diagram of ITPR1 transcript showing binding sites for TaqMan hybridisation probes used in [B] in relation to mutation sites. (B) ITPR1 expression in H9 hESCs. Bar colors match TaqMan probes shown in [A]. (C) Expression of IP3R1. Western blot performed on whole-cell lysates of five mutant H9 hES cell lines and controls using a C-terminus antibody. (D) Western blot performed on digitonin-fractionated mutant H9 hESCs using a C-terminus antibody. (E) Western blots on both whole cell lysates and digitonin fractionations using a N-terminus antibody. Cross-reactive IP3R1/IP3R2 (and possibly IP3R3) bands are highlighted by an asterisk. HEK = positive control whole-cell lysate of wild-type HEK293T cells, WT = wild-type H9 hESC.

These results confirm that, except for lines #1 and #3, all ITPR1 mutant H9 cell lines generated in previous chapters expressed significantly lower levels of ITPR1. Residual expression was not insignificant however, with all lines showing greater than 25% levels relative to controls. ITPR1 expression in lines #1 and #3 was also reduced, but not enough to achieve statistical significance. Given that #1 is likely to be a mixed population containing wild-type cells, this was not surprising. The explanation for line #3, on the other hand, is not clear.

6.3.3 Impact of Mutations on IP3R1 Expression

Given that indel-containing ITPR1 transcripts seem to partially evade NMD, mutant cells lines representing different genotypes were investigated for the presence of IP3R1 protein.

H9 whole-cell protein lysates were immunoblotted with a IP3R1-specific monoclonal antibody raised against the protein's C-terminus. This antibody would detect any proteins containing the channel domain, including alternatively-spliced variants.

No bands consistent with IP3R1 were detected in mutants A1, A4, #6, or #3 (Figure 6.2C), which collectively represent biallelic loss-of-function mutations in all three target exons. Line B1, which carries both in-frame and frameshifting indels in trans, did express full length IP3R1, as expected.

Using a N-terminus antibody which was crossreactive with IP3R2 and possibly IP3R3, no bands clearly suggestive of truncated proteins were identified (Fig. 6.2E).

These results suggested mutant lines to be IP3R1 knockouts (except B1 and probably #1, albeit untested in this experiment). Since WT H9 hESCs only showed minimal levels of IP3R1 expression, a digitonin-based crude subcellular fractionation protocol was used to enrich for microsomal and nuclear proteins, thus increasing sensitivity for lowly expressed products. Western blots showed that the fractionation protocol greatly improved detection of IP3R1 in wild-type H9 cells (Figure 6.2D).

Despite much higher sensitivity for detecting low abundance proteins, no full-length or truncated IP3R1 was detected in the mutant lines except for samples #1 and B1 (Figures 6.2D

and 6.2E). Notably, no IP3R1 was detected in line #3, despite the absence of significant NMD degradation of its ITPR1 transcript (Fig. 6.2B,D,E).

These results confirmed that 7 out of 9 mutant lines were complete, rather than partial, knockouts for full length IP3R1. It also showed that human embryonic stem cells tolerated total loss of IP3R1. Cell lines A1, A4, B4, #3, #5, #6 and #8 can therefore be used as human stem cell models of Gillespie syndrome.

6.3.4 Impact of Mutations on Expression of Other IP_3 Receptors

The previous experiments have shown that IP3R1 is expressed in wild-type undifferentiated H9 hESCs and that seven of the mutant cell lines are complete IP3R1 knockouts. Considering that H9 hESCs are known to express other types of IP_3 receptors, redundancy and/or compensatory upregulation of ITPR2/ITPR3 were investigated as possible mechanism of cellular survival and maintenance of pluripotency.

For that, expression of the three known IP_3 receptor genes (ITPR1, ITPR2 and ITPR3) was analysed by qPCR.

Expression of all three genes was detected in all samples. ITPR2 was the dominant type, closely followed by ITPR3. ITPR1 was the least expressed, at levels 2.7-fold lower than ITPR2 in wild-type cells (Figure 6.3A).

Expression of IP3R2 was confirmed on all samples using Western Blot (Figure 6.3B). Statistically significant upregulation of ITPR2 or ITPR3 was not found in any of the mutants (Figure 6.3C), suggesting that human embryonic stem cells are tolerant to total IP3R1 loss-of-function without compensatory increase from other IP_3 receptors.

6.3.5 Impact of Mutations on Neuronal Differentiation

IP3R1 is known to be the dominant isoform in the cerebral cortex, and intellectual disability is a main feature of both Gillespie Syndrome and SCA29.

To investigate whether complete loss of IP3R1 prevents neuronal differentiation or leads to

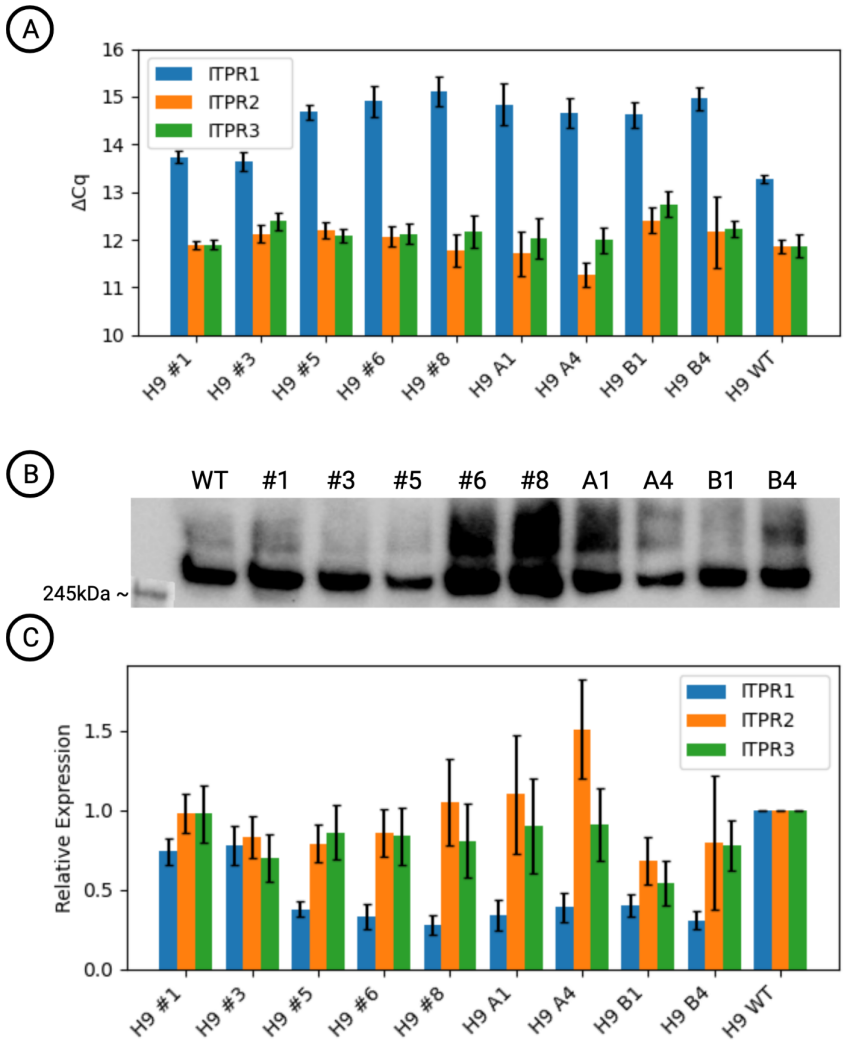

Figure 6.3: Expression of IP_3 receptors in mutant H9 hESCs. A) Quantification cycles (Cq) of ITPR1, ITPR2 and ITPR3 transcripts as measured by qPCR. Values shown have been normalized to the geometric mean of housekeeping genes GAPDH and 18S, with lower values denoting higher expression. B) Western blot for IP3R2 in digitonin-fractionated H9 hESCs. C) Expression of ITPR1, ITPR2 and ITPR3 mRNA in mutant H9 hESCs relative to wild-type.

a detectable phenotype in neuronal cells, all nine identified mutant ITPR1 H9 hES cell lines were differentiated using the forced neurogenin-2 (Ngn2) expression protocol that has been previously shown to generate cells resembling early cortical neurons[197].

H9 clones were transduced with a lentivirus carrying hNgn2. Treatment with doxycyline induced expression of Ngn2, leading to a rapid shift from undifferentiated (Fig. 6.4A) to clear neuronal morphologies (Fig. 6.4B) within 96hrs of induction. Cellular density was highly variable both inter and intra genotypes, with no particular pattern distinguishing mutants and wild-type. By day 21, all mutants showed neuronal morphological characteristics that were indistinguishable from controls (Figs. 6.4C-D).

After 23 days, both mutants and controls widely expressed the neuronal marker Tuj1 (Figs. 6.4E-F), indicating successful neuronal differentiation. MAP2, a marker of neuronal maturity, was similarly detected (Figs. 6.4G-J).

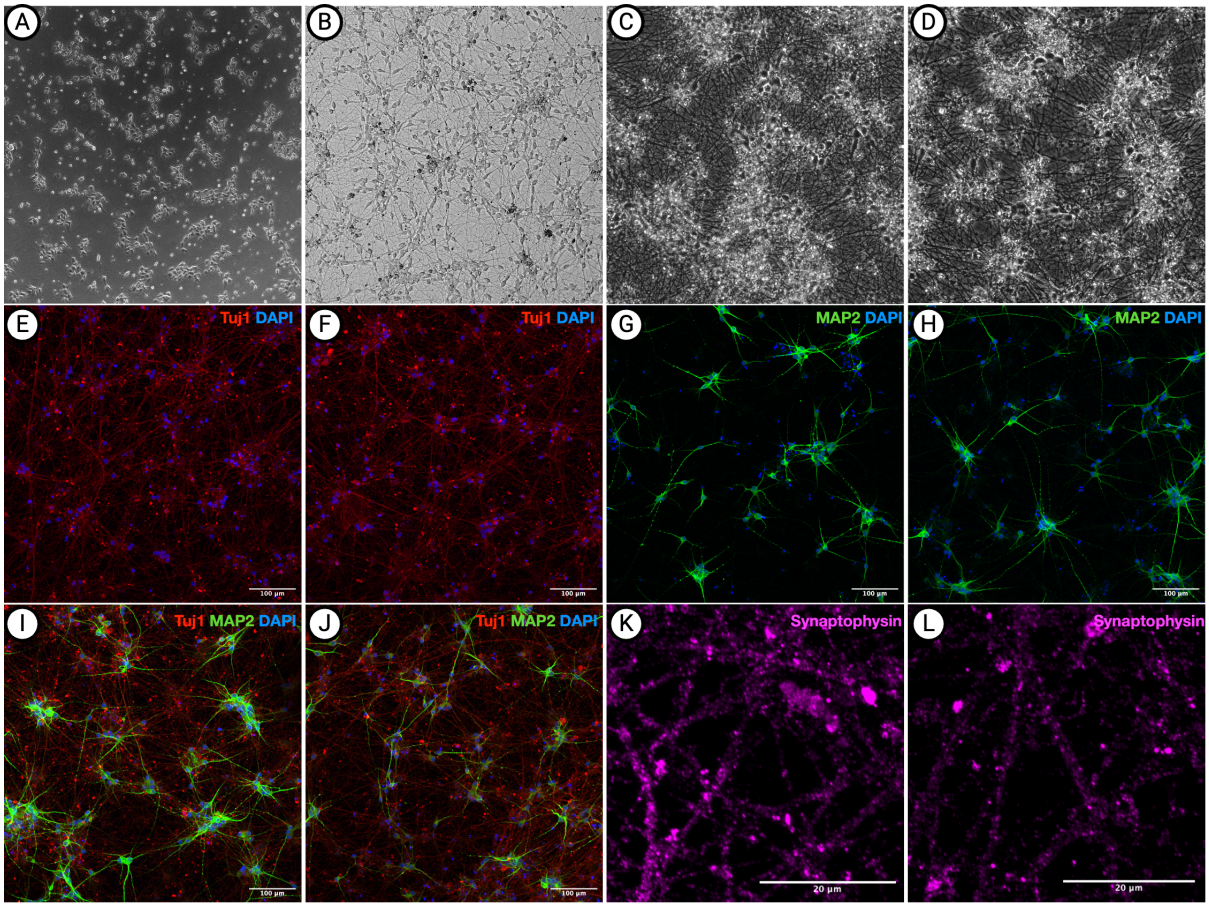

Figure 6.4: Neuronal differentiation of mutant H9 hESCs. A) Undifferentiated H9 hES cells immediately prior to neuronal induction (20x). B) On the fourth day of differentiation, virtually all cells display early neuronal morphology (20x). C) Clone #3 on differentiation day 21 (20x). D) WT on day 21 (20x). E) Immunofluorescence of day 23 clone A1 showing abundant Tuj1 staining (20x). F) WT day 23 (20x). G) Clone 06 day 23 showing MAP2 staining (20x). H) WT day 23 (20x). I) Clone A1 day 23 (20x). J). WT day 23 (20x). K) Clone A1 day 23 showing synaptophysin staining (60x). L) WT day 23 (60x).

Despite the lack of glial co-culture, which is thought to be required for development of mature synapses, both mutants and controls expressed the synaptic marker synaptophysin (Figs. 6.4K-L). These results indicated that ITPR1 knockout H9 hESC lines successfully differentiated to MAP2+ neuronal cells without any overt morphological phenotypes.

Given the absence of obvious immunocytochemical differences between knockouts and wild-type, expression of all three IP_3 Receptors was quantified in a subset of differentiated clones representing Exon 10 (A1), Exon 58 (A4) and Exon 59 (#8) biallelic mutations.

Unexpectedly ITPR1 expression was found to have decreased >30-fold during differentiation from hESCs to early neurons (Figure 6.5A), both in mutants and controls. A similar decrease was identified using different ITPR1 hybridisation probes (Figure 6.5A). Notably, expression of ITPR2 and ITPR3 also decreased in a similar magnitude and did not show compensatory upregulation in mutants relative to controls (Figure 6.5C). Despite the decrease, ITPR2 was still the most expressed IP_3 receptor gene in Ngn2-differentiated cells, followed by ITPR1 and ITPR3.

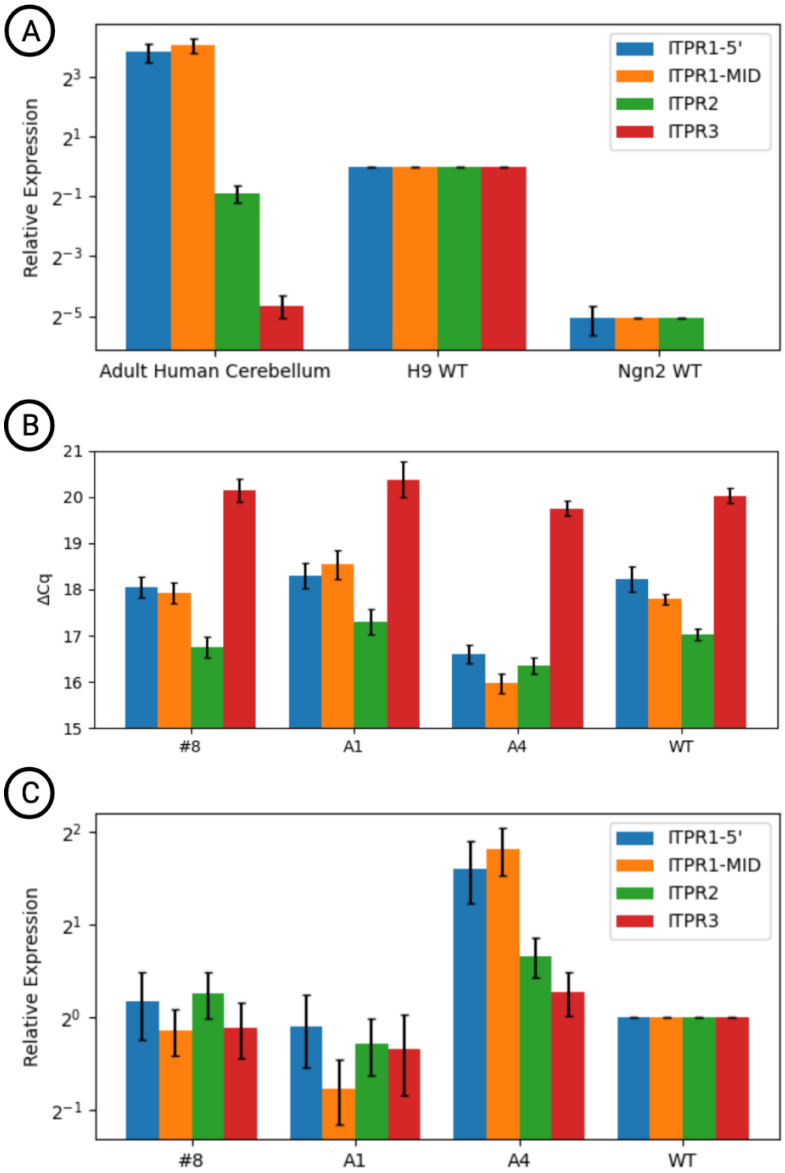

Figure 6.5: Expression of IP_3 Receptors in Ngn2-neurons. A) Expression in wild-type cells. B) Absolute expression in mutant cell lines representing all three target exons. Values shown have been normalized to the geometric mean of housekeeping genes GAPDH and 18S, with lower values denoting higher expression. C) Expression in mutant cell lines relative to wild-type Ngn2-neurons. Two different probes were used for ITPR1. Probe ITPR1-5' spans exons 3-4 while probe ITPR1-MID spans exons 36-37 (as seen in fig. 6.2A.)

As the expression level of ITPR1 in knockout lines was not reduced compared to controls, this suggested ITPR1 expression had essentially stopped. To confirm, whole-cell lysates of Ngn2-differentiated neurons were probed for IP3R1 by immunoblotting with both C- and N-terminal antibodies. No IP3R1 was detected in lysates of either wild-type or mutant neuronal cells (Figs. 6.6A,6.6B). Similar immunoblotting experiments were unable to detect IP3R2 (Figure 6.6C). These results were unchanged when long exposure times were used (Figure 6.6D). A pan-IP_3R antibody also failed to identify any signal, suggesting negligible expression of all three IP_3 receptors in these cells (Figure 6.6E).

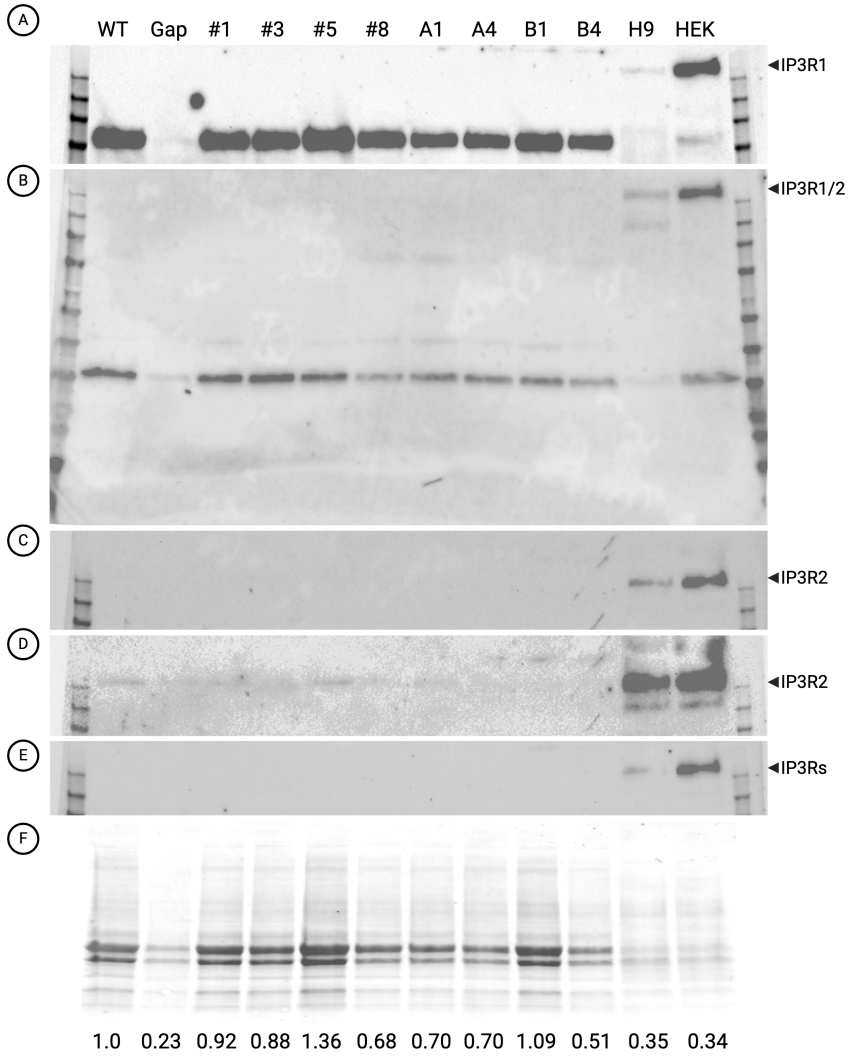

Figure 6.6: Western blots showing expression of IP_3 Receptors in whole-cell lysates of Ngn2-neurons. A) Antibody binding to the C-terminus region of IP3R1. The strong 110kDa bands are shown for reference, but are thought to be the result of nonspecific binding. B) Immunoblotting for IP3R1 using a N-terminal antibody which is cross-reactive with IP3R2. C) Antibody binding to the C-terminus region of IP3R2. D) Longer exposure of [C]. E) Pan-IP3R antibody. F) Total protein staining of PVDF membrane used for blots A-E. The numbers below each lane represent the calculated ratio of total protein present in each lane relative to the WT lane. Maximum amounts of protein were loaded per lane, limited by the total volume of the well and concentration of each lysate. Lane WT was loaded with 80μg whereas lanes H9 and HEK were loaded with 40μg each. WT = wild-type Ngn2-differentiated cells, Gap = accidental gap while loading the gel, H9 = whole-cell lysate of wild-type H9 hESCs, HEK = whole-cell lysate of wild-type HEK293T cells (positive control).

6.3.5.1 IP_3-Induced Calcium Release

The extremely low expression of IP_3 receptors in Ngn2-neurons was surprising considering the important role of IP_3-mediated calcium signalling in neurons. To functionally test whether Ngn2-neurons display calcium release in response to intracellular IP_3, two ITPR1 knockout lines and a wild-type control were loaded with a ratiometric intracellular calcium indicator dye and subjected to live-cell calcium imaging. Production of intracellular IP_3 was induced

by extracellular application of S-3,5-Dihydrophenylglycine (DHPG), a selective and potent agonist of group I metabotropic glutamate receptors (mGluR1 and mGluR5)[203].

All lines tested exhibited calcium release in response to 50μM DHPG (Figure 6.7A). The response was not significantly different between mutants and wild-type (Figure 6.7B). Calcium release was less prominent but still noticeable under calcium-free conditions (data not shown).

Taken together, these results indicated that despite the minimal presence of IP_3 receptors indicated by qPCR and Western Blot data, Ngn2-neurons displayed functional calcium release suggestive of residual IP_3 receptor activity. Knockout of IP3R1 did not seem to affect calcium release in these cells, possibly due to functional redundancy with IP3R2 even at very low expression levels.

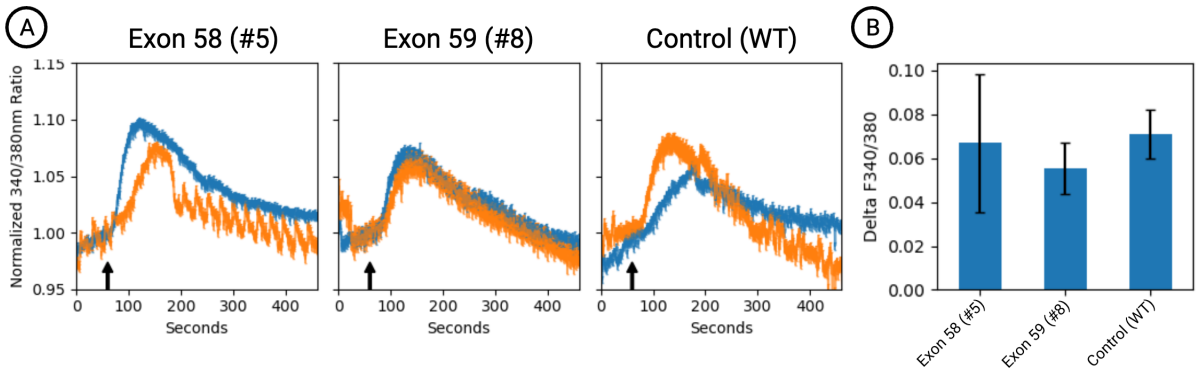

Figure 6.7: Calcium release assays in Ngn2-neurons. A) Knockout and wild-type Ngn2-neurons were loaded with Fura-2 AM and stimulated with 50μM DHPG for 2 minutes, starting at time 60 seconds (black arrows). Each coloured curve indicates the median$\pm$SEM of calcium transients of a separate coverslip, with a minimum of 139 cells per coverslip. B) Median peak amplitudes and SEM of calcium transients relative to the frame immediately prior to the DHPG stimulation.

6.4 Discussion

ITPR1 is a key gene in calcium signalling pathways that is involved in a number of ataxia syndromes. The generation of knockout human embryonic stem cell lines has the potential to further our understanding of IP3R1 in the development of ataxias and related neurological conditions.

Using the approach described in the previous chapters, several different cell lines carrying loss-of-function mutations in ITPR1 were generated, with mutations affecting three different loci in the ITPR1 gene. Nine cell lines with biallelic mutations were investigated further.

Analysis of ITPR1 transcripts in mutant lines revealed that cassette-containing alleles do not seem to be expressed, whereas indel-containing alleles are. It is however possible that whole-exon disruption as achieved by cassette insertion might lead to more complicated splicing changes that can only be detected by techniques such as northern blot or long-read next-generation sequencing. This was not done due to the complexity involved, and this represents an important limitation of this work. One clone was found to express wild-type transcripts, despite having a cassette/indel genotype on genomic DNA. Although the reasons for it are unclear, it is possible it is a mixed clone containing a sub-population of WT cells, highlighting the importance of sequencing transcripts from any isolated mutant clones.

Expression of ITPR1 transcripts was significantly reduced in most clones, consistent with nonsense-mediated degradation of mutant alleles. Residual expression was not insignificant however, with most lines expressing ITPR1 at levels approximately 25% those found in WT H9 hESCs. Interestingly, the only clone with perfect insertion of the cassette, which precisely replaced exon 59 without any additional mutations, expressed ITPR1 at levels similar to controls. As this clone did not produce shorter alternatively spliced products between exons 57 and 61, it is likely that the higher than expected expression occurs from the indel-containing allele. The reasons for it are not clear, but another possibility is that it too represents a mixed clone, containing a minor population of WT/indel amongst cassette/indel genotypes.

The variability between clones highlights one of the major benefits using the CIRCES method

to engineer knockout cell lines. As it is possible to rapidly and inexpensively generate several knockout cell lines carrying slightly different genotypes, random or artefactual inter-clone variability becomes distinguishable from *bona fide* genotype-derived cellular phenotypes. In other words, consensus results are more likely to represent true phenotypes than results from each individual clone. For example, clones #3 and #8 both have a cassette/indel genotype in ITPR1 Exon 59, yet clone #3 strongly expressed ITPR1 (fully escaping NMD) while clone #8 showed strong NMD and the lowest levels of ITPR1 transcripts. Interpreting these contradictory results was much helped by analysing the other five clones (carrying similar biallelic loss-of-function variants in other critical ITPR1 exons), which rapidly pointed towards a consensus of 60-70% reduction in ITPR1 expression in mutant clones.

Undifferentiated hESCs had previously been shown to release calcium from internal stores in response to extracellular ATP and Dopamine, suggestive of IP_3 Receptor activity[195, 204–206]. Microarray data had shown that ITPR2 and ITPR3, but not ITPR1, were expressed in undifferentiated RH1 and RCM1 hESCs[207]. Using RT-PCR, Huang et al. had detected mRNA transcripts from all three IP3R types in H9 hESCs, but could only confidently detect IP3R2 and IP3R3 by Western Blot[206]. The results presented in this chapter, however, clearly showed that wild-type H9 hESCs do express IP3R1, although at lower levels than IP3R2. Seven out of the nine clones analysed, however, were found to be complete IP3R1 knockouts using sensitive Western blot techniques, and loss of ITPR1 was not functionally compensated by increased expression of ITPR2 or ITPR3. ITPR1 hESC knockouts were viable, morphologically and functionally indistinguishable from wild-type, despite existing pharmacological evidence that calcium signalling impacts proliferation and self-renewal of embryonic stem cells[208]. These results corroborate recent studies which have shown that genetic manipulation of other calcium signalling proteins such as IP3R2[209], IP3R3[210], STIM1[211] and Orai1[211] does not impair self-renewal and proliferation of undifferentiated ES cells. Tellingly, knockout of all three IP_3 receptor genes has not shown any discernible effect in undifferentiated mouse ES cells[212].

This is not to say that IP_3 receptors have no developmental role. ITPR1 knockout mice are not born at expected Mendelian ratios, and >75% of homozygotes die *in utero*[75]. The ones that

are born, however, behave normally for the first few days and have brain and peripheral tissues that are similar to controls both anatomically and at a light-microscopy level[75]. Contrasting to other traditional cerebellar mutant mice (staggerer[213], pcd[214], nervous[215]), these animals have no detectable cerebellar malformations, yet all of them develop severe rapidly progressive ataxia that is universally fatal by their fourth week. The spontaneous *Opisthotonus* (*opt*) mice have a very similar, slightly delayed yet still fatal, phenotype, and were found to be homozygous for a large in-frame deletion that completely removes two exons that encode for 108 aminoacid residues in the regulatory domain of IP3R1[216]. While ITPR1 mRNA levels in the brain of *opt* mutants were similar to wild-type, IP3R1 protein levels were 90% lower. Careful analysis of existing animal and human data led to the hypothesis that minimal residual IP3R1 function was required for viable embryologic development, but total loss of function would be embryologically lethal. Under that hypothesis, minimal residual IP3R1 expression was present but not detected in the Western Blots of the original ITPR1 knockout mouse[75]. This would be theoretically possible as the genetic knockout in those mice was created by replacing the first *Itpr1* coding exon with an antibiotic cassette, but it is possible that downstream alternative transcription start sites were used and a minimally viable protein produced. Cerebella of homozygous mutants indeed expressed significant amounts of a slightly shorter transcript missing only the first exon[75]. Interestingly, when the genetic background of ITPR1 knockouts was shifted from the 129-C57B1/6 to the CD-1 mouse strain, the birth rate of homozygotes more than doubled[75].

Could it be that the homozygotes that survived through the gestational period represented the minority of animals in which, due to genetic background, a minimal viable threshold of residual IP3R1 function was produced? The results presented in this chapter suggest that total loss of IP3R1 might indeed be tolerated in the early stages of development. By creating multiple hESC cell lines with biallelic loss-of-function mutations in function-critical gene regions and using sensitive protein and transcript detection methods, the possibility of residual calcium-releasing protein function was confidently dismissed. But despite complete IP3R1 knockout, cells not only proliferated normally as undifferentiated stem cells but also differentiated normally to neuronal cells.

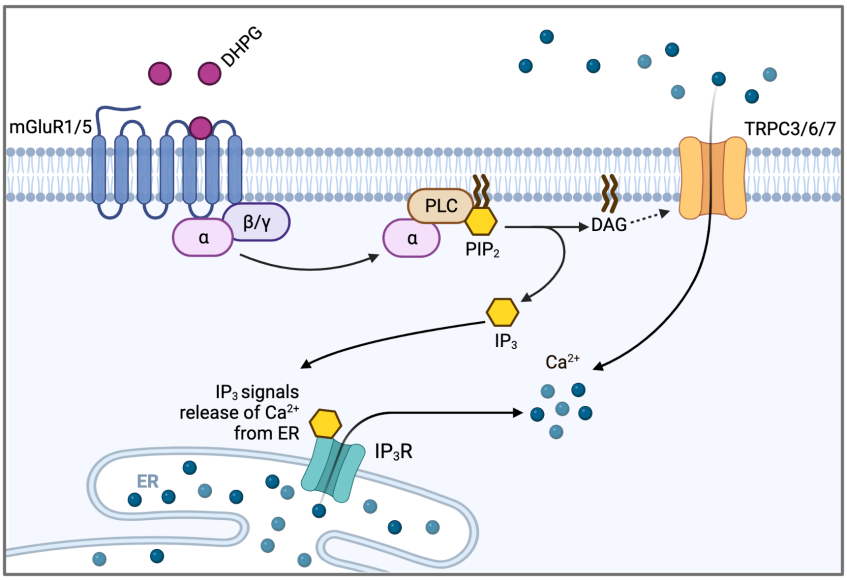

Figure 6.8: IP_3 Calcium Pathway. Upon binding of extracellular glutamate or DHPG to metabotropic glutamate receptors (mGluR1 or mGluR5), intracellular G protein activates phospholipase C (PLC), which cleaves PIP2 into IP_3 and DAG. IP_3 binding to IP_3 Receptors triggers calcium release from the ER to the cytosol. DAG can directly activate plasma membrane calcium channels such as TRPC3 to allow influx of extracellular calcium to the cytosol.

Type 1 is seen as the brain or neuronal IP_3 Receptor, and for that reason the original expectation was that its expression would increase as stem cells differentiated to neurons. Thus it was surprising to find that the opposite occurred. Three weeks after neuronal induction, expression of ITPR1 in wild-type cells had reduced nearly 30-fold relative to undifferentiated hESCs, reaching a level that was not significantly different from knockout cells. ITPR2 and ITPR3 also massively decreased, and no IP3Rs could be detected on Western blots. These results are supported by very recent, not yet through peer review[217], proteomic data in a similar glutamatergic Ngn2 neuronal differentiation protocol (Figure 6.9). At 1/8th the level found in the adult human cerebellum (Fig. 6.5A), wild-type undifferentiated hESCs barely express ITPR1 (Fig. 6.2C), which suggested that ITPR1 expression in Ngn2-neurons had all but stopped. Calcium release assays however showed that both mutant and wild-type Ngn2-neurons released intracellular calcium in response to mGluR1/5 stimulation by the selective agonist DHPG. The increase in intracellular calcium could be due to extracellular influx through DAG-activated transmembrane TRPC3/6/7 channels[42] (Figure 6.8), rather than IP_3-induced calcium release from intracellular stores through IP3Rs. But as the intracellular calcium responses to DHPG occurred even in the absence of extracellular calcium, it is more likely that residual IP3R2 and/or IP3R3 (though not detected on Western blots) provides significant functional redundancy for

IP3R1 loss even when all three receptors are expressed at very low levels.

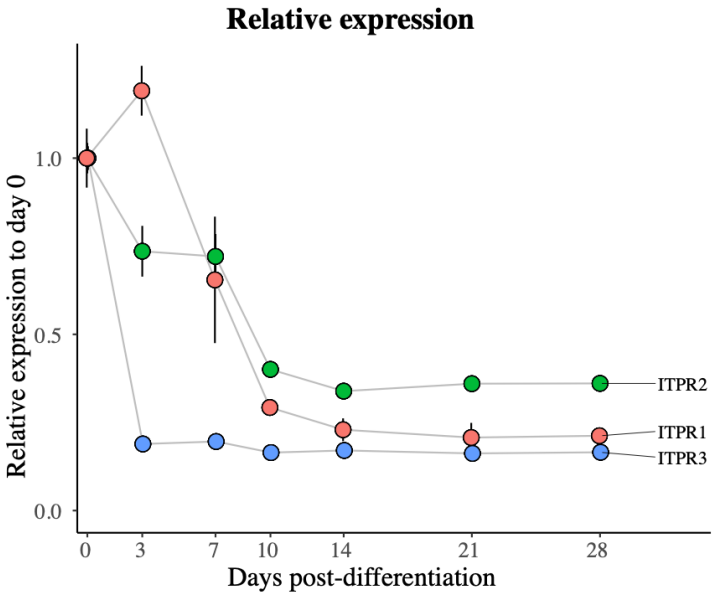

Figure 6.9: Expression of IP3Rs in proteomic analysis of Ngn2-differentiation. Data from Reilly et al, 2022.[217]

These results suggest that ITPR1 is not essential for early neuronal development. However, significant limitations apply to this interpretation. First, detailed morphological, neurophysiological or transcriptomic analyses of mutant Ngn2-neurons were not carried out, and it is possible these would have revealed subtle abnormalities. Second, the Ngn2-differentiation protocol, although fast, does not recapitulate the early transitional stages of neuronal development and does not generate glial cells. For that, dual SMAD inhibition[218] or organoid-based[219] protocols would be required. Third, as Gillespie syndrome and SCA29 are primarily diseases of the cerebellum, the ideal experimental approach would have been to use cerebellar differentiation protocols. Although these were not considered viable for this project at the time, recent work by several groups[219–221] has led to significant advances in the development of such protocols, which could be applied to these mutant H9 cell lines in the future. Finally, even if total IP3R1 knockout is shown not to affect neuronal development, it is possible that the high embryonic lethality of ITPR1 knockout animals be due to developmental defects in other organ systems, particularly in fetal-placental vascular connections [222, 223].

In summary, the novel CIRCES approach developed in the previous three chapters was used to engineer several human embryonic stem cell lines that were shown to be complete knock-

outs for IP3R1. Lack of this IP_3 receptor did not seem to affect proliferation and self-renewal of undifferentiated stem cells and was tolerated during neuronal differentiation. These mutant cell lines can be used as stem cell models of Gillespie syndrome to further advance the understanding of ITPR1-related disorders, but neuronal differentiation protocols that generate IP3R1-dominant lineages will be required to confidently establish a cellular phenotype that can be used for mechanistic research or drug development. These proof-of-concept experiments demonstrate that CIRCES is a useful tool that can rapidly and inexpensively generate knockout stem cell models. This approach can be applied to virtually any gene and should be considered especially by researchers working on autosomal recessive genetic disorders.

7 | Discussion

7.1 Development of a CRISPR-Based Genome Engineering System

Advances in CRISPR/Cas9 technology have allowed for cellular models to be generated for many human diseases, but the methodologies required are still time consuming and expensive. This included the reprogramming step if models were to be based on specific patient mutations. At the start of this project, there were few, if any methodologies that would enable rapid generation of cellular models. The experience from a previous post-doc in the Nemeth Lab demonstrated that even if the CRISPR/Cas9 step could be performed, many hundreds of stem cell clones would need to be screened before one or a few mutants could be identified. This led to to the original concept of the two-stage positive-negative selection CRISPR protocol. My project aimed to develop this idea, with ITPR1 as a model gene.

By adapting decades-old homologous recombination genome engineering techniques[86] to recent CRISPR/Cas9 advances[99], the first schematics of the CIRCES method were created. In its early designs, the cassette contained two Drosophila-specific sequences in tandem as it was intended to be paired with SpCas9n 'nickase' to achieve staggered single-stranded breaks[103]. At the time of conception, Cas9 nickases carried significantly lower rates of off-target effects[103], but had dismal rates of successful on-target editing in human stem cells[103]. In that context, highly specific single-cutting Cas9 nucleases were a welcomed development[132] just as this project started. The 'enhanced' SpCas9 (eSpCas9) showed a drastic reduction in off-target effects while maintaining similar or only marginally lower rates of on-target cutting[130]. For that reason, it was rapidly incorporated into the CIRCES protocol. A downside, which became apparent as the CRISPR field advanced, was that as a recently developed nuclease, it was not available commercially as a stable protein. It is now widely accepted that protein-based delivery of Cas9 nuclease leads to similar or potentially better on-target cutting efficiency with significantly lower rates of off-target effects, likely due to its limited half-life

once inside the cell[117]. Thus, in future applications of CIRCES, the eSpCas9-sgRNA plasmid would be better replaced by a high fidelity Cas9 ribonucleoprotein (RNP) complex with synthetic chemically-modified sgRNAs, which not only have become more affordable but also have been shown to increase on-target editing rates[133, 224]. Synthetic guides would also overcome the hU6 promoter limitation of only efficiently transcribing RNAs that natively start with a guanine or adenine[152], thus significantly increasing the number of possible genomic sites that can be targeted.

Another early feature of CIRCES was the use of Puromycin rather than Blasticidin resistance genes. Whilst positive selection of hESCs with puromycin was shown in preliminary experiments to be very fast, its effectiveness was significantly impacted by cellular density. As cells proliferated rapidly, often forming patches of high density amongst areas of low density, a fixed concentration would be too toxic to low density regions while not adequately selecting high density areas. The practical consequence was that dynamic puromycin concentration adjustments were needed on a daily basis, which was thought to go against the goal of a reliable and reproducible genome editing system. A search for alternative positive selection agents ensued, and Blasticidin was considered to be the best option. It had a very small resistance gene, it showed effective, albeit slow, selection at a range of concentrations and cell densities in preliminary tests and it was easily cloned into the cassette. Unfortunately, as reported in chapter 5, resistance to blasticidin is transferred to otherwise susceptible neighbouring cells. This critical feature, to the best of my knowledge, had not been reported in the three decades since Blasticidin was first described as a viable drug-resistance marker for mammalian cells[225].

7.1.1 On-Target Cassette Insertion

Prior to starting this project, it was unclear how efficient on-target HDR-mediated integration of the 3kb CIRCES cassette would be. While studies using ssODNs to achieve introduction of missense mutations commonly reported HDR rates of 1-2% in iPSCs/hESCs based on amplicon-based sequencing, the largest reported cassette insertion in human stem cells at that point, to my knowledge, had been 2.7kb, and it required 2kb homology arms and two separate DSBs[226].

In the event of on-target integration, how would it be identified? PCRs are extremely sensitive when amplifying from non-complex templates such as plasmids. Preliminary experiments in HEK293T cells had suggested that small amplicons generated from within the cassette itself were unreliable as they could not distinguish between amplification a correctly on-target inserted cassette from off-target integration from residual plasmid contamination. Therefore, the PCR detection strategy had to be based on amplicons containing both target and cassette sequences. To account for possible partial or imperfect HDR, these external-to-internal PCR reactions were designed to allow a significant gap between each primer and the expected cassette-target junction. This made the amplicons longer, and thus less efficient. Before these PCR reactions could be used to screen colonies, it was critical to assess whether they worked at all.

Analyses of populations of blasticidin-selected transfected H9 cells showed that from a transfected well of a 24-well plate, cassette insertion events were reliably identified for all three target exons, thus also validating the external-to-internal primers that had been designed.

In practical terms, this meant that, assuming positive selection was efficient, a population of cassette-integrated cells would be achieved following chemical transfection of a single well of a 24-well culture plate.

7.1.2 Distinguishing Cassette Integration from Overexpression

After transfection of eSpCas9 and the cassette-containing HDR template, the goal was to select a population of cells containing genomically-integrated copies of the CIRCES cassette. Cell survival to blasticidin selection, however, was only considered to indicate that a blasticidin resistance gene (BSD) was being expressed, but could not identify whether it originated from the genome or the episome. The theoretical basis for making this distinction was that episomal copies would, unlike genomically integrated ones, eventually no longer be expressed due to plasmid degradation or loss. Thus, Insertion Stage required continuous prolonged application of blasticidin selection until all episomal plasmid copies had been lost. Deciding when that point had been achieved was extremely difficult, and its uncertainty impacted interpretation of results in all phases of this project.

7.1.2.1 Plasmid Loss

Transfection of bacterial plasmids into eukaryotic cells is now commonplace in molecular biology research. Nonetheless, little data exists regarding their persistence in transfected mammalian cells. It is generally accepted that most transfected plasmids will lead to detectable expression for up to 7-10 days (e.g. GFP fluorescence), but that does not mean that 100% of cells will have lost (or no longer express) the plasmid by the following day. Using CHO cells transfected with a GFP expressing vector, a study calculated the half-life of transfected plasmid DNA to be 19.5 hours[227], which is broadly consistent with the 20-24 hrs doubling time of these cells[228]. But even as average plasmid copy number per cell rapidly approached zero after 6 days (Fig. 7.1A), the percentage of GFP expressing cells fell at a much slower rate, with 5% of cells still expressing GFP 18 days post transfection[227] (Fig. 7.1B).

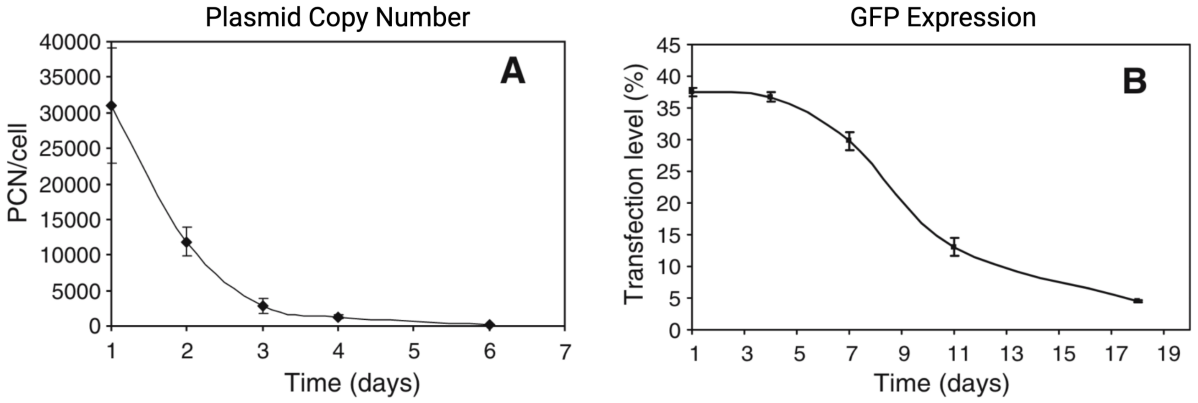

Figure 7.1: Loss of GFP-encoding plasmid in CHO cells following transfection. (A) Average plasmid copy number per cell in the first week after transfection. (B) Proportion of transfected cells showing GFP expression following transfection. Figure from ref.[227]

These and other similar figures, although helpful, cannot directly be used to guide length of blasticidin selection in CIRCES. Data from CHO cells in Figure 7.1 is presented as plasmid copies per cell, but this average hides the fact that the distribution of plasmid copies is likely to be highly uneven. During mitosis, foreign DNA is asymmetrically partitioned between daughter cells[161]. Additionally, H9 hESCs grow and proliferate quite differently from CHO or other immortalised mammalian cell lines. Despite having similar doubling times to CHOs[229, 230], hESCs (and iPSCs) are generally passaged as aggregates and grow as colonies. Cells in the centre of these colonies proliferate slower than ones at edges, and the denser the colony

becomes, the slower cells proliferate[231]. Thus any estimates of plasmid loss via replicative dilution must take into account that transfected cells in dense colonies will retain higher plasmid copy numbers than cells growing in low density contexts[231]. This phenomenon was noticed in preliminary experiments that involved transfecting GFP vectors to H9 hESCs. After 5-7 days, GFP fluorescence could only be identified (often quite intensely) in some colony centres, but never around edges. When followed serially, many of these GFP^+ cell clusters would continue to fluoresce for several days until cells were passaged.

While issues with Blasticidin selection and random integration certainly complicated their interpretation, aggregate data from different sets of preliminary experiments suggested that transient expression of the transfected cassette plasmid would cease in virtually all cells after 12-14 days when the optimised chemical transfection protocol reported in Chapter 4 was followed. After transfection, cells must be cultured at low densities to promote shorter doubling times[231] and thus accelerate plasmid dilution. hESCs and iPSCs, however, are vulnerable to Rho Kinase-dependent apoptosis and spontaneous differentiation when dissociated to single cells and cultured at very low densities[232]. These difficulties are mitigated in the CIRCES protocol by treatment with Rho-kinase inhibitors such as Y-27632[185], but regular use of CloneR, CloneR2 supplement (StemCell) or CEPT[233] are likely to provide even better results.

7.1.2.2 Random Integration of Foreign DNA

Random genomic integration of transfected plasmids occurs relatively commonly in immortalised cell lines such as HEK293T cells, with reported rates of stable integration of up to 20%[234]. Human stem cells, on the other hand, have stable genomes with low mutation rates[235, 236], high DNA repair fidelity[237] and are very sensitive to DNA damage[238]. Thus rates of random integration of foreign DNA should be very low in human stem cells.

Results presented in chapters 4 and 5 challenge this assumption. Early signs that Cas9-independent genomic integration could be higher than anticipated arose from experiments aimed at determining the duration of plasmid overexpression in transfected H9 hESCs. It was observed that cells transfected only with the cassette-containing plasmid (i.e. lacking Cas9 coding sequences)

survived blasticidin selection for several weeks, well beyond the expected length of transient plasmid expression. Repeated single-cell passaging and low density culturing reduced, but never eliminated, the number of cells displaying seemingly stable blasticidin resistance. Although these cells were not investigated further through single-cell colony isolation and genotyping, it indicated that Cas9-independent genomic integration was more frequent than anticipated.

Genotyping results from chapters 4 and 5 suggest that to be the case. Exon 10 screening results were particularly informative. In total, 46 blasticidin-resistant single-cell colonies were isolated from the first CIRCES transfection (chapter 4), but none were found on screening to have on-target cassette insertion. Though some likely had on-target cassette copies that were not detected by the PCR screening methodology used, at least 21 were shown to have a genotype in Intron 10 that was inconsistent with on-target insertion, despite stable blasticidin resistance and presence of a blasticidin-resistance gene. These observations pointed to random genomic integration. These numbers, however, cannot be taken as an objective measure of frequency, given that prolonged selection, repeated passaging, and single-cell cloning could have led to the identification of multiple daughter colonies from a rare isolated integration event. It is safe to say, nevertheless, that it is probable that at least one random cassette integration event occurred in H9 hESCs per transfection. Thus it is estimated to occur in H9 cells with a frequency higher than 4 x 10^{-5}. This makes random integration an important issue to be considered, and expected, in any genomic editing experiment using positive-selection.

7.1.2.3 Ineffective Blasticidin Selection

It is generally assumed that antibiotics used in cell culture allow for good discrimination of selectable marker expression; that is, cells expressing the resistance gene survive antibiotic selection while non-expressing cells die. This is commonly tested for using 'kill curves' to determine the lowest concentration able to kill susceptible cells while sparing those expressing the resistance gene. Kill curves done in H9 hESCs showed that blasticidin concentrations between 3-10ug/mL allowed cassette-transfected cells to survive but completely eliminated untransfected or control-transfected cells (see appendix). Unlike puromycin, these results were repro-

ducible across various cell densities, ranging from single-cell colonies to fully confluent wells. Together with its very small resistance gene[225], these features indicated Blasticidin[177] to be an ideal selection agent for CIRCES.

Issues with blasticidin selection were always considered an unlikely possibility, given its proven ability to kill fully confluent wells of untransfected cells (see Fig. 4.3). For that reason, ineffective blasticidin selection was not appreciated until very late in this project. Recurrent identification of single-cell derived colonies with wild-type alleles in the target loci was initially thought to be due to another competing cause, such as random genomic integration and/or prolonged episomal plasmid persistence. However, given the extent of proximity-based protection shown in chapter 5, it is likely that ultimately most cases can be explained by ineffective blasticidin selection.

Further work is required to establish the mechanisms behind the so far unreported passive transfer of resistance, and determining whether this feature is specific to H9 hESCs and blasticidin or if it is a widespread phenomenon across mammalian cell types and other antibiotic selection agents. For that, clones that reliably express different selection markers will have to be generated, and CIRCES will be the ideal method for it.

7.1.3 Replacement Stage

The original plan was to have CIRCES as a two-stage CRISPR protocol without an intermediary clonal step. Two main issues prevented progression to Replacement Stage: low efficiency of positive selection and high rate of indel genotypes.

7.1.3.1 Maximising Effectiveness of Positive Selection

For CIRCES to work as intended, positive selection during Insertion Stage must be near 100% efficient, i.e. 100% of surviving cells have an actively-expressed genomically-integrated cassette. As discussed in chapter 3, even small rates of positive selection failure would lead to very large reductions in overall protocol efficiency. False-discovery rate would increase nearly 100-fold for every 1% reduction in positive selection efficiency, erasing most of the protocol's

benefits. Randomly integrated cassettes are not inherently problematic considering that negative selection with ganciclovir during Replacement Stage would remove these cells. Thus the issue is how to ensure maximal positive selection efficiency.

It is possible, or even likely, that proximity-mediated antibiotic resistance is not restricted only to Blasticidin in crowded growing clusters such as hESCs. This is an important issue to be addressed by future research which has significant impact on different types of cell culture. Until further data emerges, changes to the protocol need to be considered.

The easiest to implement would be low-density culture. To an extent, it is already part of the current protocol, but to maximise positive selection efficiency it must be applied more aggressively to work around the current limitations of blasticidin selection. The aim would be to increase spatial separation of cells particularly in the period when transient expression of the BSD-TK fusion protein ceases in most cells.

Another alternative would be to use a different combination of antibiotic and resistance gene for positive selection. Puromycin is a tested antibody for which a cassette version already exists, as reported in chapter 3. Its density-dependent concentration, however, pose a challenge to the CIRCES protocol in hESCs. A better option might be Zeocin[189]. Its resistance gene (BleoR) is small at 375bp[239] and a BleoR-TK fusion protein has previously been described in a patent application[240]. A potential downside is that Zeocin has been shown to induce sub-lethal DNA damage even in resistant cells[241].

A third option would be to avoid antibiotic agents altogether and instead use fluorescent markers. This would require a complete overhaul in the protocol as fluorescence-activated cell sorting (FACS) would be needed. FACS would substantially raise the costs and complexity of running the protocol, but could potentially also substitute the ganciclovir-based negative selection step by a second round of FACS.

During the course of this project, a method very similar to CIRCES was published[186]. Also based on two-stage CRISPR HDR insertion-replacement of a positive-negative selection cassette, it used a combination of fluorescent markers and expression of a selectable surface marker (tCD19) to allow for positive selection using a two-step Magnetic-Assisted Cell Sort-

ing (MACS). Whilst expensive and requiring manual microscopy-based colony screening steps after both rounds of CRISPR editing, this approach confirmed the concept behind CIRCES is correct as the authors were able to introduce homozygous and heterozygous mutations in four different loci in both iPSCs and hESCs.

7.1.3.2 Reducing Indel Frequency in Non-Cassette Alleles

The aim of Replacement Stage is to replace the on-target inserted cassette with the second HDR donor template carrying a specific mutation. Only alleles carrying a cassette can be replaced. For that reason, heterozygous mutations (e.g. mutant/wild-type genotypes) can only be generated from cells with cassette/wild-type genotypes at the end of Insertion Stage. Similarly, homozygous mutants can only be generated from cells carrying biallelic cassette insertions.

Although issues with ineffective blasticidin selection certainly affected single-cell cloning results, the main reason that prevented progression to Replacement Stage was a lack of colonies with cassette/wild-type or cassette/cassette genotypes at the end of Insertion Stage. Colonies with on-target cassette insertions invariably had frameshifting indels *in trans*, thus the best possible genotype at the end of Replacement Stage would have been mutant/indel.

This likely happens due to repeated cycles of Cas9 double-strand cleavage and DNA repair at the target locus until the protospacer sequence has been lost[151, 163], either due to HDR (i.e. cassette integration) or NHEJ/MMEJ (i.e. indel prevents re-cutting). Most DSB's are likely repaired correctly to their original sequence[151], which can then be re-cut if Cas9 and guide RNA are present in the cell nucleus.

The issue of indels *in trans* with HDR-edited alleles had already been described by Paquet & Kwart[151]. While trying to generate heterozygous APP or PSEN1 iPS cell lines, 60 single-cell clones carrying heterozygous knock-in mutations were identified on screening. Upon investigation of the non-HDR alleles, indels were found in all but one. Their solution was similar to the one tested in chapter 6, and involved electroporating short HDR donor templates (ssODNs) containing silent blocking mutations that once incorporated would prevent Cas9 re-cutting. This approach, although successful, was not 'scarless' unless a second CRISPR step targetting

the blocking mutation is used to replace it back to the original allele.

The results presented in chapter 6 showed that co-transfection of a ssODN carrying a silent blocking mutation did lead to incorporation of the mutation in a small percentage of cells. The success rate, however, was considered far too low for incorporation into the CIRCES protocol, as the proportion of cells carrying an indel was not meaningfully reduced.

Incomplete blasticidin selection may have underestimated the effect of the ssODN strategy. Approximately 80% of non-HDR alleles in blasticidin-resistant cells were wild-type prior to clonal isolation. If most of these were bystander wild-type cells, it is possible that a significant minority of indels were replaced by the silent mutation.

The true effect of the ssODN approach may also have been underestimated due to the type of chemical transfection used. Though optimisation for different concentrations of ssODN were performed, Mirus TransIT LT1 is not a good reagent for delivering ssODNs to cells (manufacturer makes no mention of this application), and transfection efficiency of the ssODNs themselves is difficult to assess. It is possible that transfection efficiency of plasmids was high, but ssODNs were poorly packaged into transfection complexes. For best results a different transfection method may be needed. The most studied for ssODN delivery is electroporation, but the recently developed Mirus TransIT X2 is reported by the manufacturer to be able to transfect plasmid DNA, ssODNs and/or Cas9 RNP into human stem cells, an opportunity that will require future testing.

Lastly, indels in the non-HDR allele might be reduced by strategies aiming at reducing the duration of Cas9 activity in transfected cells. If the theoretical model of multiple cycles of Cas9 cutting followed by perfect end-joining repair before a NHEJ error occurs (e.g. indel) is correct, it might be possible to engineer truly scarless heterozygous mutations by shortening the potential Cas9 half-life. A simple strategy would be to reduce the amount of Cas9 plasmid transfected, which is known to both reduce on- and off-target effects[150]. Another would be to directly transfect Cas9 protein, which has been shown to have a shorter half-life, higher rates of on-target cutting and lower rates of indels[117]. A third would be to use recently-developed inducible Cas9 versions[242–246] which can be controlled after transfection.

Inducible CRISPR editing is the approach most likely to substantially reduce indels in non-HDR alleles in a way that minimally disrupts the established protocol while maintaining or improving overall HDR efficiency. In particular, the recently developed DD-Cas9[245] (which could easily be modified to DD-eSpCas9) would fit well into the CIRCES protocol. Its activity can be switched on and off by exposure to Shield-1, a nontoxic membrane-permeable small molecule[247]. This Cas9 has been fused N-terminally to a destabilising domain (DD), which is a mutant version of the rapamycin-binding protein FKBP12. Under normal conditions, DD-Cas9 is unstable and rapidly degraded. Stability is reinstated by addition of the Shield-1 ligand[245]. Critically, removal of Shield-1 leads to near complete degradation of the fusion protein in as little as 4 hours. Considering that HDR only occurs during late S and G2 phases of the cell cycle, precise pulse activation of Cas9 activity could also further improve HDR-to-NHEJ ratio if coupled to cell cycle synchronisation approaches.

7.1.4 Proposed Modifications to the Protocol

Results in this thesis provide proof of concept that numerous biallelic knockout clones can be achieved in human stem cells in a short time frame using only standard molecular biology reagents and cell culture equipment.

Nevertheless, there is space for substantial improvement in the cassette design if the overall goal is to achieve gene knockout rather than knock-in of specific mutations, as it was originally intended with two-stage CIRCES.

The current CIRCES cassette stands at 3050bp flanked by 800bp homology arms. Much of its length is due to parts not required for the generation of gene knockouts. If a Replacement Stage is not required, the HSV-TK gene and Drosophila Specific Sequence can be removed without any detriment to Insertion Stage. Similarly, using a shorter version of the EF1α promoter and a shorter polyA signal (e.g. SV40), the total cassette length could be reduced by more than three quarters to about 750bp.

This shortened cassette would bring efficiency gains to the protocol. First, a small cassette is likely to be inserted on-target with significantly higher efficiency. Second, higher integra-

tion efficiency may allow for shorter homology arms (e.g. 200-400bp) to be used, putting the overall HDR donor template at 1.2-1.5kb. Such template could be inexpensively ordered as a purified synthetic double-stranded DNA block and directly transfected as a linear HDR template to cells, thereby abolishing cassette cloning steps. Third, unlike plasmids, linear foreign DNA is not protected from cellular nucleases, and would be expected to be degraded much faster. Positive selection might thus be required for a substantially shorter period to confidently identify cassette-integrated cells. Finally, colony screening would be much easier. Primer pairs spanning the entire on-target inserted cassette and homologous regions could be designed to generate amplicons much shorter than 2kb, allowing a single PCR reaction across the insertion site to replace External-to-Internal and Screening PCR reactions.

Changes to the other components of the protocol can also bring additional gains. High-fidelity Cas9 versions (such as eSpCa9) are now commercially available as ribonucleoproteins. If coupled with commercially-ordered synthetic sgRNAs, the entire protocol, including optimisation and sgRNA testing steps, would be cloning-free. Additionally, synthetic sgRNAs do not suffer from U6-promoter-related limitations, therefore at least doubling the number of possible targets.

If these proposed changes to the protocol are implemented, it should be possible to reliably generate knockout human stem cells for any gene of interest in under 3 weeks with only basic cell culture equipment. Increased upfront costs should be mostly or completely offset by gains in overall efficiency, particularly by eliminating preliminary cloning steps and by reducing cell culture maintenance costs.

7.2 Stem Cell Models of ITPR1 Mutations

The IP_3 calcium pathway is one of the most well studied ways by which extracellular signals trigger intracellular effects. Several genes directly or indirectly involved in this pathway have been associated with neurogenetic syndromes, most prominently motor disorders such as ataxias[248].

ITPR1 encodes the IP_3 Receptor Type 1, a key ligand-gated calcium channel in the IP_3 pathway.

Over the past 15 years mutations in ITPR1 have been identified as the cause of several ataxia syndromes. These include Spinocerebellar Ataxia types 15 (SCA15)[15], 29 (SCA29)[26], Gillespie syndrome (GS)[4, 5] and a pontocerebellar hypoplasia-like disorder[68, 69]. Cerebellar ataxia due to ITPR1 mutations has also been described in dogs[249] and mice[15, 75, 216]. However, despite much research into the structure and physiology of IP_3 Receptors, the mechanisms by which mutations in ITPR1 lead to ataxia are not well understood.

Research progress is slow due to two main reasons. First, post-mortem or other types of pathological studies from patients are not available. Second, the few reported mouse models thus far do not survive past a few weeks of life[15, 75, 216]. In stark contrast to the cerebellar atrophy identified in imaging studies of human patients, mice cerebella are histologically indistinguishable from healthy littermates despite severe motor symptoms[75].

Stem cell models are one of the approaches to advance the understanding of ITPR1-related disorders. Once generated, they can be used to study the effect of different forms of ITPR1 disruption in a cellular context more similar to that found in native human tissue. However, generating human stem cell models is challenging. As discussed previously, iPSCs have several limitations and, at the time this project started, both the expense and difficulty of generating isogenic controls from iPSCs were obstacles. Therefore this thesis focused on developing a method that circumvented the expenses associated with reprogramming and obviated the need for isogenic controls by using hESCs.

The molecular tools developed in chapter 4 made possible the generation of dozens of ITPR1 mutant hESC clones. While the goal of introducing different classes of disease-causing mutations was not achieved, the proposed method was able to confidently knockout ITPR1 in at least seven different hESC lines. The optimised method was efficient and quick, and can be used to generate models for numerous loss-of-function conditions. Furthermore, as no stem cell models of ITPR1 mutations have been published to date, the results from this project and any further analyses thereof should help advance the understanding of ITPR1-related pathology. In particular, the seven knockout clones characterised in chapter 6 should be valuable models of Gillespie syndrome.

Knockout cell lines generated had biallelic loss-of-function mutations in three different exons: 10, 58 and 59. These exons were chosen because they harbour most of the identified pathogenic ITPR1 mutations. Also, they encode for sequences of residues that are critical to IP_3 binding (Exon 10) or calcium release (Exons 58 and 59), the two main functions of IP_3 Receptors.

Achieving knockouts at three separate loci confirms that the methods proposed in this thesis are not locus specific and should also be applicable to many other genetic targets. More importantly, having stem cell models with biallelic loss-of-function mutations at different regions of the transcript aids in consideration of different hypotheses regarding ITPR1-related disease mechanisms.

7.3 Mechanisms of ITPR1-Related Disorders

Haploinsufficiency is the mechanism by which large heterozygous genomic deletions are thought to cause SCA15[15]. This is based on IP3R1 levels in immortalised lymphocytes of SCA15 patients which show reduced expression without any N-truncated products[15, 16, 19].

SCA29's mechanism, on the other hand, was initially speculated to be dysregulation of IP3R1[26]. The first two identified mutations (N602D and V1562M[26]) affected residues part of binding sites for proteins such as IRBIT[250] and CA8[251], which were known inhibitors of IP_3 binding. This understanding was supported by the fact that mutations in CA8, also known as CARP, were known to cause congenital ataxia and intellectual disability[74], and by the description of a SCA29 family with a gain-of-function mutation in the suppressor domain (R36C) which was shown to enhance IP_3-binding affinity[58]. Systematic functional evaluation of published SCA29 mutations suggested a more complicated picture, however. Mutations in or near the the IP_3 binding domain (including N602D) exhibited virtually no IP3-induced calcium release[57], whereas regulatory domain mutations such as V1562M and S1502D released calcium normally[57]. More detailed investigation revealed that pathogenic mutations in the IBD drastically reduced IP_3 binding affinity, while mutations just outside the IBD (T594I and N602D) showed normal binding but likely prevented the necessary IP_3-induced conformational change to open the channel[57]. The regulatory mutations, however, were shown to prevent CA8-

mediated inhibition of calcium release[57].

The two Gillespie-causing *de novo* mutations studied (G2554R and K2611del) to date, when expressed in model systems devoid of wild-type IP3R1, have been shown to form channels that are totally unable to release calcium[4, 59]. Similarly, autosomal recessive Gillespie-causing mutations reported to date affect critical splice sites[7] or introduce stop codons[4] or frameshifts[6].

Therefore, SCA15 is caused by haploinsufficiency and Gillespie syndrome has been ascribed to loss-of-function mechanisms (recessive or dominant-negative) while SCA29 has been linked to gain-of-function, loss-of-function and regulatory mechanisms. It thus seems that disruptions in the quantity, function or regulation of IP3R1 are all capable of causing ataxia.

This general interpretation is, however, hindered by several inconsistencies. The most obvious is that parents of autosomal recessive Gillespie syndrome patients, proven carriers of heterozygous loss-of-function mutations, should be haploinsufficient for IP3R1 — and therefore develop SCA15 —, but do not. In addition, some individuals with large genomic deletions encompassing the entire ITPR1 gene also do not develop SCA15, despite the disease being reported as fully penetrant. Three generations of a family in the Oxford Ataxia cohort have been found to carry a large genomic deletion that encompasses all of ITPR1 and neighbouring genes. While the index case has mild developmental delay and hypotonia, the father and grandmother (in her eight decade of life) from whom the deletion was inherited have no evidence of ataxia (Prof Andrea Nemeth, personal communication).

Interestingly, all reported SCA15-causing deletions have involved at least the first 10 coding exons of ITPR1, without any record of small or large intragenic or 3' deletions causing SCA15. And despite ITPR1's large size, both missense and loss-of-function variants are observed in large scale genomic databases at significantly lower rates than expected. In fact, ITPR1 is one of the most intolerant human genes to heterozygous loss-of-function variants, a finding at odds with the asymptomatic carrier status of parents of AR Gillespie syndrome patients.

Given these inconsistencies, it is difficult for clinicians to predict which types of mutations are expected to lead to ataxia in humans and which are likely benign.

7.3.1 Channel Dosage Hypothesis

It has been postulated that Gillespie-causing truncating mutations may not be as detrimental as heterozygous gene ablation, and Purkinje cells of parents of AR Gillespie patients are likely to produce more ITPR1 than SCA15 patients, thereby avoiding late-onset neurodegeneration[4].

According to this hypothesis, while 50% ITPR1 expression leads to inevitable late-onset ataxia, expression levels slightly superior to that allow for near normal cerebellar function to be maintained throughout the human lifespan. Analysis of RNA samples from two AR Gillespie families support this notion. A GS patient with compound heterozygous splice site mutations expressed three different splicing products: two frameshifted predicted to be truncating and one wild-type transcript. Parents expressed both the wild-type and frameshifted transcripts[4]. Another family showed essentially the same results: Gillespie patients expressed very small levels (3.3%) of full-length ITPR1 transcript, while parents expressed slightly more than 50% (50.2-54.1%)[7].

These two families carried splice site mutations. In families that have stop codons or frameshifts in ITPR1 exons, it is suggested that mechanisms such as nonsense-associated altered splicing, basal exon skipping or translational read-through might have occurred[4]. If so, the interpretation would be that a minimal residual amount of partially functional IP3R1 exists in Gillespie patients.

Existing mouse models of ITPR1 disruption have allowed a small amount of near full length transcript to be produced. The only model considered to be a knockout disrupted ITPR1 by the insertion of a cassette in the first coding exon[75]. Approximately 80% of homozygous mice die *in utero*; of the ones that were born, no IP3R1 was detected in cerebellar samples, but significant amounts of a ITPR1 transcript missing the first exon were detectable in the cerebellum[75]. It is thus possible that small amounts of a slightly shorter N-truncated IP3R1, in the order of only a couple percent, were missed by the Western blots. Another model, the spontaneous *Opisthotonus* mouse, carries a homozygous in-frame deletion that removes two exons in the regulatory region and has a small amount (10%) of IP3R1 detectable on Western Blots of cerebellar samples[216]. Similarly, the $\Delta 18$ model carries a homozygous 18bp exonic

in-frame deletion in the regulatory region and it too had detectable levels of IP3R1 in the cerebellum[15].

Another line of evidence that supports the channel dosage hypothesis is that all other AR Gillespie syndrome mutations reported to date spare the core IP_3-binding and channel domains[4, 6, 7]. In other words, all known AR Gillespie mutations are theoretically 'escapable' by alternative splicing since none have affected function-critical domains.

There have been at least two publications which have described the generation of animal models with loss-of-function mutations in 'unescapable' regions[5, 65]. In both cases, no viable homozygous animals were reported. McEntagart et al. generated at least two heterozygous null *Itpr1* mice which developed normally. These animals had frameshifting mutations in Exon 58 which were only 6 residues away from those reported in this thesis. Surprisingly, however, no homozygous null *Itpr1* mice were reported. Kinoshita et al., more recently, tried to disrupt the equivalent of human exon 59, but all three founder mice carrying deletions in this exon died soon after birth for unknown reasons[65]. Despite having knockout mutations at high mosaic rate, these mice were reported to be histologically normal. The same authors were eventually able to generate a line of ataxic mice, although the genetic defect was in the last exon (62) and introduced a frameshift that disrupted only the last three codons of ITPR1[65].

The channel dosage hypothesis raises the possibility that true ITPR1 knockout (e.g. the equivalent of biallelic gene deletions) may not be compatible with normal embryological development. In the same line, existing CRISPR knockout screen libraries[252, 253] might not have led to complete knockout of ITPR1 and therefore their results perhaps should be interpreted with caution.

The results presented in this thesis however argue against this hypothesis, at least in embryonic stem cells and in early neuronal development. By engineering frameshifting mutations in the most function-critical regions of IP3R1, namely the channel pore (Exons 58 and 59) and the IP_3-binding pocket (Exon 10), it is theoretically impossible for these mutations to be escaped by alternative splicing. Analysis of RNA transcripts from mutant cells lines did not show evidence of alternative transcripts that could have produced a viable protein. It is known that deletion

of transmembrane regions 5 and 6 in the rat IP3R1 does not allow formation of functional channels[64]. Similarly, single aminoacid substitutions in the IP_3-binding pocket (Exon 10) are known to drastically reduce IP_3 binding[53]. It is thus expected that alternative splicing involving these regions, even if it produces a protein product, should lead to nonfunctional channels. In addition, Western Blots using sensitive detection methods did not show evidence of IP3R1 being expressed in the knockout cell lines (chapter 6). Therefore, total loss of IP3R1 was tolerated in both embryonic stem cells and cells resembling developing neurons. The reasons for this are not clear, and may include redundancy between different types of IP_3 receptors (as suggested by the normal calcium release assays) or lack of a critical function for IP_3-mediated calcium signalling in human stem cells and early stages of neurodevelopment. Indeed, a very recent publication not only confirmed the functional redundancy of IP_3 receptors in pluripotent cells but also demonstrated that human iPSCs retained normal proliferation, pluripotency and stemness despite triple knockout of IP_3 receptor genes[254].

7.3.2 5' ITPR1 Haploinsufficiency Hypothesis

Another possibility was that total IP3R1 knockout is indeed compatible with life, but disruption of another yet undiscovered shorter transcript or genetic element within ITPR1 is not. Given that all SCA15 cases have been associated with deletions involving the 5' region of ITPR1, it would be expected that disruption of a genetic element or transcript in this region is behind the neurodegeneration seen in these patients.

One such element is ITPR1-DT, an obscure two-exon divergent noncoding transcript immediately 5' of ITPR1 exon 1. Another is the 5'-UTR of ITPR1, which includes exons 1, 2 and part of 3. Undiscovered shorter protein-coding transcripts involving the first coding exons were also possibilities. For example, the latest version of GENCODE (v.43) contains two protein coding transcripts that could be of importance. The first, ITPR1-207, is a 296aa (33kDa) transcript of the first ten coding exons followed by a retained intron that was identified in a cDNA library from uterine samples. The other, ITPR1-219, is a 721aa (82kDa) cerebellar transcript that skips directly from exon 11 to exon 54. This transcript, if translated, would contain both the IP_3-binding and the channel domains and could theoretically be a functional IP_3-gated calcium

channel.

The 5' haploinsufficiency hypothesis is also contradicted by results described in chapter 6. As knockout mutations were engineered in both 5' and 3' regions of ITPR1, the two shorter transcripts (ITPR1-219 and ITPR1-207) would have been disrupted, but all lines proliferated and differentiated similar to controls. Western blots using both N- and C-terminal anti-IP3R1 antibodies also did not suggest the existence of shorter protein products.

Therefore, the fact that true knockouts of ITPR1 hESC lines reported in this thesis not only proliferated normally in the undifferentiated state but also differentiated successfully to early neurons suggests that functioning IP3R1 is not necessary for early neuroembryological development. Furthermore, the lack of IP3R1 was not compensated by upregulation of other IP3Rs in either undifferentiated stem cells or neuronal cells.

7.4 ITPR1 Expression in Neuronal Differentiation

The neuronal differentiation protocol used in chapter 6 is based on forced expression of the transcription factor Neurogenin-2 (Ngn2). It directly converts stem cells to neurons in only a few days and has been shown to generate cells transcriptionally equivalent to those found in the second trimester of brain development[197, 198]. The low ITPR1 expression in wild-type Ngn2-neurons was unexpected but mirrors the results of recent studies. In a RNAseq study, ITPR1 and ITPR3 were barely detectable in an iPSC line and their expression did not significantly change during the 35 days of neuronal differentiation[255]. ITPR2 was highly expressed in undifferentiated iPSCs but decrease more than 7 fold within the first 14 days of differentiation, maintaining similar low levels up to at least day 35[255]. A proteomic study from the same group but using a different iPSC line found results essentially identical to those reported in this thesis, with all three IP_3 receptors undergoing a significant and sustained drop in expression within 10 days of differentiation[217].

Ngn2 differentiation protocols however suffer from several limitations. They do not reproduce the very early neurodevelopmental stages such as embryoid bodies and neural rosettes which are generated in dual SMAD inhibition protocols[218]. Therefore, it is possible that ITPR1

could be important during those very early stages. Ngn2 differentiation protocols also do not produce glial cells, which are important for neuronal maturation. Moreover, while the yields of neuronal differentiation are very high when using forced expression of Ngn2[256], the neuronal cells generated are quite diverse and have transcriptomic profiles with mixed signatures between CNS and PNS-derived neurons[257].

Cortical differentiation protocols based on dual SMAD inhibition[218], though more complicated, may be better protocols for study of ITPR1 mutations in the absence of specific cerebellar protocols. For instance, data from CORTECON shows that as hESCs were differentiated to cortical neurons, ITPR1 expression halved in the first 10 days but then increased steadily through day 77, reaching levels 5-fold higher than undifferentiated cells[25]. Importantly, expression of ITPR2 and ITPR3 drops massively within the first week, staying at low levels throughout the rest of the protocol. However, considering that peak expression of ITPR1 at day 77 is approximately the same as the minimal levels of ITPR2 expression (and only double the level of ITPR3 expression), it is likely that functional redundancy of IP_3 receptors would be maintained.

The ideal cellular model to study the effects of disease-causing ITPR1 mutations will be one that dominantly expresses the type 1 receptor with little to no redundant expression of the other two types. Such pattern of expression occurs in few cell types in the human body, most notably Purkinje neurons. Therefore cerebellar differentiation protocols that can reliably generate Purkinje neurons are likely to be the best model system to study the impact of ITPR1 mutations in a disease-relevant cellular context.

7.5 Developmental Role of IP_3 Receptors

The role of IP_3 receptors during development is complex. As previously mentioned, three separate genes encode IP_3 receptors. While most ITPR1 knockout mice die *in utero* and the survivors develop ataxia and die prematurely[75], mice in which ITPR2 or ITPR3 have been knocked out individually are born at Mendelian ratios and do not show any phenotype[258].

Double ITPR2/ITPR3 knockouts are embryologically viable but rapidly become malnourished due to deficits in exocrine secretion in pancreatic and salivary gland cells[258]. If these animals

are fed a wet diet, they are able to survive despite having lower body weight than controls[258].

Double ITPR1/ITPR2 knockouts all die *in utero* by E11[222]. Since these nonviable embryos had clear heart defects and ITPR1 and ITPR2 expression was detected in myocardial and endocardial cells of wild-type embryos prior to E11, it was thought that double KOs died as a result of arrested cardiogenesis. A recent study using conditional double KO mice has shown, however, that IP3R1 and IP3R2 in cardiovascular cell lineages are dispensable for embryonic survival[223]. DKOs have allantoic-placental abnormalities that impair fetal-maternal connection and are likely to lead to secondary cardiogenesis defects through poorly understood mechanisms. Interestingly, embryos carrying at least one wild-type allele of either ITPR1 or ITPR2 did not show placental defects[223].

ITPR1 and ITPR2 therefore form a redundant system in early embryological development. That however does not explain why most ITPR1 knockout mice – with both ITPR2 alleles intact – die *in utero*, and detailed embryologic investigation of ITPR1 KO animals has not yet been undertaken. It is clear however that the genomic background greatly influences embryo viability in homozygous mutants. When the ITPR1 KO mutations were transferred from the original to other six mouse strains, embryological survival of homozygous mutants varied between 0% and 50%[259].

Regardless of the exact causes of this difference, it is not known how these findings translate to early human development. What is known is that drastic reductions in ITPR1 expression are compatible with life, albeit with variable rates of cardiovascular, skeletal and ocular malformations as well as the neurological impairments.

Although human histological data on ITPR1-related neurological diseases are yet to be reported, successful Ngn2-induced neuronal differentiation from knockout human embryonic stem cells adds to the existing results from animals models and Gillespie patients to suggest that neurological development in the context of ITPR1 knockout is not significantly abnormal until late in the prenatal period or even after birth.

7.6 Postnatal Cerebellar Dysfunction in ITPR1-Related Disorders

Gillespie Syndrome and SCA29 are generally considered to be neurodevelopmental, as opposed to neurodegenerative, conditions as they typically present in the first months of life and are clinically non-progressive. In the absence of human pathological studies on these conditions, a different interpretation can be derived from animal and cellular models that fits available clinical data.

In mice, cerebellar maturation starts at birth and is considered mostly complete by the third or fourth week of life[260]. During this period, the molecular layer of the cerebellum thickens, the granule cell layer rapidly expands and Purkinje cells (PCs) develop their characteristic arborisation with significant dendritic complexity. Therefore, it is fitting that *Itpr1* knockout mice are anatomically and behaviourally normal at birth[75] and do not develop ataxia until the second week of life. Normal expression of IP3R1 in the murine brain and cerebellum is very low prenatally and increases significantly in the second postnatal week[196, 261], temporally coinciding with the onset of ataxia in KO animals. After this point, ataxia and dystonia progress rapidly and all global knockout animals die by the third week[75]. But despite severely disabling symptoms, the light-microscopy cerebellar structure of these animals is normal throughout their entire (albeit short) lifespan. This is in contrast to traditional cerebellar mutant mice[213, 214]. Multiple studies have shown that cell numbers in *Itpr1* knockout mice are similar to wild-type, indicating that there is no Purkinje cells loss[75, 216, 262]. Electrophysiological analyses also have shown normal membrane excitability of Purkinje cells (PCs)[75] and normal synaptic transmission from parallel fibres or climbing fibres to PCs[75]. Other key pathway genes such as metabotropic glutamate receptors and protein kinase C are similarly expressed at normal levels.

The premature lethality in global knockout animals is not seen when ITPR1 is disrupted only in the brain[262, 263]. Indeed, conditional cerebellum and brainstem KO animals show ataxia and dystonia at PND9 similar to global knockouts, but survive to adulthood if fed by hand.

In these animals, at 8 weeks, the cerebral cortex was equivalent in size to wild-type controls. While their cerebella were significantly smaller than wild-type controls, microscopically they were normal with no apparent cell death[262].

If ITPR1 is knocked out only in Purkinje cells (PCs), animals still develop ataxia (but not dystonia) but are able to survive into adulthood without assisted feeding[263]. Despite severe motor symptoms, PC-specific KO animals show normal overall cerebellar morphology at 10 weeks, with cerebellar size and PC numbers comparable to controls. Therefore PC-specific *Itpr1* knockout bears the closest resemblance to the clinical trajectory of human Gillespie and SCA29 patients. Importantly, these animal models suggest there might be no cell loss in human cases of ITPR1 disorders.

More detailed investigations, however, show subtle phenotypes in knockout animals. In the global KO mice, cerebellar membranes and microsomes showed almost no IP_3 binding or IP_3-induced calcium release[75, 259], respectively, while a complete lack of long-term depression (LTD) was identified in cerebellar slices[264]. This highlights a lack of functional redundancy of IP_3 receptors in affected tissues in mice, in stark contrast to the unchanged calcium release assays results in knockout human Ngn2-neurons (as reported in chapter 6). Morphologically, dye marking of individual Purkinje cells from knockout mice reveals clearly abnormal dendritic morphology, with reduced branching point density and vacant areas lacking any dendritic structures[265]. And in culture, Purkinje cells taken from these animals on the first or second postnatal day develop abnormal dendritic outgrowth after 9 days[265]. Thus, global *Itpr1* knockout mice develop a cytoarchitectural phenotype which coincides with the period associated with cerebellar maturation.

In humans, cerebellar maturation starts prenatally at 24 weeks and progresses through the first year of life[260]. Expression of ITPR1 in the human cerebellum follows this process, increasing in the last few prenatal weeks and peaking at 1 year (Fig. 7.2). The clinical and radiological picture suggests that motor (and perhaps also cognitive) symptoms in SCA29 and Gillespie patients are temporally associated with the period of cerebellar maturation in humans, as they are in knockout mouse models. In SCA29 patients, the first symptoms are noted at birth in only

about half the cases[33]. The remainder present mostly in the first year of life. The clinical course in Gillespie syndrome is rather similar to SCA29, but the aniridia is usually detected at birth as fixed dilated pupils. Clinically, the cerebellar symptoms do not progress, with some families even reporting small improvements over time[4, 5]. Although it is tempting to speculate that lack of ITPR1 expression prevents the cerebellum from reaching its cytoarchitectural maturity, murine data from PC-specific *Itpr1* knockout models offer another possible interpretation.

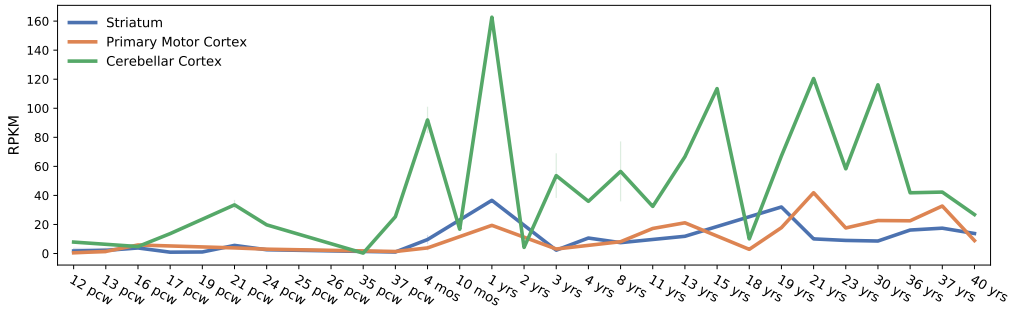

Figure 7.2: Expression of ITPR1 throughout human brain development. Cerebellar cortex ITPR1 expression is shown in green. Note the peak in expression in the first year of life. Expression in the primary motor cortex (orange) and striatum (blue) are shown for comparison. RNASeq dataset from the BrainSpan Atlas of the Developing Human Brain: Developmental Transcriptome[266].

Due to an artefact inherent to the methods used to create such PC-specific models, *Itpr1* knockout is not complete nor homogeneous: knockout does not start until postnatal day 6, reaching most PCs only by the third week. These mice therefore mostly express ITPR1 through the first two weeks of life, a period of intense cerebellar maturation, before it is almost completely lost. In these animals, ataxia is significantly delayed relative to global or cerebellum-specific knockout mice, starting only 6 weeks after birth but progressing rather rapidly thereafter, with no discernible footprints by the 8th week[263]. As mentioned, these severely ataxic animals have normal overall cerebellar morphology at 10 weeks. However, the number of dendritic branching points on PCs from these animals is significantly reduced, and spine densities are markedly increased in distal dendritic regions. Interestingly, as *Itpr1* was still expressed in a small minority of PCs at 10 weeks, these dendritic abnormalities were identified only in IP3R1$^{-/-}$ but not in IP3R1$^{+/+}$ Purkinje cells. This *Itpr1*-specific difference was not found at 3 weeks[263], and indeed global ITPR1 KO mice have normal PC spine density at a similar timepoint[265]. Thus, PCs lacking *Itpr1* appeared normal immediately *after* the critical first three postnatal weeks,

when extensive maturation of PCs, dendritic outgrowth and synapse formation occur. Nevertheless, dendritic abnormalities still developed in ITPR1 KO PCs as animals aged beyond the period associated with cerebellar maturation[263].

It is possible that IP3R1 is mostly necessary to *maintain* a highly functional state after maturation. The absence of IP3R1 leads to cerebellar dysfunction and consequent motor symptoms, but these are not reflective of Purkinje cell loss. Although human cerebellar samples from patients are required to validate the results from mouse models, existing data suggests SCA29 and Gillespie syndrome could be seen as neurodegenerative conditions of very early onset and rapid progression but limited trajectory. A minimal functional plateau is rapidly reached and clinically-irrelevant cerebellar atrophy follows more slowly. At the time of the original description of SCA29 families, patients were reported to have cerebellar hypoplasia on MRI[24]. However, serial assessments in many patients has revealed that the cerebellum, especially the vermis, progressively reduces in size[33]. Therefore, the term cerebellar atrophy has been commonly used despite a lack of clinical deterioration. By the time most patients undergo brain MRI, cerebellar atrophy is already present, but the cerebellum is often reported as normal if imaged in the first two years of life[33].

SCA15 could represent a slower, and perhaps delayed, version of the same condition, also plateauing in its functional degeneration. The mild ataxia seen in patients with longstanding disease is supportive of this view, as even patients with more than 40 years of disease duration could still be ambulatory[21]. SCA15 is generally considered an adult-onset ataxia, but the original pedigrees included an affected patient with disease onset at the age of 10[10, 267], and other pedigrees report patients showing first symptoms in their teenage years[21]. Novak et al. offer the most interesting description of three affected adult SCA15 patients who recalled clumsiness, difficulty with sports, or mild speech slurring since their school years[19]. And some patients in their 30-50s consider themselves asymptomatic despite exhibiting unequivocal ataxia and cerebellar atrophy[22], highlighting the insidious progression of the disorder. Radiologically the cerebella of some SCA15 patients show a pattern similar to that found in SCA29 and Gillespie syndrome: predominant atrophy of the superior vermis with little to no involvement of the cerebellar hemispheres[10, 19, 22]. This specific pattern is unusual

in other genetic conditions, and has been recently suggested as a hallmark of ITPR1-related disorders[9]. It is possible that atrophy starts in the superior vermis and progresses to other cerebellar regions, although serial MRI studies in pre-symptomatic patients will be required to confirm this hypothesis.

Recent updates on an Italian SCA29 family has further blurred the differences between SCA29 and SCA15. Barresi et al. originally described a case of SCA29 which presented with the classical manifestations of neonatal hypotonia, developmental delay and cerebellar atrophy on imaging[32]. The patient was found to have a heterozygous missense mutation (R241K) in the IP_3-binding domain (IBD) of ITPR1 which has been functionally shown to prevent IP_3 binding, similar to other SCA29-causing IBD mutations. Interestingly, the patient's mother was incidentally found on MRI to have cerebellar atrophy, despite being clinically unaffected in her fourth decade. A recent update on this family has shown that the mother is now reported to show slurred speech, and the clinical picture would be more compatible with SCA15[9]. Regardless of the mechanisms underlying the phenotypic variability in this family, it suggests that other very mild cases of ITPR1-related disorders are likely to exist. For example, at least two known pathogenic mutations (SCA29-causing R568G and Gillespie-causing G2554R) can be found in the genomic database GnomAD version 2.1.1 Non-Neuro dataset, which does not include samples from patients diagnosed with neurological conditions.

In aggregate, available data suggests that SCA29 and Gillespie syndrome are genetic disorders in which cerebellar dysfunction occurs mainly in the postnatal period. As different levels of residual IP3R1 function likely modulate severity of cerebellar manifestations, ITPR1-related disorders should be seen as potentially treatable. Possible therapeutic approaches include viral delivery of wild-type ITPR1 or precise correction of ITPR1 mutations using genome editing technologies such as CRISPR/Cas9 or base editors. Further research is needed to better understand the mechanism of pathogenesis, and for that adequate cellular models that can recapitulate the late stages of cerebellar development and maturation will be required. The several ITPR1 knockout human embryonic stem cell lines reported in this thesis, and the methods used to create them, will be valuable tools in the search for cellular models for mechanistic research and drug testing.

References

1. GILLESPIE, F. D. Aniridia, Cerebellar Ataxia, and Oligophrenia in Siblings. *Archives of Ophthalmology* **73,** 338–341. doi:10.1001/archopht.1965.00970030340008 (Mar. 1965).
2. Glaser, T. *et al.* Absence of PAX6 Gene Mutations in Gillespie Syndrome (Partial Aniridia, Cerebellar Ataxia, and Mental Retardation). *Genomics* **19,** 145–148. doi:10.1006/geno.1994.1024 (Jan. 1994).
3. Hall, H. N., Williamson, K. A. & FitzPatrick, D. R. The genetic architecture of aniridia and Gillespie syndrome. *Human Genetics* **138,** 881–898. doi:10.1007/s00439-018-1934-8 (Sept. 2019).
4. Gerber, S. *et al.* Recessive and Dominant De Novo ITPR1 Mutations Cause Gillespie Syndrome. *The American Journal of Human Genetics* **98,** 971–980. doi:10.1016/j.ajhg.2016.03.004 (May 2016).
5. McEntagart, M. *et al.* A Restricted Repertoire of De Novo Mutations in ITPR1 Cause Gillespie Syndrome with Evidence for Dominant-Negative Effect. *The American Journal of Human Genetics* **98,** 981–992. doi:10.1016/j.ajhg.2016.03.018 (May 2016).
6. Carvalho, D. R., Medeiros, J. E. G., Ribeiro, D. S. M., Martins, B. J. A. F. & Sobreira, N. L. M. Additional features of Gillespie syndrome in two Brazilian siblings with a novel ITPR1 homozygous pathogenic variant. *European Journal of Medical Genetics* **61,** 134–138. doi:10.1016/j.ejmg.2017.11.005 (Mar. 2018).
7. Paganini, L. *et al.* A novel splice site variant in ITPR1 gene underlying recessive Gillespie syndrome. *American Journal of Medical Genetics Part A* **176,** 1427–1431. doi:https://doi.org/10.1002/ajmg.a.38704 (2018).
8. Dell'Acqua Cassão, B., de Rezende, D., Silva, L. & Herbella, F. Esophageal dysmotility in gillespie syndrome. *Journal of Neurogastroenterology and Motility* **19,** 538–539. doi:10.5056/jnm.2013.19.4.538 (2013).

9. Romaniello, R. *et al.* Superior Cerebellar Atrophy: An Imaging Clue to Diagnose ITPR1-Related Disorders. *International Journal of Molecular Sciences* **23,** 6723. doi:10.3390/ijms23126723 (June 2022).
10. Storey, E. *et al.* A new autosomal dominant pure cerebellar ataxia. *Neurology* **57,** 1913–1915. doi:10.1212/wnl.57.10.1913 (Nov. 2001).
11. Miyoshi, Y. *et al.* A novel autosomal dominant spinocerebellar ataxia (SCA16) linked to chromosome 8q22.1-24.1. *Neurology* **57,** 96–100. doi:10.1212/wnl.57.1.96 (July 2001).
12. Knight, M. A. *et al.* Spinocerebellar ataxia type 15 (sca15) maps to 3p24.2-3pter:: exclusion of the ITPR1 gene, the human orthologue of an ataxic mouse mutant. *Neurobiology of Disease* **13,** 147–157. doi:10.1016/S0969-9961(03)00029-9 (July 2003).
13. Miura, S. *et al.* The contactin 4 gene locus at 3p26 is a candidate gene of SCA16. *Neurology* **67,** 1236–1241. doi:10.1212/01.wnl.0000238510.84932.82 (Oct. 2006).
14. Gardner, R. J. M. "SCA16" is really SCA15. *Journal of Medical Genetics* **45,** 192–192. doi:10.1136/jmg.2007.056341 (Mar. 2008).
15. Leemput, J. v. d. *et al.* Deletion at ITPR1 Underlies Ataxia in Mice and Spinocerebellar Ataxia 15 in Humans. *PLOS Genetics* **3,** e108. doi:10.1371/journal.pgen.0030108 (June 2007).
16. Hara, K. *et al.* Total deletion and a missense mutation of ITPR1 in Japanese SCA15 families. *Neurology* **71,** 547–551. doi:10.1212/01.wnl.0000311277.71046.a0 (Aug. 2008).
17. Ganesamoorthy, D. *et al.* Development of a Multiplex Ligation-Dependent Probe Amplification Assay for Diagnosis and Estimation of the Frequency of Spinocerebellar Ataxia Type 15. *Clinical Chemistry* **55,** 1415–1418. doi:10.1373/clinchem.2009.124958 (July 2009).
18. Di Gregorio, E. *et al.* Two Italian Families with ITPR1 Gene Deletion Presenting a Broader Phenotype of SCA15. *The Cerebellum* **9,** 115–123. doi:10.1007/s12311-009-0154-0 (Mar. 2010).

19. Novak, M. J. U. *et al.* An ITPR1 gene deletion causes spinocerebellar ataxia 15/16: a genetic, clinical and radiological description. *Movement Disorders: Official Journal of the Movement Disorder Society* **25,** 2176–2182. doi:10.1002/mds.23223 (Oct. 2010).

20. Castrioto, A. *et al.* A novel spinocerebellar ataxia type 15 family with involuntary movements and cognitive decline. *European Journal of Neurology* **18,** 1263–1265. doi:10.1111/j.1468-1331.2011.03366.x (2011).

21. Marelli, C. *et al.* SCA15 due to large ITPR1 deletions in a cohort of 333 white families with dominant ataxia. *Archives of Neurology* **68,** 637–643. doi:10.1001/archneurol.2011.81 (May 2011).

22. Synofzik, M. *et al.* Spinocerebellar ataxia type 15: diagnostic assessment, frequency, and phenotypic features. *Journal of Medical Genetics* **48,** 407–412. doi:10.1136/jmg.2010.087023 (June 2011).

23. Obayashi, M. *et al.* Prevalence of inositol 1, 4, 5-triphosphate receptor type 1 gene deletion, the mutation for spinocerebellar ataxia type 15, in Japan screened by gene dosage. *Journal of Human Genetics* **57,** 202–206. doi:10.1038/jhg.2012.5 (Mar. 2012).

24. Dudding, T. E. *et al.* Autosomal dominant congenital non-progressive ataxia overlaps with the SCA15 locus. *Neurology* **63,** 2288–2292. doi:10.1212/01.wnl.0000147299.80872.d1 (Dec. 2004).

25. Van de Leemput, J. *et al.* Sequencing analysis of the ITPR1 gene in a pure autosomal dominant spinocerebellar ataxia series. *Movement Disorders* **25,** 771–773. doi:10.1002/mds.22970 (2010).

26. Huang, L. *et al.* Missense mutations in ITPR1 cause autosomal dominant congenital nonprogressive spinocerebellar ataxia. *Orphanet Journal of Rare Diseases* **7,** 67. doi:10.1186/1750-1172-7-67 (Sept. 2012).

27. Shadrina, M. I. *et al.* ITPR1 gene p.Val1553Met mutation in Russian family with mild Spinocerebellar ataxia. *Cerebellum & Ataxias* **3,** 2. doi:10.1186/s40673-016-0040-8 (Jan. 2016).

28. Parolin Schnekenberg, R. *et al.* De novo point mutations in patients diagnosed with ataxic cerebral palsy. *Brain: A Journal of Neurology* **138,** 1817–1832. doi:10.1093/brain/awv117 (July 2015).

29. Ohba, C. *et al.* Diagnostic utility of whole exome sequencing in patients showing cerebellar and/or vermis atrophy in childhood. *neurogenetics* **14,** 225–232. doi:10.1007/s10048-013-0375-8 (Nov. 2013).

30. Sasaki, M. *et al.* Sporadic infantile-onset spinocerebellar ataxia caused by missense mutations of the inositol 1,4,5-triphosphate receptor type 1 gene. *Journal of Neurology* **262,** 1278–1284. doi:10.1007/s00415-015-7705-8 (May 2015).

31. Fogel, B. L. *et al.* Exome Sequencing in the Clinical Diagnosis of Sporadic or Familial Cerebellar Ataxia. *JAMA Neurology* **71,** 1237–1246. doi:10.1001/jamaneurol.2014.1944 (Oct. 2014).

32. Barresi, S. *et al.* Mutations in the IRBIT domain of ITPR1 are a frequent cause of autosomal dominant nonprogressive congenital ataxia. *Clinical Genetics* **91,** 86–91. doi:10.1111/cge.12783 (2017).

33. Zambonin, J. L. *et al.* Spinocerebellar ataxia type 29 due to mutations in ITPR1: a case series and review of this emerging congenital ataxia. *Orphanet Journal of Rare Diseases* **12,** 121. doi:10.1186/s13023-017-0672-7 (June 2017).

34. Synofzik, M. *et al.* De novo ITPR1 variants are a recurrent cause of early-onset ataxia, acting via loss of channel function. *European Journal of Human Genetics* **26,** 1623–1634. doi:10.1038/s41431-018-0206-3 (Nov. 2018).

35. Ngo, K. J., Poke, G., Neas, K. & Fogel, B. L. Spinocerebellar Ataxia type 29 in a family of Māori descent. *Cerebellum & Ataxias* **6,** 14. doi:10.1186/s40673-019-0108-3 (Oct. 2019).

36. Sakamoto, M. *et al.* Genetic and clinical landscape of childhood cerebellar hypoplasia and atrophy. *Genetics in Medicine* **0.** doi:10.1016/j.gim.2022.08.007 (Oct. 2022).

37. Streb, H., Irvine, R. F., Berridge, M. J. & Schulz, I. Release of Ca 2+ from a nonmitochondrial intracellular store in pancreatic acinar cells by inositol-1,4,5-trisphosphate. *Nature* **306,** 67–69. doi:10.1038/306067a0 (Nov. 1983).

38. Ross, C. A. *et al.* Inositol 1,4,5-trisphosphate receptor localized to endoplasmic reticulum in cerebellar Purkinje neurons. *Nature* **339,** 468–470. doi:10.1038/339468a0 (June 1989).

39. Supattapone, S., Worley, P. F., Baraban, J. M. & Snyder, S. H. Solubilization, purification, and characterization of an inositol trisphosphate receptor. *Journal of Biological Chemistry* **263,** 1530–1534 (Jan. 1988).

40. Worley, P. F., Baraban, J. M., Colvin, J. S. & Snyder, S. H. Inositol trisphosphate receptor localization in brain: variable stoichiometry with protein kinase C. *Nature* **325,** 159–161. doi:10.1038/325159a0 (Jan. 1987).

41. Berridge, M. J. Inositol trisphosphate and diacylglycerol as second messengers. *Biochemical Journal* **220,** 345–360. doi:10.1042/bj2200345 (June 1984).

42. Hofmann, T. *et al.* Direct activation of human TRPC6 and TRPC3 channels by diacylglycerol. *Nature* **397,** 259–263. doi:10.1038/16711 (Jan. 1999).

43. Berridge, M. J. Neuronal Calcium Signaling. *Neuron* **21,** 13–26. doi:10.1016/S0896-6273(00)80510-3 (July 1998).

44. Joseph, S. K. & Hajnóczky, G. IP3 receptors in cell survival and apoptosis: Ca2+ release and beyond. *Apoptosis: An International Journal on Programmed Cell Death* **12,** 951–968. doi:10.1007/s10495-007-0719-7 (May 2007).

45. Michaelsen, K. & Lohmann, C. Calcium dynamics at developing synapses: mechanisms and functions. *European Journal of Neuroscience* **32,** 218–223. doi:10.1111/j.1460-9568.2010.07341.x (2010).

46. Hagenston, A. M. & Bading, H. Calcium Signaling in Synapse-to-Nucleus Communication. *Cold Spring Harbor Perspectives in Biology* **3,** a004564. doi:10.1101/cshperspect.a004564 (Nov. 2011).

47. Fan, G. *et al.* Gating machinery of InsP 3 R channels revealed by electron cryomicroscopy. *Nature* **527,** 336–341. doi:10.1038/nature15249 (Nov. 2015).

48. Baker, M. R. *et al.* Cryo-EM structure of type 1 IP3R channel in a lipid bilayer. *Communications Biology* **4,** 1–11. doi:10.1038/s42003-021-02156-4 (May 2021).

49. Mignery, G. A., Newton, C. L., Archer, B. T. & Südhof, T. C. Structure and expression of the rat inositol 1,4,5-trisphosphate receptor. *Journal of Biological Chemistry* **265,** 12679–12685. doi:10.1016/S0021-9258(19)38397-8 (July 1990).

50. Miyawaki, A. *et al.* Structure-function relationships of the mouse inositol 1,4,5-trisphosphate receptor. *Proceedings of the National Academy of Sciences* **88,** 4911–4915. doi:10.1073/pnas.88.11.4911 (June 1991).

51. Alzayady, K. J. *et al.* Defining the stoichiometry of inositol 1,4,5-trisphosphate binding required to initiate Ca2+ release. *Science Signaling* **9,** ra35–ra35. doi:10.1126/scisignal.aad6281 (Apr. 2016).

52. Serysheva, I. I. & Ludtke, S. J. in *Current Topics in Membranes* (ed Serysheva, I. I.) 171–189 (Academic Press, Jan. 2010). doi:10.1016/S1063-5823(10)66008-5.

53. Yoshikawa, F. *et al.* Mutational Analysis of the Ligand Binding Site of the Inositol 1,4,5-Trisphosphate Receptor. *Journal of Biological Chemistry* **271,** 18277–18284. doi:10.1074/jbc.271.30.18277 (July 1996).

54. Uchida, K., Miyauchi, H., Furuichi, T., Michikawa, T. & Mikoshiba, K. Critical Regions for Activation Gating of the Inositol 1,4,5-Trisphosphate Receptor *. *Journal of Biological Chemistry* **278,** 16551–16560. doi:10.1074/jbc.M300646200 (May 2003).

55. Seo, M.-D. *et al.* Structural and functional conservation of key domains in InsP3 and ryanodine receptors. *Nature* **483,** 108–112. doi:10.1038/nature10751 (Mar. 2012).

56. Fan, G. *et al.* Conformational motions and ligand-binding underlying gating and regulation in IP3R channel. *Nature Communications* **13,** 6942. doi:10.1038/s41467-022-34574-1 (Nov. 2022).

57. Ando, H., Hirose, M. & Mikoshiba, K. Aberrant IP3 receptor activities revealed by comprehensive analysis of pathological mutations causing spinocerebellar ataxia 29. *Proceedings of the National Academy of Sciences* **115,** 12259–12264. doi:10.1073/pnas.1811129115 (Nov. 2018).

58. Casey, J. P. *et al.* A novel gain-of-function mutation in the ITPR1 suppressor domain causes spinocerebellar ataxia with altered Ca2+ signal patterns. *Journal of Neurology* **264,** 1444–1453. doi:10.1007/s00415-017-8545-5 (July 2017).

59. Terry, L. E., Alzayady, K. J., Wahl, A. M., Malik, S. & Yule, D. I. Disease-associated mutations in inositol 1,4,5-trisphosphate receptor subunits impair channel function. *Journal of Biological Chemistry* **295,** 18160–18178. doi:10.1074/jbc.RA120.015683 (Dec. 2020).

60. Galvan, D. L., Borrego-Diaz, E., Perez, P. J. & Mignery, G. A. Subunit Oligomerization, and Topology of the Inositol 1,4,5-Trisphosphate Receptor. *Journal of Biological Chemistry* **274,** 29483–29492. doi:10.1074/jbc.274.41.29483 (Oct. 1999).

61. Schug, Z. T. *et al.* Molecular Characterization of the Inositol 1,4,5-Trisphosphate Receptor Pore-forming Segment. *Journal of Biological Chemistry* **283,** 2939–2948. doi:10.1074/jbc.M706645200 (Feb. 2008).

62. Baker, O. *et al.* The contribution of homology arms to nuclease-assisted genome engineering. *Nucleic Acids Research* **45,** 8105–8115. doi:10.1093/nar/gkx497 (July 2017).

63. Nakayama, T. *et al.* The regulatory domain of the inositol 1,4,5-trisphosphate receptor is necessary to keep the channel domain closed: possible physiological significance of specific cleavage by caspase 3. *The Biochemical journal* **377,** 299–307. doi:10.1042/bj20030599 (Jan. 2004).

64. Ramos-Franco, J., Galvan, D., Mignery, G. A. & Fill, M. Location of the Permeation Pathway in the Recombinant Type 1 Inositol 1,4,5-Trisphosphate Receptor. *Journal of General Physiology* **114,** 243–250. doi:10.1085/jgp.114.2.243 (Aug. 1999).

65. Kinoshita, A. *et al.* Itpr1 regulates the formation of anterior eye segment tissues derived from neural crest cells. *Development.* doi:10.1242/dev.188755 (Aug. 2021).

66. Raslan, I. R. *et al.* Quadrupedal gait and cerebellar hypoplasia, the Uner Tan syndrome, caused by ITPR1 gene mutation. *Parkinsonism & Related Disorders* **92,** 33–35. doi:10.1016/j.parkreldis.2021.10.008 (Nov. 2021).

67. Dentici, M. L. *et al.* Identification of novel and hotspot mutations in the channel domain of ITPR1 in two patients with Gillespie syndrome. *Gene* **628,** 141–145. doi:10.1016/j.gene.2017.07.017 (Sept. 2017).

68. Dijk, T. v. *et al.* A de novo missense mutation in the inositol 1,4,5-triphosphate receptor type 1 gene causing severe pontine and cerebellar hypoplasia: Expanding the phenotype of ITPR1-related spinocerebellar ataxia's. *American Journal of Medical Genetics Part A* **173,** 207–212. doi:10.1002/ajmg.a.37962 (2017).

69. Cappuccio, G. *et al.* Long-term follow-up of an individual with ITPR1-related disorder. *American Journal of Medical Genetics Part A* **182,** 1846–1847. doi:https://doi.org/10.1002/ajmg.a.61609 (2020).

70. Hayashi, S. *et al.* Comprehensive investigation of CASK mutations and other genetic etiologies in 41 patients with intellectual disability and microcephaly with pontine and cerebellar hypoplasia (MICPCH). *PLOS ONE* **12,** e0181791. doi:10.1371/journal.pone.0181791 (Aug. 2017).

71. Gonzaga-Jauregui, C. *et al.* Exome Sequence Analysis Suggests that Genetic Burden Contributes to Phenotypic Variability and Complex Neuropathy. *Cell Reports* **12,** 1169–1183. doi:10.1016/j.celrep.2015.07.023 (Aug. 2015).

72. Miyakawa, T. *et al.* Ca2+-sensor region of IP3 receptor controls intracellular Ca2+ signaling. *The EMBO Journal* **20,** 1674–1680. doi:10.1093/emboj/20.7.1674 (Apr. 2001).

73. Keehan, L. *et al.* A novel de novo intronic variant in ITPR1 causes Gillespie syndrome. *American Journal of Medical Genetics Part A* **185,** 2315–2324. doi:10.1002/ajmg.a.62232 (2021).

74. Türkmen, S. *et al.* CA8 Mutations Cause a Novel Syndrome Characterized by Ataxia and Mild Mental Retardation with Predisposition to Quadrupedal Gait. *PLOS Genetics* **5,** e1000487. doi:10.1371/journal.pgen.1000487 (May 2009).

75. Matsumoto, M. *et al.* Ataxia and epileptic seizures in mice lacking type 1 inositol 1,4,5-trisphosphate receptor. *Nature* **379,** 168–171. doi:10.1038/379168a0 (Jan. 1996).

76. Ogura, H., Matsumoto, M. & Mikoshiba, K. Motor discoordination in mutant mice heterozygous for the type 1 inositol 1,4,5-trisphosphate receptor. *Behavioural Brain Research* **122,** 215–219. doi:10.1016/S0166-4328(01)00187-5 (Oct. 2001).

77. Kim, K. *et al.* Epigenetic memory in induced pluripotent stem cells. *Nature* **467,** 285–290. doi:10.1038/nature09342 (Sept. 2010).

78. Doetschman, T. *et al.* Targetted correction of a mutant HPRT gene in mouse embryonic stem cells. *Nature* **330,** 576–578. doi:10.1038/330576a0 (Dec. 1987).

79. Thomson, J. A. *et al.* Embryonic Stem Cell Lines Derived from Human Blastocysts. *Science* **282,** 1145–1147. doi:10.1126/science.282.5391.1145 (Nov. 1998).

80. Eiges, R. *et al.* Establishment of human embryonic stem cell-transfected clones carrying a marker for undifferentiated cells. *Current Biology* **11,** 514–518. doi:10.1016/S0960-9822(01)00144-0 (Apr. 2001).

81. Zwaka, T. P. & Thomson, J. A. Homologous recombination in human embryonic stem cells. *Nature Biotechnology* **21,** 319–321. doi:10.1038/nbt788 (Mar. 2003).

82. Leavitt, A. D. & Hamlett, I. Homologous Recombination in Human Embryonic Stem Cells: A Tool for Advancing Cell Therapy and Understanding and Treating Human Disease. *Clinical and Translational Science* **4,** 298–305. doi:10.1111/j.1752-8062.2011.00281.x (Aug. 2011).

83. Takahashi, K. & Yamanaka, S. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell* **126,** 663–676. doi:10.1016/j.cell.2006.07.024 (Aug. 2006).

84. Soldner, F. *et al.* Generation of Isogenic Pluripotent Stem Cells Differing Exclusively at Two Early Onset Parkinson Point Mutations. *Cell* **146,** 318–331. doi:10.1016/j.cell.2011.06.019 (July 2011).

85. Kuehn, M. R., Bradley, A., Robertson, E. J. & Evans, M. J. A potential animal model for Lesch–Nyhan syndrome through introduction of HPRT mutations into mice. *Nature* **326,** 295–298. doi:10.1038/326295a0 (Mar. 1987).

86. Smithies, O. Forty years with homologous recombination. *Nature Medicine* **7,** 1083–1086. doi:10.1038/nm1001-1083 (Oct. 2001).

87. Hall, B., Limaye, A. & Kulkarni, A. B. Overview: Generation of Gene Knockout Mice. *Current protocols in cell biology / editorial board, Juan S. Bonifacino ... [et al.]* **CHAPTER,** Unit–19.1217. doi:10.1002/0471143030.cb1912s44 (Sept. 2009).

88. Xia, X. & Zhang, S.-C. Genetic Modification of Human Embryonic Stem Cells. *Biotechnology & genetic engineering reviews* **24,** 297–309 (2007).

89. Hockemeyer, D. *et al.* Efficient targeting of expressed and silent genes in human ESCs and iPSCs using zinc-finger nucleases. *Nature Biotechnology* **27,** 851–857. doi:10.1038/nbt.1562 (Sept. 2009).

90. Rouet, P., Smih, F. & Jasin, M. Expression of a site-specific endonuclease stimulates homologous recombination in mammalian cells. *Proceedings of the National Academy of Sciences* **91,** 6064–6068. doi:10.1073/pnas.91.13.6064 (June 1994).

91. Choulika, A., Perrin, A., Dujon, B. & Nicolas, J. F. Induction of homologous recombination in mammalian chromosomes by using the I-SceI system of Saccharomyces cerevisiae. *Molecular and Cellular Biology* **15,** 1968–1973 (Apr. 1995).

92. Smih, F., Rouet, P., Romanienko, P. J. & Jasin, M. Double-strand breaks at the target locus stimulate gene targeting in embryonic stem cells. *Nucleic Acids Research* **23,** 5012–5019. doi:10.1093/nar/23.24.5012 (Dec. 1995).

93. Porteus, M. H. & Carroll, D. Gene targeting using zinc finger nucleases. *Nature Biotechnology* **23,** 967–973. doi:10.1038/nbt1125 (Aug. 2005).

94. Miller, J. C. *et al.* A TALE nuclease architecture for efficient genome editing. *Nature Biotechnology* **29,** 143–148. doi:10.1038/nbt.1755 (Feb. 2011).

95. Zou, J. *et al.* Gene targeting of a disease-related gene in human induced pluripotent stem and embryonic stem cells. *Cell stem cell* **5,** 97–110. doi:10.1016/j.stem.2009.05.023 (July 2009).

96. Hockemeyer, D. *et al.* Genetic engineering of human pluripotent cells using TALE nucleases. *Nature Biotechnology* **29,** 731–734. doi:10.1038/nbt.1927 (Aug. 2011).

97. Zheng, N., Li, L. & Wang, X. Molecular mechanisms, off-target activities, and clinical potentials of genome editing systems. *Clinical and Translational Medicine* **10,** 412–426. doi:10.1002/ctm2.34 (2020).

98. Corrigan-Curay, J. *et al.* Genome Editing Technologies: Defining a Path to Clinic: Genomic Editing: Establishing Preclinical Toxicology Standards, Bethesda, Maryland 10 June 2014. *Molecular Therapy* **23,** 796–806. doi:10.1038/mt.2015.54 (May 2015).

99. Jinek, M. *et al.* A Programmable Dual-RNA–Guided DNA Endonuclease in Adaptive Bacterial Immunity. *Science* **337,** 816–821. doi:10.1126/science.1225829 (Aug. 2012).

100. Gasiunas, G., Barrangou, R., Horvath, P. & Siksnys, V. Cas9–crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. *Proceedings of the National Academy of Sciences* **109,** E2579–E2586. doi:10.1073/pnas.1208507109 (Sept. 2012).

101. Shen, B. *et al.* Generation of gene-modified mice via Cas9/RNA-mediated gene targeting. *Cell Research* **23,** 720–723. doi:10.1038/cr.2013.46 (May 2013).

102. Hwang, W. Y. *et al.* Efficient genome editing in zebrafish using a CRISPR-Cas system. *Nature Biotechnology* **31,** 227–229. doi:10.1038/nbt.2501 (Mar. 2013).

103. Ran, F. A. *et al.* Double Nicking by RNA-Guided CRISPR Cas9 for Enhanced Genome Editing Specificity. *Cell* **154,** 1380–1389. doi:10.1016/j.cell.2013.08.021 (Sept. 2013).

104. Mali, P. *et al.* RNA-Guided Human Genome Engineering via Cas9. *Science* **339,** 823–826. doi:10.1126/science.1232033 (Feb. 2013).

105. Lin, S., Staahl, B. T., Alla, R. K. & Doudna, J. A. Enhanced homology-directed human genome engineering by controlled timing of CRISPR/Cas9 delivery. *eLife* **3** (ed Weigel, D.) e04766. doi:10.7554/eLife.04766 (Dec. 2014).

106. Hsu, P. D. *et al.* DNA targeting specificity of RNA-guided Cas9 nucleases. *Nature Biotechnology* **31,** 827–832. doi:10.1038/nbt.2647 (Sept. 2013).

107. Ishino, Y., Krupovic, M. & Forterre, P. History of CRISPR-Cas from Encounter with a Mysterious Repeated Sequence to Genome Editing Technology. *Journal of Bacteriology* **200,** e00580–17. doi:10.1128/JB.00580-17 (Mar. 2018).

108. Brouns, S. J. J. *et al.* Small CRISPR RNAs Guide Antiviral Defense in Prokaryotes. *Science* **321,** 960–964. doi:10.1126/science.1159689 (Aug. 2008).

109. Jansen, R., Embden, J. D. A. v., Gaastra, W. & Schouls, L. M. Identification of genes that are associated with DNA repeats in prokaryotes. *Molecular Microbiology* **43,** 1565–1575. doi:10.1046/j.1365-2958.2002.02839.x (2002).

110. Sapranauskas, R. *et al.* The Streptococcus thermophilus CRISPR/Cas system provides immunity in Escherichia coli. *Nucleic Acids Research* **39,** 9275–9282. doi:10.1093/nar/gkr606 (Nov. 2011).

111. Collias, D. & Beisel, C. L. CRISPR technologies and the search for the PAM-free nuclease. *Nature Communications* **12,** 555. doi:10.1038/s41467-020-20633-y (Jan. 2021).

112. Deltcheva, E. *et al.* CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. *Nature* **471,** 602–607. doi:10.1038/nature09886 (Mar. 2011).

113. Xue, C. & Greene, E. C. DNA Repair Pathway Choices in CRISPR-Cas9-Mediated Genome Editing. *Trends in Genetics* **37,** 639–656. doi:10.1016/j.tig.2021.02.008 (July 2021).

114. Van Overbeek, M. *et al.* DNA Repair Profiling Reveals Nonrandom Outcomes at Cas9-Mediated Breaks. *Molecular Cell* **63,** 633–646. doi:10.1016/j.molcel.2016.06.037 (Aug. 2016).

115. Shen, M. W. *et al.* Predictable and precise template-free CRISPR editing of pathogenic variants. *Nature* **563,** 646–651. doi:10.1038/s41586-018-0686-x (Nov. 2018).

116. Stadelmann, C. *et al.* mRNA-mediated delivery of gene editing tools to human primary muscle stem cells. *Molecular Therapy - Nucleic Acids* **28,** 47–57. doi:10.1016/j.omtn.2022.02.016 (June 2022).

117. Kim, S., Kim, D., Cho, S. W., Kim, J. & Kim, J.-S. Highly efficient RNA-guided genome editing in human cells via delivery of purified Cas9 ribonucleoproteins. *Genome Research* **24,** 1012–1019. doi:10.1101/gr.171322.113 (June 2014).

118. Cong, L. *et al.* Multiplex Genome Engineering Using CRISPR/Cas Systems. *Science* **339,** 819–823. doi:10.1126/science.1231143 (Feb. 2013).

119. Brinkman, E. K., Chen, T., Amendola, M. & van Steensel, B. Easy quantitative assessment of genome editing by sequence trace decomposition. *Nucleic Acids Research* **42,** e168. doi:10.1093/nar/gku936 (Dec. 2014).

120. Brinkman, E. K. & van Steensel, B. Rapid Quantitative Evaluation of CRISPR Genome Editing by TIDE and TIDER. *Methods in Molecular Biology (Clifton, N.J.)* **1961,** 29–44. doi:10.1007/978-1-4939-9170-9_3 (2019).

121. Conant, D. *et al.* Inference of CRISPR Edits from Sanger Trace Data. *The CRISPR Journal* **5,** 123–130. doi:10.1089/crispr.2021.0113 (Feb. 2022).

122. Owens, D. D. G. *et al.* Microhomologies are prevalent at Cas9-induced larger deletions. *Nucleic Acids Research* **47,** 7402–7417. doi:10.1093/nar/gkz459 (Aug. 2019).

123. Kosicki, M., Tomberg, K. & Bradley, A. Repair of double-strand breaks induced by CRISPR–Cas9 leads to large deletions and complex rearrangements. *Nature Biotechnology* **36,** 765–771. doi:10.1038/nbt.4192 (Sept. 2018).

124. Höijer, I. *et al.* CRISPR-Cas9 induces large structural variants at on-target and off-target sites in vivo that segregate across generations. *Nature Communications* **13,** 627. doi:10.1038/s41467-022-28244-5 (Feb. 2022).

125. Liu, M. *et al.* Global detection of DNA repair outcomes induced by CRISPR–Cas9. *Nucleic Acids Research* **49,** 8732–8742. doi:10.1093/nar/gkab686 (Sept. 2021).

126. Weisheit, I. *et al.* Detection of Deleterious On-Target Effects after HDR-Mediated CRISPR Editing. *Cell Reports* **31,** 107689. doi:10.1016/j.celrep.2020.107689 (May 2020).

127. Skryabin, B. V. *et al.* Pervasive head-to-tail insertions of DNA templates mask desired CRISPR-Cas9–mediated genome editing events. *Science Advances* **6,** eaax2941. doi:10.1126/sciadv.aax2941 (Feb. 2020).

128. Jensen, K. T. *et al.* Chromatin accessibility and guide sequence secondary structure affect CRISPR-Cas9 gene editing efficiency. *FEBS Letters* **591,** 1892–1901. doi:10.1002/1873-3468.12707 (2017).

129. Schep, R. *et al.* Impact of chromatin context on Cas9-induced DNA double-strand break repair pathway balance. *Molecular Cell* **81,** 2216–2230.e10. doi:10.1016/j.molcel.2021.03.032 (May 2021).

130. Slaymaker, I. M. *et al.* Rationally engineered Cas9 nucleases with improved specificity. *Science* **351,** 84–88. doi:10.1126/science.aad5227 (Jan. 2016).

131. Kleinstiver, B. P. *et al.* High-fidelity CRISPR–Cas9 nucleases with no detectable genome-wide off-target effects. *Nature* **529,** 490–495. doi:10.1038/nature16526 (Jan. 2016).

132. Chen, J. S. *et al.* Enhanced proofreading governs CRISPR–Cas9 targeting accuracy. *Nature* **550,** 407–410. doi:10.1038/nature24268 (Oct. 2017).

133. Hendel, A. *et al.* Chemically modified guide RNAs enhance CRISPR-Cas genome editing in human primary cells. *Nature Biotechnology* **33,** 985–989. doi:10.1038/nbt.3290 (Sept. 2015).

134. Fu, Y., Sander, J. D., Reyon, D., Cascio, V. M. & Joung, J. K. Improving CRISPR-Cas nuclease specificity using truncated guide RNAs. *Nature Biotechnology* **32,** 279–284. doi:10.1038/nbt.2808 (Mar. 2014).

135. Okamoto, S., Amaishi, Y., Maki, I., Enoki, T. & Mineno, J. Highly efficient genome editing for single-base substitutions using optimized ssODNs with Cas9-RNPs. *Scientific Reports* **9,** 4811. doi:10.1038/s41598-019-41121-4 (Mar. 2019).

136. Konstantakos, V., Nentidis, A., Krithara, A. & Paliouras, G. CRISPR–Cas9 gRNA efficiency prediction: an overview of predictive tools and the role of deep learning. *Nucleic Acids Research* **50,** 3616–3637. doi:10.1093/nar/gkac192 (Apr. 2022).

137. Doench, J. G. *et al.* Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9. *Nature Biotechnology* **34,** 184–191. doi:10.1038/nbt.3437 (Feb. 2016).

138. Burgio, G. & Teboul, L. Anticipating and Identifying Collateral Damage in Genome Editing. *Trends in Genetics* **36,** 905–914. doi:10.1016/j.tig.2020.09.011 (Dec. 2020).

139. Lessard, S. *et al.* Human genetic variation alters CRISPR-Cas9 on- and off-targeting specificity at therapeutically implicated loci. *Proceedings of the National Academy of Sciences* **114,** E11257–E11266. doi:10.1073/pnas.1714640114 (Dec. 2017).

140. Paix, A., Folkmann, A. & Seydoux, G. Precision genome editing using CRISPR-Cas9 and linear repair templates in C. elegans. *Methods. CRISPR-Cas systems for Genome engineering and investigation* **121-122,** 86–93. doi:10.1016/j.ymeth.2017.03.023 (May 2017).

141. Shy, B. R., MacDougall, M. S., Clarke, R. & Merrill, B. J. Co-incident insertion enables high efficiency genome engineering in mouse embryonic stem cells. *Nucleic Acids Research* **44,** 7997–8010. doi:10.1093/nar/gkw685 (Sept. 2016).

142. Byrne, S. M., Ortiz, L., Mali, P., Aach, J. & Church, G. M. Multi-kilobase homozygous targeted gene replacement in human induced pluripotent stem cells. *Nucleic Acids Research* **43,** e21. doi:10.1093/nar/gku1246 (Feb. 2015).

143. Zhang, J.-P. *et al.* Efficient precise knockin with a double cut HDR donor after CRISPR/Cas9-mediated double-stranded DNA cleavage. *Genome Biology* **18,** 35. doi:10.1186/s13059-017-1164-8 (Feb. 2017).

144. Liang, X., Potter, J., Kumar, S., Ravinder, N. & Chesnut, J. D. Enhanced CRISPR/Cas9-mediated precise genome editing by improved design and delivery of gRNA, Cas9 nuclease, and donor DNA. *Journal of Biotechnology* **241,** 136–146. doi:10.1016/j.jbiotec.2016.11.011 (Jan. 2017).

145. Chen, Y.-H. & Pruett-Miller, S. M. Improving single-cell cloning workflow for gene editing in human pluripotent stem cells. *Stem Cell Research* **31,** 186–192. doi:10.1016/j.scr.2018.08.003 (Aug. 2018).

146. Anzalone, A. V., Koblan, L. W. & Liu, D. R. Genome editing with CRISPR–Cas nucleases, base editors, transposases and prime editors. *Nature Biotechnology* **38,** 824–844. doi:10.1038/s41587-020-0561-9 (July 2020).

147. Salomon, B. *et al.* A truncated herpes simplex virus thymidine kinase phosphorylates thymidine and nucleoside analogs and does not cause sterility in transgenic mice. *Molecular and Cellular Biology* **15,** 5322–5328 (Oct. 1995).

148. Tomicic, M. T., Thust, R. & Kaina, B. Ganciclovir-induced apoptosis in HSV-1 thymidine kinase expressing cells: critical role of DNA breaks, Bcl-2 decline and caspase-9 activation. *Oncogene* **21,** 2141–2153. doi:10.1038/sj.onc.1205280 (Mar. 2002).

149. Norrman, K. *et al.* Quantitative Comparison of Constitutive Promoters in Human ES cells. *PLoS ONE* **5,** e12413. doi:10.1371/journal.pone.0012413 (Aug. 2010).

150. Ran, F. A. *et al.* Genome engineering using the CRISPR-Cas9 system. *Nature Protocols* **8,** 2281–2308. doi:10.1038/nprot.2013.143 (Nov. 2013).

151. Paquet, D. *et al.* Efficient introduction of specific homozygous and heterozygous mutations using CRISPR/Cas9. *Nature* **533,** 125–129. doi:10.1038/nature17664 (May 2016).

152. Gao, Z., Harwig, A., Berkhout, B. & Herrera-Carrillo, E. Mutation of nucleotides around the +1 position of type 3 polymerase III promoters: The effect on transcriptional activity and start site usage. *Transcription* **8,** 275–287. doi:10.1080/21541264.2017.1322170 (June 2017).

153. Kim, S., Bae, T., Hwang, J. & Kim, J.-S. Rescue of high-specificity Cas9 variants using sgRNAs with matched 5' nucleotides. *Genome Biology* **18,** 218. doi:10.1186/s13059-017-1355-3 (Nov. 2017).

154. Vasquez, K. M., Marburger, K., Intody, Z. & Wilson, J. H. Manipulating the mammalian genome by homologous recombination. *Proceedings of the National Academy of Sciences* **98,** 8403–8410. doi:10.1073/pnas.111009698 (July 2001).

155. Mortensen, R. M. & Kingston, R. E. Selection of Transfected Mammalian Cells. *Current Protocols in Molecular Biology* **86.** doi:10.1002/0471142727.mb0905s86 (Apr. 2009).

156. Lechardeur, D. *et al.* Metabolic instability of plasmid DNA in the cytosol: a potential barrier to gene transfer. *Gene Therapy* **6,** 482–497. doi:10.1038/sj.gt.3300867 (Apr. 1999).

157. Boroviak, K., Fu, B., Yang, F., Doe, B. & Bradley, A. Revealing hidden complexities of genomic rearrangements generated with Cas9. *Scientific Reports* **7,** 12867. doi:10.1038/s41598-017-12740-6 (Oct. 2017).

158. Shin, H. Y. *et al.* CRISPR/Cas9 targeting events cause complex deletions and insertions at 17 sites in the mouse genome. *Nature Communications* **8,** 15464. doi:10.1038/ncomms15464 (May 2017).

159. Adikusuma, F. *et al.* Large deletions induced by Cas9 cleavage. *Nature* **560,** E8–E9. doi:10.1038/s41586-018-0380-z (Aug. 2018).

160. Moore, J. C. *et al.* Efficient, high-throughput transfection of human embryonic stem cells. *Stem Cell Research & Therapy* **1,** 23. doi:10.1186/scrt23 (July 2010).

161. Wang, X., Le, N., Denoth-Lippuner, A., Barral, Y. & Kroschewski, R. Asymmetric partitioning of transfected DNA during mammalian cell division. *Proceedings of the National Academy of Sciences* **113,** 7177–7182. doi:10.1073/pnas.1606091113 (June 2016).

162. Scully, R., Panday, A., Elango, R. & Willis, N. A. DNA double-strand break repair-pathway choice in somatic mammalian cells. *Nature Reviews Molecular Cell Biology* **20,** 698–714. doi:10.1038/s41580-019-0152-0 (Nov. 2019).

163. Renaud, J.-B. *et al.* Improved Genome Editing Efficiency and Flexibility Using Modified Oligonucleotides with TALEN and CRISPR-Cas9 Nucleases. *Cell Reports* **14,** 2263–2272. doi:10.1016/j.celrep.2016.02.018 (Mar. 2016).

164. Jayavaradhan, R. *et al.* CRISPR-Cas9 fusion to dominant-negative 53BP1 enhances HDR and inhibits NHEJ specifically at Cas9 target sites. *Nature Communications* **10,** 2866. doi:10.1038/s41467-019-10735-7 (June 2019).

165. Liu, M. *et al.* Methodologies for Improving HDR Efficiency. *Frontiers in Genetics* **9** (2019).

166. Vartak, S. V. & Raghavan, S. C. Inhibition of nonhomologous end joining to increase the specificity of CRISPR/Cas9 genome editing. *The FEBS Journal* **282,** 4289–4294. doi:10.1111/febs.13416 (2015).

167. Miyaoka, Y. *et al.* Systematic quantification of HDR and NHEJ reveals effects of locus, nuclease, and cell type on genome-editing. *Scientific Reports* **6,** 23549. doi:10.1038/srep23549 (Mar. 2016).

168. Wienert, B. *et al.* Timed inhibition of CDC7 increases CRISPR-Cas9 mediated templated repair. *Nature Communications* **11,** 2109. doi:10.1038/s41467-020-15845-1 (Apr. 2020).

169. Yu, C. *et al.* Small Molecules Enhance CRISPR Genome Editing in Pluripotent Stem Cells. *Cell Stem Cell* **16,** 142–147. doi:10.1016/j.stem.2015.01.003 (Feb. 2015).

170. Ray, U., Vartak, S. V. & Raghavan, S. C. NHEJ inhibitor SCR7 and its different forms: Promising CRISPR tools for genome engineering. *Gene* **763,** 144997. doi:10.1016/j.gene.2020.144997 (Dec. 2020).

171. Guo, Q. *et al.* 'Cold shock' increases the frequency of homology directed repair gene editing in induced pluripotent stem cells. *Scientific Reports* **8,** 2080. doi:10.1038/s41598-018-20358-5 (Feb. 2018).

172. Richardson, C. D., Ray, G. J., DeWitt, M. A., Curie, G. L. & Corn, J. E. Enhancing homology-directed genome editing by catalytically active and inactive CRISPR-Cas9 using asymmetric donor DNA. *Nature Biotechnology* **34,** 339–344. doi:10.1038/nbt.3481 (Mar. 2016).

173. Wang, K. *et al.* Efficient Generation of Orthologous Point Mutations in Pigs via CRISPR-assisted ssODN-mediated Homology-directed Repair. *Molecular Therapy - Nucleic Acids* **5,** e396. doi:10.1038/mtna.2016.101 (Jan. 2016).

174. Clement, K. *et al.* CRISPResso2 provides accurate and rapid genome editing sequence analysis. *Nature Biotechnology* **37,** 224–226. doi:10.1038/s41587-019-0032-3 (Mar. 2019).

175. Mesnil, M., Piccoli, C., Tiraby, G., Willecke, K. & Yamasaki, H. Bystander killing of cancer cells by herpes simplex virus thymidine kinase gene is mediated by connexins. *Proceedings of the National Academy of Sciences* **93,** 1831–1835. doi:10.1073/pnas.93.5.1831 (Mar. 1996).

176. Mesnil, M. & Yamasaki, H. Bystander Effect in Herpes Simplex Virus-Thymidine Kinase/Ganciclovir Cancer Gene Therapy: Role of Gap-junctional Intercellular Communication1. *Cancer Research* **60,** 3989–3999 (Aug. 2000).

177. Yamaguchi, I., Shibata, H., Seto, H. & Misato, T. ISOLATION AND PURIFICATION OF BLASTICIDIN S DEAMINASE FROM ASPERGILLUS TERREUS. *The Journal of Antibiotics* **28,** 7–14. doi:10.7164/antibiotics.28.7 (1975).

178. Rustom, A., Saffrich, R., Markovic, I., Walther, P. & Gerdes, H.-H. Nanotubular Highways for Intercellular Organelle Transport. *Science* **303,** 1007–1010. doi:10.1126/science.1093133 (Feb. 2004).

179. Gerdes, H.-H., Bukoreshtliev, N. V. & Barroso, J. F. Tunneling nanotubes: A new route for the exchange of components between animal cells. *FEBS Letters* **581,** 2194–2201. doi:10.1016/j.febslet.2007.03.071 (2007).

180. Gousset, K. *et al.* Prions hijack tunnelling nanotubes for intercellular spread. *Nature Cell Biology* **11,** 328–336. doi:10.1038/ncb1841 (Mar. 2009).

181. Costanzo, M. *et al.* Transfer of polyglutamine aggregates in neuronal cells occurs in tunneling nanotubes. *Journal of Cell Science* **126,** 3678–3685. doi:10.1242/jcs.126086 (Aug. 2013).

182. Sharma, M. & Subramaniam, S. Rhes travels from cell to cell and transports Huntington disease protein via TNT-like protrusion. *Journal of Cell Biology* **218,** 1972–1993. doi:10.1083/jcb.201807068 (May 2019).

183. Abounit, S. *et al.* Tunneling nanotubes spread fibrillar α-synuclein by intercellular trafficking of lysosomes. *The EMBO Journal* **35,** 2120–2138. doi:10.15252/embj.201593411 (Oct. 2016).

184. Wang, Y., Cui, J., Sun, X. & Zhang, Y. Tunneling-nanotube development in astrocytes depends on p53 activation. *Cell Death & Differentiation* **18,** 732–742. doi:10.1038/cdd.2010.147 (Apr. 2011).

185. Watanabe, K. *et al.* A ROCK inhibitor permits survival of dissociated human embryonic stem cells. *Nature Biotechnology* **25,** 681–686. doi:10.1038/nbt1310 (June 2007).

186. Ikeda, K. *et al.* Efficient scarless genome editing in human pluripotent stem cells. *Nature Methods* **15,** 1045–1047. doi:10.1038/s41592-018-0212-y (Dec. 2018).

187. Sleven, H. *et al.* De Novo Mutations in EBF3 Cause a Neurodevelopmental Syndrome. *The American Journal of Human Genetics* **100,** 138–150. doi:10.1016/j.ajhg.2016.11.020 (Jan. 2017).

188. Watson, L. M. *et al.* Dominant mutations in GRM1 cause spinocerebellar ataxia type 44. *The American Journal of Human Genetics* **101,** 451–458 (2017).

189. Lanza, A. M., Kim, D. S. & Alper, H. S. Evaluating the influence of selection markers on obtaining selected pools and stable cell lines in human cells. *Biotechnology Journal* **8,** 811–821. doi:10.1002/biot.201200364 (2013).

190. Lanza, D. G. *et al.* Comparative analysis of single-stranded DNA donors to generate conditional null mouse alleles. *BMC Biology* **16,** 69. doi:10.1186/s12915-018-0529-0 (June 2018).

191. Chen, X. *et al.* Deranged Calcium Signaling and Neurodegeneration in Spinocerebellar Ataxia Type 3. *Journal of Neuroscience* **28,** 12713–12724. doi:10.1523/JNEUROSCI.3909-08.2008 (Nov. 2008).

192. Liu, J. *et al.* Deranged Calcium Signaling and Neurodegeneration in Spinocerebellar Ataxia Type 2. *Journal of Neuroscience* **29,** 9148–9162. doi:10.1523/JNEUROSCI.0660-09.2009 (July 2009).

193. Tang, T.-S. *et al.* Huntingtin and Huntingtin-Associated Protein 1 Influence Neuronal Calcium Signaling Mediated by Inositol-(1,4,5) Triphosphate Receptor Type 1. *Neuron* **39,** 227–239. doi:10.1016/S0896-6273(03)00366-0 (July 2003).

194. Shilling, D. *et al.* Suppression of InsP3 Receptor-Mediated Ca2+ Signaling Alleviates Mutant Presenilin-Linked Familial Alzheimer's Disease Pathogenesis. *Journal of Neuroscience* **34,** 6910–6923. doi:10.1523/JNEUROSCI.5441-13.2014 (May 2014).

195. Huang, J.-j. *et al.* Functional expression of the Ca 2+ signaling machinery in human embryonic stem cells. *Acta Pharmacologica Sinica* **38,** 1663–1672. doi:10.1038/aps.2017.29 (Dec. 2017).

196. Sharp, A. H. *et al.* Differential cellular expression of isoforms of inositol 1,4,5-triphosphate receptors in neurons and glia in brain. *Journal of Comparative Neurology* **406,** 207–

220. doi:10.1002/(SICI)1096-9861(19990405)406:2<207::AID-CNE6>3.0.CO;2-7 (1999).

197. Zhang, Y. *et al.* Rapid Single-Step Induction of Functional Neurons from Human Pluripotent Stem Cells. *Neuron* **78,** 785–798. doi:10.1016/j.neuron.2013.05.029 (June 2013).

198. Rosa, F. *et al.* In Vitro Differentiated Human Stem Cell-Derived Neurons Reproduce Synaptic Synchronicity Arising during Neurodevelopment. *Stem Cell Reports* **15,** 22–37. doi:10.1016/j.stemcr.2020.05.015 (June 2020).

199. Flatschacher, D., Speckbacher, V. & Zeilinger, S. qRAT: an R-based stand-alone application for relative expression analysis of RT-qPCR data. *BMC Bioinformatics* **23,** 286. doi:10.1186/s12859-022-04823-7 (July 2022).

200. Benjamini, Y. & Hochberg, Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society: Series B (Methodological)* **57,** 289–300. doi:10.1111/j.2517-6161.1995.tb02031.x (1995).

201. Holden, P. & Horton, W. A. Crude subcellular fractionation of cultured mammalian cell lines. *BMC Research Notes* **2,** 243. doi:10.1186/1756-0500-2-243 (Dec. 2009).

202. Rosas-Arellano, A. *et al.* A simple solution for antibody signal enhancement in immunofluorescence and triple immunogold assays. *Histochemistry and Cell Biology* **146,** 421–430. doi:10.1007/s00418-016-1447-2 (Oct. 2016).

203. Schoepp, D. D., Goldsworthy, J., Johnson, B. G., Salhoff, C. R. & Baker, S. R. 3,5-dihydroxyphenylglycine is a highly selective agonist for phosphoinositide-linked metabotropic glutamate receptors in the rat hippocampus. *Journal of Neurochemistry* **63,** 769–772. doi:10.1046/j.1471-4159.1994.63020769.x (Aug. 1994).

204. Malmersjö, S. *et al.* Ca2+ and cAMP signaling in human embryonic stem cell-derived dopamine neurons. *Stem Cells and Development* **19,** 1355–1364. doi:10.1089/scd.2009.0436 (Sept. 2010).

205. Apáti, Á. *et al.* Calcium signaling in pluripotent stem cells. *Molecular and Cellular Endocrinology. Cellular Signaling in Physiology and Pathology* **353,** 57–67. doi:10.1016/j.mce.2011.08.038 (Apr. 2012).

206. Huang, J. *et al.* Coupling switch of P2Y-IP3 receptors mediates differential Ca2+ signaling in human embryonic stem cells and derived cardiovascular progenitor cells. *Purinergic Signalling* **12,** 465–478. doi:10.1007/s11302-016-9512-9 (Sept. 2016).

207. Ermakov, A. *et al.* A role for intracellular calcium downstream of G-protein signaling in undifferentiated human embryonic stem cell culture. *Stem Cell Research* **9,** 171–184. doi:10.1016/j.scr.2012.06.007 (Nov. 2012).

208. Hao, B., Webb, S. E., Miller, A. L. & Yue, J. The role of Ca2+ signaling on the self-renewal and neural differentiation of embryonic stem cells (ESCs). *Cell Calcium. Calcium signalling in stem cells* **59,** 67–74. doi:10.1016/j.ceca.2016.01.004 (Mar. 2016).

209. Zhang, P. *et al.* Minimal contribution of IP3R2 in cardiac differentiation and derived ventricular-like myocytes from human embryonic stem cells. *Acta Pharmacologica Sinica* **41,** 1576–1586. doi:10.1038/s41401-020-00528-w (Dec. 2020).

210. Liang, J. *et al.* Type 3 inositol 1,4,5-trisphosphate receptor negatively regulates apoptosis during mouse embryonic stem cell differentiation. *Cell Death & Differentiation* **17,** 1141–1154. doi:10.1038/cdd.2009.209 (July 2010).

211. Hao, B. *et al.* Role of STIM1 in survival and neural differentiation of mouse embryonic stem cells independent of Orai1-mediated Ca2+ entry. *Stem Cell Research* **12,** 452–466. doi:10.1016/j.scr.2013.12.005 (Mar. 2014).

212. Wang, Y.-J. *et al.* IP3R-mediated Ca2+ signals govern hematopoietic and cardiac divergence of Flk1+ cells via the calcineurin–NFATc3–Etv2 pathway. *Journal of Molecular Cell Biology* **9,** 274–288. doi:10.1093/jmcb/mjx014 (Aug. 2017).

213. Sidman, R. L., Lane, P. W. & Dickie, M. M. Staggerer, a New Mutation in the Mouse Affecting the Cerebellum. *Science* **137,** 610–612. doi:10.1126/science.137.3530.610 (Aug. 1962).

214. Fernandez-Gonzalez, A. *et al.* Purkinje cell degeneration (pcd) Phenotypes Caused by Mutations in the Axotomy-Induced Gene, Nna1. *Science* **295,** 1904–1906. doi:10.1126/science.1068912 (Mar. 2002).

215. Campbell, D. B. & Hess, E. J. Chromosomal localization of the neurological mouse mutations tottering (tg), Purkinje cell degeneration (pcd), and nervous (nr). *Molecular Brain Research* **37,** 79–84. doi:10.1016/0169-328X(95)00275-W (Apr. 1996).

216. Street, V. A. *et al.* The Type 1 Inositol 1,4,5-Trisphosphate Receptor Gene Is Altered in the opisthotonos Mouse. *Journal of Neuroscience* **17,** 635–645. doi:10.1523/JNEUROSCI.17-02-00635.1997 (Jan. 1997).

217. Reilly, L. *et al. A fully automated FAIMS-DIA proteomic pipeline for high-throughput characterization of iPSC-derived neurons* Nov. 2021. doi:10.1101/2021.11.24.469921.

218. Shi, Y., Kirwan, P., Smith, J., Robinson, H. P. & Livesey, F. J. Human cerebral cortex development from pluripotent stem cells to functional excitatory synapses. *Nature neuroscience* **15,** 10.1038/nn.3041. doi:10.1038/nn.3041 (Feb. 2012).

219. Nayler, S., Agarwal, D., Curion, F., Bowden, R. & Becker, E. B. E. High-resolution transcriptional landscape of xeno-free human induced pluripotent stem cell-derived cerebellar organoids. *Scientific Reports* **11,** 12959. doi:10.1038/s41598-021-91846-4 (June 2021).

220. Silva, T. P. *et al.* Maturation of Human Pluripotent Stem Cell-Derived Cerebellar Neurons in the Absence of Co-culture. *Frontiers in Bioengineering and Biotechnology* **8** (2020).

221. Ishida, Y. *et al.* Vulnerability of Purkinje Cells Generated from Spinocerebellar Ataxia Type 6 Patient-Derived iPSCs. *Cell Reports* **17,** 1482–1490. doi:10.1016/j.celrep.2016.10.026 (Nov. 2016).

222. Uchida, K. *et al.* Gene Knock-Outs of Inositol 1,4,5-Trisphosphate Receptors Types 1 and 2 Result in Perturbation of Cardiogenesis. *PLOS ONE* **5,** e12500. doi:10.1371/journal.pone.0012500 (Sept. 2010).

223. Yang, F. *et al.* Inositol 1,4,5-trisphosphate receptors are essential for fetal-maternal connection and embryo viability. *PLOS Genetics* **16,** e1008739. doi:10.1371/journal.pgen.1008739 (Apr. 2020).

224. Allen, D., Rosenberg, M. & Hendel, A. Using Synthetically Engineered Guide RNAs to Enhance CRISPR Genome Editing Systems in Mammalian Cells. *Frontiers in Genome Editing* **2** (2021).

225. IzuMi, M. *et al.* Blasticidin S-resistance gene (bsr): A novel selectable marker for mammalian cells. *Experimental Cell Research* **197,** 229–233. doi:10.1016/0014-4827(91)90427-V (Dec. 1991).

226. Byrne, S. M., Mali, P. & Church, G. M. in *Methods in Enzymology* 119–138 (Elsevier, 2014). doi:10.1016/B978-0-12-801185-0.00006-4.

227. Carapuça, E., Azzoni, A. R., Prazeres, D. M. F., Monteiro, G. A. & Mergulhão, F. J. M. Time-course determination of plasmid content in eukaryotic and prokaryotic cells using Real-Time PCR. *Molecular Biotechnology* **37,** 120–126. doi:10.1007/s12033-007-0007-3 (Oct. 2007).

228. Sunstrom, N.-A. *et al.* Regulated autocrine growth of CHO cells. *Cytotechnology* **34,** 39–46. doi:10.1023/A:1008194529175 (Oct. 2000).

229. Hanna, J. *et al.* Human embryonic stem cells with biological and epigenetic characteristics similar to those of mouse ESCs. *Proceedings of the National Academy of Sciences* **107,** 9222–9227. doi:10.1073/pnas.1004584107 (May 2010).

230. Kuijk, E. *et al.* The mutational impact of culturing human pluripotent and adult stem cells. *Nature Communications* **11,** 2493. doi:10.1038/s41467-020-16323-4 (May 2020).

231. Wu, J., Fan, Y. & Tzanakakis, E. S. Increased Culture Density Is Linked to Decelerated Proliferation, Prolonged G1 Phase, and Enhanced Propensity for Differentiation of Self-Renewing Human Pluripotent Stem Cells. *Stem Cells and Development* **24,** 892–903. doi:10.1089/scd.2014.0384 (Apr. 2015).

232. Ohgushi, M. *et al.* Molecular Pathway and Cell State Responsible for Dissociation-Induced Apoptosis in Human Pluripotent Stem Cells. *Cell Stem Cell* **7,** 225–239. doi:10.1016/j.stem.2010.06.018 (Aug. 2010).

233. Chen, Y. *et al.* A versatile polypharmacology platform promotes cytoprotection and viability of human pluripotent and differentiated cells. *Nature Methods* **18,** 528–541. doi:10.1038/s41592-021-01126-2 (May 2021).

234. Lim, S., Yocum, R. R., Silver, P. A. & Way, J. C. High spontaneous integration rates of end-modified linear DNAs upon mammalian cell transfection. *Scientific Reports* **13,** 6835. doi:10.1038/s41598-023-33862-0 (Apr. 2023).

235. Thompson, O. *et al.* Low rates of mutation in clinical grade human pluripotent stem cells under different culture conditions. *Nature Communications* **11,** 1528. doi:10.1038/s41467-020-15271-3 (Mar. 2020).

236. Hasaart, K. A. L. *et al.* Human induced pluripotent stem cells display a similar mutation burden as embryonic pluripotent cells in vivo. *iScience* **25,** 103736. doi:10.1016/j.isci.2022.103736 (Feb. 2022).

237. Maynard, S. *et al.* Human Embryonic Stem Cells Have Enhanced Repair of Multiple Forms of DNA Damage. *Stem Cells* **26,** 2266–2274. doi:10.1634/stemcells.2007-1041 (Sept. 2008).

238. Liu, J. C., Lerou, P. H. & Lahav, G. Stem cells: balancing resistance and sensitivity to DNA damage. *Trends in Cell Biology* **24,** 268–274. doi:10.1016/j.tcb.2014.03.002 (May 2014).

239. Drocourt, D., Calmels, T., Reynes, J.-P., Baron, M. & Tiraby, G. Cassettes of the Streptoalloteichus hindustanus ble gene for transformation of lower and higher eukaryotes to phleomycin resistance. *Nucleic Acids Research* **18,** 4009. doi:10.1093/nar/18.13.4009 (July 1990).

240. Hejazi, M. S. GB2390606A (2004).

241. Trastoy, M. O., Defais, M. & Larminat, F. Resistance to the antibiotic Zeocin by stable expression of the Sh ble gene does not fully suppress Zeocin-induced DNA cleavage

in human cells. *Mutagenesis* **20,** 111–114. doi:10.1093/mutage/gei016 (Mar. 2005).

242. Dow, L. E. *et al.* Inducible in vivo genome editing with CRISPR-Cas9. *Nature Biotechnology* **33,** 390–394. doi:10.1038/nbt.3155 (Apr. 2015).

243. Nihongaki, Y., Kawano, F., Nakajima, T. & Sato, M. Photoactivatable CRISPR-Cas9 for optogenetic genome editing. *Nature Biotechnology* **33,** 755–760. doi:10.1038/nbt.3245 (July 2015).

244. Rose, J. C. *et al.* Rapidly inducible Cas9 and DSB-ddPCR to probe editing kinetics. *Nature Methods* **14,** 891–896. doi:10.1038/nmeth.4368 (Sept. 2017).

245. Senturk, S. *et al.* Rapid and tunable method to temporally control gene editing based on conditional Cas9 stabilization. *Nature Communications* **8,** 14370. doi:10.1038/ncomms14370 (Feb. 2017).

246. Zetsche, B., Volz, S. E. & Zhang, F. A split-Cas9 architecture for inducible genome editing and transcription modulation. *Nature Biotechnology* **33,** 139–142. doi:10.1038/nbt.3149 (Feb. 2015).

247. Banaszynski, L. A., Chen, L.-c., Maynard-Smith, L. A., Ooi, A. G. L. & Wandless, T. J. A Rapid, Reversible, and Tunable Method to Regulate Protein Function in Living Cells Using Synthetic Small Molecules. *Cell* **126,** 995–1004. doi:10.1016/j.cell.2006.07.025 (Sept. 2006).

248. Berridge, M. J. The Inositol Trisphosphate/Calcium Signaling Pathway in Health and Disease. *Physiological Reviews* **96,** 1261–1296. doi:10.1152/physrev.00006.2016 (Oct. 2016).

249. Forman, O. P., De Risio, L., Matiasek, K., Platt, S. & Mellersh, C. Spinocerebellar ataxia in the Italian Spinone dog is associated with an intronic GAA repeat expansion in ITPR1. *Mammalian Genome* **26,** 108–117. doi:10.1007/s00335-014-9547-6 (Feb. 2015).

250. Ando, H. *et al.* IRBIT Suppresses IP3 Receptor Activity by Competing with IP3 for the Common Binding Site on the IP3 Receptor. *Molecular Cell* **22,** 795–806. doi:10.1016/j.molcel.2006.05.017 (June 2006).

251. Hirota, J., Ando, H., Hamada, K. & Mikoshiba, K. Carbonic anhydrase-related protein is a novel binding protein for inositol 1,4,5-trisphosphate receptor type 1. *The Biochemical Journal* **372,** 435–441. doi:10.1042/BJ20030110 (June 2003).

252. Wang, T. *et al.* Identification and characterization of essential genes in the human genome. *Science* **350,** 1096–1101. doi:10.1126/science.aac7041 (Nov. 2015).

253. Blomen, V. A. *et al.* Gene essentiality and synthetic lethality in haploid human cells. *Science* **350,** 1092–1096. doi:10.1126/science.aac7557 (Nov. 2015).

254. Rönkkö, J. *et al.* Human IP3 receptor triple knockout stem cells remain pluripotent despite altered mitochondrial metabolism. *Cell Calcium* **114,** 102782. doi:10.1016/j.ceca.2023.102782 (Sept. 2023).

255. Tian, R. *et al.* CRISPR Interference-Based Platform for Multimodal Genetic Screens in Human iPSC-Derived Neurons. *Neuron* **104,** 239–255.e12. doi:10.1016/j.neuron.2019.07.014 (Oct. 2019).

256. Hulme, A. J., Maksour, S., St-Clair Glover, M., Miellet, S. & Dottori, M. Making neurons, made easy: The use of Neurogenin-2 in neuronal differentiation. *Stem Cell Reports* **17,** 14–34. doi:10.1016/j.stemcr.2021.11.015 (Jan. 2022).

257. Lin, H.-C. *et al.* NGN2 induces diverse neuron types from human pluripotency. *Stem Cell Reports* **16,** 2118–2127. doi:10.1016/j.stemcr.2021.07.006 (Sept. 2021).

258. Futatsugi, A. *et al.* IP3 receptor types 2 and 3 mediate exocrine secretion underlying energy metabolism. *Science (New York, N.Y.)* **309,** 2232–2234. doi:10.1126/science.1114110 (Sept. 2005).

259. Matsumoto, M. & Kato, K. Altered calcium dynamics in cultured cerebellar cells from IP3R1-deficient mice. *Neuroreport* **12,** 3471–3474. doi:10.1097/00001756-200111160-00019 (Nov. 2001).

260. Sathyanesan, A. *et al.* Emerging connections between cerebellar development, behaviour and complex brain disorders. *Nature Reviews Neuroscience* **20,** 298–313. doi:10.1038/s41583-019-0152-2 (May 2019).

261. Genazzani, A. A., Carafoli, E. & Guerini, D. Calcineurin controls inositol 1,4,5-trisphosphate type 1 receptor expression in neurons. *Proceedings of the National Academy of Sciences* **96,** 5797–5801. doi:10.1073/pnas.96.10.5797 (May 1999).

262. Hisatsune, C. *et al.* IP3R1 deficiency in the cerebellum/brainstem causes basal ganglia-independent dystonia by triggering tonic Purkinje cell firings in mice. *Frontiers in Neural Circuits* **7** (2013).

263. Sugawara, T. *et al.* Type 1 Inositol Trisphosphate Receptor Regulates Cerebellar Circuits by Maintaining the Spine Morphology of Purkinje Cells in Adult Mice. *Journal of Neuroscience* **33,** 12186–12196. doi:10.1523/JNEUROSCI.0545-13.2013 (July 2013).

264. Inoue, T., Kato, K., Kohda, K. & Mikoshiba, K. Type 1 Inositol 1,4,5-Trisphosphate Receptor Is Required for Induction of Long-Term Depression in Cerebellar Purkinje Neurons. *Journal of Neuroscience* **18,** 5366–5373. doi:10.1523/JNEUROSCI.18-14-05366.1998 (July 1998).

265. Hisatsune, C. *et al.* Inositol 1,4,5-Trisphosphate Receptor Type 1 in Granule Cells, Not in Purkinje Cells, Regulates the Dendritic Morphology of Purkinje Cells through Brain-Derived Neurotrophic Factor Production. *Journal of Neuroscience* **26,** 10916–10924. doi:10.1523/JNEUROSCI.3269-06.2006 (Oct. 2006).

266. Miller, J. A. *et al.* Transcriptional landscape of the prenatal human brain. *Nature* **508,** 199–206. doi:10.1038/nature13185 (Apr. 2014).

267. Gardner, R. J. M. *et al.* Spinocerebellar ataxia type 15. *The Cerebellum* **4,** 47–50. doi:10.1080/14734220410019029 (Mar. 2005).

A | Oligonucleotide Sequences

Oligonucleotide ID	**Sequence (5'-3')**
EMX1 sgRNA #1	GTCACCTCCAATGACTAGGGT
EMX1 sgRNA #3	GAGTCCGAGCAGAAGAAGAA
Exon10-L1R1_fwd	tatccgctcacaattccacagtaagggtagggtggagaaag
Exon10-L1R1_rev	taatcaatgtcgaccaggtgctgacagtcaacacaagaaag
Exon58_NGS_F	ACACTCTTTCCCTACACGACGCTCTTCCG ATCTTTTCGTGTGCCTGGTGAGATGG
Exon58_NGS_R	GACTGGAGTTCAGACGTGTGCTCTTCCGA TCTGAGCCTGATGAGCTAGTTCCGC
Exon58-L1R1_fwd	tatccgctcacaattccacaAGTGTAAACTGAAGACGC
Exon58-L1R1_rev	taatcaatgtcgaccaggtgTGACAATGCACATCAGCAG
Exon59_NGS_F	ACACTCTTTCCCTACACGACGCTCTTCCGATC TATGAGGTCTCACTTGAGCTGTG
Exon59_NGS_R	GACTGGAGTTCAGACGTGTGCTCTTCCGATC TGGCAGAGGGTAGAAGAGTCAAG
Exon59-L1R1_fwd	tatccgctcacaattccacaGTGAGTGTCGCTGGCCAG
Exon59-L1R1_rev	taatcaatgtcgaccaggtgAGAGGTGTCTGGATCAGTCTG
ITPR1 ex10 gDNA L1	GCAGCCCTGCAAGAGTGATA
ITPR1 ex10 gDNA L2	TATCAGCAGCCCTGCAAGAG
ITPR1 ex10 gDNA R1	CGTGCCCTTCACCAAACAAG
ITPR1 ex10 gDNA R2	CCACCCAGACATTCACCCTC
ITPR1 ex10 gRNA 1	GCACGTCTTCCTGAGAACCA
ITPR1 ex10 gRNA 2	TGTTGACTGTCAGGGTGACG
ITPR1 ex10 gRNA 3	TCTTCCTGAGAACCACGGGC
ITPR1 ex10 gRNA 4	ACTGTCAGGGTGACGTGGTG
ITPR1 ex10 gRNA 5	TTCTCACCTGTGACGAACAC

ITPR1 ex10 gRNA 6	ACTGCCGGCCCGTGGTTCTC
ITPR1 ex10 gRNA 7	TGTGGCCGACTGCCGGCCCG
ITPR1 ex58 gDNA L1	GCCTACTACTGGGCTAAGCG
ITPR1 ex58 gDNA L2	GGCAGGCCAAATGGTGATTG
ITPR1 ex58 gDNA R1	CAGCGACACTCACCACAGAT
ITPR1 ex58 gDNA R2	TCACGGCAGAGGGTAGAAGA
ITPR1 ex58 gRNA 1	AGAGGTAAATTAATCCCGAG
ITPR1 ex58 gRNA 2	AAGAGGTAAATTAATCCCGA
ITPR1 ex58 gRNA 3fix	ACTCAGGAAACCGTCCAAAG
ITPR1 ex58 gRNA 4	AAAGAGGTAAATTAATCCCG
ITPR1 ex59 gDNA L1	CGGGGGTTAGCGATGAAAAC
ITPR1 ex59 gDNA L2	TTAGCGGGGGTTAGCGATGA
ITPR1 ex59 gDNA R1	TTACTCAGCTCCCCAGATTCCT
ITPR1 ex59 gDNA R2	CTCCCCAGATTCCTCCATGATT
ITPR1 ex59 gRNA 0	TTATCTGTGGTGAGTGTCGC
ITPR1 ex59 gRNA 1	CTCACCACAGATAAAGCACG
ITPR1 ex59 gRNA 2	ATAACTCTAGCAGCAAACAG
ITPR1 ex59 gRNA 3	AAGACCACGTGCTTTATCTG
ITPR1 ex59 gRNA 7	TTGACACTTTTGCTGACCTG
ITPR1 Splice Ex10 F2	CTACTCAACAAACTGCACCACG
ITPR1 Splice Ex10 R1	AGATGCTTGAAACGGAAAAGGC
ITPR1 Splice Ex10 R2	GATGCTTGAAACGGAAAAGGCT
ITPR1_ex10_LEI_F	GGGCTGATACTCACTCTAGTGC
ITPR1_ex10_LEI_R1	cttcttgagacaaaggcttggc
ITPR1_ex10_LEI_R2	agatggggatgctgttgattgt
ITPR1_ex10_REI_F	tgaattgctgccctctggttat
ITPR1_ex10_REI_R	CTCATTGCCTCTTTCACCAAGC
ITPR1_ex56-60_F	GATGTGTGTAGGGTGGAGAGTG

ITPR1_ex58_LEI_F1	AGGCCAAATGGTGATTGCTAGA
ITPR1_ex58_LEI_F2	TGATGCAGGTGATGTTGGAAGA
ITPR1_ex58_LEI_F3	GAGTTCCTGTTCTCCGATGTGT
ITPR1_ex58_REI_R1	GGTAGAAGCTCAGTCAGAACCC
ITPR1_ex58_REI_R2	GTTCACTACTTCTCCTCCCGTG
ITPR1_ex59_LEI_F1	TGTCAGGCGTAGGAGGTTAAAC
ITPR1_ex59_LEI_F2	TCAGGGGTTGTTGGAAGATGAG
ITPR1_ex59_REI_R1	TACATGCCTCCTGATACGATGC
ITPR1_ex59_REI_R2	ATCACTGTCAGCCACTCACATT
ITPR1_ex58-60_F	AGATCTTGAAGACCACGTGCTT
ITPR1_ex58-60_R	TCCTTCAGAATCACTGCTGACC
ITPR1ex10screenF1	GCAGGAGGCTTGGTTCCATT
ITPR1ex10screenF2	AGTTGTTGTGAGGGGTACGTG
ITPR1ex10screenR1	ACTATGCCATCCCCAACCCA
ITPR1ex10screenR2	TCATTTCTGCCCCATCACCA
ITPR1ex58screenF1	ATTAGCGGGGGTTAGCGATG
ITPR1ex58screenF2	GCCCCAGACGACTTTAGCTT
ITPR1ex58screenR1	CAAAGCCAACCCTTTCCACG
ITPR1ex58screenR2	CAGCGACACTCACCACAGAT
Ex10-128-Silent	A*C*CCAGGAGCCCTTTCTCCACCCTACCCT TACCTCCACCTCCCACAGGGCTTTTGAACTG GTGGCAGATGTGGCCGACTGCCGGCCAGTG GTTCTCAGGAAGACGTGCTGCTTCTTCCTGT GTTCG*T*C

Ex10-128-R269G	A*C*CCAGGAGCCCTTTCTCCACCCTACCCT TACCTCCACCTCCCACAGGGCTTTTGAACTG GTGGCAGATGTGGCCGACTGCCCGCCAGTG GTTCTCAGGAAGACGTGCTGCTTCTTCCTGT GTTCG*T*C
Ex58-60-Silent	T*T*TGCCCTTCCTTCCCCTCGGGATTAATTT ACCTCGTTGGACGGTTTCCTGAGTACAT*C*T
Ex58-150-G2554R	A*C*CAACCACCCCACCCGCCCCCTTTGCCC TTCCTTCCCCTCGGGATTAATTTACCTCCTTG GACGGTTTCCTGAGTACATCTCCTACTCCAC CCCGGCTCCGCAGCCCGTGACTCAGCACGG TGACAATGCACATCAGCAGCGTCTCA*C*A
Ex59-60-Silent	C*C*AGACACCTCTCTTTTCCAGGAACCCTT GTTTGCTGCTAGAGTTATTTATGACCTCT*T*G
Illumina Adaptor (Forward Primer)	ACACTCTTTCCCTACACGACGCTCTTCCGATCT
Illumina Adaptor (Reverse Primer)	GACTGGAGTTCAGACGTGTGCTCTTCCGATCT

B | Validation of Cassette-Detecting PCRs

Prior to their use for colony screening, PCRs designed to detect on-target cassette insertion were validated both against untransfected WT H9 hESC DNA (negative control) and against PTD20 blasticidin-resistant H9 hESCs (positive control, described in chapter 4). DNA extraction followed the optimised ExtractPCR protocol as described in the general methods chapter. Between 1-6 sets of primers were designed for each Left External to Internal (LEI) or Right

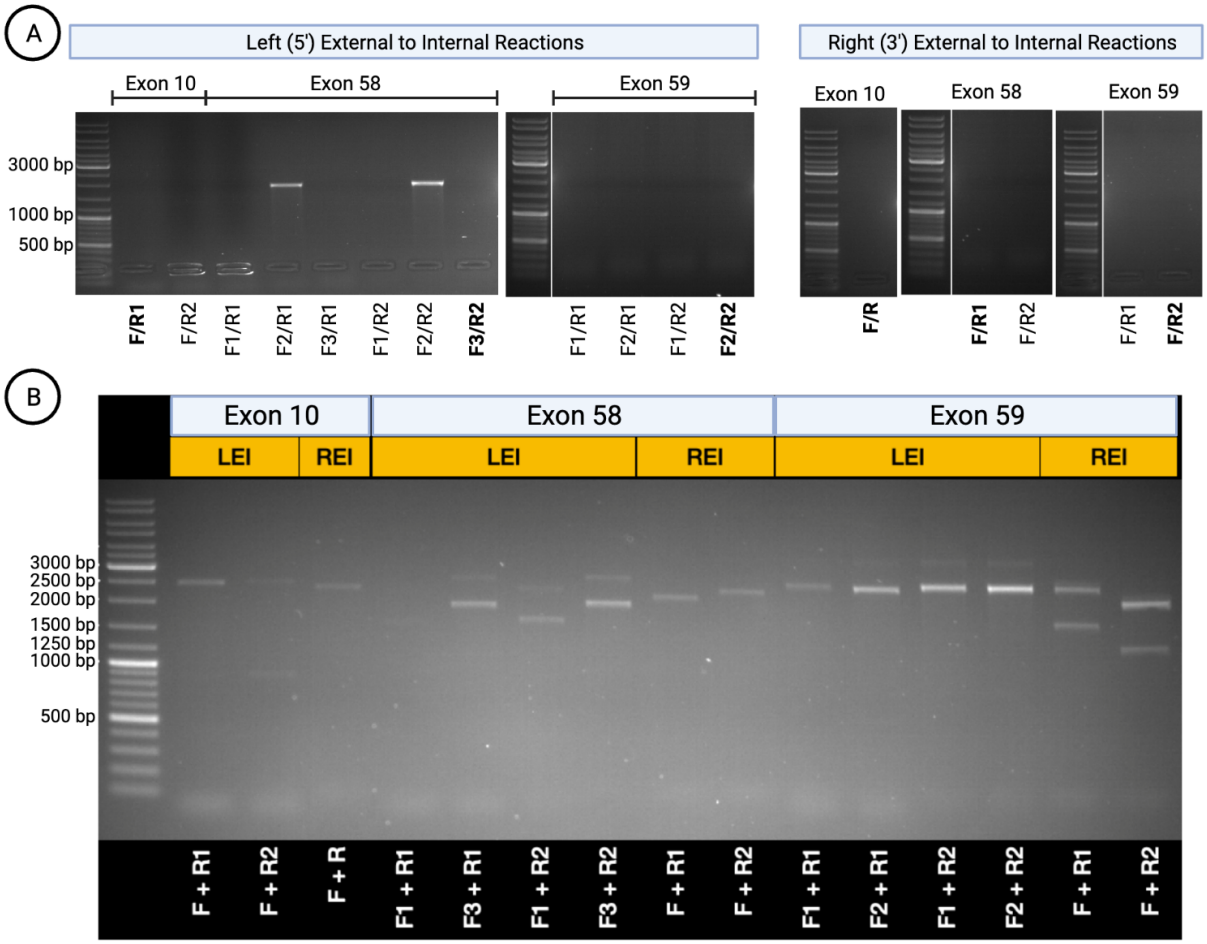

Figure B.1: Validation of PCRs used for detection of cassette insertion in the three ITPR1 target loci. (A) PCRs against untransfected wild-type H9 hESCs, therefore representing negative controls. Two primer pairs generated nonspecific amplification (Exon 58 F2/R1 and F2/R2) and were discarded. Primer pairs in bold were eventually used for screening and confirmatory experiments reported in this thesis. (B) PCRs against CIRCES Insertion Stage pooled cells 20 days post-transfection and resistant to blasticidin 3.0 μg/mL selection.

External to Internal (REI) reaction. PCR reactions were as described in the general methods section, using 0.5μL of extracted DNA per 10μL reaction, which was amplified for 35 cycles. The full 10μL reaction was loaded onto each lane of an agarose gel and subjected to electrophoresis.

Two LEI primer pairs designed to detect cassette insertion into the Exon 58 locus generated nonspecific amplification from WT H9 hESC templates (Fig. B.1A, Exon 58 primers F2/R1 and F2/R2). These were not used for any other experiments.

In CIRCES-transfected blasticidin-selected H9 hESCS, all but one (Exon 58 F1/R1) set of primer pairs led to successful generation of amplicons with sizes compatible with on-target cassette insertion (Fig. B.1B). Notably, almost all reactions identified 2 amplicons of compatible sizes, indicative of at least two different HDR-mediated on-target insertion events, one of which at least containing additional insertions or deletions.

One primer set from each target was chosen to be used for colony screening procedures. These are highlighted in bold in Fig. B.1A and listed in 7.

C | Selection Curves

Puromycin Kill Curve in HEK293T Cells

Puromycin was used to select transfected cells expressing the sgRNA/eSpCas9-T2A-puro vectors. A kill curve was performed prior to its use in order to establish the minimal effective concentration required for selection.

HEK293T cells were transfected either with a eSpCas9-T2A-puro plasmid or with a similarly-sized GFP construct (lacking the PuroR gene) and selected with different puromycin concentrations for 3 days.

One day prior to transfection, 1.3 to 1.5 x 10e5 WT HEK293T cells were seeded onto wells of 24-well-plates. The day of transfection, cells were at 80% density and were transfected with either 500ng EF-GFP construct or 500ng eSpCas9-T2A-puro/ITPR1-Ex58-sgRNA-1 using Mirus TransIT LT1.

One day after transfection, all wells close to fully confluent and over 90% of GFP-transfected cells expressed GFP, indicating excellent transfection efficiency. Positive seletion with 0.0μg/mL (control), 1.0μg/mL, 2.0μg/mL or 3.0μg/mL puromycin was initiated.

Significant selection was detected on GFP-transfected wells within 24 hrs of selection, was mostly completed by 48 hrs and by the 3rd day of selection GFP-transfected cells across all three concentrations tested had been completely killed. Cells transfected with a eSpCas9-T2A-puro vector were fully confluent at all concentrations of puromycin and had no visual differences to control (no puromycin selection) cells (Fig. C.2).

These results indicated that for HEK293T cells were naturally susceptible to puromycin concentrations equal or above to 1.0 μg/mL and that the transfected eSpCas9-T2A-puro vectors elicited resistance to a puromycin concentration of at least 3.0 μg/mL.

Based on these results, puromycin 2.0ug/mL was chosen as the optimal concentration for selecting eSpCas9-T2A-puro-transfected cells.

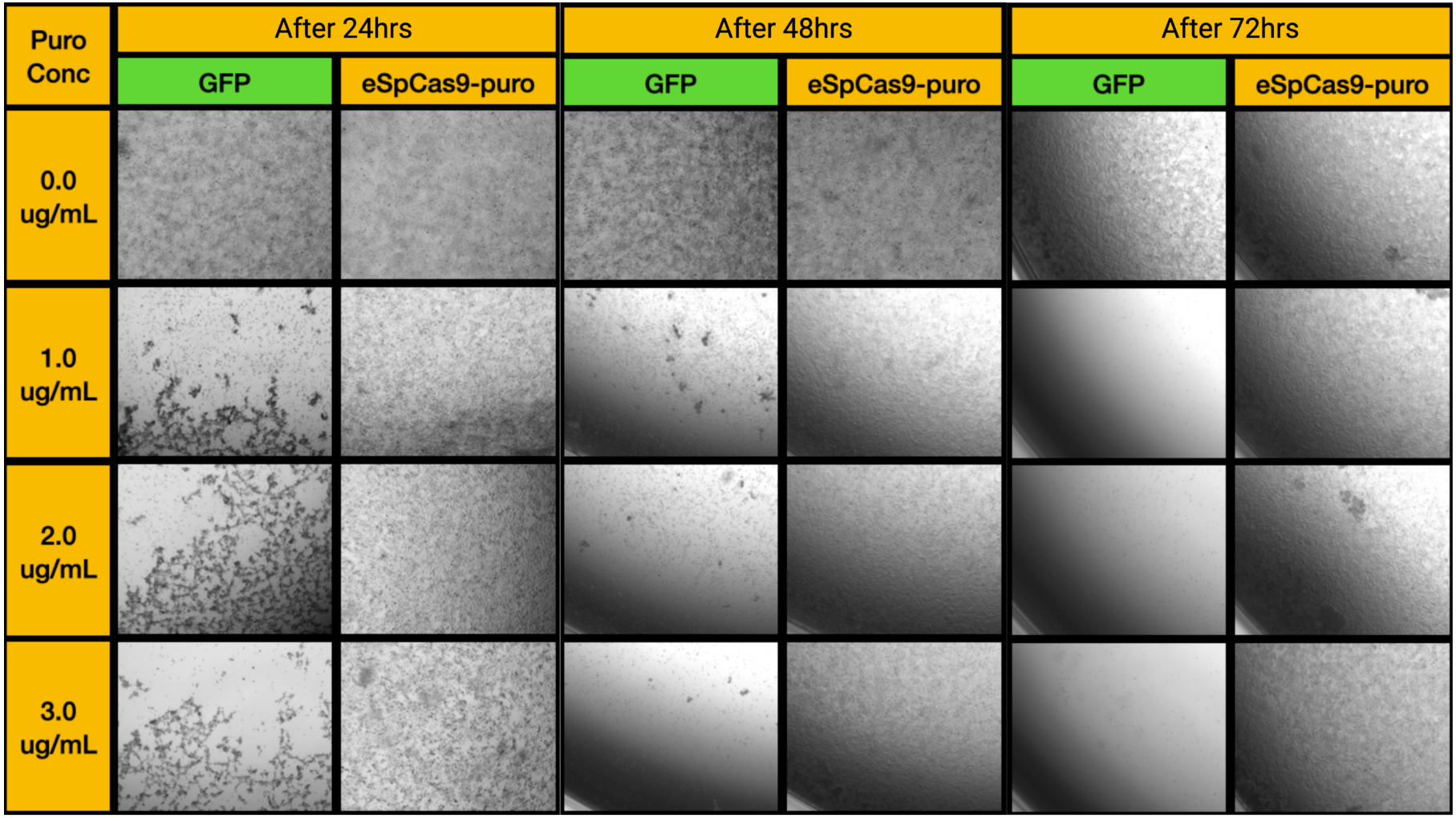

Figure C.1: Puromycin kill curve in HEK293T cells. WT cells were transfected either with 500ng of an GFP construct or with 500ng of eSpCas9-T2A-puro/sgRNA vector and selected with different concentrations of puromycin for 3 days. Brightfield 4x images were captured after 1, 2 and 3 days under selection. Due to floating cellular debris in the center of wells, images were taken near well edges. Culture medium was not changed until day 3, at which point cellular debris were aspirated and cells were carefully rinsed once with DPBS before being imaged.

Blasticidin Kill Curve in HEK293T Cells

To determine minimum effective blasticidin concentrations for adequate selection of cells expressing the Blasticidin-S-Deaminase (BSD) resistance gene, HEK293T cells were transfected either with a cassette-containing HDR plasmid or with a similarly-sized GFP construct (lacking the BSD gene) and selected with different blasticidin concentrations over a 5-day period.

One day prior to transfection, 1.4 x 10e5 WT HEK293T cells were seeded onto wells of 24-well-plates. The day of transfection, cells were at 80% density and were transfected with either 500ng EF-GFP construct or 500ng ITPR1 Exon 10-cassette HDR template using Mirus TransIT LT1 reagent.

One day after transfection, all wells close to fully confluent and over 90% of GFP-transfected cells expressed GFP, indicating excellent transfection efficiency. Positive seletion with 0.0μg/mL (control), 2.5μg/mL, 5.0μg/mL or 10.0μg/mL blasticidin was initiated.

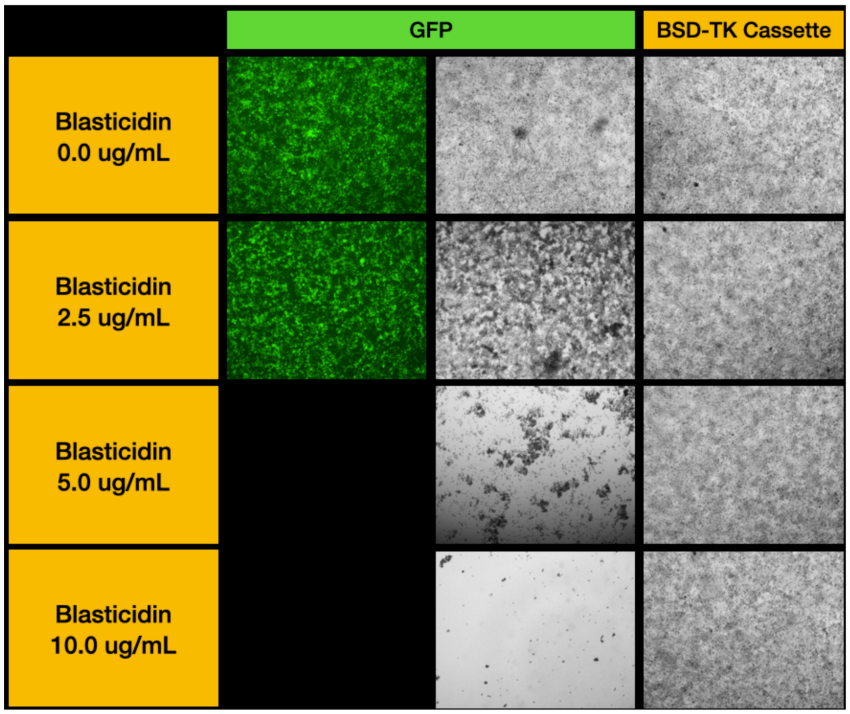

Figure C.2: Blasticidin kill curve in HEK293T cells. WT cells were transfected either with 500ng of an GFP construct or with 500ng of ITPR1 Exon 10-BSD-TK-Cassette HDR template and selected with different concentrations of blasticidin for 5 days. Brightfield and fluorescent 4x images were captured after cellular debris were aspirated and cells were carefully rinsed once with DPBS.

Initial signs of selection were detected on GFP-transfected wells after 60 hrs under selection,

and by the 3rd day of selection GFP-transfected cells in the 5.0 and 10.0μg/mL concentrations had mostly detached.

After 5 days under selection, cells transfected with a cassette HDR template were fully confluent at all concentrations of blasticidin and had no visual differences to control (no blasticidin selection) cells (Fig. C.2). In comparison, GFP-transfected cells were fully killed in wells selected with 10.0μg/mL blastiidin, >95% killed in 5.0μg/mL wells and <25% killed in 2.5μg/mL wells.

These results indicated that for HEK293T cells 10.0μg/mL blasticidin was the optimal concentration for selection of cassette-expressing cells.

Blasticidin Kill Curve in H9 hES Cells

To determine the minimum effective blasticidin concentration required to select H9 hESCs, a kill curve was performed on WT H9 hESCs transfected with a GFP construct not containing the BSD resistance gene.

Transfection followed the optimised Mirus TransIT LT1 protocol as described in the general methods section. Two days after transfection, cells were dissociated to single cells and seeded at low density in 24-well plates. The following day (PTD 3), blasticidin selection was started at concentrations ranging from 0.0 (control) to 10.0μg/mL. Maintenance medium containing blasticidin was changed daily. Cell density at each well was registered daily by visual estimation using a EVOS XL inverted microscope at 4x and 10x magnification.

After 72 hrs under selection, all wells at blasticidin concentrations equal or greater than 4.0μg/mL were completely killed (Fig. C.3). Blasticidin 2.0μg/mL was much slower, only achieving complete selection after 7 days (Fig. C.3).

Based on these results, blasticidin 3.0μg/mL was chosen to be used as a standard concentration for selecting CIRCES-transfected H9 hESCs.

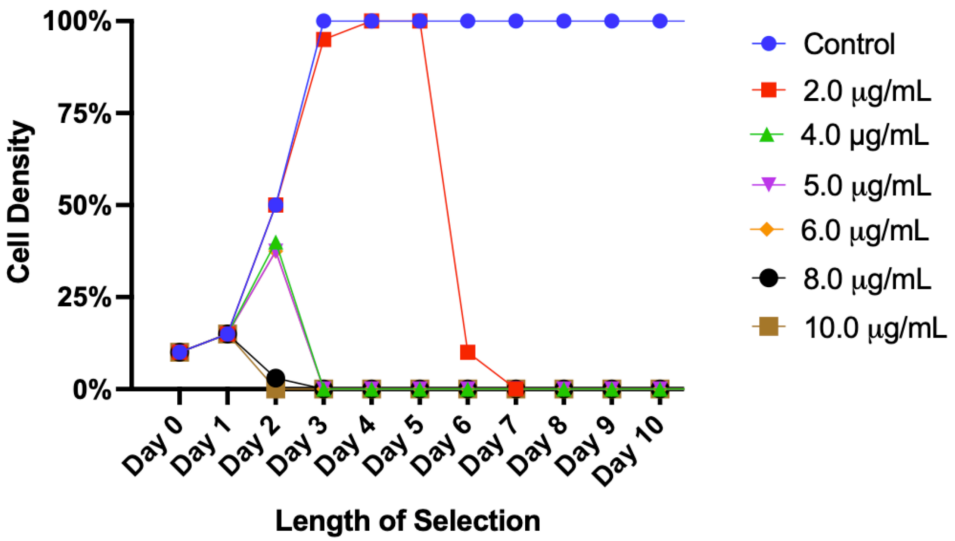

Figure C.3: Blasticidin kill curve in hESCs. WT H9 hESCs were transfected with a GFP construct and split at low density in 24-well plates. Selection with blasticidin at various concentrations was started on the 3rd post-transfection day. Density was assessed daily by visual estimation using an inverted EVOS microscope. The horizontal axis refers to the length under active blasticidin selection.

Ganciclovir Toxicity Curve in H9 hES Cells

Ganciclovir is known to show dose-dependent toxicity to human cells in vitro and in vivo (Sommadossi 1987). To evaluate the impacts of ganciclovir on human stem cells under concentrations that might be used for negative selection, WT H9 hESC were cultured under different ganciclovir concentrations and assessed for effects on cell proliferation.

WT H9 hESCs were seeded at low density in 24-well plates and started the following day on ganciclovir concentrations ranging between 0-40μM. Culture medium was changed daily and replenished with fresh ganciclovir. Cell density was visually estimated on a daily basis using an inverted EVOS XL microscope at 4x magnification.

All ganciclovir concentrations above 2μM led to reduced cell densities relative to untreated cells (Fig. C.4). Concentrations of 20μM and 40μM were extremely toxic, almost completely inhibiting cell proliferation.

Based on these results I decided to use 5μM ganciclovir for negative selection experiments in H9 hESCs.

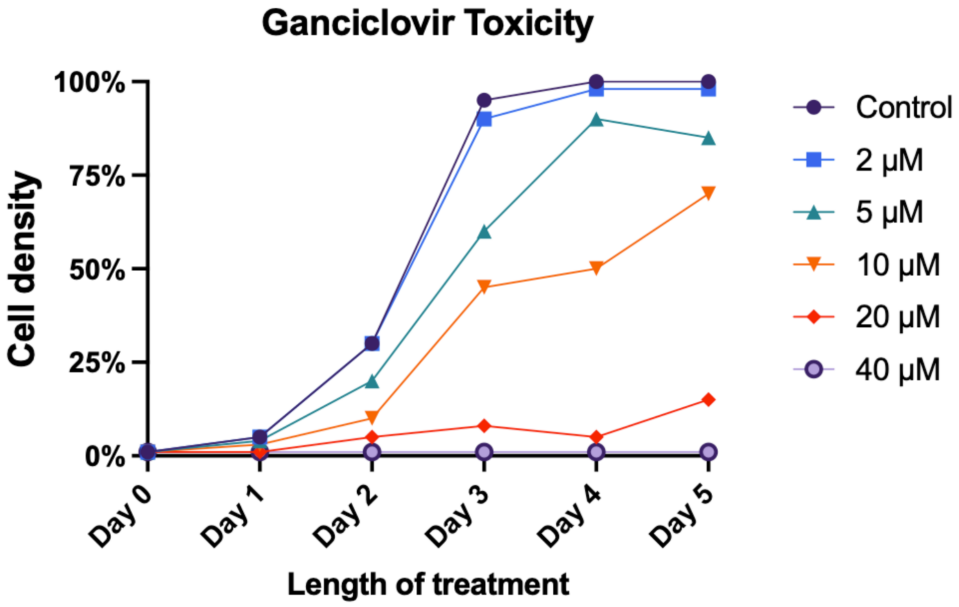

Figure C.4: Ganciclovir toxicity in H9 hESCs. Untransfected WT H9 hESCs at low densities were exposed to different concentrations of ganciclovir and assessed daily for cell density.