Understanding neurodegenerative diseases in human iPS cell models by genome editing with CRISPR-Cas9

Thesis dissertation submitted for the degree of
Master of Philosophy

Shamma Qarin

Supervisor: William C. Skarnes, PhD

Wellcome Trust Sanger Institute

Homerton College

University of Cambridge

Contents

Declaration

In accordance with the policies and guidelines of the University of Cambridge, the following declarations are made:

I hereby declare that this thesis, conducted at the Wellcome Trust Sanger Institute from January to September 2016, is my own work on the research conducted by me and contains no material which has been submitted for the award of any other degree at any university or equivalent institution. This thesis, to the best of my knowledge and belief, contains no material previously published or written by another person, except where due reference and acknowledgement is made in the text.

Name: _____

Signature: _____

Date: _____

Acknowledgements

I am grateful to The Almighty who has blessed and guided me in all my goals and aspirations that I have attempted for so far.

I would like to express my sincere gratitude and deepest appreciation towards my supervisor, Dr William Skarnes, for providing me with all the expert advice and guidance throughout my work. I am eternally grateful to him for all his support, inspiration, and encouragement, be it academically, professionally, or personally. Without him, this project would never have been possible. I would also like to sincerely thank my thesis advisory committee, consisting of Professor Roger barker and Dr Kosuke Yusa, for providing me with the expert advice to further improve my work.

I would also take the opportunity to express my sincere thanks to Dr Melanie Keene, my Graduate Tutor, and Mr Dhiru Karia, Finance Tutor, who have been very supportive and cooperative during times of difficulty and played integral roles in holding up my confidence and motivation. I would like to thank Homerton College for providing me with all the academic and pastoral support and well-being during my entire stay at Cambridge.

I place on record, my earnest thankfulness to my Graduate Administrators at the Sanger Institute, Dr Christina Hedberg-Delouka and Dr Annabel Smith, for cooperating and providing me with all the necessary technical and academic support.

I would like to express my earnest thankfulness to Dr Peri Tate who, without an exception, has always been there whenever needed and to my colleagues for always encouraging and motivating me during times of frustration!

Last but not the least, outside my workplace, I would express my earnest gratitude and heartfelt appreciation to my parents, Dr Sharif Ashrafuzzaman and Mrs Syeda Rawnak Ara, for their constant moral and financial support in every step of my life which words cannot justify, and to my late grandmother, Salma Khatun, who laid the first stone of the path leading to where I am today.

I place on record, my sense of gratitude to everyone who, directly or indirectly, have lent their helping hand in this venture.

Abbreviations

AAAS	Achalasia-Addisonianism-Alacrima syndrome
ALADIN	ALacrima-Achalasia-aDrenal Insufficiency Neurologic disorder
amp	ampicillin
BCA	bicinchoninic acid
BSA	bovin serum albumin
Cas9-RNP	Cas9-ribonucleoprotein
CRISPR	clustered regularly interspaced short palindromic repeats
crRNA	CRISPR RNA
DMSO	dimethyl sulfoxide
DNTP	deoxynucleoside triphosphate
DSB	double-stranded break
EDTA	ethylene diamine tetraacetic acid
FTH1	ferritin heavy chain
gRNA	guide RNA
HD	Huntington's disease
HDR	homology-directed repair
hESC	human embryonic stem cells
hiPSC	human induced pluripotent stem cells
HR	homologous recombination
HTT	huntingtin
kan	kanamycin
KSR	knock-out serum replacement
NHEJ	non-homologous end joining
NII	neuronal intranuclear inclusions
NPC	nuclear pore complex
OTS	off-target site
PAM	protospacer adjacent motif
PBS	phosphate buffered saline
pheS	phenylalanyl tRNA synthetase
polyQ	poly-glutamine
puro	puromycin
Ri	ROCK inhibitor
ROS	reactive oxygen species
RT	room temperature
SDS	sodium dodecyl sulphate
ssODN	single-stranded oligodeoxynucleotide
TALEN	transcription activator-like effector nucleases
TBS	tris buffered saline
TBS-T	TBS-Tween
tet	tetracycline
TE	tris-EDTA buffer
TENS	tris-EDTA-NaCl-SDS buffer
tracrRNA	trans-activating crRNA
WD-repeat	tryptophan-aspartic acid repeat
zeo	zeocin
ZFN	zinc finger nuclease

Abstract

Genetic neurodegenerative disorders are inherited diseases of the brain and nervous system, many of which are poorly understood. Human induced pluripotent stem (hiPS) cells, upon introduction of disease relevant mutations by genome editing, can be used to model diseases, gain insights into the pathophysiology of such diseases, and develop therapies. In this study, CRISPR-Cas9 system was used to edit the hiPS cell genome to generate models for two diseases – Achalasia-Addisonianism-Alacrima syndrome (AAAS) and Huntington's disease.

AAAS is a progressive neurodegenerative disorder mainly characterized by oesophageal muscle disorders, adrenal insufficiency, and tear production failure. Multiple mutations in the *AAAS* gene encoding the ubiquitously expressed ALADIN (ALacrima-Achalasia-aDrenal Insufficiency Neurologic) protein are responsible for this autosomal recessive disorder. Using two different strategies, I generated human iPS cells homozygous for a common allele of *AAAS*, a C>A point mutation in exon 1 that creates a novel splice-donor site, and for comparison, a biallelic knockout of *AAAS* exon 2. The biallelic knockout relied on replacing one allele with a puromycin selection cassette by homologous recombination (HR) and simultaneously damaging the other allele by non-homologous end-joining (NHEJ). Sequencing the non-targeted allele demonstrated frameshift indels with a biallelic targeting efficiency of 3%. The homozygous point mutation was generated by homology-directed repair (HDR) with a single-stranded oligonucleotide template. Clones homozygous for the point mutation were generated in two rounds of genome editing. A complete loss of ALADIN protein in undifferentiated human iPS cells was observed in both knockout and point mutant cells, suggesting that the disease is caused by null mutations in the *AAAS* gene. These models lay the foundation of future detailed phenotype analysis and understanding of disease manifestations by differentiating them to the disease-relevant cell types, particularly cortical neurons and adrenocortical cells.

Huntington's disease is a progressive neurodegenerative genetic disorder that mainly causes motor, behavioral, and cognitive abnormalities in an individual. This disease occurs due to an autosomal dominant expansion of CAG-repeats in the huntingtin (HTT) gene. I attempted to insert 67 Q-repeats into HTT by Cas9-assisted homologous recombination (HR) with a donor plasmid containing the extended repeat fragment. From a screen of 192 unselected clones, I was unable to recover any clones containing the expanded repeats. Presumably the rate of homologous recombination was too low in unselected clones and future experiments will require drug selection to ensure retention of the expanded Q-repeats.

Chapter 1
Introduction

Neurodegenerative disorders:

Neurodegenerative disorders are one of the current major global concerns representing a massive worldwide health burden and costing healthcare systems billions every year. According to the WHO (World Health Organization; 2016), these disorders affect up to one billion lives worldwide every year killing an estimated 6.8 million people. These are a range of nervous system disorders characterized by progressive loss of structure and function of the neurons in the brain and spinal cord leading to neuronal damage and death. The generation and cell death of neurons are critical in brain development and maintenance, hence alterations in these processes are often the reason behind neurodegenerative disorders (Winner et al. 2011). The most common and notable of the more than 600 such disorders are Alzheimer's disease, Parkinson's disease, Huntington's disease, amyotrophic lateral sclerosis, and motor neuron diseases (European Commission 2016).

Because most of these disorders are late onset, manifesting after 50-60 years of age (WHO 2016), they are becoming increasingly prevalent as average life expectancy increases (Brown et al. 2005). The late onset is because many regions of the brain are vulnerable to age-related neurodegenerative changes (Small 2003). Besides late onset, early onset of some of such disorders are also seen in children and adolescents. Almost all these disorders are genetic, in that either dominant or recessive gene mutations can lead to the gain or loss of function of associated proteins resulting in abnormal cellular functioning. Some of the vital genes which play roles in neurodegeneration include α-synuclein, presenilin-1, tau, SOD1, and huntingtin (Winner et al. 2011). Mutations in these genes often result in aggregation of toxic proteins, misfolding, mislocalization, degradation of vital proteins, and oxidative stress due to mitochondrial dysfunction (Mattson 2000; Thompson 2008). Most of these aberrant mechanisms eventually result in cellular abnormalities that trigger apoptosis of more neurons in a cascade-like manner (Mattson 2000) resulting in progressive neurodegeneration.

The pathogenesis of many of these diseases remain unclear (Brown et al. 2005). To date, researchers have mostly relied on the conventional clinicopathological approach of matching particular medical manifestations to a pattern of pathology in the brain (Ahmed et al. 2016). The cellular abnormalities and the disease mechanism pathways involved often overlap with each other which also makes it difficult to understand the progression of a disease. The aggregation of related proteins, which is often the cause of neuronal death in a certain neurodegenerative disorder, might also be the reason for other neurodegenerative disorders in

other individuals (Bertram & Tanzi 2005; Nieoullon 2011). Oxidative stress, an overproduction of reactive oxygen species (ROS) in neurons, can also play role in neuronal death alongside protein aggregation; for example, in Alzheimer's and Parkinson's disease (Barnham et al. 2004). This could partly be due to numerous cellular proteins interacting with the mitochondria in its functions and also because essential genes like, tau, SOD1, and huntingtin have all been found within mitochondria (Lin & Beal 2006). Thus, these diverse mechanisms along with interlinked pathways between multiple diseases make it difficult to differentiate one disease pathophysiology from another which is why the numbers and classification of neurodegenerative diseases still remains a challenge (Ahmed 2016).

Although neurodegenerative disorders affect a number of cells and processes in the body, the major manifestations are observed as motor, cognitive, and psychiatric abnormalities (Bertram & Tanzi 2005). Motor abnormalities are often seen as movement disorders, muscle wasting and coordination problems, gait, and balance abnormalities, while cognitive disturbance happens with variable extents of dementia, impaired reasoning, and learning difficulties. A range of psychiatric problems are also observed like depression, schizophrenia, and behavioural changes. For these disorders, there are certain drugs that can mitigate the symptoms to some extent, although they cannot absolutely be considered an effective and permanent means of treatment (Butler & Zeman 2005; Nieoullon 2011).

While there have been a few cases of family history of neurodegenerative disorders, they are largely polygenic with variants in multiple genes resulting in the phenotype (Bertram & Tanzi 2005). Most neurodegenerative disorders occur due to sporadic mutations with no family history, the causative factors of which are still unknown (Nieoullon 2011). These sporadic cases are thought to be multifactorial where certain gene mutations are susceptible to different environmental factors. Such factors often increase the risk of the diseases, but cannot actually be considered a causative agent (Bertram & Tanzi 2005). Examples include exposure to compounds such as pesticides or heavy metals, which for example, are thought to play roles in Parkinson's disease (Nieoullon 2011). Moreover, other factors like intensive chronic stress and anxiety are also thought to impact cognitive functions in many neurodegenerative disorders (Gianaros et al. 2007). However, epidemiological evidence for any association between the environment and the prevalence of such disorders is not conclusive (Brown et al. 2005), thus more research is needed. Two such genetic neurodegenerative diseases include AAAS (Achalasia-Addisonianism-Alacrima syndrome) and Huntington's disease (HD) whose

pathophysiologies are yet to be understood well enough to develop therapies and drugs for improving patient health and lifestyle.

AAAS (Achalasia-Addisonianism-Alacrima syndrome)

AAAS, Triple-A syndrome or Allgrove syndrome (MIM# 231550) is a very rare autosomal recessive disorder which was first described by Allgrove and colleagues in 1978 in two unrelated pairs of siblings. The prevalence of this disorder is unknown but has had only about 100 reported cases from its initial description in 1978 until 2013 (Bizzarri et al. 2013). AAAS is mainly characterized by a triad of features: achalasia – lower oesophageal sphincter muscle disorder, Addisonianism – adrenal insufficiency, and alacrima – tear production failure (Handschug et al. 2001), and is often associated with progressive neurodegeneration and autonomic neuropathy (Tullio-Pelet et al. 2000).

Manifestations of AAAS:

AAAS mostly exhibits itself within the first decade of life, although few cases with adulthood onset have been reported (Vishnu et al. 2014). Alacrima is the earliest manifestation of AAAS in most patients demonstrating itself as early as birth (Mazzone et al. 2013). Alacrima is the failure or reduction in tear production which is often exhibited as "crying without tears" in patients (Bizzarri et al. 2013). As suggested by lacrimal gland biopsies of patients, it happens due to small lacrimal glands and depletion of secretory acinar cells of the lacrimal gland (Mullaney et al. 1998). The second manifestation of AAAS is achalasia, which appears with advancing age or in infancy in most patients (Mazzone et al. 2013). In achalasia, the smooth muscles of the lower oesophageal sphincter fail to relax and along with absent peristalsis, makes it difficult for the patient to swallow food delaying passage of food into the stomach (Vishnu et al. 2014). Consequently, the patients suffer from problems like recurrent vomiting, abdominal pains, and failure to thrive (Bizzarri et al. 2013). The third major AAAS attribute is Addisonianism or adrenocorticotrophic hormone (ACTH) resistant adrenal insufficiency which generally arises later than the other two symptoms, however mostly within the first decade of life (Mazzone et al. 2013) and sometimes later in life. Due to ACTH resistance, ACTH levels rise extremely high causing progressive skin hyperpigmentation, and cortisol deficiency causes fatigue, loss of appetite, hypoglycemia, and other associated symptoms, and might even lead to sudden death (Bizzarri et al. 2013; Huebner et al. 2004).

Unlike the classical triad of conditions described above, AAAS patients, although not always, also suffer from central, peripheral, and autonomic nervous system abnormalities (Bizzarri et al. 2013). It is also sometimes associated with other heterogeneous clinical defects (Mazzone et al. 2013). *AAAS* gene is abundantly expressed in neurons of the cerebral cortex, cerebellum, hippocampus, brainstem, and spinal cord which might explain why *AAAS* mutations lead to 60% patients suffering from progressive neurological abnormalities (Grant et al. 1993; Huebner et al. 1999) like cognitive, motor, and sensory dysfunctions. This gene is also expressed in cells of the peripheral nervous system which leads to approximately 39% of patients suffering muscle hypotonia, weakness, muscle atrophy, and associated symptoms in patients. Besides, AAAS also causes around 30% of patients to suffer from autonomic neuropathy which is characterized by impaired cardiovascular reflexes, cardiac dysrhythmias, hypoglycemia, hyperreflexia, and others (Prpic et al. 2003; Storr et al. 2005; Mazzone et al. 2013). All the neurological symptoms, along with associated clinical features are progressive, suggesting a degenerative process (Prasad et al. 2014).

Currently there is no permanent cure for AAAS as the exact mechanisms of the progression of this disorder is not known. However, there are some treatments and medication that can alleviate the symptoms temporarily. For instance, for Addisonianism, the only treatment is the replacement of glucocorticoids, whereas for achalasia, a surgical correction like pneumatic dilation is the best management; and for alacrima, regular application of topical lubricants could improve the symptom (Sarathi & Shah 2010).

Genetics of *AAAS*:

The *AAAS* gene is located on chromosome 12q13 in the critical region between the markers KRT8 and D12S1651 (Handschug et al. 2001). The 1840bp transcript consisting of 16 exons (Tullio-Pelet et al. 2000) encodes the ALADIN (ALacrima-Achalasia-aDrenal Insufficiency Neurologic disorder) protein which comprises of 546 amino acids and has a molecular weight of 59.6kDa. However, alternative splicing of the transcript (Figure 1.1) produces several smaller isoforms of the protein (Figure 1.2). There are several other non-coding transcripts as well which do not express any functional protein.

Show/hide columns (1 hidden) Filter

Name	Transcript ID	bp	Protein	Biotype	CCDS	UniProt	RefSeq	Flags
AAAS-001	ENST00000209873.8	1840	546aa	Protein coding	CCDS8856	Q9NRG9	NM_015665 NP_056480	TSL:1 GENCODE basic APPRIS P1
AAAS-002	ENST00000394384.7	1703	513aa	Protein coding	CCDS53797	Q9NRG9	NM_001173466 NP_001166937	TSL:1 GENCODE basic
AAAS-004	ENST00000550286.5	1652	422aa	Protein coding	-	F8VZ44	-	TSL:5 GENCODE basic
AAAS-020	ENST00000548931.5	1075	358aa	Protein coding	-	H3BU82	-	CDS 5' and 3' incomplete TSL:5
AAAS-012	ENST00000547757.1	949	277aa	Protein coding	-	F8VUB6	-	CDS 3' incomplete TSL:2
AAAS-021	ENST00000549983.5	860	No protein	Processed transcript	-	-	-	TSL:3
AAAS-010	ENST00000547761.6	827	No protein	Processed transcript	-	-	-	TSL:5
AAAS-011	ENST00000549821.5	596	No protein	Processed transcript	-	-	-	TSL:3
AAAS-009	ENST00000549450.5	581	No protein	Processed transcript	-	-	-	TSL:5
AAAS-016	ENST00000551724.5	552	No protein	Processed transcript	-	-	-	TSL:4
AAAS-015	ENST00000548258.5	550	No protein	Processed transcript	-	-	-	TSL:5
AAAS-003	ENST00000552876.5	2016	No protein	Retained intron	-	-	-	TSL:2
AAAS-018	ENST00000547238.5	1154	No protein	Retained intron	-	-	-	TSL:2
AAAS-019	ENST00000550033.5	900	No protein	Retained intron	-	-	-	TSL:3
AAAS-013	ENST00000546393.6	832	No protein	Retained intron	-	-	-	TSL:5
AAAS-022	ENST00000547520.5	769	No protein	Retained intron	-	-	-	TSL:3
AAAS-005	ENST00000546562.5	627	No protein	Retained intron	-	-	-	TSL:2
AAAS-023	ENST00000552161.5	585	No protein	Retained intron	-	-	-	TSL:2
AAAS-014	ENST00000546572.1	582	No protein	Retained intron	-	-	-	TSL:2
AAAS-017	ENST00000548880.1	577	No protein	Retained intron	-	-	-	TSL:2

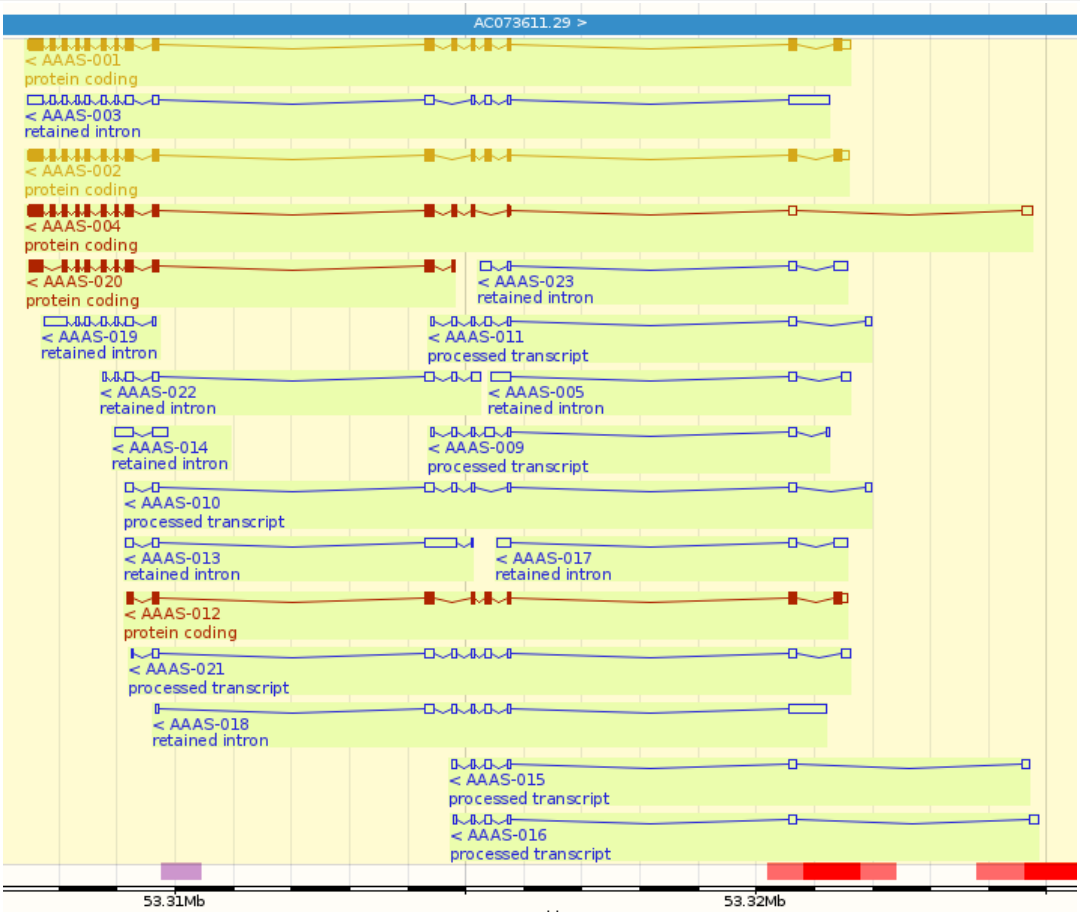


Figure 1.1: Various coding and non-coding transcripts of the *AAAS* gene encoding different ALADIN isoforms.

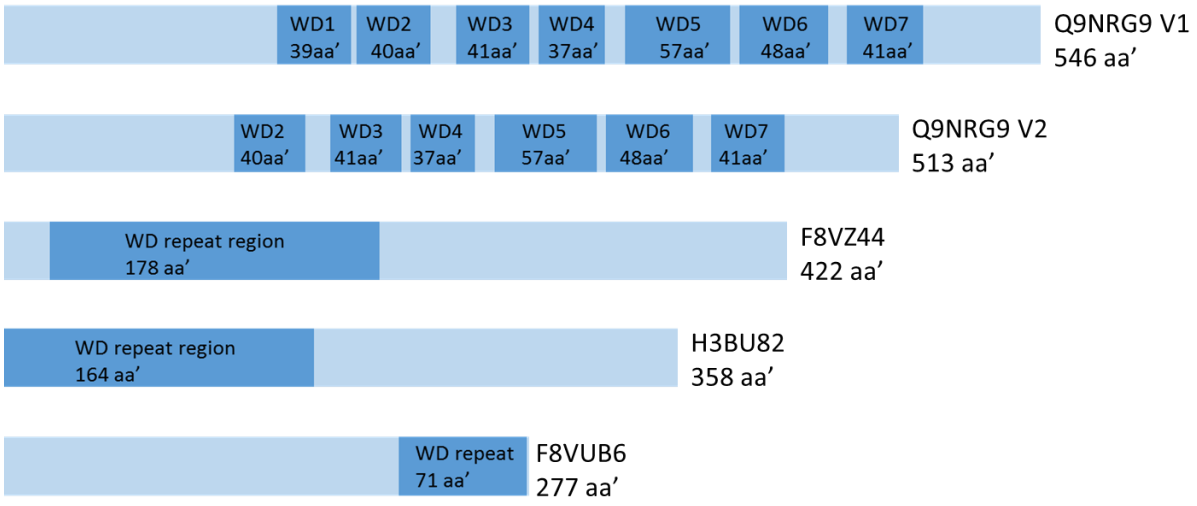


Figure 1.2: The known isoforms of ALADIN protein showing WD repeat domains (Uniprot 2016)

***AAAS* mutations**:

AAAS is caused by autosomal recessive or compound heterozygous mutations in the *AAAS* gene (Prpic et al. 2003). Over 60 different mutations scattered throughout the *AAAS* gene have been described to be responsible for the disease (Prasad et al. 2014), which predominantly affect cells of the nervous system and adrenal cortex. These mutations include missense, nonsense, deletion, frameshift, and splicing mutations (Handschug et al. 2001; Krumbholz et al. 2006), most of which result in a truncated ALADIN protein suggesting loss of function (Handschug et al. 2001). A few mutation hotspots have been identified – a missense mutation: c.787T>C in exon 8 (p.S263P) and two splice-site mutations: 43C>A in exon 1 (p.G14fs) and c.1331+1G>A in exon 14 (Krumbholz et al. 2006; Dumic et al. 2012; Kallabi et al. 2016).

Loss of function in ALADIN could either be partial or complete depending on the type of the mutation. Compound heterozygotes, for instance, with a frameshift mutation in one allele and missense in the other, lead to partial loss of function of the protein, whereas nonsense, deletion, and splice-site mutations result in complete loss of function (Houlden et al. 2002; Tullio-Pelet et al. 2000). The variability in the mutations and their locations in the gene probably explain the considerable variation in the severity of the disease among the different families studied so far. The severity of the disease is also correlated to the time the disease symptoms are manifested: it has been observed that the later the age of onset is, the less severe is the phenotype. (Houlden et al. 2002). However, no genotype-phenotype correlation has yet been established for AAAS. The phenotype has been found to be variable among patients in

the same kindred having similar mutations. Also, patients with similar phenotypes have been found to have different *AAAS* mutations or no mutation at all (Houlden et al. 2002; Prpic et al. 2003). This suggests that there could be other interacting genes or environmental factors influencing the phenotype of this disease. Besides, AAAS has not been found to have any case of family history and its incidence is unknown. However, it has been found in highly inbred kindreds or in consanguineous families. (Houlden et al. 2002).

ALADIN

The ALADIN protein belongs to the WD-repeat (tryptophan-aspartic acid repeat) containing family of regulatory proteins. WD repeat proteins are found only in eukaryotes, and there are over 136 of them described in humans (Li & Roberts 2001). These proteins are defined by the presence of a highly conserved core of around 40-60 amino acids and ending with 4-16 WD repeat domains (Neer et al. 1994). These repeat domains fold into a circularized β-propeller structure and serve as a rigid scaffold or platform for reversible protein-protein interactions and protein complex assemblies (Smith 2008). Hence, these properties enable WD repeat proteins to play roles in a range of cellular functions like signal transduction, vesicular trafficking, transcriptional regulation, cell cycle regulation, and cell-fate determination (Neer et al. 1994; Li & Roberts 2001).

The full length ALADIN protein consists of four WD repeats and is expressed ubiquitously in human tissues, and particularly abundantly in the adrenal gland, gastrointestinal structures, pituitary gland, and cerebellum (Handschug et al. 2001; Bizzarri et al. 2013) which probably explains the main AAAS manifestations. ALADIN was the first nuclear pore complex (NPC) protein described to be associated with any genetic neurodegenerative disorder and is also the only one associated with genetic adrenal disease (Cronshaw & Matunis 2003). Although the exact function of this protein is unknown, it is part of the NPC – the large multiprotein assemblies on the nuclear envelope which play roles in trafficking molecules between the cytoplasm and nucleoplasm. ALADIN is one of the approximately 30 different nucleoporins, proteins on the NPC, which function in nucleocytoplasmic transport (Cronshaw et al. 2002). It is anchored to the NPC via the membrane integrated nucleoporin, NDC1, which also plays an important role in ALADIN function. This is because a mutation in the NDC1 gene was found to affect the interaction with ALADIN, resulting in the disease pathogenesis (Kind et al. 2009).

Disease-causing mutations in *AAAS* gene often leads to the mislocalization of protein predominantly in the cytoplasm (Cronshaw & Matunis 2003) leading to functional abnormalities of the nuclear pore complex (Bizzarri et al. 2013). This suggests that correct localization of ALADIN in NPC is necessary for its function (Cronshaw & Matunis 2003). Cells derived from AAAS patients have also demonstrated functional abnormalities but no morphological features, suggesting that *AAAS* mutations affect the functions of the nuclear pore complex rather than its structure (Brooks et al. 2004). However, exceptions include the splice-site mutation Q15K (43C>A; exon 1) in which ALADIN remains in the NPC (Krumbholz et al. 2006).

ALADIN is predominantly expressed in neurons and adrenal glands and is postulated to function in intracellular trafficking, cytoskeleton assembly, transmembrane signaling, DNA replication, and transcription (Handschug et al. 2001; Cronshaw & Matunis 2003; Shen et al. 2010), which explains why these are the main tissue types exhibiting the disease manifestations. ALADIN is also involved in protein-protein interactions and in directing proteins to peroxisomes which could play roles in adrenal function and neuronal migrations (Bizzarri et al. 2013). Thus, this wide functional diversity of ALADIN could explain why mutations in the *AAAS* gene lead to a myriad of other clinical features besides the classical triad of AAAS conditions (Houlden et al. 2002).

A critical role of ALADIN is thought to be in controlling the oxidative stress in cells – an imbalance between oxidant and antioxidant levels that occurs due to increased toxic reactive oxygen species (ROS). ROS is mostly generated due to reaction between molecular oxygen and redox-active metals – calcium and iron (Barnham et al. 2004). ROS is often responsible for having deleterious effects on proteins, lipids, and nucleic acids, subsequently leading to cellular dysfunction and death. Mitochondria, because of electron leakage in respiration, are responsible for majority of ROS production in the body, and the adrenal cortex controls any risk of oxidative stress (Prasad et al. 2014). In the adrenal cortex, during steroidogenesis of glucocorticoids, expression levels of cytochrome P450 enzymes rise which often leads to electron leakage by cytochrome P450 oxidase. This leakage leads to increase in ROS levels in the adrenocortical cells. An imbalance in redox homeostasis in *AAAS* gene-mutated cells was demonstrated by a reduction in the ratio of reduced to oxidized glutathione (Prasad et al. 2013). Treatment with antioxidants like N-acetylcysteine replaced stores of reduced glutathione, improving the cell viability of knockdown cells, further validating the effect of ROS in disease pathogenesis. Since the adrenal cortex needs high levels of antioxidants to combat this

oxidative imbalance, defective nucleocytoplasmic transport of antioxidants and DNA repair proteins, as in AAAS patients, lead to a redox imbalance affecting steroidogenesis, but how this happens is still unclear (Prasad et al. 2014).

Besides being expressed in the adrenal cortex, ALADIN is also abundantly expressed in the neural cells of the central, peripheral, and autonomic nervous system. Although the exact mechanism by which neurodegeneration occurs remains unclear (Kind et al. 2010), it is thought to be the results of oxidative stress in neuronal cells, which was demonstrated by the hypersensitivity of neuroblastoma cells to oxidative stress and consequently significant cell death (Prasad et al. 2013). Neurons are more energy-expensive than the other cells of the body. Mitochondrial activity is comparatively higher in neuronal cells in order to maintain ionic gradients across cells membranes (Kind et al. 2010). However, as described earlier, when *AAAS* gene is mutated, increased activity of the respiratory chain may lead to higher ROS levels in the neurons resulting in oxidative stress ensuing neuronal death. A hypothesis of the neurodegenerative mechanism of AAAS is the Stress-induced Premature Senescence (SIPS) which is compatible with the observed pathogenesis in AAAS patients. When there are respiratory chain defects in the mitochondria that cause an increase in the metabolic activity of the mitochondria, the levels of the ROS-detoxifying enzymes superoxide dismutase (SOD), catalase, and glutathione reductase increases, which results in a rise in the mitochondrial and cellular ROS levels. This leads to an increased cellular oxidative stress, eventually leading to neuronal senescence with subsequent neurodegeneration (Kind et al. 2010).

ALADIN interacts with FTH1 (ferritin heavy chain) and is hypothesised to have a role in nucleocytoplasmic transport of FTH1 which might be important in protecting cells against oxidative imbalance (Storr et al. 2009). Besides its recognised role in iron storage, FTH1 has been demonstrated to be transcriptionally activated in an oxidative stress environment in cells (Huang et al. 2013) and is known to act as an antioxidant protecting cells against oxidative damage (Thompson et al. 2002). Absence of nuclear FTH1 in AAAS patient dermal fibroblasts and interaction of ALADIN with FTH1 suggested ALADIN could be necessary in nuclear uptake of FTH1. Co-transfection of AAAS patient neuronal cells with *AAAS* and *FTH1* genes enhanced the nuclear translocation of FTH1 by ALADIN protein reducing cellular apoptosis (Storr et al. 2009). Thus, it can be hypothesised that ALADIN could play a role in protecting the cells from iron-mediated ROS formation, hence oxidative stress. The adrenal cortex and neural tissue, having a highly oxidative cellular environment, may explain the vulnerability of these tissues in the absence of functional ALADIN (Prasad et al. 2013).

Not only for FTH1 import, ALADIN might also be involved in nuclear import of other DNA repair proteins, aprataxin and DNA ligase I, which are particularly important in DNA damage protection and repair. Aprataxin and DNA ligase I are involved in DNA single-strand break repair, so when a mutated ALADIN is mislocalized in the cytoplasm, the DNA repair proteins cannot enter the nucleus which eventually causes cell senescence. This was validated by transfecting AAAS patient dermal fibroblasts with the DNA repair proteins fused with nuclear localization signals (NLS), which restored the nuclear import of these proteins, reducing cell death induced by oxidative stress (Kiriyama et al. 2008). Another study (Hirano et al. 2006) of mutant ALADIN (I482S) in patient fibroblasts also suggested impaired nuclear import of aprataxin and DNA ligase I, which was restored by transfecting wild-type ALADIN into the cells. However, this restoration was not to wild-type levels which suggested that nuclear import of other proteins might also be necessary.

Our current knowledge of AAAS is limited. The wide functional diversity of ALADIN and the variable manifestations of the disease often confound its diagnosis and have limited our understanding of the disease mechanism so far. Lack of genotype-phenotype correlation has further acted as a barrier in the understanding of this disease. Moreover, mouse models do not recapitulate the human AAAS abnormalities suggesting functional redundancy in mouse and probable species-specific ALADIN functions. On the other hand, patient dermal fibroblasts have limitations as they are not representative of the cell types affected by the disease. The current study is to generate human iPS cell models and to gain a proper understanding of the disease mechanism pathway in the appropriate cell types, so that therapies and drugs can be developed. Cellular differentiation of the disease models and further detailed phenotypic studies are needed for a better understanding of this ubiquitous developmental gene and the disease pathophysiology.

Huntington's disease (HD)

Huntington's disease (MIM# 143100) is a progressive genetic neurodegenerative disorder that mainly causes motor, behavioural, and cognitive abnormalities in an individual. This autosomal dominant trinucleotide repeat disorder affects approximately 10 to 15 per 100,000 individuals of the European ancestry (Dayalu & Albin 2015), but is less common among other populations around the world. HD is most commonly an adult onset disorder displaying the symptoms at a mean age of 35-44 years. However, a less common childhood onset form, accounting for about 7% of the cases, appears before the age of 21 years (Nance & Myers 2001). This juvenile form

is more severe and progresses quicker than the adult form. In general, the median survival time for HD patients is 15 to 18 years after the onset of symptoms, with the average age at death being 54-55 years (Harper 2005).

Manifestations of HD

The symptoms of HD were first described by George Huntington in 1872 (Van Dellen et al. 2005). HD patients are diagnosed on the basis of family history, characteristic clinical symptoms, and genetic screening tests. Early symptoms of HD can mostly be classified under motor abnormalities including involuntary jerking movements in the body for which the disorder is also called Huntington's chorea. HD also affects muscle coordination, movement, and balance of an individual (Marshall et al. 2007). However, cognitive and behavioural problems may occur before movement problems, and include cognitive impairment, irritability, paranoia, depression, and other associated symptoms. Late onset symptoms appearing in adults often include dementia, memory loss, impaired reasoning and mental planning, personality changes, and sleep disorders (Bourne et al. 2006; Marshall et al. 2007; Rosenblatt 2007; Morton 2013). On the other hand, childhood onset often leads to clumsiness, seizures, slow movement, slurred speech, and cognitive defects in children and adolescents (Nance & Myers 2001). Besides, HD patients might also suffer from a defective metabolic system (Duan et al. 2014). All these symptoms mainly arise due to neurons of cerebral cortex, striatum, basal ganglia, and hippocampus being affected, the areas of the brain which are involved in controlling movement, cognition, and behaviour of an individual (Cowan & Raymond 2006). Brain imaging studies, like MRI and CT, and post-mortem brain analysis in pre-manifest HD patients showed small striatal volumes and atrophy suggestive of striatal neurodegeneration (Biglan et al. 2009). Caudate nucleus involved in cognition, and putamen involved in movement regulation are both parts of the dorsal striatum. Hence, striatal neurodegeneration explains learning difficulties along with other cognitive abnormalities (Mitchell et al. 1999), as well as chorea and movement problems in HD patients.

HD has no effective permanent treatment. However, some temporary management measures to mitigate the symptoms include dopamine blockers which are used to reduce the abnormal behaviours and movements. Other pharmacologic therapies include neuroleptics, amantadine and tetrabenazine which are used to control jerks and extra involuntary movements (Frank 2014). Psychotropic or antiepileptic drugs are often used to alleviate the psychiatric disturbances like, depression, psychotic symptoms, and outbursts of aggression.

Genetics of HTT

The HTT gene is the only known gene associated with HD. It is located on chromosome 4p16.3 (The Huntington's Disease Collaborative Research Group 1993). The approximately 13.5kb transcript consists of 67 coding exons encoding huntingtin protein comprising 3142 amino acids with a molecular weight of around 348kDa. The HTT gene has 13 splice variants of the transcript, only two of which are protein coding. The part of the gene involved in disease implications is the CAG (cytosine-adenine-guanine) repeats in the first exon that encode the polyglutamine (polyQ) tract.

Genotype-phenotype correlation

The only known pathogenic mutant of HTT gene responsible for the disease is the autosomal dominant expansion of the CAG repeat length in exon 1 of the gene, first described in 1993 by the Huntington's Disease Collaborative Research Group. Usually there are less than 27 CAG repeats in the wild-type gene of healthy individuals, whereas, a mutated HTT gene can contain as many as 120 CAG repeats (Yoon et al. 2006). Individuals having 28-35 CAG repeats are considered in the intermediate range because although they are healthy, they are at a risk of having children who will develop the disorder. The risk is due to instability in the repeat length during gametogenesis (Semaka et al. 2006). Reduced penetrance is observed in individuals with 36-40 repeats (Semaka et al. 2006) when they may or may not develop the disorder, while patients with more than 40 repeats demonstrate full penetrance exhibiting the classical disease manifestations (Langbehn et al. 2004). In HD and other trinucleotide repeat disorders, the number of repeats often increase with every generation, in a phenomenon known as anticipation, which leads to an earlier onset and increase in the severity of symptoms with each successive generation. Thus, the number of repeats is inversely correlated with the age of onset of the disorder and directly correlated with its severity (Langbehn et al. 2004). With an increased CAG repeat length, the rate of progression of the clinical manifestations is also higher. Typically, 40-50 CAG repeats result in a late or adult onset, while more than 60 repeats tend to give rise to the more severe juvenile form. Hence, it can be inferred that there is a strong genotype-phenotype correlation in HD (Duyao et al. 1993; Langbehn et al. 2004).

Gene-environment interactions

Besides the gene itself, some studies using mouse models suggest that environmental factors may also influence the onset and progression of HD. It is thought that environmental

modulators may influence the interaction between the abnormal huntingtin protein and other molecular mediators. Environmental factors may also affect the downstream processes involving synaptic and intraneuronal signal processing resulting in neuronal dysfunction and pathology (Van Dellen et al. 2005). The influence of environment was also demonstrated in another study with transgenic mice through environmental enrichment. Enrichment, which involves providing mice with various complex and stimulating objects in their environment, delayed the onset and progression of HD in transgenic mouse to a considerable extent. The HTT protein levels in the cerebral cortex of the enriched mice were also rescued (Spires et al. 2004). This suggested that environmental enrichment could enhance cognitive and motor stimulation, thus delaying the symptoms. However, how these mouse studies can be extrapolated to humans remains a question.

Huntingtin

Huntingtin is ubiquitously expressed with highest levels in the brain and to a lesser extent, in the heart, liver, and lungs (Walker 2007). In the brain, HTT is preferentially expressed in the neurons of the cerebral cortex, striatum, and hippocampus, areas of the brain which are involved in controlling movement, cognition, and behaviour of an individual (Cowan & Raymond 2006). Thus, impairment of these abilities in a HD patient is explained by an abnormal functioning of HTT in the neurons of those brain areas. HTT contains WW domains, involved in protein-protein interactions and binding, and caspase cleavage sites, for cleavage by caspase – the enzyme that is involved in apoptosis (Young 2003). The most vital part of the protein implicated in the disease is the polyQ tract at its N-terminal region (Li & Li 2004). Thus, a mutated huntingtin with a longer polyQ tract is cleaved more efficiently which then generates toxic NH_2-terminal fragments, hence neurodegeneration in HD (Cattaneo 2003). A mutant HTT gene with more than 35 CAG repeats is the reason behind a malformed protein with expanded polyQ repeats. These abnormally long proteins are cleaved into smaller fragments by proteases which tend to aggregate in the nerve cells, forming toxic insoluble clumps called neuronal intranuclear inclusions (NII). The role of NIIs is controversial as two contradicting hypotheses exist. In some studies, HTT has been suggested toxic, while other studies have supported that they could be beneficial.

Studies advocating NIIs to be toxic argue that these misfolded protein aggregates disrupt intracellular homeostasis by inhibiting the proteasome. This inhibition subsequently induces neurotoxicity and hinders the multiple intracellular pathways leading to neuronal

dysfunction. Proteasome inhibition by NIIs is also linked to the activation of the apoptotic machinery which leads to neuronal death. Together with the NIIs, there are post-translational modifications that also affect the structure of HTT which can cause gain of a new toxic function or loss of the beneficial properties of the protein (Bence et al. 2001; Rangone et al. 2004), eventually resulting in neuronal dysfunction. Mouse models have also displayed accumulation of the toxic HTT fragments in both the nucleus and the neuronal processes, suggesting that these subcellular sites of neurons are the hotspots for neuropathology of the disease (Li & Li 2004).

On the other hand, insoluble NIIs have also been argued to be non-pathogenic (Slow et al. 2006). One study on mouse models (Hodgson et al. 1999) showed that progressive electrophysiological abnormalities, responsible for neuronal loss, were present before the formation of aggregates. This finding suggested NIIs were not responsible for HD pathogenesis. In another study, it was demonstrated that NIIs can also exist without any disease pathology (Slow et al. 2006) suggesting there is no correlation between NII burden and neurodegeneration. A further study came up with a striking outcome stating NIIs were, in fact, a coping response which extended neuronal life (Arraste & Finkbeiner 2012). They stated that NIIs, functioning together with the autophagy systems in cells, might help to reduce the levels of diffused intracellular mutant HTT protein. This reduction alleviated the risk of neuronal death, thus extending neuronal life eventually reducing the effects of a mutant HTT.

Huntingtin is found in multiple cellular compartments and although its exact function is unknown, HTT is postulated to be involved in neuronal signalling, transport, protein binding, protein-protein interactions, transcriptional regulations (Van Dellen et al. 2005; Cowan & Raymond 2006). It is found in cell cytoplasm, within neurites, and at synapses and plays roles in interaction and binding with numerous proteins. Such proteins are vital in intracellular transport and endocytosis, which suggests that huntingtin could also be involved in these processes (Li et al. 1998). A mutant HTT protein in neuronal processes have been found to affect axonal signalling and transport, suggesting reasons behind selective neurodegeneration in HD (Li & Li 2004). Selective neurodegeneration is also thought to be a result of differential toxicity of mutant HTT protein in different neuronal subtypes (Arraste & Finkbeiner 2012).

HTT is also found in nucleus where it seems to play roles in transcriptional regulation. Mutated HTT affects the gene expression either by intranuclear aggregate formation or by sequestration of important transcription factors (De Rooij et al. 1996). Gene expression in cells

is regulated in an orchestrated manner, involving the elaborate interactions among a large number of proteins and transcription factors. HTT has been suggested to interact with many transcription factors, and a mutant HTT could result in impaired interactions, eventually disrupting the transcriptional pathways (Landles & Bates 2004). This transcriptional impairment might be the cause of neuronal dysfunction and death particularly if those genes are critical in neuronal survival. Having said that, the exact pathways or interactions influencing the disease pathogenesis remain unclear.

Due to a mutant huntingtin, calcium handling and energy metabolism in the mitochondria is impaired which activates proteases, suggesting huntingtin is involved in regulation of mitochondrial homeostasis. In HD brain and muscle cells, significant reductions in the activities of the mitochondrial respiratory chain enzymes has been measured (Turner & Schapira 2010), suggesting a role for HTT in energy metabolism. Accumulation of lactate in the cortex and basal ganglia of patients and impaired ATP production in striatal cells also support the postulation of mitochondrial impairment in HD (Lin & Beal 2006). Mitochondrial dysfunction eventually activates caspases, enzymes involved in apoptosis, which result in cell death (Rangone et al. 2004). Thus, apoptosis affects not only neurons, but also cells in other parts of the body, such as the heart, liver, and lungs, as mentioned earlier.

Besides playing indirect roles in apoptosis, as described above, HTT may be an integral anti-apoptotic protein in striatal cells. When transfected into cultured striatal neurons, mutant HTT induced neurodegeneration in an apoptotic mechanism, and antiapoptotic neurotrophic factors rescued the neurons against apoptosis induced by mutant HTT. Preventing the mutant HTT from localizing in the nucleus prevented it from forming intranuclear inclusions and inducing neurodegeneration (Saudou et al. 1998). Conversely, over-expression of wild-type HTT rescued the cells from death induced by serum deprivation. Hence, it can be implied that absence of HTT leads to apoptosis of striatal neurons explaining the disease manifestations. This anti-apoptotic function of HTT has also been observed to be essential in embryonic development (Van Dellen et al. 2005) in that mice lacking wild-type HTT demonstrated extensive cell death in the mouse embryonic ectoderm (Zeitlin et al. 1995).

The pathophysiology of Huntington's disease is still controversial and several possible molecular mechanisms have been described to be responsible for the disease. Although HTT is widely expressed in the brain, why the mutant form causes the selective neurodegeneration in the striatum, basal ganglia, and cerebral cortex remains unclear. Human stem cell models

could help address cell-type specific disease processes by differentiating cells into various neural subtypes and enable us to understand better the functions of the protein in different cell types. Careful study of the molecular mechanisms in these isogenic cell models could enable us unravel the dispute of the progression pathways of the disease. Since findings from mouse models cannot be confidently extrapolated to humans because of species-specific phenotypic differences, human stem cell-derived neurons could be the long awaited model for HD research and empower the development of neuroprotective drugs and therapeutics.

Disease models

As our knowledge of the molecular mechanisms and disease progression pathways behind these neurodegenerative diseases is not adequate, disease models are necessary. Disease models not only enhance our understanding on the diseases, but also serve as a model for developing therapies and testing novel drug targets (Merkle & Eggan 2013). *In vivo* disease models, that surpass the obstacles of ethical concerns with regards to human testing have particularly played roles in the study of neuroscience and infectious diseases (Hunter 2008). These models have also provided simplified systems that are easy to observe and manipulate and are time-saving (Hau & Schapiro 2010). Much of what we know today on human diseases and physiology have been possible by the contribution of such models.

Animal models

Use of experimental disease models dates far back in the 18th and 19th centuries. One of the first disease models to allow a remarkable scientific finding was used by Louis Pasteur. He infected sheep with anthrax to demonstrate the germ theory of disease in the 1880s (Hau & Schapiro 2010). Modern day animal models for studying human diseases include mouse, zebrafish, drosophila (*D. melanogaster*), roundworm (*C. elegans*) – all of which mimic the human body's internal environment during the progression of a certain disease to some extent. With regards to using animals as disease models, this approach has been possible due to the conservation of genetic material across species in the course of evolution (Hunter 2008). Animal models have a number of advantages. Their small size allows convenience in handling and manipulating them. They breed in large numbers making them inexpensive (Barut & Zon 2000; Benavides & Guénet 2001). Many of their genome sequences are complete, e.g. roundworm and drosophila, which enables easy genetic manipulation. More importantly, the *in vivo* environment allows the study of both cell-autonomous and non-cell-autonomous contributions to a certain disease (Merkle & Eggan 2013).

Nevertheless, there are certain limitations of animal models. One major drawback is if the organism is not a true representative of the disease or phenomenon concerned in a study (Barut & Zon 2000). Although models e.g. mouse are very similar to humans due to shared ancestry, there are significant variations in their genetic material which is often a major obstacle in extrapolating findings obtained from them to humans. Many drugs that were effective in mice models failed in human trials (Merkle & Eggan 2013). Mice are inbred and thus lack the genetic diversity of the human population. A complete knowledge of the genome of the animal used in the study is also necessary in order to understand if effects observed are attributable to any other factors (Benavides & Guénet 2001). Although they are good for observing physical symptoms, animals may not always be ideal for studying the progress of internal abnormalities, such as neuronal dysfunction as in neurodegenerative disorders – the focus of this study.

Human pluripotent stem cells (hPSC)

To overcome such limitations, the most effective model that can better represent the internal environment of human cells while allowing us to observe developmental progress are human pluripotent stem cell (hPSC). For instance, disease progression can be easily followed over the course by live-cell imaging (Merkle & Eggan 2013). These undifferentiated and unspecialized cells, having the potential to develop into any cell type, can be easily cultured in the laboratory without significant ethical issues. They can then be differentiated into specialized cells or tissues of interest. Thus, these cells hold immense possibilities to serve as disease models, in development of tissue replacement therapies and in screening of novel drugs to test their efficacy and possible toxicity (Ebert et al. 2012). Unlike animal models, iPS cells grow very quickly in large quantities allowing large-scale genetic and chemical screens (Merkle & Eggan 2013) hence, results can be observed within a relatively short time. Moreover, these *in vitro* models allow the study of genotype-phenotype correlation in a controlled environment.

The two main types of pluripotent stem cells are the human embryonic stem (hES) cells and the human induced pluripotent stem (hiPS) cells. Human ES cells are isolated from the inner cell mass of blastocyst stage human embryos and have the potential to develop and differentiate into all the three germ layers and some extraembryonic tissues (Thomson et al. 1998). Embryonic stem cells from mouse blastocysts were first described in 1981. Almost two decades later, in 1998, human ES cells were derived from pre-implantation embryos developed in culture for 5 days after oocyte fertilization (Thomson et al. 1998). Mostly spare human embryos from IVF clinics were used. However, obtaining embryos specifically for this purpose

is controversial (Wert & Mummery 2003). Since deriving human ES cells involves the destruction of the embryo, ES cell models have ethical issues associated with them (Zacharias et al. 2011). The quality of the donated embryos to be used in research is also a concern (Wert & Mummery 2003).

Human iPS cells as disease models

Another strategy in disease modelling that circumvents the ethical issues of hES cells is to use human induced pluripotent stem (iPS) cells. Human iPS cells are obtained by genetically reprogramming adult somatic cells under specific conditions that induce the cells to revert to their embryonic pluripotent state (Takahashi & Yamanaka 2006). The reprogramming strategy was first demonstrated in 2006 by Yamanaka using adult mouse fibroblasts and later, using human skin fibroblasts (Takahashi et al. 2007) which earned him the Nobel Prize. In this protocol, adult human dermal fibroblasts are induced with four transcription factors, Oct3/4, Sox2, c-Myc, and Klf4, which function in maintaining the undifferentiated state of cells and their ability to self-renew. iPS cells are very similar to ES cells in terms of morphology, physiology, gene expression capabilities, and epigenetic status of pluripotent cell-specific genes. These iPS cells have the ability to differentiate into cells of all the three germ layers *in vitro* and in teratomas (Takahashi et al. 2007). Not restricted to dermal fibroblasts, iPS cells can also be derived by reprogramming neuronal cells, hematopoietic cells, adipose cells, and others (Ebert et al. 2012). Neuronal disease models using iPSCs, as is the focus of this study, was first demonstrated in 2008 (Dimos et al. 2008). In that research, ALS (amyotrophic lateral sclerosis) patient fibroblasts were reprogrammed into iPSCs and then differentiated into functional motor neurons.

Since human iPS cells are derived directly from somatic cells, the most prominent advantage of iPS cells over ES cell models is the overcoming of the ethical concerns. Unlike ES cells, it is possible to generate disease-specific as well as patient-specific cell models with iPS cells (Takahashi et al. 2007; Zacharias et al. 2011) which enable us to gain insights into the molecular mechanisms and progression of a certain disease. Disease-specific iPSC models can also be used in a more efficient and accurate drug screening (Ebert et al. 2012). Patient-derived iPS cells have the additional advantage in treatments in that the patients' own cells can be used for therapy which avoids the risks of tissue rejection. Not only patient cells, but heathy cells can also be reprogrammed to iPS cells and genetically altered to enable wider research. iPSCs, however, have some disadvantages such as high variability in their differentiation

propensities and in that, phenotypic effects at the organism level cannot be observed. The properties of hiPSCs may also be influenced by the incomplete silencing of reprogramming factors and the presence of mutations in the parental cell line prior to reprogramming (Young et al. 2012). For tissue replacement therapy, ectopic expression of c-Myc and Klf4 genes is considered dangerous as it may potentially lead to development of cancer. These oncogenes can however be substituted by other factors like Nanog and Lin28 to avoid such risks (Medvedev et al. 2010) or be eliminated altogether as suggested in a different study (Huangfu et al. 2008). Considering all these factors, iPSCs have clearly opened new prospects for medical research (Ebert et al. 2012), disease modeling, and discovery of therapeutic agents.

Genome editing

Obtaining patient-derived iPS cells is not always feasible as it requires proper consent and is time-consuming (Zacharias et al. 2011). Thus, iPS disease models can be generated by introducing disease specific mutations into the cells by genome editing. Genome editing of iPS cells enables us to gain a better understanding on the roles a certain gene plays in the human cell and provides a window into the pathophysiology of a genetic disease (Cox et al. 2015). Genome editing is particularly useful for the study of rare diseases as it might be difficult to find donors. Phenotypes in patient-derived cells, besides the gene of interest, might be due to variable gene interactions or a conserved genetic background. Genome edited iPS cell models would surmount such obstacles as they are generated to contain only the desired mutations, thus providing an isogenic control (Kim et al. 2014). In other words, any differences in phenotype compared to wild-type controls would be attributable to that particular mutation only. This technique also circumvents the laborious process of extracting patient cells, reprogramming them to iPS cells, and research associated ethical issues (Cox et al. 2015). With genome editing, a wide range of mutations specific to a certain disease can also be simultaneously studied in an experiment (Hsu et al. 2014). Such studies would, however, not be possible with patient-specific cells as they would mostly contain only one of the mutations leading to the disease. Gene editing can also be done *in vivo* in one-cell embryos to rapidly generate animal models. Genome editing is also a potential approach for devising new gene therapies where mutations can be corrected *in vitro* and edited cells then returned to the patient (Cox et al. 2015).

Genome editing is a technique used to modify DNA sequences by introducing double-stranded breaks (DSB) at the target site with the help of programmable site-specific nucleases,

followed by repairing the DNA. Disease models are thus generated by editing the iPS cell genome to introduce mutations relevant to the disease (Kim et al. 2014). There are two mechanisms of DNA repair: non-homologous end-joining (NHEJ) and homology-directed repair (HDR). NHEJ, which can occur at any phase of the cell cycle, uses DNA repair enzymes to join the ends of DNA strands at the DSB site. Usually, NHEJ uses microhomologies, short homologous DNA sequences (<6bp) often present on single-stranded overhangs, to repair the excised DNA (Smih et al. 1995; Urnov et al. 2010). These overhangs are, however, often not compatible which makes NHEJ error-prone introducing random indels in the DNA (Gupta & Musunuru 2014). These indels might cause a frameshift in the DNA sequence eventually leading to a premature stop codon downstream, and a truncated gene product. In contrast, HDR uses an exogenous DNA molecule as the repair template to introduce precise, desired mutations at the DSB site during DNA repair. In unmanipulated cells, HDR typically occurs during the late S phase or G2 phase of mitosis, during which a sister chromatid is available to serve as the repair template (Smih et al. 1995; Urnov et al. 2010). Either plasmid vectors or single-stranded oligodeoxynucleotide (ssODN) can serve as repair templates which contain the desired mutation flanked by homology arms, homologous to the sequence at the target site (Gupta & Musunuru 2014). Although it allows high-fidelity repair, HDR has a low efficiency compared to NHEJ, which is used more frequently by cells in the repair of double-stranded breaks (Smih et al. 1995).

Zinc Finger Nuclease (ZFN)

The first programmable site-specific nuclease is the Zinc Finger Nuclease (ZFN). ZFNs are artificial hybrid molecules which contain target DNA recognition domains called Zinc Fingers and the DNA cleavage domain – FokI nuclease (Urnov et al. 2010). Each zinc finger recognizes three distinct nucleotides and multiple such fingers combine together forming the DNA binding domain which recognizes a specific 3_n sequence as the target (Pavletich & Pabo 1991). For DNA cleavage, ZFNs work in pairs that recognize two 15-18bp sequences flanking the cleavage site, one on the forward and the other on reverse strand. When the two ZFN subunits bind to their respective targets, the FokI DNA cleavage domains dimerize with each other and create DSB in the target site with 5' overhangs (Guo et al. 2010; Urnov et al. 2010; Gupta & Musunuru 2014). Subsequently, the DNA repair processes take place.

Although the ZFNs initially held promises in genome editing, and is best understood and characterized, it has several disadvantages. The major drawback is that individual ZFNs

need to be designed for every genomic target, which makes it difficult, time-consuming, and an expensive process. It is difficult to assemble zinc finger domains to bind to a DNA sequence with high affinity. Moreover, they cannot be conveniently developed for any sequence of interest in that they can be only used to target binding sites every 200bps in a random sequence (Pavletich & Pabo 1991; Gupta & Musunuru 2014). The off-target effects are considerably high which also makes it toxic for mammalian cells.

Transcription activator-like effector nucleases (TALEN)

A second site-specific nuclease is the Transcription activator-like effector nucleases (TALEN). It originated from plant pathogens which secrete transcription activator-like effector (TALEs) proteins to regulate specific host genes (Miller et al. 2011). TALEs consist of an N-terminal secretion domain, a central array of modular DNA binding repeats, and a C-terminal transcriptional activation domain. A tandem array of 10-30 of such TALE repeats serve as the DNA binding domain of the TALENs. Each repeat is 33-35 amino acids long, with two adjacent amino acids, termed repeat-variable di-residue (RVD), being specific for a DNA base pair (Cong et al. 2012; Streubel et al. 2012). Thus, TALENs are designed by fusion of a domain of TALE repeats to the FokI endonuclease cleavage domain.

TALENs have many advantages compared to ZFNs. TALENs are easier to design with multiple possible TALEN pairs for each nucleotide on any DNA sequence. The TALE repeats bind to target DNA sequences with higher affinity than the ZFNs. Also, the TALE repeat array can be easily extended to the desired length, ranging over 18bp which is typically the maximum length in the case with ZFNs. However, off-target is a major concern with TALENs as with ZFNs. An obvious disadvantage of TALEN is its large size, which makes it harder to deliver it into cells. Moreover, the repeat units impair their ability to be packaged and delivered. The molecular architecture, including RVD module assembly and cloning into expression vectors, is labour-intensive (Cong et al. 2012; Gupta & Musunuru 2014).

CRISPR-Cas9

All of the difficulties and challenges associated with ZFNs and TALENs were overcome with the advent of an RNA-programmable nuclease based on the CRISPR-Cas9 (clustered regularly interspaced short palindromic repeats) system in bacteria (Mali et al, 2013; Cong et al., 2013). This technology has revolutionized genome editing and is, by far, the most efficient, easiest to design and use, inexpensive, and time-saving procedure. It is derived from natural bacterial adaptive immune systems directed towards exogenous pathogens, such as phages. Upon initial

exposure to phages, many bacterial species incorporate fragments of the phage DNA, called "protospacer", into specific arrays, called CRISPRs. Therefore, during a secondary exposure to the pathogen, the CRISPR DNA array is transcribed into small RNAs, which the bacteria then use as guides to direct site-specific nucleases to cleave the pathogen DNA only, leaving the host genome undamaged. At least 11 different of such CRISPR-Cas9 systems have been identified in prokaryotes, however, the one emerging as a simple and useful genome editing tool is that from the Type II system of *Streptococcus pyogenes* (Doudna & Charpentier 2014; Terns & Terns 2014; Dang et al. 2015).

Three major components of the natural *S. pyogenes* CRISPR-Cas9 system (Jinek et al. 2012; Doudna & Charpentier 2014; Dang et al. 2015) are:

i) crRNA (CRISPR RNA) which is encoded by the protospacers of the repeat CRISPR arrays in the bacterial genome. These crRNA molecules are 42 bases in length and consist of the guide sequence of 20-bases which is complementary to the DNA target.

ii) tracrRNA (trans-activating crRNA), 75-bases long and anneals with crRNA forming a dual crRNA: tracrRNA complex. TracrRNA enables activity of the crRNA through crRNA maturation, stem-loop formation, and binding to the Cas9 protein.

iii) Cas9 (CRISPR-associated) nuclease which binds to the crRNA: tracrRNA duplex and cleaves DNA. Cas9 is encoded by the Cas operon lying upstream of the CRISPR arrays. It has two nuclease domains – one HNH-like and one RuvC-like, each of which cleaves one strand of the target. In particular, the HNH domain cleaves the strand complementary to the crRNA, while the RuvC-like domain cleaves the one non-complementary to crRNA. Cas9 cleaves DNA at the site complementary to the 20-nucleotide guide sequence of crRNA which is three nucleotides upstream to the PAM sequence on the target DNA. PAM (protospacer adjacent motif) is a short sequence which lies adjacent to the 3'end of the 20-bp guide RNA (gRNA) sequence and is integral in target recognition by CRISPR-Cas9. For type II *S. pyogenes* Cas9 systems, the PAM is a 3-nucleotide NGG, where N stands for any nucleotide.

In genome editing experiments, crRNA and tracrRNA molecules are often combined artificially into a chimeric single-guide RNA (sgRNA) molecule by means of a linker loop. This sgRNA consists of the crRNA at the 5'end and tracrRNA at the 3'end. SgRNAs can be

obtained by cloning CRISPR oligos into expression vectors, e.g. U6, to be expressed within transfected cells, eventually acting as guides for DNA cleavage. They are also made by *in vitro* transcription in vectors that contain a T7 promoter (Hsu et al. 2014; Nowak et al. 2016).

Compared to ZFNs and TALENs, which require designing proteins for each new target site, designing CRISPRs is easier. This is because the PAM sequence, NGG is found abundantly, in every 8 bases of a random sequence, in the human genome (Cong et al. 2013), and the target sequence needs only to lie upstream of the PAM site. Moreover, due to its high GC-content, NGG is found abundantly in protein coding exons making it easier to engineer any protein coding gene (Wu et al. 2014). If the *S. pyogenes* system were not optimal for targeting a site, there are Cas9 variants from other species with different PAMs that can be used. Multiple gRNAs can be used simultaneously to easily target multiple genes in the same cell (Wu et al. 2014). CRISPR-Cas9 also provides the major advantage of cleavage specificity over the previous systems. It is suitable and highly efficient in mammalian stem cells, as well as in mouse one-cell embryos *in vivo* with cytotoxicity being relatively low (Shen et al. 2014; Wang et al. 2016).

Although cleavage at off-target sites (OTS) is an issue, careful selection of CRISPRs with few highly-related sites in the genome can minimize the effects. A strategy that can substantially reduce off-target effects is to shorten the guide sequence. This makes the guide less tolerant to mismatches while preserving the on-target efficacy (Fu et al. 2014). Using Cas9 nickase can also alleviate the problem of OTS. Cas9 nickases, which are formed by mutating either of the cleavage domains, can make single-stranded nicks in the DNA at two closely linked sites (Ran et al. 2013; Shen et al. 2014). This way they cleave both the strands leaving 5’ overhangs. Active nickase cleavage sites are typically paired within a range of 100bp. This method avoids any probable mutation that happens in cases with wild-type Cas9 generated DSBs, in that, the repair product is only sealing off the nick (Ran et al. 2013; Shen et al. 2014). In this way, off-target effects, if any, can be efficiently alleviated by Cas9 nickase without sacrificing the specificity or robustness (Iyer et al. 2015).

CRISPR-Cas9 has numerous potential applications. Besides being used to study gene functions, they can be used in generating gene therapies for inherited diseases. They are also used in cancer studies, pharmacological studies using mouse models, and genome-wide association studies. Besides, they have proved potential solutions in agriculture by producing resistant crops, and in immunization of industrially important bacteria against phages. A dead

form of Cas9 with no cleavage activity, dCas9, can also be used in studying transcriptional regulation, targeted epigenetic changes, or live cell imaging (Wu et al. 2014; Yang 2015). In this study, CRISPR-Cas9 system was used to generate disease models in human iPS cells.

Study outline

The aim of my thesis was to generate human iPS cell disease models for two neurodegenerative disorders – Triple A syndrome and Huntington's disease by genome editing using CRISPR-Cas9. I set out to generate the following alleles:

1. Deletion of *AAAS* exon 2: establish phenotype of a biallelic knockout of the gene or complete loss-of-function allele
2. Point mutation in *AAAS* exon 1: establish phenotype of a homozygous 43C>A mutation generating a splice-donor site
3. Insertion in HTT exon 1: insertion of an extended 67 CAG-repeat fragment into the gene and compare results obtained in iPS cells with those in human ES cells

For each allele, I performed the following experiments:

- Molecular cloning to generate CRISPR gRNA plasmids and donor vectors for genome editing
- Nucleofection of human iPS cells with CRISPR-Cas9 and donor vectors to introduce the disease relevant mutations for AAAS and HD in the iPS cell genome
- Tissue culture of engineered cells with drug selection, where applicable, followed by expansion, subcloning, and archiving of the clones
- Genotyping and screening by Sanger sequencing to identify the desired mutations in the engineered cells
- Validation of the cell models by measuring protein expression levels using Western blot

The iPS cell models I generated lay the foundation for future detailed phenotype analysis and for testing potential drugs or gene therapies in the future.

Chapter 2
Methods & Materials

This research was conducted from January 2016 to September 2016 at the Wellcome Trust Sanger Institute. Tissue culture experiments with human induced pluripotent stem (iPS) cells were performed in containment level 2 laboratories of the institute. The protocols and methods used in this study are described in this chapter. I conducted all the experiments, unless otherwise stated.

Induced pluripotent stem cells

Human induced pluripotent stem (iPS) cell line, KOLF2 (ID: HPSI0114i-kolf_2), was obtained from HipSci (Human Induced Pluripotent Stem Cell Initiative; www.hipsci.org). This cell line was subcloned and karyotyped by FISH (Fluorescence in situ hybridization) prior to genome editing experiments. The subline, KOLF2-C1, obtained was used for the tissue culture experiments described herein. Both the parental KOLF2 and KOLF2-C1 subline have a normal 46; XY karyotype (Koutsourakis, Bushell & Skarnes; unpublished data).

Bioinformatics

CRISPRs were designed using the HTGT WGE CRISPR Search (Hodgkins et al. 2015) (http://www.sanger.ac.uk/htgt/wge/find_crisprs), and InFusion primers were designed using WGE Gibson Designer (http://www.sanger.ac.uk/htgt/wge/gibson_designer). All other PCR primers were designed using Primer3 (http://bioinfo.ut.ee/primer3/) or manually selected.

Synthetic DNA nucleotides, recombinant proteins, and services

All PCR, InFusion, sequencing primers, and CRISPR gRNA oligos were manufactured by Sigma. For the AAAS point mutation experiment, crRNAs, tracrRNA, and *AAAS* point mutation ssODN were all purchased commercially from Integrated DNA Technologies. Recombinant Cas9 nuclease was purchased from two commercial sources, Thermoscientific and Feldan. HTT synthetic plasmid was obtained from Invitrogen. All Sanger sequencing samples were sent to Eurofins in Germany for capillary sequencing.

Molecular biology techniques

Extraction of genomic DNA

The cells from one well of a confluent 6-well plate (BD Biosciences) were washed with 3ml DPBS (Gibco Life Technologies), and then lysed in 0.5ml TENS buffer (10mM Tris pH 8/ 1mM EDTA/ 150mM NaCl/ 0.5% SDS) containing freshly added Proteinase K (1-1.5 mg/ml; Sigma) at 60^{o}C for at least 4 hours. The cell lysate was transferred to a 1.5 ml microfuge and

an equal volume of phenol: chloroform (1:1; Amresco) was added to the cell lysate and mixed by inversion for 5 minutes. Following centrifugation for 5 minutes at 13000rpm, the aqueous phase was transferred to a new 1.5ml tube. The above step was repeated once more and the genomic DNA was precipitated with 0.8 volume of isopropanol after which the spooled DNA was washed twice in 1ml of 70% ethanol. The DNA was resuspended in TE buffer (10mM Tris/ 1mM EDTA) and placed in a 60ºC incubator for 20 minutes to evaporate any ethanol. The sample was left to resuspend overnight at room temperature (RT) and the DNA concentration was determined from its absorbance at OD_{260nm} on the following day.

Transformation of chemically competent DH5α and plasmid isolation

Plasmid DNA was added to thawed 50µl of *E. coli* DH5α cells (NEB) and incubated on ice for 20 minutes. The cells were then heat-shocked at 42ºC for 45 seconds followed by another 2-minute incubation on ice, after which the cells were allowed to recover in 500µl of SOC medium (NEB) for an hour in a 37ºC shaking incubator. The cells were then plated at two different densities on LB agar plates (Sanger media kitchen) containing the appropriate concentration of antibiotic and incubated overnight at 37ºC.

Following overnight culture of transformed cells, the desired number of colonies from the plates were used to set up overnight liquid cultures at 37ºC in 2xLB broth (Sanger media kitchen) containing any antibiotic at the appropriate concentration. Plasmid DNA was extracted from these cultures using the QIAprep spin miniprep kit (QIAGEN) as per the manufacturer's instructions. To prepare glycerol stocks, 50% glycerol (Sanger media kitchen) was added to cell cultures at a ratio of 3:2 and stored at -80ºC. Additionally, diagnostic restriction digests, where necessary, were performed using buffers and units of enzymes recommended by the manufacturer and run on 0.7% agarose (Invitrogen) gel to confirm the presence of the correct-sized DNA fragments.

Sequencing and ethanol precipitation of samples

Sanger sequencing reaction

For sequencing reactions, equal volumes of DNA and sequencing reaction master mix (Table 2.1) were added together (usually 2µl of DNA with 2µl of sequencing master mix). The reaction mixture was then subjected to sequencing reaction in a PCR machine (program in Appendix, Table A4).

Table 2.1: Sequencing reaction master mix

Component	**Volume (µl) for 1x reaction**
H_2O	0.9225
SRD (sequencing reaction diluent)	0.9225
BigDye V3.1 terminator mix (Applied Biosystems)	0.125
dGTP BigDye V3.0 terminator mix (Applied Biosystems)	0.02
Sequencing primer (100µM)	0.01
Total	**2µl**

***** SRD (0.3M Tris Base pH 9.0/ 6mM $MgCl_2$/ 0.18%v/v Tween-20/ 5.9% v/v Glycerol / 1.1% v/v Formamide (Sigma)/ 9.9% v/v Tetramethylene Sulfone (Sigma)/ 7.8mM Pottassium Glutamate (Sigma)/ 79µg/ml BSA (Sigma))

Ethanol precipitation of sequencing reaction

To the sequencing reaction from the preceding step, 30µl of precipitation mix (80% ethanol/ 50mM sodium acetate) was added and centrifuged for 25 minutes at 4000rpm and 4ºC. The plate was then inverted on a blotting pad and spun at 400rpm for a minute. To the precipitated DNA, 30µl of 80% ethanol at -20ºC was added and spun again for 10 minutes at 4000rpm and 4ºC. The plate was inverted again on a blotting pad and spun for 1 minute at 400rpm, and finally dried overnight at RT or for 20 minutes at 60ºC before sending the plates for analysis on a capillary sequencing machine (service provided by Eurofins, Germany).

Construction of CRISPR gRNA expression plasmids

Annealing CRISPR gRNA oligos

CRISPRs were chosen using the WGE CRISPR search website, in that, the target site in the genome composed of the sequence complementary to the first 20 bases of the guide RNA plus the adjacent PAM sequence. The gRNA oligos were appended at their 5'ends with a 4-base sequence compatible with the BsaI cloning site of the U6_gRNA (AU flip) expression plasmid, containing a modified guide RNA backbone (based on Chen et al. 2013 by Koutsourakis & Skarnes, unpublished; refer Appendix Figure A9). The oligos were resuspended in TE at a concentration of 1µg/µl. 2.5µg of the forward strand gRNA oligo was annealed to the same amount of reverse strand (or complementary) oligo in 50µl of annealing buffer (10mM Tris, pH 7.5-8; 50mM NaCl; 1mM EDTA). The mixture was heated at 98ºC in a heating block for 5 minutes followed by cooling down slowly to RT inside the block. The annealed oligos were then diluted 1:10 in H_2O.

Cloning annealed gRNA oligos into U6_gRNA (AU flip) expression plasmids

Ten micrograms of U6_gRNA (AU flip) plasmid was digested with BsaI (NEB), overnight at 37ºC. The linearized plasmid was then purified using the QIAquick PCR purification kit (QIAGEN) according to the manufacturer's instructions. 20ng of each of the annealed diluted gRNA oligos and the linearized plasmid were ligated with Takara ligase/ buffer mix (Takara ClonTech) and incubated at RT for 1-2 hours. The ligated plasmids were then transformed into chemically competent DH5α cells and cultured overnight at 37ºC on LB agar plates containing 50µg/ml Kanamycin (kan) (Sigma).

Following transformation, overnight liquid cultures were set up in 2xLB containing 50µg/ml kan, and plasmids were extracted from the culture. Subsequently, diagnostic restriction digests were performed on the extracted plasmids and agarose gels run to confirm the presence of correct gRNA expression plasmids. The plasmids with correctly sized bands on an agarose gel were then subjected to sequencing reaction with U6_gRNA (AU flip) forward and reverse sequencing primers (listed in Appendix, Table A8) and ethanol precipitated to send for capillary sequencing. Bacterial cultures for CRISPR gRNA plasmids with perfectly accurate sequences were then grown in maxiprep cultures (100ml of 2xLB containing 50µg/ml kan) overnight in a 37ºC shaking incubator. The plasmids were extracted from the culture on the following day using QIAGEN Plasmid Maxi Kit as per the manufacturer's instructions, and the concentration determined from OD_{260nm} in a UV spectrophotometer. The plasmids were then used for subsequent nucleofection experiments.

InFusion cloning

pUC19_RV plasmid used for InFusion cloning is a modification of pUC19 (NEB) that contains a unique EcoRV linearization site (Koutsourakis & Skarnes, unpublished; refer Appendix Figure A10). Gene-specific PCR primers for InFusion cloning were designed with the WGE Gibson Designer (for AAAS KO experiment) or Primer 3 (for HTT experiment). Each primer was appended, at its 5' end, with a 15-base extension homologous to the ends of the linear pUC19_RV plasmid. Thus, the PCR products generated by these InFusion primers would contain ends that are homologous to those of the vector. The primers were resuspended in TE and 5µM of each was used to amplify the target fragment from 50ng of wild-type KOLF2_C1 genomic DNA with CloneAmp HiFi PCR premix (Takara Clontech) according to the manufacturer's instructions (see Appendix, Table A1 for PCR program). 10µg of pUC19_RV plasmid was digested with EcoRV (NEB) restriction enzyme and the linearized plasmid

purified using QIAquick PCR purification kit according to the manufacturer's instructions. The concentration of both the PCR fragment and backbone DNA were then quantified on an agarose gel. Approximately 50ng of the PCR-amplified fragment was mixed with 50ng of linearized purified pUC19_RV plasmid, and InFusion reactions were carried out with InFusion HD enzyme (Takara Clontech) for 15 minutes at 50ºC as per manufacturer's instructions. The InFusion reaction was transformed into 50µl Stellar competent cells (Clontech) and cultured overnight at 37ºC on LB agar plates containing 50µg/ml ampicillin (amp) (Sigma).

Overnight miniprep cultures were set up in amp-supplemented 2xLB for screening the resultant InFusion plasmids. Plasmids were extracted and diagnostic restriction digests were performed to confirm the presence of the correct InFusion clones by agarose gel electrophoresis. For clones showing the correct-sized bands, sequencing reactions were performed with pUC19 forward and reverse sequencing primers (listed in Appendix, Table A8), ethanol precipitated and sent out for capillary sequencing.

Construction of targeting donor vector for knockout of exon 2 of *AAAS* gene

The donor vector for targeting the *AAAS* gene was constructed in two steps by first introducing a Gateway cassette- Zeo/PheS (zeocin/phenylalanine tRNA synthetase; Skarnes et al. 2011) into the InFusion clone by recombineering (schematic diagram in Figure 2.1) and then introducing a drug selection cassette- pL1L2 EF1α-puro-polyA (Koutsourakis & Skarnes, unpublished; see Appendix A11) by performing a Gateway exchange reaction (schematic diagram in Figure 2.2).

Recombineering

I. Amplification of Zeo/PheS (Z/P) gateway cassette fragment

Initially, 70-mer recombineering primers (U5 and D3; see schematic diagram below in Figure 2.1) were designed, each consisting of 20 bases homologous to the Z/P Gateway element and 50 bases homologous to sequences flanking exon 2 of *AAAS* (listed in Appendix, Table A5). These 70-mer primers were used for amplifying the Z/P gateway cassette fragment (program in Appendix, Table A3) which was then subsequently inserted into the AAAS InFusion plasmid by homologous recombination in bacteria. The R1R2 flanked Z/P cassette in the recombineered AAAS-Z/P plasmid would serve as an intermediate substrate for the LR Clonase Gateway exchange reaction with pL1L2 EF1α-puro-polyA, forming the resultant AAAS-EF1α-puro-polyA targeting donor plasmid.

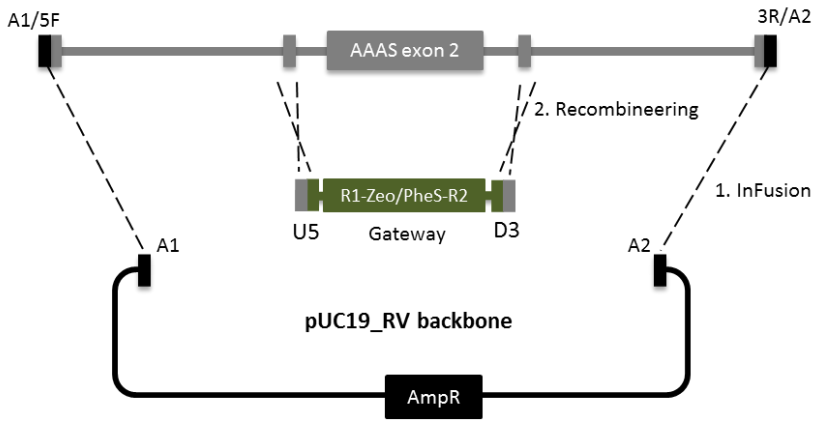


Figure 2.1: Construction of AAAS-Z/P plasmid by InFusion and recombineering for generating the final targeting donor vector for knockout of *AAAS* exon 2. 5F and 3R are InFusion PCR primers for amplification of *AAAS* exon 2.

For the amplification of Z/P cassette fragment, 10µM of each of the primers were used on 1ng of Z/P template using PrimeSTAR Max (Takara Clontech) according to the manufacturer's instructions. The PCR product fragment was then purified of any remaining primers using the Nucleospin PCR clean-up and gel extraction kit (Takara Clontech) following the manufacturer's instructions, and run on an agarose gel to estimate its concentration.

II. Transformation of recombineering-ready bacteria with the AAAS InFusion plasmid

The pBAD-gbaA plasmid (Wang et al. 2014) was transformed into DH10B bacteria and selected in tetracycline-containing 1xLB media and grown at 30^{o}C. pBAD-gbaA expresses the recombineering enzymes of the lambda phage red operon under the control of an arabinose-inducible promoter. The backbone, derived from pSC101, contains a temperature-sensitive origin of replication, and thus is lost from bacteria when grown at 37^{o}C.

a. Preparation of competent cells –

pBAD-gbaA bacterial cells were grown overnight at 30^{o}C in low salt LB (Sanger media kitchen) containing 5µg/ml tetracycline (tet) (Sigma). To prepare competent cells, the overnight culture was diluted 1:50 in 1ml low salt LB on the following day, and a 2.5-hour culture started in the same conditions as above. The culture was grown to mid-log phase (absorbance of the culture at OD_{600} = 0.2-0.4 and then incubated in ice/water slurry for 5 minutes. The cells were pelleted at 4000rpm for 5 minute at 4^{o}C and electrocompetent bacteria were made by washing the cells three times with ice cold HPLC-pure water.

b. Electroporation of AAAS InFusion plasmid

The electrocompetent cells were resuspended in 40µl water containing approximately 50ng AAAS InFusion miniprep plasmid and transferred to a 96-well cuvette placed on ice. The cells

were then electroporated at 2400V, 700Ω, and 25µF (time constants around 10msec) in a electroporator (BTX Harvard apparatus) and 50µl of 2xLB recovery media (containing 0.2% glucose without antibiotic) was added promptly into the cuvette well, while placed on ice. The media was mixed with the cell/DNA mixture by pipetting 4-5 times, and then transferred into a tube containing 500µl of recovery media (containing 0.2% glucose) without any antibiotic. The cells were then grown in 30^{o}C shaking incubator for 1 hour.

250µl of the above recovery culture was inoculated into 750µl of low salt LB containing 50µg/µl ampicillin (amp) and 5µg/µl tet, and incubated overnight in 30^{o}C shaker. On the following day, the overnight culture was diluted 1:50 in 1ml of low salt LB supplemented with tet and amp, and cultured for 2.5 hours in 30^{o}C shaking incubator.

c. Induction of the red operon with arabinose –

When the culture reached mid-log phase (absorbance at OD_{600} = 0.2-0.4), L-(+)-Arabinose 99% solution (99%; Sigma), to a final concentration of 0.2%, was added and left for 3-5 minutes in a 37^{o}C water bath. The culture was then transferred to a 37^{o}C shaking incubator and incubated for another 35 minutes. Arabinose addition induces the expression of recombineering proteins, red gamma, beta, and recA to allow subsequent homologous recombination to occur.

III. Transformation of arabinose induced bacterial cells with Z/P Gateway fragment

Electrocompetent cells were prepared from the arabinose induced cells as described above in section II (a). The pelleted competent cells were resuspended in 40µl water containing a total of 100-150ng of amplified Z/P Gateway fragment (amplified as described in section I) and electroporated as described in section II (b). However, these transformed cells were recovered at 37^{o}C shaking incubator for an hour.

250µl of the above recovery culture was inoculated into 750µl of low salt LB containing 50µg/µl amp and 6.5µg/µl zeocin (zeo) (Invitrogen), and incubated overnight in a 37^{o}C shaker. At this point, the amplified Z/P Gateway fragment is expected to have recombined with the InFusion plasmid using the 50-bp homology regions to result in the AAAS-Z/P Gateway (pAAAS-ZP) plasmid. The recombineered clones should be resistant to both zeo and amp. Plasmids were extracted from the culture using QIAprep spin miniprep and were re-transformed into DH5α and Stellar competent cells and cultured overnight at 37^{o}C on LB agar plates containing amp (50µg/µl) and zeo (6.5µg/µl); this step was performed to obtain pure colonies of cells that contain only the desired pAAAS-ZP plasmid. Miniprep cultures from single colonies were set up and plasmids were extracted. Diagnostic restriction digests with

different enzymes were performed on the extracted plasmid DNA to identify the correct pAAAS-ZP clones showing the expected sized bands on an agarose gel.

Gateway exchange reaction to produce the final AAAS donor vector

LR Clonase II enzyme (Invitrogen), containing bacteriophage lambda recombination proteins Integrase (Int) and Excisionase (Xis), catalyzes the *in vitro* recombination between *att*L-flanked EF1α-puro-polyA plasmid and *att*R-flanked Z/P cassette in the pAAAS-Z/P vector. This recombination results in the final *att*B-containing donor vector, AAAS-EF1α-puro-polyA (see Figure 2.2 below). The Gateway reaction can be represented as:

*att*L1–EF1α-puro-polyA–*att*L2 x *att*R1–Z/P–*att*R2 ⇔ *att*B1–EF1α-puro-polyA–*att*B2 x *att*P1–Z/P–*att*P2 (by product)

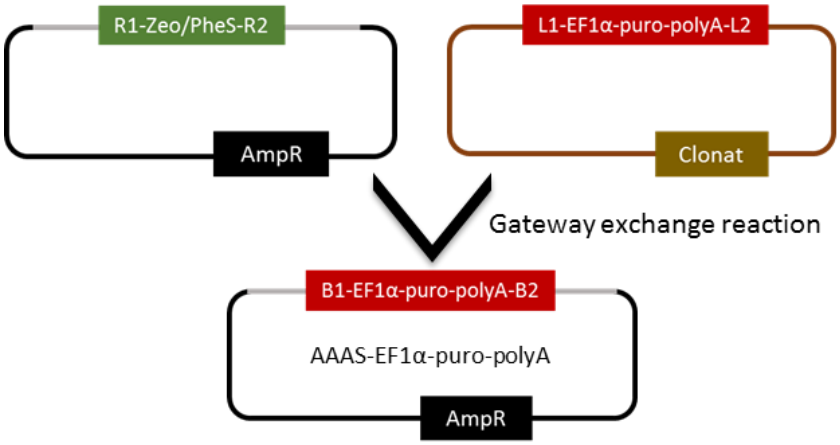


Figure 2.2: Construction of AAAS-EF1α-puro-polyA targeting donor vector by introduction of a drug selection cassette- pL1L2 EF1α-puro-polyA into AAAS-Z/P recombineered clone by LRII clonase Gateway exchange reaction

Gateway recombination reaction was set up with LR Clonase II enzyme (Invitrogen) by adding 100-200ng of AAAS-Z/P gateway plasmid DNA to 70ng of pL1L2 EF1α-puro-polyA plasmid in a total reaction volume of 10µl, and incubating at 25°C overnight as recommended by the manufacturer. After the overnight incubation, 1µl of proteinase K was added to the reaction and incubated for 10 minutes at 55°C to inactivate the Clonase enzyme. 2µl of the LR reaction mixture was then transformed into 35µl of DH10B competent cells (NEB) and cultured on YEG (yeast extract glucose plus P-chlorophenylalanine) agar plates (Sanger media kitchen) containing 50µg/ml amp. PheS (phenylalanyl tRNA synthetase) expression in *E. coli* is important in this selection in that, a mutant derivative of PheS is sensitive to p-chlorophenylalanine (Kast 1994). Thus, the mutants can be selected against in

YEG medium to kill PheS positive cells including the initial AAAS-Z/P Gateway plasmid and the by-product of LR Clonase II reaction.

From colonies on YEG-amp plate, overnight miniprep cultures were set up in 2xLB with 50µg/ml amp and plasmids extracted on the following day. Diagnostic restriction digests were set up to confirm the construction of the correct donor vectors that show the expected-sized bands on an agarose gel. The full annotated sequence of the final AAAS-EF1α-puro-polyA donor plasmid is shown in Appendix A12.

Construction of targeting donor vector for inserting extended Q-repeat in HTT gene

The HTT donor vector was constructed in two steps by InFusion cloning of exon 1 of the HTT gene and then cloning in a synthetic fragment containing an expanded Q-repeat domain.

InFusion cloning of HTT exon 1

InFusion cloning of HTT exon 1 into pUC19_RV plasmid (Figure 2.3) was performed in a similar manner as described earlier in InFusion cloning.

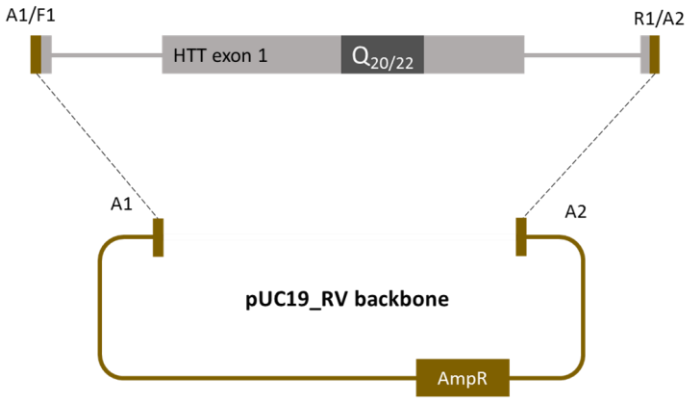


Figure 2.3: Construction of HTT InFusion clone, pHTT_exon 1

Design and synthesis of an extended Q-repeat fragment

A synthetic fragment containing 67-Q repeats in exon 1 of the HTT gene was synthesized and cloned into a kan-resistant plasmid backbone (GeneArt; refer Appendix A13). This extended repeat region was then cloned into the HTT InFusion plasmid as a BlpI/BsgI fragment (refer to Figure 2.4 below). This fragment contained mutations in PAM sites of the two CRISPRs used for the targeting experiment, so that Cas9-RNP did not cleave the donor plasmid itself, neither re-cleaved genomic DNA once the plasmid had been introduced successfully by homologous recombination.

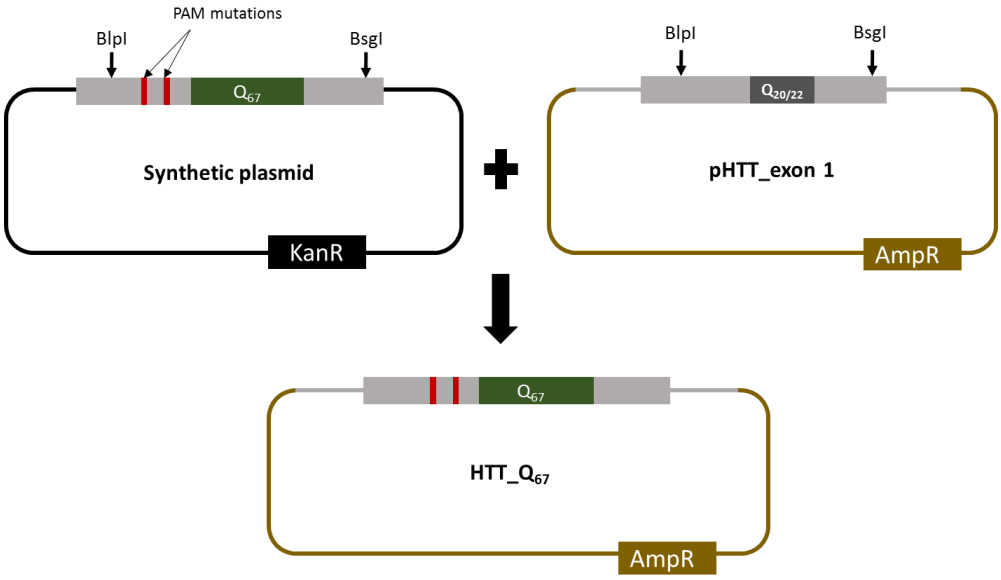


Figure 2.4: Construction of HTT_Q-67 donor plasmid by cloning in of a BlpI/BsgI digested Q67 synthetic fragment into pHTT_exon 1.

Cloning the extended Q-repeat fragment

Both synthetic and InFusion plasmids were first digested with the two restriction enzymes BlpI and BsgI. 2.5µg of HTT InFusion clone was digested in a total reaction volume of 50µl at 37°C for 3-4 hours. The backbone was then purified of the digested smaller fragment using QIAquick PCR purification kit following the manufacturer's instructions. Likewise, 5µg of the synthetic plasmid was digested in a total volume of 100µl under the same reaction conditions, and then run on 1.5% agarose gel. The extended fragment was then cut off from the gel under a UV box using a scalpel and extracted using QIAquick Gel Extraction Kit (QIAGEN) according to the instructions provided by the manufacturer. The concentrations of the fragments were then estimated on an agarose gel. Equal amounts of the synthetic fragment and the InFusion backbone were ligated together using Takara ligase/buffer mix (ClonTech) in a total reaction volume of 10µl at RT for 2 hours. This ligation reaction was then transformed into DH5α competent cells and cultured on ampicillin LB plates overnight at 37°C.

Plasmids were extracted from overnight miniprep cultures and restriction digested to diagnose the correct targeting donor plasmids from their bands on an agarose gel before they were sequenced with HTT-specific sequencing primers (listed in Appendix, Table A7). The plasmid DNA containing the perfectly matched sequences were then prepared from maxiprep cultures and extracted using QIAGEN Plasmid Maxi Kit as per the manufacturer's instructions. This donor plasmid was then used for subsequent targeting experiments in KOLF2-C1 cells.

In vitro Cas9 nuclease assay

Equimolar amounts of crRNA and tracrRNA (IDT) were annealed by heating them together at 58°C for 10 minutes and cooling down slowly to RT. The annealed cr:tracrRNA was diluted to 300nm, and added with 1µM Cas9 Nuclease NLS (NEB), both to a final concentration of 30nM in 1x Cas9 buffer. After incubating this reaction at RT for 10 minutes, InFusion plasmid DNA was added to a final concentration of 3nM and incubated at 37°C for an hour. Following this, the reaction was run on agarose gel for determination of synthetic gRNA activity.

T7 Endonuclease (T7E1) assay

T7E1 assay is often used to detect Cas9 activity in experiments. The T7E1 enzyme recognizes and cleaves non-perfectly matched DNA and heteroduplexes which are formed as a result of Cas9 activity. The cleaved products can then be run on a gel and detection of multiple bands would indicate that T7E1 cleaved the heteroduplexes which were formed due to Cas9 activity. For this assay, PCR reactions were set up with 100ng genomic DNA from Cas9-modified and wild-type cells using 20pmol of the gene-specific primer pairs (Table 2.2). The PCR amplicons were purified using QIAquick PCR purification kit and run on agarose gel to estimate their concentration. 200ng of the DNA in 19 µl NEBuffer 2 (NEB) was heated for 10 minutes at 95°C, then slowly cooled down to RT inside the block. To the reaction, 1µl of T7E1 enzyme (NEB) was added and incubated at 37°C for 15 minutes, after which the reaction was stopped by adding 0.75µl of 0.5M EDTA. The reaction was then run on an agarose gel to estimate the Cas9 activity.

PCR genotyping of human iPS cell lysates

For genotyping of cell lysates, PCR primers were first tested on wild-type KOLF2-C1 cells in order to find out the best primer combination that gave only the expected amplicon and no non-specific bands on an agarose gel.

PCR amplification of lysates

The cells in the wells of a confluent 96-well plate (BD Biosciences) were washed with 200µl DPBS per well. Discarding DPBS, 100µl of yolk sac lysis buffer (50mM KCl/ 10mM Tris-HCl pH 8.3/ 2mM $MgCl_2$/ 0.45% IGEPAL CA-630 (Sigma)/ 0.45% Tween-20 (Sigma)) with 1-1.5mg/ml of freshly added proteinase K was added per well. The plate was sealed and incubated in 58°C incubator for 4 hours to overnight, after which it was heated at 95°C on a heating block for 10 minutes to inactivate proteinase K. The crude lysates were then diluted 1:10 or 1:20 in

10mM Tris-HCl, pH 8, depending on their confluency. PCR reaction mix (Table 2.2) was used to set up a 10µl PCR reaction with 2.5µM of PCR primers and approximately 100ng of DNA. Short-range or mid-range PCR (see Appendix, Table A2 for program) was used to amplify the DNA depending on the length of the amplicon.

Table 2.2: PCR reaction mix for genotyping human iPS cell lysates

Component	**Volume (µl) for 1x reaction**
H_2O	2.1
5x LongAmp Taq buffer (NEB)	2
10mM DNTP (Thermoscientific)	0.3
100% DMSO (Sigma Aldrich)	0.2
LongAmp Taq DNA polymerase (NEB)	0.4
Total	**5 µl**

Purification of PCR amplicons

The PCR amplicons were purified with exonuclease I (NEB) and phosphatase (NEB) in successive reactions to remove the nucleotides and phosphate groups from the ends of DNA strands. In this reaction, 10µl of the PCR reaction was diluted 1:2.5 in H_2O and half of the diluted PCR amplicons was added with exonuclease reaction mix (Table 2.3). This mixture was then heated at 37ºC for 30 minutes followed by 80ºC for 15 minutes. Phosphatase reaction mix (Table 2.3) was then added to the mixture and heated again in the same manner as with exonuclease I. The resulting single-stranded DNA molecules were then diluted 1:4 in H_2O and used in subsequent sequencing reactions.

Table 2.3: Exonuclease and phosphatase reaction mixes for purifying PCR amplicons

Component	**Volume (µl) for 1x exo reaction**	**Volume (µl) for 1x phos reaction**
H_2O	3.2	2.85
10x exo-phos buffer	1.5	2
Exonuclease I (NEB)	0.3	–
Phosphatase (NEB)	–	0.15
Total	**5 µl**	**5 µl**

* 10x exonuclease-phosphatase (exo-phos) buffer (100mM $MgCl_2$/ 200mM Tris-HCl, pH 8)

Sequencing and ethanol precipitation

For the sequencing reaction, equal volumes of exonuclease-phosphatase (exo-phos) purified DNA and sequencing reaction master mix (Table 2.1) were added together in processes similar to those described earlier in the section describing sequencing. Ethanol precipitation was then

used to precipitate the samples (described in section ‘Sequencing and ethanol precipitation of samples’) prior to sending them for capillary sequencing.

Western blot

Protein lysis

The cells in a well of a 6-well plate were first washed with 3ml DPBS and the plate was placed on ice. 0.3ml of ice cold protein lysis buffer (50mM Tris, pH 7.5/ 0.5M NaCl/ 1% IGEPAL CA-630/ 1% sodium deoxycholate (Sigma Aldrich)/ 0.1% SDS (Ambion)/ 2mM EDTA/ Complete protease inhibitors (Roche)) was then added per well and incubated on ice for 5 minutes with occasional shaking. The lysate was then scraped using a cell lifter (Corning) and transferred into a 1.5ml microfuge tube to shear the DNA present in it in successive processes or stored in –70^{o}C freezer.

Shearing DNA in protein extracts

Lysates, thawed on ice, were spun through QIAshredder column (QIAGEN) at the maximum speed for 2 minutes at 4^{o}C. The flow-through was transferred into a 1.5ml tube and spun for another 10 minutes at 13000rpm and 4^{o}C. The resulting supernatant was transferred carefully into another tube, avoiding any DNA pellet at the bottom of tube, and stored on ice or at -70^{o}C until further use.

Protein quantitation

I. Quantitation of BSA (Bovin serum albumin) standards

BSA (Sigma Aldrich) with stock concentration of 1mg/ml was diluted by 1:2 in 100µl of protein lysis buffer and continued to obtain a series of dilutions, each time diluting 100µl of the preceding standard with another 100µl protein lysis buffer in the tube following it. 100µl of these serially diluted BSA standards (and a control containing 100µl protein lysis buffer only) were each mixed with 2ml Bicinchoninic Acid (BCA; Sigma Aldrich) made with BCA reagent A to reagent B ratio of 25:1. BCA contain copper (II) ions which are reduced to copper (I) ions by proteins, in a concentration dependant manner. The resulting copper (I) complex has an absorbance maximum at 562nm which is directly proportional to the concentration of the solution. Thus, absorbance of proteins is an indicator of the total protein content in a solution. The mixture was heated for 30 minutes at 37^{o}C and then cooled to RT. The OD_{562nm} of the solutions were then measured in a UV-spectrophotometer. From the absorbance values

and their corresponding concentrations, a graph was constructed, the best-fitted line of which was used for determination of concentration of the sample protein extracts.

II. Quantitation of sample protein extracts

The protein extracts were diluted 1:8 in protein lysis buffer, and 100µl of the diluted extract was mixed with 2ml BCA (like the standards) and heated likewise at 37^{o}C for 30 minutes. The samples were cooled and their absorbance at OD_{562nm} then measured in a UV-spectrophotometer. Using the graph constructed from the BSA standards, the concentrations of the sample protein extracts were determined.

SDS-PAGE

Equal amounts of each of the protein samples were measured out. Sample loading buffer (Invitrogen), equal in volume to that of the protein sample in a tube, was added. The mixture was heated in a heating block at 85^{o}C for 5 minutes, shortly spun, and loaded on a pre-cast NuPAGE^{TM} Novex^{TM} 4-12% bis-tris polyacrylamide gel (Invitrogen), along with 6µl of a protein standard (BIORAD). The gel was run in NuPAGE^{TM} Novex^{TM} SDS-PAGE gel system (Invitrogen) initially at 85V for 15 minutes and then at 100V until the dye front reached the bottom end of the gel.

Trans-blotting

One litre of 1x transfer buffer was prepared from 10x Tris/glycine buffer (BIORAD) with 200ml of methanol and MilliQ water. PVP membrane (Millipore) was soaked in methanol for 20 seconds followed by washing off the methanol in MilliQ water for 2 minutes. The membrane was then soaked in transfer buffer.

Cracking open the plates, the gel was taken off carefully on a wet filter paper soaked in transfer buffer and arranged in a sandwich (Figure 2.5) with the membrane facing the anode. The Mini Trans-Blot Electrophoretic Transfer Cell (BIORAD) was run at 150V for 2 hours for transfer of proteins onto the PVP (polyvinylpyrrolidone) membrane. After transfer, the membrane was peeled off the gel carefully and washed in transfer buffer and dried overnight, by placing between two pieces of semi-wet filter paper.

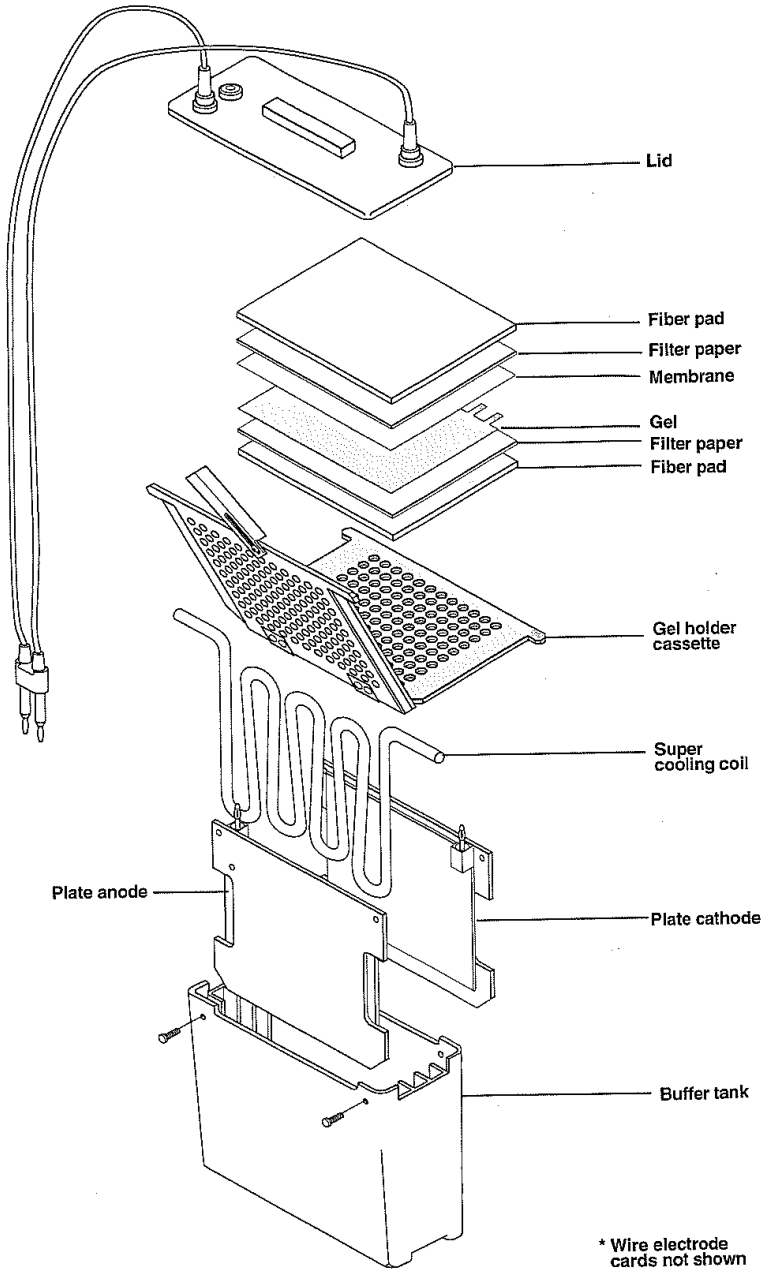


Figure 2.5: Trans-blot cell description and assembly of parts

Blocking

1xTBS (Tris buffered saline) was first prepared from 10xTBS (BIORAD) in MilliQ water. TBS-T (TBS-Tween) buffer containing 0.1% Tween-20 (Sigma) was then prepared from 1xTBS buffer. Blocking buffer was made by adding 5% w/v dried skimmed milk powder (Marvel) in TBS-T buffer and stored at 4^{o}C. The overnight dried membrane was briefly soaked in methanol and then washed in distilled water for 2 minutes. Following the wash, it was soaked

in 50ml of 1x TBS for 15 minutes and then incubated in a tray containing 30ml of 5% blocking buffer for 3-4 hours at RT on an orbital shaker.

Primary antibody binding

ALADIN polyclonal antibody, Q9NRG9 (Proteintech) was diluted 1:500 in blocking buffer from a stock concentration of ~0.5mg/ml. The membrane was then folded in a 50-ml falcon tube, and incubated in 5ml of the primary antibody solution overnight at 4^{o}C (in a cold room) on a rolling shaker. Following the incubation, the membrane was first briefly rinsed with two changes of TBS-T wash buffer. The membrane was then washed 4-6 times in approximately 25ml of TBS-T for 5-10 minutes each at RT on a rolling shaker.

Secondary antibody binding

The secondary anti-rabbit IgG antibody (Amersham Biosciences) was diluted 1:20,000 in 10ml of 5% blocking buffer and the membrane was then incubated in this antibody solution in a falcon tube for around an hour at RT on a rolling shaker. Following this secondary antibody binding, the membrane was briefly rinsed with two changes of TBS-T wash buffer, and then 4-6 times for 5-10 minutes each at RT on a rolling shaker.

Detection

Chemiluminescence detection reagent (Amersham Biosciences) was prepared by adding an equal volume of solution A, luminol enhancer to that of solution B, peroxide. The membrane was incubated in 4ml of this solution at RT for 5 minutes on a rolling shaker, then wrapped in cling film and exposed onto a chemiluminescent detection film (Roche) inside an X-ray cassette (Amersham Biosciences). The membrane was then soaked in TBS-T and stored at 4^{o}C until further use.

GAPDH loading control

The membrane was incubated in blocking buffer in a 50-ml falcon tube at RT for 2-3 hours in a rolling shaker, and the blocking buffer was then discarded. Monoclonal GAPDH-71.1 mouse primary antibody (Sigma-Aldrich; 1 mg/ml) was diluted 1:4000 in 5ml blocking buffer and the membrane was incubated in it overnight at 4^{o}C. On the following day, the membrane was briefly rinsed with two changes of TBS-T wash buffer, and then 4-6 times for 5-10 minutes each at RT on a rolling shaker.

Secondary anti-mouse IgG antibody (Amersham) was diluted 1:20,000 in 10ml of blocking buffer and the membrane was then incubated in this antibody in a falcon tube for at least an hour at RT on a rolling shaker. Following this incubation, the membrane was briefly rinsed with two changes of TBS-T wash buffer, and then 4-6 times for 5-10 minutes each at RT on a rolling shaker. The membrane was incubated in chemiluminescence detection reagents and exposed in an X-ray cassette as described in section 'Detection' above.

Tissue culture techniques

In general, gene targeting in human iPS cells was carried out using two strategies – plasmid-based and protein-based nuclease-assisted strategies. In the plasmid-based strategy, the cells were co-transfected with CRISPR gRNA expression plasmids, Cas9 nuclease expression plasmid, and targeting or donor vector plasmids. In the protein-based strategy, cells were transfected with pre-assembled Cas9-RNP and targeting plasmids or ssODN, where applicable.

Plasmid-based nuclease-assisted gene targeting of human iPS cells

Pre-nucleofection –

Plasmid preparation: The amounts of maxiprep plasmids required for nucleofection were first determined, considering the maximum amount that can be used in nucleofection to avoid significant cell death, is 12µg. 4µg of Cas9 nuclease expression plasmid (Addgene) 3µg of each of the gRNA expression plasmids, and 2µg of targeting vector plasmid (where applicable) were precipitated. NaCl was added to the plasmid mixture to a final concentration of 100mM in 100µl. 250µl of 100% ethanol was then added and incubated on ice for 5 minutes followed by centrifugation at top speed for 10 minutes. The supernatant was then carefully discarded in a tissue culture hood without disturbing the pellet. 1ml of 70% ethanol was then added to the pellet, and after spinning for 5 minutes, it was discarded; this step was repeated once more, and then the pellet was dried for 20 minutes. Finally, the semi-dried pellet was resuspended overnight in DPBS at RT at a concentration of 1µg/µl to be used subsequently for nucleofection.

Culture plates were coated with Synthemax II-SC substrate (Corning) diluted in sterile distilled water (Life Technologies) at a final concentration of 25µg/ml for 2 hours at RT. Accutase (Millipore) and DPBS were warmed to 37^{o}C in a water bath.

Nucleofection – Human iPS cells (KOLF2_C1) were grown to an almost confluent 10cm petri dish. The cells were washed with 10ml pre-warmed DPBS, and then incubated in 5ml pre-

warmed Accutase for 8-10 minutes in 37ºC/5%CO_2 to detach the cells thus obtain a single cell suspension. Following the incubation, Accutase was aspirated and 5ml of complete TeSR-E8 medium (Stem Cell Technologies) supplemented with 10µM ROCK inhibitor (Ri) (Stem Cell Technologies) was added to the plate; ROCK inhibitor enables survival of cells in single cell suspensions by preventing cell migration and dissociation-induced apoptosis. The cells were detached using a cell lifter, dissociated into a single cells by pipetting up and down 3-4 times, and collected in a universal tube (Sterilin). The cells were counted in a haemocytometer and the cell density was determined. The appropriate volume of cell suspension containing 3.5 million cells, to be used for nucleofection, was then centrifuged at 1000rpm for 3 minutes at RT to pellet the cells. Subsequently, the pellet was resuspended in 100µl of AMAXA Nucleofector Solution 2 (Solution 2 to supplement ratio of 4.5:1) (Lonza). The suspension was then mixed with plasmids, and promptly transferred into an AMAXA nucleofection cuvette (Lonza) and electroporated in the AMAXA Nucleofector 2B device (program B-016; Lonza). Using an AMAXA Pasteur pipette (Lonza), approximately 500µl of pre-equilibrated TeSR-E8 media was added to the cuvette, and finally the cell/DNA mixture was transferred immediately into a well of 6-well plate containing 3ml of pre-equilibrated TeSR-E8 media supplemented with 10µM Ri. The cells were cultured in 37ºC/5%CO_2 incubator.

Post-nucleofection – After the cells had been in ROCK inhibitor for 24-hours, the media was changed to TeSR-E8 only until the well was confluent for subsequent experiments. In case of drug selection (puromycin, as used in this study), the media was supplemented with puromycin on the 3rd day after nucleofection. In a titration experiment, 0.3µg/ml was the concentration of puromycin required to kill wild-type cells (Koutsourakis & Skarnes, unpublished data).

Protein-based nuclease-assisted gene targeting of hiPS cells

Pre-nucleofection – Pre-assembly of the annealed cr:tracrRNA to Cas9 nuclease was done in two steps:

I. Pre-annealing: crRNA and tracrRNA were added together in a 1:1 molar ratio and concentration of 2µg/µl. The mixture was then heated at 58ºC for 10 minutes and allowed to cool slowly to RT.
II. Pre-assembly: 8µg of the pre-annealed cr:tracrRNA was added with 12µg of Cas9 nuclease (Feldan) and incubated at RT for at least 15 minutes. Following the incubation, 500pmoles of ssODN or 4µg of supercoiled targeting vector plasmid,

resuspended in 2µl DPBS, was added to the pre-assembled Cas9 ribonucleoprotein complex (Cas9-RNP) and used for subsequent nucleofection.

Besides, culture plates were coated with Synthemax II-SC substrate at a final concentration of 25µg/ml for 2 hours at RT.

Nucleofection – KOLF2_C1 cell line of hiPS cells were grown to an almost confluent 10-cm dish (Thermoscientific). The cells to be used for protein nucleofection were treated in the same manner as those in the plasmid-based nucleofection described above (refer to section for plasmid-based nucleofection). However, $8x10^5$ cells were electroporated for one nucleofection which were spun at 1000rpm for 3 minutes at RT. The cell pellet was resuspended in 100µl of AMAXA Nucleofector Solution 3 (Solution 3 to supplement ratio of 4.5:1) (Lonza) and mixed with the pre-assembled Cas9-RNP and ssODN or targeting donor plasmid. The mixture was then transferred into an AMAXA nucleofection cuvette (Lonza) and electroporated using the AMAXA Nucleofector 4D device (Lonza) (program CA-137). Subsequently, using an AMAXA Pasteur pipette the transfected cells were transferred from the cuvette to a 6-well plate, pre-coated with Synthemax, containing 3ml pre-equilibrated TeSR-E8 media supplemented with 10µM Ri (TeSR-E8+Ri). The cells were grown in $37^oC/5\%CO_2$ incubator.

Post-nucleofection – The cells were fed with TeSR-E8 only, on the next day, until confluent or ready for seeding and successive experiments.

Culture and passaging of human iPS cells

The cells in a 10-cm dish or 6-well plate were first washed with DPBS (10ml for 10-cm dish; 3ml for 6-well plate) and then incubated in ReLeSR (Stem Cell Technologies), an enzyme-free reagent for gentle dissociation of cell aggregates, for 2 minutes at RT; the volume of ReLeSR used is generally 5ml in 10-cm dish and 1.5ml in a well of 6-well plate. Aspirating ReLeSR, the plate was incubated again for another 2-3 minutes and TeSR-E8 media was then added. With a cell lifter, the colonies were detached from the surface of the dish and broken into small cell clumps by triturating 2-3 times. The cell clumps were plated at the appropriate cell density onto a 10-cm dish or 6-well plate, pre-coated with 25µg/ml Synthemax II-SC substrate, and containing TeSR-E8 media. The cells were then incubated at $37^oC/5\%CO_2$, and fed until they were ready for successive experiments.

Colony picking of human iPS cells

I. Seeding single cells

The cells on a 6-well plate were first washed with 3ml of DPBS after which 1.5ml of Accutase, pre-warmed at 37ºC, was added into the well. Following incubation for 10 minutes in 37ºC/ 5% CO_2 incubator, Accutase was aspirated and the cells were collected in 3ml of TeSR-E8 medium with 10µM Ri (TeSR-E8+Ri) in a universal bottle; at this point, the cells were dissociated into a single cell suspension by triturating. The cells were then counted under a microscope, and 1500-3000 single cells were plated onto a Synthemax coated 10-cm dish containing 10ml TeSR-E8+Ri, and conditioned media to $1/10^{th}$ of the total volume. Conditioned media (CM) was collected from near-confluent cultures of KOLF2-C1 cells grown overnight and then filter sterilized. CM contains unknown secreted factors that enhance the growth and survival of iPS cells in single cell suspension (Skarnes, unpublished observation). The cells were then grown in Ri supplemented media for two days. From Day 3 onwards, the cells were fed everyday with 10ml of fresh TeSR-E8 media only, until they were ready for picking, having diameters of approximately 1mm.

II. Colony picking

For a set of 96 colonies to be picked, two flat-bottom 96-well plates were treated: one with Synthemax II-SC substrate to be used for genotyping and another with Matrigel hESC-qualified matrix (BD biosciences) for archiving the clones. 1mg/ml Synthemax was diluted 1:40 in sterile distilled water and 50µl was used to coat each well for 2 hours at RT. Matrigel was thawed on ice and diluted to a factor of approximately 1:95 in cold DMEM/F12 (Dulbecco's Modified Eagle's Medium; Life Technologies), a basal medium for iPS cell culture. 50µl of the diluted Matrigel was used to coat each well for at least an hour at RT. Matrigel was aspirated just prior to use, immediately added with 100µl of TeSR-E8+Ri per well to avoid drying out, and incubated at 37ºC/5%CO_2 until cells were plated on them. Besides, a round-bottom 96-well plate was added with 50µl TeSR-E8+Ri per well and kept in 37ºC/5%CO_2 until ready to be used in picking.

The 10-cm dish containing colonies to be picked was washed with 10ml DPBS, and 5ml of ReLeSR was added. After 5-minute incubation at RT, ReLeSR was aspirated and 10ml TeSR-E8+Ri was added to the plate. Colonies were picked in 50µl media under a stereoscope, and then transferred into the round-bottomed well plate containing 50µl media with 10µM Ri.

Following picking the required number of colonies, they were broken into smaller clumps by triturating extensively and then transferred into the Matrigel coated flat-bottomed 96-well plate containing 100µl TESR-E8+Ri per well. The cells were mixed and then 100µl media with cells was transferred into the Synthemax coated flat-bottomed 96-well plate. The cells were then fed with 200µl TeSR-E8 only until the plates were 80% confluent, and ready for archiving and genotyping. A schematic diagram for picking into each well of a 96-well plate is as follows:

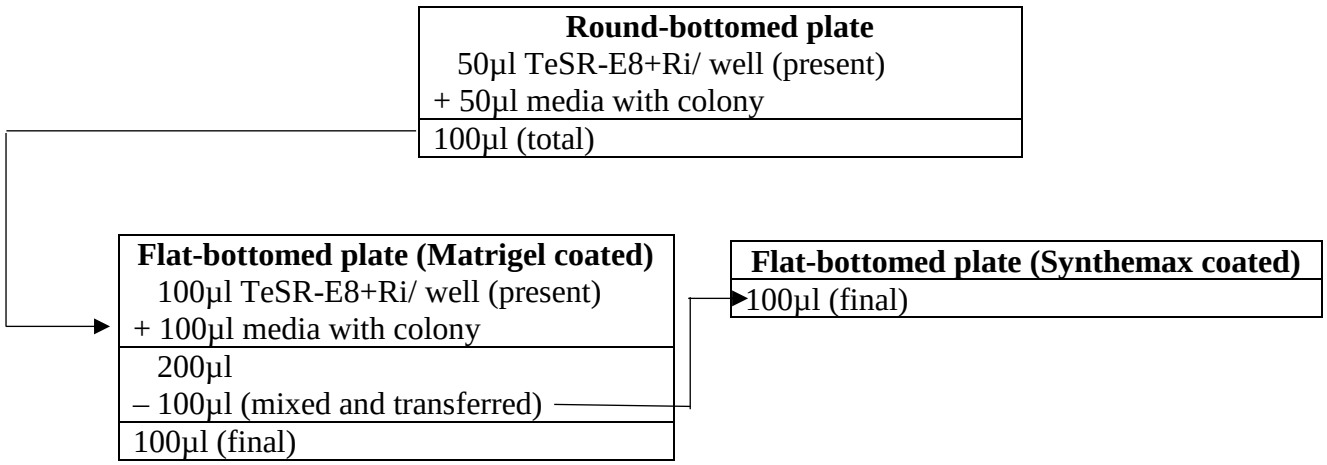


Figure 2.6: Schematic representation of a well on 96-well plate for creating plate replicas for genotyping and archiving

Colony archiving of human iPS cells into a 96-well matrix plate

The cells in the wells of a 96-well plate were first washed with 200µl of DPBS per well. 25µl of Accutase, pre-warmed at 37^{o}C, was added per well and incubated in 37^{o}C/ 5% CO_2 incubator for 10 minutes. Without aspirating Accutase, 175µl of knockout serum replacement (KSR) (Gibco Life Technologies) containing 10% dimethyl sulfoxide (DMSO) (Sigma Aldrich) and pre-sterilized through a 0.2µm filter (Whatman), was added per well; DMSO is a cryoprotectant vitrification agent which preserves cells protecting them from mechanical injury. The cell mixture was then triturated extensively to obtain a single-cell suspension. The total volume of 200µl cell suspension was then transferred into the vials of a 96-well matrix plate (Thermoscientific) and overlaid with 100µl of mineral oil (Sigma Aldrich), pre-sterilized using a 0.2µm filter unit (Corning); the purpose of mineral oil is to prevent evaporation of medium. The plate was stored in -70^{o}C freezer and then transferred into liquid nitrogen storage after 1-2 days.

Expanding hiPS cells from 96-well matrix plate cryovials into 6-well plate

The mini cryovial containing the frozen hiPS cells was thawed in hand and the contents were transferred into a universal tube containing 5ml TeSR-E8+Ri. The cells were pelleted at 1000rpm for 3 minutes and then the pellet gently resuspended in another 4ml of TeSR-E8+Ri. The cells were plated onto a Synthemax-treated well of a 6-well plate and incubated in 37^{o}C/ 5% CO_2 incubator. They were fed with TeSR-E8 media until ready for subsequent experiments.

Archiving human iPS cells into 1ml cryotubes

The cells in a well of 6-well plate were first washed with 3ml DPBS, followed by incubation in 1.5ml pre-warmed Accutase for 10 minutes at 37^{o}C/ 5% CO_2. Aspirating the Accutase, the cells were collected in 3ml of TeSR-E8 media and then pelleted at 1000rpm for 3 minutes. The pellet was resuspended in 3-4ml of KSR containing 10% DMSO pre-sterilized using a 0.2µm filter. They were stored in 1ml cryotubes (Thermoscientific), in 0.5ml aliquots, at -70^{o}C freezer and transferred to liquid nitrogen after 1-2 days for long term storage.

Expanding hiPS cells from 1ml cryotubes into 6-well plate

The cryotube was first warmed up in hand and the thawed contents were transferred into a universal tube containing 5ml TeSR-E8+Ri. The cells were centrifuged at 1000rpm for 3 minutes after which the pellet was resuspended in 4ml TeSR-E8+Ri by triturating gently. The cells were then transferred into a well of 6-well plate pre-coated with Synthemax and cultured for subsequent experiments.

Subcloning mixed clones

Mixed clones containing cells with more than one genotype were subcloned. In this process, 1500 cells were primarily seeded on a 10-cm dish (refer to section describing seeding) and grown until they were ready for picking. The cell density was kept low in this procedure so as to obtain pure and well-separated colonies. Generally, 16 colonies were picked (refer to section describing picking) and used for subsequent processes of archiving and genotyping.

Chapter 3

Triple A syndrome

3.1 Knockout of *AAAS* exon 2

To knockout the function of *AAAS*, the biallelic targeting strategy was to replace one copy of exon 2 with a puromycin (puro) selection cassette by homologous recombination (HR) and simultaneously damage the other allele by non-homologous end-joining (NHEJ; Figure 3.1.1).

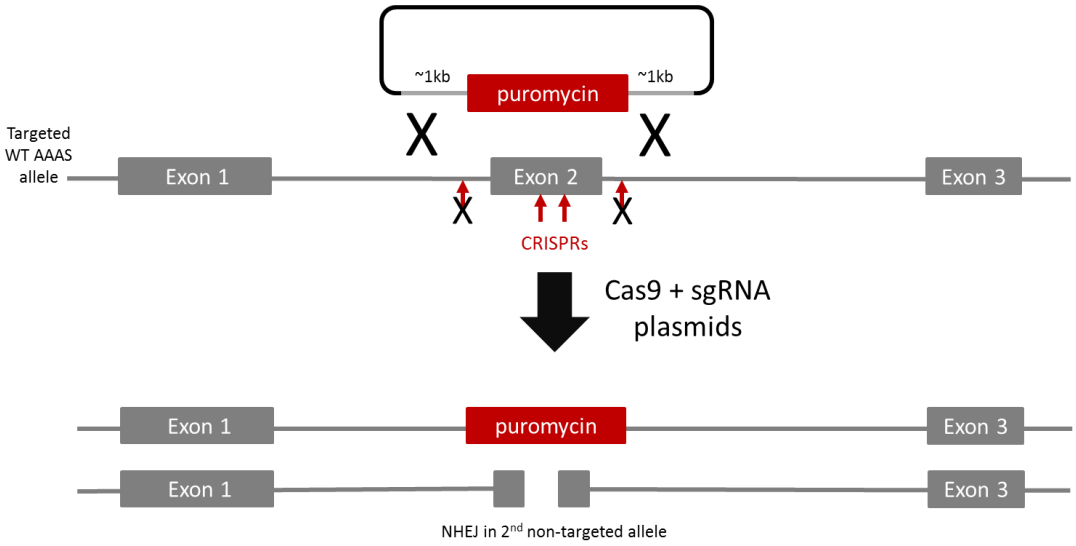


Figure 3.1.1: **CRISPR-Cas9 biallelic gene targeting strategy for AAAS**. One copy of exon 2 of *AAAS* is deleted by CRISPR-Cas9 assisted homologous recombination with a puro-resistant plasmid donor. NHEJ damage to the second non-targeted allele was identified in puro-resistant clones by PCR amplification and sequencing of *AAAS* exon 2. Arrows indicate functional CRISPR positions; crossed arrows indicate flanking CRISPRs which did not yield viable cells. "X" refers to homologous recombination.

The knockout of *AAAS* exon 2 (AAAS2) was a plasmid-based nuclease-assisted gene knockout in KOLF2-C1 hiPS cells. Cells were co-transfected with CRISPR gRNA expression plasmids, Cas9 nuclease, and targeting vector plasmids. Two experiments were conducted: one using two CRISPRs flanking exon 2 and the second using two CRISPRs lying within the exon. The targeting vector contained the EF1α-puro-polyA cassette for introduction into the alleles by homologous recombination, thus knocking out the gene, enabling for the selection of the puro resistant knockout clones. On the 3rd day after nucleofection, culture media supplemented with 0.5µg/ml puromycin was used to select for the puro-resistant colonies. The EF1a promoter, highly active in most human cells, was found to be optimal for gene editing experiments in human iPS cells (Koutsourakis & Skarnes, unpublished data). The three-day delay in supplementing the culture media with puromycin is to allow the cells enough time for expression of Cas9 protein and gRNAs and eventually for homologous recombination to occur

between genomic DNA and targeting vector, allowing expression of puro-resistance in the cells.

The targeting donor vector used for biallelic knockout of *AAAS* exon 2 was constructed in a step-wise process (Figure 3.1.2). A region of genomic DNA containing *AAAS* exon 2, with approximately 1kb homology arms on either sides of the exon, was first cloned into pUC19_RV plasmid by InFusion cloning forming pAAAS_exon 2 clone. The next step was replacing *AAAS* exon 2 of the pAAAS_exon 2 clone with Zeo/PheS (Z/P) Gateway cassette (Skarnes et al. 2011), by recombineering followed by the introduction of a drug selection cassette, pL1L2 EF1α-puro-polyA (Koutsourakis & Skarnes, unpublished), by a Gateway exchange reaction.

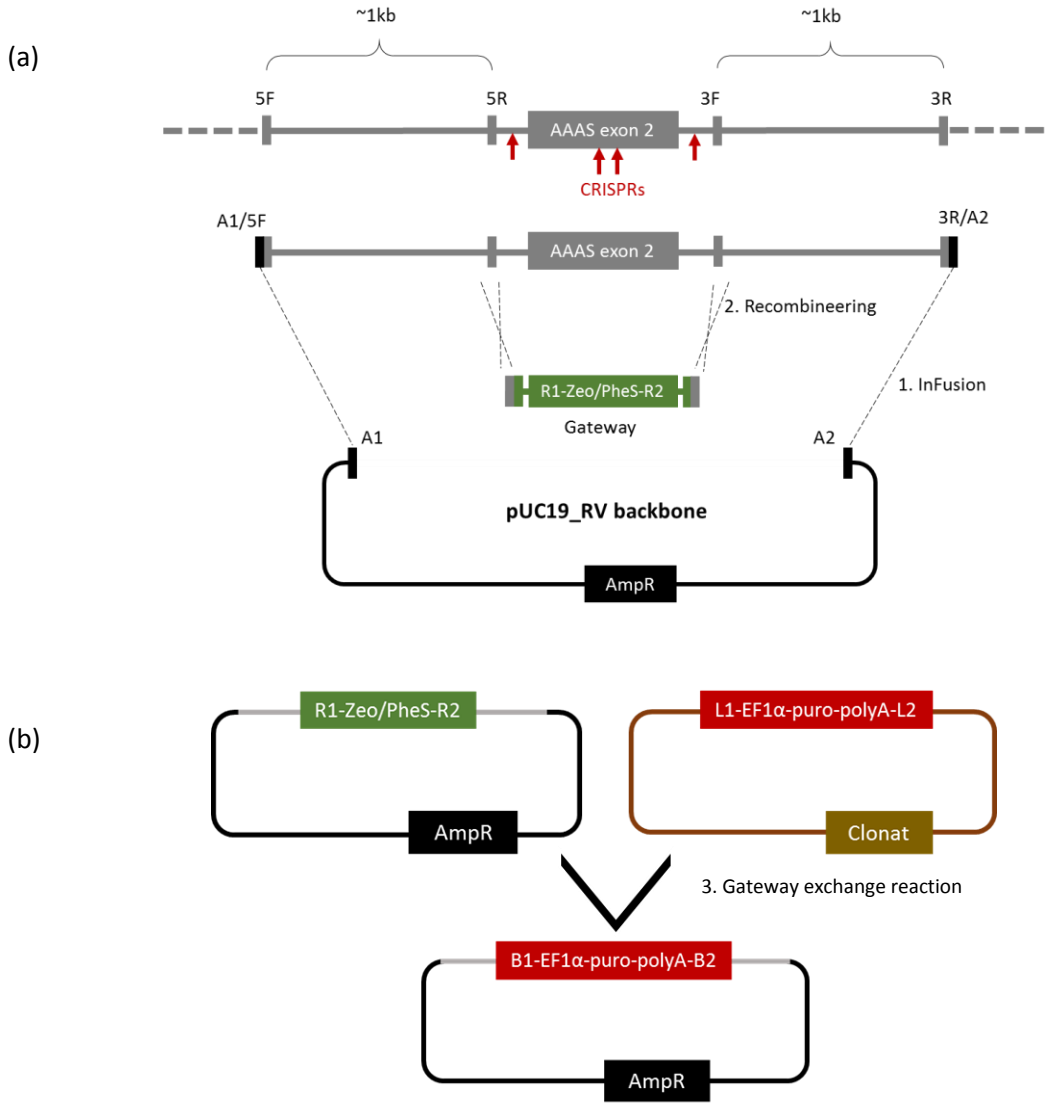


Figure 3.1.2: **Construction of targeting donor vector for knockout of *AAAS* exon 2.**
(a) Introduction of Z/P Gateway cassette into pAAAS_exon 2 by recombineering. Arrows indicate CRISPR positions. 5F and 5R are forward and reverse primers, respectively, on the

5’end of the targeted exon; 3F and 3R are on the 3’end of the exon. 5F and 3R are InFusion PCR primers for amplification of AAAS exon 2. (b) Introduction of drug selection cassette, pL1L2 EF1α-puro-polyA, into Z/P recombineered clone by LRII clonase Gateway exchange reaction.

Many puro-resistant colonies were recovered using only the CRISPRs within the exon. In contrast, only a few colonies were recovered using CRISPRs flanking exon 2. The reason for this discrepancy could be due to poor Cas9 cleavage at the intronic CRISPR sites which could not introduce the puro-cassette at the target site leading to cell death in puro-supplemented media. It is, however, also possible that this experiment failed due to other technical problems.

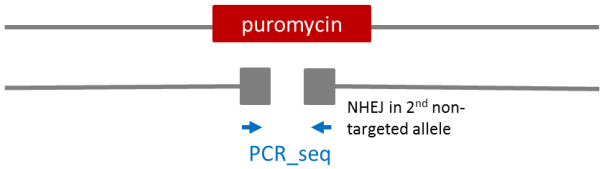
Sequencing of non-targeted allele

Analysis of 96 colonies by PCR amplification and sequencing of the non-targeted allele demonstrated that different indels were introduced by NHEJ in this allele in a small number of colonies (Table 3.1.1). Three of the clones contained frameshift mutations of which two contained deletions of 16bp and 11bp, and one had an insertion of 1bp (Figure 3.1.3). Therefore, the biallelic targeting efficiency defined as the fraction of clones with two inactive alleles was 3%. In addition, four other clones contained in-frame NHEJ indels (Figure 3.1.4). In 22 clones, overlapping sequence traces downstream of DSB site were also observed indicating they were mixed or mosaic clones. In theory, mosaic clones can be explained by cleavage of the non-targeted allele after cell division, resulting in the formation of a colony with a mixture of alleles. Almost half of the clones (52%) were wild-type containing no mutation. These clones could have suffered a DSB in their DNA, but were repaired correctly. Finally, the results suggest that the CRISPR1 exonic site was cleaved more efficiently than the CRISPR2 site as all of the indels overlapped the CRISPR1 site (Figure 3.1.3 (b)).

Table 3.1.1: Summary of the genotypes different clones obtained upon biallelic targeting of *AAAS* exon 2

No of colonies screened	**NHEJ Indel**		**Mosaic**	**Wild-type**	**No/ poor quality sequence**
	frameshift	in-frame			
96	3 (3%)	4 (4%)	22 (23%)	50 (52%)	17 (18%)

(a)



(b)

	CRISPR 1	CRISPR 2	
TCCTACAACTGACAAAGG	**ATCCCCTAAAGACCCC.TGGAAGG**	CTGGA**CCATGGCACAAGAACTGCCTTCA**	TCCATCACCGGGAGCAAGTGTGGAAGA	**WT**
TCCTACAACTGACAAAGGATCCCCTAAAGA----.	----------	CCATGGCACAAGAACTGCCTTCATCCATCACCGGGAGCAAGTGTGGAAGA	-16bp
TCCTACAACTGACAAAGGATCCCCTAAA------.	-----GGCTGGACCATGGCACAAGAACTGCCTTCATCCATCACCGGGAGCAAGTGTGGAAGA	-11bp
TCCTACAACTGACAAAGGATCCCCTAAAGACCCC**T**	TGGAAGGCTGGACCATGGCACAAGAACTGCCTTCATCCATCACCGGGAGCAAGTGTGGAAGA	+1bp

(c)

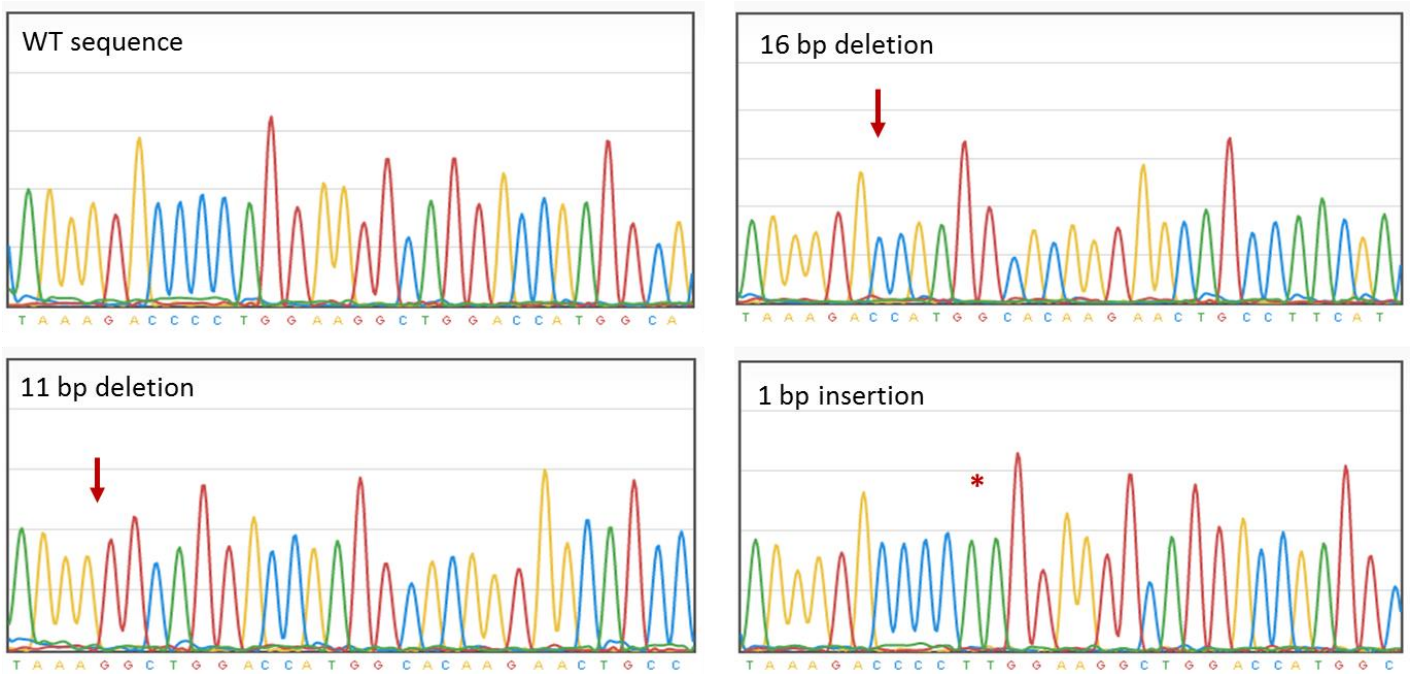


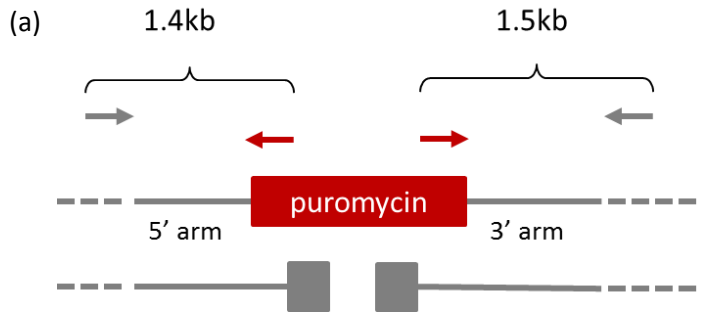
Figure 3.1.3: **Screening for NHEJ indels in non-targeted allele of *AAAS* gene.** (a) *AAAS* exon 2 showing PCR sequencing across the damaged region of non-targeted allele (sequencing primers indicated with blue arrows) (b) DNA sequences of wild-type *AAAS* exon 2 and three frameshift mutations recovered in the screen. CRISPR sites are indicated in red and PAM sites are underlined. Deleted sequences (16bp & 11bp) are indicated with dashes and inserted sequences (single base pair) in green. (c) Sequence traces of the wild-type and mutated alleles. Arrows indicate breakpoints at which deletion occurred and asterisk (*) indicates insertion of a base (T).

Figure 3.1.4: **DNA sequence alignment of in-frame NHEJ indels in non-targeted allele of *AAAS* exon 2.** CRISPR sites are indicated in red and PAM sites are underlined. Deleted sequences are indicated with dashes.

Genotyping of puro-targeted allele

The three clones containing frameshift indels in the non-targeted allele were genotyped to confirm if the puro-cassette was correctly targeted by homologous recombination. PCR amplification of the homology arms was done with gene-specific primers and puro-cassette-specific primers (Figure 3.1.5 (a); refer to Appendix, Table A8 for sequences). The gene-specific primers were designed outside of the homology arms to avoid the detection of random insertions of the plasmid donor. All three clones showed the correct-sized PCR products for both the 5' and 3' homology arms (Figure 3.1.5 (b) and (c)), suggesting that the puro-cassette was correctly inserted in each of these clones.

Two putative heterozygous clones were also chosen from the archived clones as controls. They were puro-resistant clones in which the non-targeted allele showed a wild-type sequence. However, only one of the heterozygous clones (Het1) had a correctly targeted puro cassette giving bands of expected sizes for both the 5' and 3' arms. The other clone (Het2) did not produce the expected bands and instead showed non-specific bands or smears similar to the wild-type negative control (Figure 3.1.5 (b) and (c)). This indicates that the donor plasmid was not correctly targeted but rather integrated elsewhere in the genome of this clone.



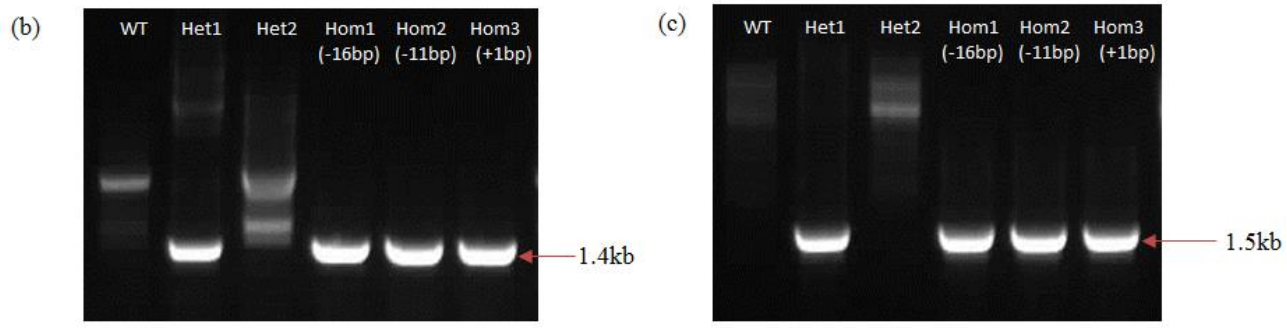


Figure 3.1.5: **Genotyping of puro-targeted allele.** (a) Schematic representation of the targeted allele showing primers for amplification of the homology arms. Red arrows indicate puro-cassette specific PCR primers, grey arrows indicate gene-specific primers outside the homology regions. Each of the homology arms are indicated along with their sizes. (b) and (c) Gel electrophoresis of PCR amplicons of (b) 5' homology arm [1.4kb] and (c) 3' homology arm [1.5kb] of the puro cassette.

3.2 *AAAS* exon 1 point mutation

This project of generating hiPS disease models for the point mutation in *AAAS* exon 1 was extended from the study by Krumbholz and colleagues (2006). Their findings (Figure 3.2.1) suggest that a single base substitution of 43C>A in exon 1 of *AAAS* gene led to the introduction of a novel splice donor site. Aberrant splicing at this site results in an exon frameshift in the gene, leading to a premature stop codon downstream. This frameshift mutation eventually resulted in a truncated ALADIN protein causing the disease manifestations as observed in patients.

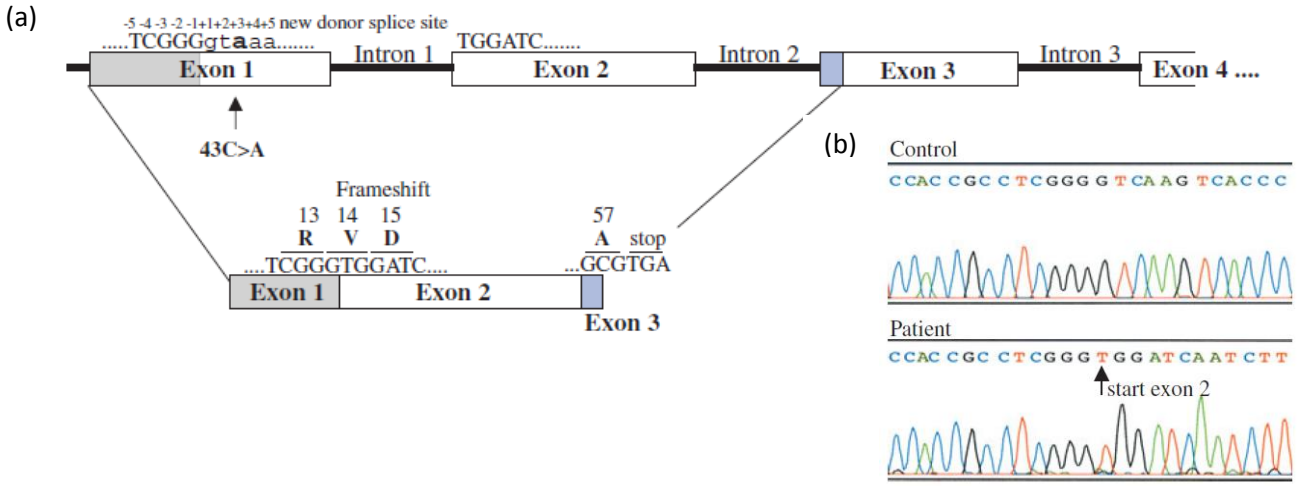


Figure 3.2.1: **Homozygous point mutation of 43C>A in *AAAS* exon 1 results in a novel splice donor site.** (a) Splicing due to the 43C>A mutation in exon 1 leads to a frameshift mutation, hence a truncated protein. (b) Electropherogram traces showing the sequences across *AAAS* exon 1 of a healthy control and that of a patient with aberrant splicing and start of exon 2 as indicated by arrow (adapted from Figure 2; Krumbholz et al. 2006)

For this part of the study, KOLF2_C1 cells were co-transfected with preassembled Cas9-RNP and ssODN (101 bases; refer to Appendix, Table A6) containing the C>A point mutation for its introduction into genomic DNA by homology-directed repair (HDR). The crRNA, tracrRNA, and Cas9 used for the Cas9-RNP preassembly were all purchased commercially. The ssODN, also bought commercially, was symmetric containing 50-base homology arms on either side of the C>A point mutation. The ssODN was non-complementary to the CRISPR gRNA oligo to prevent it from annealing to the gRNA. Two experiments were conducted, each on $8x10^5$ cells, with Cas9 protein from two commercial sources. The first experiment using 12µg of recombinant Cas9 protein (Feldan) showed decreased activity in the pool of transfected cells when compared to the second experiment done with 18µg of Cas9 (Thermofisher) based on the T7E1 assay (Figure 3.2.2).

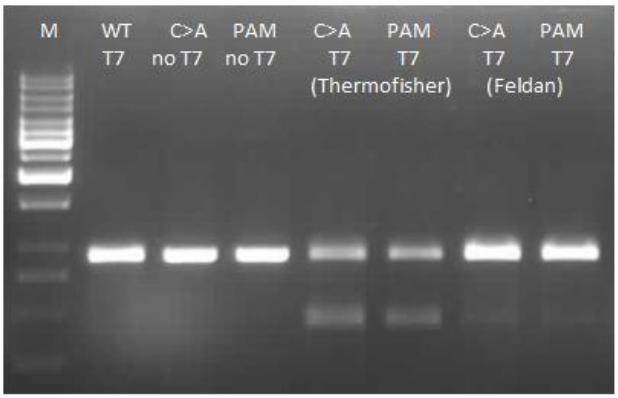
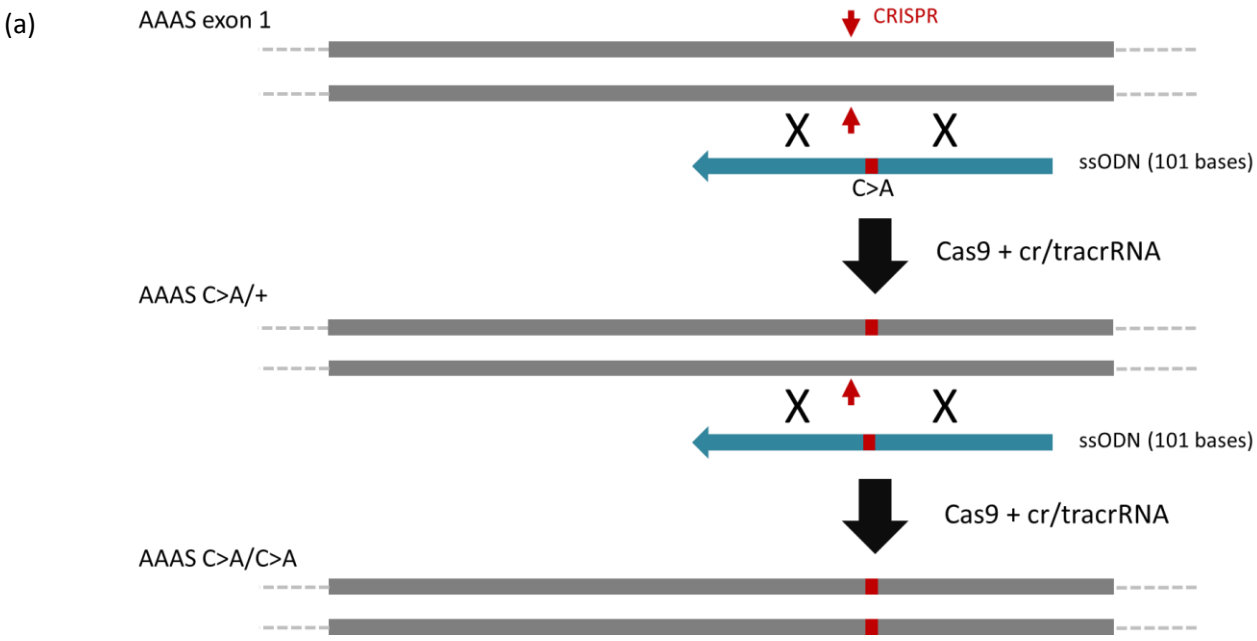


Figure 3.2.2: **T7E1 assay demonstrating Cas9 activity in a pool of transfected cells**. WT-T7: wild-type cells treated with T7E1 indicate no Cas9 activity. C>A: cells transfected with ssODN containing only the C>A point mutation; PAM: cells transfected with the control ssODN containing both C>A and PAM mutations. C>A-no T7 and PAM-no T7 are negative controls with no T7E1 treatment. C>A-T7 and PAM-T7 labelled as Thermofisher or Feldan are cells treated with T7E1 that had been transfected with Cas9 from respective manufacturers.

However, PCR genotyping of clones suggested that lesser amount of Cas9 was more efficient in producing the desired specific point mutations in the genome by HDR. The initial transfection of wild-type cells with Cas9-RNP and ssODN resulted in heterozygotes only, but no homozygotes. Thus, in order to obtain homozygotes, the experiment was repeated by re-transfecting the heterozygotes (Figure 3.2.3).



(b)

```
----CCGGCAAGATGTGCTCTCTGGGGTTGTTCCCTCCTCCACCGCCTCGGGGTCAAGTCACCCTATATGAGCACAATAACGAGCTGGTGACGGGCAGTAGCTAT----  WT
    :::::::::::::::::::::::::::::::::::::::::::::::::: :::::::::::::::::::::::::::::::::::::::::::::::::::
3'- GGCCGTTCTACACGAGAGACCCCAACAAGGGAGGAGGTGGCGGAGCCCCATTTCAGTGGGATATACTCGTGTTATTGCTCGACCACTGCCCGTCATCGATA -5' ssODN

----CCGGCAAGATGTGCTCTCTGGGGTTGTTCCCTCCTCCACCGCCTCGGGGTAAAGTCACCCTATATGAGCACAATAACGAGCTGGTGACGGGCAGTAGCTAT---- C>A
```

Figure 3.2.3: **Generation of the C>A point mutation in exon 1 of *AAAS* gene.** (a) CRISPR-Cas9 assisted gene targeting strategy. Cells were co-transfected with Cas9, crRNA, tracrRNA, and a ssODN containing the C>A mutation to recover clones heterozygous for the C>A mutation (C>A/+). The experiment was repeated in heterozygous (C>A/+) cells to recover the mutation in both alleles (C>A/C>A). "X" refers to homologous recombination; C>A mutation marked red; CRISPR positions indicated with red arrows (b) Sequence of wild-type *AAAS* exon 1 showing the CRISPR site in red (PAM sequence is underlined), the ssODN containing the C>A mutation (antisense strand; G>T change in green), and the C>A mutation (sense strand; C>A change in green) in mutated allele.

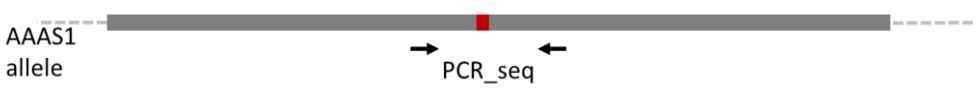
Generation of heterozygotes:

Analysis of 48 colonies by PCR amplification and sequencing revealed that 3 clones contained heterozygous C>A point mutation at the target site (Figure 3.2.4). NHEJ indels but with no C>A mutation was observed in about 14 clones, and almost half of the clones (46%) were wild-type (Table 3.2.1) which could have suffered DNA damage but were repaired correctly. These results support that NHEJ indels are the more common events in cellular DNA repair, while HDR is rare.

Table 3.2.1: Summary of the genotypes of different clones obtained in the first round of transfection of wild-type hiPS cells with the two different amounts of Cas9

No. of colonies screened after 1st transfection	**Amount of Cas9 protein; manufacturer**	**Heterozygote (C>A/+)**	**Homozygote (C>A/C>A)**	**NHEJ indels (no C>A mutation)**	**Wild-type (+/+)**	**No/ poor quality sequence**
48	12µg; Feldan	3 (6%)	0 (0%)	14 (29%)	22 (46%)	9 (19%)
96	18µg; Thermofisher	0 (0%)	0 (0%)	59 (62%)	30 (31%)	7 (7%)

(a)



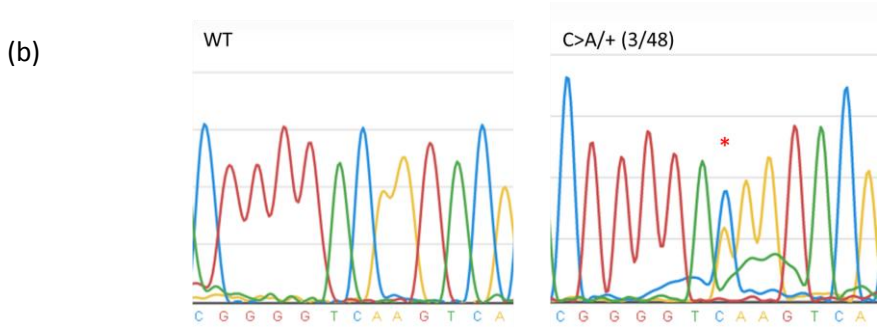


Figure 3.2.4: **Screening for the C>A point mutation in *AAAS* exon 1 after the first round of transfection.** (a) *AAAS* exon 1 allele showing PCR sequencing across the target C>A point mutation (sequencing primers indicated with black arrows). (b) Sequence traces of PCR products from representative wild-type (WT) and heterozygous (C>A/+) cell lines. The position of the C>A mutation is marked with a red asterisk.

On the other hand, clones from the other experiment in which cells were transfected with greater amount (18µg) of Cas9 nuclease were also analyzed. Screening 96 colonies revealed that none harbored the desired C>A point mutation in *AAAS* exon 1 either in the homozygous or heterozygous state (Table 3.2.1). Many of the clones (62%) had multiple sequence traces downstream of the DSB site indicative of NHEJ indels in either or both alleles. (Figure 3.2.5). Besides several colonies were wild-type containing no mutations. These results suggested that higher amounts of Cas9 led to more frequent NHEJ indels and less frequent HDR.

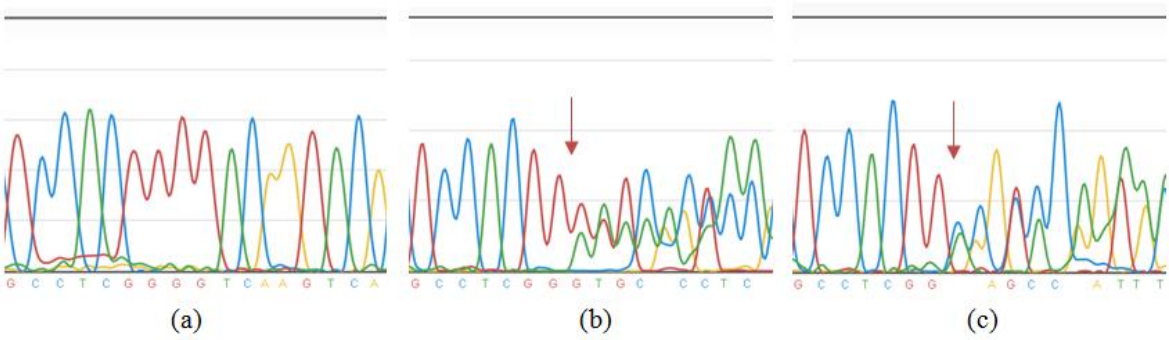


Figure 3.2.5: **Screening of *AAAS* exon 1 around DSB site showing indels.** (a) wild-type sequence, (b) and (c) show indels in one and both alleles respectively, with DSB sites indicated by arrows.

Control experiments with PAM mutations in ssODN:

Control experiments were to be conducted with ssODN containing a silent mutation in the PAM site and the C>A mutation. Mutations in the PAM site are expected to abolish re-cleavage

of the locus following incorporation of the ssODN by HDR. Thus, this experiment was conducted to test if PAM site mutations increased the rate of HDR. In exon 1 of AAAS gene, there were frequent consensus splice donor sites along the sequence, for which a silent mutation, for instance, a substitution mutation from a pyrimidine to a purine base in the PAM was likely to give rise to a splice donor site which would result in unexpected mutations and splicing in the gene. Thus, a missense mutation in the PAM sequence was done instead. This was, however, not going to affect the cell models as it was a control experiment to only test if PAM site mutations increased the rate of HDR. For this part of the experiment, it was observed that, for the reduced amount of Cas9 (12µg), there were less clones containing NHEJ indels (25/96) and more were wild-type (52/96) clones (Table 3.2.2). However, two of the clones contained heterozygous C>A and PAM mutations which agreed with the previous observations of increased HDR frequency using reduced amounts of Cas9. Contrarily, when greater amount of Cas9 was used, there were more NHEJ indel events (41/96), like earlier, however none of the clones contained the control (C>A and PAM) mutations.

Table 3.2.2: Summary of the genotypes of clones obtained upon targeting cells with two different amounts of Cas9 protein using control ssODN containing an additional PAM mutation besides the target C>A point mutation.

No. of colonies screened	**Amount of Cas9 protein; manufacturer**	**Heterozygote (C>A_PAM/+)**	**Homozygote (C>A_PAM/ C>A_PAM)**	**NHEJ indels**	**Wild-type (+/+)**	**No/ poor quality sequence**
96	12µg; Feldan	2	0	25	52	17
96	18µg; Thermofisher	0	0	41	24	31

The results indicate that more NHEJ damage, but less HDR are observed in the presence of higher amounts of Cas9 protein. This can be explained by re-cleavage of the modified locus and secondary damage to the locus. This higher frequency of NHEJ damage was probably demonstrated in the T7E1 assay as a higher activity. On the contrary, less amount of Cas9 protein exhibited decreased activity and frequency of NHEJ indels. This can be explained in that lesser amounts of Cas9 protein and sgRNA are likely to be degraded early in cells, lowering the effective concentration of the Cas9-RNP complex and its duration in cells (Wu et al. 2014). Adding mutations in the PAM unexpectedly did not improve rates of HDR. This could be explained by poorer incorporation of the ssODN oligo when more mismatches to the target

locus are present. The different manufacturers could also explain the difference in the Cas9 activities which can, however, be determined by performing experiments keeping Cas9 amounts and all the other factors, along with amount of sgRNA constant. Besides, the high frequency of NHEJ events can be prevented by inhibiting the DNA ligase IV which plays the major role in ligating Cas9-cleaved DNA strands in the NHEJ repair pathway. Scr7, an inhibitor of the DNA ligase IV can be used to block the NHEJ repair pathway forcing the damaged DNA to use ssODN to repair itself by HDR (Maruyama et al. 2015). This way, by inhibiting NHEJ, efficiency of HDR events can be enhanced eventually increasing the rates of HDR mediated precise genome editing with desired mutations.

Generation of homozygotes:

To obtain homozygotes, a second round of transfection was performed on two of the selected heterozygotes using same conditions as in the first round, that is 12µg of Cas9 (Feldan), 8µg of cr:tracrRNA, and 500pmoles of ssODN. 192 colonies were analyzed, with 96 colonies being screened by PCR and sequencing for each of the two heterozygotes transfected.

A total of 3 homozygotes were obtained out of the 192 colonies screened, which contained the C>A mutation in both the alleles following the second transfection (Figure 3.2.6). NHEJ indels were observed in 49 clones and almost half of the colonies were not modified, that is, they were heterozygotes, as were initially started with. Interestingly, two clones were observed to have reverted back from heterozygote to wild-type with both the alleles having the wild-type base-C in the position of interest. This suggests that gene conversion took place whereby the mutant DNA strand was cleaved again by the Cas9-RNP and repaired by the wild-type allele as the template, thus converting heterozygous mutant cells back to wild-type. However, it is possible that these two clones represent low level contamination by wild-type cells. A considerable number of colonies (47/192) had no or poor quality sequence which also means that the results likely underestimate the efficiency of HDR. All results are tabulated in Table 3.2.3.

Table 3.2.3: Summary of the genotypes of different clones obtained after the second round of transfection of heterozygote clones

No of colonies screened after 2nd transfection of heterozygotes	**Homozygote (C>A/C>A)**	**NHEJ indels**	**No mutation (C>A/+)**	**WT (+/+)**	**No/ poor quality sequence**
192	3 (2%)	49 (25.5%)	91 (47%)	2 (1%)	47 (24.5%)

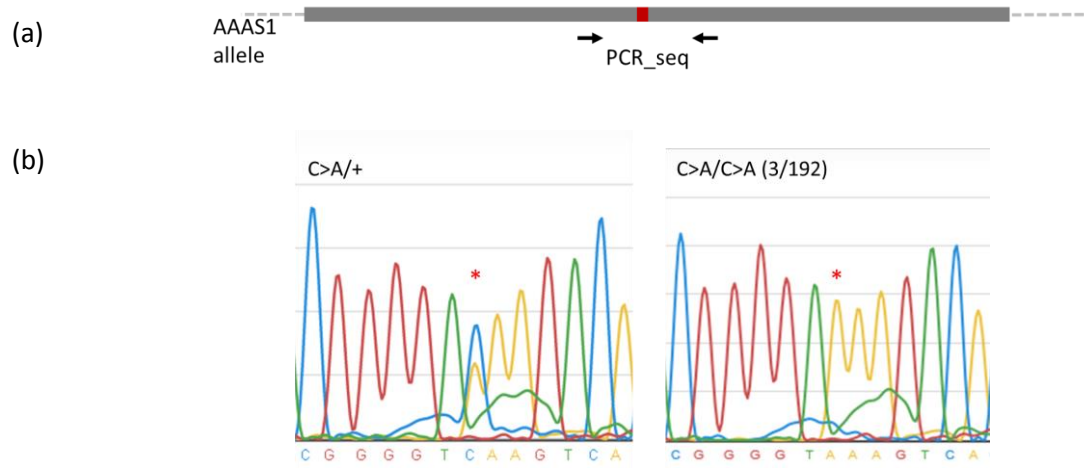


Figure 3.2.6: **Screening for the C>A point mutation in *AAAS* exon 1 after the second round of transfection.** (a) *AAAS* exon 1 allele showing PCR sequencing across the target C>A point mutation (sequencing primers indicated with black arrows). (b) Sequence traces of PCR products from representative heterozygous (C>A/+) and homozygous (C>A/C>A) cell lines. The position of the C>A mutation is marked with a red asterisk.

3.3 Expression of ALADIN protein by *AAAS* mutant clones

For the *AAAS* exon 2 knockout and *AAAS* exon 1 point mutation clones, the expression of ALADIN protein was determined by Western blot. Using rabbit polyclonal anti-AAAS antibodies, ALADIN expression was observed in wild-type and heterozygote clones, but not in the homozygotes. This observation validates that ALADIN expression was completely knocked out in these CRISPR-Cas9 gene-edited homozygotes (Figure 3.3). The anti-AAAS antibody used in this experiment recognizes the epitope present in three of the isoforms of ALADIN – Q9NRG9-V1 (60kDa), Q9NRG9-V2 (56kDa), and F8VZ44 (46kDa). Presumably, the band in the assay is not the isoform F8VZ44 because initiation of translation of this isoform happens in exon 3 and thus it should be expressed regardless of any mutations in exons 1 and 2. However, the protein observed in the assay was knocked out only in the homozygotes; it is possible that the F8VZ44 isoform is not expressed at all in undifferentiated cells, as in this case. Therefore, it can be assumed that the band could correspond to either V1 or V2 of the Q9NRG9 isoform. It can be further characterized as to which of these isoforms is being expressed by performing RT-PCR of the mRNA to its cDNA and validating which alternative transcript is being produced in these iPS cells.

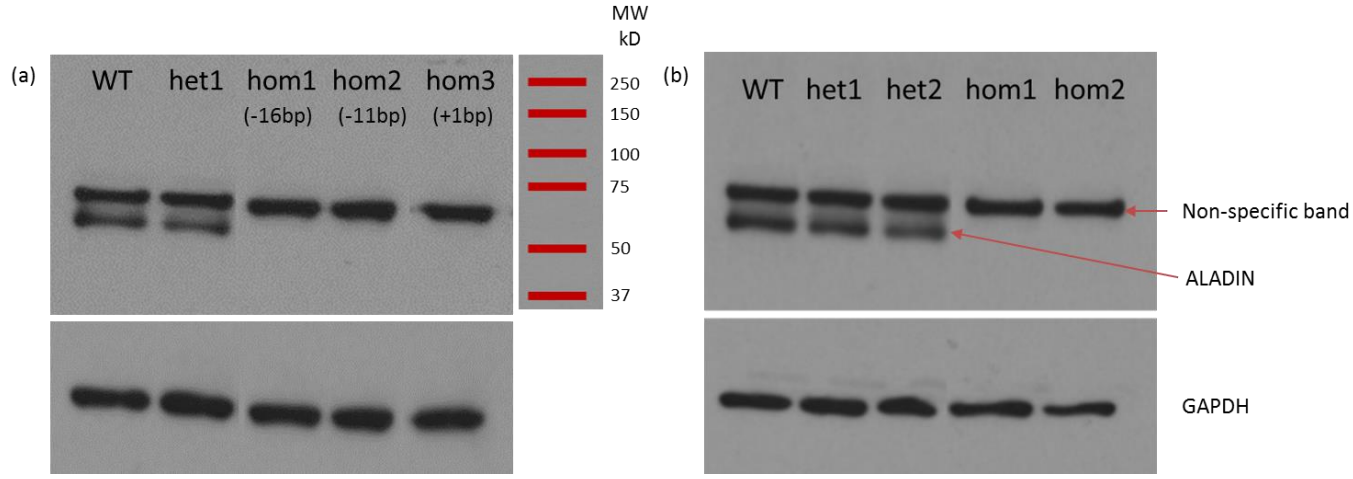


Figure 3.3: Western blot of (a) *AAAS* exon 2 knockout clones with the respective indels of the non-targeted allele indicated in brackets and (b) *AAAS* exon 1 43C>A point mutation clones. Wild-type (WT) and heterozygotes (het) express the ALADIN protein, which is however knocked out in homozygotes (hom). GAPDH was used as a loading control.

An additional slower-migrating (upper) band was observed in all the clones, which is probably due to cross-reaction to an AAAS paralog or to an unrelated protein. The multiple isoforms of ALADIN are formed by skipping of exon 6 to form 56kDa isoform, or translation

initiation in exon 3 to form the shorter 46kDa isoform. Mutations in exons 1 or exon 2, generated in this study, will result in a premature stop codon within exon 3. The mutated protein for *AAAS* exon 1 point mutation is 57 amino acids long weighing approximately 6kDa, whereas the lengths for the *AAAS* exon 2 knockout clones were 87 (for 16bp deletion), 81 (11bp deletion) and 85 (1bp insertion) amino acids weighing around 10kDa, 9kDa, and 10kDa respectively. Thus, the upper band cannot be explained by any isoform of ALADIN. BLAST results of the Q9NRG9 isoform against the human genome indicated minor identities to a processed pseudogene on chromosome 6 (refer Appendix, Figure A15). This pseudogene, however, does not encode any functional protein, and thus, rules out the possibility of a AAAS paralog. Therefore, the upper band could represent non-specific binding to an unrelated protein in the human genome, an epitope of which might cross-react with the polyclonal antibody. To resolve the origin of the upper band on the Western blot, monoclonal antibodies could be used that recognize a unique epitope of ALADIN, different to that recognized by the polyclonal antibody (Proteintech) used in this experiment (refer Appendix A14).

The observations made in this study suggest that *AAAS* gene expression and protein function was completely knocked out. The truncated protein was not detectable in the assay, either because it was too small to be detected, or due to the probable decay of the mutated transcript or the degradation of the truncated protein within the cells. In any case, it is truncated and therefore not functional.

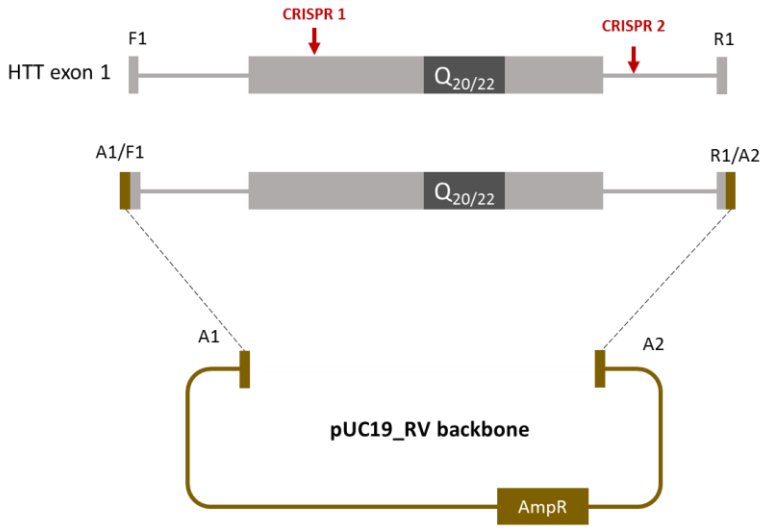
Chapter 4

Huntington’s disease

Insertion of extended Q-repeat (Q_{67}) fragment in HTT exon 1 (HTT1)

In this part of the study, the sequences of the alleles of HTT gene present in KOLF2-C1 cells were first determined. For this purpose, a genomic region containing HTT exon 1 (~1.3kb) was cloned into pUC19_RV backbone (~2.7kb) by InFusion cloning, forming an approximately 4kb plasmid, pHTT_exon 1 (Figure 4.1 (a)). Sequencing of multiple plasmids showed that there were two haplotypes for this gene with different number of Q-repeats in KOLF2_C1 cells (Figure 4.1 (b)). Compared to the reference sequence, which contains 21 Q-repeats (Ensembl), the two alleles of HTT in KOLF2-C1 cells contain 20 and 22 Q repeats, respectively.

(a)



(b)

```
5′- CTTCCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG---CAACAGCCGC Q21(ref)
5′- CTTCCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAACAGCCGC Q22
5′- CTTCCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG------CAACAGCCGC Q20
```

Figure 4.1: **Sequencing of the HTT alleles of KOLF2-C1 cells.** (a) Construction of InFusion clone, pHTT_exon 1. (b) Sequences of alleles with 20 and 22 Q-repeats (CAG-repeats; indicated in red) in KOLF2_C1 genome aligned with the reference sequence from Ensembl.

A pair of CRISPRs were tested in KOLF2-C1 cells to initially delete a region of HTT exon 1 by NHEJ before attempting to insert the extended Q_{67} repeat fragment into the exon by homologous recombination (Figure 4.2). The deletion would not have any effect on the gene function because the gene sequence deleted along with the wild-type repeat length ($Q_{20/22}$) can be determined by sequencing; this deleted gene sequence can then be inserted into the DNA along with the extended fragment by homologous recombination with a donor plasmid. The cells were co-transfected in separate plasmid-based and protein-based experiments. In the

plasmid-based experiment, the cells were co-transfected with Cas9 nuclease and CRISPR gRNA expression plasmids. This plasmid-based experiment was also attempted using trans-puromycin selection, that is, the delivery of a puro-selection plasmid along with the CRISPR and Cas9 plasmids during transfection, to select for the desired clones. On the other hand, in the protein-based experiment, cells were transfected with Cas9-RNP, pre-assembled from crRNA, tracrRNA, and Cas9 protein purchased commercially.

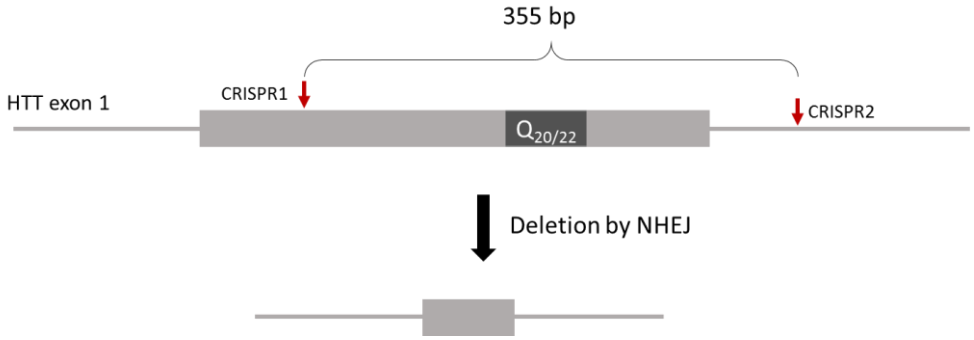
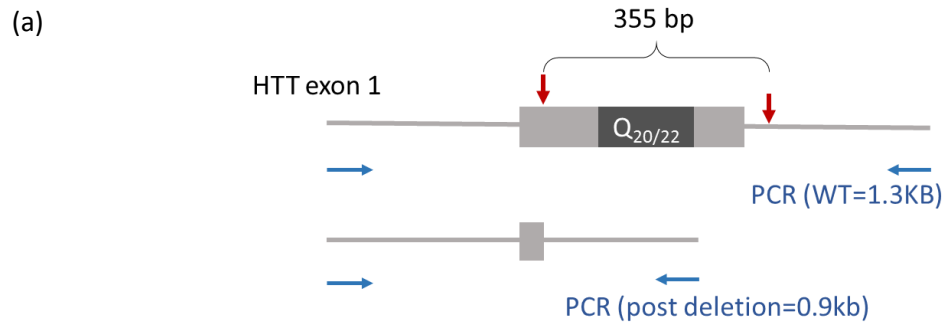


Figure 4.2: **Schematic representation of deletion of HTT exon 1 by paired CRISPRs (1 & 2)**. The deletion is for subsequent insertion of the extended Q_{67} repeat fragment

DNA extracted from transfected cells were assayed for deletion by NHEJ along with wild-type KOLF2_C1 DNA as a control. The length of the PCR amplicon from wild-type KOLF2_C1 DNA was ~1.3kb and that from the cells after deletion was expected to be 938bp (1293- 355bp). Thus, the presence of a 938bp band in a cell pool would indicate activity of both the CRISPRs and hence the 355bp deletion. However, as the *in vivo* assay demonstrates (Figure 4.3), the expected deletion did not occur in both the plasmid and protein based experiments. This implies that either or both CRISPRs were not functional in cleaving the DNA. The trans-puromycin selection also did not seem to affect in enrichment of the cells containing the deletion.



(b)

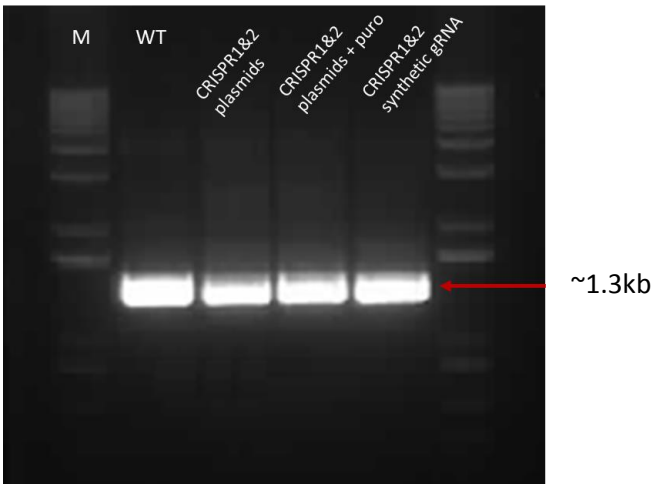
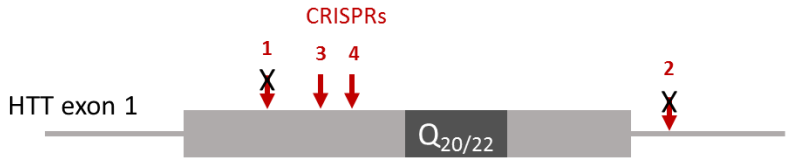


Figure 4.3: ***In vivo* assay for detection of activity of paired CRISPRs on HTT exon 1 of KOLF2_C1 genomic DNA.** (a) Schematic illustration showing the region of HTT exon 1 amplified by PCR (primer positions indicated in blue arrows) to detect for the 355bp deletion by paired CRISPRs flanking the Q-repeat region (indicated in red arrows). (b) Gel image for the assay of 355bp deletion with paired CRISPRs (1&2).

Two new CRISPRs (CRISPRs 3 and 4; Figure 4.4 (a)) were selected for introducing the extended HTT_Q_{67} repeat into the exon by homologous recombination. The activities of these CRISPRs along with the original pair were tested *in vitro* by measuring cleavage of the pHTT_exon 1 plasmid. It is worth mentioning here that *in vitro* Cas9 assay was done instead of a T7E1 assay because HTT exon 1 had two alleles in KOLF2_C1 which would have formed heteroduplexes in T7E1 assay giving misleading results. The *in vitro* Cas9 digestion assay (Figure 4.4 (b)) demonstrated greater activity for CRISPRs 3 and 4 when compared to little or no detectable activity for CRISPRs 1 and 2. This was suggested by the presence of a slower migrating band for these new CRISPR gRNAs, indicating the pHTT_exon 1 plasmid was linearized. This band was, however, very faint using the two previous CRISPR guides. Thus, the *in vitro* Cas9 assay implies that the new guides were active in Cas9 cleavage of the plasmid, and that they were suitable for transfection into iPS cells.

(a)



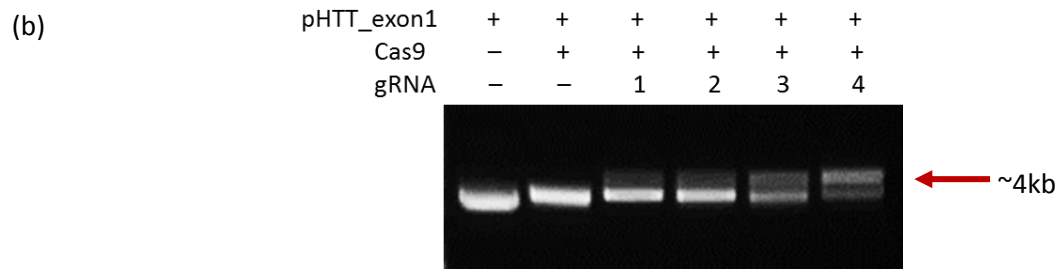


Figure 4.4: ***In vitro* Cas9 assay for detection of cleavage activity of all CRISPR gRNAs on pHTT_exon1 plasmid.** (a) CRISPR positions on the exon in pHTT_exon 1. Red arrows indicate new CRISPRs and crossed arrows indicate the paired CRISPRs used in first experiment. (b) *In vitro* Cas9 assay testing the activity of all four CRISPRs on pHTT_exon 1.

In the second targeting strategy using CRISPRs 3 and 4, the cells were co-transfected with pre-assembled Cas9-RNP and a donor plasmid containing the extended HTT_Q_{67} repeat fragment. This targeting donor plasmid was constructed by cloning in of a BlpI/BsgI digested Q_{67} synthetic fragment into the pHTT_exon 1 plasmid (Figure 4.5).

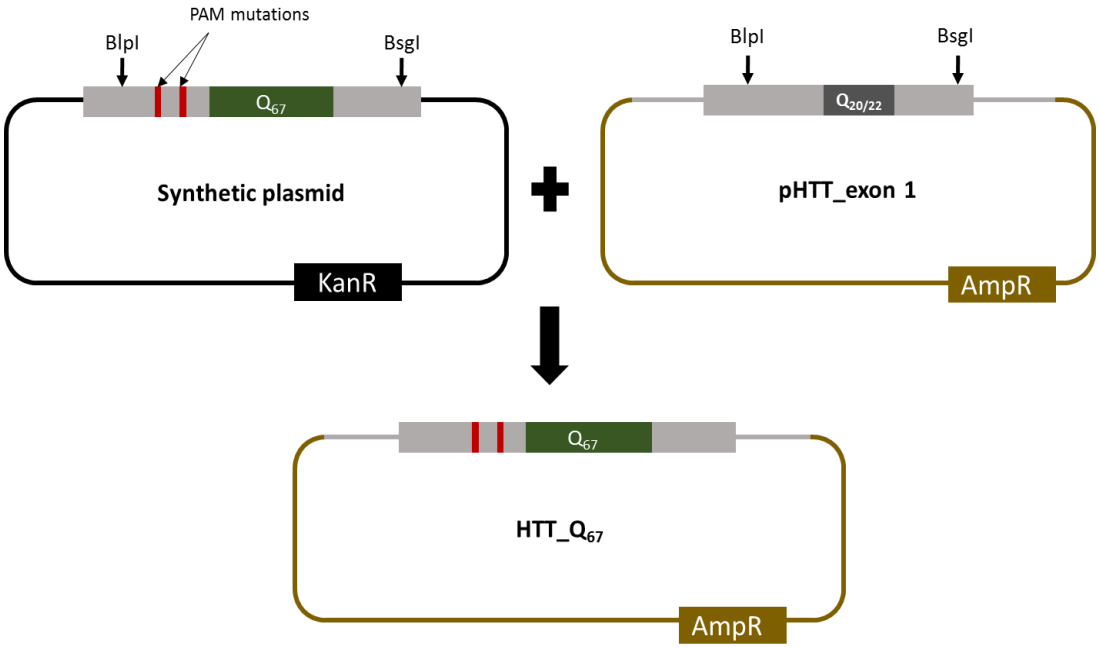


Figure 4.5: Construction of HTT_Q_{67} donor plasmid by cloning in of a BlpI/BsgI digested Q_{67} synthetic fragment into pHTT_exon 1.

The donor plasmid was expected to introduce the extended Q-repeat, by homologous recombination, into HTT exon 1 upon genomic DNA cleavage with Cas9-RNP (Figure 4.6). Two transfection experiments were conducted, one with each guide RNA.

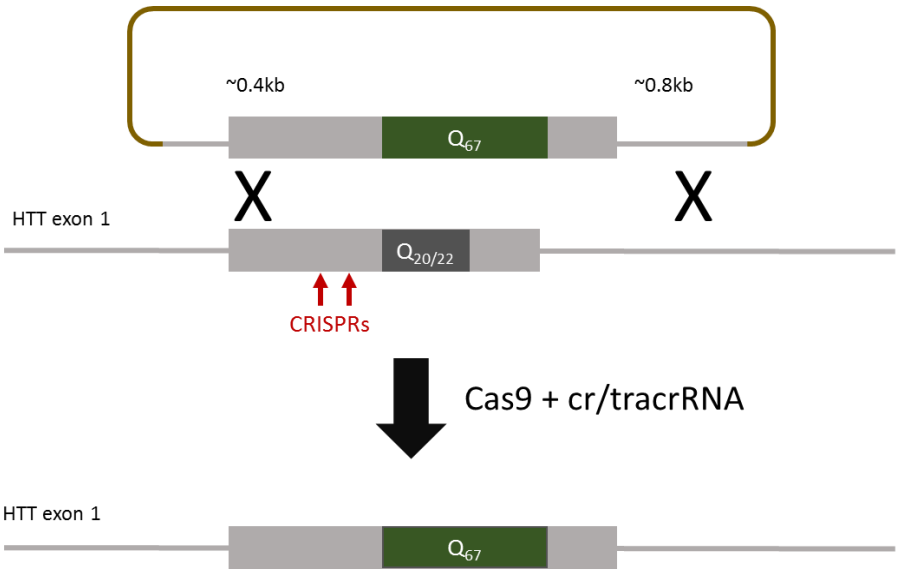


Figure 4.6: **Strategy for insertion of the extended Q_{67} repeat into HTT exon 1 in genomic DNA by homologous recombination with HTT_Q_{67} donor plasmid.** CRISPRs at which the homologous recombination occurs are indicated in red arrows. "X" denotes homologous recombination.

The genomic DNA from a pool of transfected cells was extracted and amplified with primers outside the homology arms to avoid detection of random insertions of the plasmid donor. These PCR amplicons were then amplified again with nested primers to assay for the insertion of the 67 Q-repeat fragment, a 512bp PCR amplicon (Figure 4.7 (a)). In assaying this pool of cells, it was expected that only a small fraction of the cells will contain the expanded allele. As the gel image shows (Figure 4.7 (b)), the 512bp band was not observed indicating very few or no cells contained the inserted fragment. Therefore, individual clones were picked and screened by nested PCR amplification in the same manner as described above. Detection of both the expanded (512bp) and wild-type (374bp) bands should be observed in a small subset of clones. Curiously, some colonies (for example B12, E6, and H10) exhibited only the one band approximately 500bp in size (Figure 4.7 (c)). However, sequencing these products yielded no trace. Therefore, I conclude that this band is not the expanded allele, but rather non-specific amplification or contamination which yielded a similar sized amplicon. The remaining clones amplified a wild-type band which was confirmed by sequencing. Thus, I conclude that I was unsuccessful in my attempt to engineer cells with an expanded CAG repeats.

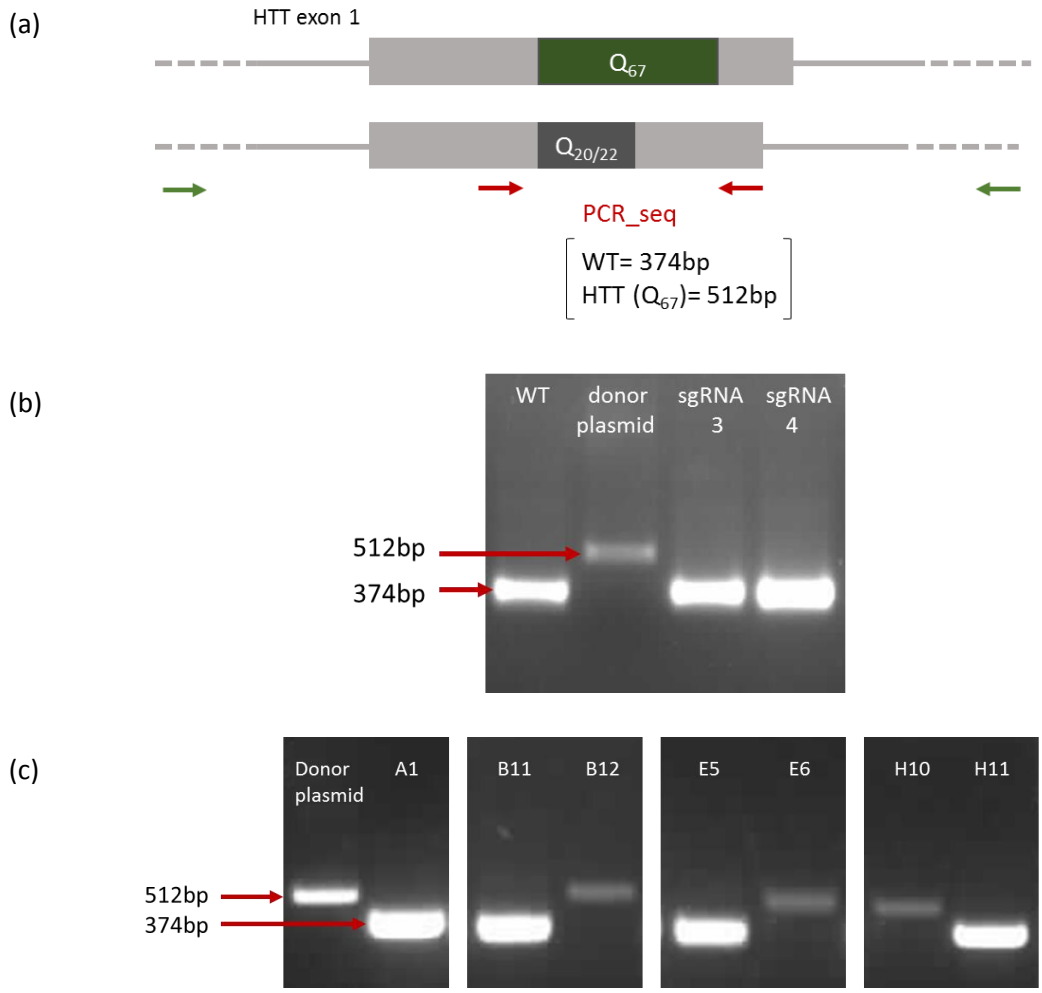


Figure 4.7: **Assay for detection of insertion of extended Q_{67} fragment into HTT exon 1 of KOLF2_C1 genomic DNA.** (a) Schematic diagram showing the region of HTT exon 1 amplified with primers outside the homology arms (shown in green arrows) and then with nested primers (shown in red arrows) to assay for the insertion. (b) Gel for detection of the 512bp band (corresponding to 67 Q-repeats) in genomic DNA of the pool of transfected cells. (c) Cropped images of gels showing the bands for individual cell colonies.

Chapter 5
Conclusions & Future studies

Conclusions:

AAAS disease models: Three AAAS biallelic knockout cell lines and two homozygous point mutant cell lines were generated and validated in this study. The knockout clones were generated using a plasmid-based nuclease-assisted gene targeting strategy in human iPS cells. At a biallelic targeting efficiency of 3%, these knockout clones contained frameshift indels in one allele, introduced by NHEJ, and puro-resistant cassette in the other allele by homologous recombination with a targeting donor plasmid. The point mutant clones were generated in a protein-based nuclease-assisted gene targeting strategy by HDR using a symmetric ssODN. Two rounds of transfection were required to recover cells homozygous for the disease allele. HDR introducing the desired point mutation was more efficient when lower amount of the Cas9 protein was used. However, higher activity was seemingly demonstrated by larger amount of Cas9 probably because this larger Cas9 amount caused re-cleavage of the modified locus as well as secondary damage to the locus, resulting in more NHEJ indels. Unexpectedly, adding mutations in the PAM did not improve rates of HDR which could be explained by a poorer rate of incorporation of the ssODN when more mismatches to the target locus are present. In future experiments, this issue with the optimum amount of Cas9 to be used for cleavage can be addressed by titrating the amount of Cas9. This would potentially avoid frequent NHEJ indels as observed with higher amount of Cas9 or reduced activity if too low amount is used.

It is worth mentioning here that upon re-transfection, the damaged wild-type copy of the gene could use either the ssODN or the mutant allele as a repair template. This could be validated by carrying out further experiments using an ssODN that contains one or more silent mutations in its sequence (Skarnes & Koutsourakis, unpublished data). Presence of the silent mutation/s in the modified DNA strand would indicate that the damage was repaired using the ssODN oligo as the template, whereas their absence would indicate gene conversion using the mutant allele. Gene conversion was evidenced in this experiment by the presence of two homozygote wild-type clones which were obtained upon the second transfection of the heterozygotes. Besides, this is possible that these homozygous point mutant clones could represent the loss of one copy of *AAAS* gene. This issue can be addressed by conducting further Southern blotting experiments to confirm amplification from both the alleles and that the clones are not representing homozygosity superficially. Another approach that can be adapted to validate the presence of both the alleles is by attempting to correct the point mutation in the gene with an ssODN containing the wild-type sequence. Homozygous wild-type cells only among all the screened clones would indicate that an allele is missing in the disease model

clones. Contrarily, if heterozygotes with a mutant allele and a wild-type allele are present among the screened clones, it would indicate that both the alleles are present and that none was deleted during genome-editing.

All these AAAS disease models were validated as a complete knockout of the ALADIN protein which suggests that the disease is caused by null or complete loss-of-function mutations in the *AAAS* gene. An additional band was detected using the polyclonal antibody (Proteintech) used in this experiment which probably represents cross-reaction with an unrelated protein in the proteome. The origin of this band needs to be investigated further using monoclonal antibodies that recognize an epitope of ALADIN different to that recognized by the polyclonal antibody used in this experiment.

HTT disease model: From a screen of 192 clones, no clones containing the extended Q_{67} repeat fragment were recovered. This is most likely because the efficiency of homologous recombination was too low in unselected clones. With plasmid-based nuclease-assisted gene targeting, transient puromycin selection did not seem to sufficiently enrich for homologous recombination events. Also, in the protein-based strategy, none of the clones exhibited insertion of the extended fragment by homologous recombination. The protein delivery should have been better than the plasmid-based approach because proteins are more efficient in specific genome modification, probably due to the transient window of Cas9 activity to which each genome is exposed (Zuris et al. 2015). Cas9-RNP is expected to modify the genomic DNA almost immediately after delivery because, unlike with plasmid-based targeting, no expression of Cas9 is needed prior to cleavage. With plasmids, Cas9 expression is slow and it can take several days before any considerable level of genome modification can occur (Hu et al. 2016). Further experiments need to be conducted using cis-delivery of a drug selection cassette within the targeting plasmid containing the extended 67 Q-repeat fragment (Figure 5.1) which would potentially increase the chances of selecting the drug-resistant clones.

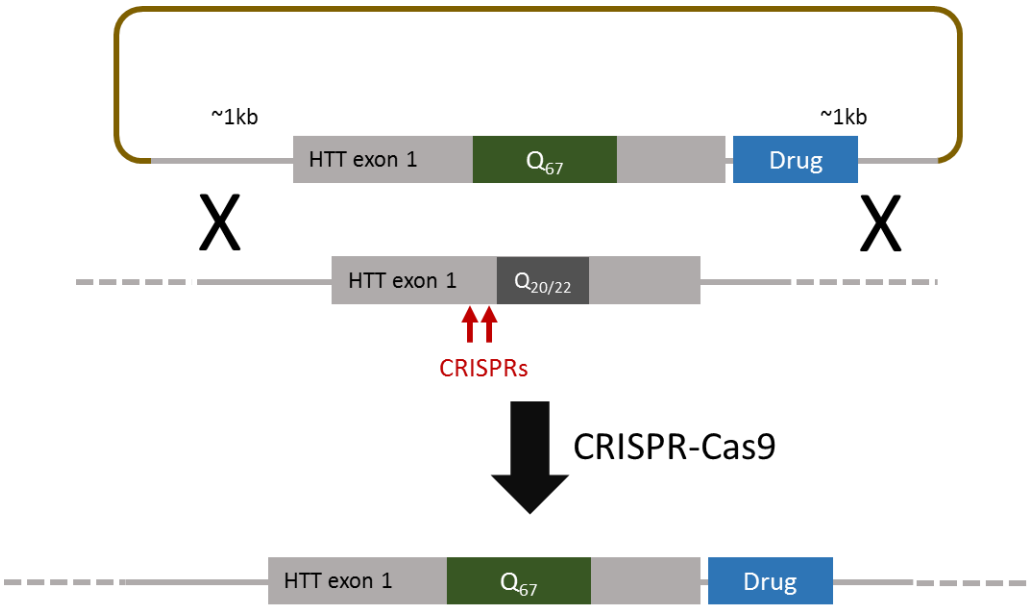


Figure 5.1: **Strategy for insertion of extended Q_{67}-repeat fragment into HTT exon 1 with cis-delivery of a drug selection cassette within the intron.** CRISPRs at which the homologous recombination occurs are indicated in red arrows. "X" indicates homologous recombination.

Also, multiple CRISPRs could be found that cleave the DNA on either side of the CAG repeat domain. Such approach would introduce multiple DSBs in the DNA and by deleting the repeat region, would increase the chances of homologous recombination between the damaged DNA and the donor plasmid.

Future studies:

The AAAS iPS cell models generated will enable us to gain a better understanding of the diverse functions of the ALADIN protein and the disease progression and pathophysiology of Triple-A syndrome. These cells can be compared to their respective isogenic wild-type control cells to identify phenotypic differences. Importantly, they can be differentiated to neurons and neural progenitor cells of the cerebral cortex, cerebellum, hippocampus, brainstem, and spinal cord, where the gene is abundantly expressed. Neural cells, affected by the disease mutations, will help us study the progressive neurological abnormalities like cognitive, motor, and sensory dysfunctions. In theory, these iPSC models can also be differentiated to adrenocortical cells, another major cell type affected in AAAS, to parasympathetic motor neurons, which mediate tear production (Dartt 2009), and to brainstem and vagal afferent neurons, which control oesophageal peristalsis (Richards & Sugarbaker 1995).

Since these *AAAS* mutants and wild-type cells are isogenic, the wild-type cells can provide us with appropriate controls to determine what other interacting genes or if any environmental factors could be confounding the phenotype of the diseased cells. Comparing the phenotypic patterns of iPSC models to patient-derived cells can also enable us gain insights into the effect of genetic background on cellular phenotypes. Moreover, these human cells complement mouse models which may not fully recapitulate the human AAAS abnormalities due to probable species-specific ALADIN functions. Further genome editing to create additional disease alleles will also help us unravel the genotype-phenotype correlation which is currently poorly understood. However, since phenotypic effects at the organism level cannot be observed with iPSC models, animal models are also needed which complement iPS cell models for studies of disease mechanisms (Merkle & Eggan 2013). An ideal animal model in many scientific studies is mouse, however, mouse models lacking the ALADIN protein have not been observed to exhibit the AAAS phenotype as humans do (Huebner et al. 2006). These mutant mice were indistinguishable from their wild-type siblings which could either be due to ALADIN functions in mice being different to that in humans or due to other external factors such as, environmental conditions or genetic modifiers. Thus, they are not suitable models to study this disease. However, other animal models genetically engineered to contain homozygous *AAAS* mutations could also be used for this disease. Such models could include primates as they are more evolutionary related to humans and have more common or shared genetic material. Thus, once the disease relevant phenotypes and their pathophysiology are identified and well-understood, these cell models can be used to perform drug screens and to develop novel therapeutics for treating this devastating disease.

References

Ahmed, RM, Devenney, EM, Irish, M, Ittner, A, Naismith, S, Ittner, LM, Rohrer, JD, Halliday, GM, Eisen, A, Hodges, JR & Kiernan, MC 2016, 'Neuronal network disintegration: common pathways linking neurodegenerative diseases', *Journal of Neurology, Neurosurgery & Psychiatry*, p. jnnp-2014-308350.

Allgrove, J, Clayden, GS, Grant, DB & Macaulay, JC 1978, 'Familial glucocorticoid deficiency with achalasia of the cardia and deficient tear production', *Lancet (London, England)*, vol. 1, no. 8077, pp. 1284–1286.

Arrasate, M & Finkbeiner, S 2012, 'Protein aggregates in Huntington's disease', *Experimental Neurology*, vol. 238, no. 1, pp. 1–11.

Barnham, KJ, Masters, CL & Bush, AI 2004, 'Neurodegenerative diseases and oxidative stress', *Nature Reviews Drug Discovery*, vol. 3, no. 3, pp. 205–214.

Barut, BA & Zon, LI 2000, 'Realizing the potential of zebrafish as a model for human disease', *Physiological Genomics*, vol. 2, no. 2, pp. 49–51.

Benavides, F & Guénet, JL 2001, '[Murine models for human diseases]', *Medicina*, vol. 61, no. 2, pp. 215–231.

Bence, NF, Sampat, RM & Kopito, RR 2001, 'Impairment of the ubiquitin-proteasome system by protein aggregation', *Science (New York, N.Y.)*, vol. 292, no. 5521, pp. 1552–1555.

Bertram, L & Tanzi, RE 2005, 'The genetic epidemiology of neurodegenerative disease', *Journal of Clinical Investigation*, vol. 115, no. 6, pp. 1449–1457.

Biglan, KM, Ross, CA, Langbehn, DR, Aylward, EH, Stout, JC, Queller, S, Carlozzi, NE, Duff, K, Beglinger, LJ & Paulsen, JS 2009, 'Motor Abnormalities in Premanifest Persons with Huntington's Disease: The PREDICT-HD Study', *Movement disorders : official journal of the Movement Disorder Society*, vol. 24, no. 12, pp. 1763–1772.

Bizzarri, C, Benevento, D, Terzi, C, Huebner, A & Cappa, M 2013, 'Triple A (Allgrove) syndrome: an unusual association with syringomyelia', *Italian Journal of Pediatrics*, vol. 39, p. 39.

Bourne, C, Clayton, C, Murch, A & Grant, J 2006, 'Cognitive impairment and behavioural difficulties in patients with Huntington's disease', *Nursing Standard (Royal College of Nursing (Great Britain): 1987)*, vol. 20, no. 35, pp. 41–44.

Brooks, BP, Kleta, R, Caruso, RC, Stuart, C, Ludlow, J & Stratakis, CA 2004, 'Triple-A syndrome with prominent ophthalmic features and a novel mutation in the *AAAS* gene: a case report', *BMC Ophthalmology*, vol. 4, p. 7.

Brown, RC, Lockwood, AH & Sonawane, BR 2005, 'Neurodegenerative Diseases: An Overview of Environmental Risk Factors', *Environmental Health Perspectives*, vol. 113, no. 9, pp. 1250–1256.

Butler, C & Zeman, AZJ 2005, 'Neurological syndromes which can be mistaken for psychiatric conditions', *Journal of Neurology, Neurosurgery & Psychiatry*, vol. 76, no. suppl 1, pp. i31–i38.

Cattaneo, E 2003, 'Dysfunction of wild-type huntingtin in Huntington disease', *News in Physiological Sciences: An International Journal of Physiology Produced Jointly by the International Union of Physiological Sciences and the American Physiological Society*, vol. 18, pp. 34–37.

Chen, B, Gilbert, LA, Cimini, BA, Schnitzbauer, J, Zhang, W, Li, G-W, Park, J, Blackburn, EH, Weissman, JS, Qi, LS & Huang, B 2013, 'Dynamic imaging of genomic loci in living human cells by an optimized CRISPR/Cas system', *Cell*, vol. 155, no. 7, pp. 1479–1491.

Cong, L, Zhou, R, Kuo, Y-C, Cunniff, M & Zhang, F 2012, 'Comprehensive interrogation of natural TALE DNA-binding modules and transcriptional repressor domains', *Nature Communications*, vol. 3, p. 968.

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

Cowan, CM & Raymond, LA 2006, 'Selective neuronal degeneration in Huntington's disease', *Current Topics in Developmental Biology*, vol. 75, pp. 25–71.

Cox, DBT, Platt, RJ & Zhang, F 2015, 'Therapeutic Genome Editing: Prospects and Challenges', *Nature medicine*, vol. 21, no. 2, pp. 121–131.

Cronshaw, JM, Krutchinsky, AN, Zhang, W, Chait, BT & Matunis, MJ 2002, 'Proteomic analysis of the mammalian nuclear pore complex', *The Journal of Cell Biology*, vol. 158, no. 5, pp. 915–927.

Cronshaw, JM & Matunis, MJ 2003, 'The nuclear pore complex protein ALADIN is mislocalized in triple A syndrome', *Proceedings of the National Academy of Sciences of the United States of America*, vol. 100, no. 10, pp. 5823–5827.

Dang, Y, Jia, G, Choi, J, Ma, H, Anaya, E, Ye, C, Shankar, P & Wu, H 2015, 'Optimizing sgRNA structure to improve CRISPR-Cas9 knockout efficiency', *Genome Biology*, vol. 16, p. 280.

Dartt, DA 2009, 'Neural Regulation of Lacrimal Gland Secretory Processes: Relevance in Dry Eye Diseases', *Progress in retinal and eye research*, vol. 28, no. 3, pp. 155–177.

Dayalu, P & Albin, RL 2015, 'Huntington Disease: Pathogenesis and Treatment', *Neurologic Clinics*, vol. 33, Movement Disorders, no. 1, pp. 101–114.

De Rooij, KE, Dorsman, JC, Smoor, MA, Den Dunnen, JT & Van Ommen, GJ 1996, 'Subcellular localization of the Huntington's disease gene product in cell lines by immunofluorescence and biochemical subcellular fractionation', *Human Molecular Genetics*, vol. 5, no. 8, pp. 1093–1099.

Dimos, JT, Rodolfa, KT, Niakan, KK, Weisenthal, LM, Mitsumoto, H, Chung, W, Croft, GF, Saphier, G, Leibel, R, Goland, R, Wichterle, H, Henderson, CE & Eggan, K 2008, 'Induced pluripotent stem cells generated from patients with ALS can be differentiated into motor neurons', *Science (New York, N.Y.)*, vol. 321, no. 5893, pp. 1218–1221.

Doudna, JA & Charpentier, E 2014, 'Genome editing. The new frontier of genome engineering with CRISPR-Cas9', *Science (New York, N.Y.)*, vol. 346, no. 6213, p. 1258096.

Duan, W, Jiang, M & Jin, J 2014, 'Metabolism in HD - Still a relevant mechanism?', *Movement disorders : official journal of the Movement Disorder Society*, vol. 29, no. 11, pp. 1366–1374.

Dumic, M, Barišic, N, Kusec, V, Stingl, K, Skegro, M, Stanimirovic, A, Koehler, K & Huebner, A 2012, 'Long-term clinical follow-up and molecular genetic findings in eight patients with triple A syndrome', *European Journal of Pediatrics*, vol. 171, no. 10, pp. 1453–1459.

Duyao, M, Ambrose, C, Myers, R, Novelletto, A, Persichetti, F, Frontali, M, Folstein, S, Ross, C, Franz, M & Abbott, M 1993, 'Trinucleotide repeat length instability and age of onset in Huntington's disease', *Nature Genetics*, vol. 4, no. 4, pp. 387–392.

Ebert, AD, Liang, P & Wu, JC 2012, 'Induced pluripotent stem cells as a disease modeling and drug screening platform', *Journal of Cardiovascular Pharmacology*, vol. 60, no. 4, pp. 408–416.

European Commission 2016, *Neurodegenerative Disorders - European Commission*, viewed 15 May 2016, <http://ec.europa.eu/health/major_chronic_diseases/diseases/brain_neurological/index_en.htm>.

Frank, S 2014, 'Treatment of Huntington's disease', *Neurotherapeutics: The Journal of the American Society for Experimental NeuroTherapeutics*, vol. 11, no. 1, pp. 153–160.

Fu, Y, Sander, JD, Reyon, D, Cascio, VM & Joung, JK 2014, 'Improving CRISPR-Cas nuclease specificity using truncated guide RNAs', *Nature Biotechnology*, vol. 32, no. 3, pp. 279–284.

Gianaros, PJ, Jennings, JR, Sheu, LK, Greer, PJ, Kuller, LH & Matthews, KA 2007, 'Prospective reports of chronic life stress predict decreased grey matter volume in the hippocampus', *NeuroImage*, vol. 35, no. 2, pp. 795–803.

Grant, DB, Barnes, ND, Dumic, M, Ginalska-Malinowska, M, Milla, PJ, von Petrykowski, W, Rowlatt, RJ, Steendijk, R, Wales, JH & Werder, E 1993, 'Neurological and adrenal dysfunction in the adrenal insufficiency/alacrima/achalasia (3A) syndrome', *Archives of Disease in Childhood*, vol. 68, no. 6, pp. 779–782.

Guo, J, Gaj, T & Barbas, CF 2010, 'Directed evolution of an enhanced and highly efficient FokI cleavage domain for zinc finger nucleases', *Journal of Molecular Biology*, vol. 400, no. 1, pp. 96–107.

Gupta, RM & Musunuru, K 2014, 'Expanding the genetic editing tool kit: ZFNs, TALENs, and CRISPR-Cas9', *The Journal of Clinical Investigation*, vol. 124, no. 10, pp. 4154–4161.

Handschug, K, Sperling, S, Yoon, SJ, Hennig, S, Clark, AJ & Huebner, A 2001, 'Triple A syndrome is caused by mutations in *AAAS*, a new WD-repeat protein gene', *Human Molecular Genetics*, vol. 10, no. 3, pp. 283–290.

Harper, B 2005, 'Huntington disease', *Journal of the Royal Society of Medicine*, vol. 98, no. 12, p. 550.

Hau, J & Schapiro, SJ 2010, *Handbook of Laboratory Animal Science, Volume I, Third Edition: Essential Principles and Practices*, CRC Press.

Hirano, M, Furiya, Y, Asai, H, Yasui, A & Ueno, S 2006, 'ALADINI482S causes selective failure of nuclear protein import and hypersensitivity to oxidative stress in triple A syndrome', *Proceedings of the National Academy of Sciences of the United States of America*, vol. 103, no. 7, pp. 2298–2303.

Hodgkins, A, Farne, A, Perera, S, Grego, T, Parry-Smith, DJ, Skarnes, WC & Iyer, V 2015, 'WGE: a CRISPR database for genome engineering', *Bioinformatics (Oxford, England)*, vol. 31, no. 18, pp. 3078–3080.

Hodgson, JG, Agopyan, N, Gutekunst, C-A, Leavitt, BR, LePiane, F, Singaraja, R, Smith, DJ, Bissada, N, McCutcheon, K, Nasir, J, Jamot, L, Li, X-J, Stevens, ME, Rosemond, E, Roder, JC, Phillips, AG, Rubin, EM, Hersch, SM & Hayden, MR 1999, 'A YAC Mouse Model for Huntington's Disease with Full-Length Mutant Huntingtin, Cytoplasmic Toxicity, and Selective Striatal Neurodegeneration', *Neuron*, vol. 23, no. 1, pp. 181–192.

Houlden, H, Smith, S, Carvalho, M de, Blake, J, Mathias, C, Wood, NW & Reilly, MM 2002, 'Clinical and genetic characterization of families with triple A (Allgrove) syndrome', *Brain*, vol. 125, no. 12, pp. 2681–2690.

Hsu, PD, Lander, ES & Zhang, F 2014, 'Development and Applications of CRISPR-Cas9 for Genome Engineering', *Cell*, vol. 157, no. 6, pp. 1262–1278.

Hu, JH, Davis, KM & Liu, DR 2016, 'Chemical Biology Approaches to Genome Editing: Understanding, Controlling, and Delivering Programmable Nucleases', *Cell Chemical Biology*, vol. 23, no. 1, pp. 57–73.

Huang, B-W, Ray, PD, Iwasaki, K & Tsuji, Y 2013, 'Transcriptional regulation of the human ferritin gene by coordinated regulation of Nrf2 and protein arginine methyltransferases PRMT1 and PRMT4', *The FASEB Journal*, vol. 27, no. 9, pp. 3763–3774.

Huangfu, D, Osafune, K, Maehr, R, Guo, W, Eijkelenboom, A, Chen, S, Muhlestein, W & Melton, DA 2008, 'Induction of pluripotent stem cells from primary human fibroblasts with only Oct4 and Sox2', *Nature Biotechnology*, vol. 26, no. 11, pp. 1269–1275.

Huebner, A, Elias, LL & Clark, AJ 1999, 'ACTH resistance syndromes', *Journal of pediatric endocrinology & metabolism: JPEM*, vol. 12 Suppl 1, pp. 277–293.

Huebner, A, Kaindl, AM, Knobeloch, KP, Petzold, H, Mann, P & Koehler, K 2004, 'The triple A syndrome is due to mutations in ALADIN, a novel member of the nuclear pore complex', *Endocrine Research*, vol. 30, no. 4, pp. 891–899.

Huebner, A, Mann, P, Rohde, E, Kaindl, AM, Witt, M, Verkade, P, Jakubiczka, S, Menschikowski, M, Stoltenburg-Didinger, G & Koehler, K 2006, 'Mice lacking the nuclear pore complex protein ALADIN show female infertility but fail to develop a phenotype resembling human triple A syndrome', *Molecular and Cellular Biology*, vol. 26, no. 5, pp. 1879–1887.

Hunter, P 2008, 'The paradox of model organisms. The use of model organisms in research will continue despite their shortcomings', *EMBO Reports*, vol. 9, no. 8, pp. 717–720.

Iyer, V, Shen, B, Zhang, W, Hodgkins, A, Keane, T, Huang, X & Skarnes, WC 2015, 'Off-target mutations are rare in Cas9-modified mice', *Nature Methods*, vol. 12, no. 6, p. 479.

Jinek, M, Chylinski, K, Fonfara, I, Hauer, M, Doudna, JA & Charpentier, E 2012, 'A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity', *Science (New York, N.Y.)*, vol. 337, no. 6096, pp. 816–821.

Kallabi, F, Ben Rhouma, B, Baklouti, S, Ghorbel, R, Felhi, R, Keskes, L & Kamoun, H 2016, 'Splicing Defects in the *AAAS* Gene Leading to both Exon Skipping and Partial Intron Retention in a Tunisian Patient with Allgrove Syndrome', *Hormone Research in Paediatrics*, vol. 86, no. 2, pp. 90–93.

Kast, P 1994, 'pKSS--a second-generation general purpose cloning vector for efficient positive selection of recombinant clones', *Gene*, vol. 138, no. 1–2, pp. 109–114.

Kim, HS, Bernitz, JM, Lee, D-F & Lemischka, IR 2014, 'Genomic Editing Tools to Model Human Diseases with Isogenic Pluripotent Stem Cells', *Stem Cells and Development*, vol. 23, no. 22, pp. 2673–2686.

Kind, B, Koehler, K, Lorenz, M & Huebner, A 2009, 'The nuclear pore complex protein ALADIN is anchored via NDC1 but not via POM121 and GP210 in the nuclear envelope', *Biochemical and Biophysical Research Communications*, vol. 390, no. 2, pp. 205–210.

Kind, B, Koehler, K, Krumbholz, M, Landgraf, D & Huebner, A 2010, 'Intracellular ROS level is increased in fibroblasts of triple A syndrome patients', *Journal of Molecular Medicine (Berlin, Germany)*, vol. 88, no. 12, pp. 1233–1242.

Kiriyama, T, Hirano, M, Asai, H, Ikeda, M, Furiya, Y & Ueno, S 2008, 'Restoration of nuclear-import failure caused by triple A syndrome and oxidative stress', *Biochemical and Biophysical Research Communications*, vol. 374, no. 4, pp. 631–634.

Krumbholz, M, Koehler, K & Huebner, A 2006, 'Cellular localization of 17 natural mutant variants of ALADIN protein in triple A syndrome - shedding light on an unexpected splice mutation', *Biochemistry and Cell Biology = Biochimie Et Biologie Cellulaire*, vol. 84, no. 2, pp. 243–249.

Landles, C & Bates, GP 2004, 'Huntingtin and the molecular pathogenesis of Huntington's disease', *EMBO Reports*, vol. 5, no. 10, pp. 958–963.

Langbehn, DR, Brinkman, RR, Falush, D, Paulsen, JS, Hayden, MR & International Huntington's Disease Collaborative Group 2004, 'A new model for prediction of the age of onset and penetrance for Huntington's disease based on CAG length', *Clinical Genetics*, vol. 65, no. 4, pp. 267–277.

Li, SH, Gutekunst, CA, Hersch, SM & Li, XJ 1998, 'Interaction of huntingtin-associated protein with dynactin P150Glued', *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, vol. 18, no. 4, pp. 1261–1269.

Li, D & Roberts, R 2001, 'WD-repeat proteins: structure characteristics, biological function, and their involvement in human diseases', *Cellular and molecular life sciences: CMLS*, vol. 58, no. 14, pp. 2085–2097.

Li, S-H & Li, X-J 2004, 'Huntingtin and its role in neuronal degeneration', *The Neuroscientist: A Review Journal Bringing Neurobiology, Neurology and Psychiatry*, vol. 10, no. 5, pp. 467–475.

Lin, MT & Beal, MF 2006, 'Mitochondrial dysfunction and oxidative stress in neurodegenerative diseases', *Nature*, vol. 443, no. 7113, pp. 787–795.

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

Marshall, J, White, K, Weaver, M, Flury Wetherill, L, Hui, S, Stout, JC, Johnson, SA, Beristain, X, Gray, J, Wojcieszek, J & Foroud, T 2007, 'Specific psychiatric manifestations among preclinical Huntington disease mutation carriers', *Archives of Neurology*, vol. 64, no. 1, pp. 116–121.

Mattson, MP 2000, 'Apoptosis in neurodegenerative disorders', *Nature Reviews Molecular Cell Biology*, vol. 1, no. 2, pp. 120–130.

Mazzone, L, Postorino, V, De Peppo, L, Vassena, L, Fatta, L, Armando, M, Giuseppe, Cappa, M, Vicari, S, Mazzone, L, Postorino, V, De Peppo, L, Vassena, L, Fatta, L, Armando, M, Scir&#xe8, Giuseppe, Cappa, M & Vicari, S 2013, 'Longitudinal Neuropsychological Profile in a Patient with Triple A Syndrome', *Case Reports in Pediatrics*, *Case Reports in Pediatrics*, vol. 2013, 2013, p. e604921.

Medvedev, SP, Shevchenko, AI & Zakian, SM 2010, 'Induced Pluripotent Stem Cells: Problems and Advantages when Applying them in Regenerative Medicine', *Acta Naturae*, vol. 2, no. 2, pp. 18–28.

Merkle, FT & Eggan, K 2013, 'Modeling Human Disease with Pluripotent Stem Cells: from Genome Association to Function', *Cell Stem Cell*, vol. 12, no. 6, pp. 656–668.

Miller, JC, Tan, S, Qiao, G, Barlow, KA, Wang, J, Xia, DF, Meng, X, Paschon, DE, Leung, E, Hinkley, SJ, Dulay, GP, Hua, KL, Ankoudinova, I, Cost, GJ, Urnov, FD, Zhang, HS, Holmes, MC, Zhang, L, Gregory, PD & Rebar, EJ 2011, 'A TALE nuclease architecture for efficient genome editing', *Nature Biotechnology*, vol. 29, no. 2, pp. 143–148.

Mitchell, IJ, Cooper, AJ & Griffiths, MR 1999, 'The selective vulnerability of striatopallidal neurons', *Progress in Neurobiology*, vol. 59, no. 6, pp. 691–719.

Morton, AJ 2013, 'Circadian and sleep disorder in Huntington's disease', *Experimental Neurology*, vol. 243, Circadian rhythms and sleep disorders, pp. 34–44.

Mullaney, PB, Weatherhead, R, Millar, L, Ayyash, LI, Ayberk, H, Cai, F & Risco, JM 1998, 'Keratoconjunctivitis sicca associated with achalasia of the cardia, adrenocortical insufficiency, and lacrimal gland degeneration: Keratoconjunctivitis sicca secondary to lacrimal gland degeneration may parallel degenerative changes in esophageal and adrenocortical function', *Ophthalmology*, vol. 105, no. 4, pp. 643–650.

Nance, MA & Myers, RH 2001, 'Juvenile onset Huntington's disease--clinical and research perspectives', *Mental Retardation and Developmental Disabilities Research Reviews*, vol. 7, no. 3, pp. 153–157.

Neer, EJ, Schmidt, CJ, Nambudripad, R & Smith, TF 1994, 'The ancient regulatory-protein family of WD-repeat proteins', *Nature*, vol. 371, no. 6495, pp. 297–300.

Nieoullon, A 2011, 'Neurodegenerative diseases and neuroprotection: current views and prospects', *Journal of Applied Biomedicine*, vol. 9, no. 4, pp. 173–183.

Nowak, CM, Lawson, S, Zerez, M & Bleris, L 2016, 'Guide RNA engineering for versatile Cas9 functionality', *Nucleic Acids Research*, p. gkw908.

Pavletich, NP & Pabo, CO 1991, 'Zinc finger-DNA recognition: crystal structure of a Zif268-DNA complex at 2.1 A', *Science*, vol. 252, no. 5007, pp. 809–817.

Prasad, R, Metherell, LA, Clark, AJ & Storr, HL 2013, 'Deficiency of ALADIN Impairs Redox Homeostasis in Human Adrenal Cells and Inhibits Steroidogenesis', *Endocrinology*, vol. 154, no. 9, pp. 3209–3218.

Prasad, R, Kowalczyk, JC, Meimaridou, E, Storr, HL & Metherell, LA 2014, 'Oxidative stress and adrenocortical insufficiency', *Journal of Endocrinology*, vol. 221, no. 3, pp. R63–R73.

Prpic, I, Huebner, A, Persic, M, Handschug, K & Pavletic, M 2003, 'Triple A syndrome: Genotype-phenotype assessment', *Clinical Genetics*, vol. 63, no. 5, pp. 415–417.

Ran, FA, Hsu, PD, Lin, C-Y, Gootenberg, JS, Konermann, S, Trevino, AE, Scott, DA, Inoue, A, Matoba, S, Zhang, Y & Zhang, F 2013, 'Double nicking by RNA-guided CRISPR Cas9 for enhanced genome editing specificity', *Cell*, vol. 154, no. 6, pp. 1380–1389.

Rangone, H, Humbert, S & Saudou, F 2004, 'Huntington's disease: how does huntingtin, an anti-apoptotic protein, become toxic?', *Pathologie-Biologie*, vol. 52, no. 6, pp. 338–342.

Richards, WG & Sugarbaker, DJ 1995, 'Neuronal control of esophageal function', *Chest Surgery Clinics of North America*, vol. 5, no. 1, pp. 157–171.

Rosenblatt, A 2007, 'Neuropsychiatry of Huntington's disease', *Dialogues in Clinical Neuroscience*, vol. 9, no. 2, pp. 191–197.

Sarathi, V & Shah, NS 2010, 'Triple-a syndrome', *Advances in Experimental Medicine and Biology*, vol. 685, pp. 1-8.

Saudou, F, Finkbeiner, S, Devys, D & Greenberg, ME 1998, 'Huntingtin acts in the nucleus to induce apoptosis but death does not correlate with the formation of intranuclear inclusions', *Cell*, vol. 95, no. 1, pp. 55–66.

Semaka, A, Creighton, S, Warby, S & Hayden, MR 2006, 'Predictive testing for Huntington disease: interpretation and significance of intermediate alleles', *Clinical Genetics*, vol. 70, no. 4, pp. 283–294.

Shen, Z, Sathyan, KM, Geng, Y, Zheng, R, Chakraborty, A, Freeman, B, Wang, F, Prasanth, KV & Prasanth, SG 2010, 'A WD-Repeat Protein Stabilizes ORC Binding to Chromatin', *Molecular Cell*, vol. 40, no. 1, pp. 99–111.

Shen, B, Zhang, W, Zhang, J, Zhou, J, Wang, J, Chen, L, Wang, L, Hodgkins, A, Iyer, V, Huang, X & Skarnes, WC 2014, 'Efficient genome modification by CRISPR-Cas9 nickase with minimal off-target effects', *Nature Methods*, vol. 11, no. 4, pp. 399–402.

Skarnes, WC, Rosen, B, West, AP, Koutsourakis, M, Bushell, W, Iyer, V, Mujica, AO, Thomas, M, Harrow, J, Cox, T, Jackson, D, Severin, J, Biggs, P, Fu, J, Nefedov, M, de Jong,

PJ, Stewart, AF & Bradley, A 2011, 'A conditional knockout resource for the genome-wide study of mouse gene function', *Nature*, vol. 474, no. 7351, pp. 337–342.

Slow, EJ, Graham, RK & Hayden, MR 2006, 'To be or not to be toxic: aggregations in Huntington and Alzheimer disease', *Trends in genetics: TIG*, vol. 22, no. 8, pp. 408–411.

Small, SA 2003, 'Measuring correlates of brain metabolism with high-resolution MRI: a promising approach for diagnosing Alzheimer disease and mapping its course', *Alzheimer Disease and Associated Disorders*, vol. 17, no. 3, pp. 154–161.

Smih, F, Rouet, P, Romanienko, PJ & Jasin, M 1995, 'Double-strand breaks at the target locus stimulate gene targeting in embryonic stem cells.', *Nucleic Acids Research*, vol. 23, no. 24, pp. 5012–5019.

Smith, TF 2008, 'Diversity of WD-repeat proteins', *Sub-cellular Biochemistry*, vol. 48, pp. 20–30.

Spires, TL, Grote, HE, Varshney, NK, Cordery, PM, van Dellen, A, Blakemore, C & Hannan, AJ 2004, 'Environmental enrichment rescues protein deficits in a mouse model of Huntington's disease, indicating a possible disease mechanism', *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, vol. 24, no. 9, pp. 2270–2276.

Storr, HL, Clark, AJL, Priestley, JV & Michael, GJ 2005, 'Identification of the sites of expression of triple a syndrome mRNA in the rat using in situ hybridisation', *Neuroscience*, vol. 131, no. 1, pp. 113–123.

Storr, HL, Kind, B, Parfitt, DA, Chapple, JP, Lorenz, M, Koehler, K, Huebner, A & Clark, AJL 2009, 'Deficiency of Ferritin Heavy-Chain Nuclear Import in Triple A Syndrome Implies Nuclear Oxidative Damage as the Primary Disease Mechanism', *Molecular Endocrinology*, vol. 23, no. 12, pp. 2086–2094.

Streubel, J, Blücher, C, Landgraf, A & Boch, J 2012, 'TAL effector RVD specificities and efficiencies', *Nature Biotechnology*, vol. 30, no. 7, pp. 593–595.

Takahashi, K, Tanabe, K, Ohnuki, M, Narita, M, Ichisaka, T, Tomoda, K & Yamanaka, S 2007, 'Induction of pluripotent stem cells from adult human fibroblasts by defined factors', *Cell*, vol. 131, no. 5, pp. 861–872.

Takahashi, K & Yamanaka, S 2006, 'Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors', *Cell*, vol. 126, no. 4, pp. 663–676.

Terns, RM & Terns, MP 2014, 'CRISPR-based technologies: prokaryotic defense weapons repurposed', *Trends in genetics: TIG*, vol. 30, no. 3, pp. 111–118.

The Huntington's Disease Collaborative Research Group 1993, 'A novel gene containing a trinucleotide repeat that is expanded and unstable on Huntington's disease chromosomes. The Huntington's Disease Collaborative Research Group', *Cell*, vol. 72, no. 6, pp. 971–983.

Thompson, KJ, Fried, MG, Ye, Z, Boyer, P & Connor, JR 2002, 'Regulation, mechanisms and proposed function of ferritin translocation to cell nuclei', *Journal of Cell Science*, vol. 115, no. Pt 10, pp. 2165–2177.

Thompson, LM 2008, 'Neurodegeneration: A question of balance', *Nature*, vol. 452, no. 7188, pp. 707–708.

Thomson, JA, Itskovitz-Eldor, J, Shapiro, SS, Waknitz, MA, Swiergiel, JJ, Marshall, VS & Jones, JM 1998, 'Embryonic Stem Cell Lines Derived from Human Blastocysts', *Science*, vol. 282, no. 5391, pp. 1145–1147.

Tullio-Pelet, A, Salomon, R, Hadj-Rabia, S, Mugnier, C, De, L, Chaouachi, B, Bakiri, F, Brottier, P, Cattolico, L, Penet, C, Bégeot, M, Naville, D, Nicolino, M, Chaussain, J-L, Weissenbach, J, Munnich, A & Lyonnet, S 2000, 'Mutant WD-repeat protein in triple-A syndrome', *Nature Genetics*, vol. 26, no. 3, pp. 332–335.

Turner, C & Schapira, AHV 2010, 'Mitochondrial matters of the brain: the role in Huntington's disease', *Journal of Bioenergetics and Biomembranes*, vol. 42, no. 3, pp. 193–198.

UniProt Knowledgebase 2016, *AAAS - Aladin - Homo sapiens (Human) - AAAS gene & protein*, UniProt Knowledgebase, viewed 11 November 2016, <http://www.uniprot.org/uniprot/Q9NRG9>

Urnov, FD, Rebar, EJ, Holmes, MC, Zhang, HS & Gregory, PD 2010, 'Genome editing with engineered zinc finger nucleases', *Nature Reviews Genetics*, vol. 11, no. 9, pp. 636–646.

Van Dellen, A, Grote, HE & Hannan, AJ 2005, 'Gene–Environment Interactions, Neuronal Dysfunction and Pathological Plasticity in Huntington's Disease', *Clinical and Experimental Pharmacology and Physiology*, vol. 32, no. 12, pp. 1007–1019.

Vishnu, VY, Modi, M, Prabhakar, S, Bhansali, A & Goyal, MK 2014, '"A" motor neuron disease', *Journal of the Neurological Sciences*, vol. 336, no. 1–2, pp. 251–253.

Walker, FO 2007, 'Huntington's disease', *Lancet (London, England)*, vol. 369, no. 9557, pp. 218–228.

Wang, H, Bian, X, Xia, L, Ding, X, Müller, R, Zhang, Y, Fu, J & Stewart, AF 2014, 'Improved seamless mutagenesis by recombineering using ccdB for counterselection', *Nucleic Acids Research*, vol. 42, no. 5, p. e37.

Wang, L, Li, F, Dang, L, Liang, C, Wang, C, He, B, Liu, J, Li, D, Wu, X, Xu, X, Lu, A & Zhang, G 2016, '*In Vivo* Delivery Systems for Therapeutic Genome Editing', *International Journal of Molecular Sciences*, vol. 17, no. 5, viewed 25 November 2016, <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4881452/>.

Wert, G de & Mummery, C 2003, 'Human embryonic stem cells: research, ethics and policy', *Human Reproduction*, vol. 18, no. 4, pp. 672–682.

Winner, B, Kohl, Z & Gage, FH 2011, 'Neurodegenerative disease and adult neurogenesis', *European Journal of Neuroscience*, vol. 33, no. 6, pp. 1139–1151.

WHO 2016, *Neurological disorders affect millions globally: WHO report* WHO, viewed 24 November 2016, <http://www.who.int/mediacentre/news/releases/2007/pr04/en/>.

Wu, X, Kriz, AJ & Sharp, PA 2014, 'Target specificity of the CRISPR-Cas9 system', *Quantitative biology*, vol. 2, no. 2, pp. 59–70.

Yang, X 2015, 'Applications of CRISPR-Cas9 mediated genome engineering', *Military Medical Research*, vol. 2, viewed 25 November 2016, <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4433013/>.

Yoon, G, Kramer, J, Zanko, A, Guzijan, M, Lin, S, Foster-Barber, A & Boxer, AL 2006, 'Speech and language delay are early manifestations of juvenile-onset Huntington disease', *Neurology*, vol. 67, no. 7, pp. 1265–1267.

Young, AB 2003, 'Huntingtin in health and disease', *Journal of Clinical Investigation*, vol. 111, no. 3, pp. 299–302.

Young, MA, Larson, DE, Sun, C-W, George, DR, Ding, L, Miller, CA, Lin, L, Pawlik, KM, Chen, K, Fan, X, Schmidt, H, Kalicki-Veizer, J, Cook, LL, Swift, GW, Demeter, RT, Wendl, MC, Sands, MS, Mardis, ER, Wilson, RK, Townes, TM & Ley, TJ 2012, 'Background mutations in parental cells account for most of the genetic heterogeneity of induced pluripotent stem cells', *Cell Stem Cell*, vol. 10, no. 5, pp. 570–582.

Zacharias, DG, Nelson, TJ, Mueller, PS & Hook, CC 2011, 'The Science and Ethics of Induced Pluripotency: What Will Become of Embryonic Stem Cells?', *Mayo Clinic Proceedings*, vol. 86, no. 7, pp. 634–640.

Zeitlin, S, Liu, JP, Chapman, DL, Papaioannou, VE & Efstratiadis, A 1995, 'Increased apoptosis and early embryonic lethality in mice nullizygous for the Huntington's disease gene homologue', *Nature Genetics*, vol. 11, no. 2, pp. 155–163.

Zuris, JA, Thompson, DB, Shu, Y, Guilinger, JP, Bessen, JL, Hu, JH, Maeder, ML, Joung, JK, Chen, Z-Y & Liu, DR 2015, 'Efficient Delivery of Genome-Editing Proteins *In Vitro* and *In Vivo*', *Nature biotechnology*, vol. 33, no. 1, pp. 73–80.

Appendices

PCR programs

Table A1: InFusion PCR program

Step	Temperature	Time
1	98ºC	10 sec
2	60ºC	10 sec
3	72ºC	30 sec
4	Go to Step 1, 34 times	
5	72ºC	10 min
6	4ºC	∞

Table A2: Short-range (SR; amplicon size< 1kb) and Mid-range (MIDR; for amplicon size= 1-5kb) PCR program

Step	Temperature	Time
1	93ºC	3 min
2	92ºC	15 sec
3	65ºC	30 sec
	-1ºC/ cycle	
4	65ºC	2 min (SR) 4min (MIDR)
5	Go to Step 2, 8 times	
6	92ºC	15 sec
7	55ºC	30 sec
8	65ºC	2 min (SR) 4min (MIDR)
	+20 sec/ cycle	
9	Go to Step 6, 30 times	
10	65ºC	9 min
11	10ºC	∞

Table A3: Recombineering PCR program

Step	Temperature	Time
1	95°C	2 min
2	95°C	30 sec
3	60°C	30 sec
4	72°C	2 min 30 sec
5	Go to Step 2, 4 times	
6	95°C	30 sec
7	72°C	2 min 30 sec
	+5 sec/ cycle	
8	Go to Step 6, 27 times	
9	72°C	7 min
10	4°C	∞

Table A4: Sequencing PCR program

Step	Temperature	Time
1	96°C	45 sec
2	92°C	10 sec
3	52°C	10 sec
4	60°C	2 min
5	Go to Step 2, 59 times	
6	10°C	∞

Primers and oligo sequences

Table A5: AAAS2 KO oligos and primers

Oligo	Sequence
CRISPR gRNA oligos (U6_gRNA (AU flip) cloning append is underlined)	
CRISPR 1	5'-ACCGGATTTGGTGAGGCATGGCAAA
CRISPR 2	5'-ACCGGCACAAAAGTGAGGAGTGTGA

CRISPR 3	5’- ACCGGATCCCCTAAAGACCCCTGGA
CRISPR 4	5’- ACCGGTGAAGGCAGTTCTTGTGCCA
InFusion PCR primers (pUC19_RV InFusion cloning append is underlined)	
5F	5'-GCCAGTGAATTCGATGGGAGGATTAATTTGGTACTTCCCT
3R	5'-TACGCCAAGCTTGATTGTAGAGAAAAGAACATGGTTTCGC
Recombineering primers (50-bp append from AAAS InFusion clone is underlined)	
U5	5’-TTTGTGGAAATGGGAAACAGACCTCTGGAATCTCTTATACTTAGCCCAGCAAGGCGCATAACGATACCAC
D3	5’ TTGAAGAACACCCCAAGGTAAAGGGGTGTTGACTCATTTTCCAGGATCTGCCGCCTACTGCGACTATAGA
PCR primers	
F1	5’- CATGGTGTTCCTGGGAAGGACAAGG
R1	5’- GAAGTGGTAATGCCATGAGGTGGGG
F2	5’- GTCTGGCGGAGTGGAAACAGCTTGG
R2	5’- GAAGTTGCAGTGAGCTGAGCACGCC
F3	5’- TCCAGCCAGCTGCCTGTGACATAGC
R3	5’- TCAGTGCCACTGCACTGGCACAATC
Sequencing primers	
Seq_F	5’- GAGATGGCACGGAATTAAAG
Seq_R	5’- TCTGCAGACTGTGACCCAGG

Table A6: AAAS1 point mutation oligos and primers

Oligo	**Sequence**
ssODN (101 bases)	
C>A mutation oligo	5’–TACTGCCCGTCACCAGCTCGTTATTGTGCTCATATAGGGTGACTT**T**ACCCCGAGGCGGTGGAGGAGGGAACAACCCCAGAGAGCACATCTTGCCGGTTCG – 3’

C>A + PAM mutation oligo	5'– TACTGCCCGTCACCAGCTCGTTATTGTGCTCATATAGGGTG ACTT**T**ACCCCGAG**A**CGGTGGAGGAGGGAACAACCCCAGAG AGCACATCTTGCCGGTTCG – 3'
crRNA (20 base guide sequence)	5' – TATAGGGTGACTTGACCCCG… – 3'
PCR primers	
F1	5'- GCAGCACTGTTCCTCCTCTCTGAGG
R1	5'- CTTGTCCTTCCCAGGAACACCATGG
F2	5'- AATTTGGTACTTCCCTGAATGTGGC
R2	5'- TAGTTCTCCCGCATCCTCACACTCC
Sequencing primers	
Seq_F	5'- GCGGTCTGTGCCGTTCCGGC
Seq_R	5'- TCCCTCCTCGCCCTGGCCAC

Table A7: HTT oligos and primers

Oligo	**Sequence (20 base guide sequence)**
CRISPR gRNA oligos (U6_gRNA (AU flip) cloning append is underlined)	
CRISPR 1	5'-ACCGGCCTCCGGGGACTGCCGTGC
CRISPR 2	5'-ACCGGTCGCCGGCCCGCAGGCTGC
crRNAs	
crRNA 1	5'-GCCTCCGGGGACTGCCGTGC…
crRNA 2	5'-GTCGCCGGCCCGCAGGCTGC…
crRNA 3	5'-GAAGGACTTGAGGGACTCGA…
crRNA 4	5'-CTTTTCCAGGGTCGCCATGG…
InFusion PCR primers (pUC19_RV InFusion cloning append is underlined)	
F1	5'-GCCAGTGAATTCGATACGCCCCTACCTCACCAC
R1	5'-TACGCCAAGCTTGATTCAGGCTGTTTTAAGTGCCAC
F2	5'-GCCAGTGAATTCGATCCATTACAGTCTCACCACGC
R2	5'-TACGCCAAGCTTGATACTCATTCAAACGCCTGCAG
PCR primers	

F3	5’- TCGCCACGCCTCCCTTACCATGCAG
R3	5’- CCACAACTCATTCAAACGCCTGCAG
F4	5’- TCACACTTGGGGTCCTCAGGTCGTG
R4	5’- AACCTCCCCATCAGCAACGTGTTGG
F5	5’- TTTTACCTGCGGCCCAGAGC
R5	5’- CAAACTCACGGTCGGTGCAG
Sequencing primers	
Seq_F	5’-CAGAGCCCCATTCATTGCC
Seq_R	5’-CCCAAACTCACGGTCGGT

Table A8: Plasmid specific PCR and sequencing primers

Puro-cassette PCR primers	
Forward (EF)	5’-CATGTCTGGATCCGGGGGTACCGCGTCGAG
Reverse (ER)	5’- GCGATCTCTGGGTTCTACGTTAGTG
pUC19_RV sequencing primers	
pUC19_for	5’- AACTGTTGGGAAGGGCGATC
pUC19_rev	5’- GTTAGCTCACTCATTAGGCAC
U6_gRNA (AU flip) sequencing primers	
U6_for	5’- TCAGGAGAGCGTTCACCGAC
U6_rev	5’- CTGTTTATGTAAGCAGACAG

Plasmid maps and sequences

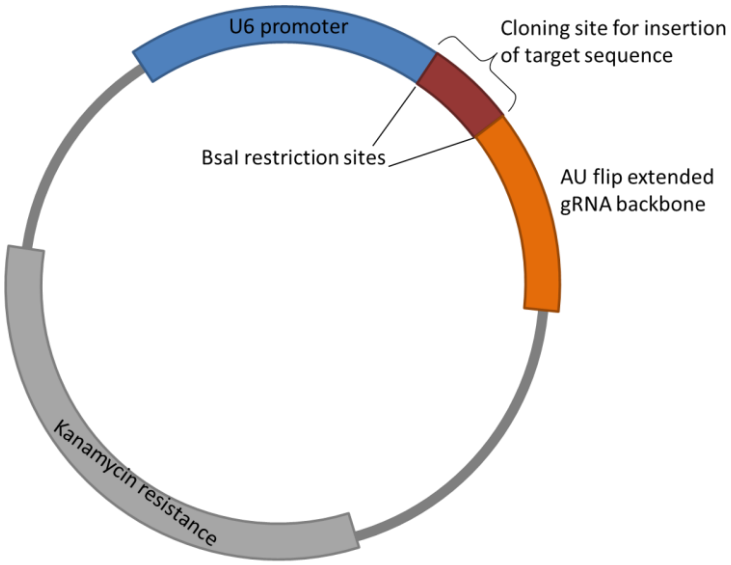


Figure A9: U6_gRNA (AU flip) expression plasmid containing extended guide RNA backbone

Sequence of U6_gRNA (AU flip) expression plasmid

CTTTCCTGCGTTATCCCCTGATTCTGTGGATAACCGTATTACCGCTAGCCAGGAAGAGT
TTGTAGAAACGCAAAAAGGCCATCCGTCAGGATGGCCTTCTGCTTAGTTTGATGCCTG
GCAGTTTATGGCGGGCGTCCTGCCCGCCACCCTCCGGGCCGTTGCTTCACAACGTTCAA
ATCCGCTCCCGGCGGATTTGTCCTACTCAGGAGAGCGTTCACCGACAAACAACAGATA
AAACGAAAGGCCCAGTCTTCCGACTGAGCCTTTCGTTTTATTTGATGCCTGGCAGTTCC
CTACTCTCGCGTTAACGctagcatggatctcgggccATTAACCCTCACTAAAGGGAAAGGTCGGG
CAGGAAGAGGGCCTATTTCCCATGATTCCTTCATATTTGCATATACGATACAAGGCTGT
TAGAGAGATAATTAGAATTAATTTGACTGTAAACACAAAGATATTAGTACAAAATACG
TGACGTAGAAAGTAATAATTTCTTGGGTAGTTTGCAGTTTTAAAATTATGTTTTAAAAT
GGACTATCATATGCTTACCGTAACTTGAAAGTATTTCGATTTCTTGGCTTTATATATCTT
GTGGAAAGGACGAAACACCGggagaccgaattcgagagggtctcaGTTTAAGAGCTATGCTGGAAA
CAGCATAGCAAGTTTAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAG
TCGGTGCTTTTTTTTTCATATGTGTCACCTAAATtgcagctcTGGCCCGTGTCTCAAAATCTC
TGATGTTACATTGCACAAGATAAAAATATATCATCATGAACAATAAAACTGTCTGCTT
ACATAAACAGTAATACAAGGGGTGTTATGAGCCATATTCAACGGGAAACGTCGAGGC
CGCGATTAAATTCCAACATGGATGCTGATTTATATGGGTATAAATGGGCTCGCGATAA
TGTCGGGCAATCAGGTGCGACAATCTATCGCTTGTATGGGAAGCCCGATGCGCCAGAG
TTGTTTCTGAAACATGGCAAAGGTAGCGTTGCCAATGATGTTACAGATGAGATGGTCA
GACTAAACTGGCTGACGGAATTTATGCCTCTTCCGACCATCAAGCATTTTATCCGTACT
CCTGATGATGCATGGTTACTCACCACTGCGATCCCCGGAAAAACAGCATTCCAGGTAT
TAGAAGAATATCCTGATTCAGGTGAAAATATTGTTGATGCGCTGGCAGTGTTCCTGCG
CCGGTTGCATTCGATTCCTGTTTGTAATTGTCCTTTTAACAGCGATCGCGTATTTCGTCT
CGCTCAGGCGCAATCACGAATGAATAACGGTTTGGTTGATGCGAGTGATTTTGATGAC
GAGCGTAATGGCTGGCCTGTTGAACAAGTCTGGAAAGAAATGCATAAACTTTTGCCAT
TCTCACCGGATTCAGTCGTCACTCATGGTGATTTCTCACTTGATAACCTTATTTTTGACG
AGGGGAAATTAATAGGTTGTATTGATGTTGGACGAGTCGGAATCGCAGACCGATACCA
GGATCTTGCCATCCTATGGAACTGCCTCGGTGAGTTTTCTCCTTCATTACAGAAACGGC
TTTTTCAAAAATATGGTATTGATAATCCTGATATGAATAAATTGCAGTTTCATTTGATG
CTCGATGAGTTTTTCTAATCAGAATTGGTTAATTGGTTGTAACACTGGCAGAGCATTAC
GCTGACTTGACGGGACGGCGCAAGCTCATGACCAAAATCCCTTAACGTGAGTTTTCGT
TCCACTGAGCGTCAGACCCCGTAGAAAAGATCAAAGGATCTTCTTGAGATCCTTTTTTT
CTGCGCGTAATCTGCTGCTTGCAAACAAAAAAACCACCGCTACCAGCGGTGGTTTGTT

TGCCGGATCAAGAGCTACCAACTCTTTTTCCGAAGGTAACTGGCTTCAGCAGAGCGCA
GATACCAAATACTGTCCTTCTAGTGTAGCCGTAGTTAGGCCACCACTTCAAGAACTCTG
TAGCACCGCCTACATACCTCGCTCTGCTAATCCTGTTACCAGTGGCTGCTGCCAGTGGC
GATAAGTCGTGTCTTACCGGGTTGGACTCAAGACGATAGTTACCGGATAAGGCGCAGC
GGTCGGGCTGAACGGGGGGTTCGTGCACACAGCCCAGCTTGGAGCGAACGACCTACA
CCGAACTGAGATACCTACAGCGTGAGCTATGAGAAAGCGCCACGCTTCCCGAAGGGA
GAAAGGCGGACAGGTATCCGGTAAGCGGCAGGGTCGGAACAGGAGAGCGCACGAGG
GAGCTTCCAGGGGGAAACGCCTGGTATCTTTATAGTCCTGTCGGGTTTCGCCACCTCTG
ACTTGAGCGTCGATTTTTGTGATGCTCGTCAGGGGGGCGGAGCCTATGGAAAAACGCC
AGCAACGCGGCCTTTTTACGGTTCCTGGCCTTTTGCTGGCCTTTTGCTCACATGTT

Legend: Red – AU flip extended gRNA backbone
Yellow highlighted: cloning site for insertion of target sequence

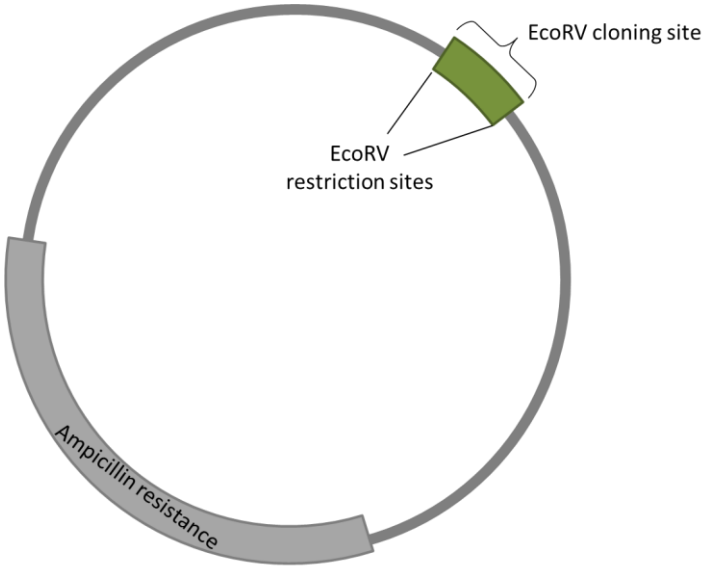


Figure A10: pUC19_RV plasmid containing EcoRV linearization site

Sequence of pUC19_RV plasmid

TCGCGCGTTTCGGTGATGACGGTGAAAACCTCTGACACATGCAGCTCCCGGAGACGGT
CACAGCTTGTCTGTAAGCGGATGCCGGGAGCAGACAAGCCCGTCAGGGCGCGTCAGC
GGGTGTTGGCGGGTGTCGGGGCTGGCTTAACTATGCGGCATCAGAGCAGATTGTACTG
AGAGTGCACCATATGCGGTGTGAAATACCGCACAGATGCGTAAGGAGAAAATACCGC
ATCAGGCGCCATTCGCCATTCAGGCTGCGCAACTGTTGGGAAGGGCGATCGGTGCGGG
CCTCTTCGCTATTACGCCAGCTGGCGAAAGGGGGATGTGCTGCAAGGCGATTAAGTTG
GGTAACGCCAGGGTTTTCCCAGTCACGACGTTGTAAAACGACGGCCAGTGAATTCGAT
ATCGGCGCGCCGATATCAAGCTTGGCGTAATCATGGTCATAGCTGTTTCCTGTGTGAAA
TTGTTATCCGCTCACAATTCCACACAACATACGAGCCGGAAGCATAAAGTGTAAAGCC
TGGGGTGCCTAATGAGTGAGCTAACTCACATTAATTGCGTTGCGCTCACTGCCCGCTTT
CCAGTCGGGAAACCTGTCGTGCCAGCTGCATTAATGAATCGGCCAACGCGCGGGGAG
AGGCGGTTTGCGTATTGGGCGCTCTTCCGCTTCCTCGCTCACTGACTCGCTGCGCTCGG
TCGTTCGGCTGCGGCGAGCGGTATCAGCTCACTCAAAGGCGGTAATACGGTTATCCAC
AGAATCAGGGGATAACGCAGGAAAGAACATGTGAGCAAAAGGCCAGCAAAAGGCCA
GGAACCGTAAAAAGGCCGCGTTGCTGGCGTTTTTCCATAGGCTCCGCCCCCCTGACGA
GCATCACAAAAATCGACGCTCAAGTCAGAGGTGGCGAAACCCGACAGGACTATAAAG
ATACCAGGCGTTTCCCCCTGGAAGCTCCCTCGTGCGCTCTCCTGTTCCGACCCTGCCGC
TTACCGGATACCTGTCCGCCTTTCTCCCTTCGGGAAGCGTGGCGCTTTCTCATAGCTCA
CGCTGTAGGTATCTCAGTTCGGTGTAGGTCGTTCGCTCCAAGCTGGGCTGTGTGCACGA

ACCCCCCGTTCAGCCCGACCGCTGCGCCTTATCCGGTAACTATCGTCTTGAGTCCAACC
CGGTAAGACACGACTTATCGCCACTGGCAGCAGCCACTGGTAACAGGATTAGCAGAG
CGAGGTATGTAGGCGGTGCTACAGAGTTCTTGAAGTGGTGGCCTAACTACGGCTACAC
TAGAAGAACAGTATTTGGTATCTGCGCTCTGCTGAAGCCAGTTACCTTCGGAAAAAGA
GTTGGTAGCTCTTGATCCGGCAAACAAACCACCGCTGGTAGCGGTGGTTTTTTTGTTTG
CAAGCAGCAGATTACGCGCAGAAAAAAAGGATCTCAAGAAGATCCTTTGATCTTTTCT
ACGGGGTCTGACGCTCAGTGGAACGAAAACTCACGTTAAGGGATTTTGGTCATGAGAT
TATCAAAAAGGATCTTCACCTAGATCCTTTTAAATTAAAAATGAAGTTTTAAATCAATC
TAAAGTATATATGAGTAAACTTGGTCTGACAGTTACCAATGCTTAATCAGTGAGGCAC
CTATCTCAGCGATCTGTCTATTTCGTTCATCCATAGTTGCCTGACTCCCCGTCGTGTAGA
TAACTACGATACGGGAGGGCTTACCATCTGGCCCCAGTGCTGCAATGATACCGCGAGA
CCCACGCTCACCGGCTCCAGATTTATCAGCAATAAACCAGCCAGCCGGAAGGGCCGAG
CGCAGAAGTGGTCCTGCAACTTTATCCGCCTCCATCCAGTCTATTAATTGTTGCCGGGA
AGCTAGAGTAAGTAGTTCGCCAGTTAATAGTTTGCGCAACGTTGTTGCCATTGCTACAG
GCATCGTGGTGTCACGCTCGTCGTTTGGTATGGCTTCATTCAGCTCCGGTTCCCAACGA
TCAAGGCGAGTTACATGATCCCCCATGTTGTGCAAAAAAGCGGTTAGCTCCTTCGGTC
CTCCGATCGTTGTCAGAAGTAAGTTGGCCGCAGTGTTATCACTCATGGTTATGGCAGCA
CTGCATAATTCTCTTACTGTCATGCCATCCGTAAGATGCTTTTCTGTGACTGGTGAGTA
CTCAACCAAGTCATTCTGAGAATAGTGTATGCGGCGACCGAGTTGCTCTTGCCCGGCG
TCAATACGGGATAATACCGCGCCACATAGCAGAACTTTAAAAGTGCTCATCATTGGAA
AACGTTCTTCGGGGCGAAAACTCTCAAGGATCTTACCGCTGTTGAGATCCAGTTCGAT
GTAACCCACTCGTGCACCCAACTGATCTTCAGCATCTTTTACTTTCACCAGCGTTTCTG
GGTGAGCAAAAACAGGAAGGCAAAATGCCGCAAAAAAGGGAATAAGGGCGACACGG
AAATGTTGAATACTCATACTCTTCCTTTTTCAATATTATTGAAGCATTTATCAGGGTTAT
TGTCTCATGAGCGGATACATATTTGAATGTATTTAGAAAAATAAACAAATAGGGGTTC
CGCGCACATTTCCCCGAAAAGTGCCACCTGACGTCTAAGAAACCATTATTATCATGAC
ATTAACCTATAAAAATAGGCGTATCACGAGGCCCTTTCGTC

Legend: Yellow highlighted: EcoRV linearization site for cloning

A11: Sequence of pL1L2_ EF1α-puro-polyA plasmid

CTTTCCTGCGTTATCCCCTGATTCTGTGGATAACCGTATTACCGCTAGCATGGATCTCG
GGGACGTCTAACTACTAAGCGAGAGTAGGGAACTGCCAGGCATCAAATAAAACGAAA
GGCTCAGTCGGAAGACTGGGCCTTTCGTTTTATCTGTTGTTTGTCGGTGAACGCTCTCC
TGAGTAGGACAAATCCGCCGGGAGCGGATTTGAACGTTGTGAAGCAACGGCCCGGAG
GGTGGCGGGCAGGACGCCCGCCATAAACTGCCAGGCATCAAACTAAGCAGAAGGCCA
TCCTGACGGATGGCCTTTTTGCGTTTCTACAAACTCTTCCTGTTAGTTAGTTACTTAAGC
TCGGGCCCCAAATAATGATTTTATTTTGACTGATAGTGACCTGTTCGTTGCAACAAATT
GATAAGCAATGCTTTTTTATAATGCCAACTTTGTACAAAAAAGCAGGCTGGCGCCGGA
ACCGAAGTTCCTATTCCGAAGTTCCTATTCTCTAGAAAGTATAGGAACTTCGAACCAA
ACGGCGCGCCGCGGCCGCCCTAAGGCGAGTAATTCATACAAAAGGACTCGCCCCTGCC
TTGGGGAATCCCAGGGACCGTCGTTAAACTCCCACTAACGTAGAACCCAGAGATCGCT
GCGTTCCCGCCCCCTCACCCGCCCGCTCTCGTCATCACTGAGGTGGAGAAGAGCATGC
GTGAGGCTCCGGTGCCCGTCAGTGGGCAGAGCGCACATCGCCCACAGTCCCCGAGAA
GTTGGGGGGAGGGGTCGGCAATTGAACCGGTGCCTAGAGAAGGTGGCGCGGGGTAAA
CTGGGAAAGTGATGTCGTGTACTGGCTCCGCCTTTTTCCCGAGGGTGGGGGAGAACCG
TATATAAGTGCAGTAGTCGCCGTGAACGTTCTTTTTCGCAACGGGTTTGCCGCCAGAAC
ACAGGTAAGTGCCGTGTGTGGTTCCCGCGGGCCTGGCCTCTTTACGGGTTATGGCCCTT
GCGTGCCTTGAATTACTTCCACGCCCCTGGCTGCAGTACGTGATTCTTGATCCCGAGCT
TCGGGTTGGAAGTGGGTGGGAGAGTTCGAGGCCTTGCGCTTAAGGAGCCCCTTCGCCT
CGTGCTTGAGTTGAGGCCTGGCTTGGGCGCTGGGGCCGCCGCGTGCGAATCTGGTGGC
ACCTTCGCGCCTGTCTCGCTGCTTTCGATAAGTCTCTAGCCATTTAAAATTTTTGATGAC
CTGCTGCGACGCTTTTTTTCTGGCAAGATAGTCTTGTAAATGCGGGCCAAGATCTGCAC
ACTGGTATTTCGGTTTTTGGGGCCGCGGGCGGCGACGGGGCCCGTGCGTCCCAGCGCA

CATGTTCGGCGAGGCGGGGCCTGCGAGCGCGGCCACCGAGAATCGGACGGGGGTAGT
CTCAAGCTGGCCGGCCTGCTCTGGTGCCTGGCCTCGCGCCGCCGTGTATCGCCCCGCCC
TGGGCGGCAAGGCTGGCCCGGTCGGCACCAGTTGCGTGAGCGGAAAGATGGCCGCTT
CCCGGCCCTGCTGCAGGGAGCTCAAAATGGAGGACGCGGCGCTCGGGAGAGCGGGCG
GGTGAGTCACCCACACAAAGGAAAAGGGCCTTTCCGTCCTCAGCCGTCGCTTCATGTG
ACTCCACGGAGTACCGGGCGCCGTCCAGGCACCTCGATTAGTTCTCGAGCTTTTGGAG
TACGTCGTCTTTAGGTTGGGGGGAGGGGTTTTATGCGATGGAGTTTCCCCACACTGAGT
GGGTGGAGACTGAAGTTAGGCCAGCTTGGCACTTGATGTAATTCTCCTTGGAATTTGCC
CTTTTTGAGTTTGGATCTTGGTTCATTCTCAAGCCTCAGACAGTGGTTCAAAGTTTTTTT
CTTCCATTTCAGGTGTCGTGATCTAGATGTCGAGCGGCCGCGGCGGCGCCCTATAAAA
CCCAGCGGCGCGACGCGCCACCACCGCCGAGACCGCGTCCGCCCCGCGAGCACAGAG
CCTCGCCTTTGCCGATCCTCTAGAGTCGAGATCCGCCGCCACCATGaccgagtacaagcccacggt
gcgcctcgccacccgcgacgacgtccccagggccgtacgcaccctcgccgccgcgttcgccgactaccccgccacgcgccacaccgtcga
tccAgaTcgccacatcgagcgggtcaccgagctgcaagaactcttcctcacgcgcgtcgggctcgacatcggcaaggtgtgggtcgcgga
cgacggcgccgcggtggcggtctggaccacgccggagagcgtcgaagcgggggcggtgttcgccgagatcggcccgcgcatggccgag
ttgagcggttcccggctggccgcgcagcaacagatggaaggcctcctggcgccgcaccggcccaaggagcccgcgtggttcctggccacc
gtcggcgtctcgcccgaccaccagggcaagggtctgggcagcgccgtcgtgctccccggagtggaggcggccgagcgcgccggggtgcc
cgccttcctggagacctccgcgccccgcaacctccccttctacgagcggctcggcttcaccgtcaccgccgacgtcgaggtgcccgaaggac
cgcgcacctggtgcatgacccgcaagcccggtgcctgaTAATGATCATAATCAGCCATATCACATCTGTAGA
GGTTTTACTTGCTTTAAAAAACCTCCCACACCTCCCCCTGAACCTGAAACATAAAATGA
ATGCAATTGTTGTTGTTAACTTGTTTATTGCAGCTTATAATGGTTACAAATAAAGCAAT
AGCATCACAAATTTCACAAATAAAGCATTTTTTTCACTGCATTCTAGTTGTGGTTTGTC
CAAACTCATCAATGTATCTTATCATGTCTGGATCCGGGGGTACCGCGTCGAGAAGTTCC
TATTCCGAAGTTCCTATTCTCTAGAAAGTATAGGAACTTCGTCGAGATAACTTCGTATA
GCATACATTATACGAAGTTATGTCGAGATATCTAGACCCAGCTTTCTTGTACAAAGTTG
GCATTATAAGAAAGCATTGCTTATCAATTTGTTGCAACGAACAGGTCACTATCAGTCA
AAATAAAATCATTATTTGCCATCCAGCTGCAGCTCTGGCCCGTGTCTCAAAATCTCTGA
TGTTACATTGCACAAGATAAAAGCTTGTATATCCATTTTCGGATCTGATCAGCACGTGT
TGACAATTAATCATCGGCATAGTATATCGGCATAGTATAATACGACAAGGTGAGGAAC
TAAACCATGGGTACCACTCTTGACGACACGGCTTACCGGTACCGCACCAGTGTCCCGG
GGGACGCCGAGGCCATCGAGGCACTGGATGGGTCCTTCACTACCGACACCGTCTTCCG
CGTCACCGCCACCGGGGACGGCTTCACCCTGCGGGAGGTGCCGGTGGACCCGCCCCTG
ACCAAGGTGTTCCCCGACGACGAATCGGACGACGAATCGGACGACGGGGAGGACGGT
GACCCGGACTCTCGGACGTTCGTCGCGTACGGGGACGACGGCGACCTGGCGGGCTTCG
TGGTCGTCTCGTACTCCGGCTGGAACCGCCGGCTGACCGTCGAGGACATCGAGGTCGC
CCCGGAGCACCGGGGGCACGGGGTCGGGCGCGCGTTGATGGGGCTCGCGACGGAGTT
CGCCCGCGAGCGGGGTGCCGGGCACCTCTGGCTGGAGGTCACCAACGTCAACGCACC
GGCGATCCACGCGTACCGGCGGATGGGGTTCACCCTCTGCGGCCTGGACACCGCCCTG
TACGACGGCACCGCCTCGGACGGCGAGCAGGCGCTCTACATGAGCATGCCCTGCCCCT
AAGCACTTCGTGGCCGAGGAGCAGGACTGACACGTTCAGAATTGGTTAATTGGTTGTA
ACATTATTCAGATTGGGCCCCGTTCCACTGAGCGTCAGACCCCGTAGAAAAGATCAAA
GGATCTTCTTGAGATCCTTTTTTTCTGCGCGTAATCTGCTGCTTGCAAACAAAAAAACC
ACCGCTACCAGCGGTGGTTTGTTTGCCGGATCAAGAGCTACCAACTCTTTTTCCGAAGG
TAACTGGCTTCAGCAGAGCGCAGATACCAAATACTGTTCTTCTAGTGTAGCCGTAGTTA
GGCCACCACTTCAAGAACTCTGTAGCACCGCCTACATACCTCGCTCTGCTAATCCTGTT
ACCAGTGGCTGCTGCCAGTGGCGATAAGTCGTGTCTTACCGGGTTGGACTCAAGACGA
TAGTTACCGGATAAGGCGCAGCGGTCGGGCTGAACGGGGGGTTCGTGCACACAGCCC
AGCTTGGAGCGAACGACCTACACCGAACTGAGATACCTACAGCGTGAGCTATGAGAA
AGCGCCACGCTTCCCGAAGGGAGAAAGGCGGACAGGTATCCGGTAAGCGGCAGGGTC
GGAACAGGAGAGCGCACGAGGGAGCTTCCAGGGGGAAACGCCTGGTATCTTTATAGT
CCTGTCGGGTTTCGCCACCTCTGACTTGAGCGTCGATTTTTGTGATGCTCGTCAGGGGG
GCGGAGCCTATGGAAAAACGCCAGCAACGCGGCCTTTTTACGGTTCCTGGCCTTTTGCT
GGCCTTTTGCTCACATGTT

Legend: Yellow highlighted: attL1 and attL2 gateway sites respectively
Red: EF1α-puro-polyA cassette

A12: Sequence of pAAAS-EF1α-puro-polyA donor plasmid

5’-
tcgcgcgtttcggtgatgacggtgaaaacctctgacacatgcagctcccggagacggtcacagcttgtctgtaagcggatgccgggagcagac
aagcccgtcagggcgcgtcagcgggtgttggcgggtgtcggggctggcttaactatgcggcatcagagcagattgtactgagagtgcaccat
atgcggtgtgaaataccgcacagatgcgtaaggagaaaataccgcatcaggcgccattcgccattcaggctgcgcaactgttgggaagggcg
atcggtgcgggcctcttcgctattacgccagctggcgaaaggggatgtgctgcaaggcgattaagttgggtaacgccagggttttcccagtca
cgacgttgtaaaacgacggccagtgaattcGATGGGAGGATTAATTTGGTACTTCCCTGAATGTGGCTTA
TTTCGGGCTTTCTAAGCGCTAAGATAGTGAGTGGGAGAAGGTGTTCGGGAATTCGGAG
CTGGGAAAAACAGTAGTCTCATGGAAAACAACAGAAAAAAACCCCACTCCAGCCTGG
TAACCAACGGTTGGCCGGGGGACGTGTCCGCGCTGCTCCCCCTCCGGCCGCCAGGGGG
CGCAGGAAGAAGCTTTGGAGGTCTGTCCCGGGACCGCGTACTCGTCCTTCTTTCCCGTT
AGTCTTTTCTTCACTTCCGTTGAGTTCCGCCTCGCCGTTTGTCCCTTGCGGTACCCGTCC
GCATACGAATCTAGCCCGGGAACCGAGTTGCGGGAGTGCGGTCTGTGCCGTTCCGGCC
AGGAGTTTGCCGACTGCAGACGTCCTGCGAACCGGCAAGATGTGCTCTCTGGGGTTGT
TCCCTCCTCCACCGCCTCGGGGTCAAGTCACCCTATATGAGCACAATAACGAGCTGGT
GACGGGCAGTAGCTATGAGAGCCCGCCCCCCGACTTCCGGGGCCAGGTGGCCAGGGC
GAGGAGGGAGGGGACCTACAGGCATGGGGGCGGAGTCGGGGACTACTGGGGAAAGG
AGGCAGTGTGACAGGGGCAGGGTCACAGAGTCTGGCGGAGTGGAAACAGCTTGGCCC
CGGCAGGAGGCTAGAAGGGCTGGAATCTGTTTTAGGACTGGGATTGGAGTATAACGG
GGGAACGGTTAAGGAGCTGTGGAGGAGGCCTAGAGGGAATGGAGAGAATGAGGGAG
AATCTAGGGAGTGTGAGGATGCGGGAGAACTATCTAAAAGTTGGGGACAGTATCTGG
GGAGGATCAGAAGGATACAACATGGAACTTTTTTCCATGGTGTTCCTGGGAAGGACAA
GGAAGGGAAACTGACATTTATTGAGCGTTTGCTATGTAAGAAAGAGGCTGAGTCAGGT
GTTTTCTAAAAATGTCATATGGTTTCATAATCACAGCAACCCTATGAAATATTATTCTA
GTTGCACAGATACGGGAGCTAAGACTTGTTTCCCATTCTCACAAGTGTTAGCTTTGTGG
AAATGGGAAACAGACCTCTGGAATCTCTTATACTTAGCCCAGCAAGGCGCATAACGAT
ACCACGATATCAACAAGTTTGTACAAAAAAGCAGGCTGGCGCCGGAACCGAAGTTCCT
ATTCCGAAGTTCCTATTCTCTAGAAAGTATAGGAACTTCGAACCAAACGGCGCGCCGC
GGCCGCCCTAAGGCGAGTAATTCATACAAAAGGACTCGCCCCTGCCTTGGGGAATCCC
AGGGACCGTCGTTAAACTCCCACTAACGTAGAACCCAGAGATCGCTGCGTTCCCGCCC
CCTCACCCGCCCGCTCTCGTCATCACTGAGGTGGAGAAGAGCATGCGTGAGGCTCCGG
TGCCCGTCAGTGGGCAGAGCGCACATCGCCCACAGTCCCCGAGAAGTTGGGGGGAGG
GGTCGGCAATTGAACCGGTGCCTAGAGAAGGTGGCGCGGGGTAAACTGGGAAAGTGA
TGTCGTGTACTGGCTCCGCCTTTTTCCCGAGGGTGGGGGAGAACCGTATATAAGTGCA
GTAGTCGCCGTGAACGTTCTTTTTCGCAACGGGTTTGCCGCCAGAACACAGGTAAGTG
CCGTGTGTGGTTCCCGCGGGCCTGGCCTCTTTACGGGTTATGGCCCTTGCGTGCCTTGA
ATTACTTCCACGCCCCTGGCTGCAGTACGTGATTCTTGATCCCGAGCTTCGGGTTGGAA
GTGGGTGGGAGAGTTCGAGGCCTTGCGCTTAAGGAGCCCCTTCGCCTCGTGCTTGAGT
TGAGGCCTGGCTTGGGCGCTGGGGCCGCCGCGTGCGAATCTGGTGGCACCTTCGCGCC
TGTCTCGCTGCTTTCGATAAGTCTCTAGCCATTTAAAATTTTTGATGACCTGCTGCGAC
GCTTTTTTTCTGGCAAGATAGTCTTGTAAATGCGGGCCAAGATCTGCACACTGGTATTT
CGGTTTTTGGGGCCGCGGGCGGCGACGGGGCCCGTGCGTCCCAGCGCACATGTTCGGC
GAGGCGGGGCCTGCGAGCGCGGCCACCGAGAATCGGACGGGGGTAGTCTCAAGCTGG
CCGGCCTGCTCTGGTGCCTGGCCTCGCGCCGCCGTGTATCGCCCCGCCCTGGGCGGCA
AGGCTGGCCCGGTCGGCACCAGTTGCGTGAGCGGAAAGATGGCCGCTTCCCGGCCCTG
CTGCAGGGAGCTCAAAATGGAGGACGCGGCGCTCGGGAGAGCGGGCGGGTGAGTCAC
CCACACAAAGGAAAAGGGCCTTTCCGTCCTCAGCCGTCGCTTCATGTGACTCCACGGA
GTACCGGGCGCCGTCCAGGCACCTCGATTAGTTCTCGAGCTTTTGGAGTACGTCGTCTT
TAGGTTGGGGGGAGGGGTTTTATGCGATGGAGTTTCCCCACACTGAGTGGGTGGAGAC
TGAAGTTAGGCCAGCTTGGCACTTGATGTAATTCTCCTTGGAATTTGCCCTTTTTGAGT
TTGGATCTTGGTTCATTCTCAAGCCTCAGACAGTGGTTCAAAGTTTTTTTCTTCCATTTC
AGGTGTCGTGATCTAGATGTCGAGCGGCCGCGGCGGCGCCCTATAAAACCCAGCGGCG
CGACGCGCCACCACCGCCGAGACCGCGTCCGCCCCGCGAGCACAGAGCCTCGCCTTTG
CCGATCCTCTAGAGTCGAGATCCGCCGCCACCATGaccgagtacaagcccacggtgcgcctcgccacccgc

gacgacgtccccagggccgtacgcaccctcgccgccgcgttcgccgactaccccgccacgcgccacaccgtcgatccAgaTcgccacat
cgagcgggtcaccgagctgcaagaactcttcctcacgcgcgtcgggctcgacatcggcaaggtgtgggtcgcggacgacggcgccgcggt
ggcggtctggaccacgccggagagcgtcgaagcgggggcggtgttcgccgagatcggcccgcgcatggccgagttgagcggttcccggct
ggccgcgcagcaacagatggaaggcctcctggcgccgcaccggcccaaggagcccgcgtggttcctggccaccgtcggcgtctcgcccg
accaccagggcaagggtctgggcagcgccgtcgtgctccccggagtggaggcggccgagcgcgccggggtgcccgccttcctggagacc
tccgcgccccgcaacctccccttctacgagcggctcggcttcaccgtcaccgccgacgtcgaggtgcccgaaggaccgcgcacctggtgcat
gacccgcaagcccggtgcctgaTAATGATCATAATCAGCCATATCACATCTGTAGAGGTTTTACTT
GCTTTAAAAAACCTCCCACACCTCCCCCTGAACCTGAAACATAAAATGAATGCAATTG
TTGTTGTTAACTTGTTTATTGCAGCTTATAATGGTTACAAATAAAGCAATAGCATCACA
AATTTCACAAATAAAGCATTTTTTTCACTGCATTCTAGTTGTGGTTTGTCCAAACTCATC
AATGTATCTTATCATGTCTGGATCCGGGGGTACCGCGTCGAGAAGTTCCTATTCCGAAG
TTCCTATTCTCTAGAAAGTATAGGAACTTCGTGAATTCtgcGCTCCGATCGCGAGATAAC
TTCGTATAGCATACATTATACGAAGTTATGTCGAGATATCTAGACCCAGCTTTCTTGTA
CAAAGTGGTTGATATCTCTATAGTCGCAGTAGGCGGCAGATCCTGGAAAATGAGTCAA
CACCCCTTTACCTTGGGGTGTTCTTCAAAAGACTTTTATTCAGGCTGTCTCCACGAGAG
CTGTATATCCTAAATCATATGAAATTTTTAGTCCTTATTATAGAACTCAAATGCTCAAT
AAATGTCAGGTTCCCTTCCTTGCCCTTCCCAGGAACACTATGGAAAAAGTACTTCCATT
TCATCTCTAGTAACTTGCCTTGCTTTCTGGCTTGATTAGGTGAACCTGCTCCCTGCCCCC
ACCTCATGGCATTACCACTTCTGAGCTTTTACCTGAAGCTCAGCAAGATCAGATATAAG
GTTGCACTGAATAATCTCCTACATAGGTCATCCACTTTGCACTTGTAGGAGTGGAGCTA
CACTACCTGGTTTCATTTTCTTTTTTTTTTTTGAGATGGAGTTTCACTCTTGTTGCCCAG
GCTGGAGTGCAGTGGCGTGCTCAGCTCACTGCAACTTCCACCTCCCAGGTTCAAGCGA
TTCTTCTGCCTCAGTCTCCCGAGTAGCTAGGATTACAGGCATGTGCCACCACGCCCAGC
TAATTTTGTATTTTTAGTAGAGACAGGGTTTCTCCATGTTGGTCAGGCTGGTCTTGAAC
TTCCGACCTCAGATGATGTGCCCGCCTTGGCCTCCCAAAGTGCTGGGATTACAGGCGT
GAGCCACCGCGCCCGGCTCATTTTCTTTTTCTTTTTTTTTTTTTTTTTTTTTTTGAGACGG
AGTCTTGCTGTGTTGCCAGGCTGGAGTGCAGAGGTGCGATCTCGGCTCACTGCAGCCT
CCGCCTCTCTGGTTCAAGCAGTTCCCCTGCCTCAGCCTCCCGAGTAGCTGGTACTATAG
GCATGTGCCACCACGCCCAGCTAACTTTTTGTATTTTAGTAGAGATGAGATTTCACCAT
GTTGGCCAGGGTGGTCTGGCCTCCTGACCTCAGGCCTGCCTTGGCCTCCCAAAGTGCTG
AGATTACAGCTACTGCACCTGGACCTATTAATGGCACATATTAATTCCTTTAAGGGTTA
TTTATGCCAGGCACAGTGACTCACGCCTATAATCCCAGCGCTTTGGTAGGCTGAGGCC
AGCAGATTGCTTGAGCTTAGTAGTTTGAGACTAGCCTGGGGAACATGGCGAAACCATG
TTCTTTTCTCTACAATCaagcttggcgtaatcatggtcatagctgtttcctgtgtgaaattgttatccgctcacaattccacacaacat
acgagccggaagcataaagtgtaaagcctggggtgcctaatgagtgagctaactcacattaattgcgttgcgctcactgcccgctttccagtcgg
gaaacctgtcgtgccagctgcattaatgaatcggccaacgcgcggggagaggcggtttgcgtattgggcgctcttccgcttcctcgctcactga
ctcgctgcgctcggtcgttcggctgcggcgagcggtatcagctcactcaaaggcggtaatacggttatccacagaatcaggggataacgcagg
aaagaacatgtgagcaaaaggccagcaaaaggccaggaaccgtaaaaaggccgcgttgctggcgtttttccataggctccgccccctgacg
agcatcacaaaaatcgacgctcaagtcagaggtggcgaaacccgacaggactataaagataccaggcgtttcccccctggaagctccctcgtg
cgctctcctgttccgaccctgccgcttaccggatacctgtccgcctttctcccttcgggaagcgtggcgctttctatagctcacgctgtaggtatct
cagttcggtgtaggtcgttcgctccaagctgggctgtgtgcacgaaccccccgttcagcccgaccgctgcgccttatccggtaactatcgtcttg
agtccaacccggtaagacacgacttatcgccactggcagcagccactggtaacaggattagcagagcgaggtatgtaggcggtgctacagag
ttcttgaagtggtggcctaactacggctacactagaagaacagtatttggtatctgcgctctgctgaagccagttaccttcggaaaaagagttggta
gctcttgatccggcaaacaaaccaccgctggtagcggtggttttttgtttgcaagcagcagattacgcgcagaaaaaaggatctcaagaagat
cctttgatcttttctacggggtctgacgctcagtggaacgaaaactcacgttaagggattttggtcatgagattatcaaaaggatcttcacctagat
ccttttaaattaaaaatgaagttttaaatcaatctaaagtatatatgagtaaacttggtctgacagttaccaatgcttaatcagtgaggcacctatctca
gcgatctgtctatttcgttcatccatagttgcctgactccccgtcgtgtagataactacgatacgggagggcttaccatctggccccagtgctgcaa
tgataccgcgagacccacgctcaccggctccagatttatcagcaataaaccagccagccggaagggccgagcgcagaagtggtcctgcaac
tttatccgcctccatccagtctattaattgttgccgggaagctagagtaagtagttcgccagttaatagtttgcgcaacgttgttgccattgctacagg
catcgtggtgtcacgctcgtcgtttggtatggcttcattcagctccggttcccaacgatcaaggcgagttacatgatcccccatgttgtgcaaaaaa
gcggttagctccttcggtcctccgatcgttgtcagaagtaagttggccgcagtgttatcactcatggttatggcagcactgcataattctcttactgtc
atgccatccgtaagatgcttttctgtgactggtgagtactcaaccaagtcattctgagaatagtgtatgcggcgaccgagttgctcttgcccggcgt
caatacgggataataccgcgccacatagcagaactttaaaagtgctcatcattggaaaacgttcttcggggcgaaaactctcaaggatcttaccg
ctgttgagatccagttcgatgtaacccactcgtgcacccaactgatcttcagcatcttttactttcaccagcgtttctgggtgagcaaaaacaggaa
ggcaaaatgccgcaaaaaagggaataagggcgacacggaaatgttgaatactcatactcttcctttttcaatattattgaagcatttatcagggttat

tgtctcatgagcggatacatatttgaatgtatttagaaaaataaacaaatagggttccgcgcacatttccccgaaaagtgccacctgacgtctaa
gaaaccattattatcatgacattaacctataaaaataggcgtatcacgaggccctttcgtc – 3’

Legend:
Purple: InFusion primers
Green: U5 and D3 recombineering primers respectively
Red: EF1α-puro-polyA cassette
Yellow highlighted: attB1 and attB2 sites respectively
Underlined: 5’ and 3’ homology arms respectively
Blue: pUC19 seq_F and seq_R respectively
Orange: ampicillin resistance

A13: Sequence of HTT synthetic plasmid

5’-
GCTGAGCGGCGCCGCGAGTCGGCCCGAGGCCTCCGGGGACTGCCGTGCCGGGCGGGA
GACaGCCATGGCGACCCTGGAAAAGCTGATGAAGGCaTTCGAGTCCCTCAAGTCCTTC
CAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG
CAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG
CAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG
CAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCCGCCACCGCCGCCGC
CGCCGCCGCCGCCTCCTCAGCTTCCTCAGCCGCCGCCGCAGGCACAGCCGCTGCTGCC
TCAGCCGCAGCCGCCCCCGCCGCCGCCCCCGCCGCCACCCGGCCCGGCTGTGGCTGAG
GAGCCGCTGCAC -3’

Legend:
Red: cloning sites BlpI and BsgI respectively
Bold: 67-Q repeats
Underlined: CRISPR sites 4 and 3 respectively
Blue lower-case: PAM mutations

A14: Sequence of Amino acid sequences of three ALADIN isoforms recognized by the Proteintech anti-AAAS antibody and by a different monoclonal antibody (Sigma, as an example)

Q9NRG9-V1 (60kDa):
MCSLGLFPPPPPRGQVTLYEHNNELVTGSSYESPPPDFRGQWINLPVLQLTKDPLKTPGRLD
HGTRTAFIHHREQVWKRCINIWRDVGLFGVLNEIANSEEEVFEWVKTASGWALALCRWA
SSLHGSLFPHLSLRSEDLIAEFAQVTNWSSCCLRVFAWHPHTNKFAVALLDDSVRVYNASS
TIVPSLKHRLQRNVASLAWKPLSASVLAVACQSCILIWTLDPTSLSTRPSSGCAQVLSHPGH
TPVTSLAWAPSGGRLLSASPVDAAIRVWDVSTETCVPLPWFRGGGVTNLLWSPDGSKILA
TTPSAVFRVWEAQMWTCERWPTLSGRCQTGCWSPDGSRLLFTVLGEPLIYSLSFPERCGE
GKGCVGGAKSATIVADLSETTIQTPDGEERLGGEAHSMVWDPSGERLAVLMKGKPRVQD
GKPVILLFRTRNSPVFELLPCGIIQGEPGAQPQLITFHPSFNKGALLSVGWSTGRIAHIPLYFV
NAQFPRFSPVLGRAQEPPAGGGGSIHDLPLFTETSPTSAPWDPLPGPPPVLPHSPHSHL

Q9NRG9-V2 (56kDa):
MCSLGLFPPPPPRGQVTLYEHNNELVTGSSYESPPPDFRGQWINLPVLQLTKDPLKTPGRLD
HGTRTAFIHHREQVWKRCINIWRDVGLFGVLNEIANSEEEVFEWVKTASGWALALCRWA
SSLHGSLFPHLSLRSEDLIAEFAQVTNCTIVPSLKHRLQRNVASLAWKPLSASVLAVACQSC
ILIWTLDPTSLSTRPSSGCAQVLSHPGHTPVTSLAWAPSGGRLLSASPVDAAIRVWDVSTET
CVPLPWFRGGGVTNLLWSPDGSKILATTPSAVFRVWEAQMWTCERWPTLSGRCQTGCWS
PDGSRLLFTVLGEPLIYSLSFPERCGEGKGCVGGAKSATIVADLSETTIQTPDGEERLGGEA

HSMVWDPSGERLAVLMKGKPRVQDGKPVILLFRTRNSPVFELLPCGIIQGEPGAQPQLITF
HPSFNKGALLSVGWSTGRIAHIPLYFVNAQFPRFSPVLGRAQEPPAGGGGSIHDLPLFTETS
PTSAPWDPLPGPPPVLPHSPHSHL

F8VZ44 (46kDa):

MKLQTQKKRLRSEDLIAEFAQVTNWSSCCLRVFAWHPHTNKFAVALLDDSVRVYNASSTI
VPSLKHRLQRNVASLAWKPLSASVLAVACQSCILIWTLDPTSLSTRPSSGCAQVLSHPGHT
PVTSLAWAPSGGRLLSASPVDAAIRVWDVSTETCVPLPWFRGGGVTNLLWSPDGSKILAT
TPSAVFRVWEAQMWTCERWPTLSGRCQTGCWSPDGSRLLFTVLGEPLIYSLSFPERCGEG
KGCVGGAKSATIVADLSETTIQTPDGEERLGGEAHSMVWDPSGERLAVLMKGKPRVQDG
KPVILLFRTRNSPVFELLPCGIIQGEPGAQPQLITFHPSFNKGALLSVGWSTGRIAHIPLYFVN
AQFPRFSPVLGRAQEPPAGGGGSIHDLPLFTETSPTSAPWDPLPGPPPVLPHSPHSHL

Legend:

Yellow highlighted: Polyclonal (Proteintech) antibody binding sequence
Red: Possible monoclonal (Sigma) antibody binding sequence

Additional figures

Show All entries | Show/hide columns (2 hidden) | Filter

Genomic Location	Overlapping Gene(s)	Orientation	Query start	Query end	Length	Score	E-val	%ID
12:53307558-53307713 [Sequence]	AAAS	Reverse	473	524	52 [Sequence]	88.1	5e-16	100.00 [Alignment]
12:53314297-53314443 [Sequence]	AAAS	Reverse	182	230	49 [Sequence]	78.0	1e-12	100.00 [Alignment]
12:53320510-53320692 [Sequence]	AAAS	Reverse	42	100	61 [Sequence]	76.5	3e-12	77.05 [Alignment]
12:53308657-53309025 [Sequence]	AAAS	Reverse	311	382	123 [Sequence]	75.6	7e-12	46.34 [Alignment]
12:53321343-53321465 [Sequence]	AAAS	Reverse	1	41	41 [Sequence]	73.5	3e-11	100.00 [Alignment]
12:53309156-53309281 [Sequence]	AAAS	Reverse	271	312	42 [Sequence]	71.7	1e-10	100.00 [Alignment]
12:53309598-53309726 [Sequence]	AAAS	Reverse	229	271	43 [Sequence]	68.5	1e-09	97.67 [Alignment]
12:53314744-53314848 [Sequence]	AAAS	Reverse	150	184	35 [Sequence]	58.1	3e-06	91.43 [Alignment]
12:53308434-53308562 [Sequence]	AAAS	Reverse	348	394	47 [Sequence]	56.9	4e-13	80.85 [Alignment]
12:53315332-53315457 [Sequence]	AAAS	Reverse	93	134	42 [Sequence]	54.2	5e-05	76.19 [Alignment]
6:95630270-95630467 [Sequence]	RP11-374I15.1	Reverse	390	457	68 [Sequence]	50.9	2e-06	52.94 [Alignment]
12:53307976-53308191 [Sequence]	AAAS	Reverse	402	485	93 [Sequence]	50.6	7e-04	46.24 [Alignment]
12:53307845-53307928 [Sequence]	AAAS	Reverse	445	472	28 [Sequence]	49.1	0.002	100.00 [Alignment]
12:53308217-53308354 [Sequence]	AAAS	Reverse	393	444	52 [Sequence]	45.6	4e-13	55.77 [Alignment]
6:95630149-95630277 [Sequence]	RP11-374I15.1	Reverse	451	496	46 [Sequence]	28.3	2e-06	47.83 [Alignment]

Figure A15: BLAST results of the Q9NRG9 isoform against the human genome indicating 52.94% and 47.83% identities (in red arrows) to a processed pseudogene, RP11-374I15.1 on chromosome 6

List of figures and tables