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

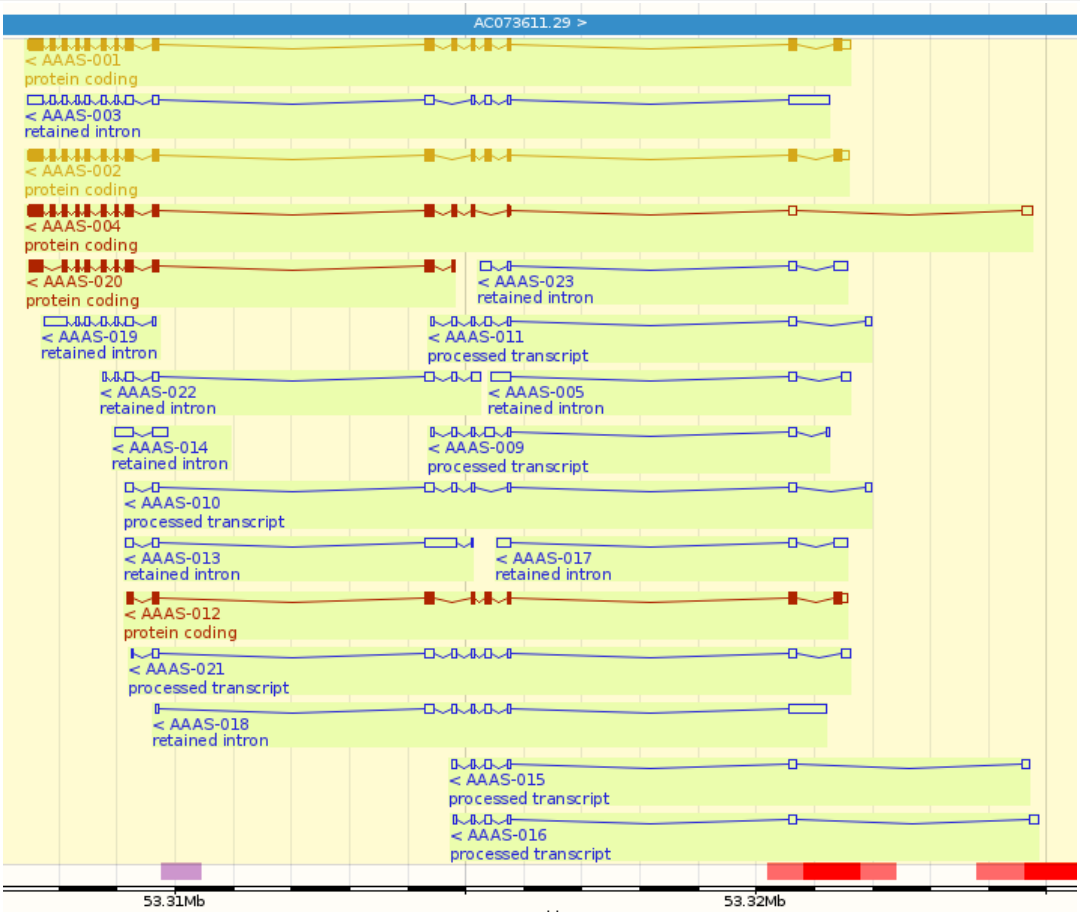

Figure 1.1: Various coding and non-coding transcripts of the *AAAS* gene encoding different ALADIN isoforms.

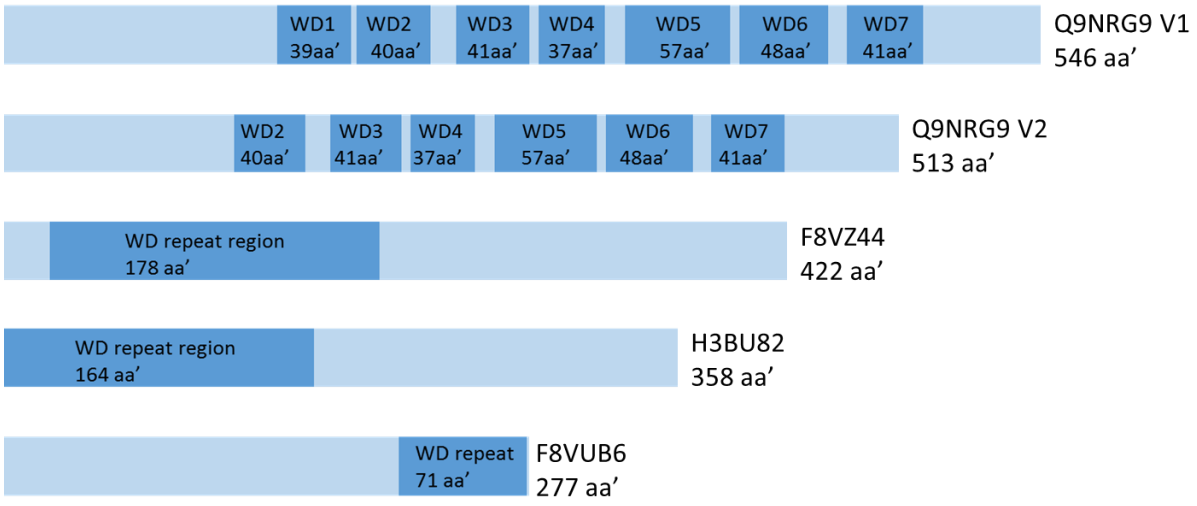

Figure 1.2: The known isoforms of ALADIN protein showing WD repeat domains (Uniprot 2016)

***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.