2019

# Genetic And Epigenetic Modifiers Of C9orf72 Expression In Neurodegenerative Disease

Christopher Cali
*University of Pennsylvania*

# Genetic And Epigenetic Modifiers Of C9orf72 Expression In Neurodegenerative Disease


## Abstract

Repeat expansion mutations in the gene C9orf72 are the most common cause of the fatal neurodegenerative diseases amyotrophic lateral sclerosis (ALS) and frontotemporal degeneration (FTD). The molecular mechanisms that contribute to these diseases are still not fully understood. In this dissertation, we explore mechanisms associated with repeat expansions in C9orf72 that alter gene expression and contribute to disease. In chapter 2, we develop a novel method of targeted DNA methylation in order to study how epigenetic changes in C9orf72 expansion carriers contribute to disease pathways. We find that C9orf72 promoter hypermethylation is sufficient to reduce gene expression and induce heterochromatin silencing of this locus in ALS patient derived cells. We also uncover a link between DNA damage repair pathways and DNA methylation where a double strand break in CpG islands can promote DNA methylation. Furthermore, we have taken advantage of these findings to develop and optimize a novel targeted DNA methylation method that utilizes homology directed repair for precise methylation editing in the absence of off-target effects. In chapter 3, we ask whether intermediate length repeat expansions in C9orf72 are associated with a different neurodegenerative disease, corticobasal degeneration (CBD). CBD shares similar clinical symptoms with Parkinson's disease but exhibits distinct tau pathology in neurons and glial cells. We find that intermediate C9orf72 repeat expansion carriers (17-30 repeats) are three times more likely to develop CBD than those with smaller repeats. While full expansion carriers with ALS/FTD tend to have reduced C9orf72 expression, we show that intermediate expansion carriers actually have increased C9orf72 mRNA and protein levels. This increase in C9orf72 expression drives aberrant gene expression profiles in vesicle trafficking, stress response and autophagy pathways. Furthermore, we find that autophagy initiated by nutrient starvation is deficient in cells that over-express C9orf72. In sum, this thesis contributes a novel method of targeted DNA methylation to the research community and shows how DNA methylation alters expression of a disease relevant gene. This work also highlights how variable lengths of the repeat expansion in C9orf72 can lead to distinctive underlying molecular mechanisms and ultimately drive risk for different neurodegenerative diseases.



## Degree Type

Dissertation

## Degree Name

Doctor of Philosophy (PhD)

## Graduate Group

Cell & Molecular Biology

## First Advisor

Edward B. Lee

## Keywords

C9orf72, DNA Methylation, HARDEN, Neurodegeneration

## Subject Categories

Genetics | Molecular Biology | Neuroscience and Neurobiology

**GENETIC AND EPIGENETIC MODIFIERS OF *C9ORF72* EXPRESSION IN NEURODEGENERATIVE DISEASE**

Christopher P. Cali

A DISSERTATION

in

Cell and Molecular Biology

Presented to the Faculties of the University of Pennsylvania

in

Partial Fulfillment of the Requirements for the

Degree of Doctor of Philosophy

2019

Supervisor of Dissertation

_____

Edward B. Lee MD, Ph.D.
Assistant Professor of Pathology and Laboratory Medicine

Graduate Group Chairperson

_____

Daniel S. Kessler, Ph.D.
Chair, Cell and Molecular Biology Graduate Group

Dissertation Committee

Marisa S. Bartolomei, Ph.D.
Perelman Professor of Cell and Developmental Biology

Nancy M. Bonini, Ph.D.
Florence R.C. Murray Professor of Biology

Roger A. Greenberg, MD, Ph.D.
J. Samuel Staub Professor of Cancer Biology

F. Bradley Johnson, MD, Ph.D.
Professor of Pathology and Laboratory Medicine

**GENETIC AND EPIGENETIC MODIFIERS OF *C9ORF72* EXPRESSION IN NEURODEGENERATIVE DISEASE**

*This work is dedicated to my grandparents for their courage in coping with devastating neurodegenerative diseases. Their experience has inspired me to pursue research that may one day help treat these terrible diseases.*

## ACKNOWLEDGMENTS

This work would not have possible without my co-workers and scientific mentors. I would like to thank my thesis advisor, Dr. Edward Lee, for his instrumental guidance and crazy (but often brilliant) ideas. I appreciate his mentorship in challenging me to do the highest quality science possible, but also knowing when to offer encouragement and confidence when things seemed bleak. I also want to specifically thank my thesis committee chair, Dr. Brad Johnson, for helping me cope with rejected papers, offering career advice and serving as an outside opinion on the quality of our work. I also thank the other members of my thesis committee, Dr. Marisa Bartolomei, Dr. Roger Greenberg and Dr. Nancy Bonini, for their enthusiasm, flexibility and helpful suggestions along the way.

To my lab-mates, past and present, I thank you for your help, encouragement and advice over the years. The original lab members, Jenny and Elaine, helped teach me numerous scientific techniques and offered helpful career advice. The more recent members, Nabil, Jess, Ashley, Aivi and Josh, each helped me in different ways. Whether it was scientific guidance, life advice or just helping me to have fun during the work day, I couldn't have gotten through the past six years without these friendships.

I also want to thank the friends I have made in my graduate group here at Penn. Having such a great group of friends has made the graduate school process more manageable. I feel fortunate to be surrounded by such smart and fun classmates who have shared in the joy (and misery) of the PhD journey. I specifically want to thank Devin, Bobby, Steve B, Steve Z, David, Colleen, Somdutta, Terra and Christin, as well as my Genetics and Epigenetics friends Mikey P, Jonathan and Priya.

I can't express enough gratitude to my wife, Meghan, for helping me to have an enjoyable experience outside of the lab these past years, despite the long hours and many disappointments. She has never once doubted my ability to finish this work, and her confidence helped me through numerous roadblocks.

To my parents, their encouragement and interest in my work was fundamental to my success. I can't thank them enough for the values they've instilled in me and for the generosity they've shown me over the years. I could not have completed this work without their support.

# ABSTRACT

## GENETIC AND EPIGENETIC MODIFIERS OF *C9ORF72* EXPRESSION IN NEURODEGENERATIVE DISEASE

Christopher P. Cali

Edward B. Lee

Repeat expansion mutations in the gene *C9orf72* are the most common cause of the fatal neurodegenerative diseases amyotrophic lateral sclerosis (ALS) and frontotemporal degeneration (FTD). The molecular mechanisms that contribute to these diseases are still not fully understood. In this dissertation, we explore mechanisms associated with repeat expansions in *C9orf72* that alter gene expression and contribute to disease. In chapter 2, we develop a novel method of targeted DNA methylation in order to study how epigenetic changes in *C9orf72* expansion carriers contribute to disease pathways. We find that *C9orf72* promoter hypermethylation is sufficient to reduce gene expression and induce heterochromatin silencing of this locus in ALS patient derived cells. We also uncover a link between DNA damage repair pathways and DNA methylation where a double strand break in CpG islands can promote DNA methylation. Furthermore, we have taken advantage of these findings to develop and optimize a novel targeted DNA methylation method that utilizes homology directed repair for precise methylation editing in the absence of off-target effects. In chapter 3, we ask whether intermediate length repeat expansions in *C9orf72* are associated with a different neurodegenerative disease, corticobasal degeneration (CBD). CBD shares similar clinical symptoms with Parkinson's disease but exhibits distinct tau pathology in neurons and glial cells. We find that intermediate *C9orf72* repeat expansion carriers (17-30 repeats) are three times more likely to develop CBD than those with smaller repeats. While full expansion carriers with ALS/FTD tend to have reduced *C9orf72* expression, we show that intermediate expansion carriers actually have increased *C9orf72* mRNA and protein levels. This increase in *C9orf72* expression drives aberrant gene expression profiles in vesicle trafficking, stress response and autophagy pathways. Furthermore, we find that

autophagy initiated by nutrient starvation is deficient in cells that over-express *C9orf72*. In sum, this thesis contributes a novel method of targeted DNA methylation to the research community and shows how DNA methylation alters expression of a disease relevant gene. This work also highlights how variable lengths of the repeat expansion in *C9orf72* can lead to distinctive underlying molecular mechanisms and ultimately drive risk for different neurodegenerative diseases.

## TABLE OF CONTENTS


**ACKNOWLEDGMENTS** ..... **IV**

**ABSTRACT** ..... **V**

**LIST OF TABLES** ..... **IX**

**LIST OF FIGURES** ..... **IX**

**CHAPTER 1: INTRODUCTION** ..... **1**

**Genetics of amyotrophic lateral sclerosis and frontotemporal degeneration** ..... **1**

**Identification of the *C9orf72* repeat expansion mutation** ..... **4**

**Microsatellite repeat expansions and disease** ..... **4**

**Loss of function mechanisms of *C9orf72* repeat expansions** ..... **7**

**C9orf72 protein function and autophagy** ..... **9**

**Gain of function mechanisms of *C9orf72* repeat expansions** ..... **11**

**Intermediate repeat expansions in *C9orf72*** ..... **14**

**Epigenetic alterations in microsatellite repeat expansion diseases** ..... **16**

**Methods of Epigenetic Editing** ..... **17**

**DNA methylation and DNA repair pathways** ..... **19**

**Objectives and Hypotheses** ..... **20**

**CHAPTER 2: TARGETED DNA METHYLATION OF *C9ORF72* VIA ENDOGENOUS DNA REPAIR PATHWAYS** ..... **23**

**Abstract** ..... **24**

**Introduction** ..... **25**

**Methods** ..... **27**

**Results** ..... **33**

**Discussion** ..... **42**

**Figures** ..... **44**

**Supplemental Figures and Tables** ..... **55**

## CHAPTER 3: C9ORF72 INTERMEDIATE REPEATS AND CORTICOBASAL DEGENERATION ..... 63

**Abstract** ..... **64**

**Introduction** ..... **65**

**Materials and Methods** ..... **67**

**Results** ..... **74**

**Discussion** ..... **83**

**Figures** ..... **87**

**Supplemental Figures and Tables** ..... **99**

## CHAPTER 4: CONCLUSIONS AND FUTURE DIRECTIONS ..... 109

**DNA methylation is linked to DNA double strand break repair** ..... **109**

**Hypermethylation of the *C9orf72* promoter leads to gene silencing** ..... **111**

**HARDEN can be applied for targeted methylation of *APP*** ..... **111**

***C9orf72* intermediate repeat expansions are a risk factor for corticobasal degeneration** ..... **114**

**Intermediate length repeat expansions increase *C9orf72* expression** ..... **115**

**Increased *C9orf72* expression disrupts autophagy** ..... **116**

**C9orf72 expression and tau aggregation** ..... **117**

**Final Remarks** ..... **119**

## APPENDIX ..... 120

**Protocols** ..... **120**

CRISPR gRNA Cloning ..... 120

DNA methylation detection via MSRE-qPCR ..... 122

Chromatin immunoprecipitation from cell lines ..... 123

Feeder Free iPSC culture ..... 126

Differentiation of iPSCs to NPCs ..... 127

## REFERENCES ..... 131

## LIST OF TABLES

**Table 2.S1: Primers and gRNA sequences……..61**
**Table 2.S2: Repair template sequences with PAM mutation highlighted……..62**
**Table 3.S1: Primers and CRISPR gRNAs……...106**
**Table 3.S2: RNA sequencing of intermediate repeat NPCs raw read counts……..107**
**Table 3.S3: Evaluation of C9orf72 Risk haplotype marker rs3849942……107**
**Table 3.S4: Repeat sizes of CBD cases with intermediate expansions……108**
**Table 3.S5: Comparison of CBD cohort vs Theuns. et al 2014 controls……108**

## LIST OF FIGURES

**Chapter 1**
**Figure 1.1: Microsatellite repeat expansions and neurodegenerative disease……5**
**Figure 1.2: Proposed mechanisms of C9orf72 repeat expansion pathogenesis……6**
**Figure 1.3: Proposed functions of C9orf72 protein……....9**
**Figure 1.4: Proposed mechanisms of targeted DNA methylation via DNA double strand break repair pathways……21**

**Chapter 2**
**Figure 2.1: Editing DNA methylation of C9orf72 using single stranded DNA templates…..45**
**Figure 2.2: Editing DNA methylation with double-stranded repair templates……47**
**Figure 2.3: DNA methylation is associated with Homology Directed Repair……..49**
**Figure 2.4: Genome-wide methylation analysis of HARDEN and Cas9-DNMT3a based methods……51**
**Figure 2.5: Editing of DNA methylation in ALS patient derived iPSCs……53**
**Supplemental Figure 2.1: Comparison of HDR efficiency……55**
**Supplemental Figure 2.2: Comparison of ssDNA vs dsDNA templates……56**
**Supplemental Figure 2.3: MSRE-qPCR of HEK293T clonal lines ……..57**
**Supplemental Figure 2.4: Effect of DNMT knockdown on targeted methylation……58**
**Supplemental Figure 2.5: Repeat Primed PCR of iPSCs derived from ALS patient……..59**
**Supplemental Figure 2.6: H3K9me3 chromatin immunoprecipitation……...60**

**Chapter 3**
**Figure 3.1: C9orf72 repeat sizing of CBD……87**
**Figure 3.2: Effect of intermediate repeats on C9orf72 expression……89**
**Figure 3.3: CRISPR/cas9 knock-in of intermediate repeats into iPSCs……...91**
**Figure 3.4: Absence of repeat pathology in intermediate expansion cases……..93**
**Figure 3.5: Elevated C9orf72 protein levels and gene expression changes in intermediate repeat NPCs……..94**
**Figure 3.6: Autophagy defects in inducible C9orf72 over-expressing cells……...96**
**Figure 3.7: Kinetics of starvation induced autophagy in C9orf72 over-expressing cells...98**
**Supplemental Figure 3.1: Repeat Expansion sizes in the CBD post mortem cerebellum cohort……99**
**Supplemental Figure 3.2: DNA methylation analysis at the C9orf72 promoter……..100**
**Supplemental Figure 3.3: Homologous template and gRNA sequence……..101**
**Supplemental Figure 3.4: Sanger sequencing of repeat edited iPSCs……102**
**Supplemental Figure 3.5: Characterization of iPSC derived neural progenitor cells…….103**

**Supplemental Figure 3.6: C9orf72 variant 3 mRNA levels from undifferentiated, repeat edited iPSCs……104**
**Supplemental Figure 3.7: Expression of C9orf72 transgene……..105**

**Chapter 4**

**Figure 4.1: Targeted DNA methylation of APP via HARDEN……..113**
**Figure 4.2: Effect of C9orf72 repeat size on expression and disease risk……115**
**Figure 4.3: Effect of C9orf72 over-expression on Tau aggregation……118**

# Chapter 1: Introduction

## Genetics of amyotrophic lateral sclerosis and frontotemporal degeneration

Amyotrophic lateral sclerosis (ALS) and frontotemporal degeneration (FTD) are fatal neurodegenerative diseases with clinical and genetic overlap[1,2]. ALS is a progressive motor neuron disease selectively affecting both upper and lower motor neurons resulting in muscle atrophy and death generally within 2-4 years after onset[3]. FTD is a disease that preferentially affects the frontal and temporal cortex in which cortical neurons degenerate resulting in language, memory and behavioral deficits[4]. Although age of onset for ALS and FTD are similar (usually 50-60 years old), patients with FTD can live 10+ years after disease onset[5]. Additionally, ALS and FTD clinical symptoms can co-occur in the same patient[3]. These diseases can be classified by whether they have a known genetic etiology. For instance, familial or hereditary disease is typically caused by a recognized genetic mutation that adheres to classic Mendelian inheritance patterns. Sporadic forms of ALS/FTD, on the other hand, usually do not have known disease-linked mutations and typically do not follow defined inheritance patterns in families, although they can still have a genetic component[6–8].

Sporadic ALS accounts for up to 90% of total ALS cases[6]. To date, several loci have been reproducibly associated with sporadic ALS, but these explain only a small percentage of the underlying genetic etiology[9]. These include an association within an intron of *UNC13A*, a gene involved in neurotransmitter realease[10–12]. Another association has been identified near the gene *SARM1*[11,13], the protein product of which has been shown to promote axonal degeneration in mice[14]. These associations occur in non-coding

regions, and thus it is still not clear whether the closest genes to the associated polymorphisms are actually those underlying ALS risk. Two loci that do have rare amino acid changes in sporadic ALS patients, however, have recently been identified[11,12]. These include polymorphisms found in *C21orf2* and *KIF5A*[11,12]. The function of C21orf2 however, is not yet clear, but it may be involved in DNA repair pathways and cilia development[15,16]. KIF5A is a microtubule-associated molecular motor that is known to function in neuronal pathways such as synaptic transmission, and knockout of the gene in mice leads to epilepsy[17]. Thus, *KIF5A* represents an interesting candidate gene to explore further. Finally, all major ALS genome-wide association studies identified variants in a region on chromosome 9, which were later found to be a risk haplotype associated with the carriers of a mutation in the gene *C9orf72*[11–13,18]. Mutations in *C9orf72* will be discussed in more detail in later sections.

As for sporadic FTD, there have only been a few identified genetic risk factors, perhaps due the clinical and neuropathological heterogeneity of this disease[19–23]. The first identified association was with *TMEM106B*[19,23], a membrane bound protein involved in lysosomal pathways[24]. Variants in *TMEM106B* were later implicated as disease modifiers in patients with familial FTD caused by *GRN* or *C9orf72* mutations, thus connecting this genetic association with known disease mechanisms[25–27]. Other associations include variants near the *HLA* locus and a locus encompassing *RAB38* and *CTSC*[21]. Additionally, variants in the ALS risk gene *UNC13A* have also been associated with FTD[2]. Several other associations have been reported but have not yet been validated in replication cohorts[28].

Much progress has been made in determining genetic causes of familial ALS and FTD. Approximately 30 genes have been implicated in familial ALS, with the most frequent being mutations in *C9orf72* and *SOD1*[29]. *SOD1,* which encodes copper-zinc superoxide dismutase, is an antioxidant enzyme that normally functions to reduce cellular superoxide levels to prevent free radical induced damage[30]. There are over 160 known mutations in *SOD1,* which account for 20% of all familial ALS cases[6]. Mutated SOD1 protein forms insoluble aggregates in mitochondria that can be detected in patient spinal cord and may be the cause of neurotoxicity[31,32]. Mutations in *TARDBP*, which encodes the RNA binding protein TDP-43, are also a cause of familial ALS and account for approximately 4% of familial ALS cases[33]. Mutations in *TARDBP* are notable because TDP-43 protein aggregates are a hallmark of sporadic ALS as well[34]. Mutations in another RNA binding protein, *FUS*¸ is also linked to familial ALS and these mutations account for 3% of cases[6]. While rare, mutations in *FUS* and *TARDBP* implicate RNA metabolism as a key disease pathway in ALS[35].

The most frequently mutated genes that cause familial FTD include *C9orf72, MAPT,* and *GRN*[22]. *MAPT* mutations, which encode the microtubule associated protein tau, account for 10-20% of familial cases[36]. Carriers of *MAPT* mutations have unique pathology in that they develop hyper-phosphorylated tau aggregates in the brain rather than TDP-43[31]. Tau aggregation is thought to be neurotoxic and is implicated in a host of other neurodegenerative diseases such as Alzheimer's disease, corticobasal degeneration, Pick's disease and progressive supranuclear palsy[37]. Mutations in *GRN* account for 5-20% of familial FTD cases and result in TDP-43 brain pathology[36]. *GRN* mutations are thought to cause disease through loss of function of the progranulin protein, which

normally may be involved in the inflammatory response to neuronal damage[38]. There are also additional very rare mutations that cause FTD in genes such as *VCP, TBK1*, *CHMP2B* and *TARDBP*[36]. Most of these rare mutations also lead to TDP-43 pathology[22].

**Identification of the *C9orf72* repeat expansion mutation**

The most recently discovered and the most common mutation underlying both ALS and FTD (~30-50% of familial cases) occurs in the gene *C9orf72*[6]. This locus was first identified as contributing to ALS and FTD by linkage analysis of large families that showed an autosomal dominant inheritance pattern for ALS and/or FTD[39,40]. The susceptibility region on chromosome nine was further refined by genome-wide association studies in European populations, which led to the discovery of a 42 variant haplotype block spanning three genes that segregated with disease risk[41,42]. The casual gene and mutation, which consisted of a $G_4C_2$ microsatellite repeat expansion in *C9orf72*, was discovered in 2011[43,44]. The hexanucleotide repeat expansion sequence can repeat hundreds to thousands of times within the first intron of the gene[45]. Several studies have attempted to find genetic or epigenetic factors that may determine whether a *C9orf72* expansion carrier develops ALS verses FTD, but these have largely been inconclusive[46,47]. To better understand the underlying pathogenic mechanisms of the *C9orf72* repeat expansion, it's important to consider related neurodegenerative diseases that are also caused by microsatellite repeat expansions.

**Microsatellite repeat expansions and disease**

Microsatellite repeat expansion diseases are a group of hereditary neurodegenerative diseases caused by repetitive DNA elements that form large

expansions within exons, introns or untranslated regions[48]. The mechanism of pathogenesis is often determined by the gene in which the expansion lies and the sequence of the expansion itself (Figure 1.1).

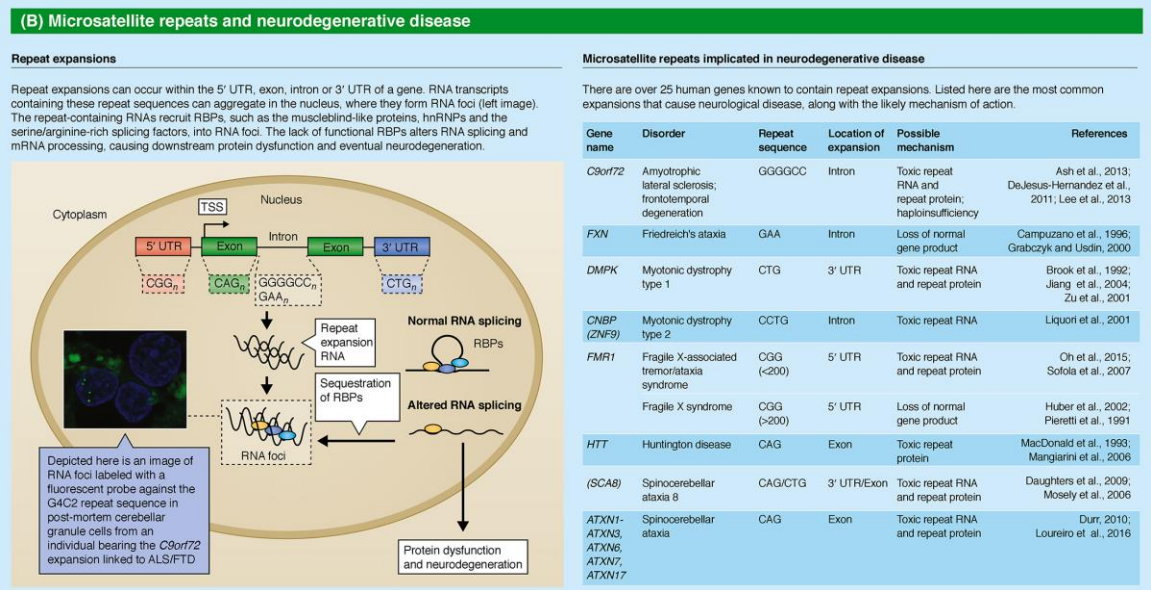



**Figure 1.1: Microsatellite repeat expansions and neurodegenerative disease.** Genomic locations and proposed mechanisms of repeat expansions that cause neurodegenerative diseases. Expansions can occur in the 5'UTR, exon, intron or 3'UTR of over 20 genes. Repetitive RNA from repeat expansions can lead to neuronal toxicity via sequestration of RNA binding proteins or through toxic protein species. Additionally, repeat expansions that are close to the transcription start site (TSS) can reduce expression of the gene in which they reside. Adapted from Liu, Cali and Lee (2017)[35]; used under the Creative Commons Attribution License.

For instance, the CAG trinucleotide repeat in Huntington's disease occurs within the first exon of *HTT*, where it is translated into aggregation-prone polyglutamine protein, leading to a toxic gain of function disease mechanism[49]. On the other hand, a CGG repeat in the 5' untranslated region of *FMR1* causes fragile X syndrome via epigenetic silencing and reduction of the normal FMRP protein[50]. When a repeat expansion occurs within an intron, however, the molecular pathogenesis is not as apparent. In Friedreich's ataxia, for example, an intronic GAA repeat reduces transcription of *FXN* and leads to neurodegeneration via loss of function mechanisms[51,52]. However, an intronic CCTG

repeat in *CNBP* is thought to cause myotonic dystrophy type 2 via a toxic RNA gain of function mechanism[53]. Thus, when the intronic hexanucleotide repeat expansion was discovered within *C9orf72*, both a toxic gain of function mechanism and a loss of function mechanism were proposed (Figure 1.2)[43,44].

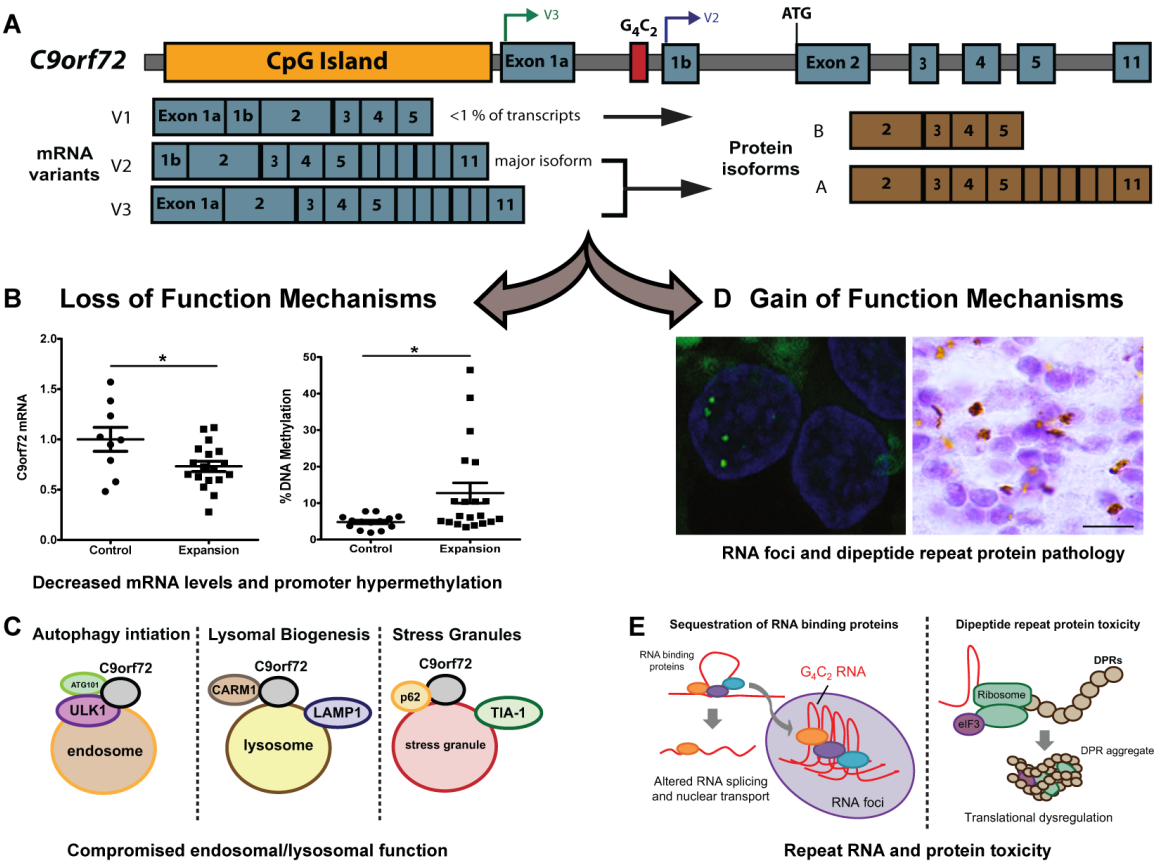



**Figure 1.2: Proposed mechanisms of *C9orf72* repeat expansion pathogenesis. A)** An intronic repeat expansion in *C9orf72* (red box) is thought to cause disease through either loss of function or gain of function mechanisms. *C9orf72* has 3 mRNA isoforms (arrows represent transcription start sites) which give rise to protein isoforms A and B. Only protein isoform A is detectable in human samples. **B)** Loss of function mechanisms include decreased mRNA levels from expansion carriers (left panel) and increased DNA methylation of the *C9orf72* promoter (right panel), leading to haploinsufficiency of the C9orf72 protein. **C)** C9orf72 is involved in endosomal/lysosomal pathways. Interaction between C9orf72 and autophagy initiation components ULK/ATG101 has been observed (left). Additionally, C9orf72 can bind CARM1 to facilitate lysomal biogenesis (middle). C9orf72 also interacts with p62 to regulate stress granule formation (right). **D)** Gain of function mechanisms include pathology derived from the repeat expansion sequence such as RNA foci (left) and dipeptide repeat (DPR) aggregates (right). **E)** Repeat RNA foci sequester RNA binding proteins and alter RNA splicing and nuclear transport (left). DPR aggregates are thought to bind elongation initiation factor 3 (eIF3) and the ribosomal subunits, leading to global translational dysregulation (right).

**Loss of function mechanisms of *C9orf72* repeat expansions**

In the years since the discovery of the *C9orf72* repeat expansion, there have been efforts to define the molecular mechanisms of *C9orf72* expansion mediated neurodegeneration. One proposed hypothesis was a loss function mechanism in which reduced C9orf72 gene product resulted in haploinsufficiency of the normal protein because reduced levels of C9orf72 mRNA and protein have been detected in repeat expansion carriers[43,44,54]. Bioinformatic analysis of the primary amino acid sequence revealed that C9orf72 is related to a class of proteins called Rab guanine exchange factors, which are important regulators of vesicle trafficking[55]. Furthermore, the loss of function hypothesis is supported by the finding that C9orf72 does regulate vesicle trafficking and autophagy in neuronal cell lines[56] and knockout in zebrafish results in motor deficits[57]. However, the loss of function model does not fully account for the results of studies using *C9orf72* knockout mouse models.

In one mouse study, global deletion of *C9orf72* resulted in several inflammatory phenotypes including splenomegaly, lymphadenopathy and altered microglial and macrophage function[58]. These mice, however, did not have neurodegenerative or behavioral phenotypes that recapitulate ALS/FTD symptoms and lived a normal lifespan[58]. In another study, *C9orf72* knockout again caused splenomegaly as well as inflammation that is consistent with autoimmune disease but no overt signs of neurodegeneration[59]. In this study, however, complete knockout of C9orf72 severely reduced lifespan, despite these mice being generated by the same methods as the

previously mentioned study[59]. A possible explanation for the difference in lifespan between the two studies could be due to environmental factors such as diet, stress levels or handling differences between the different institutions where the studies were conducted. Another knockout mouse study also reported an autoimmune phenotype consisting of myeloid expansion, T-cell activation and production of autoantibodies but no neurodegeneration or ALS/FTD related phenotypes and normal lifespan[60]. All of the *C9orf72* knockout mouse models, however, exhibit aberrant immune system phenotypes in only the homozygous knockout mouse but not in the heterozygous mouse[58–60]. Given that almost all *C9orf72* repeat expansion carriers are heterozygous for the expanded allele[61], the disease relevance of *C9orf72* bi-allelic knockout mouse models are questionable. Indeed, recent studies with novel antibodies or using mass spectrometry to better quantify protein levels have shown that C9orf72 protein levels are only reduced by ~20-40% in ALS/FTD patients[62,63]. Furthermore, the effects seen in global C9orf72 knockout mice do not demonstrate how compromised immune system function would interact with neuronal cells that are prone to degenerate in these diseases. To answer this question, an additional mouse study used a conditional allele to knockout C9orf72 expression in only neurons and glial cells at embryonic day 10.5 onward[64]. This study had the advantage of examining C9orf72 function in only disease relevant cell types and after early developmental stages. The neuron/glial specific C9orf72 knockout mouse had no signs of neurodegeneration and no indication of aberrant inflammation, suggesting that the autoimmune effects observed in whole body knockouts were likely due to C9orf72 functions in hematopoietic cells[64]. Overall, the animal models do show that the

C9orf72 protein has biological functionality that is worth exploring, but its relevance to neurodegenerative disease is not yet clear.

**C9orf72 protein function and autophagy**

The exact function of C9orf72 has been difficult to define, but it's been suggested that C9orf72 functions broadly in vesicle trafficking pathways such as the endosomal-lysosomal system and autophagy [65]. The C9orf72 protein contains a region homologous to the DENN (differentially expressed in neoplastic versus normal) domain, which allows interaction with Rab GTPases to allow exchange of GDP with GTP and thus provide the energy required for intracellular vesicle movement (Figure 1.3)[66] .

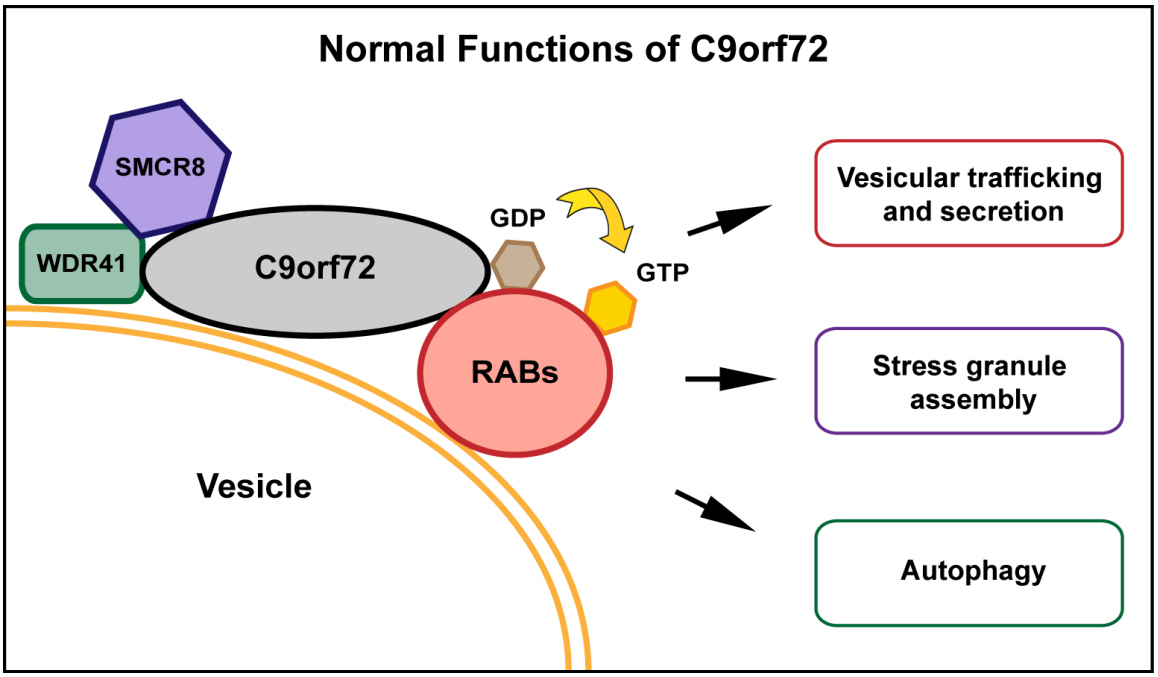



**Figure 1.3: Proposed functions of C9orf72 protein.** C9orf72 is a DENN domain containing protein that may function as a guanine exchange factor to provide energy for Rab GTPases, which are involved in numerous vesicle trafficking pathways including intracellular trafficking, vesicle secretion, stress granule assembly and autophagy. C9orf72 occurs in a complex with WDR41, SMCR8 and various Rab GTPase proteins.

Recent studies have shown that *C9orf72* is capable of interacting with numerous Rab GTPases (including Rab1a, Rab8, Rab39 and Rab7L1) that function at different steps of the endo-lysosomal and autophagy pathways[67–71]. It is also well documented that C9orf72 occurs in a complex with the proteins SCMR8 and WD41[72], and this complex interacts with the ULK1 autophagy initiation complex[67,69,73,74]. However, whether C9orf72 normally promotes autophagy induction or reduces autophagic flux is still not resolved[67,69,74,75]. This issue is complicated by the observation that C9orf72 functions differently in response to nutrient stress than it does under normal conditions[71,74] and also has other related functions such as regulation of stress granules via autophagy[76,77]. The complexity of C9orf72 function may be due to its ability to interact with many Rab GTPases, which then can carry out diverse functions within these vesicle trafficking sub-pathways[78]. For instance, C9orf72 has been demonstrated to act as guanine exchange factor (GEF) for Rab8 and Rab39[68], but its closest homolog, folliculin, is a GTPase-activating protein (GAP)[79]. GAPs and GEFs both regulate Rab GTPases that are important for vesicle trafficking, but they do so through opposite mechanisms. GEFs exchange GDP for GTP, thus activating the Rab GTPase for downstream signaling, whereas GAPs stimulate GTP hydrolysis and inactivation of the target Rab GTPase[79]. Given the homology to folliculin, it is possible that C9orf72 could have both GAP and GEF activity depending on the target, although this has not yet been shown experimentally. Additionally, the Rab GTPases that are targets of C9orf72 may change depending on the cell type, conditions or stressors present. Thus, further study is needed to define how C9orf72 functions in stress conditions and how the autophagy functions of C9orf72 interact with the repeat expansion associated pathology.

### Gain of function mechanisms of *C9orf72* repeat expansions

The other major hypothesis for understanding the molecular mechanisms of *C9orf72* repeat expansions is that neurodegeneration results from toxic RNA and protein species that are derived from the repeat expansion sequence[80]. Using fluorescent probes against the repeat RNA, RNA "foci" were first observed in the nucleus of cortex and spinal cord tissue from ALS/FTD patients[43] and have since been observed in numerous animal and cellular models[81–85]. Repeat RNA have been shown to bind and sequester RNA binding proteins such as those involved in splicing and nuclear transport[86–88]. Splicing factors that are possibly disrupted by the mutant RNA include hnRNP H and SRSF2[86–88], whereas nuclear transport defects may be due to sequestration of ALYREF and RanGAP1 by repeat expanded RNA[86,89]. This in turns leads to alterations in RNA splicing pathways as well as defective nuclear transport of RNA[87,90–93]. This RNA dysregulation model is attractive because defects in other RNA binding proteins such as FUS and TDP-43 are established mechanisms of neurodegeneration[35].

Another pathological hallmark of *C9orf72* expansion carriers is the presence of cytoplasmic protein aggregates that are generated from repeat RNA via the non-canonical translation pathway known as Repeat Associated Non-AUG translation[94–96]. These aggregates consist of chains of dipeptides which are translated from all reading frames of the $G_4C_2$ repeat sequence in both the sense and antisense direction[97]. These dipeptide repeat proteins (DPRs) are toxic in several model systems and proposed mechanisms of neurodegeneration include impaired global translation pathways, impaired RNA biogenesis, deficiencies in stress response pathways, and altered nucleocytoplasmic transport[98–103]. Finally, all *C9orf72* repeat expansion carriers develop cytoplasmic TDP-

43 aggregates, which is a hallmark of sporadic ALS and most forms of FTD[34]. TDP-43 normally resides in the nucleus, where it regulates RNA splicing, transport and stability[34]. In ALS/FTD, TDP-43 is lost from the nucleus and aggregates in the cytoplasm, resulting in thousands of gene expression changes in pathways such as RNA processing, nucleocytoplasmic transport, histone processing, and DNA damage[104]. Thus, the toxic nature of these three major pathologies (RNA foci, DPR and TDP-43) are thought to underlay the gain of function model.

More experimental evidence for the gain of function hypothesis comes from mouse models of *C9orf72* expansion carriers that were generated either by Adeno Associated Virus (AAV) delivery of $G_4C_2$ repeats or by insertion of the human repeat expansion sequence via Bacterial Artificial Chromosome (BAC)[105]. Over-expression of $G_4C_2$ repeats using AAV resulted in neurodegeneration, cognitive defects, RNA foci, and both DPR and TDP-43 pathology[84]. AAV delivery of only a DPR encoding sequence without the $G_4C_2$ RNA also induced motor and cognitive deficits in mice along with characteristic pathology[106], suggesting that DPRs alone are sufficient to cause ALS/FTD-like phenotypes. However, both of these methods drive pathology through over-expression at levels that are higher than observed in human disease[84,106]. Mice generated using the human gene within a BAC construct, which represents a more physiological model, also had characteristic RNA foci and DPR pathology[85,107–109]. While there were differences between studies in the observed phenotypes, two studies reported cognitive and motor defects in these mouse models[108,109] and one reported sparse TDP-43 pathology[109]. Thus, expression of the repeat construct alone, without loss of normal C9orf72 protein, is sufficient to recapitulate many features of ALS/FTD in mice.

In humans, however, it's not yet clear which pathology (e.g. RNA foci, DPRs or TDP-43 aggregation) are most responsible for driving neurodegeneration. TDP-43 aggregation is seen in regions of neurodegeneration in almost all ALS and FTD cases, except for familial cases caused by rare mutations[34]. This suggests that TDP-43 aggregation is linked to neurodegenerative processes. Indeed, TDP-43 pathology correlates more closely with neurodegeneration in *C9orf72* expansion carriers than DPR pathology does[110]. However, there have been several reports that FTD patients, including some that died prematurely before end stage disease, can develop neurodegeneration in the absence of TDP-43 pathology[111,112]. This has led to the idea that RNA foci and DPR pathology may be an early event in disease progression, with TDP-43 aggregation occurring later. Increased levels of DPR pathology has been reported to correlate with earlier age of onset in both ALS and FTD cases[113], suggesting that DPR aggregation may initiate disease progression. Of note, ALS cases have never been found without TDP-43 pathology, suggesting that the contribution of these pathologies may differ between ALS verses FTD[114]. Future studies will need to clarify how pathology derived from the repeat expansion interacts with TDP-43 pathology and which of these is necessary for disease progression.

A final consideration is that both a loss of function and gain function mechanism together may contribute to pathogenesis. A study using induced motor neurons from ALS patients suggests that a reduction in C9orf72 protein levels causes vesicle trafficking and lysomal defects that make neurons more susceptible to DPR aggregates from the repeat expansion[115]. This model presents a two hit hypothesis, where a loss of C9orf72 endosomal trafficking functions combine with toxicity from the repeat expansion

sequence. Although this concept is not currently supported by data from available mouse models, it is an interesting idea that should be further explored in future studies.

**Intermediate repeat expansions in *C9orf72***

Repeat expansions in *C9orf72* that cause ALS and FTD are usually hundreds to thousands of units long, with an average length of approximately 1400-1600 $G_4C_2$ repeat units in cerebellum[45]. Repeat expansions this large are almost fully penetrant, as approximately 96% of carriers will develop either ALS or FTD before age 80[45,116]. The minimum repeat number to develop ALS/FTD, however, has not been fully established. There have been reports of small repeat sizes as low as 40-65 repeats in patients with FTD[117–119]. However, there are issues that make accurate repeat sizing difficult including somatic mosaicism within individuals, differences depending on the tissue collected and technical challenges in accurately sizing a GC rich repetitive sequence[45,117]. A case report of a cognitively normal individual with a repeat size of 30-40 units in cerebellum showed that this length was enough to induce RNA foci and DPR pathology, but not TDP-43 pathology or clinical disease[120]. This was in contrast to a report indicating that a carrier with 38 repeats in several brain regions developed ALS with TDP-43 pathology and sparse DPR pathology[121]. Finally, another case report shows that an individual with an average of 70 repeats (with some larger expansions in different brain regions) did not develop clinical disease by age 90[122]. Thus, it appears that there is inter-patient variability with regards to the minimum repeat length necessary to cause ALS/FTD, perhaps due to differences in the genetic background or the degree of somatic mosaicism between individuals. In general, it seems that repeat expansions less than 30 units do not cause

ALS/FTD, whereas 30-70 repeats may represent a gray zone that depends on other genetic and/or environmental factors. The majority (~70-80%) of non-diseased individuals carry repeats of 2-8 units, whereas ~5% of the global population has repeat sizes greater than 10 units[123]. Whether intermediate repeat expansions that are too small to contribute to ALS/FTD but larger than normal (e.g. 10-30 units) are a risk factor for other diseases remains an open question.

Several studies have examined the prevalence of intermediate *C9orf72* expansions in other neurodegenerative diseases. Intermediate expansions were completely absent in Alzheimer's disease[119], but several studies reported small numbers of intermediate repeat carriers with Parkinson's Disease (PD)[124,125]. Interestingly, a follow up study that investigated *C9orf72* repeat length in only autopsy confirmed PD cases did not find any intermediate or full expansions, suggesting discrepancies between clinically and neuropathologically defined cohorts[126]. Finally, a large global meta-analysis of *C9orf72* repeat length in PD patients found a significant but very minor association of PD with intermediate length alleles ($\geq$ 17 repeats) and concluded that the risk of developing PD for intermediate carriers was negligible[123]. Thus, it is still not clear whether intermediate repeat expansions in *C9orf72* spanning 10-30 units are a risk factor for neurodegenerative disease. Additionally, a recent study with a very small cohort examined *C9orf72* repeat sizes in autoimmune diseases and found a significant association between systemic lupus erythematosus and intermediate repeats (9-30 units)[127]. While autoimmunity has been implicated in *C9orf72* knockout mice[59], additional follow up studies will be necessary to determine if there's any biological significance of *C9orf72* intermediate repeat expansions in this context.

**Epigenetic alterations in microsatellite repeat expansion diseases**

It is also important to consider epigenetic mechanisms of pathogenesis associated with microsatellite repeat expansion disorders. DNA methylation is a covalent modification of cytosine that is typically associated with transcriptional inhibition and is carried out by several DNA methyltransferase enzymes (DNMTs)[128,129]. One important function of DNA methylation is promotion of genome stability by silencing of repetitive elements[130,131]. DNA methylation patterns are known to change during aging and in the context of neurodegenerative disease[132]. Indeed, aberrant DNA methylation is a common feature of microsatellite repeat expansion diseases and has been associated with myotonic dystrophy, Friedreich's ataxia and Fragile X syndrome [133–135]. In the case of Fragile X, the main driver of disease seems to be a loss of *FMR1* expression that is associated with DNA methylation of the upstream promoter[135]. Thus, epigenetic alterations are important to understand the pathogenesis of repeat expansion mutations.

Changes in DNA methylation are also observed in *C9orf72* repeat expansion carriers. Approximately 33-36% of *C9orf72* carriers gain DNA hypermethylation of the promoter region, which is normally unmethylated in non-diseased individuals[47,136–139]. Promoter hypermethylation was not significantly different between ALS and FTD patients, but interestingly, hypermethylation correlates with longer survival in FTD patients[47]. Hypermethylation also correlates with reduced RNA foci and DPR abundance in patient tissue and patient derived cells[137], indicating that *C9orf72* promoter methylation may be protective in the setting of the repeat expansion. Additionally, FTD patients with promoter hypermethylation have reduced verbal recall decline over time

and reduced grey matter volume loss as measured by MRI[140]. Why DNA hypermethylation seems to be beneficial in FTD patients with the repeat expansion but not ALS patients is an unresolved question. One possibility, as discussed previously, is that TDP-43 pathology is driving neurodegeneration in ALS patients, whereas FTD pathogenesis is may be more connected to repeat expansion pathology that is dependent on *C9orf72* expression. Additionally, why only a subset of patients develop hypermethylation is also not clear. Studies to date have only applied correlative analysis to study DNA methylation in *C9orf72* carriers. To better understand these basic questions and to define the causal mechanisms associated with DNA hypermethylation of *C9orf72*, there is a need for disease models that accurately represent DNA hypermethylation as is observed in *C9orf72* repeat expansion carriers. Furthermore, an ideal model would be generated via targeting DNA methylation in a controlled and specific fashion to the *C9orf72* promoter in patient derived cells. Such epigenetic editing experiments would allow generation of isogenic cell lines that do not suffer from confounding factors due to genetic background.

### Methods of Epigenetic Editing

In the past decade, novel epigenetic editing tools have been developed that take advantage of DNA binding proteins fused to epigenetic effector enzymes. These include zinc finger proteins, TALE proteins and dead Cas9 proteins fused to DNA or histone modifiers[141]. These fusion protein epigenetic editing tools have helped to define basic cellular mechanisms[142–149] and have been applied as potential therapeutics in diseases with epigenetic alterations[150,151]. For targeted DNA methylation, a dead Cas9 enzyme

fused to the catalytic domain of DNA methyltransferase 3a (dCas9-DNMT3a) was the first tool generated using CRISPR/cas9[152]. CRISPR/cas9 based epigenetic editing methods are attractive alternatives to zinc finger and TALE proteins due the low cost and ease of multiplexing with guide RNAs (gRNAs) against various target loci[141]. However, a major issue has emerged with dCas9 based DNA methylation editing tools concerning specificity. Recent studies have observed that dCas9-DNMT3a tools can induce thousands of off-target methylation changes[153–155]. Numerous off-target effects have also been observed with dCas9 fused to cytosine base editing enzymes[156]. A possible explanation of these off-target effects is that Cas9 is able to bind DNA sequences with only a few base pairs of homology to the gRNA sequence, but requires more extensive homology to initiate DNA cleavage[157,158]. Thus, dCas9 fusion proteins that only require binding for DNA modification seem to initiate immensely more off-target changes than traditional Cas9 applications that rely on DNA double strand breaks[159]. A newer DNA methylation editing system has attempted to alleviate the off-target effects of dCas9 fusion proteins by using a modular system, called the SunTag system, to recruit DNMT3a to the locus where dCas9 is bound[155]. The advantage of this system is that dCas9 is not directly fused to DNMT3a and the amount of DNMT3a recruitment can be manipulated by the expression levels of an antibody that binds the SunTag sequence[155]. However, this system requires generation of multiple plasmids to target a single locus and the levels of off-target effects are dependent on the amount of DNMT3a recruitment[155]. Thus, optimal levels of the antibody encoding plasmid will need to be experimentally determined for each application in order to prevent off-target target methylation.

**DNA methylation and DNA repair pathways**

While almost all available tools for epigenetic editing involve fusion protein constructs, there is actually considerable evidence that DNA double strand break repair pathways can alter DNA methylation. Numerous studies have demonstrated that DNMTs are recruited to double strand break sites[160–163]. Upon double strand break, DNA can be repaired either through the process of non-homologous end joining (NHEJ) or homology directed repair (HDR)[164]. NHEJ occurs during G1 phase of the cell cycle and involves recognition of the DNA break by the Ku complex, recruitment of the protein DNA-PKcs, generation of overhangs by Artemis and finally ligation and fill-in by DNA polymerases[164]. The polymerases used during NHEJ are error-prone, often resulting in insertions or deletions (indels) at the repaired site[164]. Alternatively, the break can be repaired using HDR, which is error free but occurs at lower frequency than NHEJ[165]. HDR is a complex process that involves resection of the broken DNA ends, invasion of the repair template and DNA synthesis using the repair template followed by ligation to complete the newly copied DNA[166].

Several studies have utilized fluorescent reporter DNA constructs to interrogate whether DNA double strand breaks can induce DNA methylation. These studies show that upon double strand break repair via HDR, a subset of the repaired DNA becomes methylated, which leads to a reduction of reporter expression[167–170]. While there is some disagreement between studies, methylation of the exogenous construct following DNA repair appears dependent on DNMT1[167], DNMT3a[169] and UHRF1[169], a protein responsible for DNMT1 recruitment to hemi-methylated DNA[171]. Additionally, a double strand break induced in a reporter construct that is repaired via NHEJ can also lead to

DNA methylation[172]. Thus, it seems that DNA methylation is acquired following double strand break regardless of the repair pathway that is utilized.

**Objectives and Hypotheses**

The overall objectives of this dissertation were to better understand the genetic and epigenetic disease mechanisms underlying *C9orf72* repeat expansions. In Chapter 2, we aimed to generate a novel targeted DNA methylation technique that takes advantage of the endogenous DNA double strand break repair pathways. We hypothesized that double stranded breaks in the promoter CpG island of *C9orf72* would facilitate methylation near the break site. In the absence of repair template, the break would be repaired with NHEJ, which would lead to localized DNA methylation along with indels (Figure 1.4, left). In order to direct precise methylation editing, we asked whether an *in vitro* methylated repair template could be supplied such that during HDR the methylation marks would be copied from the exogenous repair template into the endogenous gene (Figure 1.4, right).

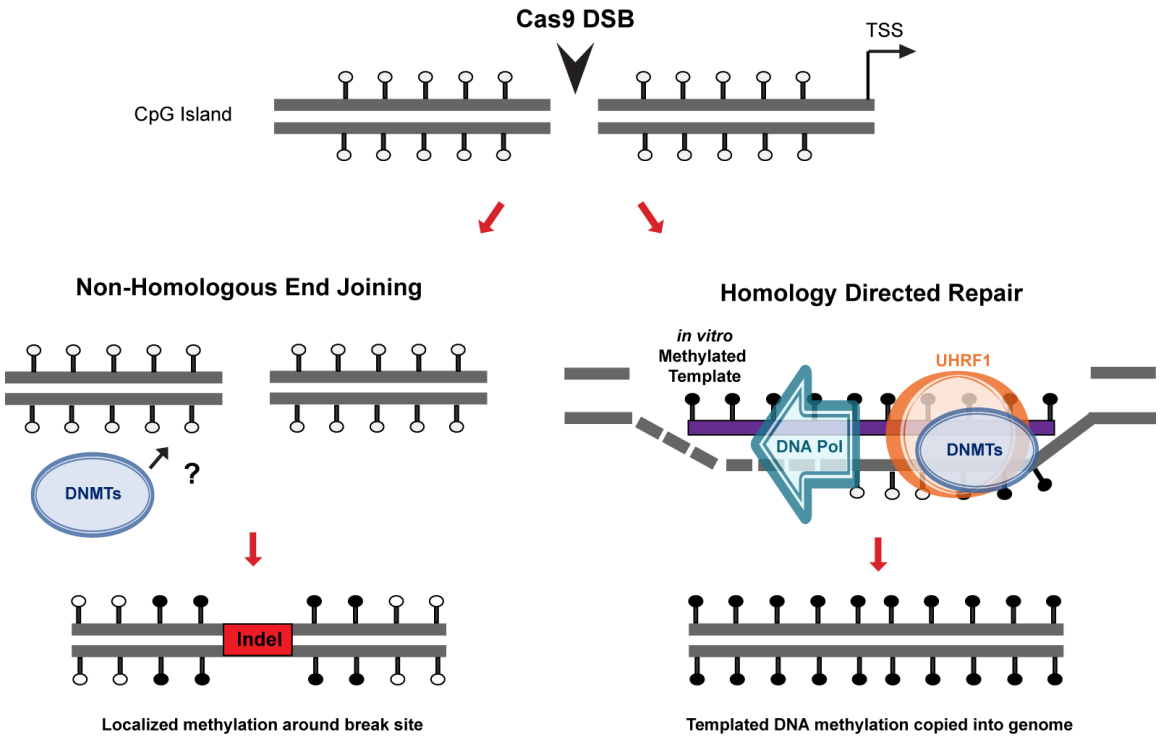



**Figure 1.4: Proposed mechanisms of targeted DNA methylation via DNA double strand break repair pathways.** Double strand breaks are induced in a CpG island via CRISPR/cas9. In the absence of repair template (left), the break is repaired via non-homologous end joining, resulting in recruitment of DNA methyltransferases (DNMTs) to the break site and introduction of localized DNA methylation. This pathway is also associated with insertions or deletions (indels) near the break site. In order to direct precise methylation editing, an *in vitro* CpG methylated template is provided (right) so that the break is repaired via homology directed repair. During repair, UHRF1 is recruited to hemi-methylated DNA, which then recruits DNMTs to copy methylation from the repair template into the endogenous sequence.

This would lead to templated methylation and should avoid the numerous off-target effects observed with dCas9 fusion protein constructs. The final aim of chapter 2 was then to apply this method, termed Homology Assisted Repair Dependent Epigenetic eNgineering (HARDEN) to ALS/FTD patient derived cells with the *C9orf72* repeat expansion in order to generate a disease model of epigenetic dysfunction that's observed in patients. Using this model, we sought to directly test how DNA hypermethylation affects *C9orf72* expression and downstream pathogenesis.

In Chapter 3, we aimed to determine the genetic and molecular mechanisms of intermediate repeat expansions in *C9orf72*. Previous studies conflicted as to whether intermediate repeats contribute to other neurodegenerative diseases such as Parkinson's disease (PD). We hypothesized that intermediate repeats in *C9orf72* may be a risk factor instead for corticobasal degeneration (CBD), a neurodegenerative disease that shares clinical features with PD but has distinct pathological protein aggregates consisting of tau[173]. In this study, we also sought to determine how intermediate repeats may affect expression of *C9orf72*. We hypothesized that intermediate repeats may alter expression differently than full repeat expansions that cause ALS/FTD. Finally, we sought to uncover the molecular pathways that are affected in intermediate repeat carriers.

# Chapter 2: Targeted DNA methylation of *C9orf72* via endogenous DNA repair pathways

This chapter has been adapted from following manuscript, which is currently in revision at *Nucleic Acids Research*, (Oxford University Press). Used under Creative Commons License

"Targeted DNA methylation of the ALS/FTD gene *C9orf72* via endogenous DNA damage repair pathways" by Christopher P Cali[1], Daniel S Park[1], and Edward B Lee[1].

[1] Translational Neuropathology Research Laboratory, Department of Pathology and Laboratory Medicine, University of Pennsylvania, Philadelphia, PA, USA

**Corresponding Author:**
Edward B. Lee
613A Stellar Chance Laboratories
422 Curie Blvd
Philadelphia, PA 19104
edward.lee@pennmedicine.upenn.edu

## Abstract

DNA methyltransferases (DNMTs) are thought to be involved in the cellular response to DNA damage, thus linking DNA repair mechanisms with DNA methylation. This study presents a novel method of targeted DNA methylation that utilizes endogenous DNA double strand break repair pathways and applies it to the neurodegenerative disease gene *C9orf72.* A double strand break induced by CRISPR/cas9 in the promoter of *C9orf72* is sufficient to induce DNA methylation, and methylation can be precisely targeted through the process of homology directed repair (HDR) via delivery of an *in vitro* methylated exogenous repair template. Long methylated double stranded DNA templates induce more methylation than shorter templates and with higher efficiency than a dCas9-DNMT3a fusion protein construct. Genome-wide methylation analysis reveals no significant off-target methylation changes when inducing methylation via HDR, whereas the dCas9-DNMT3a fusion construct causes significant off-target methylation at over 67,000 sites. This method is applied to generate a patient derived iPSC model of amyotrophic lateral sclerosis and frontotemporal dementia (ALS/FTD) that exhibits stable DNA methylation patterns similar to those seen in patients. Using this model, it's shown for the first time that DNA methylation of the 5' regulatory region directly reduces *C9orf72* expression and increases histone H3K9 tri-methylation levels.

## Introduction

Methylation of cytosine is a chemical modification of DNA associated with a myriad of regulatory processes in the mammalian genome such as regulation of gene expression, stabilization of repeat elements, X-inactivation and maintenance of genomic integrity[128]. Alterations in DNA methylation are associated with numerous diseases, including cancer and neurodegenerative diseases such as those caused by expansion of microsatellite repeat elements[132]. Aberrant methylation of the *C9orf72* promoter region has been observed in repeat expansion carriers with amyotrophic lateral sclerosis (ALS) and frontotemporal dementia (FTD)[47,137,139]. In these patients, neurodegeneration is caused by a repeat expansion in the first intron of *C9orf72* which has been proposed to cause production of toxic RNA and protein pathologies in patient tissue[94], patient derived cells[81–83,98] and animal models[84,85,88]. Interestingly, hypermethylation of the promoter occurs in ~1/3 of patients and correlates with increased survival time, reduced neuronal loss and better verbal recall ability in FTD patients[47,140]. Thus, we hypothesize that hypermethylation in the context of *C9orf72* repeat expansions is neuroprotective. However, to define causality between hypermethylation and disease pathology, there is a need to generate isogenic patient derived cell lines that accurately model this epigenetic aspect of ALS/FTD.

Current tools for targeted DNA methylation, such as the dead Cas9 (dCas9)–DNA methyltransferase (DNMT) fusion protein[152], have several features that limit their utility for generation of epigenetic disease models. Recent studies have shown that the methylation editing tools, such as the dCas9-DNMT3a fusion protein, have numerous

off-target effects[153–155,174,175], which may confound experiments or limit their use as therapeutic agents. Another limitation with available tools is that these constructs are usually large and difficult to transfect into patient derived primary cells or deliver via gene therapy vectors for therapeutic purposes[176] and they often require multiple gRNAs to achieve robust activity[152,174,175,177,178]. Additionally, DNA methylation changes made via the fusion protein constructs are often not maintained over numerous cell divisions at some loci,[148,152,153] making it difficult to generate stably edited cell lines for disease modeling.

Due to these limitations, we sought to develop a novel method for epigenetic editing that has fewer off target effects and enables generation of cellular disease models. Development of this new system was based on the observation that endogenous DNMTs are rapidly recruited to double stranded DNA (dsDNA) breaks[160,161,163,168,169]. For example, when dsDNA breaks are induced in exogenous reporter DNA constructs, a subset of cells gain DNA methylation of the exogenous construct, leading to gene silencing[167–170]. Whether this acquired DNA methylation following break repair is unique to exogenously expressed DNA or also occurs at endogenous loci is not yet known. We sought to answer this question and furthermore to determine whether precise methylation editing could be carried out via the process of homology directed repair (HDR) by providing the cells with an *in vitro* generated repair template containing the methylation marks to be copied into the genome. To our knowledge, no previous study has shown that exogenous DNA methylation marks can be copied into an endogenous gene.

This study demonstrates that double stranded DNA breaks can directly induce DNA methylation of the *C9orf72* promoter and that methylation can be copied during

HDR. These findings suggest that we have developed a novel method for editing DNA methylation that takes advantage of the endogenous cellular repair machinery and does not require over-expression of a methyltransferase enzyme. We term this method Homology Assisted-Repair Dependent Epigenetic eNgineering (HARDEN) because stable methylation editing is achieved upon DNA repair via HDR. This method exhibited vastly fewer off-target effects than dCas9-DNMT3a fusion constructs and enables generation of stably methylated cell lines that model epigenetic alterations seen in neurodegenerative disease.

## Methods

### Plasmid and Repair template Generation

CRISPR gRNAs were designed using Benchling software (http://benchling.com/) and cloned into pSpCas9(BB)-2A-Puro (PX459) V2.0 vector (Addgene, Cambridge, MA; Plasmid #62988)) for HARDEN experiments or pdCas9-DNMT3A-PuroR (ANV) (Addgene Plasmid #71684) and pdCas9-DNMT3A-PuroR_v2 (Addgene Plasmid #74407) as previously described[179]. Single stranded oligonucleotide templates including methylated oligonucleotides were synthesized by Integrated DNA Technologies, Inc. (Coralville, Iowa). The non-homologous oligonucleotide template is homologous against the unrelated gene *EMX1*. Sequences of gRNAs and templates are provided in Supplemental Tables 2.1 and 2.2. Double stranded DNA templates were made by PCR amplification of the target region using Q5 High Fidelity polymerase (New England Biolabs, Ipswich, MA) according to the manufacturer's protocol. 250ng of purified PCR products were A-tailed in a reaction containing 0.25mM dATP, 1.5mM $MgCl_2$, 1X

Buffer II and 5U Amplitaq polymerase (Applied Biosystems, Foster City, CA) for 30 minutes at $70^0$ C. A-tailed PCR products were ligated into pGEM-T Easy vector (Promega, Madison, WI) and transformed into NEB Turbo competent cells (New England Biolabs) according to the manufacturer's protocol. For introduction of PAM mutations, site directed mutagenesis was carried out using the QuickChange XL kit according to the manufacturer's protocol (Agilent Technologies, Santa Clara, CA). Non-homologous dsDNA templates consisted of the pGEM-T Easy vector without the *C9orf72* homologous region. For methylation of plasmid repair templates, 20ug of plasmid DNA was methylated in a 100ul reaction containing 50U of M.SssI CpG Methyltransferase (New England Biolabs), 640 µM SAM, 50 mM NaCl, 10 mM Tris-HCl (pH 7.9) and 10 mM EDTA for 4 hours at $37^0$C. Enzyme was inactivated for 20 minutes at $65^0$ C and then plasmid DNA was purified via phenol chloroform extraction followed by ethanol precipitation.

**Cell Culture and Transfection**

HEK293T cells were grown in DMEM high glucose + L-glutamine (Invitrogen, Carlsbad, CA) supplemented with 10% fetal bovine serum (Atlanta Biologicals, Flowery Branch, GA) and 1X Penicillin/Streptomycin (Invitrogen). For transfection of CRISPR components, 1µg Cas9/gRNA containing plasmid and 1µg repair template were co-transfected into 12 well dishes using Fugene 6 (Promega) or Lipofectamine 3000 (Invitrogen) with a 3:1 transfection to DNA ratio for 48 hours. For siRNA experiments, cells were transfected with 40nM siRNA using Lipofectamine 3000 (Invitrogen). For

time points longer than 2 days, puromycin selection was carried out at 1μg/mL for 3-4 additional days. For clonal line isolation, cells were diluted to 1cell/well in 96 well plates 7 days post transfection. DNA from clonal lines was isolated using QuickExtract Solution (Lucigen, Middleton, WI) according to manufacturer's protocol and screened for repair with the template using PCR and restriction enzyme digest. Induced pluripotent cell line CS52iALS-n6A was obtained from Cedars Sinai and maintained in mTeSR1 media (Stem Cell Technologies, Vancouver, Canada) and grown on Matrigel matrix (Corning, Corning, NY) coated plates as previously described[81]. For transfection of CRISPR components, 750ng Cas9/gRNA plasmid + 750ng repair template was co-transfected into 12 well dishes using Viafect reagent (Promega) with a 4:1 transfection to DNA ratio for 48 hours. 0.5ug/mL puromycin was used for selection of transfected iPSCs.

**Methylation sensitive restriction enzyme digest qPCR**

Methylation sensitive restriction enzyme digest quantitative PCR (qPCR) was carried out as previously described[137]. Genomic DNA was extracted using DNeasy Extraction kit (Qiagen) and qPCR was carried out using FastStart Universal SYBR Green mastermix (Roche, Basel, Switzerland) on a StepOne Plus Real-Time PCR Machine (Life Technologies) using standard cycling. For measurement of methylation in samples transfected with 1.4kb double-stranded templates, a forward primer was used that anneals outside the repair template and the qPCR cycling conditions were modified (38 cycles, annealing temperature 61.5 $^{0}$C, 1 min extension at 72$^{0}$C) to increase specificity. % methylation was calculated using the formula 2^(Ct Mock - Ct Digest). Fold change over

controls was calculated for long amplicon qPCR because percent digest verses % methylation is not 1:1. See Supplemental Table 2.1 for primers.

**T7 Endonuclease Assay and HDR digest Assay**

A T7 endonuclease assay was used to measure mutation (indel +HDR) efficiency based on T7 digestion of mismatched DNA hybrids. Genomic DNA was PCR amplified using Q5 hotstart polymerase (New England Biolabs), hybridized and digested with T7 Endonuclease I (New England Biolabs) according to manufacturer's protocol. To specifically measure HDR efficiency, 12ul of PCR product was digested with 8U SspI-HF (New England Biolabs) in 1X Cutsmart buffer for 2 hours at $37^0$ C. Digest products were run on either 2% agarose gels for T7 assay or 8% polyacrylamide gels for HDR assay and stained with SYBR Gold (Invitrogen). Bands were quantified using GelAnalyzer 2010a software (http://www.gelanalyzer.com/index.html) using valley to valley background subtraction. For calculating T7 digest efficiency, we used the formula: % digest = 100 x $(1 – (1- \text{fraction cleaved})^{1/2})$ as described previously[180]. For the restriction enzyme HDR assay: % digest = (sum digested products) / (undigested + digested products).

**Bisulfite Cloning**

1ug of genomic DNA was bisulfite converted and cleaned up using the EpiTect Bisulfite Kit (Qiagen) according to manufacturer's protocol. Bisulfite converted DNA (7.5% of total per reaction) was then PCR amplified using KAPA HiFi Uracil+ Readymix (Roche)

according to manufacturer's protocol. PCR products were purified with Ampure XP beads (Beckman Coulter), A-tailed with Amplitaq polymerase (Applied Biosystems), ligated into pGEM Easy T vector (Promega) and transformed into NEB Turbo competent cells (New England Biolabs) as described above. Amplicons were sequenced at the University of Pennsylvania Sanger Sequencing core using the T7 promoter/primer. Bisulfite analysis was carried out using BISMA software[180] and only sequences with >90% conversion rate were included in the analysis.

**Genome Wide Methylation Analysis**

Genome wide methylation analysis was carried out using the EPIC Array (Illumina, San Diego, CA), which measures CpG methylation at ~800,000 sites. Samples were transfected for 48hrs with either dCas9-DNMT3a-Puro, dCas9-DNMT3a-ANV-Puro or Cas9-Puro + 190bp methylated template and gRNA targeting *C9orf72*. Samples were puromycin selected for 48 hours and genomic DNA was extracted 4 days post transfection. Samples were then bisulfite converted and processed using the Infinium Methylation EPIC Kit (illumina) by the Center for Applied Genomics at Children's Hospital of Philadelphia. Data was analyzed in R using the Bioconductor package ChAMP[181]. Data was normalized with the beta-mixture quantile normalization method (BMIQ) and significantly differentially methylated probes were called using default parameters with the Benjamini-Hochberg procedure for false discovery rate adjustment.

**RNA Expression from iPSCs**

Cells were harvested via Accutase (Stem Cell Technologies), pelleted, washed in PBS and resuspended in 350ul RLT buffer (Qiagen). Lysates were homogenized via pipetting and vortexing and RNA was extracted using the RNeasy kit (Qiagen) with on-column DNAse digest for 15 minutes at room temperature. 1µg of RNA was converted to cDNA using the High-Capacity cDNA Reverse Transcription Kit (Applied Biosystems) according to manufacturer's protocol. cDNA was diluted tenfold and 4µl was used in a qPCR reaction with 2X FastStart Universial SYBR Green master mix (Roche) and 0.25µM of each primer. qPCR was run with default cycling conditions on a StepOne Plus qPCR machine (Applied Biosystems). Relative expression was calculated using the delta delta Ct method by normalizing to housekeeping genes (*ACTB, GPS*).

**Chromatin Immunoprecipitation**

Edited iPSC clonal lines were cross-linked with 1% formaldehyde ($37^0$ C, 10 minutes) and quenched with 125µM glycine. Cells were lysed (0.5% SDS, 10mM EDTA, 50mM Tris/HCl, pH 8) and sonicated on high power for 30 cycles using a Bioruptor 300 (Diagenode, Denville, NJ). 1µg of chromatin was incubated overnight at $4^0$ C with Protein A Dynabeads (Life Technologies) and 5µl of antibody against H3K9me3 (Cell Signaling #13969) or IgG. Beads were washed with buffers of increasing stringency (Low salt, high salt, lithium chloride and 1X TE buffer) and DNA was eluted in buffer containing 1% SDS. RNA and protein were digested using RNAse A (Thermo Scientific) and proteinase K (Promega). DNA was purified using PCR purification kit (Qiagen) and

eluted in 200µl EB Buffer (Qiagen). 4µl purified DNA was inputted into a qPCR reaction as described above and percentage of input was calculated for each sample.

## Results

### Targeted DNA methylation using oligonucleotide repair templates

As DNMTs are recruited to DNA damage sites, we tested the hypothesis that a dsDNA break induced by Cas9 in an endogenous locus would result in increased DNA methylation. HEK293T cells were transfected with a plasmid encoding for Cas9 with or without a gRNA specific to the *C9orf72* promoter, together with a negative control single stranded oligonucleotide non-homologous DNA repair template. Methylation was measured at a CpG site in the promoter region (Figure 2.1A) using a previously developed methylation sensitive restriction enzyme digest qPCR assay (MSRE-qPCR)[137]. This assay provides a quantitative measure of methylation of the *C9orf72* promoter based on resistance of DNA to digestion by the methylation-sensitive HhaI enzyme and the primers used anneal outside the repair templates and thus can only amplify genomic DNA. In the absence of gRNA, there were no measurable changes in methylation (Figure 2.1B), regardless of the presence or absence of various exogenous oligonucleotide repair templates. In contrast, cells transfected with plasmids encoding Cas9, a gRNA targeting *C9orf72* and the non-homologous repair template, exhibited an increase in promoter methylation in unselected cells (Figure 2.1B). Endogenous dsDNA breaks in this context are repaired mainly through non-homologous end joining (NHEJ)[164]. Thus, a double strand break alone, in the absence of HDR, appears to induce

methylation of the *C9orf72* promoter. Methylation in this condition, however, was transient and dissipated after only five days in culture (Figure 2.1F).

To test the hypothesis that DNA methylation can be copied from an exogenous repair template into the genome, HEK293T cells were then transfected with Cas9, gRNA targeted to the *C9orf72* promoter and a single stranded oligonucleotide homologous repair template that was either unmethylated or methylated at every CpG dinucleotide (10 sites total). Cells that received an unmethylated homologous template had a reduction in methylation compared to those transfected with the non-homologous template, and cells transfected with a methylated repair template showed increased methylation when compared to those transfected with the unmethylated template (Figure 2.1B). These results suggest that methylation of the exogenous repair template promotes methylation of the endogenous *C9orf72* promoter.

To measure NHEJ, a T7 endonuclease assay was used that digests mismatched PCR amplicons[180]. As a readout of HDR, a 2bp mutation was made in the repair template that creates an SspI restriction enzyme site (Figure 2.1A). NHEJ efficiency was similar between conditions and was consistently higher than HDR rates as expected (Figure 2.1C-E), which perhaps underlies the observation that methylation in the non-homologous template condition was the highest (as NHEJ is more frequent than HDR). The methylation associated with the methylated oligonucleotide repair template was slightly more stable, but still returned to baseline eight days post transfection (Figure 2.1G). Thus, NHEJ and HDR were associated with transient DNA methylation of the *C9orf72* promoter following Cas9-mediated DNA damage.

To test the precision of HDR dependent methylation, we designed an oligonucleotide repair template with a single methylated CpG site, corresponding to the HhaI restriction enzyme site used in our methylation assay (Figure 2.1H). If methylation of this single CpG site was copied via HDR, it should block HhaI digestion with similar efficiency as the fully methylated template. Indeed, HEK293T cells transfected with Cas9 and the single methylated CpG repair template had similar HhaI resistance as cells transfected with Cas9 and the fully methylated template (Figure 2.1H). Thus, a single stranded oligonucleotide template with a single methylated CpG site appears to be sufficient to target methylation to the *C9orf72* promoter in a transient manner.

**Longer double stranded templates improve methylation efficiency and stability**

While it appeared that methylation of the endogenous *C9orf72* promoter could be induced upon HDR with a methylated exogenous oligonucleotide repair template, methylation appeared to be relatively inefficient and transient. To enhance DNA methylation efficiency and stability, we tested the ability to introduce DNA methylation via HDR using longer double stranded DNA (dsDNA) repair templates. *In vitro* methylated dsDNA repair templates with 190bp (10 CpGs), 600bp (35 CpGs) and 1.4kb (53 CpGs) total homology to the *C9orf72* promoter were generated (Supplemental Table 3.S2) with mutations of the CRISPR PAM site to disallow repeat cutting of edited loci and to monitor HDR efficiency (Figure 2.2A). Unmethylated versus fully methylated repair templates were co-transfected with plasmids encoding CRISPR-Cas9 components into HEK239T cells. After four days, methylation was tested using a long amplicon MSRE-qPCR assay that only amplifies genomic DNA and not the repair templates.

Indeed, longer dsDNA templates induced more CpG methylation in a length dependent manner, with the longest repair template group showing more than two fold higher methylation levels than the shortest template group (Figure 2.2B).

HDR efficiency was monitored using restriction digests of genomic DNA PCR products to detect mutations of the PAM recognition site that is present in the DNA repair templates. HDR efficiency of the different templates does not explain the observed differences in DNA methylation, as the shortest 190bp methylated repair template resulted in the most HDR but the least amount of DNA methylation (Figure 2.1B and Supplemental Figure 2.1). Moreover, the methylation status of the repair template did not appear to affect HDR efficiency (Supplemental Figure 2.2). Additionally, upon comparing the 190bp dsDNA repair template versus the single stranded oligonucleotide repair template, there was a non-significant increase in methylation associated with the dsDNA template (Supplemental Figure 2.2), suggesting that DNA strandedness does not account for the increase in methylation efficiency when using longer dsDNA repair templates. Together, these results suggest that the length of the dsDNA repair template was the main factor that allowed for more efficient induction of DNA methylation.

Next, the HARDEN method of targeted DNA methylation was compared to the previously developed dead Cas9 (dCas9) DNMT3a catalytic domain fusion protein construct[152]. HEK293T cells were transfected with either Cas9+gRNA to *C9orf72* and the 1.4kb dsDNA methylated repair template or the dCas9-DNMT3a construct with the same gRNA. As negative controls, cells were transfected with catalytically inactive DNMT-dCas9 fusion enzyme (dCas9-DNMT3a-ANV) or Cas9 plasmid without gRNA. In control groups, there was very little detectable methylation using MSRE-qPCR (Figure

2.2C). Robust levels of methylation were induced using either dCas9-DNMT3a or HARDEN; however, the HARDEN method induced ~50% more methylation and thus has slightly improved efficiency at this locus. Notably, the same gRNA was used for both methods (Figure 2.2C). These results indicate that HARDEN may be an effective alternative method for targeted DNA methylation of the *C9orf72* locus.

Next, we tested whether methylation copied from a long dsDNA repair template is stably maintained after cell division, as this is important for targeted methylation editing applications. Cells were transfected with Cas9 and gRNA targeting *C9orf72* and either non-homologous repair template, unmethylated 1.4kb dsDNA repair template or methylated 1.4kb dsDNA repair template. As a control, cells transfected with Cas9 plasmid without gRNA and with a non-homologous repair template were used. Methylation was measured by MSRE-qPCR at the indicated days post transfection. Cells that were transfected with methylated template maintained higher methylation levels than controls and remained stable throughout the 30 day time course (Figure 2.2D). Additionally, cells transfected with either the unmethylated or methylated 1.4kb repair templates were sub-cloned and screened for clonal lines that underwent HDR using the repair template via PCR and restriction enzyme digest (Figure 2.2E-F). Clones were passaged from single cells and analyzed ~5 weeks after transfection of CRISPR components and repair template. 9 out of 13 clones (69.2%) that were repaired with the methylated template acquired >10% methylation, compared to only 1 clone out of 16 (6.3%) repaired with the unmethylated template (Supplemental Figure 2.3) as measured by MSRE-qPCR. Furthermore, all three clones that were repaired with the methylated template on both alleles became hypermethylated, with levels of methylation ranging

from 60-80% by both MSRE-qPCR and bisulfite sequencing of cloned amplicons (Figure 2.2G-H). Thus, repair with methylated dsDNA repair templates promotes stable and dense methylation of the *C9orf72* promoter.

**DNA Methylation is acquired via Homology Directed Repair**

Next, the underlying mechanisms of HARDEN were explored to better understand the efficiency and required proteins for this process. The first question tested was what percentage of DNA molecules that are repaired with an exogenous methylated repair template actually undergo copying of methylation into the endogenous DNA. To address this question, a ~1200bp dsDNA repair template was designed to insert a 12bp mutation that would serve as a primer site to specifically amplify genomic DNA in which the mutation is incorporated via HDR (Figure 2.3A). HEK293T cells were transfected with CRISPR components and either an unmethylated or methylated version of this template. Bisulfite PCR of genomic DNA from transfected cells showed that the mutation specific PCR only amplified DNA from cells that were co-transfected with repair template (Figure 2.3B). Bisulfite sequencing of these cloned amplicons revealed that DNA repaired with the unmethylated template was rarely methylated, with ~5% of total CpGs methylated (Figure 2.3C-D) and only 1/27 amplicons having at least 20% methylation (Figure 2.3E). In comparison, DNA repaired with the methylated template was methylated at greater than 50% of CpGs throughout much of the amplicon(Figure 2.3C-D) and had 24/31 highly methylated amplicons (Figure 2.3E). Moreover, bisulfite sequencing revealed that DNA methylation in the setting of HDR did not induce DNA methylation of the last 3' CpG site within the bisulfite amplicon (CpG site #22, Figure

2.3C). This is notable because the repair template covered the entire bisulfite amplicon except this 3' CpG site, consistent with our hypothesis that DNA methylation is copied in a templated manner. Thus, repair with a methylated homologous template facilitates methylation of the endogenous locus.

Next, siRNA knockdowns of DNA methyltransferase enzymes were carried out to determine which DNMTs are required for methylation during this process. siRNA against either DNMT1, DNMT3a, DNMT3b, and the DNMT1 recruiting factor UHRF1 were first transfected into HEK293T cells, and 24 hours later, the CRISPR components and a methylated 1.4kb repair template were introduced. Methylation was measured by MSRE-qPCR following an additional 48 hour incubation period. Knockdown of any of these proteins resulted in ~40-50% decrease in the amount of methylation acquired using HARDEN (Supplemental Figure 2.4). These results indicate that there may be non-redundant roles for DNMTs and UHRF1 for the establishment and/or maintenance of methylation during DNA repair.

**Genome wide methylation of epigenetically edited cells**

The previous results demonstrate that HARDEN can be a useful method for targeted methylation of *C9orf72*. Next, the potential genome-wide off target effects of HARDEN were explored and compared to the dCas9-DNMT3a catalytic domain fusion protein construct. Because HARDEN utilizes both the 20bp gRNA sequence and a homologous repair template to induce methylation, we hypothesized that HARDEN would be highly specific for the targeted locus. To test this, HEK293T cells were transfected with either the dCas9-DNMT3a construct, or CRISPR components and

methylated repair template (C9orf72+ Meth). As negative controls, catalytically inactive dCas9 enzyme (dCas9-DNMT3a-ANV) or Cas9 plasmid without gRNA but with a methylated repair template (No gRNA+Meth) were used. Non-transfected cells were also included as an additional control to ensure transfection did not perturb global DNA methylation. Genome-wide methylation analysis was measured using genome wide arrays that detect DNA methylation at ~800,000 CpG sites. No statistically significant off target effects were detected using the HARDEN method (Figure 2.4A). This is in contrast to the dCas9-DNMT3a fusion construct which resulted in massive off target methylation (67,246 sites with $p < 0.05$) when compared to the dCas9-DNMT3a-ANV construct (Figure 2.4B). These off-target effects are not attributable to the gRNA, as the same gRNA was used for both the dCas9-DNMT3a group and the HARDEN group. When looking at the raw methylation values across the genome, there was no difference in methylation at lowly methylated sites between non-transfected cells or cells transfected with HARDEN components (Figure 2.4C). However, cells transfected with dCas9-DNMT3a showed a pronounced gain in methylation at lowly methylated sites compared to non-transfected cells and cells transfected with dCas9-DNMT3a-ANV (Figure 2.4D). These results suggest that HARDEN is a highly specific method for DNA methylation editing and confirms previous findings that dCas9-DNMT3a fusion proteins have massive global off-target effects.

**Editing of Methylation in ALS patient derived iPSCs**

ALS and FTD patients with the *C9orf72* repeat expansion can exhibit promoter hypermethylation, which is thought to contribute to transcriptional repression of the

*C9orf72* gene[137]. To model this aspect of ALS/FTD, HARDEN was applied to previously characterized patient derived iPSCs[81] harboring the repeat expansion but with limited *C9orf72* promoter methylation (Supplemental Figure 2.5). IPSCs were transfected with CRISPR components and either unmethylated and or methylated 1.4kb dsDNA templates. Clonal lines were isolated and tested for incorporation of the PAM mutation via HDR using restriction enzyme digestion of PCR amplicons (Figure 2.5A). Methylation of edited clones was measured using MSRE-qPCR. Targeted methylation of iPSCs was slightly less efficient than in HEK293Ts, but stably methylated clones were obtained with methylation levels of ~20-50% (Figure 2.5B). DNA methylation levels were confirmed by bisulfite sequencing of cloned amplicons (Figure 2.5D). Additionally, chromatin immunoprecipitation (ChIP) was performed on clonal iPSCs using an antibody against the repressive histone mark H3K9 trimethylation (Figure 2.5C). iPSCs that gained DNA methylation also gained H3K9 trimethylation, indicating a switch to a repressive chromatin state, and this increase was specific for the *C9orf72* locus (Figure 2.5C, Supplemental Figure 2.6). Finally, RNA was extracted from these edited iPSCs and mRNA levels of the major *C9orf72* isoforms were measured via RT-qPCR (Figure 2.5E). Methylated iPSC clones had reduced mRNA levels by ~50%, with the largest effect seen for variant 3 (Figure 2.5H), which has a transcription start site closest to the targeted CpG island. These results indicate that hypermethylation of the *C9orf72* promoter CpG island does indeed cause a reduction of *C9orf72* expression, confirming previous reports that were based only on correlative analysis[47,137]. Thus, HARDEN can be used to generate a disease model of aberrant methylation and allows for defining the direct effect of DNA methylation on gene expression at this locus.

## Discussion

This study highlights a novel method for targeted DNA methylation and applies it to generate a cellular disease model. In doing so, several new findings about the interplay between DNA damage repair pathways and DNA methylation were explored. First, these results suggest that a targeted double strand break can induce DNA methylation of an endogenous locus. Previous studies have demonstrated that repair of dsDNA breaks in a GFP-based reporter construct can lead to DNA methylation following repair by either NHEJ[172] or HDR[167–170]. This study demonstrates that this process also occurs at an endogenous gene, although methylation acquired following NHEJ is not maintained through cell division and is lost within 5 days in culture. The reduction in methylation observed following NHEJ is consistent with passive demethylation, with ~50% reduction each day. These results indicate that localized DNA methylation may serve as a transient marker of dsDNA breaks, perhaps to prevent transcription and promote stability of potentially mutagenized DNA.

Another finding is that methylation appears to be copied from either a single or double stranded DNA template through the process of HDR, including from oligonucleotides harboring a single methylated CpG site. This level of specificity could be useful to study the effect of a particular CpG site on transcription factor binding. Longer double stranded templates seem to be more efficient at inducing methylation than short single stranded oligonucleotide templates and methylation induced via dsDNA templates can be stably maintained over numerous cell divisions in both HEK293T cells and iPSCs. Our results suggest that template length is more important than strandedness

for inducing methylation, as there is only a small, non-significant increase in efficiency between 190bp ssOligo templates and the corresponding 190bp dsDNA template. In contrast, there are more sizeable observed differences in efficiency between 190bp, 600bp and 1.4kb dsDNA templates. It is not clear whether this effect is due to the overall increase in the number of methylated CpGs or if the longer templates include key CpG sites necessary for stable methylation of the *C9orf72* promoter. There is evidence, for example, that specific transcription factor binding sites may be more important than the total number of methylated CpGs for determining promoter methylation state[182]. Knockdown of DNA methyltransferase enzymes revealed that both de novo (DNMT3a and 3b) and maintenance methyltransferases (DNMT1) as well as UHRF1 are required for efficient methylation using HARDEN. These enzymes have previously been shown to assemble at double strand DNA break sites[160,168,169], and UHRF1 is thought to be important for double strand break repair in addition to its role in maintenance methylation[183]. Thus, it appears that an interplay of these enzymes and cofactors function in concert to promote methylation during DNA damage repair, although they may function in different repair sub-pathways .

HARDEN was also compared to the dCas9-DNMT3a fusion protein, and it appears that HARDEN has slightly better on-target efficiency at *C9orf72* than the dCas9-DNMT3a construct. Importantly, HARDEN has no detectable off-target effects, whereas the dCas9-DNMT3a fusion protein induces tens of thousands of off-target methylation changes. While only the off-target effects of a single gRNA are compared here, other studies have documented that dCas9-DNMT3a fusion proteins have large off-target effects irrespective of the gRNA used[153,154].

HARDEN can also be used to make stably methylated cell lines that model epigenetic alterations seen in ALS/FTD patients with the *C9orf72* repeat expansion[47,136,137]. Indeed, this study confirms for the first time that the endogenous *C9orf72* promoter CpG island directly regulates expression of this gene and that DNA methylation of this region leads to an increase in histone H3K9 trimethylation. Interestingly, the overall levels of methylation in clonal iPSCs are lower than in clonal HEK293Ts. This could be due to differences in active demethylation between the cell types or selective pressure against fully methylated *C9orf72* in iPSCs. It is known that iPSCs require active demethylation via TET enzymes to maintain pluripotency[184].

We acknowledge that HARDEN may not the best targeted DNA methylation technique for some applications. Since HARDEN appears to be linked to HDR, methylation efficiency is dependent on HDR rates which can be low or absent in certain settings. For instance, targeting post-mitotic cells or cell types that have a low frequency of HDR may limit the utility of this method. Furthermore, since stable methylation editing requires both a gRNA and repair template, multiplexing to target several genes at once would be labor intensive. A final limitation is that in order to prevent cutting of the repair template, a PAM mutation is required and thus scarless methylation editing without changing the DNA sequence is difficult. Nevertheless, the cellular models established here using HARDEN can be utilized to better understand the molecular consequences of *C9orf72* hypermethylation in the context of the repeat expansion.

## Figures

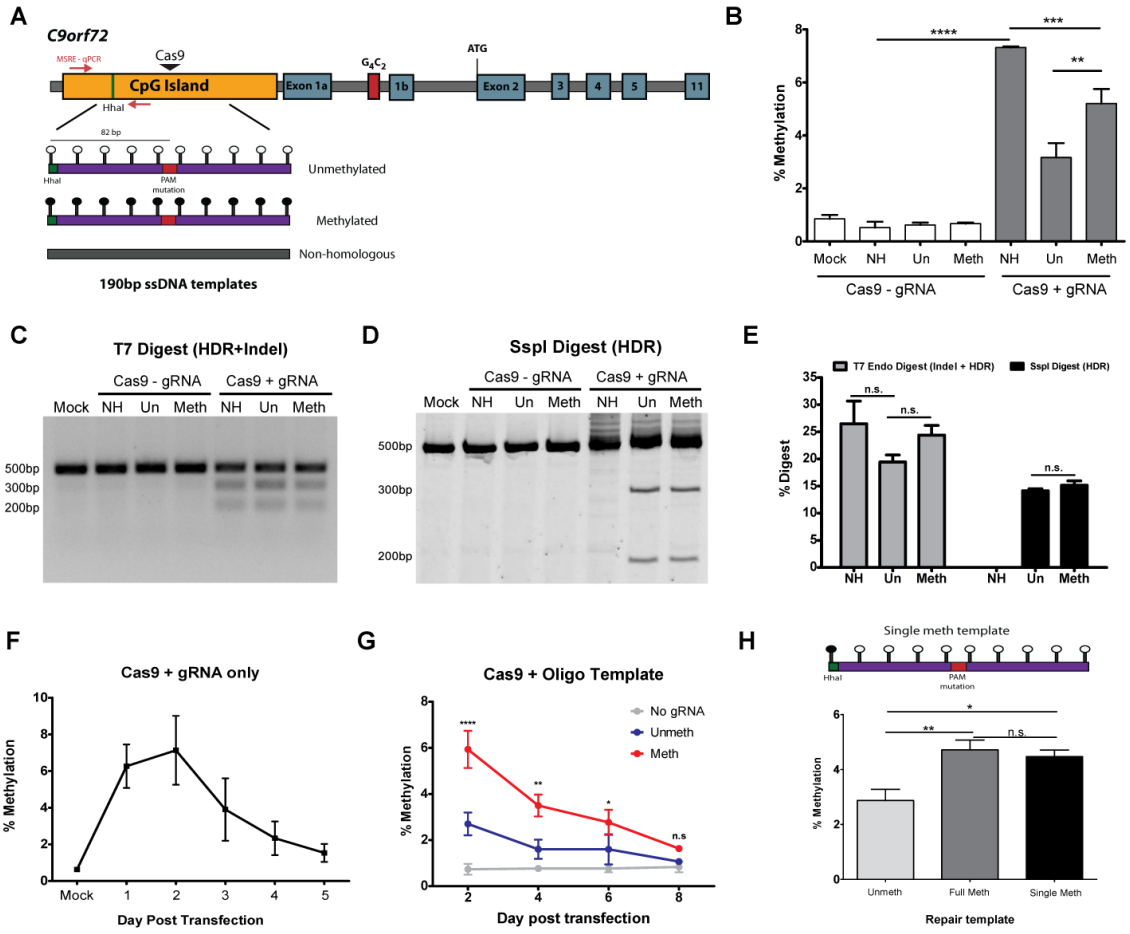



**Figure 2.1:** Editing DNA methylation of C9orf72 using single stranded DNA templates **A)** Diagram of *C9orf72* gene structure. Red arrows indicate primers used for MSRE-qPCR. The forward primer anneals outside of the repair template. Green line indicates HhaI site that is being tested for methylation in (B). **B)** MSRE-qPCR of a CpG site within the *C9orf72* promoter in HEK293T cells transfected with CRISPR/cas9 plasmid and the indicated single stranded oligonucleotide repair templates. Cells were collected 2 days post transfection. n =3 experiments. One way ANOVA followed by Bonferroni pairwise comparisons. All groups with gRNA are significantly different than those without gRNA. ****$p < 0.0001$, ***$p < 0.001$, **$p < 0.01$. **C)** Representative PCR + Digest with T7

endonuclease I from experiments shown in (B). Digest with T7 estimates the total amount of mismatched DNA (Indels and copying of the 2bp PAM mutation via HDR). **D)** Representative PCR + digest with SspI restriction enzyme from experiments shown in (B). SspI digest estimates the rate of HDR (copying of the 2bp PAM mutation creates SspI restriction enzyme site). n = 6 experiments. **E)** Quantification of T7 and SspI Digest assays. n =6 experiments. 2 way ANOVA shows significant interaction between sample group and digest assay ($p = 0.0004$). Bonferroni Post hoc tests were carried out between groups within each assay. **F)** MSRE-qPCR timecourse of HEK293T cells transfected with Cas9+gRNA and no repair template. n =3 experiments. **G)** MSRE-qPCR timecourse of HEK293T cells transfected with Cas9 and ssoligo repair templates. n =3 experiments. 2 way ANOVA shows significant interaction between time post transfection and template ($p = 0.0010$). Comparisons are made to Cas9 without gRNA transfected group. ****$p < 0.0001$, **$p < 0.01$, *$p < 0.05$. **H)** MSRE-qPCR of HEK293T cells transfected with Cas9+gRNA and either a fully unmethylated template, a fully methylated template or a template with a single methyl-CpG on the HhaI site. Cells were collected days post transfection. n =6 experiments. One way ANOVA followed by Bonferroni pairwise comparisons. **$p < 0.01$, *$p < 0.05$.

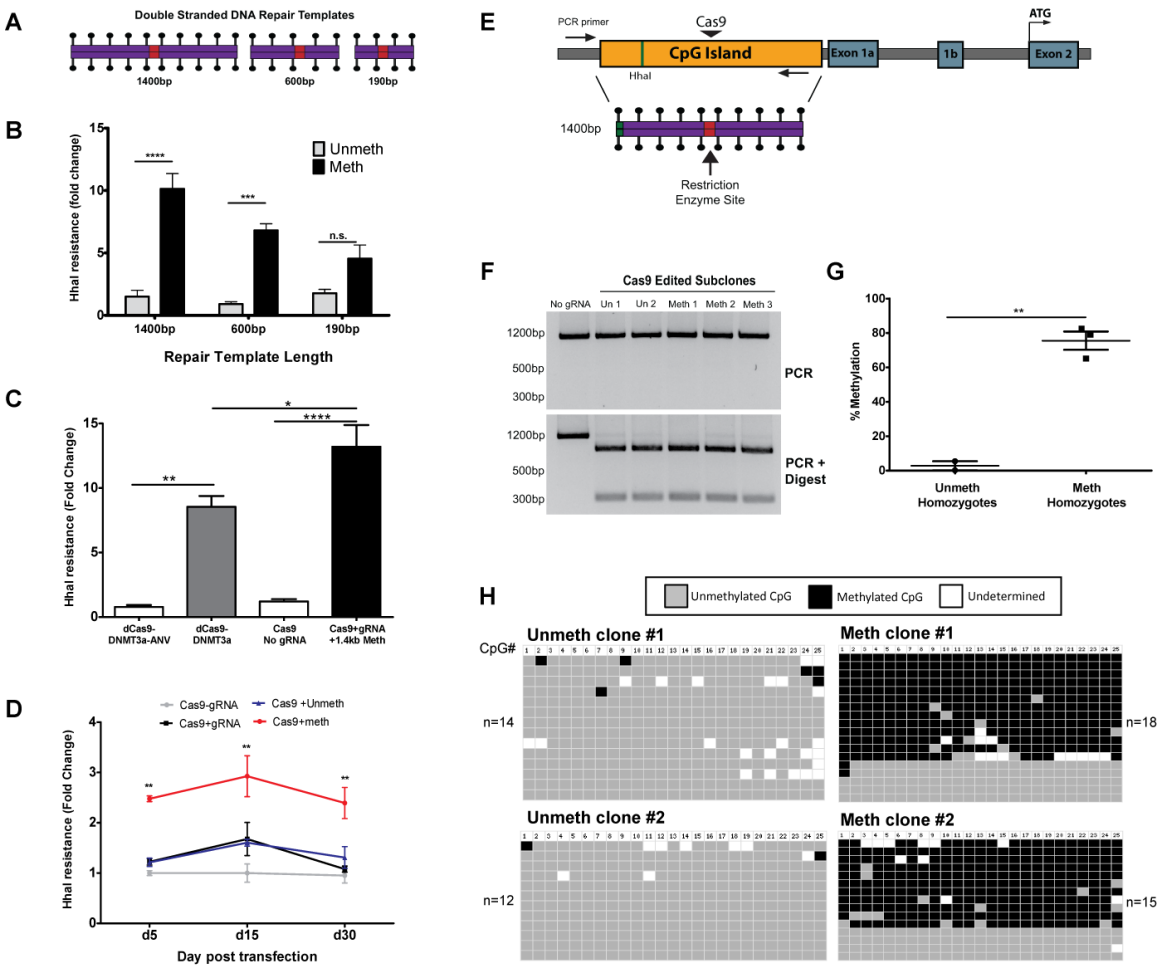



**Figure 2.2:** Editing DNA methylation with double-stranded repair templates. **A)** Diagram of double stranded DNA repair templates (either fully methylated or fully unmethylated at CpG sites) homologous to the *C9orf72* promoter. Red box indicates a 2 bp substitution that generates a restriction enzyme site and mutates the PAM sequence. **B)** MSRE-qPCR assay to measure DNA methylation using a long amplicon that anneals outside the repair templates. HEK293T cells were transfected with CRISPR/Cas9+gRNA and the indicated templates. Cells were puromycin selected for 2 days and collected at day 4 post transfection. n =3 experiments; 2 way ANOVA shows significant interaction between template length and template methylation status (p = 0.0071). Bonferroni post hoc test

between unmeth and meth template groups ****p < 0.0001, *** p < 0.001. **C)** Comparison of methylation efficiency between HARDEN method and dCas9-DNMT3a fusion protein construct. HEK293T cells were transfected with indicated plasmids and methylation was measured 2 days post transfection using long amplicon MSRE-qPCR. n =3 experiments. One way ANOVA (p < 0.0001) followed by Bonferroni comparison of indicated groups. ****p < 0.0001, **p < 0.01, *p < 0.05. **D)** Timecourse of methylation measured via long amplicon MSRE-qPCR using 1.4kb dsDNA repair templates. n =3 experiments. 2 Way ANOVA (Group p < 0.0001; day post transfection p = 0.0343) followed by Bonferroni comparison between unmethylated template and methylated template groups. **** p < 0.01. **E)** Diagram of *C9orf72* promoter region indicating the position of the repair template used to generate stably methylated HEK293T clones. Red box indicates a 2 bp substitution that generates a restriction enzyme site and mutates the PAM sequence. **F)** PCR and digest of HEK293T clonal lines that were repaired using the 1.4kb dsDNA repair templates. Cells were collected ~5 weeks post transfection. **G)** MSRE-qPCR of clonal cell lines that were homozygous for editing with either the unmethylated (n =2) or methylated 1.4kb repair template (n =3). Cells were collected ~5 weeks post transfection. Two tailed t-test; p = 0.0020. **H)** Bisulfite amplicon sequencing of homozygous clonal cell lines edited with the unmethylated or methylated 1.4kb repair templates. Cells were collected ~5 weeks post transfection. Each row is an individual DNA molecule and columns are individual CpG dinucleotides. Grey boxes indicate unmethylated CpGs, black boxes indicate methylated CpGs and white boxes are undetermined due to poor conversion or poor sequencing quality.

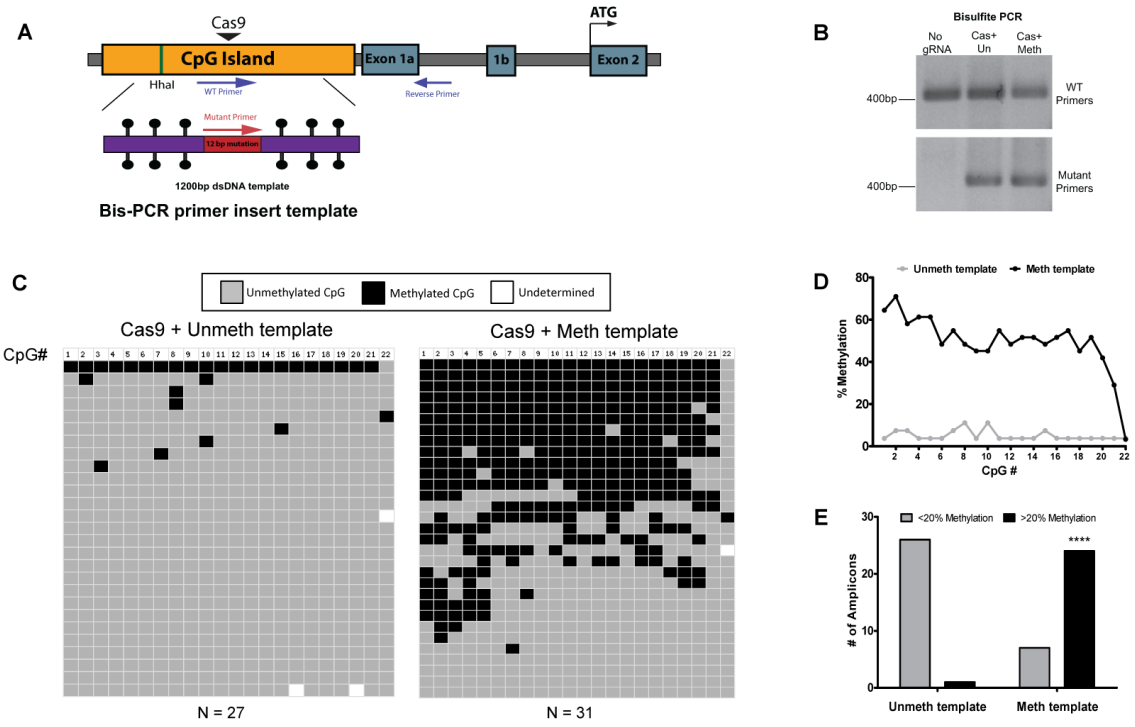



**Figure 2.3:** DNA methylation is associated with Homology Directed Repair. **A)** Diagram showing the targeted region of *C9orf72*. Purple arrows indicate primers that amplify wild type DNA, whereas the red arrow indicates the primer that specifically amplifies a 12bp mutation that was included in the repair template. **B)** Bisulfite PCR using the primers depicted in (A). Representative gel from two biological replicates. **C)** Bisulfite amplicon sequencing from two biological replicates (Unmeth: n= 27 amplicons, Meth: n =31 amplicons). Each row is an individual cloned and sequenced DNA molecule and columns are individual CpG dinucleotides. Grey boxes indicate unmethylated CpGs, black boxes indicate methylated CpGs and white boxes are undetermined due to poor conversion or poor sequencing quality. **D)** Quantification of percent methylation at each CpG site sequenced in (C). n =2 biological replicates and 22 CpG sites. **E)** Quantification of the

number of amplicons with less than or greater than 20% methylation. Data are from 2 biological replicates. Fisher's Exact Test, $p < 0.0001$.

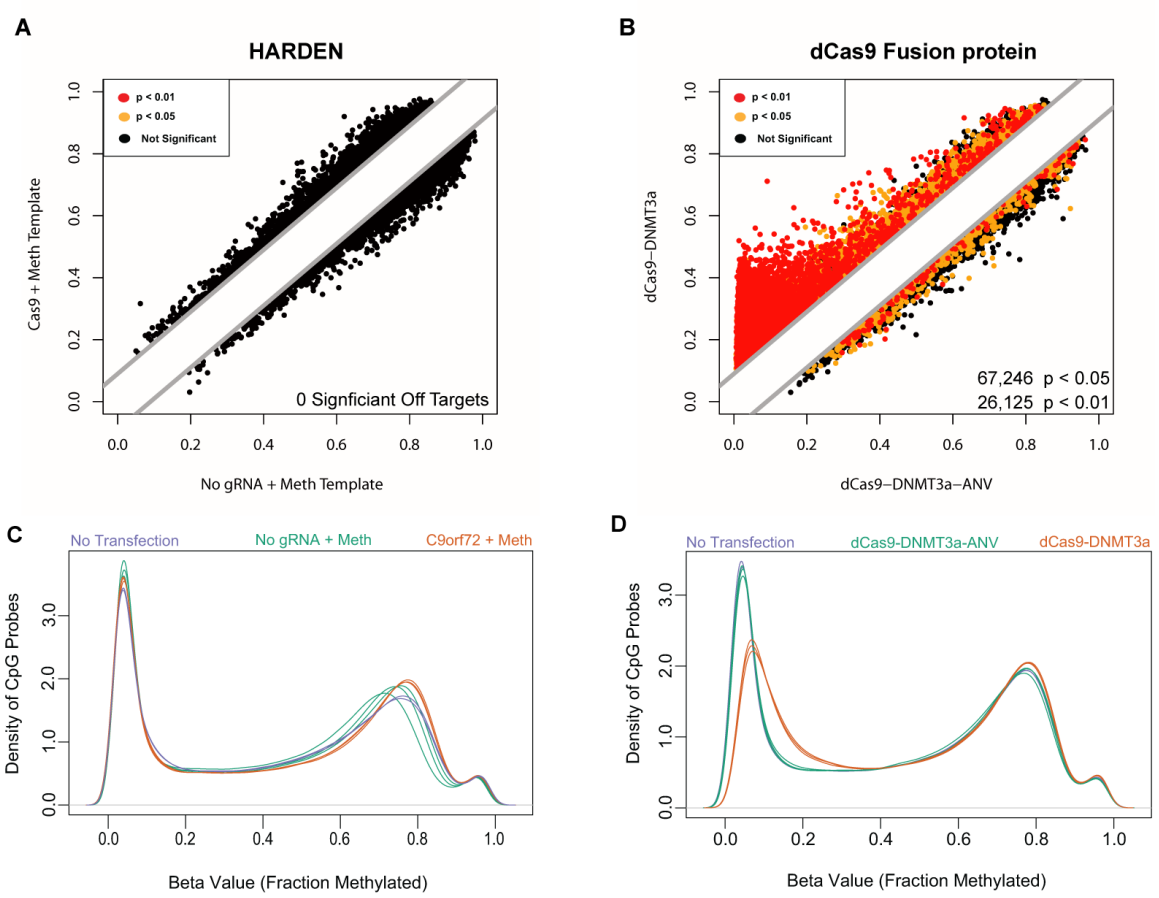



**Figure 2.4:** Genome-wide methylation analysis of HARDEN and Cas9-DNMT3a based methods. **A)** Methylation values from HEK293T cells transfected with Cas9+gRNA to *C9orf72* and a methylated repair template vs those without gRNA. Methylation was measured on an Illumina Epic Methylation Array with ~800,000 probes. Cells were puromycin selected and collected at day 4 post transfection. n =3 biological replicates; Colored dots indicate differentially methylated probes calculated using the Benjamini–Hochberg procedure. Black = not significant (adjusted p > 0.05); Yellow = adjusted p < 0.05; Red = adjusted p < 0.01. **B)** Methylation values from HEK293T cells transfected with either dCas9-DNMT3a or catalytically dead dCas9-DNMT3-ANV along with gRNA to *C9orf72*. Cells were puromycin selected and collected at day 4 post

transfection. n =3 biological replicates. Methylation was measured on an Illumina Epic Methylation Array with ~800,000 probes. Colored dots indicate differentially methylated probes calculated using the Benjamini–Hochberg procedure. Black = not significant (adjusted $p > 0.05$); Yellow = adjusted $p < 0.05$; Red = adjusted $p < 0.01$. **C)** Raw methylation values vs the density of probes genome-wide for non-transfected cells (n =2), Cas9+gRNA and methylated template (n =3) or Cas9 without gRNA and methylated template (n =3). Methylation was measured on an Illumina Epic Methylation Array with ~800,000 probes. **D)** Raw methylation values vs the density of probes genome-wide for non-transfected cells (n =2), dCas9-DNMT3a+gRNA to C9orf72 (N=3) or catalytically dead dCas9-DNMT3-ANV+gRNA to *C9orf72* (n =3). Methylation was measured on an Illumina Epic Methylation Array with ~800,000 probes.

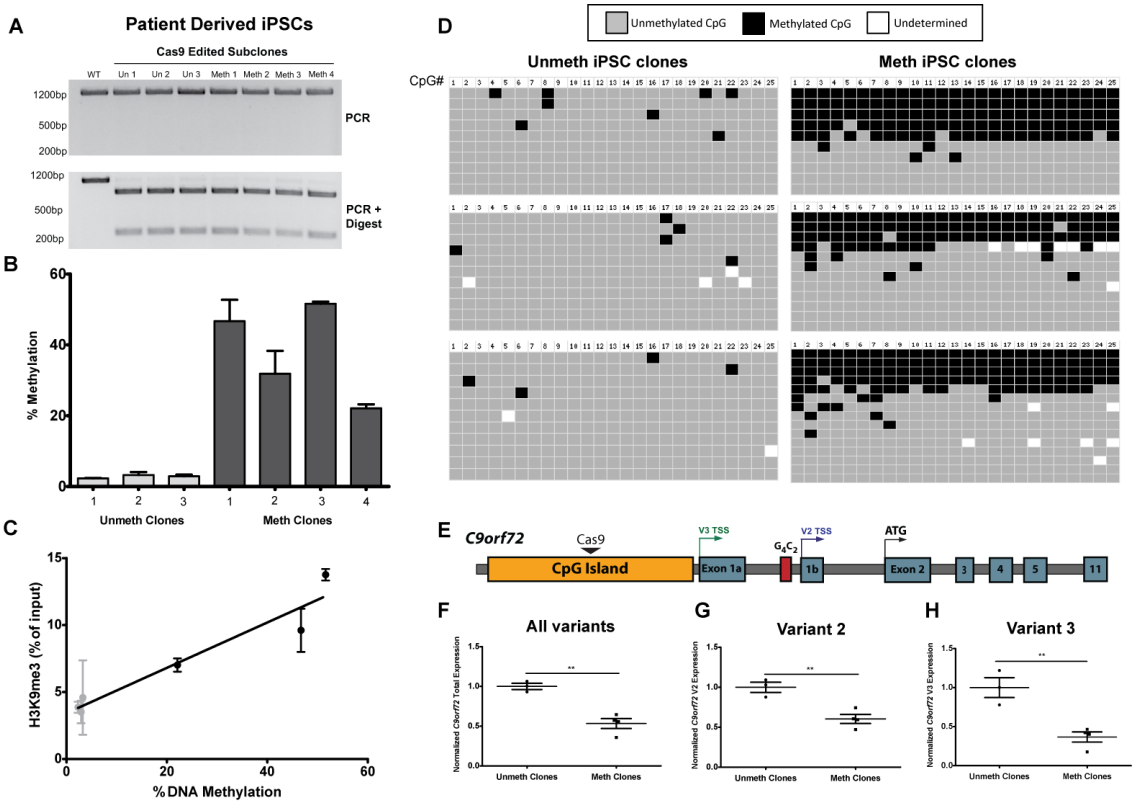
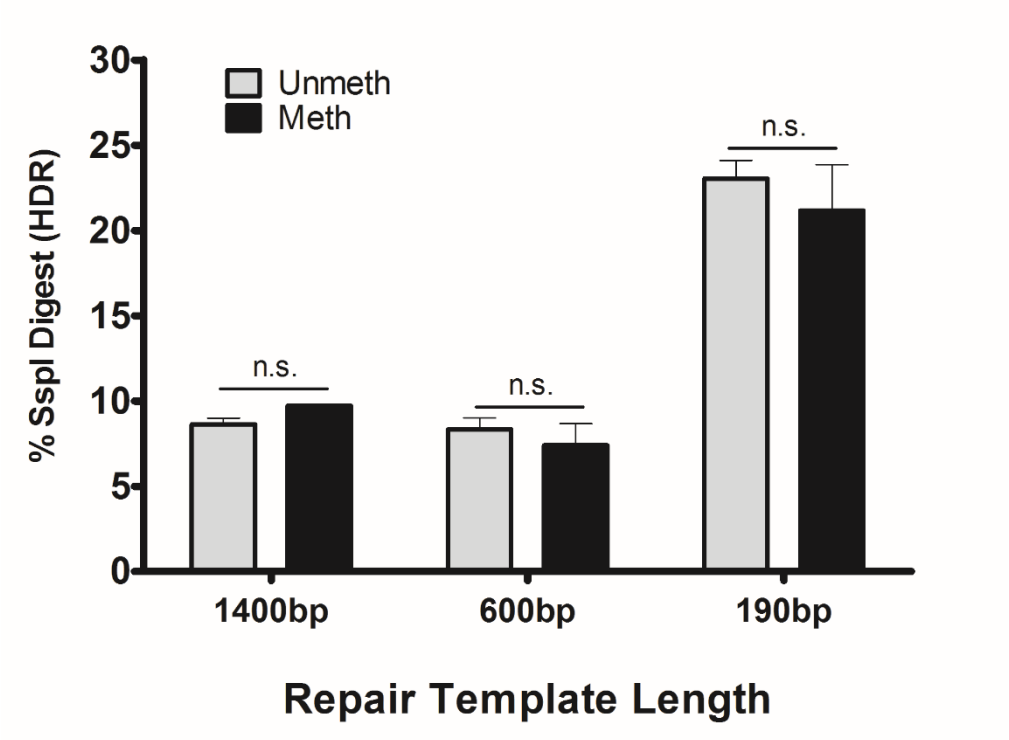


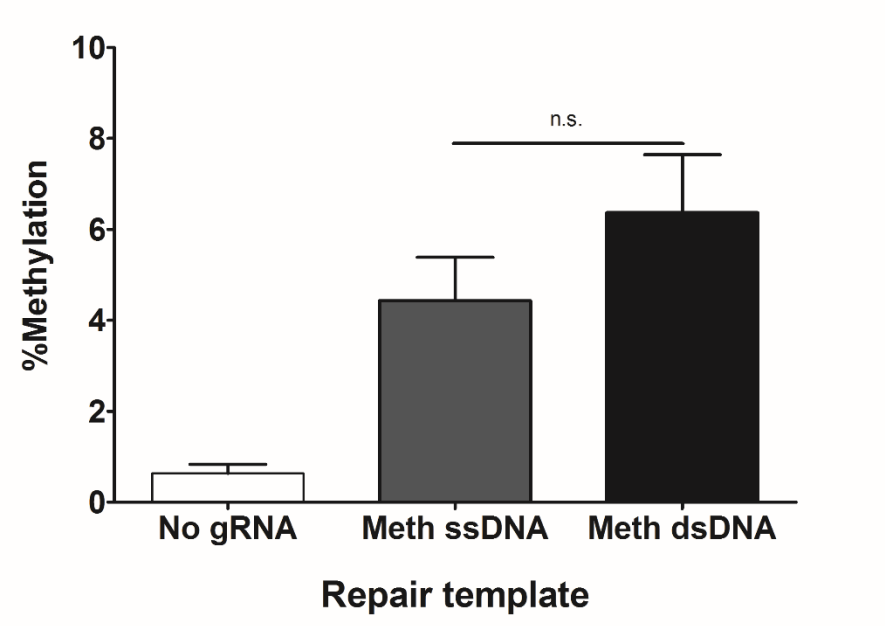

**Figure 2.5:** Editing of DNA methylation in ALS patient derived iPSCs. **A)** PCR and digest of clonal ALS patient derived iPSCs that were repaired using 1.4kb double stranded dsDNA repair templates. **B)** MSRE-qPCR to measure DNA methylation of clonal iPSCs that were homozygous for editing with the repair template. Cells were puromycin selected for 2 days, subcloned at day 7 and collected at ~6 weeks post transfection. n =2 measurements per cell line and 3-4 cell lines per group. **C)** Correlation between H3K9 trimethylation levels measured by ChIP-qPCR (n =2 experiments and 3 cell lines per group) and DNA methylation measured by MSRE-qPCR. Linear regression, $R^2 = 0.9242$; $p = 0.0022$. **D)** Bisulfite amplicon sequencing from unmethylated or methylated clonal iPSCs (n= 3 cell lines per group). Each row is an individual cloned and sequenced DNA molecule and columns are individual CpG dinucleotides. Grey boxes

indicate unmethylated CpGs, black boxes indicate methylated CpGs and white boxes are undetermined due to poor conversion or poor sequencing quality. **E)** Diagram of *C9orf72* gene structure. Black triangle indicates the Cas9 cut site, colored arrows indicate transcription start site and black arrow indicates the protein coding start site. **F)** RT-qPCR of all *C9orf72* variants in methylated or unmethylated clonal iPSCs. n =3 cell lines per group. Two-tailed t-test, p = 0.0021 **G)** RT-qPCR of *C9orf72* variant 2 in methylated or unmethylated clonal iPSCs. n = 3 cell lines per group. Two-tailed t-test, p = 0.0056. **H)** RT-qPCR of *C9orf72* variant 3 in methylated or unmethylated clonal iPSCs. n =3 cell lines per group. Two-tailed T-test, p = 0.0048.

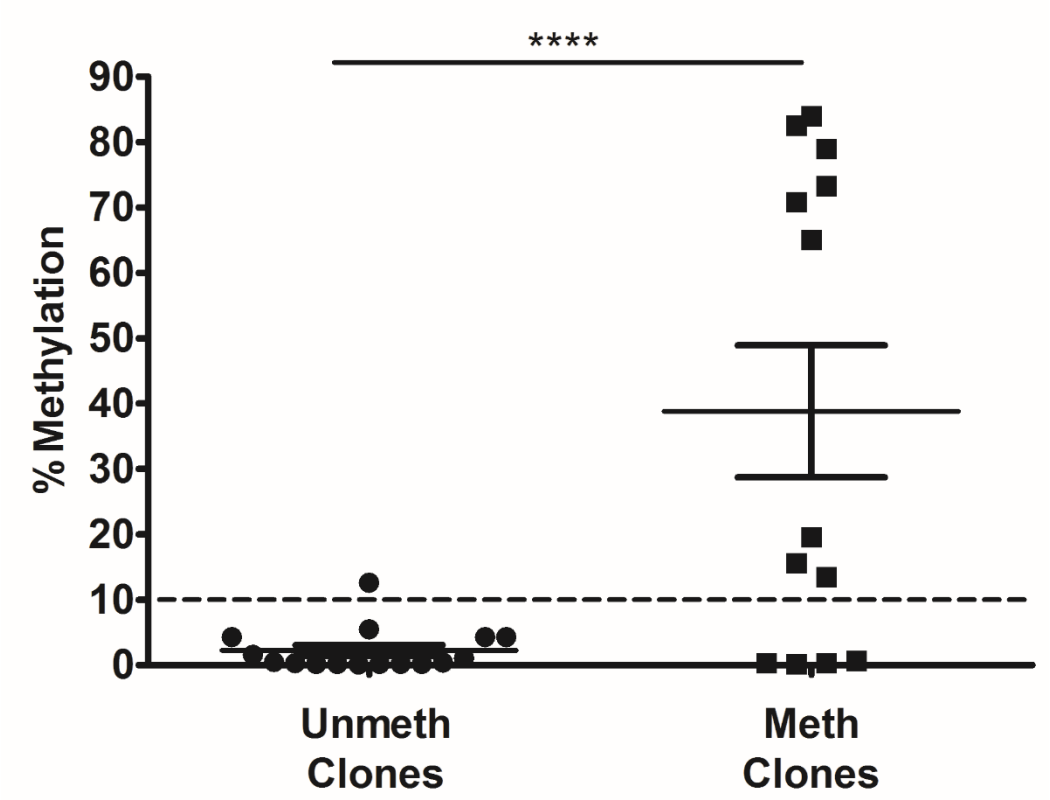
## Supplemental Figures and Tables



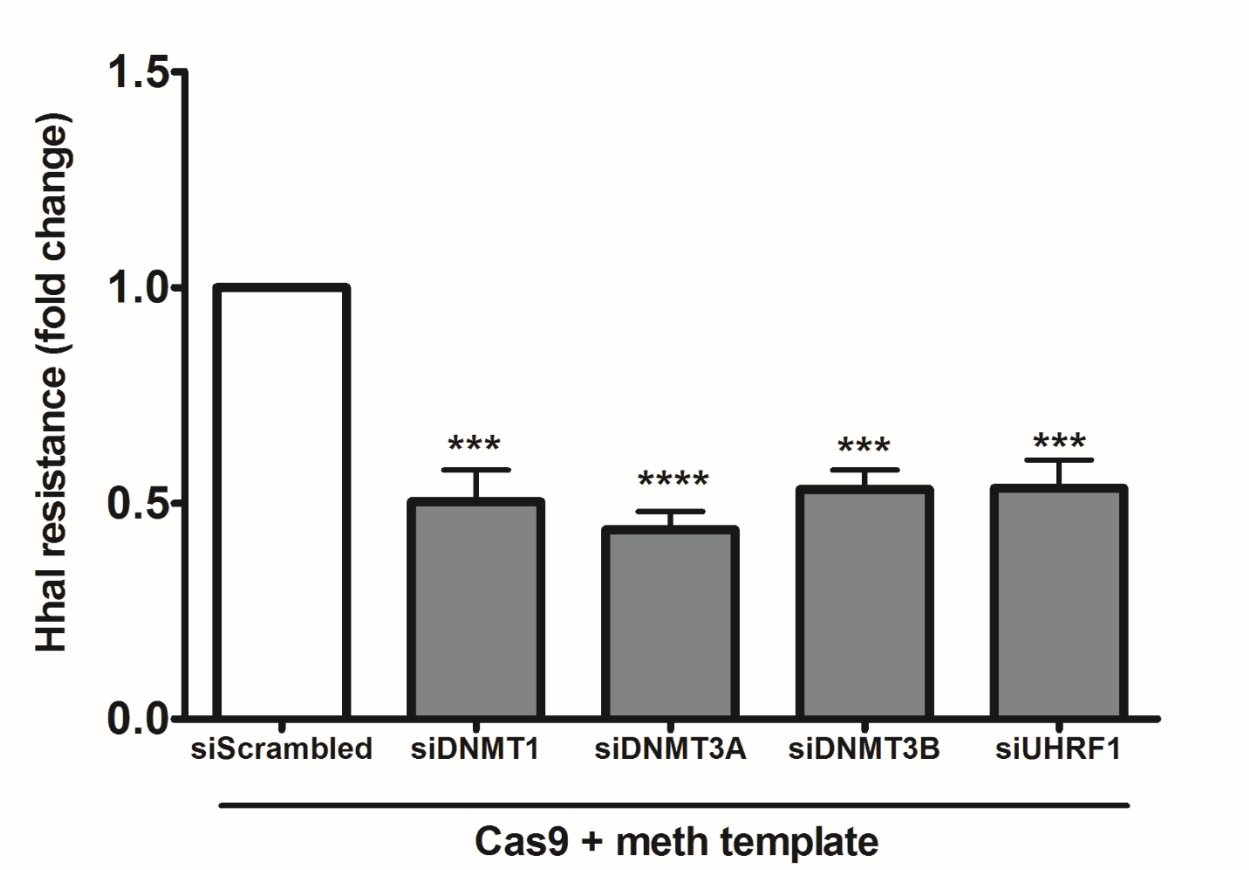

**Supplemental Figure 2.1:** Comparison of HDR efficiency in different length templates. Quantification of PCR + SspI Digest in HEK293T cells transfected with the indicated templates. Cells were puromycin selected for 2 days and collected at day 4 post transfection. n = 3 experiments. 2 way ANOVA shows template size has significant effect ($p < 0.0001$) whereas methylation status does not ($p = 0.6110$). Bonferroni post hoc tests were carried out between unmeth and meth groups at each template length.



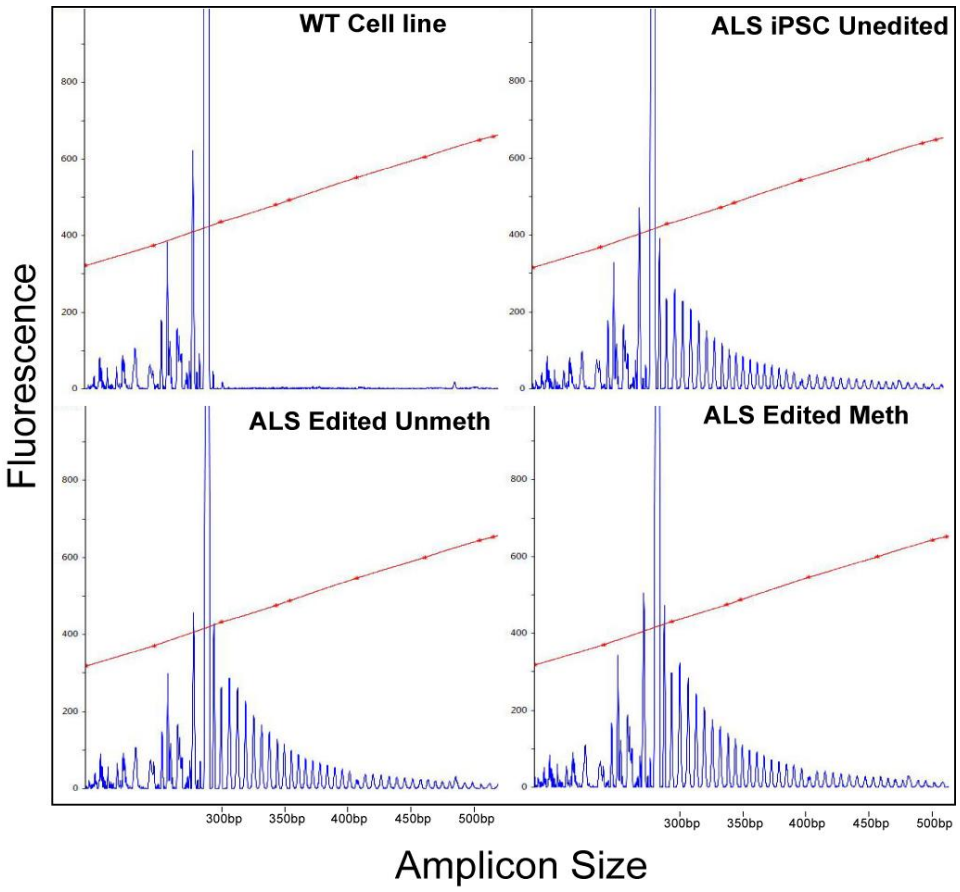

**Supplementary Figure 2.2:** Comparison of ssDNA vs dsDNA templates. MSRE-qPCR of HEK293T cells transfected with either 190bp ssOligo or 190bp dsDNA methylated templates and the CRISPR/cas9 components. Cells were puromycin selected for 2 days and collected at day 4 post transfection. n = 3 experiments. One Way ANOVA ($p = 0.0126$) followed by Bonferroni test.



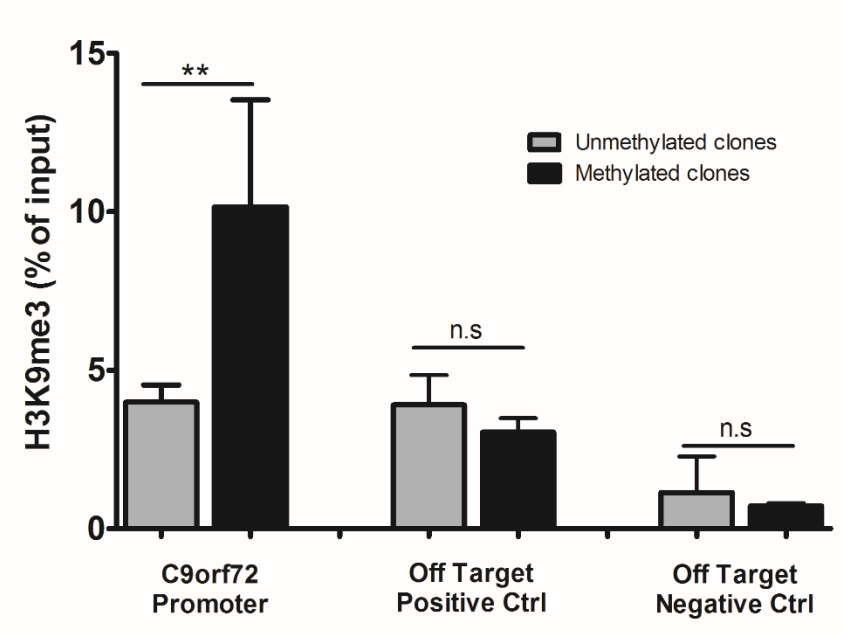

**Supplementary Figure 2.3:** MSRE-qPCR of HEK293T clonal lines repaired with either an unmethylated or methylated repair template. Cells were analyzed for HDR and methylation ~5 weeks post transfection. Shown here are clones that had at least 25% HDR (measured via SspI digest). n = 16 unmeth clones, 13 meth clones. Two-Tailed T-Test, p = 0.0004.




**Supplementary Figure 2.4:** Effect of DNMT knockdown on targeted methylation of *C9orf72*. siRNA was transfected 24 hours prior to transfection of CRISPR components and 1.4kb methylated template into HEK293T cells. Cells were collected 4 days post transfection. Methylation was measured via long amplicon MSRE-qPCR and normalized to scrambled control. n = 3 experiments; 1 way ANOVA followed by Bonferroni post hoc tests. *** $p < 0.001$; **** $p < 0.0001$.




**Supplementary Figure 2.5:** Repeat Primed PCR of iPSCs derived from ALS patient with C9orf72 repeat expansion. Capillary electrophoresis traces from fluorescently labeled repeat primed PCR on unedited or edited clonal iPSC lines repaired with either the 1.4kb unmethylated dsDNA template or methylated dsDNA template.




**Supplementary Figure 2.6:** H3K9me3 chromatin immunoprecipitation followed by qPCR of unmethylated and methylated clonal iPSCs. Off-target positive control is the *PAPC1* gene body, whereas the negative control is the *PAPC1* promoter region. n = 3 cell lines per group and 2 technical replicates. One way ANOVA ($p = 0.0001$) followed by Bonferroni post hoc tests between indicated groups. **$p < 0.01$.

## Table 2.S1: Primers and gRNA sequences

| Forward Primer Sequence (5'-3') | Reverse Primer Sequence (5'-3') | Assay |
|---|---|---|
| CAGTGTGAAAATCATGCTTGAGAGA | TTTGTGCTTGGTAGGCAGTG | DNA methylation analysis q-PCR; ChIP-qPCR |
| TGCTCTTTGAGAAAATTCATTGG | TTTGTGCTTGGTAGGCAGTG | DNA methylation analysis q-PCR (long templates) |
| TGGCACTATTAAGGATCTGAGGAG | GGAGAGAGGGTGGGAAAAAC | HDR analysis PCR (long templates) |
| TACCAGGGTTTGCAGTGGAG | GGAGAGAGGGTGGGAAAAAC | T7/HDR analysis (oligo templates) |
| GGAGTTGTTTTTTATTAGGGTTTGTAGT | TAAACCCACACCTACTCTTACTAAA | Bisulfite Cloning PCR |
| AGTTTTTTTTATTAAGGTTYGTATA | TTTCTTCTAATTAATCTTTATCAA | Bisulfite Cloning Primer Insert - WT |
| TTTAGTAGTTTTTGAATAATATTGG | TTTCTTCTAATTAATCTTTATCAA | Bisulfite Cloning Primer Insert - Mutant |
| CCCACTTCATAGAGTGTGTGTTG | TTCCATTCTCTCTGTGCCTTC | Total *C9orf72* mRNA (all variants) |
| CGGTGGCGAGTGGATATCTC | TGGGCAAAGAGTCGACATCA | *C9orf72* variant 2 mRNA |
| GCAAGAGCAGGTGTGGGTTT | TGGGCAAAGAGTCGACATCA | *C9orf72* variant 3 mRNA |
| 6FAM-AGTCGCTAGAGGCGAAAGC | TACGCATCCCAGTTTGAGACG | *C9orf72* Repeat Expansion Detection |
| TACGCATCCCAGTTTGAGACGGGGGCCGGGGCCGGGGCCGGGG | | *C9orf72* Repeat Expansion Detection |
| CACCGTAGCGTGTGCGAACCTTAAT | AAACATTAAGGTTCGCACACGCTAC | CRISPR gRNA sequences (oligos for cloning) |

**Table 2.S2: Repair template sequences with PAM mutation highlighted**

| Name | Sequence | Length (bp) |
|---|---|---|
| Non-homologous template | CAGAAGAAGAAGGGCTCCCATCACATCAACCGGTGGCGCATTGCCACGAAGCAGGCCAATGGGGAGGACATCGATGTCACCTC CAATGACAAGTTGCT AGCGGTGGGCAACCACAAACCCACGAGGGCAGAGTGCTGCTTGCTGCTGGCCAGGCCCCTGCGTGGGCCCAAGCTGGACTCT GGCCACTCCCT | 191 |
| 190bp ssOligo template | CTAGCGAAATGGGGCGGGGCAACTTGTCCTGTTCTTTTATCTTAAGACCCGCTCTGGAGGAGCGTTGGCGCAATAGCGTGTGCG AAC CTTAATATTGGAGGCTGCTGGATCTGGAGAAAGTGAAGACGATTTCGTGGTTTTGAATGGTTTTGTTTGTGCTTGGTAGGCAGTG GGCGCTCAACACATAATT | 190 |
| 600bp dsDNA template | AGGCACCGCA[ACCGCAGCCCCGCCCCGGCCCCTAGCGCGCGACTCCTGAGTTCCAGAGCTTGCTACAGGCTGCGGTTGTTTCCCTCCTTGTTTTCTTCTGGTTAACTCTTATCAGGTCTTTTCTTGTTCACCCTCAGCGAGTACTGTGAGAGCAAGTAGTGGGGAGAGAGGGTGGGAAAACAAAACACACACC TCCTA AACCCACACCTGCTCTTGCTAGACCCCGCCCCCAAAAGAGAAGCAACCGGGCAGCAGGGACGGCTGACACACCAAGCGTCACTTTTACGTGGCGGAACTTGTCGCTGTTTGACGCACCTCTCTTTCCTAGCGGGACACCGTAGGTTACGTCTGTCTGTTTCTATGTGCGATGACGTTTTCTCACGAGGCTAGCGAAATGGGGCGGG GCAACTTGTCCTGTTCTTTTATCTTAAGACACGCTCTGGATGAGCGTTGGCGCAATAGCGTGTGCGAACC[TTAATATTGGAGGC]GCTGCTGGATCCGAGAAAGTGAAGACGATTTCGTGGTTTTGAATGGTTTTGTTTGTGCTTGGTAGGCAGTG GTAAGA[CACTGTTAAG | 621 |
| 1.4kb dsDNA template | CCCCGCCCCG[GGCCCGCGCCCCGGGCCCGCCCCGACCACGCCCCGGCCCCGGCCCCCGGCCCCTAGCGCGCGACTCCTGAGTTCCAGAGCTTGCTACAGGCTGCGGTTGTTTCCCTCCTTGTTTTCTTCTGGTTAATCTTTATCAGGTCTTTTCTTGTTCACCCTCAGCGAGTACTGTGAGAGCAAGTAGTGGGG AGAGAGGGTGGGAAAACAAAACACACACCT CCTAACCCACACCTGCTCTTGCTAGACCCCGCCCCCAAAAGAGAAGCAACCGGGCAGCAGGGACGGCTGACACACCAAGCGTCACTTTTACGTGGCGGAACTTGTCGCTGTTTGACGCACCTCTCTTTCCTAGCGGGACACCGTAGGTTACGTCTGTCTGTTTCTATGTGCGATGACGTTTTCTCACGAGGCTAGCGAAATGGGGC GGGGCAACTTGTCCTGTTCTTTTATCTTAAGACCCG CTCTGGAGGAGCGTTGGCGCAATAGCGTGTGCGAACC[TTAATATTGGAGGC]GCTGCTGGATCTGGAGAAAGTGAAGACGATTTCGTGGTTTTGAATGGTTTTGTTTGTGCTTGGTAGGCAGTGGGCGCTCAACACATAATTGGTGAATGAAATTTTGTTTTACCGTAAGACACTGTTAAGTGCATTCAAAACTCCAC TGCAAACCCTGGTAGGGGACAGCTCCGGC ACTGCGGGCGGGAATCCCACGGTCCCCTGCAAAGTCACTCGCAATTTGCCTTACATGTAAGAATTCTCTCAAGCATGATTTCACACTGGCGAATGTCATTTCACACTGGCGAATGTCATTTCACACTGGCGAATGAGTTGTTTTTTTTAACTTTGAAAAACGTACCATTCTGTTTGATGTGTAAAAAACACAAAGATTTTTGAAACCTTGCGTCTTTTGGTCTG TGAGTAGCATTTACACCACTCAGATGTGTAAAAAAACAAAGGTTTTTTAAACTGTGTGCCTTTGATCTGCAAGTGTGAGATGGCACTTACTAACGATGAGTAGCATTTACACCACTCAGATGTGTAAAAAAACAAAGGTTTTTAATCATTCACCTAAAATCACACAGAACGTTTTAATCATTCACCTAAAATCACACAGAACGTTTTCATCACTGCTCAAAAGGAGAAAGGGAGGAATAAATACACAAAATGGCTCTCAACGTCTACACCTTCTGCAGA CTTTGTGAAAGTTGATCATACAAATTGGGTCATTCTTTTATACCCAACTAAAATAGTGGGGGTAGGGGGTAGAAAAGCACTTAGACAAATGACACTGCTCCCACACGTGTAATTCTCTCCAAGTCCAGCTGCTGCCAACTGCCCGTTGTGACCTGAGACCAGTTTATCTAATAGTTGCTAAAATGACCTGCTGCAGCTCTAATTTATCTACCA CCATCA[CTCACCAGTT | 1415 |
| PCR-insert template | AACTGGTGAG[TGATGGTGGTAGATAAAATTAGAGCTGCAGCAGGTCATTTTAGCAACTATTAGATAAACTGGTCTCAGGTCACAACGGGCAGTTGCAGCAGCTGGACTTGGAGAGAATTACACTGTGGGAGCAGTGTCATTTGTCCTAAGTGCTTTTCTACCCCCTACCCCCACTATTTAGTTGGGTATAAAAGA ATGAGCCACATTTGTATGATCAACTTTCACAAA GCATAGAACAGTAGGAAAAGGGTCTGTTTCTGCAGAAGGTGTAGACGTTGAAGAGCCATTTTGTGTATTTATCTCCTTCTCCTCGGTGAATGATTAAAACGTTCTGTGTGATTTTTAGTGATGAAAAAGATTAAATGCTACTCACTGTAGTAAGTGCCATCTCACACTTGCAG ATCAAAGGCACACAGTTTAAAAAACCTTTGTTTTTTTACACATCTGAGTGGTGTAAATGCTACTCACTCTGTAGTAAGTGGAATCTATACACCTGCAGACCAAAGACGCAAGGTTTCAAAAATCTTTGTGTTTTTACACATCAACAGATGGTACGTTTTCAAAAGTTAAAAAAAACAACTCATCACATATTGCAACTAGCAAAAATGACA TTCCCCAGTGT GAAACTCATGCTTGAGAGAATTCTTACATGTAAAGGCAAAATTGCGATGACTTTGCAGGGGACCCTGGGATTCCCGCCGCGACGTGCCCGGAGCTGTCCCCTACCAGGGTTTGCAGTGGAGTTTTGAATGCACTTAACAGTGTCTTACGGTAAAACAAATTTCATCCACCAATTATGTGTTGAGCGCCCACTGCCTACCAAGCACAA ACAAAACCATTCAAACCACGAAATCGTCTTCACTTTCTCCAGATCCAGCAGCCTCC[GAATAATATTGCC]GCTATTAAGGTTCGCACACGCTATTGCGCCAACGCTCCTCCAGAGCGGGTCTTAAGATAAAAGAACAGGACAAGTTGCCCGCCCCATTTCGCTAGCCTCGTGAGAAACGTCATCGCACATAGAA AACAGACAGACGTAACCTAC GGTGTCCCGCTAGGAAAGAGAGGTGCGTCAAACAGCGACAAGTTCCGCCACGTAAAAGATGACGCTGGTGTGTCAGCCGTCCCTGCTGCCCGGTTGCTTCTCTTTGGGGGCGGGGTCTAGCAAGAGCAGGTGTGGGTTTAGGAGGTGTGTGTTTTTGTTTTCCCACCCTCTCTCCCCACTACTTGCTCT[CACAGTA CTC | 1248 |

## Data Availability

Genome-wide DNA methylation array data is publically available at GEO (GSE134996).


## Acknowledgements

We thank Madlyn Kates for technical assistance.

## Funding

This study was supported by a grant from the NIH (R01 NS095793).


## Conflicts of Interest

The authors declare no conflicting interests.

## Chapter 3: C9orf72 intermediate repeats and corticobasal degeneration

This chapter is adapted from a published manuscript:



Citation:

Cali CP, Patino M, Tai YK, et al. *C9orf72* intermediate repeats are associated with corticobasal degeneration, increased *C9orf72* expression and disruption of autophagy. Acta Neuropathol. July 2019:1-17. doi:10.1007/s00401-019-02045-5.

**Corresponding Author:**
Edward B. Lee
613A Stellar Chance Laboratories
422 Curie Blvd
Philadelphia, PA 19104
edward.lee@pennmedicine.upenn.edu

## Abstract

Microsatellite repeat expansion disease loci can exhibit pleiotropic clinical and biological effects depending on repeat length. Large expansions in *C9orf72* (100s-1000s of units) are the most common genetic cause of amyotrophic lateral sclerosis (ALS) and frontotemporal degeneration (FTD). However, whether intermediate expansions also contribute to neurodegenerative disease is not well understood. Several studies have identified intermediate repeats in Parkinson's disease patients, but the association was not found in autopsy confirmed cases. We hypothesized that intermediate *C9orf72* repeats are a genetic risk factor for corticobasal degeneration (CBD), a neurodegenerative disease that can be clinically similar to Parkinson's but has distinct tau protein pathology. Indeed, intermediate *C9orf72* repeats were significantly enriched in autopsy-proven CBD (n=354 cases, odds ratio=3.59, p-value=0.00024). While large *C9orf72* repeat expansions are known to decrease *C9orf72* expression, intermediate *C9orf72* repeats result in increased *C9orf72* expression in human brain tissue and CRISPR/cas9 knockin iPSC derived neural progenitor cells. In contrast to cases of FTD/ALS with large *C9orf72* expansions, CBD with intermediate *C9orf72* repeats was not associated with pathologic RNA foci or dipeptide repeat protein aggregates. Knock-in cells with intermediate repeats exhibit numerous changes in gene expression pathways relating to vesicle trafficking and autophagy. Additionally, overexpression of *C9orf72* without the repeat expansion leads to defects in autophagy under nutrient starvation conditions. These

results raise the possibility that therapeutic strategies to reduce *C9orf72* expression may be beneficial for the treatment of CBD.

## Introduction

Large hexanucleotide repeat expansions ($G_4C_2$) in *C9orf72* are the most common genetic cause of the neurodegenerative diseases amyotrophic lateral sclerosis (ALS) and frontotemporal degeneration (FTD)[43,44,116]. ALS/FTD is characterized neuropathologically by the presence of TDP-43 inclusions which are tightly linked to neurodegeneration[34]. While the pathophysiologic mechanisms that lead to TDP-43 aggregation and neurodegeneration in the setting of large *C9orf72* repeat expansions are still unclear, both a loss of normal *C9orf72* function and a toxic gain of function attributed to repetitive RNA and protein aggregates have been proposed[43,44,81,88,115]. At the protein level, C9orf72 regulates several vesicle trafficking pathways [56,67,70] and is necessary for proper myeloid cell function in mice, but knockout of the protein does not cause neurodegeneration[58,59]. On the other hand, the presence of the repeat expansion is sufficient to induce neurodegeneration in several mouse models[84,109]. There is also uncertainty in the repeat size threshold necessary to cause disease. Repeat expansions that cause ALS/FTD are typically hundreds to thousands of units long and can differ substantially between tissues[45,185], making it difficult to accurately determine the pathogenic repeat size threshold. There have been reports of pathogenic repeats as small as 55-100 units in patient blood[117,118]. In contrast, non-diseased individuals typically harbor repeats of 2-8 units[186]. Several studies have explored whether intermediate *C9orf72* expansions (e.g. those larger than controls but not large enough to cause

ALS/FTD) are associated with risk for other neurodegenerative diseases, and these have often yielded conflicting results[119,124,125,186–191]. For instance, a significant association between intermediate expansion carriers and Parkinson's Disease (PD) was reported in clinically diagnosed PD[125] and confirmed in a large meta-analysis[123], but these associations were not found in an autopsy confirmed PD cohort[126]. Other studies with very small numbers of cases have also suggested that intermediate *C9orf72* repeats are associated with atypical Parkinsonian syndromes[192,193]. Thus, there is some evidence that intermediate expansions in *C9orf72* may confer risk for neurodegenerative diseases other than ALS/FTD, but more exploration is needed to better understand the molecular pathways and exact neurodegenerative diseases that are affected.

Here, we sought to determine whether intermediate expansions in *C9orf72* are a risk factor for corticobasal degeneration (CBD), a rare neurodegenerative disease that shares some similar clinical features with PD such as limb rigidity, bradykinesia, tremor and apraxia[194]. Patients that present with these neurologic features are often diagnosed with corticobasal syndrome (CBS) because these symptoms can also be caused by several other neurodegenerative diseases; definitive diagnosis of CBD requires neuropathology autopsy examination[195]. CBD is characterized by hyper-phosphorylated tau aggregates in neurons and astrocytes that are morphologically distinct from those found in Alzheimer's disease and other tauopathies[196]. Only ~25-50% of patients diagnosed with CBS are confirmed as CBD at autopsy[197–199], making clinically defined cohorts difficult to use for genetic studies of CBD due to underlying neuropathological heterogeneity. In this study, we determine whether *C9orf72* intermediate expansions may be a risk factor for CBD by measuring repeat size in the largest cohort of autopsy confirmed cases ever tested. In

addition, we use post mortem CBD brain tissue to confirm the genetic findings, use CRISPR/cas9 to modify the repeat number in human iPS cells at the endogenous locus, and determine the effects of *C9orf72* expression on autophagy. In doing so, we identify a novel mechanism for how intermediate repeat expansions may affect CBD pathogenesis that is distinct from large expansions that cause ALS/FTD.

## Materials and Methods

### Case Samples

Genomic DNA for repeat expansion size screening was obtained from 354 autopsy confirmed CBD cases from 17 institutions. Institutions contributing the vast majority of cases included Mayo Clinc, University College London and the University of Pennsylvania. CBD patient cerebellum tissue for RNA expression studies was obtained from the Center for Neurodegenerative Disease Research brain bank at the University of Pennsylvania and from the Mayo Clinic brain bank. Brain autopsy material is not considered human subjects research. However, legal consent for autopsy was obtained in all instances. As a control cohort, we utilized a previously published global cohort of healthy controls for which *C9orf72* repeat sizing was available[123].

### Assessment of Repeat Expansion Size and SNP Genotyping

The CBD cohort was screened for intermediate C9orf72 repeat expansions using fluorescent labeled touchdown PCR followed by fragment-length analysis on an ABI 3130xl Genetic Analyzer (Applied Biosystems, Foster City, CA). PCR used 50ng

genomic DNA in a final volume of 20ul containing 0.5U of Amplitaq Gold (Applied Biosystems) and a final concentration of 1x Amplitaq Buffer I, 5uM reverse primer, 5uM 6FAM-fluorescent labeled forward primer, 1M Betaine, 5% DMSO, 0.25mM 7-deaza-dGTP, and 0.25mM of each dNTP. For detection of full expansions using repeat-primed PCR, the Roche FastStart (Roche, Basel, Switzerland) polymerase system was used to amplify 1ug of genomic DNA in a reaction with 0.9 mM $MgCl_2$, 200nM dNTP mix, 180nM 7-deaza-dGTP, 1.4uM flanking primers, 0.175uM repeat primer, 7% DMSO, .93M Betaine and 0.125 U polymerase. For both intermediate and full expansion detection, a touchdown PCR cycling program was used where the annealing temperature was gradually lowered from 70°C to 56°C in 2°C increments with a 3-minute elongation time at 72°C for each cycle. SNP genotyping was carried out using Taqman probes as previously described[43]. All primer sequences are listed in Supplemental Table 3.1.

**Detection of RNA foci**

RNA foci detection was performed on paraffin embedded cerebellum sections as follows: Slides were deparaffinized with Xylene, followed by Xylene/Ethanol (1:1) and then rehydrated with an ethanol series (100%, 90%, 80%,70% for 3 mins each). Slides were incubated for 5 minutes in water, then digested with proteinase K (20ug/mL) for 10 minutes at 37 degrees in a buffer containing 10mM Tris-HCL pH 8 and 0.5% SDS in DEPC-treated water. Slides were then washed twice with water then dehydrated with increasing ethanol series and allowed to air dry. Slides were then prehybridized for 1 hour at 66 degrees in hybridization buffer containing 50% Foramide, 2X SSC, 50mM $NaHPO_4$ , and 10% Dextran sulfate in DEPC-treated water. Slides were hybridized with

40nM probe (LNA 5TYE563C4G2; Exiqon (Denmark) # 300500, batch # 621440) at 66 degrees overnight. Slides were then washed once with 2X SSC + 0.1% Tween for 5 mins at room temperature, then 3 times with 0.1X SSC for 10 minutes at 65 degrees. Slides were then rinsed with water, stained with DAPI and mounted with coverslips using Prolong glass antifade mountant (Invitrogen, Carlsbad, CA). Cases for RNA foci detection included 12 non-expanded CBD, 9 intermediate CBD, 4 non-expanded control cases (negative control) and 5 full expansion FTD/ALS cases (positive control).

**Detection of Dipeptide Repeat proteins and Measurement of DNA methylation**

Paraffin embedded cerebellum sections were deparaffinized and stained with antibodies against p62, GA (1:7500), GP (1:7500) or GR (1:100) as previously described[200]. 116 non-expanded CBD, 9 intermediate CBD, 4 non-expanded control cases (negative control) and 5 full expansion FTD/ALS cases (positive control) were stained. For measurement of DNA methylation in CBD patient genomic DNA, 100ng DNA was digested using the restriction enzymes HhaI and HaeIII (New England Biolabs, Ipswich, MA) followed by qPCR as previously described[137].

**RNA Expression Analysis from Postmortem Brain and repeat edited NPCs**

RNA was extracted from approximately 100mg of cerebellum tissue using Trizol reagent (Life Technologies, Carlsbad, CA). RNA integrity was tested on an Agilent 2100 bioanalyzer using the RNA Nanochip 6000 kit (Agilent, Santa Clara, CA). Only RNA samples with RIN values > 6 were included in this study. For repeat edited NPCs, cells were pelleted at day 12 post induction and RNA was extracted using the RNeasy Kit

(Qiagen). cDNA was prepared using random hexamers and Superscript III (Invitrogen) according to manufacturer's protocol. RT-qPCR was performed using SYBR Green reagents (Roche) on a StepOne Plus Real-Time PCR Machine (Life Technologies). RNA levels were normalized to the geometric mean of two housekeeping genes (ACTB, GPS1) using the ΔΔCt method. All primer sequences are listed in Supplemental Table 3.1.

**CRISPR Editing of Repeat Size**

CRISPR gRNAs were designed using the MIT CRISPR design tool (http://crispr.mit.edu/) and cloned into the Cas9-GFP PX458 vector (Addgene, Cambridge, MA) as previously described[201]. Repair templates were generated via PCR cloning using genomic DNA from intermediate repeat carriers. PCR was carried out using the FastStart Polymerase (Roche) with additives for GC rich templates (360nM 7-deaza-dGTP, 7% DMSO, 930mM Betaine). PCR products containing 2 or 28 repeats were cloned into pGEM Easy-T vector (Promega, Madison, WI) and verified by sequencing. A 2 bp substitution was made at the PAM site in the repair template using QuickChange site directed mutagenesis (Aligent) in order to prevent Cas9 cutting of the repair template and to incorporate a restriction enzyme site (See Supplemental Figure 3.3). PX458 and repair template vectors were co-transfected into WT iPSCs[202] using Viafect reagent (Promega). On day 2 post transfection, GFP+ cells were sorted via FACS (BD FACS Aria II) to 96 well plates for clonal isolation. DNA was extracted from 24 well plates of confluent clones using 0.2% SDS lysis buffer followed by phenol-chloroform extraction. DNA from clones was PCR amplified and screened for homology directed repair by restriction enzyme digest. Repeat size was measured as outlined above

on positive clones, and all CRISPR edited cell lines were Sanger sequenced via PCR cloning (10 clones/cell line).

**Differentiation of iPSCs to Neural Progenitor Cells**

IPSCs were differentiated via dual SMAD inhibition as previously described[203]. Single cells were plated at high density ($2x10^5/cm^2$) in mTeSR1 (StemCell Technologies, Vancouver, Canada) + 10uM ROCK inhibitor Y27632 (ATCC) on Matrigel (Corning, Corning, NY) coated 6 well plates. The following day, media was changed to neural induction media (see Shi et al.[203] for complete formulation) containing 10uM SB431542 (Tocris, Bristol, United Kingdom) and 1uM dorsomorphin (Sigma, St. Louis, MO). Neural induction media was changed each day. At day 6, cells were split 1:2 into neural induction media + ROCK inhibitor and media changes were continued until day 12. Day 12 NPCs were characterized for expression of NPC markers by immunofluorescence and RT-qPCR. For western blot analysis, NPCs were passaged again at day 12 and collected at day 16 post induction.

**Western Blots and Antibodies**

For detection of C9orf72, day 16 NPCs were pelleted, lysed (50mM Tris-HCl, 750mM NaCl, 5mM EDTA, 2% SDS) and sonicated using a probe sonicator (Sonics Vibra-cell), then spun at 21,000 g for 30 mins. BCA analysis (Pierce, Thermo Fisher) was performed on cleared lysate to quantify protein expression. 50ug of protein was run on an 8% Tris-Glycine gel, transferred using the Transblot Turbo system (Biorad, Hercules, CA) to a nitrocellulose membrane and blocked for 1 hour at room temperature with Odyssey Licor

Blocking buffer (Licor, Lincoln, NE). Membranes were blotted with the following antibodies overnight at 4 degrees: anti-C9orf72 1:1000 (Millipore, Burlington, MA, ABN1645), anti-GAPDH 1:1000 (Cell signaling Technology, Danvers, MA 2118S). IR dye secondary antibodies (Licor) were used to detect protein on a Licor Odyssey. Protein was quantified using Image Studio software (Licor).

For C9orf72 expression and autophagy experiments, cell lysates were prepared in ice-cold radio immunoprecipitation assay (RIPA) buffer containing 150 mM NaCl, 1% TritonX-100, 0.5% sodium deoxycholate, 0.1% SDS and 50 mM Tris (pH8.0) supplemented with protease and phosphatase inhibitors (PhosphoSTOP, Roche). Cells were lysed for 20 min and centrifuged for 10 min at 13,500 rpm. Protein concentration of the soluble fractions were determined using BCA reagent (Thermo Fisher Scientific). 20-70ug of total protein was resolved using 10% or 12% denaturing polyacrylamide gel electrophoresis and transferred to PVDF membrane (Immobilon-P, PVDF). Proteins on PVDF membranes were blocked using 5% low-fat milk in TBST containing 0.1% Tween-20 and incubated with primary antibody in SuperBlock (TBS, Thermo Fisher Scientific) overnight at 4°C. The primary antibodies used were: LC3 (ProteinTech, 14600-1-AP, 1:1000), and GAPDH (ProteinTech, 60004-1-Ig, 1:5000). The membranes were washed in TBST. Anti-rabbit or anti-mouse horse-radish peroxidase (HRP) conjugated secondary antibodies (1:3000, Thermo Fisher Scientific) in 5% milk in TBST were incubated with the membranes for 1 hour at room temperature. The membranes were incubated with WestPico or WestFemto chemiluminescent substrate (Thermo Fisher Scientific) and exposed to Amersham Hyperfilm ECL films (GE Healthcare Life Sciences). The films were developed using

Konica Minolta film developer. Protein band intensity was quantified using LI-COR Image Studio Lite.

**RNA sequencing in knockin intermediate repeat cells**

50 base pair paired end RNA sequencing was carried out on day 12 post induction isogenic NPC lines (See Supplemental Table 3.2) with 3 biological replicates per group and 2 technical replicates per cell line. Libraries were prepared using the TrueSeq Stranded mRNA and Ribo-zero rRNA removal kits (illumina, San Diego, CA). Sequencing was carried out on a Hiseq 2500 (illumina) and at least 100 million reads were generated per biological replicate. Adapter sequences were removed using Cutadapt[204]. Reads were aligned to the genome (GRCh38) using STAR (v2.6) [205] with the default mapping parameters and the read counts per gene were counted using the quantMode function and genecode v29 annotation. Differentially expressed genes between the short repeat group and intermediate repeat group were called using the DEseq2 program (v1.22)[206] with default parameters. Over enrichment analysis of biological pathways was carried out using WebGestalt (2019) on genes that were either upregulated or downregulated (adjusted p-value < 0.05) in the intermediate repeat group. For comparison of genes dysregulated in full expansion carriers, differentially expressed gene lists (adjusted $p < 0.05$ and >2 fold change) from two studies[81,83] that compared gene expression in iPSC derived neurons with full expansions vs. low repeat controls were compared against the differentially expressed genes in this study (intermediate repeat vs low repeat).

### Generation of stable C9orf72 expressing cell lines, transfection and drug treatments

The generation of an inducible single-copy C9orf72 transgene using Flp-In TRex-HeLa cells was as described previously[207]. All cell lines used in this study were maintained in DMEM (GE Healthcare HyClone, SH30243.01) with 10% FBS (GE Healthcare HyClone, SV30160.03) and 1% Pen-Strep (Thermo Fisher Scientific, 15070063) at 37°C and 5% $CO_2$. Tetracycline (Sigma-Aldrich, T7660) was added (100 ug/ml) for 18 h to induce the expression of C9orf72 transgene. Parental HeLa cells treated with tetracycline were used as controls in the experiments shown in Figure 3.6b and 3.7b. All plasmid transfections in cell lines were performed using Lipofectamine 2000 (Thermo Fisher Scientific, 11668019) or TurboFect (Thermo Fisher Scientific, R0531) as per the manufacturer's instruction. For autophagy experiments, Chloroquine (CQ, Sigma-Aldrich) was dissolved in water (20 mM) and used at 40μM in the media. Autophagy induction using Earle's Balanced Salt Solution (EBSS, Gibco) was incubated on cells following two successive PBS rinses. Incubation time for each condition is stated in the respective figure legends.

## Results

### Screening of CBD patient DNA for *C9orf72* expansions

To test the hypothesis that intermediate size expansions in *C9orf72* are a risk factor for CBD, genomic DNA from 354 autopsy confirmed CBD cases was screened using a PCR-based sizing assay (Figure 3.1a). We compared CBD patients to controls from a previously published global meta-analysis of *C9orf72* repeat size [123]. We found that 1.84% of alleles from CBD cases (corresponding to 3.7%, or 13 cases) harbored

intermediate length repeat expansions compared to 0.52% of controls (Figure 3.1b-c, OR = 3.59, p = 0.00024). This analysis was based on a minimum size cutoff of 17 which was determined a priori and based on prior independent studies[123,208]. The increase in *C9orf72* intermediate repeat allele frequency was not due to an increase in the frequency of the *C9orf72* repeat expansion haplotype between CBD patients verses non-diseased controls or general global populations (Supplemental Table 3.3), and we did not detect any expansions larger than 29 repeats (Supplemental Table 3.4). We conducted a post-hoc analysis on our CBD cohort for each repeat length cut off below 18 repeats which revealed that repeats as low as 10 units were associated with CBD (p = 0.026, OR = 1.36), where 61 out of 354 CBD cases had *C9orf72* repeat alleles that were at least 10 units. This included 5 cases which had two alleles at 10 or greater. Although this post-hoc analysis is considered exploratory[209], there appeared to be a somewhat graded increase in CBD risk with increasing repeat size (Figure 3.1d, Supplemental Table 3.5). Overall these results indicate that intermediate expansions in *C9orf72* may be a genetic risk factor for CBD, and the risk may be additive with higher repeat sizes conferring increased CBD risk.

### *C9orf72* expression from intermediate expansion carriers

Next, we determined how intermediate repeat expansions affect expression of the *C9orf72* gene. Previous studies have shown that large expansions in *C9orf72*, which cause ALS/FTD, reduces mRNA from the expanded allele[43,54,210]. This reduction in expression is often associated with hypermethylation of the *C9orf72* promoter region[47,136,137,140]. To test whether intermediate expansions alter *C9orf72* DNA

methylation and mRNA levels, we analyzed post mortem cerebellum samples from a subset of 47 CBD cases for which frozen tissue was readily available. We measured repeat expansion size, DNA methylation and *C9orf72* mRNA levels of the two major isoforms[62] (Figure 3.2a, V2 and V3 isoforms). In this subgroup analysis, we again observed high frequencies of intermediate repeat expansions between 17 and 29 units (Supplemental Figure 3.1). Intermediate size expansions did not correlate with hypermethylation of the promoter region, as is observed in full expansion cases (Supplemental Figure 3.2). Surprisingly, total *C9orf72* mRNA levels were positively correlated with total *C9orf72* repeat size (Figure 3.2b left, $R^2 = 0.276$, $p = 0.0034$).This positive correlation was not observed for the major isoform, V2, which has a transcription start site downstream from the repeat expansion (Figure 3.2c left, $R^2 = 0.00496$, $p = 0.717$). In contrast, V3 has a transcription start site upstream from the repeat expansion and shows a robust positive correlation between repeat expansion size and mRNA levels (Figure 3.2d left, $R^2 = 0.667$, $p < 0.0001$). Cases with >17 total repeats exhibited ~2.5 to 5 fold increase in variant 3 mRNA compared to cases with shorter repeats. Finally, analysis of the variant 3 to variant 2 ratio also reveals a striking positive correlation with larger repeat sizes (Figure 3.2e left, $R^2 = 0.849$, $p < 0.0001$). Similar results were obtained when using the largest repeat allele as opposed to the total *C9orf72* repeat length (Figure 3.2b-e, right). However, total repeat number produced a stronger positive correlation to *C9orf72* mRNA levels, supporting an additive effect of repeat size on expression.

To confirm the positive relationship between *C9orf72* repeat size and mRNA expression, we analyzed publicly available SNP eQTL data from the Genotype-Tissue

Expression (GTEx) Project[211]. Large *C9orf72* repeat expansion mutations invariably occur within a 140kb haplotype block with a defined series of SNPs[116,212]. The SNP rs3849942 is a commonly used marker for this haplotype block in which the G allele is associated with small non-expanded *C9orf72* alleles, and the A allele is associated with both larger intermediate to full repeat expansions[43]. We observe the same trend in our CBD genomic DNA cohort where the rs3849942 A risk allele was associated with larger *C9orf72* repeat sizes (Figure 3.2f). Using rs3849942 as a marker for longer/intermediate repeats (as full expansions are rare in unselected populations), we used GTEx data to determine if this SNP is associated with altered *C9orf72* expression. We found that the risk allele is significantly associated with increased *C9orf72* expression across numerous tissue types, with the largest effect size occurring in neural tissues (Figure 3.2g). Expanding the analysis to 20 haplotype defining SNPs[212], we found that 12 of the SNPs on the risk allele are associated with increased *C9orf72* expression in both the cortex and basal ganglia (Figure 3.2h ). In contrast, there were no significant eQTL SNPs that led to a decrease in expression. These results provide additional evidence that intermediate expansions lead to increased *C9orf72* expression.

**CRISPR Editing of *C9orf72* expansion size in iPSCs**

In order to confirm the previous findings experimentally and to determine whether the risk haplotype versus the intermediate *C9orf72* repeats themselves are driving increased *C9orf72* gene expression, CRISPR/cas9 was used to modify the repeat size in previously characterized induced pluripotent stem cells (iPSCs) from a non-diseased individual[202]. A repair template containing either 2 or 28 repeats was used to

alter the size of the repeat expansion in iPSCs that normally have alleles of either 2 or 6 repeats (Figure 3.3a-b; Supplemental Figure 3.3). Clonal lines were derived by FACS and were screened for changes in repeat size and for the presence of the knocked-in allele by PCR followed by restriction enzyme digest (Figure 3.3b). Genotype, including repeat size, was confirmed by repeat primed PCR (Figure 3.3c) and Sanger sequencing (Supplemental Figure 3.4). Despite using a repair template with 28 repeats, several different iPSC clones with repeat sizes ranging from 23 to 28 were obtained. Additional isogenic iPSC clones were also obtained with two or six *C9orf72* repeats.

Given that both neurons and glial cells are affected in CBD[194], we used the dual SMAD inhibition method[203,213] to differentiate edited iPSCs with intermediate *C9orf72* expansions and isogenic controls with small repeats into neural progenitor cells (NPCs) as determined by uniform expression of the NPC marker Pax6 (Supplemental Figure 3.5). RT-qPCR of these edited NPC lines showed a remarkably consistent pattern of higher *C9orf72* mRNA expression in clones with intermediate *C9orf72* repeats compared to isogenic controls with small repeats (Figure 3.3d). Consistent with the data from CBD post mortem brain tissue, we again saw no change in the amount of variant 2 mRNA (Figure 3.3e) but a robust increase in variant 3 mRNA levels in all intermediate *C9orf72* repeat cell lines (Figure 3.3f-g). This effect was not an artifact of neuronal differentiation, as undifferentiated iPSCs also exhibited increased *C9orf72* mRNA levels (Supplemental Figure 3.6). Thus, these experiments demonstrate a causal relationship between intermediate *C9orf72* expansions and increased *C9orf72* mRNA levels.

**Intermediate repeat carriers do not have RNA foci or DPR pathology**

Full expansion carriers have characteristic neuropathology in post mortem brain and spinal cord tissue that includes nuclear RNA foci of $G_4C_2$ transcripts[43] and aggregates of dipeptide repeat proteins (DPRs) that co-stain with p62[94]. Staining for DPRs using p62 antibody or DPR specific antibodies (GA, GP and GR) was carried out on 125 CBD cases in hippocampus and/or cerebellum, including 9 cases with intermediate repeats (repeat sizes 17-29, Figure 3.4a). Interestingly, all of these cases were negative for the aggregates typically seen in full expansion cases (Figure 3.4a). Additionally, RNA foci were not detected in CBD cerebellum tissues with intermediate repeats, nor were they detected in cases with low repeat numbers (Figure 3.4b). In comparison, full expansion FTD/ALS cases had abundant DPR aggregates together with RNA foci in ~18% of nuclei. Thus, the major pathologic hallmarks of full expansion carriers are not present in CBD cases, leading us to hypothesize that the underlying mechanisms of intermediate expansions and CBD risk may be distinct from the large repeat expansions that cause ALS/FTD.

**Elevated C9orf72 protein and global gene expression changes in intermediate repeat cell lines**

To test whether C9orf72 protein expression is higher in cells with intermediate repeats, protein expression was measured in isogenic repeat edited NPCs. NPC lines with intermediate repeats exhibited a consistent ~40% increase in C9orf72 protein levels when compared to isogenic low repeat number controls (Figure 3.5a-b). Thus, intermediate size repeat expansions increase both *C9orf72* mRNA and protein levels.

To determine whether overexpression of C9orf72 in the context of intermediate repeats alters expression of other genes, RNA sequencing was performed on 3 edited intermediate NPC lines and 3 low repeat isogenic control lines (Supplemental table 3.2, Supplemental file 1). We found 1,307 genes significantly downregulated (adjusted p-value < 0.05 and fold change > 2) in intermediate repeat cell lines and 959 genes upregulated (Figure 3.5c). Next, overrepresentation analysis was performed on gene ontology pathways in the differentially expressed genes. Genes down regulated in intermediate repeat cell lines (Figure 3.5d) are highly enriched for metabolic processes, especially those involving nutrient catabolism. Genes upregulated in intermediate repeat cell lines (Figure 3.5e) are enriched in vesicle trafficking and protein degradation pathways including golgi vesicle transport, response to ER stress and autophagy. We then looked in more detail at specific genes within the most highly dysregulated biological processes. When categorizing vesicle trafficking genes according to their function[78], there is modest increase in expression of numerous endosomal/lysomal genes, including *C9orf72* (Figure 3.5f). Interestingly, there is a consistent and sizable decrease in expression of genes involved in secretory pathways in intermediate repeat cell lines, indicating a possible defect in vesicle secretion (Figure 3.5f). Additionally, key genes involved in autophagy are modestly increased in intermediates (Figure 3.5g). However, several genes that have been shown to regulate autophagy in response to nutrient starvation[214–216] and ER stress[217] such as *TMEM74* and *TMEM74B* are markedly downregulated in intermediate repeat cells (Figure 3.5g), suggesting a defect in autophagy under stress conditions. Finally, we compared the differentially expressed genes in our dataset to previously published[81,83] gene expression studies between control

(low repeat) and full expansion cells derived from ALS/FTD patients. Interestingly, we see a distinct group of dysregulated genes between control and intermediate repeat cells than between control and full expansion cells (Figure 3.5h), indicating that different processes are affected by intermediate expansions versus full expansions.

**Over-expression of C9orf72 impairs starvation-induced autophagy**

Next, a cell culture model was used to further validate the observed changes in biological pathways in the RNA sequencing data and determine if these were in fact caused by *C9orf72* over expression. Since *C9orf72* is known to regulate autophagy[68,75,76,218] and this pathway was altered in the RNA sequencing data, we generated an inducible isogenic HeLa cell line expressing a single-copy of human *C9orf72* long isoform without the repeat expansion tagged with GFP and 6xHis-FLAG tag. The transgene expression level is comparable, i.e., 1:1, to that of endogenous *C9orf72* and is under the tight control of tetracycline (Supplemental Figure 3.7 and [207]). Importantly, the transgene also associates with the autophagy initiation complex, thereby confirming that the GFP-tagged C9orf72 recapitulates the known endogenous gene function[207]. As a readout of autophagy, we utilized an LC3 (microtubule-associated protein light chain 3)-based assay (Figure 3.6a). During autophagosome assembly, LC3-I undergoes lipidation via conjugation with a phosphatidylethanolamine group to become LC3-II [219,220]. LC3-II is normally rapidly degraded during autophagy via fusion with the lysosome, but blocking lysosome fusion allows for accumulation of LC3-II to serve as a marker of autophagic activity[221]. Immunoblotting in C9orf72 overexpressing cells revealed that basal LC3-II levels are ~20% higher in C9orf72 overexpressing cells

compared to parental HeLa cells (lane NGM, normal growth medium, Figure 3.6b-c), indicating that either autophagosome production is increased or lysomal degradation of LC3-II is reduced under normal conditions. When autophagosome-lysosome fusion was blocked with chloroquine (CQ)[222], LC3-II accumulates at a higher level in CQ-treated condition when compared to their untreated counterparts, and at a similar level between the parental HeLa and C9orf72 overexpressing cells (Lane NGM+CQ, Figure 3.6b-c). Interestingly, when autophagy was induced using nutrient-deficient medium, EBSS (Earle's balance salt solution), the accumulation of LC3-II was reduced by ~40% in the C9orf72 overexpressing cells when compared with parental HeLa cells (Lane EBSS+CQ, Figure 3.6b-c), suggesting that C9orf72 overexpression blocks starvation-induced autophagy. To further corroborate the biochemical findings, a GFP-LC3 stable HeLa cell line[223] was transfected with C9orf72 construct linked via P2A peptide to mCherry to better visualize the transfected cells. Consistent with the immunoblotting results above, ~50% more LC3-II puncta were observed in the C9orf72 overexpressing conditions in normal growth media. (NGM panel, Figure 3.6d-e). However, there was ~30% reduction of LC3-II puncta in C9orf72 overexpressing cells under starvation (EBSS+CQ panel, Figure 3.6d-e), again indicating that there is a failure of autophagy initiation in starvation conditions.

To more accurately measure the actual autophagic flux or rate, a time course without and with chloroquine treatment was performed (Figure 3.7a). With the lysosomes blocked, LC3-II accumulated at a similar rate in C9orf72 overexpressing cells under normal conditions as the parental HeLa cells (Figure 3.7b-c, top blot). Under starvation conditions, however, the rate of accumulation was markedly reduced (Figure 3.7b-c,

bottom blot) compared to the parental HeLa cells. Taken together, these results suggest that C9orf72 overexpression promotes autophagy under nutrient rich conditions but limits the cell's ability to induce autophagy during times of nutrient stress.

## Discussion

In this study, we demonstrate that intermediate repeat expansions in *C9orf72* may be a genetic risk factor for CBD, thus broadening the spectrum of neurodegenerative diseases implicated by this gene. This study highlights the importance of using autopsy confirmed cases to better understand the genetics of neurodegenerative diseases that have similar or overlapping clinical manifestations. Although intermediate repeats are found in only a small percentage of CBD patients, the magnitude of risk conferred by intermediate repeats is comparable or greater than other known genetic risk factors for CBD[224]. This study assembled the most CBD cases yet tested, and due to the rarity of CBD, a replication cohort is not yet available for follow-up analysis. While we established an a priori hypothesis based on the previously established but arbitrary cutoff of 17 repeat units or larger to define intermediate expansion carriers, our findings suggest that both CBD risk and *C9orf72* expression increases in an additive manner with increasing repeat length where perhaps C9orf72 intermediate expansions could be defined as at least 10 units. Interestingly, initial reports of *C9orf72* repeat mutation cohorts, defined at that time as at least 30 repeats, included one case of autopsy-confirmed CBD[225]. We did not detect any expansions larger than 29 repeats in our CBD cohort, and intermediate expansion carriers did not have the characteristic RNA foci and dipeptide repeat pathology seen in

full expansions. Future studies will need to further define the size threshold that determines whether carriers are more likely to develop risk for CBD versus ALS/FTD.

*C9orf72* joins a list of other repeat expansion containing genes in which the length of the repeat confers a pleiotropic effect on the disease phenotype. For instance, intermediate length CGG repeats in the gene *FMR1* result in the neurodegenerative disease Fragile X-associated tremor/ataxia syndrome (FXTAS) linked to toxicity of repeat RNA[226], whereas full expansions lead to near complete silencing of *FMR1* and the neurodevelopmental disorder Fragile X Syndrome[50]. Similarly, intermediate CAG expansions in *Ataxin-2* are a risk factor for ALS[227], whereas full expansions cause spinocerebellar ataxia type 2[228]. For *C9orf72*, large expansions cause ALS/FTD[43,44], whereas we demonstrate that intermediate expansions confer ~3.5 fold higher risk for CBD. Full expansions mutations in *C9orf72* have been associated with reduction of *C9orf72* gene expression and a gain of repeat associated pathology including nuclear repeat RNA accumulation and dipeptide repeat protein production[43,62,63,115]. In contrast, we show that intermediate expansions cause an increase in *C9orf72* mRNA levels in both CBD post mortem cerebellum and knock-in isogenic neural cell lines. This mechanism is also supported by publicly available SNP eQTL data, which shows that numerous SNPs on the expansion haplotype are associated with increased *C9orf72* expression. Thus, it appears that maintaining proper levels of C9orf72 may be necessary for proper neuronal function, with increased expression observed in CBD patients and decreased expression observed in ALS/FTD patients.

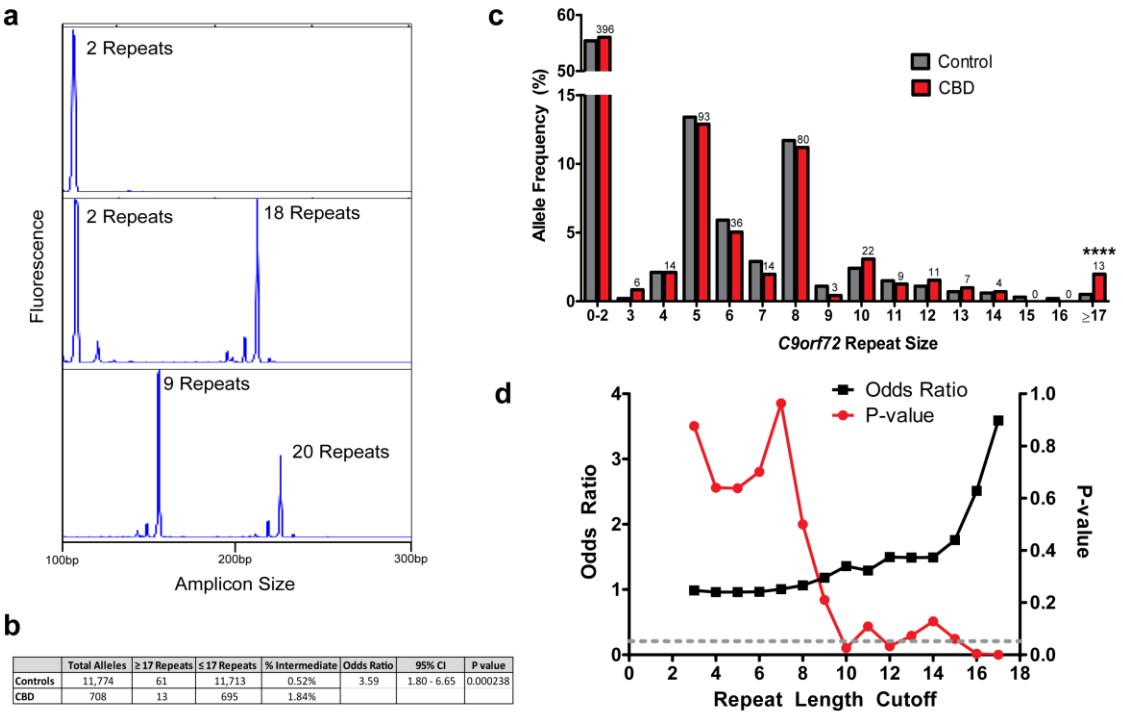
C9orf72 is thought to function as Rab guanine exchange factor[55], which participates in a host of vesicle trafficking pathways[65]. C9orf72 has been shown to

regulate endocytosis[56], vesicle secretion[70], intracellular trafficking[56,58,70,115] and autophagy[67–69,76,218,229]. Indeed, we observe numerous gene expression changes in all of these vesicle trafficking pathways in knock-in NPCs with intermediate expansions. The large number of significant gene expression changes detected in this study is likely due to the depth of sequencing (~100 million reads/sample) and the use of isogenic CRISPR edited controls, which removes variability due to genetic background. Particularly striking changes were observed in genes involved in vesicle secretion and autophagy. The involvement of autophagy pathways in numerous neurodegenerative diseases has been well documented [65], and increased levels of C9orf72 may enhance CBD risk by disrupting autophagy pathways necessary for responding to stress. *C9orf72* has been shown to regulate autophagy initiation through the *ULK1* complex[67,75,207], and we observe increased expression of autophagy initiation genes in the intermediate repeat knockin NPCs. Interestingly, over expression of C9orf72 seems to increase autophagy initiation under normal conditions, but inhibits autophagy induction under nutrient stress. Several other studies have implicated *C9orf72* as regulator of the cellular stress response[71,73,76,77], and we observe dysregulation of genes involved in the ER stress response and nutrient biosynthesis pathways in intermediate repeat knockin NPCs. In particular, the gene *TMEM74* and its paralog *TMEM74B* were highly downregulated in *C9orf72* NPCs with intermediate repeats. *TMEM74* has been shown to be a potent activator of autophagy induced via nutrient starvation[214] and is thought to function via a unique pathway that does not require initiation via the canonical *ULK1* autophagy initiation complex[215]. This alternative initiation pathway may be affected by increased C9orf72 levels and could explain why we observe an autophagy defect in C9orf72 over expressing cells under

nutrient starvation. Future studies will need to further define the molecular mechanisms by which *C9orf72* overexpression alters specific autophagy pathways.

Overall, the findings presented here illustrate the complexity of *C9orf72* repeat expansion mutations, with the length of expansion conferring risk for different neurodegenerative diseases via opposing mechanisms. Notably, the proportion of CBD cases with intermediate C9orf72 repeats is still relatively low (~4%) indicating that the overall genetic risk associated with intermediate repeats is modest. Likewise, given that CBD is a rare disease, an otherwise normal individual harboring an intermediate repeat is unlikely to develop CBD. Finally, these results indicate that autophagy pathways are dysregulated by elevated C9orf72 levels and suggest a link between *C9orf72* expression and neurodegenerative disease.

## Figures




b

| | Total Alleles | ≥ 17 Repeats | ≤ 17 Repeats | % Intermediate | Odds Ratio | 95% CI | P value |
|---|---|---|---|---|---|---|---|
| Controls | 11,774 | 61 | 11,713 | 0.52% | 3.59 | 1.80 - 6.65 | 0.000238 |
| CBD | 708 | 13 | 695 | 1.84% | | | |

**Figure 3.1**: ***C9orf72* repeat sizing of CBD. a)** Example capillary electrophoresis traces of CBD cases measured from genomic DNA using a fluorescent PCR assay with genotypes of 2,2 (top), 2,18 (middle) or 9,20 (bottom). **b)** Summary statistics for repeat size screening of CBD cases (n = 708 alleles) and previously published controls (n = 11,774 alleles)[123]. Fisher's exact test: OR = 3.59, 95% CI = 1.80 – 6.65, p = 0.000238. **c)** Allele frequencies of repeat sizes in CBD cases and controls. The number of alleles in the CBD cohort for each size is depicted above the bars. **d)** Results of Fisher's exact test

between CBD cases and controls at each allele size cutoff. Dotted gray line represents $p < 0.05$ significance level.

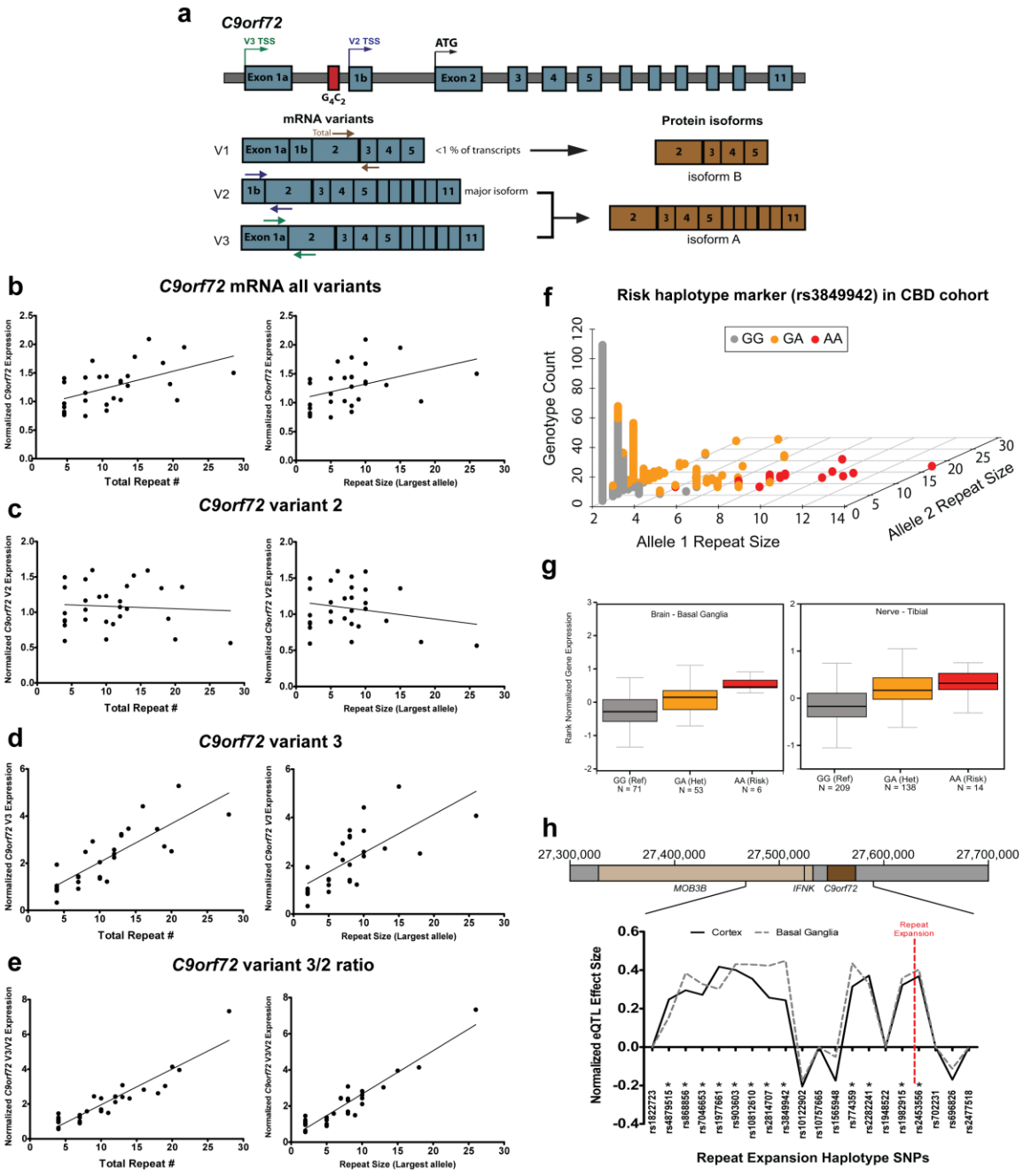



**Figure 3.2: Effect of intermediate repeats on *C9orf72* expression. a)** Diagram of the *C9orf72* locus with mRNA and protein isoforms. Transcription start sites (TSS) are labeled for each mRNA variant on the genomic DNA. Colored arrows on the mRNA depict primers used for RT-qPCR. **b)** Correlation between total *C9orf72* repeat size (left)

or largest allele (right) and *C9orf72* mRNA levels from all variants measured by RT-qPCR using RNA from CBD patient cerebellum (Total repeats: n = 29, $R^2$ = 0.276, p = 0.0034; Largest allele: n = 29, $R^2$ = 0.1613, p = 0.0308). Data is normalized to samples with 4 total repeats. **c)** Correlation between total *C9orf72* repeat size (left) or largest allele (right) and variant 2 mRNA levels (Total repeats: n = 29, $R^2$ = 0.00496, p = 0.717; Largest allele: n = 29, $R^2$ = 0.04432, p = 0.2730). **d)** Correlation between total *C9orf72* repeat size (left) or largest allele (right) and variant 3 mRNA levels (Total repeats: n = 29, $R^2$ = 0.667, p < 0.0001; Largest allele: n = 29, $R^2$ = 0.4874, p < 0.0001). **e)** Correlation between total *C9orf72* repeat size (left) or largest allele (right) and the variant 3 mRNA to variant 2 mRNA ratio (Total repeats: n = 29, $R^2$ = 0.849, p < 0.0001; Largest allele: n = 29, $R^2$ = 0.8869, p < 0.0001). **f)** Genotyping results for rs3849942 in the CBD patient genomic DNA cohort (n = 708 alleles). The A allele is associated with intermediate and full length repeat expansions. **g)** SNP eQTL data from the Gtex project showing the effect of rs3849942 on *C9orf72* expression in basal ganglia (left, p = 9.00E-07) and tibial nerve (right, p = 2.80E-16). For box plots, center line indicates median, boxes are upper and lower quartiles, error bars are standard error of the mean. Numbers of samples are indicated below each group. **h)** Normalized GTEx eQTL effect size from cortex and basal ganglia on *C9orf72* expression for all 20 SNPs in the consensus repeat expansion risk haplotype (from Mok et al. 2012[212]). * indicates SNPs with a significant effect on *C9orf72* expression. Red dotted line indicates the location of the *C9orf72* repeat expansion.

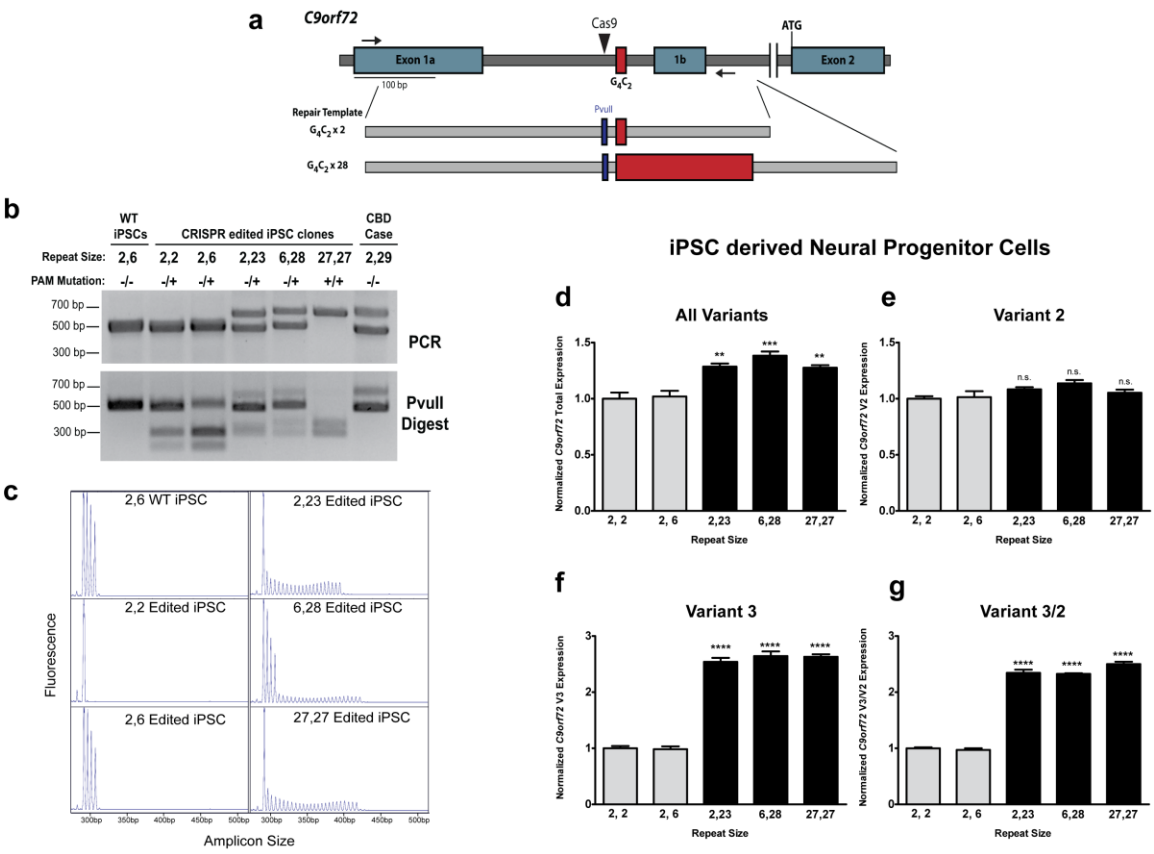



**Figure 3.3: CRISPR/cas9 knock-in of intermediate repeats into iPSCs. a)** Diagram of *C9orf72* genomic locus and double stranded repair templates used for knocking in intermediate repeats. Black triangle represents the CRISPR/cas9 cut site. Black arrows indicate the PCR primers used in (b). Red boxes indicate the $G_4C_2$ repeat. Blue boxes indicate the 2 bp substitution included in the repair templates to mutate the Cas9 PAM site and generate a PvuII restriction enzyme site. **b)** PCR (top) and PvuII digest (bottom) of CRISPR edited iPS cell lines using PCR primers spanning the CRISPR cas9 cut site and repeat expansion. Digestion of the PCR product indicates that HDR occurred. **c)** Capillary electrophoresis traces of repeat primed PCR from CRISPR edited iPSCs. **d-g)** RT-qPCR from the indicated variants of *C9orf72* from neural progenitor cells

differentiated from CRISPR edited iPSCs. Light bars correspond to low repeat lines while black bars correspond to intermediate repeat lines. N = 3 biological replicates per cell line; ANOVA (All variants: $F = 18.3$, $p = 0.0001$; Variant 2: $F= 2.9$, $p = 0.08$; Variant 3: $F = 216.0$, $p < 0.0001$; V2/V3: $F = 460.0$, $p < 0.0001$) followed by Bonferroni post hoc test between low and intermediate repeat cell lines. **$p < 0.01$, ***$p < 0.001$, ****$p < 0.0001$. Error bars are standard error of the mean.

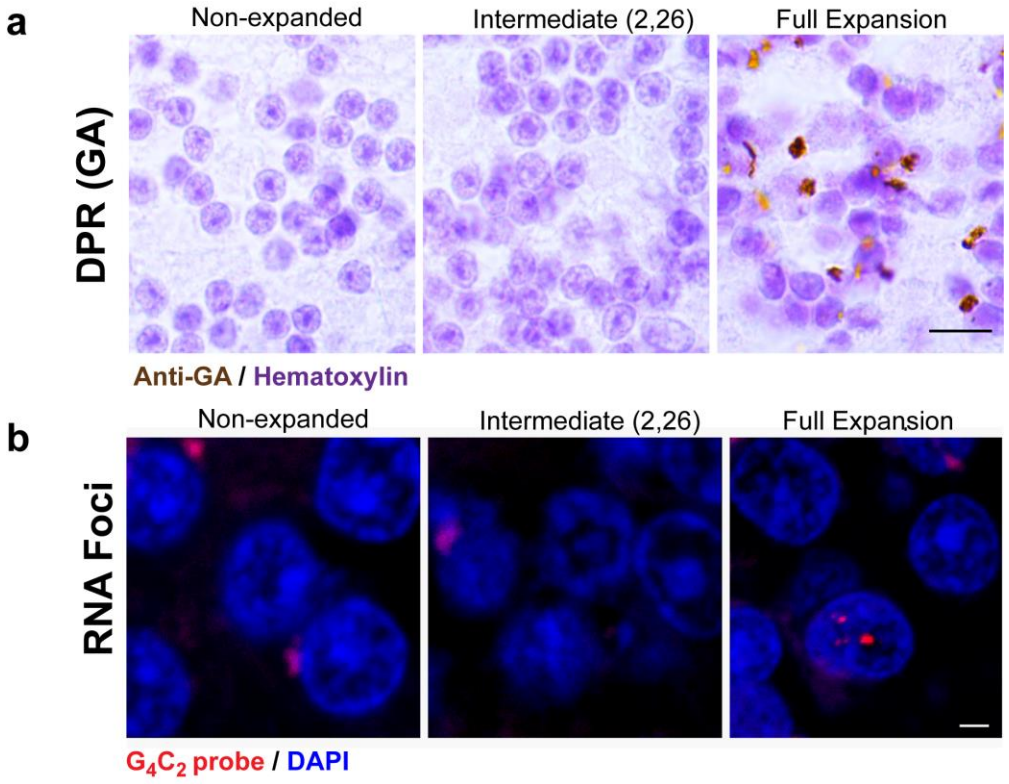



**Figure 3.4: Absence of repeat pathology in intermediate expansion cases. a)** Representative Immunohistochemistry images of non-expanded control (n=4 cases) or non-expanded CBD (n = 116 cases), intermediate expansion CBD (n = 9 cases), and full expansion FTD/ALS (n = 5 cases) post mortem cerebellar samples using anti-GA or p62 antibodies. Intermediate expansion carriers were also negative for GP/GR aggregates. Scale bar 10μm. **b)** Representative fluorescence in situ hybridization images of non-expanded control (n=4 cases) or non-expanded CBD (n = 12 cases), intermediate expansion CBD (n = 5 cases), and full expansion FTD/ALS (n = 5 cases) post mortem cerebellar samples using a probe complimentary to the $G_4C_2$ repeat. Scale bar 5μm.

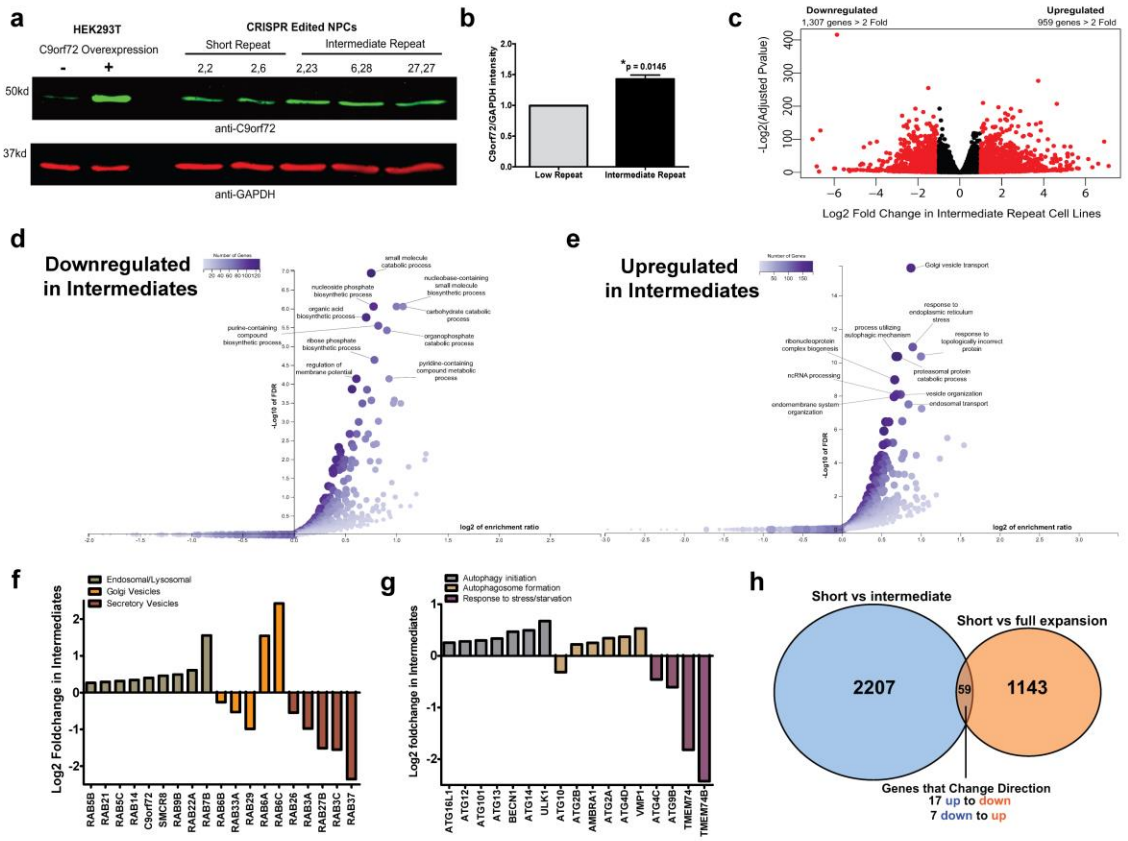



**Figure 3.5: Elevated C9orf72 protein levels and gene expression changes in intermediate repeat NPCs. a)** Western blot of protein expression from CRISPR edited NPCs. N = 2 low repeat and 3 intermediate repeat number cell lines. C9orf72 antibody was validated in HEK293T cells transfected with C9orf72 overexpression plasmid (Lanes 1-2). Representative blot of 2 technical replicates is shown. **b)** Quantification of C9orf72 protein levels from repeat edited NPCs. Protein expression was normalized to GAPDH as a loading control and the average of 2 technical replicates is reported. N = 2 low repeat and 3 intermediate repeat number cell lines. Two tailed t-test, p = 0.0145. Error bars are standard error of the mean. **c)** RNA sequencing analysis of 3 low repeat NPC lines (max repeat size 2 or 6) and 3 intermediate repeat (max repeat size 23, 27 or 28) NPC lines.

Differentially expressed genes with an adjusted p-value < 0.05 and greater than 2 fold change are depicted as red dots. **d-e)** Overrepresentation analysis of biological processes on differentially downregulated genes (d) or upregulated genes (e) in NPCs with intermediate repeats. Labels highlight the top 10 most significant biological processes. **f)** Changes in gene expression of genes involved in vesicular trafficking. All genes shown are significantly differentially expressed (adjusted $p < 0.05$) between low repeat and intermediate repeat NPCs. **g)** Changes in gene expression of genes involved in autophagy. All genes shown are significantly differentially expressed (adjusted $p < 0.05$) between low repeat and intermediate repeat NPCs. **h)** Comparison of overlap in differentially expressed genes between low repeat NPCs versus intermediate repeat NPCs (this study) or low repeat vs full expansion neurons from patient derived iPSCs from two previous studies [81,83]. Only genes with adjusted $p < 0.05$ and fold change > 2 were used in this comparison.

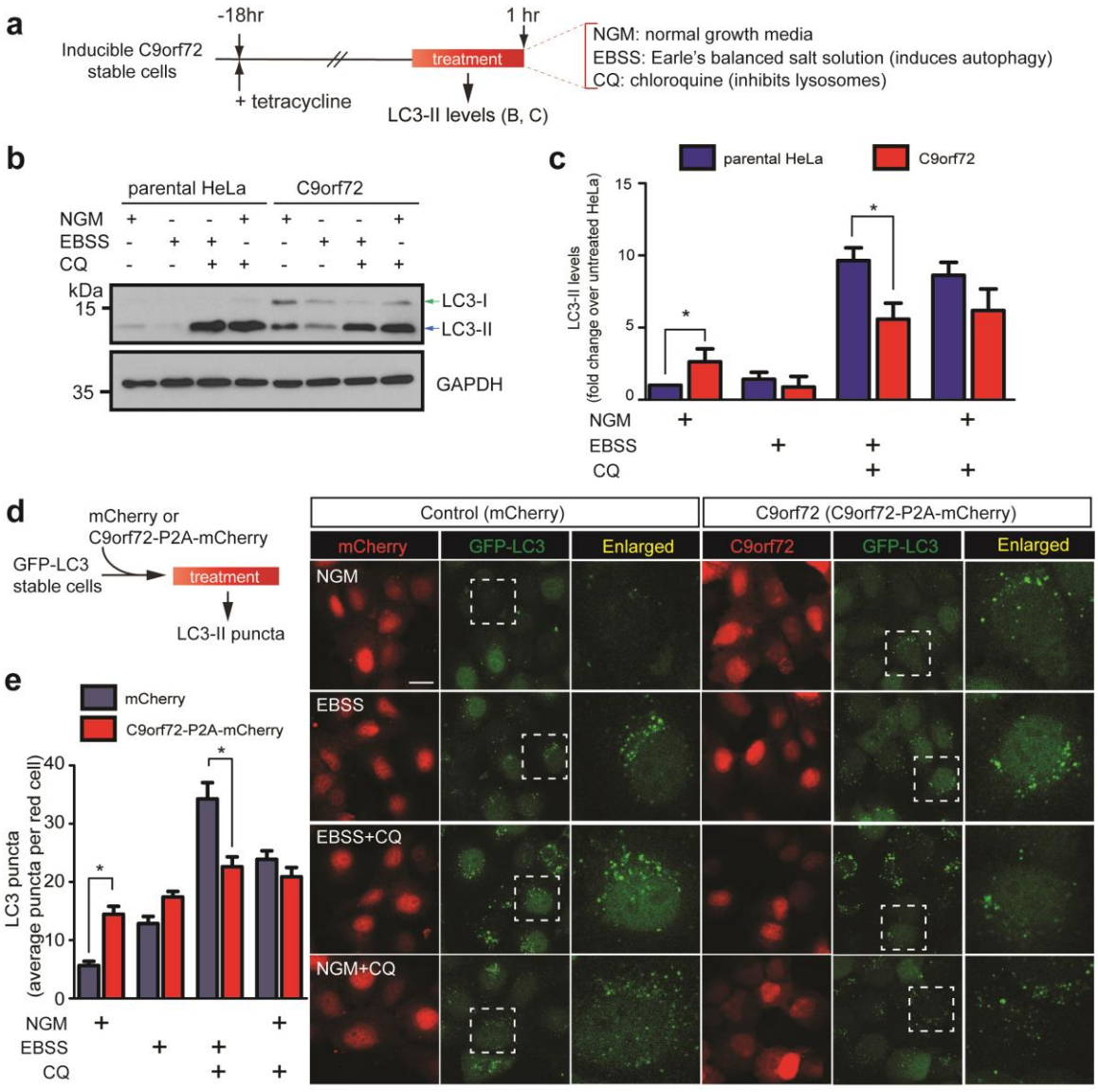



**Figure 3.6: Autophagy defects in inducible C9orf72 over-expressing cells**

**a)** Schematic of experimental procedure for LC3-II protein detection in C9orf72 HeLa stable cells. **b)** Representative western blot showing LC3-II expression levels of parental HeLa cells and C9orf72 cells first treated with treated with tetracycline to induce C9orf72 transgene expression then with EBSS and/or CQ for 1 hour. **c)** Bar chart corresponding to (b) showing pooled data of LC3-II protein levels expressed as fold change over

parental HeLa cells in normal growth media without CQ treatment. Data represents three independent experiments with * $p<0.05$, Student's t-test. **d)** Schematic of experimental procedure for the detection of LC3-II puncta using fluorescence microscopy in parental HeLa cells overexpressing GFP-LC3 and mCherry (control) or C9orf72-P2A-mCherry plasmids. The cells were treated with NGM, EBSS, EBSS+CQ or NGM+CQ for 2 hours and green puncta associated with GFP-LC3 in mCherry-overexpressing cells were analyzed using ImageJ particle count. At least 25 cells were analyzed per treatment group. Scale bar = 20 μm. **e)** Bar chart corresponds to the average LC3 puncta count in (d) with * $p < 0.05$, Student's t-test. Error bars represent standard error of the mean (S.E.M).

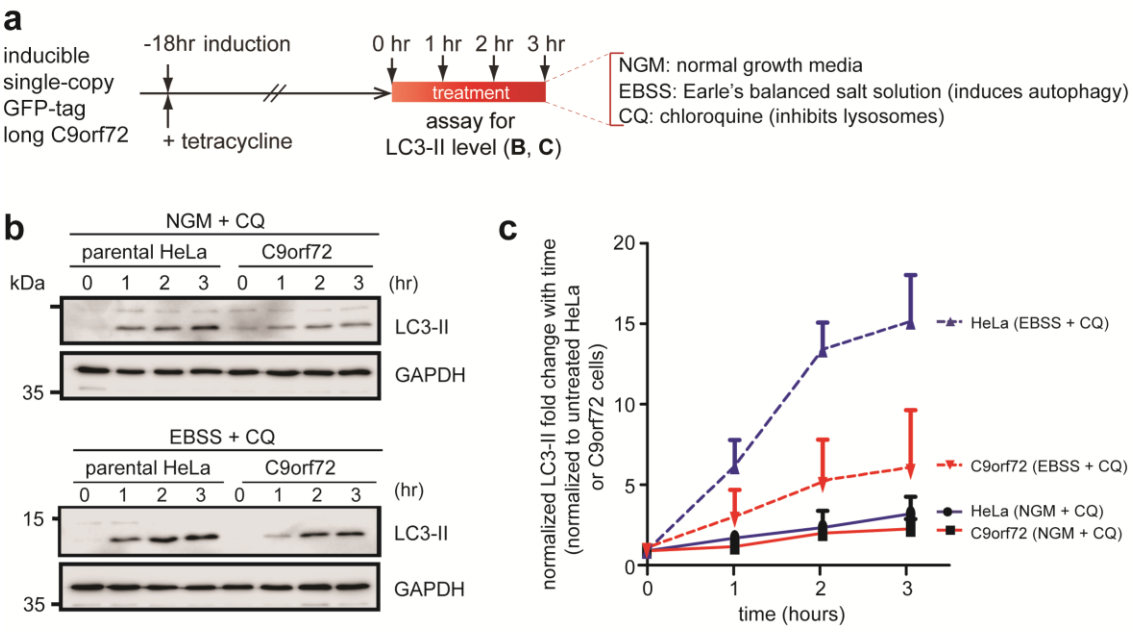
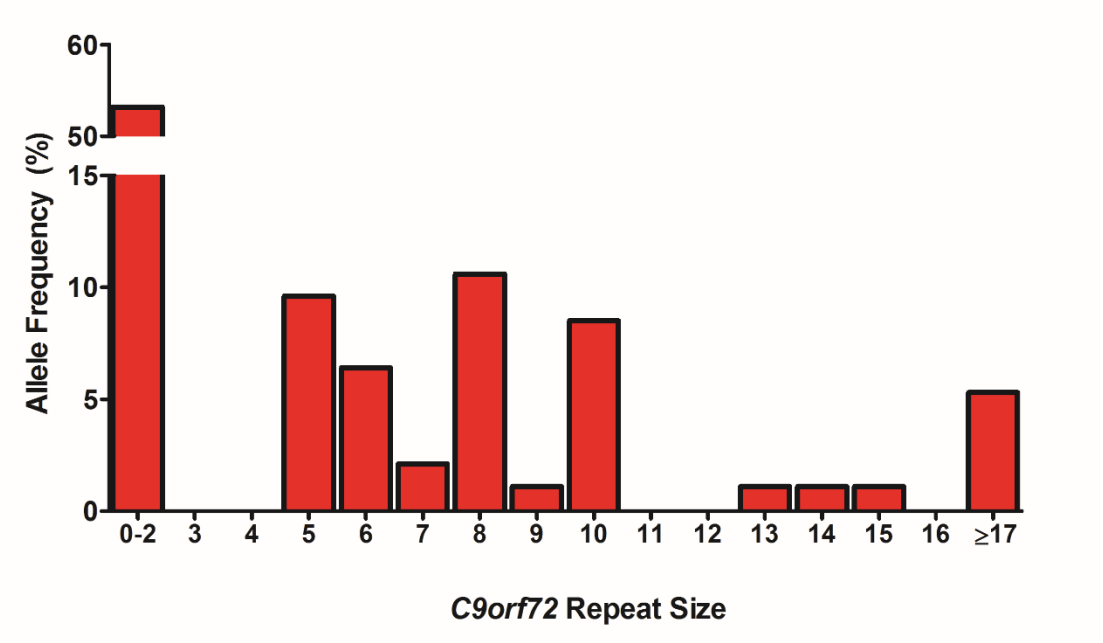
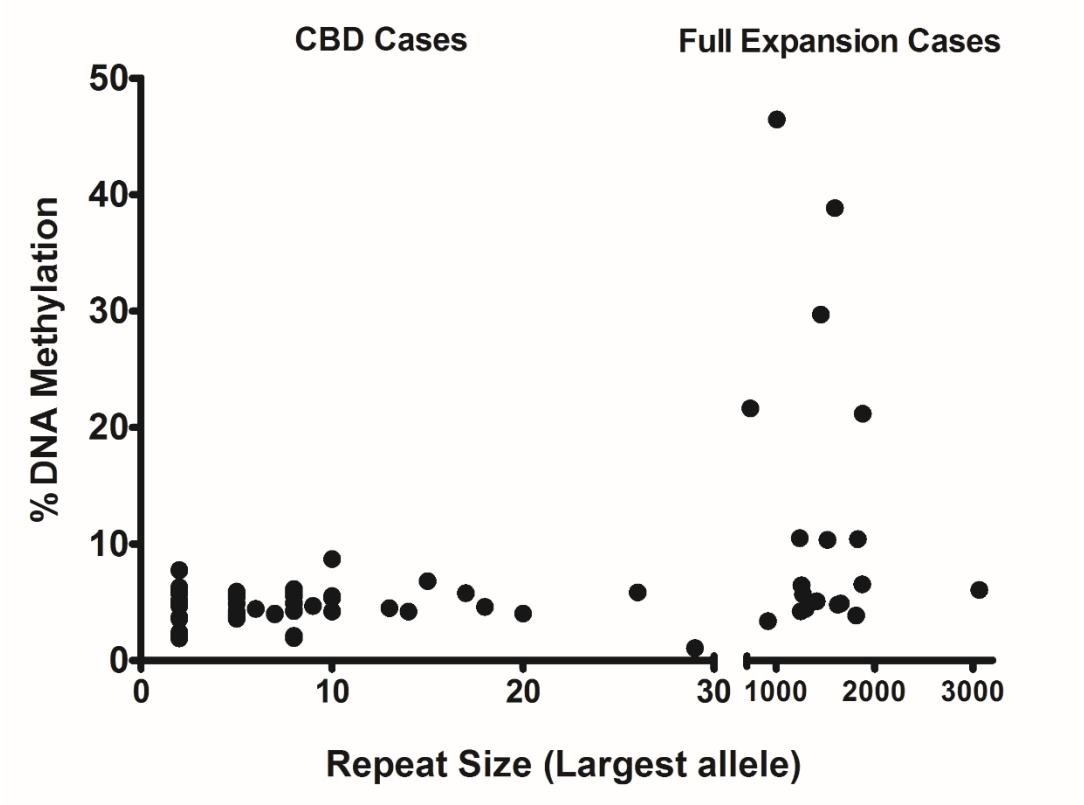



**Figure 3.7: Kinetics of starvation induced autophagy in C9orf72 over-expressing cells.** a) Schematic of time-course experiments on autophagy in C9orf72 HeLa stable cells. b) Representative western blot showing LC3-II expression levels of parental HeLa cells and C9orf72 cells first treated with tetracycline to induce C9orf72 transgene expression then with either normal growth media (NGM)+CQ or EBSS+CQ for 0, 1, 2, and 3 hours. c) Time course chart corresponding to (b) showing pooled data of LC3-II protein levels expressed as fold change over respective controls, i.e. parental HeLa cells or C9orf72 cells grown in normal growth media without CQ treatment. Data represents three independent experiments with error bars representing standard error of the mean (S.E.M).

## Supplemental Figures and Tables




**Supplemental Figure 3.1:** Repeat Expansion sizes in the CBD post mortem cerebellum cohort. Repeat sizes from CBD post mortem cerebellum detected using fluorescent PCR followed by capillary electrophoresis. N = 47 cases.

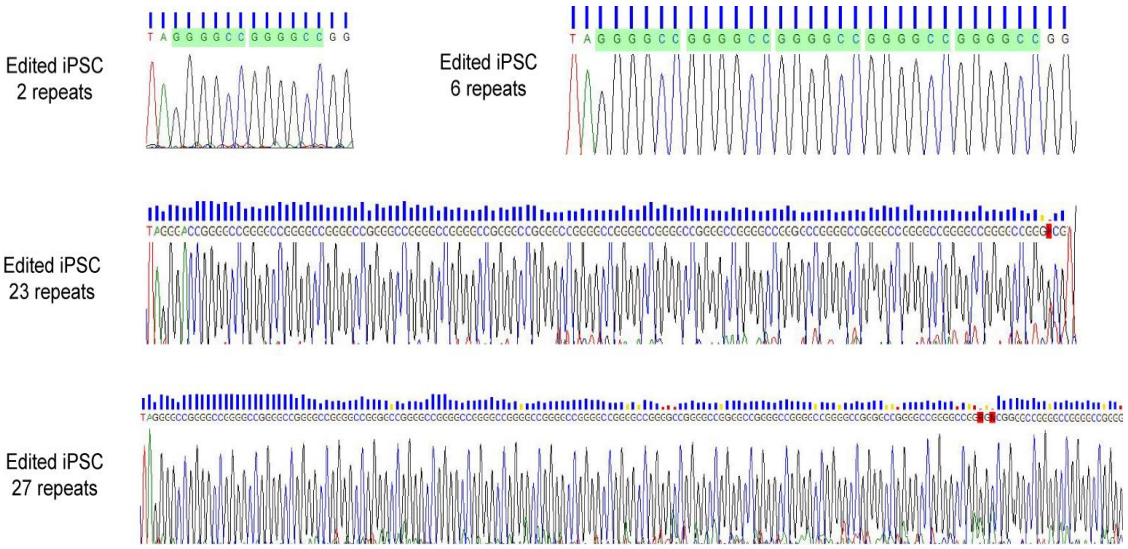



**Supplemental Figure 3.2:** DNA methylation analysis at the *C9orf72* promoter of CBD post mortem cerebellum cohort and comparison to full expansion cases. DNA methylation was measured using HhaI digest followed by qPCR on CBD cases (n = 47) and full expansion ALS/FTD cases (n= 20). Within the CBD case group, there is no significant association between repeat size and DNA methylation of *C9orf72* ($R^2$ = 0.01012, p = 0.5010).

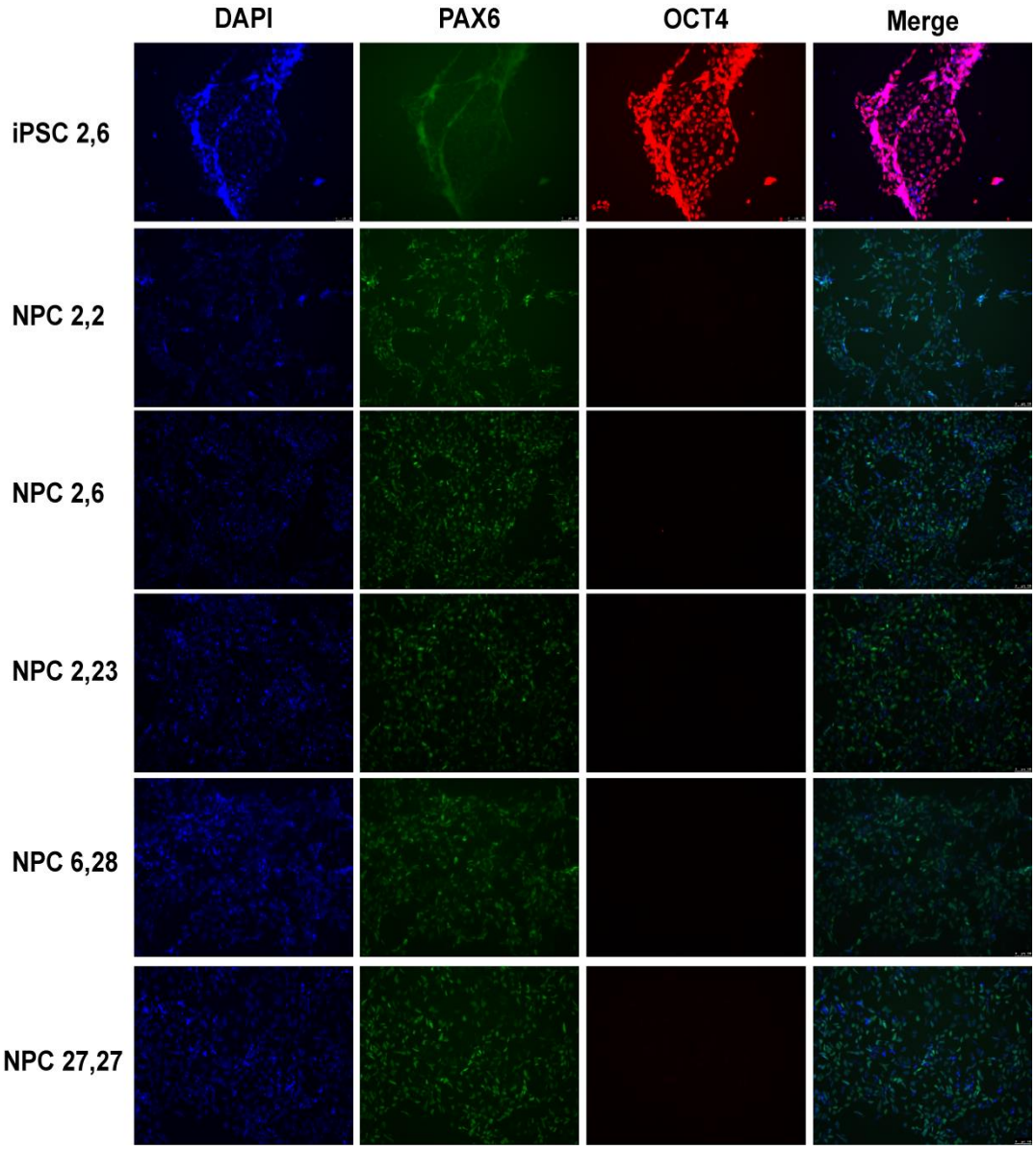
***C9orf72* Homologous Repair template (5' to 3')**

CCGCTAGGAAAGAGAGGTGCGTCAAACAGCGACAAGTTCCGCCCACGTAAAAGATGACGCTTGGTGTGTCA
GCCGTCCCTGCTGCCCGGTTGCTTCTCTTTTGGGGGCGGGGTCTAGCAAGAGCAGGTGTGGGTTTAGGAGG
TGTGTGTTTTTGTTTTTCCCACCCTCTCTCCCCACTACTTGCTCTCACAGTACTCGCTGAGGGTGAACAAG
AAAAGACCTGATAAAGATTAACCAGAAGAAAACAAGGAGGGAAACAACCGCAGCC**TGTAGCAAGCTCTGGA**
**ACTC**AGCTGTCGCGCGCTA***GGGGCCGGGGCC***GGGGCGTGGTCGGGGCGGGCCCGGGGGCGGGCCCGG
GGCGGGGCTGCGGTTGCGGTGCCTGCGCCCGCGGCGGCGGAGGCGCAGGCGGTGGCGAGTGGGTGAGTGAG
GAGGCGGCATCCTGGCGGGTGGCTGTTTGGGGTTCGGCTGCCGGGAAGAGGCGCGGGTAGAAGCGGGGGCT
CTCCTC

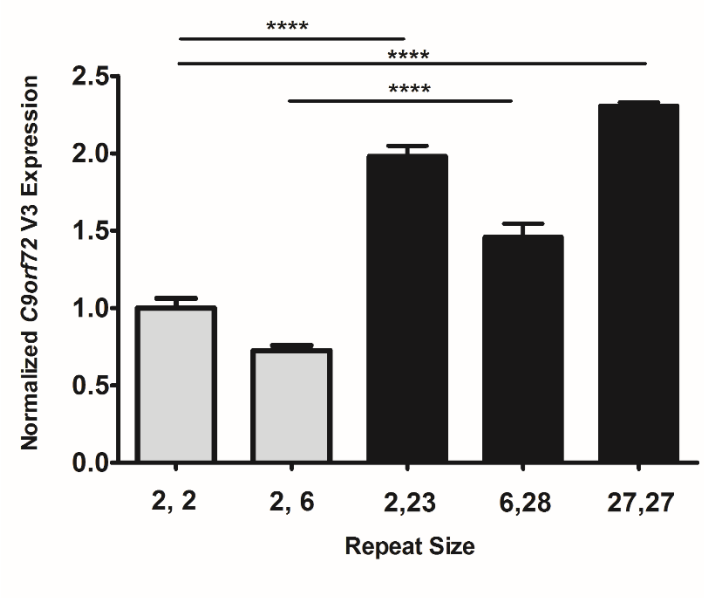
**Supplemental Figure 3.3:** Homologous template and gRNA sequence used to edit the endogenous *C9orf72* repeat expansion size. Homologous template containing either 2 or 28 $G_4C_2$ repeat units was cloned via PCR and inserted into P-gem Easy T vector. The 20bp CRISPR gRNA sequence is underlined and the 2 bp mutation which destroys the PAM site and introduces a PvuII restriction enzyme site is highlighted in red. The $G_4C_2$ expansion site is in bold (shown here is 2 repeat units).



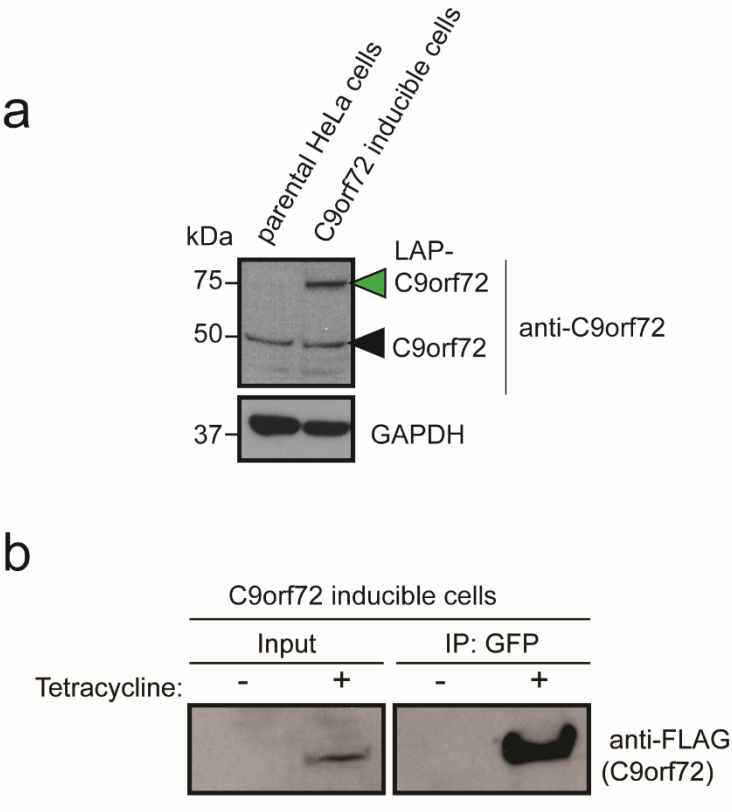

**Supplemental Figure 3.4:** Sanger sequencing of repeat edited iPSCs. Representative chromatograms of CRISPR edited iPSCs that were PCR cloned and Sanger sequenced. All CRIPSR edited cell lines included in this study were matched to have the same 8 bp deletion on the non-repeat edited allele, likely due to non-homologous end joining.




**Supplemental Figure 3.5:** Characterization of iPSC derived neural progenitor cells. Immunofluorescence images of day 12 post neural induction CRISPR edited NPCs stained with either pax6 (NPC marker) or Oct4 (iPSC marker). The number separated by a comma indicates repeat sizes of each allele in the NPC cell lines.




**Supplemental Figure 3.6:** *C9orf72* variant 3 mRNA levels from undifferentiated, repeat edited iPSCs. 1 way ANOVA followed by Bonferroni multiple comparison test. ****p < 0.0001. All intermediate repeat cell lines are significantly different than both low repeat controls




**Supplemental Figure 3.7:** Expression of C9orf72 transgene is comparable to the endogenous C9orf72 and is induced by tetracycline. a) Immunoblots of parental HeLa cells and the inducible isogenic C9orf72 cells used in this study. Both groups were treated with tetracycline for 18 hours. A polyclonal antibody (Santa Cruz Biotechnology, sc-138763) was used to detect the endogenous C9orf72 (black arrowhead) and the C9orf72 transgene (green arrowhead). The level of transgene is comparable to endogenous C9orf72 and the levels of endogenous C9orf72 are similar between parental HeLA cells and C9orf72 inducible cells. b) Immunoblots of C9orf72 transgene with or

without tetracycline induction. Inducible isogenic C9orf72 cells were cultured without (-) or with (+) tetracycline and subjected to GFP-immunoprecipitation to enrich for C9orf72 transgene. Input and GFP-immunoprecipitants were probed with FLAG antibody to detect C9orf72 transgene. C9orf72 transgene only be detected after tetracycline treatment.

**Table 3.S1: Primers and CRISPR gRNAs**

| Forward Primer Sequence | Reverse Primer Sequence | Assay |
|---|---|---|
| 6FAM-CAAGGAGGGAAACAACCGCAGCC | GCAGGCACCGCAACCGCAG | Sizing of *C9orf72* Repeats |
| 6FAM-AGTCGCTAGAGGCGAAAGC | TACGCATCCCAGTTTGAGACG | Repeat Expansion Detection |
| TACGCATCCCAGTTTGAGACGGGGGCCGGGGCCGGGGCCGGGG | | Repeat Expansion Detection |
| CCCACTTCATAGAGTGTGTGTTG | TTCCATTCTCTCTGTGCCTTC | Total *C9orf72* mRNA (all variants) |
| CGGTGGCGAGTGGATATCTC | TGGGCAAAGAGTCGACATCA | *C9orf72* variant 2 mRNA |
| GCAAGAGCAGGTGTGGGTTT | TGGGCAAAGAGTCGACATCA | *C9orf72* variant 3 mRNA |
| GAGGAGAGCCCCCGCTTCTAC | CCGCTAGGAAAGAGAGGTGCG | Template cloning and HDR analysis of repeat edited iPSCs |
| CACCGTGTAGCAAGCTCTGGAACTC | aaacGAGTTCCAGAGCTTGCTACAC | CRISPR gRNA sequences for repeat editing |
| CAGTGTGAAAATCATGCTTGAGAGA | TTTGTGCTTGGTAGGCAGTG | DNA methylation analysis |

**Table 3.S2: RNA sequencing of intermediate repeat NPCs raw read counts**

| Sample | Repeat Size | Total Reads |
|---|---|---|
| Short Repeat 1 | 2,6 | 109337673 |
| Short Repeat 2 | 2,2 | 108537553 |
| Short Repeat 3 | 2,6 | 118024548 |
| Intermediate Repeat 1 | 2,23 | 125527873 |
| Intermediate Repeat 2 | 6,28 | 121998215 |
| Intermediate Repeat 3 | 27,27 | 131039648 |

**Table 3.S3: Evaluation of *C9orf72* Risk haplotype marker rs3849942 A in CBD cases vs control populations**

| Cohort | rs3849942 A Risk Allele Frequency | Chi Square Test | Odds Ratio (95% CI) |
|---|---|---|---|
| CBD cases | 159 / 698 (22.78%) | | |
| Controls from Beck et al. 2013 | 2,597 / 10,400 (24.97%) | $X^2 = 1.568$, p = 0.2105 (n.s.) | 0.8863 (0.7385-1.064) |
| Gobal populations (dbSNP) | 35,048/162,186 (21.6%) | $X^2 = 0.4124$, p = 0.5207 (n.s.) | 1.064 (0.8913-1.271) |

**Table 3.S4: Repeat sizes of CBD cases with intermediate expansions**

| Case Number | Repeat Size Allele 1 | Repeat Size Allele 2 |
|---|---|---|
| 1 | 2 | 29 |
| 2 | 2 | 26 |
| 3 | 2 | 22 |
| 4 | 2 | 22 |
| 5 | 6 | 21 |
| 6 | 6 | 21 |
| 7 | 9 | 20 |
| 8 | 2 | 18 |
| 9 | 5 | 17 |
| 10 | 2 | 17 |
| 11 | 14 | 17 |
| 12 | 2 | 17 |
| 13 | 2 | 17 |

**Table 3.S5: Comparison of CBD cohort vs Theuns. et al 2014 controls at each allele cutoff**

| Repeat Size Cutoff | P-Value (Fisher Exact Test) | Odds Ratio |
|---|---|---|
| 3 | 0.876300 | 0.9853727 |
| 4 | 0.640200 | 0.9612321 |
| 5 | 0.638300 | 0.9609401 |
| 6 | 0.700700 | 0.9653403 |
| 7 | 0.963300 | 1.006105 |
| 8 | 0.499400 | 1.064384 |
| 9 | 0.209900 | 1.182729 |
| 10 | 0.026040* | 1.359480 |
| 11 | 0.108800 | 1.292946 |
| 12 | 0.032910* | 1.498095 |
| 13 | 0.073220 | 1.489279 |
| 14 | 0.128100 | 1.491891 |
| 15 | 0.061270 | 1.757263 |
| 16 | 0.004179** | 2.512414 |
| 17 | 0.0002376**** | 3.591016 |

# Chapter 4: Conclusions and Future Directions

This thesis explored several different pathways related to how repeat expansions in *C9orf72* contribute to neurodegenerative disease. First, we developed a novel method of targeted methylation in order to model DNA hypermethylation that is observed in ALS/FTD patients with the *C9orf72* repeat expansion. This work led to several important conclusions regarding the relationship between DNA damage and DNA methylation.

## DNA methylation is linked to DNA double strand break repair

It has long been thought that DNA methylation and DNA repair pathways were related due to the fact that chromosomal breaks correlate with changes in DNA methylation[131]. This idea was bolstered by evidence from reporter constructs that double strand breaks can directly induce DNA methylation[167–169]. Prior to this work, however, no study had shown that this process also occurs at endogenous genes. It's possible that methylation of exogenous DNA could be a defense response against foreign DNA such as viruses or transposable elements. The work presented here shows that methylation following DNA damage also occurs at endogenous genes and thus is a more generalized response than originally realized. This finding may help explain why DNA methylation at specific sites increases with age[132], as DNA damage due to oxidative stress is known to increase with age as well[230]. Future studies using multiplexed CRISPR gRNAs targeted to CpG islands could determine which loci genome-wide are prone to DNA methylation following DNA cleavage. These results would have important implications for cancer and other aging related diseases.

Another important finding relating to DNA methylation and DNA repair pathways is that methylation marks can be copied from an exogenous repair template into the genome. This was dependent on the process of homology directed repair (HDR), and the efficiency of methylation correlated with the length of the repair template. This process was also dependent on DNMT1, DNMT3a, DNMT3b and UHRF1. It was initially surprising that both the maintenance methyltransferase enzyme (DNMT1) and the *de novo* methyltransferase enzymes (DNMT3a/3b) were all involved in copying of methylation during HDR. However, it has been suggested that the methyltransferase enzymes have more overlapping functionality than previously realized[231]. Additionally, we did not determine the exact contribution of these enzymes to methylation via the different repair pathways. It is possible that the *de novo* methyltransferases are responsible for targeting methylation following NHEJ, whereas the DNMT1/UHRF1 complex is responsible for copying methylation from the repair template. Future experiments will need to clarify this point.

Additionally, it is unclear why only ~50% of CpGs are methylated in the endogenous gene following HDR. One possibility is that methylation is not copied during a subset of repair events. This is possible if some DNA is repaired via the non-canonical synthesis dependent strand annealing pathway (SDSA). SDSA differs from canonical HDR in that it only utilizes limited synthesis from one strand of the repair template rather than copying from both strands[166,232]. Since copying of methylation is likely to occur during DNA synthesis from the methylated template, repair via SDSA would restrict copying of methylation to small regions around the break site. Future studies should

explore the sub-pathways of HDR more closely in order to improve efficiency of methylation following HDR.

**Hypermethylation of the *C9orf72* promoter leads to gene silencing**

Another key finding from Chapter 2 is that DNA methylation of the *C9orf72* promoter leads directly to a gain in histone H3K9 trimethylation and a reduction of *C9orf72* mRNA levels in ALS patient derived iPS cells. Furthermore, the largest reduction in mRNA levels occurred from the transcription start site closest to the CpG island, suggesting that the proximity of DNA methylation to the transcription start site may influence transcriptional activity. This transcript (V3) also contains the repeat expanded RNA, and thus preferential reduction of this transcript may ameliorate the gain of function toxicity associated with the expanded RNA. Future studies using this newly established cellular model of DNA hypermethylation are needed to better understand how a reduction in *C9orf72* mRNA affects the balance between a loss of normal C9orf72 protein function and inhibition of the toxic gain of function pathway.

**HARDEN can be applied for targeted methylation of *APP***

A major unanswered question is whether HARDEN is a general method that can be applied to other genomic loci. It is known that genes with repeat expansions are prone to DNA methylation, so it is possible that *C9orf72* might be more amenable to targeted DNA methylation. To determine if HARDEN can be applied to another locus besides *C9orf72*, we targeted a CpG island in the promoter region of *APP*. *APP* encodes amyloid precursor protein, a key protein involved in Alzheimer's disease (AD) pathogenesis[233].

APP is cleaved by proteases to generate amyloid beta peptides, which are a major component of amyloid plaques found in AD patient brain[233]. Amyloid beta production is thought to play a central role in AD pathogenesis, as mutations in *APP* that increase amyloid beta levels are known to cause dominantly inherited early-onset AD[234]. Several studies have investigated the DNA methylation status of the CpG island upstream of *APP* but have come to differing conclusions as to whether *APP* is differentially methylated in AD patients[235–238]. Thus, it may be beneficial to generate cellular models of DNA methylation at this locus to determine the effects of DNA methylation on *APP* expression.

As a proof of principle, we applied the HARDEN system to *APP* by inducing a double strand break with Cas9 and providing a methylated repair template (Figure 4.1, A). The repair template also includes a 12 base pair mutation that allows for PCR amplification of only the DNA that has undergone HDR using the repair template (Figure 4.1, B). Bisulfite cloning of amplicons repaired with an unmethylated or methylated template was then carried out to assess the DNA methylation pattern of repaired DNA (Figure 4.1, C). This analysis revealed that DNA repaired with the methylated template is highly enriched for methylation, with the CpG sites closest to the cut site exhibiting levels of methylation exceeding 50% (Figure 4.1, C-D). In some clones, methylation was copied more than 500 base pairs downstream from the cut site (Figure 4.1, C-D). In sum, ~45% of amplicons (10/22) gain dense DNA methylation when repaired with a methylated template, compared to no amplicons (0/23) repaired with the unmethylated template (Figure 4.1, E). These results indicate that DNA methylation can be copied during HDR and confirm that HARDEN is a generalizable method that can be applied to

other genes besides *C9orf72*. Future work using these epigenetically edited cells could determine whether hypermethylation of this CpG island reduces *APP* expression and downstream amyloid beta production.

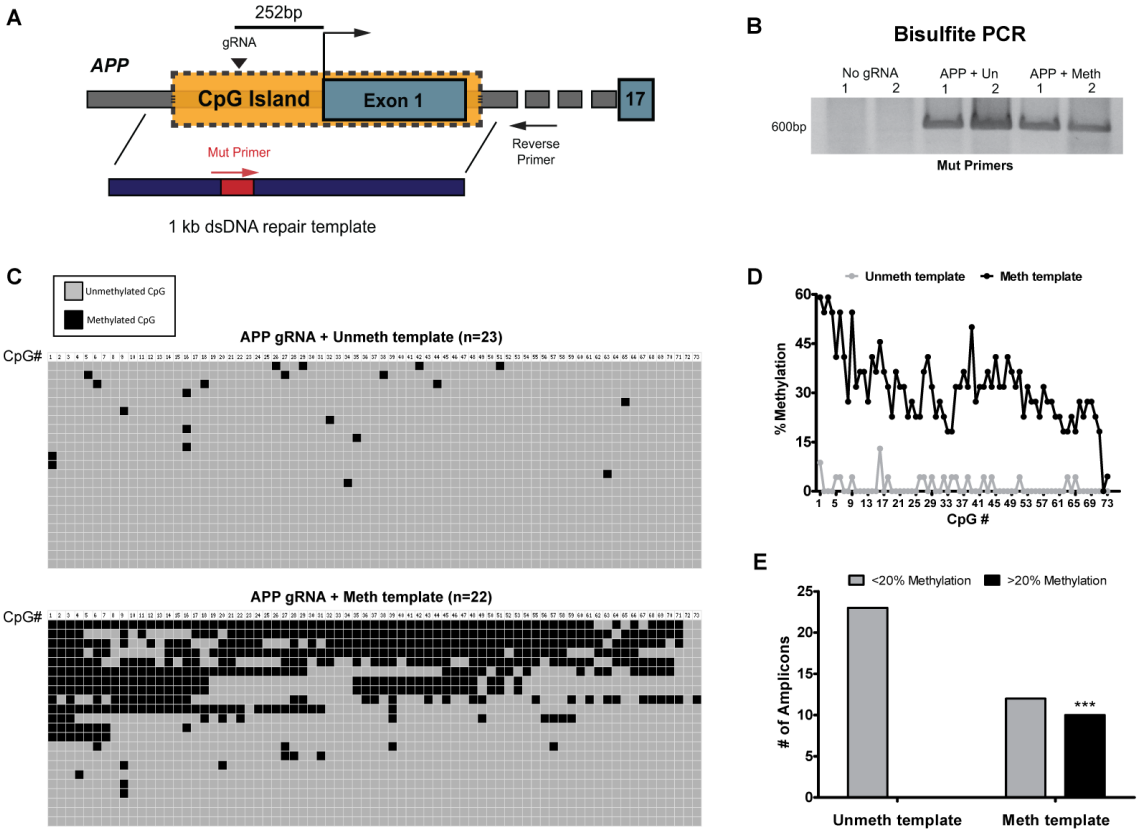



**Figure 4.1: Targeted DNA methylation of APP via HARDEN. A)** Diagram of *APP* gene locus showing CpG island, gRNA targeting site (black triangle) and bisulfite primers (Red and black arrows) used to amplify only DNA that incorporates a 12bp mutation (red box) from the template. **B)** Bisulfite PCR of HEK293T cells transfected with No gRNA plasmid, APP gRNA + unmethylated template or APP gRNA + methylated template. n = 2 biological replicates. Primer location is indicated in (A). **C)** Bisulfite sequencing of cloned amplicons from (B). Each column represents a CpG site in the ~600bp amplicon. Grey boxes indicate unmethylated cytosines whereas black boxes represent methylated cytosines. n = 23 sequences for APP gRNA+ unmethylated template group and n = 22 sequences for APP gRNA+ methylated template group. **D)** Quantification of mean methylation at each CpG site in the amplicon (73 total CpG sites) depicted in (C). n = 23 sequences from unmethylated template group and n = 22 sequences from methylated template group. **E)** Number of highly methylated amplicons (>20% methylation) from bisulfite sequencing depicted in (C). n = 23 sequences for unmethylated template group and n = 22 for methylated template group. 10 amplicons are highly methylated in the methylated template group vs 0 amplicons in the unmethylated template group. ***p = 0.0002; Fisher's exact test (two sided).

In chapter 3, we explore the genetic risk associated with *C9orf72* intermediate repeat expansions and how this relates to gene expression. Prior to this study, there was no clear indication of whether intermediate repeats in this gene were a risk factor for neurodegenerative disease. This study was not only the largest exploration of genetic risk for the neurodegenerative disease corticobasal degeneration (CBD), but it also revealed key insights into the molecular pathways that may be driving disease risk.

### *C9orf72* intermediate repeat expansions are a risk factor for corticobasal degeneration

Perhaps the most important finding of chapter 3 was that intermediate repeats in *C9orf72* are a risk factor for corticobasal degeneration. Prior to this work, there was conflicting evidence as to whether *C9orf72* was a risk factor for other neurodegenerative diseases such as Parkinson's disease[125,126]. We found that expansion carriers with intermediate repeats ranging from 17-30 units have more than 3-fold higher risk for CBD. Importantly, our study also suggests a new definition of what should be considered an "intermediate" size repeat expansion in *C9orf72*, as we observe increasing risk for CBD beginning at just 10 repeats. This work also highlights the importance of an autopsy confirmed cohort for genetic analysis of neurodegenerative diseases that have similar clinical characteristics. It remains important to note, however, that CBD is a rare disease and thus a given intermediate expansion carrier in the general population has relatively low risk for developing CBD. The rarity of CBD made it impossible to assemble a large enough replication cohort for this association. When more cases become available in the future, a replication study should be conducted.

**Intermediate length repeat expansions increase *C9orf72* expression**

Another significant finding in this thesis work is that intermediate length repeat expansions have the opposite effect on *C9orf72* expression as the full length expansions. It is well established that the full length expansions that cause ALS/FTD result in decreased *C9orf72* mRNA and protein levels[44,62,63,115]. Our work shows that *C9orf72* expression is actually increased in intermediate expansion carriers. This was true in both patient post mortem brain tissue and in knock-in isogeneic neural progenitor cells. Furthermore, it appears that expression increases with increasing total repeat number in an additive manner. This corresponds well with the observation that CBD risk also increases with increasing repeat number and suggests that increased *C9orf72* expression may be driving disease risk (Figure 4.2).

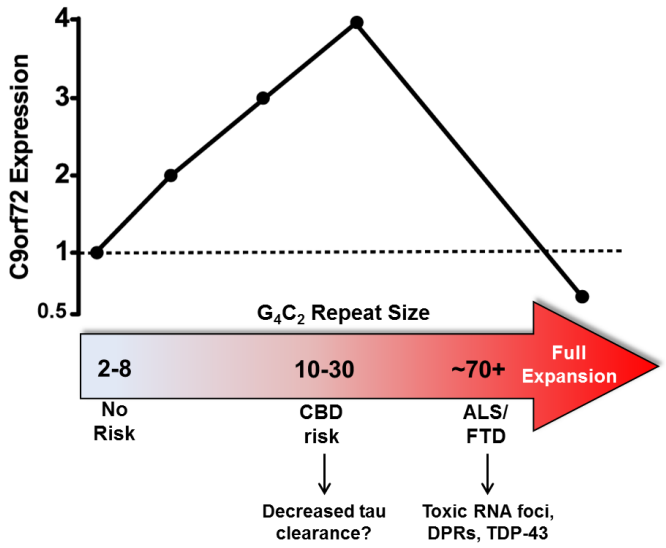



**Figure 4.2: Effect of *C9orf72* repeat size on expression and disease risk.** *C9orf72* expression increases linearly from low repeats (2-8 units) to intermediate repeats (10-30), but decreases in full expansion carriers. Risk for CBD increases with increasing repeat number (up until 30 repeats), whereas ALS/FTD develops in carriers with ~70 repeats or more. CBD risk may result from decreased tau clearance in intermediate repeat cells due to autophagy defects. ALS/FTD may be driven by toxic RNA and protein species in the context of low *C9orf72* expression.

This work reinforces previous findings in carriers with intermediate GCC repeats in the *FMR1*gene. In this gene, individuals with 55-200 repeats are at risk for the neurodegenerative disease fragile X−associated tremor/ataxia syndrome (FXTAS)[239]. FXTAS patients exhibit increased *FMR1* expression whereas larger expansions result in epigenetic silencing and the neurodevelopmental condition fragile X syndrome[50,240–242]. Future studies should delve into the mechanisms underlying how intermediate expansions cause upregulation of *C9orf72*. One possibility is that the GC rich nature of intermediate repeats promotes transcription from upstream transcription start sites (TSS) that are more active. Indeed, our data suggests that the increase in *C9orf72* mRNA levels is driven primarily through an increase in variant 3 abundance, which is normally the minor isoform and has a TSS upstream from the repeat expansion. A similar mechanism has been observed in FXTAS patients, where a shift to an upstream TSS results in higher levels of transcription[240,243]. Another possibility is that intermediate expansions lead to increased mRNA stability of repeat containing transcripts, although this is not observed in FXTAS[240]. Future work should clarify whether increased levels of C9orf72 are due to enhanced transcription, reduced mRNA degradation or increased translation.

**Increased *C9orf72* expression disrupts autophagy**

A final key finding from this thesis work is that increased expression of *C9orf72* results in an ablated autophagy response under nutrient stress conditions. We found that knock-in neuronal progenitor cells with intermediate repeats exhibit gene expression changes in key autophagy and endo-lysosomal pathways. The dysregulated genes also

included genes responsible for promoting autophagy under stress conditions. These results were reinforced by experiments in a *C9orf72* over-expression cell culture model. We found that *C9orf72* over-expression promotes autophagy under normal conditions but actually blocks autophagy induction during nutrient starvation. These finding suggest a model where proper levels of *C9orf72* are required for an appropriate stress response. While we only examined the response to nutrient deprivation, other studies have implicated C9orf72 in the response to other stressors including ER stress, oxidative stress and heat shock[71,76,77]. Thus, an ablated stress response in the context of aging or other insults may underlie the neurodegenerative disease risk associated with *C9orf72* intermediate expansions. Future studies will need to further define how differing levels of C9orf72 can both promote and inhibit autophagy under different conditions.

**C9orf72 expression and tau aggregation**

A final area for future investigation is to determine how increased expression of *C9orf72* relates to the disease pathology seen in CBD. The major pathology found in CBD patients is neuronal and astrocytic tau aggregates[244]. In preliminary experiments, we asked whether increasing *C9orf72* expression may alter tau aggregation. We utilized a previously developed[245] "tau biosensor" cell line that stably expresses a mutant tau repeat domain (P301S) fused to fluorescent reporters. When seeded with purified tau from post mortem brain, the tau reporters aggregate and generate a FRET signal that is detected by flow cytometry (Figure 4.3, A)[245]. We overexpressed C9orf72 protein at either low or high levels in these cells using constructs devoid of the intronic $G_4C_2$ repeat domain and transduced them with tau from CBD patient post-mortem brain (Figure 4.3, B). After

measurement of FRET via flow cytometry, we found that FRET intensity appeared to increase in a dose dependent manner with increasing amounts of C9orf72 over-expression (Figure 4.3, C) but was only significantly higher when using a codon optimized construct with large amounts of over-expression. These results suggest that high levels of C9orf72 over-expression may promote increased tau aggregation. Future studies in more physiological models of tau pathology will be needed to determine whether this effect is specific for C9orf72 and represents a real mechanism of disease risk.

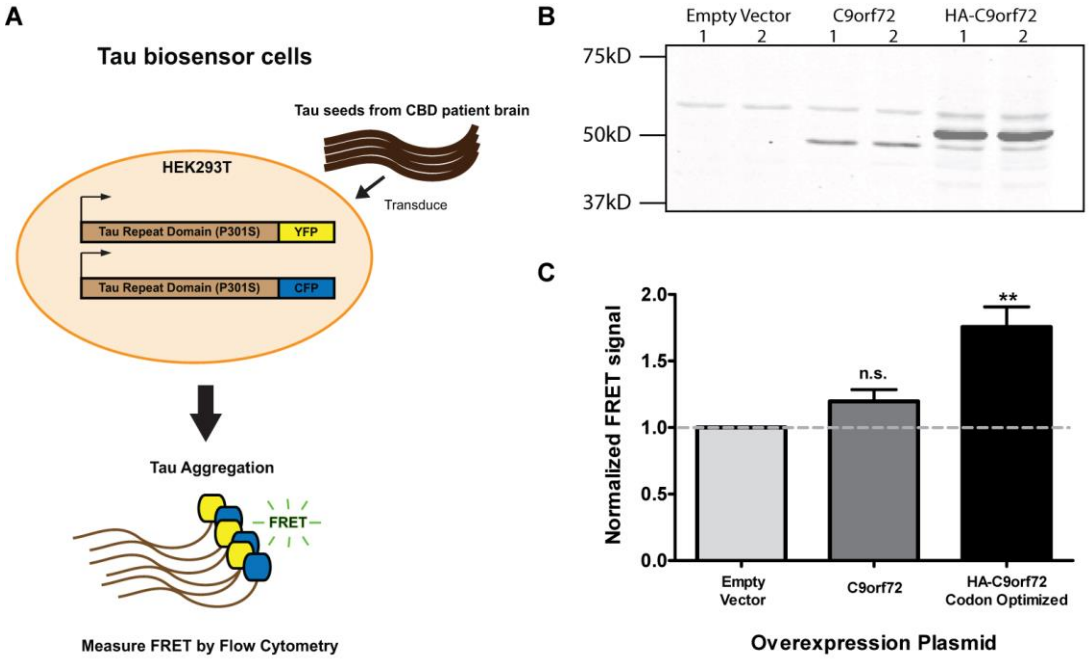



**Figure 4.3: Effect of C9orf72 over-expression on tau aggregation. A)** Overview of tau biosensor cells. HEK293T cells stably express mutant (P301S) tau repeat domain fused to a fluorescent protein. When seeded with tau from CBD patient brain, tau aggregation can be measured as a FRET signal by flow cytometry. **B)** Western blot of C9orf72 in tau biosensor cells that were transfected with either a WT or HA-tagged codon optimized C9orf72 encoding plasmid. **C)** Normalized FRET signal measured by flow cytometry of tau biosensor cells transfected with empty vector or C9orf72 overexpression constructs. Data was normalized to empty vector control. n = 4 experiments. 1 way ANOVA ($p = 0.0014$) followed by Bonferroni post hoc test. $^{**}p < 0.01$. In collaboration with Jessica Phan.

**Final Remarks**

The work presented here highlights the complexity of neurodegenerative diseases and the challenges faced by researchers in trying to define the underlying molecular mechanisms. Here, we show that mutations within a single gene that are known to cause ALS and FTD can also serve as a risk factor for a third neurodegenerative disease, CBD. Furthermore, the length of the repeat expansion and the epigenetic status of the locus serve as disease modifiers, thereby adding additional complexity. Despite this, our work demonstrates that common molecular pathways emerge in these diseases such as protein aggregation, cellular stress responses, autophagy and regulation of gene expression. We hope that the novel method of targeted DNA methylation presented here and the new insights into intermediate *C9orf72* repeat expansions may further the research community's efforts to find cures for these fatal diseases.

# Protocols

**CRISPR gRNA Cloning**

1. Design gRNAs and add BbsI overhangs:
    a. Add CACCG on 5' end of top oligo
    b. Add C on 3' end and CAAA on 5' end of bottom oligo
2. Resuspend oligos in milliq$H_20$ to 100uM
3. **Oligo Annealing/Phosphorylation Reaction:**

| Component | Amount (ul) | X 4 reactions |
|---|---|---|
| x100uM sgRNA oligo Top Strand | 1 | - |
| 100uM sgRNA oligo Bottom Strand | 1 | - |
| 10X NEB T4 DNA Ligation Buffer | 1 | 4 |
| NEB T4 PNK | 1 | 4 |
| Milliq $H_2O$ | 6 | 24 |
| **TOTAL:** | **10ul** | **8ul + 1ul F + 1ul R** |

Thermocycler program: CRISPR Anneal
Cycling Conditions: 37$^o$ for 30mins, 95$^o$ for 5mins, Ramp to 25 at 5$^o$/s (2.5% on Veriti thermocycler)

4. Dilute reaction above 1:200 in $H_2O$ (Note: diluted reaction can be stored at -20)
5. **Ligation Reaction:**

| Component | Amount (ul) | X 4 reactions |
|---|---|---|
| Cas9 Plasmid at 50ng/ul | 2 | 8 |
| Diluted guide oligos from step 4 | 2 | - |
| 10X Cutsmart Buffer | 2 | 8 |
| 10mM DTT | 1 | 4 |
| 10mM ATP | 1 | 4 |
| NEB BbsI (5U/ul) | 1.25 | 5 |
| NEB T4 DNA ligase | 1 | 4 |
| $H_20$ | 9.75 | 39 |
| **TOTAL:** | **20** | **18ul +2ul oligo** |

Thermocycler Program: CRISPR Ligation
Cycling Conditions: [37$^o$ for 5mins, 21$^o$ for 5mins ] x 6 cycles

6. **Plasmid Safe Reaction:**

| Component | Amount (ul) | X 4 reactions |
|---|---|---|
| Ligation reaction from above | 11 | - |
| 10X Plasmid Safe Buffer | 1.5 | 6 |
| 10mM Plasmid Safe ATP Solution | 1.5 | 6 |
| Plasmidsafe Exonuclease | 1 | 4 |
| **TOTAL:** | **15** | **4ul + 11ul ligation rxn** |

Thermocycler Program: Plasmidsafe
Cycling Conditions: 37$^o$ for 30mins, 70$^o$ for 30mins

7. Proceed to transformation or store at -20$^o$
8. Transformation:

a. Add 2ul of plasmidsafe reaction to 25ul of thawed NEB Turbo competent cells (NEB)
b. Incubate on ice 30mins
c. Heatshock at 42$^{o}$ for 30 sec
d. Incubate on ice 5 mins
e. Add 750ul of SOC and incubate in shaker at 37$^{o}$ for 1 hour
f. Dilute 1:3 and plate 60ul on LB amp plate

## DNA methylation detection via MSRE-qPCR

Reagents:

- HhaI (NEB R0139S)
- HaeIII (NEB R0108S)
- 2X Syber Green Master Mix (Roche 04913914001 or Thermo Fisher A25742)

-Genomic DNA must be column extracted for complete HhaI digest
-Dilute genomic DNA to 10ng/ul; Need at least 20ul per run.
-Always run 50% and 0% methylation controls
-HaeIII is needed to control for melting differences due to different repeat lengths

| # RXNs | x1 | x | 1 | x |
|---|---|---|---|---|
| **Mock** | | | **Digest** | |
| H2O | 7.9 | | 7.7 | |
| Cutsmart | 2 | | 2 | |
| HaeIII (20u/ul) | 0.1 | | 0.1 | |
| HhaI (20u/ul) | 0 | | 0.2 | |

10ul MM + 10ul DNA (100ng)

37$^{o}$ for 16hrs (overnight); 80$^{o}$ for 20mins to inactivate enzymes

**qPCR with Syber Green**

CpG1: chr9:27574058-27574271; Primers: 5PC9 F + Out5PC9 R
CpG2: chr9:27,573,299-27,573,429; Primers: 2CpG DIG F + 2CpG DIG R (always unmethylated)

| Reagent | Amount (ul) |
|---|---|
| 2X Syber Green Mix | 10 |
| 5uM Primer F | 1.2 |
| 5uM Primer R | 1.2 |
| milliH2O | 5.1 |
| Digested DNA | 2.5 |

17.5ul MM + 2.5ul DNA (12.5ng/reaction)

-Add MM to plate first in duplicates
-Add DNA and seal plate
-Run on Applied Biosystems Step One Plus qPCR machine (standard cycling for syber green reagents)

## Chromatin immunoprecipitation from cell lines

**Solutions Needed:**

- 1 M Tris-HCl pH 8 (6.05g for 50mL)
- 5 M NaCl (14.61g for 50mL)
- 0.5M EDTA (0.731g for 50mL)
- Lysis Buffer (for 200mL): 0.5% SDS (10mL of 10% SDS), 10mM EDTA (0.584g), 50mM Tris/HCl pH 8 (1.21g)
- ChiP dilution Buffer (for 200mL): 0.01% SDS (0.2mL of 10%), 1.1% TritonX (2.2mL), 1.2mM EDTA (0.07g), 16.7mM Tris/HCl pH 8 (0.40g), 167mM NaCl (1.95g)
- Low Salt Buffer (for 50mL): 0.1% SDS (0.5mL of 10%), 1% TritonX (0.5mL), 1.2mM EDTA, 16.7mM Tris/HCl pH 8, 150mM NaCl (0.44g)
- High Salt Buffer (for 50mL): 0.1% SDS (0.5mL of 10%), 1% TritonX (0.5mL), 1.2mM EDTA, 16.7mM Tris/HCl pH 8, 500mM NaCl (1.46g)
- LiCl Buffer (for 50mL): 0.25M LiCl (0.53g), 1% NP-40 (0.5mL), 1% sodium deoxycholate (0.5g), 1mM EDTA (0.1mL of 0.5M), 10mM Tris/HCl pH 8 (0.5mL of 1M)
- Elution Buffer (for 50mL): 1% SDS (0.5mL), 50mM Tris/HCl pH 8 (2.5mL of 1M), 1mM EDTA (0.1mL of 0.5M)
- BSA Blocking (for 50mL): 0.25g BSA, 1X PBS

### I) Lysate Preparation and Crosslinking

1. Cells grown to 90% confluency in 10cm dishes (~8 million cells per dish)
2. **Crosslinking**: Aspirate media and carefully add 10mL media with 1% formaldehyde; (270ul pf 37% formaldehyde stock per 10mL media)
3. Incubate for 10 mins at 37 degrees
4. **Quench**: Add 500µL 2.5 M Glycine directly to cells and incubate for 10 mins at 37 to quench; color should change to yellow
5. Add 10mL cold 1X PBS to wash cells; scrape cells with cell scrapper and add to cold 15mL conical
6. Pellet cells at 2,000 rpm for 5 mins at 4 degrees; aspirate liquid
7. Resuspend in 1.2mL cold ChIP lysis buffer (+ 1µL 1000X protease inhibitor)
8. Transfer 400µL of the cells + lysis buffer to bioruptor tubes (~$2x10^6$ cells per tube) for sonication
9. Freeze lysates at -80, or incubate 10 mins on ice and proceed directly to sonication
10. Sonicate with bioruptor on high, 30 on, 30 off for 30 cycles total; After every 10 cycles, spin for 30 sec max seed (20,000g), $4^o$ and vortex briefly.
11. After sonication, spin max speed for 10 mins, $4^o$; transfer and combine supernatant to new chilled tubes
12. Take 100µL lysate from each sample for DNA analysis and freeze the remaining lysate at -80.
13. Reverse cross-link the 100µL lysate by adding 2ul RNAse A (10mg/mL) and incubating at 37 for 30 mins; next add 1 ul proteinase K (20mg/mL) and incubate **ON** at $65^o$

14. Isolate DNA from the 100µL lysate using Qiagen PCR purfication kit; elute in 50µL EB buffer and nanodrop; Run 1-2ug of DNA on gel to determine sheer size; Optimal shear is 200-500bp.

**II) <u>Beads Preparation</u>**

Can use either agarose or Dyanabeads protein A

1. Vortex and pipette beads to resuspend
2. Add 30µL protein A beads/Dynabeads to 1.7mL tubes; one tube per IP or IgG control
3. Wash beads 2X with 500µL 0.5% BSA blocking solution
4. Resuspend in 50ul 0.5% BSA blocking solution
5. Add antibody to IP tubes (5-10µL for cell signaling antibodies; varies depending on antibody)
6. Add 2µl IgG (of 1.14 mg/mL stock or ~ 2µg)
7. Incubate on rotator, 4$^{o}$ for 5 hrs to overnight
8. Wash beads 2X with 250ul block solution and resuspend in 50µL BSA blocking solution
9. Dilute sonicated lysate with 5X ChiP dilution buffer; make enough for 20ug per IP/IgG reaction plus at least 10% extra
10. Remove 10% of the volume of diluted lysate (2 µg DNA) for the input (one per sample), store at 4$^{o}$
11. Add volume of lysate corresponding to 20ug DNA (same volume for IP and IgG samples) to the corresponding beads/antibody + BSA tubes
12. Incubate ON on rotor at 4$^{o}$

**III) <u>ChiP Washes and Reverse Cross-linking</u>**

1. Pellet beads and discard supernatant
2. Add 1mL Low-salt wash buffer, incubate on rotor for 10 mins at 4$^{o}$
3. Pellet beads and discard supernatant; Add 1mL of High Salt buffer, incubate on rotor, 10 min, 4$^{o}$
4. Pellet beads and discard supernatant; Add 1mL of LiCl buffer, incubate on rotor, 10 min, 4$^{o}$
5. Pellet beads and discard supernatant; Add 1mL of 1X TE buffer, incubate on rotor, 10 min, 4$^{o}$
6. Spin briefly on microcentrifuge and place back on magnet to remove final bit of liquid
7. Elute samples from beads by adding 200µL Elution buffer, incubate at 65$^{o}$ for 20 min with shaking at 500rpm; flick tubes every 5 mins
8. Pellet beads and collect supernatant in new tubes
9. *To inputs:* add elution buffer to total 200ul
10. Add 8ul of 10mg/ml RNAse A to each sample and incubate for 1 hour at 37$^{o}$
11. Add 3µl Proteinase K (20mg/mL), 8 µL 1M Tris-HCl, 8 µL 5M NaCl and 8 µL 0.25M EDTA to all samples

12. Incubate ON at 65$^{o}$
13. Isolate DNA using Qiagen PCR purification kit; elute 2X with 50µL EB buffer
14. Perform q-PCR (10µl 2X SyberGreen, 1.2µL each primer, 4µL DNA, 3.6µL $H_2O$ per reaction)

## Feeder Free iPSC culture

**Coating Plates with Matrigel**

- Matrigel is aliquotted in 150ul and 300ul at -20 C
- Thaw Matrigel for 30-45 mins at **4 C**; Do not leave at RT
- Add 150ul to a maximum of 12mL of **cold** DMEM/F12 media or 300ul to a max of 24mL media
    - Can make more concentrated if need fewer wells
- Use **cold** pipet to transfer 1mL of media+matrigel to each well on a 6 well plate
- Swirl plates to make sure wells are fully coated and incubate at RT for 45-60 mins

**Making mTeSR 1 Media**

- Combine 400mL Base media with 100mL Supplement to make complete mTeSR 1 media
- Aliqout 25-50mL to keep at 4 C and wrap in foil to protect from light
    - Can be used for 2 weeks
- Aliqout rest of media and freeze at -20 C
    - Stable for 6 months

**Thawing Cells**

- Make wash media: Mix 12.5 mL of DMEM/F12 media with 12.5 mL mTesR1 media
- Let cells thaw in water bath
- Thaw ROCK inhibitor
- Add cells to 25mL of wash media
- Spin at 1,2000 rpm for 3 mins (Program 1 -> Enter -> Start)
- Vacuum off media from matrigel coated plates
- Check for cell pellet and vacuum off media
- Resuspend cells in 2mL mTesR1 media per well
- Add 1ul ROCK inhibitor per 5mL media and mix by pipetting
- Plate 2mL of cells per well
- Mix plates before putting in incubator

## Differentiation of iPSCs to NPCs

### Necessary reagents:

| Reagent | Supplier | Cat# |
|---|---|---|
| 2-Mercaptoethanol | Fisher | 21985-023 |
| B27 Supplement | Cell Center (Invitrogen) | 17504044 |
| DMEM/F12 | Cell Center (Invitrogen) | 11320033 |
| Glutamax Supplement | Cell Center (Invitrogen) | 35050061 |
| Human FGF2 (Basic) | PeproTech | 100-18B |
| N2 supplement | Cell Center (Invitrogen) | 17502048 |
| Neurobasal media | Cell Center(Invitrogen) | 21103049 |
| Dorsomorphin | Sigma | P5499-5MG |
| MEM non-essential amino acids | Cell Center (Invitrogen) | 11140050 |
| L-Glutamine | Cell Center (Invitrogen) | 25030-024 |
| Insulin (human) | Sigma (cell center) | I9278-5ML |
| Pencillin/Streptomyocin | Cell Center (Invitrogen) | 15140122 |
| SB431542 | Fisher (Tocris) | 161410 |
| Matrigel | Corning | 354277 |
| ROCK inhibitor Y27632 | ATCC | ACS-3030 |
| Accutase | StemCell Technologies | 07920 |
| mTeSR1 | StemCell Technologies | 05850 |

### Media Recipes:

| N-2 Media | | |
|---|---|---|
| **Component** | **Stock Con.** | **Final Con.** |
| DMEM/F12 | 1X | 1X |
| GlutaMAX | 100X | 1X |
| N2 Supplement | 100X | 1X |
| Insulin | 11.2mg/mL | 5ug/mL |
| L-Glutamine | 200mM | 1mM |
| non-essential amino acids | 10mM | 0.1mM |
| 2-Mercaptoethanol | 55mM | 0.1mM |
| Pen/Strept | 10,000mg/mL | 50mg/mL |

| B-27 Media |
|---|

| Component | Stock Con. | Final Con. |
|---|---|---|
| Neurobasal media | 1X | 1X |
| B27 Supplement | 50x | 1X |
| L-Glutamine | 200mM | 1mM |
| Pen/Strep | 10,000mg/mL | 50mg/mL |

| | Neural Maintenance Media (1:1 mix of N-2 and B-27 media) | |
|---|---|---|
| **Component** | **Stock Con.** | **Final Con.** |
| N-2 media | 1X | 1X |
| B-27 Media | 1X | 1x |

| | Neural Induction Media - make fresh, lasts 5 days | |
|---|---|---|
| **Component** | **Stock Con.** | **Final Con.** |
| Neural maintenance Media | 1X | 1X |
| SB431542 | 100mM | 10uM |
| Dorsomorphin | 10mM | 1uM |

| | Neural Freezing media - make fresh | |
|---|---|---|
| **Component** | **Stock Con.** | **Final Con.** |
| Neural Maintenance Media | 1X | 1X |
| DMSO | 100% | 10% |
| FGF-2 | 50ug/mL | 0.02ug/mL |

**Protocol:**

Day 0: Plating of iPSCs for Neural Induction

- Grow iPSCs to confluency in 10cm dishes with Mtser1
- Remove Mtser1 and add 4mL Accutase per 10 cm dish
- Incubate at 37° for 6-8 mins until cell lift off plate
- Pipette up and down with a P1000 to break up clumps of cells, then add 4mL DMEM
- Transfer to 15mL conical and pellet cells at 1,200 rpm for 3 mins
- Resuspend cells in 8mL of Mtser + Rocki (10uM final con.)
- Remove 50ul for counting and add to 50ul trypan blue; count cells
- Seed cells at $2x10^6$ cells/well in a 6 well plate coated with matrigel (**Note: this is dependent on cell line; slowly growing lines may require higher density seeding; cells should be a confluent monolayer the next day)**

Day 1 – 6: Feed cells

- Cells should be 100% confluent before beginning induction

- Prepare Neural Induction Media by adding SB431542 (10uM final con.) and Dorsomorphin (1uM final con.); Induction media should be prepared fresh every 5 days
- Remove media, wash with 1mL of DMEM to remove dead cells
- Add 2mL Neural Induction Media

Day 5: Passage #1 (1:2 split)

- By day 5, the cells should be extremely confluent and may even start to lift off the plate
- Remove media and wash with 1 mL DMEM
- Add 1 mL accutase and incubate at 37° for 6-8 mins or until cells detach
- Add 1 mL DMEM and break up cell clumps with P1000
- Pellet cells at 1,200 rpm for 3 mins
- Resuspend in Neural Induction Media and seed into new matrigel coated 6 well dish; split cells 1:2, which will be ~$1x10^6$ cells/well

Day 6-7: Feed cells

- Feed cells with Neural induction media
- Cells should now undergo a morphological change to an epithelial like appearance (e.g. polygonal and large rather than the small and round iPSCs)

Day 8: Passage #2 (1:3 or 1:2 split depending on cell line)

- Cells should be confluent again by day 8
- Repeat passaging as on day 5, but this time split 1:3 if they are a fast growing line or 1:2 if they grow more slowly
- Wells can be pelleted for RNA extraction or used for seeding coverslips for IHC; Oct4 expression should be very low and Pax6 should be high in day 8 NPCs

Day 9-11: Feed cells

- Continue feeding with Neural Induction Media each day
- Cells should now start to form neural rosettes structures, although they may not be well defined

Day 12: Freezing of NPCs and switch to Neural Maintenance media

- NPCs can now be frozen down and used for future expansion to neurons
- Remove media and add 1mL accutase per well for 6-8 mins at 37°
- Add 1 mL media, pipette to break up clumps and transfer to conical tube
- Spin 1,200 rpm for 3 mins; remove media
- Resuspend in Neural Freezing media with freshly added FGF-2 and DMSO
- Freeze 1 mL per cryovial in Mr. Freezy at -80; transfer to liquid nitrogen the next day

Day 13-20: NPC Expansion – SWITCH to Laminin Matrix

- After day 12, add 20ng/mL FGF-2 to the media for the next 4 days to promote cell division of NPCs
- After FGF-2 treatment, media can be changed every other day rather than every day

# References


1. Lomen-Hoerth C, Anderson T, Miller B. The overlap of amyotrophic lateral sclerosis and frontotemporal dementia. *Neurology*. 2002;59(7):1077 LP - 1079. doi:10.1212/WNL.59.7.1077

2. Karch CM, Wen N, Fan CC, et al. Selective genetic overlap between amyotrophic lateral sclerosis and diseases of the frontotemporal dementia spectrum. *JAMA Neurol*. 2018;75(7):860-875. doi:10.1001/jamaneurol.2018.0372

3. van Es MA, Hardiman O, Chio A, et al. Amyotrophic lateral sclerosis. *Lancet*. 2017;390(10107):2084-2098. doi:10.1016/S0140-6736(17)31287-4

4. Rademakers R, Neumann M, MacKenzie IR. Advances in understanding the molecular basis of frontotemporal dementia. *Nat Rev Neurol*. 2012;8(8):423-434. doi:10.1038/nrneurol.2012.117

5. Kaivorinne A-L, Bode MK, Paavola L, et al. Clinical Characteristics of C9ORF72-Linked Frontotemporal Lobar Degeneration. *Dement Geriatr Cogn Dis Extra*. 2013;3(1):251-262. doi:10.1159/000351859

6. Ajroud-Driss S, Siddique T. Sporadic and hereditary amyotrophic lateral sclerosis (ALS). *Biochim Biophys Acta - Mol Basis Dis*. 2015;1852(4):679-684. doi:10.1016/j.bbadis.2014.08.010

7. Domingo A, Klein C. Genetics of Parkinson disease. *Handb Clin Neurol*. 2018;147(December 2017):211-227. doi:10.1016/B978-0-444-63233-3.00014-2

8. Bertram L, Lill CM, Tanzi RE. The genetics of alzheimer disease: Back to the future. *Neuron*. 2010;68(2):270-281. doi:10.1016/j.neuron.2010.10.013

9. Renton AE, Chiò A, Traynor BJ. State of play in amyotrophic lateral sclerosis genetics. *Nat Neurosci*. 2014;17(1):17-23. doi:10.1038/nn.3584

10. Van Es MA, Veldink JH, Saris CGJ, et al. Genome-wide association study identifies 19p13.3 (UNC13A) and 9p21.2 as susceptibility loci for sporadic amyotrophic lateral sclerosis. *Nat Genet*. 2009;41(10):1083-1087. doi:10.1038/ng.442

11. Van Rheenen W, Shatunov A, Dekker AM, et al. Genome-wide association analyses identify new risk variants and the genetic architecture of amyotrophic lateral sclerosis. *Nat Genet*. 2016;48(9):1043-1048. doi:10.1038/ng.3622

12. Nicolas A, Kenna KP, Renton AE, et al. Genome-wide Analyses Identify KIF5A as a Novel ALS Gene. *Neuron*. 2018;97(6):1268-1283.e6. doi:10.1016/J.NEURON.2018.02.027

13. Fogh I, Ratti A, Gellera C, et al. A genome-wide association meta-analysis identifies a novel locus at 17q11.2 associated with sporadic amyotrophic lateral sclerosis. *Hum Mol Genet*. 2014;23(8):2220-2231. doi:10.1093/hmg/ddt587

14. Osterloh JM, Rooney TM, Fox AN, et al. dSarm/Sarm1 is required for activation of an injury-induced axon death pathway. *Science (80- )*. 2012;337(6093):481-484. doi:10.1126/science.1223899

15. Fang X, Lin H, Wang X, Zuo Q, Qin J, Zhang P. The NEK1 interactor, C21ORF2, is required for efficient DNA damage repair. *Acta Biochim Biophys Sin (Shanghai)*.

2015;47(10):834-841. doi:10.1093/abbs/gmv076

16. Wang Z, Iida A, Miyake N, et al. Axial spondylometaphyseal dysplasia is caused by C21orf2 mutations. *PLoS One*. 2016;11(3):1-16. doi:10.1371/journal.pone.0150555

17. Nakajima K, Yin X, Takei Y, Seog DH, Homma N, Hirokawa N. Molecular Motor KIF5A Is Essential for GABAA Receptor Transport, and KIF5A Deletion Causes Epilepsy. *Neuron*. 2012;76(5):945-961. doi:10.1016/j.neuron.2012.10.012

18. Shatunov A, Mok K, Newhouse S, et al. Chromosome 9p21 in sporadic amyotrophic lateral sclerosis in the UK and seven other countries: A genome-wide association study. *Lancet Neurol*. 2010;9(10):986-994. doi:10.1016/S1474-4422(10)70197-6

19. Van Deerlin VM, Sleiman PMA, Martinez-Lage M, et al. Common variants at 7p21 are associated with frontotemporal lobar degeneration with TDP-43 inclusions. *Nat Genet*. 2010;42(3):234-239. doi:10.1038/ng.536

20. Ferrari R, Grassi M, Salvi E, et al. A genome-wide screening and SNPs-to-genes approach to identify novel genetic risk factors associated with frontotemporal dementia. *Neurobiol Aging*. 2015;36(10):2904.e13-2904.e26. doi:10.1016/j.neurobiolaging.2015.06.005

21. Ferrari R, Hernandez DG, Nalls MA, et al. Frontotemporal dementia and its subtypes: A genome-wide association study. *Lancet Neurol*. 2014;13(7):686-699. doi:10.1016/S1474-4422(14)70065-1

22. Raffaele F, Claudia M, John H. Genetics and molecular mechanisms of frontotemporal lobar degeneration: an update and future avenues. *Neurobiol Aging*. 2019;78:98-110. doi:10.1016/j.neurobiolaging.2019.02.006

23. Van Der Zee J, Van Langenhove T, Kleinberger G, et al. TMEM106B is associated with frontotemporal lobar degeneration in a clinically diagnosed patient cohort. *Brain*. 2011;134(3):808-815. doi:10.1093/brain/awr007

24. Brady OA, Zheng Y, Murphy K, Huang M, Hu F. The frontotemporal lobar degeneration risk factor, TMEM106B, regulates lysosomal morphology and function. *Hum Mol Genet*. 2013;22(4):685-695. doi:10.1093/hmg/dds475

25. Finch N, Carrasquillo MM, Baker M, et al. TMEM106B regulates progranulin levels and the penetrance of FTLD in GRN mutation carriers. *Neurology*. 2011;76(5):467-474. doi:10.1212/WNL.0b013e31820a0e3b

26. van Blitterswijk M, Mullen B, Nicholson AM, et al. TMEM106B protects C9ORF72 expansion carriers against frontotemporal dementia. *Acta Neuropathol*. 2014;127(3):397-406. doi:10.1007/s00401-013-1240-4

27. Gallagher MD, Suh E, Grossman M, et al. TMEM106B is a genetic modifier of frontotemporal lobar degeneration with C9orf72 hexanucleotide repeat expansions. *Acta Neuropathol*. 2014;127(3):407-418. doi:10.1007/s00401-013-1239-x

28. Ciani M, Benussi L, Bonvicini C, Ghidoni R. Genome Wide Association Study and Next Generation Sequencing: A Glimmer of Light Toward New Possible Horizons in Frontotemporal Dementia Research. *Front Neurosci*. 2019;13(May):1-8. doi:10.3389/fnins.2019.00506

29. Al-Chalabi A, Fang F, Hanby MF, et al. An estimate of amyotrophic lateral sclerosis heritability using twin data. *J Neurol Neurosurg Psychiatry*. 2010;81(12):1324-1326. doi:10.1136/jnnp.2010.207464

30. Valentine JS, Doucette PA, Zittin Potter S. Copper-Zinc Superoxide Dismutase and Amyotrophic Lateral Sclerosis. *Annu Rev Biochem*. 2005;74(1):563-593. doi:10.1146/annurev.biochem.72.121801.161647

31. Bosco DA, Morfini G, Karabacak NM, et al. Wild-type and mutant SOD1 share an aberrant conformation and a common pathogenic pathway in ALS. *Nat Neurosci*. 2010;13(11):1396-1403. doi:10.1038/nn.2660

32. Deng H-X, Shi Y, Furukawa Y, et al. Conversion to the amyotrophic lateral sclerosis phenotype is associated with intermolecular linked insoluble aggregates of SOD1 in mitochondria. *Proc Natl Acad Sci*. 2006;103(18):7142-7147. doi:10.1073/pnas.0602046103

33. Mathis S, Goizet C, Soulages A, Vallat JM, Masson G Le. Genetics of amyotrophic lateral sclerosis: A review. *J Neurol Sci*. 2019;399(February):217-226. doi:10.1016/j.jns.2019.02.030

34. Lee EB, Lee VMY, Trojanowski JQ. Gains or losses: Molecular mechanisms of TDP43-mediated neurodegeneration. *Nat Rev Neurosci*. 2012;13(1):38-50. doi:10.1038/nrn3121

35. Liu EY, Cali CP, Lee EB. RNA metabolism in neurodegenerative disease. *Dis Model Mech*. 2017;10(5):509-518. doi:10.1242/dmm.028613

36. Rohlfing FW, Tu RK. Genetics of frontotemporal dementia. *Am J Neuroradiol*. 2017;38(1):10-11. doi:10.3174/ajnr.A4972

37. Mackenzie IRA, Neumann M. Molecular neuropathology of frontotemporal dementia: insights into disease mechanisms from postmortem studies. *J Neurochem*. 2016;138:54-70. doi:10.1111/jnc.13588

38. Eriksen JL, Mackenzie IRA. Progranulin: Normal function and role in neurodegeneration. *J Neurochem*. 2008;104(2):287-297. doi:10.1111/j.1471-4159.2007.04968.x

39. Morita M, Andersen PM, Hosler B, et al. A locus on chromosome 9p confers susceptibility to ALS and frontotemporal dementia. *Neurology*. 2006;66:839-844.

40. Vance C, Al-Chalabi A, Ruddy D, et al. Familial amyotrophic lateral sclerosis with frontotemporal dementia is linked to a locus on chromosome 9p13.2-21.3. *Brain*. 2006;129(4):868-876. doi:10.1093/brain/awl030

41. Shatunov A, Mok K, Newhouse S, et al. Chromosome 9p21 in sporadic amyotrophic lateral sclerosis in the UK and seven other countries: A genome-wide association study. *Lancet Neurol*. 2010;9(10):986-994. doi:10.1016/S1474-4422(10)70197-6

42. Laaksovirta H, Peuralinna T, Schymick JC, et al. Chromosome 9p21 in amyotrophic lateral sclerosis in Finland: A genome-wide association study. *Lancet Neurol*. 2010;9(10):978-985. doi:10.1016/S1474-4422(10)70184-8

43. Dejesus-hernandez M, Mackenzie IR, Boeve BF, et al. Expanded GGGGCC hexanucleotide repeat in non-coding region of C9ORF72 causes chromosome 9p-linked frontotemporal dementia and amyotrophic lateral sclerosis. *Neuron*. 2011;72(2):245-256.

doi:10.1016/j.neuron.2011.09.011.Expanded

44. Renton A, Majounie E, Waite A, et al. A hexanucleotide repeat expansion in C9ORF72 is the cause of chromosome 9p21-linked ALS-FTD. *Neuron*. 2011;72(2):257-268. doi:10.1016/j.neuron.2011.09.010

45. Suh E, Lee EB, Neal D, et al. Semi-automated quantification of C9orf72 expansion size reveals inverse correlation between hexanucleotide repeat number and disease duration in frontotemporal degeneration. *Acta Neuropathol*. 2015;130(3):363-372. doi:10.1007/s00401-015-1445-9

46. Van Mossevelde S, van der Zee J, Cruts M, Van Broeckhoven C. Relationship between C9orf72 repeat size and clinical phenotype. *Curr Opin Genet Dev*. 2017;44:117-124. doi:10.1016/j.gde.2017.02.008

47. Russ J, Liu EY, Wu K, et al. Hypermethylation of repeat expanded C9orf72 is a clinical and molecular disease modifier. *Acta Neuropathol*. 2015;129(1):39-52. doi:10.1007/s00401-014-1365-0

48. Loureiro JR, Oliveira CL, Silveira I. Unstable repeat expansions in neurodegenerative diseases: Nucleocytoplasmic transport emerges on the scene. *Neurobiol Aging*. 2016;39:174-183. doi:10.1016/j.neurobiolaging.2015.12.007

49. Mangiarini L, Sathasivam K, Seller M, et al. Exon I of the HD gene with an expanded CAG repeat is sufficient to cause a progressive neurological phenotype in transgenic mice. *Cell*. 1996;87(3):493-506. doi:10.1016/S0092-8674(00)81369-0

50. Verkerk AJMH, Pieretti M, Sutcliffe JS, et al. Identification of a gene (FMR-1) containing a CGG repeat coincident with a breakpoint cluster region exhibiting length variation in fragile X syndrome. *Cell*. 1991;65(5):905-914. doi:10.1016/0092-8674(91)90397-H

51. Campuzano V, Montermini L, Moltò MD, et al. Friedreich's ataxia: Autosomal recessive disease caused by an intronic GAA triplet repeat expansion. *Science (80- )*. 1996;271(5254):1423-1427. doi:10.1126/science.271.5254.1423

52. Grabczyk E, Usdin K. The GAA•TTC triplet repeat expanded in Friedreich's ataxia impedes transcription elongation by T7 RNA polymerase in a length and supercoil dependent manner. *Nucleic Acids Res*. 2000;28(14):2815-2822. doi:10.1093/nar/28.14.2815

53. Liquori CL, Ricker K, Moseley ML, et al. Myotonic dystrophy type 2 caused by a CCTG expansion in intron 1 of ZNF9. *Science (80- )*. 2001;293(5531):864-867. doi:10.1126/science.1062125

54. Waite AJ, Bäumer D, East S, et al. Reduced C9orf72 protein levels in frontal cortex of amyotrophic lateral sclerosis and frontotemporal degeneration brain with the C9ORF72 hexanucleotide repeat expansion. *Neurobiol Aging*. 2014;35(7):1779.e5-1779.e13. doi:10.1016/j.neurobiolaging.2014.01.016

55. Levine TP, Daniels RD, Gatta AT, Wong LH, Hayes MJ. The product of C9orf72, a gene strongly implicated in neurodegeneration, is structurally related to DENN Rab-GEFs. *Bioinformatics*. 2013;29(4):499-503. doi:10.1093/bioinformatics/bts725

56. Farg MA, Sundaramoorthy V, Sultana JM, et al. C9ORF72, implicated in amytrophic lateral sclerosis and frontotemporal dementia, regulates endosomal trafficking. *Hum Mol*

*Genet*. 2014;23(13):3579-3595. doi:10.1093/hmg/ddu068

57. Ciura S, Lattante S, Le Ber I, et al. Loss of function of C9orf72 causes motor deficits in a zebrafish model of amyotrophic lateral sclerosis. *Ann Neurol*. 2013;74(2):180-187. doi:10.1002/ana.23946

58. O'Rourke JG, Bogdanik L, Yáñez A, et al. C9orf72 is required for proper macrophage and microglial function in mice. *Science (80- )*. 2016;351(6279):1324-1329. doi:10.1126/science.aaf1064

59. Burberry A, Suzuki N, Wang JY, et al. Loss-of-function mutations in the C9ORF72 mouse ortholog cause fatal autoimmune disease. *Sci Transl Med*. 2016;8(347):93. doi:10.1126/scitranslmed.aaf6038

60. Atanasio A, Decman V, White D, et al. C9orf72 ablation causes immune dysregulation characterized by leukocyte expansion, autoantibody production, and glomerulonephropathy in mice. *Sci Rep*. 2016;6(March):1-14. doi:10.1038/srep23204

61. Fratta P, Poulter M, Lashley T, et al. Homozygosity for the C9orf72 GGGGCC repeat expansion in frontotemporal dementia. *Acta Neuropathol*. 2013;126(3):401-409. doi:10.1007/s00401-013-1147-0

62. Viodé A, Fournier C, Camuzat A, et al. New Antibody-Free Mass Spectrometry-Based Quantification Reveals That C9ORF72 Long Protein Isoform Is Reduced in the Frontal Cortex of Hexanucleotide-Repeat Expansion Carriers. *Front Neurosci*. 2018;12(August):1-11. doi:10.3389/fnins.2018.00589

63. Frick P, Sellier C, Mackenzie IRA, et al. Novel antibodies reveal presynaptic localization of C9orf72 protein and reduced protein levels in C9orf72 mutation carriers. *Acta Neuropathol Commun*. 2018;6(72):1-17. doi:10.1186/s40478-018-0579-0

64. Koppers M, Blokhuis AM, Westeneng HJ, et al. C9orf72 ablation in mice does not cause motor neuron degeneration or motor deficits. *Ann Neurol*. 2015;78(3):426-438. doi:10.1002/ana.24453

65. Nassif M, Woehlbier U, Manque PA. The enigmatic role of C9ORF72 in autophagy. *Front Neurosci*. 2017;11(AUG):1-10. doi:10.3389/fnins.2017.00442

66. Zhang D, Iyer LM, He F, Aravind L. Discovery of novel DENN proteins: Implications for the evolution of eukaryotic intracellular membrane structures and human disease. *Front Genet*. 2012;3(DEC):1-10. doi:10.3389/fgene.2012.00283

67. Webster CP, Smith EF, Bauer CS, et al. The C9orf72 protein interacts with Rab1a and the ULK1 complex to regulate initiation of autophagy. *EMBO J*. 2016;35(15):1656-1676. doi:10.15252/embj.201694401

68. Corbier C, Sellier C. C9ORF72 is a GDP/GTP exchange factor for Rab8 and Rab39 and regulates autophagy. *Small GTPases*. 2016:00-00. doi:10.1080/21541248.2016.1212688

69. Yang M, Liang C, Swaminathan K, et al. A C9ORF72/SMCR8-containing complex regulates ULK1 and plays a dual role in autophagy. *Sci Adv*. 2016;(September):1-17. doi:10.1126/sciadv.1601167

70. Aoki Y, Manzano R, Lee Y, et al. C9orf72 and RAB7L1 regulate vesicle trafficking in amyotrophic lateral sclerosis and frontotemporal dementia. *Brain*. 2017;140(4):887-897.

doi:10.1093/brain/awx024

71. Liu Y, Wang T, Ji YJ, et al. A C9orf72-CARM1 axis regulates lipid metabolism under glucose starvation-induced nutrient stress. *Genes Dev*. 2018;32(21-22):1380-1397. doi:10.1101/gad.315564.118

72. Sullivan PM, Zhou X, Robins AM, et al. The ALS/FTLD associated protein C9orf72 associates with SMCR8 and WDR41 to regulate the autophagy-lysosome pathway. *Acta Neuropathol Commun*. 2016;4(1):51. doi:10.1186/s40478-016-0324-5

73. Amick J, Tharkeshwar AK, Amaya, C, Ferguson SM. WDR41 supports lysosomal response to changes in amino acid availability. Barr FA, ed. *Mol Biol Cell*. 2018;29(18):2213-2227. doi:10.1091/mbc.E17-12-0703

74. Ugolino J, Ji YJ, Conchina K, et al. Loss of C9orf72 Enhances Autophagic Activity via Deregulated mTOR and TFEB Signaling. *PLoS Genet*. 2016;12(11). doi:10.1371/journal.pgen.1006443

75. Sellier C, Campanari M-L, Julie Corbier C, et al. Loss of C9ORF72 impairs autophagy and synergizes with polyQ Ataxin-2 to induce motor neuron dysfunction and cell death. *EMBO J*. 2016;35(Icm):1-22. doi:10.15252/embj.201593350

76. Chitiprolu M, Jagow C, Tremblay V, et al. A complex of C9ORF72 and p62 uses arginine methylation to eliminate stress granules by autophagy. *Nat Commun*. 2018;9(1):2794. doi:10.1038/s41467-018-05273-7

77. Maharjan N, Künzli C, Buthey K, Saxena S. C9ORF72 Regulates Stress Granule Formation and Its Deficiency Impairs Stress Granule Assembly, Hypersensitizing Cells to Stress. *Mol Neurobiol*. 2017;54(4):3062-3077. doi:10.1007/s12035-016-9850-1

78. Stenmark H. Rab GTPases as coordinators of vesicle traffic. *Nat Rev Mol Cell Biol*. 2009;10(8):513-525. doi:10.1038/nrm2728

79. Tsun ZY, Bar-Peled L, Chantranupong L, et al. The folliculin tumor suppressor is a GAP for the RagC/D GTPases that signal amino acid levels to mTORC1. *Mol Cell*. 2013;52(4):495-505. doi:10.1016/j.molcel.2013.09.016

80. Ling SC, Polymenidou M, Cleveland DW. Converging mechanisms in als and FTD: Disrupted RNA and protein homeostasis. *Neuron*. 2013;79(3):416-438. doi:10.1016/j.neuron.2013.07.033

81. Sareen D, O'Rourke JG, Meera P, et al. Targeting RNA foci in iPSC-derived motor neurons from ALS patients with a C9ORF72 repeat expansion. *Sci Transl Med*. 2013;5(208):208ra149. doi:10.1126/scitranslmed.3007529

82. Almeida S, Gascon E, Tran H, et al. Modeling key pathological features of frontotemporal dementia with C9ORF72 repeat expansion in iPSC-derived human neurons. *Acta Neuropathol*. 2013;126(3):385-399. doi:10.1007/s00401-013-1149-y

83. Donnelly CJ, Zhang P, Pham JT, et al. RNA Toxicity from the ALS/FTD C9ORF72 Expansion Is Mitigated by Antisense Intervention. *Neuron*. 2013;80(2):415-428. doi:10.1016/j.neuron.2013.10.015.RNA

84. Chew J, Gendron TF, Prudencio M, Sasaguri H, Zhang YJ, Castanedes-Casey M, Lee CW, Jansen-West K, Kurti A, Murray ME, Bieniek KF, Bauer PO, Whitelaw EC, Rousseau

L, Stankowski JN, Stetler C, Daughrity LM, Perkerson EA, Desaro P, Johnston A, Overstreet K, PL. C9ORF72 repeat expansions in mice cause TDP-43 pathology, neuronal loss, and behavioral deficits. *Science (80- )*. 2015;348(6239):1151-1154. http://www.ncbi.nlm.nih.gov/pubmed/?term=c9orf72+repeat+expansion+in+mice+cause+TDP43.

85. O'Rourke JG, Bogdanik L, Muhammad AKMG, et al. C9orf72 BAC Transgenic Mice Display Typical Pathologic Features of ALS/FTD. *Neuron*. 2015;88(5):892-901. doi:10.1016/j.neuron.2015.10.027

86. Cooper-Knock J, Walsh MJ, Higginbottom A, et al. Sequestration of multiple RNA recognition motif-containing proteins by C9orf72 repeat expansions. *Brain*. 2014;137(7):2040-2051. doi:10.1093/brain/awu120

87. Conlon EG, Lu L, Sharma A, et al. The C9ORF72 GGGGCC expansion forms RNA G-quadruplex inclusions and sequesters hnRNP H to disrupt splicing in ALS brains. *Elife*. 2016;5(September2016). doi:10.7554/eLife.17820

88. Lee YB, Chen HJ, Peres JN, et al. Hexanucleotide repeats in ALS/FTD form length-dependent RNA Foci, sequester RNA binding proteins, and are neurotoxic. *Cell Rep*. 2013;5(5):1178-1186. doi:10.1016/j.celrep.2013.10.049

89. Zhang K, Donnelly CJ, Haeusler AR, et al. The C9orf72 repeat expansion disrupts nucleocytoplasmic transport. *Nature*. 2015;525(7567):56-61. doi:10.1038/nature14973

90. Zhang K, Donnelly CJ, Haeusler AR, et al. The C9orf72 repeat expansion disrupts nucleocytoplasmic transport. *Nature*. 2015;525(7567):56-61. doi:10.1038/nature14973

91. Freibaum BD, Lu Y, Lopez-Gonzalez R, et al. GGGGCC repeat expansion in C9orf72 compromises nucleocytoplasmic transport. *Nature*. 2015;525(7567):129-133. doi:10.1038/nature14974

92. Cooper-Knock J, Higginbottom A, Connor-Robson N, et al. Clinical/Scientific Notes C9ORF72 transcription in a frontotemporal dementia case with two expanded alleles. *Neurology*. 2013;81(19):1719-1721. doi:10.1212/01.wnl.0000435295.41974.2e

93. Cooper-Knock J, Bury JJ, Heath PR, et al. C9ORF72 GGGGCC expanded repeats produce splicing dysregulation which correlates with disease severity in amyotrophic lateral sclerosis. *PLoS One*. 2015;10(5):1-17. doi:10.1371/journal.pone.0127376

94. Ash PEA, Bieniek KF, Gendron TF, et al. Unconventional Translation of C9ORF72 GGGGCC Expansion Generates Insoluble Polypeptides Specific to c9FTD/ALS. *Neuron*. 2013;77(4):639-646. doi:10.1016/j.neuron.2013.02.004

95. Zu T, Gibbens B, Doty NS, et al. Non-ATG-initiated translation directed by microsatellite expansions. *Proc Natl Acad Sci*. 2011;108(1):260-265. doi:10.1073/pnas.1013343108

96. Mori K, Weng S-M, Arzberger T, et al. The C9orf72 GGGGCC Repeat Is Translated into Aggregating Dipeptide-Repeat Proteins in FTLD/ALS. *Science (80- )*. 2013;339(6125):1335-1338. doi:10.1126/SCIENCE.1232927

97. Cleary JD, Ranum LPW. Repeat associated non-ATG (RAN) translation: New starts in microsatellite expansion disorders. *Curr Opin Genet Dev*. 2014;26:6-15. doi:10.1016/j.gde.2014.03.002

98. Kwon I, Xiang S, Kato M, et al. Poly-dipeptides encoded by the C9orf72 repeats bind nucleoli, impede RNA biogenesis, and kill cells. *Science (80- )*. 2014;345(6201):1139-1145. doi:10.1126/science.1254917

99. May S, Hornburg D, Schludi MH, et al. C9orf72 FTLD/ALS-associated Gly-Ala dipeptide repeat proteins cause neuronal toxicity and Unc119 sequestration. *Acta Neuropathol*. 2014;128(4):485-503. doi:10.1007/s00401-014-1329-4

100. Wen X, Tan W, Westergard T, et al. Antisense proline-arginine RAN dipeptides linked to C9ORF72-ALS/FTD form toxic nuclear aggregates that initiate invitro and invivo neuronal death. *Neuron*. 2014;84(6):1213-1225. doi:10.1016/j.neuron.2014.12.010

101. Zhang YJ, Jansen-West K, Xu YF, et al. Aggregation-prone c9FTD/ALS poly(GA) RAN-translated proteins cause neurotoxicity by inducing ER stress. *Acta Neuropathol*. 2014;128(4):505-524. doi:10.1007/s00401-014-1336-5

102. Jovičić A, Mertens J, Boeynaems S, et al. Modifiers of C9orf72 dipeptide repeat toxicity connect nucleocytoplasmic transport defects to FTD/ALS. *Nat Neurosci*. 2015;18(9):1226-1229. doi:10.1038/nn.4085

103. Zhang Y-J, Gendron TF, Ebbert MTW, et al. Poly(GR) impairs protein translation and stress granule dynamics in C9orf72-associated frontotemporal dementia and amyotrophic lateral sclerosis. *Nat Med*. 2018:1. doi:10.1038/s41591-018-0071-1

104. Liu EY, Russ J, Cali CP, Phan JM, Amlie-Wolf A, Lee EB. Loss of Nuclear TDP-43 Is Associated with Decondensation of LINE Retrotransposons. *Cell Rep*. 2019;27(5):1409-1421.e6. doi:10.1016/j.celrep.2019.04.003

105. Hayes LR, Rothstein JD. C9ORF72-ALS/FTD: Transgenic Mice Make a Come-BAC. *Neuron*. 2016;90(3):427-431. doi:10.1016/j.neuron.2016.04.026

106. Zhang Y-J, Gendron TF, Grima JC, et al. C9ORF72 poly(GA) aggregates sequester and impair HR23 and nucleocytoplasmic transport proteins. *Nat Neurosci*. 2016;19(5):668-677. doi:10.1038/nn.4272

107. Peters OM, Cabrera GT, Tran H, et al. Human C9ORF72 Hexanucleotide Expansion Reproduces RNA Foci and Dipeptide Repeat Proteins but Not Neurodegeneration in BAC Transgenic Mice. *Neuron*. 2015;88(5):902-909. doi:10.1016/j.neuron.2015.11.018

108. Jiang J, Zhu Q, Gendron TF, et al. Gain of Toxicity from ALS/FTD-Linked Repeat Expansions in C9ORF72 Is Alleviated by Antisense Oligonucleotides Targeting GGGGCC-Containing RNAs. *Neuron*. 2016;90(3):535-550. doi:10.1016/j.neuron.2016.04.006

109. Liu Y, Pattamatta A, Zu T, et al. C9orf72 BAC Mouse Model with Motor Deficits and Neurodegenerative Features of ALS/FTD. *Neuron*. 2016;90(3):521-534. doi:10.1016/j.neuron.2016.04.005

110. MacKenzie IR, Arzberger T, Kremmer E, et al. Dipeptide repeat protein pathology in C9ORF72 mutation cases: Clinico-pathological correlations. *Acta Neuropathol*. 2013;126(6):859-879. doi:10.1007/s00401-013-1181-y

111. Baborie A, Griffiths TD, Jaros E, et al. Accumulation of dipeptide repeat proteins predates that of TDP-43 in frontotemporal lobar degeneration associated with hexanucleotide repeat expansions in C9ORF72 gene. *Neuropathol Appl Neurobiol*. 2015;41(5):601-612.

doi:10.1111/nan.12178

112. Vatsavayai SC, Yoon SJ, Gardner RC, et al. Timing and significance of pathological features in C9orf72 expansion-associated frontotemporal dementia. *Brain*. 2016;139(12):3202-3216. doi:10.1093/brain/aww250

113. Y.S. D, H. B, A.C. R, et al. Brain distribution of dipeptide repeat proteins in frontotemporal lobar degeneration and motor neurone disease associated with expansions in C9ORF72. *Acta Neuropathol Commun*. 2014;2:70. doi:10.1186/2051-5960-2-70

114. Scaber J, Talbot K. What is the role of TDP-43 in C9orf72-related amyotrophic lateral sclerosis and frontemporal dementia? *Brain*. 2016;139(12):3057-3059. doi:10.1093/brain/aww264

115. Shi Y, Lin S, Staats KA, et al. Haploinsufficiency leads to neurodegeneration in C9ORF72 ALS/FTD human induced motor neurons. *Nat Med*. 2018;(January). doi:10.1038/nm.4490

116. Majounie E, Renton AE, Mok K, et al. Frequency of the C9orf72 hexanucleotide repeat expansion in patients with amyotrophic lateral sclerosis and frontotemporal dementia: A cross-sectional study. *Lancet Neurol*. 2012;11(4):323-330. doi:10.1016/S1474-4422(12)70043-1

117. Gijselinck I, Van Mossevelde S, van der Zee J, et al. The C9orf72 repeat size correlates with onset age of disease, DNA methylation and transcriptional downregulation of the promoter. *Mol Psychiatry*. 2015;21(August):1-13. doi:10.1038/mp.2015.159

118. Dobson-Stone C, Hallupp M, Loy CT, et al. C9ORF72 Repeat Expansion in Australian and Spanish Frontotemporal Dementia Patients. *PLoS One*. 2013;8(2). doi:10.1371/journal.pone.0056899

119. Xi Z, Zinman L, Grinberg Y, et al. Investigation of C9orf72 in 4 Neurodegenerative Disorders. *Arch Neurol*. 2012;69(12):1583. doi:10.1001/archneurol.2012.2016

120. Gami P, Murray C, Schottlaender L, et al. A 30-unit hexanucleotide repeat expansion in C9orf72 induces pathological lesions with dipeptide-repeat proteins and RNA foci, but not TDP-43 inclusions and clinical disease. *Acta Neuropathol*. 2015;130(4):599-601. doi:10.1007/s00401-015-1473-5

121. Dedeene L, Van Schoor E, Race V, et al. An ALS case with 38 (G4C2)-repeats in the C9orf72 gene shows TDP-43 and sparse dipeptide repeat protein pathology. *Acta Neuropathol*. 2019;137(5):855-858. doi:10.1007/s00401-019-01996-z

122. McGoldrick P, Zhang M, van Blitterswijk M, et al. Unaffected mosaic C9orf72 case. *Neurology*. 2018;90(4):e323-e331. doi:10.1212/WNL.0000000000004865

123. Theuns J, Verstraeten A, Sleegers K, et al. Global investigation and meta-analysis of the C9orf72 (G4C2)n repeat in Parkinson disease. *Neurology*. 2014;83(21):1906-1913. doi:10.1212/WNL.0000000000001012

124. Jiao B, Guo J, Wang Y, et al. C9orf72 mutation is rare in Alzheimer's disease, Parkinson's disease, and essential tremor in China. *Front Cell Neurosci*. 2013;7:164. doi:10.3389/fncel.2013.00164

125. Nuytemans K, Bademci G, Kohli MM, et al. C9orf72 intermediate repeat copies are a significant risk factor for parkinson disease. *Ann Hum Genet*. 2013;77(5):351-363.

doi:10.1111/ahg.12033

126. Nuytemans K, Inchausti V, Beecham GW, et al. Absence of C9ORF72 expanded or intermediate repeats in autopsy-confirmed Parkinson's disease. *Mov Disord*. 2014;29(6):827-830. doi:10.1002/mds.25838

127. Fredi M, Cavazzana I, Biasiotto G, et al. C9orf72 Intermediate Alleles in Patients with Amyotrophic Lateral Sclerosis, Systemic Lupus Erythematosus, and Rheumatoid Arthritis. *NeuroMolecular Med*. 2019;0(0):0. doi:10.1007/s12017-019-08528-8

128. Schübeler D. Function and information content of DNA methylation. *Nature*. 2015;517(7534):321-326. doi:10.1038/nature14192

129. Jang HS, Shin WJ, Lee JE, Do JT. CpG and non-CpG methylation in epigenetic gene regulation and brain function. *Genes (Basel)*. 2017;8(6):2-20. doi:10.3390/genes8060148

130. Xu G-L, Bestor TH, Bourc'his D, et al. Chromosome instability and immunodeficiency syndrome caused by mutations in a DNA methyltransferase gene. *Nature*. 1999;402(6758):187-191. doi:10.1038/46052

131. Vogelstein B, Kinzler K, Lengauer C. DNA methylation and genetic instability in colorectal cancer cells. *Proc Natl Acad Sci*. 1997;94(March):2545-2550.

132. Lardenoije R, Iatrou A, Kenis G, et al. The epigenetics of aging and neurodegeneration. *Prog Neurobiol*. 2015;131:21-64. doi:10.1016/j.pneurobio.2015.05.002

133. Al-Mahdawi S, Pinto RM, Ismail O, et al. The Friedreich ataxia GAA repeat expansion mutation induces comparable epigenetic changes in human and transgenic mouse brain and heart tissues. *Hum Mol Genet*. 2008;17(5):735-746. doi:10.1093/hmg/ddm346

134. Steinbach P, Gläser D, Vogel W, Wolf M, Schwemmle S. The DMPK Gene of Severely Affected Myotonic Dystrophy Patients Is Hypermethylated Proximal to the Largely Expanded CTG Repeat. *Am J Hum Genet*. 1998;62(2):278-285. doi:10.1086/301711

135. Pieretti M, Zhang F, Fu Y-H, et al. Absence of expression of the FMR-1 gene in fragile X syndrome. *Cell*. 1991;66(4):817-822. doi:10.1016/0092-8674(91)90125-I

136. Xi Z, Zinman L, Moreno D, et al. Hypermethylation of the CpG island near the G4C2 repeat in ALS with a C9orf72 expansion. *Am J Hum Genet*. 2013;92(6):981-989. doi:10.1016/j.ajhg.2013.04.017

137. Liu EY, Russ J, Wu K, et al. C9orf72 hypermethylation protects against repeat expansion-associated pathology in ALS/FTD. *Acta Neuropathol*. 2014;128(4):525-541. doi:10.1007/s00401-014-1286-y

138. Belzil V V., Bauer PO, Gendron TF, Murray ME, Dickson D, Petrucelli L. Characterization of DNA hypermethylation in the cerebellum of c9FTD/ALS patients. *Brain Res*. 2014;1584:15-21. doi:10.1016/j.brainres.2014.02.015

139. Xi Z, Rainero I, Rubino E, et al. Hypermethylation of the CpG-island near the C9orf72 G4C2-repeat expansion in FTLD patients. *Hum Mol Genet*. 2014;23(21):5630-5637. doi:10.1093/hmg/ddu279

140. McMillan CT, Russ J, Wood EM, et al. C9orf72 promoter hypermethylation is neuroprotective: Neuroimaging and neuropathologic evidence. *Neurology*. 2015;84:1622-

1630. doi:10.1212/WNL.0000000000001495

141. Thakore PI, Black JB, Hilton IB, Gersbach CA. Editing the epigenome: technologies for programmable transcription and epigenetic modulation. *Nat Methods*. 2016;13(2):127-137. doi:10.1038/nmeth.3733

142. Hu J, Lei Y, Wong WK, et al. Direct activation of human and mouse Oct4 genes using engineered TALE and Cas9 transcription factors. *Nucleic Acids Res*. 2014;42(7):4375-4390. doi:10.1093/nar/gku109

143. Maeder ML, Angstman JF, Richardson ME, et al. Targeted DNA Demethylation and Endogenous Gene Activation Using Programmable TALE-TET1 Fusions. *Nat Biotechnol*. 2014;31(12):1-15. doi:10.1038/nbt.2726.Targeted

144. Hilton IB, M, D'Ippolito A, Vockley CM, et al. Epigenome editing by a CRISPR-Cas9-based acetyltransferase activates genes from promoters and enhancers. *Nat Biotechnol*. 2015;33(5):510-517. doi:10.1038/nbt.3199

145. Cui C, Gan Y, Gu L, et al. P16-specific DNA methylation by engineered zinc finger methyltransferase inactivates gene transcription and promotes cancer metastasis. *Genome Biol*. 2015;16(1):1-12. doi:10.1186/s13059-015-0819-6

146. Bernstein DL, Le Lay JE, Ruano EG, Kaestner KH. TALE-mediated epigenetic suppression of CDKN2A increases replication in human fibroblasts. *J Clin Invest*. 2015;125(5):1998-2006. doi:10.1172/JCI77321

147. Kearns N a, Pham H, Tabak B, et al. Functional annotation of native enhancers with a Cas9-histone demethylase fusion. *Nat Methods*. 2015;12(5):401-403. doi:10.1038/nmeth.3325

148. Amabile A, Migliara A, Capasso P, et al. Inheritable Silencing of Endogenous Genes by Hit-and-Run Targeted Epigenetic Editing. *Cell*. 2016;167(1):219-232.e14. doi:10.1016/j.cell.2016.09.006

149. Liu XS, Wu H, Ji X, et al. Editing DNA Methylation in the Mammalian Genome. *Cell*. 2016;167(1):233-247.e17. doi:10.1016/j.cell.2016.08.056

150. Kantor B, Tagliafierro L, Gu J, et al. Downregulation of SNCA Expression by Targeted Editing of DNA Methylation: A Potential Strategy for Precision Therapy in PD. *Mol Ther*. 2018;26(11):2638-2649. doi:10.1016/j.ymthe.2018.08.019

151. Liu X, Wu H, Krzisch M, et al. Rescue of fragile X syndrome by DNA methylation editing of the FMR1 gene. *Cell*. 2018;172(5):in revision. doi:10.1016/j.cell.2018.01.012

152. Vojta A, Dobrinic P, Tadic V, et al. Repurposing the CRISPR-Cas9 system for targeted DNA methylation. *Nucleic Acids Res*. 2016;44(12):5615-5628. doi:10.1093/nar/gkw159

153. Galonska C, Charlton J, Mattei AL, et al. Genome-wide tracking of dCas9-methyltransferase footprints. *Nat Commun*. 2018;9(1):597. doi:10.1038/s41467-017-02708-5

154. Lin L, Liu Y, Xu F, et al. Genome-wide determination of on-target and off-target characteristics for RNA-guided DNA methylation by dCas9 methyltransferases. *Gigascience*. 2018;7(3):1-19. doi:10.1093/gigascience/giy011

155. Pflueger C, Tan D, Swain T, et al. A modular dCas9-SunTag DNMT3A epigenome editing system overcomes pervasive off-target activity of direct fusion dCas9-DNMT3A constructs. *Genome Res*. 2018;28(8):1193-1206. doi:10.1101/gr.233049.117

156. Grünewald J, Zhou R, Garcia SP, et al. Transcriptome-wide off-target RNA editing induced by CRISPR-guided DNA base editors. *Nature*. 2019;569(7756):433-437. doi:10.1038/s41586-019-1161-z

157. Josephs EA, Kocak DD, Fitzgibbon CJ, McMenemy J, Gersbach CA, Marszalek PE. Structure and specificity of the RNA-guided endonuclease Cas9 during DNA interrogation, target binding and cleavage. *Nucleic Acids Res*. 2015;43(18):8924-8941. doi:10.1093/nar/gkv892

158. Wu X, Scott DA, Kriz AJ, et al. Genome-wide binding of the CRISPR endonuclease Cas9 in mammalian cells. *Nat Biotechnol*. 2014;32(7):670-676. doi:10.1038/nbt.2889

159. Hsu PD, Scott DA, Weinstein JA, et al. DNA targeting specificity of RNA-guided Cas9 nucleases. *Nat Biotechnol*. 2013;31(9):827-832. doi:10.1038/nbt.2647

160. Mortusewicz O, Schermelleh L, Walter J, Cardoso MC, Leonhardt H. Recruitment of DNA methyltransferase I to DNA repair sites. *Proc Natl Acad Sci*. 2005;102(25):8905-8909. doi:10.1073/pnas.0501034102

161. Lee GE, Kim JH, Taylor M, Muller MT. DNA methyltransferase 1-associated protein (DMAP1) is a co-repressor that stimulates DNA methylation globally and locally at sites of double strand break repair. *J Biol Chem*. 2010;285(48):37630-37640. doi:10.1074/jbc.M110.148536

162. O'Hagan HM, Wang W, Sen S, et al. Oxidative Damage Targets Complexes Containing DNA Methyltransferases, SIRT1, and Polycomb Members to Promoter CpG Islands. *Cancer Cell*. 2011;20(5):606-619. doi:10.1016/j.ccr.2011.09.012

163. Ha K, Lee GE, Palii SS, et al. Rapid and transient recruitment of DNMT1 to DNA double-strand breaks is mediated by its interaction with multiple components of the DNA damage response machinery. *Hum Mol Genet*. 2011;20(1):126-140. doi:10.1093/hmg/ddq451

164. Iyama T, Wilson DM. DNA repair mechanisms in dividing and non-dividing cells. *DNA Repair (Amst)*. 2013;12(8):620-636. doi:10.1016/j.dnarep.2013.04.015

165. Mansour WY, Schumacher S, Rosskopf R, et al. Hierarchy of nonhomologous end-joining, single-strand annealing and gene conversion at site-directed DNA double-strand breaks. *Nucleic Acids Res*. 2008;36(12):4088-4098. doi:10.1093/nar/gkn347

166. San Filippo J, Sung P, Klein H. Mechanism of Eukaryotic Homologous Recombination. *Annu Rev Biochem*. 2008;77(1):229-257. doi:10.1146/annurev.biochem.77.061306.125255

167. Cuozzo C, Porcellini A, Angrisano T, et al. DNA damage, homology-directed repair, and DNA methylation. *PLoS Genet*. 2007;3(7):1144-1162. doi:10.1371/journal.pgen.0030110

168. O'Hagan HM, Mohammad HP, Baylin SB. Double strand breaks can initiate gene silencing and SIRT1-dependent onset of DNA methylation in an exogenous promoter CpG island. *PLoS Genet*. 2008;4(8). doi:10.1371/journal.pgen.1000155

169. Morano A, Angrisano T, Russo G, et al. Targeted DNA methylation by homology-directed

repair in mammalian cells. Transcription reshapes methylation on the repaired gene. *Nucleic Acids Res*. 2014;42(2):804-821. doi:10.1093/nar/gkt920

170. Russo G, Landi R, Pezone A, et al. DNA damage and Repair Modify DNA methylation and Chromatin Domain of the Targeted Locus: Mechanism of allele methylation polymorphism. *Sci Rep*. 2016;6(August):1-14. doi:10.1038/srep33222

171. Nishiyama A, Yamaguchi L, Sharif J, et al. Uhrf1-dependent H3K23 ubiquitylation couples maintenance DNA methylation and replication. *Nature*. 2013;502(7470):249-253. doi:10.1038/nature12488

172. Allen B, Pezone A, Porcellini A, Muller MT, Masternak MM. Non-homologous end joining induced alterations in DNA methylation: A source of permanent epigenetic change. *Oncotarget*. 2017;8(25):40359-40372. doi:10.18632/oncotarget.16122

173. Armstrong MJ, Litvan I, Lang AE, et al. Criteria for the diagnosis of corticobasal degeneration. *Neurology*. 2013;80(5):496-503. doi:10.1212/WNL.0b013e31827f0fd1

174. Lei Y, Zhang X, Su J, et al. Targeted DNA methylation in vivo using an engineered dCas9-MQ1 fusion protein. *Nat Commun*. 2017;8:16026. doi:10.1038/ncomms16026

175. Xiong T, Meister GE, Workman RE, et al. Targeted DNA methylation in human cells using engineered dCas9-methyltransferases. *Sci Rep*. 2017;7(1):6732. doi:10.1038/s41598-017-06757-0

176. Falahi F, Sgro A, Blancafort P. Epigenome Engineering in Cancer: Fairytale or a Realistic Path to the Clinic? *Front Oncol*. 2015;5(February):22. doi:10.3389/fonc.2015.00022

177. Stepper P, Kungulovski G, Jurkowska RZ, et al. Efficient targeted DNA methylation with chimeric dCas9-Dnmt3a-Dnmt3L methyltransferase. *Nucleic Acids Res*. 2017;45(4):1703-1713. doi:10.1093/nar/gkw1112

178. Inoue K, Oliveira LMA, Abeliovich A. CRISPR Transcriptional Activation Analysis Unmasks an Occult γ-Secretase Processivity Defect in Familial Alzheimer's Disease Skin Fibroblasts. *Cell Rep*. 2017;21(7):1727-1736. doi:10.1016/j.celrep.2017.10.075

179. Ran FA, Hsu PD, Wright J, Agarwala V, Scott DA, Zhang F. Genome engineering using the CRISPR-Cas9 system. *Nat Protoc*. 2013;8(11):2281-2308. doi:10.1038/nprot.2013.143

180. Miller JC, Holmes MC, Wang J, et al. An improved zinc-finger nuclease architecture for highly specific genome editing. *Nat Biotechnol*. 2007;25(7):778-785. doi:10.1038/nbt1319

181. Rohde C, Zhang Y, Reinhardt R, Jeltsch A. BISMA - Fast and accurate bisulfite sequencing data analysis of individual clones from unique and repetitive sequences. *BMC Bioinformatics*. 2010;11(1):230. doi:10.1186/1471-2105-11-230

182. Lienert F, Wirbelauer C, Som I, Dean A, Mohn F, Schübeler D. Identification of genetic elements that autonomously determine DNA methylation states. *Nat Genet*. 2011;43(11):1091-1097. doi:10.1038/ng.946

183. Hahm JY, Kim JY, Park JW, et al. Methylation of UHRF1 by SET7 is essential for DNA double-strand break repair. *Nucleic Acids Res*. 2019;47(1):184-196. doi:10.1093/nar/gky975

184. Bagci H, Fisher AG. Dna demethylation in pluripotency and reprogramming: The role of Tet proteins and cell division. *Cell Stem Cell*. 2013;13(3):265-269. doi:10.1016/j.stem.2013.08.005

185. van Blitterswijk M, DeJesus-Hernandez M, Niemantsverdriet E, et al. Association between repeat sizes and clinical and pathological characteristics in carriers of C9ORF72 repeat expansions (Xpansize-72): A cross-sectional cohort study. *Lancet Neurol*. 2013;12(10):978-988. doi:10.1016/S1474-4422(13)70210-2

186. DeJesus-Hernandez M, Rayaprolu S, Soto-Ortolaza AI, et al. Analysis of the C9orf72 repeat in Parkinson's disease, essential tremor and restless legs syndrome. *Park Relat Disord*. 2013;19(2):198-201. doi:10.1016/j.parkreldis.2012.09.013

187. Harms MB, Neumann D, Benitez BA, et al. Parkinson disease is not associated with C9ORF72 repeat expansions. *Neurobiol Aging*. 2013;34(5):1519. doi:10.1016/j.neurobiolaging.2012.10.001

188. van der Zee J, Gijselinck I, Dillen L, et al. A Pan-European Study of the C9orf72 Repeat Associated with FTLD: Geographic Prevalence, Genomic Instability, and Intermediate Repeats. *Hum Mutat*. 2013;34(2):363-373. doi:10.1002/humu.22244

189. Yeh TH, Lai SC, Weng YH, et al. Screening for C9orf72 repeat expansions in parkinsonian syndromes. *Neurobiol Aging*. 2013;34(4):1311.e3-1311.e4. doi:10.1016/j.neurobiolaging.2012.09.002

190. Lesage S, Le Ber I, Condroyer C, et al. C9orf72 repeat expansions are a rare genetic cause of parkinsonism. *Brain*. 2013;136(2):385-391. doi:10.1093/brain/aws357

191. Akimoto C, Forsgren L, Linder J, et al. No GGGGCC-hexanucleotide repeat expansion in C9ORF72 in parkinsonism patients in Sweden. *Amyotroph Lateral Scler Front Degener*. 2013;14(1):26-29. doi:10.3109/17482968.2012.725415

192. Cannas A, Solla P, Borghero G, et al. C9ORF72 intermediate repeat expansion in patients affected by atypical Parkinsonian syndromes or parkinson's disease complicated by psychosis or dementia in a Sardinian population. *J Neurol*. 2015;262(11):2498-2503. doi:10.1007/s00415-015-7873-6

193. Schottlaender L V., Polke JM, Ling H, et al. The analysis of C9orf72 repeat expansions in a large series of clinically and pathologically diagnosed cases with atypical parkinsonism. *Neurobiol Aging*. 2015;36(2):1221.e1-1221.e6. doi:10.1016/j.neurobiolaging.2014.08.024

194. Armstrong MJ, Litvan I, Lang AE, et al. Criteria for the diagnosis of corticobasal degeneration. *Neurology*. 2013;80(5):496-503. doi:10.1212/WNL.0b013e31827f0fd1

195. Kouri N, Murray ME, Hassan A, et al. Neuropathological features of corticobasal degeneration presenting as corticobasal syndrome or Richardson syndrome. *Brain*. 2011;134(11):3264-3275. doi:10.1093/brain/awr234

196. Dickson DW, Bergeron C, Chin SS, et al. Office of rare diseases neuropathologic criteria for corticobasal degeneration. *J Neuropathol Exp Neurol*. 2002;61(11):935-946. doi:10.1093/jnen/61.11.935

197. Litvan I, Agid Y, Goetz C, et al. Accuracy of the clinical diagnosis of corticobasal degeneration: a clinicopathologic study. *Neurology*. 1997;48(1):119-125. doi:10.1212/WNL.48.1.119

198. Boeve BF, Maraganore DM, Parisi JE, et al. Pathologic heterogeneity in clinically diagnosed corticobasal degeneration. *Neurology*. 1999;53(4):795-800. doi:10.1212/WNL.53.4.795

199. Ling H, O'Sullivan SS, Holton JL, et al. Does corticobasal degeneration exist? A clinicopathological re-evaluation. *Brain*. 2010;133(7):2045-2057. doi:10.1093/brain/awq123

200. Mahoney CJ, Beck J, Rohrer JD, et al. Frontotemporal dementia with the C9ORF72 hexanucleotide repeat expansion: Clinical, neuroanatomical and neuropathological features. *Brain*. 2012;135(3):736-750. doi:10.1093/brain/awr361

201. Ran FA, Hsu PD, Wright J, Agarwala V, Scott DA, Zhang F. Genome engineering using the CRISPR-Cas9 system. *Nat Protoc*. 2013;8(11):2281-2308. doi:10.1038/nprot.2013.143

202. Maguire JA, Gagne AL, Jobaliya CD, Gandre-Babbe S, Gadue P, French DL. Generation of human control iPS cell line CHOPWT10 from healthy adult peripheral blood mononuclear cells. *Stem Cell Res*. 2016;16(2):338-341. doi:10.1016/j.scr.2016.02.017

203. Shi Y, Kirwan P, Livesey FJ. Directed differentiation of human pluripotent stem cells to cerebral cortex neurons and neural networks. *Nat Protoc*. 2012;7(10):1836-1846. doi:10.1038/nprot.2012.116

204. Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal*. 2011;17(1):10-12. https://doi.org/10.14806/ej.17.1.200.

205. Dobin A, Davis CA, Schlesinger F, et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*. 2013;29(1):15-21. doi:10.1093/bioinformatics/bts635

206. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol*. 2014;15(12):550. doi:10.1186/s13059-014-0550-8

207. Ho WY, Tai YK, Chang J-C, et al. The ALS-FTD-linked gene product, C9orf72, regulates neuronal morphogenesis via autophagy. *Autophagy*. 2019:1-16. doi:10.1080/15548627.2019.1569441

208. Beck J, Poulter M, Hensman D, et al. Large C9orf72 hexanucleotide repeat expansions are seen in multiple neurodegenerative syndromes and are more frequent than expected in the UK population. *Am J Hum Genet*. 2013;92(3):345-353. doi:10.1016/j.ajhg.2013.01.011

209. Bender R, Lange S. Adjusting for multiple testing - When and how? *J Clin Epidemiol*. 2001;54(4):343-349. doi:10.1016/S0895-4356(00)00314-0

210. Belzil V V., Bauer PO, Prudencio M, et al. Reduced C9orf72 gene expression in c9FTD/ALS is caused by histone trimethylation, an epigenetic event detectable in blood. *Acta Neuropathol*. 2013;126(6):895-905. doi:10.1007/s00401-013-1199-1

211. Lonsdale J, Thomas J, Salvatore M, et al. The Genotype-Tissue Expression (GTEx) project. *Nat Genet*. 2013;45(6):580-585. doi:10.1038/ng.2653

212. Mok K, Traynor BJ, Schymick J, et al. The chromosome 9 ALS and FTD locus is probably derived from a single founder. *Neurobiol Aging*. 2012;33(1):209.e3-209.e8.

doi:10.1016/j.neurobiolaging.2011.08.005

213. Chambers SM, Fasano CA, Papapetrou EP, Tomishima M, Sadelain M, Studer L. Highly efficient neural conversion of human ES and iPS cells by dual inhibition of SMAD signaling. *Nat Biotechnol*. 2009;27(3):275-280. doi:10.1038/nbt.1529

214. Yu C, Wang L, Lv B, et al. TMEM74, a lysosome and autophagosome protein, regulates autophagy. *Biochem Biophys Res Commun*. 2008;369(2):622-629. doi:10.1016/j.bbrc.2008.02.055

215. Sun Y, Chen Y, Zhang J, et al. TMEM74 promotes tumor cell survival by inducing autophagy via interactions with ATG16L1 and ATG9A. *Cell Death Dis*. 2017;8(8):e3031. doi:10.1038/cddis.2017.370

216. Antonelli M, Barilà D, Manic G, et al. ATM kinase sustains breast cancer stem-like cells by promoting ATG4C expression and autophagy. *Oncotarget*. 2017;8(13):21692-21709. doi:10.18632/oncotarget.15537

217. Wang N, Tan HY, Li S, Feng Y. Atg9b deficiency suppresses autophagy and potentiates endoplasmic reticulum stress-associated hepatocyte apoptosis in hepatocarcinogenesis. *Theranostics*. 2017;7(8):2325-2338. doi:10.7150/thno.18225

218. Ugolino J, Ji YJ, Conchina K, et al. Loss of C9orf72 Enhances Autophagic Activity via Deregulated mTOR and TFEB Signaling. *PLoS Genet*. 2016;12(11). doi:10.1371/journal.pgen.1006443

219. Nakatogawa H, Ichimura Y, Ohsumi Y. Atg8, a Ubiquitin-like Protein Required for Autophagosome Formation, Mediates Membrane Tethering and Hemifusion. *Cell*. 2007;130(1):165-178. doi:10.1016/j.cell.2007.05.021

220. Xie Z, Nair U, Klionsky DJ. Atg8 controls phagophore expansion during autophagosome formation. *Mol Biol Cell*. 2008;19(8):3290-3298. doi:10.1091/mbc.e07-12-1292

221. Tanida I, Minematsu-Ikeguchi N, Ueno T, Kominami E. Lysosomal Turnover, but Not a Cellular Level, of Endogenous LC3 is a Marker for Autophagy. *Autophagy*. 2005;1(2):84-91. doi:10.4161/auto.1.2.1697

222. Mauthe M, Orhon I, Rocchi C, et al. Chloroquine inhibits autophagic flux by decreasing autophagosome-lysosome fusion. *Autophagy*. 2018;14(8):1435-1455. doi:10.1080/15548627.2018.1474314

223. Zhou J, Tan SH, Nicolas V, et al. Activation of lysosomal function in the course of autophagy via mTORC1 suppression and autophagosome-lysosome fusion. *Cell Res*. 2013;23(4):508-523. doi:10.1038/cr.2013.11

224. Kouri N, Ross OA, Dombroski B, et al. Genome-wide association study of corticobasal degeneration identifies risk variants shared with progressive supranuclear palsy. *Nat Commun*. 2015;6:7247. doi:10.1038/ncomms8247

225. Snowden JS, Rollinson S, Thompson JC, et al. Distinct clinical and pathological characteristics of frontotemporal dementia associated with C9ORF72 mutations. *Brain*. 2012;135(3):693-708. doi:10.1093/brain/awr355

226. Hagerman RJ, Leehey M, Heinrichs W, et al. Intention tremor, parkinsonism, and generalized brain atrophy in male carriers of fragile X. *Neurology*. 2001;57(1):127-130.

doi:10.1212/WNL.57.1.127

227. Elden AC, Kim H, Hart MP, et al. Ataxin-2 intermediate-length polyglutamine expansions are associated with increased risk for ALS. *Nature*. 2011;466(7310):1069-1075. doi:10.1038/nature09320.Ataxin-2

228. Sanpei K, Takano H, Igarashi S, et al. Identification of the spinocerebellar ataxia type 2 gene using a direct identification of repeat expansion and cloning technique, DIRECT. *Nat Genet*. 1996;14(3):277-284. doi:10.1038/ng1196-277

229. Ciura S, Sellier C, Campanari ML, Charlet-Berguerand N, Kabashi E. The most prevalent genetic cause of ALS-FTD, C9orf72 synergizes the toxicity of ATXN2 intermediate polyglutamine repeats through the autophagy pathway. *Autophagy*. 2016;12(8):1406-1408. doi:10.1080/15548627.2016.1189070

230. Cencioni C, Spallotta F, Martelli F, et al. Oxidative stress and epigenetic regulation in ageing and age-related diseases. *Int J Mol Sci*. 2013;14(9):17643-17663. doi:10.3390/ijms140917643

231. Hervouet E, Peixoto P, Delage-Mourroux R, Boyer-Guittaut M, Cartron PF. Specific or not specific recruitment of DNMTs for DNA methylation, an epigenetic dilemma. *Clin Epigenetics*. 2018;10(1):17. doi:10.1186/s13148-018-0450-y

232. Verma P, Greenberg RA. Noncanonical views of homology-directed DNA repair. *Genes Dev*. 2016;30(10):1138-1154. doi:10.1101/gad.280545.116

233. Goate A, Chartier-Harlin MC, Mullan M, et al. Segregation of a Missense Mutation in the Amyloid Precursor Protein ... *Lett to Nat*. 1991;349(6311):704-706.

234. Citron M, Oltersdorf T, Haass C, et al. Mutation of the β-amyloid precursor protein in familial Alzheimer's disease increases β-protein production. *Nature*. 1992;360(6405):672-674. doi:10.1038/360672a0

235. Rogaev EI, Lukiw WJ, Lavrushina O, Rogaeva EA, St. George-Hyslop PH. The upstream promoter of the β-amyloid precursor protein gene (app) shows differential patterns of methylation in human brain. *Genomics*. 1994;22(2):340-347. doi:10.1006/geno.1994.1393

236. Barrachina M, Ferrer I. DNA methylation of Alzheimer disease and tauopathy-related genes in postmortem brain. *J Neuropathol Exp Neurol*. 2009;68(8):880-891. doi:10.1097/NEN.0b013e3181af2e46

237. Brohede J, Rinde M, Winblad B, Graff C. A DNA methylation study of the amyloid precursor protein gene in several brain regions from patients with familial alzheimer disease. *J Neurogenet*. 2010;24(4):179-181. doi:10.3109/01677063.2010.503978

238. Iwata A, Nagata K, Hatsuta H, et al. Altered CpG methylation in sporadic alzheimer's disease is associated with APP and MAPT dysregulation. *Hum Mol Genet*. 2014;23(3):648-656. doi:10.1093/hmg/ddt451

239. Hagerman RJ, Hagerman P. Fragile X-associated tremor/ataxia syndrome — features, mechanisms and management. *Nat Rev Neurol*. 2016;12(7):403-412. doi:10.1038/nrneurol.2016.82

240. Tassone F, Beilina A, Carosi C, et al. Elevated FMR1 mRNA in premutation carriers is due to increased transcription. *RNA*. 2007;13(4):555-562. doi:10.1261/rna.280807

241. Kenneson, A. Zhang, F. Hagedorn, CH. Warren S. Reduced FMRP and increased FMR1 transcription is proportionally associated with CGG repeat number in intermediate-length and premutation carriers. *Hum Mol Genet*. 2001;10(14):1449-1454. doi:10.1093/hmg/10.14.1449

242. Tassone F, Hagerman RJ, Chamberlain WD, Hagerman PJ. Transcription of the FMR1 gene in individuals with fragile X syndrome. *Am J Med Genet - Semin Med Genet*. 2000;97(3):195-203. doi:10.1002/1096-8628(200023)97:3<195::AID-AJMG1037>3.0.CO;2-R

243. Beilina A, Tassone F, Schwartz PH, Sahota P, Hagerman PJ. Redistribution of transcription start sites within the FMR1 promoter region with expansion of the downstream CGG-repeat element. *Hum Mol Genet*. 2004;13(5):543-549. doi:10.1093/hmg/ddh053

244. Armstrong MJ, Litvan I, Lang AE, et al. Criteria for the diagnosis of corticobasal degeneration. *Neurology*. 2013;80(5):496-503. doi:10.1212/WNL.0b013e31827f0fd1

245. Holmes BB, Furman JL, Mahan TE, et al. Proteopathic tau seeding predicts tauopathy in vivo. *Proc Natl Acad Sci*. 2014;111(41):E4376-E4385. doi:10.1073/pnas.1411649111