C9orf72 Intermediate Repeats Confer Risk for Corticobasal Degeneration and Increase C9orf72 Expression

Christopher P Cali[1], Maribel Patino[1], Jessica M Phan[1], PSP Genetics Consortium, Bernardino Ghetti[2], Vivianna M Van Deerlin[3], Virginia M-Y Lee[3], John Q Trojanowski[3], Kin Y Mok[4], Helen Ling[4], Dennis W Dickson[5], Gerard D Schellenberg[6], and Edward B Lee[1]

[1] Translational Neuropathology Research Laboratory, Department of Pathology and Laboratory Medicine, University of Pennsylvania, Philadelphia, PA, USA
[2] Department of Pathology and Laboratory Medicine, Indiana University School of Medicine, Indianapolis, IN, USA
[3] Center for Neurodegenerative Disease Research, Department of Pathology and Laboratory Medicine, University of Pennsylvania, Philadelphia, PA, USA
[4] Reta Lila Weston Research Laboratories and Department of Molecular Neuroscience, University College London Institute of Neurology, London, UK
[5] Department of Neuroscience, Mayo Clinic, Jacksonville, FL, USA
[6] Penn Neurodegeneration Genomics Center, 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@uphs.upenn.edu

Author Contributions: E.B.L. and conceived and designed this study. C.P.C performed RNA and protein expression experiments in patient brain and iPSCs. M.P. carried out screen for repeat size and risk allele genotype. J.M.P. performed tau biosensor cell experiments. B.G., V.M.V.D, V.M.-Y.L, J.Q.T, K.Y.M, H.L, D.W.D, G.D.S., E.B.L. and the PSP Genetics Consortium performed neuropathology analysis and provided DNA samples. C.P.C and E.B.L wrote the manuscript and all authors approved the final manuscript.

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 (>17) *C9orf72* repeats were significantly enriched in autopsy-proven CBD (n=355 cases, odds ratio=2.12, p-value=0.0094). 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. Finally, overexpression of C9orf72 protein without the repeat expansion in cells containing fluorescently tagged tau protein results in increased tau aggregation. These results raise the possibility that therapeutic strategies to reduce *C9orf72* expression may be beneficial for the treatment of tauopathies such as 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)[1–3]. ALS/FTD is characterized neuropathologically by the presence of TDP-43 inclusions which are tightly linked to neurodegeneration[4]. 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[1,2,5–7]. At the protein level, C9orf72 regulates endosomal trafficking[8–10] and is necessary for proper myeloid cell function in mice, but knockout of the protein does not cause neurodegeneration[11,12]. On the other hand, the presence of the repeat expansion is sufficient to induce neurodegeneration in several mouse models[13,14]. 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[15,16], 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[17,18]. In contrast, non-diseased individuals typically harbor repeats of 2-8 units[19]. Several studies have explored whether intermediate 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[19–27]. A significant association between intermediate expansion carriers and Parkinson's Disease (PD) was reported in clinically diagnosed PD[21], consistent with a large meta-analysis which found a small but significant increase in risk for clinically diagnosed PD with increased numbers of *C9orf72* repeats[28]. However, these associations were not found in an autopsy confirmed PD cohort[29],

raising the possibility that the association between *C9orf72* repeat number and clinically diagnosed PD may be due to contamination of clinical cohorts with individuals with diseases which mimic PD.

We considered here whether a clinically similar neurodegenerative disease, corticobasal degeneration (CBD), may underlie the observed associations between clinically diagnosed PD and *C9orf72* intermediate repeats. CBD is a rare neurodegenerative disease that shares some similar clinical features with PD such as limb rigidity, bradykinesia, tremor and apraxia[30]. 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[31]. 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[32]. Only ~25-50% of patients diagnosed with CBS are confirmed as CBD at autopsy[33–35], making clinically defined cohorts difficult to use for genetic studies of CBD due to underlying neuropathologic 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 yet assembled. In addition, we use post mortem CBD brain tissue to confirm the genetic findings and use CRISPR/cas9 to modify the repeat number in human cells at the endogenous locus. 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.

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 355 autopsy confirmed CBD cases was screened using a PCR-based sizing assay. As a control cohort, we used healthy controls from two previous studies of *C9orf72* repeat expansion length (n = 22,174)[28,36]. Using the previously defined 17 repeat units cutoff to define intermediate *C9orf72* expansions[28], we found that 1.97% of alleles from CBD cases (corresponding to 3.94% of the cases) harbored intermediate length repeat expansions (17-29 units) compared to 0.92% of controls (Figure 1, OR = 2.12, p = 0.00938). We did not detect any expansions larger than 29 repeats. These results indicate that intermediate expansions in *C9orf72* may be a genetic risk factor for CBD.

C9orf72 expression from intermediate expansion carriers

Next, we sought to determine how intermediate repeat expansions affect expression of the *C9orf72* gene. Previous studies have shown that large expansions in *C9orf72*, which cause ALS/FTD, can lead to a reduction of mRNA from the expanded allele[2,37,38]. This reduction in expression is often due to hypermethylation of the *C9orf72* promoter region[39–41]. 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[42] (Figure 2a, V2 and V3 isoforms). In CBD brain post mortem tissue, we again observed high frequencies of intermediate repeat expansions between 17 and 28 units (Supplemental Figure 1). Intermediate size expansions did not lead to hypermethylation of the

promoter region, as is observed in full expansion cases (Supplemental Figure 2). Surprisingly, total *C9orf72* mRNA levels were positively correlated with repeat expansion size (Figure 2b, $R^2 = 0.276$, $p = 0.0034$). This positive correlation was not observed for the major isoform, V2, which has a transcription start site downstream from the repeat expansion (Figure 2c, $R^2 = 0.00496$, $p = 0.717$). 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 2d, $R^2 = 0.667$, $p < 0.0001$). Cases with >12 total repeats exhibited a 2-5 fold increase in variant 3 mRNA than 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 2e, $R^2 = 0.849$, $p < 0.0001$).

To confirm the positive relationship between *C9orf72* repeat size and mRNA expression, we analyzed publically available SNP eQTL data from the Genotype-Tissue Expression (GTEx) Project[43]. Large *C9orf72* repeat expansion mutations invariably occur within a haplotype block with a defined series of SNPs[3,44]. The SNP rs3849942 is a commonly used marker for this haplotype block in which the G allele is found most frequently in non-expanded controls and the A allele is associated with both full and intermediate repeat expansions[2]. Indeed, in our CBD genomic DNA cohort, we observe an enrichment of intermediate repeats in cases with the rs3849942 A risk allele (Figure 2f). Using rs3849942 as a marker for 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 2g). Expanding our analysis to 20 haplotype defining SNPs[44], 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 2h). In contrast, there were no significant eQTL SNPs that led to a decrease in expression. These results provide additional evidence that intermediate expansions appear to increase *C9orf72* expression, even in non-disease populations.

CRISPR Editing of *C9orf72* expansion size in iPSCs

In order to confirm our findings from the CBD patient cohort and to establish a cellular model of intermediate repeat expansions, we used CRISPR/cas9 to modify the repeat size in previously characterized induced pluripotent stem cells (iPSCs) from a non-diseased individual[45]. We used a repair template containing either 2 or 28 repeats to alter the size of the repeat expansion in iPSCs that normally have alleles of either 2 or 6 repeats (Figure 3a-b). 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 3b). Genotype including repeat size was confirmed by repeat primed PCR (Figure 3c) and Sanger sequencing (Supplementary Figure 3). 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 to six *C9orf72* repeats.

The dual SMAD inhibition method[46,47] was used 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 4). 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 3d). Consistent with the data from CBD post mortem brain

tissue, we again saw no change in the amount of variant 2 mRNA (Figure 3e) but a robust increase in variant 3 mRNA levels in all intermediate *C9orf72* repeat cell lines (Figure f-g). Thus, these experiments demonstrate the causal relationship between intermediate *C9orf72* expansions and increased *C9orf72* mRNA levels.

Possible mechanisms of intermediate repeat mediated pathogenesis

Full expansion carriers have characteristic neuropathology in post mortem brain and spinal cord tissue that includes nuclear RNA foci of G_4C_2 transcripts[2] and aggregates of dipeptide repeat proteins (DPRs)[48]. Despite having elevated levels of *C9orf72* variant 3 mRNA, RNA foci were not detected in CBD brain tissues with intermediate repeats (repeat sizes 17-29, Figure 4a). In comparison, full expansion cases had RNA foci in ~18% of nuclei. We also stained for DPR aggregates (GA, GP and GR) and again found no evidence of DPR in CBD brain tissues with intermediate *C9orf72* repeats (Figure 4b). Thus, the major pathologic hallmarks of full expansion carriers are not present in CBD cases with intermediate repeats, 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.

To test whether C9orf72 protein expression is higher in CBD cases with intermediate repeat expansions, we extracted protein from frozen cerebellum from 8 low repeat control cases (4 total repeats) and 8 intermediate repeat CBD cases (19-31 total repeats). While we observed variability in C9orf72 protein levels between cases, there was a significant positive correlation between C9orf72 protein levels and increasing repeat number (Figure 4c-d; $R^2 = 0.328$, $p = 0.0204$). To confirm this finding, we tested protein expression in isogenic repeat edited NPCs, and observed a consistent ~40% increase in C9orf72 protein levels in cell lines with intermediate

repeats compared to isogenic low repeat number controls (Figure 4e-f). Thus, intermediate size repeat expansions increase both *C9orf72* mRNA and protein levels.

To better understand the potential mechanisms underlying CBD risk, we asked whether increased *C9orf72* expression alters tau aggregation. We utilized a previously developed[49] "tau biosensor cell" line that stably expresses the mutant tau repeat domain (P301S) fused to fluorescent reporters. When seeded with purified tau from post mortem brain, the tau reporters aggregate and release a FRET signal that is detected by flow cytometry (Figure 4g)[49]. We overexpressed C9orf72 protein in these cells using a construct devoid of the intronic G_4C_2 repeat domain and transduced them with tau from CBD post-mortem brain. Remarkably, we found that overexpression of C9orf72 protein caused an almost 2-fold increase in FRET signal (Figure 4h), indicating increased tau aggregation. This suggests that C9orf72 protein overexpression may be associated with increased propensity to form tau pathology.

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. Our findings supports the possibility that weak associations between clinically defined Parkinson's disease and intermediate *C9orf72* expansions may be due to small numbers of non-PD cases including CBD that were included in the PD cohorts. Interestingly, initial reports of *C9orf72* repeat mutation cohorts, defined at that time as at least 30 repeats, included one case of autopsy-confirmed CBD[50]. 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.

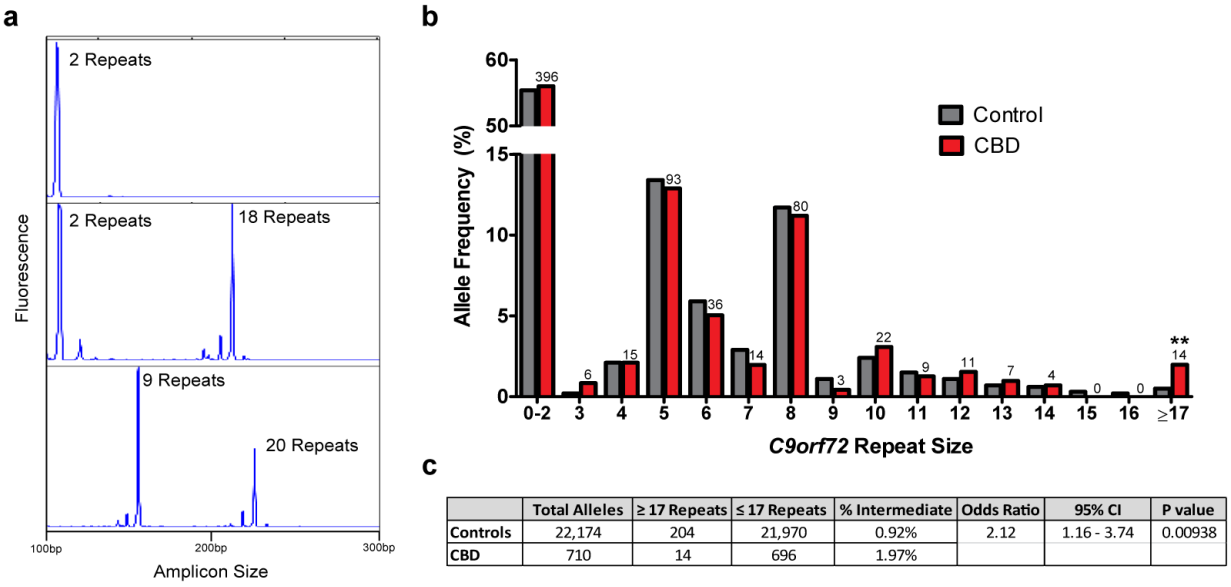
C9orf72 joins a list of other repeat expansion containing genes in which the length of the repeat confers a pleiotropic effect on the disease phenotype. For instance, intermediate length CGG repeats in the gene *FMR1* result in the neurodegenerative disease Fragile X-associated tremor/ataxia syndrome (FXTAS)[51], whereas full expansions led to the neurodevelopmental disorder Fragile X Syndrome[52]. Similarly, intermediate CAG expansions in *Ataxin-2* are a risk factor for ALS[53], whereas full expansions cause spinocerebellar ataxia type 2[54]. For *C9orf72*, large expansions cause ALS/FTD[1,2], whereas we demonstrate that intermediate expansions confer ~2 fold higher risk for CBD. Intermediate expansions in *C9orf72* seem to confer disease risk through a different mechanism than the full expansion. Here, we show that intermediate

expansions cause an increase in *C9orf72* mRNA and protein 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. This is in stark contrast to full *C9orf72* expansion carriers, which have reduced *C9orf72* mRNA and protein levels[2,7,42,55]. A similar observation has been made in FXTAS, where intermediate expansions increase FMR1 mRNA levels[56–59], but large expansions lead to epigenetic silencing of the gene[60]. Thus, it appears that maintaining proper levels of *C9orf72* is necessary for normal neuronal function, with either too little or too much gene product being associated with neurodegenerative disease risk.

To better understand how overexpression of *C9orf72* may confer risk for CBD, we overexpressed C9orf72 protein in a cellular model of tau aggregation. Overexpression of C9orf72 protein enhanced tau aggregation in this model, suggesting interplay between C9orf72 protein function and tau biology. This may be occurring through vesicular trafficking pathways that C9orf72 is thought to regulate. For instance, knockdown or knockout of *C9orf72* has been shown to reduce endocytosis[8], vesicle secretion[10] and intracellular trafficking[7,8,10,12]. Thus, overexpression in the context of tau aggregation may promote increased endocytosis of tau seeds, leading to increased intracellular tau accumulation. Additionally, upregulation of intracellular trafficking and secretion may allow for increased spread of tau seeds to nearby cells. Our results indicate that C9orf72 protein expression may affect tau aggregation and further studies in more physiologic model systems will be required to better define the cellular mechanisms which link intermediate *C9orf72* repeats and CBD.

Overall, our results illustrate the complexity of *C9orf72* repeat expansion mutations, with the length of expansion conferring different phenotypes via opposing mechanisms. The

association between intermediate expansions and CBD has implications for genetic counseling and should help to clarify the risk conferred by intermediate expansions in this gene. Finally, our findings indicate that strategies targeting *C9orf72* expression may be therapeutically beneficial for tauopathies such as CBD.



	Total Alleles	≥ 17 Repeats	≤ 17 Repeats	% Intermediate	Odds Ratio	95% CI	P value
Controls	22,174	204	21,970	0.92%	2.12	1.16 - 3.74	0.00938
CBD	710	14	696	1.97%			

Figure 1: Screening of *C9orf72* repeat size in CBD cases. A) Example capillary electrophoresis traces of CBD cases from the repeat sizing assay with genotypes of 2,2 (top), 2,18 (middle) or 9,20 (bottom). B) Allele frequencies of repeat sizes in CBD cases (n = 710 alleles) measured from genomic DNA using a fluorescent PCR assay and controls from Theuns et al. 2014[34] (n = 11,774 alleles). The number of alleles in the CBD cohort for each size is depicted above the bars. C) Summary statistics for repeat size screening of CBD cases and combined control cohort. Fisher's exact test: OR = 2.12, 95% CI = 1.16 - 3.74, p = 0.00938.

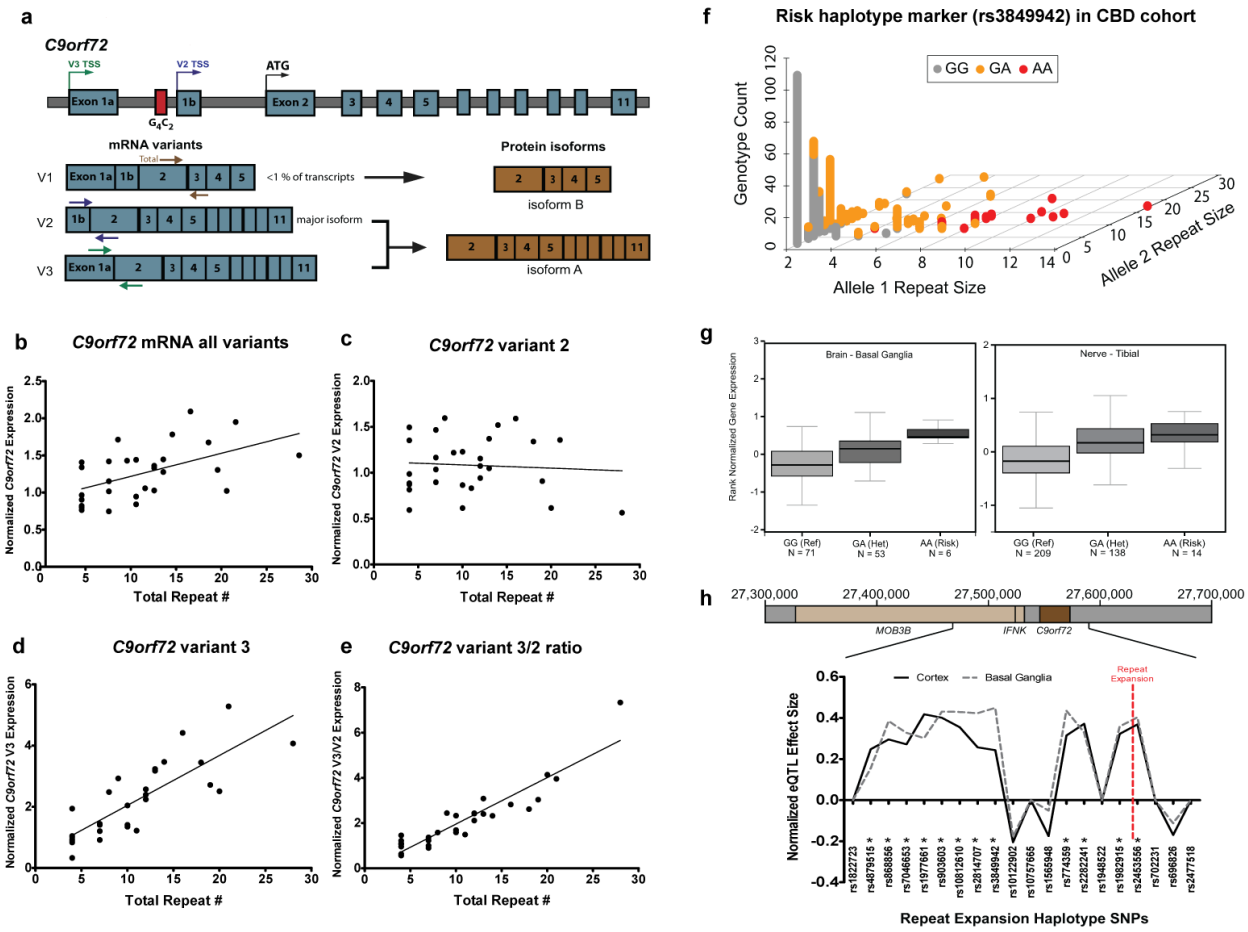


Figure 2: Effect of intermediate repeats on *C9orf72* expression. A) Diagram of the *C9orf72* locus with mRNA and protein isoforms. Transcription start sites (TSS) are labeled for each mRNA variant on the genomic DNA. Colored arrows on the mRNA depict primers used for RT-qPCR. B) Correlation between total *C9orf72* repeat size and *C9orf72* mRNA levels from all variants measured by RT-qPCR using RNA from CBD patient cerebellum (n = 29, R^2 = 0.276, p = 0.0034). Data is normalized to samples with 4 total repeats. C) Correlation between *C9orf72* repeat size and variant 2 mRNA levels (n = 29, R^2 = 0.00496, p = 0.717). D) Correlation between *C9orf72* repeat size and variant 3 mRNA levels (n = 29, R^2 = 0.667, p < 0.0001). E) Correlation between *C9orf72* repeat size and the variant 3 mRNA to variant 2 mRNA ratio (n = 29, R^2 = 0.849, p < 0.0001) F) Genotyping results for rs3849942 in the CBD patient genomic DNA cohort

(n = 710 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). 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[44]). * indicates SNPs with a significant effect on *C9orf72* expression.

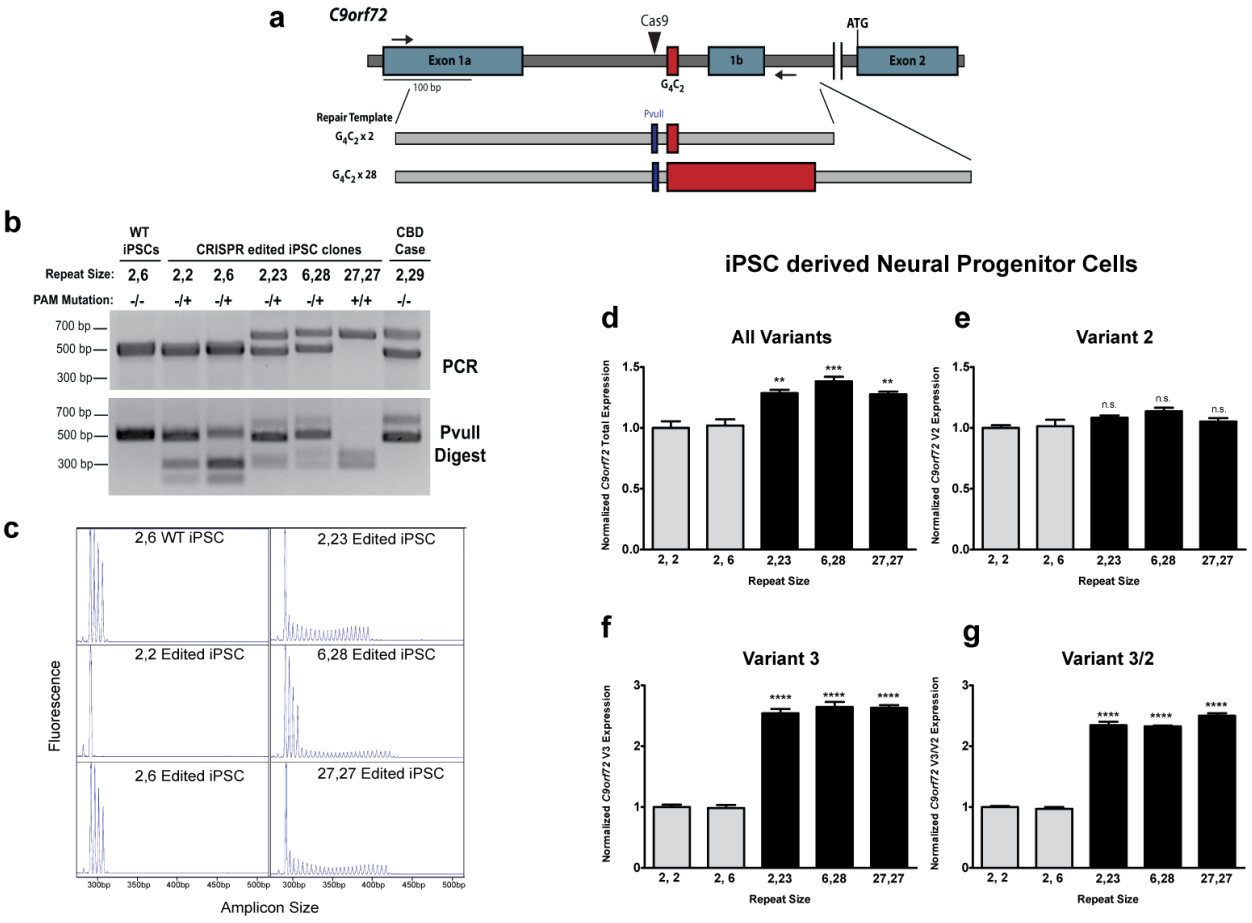


Figure 3: CRISPR/cas9 knock-in of intermediate repeats into iPSCs. A) Diagram of *C9orf72* genomic locus and double stranded repair templates used for knocking in intermediate repeats. Black triangle represents the CRISPR/cas9 cut site. Black arrows indicate the PCR primers used in (b). Red boxes indicate the G4C2 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 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.

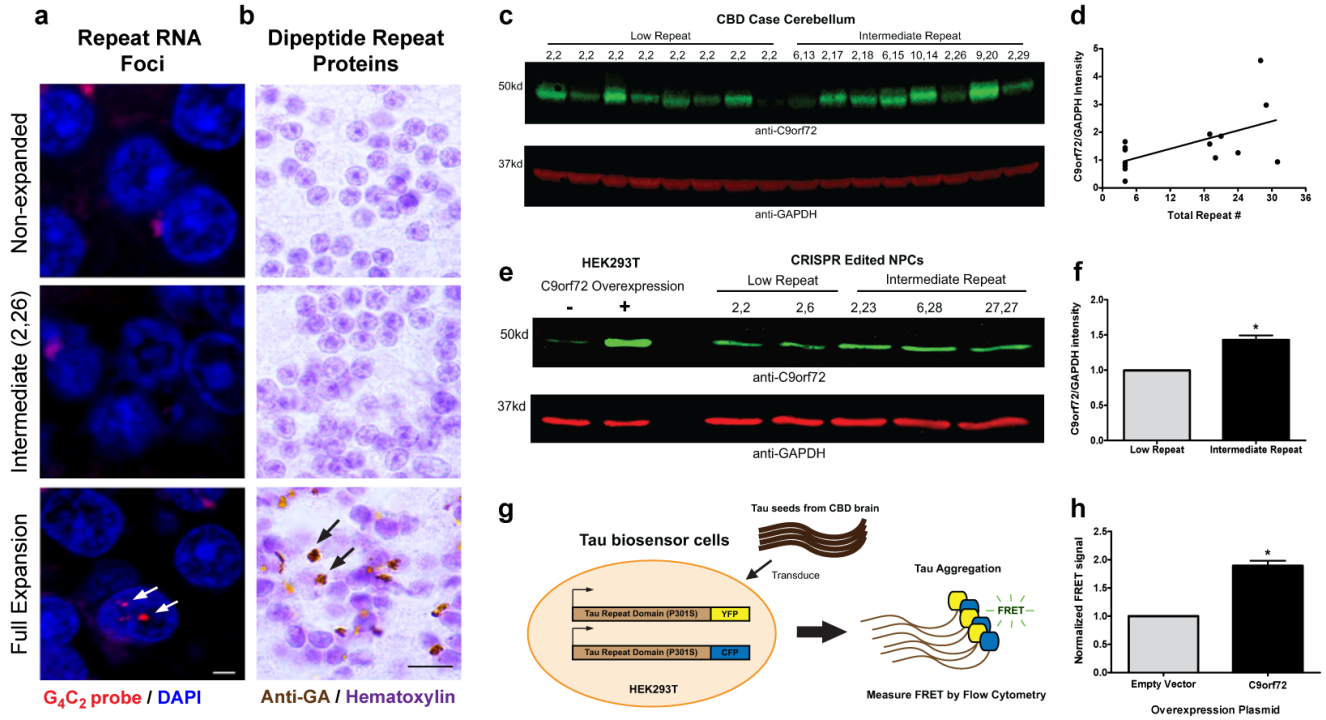


Figure 4: Mechanisms of intermediate expansion-mediated risk. A) Representative fluorescence in situ hybridization images of non-expanded (top, n = 4), intermediate expansion (center, n=5) and full expansion (bottom, n=5) carrier post mortem cerebellar samples using a probe complimentary to the G4C2 repeat. Scale bar 5µm. B) Representative immunohistochemistry images of non-expanded (top, n = 4), intermediate expansion (center, n=5) and full expansion (bottom, n=5) carrier post mortem cerebellar samples using an anti-GA antibody. Intermediate expansion carriers were also negative for GP and GR pathology. Scale bar 10µm. C) Western blot from CBD post mortem cerebellum samples. n=8 low repeat cases and 8 intermediate repeat cases. Representative blot of 3 technical replicates is shown. D) Quantification of C9orf72 protein expression vs total *C9orf72* repeat number (n=16; linear regression, $R^2 = 0.328$, p = 0.0204). Protein expression was normalized to GAPDH as a loading control and the average of 3 technical replicates is reported. E) Protein expression from CRISPR edited NPCs. N = 2 low

repeat and 3 intermediate repeat number cell lines. Representative blot of 2 technical replicates is shown. F) 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. G) Overview of tau-biosensor cells that stably express the tau repeat domain with the aggregation promoting P301S mutation coupled to either yellow fluorescent protein (YFP) or cyan fluorescent protein (CFP). Upon seeding with tau isolated from CBD post mortem brain, intracellular tau aggregates form, causing an increase in FRET signal. H) FRET signal measured by flow cytometry from tau biosensor cells transfected with either empty vector or *C9orf72* coding vector. N = 3 experiments; p = 0.010, paired t-test. Error bars are standard error of the mean.

Methods

Patient Samples

Genomic DNA for repeat expansion size screening was obtained from autopsy confirmed CBD cases as previously described[S1]. CBD patient cerebellum tissue was obtained from the University of Pennsylvania brain bank at the Center for Neurodegenerative Disease Research.

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[S2]. All primer sequences are listed in Supplemental Table 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 37^0 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 66^0 in hybridization buffer containing 50% Foramide, 2X SSC, 50mM NaHPO4 , and 10% Dextran sulfate in DEPC-treated water. Slides were hybridized with 40nM probe (LNA 5TYE563C4G2; Exiqon (Denmark) # 300500, batch # 621440) at 66^0 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 65^0. Slides were then rinsed with water, stained with DAPI and mounted with coverslips using Prolong glass antifade mountant (Invitrogen, Carlsbad, CA).

Detection of Dipeptide Repeat proteins and Measurement of DNA methylation

Paraffin embedded cerebellum sections were deparaffinized and stained with antibodies against GA (1:7500), GP (1:7500) or GR (1:100) as previously described[S3]. 4 non-expanded, 5 full expansion and 5 intermediate expansion carriers 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[S3].

RNA Expression Analysis from Postmortem Brain

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. 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 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[S4]. 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. PX458 and repair template vectors were co-transfected into WT iPSCs[S5] 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[S6]. 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.[S6] 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 Blot for C9orf72

For NPCs, cells 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 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). For CBD post-mortem tissue western blots, 150mg frozen cerebellum was lysed in 1mL 1X RIPA buffer with protease inhibitors and homogenized using an electric homogenizer (Omni TH). Lysates were then sonicated and spun in an ultracentrifuge (Beckman Optima Max-TL) at 55,000 rpm using the TLA55 rotor (Beckman, Brea, CA) for 30mins. Western blotting was performed as above using 100ug of total protein. For C9orf72 detection, monoclonal antibody 12E7, gift of Manuela Neumann (DZNE, Germany),[S7] was used at 1:2.5 dilution in TBST overnight at 4 degrees. IR dye secondary antibodies (Licor) were used to detect protein on a Licor Odyssey.

Overexpression experiments in Tau Biosensor cells

Tau RD P301S FRET Biosensor cells (ATCC CRL-3275) were transfected using Fugene 6 (Promega) with 2ug of HA-C9orf72 expressing plasmid (cloned from Addgene #74156 into pcDNA 5TO, Invitrogen) or empty vector. 24 hours later, cells were transduced with 100ng of tau lysate using Lipofectamine 2000 (Invitrogen) extracted from postmortem CBD patient brain tissue as described previously[S8]. Cells were analyzed by flow cytometry (LSR-II, BD Biosciences) 48 hours post transfection. Integrated FRET signal was determined by the product of the percent FRET positive cells and FRET median fluorescence intensity of FRET positive cells as described previously[S9]. FRET signal was normalized to cells transfected with empty vector.

References

1. Renton, A. E. *et al.* A hexanucleotide repeat expansion in C9ORF72 is the cause of chromosome 9p21-linked ALS-FTD. *Neuron* **72,** 257–268 (2011).
2. Dejesus-hernandez, M. *et al.* Expanded GGGGCC hexanucleotide repeat in non-coding region of C9ORF72 causes chromosome 9p-linked frontotemporal dementia and amyotrophic lateral sclerosis. *Neuron* **72,** 245–256 (2011).
3. Majounie, E. *et al.* Frequency of the C9orf72 hexanucleotide repeat expansion in patients with amyotrophic lateral sclerosis and frontotemporal dementia: A cross-sectional study. *Lancet Neurol.* **11,** 323–330 (2012).
4. Lee, E. B., Lee, V. M. Y. & Trojanowski, J. Q. Gains or losses: Molecular mechanisms of TDP43-mediated neurodegeneration. *Nat. Rev. Neurosci.* **13,** 38–50 (2012).
5. Sareen, D. *et al.* Targeting RNA foci in iPSC-derived motor neurons from ALS patients with a C9ORF72 repeat expansion. *Sci. Transl. Med.* **5,** 208ra149 (2013).
6. Lee, Y. B. *et al.* Hexanucleotide repeats in ALS/FTD form length-dependent RNA Foci, sequester RNA binding proteins, and are neurotoxic. *Cell Rep.* **5,** 1178–1186 (2013).
7. Shi, Y. *et al.* Haploinsufficiency leads to neurodegeneration in C9ORF72 ALS/FTD human induced motor neurons. *Nat. Med.* (2018). doi:10.1038/nm.4490
8. Farg, M. A. *et al.* C9ORF72, implicated in amytrophic lateral sclerosis and frontotemporal dementia, regulates endosomal trafficking. *Hum. Mol. Genet.* **23,** 3579–3595 (2014).
9. Webster, C. P. *et al.* The C9orf72 protein interacts with Rab1a and the ULK1 complex to regulate initiation of autophagy. *EMBO J.* **35,** 1656–1676 (2016).
10. Aoki, Y. *et al.* C9orf72 and RAB7L1 regulate vesicle trafficking in amyotrophic lateral sclerosis and frontotemporal dementia. *Brain* **140,** 887–897 (2017).

11. Burberry, A. *et al.* Loss-of-function mutations in the C9ORF72 mouse ortholog cause fatal autoimmune disease. *Sci. Transl. Med.* **8,** 93 (2016).

12. O'Rourke, J. G. *et al.* C9orf72 is required for proper macrophage and microglial function in mice. *Science (80-.).* **351,** 1324–1329 (2016).

13. 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, P. L. Neurodegeneration. C9ORF72 repeat expansions in mice cause TDP-43 pathology, neuronal loss, and behavioral deficits. - PubMed - NCBI. *Science (80-.).* **348,** 1151–4 (2015).

14. Liu, Y. *et al.* C9orf72 BAC Mouse Model with Motor Deficits and Neurodegenerative Features of ALS/FTD. *Neuron* **90,** 521–534 (2016).

15. Suh, E. R. *et al.* Semi-automated quantification of C9orf72 expansion size reveals inverse correlation between hexanucleotide repeat number and disease duration in frontotemporal degeneration. *Acta Neuropathol.* **130,** 363–372 (2015).

16. van Blitterswijk, M. *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.* **12,** 978–988 (2013).

17. Dobson-Stone, C. *et al.* C9ORF72 Repeat Expansion in Australian and Spanish Frontotemporal Dementia Patients. *PLoS One* **8,** (2013).

18. Gijselinck, I. *et al.* The C9orf72 repeat size correlates with onset age of disease, DNA methylation and transcriptional downregulation of the promoter. *Mol. Psychiatry* **21,** 1–13 (2015).

19. DeJesus-Hernandez, M. *et al.* Analysis of the C9orf72 repeat in Parkinson's disease, essential tremor and restless legs syndrome. *Park. Relat. Disord.* **19,** 198–201 (2013).

20. Xi, Z. *et al.* Investigation of C9orf72 in 4 Neurodegenerative Disorders. *Arch. Neurol.* **69,** 1583 (2012).

21. Nuytemans, K. *et al.* C9orf72 intermediate repeat copies are a significant risk factor for parkinson disease. *Ann. Hum. Genet.* **77,** 351–363 (2013).

22. Harms, M. B. *et al.* Parkinson disease is not associated with C9ORF72 repeat expansions. *Neurobiol. Aging* **34,** 1519 (2013).

23. van der Zee, J. *et al.* A Pan-European Study of the C9orf72 Repeat Associated with FTLD: Geographic Prevalence, Genomic Instability, and Intermediate Repeats. *Hum. Mutat.* **34,** 363–373 (2013).

24. Yeh, T. H. *et al.* Screening for C9orf72 repeat expansions in parkinsonian syndromes. *Neurobiol. Aging* **34,** 1311.e3-1311.e4 (2013).

25. Lesage, S. *et al.* C9orf72 repeat expansions are a rare genetic cause of parkinsonism. *Brain* **136,** 385–391 (2013).

26. Jiao, B. *et al.* C9orf72 mutation is rare in Alzheimer's disease, Parkinson's disease, and essential tremor in China. *Front. Cell. Neurosci.* **7,** 164 (2013).

27. Akimoto, C. *et al.* No GGGGCC-hexanucleotide repeat expansion in C9ORF72 in parkinsonism patients in Sweden. *Amyotroph. Lateral Scler. Front. Degener.* **14,** 26–29 (2013).

28. Theuns, J. *et al.* Global investigation and meta-analysis of the C9orf72 (G4C2)n repeat in Parkinson disease. *Neurology* **83,** 1906–1913 (2014).

29. Nuytemans, K. *et al.* Absence of C9ORF72 expanded or intermediate repeats in autopsy-

confirmed Parkinson's disease. *Mov. Disord.* **29,** 827–830 (2014).

30. Armstrong, M. J. *et al.* Criteria for the diagnosis of corticobasal degeneration. *Neurology* **80,** 496–503 (2013).

31. Kouri, N. *et al.* Neuropathological features of corticobasal degeneration presenting as corticobasal syndrome or Richardson syndrome. *Brain* **134,** 3264–3275 (2011).

32. Dickson, D. W. *et al.* Office of rare diseases neuropathologic criteria for corticobasal degeneration. *J. Neuropathol. Exp. Neurol.* **61,** 935–946 (2002).

33. Litvan, I. *et al.* Accuracy of the clinical diagnosis of corticobasal degeneration: a clinicopathologic study. *Neurology* **48,** 119–25 (1997).

34. Boeve, B. F. *et al.* Pathologic heterogeneity in clinically diagnosed corticobasal degeneration. *Neurology* **53,** 795–800 (1999).

35. Ling, H. *et al.* Does corticobasal degeneration exist? A clinicopathological re-evaluation. *Brain* **133,** 2045–2057 (2010).

36. Beck, J. *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.* **92,** 345–353 (2013).

37. Belzil, V. V. *et al.* Reduced C9orf72 gene expression in c9FTD/ALS is caused by histone trimethylation, an epigenetic event detectable in blood. *Acta Neuropathol.* **126,** 895–905 (2013).

38. Waite, A. J. *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* **35,** 1779.e5-1779.e13 (2014).

39. Xi, Z. *et al.* Hypermethylation of the CpG island near the G4C2 repeat in ALS with a

C9orf72 expansion. *Am. J. Hum. Genet.* **92,** 981–989 (2013).

40. Liu, E. Y. *et al.* C9orf72 hypermethylation protects against repeat expansion-associated pathology in ALS/FTD. *Acta Neuropathol.* **128,** 525–541 (2014).

41. McMillan, C. T. *et al.* C9orf72 promoter hypermethylation is neuroprotective: Neuroimaging and neuropathologic evidence. *Neurology* **84,** 1622–1630 (2015).

42. Viodé, 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.* **12,** 1–11 (2018).

43. Lonsdale, J. *et al.* The Genotype-Tissue Expression (GTEx) project. *Nat. Genet.* **45,** 580–585 (2013).

44. Mok, K. *et al.* The chromosome 9 ALS and FTD locus is probably derived from a single founder. *Neurobiol. Aging* **33,** 209.e3-209.e8 (2012).

45. Maguire, J. A. *et al.* Generation of human control iPS cell line CHOPWT10 from healthy adult peripheral blood mononuclear cells. *Stem Cell Res.* **16,** 338–341 (2016).

46. Chambers, S. M. *et al.* Supplement: Highly efficient neural conversion of human ES and iPS cells by dual inhibition of SMAD signaling. *Nat. Biotechnol.* **27,** 275–280 (2009).

47. Shi, Y., Kirwan, P. & Livesey, F. J. Directed differentiation of human pluripotent stem cells to cerebral cortex neurons and neural networks. *Nat. Protoc.* **7,** 1836–1846 (2012).

48. Ash, P. E. A. *et al.* Unconventional Translation of C9ORF72 GGGGCC Expansion Generates Insoluble Polypeptides Specific to c9FTD/ALS. *Neuron* **77,** 639–646 (2013).

49. Holmes, B. B. *et al.* Proteopathic tau seeding predicts tauopathy in vivo. *Proc. Natl. Acad. Sci.* **111,** E4376–E4385 (2014).

50. Snowden, J. S. *et al.* Distinct clinical and pathological characteristics of frontotemporal

dementia associated with C9ORF72 mutations. *Brain* **135,** 693–708 (2012).

51. Hagerman, R. J. *et al.* Intention tremor, parkinsonism, and generalized brain atrophy in male carriers of fragile X. *Neurology* **57,** 127–130 (2001).

52. Verkerk, A. J. M. H. *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* **65,** 905–914 (1991).

53. Elden, A. C. *et al.* Ataxin-2 intermediate-length polyglutamine expansions are associated with increased risk for ALS. *Nature* **466,** 1069–1075 (2011).

54. Sanpei, K. *et al.* Identification of the spinocerebellar ataxia type 2 gene using a direct identification of repeat expansion and cloning technique, DIRECT. *Nat. Genet.* **14,** 277–284 (1996).

55. Frick, P. *et al.* Novel antibodies reveal presynaptic localization of C9orf72 protein and reduced protein levels in C9orf72 mutation carriers. *Acta Neuropathol. Commun.* **6,** 1–17 (2018).

56. Tassone, F., Hagerman, R. J., Chamberlain, W. D. & Hagerman, P. J. Transcription of the FMR1 gene in individuals with fragile X syndrome. *Am. J. Med. Genet. - Semin. Med. Genet.* **97,** 195–203 (2000).

57. 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.* **10,** 1449–1454 (2001).

58. Allen, E. G., He, W., Yadav-Shah, M. & Sherman, S. L. A study of the distributional characteristics of FMR1 transcript levels in 238 individuals. *Hum. Genet.* **114,** 439–447 (2004).

59. Tassone, F. *et al.* Elevated FMR1 mRNA in premutation carriers is due to increased transcription. *Rna-a Publ. Rna Soc.* **13,** 555–562 (2007).

60. Pieretti, M. *et al.* Absence of expression of the FMR-1 gene in fragile X syndrome. *Cell* **66,** 817–822 (1991).

References for Methods Section

S1. Kouri, N. et al. Genome-wide association study of corticobasal degeneration identifies risk variants shared with progressive supranuclear palsy. Nat. Commun. 6, 7247 (2015).

S2. Dejesus-hernandez, M. et al. Expanded GGGGCC hexanucleotide repeat in non-coding region of C9ORF72 causes chromosome 9p-linked frontotemporal dementia and amyotrophic lateral sclerosis. Neuron 72, 245–256 (2011).

S3. Liu, E. Y. et al. C9orf72 hypermethylation protects against repeat expansion-associated pathology in ALS/FTD. Acta Neuropathol. 128, 525–541 (2014).

S4. Ran, F. A. et al. Genome engineering using the CRISPR-Cas9 system. Nat. Protoc. 8, 2281–2308 (2013).

S5. Maguire, J. A. et al. Generation of human control iPS cell line CHOPWT10 from healthy adult peripheral blood mononuclear cells. Stem Cell Res. 16, 338–341 (2016).

S6. Shi, Y., Kirwan, P. & Livesey, F. J. Directed differentiation of human pluripotent stem cells to cerebral cortex neurons and neural networks. Nat. Protoc. 7, 1836–1846

(2012).

S7. Frick, P. et al. Novel antibodies reveal presynaptic localization of C9orf72 protein and reduced protein levels in C9orf72 mutation carriers. Acta Neuropathol. Commun. 6, 1–17 (2018).

S8. Narasimhan, S. et al. Pathological tau strains from human brains recapitulate the diversity of tauopathies in non-transgenic mouse brain. J. Neurosci. 37, 1230–17 (2017).

S9. Holmes, B. B. et al. Proteopathic tau seeding predicts tauopathy in vivo. Proc. Natl. Acad. Sci. 111, E4376–E4385 (2014).