Published in final edited form as:

Nat Rev Neurosci. 2016 June ; 17(6): 383–395. doi:10.1038/nrn.2016.38.

The expanding biology of the *C9orf72* nucleotide repeat expansion in neurodegenerative disease

Aaron R. Haeusler^{1,*}, Christopher J. Donnelly^{2,3,*}, Jeffrey D. Rothstein^{4,5}

¹Department of Biochemistry and Molecular Biology, Bloomberg School of Public Health, Johns Hopkins University, Baltimore, Maryland 21205, USA.

²Department of Neurobiology, School of Medicine, University of Pittsburgh, Pittsburgh, Pennsylvania 15260, USA.

³Live Like Lou Center for ALS Research, Brain Institute, University of Pittsburgh, Pittsburgh, Pennsylvania 15261, USA.

⁴Brain Science Institute, School of Medicine, Johns Hopkins University, Baltimore, Maryland 21205, USA.

⁵Departments of Neurology and Neuroscience, School of Medicine, Johns Hopkins University, Baltimore, Maryland 21205, USA.

Abstract

A nucleotide repeat expansion (NRE) within the chromosome 9 open reading frame 72 (*C9orf72*) gene was the first of this type of mutation to be linked to multiple neurological conditions, including amyotrophic lateral sclerosis and frontotemporal dementia. The pathogenic mechanisms through which the *C9orf72* NRE contributes to these disorders include loss of *C9orf72* function and gain-of-function mechanisms of *C9orf72* driven by toxic RNA and protein species encoded by the NRE. These mechanisms have been linked to several cellular defects — including nucleocytoplasmic trafficking deficits and nuclear stress — that have been observed in both patients and animal models.

A GGGGCC (G₄C₂) nucleotide repeat expansion (NRE) in the first intron of the chromosome 9 open reading frame 72 (*C9orf72*) gene represents the most common genetic abnormality in familial amyotrophic lateral sclerosis (ALS) and in frontotemporal dementia (FTD)^{1,2}; this NRE has also been detected in 5–20% of patients with sporadic ALS³ (BOX 1). The actual size of the NRE varies dramatically among patients with ALS and in patients with FTD carrying the mutation (here termed C9 ALS/FTD). The NRE can contain as many as hundreds or thousands of repeats; however, the presence of less than 30 repeats is

Correspondence to C.J.D. and J.D.R. chrisdonnelly@pitt.edu; jrothstein@jhmi.edu.

*These authors contributed equally to this work.

Competing interests statement

The authors declare no competing interests.

SUPPLEMENTARY INFORMATION

See online article: [S1 \(table\)](#) | [S2 \(table\)](#)

ALL LINKS ARE ACTIVE IN THE ONLINE PDF

generally not associated with disease. To date, the size of the expansion does not appear to be related to differences in disease parameters, such as severity of disease or age of disease onset. Because this mutation is so commonly detected in both ALS and FTD and in both familial presentations and apparent sporadic presentations of these disorders, understanding the underlying pathophysiology that it causes is essential for both disease management and the development of future biomarkers and drug therapies.

C9orf72 encodes a protein that can be expressed as two (long and short) isoforms (FIG. 1). Initial protein bioinformatics approaches had indicated that *C9orf72* lacked structural homology with any known proteins; however, recent analyses have suggested that *C9orf72* contains a DENN (differentially expressed in normal and neoplastic cells) domain^{4,5}, which may allow it to function as a GDP–GTP nucleotide exchange factor (GEF)^{6,7}. Experiments in transgenic mice and zebrafish have demonstrated that *C9orf72* is expressed at high levels in neurons^{8,9}, and the human proteome map indicates that *C9orf72* is highly expressed in the brain¹⁰. A recent immunocytochemistry study suggests that the short isoform of *C9orf72* localizes to the nuclear membrane, whereas the long isoform is predominantly cytoplasmic¹¹. The biological significance of this subcellular distribution is unclear, particularly as the cellular function of *C9orf72* remains largely unknown. However, there is evidence to indicate that *C9orf72* may participate in cellular trafficking¹².

There are three primary mechanisms through which the *C9orf72* NRE has been proposed to contribute to neural injury: through diminished *C9orf72* protein levels and consequent loss of *C9orf72* function, through the generation of expanded toxic RNA species, and through the accumulation of dipeptide repeat proteins (DPRs) encoded by the NRE.

In this Review, we explore the current literature supporting these three proposed drivers of *C9orf72* NRE-linked disease. We present each of the possible disease mechanisms and describe the cellular defects that have been observed in patient cells and tissues and in transgenic *C9orf72* NRE models. We describe the overlapping pathways that have been identified as being altered in disease and then briefly discuss therapeutic strategies being explored to offset the cellular dysfunctions that are linked to the *C9orf72* NRE mutation.

C9orf72 loss of function

A loss-of-function mechanism in C9 ALS/FTD was first suggested by the observation that there are reduced levels of all three *C9orf72* mRNA variants in patient tissue samples^{2,13-17}. Although the transcript levels were highly variable among patients in these studies, the findings were consistent with early immunostaining studies that showed decreased nuclear levels of *C9orf72* in patient fibroblasts¹. Staining with newer and more reliable isoform-specific antibodies confirmed a reduction in the levels of the long isoform of *C9orf72*, but suggested that there was an increase in levels of the short isoform in patient cortex samples¹¹. Notably, the short isoform of *C9orf72* re-localized to the plasma membrane in motor neurons of patients with C9 ALS (although no significant difference was observed in the motor cortex)¹¹. The biological significance of this re-localization remains unclear. However, there is evidence to indicate that *C9orf72* may be involved in RAB-GTPase-mediated cellular trafficking¹². Moreover, the interaction of *C9orf72* with nuclear pore

complex (NPC) components is inversely correlated with levels of TAR DNA-binding protein 43 (TDP43) cytoplasmic inclusions in human tissue¹¹. Together, these findings provide a glimpse into the possible role of *C9orf72* loss-of-function mutations in neurodegeneration.

Mechanisms driving *C9orf72* loss of function.

Several mechanisms, including epigenetic silencing and transcriptional instability, might lead to the transcriptional silencing of *C9orf72* (REFS 16,18,19) (BOX 2). The presence of the NRE leads to increased local epigenetic modifications that cause a promoter region located upstream of and distal to the NRE (responsible for driving the production of the transcript variants V1 and V3) to contain fewer transcriptional start sites. Furthermore, the NRE results in increased numbers of transcriptional start sites immediately upstream of the NRE²⁰ (BOX 2; FIG. 1). The promoter region that is located within or downstream of the NRE (and that drives the production of transcript V2; FIG. 1) also contains fewer transcriptional start sites²⁰. An increase in repressive DNA and histone epigenetic mechanisms initiated by the NRE (BOX 2) underlies these alterations and changes the transcriptional landscape of the mutant allele. This process leads to the production of truncated and/or divergent transcripts^{18,20-22}.

The structurally polymorphic nature of the *C9orf72* NRE (BOX 2) may also directly impair polymerase processivity, causing RNA polymerases to stall, slip or release during transcriptional initiation or elongation. This impaired polymerase processivity leads to reduced transcript levels and/or aborted or truncated transcripts^{18,21,23-27}. *In vivo*, additional transcriptional elongation factors can rescue or dismantle stalled polymerase complexes²⁸. However, the need to maintain transcriptional fidelity across the NRE may sequester these essential proteins, thus increasing transcriptional and genomic instability.

Bidirectional transcription also arises as a result of the *C9orf72* NRE and is common to many NRE-linked diseases²⁹⁻³¹. Although the mechanisms underlying the bidirectional production of repeat-containing transcripts are poorly understood, it is known that bidirectional transcription occurs at most CpG islands throughout the genome and can produce varying lengths of RNA transcripts^{32,33}. Thus, a GC-rich sequence, such as the *C9orf72* NRE, may produce short abortive transcripts through bidirectional transcription. These short transcripts could act as siRNAs and drive gene repression^{34,35}, as described for other NREs^{36,37}. One could also speculate that such short abortive transcripts could hijack the molecular machinery through which microRNAs function, modulating the processing of RNAs containing sequences that are complementary to the NRE^{38,39}.

C9orf72 loss of function and neurodegeneration.

It is unclear whether the NRE-driven reduction in *C9orf72* expression levels causes neurodegeneration. In zebrafish, *c9orf72* knockdown resulted in severe axonal and behavioural deficits that were rescued by the expression of exogenous human *C9orf72* (REF. 8). Similarly, deletion of the *Caenorhabditis elegans* *C9orf72* orthologue *alfa-1* resulted in age-dependent motility deficits⁴⁰. However, knockdown of *C9orf72* in rat primary cortical and motor neurons did not affect neuronal survival over a 2-week period⁴¹. Similarly, adult mice treated with *C9orf72*-targeting antisense oligonucleotides (ASOs) exhibited reduced

C9orf72 RNA levels in the brain and spinal cord but no neuropathology or behavioural deficits⁴². In mice studied up to 24 months of age, *C9orf72* ablation in neurons and glia did not result in any motor neuron dysfunction, pathology or altered survival⁴³. The differences between findings in mice and other species might be explained by the high degree of homology between human *C9orf72* and its mouse orthologue (98%) when compared with its homology to zebrafish *c9orf72* (76%) or *C. elegans alfa-1* (50%). Mouse and human *C9orf72* protein may exhibit similar endogenous function owing to their high protein homology compared to other species; thus, loss-of-function studies in mice might be more relevant to the human disease condition.

The idea that loss of *C9orf72* function drives disease pathology was also challenged by studies of patients with C9 ALS/FTD who exhibit >50% reduction of *C9orf72* RNA levels but exhibit a phenotype similar to patients with C9 ALS/FTD who have higher *C9orf72* RNA levels^{44,45}. However, recent work that used mouse models carrying a bacterial artificial chromosome (BAC) that expresses human *C9orf72* containing the NRE provided support for a possible *C9orf72* loss-of-function mechanism. In these studies, several gain-of-function mechanisms and additional pathological features associated with disease were reproduced (see below). However, the mice showed no neurodegenerative phenotypes, suggesting that gain-of-function mechanisms alone cannot explain disease pathogenesis^{22,46}. This result hints that *C9orf72* loss of function contributes to neurodegeneration.

C9orf72 gain-of-function mechanisms

A considerable body of *in vivo* and *in vitro* evidence suggests a gain-of-function cellular mechanism as a cause of *C9orf72* NRE-mediated neurodegeneration; however, it remains unclear which of the proposed gains in function is the most important.

Sense and antisense NRE RNA foci.

Repeat expansion mutations generate expanded RNA sequences that have been linked to disease pathology. Evidence for their role in C9 ALS/FTD derives from the presence of RNA aggregates containing the (G₄C₂) repeat encoded by the NRE in nuclei within the spinal cord and cortex of patients with C9 FTD². These RNA foci were subsequently observed in the cerebellum, motor cortex, frontal cortex, hippocampus, lumbar motor neurons, interneurons and glia of patients with C9 ALS/FTD^{2,13,30,42,47-50}. (G₄C₂) RNA foci are also found in motor and cortical neurons generated from induced pluripotent stem cells (iPSCs) derived from patients with C9 ALS/FTD^{13,14,20,42}. In both patient tissue and iPSC-derived neurons the foci are predominantly nuclear^{13,50}.

The bidirectional transcription of the *C9orf72* NRE^{30,31} also results in the generation of RNA aggregates. RNA foci containing the (C₄G₂) repeat encoded by the antisense transcript of the NRE are observed in the frontal cortex, Purkinje cells, cerebellum, hippocampus, spinal cord motor neurons, astrocytes, microglia and oligodendrocytes of patients with C9 ALS/FTD^{30,42,48-50}, where they are mainly nuclear⁵⁰. (G₄C₂) and (C₄G₂) RNA foci can be detected in the same nuclei and have been reported to colocalize^{49,50}. One study showed that (G₄C₂) and (C₄G₂) foci appear in higher numbers in Purkinje neurons and motor neurons than in granule neurons and hippocampal CA4 subfield and dentate gyrus neurons⁴⁹.

However, another study found that (G₄C₂) RNA foci predominantly exist in cerebellar granule neurons, whereas (C₄G₂) RNA foci are found more frequently in cerebellar Purkinje neurons and motor neurons⁴⁹. It is important to note that a direct comparison of the levels of sense and antisense RNA foci using RNA fluorescence *in situ* hybridization is impossible unless alternative quantitative methods are used.

The appearance of *C9orf72* NRE RNA foci sometimes correlates with pathological disease features. Age of disease onset correlated with the number of (G₄C₂) and (C₄G₂) RNA foci in a small number of patients with C9 FTD⁵⁰. Furthermore, 30–60% of (G₄C₂) or (C₄G₂) RNA foci-containing cells in the frontal cortex and hippocampus showed TDP43 mislocalization. However, less than 20% of these cells showed p62 mislocalization, calling into question the correlation between RNA foci and pathological inclusions⁵⁰. Another study showed RNA foci and p62 inclusions but no TDP43 mislocalization in an asymptomatic *C9orf72* NRE carrier⁵¹. Thus, although it is tempting to speculate that RNA foci are toxic to cells, there is limited evidence to indicate that the RNA foci themselves cause cellular dysfunction. This conclusion is supported by the observations of RNA foci in the absence of neurodegenerative phenotypes in *C9orf72* BAC mice models^{22,46}. This result indicates that RNA gain-of-function mechanisms probably require an additional component to manifest neurodegenerative phenotypes. As outlined below, the sequestration of RNA-binding proteins (RBPs) by the expanded RNA within NRE RNA foci can disrupt various molecular cascades⁵² (BOX 2; FIG. 2; Supplementary information S1 (table)) and is therefore likely to contribute to disease pathogenesis.

Protein sequestration.

Protein sequestration by RNA is an established mechanism of toxicity in some NRE disorders. For example, the myotonic dystrophy (*DM1*) NRE results in the transcription of CUG repeats, which sequester the muscleblind-like protein 1 (MBNL1) protein and lead to MBNL1 loss of function⁵³.

To determine the RBPs that may be sequestered by RNA encoded by the *C9orf72* NRE, several laboratories have used various methods to enable an unbiased identification of the ribonucleoproteins that interact with the transcribed *C9orf72* NRE in different tissues^{13,18,54–56}. Many unique *C9orf72*•RBP interactomes have been identified (FIG. 3; Supplementary information S2 (table)). Comparisons among these studies identified heterogeneous nuclear ribonucleoproteins (hnRNPs) as an overrepresented protein family that bind to the *C9orf72* NRE RNA. Proteins in this family contain an RNA recognition motif that is shared with several other proteins, including TDP43 and fused in sarcoma (FUS; also known as TLS). In particular, isoforms of hnRNP H can bind to the (G₄C₂) RNA, which may lead to globally increased alternative splicing of intron exon cassettes in the motor cortex and cerebellum of patients with C9 ALS/FTD⁵⁷.

Many additional proteins, including Purα, adenosine deaminase RNA-specific, B2 (ADARB2), nucleolin, hnRNP A1, hnRNP A2/B1, hnRNP A3, hnRNP H, ALYREF and RanGAP1, colocalize with *C9orf72* (G₄C₂) RNA foci^{13,18,20,49,54,56,58,59}. However, only a handful of these proteins have been further studied, and their biological relevance has not been validated. It is also important to note that the *C9orf72*•RBP interactome is largely

dependent on the underlying structures formed by the sense *C9orf72* RNA, which may include either G-quadruplex or hairpin structures. For example, nucleolin and RanGAP1 preferentially bind to the (G₄C₂) *C9orf72* RNA G-quadruplex^{18,59}, whereas hnRNP H isoforms bind to the RNA G-quadruplex or hairpin structure with similar affinities¹⁸. Thus, the structural variability of the (G₄C₂) RNA may affect divergent RNA processing pathways and have a critical role in a multihit RNA gain-of-function model in which there are broad RNA processing alterations.

DPR pathology.

Repeat-containing RNAs that escape the nucleus can form cytoplasmic RNA foci and can also be translated into DPRs through a non-canonical translational mechanism known as repeat-associated non-ATG-dependent translation (RANT)⁶⁰ (BOX 3). RANT has been identified *in vitro* and *in vivo* in many neurological diseases, including spinocerebellar ataxia type 8 (REF. 60), myotonic dystrophy type 1 (REF. 60), fragile X syndrome⁶¹, Huntington disease and C9 ALS/FTD^{13,14,30,31,48,62-65}. RANT has also been proposed to be an important contributor to NRE-linked disease progression⁶⁶.

The sense (G₄C₂) RNA NRE of *C9orf72* encodes three DPRs, glycine–proline (GP), glycine–alanine (GA) and glycine–arginine (GR), whereas the antisense (C₄G₂) RNA encodes proline–glycine (PG), proline–arginine (PR) and proline–alanine (PA). Of these, (GP (or PG)) is the most abundant in C9 ALS/FTD neurons, perhaps because it can be produced from both the sense- and the antisense-expanded RNA³⁰. However, (PA), (PR) and (PG) DPRs have been identified in the hippocampus, cerebellum, amygdala, medulla, thalamus, frontal cortex, spinal cord and motor cortex of patients with C9 ALS/FTD^{30,48}, with the hippocampus and cerebellum (granule cells) showing the highest levels of DPRs. C9 ALS/FTD spinal cord and brainstem tissue contains the lowest DPR levels^{30,67}. Notably, (GP), (GA) and (GR) DPRs colocalize with p62 but do not correlate with TDP43 mislocalization or neurodegeneration in the neocortex, spinal cord or cerebellum⁶⁴. (GP) DPRs were also detected in C9 ALS/FTD iPSC-derived cortical and motor neuron cultures^{13,14}, and the localization of endogenous C9 ALS/FTD DPRs is predominantly cytoplasmic^{13,30,31,48,63,64,68}. Interestingly, for most cells in *C9orf72* NRE patient tissue, RNA foci and RANT products were mutually exclusive⁴⁸.

The localization of these DPRs to regions such as the cerebellum and lateral geniculate nucleus is confusing because these brain regions are not associated with obvious clinical or pathological aspects of ALS or FTD. There are changes in the RNA profile of the cerebellum of patients with C9 ALS/FTD, the significance of which is currently unclear⁵⁷. However, cerebellar (GP) levels were correlated with cognitive dysfunction in a cohort of patients with C9 ALS/FTD⁶⁹. Imaging studies also suggest that there are subtle cerebellar anatomical abnormalities in patients with C9 FTD⁷⁰. As noted above, DPRs have only occasionally been observed in spinal cord motor neurons^{30,67}, in which p62 and TDP43 cytoplasmic aggregates are common. This observation has led to confusion about the role of these DPRs in endogenous disease. Analysis of three post-mortem *C9orf72* NRE carriers who died prematurely of unexpected causes revealed DPR accumulation with minimal TDP43 mislocalization, suggesting that RANT might precede TDP43 pathology⁷¹. However,

it remains unclear whether physiologically relevant levels of DPRs can lead to neurodegeneration. Indeed, recent evidence from *C9orf72* NRE BAC mouse models show that DPR pathology does not correlate with a neurodegenerative phenotype^{22,46}. It will be imperative to standardize antibody use and protocols in future studies to better understand DPR pathobiology in C9 ALS/FTD.

What is the primary gain-of-function mechanism?

Whether ‘toxic’ protein-sequestering RNA or RANT DPRs are the primary drivers of C9 ALS/FTD neurotoxicity is highly debated (see Supplementary information S1 (table)). Exogenous overexpression of 38 or 72 copies of (G₄C₂) ((G₄C₂)₃₈ or ₇₂) produced RNA foci in the SH-SY5Y neuronal cell line and primary mouse hippocampal neuronal cultures and the presence of these foci correlated with high levels of activated caspase 3 and a length-dependent increase in apoptotic activity⁵⁸. (G₄C₂)₃₈ or ₇₂ overexpression also caused length-dependent apoptotic cell death in transfected zebrafish embryos⁵⁸.

Ubiquitous (G₄C₂)₃₀ transgene expression in *Drosophila melanogaster* was developmentally lethal, and specific expression of (G₄C₂)₃₀ in the eye or motor neurons led to progressive neurodegeneration⁵⁶. (G₄C₂)₃₀ expression in *D. melanogaster* motor neurons also induced an age-dependent reduction in locomotor activity⁵⁶. In support of a role for RBP sequestration in these effects, overexpression of Pura, which interacts with (G₄C₂) RNA, suppressed (G₄C₂)₃₀-mediated neurodegeneration in the *D. melanogaster* eye; however, it was unclear whether Pura expression attenuates motor neuron degeneration⁵⁶.

In rodents, neural-specific expression of (G₄C₂)₆₆ caused widespread CNS neurodegeneration⁷². These mice also exhibited both (G₄C₂) RNA foci and (GP) and (GA) DPRs⁷². (G₄C₂)₆₆-expressing mice exhibited cortical neuron and Purkinje cell degeneration and astrogliosis⁷² and developed significant motor deficits⁷². Although these studies support a gain-of-function mechanism of degeneration, they do not separate RNA and DPR toxicity and may not accurately recapitulate the molecular events and/or toxicity associated with physiological expression levels.

The development of neurons differentiated from iPSCs derived from patients with C9 ALS/FTD has enabled the recapitulation of endogenous allelic expression of mutant *C9orf72* RNA. C9orf72 iPSC-derived motor neurons exhibited reduced *C9orf72* mRNA expression, reduced numbers of (G₄C₂) RNA foci, reduced (GP) DPR levels and altered gene expression^{13,14,20}. Similarly, iNeurons (neurons reprogrammed directly from patient fibroblasts) from patients with C9 ALS/FTD exhibited sense and antisense RNA foci and (GP) and (PR) DPRs⁷³. C9orf72 iPSC-derived cortical neurons exhibited enhanced sensitivity to autophagy¹⁴ and iPSC-derived motor neuron cultures were sensitive to endoplasmic reticulum stressors¹⁸. Young C9 ALS/FTD iPSC-derived neuronal cultures exhibited hyperexcitability^{74,75} followed by an age-dependent shift to hypoexcitability^{20,74}. Notably, cortical hyperexcitability has also been identified in symptomatic carriers of the *C9orf92* NRE⁷⁶. Consistent with this observation, younger *C9orf72* iPSC-derived motor neuron cultures were more sensitive to extracellular glutamate stimulation¹³. Acute treatment with ASOs targeting the (G₄C₂) repeat, the intronic region upstream of the NRE or the *C9orf72* coding sequence RNA reduced the number of RNA foci and/or mRNA

expression and mitigated the sensitivity of C9 ALS/FTD iPSC-derived neurons to glutamate despite the presence of (GP) DPRs¹³. These studies point to the involvement of sense (G₄C₂) RNA in toxicity, but do not rule out (GR) or (GA) DPR toxicity, toxicity from newly translated or undetectable (GP) DPRs or toxicity from the (C₄G₂) NRE.

An RNA-based gain-of-function mechanism has also been suggested by longitudinal imaging of primary neuronal cultures. A (G₄C₂)₄₂ intronic RNA expressed in mouse cortical and motor neuron cultures reduced neuronal survival and enhanced the cumulative risk of death⁴¹. Interestingly, RNA toxicity was dose-dependent and exhibited neural-subtype specificity because (G₄C₂)₄₂ expression was not toxic to hippocampal neurons⁴¹. Notably, DPRs were not detected in intronic (G₄C₂)₄₂ RNA-expressing cells⁴¹. However, one cannot exclude a contribution of undetectable levels of DPRs.

In other studies, DPR gain-of-function mechanisms have been demonstrated. Overexpression of ATG-(PR)-(C₄G₂) or ATG-(GP)-(G₄C₂) constructs in HEK293 cells enhanced cell death compared with (G₄C₂) or (C₄G₂) expression alone³⁰. Similarly, U2OS cells showed reduced viability when incubated with (GR)₂₀ or (PR)₂₀, and U2OS cells and cultured human astrocytes exhibited morphological changes when incubated with (GR) or (PR)⁷⁷. Purified (PR)₂₀ was toxic to U2OS cells at high concentrations, whereas similar concentrations of purified (GR)₂₀, which has a short half-life⁷⁷, conferred toxicity if reapplied every 2 hours.

Longitudinal imaging of mouse primary neurons and human neurons showed that overexpression of ATG-(PR)₅₀ enhanced the cumulative risk of death and reduced survival of mouse primary motor, cortical and hippocampal neurons in a dose-dependent manner⁴¹. ATG-(GR)₅₀ was also toxic to mouse primary motor neurons; however, its effects were less severe than ATG-(PR)₅₀, and it did not reduce survival of cortical or hippocampal neurons⁴¹. ATG-(PR)₅₀ overexpression was also neurotoxic to human iPSC-derived neuronal cultures and motor neurons generated from human fibroblasts⁴¹. Although mouse primary motor neuron survival was not significantly reduced by ATG-(GA)₅₀ expression in one study, other studies found that ATG-(GA)_{50–175} expression reduces dendritic branching⁷⁸, induces endoplasmic reticulum stress⁷⁹, impairs Notch signalling⁸⁰, enhances apoptotic signalling via activated caspase 3 and caspase 7, and is toxic to primary mouse cortical neurons *in vitro*^{78,79}.

DPR-based gain of function was also demonstrated in two *in vivo D. melanogaster* studies. Targeted expression of ATG-(GR)₃₆ or ₁₀₀ and ATG-(PR)₃₆ or ₁₀₀ caused degeneration when expressed in the adult eye and reduced survival when expressed ubiquitously⁴⁷. Expression of ATG-(PA)₃₆ and ATG-(GA)₃₆ did not result in eye degeneration; however, ATG-(GA)₁₀₀ expression caused survival deficits that occurred much later than ATG-(GR)₁₀₀ or ATG-(PR)₁₀₀ (REF. 47).

In further support of DPR gain-of-function mechanisms, *D. melanogaster* with targeted expression of (G₄C₂)₃₆, ₁₀₈ or ₂₉₉ modified with stop codons to prevent the translation of long DPR products did not exhibit a degenerative eye phenotype⁴⁷ even though the short G₄C₂ RNAs could still form non-canonical structures and RNA foci. *D. melanogaster*

expressing (G₄C₂)₃₆ or 103 repeats in the absence of stop codons did exhibit neurodegeneration and production of DPRs. These DPRs could be detected in the brain, but it is unclear whether the appearance of the DPRs occurred concomitantly with eye degeneration⁴⁷. In agreement with the idea that *C9orf72* NRE toxicity is protein-based, treatment with a general inhibitor of translation led to decreased toxicity in *D. melanogaster* expressing (G₄C₂)₃₆ or 103 repeats. Additionally, ELISA (enzyme-linked immunosorbent assay) methods that detect low levels of DPRs⁷³ demonstrate that there is a correlation between neurodegeneration and DPR levels in *D. melanogaster* *C9orf72* expression models, with no or minimal correlation with *C9orf72* RNA foci⁸¹. This evidence therefore supports the notion that RANT, and maybe specific DPR species, is the primary driver of *C9orf72* NRE neurotoxicity, but still does not reconcile the recent BAC mouse models of *C9orf72* that show the presence of DPRs and RNA foci with no neurodegenerative phenotype^{22,46}.

In patients with *C9orf72* NRE ALS, DPRs are rarely observed in the spinal cord, despite the frequent presence of cytoplasmic TDP43 in this tissue⁶⁷. However, DPRs are frequently found in the frontal cortex of patients with C9 ALS/FTD, suggesting that a tissue-specific mechanism underlies RNA or DPR toxicity⁸². For example, (GA) and (GP) were found in higher levels in the cerebellum of patients with *C9orf72* FTLD and in patients with FTLD-motor neuron disease than in patients with *C9orf72* ALS, and (GP) DPR levels positively correlated with cognitive impairment⁶⁹. Thus, unknown intrinsic cellular mechanisms that promote RANT might dictate RNA or DPR neurotoxicity.

In summary, it is clear that RNA or DPR accumulation can be toxic in a dose-dependent manner. It is possible that there is a convergence of RNA and DPR toxicity. Indirect evidence for this idea was found when co-expression of intronic (G₄C₂)₄₂ and ATG-(PR)₅₀ synergistically increased the cumulative risk of death of primary mouse cortical neurons⁴¹. However, many of these experiments relied on overexpression or application of the RNA or DPRs at concentrations that probably exceed physiological levels. Therefore, it is important to understand the relative contribution of the RNA and/or DPR toxicity in relevant endogenous systems such as iPSCs and the human brain. To this end, it will be necessary to fully characterize and identify methods to specifically inhibit RANT. Furthermore, whether NRE length correlates with toxicity remains unclear. Recent work has shown that an asymptomatic (G₄C₂)₃₀ carrier exhibited sense and antisense RNA foci, DPRs and p62 inclusions but did not exhibit TDP43 mislocalization, indicating that even short repeats can initiate *C9orf72*-related pathology⁵¹.

Dysfunctional pathways in C9 ALS/FTD

How does the *C9orf72* NRE actually lead to cellular dysfunction? In recent years multiple theories regarding the pathophysiological events that underlie *C9orf72* NRE cytotoxicity have emerged (Supplementary information S1 (table)). These events provide a basic understanding of disease pathophysiology and may provide new targets for intervention.

Nuclear stress.

RNA containing the *C9orf72* NRE accumulates in the nucleus, where it can perturb many pathways and nuclear bodies. These perturbations are largely dependent on the

RBP•*C9orf72* NRE interactome that can alter the function of essential proteins, such as those involved in RNA processing and maturation. Recent transcriptome profile analyses of brain tissue from patients harbouring the *C9orf72* NRE confirmed that RNA processing is impaired and that there is increased alternative splicing and alternative polyadenylation, including mis-splicing of other known ALS- or FTD-linked genes⁵⁷.

The nucleolus — which is necessary for ribosomal RNA biogenesis — is composed of several proteins that can interact with *C9orf72* NRE RNA and specific DPRs, and nucleolar function is impaired in patients carrying the *C9orf72* NRE and models of *C9orf72* NRE^{18,22,41,77,83}. Moreover, (GR)_n and (PR)_n colocalize with the nucleolus and can induce nucleolar stress and cell death at micromolar concentrations⁷⁷. Similar nucleolar stress was observed in *in vitro* models overexpressing (G₄C₂)-containing RNA or scrambled RNAs designed to produce DPR-only cellular toxicity^{41,84}. *C9orf72* NRE BAC mouse models also recapitulate nucleolar stress pathology²². The primary cause of this nucleolar stress remains unknown; however, evidence suggests that a dual toxicity model may be applicable^{13,41} and that the nucleolus could be a convergence point for *C9orf72* NRE RNA and DPR gain of function (FIG. 2).

Nucleocytoplasmic trafficking.

A new direction in our understanding of *C9orf72* NRE pathophysiology came from four independent studies that implicated dysfunctional nucleocytoplasmic trafficking as a key event in C9 ALS/FTD. In the first study, a large interactome screen showed that (G₄C₂) RNA interacts with RanGAP1, a GTPase-activating protein that stimulates the hydrolysis of RanGTPase (Ran)^{13,59}. Ran is necessary for the active transport of proteins that contain a classical nuclear localization sequence (NLS). RanGAP1 was shown to be a potent genetic modifier (suppressor) of a neurodegenerative phenotype (rough eye) in (G₄C₂)₃₀-expressing *D. melanogaster*⁵⁹. C9 ALS iPSC-derived neurons and *D. melanogaster* *C9orf72* NRE models exhibited a disrupted Ran gradient⁵⁹ and showed reduced localization or sluggish fluorescent recovery after photobleaching (FRAP) for NLS-containing reporter molecules, implying that there are deficits in nuclear protein import⁵⁹ and disruption of functional nucleocytoplasmic transport. Indeed, TDP43, which contains a classical NLS sequence, was slightly enriched in the cytoplasm of C9 ALS iPSC neurons⁵⁹. Genetic crosses and therapeutics designed to modify nucleocytoplasmic trafficking and enhance protein import or reduce nuclear export rescued the (G₄C₂)-mediated rough eye phenotype⁵⁹. ASOs targeting the NRE sense strand rescued the nuclear transport deficits and were neuroprotective against the rough eye phenotype in the *D. melanogaster* model and rescued the Ran gradient and reduced cytoplasmic TDP43 localization in ALS iPSC neurons⁵⁹. Notably, in these studies, the flies expressed only the sense strand RNAs and DPRs were undetectable during the time of neurodegeneration, suggesting that the observed deficits were primarily caused by NRE sense strand (G₄C₂) RNA.

A second study carried out an unbiased loci deletion screen in a novel *D. melanogaster* model that expresses (G₄C₂)₅₈ and exhibits sense strand RANT products⁸⁵. Eighteen genetic modifiers of the rough eye phenotype in genes that encode for NPCs and components of the nucleocytoplasmic trafficking pathway were identified⁸⁵. One potent suppressor of the rough

eye phenotype was AlyRef, a protein involved in mRNA export^{49,54,85}. Consistent with the idea that AlyRef has a role in C9 ALS pathogenesis, (G₄C₂)₅₈-expressing *D. melanogaster* cells and C9 ALS iPSC motor neurons showed significant nuclear RNA retention, which was amplified when additional copies of (G₄C₂)₅₈ transgenes were added⁸⁵. These studies therefore support a model of dysfunction at the NPC, wherein protein import of classical NLS-containing proteins and export of RNAs is impaired.

NPC proteins were also found to be abnormally aggregated in C9 ALS iPSC neurons and throughout the motor cortex of many patients with C9 ALS⁵⁹. For example, RanGAP1 showed perinuclear aggregation and intranuclear accumulation⁵⁹. Nup107 (a modifier of neurodegeneration in the (G₄C₂)₅₈ fly model⁸⁵) and Nup205 showed similar aggregation and nuclear retention in the motor cortex of patients with C9 ALS⁵⁹, and nuclear membrane abnormalities were also observed in the (G₄C₂)₅₈ fly model⁸⁵. The mechanism (or mechanisms) underlying this aggregation is unknown. Furthermore, whether the aggregated NPC proteins affect both nuclear import and export as well as whether this pathology is reversible remains to be ascertained.

A third study showed that DPRs can also affect nucleocytoplasmic transport. Overexpression of ATG-(PR)₅₀ in yeast resulted in toxicity that could be modified by 35 genetic enhancers and 27 suppressors⁸⁶. Notably, 11 of these suppressors were involved in nucleocytoplasmic trafficking, including six importin proteins⁸⁶. iNeurons from two patients with C9 ALS also showed a reduction and/or mislocalization of a nuclear RanGEF protein, RCC1 (REF. 86), that could similarly disrupt nucleocytoplasmic trafficking. Interestingly, in contrast to the studies above, this study implicated the antisense (C₄G₂) RNA in toxicity, via (PR) accumulation⁸⁶. Therefore, although both the sense and antisense NRE lead to nucleocytoplasmic trafficking deficits, they could act through independent mechanisms.

In the fourth and most recent study, cytoplasmic protein aggregates were shown to interfere with nucleocytoplasmic transport⁸⁷. This study demonstrated that the aggregation of overexpressed artificial β -sheet proteins or mutant huntingtin or TDP43 fragments in the cytoplasm and not the nucleus lead to nucleocytoplasmic transport defects in HEK293T cells. Altered NPC pathology was also only observed when the β -sheet proteins aggregated in the cytoplasm. This work suggested that cytoplasmic protein aggregation, a pathological feature shared by several neurological disorders, can itself lead to impaired nucleocytoplasmic transport⁸⁷.

The primary cause of NPC dysfunction and nucleocytoplasmic transport defects in patients carrying the *C9orf72* NRE and in models of *C9orf72* NRE is currently unclear. The interpretation of some results may be hampered by the non-physiological overexpression systems used. It seems likely that there are many nuclear trafficking proteins that are affected by the (G₄C₂) RNA, DPRs and/or cytoplasmic protein aggregation. However, it is also likely that nucleocytoplasmic transport and NPC dysfunction in C9 ALS/FTD is a central event and that it occurs via gain-of-function mechanisms. Furthermore, it is possible that there may be a node at which different mechanisms converge. For example, (G₄C₂) RNA might initiate the cascade by interacting with RNA export factors in the nucleus and nuclear- or NPC-bound components. Simultaneously, (G₄C₂)_{exp} and (C₄G₂)_{exp} might escape

the nucleus and through RANT produce DPRs that aggregate in the cytoplasm. Arginine-rich DPRs could bind to importins and be transported to the NPC to induce further disruptions. These events could seed NPC aggregation that further compromises nuclear transport while enhancing the nuclear retention of RNAs and the cytoplasmic accumulation of NLS-containing proteins, such as TDP43. This process might lead to enhanced cryptic exon inclusion due to TDP43 mislocalization⁸⁸. Furthermore, RNA retention and nuclear import of arginine-rich DPRs could disrupt nuclear bodies such as the nucleolus^{18,41}. Finally, the short isoform of endogenous C9orf72, which localizes to the nuclear envelope and biochemically interacts with RanGTPase and importins¹¹, may also contribute to nucleocytoplasmic transport defects through a loss-of-function mechanism. Therefore, it is enticing to speculate that loss of function of C9orf72 might confer additional sensitivity to cells with gain-of-function-mediated deficits in the nucleocytoplasmic transport pathway.

Therapeutic interventions

The prospects for therapeutic interventions for nucleotide expansion disease are exciting. To date, sense ASOs have rescued *C9orf72*-linked disease phenotypes in human cellular and animal preclinical models (detailed above) and have demonstrated mitigation of C9orf72 pathophysiology^{13,20,42}. Substantial progress has been made with preclinical studies of ASOs targeting another NRE-linked disease, myotonic dystrophy⁸⁹, and Phase I trials have been successfully completed with ASOs targeting superoxide dismutase 1 (SOD1) in familial ALS^{90,91}. There are also ongoing studies using ASOs for spinal muscular atrophy⁹¹. Future studies will provide additional information on the efficacy of ASO therapies.

Recent progress has also been made in the design of nucleotide structure-specific small molecules that can alter the progression of disease. For example, small molecules designed to specifically recognize and bind to nucleotides in unique RNA spatial conformations, such as RNA hairpins, have been shown to rescue RNA-induced pathological defects and prevent the downstream accumulation of DPRs^{73,92}. Other structurally specific drugs, such as the G-quadruplex-binding porphyrin, TMPyP4, can destabilize *C9orf72* RNA G-quadruplex•RBP interactions⁹³ and can rescue transport defect pathologies and the rough eye phenotypes observed in the transgenic *D. melanogaster* (G₄C₂)₃₀ overexpression model⁵⁹. Genetic modification or compounds such as KPT-276 that inhibit nuclear export can rescue neurodegeneration and nucleocytoplasmic trafficking deficits in (G₄C₂)-overexpressing flies, suggesting that modifiers of nucleocytoplasmic trafficking might be an effective alternative therapeutic strategy^{59,85}. Similar molecules have recently been shown to be efficacious in rodent models of inflammatory demyelination⁹⁴. Inhibiting nuclear export of expanded (G₄C₂) RNA might prevent NPC dysfunction and toxic DPR accumulation.

Pharmacodynamic read outs to assess drug efficacy still remain a challenge when developing therapies for neurodegenerative disorders. SOD1 levels in the cerebrospinal fluid have been used as a pharmacodynamic marker to measure the efficacy of ASO therapeutics in patients with ALS⁹⁵. Similarly, an ELISA-based detection assay for the RANT (GP) DPR is currently in development. To date, studies have successfully quantified extracellular (GP) DPR in the cerebrospinal fluid of 14 patients with C9 ALS/FTD⁷³; perhaps this assay could

be used as a longitudinal biomarker for novel *C9orf72* NRE-targeting ASOs or small molecules.

Recent advances in genome editing using CRISPR/Cas technology have revolutionized the development of model organisms as a result of its speed and efficacy in engineering DNA. The possibility that CRISPR/Cas technology or similar genome editing might also be used as the basis of therapy to eliminate a fatal human disease is exciting; however, the ethical debate regarding treating genomic diseases in humans will be important⁹⁶. Although improvements have been made to reduce off-target effects, this technology is still in its infancy, and its efficacy *in vivo* must continue to be rigorously examined before its use as genome therapy tool could be considered.

Concluding thoughts and perspectives

The evidence that *C9orf72* NRE generated DPRs, RNA and cytoplasmic protein aggregates that may interfere with NPC and nucleocytoplasmic transport presents opportunities to enhance our understanding, and potential mitigation, of neurodegeneration. For example, does altered nucleocytoplasmic transport precede the cytoplasmic accumulation of TDP43 inclusions in patients⁶⁷? Can we create tools to mitigate DPR-based toxicity? Is the NPC more sensitive in neurons than glia and/or differentially regulated in neuron subtypes, or in neurons versus glia? What is the underlying mechanism of nuclear pore aggregation? Does the short isoform of *C9orf72* contribute to nuclear pore pathology? To answer these questions, it is necessary to gain a better understating of the chain of events and the primary mechanistic contributor that leads to the observed patient pathologies.

Substantial research focus has been placed on the toxic gain-of-function mechanisms of the *C9orf72* NRE. This focus was partially due to difficulties in understanding the role and function of *C9orf72* in normal biology. Although RNA and DPRs from the *C9orf72* NRE mutation have revealed many pathological disease features that are initiated by the gain-of-function mechanism, it is unknown which features directly correlate with neurodegeneration. However, uncoupling the *C9orf72* NRE phenotypes caused by the RNA versus DPR gain-of-function mechanisms remains a challenge. Quantitative measurements of endogenous RNA and DPR levels *in vivo* are critical to develop accurate interpretations of *in vitro* models. Moreover, current RNA and/or DPR overexpression models must be interpreted with strong caution because there can often be a disconnect between phenotypic observations in overexpression systems and actual patient phenotypes.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

The authors thank L. Ranum, A. Gitler, L. Petrucelli, P. Taylor, D. Cleveland, F. Bennet, F. Rigo, J. Wang, P. Wong and T. Lloyd for discussions that helped provide a foundation for this Review. Financial support was provided from the following institutes: the National Institutes of Health-National Institute of Neurological Disorders and Stroke (NIH-NINDS); Target ALS; the Muscular Dystrophy Association; the Robert Packard Center for ALS Research; the Live Like Lou at the Pittsburgh Foundation; and the Brain Institute at the University of Pittsburgh School of Medicine, USA.

References

1. Renton AE et al. A hexanucleotide repeat expansion in *C9ORF72* is the cause of chromosome 9p21-linked ALS-FTD. *Neuron* 72, 257–268 (2011). [PubMed: 21944779]
2. DeJesus-Hernandez M et al. Expanded GGGGCC hexanucleotide repeat in noncoding region of *C9ORF72* causes chromosome 9p-linked FTD and ALS. *Neuron* 72, 245–256 (2011). [PubMed: 21944778] References 1 and 2 are the seminal publications that defined the *C9orf72* mutation.
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). [PubMed: 22406228]
4. 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* 3, 283 (2012). [PubMed: 23248642]
5. 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* 29, 499–503 (2013). [PubMed: 23329412]
6. Allaire PD et al. The Connecdenn DENN domain: a GEF for Rab35 mediating cargo-specific exit from early endosomes. *Mol. Cell* 37, 370–382 (2010). [PubMed: 20159556]
7. Marat AL, Dokainish H & McPherson PS DENN domain proteins: regulators of Rab GTPases. *J. Biol. Chem* 286, 13791–13800 (2011). [PubMed: 21330364]
8. Ciura S et al. Loss of function of *C9orf72* causes motor deficits in a zebrafish model of amyotrophic lateral sclerosis. *Ann. Neurol* 74, 180–187 (2013). [PubMed: 23720273]
9. Suzuki N et al. The mouse *C9ORF72* ortholog is enriched in neurons known to degenerate in ALS and FTD. *Nat. Neurosci* 16, 1725–1727 (2013). [PubMed: 24185425] This report shows that the expression of endogenous *C9orf72* is elevated in neurons.
10. Kim MS et al. A draft map of the human proteome. *Nature* 509, 575–581 (2014). [PubMed: 24870542]
11. Xiao S et al. Isoform-specific antibodies reveal distinct subcellular localizations of *C9orf72* in amyotrophic lateral sclerosis. *Ann. Neurol* 78, 568–583 (2015). [PubMed: 26174152]
12. Farg MA et al. *C9ORF72*, implicated in amyotrophic lateral sclerosis and frontotemporal dementia, regulates endosomal trafficking. *Hum. Mol. Genet* 23, 3579–3595 (2014). [PubMed: 24549040]
13. Donnelly CJ et al. RNA toxicity from the ALS/FTD *C9ORF72* expansion is mitigated by antisense intervention. *Neuron* 80, 415–428 (2013). [PubMed: 24139042] A study showing that human *C9orf72* iPSC-derived motor neurons can model disease and that ASOs can mitigate neuronal injury.
14. Almeida S et al. Modeling key pathological features of frontotemporal dementia with *C9ORF72* repeat expansion in iPSC-derived human neurons. *Acta Neuropathol.* 126, 385–399 (2013). [PubMed: 23836290] This report shows that iPSC-derived cortical neurons exhibit RNA foci and are sensitive to autophagy-induced stress.
15. Gijssels I et al. A *C9orf72* promoter repeat expansion in a Flanders-Belgian cohort with disorders of the frontotemporal lobar degeneration-amyotrophic lateral sclerosis spectrum: a gene identification study. *Lancet Neurol.* 11, 54–65 (2012). [PubMed: 22154785]
16. Belzil VV 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). [PubMed: 24166615]
17. Waite AJ 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).
18. Haeusler AR et al. *C9orf72* nucleotide repeat structures initiate molecular cascades of disease. *Nature* 507, 195–200 (2014). [PubMed: 24598541]
19. 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). [PubMed: 23731538]
20. 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). This paper describes altered

transcription of *C9orf72* in iPSC-derived motor neurons expressing the expanded allele and the correction of these defects with ASOs.

21. van Blitterswijk M et al. Novel clinical associations with specific *C9ORF72* transcripts in patients with repeat expansions in *C9ORF72*. *Acta Neuropathol.* 130, 863–876 (2015). [PubMed: 26437865]
22. O'Rourke JG et al. *C9orf72* BAC transgenic mice display typical pathologic features of ALS/FTD. *Neuron* 88, 892–901 (2015). [PubMed: 26637796]
23. Grabczyk E & Usdin K The GAA*TTT triplet repeat expanded in Friedreich's ataxia impedes transcription elongation by T7 RNA polymerase in a length and supercoil dependent manner. *Nucleic Acids Res.* 28, 2815–2822 (2000). [PubMed: 10908340]
24. Ohshima K, Montermini L, Wells RD & Pandolfo M Inhibitory effects of expanded GAA. TTT triplet repeats from intron I of the Friedreich ataxia gene on transcription and replication *in vivo*. *J. Biol. Chem* 273, 14588–14595 (1998). [PubMed: 9603975]
25. Parsons MA, Sinden RR & Izban MG Transcriptional properties of RNA polymerase II within triplet repeat-containing DNA from the human myotonic dystrophy and fragile X loci. *J. Biol. Chem* 273, 26998–27008 (1998). [PubMed: 9756950]
26. Tornaletti S, Park-Snyder S & Hanawalt PC G4-forming sequences in the non-transcribed DNA strand pose blocks to T7 RNA polymerase and mammalian RNA polymerase II. *J. Biol. Chem* 283, 12756–12762 (2008). [PubMed: 18292094]
27. Li Y et al. Expanded GAA repeats impede transcription elongation through the *FXN* gene and induce transcriptional silencing that is restricted to the *FXN* locus. *Hum. Mol. Genet* 24, 6932–6943 (2015). [PubMed: 26401053]
28. Liu CR et al. Spt4 is selectively required for transcription of extended trinucleotide repeats. *Cell* 148, 690–701 (2012). [PubMed: 22341442]
29. Batra R, Charizanis K & Swanson MS Partners in crime: bidirectional transcription in unstable microsatellite disease. *Hum. Mol. Genet* 19, R77–R82 (2010). [PubMed: 20368264]
30. Zu T et al. RAN proteins and RNA foci from antisense transcripts in *C9ORF72* ALS and frontotemporal dementia. *Proc. Natl Acad. Sci. USA* 110, E4968–E4977 (2013). [PubMed: 24248382]
31. Mori K et al. Bidirectional transcripts of the expanded *C9orf72* hexanucleotide repeat are translated into aggregating dipeptide repeat proteins. *Acta Neuropathol.* 126, 881–893 (2013). [PubMed: 24132570] References 30 and 31 show that repeat-associated aggregates generated from sense and antisense *C9orf72* transcripts are detected in the human brain.
32. Core LJ, Waterfall JJ & Lis JT Nascent RNA sequencing reveals widespread pausing and divergent initiation at human promoters. *Science* 322, 1845–1848 (2008). [PubMed: 19056941]
33. Seila AC et al. Divergent transcription from active promoters. *Science* 322, 1849–1851 (2008). [PubMed: 19056940]
34. Bernstein E & Allis CD RNA meets chromatin. *Genes Dev.* 19, 1635–1655 (2005). [PubMed: 16024654]
35. Pelechano V & Steinmetz LM Gene regulation by antisense transcription. *Nat. Rev. Genet* 14, 880–893 (2013). [PubMed: 24217315]
36. Chung DW, Rudnicki DD, Yu L & Margolis RL A natural antisense transcript at the Huntington's disease repeat locus regulates *HTT* expression. *Hum. Mol. Genet* 20, 3467–3477 (2011). [PubMed: 21672921]
37. Sopher BL et al. CTCF regulates ataxin-7 expression through promotion of a convergently transcribed, antisense noncoding RNA. *Neuron* 70, 1071–1084 (2011). [PubMed: 21689595]
38. Gascon E & Gao FB The emerging roles of microRNAs in the pathogenesis of frontotemporal dementia-amyotrophic lateral sclerosis (FTD-ALS) spectrum disorders. *J. Neurogenet.* 28, 30–40 (2014). [PubMed: 24506814]
39. Freischmidt A et al. Serum microRNAs in patients with genetic amyotrophic lateral sclerosis and premanifest mutation carriers. *Brain* 137, 2938–2950 (2014). [PubMed: 25193138]
40. Therrien M, Rouleau GA, Dion PA & Parker JA Deletion of *C9ORF72* results in motor neuron degeneration and stress sensitivity in *C. elegans*. *PLoS ONE* 8, e83450 (2013). [PubMed: 24349511]

41. Wen X et al. Antisense proline-arginine RAN dipeptides linked to *C9ORF72*-ALS/FTD form toxic nuclear aggregates that initiate *in vitro* and *in vivo* neuronal death. *Neuron* 84, 1213–1225 (2014). [PubMed: 25521377]
42. Lagier-Tourenne C et al. Targeted degradation of sense and antisense *C9orf72* RNA foci as therapy for ALS and frontotemporal degeneration. *Proc. Natl Acad. Sci. USA* 110, E4530–E4539 (2013). [PubMed: 24170860] This study reveals the presence of (C₄G₂) RNA foci in ALS and FTD and shows that antisense knockdown of *C9orf72* in the adult rodent does not induce neurodegeneration.
43. Koppers M et al. *C9orf72* ablation in mice does not cause motor neuron degeneration or motor deficits. *Ann. Neurol.* 78, 426–438 (2015). [PubMed: 26044557]
44. Harms MB et al. Lack of *C9ORF72* coding mutations supports a gain of function for repeat expansions in amyotrophic lateral sclerosis. *Neurobiol. Aging* 34, 2234.e13–2234.e19 (2013).
45. Fratta P et al. Homozygosity for the *C9orf72* GGGGCC repeat expansion in frontotemporal dementia. *Acta Neuropathol.* 126, 401–409 (2013). [PubMed: 23818065]
46. Peters OM et al. Human *C9ORF72* hexanucleotide expansion reproduces RNA foci and dipeptide repeat proteins but not neurodegeneration in BAC transgenic mice. *Neuron* 88, 902–909 (2015). [PubMed: 26637797]
47. Mizielinska S et al. *C9orf72* repeat expansions cause neurodegeneration in *Drosophila* through arginine-rich proteins. *Science* 345, 1192–1194 (2014). [PubMed: 25103406] This paper reports that selected RAN translation peptides, when overexpressed in fly, can be neurotoxic.
48. Gendron TF et al. Antisense transcripts of the expanded *C9ORF72* hexanucleotide repeat form nuclear RNA foci and undergo repeat-associated non-ATG translation in c9FTD/ALS. *Acta Neuropathol.* 126, 829–844 (2013). [PubMed: 24129584] A study showing the presence of RNA foci and RAN translation aggregates in a *C9orf72* human brain.
49. Cooper-Knock J et al. Antisense RNA foci in the motor neurons of *C9ORF72*-ALS patients are associated with TDP-43 proteinopathy. *Acta Neuropathol.* 130, 63–75 (2015). [PubMed: 25943887]
50. Mizielinska S et al. *C9orf72* frontotemporal lobar degeneration is characterised by frequent neuronal sense and antisense RNA foci. *Acta Neuropathol.* 126, 845–857 (2013). [PubMed: 24170096] A highly descriptive study quantifying RNA foci and RAN products in a brain of a patient with *C9orf72* ALS.
51. Gami P 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.* 130, 599–601 (2015). [PubMed: 26347457]
52. Todd PK & Paulson HL RNA-mediated neurodegeneration in repeat expansion disorders. *Ann. Neurol.* 67, 291–300 (2010). [PubMed: 20373340]
53. Miller JW et al. Recruitment of human muscleblind proteins to (CUG)_n expansions associated with myotonic dystrophy. *EMBO J.* 19, 4439–4448 (2000). [PubMed: 10970838]
54. Cooper-Knock J et al. Sequestration of multiple RNA recognition motif-containing proteins by *C9orf72* repeat expansions. *Brain* 137, 2040–2051 (2014). [PubMed: 24866055]
55. Mori K et al. hnRNP A3 binds to GGGGCC repeats and is a constituent of p62-positive/TDP43-negative inclusions in the hippocampus of patients with *C9orf72* mutations. *Acta Neuropathol.* 125, 413–423 (2013). [PubMed: 23381195]
56. Xu Z et al. Expanded GGGGCC repeat RNA associated with amyotrophic lateral sclerosis and frontotemporal dementia causes neurodegeneration. *Proc. Natl Acad. Sci. USA* 110, 7778–7783 (2013). [PubMed: 23553836] The first study to show that (G₄C₂) RNA expression in *Drosophila* is neurotoxic.
57. Prudencio M et al. Distinct brain transcriptome profiles in *C9orf72*-associated and sporadic ALS. *Nat. Neurosci.* 18, 1175–1182 (2015). [PubMed: 26192745]
58. 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). [PubMed: 24290757]
59. Zhang K et al. The *C9orf72* repeat expansion disrupts nucleocytoplasmic transport. *Nature* 525, 56–61 (2015). [PubMed: 26308891] This study provides evidence of nucleocytoplasmic transport

defects, including impaired nuclear protein import, in (G₄C₂) RNA *Drosophila* and *C9orf72* ALS/FTD iPSC-derived neurons.

60. Zu T et al. Non-ATG-initiated translation directed by microsatellite expansions. *Proc. Natl Acad. Sci. USA* 108, 260–265 (2011). [PubMed: 21173221]
61. Todd PK et al. CGG repeat-associated translation mediates neurodegeneration in fragile X tremor ataxia syndrome. *Neuron* 78, 440–455 (2013). [PubMed: 23602499]
62. Mori K et al. The *C9orf72* GGGGCC repeat is translated into aggregating dipeptide-repeat proteins in FTL/ALS. *Science* 339, 1335–1338 (2013). [PubMed: 23393093]
63. Ash PEA et al. Unconventional translation of *C9ORF72* GGGGCC expansion generates insoluble polypeptides specific to c9FTD/ALS. *Neuron* 77, 639–646 (2013). [PubMed: 23415312] This paper describes the detection of RANT products in patients with *C9orf72* ALS.
64. Mackenzie IR et al. Dipeptide repeat protein pathology in *C9ORF72* mutation cases: clinico-pathological correlations. *Acta Neuropathol.* 126, 859–879 (2013). [PubMed: 24096617]
65. Mann DM et al. Dipeptide repeat proteins are present in the p62 positive inclusions in patients with frontotemporal lobar degeneration and motor neurone disease associated with expansions in *C9ORF72*. *Acta Neuropathol. Commun* 1, 68 (2013). [PubMed: 24252525]
66. Cleary JD & Ranum LP Repeat associated non-ATG (RAN) translation: new starts in microsatellite expansion disorders. *Curr. Opin. Genet. Dev* 26, 6–15 (2014). [PubMed: 24852074]
67. Gomez-Deza J et al. Dipeptide repeat protein inclusions are rare in the spinal cord and almost absent from motor neurons in *C9ORF72* mutant amyotrophic lateral sclerosis and are unlikely to cause their degeneration. *Acta Neuropathol. Commun.* 3, 38 (2015). [PubMed: 26108573]
68. Schludi MH et al. Distribution of dipeptide repeat proteins in cellular models and *C9orf72* mutation cases suggests link to transcriptional silencing. *Acta Neuropathol.* 130, 537–555 (2015). [PubMed: 26085200]
69. Gendron TF et al. Cerebellar c9RAN proteins associate with clinical and neuropathological characteristics of *C9ORF72* repeat expansion carriers. *Acta Neuropathol.* 130, 559–573 (2015). [PubMed: 26350237]
70. Cooper-Knock J et al. Clinico-pathological features in amyotrophic lateral sclerosis with expansions in *C9ORF72*. *Brain* 135, 751–764 (2012). [PubMed: 22366792]
71. Baborie A 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* 41, 601–612 (2014).
72. Chew J et al. Neurodegeneration. *C9ORF72* repeat expansions in mice cause TDP-43 pathology, neuronal loss, and behavioral deficits. *Science* 348, 1151–1154 (2015). [PubMed: 25977373] This report shows that adeno-associated virus-mediated overexpression of (G₄C₂) RNA in rodent brain causes RNA foci, generates RANT products and induces neurodegeneration.
73. Su Z et al. Discovery of a biomarker and lead small molecules to target r(GGGGCC)-associated defects in c9FTD/ALS. *Neuron* 83, 1043–1050 (2014). [PubMed: 25132468] This paper shows that small molecules that bind to (G₄C₂) hairpins reduce RNA foci and RANT production, and that (GP) RANT products are present in the spinal cord of a patient with *C9orf72* ALS.
74. Devlin A-C et al. Human iPSC-derived motoneurons harbouring *TARDBP* or *C9ORF72* ALS mutations are dysfunctional despite maintaining viability. *Nat. Commun* 6, 5999 (2015). [PubMed: 25580746]
75. Wainger BJ et al. Intrinsic membrane hyperexcitability of amyotrophic lateral sclerosis patient-derived motor neurons. *Cell Rep.* 7, 1–11 (2014). [PubMed: 24703839]
76. Wainger BJ & Cudkowicz ME Cortical hyperexcitability in amyotrophic lateral sclerosis: *C9orf72* repeats. *JAMA Neurol.* 72, 1235–1236 (2015). [PubMed: 26348624]
77. Kwon I et al. Poly-dipeptides encoded by the *C9orf72* repeats bind nucleoli, impede RNA biogenesis, and kill cells. *Science* 345, 1139–1145 (2014). [PubMed: 25081482]
78. May S et al. *C9orf72* FTL/ALS-associated Gly-Ala dipeptide repeat proteins cause neuronal toxicity and Unc119 sequestration. *Acta Neuropathol.* 128, 485–503 (2014). [PubMed: 25120191]
79. Zhang Y-J et al. Aggregation-prone c9FTD/ALS poly(GA) RAN-translated proteins cause neurotoxicity by inducing ER stress. *Acta Neuropathol.* 128, 505–524 (2014). [PubMed: 25173361]

80. Yang D et al. FTD/ALS-associated poly(GR) protein impairs the Notch pathway and is recruited by poly(GA) into cytoplasmic inclusions. *Acta Neuropathol.* 130, 525–535 (2015). [PubMed: 26031661]
81. Tran H et al. Differential toxicity of nuclear RNA foci versus dipeptide repeat proteins in a *Drosophila* model of *C9ORF72* FTD/ALS. *Neuron* 87, 1207–1214 (2015). [PubMed: 26402604]
82. Yamakawa M et al. Characterization of the dipeptide repeat protein in the molecular pathogenesis of c9FTD/ALS. *Hum. Mol. Genet.* 24, 1630–1645 (2015). [PubMed: 25398948]
83. Tsoi H, Lau TC-K, Tsang S-Y, Lau K-F & Chan HYE CAG expansion induces nucleolar stress in polyglutamine diseases. *Proc. Natl Acad. Sci. USA* 109, 13428–13433 (2012). [PubMed: 22847428]
84. Tao Z et al. Nucleolar stress and impaired stress granule formation contribute to *C9orf72* RAN translation-induced cytotoxicity. *Hum. Mol. Genet* 24, 2426–2441 (2015). [PubMed: 25575510]
85. Freibaum BD et al. GGGGCC repeat expansion in *C9orf72* compromises nucleocytoplasmic transport. *Nature* 525, 129–133 (2015). [PubMed: 26308899] An unbiased genetic screen of novel (G₄C₂) *Drosophila* model reveals several genetic modifiers involved in nucleocytoplasmic transport and components of the NPC.
86. Jovicic A et al. Modifiers of *C9orf72* dipeptide repeat toxicity connect nucleocytoplasmic transport defects to FTD/ALS. *Nat. Neurosci* 18, 1226–1229 (2015). [PubMed: 26308983] A genetic screen of (PR)-expressing yeast reveals many genetic modifiers involved in nucleocytoplasmic transport.
87. Woerner AC et al. Cytoplasmic protein aggregates interfere with nucleocytoplasmic transport of protein and RNA. *Science* 351, 173–176 (2016). [PubMed: 26634439]
88. Ling JP, Pletnikova O, Troncoso JC & Wong PC TDP-43 repression of nonconserved cryptic exons is compromised in ALS-FTD. *Science* 349, 650–655 (2015). [PubMed: 26250685] This paper reports that loss of TDP43 enhances the presence of cryptic exons.
89. Wheeler TM et al. Targeting nuclear RNA for *in vivo* correction of myotonic dystrophy. *Nature* 488, 111–115 (2012). [PubMed: 22859208]
90. Miller TM et al. An antisense oligonucleotide against *SOD1* delivered intrathecally for patients with *SOD1* familial amyotrophic lateral sclerosis: a Phase 1, randomised, first-in-man study. *Lancet Neurol.* 12, 435–442 (2013). [PubMed: 23541756]
91. Keil JM et al. A short antisense oligonucleotide ameliorates symptoms of severe mouse models of spinal muscular atrophy. *Mol. Ther. Nucleic Acids* 3, e174 (2014). [PubMed: 25004100]
92. Yang WY, Wilson HD, Velagapudi SP & Disney MD Inhibition of Non-ATG translational events in cells via covalent small molecules targeting RNA. *J. Am. Chem. Soc* 137, 5336–5345 (2015). [PubMed: 25825793]
93. Zamiri B, Reddy K, Macgregor RB Jr & Pearson CE TMPyP4 porphyrin distorts RNA G-quadruplex structures of the disease-associated r(GGGGCC)*n* repeat of the *C9orf72* gene and blocks interaction of RNA-binding proteins. *J. Biol. Chem* 289, 4653–4659 (2014). [PubMed: 24371143]
94. Haines JD et al. Nuclear export inhibitors avert progression in preclinical models of inflammatory demyelination. *Nat. Neurosci* 18, 511–520 (2015). [PubMed: 25706475]
95. Winer L et al. SOD1 in cerebral spinal fluid as a pharmacodynamic marker for antisense oligonucleotide therapy. *JAMA Neurol.* 70, 201–207 (2013). [PubMed: 23147550]
96. Liang P et al. CRISPR/Cas9-mediated gene editing in human triprounuclear zygotes. *Protein Cell* 6, 363–372 (2015). [PubMed: 25894090]
97. Xi Z et al. Jump from pre-mutation to pathologic expansion in *C9orf72*. *Am. J. Hum. Genet* 96, 962–970 (2015). [PubMed: 26004200]
98. Xi Z et al. The *C9orf72* repeat expansion itself is methylated in ALS and FTLN patients. *Acta Neuropathol.* 129, 715–727 (2015). [PubMed: 25716178]
99. McMurray CT Mechanisms of trinucleotide repeat instability during human development. *Nat. Rev. Genet* 11, 786–799 (2010). [PubMed: 20953213]
100. Suh E 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). [PubMed: 26022924]

101. Gijssels I et al. The *C9orf72* repeat size correlates with onset age of disease, DNA methylation and transcriptional downregulation of the promoter. *Mol. Psychiatry* 10.1038/mp.2015.159 (2015).
102. Gallagher MD et al. *TMEM106B* is a genetic modifier of frontotemporal lobar degeneration with *C9orf72* hexanucleotide repeat expansions. *Acta Neuropathol.* 127, 407–418 (2014). [PubMed: 24442578]
103. van Blitterswijk M et al. *TMEM106B* protects *C9ORF72* expansion carriers against frontotemporal dementia. *Acta Neuropathol.* 127, 397–406 (2014). [PubMed: 24385136]
104. Mirkin SM Expandable DNA repeats and human disease. *Nature* 447, 932–940 (2007). [PubMed: 17581576]
105. Orr HT & Zoghbi HY Trinucleotide repeat disorders. *Annu. Rev. Neurosci* 30, 575–621 (2007). [PubMed: 17417937]
106. La Spada AR & Taylor JP Repeat expansion disease: progress and puzzles in disease pathogenesis. *Nat. Rev. Genet* 11, 247–258 (2010). [PubMed: 20177426]
107. Pearson CE, Nichol Edamura K & Cleary JD Repeat instability: mechanisms of dynamic mutations. *Nat. Rev. Genet* 6, 729–742 (2005). [PubMed: 16205713]
108. Dols-Icardo O et al. Characterization of the repeat expansion size in *C9orf72* in amyotrophic lateral sclerosis and frontotemporal dementia. *Hum. Mol. Genet* 23, 749–754 (2014). [PubMed: 24057670]
109. Nordin A et al. Extensive size variability of the GGGGCC expansion in *C9orf72* in both neuronal and non-neuronal tissues in 18 patients with ALS or FTD. *Hum. Mol. Genet* 24, 3133–3142 (2015). [PubMed: 25712133]
110. Aguilera A & Garcia-Muse T Causes of genome instability. *Annu. Rev. Genet* 47, 1–32 (2013). [PubMed: 23909437]
111. Lopez Castel A, Cleary JD & Pearson CE Repeat instability as the basis for human diseases and as a potential target for therapy. *Nat. Rev. Mol. Cell Biol* 11, 165–170 (2010). [PubMed: 20177394]
112. Kruisselbrink E et al. Mutagenic capacity of endogenous G4 DNA underlies genome instability in FANCI-defective *C. elegans*. *Curr. Biol* 18, 900–905 (2008). [PubMed: 18538569]
113. Ribeyre C et al. The yeast Pif1 helicase prevents genomic instability caused by G-quadruplex-forming CEB1 sequences *in vivo*. *PLoS Genet.* 5, e1000475 (2009). [PubMed: 19424434]
114. Piazza A et al. Genetic instability triggered by G-quadruplex interacting Phen-DC compounds in *Saccharomyces cerevisiae*. *Nucleic Acids Res.* 38, 4337–4348 (2010). [PubMed: 20223771]
115. De S & Michor F DNA secondary structures and epigenetic determinants of cancer genome evolution. *Nat. Struct. Mol. Biol* 18, 950–955 (2011). [PubMed: 21725294]
116. Aghili L, Foo J, DeGregori J & De S Patterns of somatically acquired amplifications and deletions in apparently normal tissues of ovarian cancer patients. *Cell Rep.* 7, 1310–1319 (2014). [PubMed: 24794429]
117. Koole W et al. A polymerase Theta-dependent repair pathway suppresses extensive genomic instability at endogenous G4 DNA sites. *Nat. Commun.* 5, 3216 (2014). [PubMed: 24496117]
118. van Kregten M & Tijsterman M The repair of G-quadruplex-induced DNA damage. *Exp. Cell Res* 329, 178–183 (2014). [PubMed: 25193076]
119. Aguilera A The connection between transcription and genomic instability. *EMBO J.* 21, 195–201 (2002). [PubMed: 11823412]
120. Li X & Manley JL Cotranscriptional processes and their influence on genome stability. *Genes Dev.* 20, 1838–1847 (2006). [PubMed: 16847344]
121. Aguilera A & Garcia-Muse T R loops: from transcription byproducts to threats to genome stability. *Mol. Cell* 46, 115–124 (2012). [PubMed: 22541554]
122. Bhatia V et al. BRCA2 prevents R-loop accumulation and associates with TREX-2 mRNA export factor PCID2. *Nature* 511, 362–365 (2014). [PubMed: 24896180]
123. Skourti-Stathaki K & Proudfoot NJ A double-edged sword: R loops as threats to genome integrity and powerful regulators of gene expression. *Genes Dev.* 28, 1384–1396 (2014). [PubMed: 24990962]

124. Lin Y, Dent SY, Wilson JH, Wells RD & Napierala M R loops stimulate genetic instability of CTG. CAG repeats. *Proc. Natl Acad. Sci. USA* 107, 692–697 (2010). [PubMed: 20080737]
125. Nakamori M, Pearson CE & Thornton CA Bidirectional transcription stimulates expansion and contraction of expanded (CTG)ⁿ(CAG) repeats. *Hum. Mol. Genet* 20, 580–588 (2011). [PubMed: 21088112]
126. Reddy K et al. Processing of double-R-loops in (CAG)_n (CTG)_n and *C9orf72* (GGGGCC)_n (GGCCCC) repeats causes instability. *Nucleic Acids Res.* 42, 10473–10487 (2014). [PubMed: 25147206]
127. Moreira MC et al. Senataxin, the ortholog of a yeast RNA helicase, is mutant in ataxia-ocular apraxia 2. *Nat. Genet* 36, 225–227 (2004). [PubMed: 14770181]
128. Chen YZ et al. DNA/RNA helicase gene mutations in a form of juvenile amyotrophic lateral sclerosis (ALS4). *Am. J. Hum. Genet* 74, 1128–1135 (2004). [PubMed: 15106121]
129. Martin-Tumasch S & Brow DA *S. cerevisiae* Sen1 helicase domain exhibits 5' to 3' helicase activity with a preference for translocation on DNA rather than RNA. *J. Biol. Chem* 290, 22880–22889 (2015). [PubMed: 26198638]
130. Belzil VV et al. Characterization of DNA hypermethylation in the cerebellum of c9FTD/ALS patients. *Brain Res.* 1584, 15–21 (2014). [PubMed: 24530272]
131. Liu EY et al. *C9orf72* hypermethylation protects against repeat expansion-associated pathology in ALS/FTD. *Acta Neuropathol.* 128, 525–541 (2014). [PubMed: 24806409]
132. McMillan CT et al. *C9orf72* promoter hypermethylation is neuroprotective: neuroimaging and neuropathologic evidence. *Neurology* 84, 1622–1630 (2015). [PubMed: 25795648]
133. Russ J et al. Hypermethylation of repeat expanded *C9orf72* is a clinical and molecular disease modifier. *Acta Neuropathol.* 129, 39–52 (2015). [PubMed: 25388784]
134. Naumann A, Kraus C, Hoogeveen A, Ramirez CM & Doerfler W Stable DNA methylation boundaries and expanded trinucleotide repeats: role of DNA insertions. *J. Mol. Biol.* 426, 2554–2566 (2014). [PubMed: 24816393]
135. Simone R, Fratta P, Neidle S, Parkinson GN & Isaacs AM G-quadruplexes: emerging roles in neurodegenerative diseases and the non-coding transcriptome. *FEBS Lett.* 589, 1653–1668 (2015). [PubMed: 25979174]
136. Wang J, Haeusler AR & Simko EA Emerging role of RNA*DNA hybrids in *C9orf72*-linked neurodegeneration. *Cell Cycle* 14, 526–532 (2015). [PubMed: 25590632]
137. Lin J et al. Stabilization of G-quadruplex DNA by C-5-methyl-cytosine in *bcl-2* promoter: implications for epigenetic regulation. *Biochem. Biophys. Res. Commun* 433, 368–373 (2013). [PubMed: 23261425]
138. Brcic J & Plavec J Solution structure of a DNA quadruplex containing ALS and FTD related GGGGCC repeat stabilized by 8-bromodeoxyguanosine substitution. *Nucleic Acids Res.* 43, 8590–8600 (2015). [PubMed: 26253741]
139. Wojciechowska M, Olejniczak M, Galka-Marciniak P, Jazurek M & Krzyzosiak WJ RAN translation and frameshifting as translational challenges at simple repeats of human neurodegenerative disorders. *Nucleic Acids Res.* 42, 11849–11864 (2014). [PubMed: 25217582]
140. Colussi TM et al. Initiation of translation in bacteria by a structured eukaryotic IRES RNA. *Nature* 519, 110–113 (2015). [PubMed: 25652826]
141. Ren Q et al. Alternative reading frame selection mediated by a tRNA-like domain of an internal ribosome entry site. *Proc. Natl Acad. Sci. USA* 109, E630–E639 (2012). [PubMed: 22247292]
142. Toulouse A et al. Ribosomal frameshifting on *MJD-1* transcripts with long CAG tracts. *Hum. Mol. Genet* 14, 2649–2660 (2005). [PubMed: 16087686]
143. Gaspar C et al. CAG tract of *MJD-1* may be prone to frameshifts causing polyalanine accumulation. *Hum. Mol. Genet* 9, 1957–1966 (2000). [PubMed: 10942424]
144. Davies JE & Rubinsztein DC Polyalanine and polyserine frameshift products in Huntington's disease. *J. Med. Genet* 43, 893–896 (2006). [PubMed: 16801344]
145. Stochmanski SJ et al. Expanded *ATXN3* frameshifting events are toxic in *Drosophila* and mammalian neuron models. *Hum. Mol. Genet* 21, 2211–2218 (2012). [PubMed: 22337953]

146. Girstmair H et al. Depletion of cognate charged transfer RNA causes translational frameshifting within the expanded CAG stretch in huntingtin. *Cell Rep.* 3, 148–159 (2013). [PubMed: 23352662]
147. Yu CH, Teulade-Fichou MP & Olsthoorn RC Stimulation of ribosomal frameshifting by RNA G-quadruplex structures. *Nucleic Acids Res.* 42, 1887–1892 (2014). [PubMed: 24178029]
148. Cammas A et al. Stabilization of the G-quadruplex at the VEGF IRES represses cap-independent translation. *RNA Biol.* 12, 320–329 (2015). [PubMed: 25826664]

Box 1 |**Origins of the *C9orf72* nucleotide repeat expansion**

The chromosome 9 open reading frame 72 (*C9orf72*) nucleotide repeat expansion (NRE) was first discovered through genomic sequencing of families with high penetrance of familial amyotrophic lateral sclerosis (ALS) or familial frontotemporal dementia (FTD)^{1,2}. Geographically separated and apparently unrelated families harbouring a *C9orf72* mutation showed a shared haplotype, suggesting that this mutation arose from a single founder¹. However, the origin of the *C9orf72* NRE has been reconsidered following the identification of patients carrying *C9orf72* mutations with no apparent family history of ALS, FTD or the shared founder haplotype⁹⁷. Pre-mutation carriers (asymptomatic carriers who have relatively larger numbers (> 30) of repeats than the standard population) can produce symptomatic offspring that carry grossly expanded *C9orf72* nucleotide repeats⁹⁷. This large expansion of repeats is proposed to occur either during parent–offspring transmission or in the early stages of embryonic development^{98,99}. Multiple lines of evidence indicate that there is a range of *C9orf72* repeat sizes in both asymptomatic and symptomatic individuals, showing that *C9orf72* repeats are intrinsically unstable¹⁰⁰. Therefore, it is possible that large spontaneous expansions may be inherited by offspring when the repeat sizes present in their asymptomatic parent approach an instability threshold.

Even a relatively small number of repeats (~70) can be symptomatic²⁰, and asymptomatic carriers with few repeats (~30) do show *C9orf72*-related pathology (but not TAR DNA-binding protein 43 (TDP43) mislocalization)⁵¹. Several studies of *in vitro* *C9orf72* NRE models have suggested that phenotypes depend on the repeat size, and it was recently suggested that the age of onset is also dependent on the repeat size¹⁰¹; however, this effect may be a result of the actions of additional genetic risk factors or epigenetic differences⁹⁷. For example, single-nucleotide polymorphisms in the transmembrane protein 106B (*TMEM106B*) gene were shown to be neuroprotective and associated with later disease onset in patients with FTD (but not ALS) who are carriers of the *C9orf72* NRE^{102,103}. Future genetic and epigenetic analyses of *C9orf72* ALS and *C9orf72* FTD carriers and their families will be essential.

Box 2 |**Genome instability and loss of function**

Genomic regions that contain repetitive DNA sequences, such as nucleotide repeat expansions (NREs), have been linked to approximately 40 neurological or neuromuscular genetic disorders¹⁰⁴⁻¹⁰⁷. In tissues from patients with these disorders, the molecular mass of NRE-containing regions of DNA can be variable, indicating genome instability — a tendency for the genomic region to be prone to DNA breaks, mutations and/or expansion or contractions. Often, this instability occurs in a tissue- or cell-specific manner^{108,109}.

Several mechanisms are proposed to contribute to genome instability, including increased sensitivity to spontaneous mutations, recombination and/or accumulation of single-stranded (ss) or double-stranded (ds) DNA breaks¹¹⁰. It is currently unclear how such mechanisms contribute to cell type-specific local genome instability^{99,110,111}. However, there is evidence to indicate that the formation of secondary or tertiary nucleotide structures such as hairpins, parallel and antiparallel G-quadruplexes (G4)¹¹²⁻¹¹⁸, R loops¹¹⁹⁻¹²⁴ and I-motifs increases the exposure of ssDNA to damaging environmental agents and can initiate repeat expansion or contraction. For example, upon local dsDNA melting, the formation of these alternative structural states on at least one of the two ssDNAs can lead to insertion or removal of repeat tracts as a result of DNA repair processes⁹⁹. Additionally, bidirectional transcription, which may be mechanistically linked to R loop formation^{125,126}, may contribute to expansion or contraction of repeats¹²⁵. Given that the chromosome 9 open reading frame (*C9orf72*) NRE sequence adopts many of these structural features (see the figure, top panel), it is unsurprising that genome instability was identified in patients carrying the mutation².

Several other proteins linked to neurodegeneration have also been linked to the maintenance of genome fidelity. For example, unresolved R loops require the actions of an RNA:DNA helicase, such as the amyotrophic lateral sclerosis (ALS)-related protein senataxin^{127,128}, to resolve their stability¹²⁹. These proteins may also provide mechanistic insight into the age-related genome defects and damage, as well as chromosomal spatial reorganization, that may lead to degeneration.

There is increasing evidence to indicate that the cell can respond to the presence of the *C9orf72* NRE by epigenetically silencing the mutant *C9orf72* allele. Specifically, increased repressive cytosine hypermethylation at the *C9orf72* NRE locus has been reported^{19,98,130-133}, and has been proposed to spread outward to the flanking DNA⁹⁸ (see the figure, bottom panel). DNA methylation can be neuroprotective and act as a disease modifier^{132,133}, suggesting that DNA hypermethylation may act against the progression of disease. There is also evidence to suggest that there is increased repressive histone hypermethylation at the *C9orf72* NRE locus in patients^{130,131}. These observations are consistent with repressive epigenetic changes observed in other NRE-linked diseases¹³⁴. However, the epigenetic effects of repressive methylation may be more complicated in the case of the *C9orf72* NRE. As outlined above, the formation of alternative DNA structures may alter heterochromatin and transcriptional levels at the NRE locus^{135,136}. Cytosine hypermethylation can stabilize the formation of G-

quadruplexes, thus increasing transcriptional silencing *in vitro*^{137,138}. Therefore, there may be a battle between the repressive epigenetic mechanisms acting to decrease local genomic and transcriptional instability induced by alternative nucleotide structures and further modulation of these epigenetic effects by these same alternative nucleotide structures (FIG. 2). Thus, epigenetic modifications aimed to alleviate the potential deleterious effects of the NRE may become compromised over time. However, functional studies specifically testing for a role of epigenetics in downstream cellular toxicity have not yet been performed.

Box 3 |**The production of dipeptide repeats**

The mechanisms that underlie repeat-associated non-ATG-dependent translation (RANT) and the cellular conditions that promote it are poorly understood. Two conditional constraints have been identified for the production of detectable dipeptide repeat (DPR) levels: the nucleotide repeat expansion (NRE) must reach a threshold length before RANT will occur, and the NRE must be able to adopt RNA secondary structures such as hairpins⁶⁰. Notably, all transcribed NRE sequences that have nearly self-complementary base pairing can adopt stable or dynamic RNA secondary structures, which implies that most NREs identified to date meet this criterion.

Recently, the mechanism underlying RANT has been proposed to parallel other mechanisms of non-traditional translation, such as the translation of RNA sequences that contain internal ribosome entry sites (IRESs)¹³⁹. However, IRES-mediated translation classically uses an AUG start site, and only a few IRES sequences have been examined in the absence of AUG start sites and/or for translation in all three reading frames^{140,141}.

An unusual translation feature found to occur in other NREs is ribosomal frameshifting, caused by the shifting of the ribosomal complex units by one nucleotide in either direction during translational elongation. Large tracts of (CAG) repeats were reported to lead to increased ribosomal frameshifting in several neurological disorders¹⁴²⁻¹⁴⁵. Experiments on (CAG) repeats demonstrated that there was an equal probability of the occurrence of a frameshift at any (CAG) codon. This effect results in a mixture of frameshift hybrids in which the frequency of ribosomal frameshifting is proportional to the (CAG) length¹⁴⁶. However, whether frameshifting occurs during RANT of NREs has not been shown, and the extent to which it contributes to the diversity of DPRs generated *in vitro* or *in vivo* is unknown.

In the case of the chromosome 9 open reading frame (*C9orf72*) NRE, the formation of diverse stable RNA structures could further impair ribosomal elongation and also presumably increase the probability of frameshifting. Importantly, the formation of RNA G-quadruplexes can enhance ribosomal frameshifting *in vitro* and *in vivo*, and this process may affect the translation of and relative abundance of specific DPRs generated from the G-quadruplex-forming *C9orf72* NRE¹⁴⁷. Furthermore, the presence of an RNA G-quadruplex has been shown to increase IRES-mediated translation; however, the formation of multiple RNA G-quadruplexes on a transcript can have an inhibitory effect on translation¹⁴⁸. Therefore, the structures that enhance the recruitment of the ribosomal complex to the NREs to facilitate RANT can also modulate translation by either blocking its progression or inducing ribosomal frameshifting during translation elongation in a NRE length-dependent manner.

GDP–GTP nucleotide exchange factor

(GEF). A protein that functions to stimulate the release of guanosine diphosphate (GDP) from a GTPase to allow subsequent binding of the guanosine triphosphate (GTP) to the active site.

Epigenetic silencing

Covalent DNA or histone modifications, such as DNA or histone methylation, that act to repress or silence a genomic region by reducing access of the transcriptional machinery to the DNA.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Polymerase processivity

A relative measure of the functional efficiency and rate of polymerase to process (replicate or transcribe) nucleic acid information.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Bidirectional transcription

Transcription that occurs simultaneously on both the positive and the negative strands of DNA, where the direction of RNA polymerase progression along each strand is either convergent or divergent.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Haplotype

A combination of alleles at different loci in the genome that tend to be inherited together because they show high linkage disequilibrium (often because they are physically close).

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Induced pluripotent stem cells

(iPSCs). Cells that are created from differentiated cell types — for example, fibroblasts — that are reprogrammed by a cocktail of transcription factors (or other approaches) back to a pluripotent state. These cells can then be differentiated into cells of distinct lineages — for example, neurons.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

R loops

A nucleic acid structure that occurs when a single strand of RNA invades a double-stranded DNA, forming an RNA:DNA hybrid that is stabilized through Watson–Crick base pairs, and leaves a displaced single-stranded DNA molecule.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

dsDNA melting

The separation of base-paired DNA strands through the disruption of Watson–Crick pairing by physical or thermodynamic means.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Interactomes

Sets of physical interactions occurring between two or more components.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

G-quadruplex

A structure composed of G-quartets, four planar guanine molecules that are stabilized by Hoogsteen base pairing, that are stacked on top of each other and stabilized by a central cation and π -stacking interactions between the G-quartets.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Hairpin

A nucleic acid molecule that contains a stem region, which can be composed of complimentary base pairing and be interspersed with single-stranded or loop elements, and contains a tight looped region (hairpin turn) at the end of the base-paired stem region.

Repeat-associated non-ATG-dependent translation

(RANT). The non-canonical translation of a repetitive RNA sequence initiated by an AUG start codon.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Nucleocytoplasmic trafficking

The bidirectional protein transport between the cytoplasm and the nuclear matrix through the nuclear pore complex.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Genetic modifier

A genetic variation in (*cis*) or outside (*trans*) a gene or genetic locus that alters the phenotypic expression of the gene.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

CRISPR/Cas technology

The essential components of the bacterial adaptive immunity, Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) and CRISPR-associated (Cas) genes, that are currently being utilized and developed for precise editing, regulating and targeting of genes in model systems and organisms.

Figure 1 l. The *C9orf72* nucleotide repeat expansion.

This schematic illustrates the chromosome 9 open reading frame (*C9orf72*) gene, showing the site of the nucleotide repeat expansion (NRE; red), and three annotated transcript variants. Transcripts V2 and V3 encode the long form of the C9orf72 protein, whereas transcript V1 encodes the short form. Exons are shown as blue boxes. Transcript variants V1 and V3 contain the NRE in the first intron of the gene and therefore generate pre-mRNA that contains the repeat expansion and can produce RNA foci. The presence of the *C9orf72* NRE in these transcripts also results in decreased use of the endogenous promoter of *C9orf72* transcription (depicted) and increased use of alternative promoters identified between the endogenous promoter and the repeats. The NRE is located within the predicted promoter region of variant V2 and therefore may contribute to the reduced transcription and highly decreased levels of mRNA.

Figure 2 | Cellular processes impaired by the *C9orf72* nucleotide repeat expansion.

Cellular features and pathways that are altered by the RNA and/or dipeptide repeat proteins (DPRs) generated from the chromosome 9 open reading frame (*C9orf72*) nucleotide repeat expansion (NRE) mutation are shown. Black and grey arrows indicate pathological features linked to the *C9orf72* RNA or DPRs, respectively. The repeat-containing RNA, which forms RNA foci, has been implicated in sequestering key RNA-binding proteins (RBPs). This process may lead directly to altered nucleolar function, changes in RNA processing or maturation and nucleocytoplasmic transport defects. DPRs have also been shown to lead to nucleocytoplasmic transport defects, as well as altering autophagy, nucleolar function and increasing endoplasmic reticulum (ER) stress. As shown in the right part of the image, RNA and DPRs generated from the *C9orf72* NRE mutation converge on pathways driving nucleocytoplasmic trafficking defects, including altered nuclear pore complex dynamics, import factors and nucleolar stress. *C9orf72* RNA can sequester nucleolar proteins, leading to nucleolar stress, or bind to nuclear pore complex proteins and disrupt nucleocytoplasmic trafficking. Cytoplasmic *C9orf72* RNA can be non-canonically translated through repeat-associated non-ATG-dependent translation (RANT), leading to the production of DPRs. Specific DPRs can impair nucleocytoplasmic trafficking and, when imported into the nucleus, can associate with nucleolar proteins to cause nucleolar stress. Additionally, both the *C9orf72* RNA and specific DPRs have been proposed to affect RNA splicing and/or editing.

Figure 3 | Identification of a diverse *C9orf72*•protein interactome.

The diagram compares five different chromosome 9 open reading frame (*C9orf72*)•RNA-binding protein (RBP) interactomes discovered using RNA pulldown or protein arrays^{13,18,54-56}. As shown, many unique proteins were identified that shared no overlap among the studies (see Supplementary information S2 (table)). Numbers indicate the number of protein candidates found to overlap between the indicated studies. The protein isoforms of heterogeneous nuclear ribonucleoprotein (hnRNP) H showed the most overlap between studies. Other proteins identified in at least three studies included splicing factor proline/glutamine-rich (SFPQ), interleukin enhancer-binding factor 2 (ILF2) and myelin basic protein (MBP). Of the overlapping RBPs, only hnRNP H has been shown to contribute to the increased alternative splicing events observed in patients carrying a *C9orf72* nucleotide repeat expansion (NRE)⁵⁷. Most RBP candidates have not been explored for potential functional consequences *in vivo*. Regions assigned no number designate comparisons in which there was no protein overlap.