Published in final edited form as:
Brain Res. 2014 October 10; 0: 3–14. doi:10.1016/j.brainres.2014.03.039.

RNA-protein interactions in unstable microsatellite diseases

Apoorva Mohan, **Marianne Goodwin**, and **Maurice S. Swanson***
Department of Molecular Genetics and Microbiology, Center for NeuroGenetics and the Genetics Institute, University of Florida, College of Medicine, Gainesville, FL 32610-3610, USA

Abstract

A novel RNA-mediated disease mechanism has emerged from studies on dominantly inherited neurological disorders caused by unstable microsatellite expansions in non-coding regions of the genome. These non-coding tandem repeat expansions trigger the production of unusual RNAs that gain a toxic function, which involves the formation of RNA repeat structures that interact with, and alter the activities of, various factors required for normal RNA processing as well as additional cellular functions. In this review, we explore the deleterious effects of toxic RNA expression and discuss the various model systems currently available for studying RNA gain-of-function in neurologic diseases. Common themes, including bidirectional transcription and repeat-associated non-ATG (RAN) translation, have recently emerged from expansion disease studies. These and other discoveries have highlighted the need for further investigations designed to provide the additional mechanistic insights essential for future therapeutic development.

Keywords

neurologic disease; microsatellite; RNA-mediated toxicity; bidirectional transcription; RNA foci; protein sequestration

1. Introduction: Unstable microsatellites in neurological disease

Approximately 50% of the human genome consists of repetitive elements of which 3% consists of simple sequence repeats (SSR) (Treangen and Salzberg, 2012). These SSRs, more commonly referred to as microsatellites, are composed of short tandem repeats of 2–10 base pairs (bp) that are dispersed throughout the genome. Microsatellites have been proposed to function at multiple steps in gene expression when present within or near genes in both prokaryotes and eukaryotes (Usdin, 2008). These steps include transcription where tandem repeats act as transcriptional regulatory elements (Meloni et al., 1998; Punga and Buhler, 2010), pre-mRNA splicing with modulation of splicing activity by microsatellite polymorphisms (Pagani et al., 2000) and translation by influencing 5′ UTR ribosomal

*Correspondence to: M.S. Swanson, Department of Molecular Genetics and Microbiology, University of Florida, College of Medicine, Cancer Genetics Research Complex, 2033 Mowry Road, University of Florida, College of Medicine, Gainesville, FL 32610-3610, USA mswanson@ufl.edu.

scanning (Ludwig et al., 2011). Errors in DNA replication, recombination and mismatch repair cause microsatellite instability leading either to repeat tract expansion or contraction. Importantly, the majority of microsatellite expansion diseases are neurological/neuromuscular disorders although many of the affected genes are expressed ubiquitously. For example, polyglutamine (polyQ) disorders, including Huntington disease (HD), spinocerebellar ataxias (SCA1, 2, 3, 6, 7, 17) and spinobulbar muscular atrophy (SBMA or Kennedy's disease), are caused by protein coding region CAG repeat expansions. These polyQ expansions induce protein aggregate formation leading to altered homeostasis of multiple cellular pathways (La Spada and Taylor, 2010; Nalavade et al., 2013). Alternatively, repeat expansion mutations also occur in the non-coding regions of genes, which include 5′ and 3′ untranslated regions (UTRs) and introns, and cause dominantly inherited disorders such as myotonic dystrophy types 1 and 2 (DM1, DM2), Fragile X-associated tremor/ataxia syndrome (FXTAS) and SCA types 8, 10 and 12 (Ranum and Cooper, 2006; Poulos et al, 2011). The recent discovery of a repeat expansion mutation in chromosome 9-linked amyotrophic lateral sclerosis and frontotemporal dementia (*C9ORF72* ALS/FTD) has highlighted the importance of elucidating the molecular mechanisms underlying these non-coding expansion disorders (DeJesus-Hernandez et al, 2011; Renton et al, 2011). One disease model proposes that non-coding mutations are deleterious at the RNA level (RNAopathy) because they fold into stable RNA structures that either inhibit or enhance the normal activities of important cellular factors. In contrast, RNA repeat expansions are also prone to a non-canonical type of protein translation, or repeat-associated non-ATG (RAN) translation, and the resulting unusual peptides, which for SCA8 are composed of ATXN8 polyQ, polyserine and polyalanine tracts, may induce disease-associated pathology (proteinopathy) (Zu et al, 2011). These and other studies suggest that the full range of molecular pathways that are compromised in these diseases remains to be determined. In this review, we discuss the potential interplay between these pathogenic mechanisms with a focus on RNA gain-of-function, RNA binding proteins and the available model systems to study RNA toxicity.

2. Origins of microsatellite expansions

Microsatellite intergenerational instability is a major feature underlying expansion diseases with subsequent generations experiencing anticipation, which is characterized by increased disease severity and earlier onset age triggered by an increase in repeat length (Pearson et al., 2005). Even within an individual, repeat tract length exhibits somatic mosaicism, or heterogeneity within and between tissues, with a positive correlation existing between repeat number and tissue-specific pathology in diseases like HD and DM1 (Kennedy et al., 2003). Although our understanding of the mechanisms underlying repeat instability is incomplete, considerable evidence points to pivotal roles for errors induced during DNA replication and repair depending on tissue developmental and proliferative state (Budworth and McMurray, 2013; Lopez Castel et al., 2010; Mirkin, 2007; Usdin, 2008). During DNA replication, instability is influenced by both *cis*- and *trans*-acting factors and cell studies have shown that *cis*-elements (repeat sequence, tract length, distance from the replication origin and replication direction) can affect repeat instability by modulating replication fork dynamics during lagging strand synthesis (Cleary et al., 2002). Additionally, epigenetic factors,

including the CpG methylation status of CTCF binding sites flanking the repeat locus and histone modifications that affect chromatin structure and arrangement, impact repeat instability (Dion and Wilson, 2009; Libby et al., 2008). Repeats also have the potential to form non-canonical structures, such as non-B-form DNA-like triple helices, G-quadruplexes, intra-strand hairpins and slipped strand structures, which result in replication fork stalling and template switching (Lopez Castel et al., 2010).

Repeat instability in terminally differentiated cells is also affected by transcription. The majority (>80%) of the genome undergoes transcription (Hangauer et al., 2013) and widespread antisense transcription has been reported for many loci, including a majority of microsatellite disease-associated genes (Batra et al., 2010; Budworth and McMurray, 2013). Both CTG and CAG repeat expansions have been shown to enhance repeat instability by several fold upon unidirectional and bidirectional transcription induction in mammalian cells (Lin et al., 2006; Nakamori et al., 2011). Although the mechanisms underlying transcription-induced repeat instability are poorly defined, it is likely that the separation of DNA strands during transcription results in secondary structure formation by the repeats followed by protein-induced stabilization of these structures (McIvor et al., 2010). During bidirectional transcription, head-on collision between RNA polymerases may also occur and cause stalling and activation of downstream DNA damage response pathways (Lin and Wilson, 2011).

Trans-acting factors, including the mismatch match repair (MMR) proteins MSH2, MSH3, MSH6 and PMS2, are also critical drivers of repeat instability. Repeat contractions and stabilization occur in mice harboring CTG•CAG repeats when they are crossed with either *Msh2* or *Msh3* null mice or mice deficient in Msh2 ATPase activity indicating a role for these proteins in promoting instability in DM1 (Pearson et al., 2005). Similarly, Msh2 is important for promoting both intergenerational and somatic repeat expansions in a FXTAS model expressing CGG•CCG repeats (Lokanga et al., 2014). Naturally occurring *Msh3* polymorphisms may act as a modifier of CAG repeat instability by interfering with the stability of the Msh3 protein (Tome et al., 2013) and the resulting variations in protein levels may account for some aspects of region-specific instability seen in the striatum of HD patient brains (Gonitel et al., 2008; Pinto et al., 2013). How do these proteins influence repeat expansions? Studies in *S. cerevisiae* demonstrate that Msh2 and Msh3 alter the activities of proteins mediating Okazaki fragment processing resulting in small yet incremental expansions (Kantartzis et al., 2012). Since repeat instability is a complex phenomenon regulated by multiple DNA metabolic pathways that influence the various tissue and developmental stage specific expansion/contraction patterns, it is critical to understand the underlying mechanism. Although repeat instability plays the primary role in determining the course of disease, functional impairment for some non-coding expansion disorders lies at the next level upon transcription of the DNA to produce RNAs with expanded repeats.

3. Overview of RNA gain-of-function disease mechanisms

RNA sequence and higher order structures are two critical features that determine how the processing of a particular RNA occurs (Bugaut and Balasubramanian, 2012). Alterations in

normal RNA sequence/structure by mutations can interfere with multiple steps in RNA biogenesis as well as the functions of the mature RNA and result in widespread dysregulation. This phenomenon is exemplified in non-coding repeat expansion diseases in which repeat expansions lead to a deleterious RNA gain-of-function (Fig. 1).

3.1. RNA foci

A hallmark of most non-coding repeat expansion diseases is the formation of distinct cellular aggregates of mutant RNA, such as the nuclear foci in DM1, DM2 and *C9ORF72* ALS/FTD, or the inclusions containing mutant RNA in FXTAS (DeJesus-Hernandez et al., 2011; Greco et al., 2002; Taneja et al., 1995). RNA foci are dynamic structures characterized by random formation and dissociation with the formation phase mediated by RNA binding proteins, as shown for CUG RNA foci in stably transfected C2C12 myoblasts (Querido et al., 2011). The dynamics of RNA foci for mutant repeat RNAs associated with other repeat expansion diseases have not yet been determined. These foci contain multiple copies of mutant repeat RNA and its interacting protein partners in a complex (Taneja et al., 1995; Wojciechowska and Krzyzosiak, 2011). Morphologically, RNA foci and inclusions are distinct. The C(C)UG^{exp} foci in DM and the rGGGGCC^{exp} in *C9ORF72* ALS/FTD appear more compact, likely due to tighter interactions between the RNA and bound proteins, while rCGG^{exp}-containing inclusions in FXTAS are larger and more diffuse structures that contain many proteins, including lamins and MBNL1 (Hagerman, 2013). FXTAS brains also contain ubiquitin-positive neuronal protein inclusions that contain polyglycine (FMRpolyG) generated by RAN translation (Todd et al., 2013). An important question to address in future studies will be whether the structural distinctions between RNA foci and inclusions reflect fundamental differences in the pathogenic pathway.

Although there is a lack of information on the structures of specific RNA-protein complexes in RNA foci, several studies have analyzed synthetic repeat oligomers *in vitro*. Biochemical and enzymatic structure probing, as well as biophysical studies using CD and UV spectroscopy, reveal that $(CNG)_{20}$ transcripts form stable, but slippery, hairpins with the exception of the most stable CGG triplet (Krzyzosiak et al., 2012; Sobczak et al., 2010). Crystal structures of short CUG, CAG, CGG oligomers indicate that they adopt an A-form helical structure with non-canonical G-G, A-A and U-U wobble base pairing flanked by stabilizing GC base pairs (Kiliszek et al., 2010; Kiliszek et al., 2011; Mooers et al., 2005). DM2-associated CCUG repeats also form RNA stem-loop structures like CNG repeats but with CU mismatches and lower stability (Sobczak et al., 2003). Notably, some sequestered proteins have been suggested to preferentially recognize CNG RNA sequences and not secondary structures. For example, the MBNL proteins bind to GC base pairs and then fold the CUG^{exp} RNA into a pseudo A-form helix (Teplova and Patel, 2008). Recent reports have also proposed a more complex structure for ALS/FTD-associated GGGGCC^{exp} RNA. Studies using 1D ^{1}H NMR and CD spectroscopy suggest that this expansion RNA forms thermostable inter- and intra-molecular G-quadruplexes with a parallel orientation (Fratta et al., 2012; Reddy et al., 2013). The structural details of GGGGCC^{exp} RNA and its antisense transcript CCCCGG^{exp}, both of which form RNA foci in patient cells and tissues, requires additional characterization although limitations exist with these *in vitro* analyses. For example, prior studies have used repeats in the normal size range and the multiple RNA and

protein factors that likely influence expansion RNA structures *in vivo* are absent. Nevertheless, *in vitro* evidence for higher-order structure formation by RNA repeat expansions provides intriguing insights into the nature of RNA toxicity.

Another outstanding question is whether RNA foci have a causal role in disease progression or are innocent bystanders. For DM1, the expression of mutant RNAs above a repeat length threshold (Hamshere et al., 1997) leads to foci formation and MBNL protein sequestration followed by disruption of normal MBNL functions, including alternative splicing regulation (Ranum and Cooper, 2006). For *C9ORF72* ALS/FTD, antisense oligonucleotides (ASOs) targeting rGGGGCC^{exp} cause a reduction of sense RNA foci in patient-derived iPS cells with a concomitant rescue of the cells from glutamate-induced toxicity (Donnelly et al., 2013). These results indicate that RNA foci are not simply biomarkers but important mediators of toxicity. However, an opposing viewpoint comes from studies on a DM mouse model expressing a GFP transgene containing the DMPK 3′ UTR with a normal length $(CTG)_5$ repeat (Mahadevan, 2012). These mice fail to develop RNA foci but still exhibit pathological features associated with DM1 disease including muscle pathology, splicing deficits and cardiac conduction defects (Mahadevan et al., 2006). Interestingly, another DMPK transgenic model expressing a normal length $(CTG)_{20}$ repeat does not exhibit a pathological phenotype (Seznec et al., 2001). Since it is not currently clear if the formation of RNA foci is an essential step in pathogenesis, the possibility that overexpression of the GFP-DMPK 3′ UTR transgene also results in Mbnl functional deficiency should be tested by either rAAV-induced Mbnl1 overexpression or crossing this transgenic line to MBNL1 overexpression mice (Chamberlain and Ranum, 2012; Kanadia et al., 2006).

3.2. RNA gain-of-function

Mounting evidence indicates that microsatellite expansions in DM1/DM2, FXTAS and *C9ORF72* ALS/FTD are pathogenic at the RNA level (Echeverria and Cooper, 2012; Poulos et al., 2011; Sicot and Gomes-Pereira, 2013) (Fig. 1). These diseases are characterized by a dominant inheritance pattern and are caused by mutations in the non-coding regions of their respective genes, making a conventional protein loss-of-function mechanism unlikely. Additionally, C(C)UG^{exp}, rCGG^{exp}, rGGGGCC^{exp} RNAs accumulate in either distinct nuclear foci or inclusions in patient cells. Although full-length mutant *DMPK* mRNA accumulates in nuclear RNA foci, the majority of DM-associated disease symptoms are caused by mutant RNA expansions and not by *DMPK* haploinsufficiency (Jansen et al., 1996; Mankodi et al., 2000; Orengo et al., 2008; Reddy et al., 1996). For the CCTG^{exp} and GGGGCC^{exp} intronic expansions in DM2 and *C9ORF72* ALS/FTD, respectively, the repeats undergo splicing and accumulate in nuclear RNA foci in the absence of flanking intronic sequences (Donnelly et al., 2013; Margolis et al., 2006). Similarly, FXTAS is characterized by ubiquitin-positive, neuronal and astrocytic intranuclear inclusions (2–5 μm) containing FMR1 mRNA and various proteins coupled with a 2–10 fold elevation in FMR1 mRNA levels (Hagerman and Hagerman, 2013). In addition, *Drosophila* and mouse models expressing expansion RNAs exhibit disease-relevant symptoms irrespective of the gene context or the flanking sequence. Neurodegeneration and ubiquitin-positive intranuclear inclusions are observed in *Drosophila* expressing $rCGG_{90}$ repeats flanking an EGFP transgene (Jin et al., 2003) and mice expressing CTG_{250} repeats driven by the human

skeletal actin promoter (HSA^{LR}) develop myotonia and numerous ribonuclear foci in a transgene expression-dependent manner providing strong evidence for a gain-of-function at the RNA level (Mankodi et al., 2000). In the following sections, we discuss specific mechanisms of RNA-mediated toxicity and its potential adverse effects using three neurological diseases as examples, DM, FXTAS and *C9ORF72* ALS/FTD.

4. Myotonic dystrophy

Myotonic dystrophy has served as a paradigm for repeat expansion diseases caused by RNA gain-of-function mechanisms, also referred to as RNA dominance (Caillet-Boudin et al., 2014; Poulos et al., 2011; Ranum and Cooper, 2006). DM1 is the most common form of adult onset muscular dystrophy (1/8000 worldwide) (O'Rourke and Swanson, 2009). DM1 and DM2 are multisystemic disorders with symptoms that include progressive skeletal muscle wasting, delayed muscle relaxation (myotonia), cardiac arrhythmias, insulin resistance, gastrointestinal problems and CNS symptoms (cerebral atrophy, hypersomnolence, memory deficits, cognitive/behavioral abnormalities and intellectual disability in the congenital disease). DM is caused by unstable repeat expansions in the non-coding region of two genes: 1) a CTG^{exp} in the 3′ UTR of the *dystrophia myotonica protein kinase* (*DMPK*) gene in DM1; 2) a $CCTG^{exp}$ in the first intron of *CCHC-type zinc finger nucleic acid binding protein* (*CNBP/ZNF9*) gene in DM2 (Liquori et al., 2001; Ranum and Cooper, 2006). For DM1, the normal CTG repeat length ranges from 5–37, while individuals with >50 repeats exhibit the classical disease symptoms of DM1 and >1000 repeats is associated with the severe congenital form (CDM). For DM2, the pathogenic range of CCTG repeats varies from 75 – ~11,000 in patients compared to 11–26 repeats in normal individuals. Anticipation is a prominent feature of DM1 but not DM2, although somatic mosaicism and intergenerational instability have been widely reported in both DM1 and DM2 (Udd and Krahe, 2012).

4.1. DM pathogenic mechanisms

The expression of $C(C)UG^{exp}$ RNA, which folds into a stable stem-loop structure, alters the activities of two developmentally regulated RNA binding protein families, MBNL and CELF, causing misregulation of multiple cellular pathways (Echeverria and Cooper, 2012; Fernandez-Costa et al., 2013; Kalsotra et al., 2014; Rau et al., 2011; Wang et al., 2012) (Fig. 1). Although CELF1, an alternative splicing factor that promotes fetal splicing patterns, is not sequestered by CUG^{exp} RNAs, CELF1 protein levels increase in DM1 heart and muscle tissues through a mechanism mediated by protein kinase C (PKC) (Kuyumcu-Martinez et al., 2007). The details of how CUG^{exp} RNA activates PKC, and if this activation is specific to DM1, still needs to be clarified. In contrast, the MBNL proteins, which activate adult splicing patterns, are sequestered by $C(C)UG^{exp}$ RNAs (Miller et al., 2000) and thus the combination of an increase in CELF1 and a decrease in MBNL activity causes a shift to fetal splicing patterns of specific target transcripts in adult tissues (Osborne and Thornton, 2006). These two events make alternative splicing dysregulation a major pathogenic event in DM. In addition, abnormal DNA methylation (Castel et al., 2011), altered miRNA and mRNA expression (through decreased expression of transcription factors) (Fernandez-Costa et al., 2013; Kalsotra et al., 2014; Rau et al., 2011), bidirectional transcription of the repeats (Batra

et al., 2010), and repeat associated non-ATG translation (RAN translation) (Zu et al., 2011) have also been implicated in DM pathogenesis (Fig. 1).

4.2. RNA gain-of-function models for DM

The first convincing proof that DM is an RNA-mediated disease was obtained from the *HSA*^{LR} mouse model, which expresses a human skeletal muscle (*HSA*) transgene with CTG_{250} repeats specifically in the skeletal muscle (Mankodi et al., 2000). This model shows features reminiscent of DM including myopathy, myotonia, intranuclear CUG RNA foci and centralized myonuclei with a positive correlation between pathology and transgene expression. Additional mouse models have also been generated to overcome the limitations of this model, which include a relatively short CTG^{exp}, the absence of the endogenous DMPK 3′ UTR sequence and restricted muscle expression (Gomes-Pereira et al., 2011). These models include Cre-inducible transgenic mice expressing interrupted CTG_{960} repeats (EpA960) or CTG_0 repeats (EpA0) and transgenic mice carrying insertions of a ~45 kb human genomic region with 20, 55, and 300 CTG repeats (Seznec et al., 2000; Wang et al., 2007). DMSXL, derived from the DM300 line, is another transgenic model with >1000 CTG repeats due to intergenerational instability that led to ‘big jumps’ in CTG repeat number (Gomes-Pereira et al., 2007). These models exhibit different aspects of DM disease, such as muscle pathology, cardiac conduction problems, behavioral abnormalities accompanied by intranuclear RNA foci pathology and some tissue-specific splicing deficits (Gomes-Pereira et al., 2011; Hernandez-Hernandez et al., 2013; Huguet et al., 2012). Taken together, these mouse models serve as valuable tools to gain insight into DM disease mechanisms and for preclinical assessment of therapeutics targeting repeat expansion RNAs.

An important consequence of C(C)UG^{exp} RNA toxicity is the sequestration of MBNL proteins resulting in their functional insufficiency. The MBNL proteins were first discovered as factors that bind to CUG repeat RNA in a length-dependent manner *in vitro* and accumulate in nuclear RNA foci in patient myoblasts (Miller et al., 2000). Several *Mbnl* knockout (KO) mouse models have validated this sequestration hypothesis including *Mbnl1*$^{\Delta E3/\Delta E3}$ isoform KO mice, which develop DM-relevant and multisystemic symptoms including myotonia, ocular dust-like cataracts and abnormal splicing of developmentally regulated splicing events (Kanadia et al., 2003). Moreover, recombinant adeno-associated virus (rAAV)-mediated Mbnl1 overexpression in the *HSA*^{LR} model results in phenotypic rescue of myotonia and correction of dysregulated splicing (Kanadia et al., 2006). In contrast to Mbnl1, Mbnl2 is highly expressed in the adult CNS but not in skeletal muscle and Mbnl2 is the major Mbnl family member that regulates alternative splicing in the brain. *Mbnl2*$^{\Delta E2/\Delta E2}$ isoform KO mice develop DM-associated neurological symptoms including hypersomnia, learning/memory deficits, altered GABA sensitivity and disease specific mis-splicing (Charizanis et al., 2012). The differences observed in these *Mbnl* KO models indicate tissue-specific splicing regulation by the Mbnl family. Interestingly, Mbnl2 compensates for loss of Mbnl1 activity in heart and skeletal muscle while the absence of both proteins leads to embryonic lethality (Lee et al., 2013a). *Mbnl1; Mbnl2* conditional double KO (DKO) mice are viable but exhibit phenotypes not observed in single *Mbnl* KOs, such as muscle wasting and cardiac conduction defects, suggesting that DM is caused by

compound loss of MBNL function triggered by the production of toxic C(C)UG^{exp} RNAs (Lee et al., 2013a).

A major manifestation of DM disease in the brain is neurofibrillary degeneration associated with the aggregation of abnormally phosphorylated Tau and reduced splicing of MAPT exons 2, 3 and 10 (Jiang et al., 2004; Sergeant et al., 2001). In agreement with the possibility that DM results from compound loss of MBNL function, both MBNL1 and MBNL2 are required for enhancement of MAPT exon 2 splicing (Carpentier et al., 2014). This result suggests that heterotypic interactions between these MBNL proteins, possibly mediated by MBNL C-terminal regions (Tran et al., 2011; Yuan et al., 2007), are required for MAPT splicing regulation (Carpentier et al., 2014). Still, questions remain about the role of MBNL proteins in CDM pathogenesis and if MBNL titration affects other biochemical pathways, including RNA localization (Adereth et al., 2005; Wang et al., 2012). In addition to the MBNL family, other RNA-binding proteins have been implicated in either abnormal splicing regulation in DM1 (hnRNP H, STAU1), modulation of MBNL binding to CUG repeats (p68/DDX5) and inhibition of CUG^{exp} mRNA export from the nucleus (Kim et al., 2005; Laurent et al., 2012; O'Rourke and Swanson, 2009; Paul et al., 2006; Ravel-Chapuis et al., 2012). Further development of both loss-, and gain-of-function animal models are required to test disease-specific roles for these proteins.

5. Fragile X and Fragile X-associated tremor/ataxia syndrome

Prominent examples where microsatellite expansion length determines disease presentation and onset are the neurological disorders Fragile X syndrome (FXS) and Fragile X-associated tremor/ataxia syndrome (FXTAS) (Hagerman and Hagerman, 2013). FXS, a neurodevelopmental disorder that is the most common form of inherited intellectual disability, is triggered by $(CGG)_{>200}$ expansions in the 5′ UTR of the *FMR1* gene leading to transcriptional silencing and subsequent loss of the encoded protein, Fragile X mental retardation protein (FMRP). Interestingly, intermediate expansions of 55–200 CGG repeats, previously termed the premutation range, result in the late adult-onset neurodegenerative disorder FXTAS. The clinical phenotypes of FXTAS are distinct from FXS and >30% of these 'premutation' carriers experience gait ataxia, progressive intention tremor, parkinsonism and cerebral atrophy (Hagerman, 2013).

5.1. FXTAS pathogenic mechanisms

After the initial discovery of the FXTAS premutation, the observation that mutant *FMR1* transcripts exhibit increased levels in FXTAS (Tassone et al., 2000) and form relatively large intranuclear inclusions in multiple organs (Greco et al., 2002) led to the suggestion that an RNA toxicity mechanism underlies FXTAS pathogenesis (Hagerman and Hagerman, 2013). Further studies have shown the presence of additional proteins, including splicing regulators and miRNA processing factors in these inclusions, but the significance of the potential interactions of the factors with CGG RNAs remains elusive. Moreover, rCGG^{exp} expression has been proposed to dysregulate lamin A/C, the major component of the nuclear lamina (Arocena et al., 2005). In neurons and skin fibroblasts cultured from FXTAS patients, both the percentage of soluble lamin and the nuclear lamin architecture is disrupted suggesting that certain features of FXTAS, such as peripheral neuropathy could result from

a functional laminopathy (Garcia-Arocena et al., 2010). In large premutation carriers (>120 CGG repeats), moderate reduction in FMRP protein levels is observed, in spite of elevated *FMR1* mRNA levels (Hagerman, 2013). It has been reported that some aspects of the disease, like cognitive/behavioral dysfunction, are due to a decrease in FMRP levels in the amygdala (Hessl et al., 2011). Similar to DM and SCA8, CGG^{exp} RNA undergoes RAN translation producing polyglycine in patient tissues and $rCGG^{exp}$-induced toxicity is seen in cell and animal repeat expressing models (Todd et al., 2013). The transcription of long non-coding (lnc) RNAs and the occurrence of antisense transcription at the *FMR1* locus, as well as RAN translation of those transcripts, further widens the repertoire of processes associated with neurotoxicity (Ladd et al., 2007; Pastori et al., 2014). Therefore, a combination of factors, some of which are direct effects of the $rCGG^{exp}$ expression, may act synergistically to promote disease onset in FXTAS.

5.2. FXTAS RNA gain-of-function models

One proposed disease mechanism for FXTAS is an $rCGG^{exp}$ gain-of-function similar to that previously described for DM1, in which mutant FXTAS-associated RNAs sequester protein factors required for normal cellular pathways. To delineate the effects of RNA toxicity several transgenic and knockin model systems have been generated (Hunsaker et al., 2012). A *Drosophila* transgenic model expressing CGG_{90} repeats flanking EGFP has been developed which shows progressive eye neurodegeneration in a repeat-length and transgene dose-dependent manner (Jin et al., 2003). The CGG_{98} knockin mouse, in which the endogenous $(CGG)_8$ repeats of the *Fmr1* gene are replaced with $(CGG)_{98}$, exhibits FXTAS-relevant phenotypes including increased *Fmr1* mRNA levels, reduced Fmrp protein expression, ubiquitin-positive neuronal and glial intranuclear inclusions, age-dependent cognitive decline and behavioral abnormalities (Willemsen et al., 2003). Recent studies with another knockin line, $(CGG)_{150}$ mice, show that this expansion causes impaired migration and differentiation of various embryonic neocortical cells, thus affecting early brain development (Cunningham et al., 2011). This embryonic defect could account for the developmental problems observed in premutation carrier children. The contribution of CGG^{exp} RNA toxicity to the neurodegenerative phenotype in FXTAS is not clearly understood from these mouse models due to the decreased Fmrp protein levels. Thus, a transgenic mouse model ectopically expressing CGG_{90} repeats outside of context of *Fmr1* gene in Purkinje neurons, while maintaining normal Fmrp levels, has been generated (Hashem et al., 2009). Interestingly, in addition to intranuclear inclusions, the mice exhibit Purkinje cell loss accompanied by axonal swelling and age-dependent neuromotor learning deficits demonstrating the importance of $rCGG^{exp}$ toxicity in inclusion formation and in mediating neurotoxicity.

An early attempt to define the composition of FXTAS inclusions isolated from autopsied FXTAS brains used mass spectrometry and identified ~20 proteins, including MBNL1, hnRNP A2/B1 and lamin A/C (Iwahashi et al., 2006). Alternative experimental approaches such as RNA affinity pulldowns, have uncovered additional rCGG repeat binding factors (Sam68, Purα, DGCR8, DROSHA) (Hagerman, 2013; Jin et al., 2007; Sellier et al., 2010; Sellier et al., 2013). Overexpression of Purα, a transcriptional activator with probable roles in DNA replication and recombination, and hnRNPA2/B1, a ubiquitously expressed hnRNP

involved in pre-mRNA processing and RNA trafficking, mitigates the neurodegenerative phenotypes seen in *Drosophila* transgenic rCGG^{exp} lines (Jin et al., 2007; Sofola et al., 2007). Although Purα knockout mice undergo premature death and show severe neurological features, such as spontaneous seizures and tremors (Khalili et al., 2003), the relevance of Purα sequestration to FXTAS phenotypes remains unclear. For Sam68, a KH domain-containing alternative splicing factor, there is evidence for splicing alterations of some of its targets (SMN2, ATP11B) in FXTAS patients (Sellier et al., 2010), though the downstream effect of this dysregulation and the effect of sequestration on other Sam68 targets remains to be determined. In addition, DGCR8 and DROSHA, the proteins central to pri-miR processing are partially sequestrated in intranuclear rCGG^{exp}-containing inclusions in FXTAS patients and mammalian cells expressing premutation repeats, accompanied by global reduction of mature miRNA levels (Sellier et al., 2013). Overexpression of DGCR8, but not DROSHA, Sam68 or MBNL1, rescues CGG^{exp}-induced toxic phenotypes including reduced cell viability and decreased dendritic complexity in E18 cortical neurons suggesting that DGCR8 loss-of-function is an important feature of FXTAS pathogenesis. However, important questions remain. Are rCGG^{exp}-containing inclusions functionally equivalent to RNA foci? How do these various proteins interact with rCGG^{exp} in these inclusions and what is the extent of their sequestration? What are the pathogenic roles of the antisense transcript *ASFMR1* and does this mutant RNA also sequester factors, form inclusions and undergo RAN translation? Additional mechanistic approaches using different cellular and animal models are required to better understand these fundamental questions.

6. *C9ORF72*-linked amyotrophic lateral sclerosis and frontotemporal dementia

Recently, a non-coding GGGGCC^{exp} mutation in the first intron of the *C9ORF72* gene was discovered as a major cause of ALS/FTD (DeJesus-Hernandez et al., 2011; Renton et al., 2011). ALS is a debilitating neurodegenerative disorder in which upper and lower motor neuron death results in muscle weakness, wasting, difficulty in swallowing/breathing leading to paralysis and death usually within 3–5 years of onset. FTD causes the second most common form of presenile dementia (age of onset <65 years of age) and is a complex disorder affecting language, cognitive and behavioral skills. Approximately 40% of FTD patients suffer from ALS-like motor symptoms and 50% of ALS patients exhibit FTD-associated behavioral and personality changes (Ling et al., 2013). Although there is an overlap in patients affected with both of these conditions, it is interesting to note the degree of phenotypic variability existing in the remainder of the *C9ORF72* expansion carrier population. It is likely that genetic modifiers play a major role in determining disease presentation. For example, an allelic variant of *TMEM106B* protects carriers from developing FTD but not ALS (van Blitterswijk et al., 2014). Importantly, the differential effects of genetic modifiers is highlighted by the discovery that *TMEM106B* variants also act as risk factors for frontotemporal lobar degeneration with neuronal inclusions of hyperphosphorylated and ubiquinated TDP-43 (FTLD-TDP) (Van Deerlin et al., 2010).

6.1. C9ORF72 ALS/FTD disease mechanisms

Evidence that both C9ORF72 sense rGGGGCC^{exp} and antisense rCCCCGG^{exp} RNAs form nuclear RNA foci in neuronal and non-neuronal cells suggests a toxic RNA gain-of-function and protein sequestration mechanisms for ALS/FTD (Lagier-Tourenne et al., 2013). *In situ* hybridization experiments suggest that the incidence of sense foci is greater than antisense foci, but the number of antisense foci per cell is greater (Mizielinska et al., 2013). This result should be considered with caution because of differences in FISH probe affinity for sense and antisense RNAs. Nonetheless, it will be interesting to differentiate the pathogenic effect exerted by these individual entities. Transcriptome analysis of patient fibroblasts treated with ASOs targeting C9ORF72 sense RNA reveals that the disease-specific RNA signature is not reversed suggesting a possible role for antisense transcripts in pathogenesis (Lagier-Tourenne et al., 2013).

The observation that expression of certain *C9ORF72* transcripts are downregulated upon repeat expansion due to epigenetic modifications suggests that *C9ORF72* haploinsufficiency is a plausible disease mechanism (Belzil et al., 2013). Homology modeling predicts that C9ORF72 is related to DENN proteins belonging to the Rab GEF protein family, which are important for vesicular trafficking (Levine et al., 2013). The physiological function of the protein, and its distribution across tissues, is the focus of current studies and model systems are also being developed to address this question. LacZ reporter mice reveal *C9ORF72* expression in neuronal regions sensitive to neurodegeneration, (ventral horn of the spinal cord, cortical layers, hippocampus) but absent in non-neuronal cells like microglia and astrocytes (Suzuki et al., 2013). For *C9ORF72* haploinsufficiency models, knockdown of a zebrafish *C9ORF72* orthologue by antisense morpholinos and *C. elegans* null mutations of *Alfa-1*, a *C9ORF72* orthologue, have been developed (Ciura et al., 2013; Therrien et al., 2013). These models exhibit neurodegeneration phenotypes such as age-dependent motility defects, paralysis and motor neuron axonal degeneration. Contrary to the prediction of the haploinsufficiency model, reduction of C9ORF72 RNA levels in mice by ASOs specifically in the CNS is well tolerated with no significant pathological/behavioral changes (Lagier-Tourenne et al., 2013). However, the effect of sense RNA reduction on the corresponding C9ORF72 protein levels is not clear in this model because of the lack of specific anti-C9ORF72 antibodies. Cumulatively, these results support the need for additional *C9ORF72* loss-of-function models to clarify the contribution of haploinsufficiency to *C9ORF72* ALS/FTD.

Another pathogenic mechanism that has been linked to *C9ORF72* ALS/FTD mutations is RAN translation of GGGGCC^{exp}/ CCCCGG^{exp} RNAs and the production of six different dipeptide proteins that form intranuclear and cytoplasmic aggregates in various brain regions (Ash et al., 2013; Mori et al., 2013b; Zu et al., 2013). The majority of cells expressing these RAN proteins do not contain C9ORF72 sense/antisense RNA foci indicating a mutually exclusive mechanism in which the transcribed repeats are either sequestered into foci or are exported to the cytoplasm for RAN translation (Gendron et al., 2013). This observation suggests that some cell types may preferentially express factors that promote the formation and nuclear retention of toxic RNAs while other cells, which are deficient in these factors, are permissive for nuclear export and RAN translation. However, the detailed relationship

between RNA foci and RAN translation requires additional studies in cell and animal models of *C9ORF72* ALS/FTD disease.

6.2. RNA gain-of-function models for C9ORF72 ALS/FTD

A *Drosophila* model expressing $(GGGGCC)_{30}$ repeats flanking EGFP shows severely disrupted eye morphology and locomotor defects compared to control flies expressing $(GGGGCC)_3$ – EGFP (Xu et al., 2013). One limitation of this model is that this repeat size may be in the normal range so the effects of expanded repeats are still unknown. Efforts to generate vertebrate zebrafish and mouse models expressing expanded GGGGCC and CCCCGG repeats are ongoing.

A number of RNA binding proteins have been identified that interact with $rGGGGCC^{exp}$ RNA. Examples include: 1) hnRNP A3, a protein involved in cytoplasmic RNA trafficking; 2) Purα, a transcriptional activator; 3) ADARB2, a protein with homology to adenosine deaminases involved in adenosine to inosine RNA editing; 4) hnRNP H, a protein involved in pre-mRNA processing (Gendron et al., 2014). Both hnRNP A3 and Purα bind to GGGGCC RNA *in vitro*, but colocalization with nuclear RNA foci and loss-of-function is not observed in patients. Interestingly, hnRNP A3 is a component of the p62 positive cytoplasmic inclusions observed in patient brain and Purα forms intranuclear inclusions and rescues the neurodegenerative eye phenotype in $(rGGGGCC)_{30}$-EGFP expressing flies implicating them in pathogenesis (Mori et al., 2013a; Xu et al., 2013). A proteome array hybridized with $(rGGGGCC)_{6.5}$ RNA identified ADARB2, a protein with a possible regulatory role in RNA editing. ADARB2 colocalizes with C9ORF72 sense RNA foci in both patient and iPS cell lines and is important for RNA foci formation in iPS cells (Donnelly et al., 2013). However, the extent of ADARB2 sequestration and the downstream pathways possibly affected by sequestration of this protein requires further study. In addition, hnRNP H, a protein originally classified as a poly(G) binding protein (Swanson and Dreyfuss, 1988), has been shown to interact with $rGGGGCC^{exp}$ RNA *in vitro* and in patient brain sections (Lee et al., 2013b). In contrast, the colocalization of hnRNP H and $rGGGGCC^{exp}$ RNA is not observed in patient-derived iPS cells (Almeida et al., 2013). Further studies examining hnRNP H specific splicing alterations in patient tissues, either by microarray or RNA-seq, will be important to validate the potential effects of hnRNP H sequestration. Additionally, the presence of antisense $CCCCGG^{exp}$ RNA foci in ALS/FTD cells spotlights the need to identify proteins sequestered by C9ORF72 antisense RNA.

7. Concluding remarks and therapeutic perspective

While the mechanisms underlying non-coding repeat expansion diseases are complex, the commonalities that exist between these disorders hint that similar molecular events are involved in pathogenesis. Furthermore, the discoveries of RNA-mediated toxicity, bidirectional transcription across repeat expansions, RAN translation and miR dysregulation serve as a foundation to delineate the relative contribution of each of these mechanisms to disease pathogenesis. For some expansion diseases, it is not clear how titration of potentially sequestered proteins leads to the disease and loss-of-function animal models for these factors must be generated to validate their pathogenic roles. In diseases such as DM, where animal models have validated the sequestration hypothesis, there is widespread splicing

misregulation but the potential pathogenic effects of the majority of these splicing alternations remain unknown.

An appreciation of the pathogenic complexities of microsatellite expansion disorders has informed therapeutic development but also posed additional challenges. For example, ASO approaches have been used to correct specific mis-splicing events, such as reversal of Clcn1 mis-splicing and myotonia in a DM mouse model (Wheeler et al., 2007), but this type of targeted therapy cannot address the hundreds, or thousands, of disrupted RNA processing events that may occur following sequestration of RNA processing factors by repeat expansion RNAs. Modified ASO gapmers, composed of nuclease-resistant RNA flanking DNA designed to trigger RNase H-mediated degradation, have proven to be effective in knocking down wild-type and mutant transcripts in both C9ORF72 ALS/FTD fibroblasts and DM1 mouse models (Lagier-Tourenne et al., 2013; Wheeler et al., 2012). However, a caveat with this approach is that targeting both sense and antisense transcripts may be required for some diseases, such as ALS/FTD where current evidence suggests that bidirectional transcription plays an important pathogenic role. Haploinsufficiency is another concern for the ASO gapmer approach, since this strategy results in reduced expression of both normal and mutant alleles although nuclear-retained mutant RNAs are more susceptible to RNase H-mediated degradation (Wheeler et al., 2012).

Gene therapy approaches have been designed to replace the RNA-binding factors sequestered by toxic expansion RNAs, including Mbnl1 overexpression in a CTG^{exp} DM mouse model to rescue DM-associated mis-splicing and Pura overexpression to reverse disease phenotypes in FXTAS and *C9ORF72* ALS/FTD in *Drosophila* and mouse neuronal cell models (Jin et al., 2007; Kanadia et al., 2006; Xu et al., 2013). However, effective gene delivery and expression in multiple tissues poses a major technical challenge. Another therapeutic strategy is to deploy small molecule inhibitors of specific protein-RNA interactions to release endogenous RNA binding proteins from toxic repeat RNAs (Arambula et al., 2009; Jahromi et al., 2013; Warf et al., 2009). However, a potential problem with this approach is the possibility that altering these protein-RNA interactions may result in more deleterious effects, possibly due to enhanced RAN translation (Childs-Disney et al., 2013).

Many fundamental questions remain. An intriguing feature of these microsatellite expansion diseases is that many of them are neurological disorders despite the fact that some mutations occur in ubiquitously expressed genes. What makes certain tissues more susceptible to overt pathology? Why do repeat expansion disorders often manifest later in life and what key events trigger symptom onset? What are the roles of penetrance and genetic modifiers in these diseases? Answering these, and additional questions, will provide essential insights that should lead to the development of new therapeutic modalities for these diseases.

Acknowledgments

A.M. and M.G. are recipients of Grinter and Alumni Fellowships from the University of Florida. The authors' studies on microsatellite expansion disorders are funded by grants to M.S.S. from the National Institutes of Health (NS058901 and AR046799), the Muscular Dystrophy Association (MDA276063), the Keck Foundation and the Marigold Foundation.

References

Adereth Y, et al. RNA-dependent integrin alpha3 protein localization regulated by the Muscleblind-like protein MLP1. Nat Cell Biol. 2005; 7:1240–7. [PubMed: 16273094]

Almeida S, et al. Modeling key pathological features of frontotemporal dementia with C9ORF72 repeat expansion in iPSC-derived human neurons. Acta Neuropathol. 2013; 126:385–99. [PubMed: 23836290]

Arambula JF, et al. A simple ligand that selectively targets CUG trinucleotide repeats and inhibits MBNL protein binding. Proc Natl Acad Sci U S A. 2009; 106:16068–73. [PubMed: 19805260]

Arocena DG, et al. Induction of inclusion formation and disruption of lamin A/C structure by premutation CGG-repeat RNA in human cultured neural cells. Human Molecular Genetics. 2005; 14:3661–3671. [PubMed: 16239243]

Ash PE, et al. Unconventional translation of C9ORF72 GGGGCC expansion generates insoluble polypeptides specific to c9FTD/ALS. Neuron. 2013; 77:639–46. [PubMed: 23415312]

Batra R, Charizanis K, Swanson MS. Partners in crime: bidirectional transcription in unstable microsatellite disease. Hum Mol Genet. 2010; 19:R77–82. [PubMed: 20368264]

Belzil VV, et al. Reduced C9orf72 gene expression in c9FTD/ALS is caused by histone trimethylation, an epigenetic event detectable in blood. Acta Neuropathologica. 2013; 126:895–905. [PubMed: 24166615]

Budworth H, McMurray CT. Bidirectional transcription of trinucleotide repeats: roles for excision repair. DNA Repair (Amst). 2013; 12:672–84. [PubMed: 23669397]

Bugaut A, Balasubramanian S. 5′-UTR RNA G-quadruplexes: translation regulation and targeting. Nucleic Acids Res. 2012; 40:4727–41. [PubMed: 22351747]

Caillet-Boudin ML, et al. Brain pathology in myotonic dystrophy: when tauopathy meets spliceopathy and RNAopathy. Front Mol Neurosci. 2014; 6:57. [PubMed: 24409116]

Carpentier C, et al. Tau exon 2 responsive elements deregulated in myotonic dystrophy type I are proximal to exon 2 and synergistically regulated by MBNL1 and MBNL2. Biochim Biophys Acta. 2014

Castel AL, et al. Expanded CTG repeat demarcates a boundary for abnormal CpG methylation in myotonic dystrophy patient tissues. Human Molecular Genetics. 2011; 20:1–15. [PubMed: 21044947]

Chamberlain CM, Ranum LP. Mouse model of muscleblind-like 1 overexpression: skeletal muscle effects and therapeutic promise. Hum Mol Genet. 2012; 21:4645–54. [PubMed: 22846424]

Charizanis K, et al. Muscleblind-like 2-mediated alternative splicing in the developing brain and dysregulation in myotonic dystrophy. Neuron. 2012; 75:437–50. [PubMed: 22884328]

Childs-Disney JL, et al. Induction and reversal of myotonic dystrophy type 1 pre-mRNA splicing defects by small molecules. Nat Commun. 2013; 4:2044. [PubMed: 23806903]

Ciura S, et al. Loss of function of C9orf72 causes motor deficits in a zebrafish model of Amyotrophic Lateral Sclerosis. Ann Neurol. 2013; 74:180–87. [PubMed: 23720273]

Cleary JD, et al. Evidence of cis-acting factors in replication-mediated trinucleotide repeat instability in primate cells. Nat Genet. 2002; 31:37–46. [PubMed: 11967533]

Cunningham CL, et al. Premutation CGG-repeat expansion of the Fmr1 gene impairs mouse neocortical development. Hum Mol Genet. 2011; 20:64–79. [PubMed: 20935171]

DeJesus-Hernandez M, et al. Expanded GGGGCC hexanucleotide repeat in noncoding region of C9ORF72 causes chromosome 9p-linked FTD and ALS. Neuron. 2011; 72:245–56. [PubMed: 21944778]

Dion V, Wilson JH. Instability and chromatin structure of expanded trinucleotide repeats. Trends Genet. 2009; 25:288–97. [PubMed: 19540013]

Donnelly CJ, et al. RNA Toxicity from the ALS/FTD C9ORF72 Expansion Is Mitigated by Antisense Intervention. Neuron. 2013; 80:415–428. [PubMed: 24139042]

Echeverria GV, Cooper TA. RNA-binding proteins in microsatellite expansion disorders: mediators of RNA toxicity. Brain Res. 2012; 1462:100–11. [PubMed: 22405728]

Fernandez-Costa JM, et al. Expanded CTG repeats trigger miRNA alterations in Drosophila that are conserved in myotonic dystrophy type 1 patients. Hum Mol Genet. 2013; 22:704–16. [PubMed: 23139243]

Fratta P, et al. C9orf72 hexanucleotide repeat associated with amyotrophic lateral sclerosis and frontotemporal dementia forms RNA G-quadruplexes. Scientific Reports. 2012:2.

Garcia-Arocena D, et al. Fibroblast phenotype in male carriers of FMR1 premutation alleles. Human Molecular Genetics. 2010; 19:299–312. [PubMed: 19864489]

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. 2013; 126:829–44. [PubMed: 24129584]

Gendron TF, et al. Mechanisms of toxicity in C9FTLD/ALS. Acta Neuropathol. 2014

Gomes-Pereira M, et al. CTG trinucleotide repeat “big jumps”: large expansions, small mice. PLoS Genet. 2007; 3:e52. [PubMed: 17411343]

Gomes-Pereira M, Cooper TA, Gourdon G. Myotonic dystrophy mouse models: towards rational therapy development. Trends in Molecular Medicine. 2011; 17:506–517. [PubMed: 21724467]

Gonitel R, et al. DNA instability in postmitotic neurons. Proc Natl Acad Sci U S A. 2008; 105:3467–72. [PubMed: 18299573]

Greco CM, et al. Neuronal intranuclear inclusions in a new cerebellar tremor/ataxia syndrome among fragile X carriers. Brain. 2002; 125:1760–1771. [PubMed: 12135967]

Hagerman P. Fragile X-associated tremor/ataxia syndrome (FXTAS): pathology and mechanisms. Acta Neuropathol. 2013; 126:1–19. [PubMed: 23793382]

Hagerman R, Hagerman P. Advances in clinical and molecular understanding of the FMR1 premutation and fragile X-associated tremor/ataxia syndrome. Lancet Neurology. 2013; 12:786–798.

Hamshere MG, et al. Transcriptional abnormality in myotonic dystrophy affects DMPK but not neighboring genes. Proc Natl Acad Sci U S A. 1997; 94:7394–9. [PubMed: 9207102]

Hangauer MJ, Vaughn IW, McManus MT. Pervasive transcription of the human genome produces thousands of previously unidentified long intergenic noncoding RNAs. PLoS Genet. 2013; 9:e1003569. [PubMed: 23818866]

Hashem V, et al. Ectopic expression of CGG containing mRNA is neurotoxic in mammals. Hum Mol Genet. 2009; 18:2443–51. [PubMed: 19377084]

Hernandez-Hernandez O, et al. Myotonic dystrophy CTG expansion affects synaptic vesicle proteins, neurotransmission and mouse behaviour. Brain. 2013; 136:957–70. [PubMed: 23404338]

Hessl D, et al. Decreased fragile X mental retardation protein expression underlies amygdala dysfunction in carriers of the fragile X premutation. Biol Psychiatry. 2011; 70:859–65. [PubMed: 21783174]

Huguet A, et al. Molecular, physiological, and motor performance defects in DMSXL mice carrying >1,000 CTG repeats from the human DM1 locus. PLoS Genet. 2012; 8:e1003043. [PubMed: 23209425]

Hunsaker MR, et al. Mouse models of the fragile x premutation and the fragile X associated tremor/ataxia syndrome. Results Probl Cell Differ. 2012; 54:255–69. [PubMed: 22009357]

Iwahashi CK, et al. Protein composition of the intranuclear inclusions of FXTAS. Brain. 2006; 129:256–271. [PubMed: 16246864]

Jahromi AH, et al. Developing Bivalent Ligands to Target CUG Triplet Repeats, the Causative Agent of Myotonic Dystrophy Type 1. J Med Chem. 2013; 56:9471–81. [PubMed: 24188018]

Jansen G, et al. Abnormal myotonic dystrophy protein kinase levels produce only mild myopathy in mice. Nat Genet. 1996; 13:316–24. [PubMed: 8673131]

Jiang H, et al. Myotonic dystrophy type 1 is associated with nuclear foci of mutant RNA, sequestration of muscleblind proteins and deregulated alternative splicing in neurons. Hum Mol Genet. 2004; 13:3079–88. [PubMed: 15496431]

Jin P, et al. RNA-Mediated Neurodegeneration Caused by the Fragile X Premutation rCGG Repeats in Drosophila. Neuron. 2003; 39:739–747. [PubMed: 12948442]

Jin P, et al. Pur alpha binds to rCGG repeats and modulates repeat-mediated neurodegeneration in a Drosophila model of fragile X tremor/ataxia syndrome. Neuron. 2007; 55:556–64. [PubMed: 17698009]

Kalsotra A, et al. The Mef2 Transcription Network Is Disrupted in Myotonic Dystrophy Heart Tissue, Dramatically Altering miRNA and mRNA Expression. Cell Rep. 2014; 6:336–45. [PubMed: 24412363]

Kanadia RN, et al. A muscleblind knockout model for myotonic dystrophy. Science. 2003; 302:1978–80. [PubMed: 14671308]

Kanadia RN, et al. Reversal of RNA missplicing and myotonia after muscleblind overexpression in a mouse poly(CUG) model for myotonic dystrophy. Proc Natl Acad Sci U S A. 2006; 103:11748–53. [PubMed: 16864772]

Kantartzis A, et al. Msh2-Msh3 interferes with Okazaki fragment processing to promote trinucleotide repeat expansions. Cell Rep. 2012; 2:216–22. [PubMed: 22938864]

Kennedy L, et al. Dramatic tissue-specific mutation length increases are an early molecular event in Huntington disease pathogenesis. Hum Mol Genet. 2003; 12:3359–67. [PubMed: 14570710]

Khalili K, et al. Pur Is Essential for Postnatal Brain Development and Developmentally Coupled Cellular Proliferation As Revealed by Genetic Inactivation in the Mouse. Molecular and Cellular Biology. 2003; 23:6857–6875. [PubMed: 12972605]

Kiliszek A, et al. Atomic resolution structure of CAG RNA repeats: structural insights and implications for the trinucleotide repeat expansion diseases. Nucleic Acids Res. 2010; 38:8370–6. [PubMed: 20702420]

Kiliszek A, et al. Crystal structures of CGG RNA repeats with implications for fragile X-associated tremor ataxia syndrome. Nucleic Acids Res. 2011; 39:7308–15. [PubMed: 21596781]

Kim DH, et al. HnRNP H inhibits nuclear export of mRNA containing expanded CUG repeats and a distal branch point sequence. Nucleic Acids Res. 2005; 33:3866–74. [PubMed: 16027111]

Krzyzosiak WJ, et al. Triplet repeat RNA structure and its role as pathogenic agent and therapeutic target. Nucleic Acids Res. 2012; 40:11–26. [PubMed: 21908410]

Kuyumcu-Martinez NM, Wang GS, Cooper TA. Increased steady-state levels of CUGBP1 in myotonic dystrophy 1 are due to PKC-mediated hyperphosphorylation. Mol Cell. 2007; 28:68–78. [PubMed: 17936705]

La Spada AR, Taylor JP. Repeat expansion disease: progress and puzzles in disease pathogenesis. Nat Rev Genet. 2010; 11:247–58. [PubMed: 20177426]

Ladd PD, et al. An antisense transcript spanning the CGG repeat region of FMR1 is upregulated in premutation carriers but silenced in full mutation individuals. Hum Mol Genet. 2007; 16:3174–87. [PubMed: 17921506]

Lagier-Tourenne C, et al. Targeted degradation of sense and antisense C9orf72 RNA foci as therapy for ALS and frontotemporal degeneration. Proceedings of the National Academy of Sciences of the United States of America. 2013; 110:E4530–E4539. [PubMed: 24170860]

Laurent FX, et al. New function for the RNA helicase p68/DDX5 as a modifier of MBNL1 activity on expanded CUG repeats. Nucleic Acids Res. 2012; 40:3159–71. [PubMed: 22156369]

Lee KY, et al. Compound loss of muscleblind-like function in myotonic dystrophy. EMBO Mol Med. 2013a; 5:1887–900. [PubMed: 24293317]

Lee YB, et al. Hexanucleotide Repeats in ALS/FTD Form Length-Dependent RNA Foci, Sequester RNA Binding Proteins, and Are Neurotoxic. Cell Rep. 2013b; 5:1178–86. [PubMed: 24290757]

Levine TP, et al. The product of C9orf72, a gene strongly implicated in neurodegeneration, is structurally related to DENN Rab-GEFs. Bioinformatics. 2013; 29:499–503. [PubMed: 23329412]

Libby RT, et al. CTCF cis-regulates trinucleotide repeat instability in an epigenetic manner: a novel basis for mutational hot spot determination. PLoS Genet. 2008; 4:e1000257. [PubMed: 19008940]

Lin Y, Dion V, Wilson JH. Transcription promotes contraction of CAG repeat tracts in human cells. Nat Struct Mol Biol. 2006; 13:179–80. [PubMed: 16388310]

Lin Y, Wilson JH. Transcription-induced DNA toxicity at trinucleotide repeats: Double bubble Is trouble. Cell Cycle. 2011; 10:611–618. [PubMed: 21293182]

Ling SC, Polymenidou M, Cleveland DW. Converging mechanisms in ALS and FTD: disrupted RNA and protein homeostasis. Neuron. 2013; 79:416–38. [PubMed: 23931993]

Liquori CL, et al. Myotonic dystrophy type 2 caused by a CCTG expansion in intron 1 of ZNF9. Science. 2001; 293:864–7. [PubMed: 11486088]

Lokanga RA, Zhao XN, Usdin K. The mismatch repair protein MSH2 is rate limiting for repeat expansion in a fragile X premutation mouse model. Hum Mutat. 2014; 35:129–36. [PubMed: 24130133]

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. 2010; 11:165–70. [PubMed: 20177394]

Ludwig AL, Hershey JW, Hagerman PJ. Initiation of translation of the FMR1 mRNA Occurs predominantly through 5′-end-dependent ribosomal scanning. J Mol Biol. 2011; 407:21–34. [PubMed: 21237174]

Mahadevan MS, et al. Reversible model of RNA toxicity and cardiac conduction defects in myotonic dystrophy. Nature Genetics. 2006; 38:1066–1070. [PubMed: 16878132]

Mahadevan MS. Myotonic dystrophy: is a narrow focus obscuring the rest of the field? Curr Opin Neurol. 2012; 25:609–13. [PubMed: 22892953]

Mankodi A, et al. Myotonic dystrophy in transgenic mice expressing an expanded CUG repeat. Science. 2000; 289:1769–1772. [PubMed: 10976074]

Margolis JM, et al. DM2 intronic expansions: evidence for CCUG accumulation without flanking sequence or effects on ZNF9 mRNA processing or protein expression. Hum Mol Genet. 2006; 15:1808–15. [PubMed: 16624843]

McIvor EI, Polak U, Napierala M. New insights into repeat instability: role of RNA*DNA hybrids. RNA Biol. 2010; 7:551–8. [PubMed: 20729633]

Meloni R, et al. A tetranucleotide polymorphic microsatellite, located in the first intron of the tyrosine hydroxylase gene, acts as a transcription regulatory element in vitro. Hum Mol Genet. 1998; 7:423–8. [PubMed: 9466999]

Miller JW, et al. Recruitment of human muscleblind proteins to (CUG)(n) expansions associated with myotonic dystrophy. EMBO J. 2000; 19:4439–48. [PubMed: 10970838]

Mirkin SM. Expandable DNA repeats and human disease. Nature. 2007; 447:932–40. [PubMed: 17581576]

Mizielinska S, et al. C9orf72 frontotemporal lobar degeneration is characterised by frequent neuronal sense and antisense RNA foci. Acta Neuropathol. 2013; 126:845–57. [PubMed: 24170096]

Mooers BH, Logue JS, Berglund JA. The structural basis of myotonic dystrophy from the crystal structure of CUG repeats. Proc Natl Acad Sci U S A. 2005; 102:16626–31. [PubMed: 16269545]

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. 2013a; 125:413–23. [PubMed: 23381195]

Mori K, et al. The C9orf72 GGGGCC repeat is translated into aggregating dipeptide-repeat proteins in FTLD/ALS. Science. 2013b; 339:1335–8. [PubMed: 23393093]

Nakamori M, Pearson CE, Thornton CA. Bidirectional transcription stimulates expansion and contraction of expanded (CTG)*(CAG) repeats. Hum Mol Genet. 2011; 20:580–8. [PubMed: 21088112]

Nalavade R, et al. Mechanisms of RNA-induced toxicity in CAG repeat disorders. Cell Death Dis. 2013; 4:e752. [PubMed: 23907466]

O'Rourke JR, Swanson MS. Mechanisms of RNA-mediated disease. J Biol Chem. 2009; 284:7419–23. [PubMed: 18957432]

Orengo JP, et al. Expanded CTG repeats within the DMPK 3′ UTR causes severe skeletal muscle wasting in an inducible mouse model for myotonic dystrophy. Proc Natl Acad Sci U S A. 2008; 105:2646–51. [PubMed: 18272483]

Osborne RJ, Thornton CA. RNA-dominant diseases. Hum Mol Genet. 2006; 15:R162–9. [PubMed: 16987879]

Pagani F, et al. Splicing factors induce cystic fibrosis transmembrane regulator exon 9 skipping through a nonevolutionary conserved intronic element. J Biol Chem. 2000; 275:21041–7. [PubMed: 10766763]

Pastori C, et al. Comprehensive analysis of the transcriptional landscape of the human FMR1 gene reveals two new long noncoding RNAs differentially expressed in Fragile X syndrome and Fragile X-associated tremor/ataxia syndrome. Hum Genet. 2014; 133:59–67. [PubMed: 24005575]

Paul S, et al. Interaction of muscleblind, CUG-BP1 and hnRNP H proteins in DM1-associated aberrant IR splicing. EMBO J. 2006; 25:4271–83. [PubMed: 16946708]

Pearson CE, Nichol Edamura K, Cleary JD. Repeat instability: mechanisms of dynamic mutations. Nat Rev Genet. 2005; 6:729–42. [PubMed: 16205713]

Pinto RM, et al. Mismatch Repair Genes Mlh1 and Mlh3 Modify CAG Instability in Huntington's Disease Mice: Genome-Wide and Candidate Approaches. PLoS Genet. 2013; 9:e1003930. [PubMed: 24204323]

Poulos MG, et al. Developments in RNA splicing and disease. Cold Spring Harb Perspect Biol. 2011; 3:a000778. [PubMed: 21084389]

Punga T, Buhler M. Long intronic GAA repeats causing Friedreich ataxia impede transcription elongation. EMBO Mol Med. 2010; 2:120–9. [PubMed: 20373285]

Querido E, et al. Stochastic and reversible aggregation of mRNA with expanded CUG-triplet repeats. Journal of Cell Science. 2011; 124:1703–1714. [PubMed: 21511730]

Ranum LPW, Cooper TA. RNA-mediated neuromuscular disorders. Annual Review of Neuroscience. 2006; 29:259–277.

Rau F, et al. Misregulation of miR-1 processing is associated with heart defects in myotonic dystrophy. Nat Struct Mol Biol. 2011; 18:840–5. [PubMed: 21685920]

Ravel-Chapuis A, et al. The RNA-binding protein Staufen1 is increased in DM1 skeletal muscle and promotes alternative pre-mRNA splicing. J Cell Biol. 2012; 196:699–712. [PubMed: 22431750]

Reddy K, et al. The disease-associated r(GGGGCC)n repeat from the C9orf72 gene forms tract length-dependent uni- and multimolecular RNA G-quadruplex structures. J Biol Chem. 2013; 288:9860–6. [PubMed: 23423380]

Reddy S, et al. Mice lacking the myotonic dystrophy protein kinase develop a late onset progressive myopathy. Nat Genet. 1996; 13:325–35. [PubMed: 8673132]

Renton AE, et al. A hexanucleotide repeat expansion in C9ORF72 is the cause of chromosome 9p21-linked ALS-FTD. Neuron. 2011; 72:257–68. [PubMed: 21944779]

Sellier C, et al. Sam68 sequestration and partial loss of function are associated with splicing alterations in FXTAS patients. EMBO J. 2010; 29:1248–61. [PubMed: 20186122]

Sellier C, et al. Sequestration of DROSHA and DGCR8 by expanded CGG RNA repeats alters microRNA processing in fragile X-associated tremor/ataxia syndrome. Cell Rep. 2013; 3:869–80. [PubMed: 23478018]

Sergeant N, et al. Dysregulation of human brain microtubule-associated tau mRNA maturation in myotonic dystrophy type 1. Hum Mol Genet. 2001; 10:2143–55. [PubMed: 11590131]

Seznec H, et al. Transgenic mice carrying large human genomic sequences with expanded CTG repeat mimic closely the DM CTG repeat intergenerational and somatic instability. Human Molecular Genetics. 2000; 9:1185–1194. [PubMed: 10767343]

Seznec H, et al. Mice transgenic for the human myotonic dystrophy region with expanded CTG repeats display muscular and brain abnormalities. Hum Mol Genet. 2001; 10:2717–26. [PubMed: 11726559]

Sicot G, Gomes-Pereira M. RNA toxicity in human disease and animal models: from the uncovering of a new mechanism to the development of promising therapies. Biochim Biophys Acta. 2013; 1832:1390–409. [PubMed: 23500957]

Sobczak K, et al. RNA structure of trinucleotide repeats associated with human neurological diseases. Nucleic Acids Research. 2003; 31:5469–5482. [PubMed: 14500809]

Sobczak K, et al. Structural Diversity of Triplet Repeat RNAs. Journal of Biological Chemistry. 2010; 285:12755–12764. [PubMed: 20159983]

Sofola OA, et al. RNA-Binding proteins hnRNP A2/B1 and CUGBP1 suppress fragile X CGG premutation repeat-induced neurodegeneration in a Drosophila model of FXTAS. Neuron. 2007; 55:565–571. [PubMed: 17698010]

Suzuki N, et al. The mouse C9ORF72 ortholog is enriched in neurons known to degenerate in ALS and FTD. Nat Neurosci. 2013; 16:1725–7. [PubMed: 24185425]

Swanson MS, Dreyfuss G. Classification and purification of proteins of heterogeneous nuclear ribonucleoprotein particles by RNA-binding specificities. Mol Cell Biol. 1988; 8:2237–41. [PubMed: 3386636]

Taneja KL, et al. Foci of Trinucleotide Repeat Transcripts in Nuclei of Myotonic-Dystrophy Cells and Tissues. Journal of Cell Biology. 1995; 128:995–1002. [PubMed: 7896884]

Tassone F, et al. Elevated levels of FMR1 mRNA in carrier males: A new mechanism of involvement in the fragile-X syndrome. American Journal of Human Genetics. 2000; 66:6–15. [PubMed: 10631132]

Teplova M, Patel DJ. Structural insights into RNA recognition by the alternative-splicing regulator muscleblind-like MBNL1. Nat Struct Mol Biol. 2008; 15:1343–51. [PubMed: 19043415]

Therrien M, et al. Deletion of C9ORF72 Results in Motor Neuron Degeneration and Stress Sensitivity in C. elegans. PLoS One. 2013; 8:e83450. [PubMed: 24349511]

Todd PK, et al. CGG repeat-associated translation mediates neurodegeneration in fragile X tremor ataxia syndrome. Neuron. 2013; 78:440–55. [PubMed: 23602499]

Tome S, et al. MSH3 polymorphisms and protein levels affect CAG repeat instability in Huntington's disease mice. PLoS Genet. 2013; 9:e1003280. [PubMed: 23468640]

Tran H, et al. Analysis of exonic regions involved in nuclear localization, splicing activity, and dimerization of Muscleblind-like-1 isoforms. J Biol Chem. 2011; 286:16435–46. [PubMed: 21454535]

Treangen TJ, Salzberg SL. Repetitive DNA and next-generation sequencing: computational challenges and solutions. Nat Rev Genet. 2012; 13:36–46. [PubMed: 22124482]

Udd B, Krahe R. The myotonic dystrophies: molecular, clinical, and therapeutic challenges. Lancet Neurol. 2012; 11:891–905. [PubMed: 22995693]

Usdin K. The biological effects of simple tandem repeats: Lessons from the repeat expansion diseases. Genome Research. 2008; 18:1011–1019. [PubMed: 18593815]

van Blitterswijk M, et al. TMEM106B protects C9ORF72 expansion carriers against frontotemporal dementia. Acta Neuropathol. 2014

Van Deerlin VM, et al. Common variants at 7p21 are associated with frontotemporal lobar degeneration with TDP-43 inclusions. Nat Genet. 2010; 42:234–9. [PubMed: 20154673]

Wang ET, et al. Transcriptome-wide regulation of pre-mRNA splicing and mRNA localization by muscleblind proteins. Cell. 2012; 150:710–24. [PubMed: 22901804]

Wang GS, et al. Elevation of RNA-binding protein CUGBP1 is an early event in an inducible heart-specific mouse model of myotonic dystrophy. Journal of Clinical Investigation. 2007; 117:2802–2811. [PubMed: 17823658]

Warf MB, et al. Pentamidine reverses the splicing defects associated with myotonic dystrophy. Proc Natl Acad Sci U S A. 2009; 106:18551–6. [PubMed: 19822739]

Wheeler TM, et al. Correction of ClC-1 splicing eliminates chloride channelopathy and myotonia in mouse models of myotonic dystrophy. J Clin Invest. 2007; 117:3952–7. [PubMed: 18008009]

Wheeler TM, et al. Targeting nuclear RNA for in vivo correction of myotonic dystrophy. Nature. 2012; 488:111–5. [PubMed: 22859208]

Willemsen R, et al. The FMR1 CGG repeat mouse displays ubiquitin-positive intranuclear neuronal inclusions; implications for the cerebellar tremor/ataxia syndrome. Human Molecular Genetics. 2003; 12:949–959. [PubMed: 12700164]

Wojciechowska M, Krzyzosiak WJ. Cellular toxicity of expanded RNA repeats: focus on RNA foci. Hum Mol Genet. 2011; 20:3811–21. [PubMed: 21729883]

Xu ZH, et al. Expanded GGGGCC repeat RNA associated with amyotrophic lateral sclerosis and frontotemporal dementia causes neurodegeneration. Proceedings of the National Academy of Sciences of the United States of America. 2013; 110:7778–7783. [PubMed: 23553836]

Yuan Y, et al. Muscleblind-like 1 interacts with RNA hairpins in splicing target and pathogenic RNAs. Nucleic Acids Res. 2007; 35:5474–86. [PubMed: 17702765]

Zu T, et al. Non-ATG-initiated translation directed by microsatellite expansions. Proceedings of the National Academy of Sciences of the United States of America. 2011; 108:260–265. [PubMed: 21173221]

Zu T, et al. RAN proteins and RNA foci from antisense transcripts in C9ORF72 ALS and frontotemporal dementia. Proceedings of the National Academy of Sciences of the United States of America. 2013; 110:E4968–E4977. [PubMed: 24248382]

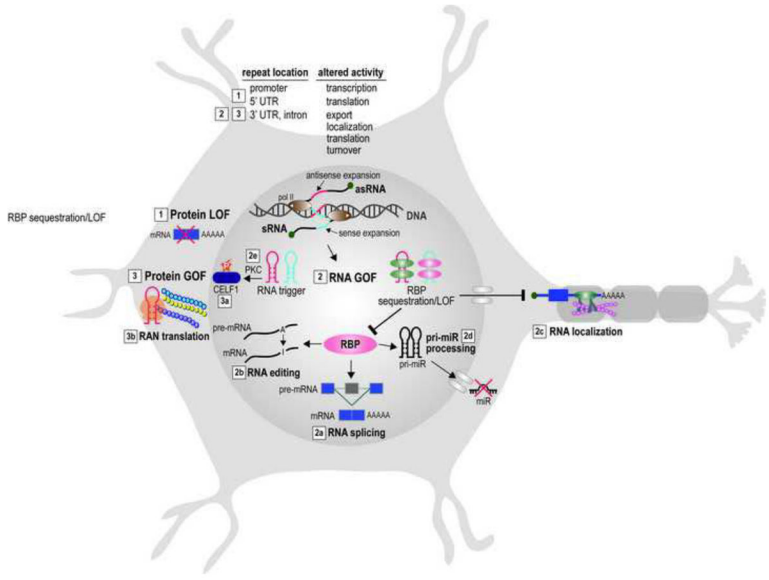

Fig. 1.

Neuronal cell model illustrating pathogenic effects of non-coding microsatellite expansions in neurologic disease. Both sense (blue line) and antisense (red line) DNA repeat expansions result in three major downstream deleterious effects (labeled 1–3): (1) protein loss-of-function (LOF) results from hypermethylation of a CG-rich expanded microsatellite repeat in the 5′ UTR and/or promoter region of an affected gene and loss of transcriptional activity (*e.g.*, FXS); (2) RNA gain-of-function (GOF) occurs following bidirectional transcription of expanded repeats and the synthesis of sense (sRNA) and antisense (asRNA), which fold into RNA hairpins (sRNA, blue; asRNA, red) or other stable structures and gain toxic functions either by sequestering an RNA binding protein(s) (RBP) and inhibiting pre-mRNA splicing (2a), pre-mRNA editing (2b), mRNA localization (2c), pri-miR processing (2d) or by triggering aberrant cellular activities such as protein kinase C (PKC) mediated CELF1 hyperphosphorylation (2e); (3) protein GOF due to altered post-translation modifications of other RNA binding proteins (3a) (CELF1 hyperphosphorylation, purple oval with white P in orange star) or RAN translation (3b) (three homopolymeric repeat proteins are shown with one undergoing translation by the ribosome (orange). In addition to these mechanisms, other pathways, including chromatin remodeling, proteome dysregulation and vesicular trafficking, may also contribute to disease pathogenesis, particularly in the CNS (Hernandez-Hernandez et al., 2013).