

Chapter 8

Deregulation of RNA Metabolism in Microsatellite Expansion Diseases

Chaitali Misra, Feikai Lin, and Auinash Kalsotra

Abstract RNA metabolism impacts different steps of mRNA life cycle including splicing, polyadenylation, nucleo-cytoplasmic export, translation, and decay. Growing evidence indicates that defects in any of these steps lead to devastating diseases in humans. This chapter reviews the various RNA metabolic mechanisms that are disrupted in Myotonic Dystrophy—a trinucleotide repeat expansion disease—due to dysregulation of RNA-Binding Proteins. We also compare Myotonic Dystrophy to other microsatellite expansion disorders and describe how some of these mechanisms commonly exert direct versus indirect effects toward disease pathologies.

Keywords Microsatellite repeat expansions · Post-transcriptional gene regulation · RNA toxicity · Alternative splicing and polyadenylation · RNA-binding proteins

8.1 Introduction

Gene expression is a highly coordinated multistep process, which allows organisms to integrate intrinsic and environmental information to exert appropriate cellular functions. The expression of most genes can be regulated at distinct stages of RNA metabolism including synthesis or transcription, post-transcriptional processing or maturation, nucleo-cytoplasmic export, translation, as well as degradation at a rate that is often dictated by transcript- and cell-type-specific cues. Although transcription is a general point of control, many co- and post-transcriptional pre-mRNA processing events add substantial capacity to tune overall gene expression [1]. The
typical pre-mRNA processing events comprise 5′ capping, splicing, and 3′ polyadenylation, which are directly linked to the nucleo-cytoplasmic export and eventual fate of mRNAs. RNA-Binding Proteins (RBPs) are essential in carrying out these processing events in both the nucleus and cytoplasm by interacting with RNA sequence or structural elements and forming distinct mRNA-protein (mRNP) complexes [2]. Disruption of RBP function(s), therefore, frequently results in deleterious RNA metabolism defects that in some cases become pathogenic [3, 4].

Neurodegenerative diseases are a heterogeneous group of neurological disorders characterized by progressive degeneration of structure and function of the central or peripheral nervous systems. Aberrant RNA metabolism is increasingly implicated in neurodegenerative diseases, a subset of which are caused by the expansion of short repetitive elements (microsatellites) within particular genes [5]. The causative repeat expansion mutation for this group of disorders is unstable because the repeat size changes through generations and even within an individual, as different tissues have cell populations with variable repeat length and in some cases the repeat length varies within the same tissue [6]. The severity of a repeat expansion disease is dependent on numerous variables, including the length of the repeat, its sequence context, and the native function of the protein-coding gene with which the repeat is associated. A typical pathogenic feature of these diseases is the accumulation of repeat-containing transcripts into aberrant RNA foci, which can sequester RBPs and prevent them from performing their normal functions [7–9]. Interestingly, once the repeat length cross a critical number, the repeat-containing RNAs can undergo phase separation—partitioning into granules due to multivalent base-pairing between repeat RNAs—or spontaneous gelation to form RNA foci, explaining why disease symptoms appear to be triggered after the expansions have reached a particular threshold number [10].

8.2 Toxicity of Coding and Noncoding Microsatellite Repeat Expansions

Over 25 human genes with tandem repeat expansions have been identified to date, and these disease-causing repeats can occur in the coding or noncoding regions [6] (Fig. 8.1 and Table 8.1). Majority of the microsatellites arise due to the expansion of trinucleotide repeats. However, expanded tetranucleotide, pentanucleotide, and hexanucleotide repeats are also detected. In the early 1990s, two microsatellites were discovered providing the first evidence that simple repeat expansions are linked to human disease. Fragile X Syndrome (FXS)—an X-linked disorder caused by CGG repeat expansions in the 5′ untranslated region (UTR) of the \textit{FMR1} gene—is the most prevalent form of inherited cognitive impairment and mental retardation [11–16]. The repeat expansion in FXS causes loss of \textit{FMR1} gene product FMRP, a polyribosome-associated RBP that binds ~4% of brain mRNAs and regulates their expression—either enhancing or suppressing translation through unknown mechanisms [17–20].
Fig. 8.1 Origin and expansion of microsatellite repeats in human disease. Schematic of the gene location for various disease-associated repeat expansions. Types of repeat expansions are indicated within the parentheses along with the range of expanded repeat numbers (UTR: untranslated region)

<table>
<thead>
<tr>
<th>Defected mRNA region</th>
<th>Disease</th>
<th>Defected gene</th>
<th>Tissue-specific clinical symptoms</th>
<th>Other tissues</th>
</tr>
</thead>
<tbody>
<tr>
<td>5′UTR</td>
<td>FXTAS</td>
<td>FMRI [158]</td>
<td>Ataxia [159], brain atrophy, white matter lesions [160, 161], cognitive decline, parkinsonism [160], peripheral neuropathy, autonomic dysfunction and short-term memory loss [162]</td>
<td>Premature ovarian failure, hypothyroidism in female [159], limb proximal muscle weakness [160]</td>
</tr>
<tr>
<td>5′UTR</td>
<td>FXS</td>
<td>FMRI [14]</td>
<td>Autism [163], mental retardation, developmental delay and increased susceptibility to seizures [15]</td>
<td>Macrorhcidism [15], cardiac murmur [164], hyperflexible joints, hernias, flat feet [165]</td>
</tr>
<tr>
<td>SCA12</td>
<td>PPP2R2B [166]</td>
<td>Ataxia, cerebral and/or cerebellar atrophy [167], seizures [22]</td>
<td>Dysarthria, action tremors in upper limbs [167]</td>
<td></td>
</tr>
<tr>
<td>Intron</td>
<td>DM2</td>
<td>ZNF9 [168]</td>
<td>Cognitive impairment [169], intellectual disability, sleepiness and fatigue [170], brain atrophy, white and grey matter abnormalities [171, 172]</td>
<td>Myotonia, muscle dysfunction, cardiac arrhythmia [40, 173], hypertrophy calf muscles [174]</td>
</tr>
<tr>
<td>ALS</td>
<td>C9ORF72 [28, 29]</td>
<td>Motor neuron degeneration, frontotemporal lobar dysfunction, dementia and cognitive impairment [175]</td>
<td>Progressive spasticity, muscle wasting, weakness and muscle atrophy [28]</td>
<td></td>
</tr>
<tr>
<td>FTD</td>
<td>C9ORF72 [28, 29]</td>
<td>Frontotemporal lobar dysfunction, motor neuron dysfunction [176], changes in personality, behavior, and language ability, dementia [175]</td>
<td>Fasciculation, muscle atrophy, weakness [177].</td>
<td></td>
</tr>
<tr>
<td>Defected mRNA region</td>
<td>Disease</td>
<td>Defected gene</td>
<td>Tissue-specific clinical symptoms</td>
<td></td>
</tr>
<tr>
<td>----------------------</td>
<td>---------</td>
<td>---------------</td>
<td>---------------------------------</td>
<td></td>
</tr>
<tr>
<td>Coding region</td>
<td></td>
<td></td>
<td>Neuronal tissues</td>
<td></td>
</tr>
<tr>
<td>Polyglutamine (PolyQ) diseases</td>
<td></td>
<td></td>
<td>Other tissues</td>
<td></td>
</tr>
<tr>
<td>SBMA</td>
<td>AR [21]</td>
<td>Lower motor neuron degeneration [178], androgen insensitivity [22]</td>
<td>Muscle weakness, gynecomastia and reduced fertility [22, 178]</td>
<td></td>
</tr>
<tr>
<td>HD</td>
<td>HTT [179]</td>
<td>Cognitive decline and dementia [22], dystonia [180]</td>
<td>Chorea [22], movement disorder [181]</td>
<td></td>
</tr>
<tr>
<td>SCA 1, 2, 3, 6, 7, 17</td>
<td>ATXN1 [186, 187], ATXN2 [188–190], ATXN3 [191], CACNA1A [192], ATXN7 [193], TBP [194]</td>
<td>Ataxia, tremor, and dysarthria, parkinsonism (SCA3), retinal dystrophy (SCA7), seizures (SCA17) [22]</td>
<td>Slurred speech (SCA1); hyporeflexia (SCA2); cardiac dysfunction (SCA7) [22]</td>
<td></td>
</tr>
<tr>
<td>Poly Alanine (Poly A) diseases</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>OPMD (OPMD)</td>
<td>PABPN1 [195]</td>
<td>– (no data)</td>
<td>Eyelid ptosis and dysphagia [195], involuntary muscle weakness [196]</td>
<td></td>
</tr>
<tr>
<td>XLMR</td>
<td>ARX [197]</td>
<td>Cognitive impairment [198], mental retardation [199], dysarthria [200]</td>
<td>Involuntary hand movements (MRXS), growth abnormality [200]</td>
<td></td>
</tr>
<tr>
<td>3’UTR DM1</td>
<td>DMPK [37]</td>
<td>Neuropsychiatric disturbances, cognitive defeats, sleepiness and fatigue; brain atrophy [169], white and grey matter abnormalities [201], mood disorder, emotion problem and memory problem</td>
<td>Myotonia, muscle wasting, cardiac arrhythmias, insulin resistance, gastrointestinal dysfunctions, posterior iridescent cataracts [54]</td>
<td></td>
</tr>
<tr>
<td>SCA8</td>
<td>ATXN8 [202]</td>
<td>Cerebellar atrophy [203], progressive ataxia [204]</td>
<td>Limb ataxia, dysarthria, nystagmus, spasticity [22]</td>
<td></td>
</tr>
</tbody>
</table>

Abbreviations: FXTAS fragile X-associated tremor/ataxia syndrome, FXS fragile X Syndrome, SCA12 spinocerebellar ataxia type 12, DM2 myotonic dystrophy type 2, ALS amyotrophic lateral sclerosis, FTD frontotemporal degeneration, SBMA spinal and bulbar muscular atrophy, HD Huntington disease, DRPLA dentatorubral pallidoluysian atrophy, SCA spinocerebellar ataxias, PolyA polyalanine diseases, OPMD oculopharyngeal muscular dystrophy, XLMR syndromic and non-syndromic X-linked mental retardation, DM1 myotonic dystrophy type 1
Spinal and bulbar muscular atrophy (SBMA)—the other microsatellite disease discovered along with FXS—arises due to a CAG repeat expansion in the coding region of the X chromosome-linked androgen receptor (AR) gene [21]. The discovery of SBMA was soon followed by the elucidation of a similar mutation as the basis for a group of disorders now known as the polyglutamine (polyQ) neurodegenerative diseases (Table 8.1). Along with SBMA, the polyQ diseases include Huntington disease (HD), dentatorubral-pallidoluysian atrophy, and six spinocerebellar ataxias (SCA) 1, 2, 3, 6, 7, and 17 [22]. As a group, these nine diseases are among the more common forms of inherited neurodegeneration. The translation of exons containing CAG repeats gives rise to elongated stretches of polyQs in mutant proteins, which aggregate into nuclear or cytoplasmic inclusions in the diseased brain [23–25]. Several observations indicate that the CAG repeat-containing RNAs, in the absence of coding for a protein, may also be a source of toxicity in polyQ diseases [26, 27]. GGGGCC hexanucleotide repeat expansion in the C9ORF72 gene has gained much attention in the past few years and is now considered the most frequent inherited cause of Amyotrophic lateral sclerosis (ALS) and Frontotemporal dementia (FTD) [28, 29]. Pathology occurs due to the toxicity of expanded repeats, which are transcribed in both the sense and antisense directions and give rise to distinct sets of intracellular RNA and protein aggregates [30–33].

Myotonic Dystrophy (DM) is part of a group of diseases characterized by repeat expansions in noncoding regions of genes. DM is defined in two clinical and molecular forms: myotonic dystrophy type 1 (DM1), and type 2 (DM2), both of which are inherited in an autosomal dominant fashion. The combined worldwide incidence of DM is approximately 1 in 8000 [34, 35]. DM1 is the most prevalent form of adult onset muscular dystrophy [36] and is caused by a CTG repeat expansion in the 3′ UTR of Dystrophia Myotonica Protein Kinase (DMPK) gene [37, 38]. DM2, on the other hand, is caused by a CCTG repeat expansion in an intron of Zinc Finger Protein 9 (ZNF9) gene [39]. While 5–37 repeats are considered normal, DM1 patients can have up to several thousand CTG repeats, which can reduce expression of DMPK [40] (Fig. 8.2a). DMPK is expressed in multiple tissues, and the major symptoms of the disease include muscle hyperexcitability (myotonia), progressive muscle wasting, cardiac defects, insulin resistance, and neuropsychiatric disturbances [41–44]. Table 8.1 provides further description of tissue-specific symptoms observed in DM and other microsatellite expansion disorders.

### 8.3 RNA Metabolism Defects in Myotonic Dystrophy

Closely after the discovery of repeats, the DMPK haploinsufficiency model was put forward to explain the DM1 pathology. However, the removal of DMPK gene in mice failed to recapitulate the major neuromuscular symptoms of DM1 [45, 46]. A separate hypothesis proposed that expanded CTG repeats might affect the expression of nearby genes. Although the adjacent gene, SIX5, exhibits reduced expression in DM1 patients [47], SIX5 knockout mice also do not reproduce DM1 muscle
Fig. 8.2  Schematic showing different pathological mechanisms for Myotonic Dystrophy type 1 (DM1) and 2 (DM2). (a) Causative mutation for DM1 is CUG repeat expansion in 3’UTR of DMPK gene and for DM2 is CCUG repeat expansion in intron 1 of ZNF9 gene. The severity of the disease is dependent on the number of repeats. Although these mutations are in two different genes, the disease mechanisms for both diseases are surprisingly similar. Most of the pathology is consistent with the toxic RNA gain-of-function mechanism and affects general RNA metabolism in both the nucleus and cytoplasm. (b) After transcription, the repeat-containing transcripts form stable hairpin loop comprising secondary structures (pink), which aggregate to form ribonuclear foci. (c) Members of the Mbnl family of RNA-binding proteins (RBPs) MBNL1/2 (purple) bind the CUG or CCUG Repeat RNA. (d) CELF1 Upregulation. (e) Misregulation of Developmentally regulated alternative splicing and polyadenylation.
pathology [48]. Instead, the CTG repeats alone, regardless of the gene context, are sufficient to induce pathogenic features of DM1 [49, 50]. The predominant pathology of DM1 actually stems from the toxic effects of expanded CUG RNA, which disrupts the normal activity of certain RBPs. Further support for the RNA toxicity model comes from the finding that although the repeat expansion in DM2 is on an entirely different gene, both diseases exhibit similar symptoms.

In both DM1 and DM2, the RNAs with expanded repeats (CUG in DM1; and CCUG in DM2) fold into stable hairpin loops that accumulate as ribonuclear foci in the nuclei of affected tissues [9] (Fig. 8.2b). These expanded RNA transcripts directly trap RBPs such as muscleblind-like proteins (MBNLs) and cause upregulation of CUG-binding protein 1 (CELF1) family of alternative splicing factors [51–54], which results in aberrant splicing of many transcripts and a broad, multi-systemic phenotype (Fig. 8.2c, d). Alternative pre-mRNA splicing generates much of the transcriptome diversity in higher eukaryotes as it enables the production of multiple transcripts with potentially different functions from each individual gene [55]. Alternative splicing decisions are generally influenced by cis-acting regulatory elements within pre-mRNAs that promote or inhibit exon recognition, as well as expression/activity of trans-acting factors (e.g., MBNL and CELF proteins) that bind to these cis elements and regulate the accessibility of the spliceosome to splice sites [56]. The misregulated splicing events in DM are usually developmentally regulated and exhibit an adult-to-embryonic switch in splicing patterns (Fig. 8.2e). Some of these embryonic isoforms fail to meet the adult tissue requirements and thus directly contribute to the overall disease pathology [54].

8.3.1 Misregulation of mRNA Processing

MBNL loss-of-function in DM1 and DM2 is a prominent example of RBP sequestration by disease-associated microsatellite expansion RNAs. The MBNL proteins were initially identified in Drosophila melanogaster for their requirement in muscle development and eye differentiation [57], and they were later shown as direct regulators of alternative splicing [58]. There are three MBNL paralogues in mammals, named MBNL1–3. MBNL1 and MBNL2 are widely expressed across many tissues,
including brain, heart, muscle, and liver, whereas MBNL3 expression is restricted to the placenta [59]. In a majority of tissues, MBNL1 and MBNL2 mRNA levels rise during differentiation [60, 61]. Besides their roles in pre-mRNA processing, MBNLs also influence gene expression by regulating cellular mRNA transport, stability as well as microRNA biogenesis [62–67]. The high expression of MBNL1 in the heart and skeletal muscle is consistent with the most severe DM phenotypes in these tissues. For instance, independent of the repeat expansion, Mbnl1 deletion in mice reproduces many of the cardinal symptoms of DM1 such as myotonia, myopathy, cataracts, and misregulation of developmentally regulated RNA processing [63, 68].

The expanded repeat-containing RNAs in DM sequester MBNL1, 2, and 3 in nuclear RNA foci [69–71], and this protein redistribution explains the inhibition of their normal functions predominantly in alternative splicing and polyadenylation, microRNA processing, and mRNA localization [58, 62, 67, 72–75]. The MBNL loss-of-function hypothesis is further supported by studies on Mbnl single- and compound-knockout mice, which recapitulate many of the DM phenotypes [68, 76–78]. The extent of symptoms, however, varies depending on the tissue context, relative concentrations of MBNL paralogues, and the degree to which they are sequestered [78]. For instance, compared to skeletal muscle, only few splicing defects are observed in the brains of Mbnl1 knockout mice [63, 79]. Alternatively, Mbnl2 knockout mice exhibit a number of DM-related central nervous system abnormalities including irregular REM sleep propensity and deficits in spatial memory [76], which is consistent with the observation that MBNL2 expression in the brain is higher than MBNL1 [59]. MBNL2 is directly sequestered by repeat expansions in the brain tissue of human DM patients resulting in misregulation of alternative splicing and polyadenylation of its normal RNA targets [80]. One of the most misspliced mRNA due to loss of MBNL2 is human microtubule-associated protein tau (MAPT) in the DM1 frontal cortex [80]. RNA toxicity mediated through MBNL2 sequestration leads to abnormal expression of tau isoforms and the progressive appearance of neurofibrillary tangles composed of intraneuronal aggregates of hyper-phosphorylated tau protein [81].

More recently, MBNL proteins were found to serve essential roles in poly(A) site selection for many transcripts (Fig. 8.2e). By integrating HITS-CLIP and RNA-seq from MBNL knockout cells and transgenic DM1 mouse model, along with minigene reporter studies, Swanson and colleagues demonstrated that MBNL proteins directly suppress or activate polyadenylation for thousands of pre-mRNAs [75, 80]. Thus, MBNL proteins coordinate multiple pre-mRNA processing steps and their sequestration in DM depletes them from their normal RNA targets.

Besides MBNL loss-of-function, there is accumulation and aberrant sub-cellular distribution of another splicing factor CELF1 in DM. CELF proteins are normally downregulated during postnatal striated muscle development, which facilitates fetal-to-adult splicing transitions in hundreds of muscle transcripts [61, 82]. CELF1 actually does not colocalize with RNA foci [83], and its upregulation in DM1 occurs through two separate mechanisms. First, CELF1 protein is stabilized through its hyper-phosphorylation [84]; and second, reduced levels of microRNAs in DM1 derepress CELF1 protein translation [85, 86] (Fig. 8.2d, g, h). The situation is less
clear in DM2, with conflicting reports of normal [87, 88] and increasing CELF1 protein levels [89] in patient tissues and cells. It is interesting to note that for many pre-mRNAs whose splicing is disrupted in DM1, CELF1 and MBNL1 regulate them in an antagonistic manner [58, 61, 90–92]. The antagonism, however, is not due to direct competition for the binding site as both CELF1 and MBNL1 bind and regulate splicing independently via distinct cis-acting RNA motifs.

In addition to MBNL and CELF proteins, other RNA splicing factors are implicated in DM. For instance, hnRNP H binds to DMPK-derived CUG-expanded RNAs in vitro and increased hnRNP H levels may also contribute toward DM pathogenesis [93]. hnRNP H forms a repressor complex with MBNL1 and nine other proteins (hnRNP H2, H3, F, A2/B1, K, L, DDX5, DDX17, and DHX9) in normal myoblast extracts but elevated hnRNP H levels in DM1 disrupt the stoichiometry of these complexes which affects splicing of specific pre-mRNAs [94, 95]. Since expanded CUG repeat RNAs fold into hairpin structures [96], the partial recruitment and colocalization of the RNA helicase p68/DDX5 with RNA foci may also have a contributing role in splicing dysregulation. Moreover, p68/DDX5 can modulate MBNL1-binding activity, and its colocalization with nuclear RNA foci can further stimulate MBNL1 binding to repeat RNAs [97].

8.3.2 Misregulation of mRNA Localization and Stability

Following transcription, newly synthesized and fully processed mRNAs are bound by specific RBPs to form export-competent mRNPs, which help their transport through the nuclear pore complex (NPC). Some pre-mRNAs are processed at the speckle periphery before being exported and repeat-containing nuclear foci can colocalize at the periphery of nuclear speckles, a non-membrane bound nuclear assembly of macromolecules including splicing factors. The presence of expanded CUG repeats may, therefore, prevent entry of other RNAs into the nuclear speckle [98, 99]. However, in DM2, the mutant ZNF9 mRNA is exported normally as the expanded CCUG repeats are removed during splicing. The nuclear foci formed by DM2 intronic repeats are widely dispersed in the nucleoplasm and not associated with nuclear speckles. Also, it is not yet clear whether the DM1 and/or DM2 nuclear foci contain partially degraded fragments of CUG or CCUG repeats or larger intact RNAs respectively.

As discussed above, CELF1 upregulation and MBNL sequestration by the CUG repeats in DM1 cause misprocessing of hundreds of transcripts. Aberrant processing results in nucleocytoplasmic export defects for many of these transcripts. Furthermore, MBNL proteins are localized both in the nucleus and cytoplasm and several studies have demonstrated their direct roles in mRNA localization [62, 100] (Fig. 8.2f). For instance, by interacting with the 3′-UTR of Integrin α3, MBNL2 moves it to the plasma membrane for its local translation [64]. Similarly, MBNL1 also plays major roles in mRNA localization and membrane-associated translation. Transcriptome-wide analyses of subcellular compartments from mouse myoblasts
showed widespread defects in mRNA localization upon combined depletion of MBNL1 and MBNL2 [62]. Many of the mislocalized mRNAs encode for secreted proteins, extracellular matrix components, and proteins involved in cell–cell communication. MBNL depletion in DM can thus have a significant impact on mRNA localization potentially affecting proper neuromuscular junction formation.

In the cytoplasm, MBNLs also regulate mRNA stability [101] (Fig. 8.2i). MBNL1 specifically recognizes YGCY-containing motifs within the 3′-UTR regions and destabilizes the target mRNAs through unknown mechanisms [65, 92]. CELF1, on the other hand, induces mRNA decay of short-lived transcripts through interactions with GU-rich elements (GREs) in their 3′-UTR and possibly recruitment of poly(A)-specific ribonuclease, which promotes deadenylation of target transcripts [102–104]. Many of the CELF mRNA targets with GREs encode proteins essential for muscle cell development and function [105–108]. Interestingly, CELF1 binds to the mRNAs coding for SRP protein subunits and promotes their decay [109]. Signal recognition particle (SRP) is a cytoplasmic ribonucleoprotein complex, which regulates the translation of secreted and membrane-associated proteins. It is likely that the CELF1 overexpression contributes to the faster turnover of SRP mRNAs and the reduced SRP levels thereby attenuate the protein secretory pathway in DM1 [109].

### 8.3.3 Misregulation of mRNA Translation

CELF1 is additionally involved in the regulation of mRNA translation [106, 110–112] (Fig. 8.2h). The affinity of CELF1 toward its mRNA targets can be modulated through phosphorylation [113]. For instance, phosphorylated CELF1 interacts with a subunit of initiation factor eIF2, leading to the recruitment of translational machinery to target mRNAs [106]. In myoblasts, AKT phosphorylates CELF1 and increases its affinity for CCND1 mRNA. During myoblast-to-myotube differentiation, cyclinD3-cdk4/6 phosphorylates CELF1, which increases CELF1 interaction with 5′-UTR of p21 mRNA (a cell cycle inhibitor) and enhances its translation. Myoblasts from DM1 patients show an increased interaction between CELF1 and AKT and have reduced cyclinD3-CDK4/6 levels during differentiation [105]. Moreover, DM1 myoblasts during differentiation show a reduced ability to withdraw from cell cycle, which may be due to the altered translation of P21 or myogenic transcription factor MEF2A by CELF1 [111, 112].

mRNA translation in DM1 is also affected due to microRNA deregulation (Fig. 8.2g). A subset of developmentally regulated microRNAs associated with cardiac arrhythmias is downregulated in the hearts of DM1 patients and mice [67, 86]. Downregulation of these microRNAs recapitulates particular gene expression deficits seen in DM1 hearts including enhanced protein levels of miR-1 targets CX43 and Cav1.2 as well as miR-23a/b target CELF1 [67, 86]. In DM1 and DM2 skeletal muscle biopsies, both the levels and cellular distribution of several evolutionarily
conserved microRNAs are altered affecting their downstream targets [114–117]. Furthermore, specific microRNAs are differentially detected in peripheral blood plasma of DM1 patients, which inversely correlate with skeletal muscle strength and may serve as noninvasive biomarkers [118]. More recently, reduced expression of miR-200c/141 tumor suppressor family was shown to correlate with increased oncologic risk in women with DM1 especially for gynecologic, brain, and thyroid cancer [119].

Besides altering cellular translation through misregulation of RBPs and microRNAs, the microsatellite expansions also promote unconventional translation of repeats in multiple reading frames producing homopolymeric peptides that aggregate in both the nucleus and the cytoplasm [120] (Fig. 8.2j). Designated as Repeat Associated Non-AUG Translation (RAN translation), it was first described for the expanded CAG and CTG repeats that cause spinocerebellar ataxia 8 (SCA8) and DM1, respectively [120]. Interestingly, the efficiency of RAN translation increases with the size of repeats and when RNA forms hairpin-like structures [121]. Additionally, the cells making the toxic RAN protein products are prone to apoptosis as detected in tissues of affected patients, indicating a potential contribution of RAN to pathogenesis. In addition to DM1, Zu et al. recently demonstrated that in DM2 the tetranucleotide expansion repeats are bidirectionally transcribed, and the resulting transcripts are RAN translated, producing tetrapeptide expansion proteins with Leu-Pro-Ala-Cys (LPAC) from the sense strand or Gln-Ala-Gly-Arg (QAGR) repeats from the antisense strand [122]. These RAN proteins were readily detected in the DM2 patient brains; however, the specific roles of these RAN proteins regarding toxicity, mechanism of action, and their regulation are yet to be determined.

Since their original discovery, RAN translation has now been observed in many other repeat-expansion diseases, including ALS/FTD, FXTAS, and Huntington’s disease [52, 123]. However, the exact mechanisms initiating translation from these repeats likely differ across diverse sequence contexts [124]. For instance, in case of FMR1, expanded CGG repeats in the 5’-UTR initiate CAP-dependent RAN translation upstream of the canonical AUG start codon, producing FMRpolyGlycine and FMRpolyAlanine in FXTAS [123, 125]. In contrast to FXTAS, the expanded repeats in DM1 exist within the 3’ UTR of DMPK mRNA, which is not in the normal path of ribosome scanning; thus, unconventional ribosome interactions must contribute in their translation. For HTT in Huntington’s disease, the CAG repeats are in the ORF, and canonical translation starts at the native AUG codon upstream of the repeats. But in some instances, HTTpolySerine and HTTpolyAlanine proteins are also produced due to RAN-translation and frame shifting from the normal HTTpolyGlutamine frame of the repeats [126]. Finally, in case of ALS/FTD, the GGGGCC repeats are within C9ORF72 intron, and the RAN-translation generates polyGlycine-Alanine, polyGlycine-Arginine, and polyGlycine-Proline dipeptide products [31, 127]. The RAN translation in this case, however, may occur from the intron retained transcript, spliced lariat, or a 3’ truncated RNA generated due to stalled transcription [124, 128].
8.4 Disrupted Function of RBPs in Other Microsatellite Expansion Disorders

Recent paradigm-shifting advances have established that defective RNA processing through disrupted function of RBPs is central to many other repeat expansion diseases (Table 8.2). For instance, RBP defects occur in both familial and sporadic cases of ALS/FTD [129, 130]. Mutations in TARDBP and FUS genes respectively encoding TDP-43 and FUS/TLS proteins result in abnormal aggregation of these proteins in neurons and are considered pathogenic for ALS/FTD. TDP-43 and FUS/TLS are RNA/DNA-binding proteins, with noticeable structural and functional similarities.

TDP-43 functions in multiple RNA processing steps including pre-mRNA splicing [131–134], RNA stability [135–137], and transport [138]. Similar to TDP-43, FUS interacts with serine-arginine (SR) proteins that serve diverse roles in splicing [139] and regulates transcription by recruiting other RBPs through noncoding RNAs [140]. Hence, the association of TDP43 and FUS/TLS with ALS and FTD is

<table>
<thead>
<tr>
<th>Diseases name</th>
<th>RBPs</th>
<th>Pathological mechanism</th>
</tr>
</thead>
<tbody>
<tr>
<td>(b) DM1/2</td>
<td>MBNL1/2/3 [49, 68, 72–74]; and CELF [39, 83, 84, 90, 91, 111, 217], HnRNP H [93], p68/DDX5 [97]</td>
<td>Nuclear Foci and RBP Sequestration, RNA splicing [58, 61, 62, 82, 218, 219] and polyadenylation misregulation [75, 80], miRNA biogenesis [67, 86, 115], Translation and cellular localization disruption [62, 99], Intracellular aggregation by non-canonical RAN translation [122]</td>
</tr>
<tr>
<td>(c) ALS/FTD</td>
<td>TDP-43 [220, 221] FUS [222, 223], TAF15 [141, 142], EWSR1 [143, 144] hnRNPA1 and hnRNPA2B1 [146], Ataxin 2 [145], TIA1 [224]</td>
<td>Nuclear foci [225], Splicing misregulation [132, 134, 226], translation, and RNA transport [227], impaired cytoplasmic localization [154, 228, 229], mutated LCD domain mediated cytoplasmic inclusions [146, 230–233]</td>
</tr>
<tr>
<td>(d) SCA 8</td>
<td>MBNL/CELF [234], Staufen [235]</td>
<td>RNA Splicing [234], RAN Translation [120]</td>
</tr>
</tbody>
</table>

Abbreviations: FXTAS fragile X-associated tremor/ataxia syndrome, FXS fragile X syndrome, DM1/2 myotonic dystrophy type 1/2, ALS amyotrophic lateral sclerosis, FTD frontotemporal degeneration, SCA spinocerebellar ataxias, FMRF fragile X mental retardation protein, CELF CUGBP Elav-like family member, mbnl Muscleblind like splicing regulator, hnRNP heterogeneous nuclear ribonucleoprotein, TAF15 TATA box-binding protein-associated factor 15, EWSR1 Ewing sarcoma breakpoint region 1, TIA1 T cell intracytoplasmic antigen
redirecting research efforts toward identifying additional RBPs that are mutated in neurological diseases, defining their normal RNA substrates and determining the misprocessed RNAs that underlie particular disease symptoms. In fact, mutations in several other RBPs that are functionally and structurally similar to FUS/TLS such as TAF15 [141, 142] and EWSR1 [143, 144], as well as the less closely related RBPs—Ataxin 2 [145], hnRNPA2B1 [146], hnRNPA1 [146], and Matrin3 [147] were recently identified. Among these RBPs, TDP-43, FUS, and hnRNPA1 harbor low complexity domains (LCDs), which can polymerize and drive phase separation to form dynamic membrane-less organelles or liquid droplets. For instance, a 57-residue segment within the FUS-LCD was recently shown to assemble into a fibril core that promotes phase-separation and hydrogel formation. Interestingly, phosphorylation of the core-forming residues by DNA-dependent protein kinase dissolves the FUS-LCD liquid droplets providing a molecular basis for the dynamics of LCD polymerization and phase separation [148].

Disease-associated mutations within LCDs of RBPs also enhance prion-like properties and accelerate the shift from liquid to solid phase disturbing proper ribonucleoprotein (RNP) formation [127, 149, 150]. These mutations likely trigger protein aggregation due to aberrant self-assembly of LCDs. The cytoplasmic aggregation of RBPs not only affects their typical functions in RNA metabolism but also diminishes general nucleocytoplasmic trafficking, a common consequence of ALS-initiating mutations [151–153]. While the exact reasons impeding nuclear/cytoplasmic transport in ALS are not yet fully established, multiple independent mechanisms have been proposed. For example, nucleocytoplasmic trafficking defects can arise due to proteotoxicity caused by cytoplasmic β-sheet containing protein aggregations [154], direct interactions between repeat RNAs and nuclear import factors [153], or inhibition by RAN translation-products of repeat RNAs [151]. Interestingly, arginine-containing dipeptide repeats produced from RAN translation of hexanucleotide GGGGCC expansions in ALS interact with LCDs of RBPs, which disrupts the dynamics and functions of membrane-less organelle formation by LCDs [155, 156]. Furthermore, subsets of these arginine-containing dipeptides frequently bind to the LCDs encoded by the nuclear pore proteins blocking the transport of macromolecules into and out of the nucleus [157]. Thus, interaction of RAN translation products with LCDs is a yet another pathogenic mechanism that interferes with the normal function of RBPs in microsatellite expansion disorders.

8.5 Conclusions

The past decade has seen remarkable progress in our understanding of the molecular pathogenesis of microsatellite repeat expansion disorders. Although the repeats may vary in terms of their length and location within a gene or the multiple ways through which they cause disease, one commonality of microsatellite expansions is the production of toxic RNA species containing repeats. Mechanistically, the
pathology arises either due to loss-of-function of the affected gene, or gain-of-function of the repeat-containing RNAs. Regarding loss-of-function, the repeats can induce transcriptional silencing of the affected gene through epigenetic modifications or produce a non-functional protein that contains a long stretch of homopolymeric amino acids. In case of gain-of-function, the RNAs with expanded repeats often sequester RBPs and thus disrupt their normal activities. Alternatively, the translated protein with a repetitive stretch of homopolymeric peptide sequence can misfold, aggregate, and trap critical cellular proteins causing nucleo-cytoplasmic export defects and further proteotoxicity. For a number of repeat expansion disorders, there is an intricate overlap of such loss- and gain-of-function mechanisms resulting in complex molecular pathologies. We envision that for many repeat expansions, the future investigations will be geared toward determining the unique versus overlapping disease mechanisms, dissecting direct versus indirect RNA metabolism defects, and finally, understanding whether alterations in RNA metabolism occur early or during late stages of the disease.

Acknowledgments A.K. is supported by grants from the US National Institute of Health (R01HL126845), Muscular Dystrophy Association (MDA514335), and the Center for Advanced Study at the University of Illinois. C.M. is supported by the American Heart Association post-doctoral fellowship (16POST29950018).

References


70. Warf MB, Berglund JA. MBNL binds similar RNA structures in the CUG repeats of myotonic dystrophy and its pre-mRNA substrate cardiac troponin T. RNA. 2007;13(12):2238–51.


