

Native functions of short tandem repeats

Shannon E Wright^{1,2,3}, Peter K Todd^{1,4*}

¹Department of Neurology, University of Michigan–Ann Arbor, Ann Arbor, United States; ²Neuroscience Graduate Program, University of Michigan–Ann Arbor, Ann Arbor, United States; ³Department of Neuroscience, Picower Institute, Cambridge, United States; ⁴VA Ann Arbor Healthcare System, Ann Arbor, United States

Abstract Over a third of the human genome is comprised of repetitive sequences, including more than a million short tandem repeats (STRs). While studies of the pathologic consequences of repeat expansions that cause syndromic human diseases are extensive, the potential native functions of STRs are often ignored. Here, we summarize a growing body of research into the normal biological functions for repetitive elements across the genome, with a particular focus on the roles of STRs in regulating gene expression. We propose reconceptualizing the pathogenic consequences of repeat expansions as aberrancies in normal gene regulation. From this altered viewpoint, we predict that future work will reveal broader roles for STRs in neuronal function and as risk alleles for more common human neurological diseases.

Introduction

At least a third of the human genome is comprised of repetitive sequences (*de Koning et al., 2011; Gemmell, 2021; Britten and Kohne, 1968*). Some of the first genomic repetitive elements were discovered in association with disease. As a result, pathogenic roles of repeats were well studied, while potential native functions of these repeats were largely dismissed. However, the conservation of genomic repeats among different eukaryotic species (*Eichler et al., 1995; Sulovari et al., 2019; Liquori et al., 2003*) and their high polymorphism rates compared to other types of genetic variations (*Willems et al., 2016*) suggests that repeats may have important biological functions in addition to the pathogenic ones. A growing body of research has revealed complex biological and evolutionary functions for repeats across the genome. Here, we summarize the important functions of one type of genomic repeat, short (2–6 base pair) tandem repeats (STRs), in DNA, RNA, and as proteins. We then reframe STR toxicity observed in repeat expansion disorders (REDs) as an aberrancy of native STR functions, rather than as solely a emergent property disconnected from the native repeat. Finally, we discuss how this alternative view of STR toxicity can improve our understanding of roles of STRs in neuronal function and human health.

A brief history of repetitive DNA

Repetitive elements in DNA were first discovered by Barbara McClintock, who observed the presence of ‘controlling elements’ randomly dispersed throughout the maize (*Zea mays*) genome (*Comfort, 2001; Ravindran, 2012; McClintock, 1950*). These interspersed repeats, which would come to be known as transposable elements (TEs), use flanking repetitive sequences to ‘jump’ around to different locations in the genome, often resulting in duplications of genetic material.

In contrast to interspersed repeats, tandem repeats (TRs) are regions in which repeating units lie in parallel (or in tandem) and are classified by size of the repeating unit as satellites (>60 base pairs), mini-satellites (10–60 base pairs), or microsatellites (<9 base pairs). Short (2–6 base pair) tandem repeats (STRs) comprise between 1% and 3% of the human genome (*Gymrek et al., 2016; Wyner et al., 2020; Lander et al., 2001*). In the early 1990s, a series of STR expansions were causally linked with human diseases, including spinobulbar muscular atrophy (*La Spada et al., 1992*), Fragile X Syndrome

*For correspondence:
petertod@umich.edu

Competing interest: See page 15

Funding: See page 16

Received: 11 October 2022

Accepted: 08 March 2023

Published: 20 March 2023

Reviewing Editor: Wenfeng Qian, Chinese Academy of Sciences, China

© This is an open-access article, free of all copyright, and may be freely reproduced, distributed, transmitted, modified, built upon, or otherwise used by anyone for any lawful purpose. The work is made available under the [Creative Commons CC0 public domain dedication](https://creativecommons.org/licenses/by/4.0/).

(*Fu et al., 1991; Heitz et al., 1991; Oberlé et al., 1991; Verkerk et al., 1991; Yu et al., 1991*), Huntington's disease (*MacDonald et al., 1993*), and myotonic dystrophy (*Brook et al., 1992; Buxton et al., 1992; Harley et al., 1992; Aslanidis et al., 1992; Fu et al., 1992; Mahadevan et al., 1992*). As such, much of the research on STRs to date has centered on the mechanisms by which repeat expansions trigger neuronal toxicity. We will use the Fragile X locus as an exemplar of this now extensive body of literature, which is reviewed in more detail elsewhere (*Malik et al., 2021a; Hagerman et al., 2017; Hagerman and Hagerman, 2015; Jacquemont et al., 2003; Glineburg et al., 2018*), as it helps us understand how STRs might function normally in the absence of expansion.

Fragile X-associated disorders: the discovery of pathogenic short tandem repeats

Fragile X Syndrome (FXS), the most common monogenic form of intellectual disability, was one of the first genetic diseases linked to an STR expansion (*Fu et al., 1991; Heitz et al., 1991; Oberlé et al., 1991; Verkerk et al., 1991; Yu et al., 1991*). In 1943, Julia Bell and James Purdon Martin described an X-linked intellectual disability primarily affecting people assigned male at birth, that could be inherited from a carrier female parent or affected male parent (*Martin and Bell, 1943*). Karyotypes of affected individuals show a folate-sensitive fragile site on the X chromosome, which causes the chromosome to bend or break at one arm (*Lubs, 1969; Proops and Webb, 1981; Chudley and Hagerman, 1987; Hagerman et al., 1986*). The fragile site associated with FXS is located at the Fragile X messenger ribonucleoprotein 1 (*FMR1*) gene, which contains a large CGG repeat in the 5' UTR of affected individuals (>200 repeats) (*Fu et al., 1991; Heitz et al., 1991; Oberlé et al., 1991; Verkerk et al., 1991; Yu et al., 1991*). In addition to intellectual disability, FXS patients commonly present with hyperactivity, anxiety, and seizures (*Hagerman et al., 2017; Chudley and Hagerman, 1987; Hagerman and Hagerman, 2002a*). Other chromosomal fragile sites also contain STRs, some of which are linked to other diseases (*Glover, 2006; Debacker and Kooy, 2007; Schwartz et al., 2006*). For example, Fragile XE syndrome (FRAXE), caused by a CGG repeat expansion in the *FMR2* gene (*Knight et al., 1993; Gu et al., 1996; Gecz et al., 1996*), manifests in an X-linked intellectual disability similar to FXS (*Mulley et al., 1995; Gecz, 2000*).

While studying pedigrees of Fragile X families, Stephanie Sherman and colleagues observed incomplete penetrance of mental impairment, affecting only 79% of males and 35% of females (*Sherman et al., 1984; Sherman et al., 1985*). This 'Sherman paradox' suggested a generational risk factor in Fragile X mental impairment, as unaffected 'normal transmitting' males (NTMs) passed on a mutant allele to unaffected female children, with disease manifestation in affected (predominantly) male grandchildren. Subsequent studies of CGG repeat length variation found that individuals from non-Fragile X families have 6–54 CGG repeats, while some unaffected individuals in Fragile X families have 55–200 repeats, a 'pre-mutation' associated with increased risk of further repeat expansion during oogenesis (*Fu et al., 1991*).

Subsequent work with Fragile X families revealed that Fragile X premutation expansion carriers often manifest clinically distinct disorders that are caused by the CGG repeat. Fragile X-associated tremor/ataxia syndrome (FXTAS) is an age-linked neurodegenerative disorder characterized by progressive intention tremor and ataxia, parkinsonism, and cognitive decline (*Hagerman and Hagerman, 2015; Jacquemont et al., 2003; Hagerman and Hagerman, 2004; Hagerman et al., 2001; Hagerman and Hagerman, 2002b; Leehey et al., 2003; Brunberg et al., 2002*). As an X-linked disorder, FXTAS primarily affects people assigned male at birth. People with two X chromosomes may develop FXTAS, but are also at risk for developing Fragile X-associated premature ovarian insufficiency (FXPOI), a disorder characterized by absent or irregular menstrual cycles, early onset of menopause, and fertility issues (*Hagerman and Hagerman, 2002a; Allingham-Hawkins et al., 1999; Murray et al., 2000a; Murray et al., 2000b*). As 'premutation disorders', FXTAS and FXPOI are thought to share similar molecular mechanisms by which the premutation CGG repeat expansion causes cytotoxicity and dysfunction.

More than 50 REDs discovered to date show common mechanisms of molecular pathology (*Malik et al., 2021a; Glineburg et al., 2018; Paulson, 2018; Rodriguez and Todd, 2019*). FXS, FXTAS, and FXPOI, collectively referred to as Fragile X-associated disorders, are revisited throughout this review to exemplify the mechanisms by which STRs can cause cellular dysfunction and toxicity. However, the often-stereotyped manifestations of REDs, in addition to the abundance of repetitive elements

throughout the genome, suggests that STRs could have native functions which become aberrant in the setting of repeat expansions. We will focus most of the rest of the review on this supposition.

Native STR functions

While overshadowed by disease-centered research, scientists have investigated functional consequences of repeat polymorphisms for decades. Studies of individual or small groups of genes showed phenotypic consequences of repeat length variation on flocculation and cell adhesion in yeast (Voynov *et al.*, 2006; Levdansky *et al.*, 2007; Verstrepen *et al.*, 2005), limb and skull morphology in dogs (Fondon and Garner, 2004) and on behavioral traits in voles (Hammock and Young, 2005). Recent advances in sequencing technology and STR-conscious alignment techniques now permit the detection and characterization of thousands of new STRs and their variation across the human genome, and have enabled genome-wide study of the effect of repeat length polymorphisms on gene expression (Willemms *et al.*, 2016; Payseur *et al.*, 2011; McIver *et al.*, 2011; O'Dushlaine and Shields, 2008; McIver *et al.*, 2013; Willemms *et al.*, 2014; Mallick *et al.*, 2016). As thousands of single-nucleotide polymorphisms (SNPs) have been linked with disease risk in Genome Wide Association Studies (GWAS; Tam *et al.*, 2019; Uffelmann *et al.*, 2021), ongoing studies of human genomes aim to link variation in STR length to phenotypic outcomes (Gymrek *et al.*, 2016; Fotsing *et al.*, 2019; Mitra *et al.*, 2021). In homage to the expression quantitative trait loci (eQTL) identified in traditional GWAS (Tam *et al.*, 2019; Uffelmann *et al.*, 2021), STRs associated with differences in expression of nearby genes are called eSTRs. In the following sections, we will first discuss evidence for evolutionary constraint on STRs linked to evolution across phylogeny and in humans. We will then showcase the mechanisms by which variation in STR length affects gene expression.

Repetitive DNA regulates transcription

Repetitive elements can impact the transcription of neighboring genes or the genes in which they reside by regulating chromatin structure and epigenetic markers. A role for repetitive DNA in facilitating 3D folding of the genome was first observed with TE-dependent formation of chromatin loops across multiple species, including yeast, *Drosophila*, and mammals (Figure 1(i); Cournac *et al.*, 2016; Bourque, 2009; Lu *et al.*, 2021). Contact maps generated using chromosome conformation capture (Hi-C) show high co-localization of repetitive elements in nuclear space in humans, mice, and *Drosophila*, demonstrating a structural function of repetitive DNA (Cournac *et al.*, 2016). The enrichment of transcription factor binding sites in proximity to spatially associated repeats suggests that repeat-mediated 3D DNA packaging may allow for context-dependent co-transcription of linearly remote genes (Cournac *et al.*, 2016).

STRs play critical roles in maintaining chromatin structure (Nikumbh and Pfeifer, 2017; Sun *et al.*, 2018; Volle and Delaney, 2012). For example, short CAG/CTG tracts avidly incorporate nucleosomes (Figure 1 (ii)), which are a basic subunit of chromatin packaging (Volle and Delaney, 2012). Nucleosome position varies with differences in STR length and flanking sequence context (Volle and Delaney, 2012), influencing chromatin structure and transcription of nearby genes. Other STRs, including CGG repeats, have the opposite effect and exclude nucleosomes in their native states, creating more open chromatin states near the transcription start sites of genes that favor local transcription (Wang *et al.*, 1996; Wang, 2007). This feature may underlie the enrichment of CGG repeats in promoters and 5'UTRs (Uesaka *et al.*, 2014).

STRs can also influence chromatin structure by modulating DNA methylation. Some STRs are prone to methylation, which can lead to gene silencing and the absence of transcription (Quilez *et al.*, 2016; Garg *et al.*, 2021; Pappalardo and Barra, 2021). One common example of repeat-mediated gene silencing, CpG islands are repeating di-nucleotide CpG sequences ranging from around 500–3000 base pairs and are located in ~40% of gene promoters across mammalian genomes (Deaton and Bird, 2011; Janitz and Janitz, 2011; Thomson *et al.*, 2010; Clouaire *et al.*, 2012; Blackledge *et al.*, 2013). STRs are commonly located near CpG islands (Sun *et al.*, 2018), and may influence their methylation states (Figure 1 (iii)). Moreover, other STRs can contain CpGs within their repetitive sequence that can undergo methylation (Bolton *et al.*, 2013).

A genome-wide study in yeast (Vinces *et al.*, 2009) estimates as many as 25% of promoters contain tandem repeats (TRs). Generally, expression of genes with TRs in their promoters increased with increasing repeat size. TRs in promoters may increase gene expression by increasing transcription

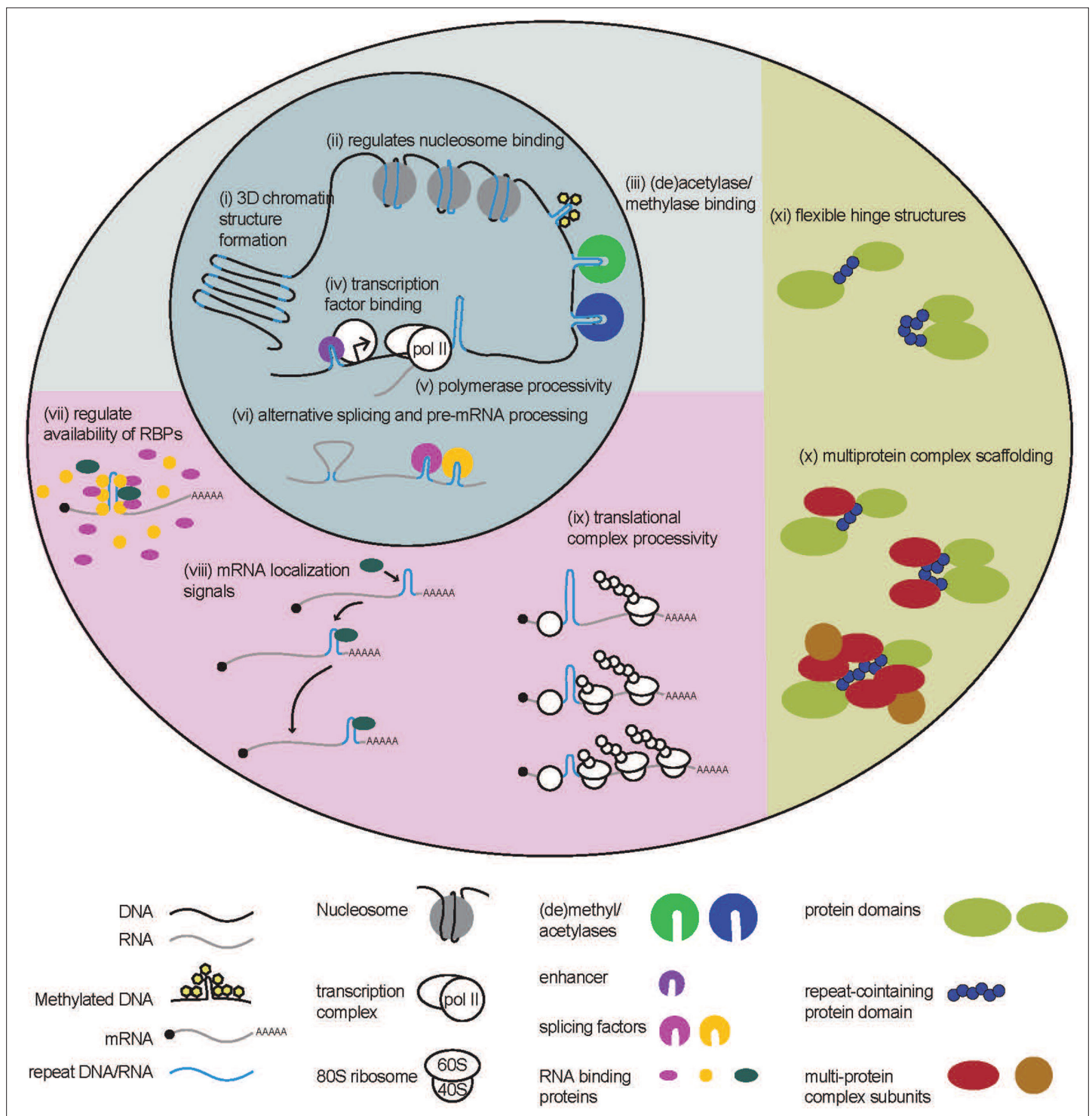


Figure 1. Native functions of genomic repeats. Repeats in DNA influence larger 3D chromatin structures, regulate binding of nucleosomes and (de)acetylases and (de)methylases. They also influence transcription factor binding and polymerase processivity to affect downstream RNA production. Repeats in RNA can affect pre-mRNA processing such as alternative splicing and can affect RNA binding protein function through direct or indirect sequestration. Repeats in 3' UTRs serve as localization signals, directing mRNA transport. Repeats in 5' UTRs regulate translational output by impeding ribosome processivity. Repeating units in proteins can provide structural flexibility within a protein or serve as binding sites for the formation of multi-protein complexes.

factor binding (**Figure 1** (iv)), blocking or reducing nucleosome density, or in the case of AT-rich repeats, by facilitating DNA melting (**Vinces et al., 2009**). However, stable secondary structures formed by TRs can also inhibit transcription by impeding procession and access of transcriptional machinery (**Figure 1** (v); **Grabczyk and Fishman, 1995**; **Belotserkovskii et al., 2010**; **Usdin and Woodford, 1995**). For example, the evolutionarily conserved THO complex is recruited to actively transcribed genes (**Kim et al., 2004**; **Abruzzi et al., 2004**; **Strässer et al., 2002**), and facilitates elongation of RNA polymerase (**Fan et al., 1996**; **Prado et al., 1997**; **Fan et al., 2001**; **Jimeno et al., 2002**; **Chávez et al., 2000**) through super-helical structures formed by long GC-rich TRs (**Voynov et al., 2006**; **Chávez et al., 2001**). Yeast strains with mutations to THO complex subunits exhibited lower levels of TR-containing FLO11 mRNA. Reduced FLO11 mRNA coincided with an accumulation of RNA polymerase at the beginning of the gene. Removal of the TR or overexpression of topoisomerase I to enhance unwinding of the structured DNA, rescued the reduction in FLO11 mRNA in THO complex mutants (**Voynov et al., 2006**).

As in yeast, human STRs can either enhance and inhibit transcription of associated genes dependent on their sequence and locations, and also affect gene expression via changes in gene methylation and chromatin structure (**Gymrek et al., 2016**; **Fotsing et al., 2019**; **Quilez et al., 2016**; **Garg et al., 2021**; **Jakubosky et al., 2020**). Together, these studies demonstrate numerous mechanisms by which TRs can enhance or inhibit gene expression.

STRs in RNA regulate pre-mRNA processing and RNA localization

Transcribed repetitive elements regulate numerous aspects of RNA biology. STRs in RNAs form complex higher order structures, including G-quadruplexes and hairpins (**Krzyzosiak et al., 2012**; **Sobczak et al., 2003**; **Malgowska et al., 2014**; **Sobczak et al., 2010**), which are thought to exert broad influence over pre-mRNA splicing (**Figure 1** (vi)) (**Muro et al., 1999**; **Tu et al., 2000**; **Black, 2003**; **Solnick and Lee, 1987**). An analysis of human introns found that sites of alternative splicing are enriched for STRs (**Lian and Garner, 2005**). STRs can facilitate alternative splicing by complementary pairing of intronic repeats, bringing exonic regions into close proximity (**Lian and Garner, 2005**). Structure-forming STRs can inhibit or enhance alternative splicing by blocking or facilitating the recruitment of splicing factors, respectively (**Lian and Garner, 2005**). For example, alternative splicing of the E11B exon in the well-conserved fibronectin gene is regulated by an intronic TGCATG repeat (**Huh and Hynes, 1994**; **Lim and Sharp, 1998**). Contractions in this STR reduce E11B exon inclusion, while overexpression of a specific splicing factor, SRp40, stimulates inclusion. While the TGCATG repeat differs from SRp40's consensus binding site, it can form a strong hairpin structure, which is a key feature of SRp40 binding site motifs (**Tacke et al., 1997**). This suggests that the TGCATG repeat may modulate alternative splicing by recruiting the SRp40 splicing factor to the intron/exon boundary (**Lim and Sharp, 1998**).

Some STRs in RNA regulate splicing in trans, by binding to and sequestering splicing factors and blocking their functions (**Figure 1** (vii)). A recent study identified a group of novel long non-coding RNAs (lncRNAs) with multiple predicted RNA binding motifs (**Yap et al., 2018**), a subset of which contained long stretches of STRs ('strRNAs'). One strRNA called the pyrimidine-rich noncoding transcript (PNCTR) contains numerous stretches of (TC)_n repeats, avidly binds to the polypyrimidine tract-binding protein (PTBP1) in cells, and negatively regulates PTBP1-mediated splicing (**Yap et al., 2018**). As such, PNCTR overexpression was sufficient to trigger mis-splicing of PTBP1 targets and trigger programmed cell death (**Yap et al., 2018**). In this way, STRs in RNA can regulate the global availability of other RNA-binding proteins (RBPs) with other functions, exerting profound control over numerous aspects of cell biology.

STRs in 3' UTRs can also serve as RNA localization signals, and via interactions with RBPs, facilitate the transport of RNAs to specified cellular compartments (**Figure 1** (viii)). A program called REPFIND was developed to analyze 3' UTRs of localized mRNAs in *Xenopus* oocytes and identified various CAC-containing repeat motifs that serve as localization elements (**Betley et al., 2002**). Mutating these CAC-containing repeats was sufficient to abolish normal RNA localization. CAC-containing repeats were also found in zebrafish and human 3' UTRs of transcripts that are known to be specifically localized within cells, suggesting that CAC-containing repeats are conserved localization elements in chordates (**Betley et al., 2002**). REPFIND was subsequently used to generate a database of repeating motifs in 3' UTRs of mammalian genes from the Mammalian Gene Collection (MGC) that revealed

hundreds of human genes containing short CAC- and CAG-rich repeats in their 3' UTRs (**Lim and Sharp, 1998**). Intriguingly, these elements facilitate RNA localization to neurites in rat hippocampal neurons (**Andken et al., 2007**).

Repetitive RNA regulates translation

STRs located in 5' UTRs and coding regions impact mRNA translational efficiency. GC-rich STRs form stable RNA structures (**Krzyzosiak et al., 2012; Sobczak et al., 2003; Malgowska et al., 2014; Sobczak et al., 2010**), which can impede the processivity of scanning translational complexes (**Figure 1 (ix)**) (**Kozak, 1980; Kozak, 1986; Kudla et al., 2009; Tuller et al., 2011; Ding et al., 2012; Bentele et al., 2013; Weinberg et al., 2016**). For example, a native GGN repeat in the 5' UTR of the potassium 2-pore domain leak channel Task3 mRNA forms a G-quadruplex structure in vivo (**Maltby et al., 2020**). This G-quadruplex is inhibitory to translation of Task3 mRNA, but can be overcome by DHX36 helicase activity to improve ribosome processivity through the stable structure (**Maltby et al., 2020**).

Indeed, libraries of synthetic (**Millette et al., 2022**) and naturally occurring (**Li et al., 2017; Niederer et al., 2022**) hairpin sequences placed within 5' UTRs can be used to precisely control translational transgene output, with potential implications for gene therapy dosing. These studies show how single unit variations in STRs can precisely modulate protein expression, generally permitting more and faster translation of mRNAs with smaller STRs, and less and slower translation of mRNAs with larger STRs.

Repeats in proteins facilitate multi-protein complex formation and structural flexibility

Eukaryotic proteins are more likely to have repeats than prokaryotic proteins, and proteins containing repeats are often unique to eukaryotes and eukaryotic functions (**Marcotte et al., 1999**). There are numerous long repeating motifs in proteins (>20 amino acids/repeat) with loose homology between repeats, that form complex tertiary structures (**Andrade et al., 2001**). These protein repeat domains are characterized by the structures they form, as all- β (i.e. β -propellers, β -trefoils), all- α (i.e. HEAT and tetratricopeptide repeats (TPRs)), or mixed α/β (i.e. leucine-rich repeats, ankyrin repeat; **Andrade et al., 2001**). Although their specific functions vary, protein repeat domains typically serve as binding sites, and are thought to have evolved in eukaryotes to aid in the formation of multi-protein complexes with advanced cellular functions (**Andrade et al., 2001; Kajava, 2012; Sharma and Pandey, 2015**).

STRs translated into proteins, are thought to have similar functions as these larger repeat-based protein domains, serving as sites for protein-protein interactions (**Figure 1 (x); Schaefer et al., 2012; Faux, 2012**). CAG repeats are enriched in coding regions and are most frequently found in the polyglutamine (polyQ) reading frame, suggesting that polyQ stretches in proteins have a native function (**Schaefer et al., 2012**). PolyQ stretches are enriched in proteins that are components of multi-protein complexes, and have functions in transcriptional control, phosphatidylinositol (PI) signaling, protein degradation, and chromatin remodeling. Evolutionary sequence comparison reveals that the location of polyQs within a protein is not always conserved (**Schaefer et al., 2012**). This suggests that polyQ stretches have evolved multiple times, and don't directly confer a protein's function, but rather modulate the protein-protein interactions necessary for those functions (**Schaefer et al., 2012; Orr, 2012**). Other CG-containing STRs (i.e. CUGs and CGGs) show similar patterns of overrepresentation in coding regions (**Schaefer et al., 2012**), and likely serve similar complex-scaffolding functions (**Nasrallah et al., 2012**).

STRs when translated into proteins can be critical for proper protein folding. For example, translation of a CAG repeat in the huntingtin gene (*HTT*) produces a polyQ tract in the HTT protein which serves as a flexible hinge, allowing the neighboring domains to fold into close proximity (**Figure 1 (xi)**) (**Caron et al., 2013**). HTT protein structure is altered with repeat expansion, demonstrating the importance of the flexibility conferred by this STR (**Caron et al., 2013**).

Pathogenic consequences of STRs: A Fragile X case study

In the previous section, we summarized how STRs in DNA, RNA, and when translated into proteins can affect gene expression and protein function. In the following section, we will draw parallels from these native functions of STRs to pathogenic mechanisms in REDs (**Figure 2**). These parallels demonstrate

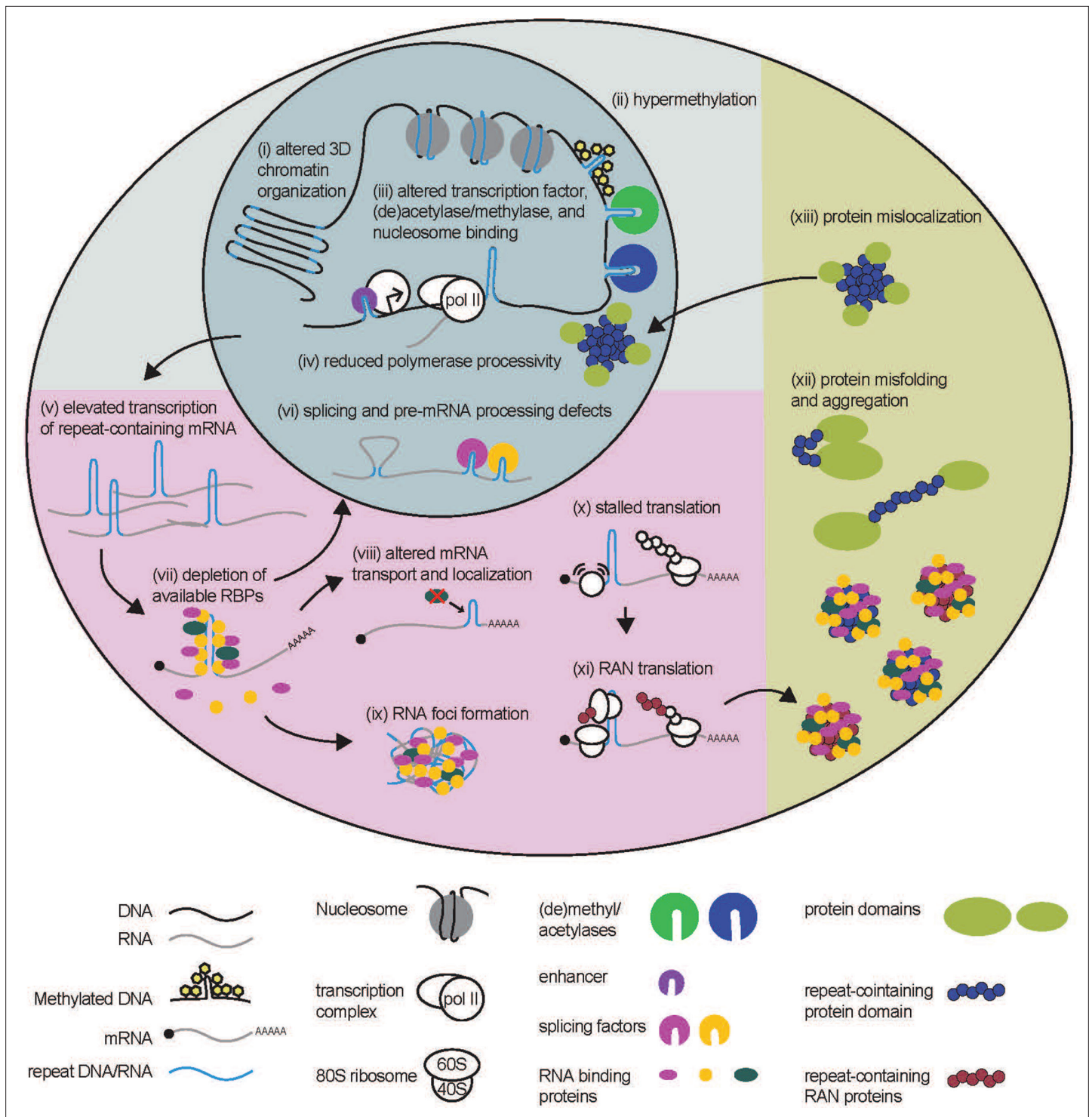


Figure 2. STR-associated toxicity in Repeat Expansion Disorders. Repeat expansions can alter global 3D chromatin structure, and influence transcription via blocking or enhancing binding of nucleosomes, (de)acetylases, (de)methylases, and transcription factors. Expanded repeats may also impede polymerase processivity. In some cases, elevated transcription of repeat expansion RNA can lead to depletion of RNA-binding proteins. Depletion of these proteins can impact many processes to which they contribute, including pre-mRNA splicing and processing, and mRNA localization. Expanded repeat RNA and bound RBPs can also aggregate into RNA foci, causing toxicity. Expanded repeat RNA can stall translational complexes, leading to repeat-associated non-AUG (RAN) translation, and contribute to the production of polymeric proteins. Polymeric proteins are aggregate prone. Longer polymeric stretches in native proteins may also cause dysfunction by preventing proper protein folding or causing the folded protein to mis-localize within the cell.

how STR toxicity can be viewed as aberrancies of native processes, rather than emergent dysfunctions. For this analysis, we will largely use the Fragile X locus discussed earlier as a well-characterized case study, although many of these principles also apply to other REDs and a few specific examples are included here (reviewed in broader detail in [Malik et al., 2021a](#); [Glineburg et al., 2018](#); [Paulson, 2018](#); [Rodriguez and Todd, 2019](#)).

Epigenetic and transcriptional dysfunction of STRs in DNA

The functional consequences of STRs on genome organization and transcription are evident when dysfunction is observed in REDs ([Dion and Wilson, 2009](#); [López-Martínez et al., 2020](#); [Yin et al., 2020](#); [Usdin, 2008](#)). Repeat expansions can alter local genome architecture and expression of neighboring genes. A prime example is observed at a CTG repeat in the 3'UTR of the *DMPK* gene associated with myotonic dystrophy type 1 (DM1), expansion of which alters local chromatin structure and suppresses transcription of neighboring gene, *Six5* ([Winchester et al., 1999](#); [Brouwer et al., 2013](#); [López Castel et al., 2011](#)). Repeat expansions also cause global alterations in chromatin structure. CGG repeat expansions in FXS patients cause severe disruptions in chromatin boundaries ([Figure 2 \(i\)](#); [Sun et al., 2018](#)). These disruptions may explain delayed DNA replication ([Subramanian et al., 1996](#)), activation of DNA replication stress pathways ([Chakraborty et al., 2020](#)) and altered local DNA replication patterns ([Gerhardt et al., 2014](#)) observed at CGG repeat expansions and the Fragile X locus in particular.

As genomic repeats influence native DNA methylation, some STRs are aberrantly methylated only upon expansion ([Figure 2\(ii\)](#); [Otten and Tapscott, 1995](#); [Steinbach et al., 1998](#); [Herman et al., 2006](#); [Greene et al., 2007](#); [Belzil et al., 2013](#); [Xi et al., 2013](#)). When the CGG repeat in the 5' UTR of *FMR1* expands beyond 200 repeats, it is susceptible to DNA methylation of both the CpG elements within the repeat and at a CpG element within the *FMR1* promoter ([Oberlé et al., 1991](#); [Sutcliffe et al., 1992](#); [Pieretti et al., 1991](#); [McConkie-Rosell et al., 1993](#); [Hansen et al., 1992](#); [Coffee et al., 2002](#); [Colak et al., 2014](#); [Willemsen et al., 2002](#)). This hypermethylation is associated with *FMR1* gene silencing, with a resulting absence of *FMR1* mRNA and FMRP, a critical RBP involved in synaptic plasticity and neuronal function ([Oberlé et al., 1991](#); [Hagerman et al., 2017](#); [Quartier et al., 2017](#); [Myrick et al., 2014](#)). How exactly repeat expansion triggers methylation and the relationship between expansion, methylation, and epigenetic silencing is not fully understood, but the locus remains transcriptionally active and unmethylated in human embryonic stem cells even in the presence of very large repeat expansions, with silencing occurring during differentiation. Some studies suggest that *FMR1* silencing requires co-transcriptional binding of CGG repeat mRNA directly to the *FMR1* promoter region as an RNA-DNA heteroduplex ([Colak et al., 2014](#); [Groh et al., 2014](#)).

STR expansions can enhance or inhibit mRNA production from nearby genes. At *FMR1*, premutation range CGG repeats which cause FXTAS or FXPOI (and which are unmethylated) result in elevated transcription of *FMR1* mRNA ([Tassone et al., 2000a](#); [Tassone et al., 2000b](#); [Entezam et al., 2007](#); [Brouwer et al., 2008](#); [Kenneson et al., 2001](#)). This may result from use of additional upstream transcription start sites ([Beilina et al., 2004](#); [Tassone et al., 2011](#)), or be associated with enrichment of acetylated histones or other chromatin activating factors at the premutation allele ([Todd et al., 2010](#)). It's possible that both hypo-expression and hyperexpression of *FMR1* stems from the complex structures formed by these CGG repeats as DNA ([Usdin and Woodford, 1995](#); [Fry and Loeb, 1994](#); [Kettani et al., 1995](#); [Patel et al., 2000](#)). As seen in native STRs, different structures formed by expanded STRs could facilitate or block binding of histone-modifying methylases, demethylases, acetylases, deacetylases, and even entire nucleosomes to affect downstream gene expression ([Figure 2 \(iii-v\)](#)) ([Wang et al., 1996](#); [Usdin and Kumari, 2015](#)).

Repeat expansions cause defects in pre-mRNA processing and mRNA localization

The native roles of STRs in RNA in regulating splicing mirror splicing dysfunction observed in numerous REDs ([Figure 2 \(vi\)](#)). Splicing of the *HTT* huntingtin gene, which contains a CAG repeat, is altered at expanded repeats associated with Huntington's Disease, resulting in the production of a transcript containing only exon 1 and the production of an exon 1 HTT protein ([Gipson et al., 2013](#); [Sathasivam et al., 2013](#); [Neueder et al., 2017](#); [Neueder et al., 2018](#); [Franich et al., 2019](#)). The exon 1 HTT protein is found in patient tissues and is toxic in model systems ([Gipson et al., 2013](#); [Sathasivam](#)

et al., 2013; Neueder et al., 2017; Neueder et al., 2018; Franich et al., 2019). Incomplete splicing of *HTT* with the CAG repeat expansion increased with overexpression and decreased with knockdown of splicing factor SRSF6. SRSF6 is predicted to bind to the 5' end of *HTT* transcripts via the CAG repeat, suggesting that SRSF6-CAG repeat interactions interfere with spliceosome formation at the nearby splice site (*Neueder et al., 2018*).

Global splicing defects in REDs result from sequestration of critical splicing factors that bind to STR-containing RNA (*López-Martínez et al., 2020; Mykowska et al., 2011; Botta et al., 2008*). STRs can be binding sites for RBPs, regulating their availability throughout the cell. RNAs with longer STRs bind more RBPs, which can cause a global cellular depletion of these factors (*Figure 2 (vii)*) (*Malik et al., 2021a; Glineburg et al., 2018; Rodriguez and Todd, 2019*). In myotonic dystrophy type 1 (DM1), the expanded CTG repeat in the 3' UTR of the *DMPK* gene (*López-Martínez et al., 2020; Korade-Mirnic et al., 1998; Udd and Krahe, 2012*) binds to muscleblind-like splicing regulator 1 protein (MBNL1) among other RBPs (*Paul et al., 2011; Jiang et al., 2004; Fardaei et al., 2001; Mankodi et al., 2001; Miller et al., 2000*), resulting in depletion of critical splicing factors (*Botta et al., 2008; Jiang et al., 2004; Pascual et al., 2006; Jog et al., 2012*) and global splicing defects (*López-Martínez et al., 2020; Paul et al., 2011; Jiang et al., 2004; Du et al., 2010; Philips et al., 1998*). RBP depletion by the CTG repeat in DM1 can also impact other aspects of pre-mRNA processing, including polyadenylation (*Thomas et al., 2017; Goodwin et al., 2015*).

In FXTAS, premutation expansion mRNA sequesters and depletes multiple RBPs that bind to the CGG repeat RNA directly (i.e. DGCR8 *Sellier et al., 2013*), Pur α (*Jin et al., 2007*), hnRNP A2/B1 (*Jin et al., 2007; Iwahashi et al., 2006; Muslimov et al., 2011; Sofola et al., 2007*) or indirectly via binding to CGG-bound proteins (i.e. Drosha *Sellier et al., 2013* and Sam68 *Sellier et al., 2010*). These RBPs are involved in a variety of functions that are affected by their sequestration, including miRNA processing (DGCR8, Drosha), mRNA transport (hnRNP A2/B1, Pur α) (*Figure 2 (viii)*), and in splicing (hnRNP A2/B1, Sam68). Splicing defects have been observed in CGG premutation expansion models (*Sellier et al., 2010; He et al., 2014*), but compensatory overexpression of CGG repeat-sequestered RBPs (*Jin et al., 2007; Sofola et al., 2007; He et al., 2014; Qurashi et al., 2011*) or blocking RBP binding to CGG repeats (*Disney et al., 2012; Verma et al., 2019; Verma et al., 2020; Verma et al., 2022*) can improve these disease-associated defects. This pathogenic sequestration of RBPs by expanded repeats mirrors the native role for STRs in RNA as RBP reserves, mediating fine-tuned dosing of RBP availability with repeat length.

In addition to the depletion of RBPs and consequent defects in RNA splicing and localization and miRNA processing, expanded STRs in RNAs may also cause toxicity by self-association (gelation) (*Figure 2(ix); Glineburg et al., 2018; Sellier et al., 2010; He et al., 2014; Jain and Vale, 2017; Ciesiolka et al., 2017; Fay et al., 2017; Tassone et al., 2004*). Yet, these processes also occur on RNAs with shorter STRs that are below the pathological threshold for disease, suggesting such that such phase separation properties of specific RNA motifs and their associated RBPs may exist on a spectrum from physiologic to pathologic.

Expanded STRs in RNA can mis-localize or be retained in the nucleus instead of transported to its functional location in the cell (*Davis et al., 1997; Mastroyiannopoulos et al., 2010; Sun et al., 2015*). This may be mediated by splicing defects (*Sun et al., 2015*), via export-inhibiting RBP interactions (*Smith et al., 2007*), or via a larger dysfunction of nucleocytoplasmic transport (*Zhang et al., 2015; Zhang et al., 2016; Jovičić et al., 2015; Freibaum et al., 2015; Grima et al., 2017; Gasset-Rosa et al., 2017; Sellier et al., 2017*). For example, SRSF proteins bind to CGG and G4C2 repeats and appear critical to their cytoplasmic transport out of the nucleus (*Malik et al., 2021b; Hautbergue et al., 2017*). In this context, lowered expression of SRSF proteins or inhibition of the SRSF protein kinase SRPK1, which regulates SRSF nuclear entry, suppress CGG repeat exit to the cytoplasm and reduce toxicity in *Drosophila* and neuronal model systems (*Malik et al., 2021b*). Together, these studies show that expanded STRs in RNA can induce toxicity via RBP depletion or by direct RNA dysfunction.

Aberrant translation of expanded STRs

Scanning translational complexes are more likely to stall at stable secondary structures formed by expanded STRs (*Figure 2 (x)*), resulting in aberrant translation initiation upstream of or within the repeat in a process known as repeat-associated non-AUG (RAN) translation (*Figure 2 (xi)*). RAN

translation produces toxic peptides that contribute to expanded STR toxicity and neurodegeneration in numerous REDs (*Glineburg et al., 2018; Ciesiolka et al., 2017; Cleary and Ranum, 2017; Kearse and Todd, 2014; Kearse et al., 2016; Todd et al., 2013; Wojciechowska et al., 2014; Mori et al., 2013a; Ash et al., 2013; Mori et al., 2013b; Bañez-Coronel et al., 2015; Zu et al., 2011; Zu et al., 2017; Soragni et al., 2018; Ishiguro et al., 2017*).

The mechanisms underlying RAN translation likely vary across different STRs and different genetic contexts (*Malik et al., 2021a; Gao et al., 2017*). At the CGG repeat of *FMR1*, RAN translation occurs in all three reading frames to produce polyarginine (FMRpolyR) (+0-frame relative to the AUG of *FMR1*), polyglycine (FMRpolyG) (+1-frame), and polyalanine (FMRpolyA) (+2-frame) peptides at different efficiencies (*Kearse et al., 2016; Todd et al., 2013*). STR-induced stalling of translation machinery is also responsible for a reduction in downstream production of the main protein produced by *FMR1* translation, FMRP, in CGG premutation carriers (*Tassone et al., 2000b; Kenneson et al., 2001*).

Repetitive proteins have pathogenic consequences

Repeat-containing peptides, produced via canonical translation of STRs in coding regions or via RAN translation, contribute to toxicity in REDs. At CGG repeats, both FMRpolyG and FMRpolyA are present within intranuclear neuronal inclusions in patient tissues (*Sellier et al., 2017; Todd et al., 2013; Buijsen et al., 2014; Krans et al., 2019; Ma et al., 2019*), and are toxic in model systems (*Sellier et al., 2017; Todd et al., 2013; Derbis et al., 2018; Gohel et al., 2019; Hoem et al., 2019*). FMRpolyG, the most abundant CGG RAN product, is necessary for CGG repeat toxicity and inclusion formation (*Sellier et al., 2017; Todd et al., 2013; Oh et al., 2015*) in overexpression models. Numerous RAN or homopolymeric peptides generated in other REDs are essential for their toxicity and formation of proteinaceous inclusions (**Figure 2** (xii); *Yamamoto et al., 2000; Schilling et al., 1999; Ordway et al., 1997; Bäuerlein et al., 2017; Paulson et al., 1997; Mizielinska et al., 2014; May et al., 2014; Zu et al., 2013*). Overall, dysfunctional aggregation of repeat derived protein products mirrors the native function of STRs in proteins as facilitators of protein-protein interactions.

Translation through large STRs that form stable secondary structures likely induces ribosome stalls and elongation errors. A growing body of work shows that disease-associated STRs undergo stall-induced translational frameshifting to produce novel chimeric polypeptides (*Gaspar et al., 2000; Toulouse et al., 2005; Davies and Rubinsztein, 2006; Tabet et al., 2018; McEachin et al., 2020; Wright et al., 2022*), and several of these studies have shown that these frameshift products have distinct contributions to neuronal dysfunction in disease (*Tablet et al., 2018; McEachin et al., 2020; Wright et al., 2022*). While there is evidence that polymeric peptides contribute to toxicity observed in REDs via aggregation, the mechanistic details of homo- and di-polymeric peptide toxicity and chimeric polypeptide toxicity remain under investigation.

Antisense transcripts contribute to REDs via multiple mechanisms

Antisense transcription from the *FMR1* locus generates multiple long-noncoding *asFMR1* mRNAs, with some including the repeat (*Ladd et al., 2007; Khalil et al., 2008; Elizur et al., 2016; Pastori et al., 2014*). One antisense transcript, *FMR4*, is thought to play a critical role in regulating the cell cycle and apoptosis (*Khalil et al., 2008*). Another antisense transcript, *FMR6*, is upregulated in premutation women, with increased repeat length correlating to elevated RNA levels and reduced number of oocytes, suggesting a relationship between antisense transcript expression and toxicity (*Elizur et al., 2016*). *FMR1* antisense transcription in general is upregulated in Fragile X premutation disorders and lost in FXS, like the sense *FMR1* mRNA (*Ladd et al., 2007*). Moreover, *asFMR1* mRNAs containing the CCG repeats can undergo RAN translation, producing additional homopolymeric proteins with toxic potential (*Kearse et al., 2016*). STR-containing antisense transcripts likely contribute to toxicity observed in many REDs, but this is best characterized in *C9ALS/FTD* and *SCA8*, where antisense transcripts are found in toxic RNA foci and contribute to RAN peptide production (*Mori et al., 2013a; Zu et al., 2011; Zu et al., 2013; Moseley et al., 2006; Gendron et al., 2013*).

Mechanisms of STR toxicity reveal novel native functions of STRs

Studies over the past three decades have delineated numerous mechanisms by which repeat expansions trigger cellular toxicity. Yet, there are striking parallels between the pathologic drivers of dysfunction elicited by repeat expansions and the native functions of STRs in regulating gene expression.

In this section, we provide examples of how mechanisms initially identified as causing STR toxicity directly inform our understanding of native functions of STRs more broadly. We also discuss how emergent pathogenic properties associated with repeat expansions might inform additional native functions of repeats that are not yet well understood.

RAN translation occurs at native repeat lengths and have native functions

While CGG repeats in the *FMR1* gene were primarily studied for their disease association, the CGG repeat is present in all humans at nonpathogenic lengths (<55 repeats) and conserved across mammals (Eichler et al., 1995; Sellier et al., 2017). Some studies suggest phenotypes associated with low CGG repeat numbers at this allele in humans, including memory difficulties and language dysfluency (Klusek et al., 2018; Mailick et al., 2014). Our group observed that CGG RAN translation, originally thought to be an aberrant toxic event, occurs in reporters with native repeat lengths (25 repeats) (Kearse et al., 2016), suggesting CGG repeats and/or translation of those repeats may have a native function in addition to the pathogenic one. CGG RAN translation at native and expanded STRs acts as an overlapping upstream open reading frame (uORF), inhibiting translation of the downstream main ORF (mORF) and thereby reducing FMRP synthesis (Rodriguez et al., 2020). Furthermore, this RAN uORF-like regulation of FMRP synthesis was critical for facilitating translational changes associated with stimulation of metabotropic glutamate receptors (mGluRs) in neurons (Rodriguez et al., 2020).

Upstream open-reading frames (uORFs) are well-characterized regulatory elements in eukaryotes that influence expression of protein produced from the main open reading frame (mORF) on the same transcript, and are typically inhibitory to downstream mORF translation (Hinnebusch et al., 2016). In this way, uORFs resulting from RAN translation of STRs may play a global role in regulating mRNA translation, presenting another mechanism by which STRs influence gene expression.

STRs facilitate protein function and localization

Expanded STRs in coding regions can fundamentally change the functions of the proteins within which they reside. In spinocerebellar ataxia type 1 (SCA1), a CAG repeat expansion in the ataxin 1 (*ATXN1*) gene changes the localization of *ATXN1* protein (Irwin et al., 2005). *ATXN1* normally shuttles between the nucleus and the cytoplasm, but an expanded polyQ region shifts *ATXN1* localization to the nucleus (Figure 2(xiii); Irwin et al., 2005). Aberrant nuclear localization of *ATXN1* underlies dysfunction in SCA1 (Lam et al., 2006; Lai et al., 2011; Klement et al., 1998; Emamian et al., 2003; Duvick et al., 2010), as modifications that favor nuclear localization are sufficient to elicit disease relevant phenotypes in the absence of the repeat expansion in mouse models.

PolyQ-associated nuclear translocation is also central to pathology in spinal and bulbar muscular atrophy (SBMA), where ligand binding and translocation to the nucleus of the expanded PolyQ-containing androgen receptor is required to elicit disease-associated transcriptional defects and cytotoxicity (Katsuno et al., 2006; Katsuno et al., 2002; Montie et al., 2009; Palazzolo et al., 2007). However, within the normal range of polyQ lengths observed in humans, Androgen receptor CAG repeat size inversely correlates with the receptor's transactivational activity and linearly correlates with infertility and decreased sperm function (Choong and Wilson, 1998; Osadchuk et al., 2022; Pan et al., 2016). These findings suggest that the CAG repeats play a normal role in testosterone activated gene cascades that become aberrant at larger repeat sizes.

STRs facilitate mRNA transport to dendrites

An investigation into dendritic mRNA localization identified a localization pathway dependent on the interaction of a CGG repeat-interacting RBP, hnRNP A2, with a GA dendritic targeting element of an RNA (Muslimov et al., 2011). This GA targeting motif was competed for by CGG repeat-containing RNAs, including *FMR1* mRNA. In addition to a native function of CGG repeats as a dendritic localization factor, this study revealed that elevated levels of CGG repeat mRNA caused by the CGG premutation expansion sequester hnRNP A2, resulting in global dysfunction in the transport of hnRNP A2-target mRNAs (Muslimov et al., 2011). Another study seeking to reveal transcriptome-wide impacts of C(C)UG repeat-mediated MBNL depletion on splicing in myotonic dystrophy (DM) also uncovered a global role for MBNL in mRNA localization (Wang et al., 2012).

PolyQ containing proteins regulate autophagy

Numerous REDs are caused by CAG repeats, including the huntingtin gene in Huntington's disease (HD) and Ataxin 3 in spinocerebellar ataxia type 3 (SCA3), with toxicity largely attributed to the aggregation of long polyQ containing proteins. Autophagy induction results in clearance of these aggregates, attenuating their toxicity (*Rubinsztein, 2006; Ravikumar et al., 2004; Menzies et al., 2010*). PolyQ tracts in ataxin 3, a deubiquitinase associated with spinocerebellar ataxia type 3 (SCA3), interact with beclin 1, a key initiator of autophagy (*Ashkenazi et al., 2017*). Ataxin 3 then deubiquitinates beclin 1, protecting it from degradation and permitting autophagy initiation. Ataxin 3 activity and interaction with beclin 1 is competitively inhibited by other polyQ tract-containing proteins in a length-dependent manner (*Ashkenazi et al., 2017*). As such, polyQ tracts may actively engage protein quality control pathways basally but then these interactions become aberrant after STR expansion, in this case inhibiting autophagy and clearance of toxic proteins. Together, these studies suggest that the pathology of disease-associated STR expansions reveal native functions of STRs, just as an improved understanding of the native functions of STRs can inform on dysfunctions in disease.

Tetranucleotide, pentanucleotide, and biallelic repeat expansion disorders

Tetranucleotide and pentanucleotide STRs are rare within coding sequences, presumably because changes in their repeat number would trigger translational frameshifts. However, they are relatively common within introns, where their expansion causes several neurological disorders that likely act through pathogenic mechanisms that are similar to those exhibited by non-coding trinucleotide STRs. For example, Myotonic dystrophy type 2 (DM2) results from a dominantly inherited intronic CCTG STR expansion in *ZNF9* (*Liquori et al., 2001*). CCTG STRs form RNA secondary structures that are like those generated by CTG STRs, and like the 3' UTR CTG repeat in DM1, the DM2 repeat binds to and sequesters the RBP muscleblind (*Botta et al., 2008; Paul et al., 2011; Fardaei et al., 2001; Mankodi et al., 2001; Miller et al., 2000; Du et al., 2010; Philips et al., 1998*). This shared mechanism explains the significant overlap in their clinical phenotypes. Perhaps more interesting, however, is how subtle differences in how these repeats underlie the phenotypic differences in these conditions. In particular, CCTG expansions in DM2 do not trigger genetic anticipation or congenital forms of disease as occurs in DM1 despite the presence of very large CCTG expansions in DM2. These phenotypic differences are thought to occur for two reasons. First, these repeats exhibit differences in how they interact with other RBPs, such as rbFOX, that modulate the effects of muscleblind sequestration (*Sellier et al., 2018*). Second, differences in the genic positioning (intron versus 3' UTR) and temporal expression of the two STRs alters their relative abilities to disrupt early developmental processes (*Thomas et al., 2017; Cerro-Herreros et al., 2017*).

An intriguing feature observed in multiple pentanucleotide repeat expansion disorders, including complex TTTTA and TTTCA repeats that cause benign adult familial myoclonic epilepsy (BAFME) in multiple genes (*Ishiura et al., 2018*), ATTC repeats in Spinocerebellar ataxia (SCA) type 31 (*Sato et al., 2009*), and AAGGG repeats that cause cerebellar ataxia, neuropathy, vestibular areflexia syndrome (CANVAS) (*Nakamura et al., 2020; Cortese et al., 2019; Tsuchiya et al., 2020*), is that the pathogenic alleles represent non-reference STRs. That is, the repeat element is not only expanded in size, but it has a different sequence than the normal allele. For example, in CANVAS, an AAAAG pentanucleotide STR normally resides within the first intron of *RFC1*. However, the pathological repeat is a qualitatively different and expanded AAGGG repeat. Moreover, CANVAS can also occur with a third pentanucleotide repeat, ACAGG, at this same genomic location. In all these cases, these pentanucleotide repeats occur within the polyA region of an Alu transposable element. Active Alu transposition requires pure polyA elements at their 3' ends (*Deininger, 2011*). As such, there is strong evolutionary selection pressure favoring mutation of these regions to pure polyA sequences. This suggests that both the reference and non-reference STRs occurred initially through a protective process that disrupted the polyA element and prevented continued transposase activity. However, stochastic differences in these interrupting mutations created some STRs that were more prone to expansion, resulting in pathogenic alleles that either create toxic STR RNAs or that interfere with local gene expression.

Conclusions and open questions

Native functions of STRs from an evolutionary perspective

Evolutionary pressures on STR copy number predict that repeat expansions will be either tolerated or selected for until an upper, deleterious limit is reached. If STRs were intrinsically deleterious, then there would be selective pressure towards repeat contractions, leading to global reductions in STR size or their selective elimination. However, several recent studies suggest that there is selective STR expansion across phylogeny, especially in primates. This is particularly true in 5' UTRs and coding regions, where constraints on repeat expansion and contraction are greatest (**Gymrek et al., 2017**). At the same time, many intergenic STRs show correlations between their size and the expression of neighboring genes. These eSTRs contribute meaningfully to population variance in gene expression profiles and disease associated Quantitative trait loci (eQTLs) in human populations (**Fotsing et al., 2019; Gymrek et al., 2017**). In general, these eSTRs are largely unconstrained unless neighboring or embedded within a gene already under strong constraint. This mutation-selection balance suggests some inherent native functions of STRs within transcript and protein space while also implying that intrinsic STR instability may allow for more rapid variation and acquisition of traits through local perturbations in gene expression than could be accomplished through single nucleotide mutations (**Figure 3A**). The highly variable methylation status, mRNA and protein expression patterns elicited by differences in *FMR1* CGG repeat lengths typify the potential for repeat variation within populations to influence gene expression (**Figure 3B**). As even subtle changes in repeat size can tune gene expression and protein function and have downstream impacts on simple and complex phenotypes, they may be an important component of the genetic differences between humans and other species, and among humans themselves.

Revisiting our understanding of and approach to Repeat Expansion Disorders

Historically, repetitive elements within human genomes have been viewed as mostly unregulated 'junk DNA' that is not under selective evolutionary pressure. As such expansions of these repetitive elements are unfortunate accidents which become apparent and important only when they elicit highly penetrant and syndromic human diseases. Consistent with this line of reasoning, the field of REDs has largely focused on emergent toxic mechanisms as drivers of disease only in the setting of large STR expansions rather than considering their pathology as alterations in the native functions played by these repeats in their normal genomic contexts. Here, we propose re-framing the discussion around repetitive elements in general- and STRs in particular- within human genomes. For each STR, we suggest first considering whether the STRs associated with a human disease have any native functions at their 'normal' size. If a native function exists, then expansion of these STRs can be viewed primarily as an aberrancy of that native function with coincident predictable impacts on gene expression dysregulation above certain repeat lengths. This reframing aligns with the approach typically taken in studying gain-of-function and loss-of-function mutations in disease associated single amino acid mutations and better ties the native functions of STRs to their pathology. It also suggests that shared regulatory rules will likely apply across REDs.

This approach to thinking about REDs leads to specific predictions. First, we predict that more REDs will be discovered in the future. For example, multiple recently described REDs are linked to CGG repeats, including neuronal intranuclear inclusion disease (NIID) (**Ishiura et al., 2019; Sone et al., 2019**), oculopharyngodistal myopathy (OPDM) and leukodystrophy (OPML) (**Ishiura et al., 2019; Deng et al., 2020; Ogasawara et al., 2020; Tian et al., 2019**), adult onset leukoencephalopathy (**Okubo et al., 2019**), and autism/intellectual disability (**Annear et al., 2022**). Most of these new CGG repeatopathies reside within the 5' UTRs, like the CGG repeat in *FMR1*, and there is already evidence of convergent disease mechanisms triggered by these new repeats with those already established in Fragile X disorders. In one particularly notable example, a CGG repeat expansion in *NOTCH2NLC* leads to the creation of an AUG-initiated upstream open reading frame in the 5' UTR that is generates a polyglycine-containing protein akin to FMRpolyG in *FMR1* (**Liu et al., 2022; Boivin et al., 2021**). This polyglycine protein is found within inclusions in patients with NIID and its generation is required to trigger inclusion formation and behavioral phenotypes in a mouse model of *NOTCH2NLC* associated NIID. As such, we know that this motif in this location within neuronally expressed genes can elicit dysfunction through predictable mechanisms. This means that we should expect other CGG

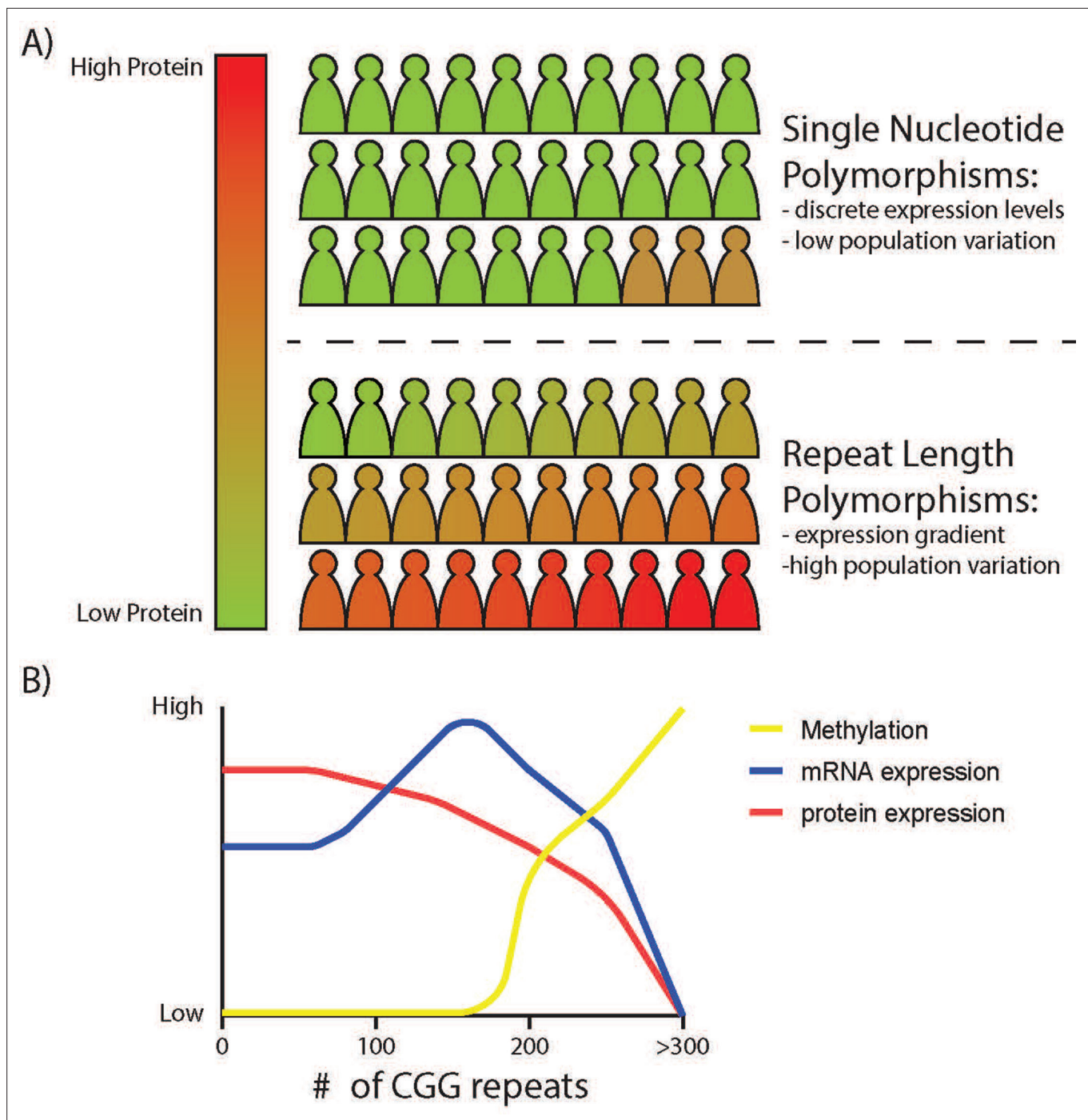


Figure 3. The effects of STR variation on local gene expressivity. **(A)** Bi-allelic variation in a gene through single nucleotide polymorphisms often result in small and discrete differences in gene expression, offering limited phenotypic differences across a population with slow evolutionary timescales. In contrast, STRs in promoters and 5'UTRs can influence protein expression over a broader dynamic range, with an inverse correlation between repeat length and protein output within transcribed regions and with differential effects on transcription dependent on the repeat and local epigenetic context. Unstable repeats change rapidly from generation to generation (and even within an individual through somatic variation), creating a mechanism by which mRNA or protein expression can vary broadly and subtly across a population, offering greater genetic and phenotypic diversity and a greater propensity for disease-causing aberrancies at the extremes. **(B)** Predicted effects of CGG repeat length on FMR1 gene expression. CGG repeat length influences FMR1 promoter epigenetic state (more open chromatin with initial expansion, then DNA methylation and closed chromatin at >200 CGG repeats), FMR1 mRNA expression, and FMRP protein expression across the polymorphic range.

repeat expansions to emerge that mirror the pathologic processes established for the *FMR1* locus and now extended to a large set of loci. Similarly, given evidence that the CGG repeat in *FMR1* 5'UTRs can serve as a functional element that regulates transcription, mRNA localization and translation, we predict that native CGG repeat elements in these disease-associated alleles may have normal functions akin to those observed for *FMR1*, and as such represent a functional motif shared among many genes.

However, these new REDs may not all fit the typical model observed to date, where highly penetrant STR expansions lead to syndromic disorders. Instead, smaller changes in repeat size at multiple loci, impacting expression of the genes in which they reside or neighboring genes, will serve as risk alleles for common conditions. This risk-allele model is already apparent, as intermediate CAG repeat expansions in *ATXN1*, *ATXN2*, and *HTT* are associated with sporadic ALS and some other common neurodegenerative disorders (Elden et al., 2010; Rosas et al., 2020). Indeed, a fair proportion of the unexplained signal within Genome Wide association Studies (GWAS) can be explained by variations within neighboring STRs (Gymrek et al., 2016; Gymrek, 2017; Hannan, 2018). To date, numerous STR variants have been linked to ASD (Mitra et al., 2021; Trost et al., 2020) and Schizophrenia (Mojarad et al., 2022). As PCR-free and long-read whole genome data becomes more abundant and available (reviewed in Mitsuhashi and Matsumoto, 2020), it will become increasingly easy to detect these dynamic repeat size/disease relationships, creating a whole new class of STR-associated conditions that will likely expand outside of neurological conditions.

Second, we predict that long-read whole genome sequencing datasets will improve our understanding of the native roles of STRs in humans, and reveal a ubiquitous impact of repeat length variation on gene expression. Once we create accurate maps of STR variation across the genome and link this variation to neighboring gene loci expression, we will be able to better discern the mechanisms by which STRs influence gene expression across cell types. We predict that many genes whose expression is affected by neighboring repeat length variation will play critical functions in the nervous system. Most known REDs present with neurological symptoms. If REDs reflect the native functions of STRs, then the overrepresentation of neurological dysfunctions linked to STR expansions suggests that STRs may play roles relevant to neuronal health and function. It is also possible that neurons, as terminally differentiated cells, may be more prone to somatic instability, leading to repeat expansion and the emergence of associated dysfunction with age.

Finally, we predict that the native functions of STRs will inform our understanding of how STR expansions cause disease and vice versa. A deeper understanding of the native functions of both disease-associated STRs and STRs in general will reveal the pathways altered in REDs, and these pathways may be areas for therapeutic intervention that can be applicable across all REDs. By studying the mechanisms by which STRs elicit disease, we can also surmise key elements of how they might function normally within nervous systems (see examples in previous section, "Mechanisms of STR toxicity reveal novel native functions of STRs"). Ultimately, research into native functions of STRs will reveal both mechanisms by which they regulate neuronal function and therapeutic targets by which their toxicity in REDs can be mitigated.

Acknowledgements

We thank current and former members of the Todd lab for helpful discussions and commentary. This work was funded by grants from the NIH to SEW (T32-NS076401 and NRSA F31NS113513) and PKT (P50HD104463, R01NS099280 and R01NS086810). PKT was also supported by the VA (BLRD BX004842 to PKT) and private philanthropic support.

Additional information

Competing interests

Peter K Todd: Dr Todd served as a consultant to Denali Therapeutics and holds a shared patent on ASOs developed with Ionis Pharmaceuticals. The other author declares that no competing interests exist.

Funding

Funder	Grant reference number	Author
National Institute of Neurological Disorders and Stroke	F31NS113513	Shannon E Wright
Eunice Kennedy Shriver National Institute of Child Health and Human Development	P50HD104463	Peter K Todd
National Institute of Neurological Disorders and Stroke	R01NS099280	Peter K Todd
National Institute of Neurological Disorders and Stroke	R01NS086810	Peter K Todd
Veterans Administration Medical Center	BX004842	Peter K Todd

The funders had no role in study design, data collection and interpretation, or the decision to submit the work for publication.

Author contributions

Shannon E Wright, Conceptualization, Investigation, Writing - original draft, Writing – review and editing; Peter K Todd, Conceptualization, Supervision, Investigation, Writing – review and editing

Author ORCIDs

Shannon E Wright  <http://orcid.org/0000-0002-7388-1617>

Peter K Todd  <http://orcid.org/0000-0003-4781-6376>

References

- Abruzzi KC**, Lacadie S, Rosbash M. 2004. Biochemical analysis of TREX complex recruitment to intronless and intron-containing yeast genes. *The EMBO Journal* **23**:2620–2631. DOI: <https://doi.org/10.1038/sj.emboj.7600261>, PMID: 15192704
- Allingham-Hawkins DJ**, Babul-Hirji R, Chitayat D, Holden JJ, Yang KT, Lee C, Hudson R, Gorwill H, Nolin SL, Glicksman A, Jenkins EC, Brown WT, Howard-Peebles PN, Becchi C, Cummings E, Fallon L, Seitz S, Black SH, Vianna-Morgante AM, Costa SS, et al. 1999. Fragile X premutation is a significant risk factor for premature ovarian failure: the International collaborative POF in fragile X study -- preliminary data. *American Journal of Medical Genetics* **83**:322–325 PMID: 10208170.
- Andken BB**, Lim I, Benson G, Vincent JJ, Ferenc MT, Heinrich B, Jarzylo LA, Man HY, Deshler JO. 2007. 3'-Utr SIRF: a database for identifying clusters of whort interspersed repeats in 3' untranslated regions. *BMC Bioinformatics* **8**:274. DOI: <https://doi.org/10.1186/1471-2105-8-274>, PMID: 17663765
- Andrade MA**, Perez-Iratxeta C, Ponting CP. 2001. Protein repeats: structures, functions, and evolution. *Journal of Structural Biology* **134**:117–131. DOI: <https://doi.org/10.1006/jsbi.2001.4392>, PMID: 11551174
- Annear DJ**, Vandeweyer G, Sanchis-Juan A, Raymond FL, Kooy RF. 2022. Non-Mendelian inheritance patterns and extreme deviation rates of CGG repeats in autism. *Genome Research* **32**:1967–1980. DOI: <https://doi.org/10.1101/gr.277011.122>, PMID: 36351771
- Ash PEA**, Bieniek KF, Gendron TF, Caulfield T, Lin WL, Dejesus-Hernandez M, van Blitterswijk MM, Jansen-West K, Paul JW, Rademakers R, Boylan KB, Dickson DW, Petrucelli L. 2013. Unconventional translation of c9orf72 GGGGCC expansion generates insoluble polypeptides specific to c9ftd/ALS. *Neuron* **77**:639–646. DOI: <https://doi.org/10.1016/j.neuron.2013.02.004>, PMID: 23415312
- Ashkenazi A**, Bento CF, Ricketts T, Vicinanza M, Siddiqi F, Pavel M, Squitieri F, Hardenberg MC, Imarisio S, Menzies FM, Rubinsztein DC. 2017. Polyglutamine tracts regulate Beclin 1-dependent autophagy. *Nature* **545**:108–111. DOI: <https://doi.org/10.1038/nature22078>, PMID: 28445460
- Aslanidis C**, Jansen G, Amemiya C, Shutler G, Mahadevan M, Tsilfidis C, Chen C, Alleman J, Wormskamp NGM, Vooijs M. 1992. Cloning of the essential myotonic dystrophy region and mapping of the putative defect. *Nature* **355**:548–551. DOI: <https://doi.org/10.1038/355548a0>, PMID: 1346925
- Bañez-Coronel M**, Ayhan F, Tarabochia AD, Zu T, Perez BA, Tusi SK, Pletnikova O, Borchelt DR, Ross CA, Margolis RL, Yachnis AT, Troncoso JC, Ranum LPW. 2015. RAN translation in huntington disease. *Neuron* **88**:667–677. DOI: <https://doi.org/10.1016/j.neuron.2015.10.038>, PMID: 26590344
- Bäuerlein FJB**, Saha I, Mishra A, Kalemánov M, Martínez-Sánchez A, Klein R, Dudanova I, Hipp MS, Hartl FU, Baumeister W, Fernández-Busnadiego R. 2017. In situ architecture and cellular interactions of polyQ inclusions. *Cell* **171**:179–187.. DOI: <https://doi.org/10.1016/j.cell.2017.08.009>, PMID: 28890085

- Beilina A**, Tassone F, Schwartz PH, Sahota P, Hagerman PJ. 2004. Redistribution of transcription start sites within the FMR1 promoter region with expansion of the downstream CGG-repeat element. *Human Molecular Genetics* **13**:543–549. DOI: <https://doi.org/10.1093/hmg/ddh053>, PMID: 14722156
- Belotserkovskii BP**, Liu R, Tornaletti S, Krasilnikova MM, Mirkin SM, Hanawalt PC. 2010. Mechanisms and implications of transcription blockage by guanine-rich DNA sequences. *PNAS* **107**:12816–12821. DOI: <https://doi.org/10.1073/pnas.1007580107>, PMID: 20616059
- Belzil VV**, Bauer PO, Prudencio M, Gendron TF, Stetler CT, Yan IK, Pregent L, Daugherty L, Baker MC, Rademakers R, Boylan K, Patel TC, Dickson DW, Petrucelli L. 2013. Reduced C9orf72 gene expression in c9FTD/ALS is caused by histone trimethylation, an epigenetic event detectable in blood. *Acta Neuropathologica* **126**:895–905. DOI: <https://doi.org/10.1007/s00401-013-1199-1>, PMID: 24166615
- Bentele K**, Saffert P, Rauscher R, Ignatova Z, Blüthgen N. 2013. Efficient translation initiation dictates codon usage at gene start. *Molecular Systems Biology* **9**:675. DOI: <https://doi.org/10.1038/msb.2013.32>, PMID: 23774758
- Betley JN**, Frith MC, Graber JH, Choo S, Deshler JO. 2002. A ubiquitous and conserved signal for RNA localization in chordates. *Current Biology* **12**:1756–1761. DOI: [https://doi.org/10.1016/s0960-9822\(02\)01220-4](https://doi.org/10.1016/s0960-9822(02)01220-4), PMID: 12401170
- Black DL**. 2003. Mechanisms of alternative pre-messenger RNA splicing. *Annual Review of Biochemistry* **72**:291–336. DOI: <https://doi.org/10.1146/annurev.biochem.72.121801.161720>, PMID: 12626338
- Blackledge NP**, Thomson JP, Skene PJ. 2013. CpG island chromatin is shaped by recruitment of ZF-cxxc proteins. *Cold Spring Harbor Perspectives in Biology* **5**:a018648. DOI: <https://doi.org/10.1101/cshperspect.a018648>, PMID: 24186071
- Boivin M**, Deng J, Pfister V, Grandgirard E, Oulad-Abdelghani M, Morlet B, Ruffenach F, Negroni L, Koebel P, Jacob H, Riet F, Dijkstra AA, McFadden K, Clayton WA, Hong D, Miyahara H, Iwasaki Y, Sone J, Wang Z, Charlet-Berguerand N. 2021. Translation of GGC repeat expansions into a toxic polyglycine protein in NIID defines a novel class of human genetic disorders: the polyG diseases. *Neuron* **109**:1825–1835. DOI: <https://doi.org/10.1016/j.neuron.2021.03.038>, PMID: 33887199
- Bolton KA**, Ross JP, Grice DM, Bowden NA, Holliday EG, Avery-Kiejda KA, Scott RJ. 2013. STARRRT: a table of short tandem repeats in regulatory regions of the human genome. *BMC Genomics* **14**:1–16. DOI: <https://doi.org/10.1186/1471-2164-14-795>, PMID: 24228761
- Botta A**, Rinaldi F, Catalli C, Vergani L, Bonifazi E, Romeo V, Loro E, Viola A, Angelini C, Novelli G. 2008. The CTG repeat expansion size correlates with the splicing defects observed in muscles from myotonic dystrophy type 1 patients. *Journal of Medical Genetics* **45**:639–646. DOI: <https://doi.org/10.1136/jmg.2008.058909>, PMID: 18611984
- Bourque G**. 2009. Transposable elements in gene regulation and in the evolution of vertebrate genomes. *Current Opinion in Genetics & Development* **19**:607–612. DOI: <https://doi.org/10.1016/j.gde.2009.10.013>, PMID: 19914058
- Britten RJ**, Kohne DE. 1968. Repeated sequences in DNA: hundreds of thousands of copies of DNA sequences have been incorporated into the genomes of higher organisms. *Science* **161**:529–540. DOI: <https://doi.org/10.1126/science.161.3841.529>, PMID: 4874239
- Brook JD**, McCurrach ME, Harley HG, Buckler AJ, Church D, Aburatani H, Hunter K, Stanton VP, Thirion JP, Hudson T. 1992. Molecular basis of myotonic dystrophy: expansion of a trinucleotide (CTG) repeat at the 3' end of a transcript encoding a protein kinase family member. *Cell* **68**:799–808. DOI: [https://doi.org/10.1016/0092-8674\(92\)90154-5](https://doi.org/10.1016/0092-8674(92)90154-5), PMID: 1310900
- Brouwer JR**, Huizer K, Severijnen LA, Hukema RK, Berman RF, Oostra BA, Willemsen R. 2008. Cgg-Repeat length and neuropathological and molecular correlates in a mouse model for fragile X-associated tremor/ataxia syndrome. *Journal of Neurochemistry* **107**:1671–1682. DOI: <https://doi.org/10.1111/j.1471-4159.2008.05747.x>, PMID: 19014369
- Brouwer JR**, Huguet A, Nicole A, Munnich A, Gourdon G. 2013. Transcriptionally repressive chromatin remodelling and CpG methylation in the presence of expanded CTG-repeats at the DM1 locus. *Journal of Nucleic Acids* **2013**:1–16. DOI: <https://doi.org/10.1155/2013/567435>
- Brunberg JA**, Jacquemont S, Hagerman RJ, Berry-Kravis EM, Grigsby J, Leehey MA, Tassone F, Brown WT, Greco CM, Hagerman PJ. 2002. Fragile X premutation carriers: characteristic MR imaging findings of adult male patients with progressive cerebellar and cognitive dysfunction. *AJNR. American Journal of Neuroradiology* **23**:1757–1766. PMID: 12427636.
- Buijsen RAM**, Sellier C, Severijnen L, Oulad-Abdelghani M, Verhagen RFM, Berman RF, Charlet-Berguerand N, Willemsen R, Hukema RK. 2014. FMRpolyG-positive inclusions in CNS and non-CNS organs of a fragile X premutation carrier with fragile X-associated tremor/ataxia syndrome. *Acta Neuropathologica Communications* **2**:162. DOI: <https://doi.org/10.1186/s40478-014-0162-2>, PMID: 25471011
- Buxton J**, Shelbourne P, Davies J, Jones C, Van Tongeren T, Aslanidis C, de Jong P, Jansen G, Anvret M, Riley B. 1992. Detection of an unstable fragment of DNA specific to individuals with myotonic dystrophy. *Nature* **355**:547–548. DOI: <https://doi.org/10.1038/355547a0>, PMID: 1346924
- Caron NS**, Desmond CR, Xia J, Truant R. 2013. Polyglutamine domain flexibility mediates the proximity between flanking sequences in huntingtin. *PNAS* **110**:14610–14615. DOI: <https://doi.org/10.1073/pnas.1301342110>, PMID: 23898200
- Cerro-Herreros E**, Chakraborty M, Pérez-Alonso M, Artero R, Llamusi B. 2017. Expanded CCUG repeat RNA expression in *Drosophila* heart and muscle trigger myotonic dystrophy type 1-like phenotypes and activate

- autophagocytosis genes. *Scientific Reports* **7**:2843. DOI: <https://doi.org/10.1038/s41598-017-02829-3>, PMID: 28588248
- Chakraborty A**, Jenjaroenpun P, Li J, El Hilali S, McCulley A, Haarer B, Hoffman EA, Belak A, Thorland A, Hehnlly H, Schildkraut CL, Chen CL, Kuznetsov VA, Feng W. 2020. Replication stress induces global chromosome breakage in the fragile X genome. *Cell Reports* **32**:108179. DOI: <https://doi.org/10.1016/j.celrep.2020.108179>, PMID: 32966779
- Chávez S**, Beilharz T, Rondón AG, Erdjument-Bromage H, Tempst P, Svejstrup JQ, Lithgow T, Aguilera A. 2000. A protein complex containing tho2, hpr1, mft1 and a novel protein, thp2, connects transcription elongation with mitotic recombination in *Saccharomyces cerevisiae*. *The EMBO Journal* **19**:5824–5834. DOI: <https://doi.org/10.1093/emboj/19.21.5824>, PMID: 11060033
- Chávez S**, García-Rubio M, Prado F, Aguilera A. 2001. Hpr1 is preferentially required for transcription of either long or G+C-rich DNA sequences in *Saccharomyces cerevisiae*. *Molecular and Cellular Biology* **21**:7054–7064. DOI: <https://doi.org/10.1128/MCB.21.20.7054-7064.2001>, PMID: 11564888
- Choong CS**, Wilson EM. 1998. Trinucleotide repeats in the human androgen receptor: a molecular basis for disease. *Journal of Molecular Endocrinology* **21**:235–257. DOI: <https://doi.org/10.1677/jme.0.0210235>, PMID: 9845666
- Chudley AE**, Hagerman RJ. 1987. Fragile X syndrome. *The Journal of Pediatrics* **110**:821–831. DOI: [https://doi.org/10.1016/s0022-3476\(87\)80392-x](https://doi.org/10.1016/s0022-3476(87)80392-x), PMID: 3295158
- Ciesiolka A**, Jazurek M, Drazkowska K, Krzyzosiak WJ. 2017. Structural characteristics of simple RNA repeats associated with disease and their deleterious protein interactions. *Frontiers in Cellular Neuroscience* **11**:97. DOI: <https://doi.org/10.3389/fncel.2017.00097>, PMID: 28442996
- Cleary JD**, Ranum LP. 2017. New developments in Ran translation: insights from multiple diseases. *Current Opinion in Genetics & Development* **44**:125–134. DOI: <https://doi.org/10.1016/j.gde.2017.03.006>, PMID: 28365506
- Clouaire T**, Webb S, Skene P, Illingworth R, Kerr A, Andrews R, Lee JH, Skalnik D, Bird A. 2012. Cfp1 integrates both CpG content and gene activity for accurate H3K4me3 deposition in embryonic stem cells. *Genes & Development* **26**:1714–1728. DOI: <https://doi.org/10.1101/gad.194209.112>, PMID: 22855832
- Coffee B**, Zhang F, Ceman S, Warren ST, Reines D. 2002. Histone modifications depict an aberrantly heterochromatinized FMR1 gene in fragile X syndrome. *American Journal of Human Genetics* **71**:923–932. DOI: <https://doi.org/10.1086/342931>, PMID: 12232854
- Colak D**, Zaninovic N, Cohen MS, Rosenwaks Z, Yang WY, Gerhardt J, Disney MD, Jaffrey SR. 2014. Promoter-Bound trinucleotide repeat mRNA drives epigenetic silencing in fragile X syndrome. *Science* **343**:1002–1005. DOI: <https://doi.org/10.1126/science.1245831>, PMID: 24578575
- Comfort NC**. 2001. From controlling elements to transposons: Barbara McClintock and the Nobel Prize. *Trends in Genetics* **17**:475–478. DOI: [https://doi.org/10.1016/s0168-9525\(01\)02383-6](https://doi.org/10.1016/s0168-9525(01)02383-6), PMID: 11485821
- Cortese A**, Simone R, Sullivan R, Vandrovцова J, Tariq H, Yau WY, Humphrey J, Jaunmuktane Z, Sivakumar P, Polke J, Ilyas M, Tribollet E, Tomaselli PJ, Devigili G, Callegari I, Versino M, Salpietro V, Efthymiou S, Kaski D, Wood NW, et al. 2019. Biallelic expansion of an intronic repeat in RFC1 is a common cause of late-onset ataxia. *Nature Genetics* **51**:649–658. DOI: <https://doi.org/10.1038/s41588-019-0372-4>, PMID: 30926972
- Cournac A**, Koszul R, Mozziconacci J. 2016. The 3D folding of metazoan genomes correlates with the association of similar repetitive elements. *Nucleic Acids Research* **44**:245–255. DOI: <https://doi.org/10.1093/nar/gkv1292>, PMID: 26609133
- Davies JE**, Rubinsztein DC. 2006. Polyalanine and polyserine frameshift products in Huntington's disease. *Journal of Medical Genetics* **43**:893–896. DOI: <https://doi.org/10.1136/jmg.2006.044222>, PMID: 16801344
- Davis BM**, McCurrach ME, Taneja KL, Singer RH, Housman DE. 1997. Expansion of a CUG trinucleotide repeat in the 3' untranslated region of myotonic dystrophy protein kinase transcripts results in nuclear retention of transcripts. *PNAS* **94**:7388–7393. DOI: <https://doi.org/10.1073/pnas.94.14.7388>, PMID: 9207101
- Deaton AM**, Bird A. 2011. CpG islands and the regulation of transcription. *Genes & Development* **25**:1010–1022. DOI: <https://doi.org/10.1101/gad.2037511>, PMID: 21576262
- Debacker K**, Kooy RF. 2007. Fragile sites and human disease. *Human Molecular Genetics* **16 Spec No. 2**:R150–R158. DOI: <https://doi.org/10.1093/hmg/ddm136>, PMID: 17567780
- Deininger P**. 2011. Alu elements: know the SINEs. *Genome Biology* **12**:1–12. DOI: <https://doi.org/10.1186/gb-2011-12-12-236>, PMID: 22204421
- de Koning APJ**, Gu W, Castoe TA, Batzer MA, Pollock DD. 2011. Repetitive elements may comprise over two-thirds of the human genome. *PLOS Genetics* **7**:e1002384. DOI: <https://doi.org/10.1371/journal.pgen.1002384>, PMID: 22144907
- Deng J**, Yu J, Li P, Luan X, Cao L, Zhao J, Yu M, Zhang W, Lv H, Xie Z, Meng L, Zheng Y, Zhao Y, Gang Q, Wang Q, Liu J, Zhu M, Guo X, Su Y, Liang Y, et al. 2020. Expansion of GGC repeat in GIPC1 is associated with oculopharyngodistal myopathy. *American Journal of Human Genetics* **106**:793–804. DOI: <https://doi.org/10.1016/j.ajhg.2020.04.011>, PMID: 32413282
- Derbis M**, Konieczny P, Walczak A, Sekrecki M, Sobczak K. 2018. Quantitative evaluation of toxic polyglycine biosynthesis and aggregation in cell models expressing expanded CGG repeats. *Frontiers in Genetics* **9**:216. DOI: <https://doi.org/10.3389/fgene.2018.00216>, PMID: 29971092
- Ding Y**, Shah P, Plotkin JB. 2012. Weak 5'-mRNA secondary structures in short eukaryotic genes. *Genome Biology and Evolution* **4**:1046–1053. DOI: <https://doi.org/10.1093/gbe/evs082>, PMID: 23034215
- Dion V**, Wilson JH. 2009. Instability and chromatin structure of expanded trinucleotide repeats. *Trends in Genetics* **25**:288–297. DOI: <https://doi.org/10.1016/j.tig.2009.04.007>, PMID: 19540013

- Disney MD**, Liu B, Yang WY, Sellier C, Tran T, Charlet-Berguerand N, Childs-Disney JL. 2012. A small molecule that targets R (CGG) (EXP) and improves defects in fragile X-associated tremor ataxia syndrome. *ACS Chemical Biology* **7**:1711–1718. DOI: <https://doi.org/10.1021/cb300135h>, PMID: 22948243
- Du H**, Cline MS, Osborne RJ, Tuttle DL, Clark TA, Donohue JP, Hall MP, Shiue L, Swanson MS, Thornton CA, Ares M Jr. 2010. Aberrant alternative splicing and extracellular matrix gene expression in mouse models of myotonic dystrophy. *Nature Structural & Molecular Biology* **17**:187–193. DOI: <https://doi.org/10.1038/nsmb.1720>, PMID: 20098426
- Duvick L**, Barnes J, Ebner B, Agrawal S, Andresen M, Lim J, Giesler GJ, Zoghbi HY, Orr HT. 2010. SCA1-like disease in mice expressing wild-type ataxin-1 with a serine to aspartic acid replacement at residue 776. *Neuron* **67**:929–935. DOI: <https://doi.org/10.1016/j.neuron.2010.08.022>, PMID: 20869591
- Eichler EE**, Kunst CB, Lugenbeel KA, Ryder OA, Davison D, Warren ST, Nelson DL. 1995. Evolution of the cryptic FMR1 CGG repeat. *Nature Genetics* **11**:301–308. DOI: <https://doi.org/10.1038/ng1195-301>, PMID: 7581454
- Elden AC**, Kim H-J, Hart MP, Chen-Plotkin AS, Johnson BS, Fang X, Armakola M, Geser F, Greene R, Lu MM, Padmanabhan A, Clay-Falcone D, McCluskey L, Elman L, Juhr D, Gruber PJ, Rüb U, Auburger G, Trojanowski JQ, Lee VM-Y, et al. 2010. Ataxin-2 intermediate-length polyglutamine expansions are associated with increased risk for ALS. *Nature* **466**:1069–1075. DOI: <https://doi.org/10.1038/nature09320>, PMID: 20740007
- Elizur SE**, Dratviman-Storobinsky O, Derech-Haim S, Lebovitz O, Dor J, Orvieto R, Cohen Y. 2016. FMR6 may play a role in the pathogenesis of fragile X-associated premature ovarian insufficiency. *Gynecological Endocrinology* **32**:334–337. DOI: <https://doi.org/10.3109/09513590.2015.1116508>, PMID: 26652047
- Emamian ES**, Kaytor MD, Duvick LA, Zu T, Tousey SK, Zoghbi HY, Clark HB, Orr HT. 2003. Serine 776 of ataxin-1 is critical for polyglutamine-induced disease in SCA1 transgenic mice. *Neuron* **38**:375–387. DOI: [https://doi.org/10.1016/s0896-6273\(03\)00258-7](https://doi.org/10.1016/s0896-6273(03)00258-7), PMID: 12741986
- Entezam A**, Biacsi R, Orrison B, Saha T, Hoffman GE, Grabczyk E, Nussbaum RL, Usdin K. 2007. Regional FMRP deficits and large repeat expansions into the full mutation range in a new fragile X premutation mouse model. *Gene* **395**:125–134. DOI: <https://doi.org/10.1016/j.gene.2007.02.026>, PMID: 17442505
- Fan HY**, Cheng KK, Klein HL. 1996. Mutations in the RNA polymerase II transcription machinery suppress the hyperrecombination mutant hpr1 delta of *Saccharomyces cerevisiae*. *Genetics* **142**:749–759. DOI: <https://doi.org/10.1093/genetics/142.3.749>, PMID: 8849885
- Fan HY**, Merker RJ, Klein HL. 2001. High-Copy-Number expression of Sub2p, a member of the RNA helicase superfamily, suppresses hpr1-mediated genomic instability. *Molecular and Cellular Biology* **21**:5459–5470. DOI: <https://doi.org/10.1128/MCB.21.16.5459-5470.2001>, PMID: 11463828
- Fardaei M**, Larkin K, Brook JD, Hamshire MG. 2001. In vivo co-localisation of MBNL protein with DMPK expanded-repeat transcripts. *Nucleic Acids Research* **29**:2766–2771. DOI: <https://doi.org/10.1093/nar/29.13.2766>, PMID: 11433021
- Faux N**. 2012. Single amino acid and trinucleotide repeats: function and evolution. *Advances in Experimental Medicine and Biology* **769**:26–40. DOI: https://doi.org/10.1007/978-1-4614-5434-2_3, PMID: 23560303
- Fay MM**, Anderson PJ, Ivanov P. 2017. ALS/FTD-associated C9orf72 repeat RNA promotes phase transitions in vitro and in cells. *Cell Reports* **21**:3573–3584. DOI: <https://doi.org/10.1016/j.celrep.2017.11.093>, PMID: 29262335
- Fondon JW**, Garner HR. 2004. Molecular origins of rapid and continuous morphological evolution. *PNAS* **101**:18058–18063. DOI: <https://doi.org/10.1073/pnas.0408118101>, PMID: 15596718
- Fotsing SF**, Margoliash J, Wang C, Saini S, Yanicky R, Shleizer-Burko S, Goren A, Gymrek M. 2019. The impact of short tandem repeat variation on gene expression. *Nature Genetics* **51**:1652–1659. DOI: <https://doi.org/10.1038/s41588-019-0521-9>, PMID: 31676866
- Franich NR**, Hickey MA, Zhu C, Osborne GF, Ali N, Chu T, Bove NH, Lemesre V, Lerner RP, Zeitlin SO, Howland D, Neueder A, Landles C, Bates GP, Chesselet M-F. 2019. Phenotype onset in Huntington's disease knock-in mice is correlated with the incomplete splicing of the mutant huntingtin gene. *Journal of Neuroscience Research* **97**:1590–1605. DOI: <https://doi.org/10.1002/jnr.24493>, PMID: 31282030
- Freibaum BD**, Lu Y, Lopez-Gonzalez R, Kim NC, Almeida S, Lee K-H, Badders N, Valentine M, Miller BL, Wong PC, Petrucelli L, Kim HJ, Gao F-B, Taylor JP. 2015. GGGGCC repeat expansion in c9orf72 compromises nucleocytoplasmic transport. *Nature* **525**:129–133. DOI: <https://doi.org/10.1038/nature14974>, PMID: 26308899
- Fry M**, Loeb LA. 1994. The fragile X syndrome D (CGG) N nucleotide repeats form a stable tetrahelical structure. *PNAS* **91**:4950–4954. DOI: <https://doi.org/10.1073/pnas.91.11.4950>, PMID: 8197163
- Fu YH**, Kuhl DP, Pizzuti A, Pieretti M, Sutcliffe JS, Richards S, Verkerk AJ, Holden JJ, Fenwick RG, Warren ST. 1991. Variation of the CGG repeat at the fragile X site results in genetic instability: resolution of the sherman paradox. *Cell* **67**:1047–1058. DOI: [https://doi.org/10.1016/0092-8674\(91\)90283-5](https://doi.org/10.1016/0092-8674(91)90283-5), PMID: 1760838
- Fu YH**, Pizzuti A, Fenwick RG Jr, King J, Rajnarayan S, Dunne PW, Dubel J, Nasser GA, Ashizawa T, de Jong P. 1992. An unstable triplet repeat in a gene related to myotonic muscular dystrophy. *Science* **255**:1256–1258. DOI: <https://doi.org/10.1126/science.1546326>, PMID: 1546326
- Gao FB**, Richter JD, Cleveland DW. 2017. Rethinking unconventional translation in neurodegeneration. *Cell* **171**:994–1000. DOI: <https://doi.org/10.1016/j.cell.2017.10.042>, PMID: 29149615
- Garg P**, Martin-Trujillo A, Rodriguez OL, Gies SJ, Hadelia E, Jadhav B, Jain M, Paten B, Sharp AJ. 2021. Pervasive cis effects of variation in copy number of large tandem repeats on local DNA methylation and gene expression. *American Journal of Human Genetics* **108**:809–824. DOI: <https://doi.org/10.1016/j.ajhg.2021.03.016>, PMID: 33794196

- Gaspar C**, Jannatipour M, Dion P, Laganière J, Sequeiros J, Brais B, Rouleau GA. 2000. CAG tract of MJD-1 may be prone to frameshifts causing polyalanine accumulation. *Human Molecular Genetics* **9**:1957–1966. DOI: <https://doi.org/10.1093/hmg/9.13.1957>, PMID: 10942424
- Gasset-Rosa F**, Chillon-Marinas C, Goginashvili A, Atwal RS, Artates JW, Tabet R, Wheeler VC, Bang AG, Cleveland DW, Lagier-Tourenne C. 2017. Polyglutamine-expanded huntingtin exacerbates age-related disruption of nuclear integrity and nucleocytoplasmic transport. *Neuron* **94**:48–57. DOI: <https://doi.org/10.1016/j.neuron.2017.03.027>, PMID: 28384474
- Gecz J**, Gedeon AK, Sutherland GR, Mulley JC. 1996. Identification of the gene FMR2, associated with FRAXE mental retardation. *Nature Genetics* **13**:105–108. DOI: <https://doi.org/10.1038/ng0596-105>, PMID: 8673085
- Gecz J**. 2000. The FMR2 gene, FRAXE and non-specific X-linked mental retardation: clinical and molecular aspects. *Annals of Human Genetics* **64**:95–106. DOI: <https://doi.org/10.1017/S0003480000007983>, PMID: 11246464
- Gemmell NJ**. 2021. Repetitive DNA: genomic dark matter matters. *Nature Reviews. Genetics* **22**:342. DOI: <https://doi.org/10.1038/s41576-021-00354-8>, PMID: 33782602
- Gendron TF**, Bieniek KF, Zhang Y-J, Jansen-West K, Ash PEA, Caulfield T, Daugherty L, Dunmore JH, Castanedes-Casey M, Chew J, Cosio DM, van Blitterswijk M, Lee WC, Rademakers R, Boylan KB, Dickson DW, Petrucelli L. 2013. Antisense transcripts of the expanded C9orf72 hexanucleotide repeat form nuclear RNA foci and undergo repeat-associated non-ATG translation in c9ftd/ALS. *Acta Neuropathologica* **126**:829–844. DOI: <https://doi.org/10.1007/s00401-013-1192-8>, PMID: 24129584
- Gerhardt J**, Tomishima MJ, Zaninovic N, Colak D, Yan Z, Zhan Q, Rosenwaks Z, Jaffrey SR, Schildkraut CL. 2014. The DNA replication program is altered at the FMR1 locus in fragile X embryonic stem cells. *Molecular Cell* **53**:19–31. DOI: <https://doi.org/10.1016/j.molcel.2013.10.029>, PMID: 24289922
- Gipson TA**, Neueder A, Wexler NS, Bates GP, Housman D. 2013. Aberrantly spliced HTT, a new player in Huntington's disease pathogenesis. *RNA Biology* **10**:1647–1652. DOI: <https://doi.org/10.4161/rna.26706>, PMID: 24256709
- Glineburg MR**, Todd PK, Charlet-Berguerand N, Sellier C. 2018. Repeat-Associated non-AUG (Ran) translation and other molecular mechanisms in fragile X tremor ataxia syndrome. *Brain Research* **1693**:43–54. DOI: <https://doi.org/10.1016/j.brainres.2018.02.006>, PMID: 29453961
- Glover TW**. 2006. Common fragile sites. *Cancer Letters* **232**:4–12. DOI: <https://doi.org/10.1016/j.canlet.2005.08.032>, PMID: 16229941
- Gohel D**, Sripada L, Prajapati P, Singh K, Roy M, Kotadia D, Tassone F, Charlet-Berguerand N, Singh R. 2019. FMRpolyG alters mitochondrial transcripts level and respiratory chain complex assembly in fragile X associated tremor/ataxia syndrome. *Biochimica et Biophysica Acta. Molecular Basis of Disease* **1865**:1379–1388. DOI: <https://doi.org/10.1016/j.bbadis.2019.02.010>, PMID: 30771487
- Goodwin M**, Mohan A, Batra R, Lee K-Y, Charizanis K, Fernández Gómez FJ, Eddarkaoui S, Sergeant N, Buée L, Kimura T, Clark HB, Dalton J, Takamura K, Weyn-Vanhenenryck SM, Zhang C, Reid T, Ranum LPW, Day JW, Swanson MS. 2015. Mbnl sequestration by toxic RNAs and RNA misprocessing in the myotonic dystrophy brain. *Cell Reports* **12**:1159–1168. DOI: <https://doi.org/10.1016/j.celrep.2015.07.029>, PMID: 26257173
- Grabczyk E**, Fishman MC. 1995. A long purine-pyrimidine homopolymer acts as a transcriptional diode. *The Journal of Biological Chemistry* **270**:1791–1797. DOI: <https://doi.org/10.1074/jbc.270.4.1791>, PMID: 7829515
- Greene E**, Mahishi L, Entezam A, Kumari D, Usdin K. 2007. Repeat-Induced epigenetic changes in intron 1 of the frataxin gene and its consequences in Friedreich ataxia. *Nucleic Acids Research* **35**:3383–3390. DOI: <https://doi.org/10.1093/nar/gkm271>, PMID: 17478498
- Grima JC**, Daigle JG, Arbez N, Cunningham KC, Zhang K, Ochaba J, Geater C, Morozko E, Stocksedale J, Glatzer JC, Pham JT, Ahmed I, Peng Q, Wadhwa H, Pletnikova O, Troncoso JC, Duan W, Snyder SH, Ranum LPW, Thompson LM, et al. 2017. Mutant huntingtin disrupts the nuclear pore complex. *Neuron* **94**:93–107. DOI: <https://doi.org/10.1016/j.neuron.2017.03.023>, PMID: 28384479
- Groh M**, Silva LM, Gromak N. 2014. Mechanisms of transcriptional dysregulation in repeat expansion disorders. *Biochemical Society Transactions* **42**:1123–1128. DOI: <https://doi.org/10.1042/BST20140049>, PMID: 25110013
- Gu Y**, Shen Y, Gibbs RA, Nelson DL. 1996. Identification of FMR2, a novel gene associated with the FRAXE CCG repeat and CpG island. *Nature Genetics* **13**:109–113. DOI: <https://doi.org/10.1038/ng0596-109>, PMID: 8673086
- Gymrek M**, Willems T, Guilmatre A, Zeng H, Markus B, Georgiev S, Daly MJ, Price AL, Pritchard JK, Sharp AJ, Erlich Y. 2016. Abundant contribution of short tandem repeats to gene expression variation in humans. *Nature Genetics* **48**:22–29. DOI: <https://doi.org/10.1038/ng.3461>, PMID: 26642241
- Gymrek M**. 2017. A genomic view of short tandem repeats. *Current Opinion in Genetics & Development* **44**:9–16. DOI: <https://doi.org/10.1016/j.gde.2017.01.012>, PMID: 28213161
- Gymrek M**, Willems T, Reich D, Erlich Y. 2017. Interpreting short tandem repeat variations in humans using mutational constraint. *Nature Genetics* **49**:1495–1501. DOI: <https://doi.org/10.1038/ng.3952>, PMID: 28892063
- Hagerman RJ**, Jackson AW, Levitas A, Rimland B, Braden M. 1986. An analysis of autism in fifty males with the fragile X syndrome. *American Journal of Medical Genetics* **23**:359–374. DOI: <https://doi.org/10.1002/ajmg.1320230128>, PMID: 3953654
- Hagerman RJ**, Leehey M, Heinrichs W, Tassone F, Wilson R, Hills J, Grigsby J, Gage B, Hagerman PJ. 2001. Intention tremor, parkinsonism, and generalized brain atrophy in male carriers of fragile X. *Neurology* **57**:127–130. DOI: <https://doi.org/10.1212/wnl.57.1.127>, PMID: 11445641
- Hagerman RJ**, Hagerman PJ. 2002a. Fragile X Syndrome: Diagnosis, Treatment, and Research. United States: Taylor & Francis. DOI: <https://doi.org/10.56021/9780801868436>

- Hagerman RJ, Hagerman PJ. 2002b. The fragile X premutation: into the phenotypic fold. *Current Opinion in Genetics & Development* **12**:278–283. DOI: [https://doi.org/10.1016/s0959-437x\(02\)00299-x](https://doi.org/10.1016/s0959-437x(02)00299-x), PMID: 12076670
- Hagerman PJ, Hagerman RJ. 2004. Fragile X-associated Tremor/Ataxia Syndrome (FXTAS). *Mental Retardation and Developmental Disabilities Research Reviews* **10**:25–30. DOI: <https://doi.org/10.1002/mrdd.20005>, PMID: 14994285
- Hagerman PJ, Hagerman RJ. 2015. Fragile X-associated tremor/ataxia syndrome. *Annals of the New York Academy of Sciences* **1338**:58–70. DOI: <https://doi.org/10.1111/nyas.12693>, PMID: 25622649
- Hagerman RJ, Berry-Kravis E, Hazlett HC, Bailey DB, Moine H, Kooy RF, Tassone F, Gantois I, Sonenberg N, Mandel JL, Hagerman PJ. 2017. Fragile X syndrome. *Nature Reviews. Disease Primers* **3**:17065. DOI: <https://doi.org/10.1038/nrdp.2017.65>, PMID: 28960184
- Hammock EAD, Young LJ. 2005. Microsatellite instability generates diversity in brain and sociobehavioral traits. *Science* **308**:1630–1634. DOI: <https://doi.org/10.1126/science.1111427>, PMID: 15947188
- Hannan AJ. 2018. Tandem repeats mediating genetic plasticity in health and disease. *Nature Reviews. Genetics* **19**:286–298. DOI: <https://doi.org/10.1038/nrg.2017.115>, PMID: 29398703
- Hansen RS, Gartler SM, Scott CR, Chen SH, Laird CD. 1992. Methylation analysis of CGG sites in the CpG island of the human FMR1 gene. *Human Molecular Genetics* **1**:571–578. DOI: <https://doi.org/10.1093/hmg/1.8.571>, PMID: 1301165
- Harley HG, Brook JD, Rundle SA, Crow S, Reardon W, Buckler AJ, Harper PS, Housman DE, Shaw DJ. 1992. Expansion of an unstable DNA region and phenotypic variation in myotonic dystrophy. *Nature* **355**:545–546. DOI: <https://doi.org/10.1038/355545a0>, PMID: 1346923
- Hautbergue GM, Castelli LM, Ferraiuolo L, Sanchez-Martinez A, Cooper-Knock J, Higginbottom A, Lin YH, Bauer CS, Dodd JE, Myszczyńska MA, Alam SM, Garneret P, Chandran JS, Karyka E, Stopford MJ, Smith EF, Kirby J, Meyer K, Kaspar BK, Isaacs AM, et al. 2017. SRSF1-dependent nuclear export inhibition of c9orf72 repeat transcripts prevents neurodegeneration and associated motor deficits. *Nature Communications* **8**:16063. DOI: <https://doi.org/10.1038/ncomms16063>, PMID: 28677678
- He F, Krans A, Freibaum BD, Taylor JP, Todd PK. 2014. Tdp-43 suppresses CGG repeat-induced neurotoxicity through interactions with hnRNP A2/B1. *Human Molecular Genetics* **23**:5036–5051. DOI: <https://doi.org/10.1093/hmg/ddu216>, PMID: 24920338
- Heitz D, Rousseau F, Devys D, Saccone S, Abderrahim H, Le Paslier D, Cohen D, Vincent A, Toniolo D, Della Valle G. 1991. Isolation of sequences that span the fragile X and identification of a fragile X-related cpg island. *Science* **251**:1236–1239. DOI: <https://doi.org/10.1126/science.2006411>, PMID: 2006411
- Herman D, Jenssen K, Burnett R, Soragni E, Perlman SL, Gottesfeld JM. 2006. Histone deacetylase inhibitors reverse gene silencing in Friedreich's ataxia. *Nature Chemical Biology* **2**:551–558. DOI: <https://doi.org/10.1038/nchembio815>, PMID: 16921367
- Hinnebusch AG, Ivanov IP, Sonenberg N. 2016. Translational control by 5'-untranslated regions of eukaryotic mRNAs. *Science* **352**:1413–1416. DOI: <https://doi.org/10.1126/science.aad9868>, PMID: 27313038
- Hoem G, Bowitz Larsen K, Øvervatn A, Brech A, Lamark T, Sjøttem E, Johansen T. 2019. The fmrpolyglycine protein mediates aggregate formation and toxicity independent of the CGG mRNA hairpin in a cellular model for FXTAS. *Frontiers in Genetics* **10**:249. DOI: <https://doi.org/10.3389/fgene.2019.00249>, PMID: 30984240
- Huh GS, Hynes RO. 1994. Regulation of alternative pre-mRNA splicing by a novel repeated hexanucleotide element. *Genes & Development* **8**:1561–1574. DOI: <https://doi.org/10.1101/gad.8.13.1561>, PMID: 7958840
- Irwin S, Vandelft M, Pinchev D, Howell JL, Graczyk J, Orr HT, Truant R. 2005. RNA association and nucleocytoplasmic shuttling by ataxin-1. *Journal of Cell Science* **118**:233–242. DOI: <https://doi.org/10.1242/jcs.01611>, PMID: 15615787
- Ishiguro T, Sato N, Ueyama M, Fujikake N, Sellier C, Kanegami A, Tokuda E, Zamiri B, Gall-Duncan T, Mirceta M, Furukawa Y, Yokota T, Wada K, Taylor JP, Pearson CE, Charlet-Berguerand N, Mizusawa H, Nagai Y, Ishikawa K. 2017. Regulatory role of RNA chaperone TDP-43 for RNA misfolding and repeat-associated translation in SCA31. *Neuron* **94**:108–124. DOI: <https://doi.org/10.1016/j.neuron.2017.02.046>, PMID: 28343865
- Ishiura H, Doi K, Mitsui J, Yoshimura J, Matsukawa MK, Fujiyama A, Toyoshima Y, Kakita A, Takahashi H, Suzuki Y, Sugano S, Qu W, Ichikawa K, Yurino H, Higasa K, Shibata S, Mitsue A, Tanaka M, Ichikawa Y, Takahashi Y, et al. 2018. Expansions of intronic TTTCA and TTTTA repeats in benign adult familial myoclonic epilepsy. *Nature Genetics* **50**:581–590. DOI: <https://doi.org/10.1038/s41588-018-0067-2>, PMID: 29507423
- Ishiura H, Shibata S, Yoshimura J, Suzuki Y, Qu W, Doi K, Almansour MA, Kikuchi JK, Taira M, Mitsui J, Takahashi Y, Ichikawa Y, Mano T, Iwata A, Harigaya Y, Matsukawa MK, Matsukawa T, Tanaka M, Shirota Y, Ohtomo R, et al. 2019. Noncoding CGG repeat expansions in neuronal intranuclear inclusion disease, oculopharyngodistal myopathy and an overlapping disease. *Nature Genetics* **51**:1222–1232. DOI: <https://doi.org/10.1038/s41588-019-0458-z>, PMID: 31332380
- Iwahashi CK, Yasui DH, An H-J, Greco CM, Tassone F, Nannen K, Babineau B, Lebrilla CB, Hagerman RJ, Hagerman PJ. 2006. Protein composition of the intranuclear inclusions of FXTAS. *Brain* **129**:256–271. DOI: <https://doi.org/10.1093/brain/awh650>, PMID: 16246864
- Jacquemont S, Hagerman RJ, Leehey M, Grigsby J, Zhang L, Brunberg JA, Greco C, Des Portes V, Jardini T, Levine R, Berry-Kravis E, Brown WT, Schaeffer S, Kissel J, Tassone F, Hagerman PJ. 2003. Fragile X premutation tremor/ataxia syndrome: molecular, clinical, and neuroimaging correlates. *American Journal of Human Genetics* **72**:869–878. DOI: <https://doi.org/10.1086/374321>, PMID: 12638084
- Jain A, Vale RD. 2017. Rna phase transitions in repeat expansion disorders. *Nature* **546**:243–247. DOI: <https://doi.org/10.1038/nature22386>, PMID: 28562589

- Jakubosky D**, D'Antonio M, Bonder MJ, Smail C, Donovan MKR, Young Greenwald WW, Matsui H, D'Antonio-Chronowska A, Stegle O, Smith EN, Montgomery SB, DeBoever C, Frazer KA, i2QTL Consortium. 2020. Properties of structural variants and short tandem repeats associated with gene expression and complex traits. *Nature Communications* **11**:2927. DOI: <https://doi.org/10.1038/s41467-020-16482-4>, PMID: 32522982
- Janitz K**, Janitz M. 2011. Chapter 12 - Assessing epigenetic information. Tollesbol T (Ed). *Handbook of Epigenetics* San Diego: Academic Press. p. 173–181. DOI: <https://doi.org/10.1016/B978-0-12-375709-8.00012-5>
- Jiang H**, Mankodi A, Swanson MS, Moxley RT, Thornton CA. 2004. Myotonic dystrophy type 1 is associated with nuclear foci of mutant RNA, sequestration of muscleblind proteins and deregulated alternative splicing in neurons. *Human Molecular Genetics* **13**:3079–3088. DOI: <https://doi.org/10.1093/hmg/ddh327>, PMID: 15496431
- Jimeno S**, Rondón AG, Luna R, Aguilera A. 2002. The yeast THO complex and mRNA export factors link RNA metabolism with transcription and genome instability. *The EMBO Journal* **21**:3526–3535. DOI: <https://doi.org/10.1093/emboj/cdf335>, PMID: 12093753
- Jin P**, Duan R, Qurashi A, Qin Y, Tian D, Rosser TC, Liu H, Feng Y, Warren ST. 2007. Pur alpha binds to rCGG repeats and modulates repeat-mediated neurodegeneration in a *Drosophila* model of fragile X tremor/ataxia syndrome. *Neuron* **55**:556–564. DOI: <https://doi.org/10.1016/j.neuron.2007.07.020>, PMID: 17698009
- Jog SP**, Paul S, Dansithong W, Tring S, Comai L, Reddy S. 2012. Rna splicing is responsive to MBNL1 dose. *PLOS ONE* **7**:e48825. DOI: <https://doi.org/10.1371/journal.pone.0048825>, PMID: 23166594
- Jovičić A**, Mertens J, Boeynaems S, Bogaert E, Chai N, Yamada SB, Paul JW, Sun S, Herdy JR, Bieri G, Kramer NJ, Gage FH, Van Den Bosch L, Robberecht W, Gitler AD. 2015. Modifiers of c9orf72 dipeptide repeat toxicity connect nucleocytoplasmic transport defects to FTD/ALS. *Nature Neuroscience* **18**:1226–1229. DOI: <https://doi.org/10.1038/nn.4085>, PMID: 26308983
- Kajava AV**. 2012. Tandem repeats in proteins: from sequence to structure. *Journal of Structural Biology* **179**:279–288. DOI: <https://doi.org/10.1016/j.jsb.2011.08.009>, PMID: 21884799
- Katsuno M.**, Adachi H, Kume A, Li M, Nakagomi Y, Niwa H, Sang C, Kobayashi Y, Doyu M, Sobue G. 2002. Testosterone reduction prevents phenotypic expression in a transgenic mouse model of spinal and bulbar muscular atrophy. *Neuron* **35**:843–854. DOI: [https://doi.org/10.1016/s0896-6273\(02\)00834-6](https://doi.org/10.1016/s0896-6273(02)00834-6), PMID: 12372280
- Katsuno M**, Adachi H, Minamiyama M, Waza M, Tokui K, Banno H, Suzuki K, Onoda Y, Tanaka F, Doyu M, Sobue G. 2006. Reversible disruption of dynactin 1-mediated retrograde axonal transport in polyglutamine-induced motor neuron degeneration. *The Journal of Neuroscience* **26**:12106–12117. DOI: <https://doi.org/10.1523/JNEUROSCI.3032-06.2006>, PMID: 17122035
- Kearse MG**, Todd PK. 2014. Repeat-Associated non-AUG translation and its impact in neurodegenerative disease. *Neurotherapeutics* **11**:721–731. DOI: <https://doi.org/10.1007/s13311-014-0292-z>, PMID: 25005000
- Kearse MG**, Green KM, Krans A, Rodriguez CM, Linsalata AE, Goldstrohm AC, Todd PK. 2016. Cgg repeat-associated non-AUG translation utilizes a cap-dependent scanning mechanism of initiation to produce toxic proteins. *Molecular Cell* **62**:314–322. DOI: <https://doi.org/10.1016/j.molcel.2016.02.034>, PMID: 27041225
- Kenneson A**, Zhang F, Hagedorn CH, Warren ST. 2001. Reduced FMRP and increased FMR1 transcription is proportionally associated with CGG repeat number in intermediate-length and premutation carriers. *Human Molecular Genetics* **10**:1449–1454. DOI: <https://doi.org/10.1093/hmg/10.14.1449>, PMID: 11448936
- Kettani A**, Kumar RA, Patel DJ. 1995. Solution structure of a DNA quadruplex containing the fragile X syndrome triplet repeat. *Journal of Molecular Biology* **254**:638–656. DOI: <https://doi.org/10.1006/jmbi.1995.0644>, PMID: 7500339
- Khalil AM**, Faghihi MA, Modarresi F, Brothers SP, Wahlestedt C, Bajic V. 2008. A novel RNA transcript with antiapoptotic function is silenced in fragile X syndrome. *PLOS ONE* **3**:e1486. DOI: <https://doi.org/10.1371/journal.pone.0001486>, PMID: 18213394
- Kim M**, Ahn SH, Krogan NJ, Greenblatt JF, Buratowski S. 2004. Transitions in RNA polymerase II elongation complexes at the 3' ends of genes. *The EMBO Journal* **23**:354–364. DOI: <https://doi.org/10.1038/sj.emboj.7600053>, PMID: 14739930
- Klement IA**, Skinner PJ, Kaytor MD, Yi H, Hersch SM, Clark HB, Zoghbi HY, Orr HT. 1998. Ataxin-1 nuclear localization and aggregation: role in polyglutamine-induced disease in SCA1 transgenic mice. *Cell* **95**:41–53. DOI: [https://doi.org/10.1016/s0092-8674\(00\)81781-x](https://doi.org/10.1016/s0092-8674(00)81781-x), PMID: 9778246
- Klusek J**, Porter A, Abbeduto L, Adayev T, Tassone F, Mailick MR, Glicksman A, Tonnsen BL, Roberts JE. 2018. Curvilinear association between language disfluency and *fmr1* CGG repeat size across the normal, intermediate, and premutation range. *Frontiers in Genetics* **9**:344. DOI: <https://doi.org/10.3389/fgene.2018.00344>, PMID: 30197656
- Knight SJJ**, Flannery AV, Hirst MC, Campbell L, Christodoulou Z, Phelps SR, Pointon J, Middleton-Price HR, Barnicoat A, Pembrey ME. 1993. Trinucleotide repeat amplification and hypermethylation of a CpG island in FRAXE mental retardation. *Cell* **74**:127–134. DOI: [https://doi.org/10.1016/0092-8674\(93\)90300-f](https://doi.org/10.1016/0092-8674(93)90300-f), PMID: 8334699
- Korade-Mirnic Z**, Babitzke P, Hoffman E. 1998. Myotonic dystrophy: molecular windows on a complex etiology. *Nucleic Acids Research* **26**:1363–1368. DOI: <https://doi.org/10.1093/nar/26.6.1363>, PMID: 9490778
- Kozak M**. 1980. Influence of mRNA secondary structure on binding and migration of 40S ribosomal subunits. *Cell* **19**:79–90. DOI: [https://doi.org/10.1016/0092-8674\(80\)90390-6](https://doi.org/10.1016/0092-8674(80)90390-6), PMID: 7357609
- Kozak M**. 1986. Influences of mrna secondary structure on initiation by eukaryotic ribosomes. *PNAS* **83**:2850–2854. DOI: <https://doi.org/10.1073/pnas.83.9.2850>, PMID: 3458245

- Krans A**, Skariah G, Zhang Y, Bayly B, Todd PK. 2019. Neuropathology of ran translation proteins in fragile X-associated tremor/ataxia syndrome. *Acta Neuropathologica Communications* **7**:152. DOI: <https://doi.org/10.1186/s40478-019-0782-7>, PMID: 31665086
- Krzyzosiak WJ**, Sobczak K, Wojciechowska M, Fiszler A, Mykowska A, Kozlowski P. 2012. Triplet repeat RNA structure and its role as pathogenic agent and therapeutic target. *Nucleic Acids Research* **40**:11–26. DOI: <https://doi.org/10.1093/nar/gkr729>, PMID: 21908410
- Kudla G**, Murray AW, Tollervey D, Plotkin JB. 2009. Coding-Sequence determinants of gene expression in *Escherichia coli*. *Science* **324**:255–258. DOI: <https://doi.org/10.1126/science.1170160>, PMID: 19359587
- La Spada AR**, Roling DB, Harding AE, Warner CL, Spiegel R, Hausmanowa-Petrusewicz I, Yee WC, Fischbeck KH. 1992. Meiotic stability and genotype-phenotype correlation of the trinucleotide repeat in X-linked spinal and bulbar muscular atrophy. *Nature Genetics* **2**:301–304. DOI: <https://doi.org/10.1038/ng1292-301>, PMID: 1303283
- Ladd PD**, Smith LE, Rabaia NA, Moore JM, Georges SA, Hansen RS, Hagerman RJ, Tassone F, Tapscott SJ, Filippova GN. 2007. An antisense transcript spanning the CGG repeat region of FMR1 is upregulated in premutation carriers but silenced in full mutation individuals. *Human Molecular Genetics* **16**:3174–3187. DOI: <https://doi.org/10.1093/hmg/ddm293>, PMID: 17921506
- Lai S**, O’Callaghan B, Zoghbi HY, Orr HT. 2011. 14-3-3 binding to ataxin-1 (ATXN1) regulates its dephosphorylation at Ser-776 and transport to the nucleus. *The Journal of Biological Chemistry* **286**:34606–34616. DOI: <https://doi.org/10.1074/jbc.M111.238527>, PMID: 21835928
- Lam YC**, Bowman AB, Jafar-Nejad P, Lim J, Richman R, Fryer JD, Hyun ED, Duvick LA, Orr HT, Botas J, Zoghbi HY. 2006. Ataxin-1 interacts with the repressor capicua in its native complex to cause SCA1 neuropathology. *Cell* **127**:1335–1347. DOI: <https://doi.org/10.1016/j.cell.2006.11.038>, PMID: 17190598
- Lander ES**, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, Devon K, Dewar K, Doyle M, FitzHugh W, Funke R, Gage D, Harris K, Heaford A, Howland J, Kann L, Lehoczky J, LeVine R, McEwan P, McKernan K, et al. 2001. Initial sequencing and analysis of the human genome. *Nature* **409**:860–921. DOI: <https://doi.org/10.1038/35057062>, PMID: 11237011
- Leehey MA**, Munhoz RP, Lang AE, Brunberg JA, Grigsby J, Greco C, Jacquemont S, Tassone F, Lozano AM, Hagerman PJ, Hagerman RJ. 2003. The fragile X premutation presenting as essential tremor. *Archives of Neurology* **60**:117–121. DOI: <https://doi.org/10.1001/archneur.60.1.117>, PMID: 12533098
- Levdansky E**, Romano J, Shadkhan Y, Sharon H, Verstrepen KJ, Fink GR, Osherov N. 2007. Coding tandem repeats generate diversity in *Aspergillus fumigatus* genes. *Eukaryotic Cell* **6**:1380–1391. DOI: <https://doi.org/10.1128/EC.00229-06>, PMID: 17557878
- Li JJ**, Chew GL, Biggin MD. 2017. Quantitating translational control: mRNA abundance-dependent and independent contributions and the mRNA sequences that specify them. *Nucleic Acids Research* **45**:11821–11836. DOI: <https://doi.org/10.1093/nar/gkx898>, PMID: 29040683
- Lian Y**, Garner HR. 2005. Evidence for the regulation of alternative splicing via complementary DNA sequence repeats. *Bioinformatics* **21**:1358–1364. DOI: <https://doi.org/10.1093/bioinformatics/bti180>, PMID: 15673565
- Lim LP**, Sharp PA. 1998. Alternative splicing of the fibronectin EIIIB exon depends on specific TGCATG repeats. *Molecular and Cellular Biology* **18**:3900–3906. DOI: <https://doi.org/10.1128/MCB.18.7.3900>, PMID: 9632774
- Liquori CL**, Ricker K, Moseley ML, Jacobsen JF, Kress W, Naylor SL, Day JW, Ranum LPW. 2001. Myotonic dystrophy type 2 caused by a CCTG expansion in intron 1 of ZNF9. *Science* **293**:864–867. DOI: <https://doi.org/10.1126/science.1062125>, PMID: 11486088
- Liquori CL**, Ikeda Y, Weatherspoon M, Ricker K, Schoser BGH, Dalton JC, Day JW, Ranum LPW. 2003. Myotonic dystrophy type 2: human founder haplotype and evolutionary conservation of the repeat tract. *American Journal of Human Genetics* **73**:849–862. DOI: <https://doi.org/10.1086/378720>, PMID: 14505273
- Liu Q**, Zhang K, Kang Y, Li Y, Deng P, Li Y, Tian Y, Sun Q, Tang Y, Xu K, Zhou Y, Wang J-L, Guo J, Li J-D, Xia K, Meng Q, Allen EG, Wen Z, Li Z, Jiang H, et al. 2022. Expression of expanded GGC repeats within NOTCH2NLC causes behavioral deficits and neurodegeneration in a mouse model of neuronal intranuclear inclusion disease. *Science Advances* **8**:eadd6391. DOI: <https://doi.org/10.1126/sciadv.add6391>, PMID: 36417528
- López Castel A**, Nakamori M, Tomé S, Chitayat D, Gourdon G, Thornton CA, Pearson CE. 2011. Expanded CTG repeat demarcates a boundary for abnormal CpG methylation in myotonic dystrophy patient tissues. *Human Molecular Genetics* **20**:1–15. DOI: <https://doi.org/10.1093/hmg/ddq427>, PMID: 21044947
- López-Martínez A**, Soblechero-Martín P, de-la-Puente-Ovejero L, Nogales-Gadea G, Arechavala-Gomez V. 2020. An overview of alternative splicing defects implicated in myotonic dystrophy type I. *Genes* **11**:1109. DOI: <https://doi.org/10.3390/genes11091109>, PMID: 32971903
- Lu JY**, Chang L, Li T, Wang T, Yin Y, Zhan G, Han X, Zhang K, Tao Y, Percharde M, Wang L, Peng Q, Yan P, Zhang H, Bi X, Shao W, Hong Y, Wu Z, Ma R, Wang P, et al. 2021. Homotypic clustering of L1 and B1/alu repeats compartmentalizes the 3D genome. *Cell Research* **31**:613–630. DOI: <https://doi.org/10.1038/s41422-020-00466-6>, PMID: 33514913
- Lubs HA**. 1969. A marker X chromosome. *American Journal of Human Genetics* **21**:231–244. PMID: 5794013.
- Ma L**, Herren AW, Espinal G, Randol J, McLaughlin B, Martinez-Cerdeño V, Pessah IN, Hagerman RJ, Hagerman PJ. 2019. Composition of the intranuclear inclusions of fragile X-associated tremor/ataxia syndrome. *Acta Neuropathologica Communications* **7**:143. DOI: <https://doi.org/10.1186/s40478-019-0796-1>, PMID: 31481131
- MacDonald ME**, Ambrose CM, Duyao MP, Myers RH, Lin C, Srinidhi L, Barnes G, Taylor SA, James M, Groot N, MacFarlane H, Jenkins B, Anderson MA, Wexler NS, Gusella JF, Bates GP, Baxendale S, Hummerich H, Kirby S,

- North M, et al. 1993. A novel gene containing A trinucleotide repeat that is expanded and unstable on huntington's disease chromosomes: the huntington's disease collaborative research group. *Cell* **72**:971–983. DOI: [https://doi.org/10.1016/0092-8674\(93\)90585-e](https://doi.org/10.1016/0092-8674(93)90585-e)
- Mahadevan M**, Tsiflidis C, Sabourin L, Shutler G, Amemiya C, Jansen G, Neville C, Narang M, Barceló J, O'Hoy K. 1992. Myotonic dystrophy mutation: an unstable CTG repeat in the 3' untranslated region of the gene. *Science* **255**:1253–1255. DOI: <https://doi.org/10.1126/science.1546325>, PMID: 1546325
- Mailick MR**, Hong J, Rathouz P, Baker MW, Greenberg JS, Smith L, Maenner M. 2014. Low-normal FMR1 CGG repeat length: phenotypic associations. *Frontiers in Genetics* **5**:309. DOI: <https://doi.org/10.3389/fgene.2014.00309>, PMID: 25250047
- Malgowska M**, Gudanis D, Kierzek R, Wyszko E, Gabelica V, Gdaniec Z. 2014. Distinctive structural motifs of RNA G-quadruplexes composed of AGG, CGG and UGG trinucleotide repeats. *Nucleic Acids Research* **42**:10196–10207. DOI: <https://doi.org/10.1093/nar/gku710>, PMID: 25081212
- Malik I**, Kelley CP, Wang ET, Todd PK. 2021a. Molecular mechanisms underlying nucleotide repeat expansion disorders. *Nature Reviews. Molecular Cell Biology* **22**:589–607. DOI: <https://doi.org/10.1038/s41580-021-00382-6>, PMID: 34140671
- Malik I**, Tseng YJ, Wright SE, Zheng K, Ramaiyer P, Green KM, Todd PK. 2021b. SRSF protein kinase 1 modulates RAN translation and suppresses CGG repeat toxicity. *EMBO Molecular Medicine* **13**:e14163. DOI: <https://doi.org/10.15252/emmm.202114163>, PMID: 34542927
- Mallick S**, Li H, Lipson M, Mathieson I, Gymrek M, Racimo F, Zhao M, Chennagiri N, Nordenfelt S, Tandon A, Skoglund P, Lazaridis I, Sankararaman S, Fu Q, Rohland N, Renaud G, Erlich Y, Willems T, Gallo C, Spence JP, et al. 2016. The simons genome diversity project: 300 genomes from 142 diverse populations. *Nature* **538**:201–206. DOI: <https://doi.org/10.1038/nature18964>, PMID: 27654912
- Maltby CJ**, Schofield JPR, Houghton SD, O'Kelly I, Vargas-Caballero M, Deinhardt K, Coldwell MJ. 2020. A 5' UTR Ggn repeat controls localisation and translation of a potassium leak channel mRNA through G-quadruplex formation. *Nucleic Acids Research* **48**:9822–9839. DOI: <https://doi.org/10.1093/nar/gkaa699>, PMID: 32870280
- Mankodi A**, Urbinati CR, Yuan QP, Moxley RT, Sansone V, Krym M, Henderson D, Schalling M, Swanson MS, Thornton CA. 2001. Muscleblind localizes to nuclear foci of aberrant RNA in myotonic dystrophy types 1 and 2. *Human Molecular Genetics* **10**:2165–2170. DOI: <https://doi.org/10.1093/hmg/10.19.2165>, PMID: 11590133
- Marcotte EM**, Pellegrini M, Yeates TO, Eisenberg D. 1999. A census of protein repeats. *Journal of Molecular Biology* **293**:151–160. DOI: <https://doi.org/10.1006/jmbi.1999.3136>, PMID: 10512723
- Martin JP**, Bell J. 1943. A pedigree of mental defect showing sex-linkage. *Journal of Neurology and Psychiatry* **6**:154–157. DOI: <https://doi.org/10.1136/jnnp.6.3-4.154>, PMID: 21611430
- Mastroiannopoulos NP**, Shammas C, Phylactou LA. 2010. Tackling the pathogenesis of RNA nuclear retention in myotonic dystrophy. *Biology of the Cell* **102**:515–523. DOI: <https://doi.org/10.1042/BC20100040>, PMID: 20690904
- May S**, Hornburg D, Schludi MH, Arzberger T, Rentzsch K, Schwenk BM, Grässer FA, Mori K, Kremmer E, Banzhaf-Strathmann J, Mann M, Meissner F, Edbauer D. 2014. C9orf72 FTLN/ALS-associated Gly-Ala dipeptide repeat proteins cause neuronal toxicity and UNC119 sequestration. *Acta Neuropathologica* **128**:485–503. DOI: <https://doi.org/10.1007/s00401-014-1329-4>, PMID: 25120191
- McClintock B**. 1950. The origin and behavior of mutable loci in maize. *PNAS* **36**:344–355. DOI: <https://doi.org/10.1073/pnas.36.6.344>, PMID: 15430309
- McConkie-Rosell A**, Lachiewicz AM, Spiridigliozzi GA, Tarleton J, Schoenwald S, Phelan MC, Goonewardena P, Ding X, Brown WT. 1993. Evidence that methylation of the FMR-1 locus is responsible for variable phenotypic expression of the fragile X syndrome. *American Journal of Human Genetics* **53**:800–809. PMID: 8213810.
- McEachin ZT**, Gendron TF, Raj N, García-Murias M, Banerjee A, Purcell RH, Ward PJ, Todd TW, Merritt-Garza ME, Jansen-West K, Hales CM, García-Sobrino T, Quintans B, Holler CJ, Taylor G, San Millán B, Teijeira S, Yamashita T, Ohkubo R, Boulis NM, et al. 2020. Chimeric peptide species contribute to divergent dipeptide repeat pathology in c9als/FTD and SCA36. *Neuron* **107**:292–305. DOI: <https://doi.org/10.1016/j.neuron.2020.04.011>, PMID: 32375063
- McIver LJ**, Fondon JW, Skinner MA, Garner HR. 2011. Evaluation of microsatellite variation in the 1000 genomes project pilot studies is indicative of the quality and utility of the RAW data and alignments. *Genomics* **97**:193–199. DOI: <https://doi.org/10.1016/j.ygeno.2011.01.001>, PMID: 21223998
- McIver LJ**, McCormick JF, Martin A, Fondon JW, Garner HR. 2013. Population-scale analysis of human microsatellites reveals novel sources of exonic variation. *Gene* **516**:328–334. DOI: <https://doi.org/10.1016/j.gene.2012.12.068>, PMID: 23274653
- Menzies FM**, Huebener J, Renna M, Bonin M, Riess O, Rubinsztein DC. 2010. Autophagy induction reduces mutant ataxin-3 levels and toxicity in a mouse model of spinocerebellar ataxia type 3. *Brain* **133**:93–104. DOI: <https://doi.org/10.1093/brain/awp292>, PMID: 20007218
- Miller JW**, Urbinati CR, Teng-Umnuay P, Stenberg MG, Byrne BJ, Thornton CA, Swanson MS. 2000. Recruitment of human muscleblind proteins to (CUG) (N) expansions associated with myotonic dystrophy. *The EMBO Journal* **19**:4439–4448. DOI: <https://doi.org/10.1093/emboj/19.17.4439>, PMID: 10970838
- Millette MM**, Holland ED, Tenpas TJ, Dent EW. 2022. A single transcript knockdown-replacement strategy employing 5' UTR secondary structures to precisely titrate rescue protein translation. *Frontiers in Genome Editing* **4**:803375. DOI: <https://doi.org/10.3389/fgene.2022.803375>, PMID: 35419562
- Mitra I**, Huang B, Mousavi N, Ma N, Lamkin M, Yanicky R, Shleizer-Burko S, Lohmueller KE, Gymrek M. 2021. Patterns of de novo tandem repeat mutations and their role in autism. *Nature* **589**:246–250. DOI: <https://doi.org/10.1038/s41586-020-03078-7>, PMID: 33442040

- Mitsuhashi S**, Matsumoto N. 2020. Long-Read sequencing for rare human genetic diseases. *Journal of Human Genetics* **65**:11–19. DOI: <https://doi.org/10.1038/s10038-019-0671-8>, PMID: 31558760
- Mizielinska S**, Grönke S, Niccoli T, Ridler CE, Clayton EL, Devoy A, Moens T, Norona FE, Woollacott IOC, Pietrzyk J, Cleverley K, Nicoll AJ, Pickering-Brown S, Dols J, Cabecinha M, Hendrich O, Fratta P, Fisher EMC, Partridge L, Isaacs AM. 2014. C9orf72 repeat expansions cause neurodegeneration in *Drosophila* through arginine-rich proteins. *Science* **345**:1192–1194. DOI: <https://doi.org/10.1126/science.1256800>, PMID: 25103406
- Mojarad BA**, Engchuan W, Trost B, Backstrom I, Yin Y, Thiruvahindrapuram B, Pallotto L, Mitina A, Khan M, Pellecchia G, Haque B, Guo K, Heung T, Costain G, Scherer SW, Marshall CR, Pearson CE, Bassett AS, Yuen RKC. 2022. Genome-Wide tandem repeat expansions contribute to schizophrenia risk. *Molecular Psychiatry* **27**:3692–3698. DOI: <https://doi.org/10.1038/s41380-022-01575-x>, PMID: 35546631
- Montie HL**, Cho MS, Holder L, Liu Y, Tsvetkov AS, Finkbeiner S, Merry DE. 2009. Cytoplasmic retention of polyglutamine-expanded androgen receptor ameliorates disease via autophagy in a mouse model of spinal and bulbar muscular atrophy. *Human Molecular Genetics* **18**:1937–1950. DOI: <https://doi.org/10.1093/hmg/ddp115>, PMID: 19279159
- Mori K**, Arzberger T, Grässer FA, Gijssels I, May S, Rentzsch K, Weng SM, Schludi MH, van der Zee J, Cruts M, Van Broeckhoven C, Kremmer E, Kretzschmar HA, Haass C, Edbauer D. 2013a. Bidirectional transcripts of the expanded c9orf72 hexanucleotide repeat are translated into aggregating dipeptide repeat proteins. *Acta Neuropathologica* **126**:881–893. DOI: <https://doi.org/10.1007/s00401-013-1189-3>, PMID: 24132570
- Mori K**, Weng SM, Arzberger T, May S, Rentzsch K, Kremmer E, Schmid B, Kretzschmar HA, Cruts M, Van Broeckhoven C, Haass C, Edbauer D. 2013b. The c9orf72 GGGGCC repeat is translated into aggregating dipeptide-repeat proteins in FTLD/ALS. *Science* **339**:1335–1338. DOI: <https://doi.org/10.1126/science.1232927>, PMID: 23393093
- Moseley ML**, Zu T, Ikeda Y, Gao W, Mosemiller AK, Daughters RS, Chen G, Weatherspoon MR, Clark HB, Ebner TJ, Day JW, Ranum LPW. 2006. Bidirectional expression of CUG and CAG expansion transcripts and intranuclear polyglutamine inclusions in spinocerebellar ataxia type 8. *Nature Genetics* **38**:758–769. DOI: <https://doi.org/10.1038/ng1827>, PMID: 16804541
- Mulley JC**, Yu S, Loesch DZ, Hay DA, Donnelly A, Gedeon AK, Carbonell P, López I, Glover G, Gabarrón I. 1995. Fraxe and mental retardation. *Journal of Medical Genetics* **32**:162–169. DOI: <https://doi.org/10.1136/jmg.32.3.162>, PMID: 7783162
- Muro AF**, Caputi M, Pariyath R, Pagani F, Buratti E, Baralle FE. 1999. Regulation of fibronectin EDA exon alternative splicing: possible role of RNA secondary structure for enhancer display. *Molecular and Cellular Biology* **19**:2657–2671. DOI: <https://doi.org/10.1128/MCB.19.4.2657>, PMID: 10082532
- Murray A**, Ennis S, MacSwiney F, Webb J, Morton NE. 2000a. Reproductive and menstrual history of females with fragile X expansions. *European Journal of Human Genetics* **8**:247–252. DOI: <https://doi.org/10.1038/sj.ejhg.5200451>, PMID: 10854106
- Murray A**, Ennis S, Morton N. 2000b. No evidence for parent of origin influencing premature ovarian failure in fragile X premutation carriers. *American Journal of Human Genetics* **67**:253–254. DOI: <https://doi.org/10.1086/302963>, PMID: 10848495
- Muslimov IA**, Patel MV, Rose A, Tiedge H. 2011. Spatial code recognition in neuronal RNA targeting: role of RNA-hnRNP A2 interactions. *The Journal of Cell Biology* **194**:441–457. DOI: <https://doi.org/10.1083/jcb.201010027>, PMID: 21807882
- Mykowska A**, Sobczak K, Wojciechowska M, Kozłowski P, Krzyżosiak WJ. 2011. Cag repeats mimic CUG repeats in the misregulation of alternative splicing. *Nucleic Acids Research* **39**:8938–8951. DOI: <https://doi.org/10.1093/nar/gkr608>, PMID: 21795378
- Myrick LK**, Nakamoto-Kinoshita M, Lindor NM, Kirmani S, Cheng X, Warren ST. 2014. Fragile X syndrome due to a missense mutation. *European Journal of Human Genetics* **22**:1185–1189. DOI: <https://doi.org/10.1038/ejhg.2013.311>, PMID: 24448548
- Nakamura H**, Doi H, Mitsuhashi S, Miyatake S, Katoh K, Frith MC, Asano T, Kudo Y, Ikeda T, Kubota S, Kunii M, Kitazawa Y, Tada M, Okamoto M, Joki H, Takeuchi H, Matsumoto N, Tanaka F. 2020. Long-Read sequencing identifies the pathogenic nucleotide repeat expansion in RFC1 in a Japanese case of canvas. *Journal of Human Genetics* **65**:475–480. DOI: <https://doi.org/10.1038/s10038-020-0733-y>, PMID: 32066831
- Nasrallah MP**, Cho G, Simonet JC, Putt ME, Kitamura K, Golden JA. 2012. Differential effects of a polyalanine tract expansion in Arx on neural development and gene expression. *Human Molecular Genetics* **21**:1090–1098. DOI: <https://doi.org/10.1093/hmg/ddr538>, PMID: 22108177
- Neueder A**, Landles C, Ghosh R, Howland D, Myers RH, Faull RLM, Tabrizi SJ, Bates GP. 2017. The pathogenic exon 1 HTT protein is produced by incomplete splicing in huntington's disease patients. *Scientific Reports* **7**:1307. DOI: <https://doi.org/10.1038/s41598-017-01510-z>, PMID: 28465506
- Neueder A**, Dumas AA, Benjamin AC, Bates GP. 2018. Regulatory mechanisms of incomplete huntingtin mRNA splicing. *Nature Communications* **9**:3955. DOI: <https://doi.org/10.1038/s41467-018-06281-3>, PMID: 30262848
- Niederer RO**, Rojas-Duran MF, Zinshteyn B, Gilbert WV. 2022. Direct analysis of ribosome targeting illuminates thousand-fold regulation of translation initiation. *Cell Systems* **13**:256–264. DOI: <https://doi.org/10.1016/j.cels.2021.12.002>, PMID: 35041803
- Nikumbh S**, Pfeifer N. 2017. Genetic sequence-based prediction of long-range chromatin interactions suggests a potential role of short tandem repeat sequences in genome organization. *BMC Bioinformatics* **18**:218. DOI: <https://doi.org/10.1186/s12859-017-1624-x>, PMID: 28420341

- Oberlé I**, Rousseau F, Heitz D, Kretz C, Devys D, Hanauer A, Boué J, Bertheas MF, Mandel JL. 1991. Instability of a 550-base pair DNA segment and abnormal methylation in fragile X syndrome. *Science* **252**:1097–1102. DOI: <https://doi.org/10.1126/science.252.5009.1097>, PMID: 2031184
- O'Dushlaine CT**, Shields DC. 2008. Marked variation in predicted and observed variability of tandem repeat loci across the human genome. *BMC Genomics* **9**:175. DOI: <https://doi.org/10.1186/1471-2164-9-175>, PMID: 18416815
- Ogasawara M**, Iida A, Kumutpongpanich T, Ozaki A, Oya Y, Konishi H, Nakamura A, Abe R, Takai H, Hanajima R, Doi H, Tanaka F, Nakamura H, Nonaka I, Wang Z, Hayashi S, Noguchi S, Nishino I. 2020. Cgg expansion in NOTCH2NLC is associated with oculopharyngodistal myopathy with neurological manifestations. *Acta Neuropathologica Communications* **8**:204. DOI: <https://doi.org/10.1186/s40478-020-01084-4>, PMID: 33239111
- Oh SY**, He F, Krans A, Frazer M, Taylor JP, Paulson HL, Todd PK. 2015. Ran translation at CGG repeats induces ubiquitin proteasome system impairment in models of fragile X-associated tremor ataxia syndrome. *Human Molecular Genetics* **24**:4317–4326. DOI: <https://doi.org/10.1093/hmg/ddv165>, PMID: 25954027
- Okubo M**, Doi H, Fukai R, Fujita A, Mitsuhashi S, Hashiguchi S, Kishida H, Ueda N, Morihara K, Ogasawara A, Kawamoto Y, Takahashi T, Takahashi K, Nakamura H, Kunii M, Tada M, Katsumoto A, Fukuda H, Mizuguchi T, Miyatake S, et al. 2019. Ggc repeat expansion of NOTCH2NLC in adult patients with leukoencephalopathy. *Annals of Neurology* **86**:962–968. DOI: <https://doi.org/10.1002/ana.25586>, PMID: 31433517
- Ordway JM**, Tallaksen-Greene S, Gutekunst CA, Bernstein EM, Cearley JA, Wiener HW, Dure LS, Lindsey R, Hersch SM, Jope RS, Albin RL, Detloff PJ. 1997. Ectopically expressed CAG repeats cause intranuclear inclusions and a progressive late onset neurological phenotype in the mouse. *Cell* **91**:753–763. DOI: [https://doi.org/10.1016/s0092-8674\(00\)80464-x](https://doi.org/10.1016/s0092-8674(00)80464-x), PMID: 9413985
- Orr HT**. 2012. Polyglutamine neurodegeneration: expanded glutamines enhance native functions. *Current Opinion in Genetics & Development* **22**:251–255. DOI: <https://doi.org/10.1016/j.gde.2012.01.001>, PMID: 22284692
- Osadchuk L**, Vasiliev G, Kleshchev M, Osadchuk A. 2022. Androgen receptor gene CAG repeat length varies and affects semen quality in an ethnic-specific fashion in young men from russia. *International Journal of Molecular Sciences* **23**:10594. DOI: <https://doi.org/10.3390/ijms231810594>, PMID: 36142533
- Otten AD**, Tapscott SJ. 1995. Triplet repeat expansion in myotonic dystrophy alters the adjacent chromatin structure. *PNAS* **92**:5465–5469. DOI: <https://doi.org/10.1073/pnas.92.12.5465>, PMID: 7777532
- Palazzolo I**, Burnett BG, Young JE, Brenne PL, La Spada AR, Fischbeck KH, Howell BW, Pennuto M. 2007. Akt blocks ligand binding and protects against expanded polyglutamine androgen receptor toxicity. *Human Molecular Genetics* **16**:1593–1603. DOI: <https://doi.org/10.1093/hmg/ddm109>, PMID: 17470458
- Pan B**, Li R, Chen Y, Tang Q, Wu W, Chen L, Lu C, Pan F, Ding H, Xia Y, Hu L, Chen D, Sha J, Wang X. 2016. Genetic association between androgen receptor gene CAG repeat length polymorphism and male infertility. *Medicine* **95**:e2878. DOI: <https://doi.org/10.1097/MD.0000000000002878>, PMID: 26962784
- Pappalardo XG**, Barra V. 2021. Losing DNA methylation at repetitive elements and breaking bad. *Epigenetics & Chromatin* **14**:25. DOI: <https://doi.org/10.1186/s13072-021-00400-z>, PMID: 34082816
- Pascual M**, Vicente M, Monferrer L, Artero R. 2006. The muscleblind family of proteins: an emerging class of regulators of developmentally programmed alternative splicing. *Differentiation; Research in Biological Diversity* **74**:65–80. DOI: <https://doi.org/10.1111/j.1432-0436.2006.00060.x>, PMID: 16533306
- Pastori C**, Peschansky VJ, Barbouth D, Mehta A, Silva JP, Wahlestedt C. 2014. 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. *Human Genetics* **133**:59–67. DOI: <https://doi.org/10.1007/s00439-013-1356-6>, PMID: 24005575
- Patel PK**, Bhavesh NS, Hosur RV. 2000. Cation-Dependent conformational switches in d-TGGCGGC containing two triplet repeats of fragile X syndrome: NMR observations. *Biochemical and Biophysical Research Communications* **278**:833–838. DOI: <https://doi.org/10.1006/bbrc.2000.3878>, PMID: 11095993
- Paul S**, Dansithong W, Jog SP, Holt I, Mittal S, Brook JD, Morris GE, Comai L, Reddy S. 2011. Expanded CUG repeats Dysregulate RNA splicing by altering the stoichiometry of the muscleblind 1 complex. *The Journal of Biological Chemistry* **286**:38427–38438. DOI: <https://doi.org/10.1074/jbc.M111.255224>, PMID: 21900255
- Paulson HL**, Perez MK, Trottier Y, Trojanowski JQ, Subramony SH, Das SS, Vig P, Mandel JL, Fischbeck KH, Pittman RN. 1997. Intranuclear inclusions of expanded polyglutamine protein in spinocerebellar ataxia type 3. *Neuron* **19**:333–344. DOI: [https://doi.org/10.1016/s0896-6273\(00\)80943-5](https://doi.org/10.1016/s0896-6273(00)80943-5), PMID: 9292723
- Paulson H**. 2018. Chapter 9 - Repeat expansion diseases. Geschwind DH, Paulson HL, Klein C (Eds). *Handbook of Clinical Neurology, Neurogenetics, Part I* Elsevier. p. 105–123. DOI: <https://doi.org/10.1016/B978-0-444-63233-3.00009-9>
- Payseur BA**, Jing P, Haas RJ. 2011. A genomic portrait of human microsatellite variation. *Molecular Biology and Evolution* **28**:303–312. DOI: <https://doi.org/10.1093/molbev/msq198>, PMID: 20675409
- Philips AV**, Timchenko LT, Cooper TA. 1998. Disruption of splicing regulated by a CUG-binding protein in myotonic dystrophy. *Science* **280**:737–741. DOI: <https://doi.org/10.1126/science.280.5364.737>, PMID: 9563950
- Pieretti M**, Zhang FP, Fu YH, Warren ST, Oostra BA, Caskey CT, Nelson DL. 1991. Absence of expression of the FMR-1 gene in fragile X syndrome. *Cell* **66**:817–822. DOI: [https://doi.org/10.1016/0092-8674\(91\)90125-i](https://doi.org/10.1016/0092-8674(91)90125-i), PMID: 1878973
- Prado F**, Piruat JI, Aguilera A. 1997. Recombination between DNA repeats in yeast hpr1Delta cells is linked to transcription elongation. *The EMBO Journal* **16**:2826–2835. DOI: <https://doi.org/10.1093/emboj/16.10.2826>, PMID: 9184227

- Proops R**, Webb T. 1981. The “ fragile ” X chromosome in the martin-bell-renpenning syndrome and in males with other forms of familial mental retardation. *Journal of Medical Genetics* **18**:366–373. DOI: <https://doi.org/10.1136/jmg.18.5.366>, PMID: 7328617
- Quartier A**, Poquet H, Gilbert-Dussardier B, Rossi M, Casteleyn AS, Portes V, Feger C, Nourisson E, Kuentz P, Redin C, Thevenon J, Mosca-Boidron AL, Callier P, Muller J, Lesca G, Huet F, Geoffroy V, El Chehadeh S, Jung M, Trojak B, et al. 2017. Intragenic FMR1 disease-causing variants: a significant mutational mechanism leading to fragile-X syndrome. *European Journal of Human Genetics* **25**:423–431. DOI: <https://doi.org/10.1038/ejhg.2016.204>, PMID: 28176767
- Quilez J**, Guilmatre A, Garg P, Highnam G, Gymrek M, Erlich Y, Joshi RS, Mittelman D, Sharp AJ. 2016. Polymorphic tandem repeats within gene promoters act as modifiers of gene expression and DNA methylation in humans. *Nucleic Acids Research* **44**:3750–3762. DOI: <https://doi.org/10.1093/nar/gkw219>, PMID: 27060133
- Qurashi A**, Li W, Zhou JY, Peng J, Jin P. 2011. Nuclear accumulation of stress response mRNAs contributes to the neurodegeneration caused by fragile X premutation rCGG repeats. *PLOS Genetics* **7**:e1002102. DOI: <https://doi.org/10.1371/journal.pgen.1002102>, PMID: 21655086
- Ravikumar B**, Vacher C, Berger Z, Davies JE, Luo S, Oroz LG, Scaravilli F, Easton DF, Duden R, O’Kane CJ, Rubinsztein DC. 2004. Inhibition of mTOR induces autophagy and reduces toxicity of polyglutamine expansions in fly and mouse models of Huntington disease. *Nature Genetics* **36**:585–595. DOI: <https://doi.org/10.1038/ng1362>
- Ravindran S**. 2012. Barbara mcclintock and the discovery of jumping genes. *PNAS* **109**:20198–20199. DOI: <https://doi.org/10.1073/pnas.1219372109>, PMID: 23236127
- Rodriguez CM**, Todd PK. 2019. New pathologic mechanisms in nucleotide repeat expansion disorders. *Neurobiology of Disease* **130**:104515. DOI: <https://doi.org/10.1016/j.nbd.2019.104515>, PMID: 31229686
- Rodriguez CM**, Wright SE, Kearse MG, Haenfler JM, Flores BN, Liu Y, Ifrim MF, Glineburg MR, Krans A, Jafar-Nejad P, Sutton MA, Bassell GJ, Parent JM, Rigo F, Barmada SJ, Todd PK. 2020. A native function for RAN translation and CGG repeats in regulating fragile X protein synthesis. *Nature Neuroscience* **23**:386–397. DOI: <https://doi.org/10.1038/s41593-020-0590-1>, PMID: 32066985
- Rosas I**, Martínez C, Clarimón J, Lleó A, Illán-Gala I, Dols-Icardo O, Borroni B, Almeida MR, van der Zee J, Van Broeckhoven C, Bruni AC, Anfossi M, Bernardi L, Maletta R, Serpente M, Galimberti D, Scarpini E, Rossi G, Caroppo P, Benussi L, et al. 2020. Role for ATXN1, ATXN2, and HTT intermediate repeats in frontotemporal dementia and alzheimer’s disease. *Neurobiology of Aging* **87**:139. DOI: <https://doi.org/10.1016/j.neurobiolaging.2019.10.017>, PMID: 31810584
- Rubinsztein DC**. 2006. The roles of intracellular protein-degradation pathways in neurodegeneration. *Nature* **443**:780–786. DOI: <https://doi.org/10.1038/nature05291>, PMID: 17051204
- Sathasivam K**, Neueder A, Gipson TA, Landles C, Benjamin AC, Bondulich MK, Smith DL, Faull RLM, Roos RAC, Howland D, Detloff PJ, Housman DE, Bates GP. 2013. Aberrant splicing of HTT generates the pathogenic exon 1 protein in huntington disease. *PNAS* **110**:2366–2370. DOI: <https://doi.org/10.1073/pnas.1221891110>, PMID: 23341618
- Sato N**, Amino T, Kobayashi K, Asakawa S, Ishiguro T, Tsunemi T, Takahashi M, Matsuura T, Flanigan KM, Iwasaki S, Ishino F, Saito Y, Murayama S, Yoshida M, Hashizume Y, Takahashi Y, Tsuji S, Shimizu N, Toda T, Ishikawa K, et al. 2009. Spinocerebellar ataxia type 31 is associated with “ inserted ” penta-nucleotide repeats containing (TGGA) N. *American Journal of Human Genetics* **85**:544–557. DOI: <https://doi.org/10.1016/j.ajhg.2009.09.019>, PMID: 19878914
- Schaefer MH**, Wanker EE, Andrade-Navarro MA. 2012. Evolution and function of CAG/polyglutamine repeats in protein-protein interaction networks. *Nucleic Acids Research* **40**:4273–4287. DOI: <https://doi.org/10.1093/nar/gks011>, PMID: 22287626
- Schilling G**, Becher MW, Sharp AH, Jinnah HA, Duan K, Kotzuc JA, Slunt HH, Ratovitski T, Cooper JK, Jenkins NA, Copeland NG, Price DL, Ross CA, Borchelt DR. 1999. Intranuclear inclusions and neuritic aggregates in transgenic mice expressing a mutant N-terminal fragment of huntingtin. *Human Molecular Genetics* **8**:397–407. DOI: <https://doi.org/10.1093/hmg/8.3.397>, PMID: 9949199
- Schwartz M**, Zlotorynski E, Kerem B. 2006. The molecular basis of common and rare fragile sites. *Cancer Letters* **232**:13–26. DOI: <https://doi.org/10.1016/j.canlet.2005.07.039>, PMID: 16236432
- Sellier C**, Rau F, Liu Y, Tassone F, Hukema RK, Gattoni R, Schneider A, Richard S, Willemsen R, Elliott DJ, Hagerman PJ, Charlet-Berguerand N. 2010. Sam68 sequestration and partial loss of function are associated with splicing alterations in FXTAS patients. *The EMBO Journal* **29**:1248–1261. DOI: <https://doi.org/10.1038/emboj.2010.21>, PMID: 20186122
- Sellier C**, Freyermuth F, Tabet R, Tran T, He F, Ruffenach F, Alunni V, Moine H, Thibault C, Page A, Tassone F, Willemsen R, Disney MD, Hagerman PJ, Todd PK, Charlet-Berguerand N. 2013. Sequestration of Drosha and DGCR8 by expanded CGG RNA repeats alters microRNA processing in fragile X-associated tremor/ataxia syndrome. *Cell Reports* **3**:869–880. DOI: <https://doi.org/10.1016/j.celrep.2013.02.004>, PMID: 23478018
- Sellier C**, Buijssen RAM, He F, Natla S, Jung L, Tropel P, Gaucherot A, Jacobs H, Meziane H, Vincent A, Champy M-F, Sorg T, Pavlovic G, Wattenhofer-Donze M, Birling M-C, Oulad-Abdelghani M, Eberling P, Ruffenach F, Joint M, Anheim M, et al. 2017. Translation of expanded CGG repeats into fmprpolyg is pathogenic and may contribute to fragile X tremor ataxia syndrome. *Neuron* **93**:331–347. DOI: <https://doi.org/10.1016/j.neuron.2016.12.016>, PMID: 28065649
- Sellier C**, Cerro-Herreros E, Blatter M, Freyermuth F, Gaucherot A, Ruffenach F, Sarkar P, Puymirat J, Udd B, Day JW, Meola G, Bassez G, Fujimura H, Takahashi MP, Schoser B, Furling D, Artero R, Allain FHT, Llamusi B, Charlet-Berguerand N. 2018. RbFOX1/MBNL1 competition for CCUG RNA repeats binding contributes to

- myotonic dystrophy type 1/type 2 differences. *Nature Communications* **9**:2009. DOI: <https://doi.org/10.1038/s41467-018-04370-x>, PMID: 29789616
- Sharma M**, Pandey GK. 2015. Expansion and function of repeat domain proteins during stress and development in plants. *Frontiers in Plant Science* **6**:1218. DOI: <https://doi.org/10.3389/fpls.2015.01218>, PMID: 26793205
- Sherman SL**, Morton NE, Jacobs PA, Turner G. 1984. The marker (X) syndrome: a cytogenetic and genetic analysis. *Annals of Human Genetics* **48**:21–37. DOI: <https://doi.org/10.1111/j.1469-1809.1984.tb00830.x>, PMID: 6712153
- Sherman SL**, Jacobs PA, Morton NE, Froster-Iskenius U, Howard-Peebles PN, Nielsen KB, Partington MW, Sutherland GR, Turner G, Watson M. 1985. Further segregation analysis of the fragile X syndrome with special reference to transmitting males. *Human Genetics* **69**:289–299. DOI: <https://doi.org/10.1007/BF00291644>, PMID: 3838733
- Smith KP**, Byron M, Johnson C, Xing Y, Lawrence JB. 2007. Defining early steps in mRNA transport: mutant mRNA in myotonic dystrophy type I is blocked at entry into SC-35 domains. *The Journal of Cell Biology* **178**:951–964. DOI: <https://doi.org/10.1083/jcb.200706048>, PMID: 17846170
- Sobczak K.**, Mezer M, Michlewski G, Krol J, Krzyzosiak WJ. 2003. Rna structure of trinucleotide repeats associated with human neurological diseases. *Nucleic Acids Research* **31**:5469–5482. DOI: <https://doi.org/10.1093/nar/gkg766>, PMID: 14500809
- Sobczak K**, Michlewski G, de Mezer M, Kierzek E, Krol J, Olejniczak M, Kierzek R, Krzyzosiak WJ. 2010. Structural diversity of triplet repeat mas. *Journal of Biological Chemistry* **285**:12755–12764. DOI: <https://doi.org/10.1074/jbc.M109.078790>, PMID: 20159983
- Sofola OA**, Jin P, Qin Y, Duan R, Liu H, de Haro M, Nelson DL, Botas J. 2007. Rna-Binding proteins hnRNP A2/B1 and CUGBP1 suppress fragile X CGG premutation repeat-induced neurodegeneration in a *Drosophila* model of FXTAS. *Neuron* **55**:565–571. DOI: <https://doi.org/10.1016/j.neuron.2007.07.021>, PMID: 17698010
- Solnick D**, Lee SI. 1987. Amount of RNA secondary structure required to induce an alternative splice. *Molecular and Cellular Biology* **7**:3194–3198. DOI: <https://doi.org/10.1128/mcb.7.9.3194-3198.1987>, PMID: 3670310
- Sone J**, Mitsuhashi S, Fujita A, Mizuguchi T, Hamanaka K, Mori K, Koike H, Hashiguchi A, Takashima H, Sugiyama H, Kohno Y, Takiyama Y, Maeda K, Doi H, Koyano S, Takeuchi H, Kawamoto M, Kohara N, Ando T, Ieda T, et al. 2019. Long-Read sequencing identifies GGC repeat expansions in NOTCH2NL associated with neuronal intranuclear inclusion disease. *Nature Genetics* **51**:1215–1221. DOI: <https://doi.org/10.1038/s41588-019-0459-y>, PMID: 31332381
- Soragni E**, Petrosyan L, Rinkoski TA, Wieben ED, Baratz KH, Fautsch MP, Gottesfeld JM. 2018. Repeat-associated non-ATG (ran) translation in fuchs' endothelial corneal dystrophy. *Investigative Ophthalmology & Visual Science* **59**:1888. DOI: <https://doi.org/10.1167/iovs.17-23265>, PMID: 29677349
- Steinbach P**, Gläser D, Vogel W, Wolf M, Schwemmler S. 1998. The DMPK gene of severely affected myotonic dystrophy patients is hypermethylated proximal to the largely expanded CTG repeat. *American Journal of Human Genetics* **62**:278–285. DOI: <https://doi.org/10.1086/301711>, PMID: 9463318
- Strässer K**, Masuda S, Mason P, Pfannstiel J, Oppizzi M, Rodríguez-Navarro S, Rondón AG, Aguilera A, Struhl K, Reed R, Hurt E. 2002. Trex is a conserved complex coupling transcription with messenger RNA export. *Nature* **417**:304–308. DOI: <https://doi.org/10.1038/nature746>, PMID: 11979277
- Subramanian PS**, Nelson DL, Chinault AC. 1996. Large domains of apparent delayed replication timing associated with triplet repeat expansion at FRAXA and FRAXE. *American Journal of Human Genetics* **59**:407–416 PMID: 8755928.
- Sulovari A**, Li R, Audano PA, Porubsky D, Vollger MR, Logsdon GA, Warren WC, Pollen AA, Chaisson MJP, Eichler EE, Human Genome Structural Variation Consortium. 2019. Human-specific tandem repeat expansion and differential gene expression during primate evolution. *PNAS* **116**:23243–23253. DOI: <https://doi.org/10.1073/pnas.1912175116>, PMID: 31659027
- Sun X**, Li PP, Zhu S, Cohen R, Marque LO, Ross CA, Pulst SM, Chan HYE, Margolis RL, Rudnicki DD. 2015. Nuclear retention of full-length HTT RNA is mediated by splicing factors MBNL1 and U2AF65. *Scientific Reports* **5**:12521. DOI: <https://doi.org/10.1038/srep12521>, PMID: 26218986
- Sun JH**, Zhou L, Emerson DJ, Phyto SA, Titus KR, Gong W, Gilgenast TG, Beagan JA, Davidson BL, Tassone F, Phillips-Cremens JE. 2018. Disease-Associated short tandem repeats co-localize with chromatin domain boundaries. *Cell* **175**:224–238. DOI: <https://doi.org/10.1016/j.cell.2018.08.005>, PMID: 30173918
- Sutcliffe JS**, Nelson DL, Zhang F, Pieretti M, Caskey CT, Saxe D, Warren ST. 1992. Dna methylation represses FMR-1 transcription in fragile X syndrome. *Human Molecular Genetics* **1**:397–400. DOI: <https://doi.org/10.1093/hmg/1.6.397>, PMID: 1301913
- Tabet R**, Schaeffer L, Freyermuth F, Jambeau M, Workman M, Lee C-Z, Lin C-C, Jiang J, Jansen-West K, Abou-Hamdan H, Désaubry L, Gendron T, Petrucelli L, Martin F, Lagier-Tourenne C. 2018. CUG initiation and frameshifting enable production of dipeptide repeat proteins from ALS/FTD C9ORF72 transcripts. *Nature Communications* **9**:152. DOI: <https://doi.org/10.1038/s41467-017-02643-5>, PMID: 29323119
- Tacke R**, Chen Y, Manley JL. 1997. Sequence-specific RNA binding by an SR protein requires RS domain phosphorylation: creation of an srp40-specific splicing enhancer. *PNAS* **94**:1148–1153. DOI: <https://doi.org/10.1073/pnas.94.4.1148>, PMID: 9037021
- Tam V**, Patel N, Turcotte M, Bossé Y, Paré G, Meyre D. 2019. Benefits and limitations of genome-wide association studies. *Nature Reviews. Genetics* **20**:467–484. DOI: <https://doi.org/10.1038/s41576-019-0127-1>, PMID: 31068683
- Tassone F**, Hagerman RJ, Loesch DZ, Lachiewicz A, Taylor AK, Hagerman PJ. 2000a. Fragile X males with unmethylated, full mutation trinucleotide repeat expansions have elevated levels of FMR1 messenger RNA.

- American Journal of Medical Genetics* **94**:232–236. DOI: [https://doi.org/10.1002/1096-8628\(20000918\)94:3<232::aid-ajmg9>3.0.co;2-h](https://doi.org/10.1002/1096-8628(20000918)94:3<232::aid-ajmg9>3.0.co;2-h), PMID: 10995510
- Tassone F**, Hagerman RJ, Taylor AK, Gane LW, Godfrey TE, Hagerman PJ. 2000b. Elevated levels of FMR1 mRNA in carrier males: a new mechanism of involvement in the fragile-X syndrome. *American Journal of Human Genetics* **66**:6–15. DOI: <https://doi.org/10.1086/302720>, PMID: 10631132
- Tassone F**, Iwahashi C, Hagerman PJ. 2004. Fmr1 RNA within the intranuclear inclusions of fragile X-associated tremor/ataxia syndrome (FXTAS). *RNA Biology* **1**:103–105. DOI: <https://doi.org/10.4161/rna.1.2.1035>, PMID: 17179750
- Tassone F**, De Rubeis S, Carosi C, La Fata G, Serpa G, Raske C, Willemsen R, Hagerman PJ, Bagni C. 2011. Differential usage of transcriptional start sites and polyadenylation sites in FMR1 premutation alleles. *Nucleic Acids Research* **39**:6172–6185. DOI: <https://doi.org/10.1093/nar/gkr100>, PMID: 21478165
- Thomas JD**, Sznajder ŁJ, Bardhi O, Aslam FN, Anastasiadis ZP, Scotti MM, Nishino I, Nakamori M, Wang ET, Swanson MS. 2017. Disrupted prenatal RNA processing and myogenesis in congenital myotonic dystrophy. *Genes & Development* **31**:1122–1133. DOI: <https://doi.org/10.1101/gad.300590.117>, PMID: 28698297
- Thomson JP**, Skene PJ, Selfridge J, Clouaire T, Guy J, Webb S, Kerr ARW, Deaton A, Andrews R, James KD, Turner DJ, Illingworth R, Bird A. 2010. CpG islands influence chromatin structure via the CpG-binding protein cfp1. *Nature* **464**:1082–1086. DOI: <https://doi.org/10.1038/nature08924>, PMID: 20393567
- Tian Y**, Wang J-L, Huang W, Zeng S, Jiao B, Liu Z, Chen Z, Li Y, Wang Y, Min H-X, Wang X-J, You Y, Zhang R-X, Chen X-Y, Yi F, Zhou Y-F, Long H-Y, Zhou C-J, Hou X, Wang J-P, et al. 2019. Expansion of human-specific GGC repeat in neuronal intranuclear inclusion disease-related disorders. *American Journal of Human Genetics* **105**:166–176. DOI: <https://doi.org/10.1016/j.ajhg.2019.05.013>, PMID: 31178126
- Todd PK**, Oh SY, Krans A, Pandey UB, Di Prospero NA, Min K-T, Taylor JP, Paulson HL. 2010. Histone deacetylases suppress CGG repeat-induced neurodegeneration via transcriptional silencing in models of fragile X tremor ataxia syndrome. *PLOS Genetics* **6**:e1001240. DOI: <https://doi.org/10.1371/journal.pgen.1001240>, PMID: 21170301
- Todd PK**, Oh SY, Krans A, He F, Sellier C, Frazer M, Renoux AJ, Chen K, Scaglione KM, Basrur V, Elenitoba-Johnson K, Vonsattel JP, Louis ED, Sutton MA, Taylor JP, Mills RE, Charlet-Berguerand N, Paulson HL. 2013. Cgg repeat-associated translation mediates neurodegeneration in fragile X tremor ataxia syndrome. *Neuron* **78**:440–455. DOI: <https://doi.org/10.1016/j.neuron.2013.03.026>, PMID: 23602499
- Toulouse A**, Au-Yeung F, Gaspar C, Roussel J, Dion P, Rouleau GA. 2005. Ribosomal frameshifting on MJD-1 transcripts with long CAG tracts. *Human Molecular Genetics* **14**:2649–2660. DOI: <https://doi.org/10.1093/hmg/ddi299>, PMID: 16087686
- Trost B**, Engchuan W, Nguyen CM, Thiruvahindrapuram B, Dolzhenko E, Backstrom I, Mirceta M, Mojarad BA, Yin Y, Dov A, Chandrakumar I, Prasolava T, Shum N, Hamdan O, Pellicchia G, Howe JL, Whitney J, Klee EW, Baheti S, Amaral DG, et al. 2020. Genome-Wide detection of tandem DNA repeats that are expanded in autism. *Nature* **586**:80–86. DOI: <https://doi.org/10.1038/s41586-020-2579-z>, PMID: 32717741
- Tsuchiya M**, Nan H, Koh K, Ichinose Y, Gao L, Shimozone K, Hata T, Kim Y-J, Ohtsuka T, Cortese A, Takiyama Y. 2020. Rfc1 repeat expansion in Japanese patients with late-onset cerebellar ataxia. *Journal of Human Genetics* **65**:1143–1147. DOI: <https://doi.org/10.1038/s10038-020-0807-x>, PMID: 32694621
- Tu M**, Tong W, Perkins R, Valentine CR. 2000. Predicted changes in pre-mRNA secondary structure vary in their association with exon skipping for mutations in exons 2, 4, and 8 of the HPRT gene and exon 51 of the fibrillin gene. *Mutation Research/Mutation Research Genomics* **432**:15–32. DOI: [https://doi.org/10.1016/S1383-5726\(99\)00011-4](https://doi.org/10.1016/S1383-5726(99)00011-4), PMID: 10729708
- Tuller T**, Vekslar-Lublinsky I, Gazit N, Kupiec M, Ruppin E, Ziv-Ukelson M. 2011. Composite effects of gene determinants on the translation speed and density of ribosomes. *Genome Biology* **12**:R110. DOI: <https://doi.org/10.1186/gb-2011-12-11-r110>
- Udd B**, Krahe R. 2012. The myotonic dystrophies: molecular, clinical, and therapeutic challenges. *The Lancet. Neurology* **11**:891–905. DOI: [https://doi.org/10.1016/S1474-4422\(12\)70204-1](https://doi.org/10.1016/S1474-4422(12)70204-1), PMID: 22995693
- Uesaka M**, Nishimura O, Go Y, Nakashima K, Agata K, Imamura T. 2014. Bidirectional promoters are the major source of gene activation-associated non-coding RNAs in mammals. *BMC Genomics* **15**:35. DOI: <https://doi.org/10.1186/1471-2164-15-35>, PMID: 24438357
- Uffelmann E**, Huang QQ, Munung NS, de Vries J, Okada Y, Martin AR, Martin HC, Lappalainen T, Posthuma D. 2021. Genome-wide association studies. *Nature Reviews Methods Primers* **1**:1–21. DOI: <https://doi.org/10.1038/s43586-021-00056-9>
- Usdin K**, Woodford KJ. 1995. Cgg repeats associated with DNA instability and chromosome fragility form structures that block DNA synthesis in vitro. *Nucleic Acids Research* **23**:4202–4209. DOI: <https://doi.org/10.1093/nar/23.20.4202>, PMID: 7479085
- Usdin K**. 2008. The biological effects of simple tandem repeats: lessons from the repeat expansion diseases. *Genome Research* **18**:1011–1019. DOI: <https://doi.org/10.1101/gr.070409.107>, PMID: 18593815
- Usdin K**, Kumari D. 2015. Repeat-Mediated epigenetic dysregulation of the FMR1 gene in the fragile X-related disorders. *Frontiers in Genetics* **6**:192. DOI: <https://doi.org/10.3389/fgene.2015.00192>, PMID: 26089834
- Verkerk AJ**, Pieretti M, Sutcliffe JS, Fu YH, Kuhl DP, Pizzuti A, Reiner O, Richards S, Victoria MF, Zhang FP. 1991. Identification of a gene (FMR-1) containing a CGG repeat coincident with a breakpoint cluster region exhibiting length variation in fragile X syndrome. *Cell* **65**:905–914. DOI: [https://doi.org/10.1016/0092-8674\(91\)90397-h](https://doi.org/10.1016/0092-8674(91)90397-h), PMID: 1710175

- Verma AK**, Khan E, Mishra SK, Jain N, Kumar A. 2019. Piperine modulates protein mediated toxicity in fragile X-associated tremor/ataxia syndrome through interacting expanded CGG repeat (R (CGG) EXP) RNA. *ACS Chemical Neuroscience* **10**:3778–3788. DOI: <https://doi.org/10.1021/acschemneuro.9b00282>, PMID: 31264835
- Verma AK**, Khan E, Mishra SK, Mishra A, Charlet-Berguerand N, Kumar A. 2020. Curcumin regulates the R (CGG) EXP RNA hairpin structure and ameliorate defects in fragile X-associated tremor/ataxia syndrome. *Frontiers in Neuroscience* **14**:295. DOI: <https://doi.org/10.3389/fnins.2020.00295>, PMID: 32317919
- Verma AK**, Khan E, Mishra SK, Kumar A. 2022. Small molecule screening discovers compounds that reduce fmrpolyg protein aggregates and splicing defect toxicity in fragile X-associated tremor/ataxia syndrome. *Molecular Neurobiology* **59**:1992–2007. DOI: <https://doi.org/10.1007/s12035-021-02697-z>, PMID: 35040038
- Verstrepen KJ**, Jansen A, Lewitter F, Fink GR. 2005. Intragenic tandem repeats generate functional variability. *Nature Genetics* **37**:986–990. DOI: <https://doi.org/10.1038/ng1618>, PMID: 16086015
- Vinces MD**, Legendre M, Caldara M, Hagihara M, Verstrepen KJ. 2009. Unstable tandem repeats in promoters confer transcriptional evolvability. *Science* **324**:1213–1216. DOI: <https://doi.org/10.1126/science.1170097>, PMID: 19478187
- Volle CB**, Delaney S. 2012. Cag/Ctg repeats alter the affinity for the histone core and the positioning of DNA in the nucleosome. *Biochemistry* **51**:9814–9825. DOI: <https://doi.org/10.1021/bi301416v>, PMID: 23157165
- Voynov V**, Verstrepen KJ, Jansen A, Runner VM, Buratowski S, Fink GR. 2006. Genes with internal repeats require the THO complex for transcription. *PNAS* **103**:14423–14428. DOI: <https://doi.org/10.1073/pnas.0606546103>, PMID: 16983072
- Wang YH**, Gellibolian R, Shimizu M, Wells RD, Griffith J. 1996. Long CCG triplet repeat blocks exclude nucleosomes: a possible mechanism for the nature of fragile sites in chromosomes. *Journal of Molecular Biology* **263**:511–516. DOI: <https://doi.org/10.1006/jmbi.1996.0593>, PMID: 8918933
- Wang YH**. 2007. Chromatin structure of repeating CTG/CAG and CGG/CCG sequences in human disease. *Frontiers in Bioscience* **12**:4731–4741. DOI: <https://doi.org/10.2741/2422>, PMID: 17485409
- Wang ET**, Cody NAL, Jog S, Biancolella M, Wang TT, Treacy DJ, Luo S, Schroth GP, Housman DE, Reddy S, Lécuyer E, Burge CB. 2012. Transcriptome-Wide regulation of pre-mRNA splicing and mRNA localization by muscleblind proteins. *Cell* **150**:710–724. DOI: <https://doi.org/10.1016/j.cell.2012.06.041>, PMID: 22901804
- Weinberg DE**, Shah P, Eichhorn SW, Hussmann JA, Plotkin JB, Bartel DP. 2016. Improved ribosome-footprint and mrna measurements provide insights into dynamics and regulation of yeast translation. *Cell Reports* **14**:1787–1799. DOI: <https://doi.org/10.1016/j.celrep.2016.01.043>, PMID: 26876183
- Willems T**, Gymrek M, Highnam G, Mittelman D, Erlich Y, 1000 Genomes Project Consortium. 2014. The landscape of human STR variation. *Genome Research* **24**:1894–1904. DOI: <https://doi.org/10.1101/gr.177774.114>, PMID: 25135957
- Willems T**, Gymrek M, Poznik GD, Tyler-Smith C, 1000 Genomes Project Chromosome Y Group, Erlich Y. 2016. Population-Scale sequencing data enable precise estimates of Y-STR mutation rates. *American Journal of Human Genetics* **98**:919–933. DOI: <https://doi.org/10.1016/j.ajhg.2016.04.001>, PMID: 27126583
- Willemsen R**, Bontekoe CJM, Severijnen LA, Oostra BA. 2002. Timing of the absence of FMR1 expression in full mutation chorionic villi. *Human Genetics* **110**:601–605. DOI: <https://doi.org/10.1007/s00439-002-0723-5>, PMID: 12107447
- Winchester CL**, Ferrier RK, Sermoni A, Clark BJ, Johnson KJ. 1999. Characterization of the expression of DMPK and SIX5 in the human eye and implications for pathogenesis in myotonic dystrophy. *Human Molecular Genetics* **8**:481–492. DOI: <https://doi.org/10.1093/hmg/8.3.481>, PMID: 9949207
- Wojciechowska M**, Olejniczak M, Galka-Marciniak P, Jazurek M, Krzyzosiak WJ. 2014. Ran translation and frameshifting as translational challenges at simple repeats of human neurodegenerative disorders. *Nucleic Acids Research* **42**:11849–11864. DOI: <https://doi.org/10.1093/nar/gku794>, PMID: 25217582
- Wright SE**, Rodriguez CM, Monroe J, Xing J, Krans A, Flores BN, Barsur V, Ivanova MI, Koutmou KS, Barmada SJ, Todd PK. 2022. Cgg repeats trigger translational frameshifts that generate aggregation-prone chimeric proteins. *Nucleic Acids Research* **50**:8674–8689. DOI: <https://doi.org/10.1093/nar/gkac626>, PMID: 35904811
- Wyner N**, Barash M, McNevin D. 2020. Forensic autosomal short tandem repeats and their potential association with phenotype. *Frontiers in Genetics* **11**:884. DOI: <https://doi.org/10.3389/fgene.2020.00884>, PMID: 32849844
- Xi Z**, Zinman L, Moreno D, Schymick J, Liang Y, Sato C, Zheng Y, Ghani M, Dib S, Keith J, Robertson J, Rogava E. 2013. Hypermethylation of the CpG island near the G4C2 repeat in ALS with a C9orf72 expansion. *American Journal of Human Genetics* **92**:981–989. DOI: <https://doi.org/10.1016/j.ajhg.2013.04.017>, PMID: 23731538
- Yamamoto A**, Lucas JJ, Hen R. 2000. Reversal of neuropathology and motor dysfunction in a conditional model of Huntington's disease. *Cell* **101**:57–66. DOI: [https://doi.org/10.1016/S0092-8674\(00\)80623-6](https://doi.org/10.1016/S0092-8674(00)80623-6), PMID: 10778856
- Yap K**, Mukhina S, Zhang G, Tan JSC, Ong HS, Makeyev EV. 2018. A short tandem repeat-enriched RNA assembles a nuclear compartment to control alternative splicing and promote cell survival. *Molecular Cell* **72**:525–540. DOI: <https://doi.org/10.1016/j.molcel.2018.08.041>, PMID: 30318443
- Yin Q**, Wang H, Li N, Ding Y, Xie Z, Jin L, Li Y, Wang Q, Liu X, Xu L, Li Q, Ma Y, Cheng Y, Wang K, Zhong C, Yu Q, Tang W, Chen W, Yang W, Zhang F, et al. 2020. Dosage effect of multiple genes accounts for multisystem disorder of myotonic dystrophy type 1. *Cell Research* **30**:133–145. DOI: <https://doi.org/10.1038/s41422-019-0264-2>, PMID: 31853004

- Yu S**, Pritchard M, Kremer E, Lynch M, Nancarrow J, Baker E, Holman K, Mulley JC, Warren ST, Schlessinger D. 1991. Fragile X genotype characterized by an unstable region of DNA. *Science* **252**:1179–1181. DOI: <https://doi.org/10.1126/science.252.5009.1179>, PMID: 2031189
- Zhang K**, Donnelly CJ, Haeusler AR, Grima JC, Machamer JB, Steinwald P, Daley EL, Miller SJ, Cunningham KM, Vidensky S, Gupta S, Thomas MA, Hong I, Chiu S-L, Hagan RL, Ostrow LW, Matunis MJ, Wang J, Sattler R, Lloyd TE, et al. 2015. The C9orf72 repeat expansion disrupts nucleocytoplasmic transport. *Nature* **525**:56–61. DOI: <https://doi.org/10.1038/nature14973>, PMID: 26308891
- Zhang Y-J**, Gendron TF, Grima JC, Sasaguri H, Jansen-West K, Xu Y-F, Katzman RB, Gass J, Murray ME, Shinohara M, Lin W-L, Garrett A, Stankowski JN, Daugherty L, Tong J, Perkerson EA, Yue M, Chew J, Castanedes-Casey M, Kurti A, et al. 2016. C9orf72 poly (GA) aggregates sequester and impair HR23 and nucleocytoplasmic transport proteins. *Nature Neuroscience* **19**:668–677. DOI: <https://doi.org/10.1038/nn.4272>, PMID: 26998601
- Zu T**, Gibbens B, Doty NS, Gomes-Pereira M, Huguet A, Stone MD, Margolis J, Peterson M, Markowski TW, Ingram MAC, Nan Z, Forster C, Low WC, Schoser B, Somia NV, Clark HB, Schmechel S, Bitterman PB, Gourdon G, Swanson MS, et al. 2011. Non-ATG-initiated translation directed by microsatellite expansions. *PNAS* **108**:260–265. DOI: <https://doi.org/10.1073/pnas.1013343108>, PMID: 21173221
- Zu T**, Liu Y, Bañez-Coronel M, Reid T, Pletnikova O, Lewis J, Miller TM, Harms MB, Falchook AE, Subramony SH, Ostrow LW, Rothstein JD, Troncoso JC, Ranum LPW. 2013. Ran proteins and RNA foci from antisense transcripts in c9orf72 ALS and frontotemporal dementia. *PNAS* **110**:E4968–E4977. DOI: <https://doi.org/10.1073/pnas.1315438110>, PMID: 24248382
- Zu T**, Cleary JD, Liu Y, Bañez-Coronel M, Bubenik JL, Ayhan F, Ashizawa T, Xia G, Clark HB, Yachnis AT, Swanson MS, Ranum LPW. 2017. Ran translation regulated by muscleblind proteins in myotonic dystrophy type 2. *Neuron* **95**:1292–1305. DOI: <https://doi.org/10.1016/j.neuron.2017.08.039>, PMID: 28910618