Myotonic dystrophy (DM) is an autosomal dominant neuromuscular disease primarily characterized by myotonia and progressive muscle weakness. The pathogenesis of DM involves microsatellite expansions in noncoding regions of transcripts that result in toxic RNA gain-of-function. Each successive generation of DM families carries larger repeat expansions, leading to an earlier age of onset with increasing disease severity. At present, diagnosis of DM is challenging and requires special genetic testing to account for somatic mosaicism and meiotic instability. While progress in genetic testing has been made, more rapid, accurate, and cost-effective approaches for measuring repeat lengths are needed to establish clear correlations between repeat size and disease phenotypes.

**Addresses**

1 Department of Biochemistry, University of Illinois, Urbana-Champaign, USA
2 Institute of Genomic Biology, University of Illinois, Urbana-Champaign, USA
3 Department of Molecular Genetics & Microbiology, Center for Neurogenetics, University of Florida, Gainesville, FL 32610, USA

Corresponding authors: Wang, Eric T (eric.t.wang@ufl.edu), Kalsotra, Auinash (kalsotra@illinois.edu)

**Current Opinion in Genetics & Development** 2017, 44:30–37

This review comes from a themed issue on **Molecular and genetic bases of disease**

Edited by Nancy Bonini, Edward Lee and Wilma Wasco

http://dx.doi.org/10.1016/j.gde.2017.01.007

0959-437X/© 2017 Elsevier Ltd. All rights reserved.

**Introduction**

DNA repeat expansions are responsible for more than 20 inherited neurological disorders—some of these include Huntington’s disease, fragile X syndrome, spinal and bulbar muscular atrophy, as well as the most common form of familial amyotrophic lateral sclerosis [1]. In multiple repeat diseases, repeat length is correlated to disease severity and age of onset [2], yet molecular pathways that go awry due to expanded repeats can differ. Studies of myotonic dystrophy (dystrophia myotonica, DM) first demonstrated the concept that microsatellite repeats in noncoding regions can be transcribed into pathogenic RNAs [3]. Expansions can occur in the germline, leading to genetic anticipation across multiple generations, and can also occur somatically during various stages of human development with preferences for distinct tissues, ages, genders, and populations [4–7]. Furthermore, the rate of expansion in somatic cells can vary within the same tissue [8].

**Overview of myotonic dystrophy**

Myotonic dystrophy exists in two clinically and molecularly defined forms: myotonic dystrophy type 1 (DM1), also known as Steinert’s disease; and myotonic dystrophy type 2 (DM2), also known as proximal myotonic myopathy, both of which are inherited in an autosomal dominant fashion [9]. DM1 is caused by a CTG expansion in the 3’ untranslated region of the dystrophia myotonica protein kinase (DMPK) gene on chromosome 19q13 [10,11], while DM2 is caused by a CCTG expansion located within intron 1 of the cellular nucleic-acid-binding protein (CNBP, formerly ZNF9) gene on chromosome 3q21 [12].

A healthy individual with normal DMPK alleles has 5–37 repeats (35 has also commonly been used as an upper threshold for normal repeat length [13]) [14]. DM1 patients who have repeats between 38 and 50 are said to have a “pre-mutation” allele and can be asymptomatic throughout their lifetime. However, they are at increased risk of having children with larger repeats [15]. Penetrance tends to grow as repeat length increases, but extreme variability in penetrance of specific symptoms exists in the patient population [14]. Somatic mosaicism and intergenerational instability are biased towards expansion in DM1 [4], although contraction can rarely occur. It is estimated that a decrease in the CTG repeat size during transmission from parents to child is about 6.4%, most frequently during paternal transmissions [16]. Children of DM1 parents typically inherit repeat lengths considerably larger than those present in the transmitting parent, the phenomenon known as “anticipation,” where disease severity increases and age of onset decreases in successive generations. Up to 5% of DM1 patients have interrupted repeats, in which the CTG repeat tract contains GGC, CCG, or CTC repeats [17,18]. Some of these interruptions have been associated with stabilization of the CTG repeat tract length [19].
The repeat expansion of DM2 in intron 1 of CNBP is found within the context of a complex (TG)\textsubscript{n}(TCTG)\textsubscript{n}(CCTG)\textsubscript{n} sequence. While non-pathogenic alleles contain up to 26 repeats, the range of repeats in patients is extremely broad, with measurements from 75 to 11 000 units (on average 5000) [12]. Unlike DM1, the size of the repeat DNA expansion in DM2 does not correlate with age of onset or disease severity [20]. Furthermore, individuals homozygous for repeat expansions have clinical features indistinguishable from that of their heterozygous siblings [21]. Phenotypes and anticipation in DM2 are almost always milder than DM1, and DM2 lacks the congenital form [22].

The combined prevalence of DM1 and DM2 is approximately 1 in 8000 (12.5 per 100 000), but this is likely an underestimate because of difficulty in clinical identification of minimally affected individuals [7]. Although DM2 is generally rarer than DM1, recent epidemiological data in Germany and Finland suggest that DM2 occurs more frequently than previously observed [23]. Similarly, the prevalence of DM1 can vary widely: in Taiwan, approximately 0.5 in 100 000 people are affected; while in the United Kingdom, the number can range from 7.1 to 10.6 in 100 000 [24]. Different factors could play a role in such variations: for instance, a founder effect is assumed to have increased the prevalence of DM1 to 1 in 500 in the Saguenay–Lac-Saint-Jean region of northeastern Quebec [25].

Despite these key differences, DM1 and DM2 share several hallmark clinical features such as myotonia, cataracts, and cardiac conduction defects [26]. The fact that two independent mutations cause similar disease pathology has led to the RNA toxicity hypothesis where the expanded repeat-containing RNAs form ribonucleic foci that sequester and disrupt the normal activities of RNA binding proteins belonging to the MBNL and CELF families [27] (for more details, see Ashizawa’s review on “RNA foci” and Thornton’s review on “DM: approaches to therapy” in the same issue, as well as Ref. [28]). In this review, we discuss mechanisms of repeat expansion, approaches for measuring repeat lengths, and the relationships between repeat length and phenotypes in DM.

Mechanisms of repeat expansion in mitotic and post-mitotic tissues

Several molecular mechanisms for repeat instability have been proposed, mainly in the context of DNA replication, recombination, transcription and/or repair (Box 1). Most of these mechanisms involve folding of microsatellite repeats into an unusual secondary structure, kinetically trapping the otherwise unstable DNA repeats [29]. In the case of DM, (CTG)\textsubscript{n}(CAG)\textsubscript{n} and (CCTG)\textsubscript{n}(CAGG)\textsubscript{n} repeats form hairpin-like secondary structures, which are stabilized by both Watson–Crick (WC) and non-WC base pairs [30,31].

Proof of principle studies from yeast have demonstrated that repeat instability can be based on replication fork stalling and restart [32,33], ruling out the classic model of strand slippage for DM [34]. In this model, the formation of a stable secondary structure during lagging strand synthesis could stall a DNA polymerase, slowing down the overall replication fork progression as the lagging and leading strand syntheses are coordinated. To minimize the stalling of replication fork, DNA polymerase can skip an Okazaki fragment to resume lagging strand synthesis (contraction pathway) or promote fork reversal (for more details on the fork reversal mechanism, please see the review by Neelsen and Lopes [35]) to generate a structure-prone single-stranded repeat extension at the 3’ end of the leading strand. If the 3’ repetitive hairpin persists when the replication restarts, repeats can expand. Recombination can also account for repeat instability in mitotically dividing cells. In bacteria, longer repeats increase the rate of recombination [36,37] while in yeast, CTG repeats cause chromosomal breakage [38]. As Mirkin described in his review [39], one possible mechanism for recombination-based instability is that repeats promote the double-strand breaks in DNA causing the invasion of fragments into sister chromatids.

The transcription and repair models of repeat instability can likewise account for expansion in both mitotic and post-mitotic cells. During transcription, the formation of slipp-outs on either strand can stall RNA polymerase II, facilitating the transcription-coupled repair. Depending on the location of the excision, subsequent patch repair could lead to expansion or contraction [40]. Similarly, studies on transgenic mouse models of Huntington’s disease and DM have demonstrated that loss of MSH2/MSH3 mismatch repair proteins can decrease the frequency of repeat expansion [41–43]. This discovery has led to a theory that the MSH2/MSH3 complex can stabilize the secondary structure and prevent the flip removal by FEN1, leading to expansion during DNA repair. Although this theory is highly supported in a yeast model [44], it is less clear in a mouse model where repeat instability was unaltered in FEN1-knockout mice [45].

As disease symptoms in DM are most prominent in post-mitotic tissues such as the heart, skeletal muscle, and central nervous system (CNS), it is thought that DNA repair-dependent mechanisms, and potentially transcription-coupled nucleotide excision repair, may drive repeat instability in these tissues. For more details on tissue-specific DNA repair mechanisms of repeat instability, see the review by Dion [46].

Diagnosis and laboratory methods to measure repeat lengths

A wide variety of DM symptoms can bring patients to the clinic, including myotonia, muscle weakness, cardiac arrhythmias, hypersomnia, gastrointestinal (GI) tract
Box 1 Mechanisms of repeat expansion and contraction.

Hairpin-like structures are formed during an out-of-register realignment of the complementary repetitive strands during DNA replication, recombination, transcription, and/or repair. Flanking DNA is shown in black. Parental repeat and complementary strands are shown in red and orange, respectively. The newly synthesized repeat strand is shown in blue. (a) A model of repeat instability generated during replication fork stalling and restart. Formation of a hairpin in the lagging-strand promotes the stalling of a DNA polymerase. To overcome this obstacle, fork reversal promotes unwinding of newly synthesized strands (as well as reannealing of parental strands), exposing a structure-prone 3’ repetitive run. A hairpin structure is formed and retained during leading strand synthesis, which leads to expansion. Repeat contraction can occur if the DNA polymerase skips the hairpin on the lagging-strand template. (b) Recombination model of repeat instability. Cleavage of stable DNA hairpin during replication generates single-stranded 3’ repetitive fragments. Invasion of these fragments with or without repeats can lead to expansion or contraction via recombination. (c) Transcription model of repeat instability. Unwinding of repetitive DNA strands during transcription can generate template strand (TS) or non-template strand (NTS) slip-outs. RNA Polymerase II (RNAPII) can stall at either slip-out and initiate the transcription-coupled repair (TCR). If RNAPII stalls at the NTS hairpin, TCR cuts the portion of TS and copies the slip-out during repair. Conversely, RNAPII can stall at the distal end of the TS slip-out where TCR can excise near the 5’ end of the NTS slip-out, subsequently removing the repeats by 3’-end cleavage. (d) A DNA-repair-dependent repeat expansion model. DNA damage induces a small gap in the structure-prone repetitive strand. Hairpin formation prevents FEN1 endonuclease from loading onto a repetitive flap. The binding of mismatch repair proteins, MSH2 and MSH3, stabilizes the hairpin leading to expansion upon completion of repair.

issues, and cataracts. Before the discovery of the genetic basis for DM, muscle biopsies were commonly used as a diagnostic tool, but typically, a definitive diagnosis is now made via genetic testing (Figure 1). If clinical features suggest DM1 but genetic testing is negative, DM2 testing is performed.

A number of approaches are taken to measure repeat lengths. A PCR across the DM1 locus is usually first conducted to determine whether there are two different short alleles, or a short allele and a longer allele of <150 repeats. Extremely long repeats are challenging to amplify by PCR. Therefore, the presence of a single short PCR product does not rule out DM1 [11]. In such cases, other methods such as triplet-repeat primed PCR (TP-PCR), small-pool PCR (SP-PCR) and/or Southern blots, on either PCR products or restriction-digested genomic DNA, are performed.

**Triplet-repeat primed PCR**

Warner et al. [47] introduced a robust and reliable method to identify (but not to measure the size of) an expanded allele using TP-PCR. This technique is typically cheaper and faster than Southern blotting for diagnostic purposes. However, TP-PCR can cause up to 9% false positive results due to sequence interruptions between repeats [48,49]. More recently, an improved bi-directionally labeled TP-PCR was developed by Radavsky et al. [50], in which TP-PCR is performed in the reverse direction with two individual fluorescently-labeled flanking primers. This method can detect expansions carrying sequence interruptions both in DM1 and DM2. In the
Schematic of genetic testing used to diagnose DM.
DM is a multisystemic disorder that primarily affects skeletal muscle, heart, brain, eyes, CNS, and GI tract. Electromyography (EMG) is used to assess myotonia in suspected individuals. Genetic testing is performed using the patient’s DNA sample, typically from the blood. There are two distinct laboratory procedures for genetic testing: SP-PCR and TP-PCR. SP-PCR requires single-molecule dilutions of extracted DNA, which are individually PCR amplified and probed with Southern blot hybridization. This method readily amplifies repeats from individual molecules, avoiding the amplification bias for the smaller allele commonly observed in conventional PCR. Similarly, TP-PCR is an improved method for detecting larger repeats. During the early amplification cycles, 5’ fluorescently labeled primer P1 and repeat-specific primer P3 with 5’ tail sequence (in yellow) generate multiple products. The primer P4, which shares the 5’ tail sequence, subsequently amplifies the products from the previous amplification cycles. A 10:1 ratio of P4 to P3 ensures that P3 is exhausted in the early amplification cycles. Flanking and repeat DNA are shown in black and red, respectively.

In the context of DM2, quadruplet-repeat primed PCR is also commonly performed to size the CCTG repeat tract [20].

Small-pool PCR
Conventional PCR and Southern blot approaches typically show expanded alleles as a smear rather than a discrete band on a gel, caused by somatic instability [51]. SP-PCR, in which small amounts of input DNA up to 2 genomic equivalents are separately PCR-amplified and detected by Southern blotting, showed that these smears are indeed CTG tracts of discrete lengths from different nuclei [4]. This method allows for assay of the entire repeat length distribution, as well as estimation of the progenitor allele length [52].

Finally, an approach combining rolling circle amplification and Southern blotting to identify expanded repeats in DM1 and DM2 with low DNA input requirements [53] and a method which utilizes a special PCR enzyme mix and machine to screen normal DMPK alleles in less than 15 min have been described [54*].

Repeat range, penetrance, age of onset, and relationship to phenotypes
DM1 patients can be subdivided into several categories, based on clinical features, age of onset, and disease severity. As larger patient cohorts are studied and more precise methods applied to measure clinical symptoms, the granularity of these categories has increased. Traditionally, DM1 has been divided into late-onset, adult, and congenital forms. However, in a recent study by De Antonio et al. [55**], separate categories of congenital, infantile, juvenile, adult and late-onset forms of DM1 are described, with respective mean repeat lengths of ~1000,
800, 600, 400, and 200 according to registry records (Table 1). Across all these forms, symptoms span a broad range of tissues, including skeletal muscle, heart, CNS, as well as the GI tract, eyes, reproductive tract, endocrine system, and immune system. Although in general many symptoms occur earlier when repeats are longer, there is extreme variation across individuals. For example, while many adult patients first present myotonia, the first symptom affecting others with DM1 may be hypersomnolence, subcapsular/iridescent “Christmas tree” cataracts, or atrial fibrillation.

Table 1

<table>
<thead>
<tr>
<th>Phenotype</th>
<th>Most prominent clinical symptoms</th>
<th>Repeat range</th>
<th>Age of onset</th>
<th>Life span</th>
<th>Penetration and anticipation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pre-mutation</td>
<td>None</td>
<td>38–49</td>
<td>–</td>
<td>Normal</td>
<td>Increased risk of penetrance</td>
</tr>
<tr>
<td>Late-onset DM1</td>
<td>Cataracts, hypersomnia, myotonia</td>
<td>100–600</td>
<td>&gt;40 years</td>
<td>Normal</td>
<td>Uncertain</td>
</tr>
<tr>
<td>Adult DM1</td>
<td>Myotonia, cardiac arrhythmias, hypersomnia, gastrointestinal difficulties, muscle weakness and wasting, cataracts, male hypogonadism, insulin resistance, cognitive challenges, left ventricular dysfunctions</td>
<td>250–750</td>
<td>20–40 years</td>
<td>Shortened</td>
<td>Full penetrance</td>
</tr>
<tr>
<td>Juvenile DM1</td>
<td>Similar symptoms as adult DM1 but more severe</td>
<td>400–800</td>
<td>10–20 years</td>
<td>Shortened</td>
<td>Full penetrance</td>
</tr>
<tr>
<td>Infantile DM1</td>
<td>Similar symptoms as congenital DM1 but less severe</td>
<td>500–1100</td>
<td>1 mo. – 10 years</td>
<td>Shortened</td>
<td>Full penetrance</td>
</tr>
<tr>
<td>Congenital DM1</td>
<td>Developmental defects, hypotonia, respiratory insufficiency, cardiac defects, severe cognitive challenges, facial dysmorphism, dysphagia</td>
<td>750–1400</td>
<td>Birth</td>
<td>Increased risk of infant mortality or shortened</td>
<td>Yes</td>
</tr>
<tr>
<td>DM2</td>
<td>Proximal muscle weakness and wasting, cognitive challenges, cardiac arrhythmias, myalgic pain, hypertension</td>
<td>100–10 000</td>
<td>Adult</td>
<td>Normal</td>
<td>Uncertain</td>
</tr>
</tbody>
</table>

There has been much discussion on whether to draw distinctions between the congenital, infantile, and juvenile forms of DM1; each is successively milder, with the congenital exhibiting profound developmental defects, including hypotonia, respiratory insufficiency, cardiac defects, severe muscle weakness, cognitive challenges, and facial dysmorphism. Congenitally affected fetuses are associated with excess amniotic fluid and decreased movement [56,57], and occur almost exclusively from maternal transmissions (although there are exceptions [58]). The infantile and juvenile forms of DM1 are less severe than congenital, but still exhibit many of the same features, in particular, cognitive challenges [59,60]. Severe myotonia is also much more prominent in the juvenile form [55].

Core features of adult and late-onset DM1 are cataracts, myotonia, GI problems, muscle weakness, and cardiac arrhythmias. Cardiac conduction defects with arrhythmias contribute to the shortened life span of adult DM1 patients [61,62]. Mild intellectual deficits can be found in both adult and late-onset patients [63*], but clinical depression and personality disorders are more common in adult DM1 [64,65]. Nocturnal apnoeic episodes and excessive daytime somnolence also have significant repercussions on the quality of adult DM1 patients’ lives [66,67].

Repeat lengths in DM2 have not been observed to correlate with specific phenotypes, such as proximal muscle weakness (often the quadriceps), cardiac arrhythmias, and cognitive decline, but this is likely in part caused by the relative paucity of studies focused on measuring these correlations. The repeat length in DM2 can be extremely long, presenting technical challenges.

Concluding remarks

As the molecular basis of DM continues to unfold, it is clear that RNA toxicity has a major role in disease pathology. Based on this framework, one would expect that repeat length and expression levels correlate with disease severity and age of onset. However, heterogeneity across tissues, somatic instability, and the relative technical difficulty of accurately measuring repeat length distributions present challenges in establishing these correlations, and importantly, identifying additional genetic modifiers that may protect or exacerbate particular disease symptoms. New sequencing methods, such as those that have been applied in the other repeat diseases (PacBio single-molecule, real-time sequencing for Fragile X [68**]), may allow for more rapid, cost-effective, and
DM repeats, age of onset, penetrance and phenotypic relationships
Yum, Wang and Kalsotra

accurate measurements of long repeat lengths at a single-nucleotide resolution, and enhance our overall understanding of expansion repeat diseases.

Conflict of interests statement
Nothing declared.

Acknowledgements
E.T.W. is supported by (NH) DP5 OD 017865. A.K. is supported by grants from the US National Institute of Health (R01HL126845), March of Dimes (5-FY14-112), and the Center for Advanced Study at the University of Illinois. The authors apologize to those whose work could not be cited because of lack of space.

References and recommended reading
Papers of particular interest, published within the period of review, have been highlighted as:

* of special interest
** of outstanding interest


35. Neelens KJ, Lopes M: Replication fork reversal in eukaryotes:

This review provides a detailed summary of in vivo mechanisms and the consequences of fork reversal during replication.


This review provides a detailed summary of tissue-specific DNA repair mechanisms responsible for trinucleotide repeat instability.


This article provides a rapid PCR screening method directly from patient’s blood samples to identify the presence of normal DMPK alleles, which can function as a negative DM1 screen if both alleles of normal size are clearly detected.


A large cohort of DM1 patients was studied from the DM-Scope registry and classified into five subtypes (including two new subtypes: infantile and juvenile) based on their distribution of CTG expansion size and age of major symptom onset.


This clinical study provides evidence that intellectual impairment in classic adult-onset DM1 patients are also common in mild late-onset DM1 patients characterized by low general intellectual functioning in both verbal and nonverbal abilities.


68. Loomis EW, Eid JS, Peluso P, Yin J, Hickey L, Rank D, McCalmon S, Hagerman RJ, Tassone F, Hagerman PJ: Sequencing the unsequenceable: expanded CGG-repeat alleles of the fragile X gene. Genome Res 2013, 23:121-128. This study demonstrates that the SMRT sequencing can generate real-time fluorescent pulse data to help sequence FMR1 alleles containing large CGG repeats.