

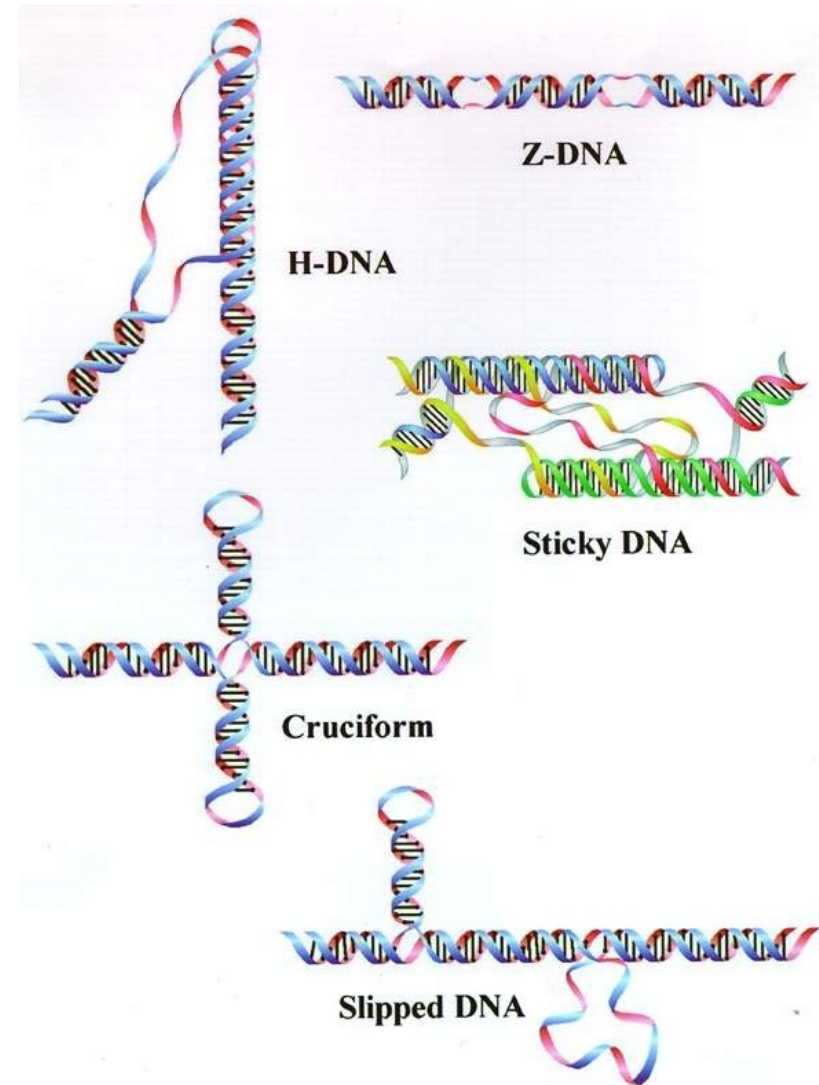
Table 1. The repeat expansion diseases

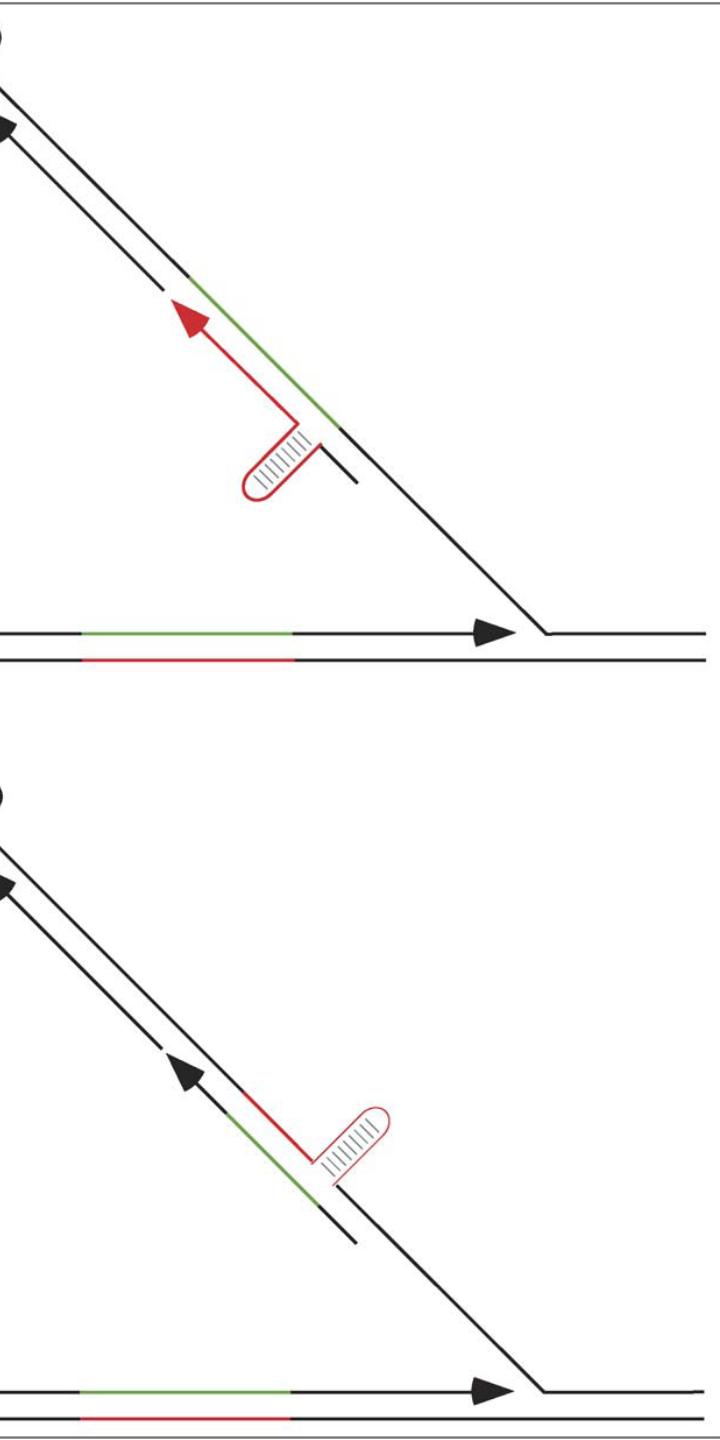
Disorder	Affected gene	Repeat ^a	Repeat location	Normal repeat no.	Symptomatic repeat no.	OMIM no.
Coding repeats						
DRPLA	<i>ATN1</i>	CAG	ORF	7–25	49–88	125370
HD	<i>HTT</i>	CAG	ORF	6–34	36–180	143100
SBMA	<i>AR</i>	CAG	ORF	11–24	40–62	313200
SCA1	<i>ATXN1</i>	CAG	ORF	6–39	39–83	164400
SCA2	<i>ATXN2</i>	CAG	ORF	15–29	34–59	183090
SCA3	<i>ATXN3</i>	CAG	ORF	13–36	55–84	109150
SCA6	<i>CACNA1A</i>	CAG	ORF	4–16	21–30	183086
SCA7	<i>ATXN7</i>	CAG	ORF	4–35	34–>300	164500
SCA17	<i>TBP</i>	CAG	ORF	25–44	45–66	607136
Noncoding repeats						
→ DM1	<i>DMPK</i>	CTG	3' UTR	5–37	>50–>2000 ←	160900
→ DM2	<i>CNBP</i>	CCTG	Intron 1	<27	75–11,000 ←	608768
→ EPM1	<i>CSTB</i>	(C) ₄ G(C) ₄ GCG	Promoter	2–3	30–75	254800
→ FXS	<i>FMR1</i>	CGG	5' UTR	6–52	~55–>2000 ←	309550
FRAXE MR	<i>AFF2/FMR3</i>	CCG	5' end	6–25	>200	309548
FRA12A MR	<i>DIP2B</i>	CGG	5' UTR	6–23	?	136630
FRDA	<i>FXN</i>	GAA	Intron 1	7–22	>66–>900	229300
SCA10	<i>ATXN10</i>	ATTCT	Intron 9	10–29	280–4500	603516
Coding and noncoding repeats						
SCA8	<i>ATXN8/ATXN8OS</i>	CAG and CTG	ORF and NCT	6–37	~107–250	603680
Repeats with uncertain location						
HDL-2	<i>JPH3</i>	CAG/CTG	?	<50	>50	606438
SCA12	<i>PPP2R2B</i>	CAG/CTG	?	<66	>66	604326

(OMIM) Online Mendelian Inheritance in Man database of human genes and genetic disorders (<http://www.ncbi.nlm.nih.gov/sites/entrez?db=omim>); (DM1) myotonic dystrophy type 1; (DM2) myotonic dystrophy type 2; (DRPLA) dentatorubral-pallidolusian atrophy; (EPM1) progressive myoclonus epilepsy type I; (FXS) Fragile X syndrome; (MR) mental retardation; (FRDA) Friedreich ataxia; (HD) Huntington disease; (HDL-2) Huntington disease-like 2; (NCT) noncoding transcript; (SBMA) spinobulbar muscular atrophy; (SCA) spinocerebellar ataxia.

^aThe strand shown corresponds to the top strand of the affected gene where it is known. The sequence of both strands is given where this is not known.

- Trinucleotide repeat sequences (TRS) - **dynamic mutations** - the number of repeats tends to increase in size over generations.
- Several factors contribute to mutational dynamics - **number of repeats, composition and length of the repeating motif, presence of interruptions within the sequence and the rate of intracellular processes such as replication, transcription, repair, or recombination.**
- **A significant feature of tandem repeats is ability to form unusual DNA structures (left-handed Z-DNA, cruciforms, slipped-stranded DNA, triplexes, and tetraplexes). Such non-B-DNA structures potentially may be hazardous for genome stability if not removed by repair mechanisms.**

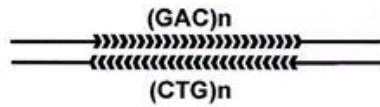




The simplest explanation of a repeat expansion – an slippage of DNA polymerase during DNA replication.

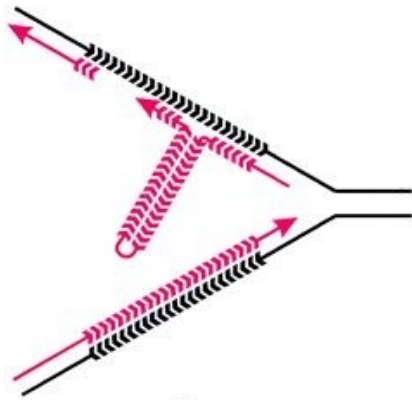
- Repeats form unusual DNA structures in ssDNA.
- The main cellular process involving DNA strand separation is DNA replication.
- **Lagging strand model of repeat expansion.** (a) Formation of the repetitive hairpin on the nascent lagging strand leads to expansions; (b) the same structure on the lagging strand template generates contractions.

ORIENTATION I



DNA synthesis

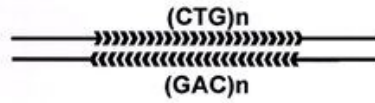
hairpin formation on the Okazaki fragment



polymerase slippage,
primer relocation,
idling synthesis

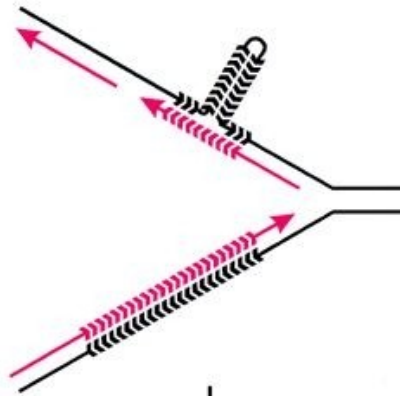
expansion

ORIENTATION II



DNA synthesis

hairpin formation in lagging template



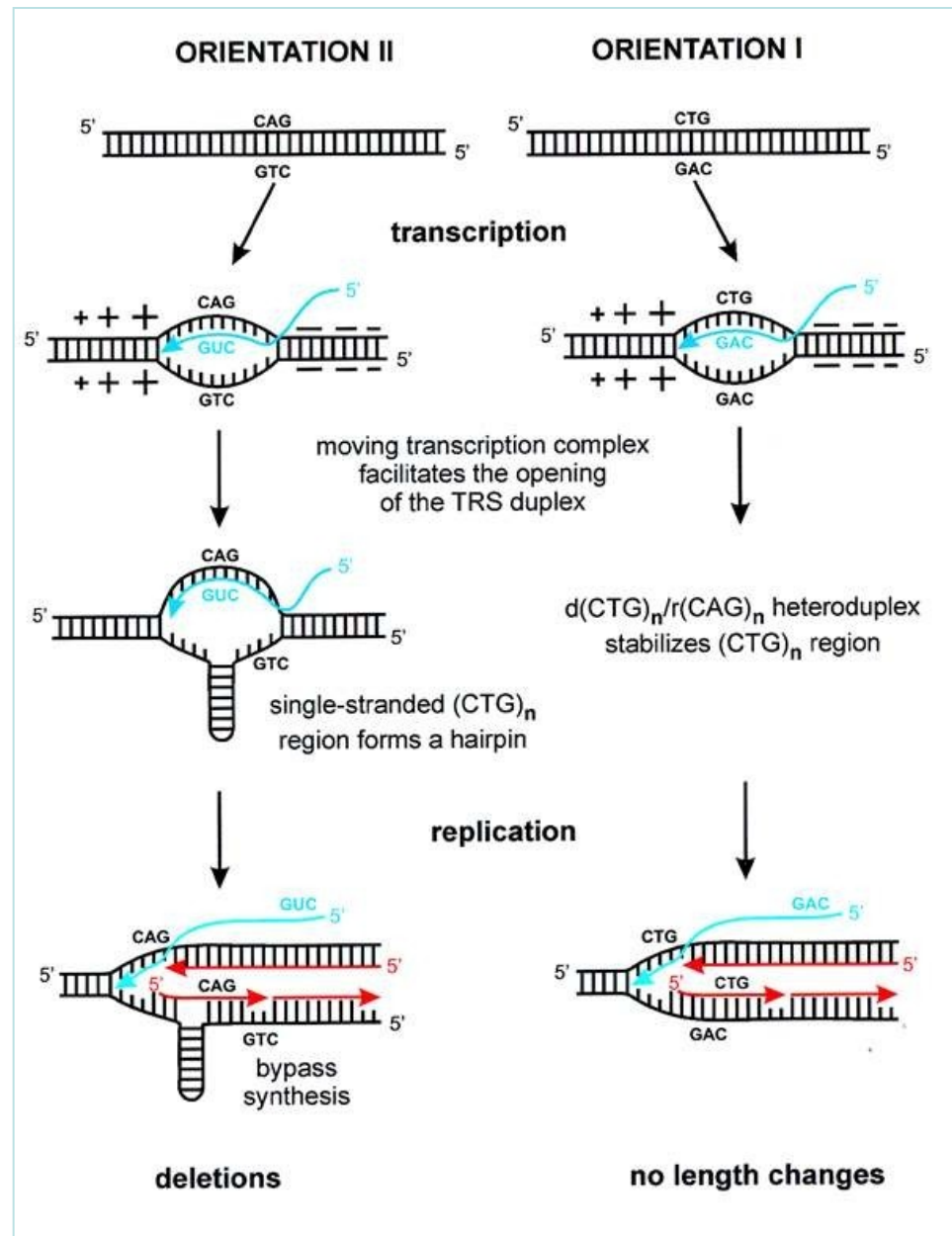
bypass
synthesis

deletion

Mechanism of the genetic instability of the TRS during replication.

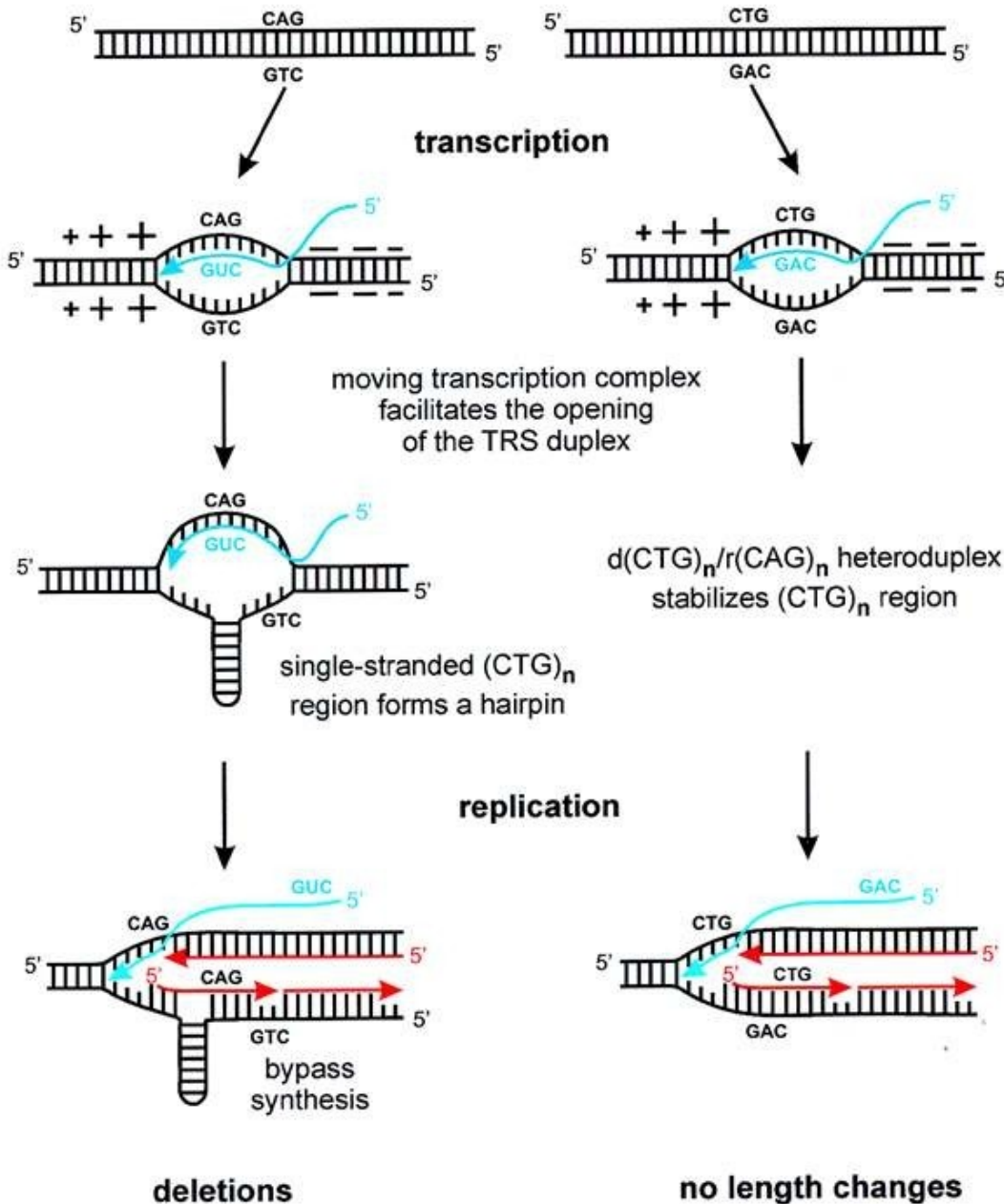
- The observed instability strongly depended on the orientation of the repeat relative to the origin of replication.
- Deletions occur if the hairpin is formed on the lagging strand template (orientation II). Expansions arise as a consequence of secondary DNA structures being formed during lagging strand synthesis (orientation I).
- Both, deletions and expansions may also happen during leading-strand synthesis but such events are substantially less frequent.

- TRS related to human diseases are actively transcribed.
- Unwinding of the double-stranded DNA by moving RNA polymerase complex introduces locally high torsional stress.
- Torsional stress leads to the formation of domains of differential DNA supercoiling.
- **It was shown that transcription could promote hairpin formation within repeating sequences and formation of such structures in TRS during transcription could lead to length changes of the repeat tract.**



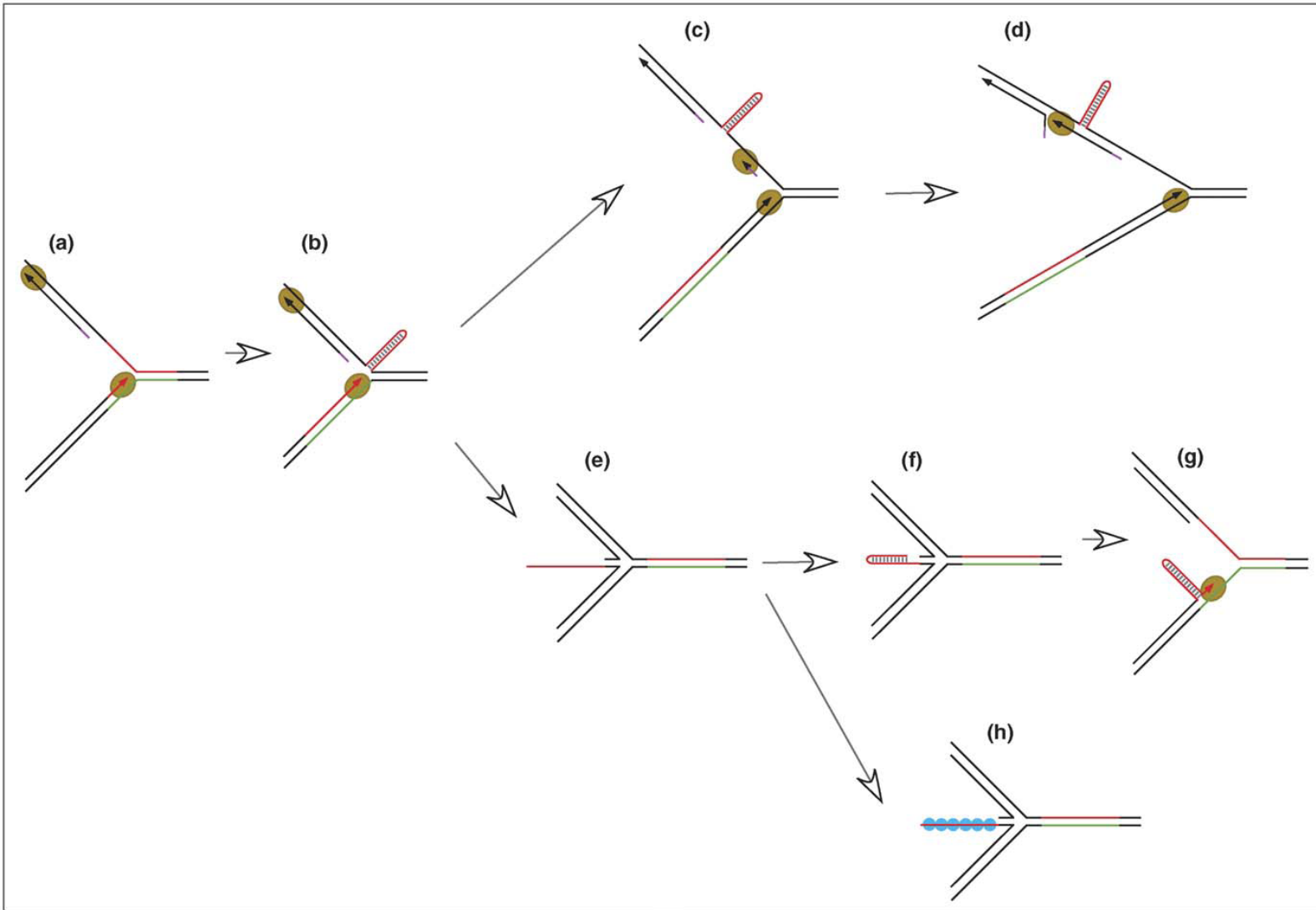
ORIENTATION II

ORIENTATION I



Correlation between replication, orientation of the repeat tract, and transcription.

- Orientation II (CTG strand serves as the lagging strand template), orientation I (CTG is within nascent lagging strand).
- Transcription of the CAG strand leads to deletions, transcription of the CTG strand elicits a much lower frequency of deletions.
- The model proposes that as the CAG strand is being transcribed, the complementary CTG strand while being single-stranded and forms a hairpin. On the other hand, the non-transcribed CAG strand in orientation I is less able to form stable hairpins. In orientation I the CTG strand is not single-stranded and cannot form stable hairpins because it is "occupied" by the RNA polymerase complex.
- The model further supposed that while TRS is transcribed, it is also replicated. In this case, the CTG hairpin in orientation II will be bypassed by the DNA polymerase complex during lagging-strand synthesis, and this will lead to deletions. Conversely, deletions in orientation I will be found rarely since there is a lower propensity to form secondary structures on the lagging-strand template by the CAG tracts and thus, no bypass synthesis occurs.



Model of repeat instability generated during replication fork stalling. (a) Entrance of the leading strand polymerase into the repetitive run. (b) Formation of unusual structure by the lagging strand template, stalling the lagging strand polymerase and the replication fork. (c) Replication continuation by skipping an Okazaki fragment. (d) Contraction of the repeat as the lagging strand polymerase skips the structure on its template. (e) Fork reversal generates a 'chicken foot' structure with a single-stranded repetitive 3' extension. (f) Folding of the repetitive extension into a hairpin-like conformation. (g) Replication restart upon flipping back the chicken foot, leading to repeat expansions. (h) Loading of the recombination proteins responsible for the strand exchange reaction onto the 3'-extension.

Model of repeat instability generated during replication fork stalling. (a) Entrance of the leading strand polymerase into the repetitive run. (b) Formation of unusual structure by the lagging strand template, stalling the lagging strand polymerase and, ultimately, the replication fork. (c) Replication continuation by skipping an Okazaki fragment. (d) Contraction of the repeat as the lagging strand polymerase skips the structure on its template. (e) Fork reversal generates a 'chicken foot' structure with a single-stranded repetitive 3' extension. (f) Folding of the repetitive extension into a hairpin-like conformation. (g) Replication restart upon flipping back the chicken foot, leading to repeat expansions. (h) Loading of the recombination proteins responsible for the strand exchange reaction onto the 3' -repetitive extension.

A structure-prone strand of the repeat is shown in red, its complementary strand is in green and flanking DNA is black. Golden ovals represent DNA polymerases, purple lines Okazaki primers and blue circles recombination proteins.

Current Opinion in Structural Biology 2006, 16:351–358

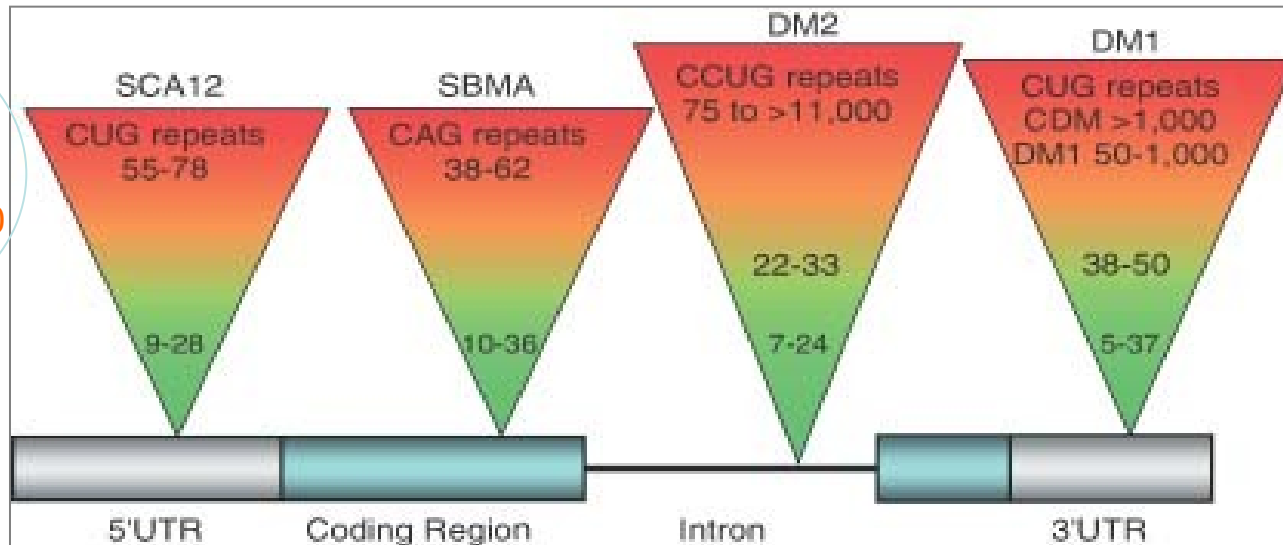
Table 1. The repeat expansion diseases

Disorder	Affected gene	Repeat ^a	Repeat location	Normal repeat no.	Symptomatic repeat no.	OMIM no.
Coding repeats						
DRPLA	<i>ATN1</i>	CAG	ORF	7–25	49–88	125370
HD	<i>HTT</i>	CAG	ORF	6–34	36–180	143100
SBMA	<i>AR</i>	CAG	ORF	11–24	40–62	313200
SCA1	<i>ATXN1</i>	CAG	ORF	6–39	39–83	164400
SCA2	<i>ATXN2</i>	CAG	ORF	15–29	34–59	183090
SCA3	<i>ATXN3</i>	CAG	ORF	13–36	55–84	109150
SCA6	<i>CACNA1A</i>	CAG	ORF	4–16	21–30	183086
SCA7	<i>ATXN7</i>	CAG	ORF	4–35	34–>300	164500
SCA17	<i>TBP</i>	CAG	ORF	25–44	45–66	607136
Noncoding repeats						
→ DM1	<i>DMPK</i>	CTG	3' UTR	5–37	>50–>2000 ←	160900
→ DM2	<i>CNBP</i>	CCTG	Intron 1	<27	75–11,000 ←	608768
→ EPM1	<i>CSTB</i>	(C) ₄ G(C) ₄ GCG	Promoter	2–3	30–75	254800
→ FXS	<i>FMR1</i>	CGG	5' UTR	6–52	~55–>2000 ←	309550
FRAXE MR	<i>AFF2/FMR3</i>	CCG	5' end	6–25	>200	309548
FRA12A MR	<i>DIP2B</i>	CGG	5' UTR	6–23	?	136630
FRDA	<i>FXN</i>	GAA	Intron 1	7–22	>66–>900	229300
SCA10	<i>ATXN10</i>	ATTCT	Intron 9	10–29	280–4500	603516
Coding and noncoding repeats						
SCA8	<i>ATXN8/ATXN8OS</i>	CAG and CTG	ORF and NCT	6–37	~107–250	603680
Repeats with uncertain location						
HDL-2	<i>JPH3</i>	CAG/CTG	?	<50	>50	606438
SCA12	<i>PPP2R2B</i>	CAG/CTG	?	<66	>66	604326

(OMIM) Online Mendelian Inheritance in Man database of human genes and genetic disorders (<http://www.ncbi.nlm.nih.gov/sites/entrez?db=omim>); (DM1) myotonic dystrophy type 1; (DM2) myotonic dystrophy type 2; (DRPLA) dentatorubral-pallidolusian atrophy; (EPM1) progressive myoclonus epilepsy type I; (FXS) Fragile X syndrome; (MR) mental retardation; (FRDA) Friedreich ataxia; (HD) Huntington disease; (HDL-2) Huntington disease-like 2; (NCT) noncoding transcript; (SBMA) spinobulbar muscular atrophy; (SCA) spinocerebellar ataxia.

^aThe strand shown corresponds to the top strand of the affected gene where it is known. The sequence of both strands is given where this is not known.

- FMR1, 5' UTR,
CGG repeats
- FXS: >200
 - FXTAS: 55-200
 - Normal: 5-54



Repeat expansions can occur in **5'UTRs, coding regions, introns, or 3'UTRs.**

Normal and premutation alleles do not show usually disease symptoms, but premutation alleles are primed to expand in the next generation. As repeats get longer, symptoms are seen at an earlier age and are more severe. **Repeat lengths in introns, 3'UTRs, and 5'UTRs can become much larger than in coding regions.**

Spinocerebellar ataxias (SCA12), Spinal Bulbar Muscular Atrophy (SBMA), myotonic dystrophy (DM)

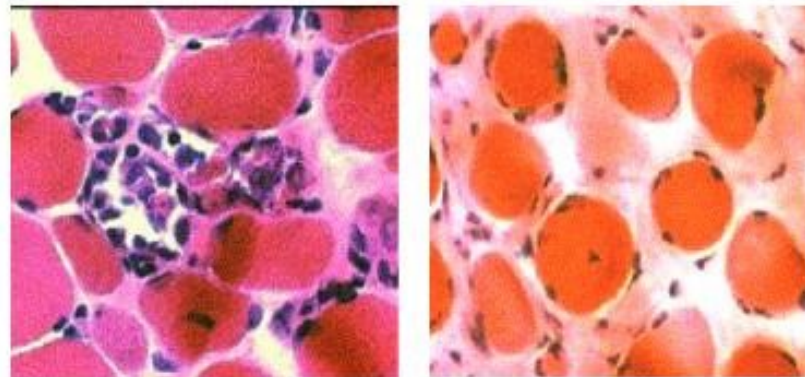
Svalové dystrofie -

charakterizovány progredující svalovou atrofií a slabostí s typickým histologickým obrazem, který prokazuje kolísání velikosti svalových vláken, jejich nekrózu a v pokročilém stadiu náhradu svalových vláken fibrózní a tukovou tkání.

Duchenne Muscular Dystrophy

2 yrs

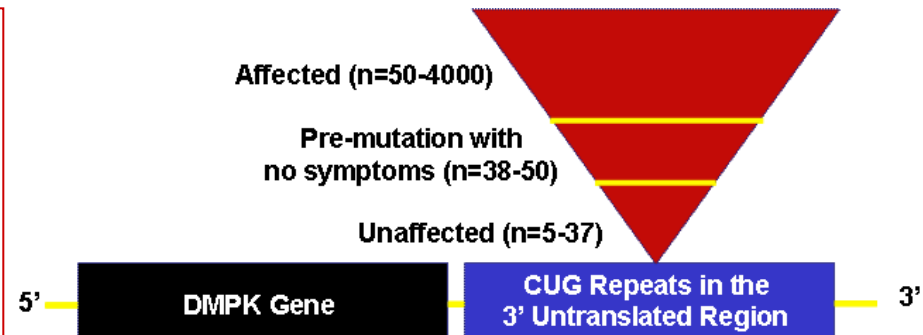
12 yrs



Myotonická dystrofie - multisystémová choroba, která postihuje kosterní i hladké svaly, ale také oči, srdce, centrální nervový systém, endokrinní, respirační, gastrointestinální a imunitní systém.

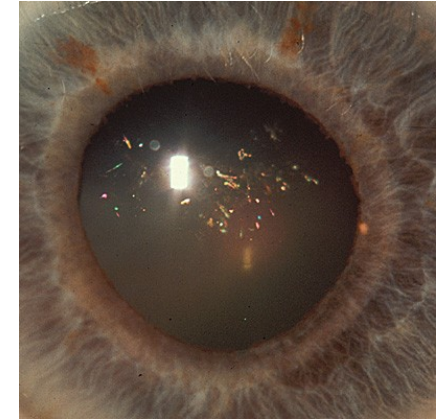
Přehled orgánových a systémových abnormit u myotonická dystrofie: **neuromuskulární:** slabost, myotonie, neuropatie; **CNS:** změny osobnosti, kognitivní deficit, hypersomnie, korová atrofie, změny bílé hmoty; **srdeční:** převodní poruchy, synkopy, náhlá smrt, městnavá srdeční slabost, prolaps mitrální chlopně; **respirační:** snížení rezistence na hypoxii, hypoventilace; **endokrinní:** diabetes, hypogonadizmus, hyperparatyreoidizmus, spontánní potraty, inkarcerovaná placenta, protrahovaný porod, poporodní krvácení; **gastrointestinální:** dysfagie, cholelithiasis, střevní pseudo-obstrukce; **oční:** zadní subkapsulární katarakta, ptóza, oftalmoparéza, pigmentová retinopatie, snížený intraokulární tlak; **kostní:** vystouplé čelo, malá sella turcica, gotické patro; **imunitní:** redukce imunoglobulinů.

DM1 is caused by an expansion of the CTG repeat in the 3'UTR of the dystrophia myotonia protein kinase gene (*DMPK*, 19q13.3), AD inheritance, frequency of DM: 1:17000.



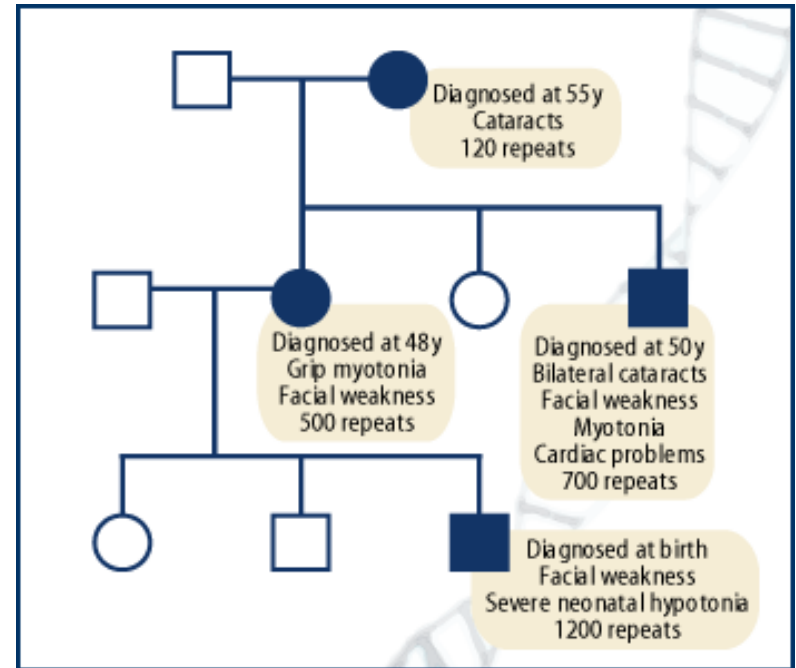
- Individuals with 5 to 37 repeats are unaffected.
- Individuals with 38-50 repeats carry the premutation. These individuals are asymptomatic. However, these repeats are unstable and can expand during meiosis. As a result, such individuals are at risk of having affected children.
- ~ 50 to 150 repeats are consistent with the mild adult-onset form of DM1,
~100 to 1000 repeats are consistent with the classic adult or childhood onset form of DM1,
> 750 repeats are consistent with the congenital form of DM1 and often result in severe neonatal complications.

- Patients with **50–150 CTG repeats (mild adult onset form of DM1)** may develop cataract, diabetes, myotonia, mild muscle weakness.
- Patients with **100–1000 CTG repeats (the classic form of DM1)** are affected earlier and more severely.
- **CTG repeat size above 1000** is associated with **the congenital form of DM1**, which may be fatal due to respiratory failure. Feeding difficulties, muscle weakness, foot deformity, and cognitive impairments are present in surviving infants.
- The very large mutations (>1,000 repeats) which result in the congenital form of myotonic dystrophy are always transmitted by an affected mother.

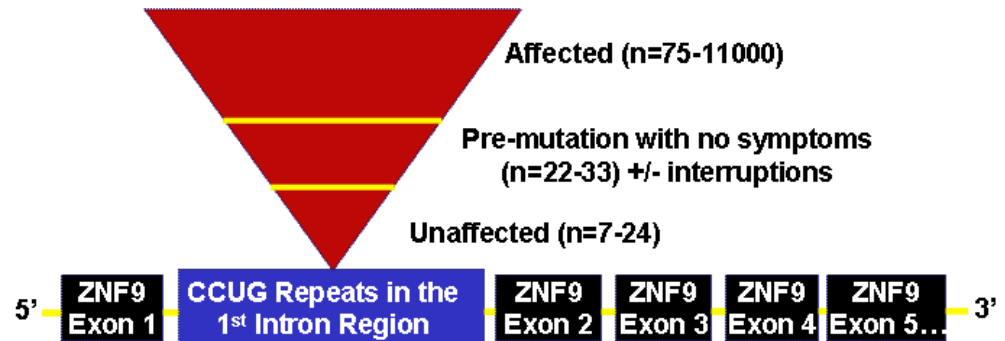


- The expanded CTG repeat - **'dynamic' mutation** - the number of repeats tends to increase in size over generations.

- Expansion of the CTG repeats commonly occurs during meiosis. As a result, children of affected individuals tend to have severe symptoms and earlier onset than their parents.

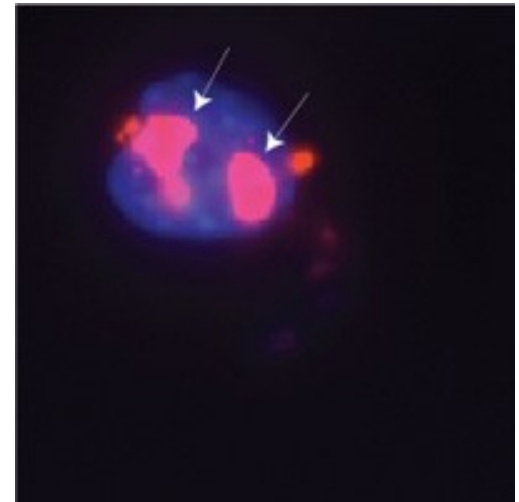
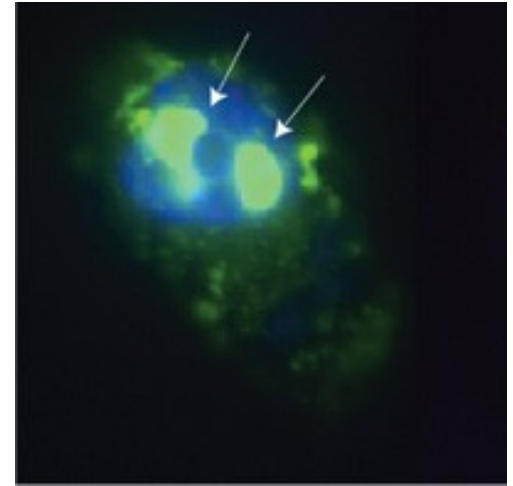


DM2 (also known as proximal myotonic myopathy [PROMM]) is caused by an expansion of the CCTG repeat in the first intron of the zinc finger 9 gene (*Znf9*, 3q21), AD inheritance.

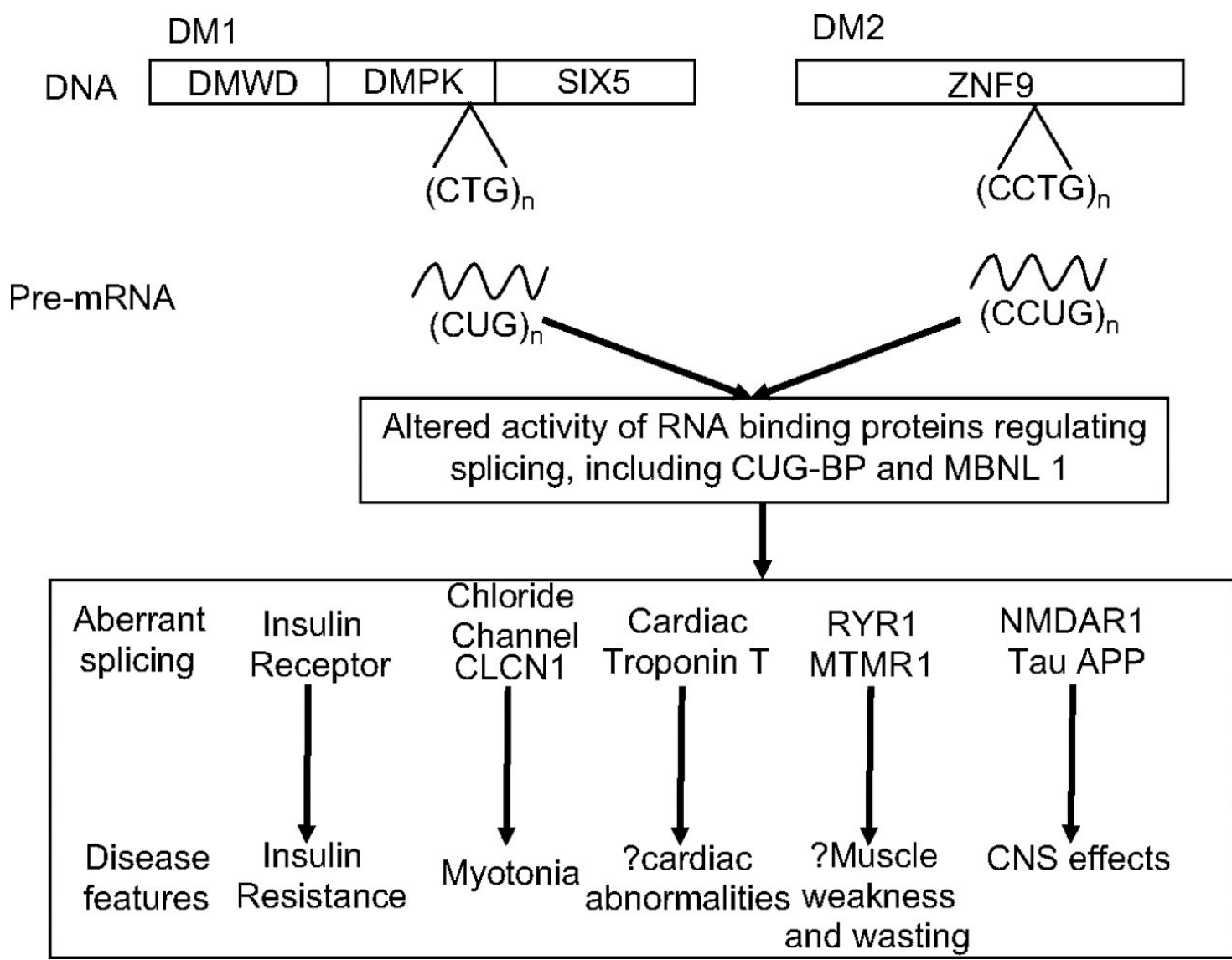


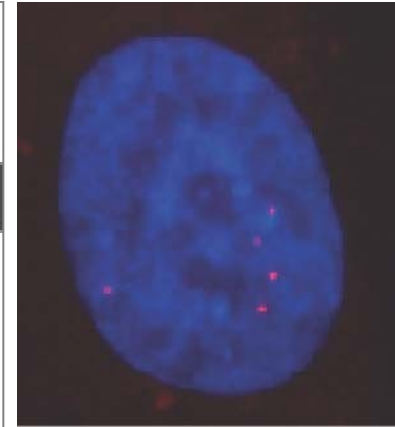
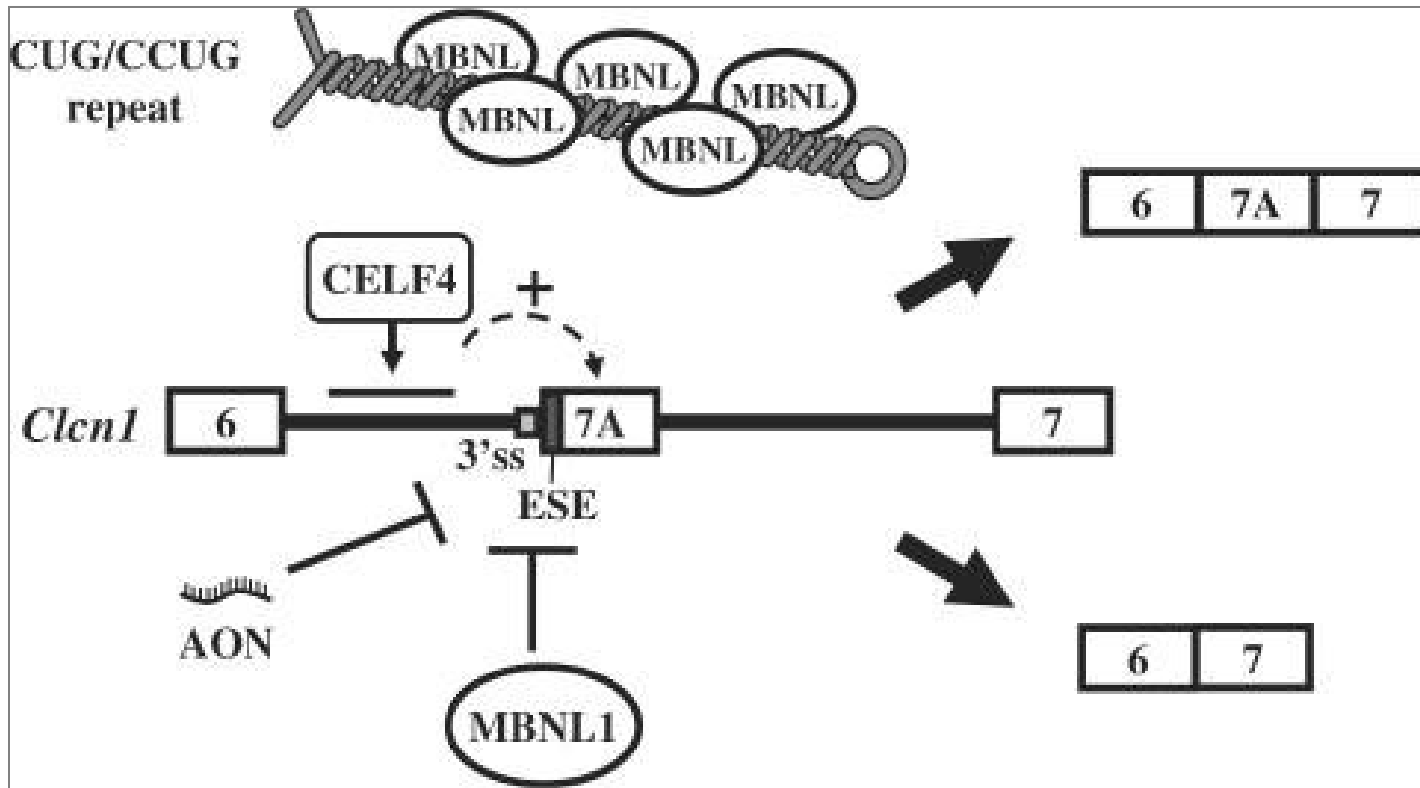
- **Unaffected individuals have less than 24 repeats.**
- **Affected individuals have between 75 and 11000 repeats.**
- The repeat structure in DM2 is more complex than the triplet repeat seen in DM1. The normal repeat structure is (TG)₁₂₋₂₆(TCTG)₇₋₁₂(CCTG)₃₋₉(g/tCTG)₀₋₄**(CCTG)₄₋₁₅**.
- **Individuals with 22-33 uninterrupted CCTG repeats carry a premutation.** These individuals are asymptomatic. However, these repeats are unstable and very likely to expand during meiosis (risk of having affected children).
- The minimum pathogenic length of the expanded region appears to be 75 uninterrupted CCTG repeats. Repeat counts can increase to over 11000 in affected individuals, with a mean repeat length of ~5000 repeats. The expanded region has been shown to display an even greater instability than the DM1 mutation.
- **Unlike DM1, the length of the DM2 expansion does not appear to correlate significantly with the age of onset or severity of disease symptoms.**

- Nuclei of cells of DM1/DM2 patients → expression of genes containing CUG/CCUG repeats → nuclear foci containing pre-mRNA with expanded CUG/CCUG repeats → **RNA gain-of-function effects** → capture of RNA binding proteins that regulate mRNA splicing: **muscleblind-like (MBNL) factor** and others → misregulation of splicing of certain genes: **chloride channel 1 (CLCN1)**, insulin receptor (IR), cardiac troponin T (cTNT/TNNT2), skeletal troponin T (TNNT3), and others.
- Insulin receptor and chloride ion channel pre-mRNAs are misspliced in DM patients → inappropriate expression of fetal isoforms and/or degradation of transcripts. **The lack of appropriate IR and CLCN1 splice isoforms in DM patients is thought to lead to the symptoms of insulin resistance and myotonia, resp.**

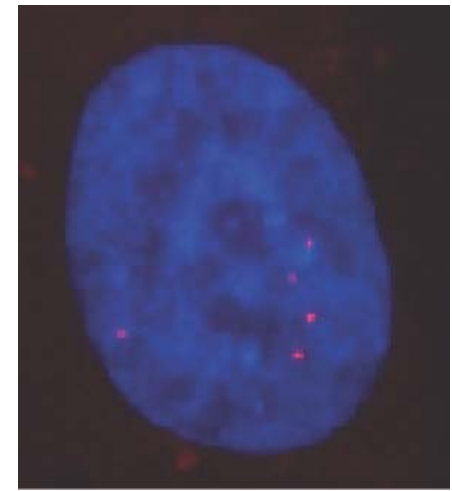
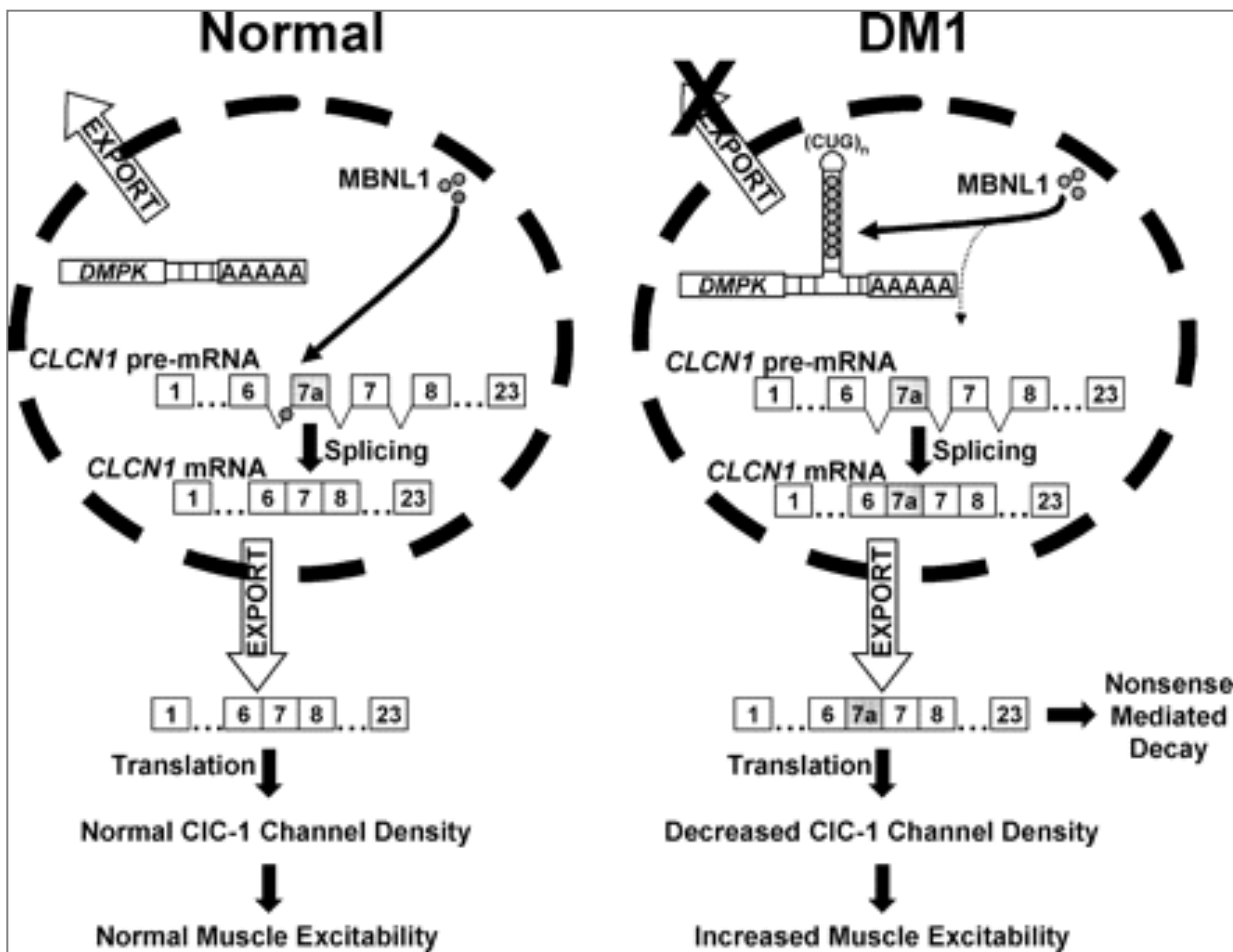


Myotonic dystrophy is thought to be caused by the binding of a protein called Mbnl1 to abnormal RNA repeats. In these two images of the same muscle precursor cell, the top image shows the location of the Mbnl1 splicing factor (green) and the bottom image shows the location of RNA repeats (red) inside the cell nucleus (blue). The white arrows point to two large foci in the cell nucleus where Mbnl1 is sequestered with RNA.





Model *Clcn1* splicing regulation by multiple factors. *Cis*- and *trans*-acting factors involved in the splicing regulation of *Clcn1* exon 7A are depicted. **An exonic splicing enhancer (ESE) is located at the 5' end of exon 7A. MBNL1 represses exon 7A inclusion through inhibiting ESE.** The facilitation of exon 7A inclusion by CELF4 is mediated by a region located in intron 6. **RNAs carrying expanded CUG/CCUG repeats deplete MBNL1 proteins, resulting in the facilitation of exon 7A inclusion.**



J Gen Physiol 2007;129:79-94

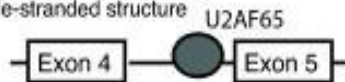
Proposed molecular model for increased muscle excitability in DM. Proper *CLCN1* pre-mRNA splicing in normal skeletal muscle is regulated by MBNL1 protein. Depletion of MBNL1 proteins results in inclusion of additional exons (e.g., exon 7a) containing premature termination codons. Aberrantly spliced *CLCN1* transcripts are exported from the nucleus, degraded through the nonsense-mediated decay pathway, and/or produce truncated proteins. These effects result in a dramatic reduction in the number of functional CIC-1 channels and a subsequent increase in muscle excitability resulting in myotonia.

Model of Regulation of cTNT Exon 5

Pathway to Exon 5 Inclusion

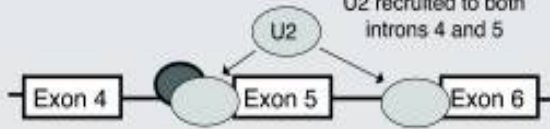
A splicing factor binding and splice site definition

U2AF65 binds the intron in a single-stranded structure



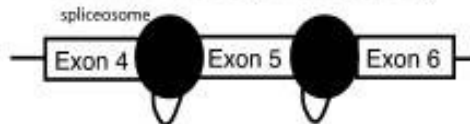
B spliceosome recruitment

U2 recruited to both introns 4 and 5

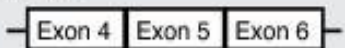


C splicing of "defined" introns

Both intron 4 and 5 spliced out separately

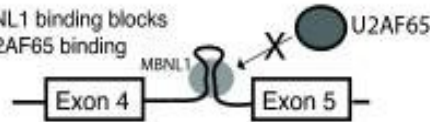


D splice product



Pathway to Exon 5 Exclusion

MBNL1 binding blocks U2AF65 binding

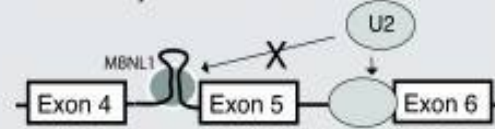


OR

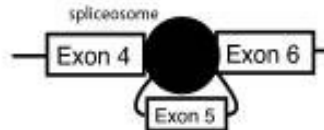
Stabilized stem-loop inhibits U2AF65 binding



no U2AF65 bound to recruit U2 to intron 4, and U2 only recruited to intron 5



Intron 4, exon 5 and intron 5 spliced out together



MBNL1 Regulates the cTNT cardiac troponin T Exon 5 Through Competition with U2AF65.

Model of regulation of cTNT exon 5 by MBNL1 and RNA structure. (A) Initial recognition of intron 4 by splicing factors. U2AF65 binds the intron in a single-stranded structure, but U2AF65 binding is inhibited if it cannot destabilize the stem because of MBNL1 binding or mutations that stabilize the stem-loop. (B) U2 snRNP recruitment to all 3' splice sites where U2AF65 is present. (C) Spliceosomal recruitment and splicing. (D) mRNA splice products. We hypothesized that MBNL1 may act through a similar mechanism to compete with U2AF65. U2AF65 is a potential competitive target of MBNL1 because a putative U2AF65 binding site appears to be in the loop portion of the stem-loop which MBNL1 binds. We found that MBNL1 does compete with U2AF65 for binding of a region within intron 4. This competition with U2AF65 is functionally important, because recruitment of the U2 snRNP is reduced by MBNL1.

**DM1 - expansion of the CTG repeat within the 3' UTR of DMPK,
DM2 - expansion of the CCTG repeat within the first intron of ZNF9.**

**• Despite the different expansions within two unrelated genes, both diseases share many common clinical manifestations (myotonia, muscle weakness, cataracts, insulin resistance, and cardiac defects).
The shared symptoms suggest the mechanisms causing the disease may also be shared.**

• Although DM1 and DM2 have similar symptoms, there are also a number of very dissimilar features making them clearly separate diseases.

• In DM1, weakness and atrophy involves distal, facial, bulbar and respiratory muscles, whereas in DM2 the proximal muscles are preferentially involved, and the patients have marked muscle pains.

• These phenotypic differences suggest that other cellular and molecular pathways are involved besides the shared molecular pathomechanisms.

- **DMPK??? ZNF9???**

- Knockout of ZNF9 in mice results in embryonic lethality. Mice heterozygous for the ZNF9 knockout display late-onset muscle wasting, cardiac abnormalities, cataracts, and mRNA expression defects similar to those seen in DM2. These defects can be rescued by reintroduction of wild type levels of ZNF9, suggesting that a loss of ZNF9 function contributes to DM2. ZNF9 has been proposed to act in a variety of cellular functions, including transcription, splicing, and translation.

- **ZNF9 protein may play a role in DM2 → RNA gain-of-function model for myotonic dystrophy is not the only pathomechanism.**

