

## Review

---

# The Contribution of Somatic Expansion of the CAG Repeat to Symptomatic Development in Huntington's Disease: A Historical Perspective

Darren G. Monckton\*

*Institute of Molecular, Cell and Systems Biology, College of Medical, Veterinary and Life Sciences, University of Glasgow, Glasgow, UK*

**Abstract.** The discovery in the early 1990s of the expansion of unstable simple sequence repeats as the causative mutation for a number of inherited human disorders, including Huntington's disease (HD), opened up a new era of human genetics and provided explanations for some old problems. In particular, an inverse association between the number of repeats inherited and age at onset, and unprecedented levels of germline instability, biased toward further expansion, provided an explanation for the wide symptomatic variability and anticipation observed in HD and many of these disorders. The repeats were also revealed to be somatically unstable in a process that is expansion-biased, age-dependent and tissue-specific, features that are now increasingly recognised as contributory to the age-dependence, progressive nature and tissue specificity of the symptoms of HD, and at least some related disorders. With much of the data deriving from affected individuals, and model systems, somatic expansions have been revealed to arise in a cell division-independent manner in critical target tissues via a mechanism involving key components of the DNA mismatch repair pathway. These insights have opened new approaches to thinking about how the disease could be treated by suppressing somatic expansion and revealed novel protein targets for intervention. Exciting times lie ahead in turning these insights into novel therapies for HD and related disorders.

**Keywords:** Somatic expansion, anticipation, CAG repeat, genotype to phenotype, mutation, DNA mismatch repair

## THE FELLOWSHIP OF THE REPEAT DISORDERS

*An unexpected discovery: unstable DNA in fragile X syndrome*

In May 1991, a remarkable series of papers began to appear that revealed an “unstable region of DNA” that was increased in size in individuals with fragile

X syndrome (FXS) [1, 2] (see Fig. 1 for a timeline of events). Within a matter of a few weeks, this unstable region was further revealed as containing a polymorphic CGG repeat [3] that mapped precisely to the position of the larger genomic DNA restriction fragments observed in FXS patients [4]. These data were particularly exciting as they provided potential explanations for two particularly unusual features of FXS and revealed previously unimagined genetic instability in humans. Firstly, as evidenced by its name, FXS had long been defined by the association of intellectual disability with a fragile site on the X-chromosome [5]. The fragile site presents as a region near the tip of the long arm of the X-chromosome that

---

\*Correspondence to: Darren G. Monckton, PhD, Professor of Human Genetics, Institute of Molecular, Cell and Systems Biology, College of Medical, Veterinary and Life Sciences, University of Glasgow, Davidson Building, University Avenue, Glasgow G12 8QQ, UK. Tel.: +44 141 330 6213; E-mail: darren.monckton@glasgow.ac.uk; ORCID: <https://orcid.org/0000-0002-8298-8264>

fails to condense during metaphase in cultured cells, particularly when the cells are grown under conditions of reduced folate [6]. The presence of a CGG repeat, and associated changes in the methylation patterns in the region, suggested that chromosomal fragility in the region was in some way directly associated with the unusual sequence nature and epigenetic consequences of the CGG repeat expansion. Secondly, and of more direct relevance to Huntington's disease (HD) and other inherited human disorders, it was also revealed that the enlarged region was highly unstable with different sizes present in individuals from the same family, consistent with intergenerational instability [1, 2, 7]. This intergenerational instability was of particular interest, as it also offered a potential explanation for the unusual inheritance patterns observed in FXS. Even in the first description of the sex-linked nature of FXS by Martin and Bell in 1943, the authors noted there were two males, that by their position within the pedigree, were obligate carriers of the mutant X-chromosome, but did not present with FXS, but could nonetheless transmit the disease to affected grandsons through their carrier daughters [8]. The existence of these so called normal transmitting males were confirmed in multiple additional families [9, 10]. Additionally, it was shown that the penetrance of the disease increased in successive generations, with the penetrance appearing to be in some way associated with the number of female transmissions relative to a normal transmitting male [9, 10]. The existence of these unusual inheritance patterns were received with a good deal of skepticism and the phenomena became known as the Sherman paradox [11, 12]. The Sherman paradox was thus resolved, when it was demonstrated that: affected FXS males inherited a full mutation expansion of greater than approximately 210 CGG repeats from their carrier mothers; that normal transmitting males carried non-penetrant premutations of 50 to 200 CGG repeats; and that such premutations were biased toward expansion in subsequent generations when transmitted by carrier females [7]. The increasing penetrance thus being a product of the fraction of mutant X-chromosomes that expand intergenerationally from premutations to full mutations. Notably, the intergenerational mutation rate of premutation FXS alleles was near an astonishing 100%. At the time such mutation rates were unprecedented in human genetics. Up until that point the most unstable sequences that had been detected were the hypervariable minisatellite tandem repeats, that were first described in 1980 [13] and subsequently

exploited as key genetic markers in DNA profiling and DNA fingerprinting [14, 15], and that had intergenerational mutation rates in the order of 1 to 5% [16]. Even these minisatellite "hypermethylation" frequencies of 1 to 5%, were still orders of magnitude greater than anything else previously observed in the human germline.

*Two is company: spinal and bulbar muscular atrophy is also caused by a repeat expansion*

At about the same time these electrifying developments in FXS were reported, a short paper in *Nature* in July 1991 [17] described "androgen receptor (*AR*) gene mutations in X-linked spinal and bulbar muscular atrophy." Specifically, the paper described the presence of an "enlargement of the CAG repeat" of 40 to 52 repeats in spinal and bulbar muscular atrophy (SBMA) patients, "roughly double" the 17 to 26 repeats observed in unaffected controls. The authors further speculated that "enlargement of the polyglutamine repeat may prevent the androgen receptor from performing an important regulatory activity in motor-neurons, thereby leading to the degeneration of these cells which is characteristic of the disease." Of note however, the SBMA *AR* CAG expansion was not initially revealed as genetically unstable. Indeed, the SBMA *AR* CAG is only moderately unstable in the germline and this was only revealed a year or two later [18, 19]. Nonetheless, the existence of a second trinucleotide repeat expansion mediating another inherited human disorder, established repeat expansion as a novel mechanism of inherited disease in humans.

*The shadow of the past: anticipation and eugenics*

Whilst it is easy now to imagine unstable DNA as explaining a variety of different phenomena, such as for instance anticipation, it is important to remember that right up until the point at which the disease-causing mutations were identified, the very existence of anticipation as a genuine biological observation was not yet firmly established. The phenomena of antedating, or as it later became known anticipation, earlier age at onset observed in successive generations, was first proposed to occur in the mid-19th century [20, 21]. Unfortunately, these ideas were taken up by some, most notably, Frederick Mott, within the burgeoning eugenics movement of the early part of the 20th century who proposed antic-

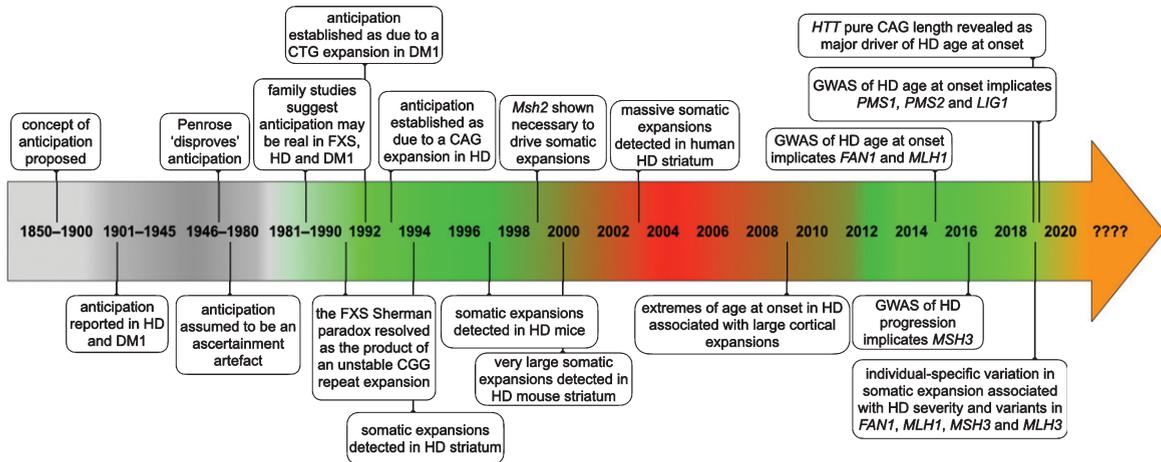


Fig. 1. Timeline of some of the key events establishing anticipation as a genuine biological phenomenon and somatic expansion as contributing toward HD pathology.

ipation to apply to “mental illness” and “insanity” in general and used this as justification for proposals to limit the reproductive rights not just of the “insane”, but also the “higher grade imbeciles” who continue to provide “fresh tainted stocks” [22]. Despite being an ardent eugenicist himself, Karl Pearson (an early pioneer of mathematical statistics and originator of the Pearson correlation coefficient) [23], and Pearson’s student, David Heron [24], noted the potential “fallacy” in the interpretation of the pedigree data used to support the existence of anticipation; namely, that disease-dependent effects on reproductive success would inevitably lead to the identification of a greater average age at onset in affected parents relative to their children. Notably, some descriptions of the inheritance of HD in the early part of the 20th century reported that this disorder also appeared to occur earlier in successive generations [25–27]. Nonetheless, it was recognised by Charles Davenport, another prominent eugenicist and advocate of the sterilisation of HD individuals, that the “law of anticipation” in HD might be “partly, if not wholly, illusory” due to the ascertainment biases as previously outlined by Pearson and Heron [28]. Around the same time anticipation was also noted in myotonic dystrophy type 1 (*dystrophia myotonica*, DM1) when apparently unrelated individuals with the adult onset form of the disease with a primarily neuromuscular presentation, were shown to be connected by prior generations with often cataracts as their only symptom [29]. Over the next few decades the existence of anticipation in DM1 appeared to be supported by additional family studies (e.g., [30, 31]) in some of which the authors at least partially considered the issue of ascertainment bias

(e.g., [32–34]). Following the atrocities of the second world war, and the backlash against the eugenic ideals of the early 20th century, Lionel Penrose published a paper in 1947 in which he described in detail the many potential biases that could account for apparent anticipation in ascertained families [35]. Penrose considered the evidence for anticipation in a number of disorders, including HD, and concluded, that even in DM1 for which the apparent evidence was strongest, the confounding effects of several different ascertainment biases could not be excluded as having yielded the level of apparent anticipation observed. For a more detailed discussion of the history of anticipation and the eugenics movement see the works by Harper et al. [36] and Judith Friedman [37]. Penrose’s publication was widely interpreted, somewhat erroneously, as having proved that anticipation did not occur, and enhanced by the absence of a plausible genetic mechanism that could yield anticipation, the existence of anticipation as a genuine biological phenomena was widely dismissed. Indeed, as late as 1989 a paper describing a detailed analysis of the inheritance patterns in DM1 was entitled “Anticipation in myotonic dystrophy: fact or fiction?” and conservatively concluded “that anticipation may be inherent in the transmission of myotonic dystrophy” [38]. In this study, all of the potential, and very real, biases described by Penrose, were carefully taken into account and corrected for, and still evidence for striking anticipation of 20 to 30 years per generation persisted. The neurologist, Chris Höweler, who conducted most of this work for his PhD, recounts how difficult it was finding two external examiners who would approve his thesis, with one expert

apparently commenting “I’m not going to read it, that’s bull\*\*\*\*, I’m not going to spend my time reading about bull\*\*\*\*” [39]. Even the two examiners who eventually approved the thesis commented “we don’t know whether you are right, but you have good arguments”. Notably, Höweler et al. considered several possible explanations for how anticipation might occur, even suggesting that “a gradual change of the mutation itself in successive generations might be assumed” [38].

*A short cut to the mutation: unstable DNA as an explanation for anticipation?*

Even prior to the identification of the disease causing mutations, the concept that the anticipation in DM1 might be analogous to the increasing penetrance observed in FXS was posited [38, 40]. Thus, building on the observations of extreme genetic instability observed in FXS and how this could resolve the Sherman paradox, it was quickly suggested that the anticipation observed in DM1 could possibly be explained by an unstable DNA fragment increasing in length from one generation to the next [41].

*A mutation unmasked: a CTG repeat expansion in myotonic dystrophy type 1*

With the discovery of the FXS mutation, the lack of a genetic mechanism to explain anticipation evaporated. Moreover, the presence of a putative expansion provided a rapid route toward identification of the causative mutation; cloned fragments from the critical region delineated by traditional linkage mapping could be used as probes on Southern blot hybridisation analyses of restriction digested genomic DNA in the absence of any knowledge of their sequence or gene content. And, within months of the reports of the identification of the FXS mutation, reports appeared in February 1992 of the presence of unstable DNA in DM1 [42–44], which was quickly confirmed as the expansion of a CTG repeat in the *DMPK* gene [45–47]. The size of the CTG repeat inherited was shown to be inversely correlated to age at onset, and highly unstable and biased toward expansion in the germline, providing a simple molecular explanation for anticipation [36].

*The old chestnut: anticipation in HD?*

As alluded to above, HD had long been posited to display anticipation [25–27, 35]. By the molecular

era of human genetics in the 1980s, it was broadly accepted that the juvenile form of HD was strongly associated with paternal transmission of the mutation [48–51]. Various explanations for the excess of transmitting fathers of juvenile HD were posited, including mitochondrial effects, the existence of X-linked modifiers, or genomic imprinting of the HD locus or modifier genes [50, 52–56]. Of note, in rejecting a mitochondrial DNA effect and positing instead the action of an imprinted modifier gene, Irwin et al. noted in 1989 that, even before the *HTT* gene had been identified, “The identification of a putative modifying gene that might be altered to retard disease onset is appealing as a possible therapeutic stratagem” [57]. Nonetheless, at this point the evidence for a broader degree of anticipation operating more generally within HD families was largely discounted as attributable to the very real biases detailed by Penrose [48–50, 58, 59]. However, by mid 1992 with three disorders now associated with the expansion of trinucleotide repeats, and the molecular basis for anticipation in DM1 established, there was much speculation that other disorders, especially those with unusual inheritance patterns, were likely to share a similar genetic basis. Indeed, with anticipation suddenly a fact, many of the old arguments were put aside and anticipation in HD was assumed to be real and direct predictions were made that it would be explained by the expansion of a trinucleotide repeat [36, 60].

*Farewell to ignorance: HD is caused by a repeat expansion and anticipation is real*

Contrary to expectations, the large expansions of hundreds of repeats that facilitated the identification of the mutations in DM1 and FXS were not found in HD. Nonetheless, within the year the predictions of unstable DNA in HD were borne out with the identification of the HD causing polyglutamine encoding CAG repeat expansion in the *HTT* gene in March 1993 [61]. The initial report of the identification of the CAG expansion immediately revealed that the number of CAG repeats was inversely associated with age at onset, and that the repeat was intergenerationally unstable with a bias toward expansions [61]. These insights were quickly verified in additional cohorts, and it was rapidly established that anticipation was a genuine biological observation in HD and was particularly associated with expansions in the male germline [62–66].

*In from the cold: anticipation legitimised*

With the discovery of unstable DNA as the basis for the conspicuous anticipation observed in DM1, the concept of anticipation was very much “legitimised” as a genuine biological phenomenon [67]. A striking consequence of this apparent liberation from the shackles of Penrose’s deconstruction of the concept of anticipation, was that over the next few years apparent anticipation was reported in numerous conditions. Whilst some of these did indeed later pan out to be associated with unstable DNA (e.g., [68–70]) (see below), many, including claims for anticipation in, for example, rheumatoid arthritis [71], uni- and bipolar affective disorder [72, 73], schizophrenia [74], rolandic epilepsy and speech dyspraxia [75], familial primary pulmonary hypertension [76], and Meniere disease [77], have not (or have not yet) been unequivocally associated with expanded simple sequence repeats. It thus seems likely that many of these reports of apparent anticipation, were indeed artefacts of the very real ascertainment biases first delineated by Pearson [23] and later expounded upon by Penrose [35]. Indeed, even after the discovery of unstable DNA, a lively debate ensued about how to fully correct for these ascertainment biases, in particular with regard to reports of anticipation in bi-polar affective disorder and schizophrenia (e.g., [78–85]). Of course, it also remains possible that there are other causes of genuine anticipation, such as for instance might be mediated by time dependent changes in environmental exposures, as has been proposed to account for the apparent anticipation observed in familial amyloidotic polyneuropathy type I for which the mutation is known to be a very static missense variant in the transthyretin gene [86]. Whilst there are many lessons to be learned from the convoluted history of genetic anticipation, some that may be of particular relevance to unravelling the role of somatic expansion in HD are that: i) we should follow where the data leads with an open mind and consider alternative explanations, even when they challenge our preconceptions or nominally established facts and mechanisms; and, most directly, ii) unstable DNA can explain previously unexplained phenomena.

*Inside information: polyglutamine encoding CAG repeat expansions cause an array of inherited neurological disorders*

In contrast to the large non-coding expansions of hundreds of repeats observed in DM1 and FXS,

HD was found to be associated with a more moderate polyglutamine encoding expansion mostly in the range of 40 to 50 CAG repeats, similar to that observed in SBMA. Following the series of remarkable breakthroughs in FXS, DM1, SBMA and HD, and using the insights gained to accelerate the search for the causative mutations, over the next few years additional similarly moderately sized genetically unstable polyglutamine encoding CAG repeat expansions were detected in spinocerebellar ataxia type 1 (SCA1) (1993) [87], Machado Joseph disease/SCA3 (1994) [88, 89], dentatorubral pallidoluysian atrophy (1994) [90, 91], SCA2 (1996) [92, 93], SCA7 (1997) [94] and SCA17 (2001) [95]. In addition to a likely shared pathogenic mechanism mediated by a gain of function of the expanded polyglutamine containing protein [96], in each disorder there is an inverse relationship between inherited repeat length and age at onset, and in each case the repeats are intergenerationally unstable, particularly during male transmission. These factors combined with an intergenerational expansion bias, leads to anticipation in each disorder, very similar to that observed in HD. The SCA8 mutation was identified in 1999 as the expansion of a CTG•CAG repeat [97]. Interestingly, the SCA8 repeat is expressed as a polyglutamine encoding CAG tract on one strand, and as a CUG repeat as part of a non-coding transcript on the other strand, suggesting pathology may involve gain-of-function at both the protein and RNA levels [98]. SCA8 is also unusual in that disease-causing expansions are typically larger than observed in the other polyglutamine encoding CAG repeat disorders (~80 to 250 repeats), yet expanded alleles show variable penetrance and are observed at relatively high frequency in the general population and/or are associated with atypical phenotypes (e.g., [97, 99–103]). Additionally, although expanded alleles are prone to expansion in the female germline, they are heavily biased toward contraction in the male germline [97, 101, 104] and consequently anticipation is not a prominent feature of SCA8 families.

*Many repeats: expanded simple sequence repeats cause a variety of disorders*

Over the ensuing decade additional disorders were associated with genetically unstable expanded trinucleotide repeats, including: a CCG expansion in fragile X E (FRAXE) (1993) [105]; a GAA expansion in Friedreich ataxia (FA) (1996) [106]; a non-coding CAG expansion in SCA12 (1999) [107]; and a CTG

repeat expansion in an alternatively spliced exon of the *JPH3* gene in Huntington disease like 2 (HDL2) (2001) [108, 109]. The expansion of unstable simple sequence repeats in disease was also shown to extend beyond triplets including: a CCCCCGCC-CGCG dodecamer repeat in progressive myoclonus epilepsy (1997) [110]; an ATTCT pentanucleotide expansion in SCA10 (2000) [111]; and, a CCTG tetranucleotide expansion in myotonic dystrophy type 2 (DM2) (2001) [112]. More latterly, at least partially facilitated by whole genome sequencing, additional disease-associated expansions have been discovered, including: a TGGAA pentanucleotide in SCA31 (2009) [113]; a GGGGCC hexanucleotide in frontotemporal dementia and amyotrophic lateral sclerosis (2011) [114, 115]; a GGCCTG hexanucleotide in SCA36 (2011) [116]; a CTG in Fuchs corneal dystrophy (2012) [117]; a CCCTCT hexanucleotide in X-linked dystonia-parkinsonism (2017) [118]; an ATTTC pentanucleotide in SCA37 (2017) [119]; a GGC expansion in Baratela-Scott syndrome GGC (2019) [120]; a GGC expansion in neuronal intranuclear inclusion disease-related disorders (2019) [121]; an AAGGG pentanucleotide in cerebellar ataxia, neuropathy, vestibular areflexia syndrome (2019) [122]; and, TTTCA/TTTTA pentanucleotide expansions in at least six different genes in benign adult familial myoclonic epilepsy (2018–2020) [123–127]. All of these disorders are associated with intergenerational instability of the repeat expansions, and many display associated atypical inheritance patterns such as anticipation.

*At the sign of the fuzzy band: somatic mosaicism in fragile X syndrome and myotonic dystrophy type 1*

In addition to intergenerational instability, it was noted in the primary FXS and DM1 studies, before the repeat expansion had even been characterised, that in many affected individuals, the enlarged fragment in blood DNA presented not as a discrete band on the Southern blot hybridisations of restriction digested DNA used to reveal the presence of the mutation, but rather as broad “fuzzy”, “smeary” or “blurred” bands [1, 2, 42–44]. These observations suggested that the intergenerationally unstable region of DNA was also somatically unstable and varied in length between cells within the individual. Indeed, in FXS, many individuals present with two or more relatively discrete pre- or full-mutation alleles (in addition to a small non-disease associated allele in females),

that are present in multiple tissues, consistent with very early embryonic mutation events [7]. Indeed, it would appear that the FXS full mutation is somatically unstable during very early development, but is then somatically stabilised, likely as a result of methylation of the region [128–132]. In DM1, however, it was quickly established that there were differences in repeat size not just within tissues, but even more dramatically between tissues [133–136]. Highly consistent with a possible role for somatic expansion in contributing toward the tissue specificity of the symptoms, it was rapidly established (1993–1994) that the repeat length observed in skeletal muscle was often thousands of repeats longer than that observed in blood DNA [133–136]. Moreover, in further contrast to FXS, whilst there is evidence for embryonic instability in DM1, this is usually only observed in congenital cases inheriting very large expansions, and in most individuals, the repeat appears to be relatively stable during embryogenesis with most somatic expansions arising postnatally [134, 135, 137–145]. Indeed, there is clear evidence that somatic expansions in blood DNA continue to accrue throughout the lifetime of the individual [141, 146–149]. Given that larger alleles are associated with earlier onset and more severe disease in DM1, it seems logical to assume that somatic expansion contributes towards the disease process. Indeed, it is now apparent that the individual-specific rate of somatic expansion is associated with both disease severity and progressive DM1 phenotypes (*i.e.* individuals in whom the repeat expands more rapidly somatically, have an earlier onset and more rapid disease course) [148, 150, 151].

*A pattern mirrored: somatic mosaicism in HD?*

Possible evidence for somatic instability was likewise reported in the primary study defining the CAG expansion in HD, with the authors noting a “diffuse fuzzy PCR product” in the blood DNA of at least one HD patient [61]. However, early follow-up studies using polyacrylamide gel electrophoresis analysis of radiolabelled PCR products revealed that in contrast to clear differences between blood and sperm DNA, the length of the primary PCR product derived from the mutant HD chromosome did not change in lymphoblastoid cell lines and primary blood DNA samples collected many years apart, between a limited number of peripheral tissues, or between different regions of the brain, leading to premature claims of “mitotic stability” and “gametic, but

not somatic instability” [65, 152]. Although repeat length variation around the primary allele was noted, these variants were conservatively interpreted as most likely representing the PCR “chatter” that is well known to generate repeat length heterogeneity even for small non-disease associated alleles [65, 152]. Whilst retrospectively it is possible to point to a greater spread of larger fragments above the primary disease associated allele in the striatal samples in particular, these data highlighted that the very large shifts in modal repeat length observed in the somatic tissues of FXS and DM1 patients were not replicated in the majority of HD patients inheriting moderately sized expansions (40 to 50 CAG repeats) [65, 152]. The following year however, a more detailed analysis of possible somatic variation was undertaken by Telenius et al. and came to a very different conclusion providing compelling evidence for “somatic mosaicism in HD which is seen predominantly in the basal ganglia and other regions of the brain selectively involved in HD” [153]. Using a wide array of peripheral tissues and different brain regions from five adult onset HD patients, the authors convincingly demonstrated that whilst a tail of PCR slippage products smaller than the primary PCR product was detected for both disease and non-disease associated alleles, a tail of larger fragments, up to approximately +5 CAG repeats, was only prominently observed in specific brain regions for the disease associated allele [153]. Notably, these putative somatic expansions were observed at a much lower level in the cerebellum and peripheral tissues and were likewise not observed using a cloned PCR template. Even more convincingly, the authors demonstrated that in two juvenile HD cases the modal length was 13 CAG repeats bigger in other brain regions relative to the cerebellum (78 versus 65 CAGs repeats respectively in one case with onset at 6 years, and 86 versus 78 CAGs repeats respectively in a second case with onset at 4 years). Unfortunately, blood DNA was not available to help identify the inherited progenitor allele length and thus determine if the *HTT* CAG repeat was particularly prone to expansions in most brain regions and/or was simply stable or actively prone to contraction in the cerebellum. Nonetheless, these data confirmed that somatic expansions were observed in the affected brain regions and were observed at a much lower level in the cerebellum, a brain region that is relatively spared from degeneration in adult-onset HD, at least in early in the disease course. This study thus provided the first direct evidence that somatic expansion represented a very plausible expla-

nation for at least some of the regional specificity of downstream neuropathology in HD. However, whilst the authors noted that these observations were consistent with somatic expansion contributing toward regional selectivity of neuronal loss, they themselves noted that they were not able to ascertain in which cell types (e.g., neurons versus glia) that the expansions occurred and noted that it remained possible that somatic expansions occurred predominantly in glia as a secondary by-product of neuronal death and active gliosis [153]. Such an interpretation was likely at least partly driven by the assumption at the time that expansion events occurred via DNA replication slippage [154–156] (see below), and thus could arise in mitotically dividing cells, but not in post-mitotic neurons. A further independent study also concluded that there was indeed evidence for a greater preponderance of somatic expansions in various brain regions relative to those observed in blood and cerebellum, but the authors noted these were not restricted to the primary affected brain regions, and concluded also that the differences were “too small to make this mechanism an obvious candidate for the cause of differential neuronal degeneration in HD” [157]. Another important observation from around this time was the demonstration, that not only was the mutant *HTT* protein expressed in HD brains, but that it appeared more diffuse in its size relative to the non-disease associated allele, with a greater spread in the cortex relative to the cerebellum [158]. These differences were most apparent in juvenile cases where they appeared to directly reflect the spread of somatic expansions detectable at the DNA level. The authors again noted that although largely absent in the cerebellum, *HTT* proteins with apparently somatically expand glutamine repeats were detected in other regions of the brain that are less affected in HD, and that their cellular origin (i.e., glia versus neuron) remained unknown [158]. A role for somatic expansion was further supported by the observation that, at least in one patient there was a correlation between the degree of regional somatic mosaicism in the brain and regional pathological severity as assessed by a qualitative assessment of neuronal loss [159].

## THE TWO TOWERS

### *The taming of *Mus domesticus*: mouse models*

In the years after the identification of the disease-causing mutations, much effort in the triplet repeat field, as it had become known, was expended gen-

erating cell and organismal models that could be used to further understand the pathologic processes. Key to these developments were the generation of various mouse models. The first mouse model of a repeat expansion disorder, an SBMA transgenic model incorporating 45 CAG repeats in an androgen receptor cDNA construct failed to replicate either disease pathology or genetic instability [160]. Excitingly however, a SCA1 transgenic incorporating a much larger 82 CAG repeat allele in an ataxin 1 cDNA transgene with high levels of expression specifically in cerebellar Purkinje cells did display a neurodegenerative ataxic phenotype [161]. However, the repeat was not detectably unstable. The first HD repeat model, an *HTT* cDNA transgenic with 44 CAG repeats unfortunately contained a frameshift mutation that likely contributed to the lack of a phenotype in these animals [162]. Notably though, these mice did not display any detectable genetic instability either.

*The parting of the ways: downstream pathology versus the mechanism of expansion*

At about this point in time, in the mid 1990s, after the identification of the disease-causing mutations, the triplet repeat expansion field (at this time none of the non-triplet repeat expansions had yet been identified) split broadly into two camps: those assailing the pathogenic pathways downstream of the repeat expansion; and those that concentrated more on surmounting the mechanisms of repeat instability *per se*. In a bizarre twist of fate however, the first transgenic model to display a robust HD-like phenotype was actually generated by Mangiarini et al. to model genetic instability [163]. The R6 lines were based on a human *HTT* exon 1 transgene that expressed a truncated HTT protein containing an expanded polyglutamine tract encoded by ~130 CAG repeats, and this proved sufficient to generate a progressive neurological phenotype. The R6/2 mice in particular have become established as one of the most widely used HD animal models to investigate a wide range of aspects of the polyglutamine pathology. Fortunately though, the original hypothesis was also proved correct and high levels of both intergenerational and somatic instability were also revealed, in particular in the R6/1 and R6/2 lines [164]. A notable feature of the R6 lines, like two series of DM1 mouse transgenics that were generated in parallel, and that also displayed genetic instability [165, 166], was the low copy number of the integrants that facilitated mutation detection relative to the large multicopy inserts

that characterised previous transgenic repeat models [160–162]. Notably, the R6 transgenics revealed a pattern of age-dependent, tissue-specific somatic expansion that was most prominent in the striatum and virtually absent in the cerebellum. However, consistent with the cell division-dependent replication slippage models that predominated in mechanistic thinking at the time, the authors suggested that the somatic expansions in brain were derived from mitotically dividing glial cells [164]. These tissue-specific patterns of somatic expansion were also later replicated in knock-in mouse models in which the expanded CAG was targeted into the endogenous mouse *Htt* gene [167, 168] (see also Wheeler and Dion, this issue [169] for additional mouse models).

*The small pool: very large striatal-specific somatic expansions in HD mice*

With the exception of a very limited single molecule analysis of blood DNA used essentially as controls for single sperm analyses of male germline dynamics [170], up until the year 2000, all the previous analyses of somatic mosaicism in HD patients or animal models had been conducted using bulk DNA PCR analyses (i.e., PCR using large amounts of input DNA, typically >10 ng, equivalent to »1,000 cellular equivalents). Although these studies revealed clear evidence for somatic instability, the size of the acquired somatic expansions beyond the length of the inherited progenitor allele detected were relatively small. In human brain samples, even in juvenile cases, the largest acquired somatic expansions detected were in the order of +13 repeats, and much less than this in patients with more typical HD germline expansions in the range of 40 to 50 CAG repeats [152, 153, 157, 159]. Somatic expansions in HD mice with much larger germline alleles of 100+ repeats were shown to have acquired somatic expansions in the order of about 20 repeats greater than the inherited progenitor allele [164, 167]. These bulk DNA analyses by standard PCR, were known to have limitations including Taq polymerase slippage that generates a tail of shorter products [171], thus masking any potential contractions, and, the preferential amplification of smaller alleles [172]. The PCR bias is further confounded by the detection method in which usually only a single radionucleotide or fluorescent moiety is incorporated into the PCR product, independent of the total fragment length. Moreover, using either radiolabelled products and polyacrylamide gel electrophoresis, or fluorescently labelled products and

capillary gel electrophoresis, yields non-zero background signals that mask low frequency variants. Thus, these approaches were unlikely to accurately detect amplification of large low frequency somatic variants. In October 2000 Kennedy et al. reported the application of the sensitive small pool PCR approach [146, 173] to investigating somatic instability in two knock-in HD mouse lines with inherited progenitor alleles of either 72 or 80 CAG repeats [168]. In small pool PCR, multiple replicate PCRs are performed with very small amounts of input DNA (typically 1 to 50 molecules per reaction) for a limited number of PCR cycles such that PCR competition between small and large alleles is reduced. Moreover, the PCR products are detected by Southern blot hybridisation using a radiolabelled repeat unit probe that binds more efficiently to larger alleles containing more repeats. Using this approach, it is thus possible to partially compensate for the amplification bias of smaller alleles, and to detect the products of individual input molecules containing up to at least 1,000 trinucleotide repeats [146, 174]. Using this small pool PCR approach Kennedy et al. revealed that a subset of cells in the striatum of their knock-in HD mice had acquired somatic expansions of up to at least 250 repeats, some three times larger than the inherited progenitor allele. Somatic instability in these mice was revealed to be age-dependent and highly tissue-specific with the striatum clearly displaying the highest frequency of large expansions relative to other brain regions. These sensitive single molecule analyses, free of the confounding effects of Taq polymerase slippage that blight bulk DNA analyses, revealed that net somatic contractions were essentially absent and somatic mosaicism was highly biased toward further expansion. Notably, the absence of overt neurodegeneration in these mice argued against the concept that somatic mosaicism was a by-product of active gliosis [168].

*The window on the brain: ultra large striatal-specific somatic expansions in HD individuals*

The detection of such large somatic expansions in the primary affected brain region of a mouse model at least partially reignited the debate that somatic expansion might be a key driver of regional pathology in HD. Skepticism remained though, as the initial data presented by Kennedy et al. were derived not just from mouse models, but from mouse models with germline alleles of 72 or 80 CAG repeats, i.e., almost

double the size of an allele inherited by a typical HD patient. There was considerable doubt that expansions anywhere near so large would be detected in humans. However, three years later, in 2003, Kennedy et al. reported the application of the same sensitive small pool PCR approach to investigating regional instability in the brains of HD individuals [175]. Astoundingly, they revealed that a subset of cells in the brain contained massive somatic expansions of many hundreds of repeats. Indeed, in two individuals who inherited either 41 or 51 repeats, some striatal and/or cortical cells were detected with more than 1,000 CAG repeats. Critically, these two individuals did not die of end-stage HD and received a neuropathological classification of Vonsattel grade 0 [176], i.e., no microscopic evidence of pathological cell loss in the striatum. In the individual with the smallest germline allele (41 repeats), who died at age 40, ~13 years prior to their predicted age at motor onset, large somatic expansions were only detected in the striatum and were absent in the cortex and hypothalamus. In the second individual who inherited a larger germline allele (51 repeats), and who died at age 27, ~6 years prior to their predicted age at motor onset, massive somatic expansions were detected in the striatum, and to a lesser extent in the cortex, but were absent in the cerebellum. Notably, analysis of another individual who inherited a much larger allele (~75 repeats) and died 10 years after diagnosis, revealed a much less obviously regional-specific pattern of somatic expansion. Indeed, in this end-stage patient, the largest expansions were observed in the cortex rather than the striatum. These data thus revealed that large somatic expansions occur in early in the disease time-course, before the onset of overt motor symptoms. These data also suggested that the regional specificity of somatic expansion may more closely follow the regional specificity of the disease earlier in the disease course, and in individuals with smaller inherited alleles. These data also suggested that neurodegeneration may selectively target striatal neurons with large expansions, rather than precipitating large expansions as a secondary endophenotype of active gliosis [175].

*Of glia and neurons: HTT somatic expansions accumulate in non-dividing neurons*

Given the clear association between CAG repeat size and age at onset in humans, and CAG repeat size and cellular toxicity, and the rapidity with which germline expansions had been accepted as an expla-

nation for anticipation, it is hard in retrospect to understand why the potential significance of somatic expansion, and in particular the significance of the detection of such large somatic expansions early in the disease course of HD [175] were not more widely considered. It is thus worth contemplating what some of the drivers for this were. As alluded to above, by the early 2000s, the basic science field had broadly gone in two main directions: one focussed on downstream pathology, which in HD was essentially polyglutamine toxicity (and the vexed question of whether aggregates are causative); and, the repeat instability field that was more focussed on the mechanisms of instability. As a consequence, relatively little attention was paid to the *in vivo* consequences of somatic expansion. More directly, questions were raised as to the reliability of the observations of such large expansions in HD brains and the possibility that such large expansions may be technical artefacts of PCR. However, pre-PCR size fractionation of genomic DNA restriction fragments containing the *HTT* repeat confirmed that such large expansions were not technical artefacts [177]. The other question that was resolved was whether expansions could occur in non-dividing striatal neurons. Using laser capture microdissection in both human HD brains and knock-in animal models, it was demonstrated that: although unstable in glia, expansions were typically larger in striatal neurons; there were more expansions in striatal, rather than cortical, neurons in less advanced HD cases; and, that in mice, smaller somatic expansions occurred in the relatively well spared nitric oxide synthase-positive interneurons compared with the overall neuronal population in the striatum [178, 179].

*The road well-travelled: striatal and cortical instability is shared among many disease loci*

Another important consideration was the fact that striatal specific expansions are not limited to HD. Notably, the striatum was revealed as the region with the largest expansions in other expanded CAG•CTG repeat mouse models of DM1 [180, 181], SCA1 [182] and DRPLA [183] (see also Wheeler and Dion, this issue [169]). Likewise in humans, high levels of somatic expansion are observed in striatal and/or cortical regions in individuals with other CAG•CTG repeat expansions including DM1 [184, 185], SCA1 [186–192], SCA2 [193, 194], MJD [187, 190, 195–197], SCA7 [94, 198] and DRPLA [188, 195, 196, 199–204]. These data suggest that regional-

specific CAG•CTG repeat somatic expansion in the brain is strongly driven by major tissue-specific *trans*-acting factors. These observations have been further borne out by a recent study that revealed very similar somatic expansion profiles of the expanded CAG repeat in both HD and SCA1 across multiple brain regions [192]. Nonetheless, it is worth noting that in SBMA overall levels of mutation length variability in somatic tissues are lower than in the other polyglutamine expansion disorders [195, 196, 205–207], with more expansions in peripheral tissues such as cardiac and skeletal muscle, skin and prostate, than in the central nervous system. Overall, however, the broadly preserved pattern of somatic mosaicism in the CNS in the polyglutamine expansion disorders sheds some doubt on whether somatic expansions really drive the regional specificity of neurodegeneration observed in HD, and the other polyglutamine encoding CAG repeat expansion disorders more broadly.

*The view from the other side: regional and cell-type specificity in HD and the spinocerebellar ataxias*

In relationship to the other polyglutamine repeat expansion disorders, particular consideration deserves to be given to the contrast between the SCAs and HD. The SCAs are characterised by early cerebellar degeneration, with loss in particular of the critical cerebellar Purkinje cells. It is very notable, however, that in addition to the six different types of SCA caused by the expansion of polyglutamine encoding CAG repeats (SCA1, 2, 3, 6, 7 and 17), the SCAs caused by other simple sequence repeat expansions (SCA8, 10, 12, 36 and 37), mutations in at least 37 additional genes involved in a wide variety of different cellular pathways can also cause SCA [208]. This is in stark contrast to HD, DRPLA and SBMA, that are each caused by a single type of mutation, the CAG expansion, in only a single gene. The massive genetic heterogeneity in the SCAs reveals that cerebellar Purkinje cells must be extremely sensitive to a wide variety of cellular insults. Thus, given that *HTT* is highly expressed throughout the brain, with very high levels in the cerebellum, including in Purkinje cells [158, 209–212], one way of viewing the dichotomy between the SCAs and HD is to ask, why is HD not SCA49, and for that matter why DRPLA is not SCA50, and SBMA SCA51? It is possible that the default state might be that in the absence of a very pronounced striatal-specific expansion process, *HTT* germline expansions might elicit a late onset

SCA like phenotype. Another important consideration in this regard relates to critical role that Purkinje cells play in the cerebellum, especially considering the relatively low density and absolute of numbers of Purkinje cells. Purkinje cells are estimated to comprise less than one in a thousand cells in an intact cerebellum. Thus, it remains very possible that investigations of somatic length variability in bulk DNA analyses have failed to account appropriately for this gross disparity in relative cell ratios. Additional application of laser capture microdissection to better investigate the mutation profiles of defined cell types within the cerebellum in both HD and the SCAs are clearly warranted. Indeed, laser-dissected cerebellar cells in DRPLA patients have demonstrated that the expansion lengths in granule cells are significantly smaller than in Purkinje cells and glia [202, 203]. Additionally, caution must be applied when interpreting studies of tissue from patients who died after a long disease course—most commonly the case in the published literature cited above. In end stage-disease tissue, the residual cell profile is much altered due to the loss of vulnerable neurons and the proliferation of cells such as astrocytes. Therefore, in order to more clearly define the relationship between somatic mutation length and pathological vulnerability in polyglutamine diseases, more detailed analysis of defined cell types (isolated by techniques such as laser capture microdissection or single cell sequencing) from candidate brain regions in the rare, early disease cases where cell loss is minimal are really needed.

Nevertheless, it is also important to recognise that the question as to whether somatic expansion drives the regional specificity of neuropathology between the disorders, is entirely separable from the question as to whether somatic expansion drives onset and progression in any one disorder. A striatal-specific expansion process may not delineate HD from the SCAs, but it is unlikely to be helping in HD, and cerebellar expansions in the SCAs, even if they occur at an overall much lower rate than in the striatum, may still be critical since cerebellar Purkinje cells are acutely sensitive to perturbation.

*The drivers of instability: somatic expansion is cell-division independent and mismatch repair dependent*

Shortly after the first expansion mutations were identified, attention quickly turned to what molecular mechanisms were driving instability. Not unrea-

sonably, it was widely assumed that expansions were most likely to be mediated by DNA replication slippage [154–156]. This concept appeared to be supported by observations that polymerase slippage *in vitro* could generate products with altered numbers of repeats—these data arising from experiments that dated back to the 1960s when repeating oligonucleotide tracts were being synthesised as part of efforts to crack the genetic code [213–215]. This concept was further reinforced by the observation of similar slippage products that arose during PCR of non-disease associated simple sequence microsatellite repeats [216, 217]. This concept was further reinforced when, as presaged by studies in microbes [218], in 1993 and 1994 it was shown that mutations in [post-replicative] DNA mismatch repair genes were associated with genome-wide microsatellite instability in the tumours of individuals with Lynch syndrome (a hereditary predisposition toward non-polyposis colon cancer) [219–223], the mechanism by which variation in non-disease associated microsatellites arose appeared to be firmly established as DNA replication slippage and mismatch repair avoidance. It seemed a very reasonable assumption that the expansion of disease-associated loci would be similarly mediated by DNA polymerase replication slippage errors that simply overwhelmed the DNA mismatch repair machinery.

Additionally, the fact that for most of the CAG•CTG repeat expansions disorders, the repeats were relatively stable in the female germline, and more unstable and prone to large expansions during male transmission, appeared to fit nicely to the greater number of premeiotic mitoses in spermatogenesis relative to oogenesis. Indeed, detailed single sperm analysis of male germline dynamics in HD did at least appear to partially support such a mitotic model [224]. Moreover, a premeiotic origin for at least some of the *HTT* CAG mutations in the male germline was directly established using laser capture microdissection of testicular cells [225]. However, these experiments also suggested that meiotic events were important too [225]. A primarily premeiotic replication dependent mechanism for male germline mutations would also predict a strong age effect. In comparing *HTT* CAG repeat length distributions in sperm between men, an age effect was not detected [224]. Likewise, there was no significant difference in *HTT* CAG repeat length distributions in two sperm samples from the same man obtained two years apart [224]. Cross-sectional and longitudinal small pool PCR analysis of sperm DNA variation in DM1 males

has similarly failed to find evidence for an age effect [146, 226, 227]. In addition to the absence of an obvious age effect in the male germline, these studies in HD [224] and DM1 [146, 226, 227], and similar studies in SBMA [228, 229] and SCA7 [230], have revealed substantive differences between the dynamics of expanded repeats in the male germline and those observed in somatic tissues, including: a much greater frequency of germline mutations than somatic mutations (at least in blood, and for DM1 males inheriting <80 CTG repeats); and, despite a bias toward expansions, a greater frequency of contractions in the male germline, including reversions into the non-disease associated range. Although male germline instability is observed in many expanded CAG•CTG repeat mouse models (e.g., [164–167, 231]; see also Wheeler and Dion, this issue [169]), the very frequent and very large expansions observed in the male germline in humans have not yet been faithfully mirrored in mice, especially when one considers the relatively large allele sizes with which the majority of the mouse models have been generated. It thus remains unclear to what extent the mechanism of expansion is shared between the germline and soma, and what mediates the obvious differences.

A clear prediction of the DNA replication slippage model for expansion would be that tissues with higher levels of cell turnover would show higher levels of somatic mosaicism. This does not appear to be borne out in either humans with, or animals models of, the CAG•CTG repeat expansion disorders, with somatic expansions accumulating in post-mitotic tissues such as skeletal muscle and brain (e.g., [94, 133–136, 164, 168, 178–181, 184, 185, 187–191, 193–204]; see also Wheeler and Dion, this issue [169]). Another clear prediction of the DNA replication slippage model is that loss of function mutations in the post-replicative DNA mismatch repair pathway should increase the frequency of expansions. This prediction was turned on its head in 1999, when Manley et al. demonstrated that the complete reverse was true and that the obligate mammalian MutS homologue *Msh2* was absolutely required to generate somatic expansions [232] (see Iyer and Pluciennik, this issue, for more details on the DNA mismatch repair pathway [233]). These insights were extended when it was shown that *Msh3*, but not *Msh6*, was also essential for the somatic expansion of CAG•CTG repeat expansions, directly implicating the MSH2/MSH3 MutSBeta complex [234]. One potential explanation for the requirement for MSH2 and MSH3 could have been that MutSBeta stabilises [232, 235] the slipped

strand [236] and/or hairpin [237] DNA structures that are the presumed length change intermediates in the expansion pathway. However, involvement of various downstream MutL homologues in CAG•CTG repeat expansion [238, 239], and the requirement for MSH2 ATPase activity [240], suggest instead that expansions may be mediated by an actual mismatch repair reaction of small slipped strand loop-outs in which the loop is preferentially incorporated [232, 238, 240]. In addition to candidate gene studies using knock-out mismatch repair gene null alleles, it is notable that naturally occurring mouse strain-specific differences in CAG•CTG repeat somatic expansion profiles could be detected [234, 241] and some of these associated with naturally occurring variants in *Mlh1* and *Msh3* [239, 242], presaging the identification of similarly acting human variants (see below). Cell division-independent inappropriate DNA mismatch repair has thus come to the fore as a likely mechanism of expansion of CAG•CTG repeats [238]. Along with other *in vitro* experiments, these animal model studies were critical in establishing the key players in the expansion pathway (for more details, see Wheeler and Dion, and Iyer and Pluciennik, in this issue [169, 233]). However, in most cases, these studies did not lead directly to insights into the accrual of somatic expansions in mediating pathology (although see below). Nonetheless, the identification of the key players in the expansion pathway, would later prove critical in providing an explanation for the results of the genome-wide associations studies for modifiers of age at onset in HD, and for providing suitable targets for candidate gene studies (see below).

*The journey to pathology: genetically suppressing somatic expansion in HD mice slows the accumulation of pathological hallmarks of disease*

The critical dependence of somatic expansion on functional *Msh2*, *Msh3* and *Mlh1* genes has been used to demonstrate that genetically slowing the rate of somatic expansions can also slow the rate of accumulation of pathological markers of HD such as polyglutamine aggregates [239, 243, 244]. One of the primary reasons that more definitive data directly linking somatic expansion and disease pathology in HD models has been difficult to generate relates to the length of the CAG repeat in mice necessary to generate sufficient levels of somatic expansion and/or a disease phenotype during the lifetime of a mouse, or the even shorter length of a typical project grant.

Simply put, mice with small germline CAG expansions in the range typically observed in humans do not develop an overt HD phenotype during their lifetime [245, 246]. HD mice with larger germline expansions typically beyond the length that is observed even in most juvenile HD cases (>80 repeats), can display robust HD phenotypes in a matter of months (e.g., [163, 167, 168]). In such cases, the repeat is likely already well beyond any cell-dependent toxic threshold and pathology can proceed in the absence of somatic expansion. Of course, that is not to say, as evidenced above [239, 243, 244], that somatic expansions may not exacerbate the phenotype in HD mice already inheriting large expansions. Indeed, such an effect may explain the greater frequency of HTT aggregates observed in a minimally CAA interrupted yeast artificial chromosome HD mouse model (presumed to be at least partially somatically unstable), relative to a bacterial artificial chromosome HD mouse model with a somatically stable highly interrupted polyglutamine encoding CAG/CAA repeat tract [247, 248]. Furthermore, detailed studies in mouse models with germline alleles that are small enough not to cause an overt early onset phenotype, but that have the capacity to somatically expand, are clearly warranted. Nonetheless, it remains unclear if short-lived mice can be effectively used to determine the absolute requirement for the somatic expansions that typically accumulate over 30 to 40 years in HD patients inheriting alleles of 40 to 50 CAG repeats.

*The passage of the extremes: the frequency of large cortical expansions is associated with variation in age at onset in HD*

A key prediction of a critical role for somatic expansion in HD pathology would be that individual-specific variation in the rate of somatic expansion would be reflected in individual differences in disease severity. To this end, the first human data demonstrating such a link were generated in 2009 [249]. Specifically, using small pool PCR analysis of cortical DNA from a cohort of HD individuals with extreme early or extreme late onset of symptoms relative to the age at onset predicted by the number of CAG repeats inherited, Swami et al. were able to demonstrate that the fraction of large somatic expansions, as quantified by the skewness of the repeat length distribution, was inversely associated with residual variation in age at onset, i.e., individuals with more large somatic expansions had an earlier age at onset than expected. These data, along with additional

data defining the repeat length dependence of somatic expansions in buccal cell DNA [250], demonstrated that the somatic expansion phenotype in humans is modifiable by factors other than repeat length, tissue and age. Through this period, animal model and human data implicating somatic expansion slowly accumulated, but even as late as 2018 somatic expansion was not widely viewed as a therapeutic target in HD (e.g., [251]).

## THE RETURN OF THE REPEAT

*The reckoning: genome-wide association studies of variation in HD age at onset reveal DNA repair gene variants*

A major limitation of the scientific process is the inherent biases and preconceptions that are inevitably brought into play in the design of an experiment. One of the beauties of genome-wide association studies (GWAS) are that they are completely unbiased, at least in terms of the having to make no predictions about the genes in which variants may modify the phenotype of interest. It was thus with great excitement that the HD community eagerly awaited the results of GWAS of modifiers of residual age at onset in HD. As detailed in the accompanying manuscript by Hong et al. [252], the results of the first GWAS to collate a large enough cohort of HD participants to achieve genome-wide significance levels was published by the GeM-HD consortium in 2015, and revealed genome-wide significant associations in two regions: the *FAN1* gene; and the *RRM2B* gene [253]. The *FAN1* gene encodes the Fanconi anaemia FANCD1/FANCD2-associated endonuclease 1 DNA repair gene. Although at the time *FAN1* was not known to be involved in the repeat expansion pathway, it is now clear that the levels of *FAN1* are important in mediating somatic expansion in cells [254] and animal models [255, 256] (see also Deshmukh et al. [257] and Zhao et al. [258], this issue). *RRM2B* encodes ribonucleotide reductase regulatory TP53 inducible subunit M2B and its role in modifying HD onset currently remains unknown. A third region encompassing the *MLH1* DNA mismatch repair gene almost reached genome-wide significance in the 2015 GWAS [253], and was subsequently replicated in an independent cohort [259]. *MLH1* has previously been shown in animal models to be essential for somatic expansions of the HD repeat [239]. Pathway analysis of the 2015 GWAS results also revealed that polymorphisms in DNA repair genes were overrepresented in

variants associated with age at onset in HD, including specifically DNA mismatch repair. These data thus strongly supported the contention that DNA mismatch repair processes mediate differences in HD age at onset not accounted for by inherited CAG length. Given the prior association of the mismatch repair proteins in the somatic expansion process, it seemed logical to assume that these polymorphisms mediate a role in HD pathology via a more direct role in the somatic expansion pathway [253].

*The land of light: the same variants in MSH3 are associated with disease severity in HD and DM1*

Providing an amazing example of the utility of careful longitudinal clinical characterisation of the disease phenotype, in 2017 Hensmann Moss et al. were able to demonstrate that a combined multi-phenotype CAG and age-adjusted disease progression score was able to reveal a genome-wide significant association with variants in the *MSH3* DNA mismatch repair gene in only just over two hundred HD participants in the TRACK-HD cohort [260]. These associations were further strengthened when it was revealed that some of the same variants in and around a polyproline/alanine encoding polymorphic 9 bp repeat in *MSH3* exon 1 were also associated with both somatic expansions rates in blood DNA, and residual variation in disease severity, in both HD and DM1 [261]. It is a bizarre coincidence that *MSH3* contains its own variable repeat, but the fact that *MSH3* is absolutely required for repeat expansion in animal models, suggests that the association between *MSH3* variants with variation in disease severity is not a coincidence, and that the causative *MSH3* variants modifying disease severity are acting directly through their effects on somatic expansion.

*The pure repeat: CAG repeat number, not encoded polyglutamine length best predicts HD severity*

The polymorphic *HTT* CAG that expands in HD is succeeded by an additional CAACAG cassette that also encodes glutamine such that the total length of the encoded polyglutamine tract equals the number of CAG repeats plus two in a typical *HTT* allele [61]. It has been known for many years that a subset of atypical *HTT* alleles can differ in this regard with some alleles containing a duplication of the CAACAG cassette, and some alleles lacking this cassette completely (see Hong et al. [252], this issue,

and Ciosi et al. [262]). These variants can give rise to CAG sizing errors as estimated using fragment length analysis, and recent high-throughput DNA sequencing analyses by Ciosi et al. have revealed that failure to take these sizing errors into account can yield highly atypical genotype to phenotype associations in HD [262]. Similar effects were also reported by Wright et al. [263] and in the latest results from the GeM HD Consortium GWAS for modifiers of age at onset in HD [264] (see Hong et al., this issue [252]). An early observation after the disease-causing mutations were first identified, was that in several disorders non-disease associated alleles were interrupted with stabilising variant repeats, whilst genetically unstable disease-causing expansions were pure (e.g., [186, 265–268]). In the DM1, the vast majority of non-disease associated alleles are pure CTG, as are most disease-causing expansions. However, a subset of approximately 5% of DM1 disease-causing expansions are interrupted by primarily CCG variant repeats [269, 270]. In addition to being genetically more stable in both the germline and soma, such alleles are typically associated with delayed onset and/or milder DM1 symptoms, thus linking variant repeat interruptions with increased somatic stability, and further linking somatic instability with disease onset [150, 269–274]. Thus, the most logical explanation for the greater predictive value of the number pure CAG repeats, rather than total glutamine number encoded, is that it is pure CAG number that drives somatic expansion and ultimately disease onset/progression. Indeed, pure CAG length accurately predicts the relative ratio of somatic expansions observed in the blood DNA of HD individuals [262], effects that appear to be mirrored in the male germline [275]. However, it should be noted that whilst correcting for pure CAG more accurately predicts age at onset and disease progression than does the number of glutamines encoded, individuals lacking the CAACAG cassette on their mutant chromosome still tend to have an earlier age at onset than expected. Likewise, individuals with the CAACAG cassette duplication on their mutant chromosome do not have a worse disease course, despite the fact they inherit alleles expressing two additional supposedly toxic glutamine codons relative to individuals with a typical expanded allele [262–264]. The reasons for these residual effects remain unknown, but could include additional effects on somatic expansion in the brain not detectable in blood DNA, or some other effect on *HTT* transcription [276], RNA stability and/or translation efficiency

for instance [252, 262, 277, 278]. Alternatively, it is possible that the ultimate pathogenic moiety in HD is not the canonical polyglutamine containing HTT protein, but some directly toxic effect of the *HTT* CAG RNA [279], an alternative truncated transcript [280] or a repeat associated non-ATG translation product [281].

*More of the same: GWAS reveals even more DNA mismatch repair gene variants associated with variation in HD age at onset*

In 2019 the GeM HD Consortium revealed the results of the latest GWAS for modifiers of age at onset in HD incorporating just over 9,000 participants [252, 264]. In addition to further confirming associations with *FAN1*, *MLH1* and *MSH3*, these data elevated variants in the *PMS2* and *PMS1* DNA mismatch repair genes, in addition to the *LIG1* DNA ligase gene, also required to complete a DNA mismatch repair reaction, to genome-wide significance. These data further highlight the critical role that DNA repair gene variants have in mediating symptomatic variation in HD, most likely through their action on somatic expansion [264].

*The mechanistic bridge: DNA repair gene variants are associated with somatic expansion scores in HD blood DNA*

The high-throughput *HTT* sequencing assay developed by Ciosi et al. also allows for the quantification of the relative ratio of somatic expansions in blood DNA (see Ciosi et al., this issue for comparison of approaches to quantifying somatic mosaicism in HD) [282], that after correcting for age and CAG length effects, results in an individual-specific somatic expansion score [262]. As expected, assuming that somatic expansion profiles in blood DNA at least broadly parallel those in the brain, the somatic expansion score was inversely associated with variation in age at onset and positively associated with individual-specific disease progression scores (i.e., individuals with a faster rate of somatic expansion have an earlier age at onset than expected and more rapid disease course). The somatic expansion score is also a molecular phenotype that can be used for association studies to reveal genetic modifiers of the expansion process. Indeed Ciosi et al. have used this phenotype to reveal direct associations between somatic expansion and variants in the *FAN1*, *MLH1*, *MLH3* and *MSH3* DNA repair genes in a candidate

gene analysis [262]. As discussed, *FAN1*, *MLH1* and *MSH3* have already been implicated as modifying age at onset and disease severity in HD by GWAS analyses [253, 260, 264]. Whilst variants in *MLH3* have not yet been significantly associated with variation in HD age at onset, the latest HD GWAS data indicate a nominal association of  $p = 0.0001$  [264]. Given the essential requirement for *MLH3* in mediating somatic expansions in HD mice [239], it seems a reasonable supposition that *MLH3* will reach genome-wide significance with a larger cohort. Nonetheless, the *FAN1*, *MLH1* and *MSH3* data already provide a mechanistic link between the HD age at onset modifiers, and direct modifiers of somatic expansion, that lends further credence to the model proposing somatic expansion as a key driver of disease pathology in HD.

*The last debate: is somatic expansion required and does size matter?*

It would now appear that the concept that somatic expansion contributes toward disease onset in HD is beyond reasonable doubt—size changes, and it clearly does matter (although see Maiuri et al., this issue, for alternative hypotheses linking DNA repair and HD [283]). More pertinent now is to consider what are the actual critical products of the somatic expansion process? Are the massive somatic expansions of hundreds or even thousands of repeats actually required? The U-shaped disease severity curve observed in the R6/2 mice where germline expansions beyond 300 CAG repeats become protective [284–286], albeit in transgenic model expressing a protein fragment, are nonetheless very intriguing. Such large expansions are at least partially hypomorphic [284–286]. However, if they were truly protective, then we might expect cells carrying such large expansions to accumulate in end-stage disease. This they clearly do not do [175]. Indeed, at any one point cells carrying such large expansions are relatively rare. However, this is exactly what might be expected if such cells have only a very short half-life and exist only transiently in such an elevated state in HD brains. It is nonetheless possible that such very large expansions may be something of a red herring. Given that transmitting an allele only one CAG repeat longer results in at least a two-year decrease in the age at onset of HD, it seems not unreasonable that even somatic gains of one or two repeats are certainly not helping and almost certainly making things worse. Some clues as to the answer to this question may come

from asking an even more fundamental question as to whether somatic expansion is absolutely required to generate HD pathology. If germline expansions in the range 40 to 50 CAG repeats are inherently toxic and capable of precipitating symptoms in the absence of somatic expansion, then that would suggest that even small changes would accelerate pathology in a meaningful way. If, however, a toxicity threshold exists at some larger size, then somatic expansion may indeed be absolutely required (see Donaldson et al., this issue, for further discussion of where the toxic CAG threshold may actually lie [287]). The concept that somatic expansion is required in HD, may be supported by the observation that individuals homozygous for HD expansions (the vast majority of whom are compound heterozygotes with disease-causing expansions of two different sizes), do not appear to have an earlier age at onset than that predicted by the larger of their two alleles [288, 289]. One possible explanation for the fully dominant nature of HD onset, supported by a modelling approach, is that onset is achieved when a particular fraction of cells somatically expand the CAG repeat beyond a higher pathological threshold [290]. As somatic expansion is highly repeat length dependent, this threshold will, in the majority of cells, be achieved first by the larger of the two inherited alleles, and hence it is the larger allele that predicts disease onset [290]. Additional insight into the issue of whether somatic expansion is actually required, could also be provided by the identification of individuals with repeat stabilising interruptions in the middle of an expanded *HTT* CAG array. For instance, if an individual with a 45 repeat allele with a single glutamine encoding variant CAA repeat (e.g., (CAG)<sub>22</sub>CAA(CAG)<sub>22</sub>) was identified, and assuming, as expected such an allele would be somatically stable, then if they were affected it would indicate that somatic expansion is not required. If however they remain asymptomatic throughout their life, then it would indicate somatic expansion is absolutely required and the true pathological threshold at the cellular level is greater than that required to be inherited (at least for a somatically unstable pure CAG repeat). Such individuals are at the least very rare in either the HD or general population, if they exist at all. However, the availability of a high-throughput *HTT* sequencing assay [262], and an improved ability to genotype simple repeats from whole-genome sequencing data [291, 292] suggest that this question is closer to being answered than it ever has been.

### *One rule to bind them: trans-modifiers of somatic expansion in the other repeat expansion disorders*

As somatic expansion is a common theme in many of the repeat expansion disorders, it seems not unreasonable to assume that somatic expansion might similarly drive disease onset in some of these disorders too. Indeed, it was established in 2012 that individual-specific variation in the rate of somatic expansion in blood DNA was inversely correlated with residual variation in age at onset in DM1 [148]. Moreover, individual-specific variation in somatic expansion was shown to be inherited as quantifiable trait consistent with an underlying genetic mechanism [148]. These data were further borne out with the demonstration in 2016 that using a candidate DNA mismatch repair gene study, variants in the *MSH3* DNA mismatch repair gene were associated directly with variation in the individual-specific rate of somatic expansion in DM1 [151]. Similarly, preliminary evidence has been generated that variants in *FAN1* and the *PMS2* DNA mismatch repair gene are also associated with residual variation in age at onset of some of the other polyglutamine encoding CAG repeat expansion disorders, including SCA1 [293]. Whilst most of the data relating directly to the role of somatic expansion in disease aetiology in humans relates to the CAG•CTG repeat disorders, there is also evidence for considerable somatic instability in Friedreich ataxia (FA) [294–299]. However, the very strong bias toward net somatic expansions observed in the CAG•CTG repeat disorders, is not seen in FA where a high frequency of somatic contractions of the GAA repeat are also observed [295–299]. Nevertheless, somatic mosaicism continues throughout life, and somatic expansions accumulate in the dorsal root ganglia, one of the primary affected tissues, suggesting somatic mosaicism may play an important part in FA [298, 299]. Moreover, it appears many of the same mismatch repair proteins are implicated in somatic instability of GAA repeats in both human cells and animal models [300–306]. Mismatch repair proteins have been similarly implicated in expansion of the CGG repeat in FXS mouse models [307–312] (for more details, see Zhao et al., this issue [258]). Given the massive cross-disorder potential of therapies aimed at suppressing somatic expansion (see Benn et al., this issue [313]), additional insights into the dynamics and mechanisms of somatic expansion in the other repeat expansion disorder are required. Suitably powered large-scale GWAS of modifiers

of somatic expansion and disease severity in the other repeat expansion disorder should be particularly informative.

*The houses of healing: somatic expansion as a therapeutic target*

In addition to inherent questions of fundamental biological interest, the ultimate goal of understanding the role of somatic expansion in HD is to evaluate it as a possible therapeutic target. The data are now clear that somatic expansion at the very least exacerbates disease onset and progression, and as such somatic expansion has to be considered a very real therapeutic target worthy of further scrutiny. Suppressing somatic expansion would be expected to be therapeutically beneficial. Even more enticing is the prospect of modulating repeat instability in such a way so as to be able to elicit somatic repeat contractions. If enabled early enough, repeat contractions raise the prospect of being not just beneficial, but potentially curative. In this light, the recent data that CAG repeat somatic instability can be modified by small molecules in HD animal models and appears to be associated with reduced markers of disease is exceptionally exciting [314–317]. However, I will leave Benn et al., this issue [313], to delve deeper into this topic, and discuss some of the technical challenges involved, and leave you here with the idea that our journey has reached a significant milestone, but it has not ended. Much remains to be done to translate these findings to new treatments for HD and related disorders, but the establishment of somatic expansion as contributory to HD pathology, and the identification of potential enzymatic targets, opens up some exciting possibilities.

## ACKNOWLEDGMENTS

The author would like to thank all those whose considerable efforts have gotten us to the point where somatic expansions is more broadly recognised as an important contributory factor in HD pathology and a *bona fide* therapeutic target. The author would also like to apologise in advance to all those whose important contributions, particularly with regard to insights into the expansion mechanism, have not been duly acknowledged in this review. Work in the D.G.M group is supported by an award from the CHDI Foundation.

## CONFLICT OF INTEREST

D.G.M. has been a scientific consultant and/or received honoraria or stock options from Biogen Idec, AMO Pharma, Charles River, Vertex Pharmaceuticals, Triplet Therapeutics, LoQus23, and Small Molecule RNA and has had research contracts with AMO Pharma and Vertex Pharmaceuticals.

## REFERENCES

- [1] Yu S, Pritchard M, Kremer E, Lynch M, Nancarrow J, Baker E, et al. Fragile X genotype characterized by an unstable region of DNA. *Science*. 1991;252(5010):1179-81.
- [2] Oberle I, Rousseau F, Heitz D, Kretz C, Devys D, Hanauer A, et al. Instability of a 550 base-pair DNA segment and abnormal methylation in fragile X syndrome. *Science*. 1991;252:1097-102.
- [3] Verkerk AJMH, Pieretti M, Sutcliffe JS, Fu Y-H, Kuhl DPA, Pizzuti A, et al. Identification of a gene (FMR-1) containing a CGG repeat coincident with a breakpoint cluster region exhibiting length variation in fragile X syndrome. *Cell*. 1991;65:905-14.
- [4] Kremer EJ, Pritchard M, Lynch M, Yu S, Holman K, Baker E, et al. Mapping of DNA instability at the fragile X to a trinucleotide repeat sequence p(CCG)n. *Science*. 1991;252:1711-4.
- [5] Lubs HA. A marker X chromosome. *Am J Med Genet*. 1969;17:133-44.
- [6] Sutherland GR. Heritable fragile sites on human chromosomes I. Factors affecting expression in lymphocyte culture. *Am J Hum Genet*. 1979;31(2):125-35.
- [7] Fu Y-H, Kuhl DPA, Pizzuti A, Pieretti M, Sutcliffe JS, Richards S, et al. Variation of the CGG repeat at the fragile X site results in genetic instability: Resolution of the Sherman paradox. *Cell*. 1991;67:1047-58.
- [8] Martin JP, Bell J. A pedigree of mental defect showing sex linkage. *J Neurol Neurosurg Psychiatry*. 1943;6:154-7.
- [9] Sherman SL, Jacobs PA, Morton NE, Froster-Iskenius U, Howard-Peebles PN, Nielsen KB, et al. Further segregation analysis of the fragile X syndrome with special reference to transmitting males. *Hum Genet*. 1985;69:3289-99.
- [10] Sherman SL, Morton NE, Jacobs PA, Turner G. The marker (X) syndrome: Cytogenetic and genetic analysis. *Hum Genet*. 1984;48:21-37.
- [11] Opitz JM. On the gates of hell and a most unusual gene. *Am J Med Genet*. 1986;23(1-2):1-10. doi: 10.1002/ajmg.1320230102
- [12] Laird CD. Proposed mechanism of inheritance and expression of the human fragile-X syndrome of mental retardation. *Genetics*. 1987;117(3):587-99.
- [13] Wyman A, White R. A highly polymorphic locus in human DNA. *Proc Natl Acad Sci U S A*. 1980;77:6754-8.
- [14] Jeffreys AJ, Wilson V, Thein SL. Individual-specific fingerprints of human DNA. *Nature*. 1985;316(6023):76-9.
- [15] Jeffreys AJ, Wilson V, Thein SL. Hypervariable minisatellite regions in human DNA. *Nature*. 1985;314(6006):67-73.
- [16] Jeffreys AJ, Royle NJ, Wilson V, Wong Z. Spontaneous mutation rates to new length alleles at tandem-

- repetitive hypervariable loci in human DNA. *Nature*. 1988;332(6161):278-81. doi: 10.1038/332278a0
- [17] La Spada AR, Wilson EM, Lubahn DB, Harding AE, Fischbeck KH. Androgen receptor gene mutations in X-linked spinal and bulbar muscular atrophy. *Nature*. 1991;352:77-9.
- [18] Biancalana V, Serville F, Pommier J, Julien J, Hanauer A, Mandel JL. Moderate instability of the trinucleotide repeat in spino bulbar muscular atrophy. *Hum Mol Genet*. 1992;1:255-8.
- [19] La Spada AR, Roling DB, Harding AE, Warner CL, Spiegel R, Hausmanowa-Petrusewicz I, et al. Meiotic stability and genotype-phenotype correlation of the trinucleotide repeat in X-linked spinal and bulbar muscular atrophy. *Nat Genet*. 1992;2:301-4.
- [20] Lucas P. *Traité philosophique et physiologique de l'hérédité naturelle dans les états de santé et de maladie du système nerveux*. Volumes 1 and 2. Paris: Chez J.B. Baillié; 1847-1850.
- [21] Morel BA. *Traité des dégénérescences*. Paris: Chez J.B. Baillié; 1857.
- [22] Mott FW. Heredity and insanity. *Eugen Rev*. 1911;2(4):257-81.
- [23] Pearson K. On an apparent fallacy in the statistical treatment of "antedating" in the inheritance of pathological conditions. *Nature*. 1912;40:334.
- [24] Heron D. An examination of some recent studies of the inheritance factor in insanity. *Biometrika*. 1914;10:356-83.
- [25] Heilbronner E. Ueber eine art progressiver hereditiit bei Huntington'scher chorea. *Arch Psychiatr Nervenkr*. 1903;36:889-94.
- [26] Lange F. Ueber chronische progressive chorea (Huntington) im jugendlichen alter. *Berl Klin Wochenschr*. 1906;43:153-6.
- [27] Curschmann H. Eine neue chorea-Huntington-familie. *Dtsch Z Nervenheilk*. 1908;35:293-305.
- [28] Davenport CB. Huntington's chorea in relation to heredity and eugenics. *Proc Natl Acad Sci U S A*. 1915;1:283-5.
- [29] Fleischer B. Über myotonische dystrophie mit katarakt. Eine hereditäre familiäre degeneration. *Albrecht Von Graefes Arch Ophthalmol*. 1918;96:91-133. doi: 10.1007/BF02018704
- [30] Adie WJ, Greenfield JG. *Dystrophia myotonica (myotonia atrophica)*. *Brain*. 1923;46:73-127.
- [31] Maas O. Observations on *dystrophia myotonica*. *Brain*. 1937;60:498-524. doi: 10.1093/brain/60.4.498
- [32] Henke K, Seeger S. Ueber die vererbung der myotonischen dystrophie. *Z Konstitutionsl*. 1927;13:371-415.
- [33] Ravin A, Waring JJ. Studies in dystrophia myotonica. I. Hereditary aspects. *Am J Med Sci*. 1939;197:593-609.
- [34] Bell J. *Dystrophia myotonica and allied diseases*. In: Penrose LS, editor. *Treasury of Human Inheritance*. Cambridge: Cambridge University Press; 1947. pp. 343-410.
- [35] Penrose LS. The problem of anticipation in pedigrees of *dystrophia myotonica*. *Ann Eugen*. 1948;14:125-32.
- [36] Harper PS, Harley HG, Reardon W, Shaw DJ. Anticipation in myotonic dystrophy: New light on an old problem. *Am J Hum Genet*. 1992;51:10-6.
- [37] Friedman JE. *Coming full circle: The development, rise, fall, and return of the concept of anticipation in hereditary disease*: University of Victoria.
- [38] Höweler CJ, Busch HFM, Geraedts JPM, Niermeijer MF, Staal A. Anticipation in myotonic dystrophy: Fact or fiction? *Brain*. 1989;112:779-97. doi: 10.1093/brain/112.3.779
- [39] Donachie J, Monckton DG. *Tomorrow belongs to me*. Glasgow, UK: University of Glasgow; 2006.
- [40] Harper PS. *Myotonic dystrophy*. 2nd ed. London: WB Saunders Co.; 1989.
- [41] Sutherland GR, Haan EA, Kremer E, Lynch M, Pritchard M, Yu S, et al. Hereditary unstable DNA: A new explanation for some old genetic questions? *Lancet*. 1991;338(8762):289-92. doi: 10.1016/0140-6736(91)90426-p
- [42] Buxton J, Shelbourne P, Davies J, Jones C, Tongeren TV, Aslanidis C, et al. Detection of an unstable fragment of DNA specific to individuals with myotonic dystrophy. *Nature*. 1992;355:547-8. doi: 10.1038/355547a0
- [43] Aslanidis C, Jansen G, Amemiya C, Shutler G, Mahadevan M, Tsilfidis C, et al. Cloning of the essential myotonic dystrophy region and mapping of the putative defect. *Nature*. 1992;355:548-50. doi: 10.1038/355548a0
- [44] Harley HG, Brook JD, Rundle SA, Crow S, Reardon W, Buckler AJ, et al. Expansion of an unstable DNA region and phenotypic variation in myotonic dystrophy. *Nature*. 1992;355:545-6. doi: 10.1038/355545a0
- [45] Brook JD, McCurrach ME, Harley HG, Buckler AJ, Church D, Aburatani H, et al. 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*. 1992;68:799-808. doi: 10.1016/0092-8674(92)90154-5
- [46] Fu YH, Pizzuti A, Fenwick RG, King J, Rajnarayan S, Dunne PW, et al. An unstable triplet repeat in a gene related to myotonic muscular dystrophy. *Science*. 1992;255:1256-8. doi: 10.1126/science.1546326
- [47] Mahadevan M, Tsilfidis C, Sabourin L, Shutler G, Amemiya C, Jansen G, et al. Myotonic dystrophy mutation: An unstable CTG repeat in the 3' untranslated region of the gene. *Science*. 1992;255:1253-5. doi: 10.1126/science.1546325
- [48] Newcombe RG, Walker DA, Harper PS. Factors influencing age at onset and duration of survival in Huntington's chorea. *Ann Hum Genet*. 1981;45(Pt 4):387-96.
- [49] Myers RH, Madden JJ, Teague JL, Falek A. Factors related to onset age of Huntington disease. *Am J Hum Genet*. 1982;34(3):481-8.
- [50] Ridley RM, Frith CD, Crow TJ, Conneally PM. Anticipation in Huntington's disease is inherited through the male line but may originate in the female. *J Med Genet*. 1988;25(9):589-95.
- [51] Bird ED, Caro AJ, Pilling JB. A sex related factor in the inheritance of Huntington's chorea. *Ann Hum Genet*. 1974;37:255-60.
- [52] Myers RH, Cupples LA, Schoenfeld M, D'Agostino RB, Terrin NC, Goldmakher N, et al. Maternal factors in onset of Huntington disease. *Am J Hum Genet*. 1985;37(3):511-23.
- [53] Laird CD. Proposed genetic basis of Huntington's disease. *Trends Genet*. 1990;6(8):242-7. doi: 10.1016/0168-9525(90)90206-1
- [54] Erickson RP. Chromosomal imprinting and the parent transmission specific variation in expressivity of Huntington disease. *Am J Hum Genet*. 1985;37(4):827-9.
- [55] Reik W. Genomic imprinting: A possible mechanism for the parental origin effect in Huntington's chorea. *J Med Genet*. 1988;25(12):805-8. doi: 10.1136/jmg.25.12.805

- [56] Boehnke M, Conneally PM, Lange K. Two models for a maternal factor in the inheritance of Huntington disease. *Am J Hum Genet.* 1983;35(5):845-60.
- [57] Irwin CC, Wexler NS, Young AB, Ozelius LJ, Penney JB, Shoulson I, et al. The role of mitochondrial DNA in Huntington's disease. *J Mol Neurosci.* 1989;1(2):129-36. doi: 10.1007/BF02896896
- [58] Farrer LA, Conneally PM, Yu PL. The natural history of Huntington disease: Possible role of "aging genes". *Am J Med Genet.* 1984;18(1):115-23. doi: 10.1002/ajmg.1320180115
- [59] Vegter-van der Vlis M, Volkers WS, Went LN. Ages of death of children with Huntington's chorea and of their affected parents. *Ann Hum Genet.* 1976;39(3):329-34.
- [60] Caskey CT, Pizzuti A, Fu YH, Fenwick RG, Jr., Nelson DL. Triplet repeat mutations in human disease. *Science.* 1992;256(5058):784-9. doi: 10.1126/science.1589758
- [61] The Huntington's Disease Collaborative Research Group. A novel gene containing a trinucleotide repeat that is expanded and unstable on Huntington's disease chromosomes. *Cell.* 1993;72:971-83. doi: 10.1016/0092-8674(93)90585-e
- [62] Duyao M, Ambrose C, Myers R, Novelletto A, Perischetti F, Frontali M, et al. Trinucleotide repeat length instability and age of onset in Huntington's disease. *Nat Genet.* 1993;4:387-92. doi: 10.1038/ng0893-387
- [63] Telenius H, Kremer HPH, Theilmann J, Andrew SE, Almqvist E, Anvret M, et al. Molecular analysis of juvenile Huntington disease: The major influence on (CAG)<sub>n</sub> repeat length is the sex of the affected parent. *Hum Mol Genet.* 1993;2(10):1535-40. doi: 10.1093/hmg/2.10.1535
- [64] Barron LH, Warner JP, Porteous M, Holloway S, Simpson S, Davidson R, et al. A study of the Huntington's disease associated trinucleotide repeat in the Scottish population. *J Med Genet.* 1993;30(12):1003-7. doi: 10.1136/jmg.30.12.1003
- [65] Zuhlke C, Riess O, Bockel B, Lange H, Thies U. Mitotic stability and meiotic variability of the (CAG)<sub>n</sub> repeat in the Huntington disease gene. *Hum Mol Genet.* 1993;2(12):2063-7. doi: 10.1093/hmg/2.12.2063
- [66] De Rooij KE, De Koning Gans PA, Skraastad MI, Bel-froid RD, Vegter-Van Der Vlis M, Roos RA, et al. Dynamic mutation in Dutch Huntington's disease patients: Increased paternal repeat instability extending to within the normal size range. *J Med Genet.* 1993;30(12):996-1002. doi: 10.1136/jmg.30.12.996
- [67] Sutherland GR, Richards RI. Anticipation legitimized: Unstable DNA to the rescue. *Am J Hum Genet.* 1992;51(1):7-9.
- [68] Pulst SM, Nechiporuk A, Starkman S. Anticipation in spinocerebellar ataxia type 2. *Nat Genet.* 1993;5(1):8-10. doi: 10.1038/ng0993-8c
- [69] Sano A, Yamauchi N, Kakimoto Y, Komure O, Kawai J, Hazama F, et al. Anticipation in hereditary dentatorubral-pallidolusian atrophy. *Hum Genet.* 1994;93:699-702.
- [70] Holmberg M, Johansson J, Forsgren L, Heijbel J, Sandgren O, Holmgren G. Localization of autosomal dominant cerebellar ataxia associated with retinal degeneration and anticipation to chromosome 3p12-p21.1. *Hum Mol Genet.* 1995;4(8):1441-5.
- [71] Deighton C, Heslop P, McDonagh J, Walker D, Thomson G. Does genetic anticipation occur in familial rheumatoid arthritis? *Ann Rheum Dis.* 1994;53(12):833-5. doi: 10.1136/ard.53.12.833
- [72] Nylander PO, Engstrom C, Chotai J, Wahlstrom J, Adolfsson R. Anticipation in Swedish families with bipolar affective disorder. *J Med Genet.* 1994;31(9):686-9. doi: 10.1136/jmg.31.9.686
- [73] Engstrom C, Thornlund AS, Johansson EL, Langstrom M, Chotai J, Adolfsson R, et al. Anticipation in unipolar affective disorder. *J Affect Disord.* 1995;35(1-2):31-40. doi: 10.1016/0165-0327(95)00035-1
- [74] Bassett AS, Honer WG. Evidence for anticipation in schizophrenia. *Am J Hum Genet.* 1994;54(5):864-70.
- [75] Scheffer IE, Jones L, Pozzebon M, Howell RA, Salting MM, Berkovic SF. Autosomal dominant rolandic epilepsy and speech dyspraxia: A new syndrome with anticipation. *Ann Neurol.* 1995;38(4):633-42. doi: 10.1002/ana.410380412
- [76] Loyd JE, Butler MG, Foroud TM, Conneally PM, Phillips JA, 3rd, Newman JH. Genetic anticipation and abnormal gender ratio at birth in familial primary pulmonary hypertension. *Am J Respir Crit Care Med.* 1995;152(1):93-7. doi: 10.1164/ajrcm.152.1.7599869
- [77] Morrison AW. Anticipation in Meniere's disease. *J Laryngol Otol.* 1995;109(6):499-502.
- [78] McInnis MG, McMahon FJ, Chase GA, Simpson SG, Ross CA, DePaulo JR, Jr. Anticipation in bipolar affective disorder. *Am J Hum Genet.* 1993;53(2):385-90.
- [79] Asherson P, Owen M. Anticipation in mental illness. *Am J Hum Genet.* 1994;54(2):386-7.
- [80] McInnis MG, McMahon FJ, Stine OC, Simpson SG, Ross CA, DePaulo JR. Reply to Asherson and Owen. *Am J Hum Genet.* 1994;54:387.
- [81] Petronis A, Sherrington R, Kennedy JL. Regression to the mean does not exclude anticipation and unstable DNA disease. *Am J Hum Genet.* 1994;55(3):589-92.
- [82] Owen M, Asherson P. Reply to Petronis et al.,. *Am J Hum Genet.* 1994;55:590-2.
- [83] McInnis MG, McMahon FJ, Stine OC, Ross CA. Reply to Petronis et al. *Am J Hum Genet.* 1994;55:590-2.
- [84] Hodge SE, Wickramaratne P. Statistical pitfalls in detecting age-of-onset anticipation: The role of correlation in studying anticipation and detecting ascertainment bias. *Psychiatr Genet.* 1995;5(1):43-7. doi: 10.1097/00041444-199521000-00007
- [85] Huang J, Vieland V. A new statistical test for age-of-onset anticipation: Application to bipolar disorder. *Genet Epidemiol.* 1997;14(6):1091-6.
- [86] Tashima K, Ando Y, Tanaka Y, Uchino M, Ando M. Change in the age of onset in patients with familial amyloidotic polyneuropathy type I. *Intern Med.* 1995;34(8):748-50. doi: 10.2169/internalmedicine.34.748
- [87] Orr HT, Chung M-Y, Banfi S, Kwiatkowski Jr. TJ, Servadio A, Beaudet AL, et al. Expansion of an unstable CAG repeat in spinocerebellar ataxia type 1. *Nat Genet.* 1993;4:221-6.
- [88] Kawaguchi Y, Okamoto T, Taniwaki M, Aizawa M, Inoue M, Katayama S, et al. CAG expansions in a novel gene for Machado-Joseph disease at chromosome 14q32.1. *Nat Genet.* 1994;8:221-7.
- [89] Cancel G, Abbas N, Stevanin G, Dürr A, Chneiweiss H, Néri C, et al. Marked phenotypic heterogeneity associated with expansion of a CAG repeat sequence at the spinocerebellar ataxia 3/Machado-Joseph disease locus. *Am J Hum Genet.* 1995;57:809-16.
- [90] Koide R, Ikeuchi T, Onodera O, Tanaka H, Igarashi S, Endo K, et al. Unstable expansion of CAG repeat in hered-

- itary dentatorubral-pallidolouysian atrophy (DRPLA). *Nat Genet.* 1994;6:9-13.
- [91] Nagafuchi S, Yanagisawa H, Sato K, Shirayama T, Ohsaki E, Bundo M, et al. Dentatorubral and pallidolouysian atrophy expansion of an unstable CAG trinucleotide on chromosome 12. *Nat Genet.* 1994;6:14-8.
- [92] Imbert G, Saudou F, Yvert G, Devys D, Trottier Y, Garnier JM, et al. Cloning of the gene for spinocerebellar ataxia 2 reveals a locus with high sensitivity to expanded CAG/glutamine repeats. *Nat Genet.* 1996;14(3):285-91.
- [93] Pulst SM, Nechiporuk A, Nechiporuk T, Gispert S, Chen XN, Lopes-Cendes I, et al. Moderate expansion of a normally biallelic trinucleotide repeat in spinocerebellar ataxia type 2. *Nat Genet.* 1996;14(3):269-76. doi: 10.1038/ng1196-269.
- [94] David G, Abbas N, Stevanin G, Dürr A, Yvert G, Cancel G, et al. Cloning of the SCA7 gene reveals a highly unstable CAG repeat expansion. *Nat Genet.* 1997;17(1):65-70.
- [95] Nakamura K, Jeong SY, Uchiyama T, Anno M, Nagashima K, Nagashima T, et al. SCA17, a novel autosomal dominant cerebellar ataxia caused by an expanded polyglutamine in TATA-binding protein. *Hum Mol Genet.* 2001;10(14):1441-8.
- [96] Gusella JF, MacDonald ME. Molecular genetics: Unmasking polyglutamine triggers in neurodegenerative disease. *Nature Rev Neurosci.* 2000;1(2):109-15.
- [97] Koob MD, Moseley ML, Schut LJ, Benzow KA, Bird TD, Day JW, et al. An untranslated CTG expansion causes a novel form of spinocerebellar ataxia (SCA8). *Nat Genet.* 1999;21(4):379-84.
- [98] Ikeda Y, Daughters RS, Ranum LP. Bidirectional expression of the SCA8 expansion mutation: One mutation, two genes. *Cerebellum.* 2007;1-9.
- [99] Vincent JB, Neves-Pereira ML, Paterson AD, Yamamoto E, Parikh SV, Macciardi F, et al. An unstable trinucleotide-repeat region on chromosome 13 implicated in spinocerebellar ataxia: A common expansion locus. *Am J Hum Genet.* 2000;66(3):819-29.
- [100] Stevanin G, Herman A, Dürr A, Jodice C, Frontali M, Agid Y, et al. Are (CTG)(n) expansions at the SCA8 locus rare polymorphisms? *Nat Genet.* 2000;24(3):213.
- [101] Moseley ML, Schut LJ, Bird TD, Koob MD, Day JW, Ranum LP. SCA8 CTG repeat: En masse contractions in sperm and intergenerational sequence changes may play a role in reduced penetrance. *Hum Mol Genet.* 2000;9(14):2125-30.
- [102] Izumi Y, Maruyama H, Oda M, Morino H, Okada T, Ito H, et al. SCA8 repeat expansion: Large CTA/CTG repeat alleles are more common in ataxic patients, including those with SCA6. *Am J Hum Genet.* 2003;72(3):704-9. doi: 10.1086/367775
- [103] Mosemiller AK, Dalton JC, Day JW, Ranum LP. Molecular genetics of spinocerebellar ataxia type 8 (SCA8). *Cytogenet Genome Res.* 2003;100(1-4):175-83. doi: 10.1159/000072852
- [104] Silveira I, Alonso I, Guimaraes L, Mendonca P, Santos C, Maciel P, et al. High germinal instability of the (CTG)n at the SCA8 locus of both expanded and normal alleles. *Am J Hum Genet.* 2000;66(3):830-40. doi: 10.1086/302827
- [105] Knight SJL, Flannery AV, Hirst MC, Campbell L, Christodoulou Z, Phelps SR, et al. Trinucleotide repeat amplification and hypermethylation of a CpG island in FRAXE mental retardation. *Cell.* 1993;74:127-34.
- [106] Campuzano V, Montermini L, Moltò MD, Pianese L, Cossée M, Cavalcanti F, et al. Friedreich's ataxia: Autosomal recessive disease caused by an intronic GAA triplet repeat expansion. *Science.* 1996;271:1423-7.
- [107] Holmes SE, O'Hearn EE, McInnis MG, Gorelick-Feldman DA, Kleiderlein JJ, Callahan C, et al. Expansion of a novel CAG trinucleotide repeat in the 5' region of PPP2R2B is associated with SCA12. *Nat Genet.* 1999;23(4):391-2.
- [108] Margolis RL, O'Hearn E, Rosenblatt A, Willour V, Holmes SE, Franz ML, et al. A disorder similar to Huntington's disease is associated with a novel CAG repeat expansion. *Ann Neurol.* 2001;50(6):373-80. doi: 10.1002/ana.1312
- [109] Holmes SE, O'Hearn E, Rosenblatt A, Callahan C, Hwang HS, Ingersoll-Ashworth RG, et al. A repeat expansion in the gene encoding junctophilin-3 is associated with Huntington disease-like 2. *Nat Genet.* 2001;29(4):377-8.
- [110] Lalioti MD, Scott HS, Buresi C, Rossier C, Bottani A, Morris MA, et al. Dodecamer repeat expansion in cystatin B gene in progressive myoclonus epilepsy. *Nature.* 1997;386(6627):847-51. doi: 10.1038/386847a0
- [111] Matsuura T, Yamagata T, Burgess DL, Rasmussen A, Grewal RP, Watase K, et al. Large expansion of the ATTCT pentanucleotide repeat in spinocerebellar ataxia type 10. *Nat Genet.* 2000;26(2):191-4.
- [112] Liquori CL, Ricker K, Moseley ML, Jacobsen JF, Kress W, Naylor SL, et al. Myotonic dystrophy type 2 caused by a CCTG expansion in intron 1 of *ZNF9*. *Science.* 2001;293(5531):864-7.
- [113] Sato N, Amino T, Kobayashi K, Asakawa S, Ishiguro T, Tsunemi T, et al. Spinocerebellar ataxia type 31 is associated with "inserted" penta-nucleotide repeats containing (TGGAA)n. *Am J Hum Genet.* 2009;85(5):544-57. doi: 10.1016/j.ajhg.2009.09.019
- [114] DeJesus-Hernandez M, Mackenzie IR, Boeve BF, Boxer AL, Baker M, Rutherford NJ, et al. Expanded GGGGCC hexanucleotide repeat in noncoding region of C9ORF72 causes chromosome 9p-linked FTD and ALS. *Neuron.* 2011;72(2):245-56. doi: 10.1016/j.neuron.2011.09.011
- [115] Renton AE, Majounie E, Waite A, Simon-Sanchez J, Rollinson S, Gibbs JR, et al. A hexanucleotide repeat expansion in C9ORF72 is the cause of chromosome 9p21-linked ALS-FTD. *Neuron.* 2011;72(2):257-68. doi: 10.1016/j.neuron.2011.09.010
- [116] Kobayashi H, Abe K, Matsuura T, Ikeda Y, Hitomi T, Akechi Y, et al. Expansion of intronic GGCCTG hexanucleotide repeat in NOP56 causes SCA36, a type of spinocerebellar ataxia accompanied by motor neuron involvement. *Am J Hum Genet.* 2011;89(1):121-30. doi: 10.1016/j.ajhg.2011.05.015
- [117] Wieben ED, Aleff RA, Tosakulwong N, Butz ML, Highsmith WE, Edwards AO, et al. A common trinucleotide repeat expansion within the transcription factor 4 (TCF4, E2-2) gene predicts Fuchs corneal dystrophy. *PLoS One.* 2012;7(11):e49083. doi: 10.1371/journal.pone.0049083
- [118] Bragg DC, Mangkalaphiban K, Vaine CA, Kulkarni NJ, Shin D, Yadav R, et al. Disease onset in X-linked dystonia-parkinsonism correlates with expansion of a hexameric repeat within an SVA retrotransposon in TAF1. *Proc Natl Acad Sci U S A.* 2017;114(51):E11020-E8. doi: 10.1073/pnas.1712526114
- [119] Seixas AI, Loureiro JR, Costa C, Ordóñez-Ugalde A, Marcelino H, Oliveira CL, et al. A pentanucleotide ATTTC repeat insertion in the non-coding region of DAB1, mapping to SCA37, causes spinocerebellar ataxia. *Am J Hum Genet.* 2017;101(1):87-103. doi: 10.1016/j.ajhg.2017.06.007

- [120] LaCroix AJ, Stabley D, Sahraoui R, Adam MP, Mehafey M, Kernan K, et al. GGC Repeat expansion and exon I methylation of XYLT1 is a common pathogenic variant in Baratela-Scott Syndrome. *Am J Hum Genet.* 2019;104(1):35-44. doi: 10.1016/j.ajhg.2018.11.005
- [121] Tian Y, Wang JL, Huang W, Zeng S, Jiao B, Liu Z, et al. Expansion of human-specific GGC repeat in neuronal intranuclear inclusion disease-related disorders. *Am J Hum Genet.* 2019;105(1):166-76. doi: 10.1016/j.ajhg.2019.05.013
- [122] Cortese A, Simone R, Sullivan R, Vandrovcova J, Tariq H, Yau WY, et al. Biallelic expansion of an intronic repeat in RFC1 is a common cause of late-onset ataxia. *Nat Genet.* 2019;51(4):649-58. doi: 10.1038/s41588-019-0372-4
- [123] Ishiura H, Doi K, Mitsui J, Yoshimura J, Matsukawa MK, Fujiyama A, et al. Expansions of intronic TTCA and TTTA repeats in benign adult familial myoclonic epilepsy. *Nat Genet.* 2018;50(4):581-90. doi: 10.1038/s41588-018-0067-2
- [124] Cen Z, Jiang Z, Chen Y, Zheng X, Xie F, Yang X, et al. Intronic pentanucleotide TTCA repeat insertion in the SAMD12 gene causes familial cortical myoclonic tremor with epilepsy type 1. *Brain.* 2018;141(8):2280-8. doi: 10.1093/brain/awy160
- [125] Yeetong P, Pongpanich M, Srichomthong C, Assawapitaksakul A, Shotelersuk V, Tantirukdham N, et al. TTCA repeat insertions in an intron of YEATS2 in benign adult familial myoclonic epilepsy type 4. *Brain.* 2019;142(11):3360-6. doi: 10.1093/brain/awz267
- [126] Florian RT, Kraft F, Leitao E, Kaya S, Klebe S, Magnin E, et al. Unstable TTTA/TTCA expansions in MARCH6 are associated with familial adult myoclonic epilepsy type 3. *Nat Commun.* 2019;10(1):4919. doi: 10.1038/s41467-019-12763-9
- [127] Corbett MA, Kroes T, Veneziano L, Bennett MF, Florian R, Schneider AL, et al. Intronic ATTTC repeat expansions in STARD7 in familial adult myoclonic epilepsy linked to chromosome 2. *Nat Commun.* 2019;10(1):4920. doi: 10.1038/s41467-019-12671-y
- [128] Wöhrle D, Hennig I, Vogel W, Steinbach P. Mitotic stability of fragile X mutations in differentiated cells indicates early post-conceptual trinucleotide repeat expansion. *Nat Genet.* 1993;4(2):140-2.
- [129] Wöhrle D, Hirst MC, Kennerknecht I, Davies KE, Steinbach P. Genotype mosaicism in fragile X fetal tissues. *Hum Genet.* 1992;89(1):114-6.
- [130] Devys D, Biancalana V, Rousseau F, Boue J, Mandell J-L, Oberle I. Analysis of full mutation fragile X mutations in fetal tissues and monozygotic twins indicates that abnormal methylation and somatic heterogeneity are established early in development. *Am J Med Genet.* 1992;43:208-16.
- [131] Tassone F, Hagerman RJ, Gane LW, Taylor AK. Strong similarities of the FMR1 mutation in multiple tissues: Postmortem studies of a male with a full mutation and a male carrier of a premutation. *Am J Med Genet.* 1999;84(3):240-4.
- [132] Reyniers E, Martin JJ, Cras P, Van Marck E, Handig I, Jorens HZ, et al. Postmortem examination of two fragile X brothers with an FMR1 full mutation. *Am J Med Genet.* 1999;84(3):245-9.
- [133] Anvret M, Ahlberg G, Grandell U, Hedberg B, Johnson K, Edstrom L. Larger expansions of the CTG repeat in muscle compared to lymphocytes from patients with myotonic dystrophy. *Hum Mol Genet.* 1993;2:1397-400. doi: 10.1093/hmg/2.9.1397
- [134] Ashizawa T, Dubel JR, Harati Y. Somatic instability of CTG repeat in myotonic dystrophy. *Neurology.* 1993;43:2674-8. doi: 10.1212/WNL.43.12.2674
- [135] Jansen G, Willems P, Coerwinkel M, Nillesen W, Smeets H, Vits L, et al. Gonosomal mosaicism in myotonic dystrophy patients: Involvement of mitotic events in (CTG)<sub>n</sub> variation and selection against extreme expansion in sperm. *Am J Hum Genet.* 1994;54:575-85.
- [136] Thornton CA, Johnson KJ, Moxley RT. Myotonic dystrophy patients have larger CTG expansions in skeletal muscle than in leukocytes. *Ann Neurol.* 1994;35(1):104-7. doi: 10.1002/ana.410350116
- [137] Hecht BK, Donnelly A, Gedeon AK, Byard RW, Haan EA, Mulley JC. Direct molecular diagnosis of myotonic dystrophy. *Clin Genet.* 1993;43(6):276-85. doi: 10.1111/j.1399-0004.1993.tb03819.x
- [138] Myring J, Meredith AL, Harley HG, Kohn G, Norbury G, Harper PS, et al. Specific molecular prenatal diagnosis for the CTG mutation in myotonic dystrophy. *J Med Genet.* 1992;29(11):785-8. doi: 10.1136/jmg.29.11.785
- [139] Tachi N, Ohya K, Chiba S, Sato T, Kikuchi K. Minimal somatic instability of CTG repeat in congenital myotonic dystrophy. *Pediatr Neurol.* 1995;12(1):81-3.
- [140] Tachi N, Ohya K, Chiba S, Sato T. Unstable DNA in a patient with a severe form of congenital myotonic dystrophy. *J Neurol Sci.* 1993;119(2):180-2.
- [141] Wong LJ, Ashizawa T, Monckton DG, Caskey CT, Richards CS. Somatic heterogeneity of the CTG repeat in myotonic dystrophy is age and size dependent. *Am J Hum Genet.* 1995;56(1):114-22.
- [142] Wong LJ, Ashizawa T. Instability of the (CTG)<sub>n</sub> repeat in congenital myotonic dystrophy. *Am J Hum Genet.* 1997;61(6):1445-8.
- [143] Martorell L, Johnson KJ, Boucher CA, Baiget M. Somatic instability of the myotonic dystrophy (CTG)<sub>n</sub> repeat during human fetal development. *Hum Mol Genet.* 1997;6(6):877-80.
- [144] Ohya K, Tachi N, Kon S, Kikuchi K, Chiba S. Somatic cell heterogeneity between DNA extracted from lymphocytes and skeletal muscle in congenital myotonic dystrophy. *Jpn J Hum Genet.* 1995;40(4):319-26.
- [145] Lavedan C, Hofmann-Radvanyi H, Shelbourne P, Rabes J-P, Duros C, Savoy D, et al. Myotonic dystrophy: Size- and sex-dependent dynamics of CTG meiotic instability, and somatic mosaicism. *Am J Hum Genet.* 1993;52:875-83.
- [146] Monckton DG, Wong LJ, Ashizawa T, Caskey CT. Somatic mosaicism, germline expansions, germline reversions and intergenerational reductions in myotonic dystrophy males: Small pool PCR analyses. *Hum Mol Genet.* 1995;4(1):1-8. doi: 10.1093/hmg/4.1.1
- [147] Martorell L, Monckton DG, Gamez J, Johnson KJ, Gich I, Lopez de Munain A, et al. Progression of somatic CTG repeat length heterogeneity in the blood cells of myotonic dystrophy patients. *Hum Mol Genet.* 1998;7(2):307-12. doi: 10.1093/hmg/7.2.307
- [148] Morales F, Couto JM, Higham CF, Hogg G, Cuenca P, Braida C, et al. Somatic instability of the expanded CTG triplet repeat in myotonic dystrophy type 1 is a heritable quantitative trait and modifier of disease severity. *Hum Mol Genet.* 2012;21:3558-67. doi: 10.1093/hmg/dds185
- [149] Martorell L, Martinez JM, Carey N, Johnson K, Baiget M. Comparison of CTG repeat length expansion and clinical progression of myotonic dystrophy over a five year period. *J Med Genet.* 1995;32(8):593-6.

- [150] Cumming SA, Jimenez-Moreno C, Okkersen K, Wenninger S, Daidj F, Hogarth F, et al. Genetic determinants of disease severity in the myotonic dystrophy type 1 OPTIMISTIC cohort. *Neurology*. 2019;93:e995-e1009. doi: 10.1212/WNL.0000000000008056
- [151] Morales F, Vasquez M, Santamaria C, Cuenca P, Corrales E, Monckton DG. A polymorphism in the *MSH3* mismatch repair gene is associated with the levels of somatic instability of the expanded CTG repeat in the blood DNA of myotonic dystrophy type 1 patients. *DNA Repair* 2016;40:57-66. doi: 10.1016/j.dnarep.2016.01.001
- [152] MacDonald ME, Barnes G, Srinidhi J, Duyao MP, Ambrose CM, Myers RH, et al. Gametic but not somatic instability of CAG repeat length in Huntington's disease. *J Med Genet*. 1993;30:982-6.
- [153] Telenius H, Kremer B, Goldberg YP, Theilmann J, Andrew SE, Zeisler J, et al. Somatic and gonadal mosaicism of the Huntington disease gene CAG repeat in brain and sperm. *Nat Genet*. 1994;6:409-14.
- [154] Richards RI, Sutherland GR. Simple DNA is not replicated simply. *Nat Genet*. 1994;6:114-6.
- [155] Tautz D, Schlotterer. Simple sequences. *Curr Opin Genet Dev*. 1994;4(6):832-7. doi: 10.1016/0959-437x(94)90067-1
- [156] Kunkel TA. Nucleotide repeats. Slippery DNA and diseases. *Nature*. 1993;365(6443):207-8.
- [157] De Rooij KE, De Koning Gans PA, Roos RA, Van Ommen GJ, Den Dunnen JT. Somatic expansion of the (CAG)<sub>n</sub> repeat in Huntington disease brains. *Hum Genet*. 1995;95(3):270-4.
- [158] Aronin N, Chase K, Young C, Sapp E, Schwarz C, Matta N, et al. CAG expansion affects the expression of mutant Huntingtin in the Huntington's disease brain. *Neuron*. 1995;15(5):1193-201.
- [159] Kono Y, Agawa Y, Watanabe Y, Ohama E, Nanba E, Nakashima K. Analysis of the CAG repeat number in a patient with Huntington's disease. *Intern Med*. 1999;38(5):407-11. doi: 10.2169/internalmedicine.38.407
- [160] Bingham PM, Scott MO, Wang S, McPhaul MJ, Wilson EM, Garben JY, et al. Stability of an expanded trinucleotide repeat in the androgen receptor gene in transgenic mice. *Nat Genet*. 1995;9(2):191-6.
- [161] Burright EN, Clark HB, Servadio A, Matilla T, Feddersen RM, Yunis WS, et al. *SCA1* transgenic mice: A model for neurodegeneration caused by an expanded CAG trinucleotide repeat. *Cell*. 1995;82(6):937-48.
- [162] Goldberg YP, Kalchman MA, Metzler M, Nasir J, Zeisler J, Graham R, et al. Absence of disease phenotype and intergenerational stability of the CAG repeat in transgenic mice expressing the human Huntington disease transcript. *Hum Mol Genet*. 1996;5:177-85.
- [163] Mangiarini L, Sathasivam K, Seller M, Cozens B, Harper A, Hetherington C, et al. Exon 1 of the *HD* gene with an expanded CAG repeat is sufficient to cause a progressive neurological phenotype in transgenic mice. *Cell*. 1996;87(3):493-506.
- [164] Mangiarini L, Sathasivam K, Mahal A, Mott R, Seller M, Bates GP. Instability of highly expanded CAG repeats in mice transgenic for the Huntington's disease mutation. *Nat Genet*. 1997;15(2):197-200.
- [165] Monckton DG, Coolbaugh MI, Ashizawa K, Siciliano MJ, Caskey CT. Hypermutable myotonic dystrophy CTG repeats in transgenic mice. *Nat Genet*. 1997;15(February):193-6. doi: 10.1038/ng0297-193
- [166] Gourdon G, Radvanyi F, Lia AS, Duros C, Blanche M, Abitbol M, et al. Moderate intergenerational and somatic instability of a 55 CTG repeat in transgenic mice. *Nat Genet*. 1997;15(2):190-2.
- [167] Wheeler VC, Auerbach W, White JK, Srinidhi J, Auerbach A, Ryan A, et al. Length-dependent gametic CAG repeat instability in the Huntington's disease knock-in mouse. *Hum Mol Genet*. 1999;8(1):115-22.
- [168] Kennedy L, Shelbourne PF. Dramatic mutation instability in HD mouse striatum: Does polyglutamine load contribute to cell-specific vulnerability in Huntington's disease? *Hum Mol Genet*. 2000;9(17):2539-44.
- [169] Wheeler VC, Dion V. Modifiers of CAG repeat instability: Insights from model systems. *J Huntingtons Dis*. 2020. doi: 10.3233/JHD-200426
- [170] Leeftang EP, Zhang L, Tavaré S, Hubert R, Srinidhi J, MacDonald ME, et al. Single sperm analysis of the trinucleotide repeats in the Huntington's disease gene: Quantification of the mutation frequency and spectrum. *Hum Mol Genet*. 1995;4:1519-26.
- [171] Shinde D, Lai Y, Sun F, Arnheim N. Taq DNA polymerase slippage mutation rates measured by PCR and quasi-likelihood analysis: (CA/GT)<sub>n</sub> and (A/T)<sub>n</sub> microsatellites. *Nucleic Acids Res*. 2003;31(3):974-80.
- [172] Mutter GL, Boynton KA. PCR bias in amplification of androgen receptor alleles, a trinucleotide repeat marker used in clonality studies. *Nucleic Acids Res*. 1995;23(8):1411-8. doi: 10.1093/nar/23.8.1411
- [173] Jeffreys AJ, Tamaki K, MacLeod A, Monckton DG, Neil DL, Armour JAL. Complex gene conversion events in germline mutation at human minisatellites. *Nat Genet*. 1994;6:136-45. doi: 10.1038/ng0294-136
- [174] Gomes-Pereira M, Bidichandani SI, Monckton DG. Analysis of unstable triplet repeats using small-pool polymerase chain reaction. *Methods Mol Biol*. 2004;277:61-76. doi: 10.1385/1-59259-804-8:061
- [175] Kennedy L, Evans E, Chen CM, Craven L, Detloff PJ, Ennis M, et al. Dramatic tissue-specific mutation length increases are an early molecular event in Huntington disease pathogenesis. *Hum Mol Genet*. 2003;12(24):3359-67. doi: 10.1093/hmg/ddg352
- [176] Vonsattel JP, Myers RH, Stevens TJ, Ferrante RJ, Bird ED, Richardson EP, Jr. Neuropathological classification of Huntington's disease. *J Neuropathol Exp Neurol*. 1985;44(6):559-77.
- [177] Hunter JM, Crouse AB, Lesort M, Johnson GV, Detloff PJ. Verification of somatic CAG repeat expansion by pre-PCR fractionation. *J Neurosci Methods*. 2005;144(1):11-7. doi: 10.1016/j.jneumeth.2004.10.006
- [178] Shelbourne PF, Keller-McGandy C, Bi WL, Yoon SR, Dubeau L, Veitch NJ, et al. Triplet repeat mutation length gains correlate with cell-type specific vulnerability in Huntington disease brain. *Hum Mol Genet*. 2007;16(10):1133-42. doi: 10.1093/hmg/ddm054
- [179] Gonitell R, Moffitt H, Sathasivam K, Woodman B, Detloff PJ, Faull RL, et al. DNA instability in postmitotic neurons. *Proc Natl Acad Sci U S A*. 2008;105(9):3467-72. doi: 10.1073/pnas.0800048105
- [180] Fortune MT, Vassilopoulos C, Coolbaugh MI, Siciliano MJ, Monckton DG. Dramatic, expansion-biased, age-dependent, tissue-specific somatic mosaicism in a transgenic mouse model of triplet repeat instability. *Hum Mol Genet*. 2000;9(3):439-45. doi: 10.1093/hmg/9.3.439
- [181] Lia AS, Seznec H, Hofmann-Radvanyi H, Radvanyi F, Duros C, Saquet C, et al. Somatic instability of the CTG

- repeat in mice transgenic for the myotonic dystrophy region is age dependent but not correlated to the relative intertissue transcription levels and proliferative capacities. *Hum Mol Genet.* 1998;7(8):1285-91.
- [182] Watase K, Venken KJ, Sun Y, Orr HT, Zoghbi HY. Regional differences of somatic CAG repeat instability do not account for selective neuronal vulnerability in a knock-in mouse model of SCA1. *Hum Mol Genet.* 2003;12(21):2789-95.
- [183] Sato T, Oyake M, Nakamura K, Nakao K, Fukusima Y, Onodera O, et al. Transgenic mice harboring a full-length human mutant DRPLA gene exhibit age-dependent intergenerational and somatic instabilities of CAG repeats comparable with those in DRPLA patients. *Hum Mol Genet.* 1999;8(1):99-106.
- [184] Ishii S, Nishio T, Sunohara N, Yoshihara T, Takemura K, Hikiji K, et al. Small increase in triplet repeat length of cerebellum from patients with myotonic dystrophy. *Hum Genet.* 1996;98(2):138-40. doi: 10.1007/s004390050176
- [185] Jinnai K, Mitani M, Futamura N, Kawamoto K, Funakawa I, Itoh K. Somatic instability of CTG repeats in the cerebellum of myotonic dystrophy type 1. *Muscle Nerve.* 2013;48(1):105-8. doi: 10.1002/mus.23717
- [186] Chong SS, McCall AE, Cota J, Subramony SH, Orr HT, Hughes MR, et al. Gametic and somatic tissue specific heterogeneity of the expanded SCA1 CAG repeat in spinocerebellar ataxia type 1. *Nat Genet.* 1995;10(july):344-50.
- [187] Lopes-Cendes I, Maciel P, Kish S, Gaspar C, Robitaille Y, Clark HB, et al. Somatic mosaicism in the central nervous system in spinocerebellar ataxia type 1 and Machado-Joseph disease. *Ann Neurol.* 1996;40(2):199-206.
- [188] Hashida H, Goto J, Kurisaki H, Mizusawa H, Kanazawa I. Brain regional differences in the expansion of a CAG repeat in the spinocerebellar ataxias: Dentatorubral pallidoluysian atrophy, Machado Joseph disease, and spinocerebellar ataxia type 1. *Ann Neurol.* 1997;41(4):505-11.
- [189] Zuhlke C, Hellenbroich Y, Schaaff F, Gehlken U, Wessel K, Schubert T, et al. CAG repeat analyses in frozen and formalin fixed tissues following primer extension preamplification for evaluation of mitotic instability of expanded SCA1 alleles. *Hum Genet.* 1997;100(3 4):339-44.
- [190] Maciel P, Lopes-Cendes I, Kish S, Sequeiros J, Rouleau GA. Mosaicism of the CAG repeat in CNS tissue in relation to age at death in spinocerebellar ataxia type 1 and Machado-Joseph disease patients. *Am J Hum Genet.* 1997;60(4):993-6.
- [191] Koefoed P, Hasholt L, Fenger K, Nielsen JE, Eiberg H, Buschard K, et al. Mitotic and meiotic instability of the CAG trinucleotide repeat in spinocerebellar ataxia type 1. *Hum Genet.* 1998;103(5):564-9.
- [192] Mouro Pinto R, Arning L, Giordano JV, Razghandi P, Andrew MA, Gillis T, et al. Patterns of CAG repeat instability in the central nervous system and periphery in Huntington's disease and in spinocerebellar ataxia type 1. *Hum Mol Genet.* 2020;29(15):2551-67. doi: 10.1093/hmg/ddaa139
- [193] Cancel G, Durr A, Didierjean O, Imbert G, Burk K, Lezin A, et al. Molecular and clinical correlations in spinocerebellar ataxia 2: A study of 32 families. *Hum Mol Genet.* 1997;6(5):709-15.
- [194] Matsuura T, Sasaki H, Yabe I, Hamada K, Hamada T, Shitara M, et al. Mosaicism of unstable CAG repeats in the brain of spinocerebellar ataxia type 2. *J Neurol.* 1999;246(9):835-9.
- [195] Ito Y, Tanaka F, Yamamoto M, Doyu M, Nagamatsu M, Riku S, et al. Somatic mosaicism of the expanded CAG trinucleotide repeat in mRNAs for the responsible gene of Machado Joseph disease (MJD), dentatorubral pallidoluysian atrophy (DRPLA), and spinal and bulbar muscular atrophy (SBMA). *Neurochem Res.* 1998;23(1):25-32.
- [196] Tanaka F, Sobue G, Doyu M, Ito Y, Yamamoto M, Shimada N, et al. Differential pattern in tissue specific somatic mosaicism of expanded CAG trinucleotide repeat in dentatorubral pallidoluysian atrophy, Machado Joseph disease, and X linked recessive spinal and bulbar muscular atrophy. *J Neurol Sci.* 1996;135(1):43-50.
- [197] Cancel G, Gourfinkel-An I, Stevanin G, Didierjean O, Abbas N, Hirsch E, et al. Somatic mosaicism of the CAG repeat expansion in spinocerebellar ataxia type 3/Machado-Joseph disease. *Hum Mutat.* 1998;11(1):23-7.
- [198] Ashery-Padan R, Marquardt T, Zhou X, Gruss P. Pax6 activity in the lens primordium is required for lens formation and for correct placement of a single retina in the eye. *Genes Dev.* 2000;14(21):2701-11.
- [199] Ueno S, Kondoh K, Kotani Y, Komure O, Kuno S, Kawai J, et al. Somatic mosaicism of CAG repeat in dentatorubral-pallidoluysian atrophy (DRPLA). *Hum Mol Genet.* 1995;4(4):663-6.
- [200] Onodera O, Oyake M, Takano H, Ikeuchi T, Igarashi S, Tsuji S. Molecular cloning of a full-length cDNA for dentatorubral-pallidoluysian atrophy and regional expressions of the expanded alleles in the CNS. *Am J Hum Genet.* 1995;57(5):1050-60.
- [201] Takano H, Onodera O, Takahashi H, Igarashi S, Yamada M, Oyake M, et al. Somatic mosaicism of expanded CAG repeats in brains of patients with dentatorubral-pallidoluysian atrophy: Cellular population-dependent dynamics of mitotic instability. *Am J Hum Genet.* 1996;58:1212-22.
- [202] Hashida H, Goto J, Suzuki T, Jeong S, Masuda N, Ooie T, et al. Single cell analysis of CAG repeat in brains of dentatorubral-pallidoluysian atrophy (DRPLA). *J Neurol Sci.* 2001;190(1-2):87-93.
- [203] Watanabe H, Tanaka F, Doyu M, Riku S, Yoshida M, Hashizume Y, et al. Differential somatic CAG repeat instability in variable brain cell lineage in dentatorubral pallidoluysian atrophy (DRPLA): A laser-captured microdissection (LCM)-based analysis. *Hum Genet.* 2000;107(5):452-7.
- [204] Aoki M, Abe K, Tobita M, Kameya T, Watanabe M, Itoyama Y. Reduction of CAG expansions in cerebellar cortex and spinal cord of DRPLA. *Clin Genet.* 1996;50(4):199-201. doi: 10.1111/j.1399-0004.1996.tb02625.x
- [205] Tanaka F, Reeves MF, Ito Y, Matsumoto M, Li M, Miwa S, et al. Tissue specific somatic mosaicism in spinal and bulbar muscular atrophy is dependent on CAG repeat length and androgen receptor gene expression level. *Am J Hum Genet.* 1999;65(4):966-73.
- [206] Watanabe M, Abe K, Aoki M, Yasuo K, Itoyama Y, Shoji M, et al. Mitotic and meiotic stability of the CAG repeat in the X-linked spinal and bulbar muscular atrophy gene. *Clin Genet.* 1996;50(3):133-7. doi: 10.1111/j.1399-0004.1996.tb02367.x
- [207] Ansved T, Lundin A, Anvret M. Larger CAG expansions in skeletal muscle compared with lymphocytes in

- Kennedy disease but not in Huntington disease. *Neurology*. 1998;51(5):1442-4.
- [208] Schols L, Bauer P, Schmidt T, Schulte T, Riess O. Autosomal dominant cerebellar ataxias: Clinical features, genetics, and pathogenesis. *Lancet Neurol*. 2004;3(5):291-304. doi: 10.1016/S1474-4422(04)00737-9
- [209] Strong TV, Tagle DA, Valdes JM, Elmer LW, Boehm K, Swaroop M, et al. Widespread expression of the human and rat Huntington's disease gene in brain and nonneural tissues. *Nat Genet*. 1993;5(3):259-65. doi: 10.1038/ng1193-259
- [210] Li SH, Schilling G, Young WS, 3rd, Li XJ, Margolis RL, Stine OC, et al. Huntington's disease gene (IT15) is widely expressed in human and rat tissues. *Neuron*. 1993;11(5):985-93. doi: 10.1016/0896-6273(93)90127-d
- [211] Landwehrmeyer GB, McNeil SM, Dure LSt, Ge P, Aizawa H, Huang Q, et al. Huntington's disease gene: Regional and cellular expression in brain of normal and affected individuals. *Ann Neurol*. 1995;37(2):218-30. doi: 10.1002/ana.410370213
- [212] Trotter Y, Devys D, Imbert G, Saudou F, An I, Lutz Y, et al. Cellular localization of the Huntington's disease protein and discrimination of the normal and mutated form. *Nat Genet*. 1995;10(1):104-10.
- [213] Kornberg A, Bertsch LL, Jackson JF, Khorana HG. Enzymatic synthesis of deoxyribonucleic acid, XVI. Oligonucleotides as templates and the mechanism of their replication. *Proc Natl Acad Sci U S A*. 1964;51:315-23. doi: 10.1073/pnas.51.2.315
- [214] Wells RD, Ohtsuka E, Khorana HG. Studies on polynucleotides. L. Synthetic deoxyribopolynucleotides as templates for the DNA polymerase of *Escherichia coli*: A new double-stranded DNA-like polymer containing repeating dinucleotide sequences. *J Mol Biol*. 1965;14(1):221-37. doi: 10.1016/s0022-2836(65)80242-x
- [215] Wells RD, Jacob TM, Narang SA, Khorana HG. Studies on polynucleotides. LXIX. Synthetic deoxyribopolynucleotides as templates for the DNA polymerase of *Escherichia coli*: DNA-like polymers containing repeating trinucleotide sequences. *J Mol Biol*. 1967;27(2):237-63. doi: 10.1016/0022-2836(67)90018-6
- [216] Tautz D, Renz M. Simple sequences are ubiquitous repetitive components of eukaryotic genomes. *Nucleic Acids Res*. 1984;12(10):4127-38. doi: 10.1093/nar/12.10.4127
- [217] Weber JL, May PE. Abundant class of human DNA polymorphisms which can be typed using the polymerase chain reaction. *Am J Hum Genet*. 1989;44(3):388-96.
- [218] Strand M, Prolla TA, Liskay RM, Petes TD. Destabilization of tracts of simple repetitive DNA in yeast by mutations affecting DNA mismatch repair. *Nature*. 1993;365(6443):274-6.
- [219] Fishel R, Lescoe MK, Rao MR, Copeland NG, Jenkins NA, Garber J, et al. The human mutator gene homolog MSH2 and its association with hereditary nonpolyposis colon cancer. *Cell*. 1993;75(5):1027-38.
- [220] Leach FS, Nicolaides NC, Papadopoulos N, Liu B, Jen J, Parsons R, et al. Mutations of a mutS homolog in hereditary nonpolyposis colorectal cancer. *Cell*. 1993;75(6):1215-25.
- [221] Nicolaides NC, Papadopoulos N, Liu B, Wei YF, Carter KC, Ruben SM, et al. Mutations of two PMS homologues in hereditary nonpolyposis colon cancer. *Nature*. 1994;371(6492):75-80.
- [222] Papadopoulos N, Nicolaides NC, Wei YF, Ruben SM, Carter KC, Rosen CA, et al. Mutation of a mutL homolog in hereditary colon cancer. *Science*. 1994;263(5153):1625-9.
- [223] Bronner CE, Baker SM, Morrison PT, Warren G, Smith LG, Lescoe MK, et al. Mutation in the DNA mismatch repair gene homologue hMLH1 is associated with hereditary non-polyposis colon cancer. *Nature*. 1994;368(6468):258-61.
- [224] Leeflang EP, Tavaré S, Marjoram P, Neal COS, Srinidhi J, MacDonald ME, et al. Analysis of germline mutation spectra at the Huntington's disease locus supports a mitotic mutation mechanism. *Hum Mol Genet*. 1999;8(2):173-83.
- [225] Yoon SR, Dubeau L, de Young M, Wexler NS, Arnheim N. Huntington disease expansion mutations in humans can occur before meiosis is completed. *Proc Natl Acad Sci U S A*. 2003;100(15):8834-8.
- [226] Martorell L, Monckton DG, Gamez J, Baiget M. Complex patterns of male germline instability and somatic mosaicism in myotonic dystrophy type 1. *Eur J Hum Genet*. 2000;8(6):423-30. doi: 10.1038/sj.ejhg.5200478
- [227] Martorell L, Gamez J, Cayuela ML, Gould FK, McAbney JP, Ashizawa T, et al. Germline mutational dynamics in myotonic dystrophy type 1 males: Allele length and age effects. *Neurology*. 2004;62(2):269-74. doi: 10.1212/WNL.62.2.269
- [228] Grewal RP, Leeflang EP, Zhang L, Arnheim N. The mutation properties of spinal and bulbar muscular atrophy disease alleles. *Neurogenetics*. 1998;1(4):249-52.
- [229] Zhang L, Fischbeck KH, Arnheim N. CAG repeat length variation in sperm from a patient with Kennedy's disease. *Hum Mol Genet*. 1995;4:303-5.
- [230] Monckton DG, Cayuela ML, Gould FK, Brock GJR, de Silva R, Ashizawa T. Very large (CAG)<sub>n</sub> DNA repeat expansions in the sperm of two spinocerebellar ataxia type 7 males. *Hum Mol Genet*. 1999;8(13):2473-8. doi: 10.1093/hmg/8.13.2473
- [231] Shelbourne PF, Killeen N, Hevner RF, Johnston HM, Tecott L, Lewandoski M, et al. A Huntington's disease CAG expansion at the murine *Hdh* locus is unstable and associated with behavioural abnormalities in mice. *Hum Mol Genet*. 1999;8(5):763-74.
- [232] Manley K, Shirley TL, Flaherty L, Messer A. *Msh2* deficiency prevents *in vivo* somatic instability of the CAG repeat in Huntington disease transgenic mice. *Nat Genet*. 1999;23(4):471-3. doi: 10.1038/70598
- [233] Iyer R, Pluciennik A. DNA mismatch repair and its role in Huntington's disease. *J Huntingtons Dis*. 2020. doi: 10.3233/JHD-200438
- [234] van den Broek WJ, Nelen MR, Wansink DG, Coerwinkel MM, te Riele H, Groenen PJ, et al. Somatic expansion behaviour of the (CTG)<sub>(n)</sub> repeat in myotonic dystrophy knock-in mice is differentially affected by Msh3 and Msh6 mismatch-repair proteins. *Hum Mol Genet*. 2002;11(2):191-8. doi: 10.1093/hmg/11.2.191
- [235] Kovtun IV, McMurray CT. Trinucleotide expansion in haploid germ cells by gap repair. *Nat Genet*. 2001;27(4):407-11.
- [236] Pearson CE, Sinden RR. Alternative structures in duplex DNA formed within the trinucleotide repeats of the myotonic dystrophy and fragile X loci. *Biochemistry*. 1996;35(15):5041-53.
- [237] Gacy AM, Goellner G, Juranic N, Macura S, McMurray CT. Trinucleotide repeats that expand in human disease form hairpin structures *in vitro*. *Cell*. 1995;81(4):533-40.

- [238] Gomes-Pereira M, Fortune MT, Ingram L, McAbney JP, Monckton DG. *Pms2* is a genetic enhancer of trinucleotide CAG•CTG repeat somatic mosaicism: Implications for the mechanism of triplet repeat expansion. *Hum Mol Genet.* 2004;13(16):1815-25. doi: 10.1093/hmg/ddh186
- [239] Pinto RM, Dragileva E, Kirby A, Lloret A, Lopez E, St Claire J, et al. Mismatch repair genes *Mlh1* and *Mlh3* modify CAG instability in Huntington's disease mice: Genome-wide and candidate approaches. *PLoS Genet.* 2013;9(10):e1003930. doi: 10.1371/journal.pgen.1003930
- [240] Tome S, Holt I, Edelmann W, Morris GE, Munnich A, Pearson CE, et al. MSH2 ATPase domain mutation affects CTG•CAG repeat instability in transgenic mice. *PLoS Genet.* 2009;5(5):e1000482. doi: 10.1371/journal.pgen.1000482
- [241] Lloret A, Dragileva E, Teed A, Espinola J, Fossale E, Gillis T, et al. Genetic background modifies nuclear mutant huntingtin accumulation and HD CAG repeat instability in Huntington's disease knock-in mice. *Hum Mol Genet.* 2006;15(12):2015-24. doi: 10.1093/hmg/ddl125
- [242] Tome S, Manley K, Simard JP, Clark GW, Slean MM, Swami M, et al. MSH3 polymorphisms and protein levels affect CAG repeat instability in Huntington's disease mice. *PLoS Genet.* 2013;9(2):e1003280. doi: 10.1371/journal.pgen.1003280
- [243] Wheeler VC, Lebel LA, Vrbanac V, Teed A, Te Riele H, MacDonald ME. Mismatch repair gene *Msh2* modifies the timing of early disease in *Hdh<sup>Q111</sup>* striatum. *Hum Mol Genet.* 2003;12(3):273-81.
- [244] Dragileva E, Hendricks A, Teed A, Gillis T, Lopez ET, Friedberg EC, et al. Intergenerational and striatal CAG repeat instability in Huntington's disease knock-in mice involve different DNA repair genes. *Neurobiol Dis.* 2009;33(1):37-47.
- [245] Morton AJ, Skillings EA, Wood NI, Zheng Z. Antagonistic pleiotropy in mice carrying a CAG repeat expansion in the range causing Huntington's disease. *Sci Rep.* 2019;9(1):37. doi: 10.1038/s41598-018-37102-8
- [246] Fossale E, Wheeler VC, Vrbanac V, Lebel LA, Teed A, Mysore JS, et al. Identification of a presymptomatic molecular phenotype in *Hdh* CAG knock-in mice. *Hum Mol Genet.* 2002;11(19):2233-41. doi: 10.1093/hmg/11.19.2233
- [247] Gray M, Shirasaki DI, Cepeda C, Andre VM, Wilburn B, Lu XH, et al. Full-length human mutant huntingtin with a stable polyglutamine repeat can elicit progressive and selective neuropathogenesis in BACHD mice. *J Neurosci.* 2008;28(24):6182-95. doi: 10.1523/JNEUROSCI.0857-08.2008
- [248] Pouladi MA, Stanek LM, Xie Y, Franciosi S, Southwell AL, Deng Y, et al. Marked differences in neurochemistry and aggregates despite similar behavioural and neuropathological features of Huntington disease in the full-length BACHD and YAC128 mice. *Hum Mol Genet.* 2012;21(10):2219-32. doi: 10.1093/hmg/dds037
- [249] Swami M, Hendricks AE, Gillis T, Massood T, Mysore J, Myers RH, et al. Somatic expansion of the Huntington's disease CAG repeat in the brain is associated with an earlier age of disease onset. *Hum Mol Genet.* 2009;18(16):3039-47.
- [250] Veitch NJ, Ennis M, McAbney JP, Shelbourne PF, Monckton DG. Inherited CAG•CTG allele length is a major modifier of somatic mutation length variability in Huntington disease. *DNA Repair.* 2007;6(6):789-96.
- [251] Dickey AS, La Spada AR. Therapy development in Huntington disease: From current strategies to emerging opportunities. *Am J Med Genet A.* 2018;176(4):842-61. doi: 10.1002/ajmg.a.38494
- [252] Hong EP, MacDonald ME, Wheeler VC, Jones L, Holmans P, Orth M, et al. Huntington's disease pathogenesis: Two sequential components. *J Huntingtons Dis.* 2020. doi: 10.3233/JHD-200427
- [253] The Genetic Modifiers of Huntington's Disease Consortium. Identification of genetic factors that modify clinical onset of Huntington's disease. *Cell.* 2015;162(3):516-26. doi: 10.1016/j.cell.2015.07.003
- [254] Goold R, Flower M, Moss DH, Medway C, Wood-Kaczmar A, Andre R, et al. *FAN1* modifies Huntington's disease progression by stabilising the expanded *HTT* CAG repeat. *Hum Mol Genet.* 2019;28:650-61. doi: 10.1093/hmg/ddy375
- [255] Zhao XN, Usdin K. FAN1 protects against repeat expansions in a Fragile X mouse model. *DNA Repair.* 2018;69:1-5. doi: 10.1016/j.dnarep.2018.07.001
- [256] Loupe JM, Pinto RM, Kim KH, Gillis T, Mysore JS, Andrew MA, et al. Promotion of somatic CAG repeat expansion by Fan1 knock-out in Huntington's disease knock-in mice is blocked by *Mlh1* knock-out. *Hum Mol Genet.* 2020;29(18):3044-53. doi: 10.1093/hmg/ddaa196
- [257] Deshmukh A, Porro A, Mohiuddin M, Lanni S, Panigrahi G, Caron M. FAN1, a DNA repair nuclease, as a modifier of repeat expansion disorders. *J Huntingtons Dis.* 2020. doi: 10.3233/JHD-200448
- [258] Zhao X, Kumari D, Miller C, Kim G, Hayward B, Vitalo AG, et al. Modifiers of somatic repeat instability in mouse models of Friedreich ataxia and the Fragile X-related disorders: Implications for the mechanism of somatic expansion in Huntington disease. *J Huntingtons Dis.* 2020. doi: 10.3233/JHD-200423
- [259] Lee JM, Chao MJ, Harold D, Abu Elneel K, Gillis T, Holmans P, et al. A modifier of Huntington's disease onset at the *MLH1* locus. *Hum Mol Genet.* 2017;26(19):3859-67. doi: 10.1093/hmg/ddx286
- [260] Hensman Moss DJ, Pardinas AF, Langbehn D, Lo K, Leavitt BR, Roos R, et al. Identification of genetic variants associated with Huntington's disease progression: A genome-wide association study. *Lancet Neurol.* 2017;16(9):701-11. doi: 10.1016/S1474-4422(17)30161-8
- [261] Flower M, Lomeikaite V, Ciosi M, Cumming S, Morales F, Lo K, et al. MSH3 modifies somatic instability and disease severity in Huntington's and myotonic dystrophy type 1. *Brain.* 2019;142:1876-86. doi: 10.1093/brain/awz115
- [262] Ciosi M, Maxwell A, Cumming SA, Hensman Moss DJ, Alshammari AM, Flower MD, et al. A genetic association study of glutamine-encoding DNA sequence structures, somatic CAG expansion, and DNA repair gene variants, with Huntington disease clinical outcomes. *EBioMedicine.* 2019;48:568-80. doi: 10.1016/j.ebiom.2019.09.020
- [263] Wright GEB, Collins JA, Kay C, McDonald C, Dolzhenko E, Xia Q, et al. Length of uninterrupted CAG, independent of polyglutamine size, results in increased somatic instability, hastening onset of Huntington disease. *Am J Hum Genet.* 2019;104(6):1116-26. doi: 10.1016/j.ajhg.2019.04.007
- [264] The Genetic Modifiers of Huntington's Disease Consortium. CAG repeat not polyglutamine length determines timing of Huntington's dis-

- ease onset. *Cell*. 2019;178(4):887-900 e14. doi: 10.1016/j.cell.2019.06.036
- [265] Chung M, Ranum LPW, Duvick LA, Servadio A, Zoghbi HY, Orr HT. Evidence for a mechanism predisposing to intergenerational CAG repeat instability in spinocerebellar ataxia type I. *Nat Genet*. 1993;5:254-8.
- [266] Kunst C, Warren ST. Cryptic and polar variation of the fragile X repeat could result in predisposing normal alleles. *Cell*. 1994;77:853-61.
- [267] Eichler EE, Holden JJ, Popovich BW, Reiss AL, Snow K, Thibodeau SN, et al. Length of uninterrupted CGG repeats determines instability in the *FMRI* gene. *Nat Genet*. 1994;8(1):88-94. doi: 10.1038/ng0994-88
- [268] Choudhry S, Mukerji M, Srivastava AK, Jain S, Brahmachari SK. CAG repeat instability at SCA2 locus: Anchoring CAA interruptions and linked single nucleotide polymorphisms. *Hum Mol Genet*. 2001;10(21):2437-46.
- [269] Musova Z, Mazanec R, Krepelova A, Ehler E, Vales J, Jaklova R, et al. Highly unstable sequence interruptions of the CTG repeat in the myotonic dystrophy gene. *Am J Med Genet A*. 2009;149A(7):1365-74. doi: 10.1002/ajmg.a.32987
- [270] Braida C, Stefanatos RK, Adam B, Mahajan N, Smeets HJ, Niel F, et al. Variatic CCG and GGC repeats within the CTG expansion dramatically modify mutational dynamics and likely contribute toward unusual symptoms in some myotonic dystrophy type 1 patients. *Hum Mol Genet*. 2010;19(8):1399-412. doi: 10.1093/hmg/ddq015
- [271] Overend G, Legare C, Mathieu J, Bouchard L, Gagnon C, Monckton DG. Allele length of the DMPK CTG repeat is a predictor of progressive myotonic dystrophy type 1 phenotypes. *Hum Mol Genet*. 2019;28(13):2245-54. doi: 10.1093/hmg/ddz055
- [272] Tome S, Dandelot E, Dogan C, Bertrand A, Genevieve D, Pereon Y, et al. Unusual association of a unique CAG interruption in 5' of DM1 CTG repeats with intergenerational contractions and low somatic mosaicism. *Hum Mutat*. 2018;39(7):970-82. doi: 10.1002/humu.23531
- [273] Pesovic J, Peric S, Brkusanic M, Brajuskovic G, Rakocevic-Stojanovic V, Savic-Pavicevic D. Repeat Interruptions modify age at onset in myotonic dystrophy type 1 by stabilizing DMPK expansions in somatic cells. *Front Genet*. 2018;9:601. doi: 10.3389/fgene.2018.00601
- [274] Cumming SA, Hamilton MJ, Robb Y, Gregory H, McWilliam C, Cooper A, et al. *De novo* gain of variant repeats in myotonic dystrophy type 1 is associated with reduced somatic instability and mild or absent clinical features. *Eur J Hum Genet*. 2018;26:1635-47. doi: 10.1038/s41431-018-0156-9
- [275] Chong SS, Almqvist E, Telenius H, LaTray L, Nichol K, Bourdelat-Parks B, et al. Contribution of DNA sequence and CAG size to mutation frequencies of intermediate alleles for Huntington disease: Evidence from single sperm analyses. *Hum Mol Genet*. 1997;6(2):301-9.
- [276] Becanovic K, Norremolle A, Neal SJ, Kay C, Collins JA, Arenillas D, et al. A SNP in the *HTT* promoter alters NF-kappaB binding and is a bidirectional genetic modifier of Huntington disease. *Nat Neurosci*. 2015;18(6):807-16. doi: 10.1038/nn.4014
- [277] Busan S, Weeks KM. Role of context in RNA structure: Flanking sequences reconfigure CAG motif folding in huntingtin exon 1 transcripts. *Biochemistry*. 2013;52(46):8219-25. doi: 10.1021/bi401129r
- [278] Krauss S, Griesche N, Jastrzebska E, Chen C, Rutschow D, Achmuller C, et al. Translation of *HTT* mRNA with expanded CAG repeats is regulated by the MID1-PP2A protein complex. *Nat Commun*. 2013;4:1511. doi: 10.1038/ncomms2514
- [279] Marti E. RNA toxicity induced by expanded CAG repeats in Huntington's disease. *Brain Pathol*. 2016;26(6):779-86. doi: 10.1111/bpa.12427
- [280] Neueder A, Landles C, Ghosh R, Howland D, Myers RH, Faull RLM, et al. The pathogenic exon 1 HTT protein is produced by incomplete splicing in Huntington's disease patients. *Sci Rep*. 2017;7:1307. doi: 10.1038/s41598-017-01510-z
- [281] Banez-Coronel M, Ayhan F, Tarabochia AD, Zu T, Perez BA, Tusi SK, et al. RAN translation in Huntington disease. *Neuron*. 2015;88(4):667-77. doi: 10.1016/j.neuron.2015.10.038
- [282] Ciosi M, Cumming SA, Chatzi A, Larson E, Tottey W, Lomeikaite V, et al. Approaches to sequence the *HTT* CAG repeat expansion and quantify repeat length variation. *J Huntingtons Dis*. 2020. doi: 10.3233/JHD-200433
- [283] Maiuri T, Hung CLK, Suart CE, Begeja N, Barba-Bazan C, Peng Y, et al. DNA repair in neurodegeneration: Somatic expansion and alternative hypotheses. *J Huntingtons Dis*. 2020;10.3233/JHD-200414. doi: 10.3233/JHD-200414
- [284] Morton AJ, Glynn D, Leavens W, Zheng Z, Faull RL, Skepper JN, et al. Paradoxical delay in the onset of disease caused by super-long CAG repeat expansions in R6/2 mice. *Neurobiol Dis*. 2009;33(3):331-41. doi: 10.1016/j.nbd.2008.11.015
- [285] Cowin RM, Roscic A, Bui N, Graham D, Paganetti P, Jankowsky JL, et al. Neuronal aggregates are associated with phenotypic onset in the R6/2 Huntington's disease transgenic mouse. *Behav Brain Res*. 2012;229(2):308-19. doi: 10.1016/j.bbr.2011.12.045
- [286] Dragatsis I, Goldowitz D, Del Mar N, Deng YP, Meade CA, Liu L, et al. CAG repeat lengths  $\geq 335$  attenuate the phenotype in the R6/2 Huntington's disease transgenic mouse. *Neurobiol Dis*. 2009;33(3):315-30. doi: 10.1016/j.nbd.2008.10.009
- [287] Donaldson J, Powell S, Rickards N, Holmans P, Jones L. What is the pathogenic CAG expansion length in Huntington's disease? *J Huntingtons Dis*. 2020;10.3233/JHD-200445. doi: 10.3233/JHD-200445
- [288] Lee JM, Ramos EM, Lee JH, Gillis T, Mysore JS, Hayden MR, et al. CAG repeat expansion in Huntington disease determines age at onset in a fully dominant fashion. *Neurology*. 2012;78(10):690-5. doi: 10.1212/WNL.0b013e318249f683
- [289] Cubo E, Martinez-Horta SI, Santalo FS, Descalls AM, Calvo S, Gil-Polo C, et al. Clinical manifestations of homozygote allele carriers in Huntington disease. *Neurology*. 2019;92(18):e2101-e8. doi: 10.1212/WNL.00000000000007147
- [290] Kaplan S, Itzkovitz S, Shapiro E. A universal mechanism ties genotype to phenotype in trinucleotide diseases. *PLoS Comput Biol*. 2007;3(11):e235.
- [291] Tang H, Kirkness EF, Lippert C, Biggs WH, Fabani M, Guzman E, et al. Profiling of short-tandem-repeat disease alleles in 12,632 human whole genomes. *Am J Hum Genet*. 2017;101(5):700-15. doi: 10.1016/j.ajhg.2017.09.013
- [292] Dolzhenko E, van Vugt J, Shaw RJ, Bekritsky MA, van Blitterswijk M, Narzisi G, et al. Detection of long repeat expansions from PCR-free whole-genome sequence data. *Genome Res*. 2017;27(11):1895-903. doi: 10.1101/gr.225672.117

- [293] Bettencourt C, Hensman-Moss D, Flower M, Wiethoff S, Brice A, Goizet C, et al. DNA repair pathways underlie a common genetic mechanism modulating onset in polyglutamine diseases. *Ann Neurol*. 2016;79(6):983-90. doi: 10.1002/ana.24656
- [294] Montermini L, Kish SJ, Jiralerspong S, Lamarche JB, Pandolfo M. Somatic mosaicism for Friedreich's ataxia GAA triplet repeat expansions in the central nervous system. *Neurology*. 1997;49(2):606-10.
- [295] Bidichandani SI, Purandare SM, Taylor EE, Gumin G, Machkhas H, Harati Y, et al. Somatic sequence variation at the Friedreich ataxia locus includes complete contraction of the expanded GAA triplet repeat, significant length variation in serially passaged lymphoblasts and enhanced mutagenesis in the flanking sequence. *Hum Mol Genet*. 1999;8(13):2425-36.
- [296] Sharma R, Bhatti S, Gomez M, Clark RM, Murray C, Ashizawa T, et al. The GAA triplet-repeat sequence in Friedreich ataxia shows a high level of somatic instability *in vivo*, with a significant predilection for large contractions. *Hum Mol Genet*. 2002;11(18):2175-87.
- [297] Hellenbroich Y, Schwinger E, Zuhlke C. Limited somatic mosaicism for Friedreich's ataxia GAA triplet repeat expansions identified by small pool PCR in blood leukocytes. *Acta Neurol Scand*. 2001;103(3):188-92. doi: 10.1034/j.1600-0404.2001.103003188.x
- [298] De Biase I, Rasmussen A, Endres D, Al-Mahdawi S, Monticelli A, Coccozza S, et al. Progressive GAA expansions in dorsal root ganglia of Friedreich's ataxia patients. *Ann Neurol*. 2007;61(1):55-60.
- [299] De Biase I, Rasmussen A, Monticelli A, Al-Mahdawi S, Pook M, Coccozza S, et al. Somatic instability of the expanded GAA triplet-repeat sequence in Friedreich ataxia progresses throughout life. *Genomics*. 2007;90(1):1-5. doi: 10.1016/j.ygeno.2007.04.001
- [300] Ku S, Soragni E, Campau E, Thomas EA, Altun G, Laurent LC, et al. Friedreich's ataxia induced pluripotent stem cells model intergenerational GAATC triplet repeat instability. *Cell Stem Cell*. 2010;7(5):631-7. doi: 10.1016/j.stem.2010.09.014
- [301] Halabi A, Ditch S, Wang J, Grabczyk E. DNA mismatch repair complex MutSbeta promotes GAA.TTC repeat expansion in human cells. *J Biol Chem*. 2012;287(35):29958-67. doi: 10.1074/jbc.M112.356758
- [302] Halabi A, Fuselier KTB, Grabczyk E. GAA\*TTC repeat expansion in human cells is mediated by mismatch repair complex MutLgamma and depends upon the endonuclease domain in MLH3 isoform one. *Nucleic Acids Res*. 2018;46(8):4022-32. doi: 10.1093/nar/gky143
- [303] Ezzatizadeh V, Pinto RM, Sandi C, Sandi M, Al-Mahdawi S, Te Riele H, et al. The mismatch repair system protects against intergenerational GAA repeat instability in a Friedreich ataxia mouse model. *Neurobiol Dis*. 2012;46(1):165-71. doi: 10.1016/j.nbd.2012.01.002
- [304] Du J, Campau E, Soragni E, Ku S, Puckett JW, Dervan PB, et al. Role of mismatch repair enzymes in GAA.TTC triplet-repeat expansion in Friedreich ataxia induced pluripotent stem cells. *J Biol Chem*. 2012;287(35):29861-72. doi: 10.1074/jbc.M112.391961
- [305] Bourn RL, De Biase I, Pinto RM, Sandi C, Al-Mahdawi S, Pook MA, et al. Pms2 suppresses large expansions of the (GAA.TTC)<sub>n</sub> sequence in neuronal tissues. *PLoS One*. 2012;7(10):e47085. doi: 10.1371/journal.pone.0047085
- [306] Ezzatizadeh V, Sandi C, Sandi M, Anjomani-Virmouni S, Al-Mahdawi S, Pook MA. MutLalpha heterodimers modify the molecular phenotype of Friedreich ataxia. *PLoS One*. 2014;9(6):e100523. doi: 10.1371/journal.pone.0100523
- [307] Lokanga RA, Zhao XN, Usdin K. The mismatch repair protein MSH2 is rate limiting for repeat expansion in a fragile X premutation mouse model. *Hum Mutat*. 2014;35(1):129-36. doi: 10.1002/humu.22464
- [308] Zhao XN, Kumari D, Gupta S, Wu D, Evanitsky M, Yang W, et al. Mutsbeta generates both expansions and contractions in a mouse model of the Fragile X-associated disorders. *Hum Mol Genet*. 2015;24(24):7087-96. doi: 10.1093/hmg/ddv408
- [309] Zhao XN, Lokanga R, Allette K, Gazy I, Wu D, Usdin K. A MutSbeta-dependent contribution of MutSalpha to repeat expansions in Fragile X premutation mice? *PLoS Genet*. 2016;12(7):e1006190. doi: 10.1371/journal.pgen.1006190
- [310] Zhao X, Zhang Y, Wilkins K, Edelmann W, Usdin K. MutLgamma promotes repeat expansion in a Fragile X mouse model while EXO1 is protective. *PLoS Genet*. 2018;14(10):e1007719. doi: 10.1371/journal.pgen.1007719
- [311] Miller CJ, Kim GY, Zhao X, Usdin K. All three mammalian MutL complexes are required for repeat expansion in a mouse cell model of the Fragile X-related disorders. *PLoS Genet*. 2020;16(6):e1008902. doi: 10.1371/journal.pgen.1008902
- [312] Hayward BE, Steinbach PJ, Usdin K. A point mutation in the nuclease domain of MLH3 eliminates repeat expansions in a mouse stem cell model of the Fragile X-related disorders. *Nucleic Acids Res*. 2020;48(14):7856-63. doi: 10.1093/nar/gkaa573
- [313] Benn C, Gibson K, D. RD. Drugging DNA damage repair pathways for trinucleotide repeat expansion diseases. *J Huntingtons Dis*. 2020. doi: 10.3233/JHD-200421
- [314] Budworth H, Harris FR, Williams P, Lee DY, Holt A, Pahnke J, et al. Suppression of somatic expansion delays the onset of pathophysiology in a mouse model of Huntington's disease. *PLoS Genet*. 2015;11(8):e1005267. doi: 10.1371/journal.pgen.1005267
- [315] Suelves N, Kirkham-McCarthy L, Lahue RS, Gines S. A selective inhibitor of histone deacetylase 3 prevents cognitive deficits and suppresses striatal CAG repeat expansions in Huntington's disease mice. *Sci Rep*. 2017;7(1):6082. doi: 10.1038/s41598-017-05125-2
- [316] Polyzos AA, Wood NI, Williams P, Wipf P, Morton AJ, McMurray CT. XJB-5-131-mediated improvement in physiology and behaviour of the R6/2 mouse model of Huntington's disease is age- and sex- dependent. *PLoS One*. 2018;13(4):e0194580. doi: 10.1371/journal.pone.0194580
- [317] Nakamori M, Panigrahi GB, Lanni S, Gall-Duncan T, Hayakawa H, Tanaka H, et al. A slipped-CAG DNA-binding small molecule induces trinucleotide-repeat contractions *in vivo*. *Nat Genet*. 2020;52(2):146-59. doi: 10.1038/s41588-019-0575-8.