Mosaicism in Short Tandem Repeat Disorders: A Clinical Perspective

Rose M. Doss [1], Susana Lopez-Ignacio [1], Anna Dischler [1], Laurel Hiatt [2], Harriet Dashnow [3], Martin W. Breuss [1,*] and Caroline M. Dias [1,4,*]

1 Section of Genetics and Metabolism, Department of Pediatrics, University of Colorado Anschutz Medical Campus, Aurora, CO 80045, USA
2 Department of Human Genetics, University of Utah School of Medicine, Salt Lake City, UT 84132, USA
3 Department of Biomedical Informatics, University of Colorado Anschutz Medical Campus, Aurora, CO 80045, USA
4 Section of Developmental Pediatrics, Department of Pediatrics, University of Colorado Anschutz Medical Campus, Aurora, CO 80045, USA
* Correspondence: martin.breuss@cuanschutz.edu (M.W.B.); caroline.dias@cuanschutz.edu (C.M.D.)

Abstract: Fragile X, Huntington disease, and myotonic dystrophy type 1 are prototypical examples of human disorders caused by short tandem repeat variation, repetitive nucleotide stretches that are highly mutable both in the germline and somatic tissue. As short tandem repeats are unstable, they can expand, contract, and acquire and lose epigenetic marks in somatic tissue. This means within an individual, the genotype and epigenetic state at these loci can vary considerably from cell to cell. This somatic mosaicism may play a key role in clinical pathogenesis, and yet, our understanding of mosaicism in driving clinical phenotypes in short tandem repeat disorders is only just emerging. This review focuses on these three relatively well-studied examples where, given the advent of new technologies and bioinformatic approaches, a critical role for mosaicism is coming into focus both with respect to cellular physiology and clinical phenotypes.

Keywords: genomic mosaicism; methylation mosaicism; short tandem repeats; *FMR1*; *HTT*; *DMPK*; fragile X; Huntington disease; myotonic dystrophy type 1

1. Introduction

Short tandem repeats (STRs), also referred to as microsatellites, are 1–6 base pair regions of repetitive DNA, accounting for approximately 3–5% of the human genome [1]. Tandem repeat disorders (TRDs), or repeat expansion disorders (REDs), occur when a tandem repeat expands or contracts from a benign to a pathogenic range, resulting in loss-of-function or toxic gain-of-function effects [2,3]. There are more than 60 known human TRDs, including fragile X syndrome (FXS), Huntington disease (HD), spinocerebellar ataxia, and myotonic dystrophy type 1 (DM1) [4–6]. Additionally, tandem repeat expansions have been linked to autism, Parkinson's disease, schizophrenia, and cancer [7–12]. Recently, genome-wide analyses have identified a role for STRs in gene expression, higher-order chromatin organization, and complex traits such as height and biomarkers of health [13–17]. Thus, STRs play central and critical roles in the cellular physiology of human health and disease.

Compared to most of the human genome, STRs exhibit higher mutation rates and instability [18,19]. In addition to germline instability, STRs also exhibit instability in somatic tissue, particularly when they exceed a pathogenic threshold: STRs can expand, contract, and acquire sequence changes. In some cases, this somatic repeat number instability may drive disease progression. Somatic alterations in epigenetic marks in different cells of the body within the

same individual can also occur. For example, DNA methylation is a critical epigenetic signature at GC-rich CpG islands that can inactivate gene expression [20]. Length and methylation mosaicism lead to significant cellular variability and have important consequences for cellular pathophysiology and clinical phenotypes. However, given historical technical limits, our understanding of the role of mosaicism in STR-linked disorders is only just emerging.

Over one million variable STR loci have been identified in the human genome, most of which are found in noncoding regions, with only 8% in coding regions [21,22]. However, coding or gene proximal STR loci contribute disproportionately to human disease. The objective of this review is to explore the pathogenic impact of STR mosaicism with a specific focus on *FMR1*, *HTT*, and *DMPK*, i.e., fragile X, HD, and DM1. These are clinically well-described TRDs, and there is an emerging appreciation for the role of mosaicism in these disorders from both the cellular and clinical perspectives.

2. Tandem Repeat Disorders

As mentioned above, there are more than 60 known TRDs in humans. While they can greatly differ in the length and composition of their repeat unit, common themes emerge (Figure 1). For instance, some neurodegenerative disorders are driven by polyglutamine tracks [23]. The overall landscape of TRDs has been expertly reviewed before [24]. Here, we focus on a summary of fragile X, HD, and DM1 as they are most pertinent to our discussion of STR mosaicism. However, there is growing appreciation for somatic instability in additional TRDs [25,26]. Additionally, distinct repetitive DNA elements, such as variable number tandem repeats, have also been linked to neurological disease, but this is beyond the scope of this review.
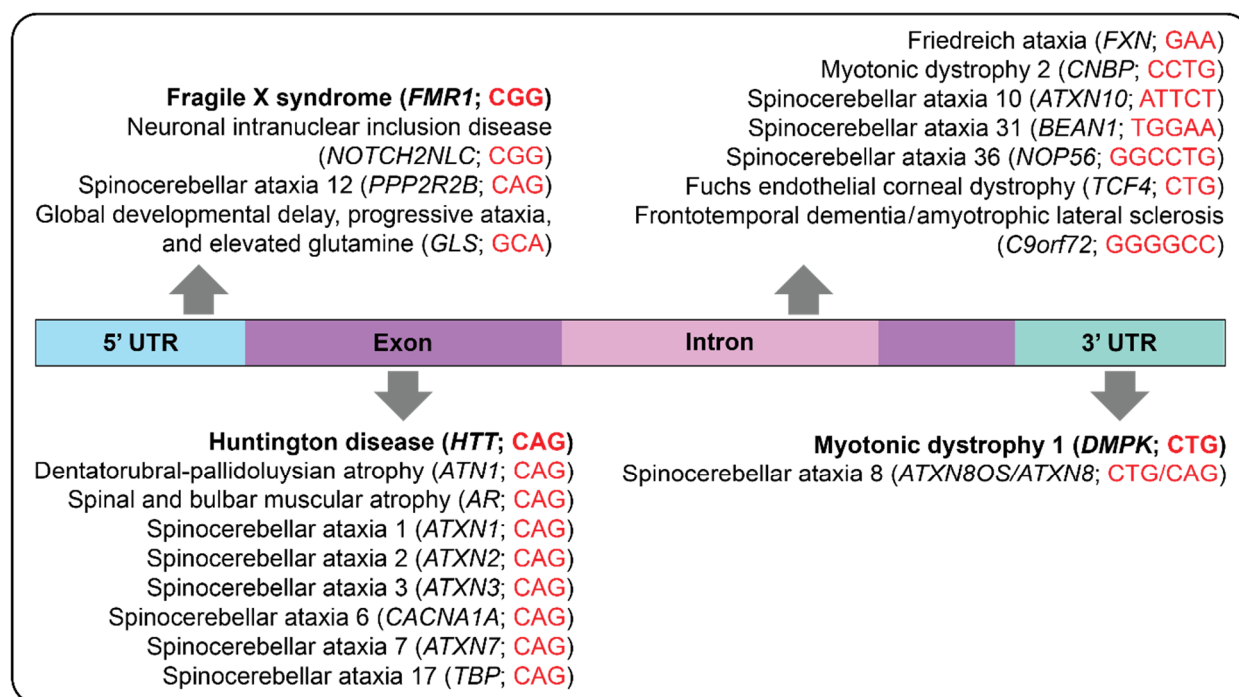

Figure 1. Genetic location and short tandem repeat motif associated with example disorders. The majority of known STR disorders are found within the exonic and intronic regions of the affected gene. Disorders of interest for this review are distinguished in bold. *FMR1*; #300624, *NOTCH2NLC*; #603472, *PPP2R2B*; #604326, *GLS*; #618412, *HTT*; #143100, *ATN1*; #125370, *AR*; #313200, *ATXN1*; #164400, *ATXN2*; #183090, *ATXN3*; #109150, *CACNA1A*; #183086, *ATXN7*; #164500, *TBP*; #607136, *FXN*; #229300, *CNBP*; #602668, *ATXN10*; #603516, *BEAN1*; #117210, *NOP56*; #614153, *TCF4*; #613267, *C9orf72*; #105550, *DMPK*; #160900, *ATXN8OS/ATXN8*; #608768. # = Phenotype MIM number..

2.1. Fragile X-Related Conditions

Fragile X-related conditions are caused by CGG expansions positioned in the 5′ untranslated region (UTR) of *FMR1* (fragile X messenger ribonucleoprotein 1), located on the X chromosome [27–32]. Among the general population of unaffected individuals, the CGG motif ranges between approximately 5 and 40 repeats. Intermediate alleles, or gray zone alleles, of approximately 41–54 repeats, are not classically considered to confer risk for fragile X-related conditions the way the full mutation and premutation do. Studies seeking to confirm separate associations of these gray zone alleles with distinct clinical outcomes, such as Parkinsonian symptoms and higher risk of death, have met with conflicting results [27,33]. Thus, the clinical impact of gray zone alleles remains controversial. The premutation range, which is ~55–200 repeats, is characterized by increased *FMR1* mRNA [28] but variably reduced *FMR1* protein (FMRP). The premutation is associated with an increased risk of acquiring multiple adult-onset conditions, including fragile X-associated primary ovarian insufficiency (FXPOI), fragile X-associated tremor/ataxia syndrome (FXTAS), and fragile X-associated neuropsychiatric disorders (FXAND) [28,29]. There are a variety of hypotheses surrounding the cellular pathophysiology of FXTAS, which are described extensively elsewhere, including DNA damage and aberrant gain of function toxic RNA and peptide formation [30,31]. Moreover, the premutation allele is particularly unstable in the human female germline, resulting in increased risk in offspring with the full mutation [32], described below.

The *FMR1* full mutation is more than 200 CGG repeats in length, which leads to the well-characterized neurodevelopmental disorder FXS [34]. FXS results from hypermethylation of the CGG repeats, leading to the inactivation of *FMR1* and absent FMRP, a protein crucial during brain development [35,36]. This disorder is a leading inherited form of intellectual disability and has been described as the most common single-gene cause of autism, accounting for 1–6% of autism cases [37]. Other notable symptoms include hyperactivity, impulsivity, anxiety, poor language development, and epilepsy [37]. FXS is inherited in an X-linked dominant pattern. Both males and females can be affected, albeit with differences in severity and progression, much of which is attributed to stochastic patterns of X-chromosome inactivation in females [38,39]. Interestingly, although there are also neurodevelopmental phenotypes, in general, the premutation does not cause FXS per se, and the full mutation does not lead to premutation-associated disorders, implicating unique cellular pathophysiology in each condition.

2.2. Huntington Disease

HD is an autosomal dominant progressive neurodegenerative disorder caused by a CAG repeat expansion in the first exon of *HTT*, located on chromosome four [40]. This coding repeat expansion produces a polyglutamine (polyQ) tract in the protein product. Characteristics of HD include uncontrollable motor movements, psychiatric symptoms, and progressive cognitive decline [41]. *HTT* encodes huntingtin, a widely expressed brain protein, which plays a role in brain development, vesicle and organelle trafficking, and neuroprotection [42–46]. In HD, the CAG repeat-associated polyglutamine tract leads to the formation of insoluble aggregates, which have been found in the striatum, cerebellum, spinal cord, and cortex [47]. These aggregates have also been found in neuronal processes, potentially disrupting synaptic transmission [48]. Striatal projection neurons are uniquely vulnerable, with specific degeneration of these cell types observed in HD neuropathology [49].

At the HD locus, unaffected individuals have 10–35 CAG repeats. Those with complete penetrance of the disease have more than 40 repeats, while incomplete penetrance is associated with an expansion size ranging from 36 to 39 [50]. Additionally, the age of onset

correlates inversely with the size of the repeat [51–53], with a rare juvenile onset form caused by CAG repeats above 60 [54].

2.3. Myotonic Dystrophy Type 1

DM1 is an autosomal dominant condition caused by a CTG repeat in the 3′ UTR of DMPK, located on chromosome 19 [55]. The typical allele size of this CTG repeat is 5–34, while premutation carriers have 35–49 repeats. Although individuals with this premutation range have no documented signs of disease, their children have an increased risk of inheriting larger repeat expansions and developing symptoms [56]. The full-penetrance pathogenic alleles are those with CTG repeats of ≥50. Clinical severity of DM1 is highly variable and associated with repeat size, ranging from mild presentations with normal lifespan all the way to severe congenital forms associated with repeats into the range of thousands; this form has a strong bias toward maternal transmission [57,58]. Overall, DM1 is slowly progressive and multisystemic, with symptoms including muscle weakness, myotonia, psychiatric symptoms, MRI abnormalities, cataracts, diabetes mellitus, hypogonadism, and changes on an electrocardiogram [59]. Care is generally supportive, and like fragile X and HD, there is no cure.

DMPK encodes the protein myotonic dystrophy protein kinase, which is involved in maintaining skeletal muscle integrity, cardiac conduction, ion-channel regulation, and metabolism [60]. It is suggested that this expansion in the 3′ UTR of the gene causes toxic RNA accumulation, resulting in multiple molecular defects, with aberrant RNA splicing thought to be a central factor contributing to the array of disease manifestations in patients [61]. A study on the neurological impact of DM1 by Dincă and colleagues found this toxicity seems to particularly impact astrocytes by interfering with transcripts that are involved in regulating synaptic function [62].

3. Features and Mechanisms of STR Instability

Historically, STRs have been reported to mutate at approximately 1×10^{-4}–1×10^{-3} nucleotides per generation, which is orders of magnitude higher than estimates of approximately 1×10^{-8} for SNVs in unique regions of DNA [63]. These mutation rates are influenced by factors such as the repeat unit size, repeat number, repeat sequence, flanking sequence, as well as sex and age [64–68]. For instance, longer repeat tracts are more prone to slippage during replication and, thus, expansion or contraction events [69–72]. Several recent studies have expanded our understanding of STR mutability genome-wide in human families, finding de novo mutation rates in the range of 5×10^{-6}–5×10^{-5} [73]. These studies considered millions of STR loci within thousands of trios, as well as dozens of multi-generational families [19,73–75]. On a genome-wide scale, this translates to an estimated 62–85 de novo STR mutations per generation. We briefly discuss below some pertinent examples of important features and mechanisms of STR instability; however, we refer the interested reader to the excellent reviews that more comprehensively address this issue [2,76].

Repeat Sequence Content: In addition to length, sequence content has also been shown to be a determinant of instability, with important clinical consequences. In the case of *FMR1*, perfect repeats appear to be more mutable across generations. For example, the *FMR1* CGG locus contains AGG insertions; however, the loss of one or more of these AGG motifs is associated with increased intergenerational expansion [77]. In HD, loss of a CAA interruption at the end of the CAG repeat tract in intermediate allele carriers is associated with a younger age of onset as well as evidence of increased somatic instability [78]. In DM1, repeat interruptions have been associated with reduced somatic instability and milder clinical phenotypes [79–81]. We should note that there are some rare examples of

repeat interruptions associated with increased instability and severity, potentially through increased stability of RNA hairpin structures [82,83]. Thus, sequence purity has important consequences on both germline and somatic instability, as well as clinical outcomes, but the nature of this relationship depends on the specific locus.

Epigenetics: The epigenetic environment may play a critical role in shaping mosaicism. For example, in fragile X syndrome, the CGG expansion is associated with not only DNA hypermethylation of the locus but also nearby histone deacetylation and histone lysine 9 methylation [84–86]. This heterochromatinization may have global effects, given recent work showing widespread heterochromatinization of long synaptic genes across the genome, associated with potential evidence of stepwise STR instability at distinct loci in fragile X syndrome [87]. Work in embryonic stem cells suggests switching *FMR1* off occurs early in development; using preimplantation human embryos, it was observed that undifferentiated human embryonic stem cells retain *FMR1* expression and unmethylated DNA and that *FMR1* silencing only occurs with differentiation—representing a critical developmental window during which mosaicism could arise [88]. Conversely, demethylation of the locus has been targeted in studies attempting to devise novel therapeutic strategies [89]. Artificial demethylation of the locus leads to R-loop formation (non-canonical DNA:RNA hybrids) that mediates contraction of the repeat in fragile X syndrome, which has relevance not only for understanding mosaicism, but also gene therapy approaches more broadly. In DM1, there is notable variability in methylation of the flanking sequence to the CTG repeat across individuals and tissues. This variation in methylation may moderate the instability of the repeat itself as well as biomarkers for clinical subgroups [90]. For example, Koehorst and colleagues found that unmethylated mothers passed on CTG expansions, but methylated mothers transmitted a contraction [91].

Mechanisms of instability: What underlies the high mutability of STRs? STR expansions are frequently linked to DNA replication, where strand slippage due to nascent and template strands misaligning during the replication process can result in changes in length [92]. However, there are also changes in length observed in post-mitotic cells; for example, there is evidence from both mouse models and human cases showing that the instability of some repeat expansions in somatic cells is influenced by the DNA repair machinery, suggesting additional distinct post-mitotic mechanisms that may contribute to disease progression. Cell-type-specific expression of repair machinery, such as mismatch repair or transcription-coupled repair, may further underlie master controls of repeat instability [93–98]. Thus, while there is evidence of critical cellular processing pathways that are likely common between STRs, there are also differences mediated by cell-type-specific gene expression profiles.

STR instability can result in changes in disease severity across generations due to mosaicism in germ cells (Figure 2A), i.e., anticipation. Likewise, mosaicism within an individual can also lead to variability across cells within an individual beyond the germline, theoretically impacting disease severity depending on where in the body this occurs (Figure 2B). Finally, given cell-type-specific expression of factors, such as DNA repair machinery, this variability may differ across different cell types, which could lead to specific, predictable phenotypes (Figure 2C).
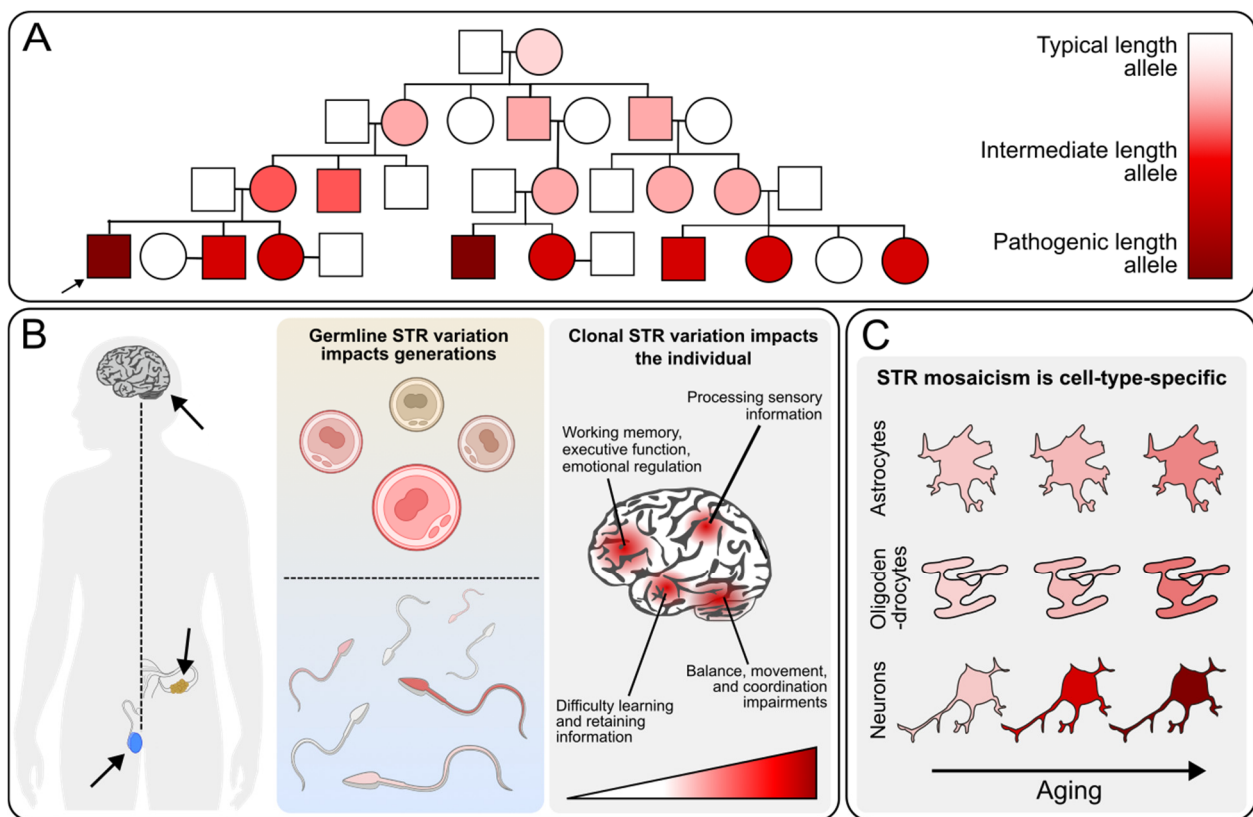
Figure 2. Illustration of the impact of STR instability over generations and the clinical significance of STR mosaicism that can be found within individuals. (**A**) Example pedigree demonstrating expansion across generations, particularly in the female germline for fragile X. (**B**) Mosaicism in the brain can impact clinical phenotypes of neurological disorders, while germline variation impacts generational changes, which can lead to differential expansion in sperm (HD) and egg cells (fragile X, DM1) depending on the specific STR. The right panel demonstrates the function of different brain regions, which could theoretically be impacted by clonal variation. (**C**) Somatic mosaicism, in some cases such as Huntington disease, may be developmentally dependent and cell-type-specific but not necessarily tied to proliferative status.

4. Mosaicism and STRs

In the past two decades, sequencing improvements have allowed us to appreciate the mosaic nature of the human body and how this affects health and disease (See Section 6 below). Broadly speaking, mosaicism (of different forms of genetic variation and not just STRs) has been shown to increase with age and contribute to cardiovascular disease, liver disease, changes in inflammation status, schizophrenia, and cancer [99–102]. Mosaic chromosomal alterations, such as aneuploidy and large-scale copy number variations (CNVs), have been associated with neurodevelopmental and neurodegenerative disorders, including autism, Alzheimer's disease, and Parkinson's disease [103–107]. Additionally, the accumulation of somatic mutations in stem cell populations has been proposed as a mechanism underlying age-related decline and regenerative capacity [108].

STR mosaicism has been more challenging to study due to obstacles in accurately measuring these repetitive regions. In the context of fragile X, although both length and methylation mosaicism are frequently described, we know surprisingly little about how mosaicism contributes to clinical phenotypes (see below). There have been several case reports identifying mosaicism of the CGG repeat length, where some cells carry the full mutation allele, and others carry premutation or normal range alleles [109–111]. There is also methylation mosaicism, where cell populations may contain either methylated,

unmethylated, or partially methylated *FMR1*. There is evidence that size and methylation mosaicism alter FMRP production, presumably driving clinical phenotypes [112,113]. However, given the differing phenotypes of premutation and full mutation carriers, as well as the difficulty of studying the human brain directly, the full clinical implications of *FMR1* mosaicism remain to be seen. It is tempting to speculate that the frequency and complexity of mosaicism in fragile X-related conditions could drive the impressive clinical heterogeneity observed in these disorders, even in carriers with measured expansions of similar sizes.

FMR1 is located on the X-chromosome and is therefore subject to the stochastic process of X-chromosome inactivation [38,39]; this process maintains gene dosage in XX female cells by turning off one X-chromosome early in development, and female carriers—as a consequence—are naturally mosaic for which X chromosome is active. The activation ratio, which is the percentage of cells that carry the benign *FMR1* allele as active, likely contributes to variable disease severity in females with fragile X-related conditions. In fact, there is evidence that the activation ratio correlates both with cellular and clinical phenotypes [114]. However, most studies that assess the activation ratio in female carriers primarily study blood, which may differ from patterns within the human brain. Thus, studies of peripheral tissues may underestimate the true impact of activation ratio on clinical outcomes. Understanding *FMR1* dynamics in relation to sex is even more important when one considers an increased predilection for instability within the female germline [115].

5. Clinical Impact of STR Mosaicism and Disease Severity

The following studies shed light on the intricate interplay between CGG repeat size and methylation patterns in shaping the clinical spectrum of FXS phenotypes, as well as the opportunities for future progress.

Fernandez and colleagues presented two unrelated families, each having one child with FXS and one sibling with an FXS-like phenotype [116]. Each of the siblings presenting with the FXS-like phenotype was revealed to have mosaicism in CGG repeat size, methylation across different tissues, or both. In Case 1, the identified alleles ranged from premutation to full mutation sizes, with varying levels of methylation. In Case 2, only premutation sizes were identified, with unexpected evidence of some methylation. The authors proposed that the size diversity of the repeat alleles and their methylation status have direct effects on mRNA and protein levels, which could explain the FXS-like phenotype [116].

Larger cohorts have also provided convincing evidence of the importance of mosaicism in fragile X-related conditions. Meng and colleagues conducted a comprehensive analysis of 487 males with FXS, illuminating the differential impact of methylation and size mosaicism on cognitive and behavioral outcomes [117]. Overall, they found evidence of size and/or methylation mosaicism in over 30% of the cohort. Their findings showed a milder phenotype in individuals with methylation mosaicism compared to those without, while size-related mosaicism showed less pronounced associations. Limitations of this study, and many others studying the clinical relationship of mosaicism, include that mosaic status was inferred from blood testing, which may or may not correlate with brain-specific changes [117]. Pretto and colleagues examined 18 individuals with FXS, 13 of whom were identified to have mosaicism when comparing peripheral blood mononuclear cells (PBMCs) and fibroblasts [110]. Among the mosaic cases, the study identified variation in methylation patterns and CGG repeat sizes between the tissues, with higher levels of *FMR1* mRNA being found in the blood than in fibroblasts. Notably, they found that cognitive improvement scores were positively correlated with the subtle increases in FMRP levels in the mosaic cases, highlighting the significance of lower methylation compared with full methylation, full mutation alleles, on clinical outcomes [110]. Interestingly, a study by Baker and colleagues found that although premutation mosaicism was associated with improved intellectual functioning in individuals

with fragile X syndrome, autistic features were similar regardless of the presence of mosaicism, once controlling for IQ [118]. Mosaicism is also well described in the *FMR1* premutation. In a recent study of premutation females, mosaicism was identified in the buccal tissue in the majority of the participants and associated with unexpected health outcomes [119]. Thus, size and methylation mosaicism in fragile X-related conditions may be the rule, as opposed to the exception, and they impact some domains of clinical severity.

In HD, instability of CAG repeats in both the testes and brain has been described for over 30 years [120]. Larger repeat lengths are associated with earlier disease onset, making somatic instability a marker for disease pathogenesis [121]. Larger tracts are also observed to become more unstable in postmitotic neurons [122]. Somatic instability is thought to be caused by differences in DNA repair proteins and pathways, for instance, mismatch repair due to the formation of secondary structures [123]. HD is also known to exhibit meiotic instability, as it is observed to contribute to repeat size growth over generations, particularly in the male germline [124,125].

In HD, somatic expansion has been shown to determine clinical outcomes beyond the germline expansion size itself [126]. Recent work [127,128] has also made tremendous advances in providing support for the Kaplan hypothesis [129], which states that the expansion of TRDs beyond pathogenic thresholds in specific cell types mediates clinical progression. Specifically, Handsaker and colleagues [128] used single-cell genomic approaches in human brain tissue and found that it was not the progression beyond 40 repeat units that led to aberrant gene expression, but rather, the cell-specific somatic expansion of these alleles to greater than 150 repeat units led to widespread gene dysregulation and markers of senescence and apoptosis in striatal neurons. Their work suggests that HD reflects a lifelong process of expansion in vulnerable cell types and has profound implications for revolutionizing our understanding of diagnosis and therapeutics more broadly for TRDs [130]. Other neuronal subtypes in HD have also been demonstrated to have unique vulnerability to pathogenic somatic expansions, including layer 5a pyramidal neurons [131]. However, these authors found that other, more resilient neuronal subtypes also demonstrated evidence of STR expansion and posited that altered connectivity is an additional "hit" necessary for neuronal loss. Why some neurons are uniquely vulnerable to expansion is an important critical area of open investigation.

In DM1, there is well-established instability of the CTG expansion in both somatic tissues and germline. For example, the sperm from male individuals with DM1 show extensive expansion size variability across individuals [132]. Additional analyses taking advantage of data from preimplantation genetic diagnostics found that oocytes and embryos from females with DM1 displayed larger increases in repeat numbers, while the sperm and embryos from males with DM1 had smaller increases at ~50–100 repeats [133], consistent with primary transmission in the female germline.

A longitudinal study observed how CTG repeat mosaicism changed over time in the blood of 43 DM1 patients [134]. They found that the range of repeat sizes and modal allele length generally increased over time, with the rate of change affected by multiple factors that included inherited progenitor allele length, age at sampling, time interval between samples, and sex of the patient. Importantly, individual variation in rates of expansion was associated with differences in age of disease onset. Several studies analyzing the somatic tissue-specificity of CTG expansion rates have shown consistent results of repeat lengths being much larger in muscle than what is observed in blood [135–137], highlighting the significance of understanding mosaic patterns when considering patient prognosis.

Collectively, these studies underscore the significance and frequency of mosaicism in shaping the clinical heterogeneity in fragile X, HD, and DM1. The findings emphasize the importance of testing multiple tissues and highlight potential challenges in genetic counsel-

ing and management for such cases. For example, could cryptic mosaicism influence the success of clinical interventions, thus necessitating more comprehensive characterization in clinical trials? These case examples also serve as a foundation for the application of similar studies in other TRDs in which mosaicism may be contributing to clinical phenotypes. This is critical as somatic mosaicism is a common feature of numerous, albeit less well-characterized REDs [138].

6. Current STR Detection Methods and Limitations

Several approaches are used to detect STRs (Table 1); however, their reliability, cost, and resolution of mosaicism detection vary. In rare disease genetics, gene panels or short-read whole-exome and genome sequencing remain the standard of care. However, pathogenic STR expansions are challenging to accurately size in short read datasets when the expansion exceeds the insert read length. This necessitates separate, targeted assessments for TRDs depending on clinical suspicion. Established current clinical methods include repeat-primed polymerase chain reaction (PCR) and Southern blot, but these are targeted and cannot accurately determine cell fraction in cases of mosaicism. Triplet-repeat primed PCR is perhaps the most used given the relative ease of implementation; although the specifics may vary, in general, it uses a third fluorescently labeled, repeat-specific primer that generates a characteristic trace following capillary electrophoresis. Multiple alleles can be clearly resolved, including very large alleles and mosaic alleles. Adaptations of this method include the incorporation of a methylation-sensitive enzyme, which allows for inference of methylation state. However, there remains a bias toward the amplification of smaller alleles, which means one cannot draw conclusions about the relative proportion of mosaic alleles. Specifically, very long alleles may demonstrate only small fluorescent peaks and yet comprise a substantial mosaic cell fraction. Furthermore, because of this bias, large expansions representing mosaic cell fractions less than 5–10% may not be reliably detected [139]. Furthermore, there is no base pair resolution, which, as described above, can have important effects on instability [140]. However, PCR "stutter", visualized as dips in the electropherogram, can be used to indirectly infer the presence of repeat interruptions.

Table 1. Overview of techniques to identify STRs and mosaicism. * Although triplet-primed PCR can detect multiple alleles, preferential amplification of smaller peaks limits mosaicism quantification and sizing of large alleles. ** Southern blotting can detect mosaicism if cell fraction is high enough but, again, may be limited with larger expansions in a smaller number of cells. *** New algorithms have been deployed to detect mosaicism of STRs in short-read data; however, this remains an emerging area. **** Caution must be exercised in interpreting mosaicism in whole genome sequencing, even with long reads. Standard 25× whole genome coverage may be insufficient to detect the presence of mosaicism for large expansions.

Method	Triplet-Primed PCR	Southern Blotting	Short-Read Sequencing	Long-Read Sequencing
Description	Modified polymerase chain reaction (PCR) using multiple primers, including one targeting repeat. Produces different-sized DNA due to multiple annealing sites within the repeat region. Visualized by capillary electrophoresis. Digestion with a methylation-sensitive enzyme can be incorporated for methylation detection.	DNA digestion, gel electrophoresis DNA detection	DNA fragmentation followed by massively parallel sequencing. Bioinformatic processing uses alignment algorithms to resolve the repetitive STR and flanking region.	High molecular weight DNA is used; single-molecule real-time or nanopore sequencing provides highly accurate, kb long reads with base pair and methylation level resolution. Bioinformatic processing uses specialized alignment algorithms to resolve the repetitive STR and flanking region, as well as DNA methylation status.

Table 1. *Cont.*

Method	Triplet-Primed PCR	Southern Blotting	Short-Read Sequencing	Long-Read Sequencing
Cost	$	$	$$	$$$
Accuracy	Lower for larger STRs	Lower for larger STRs	Lower for larger STRs	High
Limitations	No sequence information, and smaller amplicons disproportionately amplified	No sequence information, labor intensive, and less scalable than other methods	Less reliable for large or complex STR expansions	Less widely available and most expensive
Advantages	Commercial kits available	No amplification requirement	A large number of datasets amenable to analysis	Accurate and precise. Combination with other approaches (amplification-free enrichment) permits high-coverage analysis.
Methylation Detection	✔	✔	✔	✔
Mosaicism Resolution	✔ *	✔ **	✔ ***	✔ ****

In the past decade, several key advances in STR detection have created a paradigm shift in the field. First, several computational methods have been developed capable of genotyping germline STRs in short-read data, both targeted and genome-wide. These pipelines have unearthed new horizons in the detection of short tandem repeat disorders from short-read sequencing data; however, they remain limited in their sensitivity for some loci and struggle to accurately resolve very long expansions [1,141–146]. Nonetheless, there have been encouraging findings regarding both the specificity and sensitivity of such tools in identifying pathogenic repeat expansions compared to gold-standard approaches [147]. Mosaic STR variation is even more difficult to call from short-read data, although there have been recent advances in this domain as well, including prancSTR [148]. However, prancSTR is most helpful for single high-frequency mosaic alleles, which may not represent true patterns of mosaic variation for large clinically relevant expansions. Overall, these tools open up new opportunities for STR interrogation on countless pre-existing large-scale datasets, particularly for research applications in which accurate sizing may be less critical.

The second major advance that has altered our ability to detect STR expansions is long-read sequencing. Single-molecule real-time (SMRT, PacBio) sequencing and nanopore sequencing [140] read through thousands of nucleotides with sufficient accuracy, even within repetitive regions. Pacbio produces shorter but more accurate reads up to approximately 20 kilobases, while nanopore sequencing can achieve ultralong reads potentially spanning millions of bases. Thus, both are well suited to detect even the longest repeat expansions. The emergence of long-read sequencing and genotyping methods, such as TRGT, LongTR, and Straglr, has allowed for STR detection and visualization from these datasets [149–151]. Another advantage of these long-read sequencing approaches is the direct sequencing of native DNA. For example, SMRT sequencing infers 5-methylcytosine modifications through the detection of polymerase kinetics that are dependent on the presence of base modifications. Likewise, nanopore sequencing can be trained to detect various modifications from the raw conductivity data. There have been recent studies demonstrating the use of long-read sequencing to accurately genotype large STR expansions in a high-throughput capacity [152]. However, standard, genome-wide coverage may be insufficient to detect mosaicism of large expansions. Thus, methods to first enrich the targeted regions, such as with amplification-free Cas9-mediated enrichment and nanopore's adaptive sampling to perform real-time target selection [153–156], allow for higher sequencing depth of repeat expansions, which enables mosaicism detection at a reasonable cost. Other methods, such as indirect sequence capture, are also being proposed

to allow for high coverage assessment of repeat expansions [157]. These approaches as a whole, which allow for high coverage sequencing through even the largest repeats, are poised to accelerate our understanding of STR mosaicism and are likely to become standard in the field.

7. Future Directions on the Clinical Impact of STR Mosaicism

The above cases illustrate the importance of considering STR mosaicism in both cellular and clinical phenotypes, although important gaps in our understanding remain. While studies have reported the presence of STR mosaicism in various disorders, the overall prevalence and patterns of mosaicism across different STRs and tissues are not well-established [158,159]. Even in fragile X-related conditions, we know little about how mosaicism varies by cell type and brain region in the human condition as compared to HD. More comprehensive studies are needed to determine the frequency and characteristics of STR mosaicism in both health and disease, as well as across different cells and tissues in the human condition. STR mosaicism can complicate diagnostic testing and interpretation, as the detection of mosaicism may depend on the accessibility of the tissue sampled and the sensitivity of the testing method [160]. This is of particular relevance to neurologic disorders, as brain tissue is typically not available for study in a clinical population. Rather, information regarding mosaicism may be inferred from the sampling of peripheral tissues, such as blood or skin. A better understanding of the features and patterns of mosaicism from even the most fundamental perspectives should prove helpful, for example, in determining if certain peripheral tissues might better mimic patterns in the brain. Additionally, most studies on STR mosaicism have focused on coding regions, while the prevalence and consequences of STR mosaicism in non-coding regions of the genome are less explored. Investigating STR mosaicism in regulatory regions and its potential effects on gene expression and disease pathogenesis is an important area for future research. Unlike single-nucleotide variants, in which mosaicism can be represented as a binary outcome in a particular cell, STR mosaicism is potentially drastically more complex, with a large range of STR lengths potentially co-existing in a single individual. This nuance needs to be considered in both basic science studies and clinical statistical modeling, as groups are already attempting to do [127]. Finally, and most importantly, we have provided an in-depth overview of three disorders as case examples of the scope of the importance of mosaicism in STRs. However, there are dozens of disorders caused by STR expansions, and in many cases, mosaicism is also important in those contexts. Given that there are likely at least some shared mechanisms driving STR instability, research into these individually rare disorders will likely result in an improved understanding of STR biology more broadly (see Section 3 above).

Although the American College of Medical Genetics incorporates consideration of mosaicism in genetic testing for TRDs such as fragile X [160], our ability to use mosaicism to counsel a patient remains more limited, e.g., what does mosaicism mean for the individual? Establishing evidence-based guidelines for genetic counseling of individuals and families with STR mosaicism will require improved knowledge about the natural history and the mechanisms and consequences of mosaicism in these disorders across the lifespan and across tissues. Given the technical challenges in detecting low-level mosaicism and the complex inheritance patterns of repeat expansion disorders, this is not a trivial task [32,161]. For example, in FXS, the presence of methylation mosaicism can complicate the interpretation of genetic test results and the assessment of disease risk [32]. Moreover, the genetic counseling of individuals and families with STR mosaicism requires a nuanced understanding of the clinical and psychosocial implications of variable repeat expansion [32,161]. The unexpected finding that premutation females can be mosaic for the full mutation [119] is

a prime example of the importance of empirically studying the effects of mosaicism on clinical outcomes in large-scale and longitudinal clinical cohorts.

Improvement of STR sequencing will clarify the range of allele sizes, resulting in variable expression patterns that cause disease [1,119,162]. Comprehensive testing for STR mosaicism is crucial for informed reproductive decision-making, as it can have significant implications for family planning and prenatal testing. A better understanding of the precise mechanisms underlying mosaicism opens opportunities for the development of tandem-repeat targeted CRISPR gene editing, personalized antisense oligonucleotide, biologics, and other small molecule therapeutics [12,163]. Furthermore, given efforts to develop gene therapy to reintroduce *FMR1* in a subset of cells—which essentially creates a unique situation of mosaicism for *FMR1* expression dissociated from the repeat expansion itself—understanding the nuances of the effects of mosaicism could take on even new, unexpected relevance.

8. Conclusions

This review encompasses the role of STR mosaicism in human disease with a focus on fragile X, HD, and DM1 and highlights the importance of discerning mosaicism in determining disease severity and progression and developing potential therapeutics. Appropriate STR resolution methods are critical in identifying the repeats themselves and their mosaic patterns. As sequencing methods and technologies continue to advance, allowing for more efficient and cost-effective analyses, our ability to detect STR mosaicism will improve, with the current emergence of long-read sequencing set to become the standard tool in comprehensive STR analysis [148].

Further exploration into identifying the mosaic patterns of short tandem repeats has important implications for refining disease prognosis and customizing treatment interventions. Currently, the influence of STR mosaicism on disease progression and response to potential therapies is largely in its infancy and is an area that demands longitudinal studies that track changes in STR mosaicism over time and across tissues as a means to assess the impact on clinical outcomes and treatment response.

References

1. Tanudisastro, H.A.; Deveson, I.W.; Dashnow, H.; Macarthur, D.G. Sequencing and characterizing short tandem repeats in the human genome. *Nat. Rev. Genet.* **2024**, *25*, 460–475. [CrossRef] [PubMed]
2. Malik, I.; Kelley, C.P.; Wang, E.T.; Todd, P.K. Molecular mechanisms underlying nucleotide repeat expansion disorders. *Nat. Rev. Mol. Cell Biol.* **2021**, *22*, 589–607. [CrossRef] [PubMed]
3. Gall-Duncan, T.; Sato, N.; Yuen, R.K.C.; Pearson, C.E. Advancing genomic technologies and clinical awareness accelerates discovery of disease-associated tandem repeat sequences. *Genome Res.* **2022**, *32*, 1–27. [CrossRef]
4. Shi, Y.; Niu, Y.; Zhang, P.; Luo, H.; Liu, S.; Zhang, S.; Wang, J.; Li, Y.; Liu, X.; Song, T.; et al. Characterization of genome-wide STR variation in 6487 human genomes. *Nat. Commun.* **2023**, *14*, 2092. [CrossRef] [PubMed]

5. Verkerk, A.J.; Pieretti, M.; Sutcliffe, J.S.; Fu, Y.H.; Kuhl, D.P. 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*, 905–914. [CrossRef] [PubMed]

6. Hong, E.P.; Ramos, E.M.; Aziz, N.A.; Massey, T.H.; McAllister, B.; Lobanov, S.; Jones, L.; Holmans, P.; Kwak, S.; Orth, M.; et al. Modification of Huntington's disease by short tandem repeats. *Brain Commun.* **2024**, *6*, fcae016. [CrossRef]

7. Trost, B.; Engchuan, W.; Nguyen, C.M.; Thiruvahindrapuram, B.; Dolzhenko, E.; Backstrom, I.; Mirceta, M.; Mojarad, B.A.; Yin, Y.; Dov, A.; et al. Genome-wide detection of tandem DNA repeats that are expanded in autism. *Nature* **2020**, *586*, 80–86. [CrossRef]

8. Bustos, B.I.; Billingsley, K.; Blauwendraat, C.; Gibbs, J.R.; Gan-Or, Z.; Krainc, D.; Singleton, A.B.; Lubbe, S.J.; International Parkinson's Disease Genomics Consortium (IPDGC). Genome-wide contribution of common short-tandem repeats to Parkinson's disease genetic risk. *Brain* **2023**, *146*, 65–74. [CrossRef]

9. Mojarad, B.A.; Engchuan, W.; Trost, B.; Backstrom, I.; Yin, Y.; Thiruvahindrapuram, B.; Pallotto, L.; Mitina, A.; Khan, M.; Pellecchia, G.; et al. Genome-wide tandem repeat expansions contribute to schizophrenia risk. *Mol. Psychiatry* **2022**, *27*, 3692–3698. [CrossRef]

10. Wen, J.; Trost, B.; Engchuan, W.; Halvorsen, M.; Pallotto, L.M.; Mitina, A.; Ancalade, N.; Farrell, M.; Backstrom, I.; Guo, K.; et al. Rare tandem repeat expansions associate with genes involved in synaptic and neuronal signaling functions in schizophrenia. *Mol. Psychiatry* **2023**, *28*, 475–482. [CrossRef]

11. Mukamel, R.E.; Handsaker, R.E.; Sherman, M.A.; Barton, A.R.; Hujoel, M.L.A.; McCarroll, S.A.; Loh, P.R. Repeat polymorphisms underlie top genetic risk loci for glaucoma and colorectal cancer. *Cell* **2023**, *186*, 3659–3673.e3623. [CrossRef] [PubMed]

12. Erwin, G.S.; Gursoy, G.; Al-Abri, R.; Suriyaprakash, A.; Dolzhenko, E.; Zhu, K.; Hoerner, C.R.; White, S.M.; Ramirez, L.; Vadlakonda, A.; et al. Recurrent repeat expansions in human cancer genomes. *Nature* **2023**, *613*, 96–102. [CrossRef] [PubMed]

13. Fotsing, S.F.; Margoliash, J.; Wang, C.; Saini, S.; Yanicky, R.; Shleizer-Burko, S.; Goren, A.; Gymrek, M. The impact of short tandem repeat variation on gene expression. *Nat. Genet.* **2019**, *51*, 1652–1659. [CrossRef] [PubMed]

14. Bakhtiari, M.; Park, J.; Ding, Y.C.; Shleizer-Burko, S.; Neuhausen, S.L.; Halldorsson, B.V.; Stefansson, K.; Gymrek, M.; Bafna, V. Variable number tandem repeats mediate the expression of proximal genes. *Nat. Commun.* **2021**, *12*, 2075. [CrossRef]

15. Sun, J.H.; Zhou, L.; Emerson, D.J.; Phyo, S.A.; Titus, K.R.; Gong, W.; Gilgenast, T.G.; Beagan, J.A.; Davidson, B.L.; Tassone, F.; et al. Disease-Associated Short Tandem Repeats Co-localize with Chromatin Domain Boundaries. *Cell* **2018**, *175*, 224–238.e215. [CrossRef]

16. Jakubosky, D.; D'Antonio, M.; Bonder, M.J.; Smail, C.; Donovan, M.K.R.; Young Greenwald, W.W.; Matsui, H.; i2QTL Consortium; D'Antonio-Chronowska, A.; Stegle, O.; et al. Properties of structural variants and short tandem repeats associated with gene expression and complex traits. *Nat. Commun.* **2020**, *11*, 2927. [CrossRef]

17. Mukamel, R.E.; Handsaker, R.E.; Sherman, M.A.; Barton, A.R.; Zheng, Y.; McCarroll, S.A.; Loh, P.R. Protein-coding repeat polymorphisms strongly shape diverse human phenotypes. *Science* **2021**, *373*, 1499–1505. [CrossRef]

18. Ellegren, H. Heterogeneous mutation processes in human microsatellite DNA sequences. *Nat. Genet.* **2000**, *24*, 400–402. [CrossRef]

19. Porubsky, D.; Dashnow, H.; Sasani, T.A.; Logsdon, G.A.; Hallast, P.; Noyes, M.D.; Kronenberg, Z.N.; Mokveld, T.; Koundinya, N.; Nolan, C.; et al. A familial, telomere-to-telomere reference for human de novo mutation and recombination from a four-generation pedigree. *bioRxiv* **2024**. [CrossRef]

20. Moore, L.D.; Le, T.; Fan, G. DNA Methylation and Its Basic Function. *Neuropsychopharmacology* **2013**, *38*, 23–38. [CrossRef]

21. Fan, H.; Chu, J.-Y. A Brief Review of Short Tandem Repeat Mutation. *Genom. Proteom. Bioinform.* **2007**, *5*, 7–14. [CrossRef] [PubMed]

22. Weisburd, B.; Dolzhenko, E.; Bennett, M.F.; Danzi, M.C.; English, A.; Hiatt, L.; Tanudisastro, H.; Kurtas, N.E.; Jam, H.Z.; Brand, H.; et al. Defining a tandem repeat catalog and variation clusters for genome-wide analyses and population databases. *bioRxiv* **2024**. [CrossRef]

23. Tenchov, R.; Sasso, J.M.; Zhou, Q.A. Polyglutamine (PolyQ) Diseases: Navigating the Landscape of Neurodegeneration. *ACS Chem. Neurosci.* **2024**, *15*, 2665–2694. [CrossRef] [PubMed]

24. Depienne, C.; Mandel, J.-L. 30 years of repeat expansion disorders: What have we learned and what are the remaining challenges? *Am. J. Hum. Genet.* **2021**, *108*, 764–785. [CrossRef]

25. Kacher, R.; Lejeune, F.X.; David, I.; Boluda, S.; Coarelli, G.; Leclere-Turbant, S.; Heinzmann, A.; Marelli, C.; Charles, P.; Goizet, C.; et al. CAG repeat mosaicism is gene specific in spinocerebellar ataxias. *Am. J. Hum. Genet.* **2024**, *111*, 913–926. [CrossRef]

26. Hannan, A.J. Tandem repeats mediating genetic plasticity in health and disease. *Nat. Rev. Genet.* **2018**, *19*, 286–298. [CrossRef]

27. Hall, D.A.; Nag, S.; Ouyang, B.; Bennett, D.A.; Liu, Y.; Ali, A.; Zhou, L.; Berry-Kravis, E. Fragile X Gray Zone Alleles Are Associated With Signs of Parkinsonism and Earlier Death. *Mov. Disord.* **2020**, *35*, 1448–1456. [CrossRef]

28. Cabal-Herrera, A.M.; Tassanakijpanich, N.; Salcedo-Arellano, M.J.; Hagerman, R.J. Fragile X-Associated Tremor/Ataxia Syndrome (FXTAS): Pathophysiology and Clinical Implications. *Int. J. Mol. Sci.* **2020**, *21*, 4391. [CrossRef]

29. Allen, E.G.; Charen, K.; Hipp, H.S.; Shubeck, L.; Amin, A.; He, W.; Nolin, S.L.; Glicksman, A.; Tortora, N.; McKinnon, B.; et al. Refining the risk for fragile X-associated primary ovarian insufficiency (FXPOI) by FMR1 CGG repeat size. *Genet. Med.* **2021**, *23*, 1648–1655. [CrossRef]

30. Tassone, F.; Sellier, C.; Charlet-Berguerand, N.; Todd, P.K. The Molecular Biology of Premutation Expanded Alleles. In *FXTAS, FXPOI, and Other Premutation Disorders*; Springer: Cham, Switzerland, 2016; pp. 101–127. [CrossRef]

31. Hagerman, P. Fragile X-associated tremor/ataxia syndrome (FXTAS): Pathology and mechanisms. *Acta Neuropathol.* **2013**, *126*, 1–19. [CrossRef]

32. Nolin, S.L.; Brown, W.T.; Glicksman, A.; Houck, J.G.E.; Gargano, A.D.; Sullivan, A.; Biancalana, V.; Bröndum-Nielsen, K.; Hjalgrim, H.; Holinski-Feder, E.; et al. Expansion of the Fragile X CGG Repeat in Females with Premutation or Intermediate Alleles. *Am. J. Hum. Genet.* **2003**, *72*, 454–464. [CrossRef] [PubMed]

33. Chen, J.; Zhao, Y.; Zhou, X.; Xue, J.; Xiao, Q.; Pan, H.; Zhou, X.; Xiang, Y.; Li, J.; Zhu, L.; et al. Evaluation of the role of FMR1 CGG repeat allele in Parkinson's disease from the Chinese population. *Front. Aging Neurosci.* **2023**, *15*, 1234027. [CrossRef] [PubMed]

34. Hagerman, P.J.; Hagerman, R. Fragile X syndrome. *Curr. Biol.* **2021**, *31*, R273–R275. [CrossRef]

35. Pieretti, M.; Zhang, F.P.; Fu, Y.H.; Warren, S.T.; Oostra, B.A. Absence of Expression of the FMR-1 Gene in Fragile X Syndrome. *Cell* **1991**, *66*, 817–822. [CrossRef] [PubMed]

36. Yu, S.; Pritchard, M.; Kremer, E.; Lynch, M.; Nancarrow, J. Fragile X Genotype Characterized by an Unstable Region of DNA. *Science* **1991**, *252*, 1179–1181. [CrossRef]

37. Kaufmann, W.E.; Kidd, S.A.; Andrews, H.F.; Budimirovic, D.B.; Esler, A. Autism Spectrum Disorder in Fragile X Syndrome: Cooccurring Conditions and Current Treatment. *Pediatrics* **2017**, *139* (Suppl. S3), S194–S206. [CrossRef]

38. Sun, Z.; Fan, J.; Wang, Y. X-Chromosome Inactivation and Related Diseases. *Genet. Res.* **2022**, *2022*, 1391807. [CrossRef]

39. Hall, D.A.; Robertson-Dick, E.E.; O'Keefe, J.A.; Hadd, A.G.; Zhou, L.; Berry-Kravis, E. X-inactivation in the clinical phenotype of fragile X premutation carrier sisters. *Neurol. Genet.* **2016**, *2*, e45. [CrossRef]

40. McColgan, P.; Tabrizi, S.J. Huntington's disease: A clinical review. *Eur. J. Neurol.* **2018**, *25*, 24–34. [CrossRef]

41. Ghosh, R.; Tabrizi, S.J. Huntington disease. *Handb. Clin. Neurol.* **2018**, *147*, 255–278. [CrossRef]

42. Nopoulos, P.C.; Aylward, E.H.; Ross, C.A.; Mills, J.A.; Langbehn, D.R.; Johnson, H.J.; Magnotta, V.A.; Pierson, R.K.; Beglinger, L.J.; Nance, M.A.; et al. Smaller intracranial volume in prodromal Huntington's disease: Evidence for abnormal neurodevelopment. *Brain* **2011**, *134 Pt 1*, 137–142. [CrossRef] [PubMed]

43. Paulsen, J.S.; Magnotta, V.A.; Mikos, A.E.; Paulson, H.L.; Penziner, E.; Andreasen, N.C.; Nopoulos, P.C. Brain structure in preclinical Huntington's disease. *Biol. Psychiatry* **2006**, *59*, 57–63. [CrossRef] [PubMed]

44. Caviston, J.P.; Ross, J.L.; Antony, S.M.; Tokito, M.; Holzbaur, E.L.F. Huntingtin facilitates dynein/dynactin-mediated vesicle transport. *Proc. Natl. Acad. Sci. USA* **2007**, *104*, 10045–10050. [CrossRef] [PubMed]

45. Fusco, F.R.; Zuccato, C.; Tartari, M.; Martorana, A.; De March, Z.; Giampa, C.; Cattaneo, E.; Bernardi, G. Co-localization of brain-derived neurotrophic factor (BDNF) and wild-type huntingtin in normal and quinolinic acid-lesioned rat brain. *Eur. J. Neurosci.* **2003**, *18*, 1093–1102. [CrossRef]

46. Rigamonti, D.; Bauer, J.H.; De-Fraja, C.; Conti, L.; Sipione, S. Wild-Type Huntingtin Protects from Apoptosis Upstream of Caspase-3. *J. Neurosci.* **2000**, *20*, 3705–3713. [CrossRef]

47. Arrasate, M.; Finkbeiner, S. Protein aggregates in Huntington's disease. *Exp. Neurol.* **2012**, *238*, 1–11. [CrossRef]

48. Gutekunst, C.; Li, S.; Yi, H.; Mulroy, J.S.; Kuemmerle, S. Nuclear and Neuropil Aggregates in Huntington's Disease: Relationship to Neuropathology. *J. Neurosci.* **1999**, *19*, 2522–2534. [CrossRef]

49. Burrus, C.J.; McKinstry, S.U.; Kim, N.; Ozlu, M.I.; Santoki, A.V.; Fang, F.Y.; Ma, A.; Karadeniz, Y.B.; Worthington, A.K.; Dragatsis, I.; et al. Striatal Projection Neurons Require Huntingtin for Synaptic Connectivity and Survival. *Cell Rep.* **2020**, *30*, 642–657. [CrossRef]

50. Rubinsztein, D.C.; Leggo, J.; Coles, R.; Almqvist, E.; Biancalana, V. Phenotypic Characterization of Individuals with 30–40 CAG Repeats in the Huntington Disease (HD) Gene Reveals HD Cases with 36 Repeats and Apparently Normal Elderly Individuals with 36-39 Repeats. *Am. J. Hum. Genet.* **1996**, *59*, 16–22.

51. Walker, F.O. Huntington's disease. *Lancet* **2007**, *369*, 218–228. [CrossRef]

52. Langbehn, D.R.; Hayden, M.R.; Paulsen, J.S. and PREDICT-HD Investigators of the Huntington Study Group. CAG-repeat length and the age of onset in Huntington disease (HD): A review and validation study of statistical approaches. *Am. J. Med. Genet. B Neuropsychiatr. Genet.* **2010**, *153B*, 397–408. [CrossRef]

53. Ranganathan, M.; Kostyk, S.K.; Allain, D.C.; Race, J.A.; Daley, A.M. Age of onset and behavioral manifestations in Huntington's disease: An Enroll-HD cohort analysis. *Clin. Genet.* **2021**, *99*, 133–142. [CrossRef]

54. Powis, Z.; Lutz, J.; Liaquat, K.; Querubin, J.A.; Batish, S.D. Expanding the Phenotype of Extremely Early Onset Juvenile Huntington's Disease: A Case Report and Review of Previously Published Cases. *Am. J. Med. Genet. A* **2024**, *197*, e63894. [CrossRef] [PubMed]

55. Udd, B.; Krahe, R. The myotonic dystrophies: Molecular, clinical, and therapeutic challenges. *Lancet Neurol.* **2012**, *11*, 891–905. [CrossRef] [PubMed]

56. Prior, T.W.; on behalf of the American College of Medical Genetics (ACMG) Laboratory Quality Assurance Committee. Technical standards and guidelines for myotonic dystrophy type 1 testing. *Genet. Med.* **2009**, *11*, 552–555. [CrossRef]

57. Lanni, S.; Pearson, C.E. Molecular genetics of congenital myotonic dystrophy. *Neurobiol. Dis.* **2019**, *132*, 104533. [CrossRef]

58. Yanovsky-Dagan, S.; Cohen, E.; Megalli, P.; Altarescu, G.; Schonberger, O.; Eldar-Geva, T.; Epsztejn-Litman, S.; Eiges, R. DMPK hypermethylation in sperm cells of myotonic dystrophy type 1 patients. *Eur. J. Hum. Genet.* **2022**, *30*, 980–983. [CrossRef]

59. Rimoldi, M.; Lucchiari, S.; Pagliarani, S.; Meola, G.; Comi, G.P.; Abati, E. Myotonic dystrophies: An update on clinical features, molecular mechanisms, management, and gene therapy. *Neurol. Sci.* **2024**, 1–18. [CrossRef]

60. Kaliman, P.; Llagostera, E. Myotonic dystrophy protein kinase (DMPK) and its role in the pathogenesis of myotonic dystrophy 1. *Cell. Signal.* **2008**, *20*, 1935–1941. [CrossRef]

61. Lopez-Martinez, A.; Soblechero-Martin, P.; de-la-Puente-Ovejero, L.; Nogales-Gadea, G.; Arechavala-Gomeza, V. An Overview of Alternative Splicing Defects Implicated in Myotonic Dystrophy Type I. *Genes* **2020**, *11*, 1109. [CrossRef]

62. Dinca, D.M.; Lallemant, L.; Gonzalez-Barriga, A.; Cresto, N.; Braz, S.O.; Sicot, G.; Pillet, L.E.; Polveche, H.; Magneron, P.; Huguet-Lachon, A.; et al. Myotonic dystrophy RNA toxicity alters morphology, adhesion and migration of mouse and human astrocytes. *Nat. Commun.* **2022**, *13*, 3841. [CrossRef] [PubMed]

63. Sun, J.X.; Helgason, A.; Masson, G.; Ebenesersdottir, S.S.; Li, H.; Mallick, S.; Gnerre, S.; Patterson, N.; Kong, A.; Reich, D.; et al. A direct characterization of human mutation based on microsatellites. *Nat. Genet.* **2012**, *44*, 1161–1165. [CrossRef] [PubMed]

64. Chakraborty, R.; Kimmel, M.; Stivers, D.N.; Davison, L.J.; Deka, R. Relative mutation rates at di-, tri-, and tetranucleotide microsatellite loci. *Proc. Natl. Acad. Sci. USA* **1997**, *94*, 1041–1046. [CrossRef] [PubMed]

65. Glenn, T.C.; Stephan, W.; Dessauer, H.C.; Braun, M.J. Allelic Diversity in Alligator Microsatellite Loci Is Negatively Correlated with GC Content of Flanking Sequences and Evolutionary Conservation of PCR Amplifiability. *Mol. Biol. Evol.* **1996**, *13*, 1151–1154. [CrossRef]

66. Shimmin, L.C.; Chang, B.H.; Li, W.H. Male-driven evolution of DNA sequences. *Nature* **1993**, *362*, 745–747. [CrossRef]

67. Kayser, M.; Roewer, L.; Hedman, M.; Henke, L.; Henke, J.; Brauer, S.; Kruger, C.; Krawczak, M.; Nagy, M.; Dobosz, T.; et al. Characteristics and frequency of germline mutations at microsatellite loci from the human Y chromosome, as revealed by direct observation in father/son pairs. *Am. J. Hum. Genet.* **2000**, *66*, 1580–1588. [CrossRef]

68. Brinkmann, B.; Klintschar, M.; Neuhuber, F.; Huhne, J.; Rolf, B. Mutation Rate in Human Microsatellites: Influence of the Structure and Length of the Tandem Repeat. *Am. J. Hum. Genet.* **1998**, *62*, 1408–1415. [CrossRef]

69. Petruska, J.; Hartenstine, M.J.; Goodman, M.F. Analysis of strand slippage in DNA polymerase expansions of CAG/CTG triplet repeats associated with neurodegenerative disease. *J. Biol. Chem.* **1998**, *273*, 5204–5210. [CrossRef]

70. Jakupciak, J.P.; Wells, R.D. Genetic instabilities in (CTG.CAG) repeats occur by recombination. *J. Biol. Chem.* **1999**, *274*, 23468–23479. [CrossRef]

71. Lin, Y.; Dion, V.; Wilson, J.H. Transcription promotes contraction of CAG repeat tracts in human cells. *Nat. Struct. Mol. Biol.* **2006**, *13*, 179–180. [CrossRef]

72. Nakamori, M.; Pearson, C.E.; Thornton, C.A. Bidirectional transcription stimulates expansion and contraction of expanded (CTG)*(CAG) repeats. *Hum. Mol. Genet.* **2011**, *20*, 580–588. [CrossRef] [PubMed]

73. Steely, C.J.; Watkins, W.S.; Baird, L.; Jorde, L.B. The mutational dynamics of short tandem repeats in large, multigenerational families. *Genome Biol.* **2022**, *23*, 253. [CrossRef]

74. Kristmundsdottir, S.; Jonsson, H.; Hardarson, M.T.; Palsson, G.; Beyter, D.; Eggertsson, H.P.; Gylfason, A.; Sveinbjornsson, G.; Holley, G.; Stefansson, O.A.; et al. Sequence variants affecting the genome-wide rate of germline microsatellite mutations. *Nat. Commun.* **2023**, *14*, 3855. [CrossRef] [PubMed]

75. Mitra, I.; Huang, B.; Mousavi, N.; Ma, N.; Lamkin, M.; Yanicky, R.; Shleizer-Burko, S.; Lohmueller, K.E.; Gymrek, M. Patterns of de novo tandem repeat mutations and their role in autism. *Nature* **2021**, *589*, 246–250. [CrossRef] [PubMed]

76. Khristich, A.N.; Mirkin, S.M. On the wrong DNA track: Molecular mechanisms of repeat-mediated genome instability. *J. Biol. Chem.* **2020**, *295*, 4134–4170. [CrossRef]

77. Eichler, E.E.; Holden, J.J.; Popovich, B.W.; Reiss, A.L.; Snow, K. Length of uninterrupted CGG repeats determines instability in the FMR1 gene. *Nat. Genet.* **1994**, *8*, 88–94. [CrossRef]

78. Findlay Black, H.; Wright, G.E.B.; Collins, J.A.; Caron, N.; Kay, C.; Xia, Q.; Arning, L.; Bijlsma, E.K.; Squitieri, F.; Nguyen, H.P.; et al. Frequency of the loss of CAA interruption in the HTT CAG tract and implications for Huntington disease in the reduced penetrance range. *Genet. Med.* **2020**, *22*, 2108–2113. [CrossRef]

79. Cumming, S.A.; Hamilton, M.J.; Robb, Y.; Gregory, H.; McWilliam, C.; Cooper, A.; Adam, B.; McGhie, J.; Hamilton, G.; Herzyk, P.; et al. De novo repeat interruptions are associated with reduced somatic instability and mild or absent clinical features in myotonic dystrophy type 1. *Eur. J. Hum. Genet.* **2018**, *26*, 1635–1647. [CrossRef]

80. Miller, J.N.; van der Plas, E.; Hamilton, M.; Koscik, T.R.; Gutmann, L.; Cumming, S.A.; Monckton, D.G.; Nopoulos, P.C. Variant repeats within the DMPK CTG expansion protect function in myotonic dystrophy type 1. *Neurol. Genet.* **2020**, *6*, e504. [CrossRef]

81. Pesovic, J.; Peric, S.; Brkusanin, 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. [CrossRef]

82. Vinther-Jensen, T.; Ek, J.; Duno, M.; Skovby, F.; Hjermind, L.E.; Nielsen, J.E.; Nielsen, T.T. Germ-line CAG repeat instability causes extreme CAG repeat expansion with infantile-onset spinocerebellar ataxia type 2. *Eur. J. Hum. Genet.* **2013**, *21*, 626–629. [CrossRef] [PubMed]

83. Perez, B.A.; Shorrock, H.K.; Banez-Coronel, M.; Zu, T.; Romano, L.E.; Laboissonniere, L.A.; Reid, T.; Ikeda, Y.; Reddy, K.; Gomez, C.M.; et al. CCG*CGG interruptions in high-penetrance SCA8 families increase RAN translation and protein toxicity. *EMBO Mol. Med.* **2021**, *13*, e14095. [CrossRef] [PubMed]

84. Coffee, B.; Zhang, F.; Warren, S.T.; Reines, D. Acetylated histones are associated with FMR1 in normal but not fragile X-syndrome cells. *Nat. Genet.* **1999**, *22*, 209. [CrossRef]

85. Coffee, B.; Zhang, F.; Ceman, S.; Warren, S.T.; Reines, D. Histone modifications depict an aberrantly heterochromatinized FMR1 gene in fragile x syndrome. *Am. J. Hum. Genet.* **2002**, *71*, 923–932. [CrossRef] [PubMed]

86. Pietrobono, R.; Tabolacci, E.; Zalfa, F.; Zito, I.; Terracciano, A.; Moscato, U.; Bagni, C.; Oostra, B.; Chiurazzi, P.; Neri, G. Molecular dissection of the events leading to inactivation of the FMR1 gene. *Hum. Mol. Genet.* **2005**, *14*, 267–277. [CrossRef]

87. Malachowski, T.; Chandradoss, K.R.; Boya, R.; Zhou, L.; Cook, A.L.; Su, C.; Pham, K.; Haws, S.A.; Kim, J.H.; Ryu, H.S.; et al. Spatially coordinated heterochromatinization of long synaptic genes in fragile X syndrome. *Cell* **2023**, *186*, 5840–5858.e5836. [CrossRef]

88. Eiges, R.; Urbach, A.; Malcov, M.; Frumkin, T.; Schwartz, T.; Amit, A.; Yaron, Y.; Eden, A.; Yanuka, O.; Benvenisty, N.; et al. Developmental study of fragile X syndrome using human embryonic stem cells derived from preimplantation genetically diagnosed embryos. *Cell Stem Cell* **2007**, *1*, 568–577. [CrossRef]

89. Lee, H.G.; Imaichi, S.; Kraeutler, E.; Aguilar, R.; Lee, Y.W.; Sheridan, S.D.; Lee, J.T. Site-specific R-loops induce CGG repeat contraction and fragile X gene reactivation. *Cell* **2023**, *186*, 2593–2609.e2518. [CrossRef]

90. Morales, F.; Corrales, E.; Zhang, B.; Vasquez, M.; Santamaria-Ulloa, C.; Quesada, H.; Sirito, M.; Estecio, M.R.; Monckton, D.G.; Krahe, R. Myotonic dystrophy type 1 (DM1) clinical subtypes and CTCF site methylation status flanking the CTG expansion are mutant allele length-dependent. *Hum. Mol. Genet.* **2021**, *31*, 262–274. [CrossRef]

91. Koehorst, E.; Odria, R.; Capo, J.; Nunez-Manchon, J.; Arbex, A.; Almendrote, M.; Linares-Pardo, I.; Natera-de Benito, D.; Saez, V.; Nascimento, A.; et al. An Integrative Analysis of DNA Methylation Pattern in Myotonic Dystrophy Type 1 Samples Reveals a Distinct DNA Methylation Profile between Tissues and a Novel Muscle-Associated Epigenetic Dysregulation. *Biomedicines* **2022**, *10*, 1372. [CrossRef]

92. Kornberg, A.; Bertsch, L.L.; Jackson, J.F.; Khorana, H.G. Enzymatic Synthesis of Deoxyribonucleic Acid, XVI. Oligonucleotides As Templates and the Mechanism of Their Replication. *Proc. Natl. Acad. Sci. USA* **1964**, *51*, 315–323. [CrossRef] [PubMed]

93. Lia, A.; Seznec, H.; Hofmann-Radvanyi, H.; Radvanyi, F.; Duros, C. 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*, 1285–1291. [CrossRef] [PubMed]

94. Al-Mahdawi, S.; Pinto, R.M.; Ruddle, P.; Carroll, C.; Webster, Z.; Pook, M. GAA repeat instability in Friedreich ataxia YAC transgenic mice. *Genomics* **2004**, *84*, 301–310. [CrossRef] [PubMed]

95. Wheeler, V.C.; Lebel, L.A.; Vrbanac, V.; Teed, A.; te Riele, H.; MacDonald, M.E. Mismatch repair gene Msh2 modifies the timing of early disease in Hdh(Q111) striatum. *Hum. Mol. Genet.* **2003**, *12*, 273–281. [CrossRef]

96. Fortune, M.T.; Vassilopoulos, C.; Coolbaugh, M.I.; Siciliano, M.J.; Monckton, D.G. Dramatic, expansion-biased, age-dependent, tissue-specific somatic mosaicism in a transgenic mouse model of triplet repeat instability. *Hum. Mol. Genet.* **2000**, *9*, 439–445. [CrossRef]

97. Morales, F.; Vasquez, M.; Santamaria, C.; Cuenca, P.; Corrales, E.; Monckton, D.G. 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. [CrossRef]

98. Dragileva, E.; Hendricks, A.; Teed, A.; Gillis, T.; Lopez, E.T.; Friedberg, E.C.; Kucherlapati, R.; Edelmann, W.; Lunetta, K.L.; MacDonald, M.E.; et al. Intergenerational and striatal CAG repeat instability in Huntington's disease knock-in mice involve different DNA repair genes. *Neurobiol. Dis.* **2009**, *33*, 37–47. [CrossRef]

99. Uddin, M.D.M.; Nguyen, N.Q.H.; Yu, B.; Brody, J.A.; Pampana, A.; Nakao, T.; Fornage, M.; Bressler, J.; Sotoodehnia, N.; Weinstock, J.S.; et al. Clonal hematopoiesis of indeterminate potential, DNA methylation, and risk for coronary artery disease. *Nat. Commun.* **2022**, *13*, 5350. [CrossRef]

100. Marnell, C.S.; Bick, A.; Natarajan, P. Clonal hematopoiesis of indeterminate potential (CHIP): Linking somatic mutations, hematopoiesis, chronic inflammation and cardiovascular disease. *J. Mol. Cell Cardiol.* **2021**, *161*, 98–105. [CrossRef]

101. Maury, E.A.; Jones, A.; Seplyarskiy, V.; Nguyen, T.T.L.; Rosenbluh, C. Somatic mosaicism in schizophrenia brains revealsprenatal mutational processes. *Science* **2024**, *386*, 217–224. [CrossRef]

102. Risques, R.A.; Kennedy, S.R. Aging and the rise of somatic cancer-associated mutations in normal tissues. *PLoS Genet.* **2018**, *14*, e1007108. [CrossRef] [PubMed]

103. Potter, H.; Chial, H.J.; Caneus, J.; Elos, M.; Elder, N.; Borysov, S.; Granic, A. Chromosome Instability and Mosaic Aneuploidy in Neurodegenerative and Neurodevelopmental Disorders. *Front. Genet.* **2019**, *10*, 1092. [CrossRef]

104. Sherman, M.A.; Rodin, R.E.; Genovese, G.; Dias, C.; Barton, A.R.; Mukamel, R.E.; Berger, B.; Park, P.J.; Walsh, C.A.; Loh, P.R. Large mosaic copy number variations confer autism risk. *Nat. Neurosci.* **2021**, *24*, 197–203. [CrossRef] [PubMed]

105. Lee, W.P.; Tucci, A.A.; Conery, M.; Leung, Y.Y.; Kuzma, A.B.; Valladares, O.; Chou, Y.F.; Lu, W.; Wang, L.S.; Schellenberg, G.D.; et al. Copy Number Variation Identification on 3,800 Alzheimer's Disease Whole Genome Sequencing Data from the Alzheimer's Disease Sequencing Project. *Front. Genet.* **2021**, *12*, 752390. [CrossRef] [PubMed]

106. Huttenlocher, J.; Stefansson, H.; Steinberg, S.; Helgadottir, H.T.; Sveinbjornsdottir, S.; Riess, O.; Bauer, P.; Stefansson, K. Heterozygote carriers for CNVs in PARK2 are at increased risk of Parkinson's disease. *Hum. Mol. Genet.* **2015**, *24*, 5637–5643. [CrossRef] [PubMed]

107. La Cognata, V.; Morello, G.; D'Agata, V.; Cavallaro, S. Copy number variability in Parkinson's disease: Assembling the puzzle through a systems biology approach. *Hum. Genet.* **2017**, *136*, 13–37. [CrossRef]

108. Blokzijl, F.; De Ligt, J.; Jager, M.; Sasselli, V.; Roerink, S.; Sasaki, N.; Huch, M.; Boymans, S.; Kuijk, E.; Prins, P.; et al. Tissue-specific mutation accumulation in human adult stem cells during life. *Nature* **2016**, *538*, 260–264. [CrossRef]

109. Jiraanont, P.; Kumar, M.; Tang, H.-T.; Espinal, G.; Hagerman, P.J.; Hagerman, R.J.; Chutabhakdikul, N.; Tassone, F. Size and methylation mosaicism in males with Fragile X syndrome. *Expert. Rev. Mol. Diagn.* **2017**, *17*, 1023–1032. [CrossRef]

110. Pretto, D.; Yrigollen, C.M.; Tang, H.-T.; Williamson, J.; Espinal, G.; Iwahashi, C.K.; Durbin-Johnson, B.; Hagerman, R.J.; Hagerman, P.J.; Tassone, F. Clinical and molecular implications of mosaicism in FMR1 full mutations. *Front. Genet.* **2014**, *5*, 318. [CrossRef]

111. Schmucker, B.; Ballhausen, W.G.; Pfeiffer, R.A. Mosaicism of a microdeletion of 486 bp involving the CGG repeat of the FMR1 gene due to misalignment of GTT tandem repeats at chi-like elements flanking both breakpoints and a full mutation. *Hum. Genet.* **1996**, *98*, 409–414. [CrossRef]

112. Dyer-Friedman, J.; Glaser, B.; Hessl, D.; Johnston, C.; Huffman, L.C. Genetic and Environmental Influences on the Cognitive Outcomes of Children With Fragile X Syndrome. *J. Am. Acad. Child. Adolesc. Psychiatry* **2002**, *41*, 237–244. [CrossRef] [PubMed]

113. Loesch, D.Z.; Huggins, R.M.; Hagerman, R.J. Phenotypic variation and FMRP levels in fragile X. *Ment. Retard. Dev. Disabil. Res. Rev.* **2004**, *10*, 31–41. [CrossRef] [PubMed]

114. Protic, D.; Polli, R.; Hwang, Y.H.; Mendoza, G.; Hagerman, R.; Durbin-Johnson, B.; Hayward, B.E.; Usdin, K.; Murgia, A.; Tassone, F. Activation Ratio Correlates with IQ in Female Carriers of the FMR1 Premutation. *Cells* **2023**, *12*, 1711. [CrossRef]

115. Tabolacci, E.; Nobile, V.; Pucci, C.; Chiurazzi, P. Mechanisms of the FMR1 Repeat Instability: How Does the CGG Sequence Expand? *Int. J. Mol. Sci.* **2022**, *23*, 5425. [CrossRef]

116. Fernández, E.; Gennaro, E.; Pirozzi, F.; Baldo, C.; Forzano, F.; Turolla, L.; Faravelli, F.; Gastaldo, D.; Coviello, D.; Grasso, M.; et al. FXS-Like Phenotype in Two Unrelated Patients Carrying a Methylated Premutation of the FMR1 Gene. *Front. Genet.* **2018**, *9*, 442. [CrossRef]

117. Meng, L.; Kaufmann, W.E.; Frye, R.E.; Ong, K.; Kaminski, J.W.; Velinov, M.; Berry-Kravis, E. The association between mosaicism type and cognitive and behavioral functioning among males with fragile X syndrome. *Am. J. Med. Genet. A* **2022**, *188*, 858–866. [CrossRef]

118. Baker, E.K.; Arpone, M.; Vera, S.A.; Bretherton, L.; Ure, A.; Kraan, C.M.; Bui, M.; Ling, L.; Francis, D.; Hunter, M.F.; et al. Intellectual functioning and behavioural features associated with mosaicism in fragile X syndrome. *J. Neurodev. Disord.* **2019**, *11*, 41. [CrossRef]

119. Mailick, M.R.; Movaghar, A.; Hong, J.; Greenberg, J.S.; Dawalt, L.S.; Zhou, L.; Jackson, J.; Rathouz, P.J.; Baker, M.W.; Brilliant, M.; et al. Health Profiles of Mosaic Versus Non-mosaic FMR1 Premutation Carrier Mothers of Children with Fragile X Syndrome. *Front. Genet.* **2018**, *9*, 173. [CrossRef]

120. Telenius, H.; Kremer, B.; Goldberg, Y.P.; Theilmann, J.; Andrew, S.E. Somatic and gonadal mosaicism of the Huntington disease gene CAG repeat in brain and sperm. *Nat. Genet.* **1994**, *6*, 409–414. [CrossRef]

121. Swami, M.; Hendricks, A.E.; Gillis, T.; Massood, T.; Mysore, J.; Myers, R.H.; Wheeler, V.C. 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*, 3039–3047. [CrossRef]

122. Gonitel, R.; Moffitt, H.; Sathasivam, K.; Woodman, B.; Detloff, P.J. DNA instability in postmitotic neurons. *Proc. Natl. Acad. Sci. USA* **2008**, *105*, 3467–3472. [CrossRef] [PubMed]

123. Mirkin, S.M. Expandable DNA repeats and human disease. *Nature* **2007**, *447*, 932–940. [CrossRef] [PubMed]

124. Bettencourt, C.; Hensman-Moss, D.; Flower, M.; Wiethoff, S.; Brice, A.; Goizet, C.; Stevanin, G.; Koutsis, G.; Karadima, G.; Panas, M.; et al. DNA repair pathways underlie a common genetic mechanism modulating onset in polyglutamine diseases. *Ann. Neurol.* **2016**, *79*, 983–990. [CrossRef]

125. Wheeler, V.C.; Persichetti, F.; McNeil, S.M.; Mysore, J.S.; Mysore, S.S.; MacDonald, M.E.; Myers, R.H.; Gusella, J.F.; Wexler, N.S.; The US–Venezuela Collaborative Research Group. Factors associated with HD CAG repeat instability in Huntington disease. *J. Med. Genet.* **2007**, *44*, 695–701. [CrossRef]

126. Ciosi, M.; Maxwell, A.; Cumming, S.A.; Hensman Moss, D.J.; Alshammari, A.M.; Flower, M.D.; Durr, A.; Leavitt, B.R.; Roos, R.A.C.; The Track-HD Team; 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*, 568–580. [CrossRef]

127. Kacher, R.; Lejeune, F.X.; Noel, S.; Cazeneuve, C.; Brice, A.; Humbert, S.; Durr, A. Propensity for somatic expansion increases over the course of life in Huntington disease. *Elife* **2021**, *10*, e64674. [CrossRef]

128. Handsaker, R.E.; Kashin, S.; Reed, N.M.; Tan, S.; Lee, W.S.; McDonald, T.M.; Morris, K.; Kamitaki, N.; Mullally, C.D.; Morakabati, N.R.; et al. Long somatic DNA-repeat expansion drives neurodegeneration in Huntington's disease. *Cell* **2025**, *188*, 623–639.e19. [CrossRef]

129. Kaplan, S.; Itzkovitz, S.; Shapiro, E. A universal mechanism ties genotype to phenotype in trinucleotide diseases. *PLoS Comput. Biol.* **2007**, *3*, e235. [CrossRef]

130. Aldous, S.G.; Smith, E.J.; Landles, C.; Osborne, G.F.; Canibano-Pico, M.; Nita, I.M.; Phillips, J.; Zhang, Y.; Jin, B.; Hirst, M.B.; et al. A CAG repeat threshold for therapeutics targeting somatic instability in Huntington's disease. *Brain* **2024**, *147*, 1784–1798. [CrossRef]

131. Pressl, C.; Matlik, K.; Kus, L.; Darnell, P.; Luo, J.D.; Paul, M.R.; Weiss, A.R.; Liguore, W.; Carroll, T.S.; Davis, D.A.; et al. Selective vulnerability of layer 5a corticostriatal neurons in Huntington's disease. *Neuron* **2024**, *112*, 924–941.e910. [CrossRef]

132. Martorell, L.; Monckton, D.G.; Gamez, J.; Baiget, M. Complex patterns of male germline instability and somatic mosaicism in myotonic dystrophy type 1. *Eur. J. Hum. Genet.* **2000**, *8*, 423–430. [CrossRef] [PubMed]

133. Temmerman, N.D.; Sermon, K.; Seneca, S.; Rycke, M.D.; Hilven, P.; Lissens, W.; Steirteghem, A.V.; Liebaers, I. Intergenerational Instability of the Expanded CTG Repeat in the DMPK Gene: Studies in Human Gametes and Preimplantation Embryos. *Am. J. Hum. Genet.* **2004**, *75*, 325–329. [CrossRef] [PubMed]

134. Morales, F.; Vasquez, M.; Corrales, E.; Vindas-Smith, R.; Santamaria-Ulloa, C.; Zhang, B.; Sirito, M.; Estecio, M.R.; Krahe, R.; Monckton, D.G. Longitudinal increases in somatic mosaicism of the expanded CTG repeat in myotonic dystrophy type 1 are associated with variation in age-at-onset. *Hum. Mol. Genet.* **2020**, *29*, 2496–2507. [CrossRef]

135. Anvret, M.; Ahlberg, G.; Grandell, U.; Hedberg, B.; Johnson, K.; Edström, L. Larger expansions of the CTG repeat in muscle compared to lymphocytes from patients with myotonic dystrophy. *Hum. Mol. Genet.* **1993**, *2*, 1397–1400. [CrossRef] [PubMed]

136. Ashizawa, T.; Dubel, J.R.; Harati, Y. Somatic instability of CTG repeat in myotonic dystrophy. *Neurology* **1993**, *43*, 2674–2678. [CrossRef] [PubMed]

137. Thornton, C.A.; Johnson, K.; Moxley, R.T., 3rd. Myotonic dystrophy patients have larger CTG expansions in skeletal muscle than in leukocytes. *Ann. Neurol.* **1994**, *35*, 104–107. [CrossRef] [PubMed]

138. Kraus-Perrotta, C.; Lagalwar, S. Expansion, mosaicism and interruption: Mechanisms of the CAG repeat mutation in spinocerebellar ataxia type 1. *Cerebellum Ataxias* **2016**, *3*, 20. [CrossRef]

139. Hantash, F.M.; Goos, D.G.; Tsao, D.; Quan, F.; Buller-Burckle, A.; Peng, M.; Jarvis, M.; Sun, W.; Strom, C.M. Qualitative assessment of FMR1 (CGG)n triplet repeat status in normal, intermediate, premutation, full mutation, and mosaic carriers in both sexes: Implications for fragile X syndrome carrier and newborn screening. *Genet. Med.* **2010**, *12*, 162–173. [CrossRef]

140. Chintalaphani, S.R.; Pineda, S.S.; Deveson, I.W.; Kumar, K.R. An update on the neurological short tandem repeat expansion disorders and the emergence of long-read sequencing diagnostics. *Acta Neuropathol. Commun.* **2021**, *9*, 98. [CrossRef]

141. Dolzhenko, E.; Bennett, M.F.; Richmond, P.A.; Trost, B.; Chen, S.; van Vugt, J.; Nguyen, C.; Narzisi, G.; Gainullin, V.G.; Gross, A.M.; et al. ExpansionHunter Denovo: A computational method for locating known and novel repeat expansions in short-read sequencing data. *Genome Biol.* **2020**, *21*, 102. [CrossRef]

142. Willems, T.; Zielinski, D.; Yuan, J.; Gordon, A.; Gymrek, M.; Erlich, Y. Genome-wide profiling of heritable and de novo STR variations. *Nat. Methods* **2017**, *14*, 590–592. [CrossRef] [PubMed]

143. Dashnow, H.; Pedersen, B.S.; Hiatt, L.; Brown, J.; Beecroft, S.J.; Ravenscroft, G.; LaCroix, A.J.; Lamont, P.; Roxburgh, R.H.; Rodrigues, M.J.; et al. STRling: A k-mer counting approach that detects short tandem repeat expansions at known and novel loci. *Genome Biol.* **2022**, *23*, 257. [CrossRef] [PubMed]

144. Mousavi, N.; Shleizer-Burko, S.; Yanicky, R.; Gymrek, M. Profiling the genome-wide landscape of tandem repeat expansions. *Nucleic Acids Res.* **2019**, *47*, e90. [CrossRef] [PubMed]

145. Dolzhenko, E.; van Vugt, J.; Shaw, R.J.; Bekritsky, M.A.; van Blitterswijk, M.; Narzisi, G.; Ajay, S.S.; Rajan, V.; Lajoie, B.R.; Johnson, N.H.; et al. Detection of long repeat expansions from PCR-free whole-genome sequence data. *Genome Res.* **2017**, *27*, 1895–1903. [CrossRef]

146. Gymrek, M.; Golan, D.; Rosset, S.; Erlich, Y. lobSTR: A short tandem repeat profiler for personal genomes. *Genome Res.* **2012**, *22*, 1154–1162. [CrossRef]

147. Ibanez, K.; Polke, J.; Hagelstrom, R.T.; Dolzhenko, E.; Pasko, D.; Thomas, E.R.A.; Daugherty, L.C.; Kasperaviciute, D.; Smith, K.R.; WGS for Neurological Diseases Group; et al. Whole genome sequencing for the diagnosis of neurological repeat expansion disorders in the UK: A retrospective diagnostic accuracy and prospective clinical validation study. *Lancet Neurol.* **2022**, *21*, 234–245. [CrossRef]

148. Sehgal, A.; Jam, H.Z.; Shen, A.; Gymrek, M. Genome-wide detection of somatic mosaicism at short tandem repeats. *Bioinformatics* **2024**, *40*, btae485. [CrossRef]
149. Dolzhenko, E.; English, A.; Dashnow, H.; De Sena Brandine, G.; Mokveld, T.; Rowell, W.J.; Karniski, C.; Kronenberg, Z.; Danzi, M.C.; Cheung, W.A.; et al. Characterization and visualization of tandem repeats at genome scale. *Nat. Biotechnol.* **2024**, *42*, 1606–1614. [CrossRef]
150. Ziaei Jam, H.; Zook, J.M.; Javadzadeh, S.; Park, J.; Sehgal, A.; Gymrek, M. LongTR: Genome-wide profiling of genetic variation at tandem repeats from long reads. *Genome Biol.* **2024**, *25*, 176. [CrossRef]
151. Chiu, R.; Rajan-Babu, I.S.; Friedman, J.M.; Birol, I. Straglr: Discovering and genotyping tandem repeat expansions using whole genome long-read sequences. *Genome Biol.* **2021**, *22*, 224. [CrossRef]
152. Mokveld, T.; Dolzhenko, E.; Dashnow, H.; Nicholas, T.J.; Sasani, T.; van der Sanden, B.; Jadhav, B.; Pedersen, B.; Kronenberg, Z.; Tucci, A.; et al. TRGT-denovo: Accurate detection of de novo tandem repeat mutations. *bioRxiv* **2024**. [CrossRef]
153. Stevanoski, I.; Chintalaphani, S.R.; Gamaarachchi, H.; Ferguson, J.M.; Pineda, S.S. Comprehensive genetic diagnosis of tandem repeat expansion disorders with programmable targeted nanopore sequencing. *Sci. Adv.* **2022**, *8*, eabm5386. [CrossRef] [PubMed]
154. Tsai, Y.-C.; Greenberg, D.; Powell, J.; Höijer, I.; Ameur, A.; Strahl, M.; Ellis, E.; Jonasson, I.; Mouro Pinto, R.; Wheeler, V.C.; et al. Amplification-free, CRISPR-Cas9 Targeted Enrichment and SMRT Sequencing of Repeat-Expansion Disease Causative Genomic Regions. *BioRxiv* **2017**. [CrossRef]
155. Nakamura, W.; Hirata, M.; Oda, S.; Chiba, K.; Okada, A.; Mateos, R.N.; Sugawa, M.; Iida, N.; Ushiama, M.; Tanabe, N.; et al. Assessing the efficacy of target adaptive sampling long-read sequencing through hereditary cancer patient genomes. *NPJ Genom. Med.* **2024**, *9*, 11. [CrossRef]
156. Chen, Z.; Gustavsson, E.K.; Macpherson, H.; Anderson, C.; Clarkson, C.; Rocca, C.; Self, E.; Alvarez Jerez, P.; Scardamaglia, A.; Pellerin, D.; et al. Adaptive Long-Read Sequencing Reveals GGC Repeat Expansion in ZFHX3 Associated with Spinocerebellar Ataxia Type 4. *Mov. Disord.* **2024**, *39*, 486–497. [CrossRef]
157. Grosso, V.; Marcolungo, L.; Maestri, S.; Alfano, M.; Lavezzari, D.; Iadarola, B.; Salviati, A.; Mariotti, B.; Botta, A.; D'Apice, M.R.; et al. Characterization of FMR1 Repeat Expansion and Intragenic Variants by Indirect Sequence Capture. *Front. Genet.* **2021**, *12*, 743230. [CrossRef]
158. Lokanga, R.A.; Entezam, A.; Kumari, D.; Yudkin, D.; Qin, M.; Smith, C.B.; Usdin, K. Somatic expansion in mouse and human carriers of fragile X premutation alleles. *Hum. Mutat.* **2013**, *34*, 157–166. [CrossRef]
159. Long, A.; Napierala, J.S.; Polak, U.; Hauser, L.; Koeppen, A.H.; Lynch, D.R.; Napierala, M. Somatic instability of the expanded GAA repeats in Friedreich's ataxia. *PLoS ONE* **2017**, *12*, e0189990. [CrossRef]
160. Monaghan, K.G.; Lyon, E.; Spector, E.B. ACMG Standards and Guidelines for fragile X testing: A revision to the disease-specific supplements to the Standards and Guidelines for Clinical Genetics Laboratories of the American College of Medical Genetics and Genomics. *Genet. Med.* **2013**, *15*, 575–586. [CrossRef]
161. Quarrell, O.W.J.; Rigby, A.S.; Barron, L.; Crow, Y.; Dalton, A.; Dennis, N.; Fryer, A.E.; Heydon, F.; Kinning, E.; Lashwood, A.; et al. Reduced penetrance alleles for Huntington's disease: A multi-centre direct observational study. *J. Med. Genet.* **2006**, *44*, e68. [CrossRef]
162. Zhou, Y.; Kumari, D.; Sciascia, N.; Usdin, K. CGG-repeat dynamics and FMR1 gene silencing in fragile X syndrome stem cells and stem cell-derived neurons. *Mol. Autism* **2016**, *7*, 42. [CrossRef] [PubMed]
163. Hannan, A.J. Expanding horizons of tandem repeats in biology and medicine: Why 'genomic dark matter' matters. *Emerg. Top. Life Sci.* **2023**, *7*, 239–247. [CrossRef] [PubMed]