PORTLAND PRESS

Review Article

Unravelling the link between neurodevelopmental disorders and short tandem CGG-repeat expansions

Dale J. Annear and ⓘ **R. Frank Kooy**

Department of Medical Genetics, University of Antwerp, Antwerp, Belgium

Correspondence: R. Frank Kooy (frank.kooy@uantwerpen.be)

Neurodevelopmental disorders (NDDs) encompass a diverse group of disorders characterised by impaired cognitive abilities and developmental challenges. Short tandem repeats (STRs), repetitive DNA sequences found throughout the human genome, have emerged as potential contributors to NDDs. Specifically, the CGG trinucleotide repeat has been implicated in a wide range of NDDs, including Fragile X Syndrome (FXS), the most common inherited form of intellectual disability and autism. This review focuses on CGG STR expansions associated with NDDs and their impact on gene expression through repeat expansion-mediated epigenetic silencing. We explore the molecular mechanisms underlying CGG-repeat expansion and the resulting epigenetic modifications, such as DNA hypermethylation and gene silencing. Additionally, we discuss the involvement of other CGG STRs in neurodevelopmental diseases. Several examples, including *FMR1*, *AFF2*, *AFF3*, *XYLT1*, *FRA10AC1*, *CBL*, and *DIP2B*, highlight the complex relationship between CGG STR expansions and NDDs. Furthermore, recent advancements in this field are highlighted, shedding light on potential future research directions. Understanding the role of STRs, particularly CGG-repeats, in NDDs has the potential to uncover novel diagnostic and therapeutic strategies for these challenging disorders.

Neurodevelopmental disorders and short tandem repeats

In humans, neurodevelopmental disorders (NDDs) represent one of the most clinically and genetically diverse groups of disorders. These disorders arise from disruptions in early brain development and can encompass conditions such as autism spectrum disorder (ASD), intellectual disability (ID), motor impairments, seizures, learning difficulties, and attention issues. Symptoms of these disorders often encompass cognitive limitations, delays in both gross and fine motor skills, hypotonia, and difficulties in social and behavioural aspects. [1,2]. Short tandem repeats (STRs), previously known as microsatellites or simple sequence repeats, are short DNA motifs of one to six bp that are consecutively repeated in a head-to-tail-like fashion [1,2]. To date, multiple human diseases and disorders have been attributed to STRs. There appears to be a link between STRs and neurological disorders, as a large proportion of STR disorders are neuromuscular, neurodegenerative, or neurodevelopmental in nature. Currently, the primary repeat class that is associated with NDDs is the trinucleotide CGG STR. Several NDDs have been definitively associated with CGG STRs and several other NDD and neurodegenerative phenotypes have had the involvement of CGG STRs implicated (Table 1). The majority of these disorders involve some form of cognitive disorder including, but not limited to ID, developmental delay, and ASD. The typical mechanism linked to NDD-causative STRs is repeat expansion, which we will discuss further.

CGG STR expansion may manifest as a folate-sensitive fragile site (FSFS). These are cytogenetically visible breaks in the chromosomes that have been observed throughout the human genome, which are inducible by culturing cells in a folate-deficient medium [3]. To date, 22 FSFS have been identified. At the molecular level, FSFS are due to a CGG-repeat expansion. While for a long time, scientific interest

Table 1. Summary of the CGG STR expansions detected within the boundaries of the folate-sensitive fragile sites

Folate-sensitive fragile site	Chromosome [19]	Localised gene/s	Gene region	Disease phenotype	Reference
FRAXA	Xq27.3	*FMR1*	5′-UTR	Fragile X syndrome	[20]
FRAXE	Xq28	*AFF2*	5′-UTR	Fragile XE syndrome	[21]
FRAXF	Xq28	*TMEM185A*[1]	5′-UTR	Developmental delay	[22]
FRA1M	1p21.3	*ABCD3*[2]	5′-UTR	Candidate	[23,24]
FRA2A	2q11.2	*AFF3*	5′-UTR/intron	Intellectual disability	[25,26]
FRA2B	2q13	*BCL2L11*[2]	5′-UTR/intron	Candidate	[23,24]
		RGPD1[2]	5′-UTR/intron	Candidate	[24]
FRA2K	2q22.3	*ACVR2A*[2]	Promotor/5′-UTR	Candidate	[24]
FRA5G	5q35	*RASGEF1C*[2]	5′-UTR	Candidate	[6,7]
		FAM193B[2]	5′-UTR	Candidate	[24]
FRA6A	6p23	*RANBP9*[2]	Exon	Candidate	[23,24]
		GFOD1[2]	Promotor	Candidate	[6]
FRA7A	7p11.2	*ZNF713*[1]	Intron	Autism spectrum disorder	[27]
FRA8A	8q22.3	*FZD6*[2]	5′-UTR	Candidate	[23,24]
FRA9A	9p21	*C9orf72*[2]	Intron	Candidate	[23,24]
		ELAVL2[2]	5′-UTR	Candidate	[24]
FRA9B	9q32	*TMEM268*[2]	5′-UTR	Candidate	[6]
		ZNF618[2]	Promotor/5′-UTR	Candidate	[6]
FRA10A	10q23.3	*FRA10AC1*	5′-UTR	Neurodevelopmental disorder	[28–30]
FRA11A	11q13.3	*C11orf80*[1]	Exon	Candidate	[31]
FRA11B	11q23.3	*CBL*	5′-UTR	Jacobsen syndrome	[32]
FRA12A	12q13.1	*DIP2B*[1]	5′-UTR	Intellectual developmental disorder	[33]
FRA12D	12q24.13	*TPCN1*[2]	5′-UTR	Candidate	[6]
		HECTD4[2]	5′-UTR/exon	Candidate	[6]
FRA16A	16p13.11	*XYLT1*	5′-UTR	Baratela–Scott syndrome	[4]
FRA19B	19p13	*LINGO3*[2]	Promotor/5′-UTR	Candidate	[23,24]
		SAMD1[2]	5′-UTR/exon	Candidate	[6]
		GIPC1[2]	5′-UTR	Candidate	[6]
		BRD4[2]	Promotor/5′-UTR	Candidate	[6]
		CARM1[2]	Promotor/5′-UTR	Candidate	[6]
FRA20A	20p11.23	*RALGAPA2*[2]	5′-UTR	Candidate	[23,24]
FRA22A	22q13	*CSNK1E*[1]	5′-UTR	Autism spectrum disorder	[34]

[1]CGG STR expansions within the given gene are known to be causative of the associated fragile site and have been observed within individuals with a NDD phenotype, however, the association between the expansion and phenotype is not yet unequivocally linked;
[2]CGG STR expansions within this gene have been identified through either methylation epigenetic signal or bioinformatic repeat expansion detection analyses marking these loci/genes as candidates causative for NDDs and folate-sensitive fragile sites.

in STR CGG expansions has been almost exclusively limited to the fragile X syndrome (FXS), recently, interest in this class of STRs has exploded for a number of reasons. First, several FSFS due to CGG-repeat expansions that have long been considered harmless polymorphisms have now been unequivocally linked to NDD [4,5]. Second, the bioinformatic discovery that CGG-repeats are much more common than anticipated [6] and third the outcome of an epigenetic survey that discovered additional disease-relevant CGG Expansions [7]. The purpose of this review is to highlight and connect these recent developments in this rapidly emerging field.

CGG-repeat expansion-mediated epigenetic silencing

CGG STRs mutations (at least from the perspective of the *FMR1* gene) are typically divided into three different mutation classes. These are 'full mutation' >200 repeat units [8,9], 'premutation' 200–56 repeat units [10,11],

and 'gray zone' 55–41 repeat units [12,13]. Currently, full mutation repeat expansions appear to be the principal mutation mechanism linked to NDDs in humans. While the premutation and gray zone classes have been linked to human disease phenotypes, they are primarily associated with neurodegenerative disorders, such as Fragile X-associated tremor/ataxia syndrome and Neuronal intranuclear inclusion disease [14,15]. Furthermore, the pathogenic mechanisms at play differ considerably between the full mutation and the premutation and gray zone class mutations. Within this discussion, we shall focus on the full mutation class.

When a CGG STR locus expands beyond its full mutation length, for instance, 200 CGG units in the case of FXS, this triggers an epigenetic methylation event. Here, the CpG islands within and immediately surrounding the repeat tract all become methylated [16]. This hypermethylation of the repeat locus then causes the hyper-condensation of the DNA strand and the surrounding chromatin [17]. If the hypermethylated repeat in question is localised within a gene or gene promotor region (which is often the case for CGG STRs), the hyper-condensation will block access to the binding sites required by cellular transcription machinery in order to transcribe the gene in question [17]. Therefore, the gene is silenced as no mRNA is produced and consequently, no gene product can be realised. This produces an effect similar to gene loss-of-function or a gene deletion as no protein will be produced from that allele [18]. As stated above, a further consequence of CGG STR expansion may be the manifestation of a cytogenetically visible FSFS that is induced by culturing cells in a folate-deficient medium [3]. All 22 FSFS that have been described are listed in Table 1. The STRs that have been associated or linked with disease due to repeat full mutations are discussed in the following section.

FRAXA — FMR1

By far, the most well-known and comprehensively studied CGG STR is found within the 5′-untranslated region (UTR) of the Fragile X Messenger Ribonucleoprotein 1 (*FMR1*) gene on the X chromosome. Full mutation expansions of the *FMR1* CGG-repeat results in FXS. Not only was FXS the first defined CGG-repeat expansion disorder, but it is also the most common inherited form of ID and syndromic ASD [8]. FXS occurs in 1 in 4000 males and 1 in 8000 females, with males generally displaying a more severe phenotype. Females may display less severe phenotypes due to X-inactivation [35,36]. As females present with two X chromosomes, during embryonic development, in each cell, most genes on one of the X chromosomes are randomly silenced [36,37]. Therefore, in many cases, the expansion-containing allele would be silenced regardless and the healthy allele would be available for the transcription of *FMR1* [38]. To date, most research on the CGG STR is framed within the context of the *FMR1* repeat. Typically, the normal population length of the *FMR1* repeat is ∼30 CGG units but ranges from ∼6 to 40 repeat units. FXS is primarily characterised by mild to moderate ID, ASD, and particular facial appearance. Other common symptoms of FXS include joint laxity, macroorchidism (in males), hyperactivity, seizures, social and language impairments, and increased anxiety [38,39]. The highest levels of *FMR1* expression have been observed within neurons and the foetal brain in the early stages of proliferation and migration of nervous system cells [40,41]. Broadly, *FMR1* plays a role in development, mRNA binding, and translation regulation [42,43]. It seems that as *FMR1* is a crucial gene in the development of the central nervous system (CNS) that its absence during development results in FXS.

FRAXE — AFF2

Fragile XE (FRAXE) syndrome is a rare genetic condition of X-linked ID. The syndrome was initially identified in individuals who had been referred for FXS testing [21]. It is named for the FRAXE fragile site observed within the patients occurring on chromosome Xq28 distal to the FRAXA site. FRAXE syndrome is categorised by mild ID, however the phenotype can be highly variable between individuals and symptoms can include learning difficulties, language delay, hyperactivity, and a shortened attention span [44]. The cause of this disorder is a full-mutation CGG STR in the 5′-UTR of the ALF Transcription Elongation Factor 2 (*AFF2*) gene (formerly known as *FMR2*) which results in transcriptional silencing of the gene [45,46]. Though the exact molecular consequence of *AFF2* silencing and why FRAXE syndrome arises is unknown, it is suggested that *AFF2* is an RNA-binding protein that regulates alternative splicing through interaction with G quartet RNA structures and silencing of the gene may disrupt crucial gene regulation during development [44,47].

FRA2A — AFF3

ALF Transcription Elongation Factor 3 (*AFF3*) is a gene, like *AFF2*, which is part of the AF4/FMR2 gene family. Interestingly, *AFF3* contains two CGG-repeats that are capable of expanding to large sizes. There is a

CGG-repeat present within the 5′ promotor region and another repeat present in intron 2 of *AFF3*, which is a brain-active alternative promoter [6,25]. Expansion of the brain-specific promotor CGG STR beyond 200 units results in hypermethylation of the repeat and the adjacent CpG islands which accounts for gene silencing [25]. However, while the FRA2A fragile site was associated with neurodevelopmental phenotypes such as delay of motor and language skill acquisition, it was not possible to unequivocally link this expansion to neurodevelopmental delay as yet [25]. Recently, an enrichment of the *AFF3* STR expansion was observed in 6371 probands with neurodevelopmental problems [26]. cSNP analysis indicated monoallelic expression of the *AFF3* gene within FRA2A carriers, which may indicate that *AFF3* is a functionally haploinsufficient gene [27]. Haploinsufficiency of *AFF3* within the developing brain may cause an NDD phenotype. Two other independent studies by Shimizu et al. [48] and Voisin et al. [49] have indicated heterozygous *de novo* missense or deletion mutations within the *AFF3* gene within individuals displaying KINSSHIP (horseshoe kidney, Nievergelt/ Savarirayan type of mesomelic dysplasia, seizures, hypertrichosis, ID, and pulmonary involvement) syndrome. This lends further credence to the possibility that the heterozygous silencing of the *AFF3* gene through the FRA2A repeat expansion would result in an NDD or ID phenotype. A potential reason behind the phenotypic differentiation between KINSSHIP syndrome patients versus FRA2A patients may be due to the localisation of the STR expansion in the alternate brain-specific transcript of the *AFF3* gene.

FRA7A — ZNF713

Zinc Finger Protein 713 (*ZNF713*) is a gene found on chromosome 7p11.2 in relative proximity to the centromere and has a significantly enriched expression within the brain in comparison with the other body tissues [27]. It is a member of the Kruppel-type zinc finger protein family, and it contains a Kruppel-associated box (KRAB) regulatory domain [27]. While very little is known about the *ZNF713* gene, it is known to contain two large and polymorphic CGG-repeats and has been implicated in ASD. One within the 5′-UTR and another within intron 1 [6]. Using FISH analysis, the 5′repeat was identified as the causative repeat for the FRA7A FSFS [27]. While a definitive link was not established, the full mutation and subsequent silencing of the FRA7A repeat has been associated with ASD within two unrelated families [27]. Interestingly, the relatives (of ASD-affected family members) who harboured a premutation-sized CGG-repeat allele displayed increased expression of the *ZNF713* gene [27], which is consistent with what is known and previously discussed of premutation-length CGG-repeat alleles.

FRA10A — FRA10AC1

The FRA10A-associated CGG-Repeat 1 (*FRA10AC1*) gene is known to contain the CGG STR responsible for the FRA10A fragile site. The original study that identified the repeat expansion in *FRA10AC1* observed that in their tested population the repeat length was variable, however, all observed individuals that exhibited a FRA10A fragile site were heterozygous, displaying a normal-sized repeat and an expanded repeat of at least 200 CGG unit [28]. Furthermore, it was observed that in the allele containing the expanded repeat was hypermethylated and *FRA10AC1* was not transcribed [30], providing further evidence for the full-mutation mechanism of the CGG STR. Due to the lack of a disease phenotype in these individuals combined with another study where a FRA10A carrier lacked any discernible phenotype, it was concluded that (at least in the heterozygous state) the CGG-repeat full-mutation in *FRA10AC1* was not pathogenic [28,50]. This would indicate that *FRA10AC1* is not a haploinsufficient gene. However, recently it was determined by von Elsner et al. [5] that bi-allelic loss-of-function mutations of FRA10AC1 result in a NDD with growth disruption, dysmorphic features, and corpus callosum abnormalities. This was further supported by several other studies that identified homozygous loss-of-function mutations within *FRA10AC1* in NDD patients [29,30]. However, as of yet, an NDD patient exhibiting a loss-of-function mutation along with a full-mutation *FRA10AC1* repeat has not been reported. Much like the *XYLT1* repeat (to be discussed), if such a case were to arise, we would expect a similar phenotype as reported by von Elsner and colleagues to manifest.

FRA11B — CBL

The FRA11B CGG-repeat is located on chromosome 11q23.3 immediately upstream of the CBL Protooncogene (*CBL*) gene [51]. The FRA11B CGG-repeat has been implicated in Jacobsen Syndrome [52]. This is an interesting case as the proposed manner in which the full-mutation contributes to Jacobsen syndrome is unique in comparison with the other cases of CGG full-mutations. Jacobsen syndrome is a contiguous gene syndrome caused by partial deletion of the q arm of chromosome 11 [52]. The symptoms of

Jacobsen syndrome can be quite varied as the breakpoint of the deletion can vary along the q arm of chromosome 11. However, typical characteristics include particular facial features, Paris–Trousseau syndrome (a bleeding disorder), delayed motor and speech development, and cognitive impairment [53]. In a subset of Jacobsen patients, the breakpoint of the 11q arm deletion is localised at the FRA11B fragile site or within proximity to it [32]. In the *FMR1* gene, It is thought that secondary structures produced by the expanded CGG-repeat cause replication stress during DNA replication, which causes stalling of the replication fork and results in a double-stranded break (DSB) [54,55]. Then through the initiation of break-induced repair, mutations are realised downstream of the CGG-repeat [56]. It is suggested that through a similar mechanism of replication stress, a DSB is induced at the FRA11B site which then results in the loss of the chromosome 11q arm downstream of the CGG-repeat [57]. While a distinct correlation has been established between the *CBL* CCG-repeat expansion and Jacobsen syndrome [57], it is unclear if this mechanism of CGG-repeat mediated chromosomal breakage is replicated elsewhere. While deletions proximal to the *FMR1* repeat have been reported [58,59], large deletions with breakpoints at the other FSFSs have not. Heterozygous mutations in the *CBL* gene have been implicated in Noonan Syndrome, Noonan Syndrome-like Disorder, and leukaemia [60,61], however, expression of FRA11B has not.

FRA12A — DIP2B

Giraud et al. [62] first identified the FRA12A fragile site occurring on chromosome 12q13 within a male patient presenting with ID and congenital complications. It was determined that this fragile site is caused by a CGG-repeat expansion (full-mutation) within the disco-interacting protein 2 homologue B (*DIP2B*) gene [33]. Since several studies have identified the FRA12A site co-occurring with ID [33,63–65], the disorder has been termed Intellectual developmental disorder, FRA12A type. However, the association of the FRA12A fragile site with ID can be controversial, as the expression of this fragile site has been observed in healthy individuals. However, there may be a dosage effect relating to the effect of the *DIP2B* expansion. As in patients, the presence of the FRA12A site was observed in 43.7% of lymphocytes, whereas in unaffected carriers of FRA12A, observation was far lower at 16.6% of lymphocytes [23]. Furthermore, in individuals without ID but with the presence of FRA12A, *DIP2B* expression was reduced to approximately two-thirds of the levels observed in controls. However, in individuals where both FRA12A and ID were present, *DIP2B* expression was reduced to half that of controls [33]. *DIP2B* has been determined to be a regulator of axonal development and is essential for the development of neuron cells [66]. Therefore, if *DIP2B* were to be transcriptionally silenced in a large proportion of cells during early neuronal development this may be a contributing mechanism towards neurocognitive disorders.

FRA16A — XYLT1

Baratela–Scott syndrome is an interesting case. The syndrome is categorised by short stature, dysmorphic facial features, skeletal dysplasia, and developmental delay [67]. In the past, Baratela–Scott syndrome was thought to be caused by pathogenic variants or total and partial deletion of the *XYLT1* gene, which encodes for the XT1 xylosyltransferase enzyme, which catalyses the primary step in chondroitin sulfate and dermatan sulfate proteoglycan biosynthesis [4]. It was noticed in several Baratela–Scott syndrome patients that a pathogenic variant or deletion could not be detected in the *XYLT1* gene by standard sequencing techniques [4]. In these patients, a CGG-repeat expansion was detected, which was undergoing hypermethylation and silencing expression of the *XYLT1*, the mechanism observed in FXS [4]. The genomic position of this CGG STR correlates with the previously reported FSFS FRA16A [4,68]. Interestingly, what was demonstrated by LaCroix et al. [4], was that the *XYLT1* gene is not haploinsufficient, and for a patient to manifest Baratela–Scott Syndrome they must contain either a pathogenic variant, a deletion, or a repeat expansion on both alleles of the *XYLT1* gene. In this case, the repeat-associated disorder is autosomal recessive, whereas the majority of known repeat-associated disorders are autosomal dominant [4]. This case coupled with the difficulty of detecting repeat expansions in patients through standard sequencing means illustrates how repeat expansions could be an overlooked disease-causing mechanism and likely contributes towards the missing heritability problem. Furthermore, it demonstrates how STR expansions and gene silencing through CGG-hypermethylation can replicate disorders that are typically thought to be caused by either pathogenic variants or deletion events.

FRA22A — CSNK1E

The casein kinase 1 epsilon (*CSNK1E*) gene is a known member of the circadian rhythm control genes and has been previously implicated in neuropsychiatric disorders such as bipolar disorder, schizophrenia, and opiate addiction [69–71]. Through an epigenetic analysis of over 23 000 individuals, hypermethylation signatures in *CSNK1E* were detected and it was determined that *CSNK1E* contains the CGG STR that is responsible for the FRA22A fragile site through CGG STR expansion [7]. Recently, a heterozygous *de novo* variant in *CSNK1E* (p. Glu117Lys) has been reported in patients presenting with hypotonia, speech apraxia, and ASD [34]. As *CSNK1E* is expressed in all human brain regions during natal development and is co-expressed with multiple neurological disorders candidate genes this positions the FRA22A CGG STR as another strong candidate for NDD contributing repeat [34].

The discovery of CGG-STRs beyond fragile sites

In addition to the CGG-repeats that were discovered as FSFS Recent studies have identified ~6000 CGG STR loci throughout the human genome [6,24]. Interestingly, it was observed that hundreds of these CGG STR loci displayed characteristics, such as native repeat length, genetic context, and polymorphism and mutation rates, in line with those observed within the CGG STR loci that have been confirmed to be disease-causing. [6]. An Independent cohort study that aimed at identifying epivariations in the human genome identified 25 loci where rare hypermethylation coincided with the presence of an unstable CGG tandem repeat [7]. In many cases, the presence of CGG expansions at these loci was confirmed.

Taking the bioinformatic approach and the epivariation study together, we now are able to present candidate repeat expansions as the molecular defects underlying most of the known FSFSs in the genome as included in Table 1. In some cases, multiple CGG STR expansion candidates were identified within the breakpoints of the same FSFS. These may represent different CGG-repeat expansions beyond the resolution of discrimination by cytogenetic analysis. For instance, while FRAXA, FRAXE, and FRAXF are now known to be distinct repeat expansion loci, they cannot be discriminated cytogenetically, and in the past, this has led to misidentification. The identification of multiple repeat candidates within a single FSFS may be suggestive of a similar phenomenon as observed on chromosome X.

Many of these repeat-containing genes present as additional candidate genes for NDD. Altogether, the comparison of the genomic location of these genes with known disease genes illustrates that there is an overwhelming enrichment of CGG STRs within genes in which mutations are known to be causative or linked to NDDs (Figure 1). Overall, of thousands of gene-localised CGG STRs, Figure 1 highlights the genes that contain many of the largest and most polymorphic CGG STRs within the human genome. These were detected through the bioinformatic genotyping of a large cohort of whole genome samples [24].

Additionally, through the comparison of CGG STR mutation rate and mutation size in ASD-affected individuals, compared with their unaffected siblings, multiple CGG STR-containing genome regions were determined to display higher rates and degrees of STR mutation with the ASD-affected individuals [24]. Specifically in ASD, multiple other studies have established a link between the incidence of rare, large repeat expansions in ASD patients when compared with controls [72,73]. Further linking the mutation of STRs to NDDs.

Beyond neurodevelopment disorders, CGG STRs have been implicated in other forms of congenital and developmental disorders, mostly due to the expansion of exon-based repeats encoding for polyalanine tract [74–77]. In these cases, the pathogenic repeat length breakpoints appear far lower than that seen in repeat expansion-mediated epigenetic silencing, with lengths of 20 to 30 repeat units being observed. Furthermore, what is surprising is that it appears that the trinucleotide repeats are over-represented within developmental and congenital disorders [78]. While other STR classes have been associated with neuromuscular and neurodegenerative disorders, such as the GAA, CAG and CCGGGG STRs, few other classes have been definitively associated with NDDs beyond the CGG class. STRs are fascinating in the sense that though they are grouped as a class of genetic features, however depending on their nucleotide composition and the context of their position within the genome, they can exhibit wildly distinct pathogenic mechanisms. Not only that, but the pathogenic mechanism at play can vary greatly within the same class of STR depending on repeat length, the CGG STR is the classical example of this.

The CGG STR remains the primary class of STR linked to NDDs, and as each year goes by the evidence for this mounts. Beyond the known disease-causing CGG STRs, hundreds of CGG loci localised with

Figure 1. Genome distribution of all known CGG short tandem repeats localised within genes known to contribute or be associated with NDDs.

Genomic positioning (GRCh38) and HGNC names of NDD-linked genes can be observed lining the outside of the ideogram. Track A displays the largest repeat length recorded at that given locus, while Track B demonstrates the rates of heterozygosity of the given repeat locus. The chromosomes and their displayed data are separated by colour for differentiation. The data to generate this figure was taken from Annear et al. [24] where large-scale CGG STR genotyping of NDD patients and their family members was performed.

NDD-related genes have now been identified, with many of these repeat loci reflecting the characteristics exhibited by the disease-causative repeats. Furthermore, with several studies having now identified higher rates and degrees of STR mutation within individuals presenting an NDD, such as ASD, this further solidifies the link between neurological function and repeat expansion [24,72,79]. As we now have many STR loci as strong novel candidates which may contribute towards NDDs, it may only be a matter of time before these diagnoses are matched with patient phenotypes. These efforts are surely bolstered through the development of sequencing technologies, such as the long-read sequencing platforms, and the increased

availability of sophisticated DNA analysis pipelines. With the rise of these technologies, we may potentially detect new classes of STRs that contribute to NDDs through mechanisms similar to or differing from the CGG STRs.

Summary

- CGG trinucleotide STRs are the primary repeat class that is associated with NDDs.

- The principal mechanism behind CGG STR-associated NDDs is repeat expansion-mediated epigenetic silencing.

- Repeat expansion-mediated epigenetic silencing is where a given repeat expands to or beyond the full mutation threshold (>200 repeat units for the FXS CGG STR) and the CpG islands within and flanking the repeat become hypermethylated, consequently preventing gene transcription.

- 22 folate-sensitive fragile sites have been reported. These are cytogenetically visible breaks on the metaphase chromosomes, caused by STR expansions. Approximately half have had the causative CGG STR identified. Several of these CGG STRs are known to be causative of NDDs.

- Recently, hundreds of new CGG STRs have been categorised, which reflect the characteristics of known NDD-causing repeats. Therefore, this raises the question of to what degree additional CGG STRs contribute towards NDDs.

Competing Interests

The authors declare that there are no competing interests associated with the manuscript.

Abbreviations

AFF2, ALF Transcription Elongation Factor 2; ASD, autism spectrum disorder; CSNK1E, casein kinase 1 epsilon; DIP2B, disco-interacting protein 2 homologue B; DSB, double-stranded break; FMR1, Fragile X Messenger Ribonucleoprotein 1; FRA10AC1, FRA10A-associated CGG-Repeat 1; FRAXE, Fragile XE; FSFS, folate-sensitive fragile site; FXS, Fragile X syndrome; ID, intellectual disability; NDDs, neurodevelopmental disorders; STR, Short tandem repeat; UTR, untranslated region; ZNF713, Zinc Finger Protein 713.

References

1 Jeffreys, A.J., Wilson, V. and Thein, S.L. (1985) Hypervariable "minisatellite" regions in human DNA. *Nature* **314**, 67–73 https://doi.org/10.1038/314067a0

2 Pena, S.D.J., Chakraborty, R., Epplen, J.T. and Jeffreys, A.J. (1993) *DNA Fingerprinting: State of the Science*, Birkhäuser Basel, Basel

3 Sutherland, G.R. (1977) Fragile sites on human chromosomes: demonstration of their dependence on the type of tissue culture medium. *Science* **197**, 265–266 https://doi.org/10.1126/science.877551

4 LaCroix, A.J., Stabley, D., Sahraoui, R., Adam, M.P., Mehaffey, M., Kernan, K. et al. (2019) GGC repeat expansion and exon 1 methylation of XYLT1 is a common pathogenic variant in baratela-Scott syndrome. *Am. J. Hum. Genet.* **104**, 35–44 https://doi.org/10.1016/j.ajhg.2018.11.005

5 von Elsner, L., Chai, G., Schneeberger, P.E., Harms, F.L., Casar, C., Qi, M. et al. (2022) Biallelic FRA10AC1 variants cause a neurodevelopmental disorder with growth retardation. *Brain* **145**, 1551–1563 https://doi.org/10.1093/brain/awab403

6 Annear, D.J., Vandeweyer, G., Elinck, E., Sanchis-Juan, A., French, C.E., Raymond, L. et al. (2021) Abundancy of polymorphic CGG repeats in the human genome suggest a broad involvement in neurological disease. *Sci. Rep.* **11**, 2515 https://doi.org/10.1038/s41598-021-82050-5

7 Garg, P., Jadhav, B., Rodriguez, O.L., Patel, N., Martin-Trujillo, A., Jain, M. et al. (2020) A survey of rare epigenetic variation in 23,116 human genomes identifies disease-relevant epivariations and CGG expansions. *Am. J. Hum. Genet.* **107**, 654–669 https://doi.org/10.1016/j.ajhg.2020.08.019

8 Verkerk, A.J., Pieretti, M., Sutcliffe, J.S., Fu, Y.H., Kuhl, D.P., Pizzuti, A. et al. (1991) Identification of a gene (FMR-1) containing a CGG repeat coincident with a breakpoint cluster region exhibiting length variation in fragile X syndrome. *Cell* **65**, 905–914 https://doi.org/10.1016/0092-8674(91)90397-H

9   Yu, S., Pritchard, M., Kremer, E., Lynch, M., Nancarrow, J., Baker, E. et al. (1991) Fragile X genotype characterized by an unstable region of DNA. *Science* **252**, 1179–1181 https://doi.org/10.1126/science.252.5009.1179

10 Oberlé, I., Rousseau, F., Heitz, D., Kretz, C., Devys, D., Hanauer, A. et al. (1991) Instability of a 550-base pair DNA segment and abnormal methylation in fragile X syndrome. *Science* **252**, 1097–1102 https://doi.org/10.1126/science.252.5009.1097

11 Fu, Y.H., Kuhl, D.P., Pizzuti, A., Pieretti, M., Sutcliffe, J.S., Richards, S. et al. (1991) Variation of the CGG repeat at the fragile X site results in genetic instability: resolution of the Sherman paradox. *Cell* **67**, 1047–1058 https://doi.org/10.1016/0092-8674(91)90283-5

12 Nolin, S.L., Brown, W.T., Glicksman, A., Houck, G.E., Gargano, A.D., Sullivan, A. et al. (2003) Expansion of the fragile X CGG repeat in females with premutation or intermediate alleles. *Am. J. Hum. Genet.* **72**, 454–464 https://doi.org/10.1086/367713

13 Sherman, S., Pletcher, B.A. and Driscoll, D.A. (2005) Fragile X syndrome: diagnostic and carrier testing. *Genet. Med.* **7**, 584–587 https://doi.org/10.1097/01.GIM.0000182468.22666.dd

14 Tassone, F., Iwahashi, C. and Hagerman, P.J. (2004) FMR1 RNA within the intranuclear inclusions of fragile X-associated tremor/ataxia syndrome (FXTAS). *RNA Biol.* **1**, 103–105 https://doi.org/10.4161/rna.1.2.1035

15 Ishiura, H., Shibata, S., Yoshimura, J., Suzuki, Y., Qu, W., Doi, K. et al. (2019) Noncoding CGG repeat expansions in neuronal intranuclear inclusion disease, oculopharyngodistal myopathy and an overlapping disease. *Nat. Genet.* **51**, 1222–1232 https://doi.org/10.1038/s41588-019-0458-z

16 Sandberg, G. and Schalling, M. (1997) Effect of in vitro promoter methylation and CGG repeat expansion on FMR-1 expression. *Nucleic Acids Res.* **25**, 2883–2887 https://doi.org/10.1093/nar/25.14.2883

17 Brouwer, J.R., Willemsen, R. and Oostra, B.A. (2009) Microsatellite repeat instability and neurological disease. *Bioessays* **31**, 71–83 https://doi.org/10.1002/bies.080122

18 Hagerman, R.J., Berry-Kravis, E., Hazlett, H.C., Bailey, D.B., Moine, H., Kooy, R.F. et al. (2017) Fragile X syndrome. *Nat. Rev. Dis. Primers* **3**, 17065 https://doi.org/10.1038/nrdp.2017.65

19 Debacker, K. and Kooy, R.F. (2007) Fragile sites and human disease. *Hum. Mol. Genet.* **16**, R150–R158 https://doi.org/10.1093/hmg/ddm136

20 Kremer, E.J., Pritchard, M., Lynch, M., Yu, S., Holman, K., Baker, E. et al. (1991) Mapping of DNA instability at the fragile X to a trinucleotide repeat sequence p(CCG)n. *Science* **252**, 1711–1714 https://doi.org/10.1126/science.1675488

21 Knight, S.J., Flannery, A.V., Hirst, M.C., Campbell, L., Christodoulou, Z., Phelps, S.R. et al. (1993) Trinucleotide repeat amplification and hypermethylation of a CpG island in FRAXE mental retardation. *Cell* **74**, 127–134 https://doi.org/10.1016/0092-8674(93)90300-F

22 Parrish, J.E., Oostra, B.A., Verkerk, A.J., Richards, C.S., Reynolds, J., Spikes, A.S. et al. (1994) Isolation of a GCC repeat showing expansion in FRAXF, a fragile site distal to FRAXA and FRAXE. *Nat. Genet.* **8**, 229–235 https://doi.org/10.1038/ng1194-229

23 Garg, P., Joshi, R.S., Watson, C. and Sharp, A.J. (2018) A survey of inter-individual variation in DNA methylation identifies environmentally responsive co-regulated networks of epigenetic variation in the human genome. *PLoS Genet.* **14**, e1007707 https://doi.org/10.1371/journal.pgen.1007707

24 Annear, D.J., Vandeweyer, G., Sanchis-Juan, A., Raymond, F.L. and Kooy, R.F. (2022) Non-Mendelian inheritance patterns and extreme deviation rates of CGG repeats in autism. *Genome Res.* **32**, 1967–1980 https://doi.org/10.1101/gr.277011.122

25 Metsu, S., Rooms, L., Rainger, J., Taylor, M.S., Bengani, H., Wilson, D.I. et al. (2014) FRA2A is a CGG repeat expansion associated with silencing of AFF3. *PLoS Genet.* **10**, e1004242 https://doi.org/10.1371/journal.pgen.1004242

26 Jadhav, B., Garg, P., van Vugt, J.J.F.A., Garikano, K.I., Gagliardi, D., Lee, W. et al. (2023) A GCC repeat expansion in AFF3 is a significant cause of intellectual disability. *medRxiv* https://doi.org/10.1101/2023.05.03.23289461

27 Metsu, S., Rainger, J.K., Debacker, K., Bernhard, B., Rooms, L., Grafodatskaya, D. et al. (2014) A CGG-repeat expansion mutation in ZNF713 causes FRA7A: association with autistic spectrum disorder in two families. *Hum. Mutat.* **35**, 1295–1300 https://doi.org/10.1002/humu.22683

28 Sarafidou, T., Kahl, C., Martinez-Garay, I., Mangelsdorf, M., Gesk, S., Baker, E. et al. (2004) Folate-sensitive fragile site FRA10A is due to an expansion of a CGG repeat in a novel gene, FRA10AC1, encoding a nuclear protein. *Genomics* **84**, 69–81 https://doi.org/10.1016/j.ygeno.2003.12.017

29 Alsaleh, N., Alhashem, A., Tabarki, B., Mohamed, S., Alharby, E., Alkuraya, F.S. et al. (2022) A biallelic variant in FRA10AC1 is associated with neurodevelopmental disorder and growth retardation. *Neurol. Genet.* **8**, e200010 https://doi.org/10.1212/NXG.0000000000200010

30 Banka, S., Shalev, S., Park, S.-M., Wood, K.A., Thomas, H.B., Wright, H.L. et al. (2022) Bi-allelic FRA10AC1 variants in a multisystem human syndrome. *Brain* **145**, e86–e89 https://doi.org/10.1093/brain/awac262

31 Debacker, K., Winnepenninckx, B., Longman, C., Colgan, J., Tolmie, J., Murray, R. et al. (2007) The molecular basis of the folate-sensitive fragile site FRA11A at 11q13. *Cytogenet. Genome Res.* **119**, 9–14 https://doi.org/10.1159/000109612

32 Jones, C., Slijepcevic, P., Marsh, S., Baker, E., Langdon, W.Y., Richards, R.I. et al. (1994) Physical linkage of the fragile site FRA11B and a Jacobsen syndrome chromosome deletion breakpoint in 11q23.3. *Hum. Mol. Genet.* **3**, 2123–2130 https://doi.org/10.1093/hmg/3.12.2123

33 Winnepenninckx, B., Debacker, K., Ramsay, J., Smeets, D., Smits, A., FitzPatrick, D.R. et al. (2007) CGG-repeat expansion in the DIP2B gene is associated with the fragile site FRA12A on chromosome 12q13.1. *Am. J. Hum. Genet.* **80**, 221–231 https://doi.org/10.1086/510800

34 Garg, R., Badduke, C. and Slavotinek, A. (2022) Ep128: a novel CSNK1E variant in a patient with autism spectrum disorder. *Genet. Med.* **24**, S80 https://doi.org/10.1016/j.gim.2022.01.164

35 Tassone, F., Hagerman, R.J., Iklé, D.N., Dyer, P.N., Lampe, M., Willemsen, R. et al. (1999) FMRP expression as a potential prognostic indicator in fragile X syndrome. *Am. J. Med. Genet.* **84**, 250–261 https://doi.org/10.1002/(SICI)1096-8628(19990528)84:3<250::AID-AJMG17>3.0.CO;2-4

36 Loesch, D.Z., Huggins, R.M. and Hagerman, R.J. (2004) Phenotypic variation and FMRP levels in fragile X. *Ment. Retard. Dev. Disabil. Res. Rev.* **10**, 31–41 https://doi.org/10.1002/mrdd.20006

37 Del Hoyo Soriano, L., Thurman, A.J., Harvey, D.J., Ted Brown, W. and Abbeduto, L. (2018) Genetic and maternal predictors of cognitive and behavioral trajectories in females with fragile X syndrome. *J. Neurodev. Disord.* **10**, 22 https://doi.org/10.1186/s11689-018-9240-2

38 Razak, K.A., Dominick, K.C. and Erickson, C.A. (2020) Developmental studies in fragile X syndrome. *J. Neurodev. Disord.* **12**, 13 https://doi.org/10.1186/s11689-020-09310-9

39 Lozano, R., Azarang, A., Wilaisakditipakorn, T. and Hagerman, R.J. (2016) Fragile X syndrome: a review of clinical management. *Intractable Rare Dis. Res.* **5**, 145–157 https://doi.org/10.5582/irdr.2016.01048

40 Devys, D., Lutz, Y., Rouyer, N., Bellocq, J.P. and Mandel, J.L. (1993) The FMR-1 protein is cytoplasmic, most abundant in neurons and appears normal in carriers of a fragile X premutation. *Nat. Genet.* **4**, 335–340 https://doi.org/10.1038/ng0893-335

41   Abitbol, M., Menini, C., Delezoide, A.L., Rhyner, T., Vekemans, M. and Mallet, J. (1993) Nucleus basalis magnocellularis and hippocampus are the major sites of FMR-1 expression in the human fetal brain. *Nat. Genet.* **4**, 147–153 https://doi.org/10.1038/ng0693-147

42   D'Hulst, C. and Kooy, R.F. (2009) Fragile X syndrome: from molecular genetics to therapy. *J. Med. Genet.* **46**, 577–584 https://doi.org/10.1136/jmg.2008.064667

43   Maurin, T., Zongaro, S. and Bardoni, B. (2014) Fragile X syndrome: from molecular pathology to therapy. *Neurosci. Biobehav. Rev.* **46**, 242–255 https://doi.org/10.1016/j.neubiorev.2014.01.006

44   Bensaid, M., Melko, M., Bechara, E.G. Davidovic, L., Berretta, A., Catania, M.V. et al. (2009) FRAXE-associated mental retardation protein (FMR2) is an RNA-binding protein with high affinity for G-quartet RNA forming structure. *Nucleic Acids Res.* **37**, 1269–1279 https://doi.org/10.1093/nar/gkn1058

45   Gu, Y., Shen, Y., Gibbs, R.A. and Nelson, D.L. (1996) Identification of FMR2, a novel gene associated with the FRAXE CCG repeat and CpG island. *Nat. Genet.* **13**, 109–113 https://doi.org/10.1038/ng0596-109

46   Gecz, J., Gedeon, A.K., Sutherland, G.R. and Mulley, J.C. (1996) Identification of the gene FMR2, associated with FRAXE mental retardation. *Nat. Genet.* **13**, 105–108 https://doi.org/10.1038/ng0596-105

47   Melko, M., Douguet, D., Bensaid, M., Zongaro, S., Verheggen, C., Gecz, J. et al. (2011) Functional characterization of the AFF (AF4/FMR2) family of RNA-binding proteins: insights into the molecular pathology of FRAXE intellectual disability. *Hum. Mol. Genet.* **20**, 1873–1885 https://doi.org/10.1093/hmg/ddr069

48   Shimizu, D., Sakamoto, R., Yamoto, K., Saitsu, H., Fukami, M., Nishimura, G. et al. (2019) *De novo* AFF3 variant in a patient with mesomelic dysplasia with foot malformation. *J. Hum. Genet.* **64**, 1041–1044 https://doi.org/10.1038/s10038-019-0650-0

49   Voisin, N., Schnur, R.E., Douzgou, S., Hiatt, S.M., Rustad, C.F., Brown, N.J. et al. (2021) Variants in the degron of AFF3 are associated with intellectual disability, mesomelic dysplasia, horseshoe kidney, and epileptic encephalopathy. *Am. J. Hum. Genet.* **108**, 857–873 https://doi.org/10.1016/j.ajhg.2021.04.001

50   De Leon-Luis, J., Santolaya-Forgas, J., May, G., Tonk, V., Shelton, D. and Galan, I. (2005) Prenatal diagnosis of FRA10A: a case report and literature review. *Am. J. Med. Genet. A* **136**, 63–65 https://doi.org/10.1002/ajmg.a.30093

51   Jones, C., Penny, L., Mattina, T., Yu, S., Baker, E., Voullaire, L. et al. (1995) Association of a chromosome deletion syndrome with a fragile site within the proto-oncogene CBL2. *Nature* **376**, 145–149 https://doi.org/10.1038/376145a0

52   Mattina, T., Perrotta, C.S. and Grossfeld, P. (2009) Jacobsen syndrome. *Orphanet. J. Rare Dis.* **4**, 9 https://doi.org/10.1186/1750-1172-4-9

53   Jacobsen, P., Hauge, M., Henningsen, K., Hobolth, N., Mikkelsen, M. and Philip, J. (1973) An (11;21) translocation in four generations with chromosome 11 abnormalities in the offspring. A clinical, cytogenetical, and gene marker study. *Hum. Hered.* **23**, 568–585 https://doi.org/10.1159/000152624

54   Gerhardt, J., Tomishima, M.J., Zaninovic, N., Colak, D., Yan, Z., Zhan, Q. et al. (2014) The DNA replication program is altered at the FMR1 locus in fragile X embryonic stem cells. *Mol. Cell* **53**, 19–31 https://doi.org/10.1016/j.molcel.2013.10.029

55   Chakraborty, A., Jenjaroenpun, P., Li, J., El Hilali, S., McCulley, A., Haarer, B. et al. (2020) Replication stress induces global chromosome breakage in the fragile X genome. *Cell Rep.* **32**, 108179 https://doi.org/10.1016/j.celrep.2020.108179

56   Kononenko, A.V., Ebersole, T., Vasquez, K.M. and Mirkin, S.M. (2018) Mechanisms of genetic instability caused by (CGG)n repeats in an experimental mammalian system. *Nat. Struct. Mol. Biol.* **25**, 669–676 https://doi.org/10.1038/s41594-018-0094-9

57   Jones, C., Müllenbach, R., Grossfeld, P., Auer, R., Favier, R., Chien, K. et al. (2000) Co-localisation of CCG repeats and chromosome deletion breakpoints in Jacobsen syndrome: evidence for a common mechanism of chromosome breakage. *Hum. Mol. Genet.* **9**, 1201–1208 https://doi.org/10.1093/hmg/9.8.1201

58   Meijer, H., de Graaff, E., Merckx, D., Jongbloed, R., de Die-Smulders, C.D., Engelen, J. et al. (1994) A deletion of 1.6 kb proximal to the CGG repeat of the FMR1 gene causes the clinical phenotype of the fragile X syndrome. *Hum. Mol. Genet.* **3**, 615–620 https://doi.org/10.1093/hmg/3.4.615

59   Grønskov, K., Hjalgrim, H., Bjerager, M.O. and Brøndum-Nielsen, K. (1997) Deletion of all CGG repeats plus flanking sequences in FMR1 does not abolish gene expression. *Am. J. Hum. Genet.* **61**, 961–967 https://doi.org/10.1086/514872

60   Loh, M.L., Sakai, D.S., Flotho, C., Kang, M., Fliegauf, M., Archambeault, S. et al. (2009) Mutations in CBL occur frequently in juvenile myelomonocytic leukemia. *Blood* **114**, 1859–1863 https://doi.org/10.1182/blood-2009-01-198416

61   Niemeyer, C.M., Kang, M.W., Shin, D.H., Furlan, I., Erlacher, M., Bunin, N.J. et al. (2010) Germline CBL mutations cause developmental abnormalities and predispose to juvenile myelomonocytic leukemia. *Nat. Genet.* **42**, 794–800 https://doi.org/10.1038/ng.641

62   Giraud, F., Ayme, S., Mattei, J.F. and Mattei, M.G. (1976) Constitutional chromosomal breakage. *Hum. Genet.* **34**, 125–136 https://doi.org/10.1007/BF00278880

63   Sutherland, G.R. and Baker, E. (2003) Forgotten fragile sites and related phenomena. *Cytogenet. Genome Res.* **100**, 89–91 https://doi.org/10.1159/000072842

64   Smeets, D.F., Scheres, J.M. and Hustinx, T.W. (1985) Heritable fragility at 11q13 and 12q13. *Clin. Genet.* **28**, 145–150 https://doi.org/10.1111/j.1399-0004.1985.tb00374.x

65   Berg, J., Grace, E., Teik, K.W., Hammond, H., Tidman, M. and FitzPatrick, D. (2000) Bullous ichthyosiform erythroderma, developmental delay, aortic and pulmonary stenosis in association with a FRA12A. *Clin. Dysmorphol.* **9**, 213–219 https://doi.org/10.1097/00019605-200009030-00012

66   Xing, Z.-K., Zhang, L.-Q., Zhang, Y., Sun, X., Sun, X.-L., Yu, H.-L. et al. (2020) DIP2B interacts With α-tubulin to regulate axon outgrowth. *Front. Cell Neurosci.* **14**, 29 https://doi.org/10.3389/fncel.2020.00029

67   Baratela, W.A.R., Bober, M.B., Tiller, G.E., Okenfuss, E., Ditro, C., Duker, A. et al. (2012) A newly recognized syndrome with characteristic facial features, skeletal dysplasia, and developmental delay. *Am. J. Med. Genet. A* **158A**, 1815–1822 https://doi.org/10.1002/ajmg.a.35445

68   Nancarrow, J.K., Kremer, E., Holman, K., Eyre, H., Doggett, N.A., Le Paslier, D. et al. (1994) Implications of FRA16A structure for the mechanism of chromosomal fragile site genesis. *Science* **264**, 1938–1941 https://doi.org/10.1126/science.8009225

69   Lee, K.Y., Ahn, Y.M., Kim, S.H., Kang, H.-G. and Joo, E.-J. (2018) Genetic association study of CSNK1E gene in bipolar disorder and circadian characteristics. *Nord. J. Psychiatry* **72**, 599–604 https://doi.org/10.1080/08039488.2018.1509125

70   Lowrey, P.L., Shimomura, K., Antoch, M.P., Yamazaki, S., Zemenides, P.D., Ralph, M.R. et al. (2000) Positional syntenic cloning and functional characterization of the mammalian circadian mutation tau. *Science* **288**, 483–492 https://doi.org/10.1126/science.288.5465.483

71   Bryant, C.D., Parker, C.C., Zhou, L., Olker, C., Chandrasekaran, R.Y., Wager, T.T. et al. (2012) Csnk1e is a genetic regulator of sensitivity to psychostimulants and opioids. *Neuropsychopharmacology* **37**, 1026–1035 https://doi.org/10.1038/npp.2011.287

72   Trost, B., Engchuan, W., Nguyen, C.M., Thiruvahindrapuram, B., Dolzhenko, E., Backstrom, I. et al. (2020) Genome-wide detection of tandem DNA repeats that are expanded in autism. *Nature* **586**, 80–86 https://doi.org/10.1038/s41586-020-2579-z

73   Mitra, I., Huang, B., Mousavi, N., Ma, N., Lamkin, M., Yanicky, R. et al. (2021) Patterns of *de novo* tandem repeat mutations and their role in autism. *Nature* **589**, 246–250 https://doi.org/10.1038/s41586-020-03078-7

74   Nallathambi, J., Moumné, L., De Baere, E., Beysen, D., Usha, K., Sundaresan, P. et al. (2007) A novel polyalanine expansion in FOXL2: the first evidence for a recessive form of the blepharophimosis syndrome (BPES) associated with ovarian dysfunction. *Hum. Genet.* **121**, 107–112 https://doi.org/10.1007/s00439-006-0276-0

75   Amiel, J., Laudier, B., Attié-Bitach, T., Trang, H., de Pontual, L., Gener, B. et al. (2003) Polyalanine expansion and frameshift mutations of the paired-like homeobox gene PHOX2B in congenital central hypoventilation syndrome. *Nat. Genet.* **33**, 459–461 https://doi.org/10.1038/ng1130

76   Goodman, F.R., Bacchelli, C., Brady, A.F., Brueton, L.A., Fryns, J.P., Mortlock, D.P. et al. (2000) Novel HOXA13 mutations and the phenotypic spectrum of hand-foot-genital syndrome. *Am. J. Hum. Genet.* **67**, 197–202 https://doi.org/10.1086/302961

77   Brown, L.Y., Odent, S., David, V., Blayau, M., Dubourg, C., Apacik, C. et al. (2001) Holoprosencephaly due to mutations in ZIC2: alanine tract expansion mutations may be caused by parental somatic recombination. *Hum. Mol. Genet.* **10**, 791–796 https://doi.org/10.1093/hmg/10.8.791

78   Chintalaphani, S.R., Pineda, S.S., Deveson, I.W. and Kumar, K.R. (2021) An update on the neurological short tandem repeat expansion disorders and the emergence of long-read sequencing diagnostics. *Acta Neuropathol. Commun.* **9**, 98 https://doi.org/10.1186/s40478-021-01201-x

79   Wen, J., Trost, B., Engchuan, W., Halvorsen, M., Pallotto, L.M., Mitina, A. et al. (2023) Rare tandem repeat expansions associate with genes involved in synaptic and neuronal signaling functions in schizophrenia. *Mol. Psychiatry* **28**, 475–482 https://doi.org/10.1038/s41380-022-01857-4