Christopher Schröder, Bernhard Horsthemke, and Christel Depienne*

GC-rich repeat expansions: associated disorders and mechanisms

https://doi.org/10.1515/medgen-2021-2099
Received April 29, 2021; accepted November 12, 2021

Abstract: Noncoding repeat expansions are a well-known cause of genetic disorders mainly affecting the central nervous system. Missed by most standard technologies used in routine diagnosis, pathogenic noncoding repeat expansions have to be searched for using specific techniques such as repeat-primed PCR or specific bioinformatics tools applied to genome data, such as ExpansionHunter. In this review, we focus on GC-rich repeat expansions, which represent at least one third of all noncoding repeat expansions described so far. GC-rich expansions are mainly located in regulatory regions (promoter, 5′ untranslated region, first intron) of genes and can lead to either a toxic gain-of-function mediated by RNA toxicity and/or repeat-associated non-AUG (RAN) translation, or a loss-of-function of the associated gene, depending on their size and their methylation status. We herein review the clinical and molecular characteristics of disorders associated with these difficult-to-detect expansions.

Keywords: tandem repeat, repeat expansion, GC-rich repeats, regulatory regions, DNA methylation, histone modifications, RNA toxicity, RNA foci, nuclear inclusion, aggregation, RAN translation, RNA structure, monogenic disorders, long-read sequencing

The human genome is particularly enriched in repetitions of adjacent nucleotide motifs, called tandem repeats [1, 2]. This dynamic class of variation has the highest mutational rate and is consequently highly polymorphic within human populations. The instability of tandem repeats increases in a length-dependent manner and their expansion across generations is a well-known process resulting in at least 50 human monogenic disorders [3, 4, 5]. Among those, at least one third are GC-rich. Here, we aim to review the disorders and mechanisms associated with GC-rich repeat expansions, focusing mainly on well-established monogenic conditions.

***Corresponding author: Christel Depienne,** Institute of Human Genetics, University Hospital Essen, University of Duisburg-Essen, Essen, Germany, e-mail: christel.depienne@uni-due.de
Christopher Schröder, Bernhard Horsthemke, Institute of Human Genetics, University Hospital Essen, University of Duisburg-Essen, Essen, Germany

Fragile X syndrome and associated disorders caused by CGG expansions in *FMR1*

The first noncoding GC-rich expansion disorder, described in 1991, was Fragile X syndrome (FXS, MIM #300624) [6, 7]. FXS is one of the most frequent causes of intellectual disability (ID) and/or autism spectrum disorder (ASD) in males and is caused by CGG repeat expansions exceeding 200 repeats (full expansion) in the 5′ untranslated region (UTR) of the *FMR1* (FMRP translational regulator 1) gene (MIM #309550) on chromosome X. Above this threshold, CpGs contained within the CGG repeats are usually methylated and associated with an absence of *FMR1* expression [8, 9] (Figure 1A–B). Although point mutations leading to a loss-of-function of FMRP, the protein encoded by *FMR1*, are very rare, they can also cause FXS, confirming that loss-of-function of *FMR1* is the pathophysiological mechanism associated with full expansion [10]. Females with full *FMR1* expansion can also be affected depending on the X inactivation status of the mutated allele in the brain, but they usually present with milder symptoms compared to male individuals [11, 12].

Remarkably, CGG repeat expansions in *FMR1* ranging from 55 to 200 repeats (premutations) were later found to be associated with two other disorders: Fragile X-associated premature ovarian insufficiency (FXPOI, also known as Premature Ovarian Failure 1 [POF1], MIM #311360) in females [13] and Fragile X-associated tremor ataxia syndrome (FXTAS, MIM #300623) in males [14]. FXTAS is a neurodegenerative disorder mainly affecting males over 50 years of age. Ovarian insufficiency (FXPOI) occurs in 20–25 % of female *FMR1* premutation carriers and consists in absent or irregular cycles, lower fertility or infertility, and premature ovarian failure (i. e., complete cessation of menstrual periods before age 40). Contrary to full expansions, premutations are not associated with hypermethylation and do not prevent *FMR1* transcription and FMRP expression [15]. Noncoding expansions in *FMR1* have hence become a paradigm, illustrating how expansions in a single gene may have different downstream impacts and cause different disorders depending on their size [16].

Two major mechanisms have been proposed to explain FXTAS/FXPOI pathogenesis. One is a gain of func-

Open Access. © 2021 Schröder et al., published by De Gruyter. This work is licensed under the Creative Commons Attribution 4.0 International License.

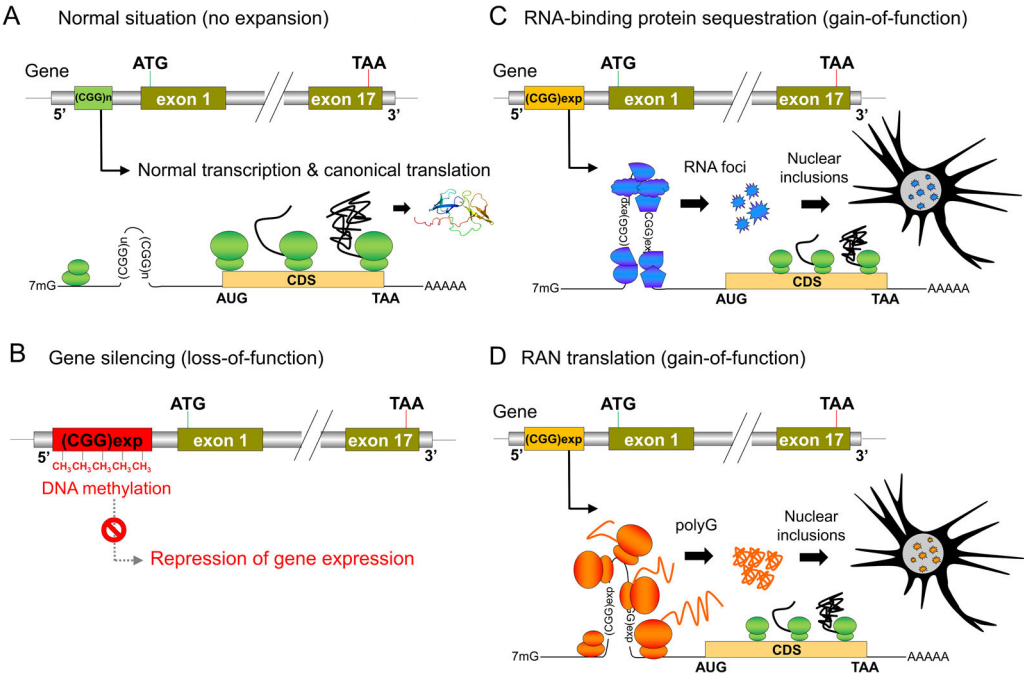

Figure 1: Main pathogenic mechanisms associated with GC-rich repeat expansions. (A) Nonpathogenic situation (e. g., less than 50 GCC repeats in *FMR1*) associated with normal transcription and canonical translation. (B) Epigenetic gene silencing. Full-length GC-rich expansions in gene promoters and/or 5′ untranslated regions (e. g., >200 CGG repeats in 5′UTR of *FMR1* causing Fragile X syndrome) are associated with DNA methylation at CpG sites. Expanded methylated alleles are locked in a chromatin configuration preventing gene transcription and protein expression. (C) Sequestration of RNA-binding splicing factors. Intermediate CGG expansions (55 to 200 repeats) causing Fragile X-associated tremor ataxia syndrome (FXTAS) can form stable RNA secondary structures able to bind specific RNA-binding proteins with high affinity. These RNA molecules accumulate to form inclusions in the nucleus and sequester bound RNA-binding proteins. (D) Repeat-associated non-AUG (RAN) translation is a noncanonical protein synthesis process in which peptide synthesis is initiated at the site of the expanded repeats in absence of an AUG codon. In the case of FXTAS, RAN translation leads to the synthesis of toxic polyglycine peptides that accumulate and form protein aggregates. Gain-of-function mechanisms described in (C) and (D) are mutually nonexclusive and can occur at the same time.

tion at the RNA level: unmethylated intermediate size CGG repeats can form stable secondary structures called G-quadruplexes and bind specific RNA-binding proteins, such as hnRNP A2B1, DROSHA, SAM68, and TDP-43 [17, 18, 19]. Aberrant RNA–protein complexes form RNA foci (also called inclusion bodies) and sequester bound proteins, preventing them from performing their normal function [20, 21, 22] (Figure 1C). The second pathogenic mechanism is the expression of toxic polypeptides directly produced by expansion by a process known as repeat-associated non-AUG (RAN) translation [23, 24, 25, 26]. RAN translation is a noncanonical protein synthesis process first described in spinocerebellar ataxia type 8 (SCA8, MIM #608768) and myotonic dystrophy type 1 (DM1, MIM #160900), in which peptide synthesis is initiated at the site of the expanded repeats in absence of an AUG codon [27] (Figure 1D). This process can theoretically occur in the three reading frames on both sense and antisense DNA strands, although only specific peptides are preferentially expressed or toxic. CGG repeats mainly lead to the abnormal expression of polyglycine (polyG) peptides that are also able to accumulate and form protein aggregates via a prion-like mechanism [24, 28, 29]. Although initially described in a pathological context, RAN translation could be a physiological process contributing to the regulation of *FMR1* expression in neurons by creating an upstream open reading frame (uORF) competing with FMRP expression [29].

RNA and protein gains-of-function are intimately linked together and probably both contribute to the pathogenesis of the disorder. Recent evidence shows that polyG peptides interact with pathogenic CGG repeat-derived RNA G-quadruplexes and that these RNA molecules could even promote the formation of polyG aggregates [28].

GC-rich expansions associated with a loss-of-function

So far, only a few disorders other than FXS have been associated with GC-rich expansions causing epigenetic gene

silencing (Table 1). Like FXS, CCG repeat expansions in the 5′UTR of *AFF2* (AF4/FMR2 family member 2, previously *FMR2*, MIM #300806) on chromosome X are associated with another X-linked ID disorder in males (FRAXE, MIM #309548) described in 1993 [30]. Intragenic deletions of *AFF2* have been identified in patients with ID [31, 32] and an excess of point mutations in *AFF2* has been described in males with ASD [33], further supporting the association of this gene with neurodevelopmental disorders.

Expansions associated with a loss-of-function have also been identified on autosomes. Most of the disorders described so far are recessive and, in this case, the disease is caused by an expansion in both alleles or by an expansion in one allele and a point variant in the other allele. The phenotype associated with expansions and point variants is usually identical. Such compound heterozygous alterations can be difficult to detect and their identification needs the combination of expansion detection and standard gene panel or exome analysis. At least three disorders corresponding to this description have been described so far.

Dodecamer (CCCCGCCCCGCG) expansions in the 5′UTR of *CSTB* (cystatin-B, MIM #601145) are responsible for progressive myoclonic epilepsy type 1 (EPM1, also known as Unverricht–Lundborg disease, MIM #254800), a recessive neurodegenerative epileptic condition characterized by tonic-clonic seizures and myoclonus [34]. Pathogenic *CSTB* expansions in both alleles or in one allele plus a point variant in the other allele cause the loss-of-function of cystatin B (stefin B), a small proteinase inhibitor, whose precise function still remains largely unknown [35].

More recently, CGG repeat expansions also leading to loss-of-function through hypermethylation have been described in *XYLT1* (xylosyltransferase 1, MIM #608124). Recessive variants in this gene had previously been described to cause Desbuquois dysplasia 2 (DBQD2, MIM #615777), a skeletal dysplasia associated with developmental delay, short stature, and facial characteristics. Expansions in *XYLT1* were uncovered using a combination of genome sequencing, microarray analysis, and Sanger sequencing in patients with Baratela-Scott syndrome (BSS), another skeletal dysplasia sharing many clinical features with DBQD2. The authors first identified homozygous or compound heterozygous pathogenic variants or deletions altering the coding region of *XYLT1* in a few patients. Segregation analysis of the variants within families revealed allelic drop-out, which prompted the authors to look for DNA methylation defects. This analysis revealed hypermethylation of alleles without point variants, consecutive to CGG expansions in the 5′UTR of *XYLT1* in a region that was incorrect in the reference genome [36]. BBS and DBQD2 are thus allelic disorders both linked to loss-of-function of *XYLT1* as a result of point variants or noncoding CGG expansions.

Finally, noncoding expansions in *GLS* (glutaminase, MIM #138280) in patients with global developmental delay, progressive ataxia, and elevated plasma glutamine (GDPAG, MIM #618412) were identified thanks to their associated biochemical phenotype. *GLS* encodes glutaminase, the enzyme catalyzing the first reaction of glutamine catabolism, an obvious candidate gene for elevated plasma glutamine. Heterozygous point variants in this gene were identified by exome sequencing in only two of three unrelated individuals with GDPAG, whereas all three had strongly impaired glutaminase activity, suggesting the existence of pathogenic variants undetected by exome sequencing. The authors then applied ExpansionHunter to genome sequence data and detected a GCA repeat expansion in the 5′UTR of *GLS* [37]. Expansions were either present in both alleles or in one allele, with a point variant in the other allele, and and exhibited a number of GCA repeats ranging from 400 to 1,500 (8–16 in control individuals).

Contrary to expansions causing FXS and BSS, expansions in neither *CSTB* nor *GLS* are hypermethylated and they both seem to lead to reduced gene transcription independently of DNA methylation [37, 38]. Unlike CGG expansions, GCA expansions in *GLS* do not contain any CpG, which is the only substrate of mammalian DNA methyltransferases, and thus cannot be methylated. Instead, they are associated with changes in histone modifications, including a decrease in transcriptionally active (H3K27ac and H3K4me3) marks and an increase in transcriptionally silent (H3K9me3) modifications [37]. These findings suggest a change in chromatin configuration as the result of the repeat expansion, as already shown for intronic GAA expansions in *FXN* (Frataxin, MIM #606829) causing Friedreich ataxia (FRDA, MIM #229300). These expansions alter the transcription of Frataxin by creating secondary DNA/RNA structures called R-loops, which block RNA polymerase and are associated with repressive histone marks [39, 40].

So far, only very few disorders have been associated with dominant repeat expansion causing a loss-of-function: CGG expansions in *DIP2B* (Disco-interacting protein 2 homolog B, MIM #611379), associated with hypermethylation and a fragile site on chr12q13.12, have been reported to lead to a dominant nonsyndromic ID disorder (FRA12A, MIM #136630) [49]. However, no additional patients have been described since the initial study and

Table 1: List of noncoding repeat expansion disorders involving GC-rich (>66 % GC) motifs. BSS, Baratela-Scott syndrome; DM1, myotonic dystrophy type 1; DM2, myotonic dystrophy type 2; EPM1, progressive myoclonus epilepsy type 1 (Unverricht–Lundborg disease); FECD3, Fuchs endothelial corneal dystrophy type 3; FRAXE, Fragile XE syndrome; ALS/FTD, amyotrophic lateral sclerosis/frontotemporal dementia; FXS, Fragile X syndrome; FXTAS, Fragile X-associated tremor ataxia syndrome; GDPAG, global developmental delay, progressive ataxia, and elevated glutamine; NIID, neuronal intranuclear inclusion disease; OPDM1–3, Oculopharyngodistal myopathy type 1–3; OPML1, oculopharyngeal myopathy with leukoencephalopathy type 1; RCPS, Richieri-Costa–Pereira syndrome; SCA8/12/36, spinocerebellar ataxia type 8/12/36; AD, autosomal dominant; XL, X-linked; NA, not available; UTR, untranslated region.

Disorder	Inheritance	Chromosome	Gene	Location	Repeat motif	Normal repeat range	Pathological repeat range	Gene silencing (methylation)	RNA foci (sequestration)	RNA translation (toxic peptide)	References
NIID/ OPDM3	AD	1q21.2	*NOTCH2NLC*	5′UTR/ exon 1	CGG	7–40	≥60-500	variable (length-dependent)	yes (likely)	likely (polyG)	[41, 42, 43]
FRA2A	AD	2q11.2	*AFF3*	Intron	CGG	8–17	≥300	yes (yes)	NA	NA	[44]
GDPAG	AR	2q32.2	*GLS*	5′UTR	GCA	8–16	≥680-1,400	yes (no)	NA	NA	[37]
DM2	AD	3q21.3	*CNBP*	Intron	CCTG/CAGG	11–30	>50 -11,000	no	yes (yes)	yes (tetrapeptides)	[45]
SCA12	AD	5q32	*PPP2R2B*	5′UTR	CAG	4–32	≥43–78	no	yes (yes)	yes (polyQ, polyS)	[46]
OPDM1	AD	8q22.3	*LRP12*	5′UTR	CGG	13–45	90-130	no	likely	likely (polyG)	[41]
ALS/FTD	AD	9p21.2	*C9ORF72*	5′UTR/ intron	GGGGCC	3–25	>30	variable (length-dependent)	yes (yes)	yes (dipeptides)	[47, 48]
OPML1	AD	10q22.3	*LOC642361/ NUTM2B-AS1*	Noncoding transcript gene	CGG/CCG	3–16	40-60	NA	yes (likely)	likely (polyG)	[41]
FRA12A	AD	12q13.12	*DIP2B*	5′UTR	CGG	NA	NA	yes (yes)	NA	NA	[49]
SCA8	AD	13q21	*ATXN8/ATXN8OS*	3′UTR	CAG/CTG	15–50	>74–250	no	yes (yes)	likely (polyQ)	[50]
BSS	AR	16p12.3	*XYLT1*	Promoter	CGG	9-20	120-800	yes (yes)	NA	NA	[36]
RCPS	AR	17q25.3	*EIF4A3*	5′UTR	18- or 20-nucleotide motifs	3-12	15-16	NA	NA	NA	[51]
FECD3	AD	18q21.2	*TCF4*	Intron	CTG	5–31	>50	no	yes (yes)	NA	[52, 53]
OPDM2	AD	19p13.12	*GIPC1*	5′UTR	CGG	12–32	≥97-120	NA	likely	likely (polyG)	[54, 55]
DM1	AD	19q13.32	*DMPK*	3′UTR	CTG	5–37	>50–10,000	no	yes (yes)	likely (polyQ)	[56]
SCA36	AD	20p13	*NOP56*	Intron	GGCCTG	5–14	≥650-2,500	possible	yes (yes)	yes (dipeptides)	[57]
EPM1	AR	21q22.3	*CSTB*	promoter/ 5′UTR	CCCCGCCCCGCG	2–3	≥30-75	yes (no)	NA	NA	[34]
FXS	XL	Xq27.3	*FMR1*	5′UTR	CGG	5–50	>200	yes (yes)	no	no	[6, 7]
FXPOI	XL	Xq27.3	*FMR1*	5′UTR	CGG	5–50	55–200	no	yes (likely)	likely (polyG)	[13]
FXTAS	XL	Xq27.3	*FMR1*	5′UTR	CGG	5–50	55–200	no	yes (yes)	yes (polyG)	[14]
FRAXE	XL	Xq28	*AFF2*	5′UTR	CCG	4–39	≥200–900	yes (yes)	NA	NA	[30, 58]

this finding thus needs to be confirmed by additional reports. Likewise, CGG expansions in the 5′UTR of *AFF3* (AF4/FMR2 family member 3, MIM #601464) result in hypermethylation associated with the FRA2A fragile site [44]. Like *DIP2B* expansions, this disease–gene association also requires further evidence but point variants in *AFF3* have recently been associated with a dominant disorder including intellectual disability, mesomelic dysplasia, horseshoe kidney, and epileptic encephalopathy [59].

GC-rich expansions associated with a gain-of-function

The first dominant noncoding repeat expansion disorders described were myotonic dystrophy type 1 (DM1) and spinocerebellar type 8 (SCA8), caused by CTG expansions in the 3′UTRs of *DMPK* (dystrophia myotonica protein kinase, MIM #605377) [56] and *ATXN8OS* (ATXN8 Opposite Strand LncRNA, MIM #603680) [50], respectively (Table 1). In the case of SCA8, a CAG expansion also exists on the reverse strand in the *ATXN8* coding gene (MIM #613289). The consequence of expansions in *DMPK* at the RNA level have extensively been studied. *DMPK* mRNA molecules containing CUG expanded repeats accumulate to form inclusions in muscle and neuron nuclei and sequester specific splicing factors such as the muscleblind-like 1 (MBNL1) protein. Consequently, the functional depletion of these RNA-binding proteins results in splicing defects of tissue-specific transcripts [60, 61]. RAN translation also occurs in DM1 and is possibly the main pathophysiological mechanism in SCA8, both mainly leading to the toxic expression of polyglutamine (polyQ) peptides, which are well known to adopt β-sheet structures prone to form insoluble fibrillar aggregates and neuronal intranuclear protein inclusions [27, 62]. RNA-dependent pathophysiological mechanisms and RAN translation of tetrapeptides (polyLPAC and polyQAGR) also coexist in myotonic dystrophy type 2 (DM2, MIM #602668), caused by CCTG/CAGG repeat expansions in *CNBP* (CCHC-type zinc finger nucleic acid-binding protein, MIM #116955). These polypeptides are able to accumulate specifically in neurons, astrocytes, and white matter structures and are toxic independently of RNA foci and nuclear sequestration of MBNL1 by CCUG transcripts. This suggests a pathophysiological model in which an RNA-dependent pathogenic mechanism first occurs in the nucleus and when sequestration capacity is exceeded, RNAs are exported to the cytoplasm where they undergo RAN translation [63].

Another example of these complexed intertwined mechanisms is exemplified by hexanucleotide (GGGGCC) expansions in *C9ORF72* (chromosome 9 open reading frame 72, MIM #614260). These expansions located in the 5′UTR (or first intron depending on the isoform) of *C9ORF72* cause a dominant disorder characterized by frontotemporal dementia, amyotrophic lateral sclerosis, or the association of both at the individual level and/or within families (ALS/FTD, MIM #105550). *C9ORF72* G_4C_2 repeats, like GCC repeats, are able to adopt G-quadruplex structures [64], and they can sequester multiple RNA-binding (mainly SRSF and hnRNP) proteins in a cell type-specific manner [65]. RAN translation of *C9ORF72* G_4C_2 repeats in sense and antisense produces five dipeptide proteins (polyGA, polyGP, polyGR, polyPA, polyPR), three of which (polyGR, polyPR, and polyGA) are highly toxic [66, 67, 68, 69, 70, 71, 72, 73]. Remarkably, polyPR and polyGA peptides are able to spread from one cell to the other via exosome-dependent but also exosome-independent mechanisms [74, 75]. Finally, *C9ORF72* G_4C_2 repeats can be methylated in a length-dependent manner and DNA methylation inversely correlates with repeat size and age at disease onset [76, 77, 78]. Loss-of-function of *C9ORF72* is insufficient to lead to ALS/FTD and no truncating mutations associated with ALS/FTD have been reported in this gene, but recent evidence suggests that *C9ORF72* haploinsufficiency could contribute to disease pathogenesis by worsening the repeat-dependent gain-of-function mechanisms [79]. The examples of DM1, SCA8, and *C9ORF72*-associated ALS/FTD show that mechanisms involving toxic RNA, RAN translation, and loss-of-function through hypermethylation are not mutually exclusive and can even have additive effects, each explaining parts of the pathophysiogenesis.

Recently, four additional dominant GC-rich repeat expansions have been identified (Table 1). The most frequent is a CGG repeat expansion in *NOTCH2NLC* (Notch 2 N-Terminal Like C, MIM #618025), one of four human-specific genes (*NOTCH2NLA*, *NOTCH2NLB*, *NOTCH2NLC*, and *NOTCH2NLR*) sharing a high degree (>99 %) of DNA homology and originating from pericentromeric tandem duplications of the 5′ part of *NOTCH2* on chromosome 1. These expansions, located in the 5′UTR/first exon of *NOTCH2NLC*, cause a dominant neurodegenerative disorder called neuronal intranuclear inclusion disease (NIID, MIM #603472). This condition is clinically variable and characterized by eosinophilic intranuclear inclusions in neurons, glial cells, fibroblasts, and muscles. The age at onset is also highly variable, ranging from infancy to late adulthood, although most patients show the first symptoms from the third decade of life. Clinical features include pyramidal and extrapyramidal symptoms, cerebellar ataxia, cognitive decline, peripheral neuropathy,

and autonomic dysfunction [80]. Moreover, most patients typically show white matter abnormalities on brain MRI reminiscent of those occasionally observed in FXTAS, including T2-weighted hyperintensity signals in the middle cerebellar peduncles and high-intensity signals in the corticomedullary junction on diffusion-weighted imaging. *NOTCH2NLC* expansions were concomitantly described in three independent studies. Ishiura et al. used TRhist and long-read Single Molecule Real-Time (SMRT) sequencing to identify *NOTCH2NLC* expansions [41]. Sone et al. and Tian et al. first performed genome-wide linkage analysis to identify overlapping intervals on chromosomes 1p22.1-q21.3 and 1p13.3-23.1 before using SMRT and/or nanopore long-read sequencing to detect and characterize *NOTCH2NLC* expansions [42, 43]. Numerous follow-up studies revealed the presence of pathogenic *NOTCH2NLC* expansions in patients with essential tremor (ETM6, MIM #618866) [81, 82], FTD and Alzheimer-like dementias [43, 83], Parkinsonism [43, 84, 85], multiple system atrophy [86], and oculopharyngodistal myopathy (OPDM3) [87, 88]. *NOTCH2NLC* expansions are more frequent in Japan and China, due to a founder effect in Asian populations, but can also be present in patients from other geographic origins including Europe and can even occur *de novo* in sporadic cases [89]. Pathogenic expansions range from 60 to more than 500 repeats whereas control individuals have less than 40 CGG repeats [41, 42, 90, 91]. *NOTCH2NLC* is correctly annotated only in the hg38 reference genome and the diagnostic testing of the expansion is complicated by the almost identical *NOTCH2NL* copies, the GC-rich nature of the expansions as well as the presence of interrupting AGG motifs present in a subset of both healthy and affected individuals [90]. *NOTCH2NCL* expansions are not associated with DNA hypermethylation and do not consistently alter the expression of *NOTCH2NLC*, but antisense transcripts are specifically produced in affected individuals, suggesting pathological mechanisms involving a toxic gain of function at the RNA level and/or the existence of RAN translation [41, 42, 92]. A recent study confirmed that RNA molecules with expanded CGG repeats can form RNA foci and sequester RNA-binding proteins into p62-positive intranuclear inclusions specifically in affected individuals [91].

Similar CGG expansions in at least three different genes have been identified in oculopharyngeal myopathy (OPML) and oculopharyngodistal myopathy (OPDM). Two of these expansions both composed of CGG repeats in *LOC642361*, a long noncoding RNA gene overlapping the *NUTM2B-AS1* antisense transcript (NUTM2B Antisense RNA 1, MIM #618639) on Chr. 10q22.3, and in the 5′UTR of *LRP12* (low-density lipoprotein receptor-related protein 12, MIM #618299) were identified by Ishiura and collaborators using the same strategy that initially detected *NOTCH2NLC* expansions [41]. *LRP12* expansions were detected in several families with oculopharyngodistal myopathy (OPDM1, MIM #164310), a neuromuscular disorder in which muscular weakness in the legs and arms is associated with external ophthalmoplegia, dysphagia, and ptosis, while expansions in *LOC642361*/*NUTM2B-AS1* were found in a single family with oculopharyngeal myopathy, limb weakness, ataxia, ptosis, and white matter abnormalities similar to those seen in NIID (OPML1, MIM #618637). Two independent studies identified CGG repeat expansions in the 5′UTR of *GIPC1* (MIM #605072) in multiple families with OPDM2 (MIM #618940) using a combination of whole-genome sequencing and long-read sequencing [54, 55]. *GIPC1* expansions lead to increased mRNA expression but do not affect protein expression [54]. Altogether, these recent studies indicate that CGG expansions in multiple genes lead to dominant neurodegenerative disorders irrespectively of the gene where they occur by mechanisms that likely resemble those described in FXTAS and *C9ORF72*-associated FTD/ALS [92].

The relationship between GC-rich repeat expansions, DNA methylation, and gene expression remains unclear. Although *FMR1* expansions have outlined a clear correlation between expansion size and hypermethylation locking the gene in an unexpressed state, the same threshold does not seem to apply to all genes equally. Indeed, expansions containing more than 200 GCC repeats in *NOTCH2NLC* or *C9ORF72* for instance are not consistently associated with DNA methylation. Contrary to full expansions in *FMR1*, large expansions in these genes still allow transcription of RNA molecules containing expanded repeats and RAN translation of toxic polypeptides able to aggregate and form inclusions. DNA methylation could even be protective in some *NOTCH2NLC*-associated NIID [91] while worsening the effect of RNA and peptide toxicity in *C9ORF72*-associated ALS/FTD [79]. In this setting, studying how DNA methylation impacts the progression of each disorder associated with GC-rich expansions is of crucial importance as this information could be used to design new treatment strategies based on Cas9 methylation editing, as recently suggested for FXS [93].

Perspectives on the identification of GC-rich expansions

Most recent studies reporting repeat expansions relied on large and/or multiple families that allowed the identifica-

tion of a genomic interval prior to the expansion. Some of the recently identified disorders turned out to occur quite frequently, suggesting that many rarer disorders associated with undetected repeat expansions exist. Indeed, because of their repetitive nature, high degree of polymorphism, and abundance in human genomes, repeat expansions remain difficult to detect by standard amplification or sequencing technologies. Repeat expansions can be looked for from short-read sequencing data using specific tools such as LobSTR [94], HipSTR [95], TREDPARSE [96], ExpansionHunter [97], STRetch [98], GangSTR [99], and exSTRa [100], but most of these tools need inputs regarding the genomic region and repeat motifs and their use is often limited to detect already known expansions. So far, only two bioinformatics tools, TRhist [101] and ExpansionHunter DeNovo [102], can detect any type of repeat expansions at a genome-wide scale. However, since short reads encompassing repeats usually map to multiple genomic regions and are clipped off or discarded, the existing tools tend to perform poorly in estimating the number of repeats on each allele and underestimate the number of repeats, especially in the case of very large expansions. For this reason, long-read technologies including Oxford Nanopore sequencing and SMRT sequencing have become the new standard to detect and characterize repeat expansions [103]. These powerful technologies detect several hundreds of structural variants per individual that mostly correspond to polymorphic repeat elements. The normal repeat ranges and associated allele frequencies of these repeat variations have not yet been described in large control populations, and the detection of expansions without hypothesis on the expanded motif or genomic region where it is located thus remains a challenge in practice. However, this field is rapidly evolving. Specific tools like NanoSatellite [104] and tandem-genotypes [105] have already been developed to specifically study repeats present in long-read data and we can hope that the identification of repeat expansions will be soon integrated to standard genetic pipelines.

Although GC-rich repeat expansions are mainly known to cause monogenic disorders, it is very likely that this type of genetic variation also largely contributes to the genetic architecture of complex disorders [106]. A recent study investigating the contribution of tandem repeats to the risk of developing ASD in cohorts of >5,000 patients showed that repeat expansions were more prevalent in subjects with ASD (23.3 %) than their healthy siblings (20.67 %). These findings suggest that repeat expansions at more than 2,500 loci account for 2.6 % of autism risk [107]. Most of the top candidate regions identified are GC-rich and include known repeat expansion loci (*FMR1*, *FXN*, and *DMPK*) as well as new loci in genes associated with monogenic disorders, such as *MBOAT7*, *CDON*, *IL1RAPL1*, and *FGF14*. Another study focusing on *de novo* repeat changes showed a significant excess of repeat expansion in ASD subjects. Interestingly, these *de novo* repeat additions mainly occur in conserved fetal brain regulatory regions [108]. Repeat expansions located in regulatory regions, which are very often GC-rich, have been shown to have a significant impact on gene expression [109, 110]. This observation holds true for copy number variation involving microsatellites (i. e., motifs less than 1–9 bp), but for variable number of tandem repeats (VNTR), involved motifs are ≥10 bp [111]. Further studies are therefore required to address this complexity and clarify the role of tandem repeat expansions in rare and more common human disorders.

Acknowledgments: We thank the three anonymous reviewers who made valuable contributions to the revision of this review.

Research funding: The research studies conducted by the authors are supported by University Hospital Essen, the Deutsche Forschungsgemeinschaft (DFG), the Bundesministerium für Bildung und Forschung (BMBF), Fondation Maladies Rares, and the Tom-Wahlig Stiftung.

Author contributions: All authors have accepted responsibility for the entire content of this manuscript and approved its submission.

Competing interests: Authors state no conflict of interest.

Informed consent: Not applicable.

Ethical approval: Not applicable.

References

[1] Gymrek M. A genomic view of short tandem repeats. Curr Opin Genet Dev. 2017;44:9–16.

[2] Willems T, Gymrek M, Highnam G, Genomes Project C, Mittelman D, Erlich Y. The landscape of human STR variation. Genome Res. 2014;24(11):1894–904.

[3] Depienne C, Mandel JL. 30 years of repeat expansion disorders: what have we learned and what are the remaining challenges? Am J Hum Genet. 2021;108(5):764–85.

[4] Hannan AJ. Tandem repeats mediating genetic plasticity in health and disease. Nat Rev Genet. 2018;19(5):286–98.

[5] Malik I, Kelley CP, Wang ET, Todd PK. Molecular mechanisms underlying nucleotide repeat expansion disorders. Nat Rev Mol Cell Biol. 2021;22(9):589–607.

[6] Verkerk AJ, Pieretti M, Sutcliffe JS, Fu YH, Kuhl DP, Pizzuti A et al. Identification of a gene (FMR-1) containing a CGG repeat coincident with a breakpoint cluster region exhibiting length variation in fragile X syndrome. Cell. 1991;65(5):905–14.

[7] Oberle I, Rousseau F, Heitz D, Kretz C, Devys D, Hanauer A et al. Instability of a 550-base pair DNA segment and abnormal methylation in fragile X syndrome. Science. 1991;252(5009):1097–102.

[8] Heitz D, Devys D, Imbert G, Kretz C, Mandel JL. Inheritance of the fragile X syndrome: size of the fragile X premutation is a major determinant of the transition to full mutation. J Med Genet. 1992;29(11):794–801.

[9] Devys D, Biancalana V, Rousseau F, Boue J, Mandel JL, Oberle I. Analysis of full fragile X mutations in fetal tissues and monozygotic twins indicate that abnormal methylation and somatic heterogeneity are established early in development. Am J Med Genet. 1992;43(1–2):208–16.

[10] Quartier A, Poquet H, Gilbert-Dussardier B, Rossi M, Casteleyn AS, Portes VD et al. Intragenic FMR1 disease-causing variants: a significant mutational mechanism leading to Fragile-X syndrome. Eur J Hum Genet. 2017;25(4):423–31.

[11] Brown WT, Jenkins EC, Goonewardena P, Miezejeski C, Atkin J, Devys D. Prenatally detected fragile X females: long-term follow-up studies show high risk of mental impairment. Am J Med Genet. 1992;43(1–2):96–102.

[12] Myers KA, van 't Hof FNG, Sadleir LG, Legault G, Simard-Tremblay E, Amor DJ et al. Fragile females: case series of epilepsy in girls with FMR1 disruption. Pediatrics. 2019;144(3):e20190599.

[13] Conway GS, Payne NN, Webb J, Murray A, Jacobs PA. Fragile X premutation screening in women with premature ovarian failure. Hum Reprod. 1998;13(5):1184–7.

[14] Hagerman RJ, Leehey M, Heinrichs W, Tassone F, Wilson R, Hills J et al. Intention tremor, parkinsonism, and generalized brain atrophy in male carriers of fragile X. Neurology. 2001;57(1):127–30.

[15] Devys D, Lutz Y, Rouyer N, Bellocq JP, Mandel JL. The FMR-1 protein is cytoplasmic, most abundant in neurons and appears normal in carriers of a fragile X premutation. Nat Genet. 1993;4(4):335–40.

[16] Oostra BA, Willemsen R. FMR1: a gene with three faces. Biochim Biophys Acta. 2009;1790(6):467–77.

[17] Sofola OA, Jin P, Qin Y, Duan R, Liu H, de Haro M et al. RNA-binding proteins hnRNP A2/B1 and CUGBP1 suppress fragile X CGG premutation repeat-induced neurodegeneration in a Drosophila model of FXTAS. Neuron. 2007;55(4):565–71.

[18] Sellier C, Rau F, Liu Y, Tassone F, Hukema RK, Gattoni R et al. Sam68 sequestration and partial loss of function are associated with splicing alterations in FXTAS patients. EMBO J. 2010;29(7):1248–61.

[19] Sellier C, Freyermuth F, Tabet R, Tran T, He F, Ruffenach F et al. Sequestration of DROSHA and DGCR8 by expanded CGG RNA repeats alters microRNA processing in fragile X-associated tremor/ataxia syndrome. Cell Rep. 2013;3(3):869–80.

[20] Sellier C, Usdin K, Pastori C, Peschansky VJ, Tassone F, Charlet-Berguerand N. The multiple molecular facets of fragile X-associated tremor/ataxia syndrome. J Neurodev Disord. 2014;6(1):23.

[21] Verma AK, Khan E, Mishra SK, Mishra A, Charlet-Berguerand N, Curcumin KA. Regulates the r(CGG) (exp) RNA hairpin structure and ameliorate defects in fragile X-associated tremor ataxia syndrome. Front Neurosci. 2020;14:295.

[22] Rosario R, Anderson R. The molecular mechanisms that underlie fragile X-associated premature ovarian insufficiency: is it RNA or protein based? Mol Hum Reprod. 2020;26(10):727–37.

[23] Glineburg MR, Todd PK, Charlet-Berguerand N, Sellier C. Repeat-associated non-AUG (RAN) translation and other molecular mechanisms in fragile X tremor ataxia syndrome. Brain Res. 2018;1693(Pt A):43–54.

[24] Sellier C, Buijsen RAM, He F, Natla S, Jung L, Tropel P et al. Translation of expanded CGG repeats into FMRpolyG is pathogenic and may contribute to fragile X tremor ataxia syndrome. Neuron. 2017;93(2):331–47.

[25] Krans A, Kearse MG, Todd PK. Repeat-associated non-AUG translation from antisense CCG repeats in fragile X tremor/ataxia syndrome. Ann Neurol. 2016;80(6):871–81.

[26] Buijsen RA, Visser JA, Kramer P, Severijnen EA, Gearing M, Charlet-Berguerand N et al. Presence of inclusions positive for polyglycine containing protein, FMRpolyG, indicates that repeat-associated non-AUG translation plays a role in fragile X-associated primary ovarian insufficiency. Hum Reprod. 2016;31(1):158–68.

[27] Zu T, Gibbens B, Doty NS, Gomes-Pereira M, Huguet A, Stone MD et al. Non-ATG-initiated translation directed by microsatellite expansions. Proc Natl Acad Sci USA. 2011;108(1):260–5.

[28] Asamitsu S, Yabuki Y, Ikenoshita S, Kawakubo K, Kawasaki M, Usuki S et al. CGG repeat RNA G-quadruplexes interact with FMRpolyG to cause neuronal dysfunction in fragile X-related tremor/ataxia syndrome. Sci Adv. 2021;7(3):eabd9440.

[29] Rodriguez CM, Wright SE, Kearse MG, Haenfler JM, Flores BN, Liu Y et al. A native function for RAN translation and CGG repeats in regulating fragile X protein synthesis. Nat Neurosci. 2020;23(3):386–97.

[30] Knight SJ, Flannery AV, Hirst MC, Campbell L, Christodoulou Z, Phelps SR et al. Trinucleotide repeat amplification and hypermethylation of a CpG island in FRAXE mental retardation. Cell. 1993;74(1):127–34.

[31] Stettner GM, Shoukier M, Hoger C, Brockmann K, Auber B. Familial intellectual disability and autistic behavior caused by a small FMR2 gene deletion. Am J Med Genet, Part A. 2011;155A(8):2003–7.

[32] Sahoo T, Theisen A, Marble M, Tervo R, Rosenfeld JA, Torchia BS et al. Microdeletion of Xq28 involving the AFF2 (FMR2) gene in two unrelated males with developmental delay. Am J Med Genet, Part A. 2011;155A(12):3110–5.

[33] Mondal K, Ramachandran D, Patel VC, Hagen KR, Bose P, Cutler DJ et al. Excess variants in AFF2 detected by massively parallel sequencing of males with autism spectrum disorder. Hum Mol Genet. 2012;21(19):4356–64.

[34] Lalioti MD, Scott HS, Buresi C, Rossier C, Bottani A, Morris MA et al. Dodecamer repeat expansion in cystatin B gene in progressive myoclonus epilepsy. Nature. 1997;386(6627):847–51.

[35] Canafoglia L, Gennaro E, Capovilla G, Gobbi G, Boni A, Beccaria F et al. Electroclinical presentation and genotype-phenotype relationships in patients with Unverricht-Lundborg disease carrying compound heterozygous CSTB point and indel mutations. Epilepsia. 2012;53(12):2120–7.

[36] LaCroix AJ, Stabley D, Sahraoui R, Adam MP, Mehaffey M, Kernan K et al. GGC repeat expansion and exon 1 methylation of XYLT1 is a common pathogenic variant in Baratela-Scott syndrome. Am J Hum Genet. 2019;104(1):35–44.
[37] van Kuilenburg ABP, Tarailo-Graovac M, Richmond PA, Drogemoller BI, Pouladi MA, Leen R et al. Glutaminase deficiency caused by short tandem repeat expansion in GLS. N Engl J Med. 2019;380(15):1433–41.
[38] Weinhaeusel A, Morris MA, Antonarakis SE, Haas OA. DNA deamination enables direct PCR amplification of the cystatin B (CSTB) gene-associated dodecamer repeat expansion in myoclonus epilepsy type Unverricht-Lundborg. Human Mutat. 2003;22(5):404–8.
[39] Grabczyk E, Kumari D, Fragile UK. X syndrome and Friedreich's ataxia: two different paradigms for repeat induced transcript insufficiency. Brain Res Bull. 2001;56(3–4):367–73.
[40] Groh M, Lufino MM, Wade-Martins R, Gromak N. R-loops associated with triplet repeat expansions promote gene silencing in Friedreich ataxia and fragile X syndrome. PLoS Genet. 2014;10(5):e1004318.
[41] Ishiura H, Shibata S, Yoshimura J, Suzuki Y, Qu W, Doi K et al. Noncoding CGG repeat expansions in neuronal intranuclear inclusion disease, oculopharyngodistal myopathy and an overlapping disease. Nat Genet. 2019;51(8):1222–32.
[42] Sone J, Mitsuhashi S, Fujita A, Mizuguchi T, Hamanaka K, Mori K et al. Long-read sequencing identifies GGC repeat expansions in NOTCH2NLC associated with neuronal intranuclear inclusion disease. Nat Genet. 2019;51(8):1215–21.
[43] Tian Y, Wang JL, Huang W, Zeng S, Jiao B, Liu Z et al. Expansion of human-specific GGC repeat in neuronal intranuclear inclusion disease-related disorders. Am J Hum Genet. 2019;105(1):166–76.
[44] Metsu S, Rooms L, Rainger J, Taylor MS, Bengani H, Wilson DI et al. FRA2A is a CGG repeat expansion associated with silencing of AFF3. PLoS Genet. 2014;10(4):e1004242.
[45] Liquori CL, Ricker K, Moseley ML, Jacobsen JF, Kress W, Naylor SL et al. Myotonic dystrophy type 2 caused by a CCTG expansion in intron 1 of ZNF9. Science. 2001;293(5531):864–7.
[46] Holmes SE, O'Hearn EE, McInnis MG, Gorelick-Feldman DA, Kleiderlein JJ, Callahan C et al. Expansion of a novel CAG trinucleotide repeat in the 5' region of PPP2R2B is associated with SCA12. Nat Genet. 1999;23(4):391–2.
[47] Renton AE, Majounie E, Waite A, Simon-Sanchez J, Rollinson S, Gibbs JR et al. A hexanucleotide repeat expansion in C9ORF72 is the cause of chromosome 9p21-linked ALS-FTD. Neuron. 2011;72(2):257–68.
[48] DeJesus-Hernandez M, Mackenzie IR, Boeve BF, Boxer AL, Baker M, Rutherford NJ et al. Expanded GGGGCC hexanucleotide repeat in noncoding region of C9ORF72 causes chromosome 9p-linked FTD and ALS. Neuron. 2011;72(2):245–56.
[49] Winnepenninckx B, Debacker K, Ramsay J, Smeets D, Smits A, FitzPatrick DR et al. CGG-repeat expansion in the DIP2B gene is associated with the fragile site FRA12A on chromosome 12q13.1. Am J Hum Genet. 2007;80(2):221–31.
[50] Koob MD, Moseley ML, Schut LJ, Benzow KA, Bird TD, Day JW et al. An untranslated CTG expansion causes a novel form of spinocerebellar ataxia (SCA8). Nat Genet. 1999;21(4):379–84.
[51] Favaro FP, Alvizi L, Zechi-Ceide RM, Bertola D, Felix TM, de Souza J et al. A noncoding expansion in EIF4A3 causes Richieri-Costa-Pereira syndrome, a craniofacial disorder associated with limb defects. Am J Hum Genet. 2014;94(1):120–8.
[52] Mootha VV, Gong X, Ku HC, Xing C. Association and familial segregation of CTG18.1 trinucleotide repeat expansion of TCF4 gene in Fuchs' endothelial corneal dystrophy. Investig Ophthalmol Vis Sci. 2014;55(1):33–42.
[53] Mootha VV, Hussain I, Cunnusamy K, Graham E, Gong X, Neelam S et al. TCF4 triplet repeat expansion and nuclear RNA foci in Fuchs' endothelial corneal dystrophy. Investig Ophthalmol Vis Sci. 2015;56(3):2003–11.
[54] Deng J, Yu J, Li P, Luan X, Cao L, Zhao J et al. Expansion of GGC repeat in GIPC1 is associated with oculopharyngodistal myopathy. Am J Hum Genet. 2020;106(6):793–804.
[55] Xi J, Wang X, Yue D, Dou T, Wu Q, Lu J et al. 5' UTR CGG repeat expansion in GIPC1 is associated with oculopharyngodistal myopathy. Brain. 2021;144(2):601–14.
[56] Mahadevan M, Tsilfidis C, Sabourin L, Shutler G, Amemiya C, Jansen G et al. Myotonic dystrophy mutation: an unstable CTG repeat in the 3' untranslated region of the gene. Science. 1992;255(5049):1253–5.
[57] Kobayashi H, Abe K, Matsuura T, Ikeda Y, Hitomi T, Akechi Y et al. Expansion of intronic GGCCTG hexanucleotide repeat in NOP56 causes SCA36, a type of spinocerebellar ataxia accompanied by motor neuron involvement. Am J Hum Genet. 2011;89(1):121–30.
[58] Knight SJ, Voelckel MA, Hirst MC, Flannery AV, Moncla A, Davies KE. Triplet repeat expansion at the FRAXE locus and X-linked mild mental handicap. Am J Hum Genet. 1994;55(1):81–6.
[59] Voisin N, Schnur RE, Douzgou S, Hiatt SM, Rustad CF, Brown NJ et al. Variants in the degron of AFF3 are associated with intellectual disability, mesomelic dysplasia, horseshoe kidney, and epileptic encephalopathy. Am J Hum Genet. 2021;108(5):857–73.
[60] Kanadia RN, Johnstone KA, Mankodi A, Lungu C, Thornton CA, Esson D et al. A muscleblind knockout model for myotonic dystrophy. Science. 2003;302(5652):1978–80.
[61] Timchenko LT, Miller JW, Timchenko NA, DeVore DR, Datar KV, Lin L et al. Identification of a (CUG)n triplet repeat RNA-binding protein and its expression in myotonic dystrophy. Nucleic Acids Res. 1996;24(22):4407–14.
[62] Cleary JD, Pattamatta A, Ranum LPW. Repeat-associated non-ATG (RAN) translation. J Biol Chem. 2018;293(42):16127–41.
[63] Zu T, Cleary JD, Liu Y, Banez-Coronel M, Bubenik JL, Ayhan F et al. RAN translation regulated by muscleblind proteins in myotonic dystrophy type 2. Neuron. 2017;95(6):1292–305 e5.
[64] Reddy K, Zamiri B, Stanley SYR, Macgregor RB Jr, Pearson CE. The disease-associated r(GGGGCC)n repeat from the C9orf72 gene forms tract length-dependent uni- and multimolecular RNA G-quadruplex structures. J Biol Chem. 2013;288(14):9860–6.
[65] Cooper-Knock J, Walsh MJ, Higginbottom A, Robin Highley J, Dickman MJ, Edbauer D et al. Sequestration of multiple RNA recognition motif-containing proteins by C9orf72 repeat expansions. Brain. 2014;137(Pt 7):2040–51.

[66] Kwon I, Xiang S, Kato M, Wu L, Theodoropoulos P, Wang T et al. Poly-dipeptides encoded by the C9orf72 repeats bind nucleoli, impede RNA biogenesis, and kill cells. Science. 2014;345(6201):1139–45.
[67] Wen X, Tan W, Westergard T, Krishnamurthy K, Markandaiah SS, Shi Y et al. Antisense proline-arginine RAN dipeptides linked to C9ORF72-ALS/FTD form toxic nuclear aggregates that initiate in vitro and in vivo neuronal death. Neuron. 2014;84(6):1213–25.
[68] Tao Z, Wang H, Xia Q, Li K, Li K, Jiang X et al. Nucleolar stress and impaired stress granule formation contribute to C9orf72 RAN translation-induced cytotoxicity. Hum Mol Genet. 2015;24(9):2426–41.
[69] Lee KH, Zhang P, Kim HJ, Mitrea DM, Sarkar M, Freibaum BD et al. C9orf72 dipeptide repeats impair the assembly, dynamics, and function of membrane-less organelles. Cell. 2016;167(3):774–88 e17.
[70] Flores BN, Dulchavsky ME, Krans A, Sawaya MR, Paulson HL, Todd PK et al. Distinct C9orf72-associated dipeptide repeat structures correlate with neuronal toxicity. PLoS ONE. 2016;11(10):e0165084.
[71] Boivin M, Pfister V, Gaucherot A, Ruffenach F, Negroni L, Sellier C et al. Reduced autophagy upon C9ORF72 loss synergizes with dipeptide repeat protein toxicity in G4C2 repeat expansion disorders. EMBO J. 2020;39(4):e100574.
[72] Chang YJ, Jeng US, Chiang YL, Hwang IS, Chen YR. The glycine-alanine dipeptide repeat from C9orf72 hexanucleotide expansions forms toxic amyloids possessing cell-to-cell transmission properties. J Biol Chem. 2016;291(10):4903–11.
[73] Sun Y, Eshov A, Zhou J, Isiktas AU, Guo JU. C9orf72 arginine-rich dipeptide repeats inhibit UPF1-mediated RNA decay via translational repression. Nat Commun. 2020;11(1):3354.
[74] Khosravi B, LaClair KD, Riemenschneider H, Zhou Q, Frottin F, Mareljic N et al. Cell-to-cell transmission of C9orf72 poly-(Gly-Ala) triggers key features of ALS/FTD. EMBO J. 2020;39(8):e102811.
[75] Westergard T, Jensen BK, Wen X, Cai J, Kropf E, Iacovitti L et al. Cell-to-cell transmission of dipeptide repeat proteins linked to C9orf72-ALS/FTD. Cell Rep. 2016;17(3):645–52.
[76] Gijselinck I, Van Mossevelde S, van der Zee J, Sieben A, Engelborghs S, De Bleecker J et al. The C9orf72 repeat size correlates with onset age of disease, DNA methylation and transcriptional downregulation of the promoter. Mol Psychiatry. 2016;21(8):1112–24.
[77] Rohilla KJ, Gagnon KT. RNA biology of disease-associated microsatellite repeat expansions. Acta Neuropathol Commun. 2017;5(1):63.
[78] Xi Z, Zhang M, Bruni AC, Maletta RG, Colao R, Fratta P et al. The C9orf72 repeat expansion itself is methylated in ALS and FTLD patients. Acta Neuropathol. 2015;129(5):715–27.
[79] Zhu Q, Jiang J, Gendron TF, McAlonis-Downes M, Jiang L, Taylor A et al. Reduced C9ORF72 function exacerbates gain of toxicity from ALS/FTD-causing repeat expansion in C9orf72. Nat Neurosci. 2020;23(5):615–24.
[80] Sone J, Mori K, Inagaki T, Katsumata R, Takagi S, Yokoi S et al. Clinicopathological features of adult-onset neuronal intranuclear inclusion disease. Brain. 2016;139(Pt 12):3170–86.
[81] Chen H, Lu L, Wang B, Hua X, Wan B, Sun M et al. Essential tremor as the early symptom of NOTCH2NLC gene-related repeat expansion disorder. Brain. 2020;143(7):e56.
[82] Sun QY, Xu Q, Tian Y, Hu ZM, Qin LX, Yang JX et al. Expansion of GGC repeat in the human-specific NOTCH2NLC gene is associated with essential tremor. Brain. 2020;143(1):222–33.
[83] Jiao B, Zhou L, Zhou Y, Weng L, Liao X, Tian Y et al. Identification of expanded repeats in NOTCH2NLC in neurodegenerative dementias. Neurobiol Aging. 2020;89:142.
[84] Ma D, Tan YJ, Ng ASL, Ong HL, Sim W, Lim WK et al. Association of NOTCH2NLC repeat expansions with Parkinson disease. JAMA Neurol. 2020;77(12):1559–63.
[85] Yau WY, Sullivan R, Rocca C, Cali E, Vandrovcova J, Wood NW et al. NOTCH2NLC intermediate-length repeat expansion and Parkinson's disease in patients of European descent. Ann Neurol. 2021;89(3):633–5.
[86] Fang P, Yu Y, Yao S, Chen S, Zhu M, Chen Y et al. Repeat expansion scanning of the NOTCH2NLC gene in patients with multiple system atrophy. Ann Clin Transl Neurol. 2020;7(4):517–26.
[87] Ogasawara M, Iida A, Kumutpongpanich T, Ozaki A, Oya Y, Konishi H et al. CGG expansion in NOTCH2NLC is associated with oculopharyngodistal myopathy with neurological manifestations. Acta Neuropathol Commun. 2020;8(1):204.
[88] Yu J, Deng J, Guo X, Shan J, Luan X, Cao L et al. The GGC repeat expansion in NOTCH2NLC is associated with oculopharyngodistal myopathy type 3. Brain. 2021;144(6):1819–32.
[89] Okubo M, Doi H, Fukai R, Fujita A, Mitsuhashi S, Hashiguchi S et al. GGC repeat expansion of NOTCH2NLC in adult patients with leukoencephalopathy. Ann Neurol. 2019;86(6):962–8.
[90] Chen Z, Xu Z, Cheng Q, Tan YJ, Ong HL, Zhao Y et al. Phenotypic bases of NOTCH2NLC GGC expansion positive neuronal intranuclear inclusion disease in a Southeast Asian cohort. Clin Genet. 2020;98(3):274–81.
[91] Deng J, Zhou B, Yu J, Han X, Fu J, Li X, et al. Genetic origin of sporadic cases and RNA toxicity in neuronal intranuclear inclusion disease. J Med Genet. 2021.
[92] Ishiura H, Tsuji S. Advances in repeat expansion diseases and a new concept of repeat motif-phenotype correlation. Curr Opin Genet Dev. 2020;65:176–85.
[93] Liu XS, Wu H, Krzisch M, Wu X, Graef J, Muffat J et al. Rescue of fragile X syndrome neurons by DNA methylation editing of the FMR1 gene. Cell. 2018;172(5):979–92 e6.
[94] Gymrek M, Golan D, Rosset S, lobSTR EY. A short tandem repeat profiler for personal genomes. Genome Res. 2012;22(6):1154–62.
[95] 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(6):590–2.
[96] Tang H, Kirkness EF, Lippert C, Biggs WH, Fabani M, Guzman E et al. Profiling of short-tandem-repeat disease alleles in 12,632 human whole genomes. Am J Hum Genet. 2017;101(5):700–15.
[97] Dolzhenko E, van Vugt J, Shaw RJ, Bekritsky MA, van Blitterswijk M, Narzisi G et al. Detection of long repeat expansions from PCR-free whole-genome sequence data. Genome Res. 2017;27(11):1895–903.

[98] Dashnow H, Lek M, Phipson B, Halman A, Sadedin S, Lonsdale A et al. STRetch: detecting and discovering pathogenic short tandem repeat expansions. Genome Biol. 2018;19(1):121.
[99] Mousavi N, Shleizer-Burko S, Yanicky R, Gymrek M. Profiling the genome-wide landscape of tandem repeat expansions. Nucleic Acids Res. 2019;47(15):e90.
[100] Tankard RM, Bennett MF, Degorski P, Delatycki MB, Lockhart PJ, Detecting BM. Expansions of tandem repeats in cohorts sequenced with short-read sequencing data. Am J Hum Genet. 2018;103(6):858–73.
[101] Doi K, Monjo T, Hoang PH, Yoshimura J, Yurino H, Mitsui J et al. Rapid detection of expanded short tandem repeats in personal genomics using hybrid sequencing. Bioinformatics. 2014;30(6):815–22.
[102] Dolzhenko E, Bennett MF, Richmond PA, Trost B, Chen S, van Vugt J et al. ExpansionHunter Denovo: a computational method for locating known and novel repeat expansions in short-read sequencing data. Genome Biol. 2020;21(1):102.
[103] Mantere T, Kersten S, Long-Read HA. Sequencing emerging in medical genetics. Front Genet. 2019;10:426.
[104] De Roeck A, De Coster W, Bossaerts L, Cacace R, De Pooter T, Van Dongen J et al. NanoSatellite: accurate characterization of expanded tandem repeat length and sequence through whole genome long-read sequencing on PromethION. Genome Biol. 2019;20(1):239.
[105] Mitsuhashi S, Frith MC, Mizuguchi T, Miyatake S, Toyota T, Adachi H et al. Tandem-genotypes: robust detection of tandem repeat expansions from long DNA reads. Genome Biol. 2019;20(1):58.
[106] Hannan AJ. Tandem repeat polymorphisms: modulators of disease susceptibility and candidates for 'missing heritability'. Trends Genet. 2010;26(2):59–65.
[107] Trost B, Engchuan W, Nguyen CM, Thiruvahindrapuram B, Dolzhenko E, Backstrom I et al. Genome-wide detection of tandem DNA repeats that are expanded in autism. Nature. 2020;586:80–6.
[108] Mitra I, Huang B, Mousavi N, Ma N, Lamkin M, Yanicky R et al. Patterns of de novo tandem repeat mutations and their role in autism. Nature. 2021;589(7841):246–50.
[109] Fotsing SF, Margoliash J, Wang C, Saini S, Yanicky R, Shleizer-Burko S et al. The impact of short tandem repeat variation on gene expression. Nat Genet. 2019;51(11):1652–9.
[110] Gymrek M, Willems T, Guilmatre A, Zeng H, Markus B, Georgiev S et al. Abundant contribution of short tandem repeats to gene expression variation in humans. Nat Genet. 2016;48(1):22–9.
[111] Garg P, Martin-Trujillo A, Rodriguez OL, Gies SJ, Hadelia E, Jadhav B et al. Pervasive cis effects of variation in copy number of large tandem repeats on local DNA methylation and gene expression. Am J Hum Genet. 2021;108(5):809–24.

Christopher Schröder
Institute of Human Genetics, University Hospital Essen, University of Duisburg-Essen, Essen, Germany

Bernhard Horsthemke
Institute of Human Genetics, University Hospital Essen, University of Duisburg-Essen, Essen, Germany

Christel Depienne
Institute of Human Genetics, University Hospital Essen, University of Duisburg-Essen, Essen, Germany
christel.depienne@uni-due.de