

# 30 years of repeat expansion disorders: What have we learned and what are the remaining challenges?

Christel Depienne<sup>1,2,\*</sup> and Jean-Louis Mandel<sup>3,4,5,6,7,\*</sup>

Tandem repeats represent one of the most abundant class of variations in human genomes, which are polymorphic by nature and become highly unstable in a length-dependent manner. The expansion of repeat length across generations is a well-established process that results in human disorders mainly affecting the central nervous system. At least 50 disorders associated with expansion loci have been described to date, with half recognized only in the last ten years, as prior methodological difficulties limited their identification. These limitations still apply to the current widely used molecular diagnostic methods (exome or gene panels) and thus result in missed diagnosis detrimental to affected individuals and their families, especially for disorders that are very rare and/or clinically not recognizable. Most of these disorders have been identified through family-driven approaches and many others likely remain to be identified. The recent development of long-read technologies provides a unique opportunity to systematically investigate the contribution of tandem repeats and repeat expansions to the genetic architecture of human disorders. In this review, we summarize the current and most recent knowledge about the genetics of repeat expansion disorders and the diversity of their pathophysiological mechanisms and outline the perspectives of developing personalized treatments in the future.

## Normal and pathological tandem repeats in human genomes

The human genome contains more than one million annotated tandem repeats (TRs).<sup>1,2</sup> Due to their repetitive nature, TRs have the highest mutational rate in the genome and are typically polymorphic and multiallelic, with the longest alleles being the most unstable. TRs are mainly scattered in non-coding regions of the genome although repetitions of triplets, usually limited in size, can also be found in coding parts of genes.<sup>1</sup> Coding TRs may occasionally become hotspots for harmful frameshift mutations.<sup>3</sup>

TRs are usually divided into microsatellites (1–9 bp repeats; or short tandem repeats, STRs) and minisatellites (10–99 bp repeats), which together form the variable number of tandem repeats (VNTRs), and satellites ( $\geq 100$  bp repeats), which mainly constitute heterochromatin and centromeres. STRs have widely been used as markers in genetic linkage analyses and forensic studies, but they are still an enigmatic and understudied class of variation. STRs present in human genomes are recent in terms of evolution and are either human specific or conserved only in very close primate species.<sup>4</sup> Most STRs have arisen from other repeated elements, such as long interspersed nuclear element (LINE) and short interspersed nuclear element (SINE), including Alu elements.<sup>5</sup> The evolutionary constraints applying to these dynamic, rapidly evolving variations and how TRs have contributed to shape human genomes are still largely unknown although

recent evidence suggests that they could play an important role in the regulation of gene expression.<sup>6,7</sup>

The expansion of TR length across generations is a well-characterized process that results in at least 50 known disorders (Tables 1, 2, and S1). The first expansions, both identified on chromosome (chr) X in 1991, were CGG repeats in the 5' untranslated region (UTR) of *FMR1* (MIM: 309550) and CAG repeats in the first exon of *AR* (MIM: 313700) as the underlying causes of fragile X syndrome<sup>8,9</sup> (FXS [MIM: 300624]) and spinal bulbar muscular atrophy (SBMA [MIM: 313200]),<sup>10</sup> respectively (Figure 1). In the next 10 years, a wave of repeat expansion discovery followed, revealing the basis of more than 25 hereditary disorders, most of which primarily affect the central nervous system.

As illustrated by the initial discoveries, two main types of STR expansions exist: expansions affecting coding regions, mainly leading to abnormally long stretches of polyglutamine (polyQ, mostly encoded by CAG codons) or polyaniline (polyA, encoded by GCN codons) within proteins (Table 1), and expansions altering non-coding regions of genes (Figure 2).<sup>6,7</sup> CAG triplet expansions, including those found in Huntington disease (HD [MIM: 143100]), spinocerebellar ataxias type 1 (SCA1 [MIM: 164400]), SCA2 (MIM: 183090), SCA3 (Machado-Joseph disease [MIM: 109150]), SCA6 (MIM: 183086), SCA7 (MIM: 164500), SCA17 (MIM: 607136), and dentatorubral-pallidoluysian atrophy (DRPLA [MIM: 125370]) are typically associated with neuronal intranuclear protein inclusions that contain

<sup>1</sup>Institute of Human Genetics, University Hospital Essen, University of Duisburg-Essen, Essen, Germany; <sup>2</sup>Institut du Cerveau et de la Moelle épinière (ICM), Sorbonne Université, UMR S 1127, Inserm U1127, CNRS UMR 7225, 75013 Paris, France; <sup>3</sup>Institut de Génétique et de Biologie Moléculaire et Cellulaire, Illkirch 67400, France; <sup>4</sup>Centre National de la Recherche Scientifique, UMR 7104, Illkirch 67400, France; <sup>5</sup>Institut National de la Santé et de la Recherche Médicale, U 1258, Illkirch 67400, France; <sup>6</sup>Université de Strasbourg, Illkirch 67400, France; <sup>7</sup>USIAS University of Strasbourg Institute of Advanced study, 67000 Strasbourg, France

\*Correspondence: [christel.depienne@uni-due.de](mailto:christel.depienne@uni-due.de) (C.D.), [jmandel@igbmc.fr](mailto:jmandel@igbmc.fr) (J.-L.M.)  
<https://doi.org/10.1016/j.ajhg.2021.03.011>

© 2021 American Society of Human Genetics.

**Table 1. List of coding repeat expansion disorders, sorted according to their year of discovery**

ID year	Disorder <sup>a</sup>	Inheritance <sup>b</sup>	Chr	Gene	Repeat motif	Normal repeat range	Pathological repeat range	Pathological mechanism(s) <sup>c</sup>	ID method <sup>d</sup>	Ref
1991	SBMA	XLR	Xq12	AR	CAG	9–36	≥ 38–68	polyQ, RNA?	L, CGA	La Spada et al. <sup>10</sup>
1993	HD	AD	4p16.3	HTT	CAG	6–35	≥ 36–250	polyQ, RNA?	L, ExScr	The Huntington's Disease Collaborative Research Group <sup>11</sup>
1993	SCA1	AD	6p22	ATXN1	CAG	6–38	≥ 39–88	polyQ, RNA?	L, ExScr	Orr et al. <sup>12</sup>
1994	DRPLA	AD	12p13.31	ATN1	CAG	3–35	≥ 48–93	polyQ, RNA?	L, ExScr	Koide et al., <sup>13</sup> Nagafuchi et al. <sup>14</sup>
1994	SCA3	AD	14q32	ATXN3	CAG	12–44	≥ 55–87	polyQ, RNA (foci)	L, cl	Kawaguchi et al. <sup>15</sup>
1996	SCA2	AD	12q24	ATXN2	CAG	13–31	≥ 32–500	polyQ, RNA?	L, cl	Pulst et al. <sup>16</sup>
1996	SCA7	AD	3p21	ATXN7	CAG	4–33	≥ 37–460	polyQ, RNA?	L, cl	Lindblad et al. <sup>17</sup>
1996	SPD1	AD	2q31.1	HOXD13	GCG	15	24	polyA (DN)	L, CGA	Akarsu et al. <sup>18</sup>
1997	SCA6	AD	19p13	CACNA1A	CAG	4–18	≥ 20–33	polyQ, RNA?	L, ExScr	Zhuchenko et al. <sup>19</sup>
1997	BCCD	AD	6p21	RUNX2	GCN	17	27	polyA (DN/LoF?)	L, CGA	Mundlos et al. <sup>20</sup>
1998	OPMD	AD	14q11.2	PABPN1	GCG	6–10	≥ 12–17	polyA (DN/LoF?)	L, ExScr	Brais et al. <sup>21</sup>
1999	SCA17	AD	6q27	TBP	CAG (or CAG/CAA)	25–40	≥ 43–66	polyQ, RNA?	L, CGA	Koide et al. <sup>22</sup>
2000, 2002	HFGS	AD	7p15.2	HOXA13	tract 1: GCN	14	22	polyA (DN/LoF?)	L, CGA	Goodman et al., <sup>23</sup> Utsch et al. <sup>24</sup>
				tract 2: GCN	12	18				
				tract 3: GCN	18	24–30				
2001	HDL2	AD	16q24.2	JPH3	CAG	6–28	≥ 41–58	polyQ, RNA (foci)	L, RED	Margolis et al. <sup>25</sup>
2001	BPES	AD	3q23	FOXL2	GCN	14	19–24	polyA (LoF)	L, CGA	De Baere et al. <sup>26</sup>
2001	HPES	AD	13q32	ZIC2	GCN	15	25	polyA (LoF)	L, CGA	Brown et al. <sup>27</sup>
2002	EIEE1	XLR	Xp21.3	ARX	tract 1: GCN	16	23	polyA (DN)	L, CGA	Strømme et al. <sup>28</sup>
				tract 2: GCN	12	20				
2002	MRGH	XLR	Xq26.3	SOX3	GCN	15	26	polyA (LoF)	L, CGA	Laumonnier et al. <sup>29</sup>
2003	CCHS	AD	4p13	PHOX2B	GCN	20	25–29	polyA (DN)	L, CGA	Amiel et al. <sup>30</sup>

ID, identification; chr, chromosome; ref, reference(s)

<sup>a</sup>BCCD, brachydactyly and cleidocranial dysplasia; BPES, blepharophimosis, ptosis and epicanthus inversus; CCHS, congenital central hypoventilation syndrome; DRPLA, dentatorubral-pallidoluysian atrophy; EIEE1, early infantile epileptic encephalopathy type 1; HD, Huntington disease; HDL2, Huntington disease-like 2; HFGS, hand-foot-genital syndrome; HPES, holoprosencephaly type 5; MRGH, mental retardation with isolated growth hormone deficiency; OPMD, oculopharyngeal muscular dystrophy; SBMA, spinal and bulbar muscular atrophy; SPD1, synpolydactyly type 1; SCA, spinocerebellar ataxia<sup>b</sup>AD, autosomal dominant; XLR, X-linked, recessive<sup>c</sup>polyA (DN), polyAlanine tract associated with dominant-negative effect; polyA (LoF), polyAlanine tract leading to a loss of function of the protein; polyQ, poly-Glutamine protein toxicity; polyG, polyGlycine peptide toxicity; RNA, RNA toxicity<sup>d</sup>cl, cloning; CGA, candidate gene analysis; ExScr, expansion screening; L, linkage; RED, repeat expansion detection

the protein (or part of it) exhibiting the expanded polyQ stretch. These inclusions result from the propensity of these abnormal polyQ chains to form  $\beta$  sheet structures leading to intermolecular cross-beta oligomerization and to the formation of insoluble fibrillar aggregates, a mechanism usually described as a toxic gain-of-function at the protein level.<sup>62</sup> PolyA peptides can also form  $\beta$  sheet structures, but polyA expansions, usually much more limited in size compared to their polyQ counterpart, have variable effects and can lead to either gain- and/or loss-of-function depending on the gene and protein where they occur.<sup>63</sup>

Non-coding repeat expansions are even more diverse and their impact strongly depends on the type, length, and location of the repeat motif within genes (Table 2).<sup>61</sup> They can occur in either the 5' UTRs, introns, or 3' UTRs of genes (Figure 2). Repeat expansions located in the 5' UTRs, promoter, or other regulatory regions are usually GC rich, like triplet (CGG) expansions in *FMR1*, dodecamer (CCCCGCCCGCG) expansions in *CSTB* (MIM: 601145),<sup>34</sup> and hexanucleotide (GGGGCC) expansions in *C9ORF72* (MIM: 614260).<sup>43,44</sup> Large, usually GC-rich expansions overlapping 5' regulatory regions are

**Table 2. List of noncoding repeat expansion disorders involving motifs up to 12 nucleotides, sorted according to their year of discovery**

ID year	Disorder <sup>a</sup>	Inheritance <sup>b</sup>	Chr	Gene	Location	Repeat motif	pathogenic motif when different	Normal repeat range	Pathological repeat range	Pathological mechanism(s) <sup>c</sup>	ID method <sup>d</sup>	Ref
1991	FXS	XLD	Xq27.3	<i>FMR1</i>	5' UTR	CGG	–	5–50	>200	Me/GS	K, FS, L, Me, cl	Oberlé et al. <sup>8</sup> Verkerk et al. <sup>9</sup>
1992	DM1	AD	19q13.32	<i>DMPK</i>	3' UTR	CTG	–	5–37	>50–10,000	RNA (foci, PS), RAN (polyQ)	L, cl	Mahadevan et al. <sup>31</sup>
1993	FRAXE	XLR	Xq28	<i>AFF2</i>	5' UTR	CCG	–	4–39	≥200–900	Me/GS	K, L	Knight et al. <sup>32</sup>
1996	FRDA	AR	9q21.11	<i>FXN</i>	intron	GAA	–	5–34	≥66–1300	EGS	L, ExScr	Campuzano et al. <sup>33</sup>
1997	EPM1	AR	21q22.3	<i>CSTB</i>	promoter / 5' UTR	CCCCGCCCGCG	–	2–3	≥30–75	EGS	L, cl, exp	Lalioi et al. <sup>34</sup>
1998	FXPOI	XL	Xq27.3	<i>FMR1</i>	5' UTR	CGG	–	5–50	55–200	RNA (foci), RAN	Fam, ExScr	Conway et al. <sup>35</sup>
1999	SCA8	AD	13q21	<i>ATXN8 / ATXN8OS</i>	3' UTR	CAG/CTG	–	15–50	>74–250	RNA (foci), RAN	L, cl	Koob et al. <sup>36</sup>
1999	SCA12	AD	5q31	<i>PPP2R2B</i>	5' UTR	CAG	–	4–32	≥43–78	RAN (polyG)?	L, RED	Holmes et al. <sup>37</sup>
2000	SCA10	AD	22q13	<i>ATXN10</i>	intron	ATTCT / ATTGT	–	10–32	>280–4,500	RNA (foci)	L, ExScr	Matsuura et al. <sup>38</sup>
2001	DM2	AD	3q21.3	<i>CNBP</i>	intron	CCTG / CAGG	–	11–30	>50–11,000	RNA (foci)	L, ExScr	Liquori et al. <sup>39</sup>
2001	FXTAS	XLR	Xq27.3	<i>FMR1</i>	5' UTR	CGG	–	5–50	55–200	RNA (foci), RAN (polyG)	Fam	Hagerman et al. <sup>40</sup>
2009	SCA31	AD	16q22	<i>BEAN1 / TK2</i>	intron	TAAAA	TGGAA (+TAGAA)	variable	≥110–760	RNA (foci, PS), RAN?	L	Sato et al. <sup>41</sup>
2011	SCA36	AD	20p13	<i>NOP56</i>	intron	GGCCTG	–	5–14	≥650–2,500	RNA (foci)	L, ExScr	Kobayashi et al. <sup>42</sup>
2011	FTD/ALS	AD	9p21.2	<i>C9ORF72</i>	5' UTR / intron	GGGGCC	–	3–25	>30	RNA (foci), RAN (polyG)	L, WGS	Renton et al. <sup>43</sup> DeJesus-Hernandez et al. <sup>44</sup>
2015	FECD3	AD	18q21.2	<i>TCF4</i>	intron	CTG	–	5–31	>50	unknown	L, GWAS	Mootha et al. <sup>45,46</sup>
2017	XDPA	XLR	Xq13.1	<i>TAF1</i>	intron (retrotransposon)	CCCTCT	–	none	30–55	unknown	L, exp	Bragg et al. <sup>47</sup>
2017	SCA37	AD	1p32	<i>DAB1</i>	intron	ATTTT	ATTTC	7-400 (ATTTT)	≥31–75 ATTTC	RNA	L, ExScr	Seixas et al. <sup>48</sup>
2018	FAME1	AD	8q24	<i>SAMD12</i>	intron	ATTTT	ATTTC	7-exp (ATTTT)	≥440–3,680 ATTTC	RNA?	L, ExScr	Ishiura et al. <sup>49</sup>
2018	FAME6	AD	16p12.1	<i>TNRC6A</i>	intron	ATTTT	ATTTC	N/A	N/A	RNA?	TRhist	Ishiura et al. <sup>49</sup>
2018	FAME7	AD	4q32.1	<i>RAPGEF2</i>	intron	ATTTT	ATTTC	N/A	N/A	RNA?	TRhist	Ishiura et al. <sup>49</sup>
2019	BSS	AR	16p12.3	<i>XYLT1</i>	promoter	CGG	–	9–20	120–800	Me/GS	ADO, Me	LaCroix et al. <sup>50</sup>

(Continued on next page)

**Table 2. Continued**

ID year	Disorder <sup>a</sup>	Inheritance <sup>b</sup>	Chr	Gene	Location	Repeat motif	pathogenic motif when different	Normal repeat range	Pathological repeat range	Pathological mechanism(s) <sup>c</sup>	ID method <sup>d</sup>	Ref
2019	CANVAS	AR	4p14	<i>RFC1</i>	intron	AAAAG / AAAGG / AAGAG / AGAGG	AAGGG	variable	≥ 400–2,000	unknown	L, WGS	Cortese et al. <sup>51</sup> Rafehi et al. <sup>52</sup>
2019	GDPAG	AR	2q32.2	<i>GLS</i>	5' UTR	GCA	–	8–16	≥ 680–1,400	EGS	CGA	van Kuilenburg et al. <sup>53</sup>
2019	NIID	AD	1q21.2	<i>NOTCH2NLC</i>	5' UTR / exon 1	CGG	–	7–60	≥ 61–500	Me/GS, RAN (polyG)?	TRhist, L, WGS	Ishiura et al. <sup>54</sup> Sone et al. <sup>55</sup> Tian et al. <sup>56</sup>
2019	OPDM1	AD	8q22.3	<i>LRP12</i>	5' UTR	CGG	–	13–45	90–130	RAN (polyG)?	TRhist	Ishiura et al. <sup>54</sup>
2019	OPML1	AD	10q22.3	<i>NUTM2B- AS1</i>	noncoding RNA gene	CGG/CCG	–	3–16	40–60	RAN (polyG)?	TRhist	Ishiura et al. <sup>54</sup>
2019	FAME4	AD	3q27.1	<i>YEATS2</i>	intron	ATTTT	ATTTC	7–400 (ATTTT)	N/A	RNA?	L, ExScr	Yeetong et al. <sup>57</sup>
2019	FAME2	AD	2q11.2	<i>STARD7</i>	intron	ATTTT	ATTTC	9–20	≥ 661–735	RNA?	L, ExScr	Corbett et al. <sup>58</sup>
2019	FAME3	AD	5p15.2	<i>MARCHF6</i>	intron	ATTTT	ATTTC	10–30	≥ 660–2,800	RNA?	L, ExScr	Florian et al. <sup>59</sup>
2020	OPDM2	AD	19p13.12	<i>GIPC1</i>	5' UTR	CGG	–	12–32	≥ 97–120	unknown	L, WGS, LRS	Deng et al. <sup>60</sup>

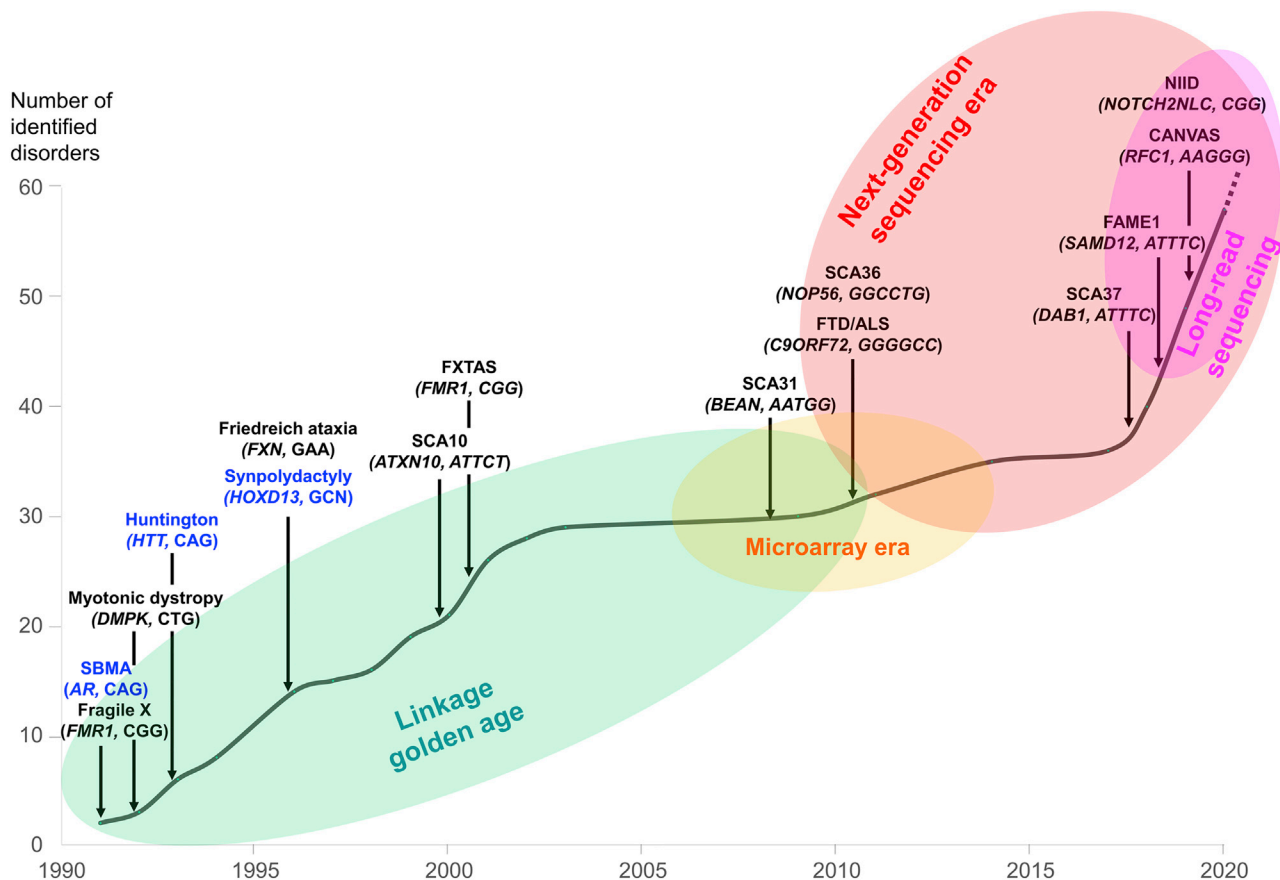
ID, identification; chr, chromosome; N/A, not available; ref, reference(s); UTR, untranslated region

<sup>a</sup>BSS, Baratela-Scott syndrome; CANVAS, cerebellar ataxia, neuropathy and vestibular areflexia syndrome; DM1, myotonic dystrophy type 1; DM2, myotonic dystrophy type 2; EPM1, progressive myoclonus epilepsy type 1 (Unverricht-Lundborg disease); FAME, familial adult myoclonic epilepsy; FECD3, Fuchs endothelial corneal dystrophy type 3; FRAXE, fragile XE syndrome; FRDA, Friedreich ataxia; FTD/ALS, frontotemporal dementia/amyotrophic lateral sclerosis; FXPOI, Fragile X-associated premature ovarian infertility; 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, oculopharyngodistal myopathy type 1; OPDM2, oculopharyngodistal myopathy type 2; OPML1, oculopharyngeal myopathy with leukoencephalopathy type 1; SCA, spinocerebellar ataxia; XDP, X-linked dystonia parkinsonism

<sup>b</sup>AD, autosomal dominant; XL, X-linked; XLR, X-linked, recessive; XLD, X-linked, dominant

<sup>c</sup>EGS, epigenetic gene silencing; Me/GS, gene silencing associated with hypermethylation; PS, protein sequestration; RNA, RNA toxicity; RAN, repeat-associated non-AUG translation

<sup>d</sup>ADO, allele drop-out; cl, cloning; CGA, candidate gene analysis; exp, expression; ExScr, expansion screening; Fam, clinical analyses in fragile X-affected families; FS, fragile site; GWAS, genome-wide association study; K, karyotyping; L, linkage; LRS, long-read sequencing; RED, repeat expansion detection; WGS, whole-genome sequencing



**Figure 1. Timeline of repeat expansion discovery in human disorders**

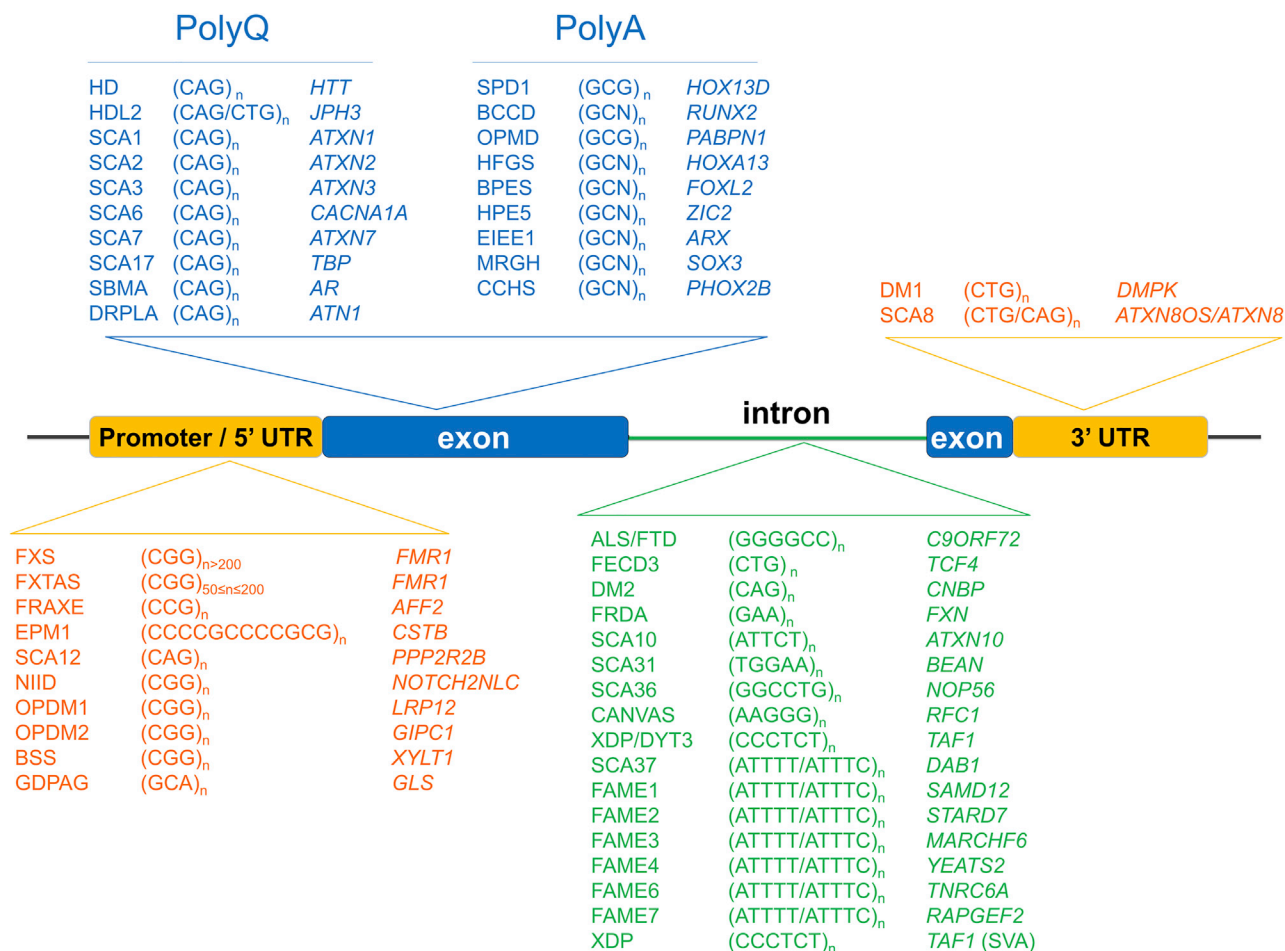
During the first twenty years, linkage analyses were the gold standard allowing to map the genomic region containing repeat expansions. The development of next-generation sequencing, and more particularly long-read sequencing, has marked a new milestone, making their identification at a genome-wide scale possible, and has allowed a new wave of repeat expansion discovery.

frequently—but not always—associated with hypermethylation of the corresponding allele and gene silencing, as exemplified by full mutations (>200 repeats) causing FXS.<sup>64</sup> Expanded repeats falling into introns correspond to the majority of recently described expansion disorders and involve motifs of different lengths (mainly tri, tetra, penta, or hexa-nucleotides) and GC content. Only two repeat expansions in 3' UTRs have been described to date, in myotonic dystrophy 1 (DM1 [MIM: 160900]; CTG expansion in the 3' UTR of *DMPK* [MIM: 605377]) and SCA8 (MIM: 608768) (CTG/CAG expansion in the 3' UTR of *ATXN8OS* [MIM: 603680]/*ATXN8* [MIM: 613289]). Remarkably, non-coding, and more particularly intronic repeat expansions act through a variety of different mechanisms, most of which are associated with dominant inheritance although a few conditions are recessive, generally resulting from the loss-of-function of the gene as a result of the expansion.

## Anticipation

A characteristic of dominant repeat expansion disorders is anticipation, a process by which the clinical manifestation

appears at an earlier age and/or becomes more severe as the disorder is transmitted from one generation to the next. Strong correlations between the size of the expansion and the age at onset and/or the severity of the disorder have been established for many repeat expansion disorders, including all polyQ as well as many of the non-coding expansions. The larger the expansion is, the earlier the age at onset and the more severe the phenotype becomes. Moreover, expanded repeats, and especially trinucleotide repeats, tend to dynamically expand further during meiosis. Therefore, the offsprings of an affected carrier receiving the pathogenic allele are at higher risk of presenting a more severe or early form of the disorder. In extreme cases, including DM1, HD, and certain types of SCAs, the disorder can occur in infancy or childhood and be fatal in a few years whereas the transmitting parent has a milder adult-onset form of the same disorder.<sup>65</sup> The risk of expansion during meiosis generally increases with the number of repeats. Contractions of repeat numbers during meiosis have also been reported but are relatively unfrequent compared to expansions.<sup>66</sup> The instability of repeat expansions during meiosis and transmission of larger alleles to offsprings is influenced by multiple factors including the size of the repeat, its structure, and the sex



**Figure 2. Location of repeat expansions within genes**

Repeat expansions may affect coding regions (mainly encoding polyglutamine or polyanaline tracts) or noncoding regions of genes. Noncoding repeat expansions are mainly located in promoters or 5' untranslated regions (UTRs) or introns. Noncoding repeats in 5' regions are often GC-rich and their expansion usually affects gene expression through methylation of other epigenetic processes. Intronic expansions and expansions affecting 3' UTRs more often lead to RNA toxicity or polypeptide synthesis via repeat-associated non-AUG (RAN) translation and subsequent formation of cellular/nuclear aggregates. BCCD, brachydactyly and cleidocranial dysplasia; BSS, Bartela-Scott syndrome; CANVAS, cerebellar ataxia, neuropathy and vestibular areflexia syndrome; CCHS, congenital central hypoventilation syndrome; DM1, myotonic dystrophy type 1; DM2, myotonic dystrophy type 2; DRPLA, dentatorubral-pallidoluysian atrophy; EIEE1, early infantile epileptic encephalopathy type 1; EPM1, progressive myoclonus epilepsy type 1 (Unverricht-Lundborg disease); FAME, familial adult myoclonic epilepsy; FECD3, Fuchs endothelial corneal dystrophy type 3; FRAXE, fragile XE syndrome; FRDA, Friedreich ataxia; FTD/ALS, frontotemporal dementia / amyotrophic lateral sclerosis; FXS, fragile X syndrome; FXTAS, fragile X-associated tremor ataxia syndrome; GDPAG, global developmental delay, progressive ataxia, and elevated glutamine; HD, Huntington disease; HDL2, Huntington disease-like 2; HFGS, hand-foot-genital syndrome; HPE5, holoprosencephaly type 5; MRGH, mental retardation with isolated growth hormone deficiency; OPMD, oculopharyngeal muscular dystrophy; NIID, neuronal intranuclear inclusion disease; OPDM1, oculopharyngodistal myopathy type 1; OPDM2, oculopharyngeal muscular dystrophy type 2; OPML1, oculopharyngeal myopathy with leukoencephalopathy type 1; SBMA, spinal and bulbar muscular atrophy; SPD1, synpolydactyly type 1; SCA, spinocerebellar ataxia; SVA, SINE-VNTR-Alu retrotransposon; XDP, X-linked dystonia parkinsonism.

and age of the transmitting parent.<sup>66,67</sup> Interrupted repeats (for instance AGG interruptions in CGG repeats, or CAA interruptions in CAG repeats) are much more stable than pure ones. Contractions may be more frequent in the male germline, accounting for the exclusive maternal transmission of congenital DM1, or of full *FMR1* expansions from premutated mothers, but not from premutated fathers. Males with FXS expansions have spermatozoa with premutation, suggesting that very long repeats may be subject to negative selection during spermatogenesis, by affecting completion of replication.<sup>68</sup> Conversely, CAG

repeat expansions associated with juvenile HD forms arise only from paternal transmission, and somatic variability can be observed in sperm of HD expansion carriers.

### Somatic instability

Repeat expansions are also mitotically unstable, each cell division being associated with a risk of error during DNA replication. This process has been initially attributed to the slippage of DNA polymerase while replicating



repetitive DNA sequences and is further amplified by the ability of some repeats to form stable folded structures such as hairpins, stem-loops, cruciform, and/or tetraplexes that favor DNA polymerase pausing.<sup>68–70</sup> However, the observations of expansions and contractions in non-dividing cells have indicated that other mechanisms are at play, and there is increasing evidence from mouse models, and very recently from affected individuals, that somatic stability or instability of repeat expansions largely depends on DNA mismatch repair processes and may impact progression of the disease.<sup>68,69,71</sup> Many repeat expansions are thus associated with somatic mosaicism, which degree is usually correlated with expansion size and also age, but can show large variation in different tissues, with a pattern that differs in the various expansion disorders.<sup>72–75</sup> Interruptions by alternative repeat motifs tend to stabilize expansions, making them less prone to alterations during DNA replication or repair.<sup>76–78</sup> Moreover, interrupted expansions are associated with milder and/or different phenotypes compared to uninterrupted expansions, as shown for DM1, SCA2, or SCA10 (MIM: 603516).<sup>79,80</sup>

## Founder effects

Many repeat expansions, especially larger ones, are associated with founder effects, where the expanded allele is associated with a single or predominant haplotype, and result in geographical clustering of the corresponding disorder or highly variable prevalences depending on the population. For instance, DM1 and Friedreich ataxia (FRDA [MIM: 229300]) are, respectively, very rare in Africa and Japan.<sup>81,82</sup> This suggests that unexpanded alleles, contrary to their expanded counterparts, are relatively stable. Furthermore, risk haplotypes are in linkage disequilibrium with the large normal alleles at the cognate repeat loci, suggesting a multistep evolution process where an initial historical event is a mutation that generates a large normal allele (protomutation) that then acts as a reservoir for further gradual expansions that will finally result in pathological alleles.<sup>82</sup> Another possibility for a rare founder event is the loss of stabilizing interruption by a variant motif in the target normal repeat (such as a loss of AGG interruptions in the CGG repeat at the fragile X locus).<sup>83</sup> Examples of repeat expansions for which founder effects have been demonstrated include DM1, FRDA, FXS, SCA3, SCA10, *C9ORF72*-related frontotemporal dementia and/or amyotrophic lateral sclerosis (FTD/ALS [MIM: 105550]), and myoclonic epilepsy of Unverricht and Lundborg (EPM1 [MIM: 254800]).<sup>84–87</sup>

## Identification strategies: Remaining challenges

Although repeat expansions were identified as an important source of human disorders in the years 1991–2003, no additional expansion disorder was discovered in the next 5 years (Figure 1). In 2011, the identification of domi-

nant hexanucleotide expansions in *C9ORF72*<sup>43,44</sup> at a locus mapped by linkage in multiple large families in 2006 revived some interest in the search for repeat expansion diseases. But attempts to identify repeat expansions remained limited by exome sequencing, which was the prevalent disease gene identification strategy.

Indeed, owing to their repetitive nature and their abundance in the human genome, TRs are very difficult to study at a genome-wide scale and have been for a long time masked by bioinformatic tools such as RepeatMasker. Short-read encompassing repeats typically map to multiple genomic regions, are clipped off, or are discarded. Furthermore, there is a computational challenge in accurately calling genotypes of polymorphic alleles composed of multiple copies of the same repeats. Detection of TRs is thus generally absent from standard NGS pipelines and requires specific tools in addition to standard SNP, indel, and structural/copy number variant pipelines. Bioinformatics tools specifically assessing repeat numbers from genome (or exome) data have been developed, including LobSTR,<sup>88</sup> HipSTR,<sup>89</sup> TREDPARSE,<sup>90</sup> ExpansionHunter<sup>91</sup> STRetch,<sup>92</sup> GangSTR,<sup>93</sup> and exSTRa.<sup>94</sup> However, until recently, most of these tools were only able to call genotypes at specific loci for a given motif. This has recently changed by the development of TRhist<sup>95</sup> and ExpansionHunter De-Novo,<sup>96</sup> which assess the presence of repeat expansions at a genome-wide scale. Nevertheless, the detection of expansions without linkage data or prior hypotheses on the expanded motif remains a challenge due to the huge number of possible motifs, the abundance of TRs in the human genome, and the difficulty in clearly distinguishing pathogenic expansions from normal polymorphic alleles based on short-read sequences.<sup>90</sup> TR expansions thus likely constitute an enormous unexplored reservoir of pathogenic variations. A recent study, which investigated tandem repeats in 17,231 individuals with autism spectrum disorders, suggested that repeat expansions at more than 2,500 loci may represent a collective contribution of 2.6% to autism risk, but the statistic analysis could not distinguish between potential highly penetrant expansions or those associated as polygenic low penetrance risk factors.<sup>97</sup>

An alternative way to detect TR expansions more accurately is to use long-read technologies, including Oxford Nanopore Technology (ONT) and Pacific Biosciences (PacBio) Single Molecule Real-Time (SMRT) sequencing.<sup>98</sup> These technologies allow researchers to obtain single read >10 kb in size, sometimes reaching up to 1–2 Mb, but specific DNA extractions from fresh blood samples or living cells are required to obtain good results. Repeat expansions are detected by pipeline calling structural variants although specific tools, like NanoSatellite<sup>99</sup> and tandem-genotypes,<sup>100</sup> have been developed to specifically study repeats present in long-read data. Remarkably, these technologies detect thousands of structural variants per individual that predominantly correspond to repeated regions. Because databases such as gnomAD are based on

short-read sequencing data, most of the detected structural variants have not yet been described in large control population datasets and filtering of pathogenic variants amid this mass of information is not an easy task.

## Overview of recently identified repeat expansion disorders

Despite the technical and computational difficulties, the last few years have clearly witnessed a second wave of repeat expansion discovery. While no expansions have been identified between 2003 and 2008, the genetic basis of more than 20 repeat expansion disorders have been unraveled since 2009, with 17 identified in the last 3 years (Figure 1 and Tables 1 and 2). In almost all studies, the identification of pathogenic expansions has taken many years and was only possible because of a large number of families that allowed reducing the candidate linkage intervals to regions containing only one or very few genes. This suggests that many more rarer expansion disorders remain to be identified. Comprehensive reviews have already covered many aspects of repeat expansion disorders identified before 2017.<sup>61,101</sup> We review in the following sections the clinical and genetic characteristics of novel disorders and recent advances made in the field since 2017.

## Dominant expansions

### Spinocerebellar ataxia 37 (SCA37)

SCAs are a clinically and genetically heterogeneous group of autosomal-dominant disorders, characterized by gait imbalance as the result of the progressive loss of cerebellar neurons. SCA37 (MIM: 615945) was first mapped to 1p32 in a large Spanish family, and the interval was confirmed and further refined to 2.8 Mb in three Portuguese families.<sup>48</sup> The systematic search for repeat expansions in this interval led to the identification of an intronic expansion composed of ATTTT repeats with an insertion of ATTTC motifs in *DAB1* (MIM: 603448).<sup>48</sup> Pathogenic expansions all contain the ATTTC insertion, ranging from 31 to 75 repeats, flanked by variable stretches of ATTTT repeats, while larger pure ATTTT expansions (up to 400 repeats) are present in healthy individuals.<sup>48</sup> Furthermore, an inverse correlation exists between the number of ATTTC repeats detected in the blood and the age at disease onset.<sup>48</sup> *DAB1* encodes a downstream effector of the reelin signaling pathway, contributing to the correct positioning of neurons in the developing brain, with highest expression in the cerebellum. A study on post-mortem brains of SCA37 expansion carriers revealed that expansions increase *DAB1* expression and trigger alternative splicing events favoring the inclusion of two exons absent from isoforms normally present in the brain.<sup>102</sup> Overexpression of ATTTC repeats led to RNA foci formation *in vitro*, and injection of ATTTC repeats *in vivo* during zebrafish development was associated with malformations and lethality,

suggesting a toxic effect of RNA molecules containing AUUUC repeats.<sup>48</sup>

### Familial adult myoclonic epilepsy (FAME)

FAME is an autosomal-dominant disorder characterized by myoclonus (cortical myoclonic tremor) and epilepsy.<sup>103,104</sup> This familial condition was first described in Japan as benign adult familial myoclonic epilepsy (BAFME) and was independently referred to as autosomal-dominant cortical myoclonus and epilepsy (ADCME) and familial cortical myoclonic tremor with epilepsy (FCMTE).<sup>104</sup> Myoclonic tremor is usually the first symptom to appear at a variable age, ranging from 10 to 60 years, with an average of 30 years.<sup>104</sup> At least half of affected individuals develop epilepsy, mainly generalized tonic-clonic seizures, concomitantly to tremor or later in life.<sup>104</sup> Myoclonus and seizures are exacerbated by photic stimulation, alcohol, sleep deprivation, or emotional stress. At an advanced age, tremors tend to worsen and be associated with gait instability and difficulties to walk.

Four loci on chromosomes 2, 3, 5, and 8 had been described between 1999 and 2013,<sup>103</sup> but the genetic cause of FAME had remained elusive for 20 years despite extensive sequencing of the candidate intervals. In 2018, the candidate interval on chr 8 was reduced to a single gene, allowing the identification of intronic repeat expansions in *SAMD12* (MIM: 618073) as the cause of BAFME (FAME1 [MIM: 601068]). Similar to SCA37 expansions, *SAMD12* expansions are composed of both ATTTT and ATTTC motifs.<sup>49</sup> In one family and one sporadic case without *SAMD12* expansion, the same expanded repeats were detected in *TNRC6A* (MIM: 610739) and *RAPGEF2* (MIM: 609530).<sup>49</sup> This discovery rapidly unlocked the other FAME loci. Similar ATTTT/ATTTC expansions were identified in *STARD7* (MIM: 616712) on chr 2,<sup>58</sup> *YEATS2* (MIM: 613373) on chr 3,<sup>57</sup> and *MARCHF6* (MIM: 613297) (formerly *MARCH6*) on chr 5.<sup>59</sup> In all six genes, FAME expansions occur at a polymorphic STR site initially composed of ATTTT repeats and contain a ATTTC insertion generally located at the 3' end of the expansion. As previously observed for SCA37, *SAMD12* expansions composed of pure ATTTT repeats exist in healthy individuals, confirming that inserted ATTTC repeats are also the pathogenic part of FAME expansions.<sup>49</sup> The size of *SAMD12* expansions varied from 2.2 to 18.4 kb (i.e., 440–3,680 repeats),<sup>49</sup> while *MARCHF6* expansions ranged from 3.3 to 14 kb (i.e., 660–2,800 repeats) on average.<sup>59</sup> However, sequencing and staining of single alleles at the *MARCHF6* locus using molecular combing revealed an important somatic mosaicism of the expansion, both in terms of length and structure in blood cells. Interestingly, micro-rearrangements occurring at the expansion site were observed in individuals with very large (>10 kb) expansions in 20% of the cells, suggesting that expansions are prone to chromosomal breakages and could constitute fragile sites.<sup>59</sup> The mean expansion size was inversely correlated with the age at epilepsy onset<sup>49,59</sup> and this effect was mainly driven



by the size of ATTTTC repeats.<sup>59</sup> However, other pathogenic motifs than ATTTTC repeats could exist, as suggested by the identification of *SAMD12* expansions composed of ATTTG instead of ATTTTC repeats in a large Chinese family affected by FAME.<sup>105</sup> *SAMD12* expansions have been reported mainly in families from Japan, China, India, and Sri Lanka,<sup>49,106</sup> while *STARD7* and *MARCHF6* expansions seem to be restricted to families from European ancestry.<sup>58,59</sup> This geographical distribution results from ancient founder effects originating from Asia or Europe, respectively, and may be useful for initial genetic testing prioritization.

Contrary to SCA37 expansions, which alter the expression and splicing of *DAB1*, a gene with an established role in cerebellum development and function, the six genes harboring FAME expansions have completely different functions and expression profiles. *MARCHF6* encodes a ubiquitous E3 ubiquitin ligase mediating the degradation of misfolded or damaged proteins in the endoplasmic reticulum. *RAPGEF2* encodes a neuron-specific GTPase activating the RAS signaling pathways in response to the activation of cell membrane receptors, such as dopaminergic receptors. *STARD7* encodes a ubiquitous protein involved in lipid transport and metabolism. *TNRC6A* encodes a subunit of a cytoplasmic ribonucleoprotein complex regulating mRNA silencing, stability, and translation, highly expressed in the cerebellum. *YEATS2* encodes a ubiquitous subunit of the ADA2A-containing histone acetyltransferase complex. The function of *SAMD12* is unknown. This strongly suggests that the ATTTTC repeats are pathogenic independently of the recipient gene and its function and that the gene is only a template for repeat expression. Reads filled with AUUUU/AUUUC repeats and RNA foci associated with abortive transcription following *SAMD12* expansions have been observed in post-mortem brains of expansion carriers from Japan, suggesting that RNA containing AUUUC repeats could form aggregates able to sequester specific RNA-binding proteins.<sup>49</sup> On the other hand, despite *MARCHF6* and *STARD7* ubiquitous expression, reads filled with AUUUU or AUUUC repeats could not be detected in lymphoblasts and fibroblasts of individuals with *MARCHF6* or *STARD7* expansions and expression and splicing of *MARCHF6* and *STARD7* were unaltered in these tissues.<sup>58,59</sup> These findings possibly indicate that abnormal transcription and accumulation of AUUUC repeats observed in *SAMD12* expansion carriers could be restricted to neuronal tissues or be absent in other FAME subtypes, and additional studies are required to determine whether expansions lead to RNA toxicity in a neuron-specific manner.

### Neuronal intranuclear inclusion disease (NIID)

NIID (MIM: 603472), also known as neuronal intranuclear hyaline inclusion disease or intranuclear inclusion body disease, is an autosomal-dominant, slowly progressive neurodegenerative disorder first described in 1968. As its name indicates, the diagnosis of this condition was

initially based on post-mortem neuropathological investigation showing eosinophilic intranuclear inclusions in neurons, but also in other cell types, such as glial cells, fibroblasts, and muscles. NIID can present as a wide and variable range of clinical manifestations, including pyramidal and extrapyramidal symptoms, cerebellar ataxia, cognitive decline, peripheral neuropathy, and autonomic dysfunction.<sup>107</sup> Furthermore, the age at onset is extremely variable, ranging from infancy to late adulthood, though most affected subjects show symptoms after the 3<sup>rd</sup> decade. Brain MRI shows characteristic white matter abnormalities including T2-weighted hyperintensity signals in the middle cerebellar peduncles and high-intensity signals in the corticomedullary junction on diffusion-weighted imaging. Skin biopsy combined with brain imaging is therefore useful for establishing an ante-mortem diagnosis.

White matter abnormalities observed in NIID are reminiscent of those occasionally observed in fragile X-associated tremor ataxia syndrome (FXTAS [MIM: 300623]), a disease caused by intermediate CGG expansions (55–200 repeats) in *FMR1*.<sup>54</sup> Taken together with the presence of neuronal intranuclear inclusions, these findings were suggestive of a repeat-expansion disorder. TRhist combined with SMRT sequencing identified CGG expansions mapping to the 5' UTR/first exon of *NOTCH2NLC* (MIM: 618025) (previously *NBP19*).<sup>54</sup> In two parallel studies, genome-wide linkage analysis first identified overlapping intervals on chromosomes 1p22.1–q21.3 and 1p13.3–23.1 and the same expansion in *NOTCH2NLC* was revealed by long-read sequencing.<sup>55,56</sup> Screening of this expansion in additional neurological disorders led to expand associated phenotypes to essential tremor (ETM6 [MIM: 618866]),<sup>108,109</sup> FTD and Alzheimer-like dementias,<sup>110</sup> and multiple system atrophy.<sup>111</sup> Many families with *NOTCH2NLC* expansions are from Japan and China, suggesting a founder effect in these populations. However, expansions can also occur *de novo* in sporadic cases<sup>112</sup> and can be present in individuals from other geographic origins although their frequency in Europe appears lower than in Asia.<sup>113</sup>

*NOTCH2NLC* is one of five genes (*NOTCH2* [MIM: 600275], *NOTCH2NLA* [MIM: 618023], *NOTCH2NLB* [MIM: 618024], *NOTCH2NLC*, and *NOTCH2NLR* [MIM: 618026]) sharing a high degree (>99%) of DNA homology in their 5' region. The *NOTCH2NL* copies are distributed as tandem duplicates on each side of the chr 1 centromere and are human specific. They have arisen from evolutionary recent duplications of *NOTCH2* that likely contributed significantly to human brain evolution, by increasing neurogenesis and cortical size.<sup>114,115</sup> *NOTCH2NL* genes are correctly annotated only in the hg38 version of the human reference genome and the presence of these multiple, almost identical copies complicates the molecular diagnosis of NIID, which is already challenging due to the GC-rich nature of these expansions. The number of pathological CGG repeats was estimated to range from 61 to

500 repeats whereas in control individuals this number varies from 6 to 60 and CGG repeats can be interrupted by AGG motifs.<sup>54,55,116</sup> CGG expansions associated with NIID are not consistently associated with DNA hypermethylation and they do not alter the expression of *NOTCH2NLC*, but anti-sense transcripts are specifically produced in affected individuals.<sup>54,55</sup> This suggests pathophysiological mechanisms possibly similar to FXTAS, with a probable gain of toxic function at the RNA level and the eventual existence of repeat-associated non-AUG (RAN) translation.

### Oculopharyngeal and oculopharyngodistal myopathies

In a family with four affected individuals presenting with oculopharyngeal myopathy, limb weakness, ataxia, ptosis, and white matter abnormalities similar to those seen in NIID (OPML1 [MIM: 618637]), Ishiura et al. used TRhist to identify CGG repeats in a long non-coding RNA gene, *LOC642361*, which overlaps on chr 10q22.3 with an anti-sense transcript, *NUTM2B-AS1* (MIM: 618639).<sup>54</sup> This expansion segregated with the disorder in the family and was highly unstable, but additional families with similar phenotypes related to this expansion are needed to confirm the pathogenicity of this expansion.

The same strategy also revealed a CGG expansion in the 5' UTR of *LRP12* (MIM: 618299) in a family with oculopharyngodistal myopathy (OPDM1 [MIM: 164310]).<sup>54</sup> Oculopharyngodistal myopathy is an adult-onset neuromuscular disorder characterized by progressive leg and arm weakness associated with external ophthalmoplegia, dysphagia, and ptosis. The CGG expansion in *LRP12* was further identified in eight families and 13 sporadic cases with OPDM or milder symptoms, including ptosis and extraocular and pharyngeal weakness.<sup>54</sup>

Independently, Deng et al. used a combination of whole-genome sequencing and long-read sequencing to identify CGG repeat expansions in the 5' UTR of *GIPC1* (MIM: 605072) in four families and three sporadic cases with OPDM2 (MIM: 618940) of Chinese origin.<sup>60</sup> This expansion was further confirmed by repeat-primed PCR in 15 additional families or sporadic cases of Chinese or Japanese origins.<sup>60</sup> Expansions increased *GIPC1* mRNA expression without altering protein levels. Like ATTTC expansions in FAME, CGG expansions seem to lead to OPDM irrespective of the gene where they occur.

### Recessive expansions

#### Cerebellar ataxia, neuropathy, and vestibular areflexia syndrome (CANVAS)

CANVAS (MIM: 614575) is an adult-onset, slowly progressive disorder characterized by cerebellar ataxia, sensory neuropathy, bilateral vestibulopathy, and autonomic dysfunction, first described in 1991 but recognized as a specific disease entity in 2011. The symptoms usually start around the 6<sup>th</sup> decade although the age at onset can vary

between 39 and 71 years. Furthermore, CANVAS is possibly a relatively common disorder although its exact prevalence remains to be determined. The description of affected siblings and additional pedigrees pointed to a recessive disorder.<sup>117</sup>

The identification of repeat expansions in CANVAS was possible through genome-wide linkage analysis in multiple small families, which identified a single locus on chromosome 4p14.<sup>51,52</sup> In this candidate interval, bi-allelic repeat expansions composed of AAGGG repeats were identified in intron 2 of *RFC1* (MIM: 102579), replacing the 11 AAAAG repeats present in the reference genome. Analysis of control individuals revealed a striking variability of repeat motifs and sizes at this locus. Non-expanded AAAAG repeats correspond to the ancestral and most frequent allele but nonpathogenic expanded AAAAG repeats are quite common in healthy individuals. Additional motifs present at this locus include AAAGG, AAGAG, and AGAGG.<sup>118</sup> The relevance of these motifs in human pathology is not yet fully understood, but the observation of an affected individual with bi-allelic ACAGG expansion suggests that other motifs than AAGGG could also be pathogenic.<sup>119</sup> Expansion size varied from 400 to 2,000 repeats (i.e., 2–10 kb) but no association between the number of AAGGG repeat unit and the age at onset was found.<sup>51</sup> AAGGG expansions are more frequent in the European population and associated with a founder effect,<sup>51</sup> although expanded alleles also exist in other populations.<sup>119</sup> Sensory neuropathy is present in nearly 100% of individuals with *RFC1* expansions, and a dry spasmodic cough is often the first manifestation, long before walking difficulty.<sup>120</sup> More rarely, bi-allelic *RFC1* expansions can be detected in individuals with multiple system atrophy.<sup>121</sup>

Remarkably, although AAGGG expansions are recessive, suggesting a loss of function of some kind, no detectable impact was observed on *RFC1* gene expression or splicing in peripheral tissues or post-mortem brain samples (including cortex and cerebellum).<sup>51</sup> In addition, no RNA molecule with expanded AAGGG or CCCUU repeats corresponding to *RFC1* pre-mRNA transcript and no RNA foci could be detected in any of the tissues examined.<sup>51</sup> *RFC1* encodes the large subunit of replication factor C, a ubiquitous DNA polymerase accessory protein required for DNA replication and DNA repair. However, the function of this complex is unaltered by expansions as fibroblasts from affected individuals showed a normal response to DNA damage.<sup>51</sup> Therefore, the mechanisms by which AAGGG expansions lead to CANVAS remains mysterious so far.

#### Recessive expansions causing gene silencing

Baratela-Scott syndrome (BSS) is a recessive skeletal dysplasia associated with developmental delay, short stature, and facial dysmorphism that clinically overlaps with Desbuquois dysplasia 2 (DBQD2 [MIM: 615777]), another skeletal dysplasia caused by bi-allelic loss-of-function variants in *XYLT1* (MIM: 608124). Using genome sequencing, microarray analysis, and Sanger sequencing, Lacroix and

collaborators identified homozygous or compound heterozygous pathogenic variants or deletions altering *XYLT1* in a few individuals with BSS while other affected individuals had only one heterozygous or no *XYLT1* variant. Based on possible allelic drop-out, they looked for DNA methylation defects and showed that alleles without identified variant or deletion were hypermethylated. This hypermethylation was secondary to a CGG expansion occurring in the 5' UTR of *XYLT1* and was associated with silencing of the corresponding allele.<sup>50</sup> These results show that BBS and DBQD2 are allelic disorders and that the most frequent loss-of-function variant in *XYLT1* is a CGG expansion in its 5' UTR.

Recessive expansions leading to a loss of function of *GLS* (MIM: 138280) were identified thanks to their associated biochemical defect. The discovery started with the identification of heterozygous variants (one missense and one nonsense) in *GLS* in two of three unrelated individuals with global developmental delay, progressive ataxia, and elevated plasma glutamine (GDPAG [MIM: 618412]). *GLS* encodes glutaminase, the enzyme catalyzing the first reaction of glutamine catabolism. Glutaminase activity was strongly impaired in all three subjects, suggesting the presence of additional variants refractory to exome detection. Using ExpansionHunter and repeat-primed PCR, the authors identified a GCA repeat expansion in the 5' UTR of *GLS*, which was either homozygous or compound heterozygous with the previously identified variants. The number of repeats ranged from 400 to 1,500 whereas the normal range in control individuals was 8–16. There was no evidence of DNA methylation associated with this expansion, but histone modifications associated with transcriptionally active regions (H3ac and H3K4m3) were reduced while histone marks of transcriptionally silent regions were increased. These findings suggest that the repeat expansion in *GLS* leads to a change in chromatin configuration that subsequently results in reduced transcription.<sup>53</sup>

### **X-linked expansion related to the insertion of a retrotransposon**

X-linked dystonia-parkinsonism (XDP or “Lubag” syndrome [MIM: 314250]) is a neurodegenerative disorder characterized by severe and progressive torsion dystonia associated with extra-pyramidal signs including tremors, bradykinesia, and rigidity (parkinsonism) described in 1976 (see GeneReviews by Evidente in [web resources](#)). XDP mainly affects males and more rarely females, and is almost exclusively limited to individuals of Filipino descent originating from Panay Island, Philippines, where its prevalence is above 5 in 100,000. Age at onset varies from 12 to 79 years.<sup>122</sup> Autopsy of XDP-affected brains has shown lesions in the striatum resembling those seen in individuals with HD.<sup>123</sup>

Genome-wide linkage in XDP-affected families had identified the DYT3 locus on Xq13.1.<sup>124</sup> Using shotgun

sequencing from BAC clones from a subject with XDP, Makino et al.<sup>122</sup> identified a SINE-VNTR-Alu (SVA) retrotransposon inserted in an intron of *TAF1* (MIM: 313650), encoding the largest component of the transcription factor IID complex. Gene expression analysis in a XDP postmortem brain showed a decrease of a neuron-specific *TAF1* isoform associated with increased DNA methylation at the site of SVA insertion.<sup>122</sup>

The inserted SVA is associated with an identical founder haplotype spanning 294 kb and including five single nucleotide substitutions and a 48-bp deletion.<sup>122</sup> Further analysis of the SVA sequence revealed a polymorphic hexanucleotide (CCCTCT) tandem repeat located at the 5' end of the SVA. Interestingly, the length of this repeat insertion is inversely correlated with the age at onset and its sequence is able to modulate the expression of a reporter gene, suggesting that this repeat expansion contributes to the pathogenesis of XDP.<sup>47,125</sup>

### **Expansions of minisatellites**

#### **A 99-bp coding expansion in *PLIN4* underlies a rare type of myopathy**

In a large family with autosomal-dominant progressive myopathy with rimmed ubiquitin-positive autophagic vacuolation (MRUPAV), Ruggieri et al.<sup>126</sup> identified a linkage interval of 5.12 Mb on chr 19p13.3 ([Table 3](#)). The authors performed a combination of exome, genome, and RNA sequencing from skeletal muscle biopsies, which was unproductive. They then performed a proteome analysis of microdissected vacuoles that revealed a 20-fold increase in periculin-4, which gene (*PLIN4* [MIM: 613247]) was contained within the linked interval. *PLIN4*, like other periculin genes, contains a domain composed of 27 to 31 repetitions of 33-amino acids (99 bp). In *PLIN4*, these 33-amino acid repeats are formed of imperfect 11-mer tracts that each form three amphipathic helices able to anchor the protein to the phospholipid monolayer of lipid droplets. A reanalysis of genome and transcriptome data showed an unusually high coverage of exon 3, encoding the repeats. Long-range PCR revealed a 1,000 pb higher band, and nanopore sequencing showed that this band corresponds to a repetition of 40 × 33-mer instead of the 27–31 normal repeats. This expansion was associated with increased aggregation of periculin-4 and aggrephagy.<sup>126</sup> This example illustrates the complexity of identifying repeat expansions involving minisatellites, due to a computational challenge in mapping and accurately calling repeats of several dozens of nucleotides.

#### **30-bp repeat array in *CACNA1C* associated with schizophrenia and bipolar disorder**

Schizophrenia and bipolar disorder are highly heritable psychiatric conditions that are part of a clinical continuum. Genome-wide association studies (GWASs) on thousands of affected and control individuals have identified more

than 150 loci contributing to these disorders.<sup>127</sup> One of the most reliable association signals lies on chr 12p13.33 within *CACNA1C* (MIM: 114205), the gene encoding the main sub-unit of the neuronal Ca<sub>v</sub>1.2 calcium channel, and consists of SNPs in linkage disequilibrium within a 100-kb interval.<sup>128</sup> In an attempt to identify the variants underlying this association, Song et al. reexamined the corresponding region and identified an intronic 30-mer repeat array that has specifically expanded in humans.<sup>129</sup> Indeed, while chimpanzees have a single copy of the 30-mer, human individuals have multiple copies, ranging from 100 to 3,000 copies. They show evidence that the human reference genome, which contains 10 repeat copies at this position, was wrongly annotated due to a repeat contraction subsequent to the BAC cloning method originally used to sequence the human genome. Using long-read sequencing they demonstrate that these 30-mer repeats are highly variable not only in size but also in their sequence, some positions being nearly invariant whereas others show significant variation. Interestingly, specific sequence variants are associated with a higher risk of developing a psychiatric disorder while others are on the contrary protective. Besides, these repeats are able to modulate the expression of a luciferase reporter in a neural progenitor cell line, indicating that these repeat arrays have variable enhancer activities that could contribute to a differential expression of *CACNA1C* during brain development.<sup>129</sup> This study shows that VNTRs are far more polymorphic than the reference genome suggests and that deciphering their contribution to gene expression and human pathology is of crucial importance but only possible using long-read technologies.

#### **25-bp repeat in *ABCA7* associated with Alzheimer disease (AD)**

AD is a frequent neurodegenerative disorder characterized by progressive dementia affecting memory, thinking, and behavior and is multifactorial in most cases. GWASs have identified an association with *ABCA7* (MIM: 605414) and truncating variants in this gene are five times more frequent in AD-affected individuals versus control individuals (these variants are present in up to 4% of affected individuals).<sup>130,131</sup> In order to identify variants underlying the GWAS signal, De Roeck et al. examined repetitive regions in *ABCA7* poorly covered by NGS and identified a 25-bp tandem repeat adjacent to the splice donor site of exon 18.<sup>132</sup> Using Southern blotting, they observed a high variability in repeat length, ranging from 12 to 427 repeat units. Expanded repeat length (>200 repeats) were significantly associated with AD, with an odd ratio of 4.5 (AD9 [MIM: 608907]). Furthermore, the VNTR length was associated with a decrease in *ABCA7* expression, as a result of alternative splicing events involving exons 18 and 19.<sup>132</sup>

#### **69-bp repeat in *WDR7* modifying the risk of amyotrophic lateral sclerosis (ALS)**

ALS is a progressive and fatal disorder characterized by the rapid neurodegeneration of motor neurons. Based on the

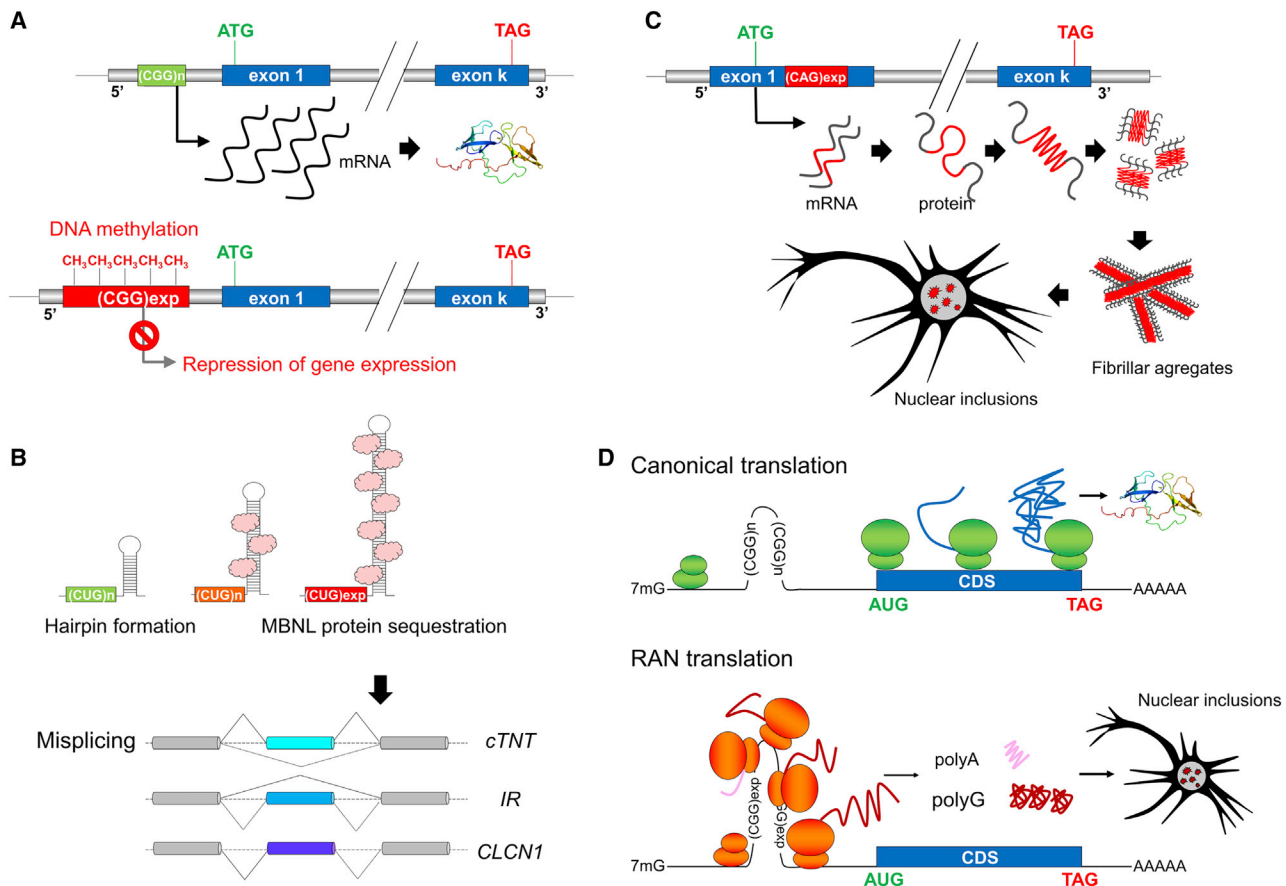
observation that repeat expansions in *C9ORF72* contribute to a significant proportion of individuals with ALS and intermediate CAG repeat number in *SCA2* are able to modify its progression,<sup>133</sup> Course et al. looked for human-specific polymorphic VNTRs in introns and identified 20 repeat-containing regions with unit length >25-mer that mapped to a unique location in human genomes. Among those was a 69-bp repeat located in *WDR7* (MIM: 613473), on chr 18q21, within the *ALS3* (MIM: 606640) locus and encoding rabconnectin-3, a protein enriched in synaptic vesicles.<sup>134</sup> As previously shown for *CACNA1C* repeats, *WDR7* repeats are specifically expanded in humans and highly polymorphic in size. In addition, these repeats are complementary and predicted to form unique secondary structures corresponding to hairpins. By comparing the repeat lengths in 376 individuals with ALS, 531 individuals with Parkinson disease, and 639 control subjects, the authors suggested a significant association of longer repeat lengths with ALS, although the distribution was variable in all three groups, ranging from 51 to 86 repeats. Furthermore, the repeat sequences are variable, and individual alleles are composed of reproducible patterns. The authors suggest that *WDR7* repeats can be transcribed in micro-RNAs *in vitro*, although there was no evidence of their presence in neuronal RNA-seq datasets, and that ectopically expressed repeats form aggregates.<sup>134</sup>

### **Pathophysiological mechanisms**

#### **Epigenetic gene silencing**

Loss-of-function by gene silencing is one possible consequence of tandem repeat expansions (Figure 3A). Generally associated with recessive (or X-linked) disorders, these expansions are easily missed. Recessive expansions can be either bi-allelic or associated with a pathogenic point variant in *trans* and the phenotype associated with expansions and point variants is usually identical. Non-coding repeat expansions can modulate the expression of their recipient gene by directly or indirectly altering chromatin conformation and availability. GC-rich expansions may create new or strengthen existing CpG islands, leading to persistent DNA hypermethylation, as observed in *FXS* (*FMRI*)<sup>135</sup> and *BSS* (*XYLT1*).<sup>50</sup> However, some GC-rich expansions seem to lead to epigenetic gene silencing independently of DNA methylation, like expansions in *CSTB* (*EPM1*)<sup>136</sup> and more recently in *GLS* (*GDPAG*).<sup>53</sup> The recent example of *GDPAG* has shown that histone acetylation and methylation marks can be altered without any association with DNA hypermethylation. This example is reminiscent of the effect of GAA expansions in *FXN* (MIM: 606829) (causing *FRDA*), which alters transcription of Frataxin by creating secondary DNA/RNA structures (R-loops) that block RNA polymerase and are associated with repressive histone marks.<sup>137,138</sup> Of note, hypermethylation and inhibition of gene transcription can be only one of several consequences of the repeat expansions, as observed for





**Figure 3. Main mechanisms associated with repeat expansions**

(A) Epigenetic gene silencing. Large, usually GC-rich expansions located in promoters and/or 5' UTRs are frequently associated with DNA hypermethylation at CpG sites. Methylated expanded alleles are locked in a chromatin configuration preventing gene expression, and therefore leading to a loss-of-function of the expanded allele. Examples of these mechanisms include full expansions in *FMR1* (fragile X syndrome) and expansions in *XYLT1* (Scott-Baratela syndrome).

(B) Sequestration of RNA-binding splicing factors. Specific expanded repeats within RNA molecules are able to form stable secondary structures such as hairpins, which can bind specific RNA-binding splicing factors (e.g., muscleblind-like proteins) with high affinity. These RNA molecules accumulate to form inclusions in the nucleus and sequester bound RNA-binding splicing factors, resulting in misplicing of a subset of transcripts requiring these factors. The illustration of this mechanism is provided by CTG/CAG expansions in *DMPK*.

(C) Protein misfolding and aggregation. Coding expansions, particularly those encoding polyQ stretches, lead to neuronal intranuclear protein inclusions. This mechanism results from the ability of abnormally long polyQ chains to form  $\beta$  sheet structures prone to form insoluble fibrillar aggregates.

(D) Repeat-associated non-AUG (RAN) translation. RAN translation is a non-canonical protein synthesis process in which peptide synthesis is initiated at the site of the expanded repeats in absence of an AUG codon. This process can theoretically occur in all reading frames on both sense and antisense DNA strands, resulting in up to six different polypeptides, although only some of these peptides are toxic and/or preferentially expressed. For example, intermediate CGG expansions (ranging from 55 to 200 repeats) associated with FXTAS mainly lead to synthesis of polyglycine peptides that are able to accumulate and form protein aggregates. The list of repeat expansion disorders associated with RAN occurs is rapidly increasing and so far includes DM2, FXTAS, FXPOI, HD, *C9ORF72*-FTD/ALS, SCA8, and SCA31.

*C9ORF72*-FTD/ALS, one of the few dominant disorders associated with hypermethylation.<sup>101</sup> Although DNA methylation inversely correlates with repeat size and age at disease onset,<sup>139</sup> loss of function of *C9ORF72* is likely insufficient to lead to FTD/ALS as no truncating mutations have been yet reported in this gene, but recent evidence suggests that it could contribute to disease pathogenesis by worsening the repeat-dependent gain-of-function mechanisms.<sup>140</sup>

#### RNA toxicity mediated by protein titration

RNA is the molecule that could mediate the pathogenicity of repeat expansion disorder in many cases. Indeed, the vast majority of repeat expansions undergo transcription,

and the transcription often even occurs bilaterally, i.e., in forward and reverse senses.<sup>101,141</sup> Repeat-containing RNA molecules can act through different mechanisms and alter multiple pathways, as previously reviewed.<sup>101</sup> The best-characterized example of these mechanisms has been studied in myotonic dystrophy: *DMPK* mRNA molecules with 3' CUG expanded repeats accumulate to form inclusions in muscle and neuron nuclei and sequester specific RNA-binding splicing factors (e.g., muscleblind-like, MBNL protein). The subsequent depletion of these proteins leads to misplicing of a subset of tissue-specific transcripts (Figure 3B).<sup>142,143</sup> It is still unclear whether recently identified repeat expansion disorders, including for instance SCA37 and FAME in



which RNA foci have been detected, could fall within the category of disorders resulting from the titration of neuronal-specific splicing factors.

### Protein misfolding and aggregation in repeat disorders

Protein aggregation has long been a characteristic feature of coding expansions (Figure 3C), in particular polyQ disorders. The recent identification of 99-mer expansions in *PLIN4* illustrates that repeat motifs other than trinucleotides can lead to protein aggregation.<sup>126</sup> Furthermore, the identification of 5' UTR CGG expansions in NIID and oculopharyngeal myopathies demonstrates that neuronal nuclear aggregates can also be a consequence of non-coding expansions although the precise mechanisms related to these aggregates remain to be determined. These disorders share many characteristics in common with FXTAS and it is hence possible that pathogenic mechanisms are alike.

### Repeat-associated non-AUG (RAN) translation

RAN translation is a non-canonical protein synthesis process, first described SCA8 and DM1 in 2011.<sup>144</sup> RAN translation (Figure 3D) initiates at the site of the expanded repeats in the absence of an AUG codon and can theoretically occur in all three reading frames using both sense and antisense directions, resulting in up to six different polypeptides.<sup>145,146</sup> For example, non-coding CAG/CTG expansions in SCA8 result in polyserine, polyglutamine, and polyalanine peptides, the two latter being able to accumulate and form aggregates. The list of repeat expansion disorders associated with RAN translation is constantly growing and now also includes myotonic dystrophy type 2 (DM2 [MIM: 602668]) caused by CCTG expansions in *CNBP* (MIM: 116955), FXTAS, fragile X-associated premature ovarian infertility (FXPOI [MIM: 311360]) in females with intermediate CGG repeat expansion in *FMR1*, HD (CAG repeat expansions in *HTT* [MIM: 613004]), *C9ORF72*-FTD/ALS, and SCA31 (MIM: 117210).<sup>145,147–151</sup> RAN translation of specific reading frames and accumulation of some peptides could occur in some cell types but not in others, providing a possible ground for the tissue specificity of some disorders.<sup>147,151</sup> Mechanisms involving toxic RNA and RAN translation are not mutually exclusive, as illustrated by DM1 and *C9ORF72*-FTD/ALS in which both RNA foci and RAN peptide aggregates have been detected in affected tissues.<sup>146</sup>

### Precision medicine in repeat expansions disorders

Precision medicine (previously referred to as personalized medicine), which consists of developing a treatment entirely adapted to an individual's disease, taking the genome and environment into account, is increasingly becoming an accessible goal in genetic disorders.<sup>152</sup> This concept has also been developed for several years for specific repeat expansion disorders, in particular HD. In this

section of the review, we will focus on the recent progress in this field.

A possible way of developing personalized treatments is to understand the pathological consequences of the expansion at the cellular level and to find molecules that are able to either reverse or delay these consequences. For example, in polyQ disorders, the accumulation of polyQ protein aggregates leads to the dysfunction of several pathways, including altered transcription, reduced mitochondrial metabolism, excitotoxicity, autophagy, and apoptosis. Molecules targeting one or several of these deregulated pathways have therefore been suggested as possible treatments.<sup>153</sup> However, these approaches are usually limited as they are not able to fully restore the cellular dysfunctions. Yet, some of these strategies can prove useful on the long term in combination with other approaches. For expansion disorders associated with RAN translation, antibodies have been designed to target specifically the toxic peptides produced and block their effect *in vivo*. This strategy has recently proven successful in a mouse model of *C9ORF72*-FTD/ALS.<sup>154</sup>

A promising way to treat dominant disorders associated with a gain of function of the expanded allele is to target and neutralize specifically this allele. This implies that one functional copy of the gene is sufficient and this strategy is not suitable for genes associated with haploinsufficiency. CRISPR-Cas9 technologies offer the possibility to edit individual genomes but so far they have limitations, including difficulties in delivering the proteins and RNA guide necessary to gene editing in the affected tissues, as well as possible off-target effects, which still restrict their use for clinical purposes.<sup>155</sup> Alternative approaches making use of antisense oligonucleotides (ASO) or siRNA delivered intrathecally have proven to be efficient in downregulating the expression of the expanded allele and have already reached phase III in clinical trials.<sup>156</sup> The designed oligonucleotides cannot directly target the repeats because of their repetitive nature and their presence in other genes and thus need to be complementary to a specific sequence located in *cis* of the expansion. In this setting, single-nucleotide polymorphisms (SNPs) are used to make the difference between the expanded and the normal alleles and target specifically the allele carrying the expansion.<sup>157,158</sup> This requires researchers to genotype and determine *cis* associations existing in each affected individual prior to designing the treatment, and this strategy can only be used for subjects with informative coding SNPs in the corresponding gene.

Another recent pioneering approach could become a game changer in HD and other CAG trinucleotide expansion disorders. A small molecule, naphthyridine-azaquinolone, was shown to specifically bind to expanded CAG repeats and induce their contraction in cells from affected individuals and the brains of mice with an expansion in *Htt*.<sup>159</sup> This approach could be used to reduce the number of CAG repeats in the brain of individuals with HD, hopefully leading to a delay in disease onset of several decades or even a lifetime protection.

## Conclusion: Future studies and remaining challenges

The identification of tandem repeat expansions in human disorders is still in its infancy and many more disorders, monogenic or complex, likely remain to be uncovered. Furthermore, in the last few years we have witnessed an important widening of phenotypes related to known repeat expansions, as illustrated by the recent identification of *HTT* expansions in individuals with FTD and ALS.<sup>160</sup> Therefore, repeat expansions should be searched for in unsolved disorders of supposedly genetic origin, even in phenotypes not specifically evoking a known expansion.

The more recent studies have highlighted the extreme diversity in repeat motifs susceptible to expand and their location in genes. However, only specific repeat motifs appear to be pathogenic while others can be expanded in human genomes without any harmful consequences. Making the difference between these pathogenic and non-pathogenic motifs is of utmost importance as this information could be used to specifically chase the pathogenic ones among the many repeats contained in our genomes. Interestingly, most of the pentanucleotide repeat expansions recently discovered are composed of pathogenic motifs distinct from the one(s) present in reference and control genomes. Studying how these alternative pathogenic repeat insertions have arisen could help finding new repeat expansion disorders. In the case of SCA37 expansions, it has been hypothesized that the thymidine to cytosine transition has occurred after expansion of ATTTT repeats to ~200 copies, followed by expansion of the mutant ATTTC sequence itself.<sup>161</sup> This mechanism is in agreement with previous studies showing the existence of susceptibility haplotypes predisposing to the disorders in specific populations.

The example of *FMR1*, in which CGG repeats in 5' UTR have been associated with three different disorders depending on the repeat size and sex (due to the location of the gene on the X chromosome), has demonstrated that expansions may have different downstream effects depending on the expansion length and that intermediate expansion sizes should also be carefully studied as they could constitute susceptibility factors for more common disorders or be associated with other or late-onset phenotypes more difficult to recognize. So far, most expansion disorders affect the central nervous system and the muscles and there are only a few examples of expansion disorders limited to other tissues. This might be attributed to an increased susceptibility of some cells or tissues to mechanisms related to these expansions, for example somatic mosaicism of the repeats with higher repeat lengths in particular tissues,<sup>162</sup> the sequestration of proteins that have a tissue-specific impact (e.g., muscle-blind proteins like), or an increased expression or specific function of the gene where the expansion occurs (e.g., *DABI*).<sup>102</sup> Neuronal and muscle cells, which are termi-

nally differentiated non-cycling cells, might also be more susceptible to somatic repeat expansions or to toxic gain-of-function mechanisms. However, it is also likely that repeat expansions may affect other tissues, as observed for FXPOI.<sup>163</sup>

A recent field of research is to understand the impact of normal and pathogenic tandem repeats on gene expression and chromatin conformation. This is now possible thanks to advanced techniques allowing researchers to study gene expression, chromatin conformation, or both at a high resolution, even from single cells. We have just started to fully acknowledge the role of intragenic microsatellites and minisatellites in modulation of gene expression and as a potential source of gene deregulation.<sup>6,129</sup> However, the study of minisatellites is associated with even greater challenges, including difficulties in determining sequence variation within repeats and integrating how both repeat number and sequence affect gene expression in a given cell type. This challenge is achievable only through novel long-read technologies. Furthermore, recent studies regarding minisatellites suggest that this category of repeated elements could act more frequently as susceptibility factors and modifiers rather than causing monogenic disorders, and establishing association with common or rare disorders might require large sample sizes. As for the role of satellites in human pathology, this is an almost untouched research topic. To date, only one disorder, facioscapulohumeral muscular dystrophy (FSHD1 [MIM: 158900]), has been linked to the heterozygous contraction of a satellite repeat array at the D4Z4 locus (encoding *DUX4* [MIM: 606009]) in the subtelomeric region of chromosome 4q35. Diagnosing this disorder still represents a challenge as only specific techniques, such as Southern blotting, Fiber FISH, or optical genome mapping, can be used to specifically detect the contraction of these very large repeats (approx. 100 kb each) and distinguish them from non-pathogenic contractions of an almost identical repeat array at 10q26 (see GeneReviews by Preston et al. in [web resources](#)).

Finally, we have to keep in mind that repeat tracts can be wrongly annotated in the human reference genome. An accumulation of short reads in a region typically suggests that this region is either variable in copy number or wrongly annotated but this observation can be made for many regions. Moreover, there are still genomic regions whose sequence remains unknown because of their high complexity. Most of these regions, including centromeres and short arms of acrocentric chromosomes, actually correspond to repeats. The telomere-to-telomere (T2T) consortium has just started sequencing all the gaps remaining in the human genome reference, a tour de force that will undoubtedly result in clarifying many human disorders in the future. The first sequence, released in 2020, was a complete (i.e., gapless) telomere-to-telomere assembly of the single chr X present in the complete hydatidiform mole CHM13.<sup>164</sup> This achievement was possible thanks to a

**Table 3. List of repeat expansions involving motifs larger than 12 nucleotides (minisatellites)**

ID year	Disorder	Inheritance	Chrom	Gene	Location	Repeat motif unit	Normal repeat range	Pathological repeat range	Pathological mechanism(s)	Method of identification	Ref
2018	SCHIZO / BD	AD (susceptibility factor)	12p13.33	CACNA1C	intron	30 bp	variable	N/A	specific 30-mer sequences associated with disease risk	GWAS, CGA	Song et al. <sup>129</sup>
2018	Alzheimer disease	AD (susceptibility factor)	19p13.3	ABCA7	intron	25 bp	12–427	>200	association of longer repeat length with AD	GWAS, WES, CGA	De Roeck et al. <sup>132</sup>
2020	MRUPAV	AD	19p13.3	PLIN4	exon	99 bp	27–31	40	protein aggregation	L, WGS, proteomics	Ruggieri et al. <sup>126</sup>
2020	ALS	AD (susceptibility factor)	18q21.31	WDR7	intron	69 bp	variable	N/A	association of longer repeat length with ALS	ExpScr, CGA	Course et al. <sup>134</sup>

AD, autosomal dominant; ALS, amyotrophic lateral sclerosis; MRUPAV, myopathy with rimmed ubiquitin-positive autophagic vacuolation; SCHIZO / BD, schizophrenia / bipolar disorders; N/A, not available; CGA, candidate gene analysis; GWAS, genome-wide association study; ExpScr, expansion screening; WES, whole-exome sequencing; WGS, whole-genome sequencing.

combination of high-coverage, ultra-long-read nanopore sequencing and alternative long-read technologies (SMRT sequencing and optical genome mapping) that improved the accuracy sequencing or facilitated the assembly. The T2T consortium has since announced the completion of the CHM13 genome sequence. The characterization of the variability of these previously unsequenceable regions and distinction between their normal and pathogenic states will require many years but this is obviously the start of a new genomic era where the systematic study of repeats will finally be possible.

### Data and code availability

This review is based on a literature search in PubMed and individual knowledge of the authors on the topic. It does not include any unpublished data or code.

### Supplemental information

Supplemental information can be found online at <https://doi.org/10.1016/j.ajhg.2021.03.011>.

### Acknowledgments

We thank the editors of the *American Journal of Human Genetics* and the three anonymous reviewers who made valuable contributions to the revision of this review. We also thank the University Hospital Essen, the Deutsche Forschungsgemeinschaft (DFG), the Tom-Wahlig-Stiftung (TWS), the Deutsche Stiftung Neurologie (DSN), and the University of Strasbourg Institute of Advanced study (USIAS) for their financial support to the research studies conducted by the authors.

### Declaration of interests

The authors declare no competing interests.

### Web resources

GeneReviews, Evidente, V.G.H. (1993). X-Linked Dystonia-Parkinsonism. <https://pubmed.ncbi.nlm.nih.gov/20301662/>  
 GeneReviews, Preston, M.K., Tawil, R., and Wang, L.H. (1993). Facioscapulohumeral Muscular Dystrophy. <https://pubmed.ncbi.nlm.nih.gov/20301616/>  
 OMIM, <https://www.omim.org/>  
 Pubmed, <https://pubmed.ncbi.nlm.nih.gov/>

### References

- Gymrek, M. (2017). A genomic view of short tandem repeats. *Curr. Opin. Genet. Dev.* 44, 9–16.
- Willems, T., Gymrek, M., Highnam, G., Mittelman, D., Erlich, Y.; and 1000 Genomes Project Consortium (2014). The landscape of human STR variation. *Genome Res.* 24, 1894–1904.
- Mattioli, F., Hayot, G., Drouot, N., Isidor, B., Courraud, J., Hinckelmann, M.V., Mau-Them, F.T., Sellier, C., Goldman, A., Telegrafi, A., et al. (2020). De Novo Frameshift Variants in the Neuronal Splicing Factor NOVA2 Result in a Common C-Terminal Extension and Cause a Severe Form of Neurodevelopmental Disorder. *Am. J. Hum. Genet.* 106, 438–452.
- Sulovari, A., Li, R., Audano, P.A., Porubsky, D., Vollger, M.R., Logsdon, G.A., Warren, W.C., Pollen, A.A., Chaisson, M.J.P., Eichler, E.E.; and Human Genome Structural Variation Consortium (2019). Human-specific tandem repeat expansion and differential gene expression during primate evolution. *Proc. Natl. Acad. Sci. USA* 116, 23243–23253.
- Grandi, F.C., and An, W. (2013). Non-LTR retrotransposons and microsatellites: Partners in genomic variation. *Mob. Genet. Elements* 3, e25674.
- Fotsing, S.F., Margoliash, J., Wang, C., Saini, S., Yanicky, R., Shleizer-Burko, S., Goren, A., and Gymrek, M. (2019). The impact of short tandem repeat variation on gene expression. *Nat. Genet.* 51, 1652–1659.
- Gymrek, M., Willems, T., Reich, D., and Erlich, Y. (2017). Interpreting short tandem repeat variations in humans using mutational constraint. *Nat. Genet.* 49, 1495–1501.

8. Oberlé, I., Rousseau, F., Heitz, D., Kretz, C., Devys, D., Hanaauer, A., Boué, J., Bertheas, M.F., and Mandel, J.L. (1991). Instability of a 550-base pair DNA segment and abnormal methylation in fragile X syndrome. *Science* 252, 1097–1102.
9. Verkerk, A.J., Pieretti, M., Sutcliffe, J.S., Fu, Y.H., Kuhl, D.P., Pizzuti, A., Reiner, O., Richards, S., Victoria, M.F., Zhang, F.P., 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.
10. La Spada, A.R., Wilson, E.M., Lubahn, D.B., Harding, A.E., and Fischbeck, K.H. (1991). Androgen receptor gene mutations in X-linked spinal and bulbar muscular atrophy. *Nature* 352, 77–79.
11. The Huntington's Disease Collaborative Research Group (1993). A novel gene containing a trinucleotide repeat that is expanded and unstable on Huntington's disease chromosomes. *Cell* 72, 971–983.
12. Orr, H.T., Chung, M.Y., Banfi, S., Kwiatkowski, T.J., Jr., Servadio, A., Beaudet, A.L., McCall, A.E., Duvick, L.A., Renum, L.P., and Zoghbi, H.Y. (1993). Expansion of an unstable trinucleotide CAG repeat in spinocerebellar ataxia type 1. *Nat. Genet.* 4, 221–226.
13. Koide, R., Ikeuchi, T., Onodera, O., Tanaka, H., Igarashi, S., Endo, K., Takahashi, H., Kondo, R., Ishikawa, A., Hayashi, T., et al. (1994). Unstable expansion of CAG repeat in hereditary dentatorubral-pallidolusian atrophy (DRPLA). *Nat. Genet.* 6, 9–13.
14. Nagafuchi, S., Yanagisawa, H., Sato, K., Shirayama, T., Oh-saki, E., Bundo, M., Takeda, T., Tadokoro, K., Kondo, I., Murayama, N., et al. (1994). Dentatorubral and pallidolusian atrophy expansion of an unstable CAG trinucleotide on chromosome 12p. *Nat. Genet.* 6, 14–18.
15. Kawaguchi, Y., Okamoto, T., Taniwaki, M., Aizawa, M., Inoue, M., Katayama, S., Kawakami, H., Nakamura, S., Nishimura, M., Akiguchi, I., et al. (1994). CAG expansions in a novel gene for Machado-Joseph disease at chromosome 14q32.1. *Nat. Genet.* 8, 221–228.
16. Pulst, S.M., Nechiporuk, A., Nechiporuk, T., Gispert, S., Chen, X.N., Lopes-Cendes, I., Pearlman, S., Starkman, S., Orzoco-Diaz, G., Lunkes, A., et al. (1996). Moderate expansion of a normally biallelic trinucleotide repeat in spinocerebellar ataxia type 2. *Nat. Genet.* 14, 269–276.
17. Lindblad, K., Savontaus, M.L., Stevanin, G., Holmberg, M., Digre, K., Zander, C., Ehrsson, H., David, G., Benomar, A., Nikoskelainen, E., et al. (1996). An expanded CAG repeat sequence in spinocerebellar ataxia type 7. *Genome Res.* 6, 965–971.
18. Akarsu, A.N., Stoilov, I., Yilmaz, E., Sayli, B.S., and Sarfarazi, M. (1996). Genomic structure of HOXD13 gene: a nine polyalanine duplication causes synpolydactyly in two unrelated families. *Hum. Mol. Genet.* 5, 945–952.
19. Zhuchenko, O., Bailey, J., Bonnen, P., Ashizawa, T., Stockton, D.W., Amos, C., Dobyns, W.B., Subramony, S.H., Zoghbi, H.Y., and Lee, C.C. (1997). Autosomal dominant cerebellar ataxia (SCA6) associated with small polyglutamine expansions in the alpha 1A-voltage-dependent calcium channel. *Nat. Genet.* 15, 62–69.
20. Mundlos, S., Otto, F., Mundlos, C., Mulliken, J.B., Aylsworth, A.S., Albright, S., Lindhout, D., Cole, W.G., Henn, W., Knoll, J.H., et al. (1997). Mutations involving the transcription factor CBEA1 cause cleidocranial dysplasia. *Cell* 89, 773–779.
21. Brais, B., Bouchard, J.P., Xie, Y.G., Rochefort, D.L., Chrétien, N., Tomé, F.M., Lafrenière, R.G., Rommens, J.M., Uyama, E., Nohira, O., et al. (1998). Short GCG expansions in the PABP2 gene cause oculopharyngeal muscular dystrophy. *Nat. Genet.* 18, 164–167.
22. Koide, R., Kobayashi, S., Shimohata, T., Ikeuchi, T., Maruyama, M., Saito, M., Yamada, M., Takahashi, H., and Tsuji, S. (1999). A neurological disease caused by an expanded CAG trinucleotide repeat in the TATA-binding protein gene: a new polyglutamine disease? *Hum. Mol. Genet.* 8, 2047–2053.
23. Goodman, F.R., Bacchelli, C., Brady, A.F., Brueton, L.A., Fryns, J.P., Mortlock, D.P., Innis, J.W., Holmes, L.B., Donnemfeld, A.E., Feingold, M., et al. (2000). Novel HOXA13 mutations and the phenotypic spectrum of hand-foot-genital syndrome. *Am. J. Hum. Genet.* 67, 197–202.
24. Utsch, B., Becker, K., Brock, D., Lentze, M.J., Bidlingmaier, F., and Ludwig, M. (2002). A novel stable polyalanine [poly(A)] expansion in the HOXA13 gene associated with hand-foot-genital syndrome: proper function of poly(A)-harbouring transcription factors depends on a critical repeat length? *Hum. Genet.* 110, 488–494.
25. Margolis, R.L., O'Hearn, E., Rosenblatt, A., Willour, V., Holmes, S.E., Franz, M.L., Callahan, C., Hwang, H.S., Troncoso, J.C., and Ross, C.A. (2001). A disorder similar to Huntington's disease is associated with a novel CAG repeat expansion. *Ann. Neurol.* 50, 373–380.
26. De Baere, E., Dixon, M.J., Small, K.W., Jabs, E.W., Leroy, B.P., Devriendt, K., Gillerot, Y., Mortier, G., Meire, F., Van Maldergem, L., et al. (2001). Spectrum of FOXL2 gene mutations in blepharophimosis-ptosis-epicanthus inversus (BPES) families demonstrates a genotype-phenotype correlation. *Hum. Mol. Genet.* 10, 1591–1600.
27. Brown, L.Y., Odent, S., David, V., Blayau, M., Dubourg, C., Apacik, C., Delgado, M.A., Hall, B.D., Reynolds, J.F., Sommer, A., 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.
28. Strømme, P., Mangelsdorf, M.E., Shaw, M.A., Lower, K.M., Lewis, S.M., Bruyere, H., Lütcherath, V., Gedeon, A.K., Wallace, R.H., Scheffer, I.E., et al. (2002). Mutations in the human ortholog of *Aristaless* cause X-linked mental retardation and epilepsy. *Nat. Genet.* 30, 441–445.
29. Laumonnier, F., Ronce, N., Hamel, B.C., Thomas, P., Lespinasse, J., Raynaud, M., Paringaux, C., Van Bokhoven, H., Kalscheuer, V., Fryns, J.P., et al. (2002). Transcription factor SOX3 is involved in X-linked mental retardation with growth hormone deficiency. *Am. J. Hum. Genet.* 71, 1450–1455.
30. Amiel, J., Laudier, B., Attié-Bitach, T., Trang, H., de Pontual, L., Gener, B., Trochet, D., Etchevers, H., Ray, P., Simonneau, M., 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.
31. Mahadevan, M., Tsilfidis, C., Sabourin, L., Shutler, G., Ame-miya, C., Jansen, G., Neville, C., Narang, M., Barceló, J., O'Hoy, K., et al. (1992). Myotonic dystrophy mutation: an unstable CTG repeat in the 3' untranslated region of the gene. *Science* 255, 1253–1255.
32. Knight, S.J., Flannery, A.V., Hirst, M.C., Campbell, L., Christodoulou, Z., Phelps, S.R., Pointon, J., Middleton-Price, H.R.,



- Barnicoat, A., Pembrey, M.E., et al. (1993). Trinucleotide repeat amplification and hypermethylation of a CpG island in FRAXE mental retardation. *Cell* 74, 127–134.
33. Campuzano, V., Montermini, L., Moltò, M.D., Pianese, L., Cossée, M., Cavalcanti, F., Monros, E., Rodius, F., Duclos, F., Monticelli, A., et al. (1996). Friedreich's ataxia: autosomal recessive disease caused by an intronic GAA triplet repeat expansion. *Science* 271, 1423–1427.
34. Lalioti, M.D., Scott, H.S., Buresi, C., Rossier, C., Bottani, A., Morris, M.A., Malafosse, A., and Antonarakis, S.E. (1997). Dodecamer repeat expansion in cystatin B gene in progressive myoclonus epilepsy. *Nature* 386, 847–851.
35. Conway, G.S., Payne, N.N., Webb, J., Murray, A., and Jacobs, P.A. (1998). Fragile X premutation screening in women with premature ovarian failure. *Hum. Reprod.* 13, 1184–1187.
36. Koob, M.D., Moseley, M.L., Schut, L.J., Benzow, K.A., Bird, T.D., Day, J.W., and Ranum, L.P. (1999). An untranslated CTG expansion causes a novel form of spinocerebellar ataxia (SCA8). *Nat. Genet.* 21, 379–384.
37. Holmes, S.E., O'Hearn, E.E., McInnis, M.G., Gorelick-Feldman, D.A., Kleiderlein, J.J., Callahan, C., Kwak, N.G., Ingersoll-Ashworth, R.G., Sherr, M., Sumner, A.J., et al. (1999). Expansion of a novel CAG trinucleotide repeat in the 5' region of PPP2R2B is associated with SCA12. *Nat. Genet.* 23, 391–392.
38. Matsuura, T., Yamagata, T., Burgess, D.L., Rasmussen, A., Grewal, R.P., Watase, K., Khajavi, M., McCall, A.E., Davis, C.F., Zu, L., et al. (2000). Large expansion of the ATTCT pentanucleotide repeat in spinocerebellar ataxia type 10. *Nat. Genet.* 26, 191–194.
39. Liquori, C.L., Ricker, K., Moseley, M.L., Jacobsen, J.F., Kress, W., Naylor, S.L., Day, J.W., and Ranum, L.P. (2001). Myotonic dystrophy type 2 caused by a CCTG expansion in intron 1 of ZNF9. *Science* 293, 864–867.
40. Hagerman, R.J., Leehey, M., Heinrichs, W., Tassone, F., Wilson, R., Hills, J., Grigsby, J., Gage, B., and Hagerman, P.J. (2001). Intention tremor, parkinsonism, and generalized brain atrophy in male carriers of fragile X. *Neurology* 57, 127–130.
41. Sato, N., Amino, T., Kobayashi, K., Asakawa, S., Ishiguro, T., Tsunemi, T., Takahashi, M., Matsuura, T., Flanigan, K.M., Iwasaki, S., et al. (2009). Spinocerebellar ataxia type 31 is associated with “inserted” penta-nucleotide repeats containing (TGGAA)<sub>n</sub>. *Am. J. Hum. Genet.* 85, 544–557.
42. Kobayashi, H., Abe, K., Matsuura, T., Ikeda, Y., Hitomi, T., Akechi, Y., Habu, T., Liu, W., Okuda, H., and Koizumi, A. (2011). Expansion of intronic GGCCTG hexanucleotide repeat in NOP56 causes SCA36, a type of spinocerebellar ataxia accompanied by motor neuron involvement. *Am. J. Hum. Genet.* 89, 121–130.
43. Renton, A.E., Majounie, E., Waite, A., Simón-Sánchez, J., Rollinson, S., Gibbs, J.R., Schymick, J.C., Laaksovirta, H., van Swieten, J.C., Myllykangas, L., et al.; ITALSGEN Consortium (2011). A hexanucleotide repeat expansion in C9ORF72 is the cause of chromosome 9p21-linked ALS-FTD. *Neuron* 72, 257–268.
44. DeJesus-Hernandez, M., Mackenzie, I.R., Boeve, B.F., Boxer, A.L., Baker, M., Rutherford, N.J., Nicholson, A.M., Finch, N.A., Flynn, H., Adamson, J., et al. (2011). Expanded GGGGCC hexanucleotide repeat in noncoding region of C9ORF72 causes chromosome 9p-linked FTD and ALS. *Neuron* 72, 245–256.
45. Mootha, V.V., Hussain, I., Cunnusamy, K., Graham, E., Gong, X., Neelam, S., Xing, C., Kittler, R., and Petroll, W.M. (2015). TCF4 Triplet Repeat Expansion and Nuclear RNA Foci in Fuchs' Endothelial Corneal Dystrophy. *Invest. Ophthalmol. Vis. Sci.* 56, 2003–2011.
46. Mootha, V.V., Gong, X., Ku, H.C., and Xing, C. (2014). Association and familial segregation of CTG18.1 trinucleotide repeat expansion of TCF4 gene in Fuchs' endothelial corneal dystrophy. *Invest. Ophthalmol. Vis. Sci.* 55, 33–42.
47. Bragg, D.C., Mangkalaphiban, K., Vaine, C.A., Kulkarni, N.J., Shin, D., Yadav, R., Dhakal, J., Ton, M.L., Cheng, A., Russo, C.T., et al. (2017). Disease onset in X-linked dystonia-parkinsonism correlates with expansion of a hexameric repeat within an SVA retrotransposon in *TAF1*. *Proc. Natl. Acad. Sci. USA* 114, E11020–E11028.
48. Seixas, A.I., Loureiro, J.R., Costa, C., Ordóñez-Ugalde, A., Marcelino, H., Oliveira, C.L., Loureiro, J.L., Dhingra, A., Brandão, E., Cruz, V.T., et al. (2017). A Pentanucleotide ATTCT Repeat Insertion in the Non-coding Region of DAB1, Mapping to SCA37, Causes Spinocerebellar Ataxia. *Am. J. Hum. Genet.* 101, 87–103.
49. Ishiura, H., Doi, K., Mitsui, J., Yoshimura, J., Matsukawa, M.K., Fujiyama, A., Toyoshima, Y., Kakita, A., Takahashi, H., Suzuki, Y., et al. (2018). Expansions of intronic TTTCa and TTTTA repeats in benign adult familial myoclonic epilepsy. *Nat. Genet.* 50, 581–590.
50. LaCroix, A.J., Stabley, D., Sahraoui, R., Adam, M.P., Mehaffey, M., Kernan, K., Myers, C.T., Fagerstrom, C., Anadiotis, G., Akkari, Y.M., et al.; University of Washington Center for Mendelian Genomics (2019). GGC Repeat Expansion and Exon 1 Methylation of *XYLT1* Is a Common Pathogenic Variant in Barata-Scott Syndrome. *Am. J. Hum. Genet.* 104, 35–44.
51. Cortese, A., Simone, R., Sullivan, R., Vandrovcova, J., Tariq, H., Yau, W.Y., Humphrey, J., Jaunmuktane, Z., Sivakumar, P., Polke, J., et al. (2019). Biallelic expansion of an intronic repeat in *RFC1* is a common cause of late-onset ataxia. *Nat. Genet.* 51, 649–658.
52. Rafahi, H., Szmulewicz, D.J., Bennett, M.F., Sobreira, N.L.M., Pope, K., Smith, K.R., Gillies, G., Diakumis, P., Dolzhenko, E., Eberle, M.A., et al. (2019). Bioinformatics-Based Identification of Expanded Repeats: A Non-reference Intronic Pentamer Expansion in *RFC1* Causes CANVAS. *Am. J. Hum. Genet.* 105, 151–165.
53. van Kuilenburg, A.B.P., Tarailo-Graovac, M., Richmond, P.A., Drögemöller, B.I., Pouladi, M.A., Leen, R., Brand-Arzamendi, K., Dobritzsch, D., Dolzhenko, E., Eberle, M.A., et al. (2019). Glutaminase Deficiency Caused by Short Tandem Repeat Expansion in *GLS*. *N. Engl. J. Med.* 380, 1433–1441.
54. Ishiura, H., Shibata, S., Yoshimura, J., Suzuki, Y., Qu, W., Doi, K., Almansour, M.A., Kikuchi, J.K., Taira, M., Mitsui, J., et al. (2019). Noncoding CGG repeat expansions in neuronal intranuclear inclusion disease, oculopharyngodistal myopathy and an overlapping disease. *Nat. Genet.* 51, 1222–1232.
55. Sone, J., Mitsuhashi, S., Fujita, A., Mizuguchi, T., Hamanaka, K., Mori, K., Koike, H., Hashiguchi, A., Takashima, H., Sugiyama, H., et al. (2019). Long-read sequencing identifies GGC repeat expansions in *NOTCH2NL* associated with neuronal intranuclear inclusion disease. *Nat. Genet.* 51, 1215–1221.
56. Tian, Y., Wang, J.L., Huang, W., Zeng, S., Jiao, B., Liu, Z., Chen, Z., Li, Y., Wang, Y., Min, H.X., et al. (2019). Expansion of Human-Specific GGC Repeat in Neuronal Intranuclear



- Inclusion Disease-Related Disorders. *Am. J. Hum. Genet.* 105, 166–176.
57. Yeetong, P., Pongpanich, M., Srichomthong, C., Assawapitaksakul, A., Shotelersuk, V., Tantirukdham, N., Chunharas, C., Suphapeetiporn, K., and Shotelersuk, V. (2019). TTTCA repeat insertions in an intron of YEATS2 in benign adult familial myoclonic epilepsy type 4. *Brain* 142, 3360–3366.
  58. Corbett, M.A., Kroes, T., Veneziano, L., Bennett, M.F., Florian, R., Schneider, A.L., Coppola, A., Licchetta, L., Franceschetti, S., Suppa, A., et al. (2019). Intronic ATTTC repeat expansions in STARD7 in familial adult myoclonic epilepsy linked to chromosome 2. *Nat. Commun.* 10, 4920.
  59. Florian, R.T., Kraft, F., Leitão, E., Kaya, S., Klebe, S., Magnin, E., van Rootselaar, A.F., Buratti, J., Kühnel, T., Schröder, C., et al.; FAME consortium (2019). Unstable TTTTA/TTTCA expansions in MARCH6 are associated with Familial Adult Myoclonic Epilepsy type 3. *Nat. Commun.* 10, 4919.
  60. Deng, J., Yu, J., Li, P., Luan, X., Cao, L., Zhao, J., Yu, M., Zhang, W., Lv, H., Xie, Z., et al. (2020). Expansion of GGC Repeat in GIPC1 Is Associated with Oculopharyngodistal Myopathy. *Am. J. Hum. Genet.* 106, 793–804.
  61. Hannan, A.J. (2018). Tandem repeats mediating genetic plasticity in health and disease. *Nat. Rev. Genet.* 19, 286–298.
  62. Stoyas, C.A., and La Spada, A.R. (2018). The CAG-polyglutamine repeat diseases: a clinical, molecular, genetic, and pathophysiologic nosology. *Handb. Clin. Neurol.* 147, 143–170.
  63. Messaëd, C., and Rouleau, G.A. (2009). Molecular mechanisms underlying polyalanine diseases. *Neurobiol. Dis.* 34, 397–405.
  64. Colak, D., Zaninovic, N., Cohen, M.S., Rosenwaks, Z., Yang, W.Y., Gerhardt, J., Disney, M.D., and Jaffrey, S.R. (2014). Promoter-bound trinucleotide repeat mRNA drives epigenetic silencing in fragile X syndrome. *Science* 343, 1002–1005.
  65. Gousse, G., Patural, H., Touraine, R., Chabrier, S., Rolland, E., Antoine, J.C., and Perrin, L. (2018). Lethal form of spinocerebellar ataxia type 7 with early onset in childhood. *Arch. Pediatr.* 25, 42–44.
  66. Nolin, S.L., Glicksman, A., Tortora, N., Allen, E., Macpherson, J., Mila, M., Vianna-Morgante, A.M., Sherman, S.L., Dobkin, C., Latham, G.J., and Hadd, A.G. (2019). Expansions and contractions of the FMR1 CGG repeat in 5,508 transmissions of normal, intermediate, and premutation alleles. *Am. J. Med. Genet. A.* 179, 1148–1156.
  67. Joosten, I.B.T., Hellebrekers, D.M.E.I., de Greef, B.T.A., Smeets, H.J.M., de Die-Smulders, C.E.M., Faber, C.G., and Gerrits, M.M. (2020). Parental repeat length instability in myotonic dystrophy type 1 pre- and protomutations. *Eur. J. Hum. Genet.* 28, 956–962.
  68. Pearson, C.E., Nichol Edamura, K., and Cleary, J.D. (2005). Repeat instability: mechanisms of dynamic mutations. *Nat. Rev. Genet.* 6, 729–742.
  69. Cleary, J.D., and Pearson, C.E. (2003). The contribution of cis-elements to disease-associated repeat instability: clinical and experimental evidence. *Cytogenet. Genome Res.* 100, 25–55.
  70. Sinden, R.R., Pytlos-Sinden, M.J., and Potaman, V.N. (2007). Slipped strand DNA structures. *Front. Biosci.* 12, 4788–4799.
  71. Jones, L., Houlden, H., and Tabrizi, S.J. (2017). DNA repair in the trinucleotide repeat disorders. *Lancet Neurol.* 16, 88–96.
  72. Matsuura, T., Fang, P., Lin, X., Khajavi, M., Tsuji, K., Rasmussen, A., Grewal, R.P., Achari, M., Alonso, M.E., Pulst, S.M., et al. (2004). Somatic and germline instability of the ATTCT repeat in spinocerebellar ataxia type 10. *Am. J. Hum. Genet.* 74, 1216–1224.
  73. Matsuura, T., Sasaki, H., Yabe, I., Hamada, K., Hamada, T., Shitara, M., and Tashiro, K. (1999). Mosaicism of unstable CAG repeats in the brain of spinocerebellar ataxia type 2. *J. Neurol.* 246, 835–839.
  74. Wong, L.J., Ashizawa, T., Monckton, D.G., Caskey, C.T., and Richards, C.S. (1995). Somatic heterogeneity of the CTG repeat in myotonic dystrophy is age and size dependent. *Am. J. Hum. Genet.* 56, 114–122.
  75. McGoldrick, P., Zhang, M., van Blitterswijk, M., Sato, C., Moreno, D., Xiao, S., Zhang, A.B., McKeever, P.M., Weichert, A., Schneider, R., et al. (2018). Unaffected mosaic *C9orf72* case: RNA foci, dipeptide proteins, but upregulated *C9orf72* expression. *Neurology* 90, e323–e331.
  76. Cumming, S.A., Hamilton, M.J., Robb, Y., Gregory, H., McWilliam, C., Cooper, A., Adam, B., McGhie, J., Hamilton, G., Herzyk, P., et al.; Scottish Myotonic Dystrophy Consortium (2018). De novo repeat interruptions are associated with reduced somatic instability and mild or absent clinical features in myotonic dystrophy type 1. *Eur. J. Hum. Genet.* 26, 1635–1647.
  77. Kraus-Perrotta, C., and Lagalwar, S. (2016). Expansion, mosaicism and interruption: mechanisms of the CAG repeat mutation in spinocerebellar ataxia type 1. *Cerebellum Ataxias* 3, 20.
  78. Pearson, C.E., Eichler, E.E., Lorenzetti, D., Kramer, S.F., Zoghbi, H.Y., Nelson, D.L., and Sinden, R.R. (1998). Interruptions in the triplet repeats of SCA1 and FRAXA reduce the propensity and complexity of slipped strand DNA (S-DNA) formation. *Biochemistry* 37, 2701–2708.
  79. Santoro, M., Masciullo, M., Silvestri, G., Novelli, G., and Botta, A. (2017). Myotonic dystrophy type 1: role of CCG, CTC and CGG interruptions within DMPK alleles in the pathogenesis and molecular diagnosis. *Clin. Genet.* 92, 355–364.
  80. Charles, P., Camuzat, A., Benammar, N., Sellal, F., Destée, A., Bonnet, A.M., Lesage, S., Le Ber, I., Stevanin, G., Dürr, A., Brice, A.; and French Parkinson's Disease Genetic Study Group (2007). Are interrupted SCA2 CAG repeat expansions responsible for parkinsonism? *Neurology* 69, 1970–1975.
  81. Acton, R.T., Rivers, C.A., Watson, B., and Oh, S.J. (2007). DMPK-associated myotonic dystrophy and CTG repeats in Alabama African Americans. *Clin. Genet.* 72, 448–453.
  82. Cossée, M., Schmitt, M., Campuzano, V., Reutenauer, L., Moutou, C., Mandel, J.L., and Koenig, M. (1997). Evolution of the Friedreich's ataxia trinucleotide repeat expansion: founder effect and premutations. *Proc. Natl. Acad. Sci. USA* 94, 7452–7457.
  83. Eichler, E.E., Macpherson, J.N., Murray, A., Jacobs, P.A., Chakravarti, A., and Nelson, D.L. (1996). Haplotype and interspersal analysis of the FMR1 CGG repeat identifies two different mutational pathways for the origin of the fragile X syndrome. *Hum. Mol. Genet.* 5, 319–330.
  84. Smith, B.N., Newhouse, S., Shatunov, A., Vance, C., Topp, S., Johnson, L., Miller, J., Lee, Y., Troakes, C., Scott, K.M., et al. (2013). The *C9ORF72* expansion mutation is a common cause of ALS+/-FTD in Europe and has a single founder. *Eur. J. Hum. Genet.* 21, 102–108.
  85. Martins, S., and Sequeiros, J. (2018). Origins and Spread of Machado-Joseph Disease Ancestral Mutations Events. *Adv. Exp. Med. Biol.* 1049, 243–254.

86. Bampi, G.B., Bisso-Machado, R., Hünemeier, T., Gheno, T.C., Furtado, G.V., Veliz-Otani, D., Cornejo-Olivas, M., Mazzeti, P., Bortolini, M.C., Jardim, L.B., Saraiva-Pereira, M.L.; and Rede Neurogenética (2017). Haplotype Study in SCA10 Families Provides Further Evidence for a Common Ancestral Origin of the Mutation. *Neuromolecular Med.* 19, 501–509.
87. Moulard, B., Genton, P., Grid, D., Jeanpierre, M., Ouazzani, R., Mrabet, A., Morris, M., LeGuern, E., Dravet, C., Mauguière, E., et al. (2002). Haplotype study of West European and North African Unverricht-Lundborg chromosomes: evidence for a few founder mutations. *Hum. Genet.* 111, 255–262.
88. Gymrek, M., Golan, D., Rosset, S., and Erlich, Y. (2012). lobSTR: A short tandem repeat profiler for personal genomes. *Genome Res.* 22, 1154–1162.
89. Willems, T., Zielinski, D., Yuan, J., Gordon, A., Gymrek, M., and Erlich, Y. (2017). Genome-wide profiling of heritable and de novo STR variations. *Nat. Methods* 14, 590–592.
90. Tang, H., Kirkness, E.F., Lippert, C., Biggs, W.H., Fabani, M., Guzman, E., Ramakrishnan, S., Lavrenko, V., Kakaradov, B., Hou, C., et al. (2017). Profiling of Short-Tandem-Repeat Disease Alleles in 12,632 Human Whole Genomes. *Am. J. Hum. Genet.* 101, 700–715.
91. Dolzhenko, E., van Vugt, J.J.F.A., Shaw, R.J., Bekritsky, M.A., van Blitterswijk, M., Narzisi, G., Ajay, S.S., Rajan, V., Lajoie, B.R., Johnson, N.H., et al.; US–Venezuela Collaborative Research Group (2017). Detection of long repeat expansions from PCR-free whole-genome sequence data. *Genome Res.* 27, 1895–1903.
92. Dashnow, H., Lek, M., Phipson, B., Halman, A., Sadedin, S., Lonsdale, A., Davis, M., Lamont, P., Clayton, J.S., Laing, N.G., et al. (2018). STRetch: detecting and discovering pathogenic short tandem repeat expansions. *Genome Biol.* 19, 121.
93. Mousavi, N., Shleizer-Burko, S., Yanicky, R., and Gymrek, M. (2019). Profiling the genome-wide landscape of tandem repeat expansions. *Nucleic Acids Res.* 47, e90.
94. Tankard, R.M., Bennett, M.F., Degorski, P., Delatycki, M.B., Lockhart, P.J., and Bahlo, M. (2018). Detecting Expansions of Tandem Repeats in Cohorts Sequenced with Short-Read Sequencing Data. *Am. J. Hum. Genet.* 103, 858–873.
95. Doi, K., Monjo, T., Hoang, P.H., Yoshimura, J., Yurino, H., Mitsui, J., Ishiura, H., Takahashi, Y., Ichikawa, Y., Goto, J., et al. (2014). Rapid detection of expanded short tandem repeats in personal genomics using hybrid sequencing. *Bioinformatics* 30, 815–822.
96. Dolzhenko, E., Bennett, M.F., Richmond, P.A., Trost, B., Chen, S., van Vugt, J.J.F.A., Nguyen, C., Narzisi, G., Gainulin, V.G., Gross, A.M., et al. (2020). ExpansionHunter Denovo: a computational method for locating known and novel repeat expansions in short-read sequencing data. *Genome Biol.* 21, 102.
97. Trost, B., Engchuan, W., Nguyen, C.M., Thiruvahindrapuram, B., Dolzhenko, E., Backstrom, I., Mirceta, M., Mojarad, B.A., Yin, Y., Dov, A., et al. (2020). Genome-wide detection of tandem DNA repeats that are expanded in autism. *Nature* 586, 80–86.
98. Mantere, T., Kersten, S., and Hoischen, A. (2019). Long-Read Sequencing Emerging in Medical Genetics. *Front. Genet.* 10, 426.
99. De Roeck, A., De Coster, W., Bossaerts, L., Cacace, R., De Pooter, T., Van Dongen, J., D’Hert, S., De Rijk, P., Strazisar, M., Van Broeckhoven, C., and Sleegers, K. (2019). NanoSatellite: accurate characterization of expanded tandem repeat length and sequence through whole genome long-read sequencing on PromethION. *Genome Biol.* 20, 239.
100. Mitsuhashi, S., Frith, M.C., Mizuguchi, T., Miyatake, S., Toyota, T., Adachi, H., Oma, Y., Kino, Y., Mitsuhashi, H., and Matsumoto, N. (2019). Tandem-genotypes: robust detection of tandem repeat expansions from long DNA reads. *Genome Biol.* 20, 58.
101. Rohilla, K.J., and Gagnon, K.T. (2017). RNA biology of disease-associated microsatellite repeat expansions. *Acta Neuropathol. Commun.* 5, 63.
102. Corral-Juan, M., Serrano-Munuera, C., Rábano, A., Cota-González, D., Segarra-Roca, A., Ispuerto, L., Cano-Organ, A.T., Adames, A.D., Méndez-Del-Barrio, C., Jesús, S., et al. (2018). Clinical, genetic and neuropathological characterization of spinocerebellar ataxia type 37. *Brain* 141, 1981–1997.
103. van den Ende, T., Sharifi, S., van der Salm, S.M.A., and van Rootselaar, A.F. (2018). Familial Cortical Myoclonic Tremor and Epilepsy, an Enigmatic Disorder: From Phenotypes to Pathophysiology and Genetics. A Systematic Review. *Tremor Other Hyperkines. Mov. (N. Y.)* 8, 503.
104. van Rootselaar, A.F., van Schaik, I.N., van den Maagdenberg, A.M., Koelman, J.H., Callenbach, P.M., and Tijssen, M.A. (2005). Familial cortical myoclonic tremor with epilepsy: a single syndromic classification for a group of pedigrees bearing common features. *Mov. Disord.* 20, 665–673.
105. Cen, Z., Chen, Y., Yang, D., Zhu, Q., Chen, S., Chen, X., Wang, B., Xie, F., Ouyang, Z., Jiang, Z., et al. (2019). Intronic (TTTGA)<sub>n</sub> insertion in SAMD12 also causes familial cortical myoclonic tremor with epilepsy. *Mov. Disord.* 34, 1571–1576.
106. Bennett, M.F., Oliver, K.L., Regan, B.M., Bellows, S.T., Schneider, A.L., Rafehi, H., Sikta, N., Crompton, D.E., Coleman, M., Hildebrand, M.S., et al. (2020). Familial adult myoclonic epilepsy type 1 SAMD12 TTTCA repeat expansion arose 17,000 years ago and is present in Sri Lankan and Indian families. *Eur. J. Hum. Genet.* 28, 973–978.
107. Sone, J., Mori, K., Inagaki, T., Katsumata, R., Takagi, S., Yokoi, S., Araki, K., Kato, T., Nakamura, T., Koike, H., et al. (2016). Clinicopathological features of adult-onset neuronal intranuclear inclusion disease. *Brain* 139, 3170–3186.
108. Chen, H., Lu, L., Wang, B., Hua, X., Wan, B., Sun, M., and Xu, X. (2020). Essential tremor as the early symptom of NOTCH2NLC gene-related repeat expansion disorder. *Brain* 143, e56.
109. Sun, Q.Y., Xu, Q., Tian, Y., Hu, Z.M., Qin, L.X., Yang, J.X., Huang, W., Xue, J., Li, J.C., Zeng, S., et al. (2020). Expansion of GGC repeat in the human-specific NOTCH2NLC gene is associated with essential tremor. *Brain* 143, 222–233.
110. Jiao, B., Zhou, L., Zhou, Y., Weng, L., Liao, X., Tian, Y., Guo, L., Liu, X., Yuan, Z., Xiao, X., et al. (2020). Identification of expanded repeats in NOTCH2NLC in neurodegenerative dementias. *Neurobiol. Aging* 89, 142.e1–142.e7.
111. Fang, P., Yu, Y., Yao, S., Chen, S., Zhu, M., Chen, Y., Zou, K., Wang, L., Wang, H., Xin, L., et al. (2020). Repeat expansion scanning of the NOTCH2NLC gene in patients with multiple system atrophy. *Ann. Clin. Transl. Neurol.* 7, 517–526.
112. Okubo, M., Doi, H., Fukai, R., Fujita, A., Mitsuhashi, S., Hashiguchi, S., Kishida, H., Ueda, N., Morihara, K., Ogasawara, A., et al. (2019). GGC Repeat Expansion of NOTCH2NLC in Adult Patients with Leukoencephalopathy. *Ann. Neurol.* 86, 962–968.

113. Yau, W.Y., Sullivan, R., Chen, Z., Lynch, D.S., Vandrovcsa, J., Wood, N.W., and Houlden, H. (2020). GGC Repeat Expansion in NOTCH2NLC Is Rare in European Leukoencephalopathy. *Ann. Neurol.* 88, 641–642.
114. Suzuki, I.K., Gacquer, D., Van Heurck, R., Kumar, D., Wojno, M., Bilheu, A., Herpoel, A., Lambert, N., Cheron, J., Polleux, F., et al. (2018). Human-Specific NOTCH2NLC Genes Expand Cortical Neurogenesis through Delta/Notch Regulation. *Cell* 173, 1370–1384.e16.
115. Fiddes, I.T., Lodewijk, G.A., Mooring, M., Bosworth, C.M., Ewing, A.D., Mantalas, G.L., Novak, A.M., van den Bout, A., Bishara, A., Rosenkrantz, J.L., et al. (2018). Human-Specific NOTCH2NLC Genes Affect Notch Signaling and Cortical Neurogenesis. *Cell* 173, 1356–1369.e22.
116. Chen, Z., Xu, Z., Cheng, Q., Tan, Y.J., Ong, H.L., Zhao, Y., Lim, W.K., Teo, J.X., Foo, J.N., Lee, H.Y., et al. (2020). Phenotypic bases of NOTCH2NLC GGC expansion positive neuronal intranuclear inclusion disease in a Southeast Asian cohort. *Clin. Genet.* 98, 274–281.
117. Szmulewicz, D.J., McLean, C.A., MacDougall, H.G., Roberts, L., Storey, E., and Halmagyi, G.M. (2014). CANVAS an update: clinical presentation, investigation and management. *J. Vestib. Res.* 24, 465–474.
118. Akçimen, F., Ross, J.P., Bourassa, C.V., Liao, C., Rochefort, D., Gama, M.T.D., Dicaire, M.J., Barsottini, O.G., Brais, B., Pedrosa, J.L., et al. (2019). Investigation of the *RFC1* Repeat Expansion in a Canadian and a Brazilian Ataxia Cohort: Identification of Novel Conformations. *Front. Genet.* 10, 1219.
119. Tsuchiya, M., Nan, H., Koh, K., Ichinose, Y., Gao, L., Shimozono, K., Hata, T., Kim, Y.J., Ohtsuka, T., Cortese, A., and Takiyama, Y. (2020). RFC1 repeat expansion in Japanese patients with late-onset cerebellar ataxia. *J. Hum. Genet.* 65, 1143–1147.
120. Cortese, A., Tozza, S., Yau, W.Y., Rossi, S., Beecroft, S.J., Jaunmuktane, Z., Dyer, Z., Ravenscroft, G., Lamont, P.J., Mossman, S., et al. (2020). Cerebellar ataxia, neuropathy, vestibular areflexia syndrome due to RFC1 repeat expansion. *Brain* 143, 480–490.
121. Wan, L., Chen, Z., Wan, N., Liu, M., Xue, J., Chen, H., Zhang, Y., Peng, Y., Tang, Z., Gong, Y., et al. (2020). Biallelic Intronic AAGGG Expansion of RFC1 is Related to Multiple System Atrophy. *Ann. Neurol.* 88, 1132–1143.
122. Makino, S., Kaji, R., Ando, S., Tomizawa, M., Yasuno, K., Goto, S., Matsumoto, S., Tabuena, M.D., Maranon, E., Dantes, M., et al. (2007). Reduced neuron-specific expression of the TAF1 gene is associated with X-linked dystonia-parkinsonism. *Am. J. Hum. Genet.* 80, 393–406.
123. Goto, S., Lee, L.V., Munoz, E.L., Tooyama, I., Tamiya, G., Makino, S., Ando, S., Dantes, M.B., Yamada, K., Matsumoto, S., et al. (2005). Functional anatomy of the basal ganglia in X-linked recessive dystonia-parkinsonism. *Ann. Neurol.* 58, 7–17.
124. Németh, A.H., Nolte, D., Dunne, E., Niemann, S., Kostrzewa, M., Peters, U., Fraser, E., Bochukova, E., Butler, R., Brown, J., et al. (1999). Refined linkage disequilibrium and physical mapping of the gene locus for X-linked dystonia-parkinsonism (DYT3). *Genomics* 60, 320–329.
125. Westenberger, A., Reyes, C.J., Saranza, G., Dobricic, V., Hansen, H., Domingo, A., Laabs, B.H., Schaake, S., Pozojevic, J., Rakovic, A., et al. (2019). A hexanucleotide repeat modifies expressivity of X-linked dystonia parkinsonism. *Ann. Neurol.* 85, 812–822.
126. Ruggieri, A., Naumenko, S., Smith, M.A., Iannibelli, E., Blasevich, F., Bragato, C., Gibertini, S., Barton, K., Vorgerd, M., Marcus, K., et al. (2020). Multiomic elucidation of a coding 99-mer repeat-expansion skeletal muscle disease. *Acta Neuropathol.* 140, 231–235.
127. Stahl, E.A., Breen, G., Forstner, A.J., McQuillin, A., Ripke, S., Trubetskoy, V., Mattheisen, M., Wang, Y., Coleman, J.R.I., Gaspar, H.A., et al.; eQTLGen Consortium; BIOS Consortium; and Bipolar Disorder Working Group of the Psychiatric Genomics Consortium (2019). Genome-wide association study identifies 30 loci associated with bipolar disorder. *Nat. Genet.* 51, 793–803.
128. Dedic, N., Pöhlmann, M.L., Richter, J.S., Mehta, D., Czamara, D., Metzger, M.W., Dine, J., Bedenk, B.T., Hartmann, J., Wagner, K.V., et al. (2018). Cross-disorder risk gene CACNA1C differentially modulates susceptibility to psychiatric disorders during development and adulthood. *Mol. Psychiatry* 23, 533–543.
129. Song, J.H.T., Lowe, C.B., and Kingsley, D.M. (2018). Characterization of a Human-Specific Tandem Repeat Associated with Bipolar Disorder and Schizophrenia. *Am. J. Hum. Genet.* 103, 421–430.
130. Le Guennec, K., Nicolas, G., Quenez, O., Charbonnier, C., Wallon, D., Bellenguez, C., Grenier-Boley, B., Rousseau, S., Richard, A.C., Rovelet-Lecrux, A., et al.; CNR-MAJ collaborators (2016). ABCA7 rare variants and Alzheimer disease risk. *Neurology* 86, 2134–2137.
131. Satoh, K., Abe-Dohmae, S., Yokoyama, S., St George-Hyslop, P., and Fraser, P.E. (2015). ATP-binding cassette transporter A7 (ABCA7) loss of function alters Alzheimer amyloid processing. *J. Biol. Chem.* 290, 24152–24165.
132. De Roeck, A., Duchateau, L., Van Dongen, J., Cacace, R., Bjerke, M., Van den Bossche, T., Cras, P., Vandenbergh, R., De Deyn, P.P., Engelborghs, S., et al.; BELNEU Consortium (2018). An intronic VNTR affects splicing of ABCA7 and increases risk of Alzheimer's disease. *Acta Neuropathol.* 135, 827–837.
133. Elden, A.C., Kim, H.J., Hart, M.P., Chen-Plotkin, A.S., Johnson, B.S., Fang, X., Armarkola, M., Geser, F., Greene, R., Lu, M.M., et al. (2010). Ataxin-2 intermediate-length polyglutamine expansions are associated with increased risk for ALS. *Nature* 466, 1069–1075.
134. Course, M.M., Gudsnek, K., Smukowski, S.N., Winston, K., Desai, N., Ross, J.P., Sulovari, A., Bourassa, C.V., Spiegelman, D., Couthouis, J., et al. (2020). Evolution of a Human-Specific Tandem Repeat Associated with ALS. *Am. J. Hum. Genet.* 107, 445–460.
135. Mor-Shaked, H., and Eiges, R. (2018). Reevaluation of *FMR1* Hypermethylation Timing in Fragile X Syndrome. *Front. Mol. Neurosci.* 11, 31.
136. Weinhaeusel, A., Morris, M.A., Antonarakis, S.E., and Haas, O.A. (2003). DNA deamination enables direct PCR amplification of the cystatin B (CSTB) gene-associated dodecamer repeat expansion in myoclonus epilepsy type Unverricht-Lundborg. *Hum. Mutat.* 22, 404–408.
137. Grabczyk, E., Kumari, D., and Usdin, K. (2001). Fragile X syndrome and Friedreich's ataxia: two different paradigms for repeat induced transcript insufficiency. *Brain Res. Bull.* 56, 367–373.
138. Groh, M., Lufino, M.M., Wade-Martins, R., and Gromak, N. (2014). R-loops associated with triplet repeat expansions promote gene silencing in Friedreich ataxia and fragile X syndrome. *PLoS Genet.* 10, e1004318.

139. Xi, Z., Zhang, M., Bruni, A.C., Maletta, R.G., Colao, R., Fratta, P., Polke, J.M., Sweeney, M.G., Mudanohwo, E., Nacmias, B., et al. (2015). The C9orf72 repeat expansion itself is methylated in ALS and FTLN patients. *Acta Neuropathol.* 129, 715–727.
140. Zhu, Q., Jiang, J., Gendron, T.F., McAlonis-Downes, M., Jiang, L., Taylor, A., Diaz Garcia, S., Ghosh Dastidar, S., Rodriguez, M.J., King, P., et al. (2020). Reduced C9ORF72 function exacerbates gain of toxicity from ALS/FTD-causing repeat expansion in C9orf72. *Nat. Neurosci.* 23, 615–624.
141. Groh, M., Silva, L.M., and Gromak, N. (2014). Mechanisms of transcriptional dysregulation in repeat expansion disorders. *Biochem. Soc. Trans.* 42, 1123–1128.
142. Kanadia, R.N., Johnstone, K.A., Mankodi, A., Lungu, C., Thornton, C.A., Esson, D., Timmers, A.M., Hauswirth, W.W., and Swanson, M.S. (2003). A muscleblind knockout model for myotonic dystrophy. *Science* 302, 1978–1980.
143. Timchenko, L.T., Miller, J.W., Timchenko, N.A., DeVore, D.R., Datar, K.V., Lin, L., Roberts, R., Caskey, C.T., and Swanson, M.S. (1996). Identification of a (CUG)<sub>n</sub> triplet repeat RNA-binding protein and its expression in myotonic dystrophy. *Nucleic Acids Res.* 24, 4407–4414.
144. Zu, T., Gibbens, B., Doty, N.S., Gomes-Pereira, M., Huguet, A., Stone, M.D., Margolis, J., Peterson, M., Markowski, T.W., Ingram, M.A., et al. (2011). Non-ATG-initiated translation directed by microsatellite expansions. *Proc. Natl. Acad. Sci. USA* 108, 260–265.
145. Banez-Coronel, M., and Ranum, L.P.W. (2019). Repeat-associated non-AUG (RAN) translation: insights from pathology. *Lab. Invest.* 99, 929–942.
146. Zu, T., Pattamatta, A., and Ranum, L.P.W. (2018). Repeat-Associated Non-ATG Translation in Neurological Diseases. *Cold Spring Harb. Perspect. Biol.* 10, a033019.
147. Zu, T., Cleary, J.D., Liu, Y., Banez-Coronel, M., Bubenik, J.L., Ayhan, F., Ashizawa, T., Xia, G., Clark, H.B., Yachnis, A.T., et al. (2017). RAN Translation Regulated by Muscleblind Proteins in Myotonic Dystrophy Type 2. *Neuron* 95, 1292–1305.e5.
148. Buijsen, R.A., Visser, J.A., Kramer, P., Severijnen, E.A., Gearing, M., Charlet-Berguerand, N., Sherman, S.L., Berman, R.F., Willemsen, R., and Hukema, R.K. (2016). 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.* 31, 158–168.
149. Todd, P.K., Oh, S.Y., Krans, A., He, F., Sellier, C., Frazer, M., Renoux, A.J., Chen, K.C., Scaglione, K.M., Basrur, V., et al. (2013). CGG repeat-associated translation mediates neurodegeneration in fragile X tremor ataxia syndrome. *Neuron* 78, 440–455.
150. Mori, K., Weng, S.M., Arzberger, T., May, S., Rentzsch, K., Kremmer, E., Schmid, B., Kretschmar, H.A., Cruts, M., Van Broeckhoven, C., et al. (2013). The C9orf72 GGGGCC repeat is translated into aggregating dipeptide-repeat proteins in FTLN/ALS. *Science* 339, 1335–1338.
151. Banez-Coronel, M., Ayhan, F., Tarabochia, A.D., Zu, T., Perez, B.A., Tusi, S.K., Pletnikova, O., Borchelt, D.R., Ross, C.A., Margolis, R.L., et al. (2015). RAN Translation in Huntington Disease. *Neuron* 88, 667–677.
152. Kim, J., Hu, C., Moufawad El Achkar, C., Black, L.E., Douville, J., Larson, A., Pendergast, M.K., Goldkind, S.F., Lee, E.A., Kuniholm, A., et al. (2019). Patient-Customized Oligonucleotide Therapy for a Rare Genetic Disease. *N. Engl. J. Med.* 381, 1644–1652.
153. Bauer, P.O., and Nukina, N. (2009). The pathogenic mechanisms of polyglutamine diseases and current therapeutic strategies. *J. Neurochem.* 110, 1737–1765.
154. Nguyen, L., Montrasio, F., Pattamatta, A., Tusi, S.K., Bardhi, O., Meyer, K.D., Hayes, L., Nakamura, K., Banez-Coronel, M., Coyne, A., et al. (2020). Antibody Therapy Targeting RAN Proteins Rescues C9 ALS/FTD Phenotypes in C9orf72 Mouse Model. *Neuron* 105, 645–662.e11.
155. Jacinto, F.V., Link, W., and Ferreira, B.I. (2020). CRISPR/Cas9-mediated genome editing: From basic research to translational medicine. *J. Cell. Mol. Med.* 24, 3766–3778.
156. Estevez-Fraga, C., Flower, M.D., and Tabrizi, S.J. (2020). Therapeutic strategies for Huntington's disease. *Curr. Opin. Neurol.* 33, 508–518.
157. Kay, C., Skotte, N.H., Southwell, A.L., and Hayden, M.R. (2014). Personalized gene silencing therapeutics for Huntington disease. *Clin. Genet.* 86, 29–36.
158. Wild, E.J., and Tabrizi, S.J. (2017). Therapies targeting DNA and RNA in Huntington's disease. *Lancet Neurol.* 16, 837–847.
159. Nakamori, M., Panigrahi, G.B., Lanni, S., Gall-Duncan, T., Hayakawa, H., Tanaka, H., Luo, J., Otabe, T., Li, J., Sakata, A., et al. (2020). A slipped-CAG DNA-binding small molecule induces trinucleotide-repeat contractions in vivo. *Nat. Genet.* 52, 146–159.
160. Dewan, R., Chia, R., Ding, J., Hickman, R.A., Stein, T.D., Abramzon, Y., Ahmed, S., Sabir, M.S., Portley, M.K., Tucci, A., et al.; American Genome Center (TAGC); FALS Sequencing Consortium; Genomics England Research Consortium; International ALS/FTD Genomics Consortium (iAFGC); International FTD Genetics Consortium (IFGC); International LBD Genomics Consortium (iLBDGC); NYGC ALS Consortium; and PROSPECT Consortium (2021). Pathogenic Huntingtin Repeat Expansions in Patients with Frontotemporal Dementia and Amyotrophic Lateral Sclerosis. *Neuron* 109, 448–460.
161. Loureiro, J.R., Oliveira, C.L., Mota, C., Castro, A.F., Costa, C., Loureiro, J.L., Coutinho, P., Martins, S., Sequeiros, J., and Silveira, I. (2019). Mutational mechanism for DAB1 (ATTTC)<sub>n</sub> insertion in SCA37: ATTTT repeat lengthening and nucleotide substitution. *Hum. Mutat.* 40, 404–412.
162. Swami, M., Hendricks, A.E., Gillis, T., Massood, T., Mysore, J., Myers, R.H., and Wheeler, V.C. (2009). Somatic expansion of the Huntington's disease CAG repeat in the brain is associated with an earlier age of disease onset. *Hum. Mol. Genet.* 18, 3039–3047.
163. Rosario, R., and Anderson, R. (2020). The molecular mechanisms that underlie fragile X-associated premature ovarian insufficiency: is it RNA or protein based? *Mol. Hum. Reprod.* 26, 727–737.
164. Miga, K.H., Koren, S., Rhie, A., Vollger, M.R., Gershman, A., Bizakadze, A., Brooks, S., Howe, E., Porubsky, D., Logsdon, G.A., et al. (2020). Telomere-to-telomere assembly of a complete human X chromosome. *Nature* 585, 79–84.