U.S. National Library of Medicine
National Center for Biotechnology Information

NLM Citation: Wallace SE, Bean LJH. Resources for Genetics Professionals — Genetic Disorders Caused by Nucleotide Repeat Expansions and Contractions. 2017 Mar 14 [Updated 2022 Oct 20]. In: Adam MP, Mirzaa GM, Pagon RA, et al., editors. GeneReviews® [Internet]. Seattle (WA): University of Washington, Seattle; 1993-2023.
Bookshelf URL: https://www.ncbi.nlm.nih.gov/books/

Resources for Genetics Professionals — Genetic Disorders Caused by Nucleotide Repeat Expansions and Contractions

Stephanie E Wallace, MD[1,2] and Lora JH Bean, PhD[3,4]

Created: March 14, 2017; Revised: October 20, 2022.

Nucleotide Repeat Expansions and Contractions

A nucleotide repeat is a sequence of nucleotides repeated a number of times in tandem; nucleotide repeats can occur within or near a gene. The size of nucleotide repeats varies: smaller numbers of repeats are common and not associated with phenotypic abnormalities; abnormally large numbers of repeats may be associated with phenotypic abnormalities and are classified as (in increasing order of size): mutable normal alleles, premutations, reduced-penetrance alleles, and full-penetrance alleles.

Molecular genetic testing used to sequence nucleotide repeats is more difficult than sequencing nonrepetitive regions of the exome because:

- Many of the known nucleotide repeats contain a higher GC content, which is difficult to amplify by PCR; and
- Repetitive regions do not align uniquely; thus, the length of the repeated sequence cannot be determined.

Specific assays are required to analyze each nucleotide repeat of interest:

- DNA containing smaller nucleotide repeats can be amplified by PCR. The amplified segments of DNA are then separated by gel or capillary electrophoresis to determine repeat length.
- Highly expanded nucleotide repeats may not be detected by PCR-based assays due to difficulty in aligning the sequence to a unique genomic position. Additional testing (e.g., Southern blot analysis or triplet repeat primed PCR) may be required to determine the length of highly expanded nucleotide repeats.

Author Affiliations: 1 Senior Editor, GeneReviews; Email: editor2@uw.edu. 2 Clinical Professor, Pediatrics, University of Washington, Seattle, Washington; Email: editor2@uw.edu. 3 Molecular Genetics Editor, GeneReviews. 4 Senior Director, Laboratory Quality Assurance, Perkin-Elmer Genomics, Inc, Pittsburgh, Pennsylvania.

Table. Genetic Disorders Caused by Nucleotide Repeat Expansions and Contractions

Gene	Disorder [1]	MOI	% of Pathogenic Variants [1]	Nucleotide Repeat (Amino Acid)	Repeat Location	Normal Repeat Number [2]	Full-Penetrance Pathogenic Repeat Number
AFF2	Fragile X syndrome, FRAXE type (OMIM 309548)	XL	Most common	CCG	5' UTR	4-39	>200
AR	Spinal and bulbar muscular atrophy	XL	100%	CAG (Gln)	Exon 1	≤34	≥38
ARX	Early-infantile epileptic encephalopathy (OMIM 308350); Partington syndrome (OMIM 309510)	XL	Most common	GCG (Ala)	Exon 2 aa 110-115	10-16	17-27
				GCG (Ala)	Exon 2 aa 144-155	12	20
ATN1	DRPLA	AD	100%	CAG (Gln)	Exon 5	6-35	≥48
ATXN1	Spinocerebellar ataxia type 1	AD	100%	CAG (Gln)	Exon 8	6-35	≥39
ATXN2	Spinocerebellar ataxia type 2	AD	100%	CAG (Gln)	Exon 1	≤31	>34
ATXN3	Spinocerebellar ataxia type 3	AD	100%	CAG (Gln)	Exon 8	12-44	~60-87
ATXN7	Spinocerebellar ataxia type 7	AD	100%	CAG (Gln)	Exon 1	7-27	37-460
ATXN8	Spinocerebellar ataxia type 8	AD	100%	CAG (Gln)	Exon 1	~80	Unknown
ATXN8OS				CTG	3' UTR	15-50 CTA/CTG	See footnote 3.
ATXN10	Spinocerebellar ataxia type 10	AD	100%	ATTCT	Intron 9	10-32	≥800
BEAN1	Spinocerebellar ataxia type 31 (OMIM 117210)	AD	100%	TGGAA	Intron 6	0	2.5- to 3.8-kb insertion
C9orf72	*C9orf72*-related amyotrophic lateral sclerosis and frontotemporal dementia	AD	100%	GGGGCC	Promotor or intron 1	2-24	>60
CACNA1A	Spinocerebellar ataxia type 6		>99%	CAG (Gln)	Exon 7	≤18	20-33
CNBP	Myotonic dystrophy type 2		100%	CCTG	Intron 1	≤26	≥75
COMP	Multiple epiphyseal dysplasia	AD	Rare [4]	GAC (Asp)	Exon 13	5	6
	Pseudoachondroplasia		~33% [4]				2-4 or 7
CSTB	Progressive myoclonic epilepsy type 1	AR	~90%	CCCCGCCCCGCG	Promoter	2-3	≥30

Table. continued from previous page.

Gene	Disorder [1]	MOI	% of Pathogenic Variants [1]	Nucleotide Repeat (Amino Acid)	Repeat Location	Normal Repeat Number [2]	Full-Penetrance Pathogenic Repeat Number
DAB1	Spinocerebellar ataxia type 37	AD	100%	ATTTC	5' UTR intron	0	31-75
DIP2B	Mental retardation, FRA12A type (OMIM 136630)	AD	100%	CGG	Promoter	6-23	>350
DMD	Duchenne muscular dystrophy	XL	1 family [5]	GAA	Intron 62	11-33	59-82
DMPK	Myotonic dystrophy type 1	AD	100%	CTG	3' UTR	5-34	>50
EIF4A3	Pierre Robin sequence with cleft mandible and limb anomalies (OMIM 268305)	AR	100%	Complex [6]	5' UTR	5-12	≥15
FMR1	*FMR1*-related disorders	XL	>99%	CGG	5' UTR	5-44	>200
FOXL2	Blepharophimosis, ptosis, and epicanthus inversus	AD	31%	GCN (Ala)	Exon 1	14	15-24
FXN	Friedreich ataxia	AR	~98%	GAA	Intron 1	5-33	≥66
GIPC1	Oculopharyngodistal myopathy 2 (OMIM 618940)	AD	100%	GGC	5' UTR	12-32	73-164
GLS	Glutaminase deficiency with impaired intellectual development and progressive ataxia (OMIM 618412)	AR	3 individuals	GCA	5' UTR	5-38	680-1500
HOXA13	Hand-foot-genital syndrome	AD	50%-60%	GCN (Ala)	Exon 1 aa 38	14	22
					Exon 1 aa 73	12	18
					Exon 1 aa 116	8, 12, or 18	22-32
HOXD13	Syndactyly type V (OMIM 186300)	AD	3 individuals	GCN (Ala)	Exon 1	15	8-11 or ≥22
HTT	Huntington disease	AD	100%	CAG (Gln)	Exon 1	≤26	≥40
JPH3	Huntington disease-like 2	AD	~100%	CTG (Ala)	Exon 2A	6-28	≥40
LRP12	Oculopharyngodistal myopathy (OMIM 164310) [7]	AD	Unknown	CGG/CGT	5' UTR	13-45	Unknown

Table. continued from previous page.

Gene	Disorder [1]	MOI	% of Pathogenic Variants [1]	Nucleotide Repeat (Amino Acid)	Repeat Location	Normal Repeat Number [2]	Full-Penetrance Pathogenic Repeat Number
MARCHF6	Familial adult myoclonic epilepsy 3 (OMIM 613608)	AD	100%	TTTTA/TTTCA	Intron 1	9-20 [8]	791-1035
MUC1	Autosomal dominant tubulointerstitial kidney disease, *MUC1*-related	AD	~95%	C [9]	Exon 2	7	8
NOP56	Spinocerebellar ataxia type 36 (OMIM 614153)	AD	100%	GGCCTG	Intron 1	3-14	≥650
NOTCH2NLC	Neuronal intranuclear inclusion disease (OMIM 603472)	AD	100%	GGC [10]	5' UTR	<38	≥66
NUTM2B-AS1	Oculopharyngeal myopathy with leukoencephalopathy 1 (OMIM 618637)	AD	100% [11]	CCG	Noncoding RNA	3-16	>35
PABPN1	Oculopharyngeal muscular dystrophy	AD	100%	GCN (Ala)	Exon 1	10	11-18
PHOX2B	Congenital central hypoventilation syndrome	AD	92%	GCN (Ala)	Exon 3	≤20	≥24
PPP2R2B	Spinocerebellar ataxia type 12 (OMIM 604326)	AD	100%	CAG	Promoter	7-31	51-78
PRDM12	Hereditary sensory and autonomic neuropathy type VIII (OMIM 616488)	AR	2 families	GCC (Ala)	Exon 5	7-14	18-19
PRNP	Creutzfeldt-Jakob disease	AD	<15%	CCTCATGGTGGTGGCTGGGGGCAG	Exon 2	4 [12]	5-16
RAPGEF2	Familial adult myoclonic epilepsy type 7 (OMIM 618075)		100%	TTTCA	Intron 14	0	Unknown
RFC1	*RFC1* CANVAS / spectrum disorder	AR	100%	AAGGG [13]	Intron 2	11-200	400 to >2000
RUNX2	Cleidocranial dysplasia spectrum disorder	AD	2 individuals [14]	GCN (Ala)	Exon 1	17	20-27
SAMD12	Familial adult myoclonic epilepsy type 1 (OMIM 601068)	AD	100%	TTTCA	Intron 4	0	≥105

Table. continued from previous page.

Gene	Disorder [1]	MOI	% of Pathogenic Variants [1]	Nucleotide Repeat (Amino Acid)	Repeat Location	Normal Repeat Number [2]	Full-Penetrance Pathogenic Repeat Number
SOX3	Panhypopituitarism and intellectual disability with growth hormone deficiency (OMIM 300123)	XL	3 families [15]	GCN (Ala)	Exon 1	15	8 or 22-26
STARD7	Familial adult myoclonic epilepsy 2 (OMIM 607876)	AD	100%	ATTTT/ATTTC	Intron 1	ATTTT ?; ATTTC 0	ATTTT(>274) ATTTC(>340)
TBP	Spinocerebellar ataxia type 17	AD	100%	CAG or CAA (Gln)	Exon 3	25-40	≥49
TBX1	Tetralogy of Fallot (OMIM 602054)	AD	1 individual	GCN (Ala)	Exon 9c	15	25
TCF4	Fuchs endothelial corneal dystrophy (OMIM 613267)	AD	~70%	CTG or CAG	Intron 3	<40	See footnote 16.
TNRC6A	Familial adult myoclonic epilepsy type 6 (OMIM 618074)	AD	100% [11]	TTTCA	Exon 1	0	29
VWA1	Hereditary motor neuropathy (OMIM 619216)	AR	80%	GGCGCGGAGC	Exon 1	2	1 or 3
XYLT1	Baratela-Scott syndrome (Desbuquois dysplasia type 2; OMIM 615777)		~50%	GGC	Promoter	9-20	~>72
YEATS2	Familial adult myoclonic epilepsy 4 (OMIM 615127)	AD	100% [11]	TTTTA/TTTTC	Intron 1	0	192
ZIC2	Holoprosencephaly type 5 (See Holoprosencephaly Overview.)	AD	~40%	GCN (Ala)	Exon 3	15	25

Table. continued from previous page.

Gene	Disorder [1]	MOI	% of Pathogenic Variants [1]	Nucleotide Repeat (Amino Acid)	Repeat Location	Normal Repeat Number [2]	Full-Penetrance Pathogenic Repeat Number
ZIC3	VACTERL (OMIM 300265)	XL	1 individual	GCC (Ala)	Exon 1	10	12

The human genome includes >32,000 trinucleotide repeats of ≥6 repeated units. The human exome contains 1030 trinucleotide repeats in exons of 878 genes [Kozlowski et al 2010].
aa = amino acid; AD = autosomal dominant; ALS = amyotrophic lateral sclerosis; AR = autosomal recessive; ORF = open reading frame; MOI = mode of inheritance; UTR = untranslated region; XL = X-linked

1. Proportion of pathogenic variants in this gene that are caused by a nucleotide repeat expansion or contraction

2. Includes data derived from the subscription-based professional view of Human Gene Mutation Database [Stenson et al 2020]

3. Penetrance is <100%; increased penetrance is reported for alleles of 54-250 CTA/CTG repeats. However, reduced penetrance has been reported at all allele sizes [Ranum et al 1999].

4. Délot et al [1999]

5. Kekou et al [2016]

6. This repeat comprises repeating units of 18 or 20 nucleotides that vary at a CA sequence.
- Normal repeat: CACA-20-nt(2-9)CA-18-nt(1)CACA-20-nt(1)CA-18-nt(1) – note, a normal allele has 5-12 total repeats.
- Abnormal allele: CACA-20-nt(1) CGCA-20-nt(12-13)CA-18-nt(1)CACA-20-nt(1)CA-18-nt(1) – note, a normal allele has 15-16 total repeats.

For the complete repeat sequence, see Favaro et al [2014].

7. Ishiura et al [2019]

8. Healthy controls were found to have 9-20 TTTTC repeats; TTTCA repeats were only present in pathogenic alleles.

9. Duplication of one cytosine in a heptanucleotide cytosine tract within one copy of a 20-125 copy number VNTR (variable number tandem repeat). The specific VNTR involved varies by family but is consistent within a family.

10. Reported as a GGC repeat [Sone et al 2019, Tian et al 2019] and as a CGG repeat [Ishiura et al 2019]

11. Only one family reported to date

12. Normal *PRNP* alleles have one nonapeptide followed by four octapeptide tandem repeat sequences, each of which comprises the following amino acids: Pro-(His/Gln)-Gly-Gly-Gly-(-/Trp)-Gly-Gln.

13. ACAGG repeat expansion (~1000 repeats) reported in three families [Scriba et al 2020, Tsuchiya et al 2020]

14. Shibata et al [2016]

15. Takagi et al [2014]

16. Penetrance is <100%; reduced penetrance has been reported in individuals with >80 CTG repeats [Wieben et al 2014].

References

Délot E, King LM, Briggs MD, Wilcox WR, Cohn DH. Trinucleotide expansion mutations in the cartilage oligomeric matrix protein (COMP) gene. Hum Mol Genet. 1999;8:123–8. PubMed PMID: 9887340.

Favaro FP, Alvizi L, Zechi-Ceide RM, Bertola D, Felix TM, de Souza J, Raskin S, Twigg SR, Weiner AM, Armas P, Margarit E, Calcaterra NB, Andersen GR, McGowan SJ, Wilkie AO, Richieri-Costa A, de Almeida ML, Passos-Bueno MR. Am J Hum Genet. A noncoding expansion in EIF4A3 causes Richieri-Costa-Pereira syndrome, a craniofacial disorder associated with limb defects. 2014;94:120-8.

Ishiura H, Shibata S, Yoshimura J, Suzuki Y, Qu W, Doi K, Almansour MA, Kikuchi JK, Taira M, Mitsui J, Takahashi Y, Ichikawa Y, Mano T, Iwata A, Harigaya Y, Matsukawa MK, Matsukawa T, Tanaka M, Shirota Y, Ohtomo R, Kowa H, Date H, Mitsue A, Hatsuta H, Morimoto S, Murayama S, Shiio Y, Saito Y, Mitsutake A, Kawai M, Sasaki T, Sugiyama Y, Hamada M, Ohtomo G, Terao Y, Nakazato Y, Takeda A, Sakiyama Y, Umeda-Kameyama Y, Shinmi J, Ogata K, Kohno Y, Lim SY, Tan AH, Shimizu J, Goto J, Nishino I, Toda T, Morishita S, Tsuji S. Noncoding CGG repeat expansions in neuronal intranuclear inclusion disease, oculopharyngodistal myopathy and an overlapping disease. Nat Genet. 2019;51:1222–32. PubMed PMID: 31332380.

Kekou K, Sofocleous C, Papadimas G, Petichakis D, Svingou M, Pons RM, Vorgia P, Gika A, Kitsiou-Tzeli S, Kanavakis E. A dynamic trinucleotide repeat (TNR) expansion in the DMD gene. Mol Cell Probes. 2016;30:254–60. PubMed PMID: 27417533.

Kozlowski P, de Mezer M, Krzyzosiak WJ. Trinucleotide repeats in human genome and exome. Nucleic Acids Res. 2010;38:4027–39. PubMed PMID: 20215431.

Ranum LPW, Moseley ML, Leppet MF, et al. Massive CTG expansions and deletions may reduce penetrance of spinocerebellar ataxia type 8. Am J Hum Genet. 1999;65:466.

Scriba CK, Beecroft SJ, Clayton JS, Cortese A, Sullivan R, Yau WY, Dominik N, Rodrigues M, Walker E, Dyer Z, Wu TY, Davis MR, Chandler DC, Weisburd B, Houlden H, Reilly MM, Laing NG, Lamont PJ, Roxburgh RH, Ravenscroft G. A novel RFC1 repeat motif (ACAGG) in two Asia-Pacific CANVAS families. Brain. 2020;143:2904–10. PubMed PMID: 33103729.

Shibata A, Machida J, Yamaguchi S, Kimura M, Tatematsu T, Miyachi H, Matsushita M, Kitoh H, Ishiguro N, Nakayama A, Higashi Y, Shimozato K, Tokita Y. Characterisation of novel RUNX2 mutation with alanine tract expansion from Japanese cleidocranial dysplasia patient. Mutagenesis. 2016;31:61–7. PubMed PMID: 26220009.

Sone J, Mitsuhashi S, Fujita A, Mizuguchi T, Hamanaka K, Mori K, Koike H, Hashiguchi A, Takashima H, Sugiyama H, Kohno Y, Takiyama Y, Maeda K, Doi H, Koyano S, Takeuchi H, Kawamoto M, Kohara N, Ando T, Ieda T, Kita Y, Kokubun N, Tsuboi Y, Katoh K, Kino Y, Katsuno M, Iwasaki Y, Yoshida M, Tanaka F, Suzuki IK, Frith MC, Matsumoto N, Sobue G. Long-read sequencing identifies GGC repeat expansions in NOTCH2NLC associated with neuronal intranuclear inclusion disease. Nat Genet. 2019;51:1215–21. PubMed PMID: 31332381.

Stenson PD, Mort M, Ball EV, Chapman M, Evans K, Azevedo L, Hayden M, Heywood S, Millar DS, Phillips AD, Cooper DN. The Human Gene Mutation Database (HGMD®): optimizing its use in a clinical diagnostic or research setting. Hum Genet. 2020;139:1197–207. PubMed PMID: 32596782.

Takagi M, Ishii T, Torii C, Kosaki K, Hasegawa T. A novel mutation in SOX3 polyalanine tract: a case of Kabuki syndrome with combined pituitary hormone deficiency harboring double mutations in MLL2 and SOX3. Pituitary. 2014;17:569–74. PubMed PMID: 24346842.

Tian Y, Wang JL, Huang W, Zeng S, Jiao B, Liu Z, Chen Z, Li Y, Wang Y, Min HX, Wang XJ, You Y, Zhang RX, Chen XY, Yi F, Zhou YF, Long HY, Zhou CJ, Hou X, Wang JP, Xie B, Liang F, Yang ZY, Sun QY, Allen EG, Shafik AM, Kong HE, Guo JF, Yan XX, Hu ZM, Xia K, Jiang H, Xu HW, Duan RH, Jin P, Tang BS, Shen L.

Expansion of human-specific GGC repeat in neuronal intranuclear inclusion disease-related disorders. Am J Hum Genet. 2019;105:166–76. PubMed PMID: 31178126.

Tsuchiya M, Nan H, Koh K, Ichinose Y, Gao L, Shimozono K, Hata T, Kim YJ, Ohtsuka T, Cortese A, Takiyama Y. RFC1 repeat expansion in Japanese patients with late-onset cerebellar ataxia. J Hum Genet. 2020;65:1143–7. PubMed PMID: 32694621.

Wieben ED, Aleff RA, Eckloff BW, Atkinson EJ, Baheti S, Middha S, Brown WL, Patel SV, Kocher JP, Baratz KH. Comprehensive assessment of genetic variants within TCF4 in Fuchs' endothelial corneal dystrophy. Invest Ophthalmol Vis Sci. 2014;55:6101–7. PubMed PMID: 25168903.

License