S1 | ***Unstable repeats and associated disorders***

Disorder or repeat (gene)	Chromosome	repeat unit	Repeat Size			Parent-of-Origin Effect	Somatic Instability	Pathogenesis (LOF, GOF, RNA)
			Normal	Premutation	Disease			
Polyglutamine (CAG) tracts								
SBMA: Spinal and Bulbar Muscular Atrophy (Androgen Receptor)	Xq11-q12	(CAG)n•(CTG)n	9 - 36	ND	40-55	ND	limited	GOF
HD: Huntington's disease (Huntington)	4p16.3	(CAG)n•(CTG)n	10 - 34	29-35	> 35	Paternal	moderate	GOF
DRPLA: Dentatorubral-pallidouysian atrophy (Dentatorubral-pallidouysian atrophy)	12p13.31	(CAG)n•(CTG)n	7 – 25	ND	49-88	Paternal	moderate	GOF
SCA1: Spinocerebellar ataxia 1 (Ataxin-1)	6p23	(CAG)n•(CTG)n	6 - 39	ND	39-81	Paternal	moderate	GOF
SCA2: Spinocerebellar ataxia 2 (Ataxin-2)	12p24	(CAG)n•(CTG)n	13 - 33	ND	> 34	Paternal	moderate	GOF
SCA3: Spinocerebellar ataxia 3 (Ataxin-3)	14p24.3-q31	(CAG)n•(CTG)n	13 - 44	ND	> 55	Paternal	moderate	GOF
SCA6: Spinocerebellar ataxia 6 (Calcium channel alpha--1A subunit)	19p13	(CAG)n•(CTG)n	4 - 18	ND	20 - 29	none	ND	LOF
SCA7: Spinocerebellar ataxia 7 (Ataxin-7)	3p21.1-p12	(CAG)n•(CTG)n	4 - 35	ND	37-306	Paternal	moderate	GOF
SCA17: Spinocerebellar ataxia 17 (TATA-box binding protein)	6q27	(CAG)n•(CTG)n	25 - 42	ND	47 - 63	ND	ND	GOF
KCNN3: no confirmed disease association (small conductance calcium-activated potassium channel)	1q21.3	(CAG)n•(CTG)n	7 – 28	ND	ND	ND	ND	ND
AIB-I: Increased prostate cancer risk (Nuclear receptor coactivator 3)	20q13	(CAG/CAA)n• (CTG/TTG)n	29/29 or 28/29	<29/29	<29/<29	ND	limited	GOF
Polyalanine (GCG) tracts								
HOXD13: Synpolydactyly (Homeobox D13)	2q31-q32	(GCG)n•(CGC)n	15	ND	22 - 29	na	none	GOF
OPMD: Oculopharyngeal Muscular Dystrophy (poly(A)-binding protein-2)	14q11.2-q13	(GCG)n•(CGC)n	10	ND	12 - 17	na	none	GOF
CBFA1:cleidocranial dysplasia (runt-related transcription factor 2)	6p21	(GCG)n•(CGC)n	17	ND	27	na	none	LOF
ZIC2:holoprosencephaly (zinc finger protein of cerebellum 2)	13q32	(GCG)n•(CGC)n	15	ND	25	na	none	LOF
HOXA13: Hand-Foot-Genital Syndrome (homeobox A13)	7p15-p14.2	(GCG)n•(CGC)n	18	ND	24 - 26	na	none	GOF ?
FOXL2:Blepharophimosis/Ptosis/Epicanthus inversus syndrome type II (forkhead transcription factor FOXL2)	3q23	(GCG)n•(CGC)n	14	ND	22 - 24	na	none	LOF
ARX: infantile spasm syndrome (aristaless-related homeobox, X linked)	Xp22.13	(GCG)n•(CGC)n	10-16	ND	17-23	na	none	LOF
Other coding repetitive elements								
CJD:Creutzfeldt-Jakob disease (Prion	20pter-p12,	24-bp repeat	4	ND	2,5 - 13	ND	ND	GOF

protein)	6p21.3							
COMP: Multiple Skeletal dysplasias (cartilage oligomeric matrix protein)	19p13.1	(GAC)n•(GTC)n	5	ND	4,6,7	ND	ND	LOF
Non-coding trinucleotide repeats								
CTG18.1: no confirmed disease association (transcription factor 4)	18q21	(CTG)n•(CAG)n	10 - 37	53 - 250	800 - 2100	ND	ND	ND
DM1: myotonic dystrophy type 1 (dystrophia myotonia protein kinase)	19q13.2-q13.3	(CTG)n•(CAG)n	5 - 37	34-90	> 90	Maternal	extensive	RNA
FRDA: Friedreich's ataxia (frataxin)	9q13	(GAA)n•(TTC)n	6 - 32	40-200	> 200	Maternal	moderate	LOF
SCA8: Spinocerebellar ataxia 8 (SCA8 gene)	13q21	(CTG)n•(CAG)n	2 > 130	45 - 109	> 110	Exp: Mat Cnt: Pat	ND	unknown
SCA12: Spinocerebellar ataxia 12 (protein phosphatase 2, regulatory subunit B)	5q31-q33	(CAG)n•(CTG)n	7 - 45	ND	55 - 78	ND	ND	ND
HDL2: Huntington's disease-like 2 (junctophilin-3)	16q24.3	(CAG)n•(CTG)n	6 - 27	ND	51 - 57	ND	ND	unknown
MAB21L1:no confirmed disease association (MAB21, C. elegans, homologue-like 1)	13q13	(CAG)n•(CTG)n	6 - 31	ND	> 50	ND	limited	ND
Rare fragile site-associated repeats *Folate-Sensitive Sites*								
FRAXA: fragile X syndrome (fragile site mental retardation 1 gene)	Xq27.3	(CGG)n•(CCG)n	6 - 52	59 - 230	230 - 2000	Exp: Mat Cnt: Pat	limited	LOF, RNA ?
FRAXE: fragile X syndrome (fragile site mental retardation 2 gene & fragile site mental retardation 3 gene)	Xq28	(CCG)n•(CGG)n	4 - 39	(31-61)	200-900	Exp: Mat Cnt: Pat	ND	LOF
FRAXF: No confirmed disease association (family with sequence similarity 11, member A)	Xq28	(CGG)n•(CCG)n	7 - 40	ND	306-1008	ND	ND	ND
FRA10A: No confirmed disease association (FRA10A candidate gene 1)	10q23	(CCG)n•(CGG)n	8 - 14	ND	> 200	ND	ND	ND
FRA11B: Jacobsen syndrome (CAS-BR-M murine ectopic retroviral transforming sequence homology)	11q23	(CCG)n•(CGG)n	11	80	100-1000	ND	ND	LOF

The relative genic location of each repeat is listed in Fig. 1a. For the listed repeat sequences the first repeat is that of the coding strand. ND: insufficient data, na: not applicable, LOF: loss of function, GOF: gain of function, RNA: toxic RNA