Table S1. Characteristics of disease-associated short tandem repeats (daSTRs) and their associated genes. Related to Figure 1, 2, 3, 4

Gene[1,4]	Disease	Repeat Sequence	Repeat Location	Repeats for Disease[2]	Repeats Found in hg19[3]	Number of Isoforms	RefSeq IDs[4]
FMR1	Fragile X syndrome	CGG	5' UTR	>200	10	5	NM_001185075, NM_001185076, NM_001185081, NM_001185082, NM_002024
PPP2R2B	Spinocerebellar ataxia 12	CAG	Promoter, 5' UTR	>50	10	8	NM_181677, NM_181676, NM_001271899, NM_181674, NM_181678, NM_001271900, NM_001271948, NM_181675
ATXN1	Spinocerebellar ataxia 1	CAG	Exon	>30	14	2	NM_000332, NM_001128164
ATXN2	Spinocerebellar ataxia 2	CAG	Exon	>30	13	1	NM_002973
ATXN3	Spinocerebellar ataxia 3	CAG	Exon	>30	8	27	NM_001127697, NR_028456, NR_028455, NM_001164778, NM_001164777, NM_001164776, NM_001164774, NM_004993, NR_028468, NR_028467, NR_028470, NR_028469, NR_028464, NR_028465, NR_028458, NR_028460, NR_028462, NM_030660, NR_031765, NR_028463, NM_001164782, NM_001164781, NR_028457, NR_028461, NM_001164779, NM_001164780, NM_001127696
CACNA1A	Spinocerebellar ataxia 6	CAG	Exon	>30	13	5	NM_000068, NM_023035, NM_001174080, NM_001127222, NM_001127221
ATXN7	Spinocerebellar ataxia 7	CAG	Exon	>30	10	3	NM_001177387, NM_000333, NM_001128149
HTT	Huntington's disease	CAG	Exon	>30	19	1	NM_002111
AR	Spinal and bulbar muscular atrophy	CAG	Exon	>30	22	2	NM_000044, NM_001011645
ATN1	Dentatorubral-pallidoluysian atrophy	CAG	Exon	>50	15	2	NM_001940, NM_001007026
FXN	Friedreich's ataxia	GAA	Intron	>50	6	3	NM_001161706, NM_000144, NM_181425

CNBP	Myotonic dystrophy 2	CCTG	Intron	>50	8	4	NM_003418, NM_001127196, NM_001127194, NM_001127192
ATXN10	Spinocerebellar ataxia 10	ATTCT	Intron	>200	14	2	NM_001167621, NM_013236
BEAN1	Spinocerebellar ataxia 31	TGGAA	Intron	>100	None	4	NM_001136106, NM_001178020, NM_001197224, NM_001197225
NOP56[5]	Spinocerebellar ataxia 36	GGCCTG	Intron	>1000	4	1	NM_006392
C9ORF72	Amyotrophic lateral sclerosis	GGGGCC	Intron	>50	3	3	NM_018325, NM_145005, NM_001256054
COMP*	Multiple skeletal dysplasias	GAC	Exon	>10	5	1	NM_000095
HOXD13*	Synpolydactyly syndrome	GCG	Exon	>20	5	1	NM_000523
HOXA13*	Hand-foot-genital syndrome	GCG	Exon	>20	3	1	NM_000522
RUNX2	Cleidocranial dysplasia	GCG	Exon	>25	5	2	NM_001024630, NM_001015051
ZIC2	Holoprosencephaly	GCG	Exon	>20	5	1	NM_007129
PABPN1*	Oculopharyngeal muscular atrophy	GCG	Exon	>10	6	1	NM_004643
FOXL2	Blepharophimosis, ptosis, epicanthus inversus syndrome	GCG	Exon	>20	3	1	NM_023067
ARX	ARX-related X-linked mental retardation	GCG	Exon	>20	10	1	NM_139058
DMPK	Myotonic dystrophy 1	CTG	3' UTR	>75	20	4	NM_001081562, NM_001081563, NM_004409, NM_001081560
ATXN8OS	Spinocerebellar ataxia 8	CTG	3' UTR	>100	15	1	NR_002717
JPH3	Huntington's disease-like 2	CTG	3' UTR, exon	>30	14	3	NM_020655, NM_001271605 NM_001271604
CSTB	Myoclonic epilepsy of Unverricht and Lundborg	CCCCGC CCCGCG	Promoter	Unknown	3	1	NM_000100
TBP[5]	Spinocerebellar ataxia 17	CAG	Exon	>30	19	2	NM_003194, NM_001172085

[1,2] From Iyer et al. Annu Rev Biochem 2015, La Spada et al. Nat Rev Genet 2010.
[3]See Methods, Table S3.
[4]From UCSC Genome Browser, accessed 12/28/2016.
[5]Genome folding not queried due to chromosomal placement, excluded from analysis.
*Repeat expansion not linked to neurological dysfunction

Table S2. Location and sequence of disease-associated short tandem repeats (daSTRs) in hg19. Related to Figure 2.

Gene	Chrom	Start	End	Strand	Sequence
FMR1	chrX	146993568	146993598	+	cggcggcggcggcggcggcggcggcggcgg
PPP2R2B	chr5	146258291	146258321	-	cagcagcagcagcagcagcagcagcagcag
ATXN1	chr6	16327866	16327908	-	cagcagcagcagcagcagcagcagcagcagcagcagcagcag
ATXN2	chr12	112036784	112036823	-	cagcagcagcagcagcagcagcagcagcagcagcag
ATXN3	chr14	92537354	92537378	-	cagcagcagcagcagcagcagcag
CACNA1A	chr19	13318672	13318711	-	cagcagcagcagcagcagcagcagcagcagcagcag
ATXN7[1]	chr3	63898361	63898391	+	cagcagcagcagcagcagcagcagcagcag
TBP[2]	chr6	170871046	170871103	+	cagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag
HTT	chr4	3076603	3076660	+	cagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag
AR[3]	chrX	66765159	66765225	+	cagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag
ATN1	chr12	7045891	7045936	+	cagcagcagcagcagcagcagcagcagcagcagcagcagcagcag
FXN	chr9	71652202	71652220	+	gaagaagaagaagaagaa
CNBP	chr3	128891467	128891499	-	cctgcctgcctgcctgcctgcctgcctgcctg
ATXN10	chr22	46191234	46191304	+	attctattctattctattctattctattctattctattctattctattctattctattct
BEAN1[4]	chr16	66524302	66524303	+	
NOP56[2]	chr20	2633379	2633403	+	ggcctgggcctgggcctgggcctg
C9ORF72[5]	chr9	27573526	27573544	-	ggggccggggccggggcc
COMP	chr19	18896844	18896859	-	gacgacgacgacgac
HOXD13	chr2	176957810	176957825	+	gcggcggcggcggcg
HOXA13	chr7	27239453	27239462	-	gcggcggcg
RUNX2	chr6	45390502	45390517	+	gcggcggcggcggcg
ZIC2	chr13	100634393	100634408	+	gcggcggcggcggcg
PABPN1	chr14	23790681	23790699	+	gcggcggcggcggcggcg
FOXL2	chr3	138665200	138665209	-	gcggcggcg
ARX	chrX	25031778	25031808	-	gcggcggcggcggcggcggcggcggcggcg
DMPK	chr19	46273462	46273522	-	ctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctg
ATXN8OS	chr13	70713515	70713560	+	ctgctgctgctgctgctgctgctgctgctgctgctgctgctg
JPH3	chr16	87637893	87637935	+	ctgctgctgctgctgctgctgctgctgctgctgctg
CSTB	chr21	45196324	45196360	-	ccccgccccgcgccccgccccgcgccccgccccgcg

[1]Isoform NM_001128149 does not contain repeat (ATXN7)
[2]Genome folding not queried due to chromosomal placement, excluded from analysis (TBP)
[3]Isoform NM_001011645 does not contain repeat (AR)
[4]Repeat not found in hg19, location for analyses determined from Sato et al 2009
[5]Isoform NM_018325 does not contain repeat (C9ORF72)

Table S3: Properties of four classes of disease-associated STRs (daSTRs). Related to Figure 2, 4

Group label	Group description	daSTR genes	Repeat unit	Location	Number of matched repeats found in hg19	Range of repeat units in matched repeats
A	CAG Exons and 5'UTR	AR, ATN1, ATXN1, ATXN2, ATXN3, ATXN7, CACNA1A, HTT	CAG	Exon	626	5 - 19
		PPP2R2B	CAG	5'UTR	160	5 - 22
B	GCG Exons and 5'UTR	ARX, HOXA13, HOXD13, FOXL2, PABPN1, RUNX2, ZIC2	GCG	Exon	531	5 - 17
		FMR1	CGG	5'UTR	737	5 - 16
C	GAA/CCTG/ ATTCT/ GGCCTG/GG CCCC Introns	FXN	GAA	Intron	2,166	5 - 32
		BEAN1	GGCCTG	Intron	10	5 - 66
		CNBP	CCTG	Intron	215	5 - 13
		ATXN10	ATTCT	Intron	26	5 - 16
		C9ORF72	GGCCCC	Intron	8	5 - 8
D	CTG 3'UTR	DMPK, ATXN8OS, JPH3	CTG	3'UTR	113	5 - 20
*	GAC Exon	COMP	GAC	Exon	30	5 - 9

*Excluded from group-based analysis. See Methods.

Table S5. Description of cell lines and tissue samples used for 5C and CTCF ChIP-seq. Related to Figures 5, 6.

Coriell catalog ID/Sample ID	Cell/Tissue type	Disease status	Number of repeats	Age	Race	Gender	Notes
GM09236	B-Lymphocyte	Healthy	<45	4	Black	Male	Brother of GM09237
GM09237	B-Lymphocyte	Disease	931-940	5	Black	Male	
GM04024	Fibroblast	Disease	645	29	Black	Male	Derived from same patient as GM04025
GM04025	B-Lymphocyte	Disease	645	29	Black	Male	
AG06103	Fibroblast	Healthy	<45	29	Black	Male	
GM06890	B-Lymphocyte	Healthy	<45	19	White	Male	Brother of GM06897
GM06897	B-Lymphocyte	Disease	477	25	White	Male	
FXS Case 1	Cerebellum	Disease	Full mutation	60	Unknown	Male	PMI* 54.5h
Control Case 1	Cerebellum	Healthy	<45	62	Unknown	Male	PMI* 9.5h
FXS Case 2	Cerebellum	Disease	Full mutation	74	Unknown	Male	PMI* 40-44h
Control Case 2	Cerebellum	Healthy	<45	69	Unknown	Male	PMI* 9.5h

*Post Mortem Interval