The American Journal of Human Genetics, Volume *104*

Supplemental Data

Length of Uninterrupted CAG, Independent of Polyglutamine Size, Results in Increased Somatic Instability, Hastening Onset of Huntington Disease

Galen E.B. Wright, Jennifer A. Collins, Chris Kay, Cassandra McDonald, Egor Dolzhenko, Qingwen Xia, Kristina Bečanović, Britt I. Drögemöller, Alicia Semaka, Charlotte M. Nguyen, Brett Trost, Fiona Richards, Emilia K. Bijlsma, Ferdinando Squitieri, Colin J.D. Ross, Stephen W. Scherer, Michael A. Eberle, Ryan K.C. Yuen, and Michael R. Hayden

SUPPLEMENTARY INFORMATION

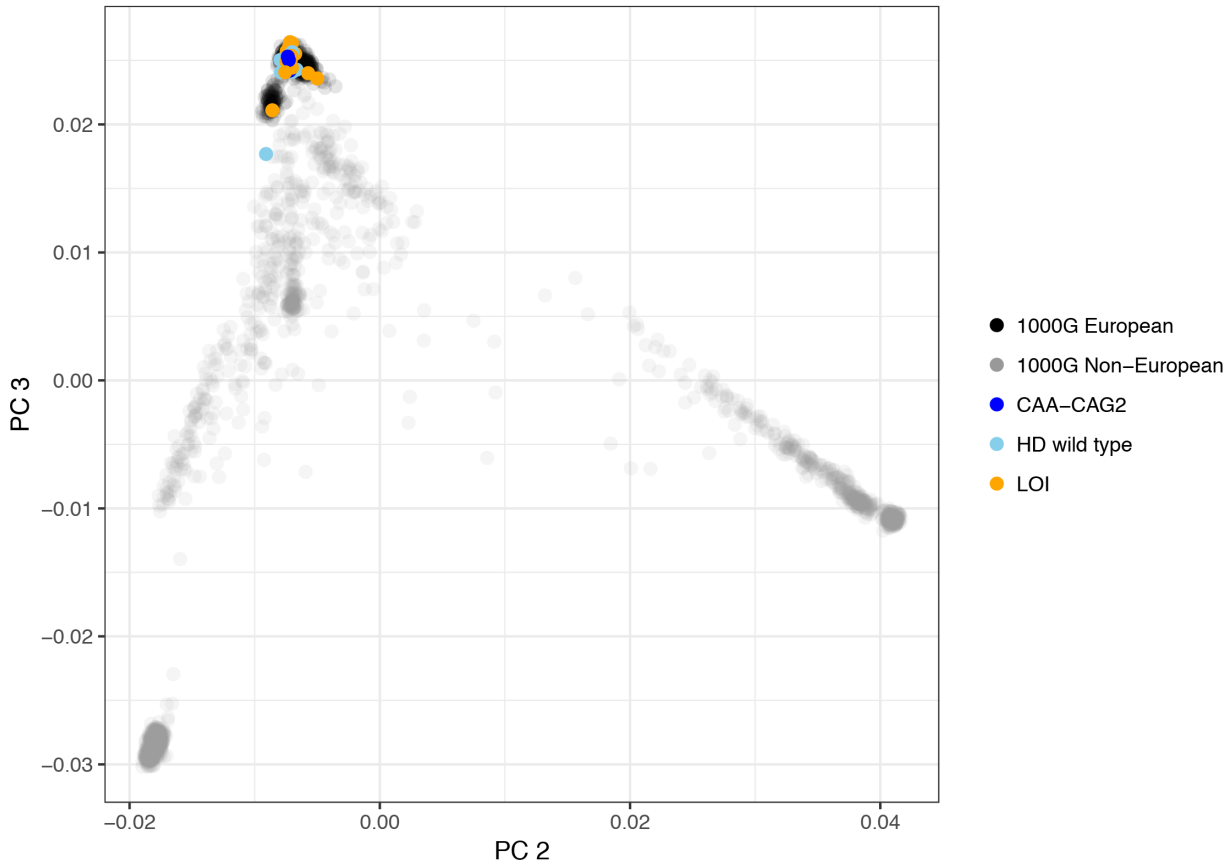

Figure S1 Principal component analysis of individuals with Huntington disease with genome-wide array data confirms that the loss of interruption (LOI) and $(CAA\text{-}CAG)_2$ carriers are of European genetic ancestry. Information for the 1000 Genomes Project Phase 3 samples are included as a reference. A total of 44 individuals with HD were assessed in this manner: LOI (*n*=21), $(CAA\text{-}CAG)_2$ (*n*=5) and HD wild type (*n*=18).

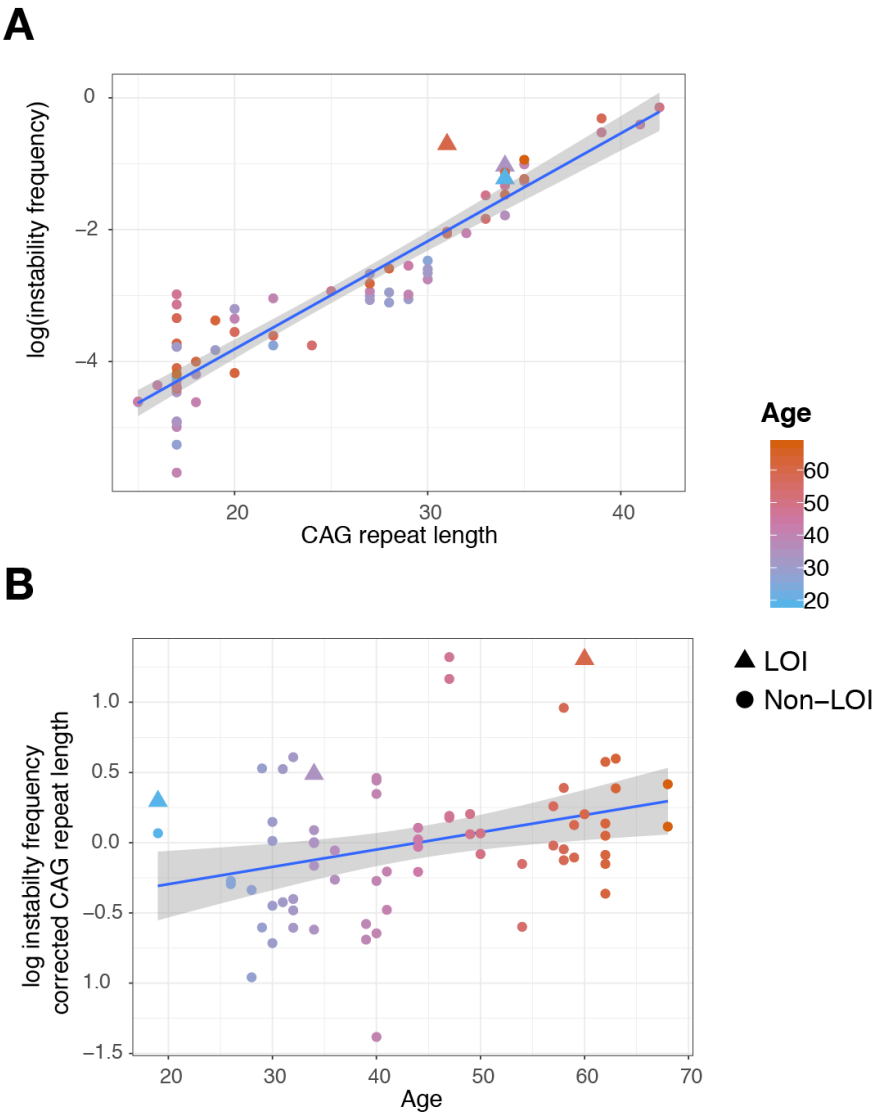

Figure S2 The *HTT* CAG-CCG loss of interruption (LOI) is associated with an increased frequency of CAG instability in sperm. **(A)** Exponential relationship between germline instability frequency and CAG repeat length, measured by small-pool PCR (variance explained by progenitor CAG repeat length, R^2 = 0.87). **(B)** Instability frequency corrected for CAG length showing the effect of age (variance explained by age, R^2 = 0.11). Instability for LOI subjects are indicated (one of the LOI individuals with HD was sampled at two separate time points). Points are colored by age at time of sampling.

Table S1. Statistical analysis of somatic and germline measures of *HTT* CAG instability. The *HTT* CAG-CCG loss of interruption (LOI), as well as age and CAG size were associated with increased instability in these semi-quantitative analyses.

Trait	Variable	*β*-coefficient	*P*-value
Expansion frequency (small-pool PCR in sperm)[a]	CAG repeat length	0.19	1.6×10^{-35}
	Age	0.01	0.01
	LOI	0.94	0.001
Expansion ratio (genomic DNA from whole blood)[a]	CAG repeat length	0.15	3.0×10^{-31}
	Age	0.006	2.5×10^{-3}
	LOI	0.43	3.5×10^{-9}

[a]log transformed; LOI, loss of interruption

Table S2 Perfect tag variants (R^2=1, D'=1) for the loss of interruption (LOI) variant sub-haplotypes. No perfect tag variant was found for LOI A1 CCG_{10}, that was observed in one of the pedigrees on reduced and fully penetrant alleles (i.e. pedigree HD-LOI-02).

Modifier variant	*HTT* haplotype	rsID	REF	ALT	R^2	D'	Distance (bp from *HTT* CAG-CCG)	HRC frequency
LOI	A1 CCG_{10}	rs145048189	C	T	1	1	772333	5.90E-03
LOI	A1 CCG_{10}	rs143157739	G	A	1	1	788115	8.78E-03
LOI	A1 CCG_{10}	rs141521686	G	A	1	1	829893	1.42E-02
LOI	A1 CCG_{10}	rs148396437	T	C	1	1	842161	8.32E-03
LOI	A1 CCG_{10}	rs143200453	C	G	1	1	891439	7.85E-03
LOI	A1 CCG_{10}	rs143751494	C	T	1	1	899366	5.71E-03
LOI	A1 CCG_{10}	rs12646393	T	C	1	1	942706	1.68E-02
LOI	C1	rs193119731	A	G	1	1	325103	5.84E-03
LOI	C1	rs764154313	G	A	1	1	379005	5.85E-04
LOI	C1	rs993019491	A	G	1	1	542551	2.93E-04
LOI	C1	rs772789339	A	G	1	1	745521	4.16E-04
LOI	C1	rs138025536	G	A	1	1	771127	4.47E-04

ALT: alternate allele; LOI: loss of interruption; HRC: Haplotype Reference Consortium; REF: reference allele
LOI A1 CCG_7 (*n*=4), LOI A1 CCG_{10} (*n*=6) and LOI C1 (*n*=12)

Table S3 Perfect tag variants (R^2=1, D'=1) for the $(CAA\text{-}CAG)_2$ HD modifier variant (*n*=5 individuals) located within 10 kb of the *HTT* CAG-CCG interrupting sequence (*n*=35)

Modifier variant	*HTT* haplotype	rsID	REF	ALT	R^2	D'	BP from *HTT* CAG-CCG	HRC frequency
$(CAA\text{-}CAG)_2$	C2	rs10006977	A	C	1	1	289	2.85E-02
$(CAA\text{-}CAG)_2$	C2	rs10009935	T	C	1	1	371	3.75E-02
$(CAA\text{-}CAG)_2$	C2	rs28571971	G	C	1	1	1340	3.49E-02
$(CAA\text{-}CAG)_2$	C2	rs28583447	T	C	1	1	1341	3.48E-02
$(CAA\text{-}CAG)_2$	C2	rs28468636	C	G	1	1	1381	3.64E-02
$(CAA\text{-}CAG)_2$	C2	rs28564368	C	A	1	1	1396	3.65E-02
$(CAA\text{-}CAG)_2$	C2	rs28485764	G	A	1	1	1425	3.71E-02
$(CAA\text{-}CAG)_2$	C2	rs77173925	A	G	1	1	1581	3.71E-02
$(CAA\text{-}CAG)_2$	C2	rs112435590	G	T	1	1	2563	3.75E-02
$(CAA\text{-}CAG)_2$	C2	rs28377140	G	C	1	1	3240	4.41E-02
$(CAA\text{-}CAG)_2$	C2	rs10014333	T	A	1	1	3310	2.85E-02
$(CAA\text{-}CAG)_2$	C2	rs10006129	G	C	1	1	4039	3.73E-02
$(CAA\text{-}CAG)_2$	C2	rs28696693	A	G	1	1	4144	3.71E-02
$(CAA\text{-}CAG)_2$	C2	rs28755900	G	C	1	1	4226	2.83E-02
$(CAA\text{-}CAG)_2$	C2	rs28393280	A	G	1	1	4976	3.57E-02
$(CAA\text{-}CAG)_2$	C2	rs6835897	G	C	1	1	5115	3.57E-02
$(CAA\text{-}CAG)_2$	C2	rs7436457	T	G	1	1	5436	3.57E-02
$(CAA\text{-}CAG)_2$	C2	rs28398130	C	G	1	1	5686	3.37E-02
$(CAA\text{-}CAG)_2$	C2	rs28682489	T	C	1	1	5702	3.38E-02
$(CAA\text{-}CAG)_2$	C2	rs4346595	T	C	1	1	5757	3.57E-02
$(CAA\text{-}CAG)_2$	C2	rs28714390	C	T	1	1	5797	3.37E-02
$(CAA\text{-}CAG)_2$	C2	rs28629394	G	A	1	1	5832	3.38E-02
$(CAA\text{-}CAG)_2$	C2	rs77099632	C	T	1	1	6497	3.38E-02
$(CAA\text{-}CAG)_2$	C2	rs141794700	A	G	1	1	6727	3.38E-02
$(CAA\text{-}CAG)_2$	C2	rs28394705	C	G	1	1	6741	3.38E-02
$(CAA\text{-}CAG)_2$	C2	rs28584232	C	T	1	1	6839	3.37E-02
$(CAA\text{-}CAG)_2$	C2	rs6830019	A	T	1	1	7217	3.31E-02
$(CAA\text{-}CAG)_2$	C2	rs7664480	C	A	1	1	7603	3.56E-02
$(CAA\text{-}CAG)_2$	C2	rs113748015	A	G	1	1	8717	2.83E-02
$(CAA\text{-}CAG)_2$	C2	rs10222986	G	A	1	1	8729	2.85E-02
$(CAA\text{-}CAG)_2$	C2	rs10222725	T	C	1	1	8730	2.85E-02
$(CAA\text{-}CAG)_2$	C2	rs80093929	G	A	1	1	9113	3.22E-02
$(CAA\text{-}CAG)_2$	C2	rs74658198	C	G	1	1	9433	2.85E-02
$(CAA\text{-}CAG)_2$	C2	rs116795936	G	A	1	1	9459	2.82E-02
$(CAA\text{-}CAG)_2$	C2	rs111382734	G	A	1	1	9553	2.91E-02

ALT: alternate allele; LOI: loss of interruption; HRC: Haplotype Reference Consortium; REF: reference allele