The American Journal of Human Genetics

Supplemental Data

Sequence-Level Analysis of the Major European Huntington Disease Haplotype

Jong-Min Lee, Kyung-Hee Kim, Aram Shin, Michael J. Chao, Kawther Abu Elneel, Tammy Gillis, Jayalakshmi Srinidhi Mysore, Julia A. Kaye, Hengameh Zahed, Ian H. Kratter, Aaron C. Daub, Steven Finkbeiner, Hong Li, Jared C. Roach, Nathan Goodman, Leroy Hood, Richard H. Myers, Marcy E. MacDonald, and James F. Gusella

Figure S1. Locations of SNPs used for haplotype analysis.

Allele definitions and approximate locations of variants defining the top HD expanded (i.e., hap.01) and normal (i.e., hap.08) haplotypes are shown. D and R denote deletion and reference, respectively. Red rectangle indicates the location of the *HTT* CAG repeat.

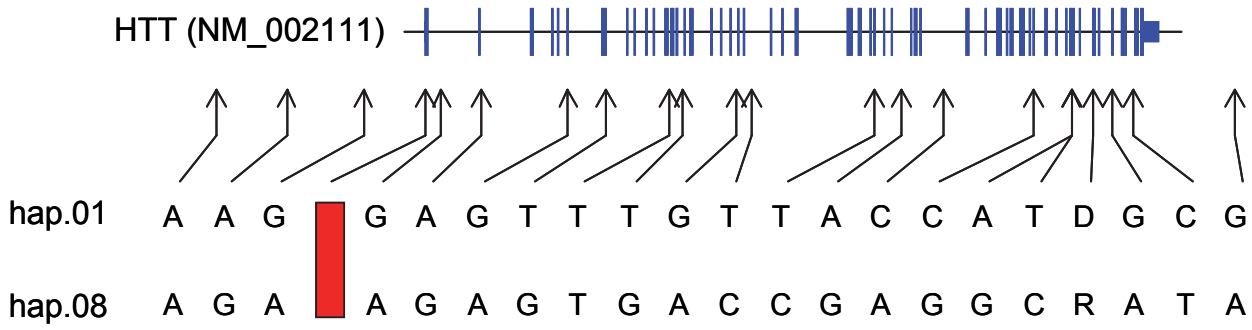


Figure S2. Ancestry of study subjects.

To assess the genetic background of our family samples, we compared our sequence data with HapMap data (http://hapmap.ncbi.nlm.nih.gov; Phase 3). We estimated pair-wise IBD using the PLINK program, then performed principal components analysis to visualize ancestry of samples relative to well characterized control samples with known ancestry. The first two PLINK-generated MDS (multi dimension scaling) values were sufficient to distinguish Europeans from Africans or Asians. Study subjects are indicated by open black circles, which co-localized with CEPH and TSI.

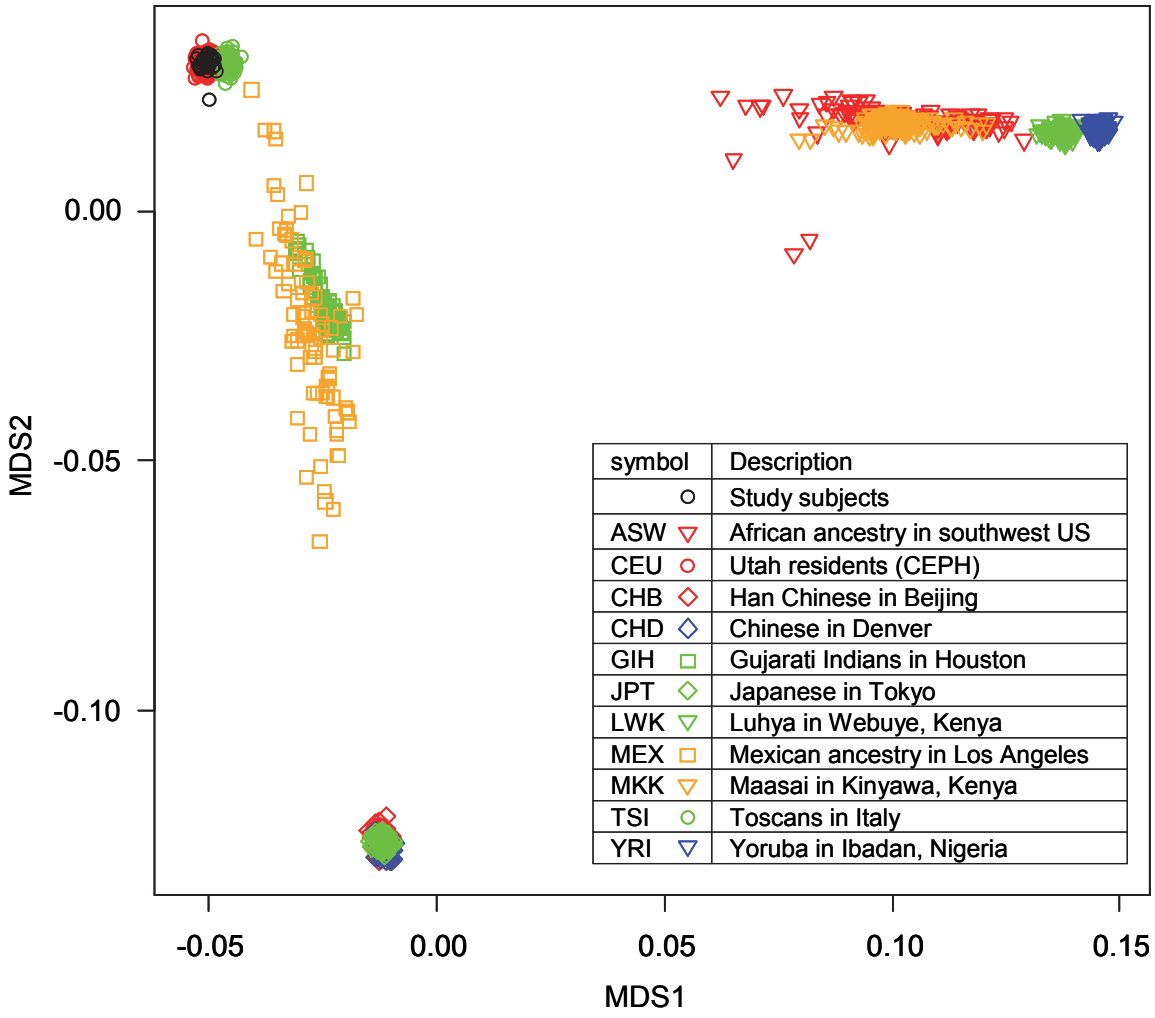


Figure S3. Summary of familial relationships of study subjects.

To confirm familial relationships within and between families, we analyzed the estimated IBD values between HD subjects estimated by PLINK program. We calculated the proportion of genome sharing between pairs of HD individuals within family and across family. All possible meaningful pairs of HD individuals were analyzed to summarize the intra-family PLINK PI_HAT values and inter-family PLINK PI_HAT values. Mean, minimum and maximum PLINK PI_HAT values are presented. Numbers in parentheses represent numbers of subjects with expanded CAG repeats in a given family regardless of membership in the trios.

Mean PI_HAT (min : max)	Family 1 (12)	Family 2 (7)	Family 3 (6)	Family 4 (4)
Family 1(12)	0.359 (0.121 : 0.593)	0.054 (0 : 0.078)	0.055 (0.046 : 0.066)	0.057 (0.047 : 0.067)
Family 2 (7)		0.438 (0.269 : 0.582)	0.055 (0.045 : 0.075)	0.058 (0.045 : 0.079)
Family 3 (6)			0.431 (0.250 : 0.562)	0.053 (0.045 : 0.061)
Family 4 (4)				0.438 (0.207 : 0.539)

Figure S4. Inconsistent alleles in 4 families.

For a given family, trio-specific haplotype phasing generated multiple HD haplotypes. Comparison of those haplotypes within a family revealed inconsistent sites which were generated by combination of missing genotypes, fully heterozygous genotypes, and sequencing errors. N, D, and I denote ambiguous allele, deletion and insertion, respectively.

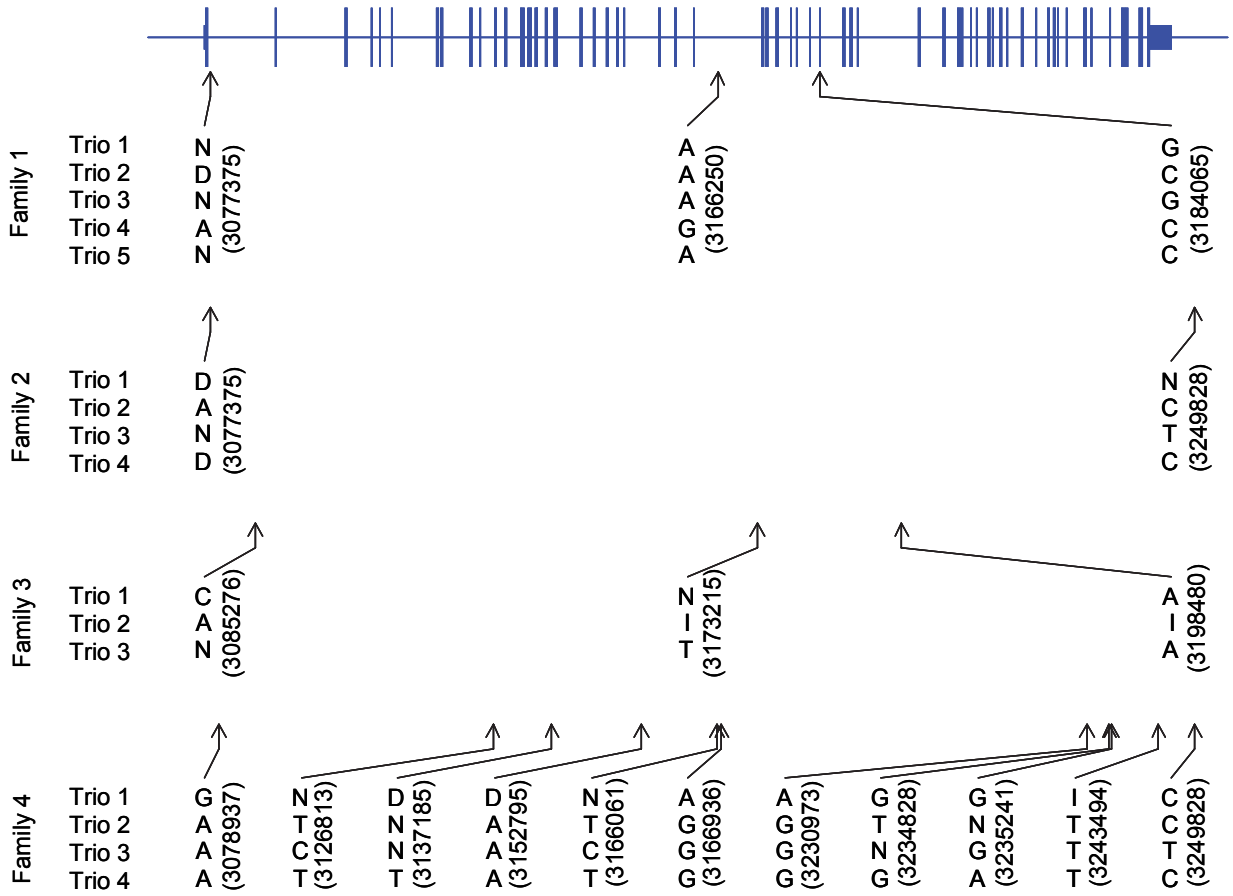


Figure S5. Validation of novel variants.

Seven family-specific variations were validated by independent Sanger sequencing or SNP genotyping. Representative Sanger sequencing results of the 3 novel variants, marked by arrows, is provided. Reverse and forward mean the direction of sequencing.

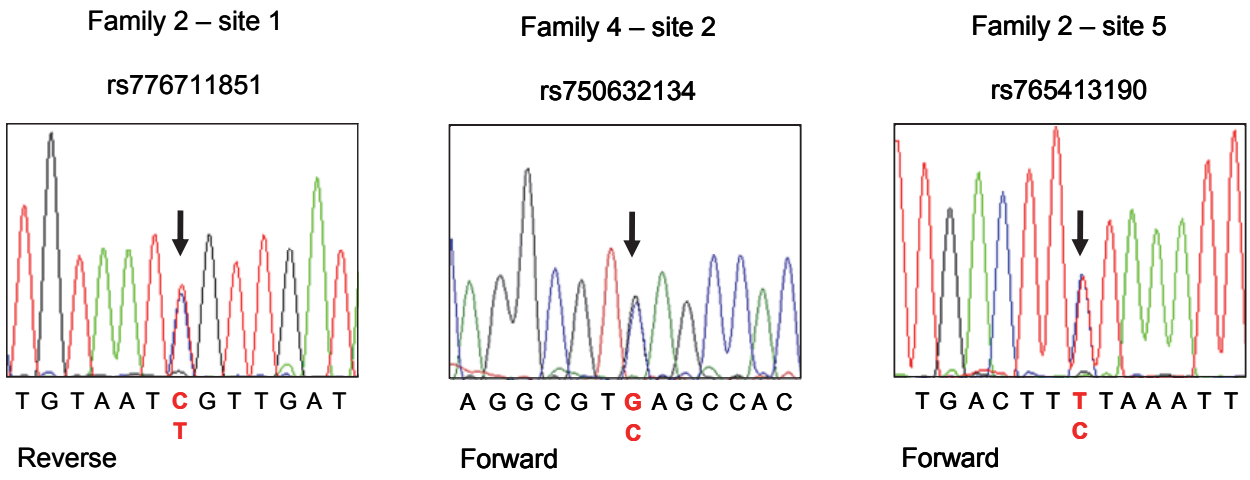
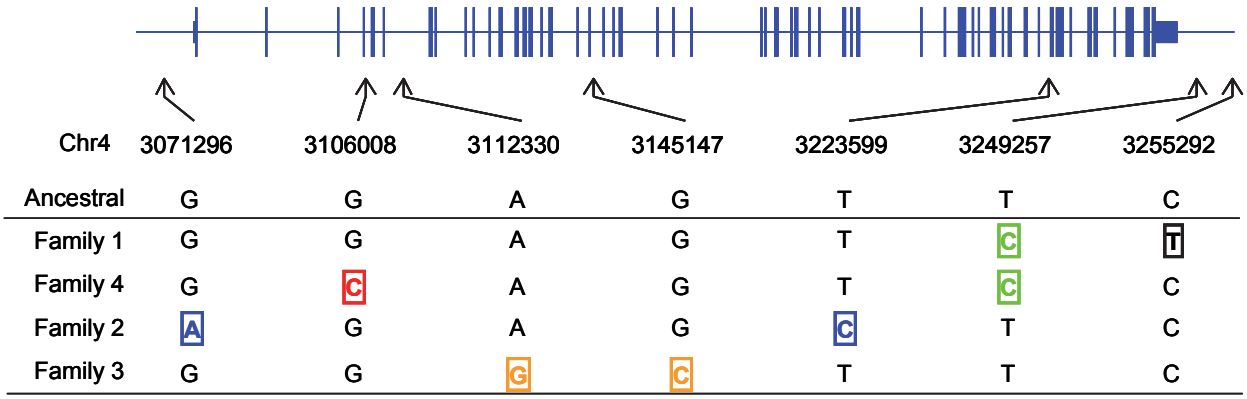


Figure S6. Historical relationship of family-specific variants.

The most parsimonious historical relationship of the disease haplotypes in the 4 families is shown. A genetic event (represented as a green line and letters) introduced a C allele at chr4: 3249257, distinguishing family 1 and family 4 from the shared ancestral hap.01 haplotype. Family 1 and family 4 experienced further genetic changes: a T allele at chr4:3255292 (a black letter and line) and a C allele at chr4:3106008 (a red letter and line). Independently, two different genetic changes on the ancestral hap.01 haplotype yielded family 2 (blue letters and lines) and family 3 (orange letters and lines) haplotypes, respectively.



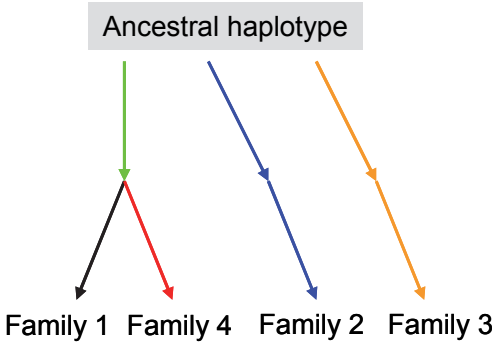


Figure S7. Recent CAG expansion mutations on normal hap.01 and the closely related hap.05.

De novo expansion mutations moving the *HTT* CAG repeat from the normal range (<36 CAGs) into the HD range (>35 CAGs) were reported previously.[22] We performed SNP genotyping, confirming that the CAG expansion new mutations occurred on haplotype hap.01 (left) and the closely related hap.05 (right). In the family on the left, in addition to square (male) and circle (female), triangular symbols are used to maintain confidentiality. Reconstruction from marker typing permitted 6 chromosomes to be distinguished (numbered arbitrarily 1-6 in brackets). The new mutation individual possess 50 CAG repeats and hap.01 on chromosome 1, which is shared by his sibling and aunt/uncle, who each have only 34 CAG repeats. In the family on the right, expansion of a 35 CAG repeat to 43 occurred during transmission of hap.05 from the father.

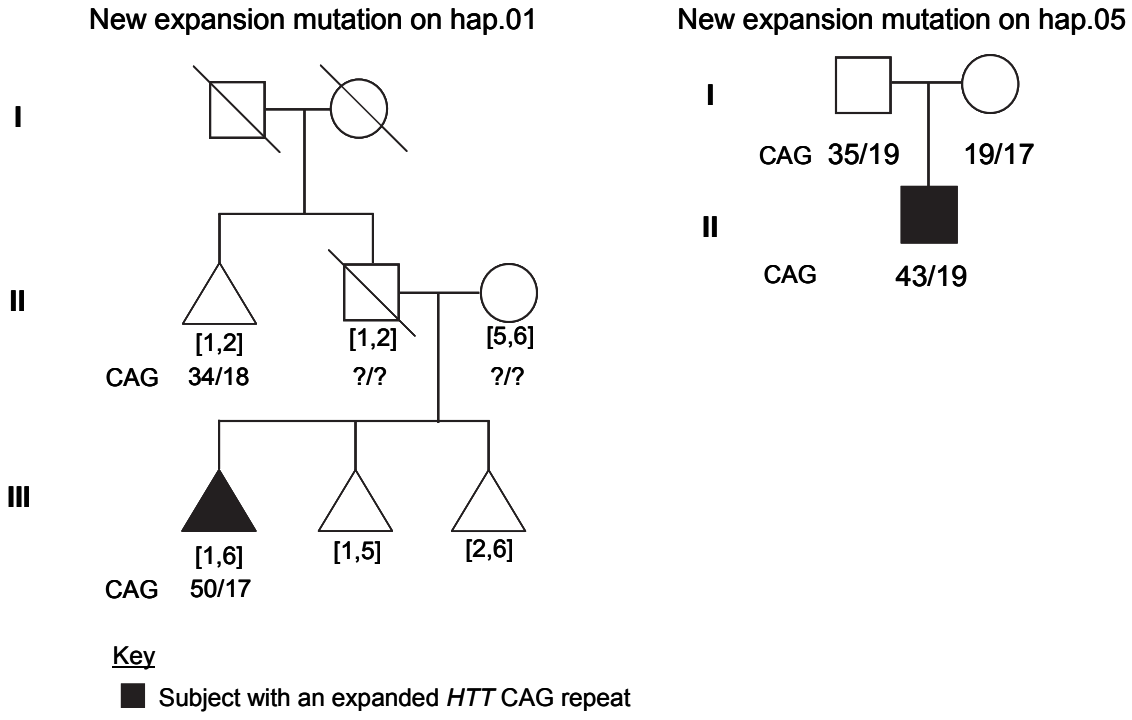


Figure S8. Sites different between hap.01 and hap.08.

A small number of individuals among our study subjects carried hap.08 as normal chromosomes, permitting reconstruction of the high resolution haplotype of this most common normal HTT haplotype. Finalized hap.01 alleles were compared to those of hap.08, revealing 109 sites that were different between them. For those 109 sites, reference, hap.01 and hap.08 alleles are provided with approximate locations. I, D, R, and S represent insertion, deletion, reference, and substitution, respectively.