The American Journal of Human Genetics, Volume *90*

Supplemental Data

Common SNP-Based Haplotype Analysis

of the 4p16.3 Huntington Disease Gene Region

Jong-Min Lee, Tammy Gillis, Jayalakshmi Srinidhi Mysore, Eliana Marisa Ramos, Richard H. Myers, Michael R. Hayden, Patrick Morrison, Martha Nance, Christopher A. Ross, Russell L. Margolis, Ferdinando Squitieri, Annamaria Griguoli, Stefano Di Donato, Estrella Gomez-Tortosa, Carmen Ayuso, Oksana Suchowersky, Ronald J. Trent, Elizabeth McCusker, Andrea Novelletto, Marina Frontali, Randi Jones, Tetsuo Ashizawa, Samuel Frank, Marie-Helene Saint-Hilaire, Steven M. Hersch, Herminia D. Rosas, Diane Lucente, Madaline B. Harrison, Andrea Zanko, Ruth K. Abramson, Karen Marder, Jorge Sequeiros, Marcy E. MacDonald, and James F. Gusella

Figure S1. Quantile-Quantile Plot of HD case:control association analysis

Among 436,185 SNPs that passed our quality control filters, 392,519 SNPs were analyzed in the Plink program after applying the default setting for the minimum sample number for each genotype (i.e., minimum 5 samples per genotype). The negative logarithms of observed p-values in the dominant model association analysis (Y-axis) were compared to those of expected p-values (X-axis) calculated based upon the theoretical distribution, revealing a genomic inflation factor of 1.04645.

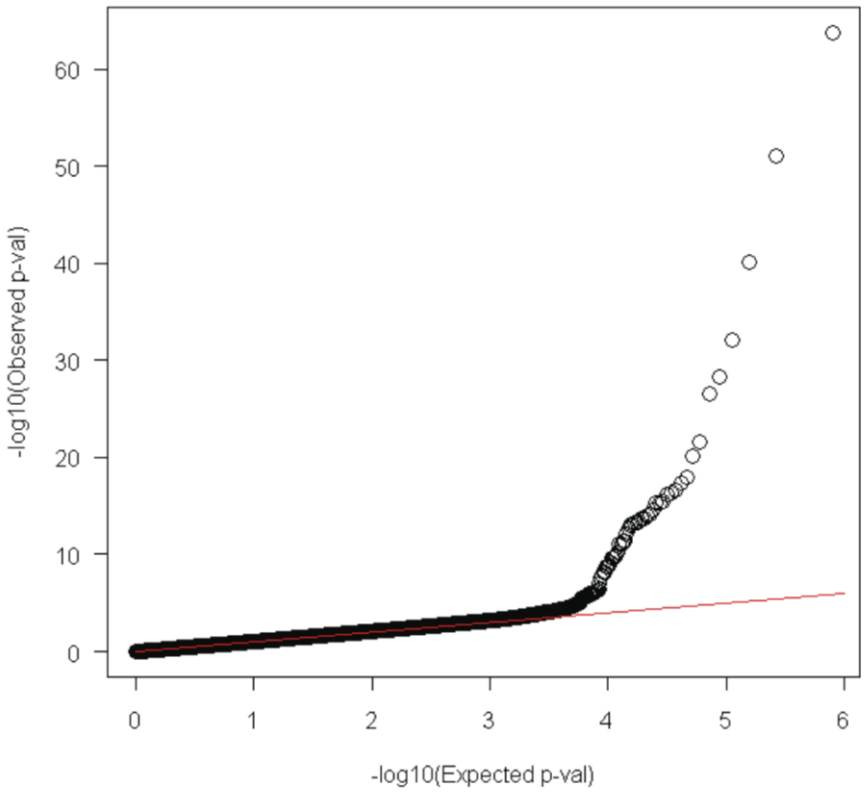

Figure S2. Effects of a SNP's physical distance from the HD mutation and allele frequency in case-control association analysis

A. The absolute physical distances of SNPs to the *HTT* CAG repeat (X-axis) were plotted against the significances of SNPs in the original dominant model association analysis (Y-axis).

B. The absolute differences of minor allele frequencies between cases and controls (X-axis) were plotted against the significance of SNPs in the original dominant model p-value (Y-axis).

C. Absolute distances of SNPs to the *HTT* CAG repeat (X-axis) were plotted against the absolute minor allele frequency differences between controls and cases.

D. Minor allele frequencies of SNPs (X-axis) in controls were plotted against significances ($-\log_{10}$(p-value)) of SNPs in the original dominant model case-control association analysis.

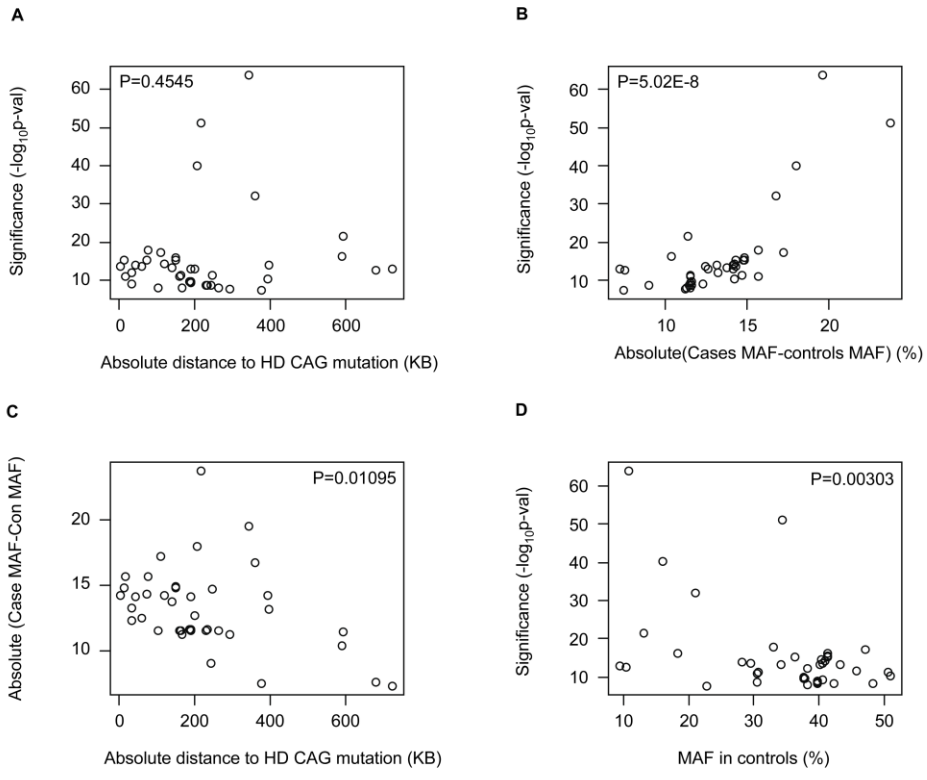

Figure S3. Effects of minor allele frequencies in controls on case-control association analysis

Cases of a disorder ascertained by phenotype may have multiple etiologies. If a proportion of cases is due to a strong effect ancestral mutation inherited from a common founder, that genetic defect may be localized by genome-wide association. To test the power of this approach at low control minor allele frequencies, we created simulation data sets (10,000 cases and 10,000 controls, red; 5,000 cases and 5,000 controls, green; 2,000 cases and 2,000 controls, blue; 1,000 cases and 1,000 controls, black) assuming the presence of a SNP minor allele of frequency 0.01 (A), 0.05 (B) or 0.10 (C) associated with the mutation-bearing chromosome. We then varied the proportion of cases attributable to this mutation-bearing chromosome and performed case:control association analysis for the target SNP (allele test) to calculate significance (-log10p-value). Dotted lines indicate genome-wide significance (5E-8). Clearly, assessing SNPs of low allele frequency in the control population permits detection of the locus even if only a small proportion of cases are due to a founder mutation-bearing chromosome, with the proportion of cases required decreasing with decreasing MAF in controls.