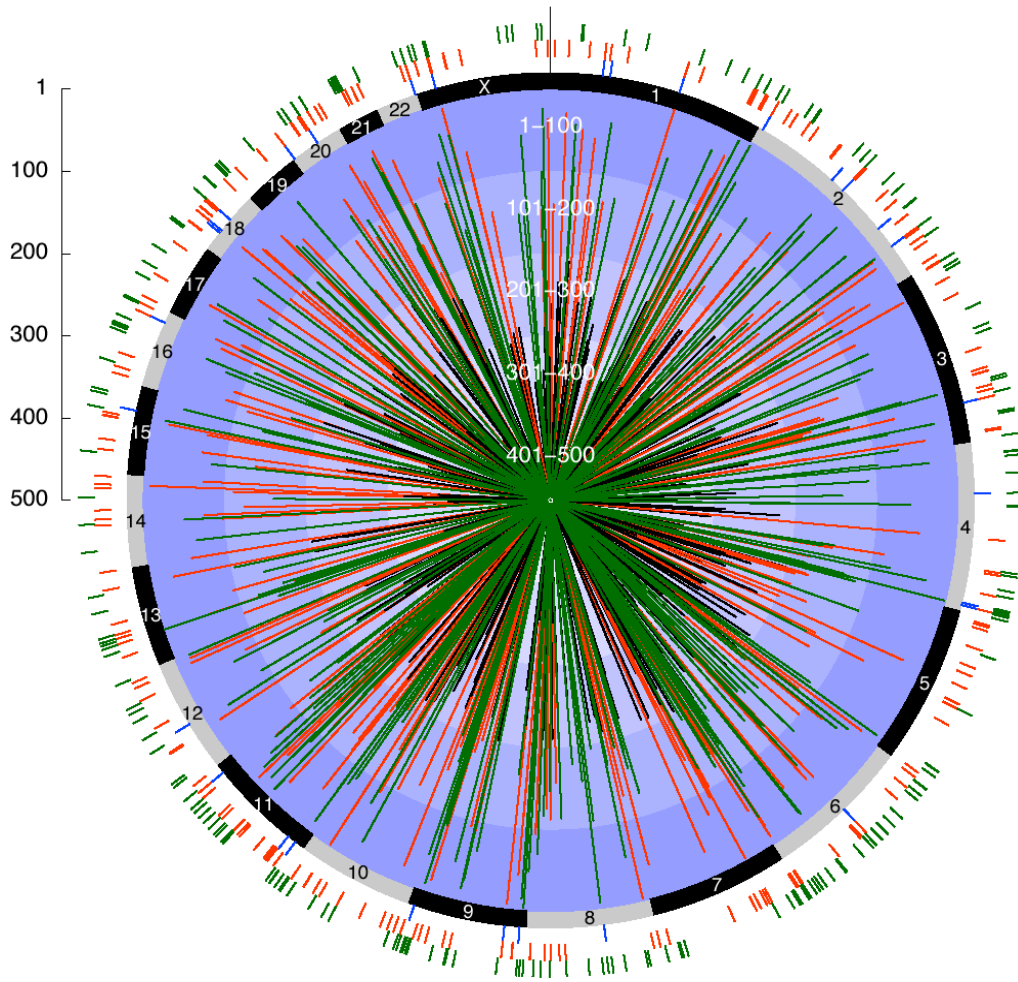


Supplementary Appendix

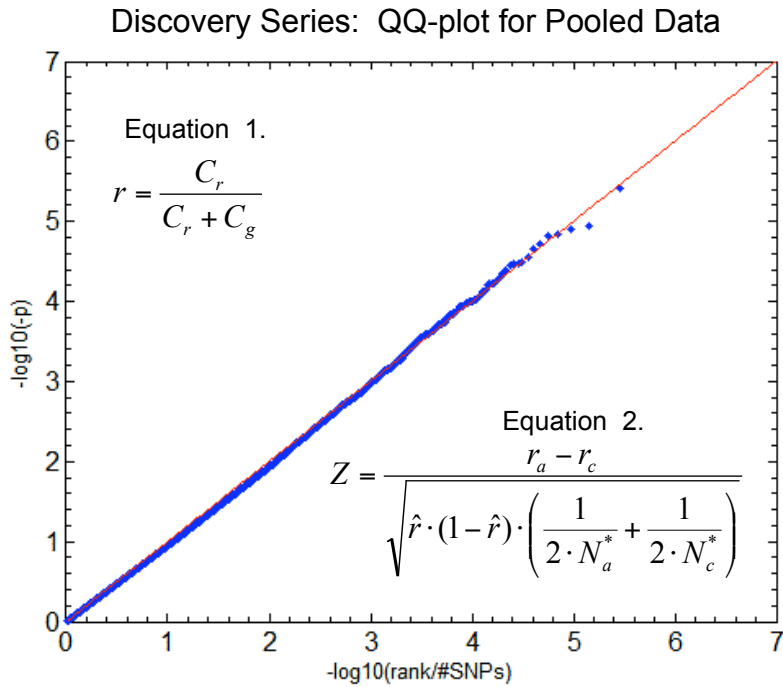
This appendix has been provided by the authors to give readers additional information about their work.

Supplement to: Dunckley T, Huentelman MJ, Craig DW, et al. Whole-genome analysis of sporadic amyotrophic lateral sclerosis. *N Engl J Med* 2007;357. DOI: 10.1056/NEJMoa070174.

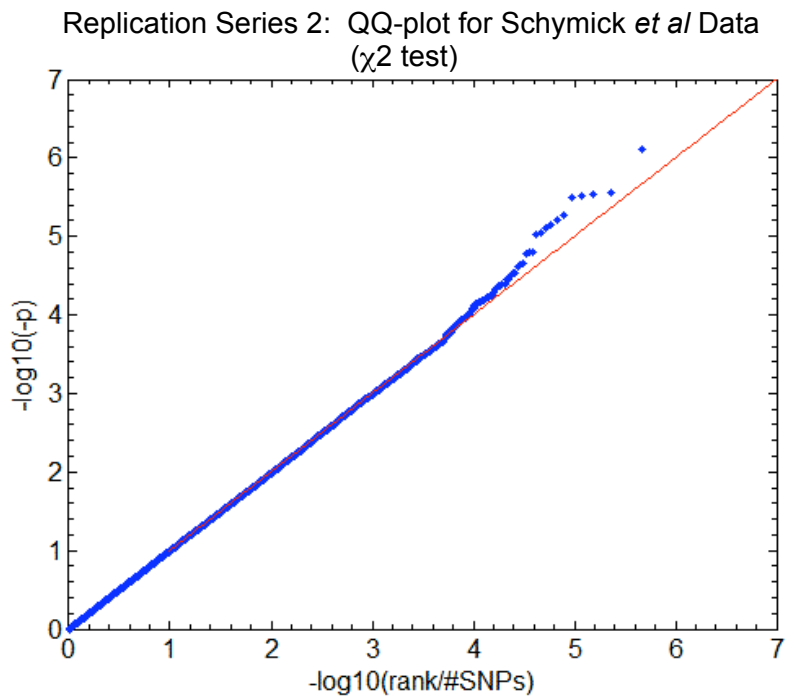


Supplementary Figure 1: The genomic location of the top 500 ranked SNPs from both the Affymetrix and Illumina pooling experiments. The genome is plotted around the circumference of a circle as a clockwise series of labeled black and gray bars representing chromosomes. Inside the chromosome ring, the individual SNPs are plotted as lines radiating from the origin where the length of the line denotes the rank so the longest line in each color represents the top ranked (number "1") SNP. The top 200 SNPs are shown in green for Illumina and red for Affymetrix while the SNPs ranked 201-500 on each platform are shown in black. Outside the chromosome ring are 3 rings of ticks. The outer rings of green and red ticks are a second representation of the location of each of the top 200 ranked Illumina and Affymetrix SNPs. The innermost ring of blue ticks shows locations where 2 or more SNPs from the combined list of top 500 Illumina and top 500 Affymetrix SNPs are within 20kb of each other.

A.



B.



Supplementary Figure 2: Quantile-Quantile plot for data generated from Discovery series of pooled genomic DNA (Panel A) and Replication series 2 (Panel B) shows no population stratification. Details of the analysis are presented in the Supplemental Methods section below.

Methods

Ancestry of Cases

Ancestry was determined through self-reporting in the presence of the physician, so that accuracy was maximized, at the time of the initial enrollment and clinical assessment. The options for ancestral origin and ethnicity (as well as clinical data options) were directly adopted from the National Institute for Neurological Disorders and Stroke form entitled “NINDS Human Genetic Resource Center Motor Neuron Clinical Data Elements” (<http://www.alsrg.org/>) so that data sharing with the ALS research community could be facilitated. The options were coded as pull-down standardized nomenclature to avoid any variance in definitions. Ancestral origin was defined as American Indian, Alaska Native, Black/African American, Asian, White, Native Hawaiian/Other Pacific Islander, Unknown, or Other. Ethnicity was defined as ‘Not Hispanic or Latino’ or ‘Hispanic or Latino’.

Assessment of population stratification

Supplemental Figure 2 illustrates Quantile-Quantile plots for data generated from Discovery series of pooled genomic DNA (Panel A) and Replication series 2 (Panel B) shows no population stratification. **A.** In this analysis, blue values show negative log base 10 values of the p-values derived from a test of two proportions using equations 1 and 2 above (Y-axis) vs. the expected values (X-axis) under the null hypothesis. The red-line shows the expected values calculated as the negative log base 10 of the SNP rank divided by the total number of SNPs. In equation 1, $r\{a|c\}$ is an approximation of allele frequency and is the ratio of signal intensity of the red channel (C_r) to the sum of the red and green channels (C_g) for the ALS and control cohorts, subscript a and c respectively. We note that at the time of analysis, k -correction factors were not available to precisely estimate allele frequency. However, we are most interested in differences and not absolute predictions of allele frequency. In equation 2, the test-statistic (Z) is calculated as essentially a test of two proportions (analogous to an allele-based χ^2 test) since only allele frequency estimates are obtained in a pooling-based study. In this equation, N^* refers to effective number of individuals in

each cohort following pooling and is calculated as the deflation in the test-statistic observed due to measurement noise as defined earlier as the ratio of expected variance to total observed variance (Pearson et al, 2007). SNPs exhibiting the greatest 5% of variance between replicate arrays were eliminated (which account for poor-performing probes) and only autosomal SNPs were plotted. We note that inflation/deflation estimates may only be useful for assessing deviation from linearity at the extremes since we do not directly calculate allele-frequency and because the effective pool size N^* incorporates deflation due to pooling-based measurement noise. A small amount of deflation of test-statistics is expected due to numerous factors such as inaccurate pipetting, degraded samples, measurement noise etc. **B.** Single-marker statistics were used to assess stratification in replication series 2. Blue dots show negative log base 10 of p-values calculated using an allele-based χ^2 test in Plink 0.99p. The Y-axis represents observed values and the X-axis represented expected values under the null hypothesis. The red-line shows the expected values calculated as the negative log base 10 of the SNP rank divided by the total number of SNPs. SNPs with a minor allele frequency (MAF) less than 5% were removed prior to plotting. As determined by STRUCTURE there is no stratification in this case-control series.

DNA extraction

Genomic DNA was isolated using the Puregene DNA isolation kit (Gentra Systems, Inc). Prior to quantitation, all DNA samples were checked for quality using 2% agarose gel electrophoresis, and degraded samples (as evidenced by characteristic smearing of DNA to low molecular weight species) were excluded from the high-density whole-genome SNP genotyping assay. Individual genomic DNA concentrations of each subject were determined in quintriplicate with the Quant-iT PicoGreen dsDNA Assay Kit (Invitrogen) according to the manufacturer's instructions.

Genotyping and Statistical Analysis

Significantly associated SNPs in the Discovery series were identified using a previously described silhouette statistic to rank differences in relative allele signals between cases and controls, implemented using GenePool Software (<http://genepool.tgen.org/>). Based on previous studies, this method is effective at identifying SNPs associated with disease, although a number of false positives would be expected, owing to measurement variance from pooling. This reinforces the need for validation using individual genotyping of separate case/control series.

An additional SNP on the same Caucasian haplotype block (defined as $r^2 > 0.8$) as the Affymetrix or Illumina top ranked SNP from the Discovery Series was selected to ensure 1) that redundancy was in place in the case of failure of any specific SNP genotypes, and 2) that the potential for identifying a positive association in the ethnically diverse replication series which may have different allele frequencies and haplotype block structure was maximized. For these 768 SNPs genotyped in the validation population, allelic χ^2 p-values were calculated using Haploview 3.32 (<http://www.broad.mit.edu/mpg/haploview/>). All SNPs failing Hardy-Weinberg equilibrium at $p=0.05$ or that had a genotype call rate of lower than 0.95 were excluded from further analyses.

Western Blotting

Frozen spinal cord tissue samples from autopsy confirmed 4 healthy non-neurologic controls and 8 ALS patients were obtained from the ALS Tissue Bank at the University of Pittsburgh. Five case and four control males and females were used in each group. The average age for the control group was 60 years and the ALS group was 58 years. The average post-mortem interval was 8 hours for the control group and 6 hours for the ALS group. In addition, cerebrospinal fluid was obtained from healthy controls and ALS patients and frozen tissue samples from multiple organs of healthy control subjects were obtained following IRB consent from the University of Pittsburgh. Frozen tissue samples were homogenized in ice-cold lysis buffer using a polytron homogenizer and disposable plastic probes (Omni International, Marietta, GA). Lysis buffer contained 50 mM Tris (pH 8.0), 600 mM NaCl, 2% CHAPS and 1% protein inhibitor cocktail II

(Sigma-Aldrich, St Louis, Mo). Protein content of the homogenates was measured using BCA Protein Assay kit (Pierce, Rockford, IL).

50 μ g of protein was loaded into each lane and electrophoresed on NuPage 12% Bis-Tris gels (Invitrogen, Carlsbad, CA) at 200 V. For cerebrospinal fluid (CSF) samples, 10 μ l of CSF was loaded per lane (\sim 1 μ g total protein per lane).

Proteins were transferred onto nitrocellulose membranes (Bio-Rad, Hercules, CA), blocked in 5% non-fat milk, immunolabeled with mouse polyclonal antibody to FLJ10986 protein (Novus Biologicals, Littleton, CO) at 1:500 concentration in Tris buffered saline, pH 7.4 (TBS), containing 0.05% Tween 20. The blot was then washed extensively in TBS and incubated with secondary anti-rabbit HRP-labeled antibody at 1:1,000 dilution in TBS. After additional washes in TBS, the secondary antibody was revealed by incubating the blot with Chemiluminescence reagent (PerkinElmer, Wellesley, MA).