A High-Density Genome-Wide Association Screen of Sporadic ALS in US Veterans

Lydia Coulter Kwee1,2,3, Yutao Liu1,2,3, Carol Haynes2,3, Jason R. Gibson1,2,3, Annjanette Stone4, Steven A. Schichman4,5, Freya Kamel6, Lorene M. Nelson7, Barbara Topol7, Stephen K. Van Den Eeden8, Caroline M. Tanner9, Merit E. Cudkowicz10, Daniela L. Grasso10, Robert Lawson10, Sumitra Muralidhar11, Eugene Z. Oddone1,3, Silke Schmidt1,2,3,9, Michael A. Hauser1,2,3,9

1 Epidemiology Research and Information Center, Durham Veterans Affairs Medical Center, Durham, North Carolina, United States of America, 2 Center for Human Genetics, Duke University Medical Center, Durham, North Carolina, United States of America, 3 Department of Medicine, Duke University Medical Center, Durham, North Carolina, United States of America, 4 Pathology and Laboratory Medicine Service and Research Service, Central Arkansas Veterans Healthcare System, Little Rock, Arkansas, United States of America, 5 Department of Pathology, University of Arkansas for Medical Sciences, Little Rock, Arkansas, United States of America, 6 Epidemiology Branch, National Institute of Environmental Health Sciences, Research Triangle Park, North Carolina, United States of America, 7 Division of Epidemiology, Department of Health Research and Policy, Stanford University School of Medicine, Stanford, California, United States of America, 8 Division of Research, Kaiser Permanente Northern California, Oakland, California, United States of America, 9 The Parkinson’s Institute, Sunnyvale, California, United States of America, 10 Massachusetts General Hospital, Neurology Clinical Trial Unit, Harvard Medical School, Boston, Massachusetts, 11 Office of Research and Development, Department of Veterans Affairs, Washington, District of Columbia, United States of America

Abstract

Following reports of an increased incidence of amyotrophic lateral sclerosis (ALS) in U.S. veterans, we have conducted a high-density genome-wide association study (GWAS) of ALS outcome and survival time in a sample of U.S. veterans. We tested ~1.3 million single nucleotide polymorphisms (SNPs) for association with ALS outcome in 442 incident Caucasian veteran cases diagnosed with definite or probable ALS and 348 Caucasian veteran controls. To increase power, we also included genotypes from 5909 publicly-available non-veteran controls in the analysis. In the survival analysis, we tested for association between SNPs and post-diagnosis survival time in 639 Caucasian veteran cases with definite or probable ALS. After this discovery phase, we performed follow-up genotyping of 299 SNPs in an independent replication sample of Caucasian veterans and non-veterans (ALS outcome: 183 cases and 961 controls; survival: 118 cases). Although no SNPs reached genome-wide significance in the discovery phase for either phenotype, three SNPs were statistically significant in the replication analysis of ALS outcome: rs6080539 (177 kb from PCSK2), rs7000234 (4 kb from ZNF704), and rs3113494 (13 kb from LOC100506746). Two SNPs located in genes that were implicated by previous GWA studies of ALS were marginally significant in the pooled analysis of discovery and replication samples: rs17174381 (p = 4.4×10^{-9}) and rs6985069 near ELP3 (p = 4.8×10^{-9}). Our results underscore the difficulty of identifying and convincingly replicating genetic associations with a rare and genetically heterogeneous disorder such as ALS, and suggest that common SNPs are unlikely to account for a substantial proportion of patients affected by this devastating disorder.


Editor: Huaibin Cai, National Institute of Health, United States of America

Received October 3, 2011; Accepted January 30, 2012; Published March 28, 2012

This is an open-access article, free of all copyright, and may be freely reproduced, distributed, transmitted, modified, built upon, or otherwise used by anyone for any lawful purpose. The work is made available under the Creative Commons CC0 public domain dedication.

Funding: The authors gratefully acknowledge support for this work from the Office of Research and Development, Cooperative Studies Program, Department of Veterans Affairs [CSP #5008]. The GENEVA study was supported by National Institute of Environmental Health Sciences/National Institutes of Health [R01 ES 013244] and the ALS Association [ALSA 1230]. The National Registry of Veterans with ALS and its DNA bank were supported by the Office of Research and Development, Cooperative Studies Program, Department of Veterans Affairs [CSP #500A, CSP #478]. The study was supported in part by the intramural research program of the National Institute of Health, by National Institute of Environmental Health Sciences [Z01-ES049005]. The GEM study was supported by the National Institute of Environmental Health Sciences [R01-ES08150] and the National Institute of Neurological Disorders and Stroke/National Institutes of Health [R01 NS 013164]. The NEALS consortium was supported by the ALS Association and the ALS Therapy Alliance. This study makes use of data generated by the Wellcome Trust Case-Control Consortium. A full list of the investigators who contributed to the generation of the data is available from www.wtccc.org.uk. Funding for the project was provided by the Wellcome Trust under award 076113 and 085475. This study was overseen by an Executive Committee appointed by the Veterans Affairs (VA): Mary Brophy, MD, MPH; MAH, PhD; Donald Humphries, PhD; LCK, PhD; EZO, MD, MHSc; SAS, MD, PhD; SS, PhD. The VA Cooperative Studies Program provided policies and requirements for conducting the research, contributed to study design, and provided administrative review of the paper but did not alter the contents. All other funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: SM set program policies as part of the Veterans Affairs Office of Research and Development’s management team. SAS owns stock in Affymetrix, Inc. CMT serves as a consultant for IMPAX Pharmaceuticals, Adamas Pharmaceuticals, Allergan Pharmaceuticals, and Genentech and receives research support from National Institute of Neurological Disorders and Stroke, Agency for Healthcare Research and Quality, United States Department of Defense, Michael J. Fox Foundation, Parkinson Disease Foundation, Parkinson Unity Walk, James & Sharron Clark. None of these consulting or funding relationships are related to ALS research. This does not alter the authors’ adherence to all the PLoS ONE policies on sharing data and materials. All other authors have declared that no competing interests exist.

* E-mail: mike.hauser@duke.edu

† These authors contributed equally to this work.
Introduction

Amyotrophic lateral sclerosis (ALS) is a fatal disease characterized by motor neuron degeneration, which leads to muscle atrophy and paralysis. Typically, this progressive muscle wasting results in death from respiratory failure within 2–4 years after onset of symptoms. ALS is the most common adult-onset motor neuron disease, with an incidence of 2–3 cases per 100,000 person-years [1], but its prevalence is low due to the poor prognosis associated with disease.

The majority of ALS cases are sporadic; approximately 5% of cases show a family history of disease [2]. Although several causative genes have been identified in familial ALS, the genetic contributors to sporadic ALS have been more difficult to identify. Genome-wide association studies of sporadic ALS have implicated a number of genes or regions (DPP6, ITPR2, UNC13A, FGGY, ELP3, KIFAP3, 9q21.2) [3–11], but replication of these findings in independent populations has proven difficult [12–16]. A possible exception is the 9q21.2 region, which was recently replicated by two large independent studies and seems to be important in both familial and sporadic ALS [17,18].

Environmental exposures also appear to play a role in ALS incidence. Some of the reported associations include cigarette smoking [19,20], head injury [21–23], exposure to lead or other heavy metals [24–27], exposure to pesticides [28,29], and physical activity [30,31], although subsequent independent studies have reported mixed results. Additionally, an increased incidence of ALS has been reported both specifically for US veterans of the Persian Gulf War [32–34] and more generally for all US veterans [35]. The nature of the relationship between military service and ALS merits further investigation into the possible aspects of military service (environmental exposures, deployment exposures, lifestyle behaviors) that may confer an increased risk of ALS.

Here, we describe a genome-wide association study (GWAS) performed in a population of US veterans. To our knowledge, this is the first genome-wide study designed to identify genetic factors that may contribute to ALS in a veteran population. Genotypes were obtained using two different arrays, which generated the highest density of single-nucleotide polymorphisms (SNPs) of all ALS GWA studies published to date. Additionally, the existence of overlapping probes on the two arrays allowed us to evaluate genotype concordance between the platforms and to reduce genotyping errors. In order to evaluate genetic factors associated with developing sporadic ALS as well as survival time after diagnosis of ALS, we performed both a case-control analysis and a survival analysis. Following the discovery phase of the study, we genotyped potentially-interesting SNPs in an independent replication study consisting of both veterans and non-veterans.

Methods

Ethics statement

This study was conducted in accordance with the Declaration of Helsinki of the World Medical Association. All study participants provided written informed consent. The research was reviewed and approved by the Institutional Review Boards from all participating institutions, including the Institutional Review Boards of the Durham VA and Duke University Medical Centers, the Parkinson’s Institute, National Institutes of Health, National Institute of Environmental Health Sciences, University of Iowa, Massachusetts General Hospital, New England Medical Center, Brigham and Women’s Hospital, Social and Scientific Systems, Inc., Battelle, Inc., Survey Research Associates, CODA, and the Human Subject Committees of the Kaiser Foundation Research Institute and Stanford University.

Participants

Case samples for this study were obtained from the National Registry of Veterans with ALS, which enrolled 2121 US veterans between April 2003 and September 2007. We refer to the veteran cases here as “NRVA cases.” Methods for recruitment, medical record review, and enrollment into the Registry have been described in detail elsewhere [36]. Briefly, cases were actively recruited through periodic searches of Veterans Administration (VA) inpatient and outpatient databases for ICD-9 (International Classification of Diseases, 9th Revision, Clinical Modification) diagnoses of motor neuron disease. Passive recruitment methods included the distribution of study brochures and mailings to ALS specialty clinics and neurologists, along with links on ALS-related websites. Following informed consent and enrollment, neurologists reviewed each case’s medical records to confirm a diagnosis of definite, probable or possible ALS according to the original El Escorial criteria [37,38]. Patients were also enrolled in the Registry if they had “suspected ALS” according to the criteria, which included a diagnosis of progressive muscular atrophy (PMA), primary lateral sclerosis (PLS) or progressive bulbar palsy (PPB). A subset of enrollees (n = 1173) consented to participate in the registry-affiliated DNA bank and donated DNA via blood (84.6%) or mouthwash (15.4%), and 1163 of these samples were available for genotyping in this study. Registry enrollees were contacted for follow-up telephone interviews every 6 months from enrollment until September 30, 2009; the ALS Functional Rating Scale (ALSFRS-R) was administered at each of the follow-up interviews.

The veteran controls described here were enrolled in the “Genes and Environmental Exposures in Veterans with ALS” (GENEVA) study. As described previously [39], controls were identified from a database of US veterans maintained by the Veterans Benefit Administration and recruited via mailings and telephone calls. We refer to these controls as “GENEVA controls.” Controls were frequency-matched to cases on age (in 5-year intervals), gender, and use of the VA for health care (prior to diagnosis, for cases). Controls passed a telephone screener to confirm the absence of ALS and other neurological diseases. All of the GENEVA controls were administered telephone interviews about environmental exposures, as was a subset of the sampled NRVA cases (57%). Controls were also asked to donate a saliva sample for DNA extraction and genotyping; 411 control samples were genotyped in the discovery phase of this study. To improve the statistical power of our discovery GWAS, we also made use of approximately 6,000 publicly-available control genotypes generated by the same two high-density chips and distributed by the Wellcome Trust Case Control Consortium (WTCCC2) (http://www.wtccc.org.uk/ccc2/) [40]. Of these controls, 51.6% were recruited from the 1958 British birth cohort; the remainder comprised blood donors from the UK Blood Service Collection. As previously demonstrated [41], publicly available controls are well suited for inclusion into the analysis of a rare disease like ALS, because they have a very low probability of misclassification.

The replication phase of this analysis included 490 additional controls ascertained through the GENEVA study and 20 NRVA cases along with samples from the Northeast ALS Consortium (52 cases and 11 controls, contributed by MEC); the Agricultural Health Study (312 controls, contributed by FK); a New England study of ALS (108 cases and 38 controls, contributed by FK); and the Genetics and Epidemiology of Motor Neuron disease (GEM) study (100 cases and 308 controls, contributed by LMN). Details of
sample ascertainment for each of these studies are given in Methods S1.

**Laboratory methods**

For NRVA case samples, DNA was obtained from peripheral blood or from buccal cells. Blood was collected in 10 ml EDTA blood tubes. The blood was centrifuged for 15 min at 1850 × g at 4°C; the plasma was removed and saved. The remaining buffy coat and red cells were transferred to a tube for DNA extraction. Buccal cells were collected in 10 ml original Mint Scope Mouthwash (Proctor & Gamble). The samples were centrifuged for 10 min at 2000 × g and the supernatant discarded. DNA from both blood and buccal cells was extracted on an Autopure instrument using Puregene reagents (Qiagen, Valencia, CA). A subset of samples (1 blood; 16 buccal) did not have sufficient DNA available for genotyping. For these samples, DNA amplification was performed using the REPLI-g DNA amplification kit (Qiagen, Valencia, CA) according to the manufacturer’s directions. Samples (20–50 ng) were amplified for 16 h and then purified by alcohol precipitation. For veteran control samples, DNA was extracted from Oragene samples (DNA Genotek, Inc.) using the Autopure LX following the manufacturer’s recommended protocol.

**Genotyping and quality control**

The discovery samples were genotyped on the Affymetrix Genome-Wide Human SNP Array 6.0 at Affymetrix, Inc. (Santa Clara, CA) and the Illumina Human1M-Duo genotyping array at the VA Pharmacogenomics Analysis Laboratory (Little Rock, AR). After calling the genotypes (see Methods S2 for a detailed description of the genotype-calling methods), we applied the following set of quality control (QC) filters to the autosomal SNPs on each array independently. We excluded SNPs with call frequency <98%, minor allele frequency ≤3%, HWE p-value <10^{-6} in controls, heterozygote calls ≥65% and p-value <10^{-5} for a test of differential missingness between cases and controls. We also applied standard QC methods to the samples genotyped on each array and removed samples based on call rate (<98%), mismatch between self-reported gender and genotypic gender, cryptic relatedness between samples, or outlying ethnicity (see Methods S3 and Table S1 for details). After applying these QC filters, our discovery sample contained 1142 cases (98.2%) and 394 controls (95.9%) with data from at least one genotyping array (Illumina 1M-Duo or Affymetrix 6.0). The mean call rate for the discovery samples varied from 99.3% (mouthwash) to 99.5% (blood). For the WTCCC2 data, we removed all samples listed in the sample exclusions file provided by WTCCC. Exclusion criteria included: cryptic relatedness between samples, gender mismatch, low call rate, questionable identity based on previous genotyping, outliers based on heterozygosity, outliers based on principal component plots using HapMap samples, and outliers based on mean A and B intensities on chromosome 22. We also removed all SNPs in the WTCCC2 SNP exclusion files provided based on these criteria: minor allele frequency <1%, information <0.975, call frequency <98% and plate association test p-value <10^{-5}. Additionally, we removed SNPs with HWE p-value <10^{-6}, which is a more stringent filter than that initially applied by the WTCCC.

The replication samples were genotyped with TaqMan assays (Applied Biosystems, Forest City, CA) using the 7900HT (Duke University Center for Human Genetics, Durham, NC) and OpenArray (HudsonAlpha Institute for Biotechnology, Huntsville, AL) platforms. Genotypes were assigned using ABI’s Genotyper software for OpenArray Taqman data, and SDS 2.4 for 7900HT Taqman data. Samples with <90% call rate on the OpenArray platform or poor duplicate concordance were removed. SNPs were removed based on call frequency <90%, discordance between duplicate samples, Mendelian inconsistencies in CEPH trios, and discordance with previously-confirmed genotypes when available.

**Statistical analysis**

For the discovery-phase analysis, we combined the markers and genotypes from the Affymetrix 6.0 and Illumina 1M-Duo chips. When a given marker was genotyped and passed QC on both chips, we followed these steps to combine the genotypes: 1) retained Illumina genotypes for symmetric (A/T or C/G) SNPs; 2) performed Fisher’s exact test on a 3 × 2 table of genotype frequencies by chip for each marker, and discarded SNPs with p <0.001; 3) set any remaining discordant genotypes for a given sample to be missing. We combined the Affymetrix 6.0 and Illumina 1M-Duo genotypes for the WTCCC2 samples following the same procedure with one exception: in step 2, due to the much larger sample size, we used a χ^2 test for different genotype frequencies between the two chips. Finally, we used Fisher’s exact test to test for genotype frequency differences between GENEVA controls and WTCCC2 controls, and removed any markers with p-value <0.001.

**ALS outcome.** To test for genetic effects on the primary endpoint of interest, ALS outcome, we restricted our analysis to definite or probable ALS cases who received an initial ALS diagnosis within the 24 months prior to Registry enrollment. We also performed a sensitivity analysis by including other diagnoses (possible ALS and PMA) and/or restricting analysis to cases diagnosed within 12 months of Registry enrollment. Subjects with diagnoses of PLS or PBP and those with longer lags (>24 months) between diagnosis and study consent were excluded from the ALS outcome analysis. Because the number of eligible GENEVA controls was limited at the time of genotyping, we added the publicly-available WTCCC2 controls in an effort to increase the power of our analysis to detect moderate associations with ALS. However, there may be important environmental or other differences between the British non-veteran subjects in the WTCCC2 sample and the US veterans in our study; for this reason, we also performed analyses using the GENEVA controls alone. We excluded all cases and controls with a first-degree family history of ALS (self-reported or, for cases, determined by a Registry neurologist) or a known SOD1 mutation. Finally, for the purpose of comparison to previously published GWA studies and to match the WTCCC2 controls, we restricted the analysis here to self-reported non-Hispanic Caucasians. After these exclusions, the sample used for ALS outcome analysis included 442 ALS cases, 348 GENEVA controls and 5909 WTCCC2 controls (see Table S2 for the number of samples meeting each exclusion criterion).

Case-control analyses were carried out using unconditional logistic regression, with genotypes coded additively based on the number of copies of the minor allele. We included sex and reference age (age at diagnosis for cases; age at interview for controls) in models using GENEVA controls only. For models including the WTCCC2 controls, we adjusted only for sex, because age data were not available for these subjects at the time of our analysis. For the discovery phase, we also adjusted all models for potential population stratification in our sample by including four components obtained from multi-dimensional scaling analysis (MDS) in PLINK. The MDS components were obtained based on a set of 84 k LD-pruned SNPs discussed in Methods S3. Statistical analyses of the data were carried out using
PLINK version 1.07. Code to create Manhattan and QQ plots was adapted from http://gettinggeneticsdone.blogspot.com/2011/04/annotated-manhattan-plots-and-qq-plots.html.

For the replication phase, we performed unconditional logistic regression on the replication samples alone (183 cases, 961 controls) and on the pooled set of samples (replication and discovery) when SNPs were successfully genotyped on both sets of samples. As with the discovery sample, we only considered non-Hispanic Caucasians and restricted analysis to definite or probable ALS cases who received an initial ALS diagnosis within the 24 months prior to study consent. Any genotypes that were discordant between the discovery and replication phases were set to missing for the combined analysis. All p-values reported in the discovery and replication phases are unadjusted for multiple testing.

Survival. Our case group for survival analysis included all non-Hispanic Caucasian cases with definite or probable ALS, regardless of the disease duration at the time of consent (n = 639, 29.5% censored). For the survival analysis, we also excluded cases who were already dependent on a ventilator at the time of study enrollment (see Table S2). Using a Cox proportional hazards regression model, we tested for genetic effects on the number of months prior to study consent. We adjusted the model for left-truncation in order to account for the fact that subjects had to survive long enough to contribute a sample to the NRVA DNA bank. All models contained an additive genotype term and included age at diagnosis and sex as covariates. For the discovery phase, the four components from the MDS analysis described above were included as well. We categorized subjects as having bulbar or non-bulbar onset; this variable did not meet the proportional hazards assumption necessary for the model, so we adjusted for site of onset using a stratified model instead of the four components from the MDS analysis described above. All models were performed using PLINK (v1.06) and R (version 2.10.1), via a PLINK plug-in.

For the replication phase, we performed a survival analysis on the replication samples alone (n = 118, 16.1% censored) and on the pooled set of samples (replication and discovery) when SNPs were successfully genotyped on both sets of samples. As with the discovery sample, we restricted analysis to non-Hispanic Caucasian cases diagnosed with definite or probable ALS.

Selection of SNPs for replication genotyping

To account for the possibility that ALS-associated SNPs might be enriched in the set of SNPs showing marginal (but not genome-wide) significance, we selected SNPs for replication genotyping using either a strict p-value criterion (p<1.0×10^{-6}) or a more lenient p-value criterion (p<1.0×10^{-4}) in the presence of other evidence of association (previous independent studies or consistent results across our sensitivity analyses). We examined the genotype cluster plots for SNPs that met either criterion, and only attempted to replicate SNPs that clustered well. In addition to SNPs from the GWAS chips, we decided to genotype additional SNPs in implicated genes/regions, because replication may be observed at the level of a gene rather than a specific SNP. We included common SNPs in coding regions or potential splice sites from 20 selected genes, tagging SNPs from 5 selected genes, and intragenic SNPs in moderate LD (r^2>0.5) with intergenic SNPs of interest. Based on previously published GWAS studies of sporadic ALS, we also genotyped the discovery and replication samples for variation in the genes DPP6 [7], ELPS3 [3], ITPR2 [5], and UNC13A [9], and for the intergenic SNP rs3849942 on chromosome 9p21.2 [9,11]. We genotyped the specific implicated SNPs in addition to common coding and possible splicing SNPs in these genes. We also included a SNP in KIFAP3 previously reported to be associated with survival [8]. The total number of SNPs that were successfully genotyped in the replication samples was 299 (286 selected for the ALS outcome phenotype, 13 selected for survival).

| Table 1. Demographic and clinical characteristics of the discovery and replication samples. |
|-----------------------------------------------|-----------------|-----------------|-----------------|-----------------|-----------------|
|                  | Discovery cases: ALS outcome | Discovery cases: Survival | Discovery controls | Replication cases: ALS outcome | Replication cases: Survival |
| n                 | 442                           | 639                           | 348                           | 183                           | 118                           | 961                           |
| Male              | 435 (98.4%)                    | 627 (98.1%)                    | 331 (95.1%)                   | 115 (62.8%)                   | 80 (67.8%)                    | 75S (78.6%)                       |
| Age (yrs, mean ± SD) | 62.3±10.2                     | 60.4±11.3                     | 62.2±10.7                     | 59.4±12.1                     | 58.8±12.2                     | 63.5±11.5                      |
| Veterans          | 442 (100%)                     | 639 (100%)                     | 348 (100%)                   | 183 (100%)                   | 118 (100%)                   | 961 (100%)                        |
| Diagnosis         |                                |                               |                               |                               |                               |                               |
| Definite ALS      | 115 (26.0%)                   | 200 (31.3%)                   | -                             | 183 (100%)*                   | 118 (100%)*                   | -                               |
| Probable ALS      | 327 (74.0%)                   | 439 (68.7%)                   | -                             |                               |                               |                               |
| Time from diagnosis to study consent          |                                |                               |                               |                               |                               |                               |
| 0–12 months       | 293 (66.3%)                   | 285 (44.6%)                   | -                             | 164 (89.6%)                   | 98 (83.1%)                    | -                               |
| 12–24 months     | 149 (33.7%)                   | 141 (22.1%)                   | -                             | 19 (10.4%)                    | 16 (13.6%)                    | -                               |
| 24–36 months     | -                            | 65 (10.2%)                   | -                             | -                             | 1 (0.8%)                      | -                               |
| >36 months       | -                            | 148 (23.2%)                   | -                             | -                             | 3 (2.5%)                      | -                               |
| Bulbar onset      | 114 (25.8%)                   | 130 (20.3%)                   | -                             | not available                | 35 (29.9%)                    | -                               |
| Median onset to diagnosis (mo)                 | 11                            | 11                            | -                             | not available                | 9.1                           | -                               |
| Median survival after diagnosis (mo)           | 24.0 (n = 335)                | 29.0 (n = 442)                | -                             | 26.8 (n = 144)                | 27.1 (n = 99)                  | -                               |

*Definite/probable ALS breakdown was not available for all replication cases.

doi:10.1371/journal.pone.0032768.t001
Results

Discovery phase

After performing the sample quality control steps described in the Methods, we obtained genotypes on at least one chip for 394 controls (95.9%) and 1142 cases (98.2%). Of these samples, 350 controls (88.8%) and 1104 cases (96.7%) passed QC on both genotyping chips, yielding extremely high-density marker data. After removing cases which did not meet the inclusion criteria for analysis of ALS outcome or survival described above, our final sample included 442 cases and 348 controls for the ALS outcome analysis and 639 cases for the survival analysis. Demographic and clinical characteristics of this discovery sample are shown in Table 1. The veteran cases and controls had similar distributions for military-specific characteristics (military branch with longest service, length of service and deployment to major conflicts) [21].

After SNP quality control and removal of monomorphic SNPs, we obtained genotypes for a total of 1,515,824 autosomal SNP markers (88.1%) from either the Affymetrix 6.0 or Illumina 1M-Duo genotyping chip. 1,280,579 of these markers (84.5%) also met our minimum minor allele frequency (MAF) criterion (≥3%).

Figure 1 shows Manhattan plots for the ALS outcome analysis, both with and without the WTCCC2 controls included. No markers met a genome-wide significance threshold of p≤5.0×10⁻⁸. Detailed results of the most significant 25 SNPs from each analysis are shown in Tables 2 and 3. Two genes, *UNC13C* and *SETBP1*, contain SNPs in the top 25 of both analyses. The QQ plots for the two ALS outcome analyses are shown in Figure 2. For the association analysis including GENEVA controls only, there is a single relatively rare SNP with an observed p-value smaller than expected, rs5762919 in *ZNRF3* (p = 1.7×10⁻⁷, MAF in cases = 0.04). Unfortunately, this SNP was not successfully genotyped in the WTCCC2 controls, so we were unable to test for a consistent result with the larger set of controls. The QQ plot for the analysis including the WTCCC2 controls deviates from the expected pattern under the null hypothesis of genome-wide absence of association. We examined whether this deviation might be due to residual population stratification by including varying numbers of MDS axes (0, 4, 10, and 20) in the logistic regression models and obtained similar QQ plots.
plots even with 20 axes included (data not shown). One potential explanation for this observation might be the age, gender and sample size differences between GENEVA and WTCCC2 controls; another is our inability to adjust for age in the logistic regression models that included the WTCCC2 controls.

Results from the survival GWAS are shown in Figure 3. Again, no SNPs met a genome-wide significance threshold of $p < 5.0 \times 10^{-8}$. The top 25 SNPs from the survival analysis are shown in Table 4. There were five SNPs from the *PARK2* gene in the top ten, all of which are in strong linkage disequilibrium with each other ($r^2 > 0.85$). The QQ plot derived from the survival analysis is shown in Figure 4 and reflects a consistent trend of observed p-values that are larger than expected. This attribute of the QQ plot seems to be related to our exclusion of rare SNPs in the survival analysis. When we included all polymorphic SNPs in the analysis, regardless of MAF, the QQ plot conformed more closely to the pattern expected under the genome-wide absence of association (data not shown).

Replication and pooled analysis

After quality control of the replication samples, we obtained genotypes for 299 SNPs on 278 cases (99.3%) and 1140 controls (98.4%). Of these samples, 183 cases and 961 controls met the inclusion criteria for the analysis of ALS outcome. In addition, 170 cases met the inclusion criteria for analysis of survival, but only 118 of these cases had data for site of onset and could be included in the analysis. The replication SNPs were not all typed on the same platform (two different OpenArray chips were used, along with single-tube TaqMan assays) and quality control was performed separately on each platform. For both ALS outcome and survival, we conducted an analysis using the independent replication samples alone, followed by a pooled analysis of all samples.

The top ten most-significant results from the pooled analysis of ALS outcome are given in Table 5. There is substantial overlap in the results with and without the WTCCC2 controls, as seven markers are in the top ten for both analyses. Although no SNPs met a genome-wide significance threshold in the pooled analysis, three SNPs produced statistically significant results in analysis of the independent replication samples alone after adjustment for multiple testing of 286 SNPs ($p > 1.75 \times 10^{-5}$): rs6080539 (177 kb from *PCSK2*), rs7000234 (4 kb from *ZNF704*), and rs3113494 (13 kb from *LOC100506746* (Table 5). Analogous results for the top ten SNPs in the replication and pooled analysis of the survival outcome are shown in Table 6. None of the SNPs tested in the survival replication analysis showed evidence of association in the replication analysis alone ($p < 0.05$), which might partially be due to the substantially smaller sample size.

Table 7 shows association results in our dataset for markers and genes implicated in previous GWAS analyses of ALS outcome. For
each gene or region of interest, we show not only the marker with the smallest p-value in the pooled analysis, but also the specific SNPs found to be associated with ALS risk in previous studies. We were unable to replicate the previously reported association of the specific SNPs in 9p21.2 and ITPR2. SNP rs1541160 in KIFAP3, which was implicated by the only other genome-wide analysis of ALS survival [8], also did not show evidence for association in our pooled dataset (HR = 0.95, 95% CI = (0.82, 1.10), p = 0.49). After correcting for multiple testing (based on 57 markers selected from genes previously reported to show association, p < 3.8 × 10^{-8}), we detected marginal association with rs6985069 near ELP3 (p = 2.7 × 10^{-7} with GENEVA controls only; p = 4.8 × 10^{-7} with GENEVA+WTCCC2 controls). This is among the top 10 results of the pooled analysis shown in Table 5, but is different from the SNP previously reported as ALS-associated (rs13268593) [3]. Similarly, the specifically-implicated SNP rs10260404 in DPP6 did not show any evidence for association, while rs17174381 in the same gene had p-values of 0.005 (GENEVA controls only) and 4.4×10^{-4} (GENEVA+WTCCC2 controls) in the pooled analysis.

**Discussion**

Our GWAS of US veterans did not identify any genetic associations that reached genome-wide significance (p <5.0×10^{-8}) for either ALS outcome or survival. After following up a smaller number of SNPs, three SNPs were significantly associated with ALS outcome in a replication-only analysis after adjusting for multiple testing. Two of these markers fall outside the transcriptional boundaries of their closest genes (PCSK2 and ZNF704), and one is 13 kb from a hypothetical non-coding RNA (LOC100506746). Several other SNPs with originally similar or stronger associations (e.g. rs11071021 from UNC13C, which was in the top 25 SNPs for the ALS outcome analysis with or without WTCCC2 controls) did not meet criteria for statistical significance in the replication phase.

Several factors may have contributed to this lack of intra-study replication. Perhaps most importantly, veterans comprised all of the cases and controls in our discovery phase analysis, but comprised only 4% of the cases and 41% of the controls in our replication phase. This also created a substantial difference in gender distribution between the two samples. Because the higher rate of ALS reported for US veterans may be partially due to military-related environmental exposures and corresponding gene-by-environment interactions, the heterogeneity in environmental exposures between the discovery and replication samples could be partly responsible for the lack of replication. In this case, a much larger veteran-only replication sample would be required to detect such interactions. Study design differences may also play a role:
the cases in the discovery study were derived from a registry-based sample that included incident and prevalent cases, while the cases in the replication study were derived from a clinic- or population-based sample and were thus enriched for incident cases. Therefore, patients analyzed in the discovery phase were, on average, enrolled longer after diagnosis than patients analyzed in the replication phase (Table 1), and patients who died within two years of diagnosis may be under-represented in the discovery sample. We accommodated this in our analysis by restricting the ALS outcome analysis to incident cases and by adjusting the survival analysis for the length-biased nature of the data. As shown in Table 1, this phenotypic restriction led to similar observed survival characteristics of the two groups of patients (median survival 24.0 vs. 26.8 months for the patients included in the ALS outcome analysis; 29.0 vs. 27.1 months for the patients included in the survival analysis). In addition to the study design differences, it is possible that diagnostic heterogeneity also contributes to the difficulty of intra-study replication. The clinical criteria used by the NRVA neurologists were carefully standardized and cross-validated to minimize diagnostic heterogeneity within the discovery sample. Such standardization was not possible in our replication dataset, because these samples were independently ascertained.

While none of the SNPs evaluated in this study reached genome-wide significance, we noted several interesting genes as potential candidates for further study. For example, results from

Figure 2. Quantile-quantile plots. a) ALS outcome with GENEVA controls only; b) ALS outcome with GENEVA and WTCCC2 controls included. doi:10.1371/journal.pone.0032768.g002

Figure 3. Manhattan plot. Survival analysis of definite/probable ALS cases. doi:10.1371/journal.pone.0032768.g003
our discovery phase ALS outcome analysis indicated weakly suggestive association with UNC13C, which, like the previously implicated UNC13A, is a homolog of the *C. elegans* gene UNC-13. Both of these genes code for proteins with neurological effects. Although not replicated, our discovery phase survival analysis suggested a marginal association with variants in the PARK2 gene; mutations in PARK2 are known to cause Parkinson’s disease [42,43]. Finally, PCSK2 is an interesting candidate gene given recent work on the potential relationship between ALS and metabolic phenotypes such as hyperlipidemia, BMI and type 2 diabetes [44,45]. SNPs in PCSK2 have been shown to be associated with type 2 diabetes in several ethnic populations [46–48]. We again note that none of these genes contained markers that were significant at a genome-wide level during the discovery phase of this analysis, and present these genes only as plausible candidates for study in independent populations.

Putting our GWAS in the context of previously published genome-wide studies, we were unable to conclusively replicate the previously-reported associations between ALS and ITPR2, UNC13A or the 9p21.2 locus. However, at the level of the gene rather than the specific previously implicated SNP, we observed some evidence of association for ELP3 and DPP6. A previous genome-wide survival analysis identified a SNP in KIFAP5 as associated with increased survival of ALS patients. Our study failed to replicate this association, consistent with another independent study [49]. We note that, although our study is one
of the larger ALS genome-wide screens performed to date, it is still
underpowered to detect common variants with modest effects on
such a rare disease. For example, it would require approximately
2450 cases and an equal number of controls to have 80% power to
detect the magnitude of the effect of rs10260404 previously
described in the DPP6 gene [7].

Perhaps the most important factor contributing to the difficulty
of replication in independent populations is the underlying genetic
heterogeneity of the disease. Linkage studies of familial ALS have
implicated twelve loci and eight genes [50]. In addition to this
locus heterogeneity, there is also extensive allelic heterogeneity;
within the SOD1 gene, more than 125 non-synonymous coding

---

Table 5. Replication and pooled analysis: SNP association with ALS outcome, including and excluding WTCCC2 controls.

<table>
<thead>
<tr>
<th>Chr</th>
<th>SNP</th>
<th>Closest gene (distance)</th>
<th>Discovery analysis</th>
<th>Replication analysis</th>
<th>Pooled analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>OR (95% CI) P</td>
<td>OR (95% CI) p</td>
<td>OR (95% CI) p</td>
</tr>
<tr>
<td>4</td>
<td>rs4833346</td>
<td>127584779 FAT4 (951 kb)</td>
<td>1.5 (1.2, 1.9) 1.2E-04</td>
<td>1.1 (0.8, 1.5) 0.43</td>
<td>1.4 (1.2, 1.6) 2.1E-05</td>
</tr>
<tr>
<td>20</td>
<td>rs6075164</td>
<td>16985876 PCSK2 (169 kb)</td>
<td>0.7 (0.5, 0.8) 1.8E-04</td>
<td>0.8 (0.7, 0.9) 0.0016</td>
<td>0.7 (0.6, 0.8) 3.7E-05</td>
</tr>
<tr>
<td>20</td>
<td>rs6080539</td>
<td>1697494 PCSK2 (177 kb)</td>
<td>0.6 (0.5, 0.7) 1.1E-04</td>
<td>0.7 (0.6, 0.8) 5.0E-05*</td>
<td>0.7 (0.6, 0.8) 5.1E-05</td>
</tr>
<tr>
<td>8</td>
<td>rs7000234</td>
<td>81699669 ZNF704 (4 kb)</td>
<td>2.0 (1.3, 3.1) 8.6E-04</td>
<td>1.8 (1.4, 2.4) 6.1E-05*</td>
<td>1.7 (1.3, 2.3) 1.5E-04</td>
</tr>
<tr>
<td>20</td>
<td>rs6080544</td>
<td>16987959 PCSK2 (167 kb)</td>
<td>0.7 (0.5, 0.8) 2.0E-04</td>
<td>1.0 (0.8, 1.4) 0.78</td>
<td>0.7 (0.6, 0.9) 1.9E-04</td>
</tr>
<tr>
<td>4</td>
<td>rs3113494</td>
<td>88051625 LOCI00506746 (13 kb)</td>
<td>0.7 (0.6, 0.8) 1.9E-04</td>
<td>0.7 (0.6, 0.9) 9.6E-05*</td>
<td>0.8 (0.6, 0.9) 2.0E-04</td>
</tr>
<tr>
<td>11</td>
<td>rs2278170</td>
<td>4453422 OR52K1 (13 kb)</td>
<td>0.7 (0.5, 0.8) 5.5E-04</td>
<td>0.7 (0.6, 0.9) 8.0E-04</td>
<td>0.7 (0.6, 0.9) 2.4E-04</td>
</tr>
<tr>
<td>6</td>
<td>rs1190270</td>
<td>105731108 POPDC3</td>
<td>0.5 (0.3, 0.7) 1.1E-04</td>
<td>0.6 (0.5, 0.8) 5.0E-05*</td>
<td>0.6 (0.4, 0.8) 2.7E-04</td>
</tr>
<tr>
<td>20</td>
<td>rs6080539</td>
<td>16977494 PCSK2 (177 kb)</td>
<td>0.6 (0.5, 0.8) 1.1E-04</td>
<td>0.7 (0.6, 0.8) 5.0E-05*</td>
<td>0.8 (0.7, 0.9) 1.9E-04</td>
</tr>
</tbody>
</table>

Table 6. Replication and pooled analysis: SNP association with ALS survival.

<table>
<thead>
<tr>
<th>Chr</th>
<th>SNP</th>
<th>Closest gene (distance)</th>
<th>Discovery analysis</th>
<th>Replication analysis</th>
<th>Pooled analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Hazard ratio (95% CI) p</td>
<td>Hazard ratio (95% CI) p</td>
<td>Hazard ratio (95% CI) p</td>
</tr>
<tr>
<td>4</td>
<td>rs6840169</td>
<td>83637534 TMEM1150C</td>
<td>1.5 (1.2, 1.7) 7.8E-06</td>
<td>1.2 (0.8, 1.9) 0.40</td>
<td>1.4 (1.2, 1.6) 3.9E-05</td>
</tr>
<tr>
<td>6</td>
<td>rs564053</td>
<td>94874715 TSG1 (332 kb)</td>
<td>1.4 (1.2, 1.6) 9.6E-06</td>
<td>1.1 (0.7, 1.6) 0.62</td>
<td>1.4 (1.2, 1.6) 5.5E-05</td>
</tr>
<tr>
<td>15</td>
<td>rs873961</td>
<td>59135496 RORA</td>
<td>0.7 (0.6, 0.8) 1.4E-05</td>
<td>1.2 (0.9, 1.6) 0.31</td>
<td>0.8 (0.7, 0.9) 7.2E-05</td>
</tr>
<tr>
<td>15</td>
<td>rs11632352</td>
<td>59145158 RORA</td>
<td>0.7 (0.6, 0.8) 3.2E-05</td>
<td>1.1 (0.8, 1.5) 0.63</td>
<td>0.8 (0.7, 0.9) 7.5E-05</td>
</tr>
<tr>
<td>6</td>
<td>rs7764218</td>
<td>162854134 PARK2</td>
<td>0.5 (0.4, 0.7) 8.6E-06</td>
<td>1.1 (0.7, 1.8) 0.65</td>
<td>0.6 (0.5, 0.8) 1.5E-04</td>
</tr>
<tr>
<td>6</td>
<td>rs7757630</td>
<td>162800228 PARK2</td>
<td>0.5 (0.4, 0.7) 7.9E-06</td>
<td>1.1 (0.7, 1.7) 0.72</td>
<td>0.6 (0.5, 0.8) 1.5E-04</td>
</tr>
<tr>
<td>15</td>
<td>rs8037669</td>
<td>59137212 RORA</td>
<td>0.7 (0.6, 0.8) 1.7E-05</td>
<td>1.3 (0.9, 1.9) 0.11</td>
<td>0.8 (0.7, 0.9) 1.6E-04</td>
</tr>
<tr>
<td>6</td>
<td>rs7740421</td>
<td>162890461 PARK2</td>
<td>0.5 (0.4, 0.7) 7.4E-06</td>
<td>1.2 (0.7, 2) 0.56</td>
<td>0.6 (0.5, 0.8) 2.5E-04</td>
</tr>
<tr>
<td>20</td>
<td>rs6027005</td>
<td>57632613 PHACTR3</td>
<td>1.7 (1.3, 2.3) 2.0E-04</td>
<td>0.8 (0.2, 3.3) 0.75</td>
<td>1.7 (1.3, 2.2) 4.1E-04</td>
</tr>
<tr>
<td>6</td>
<td>rs6904956</td>
<td>162865529 PARK2</td>
<td>0.5 (0.4, 0.7) 9.5E-06</td>
<td>0.9 (0.7, 1.4) 0.76</td>
<td>0.7 (0.5, 0.8) 5.1E-04</td>
</tr>
</tbody>
</table>

doi:10.1371/journal.pone.0032768.t005
doi:10.1371/journal.pone.0032768.t006
changes have been described [50]. If similar genetic heterogeneity exists in sporadic ALS, meta-analysis of many association studies will be needed to generate the very large sample sizes required to reliably identify causative variants. Although results from a number of GWA studies in ALS have not been successfully replicated, two recent studies did replicate an association with a common hexanucleotide repeat in the 9p21 region that accounts for a large proportion of familial and sporadic disease in European and North American populations [17,18]. This region was initially identified through a genome-wide screen [9]. Although a risk haplotype containing expanded repeats in this region is well-tagged by a particular SNP (rs3849942), the association signal from analysis of the repeats is much stronger than the association signal from the SNP [17], which underscores the difficulty of identifying causal variants other than SNPs using chip-based GWAS analysis. Given the known genetic heterogeneity in ALS and the possibility that multiple rare, highly-penetrant variants may account for a greater proportion of currently unexplained disease than common variants with lower penetrance, exome or whole genome sequencing may prove to be more successful than GWAS analysis. Given the known genetic heterogeneity in ALS, meta-analysis of many association studies will be needed to generate the very large sample sizes required to reliably identify causative variants. Although results from a number of GWA studies in ALS have not been successfully replicated, two recent studies did replicate an association with a common hexanucleotide repeat in the 9p21 region that accounts for a large proportion of familial and sporadic disease in European and North American populations [17,18]. This region was initially identified through a genome-wide screen [9]. Although a risk haplotype containing expanded repeats in this region is well-tagged by a particular SNP (rs3849942), the association signal from analysis of the repeats is much stronger than the association signal from the SNP [17], which underscores the difficulty of identifying causal variants other than SNPs using chip-based GWAS analysis. Given the known genetic heterogeneity in ALS and the possibility that multiple rare, highly-penetrant variants may account for a greater proportion of currently unexplained disease than common variants with lower penetrance, exome or whole genome sequencing may prove to be more successful than GWAS studies in revealing the genetic underpinnings of this devastating disease.

Supporting Information

Methods S1 Details of the replication genotyping samples. (ZIP)

Methods S2 Details of genotyping and SNP QC procedures. (ZIP)

Methods S3 GWAS sample QC procedures. (ZIP)

References

Table 7. Pooled analysis: ALS outcome association with genes implicated by previous GWA studies.

<table>
<thead>
<tr>
<th>Chr</th>
<th>SNP</th>
<th>pos (build 36)</th>
<th>Gene</th>
<th>OR (95% CI)</th>
<th>P</th>
<th>OR (95% CI)</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>9</td>
<td>rs3849942*</td>
<td>27543281</td>
<td>9p21.2</td>
<td>1.1 (0.9, 1.3)</td>
<td>0.17</td>
<td>1.1 (0.9, 1.2)</td>
<td>0.46</td>
</tr>
<tr>
<td>7</td>
<td>rs17174381</td>
<td>153709165</td>
<td>DPP6</td>
<td>2.1 (1.3, 3.5)</td>
<td>0.0051</td>
<td>1.9 (1.3, 2.8)</td>
<td>4.4E-04</td>
</tr>
<tr>
<td>7</td>
<td>rs10260404*</td>
<td>153841731</td>
<td>DPP6</td>
<td>1.1 (0.9, 1.2)</td>
<td>0.47</td>
<td>1.0 (0.9, 1.1)</td>
<td>0.97</td>
</tr>
<tr>
<td>8</td>
<td>rs69850569</td>
<td>28140207</td>
<td>EL3P</td>
<td>0.7 (0.6, 0.9)</td>
<td>2.7E-04</td>
<td>0.8 (0.7, 0.9)</td>
<td>4.8E-04</td>
</tr>
<tr>
<td>8</td>
<td>rs13268933*</td>
<td>28111515</td>
<td>EL3P</td>
<td>1.0 (0.8, 1.1)</td>
<td>0.83</td>
<td>1.0 (0.9, 1.1)</td>
<td>0.97</td>
</tr>
<tr>
<td>12</td>
<td>rs11048556</td>
<td>26559935</td>
<td>ITPR2</td>
<td>1.3 (0.9, 1.7)</td>
<td>0.14</td>
<td>1.4 (1.1, 1.9)</td>
<td>0.0068</td>
</tr>
<tr>
<td>12</td>
<td>rs2306677*</td>
<td>26527653</td>
<td>ITPR2</td>
<td>1.0 (0.7, 1.3)</td>
<td>0.84</td>
<td>0.9 (0.7, 1.1)</td>
<td>0.25</td>
</tr>
<tr>
<td>19</td>
<td>rs3746200</td>
<td>17573374</td>
<td>UNC13A</td>
<td>0.8 (0.6, 1.0)</td>
<td>0.050</td>
<td>0.7 (0.6, 0.9)</td>
<td>0.0068</td>
</tr>
<tr>
<td>19</td>
<td>rs12608932*</td>
<td>17613689</td>
<td>UNC13A</td>
<td>1.2 (1.0, 1.5)</td>
<td>0.039</td>
<td>1.2 (1.0, 1.4)</td>
<td>0.046</td>
</tr>
</tbody>
</table>

*Associated SNPs from previous GWAS.
+Results from discovery phase only (SNP failed in replication).

doi:10.1371/journal.pone.0032768.t007

Table S1 Quality control of discovery samples. (ZIP)

Table S2 Samples excluded from discovery phase. (ZIP)

Acknowledgments
We would like to thank the many ALS patients and controls who have generously given their time to participate in this research. We would also like to thank Valerie Loiacono and Catherine Stawnyck for recruitment and sample collection of GENEVA controls; Kelli Allen and Barbara Norman for recruitment and sample collection of NRVA cases; and the neurologists in the VA ALS Registry group, who reviewed all patient medical records included in the discovery phase of this study: Edward J. Kasarskis, MD PhD, Lexington VAMC and University of Kentucky Medical Center, Lexington, KY; Richard S. Bedlack, MD PhD, Joel C. Morgenlander, MD, and Marvin P. Rozear, MD, Durham VAMC and Duke University Medical Center, Durham, NC; Arman Sabet, MD, Neuroscience Department, Gold Coast Hospital, Southport, Australia (formerly Lexington VAMC and University of Kentucky Medical Center, Lexington, KY); and Laura Sams, MD, Cincinnati VAMC and University of Cincinnati Medical Center, Cincinnati, OH. The data from this study will be made available to VA- and non-VA researchers upon request. A VA data sharing portal is in the process of being established and will be governed by VA Cooperative Studies Program (CSP) policy. Research proposals for data use will be reviewed by an existing CSP Scientific Review Committee.

Author Contributions
Conceived and designed the experiments: SS MAH SM EZO. Performed the experiments: JRG AS. Analyzed the data: LCK CH SS. Contributed reagents/materials/analysis tools: FK LMN BT SKV CMT MEC DLG. Performed the experiments: JRG AS. Analyzed the data: LCK CH SS. Contributed reagents/materials/analysis tools: FK LMN BT SKV CMT MEC DLG. Written the paper: LCK YL SS MAH. Oversaw experiment: YL SAS MAH. Obtained funding: EZO SS.


