A common haplotype lowers *SPI1* (PU.1) expression in myeloid cells and delays age at onset for Alzheimer's disease

Kuan-lin Huang[1,2*], Edoardo Marcora[3,4*], Anna A Pimenova[4], Antonio F Di Narzo[3], Manav Kapoor[3,4], Sheng Chih Jin[5], Oscar Harari[6], Sarah Bertelsen[4], Benjamin P Fairfax[7], Jake Czajkowski[8], Vincent Chouraki[9], Benjamin Grenier-Boley[10,11,12], Céline Bellenguez[10,11,12], Yuetiva Deming[6], Andrew McKenzie[3], Towfique Raj[3,4], Alan E Renton[4], John Budde[6], Albert Smith[13], Annette Fitzpatrick[14], Joshua C Bis[15], Anita DeStefano[16], Hieab HH Adams[17], M Arfan Ikram[17], Sven van der Lee[17], Jorge L. Del-Aguila[6], Maria Victoria Fernandez[6], Laura Ibañez[6], The International Genomics of Alzheimer's Project, The Alzheimer's Disease Neuroimaging Initiative[#], Rebecca Sims[18], Valentina Escott-Price[18], Richard Mayeux[19,20,21], Jonathan L Haines[22], Lindsay A Farrer[12,16,23,24,25], Margaret A. Pericak-Vance[25,26], Jean Charles Lambert[10,11,12], Cornelia van Duijn[17], Lenore Launer[27], Sudha Seshadri[9], Julie Williams[18], Philippe Amouyel[10,11,12,28], Gerard D Schellenberg[29], Bin Zhang[3], Ingrid Borecki[30], John S K Kauwe[31], Carlos Cruchaga[6], Ke Hao[3], Alison M Goate[3,4$]

[1] Department of Medicine, [2]McDonnell Genome Institute, [6]Department of Psychiatry, [8]Department of Genetics, Washington University in St. Louis, Saint Louis, MO, USA
[3]Department of Genetics and Genomic Sciences, [4]Ronald M. Loeb Center for Alzheimer's disease, Department of Neuroscience, Icahn School of Medicine at Mount Sinai, New York, NY, USA
[5]Department of Genetics, Yale University School of Medicine, New Haven, CT, USA
[7]Wellcome Trust Centre for Human Genetics, Nuffield Department of Medicine, University of Oxford, Oxford, United Kingdom
[9]Department of Neurology, [23]Department of Medicine (Biomedical Genetics), [24]Department of Ophthalmology, Boston University School of Medicine, Boston, MA, USA
[10]Inserm, U1167, RID-AGE –Risk factors and molecular determinants of aging-related diseases, F-59000 Lille, France
[11]Univ. Lille - Excellence laboratory Labex DISTALZ, F-59000 Lille, France
[12]Institut Pasteur de Lille, F-59000 Lille, France
[13]University of Iceland, Reykjavik, Iceland
[14]Department of Epidemiology, [15]Department of Medicine, University of Washington, Seattle, Washington, USA
[16]Department of Biostatistics, [25]Department of Epidemiology, Boston University School of Public Health, Boston, MA, USA
[17]Department of Epidemiology, Erasmus University Medical Center, Rotterdam, The Netherlands
[18] Psychological Medicine and Clinical Neurosciences, Medical Research Council (MRC) Centre for Neuropsychiatric Genetics and Genomics, Cardiff University, Cardiff, UK
[19]Taub Institute on Alzheimer's Disease and the Aging Brain, [20]Gertrude H. Sergievsky Center, [21]Department of Neurology, Columbia University, New York, NY, USA
[22]Department of Epidemiology and Biostatistics, Case Western Reserve University, Cleveland, OH, USA
[25]The John P. Hussman Institute for Human Genomics, [26]Macdonald Foundation Department of Human Genetics, University of Miami, Miami, FL, USA

[27]Laboratory of Epidemiology and Population Sciences, National Institute on Aging, Bethesda, Maryland, USA
[28]Centre Hospitalier Universitaire de Lille, U1167, F-59000 Lille, France
[29]Department of Pathology and Laboratory Medicine, University of Pennsylvania Perelman School of Medicine, Philadelphia, PA, USA
[30]Regeneron Pharmaceuticals, Inc, NY, USA
[31]Department of Biology, Brigham Young University, Provo, Utah, USA

[*] These authors contributed equally to this work

[$] Corresponding Author:
Alison Goate, D.Phil.
Willard T.C. Johnson Research Professor of Neurogenetics
Director, Ronald M. Loeb Center for Alzheimer's disease
Dept. of Neuroscience, B1065
Icahn School of Medicine at Mount Sinai
1425 Madison Ave
New York, NY 10029
T: 212-659-5672
E-mail: alison.goate@mssm.edu

Data used in preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in analysis or writing of this report. A complete listing of ADNI investigators can be found at:
http://adni.loni.usc.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf

Abstract

In this study we used age at onset of Alzheimer's disease (AD), cerebrospinal fluid (CSF) biomarkers, and *cis*-expression quantitative trait loci (*cis*-eQTL) datasets to identify candidate causal genes and mechanisms underlying AD GWAS loci. In a genome-wide survival analysis of 40,255 samples, eight of the previously reported AD risk loci are significantly ($P < 5x10^{-8}$) or suggestively ($P < 1x10^{-5}$) associated with age at onset-defined survival (AAOS) and a further fourteen novel loci reached suggestive significance. Using stratified LD score regression we demonstrated a significant enrichment of AD heritability in hematopoietic cells of the myeloid and B-lymphoid lineage. We then investigated the impact of these 22 AAOS-associated variants on CSF biomarkers and gene expression in cells of the myeloid lineage. In particular, the minor allele of rs1057233 (G), within the previously reported *CELF1* AD risk locus, shows association with higher age at onset of AD ($P=8.40x10^{-6}$), higher CSF levels of $A\beta_{42}$ ($P=1.2x10^{-4}$), and lower expression of *SPI1* in monocytes ($P=1.50x10^{-105}$) and macrophages ($P=6.41x10^{-87}$). *SPI1* encodes PU.1, a transcription factor critical for myeloid cell development and function. AD heritability is enriched within the *SPI1* cistromes of monocytes and macrophages, implicating a myeloid PU.1 target gene network in the etiology of AD. Finally, experimentally altered PU.1 levels are correlated with phagocytic activity of BV2 mouse microglial cells and specific changes in the expression of multiple myeloid-expressed genes, including the mouse orthologs of AD-associated genes, *APOE*, *CLU/APOJ*, *CD33*, *MS4A4A*/*MS4A6A*, and *TYROBP*. Our results collectively suggest that lower *SPI1* expression reduces AD risk by modulating myeloid cell gene expression and function.

Introduction

Alzheimer's disease (AD) is the most prevalent form of dementia. While genome-wide association studies (GWAS) have identified more than twenty AD risk loci[1–5], most of the associated disease genes and mechanisms remain unclear. To better understand these genetic associations, additional phenotypes and endophenotypes beyond disease status can be leveraged. For example, few studies[6,7] have investigated the genetic basis of age at onset (AAO) of AD. To date, *APOE* remains the only locus repeatedly shown to associate with AAO[8–11]. *PICALM* and *BIN1* – two other AD risk loci – have also been shown to affect AAO using a candidate-gene approach[6,12,13]. A large-scale genome-wide study, including both AD cases and elderly non-demented controls with age information may reveal additional loci associated with AD. Further, we have previously used cerebrospinal fluid (CSF) biomarkers associations to demonstrate that *APOE* genotype is strongly associated with both CSF $A\beta_{42}$ and total tau levels and to identify novel loci associated with these disease-relevant quantitative traits[14,15]. Combining this information may help validate and elucidate the AD genetic association landscape.

Identifying causal genes and mechanisms underlying disease-associated loci often requires integrative analyses of expression and epigenetic datasets in disease-relevant cell types[16]. Recent genetic and molecular evidence has highlighted the role of myeloid cells of the innate immune system in AD. At the genetic level, GWAS and sequencing studies have found associations between AD and genes expressed in myeloid cells, including *TREM2*, *ABCA7*, and *CD33*[1,2,5,17–19]. At the epigenetic level, genes expressed in myeloid cells display abnormal patterns of expression and chromatin modification in AD mouse models and human samples[20–22]. In addition, we have previously shown that AD-risk alleles are polarized for *cis*-expression quantitative trait locus (*cis*-eQTL) effects in monocytes[23]. Herein, we further show that AD heritability is enriched in functional annotations for immune cells of the myeloid and B-lymphoid lineage. Integrative analyses of AD GWAS datasets with myeloid gene expression and epigenetic signatures may uncover novel AD genes and mechanisms related to the function of myeloid cells (such as monocytes and macrophages, including microglia).

In this study, we conducted a genome-wide survival analysis and subsequent CSF biomarker association analysis to uncover loci associated with AAO-defined survival (AAOS) in AD cases and non-demented elderly controls. We discovered an AAOS- and CSF $A\beta_{42}$-associated SNP, rs1057233, in the previously reported *CELF1* locus. *Cis*-eQTL analyses revealed a highly significant association of the protective rs1057233^{G} allele with reduced *SPI1* expression in monocytes and macrophages. *SPI1* encodes PU.1, a transcription factor critical for myeloid and B-lymphoid cell development and function that binds to *cis*-regulatory elements associated with several AD-associated genes in monocytes and macrophages. Moreover, we show that AD heritability is enriched within the *SPI1* cistromes in these cells, implicating a myeloid PU.1 target gene network in the etiology of AD. Together, these results indicate that genetically altered PU.1 levels may modulate AD susceptibility by affecting the expression of at least some of its target genes in myeloid cells. To validate these bioinformatic analyses, we show experimentally that altered PU.1 levels are correlated with phagocytic activity of BV2 mouse microglial cells and specific changes in the expression of multiple genes involved in a diverse array of biological processes in myeloid cells. This evidence collectively shows that lower *SPI1* expression may reduce AD risk by modulating myeloid gene expression and cell function.

Results

Genome-wide survival analysis

We analyzed data from the IGAP consortium for the genome-wide survival analysis. Samples from ADGC, CHARGE, EADI, and GERAD were included for a total of 14,406 AD cases and 25,849 controls (**Table 1a**). 8,253,925 SNPs passed all quality control criteria and were included for the final meta-analysis across all cohorts (**Supplementary Table 1**), which showed little evidence of genomic inflation ($\lambda = 1.026$). Four loci showed genome-wide significant associations ($P < 5x10^{-8}$) with AAOS: *BIN1* ($P=7.6x10^{-13}$), *MS4A* ($P=5.1x10^{-11}$), *PICALM* ($P=4.3x10^{-14}$), and *APOE* ($P=1.2x10^{-67}$) (**Supplementary Fig. 1**). While SNPs within *BIN1*[6], *PICALM*[6,12], and *APOE*[6,12,24–27] loci have previously been shown or suggested to be associated with AAO, this is the first time that the *MS4A* locus is reported to be associated with an AAO-related phenotype. The minor allele of rs7930318 near *MS4A4A* is associated with a later AAO. Four other AD risk loci previously reported in the IGAP GWAS[1] showed associations that reached suggestive significance ($P < 1.0x10^{-5}$): *CR1* ($P=1.2x10^{-6}$), *SPI1/CELF1* ($P=5.4x10^{-6}$), *SORL1* ($P=1.8x10^{-7}$), and *FERMT2* ($P=1.0x10^{-5}$). The directionalities of the effects were concordant with the previous IGAP GWAS and logistic regression of the matching cohort in all suggestive loci: previously reported AD risk-increasing alleles were all associated with a hazard ratio above 1 and earlier AAO, whereas AD risk-decreasing alleles were all associated with a hazard ratio below 1 and later AAO (**Table 1b, Supplementary Table 2**). We also identified 14 novel loci that reached suggestive significance in the survival analysis, 3 of which (rs116341973, rs1625716, and rs11074412) were nominally associated with AD risk (Bonferroni-corrected threshold: $P=0.05/22=2.27x10^{-3}$) in the IGAP GWAS (**Table 1b, Supplementary Fig. 2, 3**).

Cerebrospinal fluid biomarkers associations

To further validate the 22 loci with at least suggestive associations to AAO, we examined their associations with established CSF biomarkers, including total tau, phosphorylated tau_{181}, and $A\beta_{42}$ in a dataset of 3,646 Caucasians extended from our previous report[14] (**Table 2**). Two SNPs showed associations that reached the Bonferroni-corrected threshold ($P < 2.27\times10^{-3}$). Rs4803758 near *APOE* showed the most significant associations with levels of CSF phosphorylated tau_{181} ($P=5.81\times10^{-4}$) and CSF $A\beta_{42}$ ($P=6.75\times10^{-5}$), whereas rs1057233 in the *SPI1/CELF1* locus was significantly associated with CSF $A\beta_{42}$ ($P=4.11\times10^{-4}$). Of note, a SNP adjacent to *VLDLR*, rs7867518, showed the most significant association with CSF total tau ($P=3.02\times10^{-3}$), but failed to pass the Bonferroni-corrected threshold. The protective and deleterious effects in the survival analysis of these three SNPs were concordant with directionalities of their CSF biomarkers associations; for example, the protective $rs1057233^{G}$ allele was associated with higher CSF $A\beta_{42}$ levels and the risk $rs1057233^{A}$ allele was associated with lower CSF $A\beta_{42}$ levels.

Cis-eQTL associations and colocalization analysis

Multiple disease-associated GWAS SNPs have been identified as *cis*-eQTLs of disease genes in disease-relevant tissues/cell types[28]. We investigated *cis*-eQTL effects of the 22 AD survival-associated SNPs and their tagging SNPs ($R^2 \geq 0.8$, listed in **Supplementary Table 3**) in the BRAINEAC dataset encompassing ten different brain regions. We identified 4 significant associations (Bonferroni-corrected threshold: P=0.05/292,000 probes = 1.7×10^{-7}): rs1057233 was

associated with *MTCH2* expression in the cerebellum ($P=1.20\times10^{-9}$); rs7445192 was associated with averaged *SRA1* expression across brain regions ($P=7.0\times10^{-9}$, 1.6×10^{-7} for two probes respectively), and rs2093761 was associated with *CR1/CR1L* expression in the white matter ($P=1.30\times10^{-7}$, **Supplementary Table 4**). Further analysis using the GTEx dataset[29] identified 50 unique, associated snp-gene pairs across 44 tissues, including 11 snp-gene pairs in various brain regions (**Supplementary Table 5**). Rs1057233 was associated with *C1QTNF4* across 18 tissues and *MTCH2* in the brain cortex and nucleus accumbens/basal ganglia.

In recent years, substantial genetic and molecular evidence has implicated myeloid cells of the innate immune system in the etiology of AD, including our own finding that AD risk alleles are enriched for *cis*-eQTL effects in monocytes but not CD4+ T-lymphocytes[23]. To extend this finding and identify relevant cell types in AD, we used stratified LD score regression to estimate enrichment of AD heritability (as measured by GWAS summary statistics from the IGAP consortium[1]) partitioned by 220 cell type–specific functional annotations as described by Finucane et al.[30]. We found a significant enrichment of AD heritability in hematopoietic cells of the myeloid and B-lymphoid lineage (e.g., 14.49 fold enrichment, $P=3.49\text{x}10^{-5}$ in monocytes/CD14 enhancers/H3K4me1 and 12.33 fold enrichment, $P=1.41\text{x}10^{-6}$ in B-cells/CD19 enhancers/H3K4me1). In contrast schizophrenia (SCZ) heritability was not enriched in hematopoietic cells (1.24 fold enrichment, P=0.53, as measured by GWAS summary statistics from the Psychiatric Genomics Consortium[31]) but was significantly enriched in brain tissue (18.61 fold enrichment, $P=1.38\text{x}10^{-4}$ in fetal brain promoters/H3K4me3, **Supplementary Table 6**). These results highlight a specific contribution of myeloid cells to the modulation of AD susceptibility.

Based on these observations, we hypothesized that *cis*-eQTL effects of some AD-associated alleles may be specific to myeloid cells and thus not easily detectable in *cis*-eQTL datasets obtained from brain homogenates where myeloid cells (microglia and other brain-resident macrophages) constitute only a minor fraction of the tissue. Therefore, we analyzed *cis*-eQTL effects of the AD survival-associated SNPs and their tagging SNPs in human *cis*-eQTL datasets composed of 738 monocyte and 593 macrophage samples from the Cardiogenics consortium[32]. We identified 14 genes with *cis*-eQTLs significantly associated with these SNPs (**Table 3**). Notably, the protective rs1057233^{G} allele, located within the 3' UTR of *SPI1*, was strongly associated with lower expression of *SPI1* in both monocytes ($P=1.50\times10^{-105}$) and macrophages ($P=6.41\times10^{-87}$) with similar dose-dependent effects (**Fig. 1a, 1c**). This allele was also associated with lower expression of *MYBPC3* (monocytes: $P=5.58\times10^{-23}$; macrophages: $P=4.99\times10^{-51}$), higher expression of *CELF1* in monocytes ($P=3.95\times10^{-8}$) and lower *NUP160* expression in macrophages ($P=5.35\times10^{-22}$). Each of these genes lies within the *SPI1/CELF1* locus, suggesting complex regulation of gene expression in this chromosomal region. Within the *MS4A* locus, which contains many gene family members, the minor allele (C) of rs7930318 was consistently associated with lower expression of *MS4A4A* in monocytes ($P=8.20\times10^{-28}$) and *MS4A6A* in monocytes and macrophages (**Fig. 1b**, monocytes: $P=4.90\times10^{-23}$; macrophages: $P=1.25\times10^{-9}$). Among the novel AD survival-associated loci, rs5750677 was significantly associated with lower expression of *SUN2* in both monocytes ($P=3.66\times10^{-58}$) and macrophages ($P=3.15\times10^{-36}$), rs10919252 was associated with lower expression of *SELL* in monocytes ($P=7.33\times10^{-35}$), and rs1625716 was associated with lower expression of *CISD1* in macrophages ($P=5.98\times10^{-23}$, **Table 3**).

We then sought evidence of replication in an independent dataset of primary CD14+ human monocytes from 432 individuals of European ancestry[33]. We replicated *cis*-eQTL associations with expression of *SPI1*, *MYBPC3*, *MS4A4A*, *MS4A6A*, and *SELL* (Bonferroni-corrected threshold: P=0.05/15421 probes = 3.24×10^{-6}). We found strong evidence for the association between rs1057233 and *SPI1* expression ($P=6.39\times10^{-102}$) as well as *MYBPC3* expression ($P=5.95\times10^{-33}$, **Supplementary Table 7**). Rs1530914 and rs7929589, both in high LD with rs7930318 (R^2 = 0.99 and 0.87, respectively), were associated with expression of *MS4A4A* and *MS4A6A* ($P=3.60\times10^{-8}$, 6.37×10^{-15}), respectively. Finally, rs2272918, tagging rs10919252, was significantly associated with expression of *SELL* ($P=8.43\times10^{-16}$). Interestingly, the minor allele of all of these SNPs with replicated *cis*-eQTL associations showed protective effects in both AD risk and survival analyses, and are each correlated with lower expression of their associated gene. Further, *SPI1*, *MS4A4A*, *MS4A6A*, and *SELL* are all specifically expressed in microglia based on RNA-Seq data[34–36] from human and mouse acutely-isolated brain cell types (**Fig. 1d**, **Supplementary Fig. 4**). However, *MYBPC3*/Mybpc3 (a myosin binding protein expressed at high levels in cardiac muscle cells) is either not expressed or expressed at low levels in human and mouse microglia, respectively. Amongst all genes probed, *MYBPC3* (ILMN_1781184) expression is the most highly and significantly correlated with *SPI1* (ILMN_1696463) expression in both Cardiogenics datasets (Spearman's rho = 0.54, qval = 0.00 in monocytes and Spearman's rho = 0.42, qval = 0.00 in macrophages) suggesting that low levels of *MYBPC3* expression in human myeloid cells are possibly due to leaky transcription driven by the adjacent highly expressed *SPI1* gene[37] (**Supplementary Fig. 6d**).

We then performed the coloc statistical test[38] to determine whether AD survival-associated SNPs co-localize with myeloid *cis*-eQTLs at the *SPI1/CELF1*, *MS4A* and *SELL* loci. The results of these analyses (**Supplementary Table 8**) highlighted *SPI1* at the *SPI1/CELF1* locus as the strongest and most consistent colocalization target, and the only gene where the AD survival and gene expression association signals are likely (posterior probability $\geq$ 0.8) driven by the same causal genetic variant, in both monocytes and macrophages (PP.H4.abf of 0.85 and 0.83, respectively). *MYBPC3* in the *SPI1/CELF1* locus and *MS4A6A* in the *MS4A* locus also showed evidence of colocalization in both myeloid cell types, but they did not survive posterior probability cutoff in one of the cell types. *MS4A4A* and *MS4A6E* in the *MS4A* locus showed evidence of co-localization only in monocytes, while *SELL* did not show evidence of colocalization in either cell type.

In light of the strong *cis*-eQTL effects and colocalization results described above, we decided to focus subsequent analyses on *SPI1* at the *SPI1/CELF1* locus as the strongest candidate gene underlying the disease association in both monocytes and macrophages.

Conditional and SMR analysis of the *SPI1/CELF1* locus

The AD survival-association landscape shows that highly associated SNPs at the *SPI1/CELF1* locus span a region across multiple genes (**Fig. 1a**). In the previous IGAP GWAS logistic regression analysis of AD risk[1], rs10838725 showed the strongest association at this locus ($P=6.7\text{x}10^{-6}$, $1.1\text{x}10^{-8}$ vs. rs1057233: $P=5.4\text{x}10^{-6}$, $5.9\text{x}10^{-7}$ in IGAP stage I, stage I and II combined, respectively). Rs10838725 is located in the intron of the *CELF1* gene, which was

assigned as the putative causal gene at this locus[1] based on proximity to the index SNP, a criterion that has often proven to be erroneous[16]. In our genome-wide survival analysis, however, rs10838725 showed weak association (P=0.12, HR=1.02, 95% CI=0.99-1.05) whereas rs1057233, located in the 3'UTR of a neighboring gene, *SPI1*, showed the most significant association (**Table 1**, $P=5.4x10^{-6}$). The two SNPs exhibit only moderate linkage disequilibrium in the ADGC subset of the IGAP GWAS dataset (R^2=0.21, D'=0.96). Applying conditional logistic regression analysis of AD risk in the ADGC dataset, we found that rs1057233 remained significantly associated with AD after adjusting for rs10838725 ($P=3.2x10^{-4}$), whereas rs10838725 showed no evidence of association after adjusting for rs1057233 (P=0.66). This suggests that rs1057233 is in stronger LD with the AD risk causal variant.

The association landscape in the AD survival analysis highly resembles that of *SPI1 cis*-eQTL analysis in myeloid cells (**Fig. 1a**). We reasoned that the associations of rs1057233 with AD-related phenotypes may be explained by the regulation of *SPI1* expression in myeloid cells, and that conditional analysis of the *cis*-eQTL signal could help us further dissect this complex locus. Therefore, we conducted conditional *cis*-eQTL analyses in both Cardiogenics datasets as we did above using rs1057233 (the top SNP for AD survival) and rs10838725 (the top SNP for AD risk). In addition, we also examined rs10838698 (a SNP in high LD with rs1057233 that was directly genotyped in the Cardiogenics datasets) and rs1377416, a SNP in high LD with rs10838725 that has been proposed to be a functional variant in an enhancer near *SPI1* that is active in human myeloid cells and in the brain of a mouse model of AD[21]. It should also be noted that rs1057233 is a functional variant that has been shown in *in vitro* experiments to directly affect *SPI1* expression by changing the target sequence and binding of miR-569[39]. Rs1057233 and rs10838698 remained significantly associated with *SPI1* expression when adjusting for either of the other two SNPs in both monocytes and macrophages ($P < 8.33x10^{-3}$). On the other hand, conditioning for either of these two SNPs abolished the associations of rs1377416 and rs10838725 with *SPI1* expression (**Supplementary Table 9**). Thus, the functional variant(s) mediating the effect on *SPI1* expression and AD risk likely resides in the LD block that includes rs1057233 and rs10838698 but not rs10838725 and rs1377416 (**Supplementary Fig. 5**).

Using HaploReg[40] to annotate the top survival SNP (rs1057233) and its tagging SNPs (R^2 >= 0.8, **Supplementary Table 3**), we identified multiple SNPs (e.g, rs10838699 and rs7928163) in tight LD with rs1057233 that changed the predicted DNA binding motif of SPI1 (PU.1), raising an additional possibility of altered self-regulation and potentially decreased SPI1 (PU.1) binding and transactivation in the presence of the minor allele. Based on these results, one or more of these or other SNPs in very high LD with rs1057233, could explain the observed associations with *SPI1* expression and AD-related phenotypes.

As a complement to the colocalization and conditional analyses described above, we conducted Summary-data-based Mendelian Randomization (SMR) and Heterogeneity In Dependent Instruments (HEIDI) tests[28] to prioritize likely causal genes and variants by integrating summary statistics from our AAOS GWAS and the Cardiogenics study (**Supplementary Table 10**). SMR/HEIDI analysis was performed for the *SPI1/CELF1* locus using rs1057233, rs10838698, rs10838699, rs7928163, rs10838725 and rs1377416 as candidate causal variants. In both monocytes and macrophages, *SPI1* was consistently identified as the most likely gene whose expression levels are associated with AD survival because of causality/pleiotropy at the same

underlying causal variant (rs1057233, rs10838698, rs10838699, or rs7928163 in the same LD block) (SMR P < 4.90E-04, Bonferroni-corrected threshold for 6 SNPs tested against 17 probes and HEIDI P ≥ 0.05, **Supplementary Fig. 6**). Neither conditional analysis nor this SMR/HEIDI analysis could definitively identify a single functional variant among this set of 4 SNPs in high LD. Functional analyses will be necessary to determine which of these or other SNPs in the same LD block directly affects *SPI1* expression. Overall, rs1057233 and tagging SNPs are associated with AD risk and survival, and CSF $A\beta_{42}$. The strong *cis*-eQTL effects and colocalization results point to *SPI1* as the most likely candidate gene underlying the disease association at the *SPI1*/*CELF1* locus.

SPI1 (PU.1) cistrome and functional analysis in myeloid cells

To further evaluate *SPI1* as a candidate causal gene for AD, we investigated the functional impact of variation in *SPI1* expression. *SPI1* encodes PU.1, a transcription factor essential for the development and function of myeloid cells. We hypothesized that it may modulate AD risk by regulating the transcription of AD-associated genes that are expressed in microglia and/or other myeloid cell types. First, we tested AD-associated genes for evidence of expression in human microglia/brain myeloid cells[34] as well as presence of PU.1 binding peaks in *cis*-regulatory elements associated with these genes using ChIP-Seq datasets obtained from human monocytes and macrophages[41]. We specifically investigated 112 AD-associated genes, including the 104 genes located within the IGAP GWAS loci as defined by Steinberg et al.[42] and additionally *APOE*, *APP*[43], *TREM2* and *TREML2*[44], *TYROBP*[20], *TRIP4*[45], *CD33*[2,5], and *PLD3*[46]. Among the 112 AD-associated genes, 75 had evidence of gene expression in human microglia/brain myeloid cells, 60 of which also had evidence of association with one or more nearby PU.1 binding sites in human blood myeloid cells (monocytes or macrophages)[41] (**Supplementary Table 11**). Further examination of PU.1 binding peaks and chromatin marks/states in human monocytes and macrophages confirmed that PU.1 is bound to *cis*-regulatory elements in the proximity of many AD-associated genes, including *ABCA7*, *CD33*, *MS4A4A*, *MS4A6A*, *PILRA*, *PILRB*, *TREM2*, *TREML2*, and *TYROBP* (as well as *SPI1* itself, but notably not *APOE*) in cells of the myeloid lineage (**Fig. 1e**, **Supplementary Fig. 7**). Together, these results suggest that PU.1 may regulate the expression of multiple AD-associated genes in myeloid cells.

To further support the hypothesis that a network of PU.1 target genes expressed in myeloid cells such as microglia may be associated with AD risk, we used stratified LD score regression[30] to estimate enrichment of AD heritability (as measured by GWAS summary statistics from the IGAP consortium[1]) partitioned across the whole PU.1 cistrome, as profiled by ChIP-Seq in human monocytes and macrophages[41]. Indeed, we found a significant enrichment of AD heritability in both monocytes (47.58 fold enrichment, $P=6.94\times10^{-3}$) and macrophages (53.88 fold enrichment, $P=1.65\times10^{-3}$), but this was not the case for schizophrenia (SCZ) heritability [as measured by GWAS summary statistics from the Psychiatric Genomics Consortium (PGC)[31]] (**Supplementary Table 12**), suggesting that the contribution of the myeloid PU.1 target gene network to disease susceptibility is specific to AD. However, since PU.1 is a key myeloid transcription factor that regulates the expression of a large number of genes in monocytes and macrophages, the enrichment of AD risk alleles in PU.1 binding sites could simply reflect an enrichment of AD GWAS associations for genes that are expressed in myeloid cells, rather than

specifically among PU.1 target genes within these cells. To attempt to address this issue, we performed stratified LD score regression of AD heritability partitioned by functional annotations obtained from SPI1 (marking the PU.1 cistrome) and POLR2AphosphoS5 (marking actively transcribed genes) ChIP-Seq experiments, performed in duplicate, using a human myeloid cell line (HL60) by the ENCODE Consortium[47]. We observed a significant enrichment for SPI1 (PU.1) (34.58 fold enrichment, $P=1.31x10^{-3}$ in first replicate; 58.12 fold enrichment, $P=4.95x10^{-3}$ in second replicate) much stronger than that for POLR2AphosphoS5 (15.78 fold enrichment, $P=1.71x10^{-2}$ in first replicate; 16.34 fold enrichment, $P=1.25x10^{-1}$ in second replicate), further supporting our hypothesis that a myeloid PU.1 target gene network specifically modulates AD risk (**Supplementary Table 12**).

PU.1 target genes are implicated in various biological processes within myeloid cells that may modulate AD risk. For example, a microglial gene network for pathogen phagocytosis has been previously implicated in the etiology of AD[20] and we developed a cell-based assay to investigate the role of PU.1 in this process. We modulated levels of PU.1 by cDNA overexpression or shRNA knock-down of *Spi1* in BV2 mouse microglial cells, and used zymosan bioparticles labeled with pHrodo (a pH-sensitive dye that emits an intense fluorescent signal when internalized in acidic vesicles during phagocytosis) to measure pathogen engulfment. Analysis of zymosan uptake by flow cytometry revealed that phagocytic activity is augmented in BV2 cells overexpressing PU.1 (**Fig. 2a**), while knock-down of PU.1 resulted in a significant decrease in phagocytic activity (**Fig. 2a**). We confirmed overexpression and knock-down of PU.1 expression levels by western blotting and qPCR (**Fig. 2b, 2c, 2d, 3a**). Phagocytic activity was not changed in the population of cells with unperturbed PU.1 expression levels when analyzed by flow cytometry (**Supplementary Fig. 8d, 8e, 8f, 8g**). Taken together, these data suggest that modulation of PU.1 expression levels results in significant changes in microglial phagocytic function in response to fungal targets (mimicked by zymosan).

To further explore the functional impact of variation in *SPI1* expression in myeloid phagocytes, we performed qPCR analysis to test whether differential *Spi1* expression in BV2 mouse microglial cells modulates levels of myeloid genes that are thought to play important roles in AD pathogenesis and/or microglial cell function (**Fig. 3**, **Supplementary Fig. 9**, **Supplementary Table 13, 14**). We found that levels of some of these genes were affected in opposing directions by overexpression and knock-down of *Spi1* (**Fig. 3a**), while other genes were affected only by overexpression (**Fig. 3b**) or knock-down (**Fig. 3c**) or not affected at all (**Supplementary Fig. 9**). In particular, overexpression of *Spi1* led to upregulation of *Ccl2*, *Cxcl2*, *Aif1*, *Ms4a4a*, *Ms4a6d* (mouse ortholog of human *MS4A6A*), *Fcgr1*, *Pilrb2*, *Cd36* and down-regulation of *Il34*, *Apoe*, *Clu/ApoJ*. On the other hand, knock-down of *Spi1* led to up-regulation of *Il34*, *Apoe*, *Clu/ApoJ*, *Csf1*, *Cx3cr1*, *Axl*, *Serpinb1b* and down-regulation of *Ccl2*, *Cxcl2*, *Aif1*, *Ms4a4a*, *Ms4a6d*, *Fcgr1*, *Pilrb2*, *Il1b*, *Csf1r*, *P2ry12*, *Pilra*, *Itgam*, *Cd33*, *Tyrobp*, *Nos2*, *Ptgs2*, *Arg1*, *Ctsb*, *Nlrp3*. These data demonstrate that multiple microglial genes (many of which have already been implicated in the etiology of AD) are selectively perturbed by altered expression of *Spi1*, suggesting a collective and coordinated effect on several microglial cell functions (phagocytosis,

inflammatory response, migration/chemotaxis, proliferation/survival, lipid/cholesterol metabolism, etc.) that are thought to play a role in AD pathogenesis.

Discussion

In this study, we discovered multiple loci associated with AAO of AD in a genome-wide survival analysis (**Table 1**). The four genome-wide significantly associated loci, *BIN1* ($P=7.6x10^{-13}$), *MS4A* ($P=5.1x10^{-11}$), *PICALM* ($P=4.3x10^{-14}$), and *APOE* ($P=1.2x10^{-67}$), have all been previously reported to be associated with AD risk[1]. Notably, this is the first study showing that the *MS4A* locus is associated with AAO of AD. The most significantly associated SNP in the *MS4A* gene cluster, rs7930318, shows a protective effect (HR = 0.93, 95% CI = 0.90-.95) in the survival analysis, consistent with the result from the previous IGAP GWAS logistic regression analysis for AD risk (OR = 0.90, 95% CI = 0.87-.93).

By combining association results of AAO and CSF biomarkers, we provide evidence of AD association at additional loci (**Table 2**). In particular, rs7867518 at the *VLDLR* locus shows suggestive associations with both AD survival ($P=9.1x10^{-6}$) and CSF tau ($P=3.03x10^{-3}$). An adjacent SNP rs2034764 in the neighboring gene, *KCNV2*, has been previously reported to show suggestive association with AAO[26]. VLDLR, or the very-low-density-lipoprotein receptor, binds to APOE-containing lipoproteins in the brain[48] and physically interacts with CLU, another AD risk gene[49]. Additionally, the *VLDLR*-5-repeat allele was found to be associated with dementia[48]. Collectively, this evidence suggests that genetic variation in *VLDLR* may be linked to APOE and AD, although further investigation is required.

Cis-eQTL analyses of AD survival-associated SNPs revealed limited associations when using brain tissue homogenate data, yet identified multiple candidate genes when using data obtained from cells of the myeloid lineage, which we have found to be the most likely candidate causal cell types for AD based on stratified LD score regression analysis of AD heritability. This result calls attention to careful selection of relevant cell types in eQTL studies of disease association. In particular, by conducting *cis*-eQTL analyses using monocyte and macrophage datasets, we discovered associations of AD survival-associated SNPs with the expression of *SELL*, *SPI1*, *MYBPC3*, *NUP160*, *MS4A4A*, *MS4A6A* and *SUN2* (**Table 3**). Furthermore, we replicated the *cis*-eQTL associations of rs1057233 with *SPI1*, *MYBPC3*, rs7930318 with *MS4A4A*, *MS4A6A* and rs2272918 with *SELL* in an independent monocyte dataset. We further showed that the *SPI1* myeloid *cis*-eQTLs and AD survival-associated SNPs are not likely to be colocalized by chance and thus may be in the causal pathway to AD (**Fig. 1**), providing additional support for the hypothesis that modulation of *SPI1* expression likely contributes to the disease association at the *SPI1/CELF1* locus.

Notably, the minor allele of rs1057233 (G) at the previously reported *SPI1/CELF1* locus is suggestively associated with lower AD risk ($P=5.4x10^{-6}$, $5.9x10^{-7}$ in IGAP stage I, stage I and II combined, respectively)[1], higher AAO defined survival ($P=8.4x10^{-6}$) and significantly associated with higher CSF $A\beta_{42}$ ($P=4.11x10^{-4}$), which likely reflects decreased Aβ aggregation and β-amyloid deposition in the brain. Furthermore, it is strongly associated with lower *SPI1* expression in human monocytes ($P=1.50x10^{-105}$) and macrophages ($P=6.41x10^{-87}$, **Table 3**).

Colocalization analyses using coloc and SMR/HEIDI support the hypothesis that the same causal SNP(s) influence *SPI1* expression and AD risk. However, neither conditional nor SMR/HEIDI analyses were able to pin-point an individual SNP, but rather both approaches identified a single LD block, tagged by rs1057233, in which one or more SNPs may individually or in combination influence both *SPI1* expression and AD risk. rs1057233 directly changes the target sequence and binding of miR-569[39], and its tagging SNPs alter binding motifs of transcription factors including PU.1 itself (**Supplementary Table 3 and Supplementary Fig. 7d**). Another SNP, rs1377416, is located in a predicted enhancer in the vicinity of *SPI1* and exhibited enhancer activity when assayed *in vitro* using an episomal luciferase reporter construct transfected in BV2 mouse microglia cells[21]. However, rs1057233 remained significantly associated with AD after conditioning for either rs1377416 ($P=1.2\times10^{-3}$) or the previously reported IGAP GWAS top SNP rs10838725 ($P=3.2\times10^{-4}$) in the ADGC dataset. Further, the *cis*-eQTL association between rs1057233 and *SPI1* expression remained significant after conditioning for both of these SNPs, whereas conditioning for rs1057233 abolished their *cis*-eQTL associations with *SPI1* (**Supplementary Table 9**). Thus, rs1057233 and its tagging SNPs likely represent the underlying disease locus and may modulate AD risk through variation in *SPI1* expression. Interestingly, rs1057233 was previously found to be associated with systemic lupus erythematosus[39], body mass index[50] and proinsulin levels[51] and may potentially contribute to the connection between AD, immune cell dysfunction, obesity and diabetes.

PU.1 binds to *cis*-regulatory elements of several AD-associated genes expressed in myeloid cells, including *ABCA7*, *CD33*, *MS4A4A*, *MS4A6A*, *TREM2*, and *TYROBP* (**Fig. 1e**, **Supplementary Fig. 7**). This finding is further supported by PU.1 binding to active enhancers of *Trem2* and *Tyrobp* in ChIP-Seq experiments using BV2 mouse microglial cell line[52] or bone marrow-derived mouse macrophages[53]. PU.1 is required in mouse for the development and function of myeloid and B-lymphoid cells[54,55]. In particular, PU.1 expression is dynamically and tightly controlled during haematopoiesis to direct the specification of CD34+ hematopoietic stem and progenitor cells toward the myeloid and B-lymphoid lineage by progressively partitioning into CD14+ monocytes/macrophages, CD15+ neutrophils, and CD19+ B cells[56], which are the very same cell types that our stratified LD score regression analysis prioritized as being causal for AD. Given its selective expression in microglia in the brain (**Fig. 1d**), PU.1 may modify microglial cell function through transcriptional regulation of target genes that act as downstream modulators of AD susceptibility, as evidenced by the significant enrichment of AD heritability partitioned by PU.1 ChIP-Seq binding sites in human myeloid cells across the whole genome (**Supplementary Table 12**).

In support of this hypothesis, we also demonstrate that changes in PU.1 expression levels result in the alteration of phagocytic activity in the BV2 mouse microglial cell line (**Fig. 2, Supplementary Fig. 8**). Knock-down of PU.1 expression reduced engulfment of zymosan, whereas overexpression of PU.1 increased engulfment of zymosan, a Toll-like receptor 2 (TLR2) agonist that mimics fungal pathogens. This is in line with previous data showing decreased uptake of $A\beta_{42}$ (also a TLR2 agonist) in primary microglial cells isolated from adult human brain tissue and transfected with siRNA targeting *SPI1*[57]. Interestingly, several AD-associated genes (e.g., *CD33*[19,58], *TYROBP*, *TREM2*[59,60], *TREML2*, *CR1*, *ABCA7*[61], *APOE*[60], *CLU/APOJ*[60]) have been shown to be involved in phagocytosis of pathogens or host-derived cellular material (e.g.,

β-amyloid, apoptotic cells, myelin debris, lipoproteins, etc.), suggesting a strong link between perturbation of microglial phagocytosis and AD pathogenesis.

After knock-down of *Spi1* in BV2 microglial cells, expression of *Cd33* and *Tyrobp* decreased and expression of *Apoe* and *Clu/ApoJ* increased (**Fig. 3a, 3c**). Indeed, several other genes are dysregulated after altering *Spi1* expression, i.e. *Cd36*, *Fcgr1*, *Pilra*, *Pilrb2*, *Ms4a4a*, *Ms4a6d*, *P2ry12*, *Itgam*, *Cx3cr1*, *Axl*, *Ctsb* (**Fig. 3a, 3b, 3c**), suggesting a collective and coordinated effect of *Spi1* on the phagocytic activity of BV2 microglial cells. Furthermore, expression of *Il1b*, *Nos2*, *Ptgs2*, *Arg1*, and *Nlrp3* decreased after knock-down of *Spi1* (**Fig. 3c**), consistent with blunting of the inflammatory response that is often up-regulated in AD brains and regarded as neurotoxic. Moreover, our genetic analyses show that the protective allele within the MS4A locus is associated with lower expression of *MS4A4A* and *MS4A6A* in human monocytes or macrophages, while the BV2 experiment demonstrated that lower expression of *Spi1* (which is protective in humans) led to lower expression of *Ms4a4a* and *Ms4a6d* (mouse ortholog of *MS4A6A*), which are also associated with reduced AD risk in humans. Several large-scale transcriptomic and proteomic analyses of acutely-isolated microglial cells in animal models of aging or neurological disorders have suggested the existence of a homeostatic signature that is perturbed during aging and under pathological conditions[62–64]. It will be valuable to analyze whole-transcriptome changes in microglial cells with differential *SPI1* expression in comparison with existing datasets to test whether changes in *SPI1* levels prime microglia to exacerbate or alleviate transcriptional changes that occur during aging or disease development. Together with genetic variation in microglial specific genes associated with AD as an amplifier, *SPI1* may be a master regulator capable of distorting the cellular balance that either helps microglia to cope with and protect from the pathogenic assault or commits microglia to a neurotoxic phenotype.

PU.1 expression levels regulate several other myeloid/microglial cell functions[57,65], including proliferation, survival and differentiation, that could also modulate AD risk. Indeed, expression of *Il34* and *Csf1*, soluble factors that bind to *Csf1r* and promote differentiation of monocytes to microglia-like cells *in vitro* and are required for microglial development and maintenance *in vivo*[66,67], was elevated after knock-down of *Spi1*, while expression of *Csf1r* was reduced (**Fig. 3a, 3c**). Interestingly, inhibition of *Csf1r* in a 3xTg-AD mouse model led to a reduction in the number of microglia associated with β-amyloid plaques and improved cognition[68]. These findings suggest that it will be important to analyze cell proliferation, survival, differentiation, and migration phenotypes in microglia with differential *Spi1* expression, and in infiltrating monocytes and macrophages, because *Ccl2* and *Cxcl2* (MCP1 and MIP2α proteins) expression was directly dependent on *Spi1* levels (**Fig.3a**). Both molecules participate in recruitment of circulating monocytes and neutrophils to the brain[69,70] that can promote neuroinflammation and are detrimental in AD mouse models[71,72]. In addition, expression of a microgliosis marker *Aif1* (Iba1 protein) was dependent on *Spi1* (**Fig. 3a**), which in conjunction with changes in *Il1b*, *Nos2*, *Ptgs2*, *Arg1* and *Nlrp3* suggests that decreased *Spi1* expression may moderate the inflammatory response of microglial cells to improve disease outcomes. Interestingly, expression of *Cx3cr1* and *Axl* were markedly elevated upon knock-down of *Spi1* (**Fig. 3c**), raising the possibility that beneficial effects of changes in *Spi1* expression are exerted through modulation of synaptic or neuronal clearance[73,74]. Further experimental investigation of the proposed phenotypes will shed more light on the mechanisms of *SPI1* contribution to AD risk. Of note, overexpression and knock-down of *Spi1* in BV2 microglial cells produce different and often opposite changes in

expression of the genes profiled here, possibly driving different phenotypes that may underlie detrimental and protective functions of PU.1 in AD. Thus, exploration of PU.1 association with AD risk presents an intriguing opportunity for the discovery of novel disease mechanisms and therapeutic interventions.

In summary, by combining AD survival, CSF biomarker and myeloid *cis*-eQTL analyses, we replicated and discovered multiple genetic loci associated with AD. Specifically, we nominate *SPI1* as the candidate gene responsible for the association at the previously reported *CELF1* locus. *SPI1* encodes PU.1, a transcription factor expressed in microglia and other myeloid cells that directly regulates the transcription of other AD-associated genes expressed in these cell types. Our data suggest that lower *SPI1* expression reduces risk for AD, suggesting a novel therapeutic approach to the treatment of AD. Furthermore, we demonstrate that AD survival-associated SNPs within the MS4A gene cluster are also associated with eQTLs in myeloid cells for both *MS4A4A* and *MS4A6A*. Specifically, the allele associated with reduced AD risk is associated with lower *MS4A4A* and *MS4A6A* expression. This result is consistent with the observation that lowering *SPI1* expression, which is protective for AD risk, also lowers *MS4A4A* and *MS4A6A* expression and reduces phagocytic activity in BV2 microglial cells. These results reinforce the emerging genetic and epigenetic association between AD and a network of microglial expressed genes[2,5,19–23], highlighting the need to dissect their functional mechanisms.

Methods

Genome-wide survival association study datasets

The final meta-analysis dataset consists of samples from the Alzheimer's Disease Genetics Consortium (ADGC), Genetic and Environmental Risk in Alzheimer's Disease (GERAD), European Alzheimer's Disease Initiative (EADI), and Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE). The study cohorts consist of case-control and longitudinal cohorts. The study protocols for all cohorts were reviewed and approved by the appropriate institutional review boards. Details of ascertainment and diagnostic procedures for each dataset are as previously described[1–5] and included in the **Supplementary Information**.

CSF biomarker datasets

CSF samples were obtained from the Knight-ADRC (N=805), ADNI-1 (N=390), ADNI-2 (N=397), the Biomarkers for Older Controls at Risk for Dementia (BIOCARD) (N=184), Mayo Clinic (N=433), Lund University (Swedish) (N=293), University of Pennsylvania (Penn) (N=164), University of Washington (N=375), The Parkinson's Progression Markers Initiative (500) and Saarland University (German) (N=105). Details of ascertainment and diagnostic procedures for the dataset are included in the **Supplementary Information**.

Quality Control

For survival analysis, we excluded cases with AAO below 60 and cases with prevalent stroke. For CSF analysis, individuals under age 45 years were removed because prior studies have demonstrated that the relationship between CSF $A\beta_{42}$ levels and age appears to differ in individuals below 45 years vs. those above 45 years[75]. Of the remaining individuals in both analyses, we excluded individuals who had > 5% missing genotype rates, who showed a discrepancy between reported sex and sex estimated on the basis of genetic data, or who showed evidence of non-European ancestry based on principal component analysis using PLINK1.9[76]. We identified unanticipated duplicates and cryptic relatedness using pair-wise genome-wide estimates of proportion identity by descent (IBD) using PLINK. When duplicate samples or a pair of samples with cryptic relatedness was identified, the sample with the lower genotyping call rate was removed. We excluded potentially related individuals so that all remaining individuals have kinship coefficient below 0.05. Finally, we excluded individuals with missing disease status, age or gender information.

To control for genotype quality, we excluded SNPs with missing genotypes in > 5% of individuals in each dataset for survival analysis, and > 2% for CSF association analysis. For the EADI cohort, variants with minor allele frequency < 1%, Hardy-Weinberg P value $< 1 \times 10^{-6}$ and missingness > 2% were removed prior to imputation. Genome-wide genotype imputation was performed using IMPUTE2[77] with 1000 Genomes reference haplotypes. We excluded imputed SNPs with an IMPUTE2 quality score < 0.5 for survival analysis. For CSF association, we excluded SNPs with an IMPUTE2 quality score of < 0.3 since the dataset was only used for follow-up. In the ADGC, GERAD, CHARGE, and CSF datasets, we then removed SNPs that failed the Hardy-Weinberg equilibrium in controls calculated based on the imputed best-guess genotypes using a P value threshold of 1×10^{-6}. We excluded SNPs with minor allele frequency ≤ 0.02. Finally, we excluded SNPs with available statistics in only one consortium dataset in the meta-analysis. The number of filtered samples and SNPs in each of the above steps are recorded in **Supplementary Table 1**.

Genome-wide survival association study

We conducted a genome-wide Cox proportional hazards regression[78] assuming an additive effect from SNP dosage. The Cox proportional hazard regression was implemented in the R survival analysis package. We incorporated sex, site and the first three principal components from EIGENSTRAT[30] in all our regression models to control for their effects. For EADI, sex and four principal components were included in the model. For the Cox model, the time scale is defined as age in years, where age is age at onset for cases and age at last assessment for controls. The formula applied is as followed:

$$h(t|X) = h_0(t)\exp(\sum_{i=1}^{p} \beta_i X_i)$$

where X = (X1, X2, …, Xp) are the observed values of covariates for subject i. The Cox model has previously been shown to be applicable to case-control datasets without an elevated type 1 error rate nor overestimation in effect sizes[79,80]. The model assumes log-linearity and proportional hazards. The assumption of log-linearity is common in the additive logistic regression used in a typical GWAS. We validated the assumption of proportional hazards assumed by the Cox model by conducting the Schoenfeld test in the 22 prioritized SNPs. None of the SNPs has a Schoenfeld P value, which is the P value for Pearson product-moment correlation between the scaled Schoenfeld residuals and time, lower than 0.035 (multiple test correction threshold = 0.00227) in any of the 7 cohorts. Further, only 3 out of the 148 P values were less than 0.05, suggesting that the time proportionality assumption is unlikely to be violated in these associations (**Supplementary Table 1**). Similarly, the Schoenfeld test was conducted for all 22 SNP association models on the covariates in the ADGC and GERAD cohort (**Supplementary Table 1**). We also examined the effect sizes of our candidate SNPs in these cohorts and found consistent effect sizes (**Supplementary Fig. 3**) in the 3 retrospective case-control cohorts (ADGC, GERAD, EADI case-control) and 4 prospective cohorts (EADI-prospective, CHARGE FHS, CHS and Rotterdam).

After the analysis of each dataset, we carried out an inverse-variance meta-analysis on the results using METAL[26], applying a genomic control to adjust for inflation in each dataset. Of the 751 suggestive SNPs (P < 1*10-5), we found these SNPs to show lower standard errors and confidence intervals with the increasing number of cohorts showing consistent directionalities of effect. Particularly, the average standard error for SNPs showing 1 to 7 consistent directionalities ranges from 0.171, 0.109, 0.0744, 0.0346, 0.0234, 0.0173 to 0.01795 (**Supplementary Fig. 1b**). Thus, we limited our final analysis to SNPs that showed consistent directionalities of effect in at least 6 out of the 7 datasets included in the meta-analysis. The association graphs of results from loci of interest were plotted using LocusZoom[81].

CSF biomarker association analysis

For the CSF datasets, we performed multivariate linear regression for CSF $A\beta_{42}$ and tau, and $ptau_{181}$ association adjusting for age, gender, site, and the first three principal components using PLINK.

eQTL analysis

We examined the effect of top survival and CSF SNPs on gene expression using published databases. For general brain expression eQTL analysis, we queried the BRAINEAC eQTL data provided by the UK human Brain Expression Consortium (see URLs).

We conducted leukocyte-specific analysis using the Cardiogenics dataset[32] composed of 738 monocytes and 593 macrophages samples. For each probeset – imputed SNP pair, a simple linear regression was used to analyze the data separately for monocytes and macrophages:

$$y_i = \alpha + \beta x_i + \varepsilon_i, 1 \leq i \leq n, \varepsilon_i \sim N(0, \sigma^2)$$

where i is the subject index, x is the effective allele copy number, and y_i is the covariates-adjusted, inverse-normal transformed gene expression. Significance of *cis* (SNP within ±1Mb of the closest transcript end) eQTL effects were quantified with a Wald test on the ordinary Least Squares (OLS) estimator of the coefficient β, obtained with R. The distribution of the Wald test P values under the null hypothesis of no correlation between genotype and gene expression was estimated by rerunning the same analysis on a null dataset obtained by permuting the expression samples identifiers. For additional monocyte eQTL analysis, we queried statistics from Fairfax et al.[33] to validate findings in the Cardiogenics dataset.
For conditional analysis, we performed analysis for *SPI1* (probe: ILMN_1696463) against all SNPs within ±2Mb from the closest transcript end, by including the following SNPs effective allele copy numbers as covariates in the linear regression model, one at a time: rs1057233, rs10838698, rs7928163, rs10838699, rs10838725, rs1377416. Significance was again assessed with a two-sided Wald test on the OLS estimator of the coefficient β.

Gene expression analysis in human and mouse brain cell types
Cell-type specific gene expression in the human and mouse brain was queried from brain RNA-Seq databases described in Zhang et al.[34,35] and Bennett et al.[36] and plotted using custom R scripts (see URLs). The mouse astrocytes-FACS and astrocytes-immunopanned in mouse were collapsed into a single astrocyte cell type.

Epigenetic analysis in human myeloid cell types
We utilized HaploReg[40] to annotate the regulatory element of the significantly associated SNPs and their tagging SNPs. The myeloid chromatin marks/states and PU.1 ChIP-Seq data at genetic loci were further examined through the Washington University Epigenome browser[82] using the public Roadmap Epigenomics Consortium public tracks hub as well as custom track hubs for human monocytes and macrophages (hg19) (see URLs).

Colocalization (coloc and SMR/HEIDI) analyses
Colocalization analysis of genetic variants associated with AD and myeloid gene expression was performed using AD survival-associated (or IGAP GWAS) SNP and myeloid (monocyte and macrophage) eQTL datasets from Cardiogenics as inputs. Overlapping SNPs were retained within the hg19 region chr11:47100000-48100000 for the *SPI1/CELF1* locus, chr11:59500000-60500000 for the *MS4A* locus, and chr1:169300000-170300000 for the *SELL* locus.
Colocalization analysis of AD- and gene expression-associated SNPs was performed using the 'coloc.abf' function in the 'coloc' R package (v2.3-1). Default settings were used as prior probability of association: 1E-4 for trait 1 (gene expression), 1E-4 for trait 2 (AD) and 1E-5 for both traits. SMR/HEIDI (v0.65) analysis was performed as described in Zhu et al.[28] and the companion website (see URLs). The ADGC subset of the IGAP GWAS dataset was used to perform the LD calculations.

Partitioned heritability analysis using LD score regression
We used LDSC (LD SCore, v1.0.0) regression analysis[30] to estimate heritability of AD and schizophrenia from GWAS summary statistics (excluding the APOE [chr19:45000000-45800000] and MHC/HLA [chr6:28477797-33448354] regions) partitioned by PU.1 ChIP-Seq binding sites in myeloid cells, as described in the companion website (see URLs) and controlling for the 53 functional annotation categories of the full baseline model. GWAS summary statistics for AD and schizophrenia (SCZ) were downloaded from the IGAP consortium[1] (stage 1 dataset) and the Psychiatric Genomics Consortium (PGC)[31] (pgc.cross.scz dataset), respectively (see URLs). SPI1 (PU.1) bindings sites were downloaded as filtered and merged ChIP-Seq peaks in BED format from the ReMap database[83] (GEO:GSE31621, SPI1, blood monocyte and macrophage datasets[41]). SPI1 (PU.1) and POLR2AphosphoS5 binding sites were downloaded as broad ChIP-Seq peaks in BED format from the Encode portal[8447] (DCC:ENCSR037HRJ; GEO:GSE30567; HL60 dataset) (see URLs).

Phagocytosis assay
BV2 mouse microglial cell line was kindly provided by Marc Diamond (UT Southwestern Medical Center). BV2 cells were cultured in DMEM (Gibco 11965) supplemented with 5% FBS (Sigma F4135) and 100 U/ml penicillin-streptomycin (Gibco 15140). Routine testing of cell lines using MycoAlert PLUS mycoplasma detection kit (Lonza) showed that BV2 cells were negative for mycoplasma contamination. pcDNA3-FLAG-PU.1 was a gift from Christopher Vakoc[85] (Addgene plasmid 66974). pGFP-V-RS with either non-targeting shRNA or PU.1-targeting shRNAs was purchased from OriGene Technologies (TG502008). The pHrodo red zymosan conjugate bioparticles from Thermo Fisher (P35364) were used to assess phagocytic activity. For transient transfections, 200,000 cells were seeded in a 24-well plate. On the next day, cells were washed with PBS (Gibco 14190) and medium was changed to 400 µl DMEM supplemented with 2% FBS without antibiotic. Transfection mixes of 0.5 µg pcDNA3 or 0.5 µg pcDNA3-FLAG-PU.1 with 0.5 µg pCMV-GFP for overexpression of mouse PU.1 and 1µg pGFP-V-RS-shSCR, -shA, -shB and -shD for knock-down of mouse PU.1 were prepared with 2 µl of Lipofectamine 2000, incubated for 20 min at room temperature and added to each well. After 8 hours of incubation 1 ml of growth medium was added to each well and plates were incubated for 2 days. Then the medium was replaced with 500 µl of fresh medium, and 25 µg of bioparticles were added to cells for 3 hour incubation. Bioparticles uptake was verified with a fluorescent microscope; then the cells were collected with trypsin (Gibco #25200), washed with PBS once and re-suspended in 500 µl PBS with 1% BSA. Cells were kept on ice and phagocytic activity was analyzed on an LSR II flow cytometer (BD Biosciences). At least 30,000 events were collected in each experiment, gated on FSC-A/SSC-A and further on FSC-A/FSC-W dot plot to analyze populations of viable single cells. Data were quantified using FCS Express 5 (De Novo Software) and GraphPad Prism 7 (GraphPad Software). Cells pretreated with 2 µM Cytochalasin D for 30 minutes before and during the uptake of bioparticles were used as a negative control. The population of GFP^+/$pHrodo^+$ cells in each condition was used to quantify the phagocytic index: percentage of $pHrodo^+$ cells in GFP^+ gated population x geometric mean pHrodo intensity / 10^6; and represented as phagocytic activity. Three independent experiments were performed with two technical replicates without randomization of sample processing, n = 3. Researcher was not blinded to the samples identification. Differences between the means of preselected groups were analyzed with one-way ANOVA and Sidak's post hoc multiple comparisons test between selected groups, with a single pooled variance. Values of Cytochalasin D-treated cells were

excluded from the statistical analysis. Adjusted P values for each comparison are reported, non-significant differences are not reported.

Western blotting

BV2 cells transiently transfected as described for the phagocytosis assay were collected with trypsin after 48 hours of incubation, washed with PBS and re-suspended in PBS with 1% BSA. Cells from the same treatment were pooled and sorted on FACSARIA III (BD Biosciences) into GFP^+ and GFP^- populations, pelleted at 2,000 rpm and lysed in RIPA buffer (50 mM Tris-HCl pH 7.4, 150 mM NaCl, 1% NP-40, 0.5% sodium deoxycholate, 0.1% SDS and Complete protease inhibitor tablets (Roche)) with one freeze-thaw cycle and 1 hour incubation on ice. Protein concentration was quantified using the BCA kit (Thermo Fisher #23225). Equal amounts of protein were separated by electrophoresis in Bolt 4 – 12% Bis-Tris Plus gels with MOPS SDS running buffer and transferred using the iBlot 2 nitrocellulose transfer stack. Membranes were blocked and probed with antibodies against PU.1 (Cell Signaling #2266) and β-Actin (Sigma #A5441) in 3% non-fat dry milk in TBS / 0.1% Tween-20 buffer. Secondary antibody staining was visualized using WesternBright ECL HRP Substrate Kit (Advansta K-12045) and ChemiDoc XRS+ (BioRad). Images were quantified using ImageJ (NIH) and GraphPad Prism 7 (GraphPad Software). Two independent experiments were performed without randomization of sample processing, n = 2. Researcher was not blinded to the samples identification. Differences between every group mean were analyzed with one-way ANOVA and Sidak's post hoc multiple variance test between selected groups, with a single pooled variance. Adjusted P values for each comparison are reported.

Quantitative PCR

Sorted GFP^+ BV2 cells after overexpression or knock-down of PU.1 were collected as described for western blotting. Cell pellets were lysed in QIAzol reagent and RNA was isolated with RNAeasy Mini kit according to the manufacturer's instructions (Qiagen) including the Dnase treatment step with RNase-free DNase set (Qiagen). Quantities of RNA were measured using Nanodrop 8000 (Thermo Scientific) and reverse transcription was performed with 1-2 μg of total RNA using High-Capacity RNA-to-cDNA kit (Thermo Fisher Scientific). qPCR was performed on QuantStudio 7 Flex Real-Time PCR System (Thermo Fisher Scientific) using Power SYBR Green Master Mix (Applied Biosystems) with one-step PCR protocol. 3 ng of cDNA was used for all genes except *Ms4a4a* when 24 ng of cDNA was used in a 10 μl reaction volume. Primers were from PrimerBank[86] or designed using Primer-BLAST program (NCBI) and are listed in **Supplementary Table 14**. Ct values were averaged from two technical replicates for each gene. Geometric mean of average Ct for the housekeeping genes *GAPDH*, *B2M* and *ACTB* was used as a reference that was subtracted from the average Ct for a gene of interest (dCt). Gene expression levels were log transformed (2^{-dCt}) and related to the combined mean values of pcDNA3 and pGFP-V-RS-shSCR control samples in each sort giving relative expression for each gene of interest. Data were visualized in GraphPad Prism 7 (GraphPad Software). Four independent experiments were performed without randomization of sample processing, n = 4. Researcher was not blinded to the sample identity. Differences between means were analyzed using one-way ANOVA and Dunnett's post hoc multiple comparisons test between experimental and control groups, with a single pooled variance. Adjusted P values for each comparison are reported in **Supplementary Table 13**.

Data availability
Summary statistics for the genome-wide survival analyses are posted on the NIA Genetics of Alzheimer's Disease Data Storage (NIAGADS, see URLs).

Code availability
Codes for analyses are available upon request.

URLs
BRAINEAC, http://caprica.genetics.kcl.ac.uk/BRAINEAC; LDSC software, http://www.github.com/bulik/ldsc; baseline and cell type group annotations, http://data.broadinstitute.org/alkesgroup/LDSCORE/; stratified LD score regression companion website, https://github.com/bulik/ldsc/wiki/Partitioned-Heritability; SMR/HEIDI software and companion website, http://cnsgenomics.com/software/smr; Brain RNA-Seq, http://web.stanford.edu/group/barres_lab/brainseq2/brainseq2.html; WashU EpiGenome Browser, http://epigenomegateway.wustl.edu/browser; custom tracks for human monocytes and macrophages, http://www.ag-rehli.de/TrackHubs/hub_MOMAC.txt; International Genomics of Alzheimer's Project (IGAP) http://web.pasteur-lille.fr/en/recherche/u744/igap/igap_download.php; Psychiatric Genomics Consortium (PGC) https://www.med.unc.edu/pgc/results-and-downloads; ReMap database http://tagc.univ-mrs.fr/remap; Encode portal https://www.encodeproject.org/; NIAGADS, https://www.niagads.org.

Acknowledgements
We would like to thank the patients, control subjects, and their family members for participating in or supporting the research projects included in this study.We thank Marc Diamond (UT Southwestern Medical Center) for the BV2 cell line and Flow Cytometry CORE at the Icahn School of Medicine at Mount Sinai Hospital.
IGAP
GERAD was supported by the Wellcome Trust, the MRC, Alzheimer’s Research UK (ARUK) and the Welsh government. ADGC and CHARGE were supported by the US National Institutes of Health, National Institute on Aging (NIH-NIA), including grants U01 AG032984 and R01 AG033193. CHARGE was also supported by Erasmus Medical Center and Erasmus University.
ADNI
Data collection and sharing for this project was funded by the Alzheimer's Disease Neuroimaging Initiative (ADNI) (National Institutes of Health Grant U01 AG024904) and DOD ADNI (Department of Defense award number W81XWH-12-2-0012). ADNI is funded by the National Institute on Aging, the National Institute of Biomedical Imaging and Bioengineering, and through generous contributions from the following: AbbVie, Alzheimer’s Association; Alzheimer’s Drug Discovery Foundation; Araclon Biotech; BioClinica, Inc.; Biogen; Bristol-Myers Squibb Company; CereSpir, Inc.; Eisai Inc.; Elan Pharmaceuticals, Inc.; Eli Lilly and Company; EuroImmun; F. Hoffmann-La Roche Ltd and its affiliated company Genentech, Inc.; Fujirebio; GE Healthcare; IXICO Ltd.; Janssen Alzheimer Immunotherapy Research & Development, LLC.; Johnson & Johnson Pharmaceutical Research & Development LLC.; Lumosity; Lundbeck; Merck & Co., Inc.; Meso Scale Diagnostics, LLC.; NeuroRx Research; Neurotrack Technologies; Novartis Pharmaceuticals Corporation; Pfizer Inc.; Piramal Imaging; Servier; Takeda Pharmaceutical Company; and Transition Therapeutics. The Canadian Institutes

of Health Research is providing funds to support ADNI clinical sites in Canada. Private sector contributions are facilitated by the Foundation for the National Institutes of Health (www.fnih.org). The grantee organization is the Northern California Institute for Research and Education, and the study is coordinated by the Alzheimer's Disease Cooperative Study at the University of California, San Diego. ADNI data are disseminated by the Laboratory for Neuro Imaging at the University of Southern California

We thank the Cardiogenics (European Project reference LSHM-CT-2006-037593) project for providing summary statistics for the *cis*-eQTL-based analyses. We also thank the ENCODE Consortium and Richard Myers' lab (HAIB) for providing ChIP-Seq datasets.

This work was supported by grants from the National Institutes of Health (U01 AG049508 (AMG), R01-AG044546 (CC), RF1AG053303 (CC) and R01-AG035083 (AMG), RF-AG054011 (AMG)), the JPB Foundation (AMG) and FBRI (AMG). The recruitment and clinical characterization of research participants at Washington University were supported by NIH P50 AG05681, P01 AG03991, and P01 AG026276. Kuan-lin Huang received fellowship funding in part from the Ministry of Education in Taiwan and the Lucille P. Markey Special Emphasis Pathway in Human Pathobiology. Ke Hao is partially supported by the National Natural Science Foundation of China (Grant No. 21477087, 91643201) and by the Ministry of Science and Technology of China (Grant No. 2016YFC0206507). This work was supported by access to equipment made possible by the Hope Center for Neurological Disorders and the Departments of Neurology and Psychiatry at Washington University School of Medicine.

Competing Financial Interests

I.B. is an employee of Regeneron Pharmaceuticals, Inc. A.M.G. is on the scientific advisory board for Denali Therapeutics and has served as a consultant for AbbVie and Cognition Therapeutics.

Author Contributions

A.M.G., E.M., and K.H. conceived and designed the experiments. K.H., S.C.J., O.H., A.D., M.K., J.C., J.C.L., V.C., C.B., B.G., Y.D., A.M., T.R., A.R., J.L.D., M.V.F, L.I., B.Z., I.B., C.C. and E.M. performed data analysis. A.A.P. performed phagocytosis assays, western blotting and qPCR validation. S.B., B.P.F., J.B., R.S., V.E.P., R.M., J.L.H., L.A.F., M.A.P., S.S., J.W., P.A., G.D.S., J.S.K.K., K.H., and C.C. provided and processed the data. A.M.G. supervised data analysis and functional experiments. K.H., A.A.P., E.M., and A.M.G. wrote and edited the manuscript. All authors read and approved the manuscript.

Tables

Table 1. Genome-wide survival analysis of Alzheimer's Disease. (a) Description of Consortia samples with available phenotype and genotype data included in the genome-wide survival analysis. AAO: age at onset. AAE: age at last examination. (b) Summary of loci with significant ($P < 5x10^{-8}$) or suggestive ($P < 1x10^{-5}$) associations from the genome-wide survival analysis.

a

	Cases			Controls		
Dataset	N	Percent women	Mean AAO yrs (s.d.)	N	Percent women	Mean AAE yrs (s.d.)
ADGC	8617	58.9	74.2 (8.1)	9765	60.1	77.1 (8.4)
GERAD	2615	63.4	73.0 (8.5)	1148	62.1	76.5 (7.0)
EADI case-control study	1420	67.2	72.1 (7.1)	878	61	72.2 (7.8)
EADI longitudinal study	387	61.8	81.3 (5.6)	5416	61.1	79.3 (5.3)
CHARGE FHS	229	65.5	85.7 (6.3)	1979	54.1	80.7 (7.5)
CHARGE CHS	374	69.2	82.2 (5.0)	1675	60.6	81.1 (5.2)
CHARGE Rotterdam	764	73.2	83.1 (6.6)	4988	57.8	81.4 (6.9)
Total	14406	61.7	74.8	25849	59.6	79.0

b

SNP	Major/ minor Alleles	MAF	CHR[a]	BP	Closest Gene	Logistic OR[b]	Logistic P value	Survival HR (95% CI)[c]	Survival P value	Hetero-geneity P value
Previously reported associated loci										
rs2093761	G/A	0.2019	1	207786542	*CR1*	1.16 (1.12-1.20)	$2.6x10^{-14}$	1.07 (1.04-1.10)	$1.2x10^{-6}$	0.25
rs6431219	C/T	0.4163	2	127862133	*BIN1*	1.12 (1.09-1.15)	$7.6x10^{-13}$	1.08 (1.06-1.10)	$3.9x10^{-10}$	0.16
rs1057233	A/G	0.3194	11	47376448	*SPI1/CELF1*[d]	0.93 (0.89-0.96)	$5.4x10^{-6}$	0.94 (0.91-.97)	$8.4x10^{-6}$	0.86
rs7930318	T/C	0.4004	11	60033371	*MS4A*	0.90 (0.87-0.93)	$5.1x10^{-11}$	0.93 (0.90-.95)	$2.3x10^{-9}$	0.6
rs567075	C/T	0.3097	11	85830157	*PICALM*	0.88 (0.85-0.91)	$4.3x10^{-14}$	0.91 (0.89-.94)	$9.1x10^{-12}$	0.74
rs9665907	G/A	0.1133	11	121435470	*SORL1*	0.88 (0.83-0.93)	$1.8x10^{-7}$	0.92 (0.88-.95)	$5.5x10^{-6}$	0.96
rs17125944	T/C	0.0924	14	53400629	*FERMT2*	1.13 (1.08-1.18)	$1.0x10^{-5}$	1.10 (1.06-1.14)	$2.3x10^{-6}$	0.31
rs4803758	G/T	0.3551	19	45327423	*APOE*[e]	1.33 (1.30-1.37)	$1.2x10^{-67}$	1.21 (1.18-1.23)	$7.8x10^{-52}$	0.32
Novel loci reaching suggestive significance										
rs10919252	C/G	0.3275	1	169802956	*C1orf112*	1.04 (1.01-1.08)	$1.1x10^{-2}$	1.10 (1.06-1.14)	$8.2x10^{-7}$	0.92
rs1532244	A/G	0.0925	3	28057905	*CMC1*	0.95 (0.90-1.01)	$6.9x10^{-2}$	0.86 (0.80-.93)	$9.7x10^{-6}$	0.99
rs116341973	A/G	0.0227	3	63462893	*SYNPR*	1.20 (1.09-1.30)	$5.4x10^{-4}$	1.23 (1.15-1.31)	$2.5x10^{-7}$	0.62
rs71602496	A/G	0.1453	4	661002	*PDE6B*	1.02 (0.98-1.06)	$3.6x10^{-1}$	1.08 (1.05-1.11)	$5.0x10^{-6}$	0.11
rs1689013	T/C	0.2493	4	181048651	*LINC00290*	1.02 (0.98-1.06)	$2.7x10^{-1}$	1.07 (1.04-1.09)	$4.7x10^{-6}$	0.31
rs7445192	A/G	0.461	5	140138701	*PCDHA1*	NA	NA	1.06 (1.03-1.08)	$7.9x10^{-6}$	0.77
rs12207208	T/C	0.1034	6	40301379	*LINC00951*	1.07 (1.02-1.20)	$1.2x10^{-2}$	1.09 (1.05-1.13)	$6.8x10^{-6}$	0.78
rs17170228	G/A	0.0623	7	33076314	*NT5C3A*	1.07 (1.01-1.14)	$2.5x10^{-2}$	1.13 (1.08-1.18)	$1.0x10^{-6}$	0.94
rs2725066	A/T	0.4872	8	4438058	*CSMD1*	1.03 (1.00-1.06)	$7.3x10^{-2}$	1.10 (1.06-1.14)	$1.0x10^{-6}$	0.6
rs7867518	T/C	0.476	9	2527525	*VLDLR*	0.97 (0.94-1.00)	$6.8x10^{-2}$	0.95 (0.92-.97)	$9.1x10^{-6}$	0.79

rs1625716	T/G	0.0643	10	59960083	*IPMK*	0.87 (0.80-0.94)	$1.0x10^{-4}$	0.88 (0.82-.94)	$7.7x10^{-6}$	0.95
rs1118069	T/A	0.2805	12	84739181	*SLC6A15*	0.98 (0.94-1.01)	$2.0x10^{-1}$	0.90 (0.86-.95)	$2.7x10^{-6}$	0.8
rs11074412	A/G	0.2087	16	19833001	*IQCK*	0.94 (0.90-0.98)	$1.9x10^{-3}$	0.93 (0.90-.96)	$7.0x10^{-6}$	0.48
rs5750677	C/T	0.2885	22	39147715	*SUN2*	0.97 (0.93-1.00)	$5.1x10^{-2}$	0.94 (0.91-.97)	$5.2x10^{-6}$	0.51

aBuild 37, assembly hg19. [b]Summary statistics of the logistic regression result was obtained from stage 1 of the 2013 IGAP landmark GWAS paper[1]. [c]Calculated with respect to the minor allele. [d]*SPI1* is the nearest gene to rs1057233. The same locus is previously assigned as *CELF1* in the 2013 IGAP GWAS. [e]The nearest gene to rs4803758 is *APOE*.

Table 2. Summary of CSF biomarker-associations of suggestive and significant AD survival-associated SNPs. Associations reaching the significance threshold after Bonferroni correction for multiple testing ($P < 2.27x10^{-3}$) are bolded.

SNP	CHR	Closest gene	$Beta_{tau}$	P_{tau}	$Beta_{ptau}$	P_{ptau}	$Beta_{ab42}$	P_{ab42}
Previously reported associated loci								
rs2093761	1	*CR1*	-	>0.05	$1.46x10^{-2}$	$2.87x10^{-2}$	-	>0.05
rs6431219	2	*BIN1*	-	>0.05	-	>0.05	-	>0.05
rs1057233	11	*CELF1*	$-1.11x10^{-2}$	$6.55x10^{-2}$	$-1.25x10^{-2}$	$2.76x10^{-2}$	**$1.45x10^{-2}$**	**$8.24x10^{-4}$**
rs7930318	11	*MS4A*	$-1.24x10^{-2}$	$3.27x10^{-2}$	-	>0.05	-	>0.05
rs567075	11	*PICALM*	$-1.32x10^{-2}$	$3.22x10^{-2}$	$-1.24x10^{-2}$	$3.13x10^{-2}$	$9.10x10^{-3}$	$3.88x10^{-2}$
rs9665907	11	*SORL1*	$-1.74x10^{-2}$	$4.28x10^{-2}$	$-1.94x10^{-2}$	$1.57x10^{-2}$	-	>0.05
rs17125944	14	*FERMT2*	$2.50x10^{-2}$	$8.71x10^{-3}$	$2.09x10^{-2}$	$2.09x10^{-2}$	$-1.79x10^{-2}$	$8.90x10^{-3}$
rs4803758	19	*APOE*	$1.61x10^{-2}$	$7.42x10^{-3}$	**$2.01x10^{-2}$**	**$3.75x10^{-4}$**	**$-1.79x10^{-2}$**	**$3.12x10^{-5}$**
Novel candidate loci								
rs10919252	1	*C1orf112*	-	>0.05	-	>0.05	-	>0.05
rs1532244	3	*CMC1*	-	>0.05	$2.41x10^{-2}$	$1.23x10^{-2}$	-	>0.05
rs116341973	3	*SYNPR*	-	>0.05	-	>0.05	-	>0.05
rs71602496	4	*PDE6B*	-	>0.05	-	>0.05	-	>0.05
rs1689013	4	*LINC00290*	-	>0.05	-	>0.05	-	>0.05
rs7445192	5	*PCDHA1*	-	>0.05	$1.38x10^{-2}$	$9.98x10^{-3}$	-	>0.05
rs12207208	6	*LINC00951*	-	>0.05	-	>0.05	-	>0.05
rs17170228	7	*NT5C3A*	-	>0.05	-	>0.05	-	>0.05
rs2725066	8	*CSMD1*	$1.20x10^{-2}$	$4.53x10^{-2}$	-	>0.05	-	>0.05
rs7867518	9	*VLDLR*	$-1.58x10^{-2}$	$5.83x10^{-3}$	-	>0.05	-	>0.05
rs1625716	10	*IPMK*	-	>0.05	-	>0.05	-	>0.05
rs1118069	12	*SLC6A15*	-	>0.05	-	>0.05	$-1.07x10^{-2}$	$1.56x10^{-2}$
rs11074412	16	*IQCK*	-	>0.05	-	>0.05	-	>0.05
rs5750677	22	*SUN2*	$1.30x10^{-2}$	$3.55x10^{-2}$	-	>0.05	-	>0.05

Table 3. Significant *cis*-eQTL associations of the 22 suggestive and significant AD survival-associated SNPs. Significance threshold is determined to be 2.52×10^{-6} based on Bonferroni correction for multiple testing. The minor alleles are considered as the effective allele.

				Monocyte		Macrophage	
SNPID	CHR	Probe_Id	Gene	P value	Beta	P value	Beta
rs10919252	1	ILMN_1724422	*SELL*	7.33×10^{-35}	-0.65	-	-
rs71602496	4	ILMN_1769751	*PIGG*	5.19×10^{-10}	-0.46	9.11×10^{-13}	-0.58
rs1625716	10	ILMN_2122953	*CISD1*	5.98×10^{-23}	-1.09	7.82×10^{-8}	-0.67
rs1057233	11	ILMN_1696463	*SPI1*	1.50×10^{-105}	-1.11	6.41×10^{-87}	-1.11
rs1057233	11	ILMN_1781184	*MYBPC3*	4.99×10^{-51}	-0.83	5.58×10^{-23}	-0.62
rs1057233	11	ILMN_1686516	*CELF1*	3.95×10^{-8}	0.32	-	-
rs1057233	11	ILMN_2382083	*CELF1*	1.13×10^{-7}	0.31	1.31×10^{-4}	0.25
rs1057233	11	ILMN_1652989	*NUP160*	1.42×10^{-5}	-0.26	5.35×10^{-22}	-0.62
rs7930318	11	ILMN_2370336	*MS4A4A*	8.20×10^{-28}	-0.56	-	-
rs7930318	11	ILMN_1721035	*MS4A6A*	4.90×10^{-23}	-0.52	1.25×10^{-9}	-0.35
rs7930318	11	ILMN_1741712	*MS4A4A*	1.48×10^{-11}	-0.36	1.54×10^{-4}	-0.22
rs7930318	11	ILMN_2359800	*MS4A6A*	1.94×10^{-10}	-0.34	3.77×10^{-9}	-0.34
rs11074412	16	ILMN_1783712	*LOC400506*	6.49×10^{-17}	0.54	-	-
rs11074412	16	ILMN_2081883	*IQCK*	-	-	1.22×10^{-12}	-0.52
rs4803758	19	ILMN_2337336	*PVRL2*	1.52×10^{-8}	0.30	-	-
rs5750677	22	ILMN_2099301	*SUN2*	3.66×10^{-58}	-0.90	3.15×10^{-36}	-0.80
rs5750677	22	ILMN_1730879	*CBY1*	1.80×10^{-9}	-0.37	-	-

Figure Legends

Figure 1. Genetic and eQTL fine-mapping of AD associations and *SPI1* (PU.1) expression and ChIP-Seq analysis. (**a**) The AD-survival association landscape at the *CELF1/SPI1* locus resembles that of *SPI1* eQTL association in monocytes and macrophages. (**b**) The AD-survival association landscape resembles that of *MS4A4A/MS4A6A* eQTL association in monocytes and macrophages. (**c**) $Rs1057233^{G}$ is associated with reduced *SPI1* expression in a dosage-dependent manner. (**d**) The mouse homolog of *SPI1*, *Sfpi1* or *Spi1*, is selectively expressed in microglia and macrophages in mouse brains based on the brain RNA-Seq database[34–36]. OPCs contain 5% microglial contamination. (**e**) *SPI1* (PU.1) binds to the promoter and regulatory regions of *CD33*, *MS4A4A*, *MS4A6A*, *TREM2*, and *TREML2* in human CD14+ monocytes based on ChIP-Seq data[41].

Figure 2. PU.1 is involved in the phagocytic activity of BV2 microglial cells. (**a**) Phagocytosis of zymosan labeled with red pHrodo fluorescent dye in BV2 cells with transient overexpression and knock-down of PU.1 was measured by flow cytometry. Cytochalasin D treatment was used as a negative control. Mean phagocytic index ± SD is shown: pcDNA 0.7373 ± 0.1772, pcDNA + 1 μM Cyt 0.0236 ± 0.0242, FLAG-PU.1 1.2630 ± 0.2503, shSCR 1.014 ± 0.3656, shA 0.4854 ± 0.1209, shB 0.2579 ± 0.06967, shD 0.2002 ± 0.05168. $F(6,13) = 14.82$, pcDNA vs pcDNA + 1 μM Cyt P=0.0078, pcDNA vs FLAG-PU.1 P=0.0295, shSCR vs shA P=0.0283, shSCR vs shB P=0.0020, shSCR vs shD P=0.0010, n = 3. (**b**) BV2 cells were transiently transfected with pcDNA3 (pcDNA) or pcDNA3-FLAG-PU.1 (FLAG-PU.1) and pCMV-GFP as described for phagocytosis assay. Note a shift in mobility of the band for exogenous FLAG-PU.1 in overexpression condition compared to endogenous PU.1 in control. (**c**) BV2 cells were transiently transfected with shRNA targeting PU.1 (shA, shB and shD) or non-targeting control (shSCR) in pGFP-V-RS vector. GFP^{+} cells were sorted with flow cytometer and analyzed for levels of PU.1 in western blotting in two independent experiments (**b**, **c**). (**d**) Quantification of PU.1 levels in **c** normalized to β-Actin as a loading control. Values are presented as mean ± SD: shSCR 100 ± 2.10, shA 50.34 ± 9.52, shB 16.03 ± 14.72, shD 12.13 ± 10.03. $F(3,6) = 70.55$, shSCR vs shA P=0.0014, shSCR vs shB $P < 0.0001$, shSCR vs shD $P < 0.0001$, n = 2. * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$, one-way ANOVA with Sidak's post hoc multiple comparisons test between selected groups.

Figure 3. Genes regulated in BV2 microglial cells with differential expression of *Spi1*. qPCR analysis in transiently transfected and sorted GFP^{+} BV2 cells with overexpression (FLAG-PU.1) and knock-down (shB) of *Spi1*. Changes in expression levels are grouped for genes with altered levels after overexpression and knock-down of *Spi1* in (**a**) and genes with variable expression in BV2 cells either with overexpression (**b**) or knock-down (**c**) of *Spi1*. Values are presented as mean ± SD, n = 4 samples collected independently. * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$, one-way ANOVA with Dunnett's post hoc multiple comparisons test between experimental and control groups, detailed statistical analysis is reported in **Supplementary Table 11**.

References

1. Lambert, J.-C. *et al.* Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease. *Nat. Genet.* **45,** 1452–8 (2013).
2. Naj, A. C. *et al.* Common variants at MS4A4/MS4A6E, CD2AP, CD33 and EPHA1 are associated with late-onset Alzheimer's disease. *Nat Genet* **43,** 436–441 (2011).
3. Harold, D. *et al.* Genome-wide association study identifies variants at CLU and PICALM associated with Alzheimer's disease. *Nat Genet* **41,** 1088–1093 (2009).
4. Seshadri, S. *et al.* Genome-wide analysis of genetic loci associated with Alzheimer disease. *JAMA* **303,** 1832–40 (2010).
5. Hollingworth, P. *et al.* Common variants at ABCA7, MS4A6A/MS4A4E, EPHA1, CD33 and CD2AP are associated with Alzheimer's disease. *Nat Genet* **43,** 429–435 (2011).
6. Naj, A. C. *et al.* Effects of Multiple Genetic Loci on Age at Onset in Late-Onset Alzheimer Disease: A Genome-Wide Association Study. *JAMA Neurol* (2014). doi:10.1001/jamaneurol.2014.1491
7. Kamboh, M. I. *et al.* Genome-wide association analysis of age-at-onset in Alzheimer's disease. *Mol. Psychiatry* **17,** 1340–6 (2012).
8. Bennett, C. *et al.* Evidence that the APOE locus influences rate of disease progression in late onset familial Alzheimer's Disease but is not causative. *Am J Med Genet* **60,** 1–6 (1995).
9. Slooter, A. J. *et al.* Risk estimates of dementia by apolipoprotein E genotypes from a population-based incidence study: the Rotterdam Study. *Arch Neurol* **55,** 964–968 (1998).
10. Daw, E. W. *et al.* The number of trait loci in late-onset Alzheimer disease. *Am J Hum Genet* **66,** 196–204 (2000).
11. Tunstall, N. *et al.* Familial influence on variation in age of onset and behavioural phenotype in Alzheimer's disease. *Br J Psychiatry* **176,** 156–159 (2000).
12. Thambisetty, M., An, Y. & Tanaka, T. Alzheimer's disease risk genes and the age-at-onset phenotype. *Neurobiol Aging* **34,** 2696 e1–5 (2013).
13. Jones, E. L. *et al.* Evidence that PICALM affects age at onset of Alzheimer's dementia in Down syndrome. *Neurobiol Aging* **34,** 2441 e1–5 (2013).
14. Cruchaga, C. *et al.* GWAS of cerebrospinal fluid tau levels identifies risk variants for alzheimer's disease. *Neuron* **78,** 256–268 (2013).
15. Kauwe, J. S. K. *et al.* Alzheimer's disease risk variants show association with cerebrospinal fluid amyloid beta. *Neurogenetics* **10,** 13–17 (2009).
16. Gusev, A. *et al.* Integrative approaches for large-scale transcriptome-wide association studies. *Nat. Genet.* **48,** 245–52 (2016).
17. Jonsson, T. *et al.* Variant of TREM2 associated with the risk of Alzheimer's disease. *N Engl J Med* **368,** 107–116 (2013).
18. Guerreiro, R. *et al.* TREM2 variants in Alzheimer's disease. *N Engl J Med* **368,** 117–127 (2013).
19. Bradshaw, E. M. *et al.* CD33 Alzheimer's disease locus: altered monocyte function and amyloid biology. *Nat. Neurosci.* **16,** 848–50 (2013).
20. Zhang, B. *et al.* Integrated systems approach identifies genetic nodes and networks in late-onset Alzheimer's disease. *Cell* **153,** 707–20 (2013).
21. Gjoneska, E. *et al.* Conserved epigenomic signals in mice and humans reveal immune basis of Alzheimer's disease. *Nature* **518,** 365–369 (2015).
22. Chan, G. *et al.* CD33 modulates TREM2: convergence of Alzheimer loci. *Nat. Neurosci.*

(2015). doi:10.1038/nn.4126
23. Raj, T. *et al.* Polarization of the effects of autoimmune and neurodegenerative risk alleles in leukocytes. *Science* **344,** 519–23 (2014).
24. Corder, E. H. *et al.* Gene dose of apolipoprotein E type 4 allele and the risk of Alzheimer's disease in late onset families. *Science (80-.).* **261,** 921–923 (1993).
25. Khachaturian, A. S., Corcoran, C. D., Mayer, L. S., Zandi, P. P. & Breitner, J. C. S. Apolipoprotein E epsilon4 count affects age at onset of Alzheimer disease, but not lifetime susceptibility: The Cache County Study. *Arch. Gen. Psychiatry* **61,** 518–24 (2004).
26. Kamboh, M. I. *et al.* Genome-wide association analysis of age-at-onset in Alzheimer's disease. *Mol Psychiatry* **17,** 1340–1346 (2012).
27. Blacker, D. *et al.* ApoE-4 and age at onset of Alzheimer's disease: the NIMH genetics initiative. *Neurology* **48,** 139–47 (1997).
28. Zhu, Z. *et al.* Integration of summary data from GWAS and eQTL studies predicts complex trait gene targets. *Nat. Genet.* **48,** 481–7 (2016).
29. GTEx Consortium, T. Gte. The Genotype-Tissue Expression (GTEx) project. *Nat. Genet.* **45,** 580–5 (2013).
30. Finucane, H. K. *et al.* Partitioning heritability by functional annotation using genome-wide association summary statistics. *Nat. Genet.* **47,** 1228–1235 (2015).
31. Cross-Disorder Group of the Psychiatric Genomics Consortium. Identification of risk loci with shared effects on five major psychiatric disorders: a genome-wide analysis. *Lancet* **381,** 1371–9 (2013).
32. Garnier, S. *et al.* Genome-Wide Haplotype Analysis of Cis Expression Quantitative Trait Loci in Monocytes. *PLoS Genet.* **9,** (2013).
33. Fairfax, B. P. *et al.* Innate immune activity conditions the effect of regulatory variants upon monocyte gene expression. *Science* **343,** 1246949 (2014).
34. Zhang, Y. *et al.* Purification and Characterization of Progenitor and Mature Human Astrocytes Reveals Transcriptional and Functional Differences with Mouse. *Neuron* **89,** 37–53 (2016).
35. Zhang, Y. *et al.* An RNA-Sequencing Transcriptome and Splicing Database of Glia, Neurons, and Vascular Cells of the Cerebral Cortex. *J. Neurosci.* **34,** 11929–47 (2014).
36. Bennett, M. L. *et al.* New tools for studying microglia in the mouse and human CNS. *Proc. Natl. Acad. Sci. U. S. A.* **113,** E1738–1746 (2016).
37. Sproul, D., Gilbert, N. & Bickmore, W. a. The role of chromatin structure in regulating the expression of clustered genes. *Nat. Rev. Genet.* **6,** 775–781 (2005).
38. Giambartolomei, C. *et al.* Bayesian Test for Colocalisation between Pairs of Genetic Association Studies Using Summary Statistics. *PLoS Genet.* **10,** (2014).
39. Hikami, K. *et al.* Association of a functional polymorphism in the 3'-untranslated region of SPI1 with systemic lupus erythematosus. *Arthritis Rheum.* **63,** 755–63 (2011).
40. Ward, L. D. & Kellis, M. HaploReg: A resource for exploring chromatin states, conservation, and regulatory motif alterations within sets of genetically linked variants. *Nucleic Acids Res.* **40,** (2012).
41. Pham, T. H. *et al.* Dynamic epigenetic enhancer signatures reveal key transcription factors associated with monocytic differentiation states. *Blood* **119,** (2012).
42. Steinberg, S. *et al.* Loss-of-function variants in ABCA7 confer risk of Alzheimer's disease. *Nat. Genet.* **47,** 445–7 (2015).
43. Jonsson, T. *et al.* A mutation in APP protects against Alzheimer's disease and age-related

cognitive decline. *Nature* **488,** 96–9 (2012).
44. Benitez, B. A. *et al.* Missense variant in TREML2 protects against Alzheimer's disease. *Neurobiol. Aging* **35,** (2014).
45. Ruiz, A. *et al.* Follow-up of loci from the International Genomics of Alzheimer's Disease Project identifies TRIP4 as a novel susceptibility gene. *Transl. Psychiatry* **4,** e358 (2014).
46. Cruchaga, C. *et al.* Rare coding variants in the phospholipase D3 gene confer risk for Alzheimer's disease. *Nature* **505,** 550–4 (2014).
47. Bernstein, B. E. *et al.* An integrated encyclopedia of DNA elements in the human genome. *Nature* **489,** 57–74 (2012).
48. Sakai, K. *et al.* A neuronal VLDLR variant lacking the third complement-type repeat exhibits high capacity binding of apoE containing lipoproteins. *Brain Res.* **1276,** 11–21 (2009).
49. Bajari, T. M., Strasser, V., Nimpf, J. & Schneider, W. J. A model for modulation of leptin activity by association with clusterin. *FASEB J.* **17,** 1505–7 (2003).
50. Speliotes, E. K. *et al.* Association analyses of 249,796 individuals reveal 18 new loci associated with body mass index. *Nat. Genet.* **42,** 937–48 (2010).
51. Strawbridge, R. J. *et al.* Genome-wide association identifies nine common variants associated with fasting proinsulin levels and provides new insights into the pathophysiology of type 2 diabetes. *Diabetes* **60,** 2624–34 (2011).
52. Satoh, J.-I., Asahina, N., Kitano, S. & Kino, Y. A Comprehensive Profile of ChIP-Seq-Based PU.1/Spi1 Target Genes in Microglia. *Gene Regul. Syst. Bio.* **8,** 127–39 (2014).
53. Daniel, B. *et al.* The active enhancer network operated by liganded RXR supports angiogenic activity in macrophages. *Genes Dev.* **28,** 1562–1577 (2014).
54. McKercher, S. R. *et al.* Targeted disruption of the PU.1 gene results in multiple hematopoietic abnormalities. *EMBO J.* **15,** 5647–58 (1996).
55. Beers, D. R. *et al.* Wild-type microglia extend survival in PU.1 knockout mice with familial amyotrophic lateral sclerosis. *Proc. Natl. Acad. Sci. U. S. A.* **103,** 16021–6 (2006).
56. Mak, K. S., Funnell, A. P. W., Pearson, R. C. M. & Crossley, M. PU.1 and haematopoietic cell fate: Dosage matters. *International Journal of Cell Biology* (2011). doi:10.1155/2011/808524
57. Smith, A. M. *et al.* The transcription factor PU.1 is critical for viability and function of human brain microglia. *Glia* **61,** 929–942 (2013).
58. Griciuc, A. *et al.* Alzheimer's disease risk gene CD33 inhibits microglial uptake of amyloid beta. *Neuron* **78,** 631–43 (2013).
59. Kleinberger, G. *et al.* TREM2 mutations implicated in neurodegeneration impair cell surface transport and phagocytosis. *Sci. Transl. Med.* **6,** 243ra86 (2014).
60. Yeh, F. L., Wang, Y., Tom, I., Gonzalez, L. C. & Sheng, M. TREM2 Binds to Apolipoproteins, Including APOE and CLU/APOJ, and Thereby Facilitates Uptake of Amyloid-Beta by Microglia. *Neuron* **91,** 328–340 (2016).
61. Jehle, A. W. *et al.* ATP-binding cassette transporter A7 enhances phagocytosis of apoptotic cells and associated ERK signaling in macrophages. *J. Cell Biol.* **174,** 547–556 (2006).
62. Butovsky, O. *et al.* Identification of a unique TGF-β-dependent molecular and functional signature in microglia. *Nat. Neurosci.* **17,** 131–43 (2014).
63. Orre, M. *et al.* Isolation of glia from Alzheimer's mice reveals inflammation anddysfunction. *Neurobiol. Aging* **35,** 2746–2760 (2014).

64. Holtman, I. R. *et al.* Induction of a common microglia gene expression signature by aging and neurodegenerative conditions: a co-expression meta-analysis. *Acta Neuropathol. Commun.* **3,** (2015).
65. Back, J., Allman, D., Chan, S. & Kastner, P. Visualizing PU.1 activity during hematopoiesis. *Exp. Hematol.* **33,** 395–402 (2005).
66. Wang, Y. *et al.* IL-34 is a tissue-restricted ligand of CSF1R required for the development of Langerhans cells and microglia. *Nat. Immunol.* **13,** 753–60 (2012).
67. Noto, D. *et al.* Development of a culture system to induce microglia-like cells from haematopoietic cells. *Neuropathol. Appl. Neurobiol.* **40,** 697–713 (2014).
68. Dagher, N. N. *et al.* Colony-stimulating factor 1 receptor inhibition prevents microglial plaque association and improves cognition in 3xTg-AD mice. *J. Neuroinflammation* **12,** 139 (2015).
69. Selenica, M.-L. B. *et al.* Diverse activation of microglia by chemokine (C-C motif) ligand 2 overexpression in brain. *J. Neuroinflammation* **10,** 86 (2013).
70. Filippo, K. De *et al.* Mast cell and macrophage chemokines CXCL1 / CXCL2 control the early stage of neutrophil recruitment during tissue inflammation. *Blood* **121,** 4930–4937 (2013).
71. Zenaro, E. *et al.* Neutrophils promote Alzheimer's disease–like pathology and cognitive decline via LFA-1 integrin. *Nat. Med.* **21,** (2015).
72. Kiyota, T. *et al.* CCL2 accelerates microglia-mediated A?? oligomer formation and progression of neurocognitive dysfunction. *PLoS One* **4,** (2009).
73. Paolicelli, R. C. *et al.* Synaptic pruning by microglia is necessary for normal brain development. *Science* **333,** 1456–8 (2011).
74. Fourgeaud, L. *et al.* TAM receptors regulate multiple features of microglial physiology. *Nature* **532,** 1–15 (2016).
75. Peskind, E. R. *et al.* Age and apolipoprotein E*4 allele effects on cerebrospinal fluid beta-amyloid 42 in adults with normal cognition. *Arch. Neurol.* **63,** 936–939 (2006).
76. Purcell, S. *et al.* PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* **81,** 559–575 (2007).
77. Howie, B. N., Donnelly, P. & Marchini, J. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. *PLoS Genet.* **5,** (2009).
78. Cox, D. R. Regression Models and Life-Tables. *J. R. Stat. Soc. Ser. B-Statistical Methodol.* **34,** 187–+ (1972).
79. Prentice, R. L. & Breslow, N. E. Retrospective Studies and Failure Time Models. *Biometrika* **65,** 153–158 (1978).
80. van der Net, J. B. *et al.* Cox proportional hazards models have more statistical power than logistic regression models in cross-sectional genetic association studies. *Eur J Hum Genet* **16,** 1111–1116 (2008).
81. Pruim, R. J. *et al.* LocusZoom: Regional visualization of genome-wide association scan results. *Bioinformatics* **26,** 2336–2337 (2010).
82. Zhou, X. *et al.* Exploring long-range genome interactions using the WashU Epigenome Browser. *Nat. Methods* **10,** 375–6 (2013).
83. Griffon, A. *et al.* Integrative analysis of public ChIP-seq experiments reveals a complex multi-cell regulatory landscape. *Nucleic Acids Res.* **43,** (2015).
84. Sloan, C. A. *et al.* ENCODE data at the ENCODE portal. *Nucleic Acids Res.* **44,** D726–D732 (2016).

85. Roe, J. S., Mercan, F., Rivera, K., Pappin, D. J. & Vakoc, C. R. BET Bromodomain Inhibition Suppresses the Function of Hematopoietic Transcription Factors in Acute Myeloid Leukemia. *Mol. Cell* **58,** 1028–1039 (2015).
86. Wang, X. & Seed, B. A PCR primer bank for quantitative gene expression analysis. *Nucleic Acids Res.* **31,** e154 (2003).

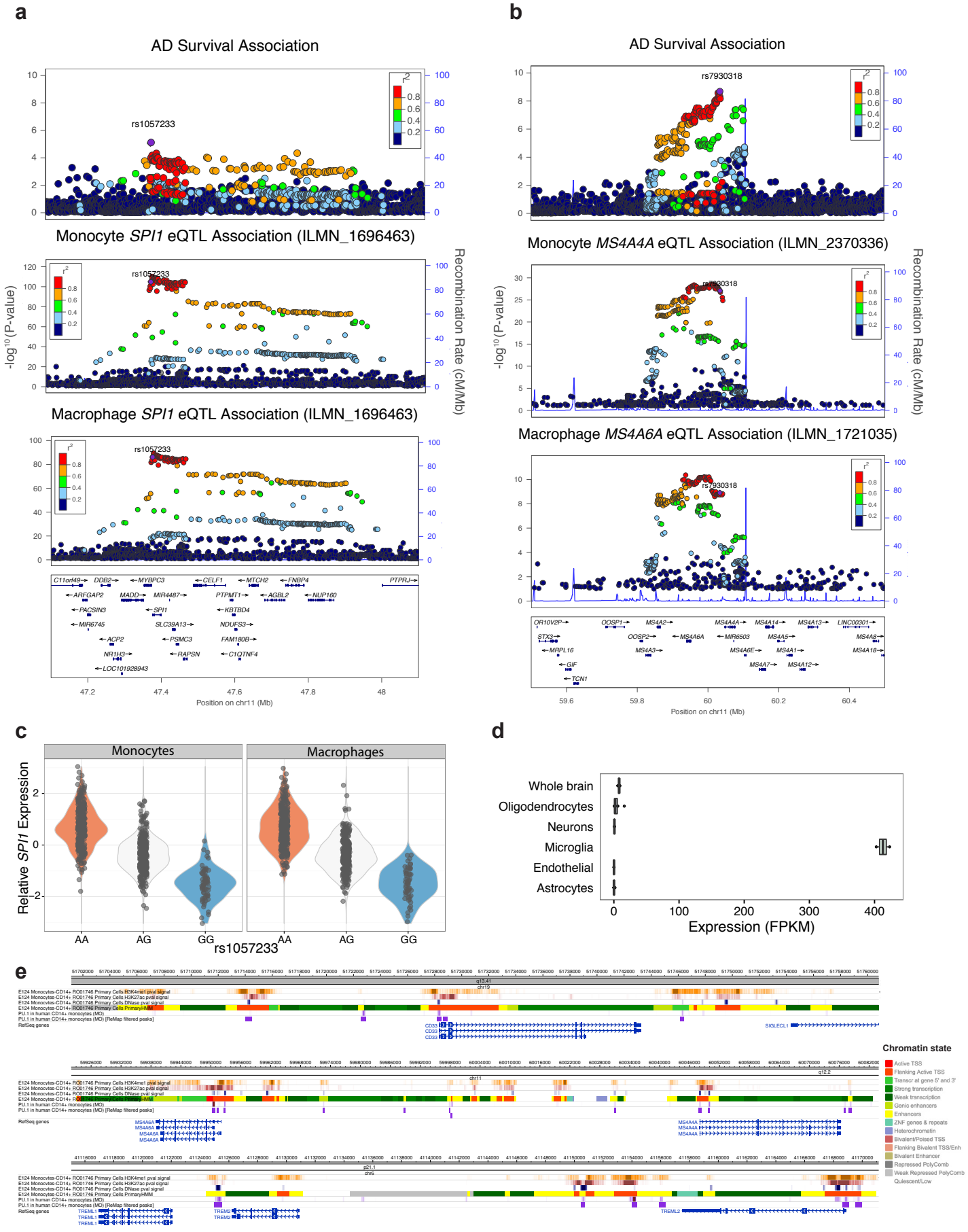

Figure 1. Genetic and eQTL fine-mapping of AD associations and SPI1 expression and ChIP-Seq analysis. (a) The AD-survival association landscape at the *CELF1/SPI1* locus resembles that of *SPI1* eQTL association in monocytes and macrophages. (b) The AD-survival association landscape resembles that of *MS4A4A/MS4A6A* eQTL association in monocytes and macrophages. (c) The Rs1057233^{G} allele is associated with reduced *SPI1* expression in a dosage-dependent manner. (d) The mouse homolog of *SPI1*, Sfpi1, is selectively expressed at microglia in mouse brains. (e) SPI1 (PU.1) binds to the promoter and regulatory regions of CD33, MS4A4A, MS4A6A, TREM2, and TREML2 in human CD14+ monocytes.

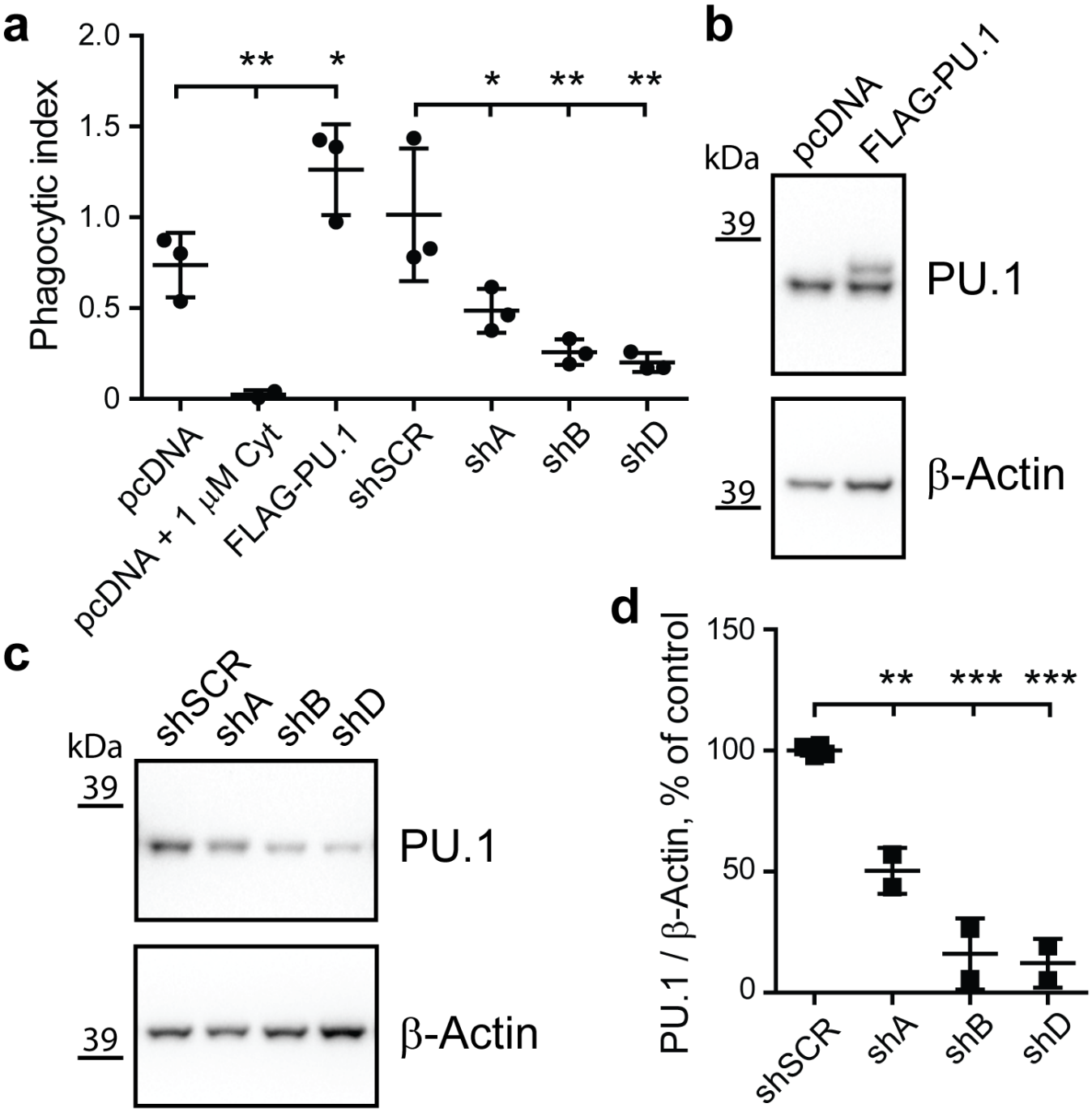

Figure 2. PU.1 is involved in the phagocytic activity of BV2 microglial cells. (a) Phagocytosis of zymosan labeled with red pHrodo fluorescent dye in BV2 cells with transient overexpression and knock-down of PU.1 was measured by flow cytometry. Cytochalasin D treatment was used as a negative control. Mean phagocytic index ± SD is shown: pcDNA 0.7373 ± 0.1772, pcDNA + 1 μM Cyt 0.0236 ± 0.0242, FLAG-PU.1 1.2630 ± 0.2503, shSCR 1.014 ± 0.3656, shA 0.4854 ± 0.1209, shB 0.2579 ± 0.06967, shD 0.2002 ± 0.05168. $F(6,13) = 14.82$, pcDNA vs pcDNA + 1 μM Cyt $P = 0.0078$, pcDNA vs FLAG-PU.1 $P = 0.0295$, shSCR vs shA $P = 0.0283$, shSCR vs shB $P = 0.0020$, shSCR vs shD $P = 0.0010$, $n = 3$. (b) BV2 cells were transiently transfected with pcDNA3 (pcDNA) or pcDNA3-FLAG-PU.1 (FLAG-PU.1) and pCMV-GFP as described for phagocytosis assay. Note a shift in mobility of the band for exogenous FLAG-PU.1 in overexpression condition compared to endogenous PU.1 in control. (c) BV2 cells were transiently transfected with shRNA targeting PU.1 (shA, shB and shD) or non-targeting control (shSCR) in pGFP-V-RS vector. GFP+ cells were sorted with flow cytometer and analyzed for levels of PU.1 in western blotting in two independent experiments (b, c). (d) Quantification of PU.1 levels in c normalized to β-Actin as a loading control. Values are presented as mean ± SD: shSCR 100 ± 2.10, shA 50.34 ± 9.52, shB 16.03 ± 14.72, shD 12.13 ± 10.03. $F(3,6) = 70.55$, shSCR vs shA $P = 0.0014$, shSCR vs shB $P = <0.0001$, shSCR vs shD $P = <0.0001$, $n = 2$. * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$, one-way ANOVA with Sidak's post hoc multiple comparisons test between selected groups.

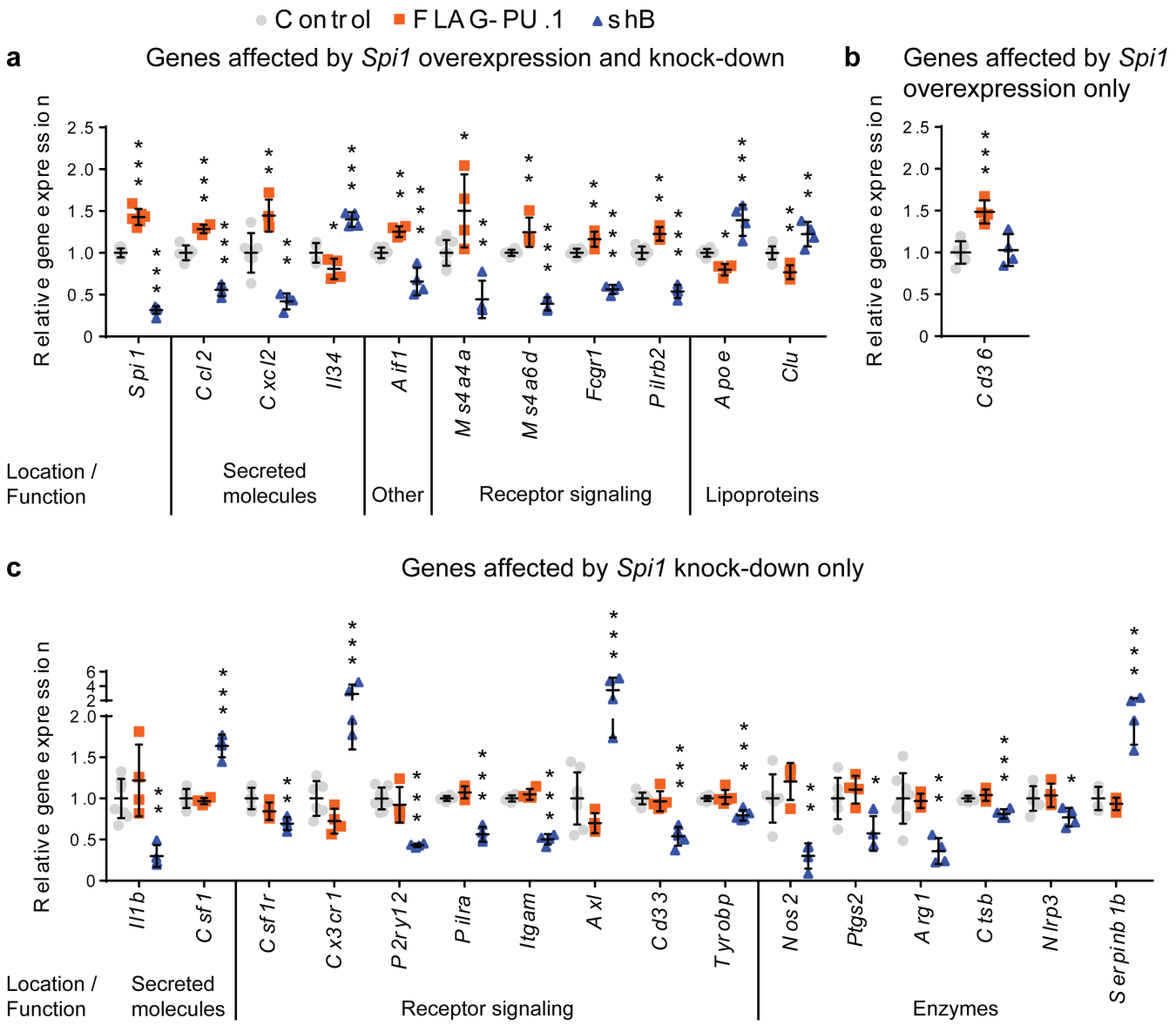

Figure 3. Genes regulated in BV2 microglial cells with differential expression of Spi1. qPCR analysis in transiently transfected and sorted GFP+ BV2 cells with overexpression (FLAG-PU.1) and knock-down (shB) of Spi1. Changes in expression levels are grouped for genes with altered levels after overexpression and knock-down of Spi1 in (a) and genes with variable expression in BV2 cells either with overexpression (b) or knock-down (c) of Spi1. Values are presented as mean ± SD, n = 4 samples collected independently. * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$, one-way ANOVA with Dunnett's post hoc multiple comparisons test between experimental and control groups, detailed statistical analysis is reported in Supplementary Table 11.