A common haplotype lowers SPI1 expression in myeloid cells and delays age at onset for Alzheimer’s disease

Kuan-lin Huang1,2*, Edoardo Marcora3,4*, Anna A Pimenova4, Antonio F Di Narzo3, Manav Kapoor3,4, Sheng Chih Jin5, Oscar Harari6, Sarah Bertelsen4, Benjamin P Fairfax7, Jake Czajkowski8, Jean Charles Lambert9,10,11, Vincent Chouraki12, Céline Bellenguez9,10,11, Benjamin Grenier-Boley9,10,11, Yuetiva Deming6, Andrew McKenzie3, Towfique Raj3,4, Alan E Renton4, John Budde6, Jorge L. Del-Aguila6, Maria Victoria Fernandez6, Laura Ibañez6, The International Genomics of Alzheimer's Project, The Alzheimer's Disease Neuroimaging Initiative6, Rebecca Sims13, Valentina Escott-Price14, Richard Mayeux15,16,17, Jonathan L Haines18, Lindsay A Farre12,19,20,21,22, Margaret A. Pericak-Vance23,24, Sudha Seshadri12, Julie Williams14, Philippe Amouyel9,10,11,25, Gerard D Schellenberg26, Bin Zhang3, Ingrid Borecki27, John S K Kauwe28, Carlos Cruchaga6, Ke Hao3, Alison M Goate3,4

1Department of Medicine, 2McDonnell Genome Institute, 6Department of Psychiatry, 8Department of Genetics, Washington University in St. Louis, Saint Louis, MO, USA 3Department of Genetics and Genomic Sciences, 4Ronald M. Loeb Center for Alzheimer’s disease, Department of Neuroscience, Icahn School of Medicine at Mount Sinai, New York, NY, USA 5Department of Genetics, Yale University School of Medicine, New Haven, CT, USA 7Wellcome Trust Centre for Human Genetics, Nuffield Department of Medicine, University of Oxford, Oxford, United Kingdom 9Inserm, U1167, F-59000 Lille, France 10Univ. Lille, U1167, Risk factors and molecular determinants of aging-related diseases, F-59000 Lille, France 11Institut Pasteur de Lille, U1167, F-59000 Lille, France 12Department of Neurology, 19Department of Medicine (Biomedical Genetics), 20Department of Ophthalmology, Boston University School of Medicine, Boston, MA, USA 13Trinity College, University of Dublin, Dublin, Ireland 14Psychological Medicine and Clinical Neurosciences, Medical Research Council (MRC) Centre for Neuropsychiatric Genetics and Genomics, Cardiff University, Cardiff, UK 15Taub Institute on Alzheimer's Disease and the Aging Brain, 16Gertrude H. Sergievsky Center, 17Department of Neurology, Columbia University, New York, NY, USA 18Department of Epidemiology and Biostatistics, Case Western Reserve University, Cleveland, OH, USA 21Department of Biostatistics, 22Department of Epidemiology, Boston University School of Public Health, Boston, MA, USA 23The John P. Hussman Institute for Human Genomics, 24Macdonald Foundation Department of Human Genetics, University of Miami, Miami, FL, USA 25Centre Hospitalier Universitaire de Lille, U1167, F-59000 Lille, France 26Department of Pathology and Laboratory Medicine, University of Pennsylvania Perelman School of Medicine, Philadelphia, PA, USA 27Regeneron Pharmaceuticals, Inc, NY, USA 28Department 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:
Abstract

In this study we used age at onset of Alzheimer’s disease (AD), cerebrospinal fluid (CSF) biomarkers, and cis-eQTL datasets to fine map AD-associated GWAS loci and investigate the underlying mechanisms. 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 and a further fourteen novel loci reached suggestive significance. One third (8/22) of these SNPs are cis-eQTLs in monocytes and/or macrophages, including rs7930318 associated with expression of MS4A4A and MS4A6A. 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 Aβ_{42} (p=1.2x10^{-4}), and lower expression of SPI1 in monocytes (p = 1.50x10^{-8}) and macrophages (p = 6.41x10^{-8}). 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 SPI1 target gene network in the etiology of AD. Finally, experimentally altered PU.1 levels are correlated with phagocytic activity of mouse BV2 microglial cells and specific changes in the expression of multiple myeloid-expressed genes, including the mouse orthologs of MS4A4A and MS4A6A. 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 associations studies (GWAS) have identified more than twenty AD risk loci\textsuperscript{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\textsuperscript{6,7} have investigated the genetic basis of age-at-onset (AAO) of AD. To date, \textit{APOE} remains the only locus repeatedly shown to associate with AAO\textsuperscript{8–11}. \textit{PICALM} and \textit{BIN1} – two other AD-risk loci – have also been shown to affect AAO using a candidate-gene approach\textsuperscript{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, cerebrospinal fluid (CSF) biomarkers, including Aβ\textsubscript{42} and tau, are tightly linked to the molecular etiology and/or pathology of the disease. Combining such information may help validate and elucidate the AD genetic association landscape. We have previously used this approach to demonstrate that \textit{APOE} genotype is strongly associated with CSF Aβ\textsubscript{42} and total tau levels and to identify novel loci associated with these disease-relevant quantitative traits\textsuperscript{14,15}.

Identifying the underlying disease genes and mechanisms requires integrative analyses of expression and epigenetic datasets in disease-relevant cell types\textsuperscript{16}. In particular, 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 multiple genes expressed in myeloid cells, including \textit{TREM2}, \textit{ABCA7}, and \textit{CD33}\textsuperscript{1,2,5,17–19}. At the epigenetic level, genes expressed in myeloid cells display abnormal patterns of gene expression and chromatin modification in AD mouse models and human samples\textsuperscript{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\textsuperscript{23}. Based on these observations, we hypothesized that 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 Aβ\textsubscript{42}-associated SNP, rs1057233, in the previously reported \textit{CELF1} locus. Cis-eQTL analyses revealed a highly significant association of the protective rs1057233\textsuperscript{G} allele with reduced \textit{SPI1} expression in monocytes and macrophages. \textit{SPI1} encodes PU.1, a transcription factor critical for myeloid 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 \textit{SPI1} cistromes in these cells, implicating a myeloid \textit{SPI1} 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 mouse BV2 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 \textit{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 (\(P < 5 \times 10^{-8}\)) with AAOS: \(BIN1\) (\(p=7.6 \times 10^{-13}\)), \(MS4A\) (\(p=5.1 \times 10^{-11}\)), \(PICALM\) (\(p=4.3 \times 10^{-14}\)), and \(APOE\) (\(p=1.2 \times 10^{-67}\)) (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 \(MS4A\)\(^{4A}\) is associated with a later AAO.

Four other AD risk loci previously reported in the IGAP GWAS showed associations that reached suggestive significance (\(p < 1.0 \times 10^{-5}\)): \(CR1\) (\(p=1.2 \times 10^{-6}\)), \(SPI1/CELF1\) (\(p=5.4 \times 10^{-6}\)), \(SORL1\) (\(p=1.8 \times 10^{-7}\)), and \(FERMT2\) (\(p=1.0 \times 10^{-5}\)). The directionalities of the effects were concordant with the IGAP GWAS 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). 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 multiple testing threshold: \(0.05/22 = 2.27 \times 10^{-3}\)) in the IGAP GWAS (Table 1b, Supplementary Fig. 2).

Cerebrospinal fluid biomarkers associations

To further validate the 22 loci with at least suggestive associations to AAOS, we examined their associations with established CSF biomarkers, including total tau, phosphorylated tau\(_{181}\), and Aβ\(_{42}\) in a dataset of 3,646 Caucasians extended from our previous report\(^14\) (Table 2). Two SNPs showed associations that reached the Bonferroni multiple-testing threshold (\(P < 2.27 \times 10^{-3}\)). Rs4803758 near \(APOE\) showed the most significant associations with levels of CSF phosphorylated tau\(_{181}\) (\(p=5.81 \times 10^{-4}\)) and CSF Aβ\(_{42}\) (\(p=6.75 \times 10^{-5}\)), whereas rs1057233 in the \(SPI1/CELF1\) locus was significantly associated with CSF Aβ\(_{42}\) (\(p=4.11 \times 10^{-4}\)). Of note, a SNP adjacent to \(VLDLR\), rs7867518, showed the most significant association with CSF total tau (\(p=3.02 \times 10^{-3}\)), but failed to pass the Bonferroni multiple-testing 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 Aβ\(_{42}\) levels and the risk rs1057233\(^A\) allele was associated with lower CSF Aβ\(_{42}\) levels.

cis-eQTL associations and colocalization analysis

Multiple disease-associated GWAS SNPs have also been identified as cis-eQTLs of disease genes and integration of these datasets obtained from disease-relevant tissues/cell types may uncover novel genes associated with disease\(^28\). First 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 2) in the BRAINEAC dataset encompassing ten different brain regions. We identified 4 significant associations (Bonferroni correction threshold: \(0.05/292,000\) probes = \(1.7 \times 10^{-7}\)); rs1057233 was associated with \(MTCH2\) expression in the cerebellum (\(P = 1.20 \times 10^{-9}\)); rs7445192 was associated with averaged \(SRA1\) expression across brain regions (\(P = 7.0 \times 10^{-9}, 1.6 \times 10^{-7}\) for two probes respectively), and rs2093761 was associated with \(CR1/CR1L\) expression in the white
matter ($P = 1.30 \times 10^{-7}$, Supplementary Table 3). Further analysis using the GTEx dataset$^{29}$ also showed potential eQTL association of rs1057233 with $C1QTNF4$ across 18 tissues and $MTC2$ in the brain cortex and nucleus accumbens/basal ganglia (Supplementary Table 4).

We have previously reported that monocyte cis-eQTLs are enriched among AD GWAS SNPs$^{23}$. Further, AD is associated with genetic variation in multiple myeloid-expressed genes, including $TREM2$, $ABCA7$, and $CD33$$^{1,2,5,17–19}$. We hypothesized the cis-eQTL effects of some AD-associated alleles may be specific to myeloid cells and thus not easily detectable in cis-eQTL datasets obtained from brain homogenates where microglial cells constitute only a minor fraction of the tissue. Therefore, we analyzed cis-eQTL effects of the AD survival-associated SNPs and their tagging SNPs in human cis-eQTL datasets composed of 738 monocyte and 593 macrophage samples from the Cardiogenics consortium$^{30}$. We identified 14 genes with cis-eQTLs significantly associated with these SNPs (Table 3). Notably, the protective rs1057233$^{G}$ allele, located within the 3' UTR of $SPI1$, was strongly associated with lower expression of $SPI1$ in both monocytes ($p = 1.50 \times 10^{-105}$) and macrophages ($p = 6.41 \times 10^{-87}$) with similar dosage-dependent effects (Fig. 1a, 1c). This allele was also associated with lower expression of $MYBPC3$ (monocytes: $p = 5.58 \times 10^{-23}$; macrophages: $p = 4.99 \times 10^{-51}$), higher expression of $CELF1$ in monocytes ($p = 3.95 \times 10^{-8}$) and lower $NUP160$ expression in macrophages ($p = 5.35 \times 10^{-22}$). Each of these genes lies within the $SPI1/CELF1$ locus, suggesting complex regulation of gene expression in this chromosomal region. The minor allele (C) of rs7930318 was consistently associated with lower expression of $MS4A4A$ in monocytes ($p = 8.20 \times 10^{-28}$) and $MS4A6A$ in monocytes and macrophages (Fig. 1b, monocytes: $p = 4.90 \times 10^{-23}$; macrophages: $p = 1.25 \times 10^{-9}$). Among the novel AD survival-associated loci, rs5750677 was significantly associated with lower expression of $SUN2$ in both monocytes ($p = 3.66 \times 10^{-58}$) and macrophages (Fig. 1c, monocytes: $p = 7.33 \times 10^{-35}$), and rs1625716 was associated with lower expression of $CISD1$ in macrophages ($p = 5.98 \times 10^{-23}$, Table 3).

We then sought evidence of replication in an independent dataset of primary CD14+ human monocytes from 432 individuals of European ancestry$^{31}$. We replicated cis-eQTL associations with expression of $SPI1$, $MYBPC3$, $MS4A4A$, $MS4A6A$, and $SELL$ (Bonferroni correction threshold: $0.05/15421$ probes $= 3.24 \times 10^{-6}$). We found strong evidence of the association between rs1057233 and $SPI1$ expression ($p = 6.39 \times 10^{-102}$) as well as $MYBPC3$ expression ($p = 5.95 \times 10^{-33}$, Supplementary Table 5). Rs1530914 and rs7929589, both in high LD with rs7930318 ($R^2 = 0.99$ and 0.87, respectively), were associated with expression of $MS4A4A$ and $MS4A6A$ ($p = 3.60 \times 10^{-8}$, $6.37 \times 10^{-15}$, respectively). Finally, rs2272918, tagging rs10919252, was significantly associated with expression of $SELL$ ($p = 8.43 \times 10^{-16}$). 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$^{32–34}$ from human and mouse acutely-isolated brain cell types (Fig. 1c, Supplementary Fig. 3). 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. $MYBPC3$ (ILMN_1781184) gene expression is most highly and significantly correlated with $SPI1$ (ILMN_1696463) expression in both Cardiogenics datasets (Spearman’s $r = 0.54$, qval = 0.00 in monocytes and Spearman’s $r = 0.42$, qval = 0.00 in macrophages) suggesting that low levels of expression in human myeloid cells are
possibly due to leaky transcription driven by the adjacent highly expressed SPI1 gene\(^{35}\) (Supplementary Fig. 5).

We then performed the coloc statistical test\(^{36}\) to further validate the colocalization of AD survival-associated SNPs with myeloid cis-eQTLs at the SPI1/CELF1, MS4A and SELL loci. The results of these analyses (Supplementary Table 6) highlighted SPI1 at the 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 CELF1 locus and MS4A6E 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 co-localization in either cell type. Similar results were obtained when using AD-associated SNPs from the IGAP GWAS to support SPI1 at the CELF1 locus as a candidate causal gene for AD in myeloid cells.

To prioritize putative functional variants underlying the colocalization of AD survival-associated SNPs and myeloid cis-eQTLs, we applied HaploReg\(^{37}\) to annotate the top survival SNP in the SPI1/CELF1 locus (rs1057233) and its tagging SNPs ($R^2 \geq 0.8$, Supplementary Table 2). Interestingly, four SNPs in tight LD with rs1057233 changed the predicted DNA binding motif of PU.1. For example, rs7928163 ($R^2$ with rs1057233=0.94) changed the known1 motif and rs10838699 ($R^2$ with rs1057233=0.96) changed the known2 motif (Supplementary Table 2), raising the possibility of altered self-regulation and potentially decreased PU.1 binding in the presence of the minor allele. Alternatively, rs1057233 was previously shown to change the target sequence and binding of miR-569 and its transcriptional repression on SPI1\(^{38}\). Based on these results, one or more of these SNPs, which are all in very high LD, could explain the observed associations with SPI1 expression and AD-related phenotypes. Overall, rs1057233 and tagging SNPs are associated with AD risk and survival, and CSF Aβ$_{42}$. The strong cis-eQTL effects and colocalization results point to SPI1 as the most likely candidate gene underlying the disease association at this locus.

**Fine-mapping 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 for AD risk\(^1\), rs10838725 showed the strongest association at this locus (rs10838725: $p = 1.1 \times 10^{-8}$ vs. rs1057233: $p = 5.9 \times 10^{-7}$ in stage I and II combined). 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 ($p=5.4 \times 10^{-6}$). The two SNPs exhibit moderate linkage disequilibrium in the ADGC subset of the IGAP GWAS dataset ($R^2=0.21$, $D'=0.96$). Applying AD risk analysis in the ADGC dataset, conditional analysis revealed that rs1057233 remained significantly associated with AD after controlling for rs10838725 ($P=3.2 \times 10^{-4}$), whereas rs10838725 showed no evidence of association after adjusting for rs1057233 ($P=0.66$). The association landscape in the AD survival analysis highly resembles that of SPI1 cis-eQTL analysis in myeloid cells (Fig. 1a). To investigate the SPI1 cis-eQTL associations across ethnicities, we used the NCBI eQTL browser to visualize these associations in different HapMap
lymphoblastoid cell lines (Supplementary Fig. 4). While the tagging SNP of rs1057233, rs10838698 ($R^2 = 0.96$), was strongly associated with $SPII$ expression in Caucasians and Asians ($p = 3.21 \times 10^{-9}, 1.03 \times 10^{-17}$ respectively), it was not associated with $SPII$ expression in Africans ($p = 0.22$). Interestingly, rs1057233 also showed no evidence of association with AD risk in African Americans ($p = 0.71$), consistent with the hypothesis that this $SPII$ eQTL may explain the disease association at this locus.

We reasoned that the associations of rs1057233 with AD-related phenotypes may be explained by the regulation of $SPII$ expression in myeloid cells, and that fine-mapping the cis-eQTL signal could help us pinpoint the functional variant. Therefore, we conducted conditional analyses based on six SNPs of interest in this locus using both Cardiogenics datasets: rs1057233 (the top survival SNP), rs10838698 (the directly genotyped SNP in high LD with rs1057233), rs10838699 (a SNP that modifies a PU.1 binding motif), rs7928163 (a SNP that modifies a PU.1 binding motif), rs1377416 (a putative enhancer SNP of $SPII$), and rs10838725 (the top SNP for AD risk in the previous IGAP GWAS). Rs1057233, rs10838698, rs10838699, and rs7928163 all remained significantly associated with $SPII$ expression when adjusting for the other two SNPs in both monocytes and macrophages ($P < 8.33 \times 10^{-3}$). On the other hand, conditioning for any of these four SNPs abolished the associations of rs1377416 and rs10838725 to $SPII$ expression (Supplementary Table 7). Thus, the functional variants mediating the effect on $SPII$ expression likely reside in the LD block that includes rs1057233, rs10838698, rs10838699 and rs7928163.

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 to prioritize likely causal genes and variants by integrating summary statistics from our AAOS GWAS and the Cardiogenics study (Supplementary Table 7). SMR/HEIDI analysis was performed for the $SPII/CELF1$ locus using rs1057233, rs10838698, rs10838699, rs7928163, rs1377416, rs10838725 as candidate causal variants. In both monocytes and macrophages, $SPII$ 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 or rs10838698, rs10838699, rs7928163 in the same LD block) (SMR $P < 4.90 \times 10^{-4}$, the multiple testing threshold for 6 SNPs tested against 17 probes and HEIDI $P \geq 0.05$). Similar results were obtained using IGAP GWAS summary statistics (Supplementary Table 7). Neither conditional analysis nor this SMR/HEIDI analysis could definitively identify a single functional variant in this locus among the set of 4 SNPs in high LD. Functional analyses will be necessary to determine which of these SNPs directly affects $SPII$ expression.

$SPII/PU.1$ cistrome and functional analysis in myeloid cells
To further evaluate $SPII$ as a candidate causal gene for AD, we investigated the functional impact of variation in $SPII$ expression. $SPII$ 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 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. We specifically investigated 112 AD-associated genes, including the 104 genes located within the IGAP GWAS loci as defined by Steinberg et al. and additionally $APOE, APP, TREM2$ and $TREM2$, $TYROBP$, $TRIP$, $CD33$, and $PLD$. Among the 112 AD-associated genes, 75 had evidence of gene expression in human microglia/brain myeloid cells, 60 of which had also evidence of association with one or more nearby PU.1 binding sites in
human blood myeloid cells (monocytes or macrophages)\textsuperscript{40} (Supplementary Table 9). Further examination of PU.1 binding peaks and chromatin marks/states in human monocytes and macrophages confirmed that PU.1 is likely bound to cis-regulatory elements in the proximity of several 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 (Supplementary Fig. 5). Together, these results suggest that PU.1 may regulate the expression of multiple AD-associated genes in disease-relevant cell types\textsuperscript{23.}

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\textsuperscript{46} to estimate enrichment of AD heritability (as measured by GWAS summary statistics from the IGAP consortium\textsuperscript{1}) partitioned across the whole PU.1 cistrome, as profiled by ChIP-Seq in human monocytes and macrophages\textsuperscript{40}. Indeed, we found a significant enrichment of AD heritability in both monocytes (56 fold enrichment, $P = 0.003$) and macrophages (60 fold enrichment, $P = 0.001$), but this was not the case for schizophrenia (SCZ) heritability [as measured by GWAS summary statistics from the Psychiatric Genomics Consortium (PGC)\textsuperscript{47}] (Supplementary Table 10), suggesting that the contribution of the myeloid PU.1 target gene network to disease susceptibility is likely specific to AD.

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\textsuperscript{48} 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 mouse BV2 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. 6d, 6e, 6f, 6g). 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 modulates levels of myeloid genes that are thought to play important roles in AD pathogenesis and/or microglial cell function (Supplementary Table 11). We found that levels of some of these genes were affected in opposing directions by over-expression and knock-down of Spi1 (Fig. 3b), while that of other genes was affected only by over-expression (Fig. 3c) or knock-down (Fig. 3d) or not affected at all (Supplementary Fig. 7). In particular, over-expression of Spi1 led to up-regulation of Ms4a4a, Ms4a6d (mouse ortholog of human MS4A4A, MS4A6A), Ccl2, Cxcl2, Aif1, C6d4, Pilrb, 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, Csfl, Cx3cr1, Axl, Serpinb1 and down-regulation of Ms4a4a, Ms4a6d, Cd33, Tyrobp, Ccl2, Cxcl2, Aif1, C6d4, Pilrb, Il1b, Csflr, P2ry12, Pilra, Itgam, Nos2, Cox2, 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\(^2\). VLDLR, or the very-low-density-lipoprotein receptor, binds to APOE-containing lipoproteins in the brain\(^4\) and physically interacts with CLU, another AD risk gene\(^5\). Additionally, the VLDLR-5-repeat allele was found to be associated with dementia\(^4\).

Collectively, this evidence suggests that genetic variation in VLDLR may be linked to APOE and AD, although further replication and investigation are 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. 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 SEL, 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 SEL 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 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 age-at-onset defined survival (p = 8.4x10^{-6}) and significantly associated with higher CSF Aβ\(_{42}\) (p = 4.11x10^{-4}), which likely reflects decreased Aβ aggregation and ß-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). Its tagging SNP was also associated with SPI1 expression in lymphoblastoid cells in Caucasians and Asians. Interestingly the eQTL association was not identified in Africans, and rs1057233 is not associated with AD risk in African Americans (P=0.71)\(^3\), suggesting that this cis-eQTL is likely responsible for the disease association at this locus and its effect may be ethnicity-specific.
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 multiple SNPs within a single LD bin, tagged by rs1057233, which may (individually or in combination) influence both SPI1 expression and AD risk. rs1057233 directly changes the target sequence and binding of miR-569 and its transcriptional repression on SPI1, and its tagged SNPs alter binding motifs of transcription factors including PU.1 itself (Supplementary Table 1 and Fig. XXX). 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 microglias21. However, rs1057233 remained significantly associated with AD after conditioning for either rs1377416 (p = 1.2x10^{-3}) or the previously reported IGAP GWAS SNP rs10838725 (p = 3.2x10^{-2}) 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 4). 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, body mass index, and proinsulin levels 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 genes expressed in myeloid cells, including ABCA7, CD33, MS4A4A, MS4A6A, TREM2, and TYROBP (Supplementary Table 5). This finding is further supported by PU.1 binding to active enhancers of Trem2 and Tyrobp in ChIP-Seq experiments using the BV2 mouse microglial cell line or bone marrow-derived mouse macrophages. PU.1 is required in mouse for the development and function of myeloid and B-lymphoid cells. Given its selective expression in microglia in the brain (Fig. 1c), 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 10).

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. 6). 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 Aβ42 (also a TLR2 agonist) in primary microglial cells isolated from adult human brain tissue and transfected with siRNA targeting SPI1. Interestingly, several AD-associated genes (e.g., CD33, TYROBP, TREM2, TREML2, CR1, ABCA7, APOE, CLU/APOJ) 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.

We show that in BV2 microglial cells expression of Cd33 and Tyrobp are decreased and expression of Apoe and Clu/Apoj are increased after knock-down of Spi1 (Fig. 3a, 3b). Indeed, several other genes are dysregulated after altering Spi1 expression, i.e. Cd36, C6d4, Pilra, Pilrb, Ms4a4a, Ms4a6d, P2ry12, Itgam, Cx3cr1, Axl, Ctsb (Fig. 3b, 3c, 3d), suggesting a collective and coordinated effect of Spi1 on the phagocytic activity of BV2 microglial cells. Furthermore, expression of Illb, Nos2, Cox2, Arg1, and Nlrp3 are decreased after knock-down of Spi1 (Fig...
3d), consistent with blunting of the inflammatory response that is often up-regulated in AD brains and regarded as neurotoxic. Furthermore our genetic analyses show that the protective alleles 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 (the 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. 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, 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, were elevated after knock-down of Spi1, while expression of Csf1r was reduced (Fig. 3b, 3d). Interestingly, inhibition of Csf1r in a 3xTg-AD mouse model led to a reduction in the number of microglia associated with ß-amyloid plaques and improved cognition. 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, as Ccl2 and Cxcl2 (MCP1 and MIP2α proteins) expression was directly dependent on Spi1 levels (Fig. 3b). Both molecules participate in recruitment of circulating monocytes and neutrophils to the brain, which can promote neuroinflammation and are detrimental in an AD mouse model. In addition, expression of a microgliosis marker Aif1 (IBA1 protein) was dependent on Spi1 (Fig. 3b), which in conjunction with changes in Il1b, Nos2, Cx3cr1 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 was markedly elevated upon knock-down of Spi1 (Fig. 3b), raising the possibility that beneficial effects of changes in Spi1 expression are exerted through modulation of synaptic or neuronal clearance. 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 mRNA 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\(^2,5,19,48,21\)–\(^23\), 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 data set 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 data set 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 Aβ42 levels and age appears to differ in individuals below 45 years vs. those above 45 years\(^7,5\). 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 PLINK\(^1,9\). 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 x 10\(^-8\) and missingness > 2% were removed prior to imputation. Genome-wide genotype imputation was performed using IMPUTE2\(^27\) 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 $1 \times 10^{-6}$. We excluded SNPs with minor allele frequency $\leq 0.02$. Finally, we excluded SNPs with available statistics in only one consortium dataset in the meta-analysis.

### Genome-wide survival association study

We conducted a genome-wide Cox proportional hazards regression 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 in all our regression models to control for their effects. For EADI, sex and four principle 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:

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

where $X = (X_1, X_2, \ldots, X_p)$ 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\cite{79,80}. After the analysis of each dataset, we carried out an inverse-variance meta-analysis on the results using METAL, applying a genomic control to adjust for inflation in each dataset. Of the 751 suggestive SNPs ($P < 1\times10^{-5}$), we found these SNPs to show lower standard errors and confidence intervals with the increasing number of cohorts showing consistent directionality of effect. Particularly, the average standard error for SNPs showing 1 to 7 consistent directionality ranges from 0.171, 0.109, 0.0744, 0.0346, 0.0234, 0.0173 to 0.0179 (Supplementary Fig. 1b). Thus, we limited our final analysis to SNPs that showed consistent directionality 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.

### CSF biomarker association analysis

For the CSF datasets, we performed multivariate linear regression for CSF $A\beta_{42}$ and tau, and $p\tau_{181}$ 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 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 $\beta$, 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.\textsuperscript{31} 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 $\beta$.

**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.\textsuperscript{32,33} and Bennett et al.\textsuperscript{34} 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\textsuperscript{37} 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\textsuperscript{82} using public Roadmap Epigenomics Consortium as well as custom tracks 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 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.\textsuperscript{28} and the companion website http://cnsgenomics.com/software/smr. 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)\textsuperscript{46} to estimate heritability of AD and schizophrenia from GWAS summary statistics (excluding the APOE and MHC regions) partitioned by PU.1 ChIP-Seq binding sites in myeloid cells, as described in https://github.com/bulik/ldsc/wiki/Partitioned-Heritability 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\textsuperscript{1} (phase1 dataset) and the Psychiatric Genomics Consortium (PGC)\textsuperscript{47} (pgc.cross.scz dataset), respectively. PU.1 bindings sites were downloaded as filtered ChIP-Seq peaks in BED format from ReMap\textsuperscript{83} (GSE31621, SPI1, blood monocyte and macrophage datasets\textsuperscript{40}) (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 (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 6 (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 repeated measures one-way ANOVA and Sidak’s post hoc multiple comparisons test, 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 labeled.

Western blotting

BV2 cells transiently transfected as described for the phagocytosis assay were collected with trypsin after 48 hour 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 repeated measures one-way ANOVA and Sidak’s
post hoc multiple variance test, 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 RNeasy 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 PrimerBank⁸⁴ or designed using Primer-BLAST program (NCBI) and are listed in **Supplementary Table 12**. 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 mean values of pcDNA3 and pGFP-V-RS-shSCR control samples in each sort giving a fold change of 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. Adjusted P values for each comparison are reported in the figure legends.

**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**

**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 data for the eQTL analysis.

This work was supported by grants from the National Institutes of Health (U01 AG049508 (AMG), R01-AG044546 (CC) and R01-AG035083 (AMG), R01-AG054011 (AMG)), the JPB Foundation (AMG) and FBRI (AMG). This research was conducted while C.C. was a recipient of a New Investigator Award in Alzheimer’s disease from the American Federation for Aging Research. C.C. is a recipient of a BrightFocus Foundation Alzheimer’s Disease Research Grant (A20133595). The recruitment and clinical characterization of research participants at Washington University were supported by NIH P50 AG05681, P01 AG03991, and P01 AG026276. K.H. received fellowship funding in part from the Ministry of Education in Taiwan and the Lucille P. Markey Special Emphasis Pathway in Human Pathobiology. 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
analysis and functional experiments. K.H., A.A.P., E.M., and A.M.G. wrote and edited the manuscript. All authors read and edited 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 \( p < 5 \times 10^{-8} \) or suggestive \( p < 1 \times 10^{-5} \) associations from the genome-wide survival analysis.

### a

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

### b

<table>
<thead>
<tr>
<th>SNP</th>
<th>Major/minor Alleles</th>
<th>MAF</th>
<th>CHR</th>
<th>BP</th>
<th>Closest Gene</th>
<th>Logistic OR ( (95% \text{ CI}) )</th>
<th>P value</th>
<th>Survival HR</th>
<th>Survival Genotype</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs2093761</td>
<td>G/A</td>
<td>0.2019</td>
<td>1</td>
<td>207786542</td>
<td>CRI</td>
<td>1.16 (1.12-1.20)</td>
<td>2.60x10^{-11}</td>
<td>1.07 (1.04-1.10)</td>
<td>1.20x10^{-6}</td>
</tr>
<tr>
<td>rs6431219</td>
<td>C/T</td>
<td>0.4163</td>
<td>2</td>
<td>127862133</td>
<td>BIN1</td>
<td>1.12 (1.09-1.15)</td>
<td>7.60x10^{-10}</td>
<td>1.08 (1.06-1.10)</td>
<td>3.90x10^{-10}</td>
</tr>
<tr>
<td>rs1057233</td>
<td>A/G</td>
<td>0.3194</td>
<td>11</td>
<td>47376448</td>
<td>SII/CEF1 ( \alpha )</td>
<td>0.93 (0.89-0.96)</td>
<td>5.40x10^{-6}</td>
<td>0.94 (0.91-0.97)</td>
<td>8.40x10^{-8}</td>
</tr>
<tr>
<td>rs7930318</td>
<td>T/C</td>
<td>0.4004</td>
<td>11</td>
<td>60033371</td>
<td>MS4A</td>
<td>0.90 (0.87-0.93)</td>
<td>5.10x10^{-11}</td>
<td>0.93 (0.90-0.95)</td>
<td>2.30x10^{-9}</td>
</tr>
<tr>
<td>rs567075</td>
<td>C/T</td>
<td>0.3097</td>
<td>11</td>
<td>85830157</td>
<td>PICALM</td>
<td>0.88 (0.85-0.91)</td>
<td>4.30x10^{-11}</td>
<td>0.91 (0.89-0.94)</td>
<td>9.10x10^{-12}</td>
</tr>
<tr>
<td>rs9665907</td>
<td>G/A</td>
<td>0.1133</td>
<td>11</td>
<td>121435470</td>
<td>SORL1</td>
<td>0.88 (0.83-0.93)</td>
<td>1.80x10^{-7}</td>
<td>0.92 (0.88-0.95)</td>
<td>5.50x10^{-6}</td>
</tr>
<tr>
<td>rs17125944</td>
<td>T/C</td>
<td>0.0924</td>
<td>14</td>
<td>53400629</td>
<td>FERMT2</td>
<td>1.13 (1.08-1.18)</td>
<td>1.00x10^{-9}</td>
<td>1.10 (1.06-1.14)</td>
<td>2.30x10^{-9}</td>
</tr>
<tr>
<td>rs4803758</td>
<td>G/T</td>
<td>0.3551</td>
<td>19</td>
<td>45327423</td>
<td>APOE ( \beta )</td>
<td>1.33 (1.30-1.37)</td>
<td>1.20x10^{-6}</td>
<td>1.21 (1.18-1.23)</td>
<td>7.80x10^{-10}</td>
</tr>
</tbody>
</table>

### Novel loci reaching suggestive significance

<table>
<thead>
<tr>
<th>SNP</th>
<th>Major/minor Alleles</th>
<th>MAF</th>
<th>CHR</th>
<th>BP</th>
<th>Closest Gene</th>
<th>Logistic OR ( (95% \text{ CI}) )</th>
<th>P value</th>
<th>Survival HR</th>
<th>Survival Genotype</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs10919252</td>
<td>C/G</td>
<td>0.3275</td>
<td>1</td>
<td>169802956</td>
<td>CLoft12</td>
<td>1.04 (1.01-1.08)</td>
<td>1.10x10^{-2}</td>
<td>1.10 (1.06-1.14)</td>
<td>8.20x10^{-4}</td>
</tr>
<tr>
<td>rs1532244</td>
<td>A/G</td>
<td>0.0925</td>
<td>3</td>
<td>28057905</td>
<td>CMC1</td>
<td>0.95 (0.90-1.01)</td>
<td>6.90x10^{-2}</td>
<td>0.86 (0.80-0.93)</td>
<td>9.70x10^{-4}</td>
</tr>
<tr>
<td>rs116341973</td>
<td>A/G</td>
<td>0.0227</td>
<td>3</td>
<td>63462893</td>
<td>SYNP1</td>
<td>1.20 (1.09-1.30)</td>
<td>5.40x10^{-4}</td>
<td>1.23 (1.15-1.31)</td>
<td>2.50x10^{-6}</td>
</tr>
<tr>
<td>rs71602496</td>
<td>A/G</td>
<td>0.1453</td>
<td>4</td>
<td>661002</td>
<td>PDE5B</td>
<td>1.02 (0.98-1.06)</td>
<td>3.60x10^{-1}</td>
<td>1.08 (1.05-1.11)</td>
<td>5.00x10^{-4}</td>
</tr>
<tr>
<td>rs1699013</td>
<td>T/C</td>
<td>0.2493</td>
<td>4</td>
<td>181048651</td>
<td>LINCO0290</td>
<td>1.02 (0.98-1.06)</td>
<td>2.70x10^{-2}</td>
<td>1.07 (1.04-1.09)</td>
<td>4.70x10^{-6}</td>
</tr>
<tr>
<td>rs7445192</td>
<td>A/G</td>
<td>0.461</td>
<td>5</td>
<td>140138701</td>
<td>PCDHA1</td>
<td>NA</td>
<td>NA</td>
<td>1.06 (1.03-1.10)</td>
<td>7.90x10^{-6}</td>
</tr>
<tr>
<td>rs12207208</td>
<td>T/C</td>
<td>0.1034</td>
<td>6</td>
<td>40301379</td>
<td>LINCO0951</td>
<td>1.07 (1.02-1.20)</td>
<td>1.20x10^{-2}</td>
<td>1.09 (1.05-1.13)</td>
<td>6.80x10^{-5}</td>
</tr>
<tr>
<td>rs17170228</td>
<td>G/A</td>
<td>0.0623</td>
<td>7</td>
<td>33076314</td>
<td>NT5C3A</td>
<td>1.07 (1.01-1.14)</td>
<td>2.50x10^{-2}</td>
<td>1.13 (1.08-1.18)</td>
<td>1.00x10^{-6}</td>
</tr>
<tr>
<td>rs2725066</td>
<td>A/T</td>
<td>0.4872</td>
<td>8</td>
<td>4438058</td>
<td>CSMD1</td>
<td>1.03 (1.00-1.06)</td>
<td>7.30x10^{-2}</td>
<td>1.10 (1.06-1.14)</td>
<td>1.00x10^{-5}</td>
</tr>
<tr>
<td>rs7867518</td>
<td>T/C</td>
<td>0.476</td>
<td>9</td>
<td>2527525</td>
<td>VLDLR</td>
<td>0.97 (0.94-1.00)</td>
<td>6.80x10^{-2}</td>
<td>0.95 (0.92-0.97)</td>
<td>9.10x10^{-4}</td>
</tr>
</tbody>
</table>

---

Previously reported associated loci:

Novel loci reaching suggestive significance:
<table>
<thead>
<tr>
<th>SNP</th>
<th>alleles</th>
<th>MAF</th>
<th>chr</th>
<th>position</th>
<th>gene</th>
<th>p-value</th>
<th>OR (95% CI)</th>
<th>OR (95% CI)</th>
<th>OR (95% CI)</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs1625716</td>
<td>T/G</td>
<td>0.0643</td>
<td>10</td>
<td>59960083</td>
<td>IPMK</td>
<td>0.87</td>
<td>(0.80-0.94)</td>
<td>1.00x10^-4</td>
<td>0.88</td>
<td>0.95</td>
</tr>
<tr>
<td>rs1118069</td>
<td>T/A</td>
<td>0.2805</td>
<td>12</td>
<td>84739181</td>
<td>SLC6A15</td>
<td>0.98</td>
<td>(0.94-1.01)</td>
<td>2.00x10^-1</td>
<td>0.90</td>
<td>0.87</td>
</tr>
<tr>
<td>rs11074412</td>
<td>A/G</td>
<td>0.2087</td>
<td>16</td>
<td>19833001</td>
<td>IQCK</td>
<td>0.94</td>
<td>(0.90-0.98)</td>
<td>1.90x10^-3</td>
<td>0.93</td>
<td>0.90</td>
</tr>
<tr>
<td>rs5750677</td>
<td>C/T</td>
<td>0.2885</td>
<td>22</td>
<td>39147715</td>
<td>SUN2</td>
<td>0.97</td>
<td>(0.93-1.00)</td>
<td>5.10x10^-2</td>
<td>0.94</td>
<td>0.91</td>
</tr>
</tbody>
</table>

*Build 37, assembly hg19. Summary statistics of the logistic regression result was obtained from stage 1 of the 2013 IGAP landmark GWAS paper. Calculated with respect to the minor allele. SPI1 is the nearest gene to rs1057233. The same locus is previously assigned as CELF1 in the 2013 IGAP GWAS. The nearest gene to rs4803758 is APOE.*
Table 2. Summary of CSF biomarker-associations of suggestive and significant AD survival-associated SNPs. Associations exceeding the multiple hypothesis-testing threshold (P < 2.27x10^{-3}) are bolded.

<table>
<thead>
<tr>
<th>SNP</th>
<th>CHR</th>
<th>Closest gene</th>
<th>Beta(\tau)</th>
<th>P(\tau)</th>
<th>Beta(\tau)</th>
<th>P(\tau)</th>
<th>Beta(p\tau)</th>
<th>P(p\tau)</th>
<th>Beta(ab\42)</th>
<th>P(ab\42)</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs2093761</td>
<td>1</td>
<td>CR1</td>
<td>-</td>
<td>&gt;0.05</td>
<td>1.51x10^{-2}</td>
<td>1.63x10^{-2}</td>
<td>-</td>
<td>&gt;0.05</td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs6431219</td>
<td>2</td>
<td>BIN1</td>
<td>-</td>
<td>&gt;0.05</td>
<td>-</td>
<td>&gt;0.05</td>
<td>-</td>
<td>&gt;0.05</td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs1057233</td>
<td>11</td>
<td>CELF1</td>
<td>-1.26x10^{-2}</td>
<td>3.85x10^{-2}</td>
<td>-1.26x10^{-2}</td>
<td>2.62x10^{-2}</td>
<td>1.53x10^{-2}</td>
<td>4.11x10^{-2}</td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs7930318</td>
<td>11</td>
<td>MS4A</td>
<td>-1.25x10^{-2}</td>
<td>3.33x10^{-2}</td>
<td>-</td>
<td>&gt;0.05</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs567075</td>
<td>11</td>
<td>PICALM</td>
<td>-1.43x10^{-2}</td>
<td>2.13x10^{-2}</td>
<td>-</td>
<td>&gt;0.05</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs9665907</td>
<td>11</td>
<td>SORL1</td>
<td>-1.70x10^{-2}</td>
<td>3.78x10^{-2}</td>
<td>-1.87x10^{-2}</td>
<td>1.50x10^{-2}</td>
<td>-</td>
<td>&gt;0.05</td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs17125944</td>
<td>14</td>
<td>FERMT2</td>
<td>2.15x10^{-2}</td>
<td>2.61x10^{-2}</td>
<td>-</td>
<td>&gt;0.05</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs4803758</td>
<td>19</td>
<td>APOE</td>
<td>1.58x10^{-2}</td>
<td>9.43x10^{-3}</td>
<td>1.96x10^{-2}</td>
<td>5.81x10^{-4}</td>
<td>-1.71x10^{-2}</td>
<td>6.75x10^{-5}</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Previously reported associated loci**

<table>
<thead>
<tr>
<th>SNP</th>
<th>CHR</th>
<th>Closest gene</th>
<th>Beta(\tau)</th>
<th>P(\tau)</th>
<th>Beta(\tau)</th>
<th>P(\tau)</th>
<th>Beta(p\tau)</th>
<th>P(p\tau)</th>
<th>Beta(ab\42)</th>
<th>P(ab\42)</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs10919252</td>
<td>1</td>
<td>C1orf112</td>
<td>-</td>
<td>&gt;0.05</td>
<td>-</td>
<td>&gt;0.05</td>
<td>-</td>
<td>&gt;0.05</td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs1532244</td>
<td>3</td>
<td>CMCI</td>
<td>-</td>
<td>&gt;0.05</td>
<td>2.23x10^{-2}</td>
<td>1.17x10^{-2}</td>
<td>-</td>
<td>&gt;0.05</td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs116341973</td>
<td>3</td>
<td>SYNPR</td>
<td>-</td>
<td>&gt;0.05</td>
<td>-</td>
<td>&gt;0.05</td>
<td>-</td>
<td>&gt;0.05</td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs71602496</td>
<td>4</td>
<td>PDE6B</td>
<td>-</td>
<td>&gt;0.05</td>
<td>-</td>
<td>&gt;0.05</td>
<td>-</td>
<td>&gt;0.05</td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs1689013</td>
<td>4</td>
<td>LINC00290</td>
<td>-</td>
<td>&gt;0.05</td>
<td>-</td>
<td>&gt;0.05</td>
<td>-</td>
<td>&gt;0.05</td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs7445192</td>
<td>5</td>
<td>PCDHA1</td>
<td>-</td>
<td>&gt;0.05</td>
<td>1.35x10^{-2}</td>
<td>1.22x10^{-2}</td>
<td>-</td>
<td>&gt;0.05</td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs12207208</td>
<td>6</td>
<td>LINC00951</td>
<td>-</td>
<td>&gt;0.05</td>
<td>-</td>
<td>&gt;0.05</td>
<td>-</td>
<td>&gt;0.05</td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs17170228</td>
<td>7</td>
<td>NT5C3A</td>
<td>-</td>
<td>&gt;0.05</td>
<td>-</td>
<td>&gt;0.05</td>
<td>-</td>
<td>&gt;0.05</td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs2725066</td>
<td>8</td>
<td>CSMD1</td>
<td>1.45x10^{-2}</td>
<td>1.70x10^{-2}</td>
<td>-</td>
<td>&gt;0.05</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs7867518</td>
<td>9</td>
<td>VLDLR</td>
<td>-1.72x10^{-2}</td>
<td>3.03x10^{-3}</td>
<td>-</td>
<td>&gt;0.05</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs1625716</td>
<td>10</td>
<td>IPMK</td>
<td>-</td>
<td>&gt;0.05</td>
<td>-</td>
<td>&gt;0.05</td>
<td>-</td>
<td>&gt;0.05</td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs1118069</td>
<td>12</td>
<td>SLC6A15</td>
<td>-</td>
<td>&gt;0.05</td>
<td>-</td>
<td>&gt;0.05</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs11074412</td>
<td>16</td>
<td>IQCK</td>
<td>-</td>
<td>&gt;0.05</td>
<td>-</td>
<td>&gt;0.05</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs5750677</td>
<td>22</td>
<td>SUN2</td>
<td>-</td>
<td>&gt;0.05</td>
<td>-</td>
<td>&gt;0.05</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Novel candidate loci**

<table>
<thead>
<tr>
<th>SNP</th>
<th>CHR</th>
<th>Closest gene</th>
<th>Beta(\tau)</th>
<th>P(\tau)</th>
<th>Beta(\tau)</th>
<th>P(\tau)</th>
<th>Beta(p\tau)</th>
<th>P(p\tau)</th>
<th>Beta(ab\42)</th>
<th>P(ab\42)</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs5750677</td>
<td>22</td>
<td>SUN2</td>
<td>-</td>
<td>&gt;0.05</td>
<td>-</td>
<td>&gt;0.05</td>
<td>-</td>
<td>&gt;0.05</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 3. Significant cis-eQTL associations of the 22 suggestive and significant AD survival-associated SNPs. Significance threshold is determined to be $2.52 \times 10^{-6}$ based on Bonferroni correction. The minor alleles are considered as the effective allele.

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

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) Rs1057233 is associated with reduced SPI1 expression in a dosage-dependent manner. (d) The mouse homolog of SPI1, Spi1, is selectively expressed in microglia and macrophages in mouse brains based on the brain RNA-Seq database. OPCs contain 5% microglial contamination. (e) SPI1 (PU.1) binds to the promoter regions of PICALM and CD33 in cells of the B-lymphoid or myeloid lineage based on ENCODE ChIP-Seq data.

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, FLAG-PU.1 1.263 ± 0.2503, shSCR 1.014 ± 0.3656, shA 0.4854 ± 0.1209, shB 0.2579 ± 0.06967, shD 0.2002 ± 0.05168. F(5,10) = 25.85, pcDNA vs FLAG-PU.1 P = 0.0049, shSCR vs shA P = 0.0120, shSCR vs shB P = 0.0002, 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. (d) Quantification of PU.1 levels in c normalized to β-Actin as a loading control. Values are presented as mean ± SD: shSCR 100 ± 0, shA 50.34 ± 9.515, shB 16.03 ± 14.72, shD 12.13 ± 10.03. F(3,3) = 87, shSCR vs shA P = 0.0120, shSCR vs shB P = 0.0026, shSCR vs shD P = 0.0023, n = 2. * P < 0.05, ** P < 0.01, *** P < 0.001, repeated measures one-way ANOVA with Sidak’s post hoc multiple comparisons test.

Figure 3. Genes regulated in BV2 microglial cells with differential expression of Spi1. (a) 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 Spi1 and its targets and other selected genes in (a), genes with altered levels after overexpression and knock-down of SPI1 in (b) and genes with variable expression in BV2 cells either with overexpression (c) or knock-down (d) of Spi1. (a) Spi1 F(2,20)=500.1: FLAG-PU.1, shB P=0.0001. Cd33 F(2,15)=37.3: shB P=0.0001. Ms4a4a F(2,13)=16.41: FLAG-PU.1 P=0.0145, shB P=0.0077. Ms4a6d F(2,11)=73.69: FLAG-PU.1 P=0.0067, shB P=0.0001. Tyrobp F(2,15)=22.83: shB P=0.0001. (b) CCL2 F(2,11)=92.5: FLAG-PU.1 P=0.0002, shB P=0.0001. Cxcl2 F(2,11)=27.73: FLAG-PU.1 P=0.0088, shB P=0.0014. Il34 F(2,11)=29.52: FLAG-PU.1 P=0.04, shB P=0.0003. Aif1 F(2,11)=34.54, FLAG-PU.1 P=0.0051, shB P=0.0006. Cldn F(2,11)=95.58: FLAG-PU.1 P=0.0043, shB P=0.0001. Pilrb F(2,11)=80.18: FLAG-PU.1 P=0.0018, shB P=0.0001. Apoe F(2,11)=30.62: FLAG-PU.1 P=0.0024, shB P=0.0004. Ctu F(2,13)=21.26: FLAG-PU.1 P=0.0039, shB P=0.0051. (c) FLAG-PU.1 vs control. Cd36 F(2,11)=13.81: P=0.0009. (d) shB vs control. Il1b F(2,13)=12.18: P=0.0025. Csf1 F(2,11)=52.19: P=0.0001. Csf1r F(2,11)=9.188:
P=0.0025. *Cx3cr1* F(2,11)=11.91; P=0.0028. *P2ryj2* F(2,13)=21.98; P=0.0001. *Pilra* F(2,11)=76.89; P=0.0001. *Itgam* F(2,13)=39.83; P=0.0001. *Axl* F(2,13)=13.49; P=0.0008. *Nos2* F(2,11)=15.54; P=0.0019. *Cox2* F(2,11)=6.717; P=0.0224. *Arg1* F(2,13)=10.31; P=0.0015. *Ctsb* F(2,13)=25.6; P=0.0001. *Nlrp3* F(2,11)=4.503; P=0.0478. *Serpinf1* F(2,11)=35.86; P=0.0001.

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.
References


