

RESEARCH ARTICLE

Tensor decomposition of stimulated monocyte and macrophage gene expression profiles identifies neurodegenerative disease-specific *trans*-eQTLs

Satesh Ramdhani^{1,2,3}, Elisa Navarro^{1,2,3}, Evan Udine^{1,2,3}, Anastasia G. Efthymiou^{1,2,3}, Brian M. Schilder^{1,2,3}, Madison Parks^{1,2,3}, Alison Goate^{1,2,3}, Towfique Raj^{1,2,3*}

1 Ronald M. Loeb Center for Alzheimer's Disease, Icahn School of Medicine at Mount Sinai, New York, New York, United States of America, **2** Nash Family Department of Neuroscience and Friedman Brain Institute, Icahn School of Medicine at Mount Sinai, New York, New York, United States of America, **3** Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, New York, United States of America

* towfique.raj@mssm.edu



OPEN ACCESS

Citation: Ramdhani S, Navarro E, Udine E, Efthymiou AG, Schilder BM, Parks M, et al. (2020) Tensor decomposition of stimulated monocyte and macrophage gene expression profiles identifies neurodegenerative disease-specific *trans*-eQTLs. *PLoS Genet* 16(2): e1008549. <https://doi.org/10.1371/journal.pgen.1008549>

Editor: Gregory S. Barsh, Stanford University School of Medicine, UNITED STATES

Received: June 11, 2019

Accepted: December 2, 2019

Published: February 3, 2020

Copyright: © 2020 Ramdhani et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All relevant data are within the manuscript and its Supporting Information files.

Funding: T.R. is supported by grants from the NIH National Institute on Aging (R01AG054005), the Alzheimer's Association, and the Michael J Fox Foundation. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Abstract

Recent human genetic studies suggest that cells of the innate immune system have a primary role in the pathogenesis of neurodegenerative diseases. However, the results from these studies often do not elucidate how the genetic variants affect the biology of these cells to modulate disease risk. Here, we applied a tensor decomposition method to uncover disease associated gene networks linked to distal genetic variation in stimulated human monocyte and macrophage gene expression profiles. We report robust evidence that some disease associated genetic variants affect the expression of multiple genes in *trans*. These include a Parkinson's disease locus influencing the expression of genes mediated by a protease that controls lysosomal function, and Alzheimer's disease loci influencing the expression of genes involved in type 1 interferon signaling, myeloid phagocytosis, and complement cascade pathways. Overall, we uncover gene networks in induced innate immune cells linked to disease associated genetic variants, which may help elucidate the underlying biology of disease.

Author summary

A steadily growing number of studies have identified and characterized *cis* expression quantitative trait loci (eQTLs) in human primary cells and tissues. However, identifying distal regulation on gene expression (*trans*-eQTLs) is far more difficult to detect due to smaller effect sizes and the large number of tests for thousands of transcripts. Here we applied a novel method to uncover gene networks linked to distal genetic variation (*trans*-eQTLs) in gene expression datasets from peripheral monocytes at baseline and stimulated with interferon (IFN) or lipopolysaccharide (LPS). We report robust evidence that some neurodegenerative disease-associated variants affect the expression of multiple genes in

Competing interests: AG served on the scientific advisory board for Denali Therapeutics from 2015-2018 and has served as a consultant to AbbVie, Biogen, Eisai, Illumina, and GSK. Other authors declare no competing financial interests.

trans. These innate immune networks illuminate potential biological mechanisms underlying Alzheimer's and Parkinson's disease. These findings provide new insights into understanding the effect of genetic variants on gene networks contributing to disease, and may help in elucidating the molecular mechanisms of neurodegenerative diseases.

Introduction

Genome-wide association studies (GWAS) have identified tens of thousands of common genetic variants associated with complex traits and diseases [1, 2]. The vast majority of these genetic variants reside in non-coding regions [3] potentially affecting the regulation of gene expression of local (*cis*) or distal (*trans*) genes. An extensive number of studies have characterized expression Quantitative Trait Loci (eQTL) across multiple tissues [4–6] and cell types [7–10] and in response to environmental stimuli [11, 12]. Nevertheless, most eQTL studies in humans have been limited almost exclusively to *cis*-effects. The few studies that have investigated *trans*-eQTLs have identified few significant associations, and there has been little replication of significant associations across data sets. One limitation is that most of these studies used complex mixtures of different cell types (e.g., brain tissue, whole blood or peripheral blood mononuclear cells [PBMCs]) as sources of RNA [13], which may result in failure to properly capture the activity of genetic variants in disease-relevant cell types. Another limitation is that these studies report *trans*-eQTLs by performing millions of Single Nucleotide Polymorphism (SNP)-by-gene tests, which can result in few significant associations due to the very stringent significance threshold imposed by multiple testing correction. Such SNP-by-gene *trans*-eQTL mapping also ignores the complex structure of gene networks. Recently, Hore *et al.* [14] developed a method to decompose a tensor (or multi-dimensional array) of multi-tissue gene expression data to uncover gene networks and map these networks with genetic variants to detect *trans*-eQTLs. This and similar methods [15–17] have been successful in mapping networks of genes regulated by genetic variants that would not have been uncovered via marginal SNP-by-gene *trans*-eQTL analysis.

Discovering such *trans*-eQTL networks may improve our understanding of the biological mechanisms underlying polygenic diseases such as Alzheimer's disease (AD) and Parkinson's disease (PD). Both are neurodegenerative diseases that share pathological hallmarks such as neuronal loss, proteinopathy, mitochondrial dysfunction and reactive microglia (the innate immune cells of the brain) [18, 19], although the specific vulnerable cells differ between disorders. GWAS have identified over 30 and 78 loci, respectively, associated with Alzheimer's disease [20–22] and Parkinson's disease [23, 24]. Nevertheless, these studies open new questions as they identify SNPs rather than genes, most of them localized in non-coding regions that are thought to modulate disease risk by influencing the expression of nearby or distal genes in a cell-type specific manner. Therefore, translating the effect of risk alleles on the genes and cell types involved in disease pathogenesis remains a challenge. Along this line, it should be noted that *cis*-eQTL analyses have implicated some of the Alzheimer's and Parkinson's disease susceptibility genes in myeloid cell function, whose expression, relative to each risk allele, is altered in the innate immune cells [7, 25]. Animal models have also pointed to an altered innate immune system as a potential driver of these diseases [26–28]. While most research has focused on the contribution of microglia in neurodegeneration, there are reports suggesting a causal role of cellular and humoral components of the peripheral innate immune system in the pathogenesis and progression of these diseases. For example, parabiosis studies suggest that peripheral blood components may influence Alzheimer's disease progression [29]. Whether peripheral blood

monocytes may themselves be drivers of disease or are simply useful proxies for the infiltrating macrophages and/or resident microglia found at the sites of neuropathology is still unknown. We hypothesize that Alzheimer's and Parkinson's disease risk alleles may modulate disease susceptibility by regulating the expression of a distal gene or set of genes in monocytes and macrophages (major cellular components of the innate immune system).

Here we use gene expression profiles from monocytes and macrophages to identify novel *trans*-eQTLs and replicate those previously known associations. We show that about one-third of *trans* regulation is significantly mediated by the expression of *cis* genes. We identified *trans*-eQTLs that colocalize with disease-associated susceptibility loci. These include GWAS loci for coronary artery disease, body mass index, and cholesterol, all traits for which peripheral blood monocytes and macrophages are functionally relevant. More interestingly, our analysis also discovered biologically meaningful *trans*-eQTL networks for neurodegenerative diseases including genes in interferon and complement signaling, as well as lysosomal function linked to Alzheimer's and Parkinson's disease susceptibility loci, respectively. These results highlight context-specific *trans*-eQTL networks in support of a role for myeloid cells of the innate immune system as key modifiers of neurodegenerative disease risk.

Results

Identification of gene networks in stimulated monocytes and macrophages

We used Sparse Decomposition of Arrays (SDA) [14] to construct gene networks by decomposing a multi-array set of gene expression measurements into latent components (see [Methods](#)). Each of these components consists of scores that indicate the relative contribution of each individual, gene, and cell or stimulation activity scores to gene networks ([Fig 1](#)). The individual scores are the magnitude of the effect for each component across individuals. We use the individual scores as phenotypes in genome-wide *trans*-eQTL analysis to identify common genetic variants that drive each component. The gene scores (or gene loadings) allow inference of the genes involved in each component. SDA is developed in a Bayesian framework and uses 'spike and slab' prior [14] to allow gene scores of each component to have a unique level of sparsity. Finally, the cell or stimulation specificity scores indicate the activity of the component for each inflammatory stimuli or cell type.

We used gene expression profiles from two previously published studies: (1) Fairfax *et al.* [12] (FF), in which CD14⁺ human monocytes were profiled with two inflammatory stimuli: naïve (CD14), and in response to either interferon- γ (IFN), or lipopolysaccharide at 2 hours (LPS2), and 24 hours (LPS24); and (2) primary human monocytes and macrophages from the Cardiogenics Consortium (CG) [30], which includes subjects from a Cardiovascular disease cohort ([Fig 1](#)). The IFN- γ and LPS stimuli were used to elicit an immune reactive phenotype in these cells that may help detect gene networks not identified in the naïve state and better approximate the state of these cells in the context of a degenerating brain. Furthermore, the monocyte and macrophage comparison may help uncover different pathogenic effects of these two cell types. We computed a maximum of 500 components in each dataset, of which the majority were non-sparse (most genes with low or zero gene scores). These non-sparse components capture technical or biological confounders in the expression data ([Fig 2A](#) and [S1 Fig](#)).

Using a ranked-based sparsity statistic (see [Methods](#)), we identified 56 and 111 components in FF and CG, respectively, that are sparse, or components with distinguishable non-zero and high scoring genes ([S1](#) and [S2](#) Tables). In the FF dataset, we identified 12 sparse components active in naïve, 5–9 active in response to IFN- γ , LPS 2hrs, or LPS 24hrs, and the remaining 25 shared across all conditions ([Fig 2B](#)). In CG, 42 sparse components were specific to monocytes and 45 specific to macrophages and 24 were shared between the two cell types ([S1 Fig](#)). Using a

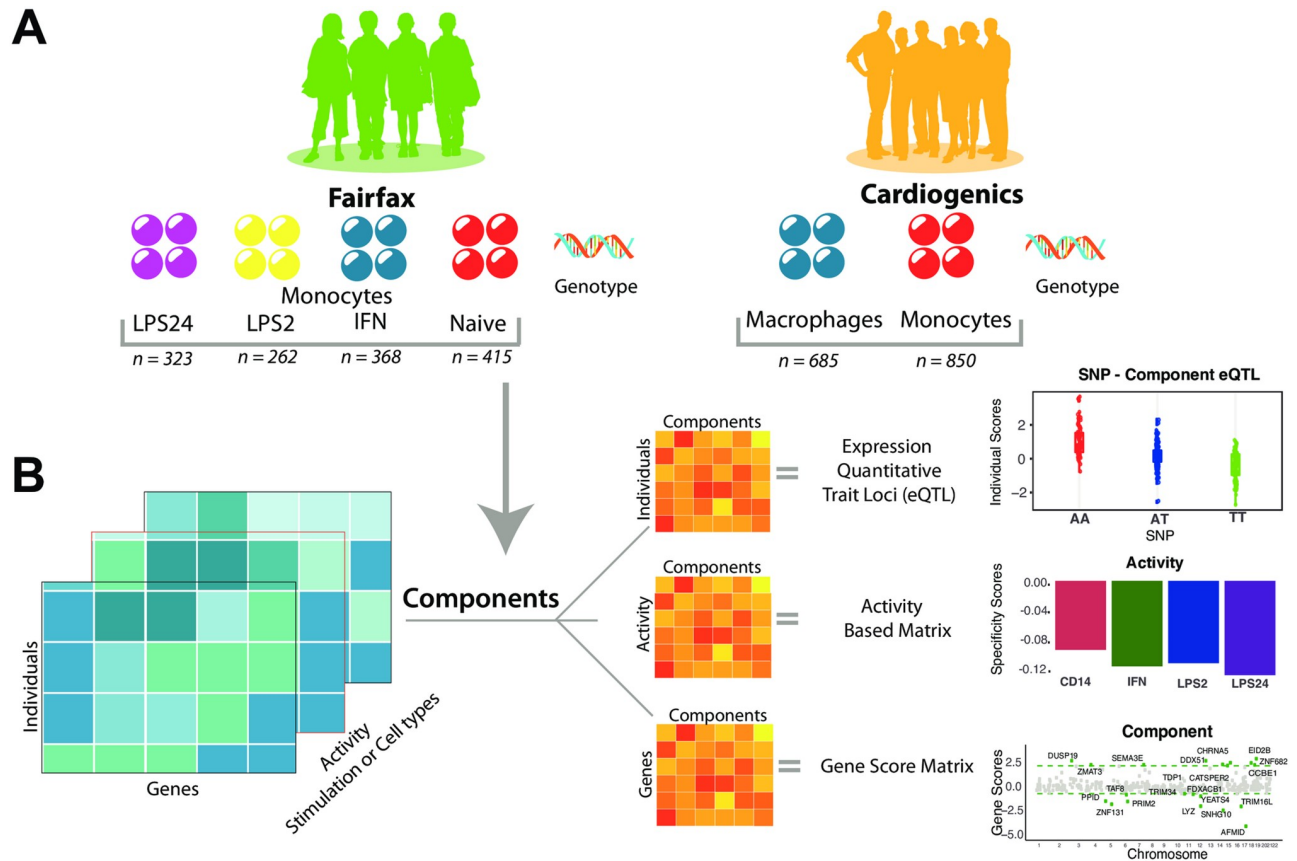


Fig 1. Overview of the study design and method. (A) The gene expression datasets used in this study. Stimulated monocyte gene expression profiles from Fairfax *et al.* [12] in four conditions: response to lipopolysaccharide at 24 hours (LPS24) and at 2 hours (LPS2), interferon- γ (IFN- γ), and naive (left panel). Peripheral blood monocytes (MP) and macrophages (MC) from the Cardiogenics Consortium (right panel). (B) Overview of the SDA approach. An illustration of decomposition of gene expression datasets to yield component vectors for relative contribution of each individual, gene and condition. The individual scores matrices are then used as phenotypes with SNP genotypes in order to identify genetic variation correlated with the components (top). The stimulation or cell-activity scores matrix is used to identify the contribution of each condition for the components (middle). The gene scores matrix is used to identify the contribution of each gene within the components (bottom).

<https://doi.org/10.1371/journal.pgen.1008549.g001>

two-way reverse correlation approach, we found 7 of the 56 sparse components in FF are conserved between FF and CG (S3 Table). An example of a shared component is shown in Fig 2C.

To assess the functional significance of the components, we tested the genes within each component for enrichment in Gene Ontology (GO) biological processes. We found significant enrichment for GO biological processes for 30 out of 56 sparse components, suggesting that the genes in each sparse component are part of coherent biological processes (S4 Table). To assess the enrichment of disease risk loci in these components, GWAS summary statistics for 18 traits and gene-sets from the sparse components were used as input for MAGMA [31]. We found three components significantly enriched for genes in GWAS loci at a Bonferroni-corrected significance threshold and four additional components enriched at a nominal P -value < 0.05 for ten traits (S5 Table). This includes a component with genes in the interferon-signaling pathway, which is enriched for genes in Alzheimer's disease susceptibility loci. Overall, these results suggest that the sparse components from stimulated monocyte gene expression data have functional significance and, in some cases, may underlie disease-relevant biological processes.

We next reasoned that the monocyte components might contain genes in disease-associated loci that disproportionately contribute to the heritability of complex diseases. We used LD score

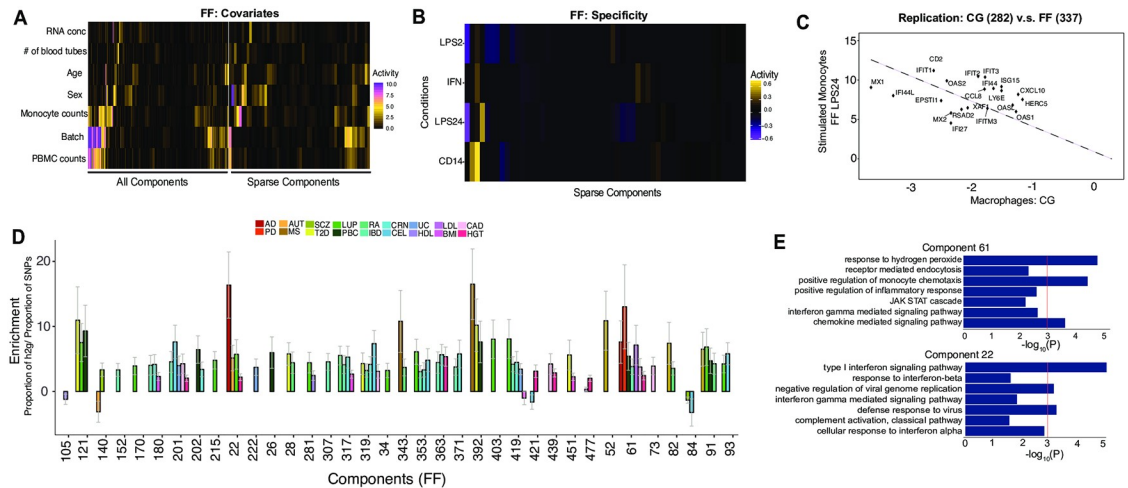


Fig 2. Discovery and reproducibility of sparse components enriched for disease heritability. (A) Heat map of known covariates and correlation with individual scores from each of the 500 components (left) and 56 sparse components (right) in FF data. (B) Heat map of condition specificity scores for the sparse components in FF data. Each row is a stimulation (naïve, LPS, or IFN) and each column is a sparse component. (C) Component replication between component 282 (CG) and component 337 (FF LPS24) for the most highly scored genes. The gene scores from the two components (in CG and FF) are highly correlated. (D) Proportion of heritability for 18 selected complex traits that can be attributed to each sparse component from the FF data. Shown here are enrichment statistics (with standard error) comparing the proportion of SNP heritability within the components divided by the proportion of total SNPs represented at FDR-corrected $P < 0.05$. (E) Gene Ontology (GO) enrichment for genes in component 61 and 22. AD: Alzheimer’s disease; PD: Parkinson’s disease; AUT: Autism; MS: Multiple Sclerosis; SCZ: Schizophrenia; T2D: Type 2 Diabetes; LUP: Lupus; PBC: Primary Biliary Cirrhosis; RA: Rheumatoid Arthritis; IBD: Inflammatory Bowel Disease; CRN: Crohn’s Disease; CEL: Celiac Disease; UC: Ulcerative Colitis; HDL: High-density lipoprotein Cholesterol; LDL: Low-density lipoprotein Cholesterol; BMI: Body Mass Index; CAD: Coronary Artery Disease; HGT: Height.

<https://doi.org/10.1371/journal.pgen.1008549.g002>

regression (LDSC) [32] to partition GWAS heritability into the contribution of SNPs located within genes from each component. We found that 36 out of 56 sparse components in FF showed significant enrichment (FDR-corrected $P < 0.05$) for 18 different traits. This includes a component in FF data (component 61) significantly enriched for SNP-based heritability for seven traits including Alzheimer’s (7.6-fold, $P < 0.03$), Parkinson’s (13-fold, $P < 0.01$) as well as a number of inflammatory and metabolic diseases (Fig 2D; S6 Table). Component 392 is enriched for three traits including Multiple sclerosis (16.5-fold, $P < 0.002$), Type 2 diabetes (10.2-fold, $P < 0.02$) and Primary biliary cirrhosis (7.6-fold, $P < 0.03$). Another component (22) is significantly enriched for SNP-based heritability for Alzheimer’s disease (16.4-fold, $P < 0.04$), autism (5.1-fold, $P < 0.03$), and lupus (5.7-fold $P < 0.03$), accounting for 6%, 2% and 2% of SNP-based heritability, respectively. Component 22 is highly enriched for a number of genes in the Type 1 interferon-signaling, defense response to virus, and complement activation pathways (Fig 2E). A number of studies have linked genes from these functional categories to Alzheimer’s disease [33–35], autism [36, 37] and lupus [13]. Interestingly, we discovered that this component is also a *trans*-eQTL for Alzheimer’s disease susceptibility loci (see below). Taken together, these results illustrate that several components identified in this study can explain a substantial amount of SNP-based genetic heritability for many common diseases and begin to implicate specific innate immune function for common variants across neurodegenerative and inflammatory diseases.

Detection of context-specific *trans*-eQTLs

To detect *trans*-eQTLs, the individual scores from each component are tested for association with genotype dosages (autosomal SNPs, Minor Allele Frequency [MAF] > 0.01 , imputed with the Haplotype Reference Consortium reference panel) across the genome. A stringent

quality control (QC) procedure was used to filter out multi-mapped gene probes and non-coding genes prior to *trans*-eQTL analysis (Methods). For clarification, hereinafter, we define the following: a *trans*-eQTL to be an SNP that maps to a component, the relevant SNP itself to be the *trans*-eSNP and any genes within the component to be *trans*-eGene(s). In FF, we identified a total of 15,856 *trans*-eQTLs (FDR < 0.05) representing 15,298 unique SNPs (833 independent SNPs; obtained by pruning SNPs in LD with $r^2 < 0.2$) and 55 unique sparse components (S7 Table). Of the significant *trans*-eQTLs, approximately 55% are stimuli-specific, detected only in response to either IFN- γ or LPS challenge. For CG, we identified a total of 3,945 *trans*-eQTLs (FDR < 0.05) representing 3,217 unique SNPs (70 independent SNPs; obtained by pruning SNPs in LD with $r^2 < 0.2$) and 57 unique sparse components (S8 Table). In each dataset, the components contain anywhere between 5–62 genes with distinguishable non-zero gene scores based on distributional cut-offs (Methods). Our analysis confirmed previously published *trans*-eQTL: a *cis*-eQTL at rs2275888 for *IFNB1* is associated with the expression of 17 genes in *trans* after 24-hour LPS stimulation, many of which are interferon (IFN- β) response genes (S2 Fig).

Trans-eQTLs in stimulated monocytes as putative drivers of disease associations

To identify disease risk alleles influencing the expression of the distal gene(s), we investigated whether the *trans*-eSNPs were previously associated with complex traits or diseases. We incorporated 20,094 SNPs from the NHGRI GWAS Catalog pertaining to 1,374 disease or complex traits that were significant at a Bonferroni-corrected significance threshold of $P < 5 \times 10^{-8}$. We identified 227 *trans*-eQTLs mapping to 29 sparse components that overlap with disease or trait-associated loci (FDR < 0.15; S9 Table). We used a more liberal FDR significance level (0.15) to identify trait-associated *trans*-eQTLs since this class of SNPs has already passed significance thresholds for a different study.

In addition, our FDR correction is applied to testing for genome-wide SNPs rather than just correcting for only the trait-associated SNPs as has been previously done [13]. Nevertheless, to ensure the robustness of the *trans*-eQTLs, we carried out a permutation scheme and found 89 *trans*-eQTLs mapping to 18 components that were significant at permutation threshold $P < 1 \times 10^{-3}$. Fig 3A provides a subset of the *trans*-eQTLs that are in disease or trait-associated loci. In CG, we found 56 *trans*-eQTLs mapping to 14 sparse components overlapping loci associated with a disease or a complex trait at FDR < 0.15 (S10 Table). Examples of colocalized *trans*-eQTLs with Parkinson's and Alzheimer's disease associated susceptibility loci are shown in Fig 3B. We found components that mapped to multiple independent loci and components that mapped to a single locus with multiple disease-associated variants. For example, disease-associated susceptibility alleles in the MHC class II are significantly associated with individual scores of component 363 in FF, which is active in response to all stimuli (S3 Fig). The component contains several MHC class II genes but also 21 non-MHC genes (e.g., *DEF8* and *AOAH*) with distinguishable non-zero scores. Some of the *trans*-associations to MHC class II alleles have been previously identified in the SNP-by-gene analysis [12], but our analysis identified a network of 21 genes pointing to a critical biological pathway underlying disease susceptibility. We identified several other examples of *trans*-eQTLs in disease-associated loci (S4–S6 Figs). Overall, we were able to not only reproduce previously identified *trans*-eQTLs and individual target genes in monocytes and macrophages but also identify biologically informative gene networks linked to distal genetic variants.

Of the disease-associated *trans*-eSNPs (223 in FF and 43 in CG), 85 and 27 in FF and CG, respectively, were also *cis*-eSNPs for nearby genes. We applied a Mendelian randomization

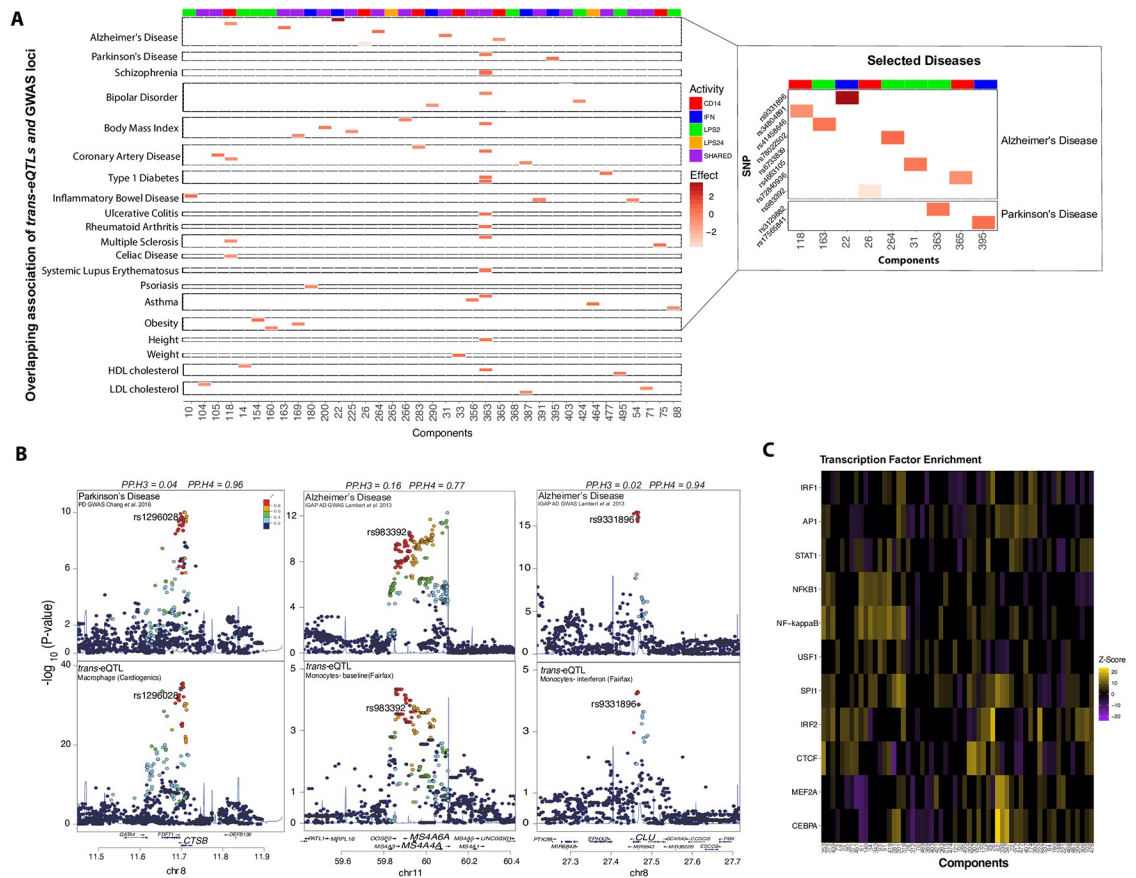


Fig 3. Trans-eQTLs colocalized in disease or trait-associated GWAS loci. (A) Significant trans-eQTLs in FF data (FDR < 0.15) in 20 selected disease or trait-associated GWAS loci (left panel). Shown are diseases-associated trans-eSNP on Y-axis and trans-eQTL components on X-axis. The red colored boxes reflect the effect size for the trans-eQTLs while the horizontal colored header reflects the condition activity scores. Trans-eQTLs that are in Alzheimer's or Parkinson's disease-associated loci (right panel). Alzheimer's disease includes GWAS susceptibility loci from Alzheimer's related traits including the age of onset, age-related cognitive decline, and APOE $\epsilon 4$ carriers. (B) Colocalization of trans-eQTLs at Parkinson's disease susceptibility locus *CTSB* (left panel) and Alzheimer's disease-associated loci *MS4A4A* (middle panel) and *CLU* (right panel). The x-axis in each panel shows the physical position on the chromosome (Mb). The y-axis shows the $-\log_{10}(P)$ association p-values for Parkinson's disease [23–24] (left panel) and Alzheimer's disease GWAS [20–22] (middle and right panels). Listed on top are 'coloc' posterior probability for hypothesis 3 (PP.H3) and 4 (PP.H4). PP.H3: Association with eQTL and GWAS, one shared SNP. PP.H4: Association with eQTL and GWAS, two independent causal SNPs. (C) Transcription factors whose binding sites occurrence is enriched in the target set of genes within each sparse component compared to the expected occurrence estimated from a background set.

<https://doi.org/10.1371/journal.pgen.1008549.g003>

method [38] to quantify support for a direct relationship between each SNP that has a cis-eQTL to their trans-eGene(s). We applied Mendelian randomization to test if the trans-eSNP mediated the effect on the trans-eGene(s) through their respective cis-gene using two different approaches. First, we used the individual scores from SDA for the 56 sparse components in the mediation analysis. At a nominal $P < 0.05$, we found 35 and 15 in FF and CG, respectively, where the cis-gene is a significant mediator of the component individual scores. In the second analysis, we directly tested for mediation using the expression level of each trans-eGene within each respective component. From this directed approach, we found a significant mediation effect for at least one trans-eGene for 79 and 21 in FF and CG trans-eSNPs, respectively. Of the significant trans mediators, we found that only 11 genes were transcription factors (TF), suggesting that TFs are not primary regulators of the trans network. However, we observed regulatory motifs for the same TF enriched among the all trans-eGenes within each network

(Fig 3C). These include many myeloid lineage-specific factors (e.g., *PU.1*, *C/EBPα* and β , and *MEF2A*) and interferon-related (e.g., *IRF1*, *IRF2*, and *STAT1*) factors (Fig 3C). In summary, our results are consistent with other findings that TFs are not the primary driver of trans-eQTL networks but are enriched for specific TF binding motifs, suggesting that the trans-eGenes are under the same regulatory control.

Macrophage-specific trans-eQTL at Parkinson’s disease susceptibility loci

Considering that GWAS have identified loci with genes related to the innate immune system both in Alzheimer’s disease and Parkinson’s disease, we hypothesize that analysis of monocytes and macrophages could reveal new gene-sets associated with disease susceptibility. In support of this hypothesis, we observed in CG macrophages that the rs1296028, one of 41 PD-associated loci interrogated in this analysis [23], affects the expression of the lysosomal protease Cathepsin B (*CTSB*) ($P = 1.46 \times 10^{-22}$; FDR = 1.31×10^{-18}) in cis [39] (S7 Fig) and the expression of 16 genes in trans ($P = 8.29 \times 10^{-35}$; FDR = 9.48×10^{-28}) (Fig 4A and 4B). The lead trans-eSNP rs1296028 (MAF = 0.11) at the *CTSB* locus is associated with Parkinson’s disease susceptibility (in LD with GWAS lead SNP rs2740594; $r^2 = 0.68$) [23]. Both the trans-eQTL and the disease association signal were consistent with a model for shared causal variants (as indicated by coloc [40] posterior probability PP3+PP4 = 0.99, PP4/PP3 = 24) (Fig 3B). Although we did not find significant enrichment of any specific biological pathways with the Ingenuity Pathway Analysis (IPA) tool, we found enrichment for biological processes such as ‘lysosomal pathway’ ($P = 5.2 \times 10^{-3}$) and ‘cholesterol degradation’ ($P = 6.22 \times 10^{-4}$) when we lowered the posterior inclusion probability (PIP) from 0.5 (default) to 0.2 for gene scores. The directions of effect for all the genes in this network were consistent: the minor allele (rs1296028-G) is associated with decreased expression of all 16 genes (Fig 4A; S8 and S9 Figs). Using a small independent macrophage eQTL dataset

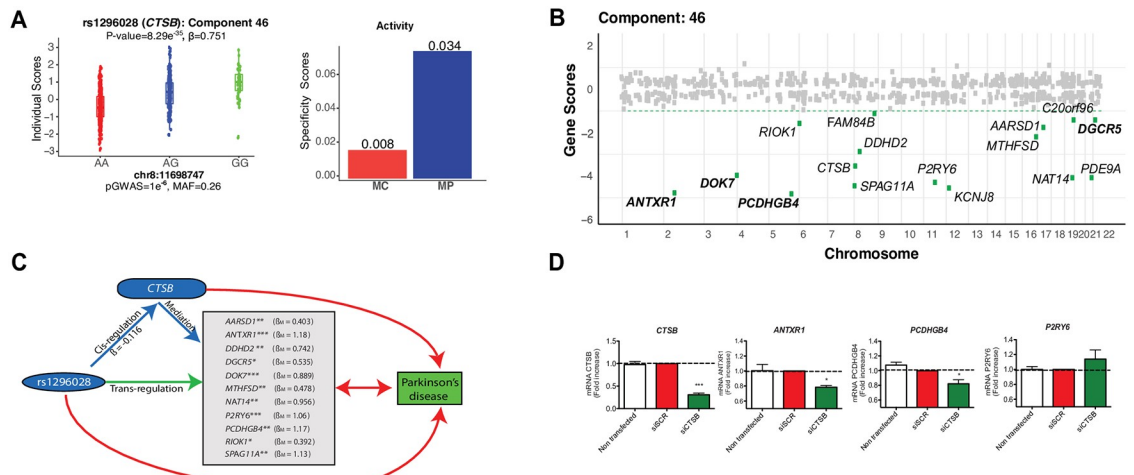


Fig 4. Macrophage-specific trans-eQTL colocalized in Parkinson’s disease associated *CTSB* locus. (A) The genotype for Parkinson’s disease susceptibility allele rs1296028 is significantly associated with the individual scores of CG component 46. The rs1296028 affects the expression of 16 genes in component 46 in macrophage (CG) (left panel). Tissue specificity scores suggest CG component 46 is active in macrophage (Right). P-value: ***: $1e-50$ | **: $1e-10$ | *: $1e-5$. (B) Gene scores for component 46 across the genome. Only the trans-eGenes with PIP > 0.5 and 2.5% distributional cut-off (green dotted line) are shown. Trans-eQTL associations that replicated (FDR < 0.20) in an independent STARNET macrophage dataset are shown in bold (C) Parkinson’s disease SNP rs1296028 mediates trans-effects to 11 genes through cis-mediator cathepsin B (*CTSB*). The beta coefficients from Mendelian randomization analysis are shown for the significant trans-eGenes. (D) Experimental validation using THP-1 derived macrophages. *CTSB* was knocked-down using siRNA during 48 h and the levels of the top-scoring genes in the component were measured by qPCR. Data was normalized against scramble siRNA (SCR). P-value: *: 0.05 | ***: 0.001.

<https://doi.org/10.1371/journal.pgen.1008549.g004>

(STARNET[5]; n = 82), we were able to replicate 4 of the 16 *trans*-eQTL associations (FDR < 0.20; S10 Fig). The direction of effect is consistent across the two datasets: rs1296028-G is associated with increased expression of the target genes as well as decreased risk for PD.

Using Mendelian randomization, we found a mediation effect for rs1296028 through *CTSB* for 11 of the genes in this network, of which the most significant mediation effects were for *ANTXR1*, *PCDHGB4*, *SPAG11A* and *P2RY6* (Fig 4C). To examine if the regulation of the *trans* target genes depended on *CTSB*, we performed siRNA-mediated knockdown of *CTSB* and measured the effect on the four previously mentioned genes that have $\beta_M > 1$ using human THP-1 derived macrophages. We observed a concomitant reduction in *ANTXR1* and *PCDHGB4* mRNA levels ($P < 0.05$) as measured by qRT-PCR (Fig 4D). A non-significant effect was observed in *P2RY6*, whereas *SPAG11A* is not expressed in the THP-1 cell line. The knockdown experiment provides an independent validation for some of the genes in the *CTSB* *trans*-eQTL network. However, additional experimental validation is necessary to fully elucidate the role of *CTSB* lysosomal network in the etiology of Parkinson’s disease.

Alzheimer’s disease loci affect the expression of genes in Type 1 interferon signaling pathway

Following the same rationale, we attempted to find altered monocyte and macrophage gene networks associated with Alzheimer’s disease susceptibility alleles. We found eight *trans*-eQTL signals that colocalize with Alzheimer’s disease (or Alzheimer’s disease-related phenotypes) susceptibility loci (Fig 3A), and we highlight two examples (Figs 5 and 6). First, we observed that rs983392 on chromosome 11q12, reported to be associated with Alzheimer’s disease ($MAF = 0.41$, $P_{GWAS} = 6 \times 10^{-16}$), is a *trans*-eQTL to a component with 54 genes of distinguishable non-zero loading scores ($P = 1.5 \times 10^{-4}$; FDR < 0.15) (Fig 5A and 5B).

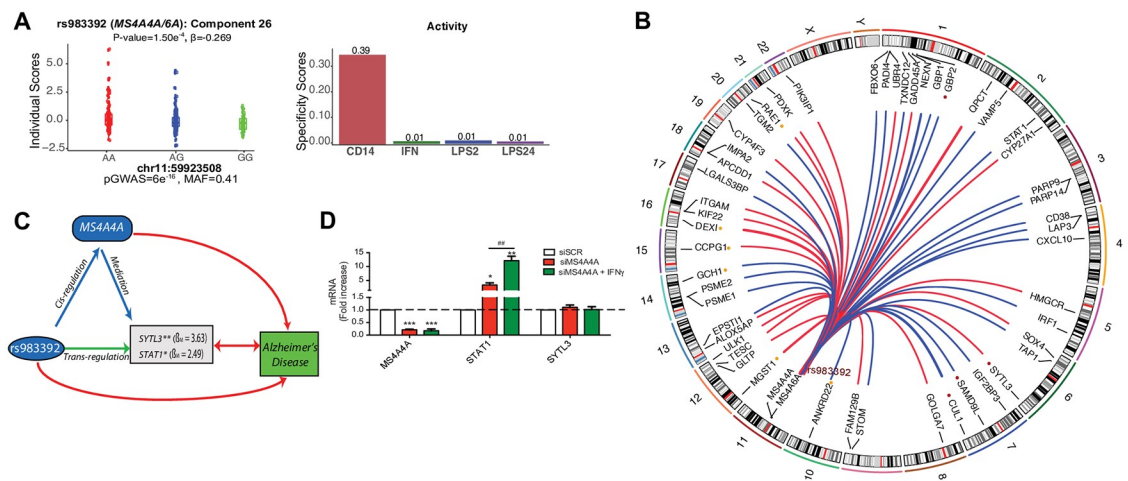


Fig 5. *Trans*-eQTL colocalized in Alzheimer’s disease associated *MS4A* locus. (A) The genotypes for Alzheimer’s disease susceptibility allele rs983392 are significantly associated with the individual scores of component 26. rs983392 is *trans*-eQTL to FF component 26 with 54 genes (left panel). The component 26 is active only at baseline (right panel). (B) Circular plot demonstrating the chromosomal position of *trans*-eSNP (rs983392) and the 54 *trans* target genes. The minor and AD-protective allele rs983392-G is associated with decreased (blue lines) and increased (red lines) expression of *trans*-eGenes. The colored dots are *trans*-eQTL that replicates in ImmVar baseline monocytes (red and yellow color dots denote *trans*-eQTL association at FDR < 0.05 and 0.20, respectively). Only the *trans*-eGenes with PIP > 0.5 and 2.5% distributional cut-off are shown. (C) Alzheimer’s disease SNP rs983392 mediates *trans*-effects to two *trans*-eGenes through *cis*-mediator *MS4A4A*. (D) Experimental validation using THP-1 derived macrophages. *MS4A4A* was knock-down using siRNA for 48 h, IFN γ 20 ng/ml was added during the last 24 h. The levels of the top-scoring genes in the component were measured by qPCR. Data was normalized against scramble siRNA (SCR). P-value: *: < 0.05 | **: < 0.01 | ***: < 0.001 vs SCR. #: < 0.01 vs siMS4A4A. (N = 5 independent experiments).

<https://doi.org/10.1371/journal.pgen.1008549.g005>

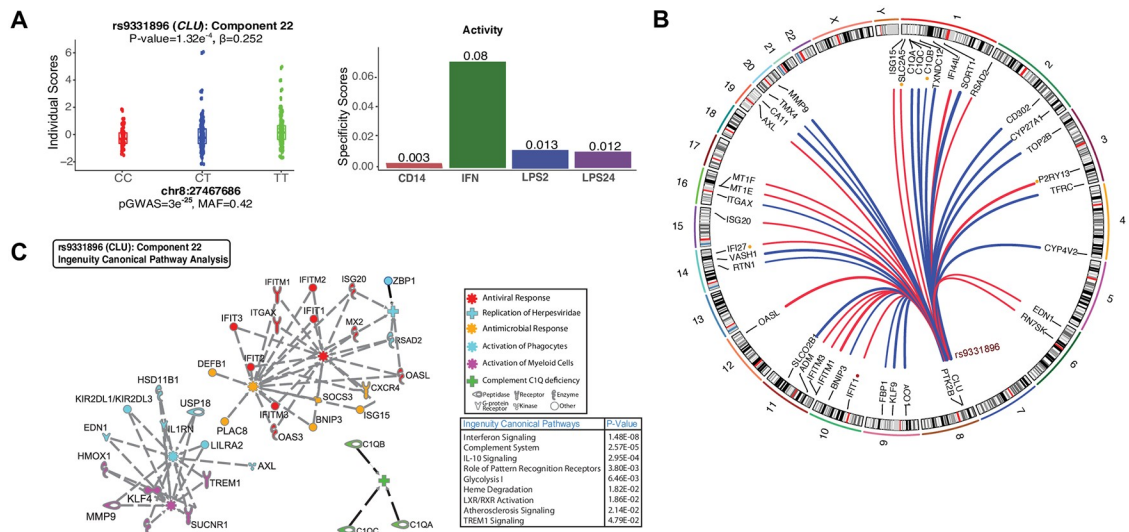


Fig 6. Trans-eQTL colocalized in Alzheimer's disease associated CLU locus. (A) The genotypes for Alzheimer's disease susceptibility allele rs9331896 are significantly associated with the individual scores of component 22 (left panel). The component is active in monocytes (FF) in response to interferon- γ (right panel). (B). Circular plot demonstrating the chromosomal position of *trans*-eSNP (rs9331896) and the 38 *trans* target genes. The minor and AD-protective allele rs9331896-C is associated with decreased (blue lines) and increased (red lines) expression of *trans*-eGenes. The colored dots are *trans*-eQTL that replicates in ImmVar IFN stimulated monocytes (red and yellow color dots denote *trans*-eQTL association at FDR < 0.05 and 0.20, respectively). Only the *trans*-eGenes with PIP > 0.5 and 2.5% distributional cut-off are shown. (C) IPA pathway analysis of the *trans*-eGenes in this component. The IPA canonical pathway enrichment P-values are shown in the lower right. The pathway network is grouped by IPA biological functions. The functional groups are defined by different colors and symbols (top right).

<https://doi.org/10.1371/journal.pgen.1008549.g006>

The component is active in the naïve condition, and stimulation with IFN- γ or LPS ablates the *trans*-eQTL signal. A significant *cis*-eQTL effect for rs983392 to both *MS4A4A* (FDR = 2×10^{-3} , $\beta = -4.4$) and *MS4A6A* (FDR = 5×10^{-4} , $\beta = -4.7$) is observed only in the naïve condition (S11 Fig). Both *trans*-eQTL and disease association signals were consistent with a model for shared causal variants (as indicated by coloc [40] posterior probability PP3+PP4 = 0.93, PP4/PP3 = 4.8) (Fig 3B). The top-scoring genes in this component encode for Interferon-Inducible Guanylate Binding Proteins (*GBP1*, *GBP2*, *GBP4*, and *GBP5*), followed by *STAT1* and *IRF1*, which are key TFs for IFN- γ activation and type 1 interferon signaling (Fig 5B and S12–S14 Figs). The genes in this component are enriched for IPA biological processes such as 'interferon signaling' ($P = 5.77 \times 10^{-16}$) and the 'antigen presentation' pathway ($P = 6.62 \times 10^{-9}$). IPA also identified enrichment for several annotated biological functions including 'antiviral response' ($P = 1.26 \times 10^{-33}$). To determine if the observed *trans*-eQTL associations are mediated by expression of either *MS4A4A* or *MS4A6A* in *cis*, we performed Mendelian randomization-based mediation using both component individual scores and gene expression levels. We observed *trans*-regulatory effects of rs983392 on *STAT1* and *SYTL3* mediated through *MS4A4A* but not through *MS4A6A* (Fig 5C). To examine if the regulation of the *trans* target genes *STAT1* and *SYTL3* depended on *MS4A4A* we performed siRNA-mediated knockdown of *MS4A4A* using human THP-1 derived macrophages. We observed an increase in *STAT1* mRNA levels with slightly higher fold change after IFN stimulation ($P < 0.01$) compared to baseline ($P < 0.05$) (Fig 5D). A non-significant effect was observed in *SYTL3* in the THP-1 cell line.

Another Alzheimer's disease-associated variant rs9331896-C ($MAF = 0.41$, $P_{GWAS} = 3 \times 10^{-25}$) located on chromosome 8p21 within intron 2 of the *CLU* gene is associated with the individual scores of a component consisting of 38 genes with non-zero scores ($P = 1.32 \times 10^{-4}$; FDR < 0.15) (Fig 6A and 6B). This component is active in IFN- γ stimulated monocytes (Fig 6B). Both

trans-eQTL and disease association signal were consistent with a model for shared causal variants (as indicated by coloc [40] posterior probability $PP3+PP4 = 0.96$, $PP4/PP3 = 47$) (Fig 3B). The SNP rs9331896-C had no *cis*-eQTL effect on *CLU* in monocytes (or in CG macrophage) but we have reported previously a splicing QTL for *CLU* in *DLPFC* [41]. Given these results, it is likely that the *cis* effect that we observed could mediate the *trans*-effect through the expression of specific isoform. The top-scoring genes in the network include the members of Interferon Induced Transmembrane Proteins (*IFITM1* and *IFITM3*) followed by *MT1E*, *MT1X*, and *MT1G* and *OASL*, all essential proteins involved in the innate immune response to viral infection (Fig 6B; S15–S17 Figs). Using IPA we found significant enrichment for ‘interferon signaling’ ($P < 1.48 \times 10^{-8}$) and ‘complement cascade’ pathways ($P < 2.57 \times 10^{-5}$) (Fig 6C). This component also contained genes in the complement components C1q family of genes (*CIQA*, *CIQB*, and *CIQC*). Interestingly, protein products of *CLU* (previously known as CLI for complement lysis inhibitor) and complement component genes physically interact and are part of a protein-protein interaction (PPI) network (S18 Fig), providing an independent validation of our *trans*-eQTL effects. Finally, we observed several genes in this component enriched for IPA biological processes such as ‘Activation of Phagocytes’ including *AXL* and *TREM1* (Fig 6C).

To assess the robustness of the *trans*-eQTLs that colocalize with Alzheimer’s disease risk alleles, we attempted to replicate the association signals using an independent dataset from the ImmVar study [7]. Of the 54 *trans*-eGenes in the *MS4A4A/6A* component, the expression of 10 *trans*-eGenes in naïve monocytes were significantly associated with rs983392 (FDR < 0.20; S19 Fig). Many of the *trans*-eQTLs were highly significant in the ImmVar data including *GBP2*, *CUL1*, and *SYTL3* (Fig 5B). Of the 38 *trans*-eGenes in the *CLU* component, expression levels of five genes were significantly associated with rs9331896 at FDR < 0.20 in the IFN stimulated monocytes (S20 Fig). Despite the small size and differences in stimuli in the replication dataset, we were able to detect suggestive association signals for a small number of *trans* effects.

Discussion

Here, we used tensor decomposition methodology [14] to detect *trans*-eQTL in naïve and stimulated human monocytes and macrophages. We detected hundreds of *trans*-acting regulatory effects on a number of genes in sparse components. These components explain a substantial amount of SNP-based genetic heritability for many common diseases. After examining the genes in each component, we observed that the target genes were involved with coherent biological processes and had regulatory motifs that are enriched for the same transcription factor. The majority of the *trans*-eQTLs were hotspot loci, each of which altered the expression of many genes within our sparse components. We detected significantly more *trans*-eQTLs in stimulated monocytes compared to naïve monocytes or macrophages despite twice the sample size for CG. Among the significant *trans*-eQTLs, we found 55% were stimuli-specific, suggesting that a larger number of *trans*-eQTLs are detectable only in the presence of specific stimuli. These observations are consistent with findings from other species such as *Caenorhabditis elegans* [42] and *Saccharomyces cerevisiae* [43] showing that environmental perturbation yields a higher number of *trans*-eQTLs compared to *cis*-eQTLs. Together, these results underscore the need to perturb primary cells with environmental stimuli to discover genotype-phenotype relationships in *trans*.

The target *trans*-eGenes reveal biological processes and downstream effects for a number of disease-associated susceptibility alleles. This includes Parkinson’s disease-associated *trans*-eQTLs with 16 target genes mediated by the lysosomal protease Cathepsin B (*CTSB*). We corroborate a previously reported *cis*-eQTL effect on *CTSB* driven by a Parkinson’s disease-associated genetic variant [39]. However, the main contribution of this study is our ability to detect

reproducible Parkinson's disease-associated *trans*-eQTLs and experimentally validate several of the target genes. Nevertheless, we are still unable to fully understand this gene network in the context of the disease due to our incomplete understanding of the functions of many genes within the network. The only gene in this network known to have any functional interaction with *CTSB* to date is *ANTXR2* (same family of *ANTXR1*, found in the network), where *CTSB*-mediated autophagy flux facilitates the delivery of toxins into the cytoplasm [44]. Further functional studies will be necessary to validate not only the *trans* targets but also to understand mechanisms underlying Parkinson's disease susceptibility at the *CTSB* locus. While alterations in autophagy and lysosomal pathways have been widely reported in neurons from Parkinson's disease [45], our results open future mechanistic studies focusing on macrophage function and gene networks in Parkinson's disease.

We have previously reported that a number of Alzheimer's disease-associated susceptibility alleles colocalize with *cis*-eQTLs in peripheral monocytes [7,46–48]. Here we hypothesized that Alzheimer's disease risk and protective alleles may modulate myeloid cell function within specific biological pathways. In support of this hypothesis, our analysis showed that the Alzheimer's disease susceptibility alleles at the *MS4A4A/6A* and *CLU* loci are associated with the individual scores of two sparse components. The component containing type 1 interferon genes explains substantial proportion (6%) of SNP-based genetic heritability for Alzheimer's disease. These *trans* target genes for both loci intersect with interferon-related functional genes that are responders of IFN- γ , key regulators of the IFN response (*IRF1* and *STAT1*), and type I interferon and antiviral effectors (*OAS*, *IFIT*, and *GBP* families). These findings confirm previous results in the p25/Cdk5 model of neurodegeneration where a reactive microglia phenotype with activated IFN pathway was found [34]. More recently, Salih *et al.* [33] identified a number of interferon signaling genes to be in co-expression network that is expressed in amyloid-responsive mouse microglia, including *Oas* and *Ifit* family members, and transcription factors such as *Irf7* and *Stat*. The authors further found that Alzheimer's disease loci colocalize with *cis*-eQTLs targeting *OAS1* in interferon- γ stimulated iPSC-derived macrophages. These studies, together with our results, suggest that dysregulation of the IFN pathway in microglia might have a role in AD pathogenesis.

The Alzheimer's disease-associated *trans*-eQTL in the clusterin (*CLU/APOJ*) locus is also associated with the expression of genes of the complement cascade. The role of the complement system in Alzheimer's disease has been suggested mostly by mouse studies in which microglia are thought to be the cellular effector of complement-mediated synaptic loss in Alzheimer's disease [49]. Indeed, C1q and oligomeric forms of amyloid- β operate in a common pathway to activate the complement cascade and drive synapse elimination by microglia [49]. In addition, a post-mortem study of Alzheimer's disease brains showed increased expression levels for complement components in Alzheimer's disease brains [50]. Other genes in this component include marker genes (*AXL* and *ITGAX*) for damage-associated microglia, or DAM, a recently identified subset of microglia found around amyloid plaques [26]. Both *AXL* and *ITGAX* are key genes involved in the Trem2-dependent DAM program, which involves the upregulation of phagocytic and lipid metabolism genes [26]. The genes in this network may have a role in debris clearance such as the removal of apoptotic neurons or A β aggregates by microglia. Along this line, network analysis of transcriptomic data from post-mortem brain tissues from Alzheimer's patients has identified immune and microglia-specific modules dominated by genes involved in pathogen phagocytosis [51]. Further work is necessary to fully understand the role of complement and interferon signaling genes in the development and progression of AD pathology.

With this study, we demonstrate that gene expression from purified immune cells are a valuable source to study gene networks associated with different diseases, including

neurodegenerative diseases. Nevertheless, this study has several limitations. First, while we were able to detect robust gene networks linked to distal genetic variants, we are still underpowered to detect *trans* target genes with smaller effect sizes. Our results suggest that many more genes have non-zero gene scores and are likely to contribute to the *trans*-eQTL networks but we are unable to detect them with the current sample size. We estimate that thousands of individuals would be needed to reliably detect effect sizes that explain a small proportion of the *trans*-QTL variance. Secondly, the reproducibility of the *trans*-eQTLs is still challenging without a stimulated dataset of comparable sample size. Thus, as larger datasets become available it will be important to validate our catalog of *trans*-eQTLs. Third, it is not clear if the *trans*-eQTLs networks identified in peripheral monocytes will be conserved in microglia during the transition to a reactive state under conditions of brain-tissue damage encountered during aging or neurodegeneration. These networks in monocytes may play a direct role in the pathogenesis of Alzheimer's or Parkinson's disease or, given the shared ontogeny of these two cell lineages, may serve as a proxy for microglial activities within the healthy, aged or diseased brain. While many genes are expressed in both peripheral monocytes and macrophages and CNS microglia, some of the genes are markers for microglia in the healthy brain, or are active during the transition from the homeostasis-associated state to a brain damage-response state. Future studies incorporating transcriptome profiles from primary human microglia from autopsied samples, or Induced Pluripotent Stem Cells (iPSC)-derived microglia challenged with disease-relevant stimuli will be an important resource in uncovering *trans*-eQTL networks underlying neurodegenerative diseases.

In summary, we identified robust *trans*-eQTL networks in peripheral myeloid cells that reveal downstream biological processes of several disease-associated loci. Although further mechanistic work is necessary to validate these gene networks our findings provide compelling human genetic evidence for a lysosomal pathway contributing to Parkinson's disease, and for myeloid phagocytosis, complement cascade and type I interferon-mediated signaling pathways contributing to Alzheimer's disease.

Methods

Overview of sparse tensor decomposition

We applied a sparse tensor decomposition model [14] to deconstruct multi-way gene expression data into latent components or *objects* of smaller dimension for simultaneous analysis. The model itself is $Y_{nlt} = \sum_{c=1}^C A_{nc} \otimes B_{tc} \otimes X_{cl} + \varepsilon_{nlt}$, where C is the number of components and A is an $N \times C$ matrix of individual scores, B is a $T \times C$ matrix of tissue scores and X is a $C \times L$ matrix of gene loadings. The error term is modeled as $\varepsilon_{nlt} \sim (0, \lambda_{lt}^{-1})$, where λ_{lt} is the precision of the error term at the l^{th} gene in the t^{th} tissue. An indicator variable I_{nt} that equals 1 when gene expression has been measured in tissue t for sample n and 0 otherwise. The likelihood is then given by $P(Y|\theta) \prod_{n,l,t} P(Y_{nlt}|\theta)^{I_{nt}}$, where θ is the vector of model parameters. This model is fit in a Bayesian framework, and place priors on the entries of the matrices A , B , X and also the precisions λ_{lt} . A key prior is the one placed on the elements of the gene loadings matrix X . A hierarchical 'spike and slab' prior is used to encourage sparsity in the rows of matrix X . The 'spike and slab' prior allows to shrink gene loadings to zero to infer more clearly which genes are involved in each component. See Hore *et al.* [14] for further details of the model.

We used the Sparse Decomposition of Arrays (SDA) software package (see URLs) to deconstruct multi-way gene expression data into latent components. The SDA model allows for non-sparse components (genes with close to '0' loading scores) that might arise as a result of

confounding effects, such as batch effects or technical artifacts. To exclude the non-sparse components from our *trans*-eQTL analysis, we implemented a sparsity ranking statistic alongside the posterior inclusion probability (PIP) > 0.5 inclusion component probability and 2.5% distributional cut-off for gene scores. The statistic is $R_i(w_i, N_i) = R_i = 1_{N_i \leq T} \times \frac{f(w_i)}{g(N_i)}$ for $i = 1, \dots, n$; where w_i is a weight function, N_i are the number of (non-zero) genes. A cutoff value is derived by simply finding a lower bound through a limiting case instead of using distributional assumptions. See [S1 Note](#) for further details.

Gene expression data

Fairfax. We obtained processed microarray gene expression data from Fairfax et al. [12] from ArrayExpress (E-MTAB-2232). In the Fairfax dataset, the gene expression of primary human monocytes was profiled in four conditions: naïve, in response to interferon- γ , and to lipopolysaccharide at 2 hours, and at 24 hours. Of the 432 total, gene expression profiles were available for 414, 367, 261 and 322 for baseline, IFN- γ , LPS2hr and LPS 24hr, respectively. The gene expression data was generated using Illumina HumanHT-12 v4 BeadChip gene expression array platform with 47,231 probes. Of these, 28,688 probes correspond to coding transcripts with well-established annotations and map unequivocally to one single genomic position were kept. We obtained the GRCh37 start and end coordinates for those genes from Ensembl for eQTL analysis. We kept the maximum of median probe expression (across individuals) for multiple probes (mapping to the same gene). This resulted in 17,509 genes used as an input analysis for SDA. See Fairfax, Humburg (12) for further details on microarray data quality control.

Cardiogenics. Cardiogenics is a European collaborative project that started in January 2007 and was funded by the European Commission through its Sixth Framework Program (reference LSHM -CT-2006-037593). As part of the Cardiogenics project, RNA from monocytes and macrophages of patients with coronary artery disease and healthy individuals was prepared and genome-wide expression was assessed in both cell types using the Illumina HumanRef 8 v3 Beadchip containing 24,516 probes corresponding to 18,311 distinct genes and 21,793 Ref Seq annotated transcripts. The DNA of all these individuals was genotyped using the Human 610 Quad custom arrays. We obtained processed microarray gene expression data from the European Genome-phenome Archive (Study: EGAS00001000411; Dataset ID: EGAD00010000446, EGAD00010000448 and EGAD00010000450). The details of the Cardiogenics datasets can be found in Rotival *et al.* [16] and Garnier *et al.* [30].

ImmVar. The Immune Variation (ImmVar) project consists of 162 African American subjects of European and African ancestry, 155 East Asian subjects of Chinese, Japanese or Korean ancestry, and 377 Caucasian subjects of European ancestry. Genome-wide genotyping was done using Illumina Infinium Human OmniExpress Exome BeadChip and subsequently imputed using the Michigan Imputation Service with Human Reference Consortium v1.1 reference panel. The mRNAs were profiled by Affymetrix GeneChip Human Gene ST 1.0 microarrays and raw data CEL files were processed using the Robust Multichip Average algorithm in Affymetrix PowerTools. For further details see Raj, Rothamel (7). In addition to array data, the stimulated and baseline monocytes from the ImmVar cohort were subsequently sequenced [52]. The RNA-seq counts of baseline and interferon-stimulated ImmVar data were downloaded from accession GSE92904. The raw fastq and genotype data is available from dbGAP under accession phs000815.v1.p1.

STARNET. The Stockholm-Tartu Atherosclerosis Reverse Network Engineering Task (STARNET) [5] macrophage gene expression data and genotype data is available from dbGAP under accession phs001203.v1.p1.

Genotype quality control and imputation

The raw genotype data were downloaded from the EGA (EGAS00000000109 and EGAD00010000450) for Fairfax and Cardiogenics, respectively, and from dbGAP (accession phs001203.v1.p1) for STARNET. The raw data was subsequently imputed using the Michigan Imputation Service with Human Reference Consortium v1.1 reference panel of European ancestry. We used 5,386,706 variants with minor allele frequency ≥ 0.01 , INFO score ≥ 0.3 and Hardy-Weinberg equilibrium chi-square $P > 1 \times 10^{-6}$ for downstream eQTL analysis.

Gene expression quality control

The gene expression array data sets were processed using the same pipeline. We performed an additional level of probe set filtering: (i) all array probes with a single nucleotide polymorphism (SNP) at minor allele frequency (MAF) greater than 0.1 in any of the 1000 Genomes populations were removed. (ii) probes that do not map to the human genome were removed. (iii) potential cross-hybridization probes, as provided by Affymetrix or Illumina, were flagged and removed prior to the *trans*-eQTL association analysis. (iv) only uniquely mapping probes and those mapping to GENCODE v19 were included. (v) probes mapping to the X and Y chromosome were excluded. The gene expression data from each study were first internally normalized by dividing the expression values for each gene in individuals of that cohort by the mean expression value across the study, with the assumption that inter-batch differences on normalized data are much lower than those on raw expression values. These normalized values for the three cohorts were assembled and \log_2 -transformed.

GWAS SNPs

The list of SNPs associated with various human complex diseases and traits was downloaded from the GWAS Catalog (see URLs; accessed October 2017). We included the SNPs with genome-wide significant association ($P < 5 \times 10^{-8}$) in our analyses. The list of SNPs was pruned to eliminate SNPs with high LD (pairwise $r^2 < 0.4$).

eQTL mapping

Prior to testing for *trans*-eQTLs association, we discarded the components that were correlated with known biological or technical covariates (see Fig 1). Only SNPs with a minor allele frequency (MAF) > 0.05 and a Hardy-Weinberg equilibrium $P > 0.001$ were included in the analyses. We performed association tests of SNP genotype or imputed allele dosage with individual scores using linear regression as implemented in the Matrix-eQTL¹⁶ software package. For genome-wide *trans* analyses, we used an FDR of 0.05 to report the significant associations. We used a more liberal FDR threshold of 0.15 for *trans*-eQTLs that colocalize with disease-associated loci. For *trans*-eQTLs that colocalized in disease-associated loci, we permuted individual scores (from SDA) 10,000 times to generate a null-distribution. We then compared the nominal eQTL P -values to the empirical distribution created from the permuted datasets [8].

The *cis* and *trans* eQTL (SNP-by-gene) was carried using linear regression to perform associations between the imputed SNPs and the normalized gene expression. *Cis*-eQTL analysis was limited to SNPs located within 1MB either side of the transcription start or end site. We used PEER [53] to account for confounding factors in the gene expression data. We fit the PEER model to gene expression data with 15 factors. The residuals from PEER-corrected gene expression data and imputed SNP dosages were used to perform linear regression using the Matrix-eQTL [54] software package. FDR for the *cis*- and *trans*-eQTL analysis was calculated following Benjamini and Hochberg [55] procedure as implemented in the Matrix-eQTL package.

Enrichment analysis

We used oPOSSUM-3 [56] to identify over-represented transcription factor binding sites (TFBS) among genes in each component. The Z-score uses the normal approximation to the binomial distribution to compare the rate of occurrence of a TFBS in the target set of genes (in each component) to the expected rate estimated from the pre-computed background set. TopGO package [57] was used to perform enrichment analysis for Gene Ontology (GO) terms. We ran enrichment for significant (PIP > 0.5 and distributional cut-off 2.5%) genes amongst all component networks. We also applied Ingenuity Canonical Pathway (IPA) analysis to perform pathway enrichment for genes within each component. We used MAGMA [31] to test for enrichment of components among GWAS traits.

LD score regression

We used stratified LD score regression [32,58] to partition SNP-based disease heritability within categories defined by the sparse components. Using GWAS summary statistics from 18 traits or complex diseases (Alzheimer's disease, Parkinson's disease, Autism, Multiple Sclerosis, Schizophrenia, Type 2 Diabetes, Lupus, Primary Biliary Cirrhosis, Rheumatoid Arthritis, Irritable Bowel Disease, Crohn's disease, Celiac Disease, Ulcerative Colitis, High-density lipoprotein Cholesterol, Low-density lipoprotein Cholesterol, Body Mass Index, Coronary Artery Disease, and Height) and LD modeled from 1000 genomes reference panel of European ancestry, we calculated the proportion of genome-wide SNP-based heritability that be attributed to SNPs within each component. Categories for each component were defined by taking all the SNPs (within each gene plus 10 kb +/- from transcript start and stop sites) for all genes within the component. To improve model accuracy, the categories defined by components categories were added to the 'full baseline model' which included 53 functional categories capturing a broad set of functional and regulatory elements. Enrichment is defined as the proportion of SNP-heritability accounted for by each component divided by the proportion of total SNPs within the module. Components with FDR-corrected enrichment *P*-values of less than 0.05 were considered significant heritability contributors.

Mendelian randomization

Mendelian Randomization is a form of instrumental variable regression used to formulate a mediating path between a variant (SNP) to a *trans*-network (*trans*-eGenes or component individual scores) through a causal gene (*cis*-gene). We exploit a possible instrument or variant/SNP that changes this causal gene but not the *trans*-gene(s) or component individual scores (aside from through the causal gene). Hence, this yields a mediating path through the causal gene where all the noise is removed except that from the variant. We implement the McDowell, Pai (38) formulation of Mendelian randomization.

Experimental validation

Human monocytic cell line, THP-1, was differentiated into macrophages upon treatment with 20 nM phorbol 12-myristate 13-acetate (PMA) for 3 days. Cells were then *transfected* with human SCR-siRNA, CTSB-siRNA or MS4A4A-siRNA (Dharmacon) using Jetprime *transfection* reagent during 8 h. Cells were collected after 48 h and gene expression levels were assessed by qPCR using Taqman primers. For the MS4A4A experiment, macrophages were exposed to 20 ng/ml of IFN γ (R&D Systems) for the last 24 h hours.

Data access

We have made a browser available for all significant *trans*-eQTLs at https://rajlab.shinyapps.io/Tensor_myeloid/. This browser also provides the list of all sparse components, activity scores, and gene scores.

URLs

LD Score Regression, <https://github.com/bulik/ldsc>.

Sparse Decomposition of Arrays (SDA), <https://jmarchini.org/sda/>.

GWAS Catalog, <http://www.ebi.ac.uk/gwas>.

Myeloid Tensor Shiny Application, https://rajlab.shinyapps.io/Tensor_myeloid/.

Supporting information

S1 Fig. Association of component individual scores with biological and technical covariates in Cardiogenics (CG). A: Shown in the heat map are $-\log_{10}$ transformed P-value of association between CG component individual scores and known covariates. Left: All 500 Components, Right: Sparse Components. B: For these sparse components, the component tissue scores were grouped based on activity in monocytes (MC) or macrophages (MP). (PDF)

S2 Fig. Replication of previously identified *trans*-eQTLs in Fairfax et. al (2014). The SNP rs2275888 maps to FF component 337 enriched with type 1 interferon-related genes with a p-value = 1.7×10^{-9} and FDR = 1×10^{-3} and gene network similar to Fairfax et. al (2014) study. A *cis*-eQTL at rs2275888 for IFNB1 is associated with the expression of 17 genes in trans after 24-hour LPS stimulation, many of which are interferon response genes. (PDF)

S3 Fig. *Trans*-eQTL mapping to a component with MHC and non-MHC genes. A: The gene expression profiles of component (363) with MHC and non-MHC genes in FF are shared among all stimuli (Top panel). Shown are the gene scores for the genes with Posterior Inclusion Probability (PIP) > 0.5 (bottom panel). *Trans*-eQTL for the MHC component co-localize with B) cholesterol associated risk variant rs9378212, C) Type 2 Diabetes risk variant rs9268645, and D) Coronary Artery Disease risk variant rs9268402. (PDF)

S4 Fig. Mendelian randomization analysis for *trans*-eQTLs mediated by a *cis*-gene in the MHC. A: Cholesterol variant rs9378212 mediates the trans-effects through HLA-DQA1, B: Type 2 Diabetes variant rs9268645 mediates the trans-effects through HLADQA1, and C: Coronary Artery Disease variant rs9268402 mediates the trans-effects through HLA-DRB1. (PDF)

S5 Fig. Replication of *trans*-eQTLs that co-localize with disease-associated susceptibility allele. A: Top: Component 391 active in FF_{LPS24} (left) maps to Crohn's disease associated variant rs503734 (Middle). The component contains genes from the members of the Zinc finger family (right). *Bottom*: The trans-eQTL is replicated in CG. B: (Top:) FF Component 105 active in FF_{LPS2} (left) maps to Coronary Artery Disease variant rs589448 (middle) and component gene scores (right). *Bottom*: The trans-eQTL is replicated in CG (Component 417 in CG_{MP}). (PDF)

S6 Fig. Mendelian randomization analysis for disease-associated trans-eQTLs in FF and CG. (A) Crohn's variant rs503734 mediate trans-effects through *cis*-gene *SENP7* and (B)

Coronary Artery Disease variant rs589448 mediate trans-effects through *cis*-gene *LYZ* and *YEATS4*.
(PDF)

S7 Fig. *Cis*-eQTL (rs1296028-CTSB) co-localizes with Parkinson's disease associated variant rs1296028.

(PDF)

S8 Fig. SNP by gene trans-eQTL association for CG component 46. Shown are box plots for Parkinson's disease associated variant rs1296028 (near *CTSB*) mapping to *AARSD1*, *DDHD2*, *DGCR5*, and *DOK7*.

(PDF)

S9 Fig. SNP by gene trans-eQTL association for CG component 46. Shown are box plots for Parkinson's disease associated variant rs1296028 (near *CTSB*) mapping to: *MTHFSD*, *NAT14*, *PCDHGB4*, and *RIOK1*.

(PDF)

S10 Fig. Independent replication of *trans*-eGenes in the *CTSB* component. SNP by Gene association analysis was performed in an independent macrophage data from the STARNET cohort. Shown here are the *trans*-eQTL for rs1296028 and selected *trans* genes in the *CTSB* component at FDR < 0.20.

(PDF)

S11 Fig. Significant *cis*-eQTL effect for rs983392 to both *MS4A4A* and *MS4A6A* in baseline monocytes.

(PDF)

S12 Fig. *Trans*-eQTLs for component FF 26 (SNP by gene analysis). FF Component 26 *trans*-eGenes: *CD38*, *CUL1*, *CYP4F3*, *FAM129B*, and *FBXO6*; *trans*-eSNP rs983392.

(PDF)

S13 Fig. *Trans*-eQTLs for component FF 26 (SNP by gene analysis). FF Component 26 *trans*-eGenes: *GBP1*, *HMGCR*, *LGALS3BP*, *PADI4* and *PSME2*; *trans*-eSNP rs983392.

(PDF)

S14 Fig. *Trans*-eQTLs for component FF 26 (SNP by gene analysis). FF Component 26 *trans*-eGenes: *QPCT*, *SERPING1*, *STAT1*, *STOM*, *SYTL3*, and *TAP1*; *trans*-eSNP rs983392.

(PDF)

S15 Fig. *Trans*-eQTLs for component FF 26 (SNP by gene analysis). FF Component 22 *trans*-eGenes: *ADM*, *CA11*, *FBP1*, *IFITM1*, and *ISG15*; *trans*-eSNP rs9331896.

(PDF)

S16 Fig. *Trans*-eQTLs for Component FF 26 (SNP by Gene analysis). FF Component 22 *trans*-eGenes: *ISG20*, *ITGAX*, *MT1E*, *MT1F*, and *OASL*; *trans*-eSNP rs9331896.

(PDF)

S17 Fig. *Trans*-eQTLs for Component FF 26 (SNP by Gene analysis). FF Component 22 *trans*-eGenes: *RSAD2*, *SLC2A5*, *TOP2B*, and *VASH1*; *trans*-eSNP rs9331896.

(PDF)

S18 Fig. Protein-protein interaction (PPI) network of protein products of *CLU*, *CIQA*, *CIQB* and *CIQC*. The PPI suggests that the products of *CLU* physically interact with

complement proteins. The PPI was generated with GeNets Meta network v1.0 database. (PDF)

S19 Fig. Independent replication of trans-eGenes in the MS4A4A/6A component. SNP by Gene association analysis was performed in an independent baseline monocytes from the ImmVar cohort. Shown here are the *trans*-eQTL for rs983392 and selected *trans* genes in the MS4A4A/6A component (FDR < 0.20). (PDF)

S20 Fig. Independent replication of trans-eGenes in the CLU component. SNP by Gene association analysis was performed in an independent stimulated monocyte data from the ImmVar cohort. Shown here are the *trans*-eQTL for rs9331896 and selected *trans* genes in the CLU component (FDR < 0.20). (PDF)

S1 Table. List of sparse components with corresponding gene scores and stimuli (IFN and LPS) activity scores identified in the FF dataset. (XLSX)

S2 Table. List of sparse components with corresponding gene scores and tissue (monocytes and macrophages) activity scores identified in the CG dataset. (XLSX)

S3 Table. Components that replicate across the three datasets: Fairfax(FF), Cardiogenics (CG) and ImmVar(IMM). Two-way reverse correlation with the gene scores was conducted on the common intersected genes from the respective components. (XLSX)

S4 Table. Enrichment of Gene Ontology (GO) categories among the genes in the sparse components. (XLSX)

S5 Table. Sparse components that are enriched for genes within disease-associated loci. The component number, P-value, and corresponding GWAS disease or traits are listed. (XLSX)

S6 Table. SNP-based heritability enrichment for each component. Proportion of heritability and enrichment statistics for 18 selected complex traits that can be attributed to each sparse component from the FF data. (XLSX)

S7 Table. Trans-eQTLs detected in Fairfax data (FF; FDR < 0.05). The matrix-eQTL output with SNP, component number, beta, t-stat, P-value and FDR are listed here. Note: The *trans*-eSNPs are not LD-pruned. (XLSX)

S8 Table. Trans-eQTLs detected in Cardiogenics data (CG; FDR < 0.05). The matrix-eQTL output with SNP, component number, beta, t-stat, P-value and FDR are listed here. Note: The *trans*-eSNPs are not LD-pruned. (XLSX)

S9 Table. Trans-eSNPs detected in the FF dataset (FDR < 0.15) that co-localize with trait-associated GWAS SNPs. The table lists SNP, P-value, trait, component number, *cis*-gene (if any), tissue or stimuli specificity scores, genes and gene loading scores for each component

and FDR.
(XLSX)

S10 Table. *Trans*-eSNPs detected in the CG dataset (FDR < 0.15) that co-localize with trait-associated GWAS SNPs. The table lists SNP, P-value, trait, component number, *cis*-gene (if any), tissue or stimuli specificity scores, genes and gene loading scores for each component, and FDR.
(XLSX)

S1 Note. Assumptions and sparsity ranking statistic formulation for gene scores for each component.
(PDF)

Acknowledgments

We thank B. Fairfax (Oxford University) for sharing the gene expression data. We thank J. Castellano, K. Alasoo, E. Marcora, and members of the Raj Laboratory and Ronald M. Loeb Center for Alzheimer's disease for critical feedback on the initial draft of the manuscript. This work was supported in part through the computational resources and staff expertise provided by Scientific Computing at the Icahn School of Medicine at Mount Sinai.

Author Contributions

Conceptualization: Towfique Raj.

Data curation: Towfique Raj.

Formal analysis: Satesh Ramdhani, Evan Udine.

Funding acquisition: Towfique Raj.

Investigation: Satesh Ramdhani, Elisa Navarro, Anastasia G. Efthymiou, Towfique Raj.

Methodology: Satesh Ramdhani, Elisa Navarro, Anastasia G. Efthymiou, Towfique Raj.

Software: Brian M. Schilder.

Supervision: Alison Goate, Towfique Raj.

Validation: Satesh Ramdhani, Elisa Navarro, Anastasia G. Efthymiou, Madison Parks.

Visualization: Brian M. Schilder.

Writing – original draft: Satesh Ramdhani, Towfique Raj.

Writing – review & editing: Satesh Ramdhani, Elisa Navarro, Towfique Raj.

References

1. Stranger BE, Stahl EA, Raj T. Progress and promise of genome-wide association studies for human complex trait genetics. *Genetics*. 2011; 187(2):367–83. <https://doi.org/10.1534/genetics.110.120907> PMID: 21115973.
2. Visscher PM, Wray NR, Zhang Q, Sklar P, McCarthy MI, Brown MA, et al. 10 Years of GWAS Discovery: Biology, Function, and Translation. *American journal of human genetics*. 2017; 101(1):5–22. <https://doi.org/10.1016/j.ajhg.2017.06.005> PMID: 28686856.
3. Maurano MT, Humbert R, Rynes E, Thurman RE, Haugen E, Wang H, et al. Systematic localization of common disease-associated variation in regulatory DNA. *Science*. 2012; 337(6099):1190–5. <https://doi.org/10.1126/science.1222794> PMID: 22955828.

4. Consortium GT. Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. *Science*. 2015; 348(6235):648–60. Epub 2015/05/09. <https://doi.org/10.1126/science.1262110> PMID: 25954001.
5. Franzen O, Ermel R, Cohain A, Akers NK, Di Narzo A, Talukdar HA, et al. Cardiometabolic risk loci share downstream cis- and trans-gene regulation across tissues and diseases. *Science*. 2016; 353(6301):827–30. <https://doi.org/10.1126/science.aad6970> PMID: 27540175.
6. Ramasamy A, Trabzuni D, Guelfi S, Varghese V, Smith C, Walker R, et al. Genetic variability in the regulation of gene expression in ten regions of the human brain. *Nature neuroscience*. 2014; 17(10):1418–28. <https://doi.org/10.1038/nn.3801> PMID: 25174004.
7. Raj T, Rothamel K, Mostafavi S, Ye C, Lee MN, Replogle JM, et al. Polarization of the effects of autoimmune and neurodegenerative risk alleles in leukocytes. *Science*. 2014; 344(6183):519–23. <https://doi.org/10.1126/science.1249547> PMID: 24786080.
8. Stranger BE, Montgomery SB, Dimas AS, Parts L, Stegle O, Ingle CE, et al. Patterns of cis regulatory variation in diverse human populations. *PLoS genetics*. 2012; 8(4):e1002639. <https://doi.org/10.1371/journal.pgen.1002639> PMID: 22532805.
9. Lappalainen T, Sammeth M, Friedlander MR, t Hoen PA, Monlong J, Rivas MA, et al. Transcriptome and genome sequencing uncovers functional variation in humans. *Nature*. 2013; 501(7468):506–11. <https://doi.org/10.1038/nature12531> PMID: 24037378.
10. Ye CJ, Feng T, Kwon HK, Raj T, Wilson MT, Asinovski N, et al. Intersection of population variation and autoimmunity genetics in human T cell activation. *Science*. 2014; 345(6202):1254665. <https://doi.org/10.1126/science.1254665> PMID: 25214635.
11. Lee MN, Ye C, Villani AC, Raj T, Li W, Eisenhaure TM, et al. Common genetic variants modulate pathogen-sensing responses in human dendritic cells. *Science*. 2014; 343(6175):1246980. <https://doi.org/10.1126/science.1246980> PMID: 24604203.
12. Fairfax BP, Humburg P, Makino S, Naranbhai V, Wong D, Lau E, et al. Innate immune activity conditions the effect of regulatory variants upon monocyte gene expression. *Science*. 2014; 343(6175):1246949. <https://doi.org/10.1126/science.1246949> PMID: 24604202.
13. Westra HJ, Peters MJ, Esko T, Yaghootkar H, Schurmann C, Kettunen J, et al. Systematic identification of trans eQTLs as putative drivers of known disease associations. *Nat Genet*. 2013; 45(10):1238–43. Epub 2013/09/10. <https://doi.org/10.1038/ng.2756> PMID: 24013639.
14. Hore V, Vinuela A, Buil A, Knight J, McCarthy MI, Small K, et al. Tensor decomposition for multiple-tissue gene expression experiments. *Nature genetics*. 2016; 48(9):1094–100. <https://doi.org/10.1038/ng.3624> PMID: 27479908.
15. Fagny M, Paulson JN, Kuijjer ML, Sonawane AR, Chen CY, Lopes-Ramos CM, et al. Exploring regulation in tissues with eQTL networks. *Proceedings of the National Academy of Sciences of the United States of America*. 2017; 114(37):E7841–E50. <https://doi.org/10.1073/pnas.1707375114> PMID: 28851834.
16. Rotival M, Zeller T, Wild PS, Maouche S, Szymczak S, Schillert A, et al. Integrating genome-wide genetic variations and monocyte expression data reveals trans-regulated gene modules in humans. *PLoS genetics*. 2011; 7(12):e1002367. <https://doi.org/10.1371/journal.pgen.1002367> PMID: 22144904.
17. Rakitsch B, Stegle O. Modelling local gene networks increases power to detect trans-acting genetic effects on gene expression. *Genome Biol*. 2016; 17:33. Epub 2016/02/26. <https://doi.org/10.1186/s13059-016-0895-2> PMID: 26911988.
18. Braak H, de Vos RA, Jansen EN, Bratzke H, Braak E. Neuropathological hallmarks of Alzheimer's and Parkinson's diseases. *Progress in brain research*. 1998; 117:267–85. [https://doi.org/10.1016/s0079-6123\(08\)64021-2](https://doi.org/10.1016/s0079-6123(08)64021-2) PMID: 9932414.
19. Ramanan VK, Saykin AJ. Pathways to neurodegeneration: mechanistic insights from GWAS in Alzheimer's disease, Parkinson's disease, and related disorders. *American journal of neurodegenerative disease*. 2013; 2(3):145–75. PMID: 24093081.
20. Lambert JC, Ibrahim-Verbaas CA, Harold D, Naj AC, Sims R, Bellenguez C, et al. Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease. *Nat Genet*. 2013; 45(12):1452–8. PMID: 24162737.
21. Sims R, van der Lee SJ, Naj AC, Bellenguez C, Badarinarayan N, Jakobsdottir J, et al. Rare coding variants in PLCG2, ABI3, and TREM2 implicate microglial-mediated innate immunity in Alzheimer's disease. *Nature genetics*. 2017; 49(9):1373–84. <https://doi.org/10.1038/ng.3916> PMID: 28714976.
22. Kunkle BW, Grenier-Boley B, Sims R, Bis JC, Damotte V, Naj AC, et al. Genetic meta-analysis of diagnosed Alzheimer's disease identifies new risk loci and implicates Abeta, tau, immunity and lipid processing. *Nat Genet*. 2019; 51(3):414–30. <https://doi.org/10.1038/s41588-019-0358-2> PMID: 30820047.

23. Chang D, Nalls MA, Hallgrimsdottir IB, Hunkapiller J, van der Brug M, Cai F, et al. A meta-analysis of genome-wide association studies identifies 17 new Parkinson's disease risk loci. *Nat Genet.* 2017; 49(10):1511–6. <https://doi.org/10.1038/ng.3955> PMID: 28892059.
24. Mike A Nalls CB, Costanza L Vallergera, Karl Heilbron, Sara Bandres-Ciga, Diana Chang, Manuela Tan, Demis A Kia, Alastair J Noyce, Angli Xue, Jose Bras, Emily Young, Ranier von Coelln, Javier Simon-Sanchez, Claudia Schulte, Manu Sharma, Lynne Krohn, Lasse Pihlstrom, Ari Siitonen, Hiro-taka Iwaki, Hampton Leonard, Faraz Faghri, J Raphael Gibbs, Dena G Hernandez, Sonja W Scholz, Juan A Botia, Maria Martinez, Jean-Christophe Corvol, Suzanne Lesage, Joseph Jankovic, Lisa M Shulman, The 23andMe Research Team, System Genomics of Parkinson's Disease (SGPD) Consortium, Margaret Sutherland, Pentti Tienari, Kari Majamaa, Mathias Toft, Alexis Brice, Jian Yang, Ziv Gan-Orr, Thomas M Gasser, Peter M Heutink, Joshua M Shulman, Nicolas A Wood, David A Hinds, John R Hardy, Huw R Morris, Jacob M Gratten, Peter M Visscher, Robert R Graham, Andrew B Singleton, International Parkinson's Disease Genomics Consortium. Expanding Parkinson's disease genetics: novel risk loci, genomic context, causal insights and heritable risk. *bioRxiv* 388165. 2019. <https://doi.org/10.1101/388165>.
25. Gagliano SA, Pouget JG, Hardy J, Knight J, Barnes MR, Ryten M, et al. Genomics implicates adaptive and innate immunity in Alzheimer's and Parkinson's diseases. *Annals of clinical and translational neurology.* 2016; 3(12):924–33. <https://doi.org/10.1002/acn3.369> PMID: 28097204.
26. Keren-Shaul H, Spinrad A, Weiner A, Matcovitch-Natan O, Dvir-Szternfeld R, Ulland TK, et al. A Unique Microglia Type Associated with Restricting Development of Alzheimer's Disease. *Cell.* 2017; 169(7):1276–90 e17. <https://doi.org/10.1016/j.cell.2017.05.018> PMID: 28602351.
27. Amor S, Puentes F, Baker D, van der Valk P. Inflammation in neurodegenerative diseases. *Immunology.* 2010; 129(2):154–69. <https://doi.org/10.1111/j.1365-2567.2009.03225.x> PMID: 20561356.
28. Krasemann S, Madoe C, Cialic R, Baufeld C, Calcagno N, El Fatimy R, et al. The TREM2-APOE Pathway Drives the Transcriptional Phenotype of Dysfunctional Microglia in Neurodegenerative Diseases. *Immunity.* 2017; 47(3):566–81 e9. <https://doi.org/10.1016/j.immuni.2017.08.008> PMID: 28930663.
29. Middeldorp J, Lehallier B, Villeda SA, Miedema SS, Evans E, Czirr E, et al. Preclinical Assessment of Young Blood Plasma for Alzheimer Disease. *JAMA neurology.* 2016; 73(11):1325–33. <https://doi.org/10.1001/jamaneurol.2016.3185> PMID: 27598869.
30. Garnier S, Truong V, Brocheton J, Zeller T, Rovital M, Wild PS, et al. Genome-wide haplotype analysis of cis expression quantitative trait loci in monocytes. *PLoS genetics.* 2013; 9(1):e1003240. <https://doi.org/10.1371/journal.pgen.1003240> PMID: 23382694.
31. de Leeuw CA, Mooij JM, Heskes T, Posthuma D. MAGMA: generalized gene-set analysis of GWAS data. *PLoS Comput Biol.* 2015; 11(4):e1004219. Epub 2015/04/18. <https://doi.org/10.1371/journal.pcbi.1004219> PMID: 25885710.
32. Finucane HK, Bulik-Sullivan B, Gusev A, Trynka G, Reshef Y, Loh PR, et al. Partitioning heritability by functional annotation using genome-wide association summary statistics. *Nat Genet.* 2015; 47(11):1228–35. Epub 2015/09/29. <https://doi.org/10.1038/ng.3404> PMID: 26414678.
33. Salih Dervis A B S, Guelfi Sebastian, Reynolds Regina H, Shoai Maryam, Ryten Mina, Brenton Jonathan, Zhang David, Matarin Mar, Botia Juan A, Shah Runil, Brookes Keeley J, Guetta-Baranes Tamar, Morgan Kevin, Bellou Eftychia, Cummings Damian M, Escott-Price Valentina, Hardy John. Genetic variability in response to amyloid beta deposition influences Alzheimer's disease risk. *Brain Communications.* 2019. <https://doi.org/10.1093/braincomms/fcz022>.
34. Mathys H, Adakkan C, Gao F, Young JZ, Manet E, Hemberg M, et al. Temporal Tracking of Microglia Activation in Neurodegeneration at Single-Cell Resolution. *Cell Rep.* 2017; 21(2):366–80. <https://doi.org/10.1016/j.celrep.2017.09.039> PMID: 29020624.
35. Sala Frigerio C, Wolfs L, Fattorelli N, Thrupp N, Voytyuk I, Schmidt I, et al. The Major Risk Factors for Alzheimer's Disease: Age, Sex, and Genes Modulate the Microglia Response to Abeta Plaques. *Cell Rep.* 2019; 27(4):1293–306 e6. <https://doi.org/10.1016/j.celrep.2019.03.099> PMID: 31018141.
36. Filiano AJ, Xu Y, Tustison NJ, Marsh RL, Baker W, Smirnov I, et al. Unexpected role of interferon-gamma in regulating neuronal connectivity and social behaviour. *Nature.* 2016; 535(7612):425–9. Epub 2016/07/15. <https://doi.org/10.1038/nature18626> PMID: 27409813.
37. Gandal MJ, Zhang P, Hadjimichael E, Walker RL, Chen C, Liu S, et al. Transcriptome-wide isoform-level dysregulation in ASD, schizophrenia, and bipolar disorder. *Science.* 2018; 362(6420). Epub 2018/12/14. <https://doi.org/10.1126/science.aat8127> PMID: 30545856.
38. McDowell I, Pai A, Guo C, Vockley CM, Brown CD, Reddy TE, et al. Many long intergenic non-coding RNAs distally regulate mRNA gene expression levels. *bioRxiv.* 2016:044719.
39. Li YI, Wong G, Humphrey J, Raj T. Prioritizing Parkinson's disease genes using population-scale transcriptomic data. *Nat Commun.* 2019; 10(1):994. <https://doi.org/10.1038/s41467-019-08912-9> PMID: 30824768.

40. Giambartolomei C, Vukcevic D, Schadt EE, Franke L, Hingorani AD, Wallace C, et al. Bayesian test for colocalisation between pairs of genetic association studies using summary statistics. *PLoS genetics*. 2014; 10(5):e1004383. <https://doi.org/10.1371/journal.pgen.1004383> PMID: 24830394.
41. Raj T, Li Yi, Wong G, Humphrey J, Wang M, Ramdhani S, et al. Integrative transcriptome analyses of the aging brain implicate altered splicing in Alzheimer's disease susceptibility. *Nature genetics*. 2018. <https://doi.org/10.1038/s41588-018-0238-1> PMID: 30297968.
42. Snoek BL, Sterken MG, Bevers RPJ, Volkers RJM, Van't Hof A, Brenchley R, et al. Contribution of trans regulatory eQTL to cryptic genetic variation in *C. elegans*. *BMC genomics*. 2017; 18(1):500. <https://doi.org/10.1186/s12864-017-3899-8> PMID: 28662696.
43. Kita R, Venkataram S, Zhou Y, Fraser HB. High-resolution mapping of cis-regulatory variation in budding yeast. *Proceedings of the National Academy of Sciences of the United States of America*. 2017; 114(50):E10736–E44. <https://doi.org/10.1073/pnas.1717421114> PMID: 29183975.
44. Ha SD, Ham B, Mogridge J, Saftig P, Lin S, Kim SO. Cathepsin B-mediated autophagy flux facilitates the anthrax toxin receptor 2-mediated delivery of anthrax lethal factor into the cytoplasm. *The Journal of biological chemistry*. 2010; 285(3):2120–9. <https://doi.org/10.1074/jbc.M109.065813> PMID: 19858192.
45. Plotegher N, Duchon MR. Crosstalk between Lysosomes and Mitochondria in Parkinson's Disease. *Frontiers in cell and developmental biology*. 2017; 5:110. <https://doi.org/10.3389/fcell.2017.00110> PMID: 29312935.
46. Raj T, Ryan KJ, Replogle JM, Chibnik LB, Rosenkrantz L, Tang A, et al. CD33: increased inclusion of exon 2 implicates the Ig V-set domain in Alzheimer's disease susceptibility. *Human molecular genetics*. 2014; 23(10):2729–36. <https://doi.org/10.1093/hmg/ddt666> PMID: 24381305.
47. Bradshaw EM, Chibnik LB, Keenan BT, Ottoboni L, Raj T, Tang A, et al. CD33 Alzheimer's disease locus: altered monocyte function and amyloid biology. *Nature neuroscience*. 2013; 16(7):848–50. <https://doi.org/10.1038/nn.3435> PMID: 23708142.
48. Huang KL, Marcora E, Pimenova AA, Di Narzo AF, Kapoor M, Jin SC, et al. A common haplotype lowers PU.1 expression in myeloid cells and delays onset of Alzheimer's disease. *Nature neuroscience*. 2017; 20(8):1052–61. <https://doi.org/10.1038/nn.4587> PMID: 28628103.
49. Hong S, Beja-Glasser VF, Nfonoyim BM, Frouin A, Li S, Ramakrishnan S, et al. Complement and microglia mediate early synapse loss in Alzheimer mouse models. *Science*. 2016; 352(6286):712–6. <https://doi.org/10.1126/science.aad8373> PMID: 27033548.
50. Veerhuis R, Janssen I, Hack CE, Eikelenboom P. Early complement components in Alzheimer's disease brains. *Acta neuropathologica*. 1996; 91(1):53–60. <https://doi.org/10.1007/s004019570001> PMID: 8773146.
51. Zhang B, Gaiteri C, Bodea LG, Wang Z, McElwee J, Podtelezhnikov AA, et al. Integrated systems approach identifies genetic nodes and networks in late-onset Alzheimer's disease. *Cell*. 2013; 153(3):707–20. <https://doi.org/10.1016/j.cell.2013.03.030> PMID: 23622250.
52. Ye CJ, Chen J, Villani AC, Gate RE, Subramaniam M, Bhangale T, et al. Genetic analysis of isoform usage in the human anti-viral response reveals influenza-specific regulation of ERAP2 transcripts under balancing selection. *Genome Res*. 2018; 28(12):1812–25. Epub 2018/11/18. <https://doi.org/10.1101/gr.240390.118> PMID: 30446528.
53. Stegle O, Parts L, Piipari M, Winn J, Durbin R. Using probabilistic estimation of expression residuals (PEER) to obtain increased power and interpretability of gene expression analyses. *Nature protocols*. 2012; 7(3):500–7. <https://doi.org/10.1038/nprot.2011.457> PMID: 22343431.
54. Shabalin AA. Matrix eQTL: ultra fast eQTL analysis via large matrix operations. *Bioinformatics*. 2012; 28(10):1353–8. <https://doi.org/10.1093/bioinformatics/bts163> PMID: 22492648.
55. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society, Series B*. 57 (1): 289–300.
56. Kwon AT, Arenillas DJ, Worsley Hunt R, Wasserman WW. oPOSSUM-3: advanced analysis of regulatory motif over-representation across genes or ChIP-Seq datasets. *G3*. 2012; 2(9):987–1002. <https://doi.org/10.1534/g3.112.003202> PMID: 22973536.
57. Alexa A, Rahnenfuhrer J, Lengauer T. Improved scoring of functional groups from gene expression data by decorrelating GO graph structure. *Bioinformatics*. 2006; 22(13):1600–7. <https://doi.org/10.1093/bioinformatics/btl140> PMID: 16606683.
58. Bulik-Sullivan BK, Loh PR, Finucane HK, Ripke S, Yang J, Schizophrenia Working Group of the Psychiatric Genomics C, et al. LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat Genet*. 2015; 47(3):291–5. Epub 2015/02/03. <https://doi.org/10.1038/ng.3211> PMID: 25642630.