Single Cell-type Integrative Network Modeling Identified Novel Microglial-specific Targets for the Phagosome in Alzheimer's disease

Authorship

Kruti Rajan Patel[1,13,14*], Kuixi Zhu[2,3,4,5*], Marc Y.R. Henrion[2,6,7]*, Noam D. Beckmann[2,3*], Sara Moein[2,3,4,5], Melissa L. Alamprese[4,5], Mariet Allen[8], Xue Wang[9], Gail Chan[15], Thomas Pertel[16], Parham Nejad[17] , Joseph S. Reddy[9], Minerva M. Carrasquillo[8], David A Bennett[19], Nilüfer Ertekin-Taner[8,10], Philip L.De Jager[11,12,14], Eric E. Schadt[2,3#],Elizabeth M. Bradshaw[11#], Rui Chang[2,3,4,5,18*#]

1. Department of Neuroscience& Opthalmology, Homology Medicines, Inc., Bedford, MA
2. Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, 10029 NY, USA
3. Institute of Genomics and Multiscale Biology, Icahn School of Medicine at Mount Sinai, New York, 10029 NY, USA
4. Department of Neurology, University of Arizona, Tucson, 85721, AZ, USA
5. The Center for Innovations in Brain Sciences, University of Arizona, Tucson, 85721, AZ, USA
6 Liverpool School of Tropical Medicine, Pembroke Place, Liverpool, L3 5QA, UK
7 Malawi - Liverpool - Wellcome Trust Clinical Research Programme, Blantyre, Malawi
8.Department of Neuroscience, Mayo Clinic Florida, Jacksonville, FL 32224, USA
9. Department of Health Sciences Research, Mayo Clinic Florida, Jacksonville, FL 32224, USA
10. Department of Neurology, Mayo Clinic Florida, Jacksonville, FL 32224, USA
11. Center for Translational & Computational Neuroimmunology, Department of Neurology, Columbia University Medical Center, New York City, NY, USA
12. Neurodegeneration Program, New York Genome Center, NY, USA
13. Ann Romney Center for Neurologic Diseases, Brigham and Women's hospital, Boston, MA
14. Cell Circuits Program, Broad Institute, Cambridge, Massachusetts, USA
15. Amgen, Cambridge, MA 02141, USA
16. Allogene, San Francisco, CA 94080, USA
17. Agios Pharmaceuticals, Cambridge, MA 02139
18. INTelico Therapeutics, AZ, 85718
19. Rush Alzheimer's Disease Center, Rush University Medical Center, Chicago, IL, USA

Co-corresponding author
* Co-first author

Summary

Late-Onset Alzheimer's Disease (LOAD) results from a complex pathological process influenced by genetic variation, aging and environment factors. Genetic susceptibility factors indicate that myeloid cells such as microglia play a significant role in the onset of LOAD. Here, we developed a computational systems biology approach to construct probabilistic causal and predictive network models of genetic regulatory programs of microglial cells under LOAD diagnosis by integrating two independent brain transcriptome and genome-wide genotype datasets from the Religious Orders Study and Rush Memory and Aging Project (ROSMAP) and Mayo Clinic (MAYO) studies in the AMP-AD consortium. From this network model, we identified and replicated novel microglial-specific master regulators predicted to modulate network states associated with LOAD. We experimentally validated three microglial master regulators (*FCER1G*, *HCK* and *LAPTM5*) in primary human microglia-like cells (MDMi) by demonstrating the molecular impact these master regulators have on modulating downstream genomic targets identified by our top-down/bottom-up method and the causal relations among the three key drivers. These master regulators are involved in phagocytosis, a process associated with LOAD. Thus, we propose three new master regulator (key driver) genes that emerged from our network analyses as robust candidates for further evaluation in LOAD therapeutic development efforts.

Introduction

Late-Onset Alzheimer's disease (LOAD) is a complex neurodegenerative disease that is characterized by neuropathology consisting of amyloid beta (Aß) plaques, neurofibrillary tangles and clinical dementia. Genome-wide association studies (GWAS) have implicated immune cell-specific genes associated with AD risk that point to microglia as a causal cell type [1-18]. Microglial cells are resident innate immune cells of the central nervous system[19] and play an important role in abolishing apoptotic cells, Aß deposits and synapse removal by phagocytosis. It has previously been discovered that microglia are associated with amyloid plaques in the brain[20]. Especially, given the recently failed clinical trials on anti-amyloid plaques[21], it becomes critically important to understand molecule mechanisms of microglia in the formation and clearance of amyloid plaques. A previous study using co-expression network analysis of post-mortem brains from patients with LOAD showed a microglial module dominated by genes implicated in phagocytosis[22].

In this study, we sought to identify cell type specific master regulators (key drivers) modulating network states underlying microglial functions, particularly those involved in phagocytosis and Aβ clearance in the context of AD. We applied the computational framework PSEA [23] to deconvolve bulk-tissue RNA sequencing (RNA-seq) data from post-mortem brain regions to isolate the microglial-specific gene expression signal. The reason we chose the PSEA method over other popular methods,

such as Cibersort [24], dtangle [25], DSA [26] and NNLS [27], is that these methods cannot directly estimate cell-type specific residuals from the bulk-tissue RNA-seq data, instead, these methods only estimate cell fraction in a bulk-tissue sample. We demonstrated the robustness of this de-convolution method using random selection of microglial biomarkers derived from single-cell RNA-seq (scRNAseq) studies [28-32]. Next, we applied a novel systems biology approach to these data to build microglial-specific probabilistic causal network models of the immune component of AD. From these models we identified master regulators of the network states in microglial cells for AD. Among the predicted microglial-specific key drivers, we experimentally validated three novel targets, *HCK, FCER1G, and LAPTM5,* that replicated across our two cohorts using human monocyte-derived microglial-like (MDMi) cells[33]: *HCK* is a member of the Src family of protein tyrosine kinases which couples to Fc receptors during phagocytosis. *FCER1G* is predominantly expressed by hematopoietic cells, encodes for a subunit of an IgE Fc receptor and is implicated as a hub gene in amyloid overexpressing models[34]. Both *FCER1G* and *HCK* were found to be hub genes of microglial modules in a LOAD transcriptomics study[22]. A study by Castillo and colleagues[35] shows that both genes are upregulated in the cortex of $App^{NL\text{-}G\text{-}F/NL\text{-}G\text{-}F}$ transgenic mice as Aβ amyloidosis progresses. *FCER1G* shows a significant association in regards to immune and microglial functions and amyloid deposits in humans and mice[34, 36]. Lastly, *LAPTM5* is associated with lysosomes organization and biogenesis[37-40]. A recent study using a murine amyloid responsive network and GWAS defined association has shown that genetic variations in *LAPTM5* are associated with amyloid deposition in AD[41].

In the present study, by using causal predictive network modeling, we showed that *HCK*, *FCER1G* and *LAPTM5* not only function together in the same co-expression module (subnetwork), but they also play a key driver role in AD through effects on phagocytosis and lysosomal function in microglial cells. In addition, our network model showed that *HCK* is downstream of both *FCER1G* and *LAPTM5*, and that *FCER1G* is downstream of *LAPTM5*. Finally, we validated these predicted relationships and their functions in the MDMi model[33]: we used lentivirus mediated targeted shRNA to knockdown *HCK*, *FCER1G* and *LAPTM5* in MDMi cells and then measured the expression of their downstream genes as predicted by our microglia-specific network model. Not only did our predictive network model accurately predict the genes that changed in response to these perturbations, but it accurately predicted the gene expression dynamics as well of the downstream genomic targets. In addition, we also confirmed a causal role for two of the key drivers as modulators of phagocytic function of microglia cells using an Aß uptake assay.

Results

Integrative Systems Biology Approach for Constructing Single Cell-Type Regulatory Networks of AD

We developed an integrative network analysis pipeline (Figure 1) to construct data-driven microglial-specific predictive networks of AD. The overall strategy for elucidating the single cell-type gene network model depicted in Figure 1 is centered on the objective, data-driven construction of predictive network models of AD that can

then be directly queried to not only identify the network components causally associated with AD, but to identify the master regulators of these components and the impact they have on the expression dynamics of the genes comprising the biological processes underlying AD, moving us towards predictive molecular models of diseases. We previously developed and applied the network reconstruction algorithm, top-down & bottom-up predictive network (predictive network for short), which statistically infers causal relationships between DNA variation, gene expression, protein expression and clinical features that are scored in hundreds of individuals or more[42].

The inputs required for this type of analysis are the molecular and clinical data generated in populations of individuals, as well as first order relationships between these data, such as QTL mapped for the molecular traits and causal relationships among traits inferred by causal mediation analysis that use the mapped QTL as a source of perturbation. These relationships are input as structure priors to the network construction algorithm, boosting the power to infer causal relationships at the network level, as we and others have previously shown [22, 43-50].

To focus on the microglial component of AD, we applied the PSEA de-convolution algorithm [23] to the transcriptomic data to identify the microglial-specific expression component of the transcriptome data (Step 1, Figure S1). We demonstrated this method is robust against random selection of microglia biomarkers derived from single-cell RNA-seq (scRNAseq) studies[28-32]. Given the microglial expression component in the ROSMAP and MAYO populations, we further focused the input of molecular traits into the network reconstruction algorithm on those traits associated

with AD, by identifying AD gene expression signatures comprised of hundreds to thousands of gene expression traits (Step 2, Figure S1). These signatures were enriched for a number of pathways, including mitochondrial and immune processes. To identify gene expression traits co-regulated with the AD signature genes, we constructed gene co-expression networks, and from these networks identified highly interconnected sets of co-regulated genes (modules) that were significantly enriched for the AD expression signatures as well as for pathways previously implicated in AD (Step 3, Figure S1). To obtain a final set of genes for input into the causal network construction process, we combined genes in the co-expression network modules enriched for AD signatures (the seed set, Step 5, Figure S1).

With our AD-centered input set of microglial genes for the network constructions defined, we mapped expression quantitative trait loci (eQTLs) for microglial-specific gene expression traits to incorporate the QTL as structure priors in the network reconstructions, given they provide a systematic perturbation source that can boost the power to infer causal relationships (Step 4, Figure S1)[22, 43, 44, 46-51]. The input gene set, and eQTL data from ROSMAP were then processed by the predictive network to construct probabilistic causal networks of AD (Step 6, Figure S1). An artificial intelligence algorithm to detect key driver genes from these network structures was then applied to identify and prioritize master regulators of the AD networks (Step 7, Figure S1). An in-silico prediction algorithm was developed and used to predict expression profiles upon perturbation. Our findings were then replicated in the MAYO

dataset. For the top regulators we identified, we performed functional and molecular validation in microglial cell systems.

The ROSMAP/Mayo Clinic Study Populations and Data Processing

Our predictive network pipeline starts by integrating whole exome sequencing (WES) and RNA sequencing (RNA-seq) data generated from the dorsolateral prefrontal cortex of 612 persons from ROSMAP[52-55] and from the temporal cortex of 266 patients from MAYO [56-58] in the Accelerating Medicines Partnership - Alzheimer's Disease (AMP-AD) consortium, spanning the complete spectrum of AD clinical and neuropathological traits (Figure 1). We processed matched genotype and RNA-seq data (Online Methods). CNS tissue consists of various cell types, including neurons, endothelial and glial cells. To discover key network drivers, which could serve as therapeutic targets, specific to a single cell type in the CNS and study their contribution to AD in that specific cell type, we utilized well-known/verified single-cell marker genes to directly de-convolve bulk-tissue gene expression data into cell type-specific gene expression for the five primary cell types in the CNS: neurons, microglia, astrocytes, endothelia and oligodendrocytes (Online Methods). In this study, we focused on investigating the role of microglial cells in AD due to their strong genetic association with AD pathogenesis [59-61]. After normalizing RNA-seq data, we evaluated the contribution of demographic, clinical, technical covariates and cell-specific markers to the gene expression variance of the 5 primary cell types using a variance partition analysis (VPA)[62] (Figure 2A). The list of cell type-specific marker genes used for neurons, microglia, astrocytes, endothelia and

oligodendrocytes were *ENO2*, *CD68*, CD34, *GFAP*, and *OLIG2* respectively, as previously published[57].

The rational of using single-gene biomarkers in the above over multi-gene biomarkers derived from scRNAseq data is as follows: 1) multi-gene biomarkers derived from various scRNAseq studies in human brain under control conditions [28-32] shows non-significant (FDR>0.05, Online Method) overlap (Figure S2, 1/0/1/1 significant pair out of 6 study-pairs in Asctrocyte/Endothelial/Microglial/Oligodendrocyte types and 2 significant pairs out of 10 study-pairs in Neuron), indicating lack of robustness and consensus in these biomarkers derived from these studies; 2) significant overlap of scRNAseq-derived biomarkers expression in ROSMAP and MAYO AD samples by PCA analysis (Figure S3), indicating the majority of scRNAseq-derived biomarkers are not ideal in distinguishing cell populations under AD conditions; 3) Significant overlap between scRNAseq-derived biomarkers and AD therapeutic targets (Figure S4) in the AMP-AD AGORA knowledge portal (https://agora.ampadportal.org/genes). This overlap is more significant than randomly selected genes from the background overlapping with the AGORA list (Online method), indicating that scRNAseq-derived biomarkers may play a significant role in AD pathology. Therefore, they are not optimal for adjusting the bulk-tissue gene expression variance by PSEA. By contrast, our single-gene biomarker, though not perfect, is derived from biological knowledge and validated by others [57]. Moreover, our single-gene biomarker had no overlap with AD therapeutic targets in AGORA, which make them good candidates for PSEA. 4) Further, our single-gene biomarker derived microglial-specific residual is significantly (p-value<2.2E-16, Figure

S5, Online Method) correlated with the "pseudo" microglial-specific residuals derived from randomly selected subset of scRNAseq biomarkers by PSEA, indicating that our single-gene biomarker derived microglial-specific residual represents a robust microglial component in the bulk-tissue RNAseq data for following analysis.

In addition to the cell-type specific markers, in ROSMAP, the covariates used in VPA included Sequencing Batch, Exonic Mapping Rate, RNA Integrity Number (RIN), Age-at-death, Age-at-first-AD-diagnosis, Post-Mortem Interval (PMI), Education, *APOE* genotype, Diagnosis, Sex and Study. In MAYO, we included Exonic Mapping Rate, RIN, Sequencing Batch, Diagnosis, Age-at-death, Tissue source, *APOE* genotype and Sex. Next, cell-type specific gene expression residuals were calculated by adjusting the bulk tissue expression data by these covariates and the cell-type specific markers using a linear regression model. To get cell-type specific gene expression, including microglial-specific gene expression, we added the estimated variance of each cell type to the residual (Online Methods). In this way, the cell-type specific gene expression can be directly derived from expression data without the need to first estimate the cell population from bulk tissue data, which could induce approximation errors.

Identifying an AD-associated gene set in microglia and mapping their eQTL

To identify an AD-centered set of microglia gene expression traits, we performed microglial-specific differential expression (DE) analysis in the ROSMAP and MAYO cohorts using the derived microglial-specific expression data (Online Methods). By

comparing expression data from AD and pathologically confirmed controls (CN), there were 513 significantly (FDR<0.05) DE microglial-specific genes in the MAYO dataset (MAYO-microglial for short) and 1,693 microglial-specific genes in the ROSMAP dataset (ROSMAP-microglial for short) (Figure 2B), with 120 significant DE genes overlapping between these two sets (Fisher Exact Test, odd ratio=3.8169, p-value<2.2E-16). To examine the biological processes that are disrupted between AD cases and controls, as reflected in the DE signatures, we performed Pathway Enrichment Analysis (PEA) on the DE gene set for each cohort. We identified 84 and 173 GO terms (Figure 2C), and 15 and 54 KEGG pathways (Figure S6) that are enriched (p-value<0.05) in the MAYO and ROSMAP-microglial data respectively. These pathways include well-known biological functions associated to AD, such as mitochondrial functions, amino acid metabolism, lipid metabolism, glial cell functions, Fc gamma R-mediated phagocytosis, Phagosome etc.

Another critical input for the construction of predictive network models, are the expression quantitative trait loci (eQTL), leveraged as a systematic source of perturbation to enhance causal inference among molecular traits, an approach we and others have demonstrated across a broad range of diseases and data types[22, 43-47, 49-51, 63-77]. We mapped cis-eQTL by examining the association of microglial-specific expression traits with genome-wide genotypes [78-82] assayed in the ROSMAP and MAYO cohorts (Online Methods). In the MAYO-microglial, 3875 (19.5%) of the genes tested were significantly correlated with allele dosage (FDR < 0.01) and in ROSMAP-

microglial, 5186 (25.6%) of the genes tested were significantly correlated with allele dosage (FDR < 0.01). Of the cis-eQTL detected in each cohort, 1785 eQTL (46% in MAYO and 34% in ROSMAP) were overlapping between the two sets.

Microglial-specific Co-expression Network in AD

While DE analysis can reveal patterns of microglial-specific expression associated with AD, the power of such analysis to detect a small-to-moderate expression difference is small. To complement the DE analyses in identifying the input gene set for the predictive network, we clustered the microglial gene expression traits into data-driven, coherent biological pathways by constructing co-expression networks, which have enhanced power to identify co-regulated sets of genes (modules) that are likely to be involved in common biological processes. We constructed co-expression networks on the AD patients for each dataset after filtering out lowly expressed genes (Online Methods). The MAYO-microglial co-expression network consists of 45 modules ranging in size from 32 to 2,201 gene members. The ROSMAP-microglial co-expression network consists of 46 modules ranging in size from 115 to 892 gene members. In comparing all pairs of modules between the datasets for overlap, we identified 133 module pairs with significant overlap (FDR<0.05 for Fisher's Exact Test, Figure 3A).

To characterize the functional relevance of the microglial-specific modules to AD pathology, two measures were employed: 1) Fold-enrichment for AD DE genes in each module; 2) Fold-enrichment for known microglial cell marker genes[31]. From these measures, we selected a seeding set of genes for input into the causal predictive

network modeling by pooling genes from microglial-specific modules: 8 modules from the MAYO-microglial co-expression network (turquoise, pink, greenyellow, tan, mediumblue, darkcyan enriched for AD microglial DE genes; red and beige enriched for microglial cell markers) and 11 modules from the ROSMAP-microglial co-expression network (black, steelblue, brown, yellow, tan, plum1, darkgreen, skyblue3, red, sienna3, and mediumpurple3 enriched for AD microglial DE genes; pink and greenyellow enriched for microglial cell markers).

It is known that microglial cells interplay with other brain cells, such as astrocytes in modulating amyloid pathology in mouse models of Alzheimer's disease[83]. We note that 2 modules (red and darkgreen) are enriched for both microglial-DE signature and astrocyte markers, 1 module (plum1) is enriched for both microglial-DE signature and neuron markers, and 2 modules (mediumpurple3, tan) are enriched for both microglial-DE signature and oligodendrocyte markers from the ROSMAP-microglial co-expression networks, indicating that these interactions were reflected in our microglial-specific data. Failing to account for these interactions will result in a compromised network model, however, over-counting these interactions will decrease the network specificity to microglial cells. Therefore, we only considered microglial-interacting cell types whose marker genes are the most significantly enriched (FDR-value<10E-4) by the same modules that are also the most significantly enriched (FDR-value<10E-4) for microglial-DE signature (Figure S7). Consequently, on top of the selected modules above, we added 2 modules (yellow and honeydrew) and 2 modules (darkgreen and red) enriched

for astrocyte cell marker genes from MAYO-microglial and ROSMAP-microglial co-expression networks. Genes in these additional modules reflecting astrocyte-microglia interactions comprised only 13.9% (583 genes) and 16.9% (873 genes) of the MAYO- and ROSMAP-microglial networks, respectively.

To further annotate the co-expression modules selected above with respect to the biological processes they participate in, we performed pathway enrichment analysis to identify overrepresented biological processes across all selected modules in each dataset. Of the 10 and 13 selected modules from the MAYO- and ROSMAP-microglial networks, 7 and 9, respectively were significantly enriched (FDR<0.05) for KEGG pathways (highlighted red in Figure 3B). There are 104 and 142 significant pathways enriched by the selected modules from the MAYO and ROSMAP networks with an overlapping of 78 significant pathways (Fish's Exact Test, OR=7.445, p-value=1.92E-15, Supplementary File S1), including the phagosome pathway.

Predictive Network Modeling of Genetic Regulations Identified Pathological Pathways and Key Drivers for Microglial Function in AD

The ultimate goal of this study was to identify upstream master regulators (referred to here as key drivers) and pathways in microglia that contribute to AD pathology. Based on our DE, eQTL and co-expression network analysis, we built causal predictive network models on the subset of genes comprising co-expression network modules selected above by integrating the eQTLs and microglial-specific RNA-seq data. To this end, we developed a novel top-down & bottom-up predictive network modeling pipeline (Online

Method) and applied it to the microglia-specific gene expression data in AD. The bottom-up component of our network reconstruction pipeline incorporates known pathway/network relationships derived from the literature and other sources, while the top-down component represents a structure-based learning algorithm that infers causal relationships supported by the eQTL and gene expression data.

To build the predictive network, we first pooled all genes from the subset of 10 and 13 selected modules in the MAYO- and ROSMAP-microglial co-expression networks respectively, to derive a set of seeding genes for each cohort (4187 for MAYO-microglial and 5152 for ROSMAP-microglial co-expression networks). The overlap between the two seeding gene sets contains 1842 genes (35.7% of ROSMAP and 43.9% of MAYO). Therefore, analysis using these two datasets increases the power to build robust networks and to discover high-confidence microglial key drivers that are associated with AD pathology. As *cis*-eQTLs causally affect the expression levels of neighboring genes, they can serve as a source of systematic perturbation to infer causal relationships among genes[51, 84, 85]. Consequently, we incorporated *cis*-eQTL genes into each network as structural priors. Of 5186 and 3875 unique *cis*-eQTL genes identified in the ROSMAP- and MAYO-specific datasets, 1978 and 687 overlapped with genes in each network respectively. The final causal predictive networks were comprised of 4600 and 4008 genes in the ROSMAP- and MAYO-microglial predictive networks (Figure 4A) respectively, with 1646 genes overlapping each network.

Identification of Microglial-specific Key Drivers indicate the Phagosome Contributes to AD

Given the predictive networks, we applied Key Driver Analysis (KDA, Online Method) to derive the list of Key Driver (KD) genes for each network (Figure 4A). KDA seeks to identify those genes in the causal network that modulate network states. In the present analysis, we applied KDA to microglial-specific predictive network models to identify those genes causally modulating the states of these networks. In total, we identified 757 and 164 KD genes identified in the ROSMAP- and MAYO-microglial networks respectively and 43 KDs replicated across both networks. We prioritized all key drivers based on i) whether a gene predicted as KD is replicated across both datasets; ii) the number of different categories of target source predicting a gene as KD used by KDA (In this study, we used 3 kinds of targets for KDA: DE genes, modules and overlapping genes of DE gene with modules); and iii) the number of different target sources used by KDA predicts a gene as a KD.

Among the pathways enriched by the 43 KDs (Figure S8), *Microglial Pathogen Phagocytosis Pathway* is the most significantly enriched (p-value=2.60E-12) biological function by these 43 replicated KDs. We identified three KD genes, *FCER1G*, *HCK* and *LAPTM5* which are known to be involved in the phagosome as driving the AD phenotype, indicating that microglia-mediated phagocytosis causally links to the AD phenotype. Though these three genes are known to be involved in phagocytosis, their molecular mechanism in phagocytosis and the causal relationship among them is still

unclear. To illustrate the molecular mechanism of these genes in phagocytosis, we extracted the downstream sub-network of these three KDs in the ROSMAP- and MAYO-microglial networks respectively (Figure 4B), and then performed CPDB pathway enrichment analysis[86], which confirmed that both sub-networks are significantly ($p<0.05$) enriched for Phagocytosis and Phagosome Pathways (Supplementary File S2). In addition, we predicted their causal relationship with the AD phenotype. In the MAYO-microglial sub-network, there are a total of 189 enriched pathways, including Microglia Pathogen Phagocytosis Pathway (p-value=0.0167), ER-Phagosome pathway (p-value=0.018) and Phagosome (p-value=0.028). In the ROSMAP-microglial sub-network, there are 534 enriched pathways, including Microglia Pathogen Phagocytosis Pathway (p-value=6.06E-11), Cross-presentation of particulate exogenous antigens (phagosomes) (p-value=1.81E-05), Fc gamma R-mediated phagocytosis (p-value=1.94E-05), Fc-gamma receptor (FCGR) dependent phagocytosis (p-value=4.03E-04), Role of phospholipids in phagocytosis (p-value=9.55E-04), Phagosome (p-value=2.24E-03), ER-Phagosome pathway (p-value=0.03). The overlapping contains 77 pathways, including Microglia Pathogen Phagocytosis Pathway, Phagosome, and ER-Phagosome pathway.

Validation of the microglial key driver and network for AD pathology by knockdown in monocyte-derived microglia-like cells

To validate the molecular mechanism and pathways of these three key drivers captured by our predictive network model, we first sought to validate the downstream and upstream genes of microglial-specific key drivers (*HCK*, *FCER1G*, *LAPTM5*) predicted by

our network model. To address this, we utilized the previously characterized highly efficient *in vitro* cell model system composed of human monocyte-derived microglia-like cells (MDMi) that recapitulates key aspects of microglia phenotype and function[33]. Using this model, we applied three different constructs of short hairpin lentiviral knockdown vectors targeting different parts of each KD gene by RNA interference (shHCK, shFCER1G and shLAPTM5) to MDMi cells differentiated from 3 healthy donors. We then picked two of the three constructs that gave at least 70-90% knockdown in the gene of interest compared to a control shRNA as verified by qPCR (Figure S9). To validate the predictive network, we measured the gene expression between MDMi cells that received the shHCK, shFCER1G or, shLAPTM5 and empty vector (shCTRL) for six network-predicted immediate downstream genes for *HCK*, *FCER1G* and *LAPTM5* respectively using Taqman real-time PCR (Figure 5A). Out of all immediate downstream genes of the KDs predicted by the two microglia networks, the six downstream genes were chosen on the basis of their expression level in the MDMi cells as determined by RNA-seq data set on MDMi (Supplementary File S3). For *HCK*, we selected *CASP1*, *FCGR1A*, *CYTH4 and NCF4*, *LAIR1*, *FCGR2A* from the MAYO- and ROSMAP-microglia networks respectively. Of the six genes, knockdown of the *HCK* gene in MDMi cells led to statistically significant reduction (p-value<0.05) (Online Method) in gene expression of five downstream genes in at least one construct: *CYTH4*, *NCF4*, *LAIR1*, *FCGR1A* and *FCGR2A* as compared to MDMi cells that received the control vector (ShCTRL). There was no significant decrease in gene expression from *CASP1* in either construct, though they trended in the same direction (p-values=5.87E-02 and 6.62E-02). For *FCER1G*, we

selected *FERMT3*, *CD14*, *SPI-1*, *LAIR1* and *S100A11*, *FXYD5* from the MAYO- and ROSMAP-microglia networks. Five out of six downstream genes showed a significant reduction in gene expression in shFCER1G MDMi cells as compared to MDMi cells that received the control vector (ShCTRL) in at least one construct. *S100A11* showed no significant decrease in gene expression with p-values (5.68E-02 and 7.69E-02) close to the significance threshold. Similarly, we also conducted experiments using shRNA targeting the third key driver *LAPTM5*. For *LAPTM5*, we selected *NFAM, TYROBP, HLA-DRA, CD68 and NFAM, TYROBP, ITGB2, SPI-1* from MAYO- and ROSMAP-microglia networks. Knockdown of *LAPTM5* significantly upregulates *ITGB2*, *HLA-DRA* and *CD68*, and significantly downregulates *SPI-1*. Furthermore, knockdown of *LAPTM5* in MDMI cells did not have any significant effect on the expression of *TYROBP* (p-value=7.61E-01 and 5.44E-01) and *NFAM* (p-value=7.27E-02 and 6.07E-02), which is close to the significance threshold. *TYROBP* in our network contains 2 and 5 parent genes respectively in the MAYO- and ROSMAP-microglia networks. In the ROSMAP-microglia network, it has significantly more parents than all other nodes in the network (p-value=2.2e-16), potentially explaining why modulation of *LAPTM5* does not regulate *TYROBP* in our model system. In addition, we measured the gene expression between MDMi cells that received the shHCK, shFCER1G or, shLAPTM5 and empty vector (shCTRL) for four network-predicted common upstream genes (*C1ORF162*, *CKLF*, *SLC37A2*, *RMDN1*) for the three KDs respectively using Taqman (Figure 5B). As predicted by the network, none of these genes are significantly changed. The experimental results are listed in Supplementary File S4 (sheet 'KD, downstream genes, results').

Validation of In-silico Prediction on Gene Expression Prediction by Microglial-specific Predictive Networks

Next, we predicted in-silico gene expression of downstream genomic targets from these three key drivers based on our predictive network model. We developed the in-silico phenotypic prediction pipeline, which is a component of the predictive network, to predict i)qualitative response, i.e. direction of gene expression fold-change and ii)quantitative response, i.e. expression fold-change of the downstream genes of three key drivers in the ROSMAP and MAYO-microglial networks. In shRNA experiments, we measured gene expression fold-change of 11 and 9 KD-target pairs (see last section) predicted by the MAYO- and ROSMAP-microglial networks with two constructs in MDMi cells from 3 healthy donors. Therefore, there are a total 66 and 54 measurements in the MAYO- and ROSMAP-microglia network respectively. First, we predicted the direction of gene expression change (up-regulation or down-regulation) and compared it to those in the shRNA experiments. Out of the 66 measurements, all predicted to be down-regulated by the MAYO network, 22 measurements were up-regulated and 44 were down-regulated in the shRNA experiment. Out of 54 measurements predicted to be down-regulated by the ROSMAP network, 18 were up-regulated and 36 were down-regulated in the shRNA experiment. The accuracy of the qualitative response prediction is 66.7% for both networks. Next, we calculated the Pearson correlation (0.423 and 0.508, p-value=3.97e-04 and 8.85e-05) and Spearman correlation (0.367 and 0.436, p-value=2.61e-03 and 1.1e-03) between predicted and experimental expression fold-change (Online method, Supplementary File S5) in the MAYO- and ROSMAP-microglia

networks (Figure 5C). Further, if we only consider measurements (p-value<0.05) in shRNA experiments, 28 and 28 pairs are used to calculate the Pearson correlation (0.549 and 0.622, p-value=2.47e-03 and 4.06e-04) and Spearman correlation (0.358 and 0.492, p-value=6.21e-02 and 8.51e-03) in the MAYO- and ROSMAP-microglia networks (Figure 5D). This result demonstrated that our predictive network model accurately predicts downstream gene expression changes upon perturbing any key driver, which is validated using Taqman.

Validation of the causal regulation among HCK, FCER1G and LAPTM5 by knockdown in MDMi cells

Next, we identified previously unknown causal regulation among the three key drivers in the process of phagocytosis using our microglia-specific predictive network models in ROSMAP and MAYO. We extracted sub-networks of phagocytosis around the 3 key drivers (highlighted in Figure 4B), and determined that *HCK* was downstream of *FCER1G*, which is, in turn, downstream of *LAPTM5*. These causal regulations are replicated in MAYO- and ROSMAP-microglial networks. Further, to validate the relationship between HCK, FCER1G and LAPTM5, we applied shRNA targeting FCER1G and *LAPTM5* using lentivirus in MDMi cells. We then measured *HCK*, *LAPTM5* and *FCER1G* expression in each knockdown. Gene expression data from the shFCER1G cells showed significant reduction (p-value=2.09E-4) in *HCK* gene but there was no significant (p-value=1.62E-01) change in LAPTM5 expression. In addition, gene expression data from shLAPTM5 showed significant reduction in HCK (p-value=3.58E-04) and FCER1G (p-value=2.12E-03)

(Figure 4C), thereby indicating that LAPTM5 was upstream of FCER1G and HCK was downstream of FCER1G, which validated our microglia-specific network. Besides HCK and FCER1G, ITGB2 and SYK gene are also significantly changed their expression level in both knockdown experiments, suggesting that they are directly or indirectly downstream of LAPTM5 and FCER1G (Figure 4C). Interestingly, both our MAYO and ROSMAP sub-networks show that ITGB2 is downstream of LAPTM5, and in the ROSMAP sub-network, ITGB2 is an indirect downstream gene of FCER1G. In addition, in the ROSMAP sub-network, SYK is direct downstream gene of LAPTM5 and indirect downstream gene of FCER1G (Figure 4B). The experimental results are listed in Supplementary Table S4 (sheet 'KD, subnetworks, results').

Functional Validation of HCK and FCER1G as key drivers of microglia function of Aβ clearance

Microglia are innate immune cells of the brain which play an important role in clearance of amyloid-beta which forms plaques in the brain, a hallmark pathology of AD. In this study, we identified key drivers in our predicted microglia-specific causal network model for phagocytosis (HCK and FCER1G) and lysosome function (LAPTM5). In order to validate the function of HCK and FCER1G as mediators of phagocytosis, we analyzed the Aß42 uptake ability of MDMi cells differentiated from 4-5 healthy donors that received lentivirus containing the short hair pin RNA targeting HCK (shHCK) and FCER1G (shFCER1G) and compared it to the MDMi cells from the same donors that received empty control virus (shCTRL) using fluorescently labelled Aß42 peptide. The MDMi cells

with both shHCK and shFCER1G show a significant decrease in Aß42 uptake for the two *HCK* constructs (paired t-test, p-value = 0.030 and 0.028) and for the two *FCER1G* constructs (p-value=0.037 and 0.045 (Figure 6). As *LAPTM5* is a lysosomal molecule, we do not expect to see an effect in our uptake assay. Thus, we functionally validated these genes as key drivers of phagocytosis in microglia cells as predicted by our network specific model. The complete results are in Supplementary File S4 (sheet 'Abeta uptake, results').

Discussion

GWAS studies in LOAD have identified several microglial specific genes. However, it is unclear how these genes interact and what cellular pathways are involved in the pathology of LOAD. Hence, a comprehensive characterization of gene regulatory networks with association to disease is important to provide insights into the underlying mechanisms of complex neurodegenerative diseases such as Alzheimer's disease. Our study uses an innovative predictive computational systems biology model to identify upstream regulators (key drivers) and cellular pathways in microglial cells that contribute to AD pathology using the Mayo Clinic and ROSMAP datasets in the AMP-AD consortium. The RNA-seq data from brain-region tissue in MAYO (TCX-Temporal Cortex, 79 AD and 76 CN) and ROSMAP (PFC-Pre-frontal Cortex, 212 AD and 200 CN) cohorts are computationally de-convoluted into single cell types including neurons, microglia, astrocytes, endothelial cells and oligodendrocytes. In this study, we focused on microglial-specific gene expression data. We performed preliminary bioinformatics

analysis including differential expression, eQTLs, co-expression networks and pathway analysis prior to building predictive causal network models. There are 4187(43.9%)/5152(35.7%) (overlap percentage in parenthesis) genes in the seeding gene list of MAYO-/ROSMAP-microglial predictive network models and 4008(41.0%)/4600(35.7%) genes in the final MAYO-/ROSMAP-microglial predictive network models. The difference of key drivers/pathways in each predictive network can be attributed to: i) different susceptibility and pathology response to AD in TCX and DLPFC regions analyzed in the MAYO and ROSMAP cohorts respectively; ii) different patient composition in the two cohorts, such as Male/Female, APOE4+/- and clinical stage; and iii) technical variance in sample extraction, RNA preparation, and RNA-sequencing as well as other covariates. However, the replicated key drivers/pathways derived by these network models identified robust biological processes and key drivers in microglial cells under AD diagnosis despite the significant variance in data and cohorts as described above. Consequently, our predictive networks identified robust key drivers for phagocytosis in microglial-specific cells associated with AD, which are validated using an *in vitro* model of microglia.

Our novel predictive network-based analysis integrated microglial cell-specific genetics and genomics data to identify key regulatory genes associated with microglial functions in AD, i.e. *LAPTM5*, *FCER1G* and *HCK*. Pathway enrichment analysis confirmed that all three key drivers and their downstream genes in the network model regulate phagosome processes. The phagosome pathway is activated upon neuronal loss or

amyloid plaque buildup, two important pathophysiological hallmarks of AD. Microglia clear synapses and neurites during development and in neurodegenerative processes using phagocytosis via C1Q, C3, CR3, and the DAP12/TYROBP cascade[87-89]. Microglia are also closely associated with amyloid beta plaques making it important to understand the cause-and-effect relationship between immune cells and AD progression[90].

Our network model predicted regulation between FCER1G, HCK and LAPTM5 genes. We highlight that HCK is downstream of FCER1G in the network with our lentiviral shRNA knockdown in MDMi cells (Figure 4C). A study by Taguchi et al[36] showed that HCK and FCER1G are up-regulated in the cortex of *App*$^{NL\text{-}G\text{-}F/NL\text{-}G\text{-}F}$ mice as Aβ amyloidosis progressed thereby associating them with plaques and phagocytosis. Furthermore, FCER1G shows significant association in AD, in regards to immune and microglial functions and amyloid deposits in humans and mice[34-36]. The reduction in uptake of the fluorescently labeled amyloid beta 1-42 in shFCER1G and shHCK microglia cells indicates the significance of these genes as phagocytic modulators in microglia cells, which validated the accuracy of our predictive network model. Furthermore, while *LAPTM5* is upstream of *FCER1G* and *HCK*, its main role is in lysosomal function, which is not captured in our assay, and thus modulating *LAPTM5* doesn't show a robust effect on uptake of Aß42 (Supplementary File S4, sheet 'Abeta uptake, results').

GWAS studies have implicated SNPs and polymorphisms in the TYROBP binding protein TREM2, ITGAM and SPI-1 genes as being associated with LOAD[91]. Using our

computational predictive network, we also highlight that TYROBP is downstream of LAPTM5 and TREM2 and hence is regulated by both genes, while ITGAM is upstream of LAPTM5, FCER1G and HCK. We further demonstrated that SPI-1, an established GWAS loci for AD, is a downstream gene of our key driver LAPTM5 in a microglial specific network from the ROSMAP/MAYO data sets. Our study demonstrates the importance of these classical immune genes in AD functions.

Comparative profiling of human cortical gene expression in AD patients[35] and mouse models with amyloid beta plaque accumulation have shown involvement of our key drivers LAPTM5, FCER1G and HCK. These three genes were identified in Zhang et al. (ref.), however here we demonstrate that they exist in one causal network, and we have delineated the upstream and downstream relationships. A recent study by Lim et al. shows that inhibition of HCK dysregulates microglial function of phagocytosis and enhances amyloid plaque build-up in the J-20 mouse model of AD[92]. In addition, a recent study[41] showed that LAPTM5 is significantly associated with the mouse amyloid response network and that it's human ortholog contains SNPs associated with AD.

In addition to network reconstruction and key driver discovery, our predictive network model is capable to perform in-silico phenotype prediction. Upon perturbing any number of genes in the network, we can predict i) whether a given gene in the network will significantly change their expression level; ii) the qualitative response, i.e. direction

of gene expression change of the downstream genes to the perturbation; iii) the quantitative response, i.e. log fold-change of gene expression in the downstream genes to perturbation. We used two shRNAs targeting different regions of the three key driver genes and tested gene expression change of 18 downstream genes of these three key drivers in the network using Taqman array. Fourteen out of 18 (78%) predicted-to-change downstream genes by our network models are validated as significantly ($p<0.05$) altered in their gene expression by knockdown experiments. In addition, 4 common upstream genes of these key drivers predicted not-to-change by the networks, are 100% validated as not altering in expression after knockdown. The accuracy of the qualitative response prediction is 66.7% for the 18 downstream target genes in MAYO and ROSMAP-microglial networks. The Pearson correlation between experimental data and quantitative prediction by the model is very significant for all measured downstream targets in MAYO- and ROSMAP-microglial network respectively. This result demonstrated that our predictive network model is capable of further predicting phenotypic changes upon perturbations in the model.

Overall, our innovative computational systems biology modeling of microglia specific networks further deepens our understanding of microglial-specific implications in AD pathology by identifying robust causal networks of key driver genes and their genomic target genes for the phagosome, including a GWAS AD-related gene (LAPTM5) that has major implications in AD. Our predictive network not only identifies FCER1G, HCK and LAPTM5 as key drivers for microglial specific genes in the phagosome pathway but also

demonstrates the functional association of HCK and FCER1G in amyloid beta uptake in microglial cells. Our approach appears to offer novel insights for drug discovery programs that can affect neurodegenerative diseases, such as LOAD.

Author contribution:

Conceptualization: R.C. E.M.B and E.E.S; ROSMAP RNA-seq and WGS pre-processing: N.D.M. M.Y.R.H.; MAYO RNA-seq and WGS pre-processing: M.A. X.W. J.S.R M. M.Y.R.H; Single Cell-type Gene Expression: R.C., M.Y.R.H., N.E.T; Data Analysis: R.C., M.Y.R.H., K.Z. S.M. M.L.A; MDMi shRNA experiment design and implementation: K.P., E.M.B. P.N. provided technical assistance. G.C., T.P., K.P. and E.M.B created and optimized the shRNA protocol. T.P. provided plasmids; Manuscript writing and figures: R.C. K.P. M.Y.R.C. K.Z. N.D.M; Manuscript Editing: E.M.B P.L.D. E.E.S. N.E.T. D.A.B.

Acknowledgement:

The authors thank the patients and their families for their participation, without whom these studies would not have been possible. This work was supported by the National Institute of Health: NIA R01 AG057457, NIA R01 AG059093 and NIA 1R01AG057931 to R.C; U01 AG046152, R01 AG048015, RF1 AG015819, R01AG056284, R01AG036836 R01AG043617 to PLD; NINDS R01 NS080820, NIA RF AG051504 and NIA U01 AG046139 to N.E.T; R01AG058852, R01NS089674, RF1AG057457, R01AG043617 to E.M.B; NIA U01AG058635 and NIMH R01MH109897 to E.E.S.

References

1. Gaikwad, S., et al., *Signal regulatory protein-beta1: a microglial modulator of phagocytosis in Alzheimer's disease.* Am J Pathol, 2009. **175**(6): p. 2528-39.
2. Jonsson, T., et al., *Variant of TREM2 associated with the risk of Alzheimer's disease.* N Engl J Med, 2013. **368**(2): p. 107-16.
3. Replogle, J.M., et al., *A TREM1 variant alters the accumulation of Alzheimer-related amyloid pathology.* Ann Neurol, 2015. **77**(3): p. 469-77.
4. Sperling, R.A., et al., *Toward defining the preclinical stages of Alzheimer's disease: recommendations from the National Institute on Aging-Alzheimer's*

Association workgroups on diagnostic guidelines for Alzheimer's disease. Alzheimers Dement, 2011. **7**(3): p. 280-92.

5. Morris, J.C., et al., *The consortium to establish a registry for Alzheimer's disease (CERAD). Part IV. Rates of cognitive change in the longitudinal assessment of probable Alzheimer's disease.* Neurology, 1993. **43**(12): p. 2457-65.
6. Brookmeyer, R., et al., *National estimates of the prevalence of Alzheimer's disease in the United States.* Alzheimers Dement, 2011. **7**(1): p. 61-73.
7. Bradshaw, E.M., et al., *CD33 Alzheimer's disease locus: altered monocyte function and amyloid biology.* Nat Neurosci, 2013. **16**(7): p. 848-50.
8. Chan, G., et al., *CD33 modulates TREM2: convergence of Alzheimer loci.* Nat Neurosci, 2015. **18**(11): p. 1556-8.
9. Raj, T., et al., *Alzheimer disease susceptibility loci: evidence for a protein network under natural selection.* Am J Hum Genet, 2012. **90**(4): p. 720-6.
10. Raj, T., et al., *Polarization of the effects of autoimmune and neurodegenerative risk alleles in leukocytes.* Science, 2014. **344**(6183): p. 519-23.
11. Guerreiro, R., et al., *TREM2 variants in Alzheimer's disease.* N Engl J Med, 2013. **368**(2): p. 117-27.
12. Hollingworth, P., et al., *Common variants at ABCA7, MS4A6A/MS4A4E, EPHA1, CD33 and CD2AP are associated with Alzheimer's disease.* Nat Genet, 2011. **43**(5): p. 429-35.
13. Naj, A.C., et al., *Common variants at MS4A4/MS4A6E, CD2AP, CD33 and EPHA1 are associated with late-onset Alzheimer's disease.* Nat Genet, 2011. **43**(5): p. 436-41.
14. Seshadri, S., et al., *Genome-wide analysis of genetic loci associated with Alzheimer disease.* JAMA, 2010. **303**(18): p. 1832-40.
15. Lambert, J.C., et al., *Genome-wide association study identifies variants at CLU and CR1 associated with Alzheimer's disease.* Nat Genet, 2009. **41**(10): p. 1094-9.
16. Wyndham, R.C., R.K. Singh, and N.A. Straus, *Catabolic instability, plasmid gene deletion and recombination in Alcaligenes sp. BR60.* Arch Microbiol, 1988. **150**(3): p. 237-43.
17. Harold, D., et al., *Genome-wide association study identifies variants at CLU and PICALM associated with Alzheimer's disease.* Nat Genet, 2009. **41**(10): p. 1088-93.
18. Hollingworth, P., et al., *Alzheimer's disease genetics: current knowledge and future challenges.* Int J Geriatr Psychiatry, 2011. **26**(8): p. 793-802.
19. Cullheim, S. and S. Thams, *The microglial networks of the brain and their role in neuronal network plasticity after lesion.* Brain Res Rev, 2007. **55**(1): p. 89-96.
20. Cameron, B. and G.E. Landreth, *Inflammation, microglia, and Alzheimer's disease.* Neurobiol Dis, 2010. **37**(3): p. 503-9.
21. Mullard, A., *Anti-amyloid failures stack up as Alzheimer antibody flops.* Nat Rev Drug Discov, 2019.
22. Zhang, B., et al., *Integrated systems approach identifies genetic nodes and networks in late-onset Alzheimer's disease.* Cell, 2013. **153**(3): p. 707-20.

23. Kuhn, A., et al., *Population-specific expression analysis (PSEA) reveals molecular changes in diseased brain.* Nat Methods, 2011. **8**(11): p. 945-7.
24. Newman, A.M., et al., *Robust enumeration of cell subsets from tissue expression profiles.* Nat Methods, 2015. **12**(5): p. 453-7.
25. Hunt, G.J., et al., *dtangle: accurate and fast cell-type deconvolution.* bioRxiv, 2018.
26. Zhong, Y., et al., *Digital sorting of complex tissues for cell type-specific gene expression profiles.* BMC Bioinformatics, 2013. **14**: p. 89.
27. Abbas, A.R., et al., *Deconvolution of blood microarray data identifies cellular activation patterns in systemic lupus erythematosus.* PLoS One, 2009. **4**(7): p. e6098.
28. Darmanis, S., et al., *A survey of human brain transcriptome diversity at the single cell level.* Proc Natl Acad Sci U S A, 2015. **112**(23): p. 7285-90.
29. Zhang, Y., et al., *Purification and Characterization of Progenitor and Mature Human Astrocytes Reveals Transcriptional and Functional Differences with Mouse.* Neuron, 2016. **89**(1): p. 37-53.
30. Lake, B.B., et al., *Neuronal subtypes and diversity revealed by single-nucleus RNA sequencing of the human brain.* Science, 2016. **352**(6293): p. 1586-90.
31. Zeisel, A., et al., *Brain structure. Cell types in the mouse cortex and hippocampus revealed by single-cell RNA-seq.* Science, 2015. **347**(6226): p. 1138-42.
32. Zhang, Y., et al., *An RNA-sequencing transcriptome and splicing database of glia, neurons, and vascular cells of the cerebral cortex.* J Neurosci, 2014. **34**(36): p. 11929-47.
33. Ryan, K.J., et al., *A human microglia-like cellular model for assessing the effects of neurodegenerative disease gene variants.* Sci Transl Med, 2017. **9**(421).
34. Matarin, M., et al., *A genome-wide gene-expression analysis and database in transgenic mice during development of amyloid or tau pathology.* Cell Rep, 2015. **10**(4): p. 633-44.
35. Castillo, E., et al., *Comparative profiling of cortical gene expression in Alzheimer's disease patients and mouse models demonstrates a link between amyloidosis and neuroinflammation.* Sci Rep, 2017. **7**(1): p. 17762.
36. Taguchi, K., et al., *Identification of hippocampus-related candidate genes for Alzheimer's disease.* Ann Neurol, 2005. **57**(4): p. 585-8.
37. Vergarajauregui, S., J.A. Martina, and R. Puertollano, *LAPTMs regulate lysosomal function and interact with mucolipin 1: new clues for understanding mucolipidosis type IV.* J Cell Sci, 2011. **124**(Pt 3): p. 459-68.
38. Pak, Y., et al., *Transport of LAPTM5 to lysosomes requires association with the ubiquitin ligase Nedd4, but not LAPTM5 ubiquitination.* J Cell Biol, 2006. **175**(4): p. 631-45.
39. Adra, C.N., et al., *LAPTM5: a novel lysosomal-associated multispanning membrane protein preferentially expressed in hematopoietic cells.* Genomics, 1996. **35**(2): p. 328-37.
40. Scott, L.M., L. Mueller, and S.J. Collins, *E3, a hematopoietic-specific transcript directly regulated by the retinoic acid receptor alpha.* Blood, 1996. **88**(7): p. 2517-30.

41. Salih, D.A.a.B., Sevinc and Guelfi, Manuel S and Reynolds, Regina H and Shoai, Maryam and Ryten, Mina and Brenton, Jonathan and Zhang, David and Matarin, Mar and Botia, Juan and Shah, Runil and Brookes, Keeley and Guetta-Baranes, Tamar and Morgan, Kevin and Bellou, Eftychia and Cummings, Damian M and Hardy, John and Edwards, Frances A and Escott-Price, Valentina, *Genetic variability in response to Abeta deposition influences Alzheimer's risk.* bioRxiv, 2018.
42. Petyuk*, V.A., et al., *THE HUMAN BRAINOME: Predictive Network Analysis identifies HSPA2 as a novel Alzheimer's disease target.* Brain, 2018. **Minor revision**.
43. Franzen, O., et al., *Cardiometabolic risk loci share downstream cis- and trans-gene regulation across tissues and diseases.* Science, 2016. **353**(6301): p. 827-30.
44. Peters, L.A., et al., *A functional genomics predictive network model identifies regulators of inflammatory bowel disease.* Nat Genet, 2017. **49**(10): p. 1437-1449.
45. Doss, S., et al., *Cis-acting expression quantitative trait loci in mice.* Genome Res, 2005. **15**(5): p. 681-91.
46. Yang, X., et al., *Validation of candidate causal genes for obesity that affect shared metabolic pathways and networks.* Nat Genet, 2009. **41**(4): p. 415-23.
47. Zhu, J., et al., *An integrative genomics approach to the reconstruction of gene networks in segregating populations.* Cytogenet Genome Res, 2004. **105**(2-4): p. 363-74.
48. Zhu, J., et al., *Stitching together multiple data dimensions reveals interacting metabolomic and transcriptomic networks that modulate cell regulation.* PLoS Biol, 2012. **10**(4): p. e1001301.
49. Zhu, J., et al., *Increasing the power to detect causal associations by combining genotypic and expression data in segregating populations.* PLoS Comput Biol, 2007. **3**(4): p. e69.
50. Zhu, J., et al., *Integrating large-scale functional genomic data to dissect the complexity of yeast regulatory networks.* Nat Genet, 2008. **40**(7): p. 854-61.
51. Schadt, E.E., et al., *An integrative genomics approach to infer causal associations between gene expression and disease.* Nat Genet, 2005. **37**(7): p. 710-7.
52. Bennett, D.A., et al., *Overview and findings from the rush Memory and Aging Project.* Curr Alzheimer Res, 2012. **9**(6): p. 646-63.
53. Bennett, D.A., et al., *Overview and findings from the religious orders study.* Curr Alzheimer Res, 2012. **9**(6): p. 628-45.
54. De Jager, P.L., et al., *A multi-omic atlas of the human frontal cortex for aging and Alzheimer's disease research.* Sci Data, 2018. **5**: p. 180142.
55. Mostafavi, S., et al., *A molecular network of the aging human brain provides insights into the pathology and cognitive decline of Alzheimer's disease.* Nat Neurosci, 2018. **21**(6): p. 811-819.
56. Allen, M., et al., *Divergent brain gene expression patterns associate with distinct cell-specific tau neuropathology traits in progressive supranuclear palsy.* Acta Neuropathol, 2018. **136**(5): p. 709-727.

57. Allen, M., et al., *Conserved brain myelination networks are altered in Alzheimer's and other neurodegenerative diseases.* Alzheimers Dement, 2018. **14**(3): p. 352-366.
58. Allen, M., et al., *Human whole genome genotype and transcriptome data for Alzheimer's and other neurodegenerative diseases.* Sci Data, 2016. **3**: p. 160089.
59. Gosselin, D., et al., *An environment-dependent transcriptional network specifies human microglia identity.* Science, 2017. **356**(6344).
60. Tansey, K.E., D. Cameron, and M.J. Hill, *Genetic risk for Alzheimer's disease is concentrated in specific macrophage and microglial transcriptional networks.* Genome Med, 2018. **10**(1): p. 14.
61. Olah, M., et al., *A transcriptomic atlas of aged human microglia.* Nat Commun, 2018. **9**(1): p. 539.
62. Hoffman, G.E. and E.E. Schadt, *variancePartition: interpreting drivers of variation in complex gene expression studies.* BMC Bioinformatics, 2016. **17**(1): p. 483.
63. Schadt, E.E., et al., *Genetics of gene expression surveyed in maize, mouse and man.* Nature, 2003. **422**(6929): p. 297-302.
64. Mehrabian, M., et al., *Integrating genotypic and expression data in a segregating mouse population to identify 5-lipoxygenase as a susceptibility gene for obesity and bone traits.* Nat Genet, 2005. **37**(11): p. 1224-33.
65. Ghazalpour, A., et al., *Integrating genetic and network analysis to characterize genes related to mouse weight.* PLoS Genet, 2006. **2**(8): p. e130.
66. Yang, X., et al., *Tissue-specific expression and regulation of sexually dimorphic genes in mice.* Genome Res, 2006. **16**(8): p. 995-1004.
67. Emilsson, V., et al., *Genetics of gene expression and its effect on disease.* Nature, 2008. **452**(7186): p. 423-8.
68. Dobrin, R., et al., *Multi-tissue coexpression networks reveal unexpected subnetworks associated with disease.* Genome Biol, 2009. **10**(5): p. R55.
69. Zhang, W., et al., *A Bayesian partition method for detecting pleiotropic and epistatic eQTL modules.* PLoS Comput Biol, 2010. **6**(1): p. e1000642.
70. Zhong, H., et al., *Liver and adipose expression associated SNPs are enriched for association to type 2 diabetes.* PLoS Genet, 2010. **6**(5): p. e1000932.
71. Zhong, H., et al., *Integrating pathway analysis and genetics of gene expression for genome-wide association studies.* Am J Hum Genet, 2010. **86**(4): p. 581-91.
72. Millstein, J., et al., *Identification of causal genes, networks, and transcriptional regulators of REM sleep and wake.* Sleep, 2011. **34**(11): p. 1469-77.
73. Schadt, E.E., S. Woo, and K. Hao, *Bayesian method to predict individual SNP genotypes from gene expression data.* Nat Genet, 2012. **44**(5): p. 603-8.
74. Tu, Z., et al., *Integrative analysis of a cross-loci regulation network identifies App as a gene regulating insulin secretion from pancreatic islets.* PLoS Genet, 2012. **8**(12): p. e1003107.
75. Roussos, P., et al., *A role for noncoding variation in schizophrenia.* Cell Rep, 2014. **9**(4): p. 1417-29.
76. Chang, R., J.R. Karr, and E.E. Schadt, *Causal inference in biology networks with integrated belief propagation.* Pac Symp Biocomput, 2015: p. 359-70.

77. Miller, C.L., et al., *Integrative functional genomics identifies regulatory mechanisms at coronary artery disease loci.* Nat Commun, 2016. **7**: p. 12092.
78. Myers, A.J., *The age of the "ome": genome, transcriptome and proteome data set collection and analysis.* Brain Res Bull, 2012. **88**(4): p. 294-301.
79. Myers, A.J., *AD gene 3-D: moving past single layer genetic information to map novel loci involved in Alzheimer's disease.* J Alzheimers Dis, 2013. **33 Suppl 1**: p. S15-22.
80. Myers, A.J., *The Genetics Of Gene Expression: Multiple Layers and Multiple Players*, in
The OMICs: Applications in Neuroscience
, G. Coppola, Editor. 2014, Oxford University Press: New York, NY. p. 132-152.
81. Myers, A.J., et al., *A survey of genetic human cortical gene expression.* Nat Genet, 2007. **39**(12): p. 1494-9.
82. Webster, J.A., et al., *Genetic control of human brain transcript expression in Alzheimer disease.* Am J Hum Genet, 2009. **84**(4): p. 445-58.
83. Lian, H., et al., *Astrocyte-Microglia Cross Talk through Complement Activation Modulates Amyloid Pathology in Mouse Models of Alzheimer's Disease.* J Neurosci, 2016. **36**(2): p. 577-89.
84. Petyuk, V.A., et al., *The human brainome: network analysis identifies HSPA2 as a novel Alzheimer’s disease target.* Brain, 2018. **141**(9): p. 2721-2739.
85. Carcamo-Orive, I., et al., *Analysis of Transcriptional Variability in a Large Human iPSC Library Reveals Genetic and Non-genetic Determinants of Heterogeneity.* Cell Stem Cell, 2017. **20**(4): p. 518-532.
86. Kamburov, A., et al., *ConsensusPathDB: toward a more complete picture of cell biology.* Nucleic Acids Res, 2011. **39**(Database issue): p. D712-7.
87. Schafer, D.P., et al., *Microglia sculpt postnatal neural circuits in an activity and complement-dependent manner.* Neuron, 2012. **74**(4): p. 691-705.
88. Wakselman, S., et al., *Developmental neuronal death in hippocampus requires the microglial CD11b integrin and DAP12 immunoreceptor.* J Neurosci, 2008. **28**(32): p. 8138-43.
89. Stevens, B., et al., *The classical complement cascade mediates CNS synapse elimination.* Cell, 2007. **131**(6): p. 1164-78.
90. Meyer-Luehmann, M., et al., *Rapid appearance and local toxicity of amyloid-beta plaques in a mouse model of Alzheimer's disease.* Nature, 2008. **451**(7179): p. 720-4.
91. Huang, K.L., et al., *A common haplotype lowers PU.1 expression in myeloid cells and delays onset of Alzheimer's disease.* Nat Neurosci, 2017. **20**(8): p. 1052-1061.
92. Lim, S.L., et al., *Inhibition of hematopoietic cell kinase dysregulates microglial function and accelerates early stage Alzheimer's disease-like neuropathology.* Glia, 2018. **66**(12): p. 2700-2718.

Online Methods

Data source

Data has been downloaded from the AMP-AD consortium database hosted on the Synapse.org data portal (doi:10.7303/syn2580853).

Mayo Clinic Transcriptome and Genome-Wide Genotype Data:

The Mayo Clinic (MAYO) transcriptome and genome-wide genotype datasets utilized in this study have previously been described[1-4]. The Mayo Clinic temporal cortex RNA sequence data (Synapse ID: syn3163039) and genome-wide genotypes (Synapse ID: syn8650953) are available from AMP-AD Knowledge portal. Below, we provide details on these datasets:

Mayo Clinic Cohort Participants:

Mayo Clinic RNAseq cohort has RNAseq-based whole transcriptome data from 278 TCX samples from subjects with the following diagnoses: 84 AD, 84 PSP, 80 controls and 30 pathologic aging. For this study, we utilized data from subjects with AD and controls. Subjects with AD had definite neuropathologic diagnosis according to the NINCDS-ADRDA criteria[5] and had Braak neurofibrillary tangle (NFT) stage of ≥4.0.

Control subjects each had Braak[6] NFT stage of 3.0 or less, CERAD[7] neuritic and cortical plaque densities of 0 (none) or 1 (sparse) and lacked any of the following pathologic diagnoses: AD, Parkinson's disease (PD), DLB, VaD, PSP, motor neuron

disease (MND), CBD, Pick's disease (PiD), Huntington's disease (HD), FTLD, hippocampal sclerosis (HipScl) or dementia lacking distinctive histology (DLDH). Within the Mayo RNAseq cohort, all AD and PSP subjects were from the Mayo Clinic Brain Bank. Thirty-one control TCX samples were from the Mayo Clinic Brain Bank, and the remaining control tissue was from the Banner Sun Health Institute. All subjects were North American Caucasians. All but control subjects, had ages at death ≥60, and a more relaxed lower age cutoff of ≥50 was applied for normal controls to achieve sample sizes similar to that of AD and PSP subjects. Brain samples for the Mayo RNAseq study underwent RNA extractions via the Trizol/chloroform/ethanol method, followed by DNase and Cleanup of RNA using Qiagen RNeasy Mini Kit and Qiagen RNase -Free DNase Set. The quantity and quality of all RNA samples were determined by the Agilent 2100 Bioanalyzer using the Agilent RNA 6000 Nano Chip. Samples had to have an RIN ≥5.0 for inclusion in the study.

All of this work was approved by the Mayo Clinic Institutional Review Board. All human subjects or their next of kin provided informed consent.

Mayo Clinic RNAseq Data:

Mayo Clinic RNAseq samples were randomized across flowcells, taking into account age at death, sex, RIN, Braak stage and diagnosis. Library preparation and sequencing of the samples were conducted at the Mayo Clinic Medical Genome Facility Gene Expression and Sequencing Cores, as previously described[8]. The TruSeq RNA Sample Prep Kit (Illumina, San Diego, CA) was used for library

preparation from all samples. The library concentration and size distribution was determined on an Agilent Bioanalyzer DNA 1000 chip. Three samples were run per flowcell lane using barcoding. All samples underwent 101 base-pair (bp), paired-end sequencing on Illumina HiSeq2000 instruments. Base-calling was performed using Illumina's RTA 1.17.21.3. FASTQ sequence reads were aligned to the human reference genome using TopHat 2.0.12 [9] and Bowtie 1.1.0[10], and Subread 1.4.4 was used for gene counting[11]. FastQC was used for quality control (QC) of raw sequence reads, and RSeQC was used for QC of mapped reads. Raw read counts were log2-transformed and normalized using Conditional Quantile Normalization (CQN) via the Bioconductor package; accounting for sequencing depth, gene length, and GC content[12].

Mayo Clinic Genome-Wide Genotype Data:

Subjects in the Mayo Clinic RNAseq cohort underwent whole genome genotyping using the Illumina Infinium HumanOmni2.5-8 BeadChip, which delivers comprehensive coverage of both common and rare SNP content from the 1000 Genomes Project (minor allele frequency>2.5%) providing genotypes for 2,338,671 markers. The genotyping was done at the Mayo Clinic Medical Genome Facility. Whole genome genotype calls were made using the auto-calling algorithm in Illumina's BeadStudio 2.0 software, subsequent to which they were converted into PLINK formats for analysis[13].

Mayo Clinic RNAseq Data Quality Control (QC):

All Mayo Clinic TCX RNAseq samples had percent mapped reads ≥ 85%. Using R statistical software, raw read counts were transformed to counts per million (CPM), which were log2 normalized. Mean expression for chromosome Y genes with non-zero counts were plotted to identify any samples with deviation from expected expression based on recorded sex. Two AD TCX samples with discordant sex were excluded. Raw read counts were then normalized using Conditional Quantile Normalization (CQN) via the Bioconductor package; accounting for sequencing depth, gene length, and GC content. GC content was calculated via the Bioconductor package, Repitools and sequencing depth was calculated as the sum of reads mapped to genes. Genes with non-zero counts across all samples were retained and principal components analysis was performed using the prcomp function implemented using R Statistical Software (R Foundation for Statistical Computing, version 3.2.3). Principal components 1 and 2 were plotted and no outliers (>6SD from mean) were identified.

Mayo Clinic Genotype Data Quality Control (QC):

Genome-wide genotypes were obtained for all subjects in the Mayo Clinic RNAseq study using Illumina Omni 2.5 Beadchips. Samples were checked for discordant sex. The same two subjects that were excluded due to discordant sex based on RNAseq data were also determined to have discordant sex based on the genome-wide genotype data. Subjects were assessed for heterozygosity rate > 3SD from the mean. One AD sample with TCX RNAseq had high heterozygosity indicating possible

sample contamination and 3 TCX RNAseq samples had low heterozygosity (2 controls and 1 AD) indicating either divergent ancestry or consanguinity. These samples were also excluded from the analysis. The dataset was filtered to include only autosomal SNPs. PLINK[13] --genome function was used to identify any sample duplicates or related pairs of subjects. Two pairs of samples were identified as > 3rd degree relatives. For each pair, the sample with the lowest SNP call rate was excluded (1 PSP and 1 control). The dataset was further filtered to remove complex genomic regions (chr8:1-12,700,000; chr2:129,900,001-136,800,000; chr17:40,900,001-44,900,000; chr6:32,100,001-33,500,000) and LD pruned using the SNPRelate (v1.4.2) package in R (v3.2.3) [14], implementing an LD threshold of 0.15 and a sliding window of 1E-07 bp. Remaining SNPs and subjects were analyzed using EIGENSOFT[15] for population outliers. Two samples were identified as population outliers (1 PSP and 1 control) using the default parameters of > 6 SD from the mean on any of the top ten inferred axes following 5 iterations and were removed from further analysis.

Clinical, genotype and processed RNA-seq read count data was obtained privately from Dr. Nilufer Tan Taner lab at the Mayo Clinic.

ROSMAP Transcriptome and Genome-Wide Genotype Data:

The ROSMAP dataset's dorsolateral prefrontal cortex gene expression (RNA-seq BAM files), genotypes and clinical covariates were downloaded from synapse (respective synapse project IDs: syn4164376, syn3157325 and syn3191087) using the synapseClient

R library (Matthew Furia (2015). synapseClient: Synapse R Client from Sage Bionetworks. R package version 1.11-1. http://www.sagebase.org). Requests for ROSMAP data can be made at www.radc.rush.edu.

ROSMAP Cohort Participants:

ROSMAP dataset contains two cohorts: The Religious Orders Study (ROS) and The Memory and Aging Project (MAP)[16]. Both ROS and MAP are a longitudinal clinical-pathologic cohort studies of aging and dementia run by the Rush Alzheimer's Disease Center. In both studies, participants enroll without known dementia and agree to annual clinical evaluation. All subjects agree to brain donation as a condition of entry. ROS enrolled individuals from religious orders (nuns, priests, brothers) from across the United States starting in 1994. MAP enrolled lay persons from across northeastern Illinois. Each study administers a battery of 21 cognitive performance tests annually of which 19 are in common. Alzheimer's Disease status was determined by a computer algorithm based on cognitive test performance with a series of discrete clinical judgments made in series by a neuropsychologist and a clinician. Persons were categorized as no cognitive impairment (NCI) if diagnosed without dementia or mild cognitive impairment (MCI). Diagnoses of dementia and AD conform to standard definitions. A clinician reviewed all cases determined by this algorithm to render a diagnosis blinded to data collected in prior years. In addition to dementia, five other diagnoses were determined by this approach including stroke, cognitive impairment due to stroke, parkinsonism, Parkinson's disease, and depression. Most other diagnoses are by self report. Upon death, a

summary diagnosis is made by a neurologist blinded to the post-mortem assessment. A post-mortem neuropathologic evaluation is performed that includes a uniform structured assessment of AD pathology, cerebral infarcts, Lewy body disease, and other pathologies common in aging and dementia. The procedures follow those outlined by the pathologic dataset recommended by the National Alzheimer's Disease Coordinating Center and pathologic diagnoses of AD use NIA-Reagan and modified CERAD criteria, and the staging of neurofibrillary pathology uses Braak Staging. Both studies are conducted by the same clinical and pathologic data collection teams with extensive item-level harmonization allowing the data to be efficiently merged.

ROSMAP Genotype data Processing

Plink2 was used to perform operations on the genotype file (see link https://www.cog-genomics.org/plink/1.9/general_usage#cite), and positions were liftovered from hg18 to hg19 (http://genome.ucsc.edu/cgi-bin/hgLiftOver). Picard was used to sort the resulting genotype file, and variants with more than 2% missing values, minor allele frequency less than 1%, Hardy-Weinberg equilibrium less than 10E-6 as well as inbred samples (inbreeding coefficient >0.15) and samples with more than 2% missing values were removed using Plink2. Starting with 750,173 variants in 1,708 individuals for ROSMAP, 736,073 variants in 1,091 individuals for ROSMAP were left after quality control.

ROSMAP RNAseq data processing

The RNA-seq BAM files were sorted using samtools[17] and converted to fastq files using the SamToFastq function (Picard 1.138, http://broadinstitute.github.io/picard/). RAPiD[18] was used to generate a count matrix for the gene expression data and generate a vcf file for each sample aligned to hg19 from the fastq files.

Imputation

After quality control, we used 1000 Genomes data[19] and IMPUTEv2[20] to impute untyped variants. Imputed variants were removed if they failed any of the previously listed quality control criteria or had information scores < 0.6. After imputation we had 7,132,687 variants in Mayo and 9,333,139 variants in ROSMAP.

De-convolute RNA-seq data into Microglia-specific Expression Residual

ROSMAP RNA-seq read count expression data was normalized using log2 counts per million (CPM) and the TMM method[21] implemented in edgeR[22]. Genes with over 1 CPM in at least 30% of the experiments were retained. We then used precision weights as implemented in the voom function from the limma[23] R package to further normalize the gene counts. MAYO expression data was normalized using the CQN R package (see above). For ROSMAP, expression residuals were obtained by correcting for the effects of technical (study, sequencing batch), sample-specific (post-mortem interval (PMI), RNA integrity number (RIN), exonic

mapping rate) and patient-specific covariates (sex, educational attainment, age at death). For Mayo we adjusted for a slightly different set of covariates due to availability of recorded measurements (source of sample, sequencing batch, RIN, exonic mapping rate, sex, age at death). For both ROSMAP and MAYO data, we computed the exonic mapping rate using RNAseQC[24]. The exonic mapping rate was also included in the covariates. Adjustment for covariates was done using the limma R package.

Further, and performed together with the above listed covariate adjustments, for both ROSMAP and Mayo, we also adjusted for 5 cell type markers[25]: ENO2 [neuron], CD68 [microglial], CD34 [endothelial], OLIG2 [oligodendrocyte], GFAP [astrocyte]. To obtain expression residuals that mimic expression patterns seen in microglial cells, we added, for every gene, the CD68 effects estimated by the linear regression models back to the expression residuals.

The final microglia-specific expression residual data available for analysis included 20,276 genes for 612 individuals (ROSMAP) and 19,885 genes for 266 individuals (Mayo), with 18,408 genes in common to the two datasets.

Rationalize and Validate Single-gene Biomarker for Bulk-tissue RNA-seq De-convolution

Compare sc-RNAseq derived Biomarker Genes per Cell Type

We assembled cell-type specific biomarkers derived from existing single-cell RNAseq data for neurons, microglial, astrocyte, endothelial and oligodendrocyte respectively. These biomarkers are included in the lists (Supplementary File S6). In

each cell type, we compared the biomarker genes from every pair of studies and calculated the significance of the overlap by Exact Fisher's test. The FDR is used to correct for multiple testing.

PCA analysis of sc-RNAseq derived Biomarker Expression in AMP-AD Data

We first merged biomarker list from different scRNA-seq studies for each cell type. Then, extracted the gene expression matrix of merged biomarkers from the ROSMAP and MAYO RNA-seq data. Next, we applied principal component analysis (PCA) on the extracted RNA-seq sub-matrix.

Compare sc-RNAseq derived Biomarker Genes with AMP-AD AGORA Targets

We merged all biomarkers for each cell type and calculated the percentage of overlapping with AGORA Targets. To evaluate the significance of this overlap, we simulated a background distribution of overlap by randomly selected the same number of genes from background genes (taking the non-duplicate union of genes in MAYO and ROSMAP RNA-seq data) per cell type, and compare the randomly generated "pseudo" biomarker list to AMP-AD AGORA Targets to generate a overlapping percentage. We repeated the random simulation 10,000 times to construct the background distribution. The p-value is then calculated by comparing true percentage to the background distribution per cell type.

Evaluate Robustness of Single-gene Biomarker derived Microglial-specific Residuals to scRNAseq Biomarker derived Microglial-specific Residuals.

By using PSEA, we estimated the variance component of the bulk-tissue RNAseq data in ROSMAP and MAYO dataset explained by our single-gene microglial biomarker (CD68). Next, we randomly select a subset of biomarkers of each cell type from the scRNAseq-

derived biomarkers (Supplementary File S6), then applied PSEA again to estimate the variance component of the bulk-tissue RNAseq data explained by the simulated subset of biomarkers. Then, we calculated the Pearson correlation of each gene between our single-gene microglial residuals with the simulated microglial residuals. We repeated this procedure 1,000 times to construct a distribution of the correlations.

Next, we intend to construct a background distribution of correlation. To this end, we again randomly select a subset of "pseudo" biomarkers of each cell type from the background genes (see above), then applied PSEA to estimate the variance component of the bulk-tissue RNAseq data explained by the simulated "pseudo" biomarkers. Then, we calculated the Pearson correlation of each gene between our single-gene microglial residuals with the simulated "pseudo" microglial residuals. We repeated this procedure 1,000 times to construct a distribution of the correlations. Lastly, we applied t-test to calculate the p-value based on the two distributions.

Computational Analysis of Microglial-specific Gene Expression Data

eQTL analysis

Expression quantitative trait loci (eQTL) analysis was performed using the R package MatrixEQTL v2.1.1[26] using QCed genotypes and normalized and covariate-adjusted celltype-specific expression residuals. cis-eQTL analysis considered markers within 1Mb of the transcription state site of each gene. False discovery rates were computed using the Benjamini–Hochberg procedure[27].

Differential Expression (DE) Analysis

We interrogated the celltype-specific residual expression data for genes differentially expressed between AD cases and healthy controls using linear models, as implemented in the R package limma[23]. Significance was assessed using Benjamini-Hochberg corrected p-values < 5%.

Co-expression networks Analysis

Co-expression networks were constructed using the coexpp R package[28] (Michael Linderman and Bin Zhang (2011). coexpp: Large-scale Co-expression network creation and manipulation using WGCNA. R package version 0.1.0. https://bitbucket.org/multiscale/coexpp). A soft thresholding parameter value of 6.5 is used to power the expression correlations. Seeding gene lists for the predictive networks were obtained by selecting genes in co-expression modules that were statistically enriched (FDR adjusted p-value < 0.05) for DE genes or astrocyte or microglial cell markers (lists of the latter two were obtained from [29]).

Key Driver Analysis

To do Key Driver Analysis, we used the R package KDA[30] (KDA R package version 0.1, available at http://research.mssm.edu/multiscalenetwork/Resources.html). The package first defines a background sub-network by looking for a neighborhood K-step away from each node in the target gene list in the network. Then, stemming from each node in this sub-network, it assesses the enrichment in its k-step (k varies from 1 to K) downstream neighborhood for the target gene list. In this analysis, we used K = 6.

Predictive networks Modeling and In-silico Prediction Validation

Though the co-expression network modules capture highly co-regulated genes operating in coherent biological pathways, these modules do not reflect the probabilistic causal information needed to identify key driver genes. Conventional Bayesian networks (BN) have been widely used to infer causal structures among genes given gene expression data, however, BN has significant limitations when it comes to infer opposite causality given the symmetry of joint probability. Recent work[31] has demonstrated that the bottom-up causality inference can accurately distinguish true causality from opposite causality in equivalent class. In this study, we developed a novel computational network model, called Predictive Network, by integrating conventional (top-down) Bayesian network with the bottom-up causality inference to address the problem of opposite causality inference in BN. Our causal predictive network pipeline incorporated multi-scale omics data, such as genotypes and transcriptomic profiles, in ROSMAP and MAYO dataset (de-convoluted microglial-specific residuals) to build causal predictive networks separately in ROSMAP and MAYO.

The predictive network model captures causal regulations among genes, which allows us to generate (in-silico) predictions upon perturbations, e.g. shRNA. Previously[32], we developed an integrative method, called Qualitative-constrained Maximal-a-Posterior (QMAP), to estimate the parameters of probabilistic graphical models. This method has been demonstrated to outperform traditional Maximal-a-Posterior (MAP) estimation

without prior information. In this paper, we extended QMAP to integrate infinite number of resources of (priori) information on genetic regulations and big training data, to estimate parameters for constructed MAYO- and ROSMAP-microglial networks. Firstly, we collected multiple resources of prior information: i) we checked each edge in the MAYO- and ROSMAP-microglial network model against pathway knowledgebase, such as CPDB and String; ii) we applied linear regression to the (continuous) residual data to estimate the interaction type of each edge; Secondly, we integrated the two resources of prior with data to derive the parameters.

To predict gene expression fold-change upon shRNA against each key driver (HCK, FCER1G, LAPTM5), we developed three-step generalization procedure. First, we extracted the total sub-network of three key drivers from the constructed MAYO- and ROSMAP-microglial predictive networks (MAYO-sub, ROSMAP-sub) and included the Markov blanket of all top nodes in these sub-networks from the original MAYO-/ROSMAP-microglial networks. Second, we simulated the predictive network under wild-type (unperturbed) AD condition where probability of every top node is initialized according to the MAYO-/ROSMAP-microglial residual data under AD condition. Thirdly, we simulated the predictive network by perturbing key drivers under AD condition where probability of each key driver is initialized according to its knockdown level measured by Tagman (Figure S9) and other top nodes are initialized according to wild-type AD condition.

To calculate simulated gene expression fold-change, we used previously developed method [33, 34] to calculate the ratio of marginal probability of 18 measured target genes and compared to the experimental gene expression fold-change by Pearson correlation.

Induction of Monocyte-Derived Microglia-like Cells (MDMi)

Peripheral blood mononuclear cells (PBMCs) were separated by Lymphoprep gradient centrifugation (StemCell Technologies). PBMCs were frozen at a concentration of 1–3 × 10^7 cells ml^{-1} in 10% DMSO (Sigma-Aldrich)/90% fetal bovine serum (vol/vol, Corning). Prior to each study, aliquots of frozen PBMCs from the PhenoGenetic cohort were thawed and washed in 10 ml PBS. Monocytes were positively selected from whole PBMCs using anti-CD14+ microbeads (Miltenyi Biotech) and plated at the following densities per well: 1 x 10^5 cells (96-well plate). To induce the differentiation of MDMi, monocytes were incubated in serum-free conditions using RPMI-1640 Glutamax (Life Technologies) with 1% penicillin/streptomycin (Lonza) and 2.5 µg/ml Fungizone (Life Technologies) and a mixture of the following human recombinant cytokines: M-CSF (10 ng/ml; Biolegend 574806), GM-CSF (10 ng/ml; R&D Systems 215-GM-010/CF), NGF-β (10 ng/ml; R&D Systems 256-GF-100), CCL2 (100 ng/ml; Biolegend 571404) at standard humidified culture conditions (37°C, 5% CO_2) for up to 10 days[35].

Lentivirus mediated shRNA triggered knockdown in primary monocyte derived microglia like cells (MDMi)

shRNA lentiviral particle preparation: Vpx viral particles were made using 293 T cells. On day 1, 293T cells were transfected using Lipofectamine 2000 (Thermo Fisher Scientific) along with envelope and packaging plasmids (Siv3+, pHEF VsVg a concentration of 1 µg/ml). On day 2, the culture medium was replaced by RPMI (Invitrogen) with 1% pennstrep and 1% fungizone (Amphotericin B) medium; the lentiviruses containing the Vpx particles were harvested 48 hours later and centrifuged at 400 *g* for 5 minutes at 4°C. The final product was filter sterilized using a 0.45-µm syringe filter (EMD Millipore). shRNA for target gene containing Lentiviral particles for each gene were obtained from the Broad Institute GPP (HCK, TRCN00000379914 and TRCN00000379408), FcER1G (TRCN0000057455 and TRCN0000057457) and LAPTM5 (TRCN0000428031 and TRCN0000429201).

Lentivirus mediated knockdown of MDMi: The monocytes were isolated from 9 healthy subjects and differentiated to MDMi using the above-mentioned protocol and plated on 96 well - temperature sensitive plate (Life Technologies #) as well as regular 384 well plates. On day 4, in the process of differentiation, Media was changed and lentivirus containing the Vpx particles and the lentivirus containing the shRNA for target gene from Broad were added to the cells, MDMi were maintained in the RPMI media with MDMi cocktail. On day 7, the transduced cells were selected using puromycin (Thermofisher scientific) at a concentration of (3ug/ml) conc. On day 10, the cells were lysed and gene expression assays (qPCR) was performed to validate expression of HCK, FcER1G and LAPTM5 as well as key downstream genes for each.

To assess statistical significance of differences in gene expression of knocked-down genes or genes downstream of these, we made use of linear mixed models, accounting for the multiple technical replicates for each biological replicate via random intercepts. These models also allowed us to deal with the fact that the experimental design was unbalanced with only two technical replicates for the empty control and three technical replicates for each biological replicates using shRNA triggered constructs.

Quantitative Real Time-Polymerase Chain Reaction (qRT-PCR)

RNA was extracted from each sample using RNeasy micro kit (Qiagen, USA). Genomic DNA contamination was minimized by spinning samples using a genomic DNA column (gDNA) according to the manufacturer's instructions. RNA was reverse transcribed into cDNA using a Taqman Reverse Transcription kit (Invitrogen). qPCR was performed using TaqMan® Fast Advanced Master Mix (Applied Biosystems) and run on a Light cycler 480 System (Roche, USA). The cycling conditions consisted of 90 °C for 10 min and 40 cycles of 95 °C for 20 s followed by 60 °C for 30 sec. Samples were assayed with 2 technical replicates. mRNA levels were normalized relative to B2M by the formula $2^{-\Delta Ct}$, where $\Delta Ct = Ct_{mRNA\text{-}X} - Ct_{B2M}$.

Aβ1-42 Uptake Assay

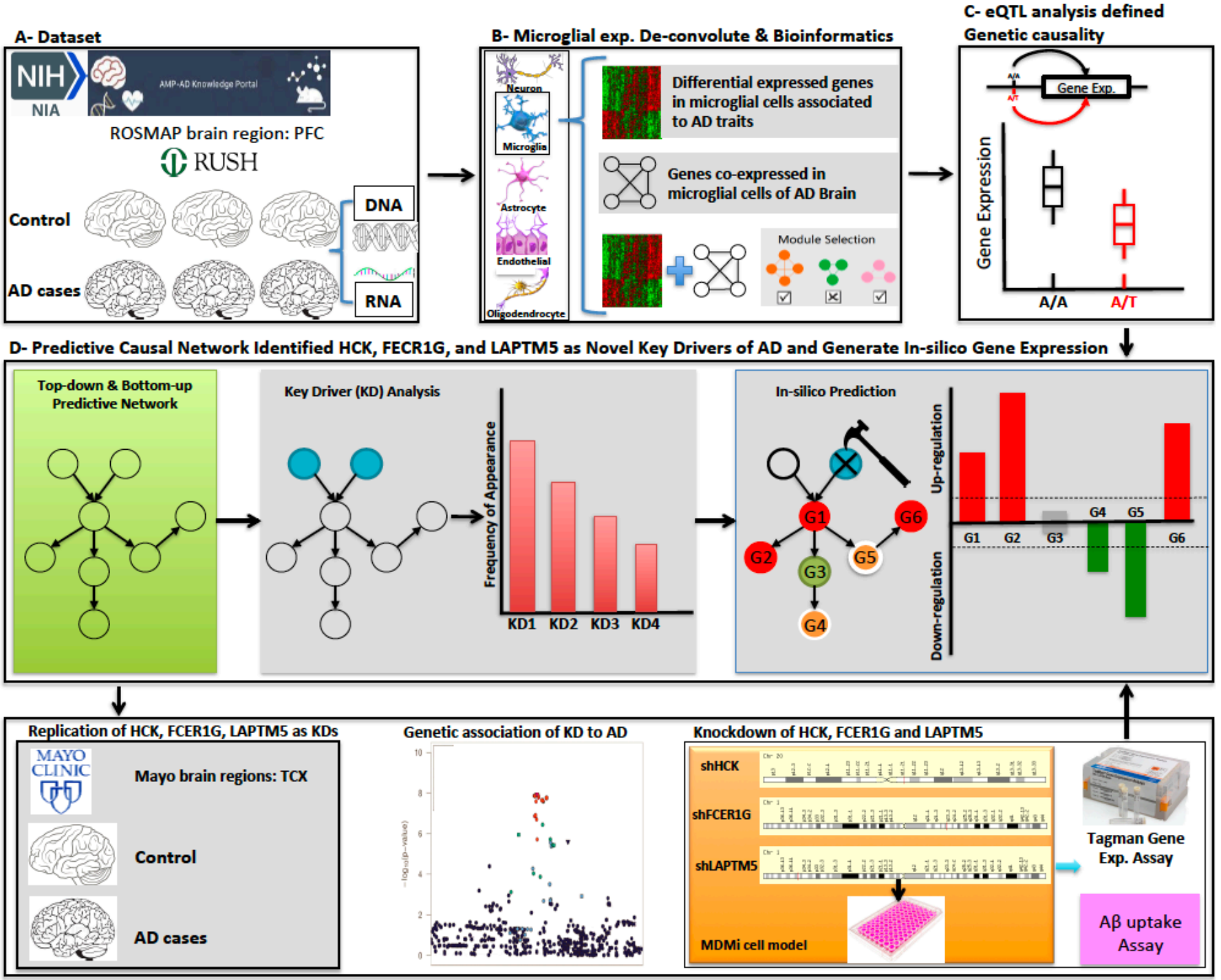
We tested the uptake ability of lentivirus mediated shRNA triggered downregulated HCK and FcER1G MDMi using HiLyte™ Fluor 488-labeled beta amyloid 1-42(Anaspec AS-60479-01) for a period of 10 days in RPMI media with MDMi cocktail (Katie's paper). On day 10, the media was replaced with media containing 1.5ug/ml of HiLyte™ Fluor 488-

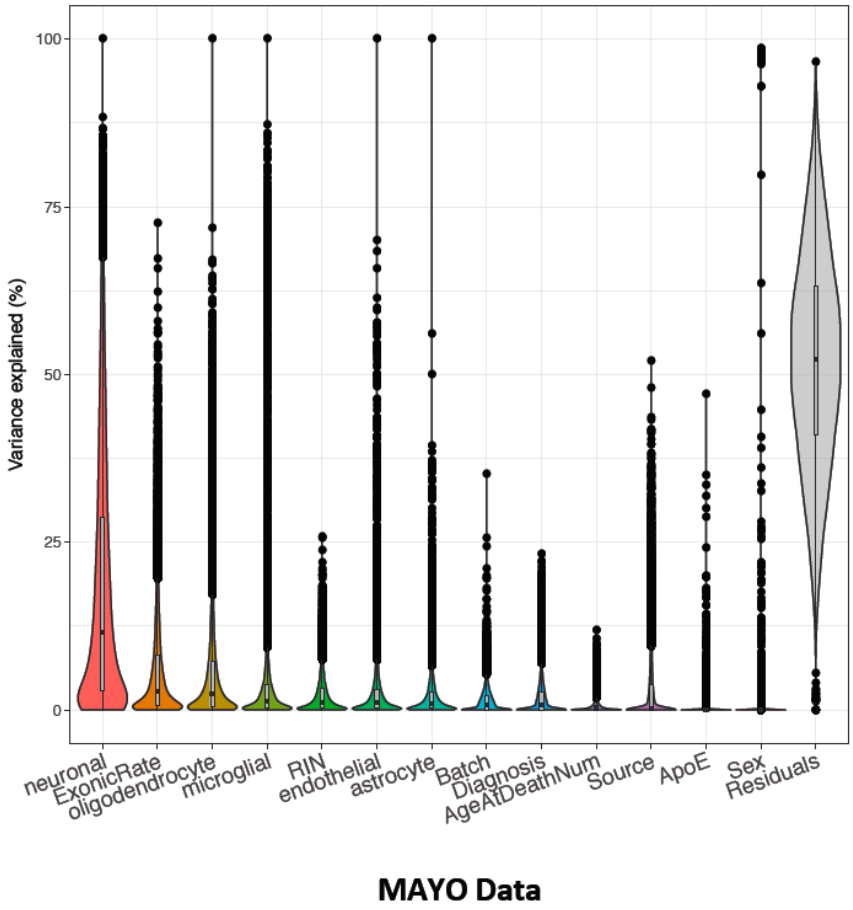
labeled beta amyloid 1-42 for 2 h at 37 °C. After 2 hours, cells were washed three times with PBS and fixed in 4% PFA for 15 min. The cells were then imaged using confocal image express C (Harvard, Longwood ICCB). The mean fluorescence intensity was measured using the Multi-wavelength scoring program. Data shown is the mean fluorescence intensity for each subject. Two-tailed, paired t-tests were used to determine statistical significance.

1. Allen, M., et al., *Human whole genome genotype and transcriptome data for Alzheimer's and other neurodegenerative diseases.* Scientific data, 2016. **3**: p. 160089.
2. Allen, M., et al., *Conserved brain myelination networks are altered in Alzheimer's and other neurodegenerative diseases.* Alzheimer's & dementia : the journal of the Alzheimer's Association, 2018. **14**(3): p. 352-366.
3. Allen, M., et al., *Divergent brain gene expression patterns associate with distinct cell-specific tau neuropathology traits in progressive supranuclear palsy.* Acta neuropathologica, 2018. **136**(5): p. 709-727.
4. Conway, O.J., et al., *ABI3 and PLCG2 missense variants as risk factors for neurodegenerative diseases in Caucasians and African Americans.* Molecular neurodegeneration, 2018. **13**(1): p. 53.
5. McKhann G, et al., *Clinical diagnosis of Alzheimer's disease: report of the NINCDS-ADRDA Work Group under the auspices of Department of Health and Human Services Task Force on Alzheimer's Disease.* Neurology, 1984. **34**(7): p. 939-44.
6. Braak, H. and E. Braak, *Neuropathological stageing of Alzheimer-related changes.* Acta neuropathologica, 1991. **82**(4): p. 239-59.
7. Mirra, S.S., et al., *Interlaboratory comparison of neuropathology assessments in Alzheimer's disease: a study of the Consortium to Establish a Registry for Alzheimer's Disease (CERAD).* Journal of neuropathology and experimental neurology, 1994. **53**(3): p. 303-15.
8. Allen, M., et al., *Gene expression, methylation and neuropathology correlations at progressive supranuclear palsy risk loci.* Acta neuropathologica, 2016. **132**(2): p. 197-211.
9. Trapnell, C., L. Pachter, and S.L. Salzberg, *TopHat: discovering splice junctions with RNA-Seq.* Bioinformatics, 2009. **25**(9): p. 1105-11.
10. Langmead, B., et al., *Ultrafast and memory-efficient alignment of short DNA sequences to the human genome.* Genome biology, 2009. **10**(3): p. R25.
11. Anders, S., P.T. Pyl, and W. Huber, *HTSeq--a Python framework to work with high-throughput sequencing data.* Bioinformatics, 2015. **31**(2): p. 166-9.

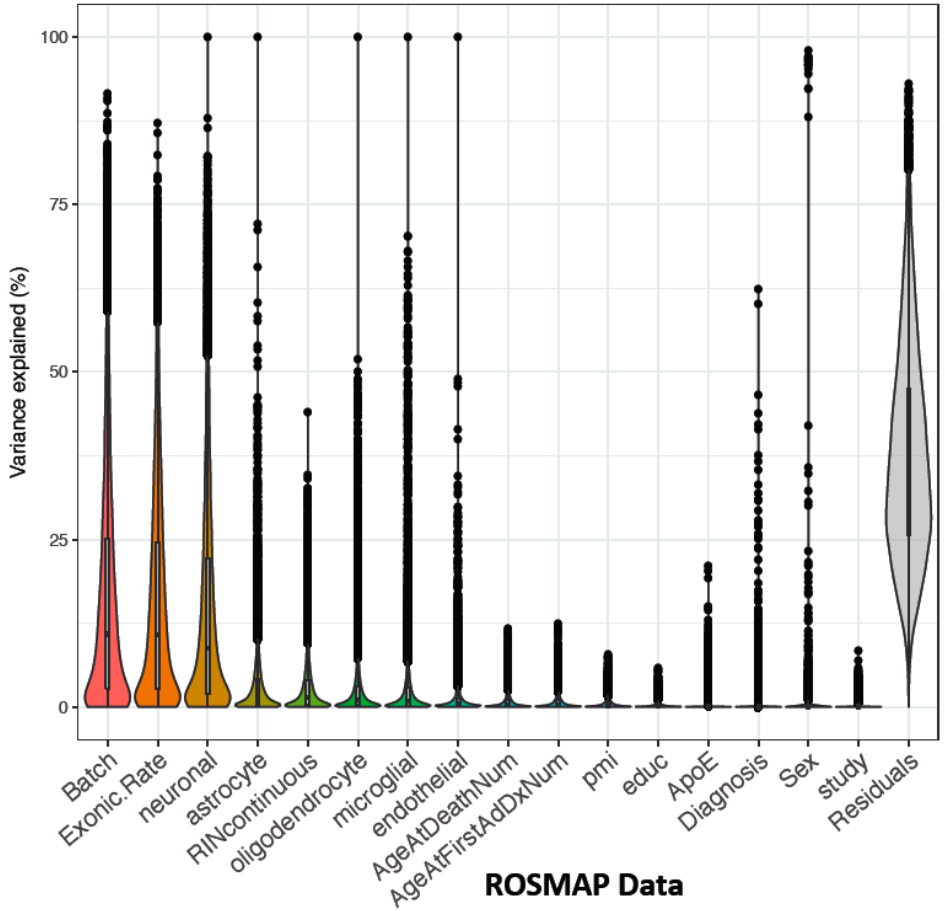
12. Hansen, K.D., R.A. Irizarry, and Z. Wu, *Removing technical variability in RNA-seq data using conditional quantile normalization.* Biostatistics, 2012. **13**(2): p. 204-16.
13. Purcell, S., et al., *PLINK: a tool set for whole-genome association and population-based linkage analyses.* Am J Hum Genet, 2007. **81**(3): p. 559-75.
14. Zheng, X., et al., *A high-performance computing toolset for relatedness and principal component analysis of SNP data.* Bioinformatics, 2012. **28**(24): p. 3326-8.
15. Patterson, N., A.L. Price, and D. Reich, *Population structure and eigenanalysis.* PLoS genetics, 2006. **2**(12): p. e190.
16. Bennett, D.A., et al., *Religious Orders Study and Rush Memory and Aging Project.* J Alzheimers Dis, 2018. **64**(s1): p. S161-S189.
17. Li, H., et al., *The Sequence Alignment/Map format and SAMtools.* Bioinformatics, 2009. **25**(16): p. 2078-9.
18. Hardik Shah, Y.-C.W., Rafael Castellanos, Chetanya Pandya, Zachary Giles, *RAPiD: An Agile and Dependable RNA-Seq Framework.* The 65th Annual Meeting of The American Society of Human Genetics, 2015.
19. Genomes Project, C., et al., *A global reference for human genetic variation.* Nature, 2015. **526**(7571): p. 68-74.
20. Howie, B.N., P. Donnelly, and J. Marchini, *A flexible and accurate genotype imputation method for the next generation of genome-wide association studies.* PLoS Genet, 2009. **5**(6): p. e1000529.
21. Robinson, M.D. and A. Oshlack, *A scaling normalization method for differential expression analysis of RNA-seq data.* Genome Biol, 2010. **11**(3): p. R25.
22. Robinson, M.D., D.J. McCarthy, and G.K. Smyth, *edgeR: a Bioconductor package for differential expression analysis of digital gene expression data.* Bioinformatics, 2010. **26**(1): p. 139-40.
23. Ritchie, M.E., et al., *limma powers differential expression analyses for RNA-sequencing and microarray studies.* Nucleic Acids Res, 2015. **43**(7): p. e47.
24. DeLuca, D.S., et al., *RNA-SeQC: RNA-seq metrics for quality control and process optimization.* Bioinformatics, 2012. **28**(11): p. 1530-2.
25. Allen, M., et al., *Conserved brain myelination networks are altered in Alzheimer's and other neurodegenerative diseases.* Alzheimers Dement, 2018. **14**(3): p. 352-366.
26. Shabalin, A.A., *Matrix eQTL: ultra fast eQTL analysis via large matrix operations.* Bioinformatics, 2012. **28**(10): p. 1353-8.
27. Benjamini, Y., et al., *Controlling the false discovery rate in behavior genetics research.* Behav Brain Res, 2001. **125**(1-2): p. 279-84.
28. Langfelder, P. and S. Horvath, *WGCNA: an R package for weighted correlation network analysis.* BMC Bioinformatics, 2008. **9**: p. 559.
29. Zeisel, A., et al., *Brain structure. Cell types in the mouse cortex and hippocampus revealed by single-cell RNA-seq.* Science, 2015. **347**(6226): p. 1138-42.
30. Zhang Bin, Z.J., *Identification of Key Causal Regulators in Gene Networks.* Lecture Notes in Engineering and Computer Science, 2013. **2**: p. 1309-1312.

31. Chang, R., J.R. Karr, and E.E. Schadt, *Causal inference in biology networks with integrated belief propagation.* Pac Symp Biocomput, 2015: p. 359-70.
32. Rui Chang, W.W. *Novel algorithm for Bayesian network parameter learning with informative prior constraints*. in *Neural Networks (IJCNN), The 2010 International Joint Conference on*. 2010.
33. Chang, R., R. Shoemaker, and W. Wang, *A Novel Knowledge-Driven Systems Biology Approach for Phenotype Prediction upon Genetic Intervention.* IEEE/ACM Transactions on Computational Biology and Bioinformatics, 2011. **8**(5): p. 1170-1182.
34. Chang, R., R. Shoemaker, and W. Wang, *Systematic search for recipes to generate induced pluripotent stem cells.* PLoS Comput Biol, 2011. **7**(12): p. e1002300.
35. Ryan, K.J., et al., *A human microglia-like cellular model for assessing the effects of neurodegenerative disease gene variants.* Sci Transl Med, 2017. **9**(421).


A- Dataset
NIH
NIA
AMP-AD Knowledge Portal
ROSMAP brain region: PFC
RUSH
Control
AD cases
DNA
RNA
B- Microglial exp. De-convolute & Bioinformatics
Neuron
Microglia
Astrocyte
Endothelial
Oligodendrocyte
Differential expressed genes in microglial cells associated to AD traits
Genes co-expressed in microglial cells of AD Brain
Module Selection
C- eQTL analysis defined Genetic causality
Gene Exp.
A/A
A/T
Gene Expression
D- Predictive Causal Network Identified HCK, FECR1G, and LAPTM5 as Novel Key Drivers of AD and Generate In-silico Gene Expression
Top-down & Bottom-up Predictive Network
Key Driver (KD) Analysis
Frequency of Appearance
KD1
KD2
KD3
KD4
In-silico Prediction
G1
G2
G3
G4
G5
G6
Up-regulation
Down-regulation
Replication of HCK, FCER1G, LAPTM5 as KDs
MAYO CLINIC
Mayo brain regions: TCX
Control
AD cases
Genetic association of KD to AD
−log10(p-value)
Knockdown of HCK, FCER1G and LAPTM5
shHCK
shFCER1G
shLAPTM5
MDMi cell model
Tagman Gene Exp. Assay
Aβ uptake Assay

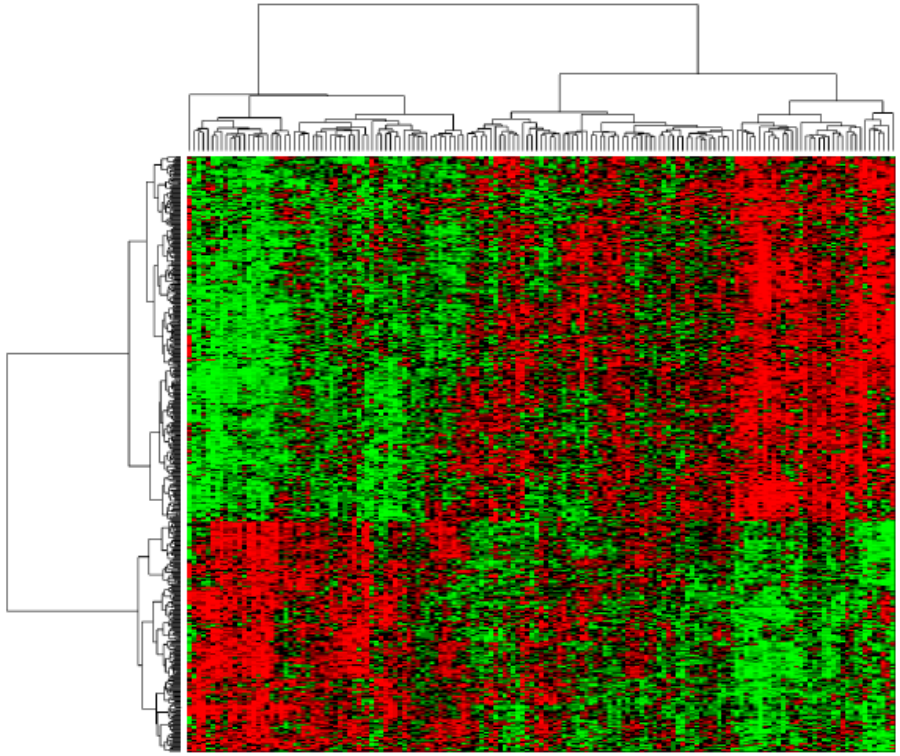


MAYO Data

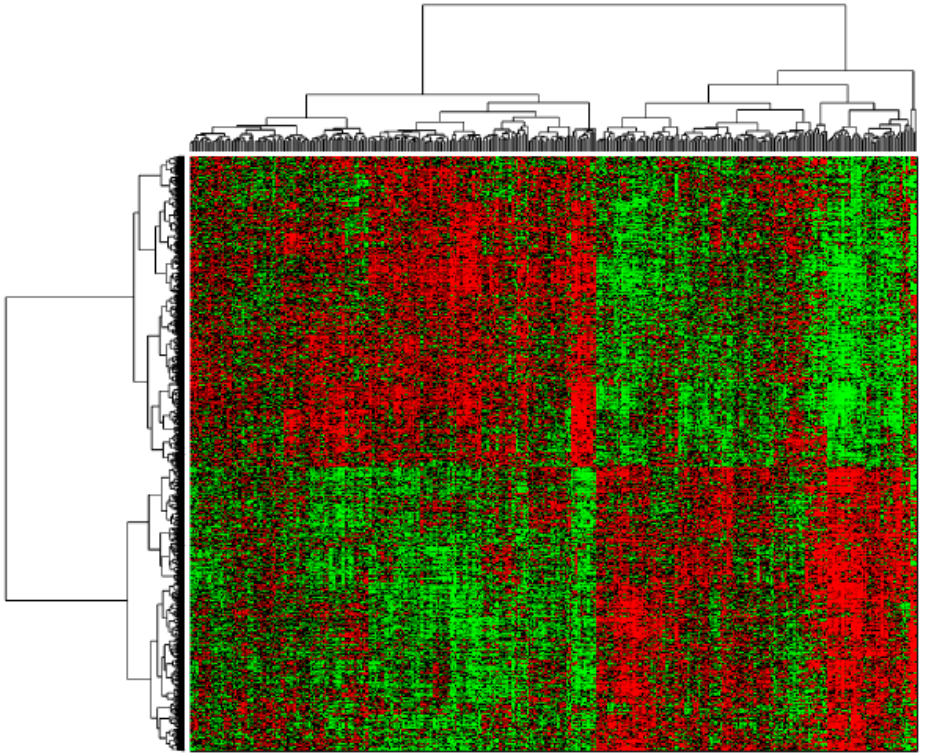


ROSMAP Data

Figure 2A



MAYO-microglial



ROSMAP-microglial

Figure 2B

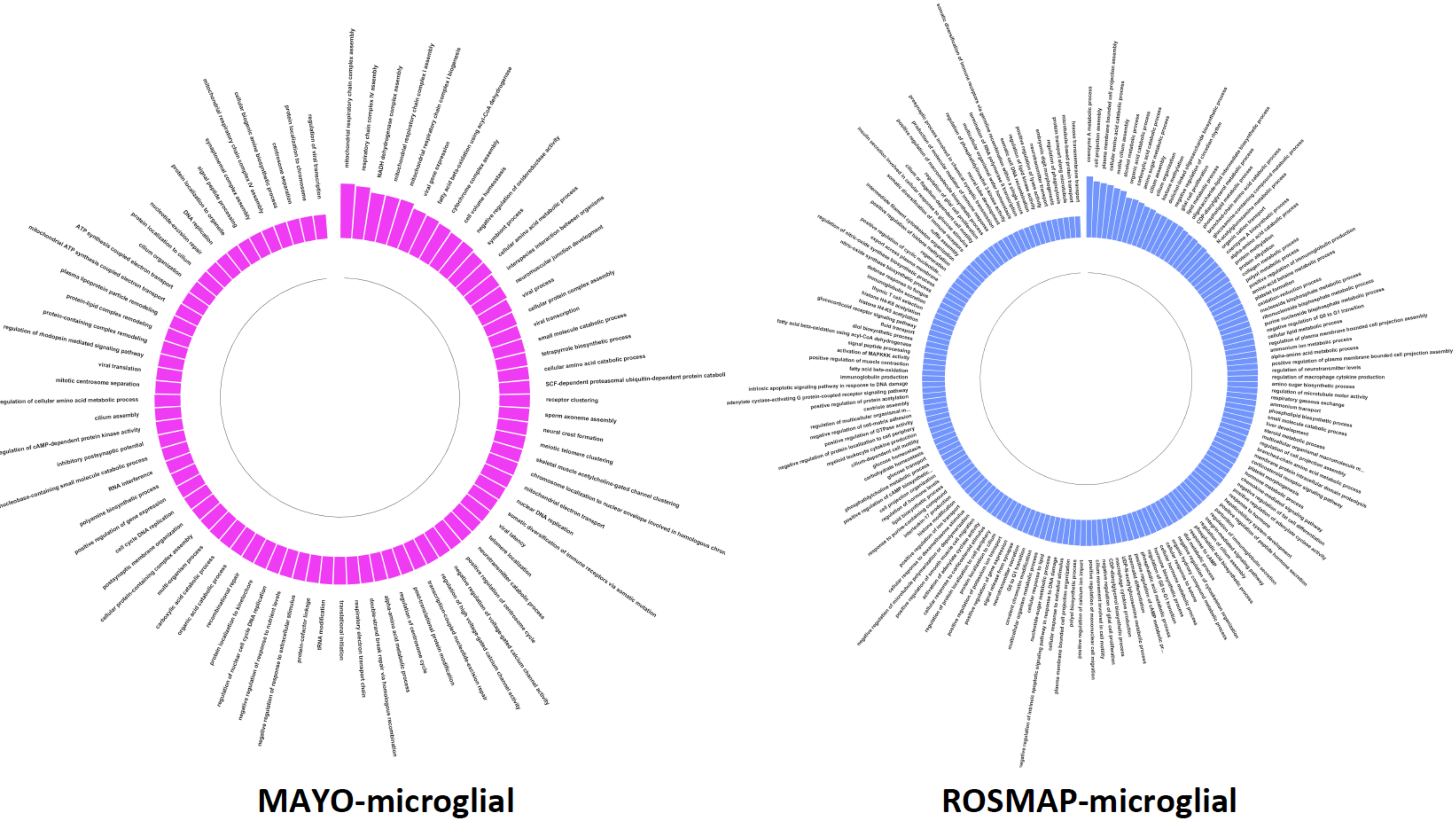


Figure 2C

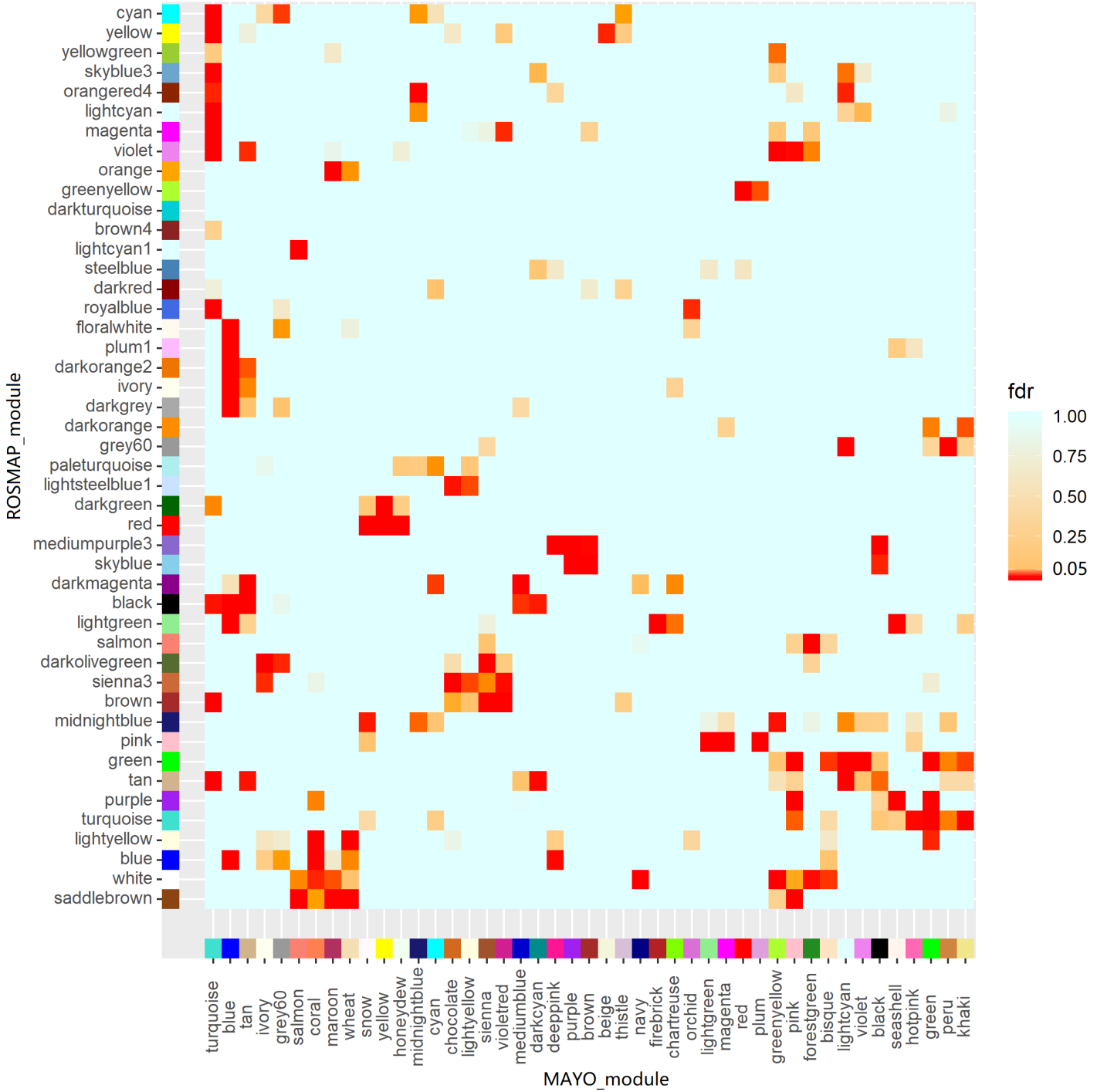
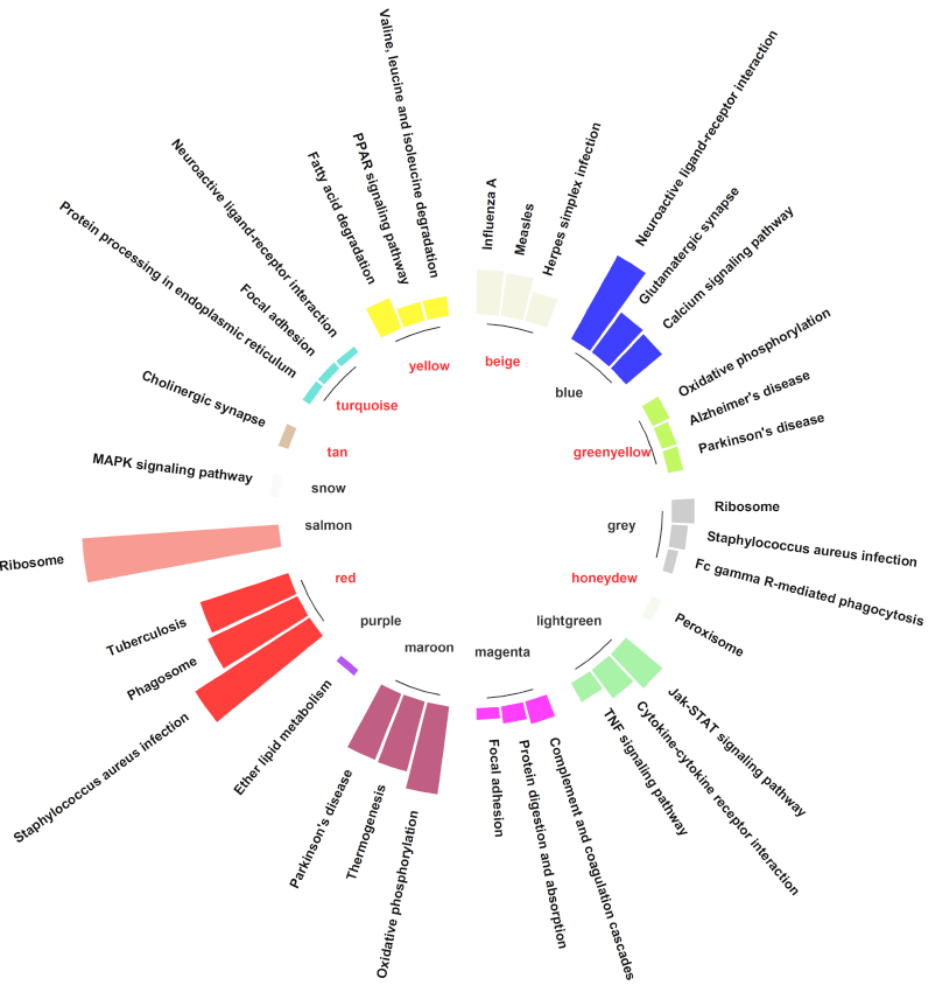
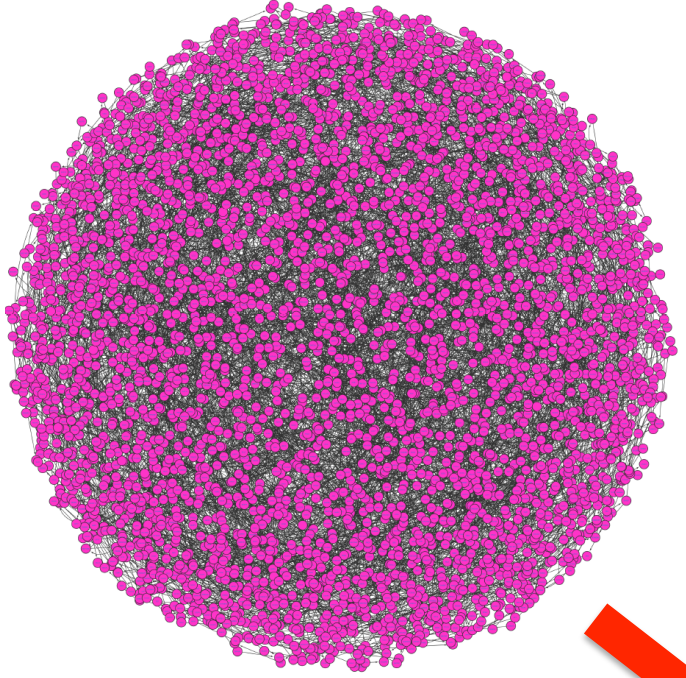

ROSMAP_module
cyan
yellow
yellowgreen
skyblue3
orangered4
lightcyan
magenta
violet
orange
greenyellow
darkturquoise
brown4
lightcyan1
steelblue
darkred
royalblue
floralwhite
plum1
darkorange2
ivory
darkgrey
darkorange
grey60
paleturquoise
lightsteelblue1
darkgreen
red
mediumpurple3
skyblue
darkmagenta
black
lightgreen
salmon
darkolivegreen
sienna3
brown
midnightblue
pink
green
tan
purple
turquoise
lightyellow
blue
white
saddlebrown
MAYO_module
turquoise
blue
tan
ivory
grey60
salmon
coral
maroon
wheat
snow
yellow
honeydew
midnightblue
cyan
chocolate
lightyellow
sienna
violetred
mediumblue
darkcyan
deeppink
purple
brown
beige
thistle
navy
firebrick
chartreuse
orchid
lightgreen
magenta
red
plum
greenyellow
pink
forestgreen
bisque
lightcyan
violet
black
seashell
hotpink
green
peru
khaki
fdr
1.00
0.75
0.50
0.25
0.05

Figure 3A

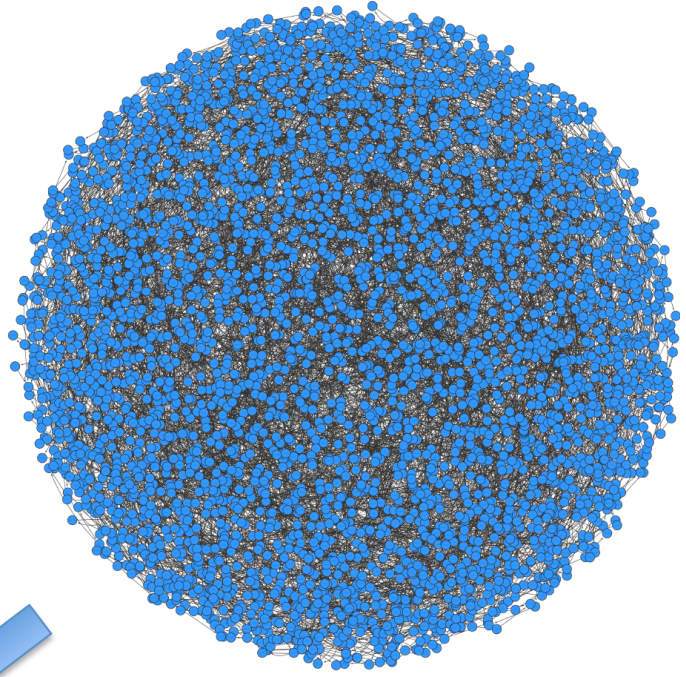

Valine, leucine and isoleucine degradation
PPAR signaling pathway
Fatty acid degradation
Neuroactive ligand-receptor interaction
Protein processing in endoplasmic reticulum
Focal adhesion
Cholinergic synapse
MAPK signaling pathway
Ribosome
Tuberculosis
Phagosome
Staphylococcus aureus infection
Ether lipid metabolism
Parkinson's disease
Thermogenesis
Oxidative phosphorylation
Focal adhesion
Protein digestion and absorption
Complement and coagulation cascades
TNF signaling pathway
Cytokine-cytokine receptor interaction
Jak-STAT signaling pathway
Peroxisome
Fc gamma R-mediated phagocytosis
Staphylococcus aureus infection
Ribosome
Parkinson's disease
Alzheimer's disease
Oxidative phosphorylation
Calcium signaling pathway
Glutamatergic synapse
Neuroactive ligand-receptor interaction
Herpes simplex infection
Measles
Influenza A
yellow
turquoise
tan
snow
salmon
red
purple
maroon
magenta
lightgreen
honeydew
grey
beige
blue
greenyellow
MAYO-microglial


Oxidative phosphorylation
Parkinson's disease
Huntington's disease
Pathways in cancer
PI3K-Akt signaling pathway
Endocytosis
Fatty acid degradation
Valine, leucine and isoleucine degradation
Arginine and proline metabolism
Glutamatergic synapse
Huntington's disease
Oocyte meiosis
Parkinson's disease
Insulin secretion
Bile secretion
Proximal tubule bicarbonate reclamation
Ubiquitin mediated proteolysis
Endocytosis
Inositol phosphate metabolism
Staphylococcus aureus infec
Phagosome
Leishmaniasis
Ribosome
Renal cell carcinoma
Sphingolipid signaling pathway
Neuroactive ligand-receptor interaction
Calcium signaling pathway
Morphine addiction
Protein processing in endoplasmic reticulum
Antigen processing and presentation
Legionellosis
Ribosome
Retrograde endocannabinoid signaling
Inflammatory mediator regulation of TRP channels
Taste transduction
Hippo signaling pathway
Focal adhesion
Pathways in cancer
Oxidative phosphorylation
Parkinson's disease
Huntington's disease
TNF signaling pathway
Jak-STAT signaling pathway
Cytokine-cytokine receptor interaction
Renal cell carcinoma
Olfactory transduction
Axon guidance
Calcium signaling pathway
Propanoate metabolism
Valine, leucine and isoleucine degradation
Parkinson's disease
Thermogenesis
Oxidative phosphorylation
Spliceosome
TNF signaling pathway
Rap1 signaling pathway
MAPK signaling pathway
Longevity regulating pathway
FoxO signaling pathway
Renal cell carcinoma
DNA replication
Ribosome
Oxidative phosphorylation
Thermogenesis
Oxidative phosphorylation
Ribosome
yellowgreen
white
violet
tan
steelblue
skyblue3
saddlebrown
red
plum1
pink
orange
midnightblue
lightgreen
lightcyan1
lightcyan
ivory
grey
greenyellow
green
floralwhite
darkorange
darkgrey
darkgreen
brown
ROSMAP-microglial

Figure 3B



MAYO-microglial Predictive Network



ROSMAP-microglial Predictive Network

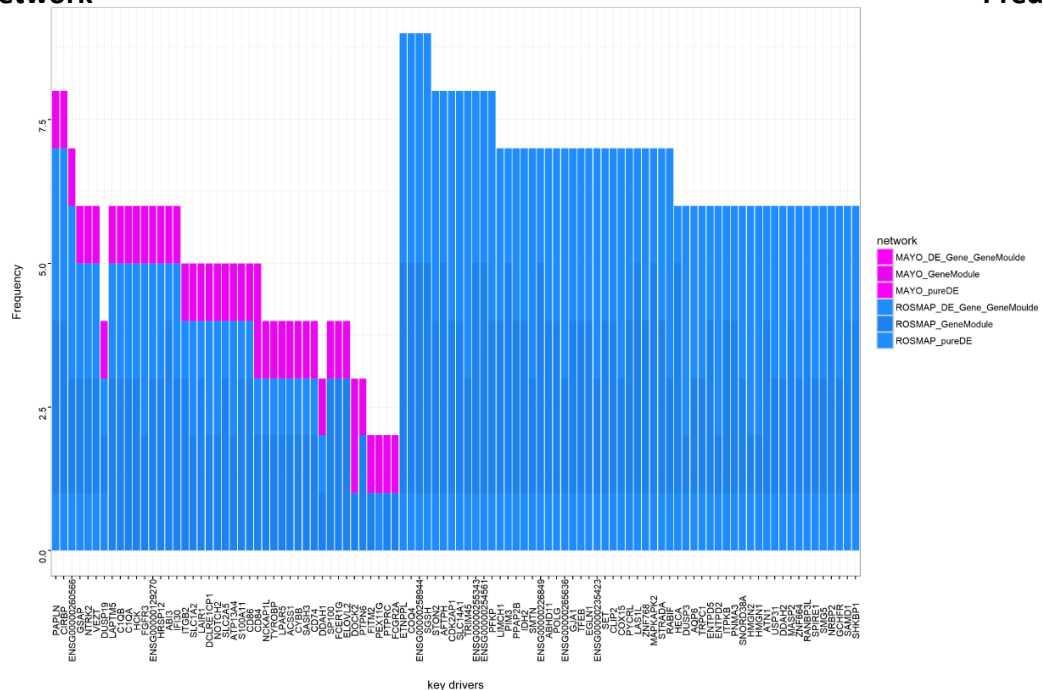


Figure 4A

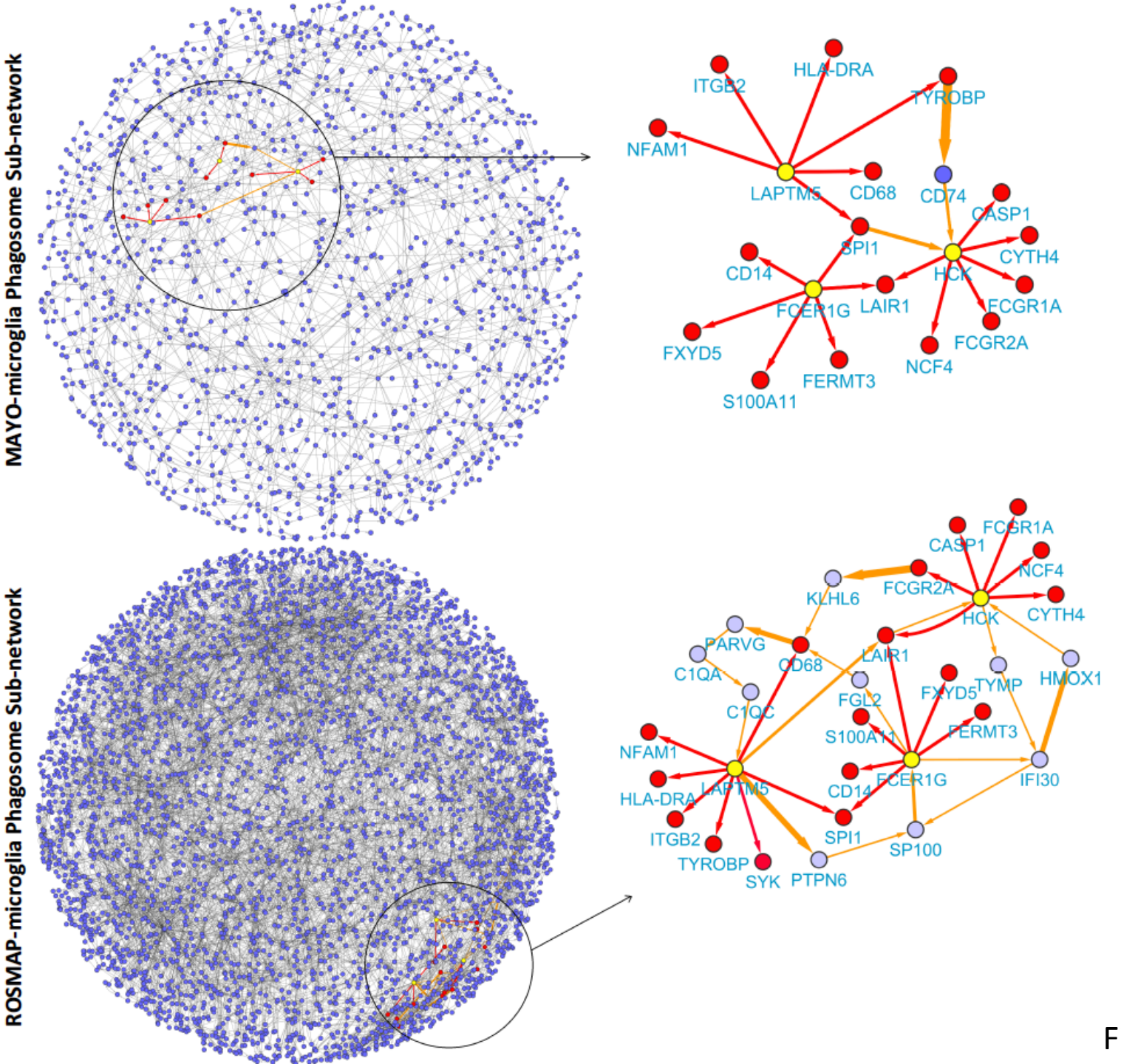
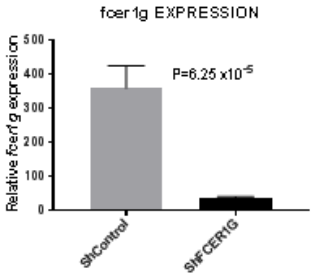
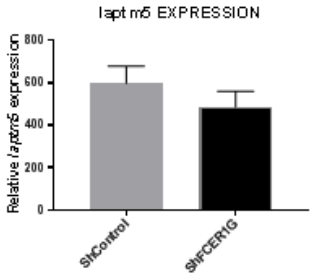
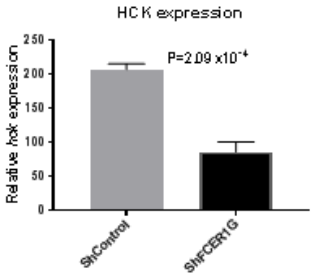

MAYO-microglia Phagosome Sub-network
ITGB2
HLA-DRA
TYROBP
NFAM1
LAPTM5
CD68
CD74
CASP1
SPI1
CYTH4
CD14
HCK
FCER1G
LAIR1
FCGR1A
FXYD5
FCGR2A
NCF4
FERMT3
S100A11
ROSMAP-microglia Phagosome Sub-network
FCGR1A
CASP1
NCF4
KLHL6
FCGR2A
HCK
CYTH4
PARVG
LAIR1
C1QA
CD68
TYMP
HMOX1
FXYD5
FGL2
C1QC
S100A11
FERMT3
NFAM1
IFI30
HLA-DRA
LAPTM5
CD14
FCER1G
ITGB2
SPI1
SP100
TYROBP
SYK
PTPN6

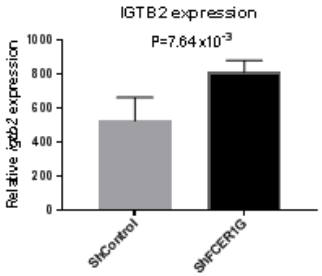
Figure 4B

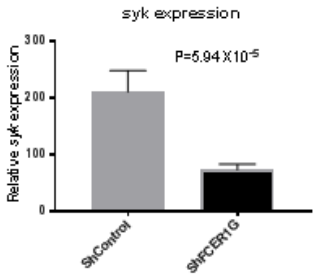
FCER1G Knockdown



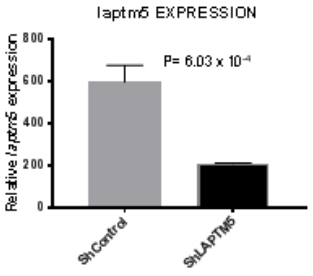


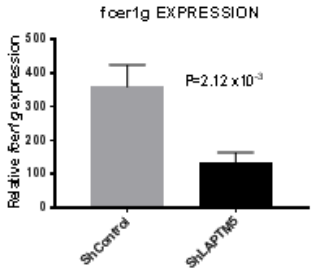


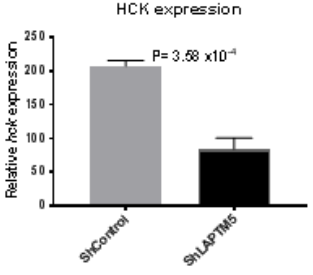


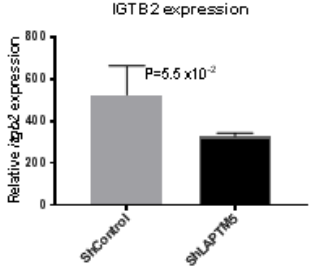


LAPTM5 Knockdown









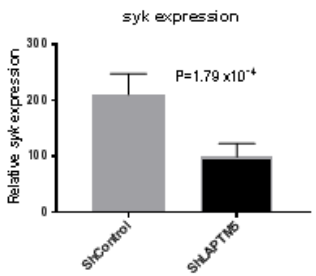
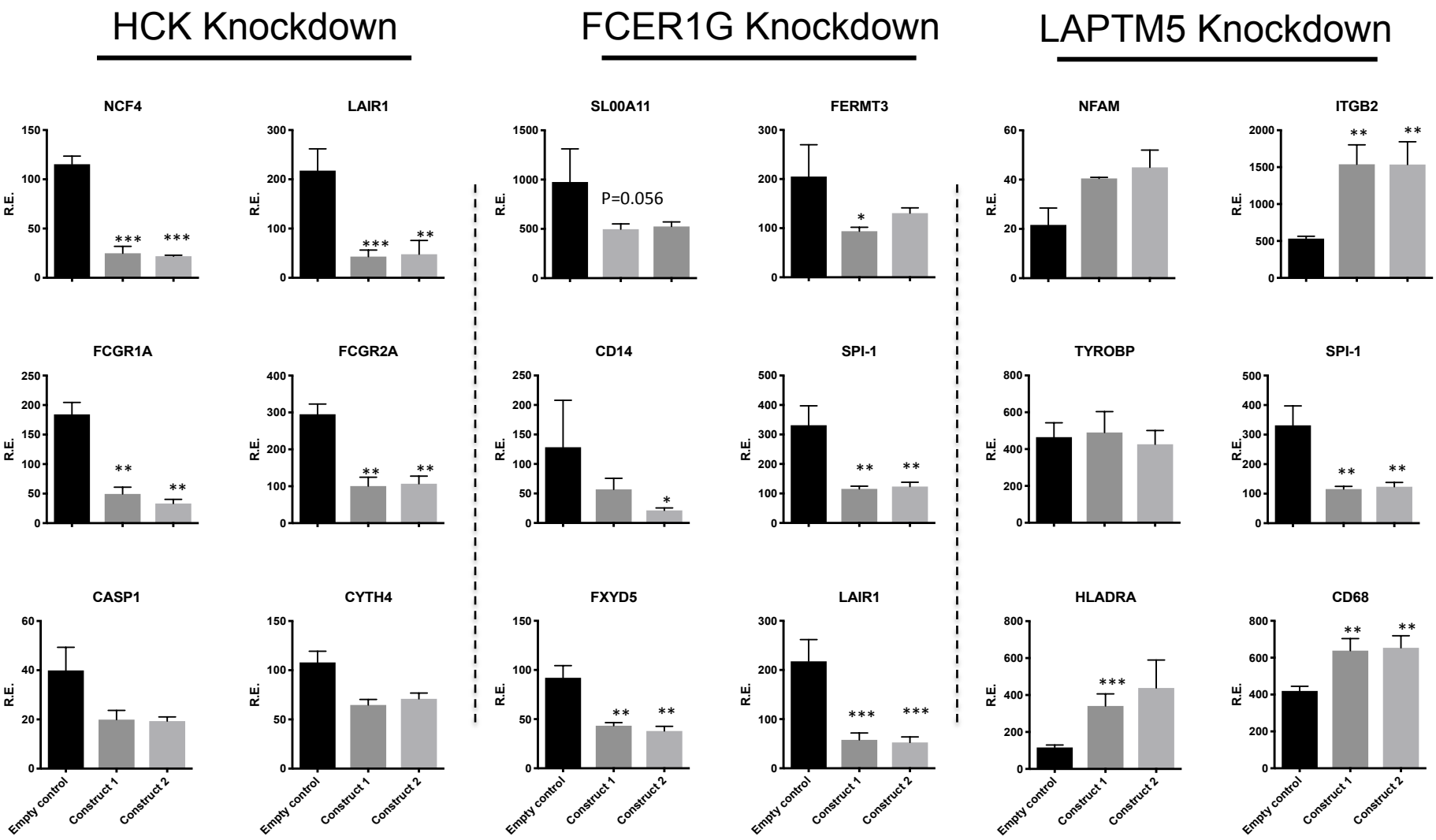


Figure 4C



* P<0.05
** P<0.01
*** P<0.001

Figure 5A

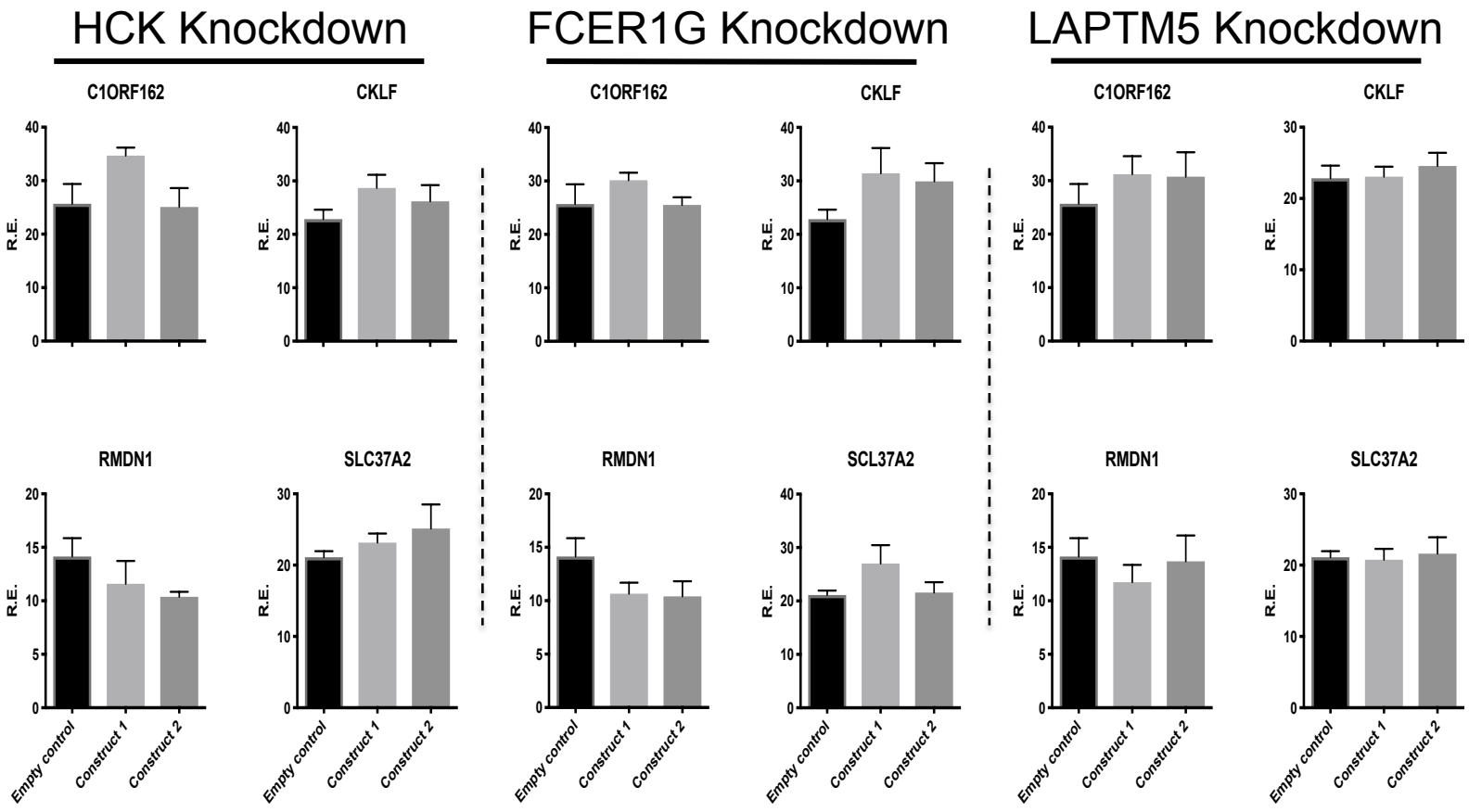
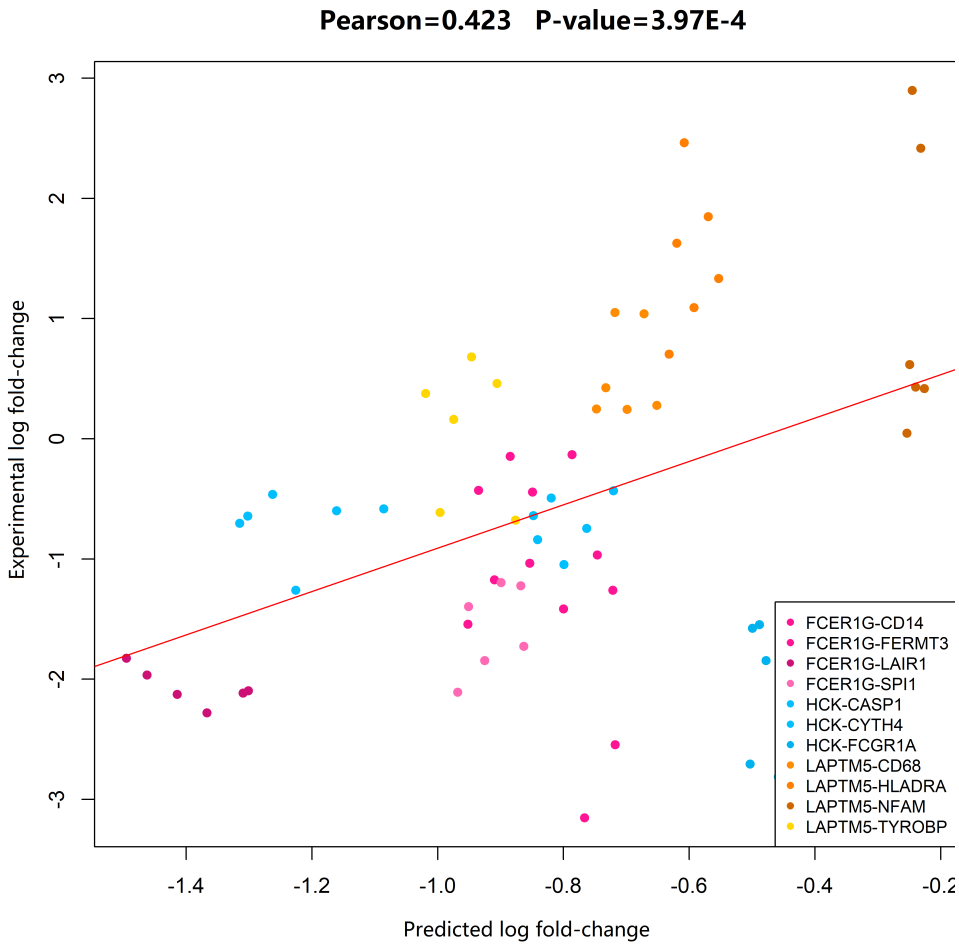

HCK Knockdown
FCER1G Knockdown
LAPTM5 Knockdown
C1ORF162
CKLF
RMDN1
SLC37A2
SCL37A2
R.E.
Empty control
Construct 1
Construct 2

Figure 5B


Pearson=0.423 P-value=3.97E-4
Experimental log fold-change
Predicted log fold-change
FCER1G-CD14
FCER1G-FERMT3
FCER1G-LAIR1
FCER1G-SPI1
HCK-CASP1
HCK-CYTH4
HCK-FCGR1A
LAPTM5-CD68
LAPTM5-HLADRA
LAPTM5-NFAM
LAPTM5-TYROBP

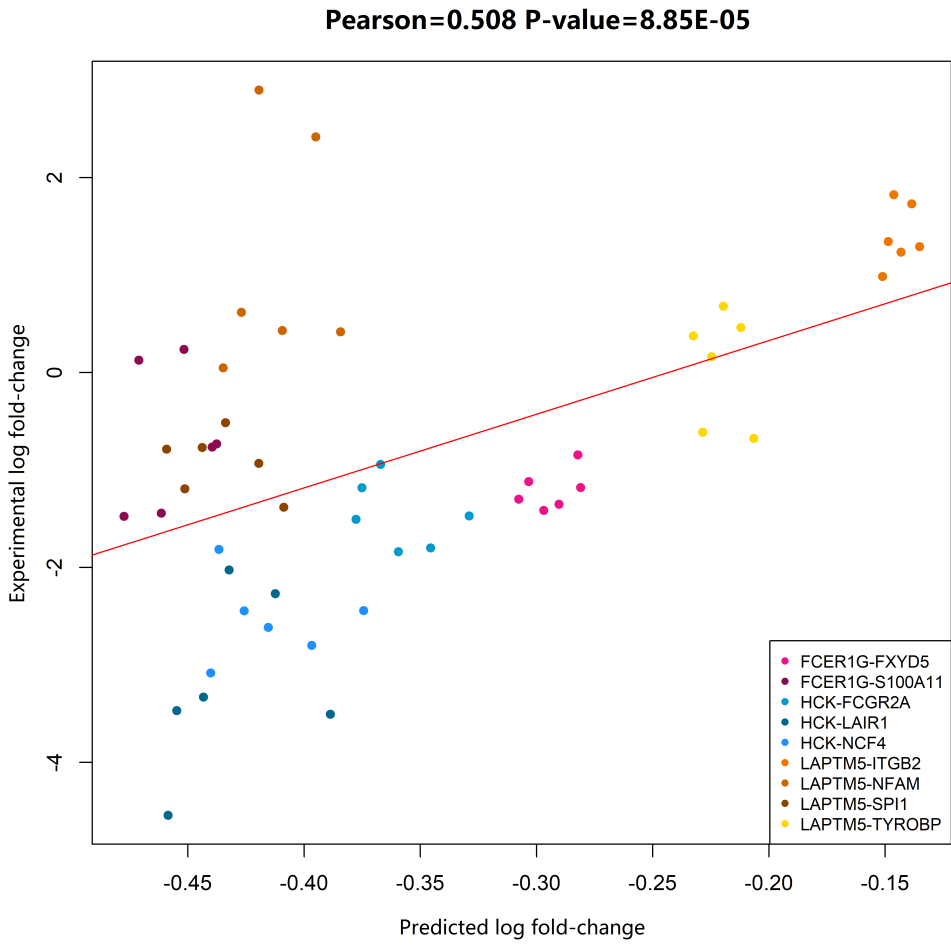
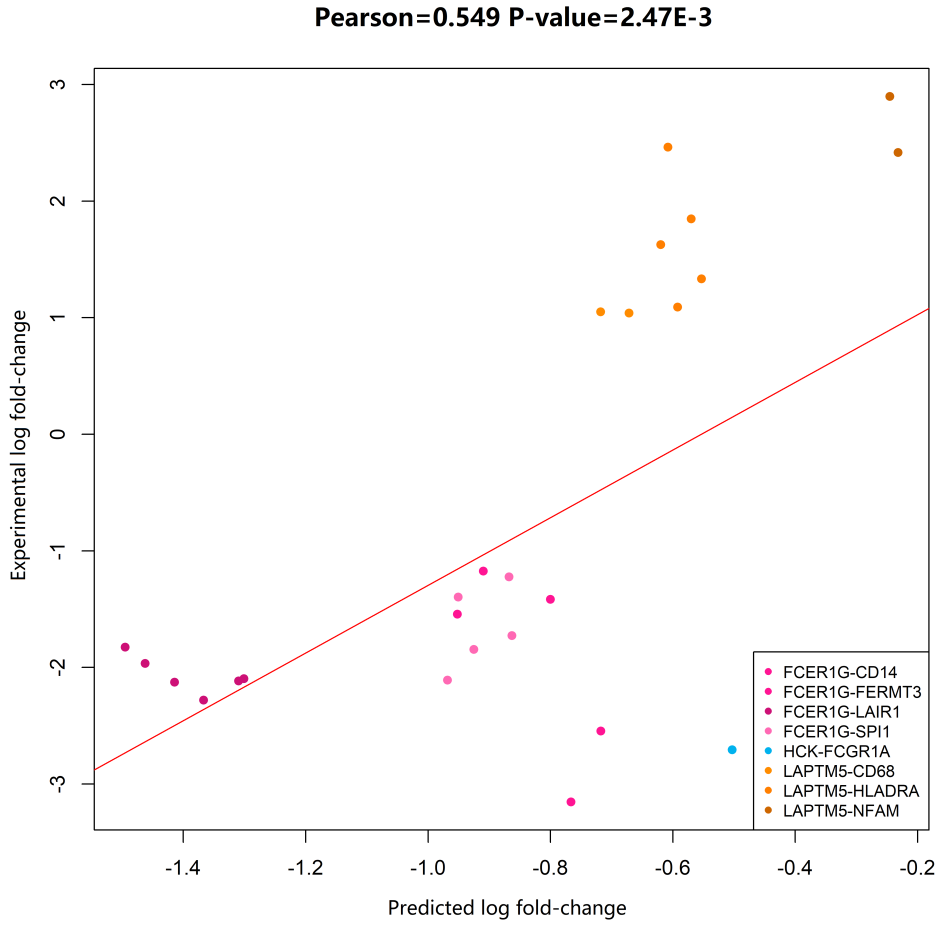

Pearson=0.508 P-value=8.85E-05
Experimental log fold-change
Predicted log fold-change
FCER1G-FXYD5
FCER1G-S100A11
HCK-FCGR2A
HCK-LAIR1
HCK-NCF4
LAPTM5-ITGB2
LAPTM5-NFAM
LAPTM5-SPI1
LAPTM5-TYROBP

Figure 5C


Pearson=0.549 P-value=2.47E-3
Experimental log fold-change
Predicted log fold-change
3
2
1
0
-1
-2
-3
-1.4
-1.2
-1.0
-0.8
-0.6
-0.4
-0.2
FCER1G-CD14
FCER1G-FERMT3
FCER1G-LAIR1
FCER1G-SPI1
HCK-FCGR1A
LAPTM5-CD68
LAPTM5-HLADRA
LAPTM5-NFAM

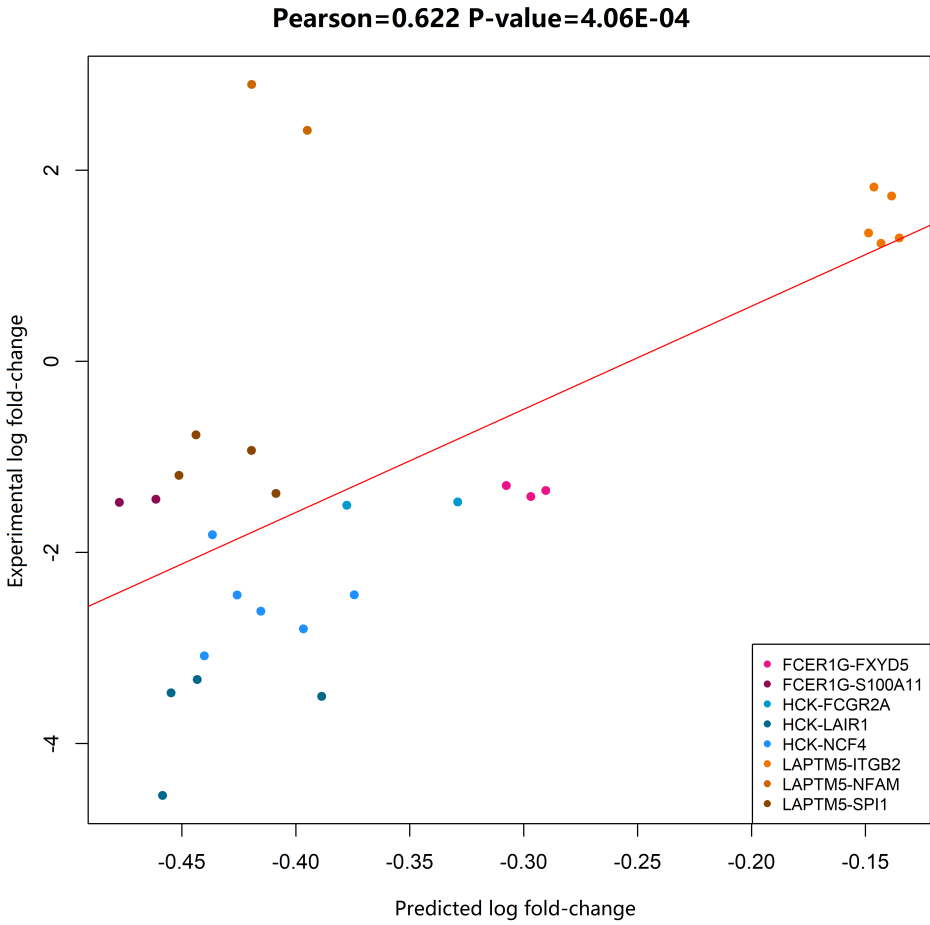
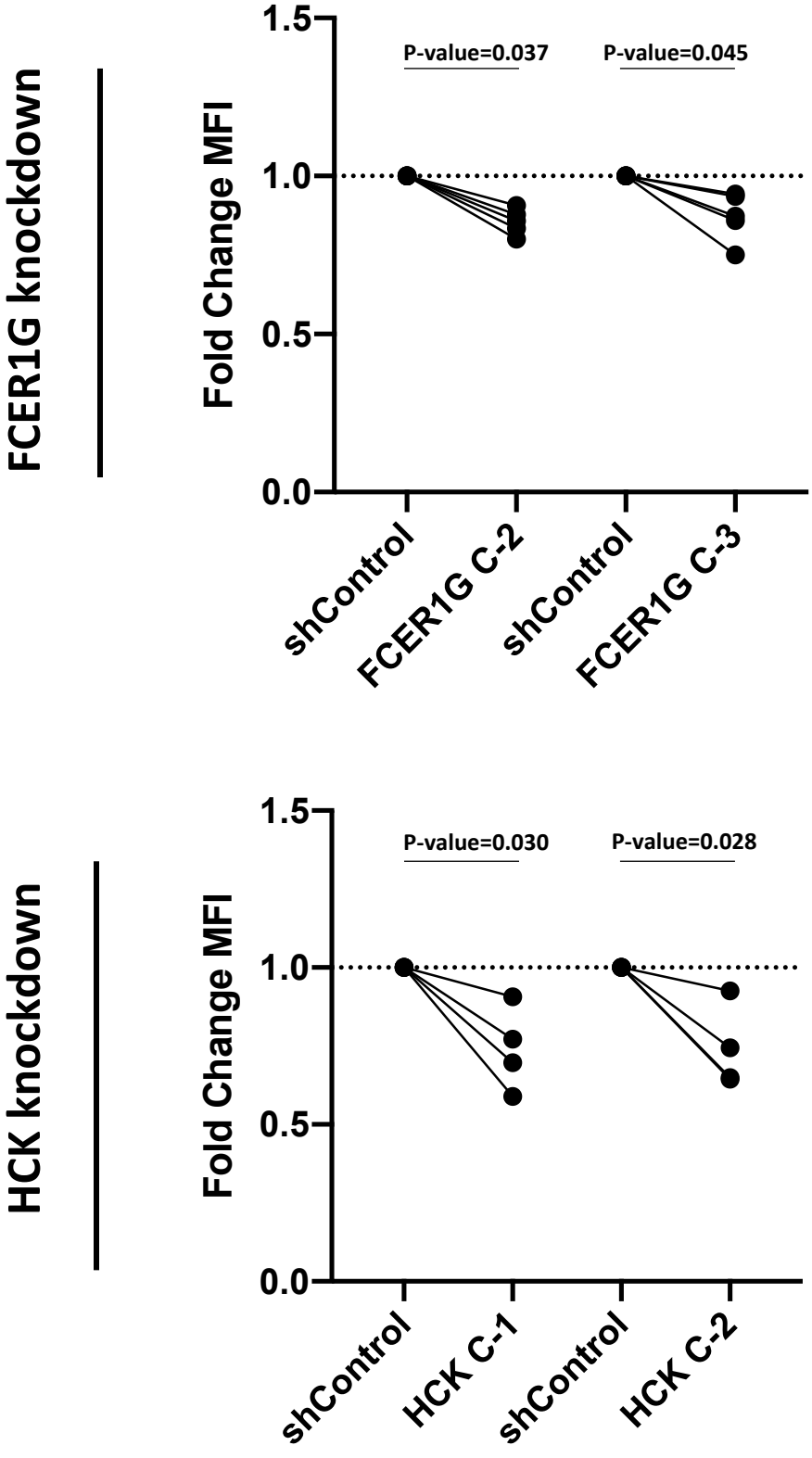

Pearson=0.622 P-value=4.06E-04
Experimental log fold-change
Predicted log fold-change
2
0
-2
-4
-0.45
-0.40
-0.35
-0.30
-0.25
-0.20
-0.15
FCER1G-FXYD5
FCER1G-S100A11
HCK-FCGR2A
HCK-LAIR1
HCK-NCF4
LAPTM5-ITGB2
LAPTM5-NFAM
LAPTM5-SPI1

Figure 5D

Aβ42 uptake



MFI=Mean Fluorescence Intensity

Figure 6

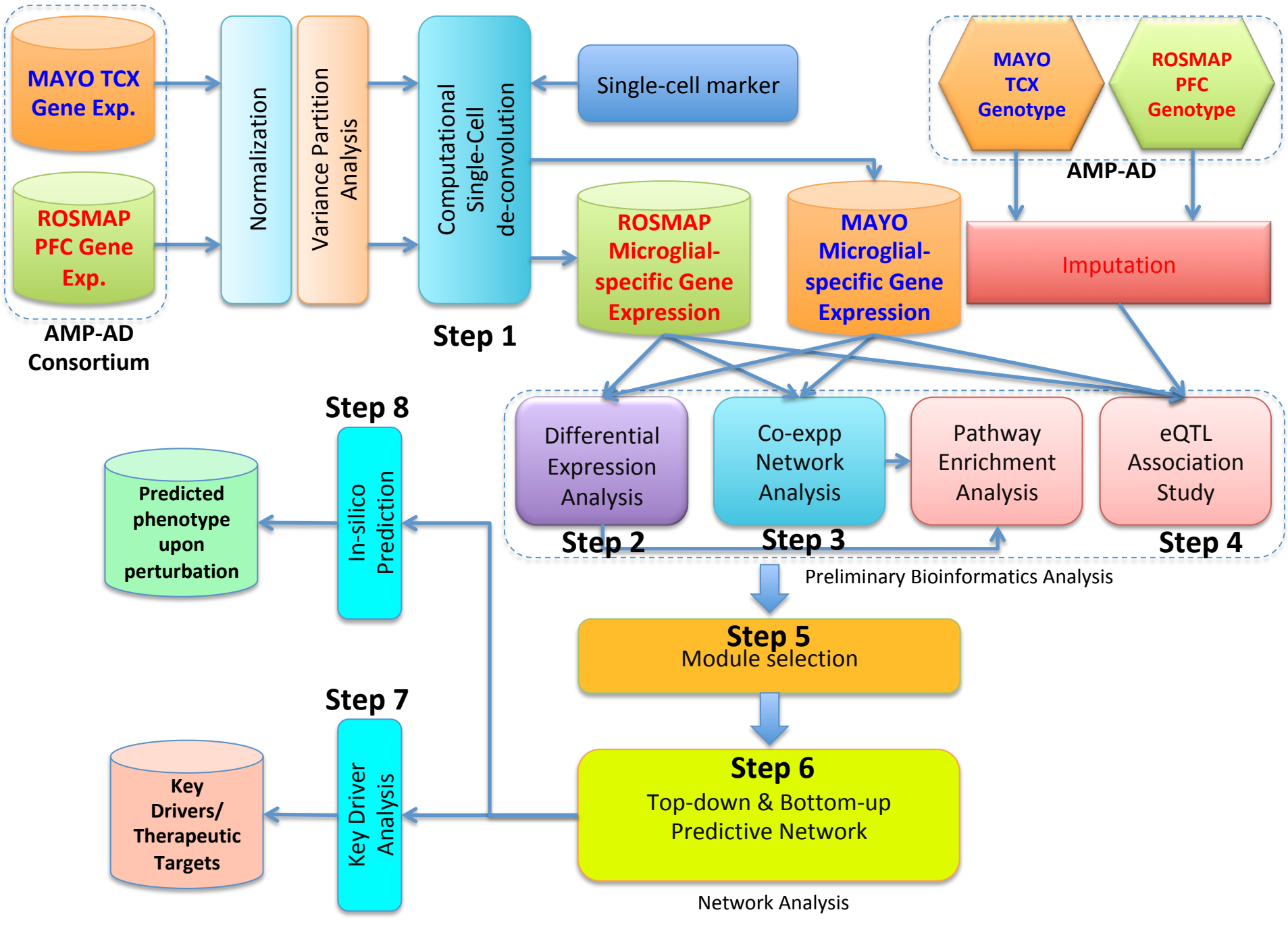

MAYO TCX Gene Exp.
ROSMAP PFC Gene Exp.
AMP-AD Consortium
Normalization
Variance Partition Analysis
Computational Single-Cell de-convolution
Single-cell marker
Step 1
ROSMAP Microglial-specific Gene Expression
MAYO Microglial-specific Gene Expression
MAYO TCX Genotype
ROSMAP PFC Genotype
AMP-AD
Imputation
Differential Expression Analysis
Step 2
Co-expp Network Analysis
Step 3
Pathway Enrichment Analysis
eQTL Association Study
Step 4
Preliminary Bioinformatics Analysis
Step 5
Module selection
Step 6
Top-down & Bottom-up Predictive Network
Network Analysis
Step 7
Key Driver Analysis
Key Drivers/ Therapeutic Targets
Step 8
In-silico Prediction
Predicted phenotype upon perturbation

Figure S1

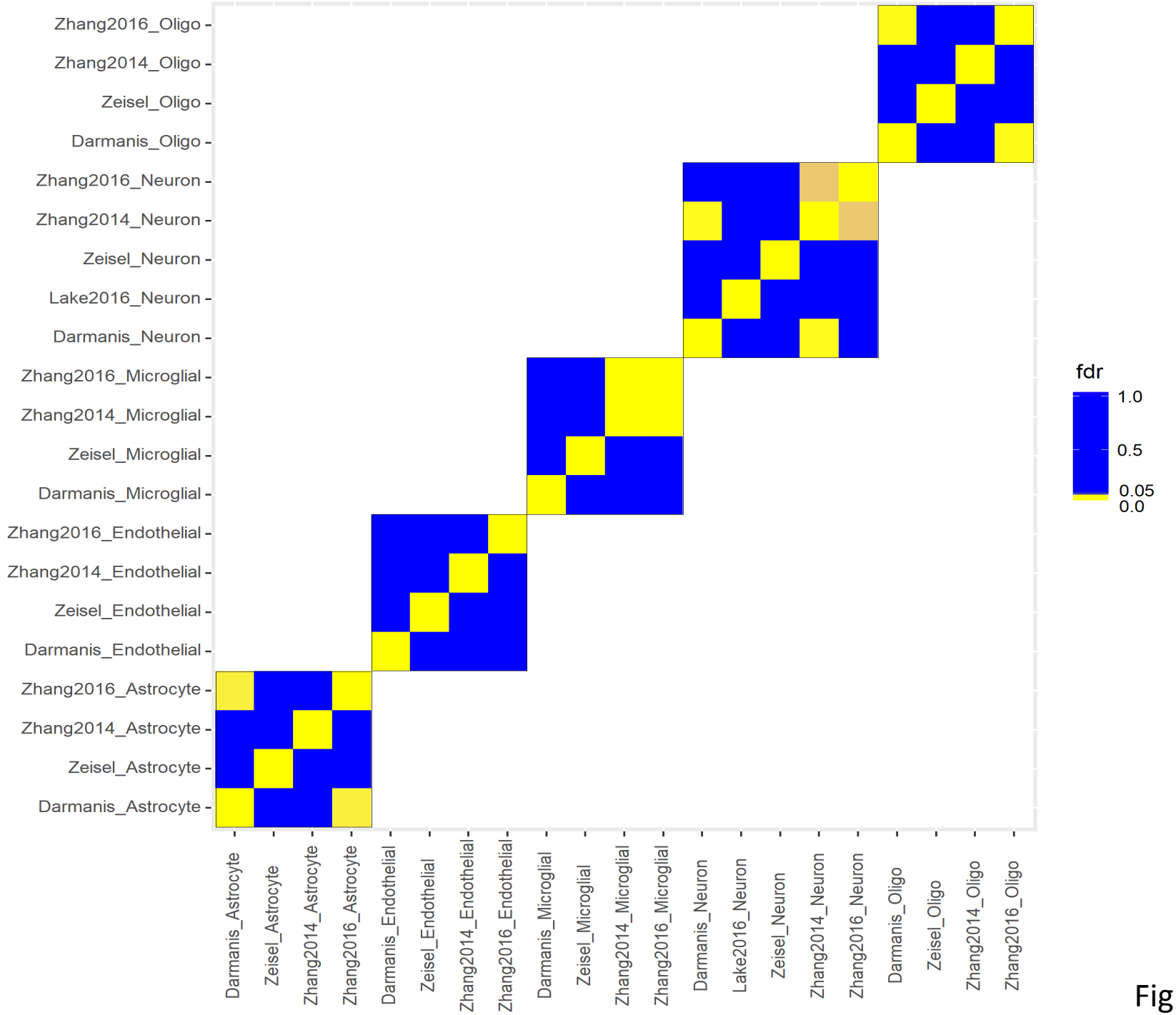
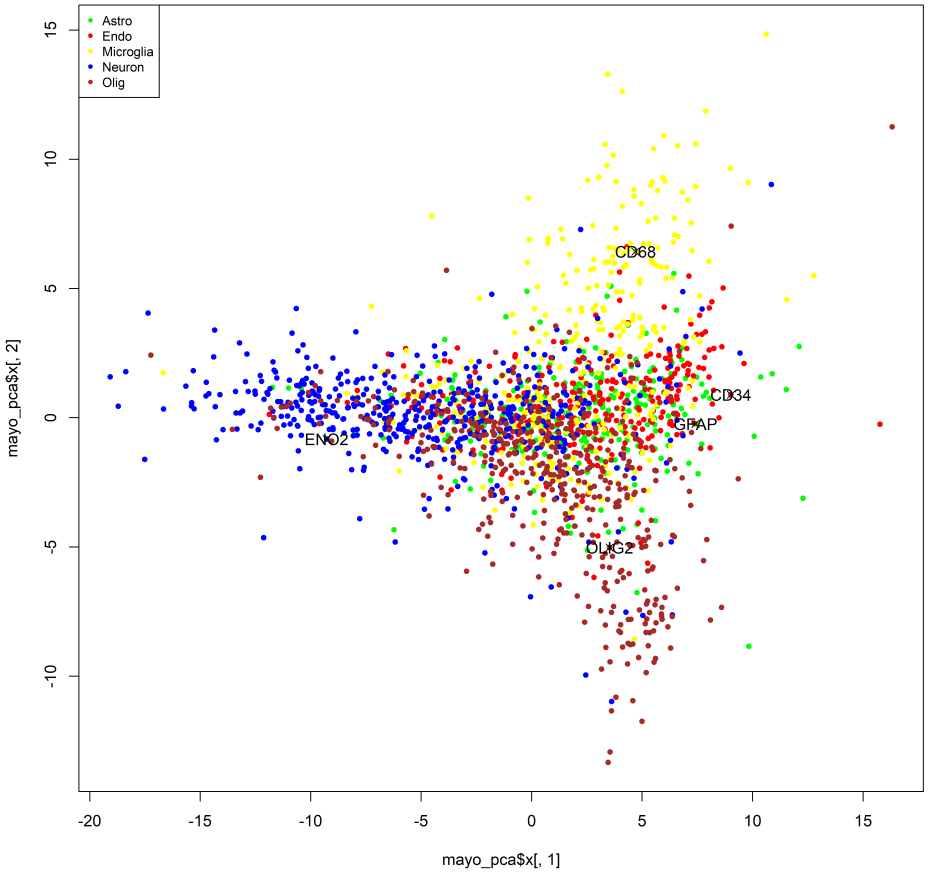
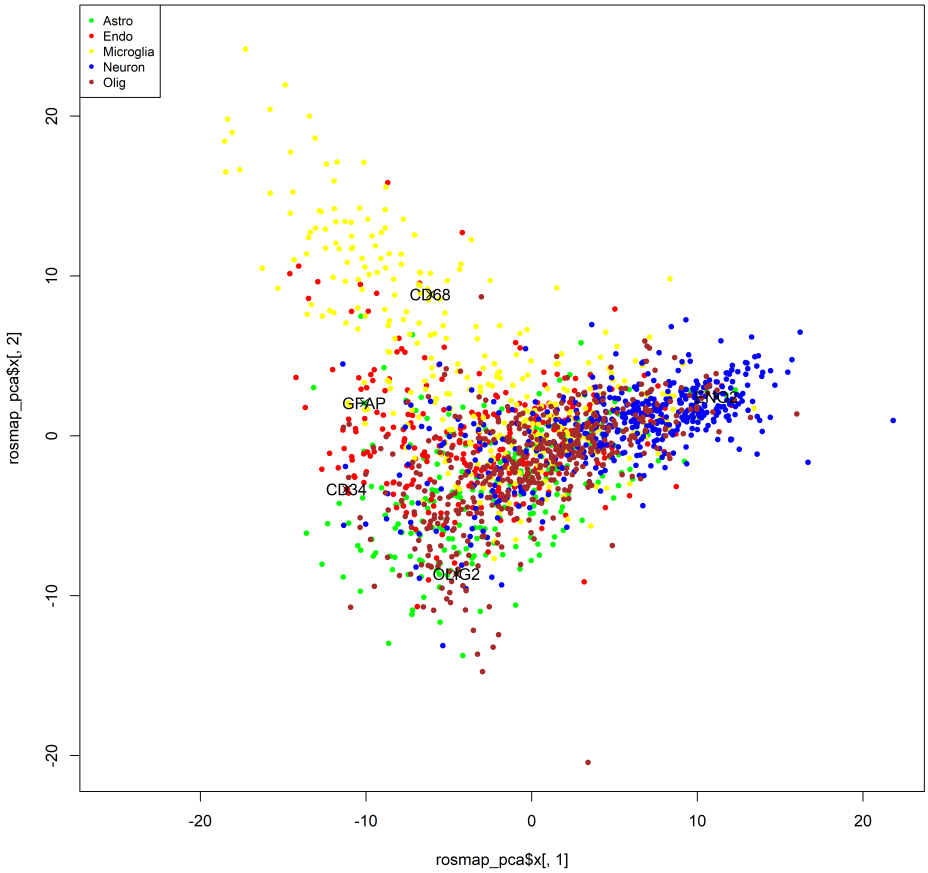

Zhang2016_Oligo
Zhang2014_Oligo
Zeisel_Oligo
Darmanis_Oligo
Zhang2016_Neuron
Zhang2014_Neuron
Zeisel_Neuron
Lake2016_Neuron
Darmanis_Neuron
Zhang2016_Microglial
Zhang2014_Microglial
Zeisel_Microglial
Darmanis_Microglial
Zhang2016_Endothelial
Zhang2014_Endothelial
Zeisel_Endothelial
Darmanis_Endothelial
Zhang2016_Astrocyte
Zhang2014_Astrocyte
Zeisel_Astrocyte
Darmanis_Astrocyte
Darmanis_Astrocyte
Zeisel_Astrocyte
Zhang2014_Astrocyte
Zhang2016_Astrocyte
Darmanis_Endothelial
Zeisel_Endothelial
Zhang2014_Endothelial
Zhang2016_Endothelial
Darmanis_Microglial
Zeisel_Microglial
Zhang2014_Microglial
Zhang2016_Microglial
Darmanis_Neuron
Lake2016_Neuron
Zeisel_Neuron
Zhang2014_Neuron
Zhang2016_Neuron
Darmanis_Oligo
Zeisel_Oligo
Zhang2014_Oligo
Zhang2016_Oligo
fdr
1.0
0.5
0.05
0.0

Figure S2



Mayo Dataset

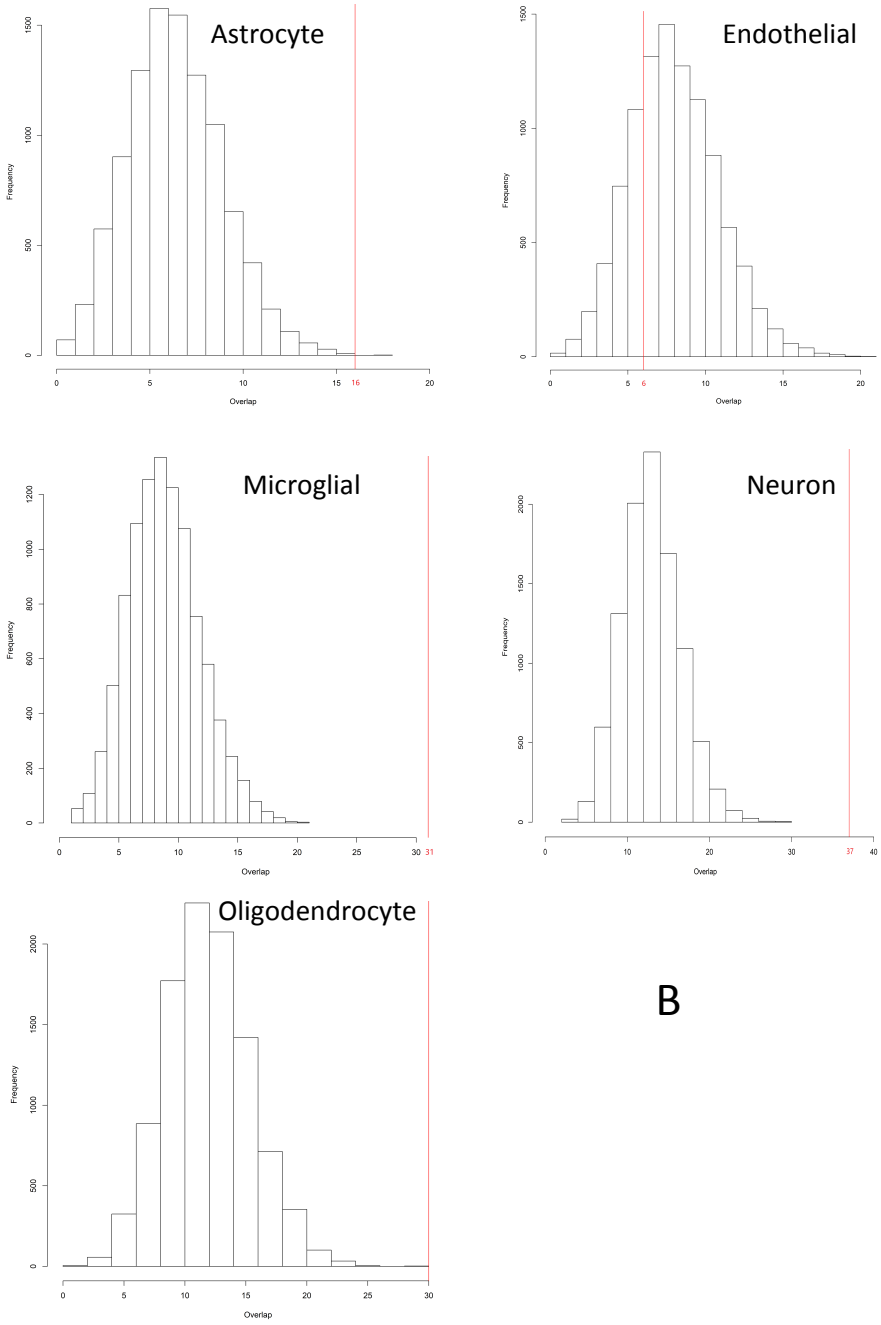


Rosmap Dataset

Figure S3

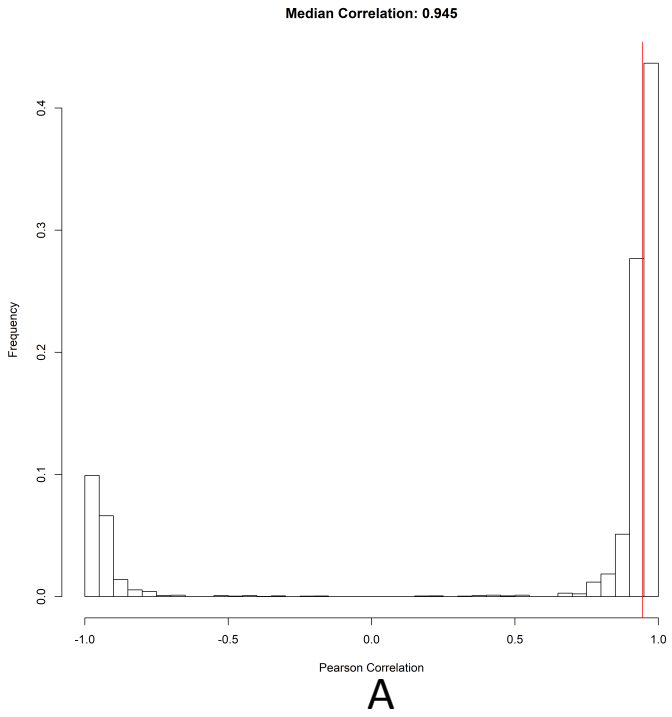
Cell Type	Overlap	P-value
Astrocyte	5.54% (16)	9E-04
Endothelial	1.68% (6)	0.855
Microglial	7.85% (31)	0
Neuron	6.53% (37)	0
Oligodendrocyte	5.77% (30)	0

A

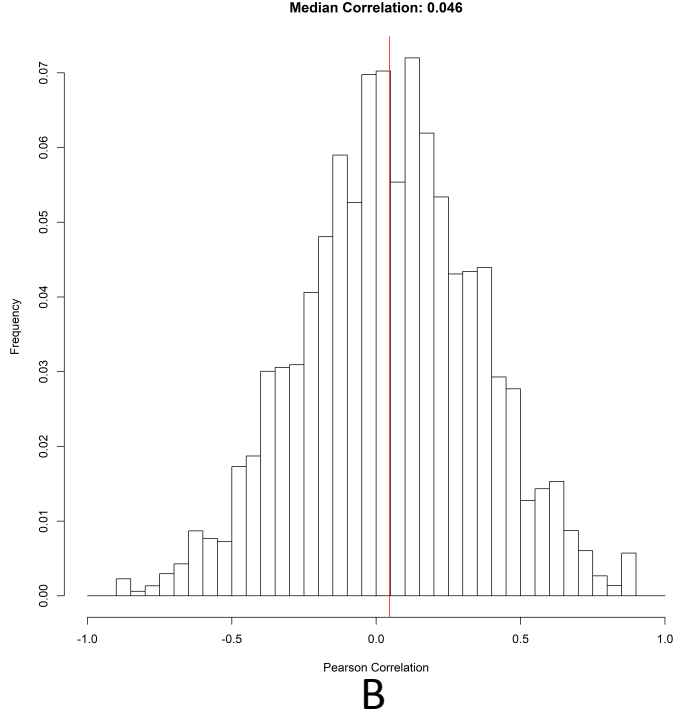


B

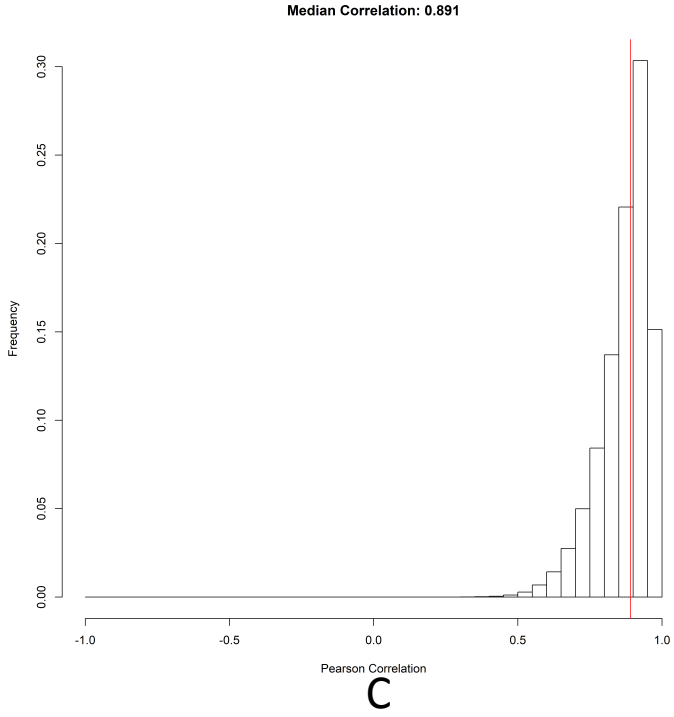
Figure S4



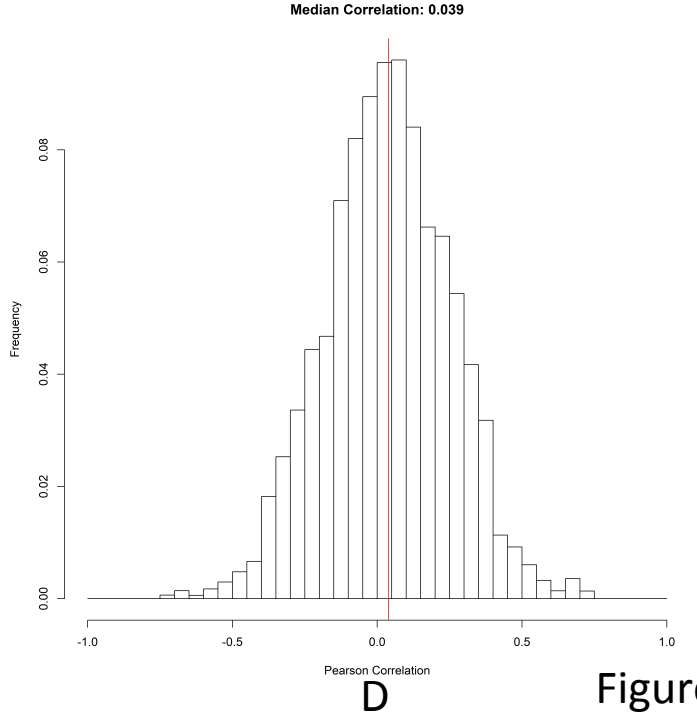
A



B

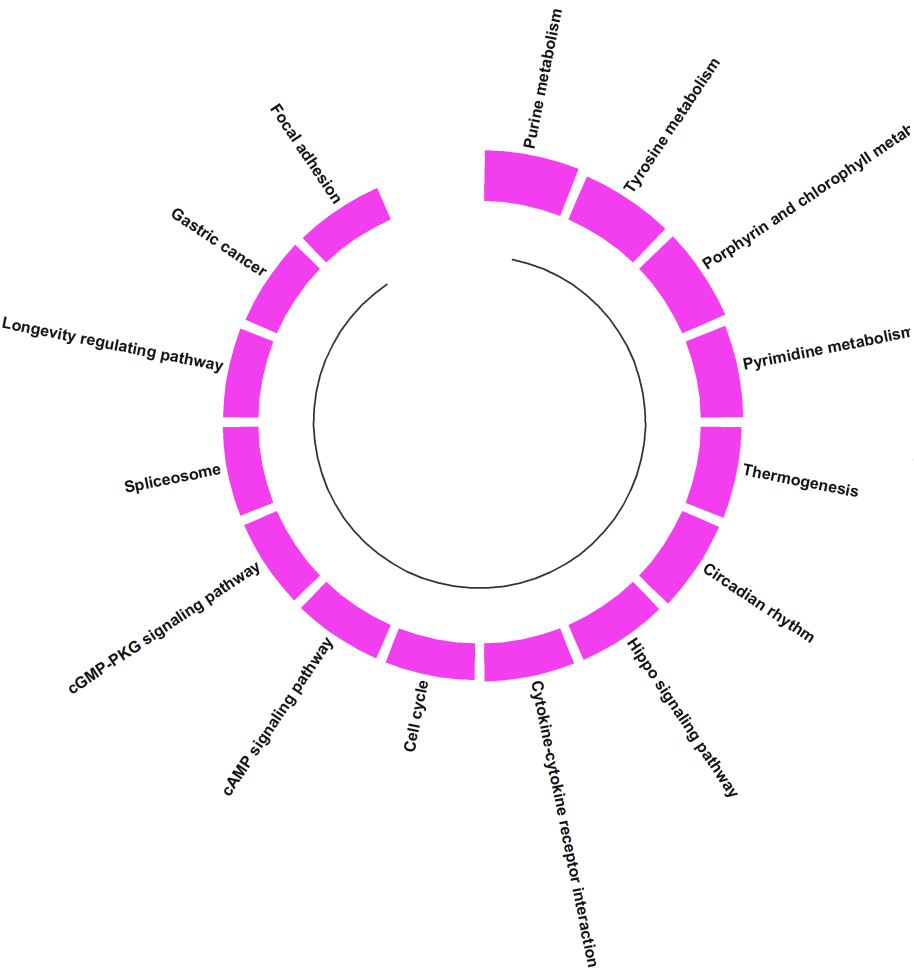


C

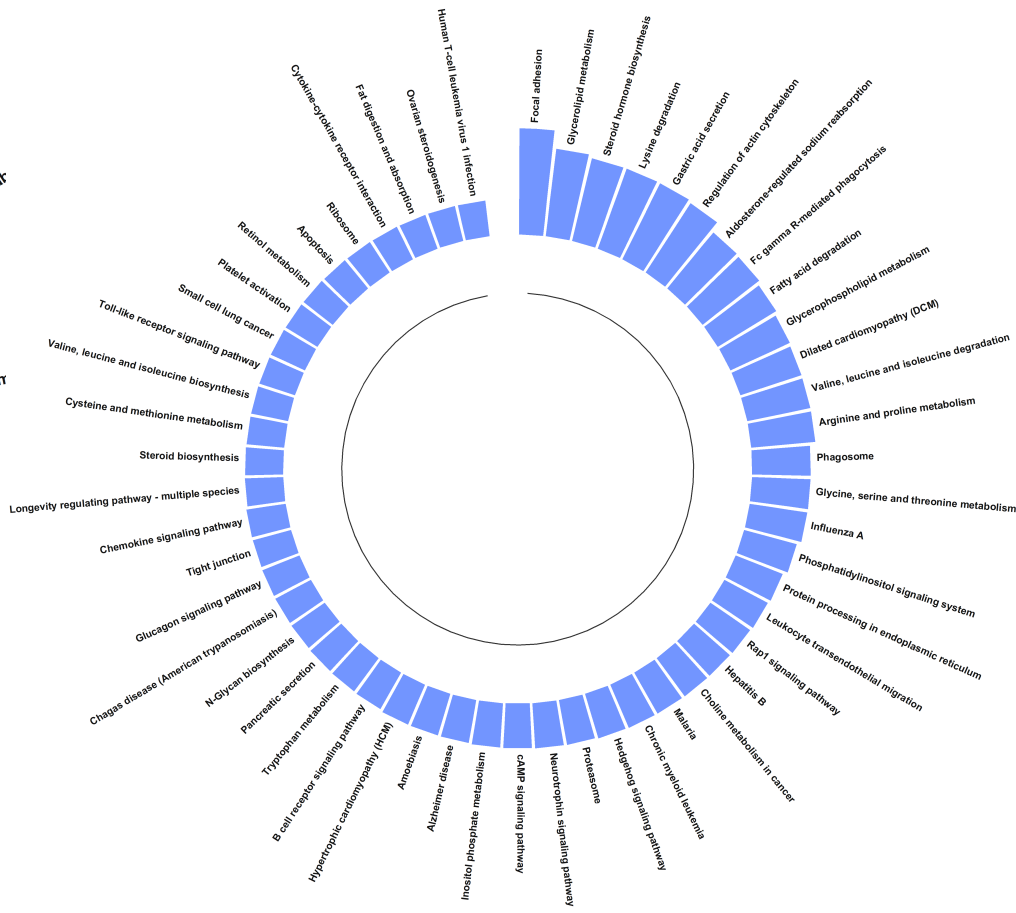


D

Figure S5

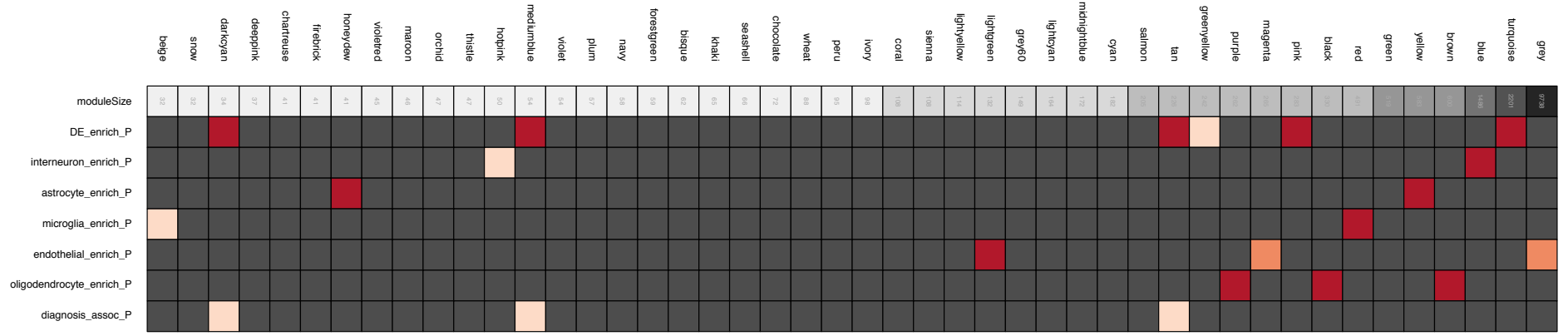


MAYO-microglial

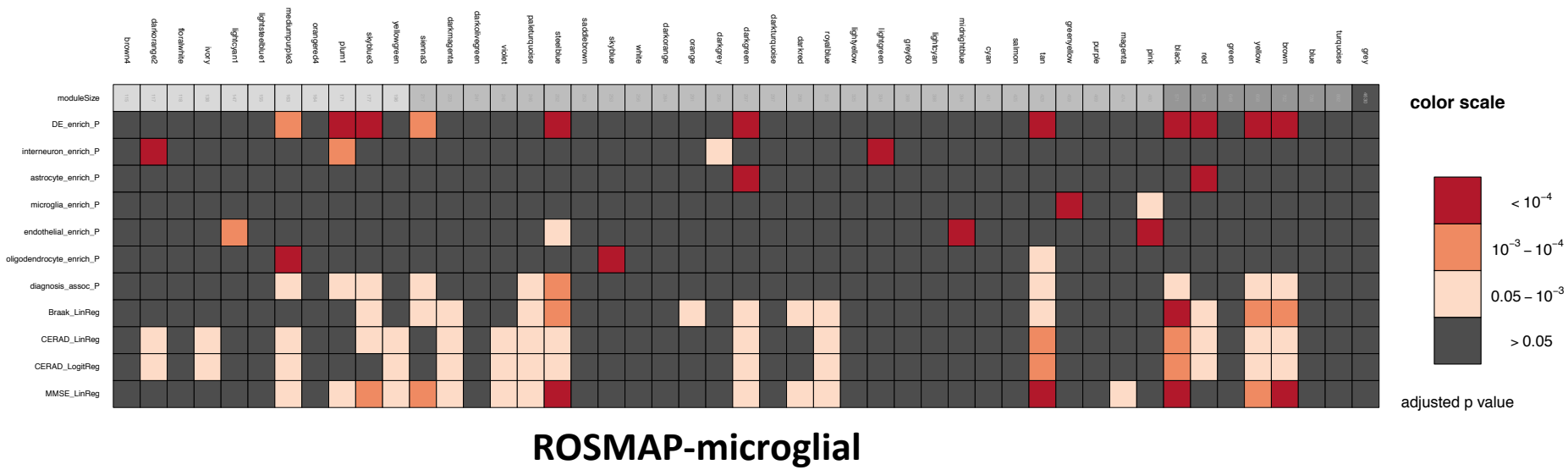


ROSMAP-microglial

Figure S6



MAYO-microglial



ROSMAP-microglial

Figure S7

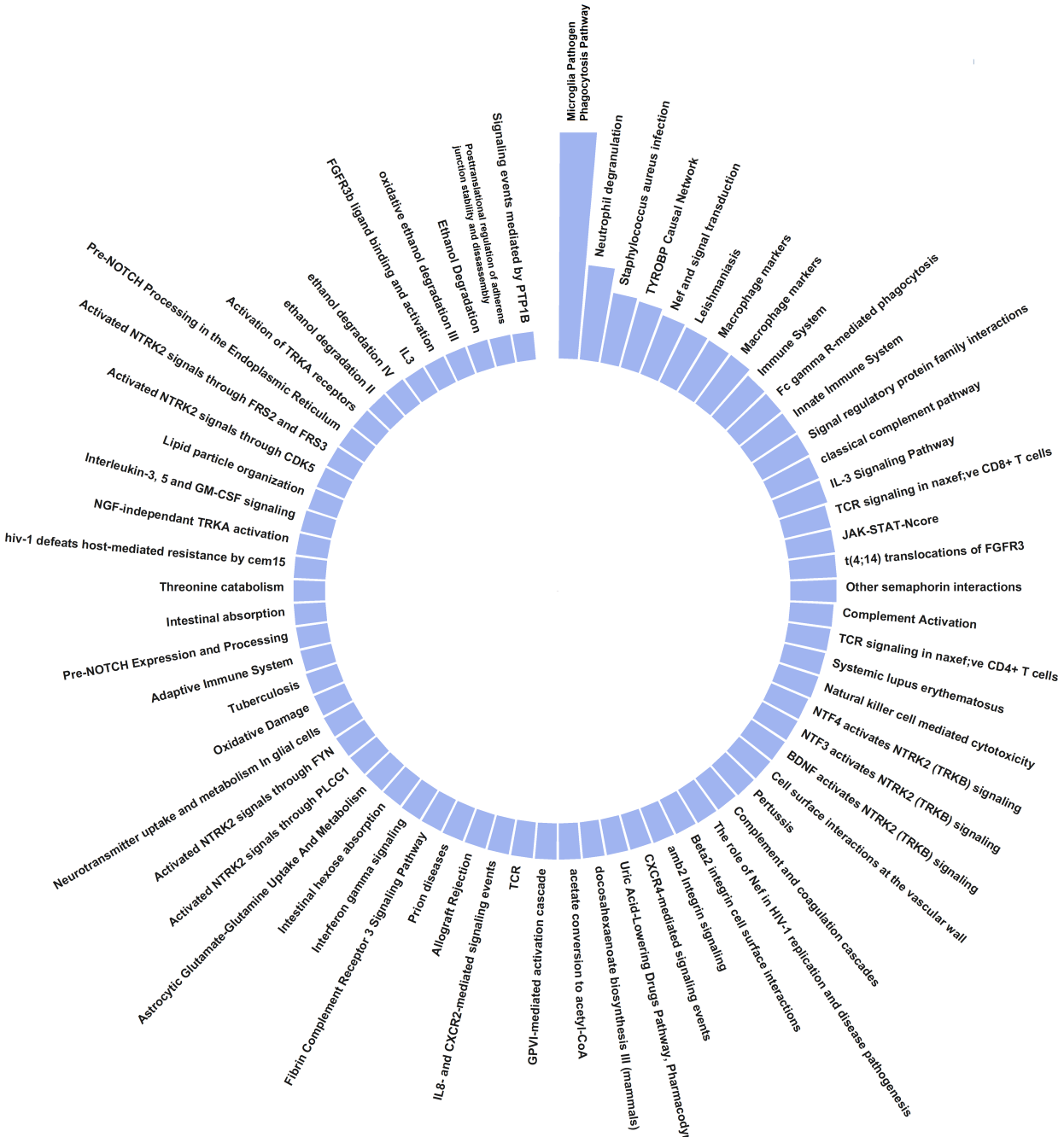
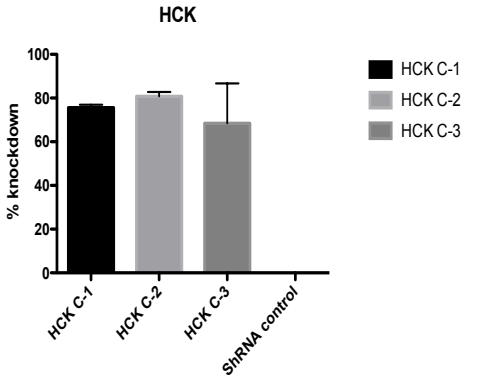
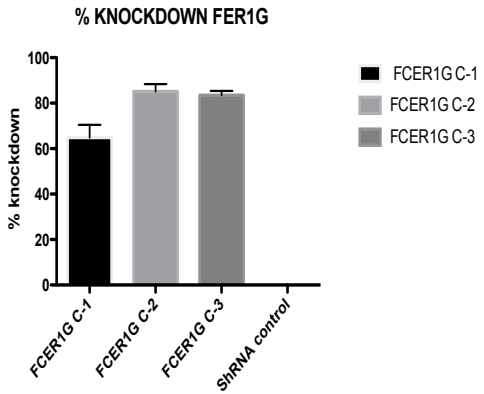
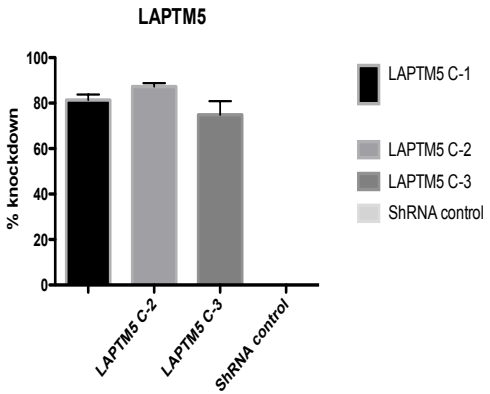
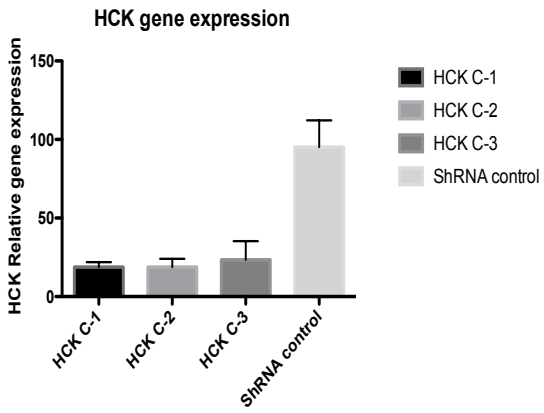

Microglia Pathogen Phagocytosis Pathway
Neutrophil degranulation
Staphylococcus aureus infection
TYROBP Causal Network
Nef and signal transduction
Leishmaniasis
Macrophage markers
Macrophage markers
Immune System
Fc gamma R-mediated phagocytosis
Innate Immune System
Signal regulatory protein family interactions
classical complement pathway
IL-3 Signaling Pathway
TCR signaling in naxef;ve CD8+ T cells
JAK-STAT-Ncore
t(4;14) translocations of FGFR3
Other semaphorin interactions
Complement Activation
TCR signaling in naxef;ve CD4+ T cells
Systemic lupus erythematosus
Natural killer cell mediated cytotoxicity
NTF4 activates NTRK2 (TRKB) signaling
NTF3 activates NTRK2 (TRKB) signaling
BDNF activates NTRK2 (TRKB) signaling
Cell surface interactions at the vascular wall
Pertussis
Complement and coagulation cascades
The role of Nef in HIV-1 replication and disease pathogenesis
Beta2 integrin cell surface interactions
amb2 Integrin signaling
CXCR4-mediated signaling events
Uric Acid-Lowering Drugs Pathway, Pharmacodynamics
docosahexaenoate biosynthesis III (mammals)
acetate conversion to acetyl-CoA
GPVI-mediated activation cascade
TCR
IL8- and CXCR2-mediated signaling events
Allograft Rejection
Prion diseases
Interferon gamma signaling
Fibrin Complement Receptor 3 Signaling Pathway
Intestinal hexose absorption
Astrocytic Glutamate-Glutamine Uptake And Metabolism
Activated NTRK2 signals through PLCG1
Activated NTRK2 signals through FYN
Neurotransmitter uptake and metabolism In glial cells
Oxidative Damage
Tuberculosis
Adaptive Immune System
Pre-NOTCH Expression and Processing
Intestinal absorption
Threonine catabolism
hiv-1 defeats host-mediated resistance by cem15
NGF-independant TRKA activation
Interleukin-3, 5 and GM-CSF signaling
Lipid particle organization
Activated NTRK2 signals through CDK5
Activated NTRK2 signals through FRS2 and FRS3
Pre-NOTCH Processing in the Endoplasmic Reticulum
Activation of TRKA receptors
ethanol degradation II
ethanol degradation IV
IL3
FGFR3b ligand binding and activation
oxidative ethanol degradation III
Ethanol Degradation
Posttranslational regulation of adherens junction stability and disassembly
Signaling events mediated by PTP1B

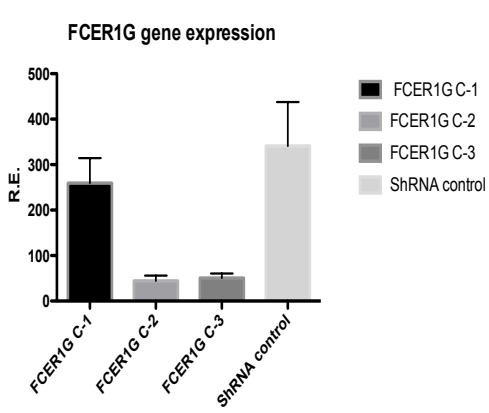
Figure S8


HCK
% knockdown
100
80
60
40
20
0
HCK C-1
HCK C-2
HCK C-3
ShRNA control


% KNOCKDOWN FER1G
% knockdown
100
80
60
40
20
0
FCER1G C-1
FCER1G C-2
FCER1G C-3
ShRNA control


LAPTM5
% knockdown
100
80
60
40
20
0
LAPTM5 C-1
LAPTM5 C-2
LAPTM5 C-3
ShRNA control


HCK gene expression
HCK Relative gene expression
150
100
50
0
HCK C-1
HCK C-2
HCK C-3
ShRNA control


FCER1G gene expression
R.E.
500
400
300
200
100
0
FCER1G C-1
FCER1G C-2
FCER1G C-3
ShRNA control

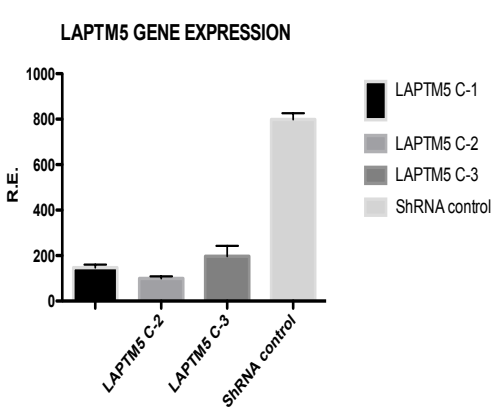

LAPTM5 GENE EXPRESSION
R.E.
1000
800
600
400
200
0
LAPTM5 C-1
LAPTM5 C-2
LAPTM5 C-3
ShRNA control

Figure S9