Integrative analysis of multi-omics reveals gene regulatory networks across brain regions from risk variants to phenotypes of Alzheimer's disease and Covid-19

Saniya Khullar[1,2], Daifeng Wang[1,2,3,*]

[1]Department of Biostatistics and Medical Informatics, University of Wisconsin – Madison, Madison, WI 53076, USA
[2]Waisman Center, University of Wisconsin – Madison, Madison, WI 53705, USA
[3]Department of Computer Sciences, University of Wisconsin – Madison, Madison, WI 53706, USA
[*]corresponding author: daifeng.wang@wisc.edu

Abstract

Background

Genome-wide association studies have found many genetic risk variants associated with Alzheimer's disease (AD). However, how these risk variants affect deeper phenotypes such as disease progression and immune response remains elusive. Also, our understanding of cellular and molecular mechanisms from disease risk variants to various phenotypes is still limited. To address these problems, we performed integrative multi-omics analysis from genotype, transcriptomics, and epigenomics for revealing gene regulatory mechanisms from disease variants to AD phenotypes.

Method

First, we cluster gene co-expression networks and identify gene modules for various AD phenotypes given population gene expression data. Next, we predict the transcription factors (TFs) that significantly regulate the genes in each module and the AD risk variants (e.g., SNPs) interrupting the TF binding sites on the regulatory elements. Finally, we construct a full gene regulatory network linking SNPs, interrupted TFs, and regulatory elements to target genes for each phenotype. This network thus provides mechanistic insights of gene regulation from disease risk variants to AD phenotypes.

Results

We applied our analysis to predict the gene regulatory networks in three major AD-relevant regions: hippocampus, dorsolateral prefrontal cortex (DLPFC), and lateral temporal lobe (LTL). These region networks provide a comprehensive functional genomic map linking AD SNPs to TFs and regulatory elements to target genes for various AD phenotypes. Comparative analyses

further revealed cross-region-conserved and region-specific regulatory networks. For instance, AD SNPs rs13404184 and rs61068452 disrupt the bindings of TF SPI1 that regulates AD gene INPP5D in the hippocampus and lateral temporal lobe. However, SNP rs117863556 interrupts the bindings of TF REST to regulate GAB2 in the DLPFC only. Furthermore, driven by recent discoveries between AD and Covid-19, we found that many genes from our networks regulating Covid-19 pathways are also significantly differentially expressed in severe Covid patients (ICU), suggesting potential regulatory connections between AD and Covid. Thus, we used the machine learning models to predict severe Covid and prioritized highly predictive genes as AD-Covid genes. We also used Decision Curve Analysis to show that our AD-Covid genes outperform known Covid-19 genes for predicting Covid severity and deciding to send patients to ICU or not. In short, our results provide a deeper understanding of the interplay among multi-omics, brain regions, and AD phenotypes, including disease progression and Covid response. Our analysis is open-source available at https://github.com/daifengwanglab/ADSNPheno.

Introduction

Alzheimer's Disease (AD), a neurodegenerative disorder and form of dementia, affects more than 50 million elders in the world[1]. In particular, late-onset AD (LOAD), which comprises 97% to 99% of all cases, usually occurs in individuals older than 65[2]. The AD patients experience memory loss, cognitive decline, and weak executive function, as reflected in their poor Mini-Mental State Examination (MMSE) results[1]. Furthermore, AD is a complex process. Many molecular changes from underlying biological mechanisms, such as an accumulation of amyloid-beta plaques, neurofibrillary tangles (NFTs) within neurons, and neuroinflammation, have been associated with AD progression phenotypes[3]. However, beyond these associations, the causal molecular mechanisms of AD, especially for the disease progression, remain elusive.

Several AD cohorts have measured the genome-wide gene expression data at the population level. Also, these data cover different brain regions in AD, such as the Hippocampus, Dorsolateral Prefrontal Cortex (DLPFC), and Lateral Temporal Lobe (LTL). For instance, the Hippocampus Cornu Ammonis 1 (CA1) region—which is crucial for autobiographical memory, mental time travel, and self-awareness—usually has the most significant loss in memory ability, neurogenesis, volume, and neuronal density in the AD Hippocampus[4]. The LTL contains the cerebral cortex (responsible for hearing, understanding language, visual processing, and facial recognition)[5] and is impacted early in AD[6]. The DLPFC is involved in executive functioning (working memory and selective attention), supports cognitive responses to sensory information[7], works with the Hippocampus to help mediate complex cognitive functions[8], and has plasticity deficits in AD patients[9]. These datasets thus enable finding genes and transcriptional activities that associate with AD phenotypes from the populations, providing molecular mechanistic insights into AD.

For example, differential expression analyses for the temporal lobes and frontal lobes found many key Differentially Expressed genes (DEGs) in AD, including ABCA1 and 2, C1R and C1S, VGF, REST, GAD1 and 2, SST, and CALB1[10]. Further, gene co-expression network analysis such as WGCNA has been widely applied to these population data to identify various gene co-

expression modules[11]. The genes in the same module show similar expression dynamics across AD phenotypes (e.g., progression stages), implying that they involve certain shared molecular mechanisms dysregulated in AD[12]. A previous study using the Hippocampal gene expression dataset, built gene co-expression networks to find enriched functions for AD and potential target genes for AD therapy and found 19 co-expression modules and key hub genes for AD, such as MT1, MT2, MSX1, NOTCH2, ADD3, and RAB31[13]. Other studies with multiscale network analysis of different AD brain regions identified cell subtype-specific AD drivers, including GABR2, LRP10, MSN, PLP1, and ATP6V1A[14]. Nevertheless, it is also vital to understand underlying gene regulatory mechanisms that control those DEGs, co-expression networks, and gene expression dynamics, which is still unclear.

Recent Genome-Wide Association Studies (GWAS) have identified various genetic variants associated with AD (e.g., explaining the variations of AD phenotypes)[15]. Linking those AD variants to genes and genomic functions provides a deeper understanding of molecular causes in AD. For instance, GWAS studies linked AD risk SNPs from 20 loci to AD genes (e.g., AP4E1, AP4M1, APBB3, BIN1, MS4A4A, MS4A6A, PILRA, RABEP1, SPI1, TP53INP1, and ZYX) by their proximity to the coding regions[16]. New Proteome-Wide Association Studies (PWAS) identified 11 additional causal AD genes involved in protein abundance[17]. However, since most AD-associated variants are located on non-coding DNA regions, identifying potentially causal AD genes from the variants is still challenging. To address this, functional genomics and gene regulatory network (GRN) analyses have been widely used to predict the biological functions and pathways that can be affected by disease variants. For example, gene regulatory network (GRN) is a crucial mechanism fundamentally controlling gene expression, such that the transcription factors (TFs) bind to the non-coding regulatory elements (e.g., promoters, enhancers) to initialize transcription. Non-coding disease SNPs may disrupt the binding sites of TFs (TFBSs) on the non-coding regulatory elements to cause the abnormal gene expression that potentially leads to diseases and disease phenotypes (e.g., disease genes). Also, many tools have been thus developed to discover such TFBS-disrupting SNPs such as SNP2TFBS[18] and atSNP[19]. Using the TFBS-disrupting SNPs, recent studies have identified many disease genes, such as for Schizophrenia[20]. However, it is still unclear how these AD-associated variants cause the gene expression for various AD phenotypes, especially during AD progression.

To address the issues as above, we performed an integrative analysis of multi-omics to reveal the functional genomics and GRNs from AD variants to AD phenotypes (**Fig. 1**, **Methods**). In particular, the input to our analysis includes the population gene expression and phenotype data, epigenomic data, and AD risk SNPs. First, using a population gene expression dataset in AD, our analysis builds the gene co-expression network, clusters the network into gene co-expression modules, and predicts the TFs that co-regulate the modular genes. The genes and modules are associated with various AD phenotypes via modular expression patterns. Then, we integrate the chromatin interaction data (e.g., Hi-C) to further link the regulatory elements (e.g., promoters, enhancers) to the genes. The binding sites of TFs are also used to link TFs to the regulatory elements. After this step, we predict a full GRN linking TFs and regulatory elements to genes and gene modules. Enrichment analysis of each gene module further links the

network to functions and pathways. Finally, we look at TFs with binding sites interrupted by AD SNPs and use them to subset the full network. This subnetwork thus reveals a map linking AD SNPs to interrupted TFs to regulatory elements to genes and modules to enriched functions/pathways to AD phenotypes, providing the mechanistic insights of AD variants to phenotypes. This paper applied this integrative analysis to population gene expression data in three major regions: Hippocampus, DLPFC, and LTL. We identified brain-specific regulatory networks for various AD phenotypes, especially on AD progression.

In addition, the recent surging SARS-CoV-2 virus (Covid-19) has widely affected elders, especially with neurogenerative diseases. For example, studies have found that Covid-19 may increase a person's risk for Alzheimer's, Parkinson's, and other brain disorders. Elders with AD are at a significantly higher risk of severe Covid-19 outcomes[21]. Moreover, Covid-19 survivors show an increased risk of neurological and psychiatric problems, known as Neuro-COVID[22]. Also, Covid-19 morbidity and mortality have been linked to an overactivated and exaggerated immune system response. Recent studies have found that the innate immune system may go awry in AD and maybe a driver of cognitive decline, neuroinflammation, neurodegeneration, and overall AD pathology[23]. Therefore, both Covid-19 and AD are associated with a dysregulation of the innate immune system response, and the excessive inflammation and severe immune response in Covid-19 could advance the progression of neuroinflammatory neurodegenerative diseases like AD[23]. For instance, several AD risk genes, like APOE4, have been identified with increased susceptibility to severe Covid-19[24]. Since Covid-19 serves as a strong marker for an exaggerated and overreactive immune system, elucidating pathways disrupted in Covid-19 and AD may provide more insights on the role of a misguided immune system in AD onset and progression. However, underlying gene functions linking the immunological functions from Covid-19 to AD are unknown. To better understand this, we looked at brain-region gene regulatory networks from our analysis that target AD genes relating to the immunological functions and pathways, including Covid-19. Using independent gene expression data for Covid, we found that many genes in the networks are significantly differentially expressed in severe Covid patients (ICU), suggesting abnormal expression activities of those genes in Covid-19. Therefore, we finally trained a machine learning model to predict severe Covid-19 from those network genes and prioritized the highly predictive genes as a set of AD-COVID marker genes. This marker set provides a potentially novel map for understanding the functional interplay between the immune system, Covid-19, and Alzheimer's Disease.

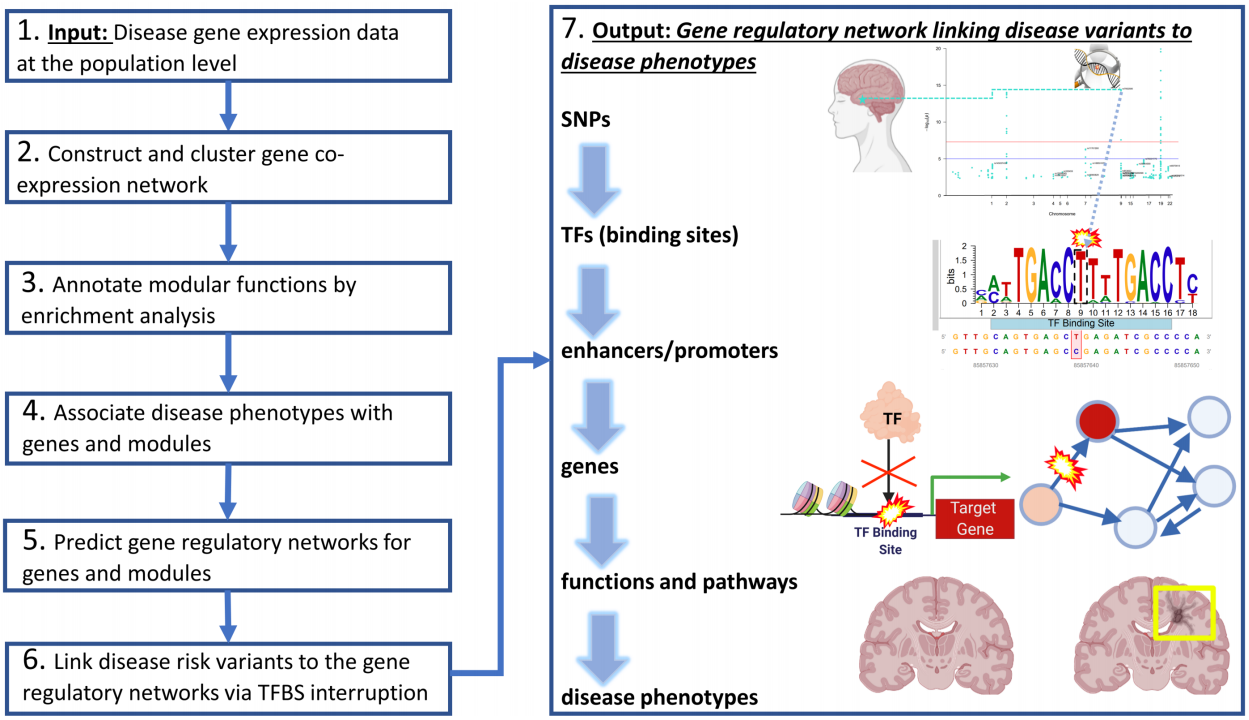


Fig. 1 – Integrative analyses to predict gene regulatory networks from disease risk variants to phenotypes. Primarily, this analysis consists of seven major steps as a pipeline. First, it inputs the population gene expression data with phenotypic information (Step 1) and constructs and clusters gene co-expression networks into gene modules (Step 2). Second, it performs enrichment analysis for modules (Step 3) and links genes and modules to various phenotypes from the population (Step 4). Third, it predicts the transcription factors (TFs) and regulatory elements (e.g., enhancers) that regulate genes and co-regulate modular genes as a gene regulatory network (Step 5). Also, it further finds disease risk variants (e.g., GWAS SNPs) that interrupt the binding sites of TFs from the network (Step 6). Finally, we output a full gene regulatory network linking disease variants to interrupted TFs and enhancers to regulated genes and modules to enriched functions and pathways to disease phenotypes (Step 7). The network thus provides a deeper understanding of gene regulatory mechanisms in diseases. As a demo, in this paper, we applied to AD population datasets from different brain regions. We predicted brain-specific gene regulatory networks for various AD phenotypes such as progression stages.

Materials and Methods

The pipeline of our integrative analysis for predicting gene regulatory mechanisms from AD risk variants to phenotypes

Our analysis can be summarized as a pipeline to predict gene regulatory networks from disease risk variants to phenotypes (**Fig. 1**). The network for specific phenotypes links disease risk variants (e.g., GWAS SNPs), non-coding regulatory elements, transcription factors (TFs) to

genes and genome functions, providing comprehensive mechanistic insights on gene regulation in disease phenotypes. Specifically, the pipeline includes the following steps. Here, our analysis is open-source available at https://github.com/daifengwanglab/ADSNPheno.

- Step 1: Input gene expression data at the population level. The input data includes gene expression data of individuals and clinical information on AD phenotypes such as Braak staging and progression.
- Step 2: Construct and cluster gene co-expression network. The pipeline constructs a gene co-expression network linking all possible gene pairs from the input data. The network edge weights are the Pearson correlations of the gene expression profiles across input samples. The gene co-expression network is further clustered into gene co-expression network modules. The genes in the same co-expression module are likely involved in similar functions and co-regulated by specific regulatory mechanisms.
- Step 3: Annotate modular functions by enrichment analyses. To annotate the functions of gene co-expression modules, we calculate enriched pathways and functions, including KEGG pathways, REACTOME pathways, and Gene Ontology (GO) terms of the genes in each gene co-expression module.
- Step 4: Associate AD phenotypes with genes and modules. We associate genes and modules with the phenotypes of input samples, revealing potential driver genes and modules for the phenotypes.
- Step 5: Predict gene regulatory networks for genes and modules. We apply multiple computational methods to predict the gene regulatory networks that link TFs, non-coding regulatory elements to genes and modules, providing regulatory mechanistic insights on AD genes and modules.
- Step 6: Link disease risk variants to the gene regulatory network. Our analysis further finds disease risk variants that interrupt the binding sites of TFs (TFBSs) in the gene regulatory networks for identifying functional variants to genes and modules to AD phenotypes.
- Step 7: Output a gene regulatory network linking disease variants to AD phenotypes. Ultimately, this network is the output that links AD genetic risk variants, non-coding regulatory elements, transcription factors (TFs) to genes and genome functions (via modules) for various phenotypes in the input data.

Population gene expression data and data processing in Alzheimer's disease

We applied this pipelined analysis to post-mortem AD population gene expression datasets for three major regions that relate to AD: Hippocampal CA1, Lateral Temporal Lobe (LTL), and Dorsolateral prefrontal cortex (DLPFC). Also, we processed the gene expression datasets as follows.

Hippocampal CA1: The microarray gene expression dataset (GSE1297)[25] was used. The dataset had total RNA expression values for 22,283 HG-U133 Affymetrix Human Genome U133 Plus 2.0 Microarray Identifier probes for 31 individual postmortem samples. These individual samples include 9 control samples (no AD), 7 initial stage samples, 8 moderate stage samples, and 7 severe stage samples. We used GEOquery[26], hgu133a.db[27], hgu133acdf[28], and Affy[29] R

packages to download the raw data and perform Robust Multichip Average (RMA) normalization[30] to account for background and technical variations among these samples. We mapped those microarray probes genes, averaging values that mapped to the same gene Entrez ID, removing probes that did not map to any known genes. Also, we transformed the resulting gene expression data by log2(x + 1) transformation and standardized that by R's scale() function. The finalized gene expression dataset in hippocampal CA1 has 13,073 unique genes for these 31 samples.

Lateral Temporal Lobe (LTL): The normalized bulk RNA-Seq dataset (GSE159699)[6] was used. This dataset had total RNA expression values for 27,130 different genes for 30 individual postmortem samples. This group of individual samples includes 8 young samples (ages 60 and below),10 old samples, and 12 old samples with advanced AD. After our data pre-processing steps, we had 25,292 genes, and we applied a log2(x+1) transformation to this gene expression data. The finalized gene expression dataset in Lateral Temporal Lobe has 25,292 unique genes for these 30 samples.

Dorsolateral prefrontal cortex (DLPFC): FPKM data from the ROSMAP Study, available on synapse.org (ID: syn3219045), was used[31]. Removing lowly expressed protein-coding genes (those with counts low for over 90% of the samples and 0 variance) shrunk the list of DLPFC genes down to 26,014 genes. In this dataset, there are 638 out of 640 individual RNA-Seq samples with mapped phenotype information. We also applied a log2(x+1) transformation to this gene expression data and then standardized it with the R function, scale(). The finalized gene expression dataset in DLPFC has 26,014 genes for these 638 samples.

Finally, there are 32,648 total unique protein-coding genes and 12,183 shared genes across these 3 brain regions. **Fig. S1** has a Venn-Diagram breakdown of gene counts across the different regions. 12,591 genes are found in the Hippocampus Ca1 and LTL, 12,444 genes are found in the Hippocampus Ca1 and DLPFC, and 18,882 genes are found in the LTL and DLPFC.

Regulatory elements and Chromatin interactions in the human brain regions

Epigenomic data has identified a variety of regulatory elements such as enhancers and promoters. Also, chromatin interaction data (e.g., Hi-C) have further revealed interacting enhancers and gene promoters. Thus, we integrated recent published epigenomic and chromatin interaction data for three brain regions to link enhancers to genes (via promoters). For Hippocampal Ca1, we obtained its enhancers and promoters from Brain Open Chromatin Atlas (BOCA)[32] and promoter-based interactions from GSE86189[33]. To identify promoters in LTL and DLPFC, we used R package, TxDb.Hsapiens.UCSC.hg19.knownGene[34], to retrieve promoter start and stop positions of genes, using a promoter length of 5,000 base pairs. Besides, we used the H3K27ac data from GSE130746[6] to find the enhancers in LTL. This dataset contains information on the target gene, distance that the H3K27ac mark is from the target gene's Transcription Start Site (TSS), and enhancer start and end positions. The enhancers in LTL that we used were at least 1,000 bases away from the TSS. Moreover, for

DLPFC, we used the enhancers and interacting enhancer-promoter pairs in DLPFC from PsychENCODE[35].

Gene co-expression network analysis

We applied WGCNA[36] to population gene expression data to construct and cluster gene co-expression networks into gene co-expression modules (minimum module size = 30 genes). Also, we further used a K-Means clustering[37] step to improve the module assignments, i.e., assigning unclustered genes (grey modules from WGCNA) into the modules. This step uses the modular eigengenes from WGCNA as initial centroids for the K-Means to build modules with a minimum size of 30 genes. In total, we obtained 29 gene co-expression modules for 13,073 genes for the hippocampal data, 56 modules for the 25,292 genes for the LTL, and 35 modules for 26,014 genes for the DLPFC. Besides, we calculated the module membership (MM) of genes for each module, which is the Pearson Correlation (r) of each gene with the modular eigengene. The MM values illustrate how similar the genes in the dataset are to a given module. Genes can have statistically significant MM values (p-value < 0.05) for multiple modules.

Enrichment analyses of gene co-expression modules

Co-expressed genes in the same module are highly likely involved in similar functions and pathways. The enrichment analysis has thus been widely used to identify such functions and pathways in a gene module. Enrichment p-values were adjusted using the Benjamini-Hochberg (B-H) correction. Given a group of genes (e.g., from a module) for each brain region, we used multiple tools and hundreds of data sources for enrichment analyses (please see **Table S1**). We used the highest enrichment -log10(adjusted p-value) scores from any source for each gene module and respective enrichment in a brain region. Then, for each enrichment for a phenotype in a region, we averaged the non-zero enrichment values for the gene modules that are statistically significantly positively correlated for that phenotype.

Association of genes and modules with AD phenotypes

We further associated genes and modules with these key AD developmental phenotypes: AD Stages and Progression (Moderate Stage, Severe Stage, and AD Progression), Healthy/Resilient (Control Stage or other resilient individuals with better cognitive abilities despite AD pathology), APOE genotype (APOE E4/E4 is a huge AD risk factor[38]), Braak Staging, neuritic plaque accumulation (measured by CERAD Score), and cognitive impairment level (based on the MMSE Score). We associated the gene co-expression modules with all possible AD phenotypes from the input data, by computing the Pearson Correlations of each modular eigengene to each phenotype. The eigengenes of modules by WGCNA are the first principal components of modular gene expression. A modular eigengene is a vector with its elements representing the expression levels of input samples and represents the most likely gene expression patterns of modular genes. Second, based on the modular eigengenes, we used the functions of moduleTraitCor() and moduleTraitPvalue() in WGCNA to find significantly associated phenotypes to the modules. Statistically significant module-phenotype associations for analysis have a p-value less than 0.05 and a positive correlation. Also, we used the gene

co-expression networks to examine the relationship between the genes and AD phenotypes and identify potential driver (hub) genes for the modules (based on the degree of connectivity for each gene in each module).

Prediction of gene regulatory networks

Gene Regulatory Networks (GRNs), a key molecular mechanism, fundamentally control gene expression. Also, co-expressed genes are likely co-regulated by similar gene regulatory networks. Thus, our analysis integrates multiple methods to predict the GRNs from gene expression data and co-expression modules. This study predicted gene regulatory networks in three brain regions that link transcription factors (TFs), Regulatory Elements (REs), and target genes/modules. First, we identified REs including enhancers and promoters that potentially interact using recent chromatin interaction data (Hi-C) and the scGRNom pipeline[39]. Second, we inferred the transcription factor binding sites (TFBSs) based on consensus binding site sequences on the interacting enhancers and promoters by TFBSTools[40] and motifmatchr[41]. This step generates a reference network linking TFs to regulatory elements (by TFBSs) to genes (by interactions). Third, using gene expression data for a given brain region, we predicted all possible TF-target gene (TG) pairs (or TF-modules) that have strong expression relationships by several widely used tools and databases: RTN[42], TReNA Ensemble Solver[43], Genie3[44], (and TF-gene-module pairs by RTN) as below. Finally, this step maps these TF-TG pairs to the reference network. It outputs a full gene regulatory network (GRN) for the region that links TFs, non-coding regulatory elements to target genes and modules.

We combined a recent list of TFs[45] with JASPAR's list[46] to generate a final list of candidate TFs for inferring TF-TG pairs with strong expression relationships. We used this final TF list to find the candidate TFs for each brain region (based on the respective gene expression data). Also, TReNA Ensemble Solver with the default parameters (geneCutoff of 0.1 and ensemble of these solvers: LassoSolver, RidgeSolver, RandomForestSolver, LassoPVSolver, PearsonSolver, and SpearmanSolver) was used to construct the transcriptional regulatory network that link TFs to target genes (TGs). Besides, we used GENIE3 to predict additional GRNs via Random Forest regression, predicting each gene's expression pattern from the expression patterns of all TFs (TF-TG pairs with weights greater than 0.0025 were retained). In addition, we used RTN to predict TFs to TGs by calculating the Mutual Information between the TFs and all genes. In particular, the permutation analysis with 1,000 permutations was applied bootstrapping and the ARACNe algorithm[47] was used to select most meaningful network edges. Finally, the TF-TG pairs found in at least 2 of the above 3 sources were combined to map to the reference network. For the DLPFC, we instead used the published PsychENCODE GRN (Elastic Net regression weight of 0.1 as a cutoff) filtered for genes found in the DLPFC gene expression data[48].

In addition to predicting TFs for individual genes, we also inferred TFs significantly co-regulating genes in a module in the Hippocampus and LTL. In particular, we performed the Master Regulatory Analysis (MRA) on the RTN-inferred network by RTN package[42]. For each gene module, MRA performed enrichment analysis using the inferred GRN, the phenotype (Module Membership correlation of all genes to that module), and hits (genes assigned to that module).

It applied the hypergeometric test for overlaps between TFs and the genes (using gene expression data) and found the statistically significant TFs for each module.

Linking GWAS SNPs for AD to gene regulatory elements

GWAS studies have identified a wide variety of genetic risk variants associated with diseases. However, most disease risk variants lie on non-coding regions, hindering finding disease genes and understanding downstream disease functions. To this end, we used gene regulatory networks as described above to link AD SNPs to the regulatory elements in the networks via interrupted TFBSs. In particular, we looked at 97,058 GWAS SNPs significantly associated with AD (i.e., AD risk SNPs with $p<0.005$)[15]. We overlapped those AD risk SNPs with the regulatory elements such as enhancers and promoters in the gene regulatory networks from Step 5. Then, we identified the variants that interrupt the TFBSs on the regulatory elements by motifbreakR[49] (using ENCODE-motif, FactorBook, HOCOMOCO, and HOMER data sources and default methodology, with a threshold of 0.001), and further linked them to the genes from the regulatory elements with interrupted TFBSs. An extension of 10 kilobase pairs was added to the start and end positions of enhancers and an extension of 2 kilobase pairs was added to the start and end positions of promoters. We found that 83,842 SNPs out of 97,058 AD GWAS SNPs interrupt the binding sites of 787 TFs.

Identification of AD-Covid genes and regulatory networks

We compared the KEGG pathways[50] for AD (hsa05010) and Covid-19 (hsa05171). We found several AD-Covid common mechanisms: Nuclear Factor Kappa B (NFkB), Inhibitor of Nuclear Factor Kappa B Kinase (IKK), c-Jun N-terminal Kinase (JNK), Interleukin-6 (IL-6), Phosphoinositide 3-Kinase (PI3K), Tumor Necrosis Factor (TNF) alpha, and TNF Receptor. This implies potential mechanistic interplays between AD and Covid-19. Further, we found 22 genes involved in those AD-Covid common mechanisms that correlate highly with AD phenotypes in different brain regions. Pathview[51] was used to visualize the correlations of those genes and AD phenotypes. Also, for each brain region, we found the subnetworks of its gene regulatory network in AD that have TFs regulating those AD-Covid common genes as the region's AD-Covid regulatory network.

Gene expression analysis and machine learning prediction for Covid-19 severity from AD-Covid regulatory networks

To gauge the clinical predictive performance of our AD-Covid genes and networks in terms of predicting Covid-19 severities, we looked at a recent population RNA-seq gene expression data in blood of Covid-19 samples (GSE157103)[52] to check whether any genes from our AD-Covid regulatory networks can predict Covid-19 severities such as being in the Intensive Care Unit (ICU) or not. To this end, we first median normalized this gene expression data (19,472 genes) and then applied differentially expression analysis by DESeq2[53] between 50 ICU and 50 non-ICU Covid patients. The Volcano plot was used to highlight differentially expressed AD-Covid genes (adjust $p<0.05$). A smoothing factor of 0.01 was added to the numerator and denominator when computing the empirical log2(fold change).

In addition to differentially expression analysis which aims to find individual associated genes, we also performed machine learning analysis for the genes from our AD-Covid regulatory networks to see if they together can predict severe Covid-19 or not.

In addition to our AD-Covid genes (from each region and combined), we also compared the machine learning prediction performance from other published Covid-19 genes. Since we are predicting Covid-19 severity, we compared predictive models from our lists with the respective performance of the benchmark list. A recent study[54] using U.K. Biobank GWAS data and Covid-19 mortality information discovered 8 Covid-19 susceptibility genes associated with an extremely high risk of Covid-19 mortality: DNAH7, CLUAP1, DES, SPEG, STXBP5, PCDH15, TOMM7, and WSB1. Another study[55] identified seven other risk genes (OAS1, OAS2 , and OAS3, TYK2, DPP9, IFNAR2, CCR2) associated with severe and life-threatening Covid-19 outcomes, including inflammatory organ damage. Numerous studies such as [56] have implicated ACE2 and TMPRSS2 as key genetic factors whose polymorphisms could be risk factors linked to greater Covid-19 susceptibility. We thus included all 17 genes, from these published studies, into a list of Covid-19 published genes, to use as our benchmark list.

The Python package, Scikit-Learn[57], was used for our machine learning analysis. The classification accuracy of each gene was calculated by 4-Fold Cross Validation (CV). For the 17 published Covid-19 susceptibility genes, we performed Recursive Feature Elimination (RFE) CV based on a Support Vector Machine (SVM) classification model (with linear kernel, outputting predicted probabilities)[58]; this calculated the classification accuracy of each gene and the optimal number of published genes to use. We fixed all models to incorporate only this optimal number of genes. Then, to build a model for each of our input gene lists, we performed RFE based on linear SVM to then select the optimal number of predictive genes from the list for classifying ICU vs. non-ICU Covid-19 patients with high accuracy (i.e., feature selection). We then trained an SVM classification model again with those select predictive genes and reported the accuracy and AUROC values of the model using 4-Fold CV. Besides Covid severity, we also calculated the correlations of gene expression with Covid and non-Covid for the genes from the AD KEGG pathway for three brain regions.

In addition, we also used Decision Curve Analysis (DCA)[59] to evaluate and compare the machine learning models of those brain-region AD-Covid genes and benchmark genes for predicting Covid severity. DCA has been widely used for making medical decisions to individuals, improving upon traditional comparison metrics (e.g., AUROC) for predictive models and other approaches that require additional information to address clinical consequences[59]. Particularly, given a model and a threshold probability pT, the patients will be sent to ICU if their percentage risks for Covid severity (i.e., ICU) from the model are greater than or equal to pT. Based on this, the true positive (TP) count is the number of Covid-19 severe individuals sent to the ICU, and the False Positive (FP) is the number of Covid-19 non-ICU individuals sent to the ICU. Thus, pT inherently represents subjective clinician preferences for FPs versus False Negatives (FNs: wrongly predicting a severe Covid-19 patient would not be severe and not sending them to the ICU). Based on TP, FP and pT, the DCA then calculates Net Benefit = TP/N – ((FP/N)*pT/(1-pT)), where N is the total number of patients (N=100 here). Thus, the Net

Benefit represents the benefit of true positive ratio (TP/N) from false positive ratio (FP/N) weighted by odds of pT (i.e., pT/(1-pT)). DCA provides a simple, personalized risk-tolerance based approach of using pT to weight the FN and FP mistakes: lower thresholds represent a fear of FNs over FPs, and vice-versa. For instance, for a clinician who sends a Covid-19 positive individual with predicted severity of at least pT = 20%, the utility of treating a Covid-19 severe individual is 4 times greater than the harm of needlessly sending a non-severe Covid-19 patient to the ICU. Also, we compared our predictive models with 2 extremes: Treat All (predict 1 for all Covid-19 positive patients and send all to the ICU regardless of severity) and Treat None (predict 0 for all positive patients and send none to the ICU). Practically, a clinician ought to opt for the predictive model or extreme intervention strategy with the highest Net Benefit based on that clinician's preferred pT; thus, two clinicians (who may have their own, different pT values) may obtain different optimal results. Thus, DCA can be used to evaluate the clinical usability of a Covid-19 severity prediction model based on its Net Benefit across clinically reasonable pT values. Finally, we performed the DCA using the codes provided by Memorial Sloan Kettering Cancer Center[60].

Results

Gene co-expression network analysis reveals gene expression dynamics for AD phenotypes across multiple brain regions

We first applied our analysis to population gene expression datasets of three major brains relating to AD: Hippocampal CA1, Lateral Temporal Lobe (LTL), and Dorsolateral prefrontal cortex (DLPFC) (**Methods and Materials**). We identified several gene co-expression modules showing specific gene expression dynamic changes for various AD phenotypes as below. These expression dynamics also imply potential underlying gene regulatory mechanisms in the phenotypes. In particular, given a brain region, we constructed and clustered a gene co-expression network for the region to a set of gene co-expression modules. In a gene co-expression network for a region, the nodes are genes and each edge represents that the two respective genes have correlated gene expression profiles during AD progression (i.e., co-expression). Also, there are likely groups of co-expressed genes within the network that form densely-connected sub-networks, also known as gene co-expression modules. Genes within a module share similar gene expression dynamics in the region for the observed AD phenotypes. We also used modular eigengenes (MEs) to represent such expression dynamics for a gene module, using the first principal components of modular gene expression matrices. The information for all modules with associated phenotypes is available in **Supplement files 1, 2** and **3** for Hippocampal CA1, Lateral Temporal Lobe and Dorsolateral prefrontal cortex, respectively.

Hippocampal CA1. We identified 29 gene co-expression modules in the Hippocampal Ca1 region (min. module size = 30 genes). We found that 21 out of 30 modules were significantly positively associated with at least one AD phenotype: AD progression, Braak Stage progression, aging, accumulation of NFTs, MMSE score cognitive impairment, AD, and being resilient. Also, their eigengenes show specific expression dynamics (**Fig. 2A** for 7 select

modules, **Fig. S2** for all modules). For instance, the pink and lightyellow modules tend to have higher expression values for individuals who do not have AD (Control Stage) and are clustered together. On the other hand, the greenyellow, yellow, magenta, and tan modules have higher expression levels for AD individuals and cluster together. In between those groups of modules is the midnightblue module, which has relatively high expression for some Control Stage and some AD individuals.

Next, using these expression dynamic patterns, we further linked these gene modules to those key AD phenotypes (**Fig. 2B**) using their significant correlations (**Fig. S3** and **Supp. file 1** for all modules). For instance, the greenyellow module has significantly highly correlations with AD, AD progression (e.g., moderate and severe stages), Braak 6 Stage and cognitive impairment. The tan module has the highest correlation (r = 0.68) with the Severe Stage, along with other AD-related phenotypes. The midnightblue module is more significant (r = 0.02) for the Braak 4 Stage, where affected individuals typically exhibit mild symptoms of dementia. The lightyellow module is significant for cognitive resilience (r = 0.5), which is the ability of individuals to exhibit stronger cognitive functioning despite AD pathology. Moreover, the lightyellow module has a strong positive correlation with better MMSE performance (r = 0.58) than the pink module (r = 0.44). Instead, the pink module is more significantly correlated with the Braak 3 Stage (r = 0.45).

Lateral Temporal Lobe. We identified 56 gene co-expression modules and found that 28 modules are significantly positively correlated with AD phenotypes of interest. We highlighted five modular eigengenes in **Fig 2C** showing specific expression dynamics for AD phenotypes in LTL (**Fig. S4** for all modules). In total, we found that 12 gene modules are positively associated with AD Progression, 1 with Aging (mediumpurple3), 1 with Gender (lightgreen), 12 with the Control Stage, and 2 with the Initial Stage (associated with Braak 1 and 2 Stages). As shown in **Fig. 2C**, the sienna3 module has higher expression values for both old and young individuals in the Control Stage. The orange, magenta, and yellow modules are clustered together and have higher expression values for AD samples. The midnightblue module is clustered between both groups and tends to be associated with higher Braak stages. As shown in **Fig. 2D**, the sienna3 module also has a statistically significant positive correlation with the Control Stage (r = 0.63) and asymptomatic based on the Braak stage (r = 0.55). The midnightblue module is associated with aging, average Braak stage, clinical Braak stage, and AD. The yellow, orange, and magenta modules are associated with aging, AD and Braak progression phenotypes, and neuritic plaque accumulation (based on the CERAD score); the orange module has a very strong correlation with dementia Braak stages (r = 0.72) and AD (r = 0.72). More module-phenotype associations are available in **Fig. S5** and **Supplementary File 2**.

Dorsolateral prefrontal cortex. We found 35 gene co-expression modules for DLPFC. **Figs. 2E** and **2F** highlighted 6 of those gene modules: darkolivegreen, yellow, red, royalblue, tan, and green (**Figs. S5-6** for all modules). Those modules are significantly associated with various AD phenotypes such as progression, MMSE, APOE genotype and neuritic plaques. Larger sample size for the DLPFC, which is over 20 times larger than that for the Hippocampal or LTL regions, likely attributes to the relatively lower correlation coefficients between modules and AD

phenotypes in DLPFC However, we still see significantly correlated modules with various AD phenotypes (Fig. 2F, $p < 0.05$). For example, the tan gene module is associated with the worst APOE genotype, APOE E4/E4 ($r = 0.082$), AD diagnosis age ($r = 0.12$) and points lost on the last MMSE ($r = 0.09$). The royalblue and green modules are statistically significantly positively correlated with the Severe Stage based on the last MMSE score, with correlations of $r = 0.22$ and $r = 0.2$, respectively. In terms of better outcomes, the darkolivegreen module is significant for the Control Stage ($r = 0.11$), better performance on the last MMSE ($r = 0.17$), and cognitive resilience ($r = 0.12$). Furthermore, the red module is also significant for better performance on the last MMSE ($r = 0.15$) and has a similar correlation as the darkolivegreen module for cognitive resilience.

Therefore, those AD-phenotype-associated gene co-expression modules uncover the specific gene expression dynamic patterns across phenotypes and suggest that those co-expressed genes in the same modules are likely involved in similar functions and pathways for the phenotypes. To understand this, we further performed the enrichments analysis of the modules as follows.

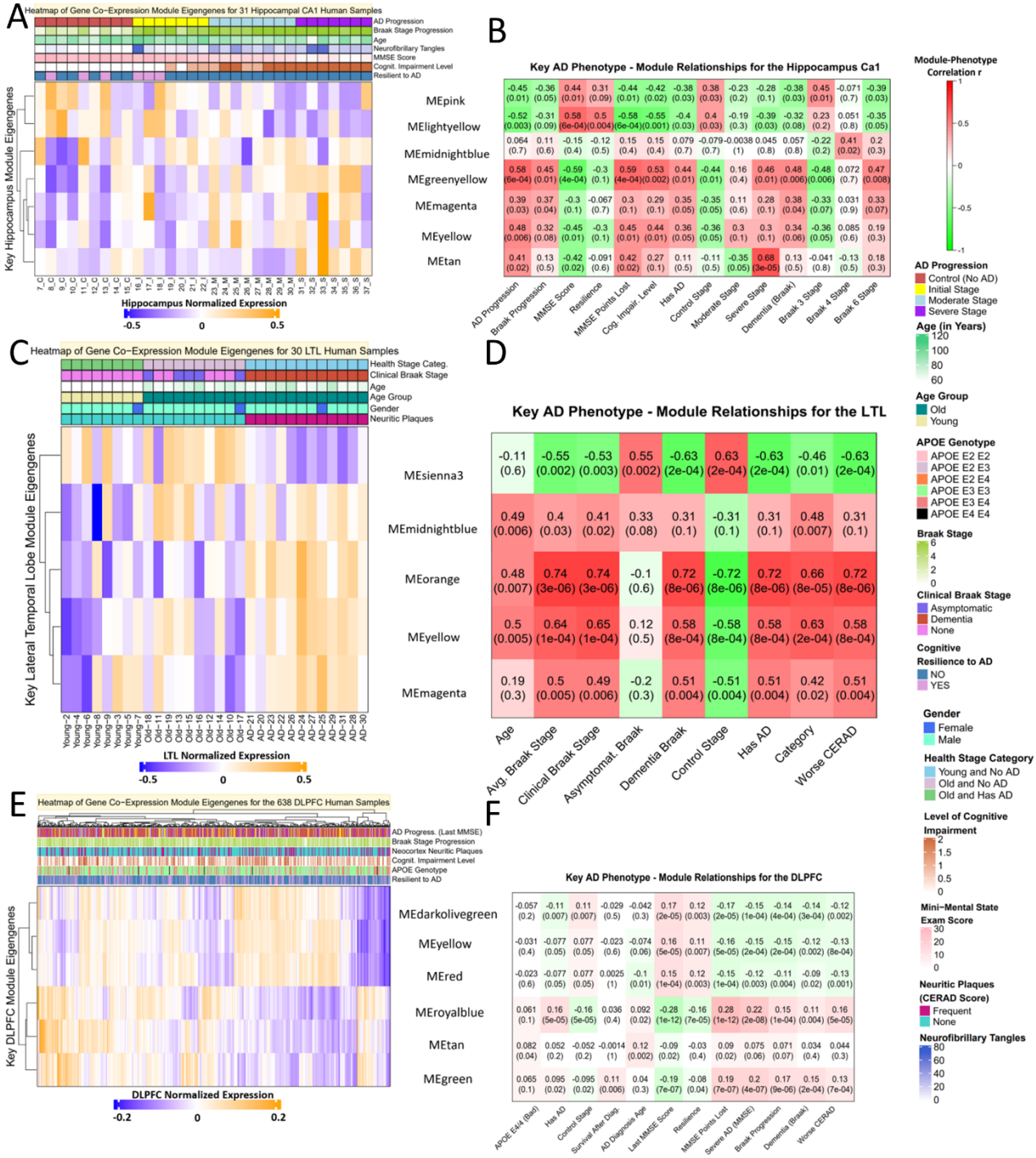


Fig. 2 – Gene co-expression modules significantly associated with AD phenotypes show specific expression dynamic patterns across phenotypes. Top heatmaps show the eigengenes of select gene co-expression modules in Hippocampal CA1 region (A), LTL (C) and DLPFC (E). Rows: modules. Columns: individual samples. Red: high expression level. Blue: low expression level. Bottom heatmaps show the correlation coefficients and p-values between select modules and AD phenotypes in Hippocampal CA1 region (B), LTL (D) and DLPFC (F).

Row: modules. Columns: AD phenotypes. Red: highly positively correlation. Green: highly negatively correlation.

Eigengenes and enrichments of co-expression modules reveal hub genes, gene functions and pathways in AD phenotypes

Gene module enrichment analysis allows us to better understand the biological functions, structures, diseases, and other observed biological phenomena associated with AD phenotypes. Through enrichment analysis (Methods), we found enriched functions and pathways of AD modules and then linked them to various AD phenotypes associated with the modules (**Fig. 3** and **Supplementary document**). Overall, healthier phenotypes are the Control Stage (No AD), cognitive resilience, and protective APOE E2/E2 genotype.

Hippocampal CA1 (**Fig. 3A**, **Supplementary file 1**): The modules for non-AD phenotypes (e.g., control and resilience) are enriched with synaptic plasticity and dendrite development, norepinephrine neurotransmitter release cycle, and calcium signaling pathway, which can be typically dysregulated in AD[61]. This also suggests that resilient individuals may be protected from microsatellite instability and amyloid accumulation that may even occur naturally[62]. Further, our analysis supports recent hypotheses on dysregulated immune systems in AD. We found that aging, NFTs, and AD developmental phenotypes are associated with abnormal innate immunity and interleukin-6 secretion. Our cognitive impairment, Braak progression, and AD modules also have enrichments for: viral genes and Covid-19 spike glycoprotein (that trigger an immune response), activated TAK1 mediating p38 MAPK activation (linked to tau phosphorylation, neurotoxicity, neuroinflammation, synaptic dysfunction, and worse AD[63]), and NFKB pathway (impaired and over-expressed during AD, leading to neuroinflammation, microgliosis[64], and suppression of Wnt Signaling[65]). Moreover, we found an association between Severe AD and immunologic memory, antigen-antibody interactions, and regulation of Interferon-Alpha Signaling. Interferon response to immunogenic amyloid may activate microglia, initiate neuroinflammation, and lead to synaptic loss[66]. Finally, many AD phenotypes are associated with Death Receptor Signaling, positive regulation of gliogenesis, Constitutive Signaling by aberrant PI3K in Cancer, and positive regulation of JNK cascade (activated in AD brains and involved in tau phosphorylation and neuronal death)[70]. Xenobiotic Metabolic Processes are specific to Severe AD modules only, and studies[67] have found links between dementia progression and various metabolic pathways.

LTL (**Fig. 3B, Supplementary file 2**): First, control modules are indeed enriched with several pathways that are typically present in healthy conditions, such as Wnt signaling, Actin organization and transmission across chemical synapses. Dysregulation of those pathways has been reported to lead to AD and neurodegeneration, e.g., Wnt signaling to inhibit amyloid-beta production and tau protein hyperphosphorylation in AD progression[68]. Second, in the LTL modules for AD and large neuritic plaques, Frontotemporal Dementia (FTD) and Loss of Nlp from Mitotic centrosomes are enriched, and the latter may lead to reduced microtubule stability, abnormal cellular morphology, and functions in AD[69]. Also, we found cell-type specific enrichments in the AD progression related phenotypes, e.g., astrocyte projection for clinical braak stage and asymptomatic. Astrocytes are increasingly activated near amyloid plaques in

LOAD, producing pro-inflammatory cytokines and reactive oxygen species[70]. Additional previously identified AD-related functions and pathways include postsynaptic differentiation, stress-activated signaling, TRAF-mediated NF-kB activation, prion pathway and epigenetic modifications. The feedback look of the prion pathway is likely disrupted during AD leading to Aβ accumulation[71]. The epigenetic modifications (e.g., Histone A4 acetylation) can be dysregulated to affect gene expression of long-term potentiation and memory formation in AD[72].

DLPFC (**Fig. 3C, Supplementary file 3**): The resident macrophage brain cells, microglia, tend to exclusively express most AD risk genes like APOE[73]. Our APOE E2 modules were shielded from neurotoxins and associated with mitochondrial inheritance ($p<1e-16$), while our E4 modules are strongly enriched for cellular response to Aβ and cognitive dysfunction. Having both high risk E4 alleles (instead of only 1 E4 allele) is not associated with mitochondrial depolarization. Still, it is more strongly associated with Serum Amyloid A Protein ($p<1e-28$ vs. $p<1e-16$), monocyte chemoattractant proteins, and neuroimmunomodulation.

We found several strong promising associations (some with $p<e-58$) for APOE4-related and cognitive impairment modules, supporting the crucial role of the immune system and reactive microglia in AD onset and pathogenesis. These include astrocyte activation (boosting production of proinflammatory cytokines and phagocytic capabilities[74]), abnormal circulating cytokine level and innate immunity, synapse pruning (excess in Schizophrenia[75]), neuroinflammatory responses, autoimmune diseases (Wegener's Granulomatosis), abnormal macrophage physiology, Microglia Pathogen Phagocytosis Pathway, and microglial cell activation (key to ALS and Multiple Sclerosis pathology[76]). During early synaptic decline in AD, microglia may change shape, functions, and pathways, express more receptors and inflammatory molecules (cytokines, chemokines), become more phagocytic and activated, and go awry, leading to neuroinflammation and cell death[73]. Except for neurotransmitter reuptake, our other AD phenotype (ex. severe stage, Braak progression, plaques, cognitive impairment, survival post diagnosis) modules share many biological associations with our APOE4 modules, like: Immune System diseases, Covid-19, inflammation mediated by chemokines and cytokines, natural killer mediated immunity. Besides, the modules for non-AD phenotypes like control and resilience are also enriched with pathways that may be dysfunctional in AD such as the Electron Transport Chain[77], neuron recognition, calcium ion regulated exocytosis, mitotic spindle checkpoint, and synaptic plasticity.

Comparison Across Brain Regions (**Fig. 3D**): We also found that many enriched pathways for AD phenotypes are shared by different brain regions. In particular, the datasets of three brain regions share major phenotypes (control, AD, and clinical Braak stage). The modules from three regions for those shared phenotypes are all enriched with physiological adaptation. Also, many of our modules from different brain regions are enriched with immunological functions that have been recently studied in AD[78]. For instance, the Hippocampal and DLPFC modules for control and AD share neuroimmunomodulation. The MAPK cascade (associated with control modules in all 3 regions) is associated with stress-activation in the AD and Braak modules in the hippocampus and LTL, which has been reported to be dysregulated in AD[79]. These AD and Braak modules in multiple brain regions are also enriched with the Ki-1 antigen, a tumor marker

of activated immune cells regulating NF-kB and apoptosis[80], proto-oncogenes, dendritic antigen-presenting cells, diabetes complications, focal adhesion (plaques), and angiogenesis VEGFA-VEGFR2 Signaling. The VEGFA-VEGFR2 pathway promotes neural cell survival, migration, and proliferation, and has altered levels in AD that may impact the role of microglia[81]. Moreover, the AD and Braak modules across regions have common brain enrichments such as Blood-Brain Barrier, virus attachment, Complement Activation (innate immune-mediated defense[82], altered in AD[74]), and Oligodendrocyte differentiation (potentially impacted by Aβ accumulation and associated with neurodegeneration[83]). Finally, the control modules across regions are enriched for higher cellular energy levels like ATP synthesis, Mitochondrial proteins. In particular, the DLPFC and Hippocampal control modules are enriched with Calcium Signaling, LTP, Presynaptic Terminals, regulation of dendrite development, and positive regulation of neurogenesis.

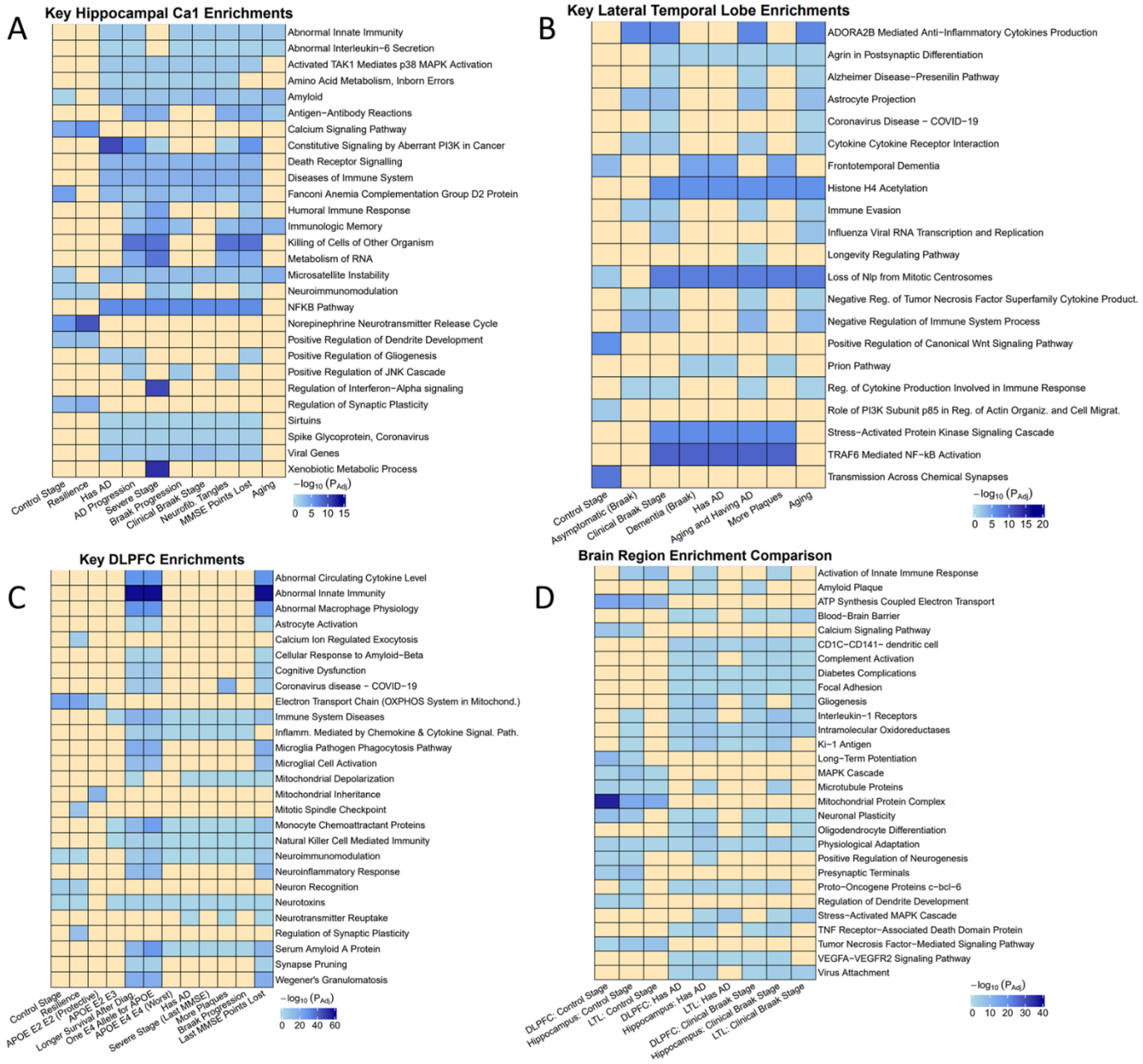
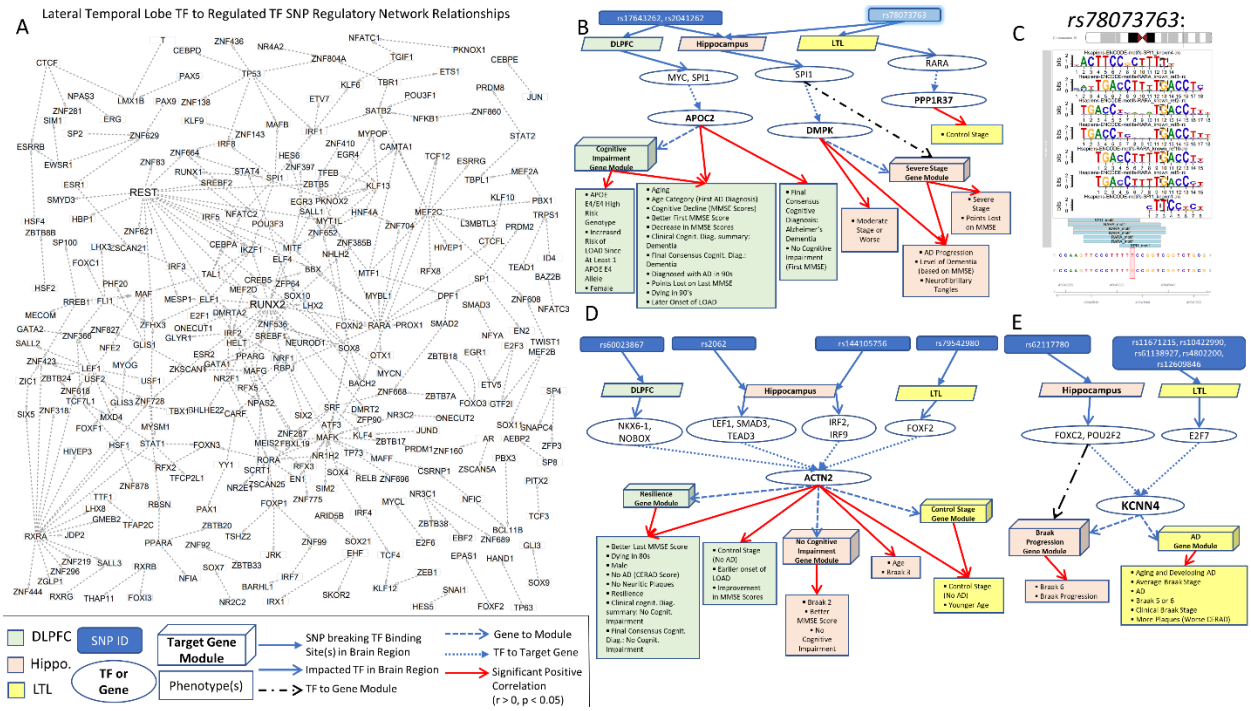


Fig. 3 – Select enriched functions and pathways of gene co-expression modules for various AD phenotypes. (**A**) Hippocampus; (**B**) Lateral Temporal Lobe (LTL); (**C**) DLPFC; (**D**) Across three regions. Rows: select enriched terms (Methods). Columns: AD phenotypes. The heatmap colors correspond to –log10(adjust p-value).

Prediction of brain-region gene regulatory networks for AD phenotypes

To understand underlying molecular mechanisms regulating gene expression associated with various AD phenotypes, we predicted the gene regulatory networks (GRNs) for genes and gene modules of brain regions, especially using multi-omics data (Methods). The brain-region GRNs link transcription factors (TFs) and regulatory elements (REs, e.g., enhancers or promoters) to target genes (TGs) and co-expressed genes (e.g., from same modules). The regulatory network edges can be activation or repression. Moreover, these GRNs can be further linked to the AD phenotypes significantly associated with TGs and modules. As described in Methods, we applied multiple widely-used approaches and public databases to predict the networks and finally used shared predictions across different approaches as highly confident GRNs. In terms of candidate TFs, we found: 1,043 in the Hippocampus, 1,580 in the LTL, and 1,588 in the DLPFC, which we input into RTN, Genie3, and Trena. As shown in **Table S2**, we obtained 6,823,631 TF-RE-TG network edges of three brain regions' GRNs, corresponding to 973,025 unique TF-TG pairs, 20,601 TGs and 709 TFs. In particular, the hippocampal GRN has 2,810,102 TF-RE-TG edges, including 169,292 unique TF-TG pairs, 11,972 TGs and 351 TFs. The GRN of Lateral temporal lobe has 161,404 TF-RE-TG edges, including 65,321 unique TF-TG pairs, 13,791 TGs and 402 TFs. The GRN of DLPFC has 3,852,125 TF-RE-TG edges, including 752,169 unique TF-TG pairs, 13,511 TGs and 670 TFs. Detailed edge lists of hippocampus and LTL GRNs are provided in **Supplementary files 4-5**.



640
641 **Fig. 4 – Select gene regulatory networks (GRN) linking AD risk variants (GWAS SNPs) to**
642 **AD phenotypes**. (**A**) Subnetwork of LTL GRN among TFs (i.e., the target genes (TGs) are TFs
643 too). Nodes are TF genes. The edges connect TFs to their TGs. Besides, TFs have the binding
644 sites interrupted by AD SNPs on the regulatory elements to TGs. (**B**) Example of 3 AD SNPs
645 that interrupt binding sites of TFs in different brain regions. The AD phenotypes and gene
646 modules positively correlated with APOC2, DMPK, and PPP1R37 expression are shown. (**C**)
647 SNP rs78073763 interrupts multiple possible binding sites of SPI1 in Hippocampus and RARA
648 in LTL. Additional regulatory links from AD SNPs to interrupted TFs to TGs along with
649 associated phenotypes and modules are shown in (**D**) regulation of ACTN2 in all three regions,
650 (**E**) FOXC2, POU2F2, and E2F7 to KCNN4 in hippocampus and LTL.

651 ## Identification of disease risk variants for AD phenotypes via integration of
652 ## GWAS and gene regulatory networks

653 Over 90% of disease risk variants (e.g., GWAS SNPs) are in the non-coding regions[18]. For
654 instance, we found that GWAS SNPs for AD are enriched in the regulatory elements of our
655 GRNs as above. Thus, it is crucial to further understand how those disease risk variants affect
656 gene regulatory mechanisms that eventually impact AD phenotypes such as progression. To
657 this end, we linked AD GWAS SNPs to our GRNs to see how those SNPs interrupt the binding
658 sites of TFs on the enhancers or promoters that regulate target genes and modules (Methods).
659 These SNPs can also be linked to various AD phenotypes of corresponding genes and modules
660 for different brain regions, i.e., “brain-region risk variants for AD phenotypes”. Specifically, we
661 found that 39,832 unique AD SNPs disrupted TFBSs on the regulatory elements of three brain-
662 region GRNs (35,940 for Hippocampus, 7,119 for LTL, 2,359 for DLPFC, **Fig. S8**). Across three
663 regions, there are 543 unique TFs whose binding sites were interrupted, regulating 11,596
664 genes (**Table S3**).

For instance, a subnetwork of the LTL GRN between TFs is shown in **Fig. 4A**, i.e., the target genes are also TFs. These subnetwork TFs also have interrupted TFBSs on the regulatory elements to their target genes. We found that several TFs are the hub genes of the subnetwork like RUNX2 (11 TFs experience difficulty in binding and regulating RUNX2) and neurogenesis TF REST (experiences difficulty in regulating 16 TFs in LTL). In fact, REST is induced by Wnt signaling, represses genes (like PLCG2) that promote cell death or AD pathology, protects neurons from Aβ-protein toxicity[84]. We found that during AD, PLCG2 is overexpressed in the LTL, which may be partly explained by REST's inability to bind to chromatin and repress its target genes[85]; this may lead to changes in autoinflammation, immune disorders, and changes in immune cell functioning[86]; moreover, we found that REST significantly regulates the turquoise gene module in the LTL (**Figs. S9-10**). The subnetwork of the DLPFC GRN between TFs (**Fig. S11**) has the hub genes: CREB3L1 (26 TFs are unable to properly regulate CREB3L1) and PAX5 (has difficulty in regulating 11 TFs). Lastly, the Hippocampal GRN SNP subnetwork between TFs has hubs such as ZNF226 (21 TFs experience difficulty regulating ZNF226) and GATA2 (unable to regulate 65 TFs in AD Hippocampus). In the Hippocampus, ZNF226 significantly regulates 3 modules (2 which are associated with the Control Stage) and GATA2 significantly regulates 4 modules (1 Control Stage module and 2 modules associated with worsening AD phenotypes) (**Figs. S12-13**).

Furthermore, we found several regulatory networks, which provide more insights on the possible association of various non-coding SNPs with AD phenotypes. **Fig. S14** provides a detailed explanation of how to interpret such networks. In **Fig. 4B**, we examine the varying effect of a given SNP on gene regulation across brain regions, and the impact on 3 target genes: APOC2, DMPK, and PP1R37. For instance, SNP rs78073763 (**Fig. 4C**) changes the DNA base from a T to a G (at chr19:45649838) and breaks the binding of RARA on the enhancer of Control Stage gene PPP1R37 in the LTL and SPI1 binding to the DMPK enhancer in the Hippocampus. A recent study found that PPP1R37 expression is strongly associated with APOE expression and has extensive cross-tissue effects on AD and that DMPK expression in the hippocampus and putamen strongly impact AD[87]. We found that increased expression of DMPK is associated with worsening AD phenotypes (ex. Moderate Stage or worse, AD progression, more severe dementia, NFTs). Two extremely statistically significant AD SNPs (rs17643262 and rs2041262; $p < 2e\text{-}13$) that disrupt SPI1 regulation of DMPK in the Hippocampus also disrupt SPI1 regulation of APOC2 in the DLPFC; APOC2 is associated with a cognitive impairment gene module, Alzheimer's dementia, cognitive decline, and having at least 1 APOE E4 allele. SPI1 is a well-known master regulator in microglial cells, plays a key roles in regulating immune functions in AD[88], is strongly correlated with AD ($r = 0.355$) and AD Progression ($r = 0.375$), Braak progression ($r = 0.437$), Braak 6 ($r = 0.407$), and belongs to a Severe AD Stage gene module ($r = 0.41$). This suggests a low-level expression of SPI1 in hippocampus control samples, which potentially reduced microglial-mediated neuroinflammatory responses and delayed AD onset[89]. SPI1 regulation in the Hippocampus by 10 TFs (ex. RXRA, RARA, NFKB1) is disrupted by several SNPs (**Fig. S15**). The regulated genes by SPI1 are also upregulated in microglia, leading to microglia-mediated neurodegeneration in AD[89]; in fact, SPI1 significantly regulates DMPK and its Severe Stage gene module in the Hippocampus. Our results further underscore the role of the microglia and immune system in AD onset and progression and

neuroinflammation. In addition, MYC regulation of APOC2 in the DLPFC is also impacted. Abnormalities in MYC functioning may lead to dysregulation in cellular processes such as cell cycle activation and Wnt Signaling and re-entry mediated neuronal cell death in AD[90].

In **Fig. 4D**, we present an example of a gene, ACTN2 (cytoskeletal alpha-actin protein), whose regulation in all 3 brain regions, is impacted by various SNPs, and whose expression is associated with healthier outcomes. Across the 3 brain regions, we observe that ACTN2 expression is positively associated with the cognitive resilience DLPFC gene module, no cognitive impairment Hippocampus gene module, and Control Stage LTL gene module. Increased expression of ACTN2 may be associated with improvements in MMSE scores and is found at the neuronal synapse, and is a highly ranked gene in a previous study on echocardiographic traits, heart function, and AD[91]. Several SNPs impact ACTN2 expression in the DLPFC and Hippocampus, but we focused on SNPs shared by several TFs in this Figure. Here, rs60023867 disrupts regulation of ACTN2 in the DLPFC by both NKX6-1 and NOBOX. In the Hippocampus, rs2062 disrupts regulation of ACTN2 by 3 TFs (LEF1, SMAD3, and TEAD3) and rs144105756 impacts IRF2 and IRF9 regulation of ACTN2. Lastly, in the LTL, rs79542980 may lead to FOXF2 dysregulation of ACTN2.

Also, we found different SNPs in the Hippocampus and LTL that impact regulation of KCNN4, a key AD drug target that is overexpressed during AD (**Fig. 4F**). In **Fig. S16**, we visualize the impact of rs62117780 on FOXC2 and POU2F2 regulation of KCNN4 in the hippocampus and rs4802200 on E2F7 regulation of KCNN4 in the LTL. FOXC2 and POU2F2 also regulate KCNN4's Hippocampal Braak Progression module. KCNN4 belongs to an AD LTL module, so increased KCNN4 expression is associated with AD progression in both regions. Previous studies found that KCNN4 is primarily expressed in macrophages and microglia and regulates microglia activation by modulating Ca2+ influx signaling and membrane potential[85]. Thus, it has low expression in healthy neurons, and is associated with neuroinflammation and reactive gliosis during AD. Blocking KCNN4 likely curbs microglial neurotoxicity, leading to slower neuronal loss and better memory levels[92]. Therefore, this link uncovers how AD SNPs regulate KCNN4 expression in AD phenotypes.

Finally, we highlighted all possible SNPs that interrupt TFBSs in our brain region GRNs via Manhattan plots (**Figs. S17-21**): **S18** (hippocampus), **S19** (LTL) and **S20** (DLPFC). 268 SNPs were found in all 3 regions and are examined in **Figs. S21-24**. Regulatory links from AD SNPs to interrupted TFBSs and regulatory elements to target genes and modules is provided in **Supplementary files 6-8**. We provide additional examples and explanations of regulatory networks linking non-coding SNPs to AD phenotypes for the networks, which we visualize in **Figs. S25-28**.

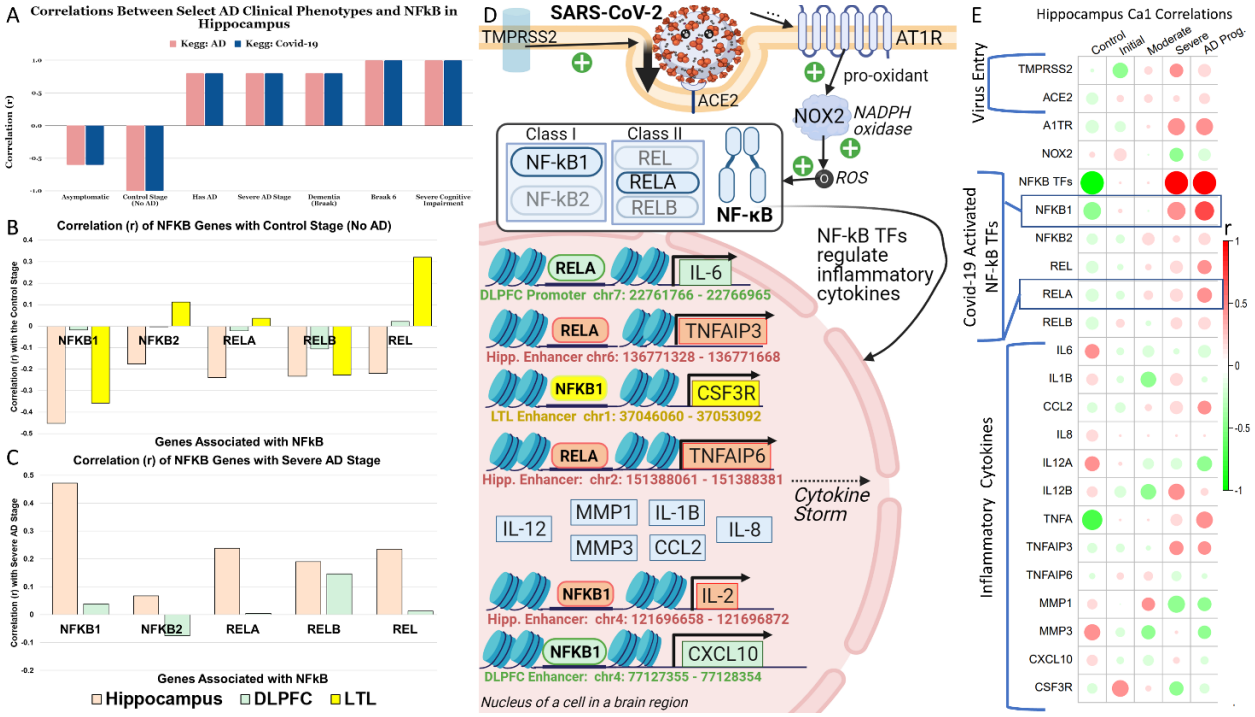


Fig. 5 – Gene regulatory networks and phenotypes for NFKB, a shared pathway of AD and Covid-19. (**A**) Pearson correlations of the NFKB pathway (KEGG:hsa05171) and AD pathway (KEGG: hsa05010) with AD phenotypes from the Pathview analysis of hippocampal expression data of pathway genes[51]. (**B**) Pearson correlations of NFKB TFs (NFKB1, NFKB2, RELA, RELB, and REL) with Control in three regions. (**C**) Pearson correlations of NFKB TFs with Severe stage in Hippocampus and DLPFC. (**D**) Covid-19 virus spike and gene regulation of NFKB TFs from our hippocampal, LTL, and DLPFC GRNs for pro-inflammatory cytokines linked with severe Covid-19 outcomes. Gray dashed arrows indicate regulation and black arrows indicate activation of cytokines by the TF. (**E**) Correlations between AD phenotypes and expression levels of genes from (**D**) in the hippocampus.

Gene regulatory networks and AD phenotypes associated with shared pathways between Covid-19 and AD

Recent Covid-19 virus has widely affected the elders with neurogenerative diseases including AD, suggesting potential links between Covid-19 and AD. We found that many significantly up/down-regulated genes in Covid-19 present in the AD KEGG pathway and positively/negatively correlated with AD in three regions. To understand molecular mechanisms across two diseases, we look at shared AD-Covid pathways such as the NFKB pathway involved in adverse effects and inflammation in both AD and Covid-19[93]. There are five TFs: NFKB1, NFKB2, REL, RELA, and RELB (proto-oncogene near APOE) involved in this pathway that regulate cellular processes including inflammation, cell growth, apoptosis[94]. Further, in both AD and Covid-19, Reactive Oxygen Species activate RELA and NFKB1 that then transcribe pro-inflammatory cytokines (typically secreted by macrophages), like: IL-6, IL-1B, and TNF, reducing LTP in AD, leading to an exaggerated and potentially lethal immune response in Covid-19 (e.g., tissue injury, Acute Respiratory Distress Syndrome (ARDS)[93]) (**Fig. S29**). We

found that in general, the gene expression levels of those NFKB TFs positively correlate with AD phenotypes such as severity but negatively with control in all three regions (**Fig. 5A** for hippocampus, **Fig. S30** for LTL, and **Fig. S31** for DLPFC). For instance, NFKB1 and RELB negatively correlate with controls in three regions, and so are NFKB2 and RELA in hippocampus and DLPFC (**Fig. 5B**). However, all five TFs positively correlate with AD severity in the hippocampus and two of them in DLPFC (**Fig. 5C**). Since the up-regulation of these NF-kB TFs is also linked to greater inflammatory responses in Covid-19 infected individuals[93], this result implies that the gene expression of NFKB TFs is a potential interplay between AD and Covid-19. Thus, we further investigated our gene regulatory networks involving NFKB TFs to understand possible regulatory mechanistic links across AD and Covid.

To this end, we looked at the impact of the NFKB pathway in SAR-CoV-2 to additional molecules (KEGG: hsa05171) (**Fig. 5D**). In particular, to enter the cell, the SARS-CoV-2 Spike protein is primed by TMPRSS2, binds to ACE2 (highly expressed in macrophages and the brain[95]), and interacts with A1TR (elevates viral entry and infection[96]). We also found that the expression levels of TMPRSS2, ACE2, and A1TR receptors are negatively correlated with controls but positively associated with late AD stages in hippocampus (**Fig. 5E**). Moreover, in our brain region GRNs, NFKB1 and RELA regulate genes of several cytokines associated with the severe Covid-19 Cytokine Storm. For instance, in the DLPFC, RELA regulates IL-6 via the promoter and particularly activates CXCL10 (biomarker whose altered levels are associated with immune dysfunction, tumor development, and disease severity[97]) by binding to the enhancer. In the LTL, NFKB1 binds to the enhancer of hematopoietic growth factor and cytokine, CSF3R, a key regulator of neutrophil cell development, proliferation, and differentiation[98]; in Severe Covid-19 patients, there are increased levels of neutrophils (immune cells involved in the first-line of defense against pathogens) and changes in their phenotype and functionality[99]. Also, our hippocampal GRN found that NFKB1 regulates IL-2 (via an enhancer on chr4:121696658-121696872) and RELA regulates TNFa-induced proteins, TNFAIP3 and TNFAIP6. Overall, NF-kBs regulate TNF-a, increasing expression during AD progression, likely triggering neurodegeneration, inflammation, neuronal death, and healthy tissue destruction[23].

RELA regulate genes of several cytokines associated with the severe Covid-19 Cytokine Storm. For instance, RELA regulates IL-6 and particularly activates IL-12A/B (recruit and activate Natural Killer cells[100]) and IL-1B via the enhancers. In **Fig. S32**, we shared additional examples of NFKB1 and RELA TFs regulating other TFs that regulate inflammatory cytokines IL-1B, IL-12B (recruit and activate Natural Killer cells[100]), CCL2, MMP 1/3, and CLGN. For instance, in the Hippocampus, NFKB1 regulates TFs SPI1 and BATF, regulating MMP1 (**Fig. S32B**). The expression of cytokines IL-2, CCL2, IL-1B IL-12B, and TNFa highly positively correlate with AD severe stages (**Fig. 5E**). IL-2 and TNFa are usually highly expressed in Covid-19 patients with severe pneumonia who are developing ARDS and need intensive care and oxygen therapy[93]. NFKB1 and RELA belong to the same Hippocampal gene module.

Further, APOE genotype is associated with differences in Complement Cascade Component C1qrs expression in Covid-19 patients[24] in the DLPFC, as it is negatively correlated with E2/E2 but positively with E4/E4 (**Fig. S33**). C1qrs activates microglia to the M1 state, where they

817 release mediators that increase inflammation and damage healthy cells[101]. We found that
818 C1qrs is negatively correlated with Control and Initial Stages, but positively correlated with
819 Moderate and Severe Stages in the Hippocampus (**Fig. S34-35**). Complement activation is
820 involved in an inflammatory feedback loop with neutrophil activation (resulting in tissue injury)[102],
821 and may be a hallmark of severe Covid-19. Many complement components are instead
822 negatively correlated with Control and Initial Stages in the Hippocampus (except MBL and
823 VWF), but positively correlated with AD progression. IgG antibodies, whose responses to
824 various epitopes are key to the immune response to Covid-19[103], are the only component
825 positively associated with Moderate AD but not for Severe AD. Fibrinogen and SELP changed
826 from negative to positive associations from Moderate to Severe AD stages.
827
828 Additionally, our GRNs help find potential roles of AD SNPs to NFKB for Covid-19. In **Fig. S27**,
829 we examine the impact of various AD SNPs on the expression of NF-kB TFs and NFKB
830 regulation of key cytokines described in **Fig. 5D**. We found several SNPs disrupting regulation
831 of RELA and NFKB1 in the Hippocampus (**Fig. S36**). Moreover, there are 4 highly significant
832 SNPs ($p < 5e-15$) impacting RELA's ability to regulate ZNF226, a hub TF from the
833 Hippocampus SNP TF-TF Subnetwork GRN (**Fig. S27B**); RELA also regulates ZNF226's gene
834 module. 99 extremely significant SNPs ($p < 1e-9$) impact regulation of RELB in DLPFC (**Fig.**
835 **S37**). In Hippocampal GRN, SNP rs71350303 disrupts RELA regulation of TNFAIP6 (**Fig.**
836 **S38B**) and in the LTL GRN SNP rs6425995 disrupts NFKB1 regulation of CSF3R **(Fig. S38C**).
837 Besides the NFKB pathway, we also found several other shared pathways in AD and Covid-19
838 from KEGG such as IKK, TNFR, PI3K, JNK, and IL6 (**Fig. S39**). We thus looked at the
839 correlations between AD phenotypes and genes from those pathways in each brain region (**Fig.**
840 **S40**). Finally, we identified highly correlated AD-COVID pathways and AD phenotypes, e.g.,
841 TNFR with severe stage and IKK with cognitive impairment in hippocampus, IKK with frequent
842 plaques in LTL, JNK with Resilience in DLPFC).
843

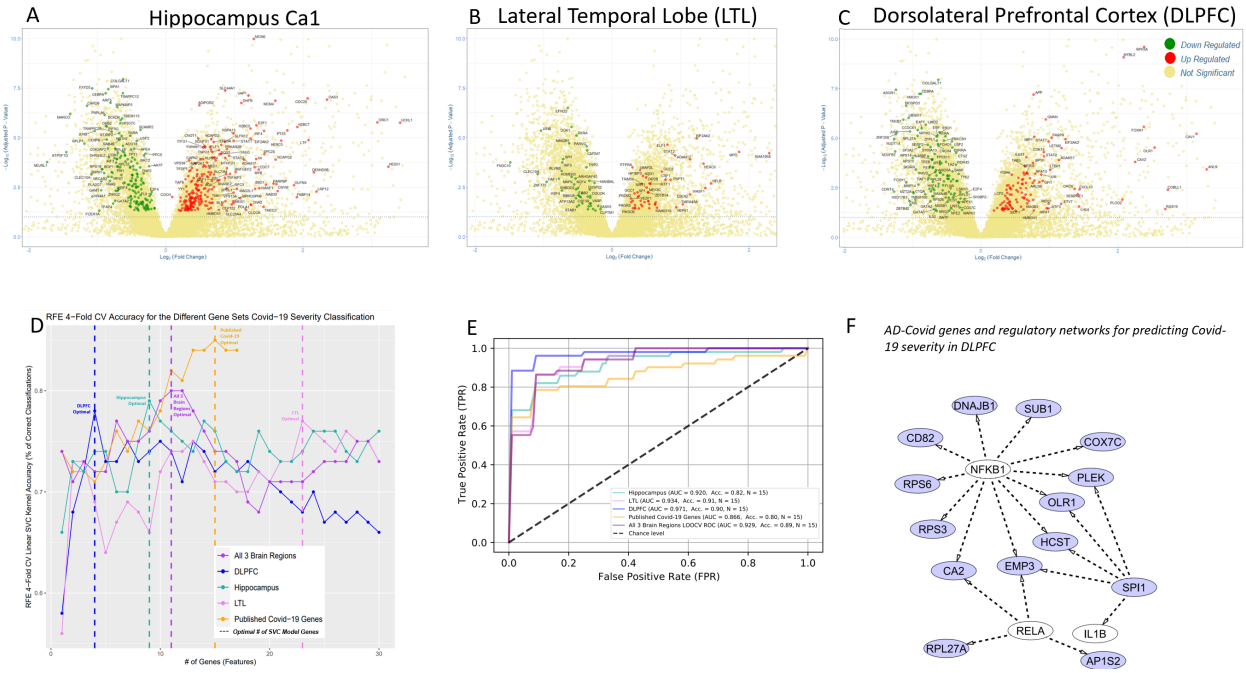


Fig. 6 – Differential expression and prediction of Covid-19 severity using AD-Covid gene regulatory networks. Top volcano plots show differential expression analyses for Covid-19 severity (i.e., Covid-19 ICU, Methods) and label the genes from the AD-Covid gene regulatory networks (GRNs) that relate to 22 AD-Covid genes: (**A**) Hippocampus, (**B**) Lateral Temporal Lobe (LTL), (**C**) Dorsolateral Prefrontal Cortex. Red: up-regulated. Green: down-regulated. Yellow: No significance. x-axis: log2(fold change). y-axis: -log10(Adjusted p-value). (**D**) Prediction accuracy of Covid-19 severity after selecting different numbers of genes from AD-Covid GRNs and recently found Covid-19 genes (benchmark genes). The accuracy was calculated based on the support vector machine classification with 4-fold cross-validation. The dashed lines correspond to the minimal numbers of select genes with highest accuracy (i.e., optimal gene sets for predicting Covid-19 severity). (**E**) Receiver Operating Characteristic (ROC) curves and area under curve (AUC) values for classifying Covid-19 severity in (**D**). (**F**) Subnetwork of DLPFC GRN relating to the 15 DLPFC optimal genes (excluding JUND) for predicting Covid-19 severity (N=15) with AD-Covid shared genes. Blue: genes/TFs found in the DLPFC final model. White: AD-Covid shared genes.

Machine learning prediction of Covid-19 severity from AD-Covid gene regulatory networks

All gene lists and machine learning prediction results for this section are available in **Supplementary file 9**. In total, we found 22 genes from the pathways that are shared between Covid-19 and AD in the KEGG database. We also looked at our brain-region gene regulatory networks (GRNs) that relate to those AD-Covid genes, including TFs that regulate them as well as their target genes, i.e., AD-Covid GRNs. We found 1,305 genes from hippocampus GRN, 670 genes from LTL GRN, and 895 genes from DLPFC GRN (2,536 unique genes in total, and 38 genes shared by three regions). As shown in **Fig. S41**, those 38 shared genes (17 found in both Covid and AD KEGG Pathways) across AD-Covid GRNs also highly correlate with AD phenotypes such as clinical Braak stage.

We found that many genes from our AD-Covid GRNs are significantly differentially expressed in the Covid-19 severity condition. Moreover, the Covid-19 positive phenotype is positively correlated with many of the mechanisms in the AD KEGG pathway (**Fig. S42**). In particular, we normalized gene expression data (**Fig. S43**) of a recent Covid-19 cohort (N=50 ICU vs. N=50 non-ICU)[108] and identified differentially expressed genes (DEGs) for Covid-19 ICU (Methods). We found 4,692 DEGs (2,490 up-regulated, 2,202 down-regulated genes) in total. Out of those DEGs, 403 DEGs are from the AD-Covid GRN in hippocampus (232 up-regulated, 171 down-regulated, **Fig. 6A**). Similarly, LTL's AD-Covid GRN has 95 DEGs (51 up-regulated, 44 down-regulated, **Fig. 6B**), and DLPFC's has 223 DEGs (113 up-regulated including APP, 110 down-regulated, **Fig. 6C**). These DEGs suggest that genes from our AD-Covid GRNs significantly associate with Covid-19 severity. Thus, beyond association, we further want to develop a model to predict Covid-19 severity from those genes.

In particular, we first gathered 17 benchmark Covid-19 susceptibility genes from recent studies[54–56]. Then, we used recursive feature elimination (RFE) to select an optimal number of genes from the list of 17 published Covid-19 genes for Covid-19 severity, i.e., 4-fold cross-validated feature selection with highest classification accuracy based on linear support vector machine (Methods, **Fig. 6D**). This resulted in 15 genes being optimal for the published genes model. Then, we performed RFE feature selection for each region's AD-Covid GRN to select the top 15 genes; building linear kernel SVM predicted probability models with the same number of genes (15) from each of the 5 gene lists would enable us to directly compare the effectiveness of our AD-Covid GRNs with that of the published genes.

Our prediction accuracy based on 15 genes for each AD-Covid GRN is also higher than the optimal benchmark of 15 Covid-19 genes (86% for Hippocampus, 89.99% for DLPFC, 90.99% for LTL, and 88.99% for combined regions, 80% for the benchmark). As shown in **Fig. 6E**, our areas under the ROC curve (AUROC) values are also larger than the benchmark (0.912 for Hippocampus, 0.971 for DLPFC, 0.934 for LTL, 0.929 for combined regions, 0.866 for benchmark). Relative to the benchmark, the optimal model (in terms of highest accuracy and highest AUC) improved accuracy by 9.99% and boosted the AUC by 0.105. Therefore, this suggests that the select genes from our AD-Covid GRNs have higher predictability than existing Covid-19 genes for predicting Covid-19 severity.

We highlight the subnetwork of the DLPFC GRN for 14 out of the 15 optimal predictive genes (excluding JUND) that are directly regulating or regulated by at least 1 of the 22 shared AD-Covid genes (**Fig. 6F**). 3 of 22 AD-Covid shared genes (NFKB1, RELA, and IL1B) were found in this network. Indeed, NFKB1 regulates 11 out of 15 DLPFC model genes (NCOR1, CCR5, and ERAP1) and RELA regulates 4 genes, further underscoring the importance of NF-kB TFs in Covid-19 outcomes. Microglia TF SPI1 regulates IL1B along with 4 other DLPFC model genes, which also supports research into immune dysregulation in both Covid-19 and AD.

We also evaluated and compared our predictive models with benchmark genes using Decision Curve Analysis (DCA). DCA enables evaluating the clinical usability of our Covid-19 severity prediction models based on their Net Benefits (**Methods**). We plotted the Decision Curves to

show how the Net Benefit of each model varies across probability thresholds (**Fig. S44**). And the threshold of each model that gives the highest Net Benefit corresponds to the optimal decision probability for sending Covid-19 patients to ICU or not, i.e., the "optimal" threshold (**Fig. S45**). In general, the models of genes from our AD-Covid GRNs have higher Net Benefits than the benchmark Covid-19 genes, especially around the optimal thresholds that achieve possible maximum Net Benefits. Note that since 50% of our Covid-19 positive patients are in the ICU, the maximum Net Benefit is 0.50. The increase in Net Benefit of our models (around 0.00121 to 0.2443 at optimal thresholds with an average increase of 0.119 in Net Benefit) compared with the benchmark could be interpreted that using the genes from our AD-Covid GRNs on average increases the number of truly severe Covid-19 patients detected by about 119 per 1000 Covid-19 positive patients, without changing the number of non-severe patients who are needlessly sent to the ICU (**Fig. S46**). Thus, those genes along with our predictive models provide potential novel strategies for helping clinical decisions on sending Covid-19 patients to ICU or not.

Discussion

In this paper, we performed an integrative analysis of multi-omics for predicting the underlying gene regulatory networks for disease phenotypes that link disease variants to TFs to regulatory elements to genes and modules. We applied our analysis to the multi-omics datasets of three AD-relevant brain regions DLPFC, Hippocampus and Lateral Temporal Lobe and predicted brain regional gene regulatory networks for various AD phenotypes such as progression The results revealed how potential causal AD risk variants lead to AD phenotypes via gene regulation. However, our analysis is open-source available and thus can serve as general-purpose for understanding functional genomics and gene regulation for other diseases.

Gene regulation typically fundamentally affects biological and disease functions at the cellular resolution. Recent sequencing data of single cells such as scRNA-seq and scATAC-seq, especially in AD and other brain diseases, enable studying the functional genomics and regulatory mechanisms at the cell type level[104]. For instance, many cell-type gene regulatory networks in the human brain, such as neuronal and glial types, have been predicted from single cell data. Shortly, we can perform integrative analysis of those cell-type networks to understand the regulatory mechanism from AD variants that cause AD for different cell types. Also, many other phenotypes are observed in AD. Increasing GWAS studies[105] have identified additional variants associated with refined AD phenotypes such as cerebrospinal fluid and psychotic symptoms. We aim to predict the gene regulatory networks of those variants for additional AD phenotypes in future.

Also, we found that some regulatory networks in AD also relate to the immunological functions and pathways for Covid-19. Further, our machine learning and Decision Curve analyses show that the genes from those AD-Covid regulatory networks better predict Covid-19 severity (i.e., ICU) than the known Covid-19 genes. With a dramatic increase of discovered variants for Covid-19, our integrative analysis will allow us to predict the regulatory mechanisms of the Covid-19 variants in AD phenotypes. If an exaggerated immune response (found in Covid-19) can lead to heightened and more severe AD, then perhaps the immune response may indeed

be rogue in AD. We can then use Covid-19 as a proxy and biomarker to better understand the role of a dysregulated immune system in AD onset and progression. This will help advance our understanding of the interplay between Covid-19, neuroimmune and AD phenotypes.

Furthermore, we integrated the data from different population studies via step by step. Many large scientific consortia have generated matched multi-omic data of individuals such as PsychENCODE[35], AMP-AD[106], TCGA[107]. Moreover, machine learning has been widely used for predicting phenotypes from individual data. Thus, we can extend our analysis to input matched omics data of individuals (e.g., genotype, gene expression, epigenomics) at the population level to train machine learning models to predict personalized phenotypes. The resulting predictive models can be further used to predict personalized phenotypes for new individual data and prioritize phenotype-specific functional genomics and gene regulatory networks in human diseases.

Supplementary Information

Supplementary file 1 – Genes, modules, phenotypes, enrichments, and Transcription Factors regulating gene modules for Hippocampus Ca1
Supplementary file 2 – Genes, modules, phenotypes, and enrichments and Transcription Factors regulating gene modules for Lateral Temporal Lobe (LTL)
Supplementary file 3 – Genes, modules, phenotypes, and enrichments for the Dorsolateral Prefrontal Cortex (DLPFC)
Supplementary file 4 – Gene Regulatory Network (GRN) for the Hippocampus Ca1
Supplementary file 5 – Gene Regulatory Network for the Lateral Temporal Lobe (LTL)
Supplementary file 6 – SNPs Interrupting Transcription Factor Binding Sites (TFBS) in Hippocampus Ca1 Gene Regulatory Network (GRN)
Supplementary file 7 – SNPs Interrupting Transcription Factor Binding Sites (TFBS) in Lateral Temporal Lobe (LTL) Gene Regulatory Network (GRN)
Supplementary file 8 – SNPs Interrupting Transcription Factor Binding Sites (TFBS) in DLPFC Gene Regulatory Network (GRN)
Supplementary file 9 – shared AD and Covid-19 genes, 5 initial gene lists (for Hippocampus Ca1, Lateral Temporal Lobe (LTL), Dorsolateral Prefrontal Cortex (DLPFC), All 3 combined, Published Covid-19 genes), final SVC linear kernel genes for 5 models, predicted probabilities from each model, and Decision Curve Analysis (Net Benefit of each model for various probability thresholds).
Supplementary document – Supplementary Figures S1 to S46 and Supplementary Tables S1 to S3.

Acknowledgements

The authors are grateful to Jonathan Edward Bryan for his assistance in tasks related to this project and Renu Khullar for her helpful comments on the manuscript.

Funding

This work was supported by the grants of National Institutes of Health, R01AG067025, R21CA237955 and U01MH116492 to Daifeng Wang and U54HD090256 to Waisman Center. Saniya Khullar was also supported by an NLM training grant to the Computation and Informatics in Biology and Medicine Training Program (NLM 5T15LM007359).

Conflict of Interest

None declared.

Contributions

D.W. conceived and designed the study. S.K. and D.W. analyzed the data and wrote the manuscript. All authors read and approved the final manuscript.

References

1. Alzheimer's Statistics. *Alzheimers.net* https://www.alzheimers.net/resources/alzheimers-statistics.
2. Rabinovici, G. D. Late-onset Alzheimer Disease. *Contin. Lifelong Learn. Neurol.* **25**, 14–33 (2019).
3. Heneka, M. T. *et al.* Neuroinflammation in Alzheimer's Disease. *Lancet Neurol.* **14**, 388–405 (2015).
4. CA1 neurons in the human hippocampus are critical for autobiographical memory, mental time travel, and autonoetic consciousness | PNAS. https://www.pnas.org/content/108/42/17562.
5. Goldstein, I. S., Erickson, D. J., Sleeper, L. A., Haynes, R. L. & Kinney, H. C. The Lateral Temporal Lobe in Early Human Life. *J. Neuropathol. Exp. Neurol.* **76**, 424–438 (2017).
6. Nativio, R. *et al.* An integrated multi-omics approach identifies epigenetic alterations associated with Alzheimer's disease. *Nat. Genet.* **52**, 1024–1035 (2020).
7. Emotional Dysfunction in Psychopathology and Neuropathology: Neural and Genetic Pathways - ScienceDirect. https://www.sciencedirect.com/science/article/pii/B9780128001059000226.

8. Brinton, R. D., Gore, A. C., Schmidt, P. J. & Morrison, J. H. 68 - Reproductive Aging of Females: Neural Systems. in *Hormones, Brain and Behavior (Second Edition)* (eds. Pfaff, D. W., Arnold, A. P., Etgen, A. M., Fahrbach, S. E. & Rubin, R. T.) 2199–2224 (Academic Press, 2009). doi:10.1016/B978-008088783-8.00068-1.

9. Kumar, S. *et al.* Dorsolateral prefrontal cortex metabolites and their relationship with plasticity in Alzheimer's disease. *Alzheimers Dement.* **16**, e045879 (2020).

10. Marques-Coelho, D. *et al.* Differential transcript usage unravels gene expression alterations in Alzheimer's disease human brains. *Npj Aging Mech. Dis.* **7**, 1–15 (2021).

11. Morabito, S., Miyoshi, E., Michael, N. & Swarup, V. Integrative genomics approach identifies conserved transcriptomic networks in Alzheimer's disease. *Hum. Mol. Genet.* **29**, 2899–2919 (2020).

12. Co-expression modules construction by WGCNA and identify potential prognostic markers of uveal melanoma - ScienceDirect. https://www.sciencedirect.com/science/article/abs/pii/S0014483517303536?via%3Dihub.

13. Liang, J.-W. *et al.* Application of Weighted Gene Co-Expression Network Analysis to Explore the Key Genes in Alzheimer's Disease. *J. Alzheimers Dis.* **65**, 1353–1364.

14. Neff, R. A. *et al.* Molecular subtyping of Alzheimer's disease using RNA sequencing data reveals novel mechanisms and targets. *Sci. Adv.* **7**, eabb5398 (2021).

15. Jansen, I. E. *et al.* Genome-wide meta-analysis identifies new loci and functional pathways influencing Alzheimer's disease risk. *Nat. Genet.* **51**, 404–413 (2019).

16. Novikova, G. *et al.* Integration of Alzheimer's disease genetics and myeloid genomics identifies disease risk regulatory elements and genes. *Nat. Commun.* **12**, 1610 (2021).

17. Wingo, A. P. *et al.* Integrating human brain proteomes with genome-wide association data implicates new proteins in Alzheimer's disease pathogenesis. *Nat. Genet.* **53**, 143–146 (2021).

1050 18. Kumar, S., Ambrosini, G. & Bucher, P. SNP2TFBS – a database of regulatory SNPs
1051 affecting predicted transcription factor binding site affinity. *Nucleic Acids Res.* **45**, D139–
1052 D144 (2017).
1053 19. Ch, Zuo, ler, Shin, S. & Keles, S. *atSNP: Affinity test for identifying regulatory SNPs*.
1054 (Bioconductor version: Release (3.12), 2021). doi:10.18129/B9.bioc.atSNP.
1055 20. Huo, Y., Li, S., Liu, J., Li, X. & Luo, X.-J. Functional genomics reveal gene regulatory
1056 mechanisms underlying schizophrenia risk. *Nat. Commun.* **10**, 670 (2019).
1057 21. Naughton, S. X., Raval, U. & Pasinetti, G. M. Potential Novel Role of COVID-19 in
1058 Alzheimer's Disease and Preventative Mitigation Strategies. *J. Alzheimers Dis. JAD* **76**,
1059 21–25 (2020).
1060 22. Neurological Manifestations of COVID-19 Feature T Cell Exhaustion and Dedifferentiated
1061 Monocytes in Cerebrospinal Fluid: Immunity. https://www.cell.com/immunity/fulltext/S1074-
1062 7613(20)30539-
1063 2?_returnURL=https%3A%2F%2Flinkinghub.elsevier.com%2Fretrieve%2Fpii%2FS107476
1064 1320305392%3Fshowall%3Dtrue.
1065 23. Landhuis, E. Could the immune system be key to Alzheimer's disease? *Knowable Mag.*
1066 *Annu. Rev.* (2021) doi:10.1146/knowable-012921-3.
1067 24. Inal, J. Biological Factors Linking ApoE ε4 Variant and Severe COVID-19. *Curr.*
1068 *Atheroscler. Rep.* **22**, (2020).
1069 25. Blalock, E. M. *et al.* Incipient Alzheimer's disease: Microarray correlation analyses reveal
1070 major transcriptional and tumor suppressor responses. *Proc. Natl. Acad. Sci.* **101**, 2173–
1071 2178 (2004).
1072 26. mail.nih.gov>, S. D. <sdavis2 at. *GEOquery: Get data from NCBI Gene Expression*
1073 *Omnibus (GEO)*. (Bioconductor version: Release (3.12), 2021).
1074 doi:10.18129/B9.bioc.GEOquery.
1075 27. hgu133a.db. *Bioconductor* http://bioconductor.org/packages/hgu133a.db/.

28. hgu133acdf. *Bioconductor* http://bioconductor.org/packages/hgu133acdf/.
29. ds.harvard.edu>, R. A. I. <rafa at *et al. affy: Methods for Affymetrix Oligonucleotide Arrays*. (Bioconductor version: Release (3.12), 2021). doi:10.18129/B9.bioc.affy.
30. RMA normalization for microarray data. https://felixfan.github.io/RMA-Normalization-Microarray/.
31. The Religious Orders Study and Memory and Aging Project (ROSMAP) Study. https://adknowledgeportal.synapse.org/Explore/Studies/DetailsPage?Study=syn3219045.
32. An atlas of chromatin accessibility in the adult human brain. https://bendlj01.u.hpc.mssm.edu/multireg/.
33. Jung, I. *et al.* A compendium of promoter-centered long-range chromatin interactions in the human genome. *Nat. Genet.* **51**, 1442–1449 (2019).
34. TxDb.Hsapiens.UCSC.hg19.knownGene. *Bioconductor* http://bioconductor.org/packages/TxDb.Hsapiens.UCSC.hg19.knownGene/.
35. Resource.PsychEncode. http://resource.psychencode.org/.
36. Langfelder, P. & Horvath, S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* **9**, 559 (2008).
37. Botía, J. A. *et al.* An additional k-means clustering step improves the biological features of WGCNA gene co-expression networks. *BMC Syst. Biol.* **11**, 47 (2017).
38. Safieh, M., Korczyn, A. D. & Michaelson, D. M. ApoE4: an emerging therapeutic target for Alzheimer's disease. *BMC Med.* **17**, 64 (2019).
39. Jin, T. *et al.* scGRNom: a computational pipeline of integrative multi-omics analyses for predicting cell-type disease genes and regulatory networks. *Genome Med.* **13**, 95 (2021).
40. Tan, G. *TFBSTools: Software Package for Transcription Factor Binding Site (TFBS) Analysis*. (Bioconductor version: Release (3.12), 2021). doi:10.18129/B9.bioc.TFBSTools.
41. Schep, A. & University, S. *motifmatchr: Fast Motif Matching in R*. (Bioconductor version: Release (3.12), 2021). doi:10.18129/B9.bioc.motifmatchr.

1102 42. Groeneveld, C. *et al. RTN: RTN: Reconstruction of Transcriptional regulatory Networks*
1103 *and analysis of regulons*. (Bioconductor version: Release (3.12), 2021).
1104 doi:10.18129/B9.bioc.RTN.
1105 43. systemsbiology.org>, S. A. <seth ament at, systemsbioloyg.org>, P. S. <pshannon at &
1106 systemsbiology.org>, M. R. <mrichard at. *trena: Fit transcriptional regulatory networks*
1107 *using gene expression, priors, machine learning*. (Bioconductor version: Release (3.12),
1108 2021). doi:10.18129/B9.bioc.trena.
1109 44. Huynh-Thu, V. A., Aibar, S. & Geurts, P. *GENIE3: GEne Network Inference with Ensemble*
1110 *of trees*. (Bioconductor version: Release (3.12), 2021). doi:10.18129/B9.bioc.GENIE3.
1111 45. The Human Transcription Factors: Cell. https://www.cell.com/cell/fulltext/S0092-
1112 8674(18)30106-
1113 5?_returnURL=https%3A%2F%2Flinkinghub.elsevier.com%2Fretrieve%2Fpii%2FS009286
1114 7418301065%3Fshowall%3Dtrue.
1115 46. JASPAR 2020: An open-access database of transcription factor binding profiles.
1116 http://jaspar.genereg.net.
1117 47. Margolin, A. A. *et al.* ARACNE: An Algorithm for the Reconstruction of Gene Regulatory
1118 Networks in a Mammalian Cellular Context. *BMC Bioinformatics* **7**, S7 (2006).
1119 48. Comprehensive functional genomic resource and integrative model for the human brain |
1120 Science. https://science.sciencemag.org/content/362/6420/eaat8464.
1121 49. Coetzee, S. G. & Hazelett, D. J. *motifbreakR: A Package For Predicting The*
1122 *Disruptiveness Of Single Nucleotide Polymorphisms On Transcription Factor Binding Sites*.
1123 (Bioconductor version: Release (3.12), 2021). doi:10.18129/B9.bioc.motifbreakR.
1124 50. KEGG PATHWAY Database. https://www.genome.jp/kegg/pathway.html.
1125 51. Luo, W. *pathview: a tool set for pathway based data integration and visualization*.
1126 (Bioconductor version: Release (3.12), 2021). doi:10.18129/B9.bioc.pathview.

52. Overmyer, K. A. *et al.* Large-Scale Multi-omic Analysis of COVID-19 Severity. *Cell Syst.* **12**, 23-40.e7 (2021).

53. Love, M., Ahlmann-Eltze, C., Forbes, K., Anders, S. & Huber, W. *DESeq2: Differential gene expression analysis based on the negative binomial distribution*. (Bioconductor version: Release (3.12), 2021). doi:10.18129/B9.bioc.DESeq2.

54. Hu, J., Li, C., Wang, S., Li, T. & Zhang, H. Genetic variants are identified to increase risk of COVID-19 related mortality from UK Biobank data. *Hum. Genomics* **15**, 10 (2021).

55. Genetic mechanisms of critical illness in COVID-19 | Nature. https://www.nature.com/articles/s41586-020-03065-y#citeas.

56. New insights into genetic susceptibility of COVID-19: an ACE2 and TMPRSS2 polymorphism analysis | BMC Medicine | Full Text. https://bmcmedicine.biomedcentral.com/articles/10.1186/s12916-020-01673-z#citeas.

57. scikit-learn: machine learning in Python — scikit-learn 0.24.1 documentation. https://scikit-learn.org/stable/.

58. sklearn.svm.SVC — scikit-learn 0.24.1 documentation. https://scikit-learn.org/stable/modules/generated/sklearn.svm.SVC.html.

59. Vickers, A. J. & Elkin, E. B. Decision Curve Analysis: A Novel Method for Evaluating Prediction Models. *Med. Decis. Making* **26**, 565–574 (2006).

60. Biostatistics: Decision Curve Analysis | Memorial Sloan Kettering Cancer Center. https://www.mskcc.org/departments/epidemiology-biostatistics/biostatistics/decision-curve-analysis.

61. Bezprozvanny, I. & Mattson, M. P. Neuronal calcium mishandling and the pathogenesis of Alzheimer's disease. *Trends Neurosci.* **31**, 454–463 (2008).

62. Rohilla, K. J. & Gagnon, K. T. RNA biology of disease-associated microsatellite repeat expansions. *Acta Neuropathol. Commun.* **5**, (2017).

1152 63. Lee, J. K. & Kim, N.-J. Recent Advances in the Inhibition of p38 MAPK as a Potential
1153 Strategy for the Treatment of Alzheimer's Disease. *Mol. J. Synth. Chem. Nat. Prod. Chem.*
1154 **22**, (2017).
1155 64. Nuclear factor-kappa β as a therapeutic target for Alzheimer's disease - Jha - 2019 -
1156 Journal of Neurochemistry - Wiley Online Library.
1157 https://onlinelibrary.wiley.com/doi/10.1111/jnc.14687.
1158 65. Pei, J. *et al.* Excessive Activation of TLR4/NF-κB Interactively Suppresses the Canonical
1159 Wnt/β-catenin Pathway and Induces SANFH in SD Rats. *Sci. Rep.* **7**, 11928 (2017).
1160 66. JCI - Type I interferon response drives neuroinflammation and synapse loss in Alzheimer
1161 disease. https://www.jci.org/articles/view/133737.
1162 67. Metabolic correlates of prevalent mild cognitive impairment and Alzheimer's disease in
1163 adults with Down syndrome - Mapstone - 2020 - Alzheimer's & Dementia: Diagnosis,
1164 Assessment & Disease Monitoring - Wiley Online Library. https://alz-
1165 journals.onlinelibrary.wiley.com/doi/full/10.1002/dad2.12028.
1166 68. Wnt signaling in the nervous system and in Alzheimer's disease | Journal of Molecular Cell
1167 Biology | Oxford Academic. https://academic.oup.com/jmcb/article/6/1/64/874321.
1168 69. Neurodegeneration and microtubule dynamics: death by a thousand cuts.
1169 https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4563776/.
1170 70. Vasile, F., Dossi, E. & Rouach, N. Human astrocytes: structure and functions in the healthy
1171 brain. *Brain Struct. Funct.* **222**, 2017–2029 (2017).
1172 71. Prion protein and Alzheimer disease.
1173 https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2807690/.
1174 72. Esposito, M. & Sherr, G. L. Epigenetic Modifications in Alzheimer's Neuropathology and
1175 Therapeutics. *Front. Neurosci.* **13**, (2019).

1176 73. Frontiers | Microglia in Alzheimer Disease: Well-Known Targets and New Opportunities |
1177 Frontiers in Aging Neuroscience.
1178 https://www.frontiersin.org/articles/10.3389/fnagi.2019.00233/full.
1179 74. Frontiers | Therapeutic Inhibition of the Complement System in Diseases of the Central
1180 Nervous System | Immunology.
1181 https://www.frontiersin.org/articles/10.3389/fimmu.2019.00362/full.
1182 75. Keshavan, M., Lizano, P. & Prasad, K. The synaptic pruning hypothesis of schizophrenia:
1183 promises and challenges. *World Psychiatry* **19**, 110–111 (2020).
1184 76. Navarro, V. *et al.* Microglia in Alzheimer's Disease: Activated, Dysfunctional or
1185 Degenerative. *Front. Aging Neurosci.* **10**, (2018).
1186 77. Ebanks, B., Ingram, T. L. & Chakrabarti, L. ATP synthase and Alzheimer's disease: putting
1187 a spin on the mitochondrial hypothesis. *Aging* **12**, 16647–16662 (2020).
1188 78. Costarelli, L., Malavolta, M., Giacconi, R. & Provinciali, M. Dysfunctional macrophages in
1189 Alzheimer Disease: another piece of the "macroph-aging" puzzle? *Aging* **9**, 1865–1866
1190 (2017).
1191 79. Du, Y. *et al.* MKP-1 reduces Aβ generation and alleviates cognitive impairments in
1192 Alzheimer's disease models. *Signal Transduct. Target. Ther.* **4**, 1–12 (2019).
1193 80. Ki-1 Antigen | Colorado PROFILES. https://profiles.ucdenver.edu/display/218590.
1194 81. Cho, S.-J., Park, M. H., Han, C., Yoon, K. & Koh, Y. H. VEGFR2 alteration in Alzheimer's
1195 disease. *Sci. Rep.* **7**, 17713 (2017).
1196 82. Complement C3 Deficiency Leads to Accelerated Amyloid β Plaque Deposition and
1197 Neurodegeneration and Modulation of the Microglia/Macrophage Phenotype in Amyloid
1198 Precursor Protein Transgenic Mice | Journal of Neuroscience.
1199 https://www.jneurosci.org/content/28/25/6333.
1200 83. Quintela-López, T. *et al.* Aβ oligomers promote oligodendrocyte differentiation and
1201 maturation via integrin β1 and Fyn kinase signaling. *Cell Death Dis.* **10**, 1–16 (2019).

1202 84. Lu, T. *et al.* REST and stress resistance in ageing and Alzheimer's disease. *Nature* **507**,
1203 448–454 (2014).
1204 85. Maezawa, I., Jenkins, D. P., Jin, B. E. & Wulff, H. Microglial KCa3.1 Channels as a
1205 Potential Therapeutic Target for Alzheimer's Disease. *Int. J. Alzheimers Dis.* **2012**, (2012).
1206 86. Magno, L. *et al.* Alzheimer's disease phospholipase C-gamma-2 (PLCG2) protective
1207 variant is a functional hypermorph. *Alzheimers Res. Ther.* **11**, 16 (2019).
1208 87. Liu, N. *et al.* Hippocampal transcriptome-wide association study and neurobiological
1209 pathway analysis for Alzheimer's disease. *PLOS Genet.* **17**, e1009363 (2021).
1210 88. Yashiro, T. *et al.* A transcription factor PU.1 is critical for Ccl22 gene expression in
1211 dendritic cells and macrophages. *Sci. Rep.* **9**, 1161 (2019).
1212 89. PU.1 regulates Alzheimer's disease-associated genes in primary human microglia |
1213 Molecular Neurodegeneration | Full Text.
1214 https://molecularneurodegeneration.biomedcentral.com/articles/10.1186/s13024-018-0277-
1215 1.
1216 90. Early induction of c-Myc is associated with neuronal cell death - ScienceDirect.
1217 https://www.sciencedirect.com/science/article/abs/pii/S0304394011013899.
1218 91. Sáez, M. E. *et al.* Genome Wide Meta-Analysis identifies common genetic signatures
1219 shared by heart function and Alzheimer's disease. *Sci. Rep.* **9**, 16665 (2019).
1220 92. Yi, M. *et al.* The potassium channel KCa3.1 constitutes a pharmacological target for
1221 astrogliosis associated with ischemia stroke. *J. Neuroinflammation* **14**, (2017).
1222 93. Kircheis, R. *et al.* NF-κB Pathway as a Potential Target for Treatment of Critical Stage
1223 COVID-19 Patients. *Front. Immunol.* **11**, (2020).
1224 94. steve. What is NF-κB pathway? *MBInfo* https://www.mechanobio.info/what-is-
1225 mechanosignaling/signaling-pathways/what-is-the-nf-%ce%bab-pathway/.
1226 95. Kwee, T. C. & Kwee, R. M. Chest CT in COVID-19: What the Radiologist Needs to
1227 Know. *RadioGraphics* **40**, 1848–1865 (2020).

96. Targeting transcriptional regulation of SARS-CoV-2 entry factors ACE2 and TMPRSS2 | PNAS. https://www.pnas.org/content/118/1/e2021450118.

97. Liu, M. *et al.* CXCL10/IP-10 in infectious diseases pathogenesis and potential therapeutic implications. *Cytokine Growth Factor Rev.* **22**, 121–130 (2011).

98. Lance, A. *et al.* Altered expression of CSF3R splice variants impacts signal response and is associated with SRSF2 mutations. *Leukemia* **34**, 369–379 (2020).

99. Frontiers | Neutrophils in COVID-19 | Immunology. https://www.frontiersin.org/articles/10.3389/fimmu.2021.652470/full.

100. Parham, P. *The Immune System*. (Garland Science, 2014).

101. Hammad, A., Westacott, L. & Zaben, M. The role of the complement system in traumatic brain injury: a review. *J. Neuroinflammation* **15**, (2018).

102. Java, A. *et al.* The complement system in COVID-19: friend and foe? *JCI Insight* **5**,.

103. Heffron, A. S. *et al.* The landscape of antibody binding in SARS-CoV-2 infection. *bioRxiv* 2020.10.10.334292 (2021) doi:10.1101/2020.10.10.334292.

104. Jiang, J., Wang, C., Qi, R., Fu, H. & Ma, Q. scREAD: A Single-Cell RNA-Seq Database for Alzheimer's Disease. *iScience* **23**, 101769 (2020).

105. GWAS Catalog. https://www.ebi.ac.uk/gwas/search?query=Alzheimer%27s%20disease.

106. AD Knowledge Portal -. https://adknowledgeportal.synapse.org/.

107. The Cancer Genome Atlas Program - National Cancer Institute. https://www.cancer.gov/about-nci/organization/ccg/research/structural-genomics/tcga (2018).