

A Network-based Deep Learning Framework Catalyzes GWAS and Multi-Omics Findings to Biology and Drug Repurposing for Alzheimer's Disease

Jielin Xu¹, Yuan Hou¹, Yadi Zhou¹, Ming Hu², Feixiong Cheng^{1,3-4}

¹Genomic Medicine Institute, Lerner Research Institute, Cleveland Clinic, Cleveland, OH 44195, USA

²Department of Quantitative Health Sciences, Lerner Research Institute, Cleveland Clinic, Cleveland, OH 44106, USA

³Department of Molecular Medicine, Cleveland Clinic Lerner College of Medicine, Case Western Reserve University, Cleveland, OH 44195, USA

⁴Case Comprehensive Cancer Center, Case Western Reserve University School of Medicine, Cleveland, OH 44106, USA

*Correspondence to: Feixiong Cheng, PhD

Lerner Research Institute, Cleveland Clinic

Tel: 216-444-7654; Fax: 216-636-0009

Email: chengf@ccf.org

Abstract

Human genome sequencing studies have identified numerous loci associated with complex diseases, including Alzheimer's disease (AD). Translating human genetic findings (i.e., genome-wide association studies [GWAS]) to pathobiology and therapeutic discovery, however, remains a major challenge. To address this critical problem, we present a **network** topology-based deep learning framework to identify disease-associated **genes** (NETTAG). NETTAG is capable of integrating multi-genomics data along with the protein-protein interactome to infer putative risk genes and drug targets impacted by GWAS loci. Specifically, we leverage non-coding GWAS loci effects on expression quantitative trait loci (eQTLs), histone-QTLs, and transcription factor binding-QTLs, enhancers and CpG islands, promoter regions, open chromatin, and promoter flanking regions. The key premises of NETTAG are that the disease risk genes exhibit distinct functional characteristics compared to non-risk genes and therefore can be distinguished by their aggregated genomic features under the human protein interactome. Applying NETTAG to the latest AD GWAS data, we identified 156 putative AD-risk genes (i.e., *APOE*, *BIN1*, *GSK3B*, *MARK4*, and *PICALM*). We showed that predicted risk genes are: 1) significantly enriched in AD-related pathobiological pathways, 2) more likely to be differentially expressed regarding transcriptome and proteome of AD brains, and 3) enriched in druggable targets with approved medicines (i.e., choline and ibudilast). In summary, our findings suggest that understanding of human pathobiology and therapeutic development could benefit from a network-based deep learning methodology that utilizes GWAS findings under the multimodal genomic analyses.

Introduction

Alzheimer's disease (AD), first described in 1907 by Alois Alzheimer, is the most common type of dementia with gradual cognitive decline and memory loss [1]. AD and AD-related dementias (AD/ADRD) are a major global health challenge and are expected to double in incidence by 2050 [2], affecting 90 million people worldwide [3]. The incidence of AD at the U.S. is expected to double by 2050 [4,5], while the attrition rate for AD clinical trials (2002-2012) is estimated at 99.6% [6]. High-throughput DNA/RNA sequencing technologies have rapidly led to a robust body of genetic and genomic data in multiple national AD genome projects, including the Alzheimer's Disease Sequencing Project (ADSP) [7] and the Alzheimer's Disease Neuroimaging Initiative (ADNI) [8]. Genome-wide association studies (GWAS) have identified over 40 AD susceptibility loci [9–12]. Despite this progress in understanding of AD genetic risk, the predisposition to AD involves a complex, polygenic, and pleiotropic genetic architecture; furthermore, massive genetic and genomic data are not effectively explored for AD drug discovery and development yet.

Recent advances in genetics and systems biology have showed that AD is governed by network-associated molecular determinants (termed disease module) of common endotypes or endophenotypes [13,14]. Approaching AD with a simplistic single-target approach has been demonstrated effective for developing symptomatic therapies but ineffective when attempted for disease modification [13]. Therapeutic approaches by specifically modulating genetic risk genes are essential for development of disease-modifying treatments in AD [14]. A recent study showed that selecting genetically supported targets can double the success rate in clinical development [15].

However, existing data, including genomics, transcriptomics, proteomics, and interactomics (protein-protein interactions), have not yet been fully utilized and integrated to explore the roles of targeted therapeutic development for AD [16–18].

Understanding AD from the point-of-view of how human interactome perturbations underlie the disease is the essence of network medicine [13,14]. The main hypothesis of the AD network medicine is that cellular networks perturbed by genetic variants gradually rewire throughout disease pathogenesis and progression [13,14]. Systematic characterization and identification of underlying pathobiology will serve as a foundation for identifying disease-modifying targets for AD. Integration of the genome, transcriptome, proteome, and the human interactome are essential for such identification. In this study, we presented a **network** topology-based deep learning framework to identify disease-associated **genes** (NETTAG) and drug targets from genetic and genomic discoveries for AD. The key premises of NETTAG are that the disease risk genes: i) exhibit distinct functional characteristics compared to non-risk genes and therefore can be distinguished by their aggregated genomic features, ii) converge to a limited number of pathobiological pathways captured by the human protein-protein interactome, and iii) include multiple AD pathology modulators and potential therapeutic targets.

Results

A network-based deep learning framework

In this study, we presented NETTAG, a network-based deep learning framework to identify risk genes from GWAS and multi-omic findings in AD. NETTG integrates multi-

omics data along with the human protein-protein interaction (PPI) network to infer likely risk genes and potential drug targets impacted by GWAS loci. Specifically, we assembled non-coding GWAS loci effects on expression quantitative trait loci (eQTLs), histone-QTLs, transcription factor binding-QTLs, enhancers and CpG islands, promoter regions, open chromatin, and promoter flanking region from GTEx [19], NIH RoadMap [20], Ensembl Regulatory Build [21], SNPnexus [22] and ENCODE [23] (**Figure 1**). The whole procedure is divided into 4 steps: i) We first utilized a deep learning model to cluster PPIs into multiple functional network modules by capturing its topological structures within the human protein-protein interactome (**Methods**). We then characterized each functional network module by linking its nodes (genes) with protein annotations from the Gene Ontology (GO) knowledgebase [24]; ii) We quantified node's (gene's) scores by integrating its functional similarity with each gene identified with multiple gene regulatory evidences via influencing GWAS loci; iii) We prioritized likely risk genes in AD by their aggregated gene regulatory features; and iv) we prioritized repurposable drugs for potential treatment of AD by evaluating network proximities between Alzheimer's risk genes (alzRGs) and known drug targets under the human protein-protein interactome network model (**Figure 1**).

A gene regulatory landscape of GWAS loci in AD

After mapping AD loci ($p < 1.0 \times 10^{-5}$) from multiple gene regulatory elements (**Methods**), we pinpointed 23 genes with CpG islands (e.g., *APOE*, *PVR*, *STK11*), 19 genes with CTCF binding sites (i.e., *BIN1*, *JPH1*, and *SYK*), 13 genes with enhancers (i.e., *BIN1*, *FARP1*, and *MARK4*), 21 genes with eQTL (i.e., *CD2AP*, *IL6*, and *PVR*), 169 genes with

histone modifications including H3K27ac, H3K27me3, H3K36me3, H3K4me1, H3K4me2, H3K4me3, H3K9ac and H4K20me1 (such as *APOE*, *CKAP5*, *DST*, and *NECTIN2*), 48 genes with open chromatin (i.e., *BIN1*, *CLU* and *INPP5D*), 23 genes with promoter (i.e., *APOE*, *IL6*, and *STK11*), 59 genes with promoter flanking region (i.e., *BIN1*, *CLU* and *MARK4*), and 20 genes (i.e., *BCAM*, *CLU* and *VSNL1*) with transcriptional factor binding site, respectively (**Figure 2A, Supplemental Table S1**). As shown in **Figure 2A**, 69 genes have AD loci with multiple gene regulatory evidences, e.g. *APOE*, *BIN1*, *CLU*, *IL6*, *PTK2B* and *etc*. Specifically, *APOE* loci have regulatory evidences with CpG island (rs429358 and rs7412), histone (rs405509 and rs769449), promoter (rs769449), and promoter flanking regions (rs75627662) (**Supplemental Figure S1A**). Bridging integrator 1 (*BIN1*), another risk factor of LOAD [25][26], is associated with multiple regulatory elements as well, including CTCF binding sites (rs12989701), enhancer (rs10207628), histone (rs6431219, rs10194375, rs72838215, rs10207628), open chromatin (rs6733839), and promoter flanking region (rs10194375). Inositol polyphosphate-5-phosphatase D (*INPP5D*), a LOAD risk gene [10], is associated with open chromatin regulation (rs10933431). Mouse model (5XFAD) study found that expression of *Inpp5d* was elevated in microglia as the disease progressed [27]. Spleen associated tyrosine kinase (*SYK*) (rs1172922) is linked with the CTCF binding site. Activation of SYK boosts inflammation [28] and modulates both A β and tau-induced pathologies [29]. Phosphatidylinositol binding clathrin assembly protein (*PICALM*) loci in AD have multiple regulatory evidences, including histone (rs867611, rs527162, rs639012, and rs17817600), promoter (rs867611) and promoter flanking region (rs10792832). *PICALM* has been suggested to regulate AD pathology with A β

generation and disorder lipid metabolism [30]. Altogether, these bioinformatics analyses highlight the crucial roles of gene regulation involving in various AD GWAS loci, which motivate us to develop NETTAG to infer new gene regulatory variants and putative risk genes in AD using network-based multi-omics evidence aggregation analyses.

Network-based prediction of likely risk genes in AD

According to NETTAG, we first clustered PPIs into functional subnetwork modules using a deep learning framework. We found that the identified subnetwork modules could reflect biological relationships as well (**Figure 2B**). Specifically, proteins with more GO terms tend to have more clusters (**Figure 2B, Supplemental Table S2**). We found that proteins in the same subnetwork module tend to have more shared GO annotations (Wilcoxon signed rank test, $p < 2.2 \times 10^{-16}$, **Methods**). This indicates that network-based fingerprints of module overlays among genes can characterize functional modularities and similarities. We therefore inferred likely risk genes by integrating PPI-derived network modules and multimodal analyses of 9 types of gene regulatory impacts by AD GWAS loci. Specifically, taking CpG island as an example, the predicted score for one particular gene regarding CpG island could approximately estimate its functional overlap (spearman correlation, $r = 0.44$, $p = 1.86 \times 10^{-24}$, **Supplemental Figure S1B, Methods**) with all 23 AD CpG island-linked genes (**Supplemental Table S1**). Finally, we inferred likely risk genes by integrating (summing up) all 9 types of gene regulatory elements. The area under curve (AUC) for receiver operating characteristic curve (ROC) using AD-associated genes collected from AlzGene [31] and DistiLD [32] (**Figure 2C** and

Supplemental Figure 2A) are 0.81 and 0.80 respectively, suggesting reasonable accuracy.

Via NETTAG (**Figure 1**), we identified 156 likely Alzheimer's risk genes (termed alzRGs), such as *APOE*, *APP*, *BIN1*, *FYN*, and *STK11*. Among 156 alzRGs, products (proteins) of 139 alzRGs formed the largest connected component within 294 PPIs (**Figure 3A, Supplemental Table S3**). Via gene and functional enrichment analyses, we found that the NETTAG-predicted alzRGs are significantly enriched by gene regulatory elements (**Figure 3B and Supplemental Figure S3A**) compared to the same number of randomly selected genes with the similar degree distribution in the human interactome network. We assembled AD-associated genes from the GWAS catalog [33], UK Biobank GWAS [34] and DisGeNET with published experimental evidences from animal models and human studies [35]. We found that alzRGs were significantly enriched in all 3 AD-associated gene sets: GWAS catalog (adjusted p-value [q] = 2.25×10^{-7}), UK Biobank GWAS (q = 8.59×10^{-3}), DisGeNET (q = 1.19×10^{-8} , Fisher's exact test, **Supplemental Table S3**). Pathway enrichment analyses [36] showed that alzRGs are significantly enriched in multiple immune pathways (**Supplemental Table S3 and Figure S3B**), including B cell (q = 5.32×10^{-4}), T cell receptor (q = 1.17×10^{-2}), cytokine signaling pathways (IL-2: q = 6.99×10^{-3} , IL-7: q = 1.43×10^{-2} , IL-18: q = 1.65×10^{-2}). In summary, NETTAG achieved reasonable accuracy (AUC=0.81) in predicting likely AD risk genes with diverse functional pathways, including key immune pathways. We next turned to perform multi-omics validation for NETTAG-predicted alzRGs, including single-cell/nuclei transcriptomics in disease-associated microglia (DAM) and astrocyte (DAA) from transgenic mouse and human brains with well-known AD neuropathology.

NETTAG-predicted genes are differentially expressed in AD

We found 95 alzRGs ($p = 2.67 \times 10^{-7}$, Fisher's exact test) that are differentially expressed regarding at least one type of transcriptomics studies in AD. Specifically, 29 ($p = 0.0185$), 67 ($p = 2.36 \times 10^{-3}$), and 39 ($p = 2.96 \times 10^{-7}$) alzRGs are differently expressed genes (DEGs) according to microarray (human AD patients and controls), bulk RNA sequencing (human AD patients and controls), and single cell/nucleus RNA sequencing (AD transgenic mouse model and human postmortem brain samples) analyses, respectively (**Methods**). Nine genes (*ACTL6B*, *ATP2B1*, *EPB41L3*, *ABCA1*, *CPLX2*, *P2RX7*, *PDE1A*, *SLC38A2*, and *VSNL1*) are DEGs based on all three types of differential transcriptomic evidences (**Supplemental Figure S4A**). Block of purinergic receptor P2X 7 (*P2RX7*) was found to reduce tau accumulation in P301S tau transgenic mice [37]. Visinin like 1 (*VSNL1*) is co-expressed with multiple genes involving in molecular mechanisms of AD, including APP [38]. Nineteen genes (*ABCA1*, *APOE*, *BCL3*, *BIN1*, *CKAP5*, *CLU*, *FARP1*, *HSPG2*, *MADD*, *MAPK7*, *MARK4*, *NCS1*, *PICALM*, *PTK2B*, *SPRED2*, *TGFB2*, *TOMM40*, *TOP1*, and *VSNL1*) have been identified by gene regulatory elements and AD GWAS studies as well [33] (**Supplemental Figure S4B**). Microtubule affinity regulating kinase 4 (*MARK4*) is the one linked with most gene regulatory elements, including CpG island (rs28469095), CTCF (rs12463049), enhancer (rs536518226), eQTL (rs8100183), histone (rs9653111 and rs10421247), open chromatin (rs138137383), promoter flanking region (rs10421247 and rs138137383), TF-binding site (rs12463049) (**Supplemental Table S1, Supplemental Figure S4B**). *MARK4* has been suggested as a potential target for AD via binding with

acetylcholinesterase (AChE) inhibitors, such as donepezil (an AChE inhibitor used for patients with AD and other types of dementia) [39]. In addition to *APOE*, *BIN1* and *PICALM*, *CLU* loci are linked with four types of regulatory elements, including histone (rs9331896 and rs1532278), open chromatin (rs2279590), promoter flanking region (rs9331888), and TF-binding sites (rs1532278) (**Supplemental Table S1**, **Supplemental Figure S4B**). Clusterin (*CLU*) is another risk gene for LOAD [40] by involving in multiple AD pathologies, including neuroinflammation, A β accumulation, and lipid metabolism [41].

Among the 67 alzRGs that are differentially expressed based on bulk RNA-seq studies, 42 alzRGs form a subnetwork within human interactome network (**Figure 4A**). And within the 23 top differentially expressed alzRGs such as HOMER2, ABCA1, HSPG2 and etc. ($|\log_2FC| > 1$, $q < 0.05$, **Supplemental Table S4**), one downregulated gene homer scaffold protein 2 (*HOMER2*) ($\log_2FC = -1.65$, $q = 2.39 \times 10^{-3}$, **Supplemental Table S4**) was found to inhibit APP production and secretion of A β peptide together with HOMER3 [42]. ATP binding cassette subfamily A member 1 (*ABCA1*) was one top upregulated gene according to bulk RNA-seq studies ($\log_2FC = 1.21$, $q = 1.37 \times 10^{-3}$, **Supplemental Table S4**), and its mutation was associated with an elevated risk of AD [43]. Heparan sulfate proteoglycan 2 (*HSPG2*) was another top upregulated gene according to bulk RNA-seq studies (**Supplemental Table S4**), it was suggested that people carrying APOE epsilon4 allele have increased risk of AD if carrying HSPG2 A allele at the same time [44].

We have collected 6 sets of proteomics data from transgenic mouse model (**Methods**). Among the 156 predicted AD-associated genes, protein products of 29

genes (i.e., *HSPA5*, *ENO1*, *FERMT2* and *VAV1*) are differentially expressed ($p = 3.98 \times 10^{-7}$, Fisher's exact test, **Supplemental Table S4**). Heat shock protein family A (Hsp70) member 5 (*HSPA5*) was suggested with important roles in tau phosphorylation and a potential target for AD treatment [45]. Study showed that oxidative inactivation of enolase 1 (*ENO1*) could accelerate development of AD from mild cognitive impairment (MCI) [46]. FERM domain containing kindlin 2 (*FERMT2*) is identified as AD risk genes by GWAS studies [10]. It is also suggested that *FERMT2* could modulate APP metabolism and A β formation, therefore linking its mechanism association with AD [47]. Another study in mouse model found that targeting vav guanine nucleotide exchange factor 1 (*VAV1*) could rescue neuron death by inhibiting JNK signaling pathway [48].

alzRGs are differentially expressed in AD-associated microglia and astrocytes

Neuroinflammation plays crucial roles in pathogenesis and disease progression of AD [49]. We found that predicted alzRGs are significantly enriched by multiple immune pathways, including B cell receptor, T cell receptor, and cytokines (IL-2, IL-7 and IL18) signaling pathways (**Supplemental Table S3, Supplemental Figure S3B**). We next turned to investigate how inflammatory pathways were impacted by predicted alzRGs using disease-associated microglia (DAM) and disease-associated astrocytes (DAA) as examples. We found that 14 alzRGs were differentially expressed ($|\log_2FC| > 0.25$, $q < 0.05$) in DAM from one 5XFAD mouse model derived single-cell RNA-seq dataset (one-sided t-test: statistic = 33.85, $p = 9.76 \times 10^{-183}$, **Supplemental Figure S4C**) and one 5XFAD mouse model derived single-nucleus RNA-seq dataset (one-sided t-test: statistic = 23.52, $p = 1.88 \times 10^{-108}$, **Figure 4B**). For DAA, 25 genes are differentially

expressed ($|\log_2FC| > 0.25$, $q < 0.05$) across 3 snRNA-seq datasets from human postmortem brains with different brain regions, including prefrontal cortex ($p < 1.0 \times 10^{-3}$, **Figure 4C**), entorhinal cortex ($p = 2.33 \times 10^{-49}$, **Figure 4D**), and super frontal gyrus ($p = 8.30 \times 10^{-27}$) (**Supplemental Figure S4D, Supplemental Table S4, Methods**). Among 39 differentially expressed alzRGs in DAM or DAA, 28 alzRGs form a subnetwork with human interactome network (**Figure 4E**). Activating transcriptional factor 3 (*ATF3*) overexpressed in DAM (one top DEGs $\log_2FC = 0.70$, $q = 2.21 \times 10^{-21}$ with 5XFAD mouse models, **Supplemental Table S4**), was also observed with elevated expression in another mouse model study [50]. The study also suggested that the elevated expression level of *ATF3* is positively correlated with A β accumulation [50]. Microtubule associated protein RP/EB family member 2 (*MAPRE2*) overexpressed in DAA (one top DEGs $\log_2FC = 0.51$, $q = 7.03 \times 10^{-20}$ with human postmortem brain tissues, **Supplemental Table S4**) was identified as a new AD-associated gene based on GWAS from ADNI cohort [51].

Discovery of high-confidence risk genes in AD

We next turned to identify high-confidence risk genes in AD via combining multiple factors: 1) high-confidence risk genes are top predicted by NETTAG (156 alzRGs); 2) high-confidence risk genes are supported with at least 3 types of multi-omics evidences (**Supplemental Table S4**); 3) high-confidence risk genes have not previously been identified by GWAS Catalog. In total we have identified 25 high-confidence AD risk genes, e.g., *BACE1*, *CDK5*, *CPLX2*, *FYN*, *MAPKAPK2*, *MEF2D*, and etc. (**Figure 5, Supplemental Table S4**). Myocyte enhancer factor 2D (*MEF2D*), an alzRG with

currently non-existing AD-associated DNA regulatory or GWAS evidence, is the one with the highest predicted score. It is differently expressed regarding to both single-nucleus RNA seq and microarray analyses (**Figure 5, Supplemental Figure S4E, Supplemental Table S4**). One study found that protocatechuic acid could rescue a cell model from okadaic acid-induced cytotoxicity (tau hyperphosphorylation) by modulating Akt/GSK-3 β /MEF2D pathway and exhibit neuroprotective effects which may suggest itself being beneficial for AD treatment [52]. Complexin 2 (*CPLX2*) is our second top predicted alzRG with currently non-existing AD-associated DNA regulatory or GWAS evidence. It is identified as AD-associated with five types of evidence and differently expressed according to both transcriptome (DAA) and proteome studies (**Figures 5, Supplemental Figure S4F, Supplemental Table S4**). Experiment with hippocampus from 3x-Tg AD mice showed abnormal lower proteomic levels of CPLX1 and CPLX2 [53] which is consistent with the transcriptomic behavior in DAA. Further decreased expression level was observed after exposure to copper, which suggested that CPLX2 together with CPLX1 may be the key factors in chronic copper overexposure-induced memory damage in AD [53].

Discovery of repurposable drugs via targeting risk genes

Among the 156 predicted alzRGs, 38 proteins have been identified as known drug targets ($p = 8.78 \times 10^{-4}$, Fisher's exact test, **Supplemental Table S4** and **Figure 5**). In total, 9 targets (i.e., APP, ATF2, BACE1, CDK5, FYN, GSK3B, MARK4, MKL1 and PTK2B, **Supplemental Table S4**) have been widely investigated as therapeutic approaches for treating AD. FYN proto-oncogene, Src family tyrosine kinase (FYN),

which contributes to A β production and tau phosphorylation, has been suggested as one potential target for AD [54]. Glycogen synthase kinase 3 beta (GSK3B) has been found in hyperphosphorylation of tau and A β production [55]. Study has found that one GSK3B inhibitor thiadiazolidinone could decrease tau phosphorylation and improve neuronal survival [55]. Beta-secretase 1 (BACE1), a β -secretase enzyme involving in A β peptide generation, has been demonstrated a promising target in AD [56]. We next turned to identify repurposable drugs by specifically targeting alzRGs (**Figure 3A**).

Using our well-established network proximity approaches [57], we computationally identified 118 candidate drugs using z-score < -2 and q < 0.05 from total 2,938 U.S. FDA-approved or clinically investigational drugs (**Supplemental Table S5, Methods**). As shown in **Figure 6A**, we grouped these top predicted 118 candidate drugs into 14 pharmacological categories based the first-level of the Anatomical Therapeutic Chemical (ATC) code. Choline, a nutrient found in many vitamins, is our 5th ranked predicted drug (**Supplemental Table S5**). Experiments with APP/PS1 mouse models showed that dietary of choline could reduce A β production and improve spatial memory by suppressing overactivation of DAM [58]. Choline does not directly target any alzRGs; yet, choline's targets interacts with multiple protein products of predicted alzRGs, including APP, BIN1, CDK5, and FYN (**Figure 6B**). Ibudilast, an anti-inflammatory drug used to attenuate multiple sclerosis, is another top prediction (**Supplemental Table S5**). Ibudilast inhibited pro-inflammatory cytokine production and blocked neuroinflammation to prevent A β -produced cognitive impairment [59]. Mechanistically, ibudilast's targets (i.e., PDE3A, PDE4B and PDE4D) have physical interactions with proteins encoded by several predicted alzRGs (i.e., BIN1, FYN, GSK3B) (**Figure 6C**). In

summary, risk genes identified by NETTAG offer potential drug targets for AD therapeutic discovery, including drug repurposing (such as ibudilast and choline). The predicted repurposable drugs offer potential candidates for future preclinical and clinical validations. Deferoxamine, a FDA-approved iron-chelating agent for treatment of iron overdose or hemochromatosis, is the top second predicted candidate drug (**Supplemental Table S5**). Treatment with deferoxamine alleviated disturbed iron homeostasis and reduced APP phosphorylation in a transgenic mouse model [60]. Mechanistically, several deferoxamine's targets (HIF1A, RGS4, RGS19 and GMNN) have physical protein interactions with multiple protein products of several predicted alzRGs, such as APOE, BIN1, CLU and GSK3B (**Supplemental Figure S4G**).

Discussion

In this study, we presented a deep-learning framework that integrates multi-omics knowledges to infer putative risk genes in AD. To avoid “black-box” deep learning models, we utilized human protein-protein interactome network to make NETTAG more transparent and interpretable in human gene discovery. For example, (**Methods**), there are 16,720 proteins in the used human protein interactome according to the GO knowledgebase [24] (**Supplemental Table S2**). We found that 97% (16,214) of annotated proteins in the human interactome have multiple GO terms, ranging from 2 to around 200 (**Supplemental Table S2, Supplemental Figure S5A**). We further found that the deep learning-predicted subnetwork modules were highly correlated with protein functions (**Figure 2B**). We performed additional spearman (r) correlation analysis to evaluate the correlation between predicted score (cumulative overlays of

divided subnetwork modules) and cumulative overlays of protein functions by considering each gene regulatory element. We found that deep learning-predicted scores showed significant correlations across all gene regulatory elements (**Supplemental Figure S1B**). For example, CpG island ($r = 0.44$, $p = 1.86 \times 10^{-24}$, enhancer ($r = 0.34$, $p = 9.07 \times 10^{-13}$), histone ($r = 0.50$, $p = 6.49 \times 10^{-37}$), and TF-binding site ($r = 0.36$, $p = 3.50 \times 10^{-17}$), show strong correlations; while promoter flanking region ($r = 0.26$, $p = 2.60 \times 10^{-12}$), open chromatin ($r = 0.22$, $p = 2.91 \times 10^{-10}$), promoter ($r = 0.20$, $p = 2.42 \times 10^{-15}$), CTCF ($r = 0.09$, $p = 1.45 \times 10^{-2}$) and eQTL ($r = 0.14$, $p = 8.62 \times 10^{-10}$) show weak correlations. These observations showed that CpG island, enhancer, histone and TF-binding sites play more important gene regulatory roles of GWAS loci in AD compared to promoter flanking region, open chromatin, promoter, CTCF and eQTL.

We performed ROC analyses to evaluate the performance of NETTAG. We found that the AUCs for predicted scores by integrating nine DNA regulatory elements are 0.23 (AlzGene) and 0.23 (DistiLD) higher than the averages of AUCs if considering each single regular element alone (**Supplemental Figures S2B-2D**). This suggested integrating multi-genomic evidences could infer like risk genes with a higher accuracy. We acknowledged several potential limitations in the current study. For example, incompleteness of the human protein-protein interactome and GWAS loci by lack of population size may influence the model performance. In addition, the genome regulatory elements used in this study is not brain-specific. More brain-specific functional genomics data should be integrated in the near future [61]. If considering all protein pairs sharing at least one common GO annotations, we found that shortest path-based distances among those protein pairs are 2 (53.97%), 3 (39.05%) and 4 (5.05%)

in decreasing order (**Supplemental Figure S5B**). Therefore, in NETTAG, we have 2 hidden layers to indirectly aggregate features from second order neighbors. We postulate that reformulating the model which could sampling second and third order neighbors directly may improve the model performance further.

In summary, we established a deep learning framework (NETTAG) that incorporates multi-genomics knowledges along with human PPIs to infer novel AD-associated genes. We showed that predicted genes are enriched with drug targets, differently expressed in disease-associated immune cell subtypes, and most importantly significantly AD-associated. We believe that the NETTAG presented here, if broadly applied, would significantly catalyze innovation in AD drug discovery.

Methods and Materials

Construction of genetic features

In this study, we collected 1,047,489 SNPs across multiple genetic traits from GWAS catalog [33], such as Alzheimer's disease, cerebral amyloid deposition measurement. Next, we performed web server SNPnexus [22] to annotate all SNPs in human genome (GRCh38) and collected the regulatory elements information from five databases, including GpG Islands [22], Ensembl Regulatory Build [21], ENCODE [23], the Genotype-Tissue Expression (GTEx) portal [19] and Roadmap [20]. Finally, nine regulatory elements (histone, open chromatin, CpG Island, TF, CTCF, eQTL, enhancer, promoter and promoter flanking region) were used as features to evaluate AD disease-associated genes. To be more specific, **step 1**: for SNPs with respect to each regulatory

elements, e.g., CpG island, we merge them with SNPs curated by GWAS Catalog with AD as the mapped traits. **Step 2:** For each AD related SNPs, the corresponding genes were identified by the “MAPPED GENE(S)” column as provided by GWAS Catalog. For SNPs with no mapped genes, if there were any reported genes (REPORTED GENE(S) column in GWAS Catalog) associated with this SNP, we then map the SNP to its reported genes. For ENCODE [23] and Ensembl Regulatory Build [21] databases, we only consider epigenomes from brain and neuron tissues and normal karyotype. The specific epigenomes with brain and neurons for each database are presented in **Supplemental Table S1**, separately. The final mapped genes for each regulatory element are provided in **Supplemental Table S1**.

Building Human Protein-protein interactome

To build the comprehensive human interactome from the most contemporary data available, we collect 18 commonly used PPI databases with experimental evidence which mainly include: (i) binary PPIs tested by high-throughput yeast-two-hybrid (Y2H) systems [16]; (ii) kinase-substrate interactions; (iii) signaling networks; (iv) binary PPIs from three-dimensional protein structures; (v) protein complexes data; and (vi) carefully literature-curated PPIs. In total, 351,444 PPIs connecting 17,706 unique proteins are free available at <https://alzgps.lerner.ccf.org>. And in this study, we only consider its largest connected components which includes 17,456 proteins and 336,549 PPIs. All details are provided in our recent studies [62,63].

Description of NETTAG

NETTAG involves 3 steps. **Step 1**: we build up a graph neural network (GNN) model to capture PPI's topology structure and establish the appropriate overlapping clustering. The GNN model is motivated by NOCD-G [64] which is one GNN-based overlapping community detection framework. And in NETTAG, we made 2 modifications with respect to NOCD-G [64].

Modification 1: The model architecture used in NETTAG is defined below (Eq.1):

$$F := \text{ReLU}[GCN_{\Theta_2}^K(A, \text{ReLU}[GCN_{\Theta_1}^2(A, X)])], \text{ with } GCN_{\Theta_l}^2 = \sum_{k=0}^2 \hat{A}^k X W^{k,l} \quad (1)$$

Here: $\hat{A}^0 = I$, $\hat{A}^1 = \hat{A} = D^{-1/2} A D^{-1/2}$, $\hat{A}^2$ is the elementwise square of $\hat{A}$ and the sub/super-script l is the layer index. A is the adjacency matrix of the PPI, D is the corresponding diagonal degree matrix. X in generate denotes the node feature matrix, and here we set $X = A$ as implemented by NOCD-G [64]. In classical graph convolution network (GCN) models [65], nodes with various degrees share the same weight matrix. Therefore in NETTAG, coupling $\hat{A}^2$ into the model is to overcome this, i.e., to assign different weight matrices for nodes with high and low degrees. To evaluate its effects, we compare the predicted genes generated by NETTAG with (test group) and without (control group) this extra 2nd order term. Each group has 10 experiments with fixed seeds. We find that predicted alzRGs generated by models from test group are more significantly AD-associated compared with those from the control group (paired seeds) with multiple sources, i.e., GWAS catalog (q without / with 2nd order term: $3.07 \times 10^{-3} \pm 7.92 \times 10^{-3}$ / $3.53 \times 10^{-5} \pm 1.07 \times 10^{-4}$, **Supplemental Figures S5C**), UK Biobank GWAS (q without / with 2nd order term: $1.22 \times 10^{-2} \pm 1.53 \times 10^{-2}$ / $7.37 \times 10^{-3} \pm 3.92 \times 10^{-3}$, **Supplemental Figures S5D**), and DisGeNET (q without / with 2nd order term: 6.30×10^{-3}

$^5 \pm 1.43 \times 10^{-4} / 8.65 \times 10^{-7} \pm 2.67 \times 10^{-6}$, **Supplemental Figures S5E**). Finally, the dimension of the final output layer interprets the clustering number.

The output matrix F is then feed into Bernoulli-Poisson model [66] (**Eq.2.1**) to learn PPI's topology. The output matrix F has N (number of total nodes in PPI) rows and C (clustering numbers) columns. Each specific row (vector) denotes the node's weights for being assigned to each cluster. Therefore, we can interpret the loss as follows: if two nodes have multiple commonly shared clusters ($F_u \cdot F_v^T$ is large), then there should exist an edge connecting each other ($1 - \exp(-F_u \cdot F_v^T)$ is close to 1) and vice versa.

$$(Loss) L = - \sum_{(u,v) \in E} \log(1 - \exp(-F_u \cdot F_v^T)) + \sum_{(u,v) \notin E} F_u \cdot F_v^T \quad (2.1)$$

$$(Loss) L = - \frac{1}{|(u, v) \in E|} \sum_{(u,v) \in E} \log(1 - \exp(-F_u \cdot F_v^T)) + \frac{1}{|(u, v) \notin E|} \sum_{(u,v) \notin E} F_u \cdot F_v^T \quad (2.2)$$

Modification 2: The PPI network is a highly sparse network which implies number of connected edges is far less than that of non-connected edges (For our PPI, the ratio between number of connected edges over number of non-connected edges = 1 / 448). To address this imbalanced training problem, the authors in [64] first uniformly subsampled certain amounts of connected edges and then subsampled the equal amounts of non-connected edges to train the averaged loss (**Eq 2.2**). Instead of uniformly sampling edges directly, we first group nodes according to their degrees into multiple bins. And in each training iteration, we first uniformly subsample same amounts of nodes from each bin, then we extract an adjacency matrix A_{sub} which compromising

only selected nodes. In this way, we can keep the topology similarities among sampled subgraphs from different iterations. Next, we use the connected and non-connected edges in A_{sub} to compute the train loss, and the rest connected and non-connected edges in $A \setminus A_{sub}$ for test loss evaluation. With this graph subsampling scheme, we find that we are capable to maintain similar connected and non-connected training edge percentages (**Supplemental Figure S5F**).

After learning the clustering affinity matrix F , we used the threshold defined in [67] as the cutoff for F to determine the node (gene) clustering membership (**Eq. 3**).

$$F_{ij}^* = \begin{cases} 1 & \text{if } F_{ij} \geq \sqrt{-\log(1 - 1/N)} \\ 0 & \text{Otherwise} \end{cases} \quad (3)$$

Step 2 of NETTAG: After clustering the whole PPI into overlapping communities, we next score each node with respect to each gene regulatory element. The detailed steps are explained in **Supplemental Materials**. Each node (gene) is firstly scored according to its clustering overlap with regulatory element evidenced genes. Next, we construct the background distribution by performing 1,000 random experiments to evaluate the statistical significance of node score computed in previous step. Finally, we evaluate the integrated node score by summing up all scores with respect to each gene regulatory element.

Step 3 of NETTAG: After inferring the nodes' score, finally we extract the disease module by considering genes which are inferred only significantly associated with AD. In detail, we collect all positive gene scores $\{S_i > 0\}$, and compute the mean μ , standard

deviation σ and the z score. We consider genes with z-scores great than 2.32 (p value = 0.01) as the predicted AD associated genes, and map those genes to the background PPI network to generate disease module.

Network proximity for drug prediction

We assembled drugs from the DrugBank database relating 2,938 compounds [68]. To predict drugs with extracted disease module from NETTAG, we adopted the closest-based network proximity measure [57] as below.

$$d_{closest}(X, Y) = \frac{1}{\|X\| + \|Y\|} \left(\sum_{x \in X} \min_{y \in Y} d(x, y) + \sum_{y \in Y} \min_{x \in X} d(x, y) \right) \quad (4)$$

where $d(x, y)$ is the shortest path length between gene x and y from gene sets X and Y , respectively. In our work, X denotes the disease module from NETTAG, Y denotes the drug targets (gene set) for each compound. To evaluate whether such proximity was significant, the computed network proximity is transferred into z score form as shown below:

$$Z_{d_{closest}} = \frac{d_{closest} - \mu_d}{\sigma_d} \quad (5)$$

Here, μ_d and σ_d are the mean and standard deviation of permutation test with 1,000 random experiments. In each random experiment, two random subnetworks X_r and Y_r are constructed with the same numbers of nodes and degree distribution as the given 2 subnetworks X and Y , separately, in the PPI network.

Agreement between clustered subnetwork modules and protein functions

With respect to each clustered subnetwork modules, considering its all-possible protein-protein pairs, we count numbers of protein-protein pairs that share at least one common GO annotations and numbers of protein-protein pairs that share no common GO annotations. Then we build up a paired statistical test with null hypothesis as proteins in the same clustered subnetwork modules have no protein functional similarities, and alternative hypothesis as proteins in the same clustered subnetwork module possess common protein functions. We apply Wilcoxon signed rank test with R.

Differentially expressed gene (DEG) with transcriptome and proteome analyses

The transcriptome analyses are performed based on microarray, bulk RNA-seq, and single-cell/nucleus (sc/sn) RNA-seq datasets. We utilize three sets of human brain microarray transcriptome data collected from late-stage AD and control donors. They are available from Gene Expression Omnibus (<https://www.ncbi.nlm.nih.gov/geo/>) database under accession numbers: GSE29378 (31 LAD and 32 controls) [69], GSE48350 (42 AD and 173 controls) [70], and GSE84422 (328 LAD and 214 controls) [71]. We also include human brain bulk RNA-seq transcriptome data collected from hippocampus region of late-stage AD and control donors with three studies including 4 LAD versus 4 controls [72], 6 LAD versus 6 controls [73], and 20 LAD versus 10 controls [74]. The DEGs for all microarray and bulk RNA-seq datasets have been analyzed in recent developed AD knowledgebase AlzGPS [75]. The complete sc/sn RNA-seq datasets used for DEG analyses in this study are available from Gene Expression Omnibus (<https://www.ncbi.nlm.nih.gov/geo/>) database under accession numbers: GSE98969 [76], GSE140511 [77], GSE147528 [78], GSE138852 [79] and

GSE157827 [80]. We have performed the corresponding differently expressed gene analyses in our previous study [63] for datasets GSE98969, GSE140511, GSE147528 and GSE138852. The more detailed bioinformatics analysis for newly add dataset GSE157827 is described in **Supplemental Materials**. The DEG analysis between DAM and non-DAM is based on mouse sc-RNA seq datasets GSE98969 and mouse sn-RNA seq dataset GSE140511. The union of DEGs between DAM and homeostasis microglia from both datasets are used for transcriptome analysis in this study. The DEG analysis between DAA and non-DAA is based on the rest three human sn-RNA seq datasets GSE147528, GSE138852 and GSE157827. The union of DEGs between DAA and homeostasis astrocytes from all these three human datasets are used for transcriptome analysis in this study. For all DEGs generated from sc/sn RNA-seq datasets, we apply uniform criterion with $q < 0.05$ and $|\log_2FC| \geq 0.25$. Proteome analyses are based on six mouse model datasets including 7- and 10-months ADLP mouse models (JNPL3 mouse model cross with 5XFAD mouse model) [81], 7- and 10-months 5XFAD mouse models [81], 12 months 5XFAD mouse model [82] and 12 months hAPP mouse model [82].

Collections of AD seed genes

We collect AD seed genes from two AD-included knowledgebase, including AlzGene and DistiLD. AlzGene [31] collected AD-associated genes via genetic association studies. Thirty-two genes supported by genetic evidences are collected from AlzGene. DistiLD made the existing GWAS studies easier for accessing disease-associated SNPs and genes [32]. We collected 19 genes ($p < 5.0 \times 10^{-8}$) with AD GWAS from DistiLD. The

complete lists of genes we used for ROC analyses are presented in **Supplemental Table S2**.

Enrichment Analysis

All pathway and disease enrichment analyses were conducted using WikiPathways [83], GWAS Catalog 2019 [33], UK Biobank GWAS v1 [34] and DisGeNET [35] from Enrichr [84], respectively.

Gene Ontology

All proteins' gene ontology annotations (human, gaf-version 2.1) are extracted from The Gene Ontology (GO) Knowledgebase [24]. For our PPI with 17,706 proteins, there are 16,736 proteins with total 268,241 GO annotations.

Software availability

All codes written for and used in this study are available from <https://github.com/ChengF-Lab/NETTAG>.

Supplementary information is available in the online version of the paper.

Competing interests.

The authors have declared no competing interest.

Author Contributions

F.C. conceived the study. J.X. performed all experiments and data analysis. Y.H. performed multi-omics data analyses. Y.Z. and M.H. interpreted the data analysis. J.X. and F.C. drafted the manuscript and critically revised the manuscript. All authors critically revised and gave final approval of the manuscript.

Acknowledgments

Funding: This work was supported by the National Institute of Aging (NIA) of the National Institutes of Health (NIH) under Award Number R01AG066707, U01AG073323, 1R56AG074001-01, 3R01AG066707-01S1, and 3R01AG066707-02S1 to F.C. This work was supported in part by the National Heart, Lung, and Blood Institute of the NIH under Award Number R00HL138272 and by the VeloSano Pilot Program (Cleveland Clinic Taussig Cancer Institute) to F.C.

References

1. Zilka N, Novak M. The tangled story of Alois Alzheimer. Bratisl Lek Listy. 2006;107:343–5.
2. Hebert LE, Beckett LA, Scherr PA, Evans DA. Annual incidence of Alzheimer disease in the United States projected to the years 2000 through 2050. Alzheimer Dis Assoc Disord. 2001;15:169–73.
3. Alzheimer's Association. 2016 Alzheimer's disease facts and figures. Alzheimers Dement. 2016;12:459–509.
4. Kodamullil AT, Zekri F, Sood M, Hengerer B, Canard L, McHale D, et al. Trial watch: Tracing investment in drug development for Alzheimer disease. Nat Rev Drug Discov. 2017;16:819.

5. Be open about drug failures to speed up research [Internet]. [cited 2020 Aug 17]. Available from: <https://www.nature.com/articles/d41586-018-07352-7>
6. Cummings JL, Morstorf T, Zhong K. Alzheimer's disease drug-development pipeline: few candidates, frequent failures. *Alzheimer's Research & Therapy*. 2014;6:37.
7. Beecham GW, Bis JC, Martin ER, Choi S-H, DeStefano AL, van Duijn CM, et al. The Alzheimer's Disease Sequencing Project: Study design and sample selection. *Neurol Genet*. 2017;3:e194.
8. Petersen RC, Aisen PS, Beckett LA, Donohue MC, Gamst AC, Harvey DJ, et al. Alzheimer's Disease Neuroimaging Initiative (ADNI). *Neurology*. 2010;74:201–9.
9. Cuyvers E, Sleegers K. Genetic variations underlying Alzheimer's disease: evidence from genome-wide association studies and beyond. *Lancet Neurol*. 2016;15:857–68.
10. Lambert J-C, Ibrahim-Verbaas CA, Harold D, Naj AC, Sims R, Bellenguez C, et al. Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease. *Nat Genet*. 2013;45:1452–8.
11. Jung YJ, Kim YH, Bhalla M, Lee SB, Seo J. Genomics: New Light on Alzheimer's Disease Research. *Int J Mol Sci*. 2018;19:3771.
12. Jansen IE, Savage JE, Watanabe K, Bryois J, Williams DM, Steinberg S, et al. Genome-wide meta-analysis identifies new loci and functional pathways influencing Alzheimer's disease risk. *Nat Genet*. 2019;51:404–13.
13. Fang J, Pieper AA, Nussinov R, Lee G, Bekris L, Leverenz JB, et al. Harnessing endophenotypes and network medicine for Alzheimer's drug repurposing. *Medicinal Research Reviews* [Internet]. [cited 2020 Aug 17];n/a. Available from: <https://onlinelibrary.wiley.com/doi/abs/10.1002/med.21709>
14. Swarup V, Hinz FI, Rexach JE, Noguchi K, Toyoshiba H, Oda A, et al. Identification of evolutionarily conserved gene networks mediating neurodegenerative dementia. *Nat Med*. 2019;25:152–64.
15. Nelson MR, Tipney H, Painter JL, Shen J, Nicoletti P, Shen Y, et al. The support of human genetic evidence for approved drug indications. *Nat Genet*. 2015;47:856–60.
16. Luck K, Kim D-K, Lambourne L, Spirohn K, Begg BE, Bian W, et al. A reference map of the human binary protein interactome. *Nature*. 2020;580:402–8.
17. Cheng F, Lu W, Liu C, Fang J, Hou Y, Handy DE, et al. A genome-wide positioning systems network algorithm for in silico drug repurposing. *Nat Commun*. 2019;10:3476.
18. Cheng F, Zhao J, Wang Y, Lu W, Liu Z, Zhou Y, et al. Comprehensive characterization of protein–protein interactions perturbed by disease mutations. *Nat Genet*. 2021;53:342–53.

19. Lonsdale J, Thomas J, Salvatore M, Phillips R, Lo E, Shad S, et al. The Genotype-Tissue Expression (GTEx) project. *Nat Genet.* 2013;45:580–5.
20. Bernstein BE, Stamatoyannopoulos JA, Costello JF, Ren B, Milosavljevic A, Meissner A, et al. The NIH Roadmap Epigenomics Mapping Consortium. *Nat Biotechnol.* 2010;28:1045–8.
21. Zerbino DR, Wilder SP, Johnson N, Juettemann T, Flicek PR. The Ensembl Regulatory Build. *Genome Biology.* 2015;16:56.
22. Oscanova J, Sivapalan L, Gadaleta E, Dayem Ullah AZ, Lemoine NR, Chelala C. SNPnexus: a web server for functional annotation of human genome sequence variation (2020 update). *Nucleic Acids Res. Oxford Academic;* 2020;48:W185–92.
23. Dunham I, Kundaje A, Aldred SF, Collins PJ, Davis CA, Doyle F, et al. An integrated encyclopedia of DNA elements in the human genome. *Nature.* 2012;489:57–74.
24. Mi H, Muruganujan A, Ebert D, Huang X, Thomas PD. PANTHER version 14: more genomes, a new PANTHER GO-slim and improvements in enrichment analysis tools. *Nucleic Acids Research.* 2019;47:D419–26.
25. Franzmeier N, Rubinski A, Neitzel J, Ewers M. The BIN1 rs744373 SNP is associated with increased tau-PET levels and impaired memory. *Nat Commun.* 2019;10:1766.
26. Tan M-S, Yu J-T, Tan L. Bridging integrator 1 (BIN1): form, function, and Alzheimer's disease. *Trends in Molecular Medicine.* 2013;19:594–603.
27. Tsai AP, Lin PB-C, Dong C, Moutinho M, Casali BT, Liu Y, et al. INPP5D expression is associated with risk for Alzheimer's disease and induced by plaque-associated microglia. *Neurobiology of Disease.* 2021;153:105303.
28. Ghosh S, Geahlen RL. Stress Granules Modulate SYK to Cause Microglial Cell Dysfunction in Alzheimer's Disease. *EBioMedicine.* 2015;2:1785–98.
29. Paris D, Ait-Ghezala G, Bachmeier C, Laco G, Beaulieu-Abdelahad D, Lin Y, et al. The Spleen Tyrosine Kinase (Syk) Regulates Alzheimer Amyloid- β Production and Tau Hyperphosphorylation. *J Biol Chem.* 2014;289:33927–44.
30. Xu W, Tan L, Yu J-T. The Role of PICALM in Alzheimer's Disease. *Mol Neurobiol.* 2015;52:399–413.
31. Bertram L, McQueen MB, Mullin K, Blacker D, Tanzi RE. Systematic meta-analyses of Alzheimer disease genetic association studies: the AlzGene database. *Nat Genet.* 2007;39:17–23.
32. Pallejà A, Horn H, Eliasson S, Jensen LJ. DistiLD Database: diseases and traits in linkage disequilibrium blocks. *Nucleic Acids Res.* 2012;40:D1036–40.

33. Buniello A, MacArthur JAL, Cerezo M, Harris LW, Hayhurst J, Malangone C, et al. The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. *Nucleic Acids Res.* 2019;47:D1005–12.
34. Sudlow C, Gallacher J, Allen N, Beral V, Burton P, Danesh J, et al. UK Biobank: An Open Access Resource for Identifying the Causes of a Wide Range of Complex Diseases of Middle and Old Age. *PLOS Medicine. Public Library of Science*; 2015;12:e1001779.
35. Piñero J, Bravo À, Queralt-Rosinach N, Gutiérrez-Sacristán A, Deu-Pons J, Centeno E, et al. DisGeNET: a comprehensive platform integrating information on human disease-associated genes and variants. *Nucleic Acids Res.* 2017;45:D833–9.
36. Slenter DN, Kutmon M, Hanspers K, Riutta A, Windsor J, Nunes N, et al. WikiPathways: a multifaceted pathway database bridging metabolomics to other omics research. *Nucleic Acids Research.* 2018;46:D661–7.
37. Ruan Z, Delpech J-C, Venkatesan Kalavai S, Van Enoo AA, Hu J, Ikezu S, et al. P2RX7 inhibitor suppresses exosome secretion and disease phenotype in P301S tau transgenic mice. *Molecular Neurodegeneration.* 2020;15:47.
38. Lin C-W, Chang L-C, Tseng GC, Kirkwood CM, Sibille EL, Sweet RA. VSNL1 Co-Expression Networks in Aging Include Calcium Signaling, Synaptic Plasticity, and Alzheimer's Disease Pathways. *Front Psychiatry [Internet]. Frontiers*; 2015 [cited 2021 Aug 9];0. Available from: <https://www.frontiersin.org/articles/10.3389/fpsy.2015.00030/full>
39. Shamsi A, Anwar S, Mohammad T, Alajmi MF, Hussain A, Rehman MT, et al. MARK4 Inhibited by AChE Inhibitors, Donepezil and Rivastigmine Tartrate: Insights into Alzheimer's Disease Therapy. *Biomolecules. Multidisciplinary Digital Publishing Institute*; 2020;10:789.
40. Sims R, van der Lee SJ, Naj AC, Bellenguez C, Badarinarayan N, Jakobsdottir J, et al. Rare coding variants in PLCG2, ABI3, and TREM2 implicate microglial-mediated innate immunity in Alzheimer's disease. *Nat Genet.* 2017;49:1373–84.
41. Uddin MdS, Kabir MdT, Begum MstM, Islam MdS, Behl T, Ashraf GM. Exploring the Role of CLU in the Pathogenesis of Alzheimer's Disease. *Neurotox Res [Internet].* 2020 [cited 2021 Aug 19]; Available from: <https://doi.org/10.1007/s12640-020-00271-4>
42. Parisiadou L, Bethani I, Michaki V, Krousti K, Rapti G, Efthimiopoulos S. Homer2 and Homer3 interact with amyloid precursor protein and inhibit A β production. *Neurobiology of Disease.* 2008;30:353–64.
43. Nordestgaard LT, Tybjærg-Hansen A, Nordestgaard BG, Frikke-Schmidt R. Loss-of-function mutation in ABCA1 and risk of Alzheimer's disease and cerebrovascular disease. *Alzheimer's & Dementia.* 2015;11:1430–8.

44. Iivonen S, Helisalmi S, Mannermaa A, Alafuzoff I, Lehtovirta M, Soininen H, et al. Heparan sulfate proteoglycan 2 polymorphism in Alzheimer's disease and correlation with neuropathology. *Neurosci Lett*. 2003;352:146–50.
45. Booth L, Roberts JL, Dent P. HSPA5/Dna K May Be a Useful Target for Human Disease Therapies. *DNA Cell Biol*. 2015;34:153–8.
46. Butterfield DA, Poon HF, St. Clair D, Keller JN, Pierce WM, Klein JB, et al. Redox proteomics identification of oxidatively modified hippocampal proteins in mild cognitive impairment: Insights into the development of Alzheimer's disease. *Neurobiology of Disease*. 2006;22:223–32.
47. Chapuis J, Flaig A, Grenier-Boley B, Eysert F, Pottiez V, Deloison G, et al. Genome-wide, high-content siRNA screening identifies the Alzheimer's genetic risk factor FERMT2 as a major modulator of APP metabolism. *Acta Neuropathol*. 2017;133:955–66.
48. He B, Chen W, Zeng J, Tong W, Zheng P. MicroRNA-326 decreases tau phosphorylation and neuron apoptosis through inhibition of the JNK signaling pathway by targeting VAV1 in Alzheimer's disease. *J Cell Physiol*. 2020;235:480–93.
49. Heneka MT, Carson MJ, El Khoury J, Landreth GE, Brosseon F, Feinstein DL, et al. Neuroinflammation in Alzheimer's Disease. *Lancet Neurol*. 2015;14:388–405.
50. Zhu M, Allard JS, Zhang Y, Perez E, Spangler EL, Becker KG, et al. Age-Related Brain Expression and Regulation of the Chemokine CCL4/MIP-1A in APP/PS1 Double-Transgenic Mice. *J Neuropathol Exp Neurol*. Oxford Academic; 2014;73:362–74.
51. Kim S, Nho K, Risacher SL, Shen L, Shaw LM, Trojanowski JQ, et al. P4-008: Mapre2 as a novel Alzheimer's disease target gene from gwas of CSF amyloid beta 1-42, tau and hyperphosphorylated tau in the ADNI cohort. *Alzheimer's & Dementia*. 2015;11:P767–8.
52. Huang L, Zhong X, Qin S, Deng M. Protocatechuic acid attenuates β -secretase activity and okadaic acid-induced autophagy via the Akt/GSK-3 β /MEF2D pathway in PC12 cells. *Molecular Medicine Reports*. Spandidos Publications; 2020;21:1328–35.
53. Yu J, Luo X, Xu H, Ma Q, Yuan J, Li X, et al. Identification of the Key Molecules Involved in Chronic Copper Exposure-Aggravated Memory Impairment in Transgenic Mice of Alzheimer's Disease Using Proteomic Analysis. *Journal of Alzheimer's Disease*. IOS Press; 2015;44:455–69.
54. Yang K, Belrose J, Trepanier CH, Lei G, Jackson MF, MacDonald JF. Fyn, a Potential Target for Alzheimer's Disease. *Journal of Alzheimer's Disease*. IOS Press; 2011;27:243–52.

55. Serenó L, Coma M, Rodríguez M, Sánchez-Ferrer P, Sánchez MB, Gich I, et al. A novel GSK-3 β inhibitor reduces Alzheimer's pathology and rescues neuronal loss in vivo. *Neurobiol Dis*. 2009;35:359–67.
56. Vassar R. BACE1 inhibitor drugs in clinical trials for Alzheimer's disease. *Alzheimer's Research & Therapy*. 2014;6:89.
57. Cheng F, Kovács IA, Barabási A-L. Network-based prediction of drug combinations. *Nat Commun*. 2019;10:1197.
58. Velazquez R, Ferreira E, Knowles S, Fux C, Rodin A, Winslow W, et al. Lifelong choline supplementation ameliorates Alzheimer's disease pathology and associated cognitive deficits by attenuating microglia activation. *Aging Cell*. 2019;18:e13037.
59. Wang H, Mei Z lin, Zhong KL, Hu M, Long Y, Miao MX, et al. Pretreatment with antiasthmatic drug ibudilast ameliorates A β 1–42-induced memory impairment and neurotoxicity in mice. *Pharmacology Biochemistry and Behavior*. 2014;124:373–9.
60. Guo C, Wang T, Zheng W, Shan Z-Y, Teng W-P, Wang Z-Y. Intranasal deferoxamine reverses iron-induced memory deficits and inhibits amyloidogenic APP processing in a transgenic mouse model of Alzheimer's disease. *Neurobiology of Aging*. 2013;34:562–75.
61. Nott A, Holtman IR, Coufal NG, Schlachetzki JCM, Yu M, Hu R, et al. Brain cell type-specific enhancer-promoter interactome maps and disease-risk association. *Science*. 2019;366:1134–9.
62. Zhou Y, Hou Y, Shen J, Mehra R, Kallianpur A, Culver DA, et al. A network medicine approach to investigation and population-based validation of disease manifestations and drug repurposing for COVID-19. *PLOS Biology*. Public Library of Science; 2020;18:e3000970.
63. Xu J, Zhang P, Huang Y, Zhou Y, Hou Y, Bekris L, et al. Multimodal single-cell/nucleus RNA sequencing data analysis uncovers molecular networks between disease-associated microglia and astrocytes with implications for drug repurposing in Alzheimer's disease. *Genome Res*. 2021;gr.272484.120.
64. Shchur O, Günnemann S. Overlapping Community Detection with Graph Neural Networks. arXiv:190912201 [cs, stat] [Internet]. 2019 [cited 2021 Jul 21]; Available from: <http://arxiv.org/abs/1909.12201>
65. Kipf TN, Welling M. Semi-Supervised Classification with Graph Convolutional Networks. arXiv:160902907 [cs, stat] [Internet]. 2017 [cited 2021 Aug 10]; Available from: <http://arxiv.org/abs/1609.02907>
66. Yang J, Leskovec J. Overlapping community detection at scale: a nonnegative matrix factorization approach. *Proceedings of the sixth ACM international conference on Web search and data mining* [Internet]. New York, NY, USA: Association for Computing

Machinery; 2013 [cited 2021 Jul 20]. p. 587–96. Available from: <https://doi.org/10.1145/2433396.2433471>

67. Yang J, McAuley J, Leskovec J. Community Detection in Networks with Node Attributes. 2013 IEEE 13th International Conference on Data Mining [Internet]. Dallas, TX, USA: IEEE; 2013 [cited 2021 Jul 21]. p. 1151–6. Available from: <http://ieeexplore.ieee.org/document/6729613/>

68. Wishart DS, Feunang YD, Guo AC, Lo EJ, Marcu A, Grant JR, et al. DrugBank 5.0: a major update to the DrugBank database for 2018. *Nucleic Acids Res.* 2018;46:D1074–82.

69. Miller JA, Woltjer RL, Goodenbour JM, Horvath S, Geschwind DH. Genes and pathways underlying regional and cell type changes in Alzheimer's disease. *Genome Med.* 2013;5:48.

70. Berchtold NC, Coleman PD, Cribbs DH, Rogers J, Gillen DL, Cotman CW. Synaptic genes are extensively downregulated across multiple brain regions in normal human aging and Alzheimer's disease. *Neurobiol Aging.* 2013;34:1653–61.

71. Wang M, Roussos P, McKenzie A, Zhou X, Kajiwar Y, Brennand KJ, et al. Integrative network analysis of nineteen brain regions identifies molecular signatures and networks underlying selective regional vulnerability to Alzheimer's disease. *Genome Med.* 2016;8:104.

72. Magistri M, Velmeshev D, Makhmutova M, Faghihi MA. Transcriptomics Profiling of Alzheimer's Disease Reveal Neurovascular Defects, Altered Amyloid- β Homeostasis, and Deregulated Expression of Long Noncoding RNAs. *J Alzheimers Dis.* 2015;48:647–65.

73. Annese A, Manzari C, Lionetti C, Picardi E, Horner DS, Chiara M, et al. Whole transcriptome profiling of Late-Onset Alzheimer's Disease patients provides insights into the molecular changes involved in the disease. *Sci Rep.* 2018;8:4282.

74. van Rooij JGJ, Meeter LHH, Melhem S, Nijholt DAT, Wong TH, Netherlands Brain Bank, et al. Hippocampal transcriptome profiling combined with protein-protein interaction analysis elucidates Alzheimer's disease pathways and genes. *Neurobiol Aging.* 2019;74:225–33.

75. Zhou Y, Fang J, Bekris LM, Kim YH, Pieper AA, Leverenz JB, et al. AlzGPS: a genome-wide positioning systems platform to catalyze multi-omics for Alzheimer's drug discovery. *Alzheimer's Research & Therapy.* 2021;13:24.

76. Keren-Shaul H, Spinrad A, Weiner A, Matcovitch-Natan O, Dvir-Szternfeld R, Ulland TK, et al. A Unique Microglia Type Associated with Restricting Development of Alzheimer's Disease. *Cell.* 2017;169:1276–1290.e17.

77. Zhou Y, Song WM, Andhey PS, Swain A, Levy T, Miller KR, et al. Human and mouse single-nucleus transcriptomics reveal TREM2-dependent and TREM2-independent cellular responses in Alzheimer's disease. *Nat Med*. 2020;26:131–42.
78. Leng K, Li E, Eser R, Piergies A, Sit R, Tan M, et al. Molecular characterization of selectively vulnerable neurons in Alzheimer's Disease [Internet]. *Neuroscience*; 2020 Apr. Available from: <http://biorxiv.org/lookup/doi/10.1101/2020.04.04.025825>
79. Grubman A, Chew G, Ouyang JF, Sun G, Choo XY, McLean C, et al. A single-cell atlas of entorhinal cortex from individuals with Alzheimer's disease reveals cell-type-specific gene expression regulation. *Nature Neuroscience*. Nature Publishing Group; 2019;22:2087–97.
80. Lau S-F, Cao H, Fu AKY, Ip NY. Single-nucleus transcriptome analysis reveals dysregulation of angiogenic endothelial cells and neuroprotective glia in Alzheimer's disease. *PNAS*. National Academy of Sciences; 2020;117:25800–9.
81. Kim DK, Park J, Han D, Yang J, Kim A, Woo J, et al. Molecular and functional signatures in a novel Alzheimer's disease mouse model assessed by quantitative proteomics. *Mol Neurodegener*. 2018;13:2.
82. Savas JN, Wang Y-Z, DeNardo LA, Martinez-Bartolome S, McClatchy DB, Hark TJ, et al. Amyloid Accumulation Drives Proteome-wide Alterations in Mouse Models of Alzheimer's Disease-like Pathology. *Cell Rep*. 2017;21:2614–27.
83. Martens M, Ammar A, Riutta A, Waagmeester A, Slenter DN, Hanspers K, et al. WikiPathways: connecting communities. *Nucleic Acids Res*. 2021;49:D613–21.
84. Kuleshov MV, Jones MR, Rouillard AD, Fernandez NF, Duan Q, Wang Z, et al. Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Res*. 2016;44:W90–7.

Figure Legends

Figure 1. A diagram illustrating NETTAG. We first (Step 1) apply a deep-learning model to capture the topological structure of the PPIs and divide it into multiple subnetwork modules (**Methods**). Then we discover that the divided subnetwork module could approximate protein functions annotated by GO knowledge portal (**Methods**). Next (step 2), we predict AD-associated genes which are functionally similar as genes that have been identified by different gene regulatory elements, i.e., CpG island, CTCF, enhancer, eQTL, histone, open chromatin, promoter, promoter flanking region and transcriptional factors. Finally (Step 3), we prioritize repurposed drugs for potential treatment of AD by evaluating network proximities between predicted AD-associated genes and drug target networks (**Methods**).

Figure 2. Gene regulatory landscape of AD GWAS loci. (A) Overview of AD GWAS loci across different chromosomes after considering nine gene regulatory elements: CpG island, CTCF, enhancer, eQTL, histone, open chromatin, promoter, promoter flanking region and transcriptional factors. (B) Proteins' cluster numbers are positively correlated with their GO annotations. We divide proteins into 10 groups according to their GO terms. For example, G1 group include the proteins that have at least one, but less than ten GO annotation. (C) ROC analyses of NETTAG based on two collected AD-association gene sets, i.e., AlzGene and DistiLD.

Figure 3. 156 Prioritized AD-associated genes (alzRGs). (A) Visualization of 156 predicted alzRGs. And 139 alzRGs are non-isolated and form a subnetwork with 294

PPIs. Prioritized alzRGs are colored with various evidences. Purple circles denote genes that are exclusively related with histone regulatory element; red circles denote genes that are exclusively related with promoter flanking region regulatory element; blue circles denote genes that are simultaneously related with multiple gene regulatory elements; grey circles denote genes with no gene regulatory element associations but have been identified as AD-associated according to GWAS studies; yellow circles denote genes that are AD-associated by other types of evidences, e.g., animal models and etc.; green circles that predicted alzRGs with no previously reported evidences. (B) Cumulative distributions of predicted scores with alzRGs and same amount of random non-alzRGs with similar degree distribution for eQTL, histone and promoter gene regulatory elements, respectively.

Figure 4. Predicted alzRGs are more likely differently expressed according to transcriptome studies (A) Visualization of 67 predicted alzRGs that are also DEGs according to human bulk RNA-seq studies with both LAD and control donors (**Methods**). (B) Violin plots show alzRGs are more likely differentially expressed in DAM according to mouse single-nucleus (GSE140511) RNA seq datasets. (C) Violin plot shows alzRGs are more likely differentially expressed in DAA according to human prefrontal cortex (PC) single-nucleus RNA seq dataset (GSE157827). (D) Violin plot shows alzRGs are more likely differentially expressed in DAA according to human entorhinal cortex (EC) single-nucleus RNA seq dataset (GSE157827). (E) Visualization of 32 predicted alzRGs that are also DEGs according to sc/sn RNA-seq studies

collected from both mouse models and human postmortem brain tissues in two disease-associated immune subtypes, i.e., DAM and DAA (**Methods**).

Figure 5. Predicted high-confidence AD-associated genes (A) Summary of multi-omics validations for all 156 predicted alzRGs with seven types of evidences. The genes are sorted in predicted score decreasing order (clockwise direction). We have collected seven types of evidences, including drug target, DEG by microarray studies, DEG by bulk RNA-seq studies, DEG in disease-associated microglia (DAM), DEG in disease-associated astrocyte (DAA), DEG by proteome studies and literature evidences. There are 126 predicted alzRGs that could be proved as AD-associated with at least one type of evidence.

Figure 6. Network-based discovery of repurposable drug candidates for AD by evaluating the network proximity between predicted alzRGs and drug-target networks (A) 118 prioritized drugs for AD treatment. Drug are grouped by fourteen different classes (e.g., immunological, respiratory, neurological, cardiovascular, and cancer (**Supplemental Table S5**)) defined by the first-level of the Anatomical Therapeutic Chemical (ATC) codes. (B) Proposed mechanism-of-action (MOAs) for Choline by drug-target network analysis. (C) Proposed MOAs for Ibudilast by drug-target network analysis.

Figure 1

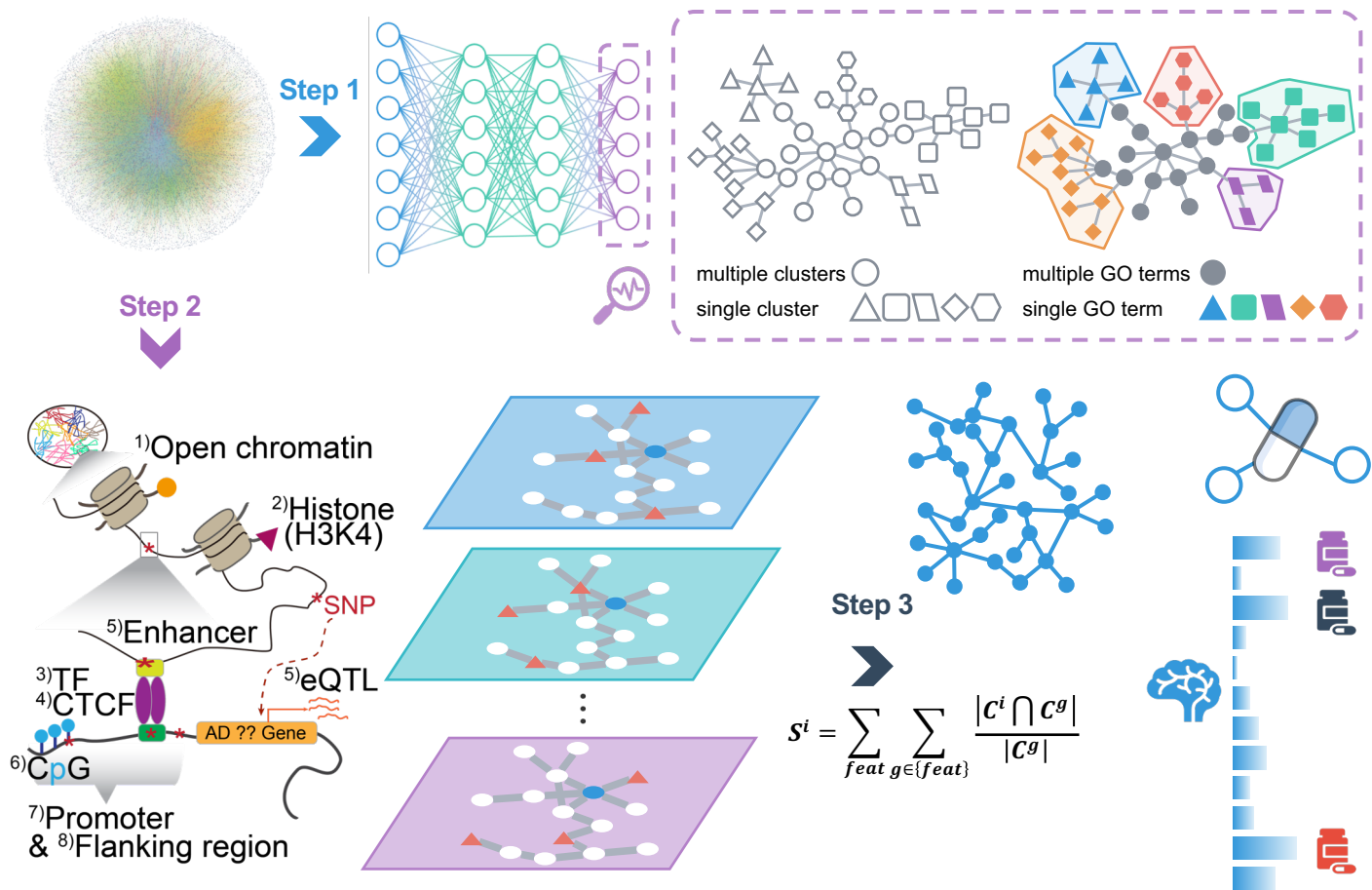
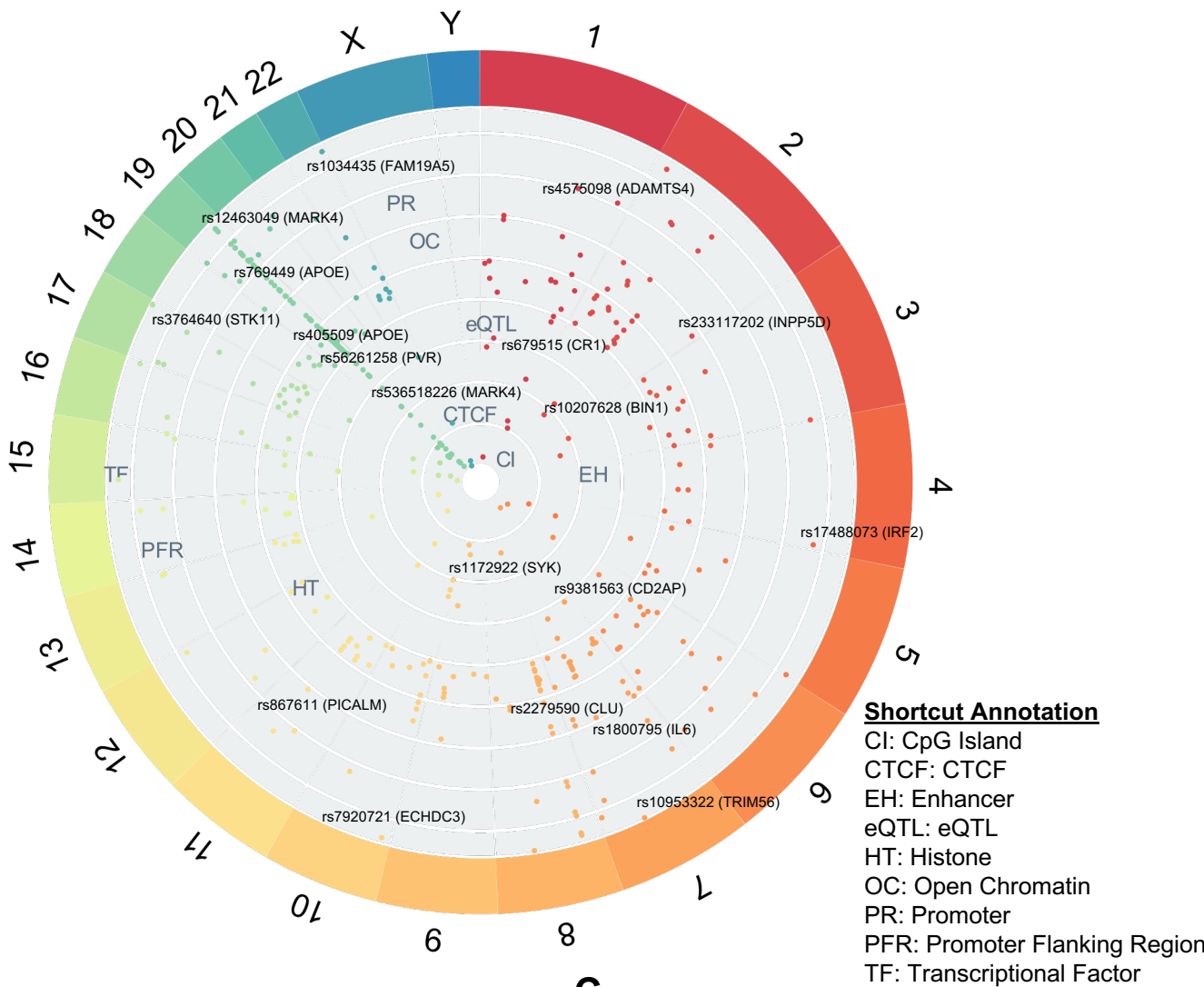
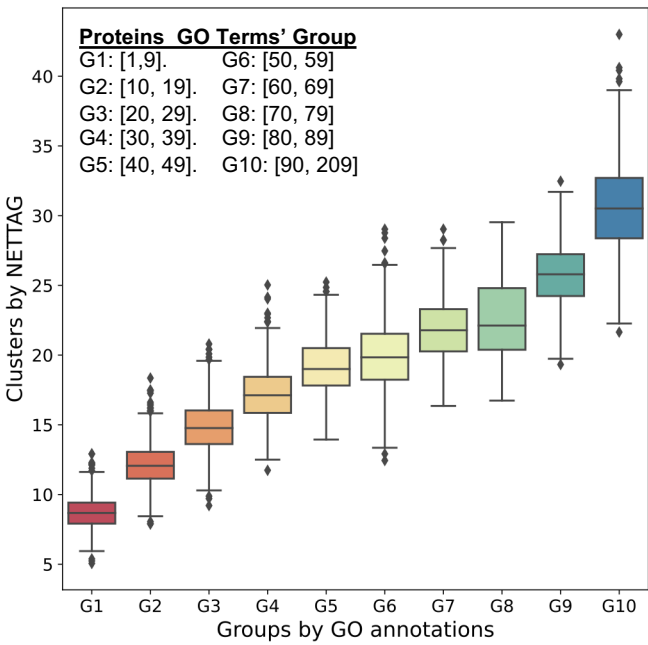


Figure 2

A



B



C

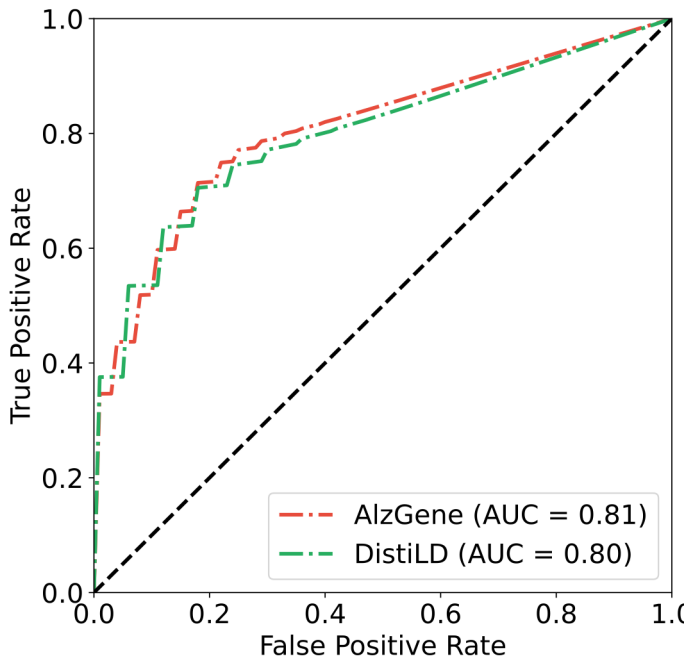
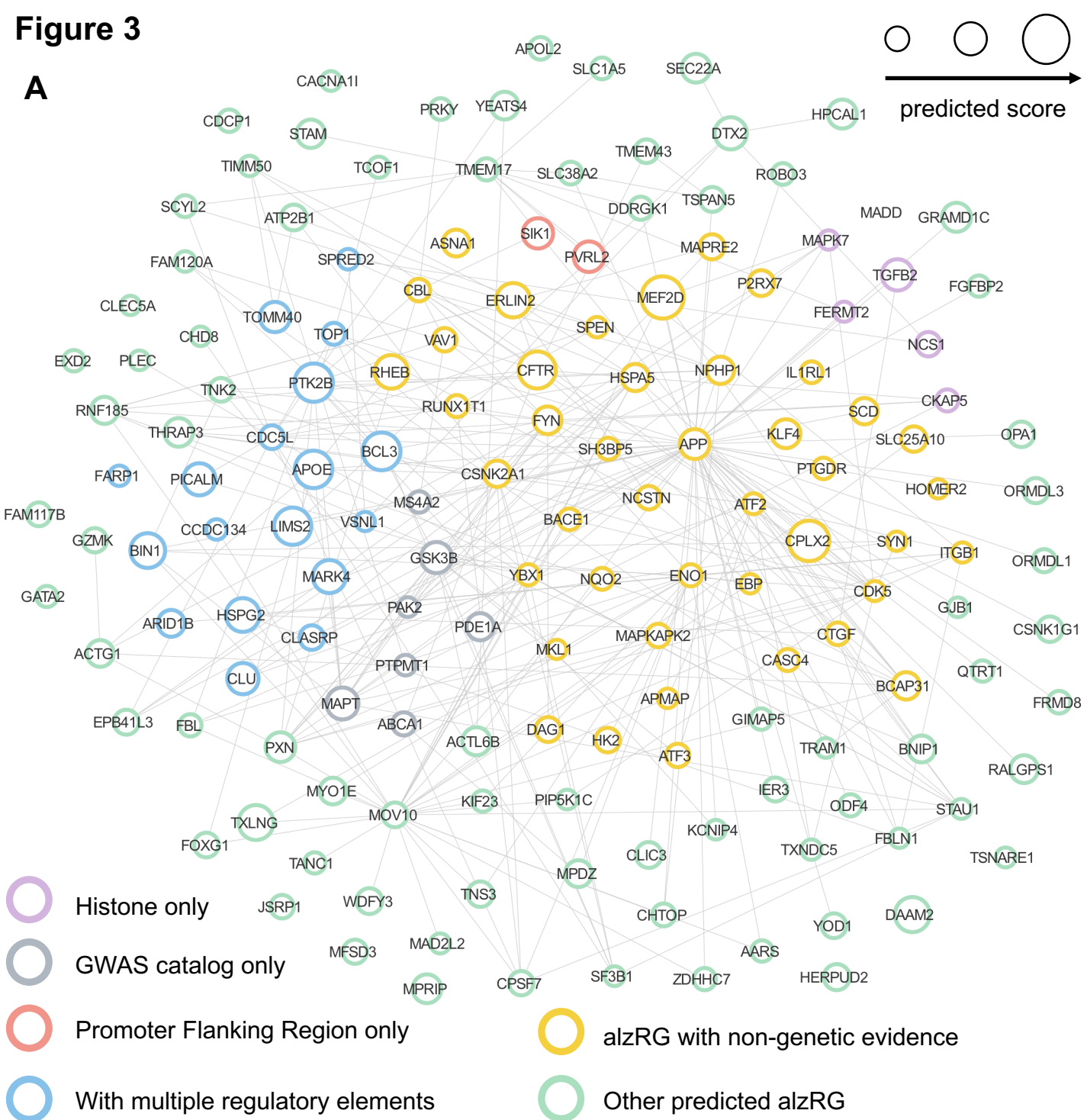


Figure 3

A



B

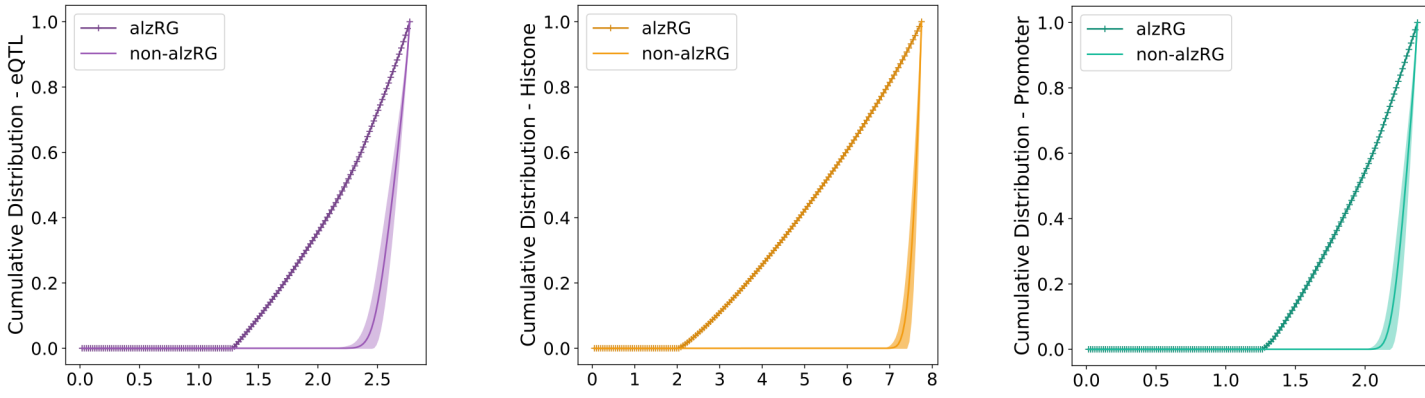
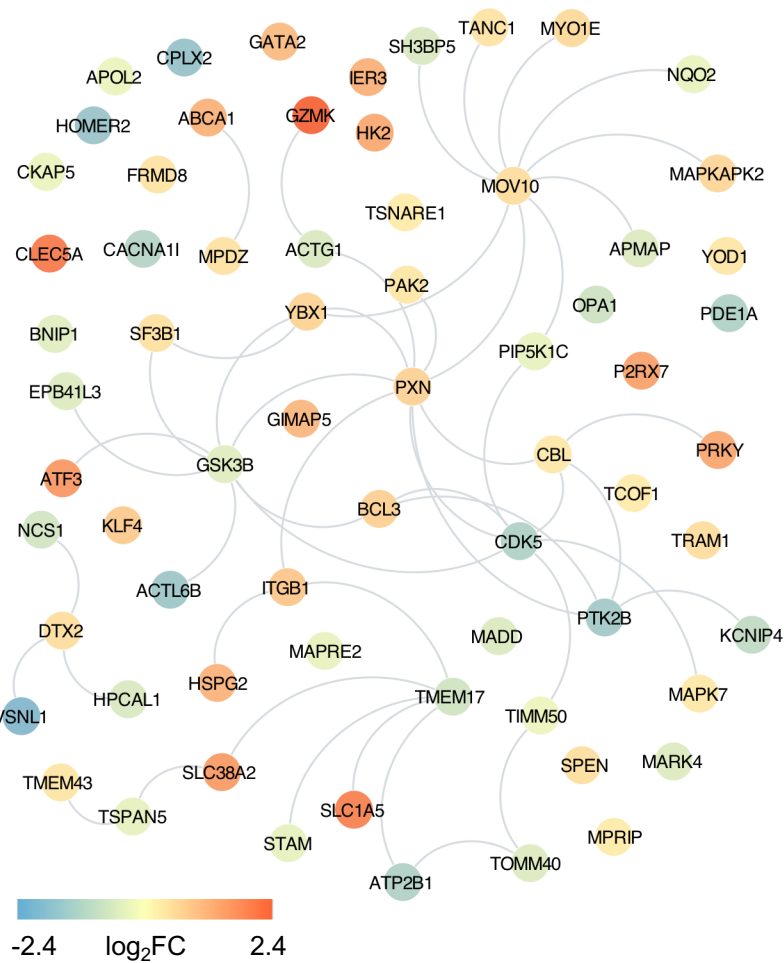
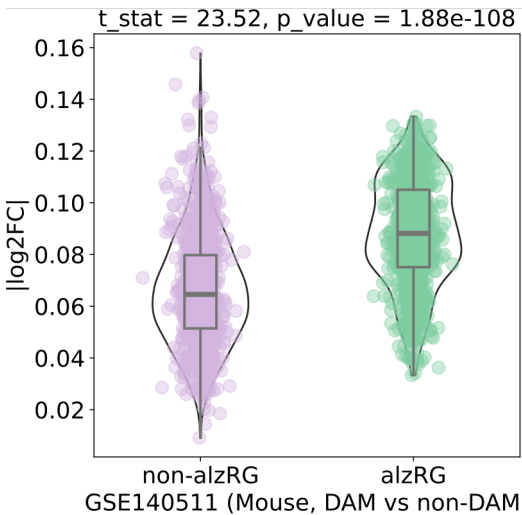


Figure 4

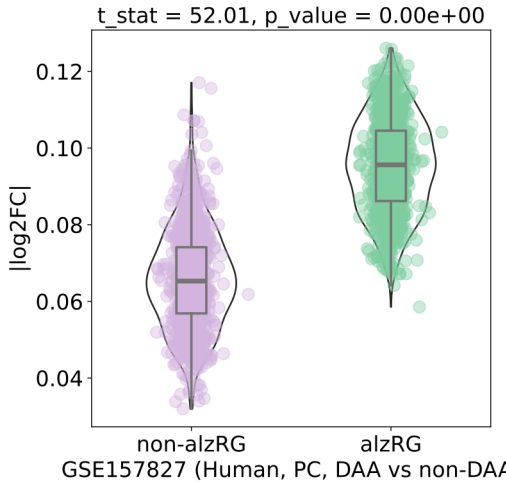
A



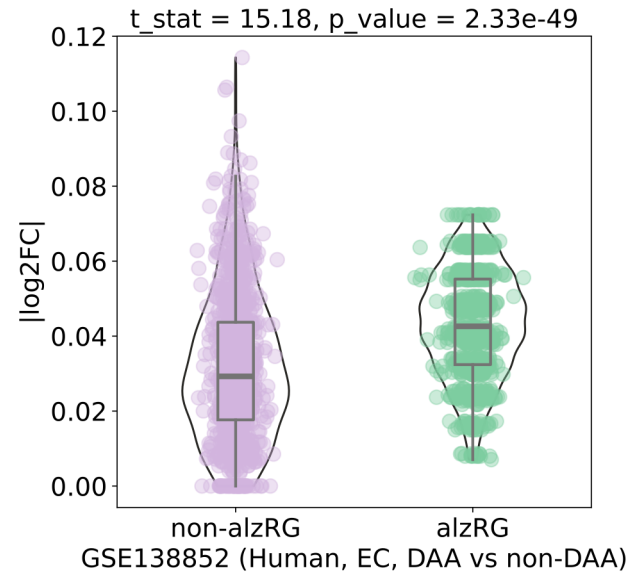
B



C



D



E

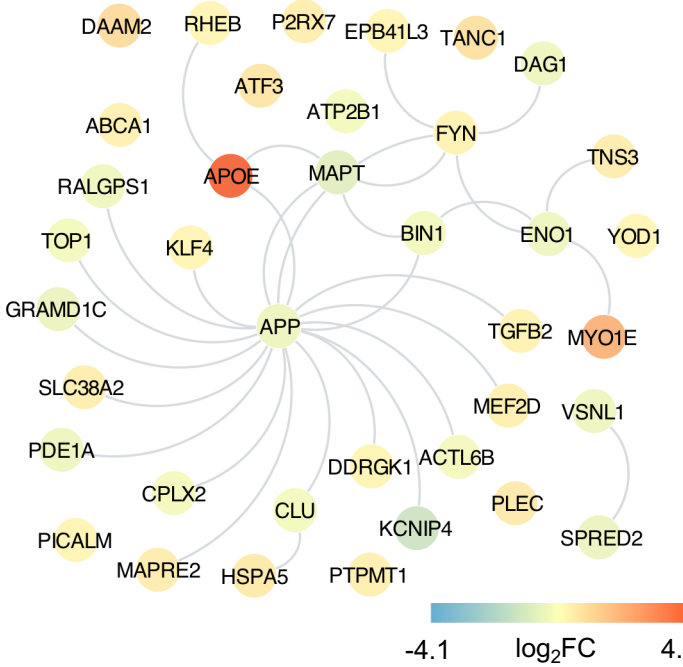
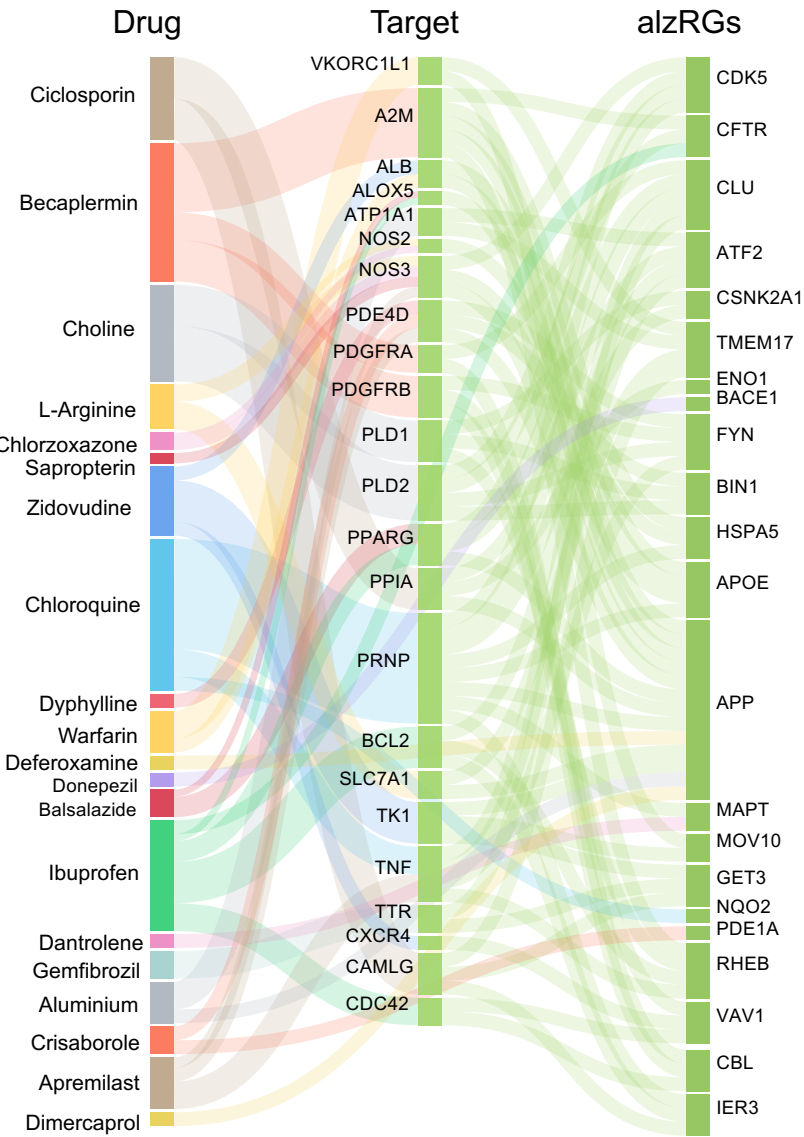


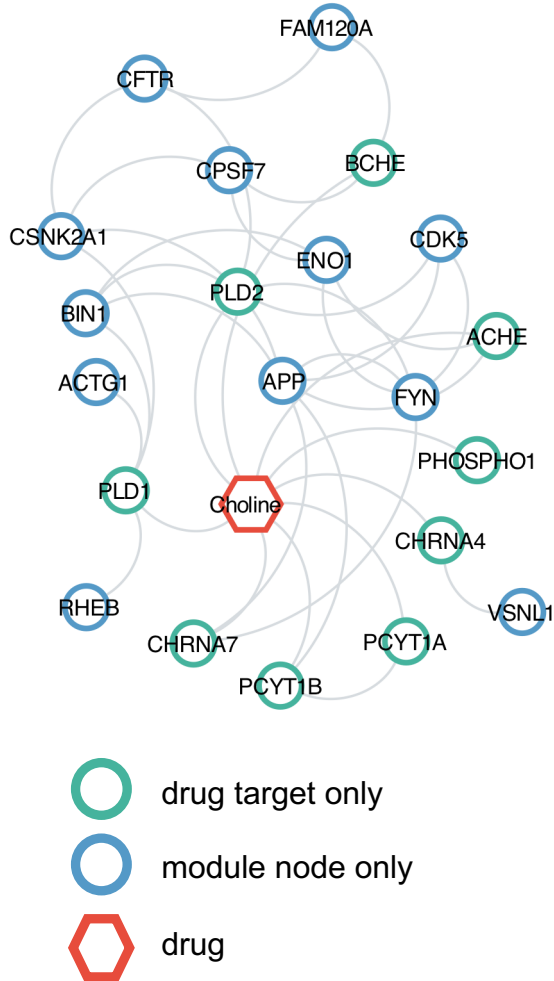
Figure 6

A



- Alimentary tract and metabolism
- Respiratory system
- Dermatologicals
- Blood and blood forming organs
- Various
- Genito-urinary system & sex hormones
- Sensory organs
- Cardiovascular system
- Antiparasitic products
- Antiinfectives for systemic use
- Nervous system
- Systemic hormonal preparations
- Musculo-skeletal system
- Antineoplastic and immunomodulating
- Not available

B



C

