

## **Network-based Translation of GWAS Findings to Pathobiology and Drug Repurposing for Alzheimer's Disease**

Jiansong Fang<sup>1,#</sup>, Pengyue Zhang<sup>2,#</sup>, Quan Wang<sup>3,4,#</sup>, Yadi Zhou<sup>1</sup>, Chien-Wei Chiang<sup>2</sup>, Rui Chen<sup>3,4</sup>, Bin Zhang<sup>1</sup>, Bingshan Li<sup>3,4</sup>, Stephen J. Lewis<sup>5</sup>, Andrew A. Pieper<sup>6</sup>, Lang Li<sup>2,\*</sup>, Jeffrey Cummings<sup>7,8</sup>, Feixiong Cheng<sup>1,9,10,\*</sup>

<sup>1</sup>Genomic Medicine Institute, Lerner Research Institute, Cleveland Clinic, Cleveland, OH 44195, USA

<sup>2</sup>Department of Biomedical Informatics, College of Medicine, Ohio State University, Columbus, OH 43210

<sup>3</sup>Department of Molecular Physiology and Biophysics, Vanderbilt University, Nashville, TN 37212, USA.

<sup>4</sup>Vanderbilt Genetics Institute, Vanderbilt University Medical Center, Nashville, TN 37212, USA.

<sup>5</sup>Department of Pediatrics, Case Western Reserve University, Cleveland, Ohio 44106, USA

<sup>6</sup>Harrington Discovery Institute, University Hospital Case Medical Center; Department of Psychiatry, Case Western Reserve University, Geriatric Research Education and Clinical Centers, Louis Stokes Cleveland VAMC, Cleveland, OH 44106, USA

<sup>7</sup>Cleveland Clinic Lou Ruvo Center for Brain Health, Las Vegas, NV 89106, USA.

<sup>8</sup>Department of Brain Health, School of Integrated Health Sciences, University of

Nevada, Las Vegas, Las Vegas, NV 89154, USA

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

<sup>9</sup>Department of Molecular Medicine, Cleveland Clinic Lerner College of Medicine,

Case Western Reserve University, Cleveland, OH 44195, USA

<sup>10</sup>Case Comprehensive Cancer Center, Case Western Reserve University School of

Medicine, Cleveland, Ohio 44106, USA.

# Co-first authors

Emails: [chengf@ccf.org](mailto:chengf@ccf.org) (F.C.) and [lang.li@osumc.edu](mailto:lang.li@osumc.edu) (L.L)

Corresponding to Feixiong Cheng, Ph.D.

Assistant Professor

Lerner Research Institute, Cleveland Clinic

Email: [chengf@ccf.org](mailto:chengf@ccf.org)

Tel: 216-444-7654

## Abstract

Genome-wide association studies (GWAS) have identified numerous susceptibility loci for Alzheimer's disease (AD). However, utilizing GWAS to identify high-confidence AD risk genes (ARGs) that can guide development of new therapeutics for patients suffering from AD has heretofore not been successful. To address this critical problem in the field, we have developed a genotype-informed, network-based methodology that interrogates pathogenesis to identify new therapeutics. When applied to AD, this approach integrates GWAS findings, multi-omics data from brain samples of AD patients and preclinical AD models, drug-target networks, and the human protein-protein interactome, along with large-scale patient database validation and *in vitro* mechanistic observations in human microglia cells. Through this approach, we identified 103 ARGs validated by various levels of pathobiological evidence in AD. Via network-based prediction and population-based validation, we then showed that pioglitazone usage is significantly associated with decreased risk of AD (hazard ratio (HR) = 0.895, 95% confidence interval [CI] 0.841-0.951,  $P = 3.97 \times 10^{-4}$ ) in a retrospective case-control validation. Pioglitazone is a peroxisome proliferator-activated receptor agonist used to treat type 2 diabetes, and propensity score matching cohort studies confirmed its association with reduced risk of AD in comparison to glipizide (HR = 0.921, 95% CI 0.861-0.983,  $P = 0.0146$ ), an insulin secretagogue that is also used to treat type 2 diabetes. *In vitro* experiments showed that pioglitazone downregulated glycogen synthase kinase 3 beta (GSK3 $\beta$ ) and cyclin-dependent kinase (CDK5) in human microglia cells, supporting a possible mechanism-of-action for its beneficial effect in AD. In summary, we present an integrated, network-based methodology to rapidly translate GWAS findings and multi-omics data to genotype-informed therapeutic discovery in AD.

## Introduction

Alzheimer's disease (AD) is a chronic neurodegenerative disorder associated with progressive cognitive decline, extracellular amyloid plaques, intracellular neurofibrillary tangles, and neuronal death (Masters *et al.*, 2015, Long and Holtzman, 2019). AD and other dementias are an increasingly important global health burden, recently estimated to affect 43.8 million people worldwide (Nichols *et al.*, 2019). Although genome-wide association studies (GWAS) have identified over 40 genome-wide significant susceptibility loci for AD (Lambert *et al.*, 2013, Cuyvers and Sleegers, 2016, Jung *et al.*, 2018, Jansen *et al.*, 2019), translating these findings into identification of high-confidence AD risk genes (ARGs) and potential therapies has eluded the field. Indeed, since Dr. Alois Alzheimer first described the condition in 1906, scientists have not developed any effective disease modifying treatments (Alzheimer, 1907, Long and Holtzman, 2019).

The number of AD patients is expected to rise to 16 million by 2050 in the United States (U.S.) alone (Kodamullil *et al.*, 2017, Alteri and Guizzaro, 2018), while the attrition rate for AD clinical trials (2002-2012) is estimated at 99.6% (Cummings *et al.*, 2014). One possible explanation for why most candidate drugs fail in later-stage clinical trials is poor target selection. Broadly in disease, drug targets with genetic support have carried a high success rate amongst U.S. Food and Drug Administration (FDA)-approved therapies (Cook *et al.*, 2014, Nelson *et al.*, 2015). However, this has not been the case with AD, and the translational application of multi-omics data such as GWAS for target identification and therapeutic development in AD remains

challenging.

We recently demonstrated the utility of network-based methodologies for accelerating target identification and therapeutic discovery by exploiting multi-omics profiles from individual patients in multiple complex diseases, including cardiovascular disease (Cheng and Desai, 2018), cancer (Cheng and Lu, 2019), and schizophrenia (Wang *et al.*, 2019). We now posit that systematic identification of likely causal genes by incorporating GWAS findings and multi-omics profiles with human interactome network models will also reveal disease-specific targets for genotype-informed therapeutic discovery in AD. This approach entails unique integration of the genome, transcriptome, proteome, and the human protein-protein interactome. Here, we report on application of this process to AD (**Fig. 1A**). Specifically, we integrated GWAS findings, multi-omics information generated from brain samples of individual AD patients and AD transgenic mouse models, publicly available drug-target networks, and the human protein-protein interactome, along with large-scale patient database validation and *in vitro* mechanistic observations in human microglia cells. By applying this network methodology to 106 AD loci, we identified a set of ARGs, most of which could be validated by multiple transcriptomic and proteomics profiles generated from AD transgenic mouse models. These ARGs represent enriched druggable targets for therapeutic discovery. Via a network-based prediction from findings of ARGs and population-based validation (**Fig. 1B and C**), we discovered that pioglitazone usage was significantly associated with decreased risk of AD in large-scale patient data. Subsequent *in vitro* mechanistic observations (**Fig.**

**1D)** revealed that pioglitazone significantly downregulates expression of glycogen synthase kinase 3 beta (GSK3 $\beta$ ) and cyclin-dependent kinase 5 (CDK5) in human microglia cells, mechanistically supporting network-based and population-based findings.

## **Materials and Methods**

### **Collection of GWAS SNPs from large-scale studies**

In this study, we assembled multiple single nucleotide polymorphisms (SNPs) associated with AD from 15 large-scale GWAS studies in diverse population groups, conducted between 2007 and 2019 (**Supplementary Table 1**). Collectively, these studies include over 270,000 AD cases and 1100,000 controls. To maximize genetic signals based on the omnigenic hypothesis (Boyle *et al.*, 2017), we adopted a loose threshold ( $P < 1 \times 10^{-5}$ ) to collect AD risk SNPs, which yielded 106 unique GWAS SNPs.

### **Construction of human protein-protein interactome**

To build a comprehensive human protein-protein interactome, we assembled data from 15 common resources with multiple levels of experimental evidence.

Specifically, we focused on high-quality protein-protein interactions (PPIs) with the following five types of experimental data: (1) binary PPIs tested by high-throughput yeast-two-hybrid (Y2H) systems; (2) kinase-substrate interactions by literature-

derived low-throughput and high-throughput experiments; (3) literature-curated PPIs identified by affinity purification followed by mass spectrometry (AP-MS), Y2H and by literature-derived low-throughput experiments, and protein three-dimensional structures; (4) signaling network by literature-derived low-throughput experiments; (5) protein complex data (see **Supplementary Methods**). The genes were mapped to their Entrez ID based on the NCBI database, and duplicated pairs were removed. Collectively, the integrated human interactome included 351,444 PPIs connecting 17,706 unique proteins. More details are provided in our recent studies (Cheng and Desai, 2018, Cheng and Lu, 2019)

### **Collection of functional genomics data**

We collected the distal regulatory element (DRE)-promoter links inferred from two studies. The first study was the capture Hi-C study of cell line GM12878 (Mifsud *et al.*, 2015). We obtained 1,618,000 DRE-promoter links predicted for GM12878 from <http://www.ebi.ac.uk/arrayexpress/experiments/E-MTAB-2323/>. The other dataset we used was from the FANTOM5 project (Andersson *et al.*, 2014), in which cap analysis of gene expression (CAGE) technology was employed to infer enhancer-promoter links across multiple human tissues. We downloaded FANTOM5 data from <http://enhancer.binf.ku.dk/presets/> and obtained 66,899 enhancer-promoter links.

### **Collection of biological and functional data**

*Disease-associated genes from Open targets.* Open targets refers to a comprehensive

platform for therapeutic target identification and validation (Koscielny *et al.*, 2017).

We collected 527 AD disease-associated genes (**Supplementary Table 2**) from the Open Targets database (Assess in September, 2019).

*Experimentally validated genes for Alzheimer's disease.* We further collected 144 high-quality experimentally validated AD genes (**Supplementary Table 2**), which combined genes involved in pathobiology of amyloidosis, tauopathy, or both, and genes characterizing other AD pathological hypothesis including neuroinflammation and vascular dysfunction (**Supplementary Method**).

*Brain specific expression.* We downloaded RNA-Seq data (RPKM value) of 32 tissues from GTEx V6 release (accessed on April 01, 2016, <https://gtexportal.org/home/>). We defined those genes with  $RPKM \geq 1$  in over 80% of samples as tissue-expressed genes and the other genes as tissue-unexpressed. To quantify the expression significance of tissue-expressed gene  $i$  in tissue  $t$ , we calculated the average expression  $\langle E(i) \rangle$  and the standard deviation  $\delta_E(i)$  of a gene's expression across all considered tissues. The significance of gene expression in tissue  $t$  is defined as

$$z_E(i, t) = (E(i, t) - \langle E(i) \rangle) / \delta_E(i) \quad (1)$$

The details have been described in previous studies (Cheng and Desai, 2018, Cheng and Lu, 2019).

*Gene Expression.* We collected human microarray data in AD cases versus controls



with brain samples from two independent datasets (GSE29378 and GSE84422) (Miller *et al.*, 2013, Wang *et al.*, 2016). We also collected mouse microarray data from AD transgenic mouse vs. controls, including brain microglia of 5XFAD mice from 2 independent datasets (GSE65067 and GSE74615) (Orre *et al.*, 2014, Wang *et al.*, 2015), and brain hippocampus of Tg4510 mice (GSE53480 and GSE57583) (Polito *et al.*, 2014).

The original microarray datasets were obtained from Gene Expression Omnibus (<https://www.ncbi.nlm.nih.gov/geo>). Detailed information of these 6 GEO datasets is provided in **Supplementary Table 3**. All raw expression data were log<sub>2</sub> transformed, and all samples were quantile normalized together. Probe IDs in each dataset were mapped to National Center for Biotechnology Information (NCBI) Entrez IDs, and probes mapping to multiple genome regions or without corresponding entrez IDs were deleted. The items were imported to R statistical processing environment using a LIMMA/Bioconductor package. All the mouse genes were further transferred into unique human-orthologous genes using the Mouse Genome Informatics (MGI) database (Eppig *et al.*, 2017). Genes with threshold fold change (FC) > 1.2 were defined as exhibiting differential expression and prioritized as predicted AD risk genes.

### **Enrichment analysis**

Differentially expressed gene/protein (DEG/DEP) sets from multiple data sources were collected for enrichment analysis. This included a total of 6 bulk RNA-seq

datasets and 10 proteomic datasets from 4 types of AD transgenic mouse models, including 5XFAD, Tg4510, ADLPAPT and hAPP (see **Supplementary Method**).

*Bulk RNA-seq.* We collected 2 RNA-seq datasets from brain or brain microglia of 5XFAD mice. In addition, we obtained 4 RNA-seq datasets from brain microglia of Tg4510 mice across different months [M] age (2M, 4M, 6 M, and 8M). Differential expression analysis was performed using DESeq (Anders, 2010), while threshold for significance of differential expression was set to  $FDR < 0.05$  using Benjamini-Hochberg's method. After mapping mouse genes to human-orthologous gene (Eppig *et al.*, 2017), we obtained 6 differentially expressed gene sets.

*Proteomics.* In total, 10 proteomic datasets were assembled from 3 types of AD transgenic mouse models in two recent publications (Savas *et al.*, 2017, Kim *et al.*, 2018). The first study performed global quantitative proteomic analysis in hAPP and hAPP-PS1 mouse models at young (3 month [M]) and old ages (12 M) (Savas *et al.*, 2017). We obtained four sets of DEPs (hAPP\_3M, hAPP\_12M, hAPP-PS1\_3M and hAPP-PS1\_12M) after merging the DEPs from different brain regions. The second study performed quantitative proteomics to uncover molecular and functional signatures in the hippocampus of three types of transgenic mice (Kim *et al.*, 2018). Two of these mouse lines, including ADLPAPT (4M, 7M, 10M) that carry three human transgenes (APP, PSEN1 and tau) and hAPP-PS1 (4M, 7M, 10M) mouse, were used in this study. After mapping mouse genes to human-orthologous gene (Eppig *et*

*al.*, 2017), we obtained 10 sets of DEPs.

## **AD risk gene prediction**

We utilized our recently developed Bayesian model selection method to predict ARGs (**Fig 1A**), based on the assumption that true risk genes are more densely connected with each other in a biological network (Wang *et al.*, 2019). Specifically, we collected at most 20 genes in the 2Mb region centered at a GWAS index SNP as the candidates for that particular locus. Assigning  $L$  as the number of GWAS loci, and we then denoted a vector of genes with length  $L$ , each being from one of the  $L$  GWAS loci, as  $(X_1, \dots, X_L)$ , and termed it as candidate risk gene set (CRGS). Assigning  $N$  to represent the biological network, we then calculated  $P(X_1, \dots, X_L | N)$ , and selected a CRGS with maximum posterior probability. Computationally, it is not feasible to enumerate all possible gene combinations, and we therefore adopted a Gibbs sampling algorithm to transition the problem into a single-dimensional sampling procedure. For example, when sampling the risk gene from candidates at the  $L$ -th locus, we assumed that the risk genes at all other  $L-1$  loci had been selected, and the sampling probability for a gene at the  $L$ -th locus was computed as conditional on the  $L-1$  risk genes, based on its closeness to other  $L-1$  risk genes in the network. For each candidate gene  $X_L$  at the  $L$ -th locus, we assigned  $M_1$  to represent the event that  $X_L$  is the risk gene at locus  $L$ ,  $M_0$  represent the event that  $X_L$  is not the risk gene at locus  $L$ , and  $X_{-L}$  to represent all the selected risk genes in the other  $L-1$  loci. The Bayesian model selection can be depicted as

$$\frac{P(M_1|X_{-L},N)}{P(M_0|X_{-L},N)} = \frac{P(M_1)}{P(M_0)} \frac{P(X_{-L}|M_1,N)}{P(X_{-L}|M_0,N)} \quad (2)$$

where  $\frac{P(X_{-L}|M_1,N)}{P(X_{-L}|M_0,N)}$  is a Bayesian Factor (BF) measuring the closeness between  $X_{-L}$  and  $X_L$  in network  $N$  and  $\frac{P(M_1)}{P(M_0)}$  is prior odds. The prior odds reflect the prior knowledge whether  $X_L$  is a risk gene or not and we assumed  $P(M_1) = P(M_0)$  in this study.

In regarding to  $\frac{P(X_{-L}|M_1,N)}{P(X_{-L}|M_0,N)}$ , we adopted the random walk with restart (RWR) algorithm to calculate the BF. Starting from any node  $n_i$  in a predefined network  $N$ , the walker faces two options at each step: either moving to a direct neighbor with a probability  $1 - r$  or jumping back to  $n_i$  with a probability  $r$ . The fixed parameter  $r$  is called the restart probability in RWR, and  $r$  was set as 0.3 in this study (Wang *et al.*, 2019). Let  $W$  be the adjacency matrix that decides which neighbor to be moved to, and  $q_t$  be the reaching probability of all nodes at step  $t$ . The RWR algorithm is formalized as

$$q_{t+1} = (1 - r)Wq_t + rs_{n_i} \quad (3)$$

$s_{n_i}$  is a vector with the  $i$ -th element as 1 and 0 for others, which means the starting node is  $n_i$ . Following the equation,  $q_t$  can be updated step by step until

$|q_{t+1} - q_t|^2 < T_{rwr}$ , where  $T_{rwr}$  is a predefined threshold. We set  $T_{rwr}$  as  $1e^{-6}$

(Wang *et al.*, 2019). The adjacency matrix  $W$  represents the distance between any

two nodes in the network and we adopted the same network and strategy in our

previous work to calculate  $W$ . We calculated  $P(X_{-L}|M_1, N)$  based on  $W$ . We

mapped  $X_L$  to the rows of  $W$  and  $X_{-L}$  to the columns of  $W$ , and obtained a vector

with the same length as  $X_{-L}$ . The sum of the vector was calculated as  $P(X_{-L}|M_1,$

$N)$ . In this study, we assumed  $P(X_{-L}|M_0, N)$  to be the same for all different

candidate genes. Through the Bayesian model selection equation,

$$\frac{P(M_1|X_{-L},N)}{P(M_0|X_{-L},N)} = \frac{P(M_1) P(X_{-L}|M_1,N)}{P(M_0) P(X_{-L}|M_0,N)} \quad (4)$$

we obtained a value for each candidate genes at locus  $L$ . We used these values as sampling probabilities for Gibbs sampling to choose a risk gene for locus  $L$ . We then repeated the sampling across the remaining loci and iterated the sampling process until convergence. Specially, in each round of Gibbs sampling, we calculated the sampling frequency for each candidate gene. The frequency was compared with that of the previous round, and if the sum of squares of frequency differences across all selected genes was smaller than a predefined threshold ( $1 \times 10^{-4}$  used in this study) then the sampling procedure was halted. Based on the sampling, we are able to assess the confidence of candidates being risk genes. In this study, we selected the gene with highest sampling frequency as risk gene for each locus as described previously (Wang *et al.*, 2019).

### **Construction of drug-target network**

We integrated six commonly used resources to collect high-quality physical drug-target interactions for FDA-approved drugs. We obtained biophysical drug-target interaction using reported binding affinity data: inhibition constant/potency ( $K_i$ ), dissociation constant ( $K_d$ ), median effective concentration ( $EC_{50}$ ), or median inhibitory concentration ( $IC_{50}$ )  $\leq 10 \mu\text{M}$ . First, we extracted the bioactivity data from the DrugBank database (v4.3) (Wishart *et al.*, 2018), the Therapeutic Target Database (TTD, v4.3.02) (Li *et al.*, 2018), and the PharmGKB database (Barbarino *et al.*,

2018). In addition, we acquired drug-target interactions related to FDA-approved drugs from three commonly used databases (Cheng and Lu, 2019). To improve data quality, we pooled only those items that satisfied the following four criteria: (i) binding affinities, including  $K_i$ ,  $K_d$ ,  $IC_{50}$ , or  $EC_{50}$ ,  $\leq 10 \mu\text{M}$ ; (ii) the target protein has a unique UniProt accession number; (iii) proteins marked as “reviewed” in the UniProt database; and (iv) proteins are from Homo sapiens. Totally, we collected 15,367 drug-target interactions connecting 1,608 FDA-approved drugs and 2,251 unique human targets (**Supplementary Table 4**).

### **Description of network proximity**

Given the set of disease proteins ( $A$ ), the set of drug targets ( $B$ ), then the closest distance  $d_{AB}$  measured by the average shortest path length of all the nodes to the other module in the human protein-protein interactome can be defined as:

$$\langle d_{AB} \rangle = \frac{1}{\|A\| + \|B\|} \left( \sum_{a \in A} \min_{b \in B} d(a, b) + \sum_{b \in B} \min_{a \in A} d(a, b) \right) \quad (5)$$

where  $d(a, b)$  denotes to the shortest path length between protein  $a$  and drug target  $b$ .

To calculate the significance of the network distance between a given drug and disease module, we constructed a reference distance distribution corresponding to the expected distance between two randomly selected groups of proteins of the same size and degree distribution as the original disease proteins and drug targets in the network. This procedure was run 1000 times. The mean  $\bar{d}$  and standard deviation ( $\sigma_d$ ) of the reference distribution were used to calculate a z-score ( $z_d$ ) by converting

an observed (non-Euclidean) distance  $d$  to a normalized distance:

$$z_d = \frac{d - \bar{d}}{\sigma_d} \quad (6)$$

### **Pharmacoepidemiologic validation**

*Patient Cohort Preparation.* The pharmacoepidemiology study utilized the MarketScan Medicare Supplemental database from 2012 to 2017. The dataset included individual-level diagnosis codes, procedure codes, and pharmacy claims for ~7 million patients per year. Pharmacy prescriptions of pioglitazone and glipizide were identified by using RxNorm and National Drug Code (NDC). For a subject exposed to the aforementioned drugs, a drug episode is defined as the time between drug initiation and drug discontinuation. Specifically, drug initiation is defined as the first day of drug supply (i.e. first prescription date). Drug discontinuation is defined as the last day of drug supply (i.e. last prescription date + days of supply) accompanied by no drug supply for the next 60 days. Gaps of less than 60-day of drug supply were allowed within a drug episode. The pioglitazone cohort included the first pioglitazone episode for each subject, as well as the glipizide cohort. Further, we excluded observations that started within 180-days of insurance enrollment. For the final cohorts, demographic variables including age, gender and geographical location were collected. Additionally, diagnoses of hypertension (HTN) and type 2 diabetics (T2D) (the ICD codes were given in **Supplementary Materials and Method**) before drug initiation were collected. These variables were specifically selected to address potential confounding biases. Lastly, a control cohort was selected from patients not

exposed to pioglitazone. Specifically, non-exposures were matched to the exposures (ratio 1:4) by initiation time of pioglitazone, enrollment history, age, and gender. The geographical location, diagnoses of HTN and T2D were collected for the control cohort as well.

*Outcome measurement.* The outcome was time from drug initiation to diagnose of AD, which was defined by using the ICD codes (**Supplementary Materials and Method**). For pioglitazone and glipizide cohorts, observations without diagnose of AD were censored at the end of drug episodes. For the control cohort, the corresponding pioglitazone episode's starting date was used as the starting time. Observations without diagnosis of AD were censored at the corresponding pioglitazone episode's end date.

*Statistical analysis.* Survival curves for time to AD were estimated using a Kaplan-Meier estimator. Additionally, propensity score stratified survival analysis was conducted to investigate the risk of AD between pioglitazone users and non-pioglitazone users, as well as pioglitazone users and glipizide users. For each comparison, the propensity score of taking pioglitazone was estimated by using a logistic regression model in which covariates included age, gender, geographical location, T2D diagnosis, and HTN diagnosis. Furthermore, propensity score stratified Cox-proportional hazards models were used to conduct statistical inference for the hazard ratios (HR) of developing AD between cohorts.



## **Experimental validation**

*Reagents.* Pioglitazone was acquired from Topscience. Lipopolysaccharides (LPS) (Cat# L2880) and 3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide (MTT) were obtained from Sigma-Aldrich. Antibodies against Phospho-GSK3B-Y216 (Cat# AP0261), GSK3B (Cat# A2081) and CDK5 (Cat# A5730), were purchased from ABclonal Technology. CDK5-Phospho-Tyr15 (Cat# YP0380) was obtained from Immunoway (Plano, Texas, USA). All other reagents were purchased from Sigma-Aldrich unless otherwise specified.

*Cell viability.* Human microglia HMC3 cells were purchased from American Type Culture Collection (ATCC, Manassas, VA). Cell viability was detected by MTT method as described previously (Huang *et al.*, 2019). 5000 cells/well were plated in 96-well plates for 12 h, and then treated with pioglitazone for 48 h. After treatment, MTT solution was added to the cells to a final concentration of 1 mg/mL and the mixture was allowed to incubate at 37 °C for 4 h. The supernatant was removed, and precipitates were dissolved in DMSO. Absorbance was measured at 570 nm using a Synergy H1 microplate reader (BioTek Instruments, Winooski, VT, USA).

*Western blot analysis.* HMC3 cells were pre-treated with pioglitazone (3 µM or 10 µM) and DMSO (control vehicle), and followed with 1 µg/mL LPS for 30 min. Cells were harvested, washed with cold PBS, and then lysed with RIPA Lysis Buffer with

1% Protease Inhibitor (Cat# P8340, Sigma-Aldrich). Total protein concentrations were measured using a standard BCA protein assay kit (Bio-Rad, CA, USA), according to the manufacturer's manual. Samples were electrophoresed by sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE), then blotted onto a polyvinylidene difluoride (PVDF; EMD Millipore, Darmstadt, Germany) membrane. After transferring, membranes were probed with specific primary antibodies (1:1000) at 4°C overnight. Specific protein bands were detected using a chemiluminescence reagent after hybridization with a horseradish peroxidase (HRP)-conjugated secondary antibody (1:3000).

## Results

### Pipeline of the network-based methodology

We utilized a Bayesian model selection method to predict ARGs (Wang *et al.*, 2019), based on the assumption that likely causal risk genes are more densely connected with each other in a biological network (**Fig. 1A**). Specifically, we collected at most 20 genes in a 2 Mb region centered at a GWAS index SNP as the candidates for that particular locus. We then calculated a posterior probability for each candidate gene based on its closeness to other risk genes in the biological network. For each GWAS locus, the candidate gene with the highest posterior probability was predicted as an ARG (see Methods). By applying this framework to the 106 AD GWAS loci we collected (Methods), we predicted 103 ARGs after merging the overlapping genes

across several different loci (**Supplementary Table 5**). Meanwhile, we also predicted a set of local background genes (LBGs) as a control group in the following analyses (Wang *et al.*, 2019). We validated our 103 ARGs using multi-omics data, including functional genomic characteristics and transcriptomics, as well as proteomic profiles generated from diverse AD transgenic mice models.

### **Multi-omics validation of network-predicted risk genes in AD**

*A significant disease module formed by ARGs in the human interactome.* AD involves a complex, polygenic, and pleiotropic genetic architecture. We found that 103 ARGs formed significantly connected subgraphs (termed disease module) rather than being scattered randomly in the human protein-protein interactome, consistent with previous disease module analyses that we demonstrated in other multiple complex diseases (Cheng and Desai, 2018, Cheng and Lu, 2019). Specifically, 68.0% of ARGs (70/103,  $P=0.015$ , permutation test) form the largest connected subnetwork (disease module), in comparison to the same number of randomly selected genes with similar connectivity (degree) as the original seed genes in the human interactome (**Supplementary Fig. S1**). This disease module (**Fig. 2A**) includes 128 PPIs (edges or links) connecting 70 unique proteins (nodes). Network analysis revealed 14 proteins with connectivity higher than 5, the top five of which were *ESR1*, *PSMC5*, *MAPK1*, *PAK1* and *NFKB1*. These same five genes have previously been implicated in AD (Granic *et al.*, 2009, Conejero-Goldberg *et al.*, 2011, Ma *et al.*, 2012, Wang *et al.*, 2016). For example, *ESR1* interacts with tau protein *in vivo*, and prevents glutamate

excitotoxic injury by A $\beta$  via estrogen signaling (Wang *et al.*, 2016). Gene expression analysis shows that PSMC5 was significantly overexpressed in patients carrying APOE4 mutations in comparison to APOE4 wild-type group (Conejero-Goldberg *et al.*, 2011). In summary, 103 predicted ARGs comprise a strong disease module in the human interactome.

*ARGs capture more gene regulatory elements.* Because a majority of GWAS SNPs lie in non-coding region and exert their function by gene regulation (Zhao *et al.*, 2018), we explored the gene regulatory elements of ARGs by testing the hypothesis that the network-predicted risk genes capture more distal regulatory elements (DREs)-promoter connections compared to background. We collected DRE-promoter connection data generated by two technologies: Cap Analysis of Gene Expression (CAGE) from Functional Annotation of the Mammalian Genome 5 (FANTOM5) project and capture Hi-C (see Methods) (Andersson *et al.*, 2014, Won *et al.*, 2016). Through this, we found that the ARGs are indeed connected to more DREs in both capture Hi-C data ( $P = 9.70 \times 10^{-4}$ , **Fig. 2B**) and FANTOM5 data ( $P = 3.52 \times 10^{-3}$ , **Fig. 2C**).

*ARGs are more likely to be differentially expressed in AD.* We next investigated differential gene expression of ARGs under different pathobiology contexts of AD. Specifically, we measured fold changes of microarray expression levels of ARGs compared to 571 local background genes (LBGs). We found that ARGs were more

likely to be differentially expressed in: (i) brain transcriptome of AD patients (**Fig. 2D** and **2E**); and (ii) mouse brain transcriptome from different AD transgenic mouse models (**Fig. 2F-2I**), including two transcriptomics datasets from 5XFAD mouse brain microglial cells (**Fig. 2F** and **2G**) and two tauopathy-related transcriptomics datasets (Tg4510) from mouse brain hippocampus (**Fig. 2H** and **2I**). We further performed differentially expressed gene enrichment analysis for network-predicted ARGs in AD. We collected bulk RNA-seq data from whole brain tissue or brain microglial cells from two common AD transgenic mouse models (5XFAD and Tg4510) and observed that ARGs were significantly differentially expressed in 5XFAD brain ( $P=0.003$ ), 5XFAD microglial cells ( $P=0.002$ ), and brain microglial of Tg4510 (**Supplementary Table 6**). This suggests that our identified ARGs are potentially involved in the pathobiology of AD.

*ARGs coding proteins are more likely to be differentially expressed in AD.* We further inspected differentially expressed proteins encoded by 103 network-predicted ARGs across 10 published proteomics datasets (see Supplementary Methods). Herein, we evaluated 3 types of AD transgenic mouse models: (a) hAPP model containing APP transgene, (b) 5XFAD model harboring human transgenes for both APP and PSEN1 mutations, and (c) ADLPAPT model carrying three human transgenes (APP, PSEN1 and tau). We found that products of ARGs were significantly differentially expressed in all 3 AD transgenic mouse models ( $P < 0.05$ , Fisher test, **Supplementary Table 6**).

Collectively, we have thus shown that network-predicted ARGs are

significantly involved in disease-related functional genomics, transcriptomics, and proteomics, supporting their role as likely causal genes for AD.

### **Incorporation of AD multi-omics data to prioritize ARGs**

To prioritize ARGs we integrated multi-omics profiles. In total, we incorporated 8 criteria that can be categorized into 5 types of biological evidence: (1) brain-expression specificity (z-score) derived from GTEx database, (2) availability of supportive experimental evidence from the literature and manually curated data from Open Targets database (Koscielny *et al.*, 2017), (3) experimentally validated AD genes, (4) differential gene expression, and (5) available drug targets. **Figure 3** shows a global view of 103 ARGs that we validated by these multiple forms of biological evidence in AD. Among 103 ARGs, 89 genes (86.4% [89/103]) satisfy at least one criterion. In addition, 15 ARGs have at least 5 forms of AD-related evidence, including 8 well-known AD genes: *APOE*, *PTK2B*, *NOS1*, *MEF2C*, *SORL1*, *EPHA5*, *ADAM10*, and *CD33*. For the rest of the ARGs, all but *BRSK1* had corresponding published literature-derived evidence. For example, *PAK1* is a predicted risk gene with 6 criteria of biological evidence: high brain expression specificity (z-score=2.83), supportive experimental evidence from the literature, druggable target data, and, differential expression in human brain of AD patients, microglial cells of 5XFAD mouse model, and brain hippocampus of a tau mouse model (**Fig. 3** and **Supplementary Table 7**). P21-activated kinase 1, encoded by the *PAK1* gene, has been implicated in AD (Zhao *et al.*, 2006), and recent studies have revealed that

inactivation of PAK1 obliterated social recognition without changing amyloid beta (A $\beta$ )/tau pathology, and also exacerbated synaptic impairment and behavioral deficits in mouse models of AD (Arsenault *et al.*, 2013, Bories *et al.*, 2017).

Among 103 ARGs, we selected 37 (**Supplementary Table 7**) using subject matter expertise based on a combination of factors: (i) high brain-expression specificity, (ii) differential expression in multiple AD transgenic mouse models; (iii) strength of the network-based prediction; and (iv) availability of supportive experimental evidence. To advance disease understanding of network-predicted high-confidence risk genes, we performed biological pathway enrichment analysis using ClueGO plugin in Cytoscape (**Supplementary Table 8**) (Bindea *et al.*, 2009). We found 4 statistically significant biological pathways in AD: (a) regulation of neurotransmitter transport, (b) A $\beta$ -related biologic process, (c) long-term synaptic potentiation, and (d) oxidative stress (**Table 1** and **Supplementary Table 9**).

*Neurotransmitter transport.* Specifically, MEF2C and RIMS1, encoding myocyte-specific enhancer factor 2C and regulating synaptic membrane exocytosis protein 1, play key roles in neurotransmitter secretion and synaptic plasticity. MEF2C (rs254776) has been reported in several GWAS studies (Allen *et al.*, 2015, Karch and Goate, 2015), and we found significantly lower expression of MEF2C in AD brain ( $P = 1.26 \times 10^{-3}$ , one side Wilcoxon test, **Supplementary Fig. 2A**). RIMS1 is a newly predicted ARG, and a recent proteome study from human hippocampus revealed its overexpression in AD (Hondius *et al.*, 2016). RIMS1 is significantly overexpressed in

5XFAD mouse microglia ( $P = 4.51 \times 10^{-3}$ , one side Wilcoxon test) compared to controls (**Supplementary Fig. 2B**).

*Beta-amyloid-related biologic process.* Five genes (APOE, ADAM10, CHRNA7, SORL1, and LRP2) are associated with beta-amyloid biologic process. Among them, APOE, ADAM10, and SORL1 are well-known AD risk genes, validated by large scale genetic studies and preclinical studies (Fjorback *et al.*, 2012, Lambert *et al.*, 2013, Kunkle *et al.*, 2019). For example, APP cleavage by ADAM10 will produce an APP-derived fragment that is neuroprotective, sAPP $\alpha$  (Peron *et al.*, 2018). CHRNA7 (neuronal acetylcholine receptor subunit alpha-7) and LRP2 (low-density lipoprotein receptor-related protein 2) are two newly identified risk genes. There is significantly lower expression of CHRNA7 in the Tg4510 mouse transcriptome ( $P = 4.55 \times 10^{-4}$ ) compared to controls (**Supplementary Fig. 2E**). CHRNA7 binds to A $\beta$  with a high affinity (Farhat and Ahmed, 2017). Finally, a previous study showed that the rs3755166 polymorphism within LRP2 is associated with susceptibility to AD in the Chinese population (Wang *et al.*, 2011).

*Long-term synaptic potentiation.* Mitogen-activated protein kinase (MAPK1) and PTK2B are two identified risk genes related to long-term synaptic potentiation. Mitogen-activated protein kinase 1, encoded by MAPK1 gene, is highly expressed in brain tissue (z-score = 3.51). The MAPK1 cascade can be activated by A $\beta$  via alpha7 nicotinic acetylcholine receptors (Dineley *et al.*, 2001), and significantly lower



expression of MAPK1 was found in Tg4510 mice ( $P = 8.27 \times 10^{-3}$ ) compared to controls (**Supplementary Fig. 2F**). PTK2B, a well-known AD gene with high expression in brain ( $z$ -score = 1.62), was identified as an early marker and *in vivo* modulator of tau pathology (Dourlen *et al.*, 2017), by mediating A $\beta$ -induced synaptic dysfunction and loss (Salazar *et al.*, 2019).

*Oxidative stress.* Oxidative stress is a prominent hypothesis in the pathogenesis of AD (Jiang *et al.*, 2016). Here we found four network-predicted ARGs (FOXO3, NOS1, NFKB1, and ESR1) that were associated with regulation of oxidative stress. FOXO3 encoding Forkhead box protein O3 transcription factor, is a direct substrate of CDK5. FOXO3 activates several genes (e.g. BCL2L11 and FASLG) to promote neuronal death and aberrant A $\beta$  processing (Shi *et al.*, 2016). Significantly lower expression of FOXO3 was found in 5XFAD mouse microglia ( $P = 8.62 \times 10^{-4}$ ) compared to controls (**Supplementary Fig. 2H**). NFKB1, encoding transcription factor nuclear factor kappa B (NF- $\kappa$ B), is implicated in oxidative stress, synaptic plasticity, and learning and memory (Snow and Albensi, 2016).

Taken together, these findings suggest that our network-predicted ARGs are involved in diverse pathobiological pathways of AD. However, experimental validations are warranted for several newly predicted ARGs.

### **High druggability of network-predicted ARGs**

To date, most disease genes generated from GWAS findings are undruggable (Okada

*et al.*, 2014). For example, a recent study revealed that none of approved and investigational AD drugs target products (proteins) of GWAS-derived genes in AD (Kwok *et al.*, 2018). We examined whether network-predicted ARGs were more druggable compared to randomly selected proteins from human protein-coding gene background. Based on drug-target networks from 6 commonly used resources (see **Methods**), we obtained 2,866 potential druggable proteins for FDA-approved or clinically investigational drugs. Surprisingly, we found that 41 out of 103 predicted ARGs (39.8 %) are known druggable proteins, which is four-fold higher than druggable proteins ( $P = 9.25 \times 10^{-11}$ , Fisher test) in the genome-wide human protein-coding gene background. High druggability of network-predicted ARGs offers more candidate targets for therapeutic discovery (such as drug repurposing) in AD. For example, ADRA2A, one of the predicted ARGs, encodes adrenoceptor alpha 2A receptor. ADRA2A is a potential target of clozapine (Philibin *et al.*, 2009), an atypical antipsychotic drugs. Long-term clozapine treatment reduces A $\beta$  deposition and improves cognitive impairment in an AD transgenic mice model (Choi *et al.*, 2017). NR1I3, encoding the nuclear receptor constitutive androstane receptor (CAR), is a potential drug target activated by the lipid-lowering drug simvastatin (Kobayashi *et al.*, 2005). Simvastatin was reported to significantly reduce levels of A $\beta$  *in vitro* and *in vivo* (Yamamoto *et al.*, 2016, Li *et al.*, 2018). In summary, network-predicted ARGs showed higher druggability compared to traditional GWAS-based analysis approaches. We next examined opportunities for drug repurposing by integrating findings from ARGs with the human protein-protein interactome network.

## **ARGs offer candidate targets for Alzheimer's drug repurposing**

We have shown that network-predicted ARGs are related to the known pathobiology of AD and offer potential druggable targets, prompting us to examine opportunities for AD therapeutic discovery. We hypothesized that for a drug with multiple targets to be beneficial for treating a disease, its target proteins should be within or in the immediate vicinity of the corresponding disease module (**Fig. 2A**) in the human interactome network. To examine the potential application of ARGs on AD drug repurposing, we applied a network proximity approach (Cheng and Desai, 2018) to quantify the interplay between AD modules from ARGs and drug targets in the human interactome network. We used the cutoff of Z score ( $Z < -1.5$ ) to select network-predicted repurposable drugs in AD. After exclusion of nutraceutical drugs, metal drugs, and radioactive diagnostic agents, 130 drug candidates were obtained. We then systematically retrieved the published anti-AD clinical, *in vitro/in vivo* reported data for the 130 predicted drugs. In total, 25 had corresponding preclinical or clinical evidence for potential application to AD (**Supplementary Table 10**). **Figure 4** shows the molecular mechanisms of the 25 predicted drug candidates with published experimental or clinical evidence for AD. These drugs are classified into 6 categories according to Anatomical Therapeutic Chemical classification (ATC) codes: musculoskeletal systems [n=6], genito urinary and hormones [n=5], cardiovascular [n=3], alimentary tract and metabolism [n=3], respiratory system [n=2], and others [n=6].

Among them, we found 4 predicted drugs having known AD clinical evidence (Galimberti and Scarpini, 2017, Lehrer and Rheinstein, 2018), including pioglitazone [NCT02913664], carvedilol [NCT01354444], febuxostat (Singh and Cleveland, 2018), and fluticasone (Lehrer and Rheinstein, 2018). Pioglitazone, an FDA approved drug for T2D, has a significant network proximity ( $Z = -1.64$ ) with the ARGs. **Figure 4** shows that pioglitazone targets six proteins by connecting with 12 neighborhoods of ARGs.

Several clinical trials conducted with pioglitazone to treat AD. A phase II study (NCT00982202) shows no statistically significant difference between controls and patients with mild to moderate AD (Geldmacher *et al.*, 2011). However, another study showed that pioglitazone was associated with cognitive and functional improvement, as well stabilization of AD in 42 diabetic patients (Sato *et al.*, 2011). Many of these studies were conducted in populations without biological confirmation of AD by biomarkers and in some case (e.g., the TOMMOW study; NCT01931566), the dose of pioglitazone was substantially lower than that used in clinical practice for the treatment of diabetes. The available clinical trial data do not exclude a beneficial effect of pioglitazone on AD. Except for the TOMMOW study that was conducted in cognitively normal at-risk individuals, other trials have examined symptomatic patients that address a question different from the risk-reduction interrogation we prosecuted. For these reasons, we chose pioglitazone to test the drug user's relationship with AD outcomes using state-of-the-art pharmacoepidemiologic analysis of large-scale patient data.

## **Network-predicted pioglitazone usage reduces risk of AD in patient data**

We selected pioglitazone, identified by the network proximity measure (closest), to assess the drug user's relationship with AD outcomes (**Supplementary Table 11**) by analyzing 7.23 million patients from the Medicare supplemental database (see **Methods**). Two comparison analyses were conducted to evaluate the predicted association based on individual level longitudinal patient data and state-of-the-art pharmacoepidemiologic methods. These included: (1) pioglitazone (n = 101,650) versus matched control population (n = 410,184), and (2) pioglitazone versus glipizide (a T2D drug, n = 191,656). **Table 2** summarizes the patient data for pharmacoepidemiologic analyses. For each comparison, we estimated the un-stratified Kaplan-Meier curves, conducted by both propensity score stratified (n strata = 10) log-rank test and Cox model. After 6 years of follow-up, pioglitazone significantly reduces risk of AD compared with match control population ( $P = 3.97 \times 10^{-4}$ , hazard ratio (HR) = 0.895, 95% confidence interval [CI] 0.841-0.951, **Fig. 5A** and **5C**). Importantly, propensity score matching cohort studies confirms that pioglitazone is associated with a reduced risk of AD in comparison to glipizide (HR = 0.921, 95% CI 0.861-0.983,  $P = 0.0146$ , **Fig. 5B** and **5C**). Thus, two comparison analyses support our network-based prediction.

## ***In vitro* observation of pioglitazone's mechanism-of-action in AD**

**Figure 5** reveals that pioglitazone significantly reduces risk of AD in patient-based

data. To investigate its mechanism-of-action in AD, we performed a network analysis through integration of drug targets and ARGs into the brain-specific PPI network (see **Methods**). Network analysis shows that pioglitazone potentially targets two tauopathy-related proteins (GSK3 $\beta$  and CDK5) in AD (**Fig. 6A**). RNA sequencing data from the GTEx database (GTEx Consortium, 2015) suggests that GSK3 $\beta$  and CDK5 are highly expressed in brain tissue. Accumulating studies suggested that inhibition of GSK3 $\beta$  and CDK5 activity is a potential therapeutic strategy for AD (Mazanetz and Fischer, 2007).

We next examined pioglitazone's mechanism-of-action on human microglia HMC3 cells. Firstly, to assess the potential cell cytotoxicity, HMC3 cells were treated with pioglitazone at various concentrations (0.03  $\mu$ M to 100  $\mu$ M) for 48 h, and cell viability was determined by MTT method. As presented in **Fig. 6B**, pioglitazone at 0.03  $\mu$ M to 10  $\mu$ M did not affect cell viability, revealing low toxicity in human cells. Thus, these optimized concentrations of pioglitazone ( $\leq$  10  $\mu$ M) were used in subsequent experiments. As shown in **Fig. 6C**, phosphorylation of GSK3 $\beta$  and CDK5 were significantly increased after LPS treatment (1  $\mu$ g/mL for 30 min) in HMC3 cells. Pre-treating with pioglitazone significantly reduced phosphorylated GSK3 $\beta$  and CDK5 in a dose-dependent manner (**Fig. 6D** and **6E** and **Supplementary Fig. 3**). Altogether, these data suggest that pioglitazone may offer potential benefits for patients with AD by reducing activation of GSK3 $\beta$  and CDK5.

## Discussion

AD risk involves a complex polygenic and pleiotropic genetic architecture (Long and Holtzman, 2019). Traditional reductionist paradigms overlook the inherent complexity of human disease and often led to treatments that are inadequate or have important adverse effects (Greene and Loscalzo, 2017). Understanding AD from the point-of-view of how cellular systems and molecular interactome perturbations underlie the disease is the essence of network medicine (Fang *et al.*, 2019). In this study, we argue that in the case of AD, cellular networks gradually evolve throughout disease initiation and progression, leading to progressive shifts of local and global network properties and systems states. It is these shifts that underlie the pathogenesis of AD. Genome alterations such as amplification, deletion, and mutation are primary events in AD. However, such events can be selected in human cells only if they encode the appropriate perturbations in the interactome network and systems properties of the affected cells. Therapeutic interventions need to be designed to deal with perturbations of AD systems properties, and have little to do, functionally speaking, with genetic and genomic events alone (Swarup *et al.*, 2019). Herein, we have proposed a genome-wide and population-based drug repurposing framework, which integrates genetic findings, functional genomics, drug-target networks, and the human protein-protein interactome, along with large-scale population-based validation and *in vitro* mechanistic observations in human microglia cells.

In total, we identified 103 ARGs by utilizing our recently developed Bayesian model selection method (Wang *et al.*, 2019). Human interactome network analysis

reveals that ARGs form a statistically significant disease module. Functional genomics enrichment analysis shows that ARGs harbor more gene regulatory elements in the human genome. In addition, large-scale transcriptomics and proteomics data analyses imply that ARGs are more likely to be differentially expressed in human AD brain and multiple AD transgenic mouse models (**Fig. 2**). These comprehensive observations suggest that ARGs potentially capture pathobiological pathways of AD (**Fig. 3**). Importantly, drug-target network analysis shows a 4-fold higher druggability compared to the known drug targets in the human genome.

A previous study showed that few products (proteins) of GWAS-derived closest genes could be applied for therapeutic discovery (Kwok *et al.*, 2018). Several factors may account for this. First, the reported significant loci occupy only a small proportion of heritability and provide limited information about underlying AD biology (Deming *et al.*, 2017). Second, many genome-wide significant loci lie in noncoding regions, and genes closest to index SNPs may not represent causal genes of AD (Zhang and Lupski, 2015). Thus, systematic identification of likely causal genes from GWAS findings using network approaches is a crucial step for understanding AD pathobiology and offers potential candidate targets for new therapeutic development as presented in this study.

Network-based drug repurposing from ARGs findings predict 4 highly repurposable drugs for AD, including pioglitazone (NCT02913664), carvedilol (NCT01354444), febuxostat and fluticasone. Carvedilol, an FDA approved drug for



hypertension that blocks the beta adrenergic receptor, significantly attenuates brain oligomeric  $\beta$ -amyloid level and cognitive deterioration in two independent AD mice models (Wang *et al.*, 2011). A propensity-matched analysis has suggested that a daily dose of 40 mg febuxostat can reduce the risk of dementia in older adults (Singh and Cleveland, 2018). Fluticasone is a non-steroidal anti-inflammatory drug (NSAID), and a recent study showed that long term use of fluticasone reduces incidence of developing AD (Lehrer and Rheinstein, 2018).

Pioglitazone, a U.S. FDA-approved anti-T2D drug, was reported to restore energy metabolism and reduce A $\beta$  levels in the brain of APP/PS1 mice (Chang *et al.*, 2019). A previous clinical study has shown that pioglitazone improves cognition and regional cerebral blood flow in patients with mild AD accompanied with T2D (Sato *et al.*, 2011). However, large-scale population-based validations have not been conducted to confirm the relationship of drug use with AD outcomes in the real-world patient data. In this study, by combining network-based prediction and population-based validation, we found that pioglitazone potentially reduced risk of AD in large-scale patient database (**Fig. 5**). In addition, *In vitro* mechanistic observations (**Fig. 6**) reveal that pioglitazone significantly downregulates expression of CDK5 and GSK3 $\beta$  in human microglia cells, mechanistically supporting network and population-based findings. However, a phase II study (NCT00982202) shows no statistically significant differences between controls and patients with mild to moderate AD for pioglitazone (Geldmacher *et al.*, 2011). One possible explanation is that pioglitazone reduces risk of AD only in patients with pre-existing diabetes or that pioglitazone may have its

effects before symptoms occur but not in more advanced patients. Thus, our findings suggest that larger clinical trials and additional mechanistic studies may be necessary to clarify pioglitazone's action in AD prevention in both a broad population and a well-defined sub-population.

Our network methodology presented here has several strengths. First, it contributes to identification of high-confidence likely causal genes, followed by multi-omics data validation, network-based drug repurposing investigation, large-scale patient data analysis, and *in vitro* mechanistic observation in human microglial cells. This work illustrates translation of GWAS findings to pathobiology and therapeutic development in AD. Second, the large patient-level longitudinal data ensures that our analyses integrate real-world patient evidence to test the drug's efficiency in AD risk reduction.

Potential weaknesses of this work should be acknowledged. First, catalogs of genetic variants from GWAS that influence human disease traits are far from complete, and this deficiency may affect the accuracy of identification of ARGs. Incompleteness of human interactome data and potential literature bias may influence performance of our network methodology. Our choice of validation models (e.g., specific type of transgenic mice models; specific cell types such as microglia) may have influenced our results. Detailed clinical information is missing for health insurance claims data regardless of high-dimensional covariate adjustment. This limits our ability to test the effects of pioglitazone on subpopulation of AD patients such as those with mild AD. In addition, although our dataset contains a geographically

diverse population of commercially-insured Americans seniors, the results are not representative of individuals who are not commercially-insured or uninsured. Finally, the phenotyping algorithms (e.g., using the ICD codes) may not capture all AD cases. Thus, this approach may need to be re-applied on a regular iterative basis as datasets are expanded, in order to offer maximum utility.

In summary, this study presents a powerful network-based methodology to translate GWAS findings to emerging therapeutic discovery by exploiting multi-omics, drug-target network, and the human protein-protein interactome, along with large-scale population-based and cell model-based validation. Approaches such as ours may minimize the translational gap between genetic discoveries and therapeutic development in AD and other complex diseases and assist in identifying urgently needed new therapies

## **Acknowledgements**

**Funding:** This work was supported by the National Heart, Lung, and Blood Institute of the National Institutes of Health (NIH) under Award Number K99 HL138272 and R00 HL138272 to F.C.

## **Author Contributions**

F.C. conceived the study. J.F., P.Z., and Q.W. performed experiments and data analysis. Y.Z., R.C., C.W.C, B.Z., B.L., S.J.L., A.A.P., L.L. and J.C., interpreted the data analysis. F.C., J.F., P.Z., and Q.W. drafted the manuscript. F.C., A.A.P., L.L., and

J.C. critically revised manuscript. All authors gave final approval of the manuscript.

## References

- Allen M, Kachadorian M, Carrasquillo MM, Karhade A, Manly L, Burgess JD, et al. Late-onset Alzheimer disease risk variants mark brain regulatory loci. *Neurol Genet*. 2015;1(2):e15.
- Alteri E, Guizzaro L. Be open about drug failures to speed up research. *Nature*. 2018;563(7731):317-9.
- Alzheimer A. Uber eine eigenartige erkrankung der hirnrinde. *Zentralbl Nervenpsych*. 1907;18:177-9.
- Anders S. Analysing RNA-Seq data with the DESeq package. *Mol Biol*. 2010;43(4):1-17.
- Andersson R, Gebhard C, Miguel-Escalada I, Hoof I, Bornholdt J, Boyd M, et al. An atlas of active enhancers across human cell types and tissues. *Nature*. 2014;507(7493):455-61.
- Arsenault D, Dal-Pan A, Tremblay C, Bennett DA, Guitton MJ, De Koninck Y, et al. PAK inactivation impairs social recognition in 3xTg-AD Mice without increasing brain deposition of tau and Abeta. *J Neurosci*. 2013;33(26):10729-40.
- Barbarino JM, Whirl-Carrillo M, Altman RB, Klein TE. PharmGKB: A worldwide resource for pharmacogenomic information. *Wiley Interdis Rev Syst Biol Med*. 2018;10(4):e1417.
- Bindea G, Mlecnik B, Hackl H, Charoentong P, Tosolini M, Kirilovsky A, et al. ClueGO: a Cytoscape plug-in to decipher functionally grouped gene ontology and pathway annotation networks. *Bioinformatics (Oxford, England)*. 2009;25(8):1091-3.
- Bories C, Arsenault D, Lemire M, Tremblay C, De Koninck Y, Calon F. Transgenic autoinhibition of p21-activated kinase exacerbates synaptic impairments and fronto-dependent behavioral deficits in an animal model of Alzheimer's disease. *Aging*. 2017;9(5):1386-403.
- Boyle EA, Li YI, Pritchard JK. An Expanded view of complex traits: From polygenic

to omnigenic. *Cell*. 2017;169(7):1177-86.

Chang KL, Wong LR, Pee HN, Yang S, Ho PC. Reverting metabolic dysfunction in cortex and cerebellum of APP/PS1 mice, a model for Alzheimer's disease by pioglitazone, a peroxisome proliferator-activated receptor gamma (PPARgamma) agonist. *Mol Neurobiol*. 2019;56(11):7267-7283.

Cheng F, Desai RJ. Network-based approach to prediction and population-based validation of in silico drug repurposing. *Nat Commun*. 2018;9(1):2691.

Cheng F, Lu W. A genome-wide positioning systems network algorithm for in silico drug repurposing. *Nat Commun*. 2019;10(1):3476.

Choi Y, Jeong HJ, Liu QF, Oh ST, Koo BS, Kim Y, et al. Clozapine Improves Memory Impairment and Reduces Abeta Level in the Tg-APP<sup>swe</sup>/PS1<sup>dE9</sup> Mouse Model of Alzheimer's Disease. *Mol Neurobiol*. 2017;54(1):450-60.

Conejero-Goldberg C, Hyde TM, Chen S, Dreses-Werringloer U, Herman MM, Kleinman JE, et al. Molecular signatures in post-mortem brain tissue of younger individuals at high risk for Alzheimer's disease as based on APOE genotype. *Mol Psychiat*. 2011;16(8):836-47.

Cook D, Brown D, Alexander R, March R, Morgan P, Satterthwaite G, et al. Lessons learned from the fate of AstraZeneca's drug pipeline: a five-dimensional framework. *Nat Rev Drug Discov*. 2014;13(6):419-31.

Cummings JL, Morstorf T, Zhong K. Alzheimer's disease drug-development pipeline: few candidates, frequent failures. *Alzheimers Res Ther*. 2014;6(4):37.

Cuyvers E, Sleegers K. Genetic variations underlying Alzheimer's disease: evidence from genome-wide association studies and beyond. *Lancet Neurol*. 2016;15(8):857-68.

Deming Y, Li Z, Kapoor M, Harari O, Del-Aguila JL, Black K, et al. Genome-wide association study identifies four novel loci associated with Alzheimer's endophenotypes and disease modifiers. *Acta Neuropathol*. 2017;133(5):839-56.

Dineley KT, Westerman M, Bui D, Bell K, Ashe KH, Sweatt JD. Beta-amyloid activates the mitogen-activated protein kinase cascade via hippocampal alpha7 nicotinic acetylcholine receptors: In vitro and in vivo mechanisms related to

- Alzheimer's disease. *J Neurosci*. 2001;21(12):4125-33.
- Dourlen P, Fernandez-Gomez FJ, Dupont C, Grenier-Boley B, Bellenguez C, Obriot H, et al. Functional screening of Alzheimer risk loci identifies PTK2B as an in vivo modulator and early marker of Tau pathology. *Mol Psychiat*. 2017;22(6):874-83.
- Eppig JT, Smith CL, Blake JA, Ringwald M, Kadin JA, Richardson JE, et al. Mouse Genome Informatics (MGI): Resources for mining mouse genetic, genomic, and biological data in support of primary and translational research. *Methods Mol Biol*. 2017;1488:47-73.
- Fang J, Pieper A, Lee G, Bekris L, Nussinov R, Leverenz JB, et al. Harnessing endophenotypes and using network medicine in Alzheimer's drug repurposing. *Med Res Rev*. 2019.
- Farhat SM, Ahmed T. Neuroprotective and neurotoxic implications of alpha7 nicotinic acetylcholine receptor and abeta interaction: therapeutic options in Alzheimer's disease. *Current Drug Targets*. 2017;18(13):1537-44.
- Fjorback AW, Seaman M, Gustafsen C, Mehmedbasic A, Gokool S, Wu C, et al. Retromer binds the FANSHY sorting motif in SorLA to regulate amyloid precursor protein sorting and processing. *J Neurosci*. 2012;32(4):1467-80.
- Galimberti D, Scarpini E. Pioglitazone for the treatment of Alzheimer's disease. *Expert Opin Investigat Drugs*. 2017;26(1):97-101.
- Geldmacher DS, Fritsch T, McClendon MJ, Landreth G. A randomized pilot clinical trial of the safety of pioglitazone in treatment of patients with Alzheimer disease. *Arch Neurol*. 2011;68(1):45-50.
- Granic I, Dolga AM, Nijholt IM, van Dijk G, Eisel UL. Inflammation and NF-kappaB in Alzheimer's disease and diabetes. *J Alzhe Dis*. 2009;16(4):809-21.
- Greene JA, Loscalzo J. Putting the patient back together - social medicine, network medicine, and the limits of reductionism. *N Engl J Med*. 2017;377(25):2493-9.
- Hondius DC, van Nierop P, Li KW, Hoozemans JJ, van der Schors RC, van Haastert ES, et al. Profiling the human hippocampal proteome at all pathologic stages of Alzheimer's disease. *Alzheimer's Dementia*. 2016;12(6):654-68.
- Huang Y, Fang J, Lu W, Wang Z, Wang Q, Hou Y, et al. A systems pharmacology

approach uncovers Wogonoside as an angiogenesis inhibitor of triple-negative breast cancer by targeting Hedgehog signaling. *Cell Chem Biol.* 2019;26(8):1143-58.e6.

Jansen IE, Savage JE, Watanabe K, Bryois J, Williams DM, Steinberg S. Genome-wide meta-analysis identifies new loci and functional pathways influencing Alzheimer's disease risk. *Nat Genet.* 2019;51(3):404-13.

Jiang T, Sun Q, Chen S. Oxidative stress: A major pathogenesis and potential therapeutic target of antioxidative agents in Parkinson's disease and Alzheimer's disease. *Progress Neurobiol.* 2016;147:1-19.

Jung YJ, Kim YH, Bhalla M, Lee SB. Genomics: New Light on Alzheimer's Disease Research. *Int J Mol Sci.* 2018;19(12).

Karch CM, Goate AM. Alzheimer's disease risk genes and mechanisms of disease pathogenesis. *Biol Psychiat.* 2015;77(1):43-51.

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 Neurodegen.* 2018;13(1):2.

Kobayashi K, Yamanaka Y, Iwazaki N, Nakajo I, Hosokawa M, Negishi M, et al. Identification of HMG-CoA reductase inhibitors as activators for human, mouse and rat constitutive androstane receptor. *Drug Met Dis.* 2005;33(7):924-9.

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(12):819.

Koscielny G, An P, Carvalho-Silva D, Cham JA, Fumis L, Gasparyan R, et al. Open Targets: a platform for therapeutic target identification and validation. *Nucleic Acids Res.* 2017;45(D1):D985-d94.

Kunkle BW, Grenier-Boley B, Sims R, Bis JC, Damotte V, Naj AC, et al. Genetic meta-analysis of diagnosed Alzheimer's disease identifies new risk loci and implicates Abeta, tau, immunity and lipid processing. *Nat Genet.* 2019;51(3):414-30.

Kwok MK, Lin SL, Schooling CM. Re-thinking Alzheimer's disease therapeutic targets using gene-based tests. *EBioMedicine.* 2018;37:461-70.

Lambert JC, Ibrahim-Verbaas CA, Harold D, Naj AC, Sims R, Bellenguez C, et al.

- Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease. *Nat Genet.* 2013;45(12):1452-8.
- Lehrer S, Rheinstein PH. Alzheimer's disease and intranasal fluticasone propionate in the FDA MedWatch adverse events database. *J Alzhe Dis Rep.* 2018;2(1):111-5.
- Li Y, Liu Q, Sun J, Wang J, Liu X, Gao J. Mitochondrial protective mechanism of simvastatin protects against amyloid beta peptide-induced injury in SH-SY5Y cells. *Int J Mol Med.* 2018;41(5):2997-3005.
- Li YH, Yu CY, Li XX, Zhang P, Tang J, Yang Q, et al. Therapeutic target database update 2018: enriched resource for facilitating bench-to-clinic research of targeted therapeutics. *Nucleic Acids Res.* 2018;46(D1):D1121-d7.
- Long JM, Holtzman DM. Alzheimer disease: An update on pathobiology and treatment strategies. *Cell.* 2019;179(2):312-39.
- Ma QL, Yang F, Frautschy SA, Cole GM. PAK in Alzheimer disease, Huntington disease and X-linked mental retardation. *Cell Logist.* 2012;2(2):117-25.
- Masters CL, Bateman R, Blennow K, Rowe CC, Sperling RA, Cummings JL. Alzheimer's disease. *Nat Rev Disease Pri.* 2015;1:15056.
- Mazanetz MP, Fischer PM. Untangling tau hyperphosphorylation in drug design for neurodegenerative diseases. *Nat Rev Drug Discov.* 2007;6(6):464-79.
- Mifsud B, Tavares-Cadete F, Young AN, Sugar R, Schoenfelder S, Ferreira L, et al. Mapping long-range promoter contacts in human cells with high-resolution capture Hi-C. *Nat Genet.* 2015;47(6):598-606.
- 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(5):48.
- 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(8):856-60.
- Nichols E, Szeke CEI, Vollset SE, Abbasi N, Abd-Allah F, Abdela J. Global, regional, and national burden of Alzheimer's disease and other dementias, 1990-2016: a systematic analysis for the Global Burden of Disease Study 2016. *Lancet Neurol.*



2019;18(1):88-106.

Okada Y, Wu D, Trynka G, Raj T, Terao C, Ikari K, et al. Genetics of rheumatoid arthritis contributes to biology and drug discovery. *Nature*. 2014;506(7488):376-81.

Orre M, Kamphuis W, Osborn LM, Jansen AHP, Kooijman L, Bossers K, et al. Isolation of glia from Alzheimer's mice reveals inflammation and dysfunction. *Neurobiol Aging*. 2014;35(12):2746-60.

Peron R, Vatanabe IP, Manzine PR, Camins A, Cominetti MR. Alpha-secretase ADAM10 regulation: insights into Alzheimer's disease treatment. *Pharmaceuticals (Basel, Switzerland)*. 2018;11(1).

Philibin SD, Walentyny DM, Vunck SA, Prus AJ, Meltzer HY, Porter JH. Further characterization of the discriminative stimulus properties of the atypical antipsychotic drug clozapine in C57BL/6 mice: role of 5-HT(2A) serotonergic and alpha (1) adrenergic antagonism. *Psychopharmacology*. 2009;203(2):303-15.

Polito VA, Li H, Martini-Stoica H, Wang B, Yang L, Xu Y, et al. Selective clearance of aberrant tau proteins and rescue of neurotoxicity by transcription factor EB. *EMBO Mol Med*. 2014;6(9):1142-60.

Salazar SV, Cox TO, Lee S, Brody AH, Chyung AS, Haas LT, et al. Alzheimer's Disease Risk Factor Pyk2 Mediates Amyloid-beta-Induced Synaptic Dysfunction and Loss. *J Neurosci*. 2019;39(4):758-72.

Sato T, Hanyu H, Hirao K, Kanetaka H, Sakurai H, Iwamoto T. Efficacy of PPAR-gamma agonist pioglitazone in mild Alzheimer disease. *Neurobiol Aging*. 2011;32(9):1626-33.

Savas JN, Wang YZ, 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(9):2614-27.

Shi C, Viccaro K, Lee HG, Shah K. Cdk5-Foxo3 axis: initially neuroprotective, eventually neurodegenerative in Alzheimer's disease models. *J Cell Sci*. 2016;129(9):1815-30.

Singh JA, Cleveland JD. Comparative effectiveness of allopurinol versus febuxostat for preventing incident dementia in older adults: a propensity-matched analysis. *Arth*

Res Ther. 2018;20(1):167.

Snow WM, Albeni BC. Neuronal Gene Targets of NF-kappaB and Their Dysregulation in Alzheimer's Disease. *Front Mol Neurosci*. 2016;9:118.

Swarup V, Hinz FI, Rexach JE, Noguchi KI, Toyoshiba H, Oda A, et al. Identification of evolutionarily conserved gene networks mediating neurodegenerative dementia. *Nat Med*. 2019;25(1):152-64.

Wang C, Zhang F, Jiang S, Siedlak SL, Shen L, Perry G, et al. Estrogen receptor-alpha is localized to neurofibrillary tangles in Alzheimer's disease. *Sci Rep*. 2016;6:20352.

Wang J, Ono K, Dickstein DL, Arrieta-Cruz I, Zhao W, Qian X, et al. Carvedilol as a potential novel agent for the treatment of Alzheimer's disease. *Neurobiol Aging*. 2011;32(12):2321.e1-12.

Wang LL, Pan XL, Wang Y, Tang HD, Deng YL, Ren RJ, et al. A single nucleotide polymorphism in LRP2 is associated with susceptibility to Alzheimer's disease in the Chinese population. *Clinica Chimica Acta*. 2011;412(3-4):268-70.

Wang M, Roussos P, McKenzie A, Zhou X, Kajiwara 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(1):104.

Wang Q, Chen R, Cheng F, Wei Q, Ji Y, Yang H, et al. A Bayesian framework that integrates multi-omics data and gene networks predicts risk genes from schizophrenia GWAS data. *Nat Neurosci*. 2019;22(5):691-9.

Wang Y, Cella M, Mallinson K, Ulrich JD, Young KL, Robinette ML, et al. TREM2 lipid sensing sustains the microglial response in an Alzheimer's disease model. *Cell*. 2015;160(6):1061-71.

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(D1):D1074-d82.

Won H, de la Torre-Ubieta L, Stein JL, Parikshak NN, Huang J, Opland CK, et al. Chromosome conformation elucidates regulatory relationships in developing human brain. *Nature*. 2016;538(7626):523-7.

Yamamoto N, Fujii Y, Kasahara R, Tanida M, Ohora K, Ono Y, et al. Simvastatin and atorvastatin facilitates amyloid beta-protein degradation in extracellular spaces by increasing neprilysin secretion from astrocytes through activation of MAPK/Erk1/2 pathways. *Glia*. 2016;64(6):952-62.

Zhang F, Lupski JR. Non-coding genetic variants in human disease. *Human Mol Genet*. 2015;24(R1):R102-10.

Zhao J, Cheng F, Jia P, Cox N, Denny JC, Zhao Z. An integrative functional genomics framework for effective identification of novel regulatory variants in genome-phenome studies. *Genome Med*. 2018;10(1):7.

Zhao L, Ma QL, Calon F, Harris-White ME, Yang F, Lim GP, et al. Role of p21-activated kinase pathway defects in the cognitive deficits of Alzheimer disease. *Nat Neurosci*. 2006;9(2):234-42.

**Table 1.** Network-predicted risk genes involved in four pathobiological pathways of Alzheimer's disease (AD).

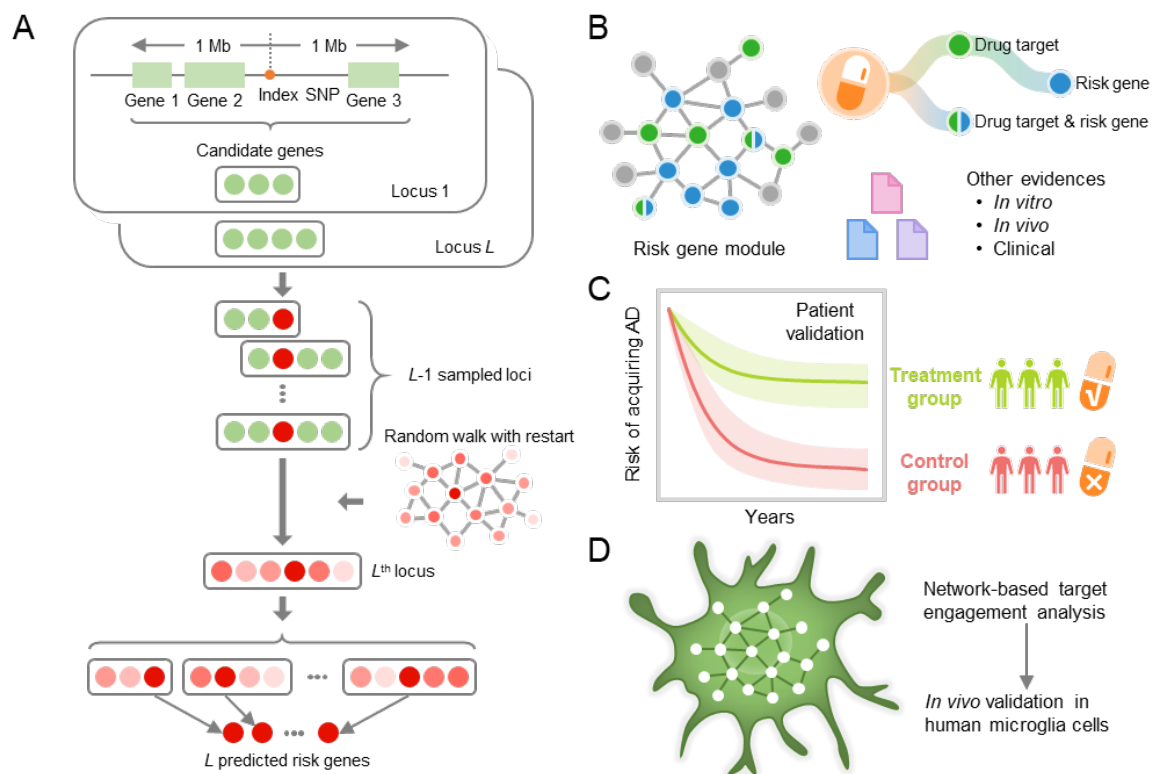
Gene	Protein	Description
<b>Neurotransmitter transport</b>		
MEF2C <sup>a</sup>	Myocyte-specific enhancer factor 2C	MEF2C mRNA expression levels were correlated with AD pathology
RIMS1	Regulating synaptic membrane exocytosis protein 1	An altered protein expression in RIMS1 during AD pathology
<b>Beta-amyloid-related biologic process</b>		
APOE <sup>a</sup>	Apolipoprotein E	Affect A $\beta$ production, aggregation, and clearance
CHRNA7	Neuronal acetylcholine receptor subunit alpha-7	Bind to A $\beta$ with very high affinity, providing therapeutic insight into AD
SORL1 <sup>a</sup>	Sortilin-related receptor	Reduce A $\beta$ generation by trafficking APP away from amyloidogenic cleavage sites
ADAM10 <sup>a</sup>	Disintegrin and metalloproteinase domain-containing protein 10	Constitutive $\alpha$ -secretase in the process of amyloid- $\beta$ protein precursor (A $\beta$ PP) cleavage
LRP2	Low-density lipoprotein receptor-related protein 2	rs3755166 polymorphism within LRP2 gene is associated with susceptibility to AD in the Chinese population
<b>Long-term synaptic potentiation</b>		
MAPK1	Mitogen-activated protein kinase 1	Beta-amyloid activates the MAPK cascade via hippocampal CHRNA7
PTK2B <sup>a</sup>	Protein-tyrosine kinase 2-beta	An <i>in vivo</i> modulator and early marker of Tau pathology.
<b>Oxidative stress</b>		
FOXO3	Forkhead box protein O3	Activate BCL2L11 and FASLG to promote neuronal death and aberrant A $\beta$ processing
NOS1	Nitric oxide synthase	Loss of endothelial NOS promotes p25 production and Tau phosphorylation
NFKB1	Nuclear factor NF-kappa-B p105 subunit	Involve in neuroinflammation, synaptic plasticity, learning, and memory implicated in AD
ESR1	Estrogen receptor	Interact with tau protein <i>in vivo</i> , and prevent glutamate excitotoxic injury by A $\beta$ through estrogen signaling mechanisms

<sup>a</sup> denotes that a gene is experimentally validated AD genes. The detailed literature information is provided in **Supplementary Table 9**.

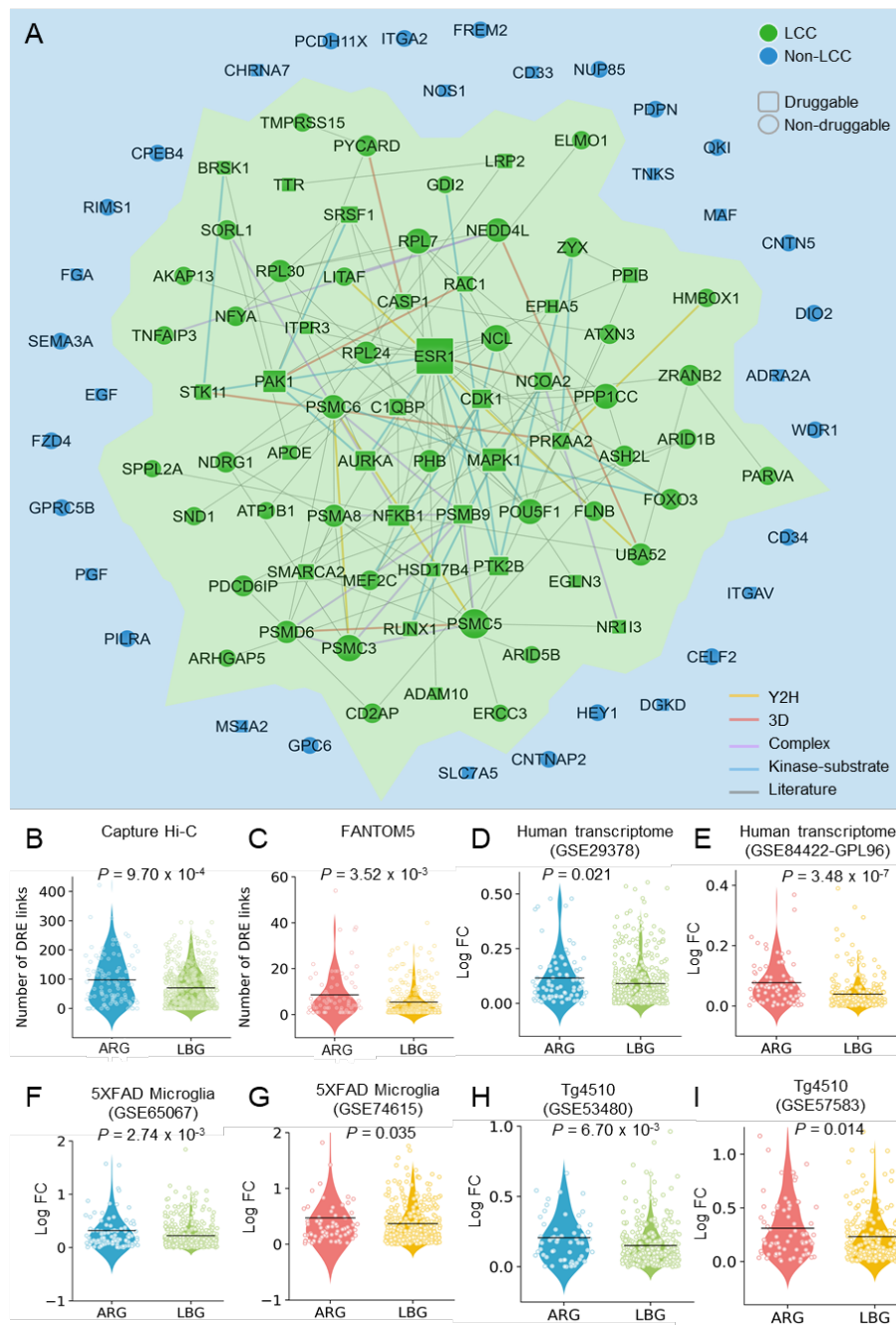
**Table 2.** Description of the dataset for the state-of-the-art pharmacoepidemiologic analysis in Alzheimer’s disease.

Drug	Sample Size	# of AD	Female (%)	Age (Mean)	Geographical location (%)					T2D (%)	HTN (%)
					Northeast	North Central	South	West	NA		
Pioglitazone	101,650	1,244	38.4	73.6	17.7	28.4	35.6	17.3	0.9	51.6	34.4
Glipizide	191,656	3,048	44.8	74.9	21.5	27.6	31.0	18.7	0.8	59.6	43.1
Control	402,184	5,498	38.4	75.46	24.6	31.6	30.3	12.4	1.0	45.0	41.4

We estimated the un-stratified Kaplan-Meier curves, conducted propensity score stratified (n strata = 10) log-rank test and Cox model. T2D: type 2 diabetes; HTN: hypertension.



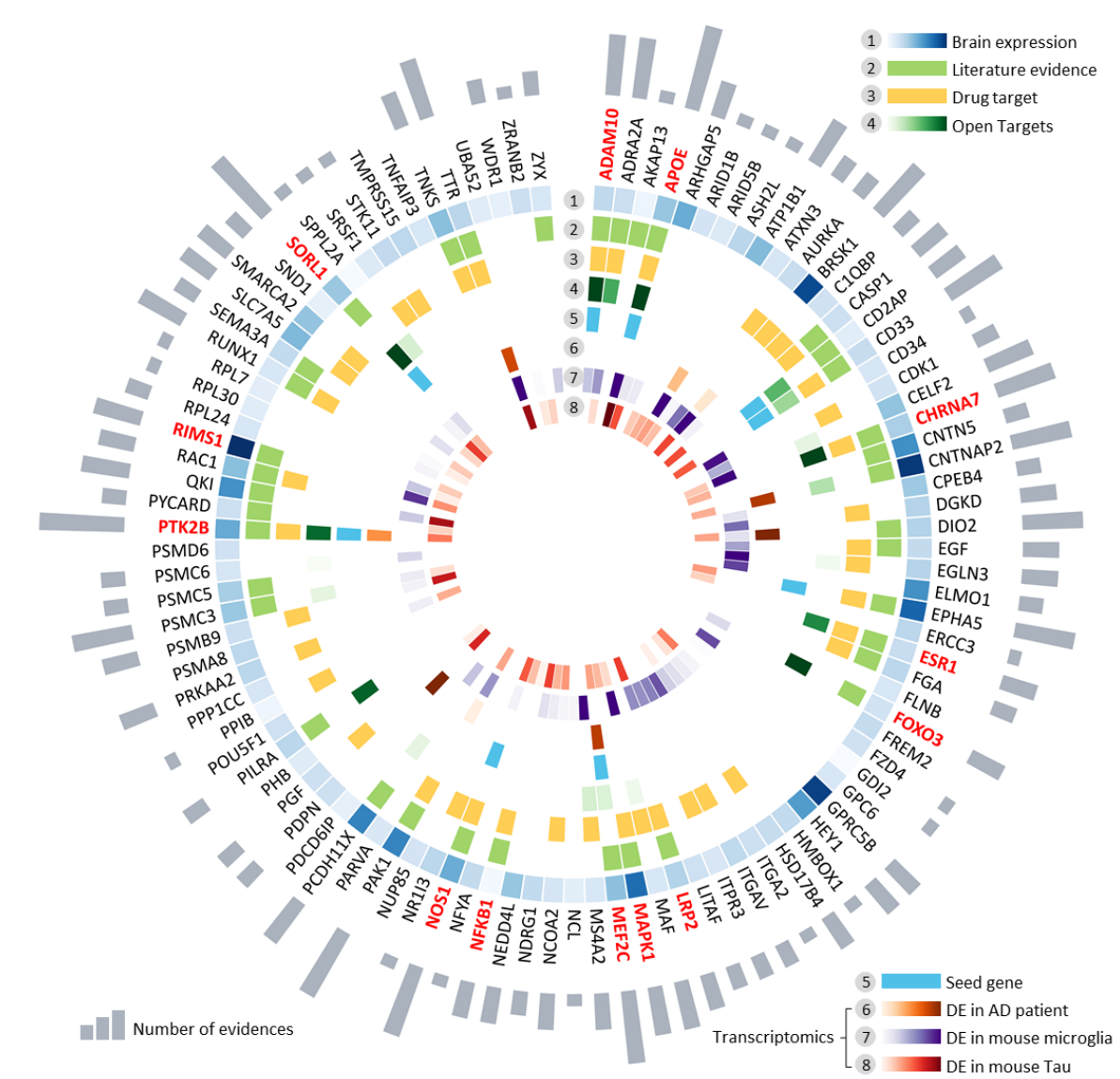
**Figure 1. A diagram illustrating a genotype-informed, network methodology and population-based validation for Alzheimer's therapeutic discovery. (A)** A framework of network-based Bayesian algorithm (see Methods) for identifying Alzheimer's disease (AD) risk genes. Specifically, this algorithm integrates multi-omics data and gene networks to infer risk genes from AD GWAS loci; **(B)** Network-based drug repurposing by incorporating ARGs and the human interactome network. **(C)** Population-based validation to test the drug user's relationship with AD outcomes; Comparison analyses were conducted to evaluate the predicted drug-AD association based on individual level longitudinal patient data and the state-of-the-art pharmacoepidemiologic methods (see Methods). **(D)** Network-based mechanistic observation. Experimental validation of network-predicted drug's proposed mechanism-of-action in human microglial cells.



**Figure 2. Network-based validation of predicted risk genes for Alzheimer's disease (AD).**

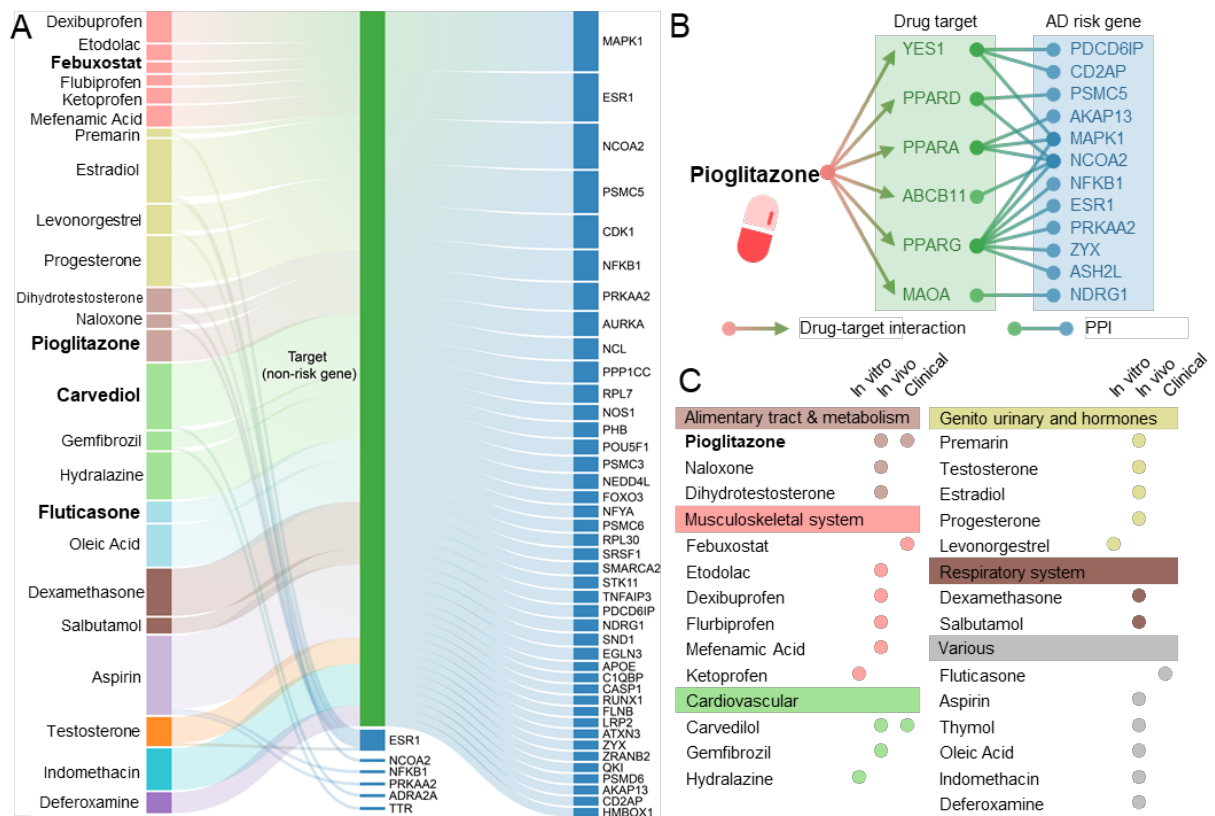
(A) A subnetwork highlighting disease module formed by predicted AD risk genes (ARGs) in the human protein-protein interactome. This disease module includes 128 protein-protein interactions (PPIs) (edges or links) connecting 70 ARGs (nodes). Larger node size highlighting the high expression level in brain compared to other tissues. (B-I) Discovery of genomic features of 103 predicted ARGs implicated in AD. ARGs capture strong distal gene regulatory elements in Hi-C (B) and FANTOM5 data (C) compared to a set of local background genes (LBGs). (D-I) ARGs are more likely differential expression across six transcriptomics datasets: (D and E) AD patient brain (GSE29378 [D] and GPL96 [E]), (F and G) brain microglia cell of 5XFAD mouse model (GSE65067 [F] and GSE74615 [G]), and (H and I) brain hippocampus of Tg4510 mouse model (GSE53480 [H] and GSE57583 [I]). *P* value was computed by one-tail T-test. LCC: Largest Connected Component.





**Figure 3. Multi-omics validation of network-predicted risk genes for Alzheimer's disease (AD).** Circle plot shows all 103 predicted AD risk genes validated by multiple-scale biological evidences. In total, 8 types of biological evidences were evaluated: (1) Brain-expression specificity derived from GTEx database (z-score>0 as a high brain-specific expressed gene); (2) literature evidence for the gene associated with AD; (3) drug target information; (4) literature-derived experimental data from Open targets database; (5) high quality experimentally validated AD-associated genes; (6) differential expression (DE) in AD patient brains; (7) differential expression in brain microglia cells of 5XFAD mouse model; (8) differential expression in brain hippocampus of Tg4510 mouse model. Gray bar denotes the number of biological evidences. 13 selected risk genes involved in four AD key pathways are highlighted by red: including regulation of neurotransmitter transport, A $\beta$  metabolic process, long-term synaptic potentiation, and oxidative stress.



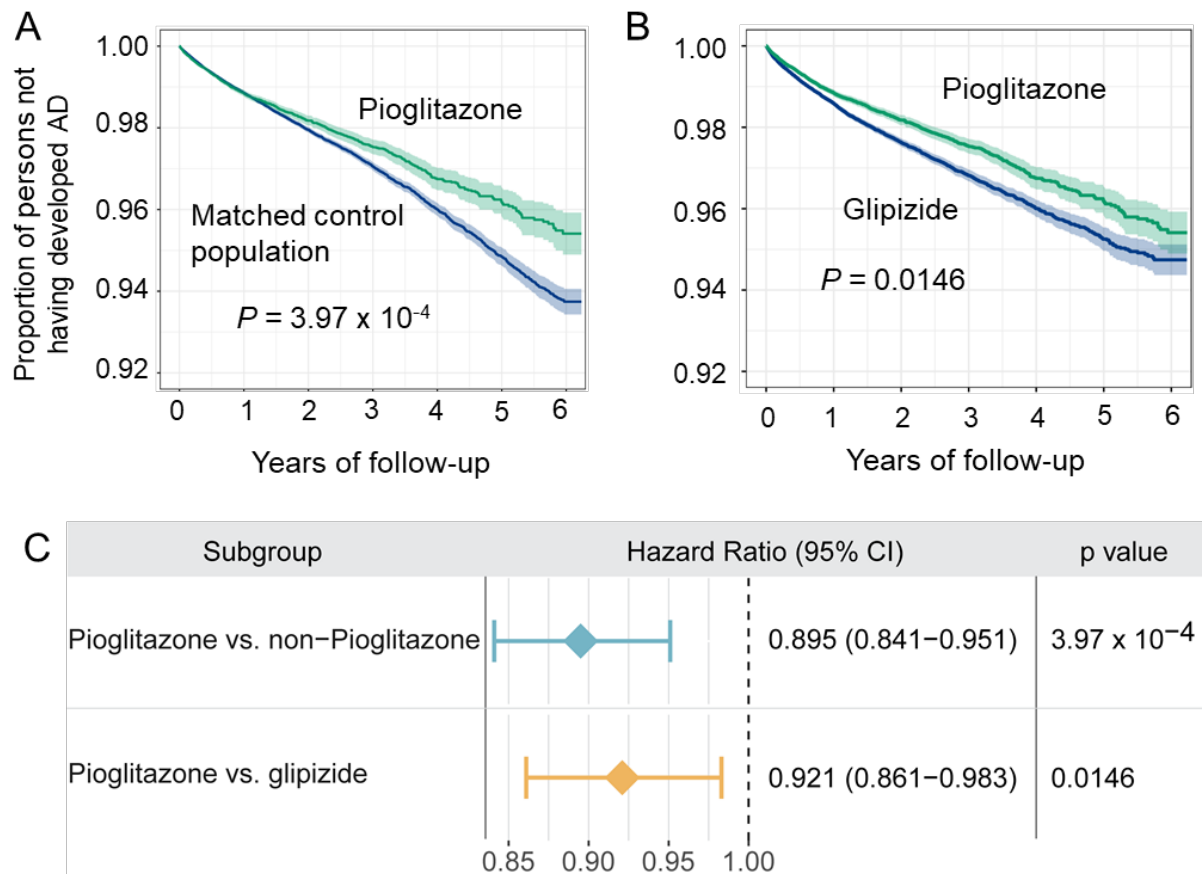


**Figure 4. Risk gene-informed drug repurposing for Alzheimer's disease (AD).** (A)

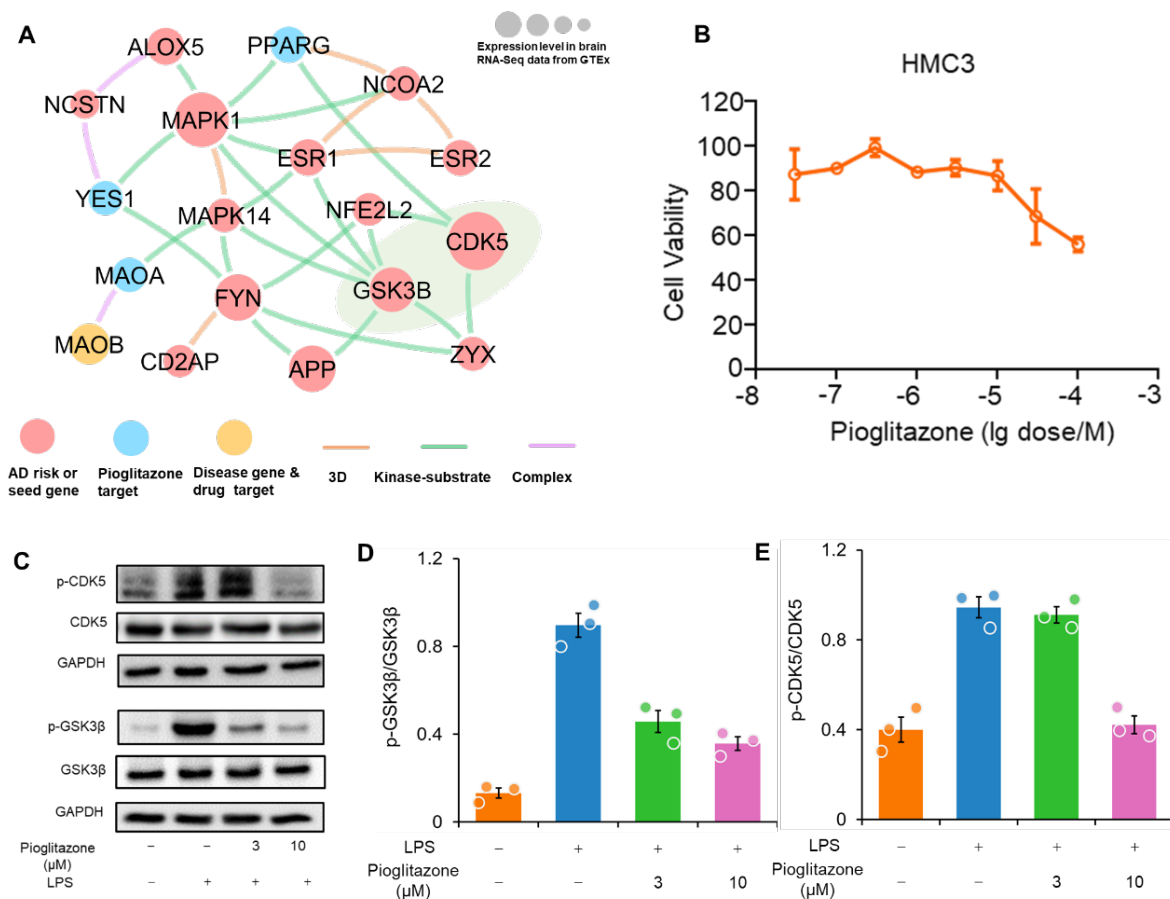
A Sankey diagram illustrates a global view of 25 repurposable drug candidates with published evidences for AD. These drugs are linked to their physical binding targets or neighborhood proteins derived from network-predicted AD risk genes. (B)

Network proximity analysis measures the network distance between disease module and drug targets in the human interactome. A subnetwork indicates the molecular mechanism of pioglitazone implicated in AD, which targets six physical binding proteins of which neighborhoods are 12 predicted AD risk genes. (C) Drugs are

grouped by their first-level Anatomical Therapeutic Chemical Classification (ATC) codes. The drugs with known anti-AD clinical status, *in vitro* and *in vivo* mouse model published data are given. Several approved drugs (pioglitazone, carvedilol, febuxostat and fluticasone) with anti-AD clinical evidence are highlighted.



**Figure 5. Longitudinal analyses reveal that pioglitazone reduces incidence of Alzheimer’s disease in patient data.** Two comparison analyses were conducted including: **(A)** pioglitazone (n = 101,650) vs. matched control population (n = 402,184), and **(B)** pioglitazone vs. glipizide (a diabetes drug, n = 191,656). First, for each comparison, we estimated the propensity score by using the variables described in **Table 2**. Then, we estimated the un-stratified Kaplan-Meier curves, conducted propensity score stratified (n strata = 10) log-rank test and Cox model. **(C)** Hazard ratios and 95% confidence interval (CI) for two cohort studies. Using propensity score stratified survival analyses, non-exposures were matched to the exposures (ratio 4:1) by adjusting the initiation time of pioglitazone, enrollment history, age and gender, and disease comorbidities (hypertension and diabetes).



**Figure 6. Experimental validation of pioglitazone's proposed mechanism-of-action in Alzheimer's disease (AD).** (A) Network analysis highlighting the inferred mechanism-of-action for pioglitazone in AD. The potential molecular mechanisms of pioglitazone were inferred through integration of known drug targets and predicted AD risk or AD seed genes into brain-specific co-expressed protein-protein interactome network (see the Methods). Node size indicates the protein-coding gene expression level in brain compared with other 31 tissues from GTEx database (GTEx Consortium, 2015). Larger size highlighting the high expression level in brain compared with other tissues. We excluded the literature-derived protein-protein interactions; (B) Effects of pioglitazone on the cell viability of HMC3 cells. HMC3 cells were treated with indicated concentrations of pioglitazone for 48 h and cell viability was determined using MTT. Data are represented as mean ± SEM (n = 3) and each experiment was performed at least three times in duplicate. (C) Effects of pioglitazone on LPS-induced activation of GSK3β (D) and CDK5 (E) in human microglia HMC3 cells. HMC3 cells were pre-treated with pioglitazone and followed LPS treatment (1 μg/mL, 30 min). The total cell lysates were collected and subjected to Western blot analysis. Quantification data represent mean ± s.d. of two independent experiments.