

Supplementary Materials and Method

1 Construction of human protein-protein interactome

To build a comprehensive human protein-protein interactome, we assembled data from 15 common resources with various kinds of experimental evidences. Specifically, we focus on high-quality PPIs with five types of data.

(1) Binary PPIs tested by high-throughput yeast-two-hybrid (Y2H) systems:

we combined binary PPIs tested from two public available high-quality Y2H datasets,^{1,2} and one in-house dataset;³ (2) Kinase-substrate interactions by

literature-derived low-throughput and high-throughput experiments from

KinomeNetworkX,⁴ Human Protein Resource Database (HPRD),⁵

PhosphoNetworks,⁶ PhosphositePlus,⁷ DbPTM 3.0,⁸ and Phospho. ELM.⁹

(3) Carefully 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 from BioGRID,¹⁰

PINA,¹¹ Instruct,¹² MINT,¹³ IntAct,¹⁴ and InnateDB.¹⁵ (4) Signaling network by

literature-derived low-throughput experiments as annotated in Signalink2.0.¹⁶

(5) Protein complexes data (56,000 candidate interactions) identified by a

robust affinity purification-mass spectrometry methodology were collected

from BioPlex V2.016.¹⁷ The genes were mapped to their Entrez ID based on

the NCBI database as well as their official gene symbols based on

GeneCards (<http://www.genecards.org/>). Duplicated pairs were removed.

Hence, inferred data, such as evolutionary analysis, gene expression data,

and metabolic associations, were excluded. The resulting updated human interactome used here includes 351,444 protein-protein interactions (PPIs) connecting 17,706 unique proteins.

2. Collection of Alzheimer's disease-associated genes

In order to collect high-quality experimentally validated genes (AD seed genes) for characterization of AD pathogenesis, we first manually curate experimentally validated (seed) genes in amyloidosis (amyloid) and tauopathy (tau). These genes satisfied at least one of the following criteria: i) gene validation in large-scale amyloid or tauopathy GWAS studies; ii) *in vivo* experimental model evidence that knockdown or overexpression of the gene leads to AD-like amyloid or tau pathology. Based on these criteria, we obtained 54 seed genes related to amyloid and 27 seed genes related to tauopathy. We further integrated a list of AD seed genes via assembling multiple data sources as below: i) 54 amyloid seed genes and 27 tauopathy seed genes; ii) 47 late-onset AD common-risk genes identified by large-scale genetic studies; iii) 35 genes with at least two AD-causing mutations from the Human Gene Mutation Database (HGMD) ¹⁸; iv) 39 disease genes curated in at least 2 of 4 following disease gene databases: HGMD (n=63), DisGeNET (n=23, score \geq 0.2) ¹⁹, MalaCards (n=79) ²⁰, and Open targets (n=81, overall score \geq 0.7 and literature score > 0) ²¹. After removing the duplicates, 144 AD seed genes were obtained (**Supplemental Table 4**).

3 Enrichment analysis

Bulk RNA-seq. We collected 2 RNA-seq datasets from brain or brain microglial of 5XFAD mouse model from two studies^{22,23}. The differential expression analysis was done with DESeq package in R²⁴. The threshold for significance of differential expression was set to a conservative statistical threshold of False Discovery Rate (FDR) < 0.05 and fold change (FC) \geq 1.2. After transferring to human-orthologous gene, we obtain two differentially expressed gene (DEG) sets for 5XFAD_brain (n=17) and 5XFAD_microglial (n=432).

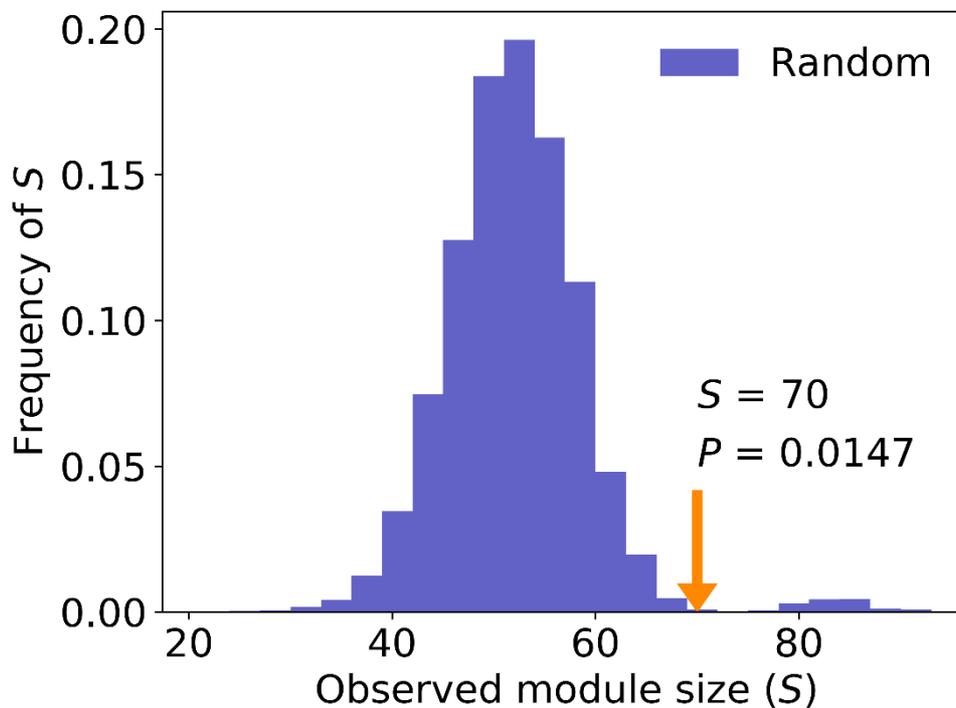
Furthermore, we collected 4 RNA-seq datasets from a recent study by Wang et al²⁵. This study performed a genome-wide RNAseq to reveal the molecular mechanisms underlying microglia activation in response to pathological tau perturbation in Tg4510 mouse model. Microglia cells were isolated from Tg4510 brain and gene expression was profiled using RNA sequencing²⁵. Four age groups of mice (2M, 4M, 6M, and 8M) were analyzed to capture longitudinal gene expression changes. Accession number for RNA-Seq data in Gene Expression Omnibus (GEO) is GSE123467. Given the larger number of DEGs identified by a threshold FC of 1.2, we used a stricter criterion (FDR < 0.05 and FC \geq 2) to determine DEGs compared with corresponding wide type mice. All DEGs identified in mice models were further mapped to unique human-orthologous genes using the Mouse Genome Informatics (MGI) database²⁶. Finally, we obtained 4 DEG sets for Tg4510, including Tg4510_2M (788), Tg4510_4M (629), Tg4510_6M (1013) and Tg4510_8M (792).

Proteomics. The ten sets of DEPs were assembled from 3 AD transgenic models in two recent publications ^{27,28}. The first publication by Savas et al. performed global quantitative proteomic analysis in hAPP and hAPP-PS1 mouse models at young (3 month [M]) and old ages (12 M) ²⁷. The samples were derived from frontal cortex (FC), hippocampal (HIP), and cerebellar (CB) extracts in mouse brain. The statistical significance of differential expression of all proteins was assessed using a two-tailed one-sample t-test on their corresponding peptide quantification ratios between both conditions. The obtained *P* values were FDR-adjusted for multiple hypothesis testing using the Benjamini–Hochberg correction ²⁹. DEPs for each brain region in mouse were determined with the threshold of FDR < 0.05 and transferred to homologous human gene ³⁰. We obtained four sets of DEPs [hAPP_3M (n=363), hAPP_12M (n=624), hAPP-PS1_3M (n=262) and hAPP-PS1_12M (n=476)] after merging the DEPs from different brain regions.

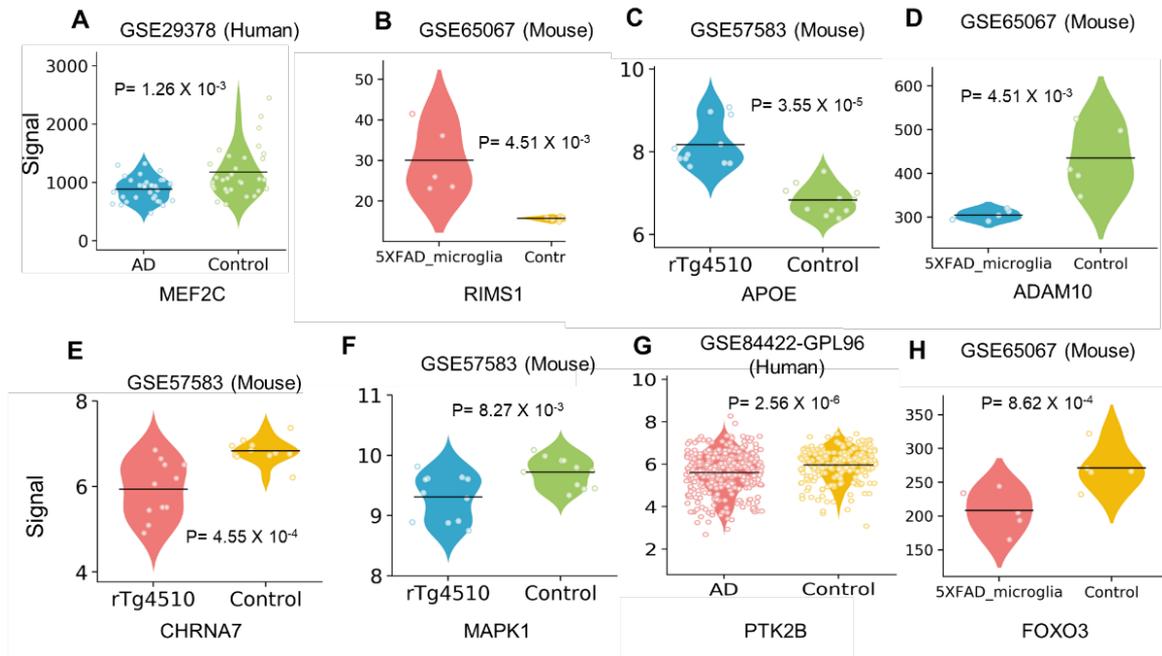
The other publication by Kim et al. performed quantitative proteomics to uncover molecular and functional signatures in HIP of two transgenic mouse, including ADLP_{APT} that carry three human transgenes (APP, PS1 and tau), and hAPP-PS1(5XFAD) mouse ²⁸. The ADLP_{APT} mice, generated by crossing the 5XFAD strain with JNPL3 tau animals, could exhibit amyloid plaques, accelerated neurofibrillary tangle formation, neuronal loss in the CA1 area, and memory deficit at an early age. The 5XFAD mouse model develops early plaque

formation, intraneuronal A β aggregation, neuron loss, and behavioral deficits.²² Three different ages of mouse were used, including young (4 M), middle (7 M), and old (10 M). The statistical cut-off value for significance was set to *P*-value <0.05 for the Student's t-test and fold-change >1.25. After transferring to homologous human gene³⁰, we obtained six sets of DEPs, including ADLP_{APT_4M} (n=53), ADLP_{APT_7M} (n=124), ADLP_{APT_10M} (n=299), hAPP-PS1_4M (n=54), hAPP-PS1_7M (n=168), and hAPP-PS1_10M (n=237).

Supplementary Figures

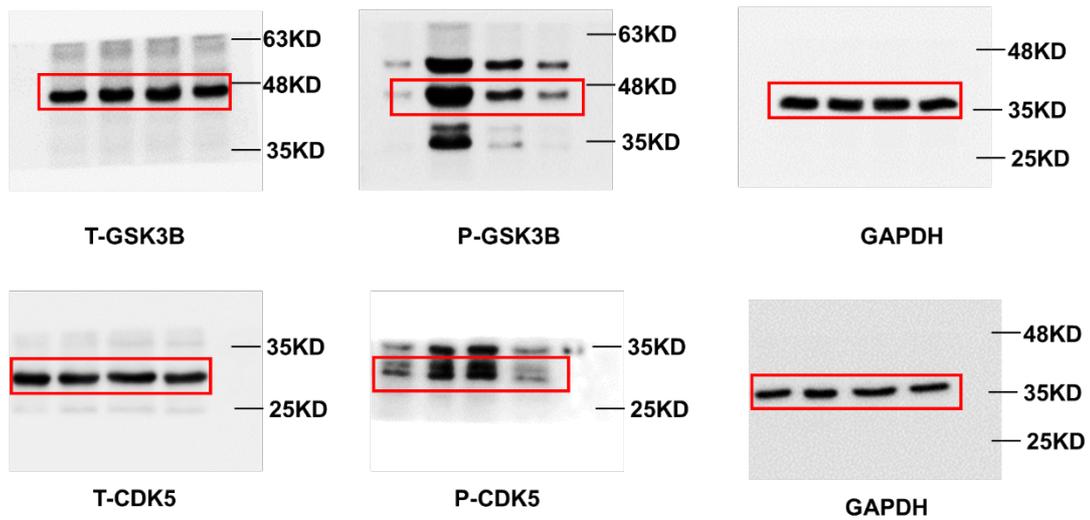


Supplementary Fig. 1. The largest connected component (LCC) analysis for Alzheimer's disease module. Experimentally identified AD genes form significantly disease module in the human interactome ($P = 0.0147$, 10,000 permutation test). Experimentally validated gene products (proteins) are likely to cluster in the same network neighborhood or disease module within the human protein-protein interactome. The observed module sizes (S), 70 (seed genes, orange line) are significantly larger than the random expectation.



Supplementary Fig. 2. Differential gene expression of the typical AD risk genes (ARGs) between AD and controls. The genes include MEF2C (A), RIMS1 (B), APOE (C), ADAM10 (D), CHRNA7 (E), MAPK1 (F), PTK2B (G), and FOXO3 (H). The signal value of AD (patient or mouse model) and controls were extracted from original microarray datasets in Gene Expression Omnibus (GEO) database. We calculated P value for each gene between AD and controls using one side Wilcoxon test.

Pioglitazone



Supplementary Fig.3. Effects of pioglitazone on LPS-induced activation of GSK3 β and CDK5 in human microglia HMC3 cells. HMC3 cells were pretreated with sildenafil and followed LPS treatment (1 μ g/mL, 30 min). The total cell lysates were collected and subjected to Western blot analysis. Quantification data represent mean \pm s.d. of two independent experiments.

Supplementary Table 11: phenotype definitions by ICD9/10 codes.

Alzheimer's disease^{31,32}

3310, F00, F000A, F001A, F002A, F009A, G30, G300, G301, G308, G309

Type 2 diabetes^{33,34}

25000, 25050, 25002, 25052, 25010, 25060, 25012, 25062, 25020, 25070, 25022, 25072, 25030, 25080, 25032, 25082, 25040, 25090, 25042, 25092, E089, E1100, E1101, E1110, E1121, E1122, E1129, E11311, E11319, E113219, E113291, E113292, E113293, E113299, E113319, E113391, E113392, E113393, E113399, E113419, E113491, E113492, E113493, E113499, E113519, E113591, E113592, E113593, E113599, E1136, E1139, E1140, E1142, E1143, E1144, E1149, E1151, E1152, E1159, E11610, E11618, E11620, E11621, E11628, E11630, E11641, E11649, E1165, E1169, E118, E119, E1300, E1310, E1311, E1322, E1329, E1339, E1349, E1351, E13628, E1365, E1369, E138, E139

Hypertension^{35,36}

4010, 4011, 4019, I10, I169

Supplementary References

- 1 Rolland, T. *et al.* A proteome-scale map of the human interactome network. *Cell* **159**, 1212-1226, doi:10.1016/j.cell.2014.10.050 (2014).
- 2 Rual, J. F. *et al.* Towards a proteome-scale map of the human protein-protein interaction network. *Nature* **437**, 1173-1178, doi:10.1038/nature04209 (2005).
- 3 Cheng, F. & Desai, R. J. Network-based approach to prediction and population-based validation of in silico drug repurposing. *Nature communications* **9**, 2691, doi:10.1038/s41467-018-05116-5 (2018).
- 4 Cheng, F., Jia, P., Wang, Q. & Zhao, Z. Quantitative network mapping of the human kinome interactome reveals new clues for rational kinase inhibitor discovery and individualized cancer therapy. *Oncotarget* **5**, 3697-3710, doi:10.18632/oncotarget.1984 (2014).
- 5 Keshava Prasad, T. S. *et al.* Human Protein Reference Database--2009 update. *Nucleic acids research* **37**, D767-772, doi:10.1093/nar/gkn892 (2009).
- 6 Hu, J. *et al.* PhosphoNetworks: a database for human phosphorylation networks. *Bioinformatics (Oxford, England)* **30**, 141-142, doi:10.1093/bioinformatics/btt627 (2014).

- 7 Hornbeck, P. V. *et al.* PhosphoSitePlus, 2014: mutations, PTMs and recalibrations. *Nucleic acids research* **43**, D512-520, doi:10.1093/nar/gku1267 (2015).
- 8 Lu, C. T. *et al.* DbPTM 3.0: an informative resource for investigating substrate site specificity and functional association of protein post-translational modifications. *Nucleic acids research* **41**, D295-305, doi:10.1093/nar/gks1229 (2013).
- 9 Dinkel, H. *et al.* Phospho.ELM: a database of phosphorylation sites--update 2011. *Nucleic acids research* **39**, D261-267, doi:10.1093/nar/gkq1104 (2011).
- 10 Oughtred, R. *et al.* The BioGRID interaction database: 2019 update. *Nucleic acids research* **47**, D529-d541, doi:10.1093/nar/gky1079 (2019).
- 11 Cowley, M. J. *et al.* PINA v2.0: mining interactome modules. *Nucleic acids research* **40**, D862-865, doi:10.1093/nar/gkr967 (2012).
- 12 Meyer, M. J., Das, J., Wang, X. & Yu, H. INstruct: a database of high-quality 3D structurally resolved protein interactome networks. *Bioinformatics (Oxford, England)* **29**, 1577-1579, doi:10.1093/bioinformatics/btt181 (2013).
- 13 Licata, L. *et al.* MINT, the molecular interaction database: 2012 update. *Nucleic acids research* **40**, D857-861, doi:10.1093/nar/gkr930 (2012).
- 14 Orchard, S. *et al.* The MIntAct project--IntAct as a common curation platform for 11 molecular interaction databases. *Nucleic acids research* **42**, D358-363, doi:10.1093/nar/gkt1115 (2014).
- 15 Breuer, K. *et al.* InnateDB: systems biology of innate immunity and beyond--recent updates and continuing curation. *Nucleic acids research* **41**, D1228-1233, doi:10.1093/nar/gks1147 (2013).
- 16 Csabai, L., Olbei, M., Budd, A., Korcsmaros, T. & Fazekas, D. Signalink: Multilayered Regulatory Networks. *Methods in molecular biology (Clifton, N.J.)* **1819**, 53-73, doi:10.1007/978-1-4939-8618-7_3 (2018).
- 17 Huttlin, E. L. *et al.* The BioPlex Network: A Systematic Exploration of the Human Interactome. *Cell* **162**, 425-440, doi:10.1016/j.cell.2015.06.043 (2015).
- 18 Stenson, P. D. *et al.* The Human Gene Mutation Database: towards a comprehensive repository of inherited mutation data for medical research, genetic diagnosis and next-generation sequencing studies. *Human genetics* **136**, 665-677, doi:10.1007/s00439-017-1779-6 (2017).
- 19 Pinero, J. *et al.* DisGeNET: a comprehensive platform integrating information on human disease-associated genes and variants. *Nucleic acids research* **45**, D833-d839, doi:10.1093/nar/gkw943 (2017).

- 20 Rappaport, N. *et al.* MalaCards: an amalgamated human disease compendium with diverse clinical and genetic annotation and structured search. *Nucleic acids research* **45**, D877-d887, doi:10.1093/nar/gkw1012 (2017).
- 21 Koscielny, G. *et al.* Open Targets: a platform for therapeutic target identification and validation. *Nucleic acids research* **45**, D985-d994, doi:10.1093/nar/gkw1055 (2017).
- 22 Bouter, Y. *et al.* Deciphering the molecular profile of plaques, memory decline and neuron loss in two mouse models for Alzheimer's disease by deep sequencing. *Frontiers in aging neuroscience* **6**, 75, doi:10.3389/fnagi.2014.00075 (2014).
- 23 Grubman, A. *et al.* Mouse and human microglial phenotypes in Alzheimer's disease are controlled by amyloid plaque phagocytosis through Hif1 α . *bioRxiv*, 639054 (2019).
- 24 Anders, S. Analysing RNA-Seq data with the DESeq package. *Mol Biol* **43**, 1-17 (2010).
- 25 Wang, H. *et al.* Genome-wide RNAseq study of the molecular mechanisms underlying microglia activation in response to pathological tau perturbation in the rTg4510 tau transgenic animal model. *Molecular neurodegeneration* **13**, 65 (2018).
- 26 Eppig, J. T. *et al.* in *Systems Genetics* 47-73 (Springer, 2017).
- 27 Savas, J. N. *et al.* Amyloid Accumulation Drives Proteome-wide Alterations in Mouse Models of Alzheimer's Disease-like Pathology. *Cell reports* **21**, 2614-2627, doi:10.1016/j.celrep.2017.11.009 (2017).
- 28 Kim, D. K. *et al.* Molecular and functional signatures in a novel Alzheimer's disease mouse model assessed by quantitative proteomics. *Molecular neurodegeneration* **13**, 2, doi:10.1186/s13024-017-0234-4 (2018).
- 29 Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal statistical society: series B (Methodological)* **57**, 289-300 (1995).
- 30 Eppig, J. T. *et al.* Mouse Genome Informatics (MGI): Resources for Mining Mouse Genetic, Genomic, and Biological Data in Support of Primary and Translational Research. *Methods in molecular biology (Clifton, N.J.)* **1488**, 47-73, doi:10.1007/978-1-4939-6427-7_3 (2017).
- 31 Wei, W. Q. *et al.* Combining billing codes, clinical notes, and medications from electronic health records provides superior phenotyping performance. *Journal of the American Medical Informatics Association : JAMIA* **23**, e20-27, doi:10.1093/jamia/ocv130 (2016).

- 32 Wilkinson, T. *et al.* Identifying dementia cases with routinely collected health data: A systematic review. *Alzheimer's & dementia : the journal of the Alzheimer's Association* **14**, 1038-1051, doi:10.1016/j.jalz.2018.02.016 (2018).
- 33 Kho, A. N. *et al.* Use of diverse electronic medical record systems to identify genetic risk for type 2 diabetes within a genome-wide association study. *Journal of the American Medical Informatics Association : JAMIA* **19**, 212-218, doi:10.1136/amiajnl-2011-000439 (2012).
- 34 Wei, W. Q. *et al.* Impact of data fragmentation across healthcare centers on the accuracy of a high-throughput clinical phenotyping algorithm for specifying subjects with type 2 diabetes mellitus. *Journal of the American Medical Informatics Association : JAMIA* **19**, 219-224, doi:10.1136/amiajnl-2011-000597 (2012).
- 35 Federman, D. G., Krishnamurthy, R., Kancir, S., Goulet, J. & Justice, A. Relationship between provider type and the attainment of treatment goals in primary care. *The American journal of managed care* **11**, 561-566 (2005).
- 36 Banerjee, D. *et al.* Underdiagnosis of hypertension using electronic health records. *American journal of hypertension* **25**, 97-102, doi:10.1038/ajh.2011.179 (2012).