

Supplementary material

Gene module-trait network analysis uncovers cell type specific systems and genes relevant to Alzheimer's Disease

Katia de Paiva Lopes¹, Ricardo A. Vialle¹, Gilad Green², Masashi Fujita³, Chris Gaiteri^{1,4}, Vilas Menon³, Julie A. Schneider¹, Yanling Wang¹, Philip L. De Jager³, Naomi Habib², Shinya Tasaki¹, David A. Bennett¹

¹ Rush Alzheimer's Disease Center, Rush University Medical Center, Chicago, IL, USA

² The Edmond and Lily Safra Center for Brain Sciences, The Hebrew University of Jerusalem, Israel

³ Center for Translational & Computational Neuroimmunology, Department of Neurology and the Taub Institute for Research on Alzheimer's Disease and the Aging Brain, Columbia University Irving Medical Center, New York, NY, USA

⁴ Department of Psychiatry, Upstate Medical University, Syracuse, NY, 13210, USA

Corresponding author:

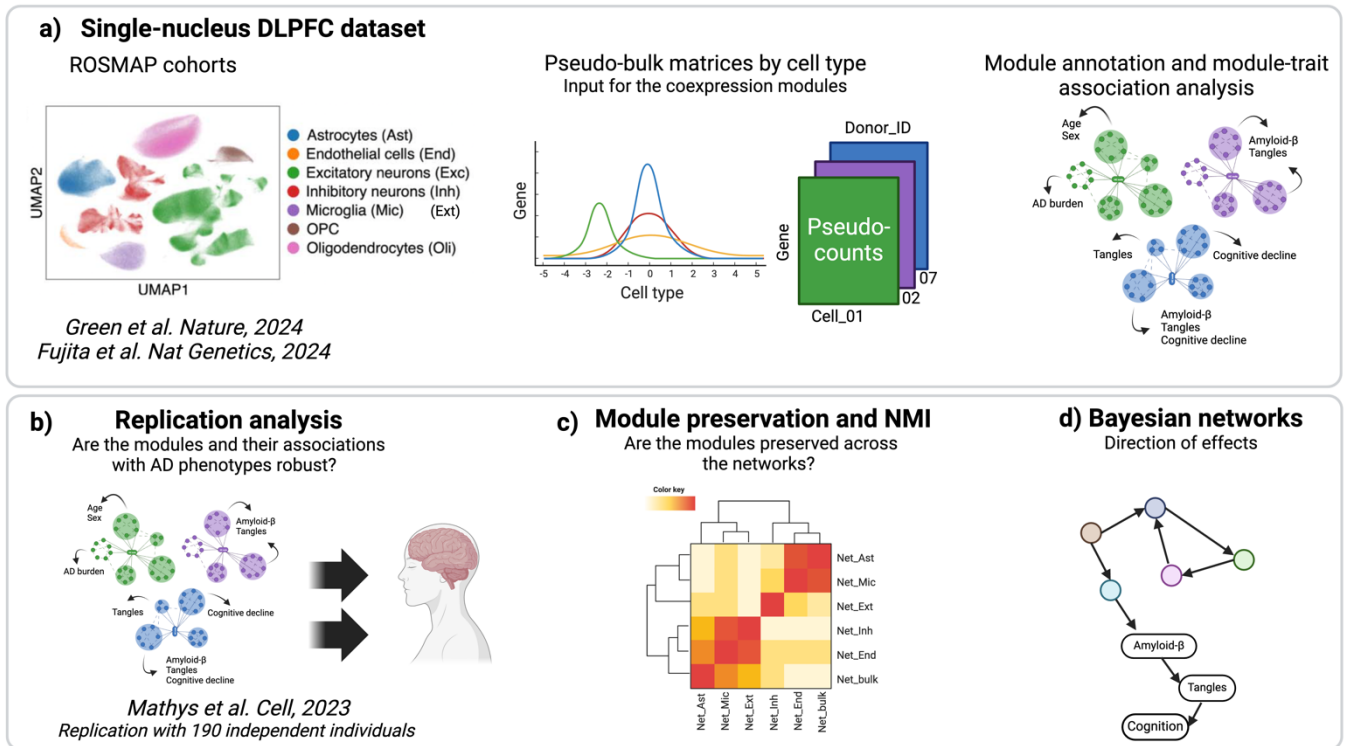
Katia de Paiva Lopes: katia_d_lopes@rush.edu

List of Supplementary Figures

Supplementary Figure 1: A graphic overview of the module-trait network (MTN) approach.....	4
Supplementary Figure 2: The percentage of snRNASeq cells by individual.	5
Supplementary Figure 4: The single-nucleus coexpression networks follow a scale-free topology.....	7
Supplementary Figure 5: The coexpression modules of the dorsolateral prefrontal cortex (DLPFC) bulk RNASeq (n=1,210 individuals).....	8
Supplementary Figure 6: Module preservation between the updated DLPFC (n=1,210 individuals) and Mostafavi et al. 2018 modules.	9
Supplementary Figure 7: Are the modules from Mostafavi et al. 2018 preserved in the updated DLPFC (n=1,210 individuals) network?	10
Supplementary Figure 8: DLPFC network distribution.....	11
Supplementary Figure 9: Heatmap showing the percentage of modules preserved across the coexpression networks.	12
Supplementary Figure 10: Comparison between module preservation against bulk RNASeq and functional enrichment.	13
Supplementary Figure 11: Do the excitatory neurons modules recapitulate sub-cell type clusters or states?	14
Supplementary Figure 12: Do the inhibitory neurons modules recapitulate sub-cell type clusters or states?	15
Supplementary Figure 13: Do the oligodendrocytes modules recapitulate sub-cell type clusters or states?	16
Supplementary Figure 14: Number of significant modules per network.....	17
Supplementary Figure 15: Replication analysis with an external dataset.	18
Supplementary Figure 16: Linear regression between the module's average expression and astrocyte cell fractions.	19
Supplementary Figure 17: The intra-module Bayesian network for the top 100 genes in the inh_M6.....	20
Supplementary Figure 18: The intra-module Bayesian network for the mic_M46.....	21
References.....	22

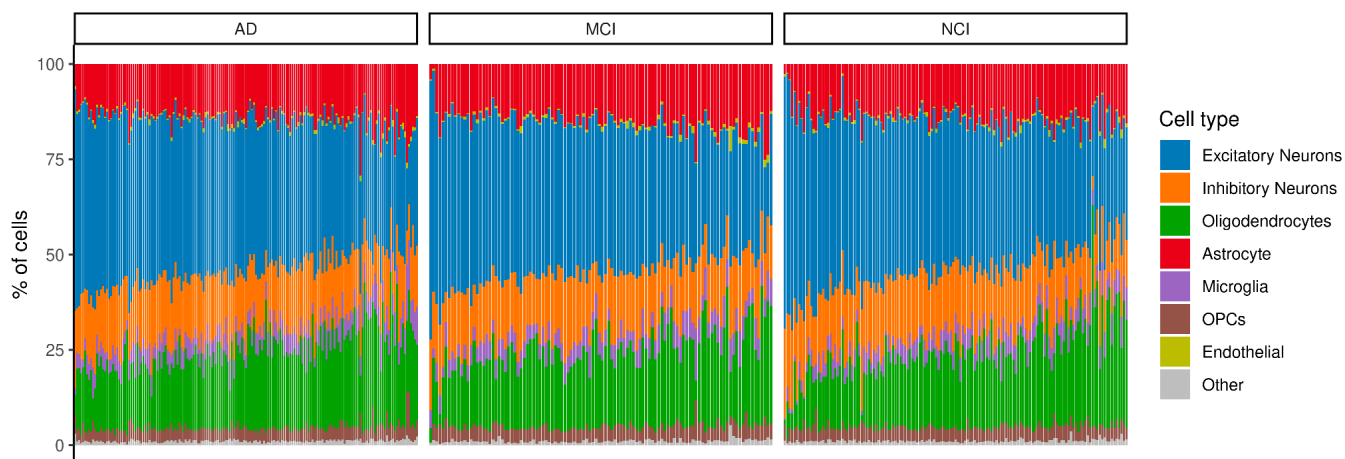
List of Supplementary Tables

- Table 1: Phenotype data for all the individuals used in the study (n = 1,210 unique donors).
- Table 2: Description of study participants with bulk RNASeq (n = 1,210).
- Table 3: Description of study participants with snRNASeq (n = 424).
- Table 4: Module assignment for the astrocytes co-expression networks.
- Table 5: Module assignment for the endothelial co-expression networks.
- Table 6: Module assignment for the excitatory neurons co-expression networks.
- Table 7: Module assignment for the inhibitory neurons co-expression networks.
- Table 8: Module assignment for the microglial co-expression networks.
- Table 9: Module assignment for the oligodendrocytes co-expression networks.
- Table 10: Module assignment for the OPCs co-expression networks.
- Table 11: Module assignment for the dorsolateral prefrontal cortex (DLPFC) co-expression networks.
- Table 12: Summarized functional enrichment analysis (FEA) for module characterization. Network = DLPFC.
- Table 13: Full summary statistics of the regression analysis between the DLPFC modules and the AD traits. "AE" indicates module average expression.
- Table 14: Normalized mutual information matrix results.
- Table 15: Full summary statistics of module preservation (MP) analysis across the 9 networks tested.
- Table 16: Zsummary statistics from the MP analysis between the single-nucleus modules and the bulk DLPFC (n=1,210).
- Table 17: Summarized Functional Enrichment Analysis (FEA) for module characterization. Network = astrocytes.
- Table 18: Summarized FEA for module characterization. Network = microglial.
- Table 19: Summarized FEA for module characterization. Network = excitatory neurons.
- Table 20: Summarized FEA for module characterization. Network = inhibitory neurons.
- Table 21: Summarized FEA for module characterization. Network = oligodendrocytes.
- Table 22: Summarized FEA for module characterization. Network = OPCs.
- Table 23: Summarized FEA for module characterization. Network = endothelial cells.
- Table 24: Fisher exact test between the modules and the damaged lipid-associated microglia and macrophages (DLAM).
- Table 25: Jaccard similarity results for the astrocyte modules versus astrocyte cells subpopulations.
- Table 26: Jaccard similarity results for the microglia modules versus microglial cells subpopulations.
- Table 27: Jaccard similarity results for the excitatory neuron modules versus excitatory neurons subpopulations.
- Table 28: Jaccard similarity results for the inhibitory neuron modules versus inhibitory neurons subpopulations.
- Table 29: Jaccard similarity results for the oligodendrocytes modules versus oligodendrocytes subpopulations.
- Table 30: Full summary statistics of the regression analysis between the module and the AD traits. "AE" indicates module average expression.
- Table 31: Replication results. Full summary statistics of the regression analysis between the module and the AD traits for the independent samples (n=190) in the MIT snRNASeq.
- Table 32: Full results for the single-nucleus Bayesian network (BN). "ME" indicates module eigengene.
- Table 33: Filtered results for the single-nucleus BN. We keep edges with minimum 3.3 of weight, after 500 permutations. "ME" indicates module eigengene.
- Table 34: Full summary statistics for the intra-BN of inh_M6.
- Table 35: Full summary statistics for the intra-BN of mic_M46.
- Table 36: Full summary statistics for the intra-BN of ast_M19.
- Table 37: Comparison of gene-pair correlation matrices before and after batch correction.
- Table 38: Rummagene enrichment results for microglia module 46 (mic_m46).



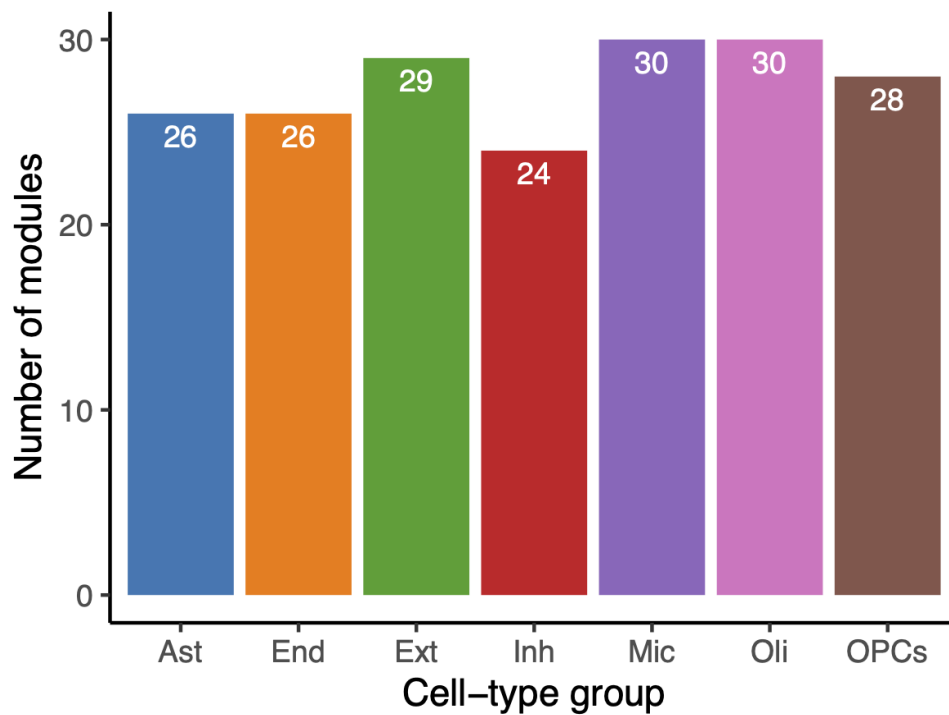
Supplementary Figure 1: A graphic overview of the module-trait network (MTN) approach.

a) The ROSMAP snRNASeq data used for the networks covers 1.6 million cells, 7 major cell-type groups and 96 sub-cell types (Fujita et al. 2024; Green et al. 2024). Pseudo-bulk matrices were created and used as input for the coexpression networks. The modules were annotated according to their biological functions and their average expression used for trait-association analysis. b) Replication analysis was performed with 190 independent participants (Mathys et al. 2023). c) Pairwise measures of module preservation and normalized mutual information were calculated to identify specific or shared systems across all the networks. d) Bayesian networks were modeled for causal inference (Tasaki et al. 2015). Figure generated with biorender.



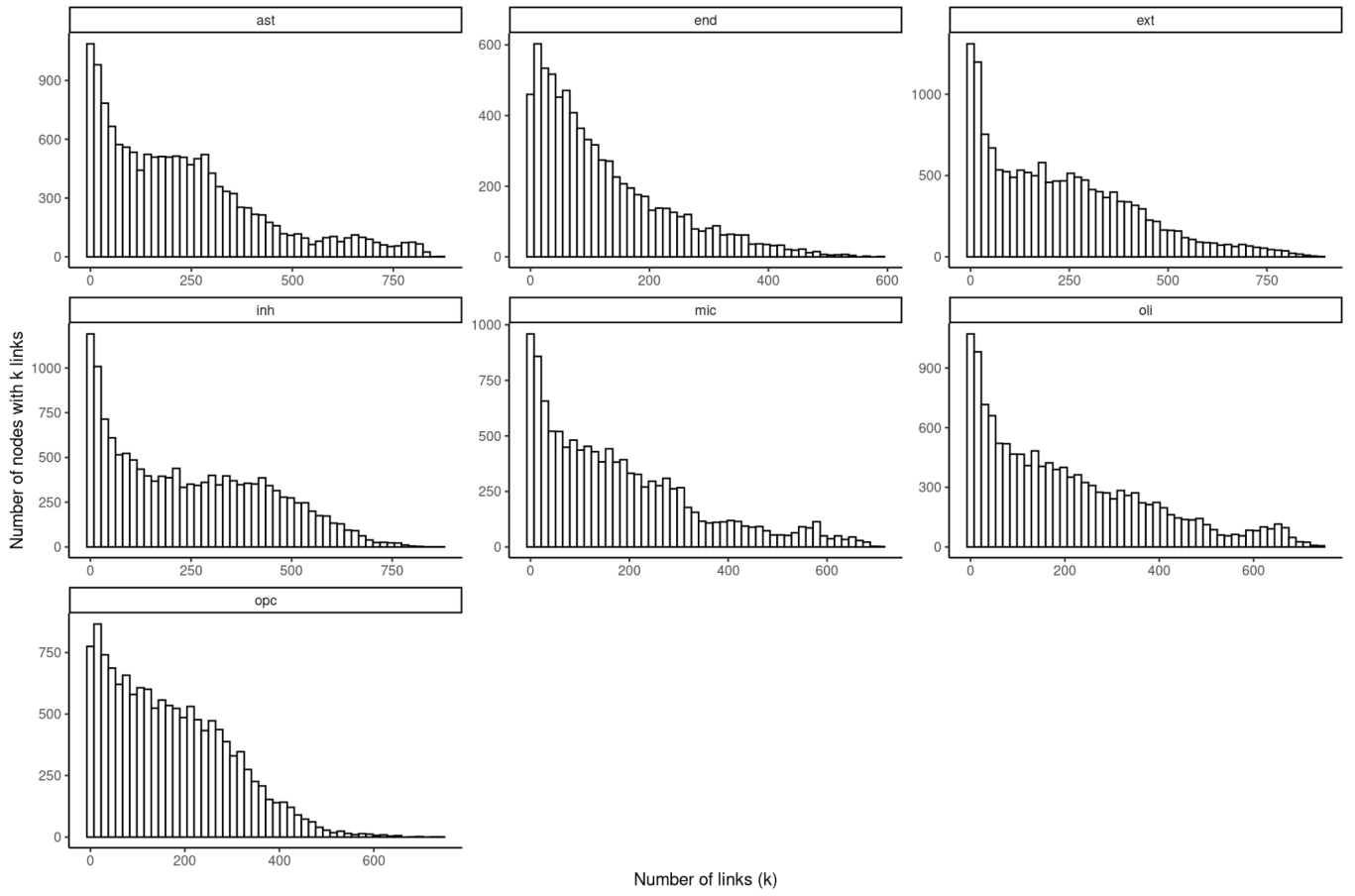
Supplementary Figure 2: The percentage of snRNASeq cells by individual.

Excitatory neurons accounts for 39.54% of the DLPFC cells, followed by Inhibitory neurons: 15.68%, Oligodendrocytes: 20.36%, Astrocyte: 13.97%, Microglia: 5.12%, Endothelial: 0.63%. AD: Clinical diagnosis of Alzheimer's dementia, MCI: Mild cognitive impairment, NCI: No cognitive impairment.



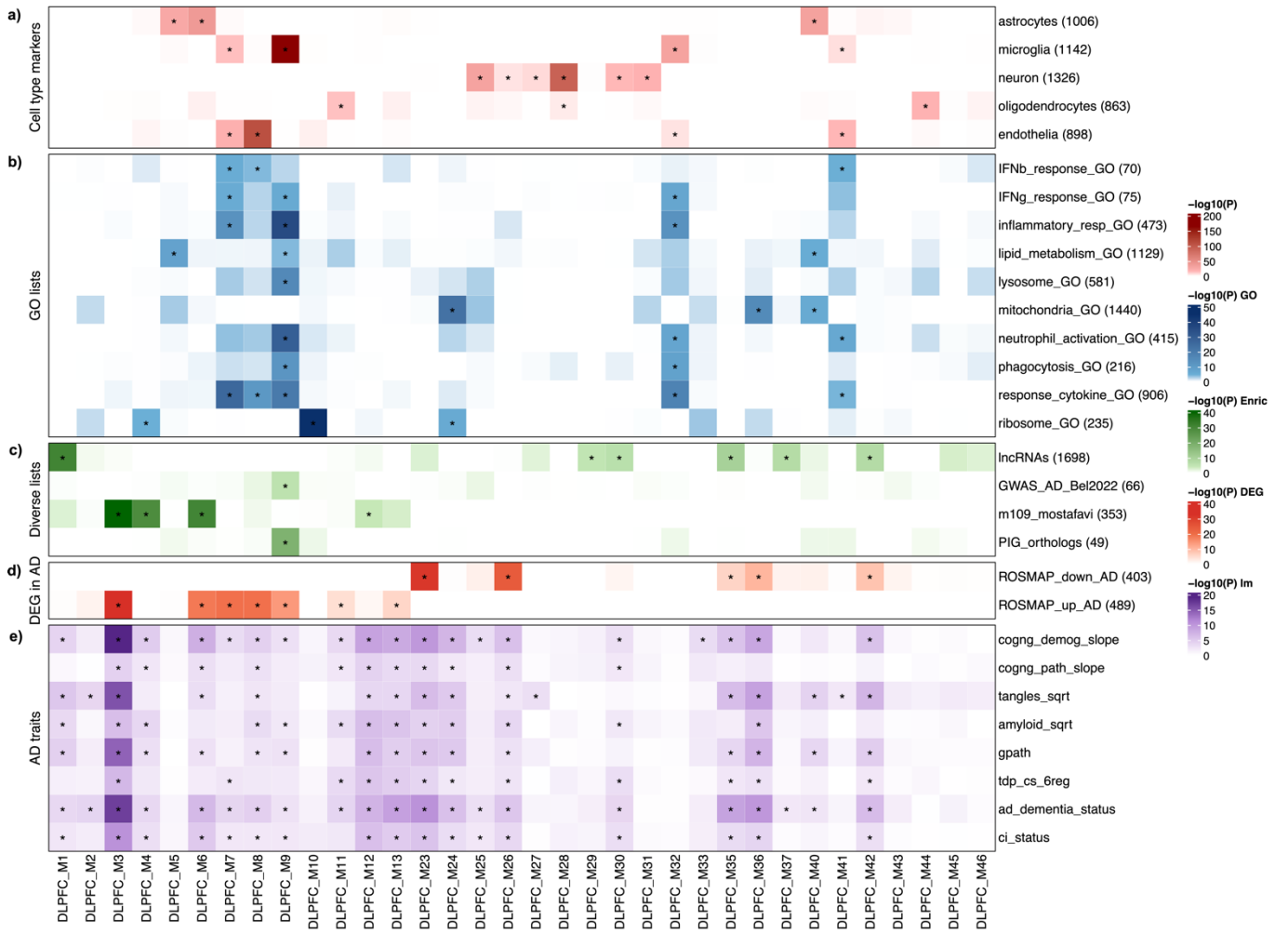
Supplementary Figure 3: Number of modules by network.

Ast: Astrocytes, end: Endothelial cells, ext: Excitatory neurons, Inh: inhibitory neurons, mic: Microglia, oli: Oligodendrocytes, and opc: Oligodendrocyte precursor cells.



Supplementary Figure 4: The single-nucleus coexpression networks follow a scale-free topology.

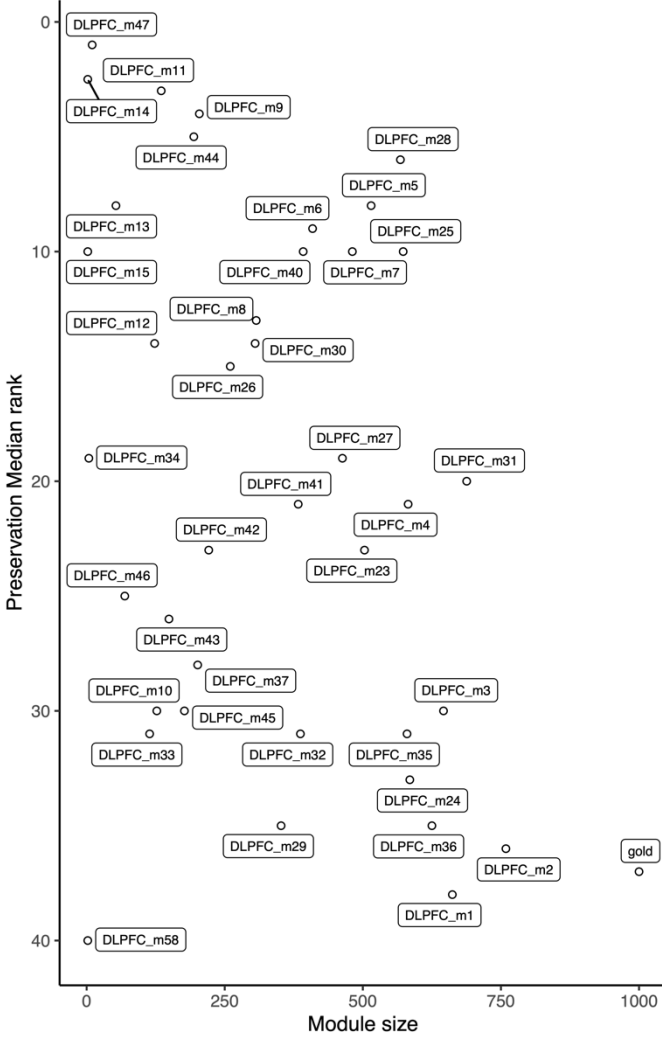
X-axis for the number of links (connections) and Y-axis for the number of nodes, for each network. Ast: Astrocytes, end: Endothelial cells, ext: Excitatory neurons, Inh: inhibitory neurons, mic: Microglia, oli: Oligodendrocytes, and opc: Oligodendrocyte precursor cells.



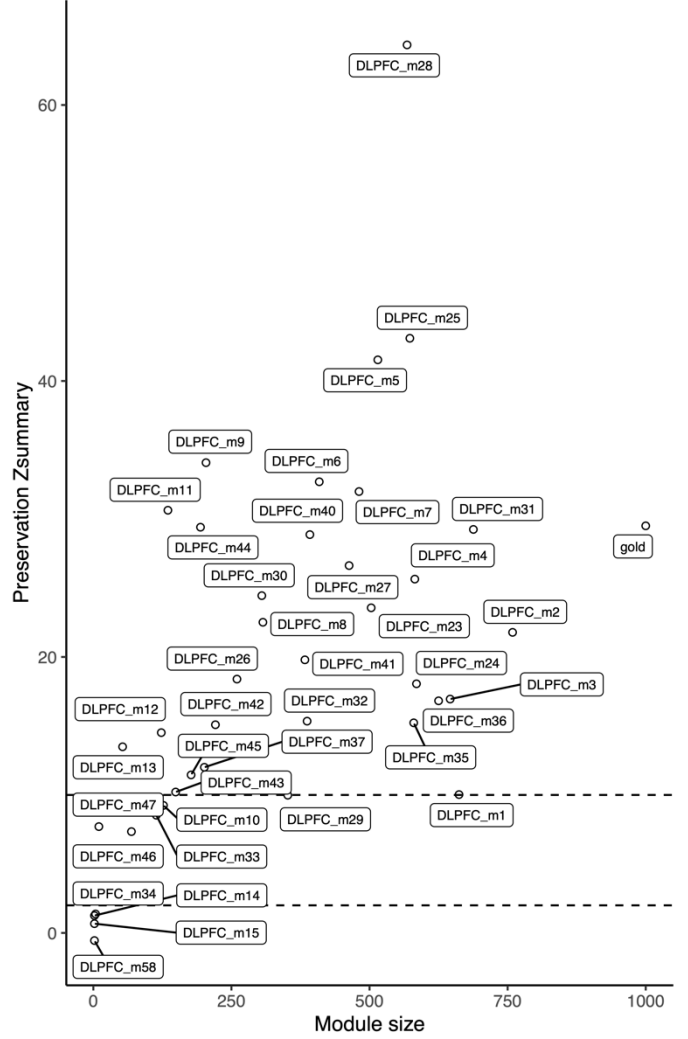
Supplementary Figure 5: The coexpression modules of the dorsolateral prefrontal cortex (DLPFC) bulk RNASeq (n=1,210 individuals).

a-d) Fisher exact test results between gene lists and the module average expression. The lists include cell marker genes from Johnson et al. 2022 (Johnson et al. 2022); gene ontology (GO) terms (Ashburner et al. 2000; Gene Ontology Consortium et al. 2023); lncRNAs (Gencode 37); AD GWAS from Bellenguez et al. 2022 (Bellenguez et al. 2022); module 109 previously published by Mostafavi et al., 2018 (Mostafavi et al. 2018); Plaque-induced gene list from Chen et al., 2022 (Chen et al. 2020); genes down and up-regulated in AD from the bulk dataset with 1,210 participants. The P -values were adjusted by computing multiple testing corrections, and asterisks highlight modules with Bonferroni P -value < 0.05 . e) Association results between the module's average expression and AD traits. The P -values are from the linear or logistic regression after adjustment by age, sex, and years of education. Asterisks indicate that the results reached the significance threshold (Bonferroni P -value < 0.05).

a) Preservation Median rank

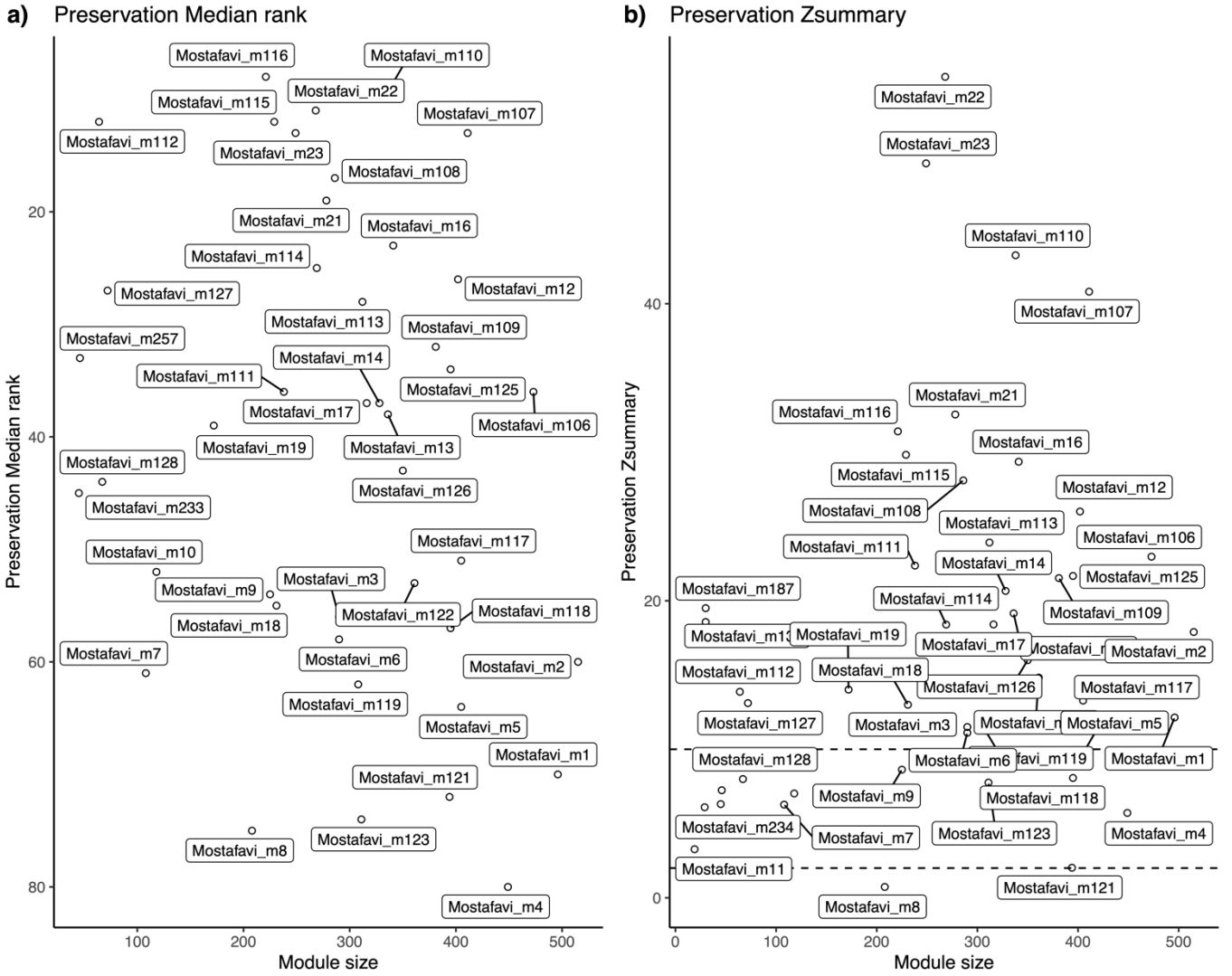


b) Preservation Zsummary



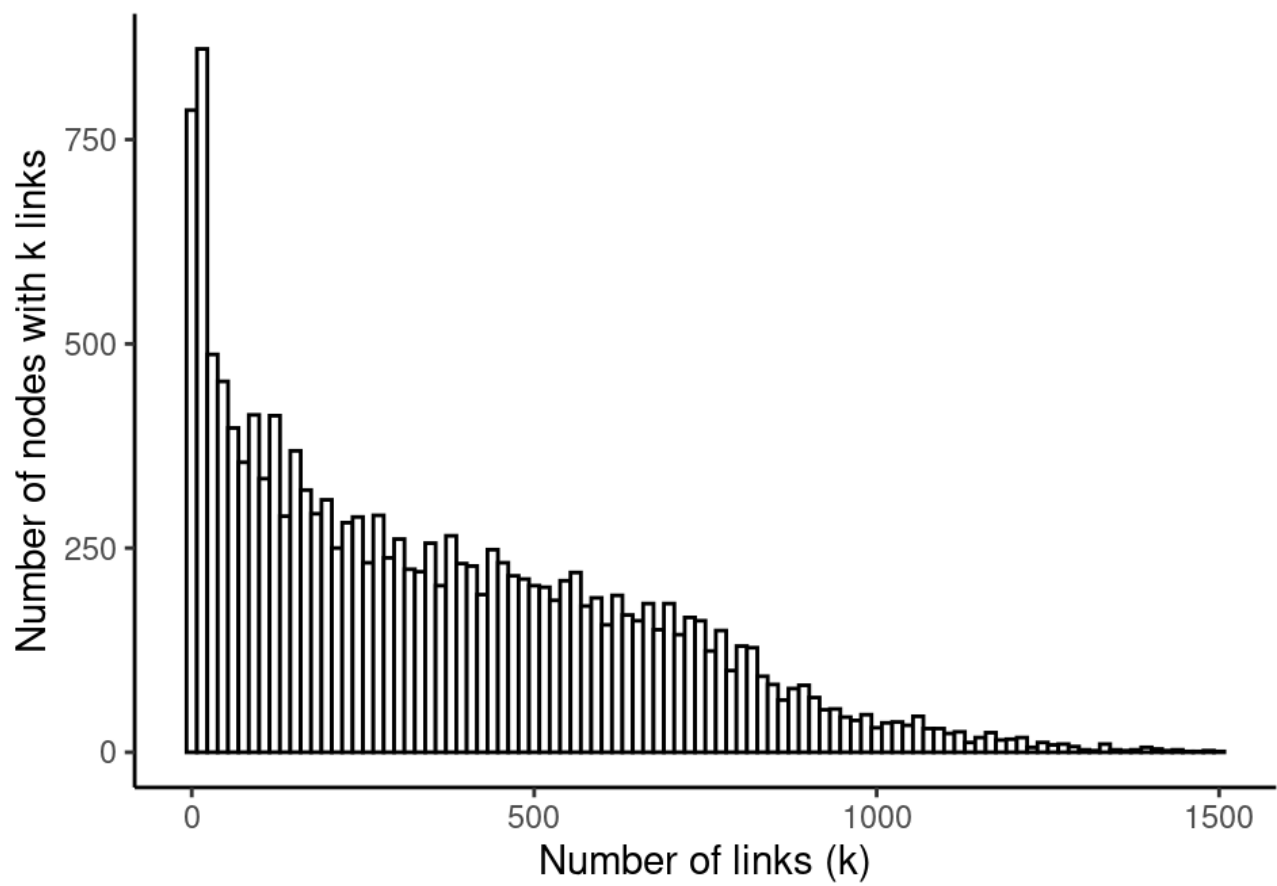
Supplementary Figure 6: Module preservation between the updated DLPFC (n=1,210 individuals) and Mostafavi et al. 2018 modules.

a) The composite statistic medianRank (y-axis) as a function of the module size. Each dot is one module labeled accordingly. b) The Zsummary statistic equation (y-axis) as a function of the module size (x-axis). Dashed lines represent the cut-off for the Kruskal–Wallis test for 2 and 10 where modules > 2 are considered preserved or > 10 highly preserved. Gold is a random module.



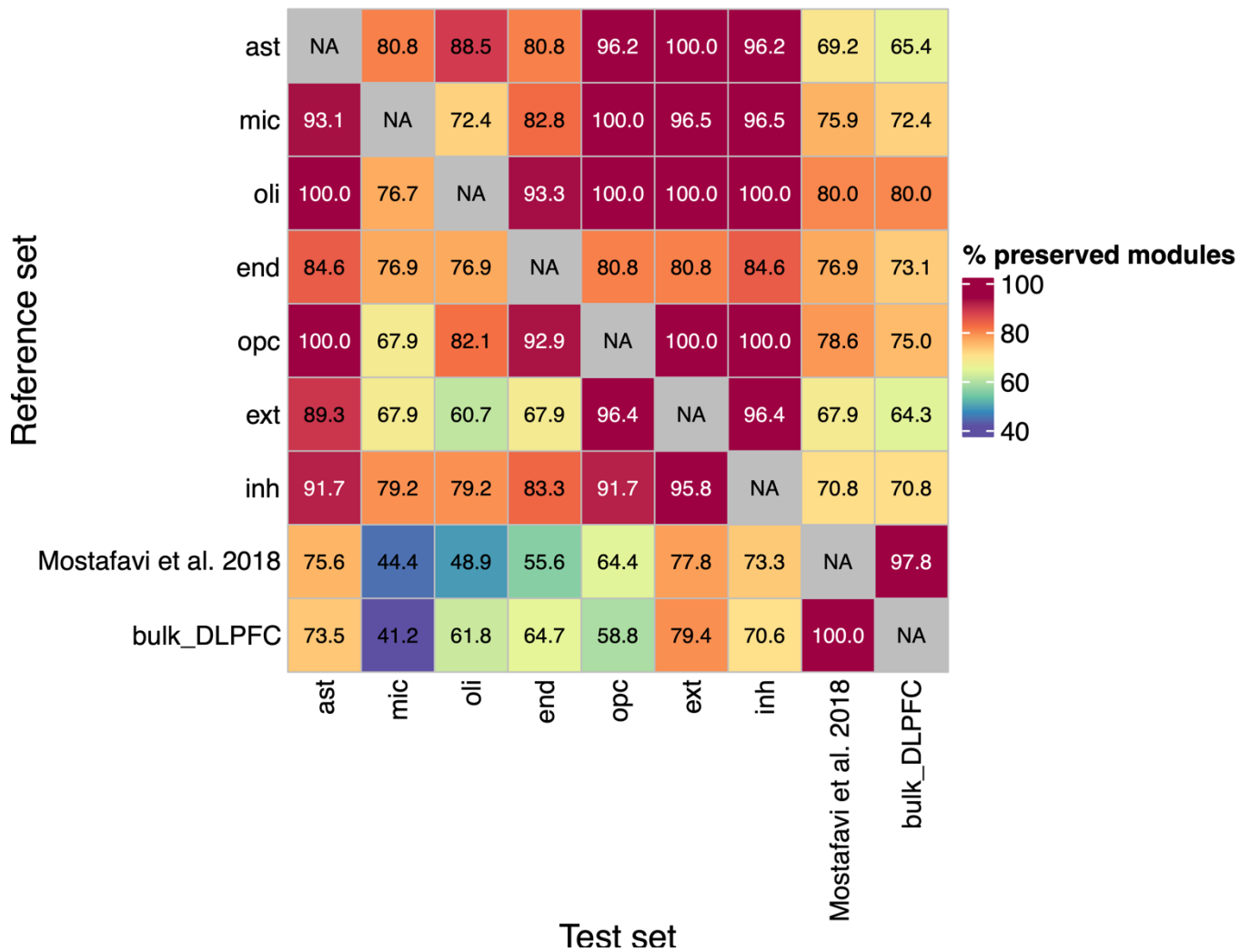
Supplementary Figure 7: Are the modules from Mostafavi et al. 2018 preserved in the updated DLPFC (n=1,210 individuals) network?

a) The composite statistic medianRank (y-axis) as a function of the module size. Each dot is one module labeled accordingly. b) The Zsummary statistic equation (y-axis) as a function of the module size (x-axis). Each dot is one module labeled accordingly. Dashed lines represent the cut-off for the Kruskal–Wallis test for 2 and 10 where modules > 2 are considered preserved or > 10 highly preserved. Gold is a random module.



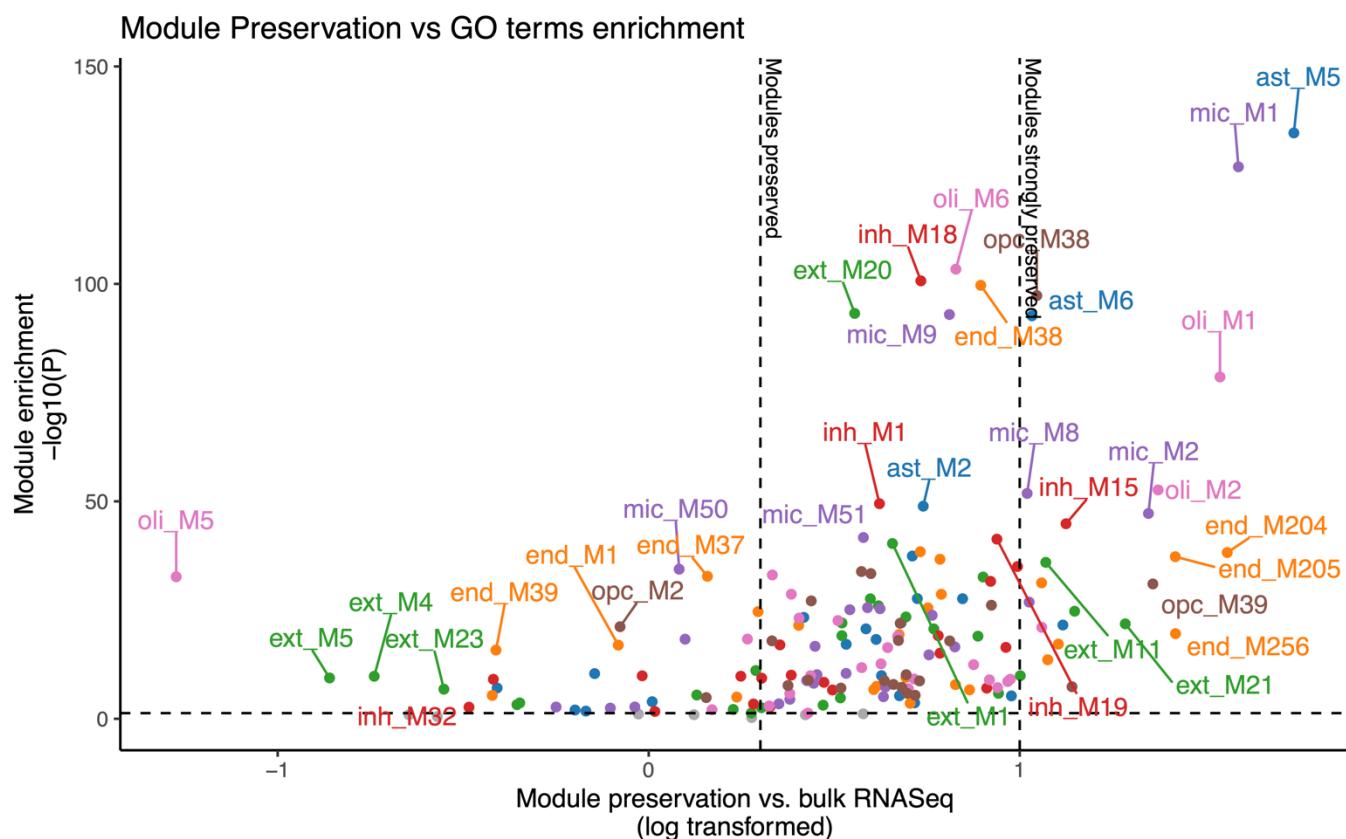
Supplementary Figure 8: DLPFC network distribution.

The number of nodes (y-axis) by the number of connections (x-axis).



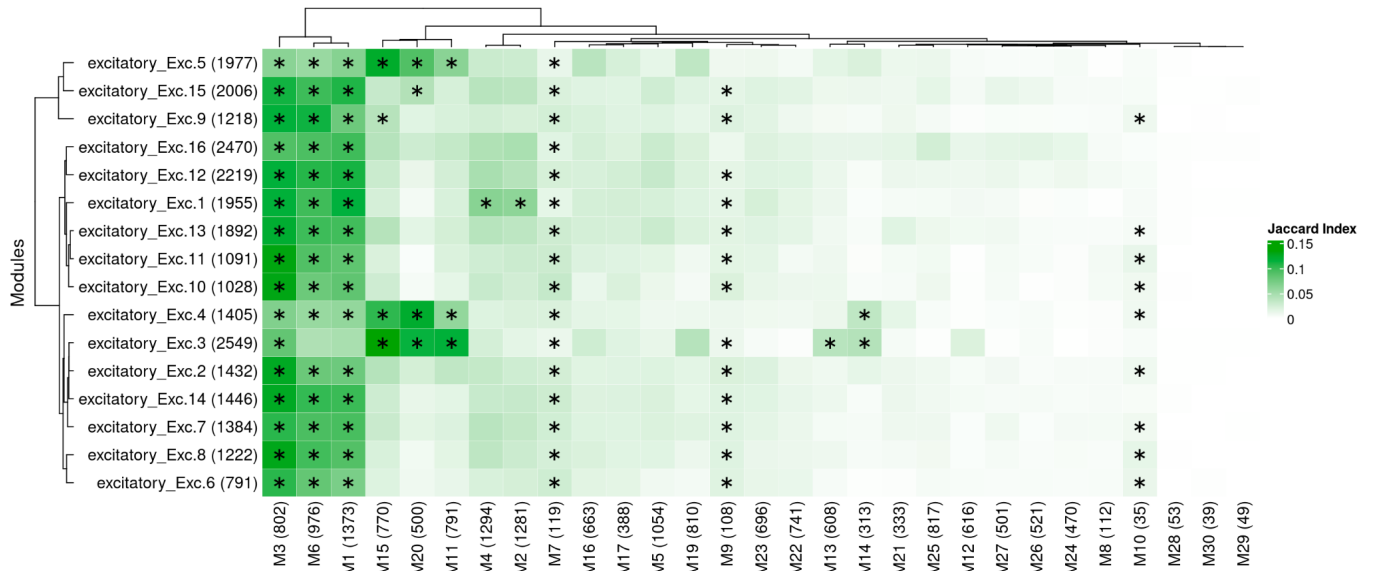
Supplementary Figure 9: Heatmap showing the percentage of modules preserved across the coexpression networks.

Heatmap showing the percentage of modules preserved across the coexpression networks according to Zsummary statistic equations based on density and connectivity. Red colors indicate high preservation while purple indicate low preservation values. Mostafavi et al. 2018 refers to the modules from the previously published Mostafavi et al. *Nat Neuroscience*, 2018. The bulk_DLPFC refers to the updated set of modules from this study which included 1,210 individuals.



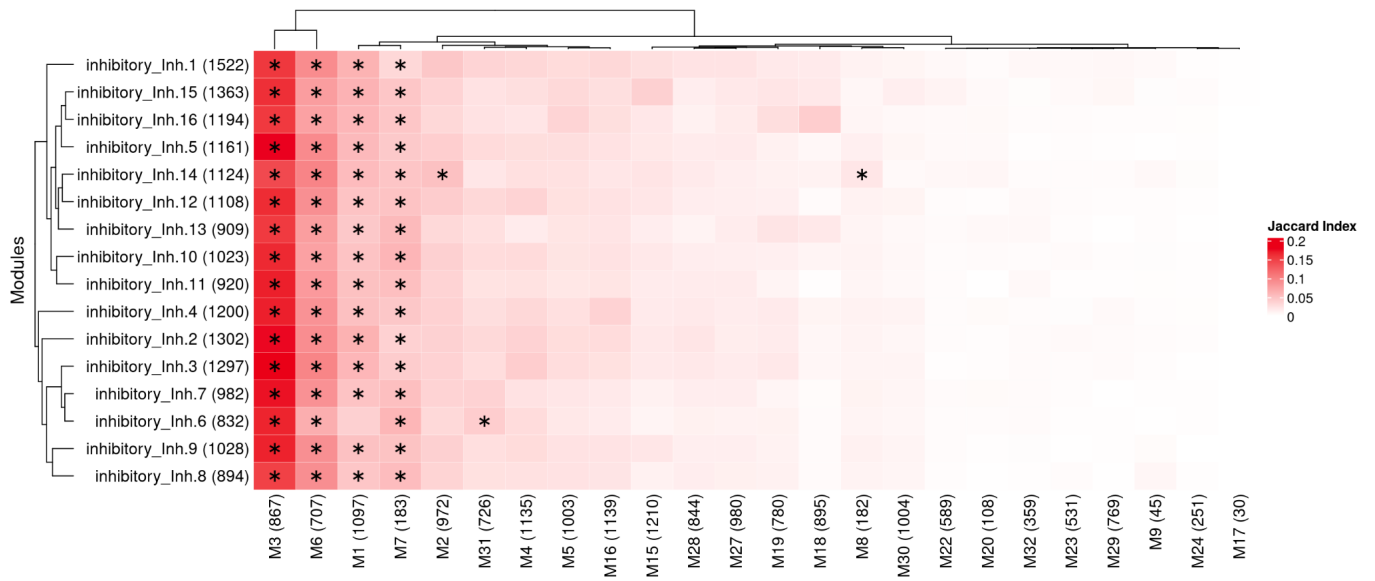
Supplementary Figure 10: Comparison between module preservation against bulk RNASeq and functional enrichment.

Module preservation is shown in the x-axis (as $\log_{10}(Z_{\text{summary}})$) while the y-axis shows the functional enrichment analysis for Gene ontology terms (best term $-\log_{10}(P\text{-value})$). Vertical dashed lines indicate module preservation thresholds: not preserved < 2 , moderately preserved if ≥ 2 and < 10 , and highly preserved if $Z_{\text{summary}} \geq 10$. Horizontal dashed lines indicate P -value threshold of 0.05 for terms from Gene ontology.



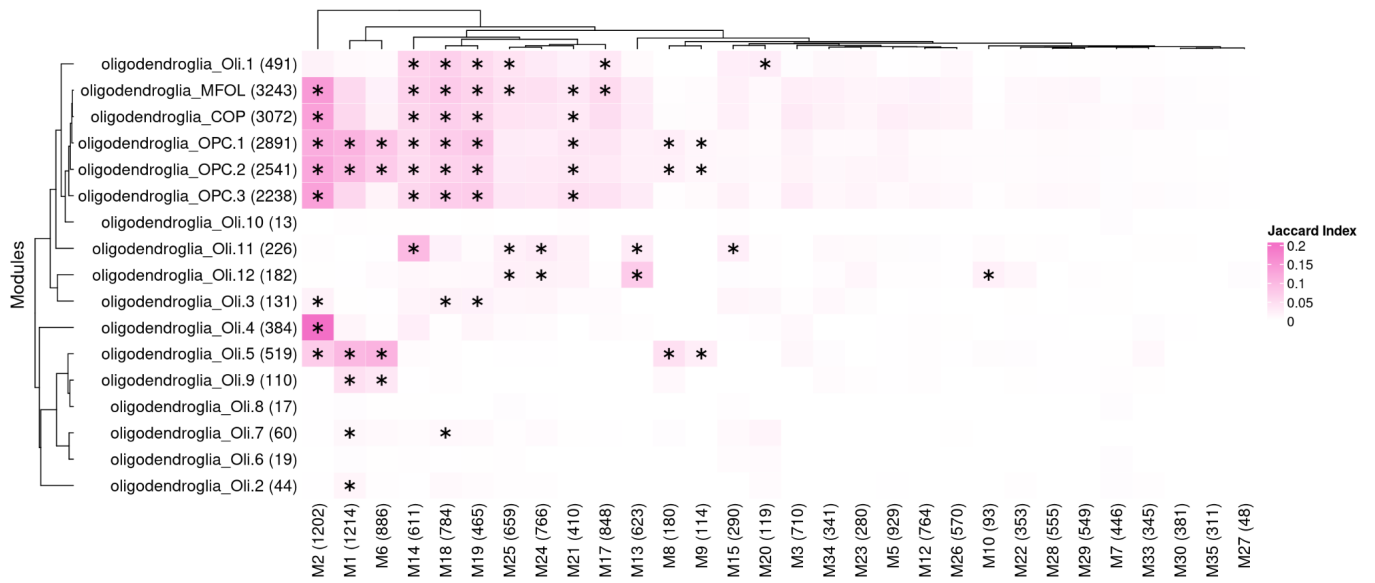
Supplementary Figure 11: Do the excitatory neurons modules recapitulate sub-cell type clusters or states?

Fisher exact test results for enrichment analysis between the excitatory neuron modules, and the sub-cell types of marker genes (Green et al. 2024). Asterisks indicate if the results reached a significance threshold, Bonferroni P -value < 0.05 . A stronger color indicates higher Jaccard index values.



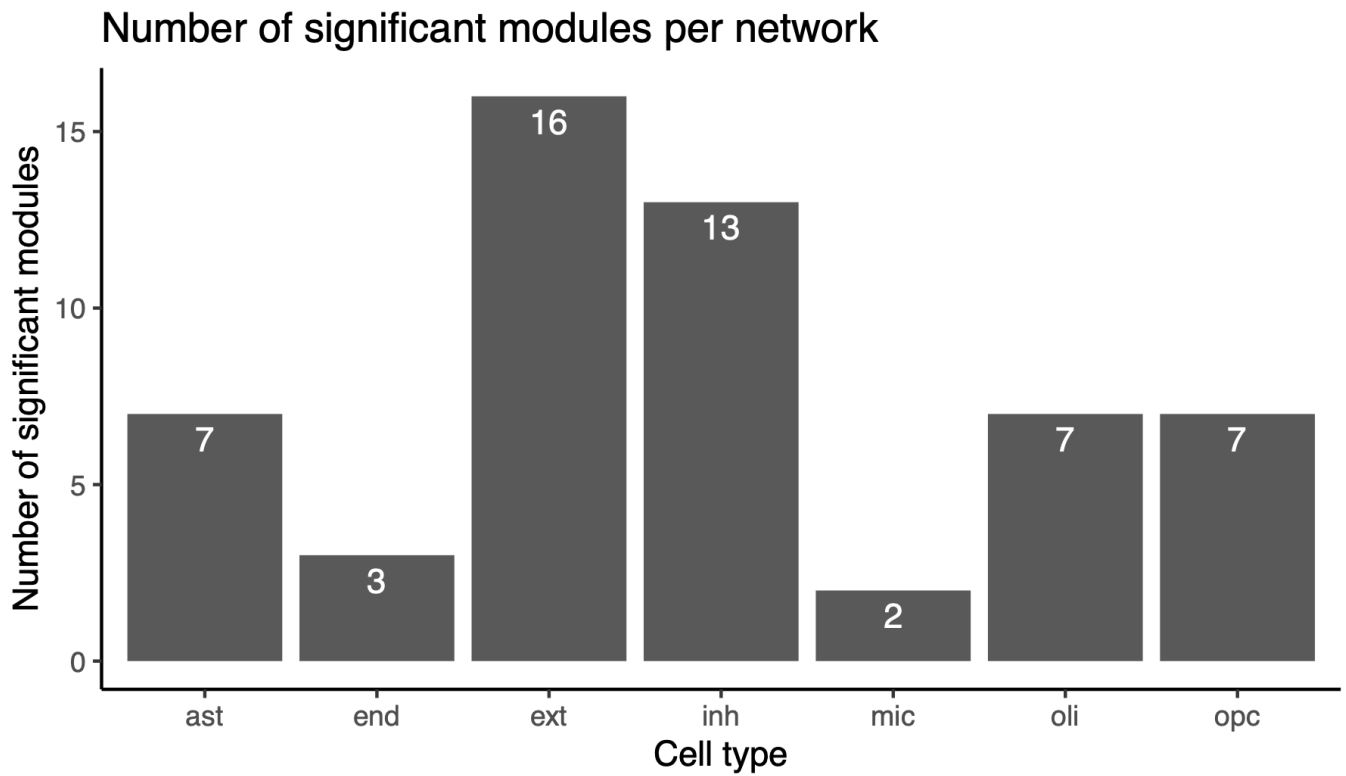
Supplementary Figure 12: Do the inhibitory neurons modules recapitulate sub-cell type clusters or states?

Fisher exact test results for enrichment analysis between the inhibitory neuron modules, and the sub-cell types of marker genes (Green et al. 2024). Asterisks indicate if the results reached a significance threshold, Bonferroni P -value < 0.05 . A stronger color indicates higher Jaccard index values.



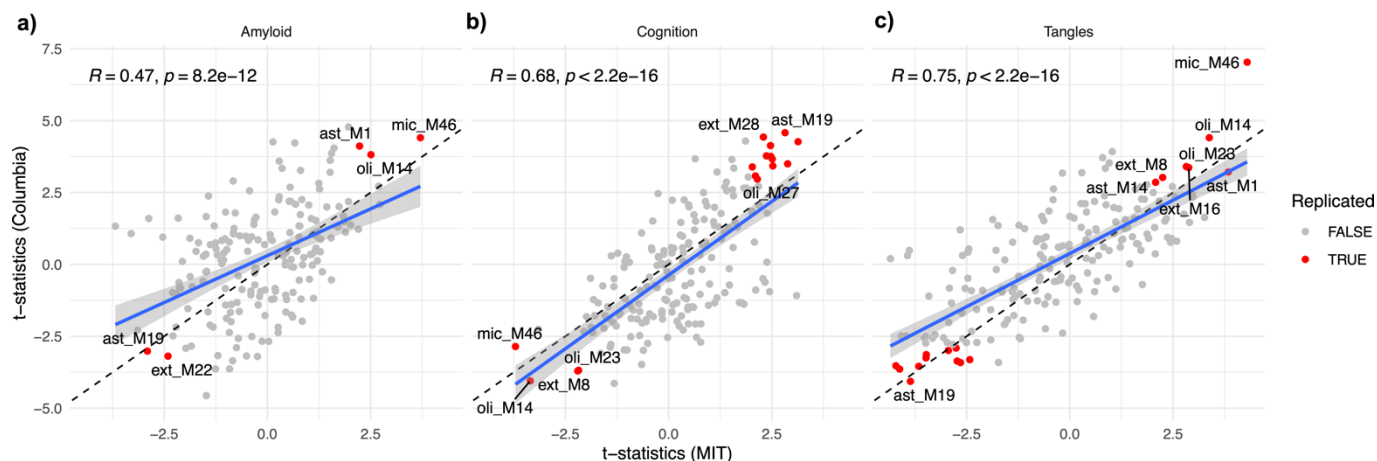
Supplementary Figure 13: Do the oligodendrocytes modules recapitulate sub-cell type clusters or states?

Fisher exact test results for enrichment analysis between the oligodendrocytes modules, and the sub-cell types of marker genes (Green et al. 2024). Asterisks indicate if the results reached a significance threshold, Bonferroni P -value < 0.05 . A stronger color indicates higher Jaccard index values.



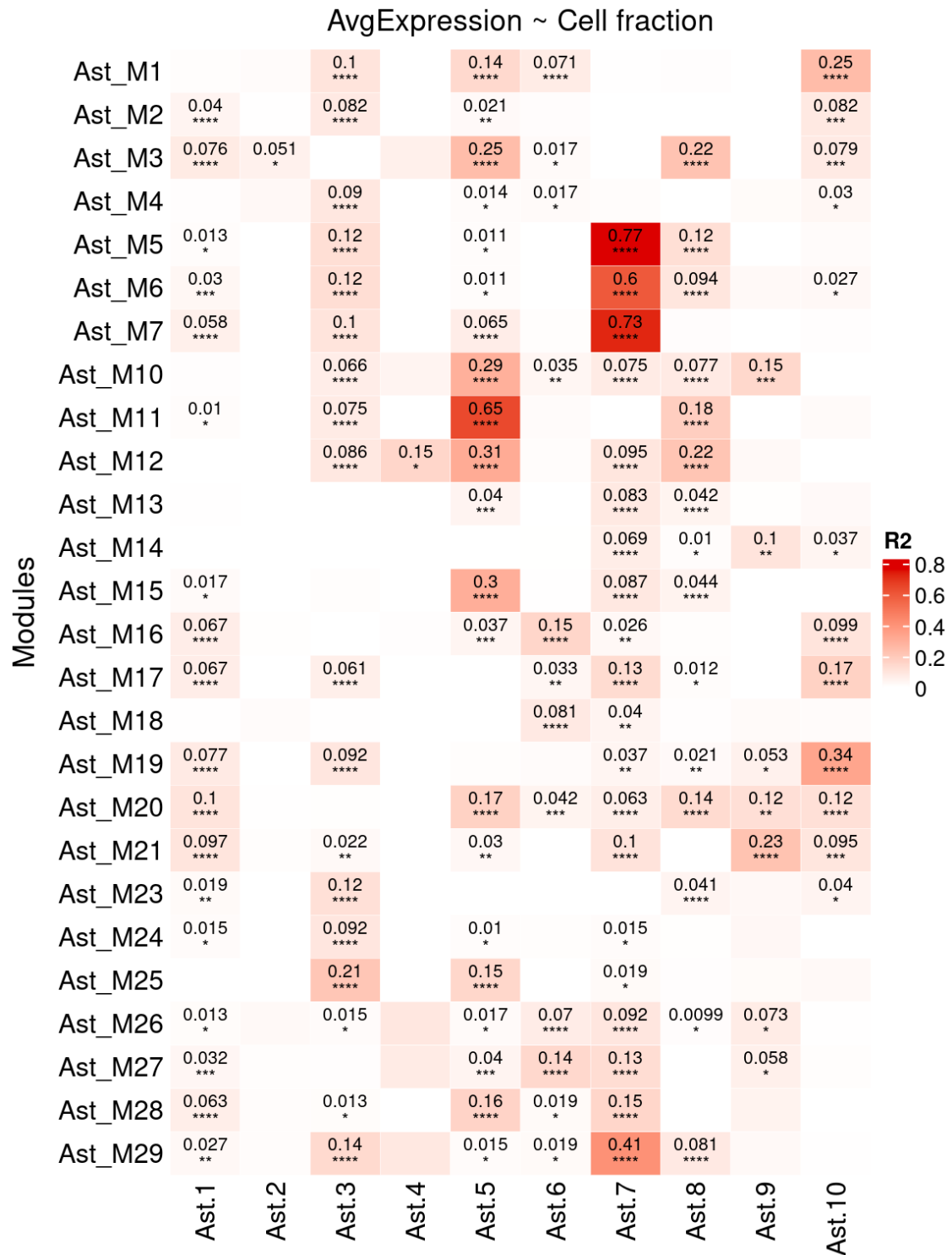
Supplementary Figure 14: Number of significant modules per network.

Barplot showing the summarized results from the regression analysis of the module's average expression with eight AD-related traits. FDR P -value < 0.05 for significance threshold. Ast: Astrocytes, end: Endothelial cells, ext: Excitatory neurons, Inh: inhibitory neurons, mic: Microglia, oli: Oligodendrocytes, and opc: Oligodendrocyte precursor cells.

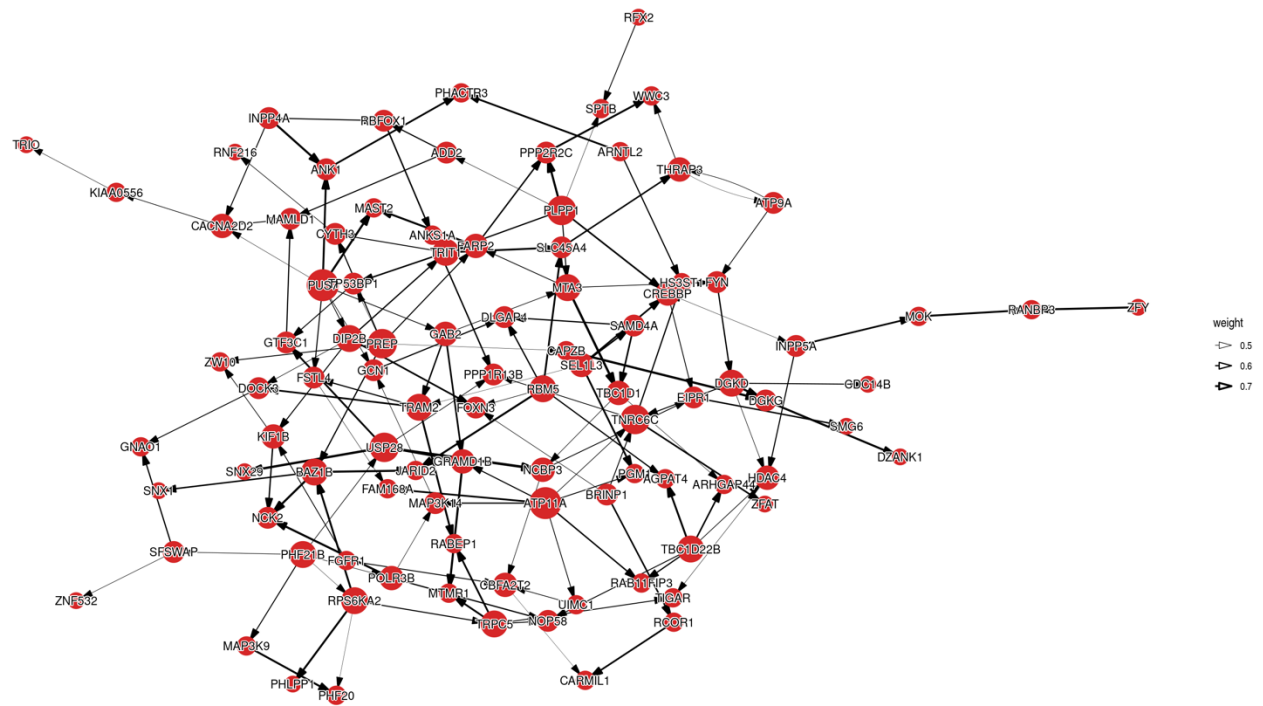


Supplementary Figure 15: Replication analysis with an external dataset.

a–c) T-statistics comparison between the modules created using the Columbia single-nucleus dataset (this study), mapped onto the non-overlapping ROSMAP participants (n=190) in the single-nucleus MIT dataset (x-axis). Each dot represents one module, and red highlights specific ones. a) T-statistics comparison for the amyloid- β phenotype, b) slope of cognitive decline, and c) tangles accumulation.



Supplementary Figure 16: Linear regression between the module's average expression and astrocyte cell fractions.



Supplementary Figure 17: The intra-module Bayesian network for the top 100 genes in the inh_M6. Top genes were selected according to their association with AD pathology and cognitive decline measurements. The arrows indicate the direction of effect, and edge thickness is related to the weights after 5,000 interactions. Node size is proportional to the node's connectivity.

References

- Ashburner, M., C. A. Ball, J. A. Blake, D. Botstein, H. Butler, J. M. Cherry, A. P. Davis, et al. 2000. "Gene Ontology: Tool for the Unification of Biology. The Gene Ontology Consortium." *Nature Genetics* 25 (1): 25–29.
- Bellenguez, Céline, Fahri Küçükali, Iris E. Jansen, Luca Klei, Luca Klei, Sonia Moreno-Grau, Najaf Amin, Adam C. Naj, et al. 2022. "New Insights into the Genetic Etiology of Alzheimer's Disease and Related Dementias." *Nature Genetics* 54 (4): 412–36.
- Chen, Wei-Ting, Ashley Lu, Kathleen Craessaerts, Benjamin Pavie, Carlo Sala Frigerio, Nikky Corthout, Xiaoyan Qian, et al. 2020. "Spatial Transcriptomics and In Situ Sequencing to Study Alzheimer's Disease." *Cell* 182 (4): 976-991.e19.
- Fujita, Masashi, Zongmei Gao, Lu Zeng, Cristin McCabe, Charles C. White, Bernard Ng, Gilad Sahar Green, et al. 2024. "Cell Subtype-Specific Effects of Genetic Variation in the Alzheimer's Disease Brain." *Nature Genetics* 56 (4): 605–14.
- Gene Ontology Consortium, Suzi A. Aleksander, James Balhoff, Seth Carbon, J. Michael Cherry, Harold J. Drabkin, Dustin Ebert, et al. 2023. "The Gene Ontology Knowledgebase in 2023." *Genetics* 224 (1). <https://doi.org/10.1093/genetics/iyad031>.
- Green, Gilad Sahar, Masashi Fujita, Hyun-Sik Yang, Mariko Taga, Anael Cain, Cristin McCabe, Natacha Comandante-Lou, et al. 2024. "Cellular Communities Reveal Trajectories of Brain Ageing and Alzheimer's Disease." *Nature* 633 (8030): 634–45.
- Johnson, Erik C. B., E. Kathleen Carter, Eric B. Dammer, Duc M. Duong, Ekaterina S. Gerasimov, Yue Liu, Jiaqi Liu, et al. 2022. "Large-Scale Deep Multi-Layer Analysis of Alzheimer's Disease Brain Reveals Strong Proteomic Disease-Related Changes Not Observed at the RNA Level." *Nature Neuroscience* 25 (2): 213–25.
- Mathys, Hansruedi, Zhuyu Peng, Carles A. Boix, Matheus B. Victor, Noelle Leary, Sudhagar Babu, Ghada Abdelhady, et al. 2023. "Single-Cell Atlas Reveals Correlates of High Cognitive Function, Dementia, and Resilience to Alzheimer's Disease Pathology." *Cell* 186 (20): 4365-4385.e27.
- Mostafavi, Sara, Chris Gaiteri, Sarah E. Sullivan, Charles C. White, Shinya Tasaki, Jishu Xu, Mariko Taga, et al. 2018. "A Molecular Network of the Aging Human Brain Provides Insights into the Pathology and Cognitive Decline of Alzheimer's Disease." *Nature Neuroscience* 21 (6): 811–19.
- Tasaki, Shinya, Ben Sauerwine, Bruce Hoff, Hiroyoshi Toyoshiba, Chris Gaiteri, and Elias Chaibub Neto. 2015. "Bayesian Network Reconstruction Using Systems Genetics Data: Comparison of MCMC Methods." *Genetics* 199 (4): 973–89.