Integrated Systems Analysis Deciphers Transcriptome and Glycoproteome Links in Alzheimer's Disease

Yusuke Matsui*[1,2], Akira Togayachi[3], Kazuma Sakamoto[1,4], Kiyohiko Angata[3], Kenji Kadomatsu[1,4], Shoko Nishihara[3]

[1]Institute for Glyco-core Research (iGCORE), Nagoya University, Furo-cho, Chikusa-ku, Nagoya 464-8601, Japan

[2]Biomedical and Health Informatics Unit, Department of Integrated Health Science, Nagoya University Graduate School of Medicine, Daiko-minami, Higashi-ku, Nagoya, 461-8673, Japan

[3]Glycan and Life Systems Integration Center (GaLSIC), Soka University, 1-236 Tangi-machi, Hachioji, Tokyo 192-8577, Japan

[4]Department of Biochemistry, Nagoya University Graduate School of Medicine, Tsurumai-cho, Showa-ku, Nagoya, 466-8550, Japan

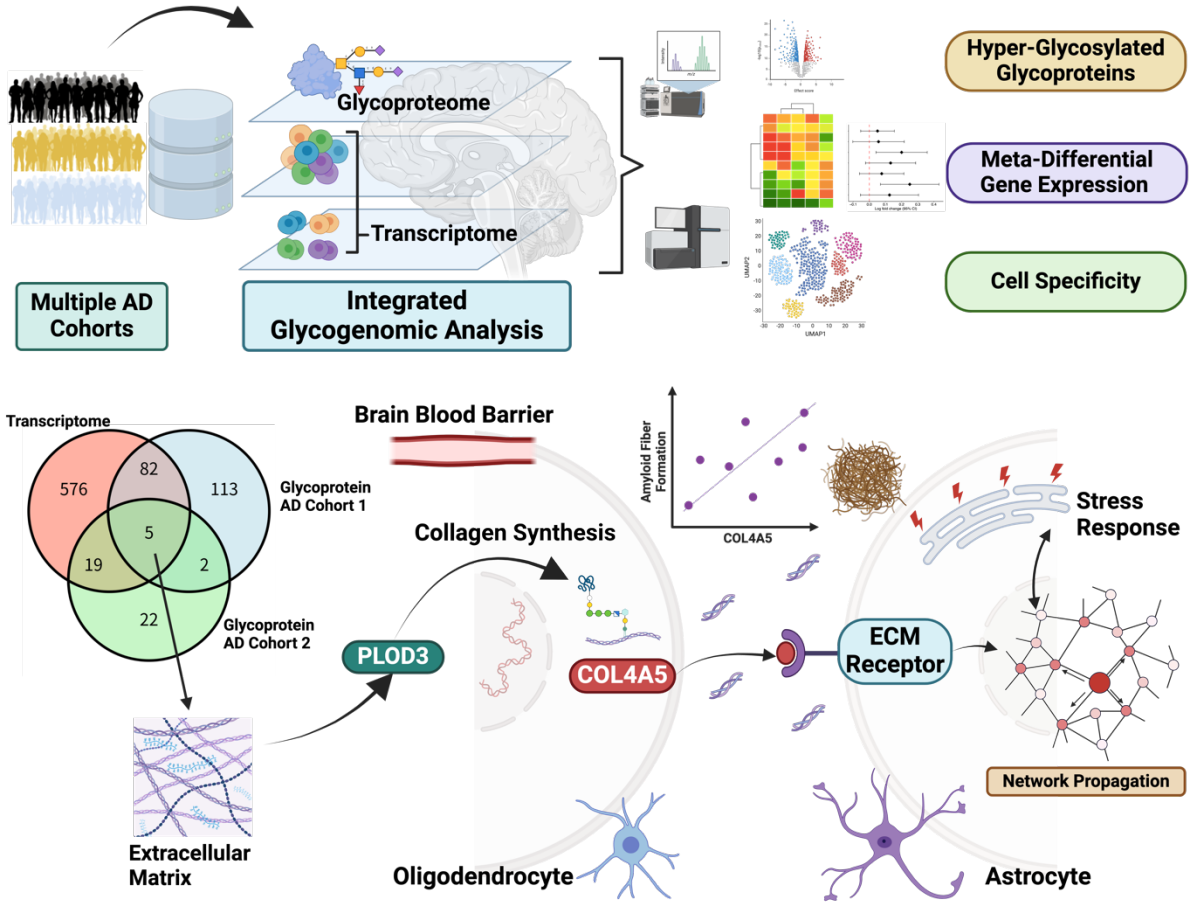
*Corresponding author

Lead Contact: matsui@met.nagoya-u.ac.jp

Running title: Glycogenomics in Alzheimer's Disease

Character count: 48,258

Graphical Abstract



Abstract

Glycosylation is increasingly recognized as a potential therapeutic target in Alzheimer's disease. In recent years, evidence of Alzheimer's disease-specific glycoproteins has been established. However, the mechanisms underlying their dysregulation, including tissue- and cell-type specificity, are not fully understood. We aimed to explore the upstream regulators of aberrant glycosylation by integrating multiple data sources using a glycogenomics approach. We identified dysregulation of the glycosyltransferase PLOD3 in oligodendrocytes as an upstream regulator of cerebral vessels and found that it is involved in COL4A5 synthesis, which is strongly correlated with amyloid fiber formation. Furthermore, COL4A5 has been suggested to interact with astrocytes via extracellular matrix receptors as a ligand. This study suggests directions for new therapeutic strategies for Alzheimer's disease targeting glycosyltransferases.

Keywords:

Glycoproteomics, Transcriptome, Glycosyltransferase, Extracellular Matrix, Alzheimer's disease

Introduction

Alzheimer's disease (AD) is an age-related neurodegenerative disease[1,2]. The primary causes are neurogenic cell loss, accumulation of misfolded proteins, oxidative stress, and inflammatory responses[3]. Genomic, transcriptomic, and epigenetic mechanisms have been intensively examined[4]. However, our knowledge of the post-translational modifications that regulate cellular functions and interactions between cells is still lacking[5]. In particular, glycosylation is the most diverse and abundant post-translational modification among protein modifications[6].

Protein glycosylation is a complex multistep process involving approximately 200 different glycosyltransferases[6–8]. There are 16 major glycosylation pathways, including lipid glycosylation, N-glycosylation, O-glycosylation, C-mannosylation, lipid glycosylation, and glycosylphosphatidylinositol (GPI)-anchored synthesis. Recently, glycomics / glycoproteomics analysis of the human AD postmortem brain[9,10], serum,[11–17] and cerebrospinal fluid[18–20] for N-glycosylation, the most abundant glycosylation pathway, has revealed dysregulated glycoproteins. In addition, the biological functions of abnormal glycans in AD pathology have been reported in some cases; for example, it is known that inhibition of BACE1 glycosylation reduces the cleavage of the amyloid ß precursor protein (APP) [21–23]. However, most biological functions of glycosylation in the pathogenesis of AD are poorly understood.

Glycan structures are not independent of the DNA template, and glycosylation depends on a combination of approximately 200 glycosyltransferases and 500 related proteins[6–8]. Thus, their dysregulation may act as an upstream regulatory factor that triggers abnormal glycosylation processes[24]. In addition, it is difficult to elucidate biological glycosylation mechanisms at the single-cell resolution using glycomics / glycoproteomics alone because current technology is limited to probing with glycosylation-specific antibodies and glycan-binding proteins, such as lectin [6]. Therefore, a glycogenomics approach that integrates genomics, or functional genomics, and glycoproteomics is critical for a comprehensive

68 understanding of biological glycosylation pathways[24,25].

69 Here, we present the factors upstream of aberrant glycosylation in AD. We performed
70 an integrated analysis of bulk and single-cell/nuclear transcriptomic and glycoproteomics data
71 from human AD brain tissues. In particular, we showed that the extracellular matrix (ECM)
72 is a common signature in the glycoproteome and transcriptome, and that ECM gene
73 expression signatures are enriched for cerebral vascular-related pathways. We identified
74 Procollagen-Lysine,2-Oxoglutarate 5-Dioxygenase 3 (PLOD3) as an upstream
75 glycosyltransferase common to ECM pathway. Through an integrated analysis of multiple
76 single-cell expression data, we showed that PLOD3 is involved in the regulation of collagen
77 type IV alpha 5 chain (COL4A5), which is strongly correlated with amyloid fiber formation.
78 Cell–cell interaction and signaling pathway analyses suggested that PLOD3-COL4A5 cascade
79 is involved in the stress response via the ECM receptor in astrocytes.

80

81 **Results**

82 **Hyperglycosylated proteins are primarily enriched in the ECM**

83 To examine the association between the molecular pathogenesis of AD and
84 glycosylation, we accessed glycoproteomic data consisting of 2 cohorts of postmortem brain
85 tissue from AD patients [9,10] (**Figure 1A**). The first dataset consisted of dorsolateral prefrontal
86 cortex tissue from 8 neuropathologically confirmed AD cases and 8 age-matched controls. The
87 second dataset comprised a subset of the ROSMAP cohort[14]. Glycoproteomic analysis was
88 performed on the postmortem brains of 10 patients with asymptomatic AD, 10 patients with
89 symptomatic AD, and 10 healthy brains in which none of the above were present.

90 In each cohort, 92 and 10 AD-specific hyperglycosylated proteins (**Supplementary**
91 **Table S1**) were identified, respectively, and pathway enrichment analysis was performed
92 (**Figure 1A**). Among the pathways significantly enriched in the 2 cohorts, we identified the
93 ECM pathway as the most common pathway among 7 pathways (**Figure 1B, C**). The

relationship between AD and ECM has recently been recognized as a new molecular pathogenesis, along with other major pathological hypotheses[26,27]. ECM components contain glycoproteins, including glycosylated proteoglycans and collagen, as major elements[27], and many glycosylations play important roles in ECM formation and maintenance.

Meta-analysis of the transcriptome reveals that glycogenes are enriched in the ECM

We explored the upstream factors that regulate ECM hyperglycosylation in AD. We accessed the AD Knowledge Portal (https://adknowledgeportal.synapse.org), which contains postmortem brain transcriptome data from multiple cohorts of patients with AD, and compiled gene expression data. The glycogene set consisting of 214 glycosyltransferases was defined using the gene list in the Glycogene Database (GGDB: https://acgg.asia/ggdb2/)[28] and literature[6,29,30] (**Figure 2A, Supplementary Table 2**). This gene set was also categorized according to its glycosylation pathway and synthesis steps (**Figure 2A**).

We derived transcriptional signatures of glycogenes based on a meta-analysis. We identified 46 differentially expressed genes (DEGs) in the glycogenes (**Figure 2B, Table S3**). We mapped glycogenes to glycosylation pathways to determine the pathways enriched for the DEGs (**Figure 2C**). Glycosyltransferases were differentially expressed in all pathways (**Figure 2C**), indicating that the signals triggering aberrant glycosylation had already been observed at the transcriptional level.

Next, we analyzed the biological functions of these glycogene signatures. The 779 globally enriched biological pathways were estimated based on the effect size from the differential expression obtained by meta-analysis using all genes (FDR < 5%) (**Figure 2D, Supplementary Table S2–S4**). Subsequently, a post-hoc enrichment analysis was performed to infer which glycosylation pathways were associated with these enriched biological pathways (**Figure 2D, Supplementary Table S5**). Significant glycosylation pathways were extracted with a hypergeometric test as the final estimation results (**Figure 2D, Table S4–S5; p-value < 5%**).

We found that the ECM is a common biological signature of the transcription and glycosylation layers in AD (**Figure 3A**). The ECM cluster was strongly associated with the hydroxyl galactose glycosylation pathway (**Figure 2D**).

PLOD3 is identified as a functional hub glycogene for ECM

Next, we performed an in-depth analysis of the glycogenes that play a central role in the ECM. Of the 779 globally enriched pathways, we constructed a bipartite graph consisting of glycogene–pathway relationships based on 48 pathways, including the differentially expressed glycogenes (**Figure 3B**). We inferred the glycogene importance based on the number of neighboring pathways, that is, the network degree (**Figure 3C**). As a result, PLOD3 was identified as a hub glycogene with the highest degree (**Figure 3C**).

PLOD3 is an enzyme that mediates essential glycosylation during the early stages of collagen formation[31]. In general, collagen is broadly modified by the hydroxylation of proline and lysine and glycosylation of specific hydroxylysine residues[32]. Hydroxylation of lysine is catalyzed by PLOD3[33,34]; hydroxylysine undergoes further glycosylation, and COLGALT1 transfers galactose, which are critical steps for maintaining collagen integrity[32].

To further confirm the results at the gene expression level, we examined whether the changes in PLOD3 expression were consistent among the AD cohorts included in the meta-analysis. We found that PLOD3 was consistently upregulated in individual cohort studies (**Figure 3D**), and the expression signatures of ECM organization and collagen formation showed a consistent overexpression trend (**Figure 3D**). Based on this analysis, we hypothesized that hyperglycosylation of the ECM in AD brain tissue is mediated by PLOD3.

PLOD3 is expressed in oligodendrocytes and co-expressed with COL4A5

We sought to determine the cellular origin of the PLOD3 and collagen genes. First, we accessed the scRNA-seq data of normal brain tissue from the Human Protein Atlas

(v22)[35,36]. We found that PLOD3 was co-expressed with COL4A5 in oligodendrocytes (**Figure 4A**). These two genes showed distinct oligodendrocyte-specific expression signatures (**Figure 4B**). We also accessed a human AD cohort of single-nucleus RNA-seq data for the entorhinal cortex (GSE138852)[37]. The entorhinal cortex is one of the brain regions that shows neurodegeneration in the early stages of AD[38–40]. The cohort included both non-cognitive impairment (NCI) and AD brain. Six cell types were identified: microglia, astrocytes, neurons, oligodendrocyte progenitor cells, oligodendrocytes, and endothelial cells (**Figure 4C**). PLOD3 and COL4A5 were highly expressed in oligodendrocytes (**Figure 4C**). These genes were also predominantly expressed in the AD group (**Figure 4C**).

COL4A5 consistently correlated with amyloid fiber formation in multiple cohort studies

COL4A5 is partially correlated with amyloid plaque accumulation[41]. However, this finding has not been validated in large clinical samples. We tested whether COL4A5 expression was significantly correlated with APP expression. We analyzed the bulk RNA-seq data used in the meta-analysis and examined their relationship with APP gene expression separately for each brain region. The results showed that COL4A5 was strongly correlated with the APP gene in all datasets (**Figure 4D**). Furthermore, we defined the gene signatures of the amyloid plaque formation pathway and analyzed the correlation between their eigengene expression and COL4A5 in the same way. As expected, a strong correlation was confirmed (**Figure 4D**). PLOD3 was evaluated similarly, showing a weaker correlation than COL4A5, but it was significant in several datasets (**Supplementary Figure 1**).

Cerebrovasculature most strongly associated with ECM dysregulation

We explored whether overexpression of the PLOD3–COL4A5 axis is involved in biological processes in the AD brain. First, we analyzed the biological pathways that best explained ECM activity. We used AES-PCA[42–44], a principal component analysis (PCA)-based

regression model with ECM activity as the outcome variable and all other biological pathway activities as predictors, for each AD cohort used in the meta-analysis (**Figure 5A, Supplementary Table S6**). The estimated p-values were statistically combined using Fisher's method (**Figure 5A**). Four of the top 10 enriched genes were associated with the vascular system (**Figure 5A**) and were overexpressed in the AD group at the expression level (**Figure 5B**). We hypothesized that the PLOD3–COL4A5 axis is involved in the cerebrovascular microenvironment.

PLOD3 and COL4A5 are expressed in oligodendrocytes of the cerebrovasculature microenvironment

We analyzed recently reported scRNA-seq data from the vascular microenvironment of the human brain (GSE16357)[45]. These data were used to quantify gene expression by VINE-seq in the cerebral blood vessels in 8 NCI and 9 AD samples (**Figure 5C**). Gene expression was quantified in 143,793 cells from 14 cell types, including vascular endothelial cells (arterial, capillary, and venous), mural smooth muscle cells (SMCs), pericytes, astrocytes, macrophages, T cells, and perivascular and medullary fibroblasts (**Figure 5C**). We examined cell types expressing PLOD3 and COL4A5, which were most strongly expressed in oligodendrocytes (**Figure 5D, E**). In contrast, other type IV collagens were mainly expressed in pericytes and SMCs, which is consistent with the fact that type IV collagen constitutes the vascular basement membrane[46].

Oligodendrocytes interact with astrocytes via the COL4A5 ligand

Next, we analyzed the biological functions and pathways mediated by the PLOD3–COL4A5 axis in the cerebrovascular microenvironment. According to the KEGG pathway analysis, COL4A5 may contribute to cell-to-cell communication via ECM ligand receptors (hsa04512). We analyzed how the PLOD3–COL4A5 axis of oligodendrocytes mediates

198 intercommunication between cell types. CellChat[47] allows for the estimation of cell–cell
199 interactions for each signaling pathway. We estimated cell–cell interactions based on collagen
200 signaling pathways in the AD group. Oligodendrocytes interacted with astrocytes via the
201 COL4A5 ligand and CD44 receptor (**Figure 6A**). This was verified using NicheNet[48], another
202 intercellular communication estimation algorithm. Among oligodendrocytes, COL4A5 was
203 identified as one of the most promising candidates (**Figure 6B**). In addition to CD44 identified
204 by CellChat, SDC4, DDR2, ITGB8, and ITGAV were predicted to be astrocyte receptors
205 (**Figure 6B**). These receptors were highly expressed in astrocytes (**Figure 6C**).
206
207 **COL4A5 ligand is involved in the regulatory cascade of the astrocyte stress response**
208 We performed a detailed analysis of signaling pathways to understand the biological
209 functions of COL4A5-mediated interactions between oligodendrocytes and astrocytes. We
210 integrated the predicted COL4A5 ligand–receptor pairs (CD44, SDC4, DDR2, ITGB8, and
211 ITGAV) into the prior knowledge of the signaling network constructed from multiple
212 perturbation experiments and databases using NicheNet. The results indicated that the
213 COL4A5 ligand targeted and activated B-cell/CLL lymphoma 6 (BCL6) and serum and
214 glucocorticoid-regulated kinase 1 (SGK1) via ECM receptors in astrocytes (**Figure 6D**). BCL6
215 is a transcription factor and master regulator of humoral immunity and B-cell
216 lymphomagenesis, while SGK 1 encodes a serine/threonine protein kinase that plays an
217 important role in cellular stress responses[49–51]. Both genes were found to be expressed in
218 astrocytes (**Figure 6E**).
219 Based on these results, we inferred the biological functions of the BCL6 and SGK1
220 gene modules in astrocytes. An astrocyte-specific co-expression network was constructed
221 based on gene expression using the hdWGCNA algorithm[52] (**Figure 6F**). Next, we applied the
222 random walk with restart (RWR) algorithm[53], which is a network propagation algorithm
223 starting from BCL6 and SGK1 on the astrocyte-specific network topology (**Figure 6F**). The

RWR allows for the evaluation of the proximity of the network between BCL6, SGK1, and other neighboring genes. Based on these results, we prioritized the top 30 neighbors (**Figure 6G**). GO analysis of these neighboring gene groups revealed that they were enriched mainly for processes involved in stress response (**Figure 6H**). These enriched pathways were also observed in the GO analysis of BCL6 and SGK1 and were independently identified using the network propagation method (**Supplementary Figure 3**).

Discussion

Our knowledge of the involvement of glycosylation, a major post-translational modification, in the pathogenesis of AD is lacking. We systematically explored the pathogenesis and driving factors based on an integrated analysis of the emerging dimensions of glycosylation in combination with transcriptomics.

In the brain tissue of patients with AD, hyperglycosylation in the ECM is the main signature shared by the glycome and transcriptome, and the glycosyltransferase PLOD3 is an upstream regulator that acts as a functional hub. PLOD3 is predominantly expressed in oligodendrocytes in AD brain tissue and the cerebrovasculature and is co-expressed with COL4A5. Importantly, COL4A5 significantly correlated with APP levels and the activity of the amyloid fiber formation pathway. Single-cell/nuclear analysis revealed that COL4A5 is a ligand for oligodendrocytes that can mediate cell–cell interactions via ECM receptors on astrocytes. In addition, signaling pathway network analysis identified BCL6 and SGK1 as its target genes, and their neighboring genes in the astrocyte-specific network analysis revealed that these two genes are involved in the regulation of the stress response.

The involvement of the ECM in AD has been supported by a large amount of literature[27,54–58]. The physiological roles of the ECM are diverse and include developmental regulation, tissue homeostasis, cell migration, cell proliferation, cell differentiation, neuronal plasticity, and neurite growth[59]. In particular, the ECM is extensively involved in the

dysregulation of perineuronal networks in AD[58,60–68], which are involved in the maintenance of spatial structure, neuronal plasticity, scaffolding,[69] and the regulation of aggregation; it is also involved in amyloid protein dynamics[27,70–75] and brain–blood barrier integrity [41,54,76–79]. As glycoproteins are the major components of the ECM[55,59,80], glycan synthesis is important for ECM homeostasis in the brain. The enrichment of dysregulated glycoproteins in the ECM is natural in this sense (**Figure 1A, B**).

We discovered that PLOD3 was enriched in the ECM and upregulated in the AD meta-analysis (**Figure 2B–D**). PLOD3 is a multifunctional enzyme, and in addition to its role as a lysyl hydroxylase, it has collagen galactosyltransferase and glucosyltransferase activity[34,81–83]. Although no direct evidence of PLOD3 in AD has been reported, it is known to play an essential role in the formation of collagen, a major component of ECM[84]. For instance, defects in PLOD3 (or lysyl hydroxylase 3; LH3) have been implicated in inherited connective tissue disorders and have been shown to cause cerebral small vessel injury[85,86], maintenance of the structural integrity of cerebral blood vessels, and the regulation of inflammatory processes[87]. This enzyme is also a promising biomarker of AD, as its expression has been reported to fluctuate in cell-free RNA expression in blood samples from patients with AD[88].

PLOD3 mediates glycosylation during early collagen formation[31]. Type IV collagen is an essential protein in the cerebral vasculature of patients with AD and is responsible for network formation in the basement membrane. Indeed, in our analysis of single-cell expression levels in cerebral vessels, many type IV collagens (COL4A1/COL4A2/COL4A3/COL4A4) were predominantly expressed in pericytes and SMCs **(Figure 5E)**. In contrast, COL4A5 behaves differently from other type IV collagens and is predominantly expressed in oligodendrocytes. Oligodendrocytes have been shown to stably bind to cerebral blood vessels by zonation analysis based on single-cell/nuclear sequencing analysis[89,90] and electron microscopy[91]. Interestingly, data from multiple studies support that COL4A5 is strongly correlated with APP and amyloid fiber formation (**Figure 4D**), suggesting

a relationship with amyloid plaque accumulation. This may be relevant because the overexpression of type IV collagen generally leads to an increase in cortical basement membrane thickness and has been implicated in the degeneration of cerebral vascular structures[55]. The functional role of type IV collagen in AD cerebrovasculature should be examined in detail in future studies.

We also performed an in silico analysis of cell–cell interactions. COL4A5 functioned as a ligand in oligodendrocyte–astrocyte interactions (**Figure 6A**). Analysis of the signaling pathway network suggested that this cell–cell interaction may contribute primarily to the stress response via SGK1 or BCL6 (**Figure 6D–H**). SGK1 is known to be transcriptionally upregulated under cellular stress[49–51]. On the other hand, both factors have also been reported to be involved in inflammatory responses in the central nervous system. Recent studies have shown that inhibition of SGK1 can suppress the NF-κB-mediated inflammatory pathway in glial cells[92]. There is also evidence that BCL6 plays a central role in regulating astrocytes and NF-κB in response to inflammatory stimuli and disorders[93]. Indeed, in our glycoprotein analysis, the immune response pathway was enriched next to the ECM (**Figure 1B, C**), and inflammatory cytokines were also significantly associated with the ECM organization pathway at the transcriptome level (**Figure 5A, Supplementary Figure S2A**). Inflammatory pathways are key signatures in the AD brain; however, their mechanisms of action in the stress response remain unclear. Further examination of the mechanisms underlying BCL6- and SGK1-mediated stress responses is required.

This study has several limitations. First, the AD glycomic analysis was limited to N-type glycans. Therefore, evidence of ECM hyperglycosylation should be verified in future studies using comprehensive glycoproteomic data. Second, the AD cohort data used in the meta-analysis were limited to those deposited on the AD knowledge portal. To establish a higher level of evidence, data from other large cohort studies should be included. Third, single-cell sequencing data were collected from several different sources; therefore, there is no

guarantee that the results reflect the differential expression results of the bulk sequencing used in the meta-analysis. It is expected that this limitation can be overcome in the future as multilayered omics data are collected. However, validation, including experimental approaches, is required.

Our results suggest that glycosylation is involved in the pathogenesis of AD through several unknown mechanisms. Our results also indicate that glycogenomics analysis integrating genetic approaches is a promising method for highlighting the biological functions of glycans and the molecular pathogenesis of diseases at a single-cell resolution. Data on AD glycoproteomics in human subjects are limited. However, as glycoproteomic analysis technology matures, it will be applied to various disease areas, and a vast amount of glycoproteomic data will be accumulated in the next decade. In the near future, the glycogenomics approach will play an important role as a bridge between the established AD genetic pathology and the emerging dimensional omics field of glycoproteomics.

Methods

Glycoproteomics enrichment analysis

The first set of glycoproteomics data[9] was used for enrichment analysis. This dataset was analyzed for glycoproteins overexpressed in the AD group (BRAAK ≥ 5) and the normal group (BRAAK ≤ 2), as defined in the original paper, using the canonical pathway collection of MSigDB (c2.cp.v2022.1.Hs.symbols.gmt). All genes were analyzed as backgrounds using the fedup package in R (https://github.com/rosscm/fedup), and the top 30 significantly enriched pathways were identified. The second set of glycoproteomics data[9] was analyzed in the same manner. Comparisons were made between the symptomatic group (BRAAK ≥ 5 and CERAD 1 or 2), the asymptomatic group (BRAAK ≥ 3 and CERAD 1 or 2), and the normal group (BRAAK ≤ 2 and CERAD 4), as defined in the original paper. Glycoproteins specifically identified in the symptomatic group were extracted. Enrichment analysis was performed to identify the top 30

328 significantly enriched pathways.

329

330 **Meta-analysis**

331 Meta-analysis using RNA-seq harmonization of AMP-AD followed the published AD-
332 CONTROL analysis protocol (https://github.com/th1vairam/ampad-DiffExp/tree/df3efa793f,
333 379730bae6d4c9e62910fb2c37e525/gene_level_analysis). First, meta-information was used
334 for data from 3 cohorts (ROSMAP, MSSM, and Mayo), including seven different brain regions,
335 to define patients with definitive late-onset AD from a clinical and neuropathological
336 perspective, that is, neurofibrillary changes, neuritic amyloid plaques, and cognitive
337 dysfunction. The AD control group consisted of patients with AD.

338 AD controls were defined as patients with few plaques and neurofibrillary changes
339 and no cognitive impairment; in ROSMAP, LOAD cases were those with a BRAAK of 4 or more,
340 a CERAD score of 2 or less, and a cognitive diagnosis of probable AD with no other causes
341 (cogdx = 4); LOAD controls were those with a BRAAK of 3 or less, a CERAD score of 3 or more,
342 and a cognitive diagnosis of "no cognitive impairment" (cogdx = 1). For the MSBB, LOAD cases
343 were defined as those with a CDR score of at least 1, a BRAAK score of at least 4, and a
344 CERAD score of at least 2. LOAD cases were similarly defined as those with a CDR score of
345 0.5 or less, a BRAAK of 3 or less, and a CERAD of 1 or less as LOAD controls. In Mayo, cases
346 were defined based on neuropathology, with LOAD cases defined as having a BRAAK score of
347 4 or higher and LOAD controls as having a BRAAK score of 3 or lower.

348 A meta-analysis using a mixed-effects model was performed to determine the
349 differences in the expression levels of each gene in each of the seven brain regions in each
350 cohort. Effect sizes were estimated using the restricted maximum likelihood method based on
351 the standard mean difference using Hedge. The Metacont function from the meta package of
352 R was used for the analysis. The p-values were corrected for multiple testing by "fdr" using
353 the p.adjust function from the stats package.

354

355 **Enrichment map**

356 Gene Set Enrichment Analysis (GSEA) was performed on all genes included in the
357 meta-analysis. The gene set was c2.cp.v2022.1.Hs.symbols from the MsigDB collection, which
358 was loaded using Enrichment Map in Cytoscape and drawn using default settings. After
359 drawing the pathways, we manually classified them into several categories and created
360 several clusters in the network. The list of glycan-related genes manually defined for each
361 glycosylation pathway was then analyzed by post-hoc analysis using the hypergeometric test
362 and the Wilcoxon test, and pathways with FDR ≤ 5% and that were significant by two tests
363 were extracted. Significant pathways in the two tests were extracted.

364

365 **Functional hub glycogene identification**

366 Among the enriched pathways based on the same GSEA results as the enrichment
367 map, only pathways containing glycogenes were extracted, and from these, a two-part graph
368 of the pathway–glycogene was extracted. Based on the two-part graphs obtained, each gene
369 was ranked according to its degree of expression. The glycogene with the highest degree was
370 defined as the functional hub glycogene. The results of querying the extracted PLOD3 to the
371 String database (v11) are shown in **Figure 2D**. Forest plots of PLOD3 are shown with
372 estimated effect sizes and 95% confidence intervals from the meta-analysis. For pathway
373 activity, GSEA was performed using the R fgsea package, with gene ranks for effect sizes for
374 each cohort and c2.cp.v2022.1.Hs.symbols for the gene set. Normalized enrichment scores
375 were used for the forest plots.

376

377 **Cell-type specificity of PLOD3**

378 For the cell-type specificity of healthy tissues, information was obtained from the
379 Human Proteome Atlas (V22) website by entering the gene name. For data on the entorhinal

cortex, information was obtained by entering gene names from http://adsn.ddnetbio.com/.

Pathway-based PCA regression and GSEA

Pathway-based PCA is a PCA-based method for analyzing pathways and phenotypic associations[43,44,94]. The R Bioconductor PathwayPCA package[42] was used for the analysis. Using region-specific gene expression data from each AD cohort (Mayo, MSSM, and ROSMAP), we specified the mean expression levels of the ECM pathway component genes as the ECM pathway activity for the objective variable and each pathway other than the ECM pathway for the explanatory variables. The gene set used was c2.cp.v2022.1.Hs.symbols from MSigDB. The pathway names containing "ECM," "Extracellular," or "Collagen" were defined as ECM pathways. The genes involved in these pathways were defined as signatures.

The p-values of the list of pathways significantly associated with ECM were combined using Fisher's method to calculate the integrated p-value. For the calculation, the log-sum function of the R metapackage[95] was used, and the p-values of the individual datasets were entered for each pathway. In addition, we cross-checked whether significantly related pathways were sufficiently enriched at the expression level. Focusing on the top 10 pathways, we applied GSEA based on the gene set c2.cp.v2022.1.Hs.symbols from MSigDB using the effect sizes of the 3 cohort meta-analysis as the gene rank. To further validate that the top 10 pathway activities tended to increase by cohort and region, the means of the effect sizes and confidence intervals were calculated for the signature genes and illustrated as forest plots.

Analysis of brain vasculature with scRNA-seq

Count data were preprocessed using the Seurat package in R. Normalization, feature selection with VST, scaling, and dimensional reduction using PCA and UMAP were performed. The cell types were visualized using those previously identified in an original paper[45]. Next, for each cell type, variation analysis between the AD and cognitively normal groups was

406 performed using Seurat's FindMarkes function, and enrichment analysis for the identified
407 groups of DEGs was performed using the R fedup package. The c2.cp.v2022.1.Hs.symbols from
408 MSigDB was used as the gene set to determine which cell types were enriched in ECM-related
409 pathways. We selected gene sets with pathway names containing "ECM," "Extracellular,"
410 "Matrisome," or "Collagen" in the pathway name. The enriched p-values were further
411 transformed as -Log10(FDR) from the multiple-test-corrected FDR, considered as
412 differentially expressed activity signals, and visualized using a heatmap. The expression
413 levels per cell type were obtained by querying https://twc-stanford.shinyapps.io/human_bbb/
414 for PLOD3.
415
416 **Cell–cell interaction and signaling network analysis**
417 Cell–cell interactions were analyzed using the R package CellChat[47]
418 (https://github.com/sqjin/CellChat). Oligodendrocytes and astrocytes identified with CellChat
419 were further analyzed using another algorithm, NicheNet[48] (https://
420 github.com/saeyslab/nichenetr). For a detailed analysis, the ligand–receptor prior information
421 was input by integrating the ligand–receptor pair information used in CellChat with the
422 ligand–receptor pair information used in NicheNet. This was also used for the signal network
423 analyses. A pre-built model was downloaded
424 (https://github.com/saeyslab/nichenetr/blob/master/vignettes/model_construction.md), and
425 both the ligand–receptor information and expression information were identified in the cell–
426 cell interactions.
427
428 **Astrocyte cell-type specific network propagation**
429 For astrocyte-specific network construction using cerebrovascular scRNA-seq, the
430 Toplogical Overlap Measure (TOM) was estimated using hdWGCNA[52], and edges were further
431 defined only if they had a TOM above the 90th percentile as a threshold. The network

propagation method was applied using the R package RandomWalkRestartMH[53]. In other words, we performed an RWR starting from SGK1 and BCL6 in the obtained network topology. The 30 most relevant neighbors were narrowed down and plotted using the R package igraph. The R package fedup was used for the enrichment analysis. To estimate the transcriptional activity of BCL6, curated regulon information was first obtained using the R package DoRothEA[96], and transcription factor target genes were estimated using the Viper[97] algorithm. The R package decoupleR[98] was used for the analysis.

Acknowledgements

This work was supported by the Human Glycome Atlas Project (HGA) and JSPS KAKENHI, Grant Number: JP20H04282. The results published herein are partly based on data obtained from the AD Knowledge Portal (https://adknowledgeportal.org). Data generation was supported by the following NIH grants: P30AG10161, P30AG72975, R01AG15819, R01AG17917, R01AG036836, U01AG46152, U01AG61356, U01AG046139, P50 AG016574, R01 AG032990, U01AG046139, R01AG018023, U01AG006576, U01AG006786, R01AG025711, R01AG017216, R01AG003949, R01NS080820, U24NS072026, P30AG19610, U01AG046170, RF1AG057440, and U24AG061340, as well as the Cure PSP, Mayo, and Michael J Fox foundations, Arizona Department of Health Services, and the Arizona Biomedical Research Commission. We thank the participants of the Religious Order Study and Memory and Aging projects for their generous donations, the Sun Health Research Institute Brain and Body Donation Program, Mayo Clinic Brain Bank, and Mount Sinai/JJ Peters VA Medical Center NIH Brain and Tissue Repository. Data and analysis contributing investigators included Nilüfer Ertekin-Taner, Steven Younkin (Mayo Clinic, Jacksonville, FL), Todd Golde (University of Florida), Nathan Price (Institute for Systems Biology), David Bennett, Christopher Gaiteri (Rush University), Philip De Jager (Columbia University), Bin Zhang, Eric Schadt, Michelle Ehrlich, Vahram Haroutunian, Sam Gandy (Icahn School of Medicine

458 at Mount Sinai), Koichi Iijima (National Center for Geriatrics and Gerontology, Japan), Scott
459 Noggle (New York Stem Cell Foundation), and Lara Mangravite (Sage Bionetworks).

460

461 **Author contributions**

462 **Yusuke Matsui:** Conceptualization, Methodology, Software, Validation, Formal analysis,
463 Investigation, Resources, Data Curation, Writing - Original draft, Writing - Reviewing and
464 Editing, Visualization, Funding acquisition. **Akira Togayachi**: Resources, Data Curation,
465 Writing - Reviewing and Editing **Kazuma Sakamoto**: Writing - Reviewing and Editing
466 **Kiyohiko Angata**: Writing - Reviewing and Editing. **Kenji Kadomatsu**: Supervision, Writing-
467 Reviewing and Editing, Project administration, Funding acquisition. **Shoko Nishihara:**
468 Conceptualization, Supervision, Writing- Reviewing and Editing, Project administration.

469 **Conflict of interest**

470 The authors declare no competing interests.

471 **Data availability**

472 The glycoproteomics datasets were compiled and used from the supplementary files
473 published in the respective papers[9,10]. For the AMP-AD transcriptome dataset, data were
474 obtained from the RNAseq Harmonization Study (RNAseq Harmonization) Repository
475 (syn21241740). Single-cell transcriptome data in the entorhinal cortex were obtained from
476 GSE138852. Brain vascular data were downloaded from GSE163577.

477

478

Refences

1. Guerreiro, R. & Bras, J. The age factor in Alzheimer's disease. *Genome Med.* **7**, 106 (2015).

2. Xia, X., Jiang, Q., McDermott, J. & Han, J.-D. J. Aging and Alzheimer's disease: Comparison and associations from molecular to system level. *Aging Cell* **17**, e12802 (2018).

3. Breijyeh, Z. & Karaman, R. Comprehensive Review on Alzheimer's Disease: Causes and Treatment. *Molecules* **25**, (2020).

4. Bennett, D. A. *et al.* Religious Orders Study and Rush Memory and Aging Project. *J. Alzheimers. Dis.* **64**, S161–S189 (2018).

5. Gaunitz, S., Tjernberg, L. O. & Schedin-Weiss, S. What Can N-glycomics and N-glycoproteomics of Cerebrospinal Fluid Tell Us about Alzheimer Disease? *Biomolecules* **11**, 858 (2021).

6. Schjoldager, K. T., Narimatsu, Y., Joshi, H. J. & Clausen, H. Global view of human protein glycosylation pathways and functions. *Nat. Rev. Mol. Cell Biol.* **21**, 729–749 (2020).

7. Narimatsu, Y. *et al.* An Atlas of Human Glycosylation Pathways Enables Display of the Human Glycome by Gene Engineered Cells. *Mol. Cell* **75**, 394-407.e5 (2019).

8. Joshi, H. J. *et al.* Glycosyltransferase genes that cause monogenic congenital disorders of glycosylation are distinct from glycosyltransferase genes associated with complex diseases. *Glycobiology* **28**, 284–294 (2018).

9. Zhang, Q., Ma, C., Chin, L.-S. & Li, L. Integrative glycoproteomics reveals protein N-glycosylation aberrations and glycoproteomic network alterations in Alzheimer's disease.

499 *Sci Adv* **6**, (2020).

500 10. Suttapitugsakul, S. *et al.* Glycoproteomics Landscape of Asymptomatic and Symptomatic
501 Human Alzheimer's Disease Brain. *Mol. Cell. Proteomics* **21**, 100433 (2022).

502 11. Yu, L. *et al.* Human Brain and Blood N-Glycome Profiling in Alzheimer's Disease and
503 Alzheimer's Disease-Related Dementias. *Front. Aging Neurosci.* **13**, 765259 (2021).

504 12. Kerdsaeng, N. *et al.* Serum Glycoproteomics and Identification of Potential Mechanisms
505 Underlying Alzheimer's Disease. *Behav. Neurol.* **2021**, 1434076 (2021).

506 13. Gizaw, S. T., Ohashi, T., Tanaka, M., Hinou, H. & Nishimura, S.-I. Glycoblotting method
507 allows for rapid and efficient glycome profiling of human Alzheimer's disease brain, serum
508 and cerebrospinal fluid towards potential biomarker discovery. *Biochim. Biophys. Acta*
509 **1860**, 1716–1727 (2016).

510 14. Gaunitz, S., Tjernberg, L. O. & Schedin-Weiss, S. The N-glycan profile in cortex and
511 hippocampus is altered in Alzheimer disease. *J. Neurochem.* **159**, 292–304 (2020).

512 15. Reyes, C. D. G. *et al.* LC-MS/MS Isomeric Profiling of N-Glycans Derived from Low-
513 Abundant Serum Glycoproteins in Mild Cognitive Impairment Patients. *Biomolecules* **12**,
514 (2022).

515 16. Frenkel-Pinter, M. *et al.* Interplay between protein glycosylation pathways in Alzheimer's
516 disease. *Sci Adv* **3**, e1601576 (2017).

517 17. Chen, C. C. *et al.* Altered serum glycomics in Alzheimer disease: a potential blood

biomarker? *Rejuvenation Res.* **13**, 439–444 (2010).

18. Schedin-Weiss, S. *et al.* Glycan biomarkers for Alzheimer disease correlate with T-tau and P-tau in cerebrospinal fluid in subjective cognitive impairment. *FEBS J.* **287**, 3221–3234 (2020).

19. Palmigiano, A. *et al.* CSF N-glycoproteomics for early diagnosis in Alzheimer's disease. *J. Proteomics* **131**, 29–37 (2015).

20. Cho, B. G., Veillon, L. & Mechref, Y. N-Glycan Profile of Cerebrospinal Fluids from Alzheimer's Disease Patients Using Liquid Chromatography with Mass Spectrometry. *J. Proteome Res.* **18**, 3770–3779 (2019).

21. Kizuka, Y. *et al.* An aberrant sugar modification of BACE1 blocks its lysosomal targeting in Alzheimer's disease. *EMBO Mol. Med.* **7**, 175–189 (2015).

22. Kizuka, Y. *et al.* Bisecting GlcNAc modification stabilizes BACE1 protein under oxidative stress conditions. *Biochem. J* **473**, 21–30 (2015).

23. Kizuka, Y., Kitazume, S., Sato, K. & Taniguchi, N. Clec4g (LSECtin) interacts with BACE1 and suppresses Aß generation. *FEBS Lett.* **589**, 1418–1422 (2015).

24. Bennun, S. V. *et al.* Systems Glycobiology: Integrating Glycogenomics, Glycoproteomics, Glycomics, and Other 'Omics Data Sets to Characterize Cellular Glycosylation Processes. *J. Mol. Biol.* **428**, 3337–3352 (2016).

25. West, C. M., Malzl, D., Hykollari, A. & Wilson, I. B. H. Glycomics, Glycoproteomics, and

537 Glycogenomics: An Inter-Taxa Evolutionary Perspective. *Mol. Cell. Proteomics* **20**, 100024
538 (2021).
539 26. Terry, A. V., Jr & Buccafusco, J. J. The cholinergic hypothesis of age and Alzheimer's
540 disease-related cognitive deficits: recent challenges and their implications for novel drug
541 development. *J. Pharmacol. Exp. Ther.* **306**, 821–827 (2003).
542 27. Sun, Y. *et al.* Role of the Extracellular Matrix in Alzheimer's Disease. *Front. Aging*
543 *Neurosci.* **13**, 707466 (2021).
544 28. Narimatsu, H. *et al.* GlycoGene Database (GGDB) on the Semantic Web. in *A Practical*
545 *Guide to Using Glycomics Databases* (ed. Aoki-Kinoshita, K. F.) 163–175 (Springer Japan,
546 2017).
547 29. Narimatsu, H. Construction of a human glycogene library and comprehensive functional
548 analysis. *Glycoconj. J.* **21**, 17–24 (2004).
549 30. *Handbook of Glycosyltransferases and Related Genes*. (Springer Japan).
550 31. Myllyharju, J. & Kivirikko, K. I. Collagens, modifying enzymes and their mutations in
551 humans, flies and worms. *Trends Genet.* **20**, 33–43 (2004).
552 32. Hennet, T. Collagen glycosylation. *Curr. Opin. Struct. Biol.* **56**, 131–138 (2019).
553 33. Heikkinen, J. *et al.* Lysyl Hydroxylase 3 Is a Multifunctional Protein Possessing Collagen
554 Glucosyltransferase Activity*. *J. Biol. Chem.* **275**, 36158–36163 (2000).
555 34. Wang, C. *et al.* The third activity for lysyl hydroxylase 3: galactosylation of hydroxylysyl

556 residues in collagens in vitro. *Matrix Biol.* **21**, 559–566 (2002).

557 35. Sjöstedt, E. *et al.* An atlas of the protein-coding genes in the human, pig, and mouse brain.
558 *Science* **367**, (2020).

559 36. Uhlén, M. *et al.* Proteomics. Tissue-based map of the human proteome. *Science* **347**,
560 1260419 (2015).

561 37. Grubman, A. *et al.* A single-cell atlas of entorhinal cortex from individuals with
562 Alzheimer's disease reveals cell-type-specific gene expression regulation. *Nat. Neurosci.*
563 **22**, 2087–2097 (2019).

564 38. Howett, D. *et al.* Differentiation of mild cognitive impairment using an entorhinal cortex-
565 based test of virtual reality navigation. *Brain* **142**, 1751–1766 (2019).

566 39. Khan, U. A. *et al.* Molecular drivers and cortical spread of lateral entorhinal cortex
567 dysfunction in preclinical Alzheimer's disease. *Nat. Neurosci.* **17**, 304–311 (2014).

568 40. Bottero, V., Powers, D., Yalamanchi, A., Quinn, J. P. & Potashkin, J. A. Key Disease
569 Mechanisms Linked to Alzheimer's Disease in the Entorhinal Cortex. *Int. J. Mol. Sci.* **22**,
570 (2021).

571 41. Lepelletier, F.-X., Mann, D. M. A., Robinson, A. C., Pinteaux, E. & Boutin, H. Early
572 changes in extracellular matrix in Alzheimer's disease. *Neuropathol. Appl. Neurobiol.* **43**,
573 167–182 (2015).

574 42. Odom, G. J. *et al.* PathwayPCA: An R/Bioconductor package for pathway based

integrative analysis of multi-omics data. *Proteomics* **20**, e1900409 (2020).

43. Chen, X., Wang, L., Smith, J. D. & Zhang, B. Supervised principal component analysis for gene set enrichment of microarray data with continuous or survival outcomes. *Bioinformatics* **24**, 2474–2481 (2008).

44. Chen, X. *et al.* Pathway-based analysis for genome-wide association studies using supervised principal components. *Genet. Epidemiol.* **34**, 716–724 (2010).

45. Yang, A. C. *et al.* A human brain vascular atlas reveals diverse mediators of Alzheimer's risk. *Nature* **603**, 885–892 (2022).

46. Xu, L., Nirwane, A. & Yao, Y. Basement membrane and blood-brain barrier. *Stroke Vasc Neurol* **4**, 78–82 (2019).

47. Guerrero-Juarez, C. F. *et al.* Single-cell analysis reveals fibroblast heterogeneity and myeloid-derived adipocyte progenitors in murine skin wounds. *Nat. Commun.* **10**, 650 (2019).

48. Browaeys, R., Saelens, W. & Saeys, Y. NicheNet: modeling intercellular communication by linking ligands to target genes. *Nat. Methods* **17**, 159–162 (2020).

49. Lang, F., Stournaras, C., Zacharopoulou, N., Voelkl, J. & Alesutan, I. Serum- and glucocorticoid-inducible kinase 1 and the response to cell stress. *Cell Stress Chaperones* **3**, 1–8 (2018).

50. Leong, M. L. L., Maiyar, A. C., Kim, B., O'Keeffe, B. A. & Firestone, G. L. Expression of

the serum- and glucocorticoid-inducible protein kinase, Sgk, is a cell survival response to multiple types of environmental stress stimuli in mammary epithelial cells. *J. Biol. Chem.* **278**, 5871–5882 (2003).

51. Webster, M. K., Goya, L., Ge, Y., Maiyar, A. C. & Firestone, G. L. Characterization of sgk, a novel member of the serine/threonine protein kinase gene family which is transcriptionally induced by glucocorticoids and serum. *Mol. Cell. Biol.* **13**, 2031–2040 (1993).

52. Morabito, S., Reese, F., Rahimzadeh, N., Miyoshi, E. & Swarup, V. hdWGCNA identifies co-expression networks in high-dimensional transcriptomics data. *Cell Rep Methods* **3**, 100498 (2023).

53. Valdeolivas, A. *et al.* Random walk with restart on multiplex and heterogeneous biological networks. *Bioinformatics* **35**, 497–505 (2019).

54. Anwar, M. M., Özkan, E. & Gürsoy-Özdemir, Y. The role of extracellular matrix alterations in mediating astrocyte damage and pericyte dysfunction in Alzheimer's disease: A comprehensive review. *Eur. J. Neurosci.* **56**, 5453–5475 (2021).

55. Ma, J. *et al.* Extracellular Matrix Proteins Involved in Alzheimer's Disease. *Chemistry* **26**, 12101–12110 (2020).

56. Levy, A. D., Omar, M. H. & Koleske, A. J. Extracellular matrix control of dendritic spine and synapse structure and plasticity in adulthood. *Front. Neuroanat.* **8**, 116 (2014).

613 57. Morawski, M., Filippov, M., Tzinia, A., Tsilibary, E. & Vargova, L. ECM in brain aging and
614 dementia. *Prog. Brain Res.* **214**, 207–227 (2014).
615 58. Soleman, S., Filippov, M. A., Dityatev, A. & Fawcett, J. W. Targeting the neural
616 extracellular matrix in neurological disorders. *Neuroscience* **253**, 194–213 (2013).
617 59. Sethi, M. K. & Zaia, J. Extracellular matrix proteomics in schizophrenia and Alzheimer's
618 disease. *Anal. Bioanal. Chem.* **409**, 379–394 (2016).
619 60. Bosiacki, M. *et al.* Perineuronal Nets and Their Role in Synaptic Homeostasis. *Int. J. Mol.*
620 *Sci.* **20**, (2019).
621 61. Li, Y., Li, Z.-X., Jin, T., Wang, Z.-Y. & Zhao, P. Tau Pathology Promotes the Reorganization
622 of the Extracellular Matrix and Inhibits the Formation of Perineuronal Nets by
623 Regulating the Expression and the Distribution of Hyaluronic Acid Synthases. *J.*
624 *Alzheimers. Dis.* **57**, 395–409 (2017).
625 62. Duncan, J. A., Foster, R. & Kwok, J. C. F. The potential of memory enhancement through
626 modulation of perineuronal nets. *Br. J. Pharmacol.* **176**, 3611–3621 (2019).
627 63. Wen, T. H., Binder, D. K., Ethell, I. M. & Razak, K. A. The Perineuronal "Safety" Net?
628 Perineuronal Net Abnormalities in Neurological Disorders. *Front. Mol. Neurosci.* **11**, 270
629 (2018).
630 64. Pantazopoulos, H. & Berretta, S. In Sickness and in Health: Perineuronal Nets and
631 Synaptic Plasticity in Psychiatric Disorders. *Neural Plast.* **2016**, 9847696 (2015).

65. Logsdon, A. F. *et al.* Decoding perineuronal net glycan sulfation patterns in the Alzheimer's disease brain. *Alzheimers. Dement.* **18**, 942–954 (2021).

66. Miyata, S. & Kitagawa, H. Chondroitin sulfate and neuronal disorders. *Front. Biosci.* **21**, 1330–1340 (2016).

67. Galtrey, C. M. & Fawcett, J. W. The role of chondroitin sulfate proteoglycans in regeneration and plasticity in the central nervous system. *Brain Res. Rev.* **54**, 1–18 (2007).

68. Yamada, J. & Jinno, S. Spatio-temporal differences in perineuronal net expression in the mouse hippocampus, with reference to parvalbumin. *Neuroscience* **253**, 368–379 (2013).

69. Brandan, E. & Inestrosa, N. C. Extracellular matrix components and amyloid in neuritic plaques of Alzheimer's disease. *Gen. Pharmacol.* **24**, 1063–1068 (1993).

70. Zhang, S. *et al.* The extracellular matrix enriched with membrane metalloendopeptidase and insulin-degrading enzyme suppresses the deposition of amyloid-beta peptide in Alzheimer's disease cell models. *J. Tissue Eng. Regen. Med.* **13**, 1759–1769 (2019).

71. Wu, A., Pangalos, M. N., Efthimiopoulos, S., Shioi, J. & Robakis, N. K. Appican expression induces morphological changes in C6 glioma cells and promotes adhesion of neural cells to the extracellular matrix. *J. Neurosci.* **17**, 4987–4993 (1997).

72. Small, D. H., Nurcombe, V., Clarris, H., Beyreuther, K. & Masters, C. L. The role of extracellular matrix in the processing of the amyloid protein precursor of Alzheimer's disease. *Ann. N. Y. Acad. Sci.* **695**, 169–174 (1993).

651 73. Stoyanov, S. *et al.* Attenuation of the extracellular matrix restores microglial activity
652 during the early stage of amyloidosis. *Glia* **69**, 182–200 (2020).
653 74. Bronfman, F. C., Soto, C., Tapia, L., Tapia, V. & Inestrosa, N. C. Extracellular matrix
654 regulates the amount of the beta-amyloid precursor protein and its amyloidogenic
655 fragments. *J. Cell. Physiol.* **166**, 360–369 (1996).
656 75. Salza, R., Lethias, C. & Ricard-Blum, S. The Multimerization State of the Amyloid-ß42
657 Amyloid Peptide Governs its Interaction Network with the Extracellular Matrix. *J.*
658 *Alzheimers. Dis.* **56**, 991–1005 (2017).
659 76. Potjewyd, G., Kellett, K. A. B. & Hooper, N. M. 3D hydrogel models of the neurovascular
660 unit to investigate blood-brain barrier dysfunction. *Neuronal Signal* **5**, NS20210027
661 (2021).
662 77. Reed, M. J., Damodarasamy, M. & Banks, W. A. The extracellular matrix of the blood-
663 brain barrier: structural and functional roles in health, aging, and Alzheimer's disease.
664 *Tissue Barriers* **7**, 1651157 (2019).
665 78. Damodarasamy, M. *et al.* The microvascular extracellular matrix in brains with
666 Alzheimer's disease neuropathologic change (ADNC) and cerebral amyloid angiopathy
667 (CAA). *Fluids Barriers CNS* **17**, 60 (2020).
668 79. Brzdak, P., Nowak, D., Wiera, G. & Mozrzymas, J. W. Multifaceted Roles of Metzincins in
669 CNS Physiology and Pathology: From Synaptic Plasticity and Cognition to

Neurodegenerative Disorders. *Front. Cell. Neurosci.* **11**, 178 (2017).

80. Bres, E. E. & Faissner, A. Low Density Receptor-Related Protein 1 Interactions With the Extracellular Matrix: More Than Meets the Eye. *Front Cell Dev Biol* **7**, 31 (2019).

81. Wang, C. *et al.* The glycosyltransferase activities of lysyl hydroxylase 3 (LH3) in the extracellular space are important for cell growth and viability. *J. Cell. Mol. Med.* **13**, 508–521 (2009).

82. Salo, A. M. *et al.* Lysyl hydroxylase 3 (LH3) modifies proteins in the extracellular space, a novel mechanism for matrix remodeling. *J. Cell. Physiol.* **207**, 644–653 (2006).

83. Valtavaara, M., Szpirer, C., Szpirer, J. & Myllylä, R. Primary Structure, Tissue Distribution, and Chromosomal Localization of a Novel Isoform of Lysyl Hydroxylase (Lysyl Hydroxylase 3)*. *J. Biol. Chem.* **273**, 12881–12886 (1998).

84. De Giorgi, F., Fumagalli, M., Scietti, L. & Forneris, F. Collagen hydroxylysine glycosylation: non-conventional substrates for atypical glycosyltransferase enzymes. *Biochem. Soc. Trans.* **49**, 855–866 (2021).

85. Zhou, J. *et al.* Cerebral small vessel disease caused by PLOD3 mutation: Expanding the phenotypic spectrum of lysyl hydroxylase-3 deficiency. *Pediatr Investig* **6**, 219–223 (2022).

86. Miyatake, S. *et al.* Biallelic COLGALT1 variants are associated with cerebral small vessel disease. *Ann. Neurol.* **84**, 843–853 (2018).

87. Li, H. *et al.* Overexpression of LH3 reduces the incidence of hypertensive intracerebral

689 hemorrhage in mice. *J. Cereb. Blood Flow Metab.* **39**, 547–561 (2019).

690 88. Toden, S. *et al.* Noninvasive characterization of Alzheimer's disease by circulating, cell-
691 free messenger RNA next-generation sequencing. *Sci Adv* **6**, (2020).

692 89. Garcia, F. J. *et al.* Single-cell dissection of the human brain vasculature. *Nature* **603**, 893–
693 899 (2022).

694 90. Vanlandewijck, M. *et al.* A molecular atlas of cell types and zonation in the brain
695 vasculature. *Nature* **554**, 475–480 (2018).

696 91. Palhol, J. S. C. *et al.* Direct association with the vascular basement membrane is a
697 frequent feature of myelinating oligodendrocytes in the neocortex. *Fluids Barriers CNS*
698 **20**, 24 (2023).

699 92. Kwon, O.-C. *et al.* SGK1 inhibition in glia ameliorates pathologies and symptoms in
700 Parkinson disease animal models. *EMBO Mol. Med.* **13**, e13076 (2021).

701 93. Linnerbauer, M., Wheeler, M. A. & Quintana, F. J. Astrocyte Crosstalk in CNS
702 Inflammation. *Neuron* **108**, 608–622 (2020).

703 94. Chen, X. Adaptive elastic-net sparse principal component analysis for pathway
704 association testing. *Stat. Appl. Genet. Mol. Biol.* **10**, (2011).

705 95. Dewey, M. Meta-Analysis of Significance Values [R package metap version 1.8]. (2022).

706 96. Garcia-Alonso, L., Holland, C. H., Ibrahim, M. M., Turei, D. & Saez-Rodriguez, J.
707 Benchmark and integration of resources for the estimation of human transcription factor

activities. *Genome Res.* **29**, 1363–1375 (2019).

97. Alvarez, M. J. *et al.* Functional characterization of somatic mutations in cancer using network-based inference of protein activity. *Nat. Genet.* **48**, 838–847 (2016).

98. Badia-I-Mompel, P. *et al.* decoupleR: ensemble of computational methods to infer biological activities from omics data. *Bioinform Adv* **2**, vbac016 (2022).

Figure 1 - Hyperglycosylated proteins are primarily enriched in the ECM.

A Analysis of glycoprotein data from 2 AD cohorts, using glycoproteins from prefrontal tissues of 2 independent AD cohorts. The first cohort (AD cohort 1) consisted of 8 samples each from healthy subjects and those with AD, and the second cohort (AD cohort 2) consisted of 10 samples each from healthy subjects, asymptomatic AD, and symptomatic AD. In each cohort, 92 and 10 AD-specific glycoproteins were identified, respectively.

B Pathway enrichment of AD-related glycoproteins. Over-representation analysis of AD-specific glycoproteins was performed.

C Significantly enriched pathways that were common in both cohorts are shown. The horizontal axis is the p-value representing the enrichment, which is the logarithm of the nominal p-value multiplied by a negative value.

Figures were created with BioRender.com.

Figure 2 - Meta-analysis of the global transcriptome reveals that glycogenes are enriched in the ECM.

A Number of glycogenes constituting the glycosylation pathway used for transcriptome analysis.

B Meta-analysis of differential gene expression in multiple AD cohorts. Transcriptome data from 3 AD cohorts: the Mayo cohort (n = 313), the MSSM cohort (n = 315), and the ROSMAP cohort (n = 1168). A meta-analysis of DEGs based on gene-level expression levels (FDR < 5%) was performed; 46 glycogenes were identified as DEGs. In the volcano plot, the horizontal axis represents the effect size summarizing the difference in expression between the non-AD and AD groups across cohorts, and the vertical axis represents the log of the p-value from the meta-analysis (bottom is 10) multiplied by a negative value.

C Mapped glycosyltransferase DEGs. In total, 46 glycogene DEGs were mapped. Genes overexpressed in the meta-analysis are shown in red, genes underexpressed are shown in blue, and genes that did not show significant mutations are shown in gray. Genes are classified into 16 major glycosylation pathways, including initiation, core elongation, elongation/branching, capping, and sulfation. Glycosyltransferases with and without pathway specificity are also distinguished.

Figure created by BioRender.com.

Figure 3 - PLOD3 is identified as a hub glycogene for the ECM.

A Comparison of AD trascriptome and glycoprotein signatures. Common pathways are shown.

B Relationship between glycogenes and globally enriched pathways. Orange nodes represent globally enriched pathways, and green nodes represent glycogen enriched in each pathway.

C Functional hub glycogenes in globally enriched pathways. To identify functional hub glycogenes involved in multiple pathways, we constructed a pathway-gene bipartite graph, calculated the degree of each glycogene (number of genes directly connected to the pathway), and ranked the importance of each glycogene. The vertical axis of the bar graph represents the order of each glycogene.

D Activity changes in PLOD3, ECM, and collagen formation in AD brains in each transcriptome cohort. Forest plots of Log2 fold changes in PLOD3, ECM organization, and collagen formation activity between non-AD and AD are plotted by cohort and brain region.

DLPFC: dorsolateral prefrontal cortex; STG: superior temporal gyrus; PHG: parahippocampal gyrus; IFG: inferior frontal gyrus; FP: frontal pole; TCX: temporal cortex; CBE: cerebellum. Dots indicate estimated mean effect sizes, bar widths are 95% confidence intervals of the estimates, and vertical lines with red dots indicate zero (no change).

Figure 4 - PLOD3 is expressed in oligodendrocytes and is co-expressed with COL4A5.

A Cell-type specificity of PLOD3 in healthy brain tissues. Cell clusters obtained from gene expression in healthy brain tissue by Human Protein Atlas (v22) scRNA-seq, and the Transcripts Per Million (TPM) in each cluster. PLOD3 and COL4A5 are highly expressed in oligodendrocytes and belong to the same cluster.

B Expression levels of PLOD3 and COL4A5 per cell type.

C Cellular specificity of PLOD3 and collagen in the enthorhinal cortex. Scatter plots show the cluster structure of cell populations projected by UMAP to 2D coordinates based on gene expression; the first panel shows cell types, the second non-AD and AD, the third and fourth panels show cell type-specific expression of PLOD3 and COL4A5 in oligodendrocytes, respectively.

D Correlation of COL4A5 with the expression of APP (upper panel) and the activity of amyloid fiber formation (lower panel) for each cohort and each region.

DLPFC: dorsolateral prefrontal cortex; STG: superior temporal gyrus; PHG: parahippocampal gyrus; IFG: inferior frontal gyrus; FP: frontal pole; TCX: temporal cortex; CBE: cerebellum.

Figure 5 - Cerebrovasculature most strongly associated with ECM dysregulation.

A Pathways significantly associated with the activity of the ECM organization were estimated for each cohort tissue using the AES-PCA model. The p-values estimated for each cohort and for each brain tissue were estimated as integrated p-values, and the top 10 pathways are shown in the figure. Figures were generated by BioRender.com.

B Enrichment of pathways involving the cerebrovasculature in AD with gene set enrichment analysis (GSEA) (FDR < 5%). Forest plots shown below each enrichment plot indicate Log2 fold change for each pathway in each cohort and each region.

C Analysis using cerebrovascular scRNA-seq data (8 NCI, 9 AD).

D Expression of PLOD3 per cell type.

E Expression of type IV collagen per cell type.

DLPFC: dorsolateral prefrontal cortex; STG: superior temporal gyrus; PHG: parahippocampal gyrus; IFG: inferior frontal gyrus; FP: frontal pole; TCX: temporal cortex; CBE: cerebellum.

Figure 6 - COL4A5 ligand is involved in the regulatory cascade of the astrocyte stress response.

A Estimated cell-to-cell communication based on ECM ligand–receptor expression.

B Receptor candidates for oligodendrocyte-derived COL4A5 ligands predicted to bind in astrocytes.

C Cell type-specific expression levels of receptors for COL4A5.

D COL4A5-mediated signaling pathways and target genes in astrocytes.

E Expression levels of target genes BCL6 and SGK1 per cell type.

F Analysis flow of the exploration of neighboring genes and functional estimation using network propagation in astrocyte-specific co-expression networks.

G Top 30 neighboring genes estimated by network propagation based on BCL6 and SGK1.

H Gene set analysis of BCL6 and SGK1 neighbor genes.

790 **Supplementary Figure 1 - Correlation of PLOD3 with APP activity (upper panel) and amyloid fiber formation**
791 **(lower panel) for each cohort and each region.**
792 DLPFC: dorsolateral prefrontal cortex; STG: superior temporal gyrus; PHG: parahippocampal gyrus; IFG:
793 inferior frontal gyrus; FP: frontal pole; TCX: temporal cortex; CBE: cerebellum.

Supplementary Figure 2 - ECM activity is strongly associated with inflammatory cytokines.

A Enrichment of pathways involving the immune system in AD with GSEA (FDR < 5%). Forest plots shown below each enrichment plot indicate Log2 fold change for each pathway in each cohort and each tissue.

B Pathways significantly associated with ECM activity obtained by applying AES-PCA for each cohort and region.

DLPFC: dorsolateral prefrontal cortex; STG: superior temporal gyrus; PHG: parahippocampal gyrus; IFG: inferior frontal gyrus; FP: frontal pole; TCX: temporal cortex; CBE: cerebellum.

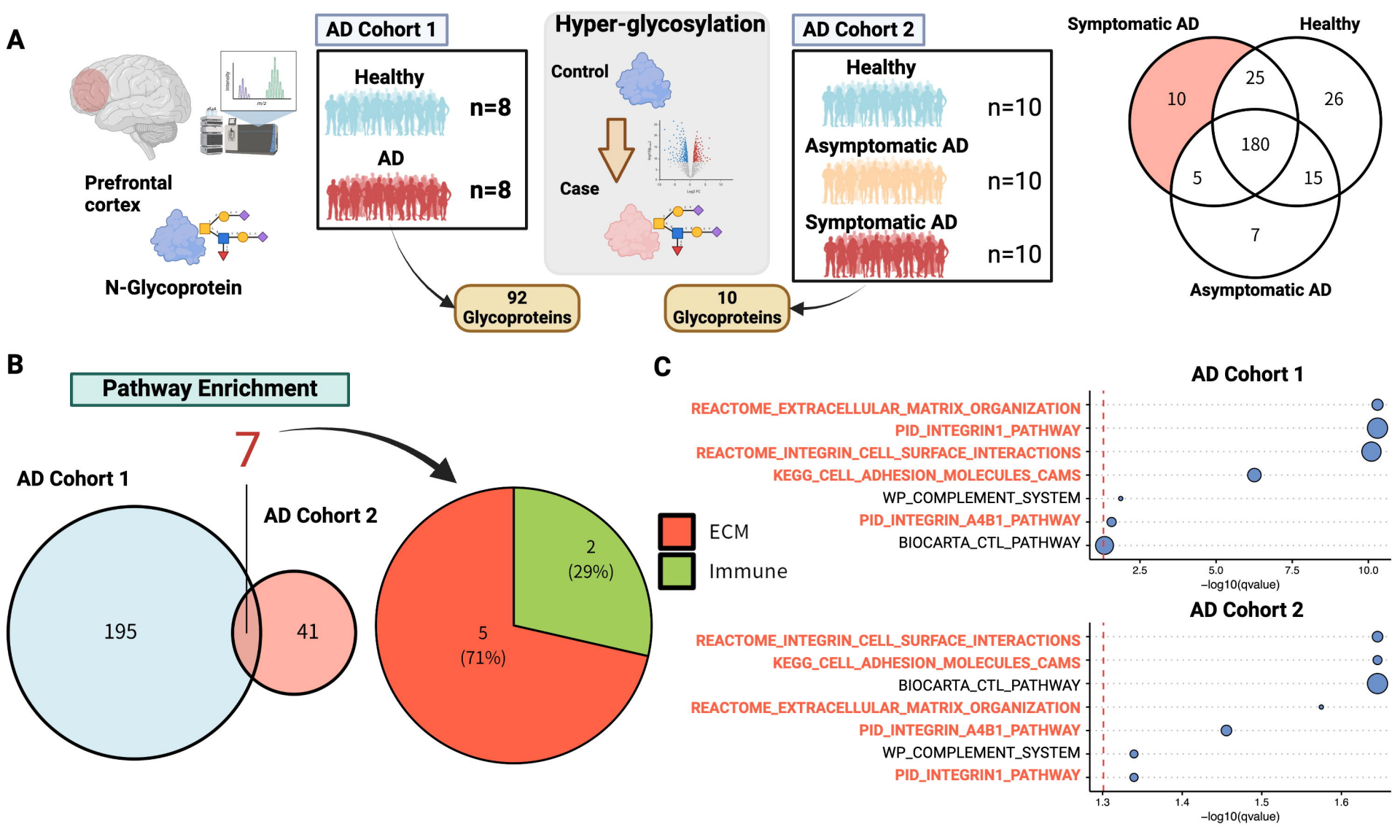
Supplementary Figure 3 - COL4A5 ligand is involved in the regulatory cascade of the astrocyte stress response.

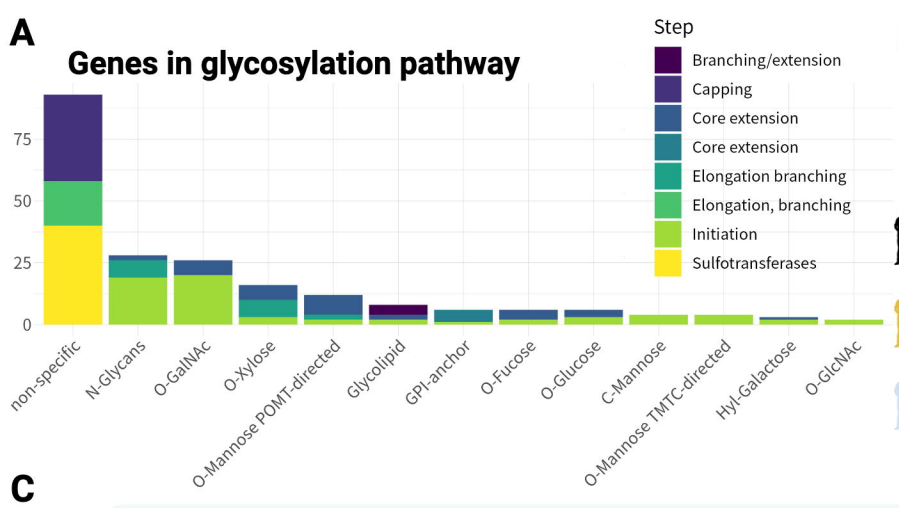
A Top 30 neighboring genes estimated by network propagation based on BCL6.

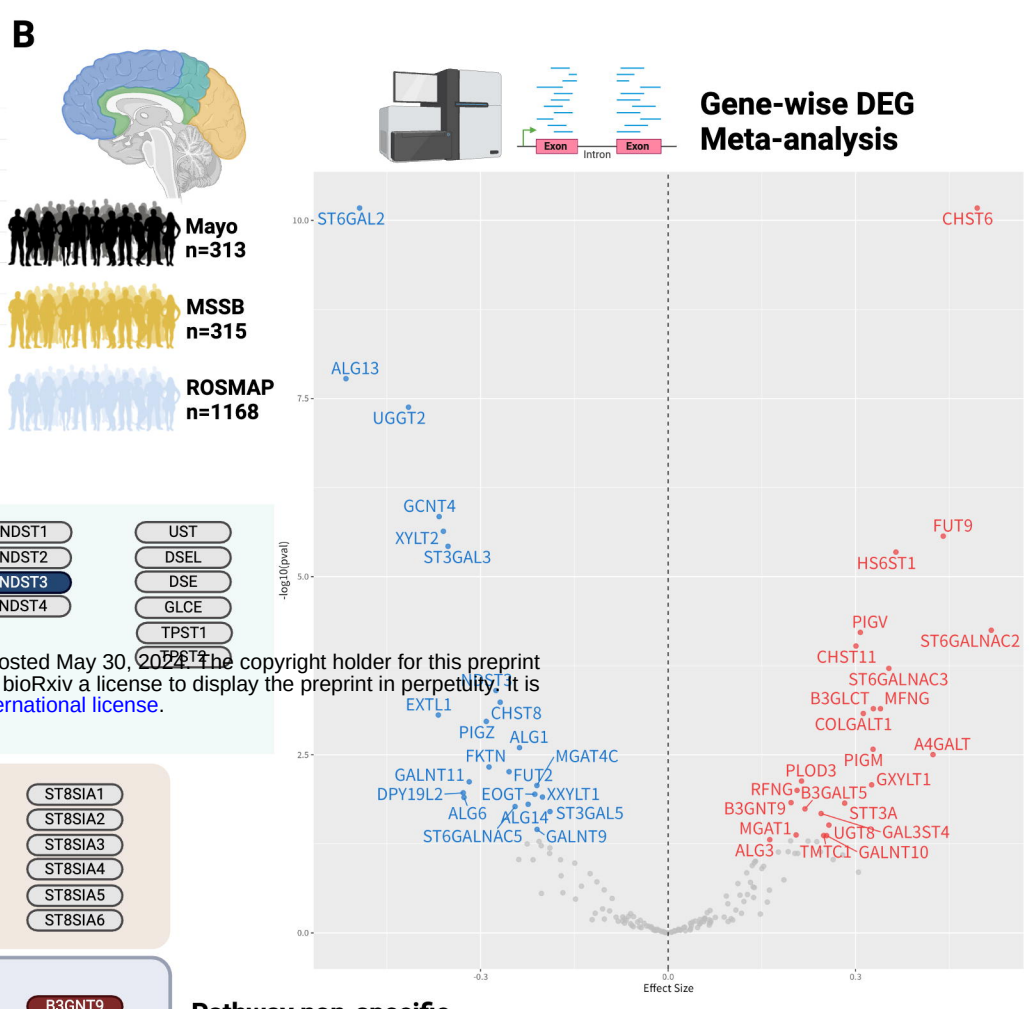
B Gene set analysis of BCL6 neighbor genes.

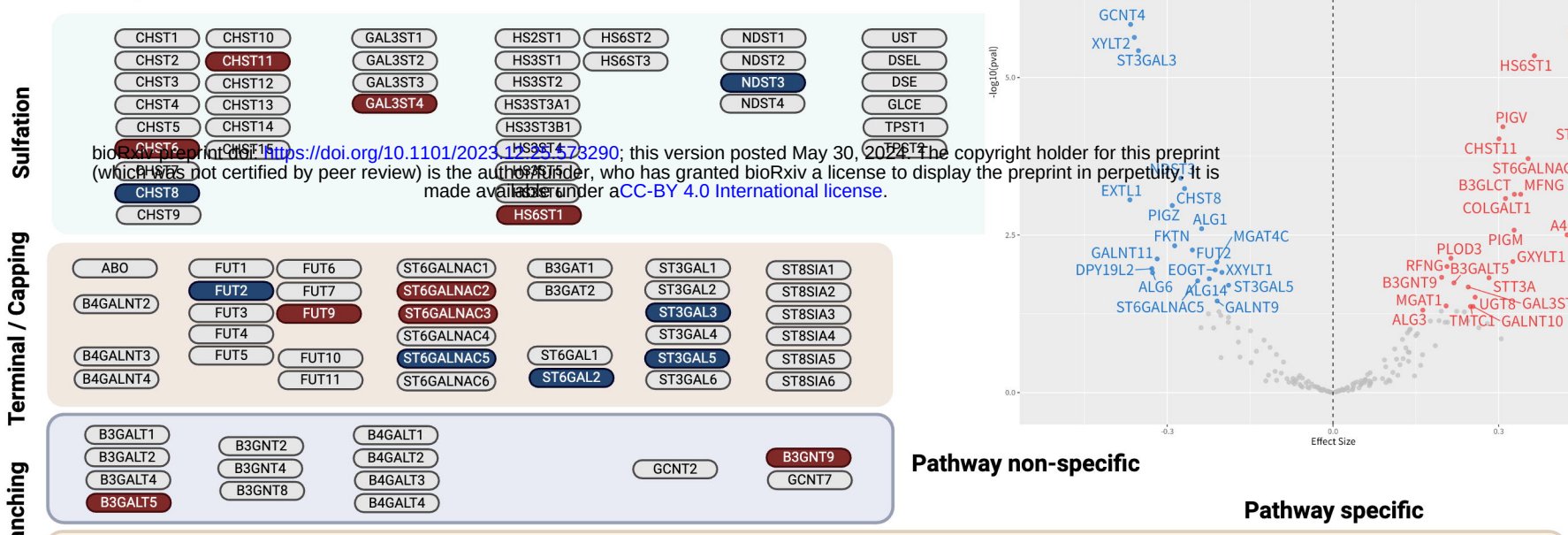
C Top 30 neighboring genes estimated by network propagation based on SGK1.

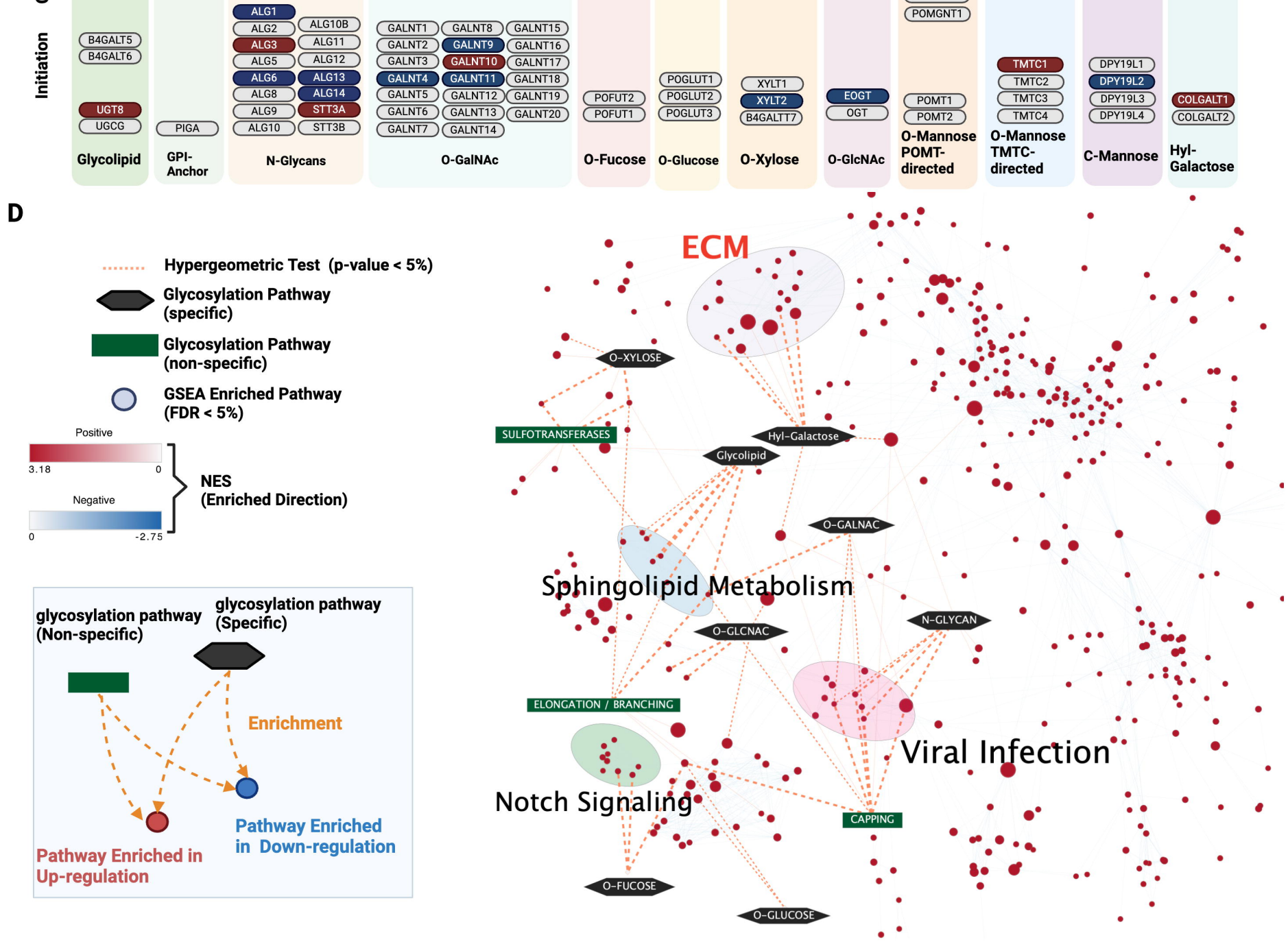
D Gene set analysis of SGK1 neighbor genes.

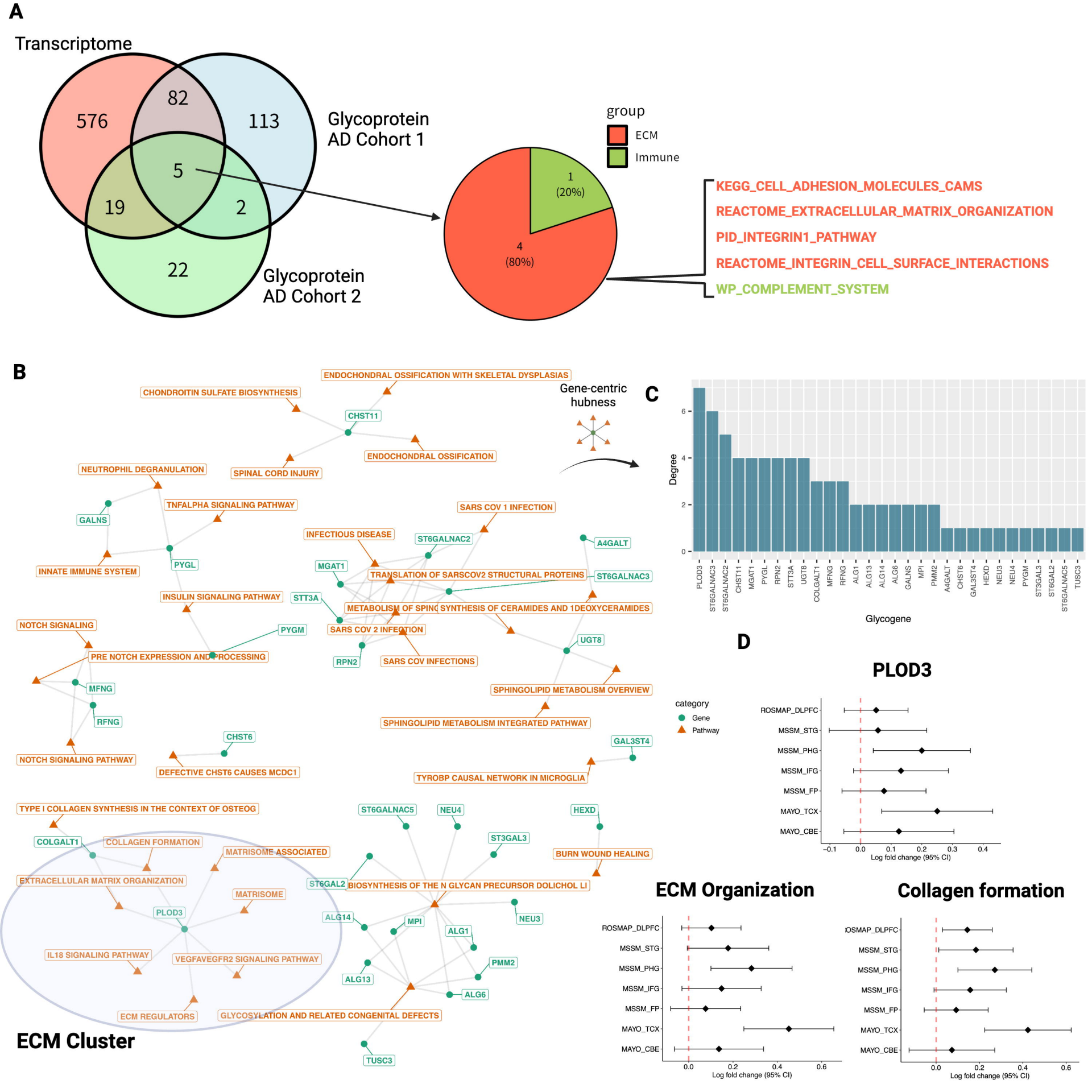

A
Prefrontal cortex
N-Glycoprotein
AD Cohort 1
Healthy
n=8
AD
n=8
92 Glycoproteins
Hyper-glycosylation
Control
Case
AD Cohort 2
Healthy
n=10
Asymptomatic AD
n=10
Symptomatic AD
n=10
10 Glycoproteins
Symptomatic AD
Healthy
Asymptomatic AD
10
25
26
180
5
15
7
B
Pathway Enrichment
AD Cohort 1
AD Cohort 2
7
195
41
5
(71%)
2
(29%)
ECM
Immune
C
AD Cohort 1
REACTOME_EXTRACELLULAR_MATRIX_ORGANIZATION
PID_INTEGRIN1_PATHWAY
REACTOME_INTEGRIN_CELL_SURFACE_INTERACTIONS
KEGG_CELL_ADHESION_MOLECULES_CAMS
WP_COMPLEMENT_SYSTEM
PID_INTEGRIN_A4B1_PATHWAY
BIOCARTA_CTL_PATHWAY
2.5
5.0
7.5
10.0
−log10(qvalue)
AD Cohort 2
REACTOME_INTEGRIN_CELL_SURFACE_INTERACTIONS
KEGG_CELL_ADHESION_MOLECULES_CAMS
BIOCARTA_CTL_PATHWAY
REACTOME_EXTRACELLULAR_MATRIX_ORGANIZATION
PID_INTEGRIN_A4B1_PATHWAY
WP_COMPLEMENT_SYSTEM
PID_INTEGRIN1_PATHWAY
1.3
1.4
1.5
1.6
−log10(qvalue)

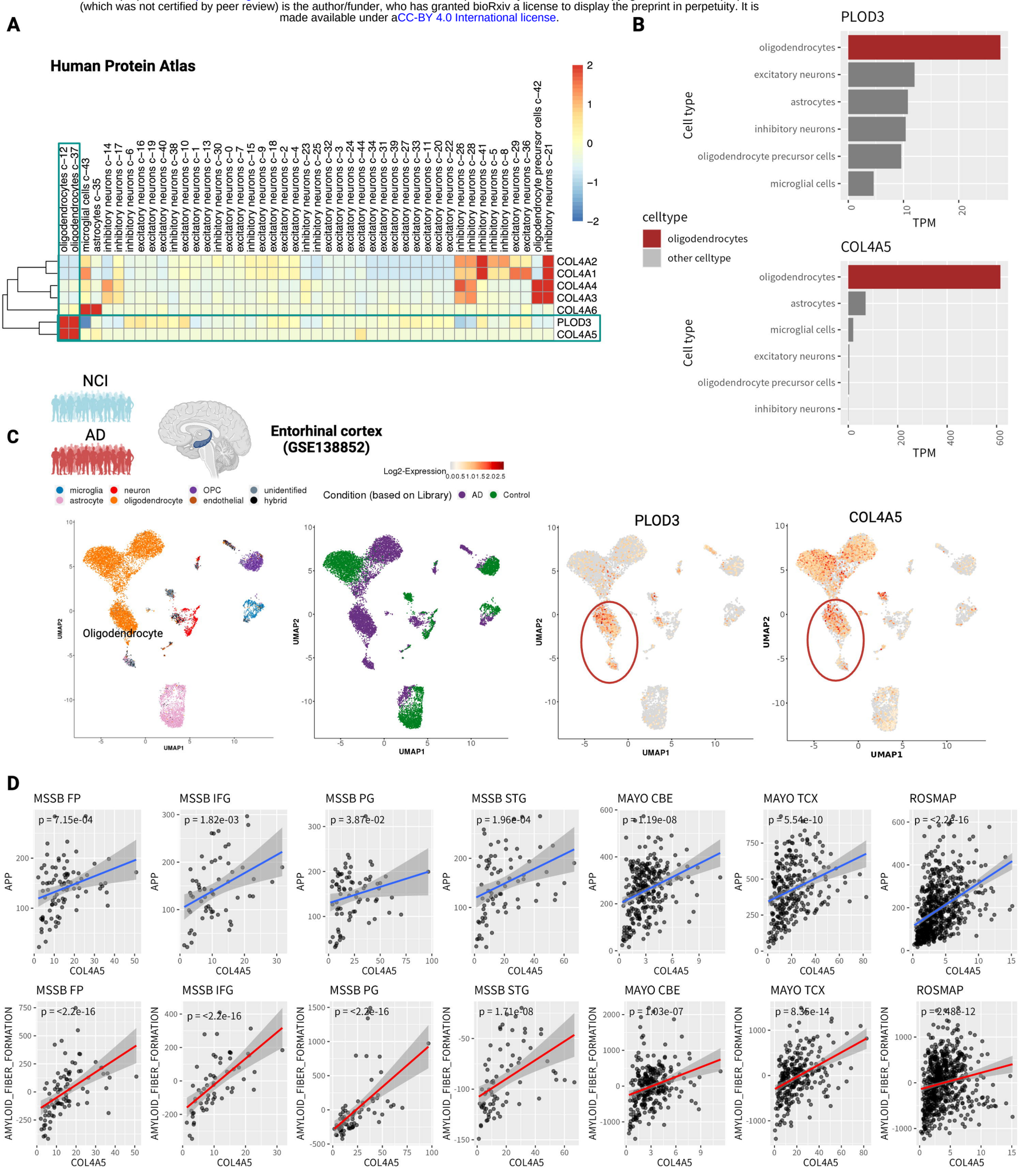


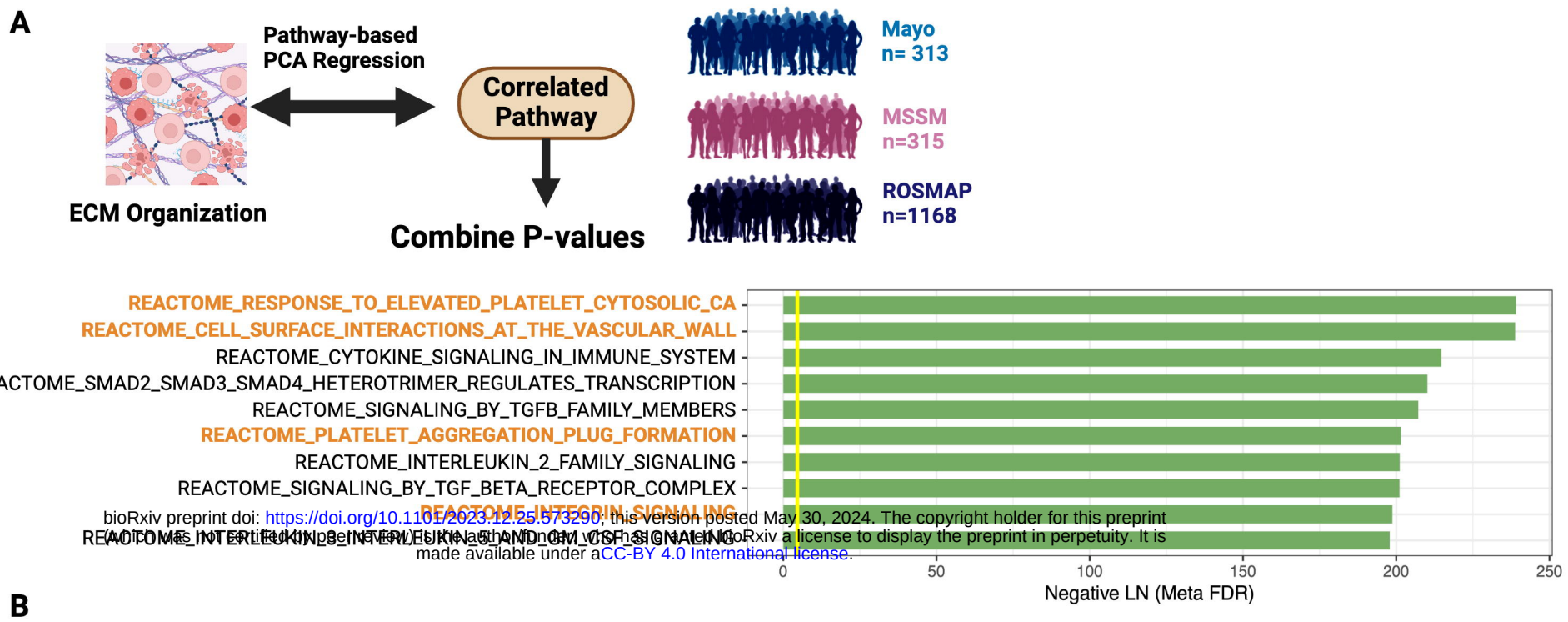


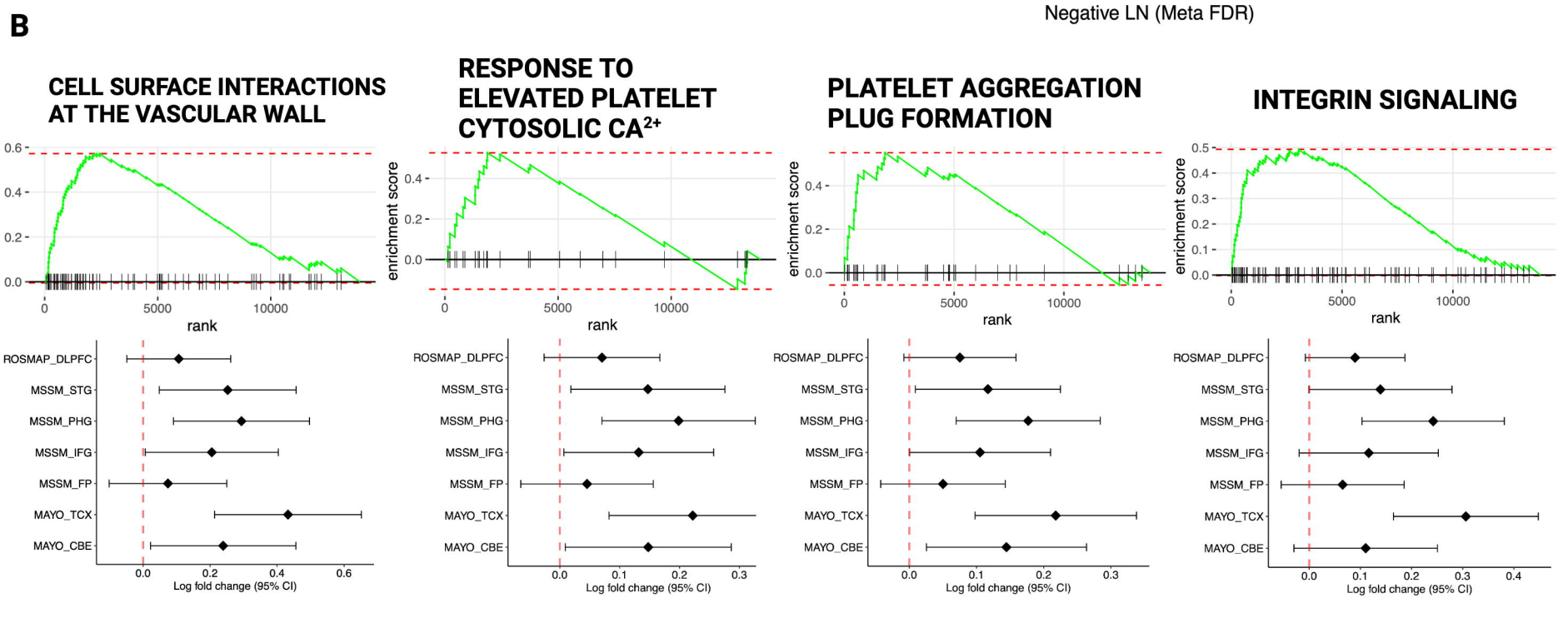


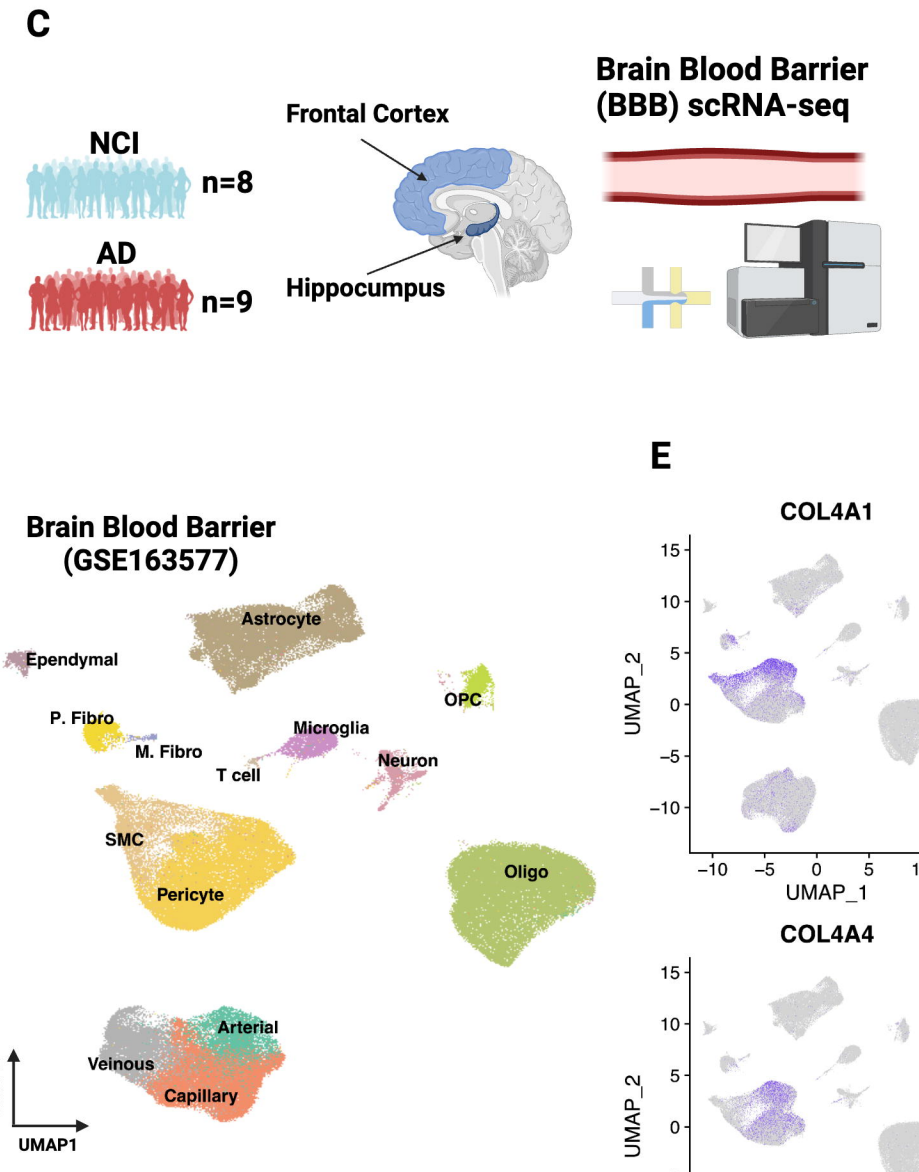


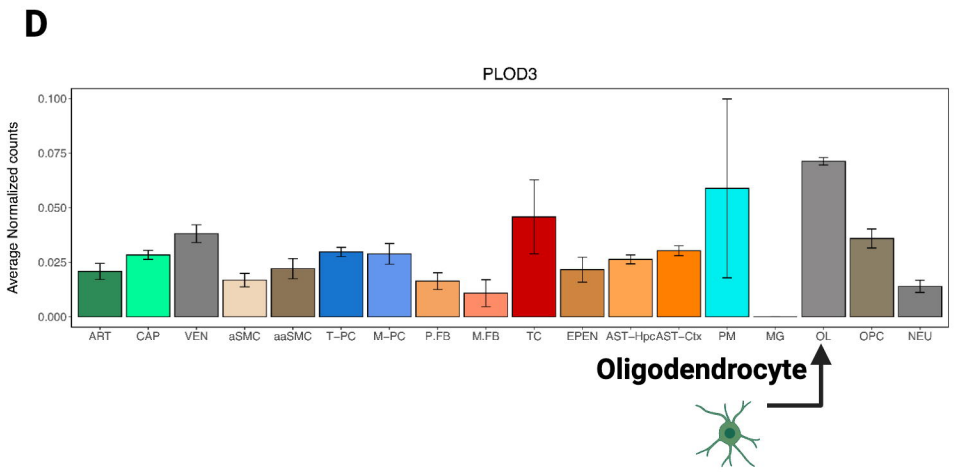

A
Transcriptome
576
82
113
5
19
2
22
Glycoprotein
AD Cohort 1
Glycoprotein
AD Cohort 2
group
ECM
Immune
1
(20%)
4
(80%)
KEGG_CELL_ADHESION_MOLECULES_CAMS
REACTOME_EXTRACELLULAR_MATRIX_ORGANIZATION
PID_INTEGRIN1_PATHWAY
REACTOME_INTEGRIN_CELL_SURFACE_INTERACTIONS
WP_COMPLEMENT_SYSTEM
B
ENDOCHONDRAL OSSIFICATION WITH SKELETAL DYSPLASIAS
CHONDROITIN SULFATE BIOSYNTHESIS
CHST11
ENDOCHONDRAL OSSIFICATION
SPINAL CORD INJURY
NEUTROPHIL DEGRANULATION
TNFALPHA SIGNALING PATHWAY
GALNS
INNATE IMMUNE SYSTEM
PYGL
SARS COV 1 INFECTION
INFECTIOUS DISEASE
ST6GALNAC2
A4GALT
MGAT1
TRANSLATION OF SARSCOV2 STRUCTURAL PROTEINS
ST6GALNAC3
INSULIN SIGNALING PATHWAY
STT3A
METABOLISM OF SPING SYNTHESIS OF CERAMIDES AND 1DEOXYCERAMIDES
NOTCH SIGNALING
PYGM
SARS COV 2 INFECTION
UGT8
PRE NOTCH EXPRESSION AND PROCESSING
RPN2
SARS COV INFECTIONS
MFNG
SPHINGOLIPID METABOLISM OVERVIEW
RFNG
SPHINGOLIPID METABOLISM INTEGRATED PATHWAY
CHST6
GAL3ST4
NOTCH SIGNALING PATHWAY
DEFECTIVE CHST6 CAUSES MCDC1
TYROBP CAUSAL NETWORK IN MICROGLIA
TYPE I COLLAGEN SYNTHESIS IN THE CONTEXT OF OSTEOG
ST6GALNAC5
NEU4
HEXD
COLGALT1
COLLAGEN FORMATION
ST3GAL3
MATRISOME ASSOCIATED
BURN WOUND HEALING
EXTRACELLULAR MATRIX ORGANIZATION
ST6GAL2
BIOSYNTHESIS OF THE N GLYCAN PRECURSOR DOLICHOL LI
MATRISOME
PLOD3
NEU3
ALG14
MPI
ALG1
IL18 SIGNALING PATHWAY
VEGFAVEGFR2 SIGNALING PATHWAY
PMM2
ALG13
ALG6
ECM REGULATORS
GLYCOSYLATION AND RELATED CONGENITAL DEFECTS
TUSC3
ECM Cluster
Gene-centric
hubness
C
Degree
Glycogene
PLOD3
ST6GALNAC3
ST6GALNAC2
CHST11
MGAT1
PYGL
RPN2
STT3A
UGT8
COLGALT1
MFNG
RFNG
ALG1
ALG13
ALG14
ALG6
GALNS
MPI
PMM2
A4GALT
CHST6
GAL3ST4
HEXD
NEU3
NEU4
PYGM
ST3GAL3
ST6GAL2
ST6GALNAC5
TUSC3
D
PLOD3
category
Gene
Pathway
ROSMAP_DLPFC
MSSM_STG
MSSM_PHG
MSSM_IFG
MSSM_FP
MAYO_TCX
MAYO_CBE
Log fold change (95% CI)
ECM Organization
Collagen formation

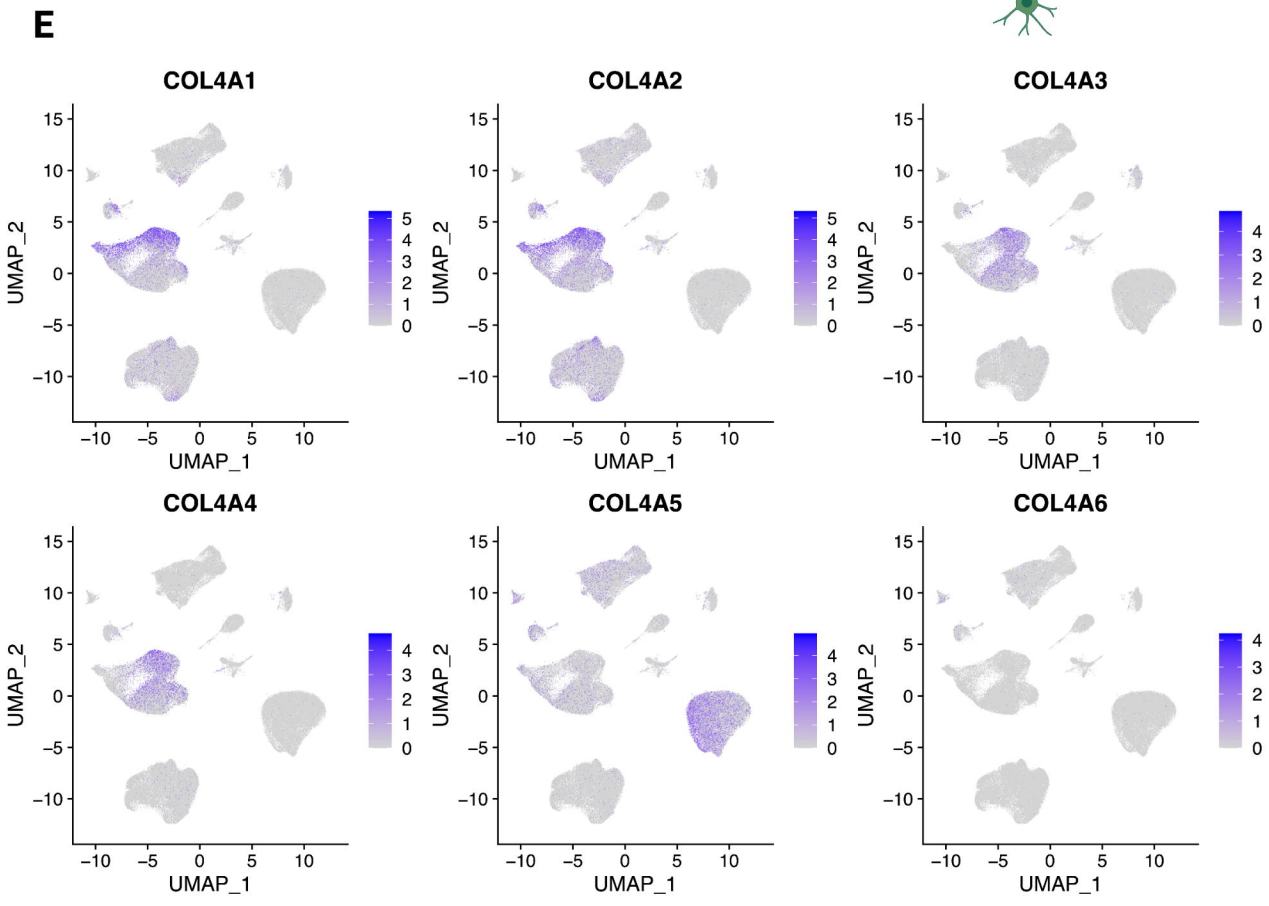


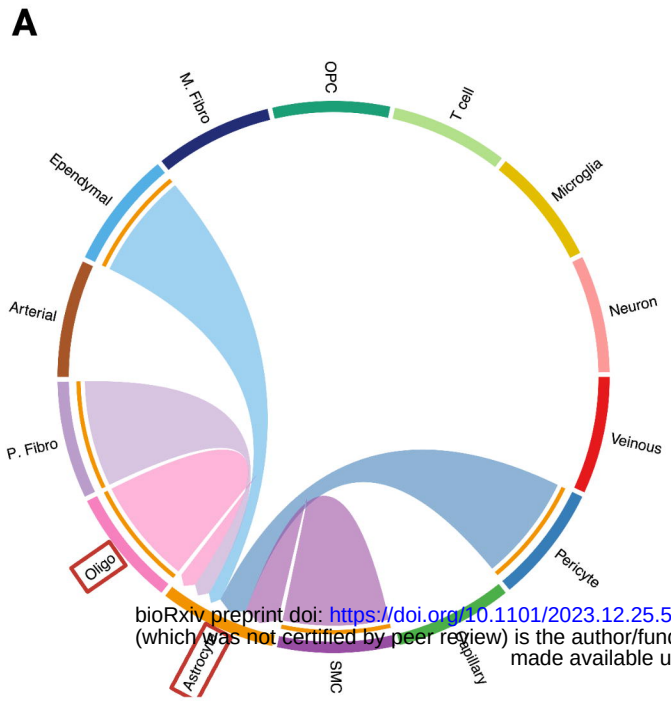


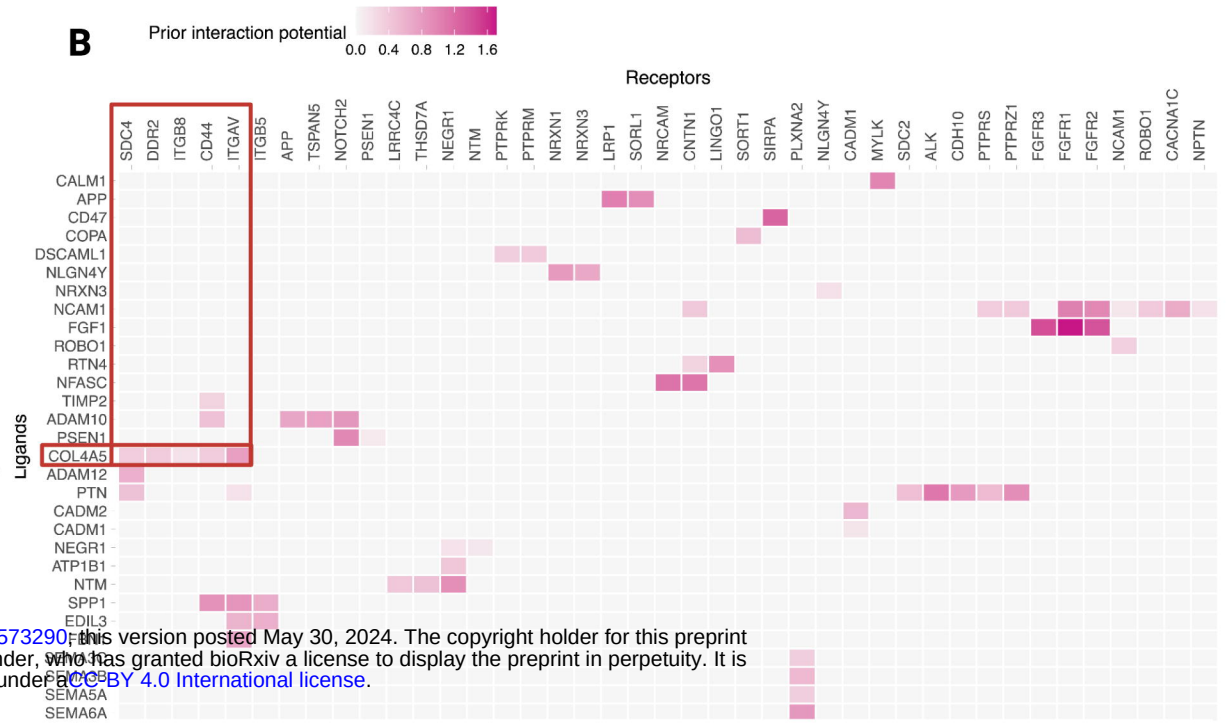


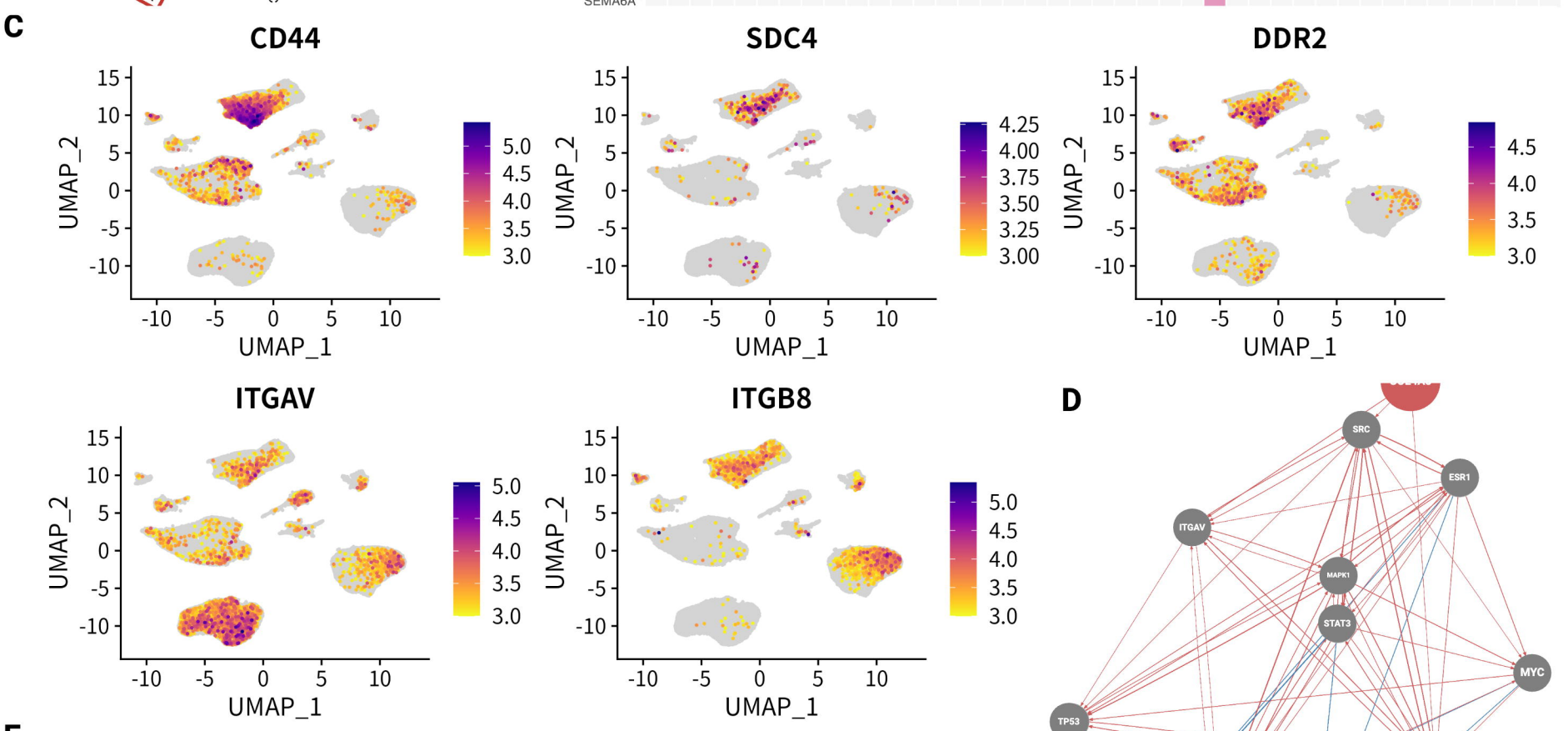


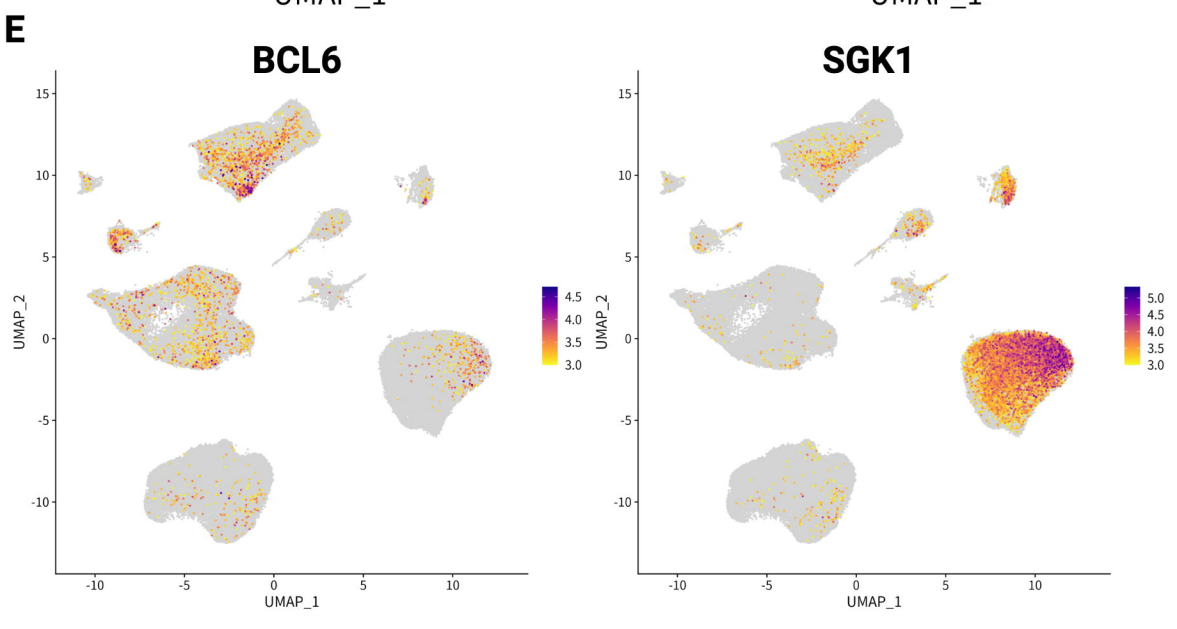


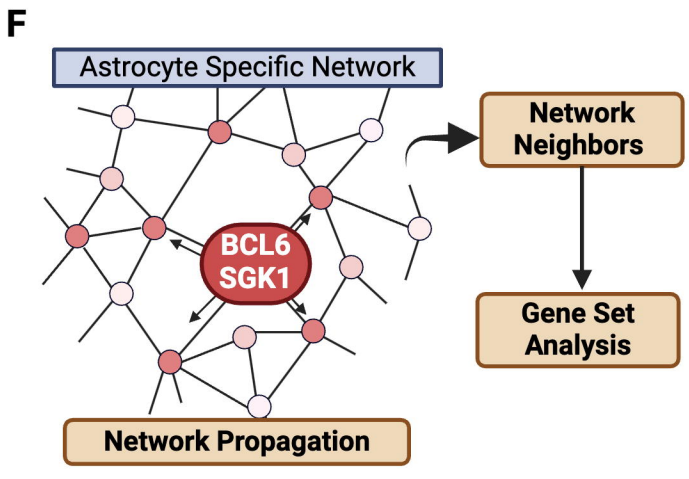


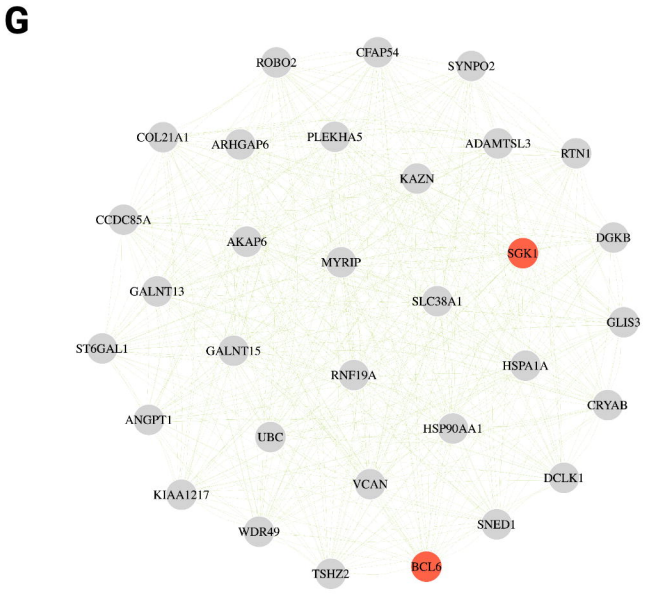


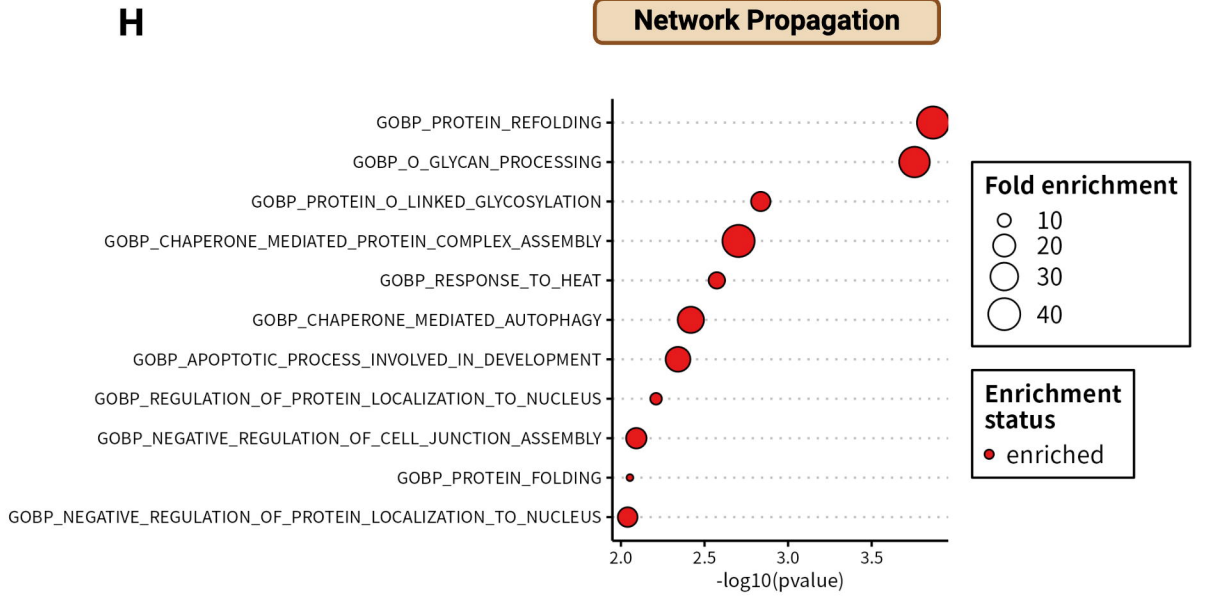


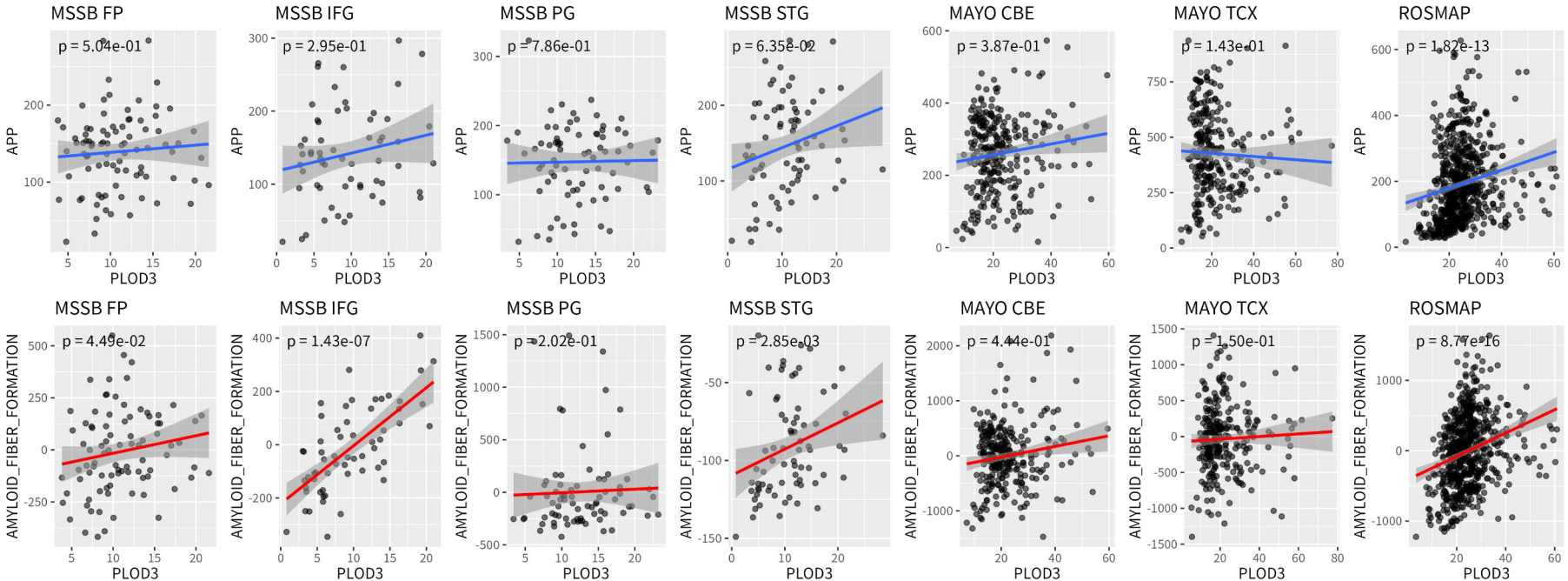




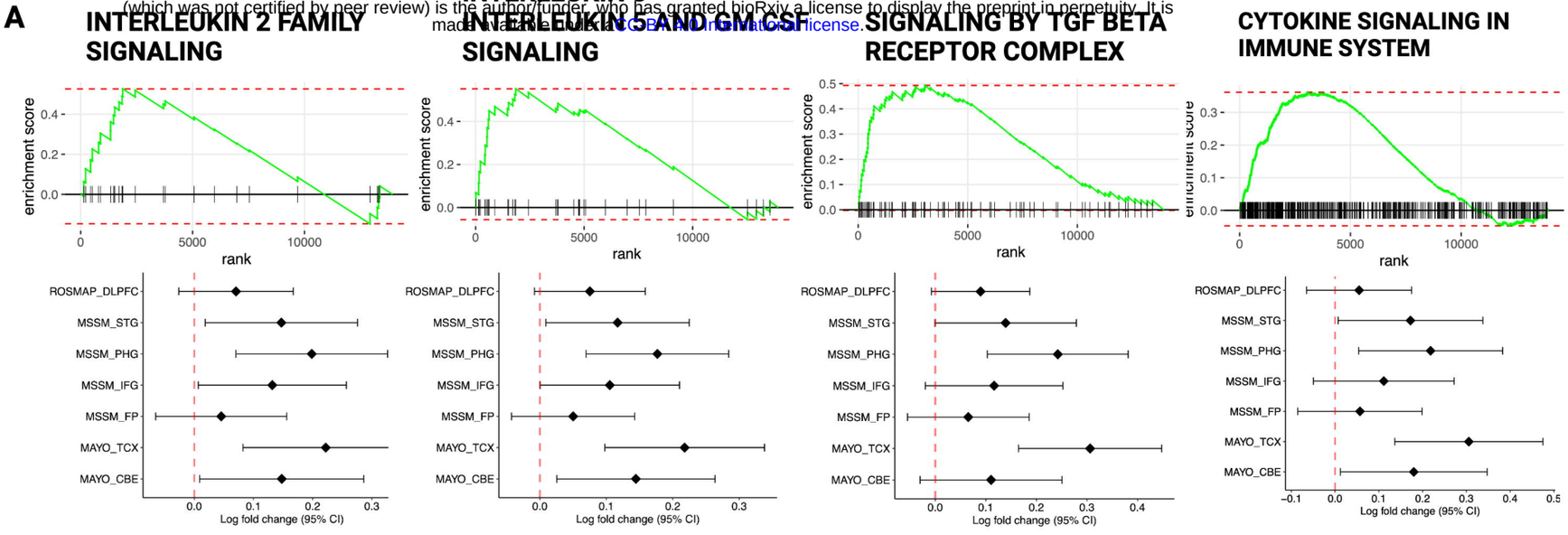




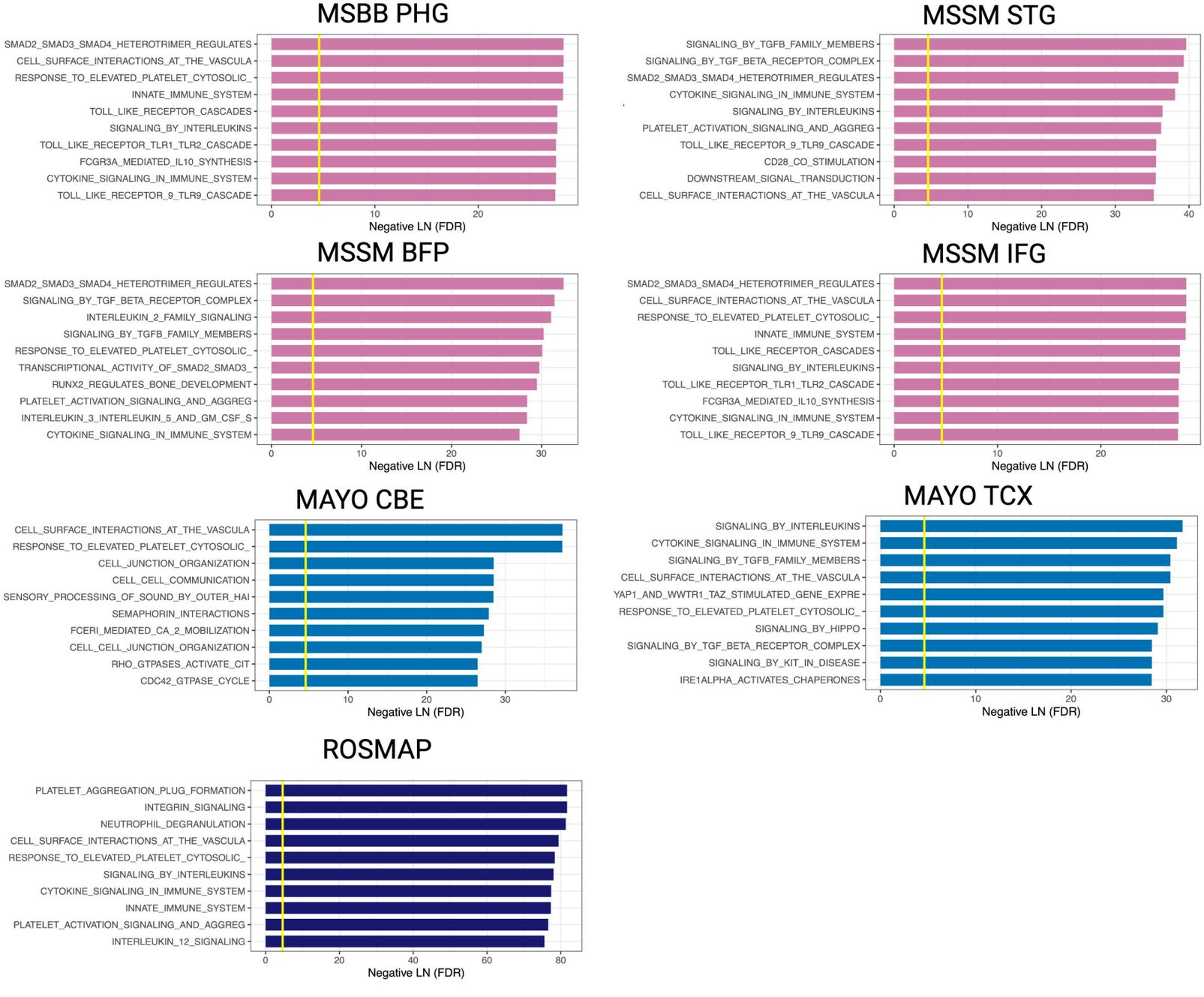


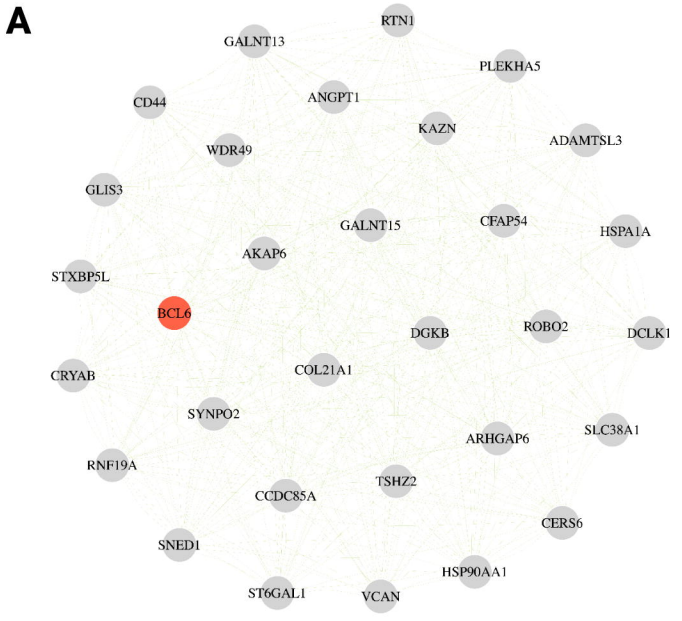

MSSB FP
p = 5.04e-01
MSSB IFG
p = 2.95e-01
MSSB PG
p = 7.86e-01
MSSB STG
p = 6.35e-02
MAYO CBE
p = 3.87e-01
MAYO TCX
p = 1.43e-01
ROSMAP
p = 1.82e-13
APP
PLOD3
MSSB FP
p = 4.49e-02
MSSB IFG
p = 1.43e-07
MSSB PG
p = 2.02e-01
MSSB STG
p = 2.85e-03
MAYO CBE
p = 4.44e-01
MAYO TCX
p = 1.50e-01
ROSMAP
p = 8.77e-16
AMYLOID_FIBER_FORMATION
PLOD3

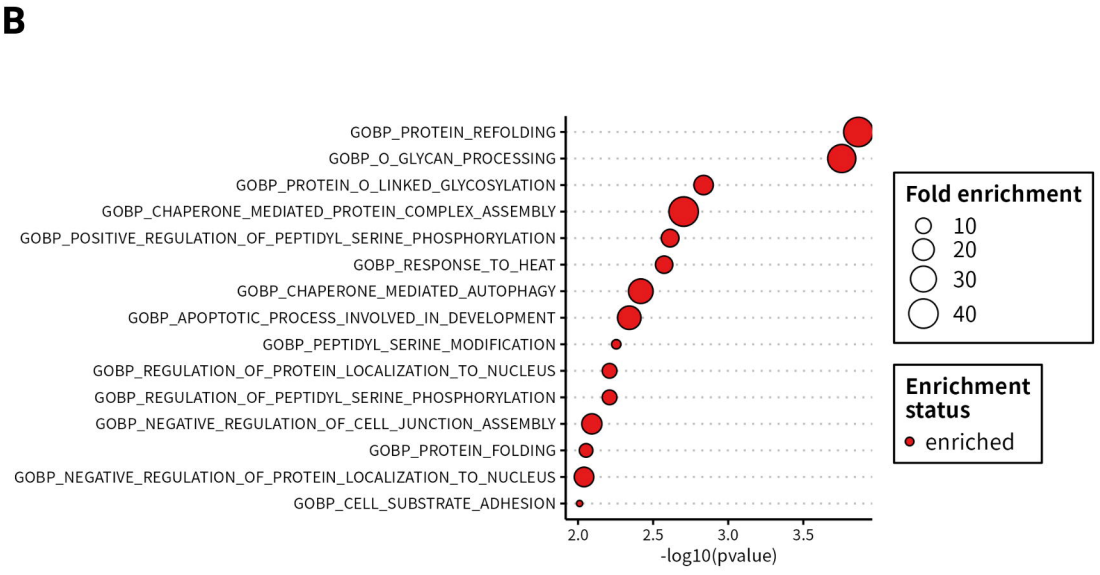
A

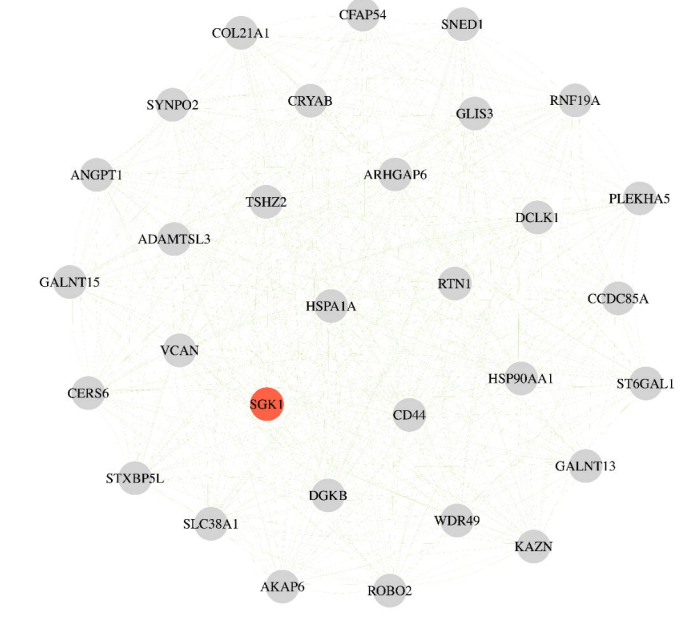


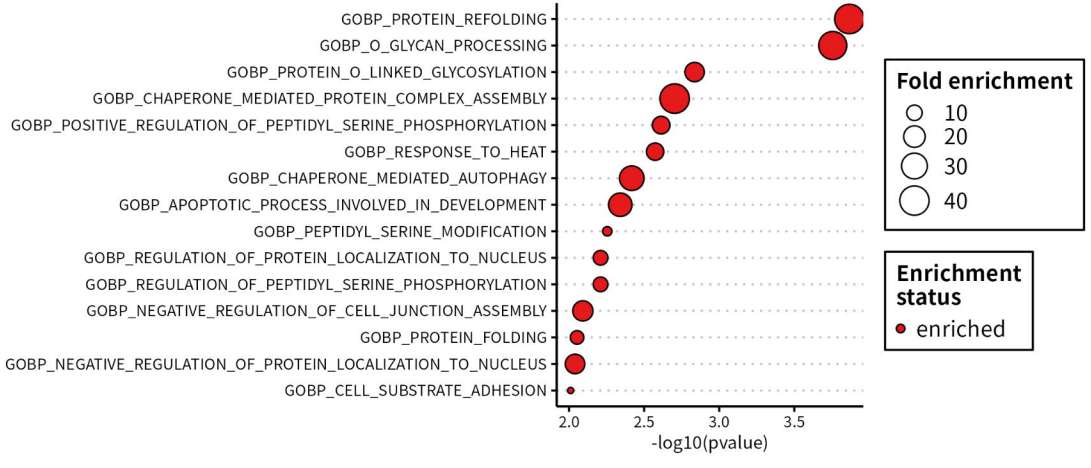
B




A
RTN1
GALNT13
PLEKHA5
CD44
ANGPT1
KAZN
ADAMTSL3
WDR49
GLIS3
GALNT15
CFAP54
HSPA1A
AKAP6
STXBP5L
BCL6
DGKB
ROBO2
DCLK1
CRYAB
COL21A1
SYNPO2
ARHGAP6
SLC38A1
RNF19A
TSHZ2
CCDC85A
CERS6
SNED1
HSP90AA1
ST6GAL1
VCAN


B
GOBP_PROTEIN_REFOLDING
GOBP_O_GLYCAN_PROCESSING
GOBP_PROTEIN_O_LINKED_GLYCOSYLATION
GOBP_CHAPERONE_MEDIATED_PROTEIN_COMPLEX_ASSEMBLY
GOBP_POSITIVE_REGULATION_OF_PEPTIDYL_SERINE_PHOSPHORYLATION
GOBP_RESPONSE_TO_HEAT
GOBP_CHAPERONE_MEDIATED_AUTOPHAGY
GOBP_APOPTOTIC_PROCESS_INVOLVED_IN_DEVELOPMENT
GOBP_PEPTIDYL_SERINE_MODIFICATION
GOBP_REGULATION_OF_PROTEIN_LOCALIZATION_TO_NUCLEUS
GOBP_REGULATION_OF_PEPTIDYL_SERINE_PHOSPHORYLATION
GOBP_NEGATIVE_REGULATION_OF_CELL_JUNCTION_ASSEMBLY
GOBP_PROTEIN_FOLDING
GOBP_NEGATIVE_REGULATION_OF_PROTEIN_LOCALIZATION_TO_NUCLEUS
GOBP_CELL_SUBSTRATE_ADHESION
2.0
2.5
3.0
3.5
-log10(pvalue)
Fold enrichment
10
20
30
40
Enrichment status
enriched


C
CFAP54
COL21A1
SNED1
SYNPO2
CRYAB
RNF19A
GLIS3
ANGPT1
ARHGAP6
TSHZ2
PLEKHA5
DCLK1
ADAMTSL3
GALNT15
RTN1
CCDC85A
HSPA1A
VCAN
HSP90AA1
ST6GAL1
CERS6
SGK1
CD44
GALNT13
STXBP5L
DGKB
SLC38A1
WDR49
KAZN
AKAP6
ROBO2


D
GOBP_PROTEIN_REFOLDING
GOBP_O_GLYCAN_PROCESSING
GOBP_PROTEIN_O_LINKED_GLYCOSYLATION
GOBP_CHAPERONE_MEDIATED_PROTEIN_COMPLEX_ASSEMBLY
GOBP_POSITIVE_REGULATION_OF_PEPTIDYL_SERINE_PHOSPHORYLATION
GOBP_RESPONSE_TO_HEAT
GOBP_CHAPERONE_MEDIATED_AUTOPHAGY
GOBP_APOPTOTIC_PROCESS_INVOLVED_IN_DEVELOPMENT
GOBP_PEPTIDYL_SERINE_MODIFICATION
GOBP_REGULATION_OF_PROTEIN_LOCALIZATION_TO_NUCLEUS
GOBP_REGULATION_OF_PEPTIDYL_SERINE_PHOSPHORYLATION
GOBP_NEGATIVE_REGULATION_OF_CELL_JUNCTION_ASSEMBLY
GOBP_PROTEIN_FOLDING
GOBP_NEGATIVE_REGULATION_OF_PROTEIN_LOCALIZATION_TO_NUCLEUS
GOBP_CELL_SUBSTRATE_ADHESION
2.0
2.5
3.0
3.5
-log10(pvalue)
Fold enrichment
10
20
30
40
Enrichment status
enriched