Supplemental materials for "Debiased personalized gene coexpression networks for population-scale scRNA-seq data"

Shan Lu[1] and Sündüz Keleş[1,2]

[1] Department of Statistics, University of Wisconsin, Madison, WI, USA.

[2] Department of Biostatistics and Medical Informatics, University of Wisconsin School of Medicine and Public Health, Madison, WI, USA.

Contents

S1 Supplemental Methods

S1.1 Implementation details of the comparison network construction methods

Since Pearson and Spearman's correlations are commonly used metrics to quantify coexpression, we considered both when computing correlation from SCTransform normalized data. The choice of SCTransform was motivated by the fact that it is by far the most widely used method for normalization as part of the `Seurat` package. Other state-of-the-art methods that we compared to have built-in coexpression estimates. For example, Noise.Reg proposed a noise regularized expression together with Spearman's correlation while SAVER used Pearson correlation for coexpression estimation.

For SCT.Pearson, SCT.Spearman, and Noise.Reg, the gene expression matrix of donor cells were normalized with `SCTransform`. coexpression matrix for SCT.Pearson (SCT.Spearman) was computed through the Pearson (Spearman's) correlation of the normalized counts in the output data slot `scale.data`. The normalized expression in slot `data` was used for Noise.Reg as it requires positive normalized expression. For Noise.Reg, a uniformly distributed random variable from the interval $[0, q_i)$, where q_i is the 1st percentile of the normalized expression of gene i, was added to the normalized expression before computing Spearman's correlation. The normalization for SAVER was performed using the `saver` function in the R package `SAVER`, and the coexpression matrix was obtained using the `cor.genes` function. Coexpression network for bigSCale2 was computed with function `compute.network`. For MetaCell, gene expression was first normalized with `Seurat`, then projected onto a low-dimensional space with UMAP. Cells were grouped into $\frac{N}{10}$ clusters, where N is the total number of cells, with the k-means algorithm using the UMAP coordinates. Expression of genes across cells in the same cluster were then aggregated into a meta cell. Coexpression network for MetaCell was constructed by Pearson correlations of meta cell expressions.

S1.2 False discovery rates for edge estimation

We designed a data-driven permutation experiment to study the false discovery rates of network edges. Given a gene expression matrix D_1, the genes are randomly split into two disjoint equal-sized

sets, S_1 and S_2. A "null" dataset D_2 is created by permuting the cell ordering of genes in S_2 while keeping the original cell ordering of genes in S_1. This permutation keeps the overall expression and sparsity levels of the genes the same as in D_1 and disassociates genes in S_2 from genes in S_1 in dataset D_2. As a result, the true correlation of gene pairs with one gene in S_1 and one gene S_2 is 0 in the permuted dataset D_2. We denote the estimated correlation matrices of gene pairs with one gene in S_1 and one gene in S_2 computed with datasets D_1 and D_2 as C_1 and C_2. C_1 and C_2 are thresholded to construct the coexpression networks using the 95th and 99th percentiles of the absolute correlations in C_1, resulting in $\widehat{C_1}$ and $\widehat{C_2}$. Then, the empirical false discovery rate is computed as

$$\text{Empirical FDR} = \min\Big(1, \frac{\#\text{ of gene pairs in } \widehat{C_2}}{\#\text{ of gene pairs in } \widehat{C_1}}\Big).$$

S1.3 PPI enrichment with the STRING database

Protein-protein interaction (PPI) enrichments of the networks were assessed using the proportion of network edges in the STRING database, i.e., the number of edges both in the coexpression network and in the STRING database divided by the number of edges in the coexpression network. A baseline enrichment proportion was computed as

$$\#\text{ of edges in STRING database } /[\#\text{ of genes } \times (\#\text{ of genes } - 1)/2].$$

S1.4 hTFtarget enrichment

For each transcription factor (TF) present in both the hTFtarget database and the gene coexpression network, we tested the enrichment of its targets from the hTFtarget database [12] among the genes with edges to the TF in the constructed coexpression network with the standard hypergeometric test. More specifically, for TF j with the number of edges k_j in the coexpression network, the number of known targets m_j in the network, the number of correctly identified targets q_j, and the total number of genes n in the network, the p-value is defined as `phyper(`$q_j - 1, m_j, n - m_j, k_j$`, lower.tail=F)` with the R function `phyper`. The hypergeometric test was performed separately for each TF and each donor. The p-values for a TF were combined with the Fisher's combined

probability test (Fisher's method) as the final p-value of enrichment for the TF.

S1.5 Differential feature analysis

We performed differential feature analysis on the basis of two types of features: gene expression and gene centrality measures in coexpression networks, including degree, pagerank, betweenness, and eigenvector centrality. Unless otherwise stated, the differential expression tests were implemented at the single cell level using the SCTransform normalized data together with `FindMarkers` function from the R package `Seurat` with the `MAST` [3] algorithm. Differential centrality tests were carried out at the donor level using the R package `Limma` [9] to accommodate small sample sizes. Donor features "Age" and "Sex" were controlled for in the differential centrality test for the dataset `Morabito_2021`. Multiplicity correction was performed with the Benjamini-Hochberg procedure [1] at a false discovery rate of 0.05 unless otherwise specified.

To investigate the potential confounding of differential network centrality with the differential expression in samples sequenced in two separate batches, we employed DESeq2 [6] to detect differential expression. Specifically, we implemented DESeq2 through the `FindMarkers` function in Seurat, specifying `test.use` as "DESeq2".

S1.6 Gene set enrichment analysis.

Gene set enrichment analysis for gene modules and differential centrality genes is conducted through package `Enrichr` [5].

S1.7 Evaluating the coexpression of Dozer-A3 module genes in immune oligodendrocytes (ODC)

To validate the findings about the Dozer-A3 module of the `Morabito_2021`, we assessed whether there was higher coexpression of genes in the Dozer-A3 module among immune ODC cells, and whether this association was more significant in the Alzheimer's group than in the Control group. To address the former, for a given donor d, we estimated the probability of a donor ODC cell expressing more than half of the genes in the Dozer-A3 module (p_d). We then counted the number

of immune ODC cells from that donor (represented by m_d) and the number of immune ODC cells expressing at least half of the genes in the Dozer-A3 module (represented by x_d). Using these estimates, we assessed the likelihood of observing x_d or more immune ODC cells expressing Dozer-A3 module genes purely by chance, with the following Binomial calculation:

$$\texttt{p-value} = \sum_{r=x_d}^{m_d} \texttt{Binomial}(m_d, r, p_d).$$

In addition, for a given donor d, the association between coexpressing Dozer-A3 module genes and being an immune ODC cell was assessed with a 2×2 contingency table. Specifically, a Fisher's exact test on the contingency table was carried out. The testing of the differences in odds ratios and negative $\log_{10}$ transformed Fisher's exact test p-values of the AD and Control diagnoses were carried out with a Wilcoxon rank sum test.

S1.8 Associating diagnosis with module eigengene expression from the scWGCNA analysis.

We used a linear mixed model (S.1) implemented through R package `lmer` to assess the association between module eigengene expression and diagnosis. In equation (S.1), y_i represents expression of eigengene of metacell i. Variables "Diagnosis", "Sex" and "Age" are fixed effects. "SampleID" was used as a random effect, because metacells in scWGCNA analysis are constructed within samples of individual donors.

$$y_i \sim \texttt{Diagnosis}_i + \texttt{Sex}_i + \texttt{Age}_i + (1|\texttt{SampleID}_i). \tag{S.1}$$

The following versions of the software were used for the analysis: `R 4.1.1`, `SCTransform 0.3.2`, `SAVER 1.1.2`, `Seurat 4.0.3`, `limma 3.48.3`, `MAST 1.18.0`, `WGCNA 1.70-3`, `bigSCale 2.0`, `cluster 2.1.4`.

S2 Supplemental results for computational experiments

S2.1 Analytical form of the gene noise ratio and correction factor under the Gamma-Poisson model

Under the Gamma-Poisson model, the noise ratio R_j and correction factor S_j, $j = 1, \cdots, G$ can be expressed as functions of the cell sequencing depths $\{\ell_i\}_{i=1}^N$ and the Gamma scale parameter u_j as:

$$R_j = \frac{1}{1 + \frac{u_j}{\frac{1}{N}\sum_{i=1}^N 1/\ell_i}}, \tag{S.2}$$

and

$$S_j = 1 + \frac{\sum_{i=1}^N 1/\ell_i}{N} \times \frac{1}{u_j}. \tag{S.3}$$

These analytical forms in equations (S.2) and (S.3) support the intuition and the empirical results that when the sequencing depth or expression is high, R_j is closer to 0 (S_j is closer to 1) and the bias of sample correlation of normalized UMI counts is expected to be small. Figure S1 displays the noise ratio as a function of gene expression and sparsity (i.e., proportion of zero counts for a gene) across both simulated and real datasets (`Jerber_2021` and `Morabito_2021`) and confirms the expected pattern of noise ratio as mean expression level and sparsity vary. We observe that for a fixed expression level (x-axis), genes with lower sparsity tend to have higher noise ratios. While this initially appears counter-intuitive, further calculations with this model support the observation. Under the Gamma-Poisson model with v_j, u_j and ℓ_i as gamma shape, scale parameters for gene j and sequencing depth for cell i, the probability of zero count for gene j is

$$\mathbb{P}(Y_{ij} = 0) = \Big(\frac{1}{1 + \ell_i u_j}\Big)^{v_j}. \tag{S.4}$$

For a fixed mean expression c_0, i.e., $u_j v_j = c_0$,

$$\mathbb{P}(Y_{ij} = 0) = \frac{1}{\Big(1 + \frac{c_0 \ell_i}{v_j}\Big)^{v_j}}. \tag{S.5}$$

The denominator of eqn. (S.5) is a monotonically increasing function of v_j for fixed $c_0\ell_i$. As a result, $\mathbb{P}(Y_{ij} = 0)$ decreases as v_j increases and, for fixed mean expression $c_0 = u_j v_j$, the sparsity level of gene j increases as u_j increases. Further utilizing eqn. (S.2), we conclude that R_j decreases as u_j increases; hence, for fixed mean expression, sparser genes have lower noise ratio, which matches the empirical observation in Figure S1.

S2.2 Bias-variance calculation of estimated correlation by Dozer

We evaluated the bias and variance of the estimated gene-gene correlations in a simulation setting with 1,000 cells and an average of 1 count per gene, mirroring the sequencing depths and sample sizes observed in the `Morabito_2021` dataset. We visualized the true vs. estimated correlation from Dozer (Figure S3) and SCT.Pearson (Figure S4, SCT.Pearson is presented since it performs closest to Dozer across all the evaluations in our manuscript) stratified by the noise ratios of each gene in a gene pair (high, medium and low). We observe that while SCT.Pearson estimates are biased across all settings (a smooth line through the data points is presented in red in all the panels), Dozer provides imperfect but significant correction across all categories, and especially for gene pairs with low and medium noise ratios. Next, we quantified the bias variance trade-off across all gene pairs stratified according to their noise ratios. Because of the bias towards zero in Figures S3, S4, we quantified the bias as the difference between the absolute estimated correlation and the absolute true correlation. We divided the genes into 10 groups based on their noise ratios, each group containing 300 genes. Next, we computed the bias (absolute estimated correlation – absolute true correlation) and standard deviation for each pair of bins by analyzing the sample of gene pairs in the strata. The resulting heatmap (Figure S5) presents our findings. As expected, gene pairs of higher noise ratios tend to have smaller absolute correlation than the true correlation and higher standard deviation.

S2.3 Simulation experiment with scDesign2

To evaluate robustness of Dozer to potential violations from the Poisson-Gamma setting, we leveraged another scRNA-seq simulation tool, scDesign2 [10], which does not solely rely on the Poisson-

Gamma setting. ScDesign2 chooses the marginal distributions of the genes adaptively among a larger set of count distributions: Poisson, Negative Binomial, Zero-inflated Poisson, and Zero-inflated Negative Binomial and employs a Copula model to generate gene-gene correlations. We used the P_FPP cells of donor "HPSI0114i-eipl_1" in Jerber_2021 dataset to estimate the simulation parameters for scDesign2. ScDesign2 parameters were set to their defaults following its vignette, with the exception of randomizing the gene ordering for marginal distributions. This was done to disentangle the relationships between gene correlations and marginal distributions. This yielded 33, 130, and 2,053 genes with Poisson, Zero-inflated Negative Binomial, and Negative Binomial as estimated distributions, respectively. Overall, in this more flexible simulation setting, Dozer performed better than the alternatives in the identification of network edges (except under the low number of cells (less than 250 cells) settings where SAVER has a slight advantage) and high centrality genes (Figures S11, S12 for AUPR and F1 scores, respectively).

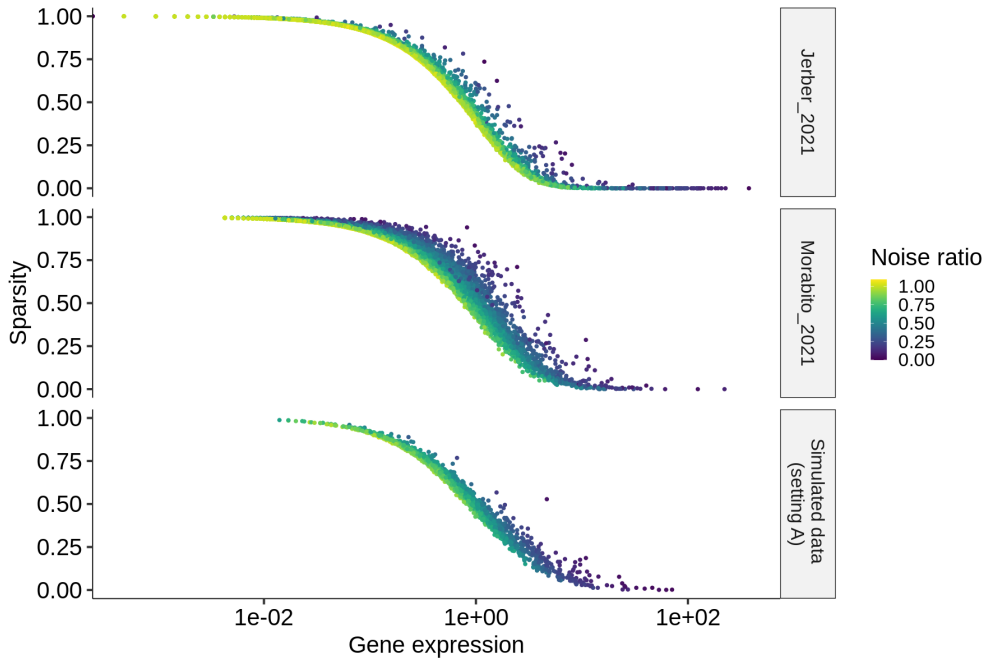

Figure S1: Noise ratio as a function of mean gene expression, sparsity level (average zero count proportion) across all genes without filtering for high noise ratio in the `Jerber_2021`, `Morabito_2021`, and simulated datasets.

Figure S2: Comparison of the analytical and estimated correction factors S_j, $j = 1, \cdots, G$ in the simulation study from the Gamma-Poisson base model. Three different estimators of the correction factors are depicted by different colors. The initial plug-in estimator (eqn. 14, green) is noisy for large values of the correction factor. The shrinkage estimator (eqn. 16, blue) steers the large scale factors down to 1. The truncated shrinkage estimator (eqn. 17, red) reduces the variation and keeps the monotonicity of the estimate.

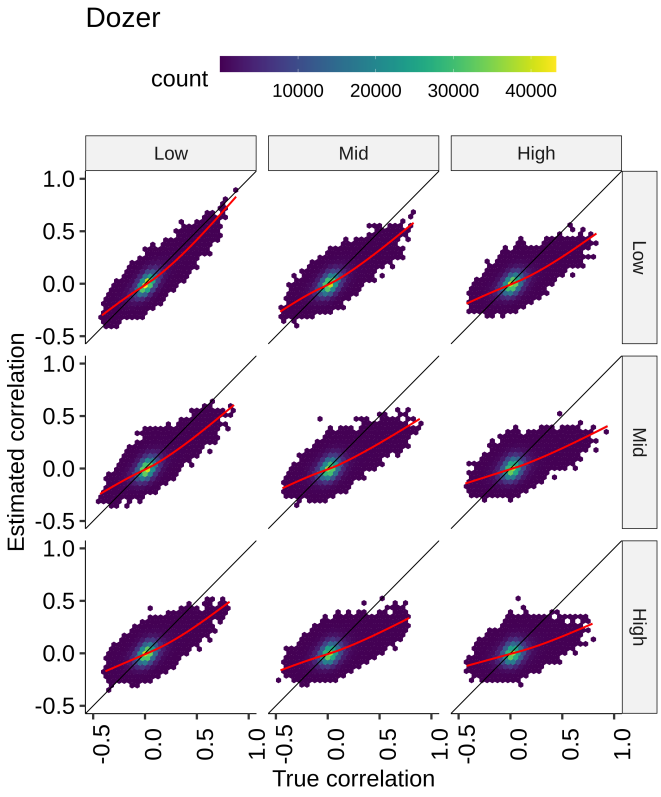

Figure S3: Scatter plot of the Dozer estimated vs. true correlations in each stratum of genes, where genes are grouped by low, medium, and high noise ratios. The distance between the 45-degree (black) line and the generalized additive model fit of the correlation estimation (red) line indicates the magnitude of bias.

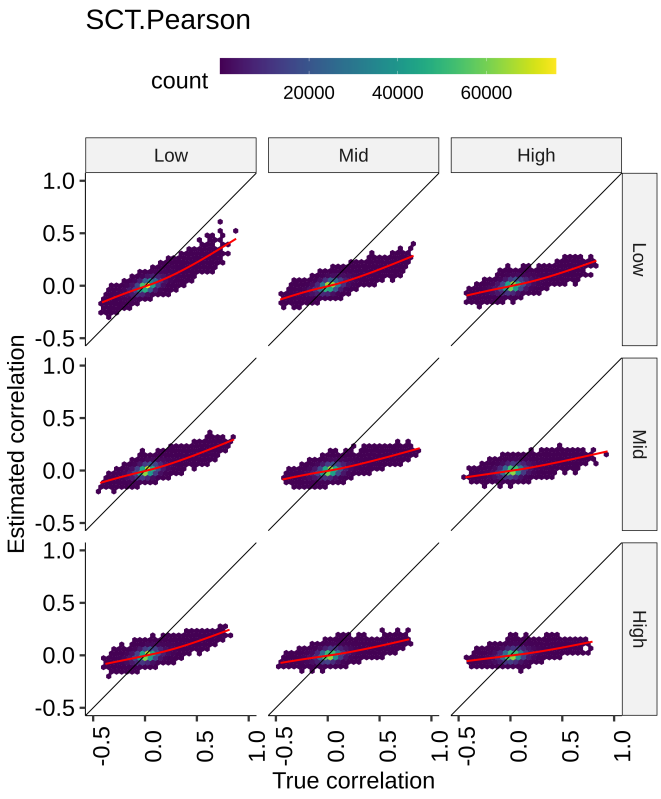

Figure S4: Scatter plot of the SCT.Pearson estimated vs. true correlations in each stratum of genes, where genes are grouped by low, medium and high noise ratio. The distance between the 45-degree (black) line and the generalized additive model fit of the correlation estimation (red) line indicates the magnitude of bias.

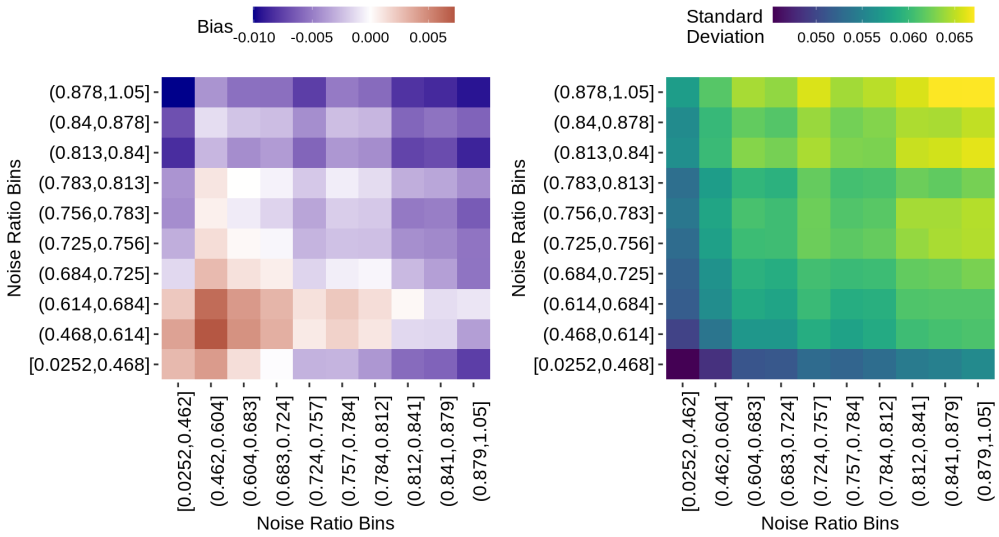

Figure S5: Bias and variance of the estimated gene pair correlations stratified by noise ratios of the gene pairs. The left heatmap presents the empirical estimate of the correlation bias for gene pairs, determined by the difference between their absolute estimated correlation and absolute true correlation, and organized according to their respective noise ratio bins on the x- and y-axes. The right heatmap presents the standard error of the absolute estimated correlation for gene pairs based on their noise ratio bins on the x- and y-axes.

Figure S6: Oversmoothing resulting from the Spearman's correlation computed with the SCTransform normalized counts of P_FPP cells in the `Jerber_2021` dataset. Genes *SYNC* and *CEP350* both have sparse UMI counts with zero expression in 72% and 80% of the cells, respectively. Left and right panels display the scatter plots of SCTranform normalized expression and the corresponding ranks of the normalized expression of the two genes, respectively. The Pearson and Spearman's correlations of the two genes are -0.03 and 0.46, respectively. Cells with zero expression for both genes are depicted in blue and lead to positive Spearman's correlation for these two genes. This is in contradiction with the Chi-square test which tests the independence of zero or positive expression between the two genes and results in a p-value of 0.49.

Figure S7: Summary of simulation setting A results in terms of AUPR scores for high centrality gene identification. The average AUPR scores for the identification of genes with top gene centralities as a function of gene noise ratios, cell sample sizes, and average sequencing depths. Gene centrality measures include pagerank, betweenness, and eigenvector centrality. The leftmost panel depicts the average AUPR score of Dozer. The other panels highlight the performances of other methods as quantified by the ratio of their AUPR scores over the AUPR score of Dozer.

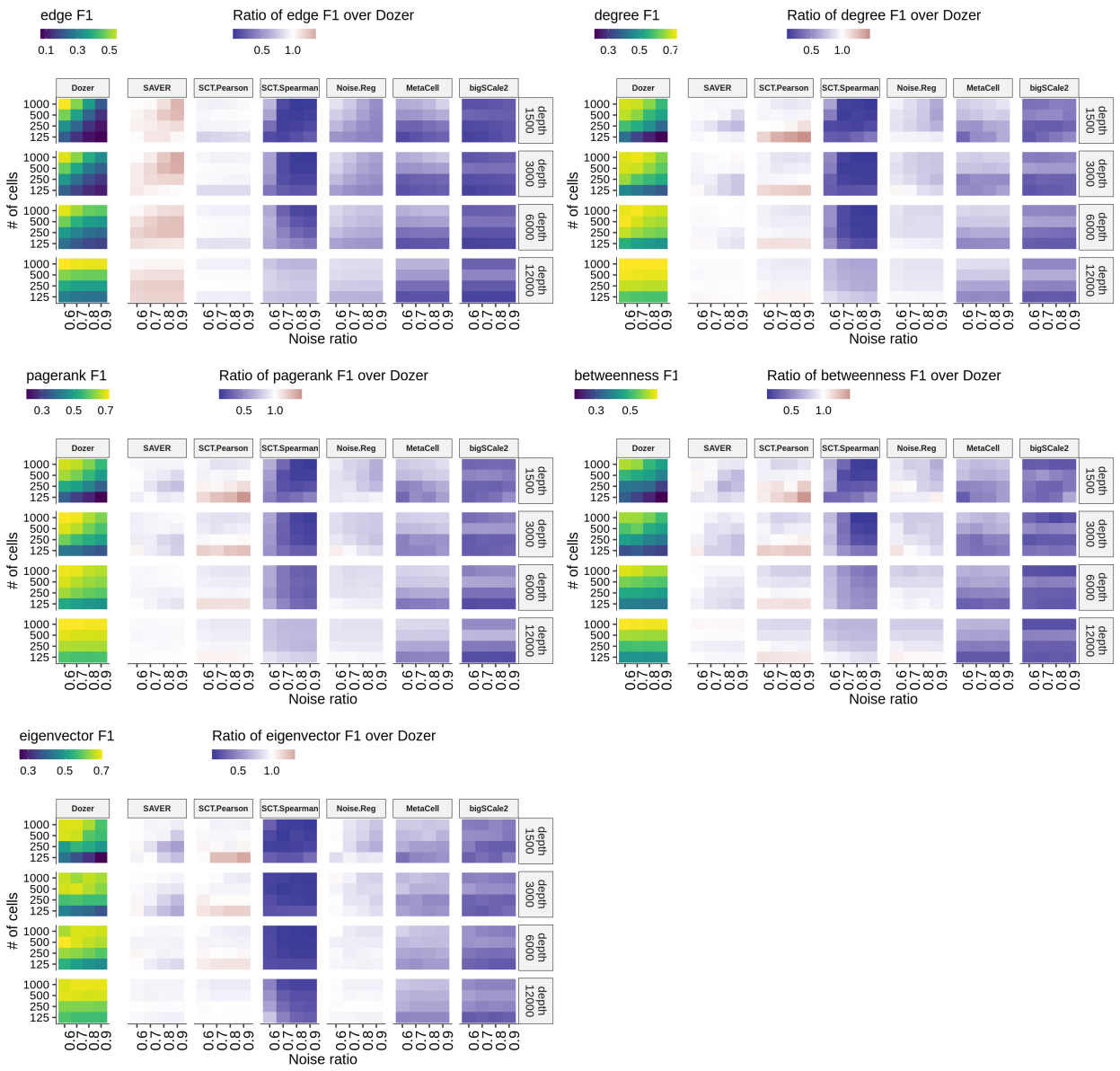

Figure S8: Summary of simulation setting A results in terms of F1 scores for edge and high centrality gene identification. The average F1 scores for edge and high centrality gene identification as a function of gene noise ratios, cell sample sizes and average sequencing depths. Gene centrality measures include degree, pagerank, betweenness, and eigenvector centralities. The left panel shows the average F1 score of Dozer. The other panels highlight the performances of other methods as quantified by the ratio of their F1 scores over the F1 score of Dozer.

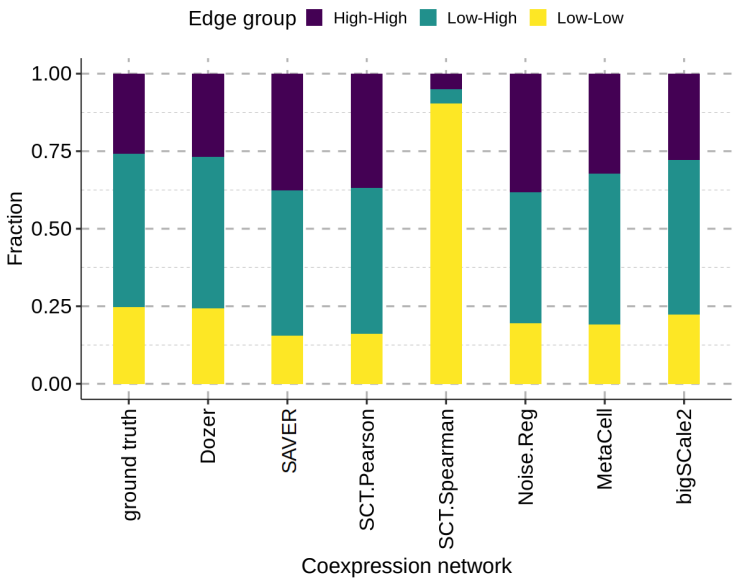

Figure S9: Stratification of network edges inferred by each method with respect to expression groups of the gene pairs: High-High: both genes in high expression group; Low-Low: both genes in low expression group; High-Low: one gene in high and the other in the low expression group. Y-axis depicts the fraction of each method's inferred network edges in each expression group. High (low) expression group is defined as the set of genes with expression larger (smaller) than the median expression among all genes. Data is generated from simulation setting A with true network edges corresponding to the gene pairs with absolute correlation exceeding the 95% percentile in the true correlation matrix in magnitude.

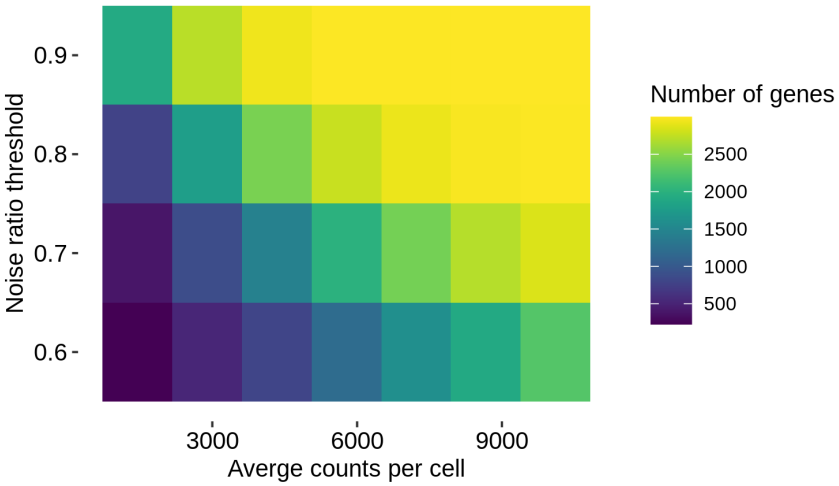

Figure S10: Quantification of the numbers of genes with varying levels of noise ratios as a function of average sequencing depth per cell. Colored entries depict the numbers of genes with noise ratios smaller than the y-axis value. The quantification is conducted on data generated from simulation setting A, which involved a total of 3,000 genes.

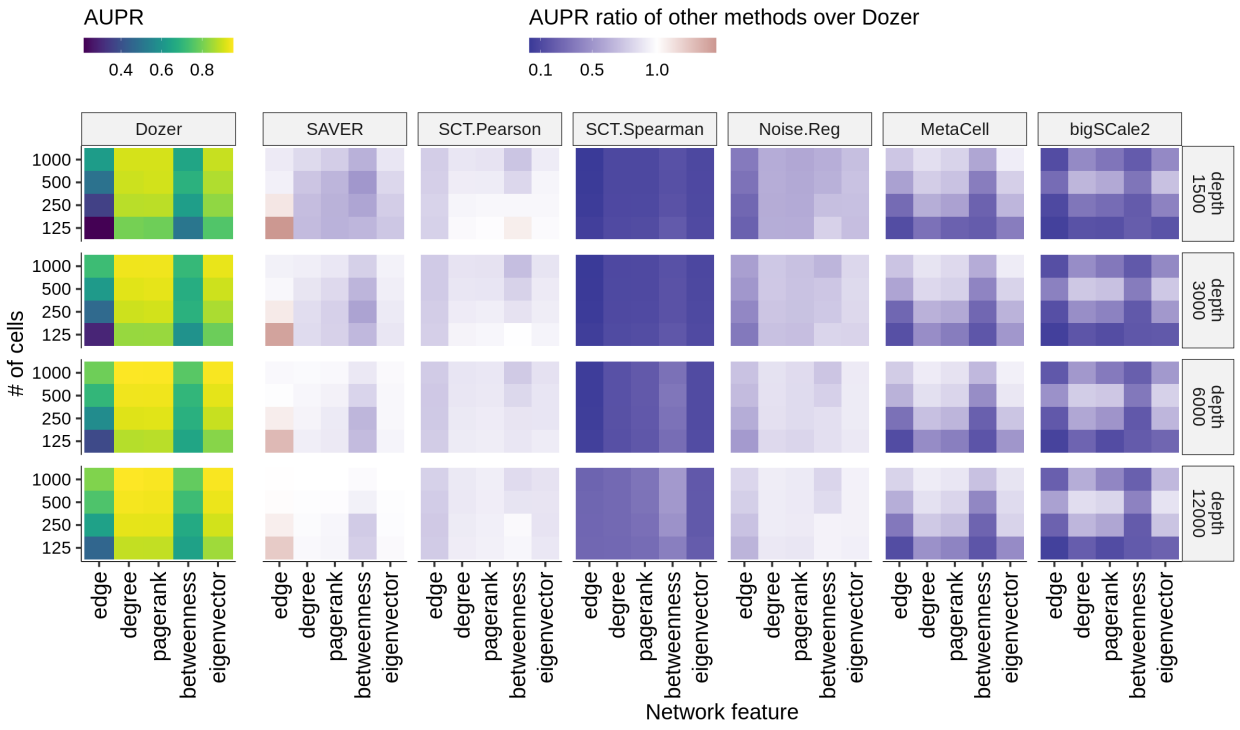

Figure S11: Summary of the scDesign2 simulations in terms of AUPR scores for edge and high centrality gene identification. The average AUPR scores for edge and high centrality gene identification as a function of cell sample sizes and average sequencing depths. Gene centrality measures include degree, pagerank, betweenness, and eigenvector centralities. The leftmost panel shows the average AUPR score of Dozer. The other panels highlight the performances of other methods as quantified by the ratio of their AUPR scores over the AUPR score of Dozer.

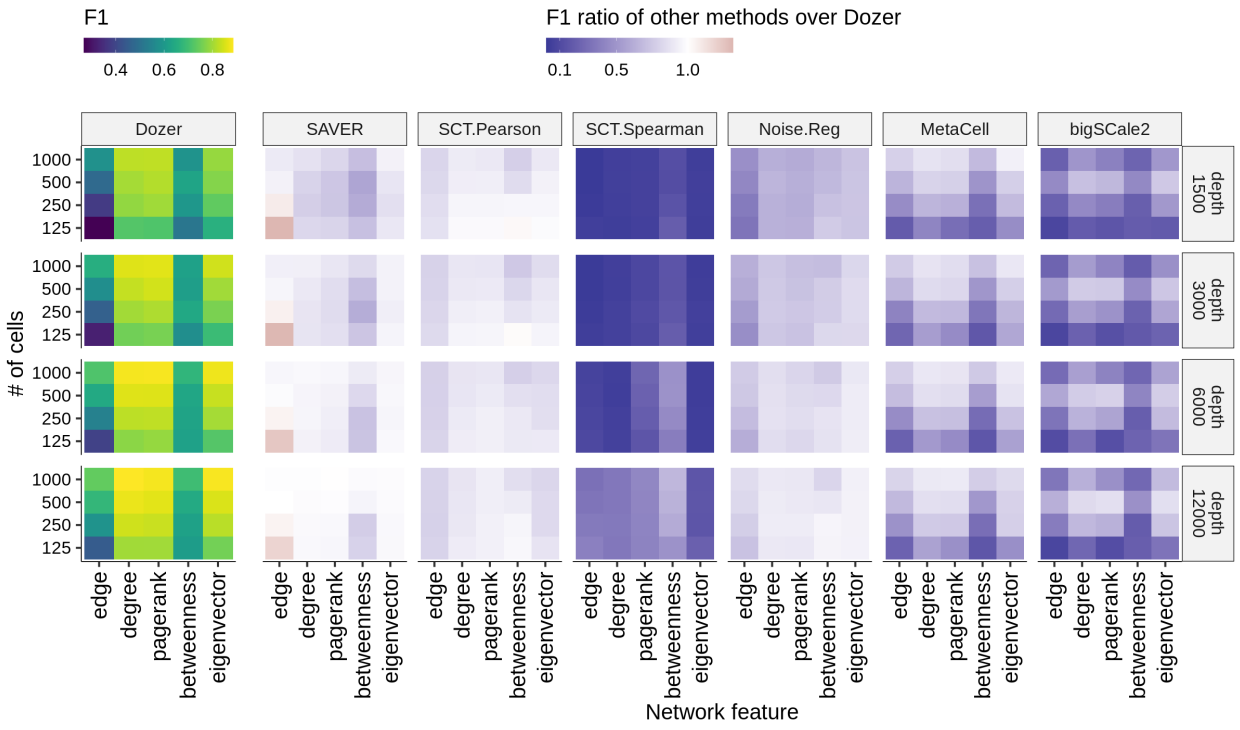

Figure S12: Summary of the scDesign2 simulations in terms of F1 scores for edge and high centrality gene identification. The average F1 scores for edge and high centrality gene identification as a function of cell sample sizes and average sequencing depths. Gene centrality measures include degree, pagerank, betweenness, and eigenvector centralities. The leftmost panel shows the average F1 score of Dozer. The other panels highlight the performances of other methods as quantified by the ratio of their F1 scores over the F1 score of Dozer.

S3 Supplemental results for the analysis of the `Jerber_2021` dataset

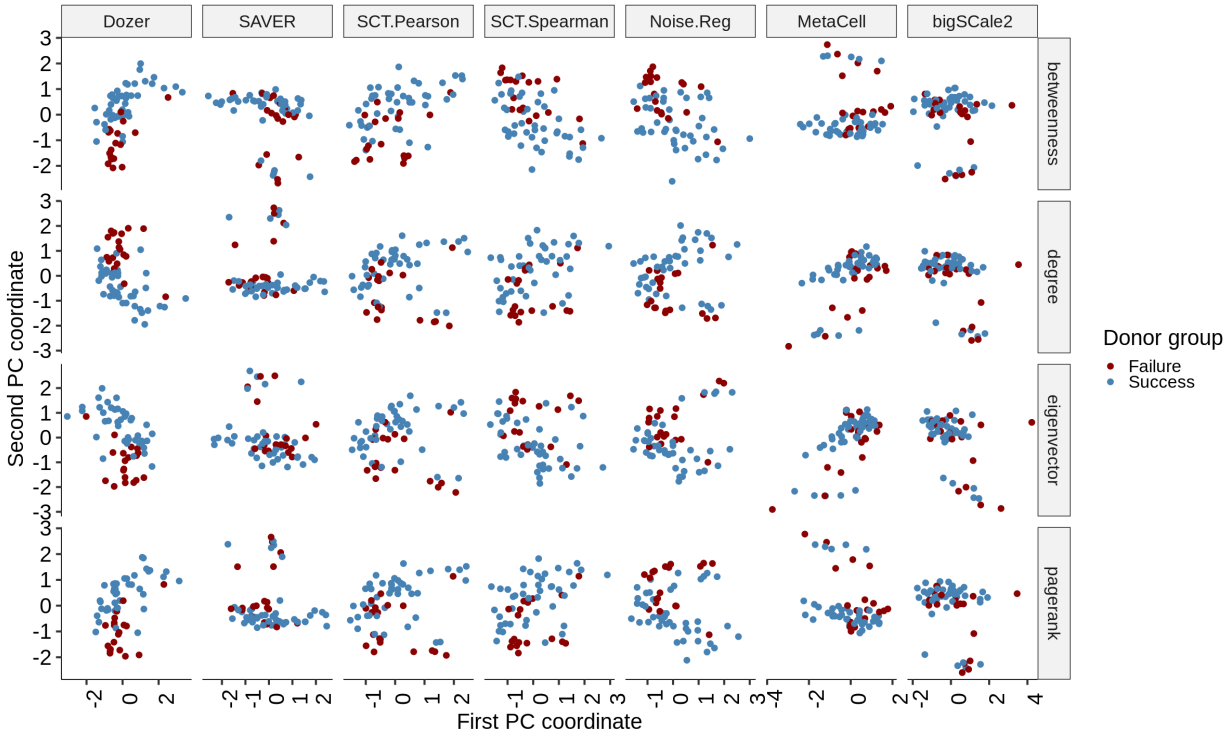

Figure S13: Two-dimensional projection of the `Jerber_2021` donor networks based on their gene centrality measures with principal components (PC). Color of the data points represents phenotypic groups, i.e., failure and success in neuronal differentiation.

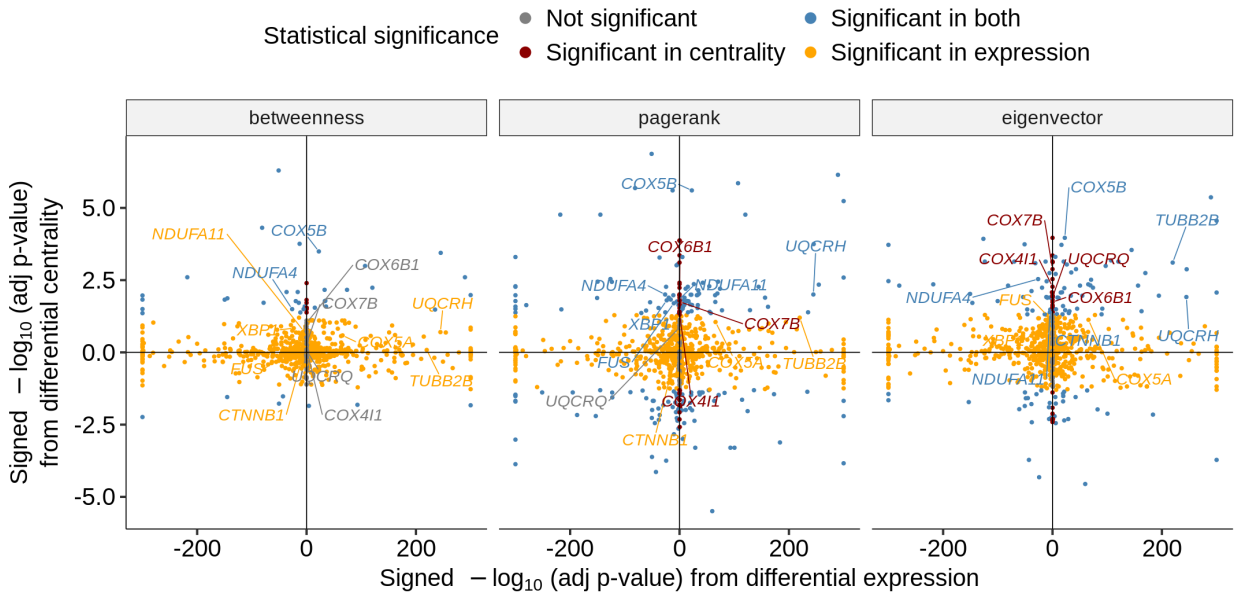

Figure S14: Comparison of differential centralities of the genes from the Dozer coexpression networks with their differential expression. Centrality measures include betweenness, pagerank and eigenvector centrality. X-axis displays the signed -$\log_{10}$(adjusted p-value) from differential expression with positive (negative) values denoting higher (lower) expression in the Failure group. Y-axis denotes the signed -$\log_{10}$(adjusted p-value) from differential degree centrality with positive (negative) values exhibiting higher (lower) centrality in the Failure group. The differential degree genes enriched in the KEGG term "Pathways of neurodegeneration" (Figure S15) are labeled in the figure, demonstrating the consistency among various centrality measures.

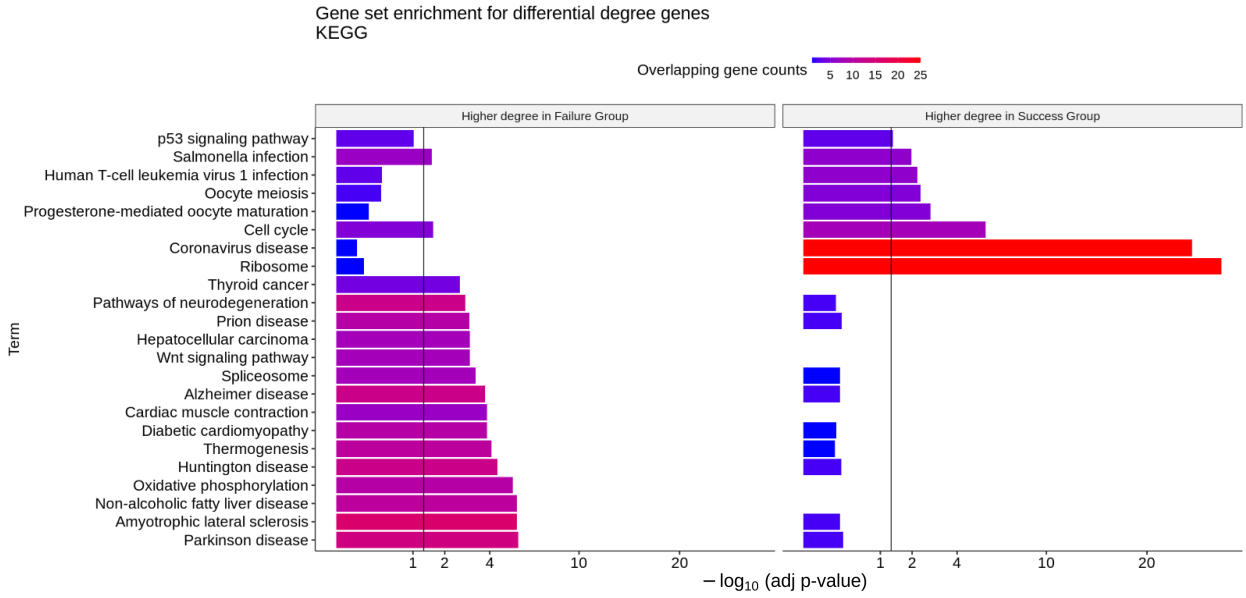

Figure S15: Gene set enrichment analysis of the Dozer identified differential degree genes with KEGG pathways.

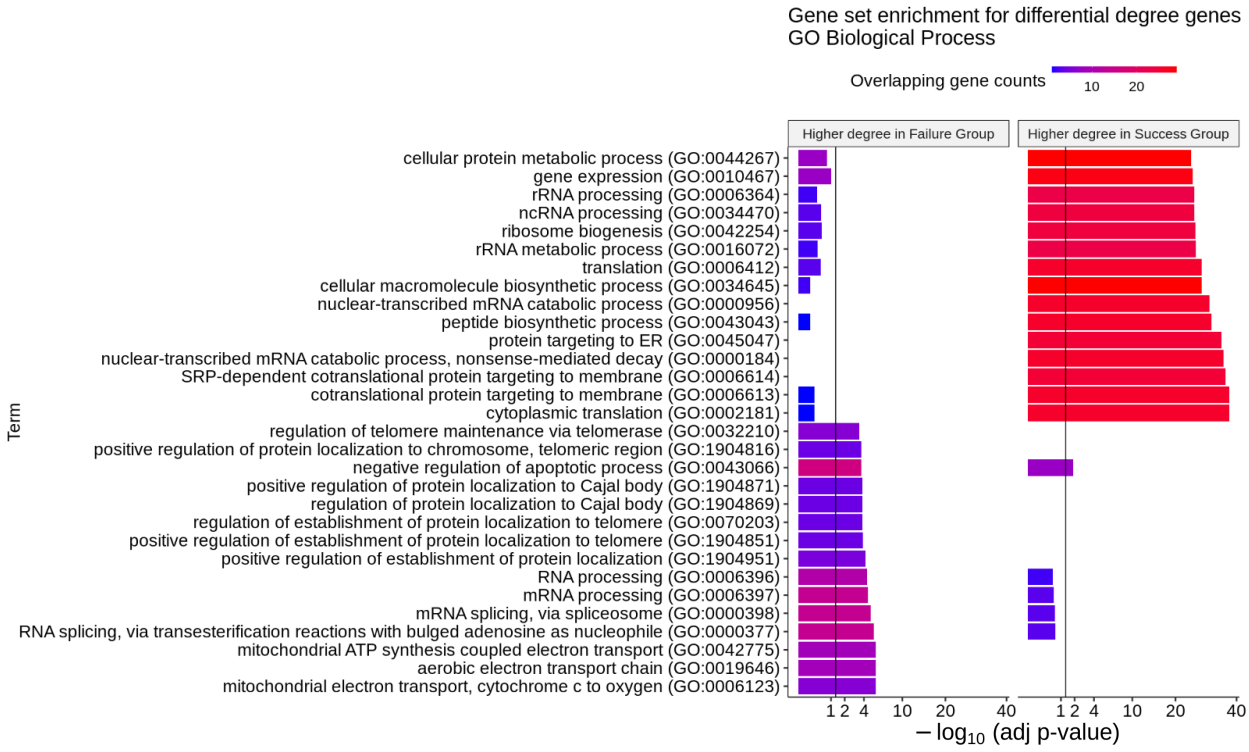

Figure S16: Gene set enrichment analysis of the Dozer identified differential degree genes with GO Biological Processes.

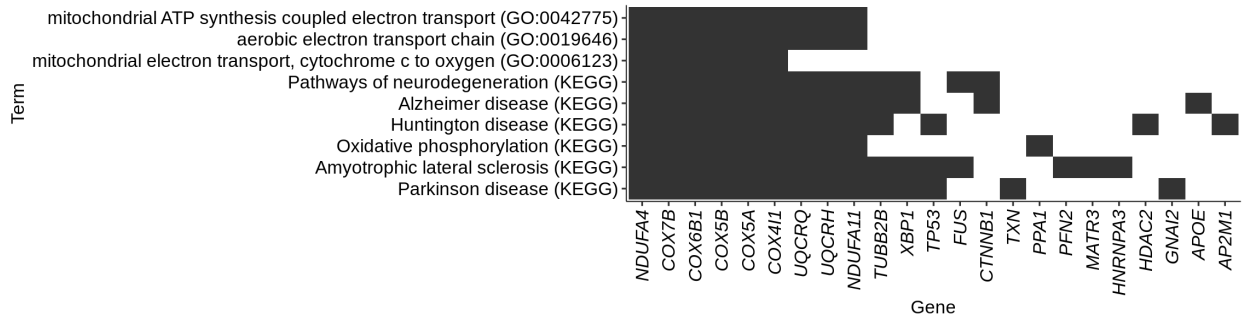

Figure S17: Membership of differential degree genes in KEGG and GO Biological Processes terms (Figures S15 and S16). Dark shade indicates that the gene is a member of the relevant KEGG or GO Biological Processes gene set.

S4 Supplemental results for the analysis of the `Morabito_2021` dataset

S4.1 Validating the largest cliques in coexpression networks

To validate the personalized coexpression networks obtained for the AD and Control donors with data from independent cohorts, we carried out a largest clique analysis. A clique (fully connected sub-network) in the coexpression network represents a group of genes with high coexpression. We computed the largest clique (largest fully connected sub-network) for each donor's network and assessed the consistency of this network structure among donors. The number of genes in largest cliques ranges from 9 to 42. Cliques from different donors have a large proportion of overlapping genes for methods Dozer, SAVER, SCT.Pearson and SCT.Spearman, with a mean Jaccard Index of 0.27 to 0.31 (Figure S19). Next, for each method, we generated a *clique-set* by pooling genes present in at least two donor largest-cliques. Clique-sets from Dozer, SAVER, SCT.Pearson, and SCT.Spearman have higher proportion of gene pairs validated by the STRING PPI database compared to other methods, with an average validation proportion of 0.12-0.17 (Figure S19), compared to a baseline of 0.027 among all gene pairs in the network. Clique-set from Dozer network also has high connectivity in an independent single cell RNA-seq of oligodendrocytes from [8], where the average number of edges in the Dozer clique-set is 12 times (22 times) larger than the average connectivity of the whole network for control (diagnosis) donors. Overall, the analysis of largest cliques in donor-specific networks supports that Dozer, SAVER, SCT.Pearson, and SCT.Spearman can consistently identify highly coexpressed gene sets in line with protein-protein interactions annotated in the STRING PPI database.

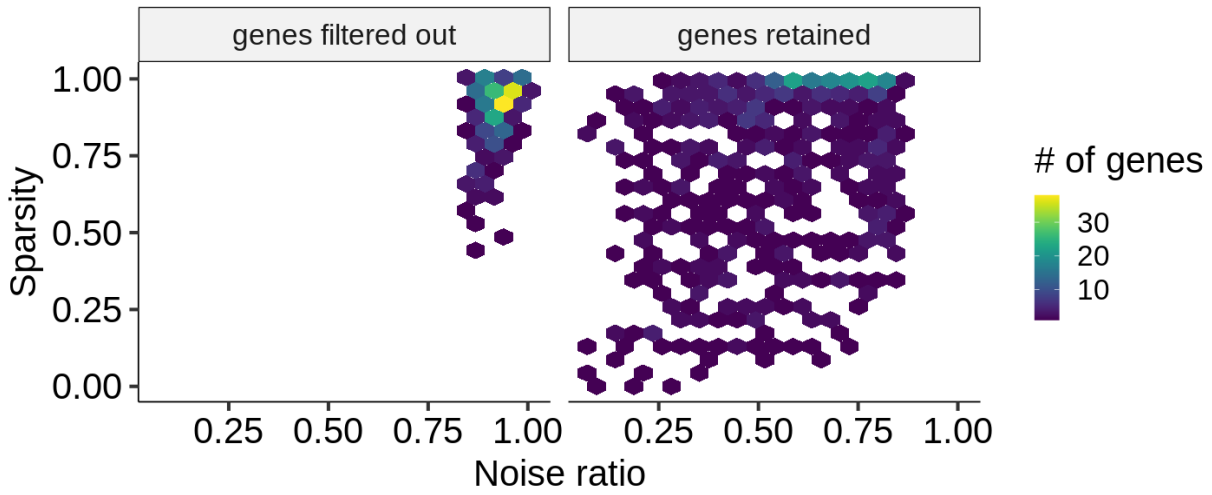

Figure S18: In the `Morabito_2021` dataset, 682 out of 1,252 genes were filtered out due to high sparsity and/or high noise ratio. The hexbin plots display the distribution of sparsity levels (proportion of zero counts) and noise ratios among the genes filtered out (left) and the genes retained (right).

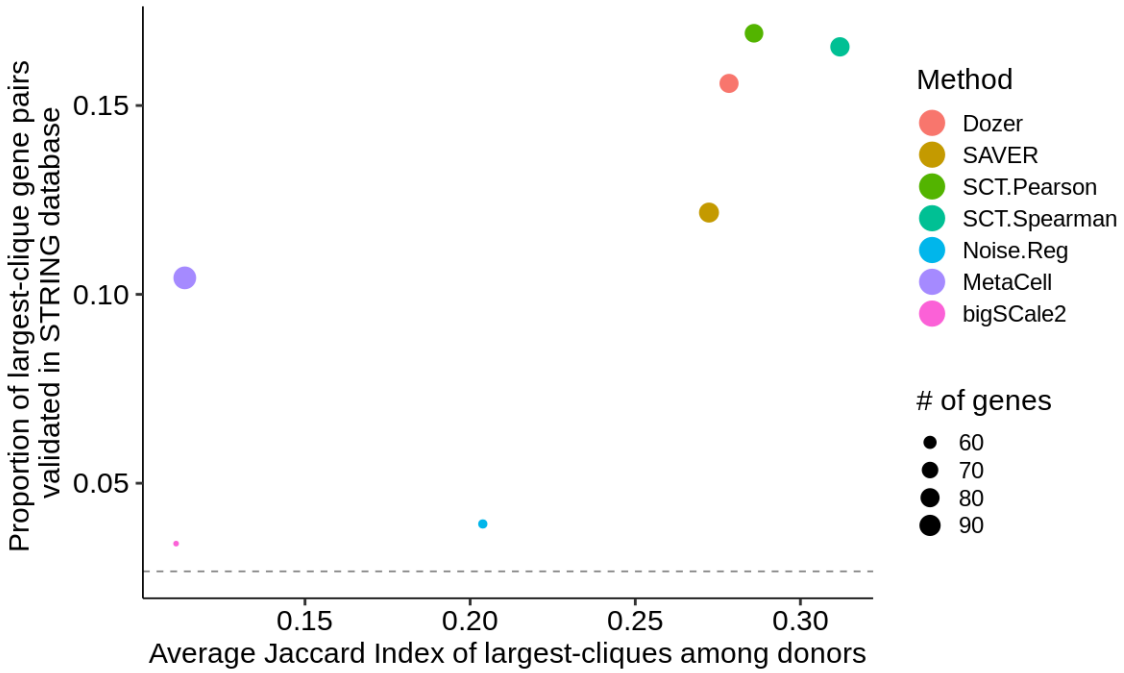

Figure S19: A largest-clique is computed from each donor for each network construction method. The reproducibility of largest-cliques among donors is evaluated using the Jaccard Index (JI) between each pair of donors. X-axis depicts the average Jaccard Index among all donor pairs. The largest-cliques were also evaluated using the STRING database by generating a clique-set across donors for each method. Y-axis depicts the proportion of gene pairs in the clique-set and are validated by the STRING database. The horizontal dashed line at 0.027 represents the STRING database validation rate for a random set of gene-pairs. The size of data points represents the number of genes in the clique-sets, ranging from 51 to 98.

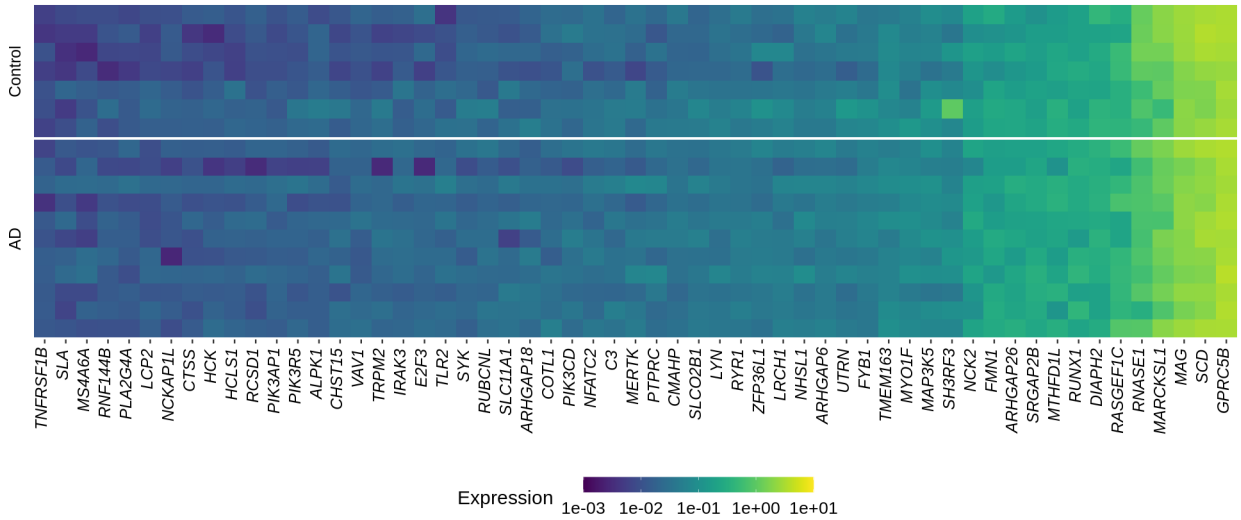

Figure S20: Heatmap of the average expression (normalized expression via SCTransform) of genes in module Dozer-A3 over cells in each donor. Rows and columns in the heatmap correspond to donors and genes, respectively.

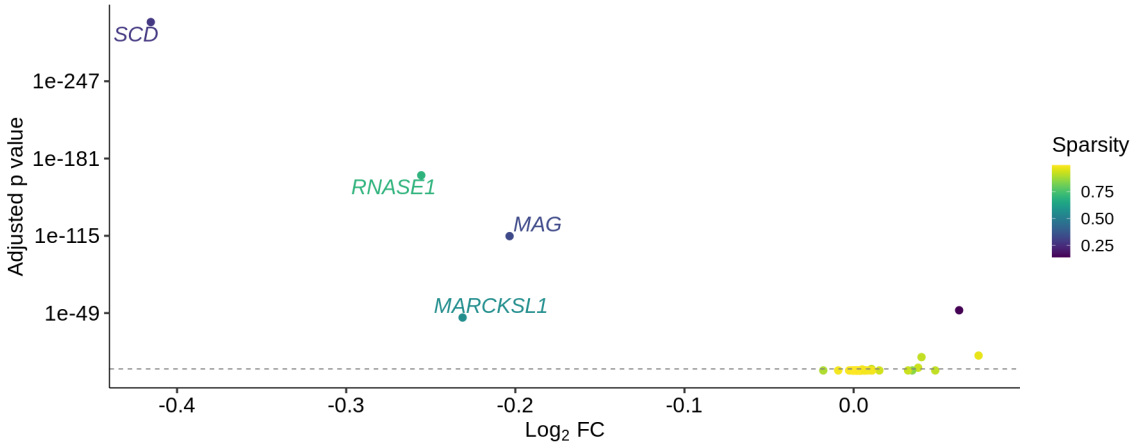

Figure S21: Volcano plot from the differential expression analysis of genes in module Dozer-A3. X-axis represents the $\log_2$ fold change of expression in AD versus Control. Y-axis represents the adjusted p-value from the differential expression test. A threshold of 0.05 for statistical significance is indicated by a dashed line. The sparsity of the genes is indicated by the color of the data points.

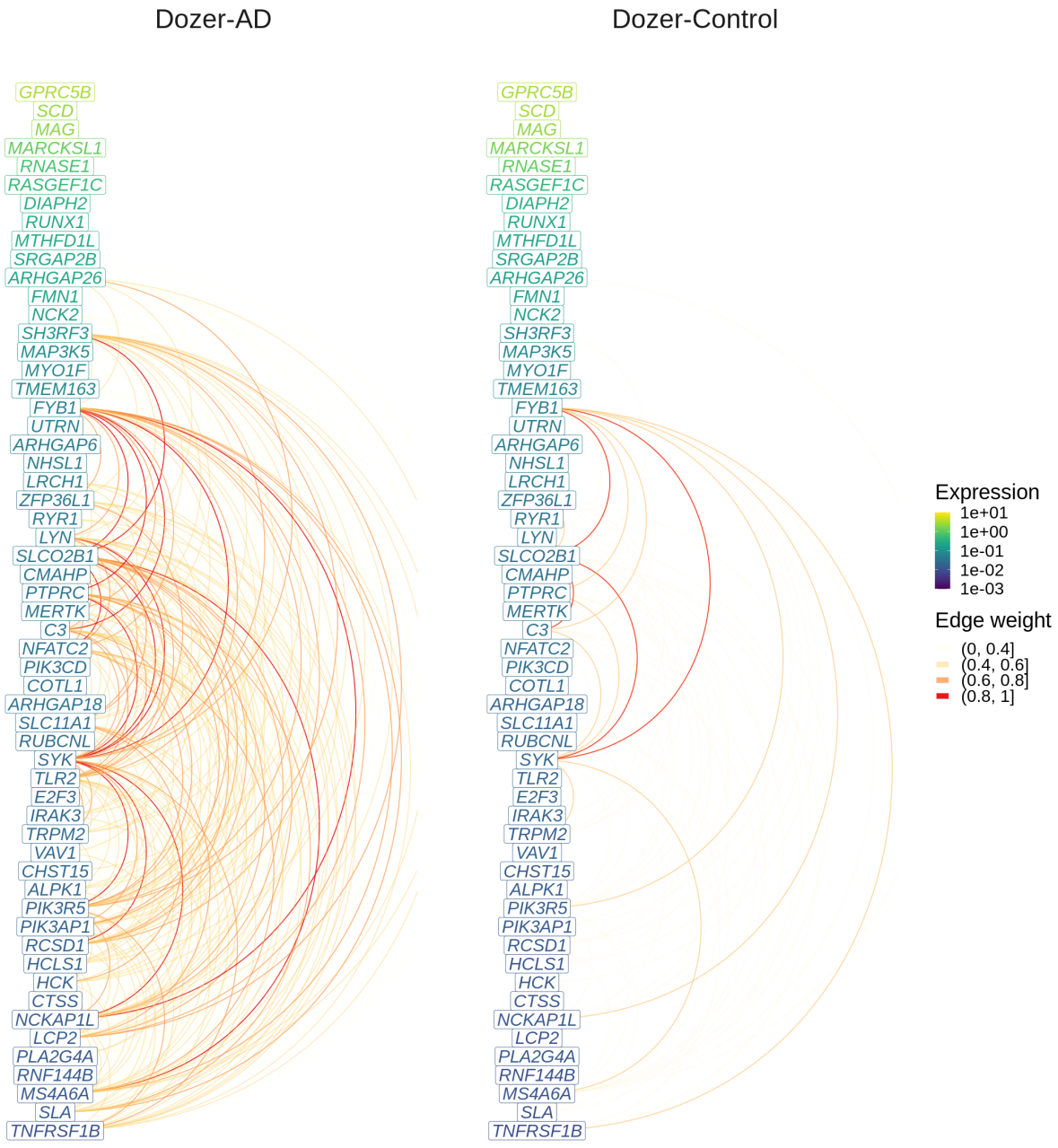

Figure S22: Hive plot visualization of module Dozer-A3 in AD and Control groups. Genes are ordered from high (top) to low (bottom) by their average expression across all donors, and colored by the average expression in the corresponding diagnosis group. The arcs between the genes in this linear layout depict the edges in the average Dozer networks of AD (left) and Control (right) donors, with colors representing the edge weights.

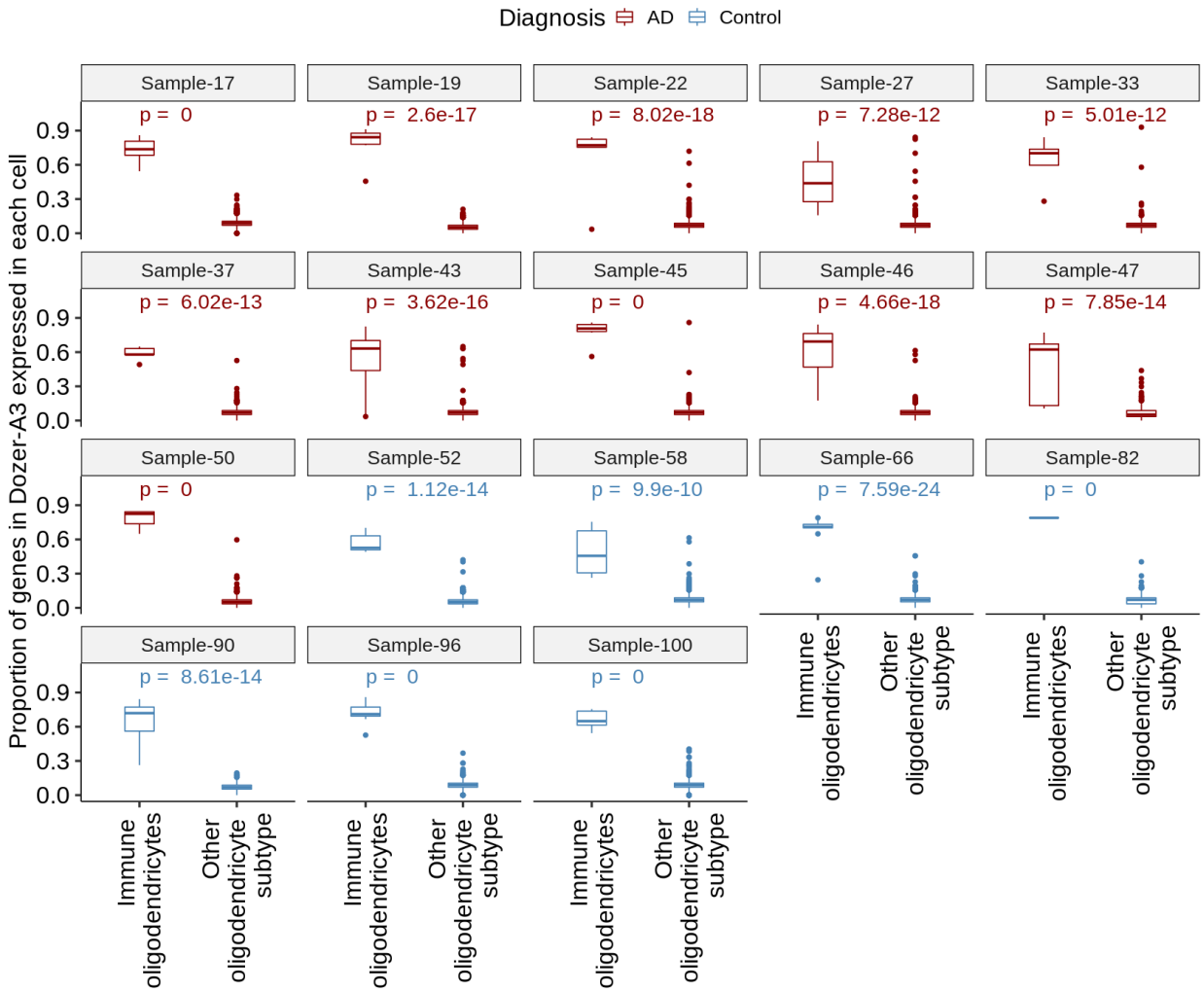

Figure S23: Proportion of the Dozer-A3 genes expressed in each cell at the donor level. Individual cells in each of the two cell populations, immune oligodendrocytes (ODC) and other ODC subtypes, are evaluated with respect to the proportion of expressed Dozer-A3 genes at the donor level. The reported p-values for each donor evaluate whether the coexpression of module Dozer-A3 in donor immune oligodendrocytes is higher than expected by chance (details in Section S1.7).

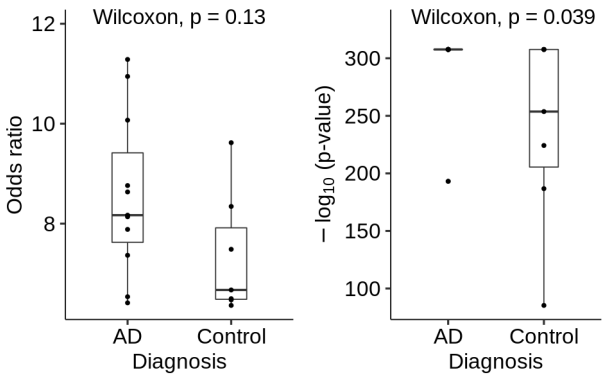

Figure S24: Boxplots of the odds ratios and -$\log_{10}$(p-values) from Fisher's exact tests that evaluate the association between coexpressing Dozer-A3 module genes and being an immune ODC cell for each donor (details in Section S1.7). Wilcoxon rank sum tests were conducted to evaluate the differences in odds ratios and -$\log_{10}$(p-values) between the two diagnosis groups (AD vs. Control).

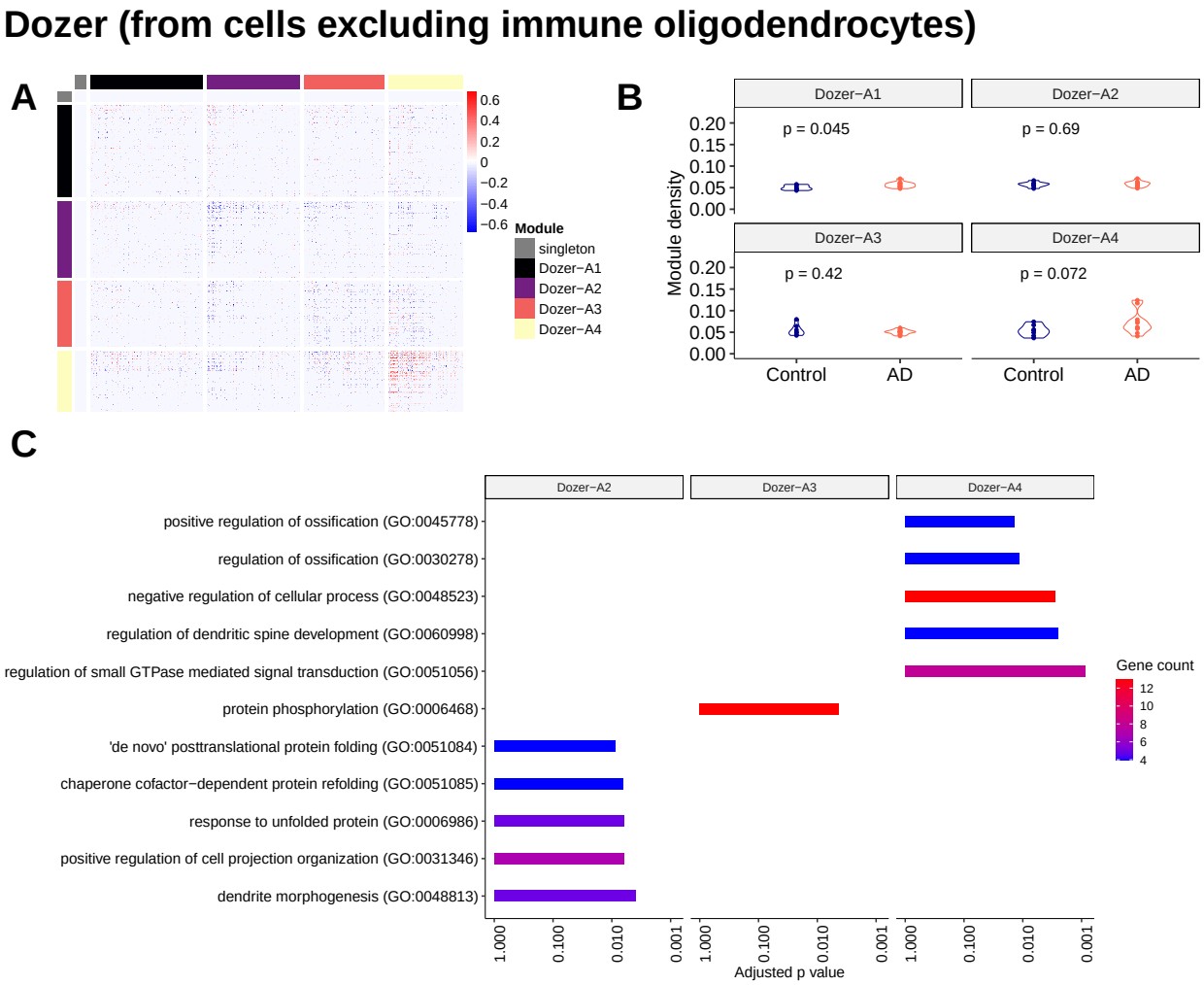

Figure S25: **(A)** Heatmap of gene modules from the Dozer "difference network" constructed with cells excluding immune oligodendrocytes. **(B)** Violin plot of module densities. **(C)** Top 5 significant GO terms from gene set enrichment analysis of each module (modules with no significant terms are excluded).

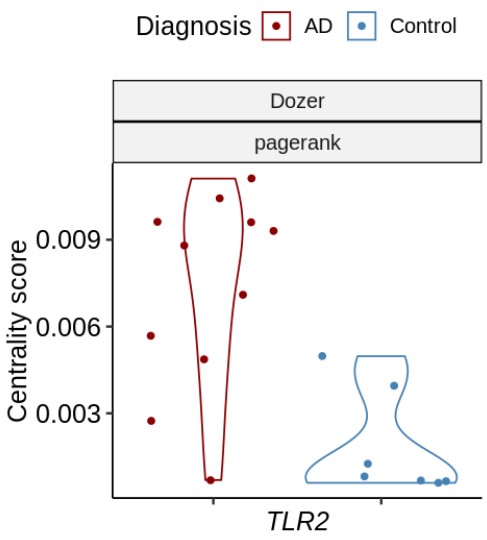

Figure S26: Violin plots of the gene centrality measures in the AD and Control groups for the genes with significant differential centrality at 5% FDR in the `Morabito_2021` dataset. The two grid labels indicate the coexpression network method and centrality metric.

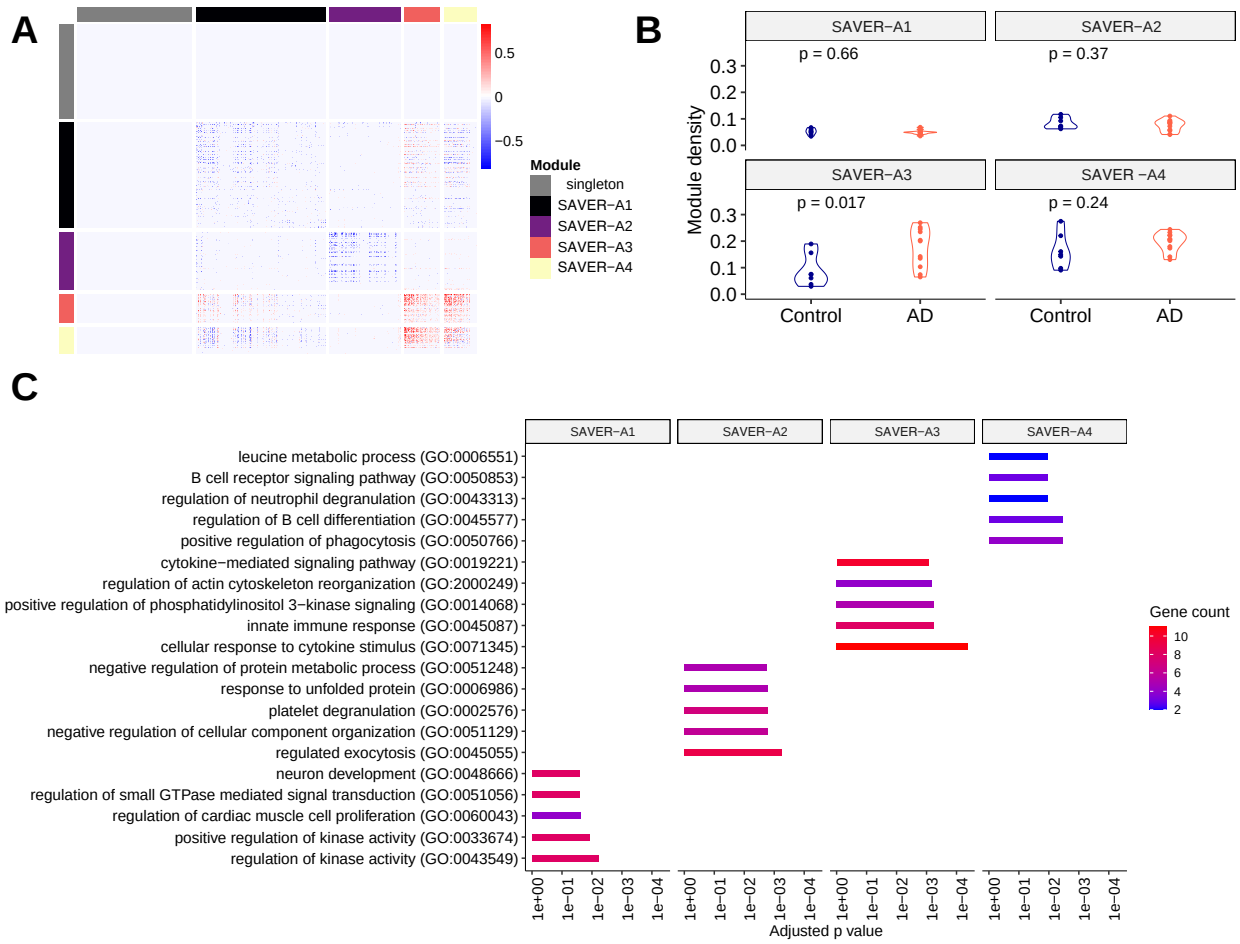

Figure S27: **(A)** Heatmap of gene modules from the SAVER "difference network". **(B)** Violin plot of module densities. **(C)** Top 5 significant GO terms from gene set enrichment analysis of each module (modules with no significant terms are excluded).

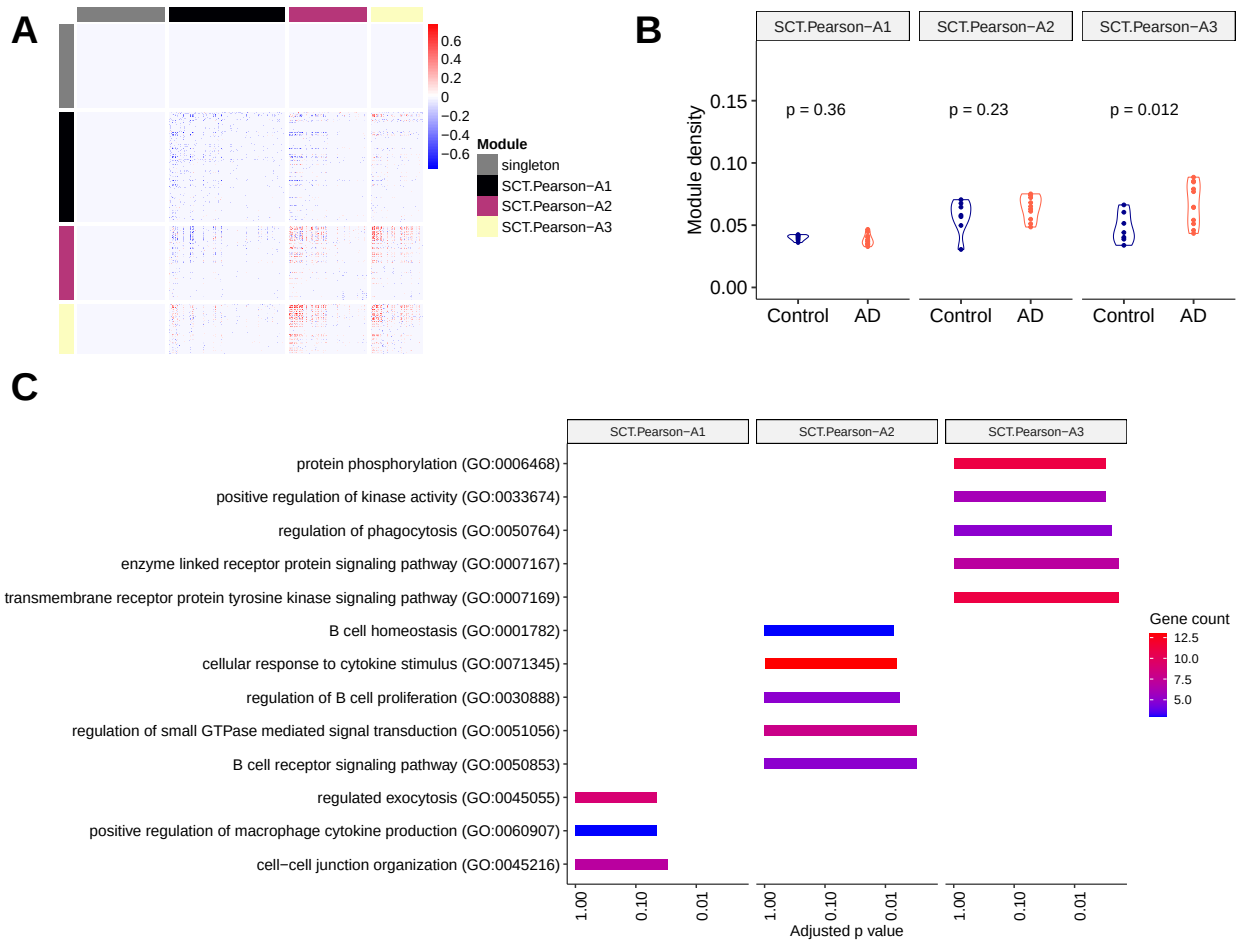

Figure S28: **(A)** Heatmap of gene modules from the SCT.Pearson "difference network". **(B)** Violin plot of module densities. **(C)** Top 5 significant GO terms from gene set enrichment analysis of each module (modules with no significant terms are excluded).

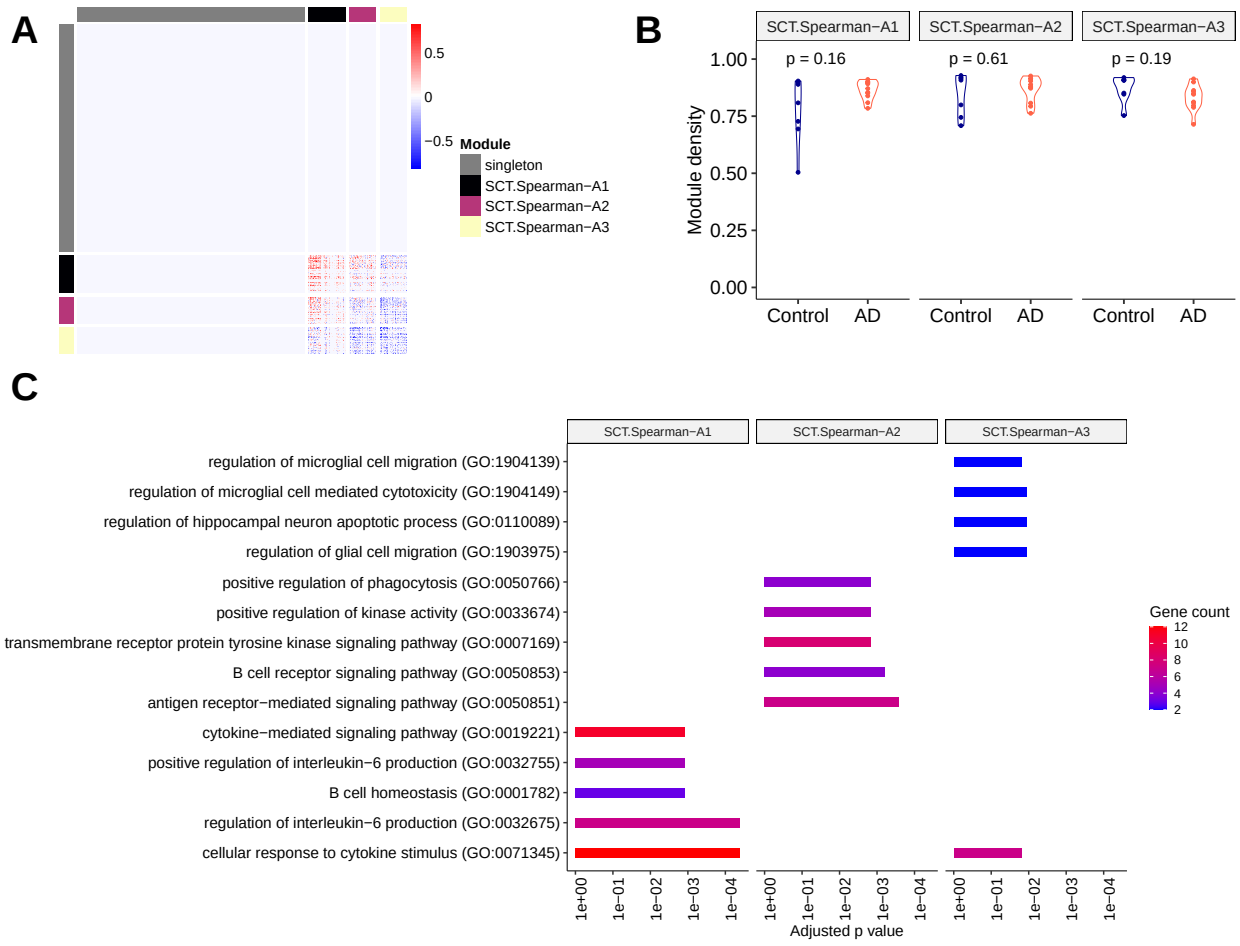

Figure S29: **(A)** Heatmap of gene modules from the SCT.Spearman "difference network". **(B)** Violin plot of module densities. **(C)** Top 5 significant GO terms from gene set enrichment analysis of each module (modules with no significant terms are excluded).

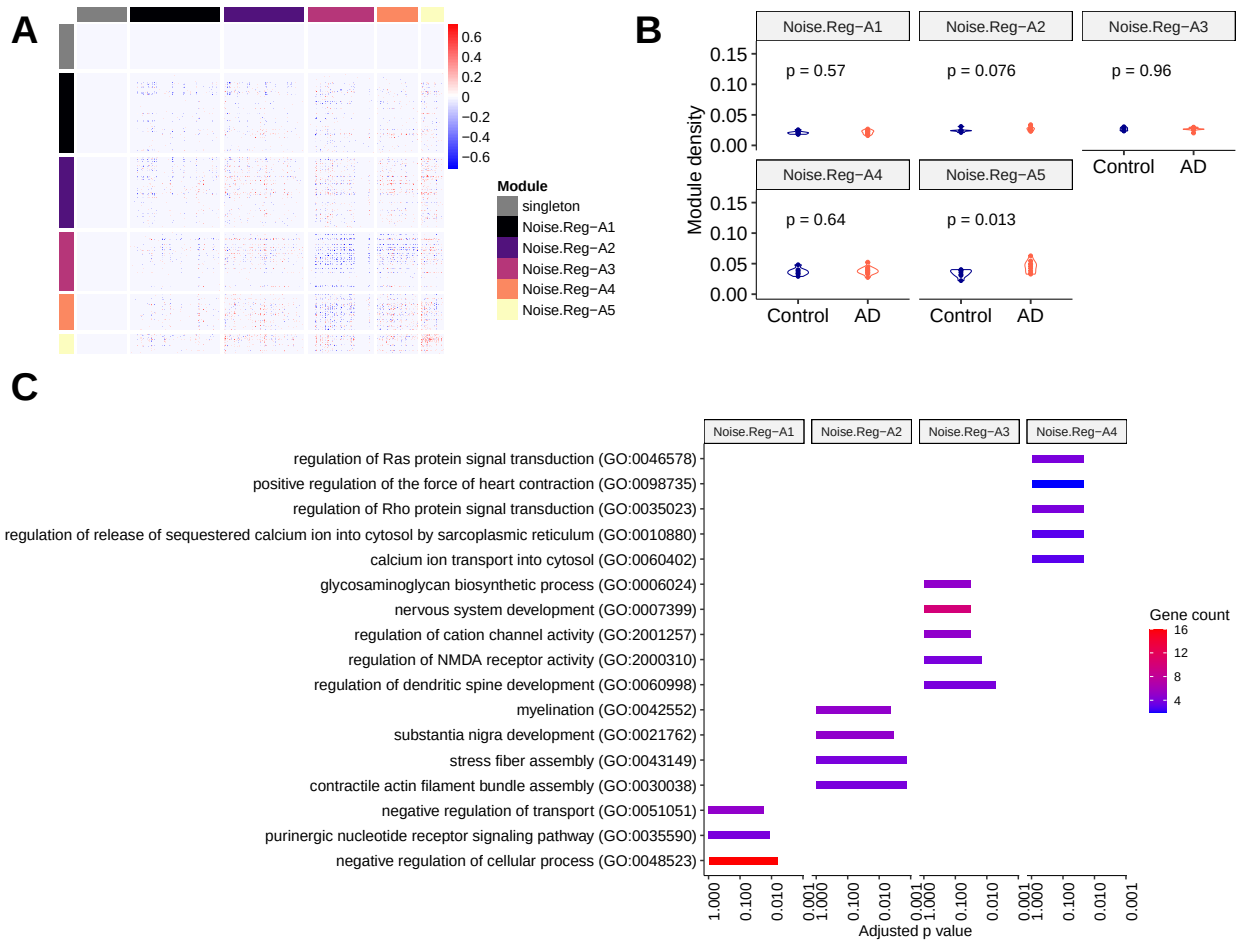

Figure S30: **(A)** Heatmap of gene modules from the Noise.Reg "difference network". **(B)** Violin plot of module densities. **(C)** Top 5 significant GO terms from gene set enrichment analysis of each module (modules with no significant terms are excluded).

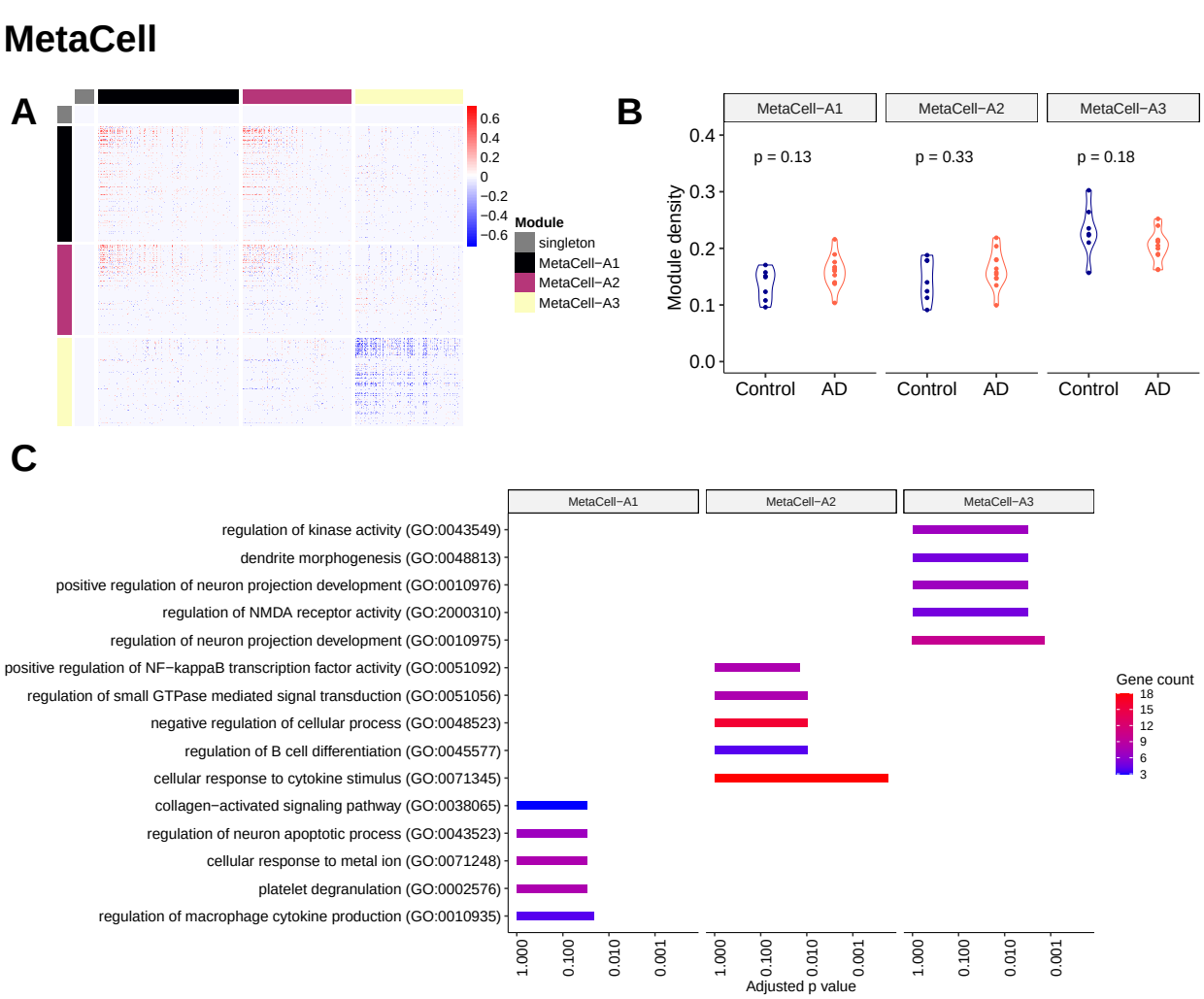

Figure S31: **(A)** Heatmap of gene modules from the MetaCell "difference network". **(B)** Violin plot of module densities. **(C)** Top 5 significant GO terms from gene set enrichment analysis of each module (modules with no significant terms are excluded).

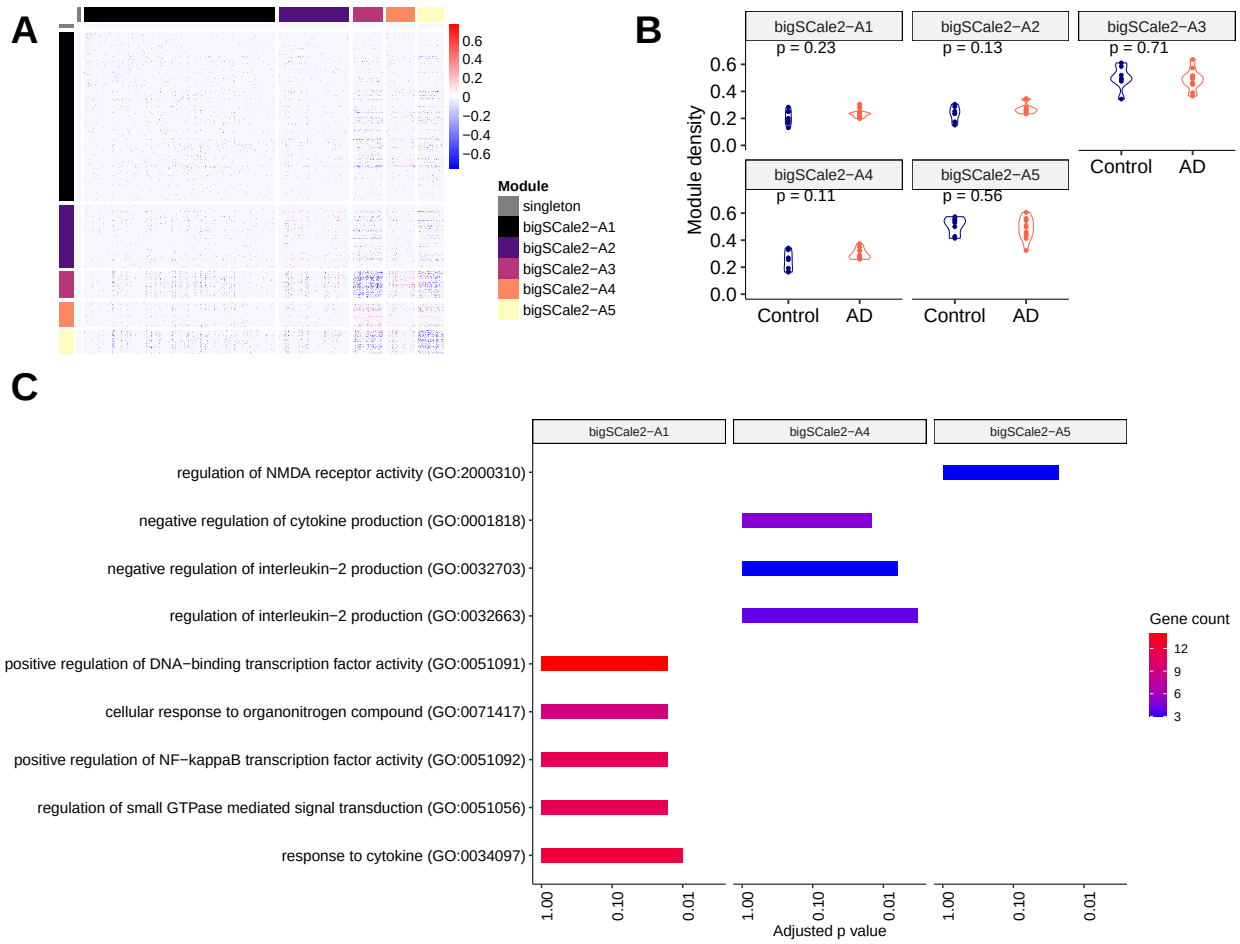

Figure S32: **(A)** Heatmap of gene modules from the bigSCale2 "difference network". **(B)** Violin plot of module densities. **(C)** Top 5 significant GO terms from gene set enrichment analysis of each module (modules with no significant terms are excluded).

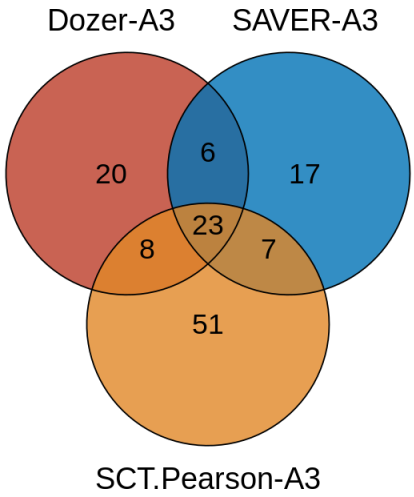

Figure S33: Venn diagram of gene modules identified by Dozer, SAVER, and SCT.Pearson, whose module densities significantly associated with the AD diagnosis.

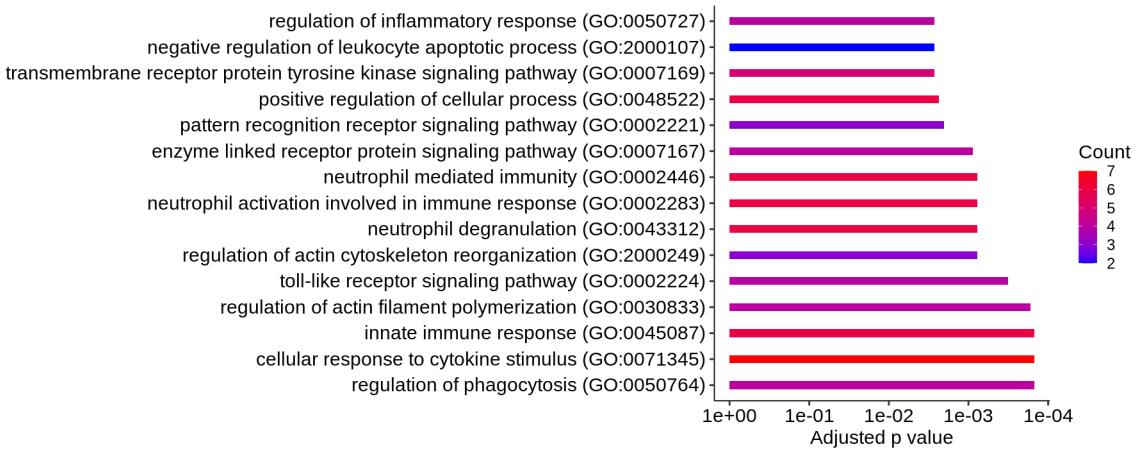

Figure S34: GO enrichment analysis of the genes shared between coexpression modules "Dozer-A3", "SAVER-A3", and "SCT.Pearson-A3".

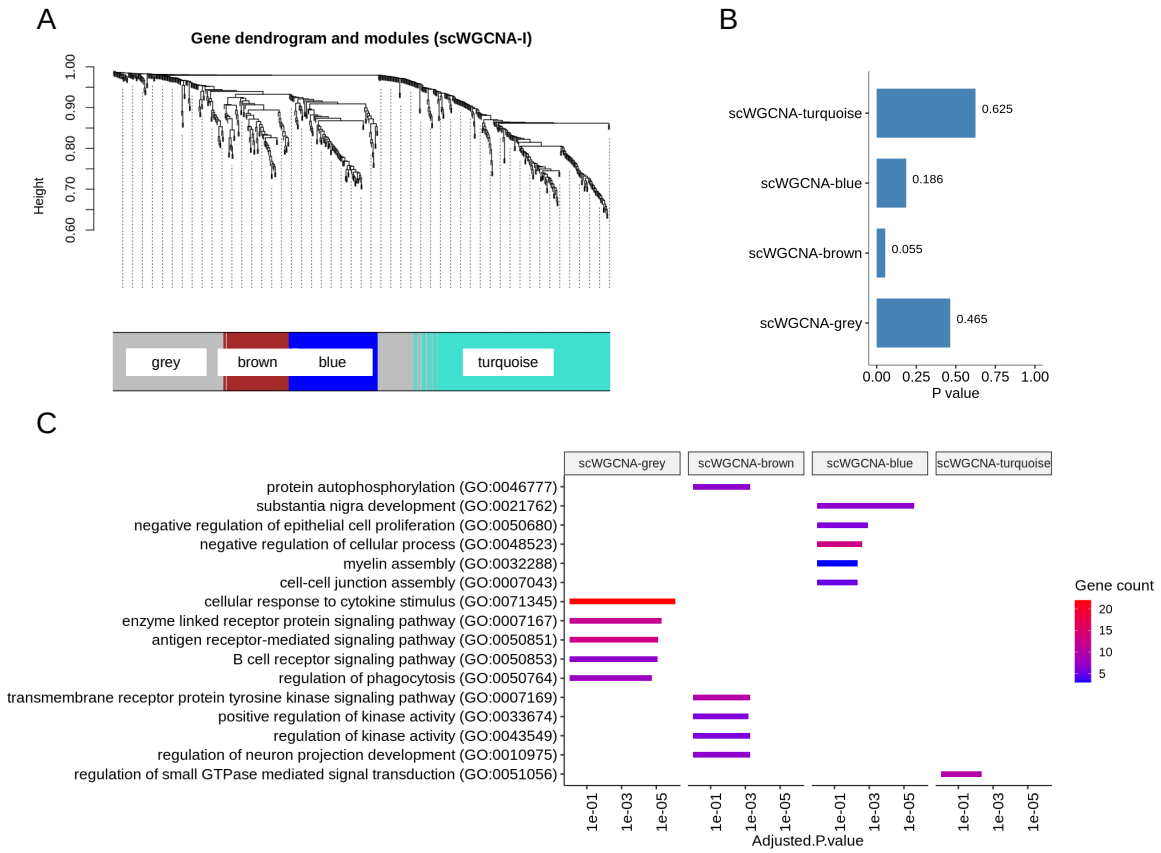

Figure S35: **(A)** Module dendrograms of scWGCNA-I. **(B)** P-values of the association tests between module eigengene expression and diagnosis. **(C)** Top 5 significant GO terms for each gene module.

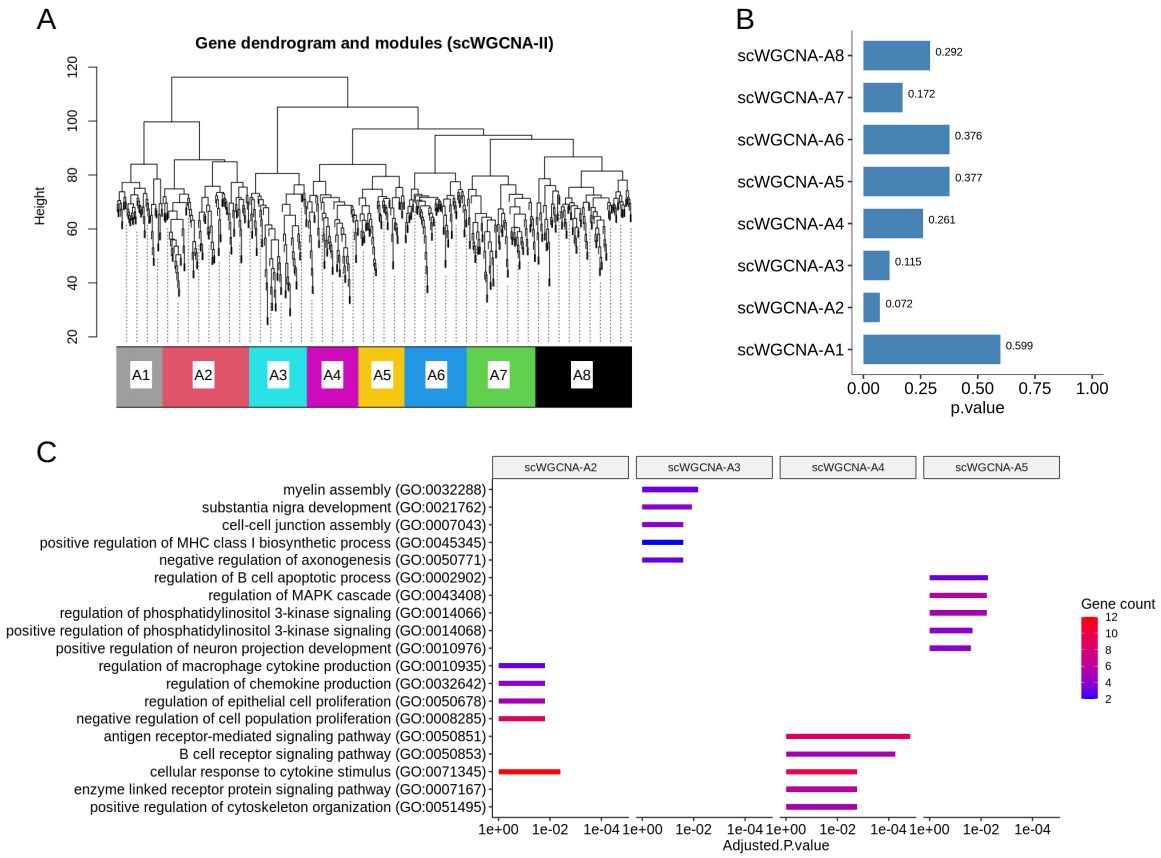

Figure S36: **(A)** Module dendrograms of scWGCNA-II. **(B)** P-values of the association tests between module eigengene expression and diagnosis. **(C)** Top 5 significant GO terms each gene module (modules with no significant terms are excluded).

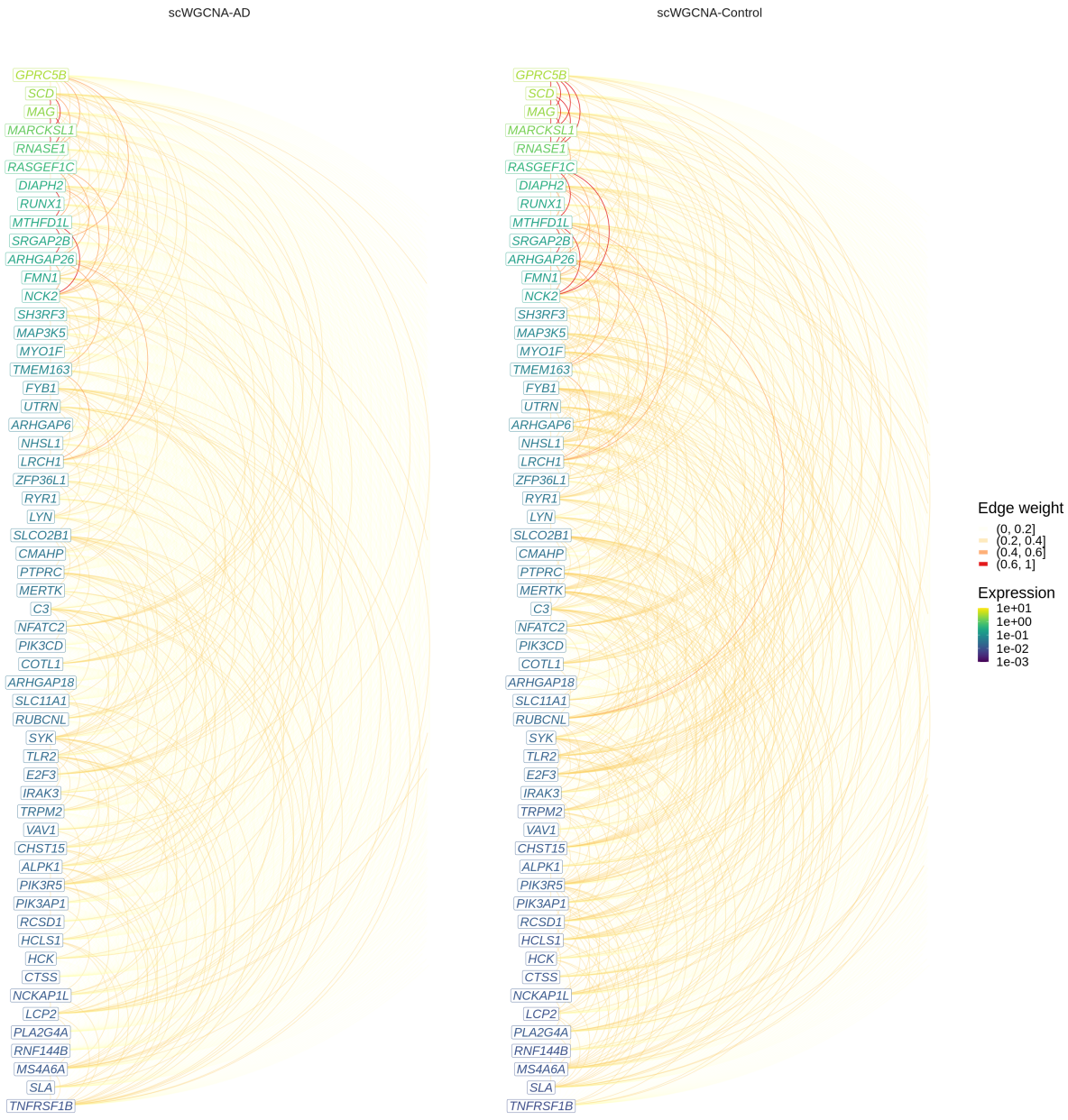

Figure S37: Hive plot visualization of scWGCNA gene correlation matrices for module Dozer-A3 in Control and AD groups. Genes are ordered from high (top) to low (bottom) by their average expression across all donors, and colored by the average expression in the corresponding diagnosis group. The color of arcs between the genes in this linear layout depicts the absolute correlation between genes in scWGCNA networks of AD (left) and Control (right) donors.

S5 Additional figures

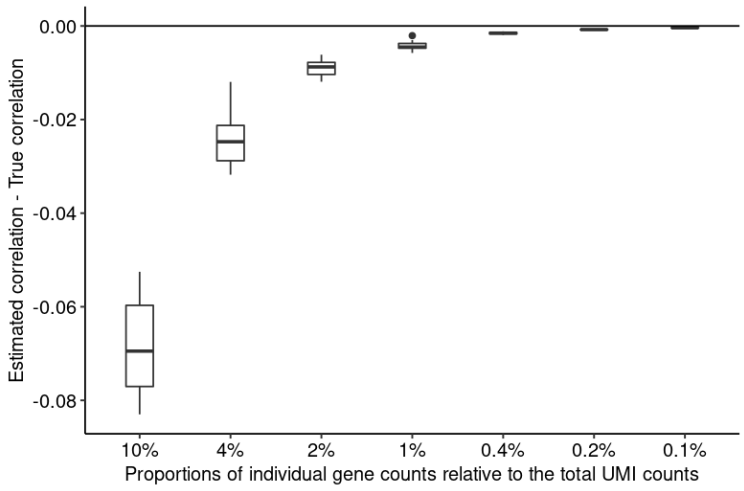

Figure S38: Boxplots for the bias in estimated gene pair correlations induced by the normalization procedure using total UMI counts as a global cell size factor. The data is simulated using the package `Splatter` [11]. The simulation parameters are taken from the example object `newSplatParams` in `Splatter`. In the simulation, the true correlations between gene pairs are set to zero. Gene pair correlations are computed using expression values normalized by total UMI counts. The estimation bias is calculated for seven different gene proportion scenarios (x-axis), ranging from genes accounting for 10% to 0.1% of the total UMI counts.

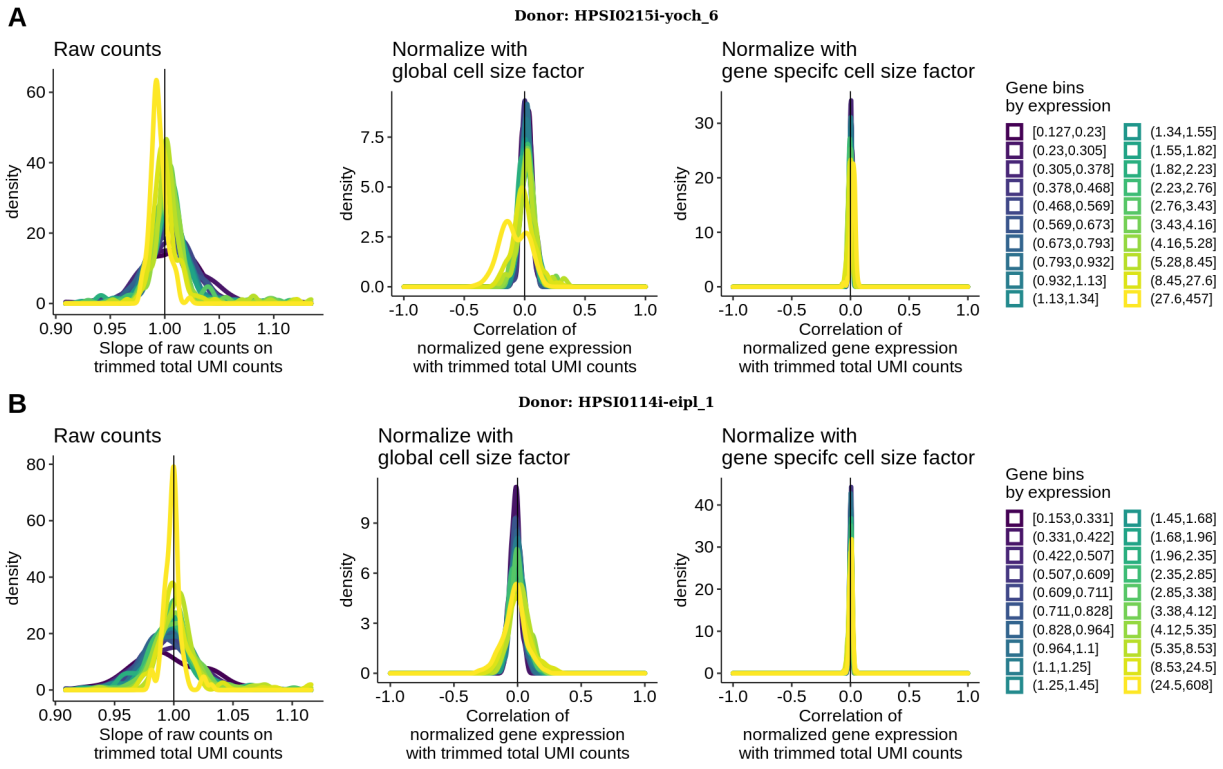

Figure S39: Diagnostics for cell size factor estimation. Panels A and B represent datasets from two different donors. The leftmost panels of A and B display the density of regression slopes between raw counts and trimmed total UMI counts in gene groups stratified by expression. The middle (right) panels of A and B show the density of correlations between normalized expression with a global cell size factor (with gene specific cell size factors) and trimmed total UMI counts in gene groups stratified by expression.

S6 Data summary

`Jerber_2021`. scRNA-seq dataset from [4] harbored iPSC cells from multiple donors under neuronal differentiation. Cells were profiled on days 11, 30, and 52 and a phenotypic trait named differentiation efficiency score was derived for each donor. Donors were further classified into two phenotype groups as "success" and "failure" in neuronal differentiation using the differentiation efficiency score. All the personalized gene coexpression network analyses utilized the P_FPP (Proliferating Floor Plate Progenitors) cells on day 11. Donors with fewer than 500 cells, with missing data in differentiation efficiency score, and six donors with extremely low sequencing depths (less than 14% of the median sequencing depth among all donors) were filtered out, leaving a total of

62 donors for the analysis. Genes with an average noise ratio among donors larger than 0.9 were filtered out, resulting in 912 genes per network.

We employed the full `Jerber_2021` dataset for personalized coexpression network analysis, and utilized two subsets of the dataset for false discovery analysis and investigating the robustness against sequencing depth differences. The first subset comprised of cells with a depth range between 12k to 17k total UMI counts from 20 donors, while the second subset consisted of 1,139 and 963 cells from a single donor, sequenced in batches labeled "pool2" and "pool3", respectively.

`Morabito_2021`. snRNA-seq data of [7] were from postmortem human tissues of 11 subjects with Alzheimer disease and and 7 age-matched control subjects. Oligodendrocytes, which accounted for 60% of the total number of cells in the datasets (with a median number of 2,005 cells per donor) were utilized in all the the coexpression analyses. To facilitate a direct comparison with the coexpression network analysis conducted in the data paper (scWGCNA), the same set of 1,252 genes were used as the starting gene set and were further filtered out if they satisfied any of the following conditions: (i) expressed in less than 2 cells in any donor; (ii) average noise ratio over donors is larger than 0.85. This resulted in 570 genes per donor for the analysis.

`Cuomo_2020`. scRNA-seq data of iPSC cells from [2] were used to generate simulation parameters. Cells in this dataset have high sequencing depths (averaging approximately 530K total counts per cell), allowing estimation of more realistic parameters to serve as ground truth in simulations.

References

[1] Benjamini Y and Hochberg Y. 1995. Controlling The False Discovery Rate - A Practical And Powerful Approach To Multiple Testing. *J. Royal Statist. Soc., Series B* **57**: 289 – 300.

[2] Cuomo AS, Seaton DD, McCarthy DJ, Martinez I, Bonder MJ, Garcia-Bernardo J, Amatya S, Madrigal P, Isaacson A, Buettner F, et al.. 2020. Single-cell RNA-sequencing of differentiating iPS cells reveals dynamic genetic effects on gene expression. *Nature communications* **11**: 1–14.

[3] Finak G, McDavid A, Yajima M, Deng J, Gersuk V, Shalek AK, Slichter CK, Miller HW, McElrath MJ, Prlic M, et al.. 2015. MAST: a flexible statistical framework for assessing transcriptional changes and characterizing heterogeneity in single-cell RNA sequencing data. *Genome biology* **16**: 1–13.

[4] Jerber J, Seaton DD, Cuomo AS, Kumasaka N, Haldane J, Steer J, Patel M, Pearce D, Andersson M, Bonder MJ, et al.. 2021. Population-scale single-cell RNA-seq profiling across dopaminergic neuron differentiation. *Nature genetics* **53**: 304–312.

[5] Kuleshov MV, Jones MR, Rouillard AD, Fernandez NF, Duan Q, Wang Z, Koplev S, Jenkins SL, Jagodnik KM, Lachmann A, et al.. 2016. Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic acids research* **44**: W90–W97.

[6] Love MI, Huber W, and Anders S. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome biology* **15**: 1–21.

[7] Morabito S, Miyoshi E, Michael N, Shahin S, Martini AC, Head E, Silva J, Leavy K, Perez-Rosendahl M, and Swarup V. 2021. Single-nucleus chromatin accessibility and transcriptomic characterization of Alzheimer's disease. *Nature Genetics* **53**: 1143–1155.

[8] Nagy C, Maitra M, Tanti A, Suderman M, Théroux JF, Davoli MA, Perlman K, Yerko V, Wang YC, Tripathy SJ, et al.. 2020. Single-nucleus transcriptomics of the prefrontal cortex in major depressive disorder implicates oligodendrocyte precursor cells and excitatory neurons. *Nature neuroscience* **23**: 771–781.

[9] Smyth GK. 2005. Limma: linear models for microarray data. In *Bioinformatics and computational biology solutions using R and Bioconductor*, pp. 397–420. Springer.

[10] Sun T, Song D, Li WV, and Li JJ. 2021. scDesign2: a transparent simulator that generates high-fidelity single-cell gene expression count data with gene correlations captured. *Genome biology* **22**: 163.

[11] Zappia L, Phipson B, and Oshlack A. 2017. Splatter: simulation of single-cell RNA sequencing data. *Genome biology* **18**: 174.

[12] Zhang Q, Liu W, Zhang HM, Xie GY, Miao YR, Xia M, and Guo AY. 2020. hTFtarget: a comprehensive database for regulations of human transcription factors and their targets. *Genomics, proteomics & bioinformatics* **18**: 120–128.