GSEA and co-expression network approach to identify molecular processes affected in Porto-sinusoidal Vascular Disease

Aishwarya Iyer[1], Martina Kutmon[2], Chris T. Evelo[1,2], Cenna Doornbos[3], Friederike Ehrhart[1*]

[1]Department of Bioinformatics - BiGCaT, NUTRIM/MHeNs, Maastricht University, Maastricht, The Netherlands
[2]Maastricht Centre for Systems Biology (MaCSBio), Maastricht University, Maastricht, The Netherlands
[3]Medical BioSciences Department, Radboud University Medical Centre, Nijmegen, The Netherlands

*corresponding author

Keywords: PSVD, INCPH, rare liver disease, pathway analysis, GSEA, WGCNA

Electronic word count: 3327
Number of Figure: 5
Number of tables: 2

Conflict of interest
The authors declare no conflicting interests.

Financial support statements
This work is supported by the funding from the European Union's Horizon 2020 research and innovation programme under the EJP RD COFUND-EJP N° 825575.

Author contribution
CE, FE, AI and MK designed the study, AI, MK and FE contributed to data analysis, AI wrote the first draft, MK, FE, CE, CD critically reviewed the first draft and provided further textual contributions, MK, CE and FE supervised the study.

Abstract

Background & Aims: Porto-sinusoidal vascular disease (PSVD) is a complex rare liver disease characterized by the absence of cirrhosis with or without the presence of portal hypertension or histological lesions. Given the knowledge gaps in the mechanisms involved in this disease with unknown etiology, we used omics-based approaches to further elucidate the pathways affected in PSVD thereby facilitating an improvement in the prognosis, diagnosis, and treatment options for these patients.

Methods: For this study we used a microarray dataset (GEO:GSE77627) of 11 histologically normal liver biopsies and 18 PSVD liver biopsies. First approach, differential gene expression analysis was performed and next gene set enrichment analysis was used to identify enriched biological pathways. A network-based approach of weighted gene coexpression analysis was implemented to identify modules of interconnected genes related to the diagnosis of the patients. We further studied the pathways enriched in these modules to allow identification of processes explaining the mechanisms involved in PSVD, while the gene network could also be used to understand the connections between the processes.

Results: Gene set enrichment of differentially expressed genes indicated an increase in signaling and cell-cycle related processes and a decrease in metabolism-related processes. Coexpression network and module analysis further validated these results by elucidating connections between GPCR signaling, energy metabolism and cell-migration related processes. Furthermore, an additional connection between fibrin clot formation processes, inflammation and immune response and cell cycle and respiration processes was identified.

Conclusion: Signaling and metabolism-related processes are deregulated in PSVD patients. Furthermore, two triangular connections (GPCR signaling- energy metabolism-cell migration and fibrin-clot formation- inflammation and immune response - cell -cycle and respiration) revealed unique unknown connections involved in PSVD etiology.

Impact and implications

PSVD is a complex rare liver disease with significant knowledge gaps in the understanding of the mechanisms and pathways affected at the molecular level. In this study, we use publicly available transcriptomics data and bioinformatics tools to elucidate pathways affected in PSVD patients. In this study, we found potential novel relations between pathways with the two triangular connections (GPCR signaling- energy metabolism-cell migration and fibrin-clot formation- inflammation and immune response - cell -cycle and respiration). These newfound connections between pathways might shed light on the etiology of this disease and help researchers develop effective diagnosis and prognosis for patients suffering from PSVD.

Introduction

Porto-sinusoidal vascular disease (PSVD) is a complex rare liver disease characterized by the absence of cirrhosis with or without the presence of portal hypertension or histological lesions. It is a recently coined term to improve the understanding of the disease by reducing the effect of heterogeneity, facilitating improved diagnosis, and simplifying comparisons between different clinical studies(1). PSVD is a rare disease with currently unknown prevalence.

Diagnosis for patients suffering from PSVD includes non-invasive imaging methods focusing on splenomegaly and portosystemic collaterals and hepatic vein venography. However, imaging by itself is insufficient and further invasive methods like biopsies are an essential part of the diagnostic routines for PSVD(1). The accuracy of analyzing the histology remains highly variable depending on the experience of the histopathologist(1). In a metabolomics study published recently, a group of metabolite markers were identified that were able to predict patients diagnosed with PSVD with an accuracy of 88%(2).

The mechanism of disease development for PSVD is not known but is dependent on the vascular developments within the liver(1). 43-48% of patients with PSVD patients have one or more associated conditions majorly classified into disorders of immunity, blood diseases and pro-thrombotic conditions, infections, congenital or familial defects, and drug exposure(1).

Because PSVD incorporates a small, heterogeneous diseased patient group with varying physiological and histological features, there is sparse information regarding molecular pathways or processes affected in this condition. A recent study by Hernandéz-Gea et al., revealed previously unknown regulatory pathways affected in PSVD using co-expression analysis using gene expression data from healthy, PSVD and liver cirrhosis patients. The study indicated deregulation of pathways specific to vascular homeostasis and oxidative phosphorylation affecting the endothelial function(3).

Omics analysis, especially transcriptomics, has been widely used to understand genes differentially regulated in a disease and next to link these genes to pathways thereby explaining the molecular mechanisms underlying the disease. Also, other approaches based on network algorithms, especially co-expression networks have been constructed from omics data to identify novel disease-specific processes.

In this study, we implemented two methods, first, gene set enrichment analysis and second, co-expression network analysis using transcriptomics to identify pathways or processes affected in patients with PSVD. Understanding the pathways or processes would shed light on the mode of action of the disease thereby allowing to improve the prognosis, diagnosis and the treatment options available to the patients suffering from this rare disorder.

Methods

Data

A previously published transcriptomics dataset by Lozano et al. was obtained from the GEO database (GEO:GSE77627)(3), ethical approval (HCB/2009/5448). The dataset contains liver mRNA expression profiles for histologically normal liver (HNL), PSVD and liver cirrhosis patients. In this study, liver cirrhosis patients were excluded given that their transcriptomics profile overlapped with the PSVD transcriptomic profile (Supplementary, Fig. S1). Additional clinical data and information was obtained from the original study authors, Clinic Barcelona. The measured variables included information on sex, wedged hepatic vein pressure (WHVP), hepatic venous pressure gradient (HVPG), Bilirubin, platelet count, spleen size, liver stiffness, PSVD-specific and non-specific biopsy markers.

Data pre-processing

The transcriptomics data was pre-processed using the 'lumi' R package(4). Background correction and quantile normalisation was performed using the *neqc* function with an offset value of 16. The data was re-annotated using NCBI gene identifiers using the Illumina HumanHT-12 DASL 4.0 R2 expression beadchip platform annotation. Samples with incomplete sex information were removed from the analysis. The data distribution for before and after normalized data is provided in Supplementary, Fig.S2.

To detect outliers, hierarchical clustering on the samples was performed and the dendrogram is provided in Supplementary Fig. S3(A). Based on the clustering, sample 'PSVD05' was removed from further analysis. The dendrogram of samples and clinical variables measured for these samples are provided in Supplementary Fig. S3 (B).

Differentially expressed gene (DEG) analysis

Differential gene expression analysis was performed to determine genes that are significantly altered (up- or down-regulated) in PSVD patients compared to healthy control after sex correction using the 'limma' R-package(5). The cut-off for significantly up-regulated gene is logFC > 1 and adjusted p-value < 0.05. For down-regulated genes the cut-off used is logFC < -1 and adjusted p-value < 0.05.

Gene set enrichment analysis for DEGs

Gene set enrichment analysis (GSEA) was performed using the clusterProfiler (v4.6.2) R-package using a maximum and minimum gene set sizes of 1000 and 30 respectively(6). The DEGs were ranked based on the product of signed log2 fold change and the negative logarithm of the adjusted *p*-value, see equation below.

ranking = sign(log2FC) * -log10(adjusted p-value)

For the enrichment analysis, the human canonical pathway gene sets from the Molecular Signatures Database (MSigDB, v2023.2.Hs) were used (7, 8). The pathway genesets from Kyoto Encyclopedia of Genes and Genomes (KEGG, 186 gene sets (9, 10)), WikiPathways (733 gene sets (11)) and Reactome (1,654 gene sets (12)) were included. Additionally, 7,751 gene sets from the biological process ontology from Gene Ontology (GO) were included for the analysis (13, 14).

Cytoscape visualisation of the enrichment analysis

The gene set enrichment analysis results were visualized in Cytoscape using a custom R script.
To calculate similarity between two enriched terms the overlap coefficient (*k*) was used (15).

$$= \frac{|\ \cap\ |}{|(,)|}$$

A cut-off score of $k > 0.5$ was used to add an edge between two enriched terms.

Co-expression network construction

The weighted gene co-expression analysis (WGCNA) is a network algorithm tool that constructs correlation networks based on similar gene expression patterns across samples. It uses an unsupervised approach to identify co-expression gene modules. This tool was implemented using the "WGCNA" R package in order to identify gene expression modules correlating to the PSVD phenotype(16).

Firstly, the lowly expressed genes, i.e. genes with average expression values below 0.05 were removed. The input was the pre-processed normalized data of all the samples used (healthy and PSVD). A step-by-step method was used to generate the consensus network and to further detect the modules.

Firstly, a similarity network was constructed using Pearson correlation for all gene pairs in the dataset. Next, a signed adjacency matrix was calculated by raising the similarity matrix to a soft-thresholding power ($\beta = 14$).

Next, the adjacency matrix was converted into a Topological Overlap Matrix (TOM). The TOM is a robust network similarity measure by calculating the effect of neighboring nodes on pairs of genes. The resulting proximity matrix is then converted to a dissimilarity TOM matrix. The dissimilarity measure works well in the clustering of gene expression profiles by identifying distinct gene modules.

From these results, a dendrogram was constructed using the dissimilarity matrix and average hierarchical clustering method (supplementary Fig. S4). To identify modules with highly interconnected genes, a dynamic tree-cut method was implemented with a minimum module size of 100.

Identification of clinically significant gene modules

The consensus co-expression network generated previously was then used to identify modules of highly interconnected genes or genes with a higher degree of co-expression using the dynamic tree-cut algorithm. For identifying modules relevant to clinical phenotypes associated with PVSD, the module eigengene, and the module membership are calculated. Finally, the correlation between the modules and the clinical phenotypes associated with PSVD was calculated to identify clinically significant gene modules.

Module over-representation analysis

Functional analysis of the clinically significant modules was performed using the over - representation analysis in the 'clusterProfiler' R package. The genesets for performing the over - representation analysis were obtained from Molecular Signatures Database (MSigDB). The pathway genesets databases included were Kyoto Encyclopedia of Genes and Genomes (KEGG), WikiPathways and Reactome, Gene Ontology: Biological process.

Data and workflow

Transcriptomics data available on GEO:GSE77627. Clinical data can be made available on request. The workflow of this study is available in WorkflowHub: https://workflowhub.eu/workflows/1040.

Results

Differentially expressed genes in PSVD

The raw transcriptomics data consisting of 29,377 probes was processed to correspond to 17,105 genes (annotated with NCBI gene identifiers). DEG analysis identified 4699 (2777 genes down-regulated and 3591 genes up-regulated) differentially expressed genes.

(Best place for Fig. 1)

The most up- and down-regulated gene is microRNA 128-1 with a log fold change of 7 and Small nucleolar RNA, C/D box 94 (SNORD 94) with a log fold change of -5.8 respectively (Fig. 1).

Gene set enrichment analysis

The results of GSEA were visualized using Cytoscape (Fig. 2). The nodes in Fig. 2 were clustered based on the genes common among two terms. Terms, which are closer together therefore share functional similarities with each other. This can be observed in Fig. 2 with terms related to metabolism and signaling clustering with similar terms. Fig. 2 shows positively enriched terms of G protein coupled receptor signaling (GPCR) pathway and Multi organism reproductive process to be connected with negatively enriched cellular lipid metabolic process indicating a potential mechanism associated with PSVD. Additionally, we observe terms of sleep regulation, cell cycle regulation (Retinoblastoma gene in cancer, Hematopoietic stem cell differentiation and cell lineage) terms to be positively enriched whereas terms of Golgi vesicle transport, peroxisome, nonsense-mediated decay and chromatin modifying enzymes to be negatively enriched. The distribution of the differentially expressed genes in the pathways around G-coupled receptor signaling (WikiPathways: WP35) and Class a-1 rhodopsin receptors signaling (WikiPathways: WP4419) were investigated in detail (Supplementary, Fig. S5 (A) and (B)). The group of thyrotropin releasing hormones was strongly upregulated while GPR4, GPR68 and GPR 18 are significantly upregulated in the pathway visualization of class A/1 rhodopsin like receptor.

(Best place for Fig. 2)

Identification of key modules using co-expression network

The co-expression network was constructed using the gene expression data measuring 17,105 genes from 27 liver biopsy samples (11 healthy and 16 PSVD patients) with 'WGCNA' R package. The patients with PSVD included in this study have clinical signs of portal hypertension. The two most frequent signs of portal hypertension in PSVD patients being splenomegaly and the presence of gastroesophageal varices. Splenomegaly was present in all PSVD patients with a mean size of 15.3 ± 2.7 cm (Table 1). Additionally, 68% of PSVD patients show clinically elevated

portal hypertension with a mean hepatic venous pressure gradient (HVPG) of 7.9 ± 3.8 (Table 1). The clinical and histological data for all the samples are in Supplementary Table 1.

Table 1: Clinical characteristics

	HNL group (n=11)	PSVD group (n = 16)
Sex (Male)	4 (36%)	11 (69%)
Presence GEV	0	5 (31%)
HVPG (mmHg)	3.9 ± 0.8	7.9 ± 3.8
WHVP (mmHg)	7.0 ± 0.7	13.3 ± 4.1
Platelet count (x10^9/l)	236.5 ± 63.8	163.6 ± 143.0
Total Bilirubin (mg/dl)	0.8 ± 0.4	1.2 ± 1.2
Spleen size (cm)	9.5 ± 0.7	15.3 ± 2.7
Liver stiffness (kPa)	5.8 ± 0.9	7.8 ± 3.1
Direct Bilirubin (mg/dl)	0.4 ± 0.2	0.4 ± 0.3

* Mean ± SD

GEV, Gastroesophageal varices; HVPG, hepatic venous pressure gradient; WHVP, wedged hepatic venous pressure, HNL, healthy normal liver; PSVD, porto-sinusoidal vascular disease.

A range of soft-thresholding powers (*β*) were used to assess the scale free topology of the network constructed. For this analysis, *β* of 14 was selected which had a scale-free topology fit (R^2) of 0.82; shown in Fig. 3(A). Using the average hierarchical clustering and dynamic tree cut method, a total of 24 distinct gene modules and the corresponding coexpressed genes for each module were identified. To explore the relationship between the identified coexpressed modules and the clinical variables associated with PSVD phenotype, such as diagnosis, sex, gastroesophageal varices, spleen size, HVPG, WHVP, platelet count, PHT-specific and PSVD-specific histological markers. Out of the 24 distinctly identified modules, 14 modules were selected which significantly correlated to the diagnosis of PSVD, shown in Fig. 3(B) (see Supplementary, Fig. S6 for module-trait relationship of all 24 modules). The number of genes for the 14 modules is summarized in Table 2 including the percentage of up-regulated (fold change > 1 and adjusted p value < 0.05) and down-regulated genes (fold change < 1 and adjusted p value < 0.05).

(Best place for Fig. 3)

Table 2: Number of genes significantly expressed in modules significantly correlating with diagnosis of PSVD patients.

	% up-regulated genes	% down-regulated genes	Total genes in module
grey60	83.0	0	259
red	59.8	0	515
black	36.9	0	482
turquoise	41.1	0	4150
yellow	26.6	0	1156
darkturquoise	13.8	0	160
lightyellow	19.9	0	201
blue	0	58.1	3693
brown	0	28.4	1395
green	46.2	0	876
salmon	58.8	0	316
cyan	47.0	0	296
lightcyan	0	32.8	287
tan	0	22.6	345

Pathway analysis for key modules

To understand the molecular processes affected in PSVD, we performed functional over-representation analysis for the 14 selected modules. Fig. 4 elucidates a broader understanding of processes affected in each module along with the relations these modules have with each other. Fig. 4 shows negative correlation between the brown module representing energy

metabolism with the turquoise and yellow modules representing the GPCR signaling process. Additionally, the red module representing the inflammation and immune response processes is positively correlated with the grey60 module which represents sleep regulation. Blue module representing cell growth and migration related processes, having the highest degree, is negatively correlated with the grey60 module and the red module.

Hereafter, we selected the top three positively correlating modules being the grey60, red and black module; and the top two negatively correlating modules being black and blue modules for further understanding of the processes involved in PSVD. The grey60 module with highest significant positive correlation to PSVD diagnosis had one significant over-represented term belonging to the 'sleep regulation' pathway (WP3591). The next highest positive correlated red module had gene ontology term of 'cellular starvation' and canonical pathway term involving immune response processes like the 'toll-like receptor signaling pathway', 'pattern recognition receptor signaling pathway' and 'Ikk complex recruitment mediated by RIP1'. The black module captured processes involving 'complement and coagulation cascades', 'formation of fibrin clot' and 'Intrinsic pathway of fibrin clot formation'. Fig. 5 represents the enrichment analysis results for the blue and brown module with clustering of similar processes visualized using the dendrogram. The blue module mainly indicates processes related to cell growth, endothelial cell migration and protein translation. The brown module with 46 enriched processes mainly involved in general metabolic processes. The results for the enrichment analysis for the rest of the modules are in Supplementary Fig. S7.

(Best place for Fig. 4 and 5)

Conclusions

In this study, we re-analyzed a transcriptomics dataset on PSVD originally produced by Lonzano et al. and focused on the in-depth comparison of transcriptome changes between the PSVD and the HNL group.

The DEG analysis showed that, among others, microRNA 128-1 was heavily up and SNORD94 down-regulated. For the microRNA 128-1 there is some evidence that it is involved in cardiac hypoxia, ischemia and reperfusion injury mediated apoptosis (17) which may play a role in PSVD, too.

One previous study had found SNORD94 being robustly down-regulated as well when comparing between non-alcoholic fatty liver (NAFL) and non- alcoholic steatohepatitis (NASH) transcriptomics profiles (18). Being the second study that identifies this marker in liver diseases it makes it interesting for future studies. Additionally, knockdown studies of SNORD94 have identified its potential role in modifying alternate splicing of mRNA due its role in the

methylation of spliceosomal RNA U6 (19) which could provide a mechanistic explanation to changes in chromatin methylation and protein translation processes found in this study.

Investigating the whole genes set using enrichment analysis, general signaling pathways (especially GPCR class A rhodopsin-like signaling), cell cycle processes are upregulated while metabolic processes, methylation and chromatin modification, protein translation and transport are downregulated (Fig. 3).
Sleep regulation and GPCR imbalance were identified in both GSEA and WGCNA analyses. Sleep regulation is generally found to be disturbed in patients with liver diseases and it is presumed to be bi-directional (20, 21). It was not reported in PSVD before.
The changes in the cell cycle processes could possibly be due to vascular injury, endothelial cell dysfunction and proliferation as described previously in Lozano et al. (3).
There are some prior studies on changes in metabolic processes but there, the processes need to be investigated separately. For lipid biosynthesis, e.g. a metabolomics study on INCPH patients revealed a decrease in metabolites associated with lipid metabolism compared to healthy patients (2). Thereby, further confirming the association between lipid metabolic pathways, which also links to obesity-caused metabolic syndrome, a risk factor of PSVD, and PSVD (1).
Changes in energy metabolism were identified by WGCNA but not in GSEA in this study. In the original study, which compared PSVD to both liver cirrhosis and HNL tissue, energy metabolism, respectively oxidative phosphorylation and TCA cycle, were upregulated (3).
Decreased protein translation and transport have been linked to endoplasmic reticulum stress in liver diseases. Furthermore, recent studies have elucidated the role of endoplasmic stress on the regulation of metabolism of lipids and its homeostasis in hepatocytes (22).
Changes in gene methylation patterns were previously identified in cirrhotic liver and hepatocellular carcinoma (23). Further studies should investigate if SNORD94 downregulation is causing or contributing to these changes in methylation.

The coexpression network analysis identified 14 modules with significant correlation with PSVD diagnosis (Fig. 4). The module eigengene correlation network elucidates the triangular relation between energy metabolism (Brown module), GPCR signaling (Turquoise and yellow modules) and the cell growth and migration processes (Blue module). This connects to the previous study by Lonzano et al. which identified processes of vascular homeostasis and oxidative phosphorylation as the key drivers of PSVD (3). Vascular homeostasis has been linked to cellular remodeling during vascular injury in portal hypertension and is a common histopathological finding linked to PSVD (3).
GPR4, GPR68 and GPR 18 are significantly upregulated in the pathway visualization of class A/1 rhodopsin-like receptors (Supplementary, Fig. S5 (B)). Interestingly, we also observed the

significant positive upregulation of a group of genes around Thyrotropin releasing hormone, which is linked to increased blood pressure (24). Understanding this triangular relation between blood pressure, vascular injury and GPCRs may contribute to the understanding of the disease mechanism and may provide potential biomarkers for diagnosis and treatment.
The positive relation between fibrin clot formation, inflammation and immune response and organic acid and lipid metabolic processes are also supported by the previous study. Note that the previous study used both cirrhotic liver and healthy liver as control groups vs. PSVD.

The weakness of this study is the relatively low number of patients but that is a general problem with rare disease studies. A way to overcome this problem is pooling and re-using of data from different studies but changes in the diagnosis criteria and definition of PSVD or INCPH over time makes this challenging. However, in re-investigating this subset of a previously published transcriptomics dataset we got generally similar but more explicit conclusions about PSVD and how it specifically compares to HNL. Additionally, the network analysis helped in identifying connections between known processes associated with PSVD, the energy metabolism - GPCR signaling - cell migration connection which could be used as a target for treatment. Also the link between SNORD94 and liver diseases requires further investigation as a potential treatment target. To understand the disease better, multi-omics data might help elaborate the regulation and mechanisms, especially investigating the number of changes in metabolic processes that are also linked to metabolic syndrome and other risk factors. Given the heterogeneity in patient risk factors, genetic and environmental backgrounds, studies focusing on personalized models could help in developing better treatment options for the patients.

To conclude, in this study we provide further evidence to the triangle of GPCR signaling - inflammation - energy metabolism as potential drivers of PSVD and identified SNORD94 as a marker and potential treatment target for liver disease.

Abbreviations (Alphabetical)

DEG - Differentially Expressed Gene
GSEA - Gene Set Enrichment Analysis
GPCR - G protein-coupled receptor
GPR4 - G protein-coupled receptor 4
GPR68 - G protein-coupled receptor 68
GPR 18 - G protein-coupled receptor 18
GEV - Gastroesophageal Varices

HNL – Histologically Normal Liver
HVPG - Hepatic Venous Pressure Gradient
INCPH – Idiopathic Non-Cirrhotic Portal Hypertension
logFC – logarithmic fold-change base 2
MSigDb – Molecular Signatures Database
PSVD – Porto- Sinusoidal Vascular Disease
RIP1 – Receptor Interacting Protein 1
SNORD94 - Small nucleolar RNA, C/D box 94
TCA – Tricarboxylic Acid cycle
TOM - Topological Overlap Matrix
WGCNA - Weighted Gene Co-expression Analysis
WHVP - Wedged Hepatic Vein Pressure

Acknowledgements
This work is supported by the funding from the European Union's Horizon 2020 research and innovation programme under the EJP RD COFUND-EJP N° 825575. The authors would also like to thank the data providers Juan Carlos Garcia-Pagan and his team for their support and helpful discussions.

References
1. De Gottardi A, Rautou P-E, Schouten J, Rubbia-Brandt L, Leebeek F, Trebicka J, et al. Porto-sinusoidal vascular disease: proposal and description of a novel entity. The lancet Gastroenterology & hepatology. 2019;4(5):399-411.
2. Seijo S, Lozano JJ, Alonso C, Miquel R, Berzigotti A, Reverter E, et al. Metabolomics as a diagnostic tool for idiopathic non-cirrhotic portal hypertension. Liver International. 2016;36(7):1051-8.
3. Hernández-Gea V, Campreciós G, Betancourt F, Pérez-Campuzano V, Seijo S, Díaz A, et al. Co-expression gene network analysis reveals novel regulatory pathways involved in porto-sinusoidal vascular disease. Journal of hepatology. 2021;75(4):924-34.
4. Du P, Kibbe WA, Lin SM. lumi: a pipeline for processing Illumina microarray. Bioinformatics. 2008;24(13):1547-8.
5. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, Smyth GK. limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic acids research. 2015;43(7):e47-e.
6. Yu G, Wang L-G, Han Y, He Q-Y. clusterProfiler: an R package for comparing biological themes among gene clusters. Omics: a journal of integrative biology. 2012;16(5):284-7.
7. Liberzon A, Birger C, Thorvaldsdóttir H, Ghandi M, Mesirov JP, Tamayo P. The molecular signatures database hallmark gene set collection. Cell systems. 2015;1(6):417-25.

8. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proceedings of the National Academy of Sciences. 2005;102(43):15545-50.
9. Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes. Nucleic acids research. 2000;28(1):27-30.
10. Kanehisa M. Toward understanding the origin and evolution of cellular organisms. Protein Science. 2019;28(11):1947-51.
11. Agrawal A, Balcı H, Hanspers K, Coort SL, Martens M, Slenter DN, et al. WikiPathways 2024: next generation pathway database. Nucleic acids research. 2024;52(D1):D679-D89.
12. Milacic M, Beavers D, Conley P, Gong C, Gillespie M, Griss J, et al. The reactome pathway knowledgebase 2024. Nucleic acids research. 2024;52(D1):D672-D8.
13. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, et al. Gene ontology: tool for the unification of biology. Nature genetics. 2000;25(1):25-9.
14. Aleksander SA, Balhoff J, Carbon S, Cherry JM, Drabkin HJ, Ebert D, et al. The gene ontology knowledgebase in 2023. Genetics. 2023;224(1):iyad031.
15. Merico D, Isserlin R, Stueker O, Emili A, Bader GD. Enrichment map: a network-based method for gene-set enrichment visualization and interpretation. PloS one. 2010;5(11):e13984.
16. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. BMC bioinformatics. 2008;9:1-13.
17. Wan X, Yao B, Ma Y, Liu Y, Tang Y, Hu J, et al. MicroRNA-128-1-5p attenuates myocardial ischemia/reperfusion injury by suppressing Gadd45g-mediated apoptotic signaling. Biochemical and Biophysical Research Communications. 2020;530(1):314-21.
18. Li X, Zhong S, Huang X, Li Y, Yang M, Yuan H-X, et al. Integration analysis identifies the role of metallothionein in the progression from hepatic steatosis to steatohepatitis. Frontiers in Endocrinology. 2022;13:951093.
19. Ogren A, Kibiryeva N, Marshall J, O'Brien Jr JE, Bittel DC. Snord94 expression level alters methylation at C62 in snRNA U6. Plos one. 2019;14(12):e0226035.
20. Marjot T, Ray DW, Williams FR, Tomlinson JW, Armstrong MJ. Sleep and liver disease: a bidirectional relationship. The Lancet Gastroenterology & Hepatology. 2021;6(10):850-63.
21. Montagnese S, De Pittà C, De Rui M, Corrias M, Turco M, Merkel C, et al. Sleep-wake abnormalities in patients with cirrhosis. Hepatology. 2014;59(2):705-12.
22. Malhi H, Kaufman RJ. Endoplasmic reticulum stress in liver disease. Journal of hepatology. 2011;54(4):795-809.
23. Ammerpohl O, Pratschke J, Schafmayer C, Haake A, Faber W, von Kampen O, et al. Distinct DNA methylation patterns in cirrhotic liver and hepatocellular carcinoma. International journal of cancer. 2012;130(6):1319-28.
24. Paakkari I, Siren A-L, Nurminen M-L, Svartström-Fraser M. Injection of thyrotropin releasing hormone into the locus coeruleus increases blood pressure. European Heart Journal. 1987;8(suppl_B):147-51.

Figure legends (see high resolution figures as pdf)

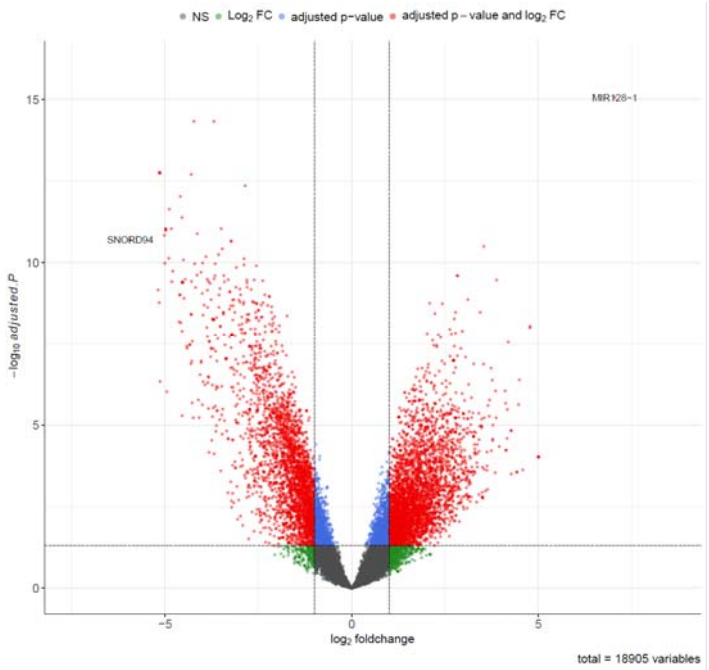


Fig. 1. Volcano plot for differential gene expression. Volcano plot of all the differentially expressed genes between PSVD and healthy samples in liver tissues. Dots represent individual genes and the color represents the not significant genes (NS) - grey, genes greater than -1 < log2 fold change (log2FC) > 1 - green, genes with significant adjusted p-value - blue and genes with both -1 < log2FC > 1 and significant adjusted p-value - red.

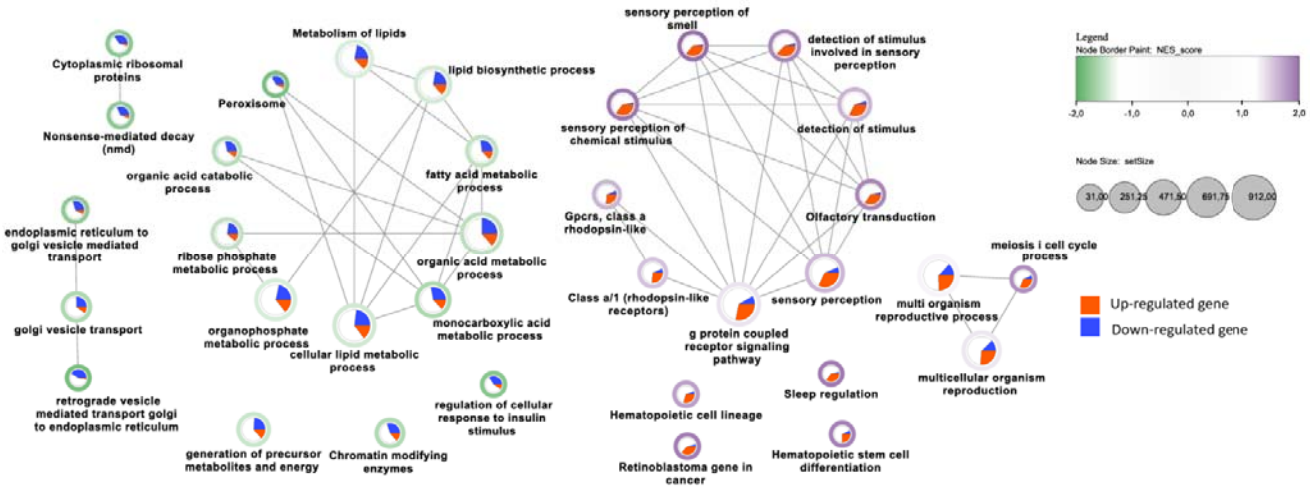


Fig. 2. Gene set enrichment analysis (GSEA) results. A node represents each enriched gene set of the gene ontology class biological process and canonical pathways with a false discovery rate cut-off of < 0.05. The node border color indicates normalized enrichment scores of the terms. The pie chart displayed within the node indicates the number of down-regulated genes (blue)

and up-regulated genes (red) out of the total genes in a gene set. The node size is assigned based on the setSize (number of genes in a term). A gene is up- regulated or down-regulated in PSVD when the fold-change vs HNL was > 1 or < 1 respectively and its adjusted p-value < 0.05. HNL, Histologically normal liver; PSVD, porto-sinusoidal vascular disease.

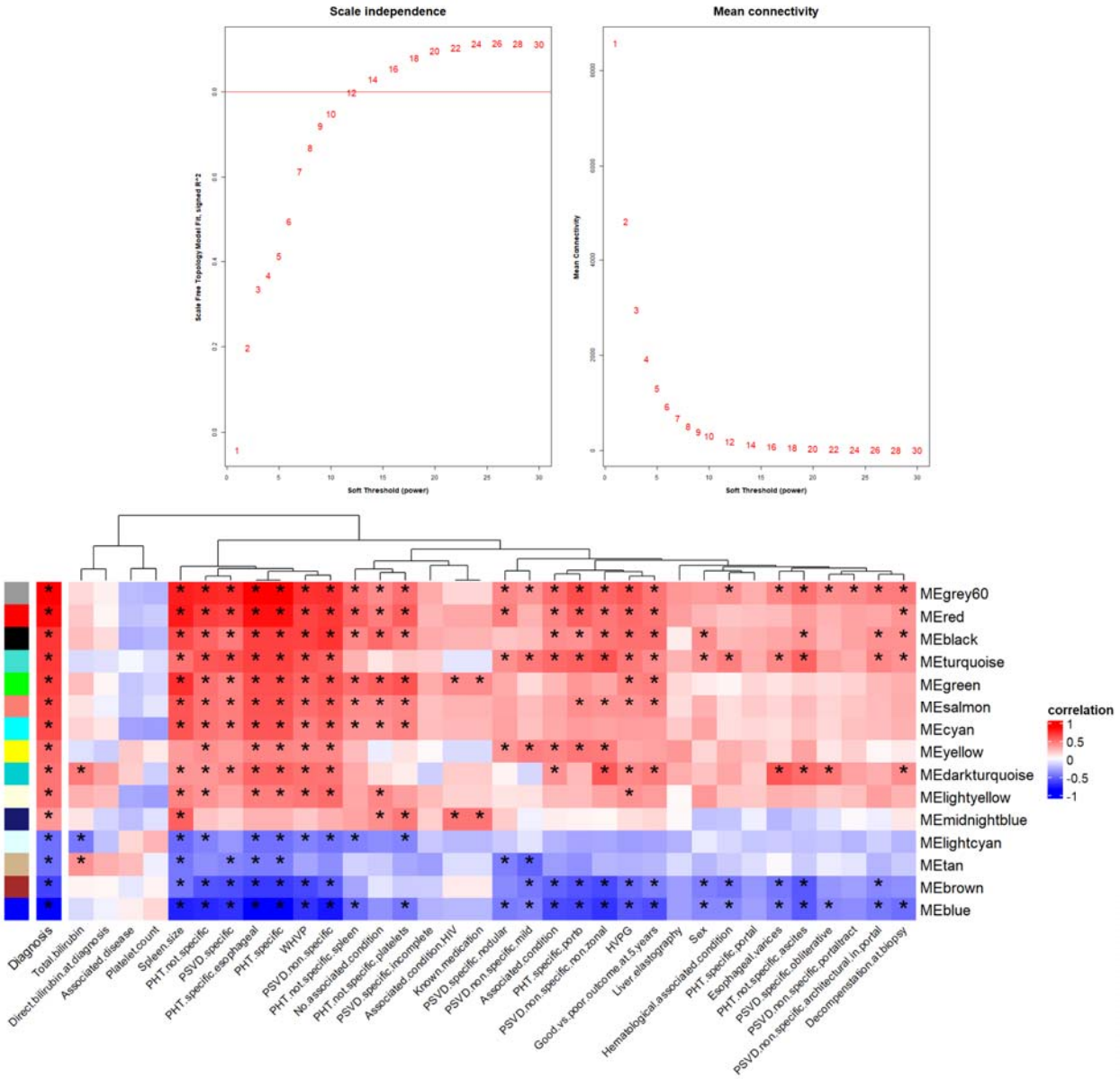


Fig. 3. Scale-free topology and module trait relationship for the coexpression network. (A) Determination of soft thresholding power for coexpression network construction: 1) Analysis of scale-free index for a range of soft thresholding values (*β*). 2) Analysis of the mean connectivity for a range of soft thresholding values (*β)*. (B) A heatmap of the module-trait relationship for modules significantly correlating to the diagnosis of PSVD on the y- axis and clinical variables on the x-axis. The color gradient on the heatmap represents the strength of the Pearson

correlation coefficients.* represents the modules significantly correlating to the respective clinical variables ($p < 0.05$). Hepatic venous pressure gradient (HVPG), wedged hepatic vein pressure (WHVP), portal hypertension (PHT), porto-sinusoidal vascular diseases (PSVD).

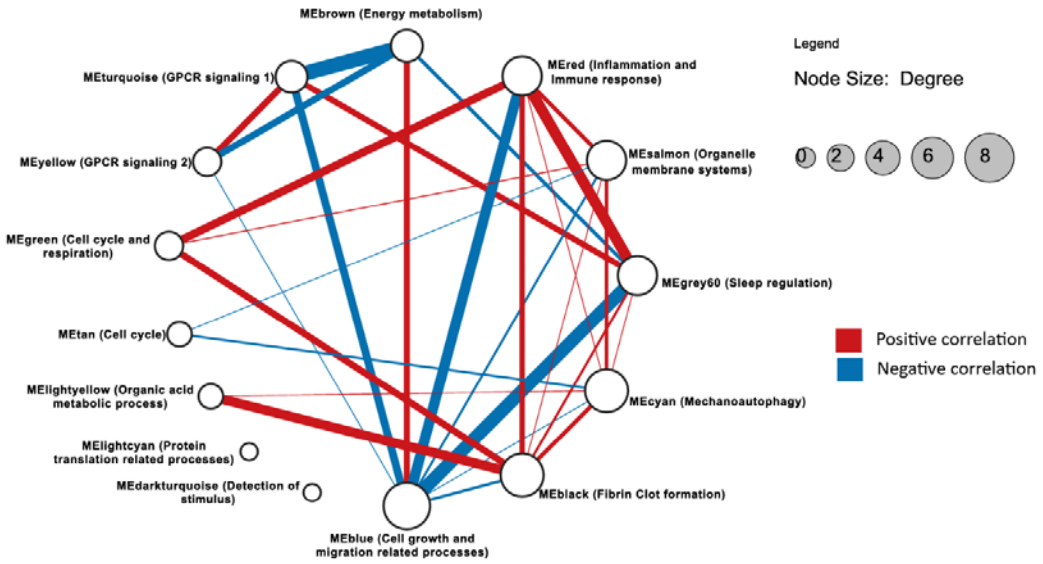


Fig. 4. Module eigengene correlation network. A network representation of the correlation of the eigengene of the modules significantly correlating to the diagnosis of PSVD. The size of the nodes is based on their degree. The color of the edges represents the direction of the Pearson correlation coefficient where red and blue represent positive and negative correlation respectively. The width of the edges is a gradient of the absolute Pearson correlation coefficient (0.6- 0.98).

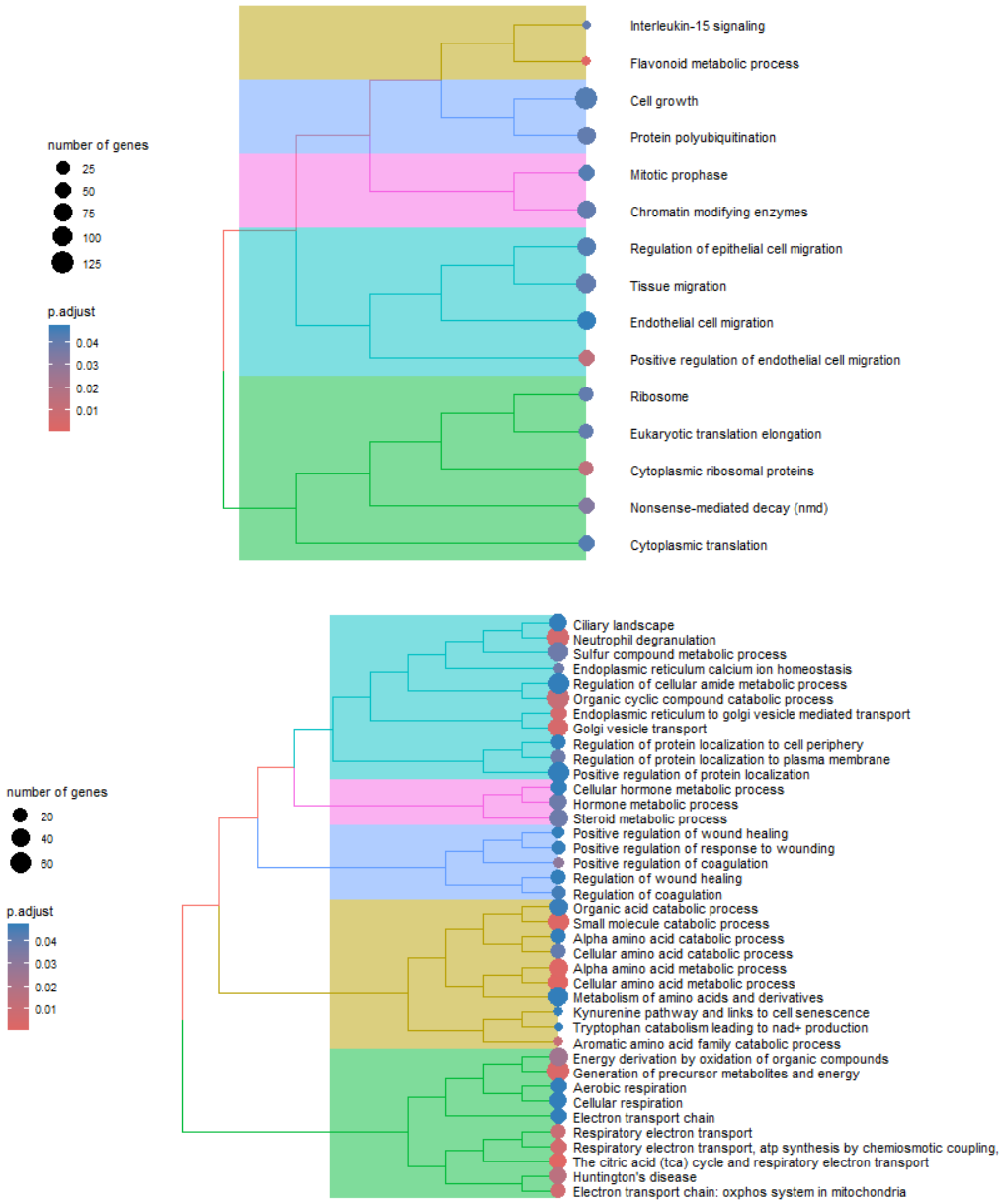
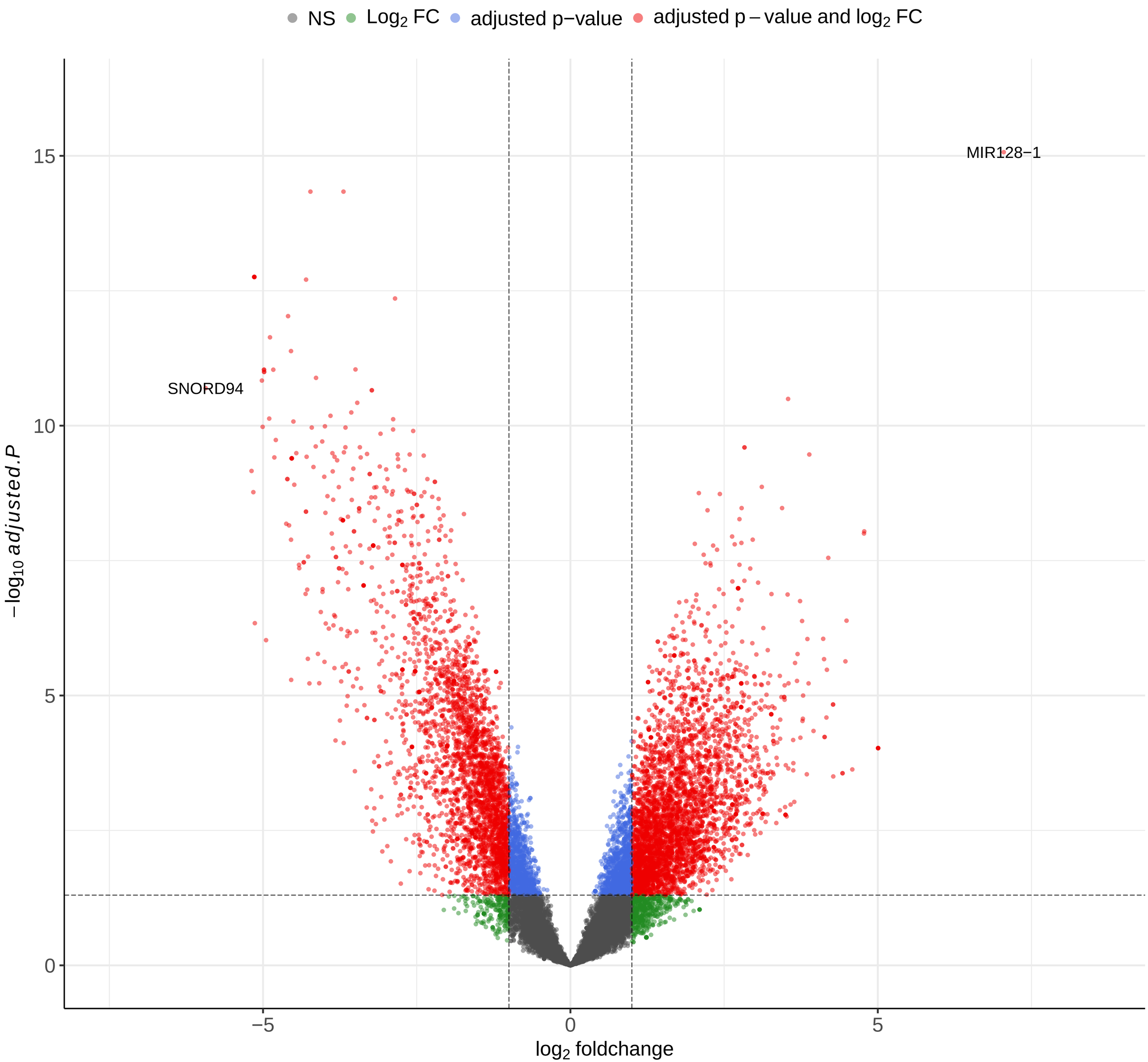


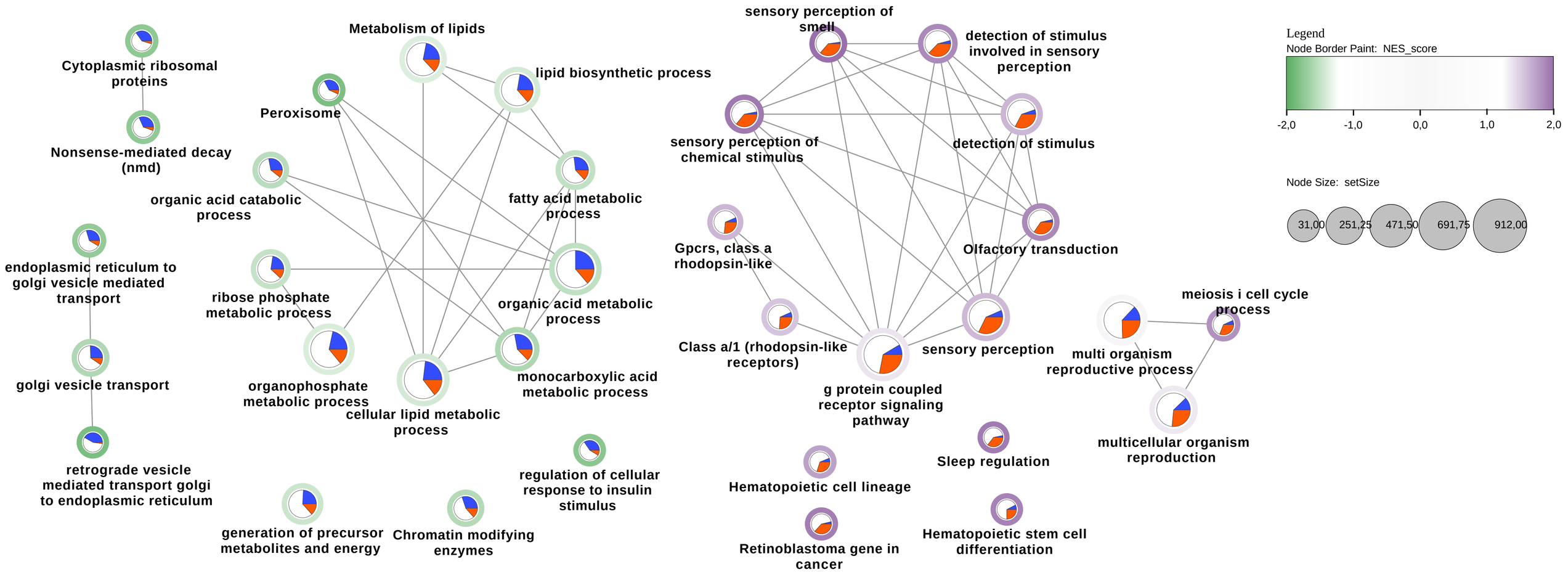
Fig. 5. Treeplot for blue and brown modules. Visualizing the functional over representation analysis of the blue (A) and brown modules (B) respectively.

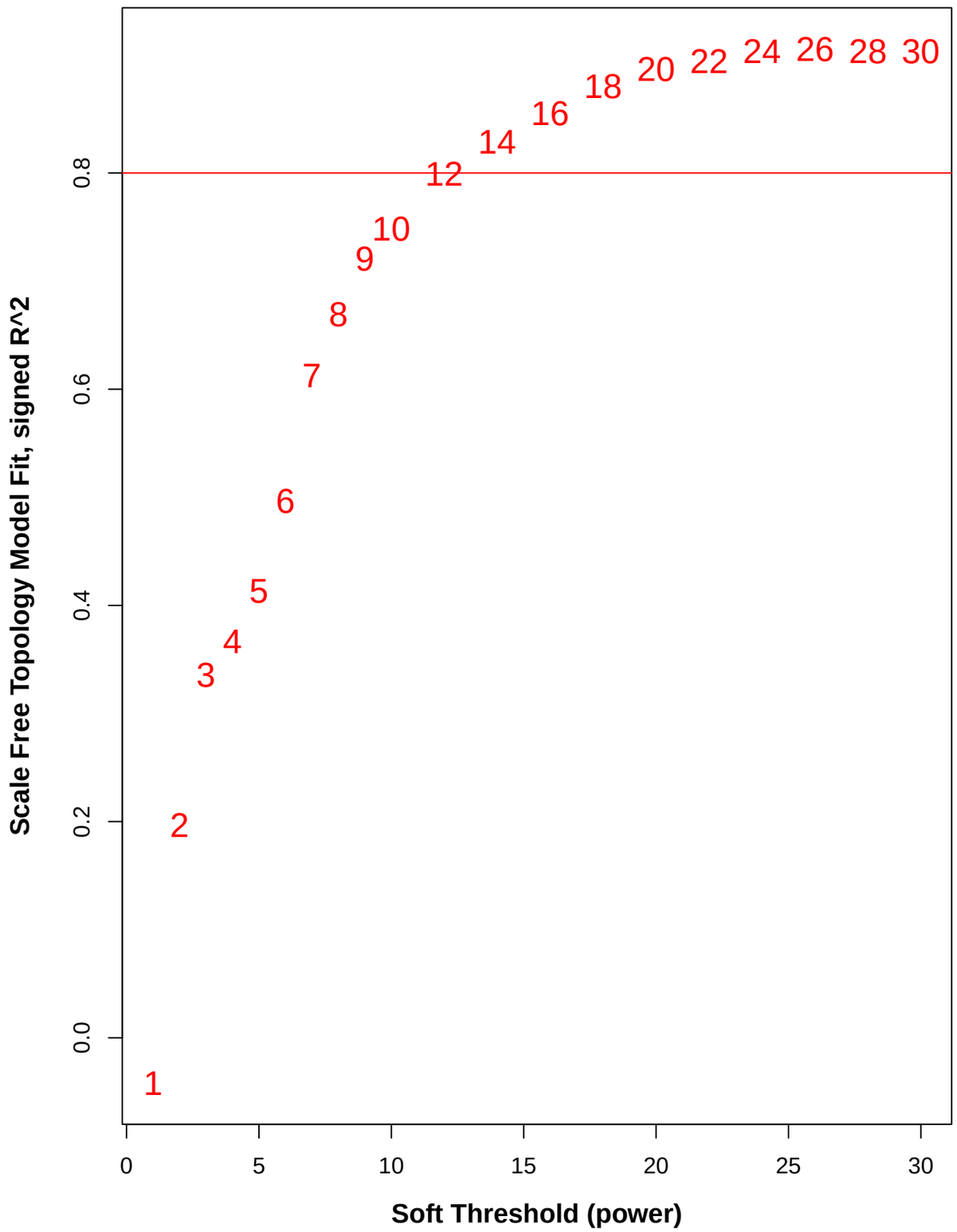
Volcano plot

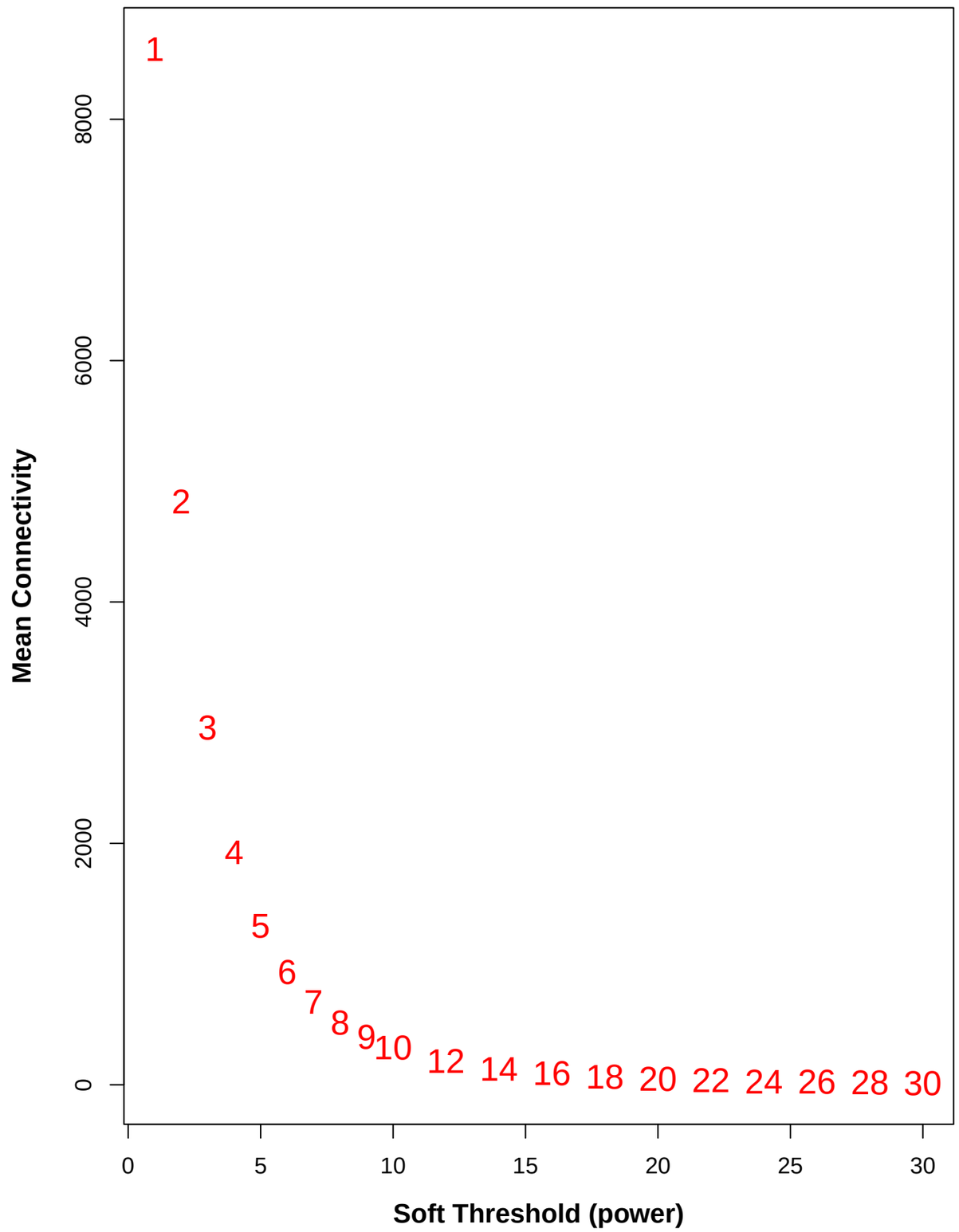
EnhancedVolcano

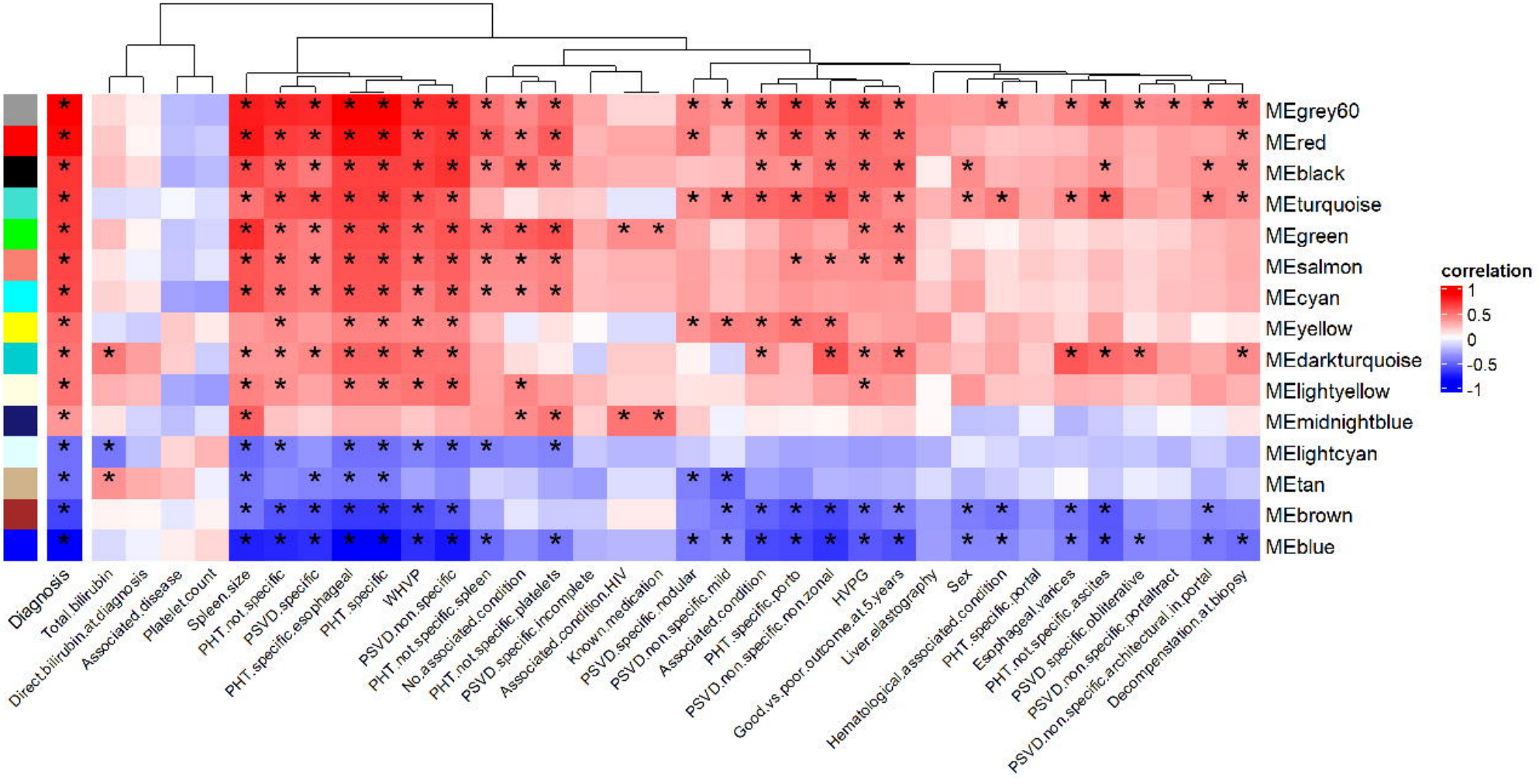


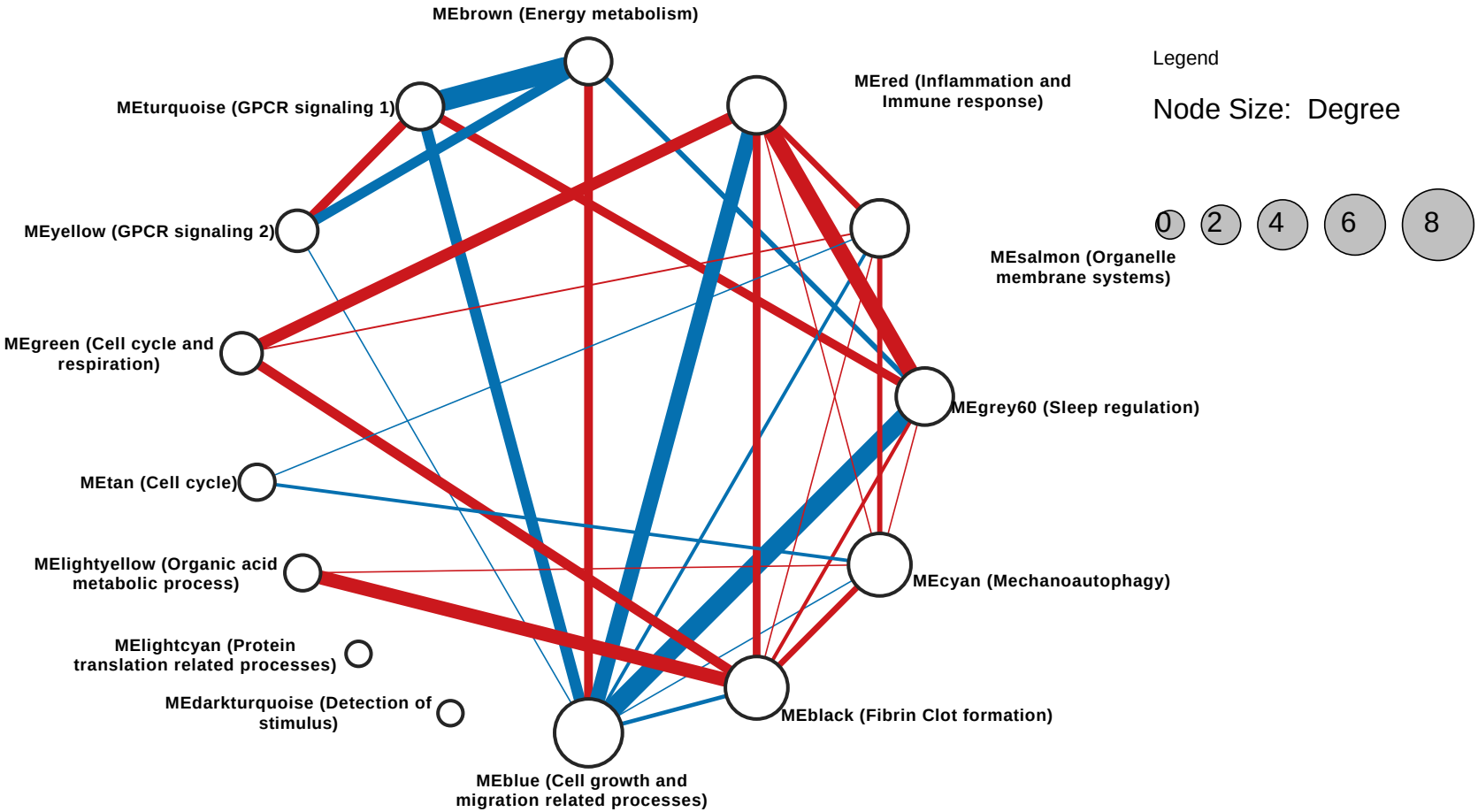
total = 18905 variables

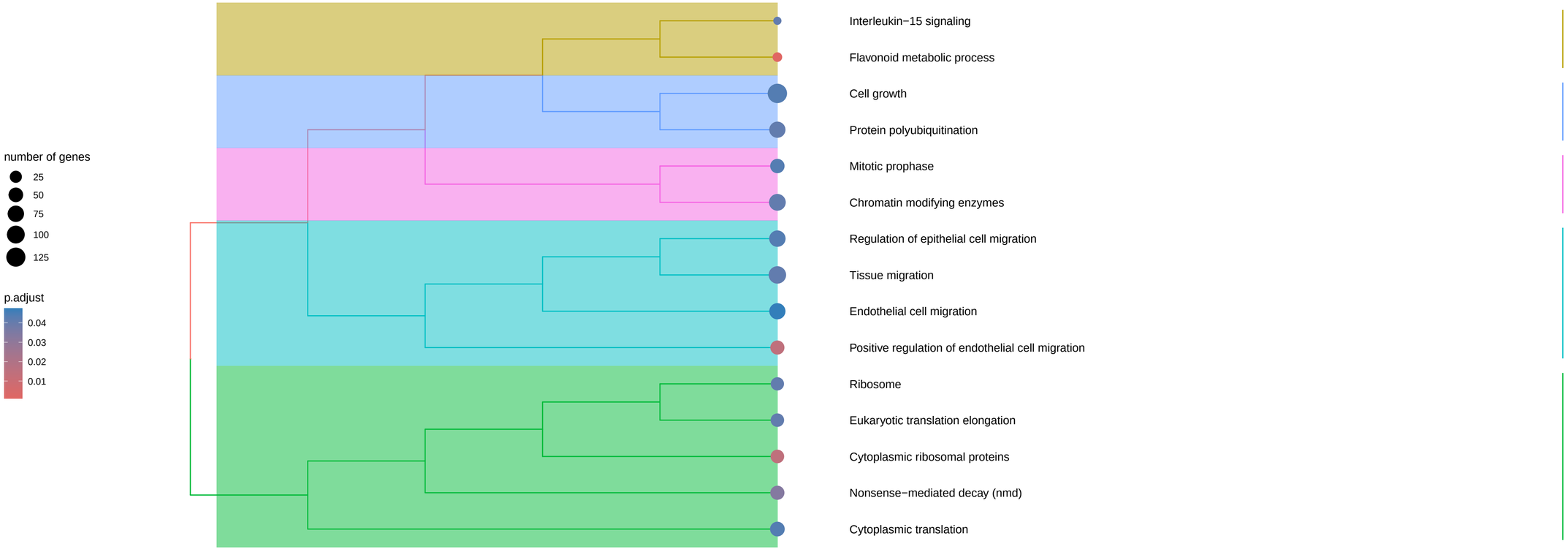

Cytoplasmic ribosomal proteins
Nonsense-mediated decay (nmd)
endoplasmic reticulum to golgi vesicle mediated transport
golgi vesicle transport
retrograde vesicle mediated transport golgi to endoplasmic reticulum
Metabolism of lipids
Peroxisome
lipid biosynthetic process
organic acid catabolic process
fatty acid metabolic process
ribose phosphate metabolic process
organic acid metabolic process
organophosphate metabolic process
monocarboxylic acid metabolic process
cellular lipid metabolic process
regulation of cellular response to insulin stimulus
generation of precursor metabolites and energy
Chromatin modifying enzymes
sensory perception of smell
detection of stimulus involved in sensory perception
sensory perception of chemical stimulus
detection of stimulus
Gpcrs, class a rhodopsin-like
Olfactory transduction
Class a/1 (rhodopsin-like receptors)
sensory perception
g protein coupled receptor signaling pathway
meiosis i cell cycle process
multi organism reproductive process
multicellular organism reproduction
Sleep regulation
Hematopoietic cell lineage
Hematopoietic stem cell differentiation
Retinoblastoma gene in cancer
Legend
Node Border Paint: NES_score
-2,0
-1,0
0,0
1,0
2,0
Node Size: setSize
31,00
251,25
471,50
691,75
912,00

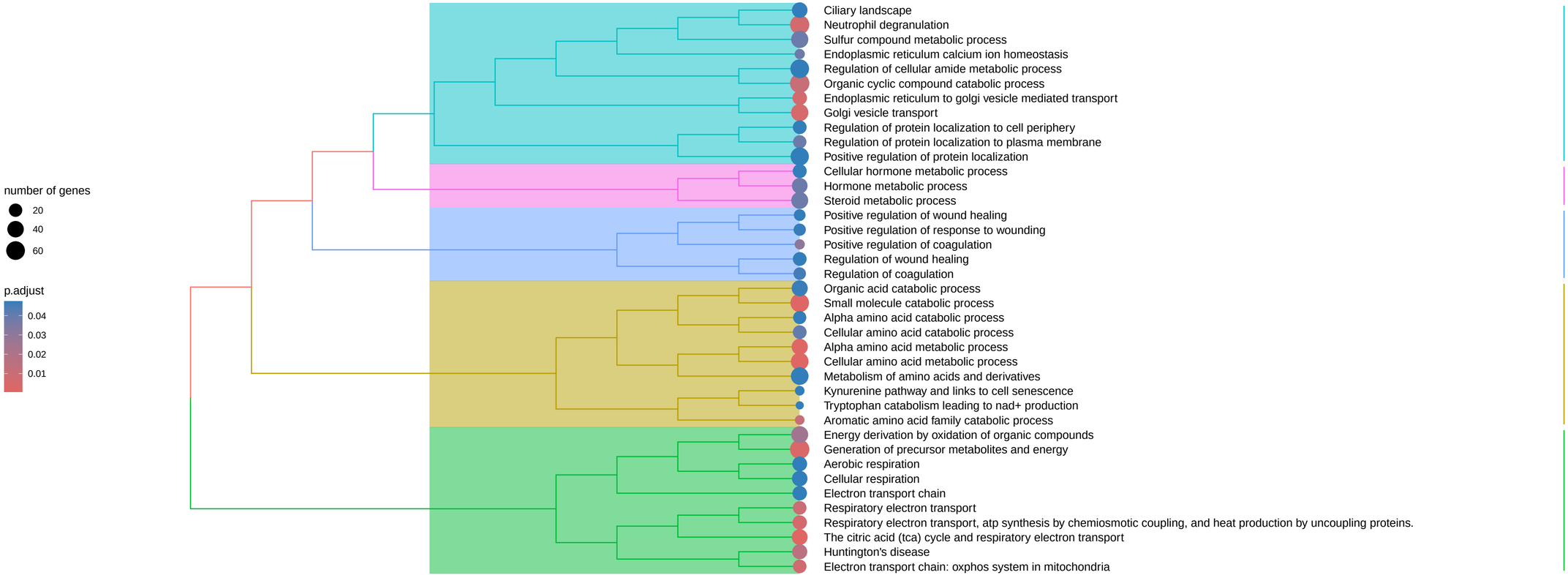

Scale independence
Scale Free Topology Model Fit, signed R^2
Soft Threshold (power)
0.0
0.2
0.4
0.6
0.8
0
5
10
15
20
25
30
1
2
3
4
5
6
7
8
9
10
12
14
16
18
20
22
24
26
28
30


Mean connectivity
Mean Connectivity
Soft Threshold (power)
0
2000
4000
6000
8000
0
5
10
15
20
25
30
1
2
3
4
5
6
7
8
9
10
12
14
16
18
20
22
24
26
28
30


MEgrey60
MEred
MEblack
MEturquoise
MEgreen
MEsalmon
MEcyan
MEyellow
MEdarkturquoise
MElightyellow
MEmidnightblue
MElightcyan
MEtan
MEbrown
MEblue
correlation
1
0.5
0
-0.5
-1
Diagnosis
Total.bilirubin
Direct.bilirubin.at.diagnosis
Associated.disease
Platelet.count
Spleen.size
PHT.not.specific
PSVD.specific
PHT.specific.esophageal
PHT.specific
WHVP
PSVD.non.specific
PHT.not.specific.spleen
No.associated.condition
PHT.not.specific.platelets
PSVD.specific.incomplete
Associated.condition.HIV
Known.medication
PSVD.specific.nodular
PSVD.non.specific.mild
Associated.condition
PHT.specific.porto
PSVD.non.specific.non.zonal
HVPG
Good.vs.poor.outcome.at.5.years
Liver.elastography
Hematological.associated.condition
Sex
PHT.specific.portal
Esophageal.varices
PHT.not.specific.ascites
PSVD.specific.obliterative
PSVD.non.specific.portaltract
PSVD.non.specific.architectural.in.portal
Decompensation.at.biopsy


MEbrown (Energy metabolism)
MEturquoise (GPCR signaling 1)
MEred (Inflammation and Immune response)
MEyellow (GPCR signaling 2)
MEsalmon (Organelle membrane systems)
MEgreen (Cell cycle and respiration)
MEgrey60 (Sleep regulation)
MEtan (Cell cycle)
MElightyellow (Organic acid metabolic process)
MEcyan (Mechanoautophagy)
MElightcyan (Protein translation related processes)
MEdarkturquoise (Detection of stimulus)
MEblack (Fibrin Clot formation)
MEblue (Cell growth and migration related processes)
Legend
Node Size: Degree
0
2
4
6
8


number of genes
25
50
75
100
125
p.adjust
0.04
0.03
0.02
0.01
Interleukin−15 signaling
Flavonoid metabolic process
Cell growth
Protein polyubiquitination
Mitotic prophase
Chromatin modifying enzymes
Regulation of epithelial cell migration
Tissue migration
Endothelial cell migration
Positive regulation of endothelial cell migration
Ribosome
Eukaryotic translation elongation
Cytoplasmic ribosomal proteins
Nonsense−mediated decay (nmd)
Cytoplasmic translation


number of genes
20
40
60
p.adjust
0.04
0.03
0.02
0.01
Ciliary landscape
Neutrophil degranulation
Sulfur compound metabolic process
Endoplasmic reticulum calcium ion homeostasis
Regulation of cellular amide metabolic process
Organic cyclic compound catabolic process
Endoplasmic reticulum to golgi vesicle mediated transport
Golgi vesicle transport
Regulation of protein localization to cell periphery
Regulation of protein localization to plasma membrane
Positive regulation of protein localization
Cellular hormone metabolic process
Hormone metabolic process
Steroid metabolic process
Positive regulation of wound healing
Positive regulation of response to wounding
Positive regulation of coagulation
Regulation of wound healing
Regulation of coagulation
Organic acid catabolic process
Small molecule catabolic process
Alpha amino acid catabolic process
Cellular amino acid catabolic process
Alpha amino acid metabolic process
Cellular amino acid metabolic process
Metabolism of amino acids and derivatives
Kynurenine pathway and links to cell senescence
Tryptophan catabolism leading to nad+ production
Aromatic amino acid family catabolic process
Energy derivation by oxidation of organic compounds
Generation of precursor metabolites and energy
Aerobic respiration
Cellular respiration
Electron transport chain
Respiratory electron transport
Respiratory electron transport, atp synthesis by chemiosmotic coupling, and heat production by uncoupling proteins.
The citric acid (tca) cycle and respiratory electron transport
Huntington's disease
Electron transport chain: oxphos system in mitochondria