bioRxiv preprint doi: https://doi.org/10.1101/2024.06.14.599028; this version posted October 21, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

GSEA and the coexpression network approach identify novel pathway connections of molecular processes affected in Porto-sinusoidal vascular disease

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

1 Department of Translational Genomics NUTRIM/MHeNs, Maastricht University, Maastricht, The Netherlands
2 Medical BioSciences Department, Radboud University, Nijmegen, The Netherlands
3 Maastricht Centre for Systems Biology and Bioinformatics(MaCSBio), Maastricht University, Maastricht, The Netherlands

* friederike.ehrhart@maastrichtuniversity.nl

1 Abstract

1.1 Background

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 by PSVD, facilitating improvements in the prognosis, diagnosis, and treatment options for these patients.

1.2 Methods

We applied gene set enrichment analysis (GSEA) and weighted gene coexpression network analysis (WGCNA) to identify pathways dysregulated in PSVD. Network construction and visualization were performed in Cytoscape to explore interconnectivity among enriched processes. Within key modules, candidate genes were prioritized by ranking approaches and cross-referenced with findings from previous studies.

1.3 Results and Conclusion

This study revealed that PSVD is characterized by coordinated dysregulation, with immune and signaling pathways activated alongside the suppression of metabolic, ribosomal, and mitochondrial programs. Alterations in ribosomal proteins, ATP synthase subunits, and serpin family members highlight translational, bioenergetic, and anticoagulant dysfunction as core mechanisms. Together, these findings define PSVD as a disorder of integrated immune, vascular, and metabolic imbalance.

2 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 [1]. This term was recently coined to improve the understanding of the disease by reducing the effect of heterogeneity, facilitating improved diagnosis, and

bioRxiv preprint doi: https://doi.org/10.1101/2024.06.14.599028; this version posted October 21, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

simplifying comparisons between different clinical studies [1]. PSVD is a rare disease with a currently unknown prevalence.

The diagnosis of patients suffering from PSVD includes noninvasive imaging methods focused on splenomegaly,rtosystemic colcollaterals,nd hepatic vein venography. However, imaging by itself is insufficient, and further invasive methods, such as biopsies, are an essential part of the diagnostic routines for PSVD [1]. The accuracy of analysis remains highly variable depending on the experience of the histopathologist [1]. In a recently published metabolomics study, a group of metabolite markers was identified that could 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 prothrombotic 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 mechanisms by identifying genes that are coexpressed or change [4, 5].

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 for improved prognosis, diagnosis, and the treatment options available to the patients suffering from this rare disorder.

3 Materials and methods

3.1 Data

A previously published transcriptomics dataset by Lozano et al. was obtained from the GEO database (GEO:GSE77627) [3], including their 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 transcriptomic profile overlapped with the PSVD transcriptomic profile (see supplementary Fig. **Principal component analysis (PCA) for healthy, liver cirrhosis, and PSVD liver biopsy samples.** PCA was performed on normalized transformed gene expression data. The PCA plot above represents the variance explained by the top two components). Additional clinical data and information were obtained from the original study authors, Hospital Clinic of the University of 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.

bioRxiv preprint doi: https://doi.org/10.1101/2024.06.14.599028; this version posted October 21, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

3.2 Data pre-processing

The raw Illumina probe data using the Illumina HumanHT-12 DASL 4.0 R2 expression beadchip platform annotation was first filtered for protein-coding genes using the biomaRt(v2.64.0) R package with the filters: biotype (protein coding), chromosome name (22 chromosomes, mitochondrial chromosome and sex chromosomes) using the gene identifiers provided in the annotation file. Next, the filtered probe data was pre-processed using the lumi(v) R package [6]. Background correction and quantile normalisation was performed using the *neqc* function with an offset value of 16. The data was re-annotated using ENSEMBL gene identifiers. A misdiagnosed patient (PSVD17) was removed from the analysis. Samples with incomplete sex information were removed from the analysis. The data distribution for before and after normalized data is provided in supplementary Fig. **Distribution of the gene expression data across samples.** (A)Distribution of the gene expression data before and after normalization. X-axis represents the samples (pink – Healthy liver biopsies and blue- PSVD liver biopsies). Y-axis represents the genes expression values in logarithmic scale. (B) Distribution of the detection p-value before and after normalization. X-axis represents the samples (pink – Healthy liver biopsies and blue- PSVD liver biopsies). Y-axis represents the p-values of the probes used to measure the gene expression of the samples.

To detect outliers, hierarchical clustering on the samples was performed and the dendrogram is provided in supplementary Fig. **Hierarchical clustering of the samples.** (a) Dendogram of the sample clustering. Sample PSVD05 (shown in red) was removed from the analysis given that it was clustering with the healthy liver samples. (b) Dendogram representing the sample clustering after outlier removal against the clinical variables visualized in the rows(a). Based on the clustering, sample 'PSVD05' was removed from further analysis. The dendrogram of samples and clinical variables measured for these samples is provided in supplementary Fig. **Hierarchical clustering of the samples.** (a) Dendogram of the sample clustering. Sample PSVD05 (shown in red) was removed from the analysis given that it was clustering with the healthy liver samples. (b) Dendogram representing the sample clustering after outlier removal against the clinical variables visualized in the rows(b).

3.3 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 controls after sex correction using the limma(v3.64.1) R package [7]. The cut-off for significantly upregulated genes is logFC > 1 and adjusted p-value < 0.05, significantly weak upregulated genes is $0.58 <$ 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. A cut-off of $0.58 <$ logFC > 1 and adjusted p-value > 0.05 is used for significantly downregulated genes. The *EnhancedVolcano* function in EnhancedVolcano(v1.26.0) R package was used for creating the volcano plot for differentially regulated genes identified in PSVD vs HNL comparison [8].

3.4 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 500 and 10 respectively [9, 10]. The DEGs were ranked based on the product of signed log fold change and the negative logarithm of the adjusted p-value, see the equation below.

bioRxiv preprint doi: https://doi.org/10.1101/2024.06.14.599028; this version posted October 21, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

$$ranking = log_2FC * -log_{10}adjustedp - value \quad (1)$$

For the enrichment analysis, the human canonical pathway gene sets from the Molecular Signatures Database (MSigDB, v2023.2.Hs) were used ([10, 11]). The pathway genesets from Kyoto Encyclopedia of Genes and Genomes (KEGG, 186 gene sets ([12, 13])), WikiPathways (733 gene sets ([14])) and Reactome (1,654 gene sets ([15])) were included. Additionally, 7,751 gene sets from the Biological Process ontology from Gene Ontology (GO) were included for the analysis ([16, 17]).

3.5 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. [18].

$$k = \frac{|A \cap B|}{|min(A, B)|} \quad (2)$$

A cut-off score of k > 0.4 was used to add an edge between two enriched terms.

3.6 Coexpression 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(v1.73) R package to identify gene expression modules correlating to the PSVD phenotype [19].

Normalized gene expression data was adjusted for sex effects using the *removeBatchEffect* function from the limma(v) R package [7]. Next, 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 ($\beta = 18$).

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 (see Supplementary Fig. **Cluster dendrogram and modules detected.** Hierarchical tree (average linkage) using dynamic tree cutting method was used for module detection. The Dynamic Tree Cut band represents the genes assigned to particular modules. 24 modules were detected. To identify modules with highly interconnected genes, a dynamic tree-cut method was implemented with a minimum module size of 150.

bioRxiv preprint doi: https://doi.org/10.1101/2024.06.14.599028; this version posted October 21, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

3.7 Identification of key clinically significant modules and core genes for the key 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, a significant correlation between the modules and the clinical phenotype 'Diagnosis of PSVD' was used to identify clinically significant gene modules. Core genes from significant modules were retained based on thresholds of gene significance > 0.5 and $|\text{module membership}| > 0.5$.

3.8 Over-representation analysis (ORA) and functional annotation for key modules

Functional analysis of the core genes for the key clinically significant modules was performed using the *enricher* function in the clusterProfiler R package [9]. The Gene Ontology: Biological Process geneset for performing the over-representation analysis was obtained from the Molecular Signatures Database (MSigDB, v2023.2.Hs)([10,11,16,17]). For certain modules where enriched terms were not obtained using the above gene set, WikiPathways and the Reactome gene sets were used for performing enrichment analysis. The key modules were then manually functionally annotated by assigning an appropriate biological term based on the enriched terms for each module.

3.9 Module Eigengene Correlation Network

Pearson correlation using the *cor* function from the stats(v3.6.2) R package was implemented [20]. The Pearson correlation of the module eigengenes obtained from coexpression network analysis was calculated for the 14 key modules. The network was then exported to Cytoscape and the key modules (as nodes in the network) were annotated using the results from Section 3.8.

3.10 Cytoscape network analysis of the PPI network for the module clusters

Module clusters identified from Section 4.3 were then used for generating PPI networks. Core genes from the immune (grey60 and red) and signaling (yellow and turquoise) module clusters identified in section 4.3 were exported to Cytoscape using the Ensembl identifiers with a STRING confidence score of 0.7 [21]. For the metabolic module cluster, the core genes from black, blue, and lightgreen modules were exported to Cytoscape using Ensembl identifiers with a STRING confidence score of 0.9 [21]. Next, Glay community detection algorithm using the Cluster Network option in the clusterMaker app in Cytoscape was implemented to detect clusters within the PPI networks [22, 23]. Clusters with less than 20 nodes were removed from the network. The top five genes in the network were selected based on the ranking measure of the product of the absolute log fold change and the node degree.

3.11 Open source code

All analysis steps described were fully automated, and the scripts used for this study are available on WorkflowHub for reproduction and further exploration: `https://workflowhub.eu/workflows/1040?version=1`.

bioRxiv preprint doi: https://doi.org/10.1101/2024.06.14.599028; this version posted October 21, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

4 Results

4.1 Differentially expressed genes in PSVD

The raw transcriptomics data consisting of 26,776 Illumina probes was processed to correspond to 15,551 protein-coding genes (annotated with ENSEMBL gene identifiers). DEG analysis identified 3,152 significantly upregulated genes, 803 significantly and weakly upregulated genes, 2,412 significantly downregulated genes, and 788 significantly weak downregulated genes in PSVD patients.

The top 3 up-regulated genes were erythropoietin (EPO) - logFC 4.8, Ankyrin Repeat Domain 1 (ANKRD1) - logFC 4.8 and G Antigen 12J (GAGE12J) - logFC 4.5 while the top 3 down-regulated genes were ephrin A2 (EFNA 2) logFC −5.1, Nuclear Factor I C (NFIC)- logFC −5.0 and meteorin, glial cell differentiation regulator (METRN) - logFC −4.9 (Fig. 1).

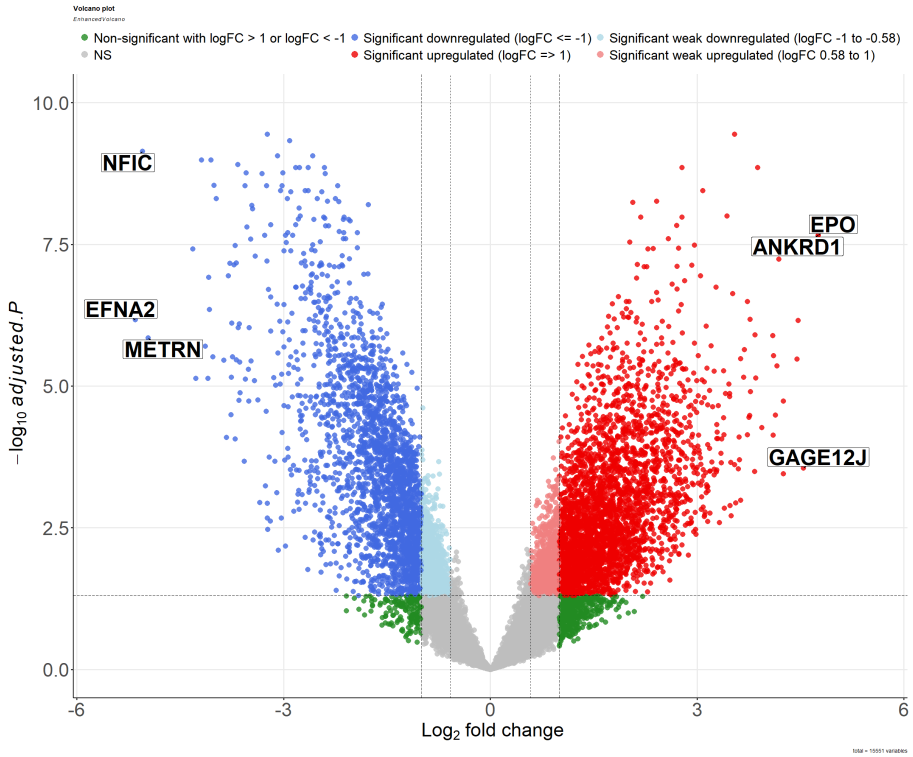

Fig 1. Volcano plot for differential gene expression. Volcano plot of all the differentially expressed genes between PSVD and healthy samples in liver tissues. Dots represent individual genes, and the color represents the not significant genes (NS) - grey, Non-significant genes with -1 < logFC > 1 and adjusted p- value > 0.05 - green , Significant upregulated genes with logFC => 1 and adjusted p-value < 0.05 - red, Significant weak upregulated gene with 0.58 < logFC > 1 and adjusted p-value > 0.05 - light red, Significant downregulated genes with logFC <= −1 and adjusted p-value < 0.05 - red. The top six differentially regulated genes are highlighted in the figure.

4.2 Gene set enrichment analysis

Gene set enrichment analysis (GSEA) was performed on the differentially expressed genes in PSVD. Using the Gene Ontology, KEGG, Reactome and WikiPathways gene

bioRxiv preprint doi: https://doi.org/10.1101/2024.06.14.599028; this version posted October 21, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

sets, 69 significantly enriched terms were obtained, in which 9 positively enriched terms and 60 negatively enriched terms were visualised in Cytoscape (Fig. 2). The top positively significantly enriched term was the Olfactory signaling pathway (Reactome) with a normalised enrichment score (NES) of 2 while the Peroxisome pathway (KEGG) was the top negatively significantly enriched term with a NES of −1.9. A similarity threshold of 0.4 was applied to cluster the enriched terms based on the number of genes common between the terms, thereby sharing similar functional profiles.

From Fig. 2 we observe that the lipid and fatty acid metabolism processes, insulin-related processes, tissue and epithelial cell migration processes, endoplasmic reticulum and Golgi associated processes, organic compounds and tricarboxylic acid metabolic process, organic hydroxy compound metabolic process, monosaccharide metabolic process, glycerolipid metabolic process, nucleoside biphosphate metabolic process, immune-related processes like neutrophil degranulation, viral process, SARS-CoV infection, and protein translation were negatively enriched. Processes related to olfactory and sensory stimulus, sleep regulation, meiosis cell cycle and retinoblastoma were positively enriched.

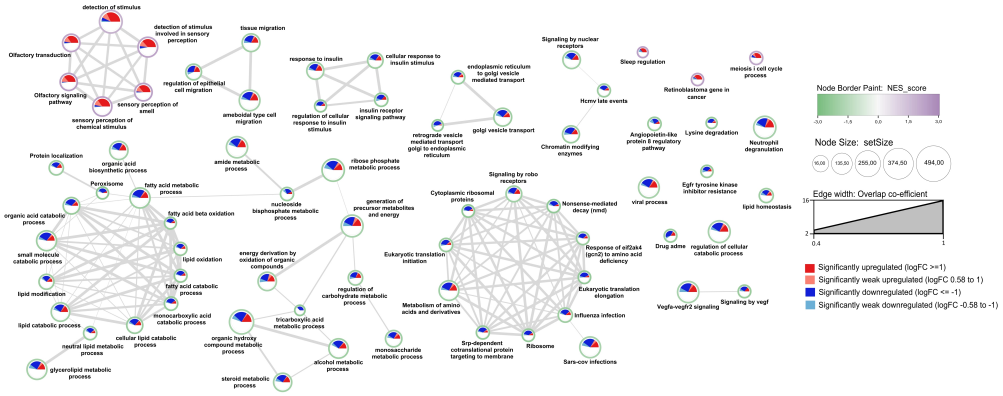

Fig 2. Gene set enrichment analysis (GSEA) results. A node represents each enriched gene set of the Gene Ontology class Biological Process and canonical pathways with a false discovery rate cut-off of < 0.05. The node border color indicates normalized enrichment scores of the terms. The pie chart displayed within the node indicates the number of significantly downregulated and weak downregulated genes (darkblue and lightblue respectively) and significantly upregulated and weak upregulated genes (red and light red respectively) out of the total genes in a gene set. The node size is assigned based on the setSize (number of genes in a term). The edge weight representing the overlap coefficient (similarity index between two terms) has filtered with a cut-off value of 0.4.

4.3 Identification of 15 key PSVD modules using co-expression network analysis

The co-expression network was constructed using 15,551 protein-coding genes from 27 liver biopsy samples (11 healthy and 16 PSVD patients) using the '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).

bioRxiv preprint doi: https://doi.org/10.1101/2024.06.14.599028; this version posted October 21, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

A range of soft-thresholding powers (β) were used to assess the scale free topology of the network constructed. For this analysis, β of 18 was selected which had a scale-free topology fit (R^2) of 0.81; shown in Fig. 3(a). Using the average hierarchical clustering and dynamic tree cut method, a total of 35 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 35 distinctly identified modules, 15 modules were selected which significantly correlated to the diagnosis of PSVD, shown in Fig. 3(b) (see Supplementary Fig. **Module-trait relationship heatmap.** 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. Number in each cell is the correlation and the p-value (in brackets). Hepatic venous pressure gradient (HVPG), wedged hepatic vein pressure (WHVP), portal hypertension (PHT), porto-sinusoidal vascular diseases (PSVD) for module-trait relationship for all 35 modules).

bioRxiv preprint doi: https://doi.org/10.1101/2024.06.14.599028; this version posted October 21, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

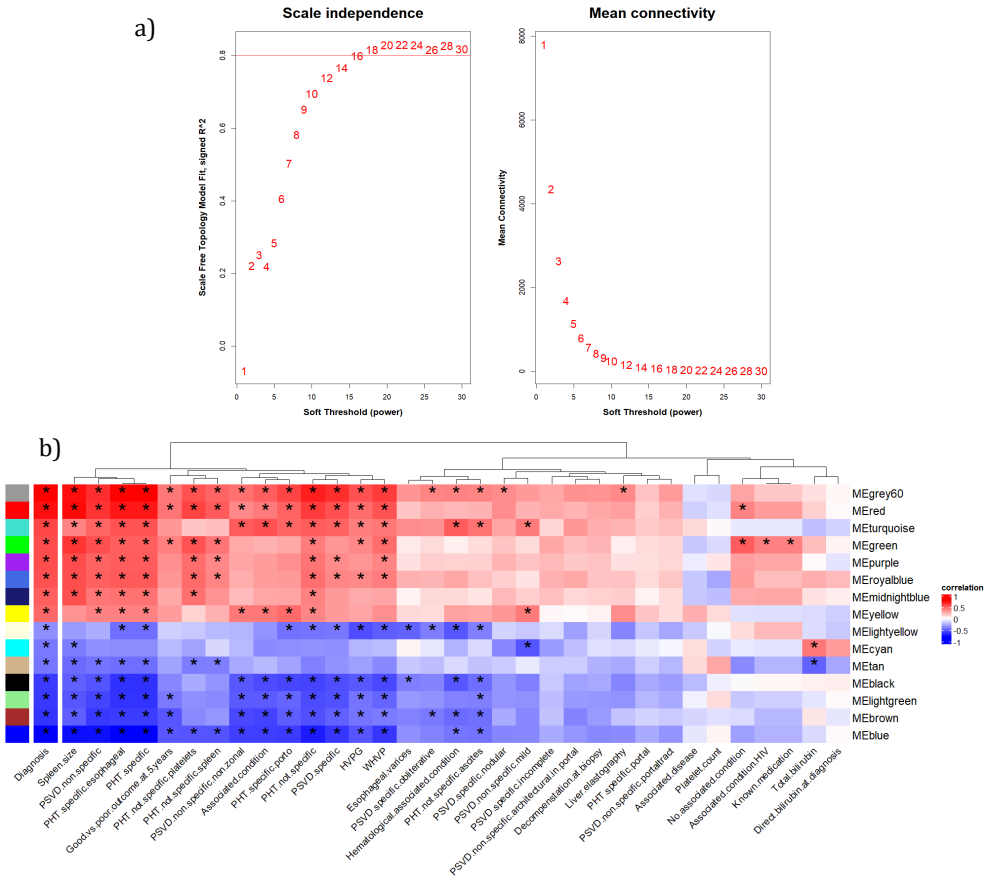

Fig 3. Scale-free topology and module trait relationship for the coexpression network. (a) Determination of soft thresholding power for coexpression network construction: 1) Analysis of scale-free index for a range of soft thresholding values (β). 2) Analysis of the mean connectivity for a range of soft 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).

4.4 Selection of Core Genes in PSVD-Associated Modules

Core genes for each module were selected based on a threshold of 0.5 for both Gene Significance (GS) and Module Membership (MM). Scatter plots of GS versus MM for each module are shown in Supplementary Fig. The scatter plots depict the relationship between gene significance (GS) for PSVD diagnosis and module membership (MM) within each key co-expression module identified by WGCNA. Each point represents a single gene, where GS reflects the correlation between gene expression and PSVD diagnosis, and MM represents the correlation between the gene and the module eigengene, indicating its connectivity within the module. Genes in the upper-right quadrant ($GS > 0.5$ and $MM > 0.5$) were designated as core (hub) genes, as they are both strongly associated with the trait and highly central within the module network.

bioRxiv preprint doi: https://doi.org/10.1101/2024.06.14.599028; this version posted October 21, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Table 1. Clinical characteristics

	HNL (n=11)	PSVD (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 ($10^9/l$)	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

$Mean \pm SD$

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

Data points are color-coded to reflect differential expression in PSVD: significantly upregulated (red), weakly upregulated (light red), significantly downregulated (blue), weakly downregulated (light blue), and non-significant genes (grey). The lightyellow module was excluded from downstream analyses primarily because none of the genes with the module met the 0.5 cut-off for GS, and additionally, the module showed a low correlation with PSVD diagnosis ($r = -0.40$). Fig. 4 illustrates the distribution of genes and core genes across modules, as well as their expression profiles.

bioRxiv preprint doi: https://doi.org/10.1101/2024.06.14.599028; this version posted October 21, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

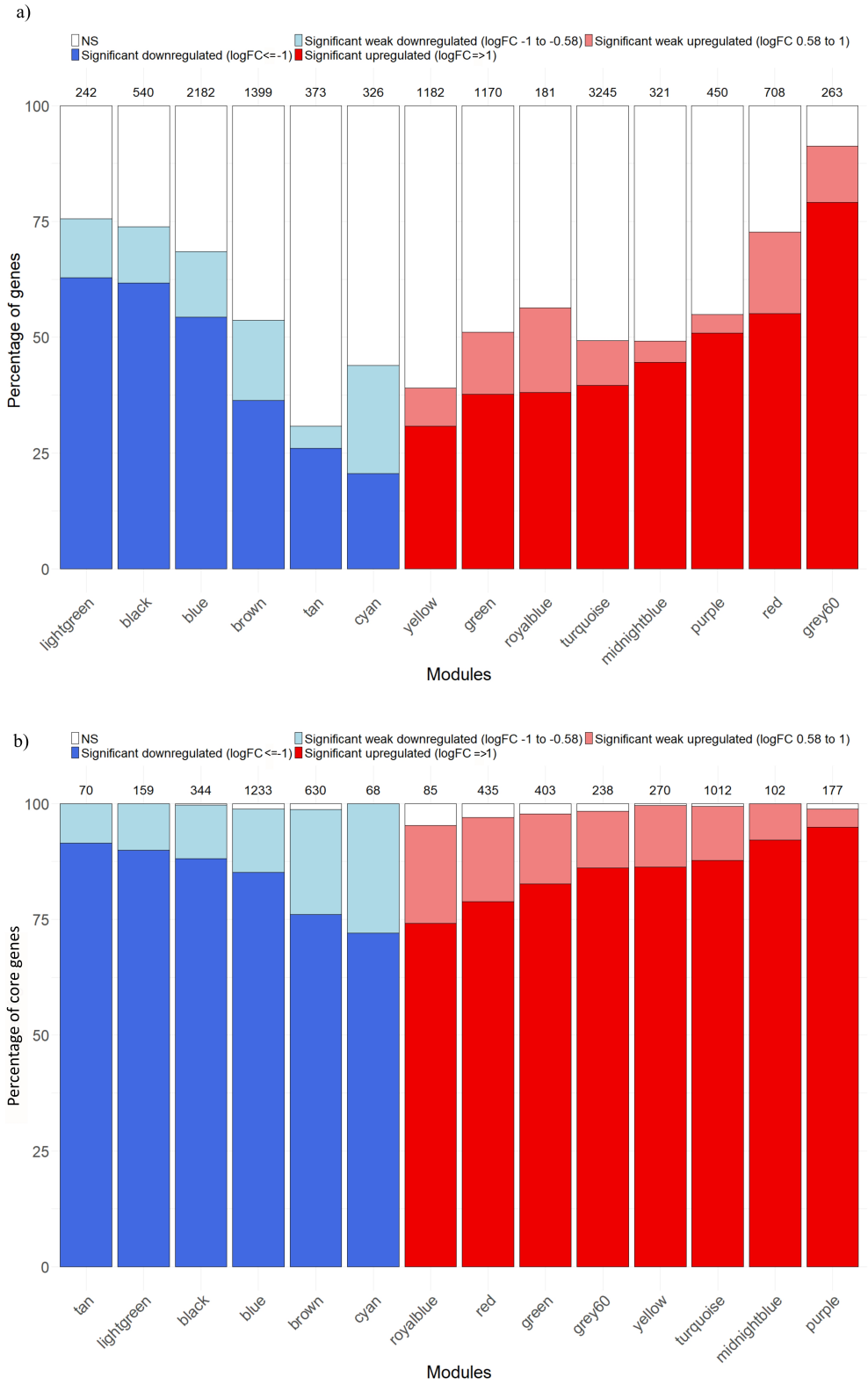

Fig 4. Barplots visualizing the distribution of differentially expressed genes within key modules significantly correlated with the diagnosis of PSVD patients. a) Percentage distribution of all genes in the key modules. The x-axis represents modules, with the total number of genes per module indicated above each bar. The y-axis shows the proportion of genes that are significantly downregulated (dark blue, $\log \mathrm{FC} < -1$), weakly downregulated (light blue, $-1 \leq \log \mathrm{FC} \leq -0.58$), weakly upregulated (light red, $0.58 \leq \log \mathrm{FC} \leq 1$), or significantly upregulated (dark red, $\log \mathrm{FC} > 1$). (b) Percentage distribution restricted to the core genes of each module. Modules are shown on the x-axis, with the number of core genes per module indicated above each bar. The color coding reflects the same categories of differential expression as in panel (a).

bioRxiv preprint doi: https://doi.org/10.1101/2024.06.14.599028; this version posted October 21, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

4.5 Module Eigengene Correlation Network analysis identifies immune, signaling, and metabolic pathways as core dysregulated modules in PSVD

To understand the biological processes affected in PSVD, we performed functional over-representation analysis for the 14 selected modules using the Gene Ontology - Biological Process gene sets from MSigDB. grey60 module with the highest positive correlation and significance ($r = 0.91, p = 6.255351e-11$) to PSVD diagnosis trait identified processes related to immune cell activation and differentiation, involving the T cells, leukocytes and lymphocytes. This module highlights the adaptive immune system related processes due to terms involving acute and antigenic inflammatory response, positive regulation of cell-cell adhesion and cytokine signaling pathways. (The enrichment of cardiac and placental morphogenesis terms may reflect shared developmental signaling pathways active during hepatic organogenesis or in hepatic stromal or endothelial compartments) - part of the discussion .

Blue module with highest significant negative correlation ($r = -0.91, p = 6.298297e-11$) to PSVD diagnosis is strongly enriched for metabolic, translational and insulin-responsive processes. This module reflects the core metabolic functions associated with liver tissue. Key metabolic processes part of the blue module enrichment are fatty acid β-oxidation, lipid catabolism and biosynthesis, steroid and ketone metabolism, and cellular energy production via oxidative phosphorylation and aerobic respiration processes. Additionally, enriched terms related to cytoplasmic translation, protein ubiquitination and TOR signaling regulation suggest the nutrient sensing, energy balance, and stress response role of liver. Vesicle transport and ER-Golgi trafficking terms hint at active protein and lipid processing, critical in hepatocytes.

Fig. 5 presents the Module Eigengene Correlation Network for key modules significantly associated with PSVD diagnosis. The immune-related modules, grey60 (Immune cell Activation and Adhesion) and red (Innate Immune & Vitamin Biosynthesis), exhibit a strong positive correlation with each other, indicating coordinated expression patterns, and both modules show a positive association with PSVD diagnosis. Metabolic modules, including blue (Lipid Metabolism, Energy Production & Insulin Signaling), black (Hepatic Detoxification, and Amino Acid & Lipid Catabolism), and lightgreen (Glycoprotein Metabolism & Endothelial Regulation), also display strong positive correlations among themselves; however, they are negatively associated with PSVD, suggesting a potential downregulation of metabolic pathways in PSVD patients. Signaling-related modules, turquoise (Chemosensory & ciliary motility) and yellow (Sensory Perception & GPCR Signaling), show a strong positive correlation with each other and are positively associated with PSVD, indicating co-regulated signaling processes in PSVD patients. Interestingly, there is a pronounced negative correlation between the metabolic modules (black, blue, lightgreen) and the immune modules (grey60, red), as well as between the metabolic modules and the signaling modules (turquoise, yellow). These patterns reveal the existence of three functionally distinct module clusters: immune, metabolic, and signaling clusters with opposing correlations to the diagnosis of PSVD. This organization may reflect antagonistic or complementary regulatory mechanisms, suggesting that upregulation of immune and signaling pathways may occur concurrently with downregulation of metabolic pathways in PSVD pathogenesis.

bioRxiv preprint doi: https://doi.org/10.1101/2024.06.14.599028; this version posted October 21, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

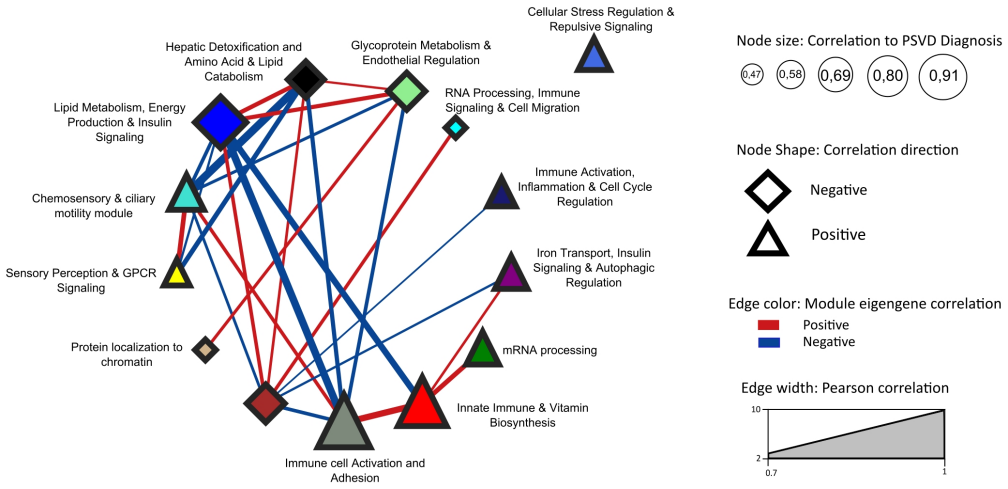

Fig 5. Module Eigengene Correlation Network of PSVD-associated key modules. The network visualization depicts the correlations among module eigengenes for modules significantly associated with PSVD diagnosis, highlighting pathways relevant to PSVD. Node size reflects the strength of correlation with PSVD (larger nodes indicate higher correlation, ranging from 0.47 to 0.91), while node shape indicates the direction of the correlation (triangle = positive, diamond = negative). Edges represent correlations between module eigengenes, with color indicating the correlation direction (red = positive, blue = negative) and edge width proportional to the Pearson correlation strength (0.7–1.0). Modules are annotated with enriched biological functions based on Gene Ontology: Biological Processes, providing insights into the functional relevance of PSVD-associated co-expression modules.

4.6 Protein-protein Interaction (PPI) networks for the Immune, Signaling and Metabolic module clusters

Fig. 6 shows the PPI network for the immune and signaling module clusters. Fig. 6a, representing the PPI network of the immune module cluster for grey60 and red module core genes, indicates five sub clusters identified by the Glay community detection algorithm and functionally annotated using the STRINGapp enrichment function. The large red-bordered node, A2M, has been previously linked to PSVD pathogenesis in the study by Hernandez-Gea V et al. [3]. The large black-bordered nodes (CCR7, JUN,BARD1, RPS27A, and SNRPG) are the top five ranked genes based on the product of node degree and gene log-fold change.

Fig. 6b, representing the PPI network of the signaling module cluster for turquoise and yellow module core genes, indicates six sub clusters identified by the Glay community detection algorithm and functionally annotated using the STRINGapp enrichment function. The large red-bordered node, SERPINA12, has been previously linked to PSVD pathogenesis in the study by Hernandez-Gea V et al. [3]. The large black-bordered nodes (GHRL, TOP2A, CDC6, IL6, and CD19) are the top five ranked genes based on the product of node degree and gene log-fold change.

Supplementary Fig. **Protein–Protein Interaction (PPI) Networks for the Metabolic Module cluster (blue, black, lightgreen).** Nodes represent core genes, colored by log-fold change (logFC). The red-bordered node is a previously reported gene, while large black-bordered nodes are the top five genes ranked by absolute(logFC) × node degree. Subclusters (circles) were identified using the Glay community detection algorithm, with functional enrichment and annotation performed via STRINGapp

bioRxiv preprint doi: https://doi.org/10.1101/2024.06.14.599028; this version posted October 21, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

represents the PPI network of the metabolic module cluster (blue, black, and lightgreen). The network shows 11 sub clusters identified by the Glay community detection algorithm and functionally annotated using the STRINGapp enrichment function. The large red-bordered node, (ATP5MG, ATP5PF, ATPV0C, ATP5F1C, ATP6V0E1, ATP5PO, ATP5F1A, SERPIND1, APOE and APOA2), has been previously linked to PSVD pathogenesis in the study by Hernandez-Gea V et al. [3]. The large black-bordered nodes (RPS6, RPS8, RPS11, MRPL12 and RPL12) are the top five ranked genes based on the product of node degree and gene log-fold change.

bioRxiv preprint doi: https://doi.org/10.1101/2024.06.14.599028; this version posted October 21, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

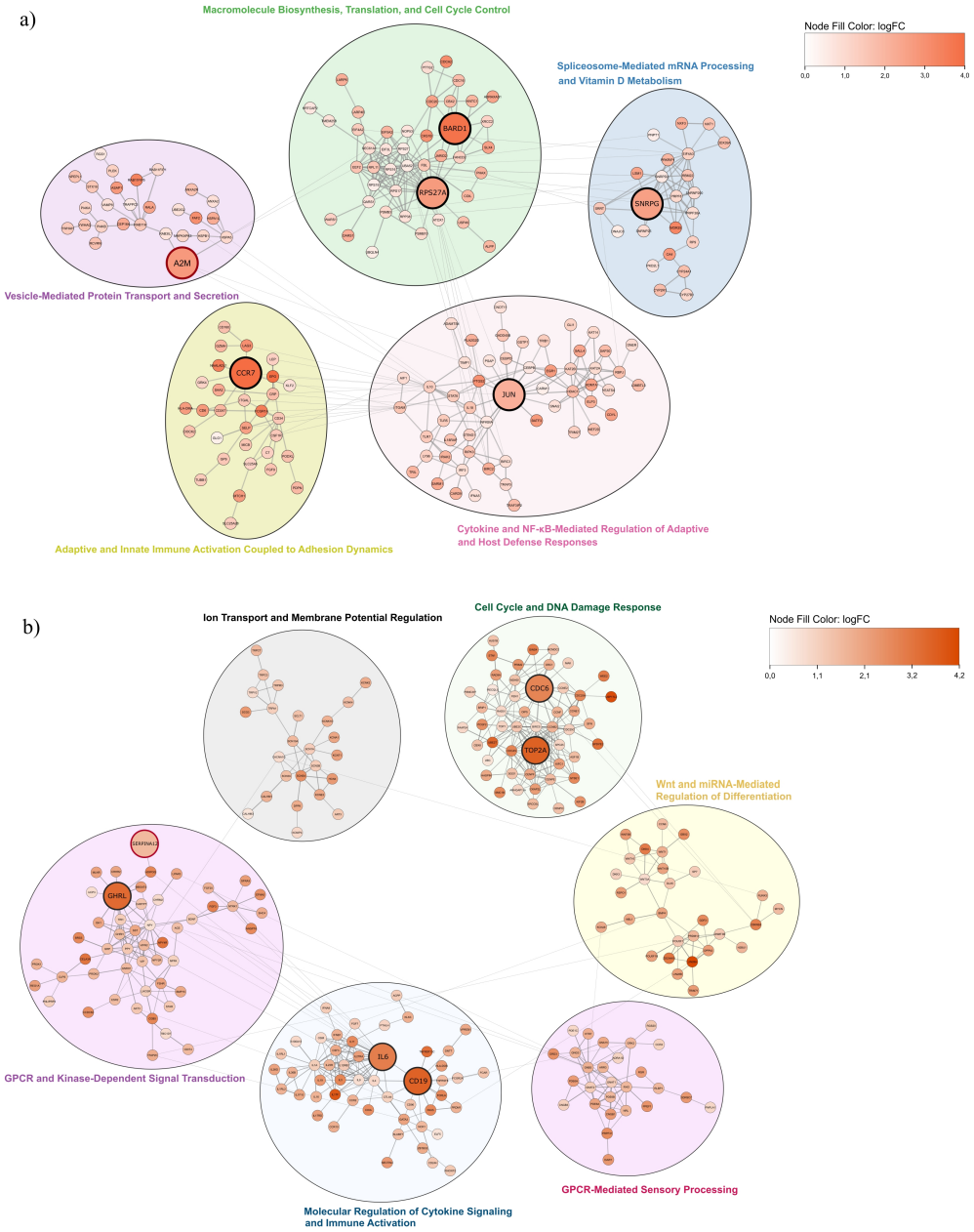

Fig 6. Protein–Protein Interaction (PPI) Networks for Immune and Signaling Module Clusters. a) Immune module cluster (grey60 and red): Nodes represent core genes, colored by log-fold change (logFC). The red-bordered node is a previously reported gene, while large black-bordered nodes are the top five genes ranked by absolute(logFC) × node degree. Subclusters (circles) were identified using the Glay community detection algorithm, with functional enrichment and annotation performed via STRINGapp. b) Signaling module cluster (turquoise and yellow): Nodes represent core genes, colored by logFC. The red-bordered node is a previously reported gene, while large black-bordered nodes are the top five ranked genes. Sub clusters and functional enrichment and annotation were done as above.

bioRxiv preprint doi: https://doi.org/10.1101/2024.06.14.599028; this version posted October 21, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

5 Discussion

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 transcriptomic changes between the PSVD and HNL groups.

The top differentially expressed genes indicate early cellular stress and vascular dysfunction in PSVD. Upregulated EPO, ANKRD1, and GAGE12J suggests the activation of erythropoietic and mechanotransductive stress pathways [24–30]. Elevated EPO likely reflects IL-6–dependent hepatic signaling under hypoxic or inflammatory conditions [31]. ANKRD1, a YAP/Hippo-responsive mechanosensor induced by pro-inflammatory cytokines, may mirror endothelial strain and extracellular-matrix remodeling, both central to sinusoidal injury [32–34]. GAGE12J, lacking functional annotation, represents a novel transcript of potential relevance to PSVD. Downregulated EFNA2, NFIC, and METRN indicate impaired vascular stability and regeneration. EFNA2 loss may weaken angiogenic and immune–endothelial communication [35–37]. Unlike hepatocellular carcinoma, where EFNA2 is upregulated and pro-angiogenic [38], this downregulation may represent a PSVD-specific maladaptive response of the portal microcirculation. NFIC reduction suggests diminished hepatocyte proliferation and matrix regulation through TGF-β-dependent signaling [39, 40]. METRN downregulation implies disturbed endothelial–immune signaling and vascular repair [41]. Together, these genes define early molecular events linking inflammation, vascular stress, and regenerative failure.

Using both GSEA and WGCNA, we examined altered processes in PSVD and their interrelationships to gain mechanistic insights into disease pathogenesis. The Module Eigengene Correlation Network (see Fig. 5) highlighted three major clusters. The immune cluster (immune cell activation, adhesion (grey60) and innate immune and vitamin biosynthesis (red) modules) was positively associated with PSVD. Interestingly, immune enrichment was not detected by GSEA, suggesting that coexpression analysis may capture subtler immune dysregulation. The signaling cluster (chemosensory and ciliary motility (turquoise) and sensory perception and GPCR signaling (yellow) modules) was also positively associated, consistent with GSEA findings. In contrast, the metabolic cluster (hepatic detoxification, amino acid and lipid catabolism (black), lipid metabolism, energy production and insulin signaling (blue), and glycoprotein metabolism with endothelial regulation (lightgreen) modules) was negatively associated, corroborated by GSEA. Importantly, immune and signaling modules were positively correlated with each other but negatively correlated with metabolic modules. This pattern suggests a coordinated dysregulation in PSVD, where heightened immune and signaling activity occurs in parallel with the suppression of metabolic pathways, highlighting novel pathway interconnections that may underlie disease mechanisms.

Within the immune module cluster network (see Fig. 6a), we identified the top five ranked genes based on their node degree and log fold change. No direct link between BARD1 and PSVD has been reported; BARD1 is mainly studied as a BRCA1 partner in DNA repair and is overexpressed in hepatocellular carcinoma, where it promotes tumor progression [42].Although RPS27A has not been linked to PSVD to date, its roles in translational control, ubiquitin signaling, and inflammation via NF-κB suggest it is a plausible candidate for influencing endothelial stress or regenerative pathways in the portal microvasculature. No known direct evidence links JUN (c-Jun) to PSVD. However, in the liver, c-Jun regulates hepatocyte survival, proliferation, and fibrosis in injury contexts [43]. Though not studied in PSVD, CCR7's role in modulating hepatic T-cell homeostasis and its immunoregulatory function make it a plausible candidate for involvement in portal microvascular inflammation and immune infiltration in PSVD. In a PSVD co-expression network study, A2M was identified as a highly connective gene, suggesting it may contribute to disease pathogenesis through roles in protease inhibition

bioRxiv preprint doi: https://doi.org/10.1101/2024.06.14.599028; this version posted October 21, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

and matrix remodeling [3].

For the signaling module cluster network (see Fig. 6b), the top 5 ranked genes, GHRL, IL6, CD19, CDC6, and TOP2A, have not been directly linked to PSVD. However, their roles in hepatoprotection and fibrosis (GHRL) [44], inflammation and endothelial injury (IL6, CD19) [45], and liver inflammation and immune cell regulation (CDC6, TOP2A) [46] suggest that their dysregulation could contribute to vascular remodeling and impaired regeneration in PSVD. In contrast to Hernández-Gea et al., SERPINC1 was found to be downregulated when comparing PSVD to liver cirrhosis and healthy controls; our PSVD versus healthy liver biopsy analysis revealed SERPINA12 upregulation within the signaling module cluster [3]. Given that serpins act as key regulators of protease activity, coagulation balance, and inflammatory signaling, this pattern suggests that distinct serpin family members may differentially contribute to PSVD pathogenesis. SERPINC1 reflects altered anticoagulant activity, while SERPINA12 may participate in modulating vascular inflammation and endothelial–immune interactions.

In the metabolic module cluster (see Supplementary Fig. **Protein–Protein Interaction (PPI) Networks for the Metabolic Module cluster (blue, black, lightgreen).** Nodes represent core genes, colored by log-fold change (logFC). The red-bordered node is a previously reported gene, while large black-bordered nodes are the top five genes ranked by absolute(logFC) $\times$ node degree. Subclusters (circles) were identified using the Glay community detection algorithm, with functional enrichment and annotation performed via STRINGapp), the top 5 ranked genes, RPS6, RPS8, and RPS11, part of the translation and ribosome biogenesis subcluster, were downregulated, indicating impaired ribosome biogenesis and translational activity. Ribosomal stress is known to activate p53, disrupt hepatocyte viability, and modulate immune regulation [47–49], suggesting that suppressed ribosomal protein expression may contribute to endothelial dysfunction and microvascular remodeling in PSVD. In PSVD, SERPINC1 (antithrombin III) has been highlighted as a hub gene in co-expression network analysis, underscoring the role of anticoagulant serpins in disease pathogenesis [3]. SERPIND1 (heparin cofactor II), part of the complement, proteolysis, and vesicle-mediated immune regulation subcluster, although not yet studied in PSVD, shares functional overlap with SERPINC1 as a liver-derived serpin that inhibits thrombin activity in the presence of heparin or dermatan sulfate [50, 51]. Together, these findings suggest that the dysregulation of multiple anticoagulant serpins may collectively influence coagulation balance and vascular remodeling in PSVD. APOE and APOA2, part of the complement, proteolysis, and vesicle-mediated immune regulation subcluster in the metabolic module network, were downregulated in PSVD, contrasting with their upregulation in Hernández-Gea et al. [3]. This difference likely reflects variations in the control groups (HNL vs. cirrhosis+HNL), but consistently implicates lipid metabolism in PSVD pathogenesis. In the network, we additionally observed a consistent downregulation of multiple ATPase subunits, including mitochondrial ATP synthase components (ATP5MG, ATP5PF, ATP5F1C, ATP5PO, ATP5F1A) and vacuolar proton pump subunits (ATP6V0E1, ATPV0C), which are part of the mitochondrial energy metabolism and biosynthesis subcluster. This points to deficits in oxidative phosphorylation and vesicular acidification. Notably, Hernández-Gea et al. (2021) also flagged ATP synthases (e.g., ATP5G1, ATP5B) as highly connective genes within the PSVD transcriptomic network, implicating mitochondrial energy dysfunction as a shared pathogenic axis [3]. The coordinated downregulation of ribosomal proteins (RPS6, RPS8, RPS11) and ATPase subunits (ATP5MG, ATP5PF, ATP5F1C, ATP5PO, ATP5F1A, ATP6V0E1, ATPV0C) points to a combined translational and bioenergetic dysfunction. Together with Hernández-Gea et al. (2021) highlighting these pathways as network hubs, this underscores impaired protein synthesis and mitochondrial energy

bioRxiv preprint doi: https://doi.org/10.1101/2024.06.14.599028; this version posted October 21, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

metabolism as central drivers of PSVD pathogenesis.

A limitation of this study is the relatively small patient cohort, a common issue in studying rare diseases. Although pooling transcriptomic data across studies could help, evolving diagnostic criteria for PSVD and INCPH complicate integration. Nonetheless, our re-analysis provides complementary insights to earlier work by Hernández-Gea et al. (2021). By directly comparing PSVD with HNL, we refined existing observations and uncovered additional interconnected processes and candidate genes. A central pattern emerging from our analysis is a coordinated dysregulation in PSVD of the immune and signaling pathways being upregulated in parallel with the suppression of metabolic processes. While these findings describe processes at the population level, patient heterogeneity in risk factors, genetics, and environmental exposures remains a challenge. Emerging evidence indicates that gut dysbiosis contributes to PSVD onset and progression by driving porto-sinusoidal abnormalities and intrahepatic thrombosis. Disruption of the gut–liver axis permit translocation of microbial products, such as LPS and metabolites, into the liver. Future studies integrating microbiome profiling with transcriptomic and metabolic analyses are needed to uncover gut–liver–vascular interactions in disease pathogenesis.

6 Conclusion

This re-analysis of PSVD transcriptomics, focusing on the comparison with histologically normal liver, uncovers a coordinated imbalance between upregulated immune and signaling processes and suppressed metabolic, translational, and bioenergetic pathways. GSEA demonstrated upregulation of signaling and chemosensory pathways and, broad suppression of metabolic and protein translation associated processes. Coexpression network analysis further revealed consistent downregulation of ribosomal proteins (RPS6, RPS8, RPS11) and ATP synthase subunits (ATP5MG, ATP5PF, ATP5F1C, ATP5PO, ATP5F1A, ATP6V0E1, ATPV0C), implicating combined deficits in protein synthesis and mitochondrial energy metabolism as central drivers of vascular remodeling. The contrasting regulation of serpin family members (SERPINC1, SERPINA12, SERPIND1) highlights disrupted anticoagulant and inflammatory control as additional disease mechanisms. Differences from prior studies likely reflect distinct control group selection but consistently point to lipid metabolism and coagulation as core pathogenic axes.

Together, these findings position PSVD as a disorder of integrated immune, vascular, and metabolic dysregulation. They also underscore the need for effective multi-omics studies to validate candidate genes and pathways uncovered in this and previous studies, while emerging evidence on the gut–liver axis suggests that incorporating microbiome analysis may uncover novel methods of non-invasive patient diagnosis and prognosis. This systems-level framework would refine our understanding of PSVD pathogenesis and open up further avenues for mechanistic exploration and biomarker discovery.

7 Supporting information

S1 Fig.

bioRxiv preprint doi: https://doi.org/10.1101/2024.06.14.599028; this version posted October 21, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

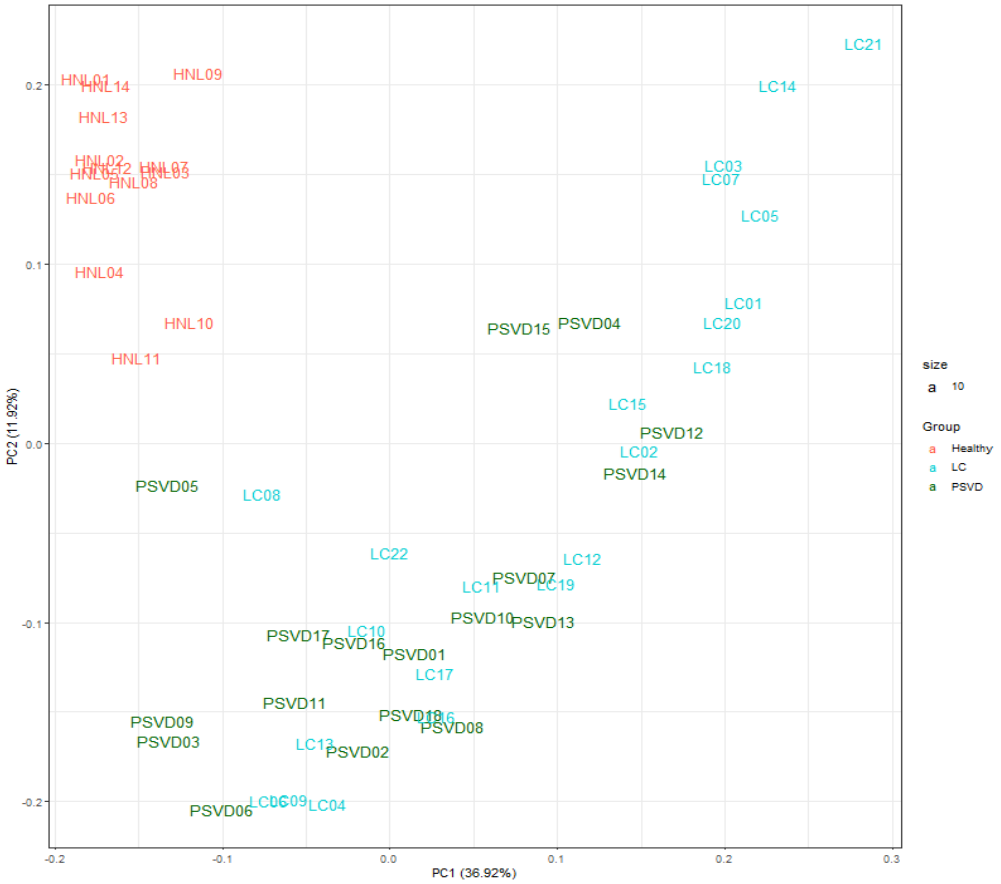

Figure S1. Principal component analysis (PCA) for healthy, liver cirrhosis, and PSVD liver biopsy samples. PCA was performed on normalized transformed gene expression data. The PCA plot above represents the variance explained by the top two components.

bioRxiv preprint doi: https://doi.org/10.1101/2024.06.14.599028; this version posted October 21, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

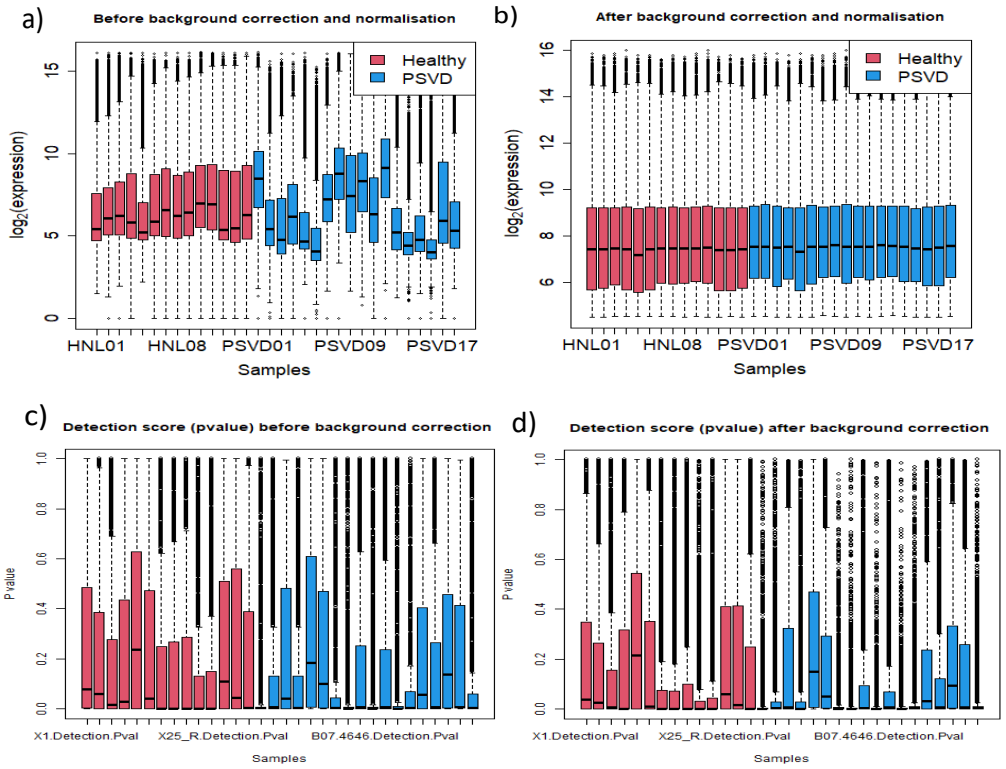

Figure S2. Distribution of the gene expression data across samples. (A)Distribution of the gene expression data before and after normalization. X-axis represents the samples (pink – Healthy liver biopsies and blue- PSVD liver biopsies). Y-axis represents the genes expression values in logarithmic scale. (B) Distribution of the detection p-value before and after normalization. X-axis represents the samples (pink – Healthy liver biopsies and blue- PSVD liver biopsies). Y-axis represents the p-values of the probes used to measure the gene expression of the samples.

S2 Fig.

S3 Fig.

bioRxiv preprint doi: https://doi.org/10.1101/2024.06.14.599028; this version posted October 21, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

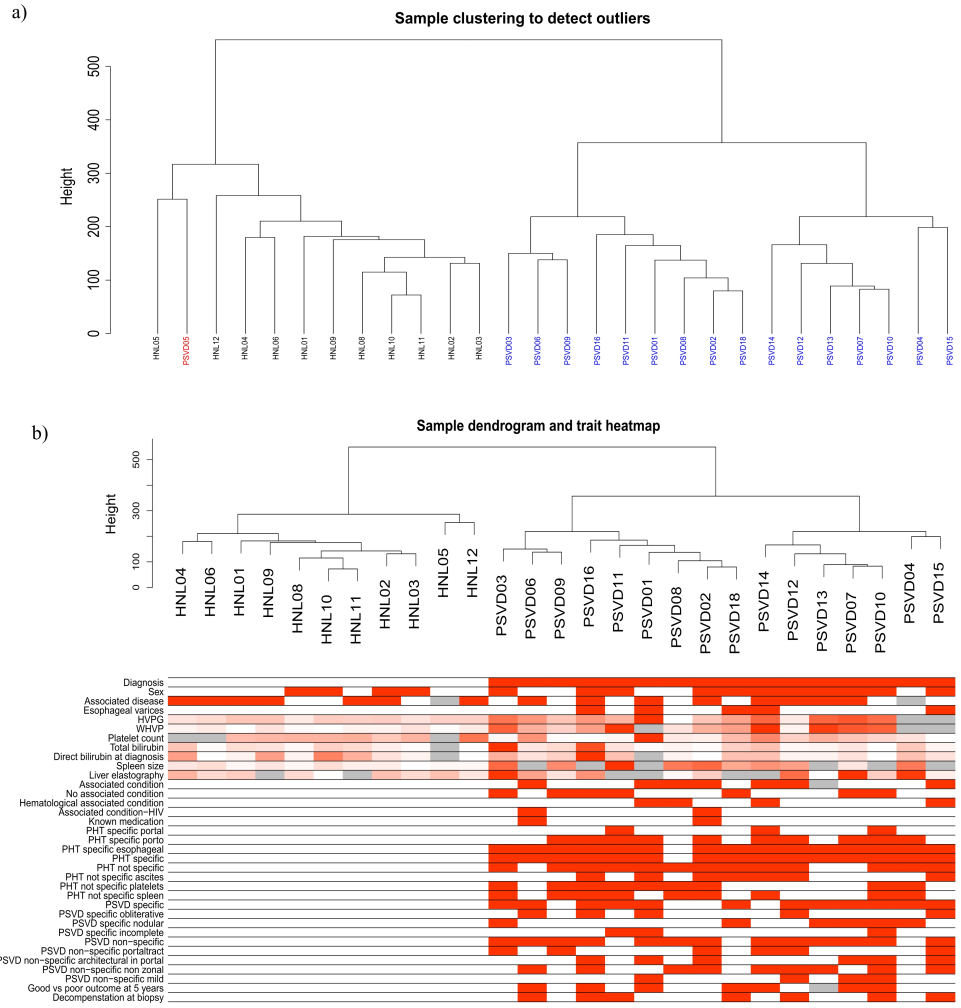

Figure S3. Hierarchical clustering of the samples. (a) Dendogram of the sample clustering. Sample PSVD05 (shown in red) was removed from the analysis given that it was clustering with the healthy liver samples. (b) Dendogram representing the sample clustering after outlier removal against the clinical variables visualized in the rows.

bioRxiv preprint doi: https://doi.org/10.1101/2024.06.14.599028; this version posted October 21, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

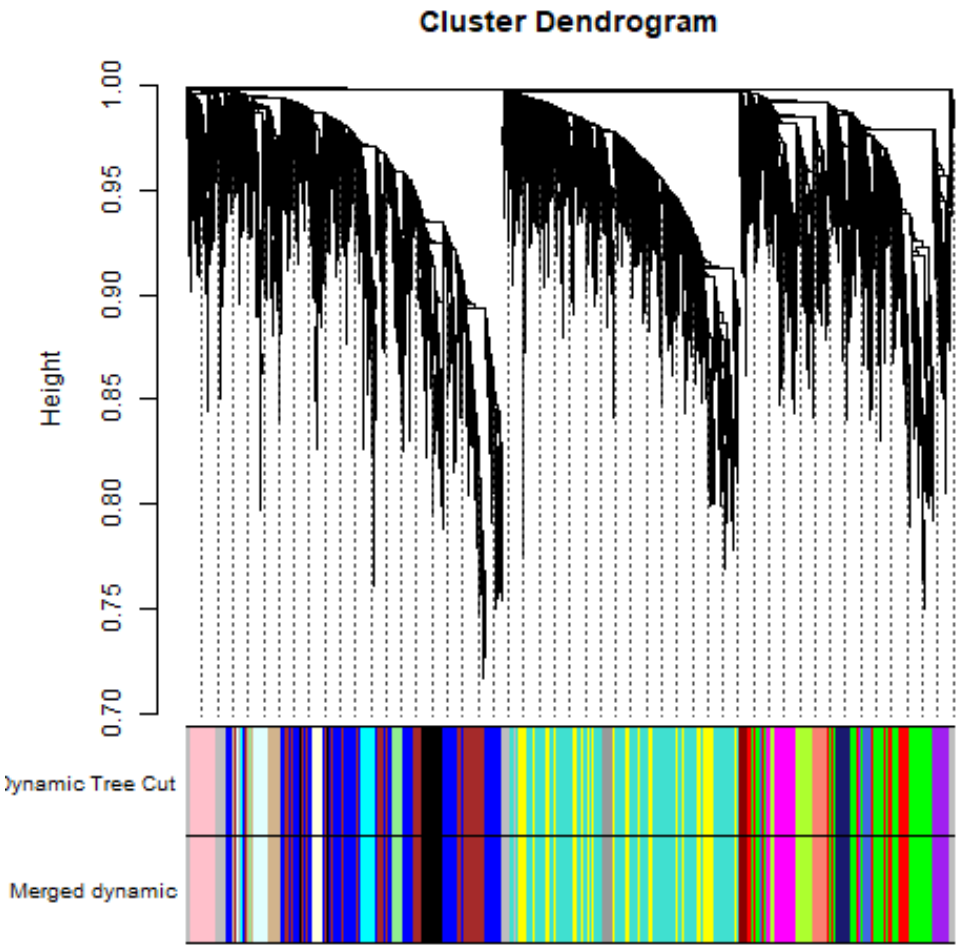

Figure S4. Cluster dendrogram and modules detected. Hierarchical tree (average linkage) using dynamic tree cutting method was used for module detection. The Dynamic Tree Cut band represents the genes assigned to particular modules. 24 modules were detected.

S4 Fig.

S5 Fig.

bioRxiv preprint doi: https://doi.org/10.1101/2024.06.14.599028; this version posted October 21, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

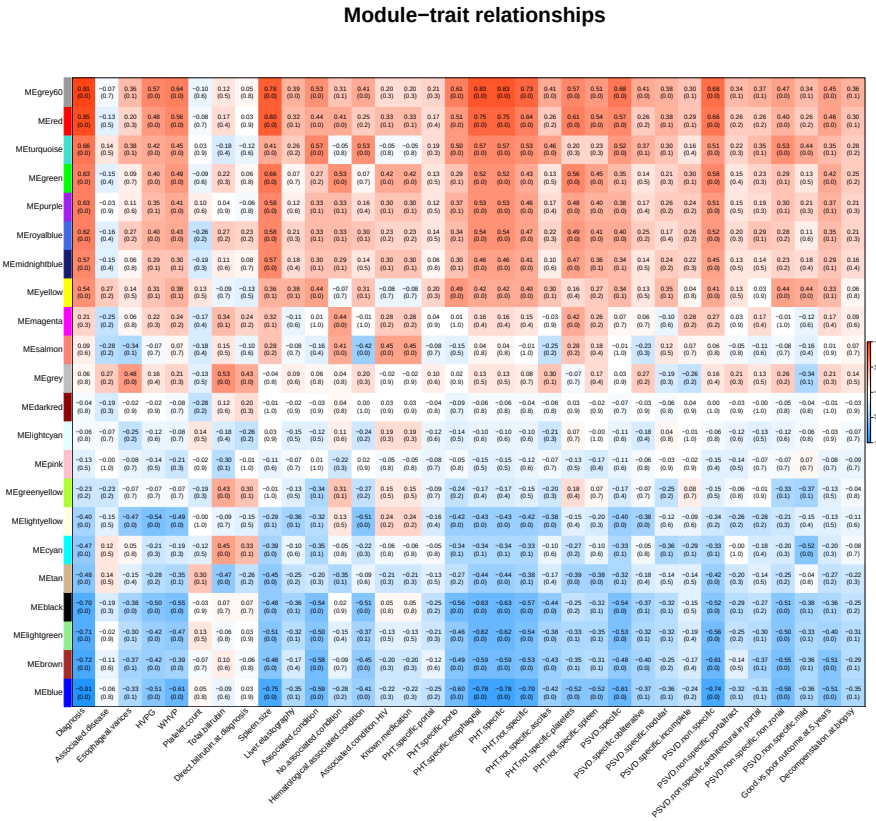

Figure S5. Module-trait relationship heatmap. A heatmap of the module-trait relationship for modules significantly correlating to the diagnosis of PSVD on the y-axis and clinical variables on the x-axis. The color gradient on the heatmap represents the strength of the Pearson correlation coefficients. Number in each cell is the correlation and the p-value (in brackets). Hepatic venous pressure gradient (HVPG), wedged hepatic vein pressure (WHVP), portal hypertension (PHT), porto-sinusoidal vascular diseases (PSVD).

bioRxiv preprint doi: https://doi.org/10.1101/2024.06.14.599028; this version posted October 21, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

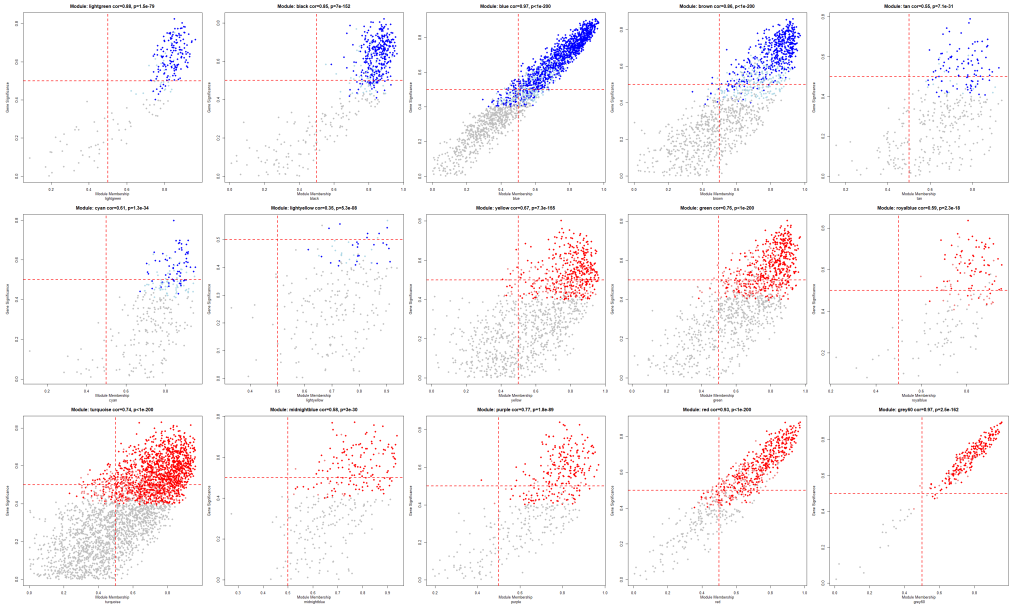

Figure S6. The scatter plots depict the relationship between gene significance (GS) for PSVD diagnosis and module membership (MM) within each key co-expression module identified by WGCNA. Each point represents a single gene, where GS reflects the correlation between gene expression and PSVD diagnosis, and MM represents the correlation between the gene and the module eigengene, indicating its connectivity within the module. Genes in the upper-right quadrant (GS > 0.5 and MM > 0.5) were designated as core (hub) genes, as they are both strongly associated with the trait and highly central within the module network. Data points are color-coded to reflect differential expression in PSVD: significantly upregulated (red), weakly upregulated (light red), significantly downregulated (blue), weakly downregulated (light blue), and non-significant genes (grey).

S6 Fig.

S7 Fig.

bioRxiv preprint doi: https://doi.org/10.1101/2024.06.14.599028; this version posted October 21, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

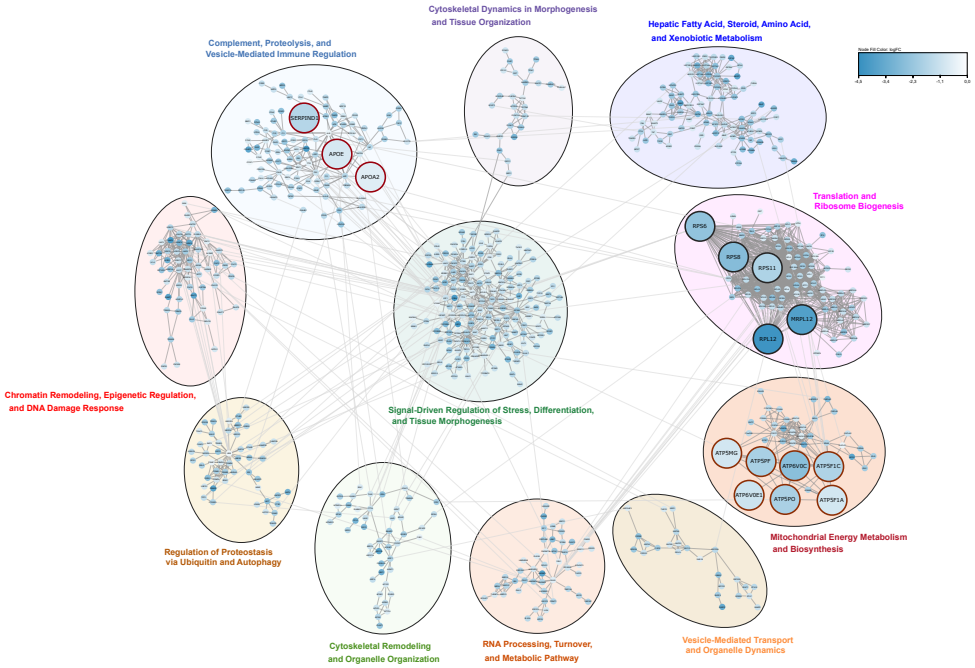

Figure S7. Protein–Protein Interaction (PPI) Networks for the Metabolic Module cluster (blue, black, lightgreen). Nodes represent core genes, colored by log-fold change (logFC). The red-bordered node is a previously reported gene, while large black-bordered nodes are the top five genes ranked by absolute(logFC) × node degree. Subclusters (circles) were identified using the Glay community detection algorithm, with functional enrichment and annotation performed via STRINGapp.

bioRxiv preprint doi: https://doi.org/10.1101/2024.06.14.599028; this version posted October 21, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

S1 Table. Enrichment output for all the 14 key modules. See S1_Table.xlsx for the complete dataset.

8 Acknowledgments

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.

9 Ethical approval declarations

For this study, a previously published transcriptomics dataset by Lozano et al. was obtained from the GEO database (GEO:GSE77627) [3]. The protocol was therefore already reviewed and approved by the ethical committee at their institution (HCB/2009/5448).

10 Declarations

- Funding : This work is supported by funding from the European Union's Horizon 2020 research and innovation program under the EJP RD COFUND-EJP N° 825575.
- Conflict of interest/Competing interests : The authors declare no conflicting interests.
- Ethics approval and consent to participate : A previously published transcriptomics dataset by Lozano et al. was obtained from the GEO database (GEO:GSE77627) [3], ethical approval (HCB/2009/5448). All patients registered in their database gave specific informed consent to use their clinical data and the remnants of their liver biopsies for research studies approved by the Ethical Committee.
- Consent for publication : Not applicable
- Data availability : The transcriptomics dataset analysed during the current study is available in the Gene Expression Omnibus (GEO) database under the accession number (GSE77627)(`https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE77627`).Clinical data used in this analysis can be made available on reasonable request.
- Materials availability : Not applicable
- Code availability : The workflow for this study has been implemented in the Common Workflow Language and can be accessed on WorkflowHub (`https://doi.org/10.48546/WORKFLOWHUB.WORKFLOW.1040.1`).
- 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.

bioRxiv preprint doi: https://doi.org/10.1101/2024.06.14.599028; this version posted October 21, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

References

1. De Gottardi A, Rautou PE, Schouten J, Rubbia-Brandt L, Leebeek F, Trebicka J, et al. Porto-sinusoidal vascular disease: proposal and description of a novel entity [Journal Article]. 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 [Journal Article]. 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 Article]. Journal of hepatology. 2021;75(4):924-34.

4. Gaiteri C, Ding Y, French B, Tseng GC, Sibille E. Beyond modules and hubs: the potential of gene coexpression networks for investigating molecular mechanisms of complex brain disorders [Journal Article]. Genes, brain and behavior. 2014;13(1):13-24.

5. Guo X, Xiao H, Guo S, Dong L, Chen J. Identification of breast cancer mechanism based on weighted gene coexpression network analysis [Journal Article]. Cancer gene therapy. 2017;24(8):333-41.

6. Du P, Kibbe WA, Lin SM. lumi: a pipeline for processing Illumina microarray [Journal Article]. Bioinformatics. 2008;24(13):1547-8.

7. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies [Journal Article]. Nucleic acids research. 2015;43(7):e47-7.

8. R Core Team. R: A Language and Environment for Statistical Computing. Vienna, Austria; 2025. Available from: `https://www.R-project.org/`.

9. Yu G, Wang LG, Han Y, He QY. clusterProfiler: an R package for comparing biological themes among gene clusters [Journal Article]. Omics: a journal of integrative biology. 2012;16(5):284-7.

10. 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 [Journal Article]. Proceedings of the National Academy of Sciences. 2005;102(43):15545-50.

11. Liberzon A, Birger C, Thorvaldsdóttir H, Ghandi M, Mesirov JP, Tamayo P. The molecular signatures database hallmark gene set collection [Journal Article]. Cell systems. 2015;1(6):417-25.

12. Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes [Journal Article]. Nucleic acids research. 2000;28(1):27-30.

13. Kanehisa M. Toward understanding the origin and evolution of cellular organisms [Journal Article]. Protein Science. 2019;28(11):1947-51.

14. Agrawal A, Balcı H, Hanspers K, Coort SL, Martens M, Slenter DN, et al. WikiPathways 2024: next generation pathway database [Journal Article]. Nucleic acids research. 2024;52(D1):D679-89.

bioRxiv preprint doi: https://doi.org/10.1101/2024.06.14.599028; this version posted October 21, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

15. Milacic M, Beavers D, Conley P, Gong C, Gillespie M, Griss J, et al. The reactome pathway knowledgebase 2024 [Journal Article]. Nucleic acids research. 2024;52(D1):D672-8.

16. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, et al. Gene ontology: tool for the unification of biology [Journal Article]. Nature genetics. 2000;25(1):25-9.

17. Aleksander SA, Balhoff J, Carbon S, Cherry JM, Drabkin HJ, Ebert D, et al. The gene ontology knowledgebase in 2023 [Journal Article]. Genetics. 2023;224(1):iyad031.

18. Merico D, Isserlin R, Stueker O, Emili A, Bader GD. Enrichment map: a network-based method for gene-set enrichment visualization and interpretation [Journal Article]. PloS one. 2010;5(11):e13984.

19. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis [Journal Article]. BMC bioinformatics. 2008;9:1-13.

20. Doncheva NT, Morris JH, Gorodkin J, Jensen LJ. Cytoscape StringApp: network analysis and visualization of proteomics data. Journal of proteome research. 2018;18(2):623-32.

21. Utriainen M, Morris JH. clusterMaker2: a major update to clusterMaker, a multi-algorithm clustering app for Cytoscape. BMC bioinformatics. 2023;24(1):134.

22. Su G, Kuchinsky A, Morris JH, States DJ, Meng F. GLay: community structure analysis of biological networks. Bioinformatics. 2010;26(24):3135-7.

23. Erslev AJ. Erythropoietin. New England Journal of Medicine. 1991;324(19):1339-44.

24. Krantz SB. Erythropoietin. Blood. 1991;77(3):419-34.

25. Nielsen OJ. Recombinant human erythropoietin: experimental and clinical applications. Scandinavian Journal of Clinical and Laboratory Investigation. 1991;51(1):1-10.

26. Jelkmann W. Regulation of erythropoietin production. The Journal of physiology. 2011;589(6):1251-8.

27. Fisher JW. Erythropoietin: physiology and pharmacology update. Experimental biology and medicine. 2003;228(1):1-14.

28. Mia MM, Singh MK. New insights into Hippo/YAP signaling in fibrotic diseases. Cells. 2022;11(13):2065.

29. Russell JO, Camargo FD. Hippo signalling in the liver: role in development, regeneration and disease. Nature reviews Gastroenterology & hepatology. 2022;19(5):297-312.

30. Pasquale EB. Eph receptors and ephrins in cancer: bidirectional signalling and beyond. Nature Reviews Cancer. 2010;10(3):165-80.

31. Xu X, Wang X, Li Y, Chen R, Wen H, Wang Y, et al. Research progress of ankyrin repeat domain 1 protein: an updated review. Cellular & Molecular Biology Letters. 2024;29(1):131.

bioRxiv preprint doi: https://doi.org/10.1101/2024.06.14.599028; this version posted October 21, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

32. Zheng C, Luo J, Yang Y, Dong R, Yu FX, Zheng S. YAP activation and implications in patients and a mouse model of biliary atresia. Frontiers in Pediatrics. 2021;8:618226.

33. Mannaerts I, Leite SB, Verhulst S, Claerhout S, Eysackers N, Thoen LF, et al. The Hippo pathway effector YAP controls mouse hepatic stellate cell activation. Journal of hepatology. 2015;63(3):679-88.

34. Xu X, Wang X, Li Y, Chen R, Wen H, Wang Y, et al. Research progress of ankyrin repeat domain 1 protein: an updated review. Cellular & Molecular Biology Letters. 2024;29(1):131.

35. Kohara S, Ogawa K. Eph/Ephrin promotes the adhesion of liver tissue-resident macrophages to a mimicked surface of liver sinusoidal endothelial cells. Biomedicines. 2022;10(12):3234.

36. Barquilla A, Pasquale EB. Eph receptors and ephrins: therapeutic opportunities. Annual review of pharmacology and toxicology. 2015;55(1):465-87.

37. Huang S, Dong C, Zhang J, Fu S, Lv Y, Wu J. A comprehensive prognostic and immunological analysis of ephrin family genes in hepatocellular carcinoma. Frontiers in Molecular Biosciences. 2022;9:943384.

38. Edelmann S, Fahrner R, Malinka T, Song BH, Stroka D, Mermod N. Nuclear Factor I-C acts as a regulator of hepatocyte proliferation at the onset of liver regeneration. Liver International. 2015;35(4):1185-94.

39. Plasari G, Calabrese A, Dusserre Y, Gronostajski RM, Mcnair A, Michalik L, et al. Nuclear factor IC links platelet-derived growth factor and transforming growth factor β1 signaling to skin wound healing progression. Molecular and cellular biology. 2009;29(22):6006-17.

40. Nishino J, Yamashita K, Hashiguchi H, Fujii H, Shimazaki T, Hamada H. Meteorin: a secreted protein that regulates glial cell differentiation and promotes axonal extension. The EMBO journal. 2004;23(9):1998-2008.

41. Liao Y, Yuan S, Chen X, Zhu P, Li J, Qin L, et al. Up-regulation of BRCA1-associated RING domain 1 promotes hepatocellular carcinoma progression by targeting Akt signaling. Scientific reports. 2017;7(1):7649.

42. Schulien I, Hockenjos B, Schmitt-Graeff A, Perdekamp MG, Follo M, Thimme R, et al. The transcription factor c-Jun/AP-1 promotes liver fibrosis during non-alcoholic steatohepatitis by regulating Osteopontin expression. Cell Death & Differentiation. 2019;26(9):1688-99.

43. Mao Y, Zhang S, Yu F, Li H, Guo C, Fan X. Ghrelin attenuates liver fibrosis through regulation of TGF-β1 expression and autophagy. International journal of molecular sciences. 2015;16(9):21911-30.

44. Wang MJ, Zhang HL, Chen F, Guo XJ, Liu QG, Hou J. The double-edged effects of IL-6 in liver regeneration, aging, inflammation, and diseases. Experimental Hematology & Oncology. 2024;13(1):62.

45. Jia W, Liu X, Zhang Z. Role of TOP2A and CDC6 in liver cancer. Medicine. 2023;102(42):e35604.

bioRxiv preprint doi: https://doi.org/10.1101/2024.06.14.599028; this version posted October 21, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

46. Comerford SA, Hinnant EA, Chen Y, Hammer RE. Hepatic ribosomal protein S6 (Rps6) insufficiency results in failed bile duct development and loss of hepatocyte viability; a ribosomopathy-like phenotype that is partially p53-dependent. PLoS genetics. 2023;19(1):e1010595.

47. Fu W, Lin Y, Bai M, Yao J, Huang C, Gao L, et al. Beyond ribosomal function: RPS6 deficiency suppresses cholangiocarcinoma cell growth by disrupting alternative splicing. Acta Pharmaceutica Sinica B. 2024;14(9):3931-48.

48. Lindström MS. The Central Role of Ribosomal Proteins in p53 Regulation. Cancers. 2025;17(10):1597.

49. Tollefsen DM. Heparin cofactor II deficiency. Archives of pathology & laboratory medicine. 2002;126(11):1394-400.

50. Tollefsen D, Pestka CA, Monafo W. Activation of heparin cofactor II by dermatan sulfate. Journal of Biological Chemistry. 1983;258(11):6713-6.

51. He L, Vicente CP, Westrick RJ, Eitzman DT, Tollefsen DM, et al. Heparin cofactor II inhibits arterial thrombosis after endothelial injury. The Journal of clinical investigation. 2002;109(2):213-9.