Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Identification of Non-Invasive Tumor-specific Biomarkers for Hepatocellular Carcinoma through Multi-transcriptomics Data Integration

Zahra Golshan
Tarbiat Modares University
Hanieh Fadakar
Shiraz University of Medical Sciences
Saeedeh Salehi
Tehran University of Medical Sciences
Yasaman Najafi
Razi University
Sanaz Akbari
Karaj Islamic Azad University
Maryam Momeni
University of Isfahan
Alieh Gholaminejad
a.gholaminejad@res.mui.ac.ir

Isfahan University of Medical Sciences

Research Article

Keywords: Hepatocellular carcinoma, Transcriptomic analysis, Cancer biomarkers, Non-invasive diagnostics, WGCNA, Blood-based biomarkers, Exosomal miRNAs

Posted Date: November 1st, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10946987/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License. Read Full License

Additional Declarations: No competing interests reported.

Abstract

Background

Hepatocellular carcinoma (HCC) is the most prevalent primary liver cancer and a leading cause of cancer-related mortality worldwide. Early detection is critical for improving patient outcomes; however, current surveillance modalities exhibit insufficient sensitivity for early-stage disease. While tissue biopsy remains the diagnostic gold standard, its invasive nature limits its routine application. This study seeks to identify non-invasive blood-based biomarkers for HCC through transcriptomic analysis.

Methods

We analyzed transcriptomic datasets from The Cancer Genome Atlas (TCGA) and the Gene Expression Omnibus (GEO), focusing on HCC tissues, plasma, and peripheral blood mononuclear cells (PBMCs). Weighted Gene Co-expression Network Analysis (WGCNA) and differential expression profiling were employed to identify candidate biomarkers, with key genes prioritized using STRING network analysis. Additionally, exosomal miRNA profiles were integrated to capture systemic regulatory signatures. Pathway enrichment analysis was conducted via Reactome 2024 to identify consistently dysregulated "tissue fundamental pathways" across liver datasets, which were then compared with blood-derived pathways. Candidate genes within these pathways were analyzed to define potential blood-based biomarker panels.

Results

We identified 11 plasma-derived genes (e.g., RPS27A, CCND1, HSP90AB1, POLR2F, H2BC13, H2BC14, POLR2C, POLR2L, NPM1, FEN1, XRCC6), 6 PBMC-derived genes (e.g., JUN, POLR2H, CUL1, DICER1, AURKA, CCNA1), and 9 exosomal miRNAs (e.g., hsa-miR-671-5p, hsa-miR-184, hsa-miR-425-5p, hsa-miR-193a-5p, hsa-miR-103a-3p, hsa-miR-371a-5p, hsa-miR-874-3p, hsa-miR-663a, hsa-miR-18a-5p) significantly correlated with HCC. Many identified biomarkers align with established cancer hallmark pathways, underscoring their mechanistic relevance. Notably, several of these genes (e.g., JUN, POLR2C) and miRNAs (e.g., miR-874-3p) have not been previously reported in association with HCC, highlighting their potential as novel biomarker candidates.

Conclusion

This integrative multi-transcriptomic analysis has unveiled novel, clinically relevant blood-based biomarkers for HCC, providing a promising foundation for sensitive, non-invasive detection strategies. Prospective validation in clinical cohorts is crucial for advancing the translation of these findings into early diagnostic protocols.

Key points

- Discovered novel blood gene and exosomal miRNA biomarkers for HCC detection.

- Found 90 common pathways were identified; 67 were shared with plasma and 12 with PBMC.

- Plasma biomarkers show mitotic dysregulation; PBMCs reveal immune roles.

- Novel genes and miRNAs align with cancer hallmarks, novel in HCC context.

- Panels offer promising non-invasive targets for early HCC diagnosis.

1. Introduction

Liver cancer is the sixth most commonly diagnosed cancer and the third leading cause of cancer-related mortality worldwide, with incidence projected to exceed 1 million new cases annually by 2025. Hepatocellular carcinoma (HCC), the most common primary liver malignancy, accounts for approximately 80% of these cases and ranks among the leading causes of cancer-related death globally [1–4]. Early detection is essential for effective intervention, enabling curative treatments such as surgical resection, radiofrequency ablation, trans-arterial chemoembolization, or liver transplantation. However, because early-stage HCC is frequently asymptomatic, most patients are diagnosed at advanced stages when curative therapies are no longer feasible, contributing to a 5-year survival rate that remains below 20% in most regions [5, 6].

Current surveillance methods, including abdominal ultrasonography in combination with serum α-fetoprotein (AFP) testing, remain the standard approach for early HCC detection. While this strategy offers some diagnostic benefit, it can miss up to one-third of early-stage tumors, limiting opportunities for curative treatment. As such, the identification of alternative, highly accurate imaging- and blood-based surveillance methods is a high-priority goal to improve early detection rates. Although tissue biopsy remains the gold standard for definitive diagnosis, its invasiveness, procedural risks (pain, bleeding, liver injury), and rare potential for intrahepatic metastasis [7, 8] highlight the urgent need for reliable, non-invasive diagnostic tools.

Liquid biopsy has emerged as a promising alternative, enabling real-time, minimally invasive monitoring of tumor-specific biomarkers in blood and other body fluids [9, 10]. Compared to traditional tissue biopsy, liquid biopsy enables dynamic tracking of disease progression and therapeutic response, offering valuable insights for personalized treatment strategies [8, 11]. Tumor-derived components—such as circulating tumor cells, nucleic acids, and exosomes—have shown promise as novel liquid biopsy biomarkers for HCC diagnosis. The recent accessibility of high-throughput laboratory and bioinformatics methods now facilitates the development of robust, multi-marker serological panels for early detection [8, 12].

Among omics technologies, transcriptomics has proven particularly powerful in identifying molecular signatures associated with cancer [13–18]. While emerging approaches such as spatial proteomics and single-cell sequencing are gaining attention, bulk RNA sequencing (RNA-seq) and microarray platforms remain widely accessible, enabling large-scale profiling of differentially expressed genes (DEGs) linked to tumor initiation and progression [19, 20]

However, serum transcriptomics alone may not accurately reflect tumor-specific molecular events. To overcome this limitation, an integrated approach that links DEGs in serum with those involved in tumor tissue biology is essential. Specifically, by performing pathway enrichment analysis on DEGs from both tissue and serum and identifying shared pathways, it is possible to isolate tumor-specific biomarkers that can be measured in blood [21–23].

In the present study, we applied this integrative strategy using transcriptomic datasets from The Cancer Genome Atlas (TCGA) and the Gene Expression Omnibus (GEO), encompassing HCC tissue, peripheral blood mononuclear cells (PBMCs), and plasma samples. Rigorous data preprocessing, normalization, and quality control were performed to ensure reliability. Differential expression analysis identified key DEGs, followed by Weighted Correlation Network Analysis (WGCNA) [24] to construct co-expression networks and pinpoint clinically relevant hub genes. Pathway enrichment analysis revealed shared biological pathways across tissue and liquid biopsy samples to identify tumor-specific pathways detectable in blood compartments. From these, we derived candidate gene and miRNA biomarker panels for non-invasive diagnosis. Our main goal is to discover robust, tumor-specific biomarkers for HCC that can be measured through liquid biopsy, thereby improving early detection and clinical management. The workflow employed for biomarker identification is illustrated in Fig. 1.

2. Methods

2.1 Data Sources

We analyzed publicly available transcriptomic datasets derived from liver tissues, peripheral blood mononuclear cells (PBMCs), plasma, and exosomal miRNAs from the Gene Expression Omnibus (GEO) and The Cancer Genome Atlas (TCGA) databases (https://portal.gdc.cancer.gov/projects/TCGA-LIHC). Liver tissue datasets included RNA-seq data from TCGA-LIHC and microarray data from GEO: GSE14520 (GPL571 and GPL3921; https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE14520) and GSE25097 (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE25097). Blood-based datasets comprised GSE49515 (PBMC; https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE49515), GSE142987 (plasma; https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE142987), and GSE82303 (blood exosomal miRNAs; https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE82303). All datasets were selected based on sample size, data quality, and availability of both tumor and control samples. Data were retrieved in December 2024.

2.2 Data Preprocessing and Quality Control

For microarray data, raw expression matrices were $\log_2$-transformed and quantile normalized using the “Limma” R package. RNA-seq data from GEO were analyzed through DESEq2 standard pipeline. RNA-seq data from TCGA were downloaded as raw counts using the “TCGAbiolinks” package and preprocessed via “edgeR”, including library size normalization (TMM) and dispersion estimation. Quality control was performed by assessing sample distributions, inter-array correlations, and outlier detection using principal component analysis (PCA). Only high-quality samples with consistent clustering were retained for further analysis.

2.3 Differential Expression Analysis (DEA)

Differential expression analysis (DEA) was carried out using limma for microarray datasets and edgeR for TCGA RNA-seq data following standard normalization and dispersion estimation workflows. Genes with $|\log_2$ fold-change (logFC)$| \geq 0.585$ and an adjusted p-value < 0.05 (Benjamini-Hochberg method) were considered differentially expressed (DEGs). Visualizations, including volcano plots and PCA plots, were generated using “ggplot2”.

2.4 Weighted Gene Co-Expression Network Analysis (WGCNA)

WGCNA was performed separately for each of the four liver tissue datasets (TCGA- LIHC, GSE25097, GSE14520-GPL3921, and GSE14520-GPL571) to identify co-expression modules. Using the "WGCNA" R package, a soft-thresholding power (β) was selected to approximate scale-free network topology ($R^2 > 0.9$). Adjacency matrices were converted to topological overlap matrices (TOMs), and modules were identified via hierarchical clustering and dynamic tree cut. A minimum module size of 30 genes was specified, and modules with highly similar eigengenes (dissimilarity < 0.25) were merged. Module–trait relationships were calculated using Pearson correlation with tumor/normal status. To ensure biological relevance, only **positively correlated modules** (Pearson's r > 0.5 with tumor status) were retained and defined as **tumor-associated modules**, which were selected for downstream pathway enrichment analysis. Module Membership (MM), defined as the correlation between individual gene expression and the corresponding module eigengene (ME), was calculated to identify genes within each tumor-associated module. Gene Significance (GS), reflecting the association between gene expression and sample phenotype (normal vs. tumor), was used to prioritize modules with strong clinical relevance. Key genes within tumor-associated modules were prioritized based on high MM and GS.

2.5 Pathway Enrichment Analysis

Pathway enrichment analysis was conducted using Enrichr, querying the Reactome 2024 pathway database. DEGs from each blood dataset (GSE49515, GSE142987) and key genes from the tumor-associated modules of the liver tissue datasets (TCGA- LIHC, GSE25097, GSE14520-GPL3921, and GSE14520-GPL571) were subjected to pathway enrichment analysis separately. Only pathways with adjusted p-values < 0.05 were retained. Next, the non-redundant and overlapping pathways across the four liver tissue datasets were identified and unified. These consolidated pathways were termed "tissue fundamental pathways", representing robust biological processes consistently dysregulated in HCC liver tissues. Subsequently, the overlap between these tissue fundamental pathways and the enriched pathways in blood compartments (PBMC and plasma) was computed to identify shared biological mechanisms that are detectable in the circulation. These shared pathways served as the basis for constructing blood-based biomarker panels.

2.6 Mapping Shared Pathway Gene Sets to Cancer Hallmarks

The HUGO Gene Nomenclature Committee (HGNC)-approved gene symbols from the shared-pathway gene sets —comprising plasma- and PBMC-associated genes—were uploaded to the Cancer Hallmarks Analytics Tool (https://cancerhallmarks.com) [27] to evaluate their biological relevance in the context of tumorigenesis. The "Integrated Cancer Hallmark Gene Set" (human) was selected for analysis, and radar plots were generated to determine the representation of each hallmark category. This classification provided a structured framework for linking circulating molecular alterations to fundamental processes of cancer biology and for prioritizing biomarker candidates with strong mechanistic connections to HCC pathogenesis.

2.7 Construction of Blood-Based Biomarker Panels

To identify non-invasive, blood-based biomarkers reflective of tumor biology, we first identified DEGs from the PBMC (GSE49515) and plasma (GSE142987) datasets that were involved in the shared pathways (i.e., the intersection between blood-enriched pathways and the tissue fundamental pathways). These DEGs were subjected to protein–protein interaction (PPI) network analysis using the Search Tool for the Retrieval of Interacting Genes/Proteins (STRING) database to determine their functional centrality.

From the resulting interaction networks, the top 50 hub genes, ranked by degree of connectivity, were selected for each dataset. To refine these candidates, we next intersected these hub gene lists with the subset of significantly dysregulated DEGs (defined as genes with |logFC| ≥ 2 and adjusted p-value < 0.05) from their respective datasets. The resulting gene sets were defined as PBMC-derived and plasma-derived biomarker panels, respectively. These panels were prioritized for their potential to serve as non-invasive, HCC-specific diagnostic biomarkers, as they represent highly dysregulated, functionally central genes that are both detectable in circulation and mechanistically linked to tumor-related pathways identified in liver tissue.

2.8 Construction of an Exosomal miRNA Biomarker Panel

To elucidate plasma-specific regulatory mechanisms, miRNA–target interactions were retrieved from miRTarBase (https://mirtarbase.cuhk.edu.cn/), focusing exclusively on experimentally validated interactions supported by strong evidence, such as Western blotting, qRT-PCR, and reporter assays. The interaction data were obtained from the latest version of miRTarBase (https://mirtarbase.cuhk.edu.cn/~miRTarBase/miRTarBase_2025/index.php). MiRNAs targeting genes enriched in plasma-associated pathways (identified in Section 2.3) were subsequently determined. To ensure consistency with current nomenclature, outdated miRNA identifiers were updated using the dbDEMC ID Converter tool (https://www.biosino.org/dbDEMC/idConvertor). The overlap between the predicted plasma-associated miRNAs and differentially expressed miRNAs (DEMs) from the exosomal dataset (GSE82303) was evaluated using Venn diagram analysis. MiRNAs common to both datasets were considered promising candidates. Finally, a panel of HCC-specific exosomal miRNA biomarkers was established, comprising miRNAs that targeted plasma pathway genes and demonstrated significant differential expression in exosomal samples, thus representing potential non-invasive biomarkers for HCC detection.

2.9 Hub genes expression levels in immune cells at single cell level in HCC

To validate cellular sources of identified biomarkers, we analyzed the HCC_IMM_10X_GSE140288 dataset from the IMMUcan single-cell RNA-seq database (https://immucanscdb.vital-it.ch/)[25].This 10X Chromium-based HCC dataset comprises 66,187 cells, including T cells (n = 6,555), PBMC/other lymphoid cells (n = 5,650), B cells (n = 528), myeloid cells (n = 1,582), NK cells (n = 1,014), endothelial cells (n = 40), and other tumor-associated populations. Log-normalized expression values of the 11 plasma and 6 PBMC-derived biomarkers were extracted across annotated cell types to determine which tumor and immune populations contribute to the circulating gene signatures defined in our bulk and pathway analyses.

2.10 Use of Artificial Intelligence Technologies

Large Language Models (LLMs) were employed solely to correct grammatical errors and improve the readability of the manuscript. Following their use, the authors reviewed the text, made any necessary revisions, and take full responsibility for the content of the published article.

3. Results

3.1 Differential Expression Analysis Across Tissue and Blood-Based Datasets

To ensure consistency across platforms, microarray and RNA-seq datasets were preprocessed according to platform-specific pipelines. Microarray datasets were log_2-transformed and quantile normalized using the "Limma" package. RNA-seq data from GEO were processed using the standard DESeq2 pipeline, while RNA-seq data from TCGA-LIHC were downloaded as raw counts via the "TCGAbiolinks" package and normalized using "edgeR". Quality control (QC) was rigorously applied, including assessment of sample distribution, inter-array correlation, and PCA to detect and exclude outliers. Only high-quality samples exhibiting consistent clustering patterns were retained for downstream analyses. Following preprocessing and QC, differential expression analysis (DEA) was performed across all selected datasets, encompassing liver tissue, plasma, PBMCs, and blood-derived exosomal miRNAs (Online Resource1 Fig S1). Specifically, one plasma dataset (GSE142987), one PBMC dataset (GSE49515), one exosomal miRNA dataset (GSE82303), two tissue microarray datasets (GSE25097 and GSE14520 with GPL571 and GPL3921 platforms), and TCGA-LIHC RNA-seq data were analyzed.

As summarized in Table 1, the number of differentially expressed genes (DEGs) varied across datasets. The blood-based datasets, including GSE142987 (plasma) and GSE49515 (PBMC), exhibited extensive gene dysregulation. Specifically, 5,286 genes were upregulated and 4,558 downregulated in plasma, while 2,612 upregulated and 1,395 downregulated genes were identified in PBMCs. In the exosomal miRNA dataset GSE82303, 18 upregulated and 16 downregulated miRNAs were detected. In addition to these, DEG analysis of liver tissue datasets, including GSE25097, GSE14520 (GPL571 and GPL3921), and TCGA-LIHC, was performed and the results are reported in Table 1. However, these tissue DEGs were not directly used in downstream biomarker selection. Overall, these findings highlight the widespread molecular dysregulation across both systemic (blood) and hepatic (tissue) compartments in HCC and form the basis for subsequent pathway and biomarker analyses.

Table 1
Summary of datasets characteristics, number of differentially expressed genes (DEGs), and differentially expressed miRNAs (DEMs)

Accession Number	Technology	Platform	Published Year	Sample Type	Tumor Samples	Normal Samples	DEGs	
							Up	Down
GSE25097	Microarray	GPL10687	2011	Tissue	268	243	225	627
GSE14520	Microarray	GPL3921	2010	Tissue	225	220	1762	1125
GSE14520	Microarray	GPL571	2010	Tissue	22	21	1776	1087
TCGA-LIHC	RNAseq	Illumina		Tissue	374	50	438	1290
GSE142987	RNAseq	GPL20795	2020	Plasma	35	30	5286	4558
GSE49515	Microarray	GPL570	2013	PBMC	10	10	2612	1395
GSE82303	Microarray	GPL14943	2023	Blood exosome	27	76	18	16

3.2 Construction of Weighted Gene Co-Expression Networks and Identification of Tumor-Associated Modules

To explore the gene co-expression patterns associated with tumor development, we performed Weighted Gene Co-Expression Network Analysis (WGCNA) on four liver tissue datasets —GSE14520-GPL3921, GSE14520-

GPL571, GSE25097, and TCGA-LIHC—to identify gene modules associated with tumor status (Fig. 2, Online Resource1 FigS2-S5). Following initial sample clustering to ensure data quality, an appropriate soft-thresholding power was selected for each dataset (β = 6 for TCGA- LIHC and GSE25097; β = 13 for GSE14520-GPL3921 and GSE14520-GPL571) to achieve scale-free topology network ($R^2 \geq 0.9$). Subsequent network construction and hierarchical clustering based on the topological overlap matrix (TOM) identified 19 modules in TCGA- LIHC, 13 modules in GSE14520-GPL3921, 24 modules in GSE14520-GPL571, and 33 modules in GSE25097. Module–trait correlation analysis revealed several modules significantly associated with tumor phenotype. Across all datasets, several modules showed significant positive or negative correlations with tumor status. Specifically, TCGA- LIHC exhibited five modules with significant positive correlations and four modules with significant negative correlations. Similarly, GSE14520-GPL3921 revealed three positively correlated and two negatively correlated modules; GSE14520-GPL571 showed six positively correlated and five negatively correlated modules; and GSE25097 displayed seven modules with positive correlations and six modules with negative correlations. Positively correlated modules (Pearson's r > 0.5) were prioritized for downstream pathway analysis, as they were likely to contain genes implicated in HCC development. From these modules, key genes were extracted based on intramodular connectivity. Only these highly connected genes from the positively correlated modules were subjected to pathway enrichment analysis, ensuring that the subsequent biological interpretation was driven by the most central and functionally relevant genes within each tumor-associated network.

3.3 Identification of Key Pathways Underlying HCC Development

Pathway enrichment analysis of the tumor-associated modules from each tissue dataset (using Reactome 2024) revealed a total of 185 (TCGA-LIHC), 337 (GSE25097), 497 (GSE14520-GPL3921), and 778 (GSE14520-GPL571) significantly enriched pathways (Online Resource2 Tables S1–S4). Intersecting these results identified 90 non-redundant, consistently enriched pathways across all datasets, termed "tissue fundamental pathways" (Fig. 3a, Online Resource2 Table S5), representing key molecular mechanisms implicated in HCC pathogenesis.

To prioritize biologically relevant processes, the top 15 tissue fundamental pathways were ranked according to gene ratio and adjusted p-value (Fig. 3b). These pathways highlight critical oncogenic mechanisms, including DNA damage bypass, TP53-mediated transcriptional regulation of cell cycle genes, and cellular senescence, reflecting genomic instability and tumor suppressor dysfunction. Cell cycle–associated events were prominently represented, such as mitotic spindle checkpoint control, termination of translesion DNA synthesis, G0/G1 regulation, and cyclin A/B1/B2-associated transitions during G2/M phase. Additional enriched processes included Polo-like kinase–mediated events, aberrant regulation of mitotic cell cycle due to RB1 defects, and defective apoptosis pathways, all of which are known to drive hepatocarcinogenesis. The recurrence of these pathways across independent datasets highlights their central role in HCC biology and provides a robust mechanistic foundation for subsequent identification of blood-based surrogate biomarkers.

3.4 Shared Pathway Identification Between Blood-Derived and Tissue Datasets

To identify circulating molecular features reflective of tumor-specific biology, pathway enrichment analysis was also conducted on the plasma (GSE142987, Online Resource2 Table S6) and PBMC (GSE49515, Online Resource2 Table S7) datasets. Enriched pathways from each blood-derived dataset were compared with the

previously defined “tissue fundamental pathways”. Comparison of plasma pathways (n = 586) with the tissue fundamental pathways revealed 67 shared pathways (Fig. 3c, Online Resource2 Table S8) comprising 1,067 genes, while comparison of PBMC pathways (n = 86) identified 15 shared pathways (Fig. 3d, Online Resource2 Table S9) involving 536 genes.

The top 15 shared pathways for each comparison were visualized as bubble plots to highlight functional relevance (Fig. 3e). PBMC–tissue shared pathways (left panel) were enriched for transcriptional regulation and mitotic cell-cycle control, including Gene Expression (Transcription), RNA Polymerase II Transcription, SUMOylation, and TP53-mediated transcriptional regulation. Plasma–tissue shared pathways (right panel), in contrast, showed strong enrichment for APC/C complex-mediated degradation of mitotic proteins and other cell-cycle processes such as Activation of APC/C and Cdc20 Phospho-APC/C-Mediated Degradation of Cyclin A. This pattern highlights that plasma-derived signals predominantly capture tumor-driven mitotic dysregulation, whereas PBMCs reflect transcriptional control programs more broadly. Collectively, both plasma and PBMCs capture distinct yet complementary aspects of tumor-associated molecular activity, reinforcing their potential as minimally invasive sources for HCC biomarker panel development.

3.5 Biological Context of Shared Pathway Genes via Cancer Hallmarks

To better understand the biological implications of these shared pathways, we next assessed their representation across the canonical hallmarks of cancer. The shared-pathway gene sets (1,067 plasma-associated genes and 536 PBMC-associated genes) were evaluated for enrichment across the ten canonical hallmarks of cancer using the Cancer Hallmarks Online Tool (https://cancerhallmarks.com). In plasma, significant enrichment (adjusted $p < 0.05$) was observed for Replicative Immortality, Reprogramming Energy Metabolism, Resisting Cell Death, and Genome Instability, with Replicative Immortality showing the strongest association (adjusted $p \approx 0.0001$). In PBMCs, these same four hallmarks were enriched, but additional hallmarks emerged, including Evading Growth Suppressors, Evading Immune Destruction, Tumor-Promoting Inflammation, and Tissue Invasion and Metastasis. These findings suggest that plasma-derived genes predominantly reflect proliferative and survival-related tumor processes, while PBMC-derived genes capture a wider spectrum of cancer-associated mechanisms, including immune modulation, inflammation, and metastasis. Together, these complementary profiles underscore the potential of plasma and PBMC gene signatures as minimally invasive biomarkers that reflect distinct yet interconnected aspects of HCC biology (Fig. 4).

3.6 Identification of Tumor Specific Blood-derived Biomarker Panels for HCC Monitoring

DEGs from the plasma and PBMC datasets that mapped to the shared pathways were extracted and analyzed for protein–protein interaction (PPI) network connectivity using the STRING database (Fig. 4c,d). For each dataset, the top 50 hub genes—defined as nodes with the highest degree of connectivity—were identified (Online Resource2 Tables S10 and S11). These hub genes were then intersected with the subset of highly differentially expressed genes from their respective datasets ($|\log_2$ fold change$| \geq 2$, adjusted p-value < 0.05) (Fig. 4e-g). Genes that overlapped between the two criteria (high connectivity and strong differential expression) were prioritized as candidate biomarker panels. This approach yielded: 11 “HCC-specific plasma-derived biomarker” candidates—NPM1, CCND1, POLR2F, HSP90AB1, POLR2L, H2BC13, XRCC6, RPS27A, H2BC14, FEN1, and

POLR2C—and six “HCC-specific PBMC-derived biomarker” candidates—CCNA1, DICER1, JUN, AURKA, POLR2H, and CUL1. These panels represent robust, non-invasive biomarker candidates with the potential to enhance HCC diagnosis and monitoring (Table 2).

3.7 Identification of Plasma-Associated miRNAs and Construction of an Exosomal miRNA Biomarker Panel

Having identified strong candidate gene-based biomarkers, we next sought to investigate whether circulating microRNAs, which play crucial regulatory roles, could complement these findings by forming an integrated exosomal miRNA biomarker panel. To explore the regulatory mechanisms underlying the 67 plasma-tissue common pathways (enriched with 1067 DEGs), we retrieved miRNA-target interaction data from miRTarBase, focusing on experimentally validated interactions supported by strong evidence. We then examined the overlap between miRNAs targeting plasma-associated DEGs and DEMs from the exosomal dataset GSE82303. Using Venn diagram analysis, we identified a total of 16 common DEMs (Fig. 4g, Online Resource2 Table S12), all of which passed the logFC cutoff of ± 1.0, and 9 of which passed the stricter cutoff of ± 1.5. These DEMs were prioritized based on their logFC change values, with higher values indicating stronger regulatory potential within the context of the study. The list of DEMs with a cutoff of ± 1.5 as “HCC specific blood-based miRNA panel” includes hsa-miR-671-5p, hsa-miR-184, hsa-miR-425-5p, hsa-miR-193a-5p, hsa-miR-103a-3p, hsa-miR-371a-5p, hsa-miR-874-3p, hsa-miR-663a, and hsa-miR-18a-5p. We have therefore introduced a panel consisting of these 9 miRNAs as blood exosomal-based biomarkers, reflecting events occurring in the tumor tissue. This proposed biomarker panel represents key regulators that connect tissue-level pathways with systemic changes observed in plasma and blood exosomes (Table 2).

Table 2
Overview of HCC specific Blood-Based Biomarker Panels

Dataset	Shared Pathways with Tissue	Number of Genes	Top Biomarker Candidates	Node-Degrees
Plasma (GSE142987)	67 shared pathways	1,067	RPS27A	237
			CCND1	184
			HSP90AB1	178
			POLR2F	153
			H2BC13	151
			H2BC14	151
			POLR2C	150
			POLR2L	150
			NPM1	139
			FEN1	132
			XRCC6	131
PBMC (GSE49515)	15 shared pathways	536	JUN	116
			POLR2H	66
			CUL1	64
			DICER1	59
			AURKA	57
			CCNA1	56
Exosomal (GSE82303)	16 common DEMs (9 with a cutoff of ± 1.5 FC)	-	hsa-miR-671-5p, hsa-miR-184, hsa-miR-425-5p, hsa-miR-193a-5p, hsa-miR-103a-3p, hsa-miR-371a-5p, hsa-miR-874-3p, hsa-miR-663a, hsa-miR-18a-5p	-

3.8 Analysis of correlation between hub gene expression and tumor immune microenvironment at single-cell level

To elucidate cellular sources of the signature genes, we analyzed single-cell RNA-seq data from HCC tumors using the IMMUcan dataset HCC_IMM_10X_GSE140288, which comprises 66,187 cells annotated as B, T, NK, myeloid, endothelial, plasma, and unknown clusters. Mean expression values for each gene across major cell types were extracted from the IMMUcan annotation matrix and mapped to the 11 plasma-derived (RPS27A, CCND1, HSP90AB1, POLR2F, POLR2C, POLR2L, NPM1, FEN1, XRCC6) and 6 PBMC-derived (JUN, POLR2H, CUL1, DICER1, AURKA, CCNA1) biomarkers defined by our bulk, pathway, and PPI analyses.

Plasma biomarkers showed high expression in transcriptionally active hematopoietic and endothelial compartments, consistent with their enrichment in cell-cycle, DNA-repair, and transcriptional pathways in bulk data. RPS27A and NPM1 reached maximal levels in B-cell clusters, indicating strong involvement in lymphoid translational and nucleolar activity, while POLR2C, FEN1, and XRCC6 peaked in NK cells, linking these RNA polymerase II and DNA-repair components to cytotoxic lymphocyte populations. CCND1, HSP90AB1, POLR2F, and POLR2L were most highly expressed in endothelial clusters, suggesting that proliferative and stress-response programs captured by the plasma panel extend to tumor-associated vascular and stromal compartments as well (Fig. 5).

PBMC biomarkers were preferentially expressed in immune lineages, supporting their interpretation as markers of systemic immune activation rather than purely tumor-intrinsic signals. JUN and POLR2H exhibited broadly elevated levels across T, NK, B, and endothelial clusters with maxima in endothelial cells, consistent with their roles in activation-induced transcription. CUL1 was highest in NK cells, whereas DICER1 showed co-dominant expression in B and myeloid clusters, aligning with ubiquitin–proteasome and microRNA-processing functions in effector and antigen-presenting cells. AURKA and CCNA1 displayed lower but selective expression in myeloid cells, matching their cell-cycle–related roles in proliferating immune populations (Fig. 5).

4. Discussion

In this study, we employed an integrative multi-transcriptomic approach to identify non-invasive, tumor-specific biomarkers for hepatocellular carcinoma (HCC), analyzing datasets derived from various sources, including tissue samples, plasma, peripheral blood mononuclear cells (PBMCs), and exosomal microRNAs (miRNAs). Through differential expression and co-expression network analyses, we uncovered crucial genes and pathways associated with HCC. Specifically, we identified eleven plasma-derived genes (NPM1, CCND1, POLR2F, HSP90AB1, POLR2L, H2BC13, XRCC6, RPS27A, H2BC14, FEN1, and POLR2C) and six PBMC-derived genes (CCNA1, DICER1, JUN, AURKA, POLR2H, and CUL1) as potential biomarkers. Furthermore, we proposed a panel consisting of nine exosomal miRNAs (hsa-miR-671-5p, hsa-miR-184, hsa-miR-425-5p, hsa-miR-193a-5p, hsa-miR-103a-3p, hsa-miR-371a-5p, hsa-miR-874-3p, hsa-miR-663a, and hsa-miR-18a-5p), which reflect systemic regulatory mechanisms involved in tumorigenesis. Together, these biomarker panels capture key molecular features of HCC. Rather than representing isolated findings, they provide complementary insights into tumor biology—plasma biomarkers reflect tumor-intrinsic molecular programs, while PBMC biomarkers capture the systemic immune and inflammatory responses to malignancy.

Importantly, hallmark analysis of the shared-pathway gene sets provides a conceptual framework for interpreting their potential functional significance. The pattern emerging from plasma-associated genes suggests a predominance of hallmarks linked to tumor growth and survival—replicative immortality, reprogramming of energy metabolism, resistance to cell death, and genomic instability—indicating that circulating molecular signals are heavily shaped by core proliferative processes. In contrast, PBMC-associated genes show enrichment not only in these proliferative hallmarks but also in those related to immune evasion, tumor-promoting inflammation, and metastatic competence. This divergence implies that the two blood compartments capture distinct yet complementary biological dimensions of HCC: plasma largely mirrors tumor-cell–intrinsic dysregulation, whereas PBMCs reflect host–tumor interactions, particularly those involving immune modulation. Together, these findings support the integration of both plasma and PBMC transcriptomic profiles in biomarker discovery to achieve a more comprehensive picture of tumor progression.

The hub genes emerging from our network-based prioritization are not only potential non-invasive diagnostic candidates but also promising therapeutic targets. For instance, CCND1 is known to play a significant role in HCC development by regulating cell cycle progression [26]. NPM1 overexpression has been associated with advanced tumor grades and poor prognoses [27]. Meanwhile, AURKA enhances cancer metastasis in HCC by promoting epithelial-mesenchymal transition and cancer stem cell behaviors via the PI3K/AKT pathway [28]. Additionally, mutations in DICER1 have been linked to liver carcinogenesis in certain familial and sporadic HCC subtypes due to β-catenin activation [29]. CUL1 is notably upregulated in HCC tissues compared to adjacent non-tumorous tissue, correlating with larger tumor sizes, higher histological grades, advanced TNM stages, and poorer five-year overall and disease-specific survival rates. Furthermore, FEN1 promotes the proliferation, migration, and invasion of HCC cells by facilitating the transition from G2 to M phase of the cell cycle, positioning it as a potential therapeutic target [30]. The functional polymorphism XRCC6 T-991C has also been related to HCC susceptibility, potentially resulting in differential expression levels of XRCC6 mRNA and protein, making it a candidate for early prediction and pharmacogenomic therapy [31]. Additionally, inhibiting HSP90AB1 has been shown to enhance the infiltration of CD8 + T cells and other immune cell populations within HCC tissues, correlating with improved prognosis, thereby suggesting a novel therapeutic approach [32].

Notably, JUN, CCNA1, POLR2C, POLR2H, POLR2F, POLR2L, RPS27A, H2BC14, and H2BC13 have not been previously reported as HCC biomarkers, highlighting their novelty and potential significance in expanding the diagnostic and therapeutic landscape. Furthermore, the integration of tissue-derived pathways with blood-based gene expression profiles underscores the systemic nature of HCC, indicating these biomarkers could significantly improve non-invasive diagnostic strategies for early detection.

To elucidate the cellular sources of the signature genes, we analyzed single-cell RNA-seq data from HCC tumors using the IMMUcan dataset HCC_IMM_10X_GSE140288. This analysis indicated that plasma- and PBMC-derived biomarkers localize to distinct malignant and immune compartments, mirroring the tumor-intrinsic and immune-related biology inferred from bulk, pathway, and hallmark analyses.

The 11 plasma-derived genes, previously selected as hub nodes in plasma–tissue shared pathways enriched for cell cycle, DNA repair, and transcriptional control, exhibited high expression in tumor cell-containing clusters, with minimal expression in major immune subsets. CCND1 and AURKA were most abundant in highly proliferative tumor compartments, supporting their role as markers of cell-cycle dysregulation. POLR2C, POLR2F, and POLR2L were co-expressed in these clusters, consistent with enhanced RNA polymerase II-mediated transcription. Additionally, NPM1 and RPS27A were elevated in metabolically active tumor cells, while FEN1, XRCC6, H2BC13, H2BC14, and HSP90AB1 were enriched in tumor regions with high replicative and stress-related activity. Together, these patterns indicate that the plasma panel predominantly reflects tumor-intrinsic proliferation and genome instability, anchoring bulk-derived signatures in malignant hepatocyte populations.

In contrast, the 6 PBMC-derived biomarkers demonstrated a preferential expression in immune cell subsets. JUN was enriched in T cells, especially activated T-cell subsets, and CCNA1 and AURKA were associated with proliferating T and B cells, indicating that these markers report on lymphocyte activation rather than tumor cell cycling. DICER1, CUL1, and POLR2H were elevated in B cells and myeloid populations, aligning with active RNA processing, protein turnover, and transcription in antigen-presenting and innate immune cells. This immune-restricted expression pattern corresponds with PBMC–tissue shared pathways and hallmark analyses, linking PBMC genes to immune evasion, inflammation, and tissue invasion. These finding echoes prior cancer studies

where single-cell analysis is employed to map hub genes to specific immune subsets within the tumor microenvironment.

The single-cell data provide orthogonal validation of the tumor–immune biomarker architecture. By localizing plasma biomarkers to tumor-rich clusters and PBMC biomarkers to defined immune cell subsets, the single-cell analysis offers independent, cell-resolved evidence supporting the division of the panels into tumor-intrinsic and immune-response components. This multi-scale consistency reinforces the biological plausibility of the biomarker architecture and supports its potential use as a non-invasive tumor–immune axis for HCC detection and monitoring.

MicroRNAs add an additional regulatory layer to this biomarker framework. As stable, cell-free molecules detectable in body fluids, they are well suited for non-invasive diagnostics [33, 34]. It is evident that miRNAs target genes in a systematic manner and show targeting specificity down to the pathway level [35]. The exosomal miRNAs identified in this study reflect systemic regulatory mechanisms that parallel tissue-level changes. For example, hsa-miR-184 has been shown to facilitate migration and proliferation in hepatocellular carcinoma [36]. Likewise, hsa-miR-425-5p targets tumor suppressor pathways, contributing to HCC metastasis [37], while hsa-miR-18a-5p is associated with more aggressive tumor behavior and worse patient outcomes [38]. Additionally, miR-671-5p and miR-663a enhance cell proliferation, invasion, and migration by targeting ALDH2 [39, 40]. miR-193a-5p serves as a crucial regulator of the cell cycle, apoptosis, and metastasis [41], while miR-371a-5p is involved in tumor growth regulation [42]. Furthermore, downregulation of the tumor suppressor EVA1A by miR-103a-3p may act as a key mediator in HCC, presenting new diagnostic and therapeutic avenues [43]. Conversely, miRNAs like hsa-miR-874-3p, which have not been extensively studied in relation to HCC, could serve as novel exosomal biomarkers.

While the findings provide a compelling rationale for multi-compartment biomarker integration, several limitations should be acknowledged. The reliance on publicly available datasets introduces variability due to potential differences in sample processing and data acquisition methods. Therefore, validating these findings in independent cohorts through functional assays remains imperative to establish their clinical utility. Moreover, mechanistic studies dissecting how these gene and miRNA signatures influence hallmark-level processes in HCC will be crucial for translating them into clinically actionable tools. Future research should prioritize the validation of these biomarker panels across larger and more diverse populations while further exploring the mechanistic pathways involving these genes and miRNAs. Such investigations could provide invaluable insights into targeted therapies and improve early detection strategies for HCC, ultimately enhancing patient outcomes.

5. Limitations

Several limitations of this study should be acknowledged. **First**, the reliance on publicly available datasets introduces variability due to potential differences in sample processing, platform-specific normalization, and data acquisition methods across the TCGA-LIHC, GSE14520 (GPL571 and GPL3921), GSE25097, GSE49515, GSE142987, and GSE82303 datasets. These technical differences may influence the reproducibility and comparability of the derived gene signatures. **Second**, the sample sizes of the blood-based datasets were relatively modest. The PBMC dataset (GSE49515) comprised 20 samples (10 tumor and 10 control), and the plasma dataset (GSE142987) and exosomal miRNA dataset (GSE82303) also had limited sample numbers. This restricts the statistical power of the differential expression and network analyses and may reduce the

generalizability of the identified biomarker panels. **Third**, the candidate gene and miRNA biomarkers were identified through integrative bioinformatics analyses and were not experimentally validated in independent clinical cohorts. No functional assays (e.g., qRT-PCR, Western blotting, or miRNA reporter assays) were performed to confirm the expression levels or biological roles of the identified biomarkers. Prospective validation in larger, ethnically diverse patient cohorts with standardized sample collection and processing protocols is essential. **Fourth**, the single-cell RNA-seq analysis was restricted to a single dataset (HCC_IMM_10X_GSE140288, 66,187 cells), which may not fully capture the heterogeneity of the tumor microenvironment across different HCC etiologies (e.g., HBV, HCV, alcohol-related) or disease stages. **Fifth**, the exosomal miRNA biomarker panel was derived from a single dataset (GSE82303) and the miRNA-target interactions were based on experimentally validated entries in miRTarBase. While this ensures high confidence in the predicted interactions, the panel requires validation in independent exosomal miRNA datasets and clinical samples before clinical translation. **Finally**, mechanistic studies dissecting how these gene and miRNA signatures influence hallmark-level processes in HCC will be crucial for translating them into clinically actionable tools. Future research should prioritize the validation of these biomarker panels across larger and more diverse populations while further exploring the mechanistic pathways involving these genes and miRNAs. Such investigations could provide invaluable insights into targeted therapies and improve early detection strategies for HCC, ultimately enhancing patient outcomes.

6. Conclusion

Our integrative analysis highlights distinct yet complementary tumor-specific signals in plasma, PBMCs, and exosomal miRNAs, with strong associations to key hallmarks of cancer. The identified genes and miRNAs represent promising candidates for non-invasive HCC biomarker panels, and their functional diversity underscores the importance of a multi-source approach to capture the full spectrum of tumor biology. Validation of these markers in independent and diverse populations will be a crucial next step toward their incorporation into precision diagnostic strategies.

Abbreviations

HCC
Hepatocellular carcinoma
TCGA
Cancer Genome Atlas
GEO
Gene Expression Omnibus
PBMC
peripheral blood mononuclear cells
WGCNA
Weighted Gene Co-expression Network Analysis
RNA-Seq
RNA sequencing
DEGs
differentially expressed genes
PCA

principal component analysis
DEA
Differential Expression Analysis
MM
Module Membership
ME
module eigengene
GS
Gene Significance
HGNC
The HUGO Gene Nomenclature Committee
PPI
protein–protein interaction
DEMs
differentially expressed miRNAs
LLMs
Large Language Models
QC
Quality control

Declarations

Ethical Approval and Accordance: Not applicable. This study is a secondary analysis of publicly available, de-identified transcriptomic datasets. No human participants or animals were involved, and no new data were generated.

Consent to participate: Not applicable.

Consent to publish: Not applicable.

Clinical trial number: Not applicable.

Competing interests: The authors declare no competing interests.

Funding: The authors did not receive support from any organization for the submitted work.

Authors' contributions: Conceptualization, Methodology, Data curation, Formal analysis, Investigation, Writing — Original Draft, Writing — Review & Editing, Supervision, Project administration: A.G., S.S., Z.G., H.F., Y.N., S.A., M.M. All authors have read and approved the final version of the manuscript.

Use of Artificial Intelligence: Large Language Models (LLMs) were employed solely to correct grammatical errors and improve the readability of the manuscript. Following their use, the authors reviewed the text and take full responsibility for the content.

Data Availability Statement

The datasets analyzed in this study are publicly available from the following repositories:

TCGA-LIHC: RNA-seq data for hepatocellular carcinoma were obtained from The Cancer Genome Atlas Liver Hepatocellular Carcinoma (TCGA-LIHC) cohort via the Genomic Data Commons (GDC) Data Portal (https://portal.gdc.cancer.gov/projects/TCGA-LIHC).

GEO datasets:

- GSE14520 (GPL571 and GPL3921 platforms): https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE14520
- GSE25097: https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE25097
- GSE49515 (PBMC): https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE49515
- GSE142987 (plasma): https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE142987
- GSE82303 (blood exosomal miRNAs): https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE82303

Single-cell RNA-seq dataset: HCC_IMM_10X_GSE140288 was obtained from the IMMUcan single-cell RNA-seq database (https://immuncanscdb.vital-it.ch/) and is available from GEO under accession GSE140288 (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE140288).

Downstream analysis resources:

- miRTarBase 2025 (experimentally validated miRNA-target interactions): https://mirtarbase.cuhk.edu.cn/~miRTarBase/miRTarBase_2025
- Cancer Hallmarks Analytics Tool: https://cancerhallmarks.com
- STRING database (protein-protein interaction analysis): https://string-db.org
- Reactome 2024 pathway database (accessed via Enrichr): https://maayanlab.cloud/Enrichr/

Data were retrieved in December 2024. The scripts and processed data generated during this study are available from the corresponding author upon reasonable request.

References

1. Llovet JM, Kelley RK, Villanueva A, Singal AG, Pikarsky E, Roayaie S et al. Hepatocellular carcinoma. Nat Reviews Disease Primers. 2021 2021/01/21;7(1):6.
2. Singal AG, Kanwal F, Llovet JM. Global trends in hepatocellular carcinoma epidemiology: implications for screening, prevention and therapy. Nat Reviews Clin Oncol. 2023 2023/12/01;20(12):864–84.
3. Singh SP, Madke T, Chand P. Global Epidemiology of Hepatocellular Carcinoma. J Clin Experimental Hepatol. 2025 2025/03/01/;15(2):102446.
4. Chan SL, Sun H-C, Xu Y, Zeng H, El-Serag HB, Lee JM et al. The Lancet Commission on addressing the global hepatocellular carcinoma burden: comprehensive strategies from prevention to treatment. The Lancet.
5. Chidambaranathan-Reghupaty S, Fisher PB, Sarkar D. Hepatocellular carcinoma (HCC): Epidemiology, etiology and molecular classification. Adv Cancer Res. 2021;149:1–61.
6. Tabori NE, Sivananthan G. Treatment Options for Early-Stage Hepatocellular Carcinoma. Semin Intervent Radiol. 2020;37(5):448–55.

7. Forner A, Reig M, Bruix J. Hepatocellular carcinoma. Lancet. 2018;391(10127):1301–14.
8. Johnson P, Zhou Q, Dao DY, Lo YD. Circulating biomarkers in the diagnosis and management of hepatocellular carcinoma. Nat reviews Gastroenterol Hepatol. 2022;19(10):670–81.
9. Chan Y-T, Zhang C, Wu J, Lu P, Xu L, Yuan H, et al. Biomarkers for diagnosis and therapeutic options in hepatocellular carcinoma. Mol Cancer. 2024;23(1):189. 2024/09/06.
10. Wu ST, Zhu L, Feng XL, Wang HY, Li F. Strategies for discovering novel hepatocellular carcinoma biomarkers. World J Hepatol. 2025;17(2):101201.
11. Lee YT, Fujiwara N, Yang JD, Hoshida Y. Risk stratification and early detection biomarkers for precision HCC screening. Hepatology. 2023;78(1):319–62.
12. Maravelia P, Silva DN, Rovesti G, Chrobok M, Stål P, Lu YC et al. Liquid Biopsy in Hepatocellular Carcinoma: Opportunities and Challenges for Immunotherapy. Cancers (Basel). 2021;13(17).
13. Kalaki NS, Ahmadzadeh M, Najafi M, Mobasheri M, Ajdarkosh H, Niya MHK. Systems biology approach to identify biomarkers and therapeutic targets for colorectal cancer. Biochem Biophys Rep. 2024;37:101633.
14. Momeni M, Rashidifar M, Balam FH, Roointan A, Gholaminejad A. A comprehensive analysis of gene expression profiling data in COVID-19 patients for discovery of specific and differential blood biomarker signatures. Sci Rep. 2023;13(1):5599.
15. Hojjati F, Roointan A, Gholaminejad A, Eshraghi Y, Gheisari Y. Identification of key genes and biological regulatory mechanisms in diabetic nephropathy: Meta-analysis of gene expression datasets. Nefrología. 2022.
16. Gholaminejad A, Roointan A, Gheisari Y. Transmembrane signaling molecules play a key role in the pathogenesis of IgA nephropathy: a weighted gene co-expression network analysis study. BMC Immunol. 2021;22(1):1–17.
17. Gholaminejad A, Ghaeidamini M, Simal-Gandara J, Roointan A. An Integrative in silico Study to Discover Key Drivers in Pathogenicity of Focal and Segmental Glomerulosclerosis. Kidney Blood Press Res. 2022:1–13.
18. Roointan A, Ghaeidamini M, Yavari P, Naimi A, Gheisari Y, Gholaminejad A. Transcriptome meta-analysis and validation to discovery of hub genes and pathways in focal and segmental glomerulosclerosis. BMC Nephrol. 2024;25(1):293. 2024/09/04.
19. Mi N, Cao J, Zhang J, Fu W, Huang C, Gao L, et al. Identification of hub genes involved in the occurrence and development of hepatocellular carcinoma via bioinformatics analysis. Oncol Lett. 2020;20(2):1695–708.
20. Naghdibadi M, Momeni M, Yavari P, Gholaminejad A, Roointan A. Clear Cell Renal Cell Carcinoma: A Comprehensive in silico Study in Searching for Therapeutic Targets. Kidney Blood Press Res. 2023;48(1):135–50.
21. Chauhan R, Lahiri N. Tissue- and Serum-Associated Biomarkers of Hepatocellular Carcinoma. Biomark Cancer. 2016;8(Suppl 1):37–55.
22. Debes JD, Romagnoli PA, Prieto J, Arrese M, Mattos AZ, Boonstra A et al. Serum Biomarkers for the Prediction of Hepatocellular Carcinoma. Cancers (Basel). 2021;13(7).
23. Huang H, Zhang M, Lu H, Chen Y, Sun W, Zhu J, et al. Identification and evaluation of plasma exosome RNA biomarkers for non-invasive diagnosis of hepatocellular carcinoma using RNA-seq. BMC Cancer. 2024;24(1):1552. 2024/12/18;.

24. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. BMC Bioinformatics. 2008;9:1–13.
25. Camps J, Noël F, Liechti R, Massenet-Regad L, Rigade S, Götz L, et al. Meta-analysis of human cancer single-cell RNA-seq datasets using the IMMUcan database. Cancer Res. 2023;83(3):363–73.
26. Zhang H. CCND1 silencing suppresses liver cancer stem cell differentiation through inhibiting autophagy. Hum Cell. 2020;33(1):140–7.
27. Yang G, Li H, Dong Z, Deng K, Lu Y. Nucleophosmin 1 associating with engulfment and cell motility protein 1 regulates hepatocellular carcinoma cell chemotaxis and metastasis. Open Med (Wars). 2023;18(1):20230708.
28. Chen C, Song G, Xiang J, Zhang H, Zhao S, Zhan Y. AURKA promotes cancer metastasis by regulating epithelial-mesenchymal transition and cancer stem cell properties in hepatocellular carcinoma. Biochem Biophys Res Commun. 2017;486(2):514–20.
29. Caruso S, Calderaro J, Letouzé E, Nault JC, Couchy G, Boulai A, et al. Germline and somatic DICER1 mutations in familial and sporadic liver tumors. J Hepatol. 2017;66(4):734–42.
30. Wang R, Zhang H, Huang D, Xu J, Zhang Y, Wang T. FEN1 Promotes Hepatocellular Carcinoma Progression by Activating Cell Cycle Transition from G2 To M Phase. J Cancer. 2024;15(4):981–9.
31. Hsu CM, Yang MD, Chang WS, Jeng LB, Lee MH, Lu MC, et al. The contribution of XRCC6/Ku70 to hepatocellular carcinoma in Taiwan. Anticancer Res. 2013;33(2):529–35.
32. Yudong W. HSP90AB1's effect on hepatocellular carcinoma's local immune response and prognosis. Mol Cell Biomech. 2024 08/13;21:138.
33. Roointan A, Gholaminejad A, Shojaie B, Hudkins KL, Gheisari Y. Candidate MicroRNA Biomarkers in Lupus Nephritis: A Meta-analysis of Profiling Studies in Kidney, Blood and Urine Samples. Mol Diagn Ther. 2022 Dec 15.
34. Li C, Zhou T, Chen J, Li R, Chen H, Luo S, et al. The role of Exosomal miRNAs in cancer. J Translational Med. 2022;20(1):6. 2022/01/03.
35. Kern F, Krammes L, Danz K, Diener C, Kehl T, Küchler O, et al. Validation of human microRNA target pathways enables evaluation of target prediction tools. Nucleic Acids Res. 2021;49(1):127–44.
36. Guo X, Wang Z, Deng X, Lu Y, Huang X, Lin J, et al. Circular RNA CircITCH (has-circ-0001141) suppresses hepatocellular carcinoma (HCC) progression by sponging miR-184. Cell Cycle. 2022;21(15):1557–77.
37. Fang F, Song T, Zhang T, Cui Y, Zhang G, Xiong Q. MiR-425-5p promotes invasion and metastasis of hepatocellular carcinoma cells through SCAI-mediated dysregulation of multiple signaling pathways. Oncotarget. 2017;8(19):31745–57.
38. Quan B, Liu W, Yao F, Li M, Tang B, Li J, et al. LINC02362/hsa-miR-18a-5p/FDX1 axis suppresses proliferation and drives cuproptosis and oxaliplatin sensitivity of hepatocellular carcinoma. Am J Cancer Res. 2023;13(11):5590–609.
39. Chen X, Luo J, Gao S, Jiang J, Yang B, Zhang Z. miR-671-5p Promotes Cell Proliferation, Invasion, and Migration in Hepatocellular Carcinoma through Targeting ALDH2. Crit Rev Eukaryot Gene Expr. 2022;32(4):73–82.
40. Huang W, Li J, Guo X, Zhao Y, Yuan X. miR-663a inhibits hepatocellular carcinoma cell proliferation and invasion by targeting HMGA2. Biomed Pharmacother. 2016;81:431–8.

41. Mahboobnia K, Beveridge DJ, Yeoh GC, Kabir TD, Leedman PJ. MicroRNAs in Hepatocellular Carcinoma Pathogenesis: Insights into Mechanisms and Therapeutic Opportunities. Int J Mol Sci. 2024;25(17):9393.
42. Visalli M, Bartolotta M, Polito F, Oteri R, Barbera A, Arrigo R, et al. miRNA expression profiling regulates necroptotic cell death in hepatocellular carcinoma. Int J Oncol. 2018;53(2):771–80.
43. Xu Q, Liao Z, Gong Z, Liu X, Yang Y, Wang Z, et al. Down-regulation of EVA1A by miR-103a-3p promotes hepatocellular carcinoma cells proliferation and migration. Cell Mol Biol Lett. 2022;27(1):93.

Figures

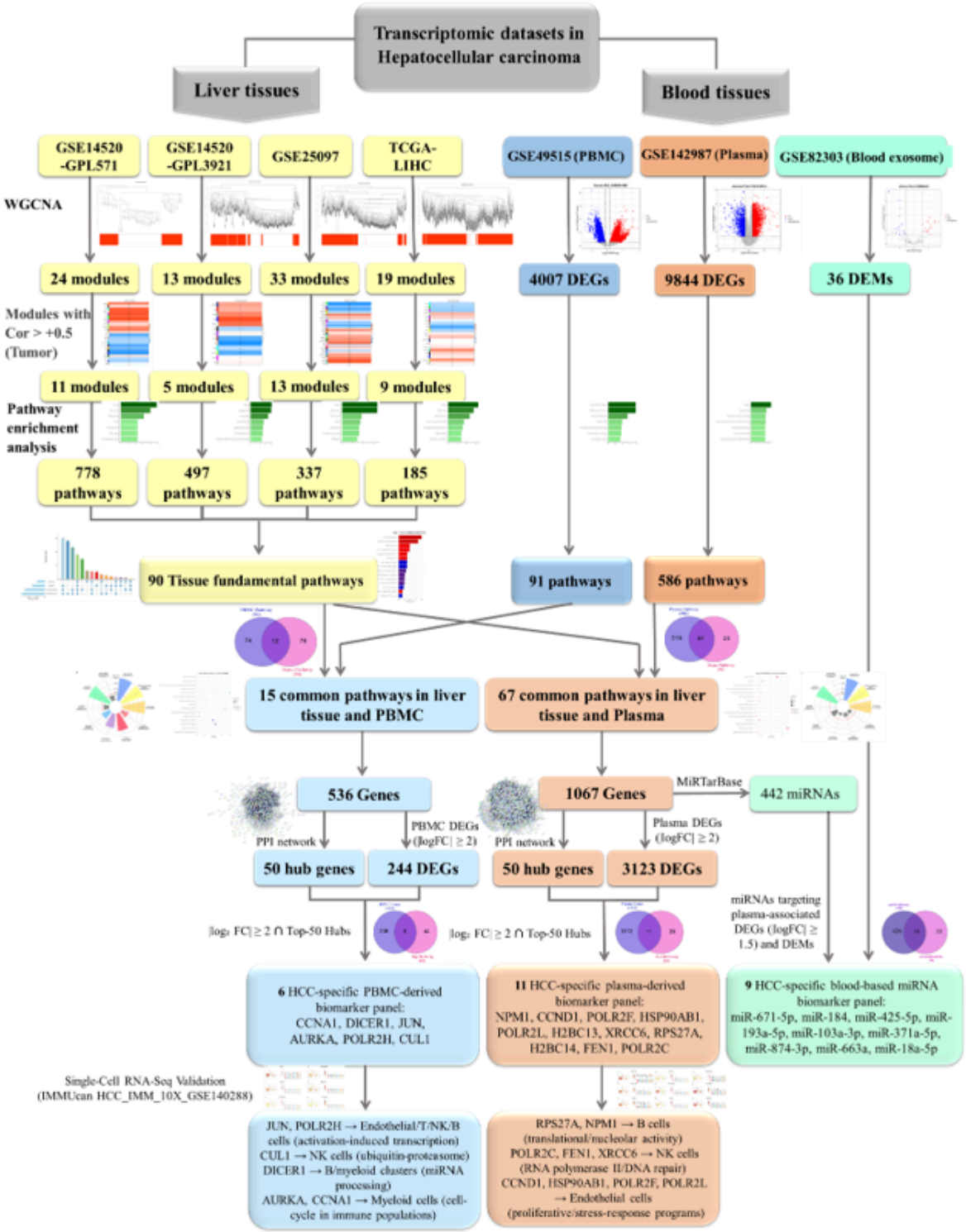


Figure 1

Workflow for the integrative transcriptomic analysis and biomarker discovery in hepatocellular carcinoma (HCC). Transcriptomic datasets from liver tissues (TCGA-LIHC, GSE14520-GPL571, GSE14520-GPL3921, GSE25097) and blood compartments (plasma: GSE142987; PBMCs: GSE49515; blood exosomal miRNAs: GSE82303) were analyzed through differential expression, WGCNA, and pathway enrichment. Tissue datasets yielded 90 fundamental pathways consistently dysregulated in HCC, which were intersected with blood-enriched pathways to identify 15 PBMC–liver shared pathways and 67 plasma–liver shared pathways. These were

mapped to genes, hub genes, and DEGs, leading to the identification of HCC-specific PBMC-derived (6 genes), plasma-derived (11 genes), and exosomal miRNA-derived (9 miRNAs) biomarker panels

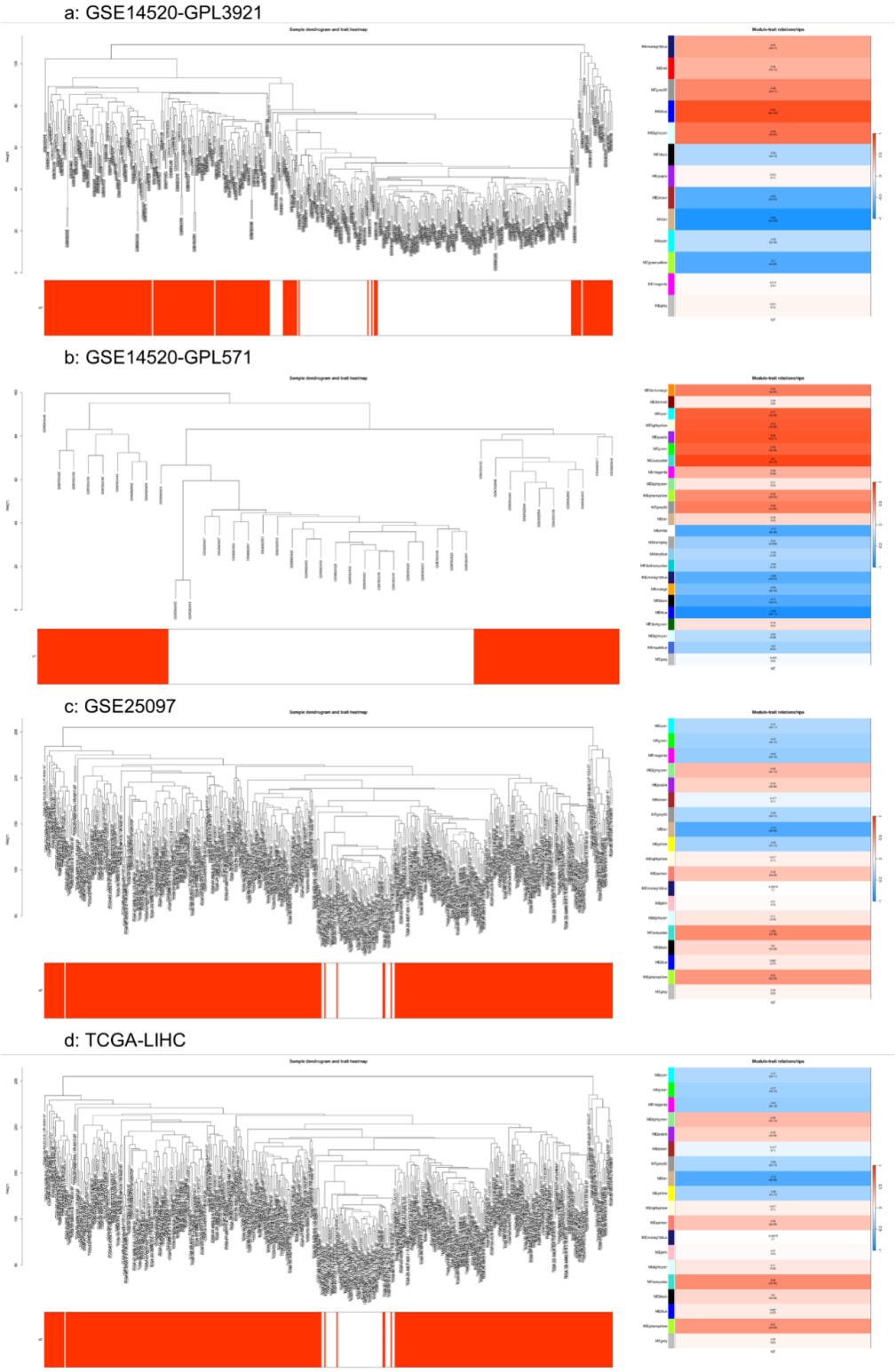


Figure 2

Weighted Gene Co-expression Network Analysis (WGCNA) across four liver cancer datasets. Sample clustering dendrograms and associated trait heatmaps are shown for **(a)** GSE14520-GPL3921, **(b)** GSE14520-GPL571, **(c)** GSE25097, and **(d)** TCGA-LIHC. In each panel, the upper portion displays a hierarchical clustering of samples

based on their gene expression profiles, enabling detection of outliers and major expression-based groupings. The corresponding trait heatmap below each dendrogram shows the distribution of key clinical traits across the clustered samples. To the right of each panel, heatmaps of module–trait relationships are presented, where each row represents a co-expression module (identified by color), and each column corresponds to a clinical trait. The cells contain correlation coefficients and associated p-values, with the color gradient indicating the strength (red: positive, blue: negative) and significance of correlations. These analyses reveal modules significantly associated with clinical parameters, providing insights into biologically relevant gene networks involved in hepatocellular carcinoma pathogenesis

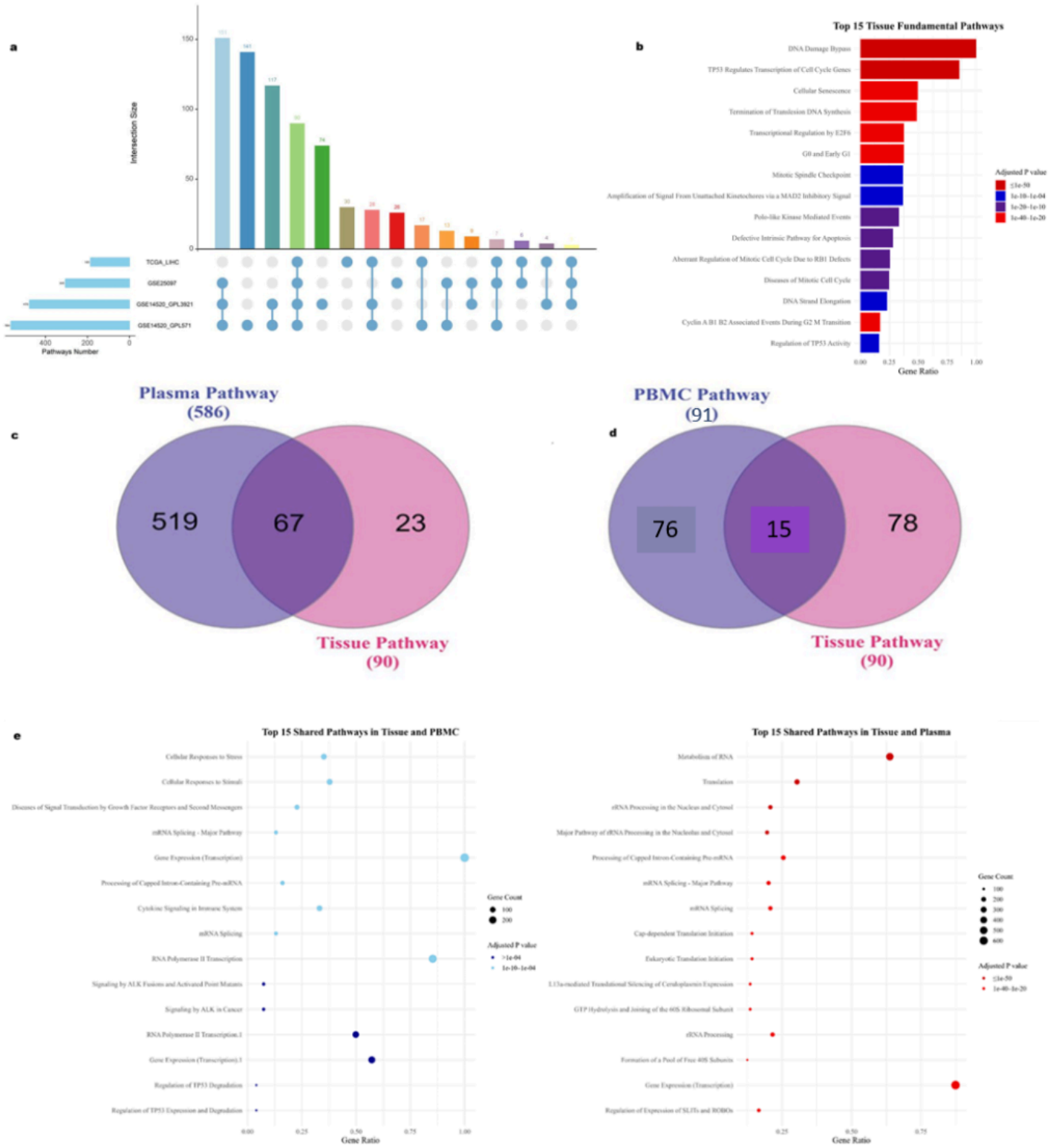


Figure 3

Identification and visualization of tissue fundamental pathways. (a) Upset plot illustrating the overlap of significantly enriched pathways among the four tissue datasets. A total of 90 shared, non-redundant pathways were identified, categorized as “tissue fundamental pathways”. This visualization highlights the complexities of pathway relationships, revealing critical intersections across the datasets”. **(b)** Bar plot showing the top 15 most significantly enriched pathways among the tissue fundamental pathways, ranked by gene ratio. Bar color

intensity corresponds to adjusted p-values, and gene ratio represents the proportion of input genes mapped to each pathway. **(c)** Venn diagram of shared pathways between plasma and tissue datasets. **(d)** Venn diagram of shared pathways between PBMC and tissue datasets. **(e)** Bubble plot of the top 15 pathways shared between tissue and blood (plasma and PBMC), based on Reactome enrichment scores

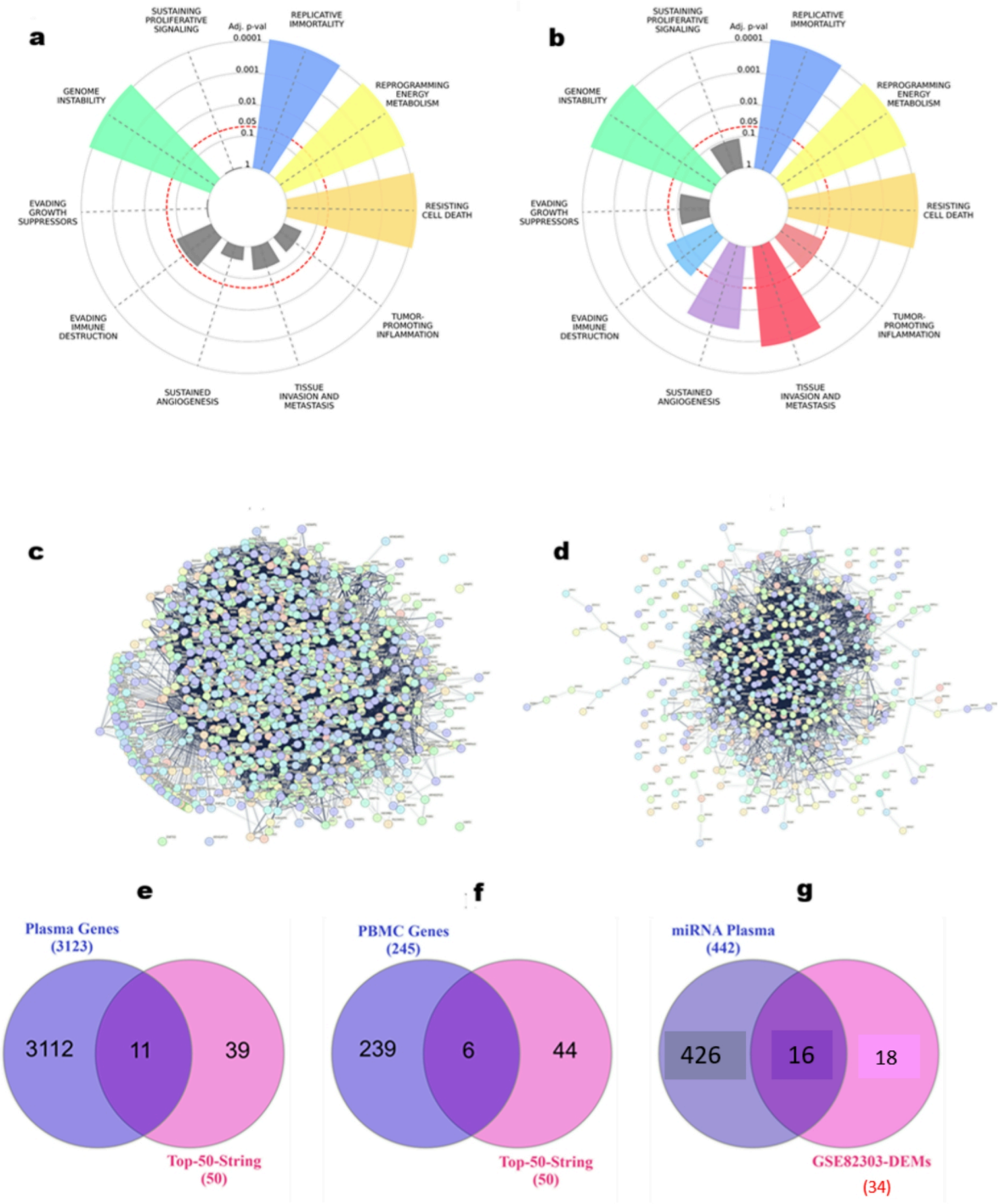


Figure 4

Enrichment of shared-pathway genes in the ten canonical hallmarks of cancer. Radar plots showing hallmark enrichment for **(a)** plasma-associated genes (n = 1,067) and **(b)**PBMC-associated genes (n = 536) derived from pathways shared with HCC tissue datasets. Significance is indicated relative to the red dashed line (adjusted p = 0.05). PBMC, peripheral blood mononuclear cells; HCC, hepatocellular carcinoma. **Network connectivity analysis of DEGs from plasma and PBMC datasets and Overlap analysis for prioritizing candidate blood-derived biomarker panels.** PPI network of DEGs from the **(c)** plasma and **(d)**PBMC datasets that were involved in the shared pathways. Networks were constructed using the STRING database. DEGs, differentially expressed genes; PPI, protein–protein interaction; STRING, Search Tool for the Retrieval of Interacting Genes/Proteins. **(e)** Venn diagram showing the overlap between the top 50 most connected genes from the plasma shared-pathway network and the highly differentially expressed plasma genes. **(f)** Venn diagram showing the overlap between the top 50 most connected genes from the PBMC shared-pathway network and the highly differentially expressed PBMC genes. **(g)** Venn diagram showing the overlap between miRNAs predicted to target plasma-associated DEGs (logFC cutoff of ±1.0) and differentially expressed miRNAs (DEMs). These candidate panels represent promising non-invasive biomarkers reflecting key molecular events associated with HCC development. PBMC, peripheral blood mononuclear cell; DEGs, differentially expressed genes

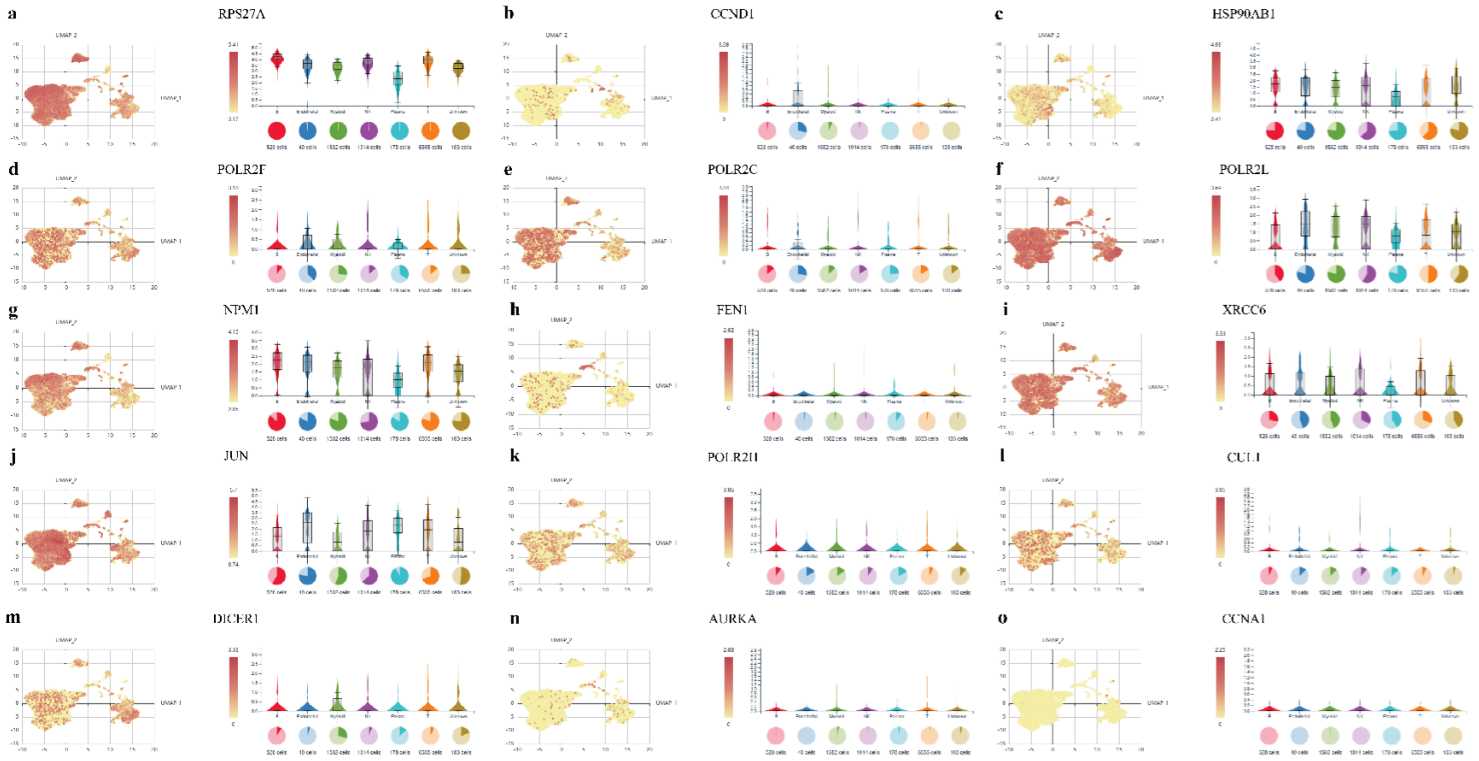


Figure 5

Single-cell RNA-sequencing analysis of plasma-derived biomarker genes across major cell-type clusters (T, B, NK, myeloid, endothelial, plasma, and unknown) (a-i). Single-cell RNA-sequencing analysis of PBMC-derived biomarker genes across major cell-types clusters (T, B, NK, myeloid, endothelial, plasma, and unknown) (j-o) **in HCC tumors.**

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- OnlineResource1.pdf
- OnlineResource2.xlsx