BMC Cancer

RESEARCH

Open Access

Multi-omics SMR and experimental supportive analyses decipher causal drivers hepatocellular carcinoma

Zhiya Yang[1†], Tingyang Li[2†], Jiayun Shen[2†], Yao Huang[3], Chao Lei[4], Jiarui Yu[1], Zhaochen Ma[2], Ming Zhang[1*] and Ying Li[5*]

Abstract

Background Hepatocellular carcinoma (HCC) is a highly prevalent and fatal digestive system malignancy, challenging to treat due to its latent onset and non-specific symptoms in advanced stages. Somatic mutations play a crucial role in hepatocarcinogenesis, with nearly half of HCC patients carrying oncogenic driver mutations such as TP53, CTNNB1, or TERT. In parallel, germline susceptibility variants identified by genome-wide association studies (GWAS) — including loci near TERT, MBOAT7, TM6SF2, and PNPLA3 — reveal inherited predisposition that shapes the molecular landscape for HCC development. Despite recent therapeutic advancements, long-term survival remains suboptimal, necessitating a deeper understanding of its pathogenesis and the identification of precise molecular targets. Traditional genomic studies, such as genome-wide association studies (GWAS), have successfully identified associated variants; however, due to their statistical design, they do not provide direct causal inference, functional supportive analyses, or comprehensive insight into multi-level molecular regulation and tumor microenvironment heterogeneity, serving instead as a critical starting point for subsequent functional and integrative analyses.

Methods To address these gaps, this study employed an integrated multi-omics approach combining HCC GWAS summary data (FinnGen) with expression (eQTL from GTEx V8), methylation (mQTL), and protein (pQTL from ARIC, UKBPPP, DECODE) quantitative trait loci data. We utilized Summary-data-based Mendelian Randomization (SMR) to infer causal associations between molecular traits and HCC risk, prioritizing candidates with higher clinical translation potential. To refine SMR-based prioritization of candidate genes, bulk transcriptome sequencing and ELISA-based quantification were performed as complementary analyses on peripheral blood samples from 10 HCC patients and 10 healthy controls. Following SMR-based gene prioritization, bulk transcriptome and spatial transcriptomic analyses were first used to refine candidate selection and guide subsequent quantification, thereby avoiding unnecessary assays and optimizing the use of clinical samples and research resources. These analyses aimed to assess whether

[†]Zhiya Yang, Tingyang Li and Jiayun Shen are contributed equally to this work.

*Correspondence:
Ming Zhang
zyyzhangming@163.com
Ying Li
liying1251@163.com

Full list of author information is available at the end of the article

expression changes were directionally consistent with eQTL and pQTL effects, providing supportive—rather than confirmatory—evidence for the inferred genetic associations. Spatial transcriptomics was applied to HCC tissue sections to map region-specific expression patterns of candidate genes. Finally, publicly available single-cell RNA sequencing (scRNA-seq) data was analyzed to resolve cell composition changes, cell-type-specific expression, and intercellular communication networks within the HCC tumor microenvironment.

Results Multi-omics SMR analysis identified numerous loci causally associated with HCC risk, with eQTL SMR revealing enrichment in critical cancer pathways ("Signal transduction," "Cancer: overview," "Immune system"). A robust and replicated causal signal for proteins was found on chromosome 19 across three independent pQTL cohorts, with a secondary signal on chromosome 2. The intersection of mQTL, eQTL, and pQTL SMR analyses yielded a core set of 16 candidate genes. Peripheral blood transcriptome profiling showed a clear separation between HCC and controls, with 13 of these 16 genes (e.g., LY9, ST6GAL1, SHMT1) significantly differentially expressed. ELISA validated elevated protein levels of ST6GAL1, PSMB1, LY9, and JUND, and decreased SOD4 in HCC patients. Spatial transcriptomics revealed significant intra-tumoral heterogeneity and distinct, localized expression patterns for genes like ST6GAL1, LGALS1, and JUND. Single-cell RNA sequencing unveiled shifts in cell type composition (e.g., increased Cytotoxic CD4 + T cells and MDSCs, decreased hepatocytes in tumors), cell-type specific expression of candidate genes, and complex intercellular communication networks.

Conclusion By integrating germline (GWAS-based) and somatic evidence, this study provides a comprehensive view of HCC pathogenesis. This integrated strategy successfully identified a core set of genes and proteins with potential causal links to HCC, elucidating their functional convergence in cancer biology. These findings offer novel molecular insights and candidate targets for precise diagnosis, prognostic assessment, and targeted therapy of HCC, laying a solid foundation for future translational research.

Introduction

Liver cancer is a malignant tumor of the digestive system with a high incidence rate and mortality rate, and it is one of the most difficult cancers to treat [1]. Every year, there are 870,000 new cases of liver cancer worldwide, and 760,000 deaths. Some studies indicate that the five-year survival rate of patients is only around 5% to 30%. If no urgent intervention measures are taken to reverse this trend, the number of liver cancer cases and deaths in the next 25 years may double. Liver cancer has a latent onset, and early-stage liver cancer often has no obvious symptoms. In the middle and advanced stages, the clinical manifestations often lack specificity and are easily overlooked or misdiagnosed. More and more studies are focusing on the diagnosis and screening of liver cancer [2]. Both inherited genetic susceptibility and somatic oncogenic alterations contribute to hepatocellular carcinoma (HCC) development. Nearly half of HCC patients harbor somatic driver mutations such as TP53, CTNNB1, or TERT, while germline variants identified through GWAS — including loci at TERT, MBOAT7, TM6SF2, and PNPLA3 — reflect inherited predisposition that shapes baseline susceptibility and influences how hepatocytes respond to carcinogenic insults [3–6].

Many studies have shown that genetic inheritance is closely related to the occurrence of hepatocellular carcinoma [7, 8]. Nearly half of HCC patients carry at least one frequent oncogenic mutation, such as TP53, CTNNB1 or TERT. Despite significant advancements in HCC treatment strategies in recent years, the long-term survival rate remains suboptimal, underscoring the urgent need for a deeper understanding of its complex pathogenesis and the identification of precise molecular targets [9–11]. Integrating germline causal inference with somatic mutation–based frameworks provides complementary insight, bridging the gap between inherited risk and tumor-acquired alterations in hepatocarcinogenesis.

While somatic mutations such as *TP53*, *CTNNB1*, and *TERT* have been well established as tumor-acquired drivers that promote malignant transformation, they do not fully explain why only certain individuals develop HCC under similar environmental exposures or viral infections. Inherited germline variants, by contrast, shape the baseline susceptibility and molecular context that influence how hepatocytes respond to carcinogenic insults. Therefore, integrating germline-level causal inference with molecular QTL data enables us to bridge the gap between inherited risk and tumor-acquired mutations, providing a complementary perspective to the somatic mutation–driven understanding of HCC.

Previous genomic studies, particularly genome-wide association studies (GWAS), have successfully identified numerous genetic variants associated with HCC risk. However, these studies primarily focused on associative analyses, often failing to directly elucidate the causal relationships between genetic variants and disease. GWAS findings generally reveal correlations rather than causations, and the biological mechanisms linking non-coding variants to gene function or protein activity remain obscure. They also fell short in comprehensively

explaining how these genetic signals drive HCC development through multi-level molecular regulation (e.g., gene expression, DNA methylation, and protein abundance). Furthermore, traditional bulk tissue analyses frequently overlooked the intricate cellular heterogeneity of the tumor microenvironment and its critical role in disease progression. Consequently, although GWAS has provided valuable association maps, it lacks the power to determine whether these associations are truly causal, limiting their translational potential. There was also a notable lack of multi-dimensional supportive analyses of key target in both the circulatory system and the in situ tumor microenvironment. These limitations collectively pose significant challenges in sifting through vast genetic data to identify truly causal driver genes and proteins and translating them into effective clinical applications.

To overcome these inherent limitations of traditional GWAS, our study introduces an innovative causal inference framework based on Summary-data-based Mendelian Randomization (SMR). SMR leverages genetic variants as instrumental variables to infer causal associations between molecular traits (such as gene expression, DNA methylation, and protein levels) and HCC risk, This approach bridges genetic associations with biological mechanisms, revealing molecular alterations that are not only correlated with but also potentially drive hepatocarcinogenesis.

Specifically, our study first integrated HCC GWAS summary data from the FinnGen consortium, expression quantitative trait loci (eQTL) data from the GTEx V8 project, and methylation quantitative trait loci (mQTL) data. Critically, to address the deficiency of limited coverage from single-omics data, we further incorporated protein quantitative trait loci (pQTL) data from three independent cohorts: ARIC, UK Biobank Pharma Proteomics Project (UKBPPP), and DECODE. This multi-omics SMR framework represents a substantial methodological advance beyond traditional GWAS by combining genomic, epigenomic, and proteomic evidence to enhance both the robustness and causal credibility of detected signals.

Moreover, unlike purely statistical analyses, we complemented our SMR-based causal inference with comprehensive experimental supportive analyses. To explore the clinical and biological relevance of SMR-identified targets, we performed peripheral blood transcriptome sequencing, ELISA-based protein quantification, spatial transcriptomics, and single-cell RNA sequencing. These multi-level supportive analyses ensure that the causal signals inferred from genetic data correspond to real biological differences at the transcriptional, proteomic, and spatial cellular levels. Together, these innovations establish a new integrative paradigm for uncovering truly causal driver genes in HCC, fundamentally surpassing the descriptive limitations of conventional GWAS.

Through this integrated strategy of multi-omics causal inference and multi-dimensional experimental supportive analyses, our study successfully identified a core set of genes and proteins with potential causal links to HCC, and elucidated their functional convergence in cancer biology. These findings not only provide novel molecular insights and candidate targets for precise diagnosis, prognostic assessment, and targeted therapy of HCC but also lay a solid foundation for future translational research in HCC.

Methods

Data sources for HCC GWAS and Multi-Omics QTLs

Summary-level data for hepatocellular carcinoma (HCC) were obtained from the FinnGen consortium (release R12, ID: finngen_R12_C3_HEPATOCELLU_CARC_EXALLC). For causal inference, we integrated multi-omics quantitative trait loci (QTL) data. Expression quantitative trait loci (eQTL) summary data ($P < 1 \times 10^{-5}$) were obtained from the GTEx V8 project, integrating cross-tissue eQTL effects (pan-tissue) to capture systemic regulatory influences beyond liver-specific expression. This design reflects the systemic and metastatic characteristics of hepatocellular carcinoma and enables identification of causal regulatory variants with multi-tissue impact. Methylation QTL (mQTL) data were sourced from. Protein QTL (pQTL) data were consolidated from three independent studies: the Atherosclerosis Risk in Communities (ARIC), UK Biobank Pharma Proteomics Project (UKBPPP), and DECODE studies.

To ensure data compatibility and reduce population or technical bias, all GWAS and QTL summary statistics were derived from European ancestry cohorts. Prior to SMR analysis, all datasets were harmonized to the same human genome reference build (GRCh38/hg38) to guarantee coordinate alignment, as mismatched genome assemblies (e.g., hg19 vs. hg38) would result in computational errors [12]. Furthermore, SNPs were quality-controlled to remove ambiguous alleles and strand inconsistencies, and allele frequencies were cross-checked across datasets to ensure consistency before integration.

Causal inference and functional annotation

We performed Summary-data-based Mendelian Randomization (SMR) using the SMR package (v.1.3.1) to identify potential causal links between molecular traits and HCC risk (significance threshold: p_SMR < 0.05). Given the exploratory and discovery-oriented nature of this study, we applied a nominal threshold of p_SMR < 0.05 to identify putatively causal associations. HEIDI heterogeneity filtering, a standard component of

the SMR framework, was not applied here to maximize sensitivity for candidate discovery and facilitate subsequent biological validation.

Transcriptome sequencing of peripheral blood samples

Peripheral blood was collected from 10 HCC patients and 10 healthy controls. Following quality assessment, total RNA was extracted and used for library preparation. Eukaryotic mRNA was enriched using Oligo(dT) magnetic beads, fragmented, and reverse-transcribed to synthesize first-strand cDNA. After second-strand synthesis, the double-stranded cDNA underwent end-repair, A-tailing, and ligation to sequencing adapters. Fragments of approximately 300 bp were size-selected and amplified via PCR. Library quality and concentration (> 2 nM) were verified using an Agilent 2100 Bioanalyzer and qRT-PCR. Qualified libraries were pooled and sequenced on an Illumina platform.

Differential expression and bioinformatics analysis

After quality control of raw sequencing data, normalized expression data were used for Principal Component Analysis (PCA) to assess sample clustering. Differential expression analysis was performed to identify differentially expressed genes (DEGs) between HCC and control cohorts. The results were visualized using heatmaps and volcano plots generated in R. Multiple testing correction was performed using the Benjamini–Hochberg false discovery rate (FDR) procedure to control for type I errors. Adjusted p-values (FDR < 0.05) were considered statistically significant. Effect sizes were evaluated using Cohen's d for gene expression differences and standardized mean differences (SMD) for protein abundance.

Spatial transcriptomics analysis of HCC tissue

Spatial transcriptomics was conducted on an HCC tissue section using the 10x Genomics Visium platform. Raw data were processed with the Seurat R package (v4.0). After creating a Seurat object, we performed quality control, filtering out spots with fewer than 200 or more than 8000 detected features. Data were normalized, and highly variable genes were identified. We then performed Principal Component Analysis (PCA), followed by Uniform Manifold Approximation and Projection (UMAP) and unsupervised clustering. The SpatialFeaturePlot function was used to visualize the expression of candidate genes on the tissue section.

Protein quantification by ELISA

Targeted protein quantification for candidate genes was performed using enzyme-linked immunosorbent assay (ELISA) kits. Human ST6GAL1 ELISA Kit, Human SOD4 ELISA Kit, Human PSMB1 ELISA Kit, Human Ly9 ELISA Kit, and Human JUND ELISA Kit were all purchased from Jingmei Biotechnology (Jiangsu, China). Briefly, samples and standards were added to antibody-coated microplates. After incubation, HRP-conjugated detection antibodies were added, followed by substrate incubation. The reaction was terminated, and absorbance was measured at 450 nm. Protein concentrations were calculated based on standard curves.

Single-Cell RNA sequencing data analysis

Single-cell RNA sequencing (scRNA-seq) data (GSE245906) were obtained from the Gene Expression Omnibus (GEO) database. The dataset comprised cells from both "NonTumoral" and "Tumor" samples. Following standard preprocessing, which included quality control, normalization, and batch correction, the data were integrated into a Seurat object. Cell types were annotated based on known marker genes, and dimensionality reduction techniques such as UMAP and t-SNE were employed to visualize cell population distributions. Cellular proportions were calculated to assess differences in cell type composition between "NonTumoral" and "Tumor" samples. The expression patterns of candidate genes were profiled across various cell types using dot plots and visualized on t-SNE plots. Intercellular communication networks within the HCC microenvironment were analyzed using the CellChat R package to infer and visualize ligand-receptor interactions between different cell populations.

Statistical analysis

All statistical analyses were performed using R software (version 4.2.0). Summary-data-based Mendelian Randomization (SMR) analyses were conducted using the SMR package (v.1.3.1) with a nominal threshold of p_SMR < 0.05 to maximize discovery sensitivity. While we applied a nominal p-value threshold to maximize sensitivity in this discovery phase, we complemented this with stringent biological supportive analyses across multiple orthogonal platforms (cross-omics convergence, independent cohort replication, and experimental confirmation) to control false discovery.

Differential gene expression was analyzed using DESeq2 (v.1.34.0) with Benjamini-Hochberg FDR correction (adjusted $p < 0.05$). Effect sizes were evaluated by log2 fold-change, with |log2FC| > 1 considered biologically meaningful. ELISA data were compared using two-tailed Student's t-tests or Mann-Whitney U tests after normality assessment (Shapiro-Wilk test), with $p < 0.05$ considered significant.

Results

Integrated multi-omics SMR identifies causal loci and convergent candidate genes in hepatocellular carcinoma

To elucidate causal molecular drivers of hepatocellular carcinoma (HCC), we applied an integrated multi-omics

framework combining GWAS summary statistics with methylation (mQTL), expression (eQTL), and protein (pQTL) quantitative trait loci through summary-data-based Mendelian randomization (SMR).

As shown in Fig. 1A–B, the genome-wide SMR analyses of mQTL and eQTL datasets identified widespread associations across chromosomes. A total of 93,011 mQTL sites were analyzed, yielding 4,487 significant associations and 1,606 mQTL-significant genes, while 15,639 eQTL loci produced 841 significant associations and 696 eQTL-significant genes.

At the protein level, pQTL-SMR analyses across three independent cohorts—ARIC (1,616 sites, 72 significant loci), DECODE (1,739 sites, 84 loci), and UKBPPP (2,010 sites, 93 loci)—revealed consistent causal protein signals (Fig. 1C–E). Integrative comparison of mQTL-, eQTL-, and pQTL-derived gene sets identified 16 overlapping genes exhibiting convergent evidence across all molecular layers (Fig. 1F). These candidates, including *ST6GAL1*, *PSMB1*, *JUND*, *LY9*, *SHMT1*, *and LGALS1*, represent high-confidence causal mediators in HCC pathogenesis. Complete SMR statistical outputs are provided in Supplement File.

Differential gene expression analyses support SMR-prioritized candidate genes

Principal component analysis (PCA) demonstrated a clear separation between HCC patients and healthy controls based on peripheral blood transcriptomic profiles (Fig. 2A). A total of 2,746 differentially expressed genes (DEGs) were identified (adjusted $p < 0.05$, $|\log_2 FC| > 1$), with a distinct chromosomal distribution pattern (Fig. 2B). Heatmap and volcano plot analyses (Fig. 2C–D) revealed significant upregulation of *ST6GAL1*, *PSMB1*, *JUND*, and *LY9*, consistent with SMR predictions. Among the 16 SMR-prioritized genes, 13 showed differential expression in peripheral blood (Fig. 3).

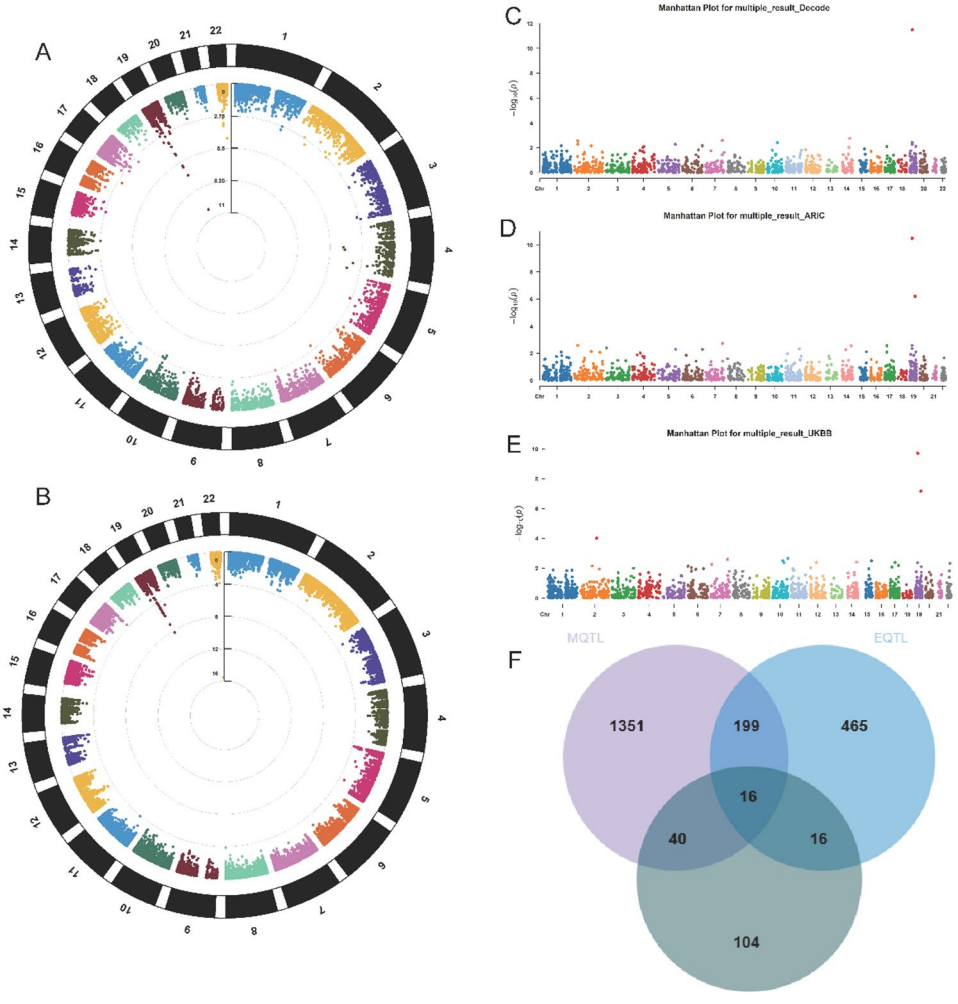

Fig. 1 Summary-data-based Mendelian Randomization identifies multi-omics causal loci and genes associated with HCC. **A–B** Circular Manhattan plots of genome-wide SMR results from mQTL and eQTL analyses. **C–E** Manhattan plots from pQTL SMR in ARIC, DECODE, and UKBPPP cohorts. **F** Venn diagram illustrating overlaps across mQTL (1,606), eQTL (696), and pQTL (176) gene sets

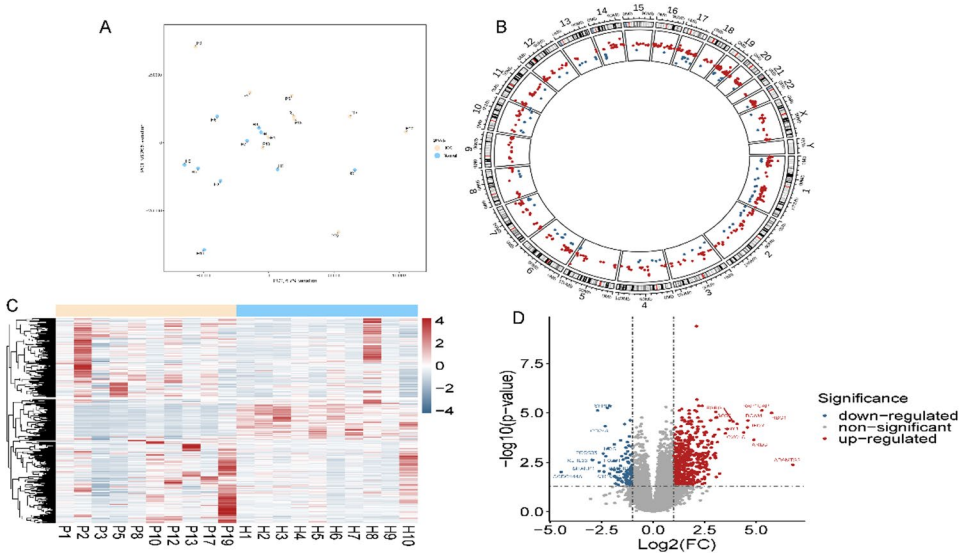

Fig. 2 Peripheral blood transcriptome reveals distinct gene expression profiles between HCC and controls. **A** PCA plot of 10 HCC and 10 controls. **B** Chromosomal DEG distribution. **C** Heatmap of DEGs. **D** Volcano plot showing significant up- and down-regulated genes

Spatial transcriptomics refines candidate gene localization and microenvironmental context

Spatial transcriptomics revealed substantial intratumoral heterogeneity in gene expression across HCC tissue sections. Genes such as *ST6GAL1*, *PSMB1*, *LGALS1*, *LY9*, and *JUND* exhibited regionally localized expression patterns within tumor subregions (Figs. 4 and 5). These findings indicate that tumor microenvironmental context drives spatially confined expression of key causal genes. The protein quantification based on enzyme-linked immunosorbent assay (ELISA) provided strong evidence for the SMR gene, indicating that there is a largely consistent directional relationship between the predicted gene results and the observed expression differences (Fig. 6). Supplement Fig. 1 provides extended supportive analyses data, demonstrating reproducibility and interindividual variation across samples.

Single-cell RNA sequencing reveals cell-type–specific expression and immune remodeling

Single-cell RNA sequencing (scRNA-seq) identified 14 major cell clusters in tumor and non-tumor samples, including hepatocytes, Kupffer cells, endothelial cells, T cells, and MDSCs (Fig. 7A–C). The tumor samples exhibited increased cytotoxic $CD4^+$ T cells and MDSCs, with reduced hepatocyte proportion. Candidate genes such as *ST6GAL1*, *JUND*, *PSMB1*, and *LY9* showed cell-type–specific expression: *ST6GAL1* was enriched in macrophages and tumor endothelial cells, while *JUND* and *PSMB1* were predominantly expressed in proliferative hepatocytes and activated immune cells.

CellChat analysis demonstrated that these genes are involved in critical ligand–receptor interactions regulating immune suppression and metabolic signaling (Supplement Figs. 2 and 3).

Discussion

This study applied an integrated multi-omics strategy combining GWAS, eQTL, mQTL, and pQTL data through Summary-data-based Mendelian Randomization (SMR) to identify potential causal genes and proteins associated with hepatocellular carcinoma (HCC). The integration of multi-level omics and experimental supportive analysis provides robust causal inference and biological interpretation for novel molecular targets.

Our analyses revealed strong causal signals across multiple QTL layers, particularly a consistent protein-level association on chromosome 19 validated in three independent cohorts (ARIC, DECODE, UKBPPP). Functional enrichment demonstrated that these candidate genes were primarily involved in "Signal transduction," "Immune system," and other canonical cancer pathways, indicating that genetically driven dysregulation of molecular traits contributes to HCC pathogenesis. Integrating mQTL, eQTL, and pQTL results identified a core set of 16 genes, including *JUND*, *TDRKH*, *ST6GAL1*, *PSMB1*, and *LY9*, whose biological significance was further validated at transcriptomic and proteomic levels.

JUND and *TDRKH* have been identified as key driver genes that may act on convergent oncogenic pathways. As a core component of the AP-1 transcription factor, *JUND* is upregulated at both mRNA and protein levels in HCC patients, and is enriched in tumor subregions. It can activate the JNK/MMP9 signaling pathway, promoting extracellular matrix degradation, migration, and invasion [13, 14]. Although *TDRKH* has been less studied

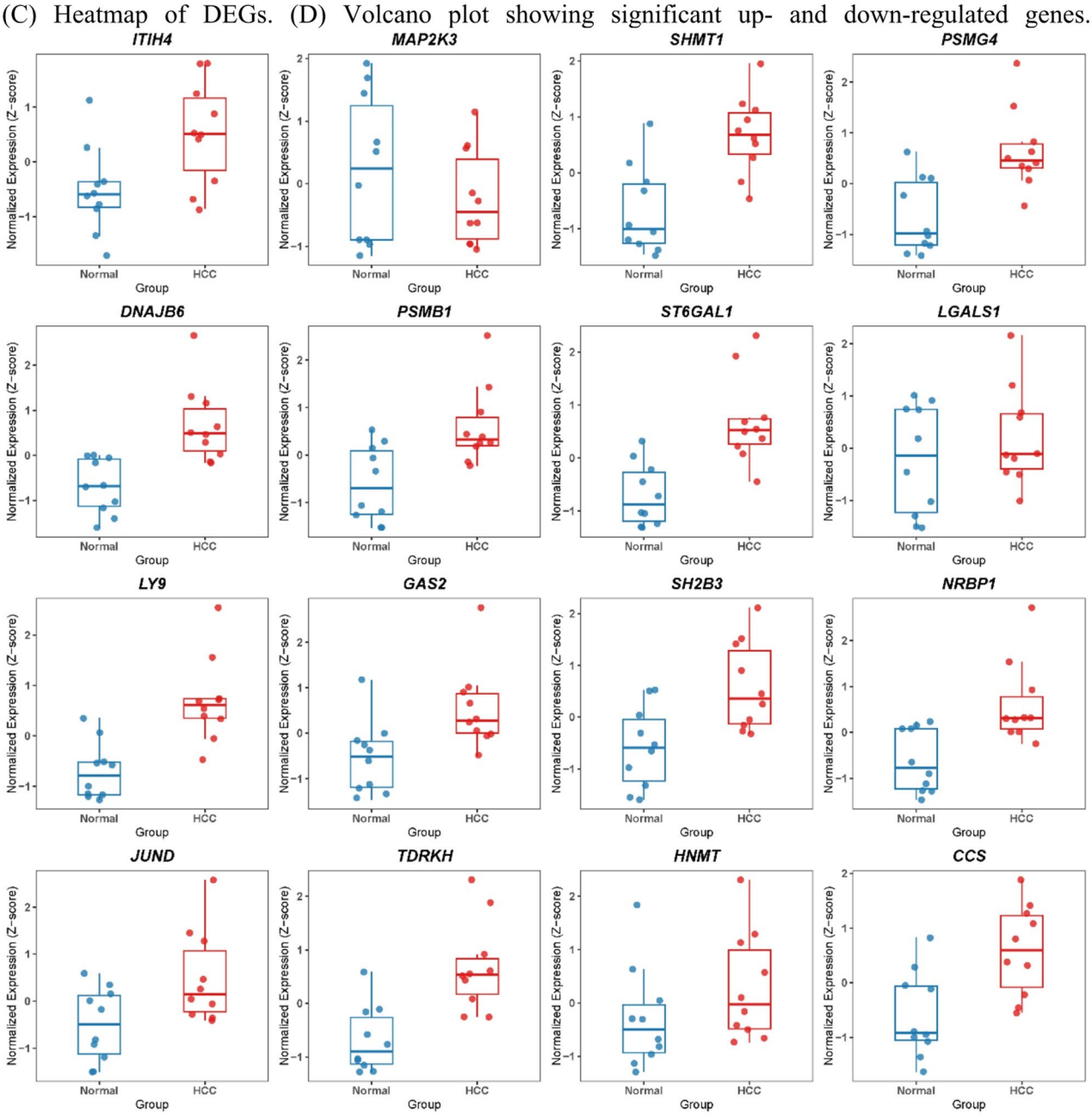

Fig. 3 Supportive analyses of 16 SMR candidate genes in peripheral blood. Boxplots showing normalized expression (Z-score) for each candidate gene between HCC and normal groups

in HCC, multi-omics Mendelian randomization analysis has classified it as a core oncogenic gene. Its antisense transcript *TDRKH-AS1* can promote proliferation and inhibit apoptosis through the Akt signaling pathway, suggesting that the sense and antisense transcripts share a common survival-promoting function [11]. These findings are consistent with the stress-related and Akt activation signals observed in single-cell studies [14].

ST6GAL1 exhibits context-dependent functions: it can promote proliferation and invasion through the PI3K/Akt pathway, and can also inhibit metastasis by lysophosphatidic acidification mediated by *MCAM* [15–17]. In this study, *ST6GAL1* was significantly upregulated at both transcriptional and protein levels, and showed spatial enrichment in tumor subregions, suggesting that its function is regulated by the microenvironment. As a key sialyltransferase, *ST6GAL1* can stabilize PD-L1 through α2,6-sialylation [18], This promotes immune evasion and regulation of the tumor microenvironment. Its dual effects suggest that circulating *ST6GAL1* has diagnostic

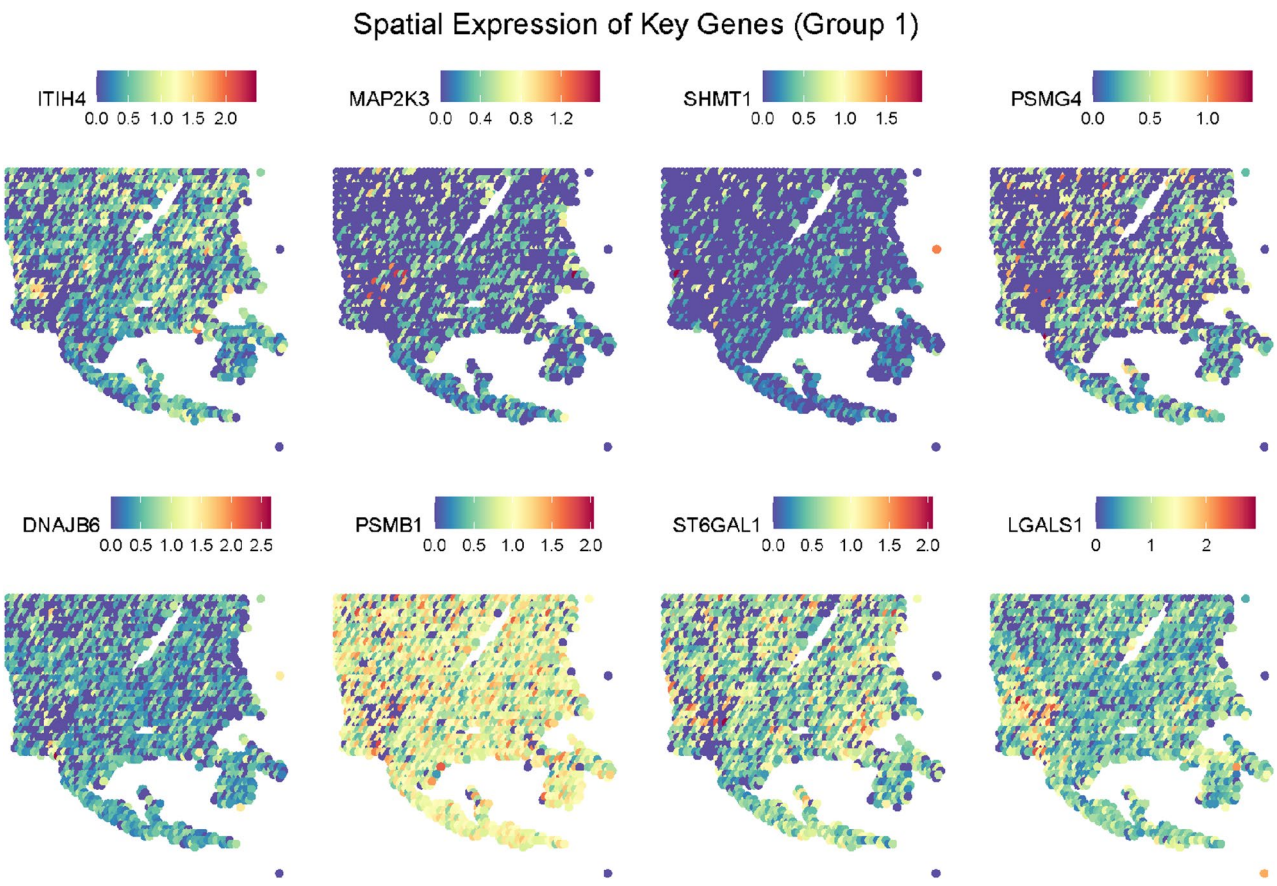

Fig. 4 Spatial transcriptomics reveals heterogeneous and localized expression of candidate genes in HCC tissue. Spatial feature plots of ITIH4, MAP2K3, SHMT1, PSMG4, DNAJB6, PSMB1, ST6GAL1 and LGALS1 demonstrating localized expression in tumor subregions

potential, while the tumor-intrinsic activity may be jointly shaped by the PI3K/Akt signaling and immune-matrix interactions in the immunosuppressive niche [19, 20].

Through multi-omics SMR analysis and proteomic supportive analysis, *PSMB1*, a proteasome β1 subunit, was identified as upregulated in HCC. Although direct evidence for *PSMB1* in HCC remains limited, studies on related proteasome components (e.g., *PSMB8*, *PSMD13*) highlight the proteasome's role in metabolic reprogramming, immune evasion, and proteostasis maintenance [21]. Our findings extend this concept, suggesting that *PSMB1* may contribute to HCC progression by regulating metabolic enzyme stability, redox balance, and immune microenvironment remodeling [22, 23].

LY9 exhibits tissue-specific expression differences: it is significantly upregulated in peripheral blood, but significantly downregulated in liver tissue. This pattern is consistent with previous reports, which regarded *SLAMF3/LY9* as a tumor suppressor receptor for hepatocytes [24]. The upregulation of *LY9* in peripheral blood is more likely to reflect changes in the ratio and function of immune cells (such as activated $CD4^+$ T cells), rather than the expression of liver cells themselves [25]. Therefore, this phenomenon suggests that the systemic immune status should be considered when interpreting key target changes. Overall, the changes in peripheral *LY9* can serve as a powerful supplement to tissue-level diagnosis and enhance its potential application value in immune monitoring [26].

Among the five genes validated by ELISA (ST6GAL1, CCS, PSMB1, LY9, and JUND), their selection was based on significant differential expression at the transcriptome level and spatial transcriptomics co-localization within tumor subregions, consistent with previous integrative omics screening strategies reported in liver cancer studies. Notably, although SOD4 (regulated by CCS) showed upregulated gene expression, its protein abundance decreased, a discrepancy likely explained by post-transcriptional and oxidative stress–induced mechanisms such as impaired copper chaperone activity, enhanced protein degradation, or miRNA-mediated suppression. These mechanisms have been widely reported in oxidative stress and neurodegeneration models [27–29].

By integrating multi-omics causal inference with transcriptomic, proteomic, and spatial supportive analysis, this study advances our understanding of HCC biology.

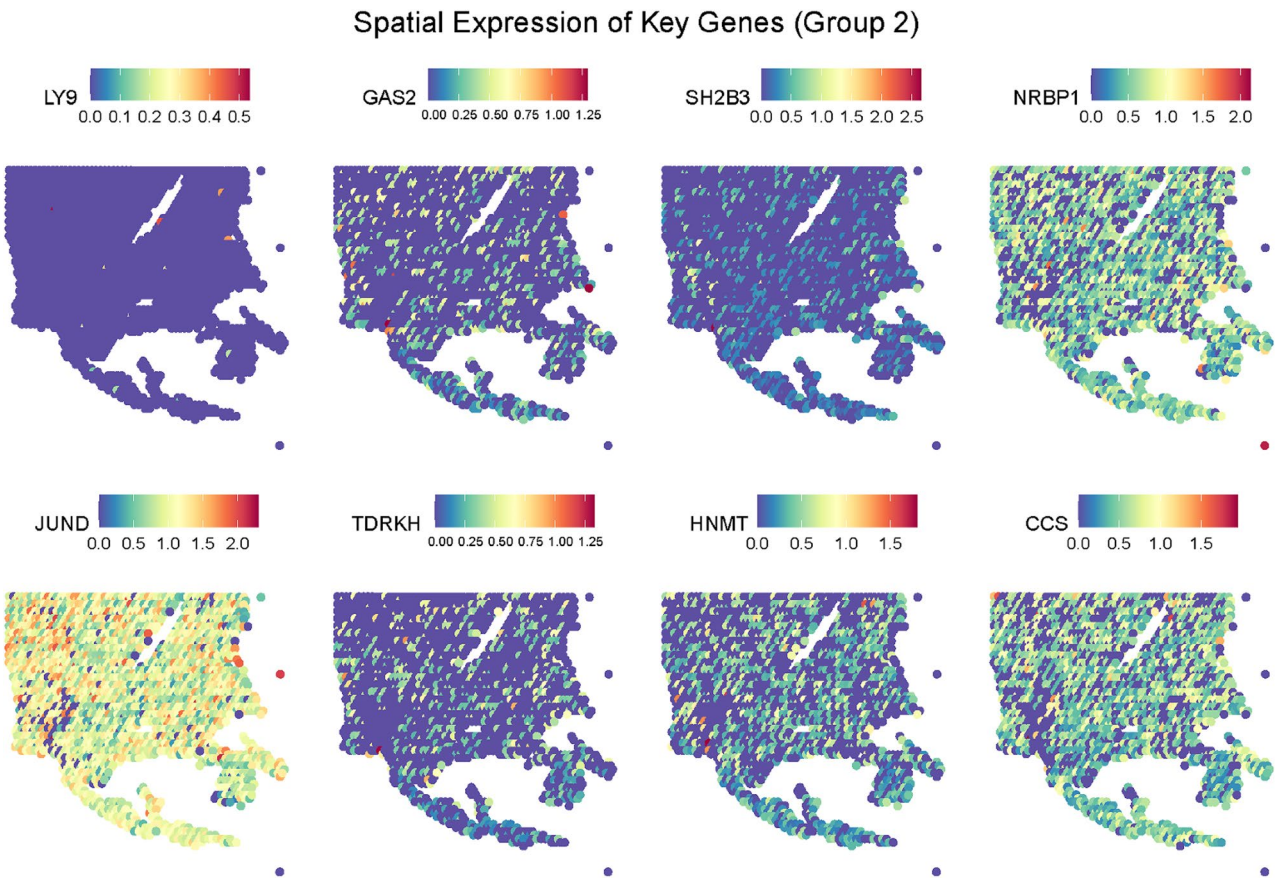

Fig. 5 Spatial transcriptomics reveals heterogeneous and localized expression of candidate genes in HCC tissue. Spatial feature plots of LY9, GAS2, SH2B3, NRBP1, JUND, TDRKH, HNMT and CCS demonstrating localized expression in tumor subregions

Known oncogenic factors such as *JUND* were reaffirmed, while novel causal genes such as *PSMB1* and context-dependent regulators like *ST6GAL1* were uncovered. The convergence of multi-layered evidence strengthens their potential as candidate key target for diagnosis and as molecular targets for therapeutic development.

Furthermore, the SMR framework of this study demonstrates a remarkable multi-omics integration advantage in terms of methodology. Unlike previous studies that relied solely on single-level GWAS or co-localization analysis, this study achieved cross-level causal inference by integrating mQTL, eQTL, and pQTL data. It also combined multi-dimensional supportive analysis using peripheral blood transcriptomes, ELISA, and spatial transcriptomes, significantly enhancing the biological credibility and clinical translational potential of the causal inference [30, 31]. Concurrently, single-cell transcriptomics further revealed the specific expression and signaling communication patterns of candidate genes (such as ST6GAL1, JUND) in immune cell subsets, suggesting their crucial roles in the remodeling of the immune microenvironment and the progression of HCC.

Nevertheless, several limitations remain. We sincerely acknowledge that the HEIDI heterogeneity test, an essential component of the SMR framework, was not implemented in the present analysis. This omission represents a methodological limitation of our study. In order to more comprehensively explore potentially relevant genes and facilitate downstream biological screening and validation, we adopted a nominal threshold of $p\_SMR < 0.05$ without HEIDI filtering. We fully recognize that this exploratory approach may include non-causal associations, and we plan to incorporate standard HEIDI testing and more stringent filtering criteria in future analyses to further strengthen causal inference. The causal inference of SMR requires functional supportive analyses(e.g., knockdown or overexpression) in cellular or animal models. The relatively small supportive analyses cohort ($n = 10$ per group) may limit generalizability, and future studies should therefore expand the sample size (for instance, to at least 50 participants per group) and include independent cohort validation to enhance the robustness and reproducibility of the results. Expanding the multi-omics framework to include metabolomics and single-cell epigenomics would offer a more holistic view of molecular

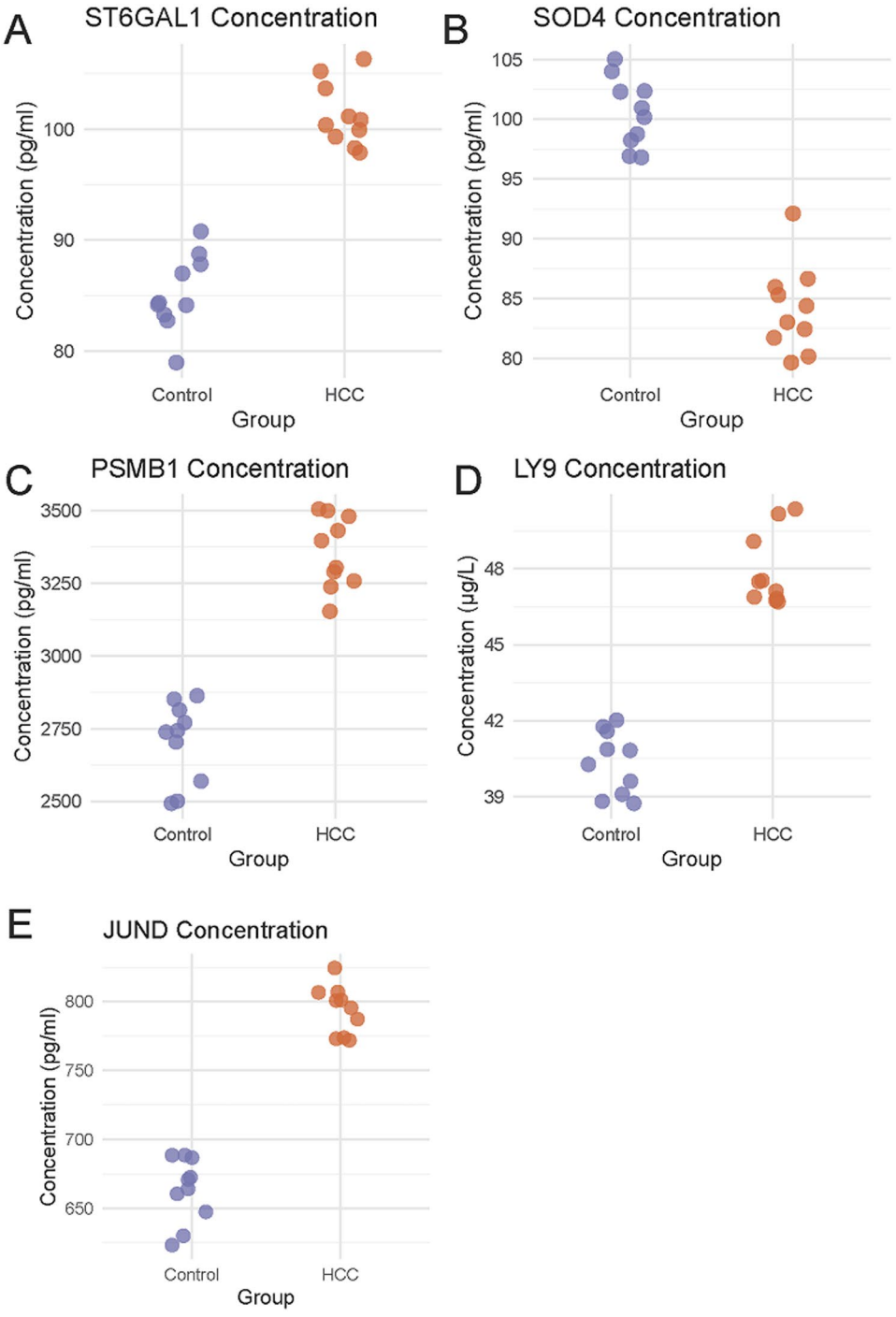

Fig. 6 ELISA analysis supports confirms altered circulating protein abundance in HCC. **A–E** Protein concentrations of ST6GAL1, SOD4, PSMB1, LY9, and JUND between HCC and controls

regulation. Ultimately, clinical trials are warranted to determine the diagnostic and prognostic value of these key target in diverse HCC populations. Another limitation of this study is the composition of the control group. The peripheral blood samples used for transcriptomic and proteomic supportive analyses were obtained from healthy individuals rather than high-risk populations such as patients with cirrhosis or chronic liver disease. Given that hepatocellular carcinoma (HCC) often arises in the context of chronic liver injury, it is possible that some of the molecular signals identified here reflect underlying hepatic dysfunction rather than cancer-specific alterations. The absence of high-risk controls may therefore lead to an overestimation of the differences

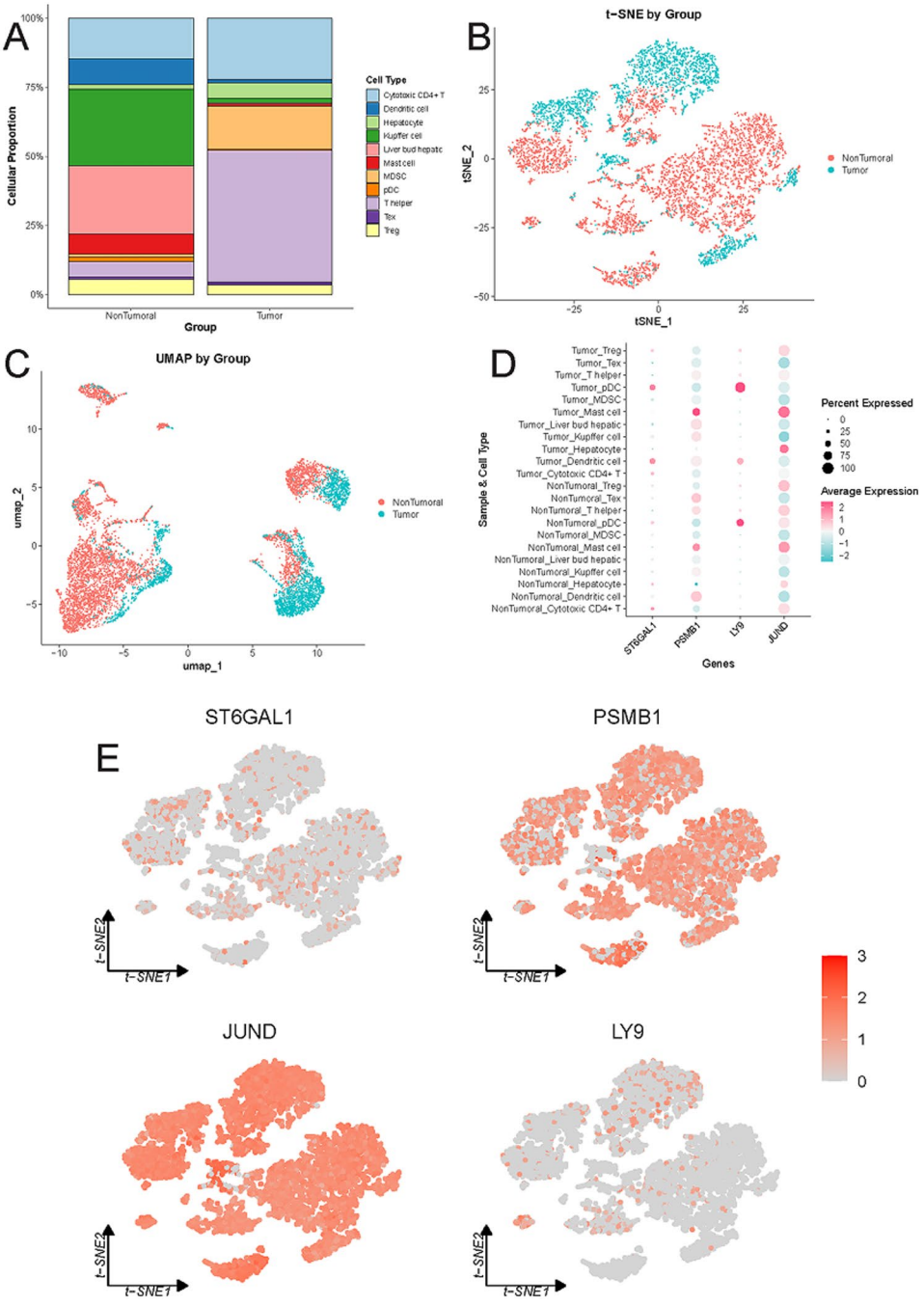

Fig. 7 Single-cell RNA-seq identifies cell-type-specific expression of SMR candidate genes. **A–C** UMAP/t-SNE plots showing cell-type clusters in tumor and non-tumor samples. **D–E** Dot and feature plots illustrating expression of ST6GAL1, PSMB1, LY9, and JUND across annotated immune and stromal cell populations

between HCC and normal groups. Future studies should include cirrhotic or chronic liver disease cohorts as intermediate controls to better distinguish HCC-specific key target from those associated with hepatic inflammation or fibrosis.

Although the spatial transcriptomics analysis was performed on a single representative HCC tissue section, it served as a spatial supportive analyses of key genes identified through multi-omics SMR and blood-based assays. The convergent evidence from multi-platform analyses supports the biological relevance of these findings, though future studies with larger spatial sampling will be needed to confirm tumor-wide representativeness.

Conclusion

This study employed a comprehensive multi-omics SMR framework coupled with multi-dimensional experimental supportive analyses to precisely identify potential driver genes and proteins causally linked to hepatocellular carcinoma (HCC). By integrating GWAS, eQTL, mQTL, and pQTL data, we robustly inferred causal associations, highlighting a core set of 16 candidate genes, including JUND, TDRKH, ST6GAL1, PSMB1, and LY9, whose roles were further supported by differential expression and protein abundance in peripheral blood, distinct spatial localization within tumor tissues, and cell-type specific patterns in the tumor microenvironment. These findings not only advance our understanding of HCC pathogenesis beyond mere associations, providing novel molecular insights into known and emerging drivers, but also establish a strong foundation for developing precise diagnostic, prognostic, and therapeutic strategies, paving the way for future translational research in HCC.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12885-026-15667-2.

Supplementary Material 1.

Authors' contributions

Y.L. and M.Z. conceived the study, acquired funding, and supervised the overall project. Z.Y., T.L., and J.S. contributed equally to this work, including methodology development, data analysis, and drafting of the initial manuscript. T.L., J.S., and Z.M. performed the multi-omics SMR analysis and single-cell RNA sequencing data analysis. Z.Y., J.Y., and Y.L. carried out the peripheral blood transcriptome sequencing, ELISA, and spatial transcriptomics experiments. Y.H. and C.L. assisted with patient sample collection and clinical data provision. All authors reviewed and edited the manuscript.

Funding

This study was supported by the China Academy of Chinese Medical Sciences Science and Technology Innovation Project, "Development of a Traditional Chinese Medicine Self-Medication System Based on Large Language Models" (Project No.: ZN2023A02).

Data availability

The raw RNA sequencing data and processed expression matrices from peripheral blood transcriptome analysis have been deposited in the Gene Expression Omnibus (GEO) database under the accession number GSE317443. All other relevant data supporting the findings of this study, including detailed statistical analyses and additional quality control metrics, are available within the article and its Supplementary Information files, or from the corresponding authors upon reasonable request.

Declarations

Ethics approval and consent to participate

All experiments involving human tissues were conducted in accordancewith the principles of the Declaration of Helsinki. This study was approved byt the Medical Ethics Committee of Wangjing Hospital, China Academy of Chinese Medical Sciences (Approval No.: WJEC-YJS-2025-001-P001) and the Ethics Committee Office of Mengchao Hepatobiliary Hospital of Fujian Medical University (Approval No.: 2025_010_01).Written informed consent was obtained from alparticipants.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Author details

[1]Wangjing Hospital, China Academy of Chinese Medical Sciences, Beijing 100102, China
[2]Institute of Chinese Materia Medica, China Academy of Chinese Medical Sciences, Beijing 100700, China
[3]Department of Hepatobiliary Surgery, Mengchao Hepatobiliary Hospital of Fujian Medical University, Fuzhou, China
[4]Guang'anmen Hospital, China Academy of Chinese Medical Sciences, Beijing 100053, China
[5]Experimental Research Center, China Academy of Chinese Medical Sciences, 100700 Beijing, China

Received: 23 October 2025 / Accepted: 28 January 2026
Published online: 16 February 2026

References

1. Sharman JE, O'Brien E, Alpert B, Schutte AE, Delles C, Hecht Olsen M, et al. Lancet commission on hypertension group position statement on the global improvement of accuracy standards for devices that measure blood pressure. J Hypertens. 2020;38(1):21–9.
2. Zheng J, Wang S, Xia L, Sun Z, Chan KM, Bernards R, et al. Hepatocellular carcinoma: signaling pathways and therapeutic advances. Signal Transduct Target Therapy. 2025;10(1):35.
3. Ghouse J, Gellert-Kristensen H, O'Rourke CJ, Seidelin AS, Thorleifsson G, Sveinbjornsson G, et al. Genome-wide meta-analysis identifies nine loci associated with higher risk of hepatocellular carcinoma development. JHEP Reports: Innov Hepatol. 2025;7(9):101485.
4. Yu HP, Xiang BD, Qian J, Liu H, Li YF, Lin Q, et al. Genome-wide association study of HBV-related hepatocellular carcinoma identifies a functional variant at the FAM114A1 locus. Journal of genetics and genomics=Yi chuan xue bao. 2025.
5. Tavaglione F, Pennisi G, Pelusi S. PNPLA3 I148M and hepatocellular carcinoma. Liver International: Official J Int Association Study Liver. 2025;45(4):e70051.
6. Du M, Wu T, Yuan H, Zhang T, Liu Z, Chen X. Integrated genetic information of metabolic Dysfunction-Associated steatotic liver Disease-Related traits improves hepatocellular carcinoma risk stratification and screening. JCO Precision Oncol. 2025;9:e2500638.
7. Gao Q, Zhu H, Dong L, Shi W, Chen R, Song Z, et al. Integrated proteogenomic characterization of HBV-Related hepatocellular carcinoma. Cell. 2019;179(2):561–e577522.
8. Tümen D, Heumann P, Gülow K, Demirci CN, Cosma LS, Müller M, et al. Pathogenesis and current treatment strategies of hepatocellular carcinoma. Biomedicines 2022, 10(12).
9. Hassan MM, Li D, Han Y, Byun J, Hatia RI, Long E, et al. Genome-wide association study identifies high-impact susceptibility loci for HCC in North America. Hepatology (Baltimore MD). 2024;80(1):87–101.
10. Liu Q, Liu G, Lin Z, Lin Z, Tian N, Lin X, et al. The association of LncRNA SNPs and SNPs-environment interactions based on GWAS with HBV-related HCC risk and progression. Mol Genet Genom Med. 2021;9(2):e1585.
11. Zhu W, Fan C, Liu B, Qin J, Fan A, Yang Z, et al. Therapeutic targets for hepatocellular carcinoma identified using proteomics and Mendelian randomization. J Gastroenterol Hepatol. 2025;40(1):282–93.
12. Feng L, Yuan Q, Yu H, Ye R, Xie Z, Xu J, et al. Identification of biomarkers associated with exhausted CD8+T cells in the tumor microenvironment of intrahepatic cholangiocarcinoma based on Mendelian randomization and bioinformatics analysis. Discover Oncol. 2025;16(1):1092.
13. Ruiz EJ, Lan L, Diefenbacher ME, Riising EM, Da Costa C, Chakraborty A, et al. JunD, not c-Jun, is the AP-1 transcription factor required for Ras-induced lung cancer. JCI Insight. 2021; 6(13).
14. Wang X, Li P, Ji H, Xu Z, Xing H. Single-cell transcriptomics reveals overactivated reactive oxygen species pathway in hepatocytes in the development of hepatocellular carcinoma. Sci Rep. 2024;14(1):29809.

15. Zou X, Lu J, Deng Y, Liu Q, Yan X, Cui Y, et al. ST6GAL1 inhibits metastasis of hepatocellular carcinoma via modulating sialylation of MCAM on cell surface. Oncogene. 2023;42(7):516–29.
16. Liu Q, Ma H, Sun X, Liu B, Xiao Y, Pan S, et al. The regulatory ZFAS1/miR-150/ST6GAL1 crosstalk modulates sialylation of EGFR via PI3K/Akt pathway in T-cell acute lymphoblastic leukemia. J Experimental Clin Cancer Research: CR. 2019;38(1):199.
17. Wei A, Fan B, Zhao Y, Zhang H, Wang L, Yu X, et al. ST6Gal-I overexpression facilitates prostate cancer progression via the PI3K/Akt/GSK-3β/β-catenin signaling pathway. Oncotarget. 2016;7(40):65374–88.
18. Xu X, Liu J, Qin W, Ding C, Cai Z, Shu D, et al. ST6GAL1-Mediated sialylation stabilizes PD-L1 and drives immunosuppressive tumor microenvironment in colorectal Cancer. *Advanced science (Weinheim*. Baden-Wurttemberg Germany). 2025;12(42):e06225.
19. Zhang X, Lao M, Sun K, Yang H, He L, Liu X, et al. Sphingolipid synthesis in tumor-associated macrophages confers immunotherapy resistance in hepatocellular carcinoma. Sci Adv. 2025;11(21):eadv0558.
20. Xie D, Liu Y, Xu F, Dang Z, Li M, Zhang Q, et al. Immune microenvironment and immunotherapy in hepatocellular carcinoma: mechanisms and advances. Front Immunol. 2025;16:1581098.
21. Gao B, Lu Y, Lai X, Xu X, Gou S, Yang Z, et al. Metabolic reprogramming in hepatocellular carcinoma: mechanisms of immune evasion and therapeutic implications. Front Immunol. 2025;16:1592837.
22. Weng X, Huang Y, Fu Z, Liu X, Xie F, Wang J, et al. METTL1-driven nucleotide metabolism reprograms the immune microenvironment in hepatocellular carcinoma: a multi-omics approach for prognostic biomarker discovery. Front Immunol. 2025;16:1582203.
23. Wang Q, Liu J, Yang M, Zhou J, Li Y, Zheng J, et al. Targeting AKR1B1 inhibits metabolic reprogramming to reverse systemic therapy resistance in hepatocellular carcinoma. Signal Transduct Target Therapy. 2025;10(1):244.
24. Luo Y, Jin X, Huang L, Zeng D, Zhang N, Tang S, et al. RUNX1/SLAMF3 Axis Drives Immunosuppression to Contribute to Colorectal Cancer Liver Metastasis by Blocking Phagocytosis and Depleting C1QC(+) Tumor-Associated Macrophages. Advanced science (Weinheim, Baden-Wurttemberg, Germany). 2025;12(32):e06641.
25. Ogishi M, Puchan J, Yang R, Arias AA, Han JE, Nguyen T, et al. Human LY9 governs CD4(+) T cell IFN-gamma immunity to Mycobacterium tuberculosis. Sci Immunol. 2025;10(107):eads7377.
26. Kim JE, Kim HS, Kim W, Lee EH, Kim S, Kim T, et al. Isoxazole-based molecules restore NK cell immune surveillance in hepatocarcinogenesis by targeting TM4SF5 and SLAMF7 linkage. Signal Transduct Target Therapy. 2025;10(1):15.
27. Ercan H, Reumiller CM, Muhlberger J, Hsu F, Schmidt GJ, Umlauf E, et al. Platelets mirror changes in the frontal lobe antioxidant system in alzheimer's disease. Alzheimer's Dement J Alzheimer's Assoc. 2025;21(4):e70117.
28. Araya-Osorio R, Dominguez M, Thallmair S, Marrink SJ, Souza PCT, Mera-Adasme R. Allosteric pathway connects Zn(II) loss from SOD1 to known pathogenic mechanisms. Int J Biol Macromol. 2025;331(Pt 1):148349.
29. Liu D, Zhao X, Wang Z, Wang G, Chen Z, Ning S, et al. A novel protein encoded by circARHGAP12 attenuates DNA damage and apoptosis by regulating MDC1 in intestinal ischemia/reperfusion injury. Int J Biol Macromol. 2025;286:138374.
30. Lee SK, Lim J, Jhun JY, Moon J, Kim HS, Choi JY, et al. Landscape of T-cell exhaustion heterogeneity and HBV integration in virus-related HCC revealed by whole-exome, transcriptome, and single-cell sequencing. JHEP Reports: Innov Hepatol. 2025;7(11):101518.
31. Wei S, Quan Y, Li X, Zhong S, Xiao L, Yang C, et al. Unveiling migraine subtype heterogeneity and risk loci: integrated genome-wide association study and single-cell transcriptomics discovery. J Headache Pain. 2025;26(1):185.

Publisher's note