Targeting VCX2 in Hepatocellular Carcinoma: An Omics-Guided Druggability Study Using Peruvian Natural Products

Luis Daniel Goyzueta-Mamani
Universidad Católica de Santa María
Haruna Luz Barazorda-Ccahuana
Universidad Católica de Santa María
Mayron Antonio Candia-Puma
Universidad Católica de Santa María
Nadia M. Hamdy
Ain Shams University
Miguel Angel Chávez-Fumagalli
mchavezf@ucsm.edu.pe

Universidad Católica de Santa María

Article

Keywords: Hepatocellular carcinoma (HCC), single-cell RNA sequencing, VCX2, cancer biomarker, Luteolin-5-O-glucoside, Single-cell transcriptomics, Computational Analysis, Natural products, Molecular Docking

Posted Date: October 9th, 2025

DOI: https://doi.org/10.21203/rs.3.rs-7601713/v1

Additional Declarations: No competing interests reported.

Abstract

Hepatocellular carcinoma (HCC) is among the deadliest cancers, and current biomarkers offer limited diagnostic and therapeutic utility. To uncover novel druggable targets, we employed an omics-guided computational pipeline that combines single-cell RNA sequencing (scRNA-seq), differential gene expression (DGE) analysis, UMAP clustering, and protein–protein interaction (PPI) network mapping. Among the prioritized genes—TMBIM4, RGS5, CEACAM7, and VCX2—the cancer/testis antigen VCX2 stood out for its aberrant expression and potential role in chromosomal instability. The VCX2 structure was modeled via AlphaFold2 and refined to reveal druggable pockets for virtual screening using the PeruNPDB (Peruvian Natural Products Database). Docking identified luteolin-5-O-glucoside from *Equisetum arvense* as the top ligand (Docking score: − 7.42 kcal/mol; ΔG_{bind}: − 40.13 ± 1.12 kcal/mol). MM-GBSA and interaction analysis showed stable hydrogen bonds with PRO91, GLU97, and GLU109. These findings highlight VCX2 as a promising target for HCC and suggest luteolin-5-O-glucoside as a potential lead scaffold. Further molecular dynamics and SAR studies are needed to validate and optimize its efficacy.

INTRODUCTION

HCC represents approximately 75–85% of primary liver cancer, causes more than 830,000 fatalities annually[1,2], and ranks as the third leading cause of cancer-related mortality worldwide[3]. Its burden reflects both global incidence and regional disparities: East Asia and Sub-Saharan Africa, where hepatitis B virus (HBV) infection is endemic [4,5], show the highest rates[6], while Western countries report increasing incidence due to non-alcoholic fatty liver disease (NAFLD) and obesity[7]. Chronic hepatic inflammation, fibrosis, and cirrhosis underlie most cases, driven by viral hepatitis, alcohol, aflatoxin exposure, and metabolic disorders[8]. Clinically, early-stage HCC is often asymptomatic, and current surveillance (imaging plus Alpha-Fetoprotein, AFP) lacks sensitivity and specificity, underscoring the need for more robust biomarkers and targeted strategies[9].

At a molecular level, HCC pathogenesis involves a complex interplay of genetic and epigenetic alterations that promote tumor progression, immune evasion, and therapeutic resistance.

Recognized biomarkers, such as AFP [10], glypican-3 (GPC3) [11], and Epithelial cell adhesion molecule (EpCAM), have clinical utility but show limited performance—particularly in early disease—motivating the discovery of more selective targets. Recent single-cell and bulk transcriptomic studies, coupled with network-based prioritization, have highlighted underexplored drivers (e.g., SPP1, AKR1B10, LMNB1), illustrating how integrative omics can reveal mechanistically informative biomarkers.

Natural products continue to be a rich source of anti-cancer scaffolds with multi-target activity relevant to HCC biology. Compounds such as curcumin, resveratrol, quercetin, prodigiosin, and berberine modulate apoptosis, angiogenesis, and immune signaling; for example, curcumin attenuates TGF-β–driven EMT/fibrosis[10], resveratrol suppresses VEGF-mediated angiogenesis[11], and berberine

downregulates GPC3 associated with Wnt/β-catenin signaling[12]. Hepatoprotective phytochemicals from Silybum marianum (flavonolignans such as silymarin) [13], *Glycyrrhiza glabra* (glycyrrhizin), and *Phyllanthus* species have further motivated biomarker-guided natural-product discovery[14].

Here, we integrate scRNA-seq with pan-cancer transcriptomic profiling to identify Variable Charge X-linked 2 (VCX2) as an underreported, cancer/testis antigen (CTA)– like candidate in HCC. CTAs are typically restricted to immune-privileged germline tissues, and their aberrant expression in tumors provides both diagnostic specificity and therapeutic tractability. We then evaluate VCX2 druggability through structural modeling and molecular dynamics (MD), predict a binding pocket with SiteMap, and perform structure-based screening of the Peruvian Natural Products Database (PeruNPDB) to prioritize ligand candidates. Luteolin-5-O-glucoside emerges as the top binder and is subsequently examined by Glide docking and MM-GBSA. The overall workflow—from scRNA-seq–based prioritization to *in silico* ligand evaluation—is summarized in the Graphical Abstract.

MATERIAL AND METHODS

These calculations provided refined binding free energies (ΔG_{bind}) and highlighted hotspot residues within the VCX2 binding pocket, supporting robust prediction of ligand–protein stability. Full computational details and energy decomposition equations are provided in the Supplementary Materials.

Single-cell RNA sequencing data processing and integration

The sequential steps of this pipeline are outlined in the Graphical Abstract and described in detail below. scRNA-seq datasets were collected to characterize transcriptional differences between healthy and cancerous liver tissues. Data from five relatively healthy individuals were obtained from the Gene Expression Omnibus (GEO) (GSE115469)[15] (https://www.ncbi.nlm.nih.gov/geo/, accessed 14 February 2024)[16], while hepatocellular carcinoma (HCC) samples were retrieved from the CancerSCEM database (https://ngdc.cncb.ac.cn/cancerscem/index, accessed 20 February 2024)[17,18]. Additional tumor datasets (bladder[19], breast[20], colorectal[21], lung[22,23], ovarian [24], pancreatic[25], and gastric[26] cancers) were included to enable broader comparative assessments. Raw count matrices were imported into R (v4.1.0)[27] and processed with Seurat (v4.0.3)[28]. Cells with fewer than 100 detected genes, more than 25% mitochondrial transcripts, or genes expressed in fewer than three cells were excluded. Normalization and variance stabilization were carried out with Seurat's built-in functions, while batch effects were regressed using the number of detected UMIs and mitochondrial expression levels. Principal component analysis (PCA) was performed on highly variable genes, and graph-based clustering was applied to group cells with similar expression profiles. UMAP embeddings were generated to visualize the resulting clusters and cellular heterogeneity. Differentially expressed genes (DEGs) were identified using the Wilcoxon rank-sum test with Bonferroni-adjusted $p<0.05$. The resulting clusters were annotated using the SingleR package, which compares test profiles to reference transcriptomes to

assign likely cell identities[29]. Functional enrichment analyses for Biological Processes, Molecular Functions, and Cellular Components, as well as Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways, were performed with clusterProfiler[30] and enrichplot[31]. Visualizations, including UMAP projections, volcano plots, dot plots, and heatmaps, were generated with SCpubr[32].

Protein-Protein Interaction (PPI) Network and Tissue-Specific Expression Analysis

To better understand the biological context of candidate biomarkers, PPI networks were constructed using the STRING database (v12.0) (https://string-db.org, accessed 20 April 2024)[33], retaining edges with confidence scores ≥0.70. Network visualization was performed in Cytoscape (v3.9.1)[34], and centrality metrics were computed with the cytoHubba plugin[35] using the Maximal Clique Centrality (MCC) algorithm. Hub nodes were prioritized for their potential relevance to hepatocarcinogenesis. Functional enrichment of the network was performed using the StringApp plugin [36], which retrieved Gene Ontology (GO) terms and KEGG annotations directly from STRING. To assess expression specificity, candidate genes were also cross-validated in the TISSUES database (v2.0) (https://tissues.jensenlab.org/Search, accessed on 25 April 2024), which integrates transcriptomic and proteomic evidence of tissue-specificity[37].

Structural Refinement of VCX2: Molecular Dynamics and Protein Preparation

The amino acid sequence of VCX2 (UniProt ID: Q9H322) was retrieved from UniProt, and the initial three-dimensional model was obtained from the AlphaFold Protein Structure Database (v4). To improve structural fidelity before docking, the model was refined with GROMACS 2022.4 using the CHARMM36m force field. The refinement protocol consisted of energy minimization to resolve steric clashes, simulated annealing cycles (heating to 500 K and cooling to 300 K), equilibration under NVT conditions, and a 1 μs production simulation under NPT conditions at 309 K and 1 bar. Three independent replicates with different initial velocities were conducted to capture conformational variability and ensure reproducibility.

Trajectory analysis confirmed model stability, with root mean square deviation (RMSD) convergence and consistent PCA profiles across replicates. Model validation with ProSa-web (https://prosa.services.came.sbg.ac.at/prosa.php) and PDBsum (https://www.ebi.ac.uk/thornton-srv/databases/pdbsum/) confirmed stereochemical quality, with residues predominantly located in favored Ramachandran regions.

The final refined structure was prepared for docking in Schrödinger Maestro (v12.8.117) [38] using the Protein Preparation Wizard. Bond orders were corrected, hydrogen atoms added, side chains optimized, protonation states predicted at pH 7.0 with Epik, and hydrogen bond networks refined. A final energy minimization was performed with the OPLS4 force field [39], yielding a high-quality receptor model suitable for downstream binding site mapping and docking simulations.

Virtual screening with the Peruvian Natural Products Database (PeruNPDB)

A structure-based virtual screening was performed with phytochemicals from the PeruNPDB (https://perunpdb.com.pe, accessed on April 15, 2024)[40], a curated collection of compounds from native Peruvian biodiversity. In total, 280 molecules were retrieved in SMILES format[41] from the official web server (https://perunpdb.com.pe, accessed April 15, 2024). SMILES strings were converted into three-dimensional structures and energy minimized using OpenBabel (v 3.1.1) within the Python Prescription Virtual Screening Tool[42]. The ligands were imported into PyRx v0.9.8, which uses AutoDock Vina(v 1.2.5)[43] for docking, with the MD-refined VCX2 structure as the receptor. Grid parameters were defined with the “Run AutoGrid” option, and simulations were executed with the Lamarckian Genetic Algorithm, setting exhaustiveness to 20. For each ligand, only the lowest-energy pose was retained. Binding affinities were standardized by Z-score normalization, and ligands within ±1.645 ($p < 0.05$) were considered significant. Results were visualized as violin plots in GraphPad Prism (v10.0.2; GraphPad Software, San Diego, CA, USA; http://www.graphpad.com)

Binding Site Prediction, Grid Generation, and Ligand Preparation

Following structural refinement of the VCX2 model, potential ligand-binding pockets were identified using the SiteMap module in Schrödinger Maestro (v12.8.117). The algorithm predicted the five top-ranked binding sites on the protein surface based on site score, volume, enclosure, exposure, and the balance of hydrophobicity and hydrophilicity. The top-ranked pocket, with the highest composite site score, was selected for docking. Site points were cropped at a 4.0 Å radius from the binding site centroid.

A receptor grid was generated using the Receptor Grid Generation wizard in Glide (Schrödinger Suite), centered on the centroid of the selected SiteMap pocket. The grid box was adjusted to enclose the predicted binding cavity while minimizing excess volume fully. No positional or pharmacophore constraints were applied during grid preparation.

The selected ligand from virtual screening was prepared using the LigPrep wizard (Schrödinger Maestro v12.8.117). Protonation states and tautomeric forms were generated using Epik at pH 7.0 ± 2.0, with a maximum of 32 stereoisomers per compound generated to account for conformational diversity. Ligand structures were energy-minimized using the OPLS4 force field before docking.

Molecular Docking (MD) and MM-GBSA Binding Free Energy Calculations

Structure-based molecular docking was performed with the Glide module (Schrödinger Suite v12.8.117) using the receptor grid defined by the top-ranked SiteMap pocket. Ligands were docked in a two-stage precision protocol: Standard Precision (SP) to identify plausible binding poses, followed by Extra Precision (XP) for high-resolution scoring and pose refinement. Enhanced sampling was enabled, and a van der Waals scaling factor of 0.80 was applied to nonpolar atoms to improve discrimination of sterically unfavorable interactions.

Docking results were ranked by GlideScore, and key protein–ligand interactions (hydrogen bonds, salt bridges, π–π stacking) were analyzed using Maestro's Interaction Diagram. To complement docking, MM-GBSA binding free energy calculations were performed with the Prime module (v4.8), applying the VSGB 2.0 solvation model and OPLS4 force field. Snapshots were extracted from the equilibrated portion of the 100 ns MD trajectories (90–100 ns), ensuring thermodynamic convergence.

RESULTS AND DISCUSSIONS

We identified VCX2 as a novel biomarker in HCC, representing the first report of this cancer/testis antigen in liver malignancy. scRNA-seq analysis revealed its differential expression between healthy and diseased liver tissues, supported by pan-cancer validation that confirmed its relative specificity to HCC. Structural modeling of VCX2 uncovered a druggable binding pocket, which enabled virtual screening of the PeruNPDB natural products library. From this screen, Luteolin-5-O-glucoside emerged as the top inhibitor candidate, showing a docking score of –7.42 kcal/mol and favorable MM-GBSA binding free energy. These findings link the discovery of VCX2 with therapeutic exploration, establishing a framework that integrates omics-guided biomarker identification and natural product–based inhibitor prediction (Figures 1 and 5).

Our analysis focused on genes associated with fibrosis, angiogenesis, immune modulation, and apoptosis regulation—processes central to HCC initiation and progression. The scRNA-seq dataset enabled high-resolution profiling of hepatic cell populations, including hepatocytes, endothelial cells, Kupffer cells, and stellate cells, providing a foundation for biomarker prioritization. UMAP projections (Figure 1A) revealed distinct transcriptional clusters reflecting these cellular identities, while overlay analysis of healthy (green) and diseased (red) samples (Figure 1B) showed marked compositional shifts. HCC tissues exhibited enrichment of fibroblasts and immune cell populations, consistent with stromal activation, immune infiltration, and metabolic reprogramming that drive hepatic tumorigenesis.

HCC progression involves chronic inflammation, immune dysregulation, fibrogenesis, and extracellular matrix (ECM) remodeling, which together establish a tumor-permissive microenvironment[44,45]. To capture these transitions, we compared the cellular and transcriptional architecture of healthy and diseased liver tissues (Figure 2).

Figure 1C shows a heatmap summarizing the top three marker genes per cluster. The upper panel highlights co-expression patterns, revealing a strong correlation between TMBIM4 and RGS5, genes implicated in apoptosis inhibition and vascular remodeling[46]. The middle panel reports statistical significance (adjusted $p < 0.05$), with markers such as TMBIM4, RGS5, and CEACAM7 displaying robust shifts. CEACAM7, normally tumor-suppressive, exhibited aberrant expression in HCC[47], consistent with its role in malignant transformation[48]. The lower panel illustrates $\log_2$ fold-changes, where TMBIM4, RGS5, and Cytochrome B5 Reductase 3 (CYB5R3) were upregulated, indicating their roles in apoptosis resistance, oxidative stress adaptation, and metabolic reprogramming[49,50]. Downregulated hepatocyte-

specific genes reflected impaired metabolic capacity, while VCX2 showed unexpectedly low expression, underscoring the need for broader contextual analyses[51].

The volcano plot in Figure 1D provides a global overview of differential expression, with TMBIM4 and RGS5 among the most significantly upregulated markers, supporting their roles in survival, immune evasion, and vascular remodeling. Conversely, downregulated genes clustered on the left side of the distribution were enriched in hepatocyte metabolic pathways that are suppressed during tumorigenesis [52].

Finally, Figure 1E illustrates gene expression patterns using a dot plot. Dot size reflects the proportion of expressing cells, while color intensity indicates mean expression. TMBIM4, RGS5, and VCX2 stood out for their elevated expression in fibroblasts, macrophages, T cells, and B cells within diseased livers. Notably, TMBIM4 showed marked enrichment in tumor-associated fibroblasts and immune cells, suggesting a role in apoptosis resistance within the fibrotic and inflammatory microenvironment characteristic of HCC[53].

In Figure 2A, bar plots show the relative abundance of major hepatic cell types. Healthy livers were primarily composed of hepatocytes (green), with smaller contributions from macrophages (blue), endothelial cells (yellow), fibroblasts (brown), T cells (dark pink), and B cells (light pink). Diseased tissues displayed a sharp decline in hepatocytes alongside expansion of fibroblasts, macrophages, and immune cells, reflecting fibrotic remodeling and chronic inflammation typical of HCC[54,55].

Figure 2B shows signaling pathways enriched in diseased liver tissue. The PI3K–Akt pathway was most prominent, driving survival, proliferation, and angiogenesis, while cytokine–cytokine receptor interactions reflected immune activation by macrophages and T cells[56–59]. ECM–receptor and focal adhesion pathways indicated fibroblast-driven remodeling that supports invasion, and enrichment of ribosome-related and immune pathways pointed to increased protein synthesis and systemic immune dysregulation[58,59]. Together, these results outline a network of proliferative, fibrotic, inflammatory, and angiogenic programs in HCC[60,61].

Figure 2C highlights increased chemotaxis, consistent with immune recruitment, and enhanced ECM organization from fibroblast activity, along with altered cell adhesion that facilitates tumor spread[62]. Elevated translation and ribosome activity reflected the biosynthetic needs of malignant and stromal cells[63].

Structural changes included collagen-rich ECM remodeling, focal adhesion formation, and basement membrane alterations that enable angiogenesis and tumor migration (Figure 2D). Increased ribosomes and endoplasmic reticulum further emphasized the high protein synthesis required for tumor progression[64,65].

Figure 2E highlights molecular functions enriched in diseased liver tissue, with a strong emphasis on ECM remodeling. Upregulation of structural components, such as collagen, reflects the activity of fibroblasts and stellate cells [68]. In contrast, increased integrin binding and adhesion molecule activity support the greater motility and invasive potential of tumor and stromal cells [66]. Enrichment of heparin and glycosaminoglycan binding indicates modulation of growth factor availability and cytokine signaling, sustaining angiogenesis and inflammation[67]. Finally, elevated platelet-derived growth factor (PDGF) binding underscores persistent fibroblast activation and their central role in ECM production and remodeling of the tumor microenvironment[68].

These findings illustrate the profound cellular and molecular reprogramming that drives HCC progression. The loss of hepatocytes, alongside the expansion of fibroblasts, macrophages, and immune cells, reflects both structural disintegration and immune dysregulation in the liver microenvironment. Enrichment of pathways such as PI3K–Akt, ECM–receptor interactions, and cytokine signaling underscores the interplay of proliferation, fibrosis, inflammation, and angiogenesis, while processes like cell adhesion, ECM biosynthesis, and chemotaxis highlight stromal and immune contributions to a pro-tumorigenic niche. At the molecular level, PDGF signaling, integrin-mediated communication, and matrix remodeling emerge as key mechanisms of invasion and metastasis, offering translational opportunities for therapies targeting fibrotic, immune, and angiogenic axes of HCC.

To broaden relevance, a pan-cancer analysis was performed across bladder, breast, colon, lung, ovarian, pancreatic, and gastric tumors. This comparison distinguished liver-specific from shared oncogenic biomarkers, providing insight into their potential as diagnostic markers or therapeutic targets beyond HCC. Through integrated analyses of clustering, differential expression, and cross-tumor validation, the study builds a functional landscape that situates liver biomarkers within wider oncogenic networks.

Figures 3A and 3B display the clustering and expression profiles of key genes across eight cancer datasets, highlighting both the distinctiveness of liver cancer and the pathways it shares with other malignancies. In Figure 3A, HCC forms a separate cluster enriched in genes related to apoptosis resistance, fibrosis, and vascular remodeling, hallmarks of its progression[69]. These liver-specific patterns reflect the impact of chronic injury, metabolic reprogramming, and persistent inflammation, distinguishing HCC from pancreatic, gastric, and lung cancers. While pathways such as ECM remodeling and immune modulation appear across multiple tumors, their expression in non-hepatic cancers lacks the tissue-specific context that characterizes HCC[70].

Figure 3B shows that liver-specific clusters exhibit markedly higher expression of genes related to **metabolism and fibrosis** in HCC compared to other cancers, where these genes are minimally expressed, supporting their value as diagnostic biomarkers. In contrast, genes regulating **angiogenesis** and **ECM degradation** were also expressed in gastric and pancreatic cancers, underscoring the shared tumorigenic pathways [71]. This highlights the dual challenge of identifying truly liver-specific markers while recognizing the convergence of oncogenic mechanisms across malignancies.

To expand this analysis, Figure 3C presents a heatmap of gene expression across bladder, breast, colon, lung, ovary, pancreas, and stomach cancers. Several markers, including Metastasis-Associated Lung Adenocarcinoma Transcript 1 (MALAT1) (a metastasis-promoting lncRNA) and mitochondrial genes Mitochondrially Encoded NADH Dehydrogenase 3 (MT-ND3) and Mitochondrially Encoded NADH Dehydrogenase 2 (MT-ND2), were widely expressed, reflecting common tumor strategies such as angiogenesis, immune modulation[72], and metabolic adaptation through oxidative phosphorylation[73,74]. Other genes of interest include S100 Calcium-Binding Protein A6 (S100A6) and S100 Calcium-Binding Protein A11 (S100A11), both linked to stress responses, epithelial-to-mesenchymal transition (EMT), and increased motility, as well as Mitochondrially Encoded Cytochrome B (MT-CYB), which supports cancer-related metabolic reprogramming through its role in the electron transport chain[75–77]. Ribosomal and metabolic genes such as Glyceraldehyde-3-Phosphate Dehydrogenase (GAPDH), Ribosomal Protein L41 (RPL41), and Ribosomal Protein S29 (RPS29) also showed strong expression, consistent with the biosynthetic demands of rapidly proliferating tumors[78–80].

Notably, while VCX2 was detected in the dataset, its expression was lower compared to dominant markers such as TMBIM4 and RGS5. Nevertheless, the functional importance of TMBIM4 in apoptosis inhibition and RGS5 in vascular remodeling reinforces their roles in promoting tumor survival and establishing permissive microenvironments for metastatic progression[81,82].

Figure 3D presents a tree plot summarizing biomarker enrichment across the dataset, emphasizing pathways and processes linked to tumor growth. Enriched categories included DNA catabolism, nucleocytoplasmic transport, ncRNA processing, ribosome biogenesis, Ras GTPase signaling, and immune regulation, representing core mechanisms of tumor proliferation and genomic instability. Markers such as TMBIM4, RGS5, and MALAT1 were central to processes regulating apoptosis, vascular remodeling, and immune modulation, while ribosomal genes (RPL41, RPS29) reinforced the increased protein synthesis required by cancer cells[83].

The visualization encodes statistical significance by color intensity and pathway size by dot area, highlighting the relative weight of each process. Pathways involving DNA catabolism and nucleocytoplasmic transport involved a large number of genes, underscoring their extensive role in driving genomic instability and tumor growth[84]. Immune regulation pathways, influenced by CEACAM7 and MALAT1, reflected the balance between immune activation, evasion, and tumor-promoting inflammation[85]. Meanwhile, RGS5-associated Ras GTPase signaling pointed to angiogenesis and metastasis[86].

Overall, the tree plot provides a clear and statistically robust overview of transcriptional reprogramming across cancers. It highlights both shared mechanisms—such as ribosome biogenesis and immune modulation—and context-specific pathways, underscoring the complexity of transcriptional networks that collectively sustain cancer progression[87–89].

Building on this analysis, Figures 4A–4G compare differentially expressed genes (DEGs) in HCC with those in seven other cancer types, underscoring the molecular distinctiveness of liver cancer. In Figure 4A, Mitochondrial 2-Like 1 **(MTRNR2L1)** is upregulated in HCC, reflecting mitochondrial stress responses, while Apolipoprotein A2 **(APOA2)** highlights liver-specific lipid metabolism reprogramming[90,91]. Conversely, the marked downregulation of **Albumin (ALB)** compared to bladder cancer indicates impaired synthetic capacity in diseased liver tissue[92]. An immune-related marker, Immunoglobulin Heavy Chain Protein **(IGHGP),** further emphasizes the contribution of immune responses to HCC pathogenesis[93]. In Figure 4B, the comparison with breast cancer shows elevated X-Inactive Specific Transcript **(XIST)** expression in breast tumors, reflecting epigenetic regulatory mechanisms, while HCC maintains higher levels of **APOA2** and **IGHGP**, reinforcing its unique reliance on metabolic reprogramming and immune modulation[94].

In Figure 4C, the comparison between liver and colon cancers shows higher expression of Trefoil Factors (TFF1 and TFF2) in colon cancer, consistent with their roles in epithelial maintenance and mucosal integrity[95]. In contrast, liver cancer exhibits elevated ALB, APOA2, and Immunoglobulin Heavy Constant Gamma 3 (IGHG3), highlighting liver-specific metabolic alterations and B-cell–mediated immune responses[96]. Figure 4D contrasts liver and lung cancers. Lung tumors display high levels of Eukaryotic Translation Initiation Factor 1A, Y-linked (EIF1AY)[97], reflecting enhanced protein synthesis, while HCC shows pronounced expression of fibrinogen components (FGA, FGB, FGG) and Transthyretin (TTR), pointing to hypercoagulability and disrupted metabolic regulation[98].

The role of regulatory genes in HCC progression is reinforced by markers such as RGS5, which is strongly expressed in endothelial cells and macrophages, where it promotes vascular remodeling and neovascularisation[99]. Although VCX2 is less abundant than TMBIM4 or RGS5, its distinct profile in diseased liver tissue suggests a role in chromosomal stability and tumor proliferation, potentially acting synergistically with other markers[100]. Additionally, CEACAM7, expressed in immune cells, contributes to inflammation and immune evasion, further supporting a pro-tumorigenic environment[101].

Figure 4E compares liver and ovarian cancers, highlighting elevated glutathione peroxidase 1 (**GPX1)** in HCC[102], a regulator of oxidative stress, alongside **APOA2** and **IGHGP**, which reinforce its distinct metabolic and immunological profiles. In contrast, ovarian cancer shows high expression of **XIST**, reflecting differences in epigenetic regulation.

In Figure 4F, liver and pancreatic cancers are contrasted. HCC shows higher expression of ATP Synthase Subunit E **(ATP5E)** and Guanine Nucleotide-Binding Protein Subunit Beta-2-Like 1 **(GNB2L1),** genes tied to mitochondrial bioenergetics and metabolic reprogramming[103], while pancreatic cancer relies more on **Osteopontin (SPP1),** associated with ECM remodeling and metastasis, though it is also moderately elevated in HCC[104]. Figure 4G shows that stomach cancer is enriched in **TFF1** and **TFF2**, linked to mucosal defense, whereas HCC exhibits higher **ALB, APOA2, IGHG1, and fibrinogen components (FGA, FGB, FGG),** consistent with inflammation and a pro-thrombotic microenvironment.

Figure 4H examines the biomarker distribution across different cell types. TMBIM4 is broadly expressed, underscoring its role in apoptosis inhibition and stress resilience[81]. While VCX2 is confined mainly to gametocytes and hepatocytes, suggesting a specialized function in chromosomal stability and liver-specific processes[51]. RGS5 shows high expression in osteoblasts, smooth muscle, and stem cells, reflecting its role in vascular remodeling and angiogenesis[47].

The restricted expression of **VCX2**, a CTA, in hepatocytes has significant implications[51,105]. According to Zhang et al. (2023) [109], which usually remains silent in somatic tissues due to DNA methylation, VCX2 can be reactivated in cancer through epigenetic dysregulation, thereby supporting immune evasion and intratumorally heterogeneity. This plasticity suggests potential for VCX2 as both a biomarker and a therapeutic target for epigenetic and immune-based interventions.

Additional genes of note include Keratin 17 (KRT17), linked to epithelial integrity and invasiveness[106]; Keto Reductase Family 1 Member C3 (AKR1C3), associated with steroid metabolism and differentiation[107]; and Endothelial cell-specific chemotaxis receptor (ECSCR), reinforcing endothelial signaling and angiogenesis[108]. Although overshadowed by dominant markers like TMBIM4 and RGS5, low-abundance transcripts such as Ataxin 8 Opposite Strand (ATXN8OS)[109] and RP11-661A12.4 may also play roles in transcriptional regulation and warrant further study.

In summary, Figure 4H highlights the distinct expression patterns of key biomarkers across cell types. TMBIM4 shows broad distribution across nearly all lineages, underscoring its universal role in apoptosis regulation and stress resilience. In contrast, markers such as VCX2, RGS5, KRT17, AKR1C3, and ECSCR display restricted, cell-type–specific expression, suggesting specialized functions[108]. Notably, the enrichment of VCX2 in hepatocytes reinforces its potential as a liver-specific marker in HCC, linking it to chromosomal stability and tumor adaptation. The detection of low-abundance transcripts such as ATXN8OS and RP11-661A12.4 further points to the importance of exploring less prominent markers that may hold significant regulatory roles.

Figure 4I extends this analysis by evaluating cross-cancer gene expression. Expression is quantified by average intensity (color gradients) and the percentage of cells expressing each marker (dot size). VCX2 emerges with pronounced, liver-specific expression, particularly in HCC, underscoring its potential in chromosomal regulation, immune evasion, and tumor progression. As a CTA, its restricted expression makes it especially promising for immunotherapies, including monoclonal antibodies, peptide-based vaccines, or chimeric antigen receptor T-cell (CAR-T) cell approaches. In contrast, TMBIM4 maintains broad expression across organs and cancers, confirming its function as a universal regulator of apoptosis and stress response[110].

Within the liver, **VCX2, RGS5, RP11-661A12.4, and ATXN8OS** stand out as significant biomarkers. VCX2 shows strong specificity for HCC, while **RGS5** contributes to angiogenesis and vascular remodeling[111]. The lncRNAs **RP11-661A12.4** and **ATXN8OS** exhibit liver-restricted expression, indicating potential regulatory roles in metabolic reprogramming and oncogenesis. Together, these markers highlight the

interplay between metabolic, vascular, and epigenetic mechanisms in shaping liver cancer biology and reinforce their potential as diagnostic and therapeutic targets[112].

Figure 4J examines the combinatorial expression of **VCX2, ATXN8OS, and RP11-661A12.4** to identify synergistic interactions driving HCC progression. The strongest effect was observed for **VCX2 + ATXN8OS**, where VCX2's role in chromosomal stability appears to enhance the oncogenic transcriptional activity of ATXN8OS[113,114]. The **VCX2 + RP11-661A12.4** pair also showed synergy, likely through integration of genomic regulation and transcriptional control, whereas the **ATXN8OS + RP11-661A12.4** combination was comparatively weaker. Interestingly, while the complete triad remained highly expressed in pathological tissue, it did not exceed the VCX2 + ATXN8OS pair, underscoring the central influence of VCX2 in liver tumorigenesis. These results position VCX2 as a potential master regulator and immunotherapeutic target in HCC.

Building on these findings, subsequent analyses reinforced VCX2's involvement in chromosomal regulation, transcriptional networks, and tumor progression, supporting its candidacy for targeted therapies. The broader synthesis of Figure 4 emphasizes how transcriptional reprogramming and cellular restructuring—characterized by hepatocyte loss and the expansion of fibroblasts, macrophages, and immune cells—contribute to the pathological microenvironment of HCC. Within this context, TMBIM4 and RGS5 emerge as complementary regulators, promoting survival, angiogenesis, and immune evasion. Together, these biomarkers highlight key oncogenic mechanisms and therapeutic vulnerabilities in liver cancer.

To explore VCX2's broader functional role, we examined its integration within gene and protein interaction networks (Figure 5). The upper panel displays three network diagrams, with Figure 5A illustrating a node-link structure derived from human liver expression data. Here, **VCX2** (light pink) is positioned as a central hub connected to multiple interacting genes and proteins (blue and gray nodes). These associations indicate links to **chromatin organization, immune modulation, and transcriptional regulation,** underscoring VCX2's potential to coordinate chromosomal stability with liver-specific oncogenic signaling. The clustering pattern within the network highlights its role as an integrative regulator in HCC biology.

Figure 5B further establishes VCX2 as a central hub within the interaction network, represented as a red node with the highest centrality scores, underscoring its dominant regulatory potential. VCX2 directly connects to functionally significant genes, including Steroid Sulfatase (STS), Arylsulfatase F (ARSF), Pseudouridine 5'-Phosphatase (PUDP), and Patatin-Like Phospholipase Domain Containing 4 (PNPLA4) (orange/yellow nodes). These interactors span key processes relevant to tumor biology: STS in steroid metabolism and hormone regulation[115], ARSF in glycosaminoglycan metabolism and ECM integrity[116], PUDP in RNA modification and stability[117], and Patatin-Like Phospholipase Domain Containing 4 (PNPLA4) in lipid metabolism, particularly relevant in liver cancer[118]. The size and prominence of these nodes highlight the strength of these associations and reinforce VCX2's role in integrating diverse regulatory pathways in HCC.

The interaction network in Figure 5C highlights the complexity of **VCX2's molecular associations,** linking it to diverse processes including chromosomal regulation, transcriptional control, lipid metabolism, and tumor progression[51]. Color gradients and segmented nodes denote overlapping functions across datasets, while interactions with Heat Shock Transcription Factor Family, X-linked 2 (**HSFX2),** and Melanoma-Associated Antigen 10 (**MAGEA10)** connect VCX2 to other CTAs implicated in immune evasion and proliferation[119]. These associations reinforce VCX2's potential functional relevance but also raise the need for structural validation to ensure that downstream modeling and screening efforts rely on a biologically meaningful conformation.

To evaluate the structural fidelity of VCX2, we performed molecular dynamics simulations across three independently generated models, followed by PCA. The eigenvalue decay curves (Figure S2A, left) revealed consistent variance profiles across replicates, while the PCA projection of the 1000 ns trajectory (Figure S2A, right) identified distinct conformational clusters, indicating that metastable states were sampled during the simulation. Stereochemical validation using Ramachandran plots (Figure S2B) showed a substantial improvement in model quality: residues in favored regions increased from 54% in the AlphaFold-predicted model to 89% after refinement, with disallowed conformations eliminated[120–122]. Global structural evaluation with ProSA-web further supported this refinement, as the Z-score of the optimized model fell within the range of experimentally solved X-ray structures, and local energy distributions were markedly improved (Figure S2C). Finally, structural comparison of the initial and refined models (Figure S2D) demonstrated increased compactness, improved secondary structure definition, and reduced conformational disorder, validating the suitability of the refined VCX2 structure for downstream virtual screening. This refined model served as the foundation for subsequent binding pocket prediction and docking analyses.

Building on this validated model, the three-dimensional model in Figure 5D further highlights VCX2 as a druggable target, revealing distinct ligand-binding pockets (blue, orange, pink). The **blue pocket**, with the highest site score (1.009, Maestro SiteMap), was prioritized for MD simulations and binding energy analysis. Structural refinement produced a stable conformation, with RMSD values stabilizing between 0.05–0.31 nm and Ramachandran plot analysis confirming improved geometry (89% residues in favored regions vs. 54% in the initial AlphaFold model) (Supplementary Information Table S1 and Figure S2). Together, these results establish a structurally robust and druggable model of VCX2, though high-resolution data from X-ray crystallography or cryo-EM remain necessary to fully validate binding site predictions and optimize docking accuracy[123].

A structure-based virtual screening was performed using the **PeruNPDB** database to identify natural inhibitors of VCX2, focusing on the top-ranked binding pocket identified by site mapping. Interactions with binding energies weaker than –2 kcal/mol were filtered out. Figure 5E shows the Z-score distribution of binding affinities, highlighting several compounds with strong predicted interactions. Among these, **Luteolin-5-O-glucoside (PeruNPDB277)** emerged as the top candidate, falling within the high-affinity region. Figure 5F depicts its chemical structure, classifying it as a flavonoid glycoside widely

present in plants such as *Cirsium maackii* and *Equisetum arvense*[124,125]. Notably, *E. arvense* is traditionally used in Peruvian medicine for its diuretic, wound-healing, and anti-inflammatory properties, consistent with the reported anticancer activities of luteolin derivatives[126].

To further characterize its inhibitory potential, we examined the molecular basis of Luteolin-5-O-glucoside binding to VCX2. The multiple hydroxyl groups of the compound facilitate extensive hydrogen bonding within the protein pocket, strengthening the stability of the complex. Consistent with this structural profile, Luteolin-5-O-glucoside achieved a docking score of −7.42 kcal/mol, ranking highest among screened compounds. To refine these predictions, **MM-GBSA** calculations were performed, providing a thermodynamic assessment of binding. The calculated free energy (ΔG_{bind}) was −40.13 ± 1.12 kcal/mol, with a strong Coulombic contribution (−45.97 kcal/mol), indicating spontaneous and highly stable binding. Energy decomposition revealed electrostatic interactions as the dominant stabilizing force, supported by van der Waals contacts and favorable solvation contributions. Together, docking and binding free energy analyses establish Luteolin-5-O-glucoside as a strong inhibitor candidate.

To characterize the molecular determinants of binding, Figure 6 highlights the residues mediating ligand interactions. Key hydrogen bonds were formed with PRO91, GLU97, and GLU109 (Figure 6A), anchoring the compound within the active site. Additional contributions from ARG46, ARG49, ARG50, LYS34, THR35, VAL38, SER24, PRO26, PRO94, GLU92, and GLU93 created a dense network of polar and hydrophobic contacts (Figure 6B), further stabilizing the complex.

MD simulations were performed to assess the dynamic stability of Luteolin-5-O-glucoside within the VCX2 binding pocket. Despite the absence of a crystal structure, the refined homology model provided a reliable framework for long-timescale simulations. Trajectory analysis revealed that key residues—GLU109, GLU96, and PRO91—were consistently critical for ligand stabilization, with interaction energies of −109.46, −105.56, and −79.61 kcal/mol, respectively (Supporting Information Figure S1). These residues, conserved across the VCX family, likely play essential structural and functional roles in mediating ligand recognition and may contribute to oncogenic signaling when dysregulated.

Overall, the identification of Luteolin-5-O-glucoside as a high-affinity VCX2 ligand introduces a promising strategy for therapeutic interference in VCX2-driven pathways. Its strong docking score, favorable binding free energy, and stable MD profile support its potential as a lead scaffold for small-molecule inhibition of VCX2-expressing malignancies. Further experimental validation, structure-activity relationship (SAR) optimization, and preclinical studies will be necessary to advance its candidacy toward targeted therapeutic development. To enhance efficacy and drug-likeness, ultimately advancing its candidacy for targeted therapy.

CONCLUSIONS

This study was designed to identify a specific biomarker for HCC, leading to the selection of VCX2 as a promising candidate. Transcriptomic analyses, including scRNA-seq, revealed differential expression of VCX2 in HCC compared to healthy liver tissue. Pan-cancer comparisons across bladder, breast, colon, lung, ovarian, pancreatic, and stomach cancers further validated VCX2 as a potentially unique marker for liver malignancy, linking it to tumor progression, chromosomal regulation, and oncogenic adaptation.

Building on this validation, we conducted advanced *in silico* studies to assess the druggability of VCX2. Structural modeling and site-mapping identified multiple binding pockets, with the highest-scoring site (site score 1.009) prioritized for docking against the PeruNPDB natural product database. Among the 280 screened compounds, Luteolin-5-O-glucoside (PeruNPDB277), derived from *Equisetum arvense*, emerged as the top candidate. It achieved a docking score of − 7.42 kcal/mol and a binding free energy (ΔG_{bind}) of − 40.13 ± 1.12 kcal/mol, with strong Coulombic contributions (–45.97 kcal/mol). Stabilizing interactions included hydrogen bonds with PRO91, GLU97, and GLU109, supported by electrostatic and van der Waals forces, underscoring its potential as a VCX2 inhibitor.

Although VCX2 may act as an oncogenic driver, its inhibition via small molecules could provide a therapeutic strategy to reduce tumor progression or therapy resistance. However, reliance on AlphaFold and homology-based structural models highlights current limitations, reinforcing the need for experimental structural characterization (e.g., X-ray crystallography, cryo-EM) and *in vitro* validation.

In conclusion, this work integrates single-cell transcriptomics, Pan-cancer validation, structural modeling, and virtual screening to identify VCX2 as a novel HCC biomarker and druggable target. The discovery of Luteolin-5-O-glucoside as a lead compound provides proof-of-concept for precision medicine approaches aimed at modulating VCX2-driven oncogenic pathways. Future efforts should focus on experimental validation, structure-based drug design, and therapeutic development to advance VCX2 inhibitors toward clinical application.

Declarations

Author Contributions: Conceptualization, L.D.G.M., M.A.C.F.; methodology, L.D.G.M., H.L.B.C., and; validation, L.D.G.M., H.L.B.C., M.A.C.F., N.M.H., ; formal analysis, L.D.G.M., H.L.B.C., M.A.C.F.; investigation, L.D.G.M., H.L.B.C., and; resources, M.A.C.F.; writing—original draft preparation, L.D.G.M., M.A.C.F.; writing—review and editing, L.D.G.M., H.L.B.C., M.A.C.P., M.A.C.F., N.M.H., and; supervision, M.A.C.F.; and All authors have read and agreed to the published version of the manuscript.

Funding: This work was funded by the Consejo Nacional de Ciencia, Tecnología e Innovación Tecnológica (CONCYTEC), the Programa Nacional de Investigación Científica y Estudios Avanzados (PROCIENCIA), by the call “E067-2023-01 Proyectos Especiales: Proyectos de Incorporación de Investigadores Postdoctorales en Instituciones Peruanas” [número de contrato PE501084367-2023], and Grant PE501088204-2024. Also, it was funded by the Universidad Católica de Santa María (grants 27499-R-2020, 27574-R-2020, 7309-CU-2020, and 28048-R-2021) and by the Research Management Office from the Universidad Católica de Santa María.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The software used in this study includes Maestro (Schrödinger Suite), Desmond, SiteMap, and Glide, all of which require a paid license for both academic and commercial use. Additionally, STRING, Cytoscape, and PASSer were utilized, which are freely available for non-commercial use. All software tools are referenced throughout the Methods and Results sections.

Furthermore, protein structures, docking grids, molecular dynamics trajectories, ligand preparation files, energy minimization data, and analysis scripts generated during this study are documented accordingly. All simulation scripts, docking results, molecular interaction analyses, and processed datasets can be found at https://figshare.com/s/d13d8bcd01fbf5487038 and https://github.com/CompBioChemRG/SingleCell_CBCRG.

ACKNOWLEDGMENTS:

The authors express their gratitude for the financial support from the Programa Nacional de Investigación Científica y Estudios Avanzados—PROCIENCIA (PE501088204-2024 and PE501084367-2023).

Conflicts of Interest: The authors declare no conflicts of interest.

SUPPORTING INFORMATION

Additional structural and simulation data (PDF): Includes MM-GBSA per-residue interaction energy analysis (Figure S1); Ramachandran plots, ProSA-web Z-scores, residue-wise energy profiles for VCX2 models, principal component analysis of VCX2 dynamics (Figure S2); and RMSD values over a 999 ns MD trajectory (Table S1).

References

1. Toh, M. R. *et al.* Global epidemiology and genetics of hepatocellular carcinoma. *Gastroenterology***164**, 766–782 (2023).
2. Siegel, R. L., Kratzer, T. B., Giaquinto, A. N., Sung, H. & Jemal, A. Cancer statistics, 2025. *CA. Cancer J. Clin.***n/a**, (2025).
3. Guo, Q. *et al.* Projected epidemiological trends and burden of liver cancer by 2040 based on GBD, CI5plus, and WHO data. *Sci. Rep.***14**, 28131 (2024).
4. Perz, J. F., Armstrong, G. L., Farrington, L. A., Hutin, Y. J. F. & Bell, B. P. The contributions of hepatitis B virus and hepatitis C virus infections to cirrhosis and primary liver cancer worldwide. *J. Hepatol.***45**, 529–538 (2006).

5. Hammad, R. *et al.* Cytotoxic T cell expression of leukocyte-associated immunoglobulin-like receptor-1 (LAIR-1) in viral hepatitis C-mediated hepatocellular carcinoma. *Int. J. Mol. Sci.***23**, 12541 (2022).
6. Kafeero, H. M. *et al.* Prevalence and predictors of hepatitis B virus (HBV) infection in east Africa: evidence from a systematic review and meta-analysis of epidemiological studies published from 2005 to 2020. *Arch. Public Heal.***79**, 1–19 (2021).
7. Fan, J.-G., Kim, S.-U. & Wong, V. W.-S. New trends on obesity and NAFLD in Asia. *J. Hepatol.***67**, 862–873 (2017).
8. Fattovich, G., Stroffolini, T., Zagni, I. & Donato, F. Hepatocellular carcinoma in cirrhosis: incidence and risk factors. *Gastroenterology***127**, S35–S50 (2004).
9. Galle, P. R. *et al.* Biology and significance of alpha-fetoprotein in hepatocellular carcinoma. *Liver Int.***39**, 2214–2229 (2019).
10. Zhao, Y. *et al.* Functional drug–target–disease network analysis of gene–phenotype connectivity for curcumin in hepatocellular carcinoma. *PeerJ***9**, e12339 (2021).
11. Liu, Y. *et al.* Oxyresveratrol prevents murine H22 hepatocellular carcinoma growth and lymph node metastasis via inhibiting tumor angiogenesis and lymphangiogenesis. *J. Nat. Med.***72**, 481–492 (2018).
12. Singh, A. K. *et al.* Current therapeutic modalities and chemopreventive role of natural products in liver cancer: Progress and promise. *World J. Hepatol.***15**, 1 (2023).
13. Seeff, L. B., Bonkovsky, H. L., Navarro, V. J. & Wang, G. Herbal products and the liver: a review of adverse effects and mechanisms. *Gastroenterology***148**, 517–532 (2015).
14. Kaur, N., Kaur, B. & Sirhindi, G. Phytochemistry and pharmacology of Phyllanthus niruri L.: a review. *Phyther. Res.***31**, 980–1004 (2017).
15. MacParland, S. A. *et al.* Single cell RNA sequencing of human liver reveals distinct intrahepatic macrophage populations. *Nat. Commun.***9**, 4383 (2018).
16. Clough, E. *et al.* NCBI GEO: archive for gene expression and epigenomics data sets: 23-year update. *Nucleic Acids Res.***52**, D138–D144 (2024).
17. Zeng, J. *et al.* CancerSCEM: a database of single-cell expression map across various human cancers. *Nucleic Acids Res.***50**, D1147–D1155 (2022).
18. D'Avola, D. *et al.* High-density single cell mRNA sequencing to characterize circulating tumor cells in hepatocellular carcinoma. *Sci. Rep.***8**, 11570 (2018).
19. Lee, H. W. *et al.* Single-cell RNA sequencing reveals the tumor microenvironment and facilitates strategic choices to circumvent treatment failure in a chemorefractory bladder cancer patient. *Genome Med.***12**, 1–21 (2020).
20. Gao, R. *et al.* Delineating copy number and clonal substructure in human tumors from single-cell transcriptomes. *Nat. Biotechnol.***39**, 599–608 (2021).
21. Lee, H.-O. *et al.* Lineage-dependent gene expression programs influence the immune landscape of colorectal cancer. *Nat. Genet.***52**, 594–603 (2020).

22. Laughney, A. M. *et al.* Regenerative lineages and immune-mediated pruning in lung cancer metastasis. *Nat. Med.***26**, 259–269 (2020).
23. Lambrechts, D. *et al.* Phenotype molding of stromal cells in the lung tumor microenvironment. *Nat. Med.***24**, 1277–1289 (2018).
24. Nelson, L. *et al.* A living biobank of ovarian cancer ex vivo models reveals profound mitotic heterogeneity. *Nat. Commun.***11**, 822 (2020).
25. Peng, J. *et al.* Single-cell RNA-seq highlights intra-tumoral heterogeneity and malignant progression in pancreatic ductal adenocarcinoma (vol 29, pg 777, 2019). *CELL Res.***29**, 777 (2019).
26. Zhang, P. *et al.* Dissecting the single-cell transcriptome network underlying gastric premalignant lesions and early gastric cancer. *Cell Rep.***30**, 4317 (2020).
27. Team, R. C. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria; 2021. R version 4.1. 0.
28. Team, R. C. RA language and environment for statistical computing, R Foundation for Statistical. *Computing* (2020).
29. Aran, D. *et al.* Reference-based analysis of lung single-cell sequencing reveals a transitional profibrotic macrophage. *Nat. Immunol.***20**, 163–172 (2019).
30. Xu, S. *et al.* Using clusterProfiler to characterize multiomics data. *Nat. Protoc.* 1–29 (2024).
31. Yu, G. Enrichplot: visualization of functional enrichment result. *R Packag. version***1**, (2021).
32. Blanco-Carmona, E. Generating publication ready visualizations for Single Cell transcriptomics using SCpubr. *bioRxiv* 2002–2022 (2022).
33. Szklarczyk, D. *et al.* STRING v10: protein–protein interaction networks, integrated over the tree of life. *Nucleic Acids Res.***43**, D447–D452 (2015).
34. Shannon, P. *et al.* Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.***13**, 2498–2504 (2003).
35. Chin, C.-H. *et al.* cytoHubba: identifying hub objects and sub-networks from complex interactome. *BMC Syst. Biol.***8**, 1–7 (2014).
36. Gorodkin, J., Jensen, L., Doncheva, N. & Morris, J. Cytoscape StringApp: Network Analysis and Visualization of Proteomics Data. *J. Proteome Res.***18**, (2019).
37. Palasca, O., Santos, A., Stolte, C., Gorodkin, J. & Jensen, L. J. TISSUES 2.0: an integrative web resource on mammalian tissue expression. *Database***2018**, bay003 (2018).
38. Release, S. Glide: Schrödinger. *LLC, NY* (2020).
39. Lu, C. *et al.* OPLS4: Improving force field accuracy on challenging regimes of chemical space. *J. Chem. Theory Comput.***17**, 4291–4300 (2021).
40. Barazorda-Ccahuana, H. L. *et al.* PeruNPDB: the Peruvian Natural Products Database for in silico drug screening. *Sci. Rep.***13**, 7577 (2023).
41. Weininger, D. SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules. *J. Chem. Inf. Comput. Sci.***28**, 31–36 (1988).

42. Dallakyan, S. & Olson, A. J. Small-molecule library screening by docking with PyRx. *Methods Mol. Biol.***1263**, 243–250 (2015).
43. Trott, O. & Olson, A. J. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J. Comput. Chem.***31**, 455–461 (2010).
44. Carloni, V., Luong, T. V. & Rombouts, K. Hepatic stellate cells and extracellular matrix in hepatocellular carcinoma: more complicated than ever. *Liver Int.***34**, 834–843 (2014).
45. Ali, N. A., Hamdy, N. M., Gibriel, A. A. & El Mesallamy, H. O. Investigation of the relationship between CTLA4 and the tumor suppressor RASSF1A and the possible mediating role of STAT4 in a cohort of Egyptian patients infected with hepatitis C virus with and without hepatocellular carcinoma. *Arch. Virol.***166**, 1643–1651 (2021).
46. Rojas-Rivera, D. & Hetz, C. TMBIM protein family: ancestral regulators of cell death. *Oncogene***34**, 269–280 (2015).
47. Arnold, C. *et al.* RGS 5 promotes arterial growth during arteriogenesis. *EMBO Mol. Med.***6**, 1075–1089 (2014).
48. Beauchemin, N. & Arabzadeh, A. Carcinoembryonic antigen-related cell adhesion molecules (CEACAMs) in cancer progression and metastasis. *Cancer Metastasis Rev.***32**, 643–671 (2013).
49. Im, J.-Y. *et al.* CYB5R3 functions as a tumor suppressor by inducing ER stress-mediated apoptosis in lung cancer cells via the PERK-ATF4 and IRE1α-JNK pathways. *Exp. Mol. Med.***56**, 235–249 (2024).
50. Rodríguez López, S. Mechanisms of longevity extension and healthy aging by over-expression of cytochrome b5 reductase 3. (2021).
51. Jakobsen, M. K. *et al.* The cancer/testis antigen gene VCX2 is rarely expressed in malignancies but can be epigenetically activated using DNA methyltransferase and histone deacetylase inhibitors. *Front. Oncol.***10**, 584024 (2021).
52. Gilgenkrantz, H., Paradis, V. & Lotersztajn, S. Cell metabolism-based therapy for liver fibrosis, repair, and hepatocellular carcinoma. *Hepatology***81**, 269–287 (2025).
53. Chen, Y. *et al.* TMBIM4 deficiency facilitates NLRP3 inflammasome activation-induced pyroptosis of trophoblasts: a potential pathogenesis of preeclampsia. *Biology (Basel).***12**, 208 (2023).
54. Yang, C. *et al.* Liver fibrosis: insights into migration of hepatic stellate cells in response to extracellular matrix and growth factors. *Gastroenterology***124**, 147–159 (2003).
55. Capece, D. *et al.* The inflammatory microenvironment in hepatocellular carcinoma: a pivotal role for tumor-associated macrophages. *Biomed Res. Int.***2013**, 187204 (2013).
56. Xue, C., Li, G., Lu, J. & Li, L. Crosstalk between circRNAs and the PI3K/AKT signaling pathway in cancer progression. *Signal Transduct. Target. Ther.***6**, 400 (2021).
57. Franke, T. F., Hornik, C. P., Segev, L., Shostak, G. A. & Sugimoto, C. PI3K/Akt and apoptosis: size matters. *Oncogene***22**, 8983–8998 (2003).
58. Norhalifah, H. K., Mat, N. F. C. & Edinur, H. A. Cytokine gene polymorphisms in cancer and inflammatory disorders. *Curr. Immunol. Rev.***14**, 81–93 (2018).

59. Munn, L. L. Cancer and inflammation. *Wiley Interdiscip. Rev. Syst. Biol. Med.***9**, e1370 (2017).
60. Hora, S. & Wuestefeld, T. Liver Injury and Regeneration: Current Understanding, New Approaches, and Future Perspectives. *Cells***12**, 2129 (2023).
61. Ma, L., Li, C. C. & Wang, X. W. Roles of Cellular Neighborhoods in Hepatocellular Carcinoma Pathogenesis. *Annu. Rev. Pathol. Mech. Dis.***20**, 169–192 (2025).
62. Dhar, D., Baglieri, J., Kisseleva, T. & Brenner, D. A. Mechanisms of liver fibrosis and its role in liver cancer. *Exp. Biol. Med.***245**, 96–108 (2020).
63. Scagliola, A., Miluzio, A. & Biffo, S. Translational control of metabolism and cell cycle progression in hepatocellular carcinoma. *Int. J. Mol. Sci.***24**, 4885 (2023).
64. Yilmaz, M. & Christofori, G. EMT, the cytoskeleton, and cancer cell invasion. *Cancer Metastasis Rev.***28**, 15–33 (2009).
65. Malhi, H. & Kaufman, R. J. Endoplasmic reticulum stress in liver disease. *J. Hepatol.***54**, 795–809 (2011).
66. Li, J., Jung, W., Nam, S., Chaudhuri, O. & Kim, T. Roles of interactions between cells and extracellular matrices for cell migration and matrix remodeling. *Multi-scale Extracell. matrix Mech. Mechanobiol.* 247–282 (2020).
67. Taylor, K. R. & Gallo, R. L. Glycosaminoglycans and their proteoglycans: host-associated molecular patterns for initiation and modulation of inflammation. *FASEB J.***20**, 9–22 (2006).
68. Bonner, J. C. Regulation of PDGF and its receptors in fibrotic diseases. *Cytokine Growth Factor Rev.***15**, 255–273 (2004).
69. Baglieri, J., Brenner, D. A. & Kisseleva, T. The role of fibrosis and liver-associated fibroblasts in the pathogenesis of hepatocellular carcinoma. *Int. J. Mol. Sci.***20**, 1723 (2019).
70. Pickup, M. W., Mouw, J. K. & Weaver, V. M. The extracellular matrix modulates the hallmarks of cancer. *EMBO Rep.***15**, 1243–1253 (2014).
71. Eble, J. A. & Niland, S. The extracellular matrix in tumor progression and metastasis. *Clin. Exp. Metastasis***36**, 171–198 (2019).
72. Goyal, B. *et al.* Diagnostic, prognostic, and therapeutic significance of long non-coding RNA MALAT1 in cancer. *Biochim. Biophys. Acta (BBA)-Reviews Cancer***1875**, 188502 (2021).
73. Xu, Y., Su, J., Zhou, J., Yuan, Q. & Han, J. Roles of MT-ND1 in Cancer. *Curr. Med. Sci.***43**, 869–878 (2023).
74. Mao, X. *et al.* Genetic diversities of MT-ND3 and MT-ND4L genes are associated with high-altitude adaptation. *Mitochondrial DNA Part B***4**, 324–328 (2019).
75. Melle, C., Ernst, G., Schimmel, B., Bleul, A. & von Eggeling, F. Colon-derived liver metastasis, colorectal carcinoma, and hepatocellular carcinoma can be discriminated by the Ca2+-binding proteins S100A6 and S100A11. *PLoS One***3**, e3767 (2008).
76. Cross, S. S., Hamdy, F. C., Deloulme, J.-C. & Rehman, I. Expression of S100 proteins in normal human tissues and common cancers using tissue microarrays: S100A6, S100A8, S100A9 and S100A11 are

all overexpressed in common cancers. *Histopathology***46**, 256–269 (2005).

77. Yi, Y. *et al.* Mitochondrial-cytochrome c oxidase II promotes glutaminolysis to sustain tumor cell survival upon glucose deprivation. *Nat. Commun.***16**, 212 (2025).
78. Das, M. R. *et al.* Molecular association of glucose-6-phosphate isomerase and pyruvate kinase M2 with glyceraldehyde-3-phosphate dehydrogenase in cancer cells. *BMC Cancer***16**, 1–12 (2016).
79. Molavi, G., Samadi, N. & Hosseingholi, E. Z. The roles of moonlight ribosomal proteins in the development of human cancers. *J. Cell. Physiol.***234**, 8327–8341 (2019).
80. Mao-De, L. & Jing, X. Ribosomal proteins and colorectal cancer. *Curr. Genomics***8**, 43–49 (2007).
81. Liu, Q. TMBIM-mediated Ca2+ homeostasis and cell death. *Biochim. Biophys. Acta (BBA)-Molecular Cell Res.***1864**, 850–857 (2017).
82. Silini, A. *et al.* Regulator of G-protein signaling 5 (RGS5) protein: a novel marker of cancer vasculature elicited and sustained by the tumor's proangiogenic microenvironment. *Cell. Mol. life Sci.***69**, 1167–1178 (2012).
83. Michalik, K. M. *et al.* Long noncoding RNA MALAT1 regulates endothelial cell function and vessel growth. *Circ. Res.***114**, 1389–1397 (2014).
84. Talens, F. & Van Vugt, M. A. T. M. Inflammatory signaling in genomically instable cancers. *Cell Cycle***18**, 1830–1848 (2019).
85. Li, K. *et al.* Micro electrical fields induced MSC-sEVs attenuate neuronal cell apoptosis by activating autophagy via lncRNA MALAT1/miR-22-3p/SIRT1/AMPK axis in spinal cord injury. *J. Nanobiotechnology***21**, 451 (2023).
86. Crosas-Molist, E. *et al.* Rho GTPase signaling in cancer progression and dissemination. *Physiol. Rev.***102**, 455–510 (2022).
87. Ala, U. & Fagoonee, S. RNA-binding protein transcripts as potential biomarkers for detecting Primary Sclerosing Cholangitis and for predicting its progression to Cholangiocarcinoma. *Front. Mol. Biosci.***11**, 1388294 (2024).
88. Wang, S. *et al.* RPL41, a small ribosomal peptide deregulated in tumors, is essential for mitosis and centrosome integrity. *Neoplasia***12**, 284-IN8 (2010).
89. Lallemant, B. *et al.* Reference gene selection for head and neck squamous cell carcinoma gene expression studies. *BMC Mol. Biol.***10**, 1–10 (2009).
90. Rochette, L., Meloux, A., Zeller, M., Cottin, Y. & Vergely, C. Role of humanin, a mitochondrial-derived peptide, in cardiovascular disorders. *Arch. Cardiovasc. Dis.***113**, 564–571 (2020).
91. Zaki, M. E., Amr, K. S. & Abdel-Hamid, M. APOA2 polymorphism in relation to obesity and lipid metabolism. *Cholesterol***2013**, 289481 (2013).
92. Fu, X., Yang, Y. & Zhang, D. Molecular mechanism of albumin in suppressing invasion and metastasis of hepatocellular carcinoma. *Liver Int.***42**, 696–709 (2022).
93. Dunn-Walters, D., Townsend, C., Sinclair, E. & Stewart, A. Immunoglobulin gene analysis as a tool for investigating human immune responses. *Immunol. Rev.***284**, 132–147 (2018).

94. Vincent-Salomon, A. *et al.* X inactive–specific transcript RNA coating and genetic instability of the X chromosome in BRCA1 breast tumors. *Cancer Res.***67**, 5134–5140 (2007).
95. Hoffmann, W. Trefoil factors: TFF (trefoil factor family) peptide-triggered signals promoting mucosal restitution. *Cell. Mol. Life Sci. C.***62**, 2932–2938 (2005).
96. Zhao, J. *et al.* Current insights into the expression and functions of tumor-derived immunoglobulins. *Cell Death Discov.***7**, 148 (2021).
97. Fan, H. *et al.* Gender differences of B cell signature in healthy subjects underlie disparities in incidence and course of SLE related to estrogen. *J. Immunol. Res.***2014**, 814598 (2014).
98. Kenarangi, T., Bakhshi, E., InanlooRahatloo, K. & Biglarian, A. Identification of gene signature in RNA-Seq hepatocellular carcinoma data by Pareto-optimal cluster algorithm. *Gastroenterol. Hepatol. From Bed to Bench***15**, 387 (2022).
99. Kong, P. *et al.* RGS5 maintaining vascular homeostasis is altered by the tumor microenvironment. *Biol. Direct***18**, 78 (2023).
100. Cuevas-Covarrubias, S. A. & González-Huerta, L. M. Analysis of the VCX3A, VCX2 and VCX3B genes shows that VCX3A gene deletion is not sufficient to result in mental retardation in X-linked ichthyosis. *Br. J. Dermatol.***158**, 483–486 (2008).
101. Kelleher, M., Singh, R., O'Driscoll, C. M. & Melgar, S. Carcinoembryonic antigen (CEACAM) family members and inflammatory bowel disease. *Cytokine Growth Factor Rev.***47**, 21–31 (2019).
102. Lei, X. G., Cheng, W.-H. & McClung, J. P. Metabolic regulation and function of glutathione peroxidase-1. *Annu. Rev. Nutr.***27**, 41–61 (2007).
103. Huang, Y.-J. *et al.* ATP synthase subunit epsilon overexpression promotes metastasis by modulating AMPK signaling to induce epithelial-to-mesenchymal transition and is a poor prognostic marker in colorectal cancer patients. *J. Clin. Med.***8**, 1070 (2019).
104. Nallasamy, P. *et al.* Pancreatic tumor microenvironment factor promotes cancer stemness via SPP1–CD44 axis. *Gastroenterology***161**, 1998–2013 (2021).
105. Yin, J., Gu, T., Chaudhry, N., Davidson, N. E. & Huang, Y. Epigenetic modulation of antitumor immunity and immunotherapy response in breast cancer: biological mechanisms and clinical implications. *Front. Immunol.***14**, (2024).
106. Zhang, H. *et al.* The role of Keratin17 in human tumours. *Front. Cell Dev. Biol.***10**, 818416 (2022).
107. Li, M. *et al.* AKR1C3 in carcinomas: from multifaceted roles to therapeutic strategies. *Front. Pharmacol.***15**, 1378292 (2024).
108. Verma, A. *et al.* Endothelial cell–specific chemotaxis receptor (ecscr) promotes angioblast migration during vasculogenesis and enhances VEGF receptor sensitivity. *Blood, J. Am. Soc. Hematol.***115**, 4614–4622 (2010).
109. Deng, Z. *et al.* lncRNA ATXN8OS promotes breast cancer by sequestering miR-204. *Mol. Med. Rep.***20**, 1057–1064 (2019).

110. Carrara, G., Parsons, M., Saraiva, N. & Smith, G. L. Golgi anti-apoptotic protein: a tale of camels, calcium, channels and cancer. *Open Biol.***7**, 170045 (2017).
111. Hamzah, J. *et al.* Vascular normalization in Rgs5-deficient tumours promotes immune destruction. *Nature***453**, 410–414 (2008).
112. Roh, J., Im, M., Chae, Y., Kang, J. & Kim, W. The involvement of long non-coding RNAs in glutamine-metabolic reprogramming and therapeutic resistance in Cancer. *Int. J. Mol. Sci.***23**, 14808 (2022).
113. Kumar, S., Gonzalez, E. A., Rameshwar, P. & Etchegaray, J.-P. Non-coding RNAs as mediators of epigenetic changes in malignancies. *Cancers (Basel).***12**, 3657 (2020).
114. Prabhakaran, R. *et al.* Epigenetic frontiers: miRNAs, long non-coding RNAs and nanomaterials are pioneering to cancer therapy. *Epigenetics Chromatin***17**, 31 (2024).
115. Reed, M. J., Purohit, A., Woo, L. W. L., Newman, S. P. & Potter, B. V. L. Steroid sulfatase: molecular biology, regulation, and inhibition. *Endocr. Rev.***26**, 171–202 (2005).
116. Feferman, L. *et al.* Arylsulfatase B (N-acetylgalactosamine-4-sulfatase): potential role as a biomarker in prostate cancer. *Prostate Cancer Prostatic Dis.***16**, 277–284 (2013).
117. Yu, J. *et al.* Bioinformatics analysis combined with experiments predicts PUDP as a potential prognostic biomarker for hepatocellular carcinoma through its interaction with tumor microenvironment. *Front. Oncol.***12**, 830174 (2022).
118. Björnson, E. *et al.* Stratification of hepatocellular carcinoma patients based on acetate utilization. *Cell Rep.***13**, 2014–2026 (2015).
119. Zendman, A. J. W., Van Kraats, A. A., Weidle, U. H., Ruiter, D. J. & Van Muijen, G. N. P. The XAGE family of cancer/testis-associated genes: alignment and expression profile in normal tissues, melanoma lesions and Ewing's sarcoma. *Int. J. cancer***99**, 361–369 (2002).
120. Lyu, J. *et al.* AlphaFold2 structures template ligand discovery. *BioRxiv* 2012–2023 (2024).
121. Baselious, F. *et al.* Comparative Structure-Based Virtual Screening Utilizing Optimized AlphaFold Model Identifies Selective HDAC11 Inhibitor. *International Journal of Molecular Sciences* vol. 25 at https://doi.org/10.3390/ijms25021358 (2024).
122. Alhumaid, N. K. & Tawfik, E. A. Reliability of AlphaFold2 Models in Virtual Drug Screening: A Focus on Selected Class A GPCRs. *International Journal of Molecular Sciences* vol. 25 at https://doi.org/10.3390/ijms251810139 (2024).
123. García-Nafría, J. & Tate, C. G. Cryo-electron microscopy: moving beyond X-ray crystal structures for drug receptors and drug development. *Annu. Rev. Pharmacol. Toxicol.***60**, 51–71 (2020).
124. Jung, H. A., Jin, S. E., Min, B.-S., Kim, B.-W. & Choi, J. S. Anti-inflammatory activity of Korean thistle Cirsium maackii and its major flavonoid, luteolin 5-O-glucoside. *Food Chem. Toxicol.***50**, 2171–2179 (2012).
125. Carneiro, D. M. *et al.* Equisetum arvense: new evidences supports medical use in daily clinic. *Pharmacogn. Rev***13**, 50–58 (2019).

126. Parameshwaran, K., Almaghrabi, M., Govindarajulu, M., Clark, R. C. & Dhanasekaran, M. Bioactives and Pharmacology of Equisetum arvense L. in *Bioactives and Pharmacology of Medicinal Plants* 429–440 (Apple Academic Press, 2022).

Figures

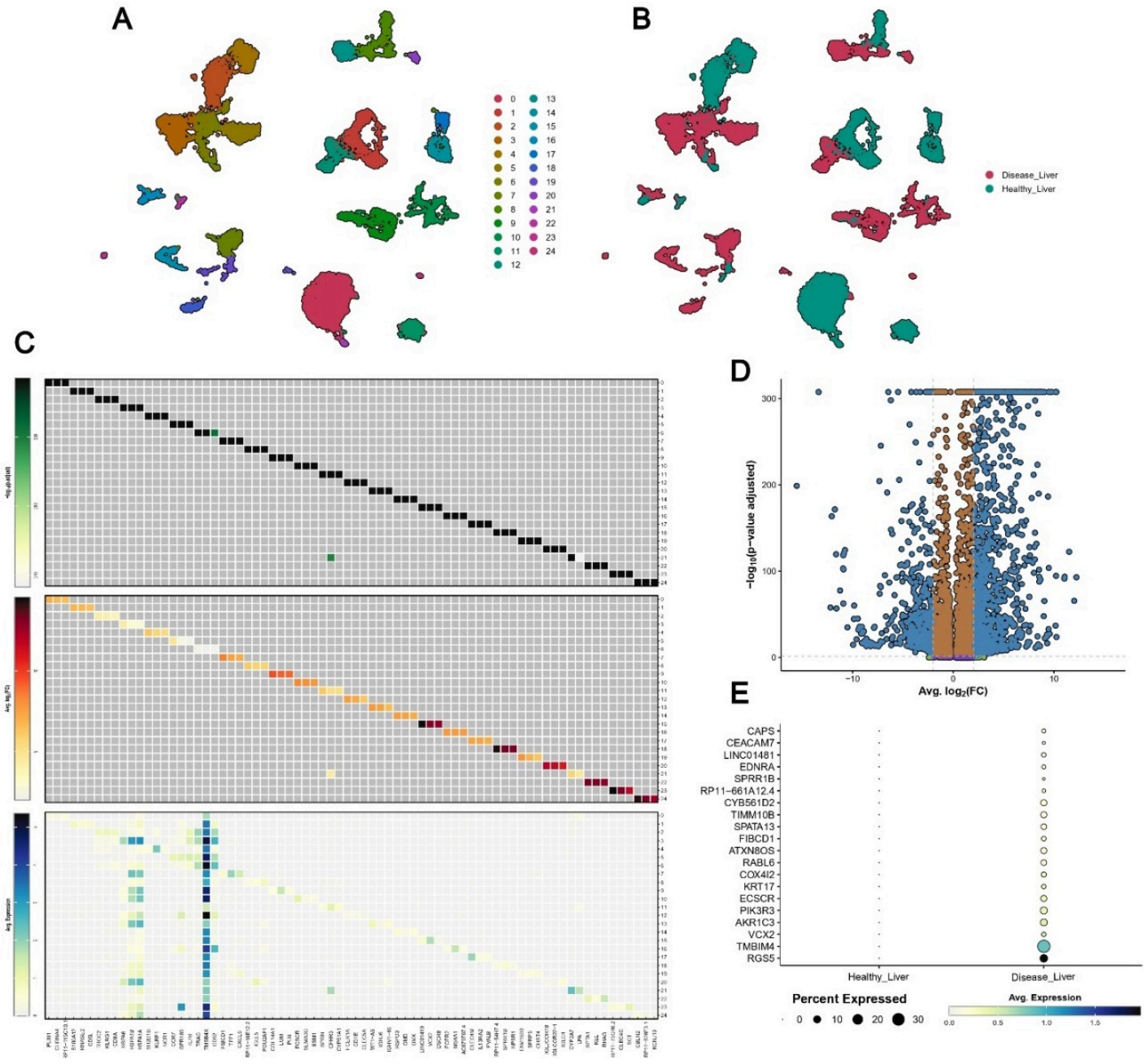

Figure 1

Integrated single-cell analysis of liver cell populations and transcriptional changes in health and disease. **(A)**UMAP plot of single-cell RNA-seq data showing major liver cell types. Healthy liver is dominated by

hepatocytes (green), whereas diseased tissue shows expansion of fibroblasts (brown), macrophages (blue), and immune subsets, including T cells (dark pink) and B cells (light pink). **(B)** UMAP overlay of healthy (green) versus diseased (red) samples highlights transcriptional divergence and shifts in cellular composition. **(C)** Heatmaps depict gene relationships and cluster-specific markers, with the upper heatmap showing correlations among the top three genes per cell cluster. **(D)** Volcano plot of differential expression between healthy liver and HCC. **(E)** Dot plot summarizes expression dynamics across liver cell types, with color intensity representing average expression and dot size reflecting the percentage of expressing cells.

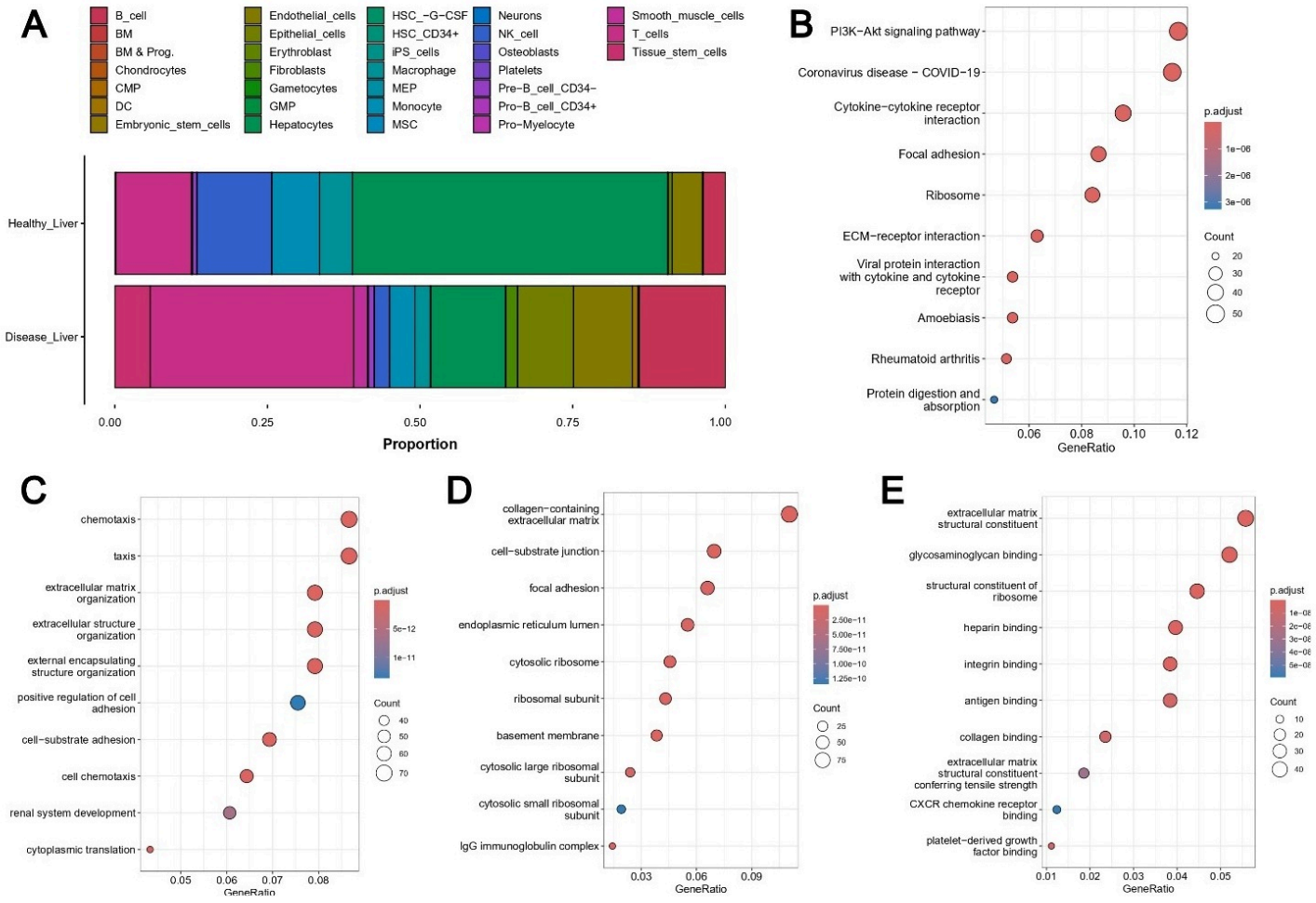

Figure 2

Comprehensive analysis of molecular, cellular, and functional alterations in healthy and diseased liver tissues. **(A)**Proportional representation of cell types in healthy and diseased liver tissues. **(B)** KEGG pathway enrichment analysis. **(C)** Biological process enrichment. **(D)** Cellular component enrichment. **(E)**Molecular function enrichment.

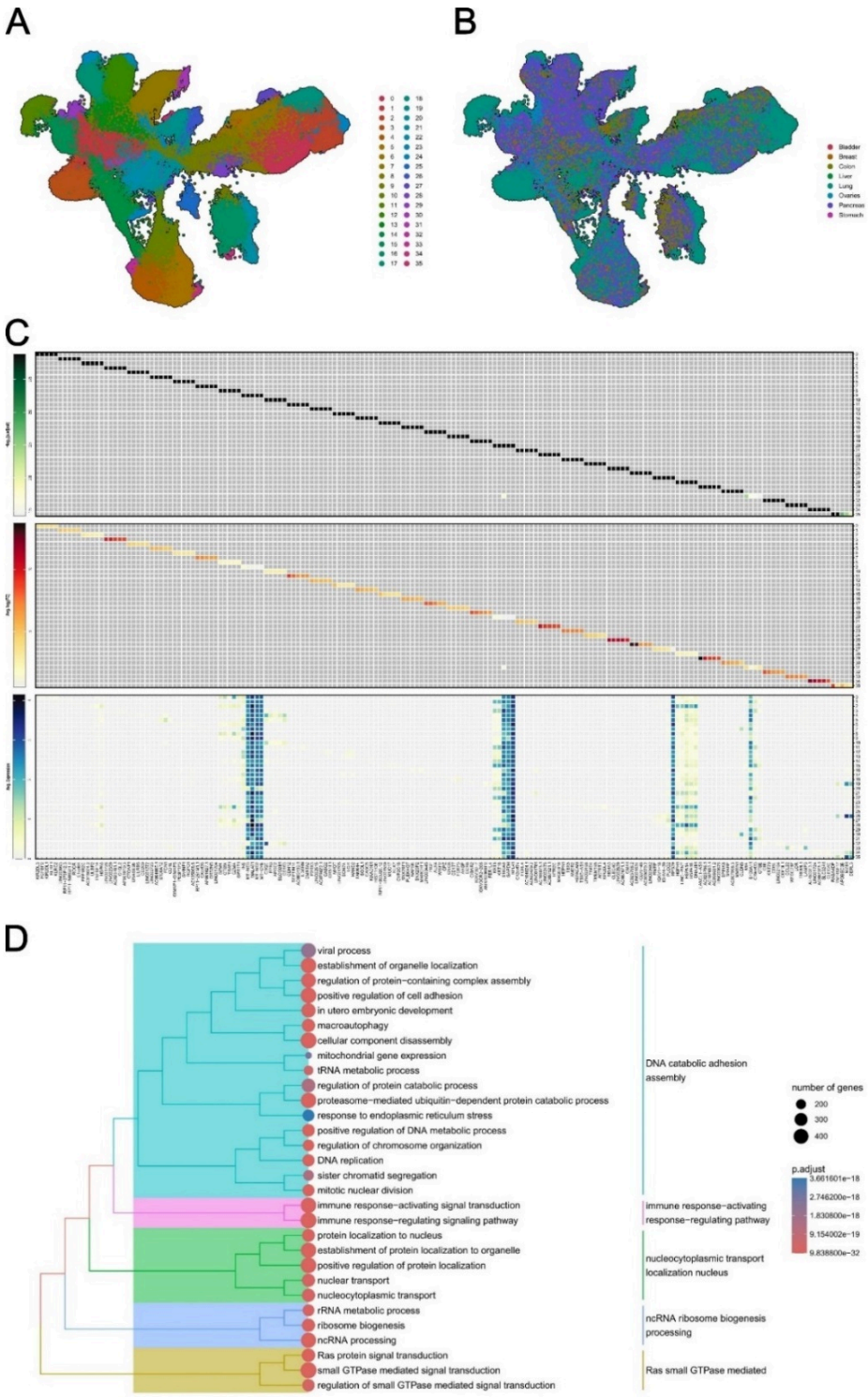

Figure 3

Comprehensive analysis of cancer biomarker expression and pathway enrichment across multiple cancer types. **(A)** and **(B)** present clustering maps derived from eight cancer datasets, identifying transcriptional patterns and distinct gene expression clusters associated with specific cancer types. **(C)** showcases a heatmap highlighting the top markers per cluster, illustrating their expression levels and

functional roles across cancer datasets. **(D)**displays a tree plot of enrichment analysis for all biomarkers in the dataset, with pathways color-coded to represent their involvement in biological processes.

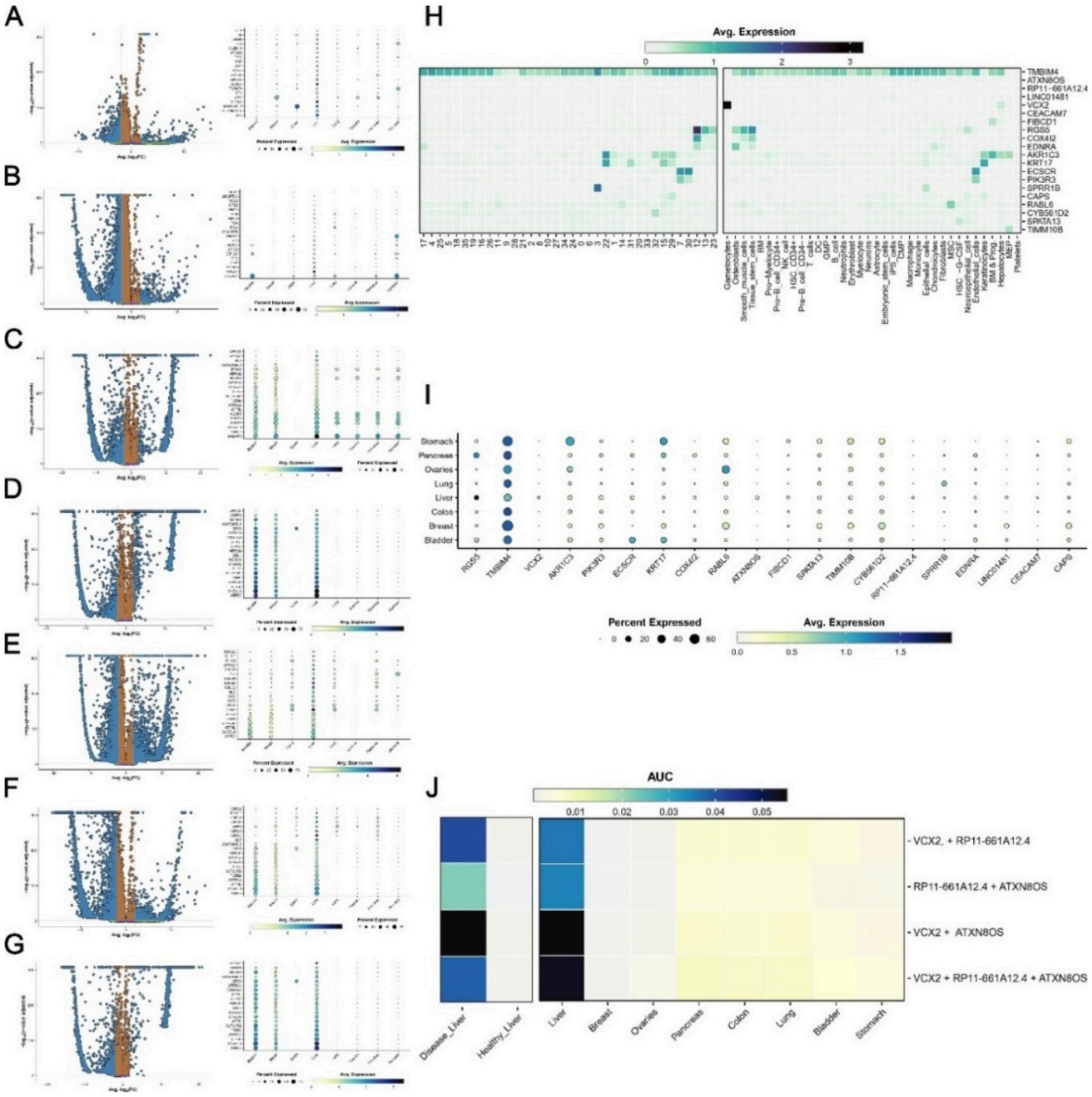

Figure 4

The volcano plots show differentially expressed genes in liver cancer than **(A)** bladder, **(B)** breast, **(C)** colon, **(D)** lung, **(E)** ovaries, **(F)** pancreas, and **(G)** stomach cancers. The dot plots represent the results of the top 20 genes per comparison. **(H)** heatmap showing the results for each cluster, cell annotations for the genes expressed differently in HCC vs healthy liver, and **(I)** the dot plot of the same genes by cancer

organ dataset. **(J)**enrichment heatmap for a combination of markers against healthy and diseased liver datasets and the eight cancer tissue datasets.

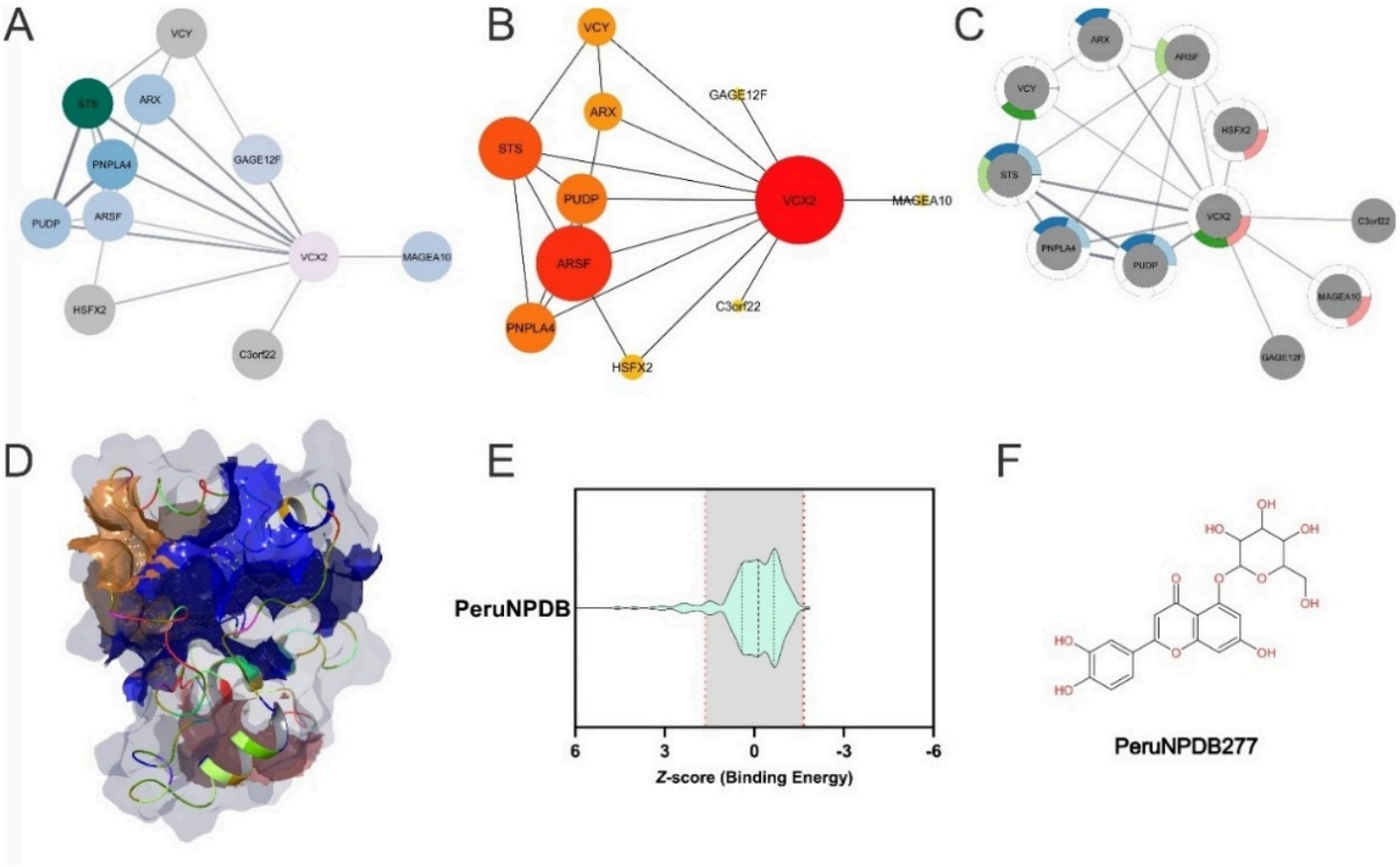

Figure 5

Network analysis and therapeutic exploration of VCX2. (Top) Three network diagrams illustrating VCX2's molecular interactions. **(A)** Node-link diagram depicting VCX2 (pink) and its associated genes. **(B)** Centralized network emphasizing key partners. **(C)**Expanded interaction network, where color gradients highlight shared genes or proteins. **(D)** Identification of druggable pockets in VCX2. **(E)**Molecular docking results and binding affinity analysis (Z-score) to assess candidate compounds. **(F)** Structural representation of PeruNPDB277 (Luteolin-5-O-glucoside).

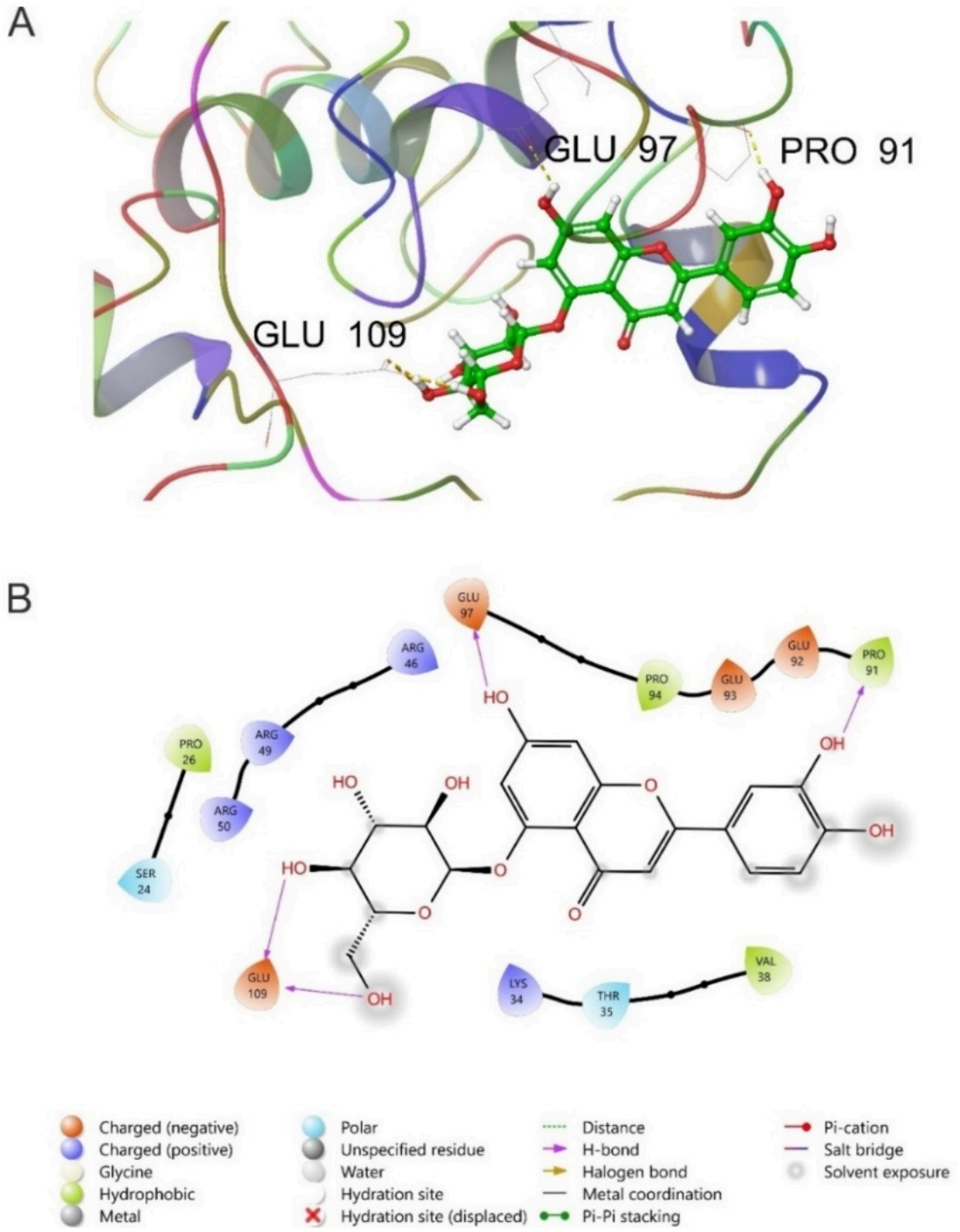

Figure 6

Molecular interactions of Luteolin-5-O-glucoside with VCX2. **(A)** 3D binding conformation of Luteolin-5-O-glucoside (green) within the VCX2 active site, showing hydrogen bonds (yellow dashed lines) with PRO91, GLU97, and GLU109. **(B)** 2D interaction map highlighting key ligand-protein interactions, including hydrogen bonding, pi-pi stacking, and electrostatic interactions.

Supplementary Files

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

- supplementaryScientificreportsGoyzuetaD2025c.docx