Cell-cycle reactivation and hepatocyte identity loss in hepatocellular carcinoma: Transcriptomic hallmarks and validation strategies (Review)

JICHEN HOU[1], BING ZHU[2] and LU LIANG[1]

[1]Department of Hepatobiliary Surgery, Affiliated Baotou Clinical College of Inner Mongolia Medical University, Baotou Central Hospital, Baotou, Inner Mongolia 014040, P.R. China;
[2]Translational Medicine Center, Baotou Central Hospital, Baotou, Inner Mongolia 014040, P.R. China

Received May 20, 2026; Accepted June 18, 2026

DOI: 10.3892/ol.2026.15758

Abstract. Hepatocellular carcinoma (HCC) is molecularly diverse, but repeated transcriptomic surveys point to a useful recurring contrast: Cell-cycle and DNA-replication programs become prominent while differentiated hepatocyte transport, as well as metabolic and innate-defense functions decline. In the present study, this paired change was used as a practical lens for reading HCC transcriptomes. Rather than treating individual proliferation-, transporter- or metabolism-related genes as isolated markers, the review asks what cellular state these modules represent and what evidence is needed before they are discussed as mechanisms or therapeutic vulnerabilities. The present review outlines how proliferation-identity imbalance can guide model choice, RT-qPCR or RNA-seq panels, protein and pathway readouts and pharmacological testing. The current review also includes a worked example showing how a published HCC transcriptomic signal can be interpreted through this framework. The framework is deliberately conservative: Transcriptomic modules can nominate states and experiments, but they do not by themselves establish target dependency, mechanisms or therapeutic efficacy.

Correspondence to: Dr Lu Liang, Department of Hepatobiliary Surgery, Affiliated Baotou Clinical College of Inner Mongolia Medical University, Baotou Central Hospital, 61 Huancheng Road, Donghe, Baotou, Inner Mongolia 014040, P.R. China
E-mail: 1291201364@qq.com

Abbreviations: AURKA, aurora kinase A; CCNB1, cyclin B1; CYP, cytochrome P450; DEG, differentially expressed gene; EdU, 5-ethynyl-2'-deoxyuridine; GLS2, glutaminase 2; HCC, hepatocellular carcinoma; MKI67, marker of proliferation Ki-67; OCT1, organic cation transporter 1; PCNA, proliferating cell nuclear antigen; PI, propidium iodide; RNA-seq, RNA sequencing; RRM2, ribonucleotide reductase regulatory subunit M2; RT-qPCR, reverse transcription-quantitative PCR; SLCO, solute carrier organic anion transporter; TOP2A, DNA topoisomerase IIα

Key words: hepatocellular carcinoma, transcriptomics, cell cycle, hepatocyte identity, DNA replication, metabolic reprogramming, bioinformatics, model selection, validation ladder, single-cell transcriptomics

Contents

1. Introduction

Background. Hepatocellular carcinoma (HCC) still accounts for a large share of cancer-related mortality, even though surveillance, locoregional treatment, systemic therapy and immunotherapy have improved the outlook for selected patients (1-6). Treatment decisions depend on tumour burden, liver reserve, performance status, molecular heterogeneity and tolerability, and current clinical guidelines continue to emphasize that biological interpretation must be integrated with liver-function context (2-6). Biologically,

most HCCs arise in a chronically injured liver. Chronic viral hepatitis, metabolic injury, inflammation, fibrosis and repeated regeneration create a background in which malignant transformation is difficult to describe by mutation status alone (2,7,8). Non-coding RNA regulation and transcriptomic remodeling further add regulatory layers that are relevant to HCC biology and precision interpretation (7-13). Transcriptomic profiles therefore remain useful, because they record malignant-cell programs together with residual signals from the diseased liver environment (7-10).

Transcriptomic cell-state remodeling in HCC. Across many HCC transcriptomic datasets, the most recognizable signal is not a single hub gene. It is a paired shift: Mitotic, DNA-replication and replication-stress programs become stronger, whereas transporter, detoxification, metabolic and defense-related transcripts become weaker (11,12,14-27). Representative proliferation-associated examples include DNA topoisomerase IIα (TOP2A), cyclin B1 (CCNB1) and ribonucleotide reductase regulatory subunit M2 (RRM2), whereas hepatocyte-identity-related examples include solute carrier family 22 member 1 (SLC22A1), solute carrier organic anion transporter family member 1B3 (SLCO1B3), glutaminase 2 (GLS2) and cytochrome P450 (CYP) family genes (14-27). Reading these two arms together is often more informative than ranking isolated differentially expressed genes (DEGs), because the paired signal links tumour growth, differentiation state, model suitability and validation design. Representative transcriptomic modules relevant to this proliferation-identity imbalance are summarized in Table I.

This view builds on earlier HCC molecular-classification studies. Transcriptome-based subclassifications separated proliferative, poorly differentiated and more hepatocyte-like tumour groups, showing that HCC expression states have biological structure rather than forming random marker collections (11,12). The present review uses those state-level observations for a different purpose: To connect recurrent expression patterns with model selection, assay design and claim strength.

No connectivity-mapping platform was used to generate data, perform compound ranking or support the conclusions of the present narrative review. Connectivity Map (https://clue.io/cmap), the Library of Integrated Network-based Cellular Signatures (https://lincsproject.org/) and L1000 Fireworks Display (https://maayanlab.cloud/l1000fwd/) are mentioned only to distinguish the present state-interpretation framework from platform-centred drug-repositioning studies.

Aim and objectives of this review. This review aims to clarify how the recurring proliferation-identity pattern in HCC transcriptomes should be interpreted and translated into validation strategies without overstating what expression data can prove. Specifically, it will: i) Summarize recurrent transcriptomic hallmarks involving cell-cycle reactivation and hepatocyte identity loss; ii) explain how these paired signals can guide experimental model selection and validation design; iii) distinguish module-level hypotheses from mechanistic or therapeutic claims; and iv) demonstrate the practical use of the framework through a worked example based on published HCC transcriptomic evidence.

Novelty and contribution. The recurring HCC transcriptomic signal is considered to be a cell-state problem. In that reading, cell-cycle activation and erosion of hepatocyte identity are not two unrelated findings; together, they help decide which model is appropriate, which reverse transcription-quantitative PCR (RT-qPCR) or RNA sequencing (RNA-seq) readout is worth using and when a result is still only a hypothesis rather than evidence for a mechanism or therapy.

2. Scope, conceptual approach and bioinformatics considerations

Scope and conceptual approach. Across bulk and integrated HCC transcriptomic studies, proliferative tumours commonly show enrichment of cell-cycle and DNA-replication programs, whereas differentiated hepatocyte transport, metabolic and defense-related programs are frequently reduced (11,12,14-27). Single-cell and spatial studies further indicate that the cellular origin of these signals can vary across malignant, immune, stromal and adjacent-liver compartments (28-34). These observations motivate the practical questions addressed in this review: Which patterns repeat in bulk or integrated datasets; whether proliferative and hepatocyte-identity genes move together or apart; what single-cell and spatial studies add to bulk signatures; which experimental model preserves the state being discussed; and at what point a module-level observation becomes an overclaim about mechanism or treatment. Table I summarizes representative modules and cautions, whereas Table II links selected modules to model features and claim boundaries.

Review design rationale. This article is a structured narrative review rather than a systematic review or meta-analysis and was not prepared according to PRISMA 2020 (35). The evidence comes from different literatures, including bulk transcriptomics, single-cell and spatial studies, pathway papers, model systems and validation reports. These materials do not share a common effect measure that could be pooled usefully; therefore, no PRISMA flow diagram or pooled estimate was prepared. The aim is an interpretive framework with explicit limits.

Search strategy and selection criteria. Searches were updated on 14 May 2026 in PubMed (https://pubmed.ncbi.nlm.nih.gov/), Web of Science (https://www.webofscience.com/) and Scopus (https://www.scopus.com/). Google Scholar (https://scholar.google.com/) was used as a secondary tool to follow citations from key HCC transcriptomic, single-cell, spatial-omics and model-system papers. Search terms combined HCC-related words ('hepatocellular carcinoma', 'HCC', 'liver cancer'), transcriptomic terms ('transcriptome', 'gene signature', 'cell cycle', 'DNA replication', 'hepatocyte identity', 'single-cell', 'spatial transcriptomics') and translational terms ('organoid', 'xenograft', 'drug response', 'transporter', 'SLC22A1', 'OCT1', 'metabolic reprogramming'). Original studies, relevant reviews and papers informative for experimental model choice were prioritized. Purely prognostic studies were retained only when they helped interpret cell state, model suitability or validation logic. Articles were excluded from narrative emphasis when

Table I. Transcriptomic modules relevant to HCC cell-state remodeling.

Module	Representative genes/programs	Biological interpretation	Main caution
Cell cycle/mitosis	TOP2A, CCNB1, AURKA, CDK1	Active cycling and mitosis (14-19)	May mainly mark cycling fraction
DNA replication/repair	RRM2, MCM-family genes, PCNA-related programs	Nucleotide demand and replication stress (14-19,42)	Non-specific if read alone
Hepatocyte transport	SLC22A1, SLCO1B3	Reduced liver transport and drug handling (20-22)	Transporter biology varies by model
Metabolism/detoxification	GLS2, CYP-family genes	Remodeled liver metabolic identity (23-27)	May reflect tissue context or purity
Immune/defense	Complement and defense-response genes	Innate-defense and microenvironment signal (9,28-34)	Bulk origin may be non-malignant

The listed genes are examples of broader programs and should not be read as individually validated targets. AURKA, aurora kinase A; CCNB1, cyclin B1; CDK1, cyclin-dependent kinase 1; CYP, cytochrome P450; GLS2, glutaminase 2; HCC, hepatocellular carcinoma; MCM, minichromosome maintenance; PCNA, proliferating cell nuclear antigen; RRM2, ribonucleotide reductase regulatory subunit M2; SLC22A1, solute carrier family 22 member 1; SLCO1B3, solute carrier organic anion transporter family member 1B3; TOP2A, DNA topoisomerase IIα.

Table II. Linking transcriptomic modules to testable vulnerabilities.

Transcriptomic signal	Possible vulnerability	Useful model feature	Claim boundary
High cell-cycle/mitotic module	Cell-cycle kinase or mitotic dependency (19,41)	Proliferative HCC line with measurable cycling fraction	Needs perturbation assays
High DNA replication/ RRM2 module	Replication-stress or nucleotide-synthesis dependency (42)	Model with robust S-phase fraction	Needs functional assays
Low hepatocyte transporter module	Drug uptake/exposure limitation or differentiation loss (20-22)	Model preserving transporter biology	Needs exposure testing
Metabolic gene rewiring	Nutrient/redox/lipid metabolic dependency (23-27)	Model matched to metabolic phenotype	Avoid broad metabolism claims
Immune/defense shift	Microenvironmental or inflammatory interaction (9,28-34)	Co-culture, organoid or spatial context	Needs cellular-source evidence

The table separates plausible hypotheses, model features and claim boundaries. AURKA, aurora kinase A; CCNB1, cyclin B1; CYP, cytochrome P450; EdU, 5-ethynyl-2'-deoxyuridine; HCC, hepatocellular carcinoma; PI, propidium iodide; RRM2, ribonucleotide reductase regulatory subunit M2; RT-qPCR, reverse transcription-quantitative PCR; SLC22A1, solute carrier family 22 member 1; SLCO1B3, solute carrier organic anion transporter family member 1B3.

they lacked direct relevance to HCC or HCC-relevant experimental models, provided insufficient methodological detail, focused solely on prognostic associations without informing cell-state interpretation or validation design, or made mechanistic or therapeutic claims unsupported by functional evidence.

Reporting boundary. This review synthesizes published evidence without generating a new dataset, pooling effect estimates or performing *de novo* statistical analyses. Reported findings were interpreted according to the original study design, tissue source, analytical platform, model system and validation depth. Therefore, the conclusions represent evidence-informed narrative guidance rather than meta-analytic estimates of effect.

Bioinformatics considerations for interpreting HCC transcriptomic signatures. For HCC transcriptomic interpretation, the first task is not only to rank genes by fold change, but also to determine whether the signal represents a coherent biological state. A stable proliferative module should usually include several concordant markers of DNA replication, mitotic entry

or chromosome segregation, such as TOP2A, RRM2, CCNB1, cyclin-dependent kinase 1 (CDK1), marker of proliferation Ki-67 (MKI67), proliferating cell nuclear antigen (PCNA) and aurora kinase A (AURKA), rather than one isolated gene. A hepatocyte-identity module should likewise be evaluated as a coordinated liver-function program involving transporters, metabolic enzymes and liver-enriched transcriptional or secretory markers, rather than a single downregulated transcript (11,12,14-27).

A second task is to consider cellular origin. Bulk RNA sequencing (RNA-seq) can detect strong recurrent signals, but it cannot by itself determine whether a signal originates from malignant hepatocytes, cycling stromal or immune cells, or endothelial cells, or whether it reflects differences between tumour and adjacent chronically injured liver. Single-cell and spatial evidence should therefore be used, where available, to test whether proliferation-associated and hepatocyte-identity signals map to the malignant epithelial compartment or to the microenvironment (28-34).

A third task is to align the analysis with validation. If the computational result is a proliferation-identity imbalance, then model choice and assay choice should preserve and test that state. This means that a cell line, organoid or xenograft should be selected because it retains the relevant proliferative, metabolic or transporter phenotype. The same logic applies to validation readouts: reverse transcription-quantitative PCR (RT-qPCR) or RNA-seq can confirm coordinated transcript changes, whereas protein assays, cell-cycle profiling, 5-ethynyl-2'-deoxyuridine (EdU) incorporation and rescue or perturbation experiments are needed before mechanistic language becomes defensible (14-19,36-39).

3. Cell-cycle reactivation as a recurrent HCC transcriptomic hallmark

Cell-cycle and DNA-replication programs are among the most consistently upregulated signals in HCC. TOP2A, CCNB1, RRM2, AURKA and CDK1 often appear together within mitotic, chromosome-segregation and replication-stress modules (14-19). Their recurrence makes them useful markers of a proliferative state. It does not, however, mean that each gene is a primary driver or a stand-alone therapeutic target.

TOP2A participates in resolving topological stress during DNA replication and chromosome segregation. In HCC and other cancers, high TOP2A expression generally tracks with proliferation and aggressive biology, which provides a rationale for considering topoisomerase-related mechanisms in an appropriate experimental setting (40). CCNB1 and cyclin B2 mark cell-cycle transition and mitotic entry. AURKA is involved in centrosome and spindle regulation, and cyclin-dependent kinase 4/6-related evidence illustrates that cell-cycle perturbation can intersect with immune-related effects in HCC models (41). RRM2 links DNA-replication demand with nucleotide metabolism and replication stress and has been examined as a candidate HCC vulnerability in integrated in silico and *in vitro* work (42). Read as a module, these genes describe a cycling tumour state more reliably than any single marker.

The risk is familiar in bioinformatics studies: a frequently upregulated cell-cycle gene can be quickly promoted into an independent target, even though high expression may instead reflect tumour purity, the fraction of cycling cells, sampling of an aggressive region or a shared proliferation program. A more defensible use of the signal is to select assays and models. EdU incorporation, colony formation, propidium iodide (PI)-based cell-cycle profiling and protein-level assessment of CCNB1, AURKA, TOP2A or RRM2 provide stronger evidence than expression differences alone (14-19,28-34).

For that reason, cell-cycle genes should first be treated as state markers. Therapeutic language is justified only after functional dependency, pathway engagement and model-consistent effects have been shown.

4. Hepatocyte identity loss and downregulated liver-function programs

The other arm of the paired pattern is loss of hepatocyte identity. Downregulated transcripts often include transporters, detoxification enzymes, complement and coagulation genes, metabolic regulators and other liver-enriched functions. Representative examples include SLC22A1, SLCO1B3, CYP-family genes and GLS2. These signals matter because they influence tumour classification, drug uptake, metabolic competence and model interpretation (20-27).

SLC22A1 encodes organic cation transporter 1 (OCT1), which is relevant to hepatic uptake of endogenous metabolites and several drugs (20-22). CYP enzymes and SLCO-family transporters add a second pharmacological layer by linking the expression state to metabolism and liver-specific exposure. GLS2, a glutaminase associated with mitochondrial metabolism and redox balance, may represent a liver-metabolic component that should be interpreted separately from generic proliferation-associated nutrient demand (23-27).

This matters clinically because loss of transporter and metabolic programs can change how a candidate intervention is interpreted. Reduced OCT1/SLC22A1 or SLCO expression may limit intracellular drug exposure, whereas altered CYP and detoxification programs can shift assumptions about dose relevance and toxicity. In HCC, where cirrhosis and impaired liver function are common, transporter competence and metabolic identity are part of the pharmacological context rather than background annotation (20-27).

The same point applies to experimental interpretation. A compound may reduce proliferation in a poorly differentiated monolayer cell line without testing transporter-dependent uptake, hepatocyte metabolism or liver-specific toxicity. Conversely, restoration of one liver-function transcript is not enough to claim restored hepatocyte identity; stronger language requires coordinated changes across transporter, metabolic and functional readouts in models suited to the question (36-39).

Taken together, loss of hepatocyte identity is not only a diagnostic or descriptive signal. It changes the meaning of drug exposure, model suitability and liver-relevant pharmacology. A downregulated signature can be listed in a table, but its practical value lies in deciding what the model can and cannot test.

5. Proliferation-identity coupling: A more useful framework than single-gene interpretation

Proliferation-identity imbalance is a convenient name for a recurring, but not universal, HCC contrast. Replication and

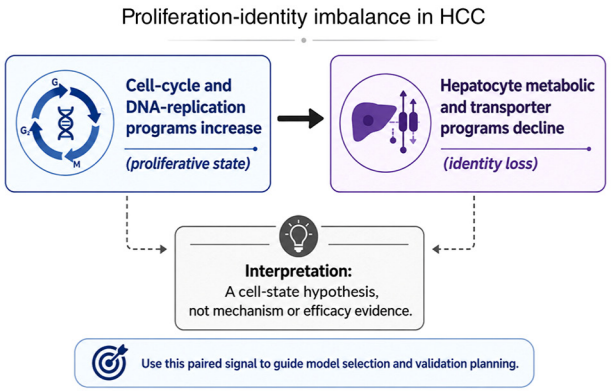

Figure 1. Conceptual summary of proliferation-identity imbalance in HCC. The schematic illustrates a recurrent transcriptomic pattern in HCC in which cell-cycle progression and DNA-replication programs increase, whereas hepatocyte-associated metabolic and transporter programs decline. The lower interpretation panel emphasizes that this paired pattern is intended as a cell-state hypothesis that can guide module-level interpretation, model selection and validation planning, rather than as direct evidence of a specific mechanism or therapeutic efficacy. HCC, hepatocellular carcinoma.

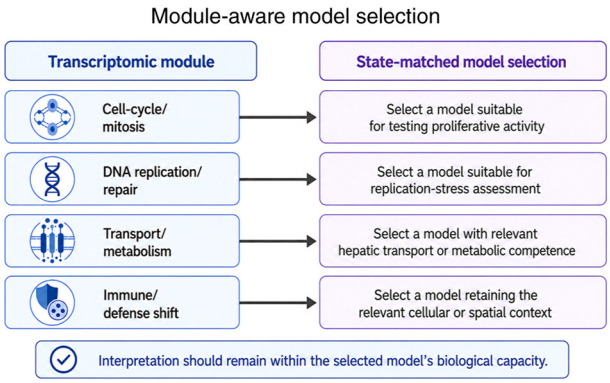

Figure 2. Module-aware model selection framework for HCC transcriptomic programs. The schematic presents the high-level relationship between transcriptomic modules and state-matched model-selection principles. Module-specific hypotheses, useful model features and claim boundaries are detailed in Table II, whereas assay categories and validation evidence are detailed in Table III. HCC, hepatocellular carcinoma.

mitotic programs tend to rise as differentiated liver functions weaken. The phrase is useful because it puts two findings that are often discussed separately into the same frame and slows the common move from 'gene is differentially expressed' to 'gene is a mechanism or target'.

The concept does not flatten HCC heterogeneity. A tumour with intense cell-cycle activity and marked loss of liver-function transcripts may behave differently from a tumour with moderate proliferation but stronger immune or metabolic remodeling. The model is therefore a guide for comparison, model choice and claim wording, not a new universal class of HCC.

Fig. 1 gives a schematic version of the idea. It should not be read to mean that proliferation and hepatocyte identity always change in lockstep. Its main use is more modest: Examining both arms together is usually more informative than judging either arm in isolation.

Used in that limited sense, the proliferation-identity frame helps organize recurring expression patterns while leaving room for subtype, etiology and microenvironmental variation (11,12,14-19,28-34).

6. Single-cell and spatial transcriptomics refine bulk signatures

Bulk transcriptomics is useful for detecting reproducible signals, but it averages over cell types. A change seen in a bulk HCC sample may come from malignant hepatocytes, stromal cells, immune cells, endothelial cells or the injured non-tumour liver. Single-cell and spatial approaches can separate some of these sources, which is particularly important for immune, complement, matrix and endothelial programs (28-34).

Recent single-cell and spatial studies make this point concrete. Some prognostic or candidate therapeutic signals are concentrated in malignant-cell subsets, immune interfaces or tumour-host metabolic niches rather than spread evenly through the tumour mass (28-34). For bulk-signature work, this means that the cellular source of a signal should be checked before it is used to justify a target, model or validation assay.

Spatial location also matters. HCC arises in a diseased organ, and the tumour core, invasive margin and cirrhotic liver around the tumour may carry different expression states. A tumour-vs.-adjacent comparison compresses these differences into one average. In translational work, that average can be misleading if the candidate signal is absent from malignant cells or is mainly stromal or immune (28-34).

A more defensible workflow is therefore iterative: start with stable bulk or multi-cohort patterns; use single-cell or spatial evidence to ask where the signal originates; then choose a validation model because it preserves the relevant state rather than because it is convenient (14-19,28-34,36-39). Fig. 2 presents this high-level model-selection logic.

Single-cell and spatial data are valuable when they change the interpretation or the experiment. If they are presented without a link to the cell type, tissue location or assay being tested, their contribution is limited (28-34).

7. Candidate therapeutic vulnerabilities inferred from transcriptomic modules

Transcriptomic patterns can suggest candidate vulnerability hypotheses, but the evidence level must remain explicit. Cell-cycle activation points toward cyclin-dependent kinase, Aurora kinase, checkpoint, mitotic-spindle and replication-stress biology. An expression pattern alone is still not a validated target. A vulnerability claim requires a functional phenotype, dose-response behaviour, pathway or protein readout and, preferably, consistency across more than one HCC model (19,41,42).

Metabolic remodeling in HCC encompasses altered glycolysis, lipid synthesis, amino-acid metabolism, redox homeostasis and mitochondrial function, although the contribution of each program varies across tumours and experimental contexts (23-27). These changes can intersect with proliferation because a dividing tumour requires nucleotides, lipids, amino acids and reducing power. Metabolic signals should therefore be interpreted together with proliferative state. A

useful translational question is whether the signal implies nutrient dependence, redox vulnerability, transporter-mediated exposure or immune suppression that can be tested (23-27).

Model selection is the third translational axis. Huh7, PLC/PRF/5, Hep3B, MHCC97H and other HCC models differ in differentiation state, TP53 status, viral background, transporter expression and proliferation rate (43). HepG2 is frequently used in liver-cancer experiments, but it has been shown to be a hepatoblastoma-derived cell line rather than a hepatocellular carcinoma cell line (44); it should therefore not be regarded as a representative HCC model. Model choice should follow the expression state under study rather than convenience alone. For example, Table II distinguishes a high cell-cycle module, for which a proliferative model and functional perturbation assays are required, from a low-transporter module, for which transporter-competent models and uptake testing are needed before altered drug exposure can be inferred.

Patient-derived tumour organoids, patient-derived xenograft models and related patient-derived systems can partly bridge this gap because they retain more genomic, histological or treatment-response context than conventional monolayer cultures (36-39,43). They are particularly useful when the question involves transporter competence, metabolic phenotype, intratumoural heterogeneity or drug-sensitivity ranking. Even these models need caution: organoid culture success, stromal and immune loss, medium composition and growth-factor dependence can alter the cell state being tested.

A model suitable for testing a mitotic phenotype may therefore be inadequate for evaluating transporter-dependent drug exposure, hepatocyte-identity restoration or microenvironment-dependent biology (36-39,43).

8. A minimal validation package for HCC transcriptomic modules

A practical RT-qPCR panel needs both arms of the proposed state. TOP2A, CCNB1, RRM2 and AURKA can represent proliferation, whereas SLC22A1, SLCO1B3, GLS2 and CYP1A2 can represent hepatocyte identity or metabolic-function loss. The panel does not replace RNA-seq; its purpose is to test, at a manageable scale, whether a perturbation shifts a coordinated expression state rather than one isolated gene (14-27).

Validation should be scaled to the claim. A viability assay alone supports only preliminary growth inhibition. If EdU incorporation and cell-cycle profiling also change, the interpretation can move toward proliferation or cell-cycle modulation. If selected module genes shift in the expected direction by RT-qPCR or RNA-seq, it becomes reasonable to discuss a module-level transcriptional response. Mechanistic or translational wording should wait for protein or pathway readouts, target perturbation, model extension and, where relevant, *in vivo* support (14-19,36-43).

Fig. 3 and Table III present a validation ladder rather than a mandatory checklist. Fig. 3 provides the stage sequence, whereas Table III specifies the assay categories, the evidence each stage can support and the limitations that remain. The purpose is to prevent a common shortcut: treating altered expression or a short viability experiment as if it already proved a mechanism (14-19,36-43).

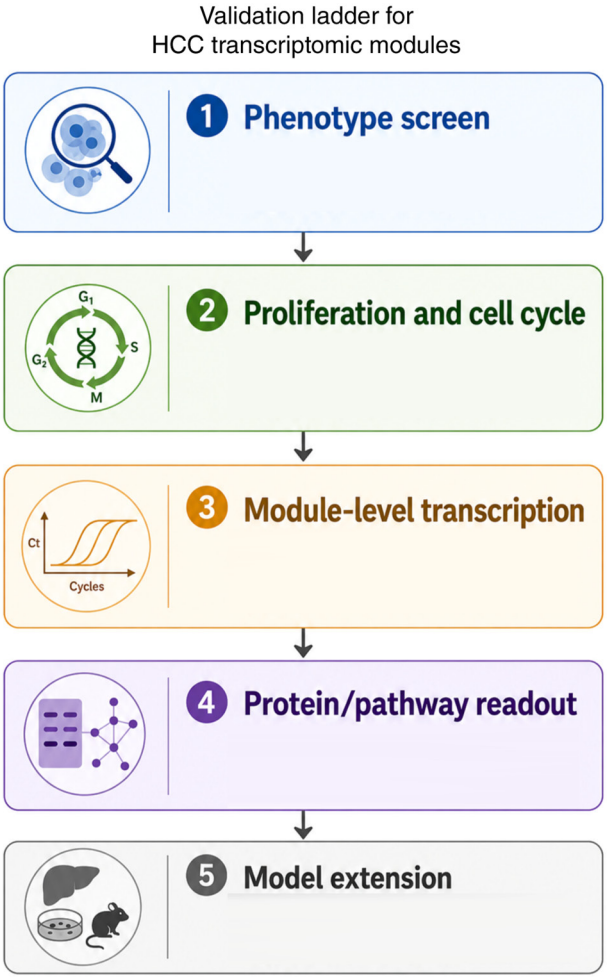

Figure 3. Stepwise validation ladder for HCC transcriptomic modules. The figure displays only the sequence of five validation levels from phenotype screening to model extension. Detailed assay categories, evidentiary support and remaining limitations are provided in Table III. HCC, hepatocellular carcinoma; IF, immunofluorescence; PDX, patient-derived xenograft; RNA-seq, RNA sequencing; RT-qPCR, reverse transcription-quantitative PCR.

In practical terms, a claim should move stepwise. First comes phenotype screening, followed by proliferation and cell-cycle assessment, then a coordinated transcriptional response, and only then protein, pathway or target-level support. A single viability assay, or one gene measured alone, can justify a preliminary observation, not a mechanistic conclusion (14-19,36-43).

9. Worked example: Applying the proliferation-identity framework to The Cancer Genome Atlas Liver Hepatocellular Carcinoma (TCGA-LIHC)

A concrete example can be drawn from the TCGA-LIHC cohort, a widely used bulk transcriptomic resource for HCC (7). In this dataset-level setting, the first step is not to rank a single gene as the most important finding, but to ask whether the tumour profile shows a coordinated state shift. Published TCGA-LIHC and integrated HCC analyses

Table III. Minimal validation ladder for HCC transcriptomic cell-state hypotheses.

Validation step	Core assays	What it supports	What it does not prove
Phenotype screen	Cell Counting Kit-8, CellTiter-Glo and colony formation	Growth or survival effect	Mechanism or module reversal
Proliferation and cell cycle	EdU, PI cell-cycle flow cytometry	DNA synthesis and cell-cycle redistribution	Target engagement
Module RT-qPCR	TOP2A, CCNB1, RRM2, AURKA, SLC22A1, GLS2 and selected context genes	Partial transcriptional support for module shift	Whole-transcriptome reversal
Pathway/protein readout	Western blot, immunofluorescence and kinase/pathway markers	Protein-level pathway modulation	Clinical relevance
Model extension	Second HCC cell line, organoid, xenograft or PDX model	Context robustness and translational plausibility	Patient benefit

Each step indicates what the assay can support and what remains unproven without further evidence. The staged framework is informed by published HCC transcriptomic, cell-cycle and preclinical-model studies (14-19,36-43). AURKA, aurora kinase A; CCNB1, cyclin B1; EdU, 5-ethynyl-2'-deoxyuridine; GLS2, glutaminase 2; HCC, hepatocellular carcinoma; IF, immunofluorescence; PDX, patient-derived xenograft; PI, propidium iodide; RNA-seq, RNA sequencing; RRM2, ribonucleotide reductase regulatory subunit M2; RT-qPCR, reverse transcription-quantitative PCR; SLC22A1, solute carrier family 22 member 1; TOP2A, DNA topoisomerase IIα.

consistently identify TOP2A, CCNB1 and RRM2 as components of proliferation-associated programmes (11,12,14-16). In contrast, reduced albumin and transthyretin expression reflects diminished hepatocyte differentiation (14,15), whereas reduced SLC22A1/OCT1 and SLCO-family transporter expression has been documented in HCC (20-22). Under the proposed framework, this pattern is interpreted as a proliferation-identity imbalance rather than as evidence that any one highly ranked gene is automatically a driver or drug target.

The second step is to interrogate cellular origin. A TCGA-LIHC tumour-sample bulk signal averages malignant hepatocytes, immune cells, endothelial cells and fibroblasts, In tumour-versus-adjacent tissue analyses, adjacent diseased liver constitutes a separate comparator and may influence the observed differential signal. Therefore, the proliferation-identity pattern should be cross-checked against single-cell or spatial evidence. If proliferative genes are concentrated in malignant epithelial compartments and hepatocyte-identity transcripts are reduced in those same compartments, the bulk signal supports tumour-cell state remodeling. If the apparent signal is mainly stromal, immune or derived from adjacent cirrhotic tissue, the validation question changes and a conventional HCC cell-line experiment may not test the relevant biology (28-34).

The third step is state-matched model selection. If the TCGA-LIHC-like signal represents a high-proliferation and low-hepatocyte-identity tumour state, the first experimental model should preserve strong cycling activity and allow cell-cycle readouts. Such a model can be used to test EdU incorporation, PI-based cell-cycle distribution, colony formation and protein expression of CCNB1, CDK1, TOP2A, RRM2 or PCNA. By contrast, if the working hypothesis concerns transporter-dependent drug exposure or restoration of hepatocyte metabolism, the model should retain measurable transporter and metabolic competence, such as selected differentiated models, organoids or patient-derived systems (36-39,43). The model is therefore chosen because it matches the state being tested, not simply because it is available.

The fourth step is to apply the validation ladder. For TCGA-LIHC example, a first tier would assess phenotype, including growth or survival. A second tier would examine proliferation and cell-cycle readouts, such as EdU incorporation, PI-based cell-cycle distribution or colony formation assays. A third tier would confirm the coordinated mRNA pattern by RT-qPCR or RNA-seq using both proliferation markers and hepatocyte-identity markers. A fourth tier would examine protein or pathway readouts, such as CCNB1, CDK1, PCNA, TOP2A, RRM2, OCT1/SLC22A1 or selected CYP proteins when appropriate. Stronger mechanistic claims would require perturbation or rescue experiments, and translational claims would require extension across additional models or patient-derived systems (14-19,36-43). This example shows how a bulk TCGA-LIHC-type DEG list can be converted into a state-aware experimental plan rather than a direct therapeutic conclusion.

10. Pitfalls in interpreting HCC transcriptomic modules

Interpretation can drift for several reasons. Adjacent liver is not a neutral control in many HCC datasets; it may be cirrhotic, inflamed, steatotic, fibrotic or virally injured. Tumour purity also changes bulk immune, stromal, endothelial, complement and matrix signals. Platform differences add another layer because microarray and RNA-seq studies differ in probe design, normalization, dynamic range and annotation. Original studies can use meta-analysis, robust rank aggregation or direction-consistency filters to reduce noise, but none of these steps removes the underlying biology and sampling issues (7,11,12,14-19,28-34).

Hub genes create another trap. Genes that sit near the centre of protein-interaction or co-expression networks are

often essential, abundant or well studied. That does not make them uniquely disease-driving. Similarly, a high-risk prognostic signature may describe aggressive tumours without defining a core HCC mechanism or a general therapeutic vulnerability (14-19).

Validation can be too thin as well. One cell line, one dose and one or two genes may be enough for a first observation, but not for a coordinated mechanism. These limits do not weaken transcriptomics; they mark the point where interpretation should stop (36-39,43).

Expression patterns are most useful when readers can see the cohort context, platform constraints, cell-type origin and validation depth. Without that information, stable-looking signatures can easily harden into therapeutic stories that the data do not yet support (7,14-19,28-34,36-39,43).

11. Recommendations and future perspectives

Future work should join robust bulk signatures with cell-type resolution more deliberately. Multi-cohort analyses can show which patterns recur, whereas single-cell and spatial studies can show where those patterns arise. Stronger experimental designs will then choose a cell line, organoid, co-culture or xenograft because it retains the state under study, not because it is the nearest available model (14-19,28-34,36-39,43).

A first recommendation is to report transcriptomic claims in evidence tiers. A recurrent DEG module should be described as a state hypothesis. A coordinated RT-qPCR or RNA-seq response can support module-level transcriptional modulation. Protein, pathway and phenotype assays are needed before mechanistic language is used. Perturbation, rescue and model extension are needed before translational vulnerability claims are justified (14-19,36-43).

A second recommendation is to make model-selection rationale explicit. For proliferation-associated modules, the model should retain measurable cell-cycle activity. For hepatocyte-identity or transporter modules, the model should preserve relevant uptake, detoxification or metabolic features. For immune, complement or stromal modules, the model should include the cellular context required to test the signal. This recommendation is especially important because HCC cell lines, organoids and xenografts preserve different aspects of tumour biology (28-34,36-39,43).

For translation, background liver disease and drug handling need to be part of the same discussion. A proliferative module may point to cell-cycle or replication-stress testing, but transporter loss, CYP remodeling and weak hepatocyte identity can affect intracellular exposure and liver-specific toxicity. Future studies should therefore report not only whether tumour cells are inhibited, but also why the chosen model can support the pharmacological interpretation (20-27,36-39,43).

Future translational strategies should remain hypothesis-generating until they are tested in HCC-relevant models with state-matched readouts. Transcriptomic modules can prioritize questions about cell-cycle, replication-stress or metabolic dependencies, but they do not establish therapeutic efficacy by themselves (19,23-27,40-42).

Reporting habits also need to improve. Gene lists, preprocessing choices, ranking methods, cohort composition, model-selection rationale and validation status should be visible to readers. Negative validation results are worth reporting because they show where public transcriptomic predictions fail. The language should follow the evidence: a recurrent cell-cycle signal supports a proliferative-state hypothesis; it does not make every gene in the module a target. Loss of transporter or metabolic transcripts supports hepatocyte-identity erosion; it does not by itself prove drug resistance (7,14-19,28-34,36-39,43).

12. Conclusions

Across HCC transcriptomes, a recurring paired pattern is evident: Cell-cycle and DNA-replication programs often rise while differentiated hepatocyte transport, metabolic and defense functions decline. The proliferation-identity frame offers a practical way to read that pattern, select experimental models and plan validation without overstating what expression data can prove.

Expression data are most useful when they lead to testable biological questions. They are weakest when treated as proof of mechanism or therapeutic effect. A state-aware reading asks a limited set of questions before stronger claims are made: Which cell state is being represented, which model retains it, what phenotype is expected to change and what evidence would be sufficient for the claim.

Acknowledgements

Not applicable.

Funding

No funding was received.

Availability of data and materials

Not applicable.

Authors' contributions

JH and LL developed the review concept, including the proliferation-identity pairing as the central focus of the manuscript. JH performed the literature search and synthesis and wrote the first draft. BZ and LL reviewed and edited the manuscript. LL supervised the work. Data authentication is not applicable. All authors read and approved the final manuscript. All authors take responsibility for the accuracy of the literature synthesis and interpretation presented.

Ethics approval and consent to participate

Not applicable.

Patient consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Use of artificial intelligence tools

During the preparation of this work, AI tools were used to improve the readability and language of the manuscript, and subsequently, the authors revised and edited the content produced by the AI tools as necessary, taking full responsibility for the ultimate content of the present manuscript.

References

1. Bray F, Laversanne M, Sung H, Ferlay J, Siegel RL, Soerjomataram I and Jemal A: Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin 74: 229-263, 2024.
2. Llovet JM, Kelley RK, Villanueva A, Singal AG, Pikarsky E, Roayaie S, Lencioni R, Koike K, Zucman-Rossi J and Finn RS: Hepatocellular carcinoma. Nat Rev Dis Primers 7: 6, 2021.
3. Reig M, Forner A, Rimola J, Ferrer-Fàbrega J, Burrel M, Garcia-Criado Á, Kelley RK, Galle PR, Mazzaferro V, Salem R, *et al*: BCLC strategy for prognosis prediction and treatment recommendation: The 2022 update. J Hepatol 76: 681-693, 2022.
4. Singal AG, Llovet JM, Yarchoan M, Mehta N, Heimbach JK, Dawson LA, Jou JH, Kulik LM, Agopian VG, Marrero JA, *et al*: AASLD practice guidance on prevention, diagnosis, and treatment of hepatocellular carcinoma. Hepatology 78: 1922-1965, 2023.
5. European Association for the Study of the Liver: EASL clinical practice guidelines on the management of hepatocellular carcinoma. J Hepatol 82: 315-374, 2025.
6. Gordan JD, Kennedy EB, Abou-Alfa GK, Beal E, Finn RS, Gade TP, Goff L, Gupta S, Guy J, Hoang HT, *et al*: Systemic therapy for advanced hepatocellular carcinoma: ASCO guideline update. J Clin Oncol 42: 1830-1850, 2024.
7. Cancer Genome Atlas Research Network. Electronic address: wheeler@bcm.edu and Cancer Genome Atlas Research Network: Comprehensive and integrative genomic characterization of hepatocellular carcinoma. Cell 169: 1327-1341.e23, 2017.
8. Villanueva A: Hepatocellular carcinoma. N Engl J Med 380: 1450-1462, 2019.
9. Sia D, Jiao Y, Martinez-Quetglas I, Kuchuk O, Villacorta-Martin C, Castro de Moura M, Putra J, Camprecios G, Bassaganyas L, Akers N, *et al*: Identification of an immune-specific class of hepatocellular carcinoma, based on molecular features. Gastroenterology 153: 812-826, 2017.
10. Zucman-Rossi J, Villanueva A, Nault JC and Llovet JM: Genetic landscape and biomarkers of hepatocellular carcinoma. Gastroenterology 149: 1226-1239.e4, 2015.
11. Boyault S, Rickman DS, de Reyniès A, Balabaud C, Rebouissou S, Jeannot E, Hérault A, Saric J, Belghiti J, Franco D, *et al*: Transcriptome classification of hepatocellular carcinoma is related to gene alterations and to new therapeutic targets. Hepatology 45: 42-52, 2007.
12. Hoshida Y, Nijman SMB, Kobayashi M, Chan JA, Brunet JP, Chiang DY, Villanueva A, Newell P, Ikeda K, Hashimoto M, *et al*: Integrative transcriptome analysis reveals common molecular subclasses of human hepatocellular carcinoma. Cancer Res 69: 7385-7392, 2009.
13. Hamdy NM, Sallam AM, Elazazy O, Kabel AM, Salama RM, Gouhar SA, El-Daly SM and Darwish SF: LincRNA-miR interactions in hepatocellular carcinoma: Comprehensive review and in silico analysis: A step toward ncRNA precision. Naunyn Schmiedebergs Arch Pharmacol 398: 14785-14812, 2025.
14. Allain C, Angenard G, Clement B and Coulouarn C: Integrative genomic analysis identifies the core transcriptional hallmarks of human hepatocellular carcinoma. Cancer Res 76: 6374-6381, 2016.
15. Khorsand B, Naderi N, Karimian SS, Mohaghegh M, Aghaahmadi A, Hadisadegh SN, Owrang M and Houri H: Comprehensive transcriptomic analysis of hepatocellular carcinoma: Uncovering shared and unique molecular signatures across diverse etiologies. Biochem Biophys Rep 43: 102123, 2025.
16. Yang Z, Wu X, Li J, Zheng Q, Niu J and Li S: CCNB2, CDC20, AURKA, TOP2A, MELK, NCAPG, KIF20A, UBE2C, PRC1, and ASPM may be potential therapeutic targets for hepatocellular carcinoma using integrated bioinformatic analysis. Int J Gen Med 14: 10185-10194, 2021.
17. Qi F, Zha G, Zhang Y, Liu S, Yang Y, Sun W, Wang D, Liu Z, Lu Z and Zhang D: Integrative analysis of bulk and single-cell transcriptomic data reveals novel insights into lipid metabolism and prognostic factors in hepatocellular carcinoma. Discov Oncol 15: 591, 2024.
18. Moghimi A, Bani Hosseinian N, Mahdipour M, Ahmadpour E, Miranda-Bedate A and Ghorbian S: Deciphering the molecular complexity of hepatocellular carcinoma: Unveiling novel biomarkers and therapeutic targets through advanced bioinformatics analysis. Cancer Rep 7: e2152, 2024.
19. Bisteau X, Caldez MJ and Kaldis P: The complex relationship between liver cancer and the cell cycle: A story of multiple regulations. Cancers 6: 79-111, 2014.
20. Schaeffeler E, Hellerbrand C, Nies AT, Winter S, Kruck S, Hofmann U, van der Kuip H, Zanger UM, Koepsell H and Schwab M: DNA methylation is associated with downregulation of the organic cation transporter OCT1 (SLC22A1) in human hepatocellular carcinoma. Genome Med 3: 82, 2011.
21. Heise M, Lautem A, Knapstein J, Schattenberg JM, Hoppe-Lotichius M, Foltys D, Weiler N, Zimmermann A, Schad A, Gründemann D, *et al*: Downregulation of organic cation transporters OCT1 (SLC22A1) and OCT3 (SLC22A3) in human hepatocellular carcinoma and their prognostic significance. BMC Cancer 12: 109, 2012.
22. Herraez E, Lozano E, Macias RIR, Vaquero J, Bujanda L, Banales JM, Marin JJG and Briz O: Expression of SLC22A1 variants may affect the response of hepatocellular carcinoma and cholangiocarcinoma to sorafenib. Hepatology 58: 1065-1073, 2013.
23. Park S and Hall MN: Metabolic reprogramming in hepatocellular carcinoma: Mechanisms and therapeutic implications. Exp Mol Med 57: 515-523, 2025.
24. Dai P, Feng J, Dong Y, Zhang S, Cao J, Cui X, Qin X, Yang S and Fan D: Metabolic reprogramming in hepatocellular carcinoma: An integrated omics study of lipid pathways and their diagnostic potential. J Transl Med 23: 644, 2025.
25. Liu Q, Zhang X, Qi J, Tian X, Dovjak E, Zhang J, Du H, Zhang N, Zhao J, Zhang Y, *et al*: Comprehensive profiling of lipid metabolic reprogramming expands precision medicine for HCC. Hepatology 81: 1164-1180, 2025.
26. Hu N, Li H, Tao C, Xiao T and Rong W: The role of metabolic reprogramming in the tumor immune microenvironment: Mechanisms and opportunities for immunotherapy in hepatocellular carcinoma. Int J Mol Sci 25: 5584, 2024.
27. Tang D, Kroemer G and Kang R: Ferroptosis in hepatocellular carcinoma: From bench to bedside. Hepatology 80: 721-739, 2024.
28. Lu Y, Yang A, Quan C, Pan Y, Zhang H, Li Y, Gao C, Lu H, Wang X, Cao P, *et al*: A single-cell atlas of the multicellular ecosystem of primary and metastatic hepatocellular carcinoma. Nat Commun 13: 4594, 2022.
29. Sun Y, Wu L, Zhong Y, Zhou K, Hou Y, Wang Z, Zhang Z, Xie J, Wang C, Chen D, *et al*: Single-cell landscape of the ecosystem in early-relapse hepatocellular carcinoma. Cell 184: 404-421.e16, 2021.
30. Wu R, Guo W, Qiu X, Wang S, Sui C, Lian Q, Wu J, Shan Y, Yang Z, Yang S, *et al*: Comprehensive analysis of spatial architecture in primary liver cancer. Sci Adv 7: eabg3750, 2021.
31. Ye J, Lin Y, Liao Z, Gao X, Lu C, Lu L, Huang J, Huang X, Huang S, Yu H, *et al*: Single cell-spatial transcriptomics and bulk multi-omics analysis of heterogeneity and ecosystems in hepatocellular carcinoma. NPJ Precis Oncol 8: 262, 2024.
32. Pang W, Wang Y, Lu X, Li M, Long F, Chen S, Yu Y, Li M and Lin H: Integrated spatial and single cell transcriptomics identifies PRKDC as a dual prognostic biomarker and therapeutic target in hepatocellular carcinoma. Sci Rep 15: 14834, 2025.
33. Jia G, He P, Dai T, Goh D, Wang J, Sun M, Wee F, Li F, Lim JCT, Hao S, *et al*: Spatial immune scoring system predicts hepatocellular carcinoma recurrence. Nature 640: 1031-1041, 2025.
34. Chen P, Geng H, Ma B, Zhang Y, Zhu Z, Li M, Chen S, Wang X and Sun C: Integrating spatial omics and single-cell mass spectrometry imaging reveals tumor-host metabolic interplay in hepatocellular carcinoma. Proc Natl Acad Sci USA 122: e2505789122, 2025.
35. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, Shamseer L, Tetzlaff JM, Akl EA, Brennan SE, *et al*: The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. BMJ 372: n71, 2021.

36. Broutier L, Mastrogiovanni G, Verstegen MM, Francies HE, Gavarró LM, Bradshaw CR, Allen GE, Arnes-Benito R, Sidorova O, Gaspersz MP, *et al*: Human primary liver cancer-derived organoid cultures for disease modeling and drug screening. Nat Med 23: 1424-1435, 2017.
37. Qureshi AA, Wehrle CJ, Ferreira-Gonzalez S, Jiao C, Hong H, Dadgar N, Arpi-Palacios J, Phong YP, Kim J, Sun K, *et al*: Tumor organoids for primary liver cancers: A systematic review of current applications in diagnostics, disease modeling, and drug screening. JHEP Rep 6: 101164, 2024.
38. Airola C, Pallozzi M, Cesari E, Cerrito L, Stella L, Sette C, Giuliante F, Gasbarrini A and Ponziani FR: Hepatocellular-carcinoma-derived organoids: Innovation in cancer research. Cells 13: 1726, 2024.
39. Wong AM, Huang H, Wong AM, Leung HHW, Cheung WWY, Lam YK, Ding X, Zhang L, Chu C, Liu Y, *et al*: Patient-derived organoids inform pharmacogenomic vulnerabilities in liver cancer. JHEP Rep 7: 101426, 2025.
40. Bondarev AD, Jonsson J, Chubarev VN, Tarasov VV, Lagunas-Rangel FA and Schiöth HB: Recent developments of topoisomerase inhibitors: Clinical trials, emerging indications, novel molecules and global sales. Pharmacol Res 209: 107431, 2024.
41. Cai X, Yin G, Chen S, Tacke F, Guillot A and Liu H: CDK4/6 inhibition enhances T-cell immunotherapy on hepatocellular carcinoma cells by rejuvenating immunogenicity. Cancer Cell Int 24: 215, 2024.
42. Ibrahim L, Mohamed RH, Tolba MM, Radwan SM, Hamdy NM and Elhefnawi M: Unveiling the therapeutic potential of targeting RRM2 in hepatocellular carcinoma: An integrated in silico and in vitro study. Funct Integr Genomics 25: 123, 2025.
43. Cigliano A, Liao W, Deiana GA, Rizzo D, Chen X and Calvisi DF: Preclinical models of hepatocellular carcinoma: Current utility, limitations, and challenges. Biomedicines 12: 1624, 2024.
44. López-Terrada D, Cheung SW, Finegold MJ and Knowles BB: Hep G2 is a hepatoblastoma-derived cell line. Hum Pathol 40: 1512-1515, 2009.

Copyright © 2026 Hou et al. This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International (CC BY-NC-ND 4.0) License.