



Transcriptomic trajectory of hepatocyte ballooning reveals a lipotoxic point of no return and initiates oncogenic regulatory rewiring in MASLD

Debopriyo Ganguly^{a,c}, Bandana Mondal^{a,c}, Subhas Das^d, Indrashish Dey^d, Ayan Mandal^{a,c}, Trinath Ghosh^a, Monalisha Deo^a, Misba Ahmed^a, Soma Banerjee^d, Ankita Chatterjee^{b,*}, Priyadarshi Basu^{a,c,**}

^a Biotechnology Research Innovation Council – National Institute of Biomedical Genomics (BRIC-NIBMG), Kalyani, West Bengal, 741251, India

^b John C. Martin Center for Liver Research and Innovations, Kolkata, West Bengal, India

^c Regional Centre for Biotechnology, Government of India, Faridabad, Haryana, India

^d Centre for Liver Research, School of Digestive and Liver Diseases, Institute of Post Graduate Medical Education and Research (IPGMER), A.J.C. Bose Road, Kolkata, 700020, India

ARTICLE INFO

Keywords:

Hepatocyte ballooning
MASLD
MASH
WGCNA
Machine learning
Single-nucleus RNA sequencing
Pseudotime

ABSTRACT

Background: Hepatocyte ballooning is a defining histological marker of progressive steatohepatitis, but its molecular basis remains poorly defined.

Objective: To identify a molecular signature and mechanistic trajectory underlying hepatocyte ballooning across the MASLD-to-hepatocellular carcinoma spectrum.

Methods: Liver microarray data from 43 patients stratified by histological ballooning score were combined with machine-learning feature selection and publicly available single-nucleus RNA sequencing data (87,244 hepatocyte nuclei) to derive a gene signature, tested in an independent bulk transcriptomic cohort (GSE130970, n = 75); pseudotime, transcription-factor regulon and cell-cell communication analyses characterised the underlying biology, with in vitro validation in palmitic acid-treated Huh7 cells and prognostic testing in 510 TCGA-LIHC patients.

Results: An eight-gene signature (*SPP1*, *MFSD2A*, *PRAP1*, *ANXA2*, *KRT8*, *MACROH2A2* (*H2AFY2*), *JUN* and *FABP4*) characterised ballooned hepatocytes and discriminated moderate-to-severe from absent ballooning with an AUC of 0.768 (95% CI 0.654–0.883). Six of these genes formed a Hepatocyte Ballooning–Hepatocyte injury (HBHi) score that was highest in nuclei from cirrhotic (F4) liver. Pseudotime analysis resolved a four-stage molecular collapse culminating at pseudotime ≈ 25, marked by loss of the chromatin stabiliser *MACROH2A2*, co-induction of *KRT8* and *SPP1*, and replacement of the *HNF4A*/*PPARA*-driven programme by the stress factors *JUN*, *ATF3* and *EGR1*. Ballooned hepatocytes transmitted pro-fibrogenic signals to stellate cells and macrophages via *FN1*–*CD44* and *FN1*–(*ITGAV* + *ITGB1*) axes; high *KRT8*, *JUN* and *SPP1* expression was each associated with poorer survival in hepatocellular carcinoma.

Conclusions: Hepatocyte ballooning is a transcriptionally active, stepwise process with a lipotoxic point of no return and prognostic relevance extending to hepatocellular carcinoma.

1. Introduction

Metabolic dysfunction–associated steatotic liver disease (MASLD) is the most common chronic liver disorder worldwide. A meta-analysis of 92 population-based studies encompassing more than 9.3 million individuals estimated global prevalence at 30.05% overall, rising from

25.26% in 1990–2006 to 38.00% in 2016–2019 [1], and the burden is projected to continue increasing [2]. A subset of patients with simple steatosis progresses to metabolic dysfunction–associated steatohepatitis (MASH), which may in turn advance to cirrhosis and hepatocellular carcinoma (HCC) [2,3]. In this paper the term MASH is used interchangeably with the earlier designation NASH where required for

* Corresponding author.

** Corresponding author. BRIC-NIBMG, Kalyani, 741251, West Bengal, India.

E-mail addresses: ankita.chatterjee88@gmail.com (A. Chatterjee), pb1@nibmg.ac.in (P. Basu).

<https://doi.org/10.1016/j.bbrep.2026.102785>

Received 25 June 2026; Received in revised form 4 August 2026; Accepted 7 September 2026

Available online 12 September 2026

2405-5808/© 2026 Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

consistency with cited datasets and literature.

Hepatocyte ballooning (HB) is a cardinal feature of MASH and refers to a form of hepatocyte degeneration characterised by cell enlargement, swelling, rounding and a distinctive reticulated cytoplasm [3]. Ballooned cells are typically larger and rounder than normal hepatocytes and display a pale, reticulated cytoplasm that gives them a “spider web” appearance on haematoxylin and eosin staining [4]. Histopathological studies have shown co-localisation of steatotic hepatocytes and areas of perisinusoidal fibrosis with ballooned hepatocytes, implicating HB in MASH progression [4,5]. The presence of hepatocyte ballooning has been linked to greater severity of liver disease and a higher likelihood of hepatic, cardiovascular and chronic kidney complications [4], and a higher frequency of liver-related death has been observed in patients whose biopsies show ballooning, Mallory–Denk bodies or fibrosis [6].

Because hepatocyte ballooning correlates with inflammatory activity, fibrosis progression and adverse clinical outcomes, it represents a key pathological transition within the MASLD–MASH continuum [7]. Both of the principal histological scoring systems grade ballooning semi-quantitatively: the NASH Clinical Research Network NAFLD Activity Score (NAS) scores hepatocellular ballooning on a 0–2 scale [8], and the SAF (Steatosis–Activity–Fibrosis) system incorporates ballooning into a composite activity score [9]. Despite this diagnostic importance, hepatocyte ballooning remains defined principally by morphological criteria, and the molecular mechanisms linking the phenotype to malignant progression are incompletely understood [4,6,7,10].

Here we address this gap through an integrative analysis combining bulk transcriptomics, machine-learning feature selection and single-nucleus RNA sequencing to reconstruct the molecular trajectory of hepatocyte injury across the healthy–MASLD–MASH–HCC spectrum. We identify transcriptional programmes associated with hepatocyte ballooning, resolve a continuous hepatocyte disease axis, and characterise hepatocyte ballooning as a transcriptionally governed state that links metabolic liver injury to hepatocellular carcinoma.

2. Materials and methods

2.1. Microarray data processing and weighted gene co-expression network analysis

Hepatic microarray data generated previously by our group [11] were re-analysed. Following quality filtering, 43 samples were retained and the 10,000 most variable probe sets were carried forward. A co-expression network was constructed with WGCNA [12] using a topological overlap matrix-based adjacency computed at a soft-thresholding power of five, and modules were detected by dynamic tree cut. Closely related modules were merged at a dissimilarity threshold of 0.25. Module–trait associations were calculated as Pearson correlations between module eigengenes and clinical variables.

2.2. Differential expression analysis

Differential expression was assessed with limma using linear modelling and empirical Bayes moderation [13]. The Benjamini–Hochberg procedure was applied to control the false discovery rate. Differentially expressed genes (DEGs) were defined as $|\log_2FC| > 0.5$ and $FDR < 0.05$. The comparison was made between patients with a hepatocyte ballooning score (HBS) of 1 or 2 ($n = 19$) and those with an HBS of 0 ($n = 24$).

2.3. Feature selection and stability analysis

LASSO logistic regression was performed with the glmnet package, and the penalty parameter was selected by 10-fold cross-validation [14]. To assess the stability of gene selection, 10,000 bootstrap iterations were performed and genes with a bootstrap selection frequency above the

50th percentile were retained. *HSPC157* and *C10orf116* (*ADIRF*) were excluded despite high selection frequencies because neither has a documented biological function in hepatocyte injury.

2.4. Independent validation and clinical translation

Bulk transcriptomic data from an independent cohort (GSE130970, $n = 75$) were normalised [15]. Samples were scored using a ridge-penalised logistic regression model built exclusively on the final eight genes [14]. GSE130970 was not used at any stage of feature selection or model fitting. The area under the receiver-operating-characteristic curve (AUC) was estimated with bootstrapping to generate a 95% confidence interval. A nomogram was constructed with the rms package [16]. Clinical net benefit was evaluated by decision curve analysis using the rmda package [17]. Agreement between predicted probabilities and observed outcomes was assessed with $B = 200$ bootstrap calibration repetitions.

2.5. Single-nucleus RNA sequencing processing and integration

Two publicly available single-nucleus RNA sequencing (snRNA-seq) datasets, GSE189175 [18–20] and GSE202379 [21], were processed with Seurat v5 [22]. Nuclei with fewer than 200 or more than 6000 detected genes, or with more than 25% mitochondrial reads, were removed. Computational doublets were identified and removed with scDblFinder [23]. Datasets were integrated by reciprocal principal component analysis (RPCA) and clustered using a shared nearest-neighbour graph. The hepatocyte population was subclustered at a Leiden resolution of 0.4, yielding 11 subpopulations.

2.6. Selection of six genes for single-nucleus analysis

The signature was narrowed to six genes (*JUN*, *ANXA2*, *KRT8*, *SPPI*, *MFSD2A* and *MACROH2A2*) for single-nucleus analysis on two criteria: detection in more than 10% of hepatocyte nuclei across both datasets (Supplementary Fig. S2B) and published evidence of hepatocyte-specific function [24–29]. *FABP4* was excluded because it is expressed predominantly in macrophages and adipocytes, giving subthreshold nuclear detection in hepatocytes [30]. *PRAP1* was excluded because of sparse nuclear capture, consistent with its predominantly cytoplasmic localisation.

2.7. Trajectory inference and the HBHi score

Pseudotime trajectory inference was performed on the hepatocyte subset with Monocle 3 [31]. The trajectory root was anchored to the centroid of the healthy hepatocyte cluster. Gene expression kinetics along pseudotime were smoothed with generalized additive models. The Hepatocyte Ballooning–Hepatocyte injury (HBHi) score was defined as the first principal component (PC1) of the expression matrix of the six signature genes computed exclusively within snRNA-seq hepatocyte nuclei. The relationship between HBHi score and pseudotime was assessed by Spearman rank correlation.

2.8. Metabolic pathway activity analysis

KEGG metabolic pathway activity was scored for each nucleus with the scMetabolism package [32] using VISION-based enrichment. Metabolic scores were mapped onto the pseudotime axis and trends visualised with GAM smoothing.

2.9. Gene regulatory network analysis

Regulatory networks were inferred with pySCENIC [33,34]. Co-expression modules were inferred with GRNBoost2 and pruned against the RcisTarget database using the hg38 500 bp and 10 kb motif

rankings. Regulon activity was scored per nucleus with AUCell, and GRNBoost2 edge weights were used to represent regulatory strength. The relationship between regulon activity and the HBHI score was assessed by Pearson correlation.

2.10. Cell–cell communication analysis

CellChat [35] was applied to the integrated snRNA-seq dataset using the human CellChat ligand–receptor database. Communication probabilities were computed for all cell-type pairs across all disease stages, and the sender, receiver, mediator and influencer roles of each cell type were assessed. Significant ligand–receptor pairs were identified by permutation testing ($n = 100$) at $p < 0.05$.

2.11. TCGA-LIHC survival analysis

RSEM-normalised expression and matched clinical survival data were obtained for 510 patients with HCC from the TCGA-LIHC cohort [36] via the UCSC Xena browser (<https://xenabrowser.net/datapages/?cohort=GDC+TCGA+Liver+Cancer+%28LIHC%29>). For each signature gene, patients were dichotomised at the median expression value. Kaplan–Meier curves were plotted for overall and disease-free survival, groups were compared by log-rank test, and hazard ratios (HR) with 95% confidence intervals were obtained by Cox proportional hazards regression. All hazard ratios reported are univariable. Analyses were performed with the survival and survminer packages.

2.12. Cell culture, treatment, RNA extraction and RT-qPCR

Huh7 hepatocellular carcinoma cells (JCRB Cell Bank, JCRB0403; RRID: CVCL_0336) were cultured in DMEM (Thermo Fisher Scientific, Cat# 11965092) supplemented with 10% FBS (Gibco, Thermo Fisher Scientific, Cat# A5670501) and 1% penicillin–streptomycin (Gibco, Thermo Fisher Scientific, Cat# 15140122) at 37 °C in 5% CO₂. Cells (5×10^4 per well) were seeded in 12-well plates and treated with 150 μ M palmitic acid (PA; Sigma-Aldrich, Cat# P0500) conjugated to fatty-acid-free bovine serum albumin (Sigma-Aldrich, Cat# A2153) for 24 h or 48 h. Cells treated with the vehicle alone (0.1% DMSO; Sigma-Aldrich, Cat# D2650) served as controls. Total RNA was extracted with TRIzol Reagent (Invitrogen, Thermo Fisher Scientific, Cat# 15596026) [37] and reverse transcribed with the High-Capacity cDNA Reverse Transcription Kit (Applied Biosystems, Thermo Fisher Scientific, Cat# 4368814) according to the manufacturer's instructions. RT-qPCR was performed with SYBR Green Master Mix (Applied Biosystems, Thermo Fisher Scientific, Cat# 4309155) using 18S rRNA as the reference gene, and relative expression was calculated by the $2^{-\Delta\Delta CT}$ method [38].

2.13. Protein extraction and western blotting

Huh7 cells (1×10^5 per well) were seeded in 6-well plates, treated with 150 μ M PA for 24 h or 48 h as described above, and lysed in 150 μ L ice-cold RIPA buffer (Sigma-Aldrich, Cat# R0278) supplemented with protease inhibitor cocktail (Roche, Cat# 11836153001) and phosphatase inhibitor cocktail (Sigma-Aldrich, Cat# P5726). Protein concentration was determined by the Bradford assay (Bio-Rad Laboratories, Cat# 5000006) [39]. Equal amounts of protein were resolved by SDS-PAGE, transferred to 0.45 μ m PVDF membranes (Merck Millipore, Cat# IPVH00010) and blocked in 5% BSA in TBS containing 0.1% Tween-20 (Sigma-Aldrich, Cat# P1379). Membranes were incubated overnight at 4 °C with primary antibodies against osteopontin/SPP1 (Cat# A1499), cytokeratin 8/KRT8 (Cat# A1024) and β -actin (Cat# AC004; loading control), all from Abclonal Technology, followed by HRP-conjugated goat anti-rabbit IgG (H + L) (Abclonal Technology, Cat# AS014) or HRP-conjugated goat anti-mouse IgG (H + L) (Abclonal Technology, Cat# AS003) as appropriate. Blots were developed with an

ECL western blotting substrate (Thermo Fisher Scientific, Cat# 32106) on a chemiluminescence documentation system. Full details of all reagents, antibodies, cell lines and software are given in the Key Resources Table (Supplementary Table S1).

2.14. Quantification and statistical analysis

All computational and statistical analyses were performed in R (v4.3.1) and Python (v3.10). Key software versions were: WGCNA v1.73, limma v3.58.1, glmnet v4.1.10, rms v6.7.1, rmda v1.6, Seurat v5.4.0, scDblFinder v1.16.0, Monocle v3-1.4.26, scMetabolism v0.2.1, pySCENIC, AUCell v1.24.0, RcisTarget v1.20.0, CellChat v1.6.1, survival v3.8.3 and survminer. Variables were compared across ballooning categories with the Kruskal–Wallis test with Dunn's post-hoc correction and Benjamini–Hochberg adjustment. Continuous clinical trait associations were tested by one-way ANOVA. Spearman rank correlation was used for trajectory data, and Pearson correlation for module–trait and regulon associations. RT-qPCR data are presented as mean $\pm$ SD of three independent experiments and compared by unpaired two-tailed Student's t-test. CellChat interactions were considered significant on permutation testing at $p < 0.05$. TCGA survival statistics were obtained by log-rank test and Cox regression. Significance levels are denoted * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, **** $p < 0.0001$. Exact sample sizes, tests and p values are given in the figure legends.

3. Results

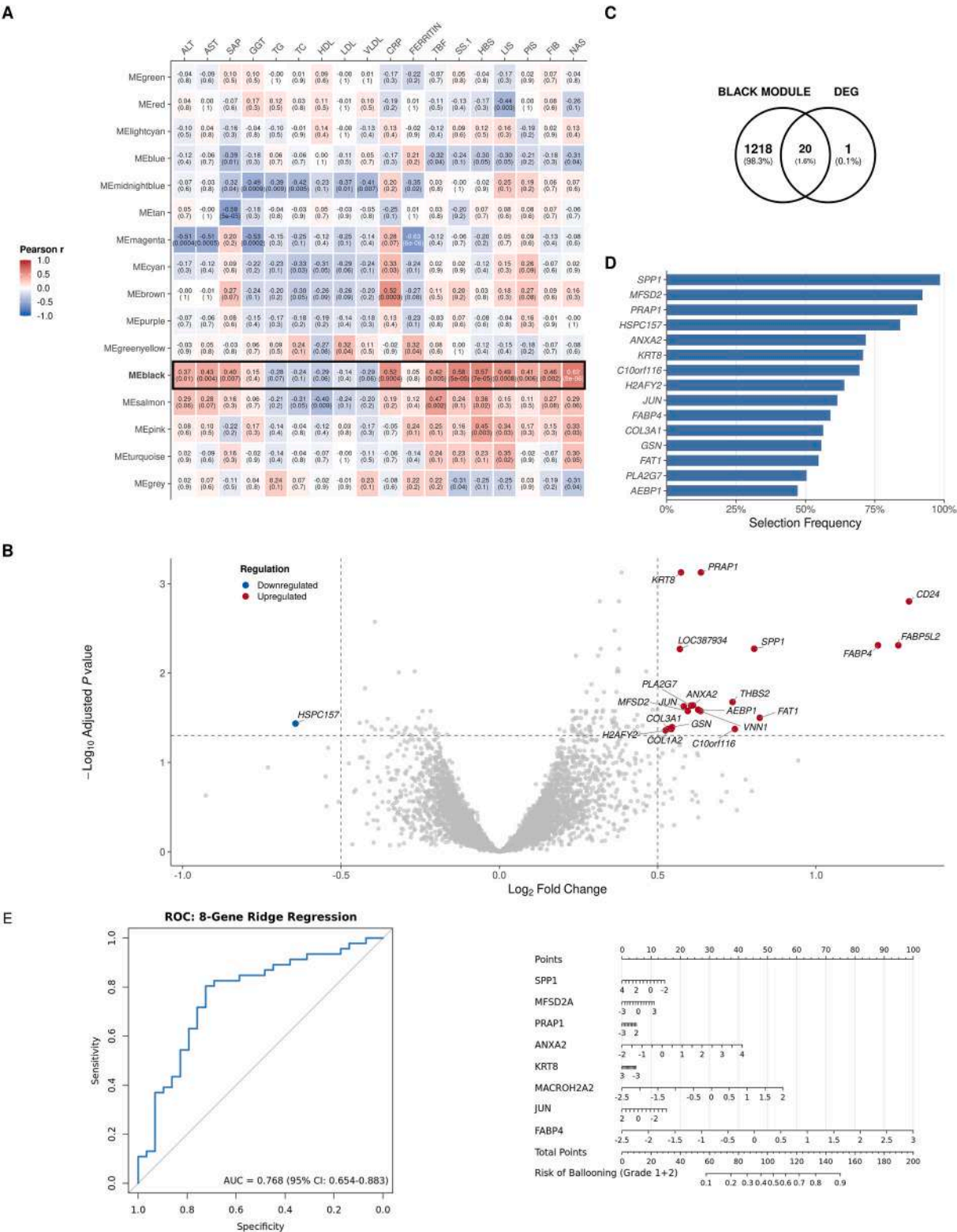
3.1. Identification of a hepatocyte ballooning gene signature from the liver tissue transcriptome

Ballooned hepatocytes are widely recognised as the histological feature that most specifically distinguishes benign steatosis from progressive steatohepatitis and fibrosis. Ballooning has, however, historically been interpreted as a morphological observation rather than as a transcriptionally altered cellular state. We examined liver microarray data from 43 patients categorised by histological hepatocyte ballooning score (HBS) from our previously published cohort [11] (Supplementary Fig. S1A, Supplementary Table S2). Patients were categorised as HBS 0 (no ballooning; $n = 24$) or HBS 1 + 2 (moderate and severe; $n = 19$). Patients without ballooning were largely free of fibrosis, the great majority being at fibrosis stage 0 and the lowest levels of portal and lobular inflammation (Supplementary Fig. S1B–C). Alanine aminotransferase and aspartate aminotransferase were significantly elevated in the moderate (adjusted $p = 2.6 \times 10^{-5}$) and severe (adjusted $p = 1.8 \times 10^{-7}$) groups relative to HBS 0 (Supplementary Fig. S1D).

WGCNA [12] applied to the normalised expression matrix identified the MEblack module (1238 genes), whose eigengene was more strongly correlated with the hepatocyte ballooning score than that of any other module ($r = 0.57$). MEblack was also correlated with the NAFLD activity score ($r = 0.62$), steatosis grade ($r = 0.58$), lobular inflammation ($r = 0.49$), fibrosis stage ($r = 0.46$), portal inflammation ($r = 0.41$), AST ($r = 0.43$) and ALT ($r = 0.37$; Fig. 1A; Supplementary Table S3). Differential expression analysis with limma [13] comparing HBS ≥ 1 with HBS 0 identified 21 significant DEGs at $|\log_2 FC| > 0.5$ and FDR < 0.05 (Fig. 1B; Supplementary Table S5). Intersecting the 21 DEGs with the MEblack module yielded 20 high-confidence candidate genes (Fig. 1C; Supplementary Table S4–S6).

LASSO regression [14] with 10,000 bootstrap iterations was applied to the 20 candidates and selected *SPP1*, *MFS2A*, *PRAP1*, *ANXA2*, *KRT8*, *MACROH2A2* (*H2AFY2*), *JUN* and *FABP4* (Fig. 1D). *HSPC157* (a pseudogene) and *C10orf116* (*ADIRF*) also scored highly but were excluded because neither has an established role in hepatocyte injury. Among the retained genes, *SPP1* (osteopontin) drives pro-fibrogenic signalling from injured hepatocytes to stellate cells [24], and *KRT8* protein aggregates form the Mallory–Denk bodies that characterise ballooned cells [25].

The panel was evaluated in an independent bulk transcriptomic



cohort (GSE130970, $n = 75$) [15] using ridge regression. The model achieved an AUC of 0.768 (95% CI 0.654–0.883). Model weights were used to construct a nomogram [16] converting the eight-gene profile into a risk score for grade 1 + 2 ballooning (Fig. 1E). The three genes contributing most to the score were *MACROH2A2*, *FABP4* and *ANXA2*. Decision curve analysis [17] indicated superior net clinical benefit relative to treat-all and treat-none strategies across thresholds from 0 to 0.92, with net benefits of 0.517, 0.373 and 0.133 at the 20%, 50% and 80% thresholds respectively. Bootstrap calibration ($B = 200$) gave a low mean absolute error of 0.046 (Supplementary Fig. S1E).

3.2. Pseudotime trajectory analysis resolves a four-stage molecular collapse

To determine the cellular origin of this transcriptional programme, the snRNA-seq datasets GSE202379 [21] and GSE189175 [18–20], spanning the healthy–MASLD–MASH–HCC spectrum, were integrated. Six genes (*JUN*, *ANXA2*, *KRT8*, *SPP1*, *MFSD2A* and *MACROH2A2*) were retained because they were detected in more than 10% of hepatocyte nuclei in both datasets and each has supporting literature for a role in hepatocyte injury and MASH. These six genes were used for all subsequent single-cell analyses (trajectory scoring, HBHi score and regulatory network analysis). *FABP4* was excluded because its expression is predominantly in liver macrophages and adipocytes rather than hepatocytes [30] and nuclear detection was low and inconsistent. *PRAP1* was excluded because nuclear transcript counts were very low across all hepatocyte clusters in both datasets, consistent with its primarily cytoplasmic localisation (Supplementary Tables S7–S8).

After integration, five major liver cell types were recovered: hepatocytes, stellate cells, endothelial cells, cholangiocytes and immune cells (Supplementary Fig. S2A). The hepatocyte subset was isolated and re-clustered, yielding 11 subpopulations (Supplementary Fig. S2B), and all six signature genes were confirmed to be detected within these clusters by dot plot.

Trajectory inference with Monocle 3 [31] ordered hepatocytes along a pseudotime axis spanning healthy to HCC (Fig. 2A). The defined disease stages fell along this axis in the expected chronological order (Fig. 2A, lower panels). The HBHi score correlated with pseudotime (Spearman's $\rho = 0.40$, $p < 2.2 \times 10^{-16}$; Fig. 2D; Supplementary Fig. S3A). When nuclei were grouped by fibrosis stage, the HBHi score was highest in cirrhotic (F4) liver, while stages F0–F3 were broadly comparable (Supplementary Fig. S3B).

A progressive decline in *MFSD2A* expression was observed. Because *Mfsd2a* marks periportal hepatocytes that repopulate the liver during injury and regeneration [26], its loss indicates a breakdown of hepatic spatial zonation (Fig. 2E). A transient increase in *JUN* and *ANXA2* was noted, consistent with activation of AP-1 stress signalling [27] and a membrane injury response [28]. The chromatin stabiliser *MACROH2A2* [29] declined sharply at pseudotime ≈ 25 , at which point the structural (*KRT8* [25]) and pro-fibrogenic (*SPP1* [24]) programmes were simultaneously activated. Late-trajectory hepatocytes occupied a state characterised by loss of normal hepatocyte identity and structural collapse relative to healthy hepatocytes.

3.3. Identification of a lipotoxic tipping point along the MASLD spectrum

To characterise the functional transition accompanying hepatocyte injury and ballooning, metabolic activity was analysed along the disease trajectory. scMetabolism [32] was used to score metabolic pathway activity in each nucleus. Early in pseudotime, hepatocytes retained normal functionality, including macronutrient and xenobiotic metabolism (Fig. 3A). As disease progressed along the pseudotime axis, these pathways declined in a coordinated manner. Fatty acid degradation capacity fell sharply around the 30th percentile of the trajectory, the point at which the rising HBHi injury burden overtook it (Fig. 3B). Reactive oxygen species and TNF/NF- κ B signalling were the earliest

injury pathways to rise, whereas apoptosis and unfolded protein response activity increased later and peaked in the final third of the trajectory (Fig. 3C).

3.4. Signature genes are acutely induced by palmitic acid-driven lipotoxicity in vitro

Having identified a lipotoxic tipping point along the pseudotime trajectory, we asked whether this transcriptional response could be induced experimentally by acute lipotoxic stress alone. Huh7 cells [40] were treated with 150 μ M palmitic acid (PA), a saturated fatty acid that induces endoplasmic reticulum stress, mitochondrial dysfunction, reactive oxygen species production and apoptosis in hepatocytes [41, 42]. RT-qPCR revealed a time-dependent induction of *JUN*, *KRT8* and *SPP1* transcripts in PA-treated cells relative to vehicle controls. *JUN* and *KRT8* were already significantly upregulated at 24 h (1.6-fold, $p < 0.01$, and 2.3-fold, $p < 0.01$, respectively) and rose further by 48 h (3.5-fold and 5.1-fold, both $p < 0.001$). *SPP1* was unchanged at 24 h and significantly induced only at 48 h (2.4-fold, $p < 0.01$; Fig. 3D). At the protein level, Western blot analysis confirmed increased expression of osteopontin (*SPP1*) and *KRT8* in PA-treated cells (Fig. 3D). Together these findings indicate that acute lipotoxic stress is sufficient to activate the hepatocyte ballooning gene programme in vitro.

3.5. Stress transcription factors replace the metabolic regulatory programme

The HBHi trajectory identified a metabolic-to-stress transition at pseudotime ≈ 25 , but the transcription factors orchestrating this switch were unknown. pySCENIC [33,34] was therefore used to map transcription factor regulon activity across the snRNA-seq hepatocyte subset and to correlate it with the HBHi score.

Two opposing groups of transcription factors emerged (Fig. 4A). Factors whose activity increased with HBHi — led by *EGR1* ($r = 0.56$), *JUN* ($r = 0.52$) and *ATF3* ($r = 0.45$) — constituted a stress-driven arm. Factors whose activity decreased — *HNF4A* ($r = -0.40$), *PPARA* ($r = -0.38$) and *NR1H4* ($r = -0.33$) — constituted a collapsing protective arm. Along the pseudotime axis, *CEBPB*, *FOS*, *ATF3* and *JUN* rose progressively while *HNF4A*, *NR1H3* and *RXR α* declined, confirming a coordinated regulatory switch (Fig. 4B). Notably, *EGR1* transcript counts were sparse across disease stages, yet its regulon activity rose sharply in HCC, illustrating the advantage of regulon-based inference over raw transcript counts (Supplementary Fig. S3C).

We next asked whether these HBHi-correlated transcription factors directly regulate the six signature genes. Regulatory weights were mapped between correlated factors and their downstream targets (Fig. 4C). On the stress side, *ATF3* directly targeted *JUN* with the strongest regulatory weight in the network (weight = 5.2), and *EGR1* also directly regulated *JUN* (weight = 2.7), indicating that stress factors amplify one another (Fig. 4C, D, 4G). On the protective side, *SOX5* sustained *MACROH2A2* (weight = 3.2) and *NR1H4* sustained *MFSD2A* (weight = 2.5), both confirmed at single-cell level (Fig. 4C, E, 4F). As the HBHi score increased, both protective axes progressively lost activity, withdrawing two key brakes on healthy hepatocyte identity.

Collectively these findings describe a two-arm transcriptional collapse: a stress arm (*ATF3/EGR1* $\rightarrow$ *JUN*) that drives injury forward, and a protective arm (*NR1H4* $\rightarrow$ *MFSD2A*, *SOX5* $\rightarrow$ *MACROH2A2*) that progressively loses its hold on healthy hepatocyte identity along the HBHi trajectory.

3.6. Ballooned hepatocytes send pro-fibrogenic signals to surrounding cells

CellChat [35] was applied to model ligand–receptor communication across all cell types at every disease stage. Cell type proportions shifted significantly with disease progression, with the HB_{high} hepatocyte

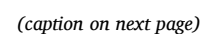


Fig. 2. Single-nucleus trajectory and temporal gene kinetics across the MASLD-to-HCC spectrum.

A. Global pseudotime trajectory and disease-stage distribution. Top: UMAP projection of the inferred hepatocyte lineage from Monocle 3 (left) and the distribution of Seurat clusters along the trajectory (right); 87,244 hepatocyte nuclei from GSE189175 and GSE202379. Bottom: split UMAP plots showing progression of cell states from Healthy through NAFLD, NASH and Advanced NASH to HCC.

B. Feature plots. Expression of *MFSD2A*, *MACROH2A2* (*H2AFY2*), *ANXA2*, *JUN*, *KRT8* and *SPP1* across the pseudotime manifold.

C. HBHi signature score. UMAP visualization of the HBHi signature score across the hepatocyte manifold.

D. HBHi score versus pseudotime. Scatter plot of HBHi score against pseudotime, coloured by histological category, with a smoothed fit (Spearman's $\rho = 0.40$, $p < 2.2 \times 10^{-16}$).

E. Temporal gene kinetics. Spline plots (left) of scaled expression of signature genes over pseudotime, smoothed by generalized additive model, and a schematic summary (right) of the four-stage transition culminating in the Catastrophic Nexus at pseudotime ≈ 25 .

subpopulation expanding progressively and most markedly in advanced NASH and HCC (Supplementary Fig. S4). Key ligand–receptor interactions from HB_{high} hepatocytes to stellate cells and macrophages included FN1–CD44 and FN1–(ITGAV + ITGB1), indicating active extracellular matrix remodelling, and LAMC1–(ITGA1+ITGB1) and LAMC1–(ITGA9+ITGB1), indicating basement membrane reprogramming (Supplementary Fig. S5).

3.7. Ballooning-associated genes are prognostic in hepatocellular carcinoma

We next asked whether the genes activated at the Catastrophic Nexus carry prognostic weight in human liver cancer. Expression and survival data from 510 patients with HCC in the TCGA-LIHC cohort [36] were analysed. Expression levels were dichotomised at the median for each gene and outcomes evaluated by Kaplan–Meier curves and Cox proportional hazards regression (Supplementary Fig. S6). High *KRT8* expression was associated with worse overall survival (log-rank $p = 0.038$; HR = 1.5, $p = 0.04$). High *JUN* expression similarly predicted worse overall survival (log-rank $p = 0.02$; HR = 1.5, $p = 0.021$). High *SPP1* expression carried the strongest prognostic signal of all genes tested (log-rank $p = 1.6 \times 10^{-5}$; HR = 2.2, $p = 2.4 \times 10^{-5}$). Conversely, high *MFSD2A* expression — the gene lost earliest along the trajectory — trended toward better disease-free survival (log-rank $p = 0.1$; HR = 0.75), consistent with its role as a marker of healthy hepatocyte zonation. All hazard ratios reported are univariable.

4. Discussion

This study investigated the molecular basis of hepatocyte ballooning in MASLD/MASH and placed it within the disease continuum from healthy liver to hepatocellular carcinoma. Using an integrative approach combining bulk transcriptomics, machine learning, single-nucleus RNA sequencing and cell–cell communication analysis, our findings define hepatocyte ballooning as a transcriptionally active and dynamic cellular state rather than a passive morphological observation.

An eight-gene signature (*SPP1*, *MFSD2A*, *PRAP1*, *ANXA2*, *KRT8*, *MACROH2A2*, *JUN* and *FABP4*) was identified from microarray data and evaluated in an independent cohort (AUC = 0.768). This level of discrimination is clinically informative though not definitive, and it illustrates the value of combining multiple markers rather than relying on a single gene. Each gene has a supported biological role. *SPP1* (osteopontin) is induced by hedgehog pathway activation and promotes fibrosis progression in steatohepatitis by activating hepatic stellate cells [24]. *KRT8* aggregates form Mallory–Denk bodies, a structural hallmark of ballooned hepatocytes [25]. *MACROH2A2*, a chromatin-stabilising histone variant, is a new addition in this context; macroH2A variants regulate epigenetic plasticity and the silencing of cancer-promoting expression programmes [29], and its loss may therefore permit de-repression of stress and oncogenic genes, although this remains a hypothesis that our data do not directly test. AP-1 dimers containing c-Jun regulate steatohepatitis and PPAR γ signalling in dietary models of fatty liver disease [27]. The three genes contributing most to the nomogram score were *MACROH2A2*, *FABP4* and *ANXA2*, pointing to chromatin instability, lipid handling and membrane stress as central

features of ballooned hepatocyte identity. The HBHi score, derived from six of these genes, was highest in cirrhotic (F4) liver, indicating that the ballooning transcriptional state is most pronounced once structural liver damage is advanced.

Pseudotime trajectory analysis placed hepatocytes along a continuous axis from healthy to HCC and recovered the expected chronological ordering of disease stages. Four sequential events were identified. First, in early disease, hepatocytes maintained normal metabolism and *MFSD2A* was highly expressed, marking periportal identity [26]; loss of *MFSD2A* was therefore an early indicator that hepatic spatial organisation had begun to break down. Second, *JUN* and *ANXA2* were transiently activated, reflecting AP-1-mediated stress responses [27] and membrane injury signalling [28]. Third, at approximately pseudotime 25, three events occurred concurrently: *MACROH2A2* expression fell sharply, *KRT8* was induced, and *SPP1* was activated. This coincidence of chromatin de-repression, cytoskeletal injury and fibrogenic signalling defines what we term the Catastrophic Nexus, marking the transition from a recoverable injury state to structural and pro-fibrogenic collapse. Fourth, late-stage hepatocytes entered a state characterised by loss of normal hepatocyte identity. This ordered sequence offers a more granular and mechanistic account of MASH progression than histological staging alone, and may provide a framework for defining intermediate disease states not captured by current scoring systems.

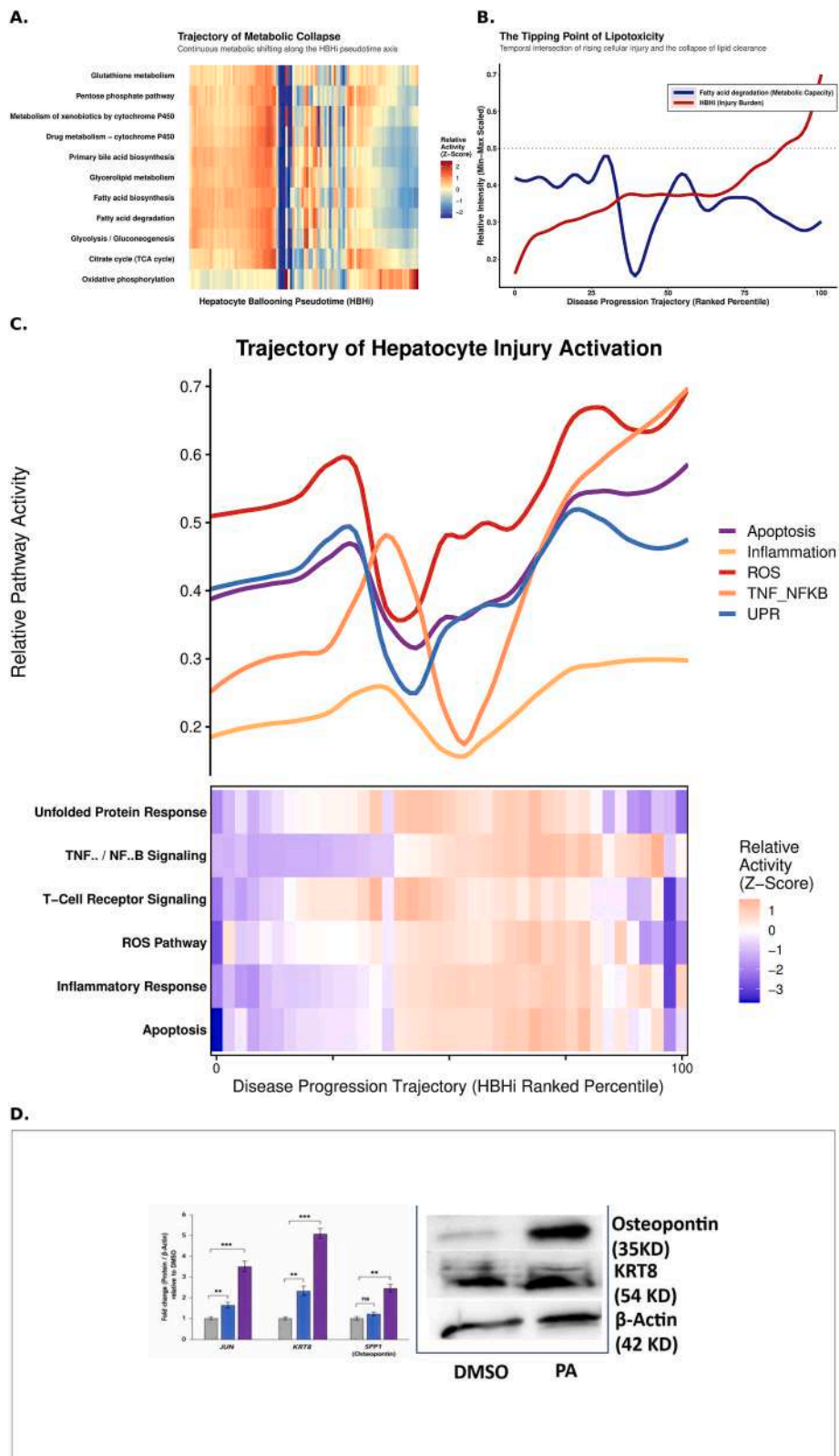
The regulatory network analysis identified two opposing groups of transcription factors. The protective arm — HNF4A, PPARA and NR1H4 (FXR) — was highly active in healthy hepatocytes and declined progressively as the HBHi score rose. The stress arm — EGR1, JUN and ATF3 — became increasingly active in parallel. Within the stress arm, ATF3 directly drove *JUN* expression (regulatory weight = 5.2) and EGR1 also targeted *JUN* (weight = 2.7), creating a self-reinforcing loop that is difficult to interrupt once activated. On the protective side, SOX5 sustained *MACROH2A2* and NR1H4 maintained *MFSD2A* in healthy cells; as these regulators lost activity, their downstream targets were coordinately lost. This two-arm switch, in which the stress arm drives forward while the protective arm progressively fails, offers a mechanistic account of why late-stage MASH tends toward irreversibility. Notably, EGR1 transcript levels were sparse across disease stages while its regulon activity rose sharply in HCC, illustrating the value of regulon-based inference over raw transcript counts for identifying functionally active transcription factors.

Cell–cell communication analysis showed that heavily ballooned hepatocytes were the dominant signal senders across all disease stages, and that the HB_{high} hepatocyte subpopulation expanded progressively as disease advanced toward HCC. Key ligand–receptor interactions from ballooned hepatocytes to stellate cells and macrophages included FN1–CD44, FN1–(ITGAV + ITGB1) and LAMC1–(ITGA1+ITGB1), indicating extracellular matrix remodelling and basement membrane reprogramming driven by injured hepatocytes. These findings are clinically relevant because they indicate that ballooned hepatocytes are not passive bystanders but active drivers of fibrogenesis in neighbouring cells, consistent with reports that hepatic osteopontin expression correlates with fibrosis stage in patients with steatohepatitis [24] and with the broader landscape of intercellular signalling described in NASH liver [43]. The signalling axis between ballooned hepatocytes, stellate cells and macrophages may represent a therapeutic target that could be

addressed independently of resolving ballooning itself. We note that all cell–cell communication findings here are computationally inferred and require experimental confirmation.

Huh7 cells treated with 150 μ M palmitic acid showed a time-

dependent induction of JUN, KRT8 and SPP1 mRNA — significant for JUN and KRT8 by 24 h and for all three by 48 h — and western blotting confirmed increased SPP1 and KRT8 protein. These results indicate that acute lipotoxic stress alone is sufficient to activate the core signature



(caption on next page)

Fig. 3. Metabolic collapse, the lipotoxic tipping point, and in vitro validation.
A. Metabolic trajectory heatmap. Z-score normalised activity of KEGG metabolic pathways (including oxidative phosphorylation, TCA cycle and fatty acid degradation) declining along the HBHi pseudotime axis; scored per nucleus with scMetabolism.
B. The lipotoxic tipping point. Line plot showing the crossover between declining fatty acid degradation capacity and rising HBHi injury burden near the 30th percentile of the trajectory.
C. Injury pathway dynamics. Line plot (top) and heatmap (bottom) of relative activity of injury-related pathways, including reactive oxygen species, TNF/NF- κ B signalling, apoptosis and the unfolded protein response, along pseudotime.
D. In vitro validation. Left: RT-qPCR analysis of JUN, KRT8 and SPP1 in Huh7 cells treated with 150 μ M palmitic acid for 24 h or 48 h relative to 0.1% DMSO vehicle control; expression normalised to 18S rRNA and calculated by the $2^{-\Delta\Delta CT}$ method. Right: representative western blots for osteopontin (SPP1), KRT8 and β -actin (loading control) in vehicle- and PA-treated cells. Bars show mean $\pm$ SD of three independent experiments; comparisons by unpaired two-tailed Student's t-test (ns, not significant; ** $p < 0.01$; *** $p < 0.001$).

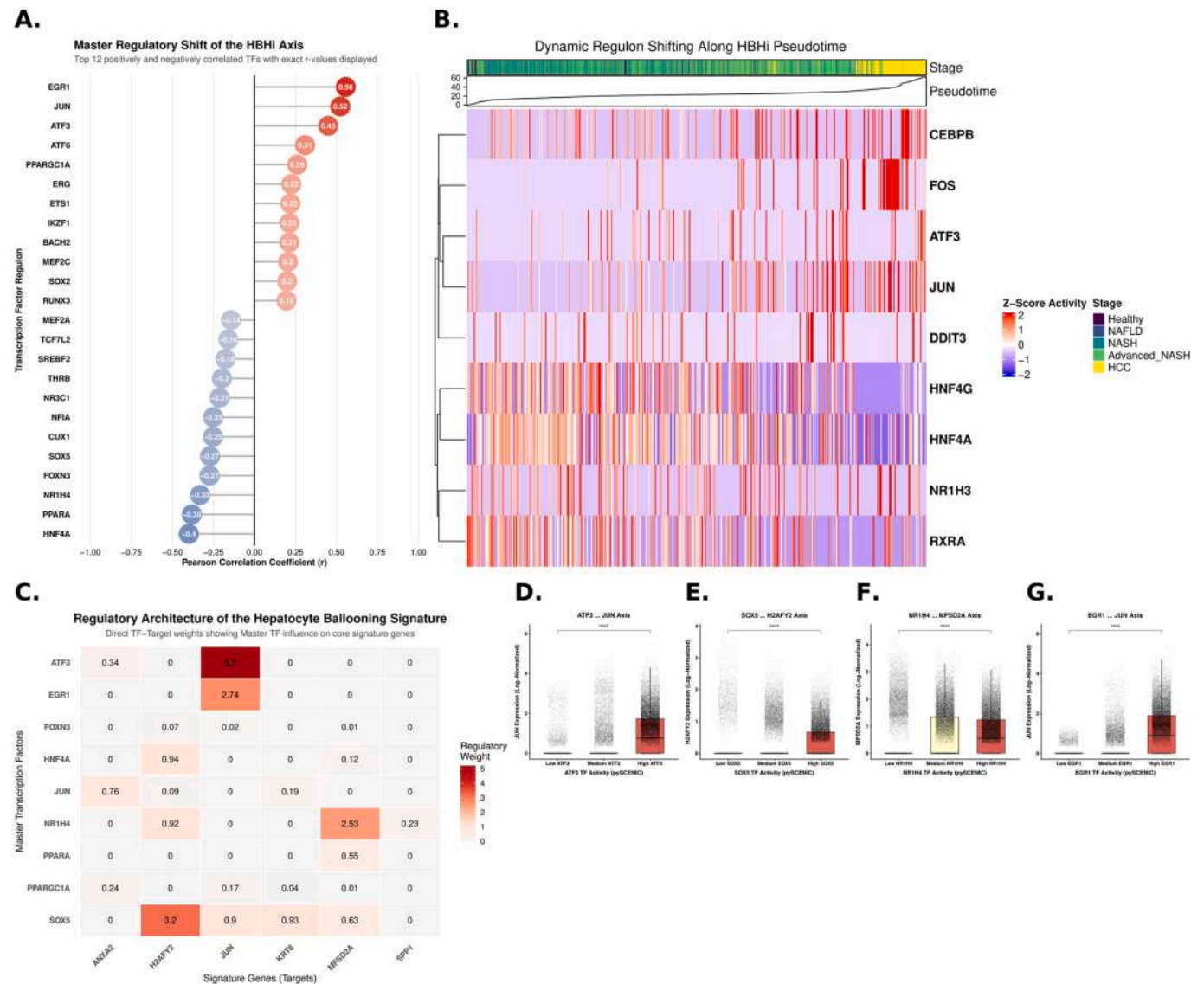


Fig. 4. Regulatory landscape and master transcription factor drivers of the HBHi axis.
A. Transcription factor correlation plot. Top 12 positively (red) and negatively (blue) correlated transcription factor regulons with the HBHi score, by Pearson correlation.
B. Dynamic regulon heatmap. Z-score normalised regulon activity along pseudotime and across disease stages (Healthy to HCC), highlighting late-acting factors including JUN, ATF3 and FOS.
C. Regulatory architecture heatmap. Matrix of direct transcription factor–target GRNBoost2 weights, identifying master regulators (SOX5, ATF3, EGR1) and their influence on the signature genes. Targets are shown for every gene with at least one inferred regulatory edge, including CXCL10 and FABP4; ANXA2 had no inferred edge and is therefore not displayed.
D–G. Transcription factor–target relationships. Boxplots of log-normalised expression of target genes (JUN, MACROH2A2, MFSD2A) stratified by regulon activity tercile (Low, Medium, High) of their respective master transcription factors (ATF3, SOX5, NR1H4, EGR1). **** $p < 0.0001$.

genes, subject to the limitations of the model discussed below.

High expression of *SPP1* (HR = 2.2, $p = 2.4 \times 10^{-5}$), *KRT8* (HR = 1.5, $p = 0.04$) and *JUN* (HR = 1.5, $p = 0.021$) was each associated with worse overall survival in TCGA-LIHC ($n = 510$). Conversely, high *MFSD2A* expression trended toward better disease-free survival (HR = 0.75), consistent with its role as a marker of healthy hepatocyte identity and periportal zonation. These associations suggest that the ballooning transcriptional state shares molecular features with oncogenic progression, which is biologically plausible given that c-Jun is required for the early stages of chemically induced hepatocellular carcinoma in mice, where it antagonises p53-dependent apoptosis [44]. These findings are nonetheless observational and derived from HCC patients rather than MASH patients; they do not establish that hepatocyte ballooning causes HCC. All hazard ratios are univariable and, given the number of genes tested, the associations for *KRT8* and *JUN* in particular should be regarded as provisional.

4.1. Limitations

This study has several important limitations. First, the discovery cohort was small ($n = 43$); although the signature was evaluated in an independent cohort (GSE130970, $n = 75$), larger and more ethnically diverse prospective cohorts are needed to establish clinical utility. Second, the bulk microarray data used for signature discovery lack single-cell resolution, and co-expression patterns observed at bulk level may mask cell-type-specific effects or reflect shifts in cell type proportions between samples. Third, the publicly available snRNA-seq datasets were collected at different centres and processed with different protocols; RPCA integration was used to reduce batch effects, but residual technical variation cannot be excluded. Fourth, the in vitro validation used Huh7 cells, a hepatocellular carcinoma line with substantially altered metabolism [40] that lacks primary hepatocyte polarity and cannot reproduce paracrine signalling. Acute palmitic acid treatment models only the immediate cellular response to lipotoxicity and does not replicate the chronic, multi-year trajectory observed in patients; in particular, sustained silencing of *MACROH2A2* and the progressive collapse of HNF4A and NR1H4 are long-term processes that cannot be reproduced in short-term culture. Fifth, the TCGA-LIHC survival analysis was performed in HCC patients rather than MASH patients; the prognostic value of the HBHi score in the MASH-to-HCC transition specifically remains to be tested in longitudinal cohorts capturing disease progression from its earliest stages. Sixth, all cell–cell communication findings are computationally inferred ligand–receptor interactions and require experimental validation.

4.2. Conclusions

This study provides a molecular model of hepatocyte ballooning as a progressive, transcriptionally active state. The eight-gene signature and HBHi score offer tools for objective, transcriptomics-based staging that may complement histological scoring. The Catastrophic Nexus — defined by the simultaneous loss of *MACROH2A2*, induction of *KRT8* and activation of *SPP1* at pseudotime ≈ 25 — and its upstream drivers ATF3 and EGR1 may define a therapeutic window. Intercepting this transition before irreversible fibrosis or malignant transformation occurs could represent a new strategy for MASH treatment. Future work should focus on prospective validation of the HBHi score in MASH cohorts, functional studies to confirm the causal roles of the identified transcription factors, and preclinical testing of interventions targeting the FN1–CD44 and FN1–ITGB1 signalling axes.

Ethics statement

The study was approved by the Institutional Ethics Committee of BRIC–National Institute of Biomedical Genomics, Kalyani, India (approval number NIBMG/2012/1/6). All procedures conformed to the

Declaration of Helsinki. Written informed consent was obtained from every participant prior to liver biopsy and sample collection.

Data availability

All single-nucleus RNA sequencing datasets analysed in this study are publicly available from the NCBI Gene Expression Omnibus under accessions GSE189175 and GSE202379. The independent bulk transcriptomic validation cohort is available under accession GSE130970. The microarray discovery dataset re-analysed here was originally described in Ref. [11]. TCGA-LIHC expression and clinical data were obtained from the UCSC Xena browser (<https://xenabrowser.net>). Custom analysis code is available from the corresponding authors upon reasonable request. No new unique reagents were generated in this study. All the codes used in this manuscript will be available from the corresponding authors with reasonable request.

CRediT authorship contribution statement

Debopriyo Ganguly: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Software, Validation, Visualization, Writing – original draft. Bandana Mondal: Data curation, Formal analysis, Investigation, Methodology, Writing – review & editing. Subhas Das: Investigation, Methodology, Validation, Writing – review & editing. Indrashish Dey: Investigation, Methodology, Validation, Writing – review & editing. Ayan Mandal: Data curation, Formal analysis, Investigation, Writing – review & editing. Trinath Ghosh: Formal analysis, Writing – review & editing. Monalisha Deo: Formal analysis, Methodology, Writing – review & editing. Misba Ahmed: Formal analysis, Methodology, Writing – review & editing. Soma Banerjee: Resources, Supervision, Writing – review & editing. Ankita Chatterjee: Conceptualization, Funding acquisition, Investigation, Project administration, Resources, Supervision, Writing – review & editing. Priyadarshi Basu: Conceptualization, Funding acquisition, Investigation, Project administration, Resources, Supervision, Writing – review & editing. All authors read and approved the final manuscript.

Ethics declaration

Written informed consent to take part in the study and to publish the article has been obtained from all participants or their legal representatives. The privacy rights of participants have been observed.

This study included organ or tissue donors. This study includes human biological material and consent was obtained by donors, or their next of kin or legal representatives, for use in this study and for publication of the article. The samples used in this research were not sourced from executed prisoners or prisoners of conscience.

This study was performed in compliance with relevant laws, regulatory frameworks and guidelines where the research took place. This study was approved by the Institutional Ethics Committee of BRIC–National Institute of Biomedical Genomics, Kalyani, India. (Approval No. NIBMG/2012/1/6)

AI usage declaration

During the preparation of this work, the authors used AI to assist with language editing, formatting and organisation of the manuscript. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the content of the published article.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors. D.G. and B. M. were supported by research fellowships from the University Grants

Commission (UGC), Government of India. A.M. was supported by a research fellowship from the Indian Council of Medical Research (ICMR), Government of India. The funding bodies had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

The authors thank all study participants. The authors thank Prof. Analabha Basu, BRIC-NIBMG, for his valuable suggestions and constant support.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.bbrep.2026.102785>.

Data availability

Data will be made available on request.

References

- [1] Z.M. Younossi, P. Golabi, J.M. Paik, A. Henry, C. Van Dongen, L. Henry, The global epidemiology of nonalcoholic fatty liver disease (NAFLD) and nonalcoholic steatohepatitis (NASH): a systematic review, *Hepatology* 77 (4) (2023) 1335–1347, <https://doi.org/10.1097/HEP.0000000000000004>.
- [2] M.H. Le, Y.H. Yeo, X. Li, J. Li, B. Zou, Y. Wu, et al., Forecasted 2040 global prevalence of nonalcoholic fatty liver disease using hierarchical Bayesian approach, *Clin. Mol. Hepatol.* 28 (4) (2022) 841–850, <https://doi.org/10.3350/cmh.2022.0239>.
- [3] T. Singla, K.N. Muneshwar, A.G. Pathade, S. Yelne, Hepatocytic ballooning in non-alcoholic steatohepatitis: bridging the knowledge gap and charting future avenues, *Cureus* 15 (9) (2023) e45884, <https://doi.org/10.7759/cureus.45884>.
- [4] Y. Li, J. Wang, Y. Tang, W. Wang, J. Zhang, Q. Zhou, et al., Hepatocytic ballooning in non-alcoholic steatohepatitis: dilemmas and future directions, *Liver Int.* 43 (6) (2023) 1170–1182, <https://doi.org/10.1111/liv.15571>.
- [5] S. Caldwell, Y. Ikura, D. Dias, K. Isomoto, A. Yabu, C. Moskaluk, et al., Hepatocellular ballooning in NASH, *J. Hepatol.* 53 (4) (2010) 719–723, <https://doi.org/10.1016/j.jhep.2010.04.031>.
- [6] E.M. Brunt, D.E. Kleiner, L.A. Wilson, P. Belt, B.A. Neuschwander-Tetri, Nonalcoholic fatty liver disease (NAFLD) activity score and the histopathologic diagnosis in NAFLD: distinct clinicopathologic meanings, *Hepatology* 53 (3) (2011) 810–820, <https://doi.org/10.1002/hep.24127>.
- [7] E.M. Brunt, C.G. Janney, A.M. Di Bisceglie, B.A. Neuschwander-Tetri, B.R. Bacon, Nonalcoholic steatohepatitis: a proposal for grading and staging the histological lesions, *Am. J. Gastroenterol.* 94 (9) (1999) 2467–2474, <https://doi.org/10.1111/j.1572-0241.1999.01377.x>.
- [8] D.E. Kleiner, E.M. Brunt, M. Van Natta, C. Behling, M.J. Contos, O.W. Cummings, et al., Design and validation of a histological scoring system for nonalcoholic fatty liver disease, *Hepatology* 41 (6) (2005) 1313–1321, <https://doi.org/10.1002/hep.20701>.
- [9] P. Bedossa, C. Poitou, N. Veyrie, J.L. Bouillot, A. Basdevant, V. Paradis, et al., Histopathological algorithm and scoring system for evaluation of liver lesions in morbidly obese patients, *Hepatology* 56 (5) (2012) 1751–1759, <https://doi.org/10.1002/hep.25889>.
- [10] Y. Takahashi, T. Fukusato, Histopathology of nonalcoholic fatty liver disease/nonalcoholic steatohepatitis, *World J. Gastroenterol.* 20 (42) (2014) 15539–15548, <https://doi.org/10.3748/wjg.v20.i42.15539>.
- [11] A. Chatterjee, A. Basu, K. Das, P. Singh, D. Mondal, B. Bhattacharya, et al., Hepatic transcriptome signature correlated with HOMA-IR explains early nonalcoholic fatty liver disease pathogenesis, *Ann. Hepatol.* 19 (5) (2020) 472–481, <https://doi.org/10.1016/j.aohep.2020.06.009>.
- [12] P. Langfelder, S. Horvath, WGCNA: an R package for weighted correlation network analysis, *BMC Bioinf.* 9 (2008) 559, <https://doi.org/10.1186/1471-2105-9-559>.
- [13] M.E. Ritchie, B. Phipson, D. Wu, Y. Hu, C.W. Law, W. Shi, et al., Limma powers differential expression analyses for RNA-seq and microarray studies, *Nucleic Acids Res.* 43 (7) (2015) e47, <https://doi.org/10.1093/nar/gkv007>.
- [14] J. Friedman, T. Hastie, R. Tibshirani, Regularization paths for generalized linear models via coordinate descent, *J. Stat. Software* 33 (1) (2010) 1–22, <https://doi.org/10.18637/jss.v033.i01>.
- [15] S.A. Hoang, A. Oseini, R.E. Feaver, B.K. Cole, A. Asgharpour, R. Vincent, et al., Gene expression predicts histological severity and reveals distinct molecular profiles of nonalcoholic fatty liver disease, *Sci. Rep.* 9 (2019) 12541, <https://doi.org/10.1038/s41598-019-48746-5>.
- [16] F.E. Harrell Jr., Regression modeling strategies: with applications to linear models. Logistic and Ordinal Regression, and Survival Analysis, second ed., Springer, Cham, 2015 <https://doi.org/10.1007/978-3-319-19425-7>.
- [17] A.J. Vickers, E.B. Elkin, Decision curve analysis: a novel method for evaluating prediction models, *Med. Decis. Mak.* 26 (6) (2006) 565–574, <https://doi.org/10.1177/0272989X06295361>.
- [18] S.H.T. Lee, K.M. Garske, U.T. Arasu, A. Kar, Z. Miao, M. Alvarez, et al., Single nucleus RNA-sequencing integrated into risk variant colocalization discovers 17 cell-type-specific abdominal obesity genes for metabolic dysfunction-associated steatotic liver disease, *EBioMedicine* 106 (2024) 105232, <https://doi.org/10.1016/j.ebiom.2024.105232>.
- [19] M. Alvarez, J.N. Benhammou, N. Darci-Maher, S.W. French, S.B. Han, J. S. Sinheimer, et al., Human liver single nucleus and single cell RNA sequencing identify a hepatocellular carcinoma-associated cell-type affecting survival, *Genome Med.* 14 (2022) 50, <https://doi.org/10.1186/s13073-022-01055-5>.
- [20] N. Darci-Maher, M. Alvarez, U.T. Arasu, I. Selvarajan, S.H.T. Lee, D.Z. Pan, et al., Cross-tissue omics analysis discovers ten adipose genes encoding secreted proteins in obesity-related non-alcoholic fatty liver disease, *EBioMedicine* 92 (2023) 104620, <https://doi.org/10.1016/j.ebiom.2023.104620>.
- [21] C. Gribben, V. Galanakis, A. Calderwood, E.C. Williams, R. Chazarra-Gil, M. Larraz, et al., Acquisition of epithelial plasticity in human chronic liver disease, *Nature* 630 (8015) (2024) 166–173, <https://doi.org/10.1038/s41586-024-07465-2>.
- [22] Y. Hao, T. Stuart, M.H. Kowalski, S. Choudhary, P. Hoffman, A. Hartman, et al., Dictionary learning for integrative, multimodal and scalable single-cell analysis, *Nat. Biotechnol.* 42 (2) (2024) 293–304, <https://doi.org/10.1038/s41587-023-01767-y>.
- [23] P.L. Germain, A. Lun, C. Garcia Meixide, W. Macnair, M.D. Robinson, Doublet identification in single-cell sequencing data using scDblFinder, *F1000Res* 10 (2021) 979, <https://doi.org/10.12688/f1000research.73600.2>.
- [24] W.K. Syn, S.S. Choi, E. Liaskou, G.F. Karaca, K.M. Agboola, Y.H. Oo, et al., Osteopontin is induced by hedgehog pathway activation and promotes fibrosis progression in nonalcoholic steatohepatitis, *Hepatology* 53 (1) (2011) 106–115, <https://doi.org/10.1002/hep.23998>.
- [25] P. Strnad, C. Stumptner, K. Zatloukal, H. Denk, Intermediate filament cytoskeleton of the liver in health and disease, *Histochem. Cell Biol.* 129 (6) (2008) 735–749, <https://doi.org/10.1007/s00418-008-0431-x>.
- [26] W. Pu, H. Zhang, X. Huang, X. Tian, L. He, Y. Wang, et al., Mfsd2a+ hepatocytes repopulate the liver during injury and regeneration, *Nat. Commun.* 7 (2016) 13369, <https://doi.org/10.1038/ncomms13369>.
- [27] S.C. Hasenfuss, L. Bakiri, M.K. Thomsen, E.G. Williams, J. Auwerx, E.F. Wagner, Regulation of steatohepatitis and PPAR γ signaling by distinct AP-1 dimers, *Cell Metab.* 19 (1) (2014) 84–95, <https://doi.org/10.1016/j.cmet.2013.11.018>.
- [28] V. Gerke, S.E. Moss, Annexins: from structure to function, *Physiol. Rev.* 82 (2) (2002) 331–371, <https://doi.org/10.1152/physrev.00030.2001>.
- [29] M. Buschbeck, S.B. Hake, Variants of core histones and their roles in cell fate decisions, development and cancer, *Nat. Rev. Mol. Cell Biol.* 18 (5) (2017) 299–314, <https://doi.org/10.1038/nrm.2016.166>.
- [30] G.S. Hotamisligil, D.A. Bernlohr, Metabolic functions of FABPs — mechanisms and therapeutic implications, *Nat. Rev. Endocrinol.* 11 (10) (2015) 592–605, <https://doi.org/10.1038/nrendo.2015.122>.
- [31] J. Cao, M. Spielmann, X. Qiu, X. Huang, D.M. Ibrahim, A.J. Hill, et al., The single-cell transcriptional landscape of mammalian organogenesis, *Nature* 566 (7745) (2019) 496–502, <https://doi.org/10.1038/s41586-019-0969-x>.
- [32] Y. Wu, S. Yang, J. Ma, Z. Chen, G. Song, D. Rao, et al., Spatiotemporal immune landscape of colorectal cancer liver metastasis at single-cell level, *Cancer Discov.* 12 (1) (2022) 134–153, <https://doi.org/10.1158/2159-8290.CD-21-0316>.
- [33] S. Aibar, C.B. González-Blas, T. Moerman, V.A. Huynh-Thu, H. Imrichova, G. Hulselms, et al., SCENIC: single-cell regulatory network inference and clustering, *Nat. Methods* 14 (11) (2017) 1083–1086, <https://doi.org/10.1038/nmeth.4463>.
- [34] B. Van de Sande, C. Flerin, K. Davie, M. De Waegeneer, G. Hulselms, S. Aibar, et al., A scalable SCENIC workflow for single-cell gene regulatory network analysis, *Nat. Protoc.* 15 (7) (2020) 2247–2276, <https://doi.org/10.1038/s41596-020-0336-2>.
- [35] S. Jin, C.F. Guerrero-Juarez, L. Zhang, I. Chang, R. Ramos, C.H. Kuan, et al., Inference and analysis of cell-cell communication using CellChat, *Nat. Commun.* 12 (1) (2021) 1088, <https://doi.org/10.1038/s41467-021-21246-9>.
- [36] The Cancer Genome Atlas Research Network, Comprehensive and integrative genomic characterization of hepatocellular carcinoma, *Cell* 169 (7) (2017) 1327–1341.e23, <https://doi.org/10.1016/j.cell.2017.05.046>.
- [37] P. Chomczynski, N. Sacchi, Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction, *Anal. Biochem.* 162 (1) (1987) 156–159, [https://doi.org/10.1016/0003-2697\(87\)90021-2](https://doi.org/10.1016/0003-2697(87)90021-2).
- [38] K.J. Livak, T.D. Schmittgen, Analysis of relative gene expression data using real-time quantitative PCR and the 2 $^{-\Delta\Delta CT}$ method, *Methods* 25 (4) (2001) 402–408, <https://doi.org/10.1006/meth.2001.1262>.
- [39] M.M. Bradford, A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding, *Anal. Biochem.* 72 (1976) 248–254, [https://doi.org/10.1016/0003-2697\(76\)90527-3](https://doi.org/10.1016/0003-2697(76)90527-3).
- [40] H. Nakabayashi, K. Taketa, K. Miyano, T. Yamane, J. Sato, Growth of human hepatoma cell lines with differentiated functions in chemically defined medium, *Cancer Res.* 42 (9) (1982) 3858–3863.

- [41] M. Ricchi, M.R. Odoardi, L. Carulli, C. Anzivino, S. Ballestri, A. Pinetti, et al., Differential effect of oleic and palmitic acid on lipid accumulation and apoptosis in cultured hepatocytes, *J. Gastroenterol. Hepatol.* 24 (5) (2009) 830–840, <https://doi.org/10.1111/j.1440-1746.2008.05733.x>.
- [42] P. Hirsova, S.H. Ibrahim, G.J. Gores, H. Malhi, Lipotoxic lethal and sublethal stress signaling in hepatocytes: relevance to NASH pathogenesis, *J. Lipid Res.* 57 (10) (2016) 1758–1770, <https://doi.org/10.1194/jlr.R066357>.
- [43] X. Xiong, H. Kuang, S. Ansari, T. Liu, J. Gong, S. Wang, et al., Landscape of intercellular crosstalk in healthy and NASH liver revealed by single-cell secretome gene analysis, *Mol. Cell* 75 (3) (2019) 644–660.e5, <https://doi.org/10.1016/j.molcel.2019.07.028>.
- [44] R. Eferl, R. Ricci, L. Kenner, R. Zenz, J.P. David, M. Rath, et al., Liver tumor development: c-Jun antagonizes the proapoptotic activity of p53, *Cell* 112 (2) (2003) 181–192, [https://doi.org/10.1016/s0092-8674\(03\)00042-4](https://doi.org/10.1016/s0092-8674(03)00042-4).