


## RESEARCH ARTICLE

# Dysregulated cholesterol regulatory genes as a diagnostic biomarker for cancer

Seema Kuldeep<sup>1</sup> | Sneha Soni<sup>1</sup> | Anubhav Srivastava<sup>2</sup> | Anjali Mishra<sup>3</sup> | Lokendra Kumar Sharma<sup>2</sup> | Chandi C. Mandal<sup>1</sup> 

<sup>1</sup>Department of Biochemistry, School of Life Sciences, Central University of Rajasthan, Ajmer, Rajasthan, India

<sup>2</sup>Department of Molecular Medicine and Biotechnology, Sanjay Gandhi Post Graduate Institute of Medical Sciences-, Lucknow, Uttar Pradesh, India

<sup>3</sup>Department of Endocrine and Breast Surgery, Sanjay Gandhi Post Graduate Institute of Medical Sciences-, Lucknow, Uttar Pradesh, India

## Correspondence

C.C. Mandal, Department of Biochemistry, School of Life Sciences, Central University of Rajasthan, NH-8, Bandarsindri, Kishangarh-305817, Ajmer, Rajasthan, India.  
Email: [chandimandal@gmail.com](mailto:chandimandal@gmail.com); [ccmandal@curaj.ac.in](mailto:ccmandal@curaj.ac.in)

## Abstract

**Background:** A dysregulation of cholesterol homeostasis is often seen in various cancer cell types, and elevated cholesterol content and that of its metabolites appears to be crucial for cancer progression and metastasis. Cholesterol is a precursor of various steroid hormones and a key plasma membrane component especially in lipid-rafts, also modulating many intracellular signaling pathways.

**Methods:** To provide an insight of dysregulated cholesterol regulatory genes, their transcript levels were analyzed in different cancers and their influence was correlated with the overall survival of cancer patients using cancer database analysis.

**Results:** This analysis found a set of genes (e.g., ACAT1, RXRA, SOAT1 and SQLE) that were not only often dysregulated, but also had been associated with poorer overall survival in most cancer types. Quantitative reverse transcriptase-polymerase chain reaction analysis revealed elevated SQLE and SOAT1 transcript levels and downregulated expression of RXRA and ACAT1 genes in triple negative breast cancer tissues compared to adjacent control tissues, indicating that this dysregulated expression of the gene signature is a diagnostic marker for breast cancer.

**Conclusion:** For the first time, the present study identified a gene signature associated with the dysregulation of cholesterol homeostasis in cancer cells that may not only be used as a diagnostic marker, but also comprise a promising drug target for the advancement of cancer therapy.

## KEYWORDS

biomarkers, cancer, cancer therapy, cholesterol, cholesterol regulatory genes, gene signature

**Abbreviations:** ABCG1, ATP binding cassette subfamily G member 1; ACAT1, Acetyl-CoA acetyltransferase 1; APOE, Apolipoprotein E; BLCA, Bladder urothelial carcinoma; BRCA, Breast invasive carcinoma; C5ORF4, Chromosome 5 open reading frame 4; COAD, Colon adenocarcinoma; DCIS, Ductal carcinoma in situ; DHCR7, 7-Dehydrocholesterol reductase; ER, Endoplasmic reticulum; GPCR, G protein-coupled receptor; HDL, High-density lipoprotein; HMGCR, 3-Hydroxy-3-methylglutaryl-CoA reductase; HNSC, Head and neck squamous cell carcinoma; HR, Hazard ratio; HSD17B7, Hydroxysteroid 17-beta dehydrogenase 7; KEGG, Kyoto encyclopedia of genes and genomes; KICH, Kidney chromophobe; KIRC, Kidney renal clear cell carcinoma; KIRP, Kidney renal papillary cell carcinoma; LDL, Low-density lipoprotein; LIHC, Liver hepatocellular carcinoma; LUAD, Lung adenocarcinoma; LUSC, Lung squamous cell carcinoma; LXR, Liver X receptor; NADA, N-arachidonoyl dopamine; PRAD, Prostate adenocarcinoma; RSEM, RNA-Seq by Expectation-Maximization; RXRA, Retinoid X receptor alpha; SCAP, SREBP cleavage activating protein; SOAT1, Sterol O-acyltransferase 1; SQLE, Squalene epoxidase; SRE, Sterol response element; SREBF1, Sterol regulatory element binding transcription factor 1; SREBP, Sterol regulatory element-binding protein; STAD, Stomach carcinoma; TCGA, The Cancer Genome Atlas; THCA, Thyroid carcinoma; UCEC, Uterine corpus endometrial carcinoma.

## 1 | INTRODUCTION

Cancers not only can develop in various organs/tissues, but also can invade almost every tissue/organ of our body. Although cancers are of various types, they have similar hallmarks, such as the ability of self-dependency for growth, ability to evade from the process of cell death (i.e., apoptosis), uncontrolled proliferation, induced angiogenesis, resistance toward growth suppressing factor, the ability to invade the nearby tissue and metastasize to distant organs,<sup>1</sup> evasion of the immune system, and reprogramming of cellular metabolism.<sup>2</sup> Hypercholesterolemia is associated with cancer development because exogenous cholesterol activates the Hedgehog signaling and endogenous cholesterol induces mTORC1 signaling.<sup>3</sup> Cholesterol also activates the Wnt/ $\beta$  catenin and phosphoinositide 3-kinase/AKT pathways, which are linked with tumor formation in various cancer types.<sup>4,5</sup> Cholesterol is a precursor molecule for various steroid hormones, a key component of lipid rafts and plasma membrane, and is also involved in regulation of various intracellular signal transduction pathways.<sup>6</sup> Cholesterol homeostasis is tightly regulated by various enzymes and transcription factors that are involved in cholesterol uptake, synthesis, export, metabolism, esterification and transcription regulation of various cholesterol regulatory genes.<sup>6</sup> Various studies, including our own, reported an elevated cellular cholesterol content in cancer cells and increased levels of expression of various key genes, including low-density lipoprotein (LDL) receptor (LDLR), 3-hydroxy-3-methylglutaryl-CoA reductase (HMGCR) and sterol regulatory element-binding protein factor 1 (SREBP1), and also a link with cancer mortality.<sup>3,7,8</sup> Cholesterol homeostasis is found to be dysregulated and can modulate the signaling pathways in cancer cells.<sup>9</sup> Cancer progression, proliferation, migration and invasion are associated with cholesterol metabolism and, accordingly, this dysregulation is linked with the hallmark in many cancers.<sup>9</sup>

Many enzymes in the *de novo* biosynthesis of cholesterol or the mevalonate pathway are dysregulated in different cancer types.<sup>1</sup> There are approaching 30 enzymatic reactions involved in cholesterol biosynthesis. Among them, HMGCR<sup>10</sup> and squalene epoxidase (SQLE) are the two key enzymes involved in rate determining steps for cholesterol biosynthesis.<sup>11,12</sup> HMGCR converts the HMG-CoA into mevalonate and SQLE catalyzes the oxidation reaction converting squalene to 2,3-epoxysqualene.<sup>13</sup> Cholesterol synthesis, transport and regulation is regulated by two major nuclear receptor systems: sterol regulatory element-binding proteins (SREBPs) and liver X receptor (LXRs). The major role of SREBP is to increase cholesterol level by transcription activation,<sup>14</sup> whereas LXR promotes reverse cholesterol transport.<sup>15,16</sup>

As a result of the non-polar nature of cholesterol, it is transported with the lipoproteins. These lipoproteins remove the cholesterol from the peripheral tissue and transport it to the liver thus maintaining the cholesterol homeostasis. Accordingly, lipoproteins have a fundamental role in cancer development by supplying cholesterol to the tumor cells.<sup>17</sup> The major cholesterol transporters are LDL and high-density lipoprotein (HDL). LDL transports the cholesterol to the peripheral tissue via LDL receptor (LDLR) present on the plasma membrane, whereas

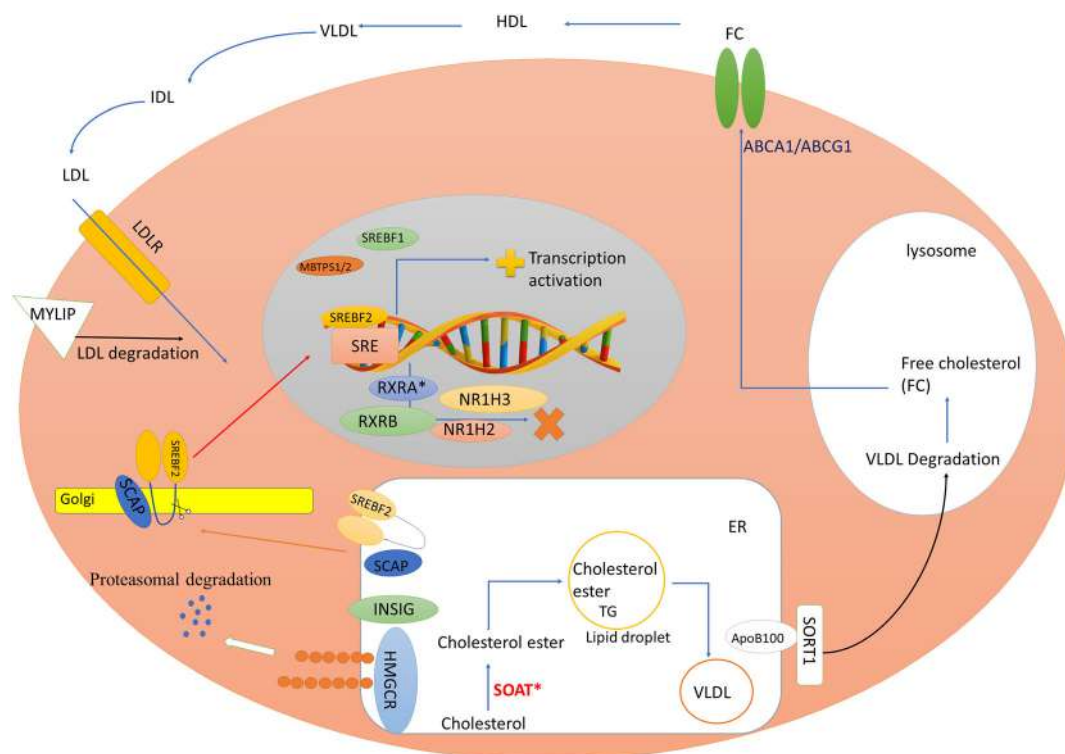
HDL removes the excess cholesterol from the peripheral tissue and delivers it back to the liver, this is called reverse cholesterol transport.<sup>17-19</sup> The efflux of cholesterol from peripheral tissue and converted into HDL is facilitated by ABCA1 and ABCG1 transporters.<sup>17</sup>

Increased levels of cholesterol esters are often observed in breast cancer, leukemia, glioma and prostate cancer.<sup>20</sup> However, the contribution of cholesterol esters in tumor progression and metastasis is not well understood.<sup>20</sup> Our previous study found a presence of elevated cholesterol in high-metastatic breast cancer and malignant breast cancer tissues.<sup>8</sup> Moreover, elevation of cholesterol has been linked with cancer mortality.<sup>7</sup> In addition, our research group had also documented that cholesterol lowering drug statins inhibit cancer growth and osteolytic metastasis of breast cancer, as demonstrated in animal models.<sup>21-23</sup> Moreover, our recent studies documented that antidiabetic metformin and *N*-arachidonoyl dopamine showed anticancer activity by decreasing cellular cholesterol content in breast cancer cells.<sup>8,24</sup> Thus, taking the idea from the past evidence for the dysregulated cholesterol biosynthesis genes in different cancers, the present study focuses on the impact of the dysregulated gene involved in cholesterol biosynthesis, esterification, efflux, transport, uptake, transcription regulation on the patient's survival in various cancer types. This is a holistic approach aiming to identify the dysregulated genes of cholesterol metabolism and regulation in cancer types and to analyze its effect on a patient's survival. The study addresses a gene signature of the dysregulated cholesterol metabolism and regulation pathway for diagnosis and predictive biomarker in cancer types and provides a possible therapeutic target for improvement in cancer therapy. Experimental quantitative reverse transcriptase-polymerase chain reaction (qRT-PCR) analysis of breast cancer tissues had validated this gene signature (e.g., ACAT1, RXRA, SOAT1 and SQLE) as a diagnostic marker for triple negative breast cancer.

## 2 | MATERIALS AND METHODS

### 2.1 | Selection of genes involved in the cholesterol pathway

Primarily, 48 genes involved in cholesterol regulation were taken from Kyoto Encyclopedia of Genes and Genomes (KEGG) (<https://www.genome.jp/kegg>) pathway database and various literature sources.<sup>25</sup> Here, 25 genes involved in cholesterol pathway were obtained from KEGG, whereas four genes linked with cholesteryl ester synthesis and 19 genes associated with the cholesterol regulation, uptake, efflux and transport were selected from various literature sources.<sup>25</sup> The source of genes associated with cholesterol pathway from the KEGG database was the steroid biosynthesis and mevalonate pathway under the category of terpenoid backbone biosynthesis. Furthermore, the selected genes have been verified again by other databases such as REACTOME<sup>26,27</sup> and wikipathways (<https://wikipathways.org>), and, finally, 51 genes were selected for study. The cholesterol regulatory pathway and the complete list of genes participating in cholesterol regulation are shown in Figure 1 and Table 1.



**FIGURE 1** Cholesterol regulatory mechanism. Genes that are involved in regulating the cholesterol synthesis, uptake and transport are shown. There is a transcription factor that regulates the cholesterol synthesis pathway in the cell-sterol regulatory element-binding proteins (SREBP1 and SREBP2). When the endoplasmic reticulum (ER) membrane has high cholesterol, SREBP cleavage activating protein (SCAP) undergoes conformational changes and binds to INSIG-1, which is a ER membrane anchor protein; this allow the SCAP/SREBPs complex to stay in ER membrane but, under low cholesterol content in ER membrane, INSIG-1 dissociate from the SCAP, which allows SCAP/SREBP complex to move into the Golgi complex, where SREBP is proteolytically cleaved by S1P and S2P. This results into nuclear translocation of N-terminal domain of SREBP where it binds to the sterol response element (SRE) region in the DNA and activate the transcription of cholesterol biosynthesis gene, uptake and fatty acid synthesis.<sup>13,46,47</sup> Among lipoproteins, low-density lipoprotein (LDL) and high-density lipoprotein (HDL) are the important cholesterol transporters, with LDL transporting the cholesterol to the peripheral tissue via LDL receptor present on the plasma membrane, whereas HDL removes the excess cholesterol from the peripheral tissue and transports to the liver; this is called as reverse cholesterol transport. The efflux of cholesterol is facilitated by ABCA1 and ABCG1 transporters. HMGCR degradation is mediated by INSIG1, INSIG2, AMFR. Transcription repression factor are RXRA, RXRB, NR1H3 and NR1H2. ACAT1\*, SQLE\*, HSD17B7\*, SOAT1\* and RXRA\* are constituents of the gene signature.

## 2.2 | Transcript analysis in various cancer types

We had aimed to analyze clinical data of cancer patients of 31 cancer types available in cancer data portals such as UALCAN (<http://ualcan.path.uab.edu>) and firebrowse (<http://firebrowse.org>) based on The Cancer Genome Atlas (TCGA) database.<sup>28</sup> The transcript levels of the cholesterol metabolism and regulatory genes shown in the Table 1 were compared between tumor and control tissues using both the UALCAN and firebrowse servers. Fourteen cancer types out of 31 were considered for this transcript analysis considering those cancer types that had at least 20 or above samples in the corresponding control. Fold change in transcript level was computed by considering the median values for tumor and control in all selected cancer types.

The UALCAN server based on RNA-sequencing (RNA-seq) level 2 was used to calculate transcript levels of various genes.<sup>28</sup> In the present study, UALCAN was used to analyze the change in the transcript

levels of cholesterol metabolism and regulatory genes in tumors compared to control tissues.

RNA-seq data are also available in the firebrowse database, comprising one of the computational tools of TCGA.<sup>29</sup> It provides graphical tools to explore expression level and uses RNA-seq by the expectation-maximization estimation method (<https://www.cancer.gov/tcga>). We used the firebrowse to compare the mRNA expression of the cholesterol metabolism and regulatory genes between tumor and control tissues. Supporting data for mRNA expression analysis available are provided in the Supporting information (Table S1).

## 2.3 | Overall survival analysis

Overall survival is defined as fraction of people in a study who are alive for a period of time after diagnosed with a disease such as cancer.<sup>30</sup> cBioportal allows us to explore, visualize, and analyze genomic

**TABLE 1** List of genes involved in cholesterol metabolism and regulation.

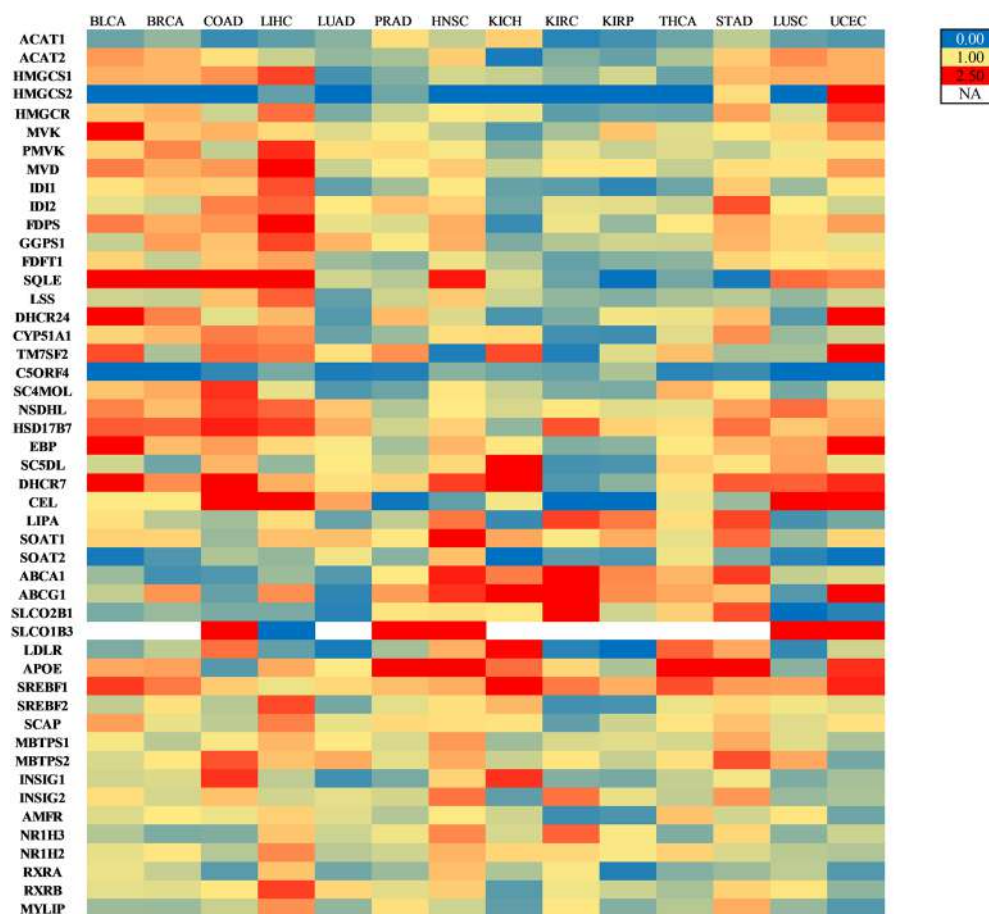
S. no	Gene	Gene name	Function
1	ACAT1	acetyl-CoA acetyltransferase 1	Synthesis
2	ACAT2	acetyl-CoA acetyltransferase 2	
3	HMGCS1	3-hydroxy-3-methylglutaryl-CoA synthase 1	
4	HMGCS2	3-hydroxy-3-methylglutaryl-CoA synthase 2	
5	HMGCR	3-hydroxy-3-methylglutaryl-CoA reductase	
6	MVK	mevalonate kinase	
7	PMVK	phosphomevalonate kinase	
8	MVD	mevalonate decarboxylase	
9	IDI1	isopentenyl-diphosphate delta isomerase 1	
10	IDI2	isopentenyl-diphosphate delta isomerase 2	
11	FDPS	farnesyl diphosphate synthase	
12	GGPS1	geranylgeranyl diphosphate synthase 1	
13	FDFT1	farnesyl-diphosphate farnesyltransferase 1	
14	SQLE	squalene epoxidase	
15	LSS	lanosterol synthase	
16	DHCR24	24-dehydrocholesterol reductase	
17	CYP51A1	cytochrome P450 family 51 subfamily A polypeptide 1	
18	TM7SF2	transmembrane 7 superfamily member 2	
19	FAXDC2/C5ORF4	fatty acid hydroxylase domain containing 2/chromosome 5 open reading frame 4	
20	MSMO1	methyl sterol monooxygenase	Esterification
21	NSDHL	NAD(P) dependent steroid dehydrogenase-like	
22	HSD17B7	hydroxysteroid (17-beta) dehydrogenase 7	
23	EBP	emopamil binding protein (sterol isomerase)	
24	SC5D	sterol-C5-desaturase	
25	DHCR7	7-dehydrocholesterol reductase	
26	LBR	Lamin B receptor	
27	PLPP6	Phospholipid phosphatase 6	
26	CEL	carboxyl ester lipase	
29	LIPA	lipase A, lysosomal acid, cholesterol esterase	
30	SOAT1	sterol O-acetyltransferase 1	Efflux
31	SOAT2	sterol O-acetyltransferase 2	
32	ABCA1	ATP-binding cassette, sub-family A	
33	ABCG1	ATP-binding cassette, sub-family G	
34	SLCO2B1	solute carrier organic anion transporter family member 2B1	
35	SLCO1B3	solute carrier organic anion transporter family member 1B3	Transport
36	ARV1	ARV1 homolog, fatty acid modulator	
37	LDLR	low density lipoprotein receptor	Uptake
38	APOE	apolipoprotein E	
39	SREBF1	sterol binding element transcription factor 1	Transcriptional activation
40	SREBF2	sterol binding element transcription factor 2	
41	SCAP	SREBF chaperone	
42	MBTPS1	membrane bound transcription factor peptidase site 1	
43	MBTPS2	membrane bound transcription factor peptidase site 2	
44	INSIG1	insulin induced gene 1	HMGCR degradation

**TABLE 1** (Continued)

S. no	Gene	Gene name	Function
45	INSIG2	insulin induced gene 2	
46	AMFR	autocrine motility factor receptor E3 ubiquitin protein ligase	
47	NR1H3	nuclear receptor subfamily 1 group H member 3	Transcriptional repression
48	NR1H2	nuclear receptor subfamily 1 group H member 2	
49	RXRA	retinoid X receptor alpha	
50	RXRB	retinoid X receptor beta	
51	MYLIP	myosin regulatory light chain interacting protein	Degradation of LDLR

Twenty-five genes are involved in cholesterol biosynthesis. HMGCR and SQLE are the two-rate limiting enzyme in cholesterol biosynthesis. Four genes are involved in cholesteryl ester formation, and 19 are involved in efflux, uptake, transport and cholesterol regulation.

**FIGURE 2** Heatmap representing fold change of cholesterol related gene in different cancer type using UALCAN with data normalized from 0 (blue) to 2.5 (red). The gene expression data were from UALCAN and firebrowse. Fold change was calculated by taking ratio of tumor/control. BLCA, bladder urothelial carcinoma; BRCA, breast invasive carcinoma; COAD, Colon adenocarcinoma; LIHC, Liver hepatocellular carcinoma; LUAD, Lung adenocarcinoma; PRAD, Prostate adenocarcinoma; HNSC, Head and neck squamous cell carcinoma; KICH, Kidney chromophobe; KIRC, Kidney renal clear cell carcinoma; KIRP, Kidney renal papillary cell carcinoma; THCA, Thyroid carcinoma; STAD, Stomach carcinoma; LUSC, Lung squamous cell carcinoma; UCEC, Uterine corpus endometrial carcinoma. NA; not available (white colour).



data of cancer.<sup>31,32</sup> In the present study, we used cBioportal to check the effect of dysregulated genes of cholesterol metabolism and regulatory pathway on the overall survival of the patients having different cancer types. For all of the dysregulated genes, the influence of both high expression and low expression on the overall survival (in terms of months) was analyzed.

Kaplan–Meier plotter (<http://kmplot.com/analysis>) was used to analyze clinical data to assess the prognosis.<sup>33</sup> To assess the prognostic value, patients are categorized into two cohorts on the basis of

median expression (higher vs. lower). We had examined the individual overall survival of the dysregulated genes and combinatory overall survival for various cancer types.

### 2.3.1 | Patient sample information

In total, 18 histopathologically confirmed Triple Negative Breast Cancer (TNBC: ER-/PR-/HER2-) paired tissue specimens (tumor and



adjacent non-tumor breast tissue) were used, which were obtained during surgical procedure from the Department of Endocrine and Breast Surgery, Sanjay Gandhi Post Graduate Institute of Medical Sciences, Lucknow, India. Patient samples were collected following institutional ethical guidelines with ethical approval from institute's ethical committee (IEC code: 2019-43-IP-108, Ref. PGI/BE/311/2019). Tumor (tissue within the tumor boundary) and adjacent non-tumor (distal normal tissue at least 10 mm from the outer tumor boundary) tissues were collected. The information of each human subjects is provided in the Supporting information (Table S2). A fraction of the resected tissues was stored in RNeasy<sup>TM</sup> (Thermo-Fisher Scientific Inc., Waltham, MA, USA) and then stored at  $-80^{\circ}\text{C}$  for RNA extraction.

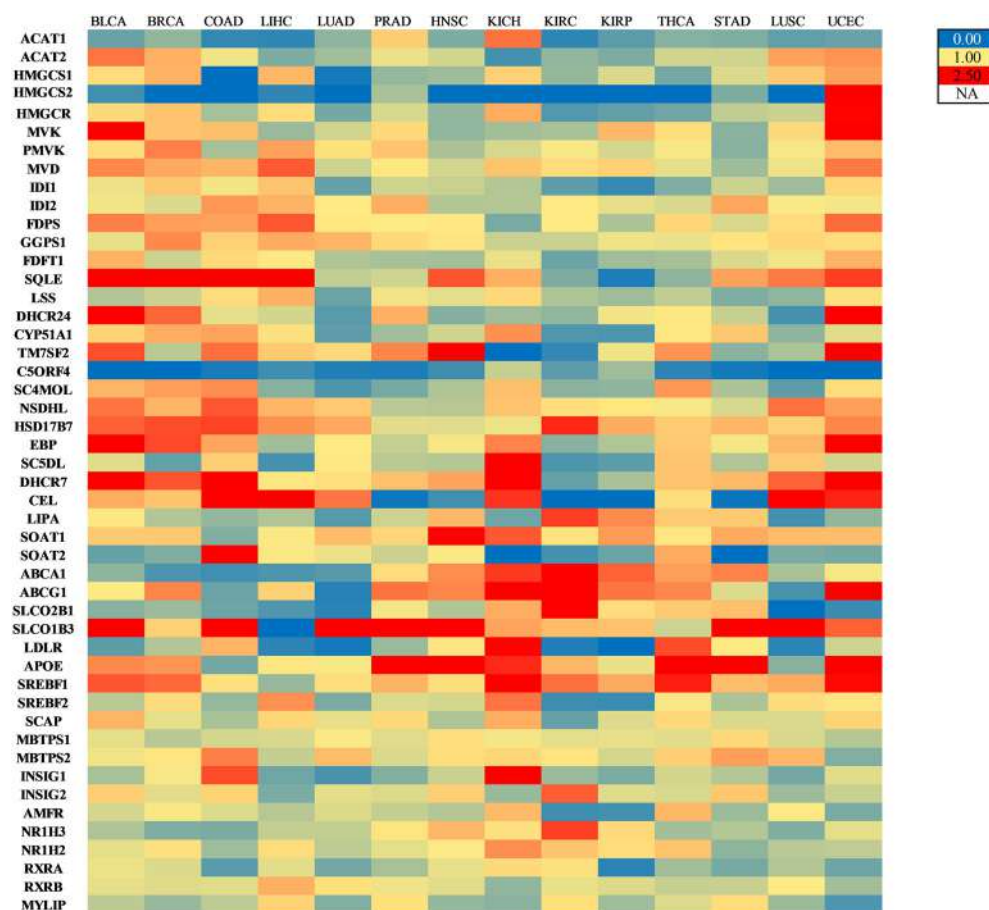
### 2.3.2 | qRT-PCR analysis

For gene expression studies, total RNA was isolated from 25–30 mg of the tissue. Tissue was homogenized in liquid nitrogen followed by RNA isolation using RNeasy Plus (Takara Bio Inc., Shiga, Japan) in accordance with the manufacturer's instructions. The RNA was quantified using a Nanodrop UV-visible spectrophotometer (Thermo-Fisher Scientific Inc.) at  $A_{260}$  and RNA quality was assessed

via the  $A_{260/280}$  ratio (range 1.8–2). In total, 1  $\mu\text{g}$  of RNA was reversed transcribed into cDNA using the RevertAid First Strand cDNA Synthesis Kit (Thermo Fisher Scientific Inc.) in accordance with the manufacturer's instructions. qPCR was performed using TB Green Premix Ex Taq<sup>TM</sup> (Takara Bio Inc.) in accordance with the manufacturer's instructions. The qPCR conditions were: initial denaturation for 3 min at  $95^{\circ}\text{C}$ , followed by 40 cycles at  $95^{\circ}\text{C}$  for 15 s and  $60^{\circ}\text{C}$  for 60 s. Relative gene expression of target genes was normalized to  $\beta$ -actin expression (reference gene) using the  $2^{-\Delta\Delta C_T}$  method.<sup>34</sup> Data were compiled and statistical analysis was conducted using Prism, version 9 (GraphPad Software Inc., San Diego, CA, USA). Data are presented as the mean  $\pm$  SEM. The list of primers used for RT qPCR of the various genes is provided in the Supporting information (Table S3).

### 2.4 | Statistical analysis

In the present study, we analyzed the transcript level of 51 genes involved in cholesterol metabolism and regulatory pathways and fold change data were calculated. We performed Student's *t* test to inspect the genes that are dysregulated significantly compared to control samples.  $p < 0.05$  was considered statistically significant.



**FIGURE 3** Heatmap representing fold change of cholesterol related gene in different cancer type using firebrowse with data normalized from 0 (blue) to 2.5 (red). The gene expression data were from firebrowse. Fold change was calculated by taking ratio of tumor/control. BLCA, bladder urothelial carcinoma; BRCA, breast invasive carcinoma; COAD, colon adenocarcinoma; HNSC, head and neck squamous cell carcinoma; KICH, kidney chromophobe; KIRC, kidney renal clear cell carcinoma; KIRP, kidney renal papillary cell carcinoma; LIHC, liver hepatocellular carcinoma; LUAD, lung adenocarcinoma; LUSC, lung squamous cell carcinoma; NA, not available (white colour); PRAD, prostate adenocarcinoma; STAD, stomach carcinoma; THCA, thyroid carcinoma; UCEC, uterine corpus endometrial carcinoma.

**TABLE 2** Hazard ratio and log-rank *p* value for the dysregulated gene of cholesterol metabolism in cancer type using Kaplan–Meier plotter.

Cancer	ACAT1		SQLE		SREBF1		RXRA		FAXDC2/C5ORF4	
	HR	Logrank P	HR	Logrank P	HR	Logrank P	HR	Logrank P	HR	Logrank P
Bladder carcinoma	1.54	0.0095	1.51	0.027	1.59	0.0019	0.88	0.42	1.13	0.42
breast cancer	1.63	0.0055	1.5	0.02	0.72	0.091	1.18	0.32	0.62	0.0066
Head and neck squamous cell carcinoma	1.26	0.11	1.6	$7.00E \times 10^{-4}$	0.67	0.008	0.77	0.076	0.74	0.027
Kidney renal clear cell carcinoma	0.41	$4.50 \times 10^{-9}$	1.5	0.011	1.16	0.32	0.5	$2.00 \times 10^{-5}$	0.49	$8.60 \times 10^{-6}$
Kidney renal papillary cell carcinoma	0.29	$1.40 \times 10^{-5}$	4.24	$6.50 \times 10^{-7}$	0.45	0.0089	3.12	$7.30 \times 10^{-5}$	0.47	0.011
Liver hepatocellular carcinoma	0.47	$3.00 \times 10^{-5}$	1.58	0.0095	1.35	0.098	0.77	0.15	0.51	0.00021
Lung adenocarcinoma	0.64	0.0028	1.7	0.00029	0.87	0.35	0.66	0.018	0.73	0.036
Lung squamous cell carcinoma	0.85	0.31	1.31	0.061	1.41	0.041	1.41	0.013	1.28	0.096
Stomach adenocarcinoma	1.23	0.22	1.3	0.11	0.64	0.011	0.67	0.016	1.5	0.015
Thyroid carcinoma	2.34	0.082	12.87	0.0014	0.42	0.077	2.28	0.1	4.23	0.002
Uterine corpus endometrial carcinoma	0.66	0.048	1.72	0.0095	0.47	0.00029	0.73	0.15	1.46	0.076

The prognostic value of mRNA levels dysregulated genes in different cancer types is shown with the respective log-rank *p* values. Low expression of ACAT1, RXRA and C5ORF4 is associated with poor overall survival, whereas higher expression of SQLE, SREBF1, DHCR7, SOAT1, HSD17B7, ABCG1 and APOE is associated with shorter overall survival in most of the cohorts. COAD, colon adenocarcinoma; KICH, kidney chromophobe; and PRAD, prostate adenocarcinoma cohorts are unavailable in Kaplan–Meier plotter.

**TABLE 2** (Continued)

Cancer	DHCR7		SOAT1		HSD17B7		ABCG1		APOE		HMGCS2	
	HR	Logrank P	HR	Logrank P	HR	Logrank P	HR	Logrank P	HR	Logrank P	HR	Logrank P
Bladder carcinoma	1.66	0.00064	1.21	0.21	0.74	0.055	0.72	0.038	1.34	0.11	0.54	0.0019
breast cancer	0.77	0.13	0.74	0.13	0.82	0.25	1.34	0.1	0.74	0.065	1.21	0.3
Head and neck squamous cell carcinoma	1.54	0.0027	1.52	0.0018	1.55	0.0017	0.81	0.17	0.84	0.2	1.24	0.16
Kidney renal clear cell carcinoma	0.64	0.003	1.41	0.028	1.2	0.29	0.5	$5.40 \times 10^{-6}$	1.42	0.024	0.49	$1.4e-06$
Kidney renal papillary cell carcinoma	1.84	0.042	1.98	0.022	2.32	0.0049	0.41	0.004	0.56	0.063	0.68	0.21
Liver hepatocellular carcinoma	1.59	0.01	1.62	0.0078	1.25	0.28	1.5	0.057	0.62	0.011	0.41	$2.4e-07$
Lung adenocarcinoma	1.54	0.0073	1.3	0.1	0.62	0.0028	1.27	0.12	1.35	0.092	0.64	0.0031
Lung squamous cell carcinoma	0.8	0.11	1.27	0.11	0.88	0.37	0.89	0.39	1.21	0.17	1.18	0.23
Stomach adenocarcinoma	1.2	0.28	1.54	0.0085	1.31	0.14	1.52	0.011	1.33	0.16	1.36	0.075
Thyroid carcinoma	5.04	0.082	2.69	0.047	0.32	0.063	0.45	0.28	0.21	0.0017	2.78	0.048
Uterine corpus endometrial carcinoma	1.75	0.01	0.62	0.028	0.75	0.16	0.48	0.0045	1.83	0.03	1.43	0.1

The prognostic value of mRNA levels dysregulated genes in different cancer types is shown with the respective log-rank *p* values. Low expression of ACAT1, RXRA and C5ORF4 is associated with poor overall survival, whereas higher expression of SQLE, SREBF1, DHCR7, SOAT1, HSD17B7, ABCG1 and APOE is associated with shorter overall survival in most of the cohorts. COAD, colon adenocarcinoma; KICH, kidney chromophobe; and PRAD, prostate adenocarcinoma cohorts are unavailable in Kaplan–Meier plotter.

### 3 | RESULTS

#### 3.1 | Dysregulation of genes involved in cholesterol metabolism and regulation in different cancer types

To explore the diagnostic potential of cholesterol metabolism and regulatory genes, mRNA expression was analyzed using the UALCAN and firebrowse databases. mRNA expression of a total of 51 genes involved in cholesterol metabolism and its regulation in 14 cancer types was analyzed by comparing patient tumor samples with normal samples, and fold changes (tumor: control) were calculated. On the basis of fold change, a heat map was constructed of 51 genes in 14 different cancer types by normalizing the data in the range 0–2.5 as shown in Figures 2 and 3. The range 0–2.5 depicts dysregulation at the transcript level. Moreover, fold change  $< 1$ , fold change  $> 1$  and fold change = 1 were considered as downregulation, upregulation and no change in transcript levels, respectively, when calculating the ratio between tumor and normal tissues. The fold change for all 51 genes in 14 different cancer types obtained via both UALCAN and firebrowse is shown in Figures 2 and 3. We noted that the transcript levels of many genes were dysregulated in these 14 cancer types. Out of these genes, 11 genes were selected that were perturbed. The basis for the selection of these 11 genes was the consistent dysregulation in the majority of the cancer types. In brief, acetyl-CoA acetyltransferase 1 (ACAT1), chromosome 5 open reading frame 4 (C5ORF4), retinoid X receptor alpha (RXRA) and 3-hydroxy-3-methylglutaryl-CoA synthase 2 (HMGCS2) showed a consistent lower transcript levels in cancer tissues compared to normal tissues, whereas higher mRNA expression of ATP binding cassette subfamily G member 1 (ABCG1), apolipoprotein E (APOE), 7-dehydrocholesterol reductase (DHCR7), hydroxysteroid 17-beta dehydrogenase 7 (HSD17B7), sterol O-acyltransferase 1 (SOAT1), sterol regulatory element binding transcription factor 1 (SREBF1) and squalene epoxidase (SQLE) was observed in tumor samples compared to control samples in most of the 14 cancer types.

#### 3.2 | Dysregulated mRNA expression associated with overall survival of cancer patients

Furthermore, we used Kaplan–Meier (KM) plotter to analyze the impact of dysregulated gene expression on patient overall survival to assess the prognosis role of dysregulated genes. Out of 14 selected cancer studies preferred for transcript analysis, 11 cancer types were available in the KM plotter server for the analysis of overall survival (Table 2) that depicts the hazard ratio (HR) and significance of the KM plots. If HR is  $> 1$ , this indicates that the patient has a shorter or poor overall survival, whereas HR  $< 1$  indicates that the patient has less likely to have poor survival when comparing between high and low expression genes. SQLE showed that a higher HR value predicts that it may be associated with poorer overall survival in all 11 cancer types. Similarly, SOAT1 showed higher HR value in nine cancer types.

Indeed, the downregulated genes, namely RXRA, ACAT1, C5ORF4 and HMGCS2, were shown to affect overall survival in few cancer types, as depicted in Table 2. To validate these findings enable a proper conclusion in context with overall survival, cBioportal was utilized to analyze the effect of these dysregulated genes on the overall survival of the cancer patients. SQLE and SOAT1 showed decreased median month overall survival when their mRNA expression was high in majority of cancer types. For ACAT1, low expression also impacted the overall survival in few cancer types and RXRA low expression consistently affected the overall survival, as shown in Table 3.

#### 3.3 | A combination of four dysregulated genes associated with shorter overall survival in all 11 cancer types

To assess the cumulative effect of dysregulated gene (both upregulated and downregulated) on overall patient survival, the KM plots were plotted by selecting different possible combination of two or three genes in KM plotter. Different possible combinations were made by selecting consistently one upregulated and one downregulated gene. Here, we found that a three gene combination (RXRA, SQLE and SOAT1) was associated with the most poorer overall survival in all 11 cancer types. Further to this combination, we attempted to make a four gene combination by adding one upregulated gene and one downregulated gene to this combination of three genes and found that two combinations of four genes (i.e., RXRA, SQLE, SOAT1 and ACAT1; RXRA, SQLE, SOAT1 and HSD17B7) were strongly associated with the poorer overall survival in all 11 cancer studies (Table 4). These two combinations (RXRA, SQLE, SOAT1 and ACAT1/HSD17B7) showed a decreased overall survival in all 11 cancer types. We then determined a five gene combination, but none showed poorer survival in all 11 cancer types. A four gene (RXRA, SQLE, SOAT1 and ACAT1/HSD17B7) combination showed minimum violation of poorer survival in 11 cancer cohorts (Table 4). These findings suggested that a four gene (RXRA, SQLE, SOAT1 and ACAT1/HSD17B7) combination appears to be a gene signature for the prognosis in all 11 cancer types. Overall survival graphs for the two gene signature combinations are shown in Figures 4 and 5.

#### 3.4 | Dysregulated expression of the gene signature associated with node metastases and advanced cancer stages

Because dysregulated expression of RXRA, SQLE, SOAT1, ACAT1/HSD17B7 genes was linked with the overall survival of patients in different cancer types, with the aim of further evaluating the rationale behind the association of these dysregulated genes and poorer survival, we analyzed gene expression in nodes and cancer stages. Here, we noted that transcripts of ACAT1 and RXRA were downregulated in node positive and higher cancer stage tumor samples compared to normal tumors in most of the cancer types (Figure 6). Similarly, a



**TABLE 3** Overall survival data analyzes using cBioportal.

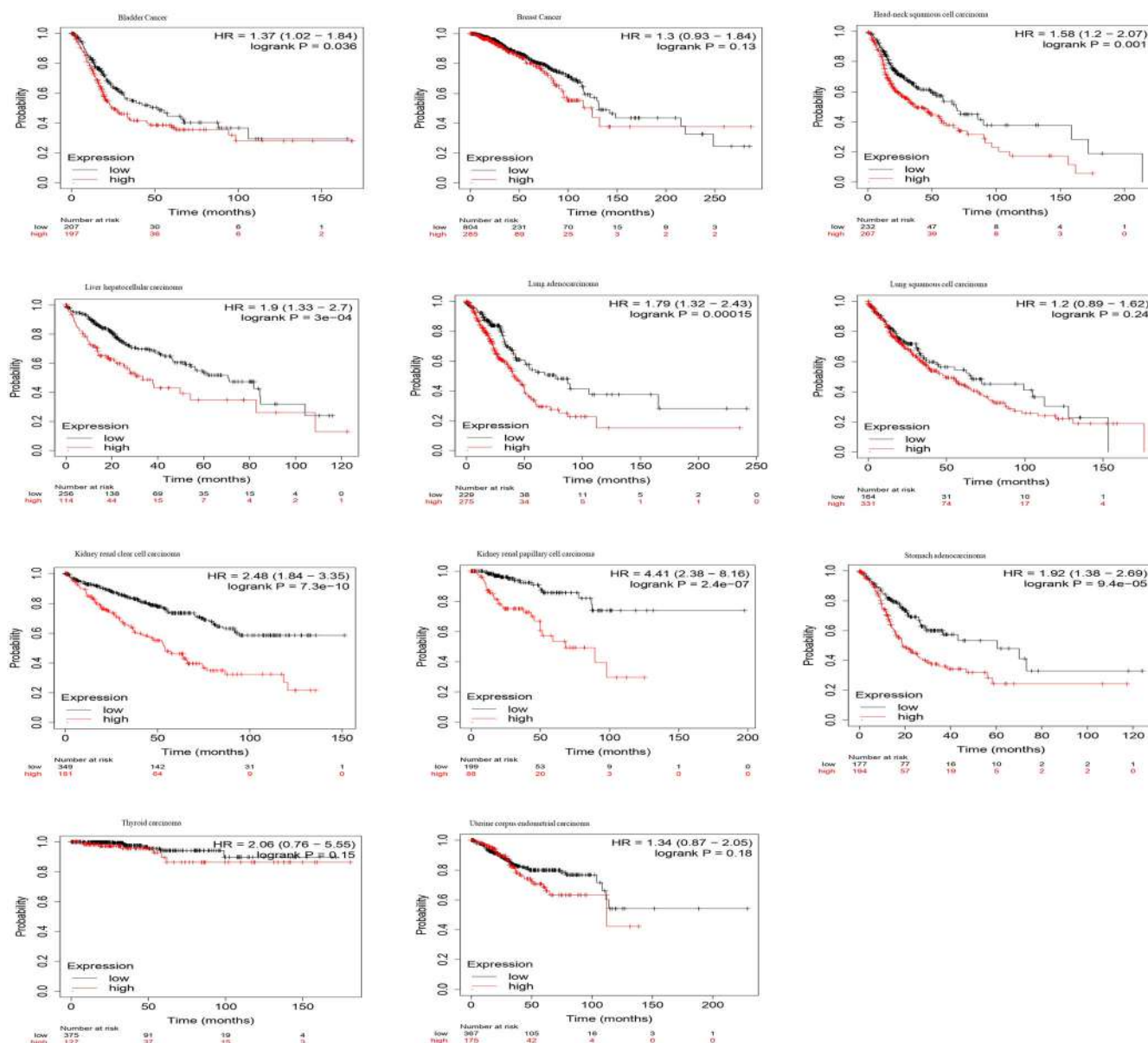
Genes	Cancer	BLCA	BRCA	HNSC	KIRC	LIHC	LUAD	LUSC	STAD
ABCG1	High	41.75	129.07	37.25	NA	NA	53.33	36.39	16.31
	Low	21.86	212.25	42.38	39.13	102.66	42.34	44.19	55.43
ACAT1	High	17.98	129.7	10.75	NA	70.01	59.11	65.23	9.67
	Low	41.75	NA	57.47	56.68	13.76	34.29	75.09	28.96
APOE	High	59.31	NA	67.81	NA	107.03	86.04	64.21	28.73
	Low	104.65	122.73	100.49	NA	20.6	50.24	73.12	NA
FAXDC2/C5ORF4	High	34.06	129.47	65.82	NA	40.37	50.24	54.44	19.96
	Low	44.32	128.98	53.95	56.31	20.11	40.41	55.73	NA
DHCR7	High	24.3	90.77	19.91	69.15	39.75	41.36	46.78	22.19
	Low	92.97	128.98	NA	73.15	107.03	53.65	64.93	28.57
HSD17B7	High	44.32	NA	49.41	NA	19	58.85	39.12	21.04 (16.14 - NA)
	Low	19.4	127.23	76.18	NA	51.25	37.71	64.93	57.43 (22.19 - NA)
RXRA	High	51.16	244.91	30.06	NA	83.18	104.19	110.99	42.54
	Low	33.04	129.6	32.19	75.58	37.75	49.25	68.58	22.19
SOAT1	High	20	NA	47.01	53.38	29.53	39.25	38.17	21.44 (15.58–42.54)
	Low	17	140.28	56.48	52.2	70.01	39.35	75.09	55.43 (34.29 - NA)
SQLE	High	20.22	NA	32.36	53.46	45.07	42.34	33.07	23.41 (17.85 - NA)
	Low	NA	140.18	57.42	93.04	58.84	66.64	64.21	28.57 (18.05 - NA)
SREBF1	High	29.7	129.7	68.43	42.42	53.35	50.33	53.92	42.54
	Low	33.02	112.37	100.49	62.82	39.75	42.51	70.13	59.54
HMGCS2	high	NA	150.60	NA	NA	70.01	NA	NA	NA
	low	NA	168.60	NA	52.89	37.29	20.63	NA	NA

High expression of mRNA is considered at more than one ( $\exp > 1$ ) and low expression at less than minus one ( $\exp < -1$ ). Median month data analyzed for overall survival of patients from different cohorts under expression of dysregulated cholesterol related genes. BLCA, bladder urothelial carcinoma; BRCA, breast invasive carcinoma; LIHC, liver hepatocellular carcinoma; LUAD, lung adenocarcinoma; HNSC, head and neck squamous cell carcinoma; KIRC, kidney renal clear cell carcinoma; STAD, stomach adenocarcinoma; LUSC, lung squamous cell carcinoma. Data for COAD, colon adenocarcinoma; KICH, kidney chromophobe; KIRP, kidney renal papillary cell carcinoma; PRAD, prostate adenocarcinoma; THCA, thyroid carcinoma; and UCEC, uterine corpus endometrial carcinoma cohort is unavailable at expression  $> 1$  and  $< -1$ .

**TABLE 4** Combinatory gene analysis by Kaplan–Meier in 11 cancer studies.

	RXRA, SQLE, SOAT1		RXRA, SQLE, SOAT1, ACAT1		RXRA, SQLE, SOAT1, HSD17B7		RXRA, SQLE, SOAT1, HSD17B7, ACAT1	
	HR	Log-rank <i>p</i>	HR	Log-rank <i>p</i>	HR	Log-rank <i>p</i>	HR	Log-rank <i>p</i>
Bladder Carcinoma	1.28	0.11	1.37	0.036	1.35	0.051	1.33	0.06
Breast cancer	1.34	0.073	1.3	0.13	1.31	0.1	1.21	0.27
Head–neck squamous cell carcinoma	1.68	0.00014	1.58	0.001	1.73	$5.00 \times 10^{-5}$	1.59	0.00082
Kidney renal clear cell carcinoma	1.74	$3.00 \times 10^{-4}$	2.48	$7.30 \times 10^{-10}$	1.72	0.00054	2.47	$8.60 \times 10^{-10}$
Kidney renal papillary cell carcinoma	2	0.019	4.41	$2.40 \times 10^{-7}$	2.12	0.012	4.49	$1.60 \times 10^{-7}$
Liver hepatocellular carcinoma	1.72	0.0018	1.9	$3.00 \times 10^{-4}$	1.73	0.0017	1.88	0.00033
Lung adenocarcinoma	1.81	0.00065	1.79	0.00015	1.82	0.00085	1.74	$3.00 \times 10^{-4}$
Lung squamous cell carcinoma	1.17	0.33	1.2	0.24	1.17	0.3	1.22	0.21
Stomach adenocarcinoma	1.83	0.00026	1.92	$9.40 \times 10^{-5}$	1.83	0.00028	1.85	0.00025
Thyroid carcinoma	1.87	0.22	2.06	0.15	1.74	0.28	0.46	0.11
Uterine corpus endometrial carcinoma	1.27	0.25	1.34	0.18	1.33	0.17	1.4	0.12

Multiple genes were selected for analyzing the overall survival in different cancer, sequential addition of a dysregulated gene to the initial gene signature of three genes. Among four gene combinations, “RXRA, SOAT1, SQLE and ACAT1” and “RXRA, SOAT1, SQLE and HSD17B7” showed a high hazard ratio (HR) value with a significant log-rank *p* value in most of the cancers. A five gene combination shows a high HR in cancer types expect for thyroid carcinoma (HR = 0.46).

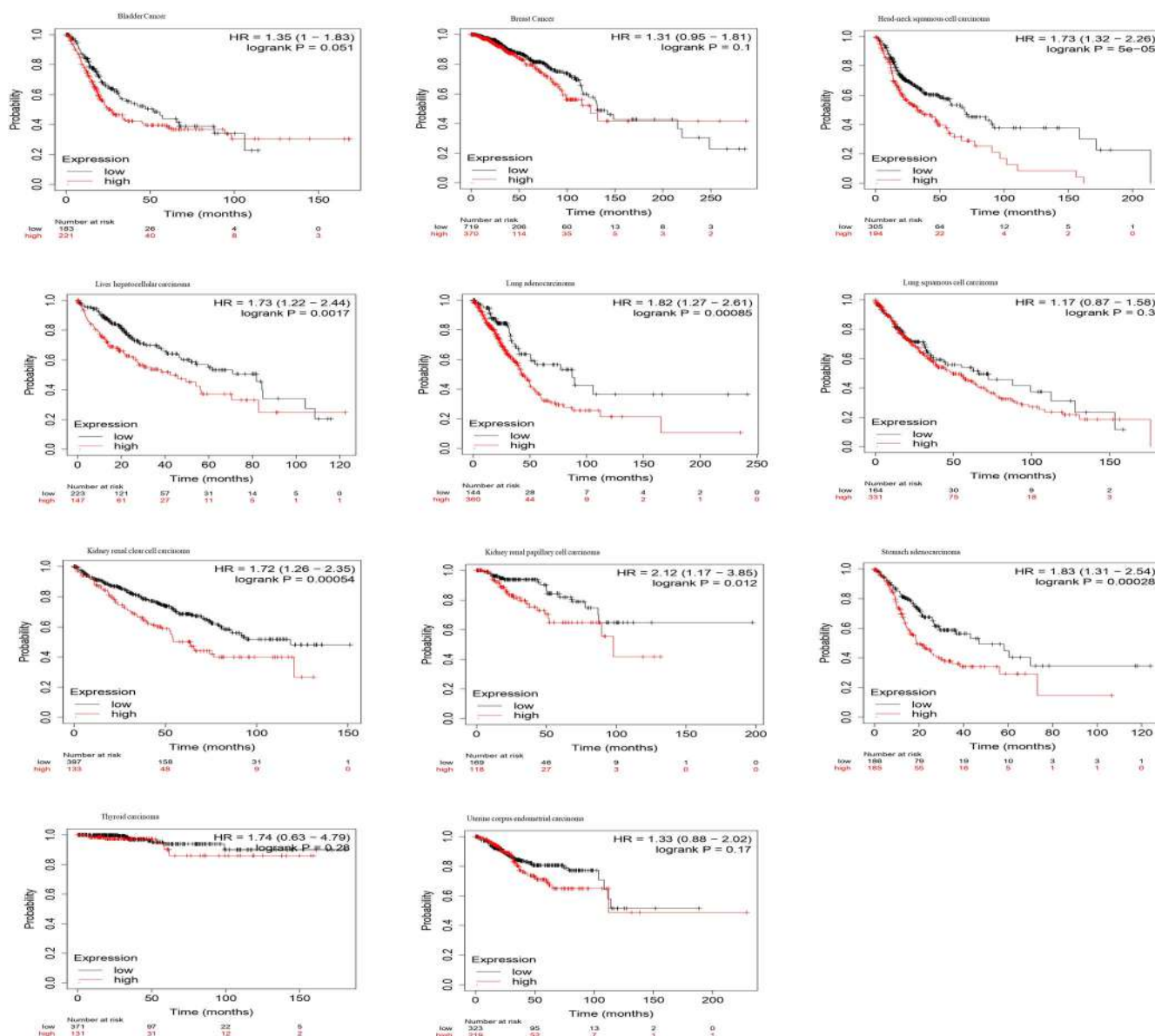


**FIGURE 4** Combination of RXRA, SOAT1, SQLE, and ACAT1 showing poor prognosis in 11 cancer types. Multiple gene selected and analyzed the mean expression of RXRA, SOAT1, SQLE, and ACAT1 to look over the overall survival in different cancer types. The expression of these genes is showing correlation with prognosis in eleven cancer types. RXRA and ACAT1 are placed under inverted and SOAT1 and SQLE under not inverted.  $n$  = number of patients where  $n$  = 405 for bladder carcinoma,  $n$  = 1090 for breast cancer,  $n$  = 500 for head-neck squamous cell carcinoma,  $n$  = 530 for kidney renal clear cell carcinoma,  $n$  = 288 for kidney renal papillary cell carcinoma,  $n$  = 371 for liver hepatocellular carcinoma,  $n$  = 513 for lung adenocarcinoma,  $n$  = 501 for lung squamous cell carcinoma,  $n$  = 375 for stomach adenocarcinoma,  $n$  = 502 for thyroid carcinoma, and  $n$  = 543 for uterine corpus endometrial carcinoma. Higher expression of SQLE and SOAT1 and lower expression of ACAT1 and RXRA are significantly associated with shorter overall survival in all these cancer types.

higher expression of SQLE was also associated with nodal metastasis and higher cancer stages. SOAT1 showed a higher expression in bladder urothelial carcinoma (BLCA), kidney chromophobe (KICH), kidney renal clear cell carcinoma (KIRC) and liver hepatocellular carcinoma (LIHC) based on node  $N_1$  status. These observations further suggest that the gene signature may be linked with a poorer prognosis for cancer patients.

### 3.5 | Experimental validation of the gene signature as a diagnostic marker in breast cancer tissues

To validate the above findings, we analyzed the expression of the gene signature RXRA, SQLE, SOAT1 and ACAT1 in human breast cancer tissue samples. Here, we compared the transcript levels of the above genes between triple negative breast tissue samples and



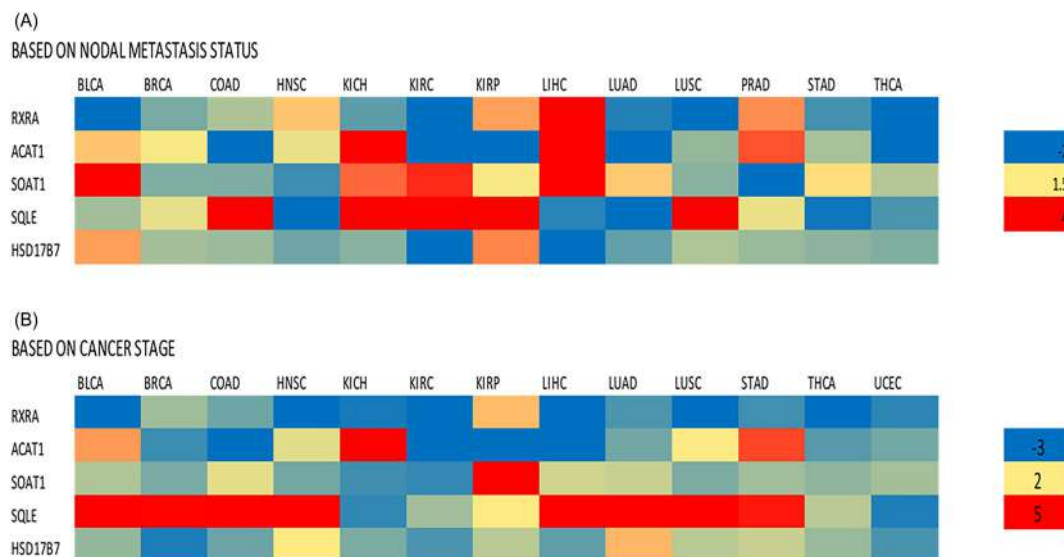
**FIGURE 5** Combination of RXRA, SOAT1, SQLE and HSD17B7 showing poor prognosis in 11 cancer types. Multiple gene selected and analyzed the mean expression of RXRA, SOAT1, SQLE, and HSD17B7 to look over the overall survival in different cancer types. The expression of these genes is showing correlation with prognosis in eleven cancer types. RXRA is placed under inverted and SOAT1, SQLE, and HSD17B7 under not inverted.  $n$  = number of patients where  $n$  = 405 for bladder carcinoma,  $n$  = 1090 for breast cancer,  $n$  = 500 for head-neck squamous cell carcinoma,  $n$  = 530 for kidney renal clear cell carcinoma,  $n$  = 288 for kidney renal papillary cell carcinoma,  $n$  = 371 for liver hepatocellular carcinoma,  $n$  = 513 for lung adenocarcinoma,  $n$  = 501 for lung squamous cell carcinoma,  $n$  = 375 for stomach adenocarcinoma,  $n$  = 502 for thyroid carcinoma, and  $n$  = 543 for uterine corpus endometrial carcinoma. Higher expression of SQLE, SOAT1, and HSD17B7 and lower expression of RXRA are significantly associated with shorter overall survival in all these cancer types.

adjacent control tissues by qRT-PCR. These PCR analyses showed a significant elevation of SOAT1 and SQLE transcripts in triple negative breast tumor tissues compared to adjacent control tissues (Figure 7). However, the transcript levels of RXRA and ACAT1 were found to be significantly downregulated in triple negative breast cancer tissues compared to the adjacent control tissues of breast cancer patients. Moreover, cancer database analysis also found elevated transcripts of SQLE and SOAT 1 and downregulated transcripts of RXRA and ACAT1 genes in triple negative breast cancer subtype compared to other breast cancer subtypes (Table 5). Thus, these experimental

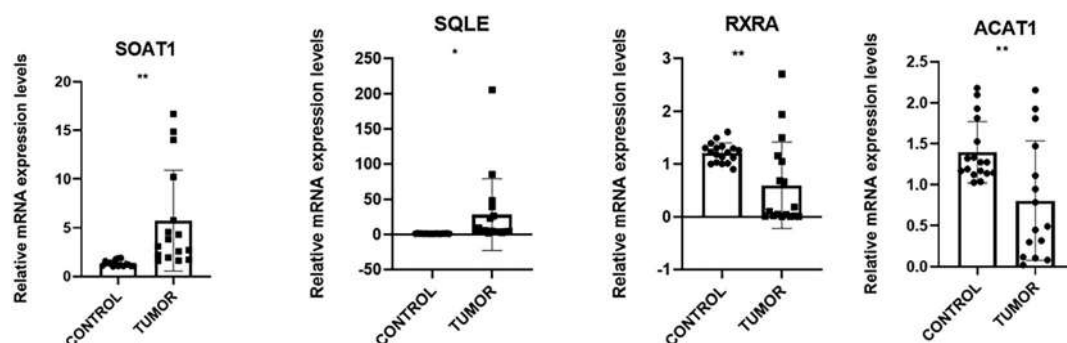
findings tally with the results obtained from the analysis of the cancer databases. All of these data suggest that elevation of SQLE and SOAT1 transcripts, and downregulation of RXRA and ACAT1 mRNAs could serve as diagnostic markers for triple negative breast cancer.

## 4 | DISCUSSION

From the 1980s onward, the association of cholesterol and cancer has been examined in different cohort studies.<sup>35</sup> Cholesterol and its



**FIGURE 6** Heatmap showing change in mRNA expression in nodal metastasis and advance cancer stage tumor of different cancer types. (A) Based on nodal metastasis status, mRNA expression change is the difference of node positive (N1) and node negative (N0); the expression change data are normalized from  $-2$  (blue) to  $4$  (red) in nodal metastasis status. (B) Based on cancer stage, difference between mRNA expression at higher stage to the lower stage and values is normalized from  $-3$  (blue) to  $5$  (red) in higher cancer stage. Here, negative values indicated the decreased in the expression compared with N0/lower cancer stage. BLCA, bladder urothelial carcinoma; BRCA, breast invasive carcinoma; COAD, colon adenocarcinoma; HNSC, head and neck squamous cell carcinoma; KICH, kidney chromophobe; KIRC, kidney renal clear cell carcinoma; KIRP, kidney renal papillary cell carcinoma; LIHC, liver hepatocellular carcinoma; LUAD, lung adenocarcinoma; LUSC, lung squamous cell carcinoma; STAD, stomach carcinoma; THCA, thyroid carcinoma; UCEC, uterine corpus endometrial carcinoma.



**FIGURE 7** Gene expression profiling of (A) RXRA (B) ACAT1 (C) SQLE, and (D) SOAT1 genes in human TNBC tissue samples. The relative mRNA expression of these genes was measured using TB green-based reverse transcription\_quantitative PCR and normalized to  $\beta$ -actin reference gene expression. Results are presented as the fold change in the mRNA expression relative to adjacent nontumor tissue. Experiments were performed with  $\geq 3$  replicates for each condition. Statistical significance was calculated using two-tailed unpaired students t-test. \*P < 0.05, \*\*P < 0.01.

**TABLE 5** Comparative transcript levels of gene signature between subtypes of breast cancer through UALCAN server using the BRCA (breast cancer) cohort.

Gene name	Median expression (transcripts/million)			
	Normal (n = 114)	Luminal (n = 566)	HER2 positive (n = 37)	Triple negative (n = 116)
RXRA	36.71	32.85	24.938	18.332
ACAT1	61.228	41.469	32.102	33.72
SOAT1	14.121	16.094	17.581	21.42
SQLE	12.623	39.591	61.671	55.189
HSD17B7	10.59	21.154	15.239	14.509

n, number of samples present in the cohort.

derivatives have strong affinity toward G protein-coupled receptor, which activates the sonic hedgehog pathway that induces tumor formation.<sup>35</sup> Dysregulated cholesterol homeostasis not only induces the oncogenic pathway, but also associates with inflammation and miRNA-mediated cancer development.<sup>35</sup> A higher LDL-cholesterol level has been reported as a prognostic factor of breast cancer progression.<sup>36</sup> Our previous studies revealed a positive link between cholesterol and cancer mortality in various cancer types.<sup>7</sup> We had noted that simvastatin, a cholesterol lowering drug that targets HMG-CoA reductase, inhibits bone metastasis and breast cancer development in animal models.<sup>22,23</sup> Moreover, simvastatin prevents osteoclastogenic colony stimulating factor-1 to prevent breast cancer-induced osteolytic activity.<sup>21,37</sup> Our recent study has also found elevated level of cholesterol in malignant breast cancer tissues compared to benign breast tissues.<sup>8</sup> The drug metformin and NADA (N-Acyl-Dopamine) prevent cancer proliferation and migration in breast cancer by lowering cellular cholesterol content.<sup>8,24</sup> Thus, the cellular cholesterol pathway appears to be a key therapeutic target for cancer treatment.<sup>38</sup> Other research groups have reported that SQLE, a key rate limiting enzyme of cholesterol biosynthesis pathway, was overexpressed in ductal carcinoma *in situ* and breast cancer tissue compared to normal adjacent tissue.<sup>24</sup> A study reported that ACAT-1 overexpression is associated with poor patient survival in pancreatic cancer.<sup>20</sup> Our current results show the overexpression of ACAT-1 in prostate adenocarcinoma (PRAD) and kidney chromophobe (KICH) cohorts (Figures 2 and 3), whereas other cohorts selected for the present study support the strong downregulation of ACAT-1 gene in cancer tissues. Among the transcription regulation of the cholesterol homeostatic pathway NR1H3, NR1H2, RXRA and RXRB are involved in transcription repression. Our cancer database study shows RXRA to be significantly downregulated in most of the cohorts, except for liver hepatocellular carcinoma (LIHC) and head and neck squamous cell carcinoma (HNSC) and to be strongly associated with patient overall survival (Figure 2 and Table 3). A previous study suggested that HSD17B7 demonstrated high expression in human ductal carcinoma and breast cancer cells, with this providing evidence for the upregulation of this enzyme by estradiol.<sup>39</sup> However, our results give a similar pattern of expression of HSD17B7 in 12 cohort studies. SOAT1 mediates the cholesterol esterification to form cholesterol ester and is stored as lipid droplets.<sup>40</sup> SOAT1 transcripts were found in most of the tissues, whereas SOAT2 is present only in a few tissue types, including liver and small intestine. Studies have revealed that SOAT1 is over expressed in the hepatocellular carcinoma, pancreatic, colon and prostate cancers.<sup>40</sup> Our findings also reveal that SOAT1 is over expressed in most of the cancer types (Figures 2 and 3). In this context, our experimental qRT-PCR analysis found a significant upregulation of both SQLE and SOAT1 mRNA levels in triple negative breast cancer tissues compared to the corresponding adjacent control tissues; however, both RXRA and ACAT1 mRNA levels were significantly decreased in tumor samples (Figure 7). These experimental findings further enhance the assurance of the gene signature (RXRA, SQLE, SOAT1 and ACAT1) as a diagnostic marker.

Statin drugs are used to target the HMG-CoA and reduce the plasma cholesterol level.<sup>41</sup> They are effective in the treatment of cardiovascular, atherosclerosis, dyslipoproteinemia and liver diseases. They are also considered as an anticancer drug and reduce cancer cell viability.<sup>42,43</sup> However, there are some studies that reject the association of statin use with cancer risk.<sup>42,43</sup> In the present study, we found a gene signature that might be a new target for the treatment in cancer because of dysregulation and poor overall survival in most of the cancer types. Along with the statin drug, the drug against gene signature (especially those that are upregulated in the gene signature; i.e., SOAT1, SQLE) might give a better result for suppressing tumor growth and metastasis, and the effectiveness of cholesterol lowering drugs may be enhanced. Indeed, cell specific delivery approaches can be utilized to target the deregulated cholesterol pathway in cancer cells without perturbing the normal cellular cholesterol homeostasis. Therefore, future research must be based on validating the mRNA expression analysis in cancer cell lines of respective cancer types and target specific delivery must reveal new therapeutic options for cancer therapy. Our findings collectively offer novel therapeutic targets for cancer treatment and prevention, and a gene signature in cholesterol regulatory genes is identified that can act as biomarker for cancer diagnosis and prognosis. Moreover, this preliminary study shows that RXRA, SQLE, SOAT1 and ACAT1 could be a generic diagnostic marker in selected 14 cancer types on the basis of cancer data base analysis. The present study did not focus on the particular cancer type in the case of cancer database analysis. In addition, we did not consider the cohort based on age, gender, disease-free state or diseases condition, chemotherapy treatment status, etc. We have experimentally validated the dysregulated expressions of the gene signature (RXRA, SQLE, SOAT1, ACAT1) in triple negative breast cancer. This strong result of “gene signature” needs further experimental validation in different cancer types so that these genes can be considered as diagnostic markers for various cancer types. As a result of the unavailability of sample size, the present study did not reveal whether the concept of gene signature is justified in all the site-specific cancer types. A similar kind of analysis has screened many differentially expressed genes using such bioinformatics analysis and reported that 11 genes are associated with progression and prognosis of endometrial cancer.<sup>44</sup> Another study identified a six gene signature used to predict breast cancer lung metastasis.<sup>45</sup> For the first time, the present study provides a gene signature related to cholesterol metabolism and regulation, acting as a biomarker for diagnosis and prognosis, and this might have potential therapeutic option for cancer therapy.

## 5 | CONCLUSIONS

Dysregulation of cholesterol homeostasis appears to be a key factor in cancer development through activation of various signaling pathways. Our study provides a glimpse of dysregulated gene of cholesterol metabolism and regulation in different cancer types, as well as



their effect on the overall survival of the cancer patients. We also identified a novel gene signature related to cholesterol metabolism that can act as diagnostic and prognostic biomarker in cancer and also as a potential drug target for cancer therapy.

## AUTHOR CONTRIBUTIONS

SK collected data, prepared tables and figures, and drafted the manuscript. SS contributed to collecting and analysing data. AS performed experiments and prepared figures. AM collected clinical samples. LKS collected samples and edited the manuscript. CCM formulated the whole study and the written manuscript.

## ACKNOWLEDGEMENTS

CCM is supported by UGC (30-49/2014 [BSR]), DBT (6242-P9/RGCB/PMD/DBT/CCML/2015), DST-SERB (DST/CRG/2021/002963) and DST-RFBR (INT/RUS/RFBR/P-256). LKS is supported by ICMR EMR-2019-665/CMB/Adhoc/BMS.

## CONFLICT OF INTEREST STATEMENT

The authors declare that they have no conflicts of interest.

## DATA AVAILABILITY STATEMENT

Supporting data for mRNA expression analysis are available in the Supporting information (Table S1).

## ORCID

Chandi C. Mandal  <https://orcid.org/0000-0002-2292-6635>

## REFERENCES

- Gobel A, Rauner M, Hofbauer LC, Rachner TD. Cholesterol and beyond - the role of the mevalonate pathway in cancer biology. *Biochim Biophys Acta Rev Cancer*. 2020;1873(2):188351. doi:10.1016/j.bbcan.2020.188351
- Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. *Cell*. 2011;144(5):646-674. doi:10.1016/j.cell.2011.02.013
- Ding X, Zhang W, Li S, Yang H. The role of cholesterol metabolism in cancer. *Am J Cancer Res*. 2019;9(2):219-227.
- Lee JJ, Loh K, Yap YS. PI3K/Akt/mTOR inhibitors in breast cancer. *Cancer Biol Med*. 2015;12(4):342-354. doi:10.7497/j.issn.2095-3941.2015.0089
- Murillo-Garzon V, Kypka R. WNT signalling in prostate cancer. *Nat Rev Urol*. 2017;14(11):683-696. doi:10.1038/nrurol.2017.144
- Kuzu OF, Noory MA, Robertson GP. The role of cholesterol in cancer. *Cancer Res*. 2016;76(8):2063-2070. doi:10.1158/0008-5472.CAN-15-2613
- Mandal CC, Sharma A, Panwar MS, Radosevich JA. Is cholesterol a mediator of cold-induced cancer? *Tumour Biol*. 2016;37(7):9635-9648. doi:10.1007/s13277-016-4799-2
- Sharma A, Bandyopadhyaya S, Chowdhury K, et al. Metformin exhibited anticancer activity by lowering cellular cholesterol content in breast cancer cells. *PLoS One*. 2019;14(1):e0209435. doi:10.1371/journal.pone.0209435
- Huang B, Song BL, Xu C. Cholesterol metabolism in cancer: mechanisms and therapeutic opportunities. *Nat Metab*. 2020;2(2):132-141. doi:10.1038/s42255-020-0174-0
- Ma S, Sun W, Gao L, Liu S. Therapeutic targets of hypercholesterolemia: HMGCR and LDLR. *Diab Metab Syndrome Obes: Targets Ther*. 2019;12:1543-1553. doi:10.2147/DMSO.S219013
- Luo J, Yang H, Song B-L. Mechanisms and regulation of cholesterol homeostasis. *Nat Rev Mol Cell Biol*. 2020;21(4):225-245. doi:10.1038/s41580-019-0190-7
- Paolicelli RC, Widmann C. Squalene: friend or foe for cancers. *Curr Opin Lipidol*. 2019;30(4):353-354. doi:10.1097/MOL.0000000000000619
- Xu H, Zhou S, Tang Q, Xia H, Bi F. Cholesterol metabolism: new functions and therapeutic approaches in cancer. *Biochim Biophys Acta Rev Cancer*. 2020;1874(1):188394. doi:10.1016/j.bbcan.2020.188394
- Goldstein JL, DeBose-Boyd RA, Brown MS. Protein sensors for membrane sterols. *Cell*. 2006;124(1):35-46. doi:10.1016/j.cell.2005.12.022
- Ikonen E. Cellular cholesterol trafficking and compartmentalization. *Nat Rev Mol Cell Biol*. 2008;9(2):125-138. doi:10.1038/nrm2336
- Relas H, Gylling H, Miettinen TA. Dietary squalene increases cholesterol synthesis measured with serum non-cholesterol sterols after a single oral dose in humans. *Atherosclerosis*. 2000;152(2):377-383. doi:10.1016/S0021-9150(99)00478-5
- Cruz PM, Mo H, McConathy WJ, Sabnis N, Lacko AG. The role of cholesterol metabolism and cholesterol transport in carcinogenesis: a review of scientific findings, relevant to future cancer therapeutics. *Front Pharmacol*. 2013;4:119. doi:10.3389/fphar.2013.00119
- Brown MS, Goldstein JL. A receptor-mediated pathway for cholesterol homeostasis. *Science*. 1986;232(4746):34-47. doi:10.1126/science.3513311
- Simons K, Ikonen E. How cells handle cholesterol. *Science*. 2000;290(5497):1721-1726. doi:10.1126/science.290.5497.1721
- Li J, Gu D, Lee SS, et al. Abrogating cholesterol esterification suppresses growth and metastasis of pancreatic cancer. *Oncogene*. 2016;35(50):6378-6388. doi:10.1038/onc.2016.168
- Chowdhury K, Sharma A, Sharma T, Kumar S, Mandal CC. Simvastatin and MBDC inhibit breast cancer-induced osteoclast activity by targeting Osteoclastogenic factors. *Cancer Invest*. 2017;35(6):403-413. doi:10.1080/07357907.2017.1309548
- Ghosh-Choudhury N, Mandal CC, Ghosh-Choudhury N, Ghosh Choudhury G. Simvastatin induces derepression of PTEN expression via NFκB to inhibit breast cancer cell growth. *Cell Signal*. 2010;22(5):749-758. doi:10.1016/j.cellsig.2009.12.010
- Mandal CC, Ghosh-Choudhury N, Yoneda T, Choudhury GG, Ghosh-Choudhury N. Simvastatin prevents skeletal metastasis of breast cancer by an antagonistic interplay between p53 and CD44. *J Biol Chem*. 2011;286(13):11314-11327. doi:10.1074/jbc.M110.193714
- Bandyopadhyaya S, Akimov MG, Verma R, et al. N-acridonoyl dopamine inhibits epithelial-mesenchymal transition of breast cancer cells through ERK signaling and decreasing the cellular cholesterol. *J Biochem Mol Toxicol*. 2021;35(4):e22693. doi:10.1002/jbt.22693
- Rye MB, Bertilsson H, Andersen MK, et al. Cholesterol synthesis pathway genes in prostate cancer are transcriptionally downregulated when tissue confounding is minimized. *BMC Cancer*. 2018;18(1):478. doi:10.1186/s12885-018-4373-y
- Fabregat A, Sidiropoulos K, Viteri G, et al. Reactome diagram viewer: data structures and strategies to boost performance. *Bioinformatics*. 2018;34(7):1208-1214. doi:10.1093/bioinformatics/btx752
- Jassal B, Matthews L, Viteri G, et al. The reactome pathway knowledgebase. *Nucleic Acids Res*. 2020;48(D1):D498-d503. doi:10.1093/nar/gkz1031
- Chandrashekar DS, Bashel B, Balasubramanya SAH, et al. UALCAN: a portal for facilitating tumor subgroup gene expression and survival analyses. *Neoplasia*. 2017;19(8):649-658. doi:10.1016/j.neo.2017.05.002
- Deng M, Braegelmann J, Kryukov I, Saraiva-Agostinho N, Perner S. FirebrowseR: an R client to the broad Institute's firehose pipeline. *Database: J Biol Databases Curation*. 2017;2017:2017. doi:10.1093/database/baw160



30. Cheema PK, Burkes RL. Overall survival should be the primary end-point in clinical trials for advanced non-small-cell lung cancer. *Curr Oncol (Toronto, Ont)*. 2013;20(2):e150-e160. doi:[10.3747/co.20.1226](https://doi.org/10.3747/co.20.1226)
31. Cerami E, Gao J, Dogrusoz U, et al. The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. *Cancer Discov*. 2012;2(5):401-404. doi:[10.1158/2159-8290.CD-12-0095](https://doi.org/10.1158/2159-8290.CD-12-0095)
32. Gao J, Aksoy BA, Dogrusoz U, et al. Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Sci Signal*. 2013;6(269):pl1. doi:[10.1126/scisignal.2004088](https://doi.org/10.1126/scisignal.2004088)
33. Nagy A, Munkacsy G, Gyorffy B. Pancancer survival analysis of cancer hallmark genes. *Sci Rep*. 2021;11(1):6047. doi:[10.1038/s41598-021-84787-5](https://doi.org/10.1038/s41598-021-84787-5)
34. Schmittgen TD, Livak KJ. Analyzing real-time PCR data by the comparative C(T) method. *Nat Protoc*. 2008;3(6):1101-1108. doi:[10.1038/nprot.2008.73](https://doi.org/10.1038/nprot.2008.73)
35. Mok EHK, Lee TKW. The pivotal role of the dysregulation of cholesterol homeostasis in cancer: implications for therapeutic targets. *Cancers (Basel)*. 2020;12(6):1410. doi:[10.3390/cancers12061410](https://doi.org/10.3390/cancers12061410)
36. Rodrigues dos Santos C, Fonseca I, Dias S, Mendes de Almeida JC. Plasma level of LDL-cholesterol at diagnosis is a predictor factor of breast tumor progression. *BMC Cancer*. 2014;14(1):132. doi:[10.1186/1471-2407-14-132](https://doi.org/10.1186/1471-2407-14-132)
37. Mandal CC. Osteolytic metastasis in breast cancer: effective prevention strategies. *Expert Rev Anticancer Ther*. 2020;20(9):797-811. doi:[10.1080/14737140.2020.1807950](https://doi.org/10.1080/14737140.2020.1807950)
38. Mandal CC, Rahman MM. Targeting intracellular cholesterol is a novel therapeutic strategy for cancer treatment. *J Cancer Sci Ther*. 2014;6(12):510-513. doi:[10.4172/1948-5956.1000316](https://doi.org/10.4172/1948-5956.1000316)
39. Shehu A, Albarracin C, Devi YS, et al. The stimulation of HSD17B7 expression by estradiol provides a powerful feed-forward mechanism for estradiol biosynthesis in breast cancer cells. *Mol Endocrinol*. 2011;25(5):754-766. doi:[10.1210/me.2010-0261](https://doi.org/10.1210/me.2010-0261)
40. Xu H, Xia H, Zhou S, Tang Q, Bi F. Cholesterol activates the Wnt/PCP-YAP signaling in SOAT1-targeted treatment of colon cancer. *Cell Death Discov*. 2021;7(1):38. doi:[10.1038/s41420-021-00421-3](https://doi.org/10.1038/s41420-021-00421-3)
41. Boudreau DM, Yu O, Johnson J. Statin use and cancer risk: a comprehensive review. *Expert Opin Drug Saf*. 2010;9(4):603-621. doi:[10.1517/14740331003662620](https://doi.org/10.1517/14740331003662620)
42. Jacobs EJ, Rodriguez C, Brady KA, Connell CJ, Thun MJ, Calle EE. Cholesterol-lowering drugs and colorectal cancer incidence in a large United States cohort. *J Natl Cancer Inst*. 2006;98(1):69-72. doi:[10.1093/jnci/djj006](https://doi.org/10.1093/jnci/djj006)
43. Vinogradova Y, Hippisley-Cox J, Coupland C, Logan RF. Risk of colorectal cancer in patients prescribed statins, nonsteroidal anti-inflammatory drugs, and cyclooxygenase-2 inhibitors: nested case-control study. *Gastroenterology*. 2007;133(2):393-402. doi:[10.1053/j.gastro.2007.05.023](https://doi.org/10.1053/j.gastro.2007.05.023)
44. Liu J, Zhou S, Li S, et al. Eleven genes associated with progression and prognosis of endometrial cancer (EC) identified by comprehensive bioinformatics analysis. *Cancer Cell Int*. 2019;19(1):136. doi:[10.1186/s12935-019-0859-1](https://doi.org/10.1186/s12935-019-0859-1)
45. Landemaine T, Jackson A, Bellahcène A, et al. A six-gene signature predicting breast cancer lung metastasis. *Cancer Res*. 2008;68(15):6092-6099. doi:[10.1158/0008-5472.CAN-08-0436](https://doi.org/10.1158/0008-5472.CAN-08-0436)
46. Cheng X, Li J, Guo D. SCAP/SREBPs are central players in lipid metabolism and novel metabolic targets in cancer therapy. *Curr Top Med Chem*. 2018;18(6):484-493. doi:[10.2174/1568026618666180523104541](https://doi.org/10.2174/1568026618666180523104541)
47. Ye J, DeBose-Boyd RA. Regulation of cholesterol and fatty acid synthesis. *Cold Spring Harb Perspect Biol*. 2011;3(7):a004754. doi:[10.1101/cshperspect.a004754](https://doi.org/10.1101/cshperspect.a004754)

## SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

**How to cite this article:** Kuldeep S, Soni S, Srivastava A, Mishra A, Sharma LK, Mandal CC. Dysregulated cholesterol regulatory genes as a diagnostic biomarker for cancer. *J Gene Med*. 2023;e3475. doi:[10.1002/jgm.3475](https://doi.org/10.1002/jgm.3475)

# Thrombospondin 2 is a key determinant of fibrogenesis in non-alcoholic fatty liver disease

Takefumi Kimura<sup>1,2</sup>  | Takanobu Iwadare<sup>1</sup> | Shun-ichi Wakabayashi<sup>1</sup> | Seema Kuldeep<sup>3</sup> | Tomoyuki Nakajima<sup>4</sup> | Tomoo Yamazaki<sup>1,5</sup> | Daiki Aomura<sup>6</sup> | Hamim Zafar<sup>7</sup> | Mai Iwaya<sup>4</sup> | Satoru Joshita<sup>1</sup> | Takeshi Uehara<sup>4</sup> | Sai P. Pydi<sup>3</sup> | Naoki Tanaka<sup>8,9,10</sup> | Takeji Umemura<sup>1,2</sup> 

<sup>1</sup>Department of Medicine, Division of Gastroenterology and Hepatology, Shinshu University School of Medicine, Matsumoto, Japan

<sup>2</sup>Consultation Center for Liver Diseases, Shinshu University Hospital, Matsumoto, Japan

<sup>3</sup>Department of Biological Sciences and Bioengineering, Indian Institute of Technology, Kanpur, India

<sup>4</sup>Department of Laboratory Medicine, Shinshu University School Hospital, Matsumoto, Japan

<sup>5</sup>Department of Medicine, University of California San Diego, San Diego, La Jolla, USA

<sup>6</sup>Department of Medicine, Division of Nephrology, Shinshu University School of Medicine, Matsumoto, Japan

<sup>7</sup>Department of Computer Science and Engineering and Biological Sciences and Bioengineering, Indian Institute of Technology, Kanpur, India

<sup>8</sup>Department of Global Medical Research Promotion, Shinshu University Graduate School of Medicine, Matsumoto, Japan

<sup>9</sup>International Relations Office, Shinshu University School of Medicine, Matsumoto, Japan

<sup>10</sup>Research Center for Social Systems, Shinshu University, Matsumoto, Japan

## Correspondence

Takefumi Kimura, Department of Medicine, Division of Gastroenterology and Hepatology, Shinshu University School of Medicine, Asahi 3-1-1, Matsumoto, Nagano 390-8621, Japan.  
Email: [kimuratakefumi@yahoo.co.jp](mailto:kimuratakefumi@yahoo.co.jp); [t\\_kimura@shinshu-u.ac.jp](mailto:t_kimura@shinshu-u.ac.jp)

## Funding information

AMED, Grant/Award Number: JP23fk0210125; JSPS KAKENHI, Grant/Award Number: JP22K20884; Aiba Works Medical Research; Indian Council of Medical Research, Grant/Award Number: 5/4/8-18/Obs/SPP/2022-NCD-II

Handling Editor: Luca Valenti

## Abstract

**Objective:** Hepatic overexpression of the thrombospondin 2 gene (THBS2) and elevated levels of circulating thrombospondin 2 (TSP2) have been observed in patients with chronic liver disease. This study aimed to identify the specific cells expressing THBS2/TSP2 in non-alcoholic fatty liver disease (NAFLD) and investigate the underlying mechanism behind THBS2/TSP2 upregulation.

**Design:** Comprehensive NAFLD liver gene datasets, including single-cell RNA sequencing (scRNA-seq), in-house NAFLD liver tissue, and LX-2 cells derived from human hepatic stellate cells (HSCs), were analysed using a combination of computational biology, genetic, immunological, and pharmacological approaches.

**Results:** Analysis of the genetic dataset revealed the presence of 1433 variable genes in patients with advanced fibrosis NAFLD, with THBS2 ranked among the top 2 genes. Quantitative polymerase chain reaction (qPCR) examination of NAFLD livers showed a significant correlation between THBS2 expression and fibrosis stage ( $r=.349$ ,

**Abbreviations:** COL1A1, Collagen type I alpha 1; COL1A2, Collagen type I alpha 2; COL3A1, Collagen type III alpha 1; COL4A1, Collagen type IV alpha 1; DAPI, 4',6-diamidino-2-phenylindole; DEGs, differentially expressed genes; ECM, extracellular matrix; GEO, gene expression omnibus; HSCs, hepatic stellate cells; NAFLD, non-alcoholic fatty liver disease; NASH, non-alcoholic steatohepatitis; PBS, phosphate-buffered saline; qPCR, quantitative polymerase chain reaction; scRNA-seq, single-cell RNA sequencing; SMAD, small mothers against decapentaplegic; TGF $\beta$ , transforming growth factor beta; THBS2, Thrombospondin 2 gene; TSP2, Thrombospondin 2.

Takefumi Kimura and Takanobu Iwadare contributed equally.

This is an open access article under the terms of the [Creative Commons Attribution-NonCommercial-NoDerivs](https://creativecommons.org/licenses/by-nc-nd/4.0/) License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

© 2023 The Authors. *Liver International* published by John Wiley & Sons Ltd.



$p < .001$ ). In support of this, scRNA-seq data and in situ hybridization demonstrated that the THBS2 gene was highly expressed in HSCs of NAFLD patients with advanced fibrosis. Pathway analysis of the gene dataset revealed THBS2 expression to be associated with the transforming growth factor beta (TGF $\beta$ ) pathway and collagen gene activation. Moreover, the activation of LX-2 cells with TGF $\beta$  increased THBS2/TSP2 and collagen expression independently of the TGF $\beta$ -SMAD2/3 pathway. THBS2 gene knockdown significantly decreased collagen expression in LX-2 cells.

**Conclusions:** THBS2/TSP2 is highly expressed in HSCs and plays a role in regulating fibrogenesis in NAFLD patients. THBS2/TSP2 may therefore represent a potential target for anti-fibrotic therapy in NAFLD.

#### KEYWORDS

Fibrosis, NAFLD, THBS2, Thrombospondin 2, TSP2

## 1 | INTRODUCTION

Non-alcoholic fatty liver disease (NAFLD) is a global health concern<sup>1,2</sup> that encompasses a spectrum of conditions, ranging from non-alcoholic fatty liver to non-alcoholic steatohepatitis (NASH).<sup>3</sup> NASH is characterized by the accumulation of fat in the liver accompanied by inflammation and scarring, which can lead to cirrhosis and liver cancer.<sup>4</sup> Several large clinical trials have suggested that the degree of liver fibrosis is closely related to prognosis in NAFLD.<sup>5-7</sup> However, the mechanisms underlying the progression of liver fibrosis are not fully understood, and effective therapeutic strategies for NAFLD have yet to be established.<sup>8,9</sup>

Thrombospondins are a group of proteins that are produced and secreted by various cells.<sup>10,11</sup> The glycoproteins are characterized by multiple domains and exhibit calcium-binding capability.<sup>12</sup> Thrombospondins interact with a wide range of substances in the body, including cytokines, growth factors, receptors, and components of the extracellular matrix (ECM).<sup>13</sup> Of the five different types of thrombospondins, thrombospondin 2 (TSP2) stands out as a particularly unique member. TSP2 is encoded by the thrombospondin 2 (THBS2) gene and is involved in such processes as fibrin formation, bone growth, maintenance of normal blood vessel density, blood clotting, and cell adhesion.<sup>14</sup> In the skin, TSP2 is mainly produced by fibroblasts and smooth muscle cells and is believed to play a role in the process of wound repair and tissue remodelling.<sup>14</sup> Recently, several studies have reported that THBS2 is upregulated in the fibrotic liver in NAFLD, with secreted TSP2 emerging as a potential biomarker for disease progression.<sup>15,16</sup> These results were supported by a large investigation of NAFLD patients with diabetes mellitus or metabolic-associated fatty liver disease.<sup>17,18</sup> Furthermore, a correlation between the degree of fibrosis and inflammation and serum TSP2 levels was observed not only in NAFLD, but also in hepatitis C virus-infected patients.<sup>19,20</sup> Despite these advances, however, the specific liver cell populations responsible for THBS2/TSP2 expression in NAFLD, the

#### Key points

This study identified the elevated expression of the thrombospondin 2 gene (THBS2) in hepatic stellate cells of NAFLD patients with advanced fibrosis. THBS2 appears to play a crucial role in promoting fibrogenesis in NAFLD, suggesting its potential as a therapeutic target for anti-fibrotic interventions.

underlying mechanisms governing expression dynamics, and the precise role of THBS2/TSP2 in the pathogenesis of NAFLD remain largely unknown.

To address the above issues, we used comprehensive genetic data, single-cell RNA sequencing (scRNA-seq) datasets, in situ hybridization data, and in vitro models to identify THBS2/TSP2-expressing cells in NAFLD, elucidate their expression mechanisms, and investigate their potential therapeutic applications. We go on to show that THBS2/TSP2 is highly expressed in hepatic stellate cells (HSCs) and plays a role in regulating the fibrotic process in patients with NAFLD.

## 2 | MATERIALS AND METHODS

### 2.1 | Liver mRNA data collection and processing

The mRNA expression data from the Gene Expression Omnibus (GEO) database was downloaded through a microarray dataset (GSE49541) for processing by GEO2R and incorporation into this study.<sup>21,22</sup> The data compared the expression of various human genes at different liver fibrosis stages in NAFLD patients. Differentially expressed genes were screened from the GEO dataset with a threshold of  $|\log_2\text{FC}| > 1$  and adjusted  $p < .05$ .<sup>19</sup>

## 2.2 | Pathway enrichment analysis

The GSE49541 dataset was subjected to enrichment and network analysis using Metascape (<https://Metascape.org/>).<sup>21–23</sup> Metascape is a web-based portal designed to provide a comprehensive gene list annotation and analysis resource for biologists.<sup>23</sup> To gain insights into the biological roles of identified differentially expressed genes (DEGs) and differentially expressed proteins (DEPs), we conducted pathway enrichment analysis of Gene Ontology Biological Process, Kyoto Encyclopedia of Genes and Genomes, Reactome, and Canonical pathway in Metascape tools.<sup>23–26</sup> By inputting the lists of DEGs and DEPs simultaneously, Metascape can identify commonly enriched and selectively enriched pathways from two levels, which enables a comprehensive assessment of the molecular features of the biological process.

## 2.3 | scRNA-seq dataset processing

In this study, the scRNA-seq datasets GSE174748 and GSE189175<sup>27,28</sup> were analysed using the Seurat R toolkit 4.5. Initially, quality control and filtering were performed to remove lower quality cells (>10% mitochondrial genes), followed next by normalization and scaling. Based on the information provided, condition, sex, and tissue type were annotated in the metadata, which was later used in downstream analysis. After quality control and normalization, the datasets were merged using the canonical correlation analysis function and subsequently integrated using harmony to remove any batch effects. Upon integration, the final dataset was subjected to normalization and scaling once again. The top 2000 variable feature was employed using Seurat's built-in function to construct principal components, with the top 15 principal components used for clustering and manifold approximation and projection for dimension reduction visualization. Clusters were annotated using known canonical cell-specific markers (Table S1).

## 2.4 | Liver samples from NAFLD patients and histological findings

This study included liver tissue samples from 96 biopsy-proven Japanese NAFLD patients admitted to Shinshu University Hospital (Matsumoto, Japan) between 2015 and 2018. The clinical data of the patients are summarized in Table 1. This study was reviewed and approved by the Institutional Review Board of Shinshu University Hospital (Matsumoto, Japan) (approval number: 3021), and written informed consent was obtained from all participating subjects. The investigation was conducted according to the principles of the Declaration of Helsinki. Liver specimens of at least 1.5 cm in length were obtained from segments 5 or 8 using a 14-gauge needle as described previously and immediately fixed in 10% neutral formalin.<sup>29</sup> Sections of 4 µm in thickness were cut and stained using the haematoxylin and eosin and Azan–Mallory methods. The histological

**TABLE 1** Clinico-pathological background of patients from own institution used for qPCR studies.

	All (n = 96)
	Median (IQR)/n (%)
Age (years)	56 (37–65)
Male	50 (52%)
Body mass index (kg/m <sup>2</sup> )	26.1 (24–30)
Laboratory data	
Albumin (mg/dL)	4.6 (4.3–4.8)
Bilirubin (mg/dL)	0.91 (0.71–1.18)
AST (U/L)	44 (29–64)
ALT (U/L)	62 (19–273)
GGTP (IU/L)	50 (36–86)
Cholinesterase (U/L)	375 (298–433)
Fasting glucose (mg/dL)	106 (95–116)
Insulin (µU/mL)	13.5 (8.9–26)
HOMA-IR	3.7 (2.4–7.6)
Haemoglobin A1c (%)	5.6 (5.2–6.0)
Total cholesterol (mg/dL)	209 (179–237)
LDL cholesterol (mg/dL)	130 (118–154)
HDL cholesterol (mg/dL)	49 (43–57)
Triglycerides (mg/dL)	134 (97–176)
Alpha-fetoprotein (ng/mL)	2.6 (1.9–3.6)
Platelets (×10 <sup>4</sup> /µL)	20.4 (16.0–25.6)
APRI	0.67 (0.46–1.28)
FIB-4	1.52 (0.76–2.5)
Pathology	
NAFLD activity score	5 (3–5)
Fibrosis stage (0/1/2/3/4)	15/44/12/19/6

activity of NAFLD was assessed by an independent expert pathologist in a blinded manner according to the NAFLD scoring system.<sup>30</sup> Fibrosis stage was scored as follows: F0, none; F1, perisinusoidal or periportal; F2, perisinusoidal and portal/periportal; F3, bridging fibrosis; and F4, cirrhosis.<sup>30</sup>

## 2.5 | THBS2 RNA in situ hybridization

Detection of THBS2 mRNA was performed using an RNAscope kit (Cosmo Bio), according to the manufacturer's instructions using unstained tissue sections as described in a previous report.<sup>31</sup> Brown punctate dots in the nucleus and/or cytoplasm indicated positive staining.<sup>31</sup>

## 2.6 | Quantitative reverse-transcription polymerase chain reaction analysis of gene expression

Total RNA was extracted from frozen tissues or cultured cells using the RNeasy Mini Kit (Qiagen). SuperScript III First-Strand



Synthesis SuperMix (Invitrogen) was used to prepare cDNA. The SYBR Green method (Applied Biosystems) was employed for qPCR studies.<sup>32</sup>  $\Delta\Delta C_t$  was used to normalize the gene expression data collected in this study relative to the expression of GAPDH.<sup>33</sup> The PCR primers used in this study are listed in Table S2.

## 2.7 | Western blotting studies

LX-2 cells were lysed and proteins were extracted using radioimmunoprecipitation assay buffer supplemented with complete EDTA-free protease inhibitor cocktail (Sigma-Aldrich). After centrifugation of cell lysates at 12000g for 10 min, protein concentrations were determined by means of a BCA protein assay kit (Pierce). Subsequently, protein samples were denatured at 95°C using NuPAGE LDS sample buffer (Thermo Fisher Scientific) and separated by 4%–12% sodium dodecyl sulphate–polyacrylamide gel electrophoresis. The proteins were then transferred to nitrocellulose membranes and incubated overnight at 4°C with primary antibodies. On the next day, the membranes were washed thoroughly and then incubated with horseradish peroxidase-conjugated anti-rabbit or anti-goat secondary antibodies, followed by visualization of the separated protein bands using SuperSignal West Pico Chemiluminescent Substrate (Thermo Fisher Scientific) on a c600 Imaging System Imager (Azure Biosystems).<sup>34</sup> Immunoreactive bands were quantified using ImageJ Software (NIH).<sup>35</sup>

The following antibodies (source, catalogue #, and dilution are indicated) were used: TSP2 (Abcam, #ab112543, 1:1000), small mothers against decapentaplegic (SMAD) 2/3 (D7/G7) (Cell Signaling Technology, #8685, 1:1000), Phospho-Smad2 (Ser465/467)/Smad3 (Ser423/425) (D27/F4) (Cell Signaling Technology, #8828, 1:1000), GAPDH (D16/H11) (Cell Signaling Technology, #5174, 1:1000), anti-rabbit IgG, and HRP-linked antibody (Cell Signaling Technology, #7074, 1:2000).

## 2.8 | Cells

LX-2 cells were obtained from Sigma-Aldrich, and HepG2 cells were procured from Cellular Engineering Technologies. Both cell lines were cultured in Dulbecco's Modified Eagle Medium with 2% foetal bovine serum.<sup>36,37</sup>

## 2.9 | siRNA-mediated knockdown of THBS2 expression in LX-2 cells

Human THBS2 siRNA and control siRNA were purchased from Santa Cruz. On day 0, LX-2 cells were transfected with siRNA using Lipofectamine RNAiMAX (Thermo Fisher Scientific) according to the manufacturer's instructions.<sup>35</sup> Four hours later, the cells

were transferred to normal medium and used for further experiments on day 3.

## 2.10 | Immunocytochemistry

LX-2 cells were plated on glass coverslips (Thermo Fisher Scientific) in 12-well culture dishes and grown to approximately 50% confluence for 3 days to promote cell adherence. The cells were then washed twice with cold serum-free medium and then fixed in 4% paraformaldehyde in phosphate-buffered saline (PBS) for 10 min. After fixation, the cells were washed twice with PBS followed by permeabilization with PBS containing 0.1% Triton X-100 for 15 min. The cells were next washed twice with PBS and incubated with blocking solution (5% BSA in PBS) for 30 min. Primary antibodies (collagen type I obtained from Cosmo Bio, diluted 1:100 in blocking solution) were incubated with the cells for 1 h. After three washes with 0.2% Tween 20 in PBS (PBST), the cells were incubated with fluorescein-labelled anti-mouse IgG (Vector Laboratories, diluted 1:200 in blocking buffer) for 1 h. The cells were washed three times with PBST, stained with DAPI (4',6-diamidino-2-phenylindole molecular probes, diluted 1:5000 in PBS) for 1 min, and then washed three times with PBST and twice with PBS. The cells were viewed with a Nikon Eclipse E600 fluorescence microscope.<sup>38</sup> ImageJ technology (NIH) was employed for the quantification of collagen.<sup>39</sup>

## 2.11 | Compounds and enzyme-linked immunosorbent assays

The recombinant human TGF $\beta$  protein and recombinant human TSP2 protein were obtained from R&D Systems. The SMAD3 Inhibitor SIS3 was procured from Cayman Chemical Company. Compound concentrations and treatment times are shown in each figure legend. TSP2 concentrations in LX-2 cell lysates were determined using enzyme-linked immunosorbent assays following the manufacturer's instructions (Quantikine® ELISA, #DTSP20, R&D Systems).<sup>16</sup> The cell lysate method for LX-2 cells was identical to that for western blotting.

## 2.12 | Statistics

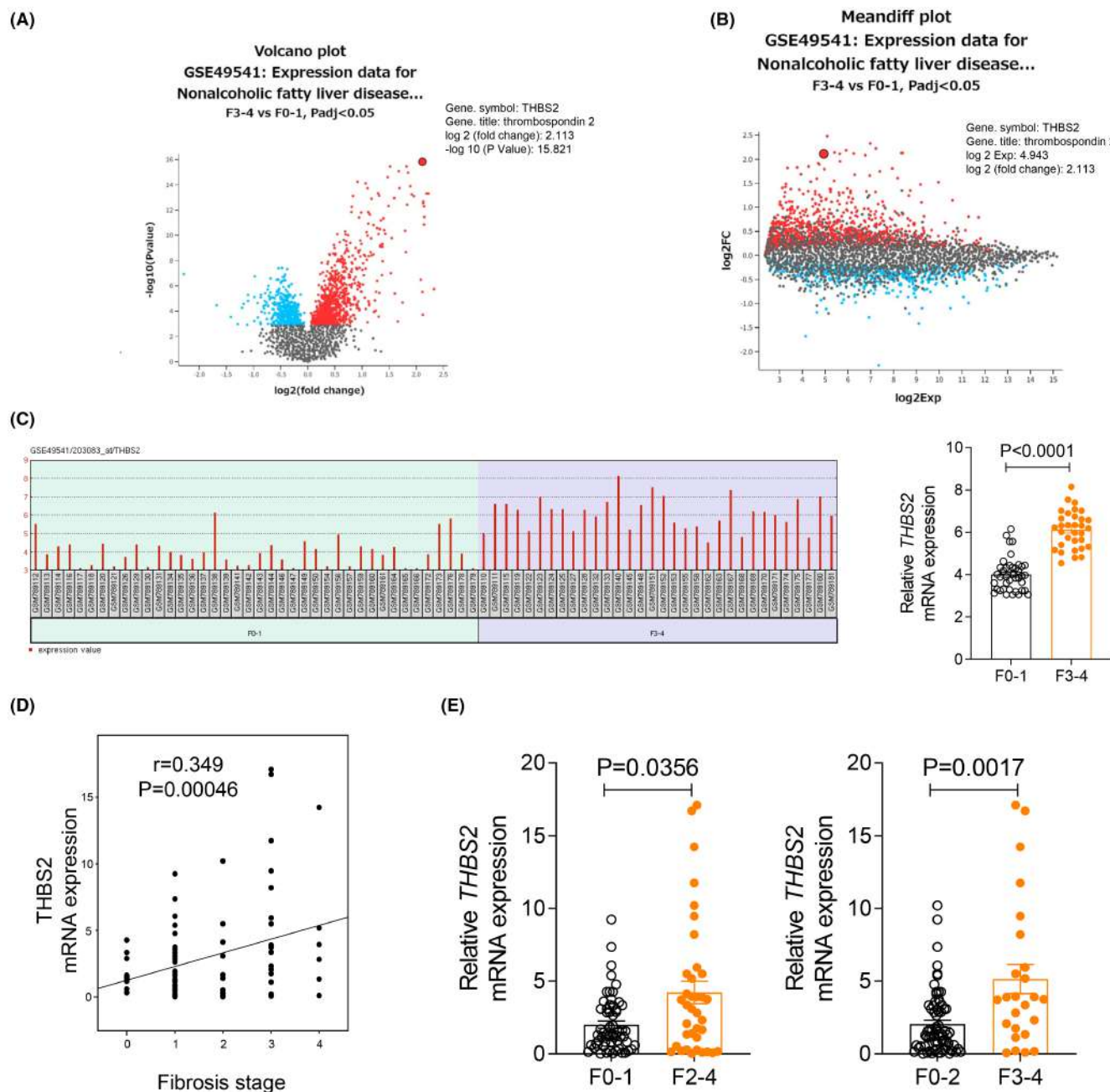
Data were collected and analysed using Prism 8 (GraphPad) and StatFlex Ver. 7.0. All data are expressed as the mean  $\pm$  standard error of the mean for the indicated number of observations. Prior to the specific statistical tests, we performed testing for normality and homogeneity of variance. The data were then evaluated for statistical significance by one-way ANOVA followed by the indicated post-hoc test or two-tailed unpaired Student's *t* test or Mann–Whitney *U* test, as appropriate. Correlation analysis was conducted by Spearman's test or Kendall's tau test. A *p* < .05 was considered statistically significant.

### 3 | RESULTS

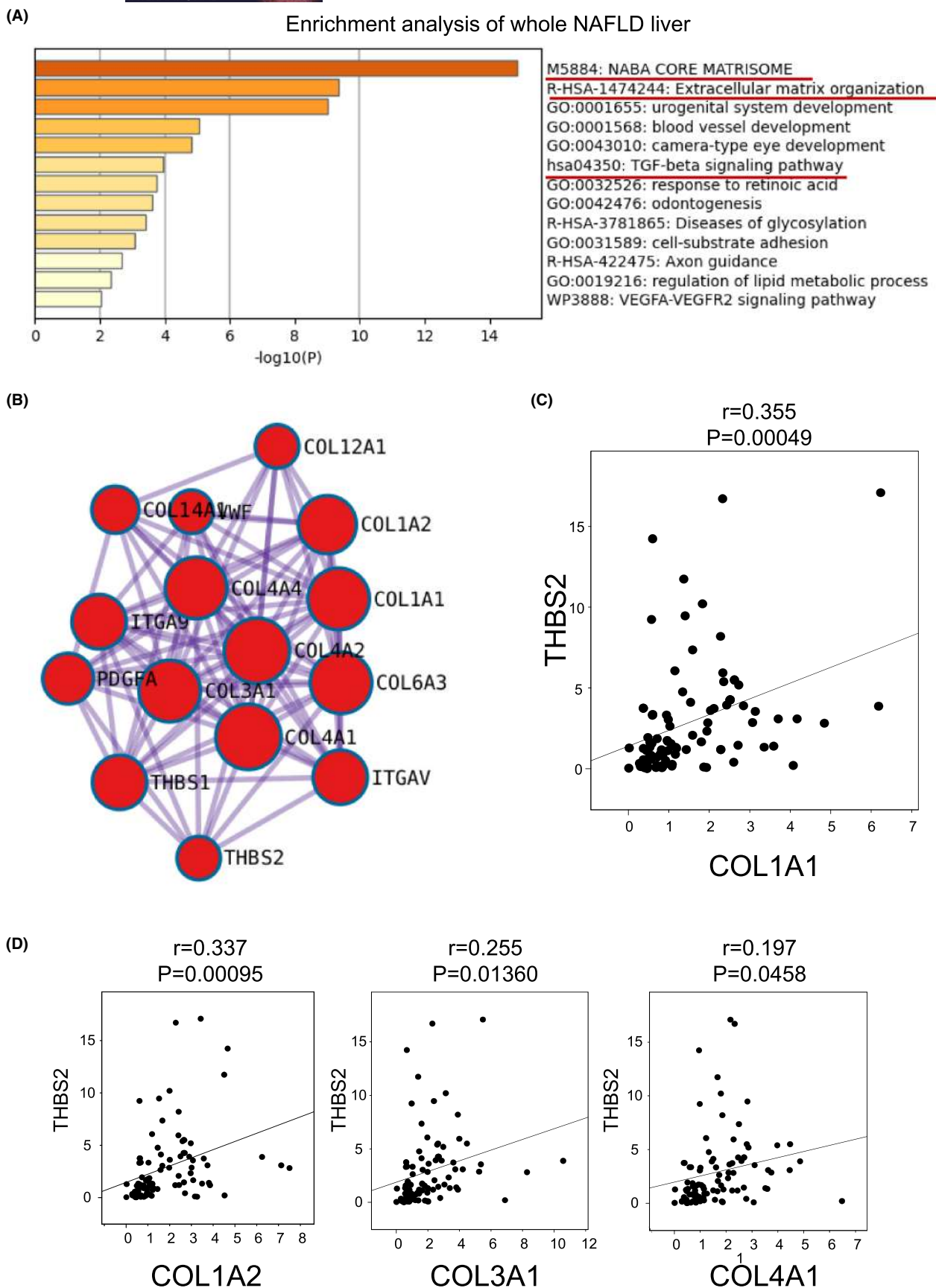
#### 3.1 | Hepatic THBS2 expression is markedly increased in NAFLD patients with advanced fibrosis

We first analysed a microarray dataset (GSE49541) to determine the extent to which THBS2 was highly expressed in the liver of

patients with NAFLD.<sup>21,22</sup> When comparing mild fibrosis (F0-1) with advanced fibrosis (F3-4), 1433 out of 53242 genes showed significant variation. Characteristically, THBS2 ranked second of the 1433 variable genes exhibited by advanced fibrosis NAFLD (Figure 1A). The log<sub>2</sub> fold change for THBS2 was 2.113, and -log<sub>10</sub> (*p* value) was 15.821. The expression values of THBS2 in mild (F0-1) and advanced (F3-4) fibrosis cases are shown in Figure 1C. While



**FIGURE 1** THBS2 gene expression is upregulated in NAFLD patients with advanced fibrosis. (A–C) Gene expression analysis using a NAFLD liver-derived microarray dataset (GSE49541). Volcano plot (A) and mean differentiation plot (B) with an adjusted *p* < .05 when comparing F3-4 cases with F0-1 cases. The large red dot indicates THBS2. (C) Left panel: THBS2 expression levels by NAFLD case in GSE49541 (F0-1, *n*=40; F3-4, *n*=32). Right panel: Comparison of THBS2 expression in NAFLD cases in the left panel (F0-1 vs. F3-4). (D, E), Relative THBS2 mRNA expression levels in NAFLD liver tissue samples from our institution. Relative THBS2 mRNA expression levels (*n*=96) by fibrosis stage (D), F0-1 (*n*=59) vs. F2-4 (*n*=37) and F0-2 (*n*=71) vs. F3-4 (*n*=25) (E). Data are presented as the mean ± standard error of the mean. Correlation analysis was conducted by Kendall's tau test (D) and two-group comparisons by the Mann-Whitney *U* test (C, E).



**FIGURE 2** Hepatic THBS2 is associated with collagen genes in NAFLD. (A, B) Enrichment analysis (A) and network analysis (B) in a NAFLD liver gene set (GSE49541) using Metascape.<sup>23</sup> (B) Group of genes showing strong linkage to THBS2. This is a partial expansion of Figure S1C,D. Correlation of relative mRNA expression levels of THBS2 and COL1A1 (C), as well as of THBS2 with COL1A2, COL3A1, and COL4A1 (D) in NAFLD liver tissue samples from our institution ( $n=96$ ). Correlation analysis was conducted by Spearman's test (C).

THBS2 expression was not high for mild fibrosis, elevated THBS2 expression was evident in many advanced fibrosis cases ( $p < .0001$ ). qPCR using 96 biopsied NAFLD samples also revealed a correlation between fibrosis stage and hepatic THBS2 expression ( $r = .349$ ,  $p = .00046$ ) (Figure 1D), thus validating the RNA sequencing data (Figure 1B,C). Similarly, THBS2 expression was significantly higher in patients with greater fibrosis in comparisons of F0-1 vs. F2-4 and F0-2 vs. F3-4 ( $p = .0356$  and  $p = .0017$ , respectively) (Figure 1E).

### 3.2 | Enrichment analysis of comprehensive genetic data: association of THBS2 with collagen-related genes and TGF $\beta$ pathway

Enrichment analysis was performed using Metascape<sup>23</sup> using the same gene dataset (GSE49541) as in the previous section.<sup>21,22</sup> In NAFLD with advanced fibrosis, the TGF $\beta$  signalling pathway was activated in addition to the organization of collagen-containing NABA core matrisome and ECM (Figure 2A, underlined).

Figure S1 shows the THBS2 gene network in the progression of fibrotic NAFLD, with Figure 2B providing partial enlargement of the THBS2 nearside. The established MCODE algorithm of Metascape was employed to find the densely connected genes in the network, showing the biological role of each component. Network analysis revealed THBS2 to be associated with collagen-related genes, including collagen type I alpha 1 (COL1A1), collagen type I alpha 2 (COL1A2), collagen type III alpha 1 (COL3A1), and collagen type IV alpha 1 (COL4A1). Supportive qPCR data from liver biopsy tissue from our own cohort also showed a significant correlation between THBS2 and collagen-related genes (vs. COL1A1:  $r = .355$ ,  $p = .00049$ ; vs. COL1A2:  $r = .337$ ,  $p = .00095$ ; vs. COL3A1:  $r = .225$ ,  $p = .01360$ ; vs. COL4A1:  $r = .197$ ,  $p = .0458$ ) (Figure 2C,D). The clinical data of the patients used for qPCR are shown in Table 1.

### 3.3 | scRNA-seq data and in situ hybridization analysis of fibrotic NAFLD cases confirm the overexpression of THBS2 in HSCs

scRNA-seq data obtained from the GSE174748 and GSE189175 datasets were analysed using Seurat 4.5 to determine the THBS2-expressing cells in healthy individuals and patients with NAFLD (Figure 3A).<sup>27,28</sup> The THBS2 gene was particularly enriched in HSC populations over other cell types (Figure 3B,C, Figure S2A,B). Notably, we observed a significant upregulation of THBS2 expression in fibrotic NAFLD patients, particularly in HSCs, when compared with healthy controls and subjects with NAFLD but no fibrosis (Figure 3D, Figure S2C). These findings suggested that THBS2 might play a role in the development of fibrosis in NAFLD.

To further confirm the expression of THBS2 in specific cell types, we performed in situ hybridization analysis using a THBS2 mRNA probe of liver tissue sections from severely fibrotic NAFLD cases,

which revealed positive brown signals indicating THBS2 expression in HSCs in F3 NAFLD livers (Figure 4A–C, left). In contrast, no THBS2 signals were observed in control livers (Figure 4D, left). Upon closer examination (Figure 4A), THBS2 signalling was mainly present around hepatocytes and collagen fibres. These results supported the scRNA-seq data, confirming that HSCs contributed to the elevated expression of THBS2 in fibrotic NAFLD livers.

### 3.4 | THBS2 and COL1A1 are upregulated in LX-2 cells upon TGF $\beta$ treatment

To investigate the mechanisms underlying the association between THBS2 and fibrosis development, in vitro studies were performed using LX-2 cells as immortalized HSCs<sup>38</sup> based on our finding that HSCs were the primary cells responsible for THBS2 in fibrotic NAFLD cases. COL1A1 mRNA levels were significantly increased by co-treatment with TGF $\beta$  in LX-2 cells ( $p < .0001$ ), although HepG2 was not (Figure 5A). Since COL1A1 mRNA increased according to TGF $\beta$  concentration (Figure 5B), a TGF $\beta$  concentration of 4 ng/mL was used throughout this study.

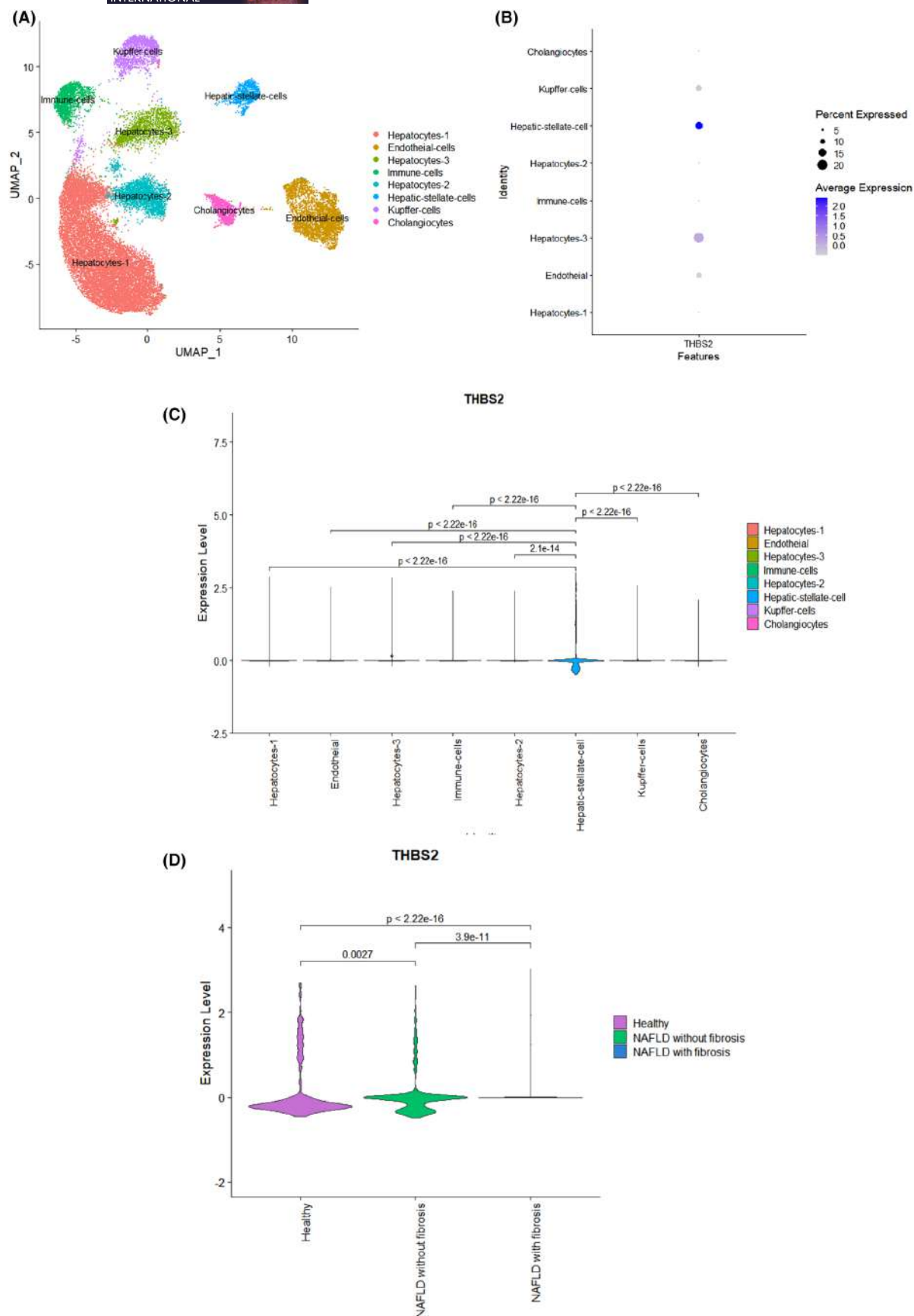
LX-2 cells expressed greater than 160-fold THBS2 mRNA levels compared with HepG2 cells (Figure 5C). TGF $\beta$ -treated LX-2 cells showed significantly higher THBS2 mRNA expression levels than in control conditions ( $p < .0001$ ) (Figure 5C). Furthermore, TGF $\beta$  treatment increased TSP2 protein levels in LX-2 cells (Figure 5D–F) and LX-2 cell culture medium (Figure 5G). These results indicated that THBS2/TSP2 expression was enhanced in HSCs and TSP2 was secreted by TGF $\beta$  treatment.<sup>16</sup>

### 3.5 | Knockdown of THBS2 gene suppresses COL1A1 expression in LX-2 cells

To evaluate whether COL1A1 was regulated by THBS2, THBS2 gene knockdown was evaluated in LX-2 cells using siRNA technology. As shown in Figure 6A,B, THBS2-siRNA treatment reduced the expression of THBS2 mRNA and TSP2 protein by roughly half versus control siRNA treatment ( $p = .0003$ ). THBS2 gene knockdown suppressed constitutive COL1A1 mRNA levels ( $p = .0046$ ) (Figure 6C), which was also observed in TGF $\beta$ -treated conditions ( $p = .005$ ) (Figure 6D). Conversely, treatment of LX-2 with recombinant TSP2 increased COL1A1 mRNA levels relative to control conditions ( $p = .0051$ ) (Figure 6E).

Fluorescent staining of type 1 collagen was performed to verify the above results. Green collagen fibres were significantly reduced by THBS2-siRNA ( $p = .0004$ ) (Figure 6F,G). TGF $\beta$  treatment resulted in prominent collagen accumulation inside and outside of cells, whereas THBS2-siRNA treatment significantly decreased collagen ( $p = .0001$ ) (Figure 6F,G). These findings strongly implicated THBS2 as a regulatory gene for collagen formation in LX-2 cells. Examining the time course of TGF $\beta$  treatment in LX-2 cells and the expression of THBS2 and COL1A1, THBS2 mRNA expression peaked at

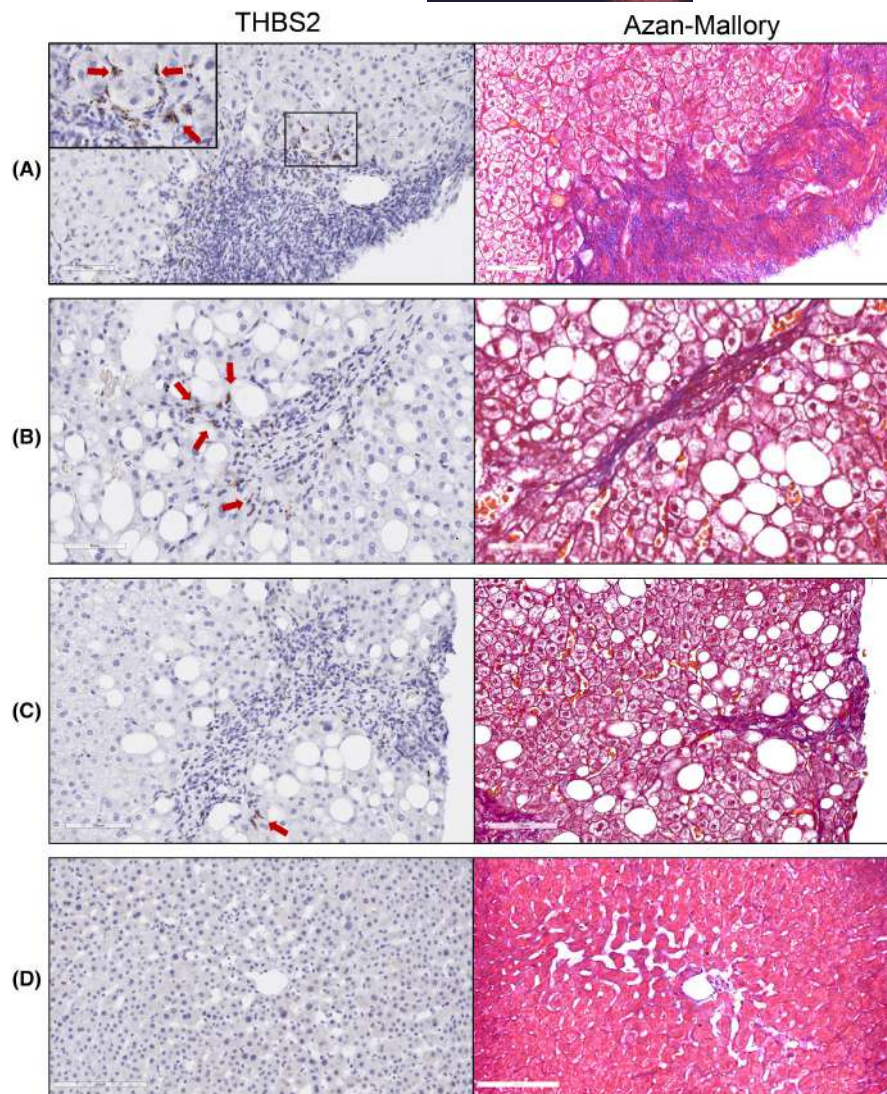




**FIGURE 3** Predominant expression of THBS2 in HSCs by scRNA-seq data analysis. (A–D) GSE174748 and GSE189175 scRNA-seq datasets were used in this study. See Methods for details. (A) UMAP plot of liver cell clusters for the whole dataset. Cells are coloured by inferred liver cell types. (B) The dot plot showing THBS2 expression levels by cell clusters. (C) Comparison of THBS2 expression levels in different cell clusters. (D) Violin plot showing THBS2 expression levels of HSCs in three groups: healthy, NAFLD with fibrosis, and NAFLD without fibrosis. Statistical significance was applied using the Mann–Whitney  $U$  test in the `stat_compare_means` function of Seurat (C, D).



**FIGURE 4** In situ hybridization analysis of NAFLD liver reveals THBS2 expression in HSCs around collagen fibres. (A–D) In situ hybridization analysis of liver tissue samples using a THBS2 mRNA probe (left) and Azan–Mallory staining (right). NAFLD cases with fibrosis (F3,  $n=3$ ) (A–C) and control case without chronic liver disease ( $n=1$ ) (D). Arrows indicate representative positive sites.



10 hours, followed next by an increase in COL1A1 (Figure 6H). This result was consistent with the notion of THBS2 being a regulator of COL1A1 in LX-2 cells.

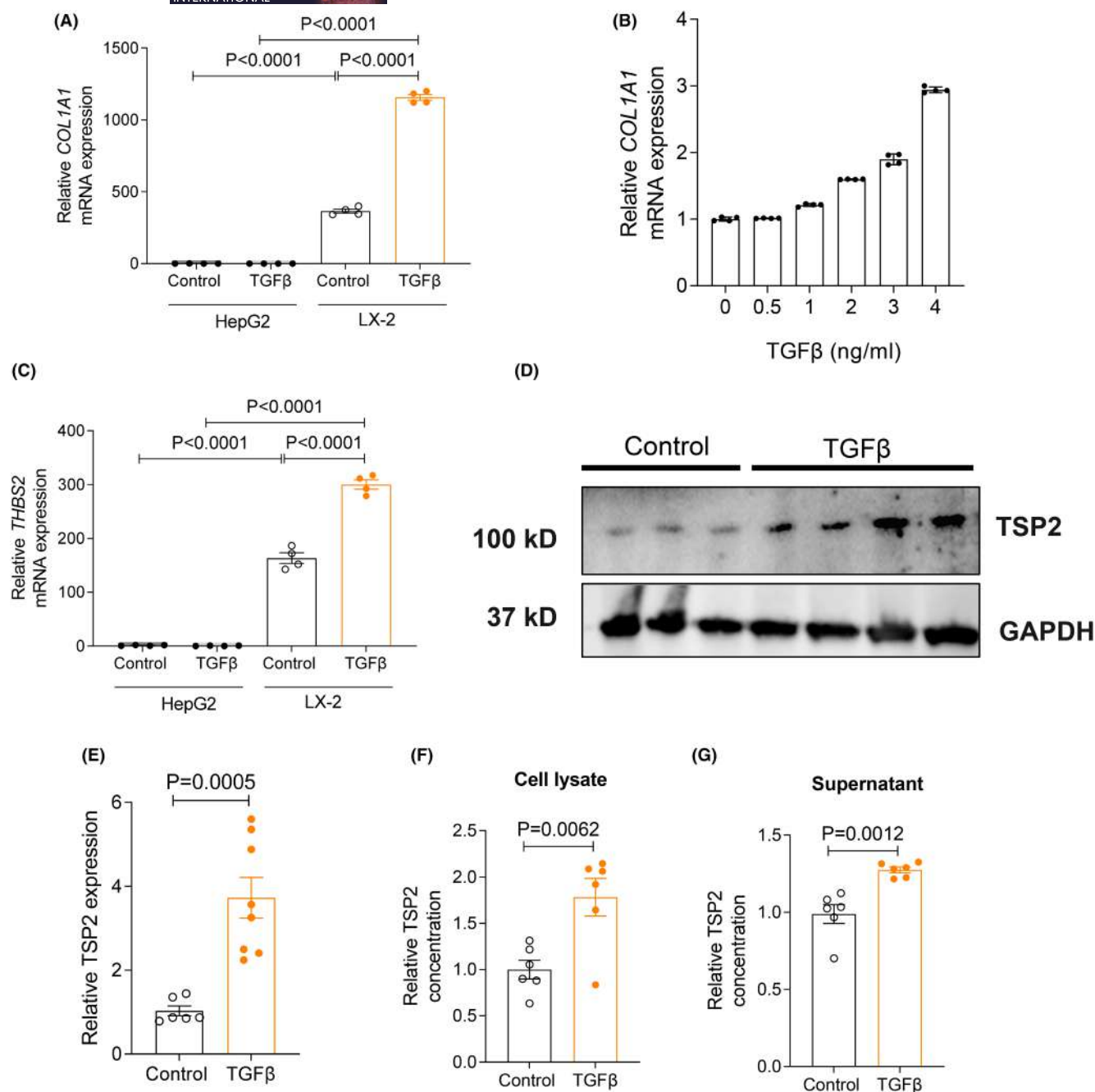
### 3.6 | THBS2/TSP2 expression is independent of the classical TGF $\beta$ -SMAD pathway in LX-2 cells

It is well known that TGF $\beta$  expresses collagen via the phosphorylation of SMAD2/3,<sup>40</sup> and so we investigated the association between THBS2/TSP2 expression and SMAD in LX-2 cells. As shown in Figure 7A,B, the phosphorylation of SMAD2/3 induced by TGF $\beta$  treatment in LX-2 cells was not affected by the presence or absence of THBS2-siRNA. By examining the changes in THBS2 and COL1A1 mRNA expression levels by SMAD inhibitor treatment in LX-2 cells, we observed that while COL1A1 mRNA levels were significantly reduced ( $p=.0005$ ) by the SMAD inhibitor, THBS2 levels were not (Figure 7C). These findings indicated that THBS2/TSP2 expression in LX-2 cells was independent of the classical TGF $\beta$ -SMAD pathway (Figure 7D).

## 4 | DISCUSSION

In this study, a comprehensive hepatic gene dataset and in-house biopsy tissue from NAFLD patients revealed significantly higher expression of THBS2 in the livers of patients with advanced fibrosis than in those with mild fibrosis. This result is consistent with several previous reports and highlights the significance of THBS2 in NAFLD fibrosis.<sup>15,16</sup> Interestingly, detailed gene network analysis and qPCR examination showed that the THBS2 gene was strongly associated with COL1A1, COL1A2, COL3A1, and COL4A1 expression, which supported the relationship between THBS2 and collagen expression witnessed in cell experiments.

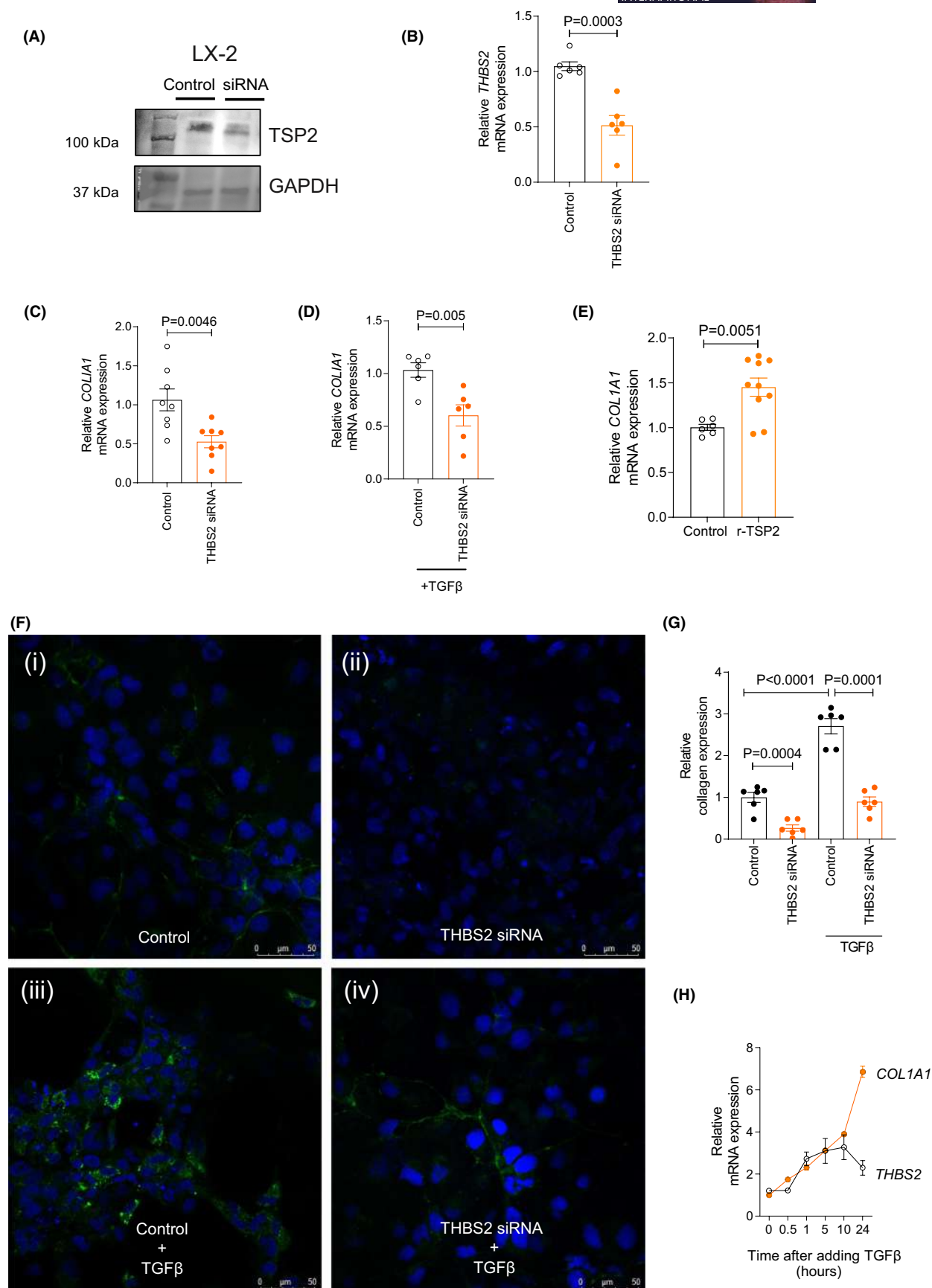
In the analysis of scRNA-seq data, a greater number of THBS2-expressing cells were observed in the HSC cluster. Additionally, the THBS2 expression level of HSCs in NAFLD cases with fibrosis was higher than that of HSCs in NAFLD/healthy individuals without fibrosis. This was consistent with a very recent scRNA-seq data analysis study showing THBS2 expression to be more abundant in fibroblasts in the liver.<sup>41</sup> To confirm this notion, we performed in situ hybridization analysis of liver tissue from NAFLD patients



**FIGURE 5** TGF $\beta$ -mediated activation of LX-2 cells further raises expression of THBS2/TSP2 and COL1A1. (A) COL1A1 mRNA expression levels in HepG2 and LX-2 cells in the presence and absence (control) of TGF $\beta$ . (B) COL1A1 mRNA expression levels at different TGF $\beta$  concentrations. (C) THBS2 mRNA expression levels in HepG2 and LX-2 cells in the presence and absence (control) of TGF $\beta$ . (D) Western blotting studies with LX-2 cells treated in the presence and absence (control) of TGF $\beta$ . (E–G) Quantitative analysis of TSP2 protein expression levels in cell lysates by western blotting (E) and ELISA (F), as well as in the supernatant (G) by ELISA of LX-2 cells in the presence and absence (control) of TGF $\beta$ . (A, C) Data were expressed relative to basal levels measured in the absence of TGF $\beta$  with HepG2 cells. (B) Data were expressed relative to basal levels measured in the absence of TGF $\beta$ . (E–G) TSP2 protein expression levels were normalized relative to protein expression levels of the control. TGF $\beta$  treatment was 4 ng/mL for 10h, except for (B) Data are presented as the mean  $\pm$  standard error of the mean of at least three independent experiments. Correlation analysis was conducted by one-way ANOVA followed by Bonferroni's post-hoc test (A, C) and two-tailed Student's *t*-test (E–G).

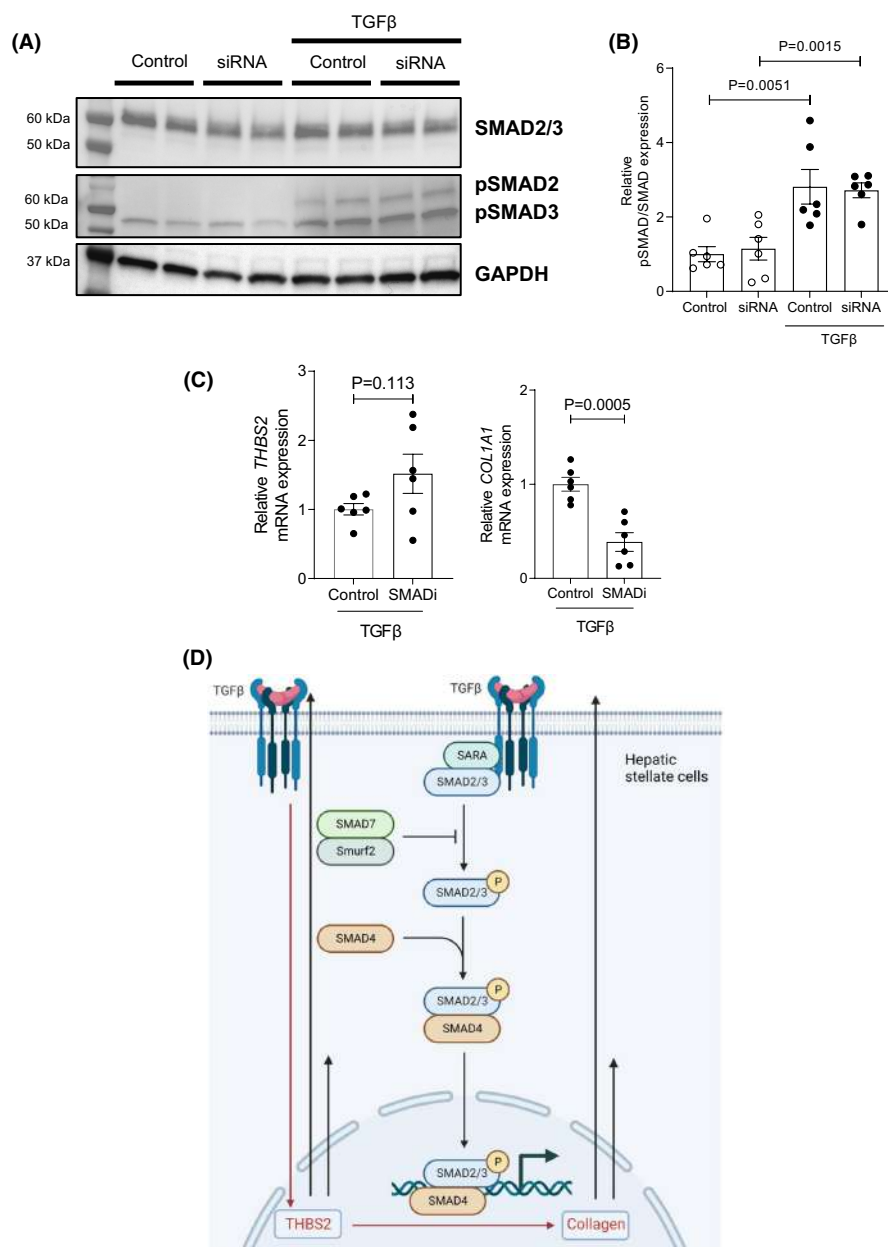
with severe fibrosis to reveal high THBS2 expression in HSCs at the margins of collagen fibre accumulation. These results prompted us to hypothesize that HSCs played a certain role in collagen formation; indeed, TGF $\beta$  treatment markedly increased THBS2/TSP2

expression and simultaneously promoted collagen formation in experiments with LX-2 cells. The suppression of THBS2 reduced collagen expression, identifying it as a possible upstream regulator of collagen expression.





**FIGURE 6** Knockdown of THBS2 gene decreases collagen formation in LX-2 cells. (A, B) Western blotting studies (A) and relative THBS2 mRNA expression levels (B) in LX-2 cells treated with control siRNA and THBS2 siRNA. (C, D) Relative COL1A1 mRNA expression levels treated with control siRNA and THBS2 siRNA in the presence (D) and absence (C) of TGF $\beta$ . (E) Relative COL1A1 mRNA expression levels in the presence and absence of recombinant TSP2 protein (30 ng/mL). (F) Fluorescent staining of LX-2 cells treated with control siRNA or THBS2 siRNA or in the presence and absence of TGF $\beta$ : (i) control siRNA, (ii) THBS2 siRNA, (iii) control siRNA and TGF $\beta$ , and (iv) THBS2 siRNA and TGF $\beta$ . Collagen expression was identified as green and cell nuclei as blue DAPI. (G) Relative quantification of collagen expression values per cell. (H) Expression levels of THBS2 mRNA and COL1A1 mRNA over time after addition of TGF $\beta$  to LX-2 cells. (B–E, G) Data were normalized relative to mRNA expression levels of the control. TGF $\beta$  treatment was performed with 4 ng/mL for 10 h, except for (H). Data are presented as the mean  $\pm$  standard error of the mean of at least three independent experiments. Correlation analysis was conducted by two-tailed Student's *t*-test (B–E) and one-way ANOVA followed by Bonferroni's post-hoc test (G).



**FIGURE 7** THBS2 expression is independent of TGF $\beta$ -SMAD2/3 pathway in LX-2 cells. (A, B) Western blotting studies of SMAD 2/3 (A) and quantitative analysis (B) in LX-2 cells treated with control siRNA and THBS2 siRNA in the presence and absence of TGF $\beta$ . Control: control siRNA, siRNA: THBS2 siRNA. (C, D) SMAD inhibitors downregulate COL1A1 mRNA expression levels without affecting THBS2 mRNA expression levels. SMADi: SMAD inhibitor, SIS3 (10  $\mu$ M). (D) Proposed mechanism of TGF $\beta$ -THBS2/TSP2 expression in LX-2 cells. (B, C) Data were normalized relative to expression levels of the control. TGF $\beta$  treatment was performed with 4 ng/mL for 10 h. Data are presented as the mean  $\pm$  standard error of the mean of at least three independent experiments. Correlation analysis was conducted by one-way ANOVA followed by Bonferroni's post-hoc test (B) and two-tailed Student's *t*-test (C).

ECM is composed of numerous proteins, including collagen, fibronectin, and laminin.<sup>42</sup> Genetic mutations of ECM proteins can profoundly alter ECM properties, as evidenced by Marfan and Ehlers-Danlos syndromes.<sup>14</sup> On the other hand, TSP2 does not directly contribute to ECM structure or stability, but may influence cell-matrix interactions and cell signalling/behaviour.<sup>43</sup> For instance,

although whole-body TSP2-deficient mice had a normal appearance and were capable of normal reproduction, the animals had loosened connective tissue, including the skin, ligaments, and tendons, which appeared as disorganized and abnormal collagen fibres on microscopy.<sup>44</sup> Such results emphasize that TSP2 is required for proper formation and organization of collagen fibres in the skin and tendons.

It is reasonable to presume that TSP2 can also affect collagenogenesis in the liver in a similar manner.

Little is known on the transcriptional regulation of TSP2.<sup>14</sup> However, the regulatory mechanism of thrombospondin 1 (TSP1), which shares 85% of amino acids along with structural similarities and a similar binding domain, may provide a stepping stone to understand the regulatory mechanism of TSP2.<sup>14</sup> The binding domain of TSP1 is postulated to interact with cell surface receptors (LRP, CD36, and CD47), ECM components (decorin, fibronectin, and heparan sulphate proteoglycan), enzymes (matrix metalloproteinase, elastase, and cathepsin G), and calcium as well as TGF $\beta$ .<sup>14</sup> The interaction of TGF $\beta$  with this binding domain is consistent with the increased expression of THBS2/TSP2 upon TGF $\beta$  loading in LX-2 cells in the present study.

Lastly, TGF $\beta$  is a multifunctional cytokine that plays an important role in regulating cell proliferation, differentiation, migration, and ECM deposition.<sup>45</sup> TGF $\beta$  signalling is essential for the formation and maintenance of tissue homeostasis, and its dysregulation is strongly implicated in NAFLD fibrosis.<sup>46,47</sup> In the liver, TGF $\beta$  is produced primarily by Kupffer cells and hepatic sinusoidal endothelial cells.<sup>48</sup> During liver injury, TGF $\beta$  promotes the activation and proliferation of HSCs as well as the formation of ECM components, especially collagen.<sup>49,50</sup> The collagen formation in HSCs is mainly regulated by the TGF $\beta$ -SMAD2/3 pathway.<sup>51</sup> However, in the present experiments, whereas TGF $\beta$ -induced collagen synthesis was inhibited by disrupting THBS2/TSP2 expression, SMAD2/3 phosphorylation was unaltered in LX-2 cells. The detailed regulation of collagen by TSP2 is complex and may involve a variety of regulatory systems, and thus is a subject for future studies.

In summary, this investigation revealed that HSCs express THBS2/TSP2 in the liver of NAFLD patients with advanced fibrosis using in situ hybridization and a scRNA-seq dataset. We also provided evidence that THBS2 knockdown inhibited collagen formation in the LX-2 cell line of HSCs. These results suggest the potential clinical utility of THBS2 as a therapeutic target to halt the progression of liver fibrosis in NAFLD.

## AUTHOR CONTRIBUTIONS

Takefumi Kimura, Takanobu Iwadare, and Naoki Tanaka designed the study and researched data. Takefumi Kimura, Takanobu Iwadare, Tomoo Yamazaki, and Daiki Aomura carried out experiments and interpreted and analysed experimental data. Satoru Joshita collected patient samples and analysed data. Shun-ichi Wakabayashi, Seema Kuldeep, Hamim Zafar, and Sai P Pydi analysed scRNA-seq data. Mai Iwaya and Takeshi Uehara provided pathological guidance. Takefumi Kimura and Takanobu Iwadare wrote the manuscript. Sai P Pydi and Takeji Umemura supervised the entire study.

## ACKNOWLEDGEMENTS

The authors thank Asami Yamazaki and Trevor Ralph for their assistance in sample preparation and English proofreading, respectively.

## FUNDING INFORMATION

This research was supported by AMED under grant number JP23fk0210125 and by JSPS KAKENHI under grant number JP22K20884. T.K. was the recipient of the Aiba Works Medical Research grant for carrying out several experiments. S.P.P. lab is supported by the Indian Council of Medical Research (5/4/8-18/Obs/SPP/2022-NCD-II).

## CONFLICT OF INTEREST STATEMENT

The authors declare no competing interests.

## DATA AVAILABILITY STATEMENT

The gene data referenced during the study have been deposited in GEO under the accession codes GSE49541, GSE174748, and GSE189175. All the other data supporting the findings of this study are available within the article and its Supplementary Information files or from the corresponding author on reasonable request.

## ORCID

Takefumi Kimura  <https://orcid.org/0000-0002-1481-1029>

Takeji Umemura  <https://orcid.org/0000-0001-7985-919X>

## REFERENCES

1. Lazarus JV, Mark HE, Anstee QM, et al. Advancing the global public health agenda for NAFLD: a consensus statement. *Nat Rev Gastroenterol Hepatol*. 2022;19(1):60-78.
2. Powell EE, Wong VW, Rinella M. Non-alcoholic fatty liver disease. *Lancet*. 2021;397(10290):2212-2224.
3. Younossi Z, Anstee QM, Marietti M, et al. Global burden of NAFLD and NASH: trends, predictions, risk factors and prevention. *Nat Rev Gastroenterol Hepatol*. 2018;15(1):11-20.
4. Tanaka N, Kimura T, Fujimori N, Nagaya T, Komatsu M, Tanaka E. Current status, problems, and perspectives of non-alcoholic fatty liver disease research. *World J Gastroenterol*. 2019;25(2):163-177.
5. Vilar-Gomez E, Calzadilla-Bertot L, Wai-Sun Wong V, et al. Fibrosis severity as a determinant of cause-specific mortality in patients with advanced nonalcoholic fatty liver disease: a multi-national cohort study. *Gastroenterology*. 2018;155(2):443-457.e417.
6. Kimura T, Tanaka N, Tanaka E. What will happen in patients with advanced nonalcoholic fatty liver disease? *Hepatobiliary Surg Nutr*. 2019;8(3):283-285.
7. Sanyal AJ, Van Natta ML, Clark J, et al. Prospective study of outcomes in adults with nonalcoholic fatty liver disease. *N Engl J Med*. 2021;385(17):1559-1569.
8. Tokushige K, Ikejima K, Ono M, et al. Evidence-based clinical practice guidelines for nonalcoholic fatty liver disease/nonalcoholic steatohepatitis 2020. *J Gastroenterol*. 2021;56(11):951-963.
9. Kimura T, Singh S, Tanaka N, Umemura T. Role of G protein-coupled receptors in hepatic stellate cells and approaches to anti-fibrotic treatment of non-alcoholic fatty liver disease. *Front Endocrinol (Lausanne)*. 2021;12:773432.
10. Zhang K, Li M, Yin L, Fu G, Liu Z. Role of thrombospondin-1 and thrombospondin-2 in cardiovascular diseases. *Int J Mol Med*. 2020;45(5):1275-1293.
11. Gao F, Chen W, Zhao T, et al. Diagnostic and prognostic roles of thrombospondin-2 in digestive system cancers. *Biomed Res Int*. 2022;2022:3749306.
12. Simantov R, Febbraio M, Silverstein RL. The antiangiogenic effect of thrombospondin-2 is mediated by CD36 and modulated by histidine-rich glycoprotein. *Matrix Biol*. 2005;24(1):27-34.



13. Rienks M, Papageorgiou AP, Frangogiannis NG, Heymans S. Myocardial extracellular matrix: an ever-changing and diverse entity. *Circ Res*. 2014;114(5):872-888.
14. Calabro NE, Kristofik NJ, Kyriakides TR. Thrombospondin-2 and extracellular matrix assembly. *Biochim Biophys Acta*. 2014;1840(8):2396-2402.
15. Kozumi K, Kodama T, Murai H, et al. Transcriptomics identify thrombospondin-2 as a biomarker for NASH and advanced liver fibrosis. *Hepatology*. 2021;74(5):2452-2466.
16. Kimura T, Tanaka N, Fujimori N, et al. Serum thrombospondin 2 is a novel predictor for the severity in the patients with NAFLD. *Liver Int*. 2021;41(3):505-514.
17. Lee CH, Seto WK, Lui DT, et al. Circulating thrombospondin-2 as a novel fibrosis biomarker of nonalcoholic fatty liver disease in type 2 diabetes. *Diabetes Care*. 2021;44(9):2089-2097.
18. Wu X, Cheung CKY, Ye D, et al. Serum thrombospondin-2 levels are closely associated with the severity of metabolic syndrome and metabolic associated fatty liver disease. *J Clin Endocrinol Metab*. 2022;107(8):e3230-e3240.
19. Iwadare T, Kimura T, Tanaka N, et al. Circulating thrombospondin 2 levels reflect fibrosis severity and disease activity in HCV-infected patients. *Sci Rep*. 2022;12(1):18900.
20. Matsumae T, Kodama T, Tahata Y, et al. Thrombospondin-2 as a predictive biomarker for hepatocellular carcinoma after hepatitis C virus elimination by direct-acting antiviral. *Cancers (Basel)*. 2023;15(2):463.
21. Moylan CA, Pang H, Dellinger A, et al. Hepatic gene expression profiles differentiate presymptomatic patients with mild versus severe nonalcoholic fatty liver disease. *Hepatology*. 2014;59(2):471-482.
22. Murphy SK, Yang H, Moylan CA, et al. Relationship between methyloyme and transcriptome in patients with nonalcoholic fatty liver disease. *Gastroenterology*. 2013;145(5):1076-1087.
23. Zhou Y, Zhou B, Pache L, et al. Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. *Nat Commun*. 2019;10(1):1523.
24. Gene Ontology Consortium. The gene ontology project in 2008. *Nucleic Acids Res*. 2008;36(suppl\_1):D440-D444.
25. Kanehisa M, Sato Y, Kawashima M, Furumichi M, Tanabe M. KEGG as a reference resource for gene and protein annotation. *Nucleic Acids Res*. 2016;44(D1):D457-D462.
26. Jassal B, Matthews L, Viteri G, et al. The reactome pathway knowledgebase. *Nucleic Acids Res*. 2020;48(D1):D498-D503.
27. Filliol A, Saito Y, Nair A, et al. Opposing roles of hepatic stellate cell subpopulations in hepatocarcinogenesis. *Nature*. 2022;610(7931):356-365.
28. Alvarez M, Benhammou JN, Darci-Maher N, et al. Human liver single nucleus and single cell RNA sequencing identify a hepatocellular carcinoma-associated cell-type affecting survival. *Genome Med*. 2022;14(1):50.
29. Fujimori N, Kimura T, Tanaka N, et al. 2-step PLT16-AST44 method: simplified liver fibrosis detection system in patients with non-alcoholic fatty liver disease. *Hepatol Res*. 2022;52(4):352-363.
30. Kleiner DE, Brunt EM, Van Natta M, et al. Design and validation of a histological scoring system for nonalcoholic fatty liver disease. *Hepatology*. 2005;41(6):1313-1321.
31. Nakajima T, Uehara T, Maruyama Y, Iwaya M, Kobayashi Y, Ota H. Distribution of Lgr5-positive cancer cells in intramucosal gastric signet-ring cell carcinoma. *Pathol Int*. 2016;66(9):518-523.
32. Jia F, Diao P, Wang X, et al. Dietary restriction suppresses steatosis-associated hepatic tumorigenesis in hepatitis C virus core gene transgenic mice. *Liver Cancer*. 2020;9(5):529-548.
33. Jain S, Ruiz de Azua I, Lu H, White MF, Guettier JM, Wess J. Chronic activation of a designer G(q)-coupled receptor improves beta cell function. *J Clin Invest*. 2013;123(4):1750-1762.
34. Diao P, Wang Y, Jia F, et al. Dietary fat composition affects hepatic angiogenesis and lymphangiogenesis in hepatitis C virus core gene transgenic mice. *Liver Cancer*. 2023;12(1):57-71.
35. Kimura T, Pydi SP, Wang L, et al. Adipocyte G(q) signaling is a regulator of glucose and lipid homeostasis in mice. *Nat Commun*. 2022;13(1):1652.
36. Han CY, Koo JH, Kim SH, et al. Hecpudin inhibits Smad3 phosphorylation in hepatic stellate cells by impeding ferroportin-mediated regulation of Akt. *Nat Commun*. 2016;7:13817.
37. Wu X, Wu X, Ma Y, et al. CUG-binding protein 1 regulates HSC activation and liver fibrogenesis. *Nat Commun*. 2016;7:13498.
38. Xu L, Hui AY, Albanis E, et al. Human hepatic stellate cell lines, LX-1 and LX-2: new tools for analysis of hepatic fibrosis. *Gut*. 2005;54(1):142-151.
39. Collins TJ. ImageJ for microscopy. *Biotechniques*. 2007;43(1 Suppl):25-30.
40. Khalil H, Kanisicak O, Prasad V, et al. Fibroblast-specific TGF-beta-Smad2/3 signaling underlies cardiac fibrosis. *J Clin Invest*. 2017;127(10):3770-3783.
41. Govaere O, Hasoon M, Alexander L, et al. A proteo-transcriptomic map of non-alcoholic fatty liver disease signatures. *Nat Metab*. 2023;5:572-578.
42. Halper J, Kjaer M. Basic components of connective tissues and extracellular matrix: elastin, fibrillin, fibulins, fibrinogen, fibronectin, laminin, tenascins and thrombospondins. *Adv Exp Med Biol*. 2014;802:31-47.
43. Bornstein P, Sage EH. Matricellular proteins: extracellular modulators of cell function. *Curr Opin Cell Biol*. 2002;14(5):608-616.
44. Kyriakides TR, Zhu YH, Smith LT, et al. Mice that lack thrombospondin 2 display connective tissue abnormalities that are associated with disordered collagen fibrillogenesis, an increased vascular density, and a bleeding diathesis. *J Cell Biol*. 1998;140(2):419-430.
45. Gough NR, Xiang X, Mishra L. TGF-beta signaling in liver, pancreas, and gastrointestinal diseases and cancer. *Gastroenterology*. 2021;161(2):434-452.e415.
46. Ahmed H, Umar MI, Imran S, et al. TGF-beta1 signaling can worsen NAFLD with liver fibrosis backdrop. *Exp Mol Pathol*. 2022;124:104733.
47. Fabregat I, Moreno-Caceres J, Sanchez A, et al. TGF-beta signalling and liver disease. *FEBS J*. 2016;283(12):2219-2232.
48. Carambia A, Freund B, Schwinge D, et al. TGF-beta-dependent induction of CD4(+)CD25(+)Foxp3(+) Tregs by liver sinusoidal endothelial cells. *J Hepatol*. 2014;61(3):594-599.
49. Tsuchida T, Friedman SL. Mechanisms of hepatic stellate cell activation. *Nat Rev Gastroenterol Hepatol*. 2017;14(7):397-411.
50. Jia F, Hu X, Kimura T, Tanaka N. Impact of dietary fat on the progression of liver fibrosis: lessons from animal and cell studies. *Int J Mol Sci*. 2021;22(19):10303.
51. Ellis LR, Warner DR, Greene RM, Pisano MM. Interaction of Smads with collagen types I, III, and V. *Biochem Biophys Res Commun*. 2003;310(4):1117-1123.

## SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

**How to cite this article:** Kimura T, Iwadare T, Wakabayashi S-i, et al. Thrombospondin 2 is a key determinant of fibrogenesis in non-alcoholic fatty liver disease. *Liver Int*. 2024;44:483-496. doi:[10.1111/liv.15792](https://doi.org/10.1111/liv.15792)