

Dedicated to my beloved parents

DECLARATION BY THE CANDIDATE

I hereby declare that the thesis “**Integrative systems biology approaches to identify molecular signatures in gallbladder carcinoma**” being submitted to the Department of Molecular Biology and Biotechnology, Tezpur University, Tezpur, Assam in partial fulfillment for the award of the degree of Doctor of Philosophy in Molecular Biology and Biotechnology, has previously not formed the basis for the award of any degree, diploma, associateship, fellowship or any other similar title or recognition.

Date: 25.04.2024.

Place: Tezpur

Nabanita Roy

(Nabanita Roy)

School: School of Sciences

Department: Molecular Biology and Biotechnology,
Tezpur University, Tezpur 784028




TEZPUR UNIVERSITY

CERTIFICATE OF SUPERVISOR

This is to certify that the thesis entitled “**Integrative systems biology approaches to identify molecular signatures in gallbladder carcinoma**” submitted to the School of Sciences, Tezpur University in requirement of partial fulfillment for the award of the degree of Doctor of Philosophy in Molecular Biology and Biotechnology is a record of research work carried out by **Ms. Nabanita Roy** under my supervision and guidance. All help received by her from various sources has been duly acknowledged. No part of this thesis has been submitted elsewhere for award of any other degree.

Date: 25.04.2024

Place: Tezpur


(Dr. Pankaj Barah)

Designation: Assistant Professor

School: School of Sciences

Department: Molecular Biology and Biotechnology,

Tezpur University, Tezpur 784028

Acknowledgment

First and foremost, I bow before the Almighty, expressing my deepest gratitude for granting me the strength and courage to make a humble contribution to society through my research work.

I express my sincere gratitude to my supervisor, Dr. Pankaj Barah for giving me the opportunity to pursue Ph.D. under his guidance. His insightful advice, encouragement, constant vigilance, and constructive criticism during my Ph.D. tenure have enabled me to perform better and develop good research practices. I am extremely grateful for his dedicated support, which has played a crucial role in my academic development and successful completion of my doctoral thesis.

I offer my gratitude to Prof. V. K Jain and Prof. Shambhu Nath Singh, the former and present Vice Chancellor of Tezpur University, respectively for granting me the opportunity to work in this esteemed institution. I always feel fortunate to be a part of this University which has an excellent environment for research among the students.

I deeply acknowledge the Department Heads of Molecular Biology and Biotechnology, Tezpur University for extending all possible facilities for carrying out my research work. I am also thankful to my Doctoral Committee members- Prof. Dhruva Kumar Bhattacharyya, Prof. Robin Doley, Dr. Rupak Mukhopadhyay, and Prof. (Dr.) Anupam Sarma as well as all the faculty members of the Department of Molecular Biology and Biotechnology for their valuable suggestions throughout my Ph.D. tenure.

I want to express my sincere gratitude to Prof. (Dr.) Anupam Sarma, Professor and Head of the Department of Oncopathology, and Dr. Avdhesh Rai, Scientific Officer at Dr. B. Borooah Cancer Institute (BBCI), Guwahati, for the generous support during my visit for sample collection and qRT-PCR analysis. I would also like to acknowledge Prof. (Dr.) Subhash Khanna, CEO of Swagat Super Speciality Surgical Hospital (SSSSH), Guwahati for his dedicated support in the collection of surgically resected gallbladder tissue samples. I am thankful to all the nursing staff, lab technicians, and project staffs of BBCI and SSSSH for their help and cooperation.

I would also like to acknowledge Dr. Mohit Kumar Jolly, Assistant Professor at IISC Bangalore and his group for extending his support in performing the EMT analysis.

I extend my sincere gratitude to the patients who graciously provided their consent for the collection of samples crucial to my Ph.D. work. Their willingness to participate has been invaluable to the success of my research, and their contribution is deeply appreciated.

I also take this opportunity to thank Dr. K. K. Hazarika, Dr. N. K. Bordoloi, Mr. P. Mudoi, Mrs. Pranita S. Talukdar, Mr. Bijoy Mech, Mr. Guna Das, and all the non-teaching staff for all the help they have provided me during my Ph.D. tenure.

I sincerely thank all the past and present members of the EvolOMICs lab- Ankur, Mithil, Drishtee, Chiranjeev, Ria, Barasha, Akash, Cinmoyee, and Prangan for their immense help and support during my Ph.D. journey. I thank all my seniors and juniors of the MBBT department for helping me in some way or the other.

There are no words to express the boundless sacrifice and hard work that my parents put forth for me to achieve my current position. I want to express my deepest sense of gratitude to my Maa, Swati Roy, and Baba, Nishit Ranjan Roy for the emotional, financial, and moral support to achieve all my pursuits. I extend my heartfelt gratitude to my late grandfather, Dr. Samarendra Kumar Roy, and grandmother, Bithika Roy, for their blessings, which have been a source of support for my well-being and success. Special thanks to my litter sister, Nabarati for always motivating me. I also express my gratitude to my in-laws for silently supporting me in my Ph.D. journey and letting me be the person I am. Last but not least; I express my heartfelt appreciation to my husband, Mrinmoy Purkayastha, for providing me the strength and moral support. I am grateful for his belief in my capabilities and the encouragement he provided throughout my Ph.D. journey.

Finally, I would like to thank all the people, whose direct and indirect support has helped me complete my research work.

Date: 25/04/2024
Place: Tezpur

(Nabanita Roy)

LIST OF FIGURES

Figure No.	Figure captions	Page No.
CHAPTER- I		
1.1	The anatomical position of the gallbladder in the hepatobiliary system	2
1.2	Histological representation of the gallbladder wall with the corresponding pathological stages of GBC.	5
CHAPTER- II		
2.1	Estimated age-standardized global incidence (A) and mortality (B) rates of GBC patients in both sexes. (The countries with the highest GBC incidence and mortality rates are highlighted). Figure generated from GLOBOCAN 2020 (https://gco.iarc.fr/).	11
2.2	Incidence of GBC in India. (A) States with high incidence rates are indicated in red dots. (B) The gender-specific burden of GBC cases in India from 2020-2040.	12
2.3	Mechanistic diagram showing the progression of GSD to GBC through different pathological spectra resulting in multiple genomic changes/dysregulations and finally leading to invasive GBC progression.	14
2.4	The roadmap of development of RNA-sequencing technologies.	29
CHAPTER- III		
3.1	Illustrative representation of the overall methodologies employed to identify molecular signatures associated with GBC pathogenesis.	50
3.2	Collection of GBC and GSD clinical tissue samples from the population of Assam.	51
3.3	Schematic workflow for identification of predicted novel lncRNAs from the transcriptomic dataset.	57
3.4	Schematic representation of the TF binding to target gene through PWM scanning. It is a scoring matrix for representing TF binding motifs. It represents a matrix of N rows and four	62

	columns, in which the matrix score of each base at each position is described.	
CHAPTER- IV		
4.1	Schematic representation of the details of four different case studies performed to identify molecular signatures and pathways in GBC.	68
4.2	Outline of the workflow for identification of potential overlapping and unique DEGs in three aggressive cancers of the hepatobiliary system (GBC, HCC, and ICC).	70
4.3	Identification of overlapping gene expression signatures among GBC, HCC, and ICC. (A) Volcano plot representing the identification of significant DEGs in GBC, HCC, and ICC. (B) Upset plot showing the number of shared and unique DEGs in GBC, HCC, and ICC. (C) The heatmap depicts the hierarchical clustering of the 256 shared DEGs between the three cancers of HBCs. (D) Bar plot representing the top ten significant biological processes associated with the shared DEGs between GBC, HCC, and ICC.	72
4.4	Identification of unique gene expression profiles in GBC, HCC, and ICC. (A) Bar plot showing the number of unique DEGs identified in each cancer type. (B) Complete linkage hierarchical clustering analysis of the expression profile of the top 500 unique DEGs identified in GBC, HCC, and ICC.	73
4.5	Construction of gene co-expression networks and identification of nonpreserved modules. Clustering dendrogram (left) of genes based on distances between gene pairs that are subsequently grouped into modules (minClusterSize=30), designated by different colors based on the similarity of the magnitude of gene expression. Preservation analysis (right) of modules was done based on <i>Z-summary</i> and <i>medianRank</i> . Modules whose topological properties changed in a cancer network compared to normal networks are termed non-preserved modules. The black,	74-75

	brown, and lightyellow modules are identified as non-preserved in (A) GBC, (B) HCC, and (C) ICC respectively.	
4.6	Construction of PPI networks with gene sets identified from nonpreserved modules of GBC, HCC, and ICC. The large brown, black, and yellow nodes represent the hub DEGs in GBC, HCC, and ICC respectively based on degree centrality. Significant PPIs were filtered using a combined score > 0.70.	78
4.7	Outline of the workflow for identification of potential DEGs in GBC compared to normal.	81
4.8	Differential gene expression in GBC as compared to control. (A) Gene-wise complete linkage hierarchical clustering heatmap of DEGs identified in GBC compared to control. (B) Identification of the top ten enriched biological processes associated with the significant DEGs. The x-axis represents the enrichment ratio between the number of DEGs and all UniGenes enriched in particular GO terms. The size of the dot represents the number of DEGs assigned to the particular GO term and the color of the dot represents the <i>Padj</i> . The left panel of the dot plot represents terms/pathways upregulated in GBC and the right panel represents terms downregulated in GBC as compared to the control.	82
4.9	Construction of differential gene co-expression networks in GBC and control samples. (A) Hierarchical clustering dendrogram of DEGs based on dissimilarity measure (the 1-TOM) matrix. The co-expressed modules identified in the GBC and control network are represented by different colors. (B) Clustering network heatmaps of co-expressed modules identified in GBC and control co-expression networks.	83
4.10	Identification of nonpreserved modules from GBC and control networks based on <i>Z-summary</i> and <i>medianRank</i> . (A) Identification of nonpreserved modules in the control condition. The modules in midnightblue and royalblue colors are identified as non-preserved. (B) Identification of modules in the GBC	84-85

	condition, where the modules in tan, salmon, and grey60 color are identified as non-preserved.	
4.11	PPI network analysis of the significant non-preserved modules identified in GBC and control gene co-expression networks. Construction of the PPI networks with the DEGs identified from the non-preserved modules of the control network (midnightblue and royal blue modules) and GBC network (salmon, tan, and grey60 modules). The small blue circles represent the proteins and the large red node represents the genes in the modules.	88
4.12	Outline of the workflow for identification of potential DEGs in GBC compared to GSD with three different follow-up periods.	92
4.13	Identification of differential gene expression and pathways in GBC compared to normal. (A) Bar plot showing the number of total, upregulated, and downregulated DEGs identified in GBC as compared to adjacent normal samples. (B) Gene-wise hierarchical clustering of significant DEGs identified in GBC samples as compared to control samples. (C) Dot plot showing top ten significantly enriched pathways associated with DEGs identified in GBC vs. Normal.	93-94
4.14	Identification of shared and unique differential gene expression profiles in GBC compared to GSD with three different follow-up periods. (A) Venn diagram showing the overlapping DEGs between GBC vs. GSD with 3 different follow-up periods. (B) Top 10 enriched pathways associated with the overlapping DEGs. (C) Complete linkage gene-wise hierarchical clustering of unique DEGs identified in GBC compared to GSD with 3 different follow-up periods.	95
4.15	Signaling network showing the key signaling pathways and complexes associated with the hub DEGs identified in GBC. The green node represents associated proteins; light blue nodes indicate signaling pathways; circled blue nodes represent hub genes; squared blue nodes represent signaling complex and the yellow nodes indicate protein families.	97

4.16	Validation of the expression and genetic alteration of the potential DEGs identified from case studies involving analysis of public transcriptomic datasets (case studies 1, 2, and 3). (A) The genetic alteration associated with hub DEGs identified in GBC. (B) Boxplot showing the gene expression level of the selected hub DEGs in tumor samples compared to normal from four TCGA gastrointestinal cancer datasets. The red and green boxes represent tumor and normal samples respectively, whereas; the black line represents metastatic samples.	98-99
4.17	Schematic outline of the overall workflow carried out for identification of crucial molecular signatures in GBC from in-house generated GBC and GSD transcriptomic dataset.	101
4.18	Visualization of the RNA-seq libraries prepared from each sample in the TapeStation system.	103
4.19	Identification of significant DEGs in GBC and GBC+GS group as compared to control. (A) Bar plot showing the number of DEGs identified in GBC+GS and GBC groups. (B) Hierarchical clustering analysis showing the expression profile of DEGs identified in the GBC and GBC+GS group compared to control (GSD).	105
4.20	Identification of shared DEGs and pathways between GBC and GBC+GS groups. (A) The Venn diagram represents 188 shared DEGs between the two GBC groups. (B) Hierarchical clustering analysis showing heatmap of the expression level of the shared DEGs. (C) Bar plot representing the top ten significant pathways ($P_{adj} < 0.05$) associated with the shared DEGs.	106
4.21	Functional enrichment of significant DEmRNAs. Bar plots representing the top ten significantly enriched biological processes and pathways associated with the upregulated (activated) and downregulated (suppressed) DEmRNAs in GBC+GS (A) and GBC (B) groups. The x-axis and y-axis represent the significantly enriched processes and pathways and p-values (\log_{10} transformed) respectively.	108-109

4.22	Construction of PPI networks and identification of significant module clusters from the whole PPI network. The color scale represents the degree centrality of the interconnected nodes.	110
4.23	Bar plot showing the top five enriched pathways from KEGG, MsigDB, and Reactome database in (A) GBC+GS and (B) GBC group. The X-axis and Y-axis represent the significant pathways and gene ratio respectively.	113
4.24	qRT-PCR validation of the hub DEGs identified in the GBC and GBC+GS group. (A) Bar plot representing the relative expression of <i>LMOD1</i> and <i>SMAD4</i> identified using qRT-PCR data analysis in GBC compared to control. (B) Bar plot showing the gene expression level (log2FoldChange) of <i>LMOD1</i> and <i>SMAD4</i> identified through RNA-seq and qRT-PCR.	114
4.25	Schematic diagram showing gallstone-dependent and gallstone-independent GBC development progresses through distinct molecular pathogenesis involving distinct gene sets and diverse biological pathways. The upregulated and downregulated genes are highlighted in red and green font respectively.	119
CHAPTER- V		
5.1	Overall schematic workflow for identification of crucial lncRNAs involved in GBC pathogenesis.	125
5.2	Identification of differential expression profile of lncRNA in GBC and GBC+GS group. (A) Bar graph showing the total, up and down-regulated DElncRNA and DENlncRNA identified in GBC and GBC+GS groups. (B) Pie chart representing the percentage of novel lncRNAs classified under various categories based on their genomic positions. C. Hierarchical clustering pattern of DElncRNA and DENlncRNA expression profile in GBC and GBC+GS cases compared to control.	126-127
5.3	Identification of overlapping DE-lncRNAs and DE-nlncRNA between GBC & GBC+GS group. Heatmap representing hierarchical clustering of 36 common DElncRNAs (A) and 6 common DENlncRNAs (B) between two GBC groups. (C) Pie	128

	chart representing the experimentally validated common DE-lncRNAs and their associated cancer functional state in GBC.	
5.4	lncRNA-mRNA expression correlation network. Construction of DElncRNA-DEmRNA and DENlncRNA-DEmRNA expression correlation networks in GBC and GBC+GS groups respectively. The blue and red nodes indicate DEmRNAs and DElncRNAs respectively. The bar plots representing the top ten hub DElncRNAs and DENlncRNAs determined through correlation networks based on degree centrality in GBC and GBC+GS groups respectively. The blue bars represent downregulation and the yellow bars indicate upregulation.	130
5.5	Construction of CeRNA networks. (A-B) CeRNA regulatory networks showing the highly connected miRNAs and their target DElncRNAs and DEmRNAs in GBC and GBC+GS groups respectively. (C-D) CeRNA regulatory networks showing the highly connected miRNAs and their target DENlncRNAs and DEmRNAs in GBC and GBC+GS groups respectively. DEmRNAs, DElncRNAs, and miRNAs are represented by blue octagon shapes, red diamonds, and green rectangles respectively.	134
5.6	Pathway analysis of the interacting mRNAs in the ceRNA networks. (A-B) The bar plot represents the top five significantly enriched pathways linked to interacting DEmRNAs in respective GBC and GBC+GS ceRNA networks constructed with DElncRNAs. (C-D) The bar plot represents the top five significantly enriched pathways linked to interacting DEmRNAs in respective GBC and GBC+GS ceRNA networks constructed with DENlncRNA	136
5.7	ceRNA network clusters involving the hub DElncRNAs and DENlncRNAs targeting multiple hub miRNAs in GBC and GBC+GS groups. (A-B) ceRNA network clusters involving hub DE-lncRNA in GBC and GBC+GS groups respectively. (C-D) ceRNA network clusters involving hub DE- novel lncRNA and in GBC and GBC+GS groups respectively.	138

5.8	Cross-validation of the hub lncRNAs with independent public datasets. (A) Boxplot representing the log ₂ foldchange expression of the hub lncRNAs in TCGA-CHOL, TCGA-LIHC, and TCGA-PAAD datasets. (B) Venn diagram showing the shared DElncRNAs identified between public and in-house datasets. The corresponding heatmaps showing the expression patterns of the common DE-lncRNAs identified.	139-140
5.9	qRT-PCR validation of the hub lncRNAs identified in the GBC and GBC+GS group. (A) Bar plot representing the relative expression of lncRNAs identified using qRT-PCR data analysis in GBC compared to control. (B) Bar plot showing the gene expression level (log ₂ FoldChange) of lncRNAs identified through RNA-seq and qRT-PCR.	141
5.10	Predicted secondary structures of MSTRG.16633.1 and MSTRG.53675.1 identified in the GBC and GBC+GS groups respectively.	142
5.11	Mechanistic illustration of the action of hub DElncRNA and DElncRNA in GBC and GBC+GS pathogenesis. (A) Dysregulated lncRNAs (LINC00852 and MSTRG.53675) identified in the gallstone-associated GBC group show that the hub lncRNAs were found to interact with potential hub DElncRNAs involved in FoxO signaling pathways which induce pro-tumorigenic effects by modulating crucial regulatory pathways and promote tumorigenesis in gallbladder with gallstones. (B) In the GBC group, the lncRNAs (DIO3OS and MSTRG.16633) identified indirectly interact with the cell-adhesion molecules and contribute to tumor invasion through EMT and tumor-microenvironment interaction, which ultimately leads to increased cell proliferation and metastatic gallbladder carcinogenesis.	147

CHAPTER VI

6.1	The overall schematic workflow of transcriptional regulatory network analysis and identification of potential TFs in GBC pathogenesis.	154
6.2	Identification of differentially expressed TFs. (A) Venn diagram representing the number of DETFs identified in GBC and GBC+GS cases respectively. (B) Expression heatmap of the DETFs identified in GBC and GBC+GS cases compared to controls. The heatmaps were plotted using the Zscore.	155
6.3	Identification of significant pathways involved with DETFs identified in GBC and GBC+GS groups. Bar plot showing the top five significantly enriched KEGG and hallmark pathways associated with DETFs identified in GBC (A-B) and GBC+GS (C-D) groups respectively. The x-axis and y-axis represents the p-values and the significant KEGG and molecular signature hallmark pathways respectively.	157
6.4	Transcriptional regulatory networks construction and identification of hub TFs. (A-B) Construction of TF-TG regulatory network and identification of hub TFs in GBC and GBC+GS group respectively. (C-D) Construction of TF-lncRNA regulatory network and identification of hub TFs in GBC and GBC+GS group respectively. The blue triangular nodes represent the hub TFs identified using the highest degree of centrality. The red triangles and V-shaped nodes represent the DETFs and DE-lncRNAs respectively.	159
6.5	Pathway enrichment of the target genes identified in TF-TG regulatory networks. The bar plot (right panel) represents the top five identified significant (p-value < 0.05) KEGG pathways associated with hub DETFs identified from (A) GBC and (B) GBC+GS regulatory networks (left panel).	162
6.6	Enriched cancer hallmark processes associated with DElncRNAs identified in TF-lncRNA networks. The pie chart showing the	163

	enriched cancer hallmarks linked to DE-lncRNA targeted by hub TFs identified in the GBC (A) and GBC+GS (B) networks.	
6.7	Construction of a transcriptional regulatory network of DETFs identified in GBC compared to normal. (A) Identification of hub TFs in GBC based on degree centrality. The red node represents the top hub TFs and the small blue nodes represent target genes. (B) Pairwise correlation of EMT score of hub TFs identified through TF-TG interactions. The significance of each hub TF is represented with a symbol- p -value < 0.001 (***) ; p -value < 0.01 (**) and p -value < 0.05 (*).	164
6.8	Validation of <i>KLF15</i> and <i>MECOM</i> expression in TCGA datasets. (A) Box plot showing the expression level (log2foldchange) of <i>KLF15</i> and <i>MECOM</i> in four different TCGA datasets of gastrointestinal cancers. (B) Bar plot representing the mutational profile of <i>KLF15</i> and <i>MECOM</i> in five different cancers of the gastrointestinal tract.	165-166
6.9	qRT-PCR validation of the hub lncRNAs identified in the GBC and GBC+GS group. (A) Bar plot representing the relative expression of <i>KLF15</i> and <i>MECOM</i> TFs identified using qRT-PCR data analysis in GBC and GBC+GS group compared to control. (B) Bar plot showing the gene expression level (log2FoldChange) of <i>KLF15</i> and <i>MECOM</i> identified through RNA-seq and qRT-PCR.	167

LIST OF TABLES

Table. No.	Table legends	Page No.
Chapter II		
2.1	List of reported key genes dysregulated in GBC patients.	17
2.2	List of reported altered tumor suppressors and oncogenic miRNA identified in GBC.	20
2.3	List of OMICs scale GBC datasets available at NCBI and ENA databases.	30
Chapter III		
3.1	Cycling conditions for Real-Time PCR.	53
3.2	List of primer sequences of the hub DEGs and DE-lncRNA used in qRT-PCR.	53
3.3	Filtration steps to identify novel lncRNAs from RNA-Seq datasets.	55-56
Chapter IV		
4.1	The sample information of the RNA-seq datasets selected for Case Study 1.	71
4.2	The <i>Zsummary</i> and <i>medianRank</i> preservation (pres) values of the significant nonpreserved modules identified in GBC, HCC, and ICC networks.	76
4.3	The top five significantly enriched KEGG pathways associated with nonpreserved modules identified from the GBC, HCC, and ICC networks.	77
4.4	List of top five hub DEGs identified from nonpreserved modules of GBC, HCC, and ICC through intramodular connectivity analysis.	78
4.5	List of top five hub DEGs identified from nonpreserved modules of GBC, HCC, and ICC through PPI network analysis.	79

4.6	The Top five significant biological processes associated with DEGs identified from nonpreserved modules of GBC and control networks.	86
4.7	The Top five significant KEGG pathways associated with DEGs identified from nonpreserved modules of GBC and control networks.	87
4.8	List of top five hub DEGs identified in GBC from nonpreserved modules using intra-modular connectivity analysis. The DEGs with the highest interaction (weight) with the other DEGs in the modules are considered as hubs.	88
4.9	List of top five hub DEGs identified in GBC from nonpreserved modules using PPI network analysis.	89
4.10	List of significant hub DEGs identified in GBC compared to GSD with three different follow-up periods through PPI network analysis.	96-97
4.11	Clinical information on the GBC and GSD tissue samples collected through surgical resection and USG-guided biopsy	102
4.12	Summary of reads generated from GBC and GSD samples generated through transcriptomic sequencing.	104
4.13	List of hub DEGs identified through PPI network analysis by taking the consensus of five network topology measures.	111
Chapter V		
5.1	The statistics of correlation networks constructed with DEmRNA -DE-lncRNA pairs and DEmRNA-DE-nlncRNA pairs.	129
5.2	List of top five enriched pathways linked with DE-mRNAs coexpressed with hub DElncRNA identified in GBC and GBC+GS group. The pathways in bold font represent common enriched pathways identified between KEGG and MsigDB database.	132

5.3	List of top five enriched pathways linked with DE mRNAs coexpressed with hub DE lncRNA identified in GBC and GBC+GS group. The pathways in bold font represent common enriched pathways identified between the KEGG and MsigDB database.	133
5.4	List of hub DE lncRNAs and their parental gene information and genomic location.	135
5.5	List of top five miRNAs identified in GBC and GBC+GS groups through the construction of ceRNA networks. The asterisk symbol represents the miRNAs conserved in all the ceRNA networks.	137
Chapter VI		
6.1	List of top ten hub DE-TFs identified from TF-TG and TF-lncRNA regulatory networks in the GBC group. The asterisk symbol indicates the shared DE-TFs identified in both TF-TG and TF-lncRNA networks.	160
6.2	List of top ten hub DE-TFs identified from TF-TG and TF-lncRNA regulatory networks in the GBC+GS group. The asterisk symbol indicates the shared DE-TFs identified in both TF-TG and TF-lncRNA networks.	161

LIST OF ABBREVIATIONS

Abbreviations	Full form
76GS:	76 gene signatures
AAR:	Age-adjusted rates
AJCC:	American Joint Committee on Cancer
ASR:	Age-standardized rates
BBCI:	Dr. B. Borooah Cancer Institute
BLAST:	Basic local alignment search tool
BP:	Biological processes
BTS:	Biliary tract system
CAMs:	Cell adhesion molecules
cDNA:	Complementary DNA
ceRNA	Competitive endogenous RNA
CHOL:	Cholangiocarcinoma
CIS-BP:	Regulatory Sequence Analysis Tool
COAD:	Colorectal adenocarcinoma
CPC2:	Coding potential calculator 2
DEGs:	Differentially expressed genes
DE-lncRNA:	Differentially expressed lncRNA
DEmRNA:	Differentially expressed mRNA
DE-nlncRNA:	Differentially expressed novel lncRNA
ECM:	Extracellular matrix
EMT:	Epithelial-mesenchymal transition
ENA:	European nucleotide archive
EPC:	Edge percolated centrality
ETENBR:	End-to-end and Beyond RNA-Seq analysis pipeline
ETENLNC:	End-to-end Novel LncRNA analysis pipeline
FASTP:	Fast preprocessing
FASTQC:	FAST Quality Check
GB:	Gallbladder
GBC:	Gallbladder cancer
GBC+GS:	Gallbladder cancer with gallstones

GCN:	Gene coexpression network
GEO:	Gene expression omnibus
GLM:	Generalized linear model
GLOBOCAN:	Global cancer observatory
GS:	Gallstones
GSD:	Gallstone disease
GSD10:	GSD with a follow-up period of more than 10 years
GSD3:	GSD with a follow-up period of 1-3 years
GSD5:	GSD with a follow-up period of 3-5 years
HBCs:	Hepatobiliary cancers
HCC:	Hepatocellular carcinoma
HISAT2:	Hierarchical indexing for spliced alignment of transcripts
HSV1:	Herpes simplex virus 1
ICC:	Intra hepatic cholangiocarcinoma
KEGG:	Kyoto Encyclopedia of Genes and Genomes
KLF:	Kruppel-like factors
KS:	Kolmogorov Smirnov test
LIHC:	Liver hepatocellular carcinoma
LncRNA:	Long noncoding RNA
Log2FC:	Log2 Fold Change
LoH:	Loss of heterozygosity
MCC:	Maximum clique centrality
MCODE:	Molecular Detection Complex
MF:	Molecular functions
miRNA:	microRNA
MLR:	Multinomial logistic regression
MNC:	Maximum neighbourhood component
MRE:	miRNA response element
mRNAs:	Messenger RNAs
MSI:	Microsatellite instability
MsigDB:	Molecular Signature Hallmark Database
NATs:	Natural antisense transcript
NCBI:	National Center for Biotechnology Information

NCCN:	The National Comprehensive Cancer Network
ncRNAs:	Noncoding RNAs
NE:	North-East
NF-kB:	Nuclear factor-kappa B
NGS:	Next generation sequencing
ORF:	Open reading frame
PAAD:	Pancreatic adenocarcinoma
Padj:	P-adjusted value
pcRNAs:	Protein-coding RNAs
PDJ:	Pancreaticobiliary duct junction
PPI:	Protein-Protein Interactions
PROMPTs:	Promoter upstream transcripts
PWM:	Position weight matrix
qRT-PCR:	Quantitative Real-Time PCR
RIN:	RNA integrity number
RSAT:	Regulatory Sequence Analysis Tool
SSSSH:	Swagat Super Specialty Surgical Hospital
STAD:	Stomach adenocarcinoma
TCGA:	The Cancer Genome Atlas
TERC:	Telomerase RNA component
TFs:	Transcription factors
TF-TG:	Transcription factor-Target gene
TF-lncRNA:	Transcription factor-Target lncRNA
TGF-β:	Transforming growth factor-beta
TNF-alpha:	Tumor necrosis factor-alpha
TNM:	Tumor Node Metastasis
TOM:	Topological Overlap Matrix
TRN:	Transcriptional Regulatory network
TSG:	Tumor suppressor gene
USG:	Ultrasonography
WGCNA:	Weighted Gene Co-expression Network Analysis
XMEs:	Xenobiotic metabolism enzymes
