Systems-level investigation of liver transcriptome in physiology and pathology

Thesis submitted in partial fulfilment of the requirements for the degree of

Doctor of Philosophy in

BIOINFORMATICS

by

MANISRI PORUKALA

20162147

manisri.porukala@research.iiit.ac.in

International Institute of Information Technology, Hyderabad

(Deemed to be University)

Hyderabad - 500 032, INDIA

MAY, 2024

International Institute of Information Technology

Hyderabad, India

CERTIFICATE

It is certified that the work contained in this thesis, titled "Systems-level investigation of liver transcriptome in physiology and pathology" by Manisri Porukala (Roll no: 20162147), has been carried out under my supervision and is not submitted elsewhere for a degree.

_____ _____

Date Adviser: Dr. Vinod P. K.

To Amma, Daddy, chelli

ACKNOWLEDGMENTS

My journey of PhD is a contribution of many people. It's my mother, Ramadevi and my supervisor, Dr. Vinod P. K., who stood by me every single day and helped me pull myself together. I am grateful to my father, Ravinder, who imbibed discipline in me and constantly supported me in pursuing higher education. My special thanks to my younger sister, Kranthi Sree, who kept me motivated and supported me through thick and thin. I am thankful to my schoolteachers, Siddhi Kolarkar, Indira, Kranthi Deshpande and Grace Wilson, who laid the foundation for my career.

I express my deepest gratitude to my guide, Dr. Vinod P. K., who taught me what research truly means. He helped me recognize my potential to pursue research. His timely suggestions and valuable insights streamlined my problem-solving ability. His dedication and patience in detailed and thoughtful discussions led to new ideas and shaped the thesis work. I am thankful to him for introducing me and allowing me to explore research areas beyond the thesis work. His personality qualities of discipline, composure, and respect towards elders have undoubtedly set an example for me to become a better person.

I thank my thesis examiners, Prof. K. V. Venkatesh, Dr. Gopalakrishnan Bulusu, Dr. Ram Rup Sarkar, and Prof. Shekhar Mande for their critical reviews and suggestions. I thank the faculty of the Centre for Computational Natural Sciences and Bioinformatics for their guidance and for fostering the research environment. I thank the Dept. of Biotechnology, Govt. of India, and International Institute of Information Technology, Hyderabad, for providing financial support during my PhD. I thank my interview panel members, Dr Nita Parekh and Prof. Deva Priya Kumar, for approving my selection and providing an opportunity to be a part of IIIT-Hyderabad. I am extremely thankful to doctoral committee member Dr Nita Parekh for defending my inconsistent progress when required, which allowed me to demonstrate my potential towards the closing of my PhD. I thank Dr. Semparithi Aravindan for his technical assistance. I am thankful to the administration staff of IIIT-Hyderabad for their official assistance.

I strongly feel my PhD is an outcome of the long-term collective efforts of my late grannies Savitri and Dr C. N. Sunandha, who inculcated the importance of education in my mother.

I thank my group members of the Computational Systems Biology Lab, Dr Nishtha Pandey, Rami, Akansha and Ramya, for providing a pleasant work environment. Nishtha has been a great wall of support personally and professionally throughout my journey at IIIT-Hyderabad. I am lucky to have fellow PhD students Prashanthi, Ramya, Sanchari and Broto for their guidance when I was going through a rough patch. My evening walking partner, Sri Lakshmi, has been a great source of joy. Antarip, Sohini, Navneet, Mohan, Shruti, Subba, Shwetha, Tanashri, Vijay, Rohit, Pradeep, Aravind, Cyrin, and Vishal for creating a positive lab environment. I am fortunate to have decade-old friendships with Mansa, Meghana, Hemanth, Akhil, Sreenu, Pankaj, Krishnakanth, Srinath, Yashaswi, and Dr Sai Kiran, who stood by my side.

Needless to mention the *Prof. C.R. Rao Road* that I take pride in quoting "The Road *Taken*". I am privileged to have been a part of this road for 14 years, right from my Bachelor's days (August - 2010) to PhD (June - 2024) covering three different institutes: University of Hyderabad (I. M. Sc. Systems Biology, 2010 - 2015), C.R.Rao Advanced Institute of Mathematics, Statistics and Computer Science (AIMSCS) (Junior Research Fellow, 2015 - 2016), and International Institute of Information Technology (PhD Bioinformatics, 2017 - 2024) on its way.

List of publications

Journal Publications

1. **Porukala M**, Vinod P. K., Gene expression signatures of stepwise progression of Hepatocellular Carcinoma, *PLoS ONE, 18(12): e0296454, (2023)*

2. **Porukala M**, Vinod P. K., Network-level analysis of ageing and its relationship with diseases and tissue regeneration in the mouse liver, *Scientific Reports* (13), 4632 (2023)

3. **Porukala M**, Vinod P. K., Systems-level analysis of transcriptome reorganization during liver regeneration, *Molecular Omics*, (18), 315-327 (2022)

Poster presentations

1. "Mathematical modelling of transcriptional network of liver regeneration", Manisri Porukala, Vinod P. K., at Society for Mathematical Biology (SMB) 2021 Annual Meeting (Virtual), University of California Riverside (UCR), June 13-17, 2021.

2. "Rewiring of liver tissue network with ageing and its crosstalk with associated liver pathologies and regeneration", Manisri Porukala, Vinod P. K., at 21st International Conference on Bioinformatics (InCoB) (Virtual), King Abdullah University of Science and Technology (KAUST), November 21 - 23, 2022.

3. "Molecular insights into the progression of precancerous to cancer state in Hepatocellular carcinoma using gene co-expression network analysis", Manisri Porukala, Vinod P. K., at 42nd Annual Conference of the Indian Association of Cancer Research, Tata Memorial Centre, Navi-Mumbai, India, January 12 – 16, 2023.

Abstract

Metabolism is an integral part of cellular physiology, with the liver as the central organ for a wide range of metabolic functions and homeostasis. The liver, unlike other organs, has a remarkable capacity to regenerate after partial loss of its mass, thus maintaining a constant liver-to-body weight ratio to preserve homeostasis. In an injury-free liver, the turnover of regeneration is very slow, but in the presence of an injury or perturbation, the regenerative response is triggered as a reparative strategy. Perturbations interfering with liver functions and disturbing the homeostatic state have serious repercussions leading to metabolic disorders like hepatic steatosis, non-alcoholic fatty liver disease (NAFLD), non-alcoholic steatohepatitis (NASH) and hepatocellular carcinoma (HCC). Ageing is one risk factor that increases the susceptibility to these diseases. The regenerative ability of the liver has been used to treat these diseases in the form of liver transplantation and partial liver resection, but recurrence of the disease is often observed after a few years. Understanding the molecular network that controls liver function and its regenerative ability in health and disease is crucial for developing clinical applications. The emergence of high throughput omics technologies provides a scope to develop a systems-level understanding of the disease. In this direction, we attempt to understand the pathophysiology of the liver by adopting systems biology approach for omics data interpretation.

Tissue homeostasis and regeneration depend on the reversible transitions between quiescence (G0) and cell proliferation. During regeneration, the liver needs to maintain the essential metabolic tasks along with the fulfilment of metabolic requirements for hepatocyte growth and division. To understand the regulatory mechanisms involved in balancing the liver function and proliferation demand after injury or resection, we analyzed RNA sequencing temporal data of liver regeneration after two-thirds partial hepatectomy (PH) using network inference and mathematical modelling approaches. The reconstruction of the dynamic regulatory network revealed the overall temporal coordination of metabolism, RNA splicing and cell cycle during liver regeneration. A temporal shift in the gene expression pattern corresponding to increased hepatocyte proliferation and decreased hepatocyte function is observed with HNF4A as a key transcriptional activator. Based on these key observations, we developed a mathematical model of the HNF4A regulatory circuit, which showed the emergence of different states corresponding to compensatory

metabolism, proliferation, and epithelial-to-mesenchymal transition. We showed that a mutually exclusive behaviour emerges due to the bistable inactivation of HNF4A, which controls the initiation and termination of liver regeneration and different population-level behaviour. Through our approach of modelling a regulatory circuit from high-throughput gene expression data, we proposed a mechanistic explanation of different states observed in single-cell RNA sequencing data of liver regeneration.

The functional impairment of the liver with ageing reduces its regenerative capability and predisposes it to NAFLD and HCC. Mapping the molecular network of the liver encompassing these physiological (ageing) and pathological conditions may help to understand the crosstalk of ageing with different liver diseases. We performed network-level analyses by integrating mouse transcriptomic data with protein-protein interaction (PPI) network. A sample-wise analysis using network entropy measure was performed, which showed an increasing trend with ageing and helped to identify ageing genes based on local entropy changes. To gain further insights, we also integrated the differentially expressed genes (DEGs) between young and different age groups with the PPI network and identified core modules and nodes associated with ageing. Finally, we computed the network proximity of the ageing network with different networks of liver diseases and regeneration to quantify the effect of ageing. Our analysis revealed the complex interplay of immune, cancer signalling, and metabolic genes in the ageing liver. We found significant network proximities between ageing and NAFLD, HCC, liver damage conditions, and the early phase of liver regeneration with common nodes, including NLRP12, TRP53, GSK3B, CTNNB1, MAT1 and FASN. A common theme involving pathways in cancers connects ageing, regeneration and liver diseases. Overall, our study maps the network-level changes of ageing and their interconnections with the physiology and pathology of the liver.

While the regulated process of liver regeneration is crucial for damage-induced repair, its dysregulation may lead to HCC, the most common type of liver cancer. Understanding the molecular pathogenesis of HCC sequentially from precancerous state to cancer may improve prognosis and treatment strategies. In this direction, we studied the transcriptomics data of tumour samples and their adjacent normal samples in different premalignant states from HCC patients undergoing transplantation or partial hepatectomy. A hierarchical approach was adopted to identify modules, pathways, and genes relevant to the prediction of disease-free survival (DFS). Modules of co-expressed genes were identified from (a)

only tumour samples, (b) premalignant and tumour samples collectively (premalignant-to-tumour), and (c) all normal and premalignant samples and their association to patient clinical characteristics was studied. Modules and genes related to the cell cycle, immune system, ribosome, and liver metabolic pathways served as good predictors for DFS using tumour samples. DFS modules were also associated with treatment (transplantation and resection) given to patients. An overall decrease in liver function and immune pathways but an increase in cell cycle activity was observed. The progression from premalignant to tumour is accompanied by variations in the extent of downregulation of liver function and immune system and an increase in cell cycle activity, bringing about variability in patient outcomes. Interestingly, we showed that modules and genes based on normal and premalignant samples also serve as good predictors of DFS. An increase in immune and cell cycle activity was observed in premalignant conditions, which suggests that tumour-matched normal samples already contain multiple signatures relevant to predict the DFS. This analysis revealed a shift in immune activity from premalignant to tumour state. THBD, a classical marker for dendritic cells, is a good predictor of DFS at the premalignant stage. Further, cell cycle genes related to microtubules, kinetochores, and centromere are altered in the premalignant stage, which are DFS genes in tumour samples. This study captured the dynamic changes in gene expression of various biological processes in the stepwise progression of HCC.

Although liver regeneration capacity has been applied as a clinical intervention tool in the form of liver transplantation, its functional stability determines the post-treatment outcome. Understanding the molecular mechanisms driving long-term stability (normal) or rejection of the transplanted graft may play an important role in improving the post-transplantation outcome. This may help identify molecular markers to predict post-operative rejection. We performed differential gene coexpression analysis of transcriptomic data from post-operative liver biopsies of normal and rejection patients. This analysis revealed the rewiring of gene coexpression patterns pertaining to liver function, immune pathways and cell cycle. The modules of immune and cell cycle genes showed intact within-module coexpression in rejection samples compared to normal samples. We identified EGR2, MTHFD2, CD52, and CD38 from immune module and RRM2, TOP2A, ZWINT, TYMS, MAD2L1, ANLN, PRC1, and CDKN3 from cell cycle module as novel features distinguishing normal and rejection samples which require further validation.

Overall, this thesis attempts to generate systems-level insights into liver physiology and pathology that may have implications in clinical settings. The pathophysiology of the liver is studied using transcriptome data from experimental mouse models to data from HCC patients. The work on liver regeneration provided insights into regulatory mechanisms governing the balance between liver function and proliferation. The network-level analysis of liver ageing, regeneration and pathological conditions showed the network-level influence of ageing on different liver-associated conditions and helped to identify key candidate pathways and genes commonly dysregulated across these conditions. Further, the work on HCC patients helped to characterise the different trajectories for progression from premalignant to tumour state and predict DFS based on both premalignant and tumour samples. This provides a scope for early detection and prognostication of HCC patients. We also analysed the liver transplantation dataset to identify features distinguishing normal and rejection samples.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABBREVIATIONS

BA	Bile acid
BCAA	Branched-chain amino acid
BIoGRID	Biological General Repository for Interaction Datasets
BN	Bottleneck
BW	Betweenness
CS	Cirrhosis
DBN	Dynamic Bayesian network
DC	Dendritic cell
DD	Deceased donor
DEG	Differentially expressed gene
DFS	Disease-free survival
DGCA	Differential gene co-expression analysis
DH	Dysplastic nodule high grade
DIP	Database of Interacting Proteins
DL	Dysplastic nodule low grade
DREM	Dynamic regulatory events miner
EAD	Early allograft dysfunction
EC	Endothelial cells
ECM	Extracellular matrix
EGF	Epidermal growth factor
ER	Endoplasmic reticulum
ES	Enrichment score
FBA	Flux balance analysis
FDR	False discovery rate
FH	Fibrosis high
FL	Fibrosis low
G0 phase	Quiescent phase
G1 phase	Gap 1 phase
GCN	Gene coexpression network
GEO	Gene expression omnibus
GO	Gene Ontology
GRN	Gene regulatory network
GSEA	Gene set enrichment analysis
GSH	Glutathione
HBV	Hepatitis B virus
HCC	Hepatocellular carcinoma
HCV	Hepatitis C virus
HFD	High-fat diet
HGF	Hepatic growth factor
HIR	Hepatic injury and regeneration
HMM	Hidden Markov Model
HSC	Hepatic Stellate Cells
IOHMM	Input output Hidden Markov Model
IRI	Ischemic reperfusion injury
KC	Kupffer cells

KEGG	Kyoto Encyclopedia of Genes and Genomes
LD	Live donor
LT	Liver transplantation
M phase	Mitotic phase
MCC	Maximal clique centrality
MCODE	Molecular Complex Detection
MM	Module membership
MMP	Matrix metalloproteinases
NAFLD	Non-alcoholic fatty liver disease
NASH	Non-alcoholic steatohepatitis
ODE	Ordinary differential equation
PH	Partial hepatectomy
PHLF	Post hepatectomy liver failure
PPI	Protein-protein interaction
RAT	Rejection associated transcripts
S phase	Synthetic phase
SAH	S-adenosylhomocysteine
SAM	S-adenosylmethionine
STRING	Search Tool for the Retrieval of Interacting Genes/Proteins
SVD	Singular value decomposition
TAA	Thioacetamide
TCMR	T-cell mediated rejection
TF	Transcription factor
TFG	Transforming growth factor
TH	Total hepatectomy
TOM	Topological overlap matrix
t-SNE	t-distributed Stochastic Neighbour Embedding (t-SNE)
VAT	Visceral Adipose Tissue
WGCNA	Weighted gene co-expression network analysis

Chapter 1 Introduction

1.1 Background

Metabolism is an integral part of cellular physiology, with the liver being a major metabolic organ performing various functions. The liver also has a remarkable ability to regenerate after a partial loss of its tissue or toxin-induced chemical insult instead of scar formation. This regenerative capacity may play a crucial role in host survival in all cases of liver damage. Perturbations that interfere with liver functions or its regenerative capacity following an injury may lead to serious metabolic disorders. Therefore, it is crucial to understand the molecular mechanisms of the liver's regenerative capacity and metabolism to reduce the impact of metabolic disorders.

1.2 Functions of the liver

The liver performs a wide variety of functions that are broadly categorized into three classes: biotransformation, storage, and synthesis[1] **(Figure 1.1)**. Major functions of the liver include the production of hormones, bile acids and vitamins, clearance of toxins, decomposition of erythrocytes and synthesis of plasma proteins[2]. It also regulates blood volume, endocrine signalling pathways and macronutrient metabolism. The liver efficiently maintains post and pre-prandial blood glucose levels by ensuring a proper balance between the storage and synthesis of glucose[3]. It is a major site for the synthesis of cholesterol that is converted to bile acids. The active secretion of bile acids by the liver into the intestine is essential for the absorption of dietary lipids and fat-soluble vitamins[4]. The liver is unique in its capacity to dispose of nitrogenous waste from the body through the urea cycle since it is the only organ that harbours all the required enzymes[5].

The liver is the first tissue to get directly exposed to gut-absorbed products via the portal blood due to its strategic location between the intestinal tract and the rest of the body, hence making it crucial for the clearance of toxins or ingested drugs[4,6]. In addition, there is a constant influx of nutrients from the intestinal tract into the liver. The anatomical location of visceral adipose tissue (VAT) in the mesentery connects it directly to the liver, forming the adipose tissue-liver axis of communication[7]. This axis forms the key metabolic circuit handling macronutrient uptake, processing, transport, and storage. The "Cori cycle" operating between the liver and skeletal muscle is an adaptation mechanism that recycles lactate produced during anaerobic glycolysis in muscles to glucose in the liver by the action

of gluconeogenic enzymes[8]. Tissue-to-tissue crosstalk occurs between the liver and pancreas (β-cells), both these tissues being involved in glucose homeostasis and lipid metabolism[9]. Thus, crosstalk between the liver and other organs helps maintain homeostasis and energy balance, the derailing of which leads to multiple pathologies.

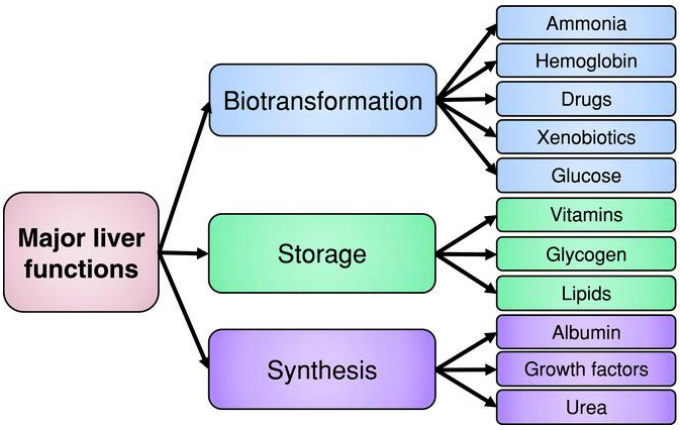


Figure 1.1: Major functions of the liver in the body[1].

1.3 The microarchitecture of liver

The microarchitecture of the liver consists of repeated hexagonal units called lobules ranging between 0.5 mm in diameter in mice and 1 mm in humans[10,11]. Each lobule has a central vein and hepatocyte cords radiating from it towards the portal triads consisting of the portal vein, hepatic artery, and bile duct **(Figure 1.2)**. Hepatocyte cords comprise hepatocytes (parenchymal cells), the major cell type in the liver, that are arranged in a single layer spanning the radial lobule axis and interspersed with sinusoids, which carry blood from hepatocyte cords to the central vein **(Figure 1.2)**. Oxygenated blood received via the hepatic artery mixes with deoxygenated blood from the portal vein. The sinusoids connecting the hepatic artery and portal vein with the central vein can be divided into different metabolic zones: periportal zone (zone 1), intermediate zone (zone 2) and perivenous/pericentral zone (zone 3). As blood moves through, the oxygen concentration decreases, and it is lowest in the perivenous zone. Other non-parenchymal cell types in the liver include endothelial cells, stellate cells, and Kupffer cells. Endothelial cells are present in the sinusoids. Kupffer cells are macrophages present within the lumen of the liver sinusoids. Stellate cells are present between the region of hepatocytes and sinusoidal endothelial cells called the Space of Disse.

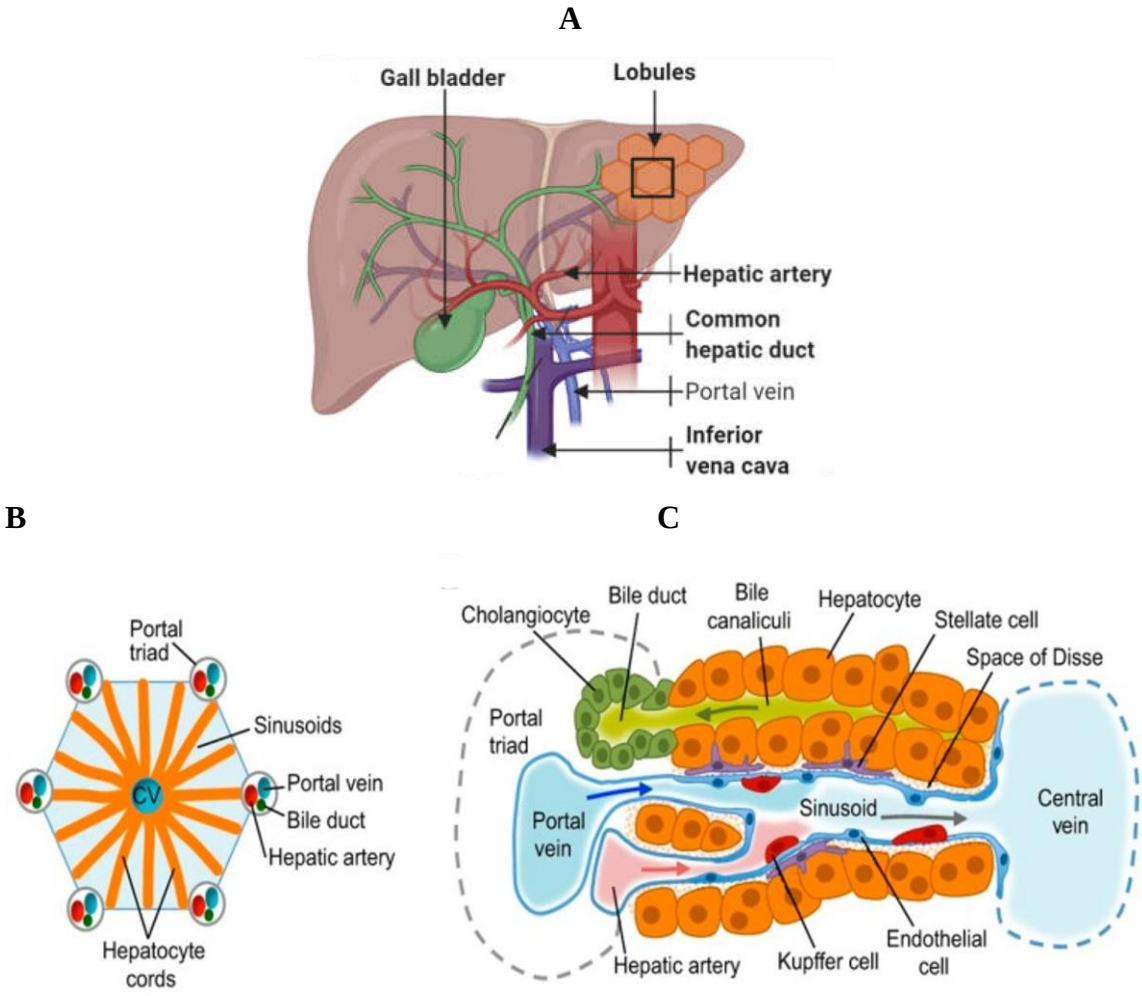


Figure 1.2: (A) Structural organization in adult liver[12]. (B) The functional unit of the liver lobule consists of a central vein (CV), portal triad, biliary duct and hepatic artery[13]. (C) Organization of liver cell types within the lobule[13].

1.4 Liver regeneration

Tissue regeneration involves repeated cellular renewal occurring during normal ageing (tissue homeostasis) and the restoration of damaged tissue after injury or disease (reparative regeneration)[14]. The homeostatic renewal in adult organs can be driven by different processes, including differentiation of recruited or resident stem cells and progenitor cells, cell proliferation and transdifferentiation of existing cells. The liver has the innate ability to regenerate after injury, allowing it to restore the original liver size. The homeostatic process in the absence of injury can be viewed as maintaining the normal liver size. The regenerative capacity of the liver has long been recognized both in animals and humans[15]. In the normal liver, less than 2% of hepatocytes proliferate at a given moment, and it is a very slow process[16].

The regenerative capacity of the liver plays a crucial role in host survival to liver insult by toxins or viruses[14]. Despite being terminally differentiated, parenchymal cells in the residual liver tissue quit the G0 phase of the cell cycle and undergo a transition from quiescence to proliferation[17]. Partial hepatectomy (PH) of rodent livers is frequently used to study liver regeneration mechanisms[18]. The liver can use distinct mechanisms for repair depending on the mode of injury[14,19]. Deciphering the source of cells for liver regeneration using novel lineage tagging of hepatic cells and single-cell RNA transcriptomics has received considerable attention, and there is a debate about the source of new cells in the liver[20,21]. Different regions of the liver lobule exhibit differences in hepatocyte turnover. Recent studies indicate that zone 2 has the highest proliferative activity and is a primary source of new hepatocytes during homeostasis and regeneration[22].

1.4.1 Different models of liver regeneration

The animal models of liver regeneration are continuously evolving and are helping advance clinical practice **(Figure 1.3)**. Hepatotoxins such as CCl_4, D-galactosamine, and acetaminophen are usually employed to study drug-induced liver regeneration[23]. Exposure of the liver to hepatotoxin results in a necrotic injury that mimics certain liver diseases[24]. Administration of CCl_4 is the widely used animal model to understand liver injury, repair and regenerative response. A single dose of CCl_4 causes hepatotoxicity mediated by the action of Cyp2e1, resulting in oxidative damage to DNA, proteins, lipids, and carbohydrates. A peak of parenchymal necrosis rises 24 hours after CCl_4 exposure, followed by regeneration. Long-term exposure to CCl_4 activates quiescent hepatic stellate cells (HSC) and impairs hepatocyte proliferation, leading to fibrosis. D-galactosamine is a potent hepatotoxin that depletes UTP, inhibiting RNA and protein synthesis. It causes diffuse inflammation and necrosis of liver cells, as seen in viral hepatitis. Spotty necrosis and lymphocyte infiltration are observed 6-24 hours after exposure, followed by a recovery phase at 72 hours. The regenerative capacity is weaker than that of the CCl_4 model. Acetaminophen, a commonly used antipyretic and analgesic, causes acute liver failure upon overdosage due to the accumulation of free radicals and Kupffer cells (KC) activation. Cell proliferation is observed between 12-48 hours after its overdose. TAA is a hepatocarcinogen known to induce perivenous necrosis followed by regeneration. Ethanol is another hepatotoxin used to investigate liver regeneration.

In addition to drug-induced liver regeneration, 2/3rd PH is also known for its remarkable capacity to regenerate into a functional organ. It is a model of acute liver injury[25]. Unlike the drug-induced models, the PH model is an ideal case of regeneration process that proceeds in a disease-free environment with less chance of post-operative liver failure[24]. The multi-lobar structure of the liver makes the surgical resection amenable[26]. Apart from this, liver regeneration post-PH is characterized by a definite starting point (the time of resection), synchronized cell cycle, less injury and low infiltration of inflammatory cells[27]. Although pharmacological models are clinically relevant in mimicking liver diseases, they require dose standardization for better reproducibility[24]. While drug-induced liver injury provides a wealth of information on aberrant liver regeneration under diseased conditions, the PH model sets the stage for normal liver regeneration. Understanding the differences between these two models of regeneration may have implications for developing healthy regeneration and remodelling strategies in chronic liver disease[28]. For the present study, we investigated the mode of liver regeneration using the PH model.

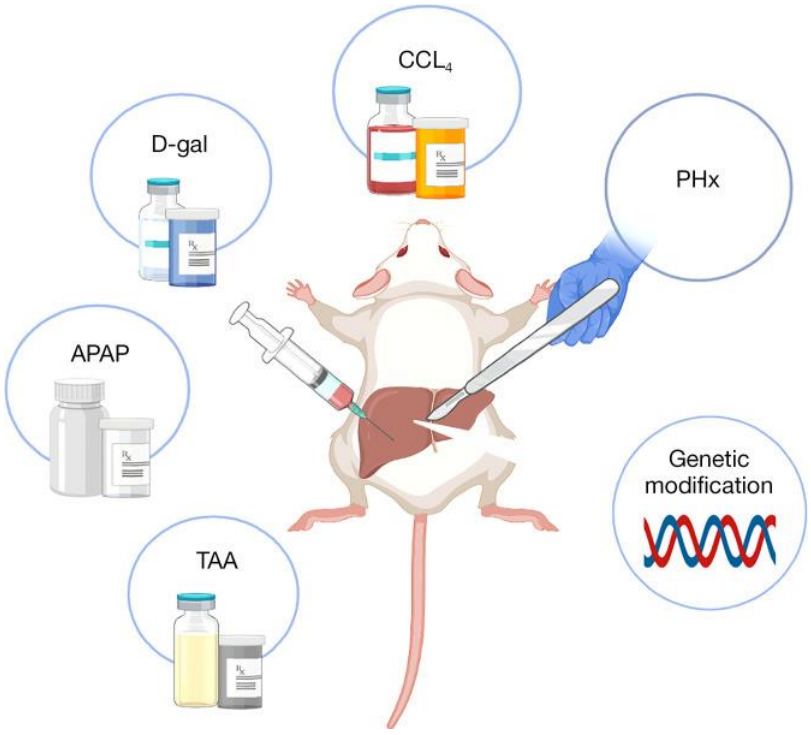


Figure 1.3: Different models of regeneration[23]

1.5 Partial hepatectomy model of liver regeneration

The landmark event showcasing the regenerative potential of the liver dates back to the year 1931 when the model of liver resection was carried out by Higgins and Anderson[18]. The

first model of 2/3rd PH of liver lobes was performed on white rats, which showed that remnant liver stimulates a synchronous hepatocyte proliferation to gradually restore the lost mass. This classic rat model of 70% PH involved the removal of two anterior lobes - the median lobe, ML and the left lateral lobe, LLL **(Figure 1.4)**.

The remnant rat liver, after 2/3rd PH, increases its weight to 45% of the original weight within 24 hours. It takes another 48 hours to attain 70% of the original liver weight. After this quick response, the rate of increase slows down, reaching the preoperative liver weight 1-2 weeks after PH[15,29] **(Figure 1.5)**. An important characteristic feature of liver re-generation post partial resection is 1) the resected lobes do not grow back; instead, the remaining lobes take up the responsibility of restoring the lost mass, and 2) the remnant liver can also respond to repeated PHs. Therefore, liver regeneration is regarded as a process of compensatory growth to restore function rather than attaining the form[15]. In addition to 70% PH, other models of 90, 95 and 97% PH are also used to study liver regeneration and acute liver failure. The pace of liver regeneration is proportional to the amount of hepatic tissue removed. However, it is not uniform within different ranges of resection. In rats, resection ranging between 40-80% mounts a sharp response in DNA replication with synchronized cell proliferation, while too large (>85%) or too small (<30%) PH models show delayed and unsynchronized cell proliferation.

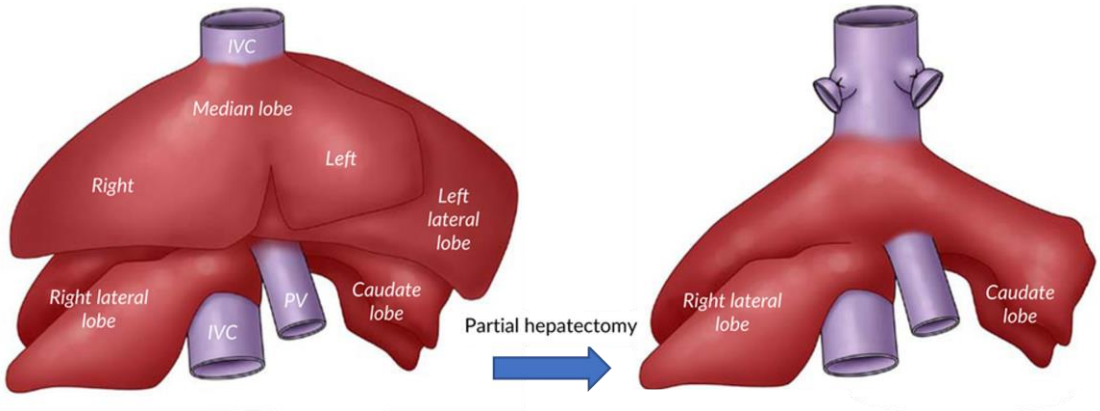


Figure 1.4: Rodent model of liver resection. The classical model of 70% partial hepatectomy with the removal of the median lobe (ML) and left lateral lobe (LLL)[24].

Continuous labelling experiments in weanling rats showed that hepatocytes enter the cell cycle after PH in weanling rats with two peaks of hepatocyte replication 20 and 35 hours post-PH[30]. In young adult rats, the peak of proliferation is 24 hours after PH[31], while in mice, it is attained between 36 and 40 hours[15,32]. The process of liver regeneration is characterized by three phases: priming, proliferative and termination phases. PH performed

on genetically modified animals provides a wealth of information on the molecular underpinnings spanning these three phases of liver regeneration.

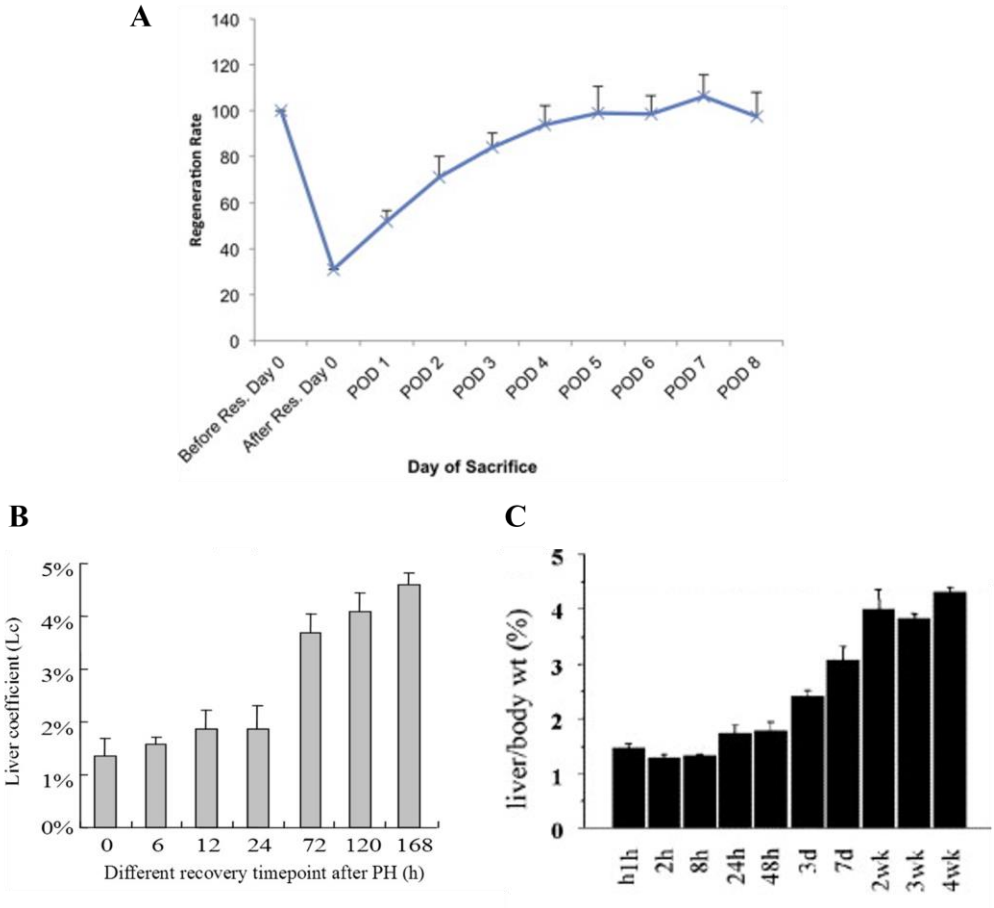


Figure 1.5: Temporal dynamics of liver regeneration after 2/3rd PH. (A) Rate of liver regeneration (in %) after 2/3rd PH in the rat[33]. (B) & (C) Liver-to-body weight ratio after 2/3rd PH in rat[34] and mouse[35] respectively. POD – Post-operative day.

1.5.1 The priming phase

After PH, hepatocytes are the first of the cell types to quit the quiescent G0 phase and enter the cell cycle. The loss of 2/3 of the liver tissue with PH alters the portal supply, resulting in an increase in portal blood pressure per unit of tissue on the remnant 1/3 of the tissue[24]. This marks the beginning of the priming phase by increased activity of uPA in the endothelial cells within 5 minutes of PH. Activated uPA triggers a fibrinolytic cascade mediated by MMPs and plasmin to degrade the extracellular matrix (ECM). The complement cascade proteins of innate immunity also send out early signals that bind to their receptors on Kupffer cells (KC). This activates the NFKB signalling within KC to alleviate the inhibiting effect of IKB on NFKB[36,37], thus enabling NFKB to translocate to the nucleus and turn on transcription of its target genes TNFα, IL6 and Cyclin D1[38]. TNFα

and IL6 are important proinflammatory cytokines in liver regeneration. TNFα triggers an autocrine signal on KC to amplify NFKB signalling while it acts on hepatocytes in a paracrine fashion[39]. TNFα in hepatocytes activates the JNK pathway and immediate early genes[40]. Likewise, the binding of IL6 to its receptor activates STAT3 and ERK pathways to target the activation of Cyclin D1 and p21. This sequence of events takes place within 30 minutes to 1 hour of PH, culminating in the expression of about a hundred genes that are otherwise dormant in the normal liver[37,38,41]. Although these initial signals push hepatocytes from G0 to the G1 phase, cells cannot move past "the restriction point" in G1 and proceed further in the cell cycle. Moving beyond the restriction point is brought by growth factors[42]. Therefore, the priming phase is reversible, which by itself is incapable of triggering DNA replication and thus sensitizes hepatocytes to proliferation signals so that they move further in the cell cycle.

1.5.2 The replicative phase

The proliferative phase of liver regeneration begins 5 hours after PH once hepatocytes cross the restriction point[43]. Parenchymal hepatocytes respond to hepatocyte growth factor (HGF) secreted by non-parenchymal cells[44]. HGF adhered to ECM is released due to matrix remodelling and is activated by uPA. Activated HGF binds to its receptor c-Met on hepatocytes after 1 hour of PH, subsequently inducing the downstream pathways PI3/Akt/mTOR, Ras-Raf-MEK-ERK[45]. This ultimately leads to the activation of cyclins and CDKs required for DNA synthesis. At the same time, EGFR is stimulated by EGF[24]. There is also crosstalk between these two activation pathways, with c-Met augmenting EGFR activity. Combined disruption of these two signalling pathways completely abolishes liver regeneration[46].

Apart from proinflammatory cytokines and growth-factor mediated proliferation, other signalling pathways essential for successful liver regeneration include Wnt/β-catenin pathways, Notch, and Hippo signalling pathways. Notch signalling is activated as early as 15-30 minutes after PH and brought by cell-cell interaction since Notch receptors couple with membrane-bound ligands. This interaction culminates in the transcription of proliferative response genes Myc, and Cyclin D1[38]. Wnt signalling is activated 1-3 hours post-PH with ligand Wnt synthesized by KC in response to TNFα and activates downstream signalling when it binds to the receptor on hepatocytes[38,47]. This liberates the β-catenin from its inhibitory complex, and finally, activated β-catenin triggers the transcription of target

genes. The Hippo signalling pathway has both proliferative and anti-proliferative roles in liver regeneration[48]. The Hippo pathway is a repertoire of kinases that, in the absence of injury/PH, inhibits Yap from inducing transcription of proliferation-related target genes.

1.5.3 The termination phase

Restoration of normal liver-to-body mass ratio of ~2.5% three days post-PH marks the beginning of the termination phase[35,49]. Within a week after PH, the proliferation ceases[50]. About 85% of the liver-to-body mass ratio is attained by 14 days of resection[18,46]. While the first two phases are dedicated to recovering the lost liver mass, the termination phase acts as a brake to arrest cell proliferation when the remnant liver reaches the "required size". Unlike other organs, the liver has an inbuilt tracker, "hepatostat", that always adjusts its size to 100%, which is required to maintain homeostasis[51]. Moreover, the termination phase ensures that the regeneration process is in the right direction, preventing the onset of cancer[52]. TGFβ is the main regulator of the termination phase and is secreted by KCs and HSCs. Its levels peak around 72 hours post-PH, and it binds to TGFR on hepatocytes to phosphorylate R-Smad proteins, which stimulate the transcription of cell cycle inhibitors that inhibit CDK activity[52,53]. Another layer of regulation aiding termination is brought by the negative feedback loop imposed by SOCS3 on STAT3[54–56]. In the early stage, activation of YAP with concomitant downregulation of Mst1/2 and Lats1/2 was shown to support proliferation, while the opposite trend is seen towards the termination phase. YAP mRNA levels peak around 32-48 hours post-PH, coinciding with S and M phase peaks along with its targets[57]. Therefore, the Hippo signalling pathway plays an important role in controlling the liver size post-PH. Detailed mechanisms pertaining to the termination phase are still under the active investigation stage.

1.5.4 Mathematical model of liver regeneration

The information on different phases of the liver regeneration process was used to develop the mathematical models. The model proposed for rat liver regeneration incorporates the regulation by cytokines, growth factor and ECM and reproduces the known phenomenology[58] (Furchtgott *et al*., 2009). Modifications of this model consider cell growth, bone marrow cells, species-specific differences, liver repair scenarios and non-parenchymal cell states[59–62] (Cook *et al*., 2018; Cook *et al*., 2015; Pedone *et al*., 2017; Young and Periwal, 2016). However, the functional relationships between metabolism and liver regeneration remain unexplored at the systems level[63]. These models mostly

considered data from transgenic and gene-knock studies in rodents. In recent years, efforts have been made to comprehensively study the liver regeneration process using high-throughput omics technology to obtain mechanistic insights into liver regeneration.

1.6 Clinical and translational relevance of liver regeneration

The experimental regeneration model of 2/3rd PH has an important clinical application. Rodents accomplish the restoration of full functional liver mass post-PH within 7 – 10 days, while in humans, it takes about 60 – 90 days[32,64]. Liver resection is a choice of intervention for HCC and benign liver disease[65]. In HCC patients, resection of the liver mass containing tumours is widely used[62]. The unique regenerative ability of the liver tissue post-PH also showed the road ahead for organ transplantation. Patients affected with liver diseases such as cirrhosis can be treated using the transplantation of partial liver tissue from a live donor, which is followed by regeneration in both the donor and recipient[62] **(Figure 1.6)**. However, the dynamics of normal recovery may be impaired under the conditions of acute liver failure, pre-existing fibrotic condition or if the transplanted organ or the remnant liver post-resection is too small[28,66]. Transplantation from deceased donors is also prevalent.

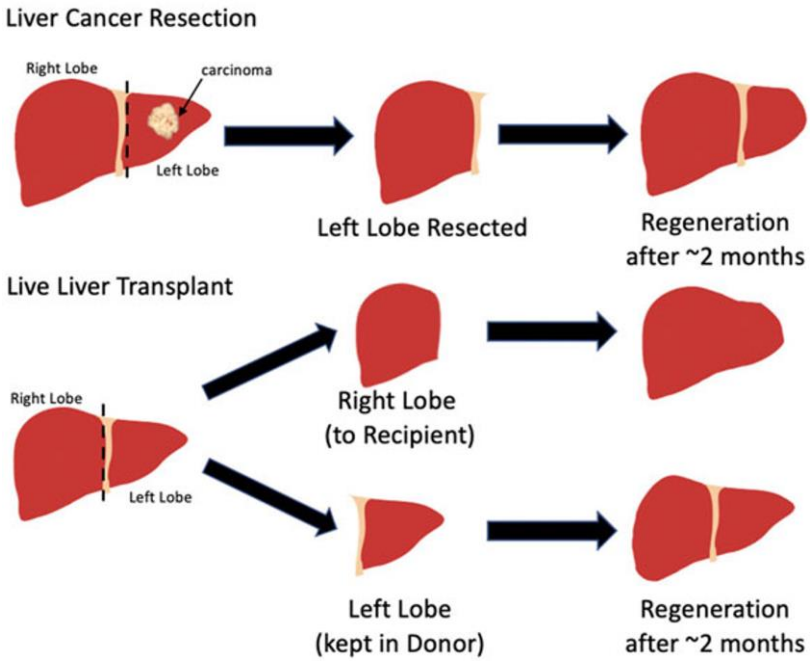


Figure 1.6: Clinical application of liver regeneration. Partial resection following identification of a tumour (top) and live liver transplantation (bottom) from donor to recipient[67].

The regenerative capacity of the remnant liver determines outcomes ranging from normal restoration of the tissue to impaired restoration or liver failure[68]. Deviation in the ability of

the post-operated liver to perform its functions and maintain homeostasis results in post-hepatectomy liver failure (PHLF)[69]. Patient's clinical history, such as weight, diabetes, and age, contribute to PHLF[70,71]. The fraction of the remnant liver and its functional capacity plays a key role in determining PHLF. Therefore, candidates for surgery are meticulously chosen to minimize PHLF in patients (recipient) and avoid detrimental outcomes in living donors. On the other hand, the donor's liver is carefully scrutinized to ensure it is free from liver diseases prior to transplantation.

Studies on human liver regeneration are mostly based on non-invasive techniques of whole-organ imaging and serum analysis, which attempt to associate clinical outcomes with pre-operative markers. The inability to readily access liver tissue from human subjects limited the study of molecular mechanisms of liver regeneration mostly to animal models. To be able to fill the void between in vivo models and clinical realism, constant refinement of animal models based on feedback from human studies is required[28].

1.7 Liver ageing and diseases

The liver's ability to maintain homeostasis and capacity to regenerate declines progressively with ageing. A decrease in liver function with ageing disturbs metabolic homeostasis, leading to lipid accumulation and increased prevalence of steatosis in old age[7]. This can lead to a spectrum of liver damage conditions called non-alcoholic fatty liver disease (NAFLD), which ranges from benign form steatosis (NAFL) to more aggressive form, non-alcoholic steatohepatitis (NASH)[72]. NAFLD is linked to multiple factors, including obesity, insulin resistance and Type 2 diabetes (T2D). NASH is characterized by lobular inflammation and hepatocellular ballooning in the presence or absence of fibrosis[73]. NASH can progress into progressive fibrosis and cirrhosis, which can increase the risk of HCC. Persistent inflammation can scar the tissue around the liver, leading to fibrosis. Cirrhosis occurs after years of inflammation, and the liver gets permanently scarred and lumpy, which is irreversible and can lead to liver failure **(Figure 1.7)**. HCC is the most common type of liver cancer, and the main risk factors for its development include infection by hepatitis B virus (HBV) and hepatitis C virus (HCV). NAFLD is also becoming a more frequent risk factor for the development of HCC[74].

The early stage of chronic liver diseases remains undetected until they emerge in a symptomatic state where the liver is sufficiently damaged, impacting the function. At this stage, the treatment that is available is often limited to liver transplantation[72]. Currently,

there are no FDA-approved drugs due to limited knowledge of mechanisms that drive the progression from steatosis to HCC.

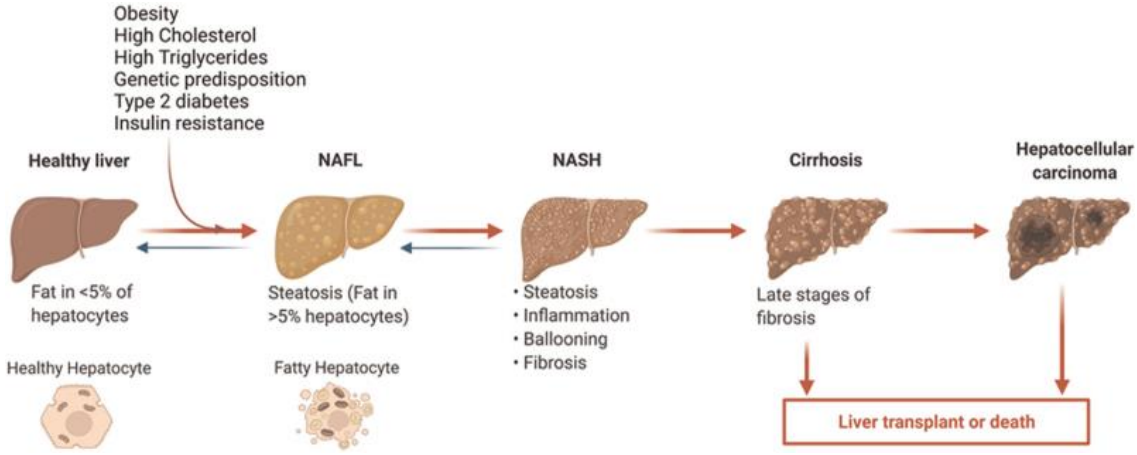


Figure 1.7: Spectrum of liver damage conditions[75]

1.8 Omics data to understand the genome-wide changes in diseases

Model organisms such as *Mus musculus* and *Rattus norvegicus* have been used as surrogates in translational research to study human diseases. While most of the studies deciphered underlying mechanisms from knockout studies, the emergence of high throughput omics technologies provides a scope to develop a systems-level understanding of the disease. The genome-wide changes can be mapped at different levels, including genomics, epigenomics, transcriptomics, proteomics and metabolomics at tissue and single-cell levels **(Figure 1.8)**. The variation in DNA sequence can be mapped to obtain insights into the contribution of genetic variants to a particular disease. The non-genetic (DNA methylation, histone modification and chromatin organization) contribution to the phenotype can also be explored using epigenomics. These genetic and epigenetic variations can impact the gene expression that can be quantified using transcriptomics, which measures mRNA level changes occurring in the cells. Proteomics measures protein expression levels and different types of posttranslational modifications. The various biological functions are regulated by proteome, which is also targeted by drugs. Metabolomics quantifies the changes in metabolites due to changes at the expression level of enzymes that catalyze various biochemical reactions.

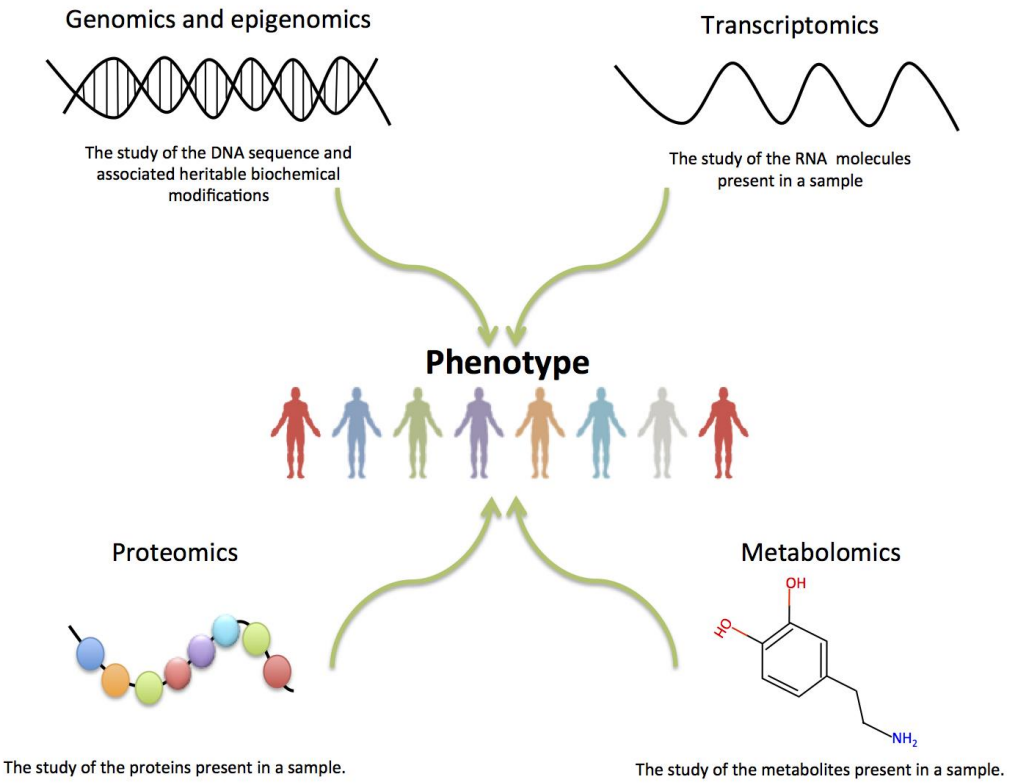


Figure 1.8: Integration of omics technologies at different levels to provide a complete model of the biological system under study[76].

1.8.1 Transcriptome data

Quantifying the transcript levels bridges the gap between the genome and functional molecules that regulate the cells[77]. Transcriptome data of different liver-associated conditions obtained using Microarrays and RNA sequencing (RNA-Seq) techniques from mouse models and patients are widely available compared to proteomic data, the other product of the genome.

Microarray

This technology is based on the principle of hybridization between each strand of DNA from a tissue or cell line, referred to as 'target', and its corresponding complementary DNA sequences immobilized on a solid surface. This solid surface is an ordered array of microscopic DNA spots, each containing thousands of copies of a specific DNA sequence, usually corresponding to a short segment of a gene. The arrayed material is equivalent to the probe used in a typical Northern blot analysis and hence generally called a 'probe'. The array platforms fall into categories, namely, printed cDNA microarray and in situ synthesized oligonucleotide microarray. Probes of cDNA microarray are amplified products

of polymerase chain reaction, while oligonucleotide probes are synthesized by photolithography[78].

A microarray experiment begins with the extraction of mRNA or amplified cDNA from the tissue or cell samples to be analyzed. This is followed by fluorescent labelling and hybridization onto the DNA array, leaving it for incubation overnight. Non-specific hybrids are removed by washing. Laser illumination of fluorescent-tagged hybrids causes excitation that is detected by a confocal scanner. The digital image generated by the scanner is processed by specialized software to convert each spot intensity to a numerical reading. The scanner first locates each spot by its shape, then integrates the intensities inside and finally estimates the surrounding background noise. This gives an integer-valued reading as the readout of the concentration of the target sequence in the sample to which the probe is spotted[79].

The dual-channel assay is the widely employed experimental design for microarray to compare the relative abundances of mRNA between two different samples, for example, control and test **(Figure 1.9)**. Two different target samples are labelled with different colour dyes (Cy3 and Cy5) and co-hybridized to the same array, resulting in competitive binding of the target to the arrayed probes. The reading obtained from image scanning is transformed to a ratio of relative abundance of the target sequence from the test sample with respect to the reference (control) sample. In a single-channel assay, both samples are labelled with the same dye and hybridized to different arrays. The intensity data for each probe indicates the relative abundance with respect to the labelled target **(Figure 1.9)**. The co-hybridization of target samples in dual-channel allows direct comparison of gene abundance between two samples, minimizing the error due to variability in manufacturing of multiple arrays. The advantage of one-channel microarray lies in its design simplicity and flexibility, facilitating comparison across samples and between groups of samples. Due to single-sample hybridization per array, the effect of the aberrant sample affecting the data derived from other samples is minimized in a one-channel assay. The comparison of data generated from arrays of different experiments is possible in a one-channel assay. The variability in array processing is the main source of error due to variability in one-channel assay, which can be reduced by introducing sufficient biological and technical replicates[80].

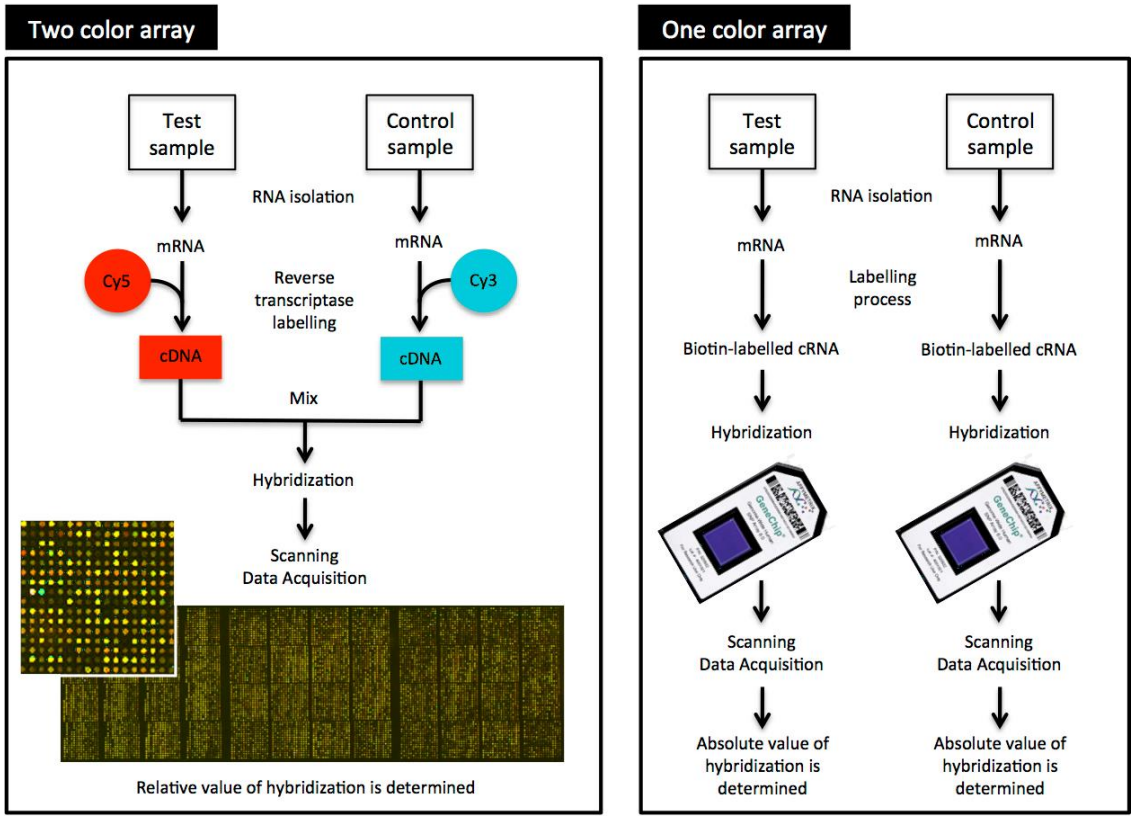


Figure 1.9: Expression analysis by two- and one-colour microarray: Two-colour microarray uses two different fluorescent dyes hybridized on a single chip, while one-colour microarray uses a single dye and two chips to generate expression profiles[81].

RNA-seq

It is a sequence-based approach to quantifying and studying the transcriptome of a sample. The total RNA is extracted from the tissue of interest and converted to a library of fragmented cDNA (200-500 base pairs). Sequencing adapters are ligated to both ends of cDNA. This is followed by an amplification step, and the fragments are sequenced either from one end or both ends, resulting in single-end or paired-end reads, respectively. These reads range between 30-400 base pairs (bp) in length, depending on the technology. With the data obtained in the form of sequenced reads, the computationally intense sequence analysis follows **(Figure 1.10)**. The reads can be mapped to reference genome or undergo de novo assembly. De novo assembly of reads is useful to identify a genome-scale transcription map that gives information on transcriptional structure as well as gene expression levels[82].

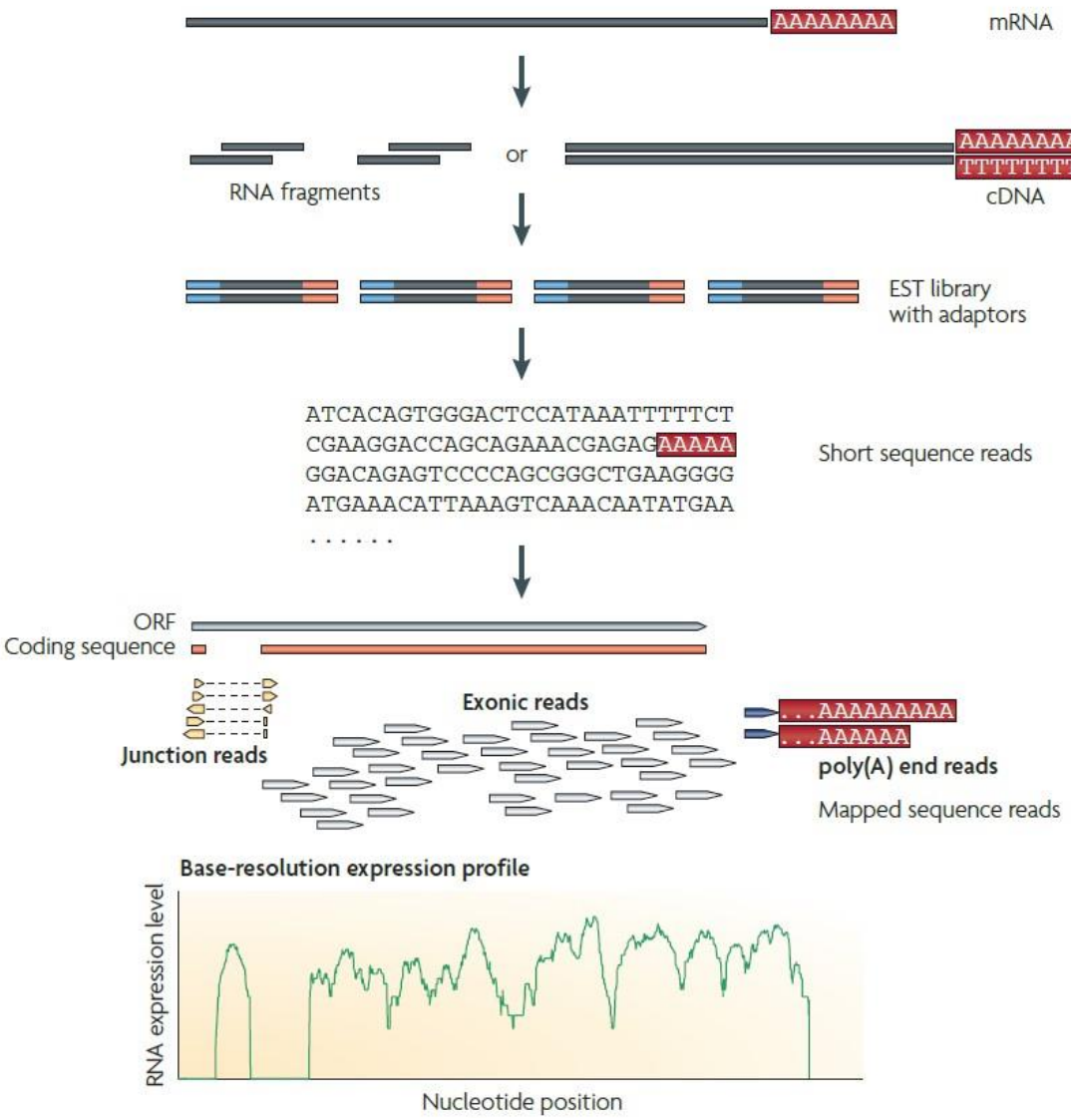


Figure 1.10: Outline of a typical RNA-seq experiment[82].

The quality of RNA extracted is verified using RNA Integration Number (RIN > 8) to minimize errors in the downstream steps. Prior to cDNA library construction, several manipulations are performed before proceeding with further steps. In order to catalogue different types of small RNAs (miRNA, piRNA, siRNA), the sequencing step can directly follow after the ligation of adapters. To enrich mRNA from rRNA, the fact that mature mRNA carries a poly(A) tail is used and captured by hybridizing to oligo dT beads. Larger RNA molecules must be fragmented into smaller pieces ranging between 200 and 500 bp to make them amenable to sequencing, which is important for accurate transcript assembly[83].

Comparing transcriptomic technologies

Microarray transcript detection is based on molecular hybridization, while RNA-seq captures the transcripts in samples by directly sequencing them. The level of gene expression corresponding to a transcript is given as a measure of continuous probe intensities in microarray as opposed to discrete digital sequence read counts provided by RNA-seq. This facilitates unlimited dynamic range giving a near true abundance of readouts of transcripts within cells and quantification of subtle changes in gene expression important for biological processes. The manufacturing of arrays (probes) for microarray requires prior knowledge of genome sequence. Moreover, standard microarray probes can cover only 20% of a gene on average.

RNA-seq enables quantification of expression from previously unannotated regions of the genome, unlike microarrays and helps discover novel transcripts. Although this limitation of microarray can be overcome by designing probes with overlapping regions of the genome, as in tiling arrays, it lacks full genome coverage. The single-base resolution of RNA-seq has given a revised view of known gene annotation by identifying gene boundaries and introns, new isoforms of the genes, and splicing events. In addition, isoform abundances can also be obtained. Splicing arrays designed for this purpose, on the other hand, requires prior information on genes and their known isoforms.

The advantage microarrays provide over RNA-seq is the low cost of implementation. Nevertheless, the cost of sequencing is decreasing. An important concern with RNA-seq is the depth of sequencing, which is defined as the average number of times a nucleotide base is measured in the pool of random raw sequences[84]. It gives an estimate of how many times a sample must be sequenced. Highly expressed transcripts can be detected at low sequencing depth, while moderate and rarely expressed transcripts require much higher depth. Another concern is the humongous amount of data generated by RNA-seq compared to that of microarrays[85]. Despite these differences, both platforms show a good correlation of gene expression profiles[86] with more DEGs being identified by RNA-seq[87].

1.9 Network-level analysis of transcriptome

Biological networks are widely used in systems-level analysis. They provide a framework for integrating omics data and extracting the underlying mechanism in physiology and pathology. In biological networks, nodes are represented by components such as genes, proteins etc, and edges define the relationship between nodes. These networks are reported to exhibit graph topological properties such as a scale-free topology, where few nodes have

a number of connections (degree k) that follow a power-law distribution. This means that there are only a few nodes that act as "hubs" with many connections, which are hypothesized to be important. Betweenness is another topological measure which characterizes the nodes based on the shortest path. Nodes with high betweenness are considered to be "bottleneck hubs" since they bridge a high number of shortest paths with other nodes. Thus, graph theoretical measures can potentially help to identify disease biomarkers and novel drug targets. The biological network analysis also aims to understand network organization into functional modules and their constituent genes[88], which can inform the biological mechanism related to the disease. Different biological networks include protein-protein interaction networks, gene regulatory networks, metabolic networks, and integrated networks. The integrated view of the biological network is given below **(Figure 1.11)**.

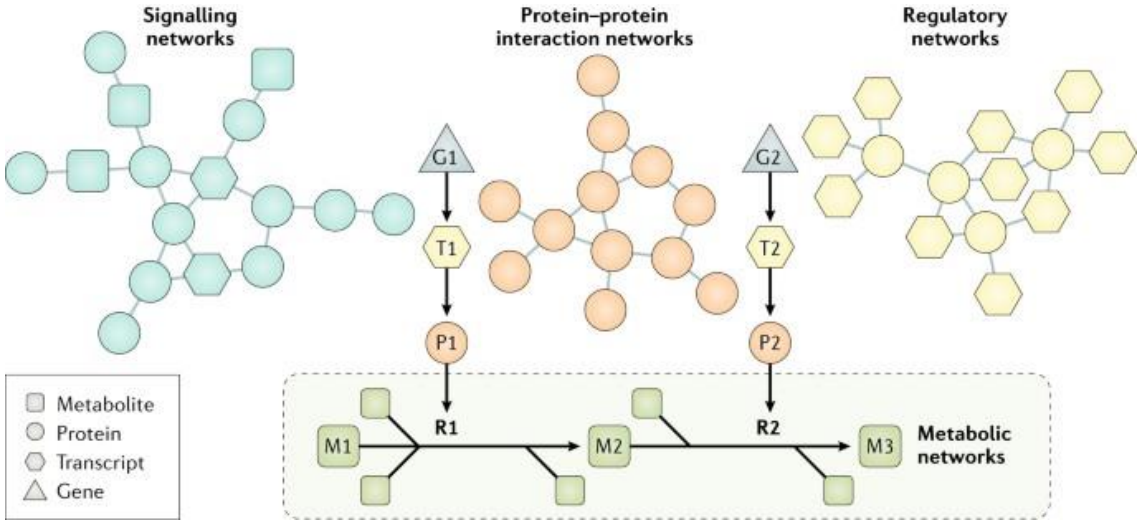


Figure 1.11: Integration of biological networks to understand the underlying mechanisms of biological functions[89].

1.9.1 Protein-protein interaction network

Many biological functions and disease-state may not be explained solely by considering individual genes or proteins. Differential gene expression analysis, to some extent, helps identify genes involved in a molecular mechanism but provides little insight into how they collectively work towards a physiological state. The underlying complex relationship among genes is not obvious from analysis based on single genes. Under the contexts where disruption of a process results from slight insignificant changes in the expression of numerous genes, it is necessary to look at the interactions[90]. A protein-protein interaction

(PPI) network encompasses the associations between two or more proteins that may be either due to physical interaction or participation within a common pathway[91].

Human PPI networks have been constructed into databases such as STRING, BioGRID, Database of Interacting Proteins (DIP) etc[89]. These PPINs have been used to sieve potential drug targets. Graph theoretical network measures such as degree have been applied to identify essential proteins. The PPINs provide general static interactome built from multiple contexts, of which only a subset may be relevant for a particular condition at a given time. Therefore, PPINs can be integrated with gene expression to obtain a dynamic picture of the interactome.

1.9.2 Metabolic networks

Cellular metabolism provides the cell with the energy required for survival and cell growth. For its functions, a cell takes up essential small molecules which are metabolized by cellular reactions either spontaneously or with the help of enzyme catalysis. These small molecules are interconverted from one form to another in a cascade of reactions called metabolic pathways. The ability of small molecules to participate in different reactions and the non-specificity of enzymes catalyzing multiple small molecules brings nonlinearity in metabolic pathways. This ultimately leads to strongly interwoven metabolic pathways, and the interplay between them forms a metabolic network. Therefore, a metabolic network is made up of all the biochemical reactions describing the relationship between metabolites and catalytic enzymes[92].

Graphically, the metabolic network of a cell can be represented by a bipartite graph of two node sets corresponding to metabolites and enzymes with edges connecting a metabolite node to an enzyme node if they participate in the same biochemical reactions. In a bipartite metabolic network, no two enzymes or no two metabolites are directly connected to each other. A metabolic network can also be represented by a unipartite enzyme (or rection) interaction graph **(Figure 1.12)**. The nodes of this network are enzymes, and an edge connects two enzymes if they share a common substrate/metabolite in the corresponding reactions[93].

The central focus of building networks for cellular metabolism is to understand network function from its structure using topological analysis. The degree distribution (probability distributions of edges per node) of the network was used to show the existence of scale-free

topology in metabolic networks as its global structural organization. Further, to capture the dynamic properties of the metabolic network, stoichiometric and kinetic modelling has been applied. These dynamical models help investigate fluxes of metabolites within metabolic pathways. Flux balance analysis (FBA), a constraint-based analysis of metabolic networks, is widely used and is based on the principle of conservation of mass. In FBA analysis, optimal reaction rates are calculated at a steady state under the constraint of maximizing biomass.

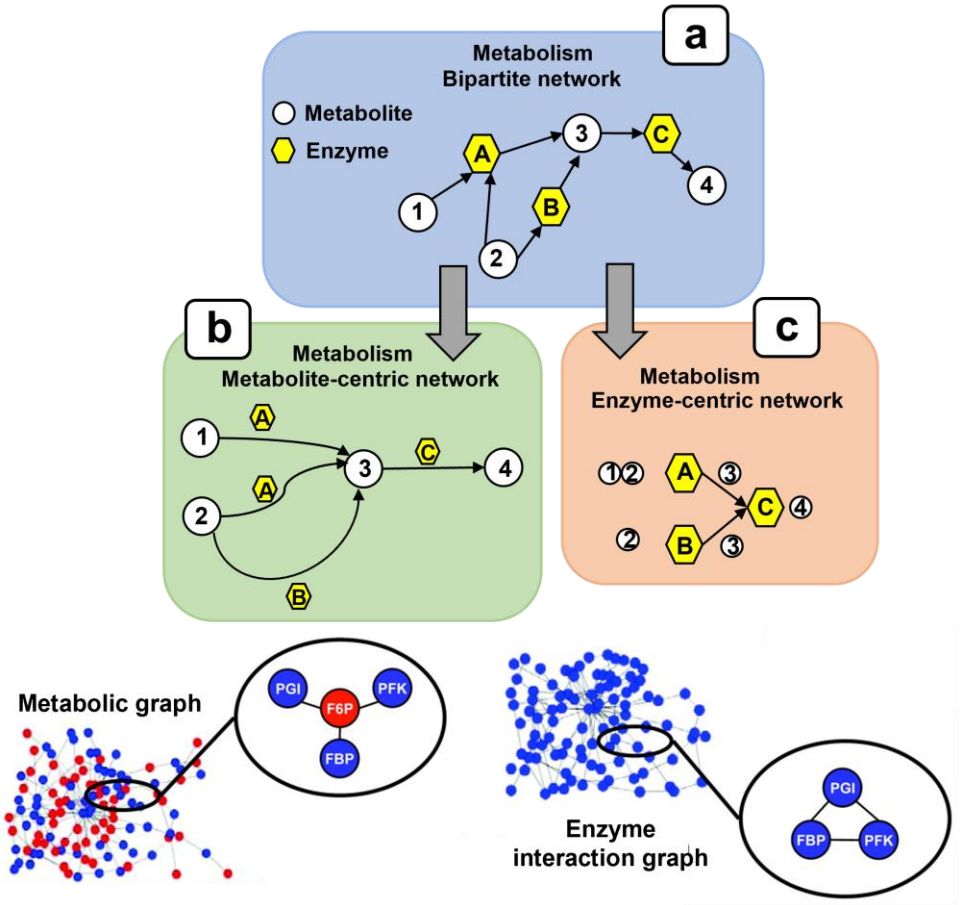


Figure 1.12: Transformation of a bipartite metabolic network (a) to a metabolite-centric unipartite network (b) and an enzyme-centric unipartite network (c)[94]. Below is an illustration of a metabolic and enzyme interaction graph depicting three enzymes and one metabolite of glycolysis[93].

1.9.3 Co-expression network

The high-throughput genome-wide expression techniques have undoubtedly laid the foundation stones for acquiring systems-level knowledge but pose a challenge to deriving the system structure from its components. The accumulation of large amounts of

transcriptomic data using microarray and RNA-seq techniques led to the reconstruction of the gene co-expression network (GCN) in different biological conditions. GCN gives an overview of the similarity between expression profiles of genes from a particular group of biological samples. The edge relationship is not determined in GCN, which contrasts with the gene regulatory network (GRN) that includes directed edges representing the causal relationships, including biochemical interactions and activation/inactivation[95]. However, co-expression modules are of biological interest since genes of a module can be regulated by the same transcriptional program, functionally related or map to the same pathway or protein complex.

Different methods have been developed for the construction of GCN. The following steps are involved in the construction of GCN from gene expression data. After initial preprocessing and filtering of genes, a similarity matrix (every pair of genes) is built by selecting suitable co-expression measures, followed by thresholding and network construction, identifying modules using clustering techniques and finally downstream analysis for biological insights **(Figure 1.13)**. Commonly used coexpression measures are Pearson correlation, Spearman's rank correlation, mutual information, and Euclidean distance[91].

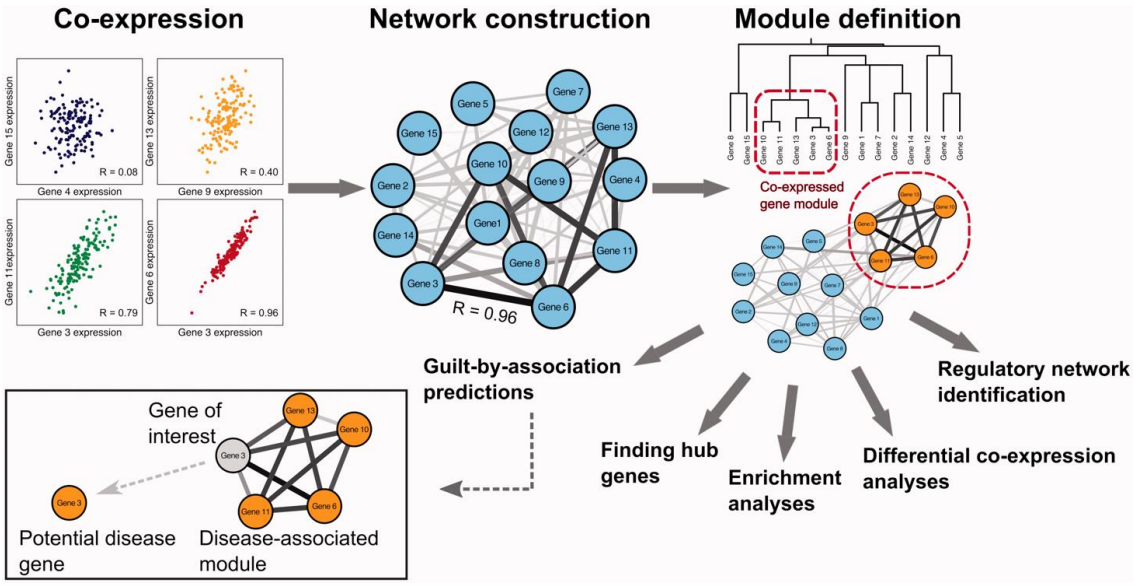


Figure 1.13: Step-by-step procedure for building gene co-expression networks from gene expression data[96].

After choosing a correlation measure, the GCN is built based on $m \times m$ similarity matrix S with entries $s_{ij} = |\, cor(i,j) \,|$, the correlation between two genes i and j scaled between 0 and 1. Such a similarity matrix based on absolute correlation is "unsigned". For a signed

similarity matrix $s_{ij} = \frac{1+ cor(i,j)}{2}$ with $s_{ij} < 0.5$ meaning negative correlation and $s_{ij} \geq 0.5$ meaning positive correlation. The similarity matrix is transformed to an adjacency matrix using an adjacency function. The choice of adjacency function gives an unweighted or weighted network, respectively, depending on whether hard thresholding or soft thresholding is used. Unweighted network using hard threshold τ with $a_{ij} \in \{0,1\}$ is given as

$$a_{ij} = \begin{cases} 1, & s_{ij} \geq \tau \\ 0, & s_{ij} < \tau \end{cases} \quad (1.1)$$

and weighted network using power-law-based soft threshold β with $a_{ij} \in [0,1]$ is given as

$$a_{ij} = |s_{ij}|^{\beta} \quad (1.2)$$

Soft thresholding reduces noisy correlations and emphasizes the disparity between strong and weak correlations.

The primary goal of GCN is to identify a subset of tightly connected nodes (modules). Topological overlap dissimilarity measure that quantifies interconnectedness between nodes was found to identify functionally relevant modules[97]. The topological overlap of two nodes is a reflection of their interconnectedness. It is given by

$$t_{ij} = \frac{l_{ij} + a_{ij}}{min\{k_i, k_j\} + 1 - a_{ij}} \quad (1.3)$$

where $l_{ij} = \sum_p a_{ip} a_{pj}$, which gathers all the shared neighbours between nodes i and j and sums up to zero if there are no shared neighbours, $k_i = \sum_p a_{ip}$ is the node connectivity of node i. Since $l_{ij} \leq \min(\sum_{p \neq j} a_{ip}, \sum_{p \neq i} a_{pj})$, it implies that $l_{ij} \leq \min(k_i, k_j) - a_{ij}$, finally rendering $0 \leq t_{ij} \leq 1$ like $0 \leq a_{ij} \leq 1$. If two nodes are connected, and the node with fewer links shares the same neighbours as the other node, then $t_{ij} = 1$. If the two nodes are not connected and do not share any neighbours $t_{ij} = 0$. Therefore, a_{ij} which is simply based on the correlation between nodes i and j is transformed into t_{ij} which also accommodates their connection strength with common neighbouring nodes. A toy example of a network with 11 nodes and the corresponding topological overlap matrix (TOM) is shown in **Figure 1.14**. This depicts how nodes with similar expression patterns can be clustered together. If two genes have a strong connection strength with the same set of

genes, they have a high TOM score. To apply TOM for extracting modules, the similarity matrix t_{ij} is converted to dissimilarity matrix $d_{ij}^{t} = 1 - t_{ij}$.

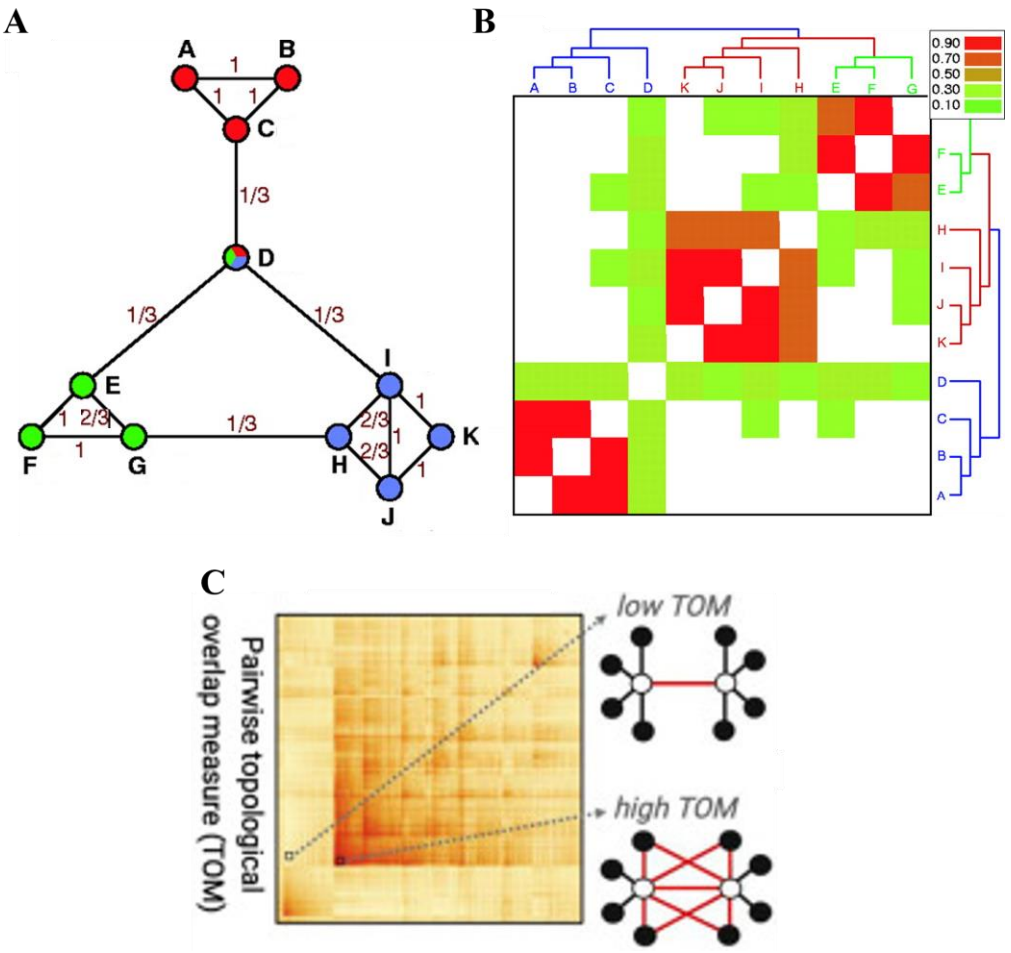


Figure 1.14: The underlying modularity of a complex network. (A) Topological overlap of a small hypothetical network. The edge weights represent TOM between every pair of nodes, and the parenthesis next to each node corresponds to the clustering coefficient[97]. (B) Clustering of the heatmap of TOM for the network shown in (A). The tree structure in the dendrogram shows three distinct modules. The module 'EFG' is topologically closer to the module 'HIJK' than to the module 'ABC'. (C)An illustration of TOM highlighting low and high-scoring TOM regions[98].

1.9.4 Gene-regulatory (GRN) network

Transcriptional regulation is at the core of shaping organismal complexity and diversity[99]. Transcription factors (TF) govern the activity of genes in the nucleus, which in turn guide protein synthesis from the ribosomal machinery in the cytoplasm, the prime location for biochemical reactions within the cell. These sequences of events eventually influence the activity of the cell. Some of the proteins themselves act as TF and return to the nucleus to

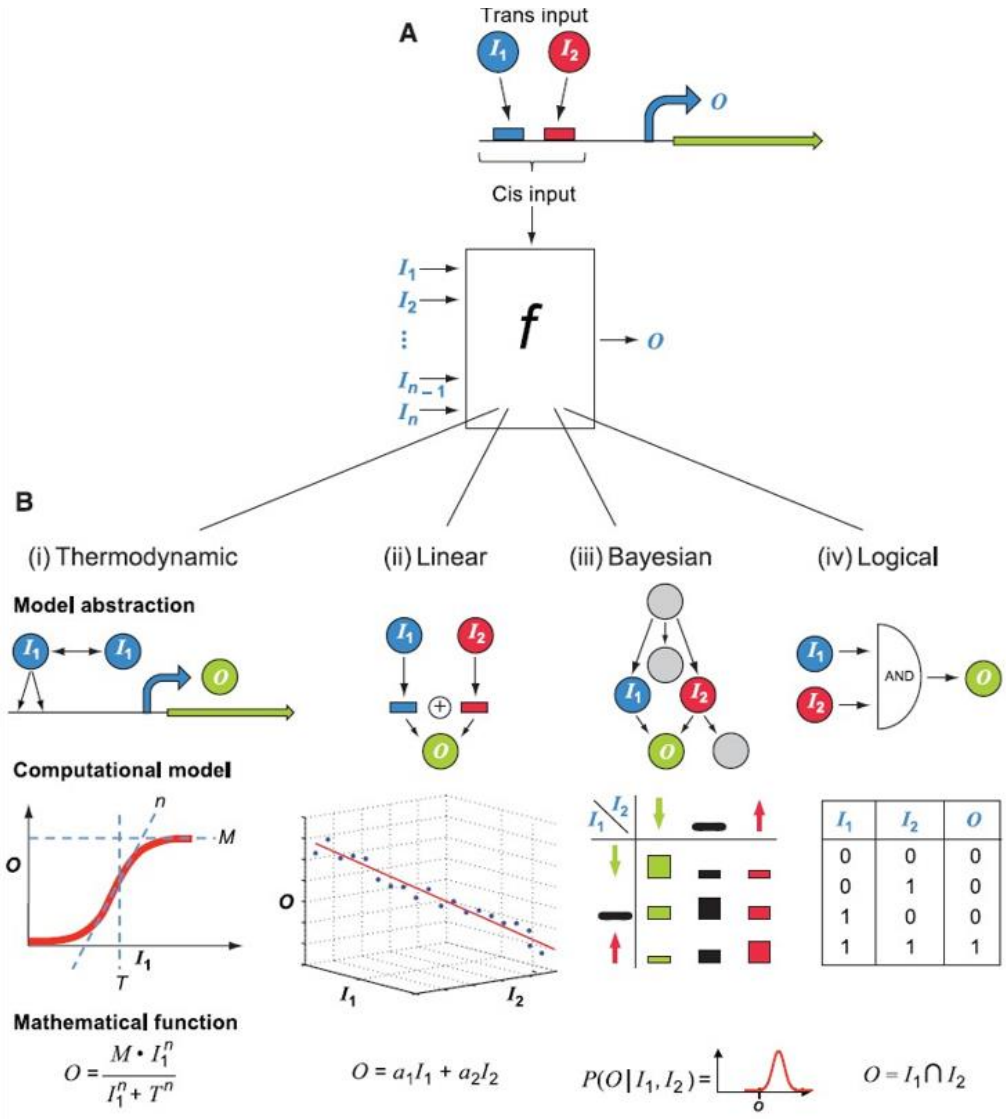
control several other genes(s) by binding to the regulatory elements. A plethora of transcription factors (trans inputs) can bind to the regulatory sequences (cis inputs) of the target gene, with the final expression (output) being determined by the combination of bound TFs[100]. The amounts of gene products and their temporal order of synthesis are crucial for the functioning of the cell[101]. This complex process of controlling gene expression is represented as a gene regulatory network[102], which is a collection of molecular species and their interactions[101].

The advancement of powerful experimental techniques targeting gene regulation studies has led to an unmet demand to model gene regulatory networks. Models have been built in this direction, defining the biological system at different levels of abstraction to answer relevant questions pertaining to a) the behaviour of a system under different conditions, b) the dynamics of the system under the loss/change of its component(s), c) robustness of the system under extreme circumstances[101]. A wide variety of methods like Boolean networks, probabilistic Boolean networks, Petri nets, Bayesian networks, and non-linear ordinary differential equations have been developed for modelling gene regulation[103] **(Figure 1.15)**.

1.9.4.1 Boolean network

These are the simplest models of GRN construction where a single gene is represented by a binary/Boolean variable ON (1) or OFF (0), depending on whether it is active or inactive, respectively. A Boolean network is a directed acyclic graph (DAG). It is represented as a tuple G(V, F) where $V = (Y_1, Y_2, \dots\dots, Y_n) \epsilon \{1,0\}^n$ is a vector of Boolean variables, F is a set of Boolean functions $F = \{f_1, f_2, \dots, f_n\}, f_i: \{0,1\}^n \mapsto \{0,1\}$. Each node $Y_i, i \in \{1,2, \dots, n\}$ corresponds to a gene in the network and is individually associated with a Boolean function $f_i = f_i(Y_{i_1}, Y_{i_2}, \dots\dots, Y_{i_k}) \in F = \{f_1, f_2, \dots\dots, f_n\}$. Parent nodes $Y_{i_1}, Y_{i_2}, \dots, Y_{i_k}$ of each node Y_i, serves as the input to the corresponding Boolean function f_i. Boolean operations on the genes are based on "AND", "OR", and "NOT" logical gates, which show the regulatory action of TF on the target genes **(Figure 1.16)**. At a given time *t,* the state of the Boolean network is defined as $S(t) = (Y_1(t), Y_2(t), \dots\dots, Y_n(t))$, which is a 0-1 vector describing levels of all genes called the global state. This state *S(t)* is updated to *S(t+1)* in the next time step *t+1* based on Boolean functions *F*. The level of every gene Y_i at time *t+1* is updated as $S_{t+1}(Y_i) = f_i(S_t(Y_{i_1}), \dots\dots, S_t(Y_{i_k}))$. The global state changes synchronously in discrete and equally spaced time steps, with states of all genes being updated simultaneously depending on the levels of their regulators in the previous time step.

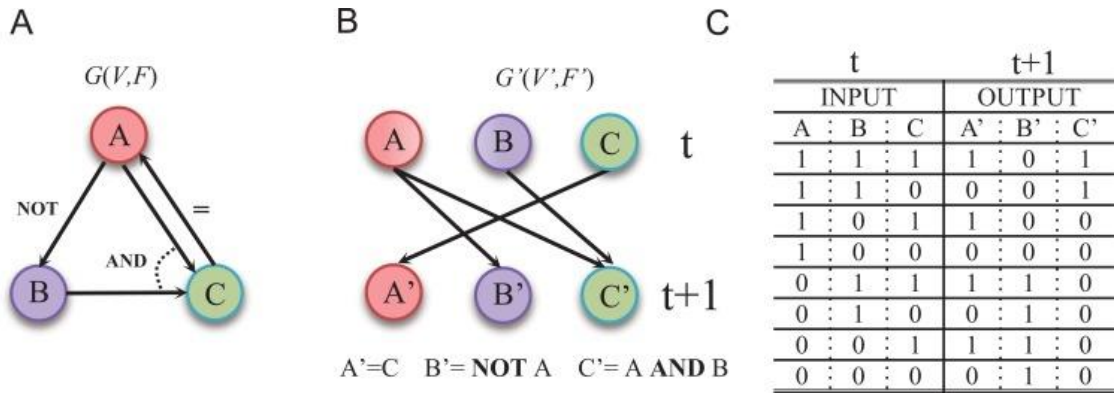
Thus, updates are deterministic in nature. This transition state is represented by wiring diagram G`(V`, F`) by adding an additional node $V`_i$ for every node V_i and linking the parent



I_1	I_2	O
0	0	0
0	1	0
1	0	0
1	1	1

Figure 1.15: Different models for Gene regulatory networks. (A) A gene-regulation function describing the action of inputs elements TF, regulatory sequences (I) to generate output (O), i.e. mRNA expression level of the gene. (B) i) Thermodynamic equilibrium-based TF-binding models using Hill functions, for example, ii) Linear models with output as a linear combination of input variables, iii) Bayesian networks providing the probability distribution of output given the input, and iv) Logical or Boolean models with output based on logical operations between input variables[104].

nodes $V_{i_j}(1 \leq j \leq k)$ to $V`_i$ resulting in the network with $V` = \{V_1, \ldots, V_n, V`_1, \ldots, V`_n\}$. The state of the additional node set $\{V`_1, \ldots, V`_n\}$ follows from $V`_i = f_i(V_{i_1}, \ldots, V_{i_k})$ for every i $\epsilon$ {1,…,n} treating $\{V_1, \ldots, V_n\}$ as the input and $\{V`_i, \ldots, V`_n\}$ as the output. The states obtained at every time step are considered as dynamic sequence of system states which are compared with biological evidence to arrive at the final successful model[105,106].



t			t+1		
INPUT			OUTPUT		
A	B	C	A'	B'	C'
1	1	1	1	0	1
1	1	0	0	0	1
1	0	1	1	0	0
1	0	0	0	0	0
0	1	1	1	1	0
0	1	0	0	1	0
0	0	1	1	1	0
0	0	0	0	1	0

Figure 1.16: A toy model of the Boolean network. (A) A Boolean network G(V,F). (B) The wiring diagram for the Boolean network in (A). (C) The logical operations and state transition table[106].

Most models focus on the behaviour of a single pathway or an outcome. Briefly, model construction begins with the compilation of nodes (relevant to the question) and their interacting partners based on previous experimental literature. Boolean functions for each node are laid based on its incoming nodes (activation or inactivation excluding the actual concentration levels). The initial state of each node is defined by its pre-stimulus or resting state (steady states will be used for validation). Having defined these above steps, the model is simulated, and its state change over time is evaluated. The steady-state obtained is compared with the biological information, and the model is revised in case of qualitative differences. A good consensus between biological knowledge and the model has important applications for a higher degree of understanding. Subsequently, the revised model can be used, for instance, to analyze the effect of node perturbations[107].

Due to the ease of implementation, Boolean models are an attractive choice for inferring gene regulation. Despite their crude simplification of the system, Boolean networks are likely to reproduce qualitative behaviour. They have been fruitful in capturing qualitative biological phenomena such as switch-like behaviour, oscillations, hysteresis, multistationarity etc[108]. Limitations of Boolean networks arise due to the strict assumption of genes to be either in on or off-state and synchronous nature, which is not true for biological systems.

While discrete models provide insights into the essential qualitative behaviour of the system, they seldom offer information on systems that evolve continuously in time and space. To model such scenarios, nonlinear ordinary differential equations (ODEs) are applied. Typically, the lack of information on mechanistic intricacies of regulatory networks and the unavailability of reaction rate constant creates a roadblock for modelling gene

regulation using ODE-based continuous models[109]. To overcome these shortcomings in discrete and continuous models, Reinitz and colleagues proposed an intermediate modelling strategy[110].

1.9.4.2 Bayesian network

Intuitive inferences from studies investigating individual components of pathways under various conditions play a crucial role in mapping signalling pathways. Further, to account for the underlying complexities of the system, a global multivariate approach is required. Bayesian networks are typically probabilistic graphical models that provide a suitable environment to model complex systems, helpful in inferring the dependence of pathway components on each other[111].

A Bayesian network is represented by a directed acyclic graph G (X, B) with vertices $X = \{X_1, X_2, \dots, X_n\}$ corresponding to variables, directed edge set B represents probabilistic dependence relations between the variables. Each node is attached with probabilities that define the chance of finding it in a particular state. Conditional probabilities are used to define the dependence of the state of a node on the state of another. These dependencies translate throughout the network, influencing the probabilities of other nodes. Graph G follows the Markov property, so each node X_i , is independent of its non-descendants given its parents[105,106,108]. The joint distribution of nodes is given by applying the chain rule of probability and conditional independence as:

$$P(X_1, X_2, \dots, X_n) = \prod_{i=1}^{n} P(X_i \mid parent\{X_i\}) \ , parent\{X_i\} \ is \ the \ set \ of \ parents \ of \ X_i \ in \ G \qquad (1.4)$$

The dependence structure between multiple interacting nodes gives the global network probability. For the Bayesian network in **Figure 1.17A**, the joint probability distribution of the network is $P(A, B, C, D, E) = P(A)P(B|A)P(C|B)P(D|A, E)P(E)$.

Inferring regulatory network from experimental data (transcriptomics data) involves two steps. Structure learning and parameter estimation are two major steps involved in Bayesian network modelling. The structural component is represented by a DAG G(V, E) consisting of causal relationships of regulations (set of edges E) among set of genes V. An edge exists

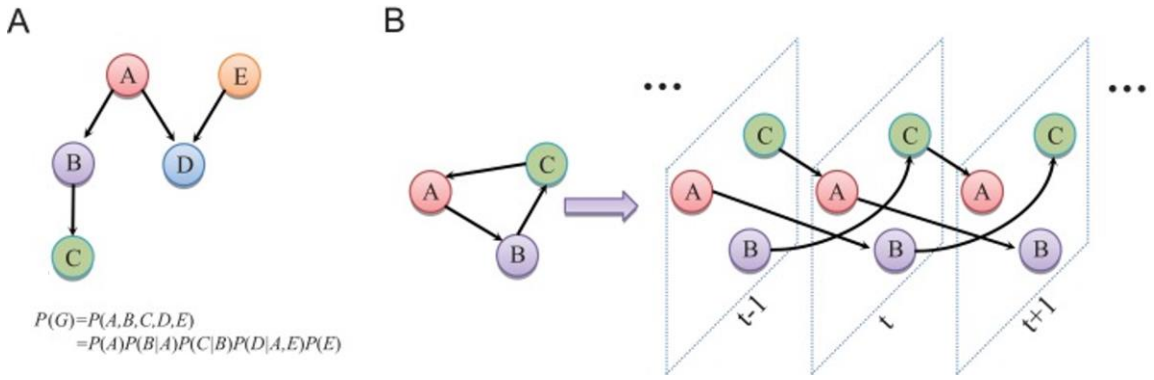


Figure 1.17: The graphical representation of Bayesian and Dynamic Bayesian network. (A) An example of a Bayesian network showing the joint probability distribution of network. (B) Dynamic Bayesian Network showing static and dynamic representations allowing cyclic structures[106].

from node X_i to node X_j if and only X_i directly regulates X_j. The parameter component θ accommodates the conditional probability distribution of each node given its parents. The network structure that best describes the expression data is obtained by choosing the appropriate parameter set θ that renders the highest posterior probability given the data, i.e. $P(G|D)$. The output of the Bayesian modelling on expression data is a hierarchical graph revealing the most plausible causal regulations between genes.

Bayesian modelling is an attractive choice for studying gene regulation due to its probabilistic nature that can account for inherent noise in biological systems[111]. It can efficiently describe direct molecular interactions and indirect influences and thus help in discovering previously unknown mechanisms. Arriving at an optimal Bayesian network is computationally expensive that can be handled by using experimentally validated predefined regulations as a prior to reducing the search space. The main drawback is its inability to decompose joint probability due to its acyclic nature and, hence cannot explain dynamical aspects like feedback regulation and oscillations.

To tackle the above shortcoming, a Dynamic Bayesian Network (DBN) was developed by duplicating nodes in the network to introduce the concept of time incorporating the temporal (and cyclic) regulation of dependencies between genes[105,106] **(Figure 1.17B)**. The probability distribution in DBN is given by a two-time-slice BN (2TBN) consisting of variables X_t with parents from X_{t-1} and/or X_t as:

$$P(X_t|X_{t-1}) = \prod_{i=1}^{n} P\left(X_{i,t}\middle|parent(X_{i,t})\right) \quad (1.5)$$

A DBN supports both intra-slice and inter-slice edges. DBNs incorporate the ordering of time, giving improved insights into temporal transcriptional regulation and have been applied to model time-series gene expression data.

Methods utilizing TF-gene interactions (static interaction such as those from ChIP-chip data) with expression data have been shown to be relatively more insightful than expression-only methods. There is a growing need to develop strategies that allow explicit integration of TF-gene interactions with expression data. Luscombe *et al*., (2004) proposed a method to integrate time series gene expression data with static ChIP-chip data based on differential expression of genes and mapping the ordered sequence of static regulatory graphs using trace-back algorithm[112]. Another group used Input-output Hidden Markov Model (IOHMM) to model gene expression patterns as a series of bifurcation events and identify the time when a TF(s) imposes its influence on these events[113] **(Figure 1.18).**

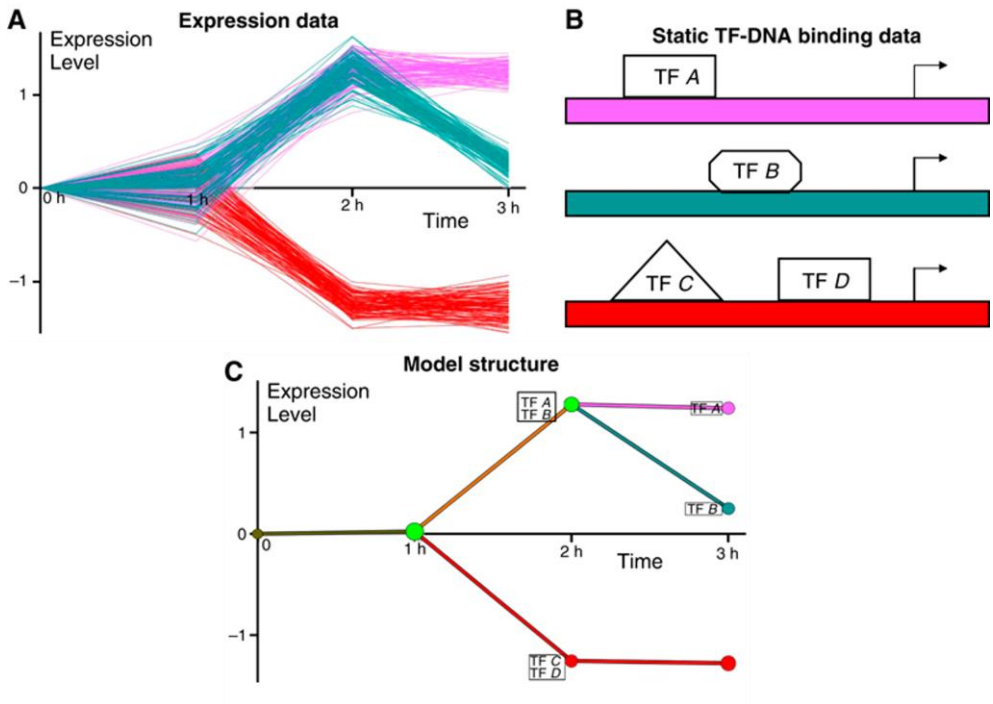


Figure 1.18: Inferring gene regulatory network by integration of transcriptomics with TF-gene interaction data. (A) Plots of time-series gene expression data. (B) Static TF-gene interaction data. (C) Gene regulatory network inferred from dynamic gene expression profiles and static TF-gene interaction data[113].

1.9.4.3 Input-output Hidden Markov Models (IOHMM)

An HMM is a simple example of DBN with one hidden state and one observed state in each time-slice[114–116]. It is a probabilistic model used to define a sequence of observed events that depend on internal factors that are not directly observed. HMMs are widely used for

DNA sequence analysis[117]. An HMM can be represented as a system undergoing probabilistic transitions from one state to another while emitting a symbol/character from each state **(Figure 1.19A)**. If $\boldsymbol{x} = x_1 x_2 \ldots\ldots x_L$ is observed output sequence from an underlying hidden state sequence $\boldsymbol{\pi} = \pi_1 \pi_2 \ldots . \pi_L$, where each x_n can take finite number of possible values from a set of observations $O = \{O_1, O_2, \ldots, O_N\}$ and each state π_n can take values from a set of states $S = \{1,2, \ldots . , M\}$, then the probability that HMM generates the output sequence $\boldsymbol{x} = x_1 x_2 \ldots\ldots x_L$ from unobserved state sequence $\boldsymbol{\pi} = \pi_1 \pi_2 \ldots . \pi_L$ is given by

$$P(x_1, \ldots . , x_L, \pi_1, \ldots . , \pi_L) = a(\pi_1) e(x_1|\pi_1) t(\pi_2|\pi_1) e(x_2|\pi_2) \ldots . t(\pi_L|\pi_{L-1}) e(x_L|\pi_L) \quad (1.6)$$

$a(i)$ is the initial state probability $P\{\pi_1 = i\}$, $e(x|i)$ is the probability of emitting a symbol x from state i, $t(j|i)$ is the probability of transition from state i to state j.

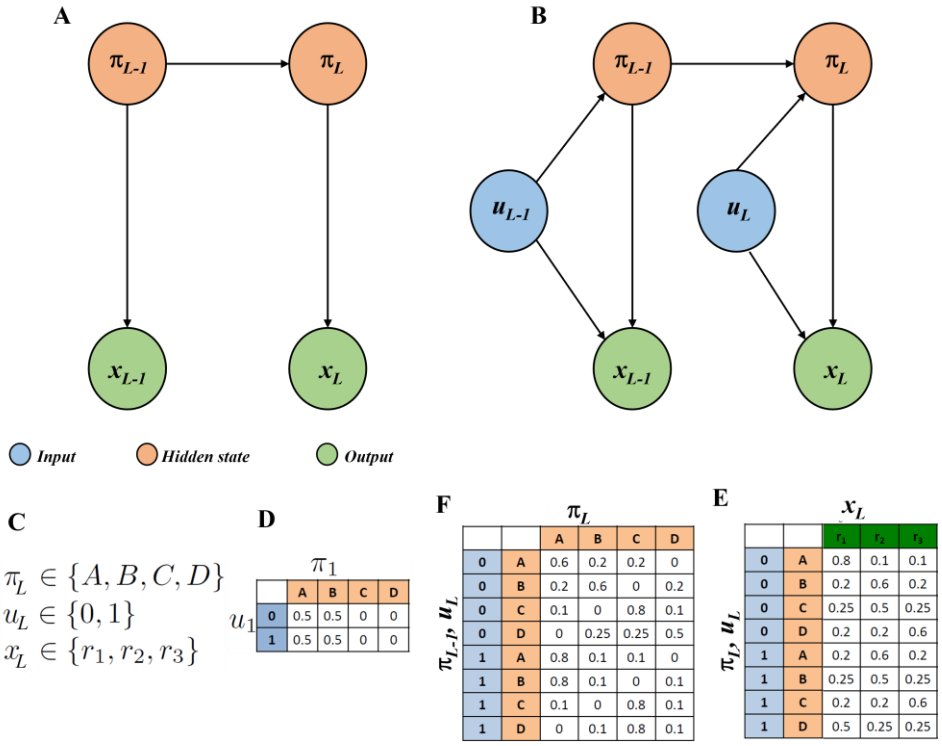


Figure 1.19: Hidden Markov Model and Input-Output Hidden Markov Model. (A) State transition and emission symbol between two consecutive steps in HMM. (B) State transition, emission symbol and input state between two consecutive steps in IOHMM. (C) IOHMM with hidden state set S = {A, B, C, D, E}, binary input U = {1,0} and output observations O = {r_1, r_2, r_3}. (D) Initial probabilities of IOHMM. (E) Transition probabilities of IOHMM. (F) Emission probabilities of IOHMM[118].

IOHMM, like HMM, has states, transitions, emissions, and an additional sequence of inputs $\{u_1, u_2, \dots, u_L\}$ **(Figure 1.19B)**. The transitions and emissions depend on these inputs and are given as[118]

$$P(x_1, \dots., x_L, \pi_1, \dots., \pi_L | u_1, u_2, \dots, u_L) \\ = a(\pi_1|u_1)e(x_1|\pi_1, u_1)t(\pi_2|\pi_1, u_2)e(x_2|\pi_2, u_2) \dots. t(\pi_L|\pi_{L-1}, u_L)e(x_L|\pi_L, u_L) \quad (1.7)$$

IOHMM framework has been adopted to track dynamic regulatory events by integrating TF-gene association data with time series gene expression data **(Figure 1.20).**

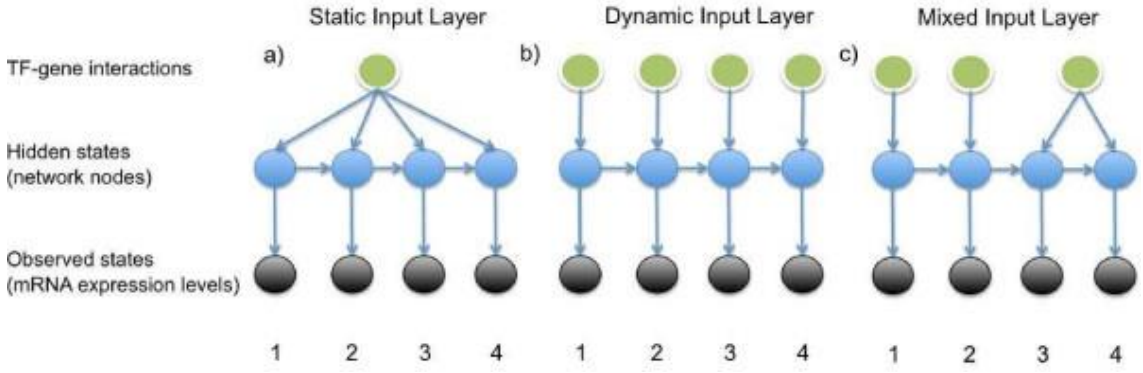


Figure 1.20: Integration of transcriptomics with TF-gene interaction data using the IOHMM framework. The observations (black nodes) correspond to gene expression at different time points given as input. The hidden states (blue nodes) are split nodes where genes diverge into different paths. The input states (green nodes) correspond to TF-gene interaction data. (A) IOHMM with all hidden nodes connected to one static input node. (B) IOHMM with time-dependent dynamic input nodes. (C) IOHMM with a mix of static and dynamic input nodes[119].

1.10 Structure of the thesis

The focus of the thesis is to understand the systems-level changes of the liver transcriptome in physiology and pathology. This includes mapping the temporal regulation of liver regeneration and ageing in mouse models, followed by an understanding of the dysregulation of liver regeneration in the clinical setting (patients). Chapter 2 presents a comprehensive picture of liver regeneration dynamics by applying probabilistic graphical modelling and gene co-expression network analysis pipeline to the time series gene expression data of mice. A discrete-continuous modelling framework was employed to model the circuit proposed to regulate the balance between liver function and proliferation

during regeneration. Chapter 3 maps the ageing-induced network-level changes in the liver and explores the crosstalk of ageing with regeneration and liver disease conditions NAFLD and HCC in the mouse liver. The change in statistical and topological properties of the ageing network is presented. The proximity of the ageing network to different liver condition-specific networks is studied to identify molecular players influencing regeneration and diseases with ageing. In Chapter 4, we addressed the scenario of impaired regeneration leading to the development of HCC. A detailed co-expression network analysis of HCC from normal to tumour through different premalignant states is presented. We also reported a molecular basis for disease-free survival prediction post-treatment by transplantation or partial hepatectomy. Liver regeneration after transplantation to treat liver disease provides a scope to study liver regeneration in humans. We analysed liver transplantation datasets to generate systems-level insights into rejection or tolerance (chapter 5). In the last chapter, we will present major conclusions and the future scope of the thesis.

Chapter 2 Systems-level analysis of transcriptome reorganization during liver regeneration

2.1 Introduction

The liver is bestowed with an impeccable capacity to restore its lost mass following an injury or partial resection by coordinated cell growth (hypertrophy) and proliferation (hyperplasia). The ability to maintain and recover the original liver-to-body mass ratio is inferred as the thermostat-like regulator "Hepatostat"[120]. Different studies have used the surgical procedure of two-thirds partial hepatectomy (PH) in rodents *(Mus musculus, Rattus norvegicus)* to understand liver regeneration[50]. These studies have revealed the sources of regenerated liver mass and described the three phases of liver regeneration (priming, proliferation, and termination). The liver is the metabolically active organ and is at a crossroads of lipid and carbohydrate metabolism. The regenerating liver not only needs to maintain the essential metabolic function but also needs to meet the metabolic requirement of hepatocyte growth and division.

Liver regeneration depends on the control mechanisms regulating the reversible transition between quiescence and proliferation. Hepatocytes shift from quiescent to primed state with the expression of immediate-early (IE) genes in response to cytokines (IL6 and TNFα) derived from non-parenchymal cells[36,50,121,122]. The second phase of regeneration involves the activation of growth factor signalling. Non-parenchymal cells synthesize and release growth factors and promote the release of extracellular matrix (ECM)-bound reservoir of growth factors. These include growth factors HGF and EGF, which activate c-met and EGFR receptors, respectively[44,123]. The last step involves cessation of proliferation by integrin signalling that promotes communication between ECM and epithelial cells[124,125]. The liver-to-body mass ratio is maintained by controlling the rate of cell division and apoptosis. After two-thirds PH, the hepatocytes also increase in size, followed by cellular division. An increase in the hepatocyte size alone is sufficient to recover the lost mass after PH in Cdk1 knockout[126]. A significant decrease in NADH concentration and mitochondrial function is observed. This kind of compensatory mechanism is not without consequences since there is an increase in liver damage markers.

In addition to cytokines and growth factors, metabolic signals play a role in liver regeneration[127,128]. The change in metabolic demand under liver regeneration leads to

systemic reorganization of metabolism. Animals subjected to PH display hypoglycemia in the initial phase since the liver plays a major role in maintaining systemic glucose levels[129]. There is an increase in the systemic influx of lipids and triglycerides from extrahepatic adipose tissue after PH, leading to transient steatosis, which provides the energy currency required for regeneration[128,130,131]. Other systemic cues include increased bile acid (BA) levels since the remnant liver cannot handle the BA returning via portal flow[132]. Blocking these metabolic alterations is shown to impair regeneration. Thus, regeneration is tightly intertwined with alterations in systemic metabolism.

Whole transcriptomic studies using microarray and RNA sequencing have mapped the gene expression pattern and transcriptional regulation in different liver regeneration models of rodents[126,133–141]. Metabolic genes are shown to be repressed, while the cell cycle genes are upregulated during liver regeneration. This raises the question of how the liver maintains metabolic homeostasis during liver regeneration. A division of metabolism into oxidative and biosynthesis phases has been proposed during liver regeneration[142]. It is not clear how the liver achieves the dynamic balance between various cellular processes, including metabolism and cell cycle. Since the hepatocytes in the liver lobule are exposed to different microenvironments, there is a zonation (spatial heterogeneity) of gene expression. The liver lobule is metabolically partitioned to periportal, mid-lobular, and pericentral zones, with different zones exhibiting differences in proliferative capability[143–145]. Recently, single-cell RNA sequencing (scRNA-seq) studies have started to reveal the division of labour with one population of hepatocytes activating early-postnatal-like gene expression and other compensating for metabolic function during liver regeneration[146,147]. In response to PH, a wave of hepatocyte proliferation starting from zone 1 to zone 3 has been observed, with midzone 2 representing the primary source of new hepatocytes during liver homeostasis and regeneration[148,149].

In this study, we modelled the temporal reorganization of the transcriptome of liver regeneration after PH to understand the coordination of liver function and regeneration using a schema outlined in **Figure 2.1**. The inference of dynamic regulatory network from RNA-seq data was performed, which shows the interplay of different cellular processes at different time points during liver regeneration. The co-expression pattern of genes reveals the coordination of metabolism and the cell cycle. We also developed a mathematical model of the integrated circuit of liver regeneration, which accounts for the dynamic balance

between requirements of liver function and regeneration as observed in scRNA-seq studies of liver regeneration.

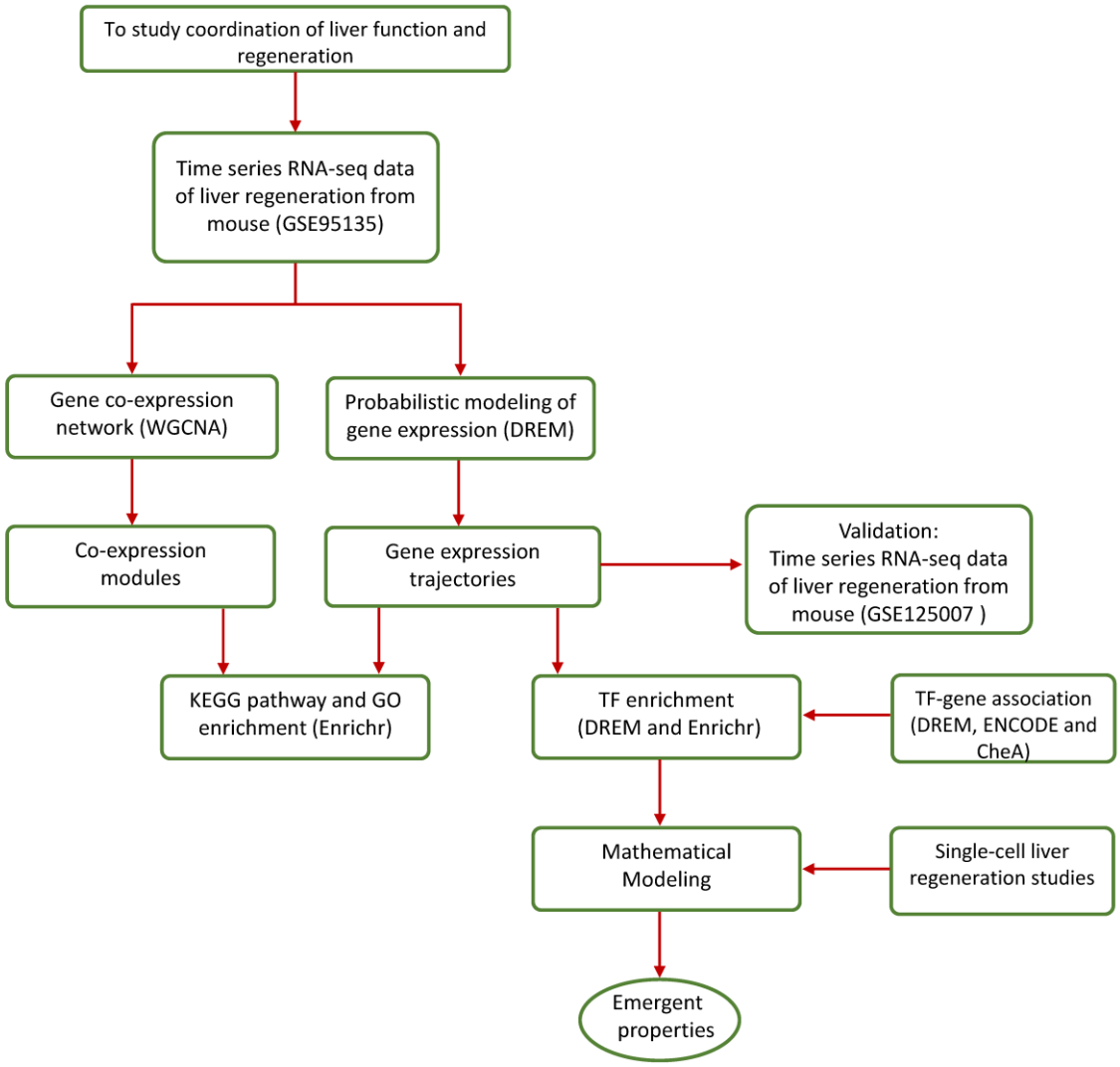


Figure 2.1: The workflow used to study the temporal reorganization of the transcriptome during liver regeneration. The workflow is divided into three parts. Gene co-expression analysis of temporal RNA-seq data of liver regeneration to identify co-expressed modules. Probabilistic modelling of temporal gene expression to extract different trajectories of genes and associated TFs. Mathematical modelling based on trajectories, TF-gene association, single-cell studies to explain the qualitative behaviour of emergent properties of liver regeneration.

2.2 Methods

2.2.1 Transcriptomics data

We used the publicly available high-resolution temporal RNA-seq data (Illumina HISeq 2000) of liver regeneration after PH from Gene Expression Omnibus (with accession

number GSE95135)[139]. The samples in the dataset correspond to PH operated 12- to 14-week-old male C57/BL6 mice entrained with 12 hours light-dark, fasting-feeding cycles. PH samples include time points 0, 1, 4, 10, 20, 28, 36, 44, 48, 72, 168 and 672 hours. Log_2(RPKM +1) values were used for the downstream analysis. We also verified our findings using another RNA-seq data of liver regeneration (with accession number GSE125007)[140]. PH samples include time points 0, 24, 30, 40, 48, 96, 168 and 672 hours.

2.2.2 Reconstruction of the co-expression network of liver regeneration

The co-expression network of liver regeneration was constructed using the weighted gene co-expression network analysis (WGCNA) package in R[150]. Top 5000 varying genes across time points were selected (using rowVars function in R) to construct the correlation (Pearson) matrix for WGCNA.

A soft power adjacency function, $a_{ij} = s_{ij}^{\beta}$, was used to construct an adjacency matrix. We used the scale-free topology criterion to choose power β. This was obtained by computing the square of the correlation (R2) between log(p(k)) and log(k), where p(k) is the frequency distribution of the connectivity k. A plot of R2 and β, which shows a saturation characteristic, was used to choose the β value of 15. This corresponds to the point where the saturation is reached with mean connectivity ≥100. A topological overlap matrix (TOM) was constructed from the adjacency matrix, and 1-TOM was used to construct the dendrogram[151]. The modules were identified using the dynamic tree cut algorithm with a minimum module size of 100. The module eigengene (ME) expression was obtained by singular value decomposition (SVD). Enrichr was used to identify GO terms and KEGG pathways associated with each module[152]. The co-expression network was visualized using Cytoscape[153].

2.2.3 Probabilistic graphical modelling

We further reconstructed the dynamic regulatory network using the DREM method[119], to integrate time-series gene expression data with the transcription factor (TF)-gene association data. This approach clusters patterns of gene expression into paths and bifurcation points. Each bifurcation point represents a divergence in the expression of co-expressed genes under TF(s) influence. Log_2fold change in the expression of genes at every time point with respect to the reference time point (0 hours) was used as an input. We used the generated TF-gene association available for mouse[119]. The following parameters were

used: a) Minimum $\log_2$ fold change of 1, b) The expression scaling weight set to 0.5, and c) TFs associated with a bifurcation point in the model were chosen with a hypergeometric distribution score less than 0.001. We also used ChEA and ENCODE_and_ChEA_consensus libraries from Enrichr[152] to identify TFs significantly (adjusted p-value < 0.05) associated with the clusters. ChEA (ChIP-x Enrichment analysis) database is a collection of putative TF-gene association data extracted from TF-binding studies (ChIP-chip, ChIP-seq, etc). ChIP-seq data from ENCODE (Encyclopedia of DNA Elements) project is also used for the TFs identification. The target genes and TFs from ChEA and ENCODE databases are used for consensus TFs.

The DREM method proposed by Schulz *et al.,* (2012) to integrate time-series gene expression data with TF-gene association data is based on Input Output Hidden Markov Model (IOHMM). The hidden states of IOHMM are used to group the genes into paths. Each of these paths pass through the hidden states over time. Transitions among the hidden states are constrained to obtain the tree-like structure resulting in the bifurcation points. Many possible tree-like structures are searched and scored to select the best one.

The DREM model M is a tuple (H, E, Ψ, Θ, n, γ), where n corresponds to the number of discrete time points, H is the set of hidden states 'h' such that every h is characterized by a Gaussian output distribution f_h and is associated with one time point. Θ is the set of parameters (μ_h, σ_h) corresponding to the Gaussian distribution of each hidden state. Every hidden state h_t with t < n can have at most γ child nodes (hidden states of H in the next time point). E is the set of directed edges between h_a and h_b s.t a + 1 = b, reflecting the transitions among the hidden states (i.e. hidden states of the consecutive time points are connected). Ψ consists of parameters controlling the transition probabilities between the hidden states (and their child nodes). Ψ_h for every hidden state h is a vector of parameters for logistic regression classifier which makes use of TF-gene association information. Let I_g be the vector defining TF-gene association for gene g and all its regulating TFs (1 for activating, -1 for inhibiting, 0 for no regulation). If gene g is in the hidden state h_a at time t-1 and has h_b and h_c as child states at time t, then the probability that g undergoes a transition from h_a to h_b (or h_c) is given by logistic function with intercept INT as follows:

$$\frac{1}{1 + e^{-\Psi_h.INT - \sum_x \Psi_h x \,.\, I_g x}} \qquad (2.1)$$

The expression level of TF is also incorporated to influence the learning of the classifier using the shifted version of the logistic function as follows

$$f_w^*(x) = sign(x)(\frac{2}{1+e^{-xw}} - 1) \quad (2.2)$$

where x is the expression ratio of a TF between two time points, w is the expression scaling weight that controls the steepness of the function. To incorporate the efficiency of TF, it is selected based on the minimum threshold

$$select(x) = \begin{cases} f_w^*(x), & if\ abs(f_w^*(x)) \geq minExpTF, \\ sign(x).minExpTF, & else \end{cases} \quad (2.3)$$

Therefore, pairwise binding information of a TF t to gene g (B_g Є {-1, 1, 0}) is now transformed to

$$B_g' = select(x_t).B_g \quad (2.4)$$

If the vector $O_g = (o_g(1), o_g(2),\ldots,o_g(n-1))$ denotes the log ratio of expression values for a gene g at time points 1 to n-1 with respect to zeroth time point, with I_g as its TF-gene association mapping vector and H_i is its hidden state variable at time t, then the probability it is in state h_b at time t given that it is in state h_a at time t-1 is given by $P(H_t = h_b|H_{t-1} = h_a, I_g)$. This probability is 0 if h_b is not the child of h_a and 1 if it is the only child. If number of child nodes is greater than or equal to 2, then the transitions are probabilistic and are based on the vector I_g.

The likelihood density r, for a set of genes G for the model, is given by

$$r(G|M) = \underbrace{\sum_{g \in G} log}_{all\ genes} \underbrace{\sum_{q \in Q}}_{all\ paths} \underbrace{\prod_{t=1}^{n-1} f_{q(t)}(o_g(t))}_{output\ densities} \underbrace{\prod_{t=1}^{n-1} P(H_t = q(t)|H_{t-1} = q(t-1), I_g)}_{transition\ probabilities} \quad (2.5)$$

where, Q is the set of all paths of hidden states of length n beginning from the root with non-zero probability. For a path q Є Q, q(i) is the hidden state of the path at time point i. The first product corresponds to the product of output densities for the expression values and a given sequence of hidden states. Given a tree structure determined by H and E, the parameters that maximize the likelihood density 'r' are estimated using Baum-Welch algorithm.

To build the dynamic regulatory map, the algorithm begins with a search over structures of trees. Parameters of Gaussian distribution and the classifiers corresponding to a tree structure are learnt by randomly selecting the subset of genes as the training set using a version of the Baum-Welch algorithm. The remaining genes are used as test set to assign scores to various tree structures considered. The best scoring tree structure is selected, and all genes are used to train the parameters and arrive at a final model.

2.2.4 Mathematical Modelling

A simple regulatory circuit controlling significant TFs associated with clusters was constructed to model the transition between hepatocyte function and proliferation states. Since DREM analysis was carried out using only the TF-gene association data, we also considered interactions that control TFs from the literature. We studied the functional bifurcation of hepatocytes into different states characterized by single-cell transcriptomic studies of liver regeneration[146]. We adopted the framework proposed by Reinitz and colleagues to model the regulatory circuit[109,110]. This framework combines the best features of discrete and continuous approaches to simplify the complexity of the interactions in the network. We formulated a set of non-linear Ordinary differential equations (ODEs) of the form:

$$\frac{dY_i}{dt} = \beta_i[F(\delta_i W_i) - Y_i], \qquad W_i = w_{i0} + \sum_{j=1}^{N} w_{ij} Y_j, \quad i = 1, \dots, N \tag{2.6}$$

Y_i is expression level of the gene ($0 \leq Y_i \leq 1$), $F(\delta W) = 1/(1 + e^{-\delta W})$ is "soft-Heaviside" function that varies from 0 ($W << -1/\delta$) to 1 ($W >> 1/\delta$), δ determines the steepness of the function and Wi is the net effect on gene i of all genes in the network. The steepness in biological terms can be related to sensitivity of the response to an input signal. The coefficient ω_{ij} can take values less than 0 (gene$_j$ inhibits the gene$_i$), more than 0 (gene$_j$ activates gene$_i$) or equal to 0 (no effect of gene$_j$ on gene$_i$). This equation also behaves like a discrete boolean for a large value of δ_i's **(Figure 2.2A)**. For δ_i values greater than 1, Y_i flips between 0 and 1 on a timescale $\approx \beta_i^{-1}$. For larger δ values, the system shows ultrasensitive behaviour with threshold for turning ON/OFF the response with input stimuli. Ultrasensitivity is of biological significance in cell signalling network[154].

Consider a circuit with two components, Y_1 and Y_2, with input signal S activating on Y_1

and further Y_1, inhibiting Y_2. The wiring diagram describing this system is $S \rightarrow Y_1 \dashv Y_2$. The equations describing the dynamics of Y_1 and Y_2 are as follows:

$$\frac{dY_1}{dt} = \beta_1 \left[\frac{1}{1 + e^{-\delta_1 (S + w_{10})}} - Y_1 \right] \tag{2.7}$$

$$\frac{dY_2}{dt} = \beta_2 \left[\frac{1}{1 + e^{-\delta_2 (w_{20} + w_{21} Y_1)}} - Y_2 \right] \tag{2.8}$$

The above two differential equations can be solved with parameter values assigned to arrive at the solutions $Y_1(t)$ and $Y_2(t)$, which can be further visualised to study their dynamics as a function of time t. The dynamics of the system can also be studied by steady-state analysis, a state where $\frac{dY_1}{dt} = 0$ and $\frac{dY_2}{dt} = 0$. By plugging the parameter values, the steady-state values of Y_1 and Y_2 can be obtained by using the condition $\frac{dY_1}{dt} = 0$ and $\frac{dY_2}{dt} = 0$. This gives steady-state value of Y_1 as $Y_1^* = \frac{1}{1 + e^{-\delta_1 (S + w_{10})}}$ and subsequently this value can be used to obtain steady-state value of Y_2 as $Y_2^* = \frac{1}{1 + e^{-\delta_2 (w_{20} + w_{21} Y_1^*)}}$. The dynamics of the system can be seen in **Figure 2.2B.**

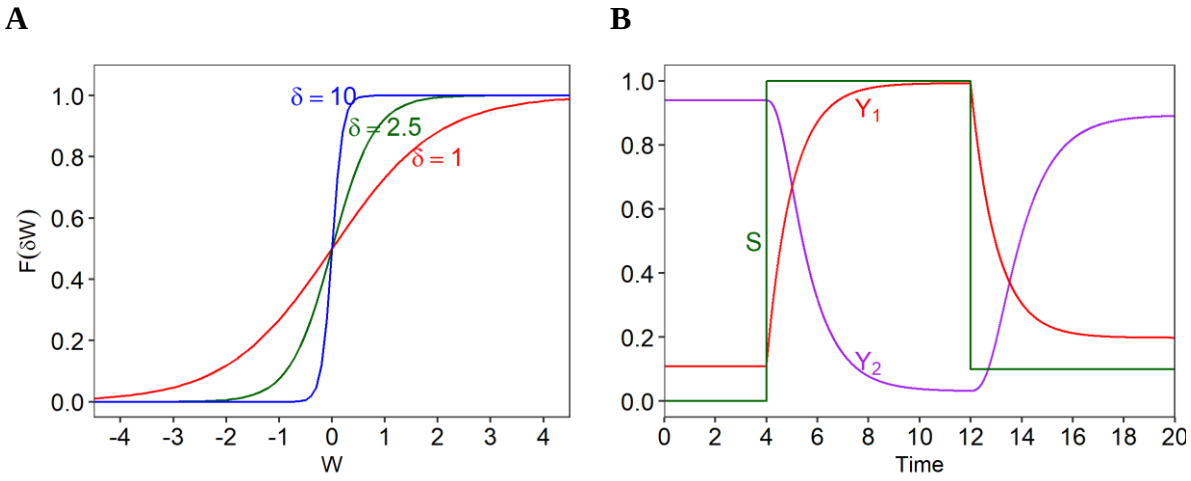


Figure 2.2: Discrete-continuous ODE framework. A) Graph depicting how $F(\delta W) = \frac{1}{1+e^{-\delta W}}$ varies from 0 to 1 as a function of W, and δ determines the steepness. B) Dynamics of the system $S \rightarrow Y_1 \dashv Y_2$. The simulation is shown for the parameter values: $\beta_1 = \beta_2 = 1$, $\delta_1 = \delta_2 = 7$, $w_{10} = -0.3$, $w_{20} = 0.5$, $w_{21} = -1$. Since S activates Y_1, the value of Y_1 rises as the input signal is increased. Simultaneously, levels of Y_2 drop due to the inhibitory effect from Y_1. The opposite is seen when the signal is removed. The activation/inactivation of Y_1 and Y_2 are sharp due to the high δ values.

The values of the parameter β and δ in the above equations can be varied to study the effect on the response (sensitivity analysis) or to study how the qualitative behaviour of the

system changes from one state to the other state (bifurcation analysis). We focused on studying how the qualitative behaviour (steady state) of the system changes with respect to parameter changes by performing phase plane and bifurcation analyses using XPPAUT (available from http://www.math.pitt.edu/~bard/xpp/xpp.html). The emergent properties of the liver regeneration circuit were analyzed to understand the emergence of different states and their transitions during liver regeneration.

2.3 Results

2.3.1 Dynamic regulatory network of liver regeneration

We first constructed the dynamic co-expression network of liver regeneration to study the transcriptome organization into functional modules using time-series expression data. We performed WGCNA and identified modules related to liver regeneration after PH. The modular organization of the co-expression network is shown in **Figure 2.3A**. Modules blue (M1), green (M2), red (M3), pink (M4), and purple (M5) that are clustered together show a positive correlation with pre-and post-PH stages **(Figure 2.3B)**. On the other hand, modules black (M6), yellow (M7), brown (M8) and magenta (M9) that are clustered together show a negative correlation with stages. However, the correlation of most modules

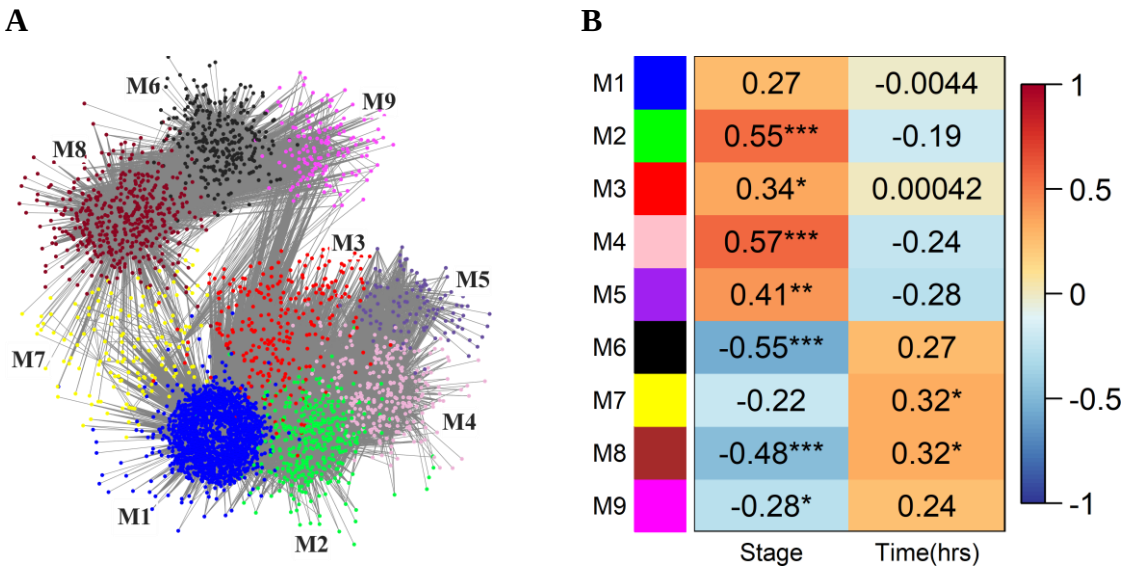


Figure 2.3: Co-expression network of liver regeneration. (A) Modular organization of transcriptome of liver regeneration. Modules M1 – M9 identified by WGNCA are indicated by different colours. (B) Correlation of module eigen (ME) gene expression of modules M1 – M9 with stages (pre-and post-PH) and different time points. * indicates p-value < 0.001, ** indicates 0.001 ≤ p-value < 0.01, * indicates 0.01 ≤ p-value < 0.05.**

with different time points of liver regeneration decreases. The eigengene expression of each module shows the transient nature of gene expression with a change in direction occurring at different time points and recovery to the pre-PH condition in the termination phase of liver regeneration **(Figure 2.4)**. The eigengene expression of M4 and M5 modules increases between 1 and 4 hours, and of M2 and M3 modules increases between 4 and 10 hours post-PH (priming phase). The eigengene expression of the M1 module increases between 28 and 36 hours post-PH (proliferative phase). The M5 module shows early recovery between 36 and 44 hours compared to other modules. The M8 and M9 modules are downregulated between 1 and 4 hours, while the M6 module is downregulated between 4 and 10 hours.

We also identified the biological processes and KEGG pathways associated with each module **(Table A1).** In the upregulated modules, the M1 module is associated with the cell cycle, DNA replication, and p53 signalling pathway; the M3 module is associated with protein processing in ER and protein export; the M2 module is associated with complement and coagulation cascades, and the M4 module is associated with TNF signalling and ribosome biogenesis. The M1 module captures the proliferative response of hepatocytes that peaks after 36 hours, while the M3 module captures the role of endoplasmic reticulum (ER) stress. The upregulated modules also show a link with metabolic pathways: glutathione metabolism (M1), arginine and proline metabolism (M1), amino sugar, and nucleotide sugar metabolism (M3, M1), fatty acid degradation (M5 and M2), PPAR signalling pathway and peroxisome (M5).

The downregulated modules are primarily associated with metabolism **(Table A1)**. The M6 module is associated with cholesterol metabolism, steroid hormone biosynthesis, bile acid biosynthesis, bile secretion, and PPAR signalling pathway. The M8 and M9 modules are associated with retinol and amino acid metabolism (branched-chain amino acids; glycine, serine, and threonine; tryptophan; cysteine and methionine; histidine). We also found glutathione metabolism, folate metabolism, pentose and glucuronate interconversions, glycoxylate and dicarboxylate metabolism, and arachidonic acid metabolism as part of the downregulated modules. WGCNA revealed the global organization of liver transcriptome into modules, which are obtained based on the scale-free topology criteria.

To further generate insights into the dynamic organization and regulatory mechanism of liver regeneration, we performed probabilistic modelling of gene expression (**section 2.2.3**). This dynamic modelling approach revealed three core clusters that are upregulated

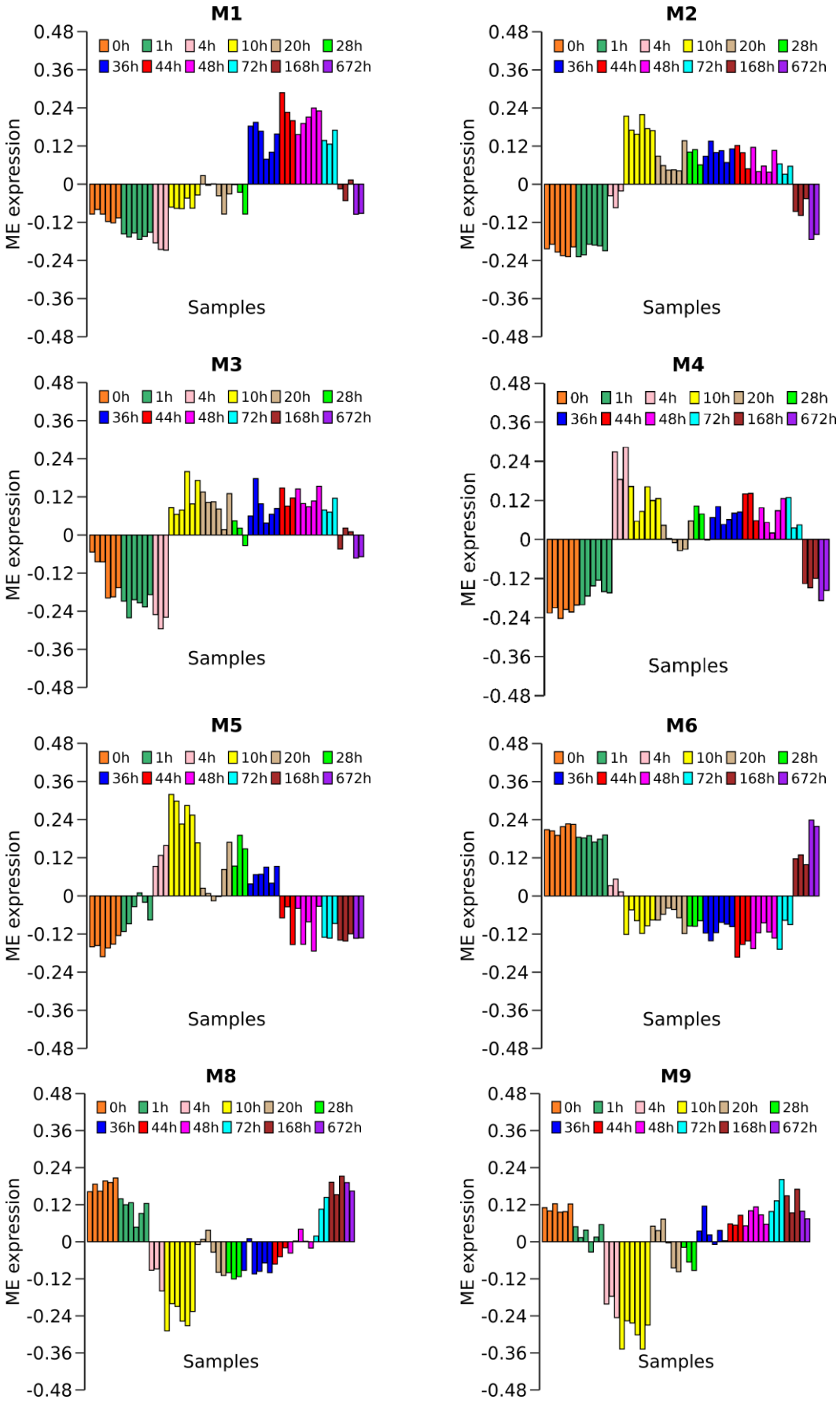


Figure 2.4: Eigengene expression profile of individual modules with respect to different time points of liver regeneration.

immediately (cluster 1 - **Figure 2.5A**), upregulated after 28 hours (cluster 2 - **Figure 2.6A**), and downregulated immediately (cluster 3 - **Figure 2.7A**) post-PH and their bifurcation into 17 sub-clusters (named paths A to Q). We identified transcription factors (TFs) associated with the three core clusters.

TFs regulating cluster 1 include FOS, JUN, CEBPB, NFKB, and STATs **(Table 2.1)**. These TFs are known to be involved during the priming phase of liver regeneration. Other TFs include HNF4A, XBP1, LEF1, USF1, GATA4, EGR1, ESR1, and NFATCs. LEF1 is a

Table 2.1: Transcription factors (TFs) associated with DREM core clusters. Significant TFs (adj p-value < 0.05) are identified based on different databases for transcription factor enrichment analysis. DREM results are based on the generated TF-gene association for the mouse (score < 0.001). * represents uncorrected p-value.

Cluster	ChEA	ENCODE and ChEA	DREM	DREM, ENCODE and ChEA/ChEA
1	RXR(1.35e-12), LXR(2.21e-12), PPARA(4.7e-07) RELA(2.62e-06), CEBPB (1.4e-04), CLOCK (3.8e-03)	HNF4A(7e-03), STAT3(6e-03), GATA1(2e-03), EGR1(0.04), ESR1(0.01), FOSL2(0.04), FOS*(0.01), USF1*(0.02),	CEBPB (2.6e-06), XBP1(1.3e-04), NFATCs(5.41e-06), STATs(4.74e-04), GATA4(1.44e-04), JUN(1.35e-05), JUND(1.83e-05), NFKB1(7.82e-04), LEF1(6.3e-04), NFE2L1(1.31e-04), FOS(1.83e-05), FOSL2(6.44e-05)	CEBPB, FOS, FOSL2
2	E2F1 (5.75e-06), E2F4 (4.5e-17),MYC (3.44e-05), FOXM1(4.19e-20)	E2F1(1.05e-19), E2F4(1.55e-38),FOXM1(4.07e-25), MYC(2.27e-05), IRF3(3e-04), NFYA(3.65e-17), NFYB(1.75e-06), NRF1(1.72e-05)	E2F1(3.3e-07), E2F4(1.29e-07), TFDP1 (3.14e-07), TFDP2 (8.1e-06), NFYA (1.35e-07), NFYB(1.75e-06), NRF1(1.74e-04)	E2Fs, NRF1, NFYA, NFYB
3	RXR(4.01e-08) LXR(1.94e-05), PPARA(5.14e-15), EGR1(2.2e-09), ESR1(1.85e-05), FOXO1(5e-06)	HNF4A(3e-02), EGR1(3e-03), ESR1(4e-03)*		

downstream effector of the Wnt pathway important for hepatic periportal gene expression[155], and XBP1 is a regulator of unfolded protein response (UPR). Further, six sub-clusters (paths A to F) come under cluster 1 **(Figure 2.5A)**. Paths A, B, C, and E are

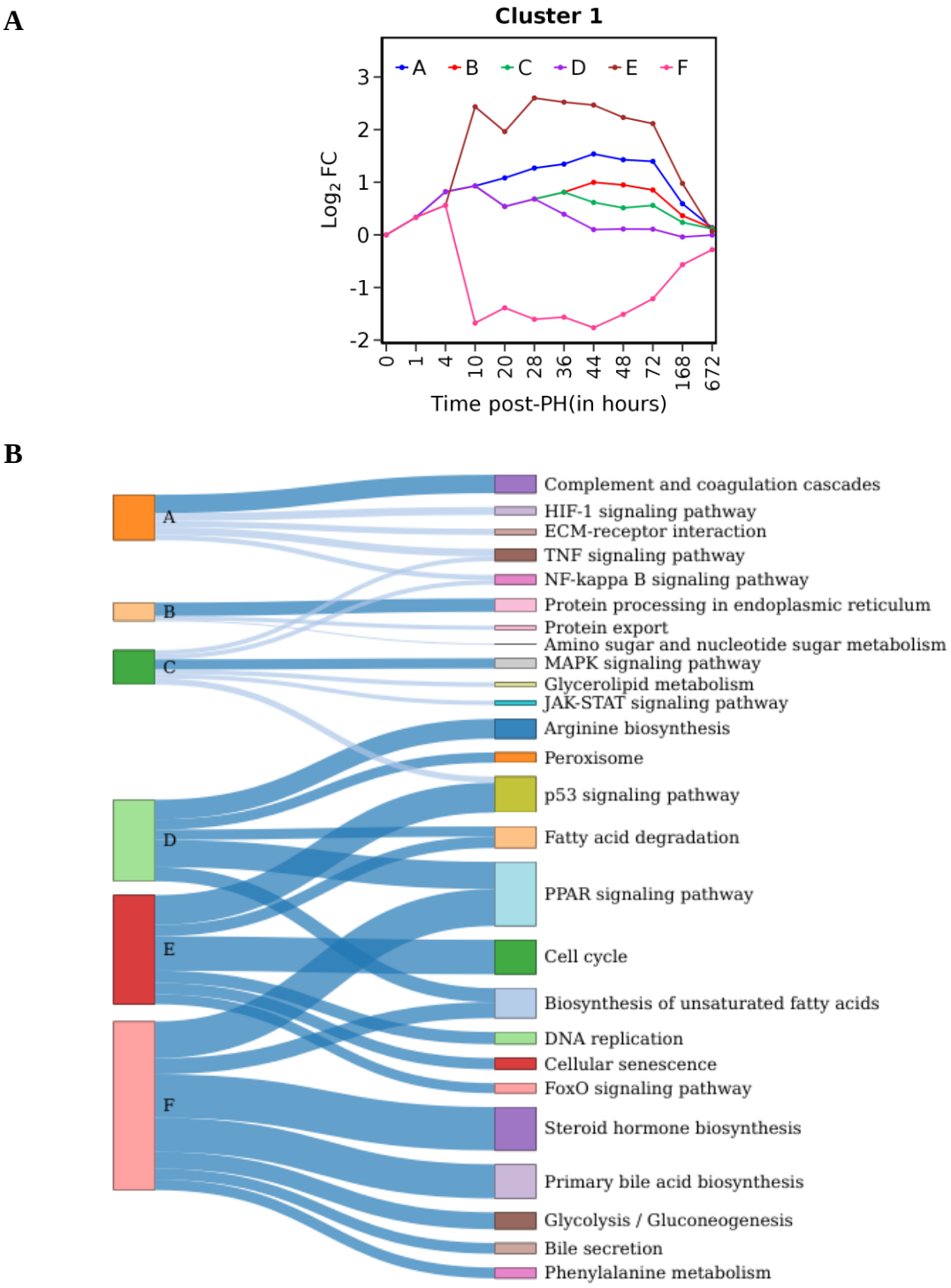


Figure 2.5: The regulatory paths of the set of co-expressed genes in the cluster 1 of liver regeneration. A) The x-axis represents the time points of sample collection, and the y-axis represents the mean log2 fold change (log_2FC) in mRNA expression post-PH for each path. B) A path is split into multiple paths (split nodes) based on the

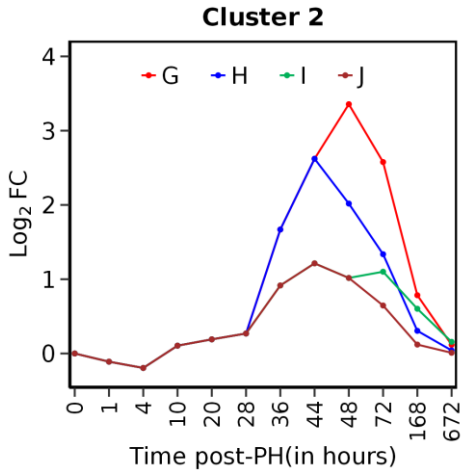
divergence in gene expression. Sankey plot showing the significant KEGG pathways (dark blue: *adj p-value* < 0.05; light blue: *p-value* < 0.05; thickness represents $-log_{10}(adj\ p-value)$) associated with different paths of cluster 1 (A to F).

upregulated throughout the regeneration period up to 1-week post-PH and return to the baseline at 4 weeks. Path A is enriched for complement and coagulation cascade, HIF and TNF signalling pathways **(Figure 2.5B)**. Path B captures pathways related to protein processing in ER, amino sugar and nucleotide sugar metabolism, and protein export, which may play an essential role during the initial response soon after the resection. Path E is significantly upregulated post-10 hours and is enriched for fatty acid degradation, p53 signalling, and DNA replication. Path D shows an initial rise in gene expression, which returns to baseline 36 hours post-PH **(Figure 2.5A)** and is enriched for metabolic pathways related to fatty acid and amino acid metabolism **(Figure 2.5B)**. Path F initially shows an increasing trend in gene expression but gets downregulated from 4 hours throughout the regeneration period, which makes it different from other paths in cluster 1. This path is enriched in steroid hormone biosynthesis and bile acid biosynthesis. The enrichment of paths D, E, and F suggests alterations in lipid metabolism in the priming phase.

TFs regulating cluster 2 include E2Fs, FOXM1, MYC, NRF1, IRF3, NFYA, NYFB, and TFDP1/2 **(Table 2.1)**. The gene expression of cluster 2 is relatively uniform compared to the other two core clusters. There are 4 paths (G to J) that come under cluster 2, showing an increasing trend post 28 hours (**Figure 2.6A**). Path G continues to increase until 48 hours and returns to baseline at 4 weeks. Paths G, H, and J of this cluster are significantly enriched for cell cycle events, spliceosome, and DNA damage repair pathways (**Figure 2.6B**). Path I is associated with neutrophil degranulation.

Cluster 3 shows both transient and sustained downregulation (paths K to Q) throughout the regeneration period, returning to baseline after 1 week (**Figure 2.7A**). TFs regulating this cluster include HNF4A, RXR, LXR, EGR1, and ESR1 (**Table 2.1**). The paths of this cluster are mainly enriched for metabolic pathways (**Figure 2.7B**). Paths K, L, and M show transient downregulation at 4 hours; however, path K further rises during the proliferative phase. Path L is associated with negative regulation of JAK-STAT and cell size. Paths Q and P are downregulated throughout liver regeneration. Paths K, M, P, and Q are associated with the lipid and amino acid metabolism, with path K and P also associated with the glutathione metabolism. Paths N and O also show transient downregulation at 4 hours and are associated with amino acid metabolism and ribosome, respectively.

A



B

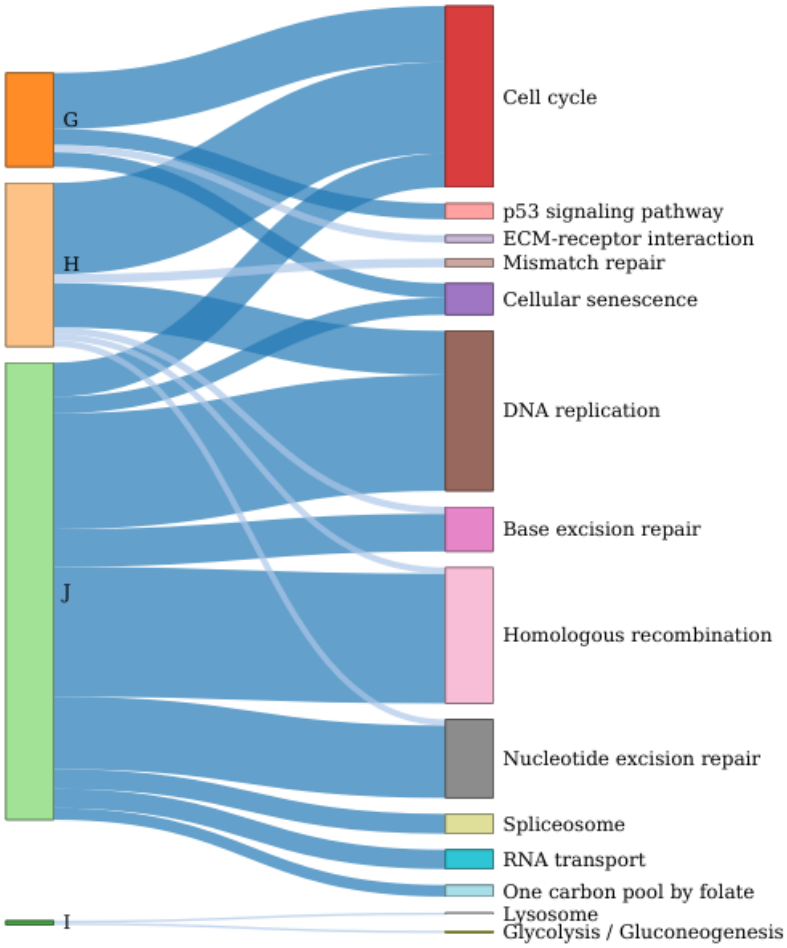
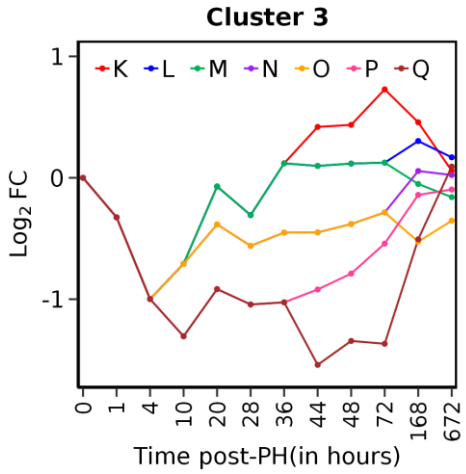


Figure 2.6: The regulatory paths of the set of co-expressed genes in the cluster 2 of liver regeneration. A) The x-axis represents the time points of sample collection, and the y-axis represents the mean log2 fold change (log₂FC) in mRNA expression post-PH for each path. B) A path is split into multiple paths (split nodes) based on the divergence in gene expression. Sankey plot showing the significant KEGG pathways (dark blue: adj p-value < 0.05; light blue: p-value < 0.05; thickness represents $-log_{10}(adj\ p-value)$) associated with different paths of cluster 2 (G to J).

A



B

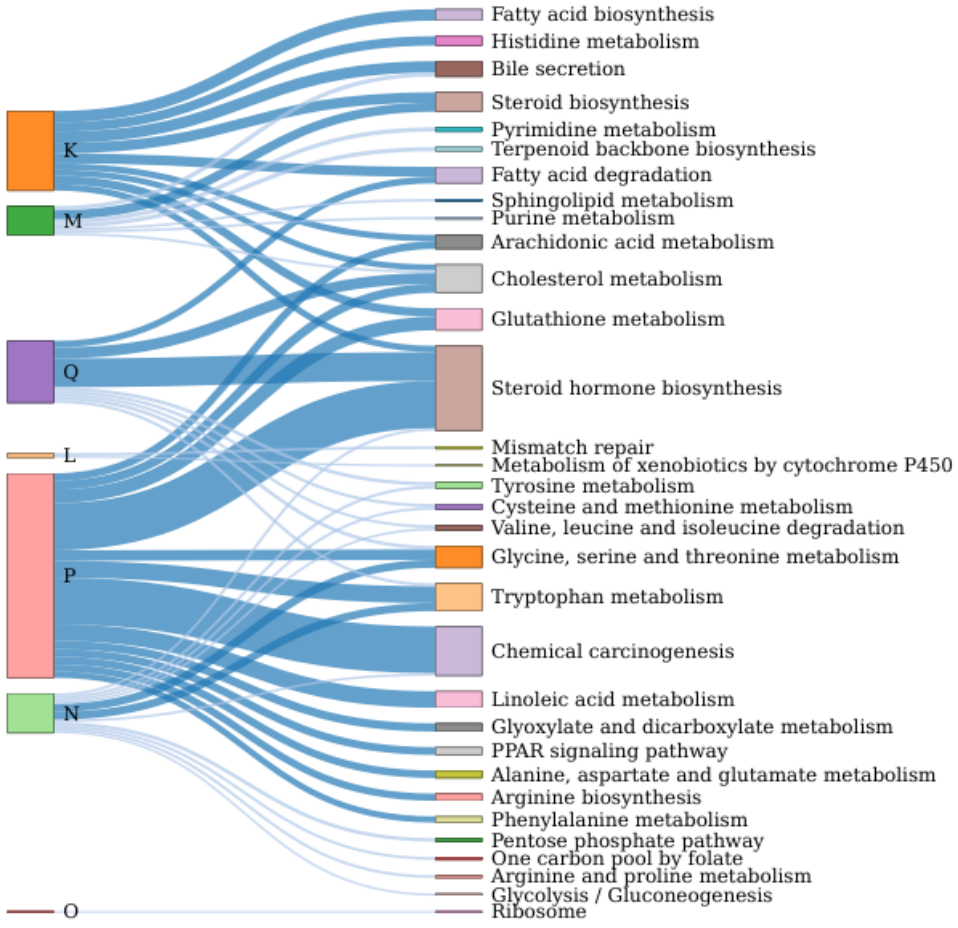


Figure 2.7: The regulatory paths of the set of co-expressed genes in the cluster 3 of liver regeneration. A) The x-axis represents the time points of sample collection, and the y-axis represents the mean log2 fold change (log_2FC) in mRNA expression post-PH for each path. B) A path is split into multiple paths (split nodes) based on the divergence in gene expression. Sankey plot showing the significant KEGG pathways (dark blue: adj p-value < 0.05; light blue: p-value < 0.05; thickness represents $-log_{10}(adj\ p-value)$) associated with different paths of cluster 3 (K to Q).

Further, we analyzed another recent RNA-seq data (dataset 2 – validation dataset) of liver regeneration using probabilistic modelling to confirm our findings. Although this dataset has a starting time point of 24 hours, we consistently observed the organization of the transcriptome into three core clusters pertaining to cell cycle, immune response, and metabolism with a similar set of transcriptional factors capturing the events of liver regeneration **(Figure 2.8 and Table A2)**. The changes related to the immune response are a continuum from the priming phase, as observed in **Figure 2.5**. We also found pathways related to RNA transport and spliceosome upregulated in both datasets, while ER stress pathway is not enriched **(Table A2).** Lipid and glutathione metabolism are downregulated

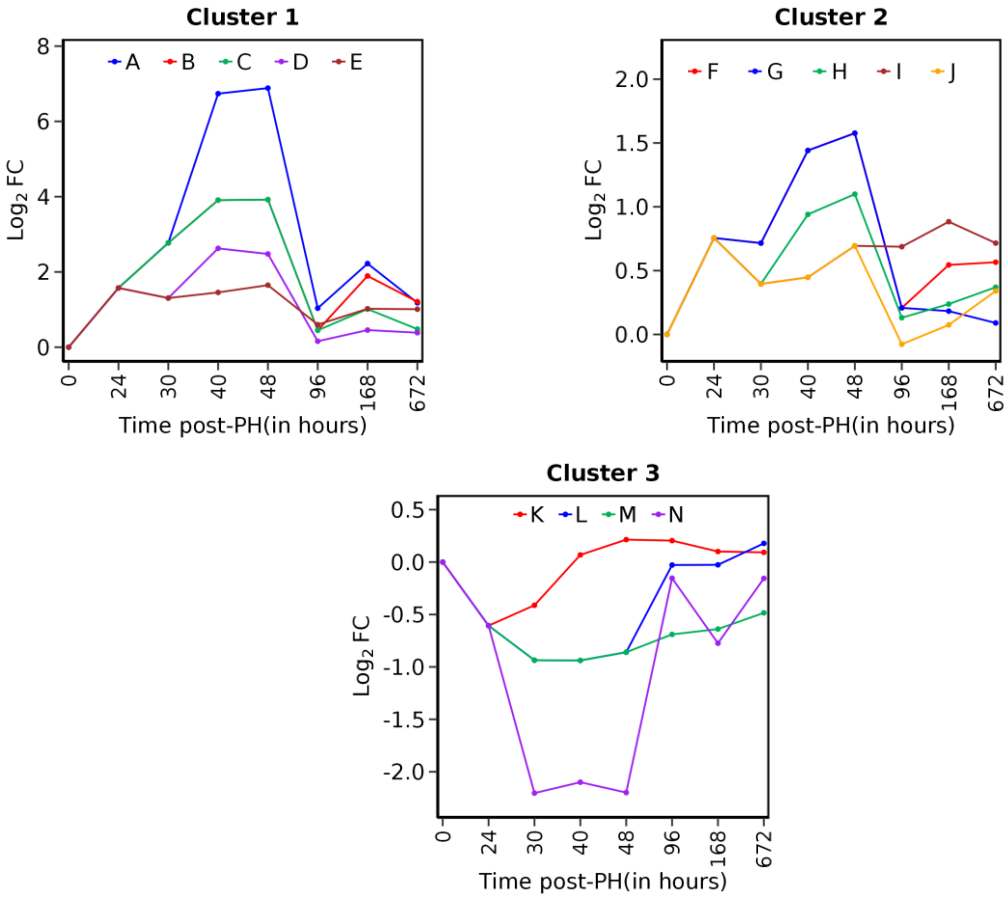


Figure 2.8: The regulatory paths of the set of co-expressed genes in the three core clusters of liver regeneration (validation dataset). The x-axis represents the time points of sample collection, and the y-axis represents the log2 fold change (log_2FC) in mRNA exp expression post-PH. A path is split into multiple paths (split nodes) based on the divergence in gene expression.

and co-cluster in a single path in both datasets (paths P and L in datasets 1 and 2, respectively). Downregulated steroid hormone biosynthesis and glycolysis/gluconeogenesis (cluster 3, paths L and N) are reset to the baseline by 96 hours.

Cysteine and methionine metabolism and one-carbon folate pool are upregulated in the validation dataset.

2.3.2 Co-expression pattern of lipid metabolism genes shows further rise during the proliferative phase

We further analyzed the co-expression pattern of metabolic pathway genes. In path K, we observed that genes related to the *de novo* lipogenesis pathway (SREBF1, FASN, ACACA, ACLY, ACSL3) and hydrolysis of fat (LPL) are further upregulated by 72 hours **(Figure A1)** after initial downregulation. Path K also includes genes of cholesterol metabolism (HMGCR, SQLE, and LDLR) that are initially downregulated **(Figure A2)**. The initial phase of lipid metabolism downregulation coincides with the cell growth phase (hypertrophy before the initiation of the proliferative phase) observed after PH. A negative correlation between lipid metabolism and cell size has been reported[156]. Further, the RNAi of SREBs that are involved in lipid metabolism results in cell size increase.

On the other hand, genes that control β-oxidation follow the inverse profile to *de novo* lipogenesis pathway, with its expression returning to baseline during the proliferative phase (cluster 1, path D) **(Figure A3)**. PDK4, a key regulator coordinating lipid metabolism with liver growth[157], is upregulated (cluster 1, path D) **(Figure A4)**. UCP2 that controls proliferation by inhibiting the switch from fatty acid oxidation to pyruvate utilization, is also upregulated (cluster 1, path A)[158]. This is accompanied by the upregulation of LDHA (cluster 1, path A) that can promote pyruvate utilization.

2.3.3 Gene expression pattern affecting the levels of GSH, NAD and SAH

Genes of glutathione metabolism are also transiently downregulated (paths P and L) **(Figure 2.9)**. These include GCLC and GCLM involved in *de novo* synthesis of GSH, which plays an important role in scavenging reactive oxygen species and maintaining the redox balance. GCLC and GCLM encode the catalytic and modifier subunits of glutamate cysteine ligase (GCL), respectively. GCL catalyzes the rate-limiting step involved in the generation of γ-glutamyl cysteine (γ-GC) from glutamate and cystine. γ-GC and GSH levels are known to be regulated by inflammatory signalling[159]. However, glutathione S-transferases (GSTM1, GSTM2, GSTM3, GSTM4, GSTM6, and GSTM7) are transiently downregulated in 10 hours. Increased GSH levels are reported in HCC and during liver regeneration[160]. GSH deficiency interferes with liver regeneration after PH[161]. In the NAD

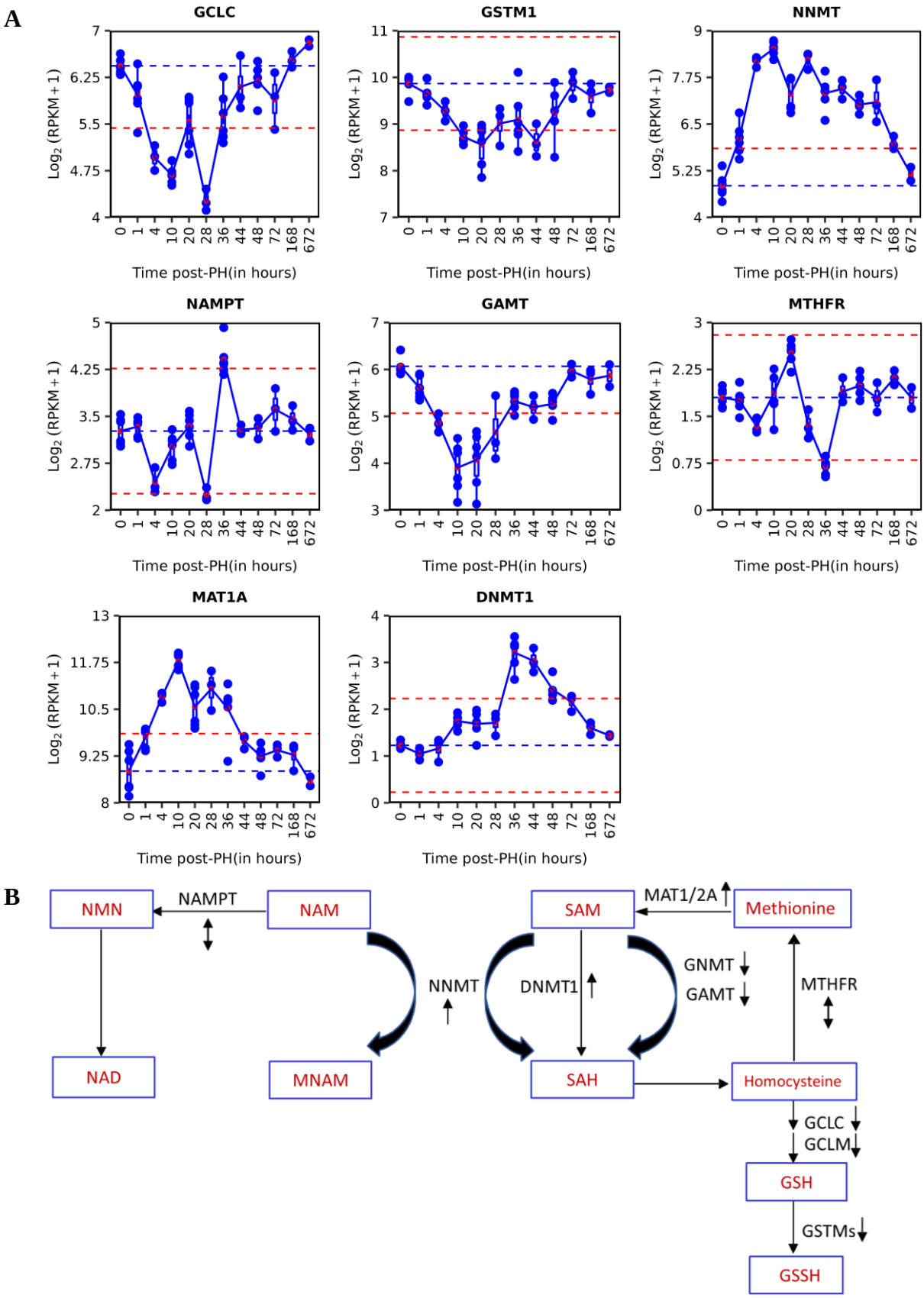


Figure 2.9: (A) Expression profile of genes affecting GSH, NAD, and SAM. The blue line represents the baseline (0 hours), and the red line represents a two-fold change with respect to the baseline. (B) Mapping of gene expression to metabolites. The arrow indicates the direction of change in gene expression over the period of regeneration.

salvage pathway, we observed that NNMT is upregulated (cluster 1, path E) during liver regeneration, while NAMPT levels fluctuate with downregulation at 4 and 28 hours (cluster 3, path L) **(Figure 2.9A)**. However, NAMPT is upregulated at 36 hours, as shown previously[126]. Deletion of NAMPT is shown to affect cell proliferation during liver regeneration[162]. NNMT knockdown leads to an increase in lipogenic gene expression and a decrease in gluconeogenic gene expression. Both NAMPT and NNMT control lipid, cholesterol, and glucose metabolism by stabilizing SIRTs[163]. NNMT is at the crossroads of metabolism and epigenetic regulation, but it is not the major methyltransferase in the liver to maintain S-adenosyl-methionine (SAM) to S-adenosyl homocysteine (SAH) ratio[164]. However, NNMT overexpression can decrease NAD levels, reduce methylation capacity, and promote liver steatosis and fibrosis[165].

We observed that the expression of main methyltransferases of the liver (GNMT, GAMT) involved in the conversion of SAM to SAH is downregulated (cluster 3, path P), and genes of methionine catabolism (MAT1A and MAT2A) involved in the production of SAM are upregulated (cluster 1, paths B and E). This shows that SAM levels may increase during regeneration and contribute to epigenetic control. This is further supported by the upregulation of genes involved in DNA methylation (DNMT1 and UHRF1) (cluster 2, path H) **(Figure 2.9A)**, which is consistent with the observation by Wang *et al*., (2019)[140]. Both these genes are co-expressed with cell cycle genes in the mid-phase, while genes involved in SAM production are upregulated very early on. The expression of MTHFR involved in the generation of methionine fluctuates with an increase at 20 hours and decrease at 36 hours. Overall, our analysis captures the systems-level changes in gene expression affecting the GSH, NAD and SAH levels **(Figure 2.9B)**.

2.3.4 The role of alternative splicing in liver regeneration

Genes related to the mRNA cleavage and polyadenylation are co-expressed with cell cycle genes in clusters 1 and 2. PTBP1 and RBMX involved in the regulating alternate splicing events are co-expressed with the cell cycle genes, and fetal liver gene AFP (cluster 2, path J) **(Figure 2.10)**. The path J also includes genes encoding the components of the spliceosome complex and the factors involved in its assembly (EFTUD2, SNRPD1, SNRPA1, LSM2/3, PPIL1, PPIH). On the other hand, splicing factor genes SRSF3 and SRSF7 are co-expressed and immediately upregulated (cluster 1, path B) post-PH compared to PTBP1 and RBMX. SRSF3 and SRSF7 are also involved in alternate polyadenylation of

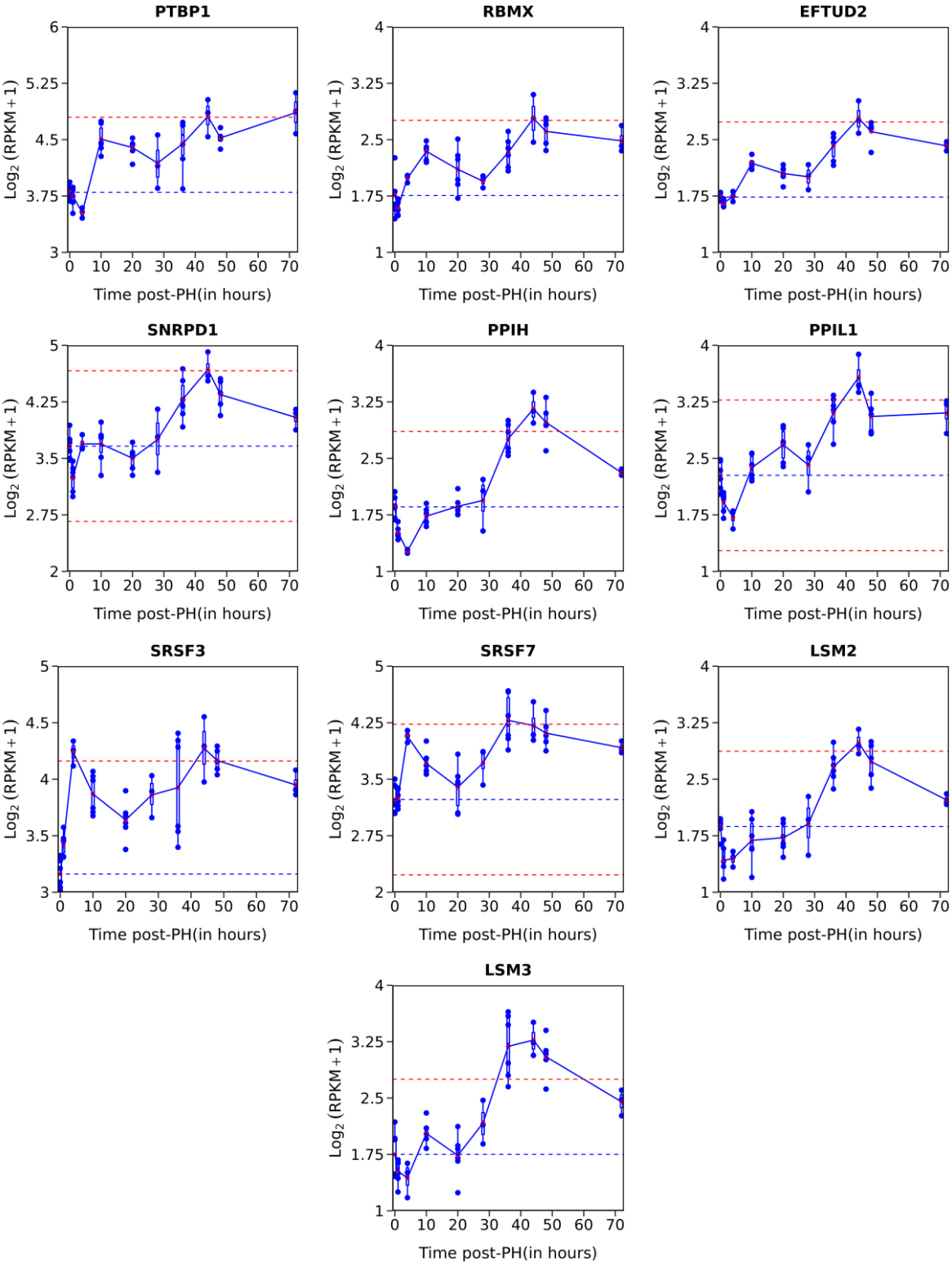


Figure 2.10: Expression profile of genes involved in the RNA splicing. The blue line represents the baseline (0 hours), and the red line represents a two-fold change with respect to the baseline.

mRNA precursors[166]. SRSF3 is known to promote metabolic homeostasis and limit cell proliferation[167]. A decrease in SRSF3 expression is observed in mouse models of NAFLD

and NASH[168]. A global change in alternate splicing machinery is also observed in HCC and NAFLD[169,170].

2.3.5 A model of balance between liver identity and proliferation during liver regeneration

Our results indicate that metabolic functions such as cholesterol biosynthesis, bile biosynthesis, and lipogenesis are downregulated, while proliferation markers are upregulated. These results suggest that a transient decrease in hepatic metabolism may counterbalance hepatocyte function vs. hepatocyte growth and proliferation. The transcription factor enrichment analysis showed HNF4A as a TF governing cluster 1 and 3 **(Table 2.1)**. The staining for HNF6 (Onecut1) and HNF4A, two key hepatocyte TFs, show a decrease in replicating cells[171]. Deletion of HNF4A in mice leads to an increase in proliferation[172]. We observed that liver identity genes (223 out of 622 identified genes)[173] overlap with immediately upregulated cluster 1 and downregulated cluster 3. On the other hand, there is no overlap with cluster 2, the upregulated cell cycle cluster activated post 28 hours. We also observed that there is less overlap between liver identity genes and cell cycle clusters in the validation dataset. This pattern of gene expression may be due to the heterogeneity in the hepatocyte functions. Single-cell transcriptomic data of liver regeneration has shown that there is a functional bifurcation of hepatocytes into proliferative and non-proliferative cells[146]. The hepatic function genes are upregulated predominately in non-proliferating cells, while these genes are downregulated in dividing cells.

A mathematical model of a simple HNF4A regulatory circuit **(Table A3)** was developed to demonstrate how the division of labour occurs. We hypothesize that a mutually exclusive behaviour between liver function and cell division can be established by the feedback loop regulation between these processes during liver regeneration. The regulatory circuit connecting HNF4A and cell cycle was established based on the literature[174–177]. This includes mutual antagonism between CCND1 (CYCLIN D) and HNF4A **(Figure 2.11A)**. Deletion of HNF4A results in increased expression of MYC and CYCLIN D, while CYCLIN D represses the transcriptional activity of HNF4A. This double negative feedback loop circuit can be regulated by a plethora of signals (local and in circulation) activated during the priming phase of liver regeneration. The PH can alter the balance between

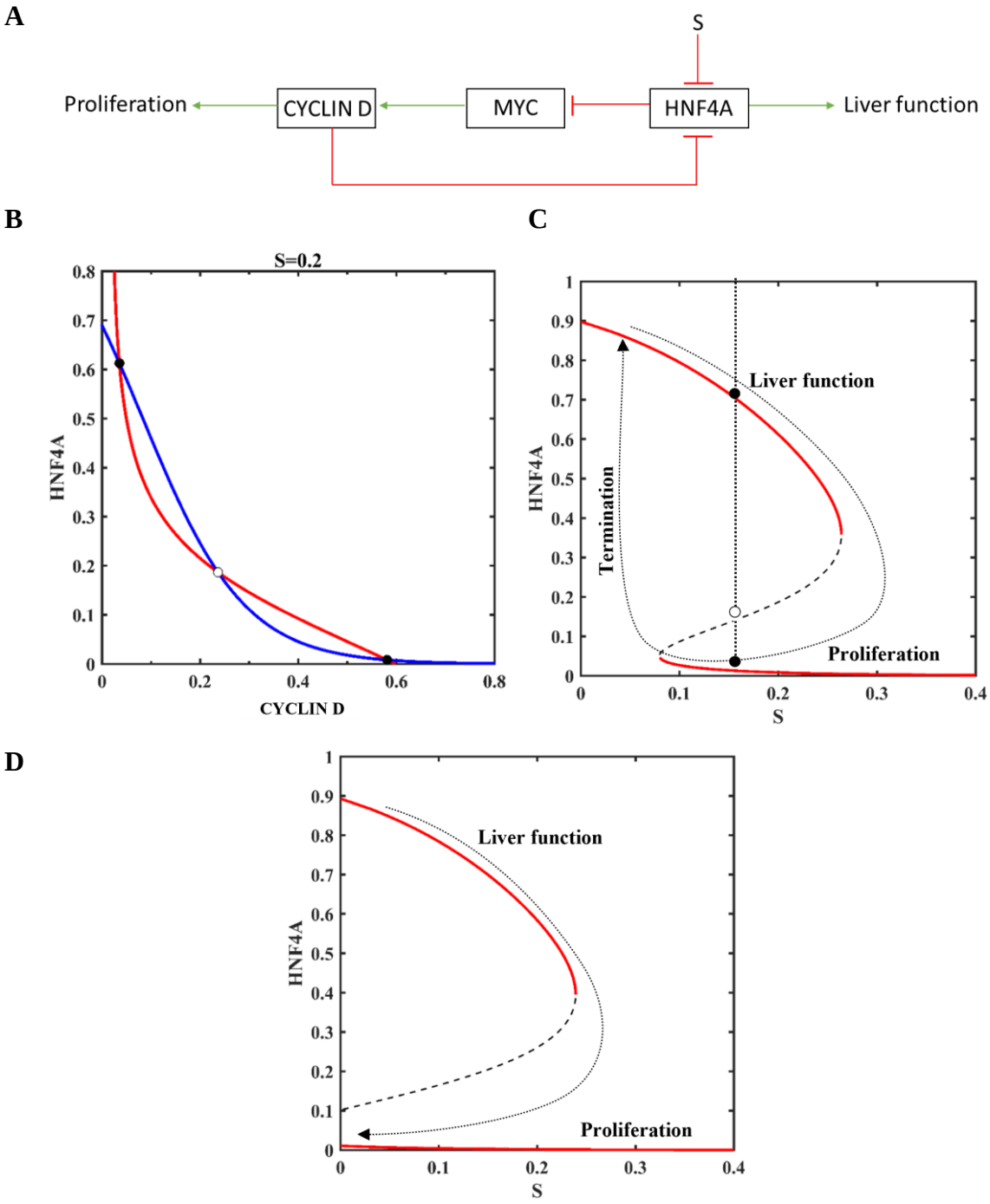


Figure 2.11: A model of balance between liver function and regeneration. (A) A core regulatory circuit of mutual antagonism between CYCLIN D and HNF4A controlled by inhibitory input stimuli S is shown. The activation is shown in green, and inhibition in red. (B) Phase plane showing the nullclines of HNF4A (blue) and CYCLIN D (red) for S=0.2. (C) Bistable inactivation of HNF4A with an increase in input stimuli S. (D) Irreversible inactivation of HNF4A with an increase in feedback loop strength (w_{HNF4A_CycD}= -1.5). The solid circle represents the stable steady state, and the open circle represents the unstable steady state.

mitogen and mitoinhibitors by matrix remodelling, induce secretion of ligands and cytokines and change the circulating levels of metabolites in plasma. Further, the underlying

mechanism of control of this core double negative circuit can vary since hepatocytes express different genes depending on their location in the hepatic lobule along the periportal-pericentral axis. We propose two kinds of inputs: proliferative and compensation signals that can act on the core circuit in a context-dependent manner in different hepatocytes to bring about different outcomes. The proliferative input suppresses HNF4A and activates CYCLIN D expression, while compensation signals promote HNF4A activation.

The HNF4A circuit was translated into a discrete-continuous model with parameter values chosen to yield two distinct attractor states corresponding to hepatocyte function and proliferation. The phase plane of the model shows that the nullclines of HNF4A and CYCLIN D can intersect at three points creating two stable and one unstable steady states **(Figure 2.11B)** for the parameter values given in **Table 2.2 (Model 1)**. This circuit exhibits bistable characteristics depending on the strength of the proliferation signal S **(Figure 2.11C)**. Bistability depends on the higher δ values that make the sigmoidal function sharp **(Table 2.2).** The two stable states correspond to hepatocytes in proliferative (high CYCLIN D) and differentiated (high HNF4A) states. At the intermediate signal strength, two populations of hepatocytes (differentiated and proliferation states) co-exist. The flipping of HNF4A and CYCLIN D occurs on the timescale of 1/β. It takes about 10 time units (S = 0.3) for the system to go to a new steady state, which approximately coincides with the experimental gene expression profile of CCND1 and the targets of HNF4A **(Figure 2.12)**. Trajectories in cluster 1 and cluster 2 correspond to the proliferative state, while trajectories in cluster 3 correspond to hepatocyte function state. Cell cycle genes in cluster 1 (CYCLIN D) are required to activate transcription factor E2F corresponding to cluster 2. CYCLIN D is co-expressed with Cdk inhibitor (CDKN1A); their ratio controls the passage through the restriction point and E2F activation.

The liver regeneration program after PH can be viewed as changes occurring around the bistable switch. This leads to transient loss of hepatocyte identity which is characterized by a transient increase in the expression of CYCLIN D (CCND1) and a decrease in HNF4A targets. Time-series gene expression data show an immediate rise in Cyclin D (cluster 1, path E) and a decrease in HNF4A targets: GJB1, IVD (cluster 3, path P) and CYP2C37, ALAS2 (cluster 3, path Q) **(Figure 2.12)**. The re-activation of HNF4A with the decrease in the input signal (due to repair) becomes essential for the termination of liver regeneration

Table 2.2: Parameter values used for the phase plane and bifurcation analyses. Model 1 corresponds to Figure 2.11A, and Model 2 corresponds to Figure 2.15A. The input S (proliferative) and M (compensatory) are varied.

Parameter	Description	Model 1	Model 2
β_{HNF4A}	Time scale for rate of change of HNF4A	1	1
δ_{HNF4A}	Steepness of soft-Heaviside function for HNF4A	8	8
w_{HNF4A_0}	Basal coefficient for HNF4A	0.3	0.5
w_{HNF4A_CycD}	Coefficient for HNF4A inhibition by CYCLIN D	-1.2	-1.2
w_{HNF4A_SNAIL}	Coefficient for HNF4A inhibition by SNAIL	-	-0.2
β_{CycD}	Time scale for rate of change of CYCLIN D	1	1
δ_{CycD}	Steepness of soft-Heaviside function for CYCLIN D	8	8
w_{CycD_0}	Basal coefficient for CYCLIN D	-0.5	-0.5
w_{CycD_Myc}	Coefficient for activation of CYCLIN D by MYC	1.1	1.1
δ_{Myc}	Steepness of soft-Heaviside function	8	8
w_{Myc_HNF4A}	Coefficient for MYC inhibition by HNF4A	-0.5	-0.5
β_{SNAIL}	Time scale for rate of change of SNAIL	-	1
δ_{SNAIL}	Steepness of soft-Heaviside function	-	8
w_{SNAIL_0}	Basal coefficient for SNAIL	-	-0.1
w_{SNAIL_HNF4A}	Coefficient for SNAIL inhibition by HNF4A	-	-1
w_{SNAIL_SNAIL}	Coefficient for self-activation of SNAIL	-	1.3
w_{S_HNF4A}	Coefficient for HNF4A inhibition by S (input)	-1	
w_{M_HNF4A}	Coefficient for HNF4A activation by M (input)	1	

(shown by an arrow in **Figure 2.11C**), which is characterized by downregulation of CYCLIN D and upregulation of HNF4A targets. The threshold for inactivation and re-activation (saddle nodes – SN1 and SN2) of HNF4A is less sensitive to variations in each parameter values (for a 10% increase or decrease from default parameter values) **(Figure 2.13)**. The proposed mechanism is minimalistic and may involve additional feedback loops that can further contribute to the robustness of the model. The development of HCC can be explained by the shift in re-activation threshold to a negative regime with the change in feedback loop strength, making the transition irreversible **(Figure 2.11D)**. On the other hand, the compensation signal shifts the HNF4A nullcline up resulting in one stable steady state corresponding to hyperactivation of HNF4A as observed by Chembazhi *et al*., (2021) **(Figure 2.14)**.

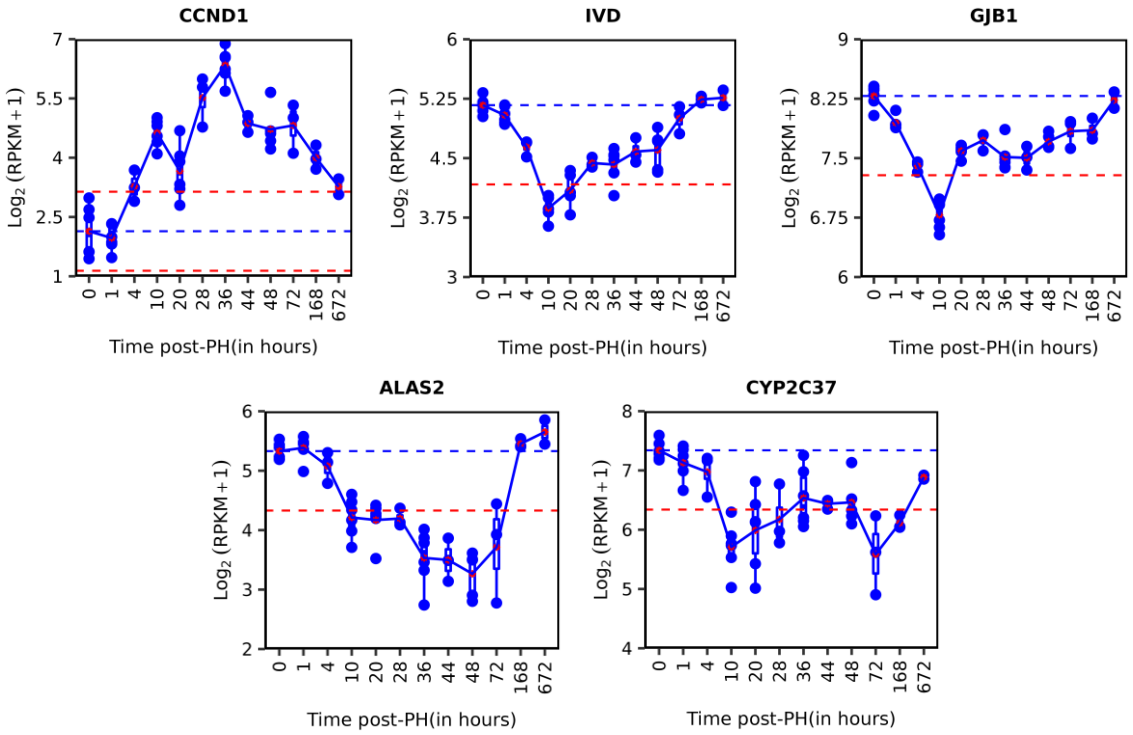


Figure 2.12: Transient activation of CYCLIN D (CCND1) and inactivation of HNF4A (IVD, GJB1, ALAS2 and CYP2C37) targets during liver regeneration. The blue line represents the baseline (0 hours), and the red line represents a two-fold change with respect to the baseline.

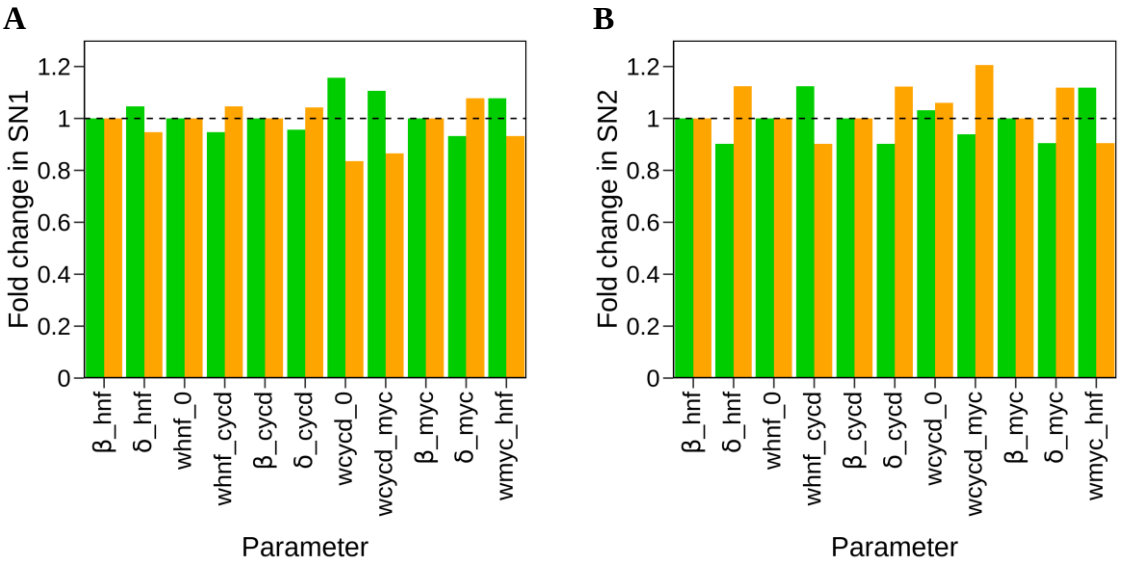


Figure 2.13: Effect of parameters on the HNF4A bistable response. A) The fold change in threshold for inactivation (saddle node, SN1) and B) re-activation (saddle node, SN2) of HNF4A in Figure 2.11C with respect to variations in each parameter values is shown. The parameter values were varied in the range 10% from default parameters values (as given in Table S3). Green and orange represent the increase and decrease in the corresponding parameter value. The fold change is SN (new)/SN (default).

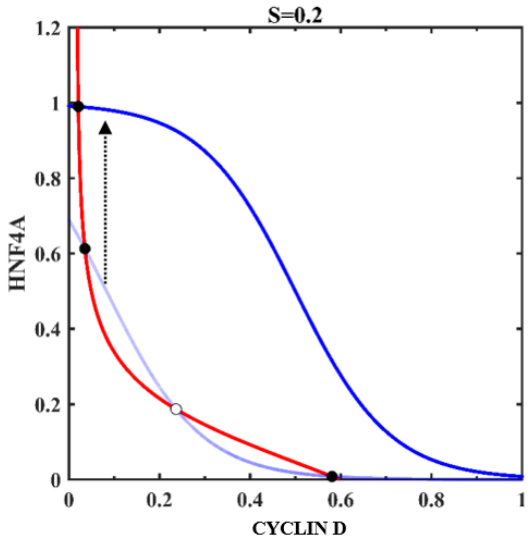


Figure 2.14: Phase plane showing the hyperactivation of Hnf4a during liver regeneration. The nullcline of HNF4A (blue) shifts above (arrow) in the presence of activatory stimuli (M=0.5). Nullclines of CYCLIN D (red) and HNF4A (light blue) for inhibitor stimuli S=0.2 are given for reference. The solid circle represents the stable steady state, and the open circle represents the unstable steady state.

The single-cell omics study of liver regeneration has also shown the existence of a hybrid cluster enriched for epithelial and hepatocyte-specific features and markers of mesenchymal cells[146]. The repression of the mesenchymal program is also required to maintain liver identity[178]. HNF4A and epithelial-to-mesenchymal transition (EMT) master regulatory gene SNAIL form a mutual inhibition circuit, which controls the balance between liver differentiation and mesenchymal program **(Figure 2.15)**. EMT control network involves SNAIL-induced self-activation of ZEB1[179]. An inhibition of HNF4A by proliferative signal can activate EMT in some hepatocytes. The phase plane analysis of this control circuit **(Table A4)** shows tristability (co-existence of three stable steady states) depending on the strength of proliferation signal S **(Figure 2.15 B - D)** for the parameter values given in **Table 2.2**. These states correspond to (1) high HNF4A with low SNAIL/ZEB1, (2) low HNF4A with high SNAIL/ZEB1, and (3) high HNF4 and SNAIL/ZEB1 (hybrid state). The auto-activation loop on SNAIL/ZEB1 helps in generating an intermediate hybrid state by making the nullcline of SNAIL/ZEB1 bistable (**Figure 2.15**). The proposed regulatory circuit accounts for the cellular plasticity during liver regeneration. Multistable characteristics may be relevant to understand the emergence of different states and their transitions during liver regeneration.

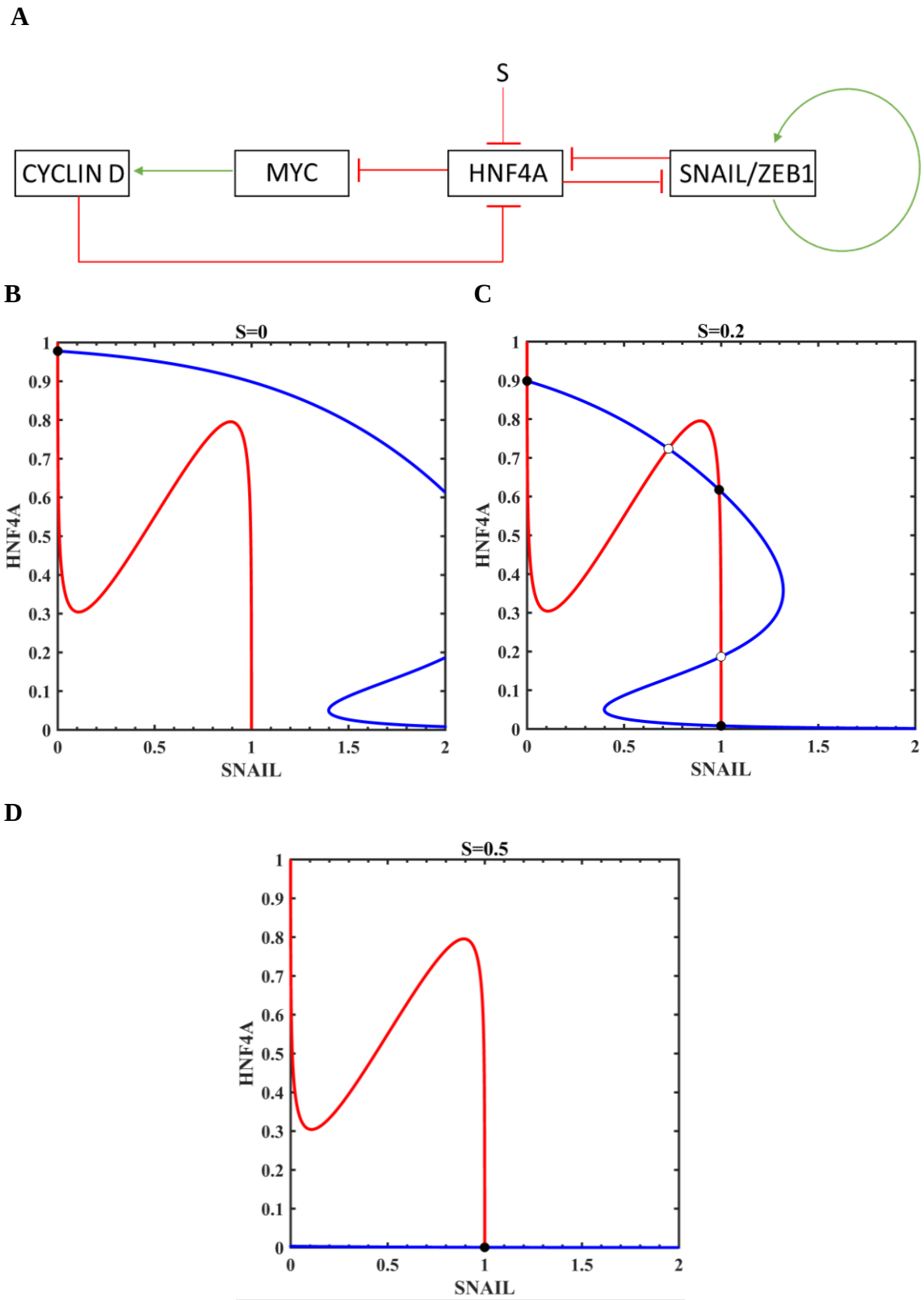


Figure 2.15: Multistability of the underlying circuit of liver regeneration. (A) The regulatory circuit controlling the balance between liver function and EMT during liver regeneration is shown. Activation is shown in green and inhibition in red. Nullclines of HNF4A (blue) and SNAIL/ZEB1(red) are shown for different input stimuli (B) S=0, (C) S=0.2, and (D) S=0.5. The solid circle represents the stable steady state, and the open circle represents the unstable steady state.

2.4 Discussion

The liver balances its function and proliferation demand after injury or resection. Recent advancement in high throughput techniques is helping to understand further the regulatory mechanisms involved in the regulation of liver regeneration. In this work, we performed a comprehensive analysis of how the transcriptome is reorganized into modules/trajectories and reported co-regulation of different biological processes using bulk RNA-Seq data of liver regeneration after PH. We also analysed the results obtained based on scRNA-seq data of liver regeneration. The dynamic network reconstruction revealed the trajectories of major pathways that are upregulated and downregulated (transient vs. sustained) during liver regeneration **(Figure 2.4** to **Figure 2.7).** We found three clusters specific to liver metabolism, cell cycle and immune response. This is consistent with the original study[139] that showed genes involved in liver function decrease in expression post-PH, whereas cell cycle, RNA metabolism, and protein modification genes display higher expression levels. The clustering also mirrors the reported difference in recruitment of RNA polymerase and histone modifications found in the early and proliferation phases. Our analysis supports the model of mutual antagonism between liver function and proliferation in liver regeneration. We show multistability of the underlying network of liver regeneration.

While overall metabolic downregulation suggests a decrease in liver function, the dynamics of metabolic pathways suggest that maintaining the levels of specific metabolites is required for liver regeneration. We observed that fine-tuning of SAM levels might be required for the methyltransferase reactions in liver regeneration (**Figure 2.9**). This is supported further by the downregulation of major liver methyltransferases. The co-expression pattern of cell cycle and DNA methylation genes highlights the scenario for crosstalk between cell cycle and chromatin regulatory proteins. Genes of NAD, glutathione, and lipid metabolic pathway decreased and reappeared immediately in the priming phase of liver regeneration. This suggests a possible requirement of these pathways for the cell cycle progression. Although NNMT expression correlates with adiposity, its expression during liver regeneration may be beneficial. Hepatic steatosis is shown to alter the demand for NAD and GSH[180]. Lipid metabolic genes are also further upregulated at 36 to 72 hours, respectively, coinciding with the proliferative phase **(Figure A1).** Along with a decrease in the liver's metabolic function, our analysis also captures the dynamic changes in metabolism that may indicate the shift from growth to proliferative phase during liver regeneration.

Another pathway that is upregulated during liver regeneration is protein processing in ER and protein transport **(Figure 2.5B and Table A1)**. ER stress plays a role in liver metabolism, damage, and inflammation[181]. The knockdown of XBP1 results in liver injury and impairment of liver regeneration[182]. Loss of IRE1, an upstream activator of XBP1, impairs liver regeneration with activation of STAT3 affected[183]. ER stress is also shown to suppress the liver identity genes in the damaged liver[173]. On the other hand, an increase in ER stress under HFD conditions can impair liver regeneration[184]. We also observed ribosome biogenesis and RNA processing as important features of liver regeneration. Ribosome biogenesis is known to increase during cell growth and proliferation[185]. Liver maturation and dedifferentiation are associated with alternative splicing mechanisms[167,170]. 3,5 diethoxicarbonyl-1,4 dihydrocollidine (DDC) treatment leads to liver regeneration and a switch to a fetal splicing program[138]. These observations suggest that mRNA cleavage and polyadenylation may also control gene expression during liver regeneration after PH in addition to epigenetic regulation.

The co-expression pattern of genes suggests a mutually exclusive behaviour of the cell cycle and liver identity genes during liver regeneration. Few liver identity genes were upregulated and were co-expressed with Cdk inhibitor (CDKN1A) and activator (CCND1) (cluster 1). An increase in Cdk inhibitor level may provide a window of opportunity for hepatocytes to grow before dividing. Alternatively, the co-expression of liver identity genes with Cdk inhibitor may suppress the cell division and maintain liver function[186]. On the other hand, CYCLIN D expression alone influences the transcriptional regulation of liver metabolism[175]. Distinguishing these effects requires single-cell level quantification. Mathematical modelling showed that interaction between regulators of cell cycle and liver function could make the system bistable (**Figure 2.11**), which accounts for the co-existence of two populations of hepatocytes, with one undergoing cell division while the other helping to maintain liver function[146,147]. The bistable switch accounts for transient inactivation of HNF4A with dynamic change in input signals during liver regeneration. We highlighted that the transition from liver function to proliferation could become irreversible with a change in the feedback loop strength. In this picture, the termination of liver regeneration depends on the re-activation of HNF4A, which is consistent with Huck *et al.,* (2019)[187]. Different studies have reported HNF4A inactivation in HCC[188–190]. We also showed that multistability emerges by coupling the HNF4A feedback loop with the EMT circuit **(Figure 2.15).** The EMT circuit is also known to exhibit tristability in cancer progression[191].

We propose an integrated circuit of liver regeneration by extending the core circuit **(Figure 2.16)**. The cell cycle and EMT control during liver regeneration may involve the YAP1/Hippo and Wnt/β-catenin signalling pathways converging on HNF4A inactivation. YAP1, a mechanical rheostat acting downstream of the Hippo signalling pathway, has a direct role in hepatocyte differentiation by inhibiting HNF4A and activating SNAIL[192]. Reciprocal control of YAP1 by SNAIL (activating) and HNF4A (inhibiting) has been shown, resulting in a complex circuit of multiple feedback loops controlling liver identity. The ectopic activation of YAP1 is sufficient to de-differentiate hepatocytes into cells with stem cell-like characteristics[193]. The early phase of liver regeneration is accompanied by YAP1 activation and nuclear location[48]. YAP1 also cooperates with MYC in the control of proliferation[194]. This mechanism explains the existence of proliferative hepatocytes undergoing EMT during liver regeneration[146].

Wnt/β-catenin signalling pathway is also induced in response to liver regeneration under PH[47,195]. WNT can also control the same circuit of HNF4A and EMT. It is known that WNT and HNF4A mutually inhibit each other, and WNT activates SNAIL to control the EMT[196]. MYC and β-catenin cooperate in liver carcinogenesis with YAP1 as a mediator[197]. It is also shown that sinusoidal endothelial cell Wnts drive proliferation, while macrophage Wnts drive functional compensation[186]. This integrated circuit of liver regeneration controlled by YAP1 and WNT may provide underlying features of proliferation, compensatory metabolism, and EMT states as observed in single-cell studies. Thus, simultaneous control of HNF4A may drive the bifurcation of hepatocytes into different activity states. On the other hand, both WNT and YAP1 have opposing functions to establish liver zonation[47]. Wnt/β-catenin signalling is active in pericentral hepatocytes, while YAP1 is expressed in the periportal region. It will be interesting to study further the factors that control the dual role of YAP1 and WNT in liver regeneration and zonation. Overall, our study provides a systems-level view of liver regeneration post-PH. The underlying gene modules identified here can be connected to the phenomenological model of liver regeneration[58] to obtain the dynamical characteristics of entry and exit from liver regeneration.

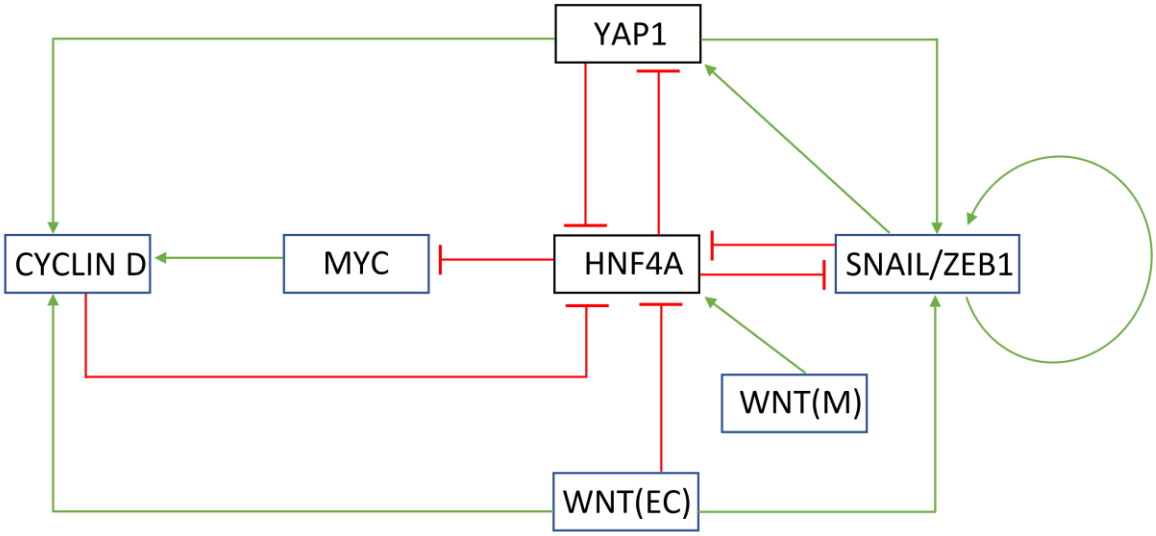


Figure 2.16: The proposed integrated circuit of liver regeneration controlled by YAP1 and WNT. Activation is shown in green, and inhibition in red. EC-endothelial cell, M-macrophage. Circular arrow represents autoregulation.

Chapter 3 Network-level analysis of ageing and its relationship with diseases and tissue regeneration in the mouse liver

3.1 Introduction

Ageing is an inevitable complex process altering a multitude of cellular processes. Several studies employing animal models across different organs have outlined the general hallmarks of ageing related to epigenetic modifications, cellular senescence, altered intercellular communication, telomere shortening, nutrient sensing deregulation, mitochondrial dysfunction, stem cell exhaustion, loss of proteostasis, genomic instability, which culminate in the loss of tissue homeostasis[198]. The complexity of the ageing process is further heightened by the interconnected features of some of these processes[199]. Different factors are suggested to cause or contribute to ageing, including DNA damage, free radical accumulation and metabolic dysfunction[200]. The oxidative theory of ageing proposes macromolecular damage by the products of metabolism and inefficient repair.

Molecular pathways involving IGF1/GH and mTOR have been implicated in the ageing process[201,202]. Caloric restriction and mTOR inhibition by rapamycin slow down many age-dependent processes and extend lifespan[203,204]. With the advent of high-throughput techniques, biological processes underlying the initiation and progression of ageing can be unfolded at the systems level. However, most studies focused on identifying DEGs and patterns of gene expression in ageing to characterize the transcriptomic changes[205–207]. The upregulation of inflammatory, immune and stress response genes has been reported in different microarray and RNA-seq experiments of ageing in mice[208,209]. The inflammaging theory postulates ageing accrues inflammation[210]. Tissue-wise transcriptomics study across multiple age groups in mice shows distinct gene expression signatures in different organs, with the liver undergoing extensive changes over time compared to other tissues[206]. The liver is an important metabolic organ that plays a vital role in synthesizing plasma proteins, clotting factors, triglycerides, cholesterol, glycogen, and detoxification[2,211]. Therefore, understanding how ageing rewires the regulatory network of the liver is crucial.

The impairment of structure and function of liver tissue with ageing exacerbates the risk of liver diseases and affects its regeneration potential after damage[135]. Non-alcoholic fatty liver disease (NAFLD) is the commonly seen pathological condition of the liver that evolves into non-alcoholic steatohepatitis (NASH), cirrhosis and hepatocellular carcinoma

(HCC). The progression of NAFLD to NASH and further to HCC is favoured by increased inflammation in old age[210]. Interwinding nature of liver ageing and age-related diseases may create a futile cycle of each fuelling the other, leading to a transition from chronological ageing to pathological ageing. In addition to increasing the disease risks, ageing also delays regeneration after partial hepatectomy[135]. Most of the studies designed to understand liver diseases were dealt independent of each other and without involving the intrinsic process of ageing[212]. Delineating the shared mechanisms inherent to the ageing process and age-related diseases shows a road ahead, thereby suggesting therapeutics for liver diseases that are influenced by age.

Network-based approaches can be applied to understand the dynamic changes in gene expression patterns with lifespan and to study the crosstalk between ageing and ageing-related diseases. This provides a systems-level understanding and helps to map dynamical changes. The PPI network provides a scaffold to integrate gene expression data and study the statistical and topological properties of the network in the context of liver ageing and its related diseases. The usage of the PPI network helps to distinguish direct and indirect control compared to the correlation-based co-expression network.

In this work, we studied how the statistical properties of the liver network change with ageing by integrating the PPI network and mRNA expression profiles of mouse liver samples across ten different age groups available from Tabula Muris Consortium[207]. Network entropy quantifying randomness offers a new perspective for studying ageing and diseases. We show that entropy of the liver network increases with ageing, indicating the increase in randomness due to network disruption by genomic alterations. We computed the local entropy measure to identify genes and pathways associated with ageing. The genomic alterations in ageing may either increase or decrease the randomness of the local connectivity patterns (change the probability of interactions)[213–215]. A decrease in entropy signifies specific signalling interactions with higher weights, while an increase in entropy signifies the unpredictable nature of interactions. To gain further insights, we integrated the DEGs between young and different age groups with the PPI network to identify core modules and nodes that show changes in local and global topological network measures with ageing. Finally, we computed the network proximity of the ageing network with different networks of liver diseases and regeneration to study the effect of ageing. The workflow of the study is shown in **Figure 3.1**.

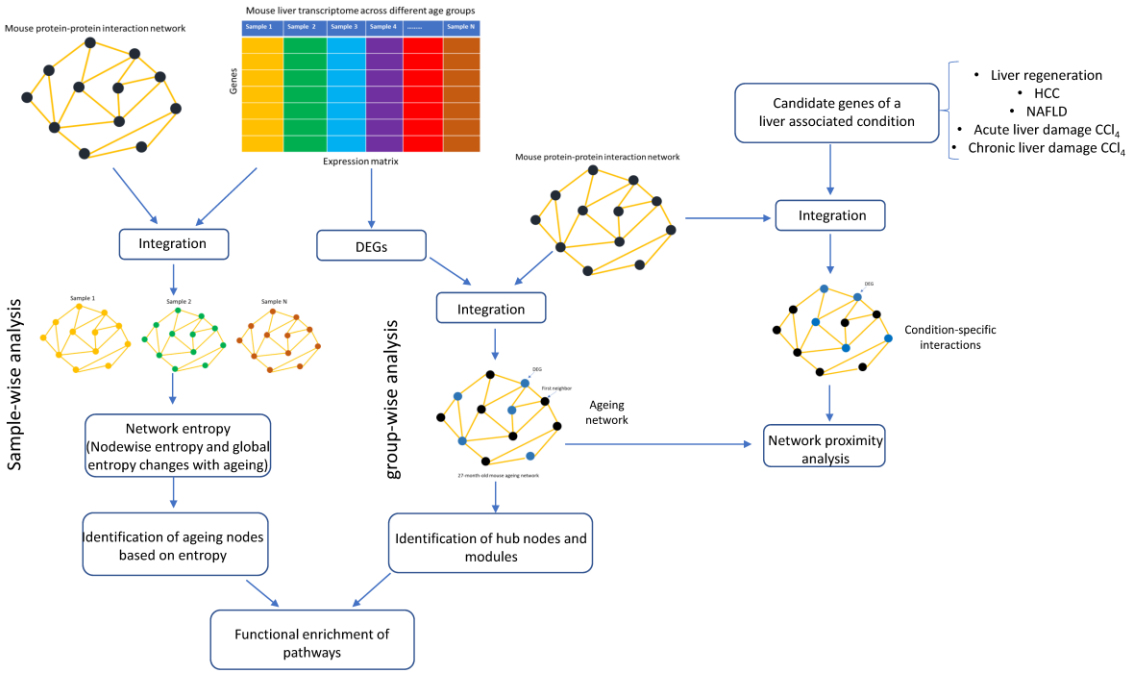


Figure 3.1: The workflow to study the network-level changes of ageing and its association with tissue regeneration and diseases in the mouse liver. The workflow is divided into three parts. Network entropy analysis was performed for each sample by integrating the mouse protein-protein (PPI) interaction network with liver ageing gene expression data. Topological analysis was performed by integrating differentially expressed genes (between two age groups) with PPI network. Network proximity analyses of ageing with different pathophysiological conditions (regeneration, NAFLD, HCC, acute and chronic damage)were performed using differentially expressed genes in each condition. The individual nodes were identified based on different network measures and enrichment of functional pathways performed.

3.2 Methods

3.2.1 Network entropy-based approach to analyze liver ageing

Transcriptomics data (bulk RNA-seq) of mouse liver tissue with age groups 3, 6, 9, 12, 15, 18, 21, 24, and 27 months was obtained from GEO with accession number GSE132040 (Tabula Muris Consortium). Each age group has 3-4 replicate samples. The raw count data was normalized using variance stabilizing transformation (VST)[216]. An entropy-based approach was used to integrate the gene expression data with the PPI network. Mouse-specific STRING PPI network (10596 nodes and 86074 edges) with interaction confidence-score cut-off ≥ 0.9 was used as the initial PPI network. A network characterised by a specific number of nodes, edges and edge weights is considered an instance in an ensemble of large number of networks with similar features. This system has two sets of observables related to degree sequence and distribution of edge weights. The entropy metric of a network is

given by calculating the maximum entropy of the ensemble satisfying the given constraints (with the identical topological and spatial structure of the network) rather than the original network **(section 3.2.1.1)**[215]. For the integration of gene expression and PPI network, nodes in the PPI network are assigned their corresponding gene expression values specific to a particular sample. The edges connecting nodes are weighted as the distance between gene expression values. The edge weights are converted to a distribution by partitioning them into number of bins equal to the square root of number of nodes in the network. While building the network, nodes with zero gene expression value are removed from the network. Hence, the final network, which is subjected to entropy maximization, differs from the original PPI network and is sample-specific. Therefore, the static PPI network evolves when it is integrated with sample-specific gene expression.

3.2.1.1 Entropy calculation by integration of PPI with transcriptomics data

In the case of PPI network integration with transcriptomics, a system is a network with N nodes forming an adjacency matrix $A = \{a_{ij}\}\ \forall\ i, j \in 1,2,\ldots,N$, weights d_{ij} as edge weights. This system has two sets of observables related to topology and distribution of edge weights. Topology is an N-dimensional vector $\{k_i\}\ \forall\ i \in 1,2,\ldots,N$, called as degree sequence with number of connections of each node. Distribution of edge weights is related to the partitioning of distance between nodes into N_b bins, where $N_b = \sqrt{N}$. Distance between nodes i and j of a sample a is given as the Euclidean distance of the gene expression (g) values i.e.,

$$d_{ij} = \sqrt{\left(g_i^a - g_j^a\right)^2} = \left|g_i^a - g_j^a\right| \tag{3.1}$$

where a corresponds to the sample identity in the dataset and i runs through all the nodes of a network. Each of the distance bin N_b takes a value corresponding to number of links whose associated edge weights fall inside its boundaries. A network ensemble is imposed to have topological constraint similar to the degree sequence of the sample-specific network and the spatial constraint to have same average value of links per bin. Therefore, topological and spatial constraints based on the expression profile are specific to each sample and hence the entropy values. The entropy of the network ensemble is defined as

$$S = -\sum_{i<j} p_{ij}\, log\, p_{ij} - \sum_{i<j} \left(1 - p_{ij}\right) log\left(1 - p_{ij}\right) \tag{3.2}$$

where p_{ij} is the probability of finding a link between nodes i and node j. As with the canonical entropy, the members of an ensemble on an average satisfy the constraints.

The constraints on the degree sequence k_i and link distribution B_l are defined as follows:

$$k_i = \sum_{j}^{N} p_{ij}\,;\;\; i = 1,2,\ldots.,N \tag{3.3}$$

$$B_l = \sum_{i<j}^{N} \chi_l(d_{ij}) p_{ij}\,;\;\; l = 1,2,\ldots,N_{b,} \tag{3.4}$$

χ_l is the characteristic function of each bin of width $(\Delta d)_l$ such that

$$(\Delta d)_l : \chi_l(x) = 1\; if\; x \epsilon [d_l\,,\, d_l + (\Delta d)_l],\; \chi_l = 0\; otherwise \tag{3.5}$$

The probability matrix $\{p_{ij}\}$ is obtained by constrained maximisation of entropy function as:

$$\frac{\partial}{\partial p_{ij}}\left\{S + \sum_{i}^{N} \lambda_i \left(k_i - \sum_{j} p_{ij}\right) + \sum_{l}^{N_b} g_l \left(B_l - \sum_{i<j}^{N} \chi_l(d_{ij}) p_{ij}\right)\right\} = 0 \tag{3.6}$$

where λ_i and g_l are the Lagrangian multipliers related to the constraints. For each pair of i and j, the marginal probability is given as

$$p_{ij} = \sum_{l}^{N_b} \chi_l(d_{ij}) \frac{e^{-(\lambda_i+\lambda_j+g_l)}}{1 + e^{-(\lambda_i+\lambda_j+g_l)}} = \sum_{l}^{N_b} \chi_l(d_{ij}) \frac{z_i z_j W_l}{1 + z_i z_j W_l} \tag{3.7}$$

where, $z_i = e^{-\lambda_i}$ and $W_l = e^{-g_l}$ are functions of the Lagrangian multipliers λ_i and g_l.

3.2.1.2 Estimating single node entropy

The link probabilities p_{ij} obtained from the full PPI network of a sample can be further used to derive the entropy associated with a single node i that takes the form of Shannon entropy. This is valid since $p_{ij} > 0,\ \forall\ i = 1,2,\ldots.,N$ and $\sum_j p_{ij} = k_i$, the degree of node i. Single-node entropy S_i of *i-th* node is given as:

$$S_i = -\sum_{j} p'_{ij}\ log\, p'_{ij}\,;\quad p'_{ij} = \frac{p_{ij}}{k_i} \tag{3.8}$$

The Wilcoxon rank sum test was applied to identify nodes showing significant differences (FDR < 0.05) in entropy between groups of samples at the single-node level. This analysis was performed by considering samples of 3-6 months old mice as the younger age group

and samples of 24-27 months old mice as the older age group. The pathway enrichment analysis of nodes that display significant differences in single-node was performed using Enrichr[152] to obtain significantly affected pathways (adjusted p-value < 0.05).

3.2.2 Graph theoretical analysis of ageing PPI network

Unlike the previous approach, which integrates sample-specific gene expression with the PPI network for entropy calculation, we alternatively constructed the age group-wise networks to compare the local and global network measure changes with ageing. For this, DEGs comparing 3 months old mice with 18, 24 and 27 months old were used for building individual PPI networks. DEGs identified using the DESeq2 pipeline were integrated with STRING PPI (confidence-score cut-off ≥ 0.9) to build individual networks for comparison. Each PPI network was further expanded by including the first neighbours of DEGs, and this network was considered for all the downstream analyses.

Each PPI network was analyzed using the CytoHubba plugin in Cytoscape 3.9.0[153]. A PPI network is assumed to be an undirected network $G = (V, E)$ with V as a set of nodes and E as a set of edges connecting the nodes. CytoHubba identifies essential hub nodes and subnetworks within the PPI network using various local and global metrics[217]. Each of these metrics is associated with a function F, which assigns every node v a numeric value $F(v)$. A node u is awarded a higher rank compared to another node v if $F(u) > F(v)$. A local ranking method only considers the relationship between the node and its direct neighbours to calculate the score. On the other hand, a global ranking method assigns a score to a node based on its relationship with the entire network.

For local measure, we used Maximal Clique Centrality (MCC), which is based on the concept of a clique that emphasizes the highly connected clusters within a network. A clique C in a network is a subset of nodes ($C \subseteq V$) such that every pair of nodes is connected. Further, if such a clique cannot be extended by adding one or more other nodes (for any $x \in V \backslash C$, $C \cup \{x\}$ is not a clique), it becomes a maximal clique. MCC score for a node v is given as

$$MCC(v) = \sum_{C \in S(v)} (|C| - 1)! \tag{3.9}$$

where $S(v)$ is the collection of maximal cliques C, which contains v, and $(|C|-1)!$ is the product of all positive integers less than $|C|$. Therefore, a node with a higher MCC score implies that it is part of larger cliques or, many smaller cliques or both.

In addition to the connectivity of a node, its spatial position in the network also influences communication among other nodes. To capture the nodes that regulate the information flow within the network, we used two shortest path-based global measures, Bottleneck centrality (BN) and Betweenness centrality (BW), for each node. Bottlenecks are considered to act as bridges holding crucial functional and dynamic properties in the network[218]. While the *BN(v)* score of node v is based on the shortest path trees of all other nodes in the network, *BW(v)* is based on the number of shortest paths between every pair of nodes traversing the node v. Scoring of *BN(v)* for a node v begins with the construction of tree T_s of shortest paths from a node s to all other nodes in the network, followed by enumeration of the number of these shortest paths going through node v. This process is iterated for all $s \in V$. A node v in the shortest path tree T_s is considered as a bottleneck if more than $\frac{|V(T_s)|}{4}$ of the paths in the tree cross it[219], where $|V(T_s)|$ is the number of nodes in the tree. Finally, *BN(v)* of node v is scored as the number of such shortest path trees where it is considered as a bottleneck and is given by

$$BN(v) = \sum_{s \in V} p_s(v) \tag{3.10}$$

where $p_s(v) = \begin{cases} 1, & if\ number\ of\ paths\ from\ s\ to\ V(T_s)\backslash v\ through\ v > \frac{|V(T_s)|}{4} \\ 0, & otherwise \end{cases}$

Betweenness centrality *BW(v)* of node v in the connected component *C(v)* containing v is the sum of the fraction of shortest paths between every pair of nodes s and t traversing through v, $\sigma_{st}(v)$, to the total number of shortest paths between every pair of nodes s and t, σ_{st}, and is given by

$$BW(v) = \sum_{s \neq t \neq v \in C(v)} \frac{\sigma_{st}(v)}{\sigma_{st}} \tag{3.11}$$

Further, the densely connected components of the network that are likely to form molecular complexes were identified using the Molecular Complex Detection (MCODE)[220] program's default settings in Cytoscape. MCODE clusters with scores ≥ 5 were further analysed by using Enrichr for pathway enrichment.

3.2.3 Network-based proximity analysis

Gaining insights into the interconnectedness of disease genes with ageing within the PPI network helps to understand the risk of ageing. If disease modules in an interactome overlap or are significantly closer to ageing modules, then perturbations due to ageing may affect pathways in the disease or drive its progression. Proximity analysis was performed to study the associations between the ageing liver and each of the perturbed liver conditions (liver regeneration post-PH, NAFLD, HCC, acute liver damage by CCl_4 and chronic liver damage by CCl_4). The association between two conditions was quantified using a network proximity metric[221]:

$$< d_{AB}^{c} > = \frac{1}{||A|| + ||B||}\left(\sum_{a \in A} min_{b \in B} d(a,b) \; + \sum_{b \in B} min_{a \in A} d(a,b)\right) \qquad (3.12)$$

where *d(a, b)* represents the shortest path length between gene *a* from condition A and gene *b* from condition B in the interactome. The significance of this distance metric was evaluated using the Z-score of the permutation test by randomly selecting nodes from the whole network with degree distributions similar to that of the nodes in the two sets. Z-scores were calculated by permutation tests of 1,000 repetitions as follows:

$$Z_{d_{AB}} = \frac{d_{AB} - d_m}{\sigma_m} \qquad (3.13)$$

where d_m and σ_m are the mean and standard deviation of the permutation test.

Candidate gene lists for ageing and other conditions were selected from different studies with similar mouse strains **(Table 3.1)**. DEGs between 3- and 27-months old mice were considered as signatures of ageing for proximity analysis. We also performed proximity analysis using DEGs from different age groups, including 12, 18 and 21 months, for comparison. Candidates for different phases of liver regeneration were considered by taking the union of DEGs of early-phase (1, 4, 10h post-PH compared to pre-PH), mid-phase (36, 44, 48h post-PH compared to pre-PH), and late-phase (1- and 4-weeks post-PH compared to pre-PH). We also included DEGs of sham-operated control samples at different phases, i.e., early-phase (1, 4, 10h post sham surgery) and mid-phase (48h post sham surgery). Candidate genes for NAFLD and HCC (DEGs between healthy control and disease) were pooled from their respective studies **(Table 3.1).** The proximity analysis was performed using the high confidence mouse-specific STRING PPI network (confidence score ≥ 0.9).

Two conditions with Z-score < − 1.5 and FDR < 0.05 were considered significantly proximal. To infer the biological significance of proximity of ageing signatures with

Table 3.1: Datasets used to define the list of candidate genes for different liver-associated conditions.

Liver condition (Accession no.)	Accession no.	Experimental mouse model (Age of mice during sample collection)	Strain
Ageing	GSE132040	Age group spanning 3 – 27 months (3, 6, 9, 12, 15, 18, 21, 24, 27 months old)	C57/BL6J
Regeneration, sham-operated control	GSE95135	12-14 weeks old mice (3 months old)	C57/BL6J
NAFLD	GSE148080	Normal diet beginning at 8 weeks followed by 8-16 weeks of normal diet/high sucrose diet (8 months old)	C57/BL6J
	GSE184019	Normal diet at 8 weeks followed by 3 weeks of normal diet/high sucrose diet. Samples collected at 11 weeks. (3 months old)	C57/BL6J
HCC	GSE132728	Single dose of DEN at 2 weeks followed by weekly dose of CCl_4 from 8 weeks to 24 weeks (6 months old)	C57/BL6J
	GSE89689	Single dose of DEN at 2 weeks followed by first dose of CCl_4 dose 4 weeks later. Further weekly dose of CCL_4 for 15 weeks. Final samples were collected 10 weeks after the last dose of CCl_4 (8 months old)	C57/BL6J
Acute damage (CCl_4)	GSE167033	8-10 weeks old mice were administered with CCl_4. Samples were collected 2 and 8 hours post treatment, 1, 2, 4, 8, 16 days post treatment (2-3 months old)	C57/BL6/N
Chronic damage (CCl_4)	GSE167216	8-10 weeks old mice were treated with CCl_4 twice a week for 2, 6, 12 months (4, 8, 12 months old)	C57/BL6/N

other conditions, the shortest path connecting each pair of DEG sets was identified as depicted in **Figure 3.2**.

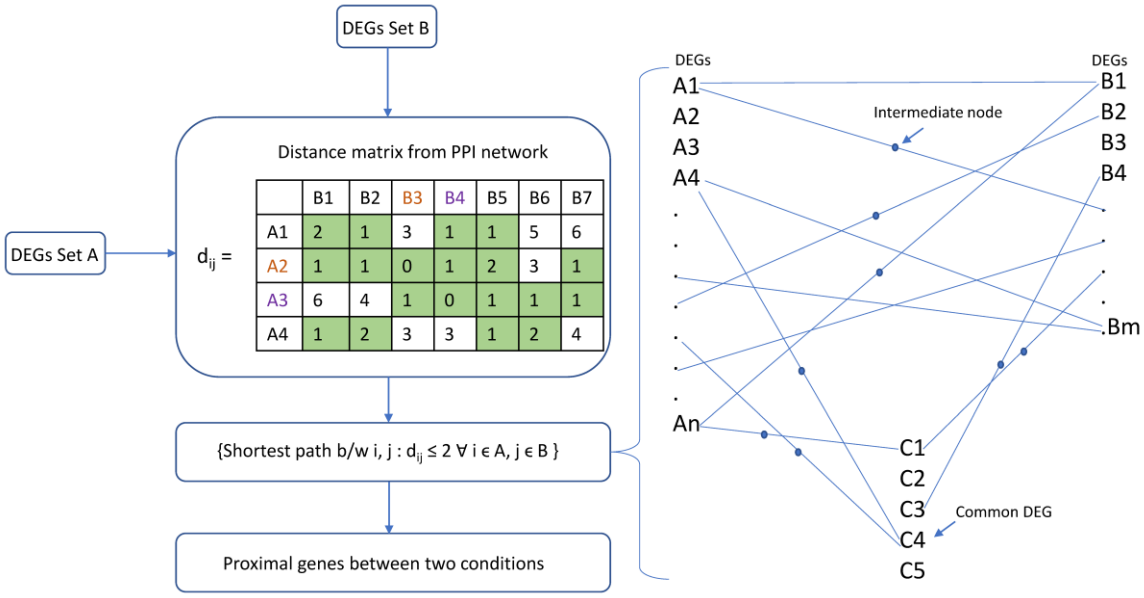


Figure 3.2: The flowchart to identify proximity genes between two conditions A and B. The shortest path connecting each pair of DEGs in condition A and B was identified. The nodes (A1, A2…An and B1, B2…Bm) can be directly connected or through an intermediate node. C represents common nodes between two conditions.

3.3 Results

3.3.1 Alteration in network entropy with ageing in the mouse liver

We used the network entropy measure to study ageing. The sample-wise gene expression was integrated with the PPI network. The estimation of network entropy from the liver gene expression data shows that entropy increases in old age (18-21 months) compared to the young (3-6 months) mice **(Figure 3.3)**. This is in agreement with other studies that used a similar approach to study the progression of ageing in the context of skeletal muscle and T-lymphocytes[215,222]. The 18-21 months is a tipping point, after which entropy slightly decreases in the oldest age group (24-27 months). This reveals that the liver tissue undergoes network disorganization with ageing, increasing the disorderness or randomness. We also performed local differential entropy analysis between young and old age groups to identify nodes showing a significant increase in randomness. We identified 684 nodes with significantly differing single node entropies (Wilcoxon Rank sum test q-value < 0.05, absolute difference in median > 0.03) between young (3m-6m age) and oldest age (24m-27m) groups. The pathway enrichment of these genes revealed that complement and

coagulation cascades, cytokine-cytokine receptor interaction, xenobiotics metabolism, steroid hormone biosynthesis, NFKβ signalling pathway, PI3-AKT signalling pathway, and MAPK signalling are significantly affected **(Table 3.2)**. The entropy-based approach captures relevant pathways associated with ageing.

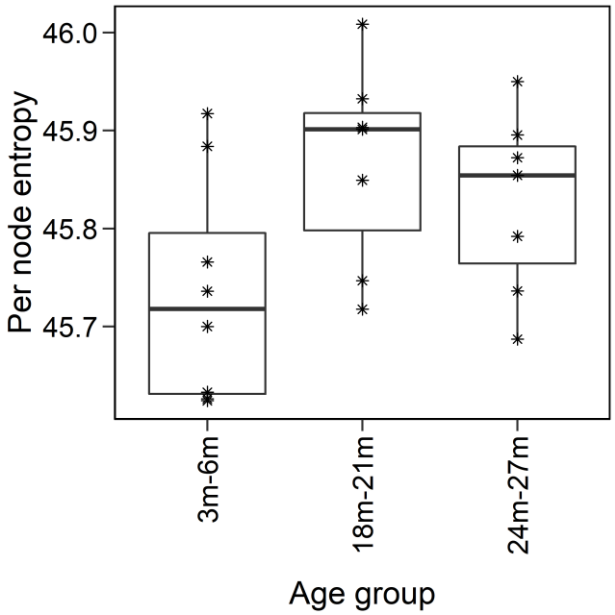


Figure 3.3: Network entropy-based analysis of liver ageing network. Boxplot showing the change in network entropy across different age groups. Sample-wise entropy is calculated and is normalized by number of nodes in its corresponding network.

Table 3.2: KEGG Pathway enrichment using Enrichr with genes showing significant change in single node entropy b/w young (3-6 months) and old (24-27 months) mice with FDR < 0.05 and absolute difference in the median between 2 groups > 0.03.

S. no	KEGG Pathway	Overlap	p-value	Adjusted p-value
1	Complement and coagulation cascades	18/88	8.46E-10	2.42E-07
2	Cytokine-cytokine receptor interaction	30/292	8.91E-08	1.27E-05
3	Primary immunodeficiency	10/36	2.33E-07	2.15E-05
4	Metabolism of xenobiotics by cytochrome P450	13/66	3.01E-07	2.15E-05
5	Chemical carcinogenesis	15/94	6.66E-07	3.81E-05
6	Steroid hormone biosynthesis	14/89	1.87E-06	8.88E-05
7	PI3K-Akt signalling pathway	31/357	2.17E-06	8.88E-05
8	Pathways in cancer	39/535	8.15E-06	0.000291
9	Pentose and glucuronate interconversions	8/34	1.48E-05	0.000471
10	MAPK signalling pathway	25/294	2.93E-05	0.000796
11	Cholinergic synapse	14/113	3.16E-05	0.000796
12	Drug metabolism	14/114	3.49E-05	0.000796
13	Th1 and Th2 cell differentiation	12/87	3.94E-05	0.000796
14	T cell receptor signalling pathway	13/101	4.01E-05	0.000796
15	Fatty acid degradation	9/50	4.33E-05	0.000796

The top-ranking nodes based on increase in entropy belong to the cytochrome P450 superfamily (CYP7B1, CYP2D9, CYP2F2, CYP2C29**)** and UDP-glucuronosyltransferases (UGT2B5, UGT2B36 and UGT2B1), which are linked to drug metabolism and steroid hormone synthesis **(Figure 3.4)**. The entropy increase is observed with FGG, FGB and VTN, which are associated with ECM and wound healing. VTN encodes for a secreted protein vitronectin that inhibits the membrane-damaging effect of the terminal cytolytic complement pathway (endothelial cells)[223]. TDO2 shows an increase in entropy and is linked to changes in tryptophan and kynurenine (Kyn). Tryptophan metabolism controls the inflammation-associated decline in age-related tissue homeostasis (inflammaging)[224].

Fatty acid oxidation genes ACSL1, ACADVL, ETFDH, ACOX2, HADHA, HSD17B4 and fatty acid transport gene SLC27A2 show an increase in entropy. The involvement of mitochondrial and peroxisome genes linked to fatty acid oxidation suggests an interplay between peroxisome-mitochondria in liver ageing[225]. CREB3L3, which cooperates with PPARA to regulate the expression of genes involved in fatty acid metabolism, also shows an increase in entropy **(Figure 3.4).** On the other hand, the entropy of lipid synthesis genes FASN, SREBF1, FADS1 and AACS and lipid transport gene LDLR decrease with ageing. Interestingly, the entropy of PGRMC1 and INSIG2 that regulate hepatic *de novo* lipogenesis via SREBF1 increases. Similarly, PLIN2, a gene associated with the metabolism of intracellular lipid droplets (LDs), also shows an increase with ageing.

Further, genes of glutathione metabolism show a change in entropy with ageing. Glutathione S-transferase (GSTs) GSTP1 shows an increase in entropy, while GSTM1 shows a decrease in entropy. GSTs are the Phase-II enzymes that protect the cells against damage induced by electrophiles and products of oxidative stress. They are shown to have anti-ageing effect[226]. GPTX1, which catalyzes the reduction of hydrogen peroxide (H_2O_2) by GSH, also shows an increase in entropy along with GCLC, an essential gene for GSH synthesis. RARRES2, which encodes a chemoattractant protein (Chemerin) secreted by the liver, shows a decrease in entropy with ageing. Chemerin is a modulator of immune response by promoting the chemotaxis of numerous immune cell types and it has a role in pathophysiological conditions, including HCC and NAFLD[227].

The overlap of entropy-based genes with DEGs between 3- and 27-months old mice shows only a few overlaps, indicating that genes identified based on statistical properties of the

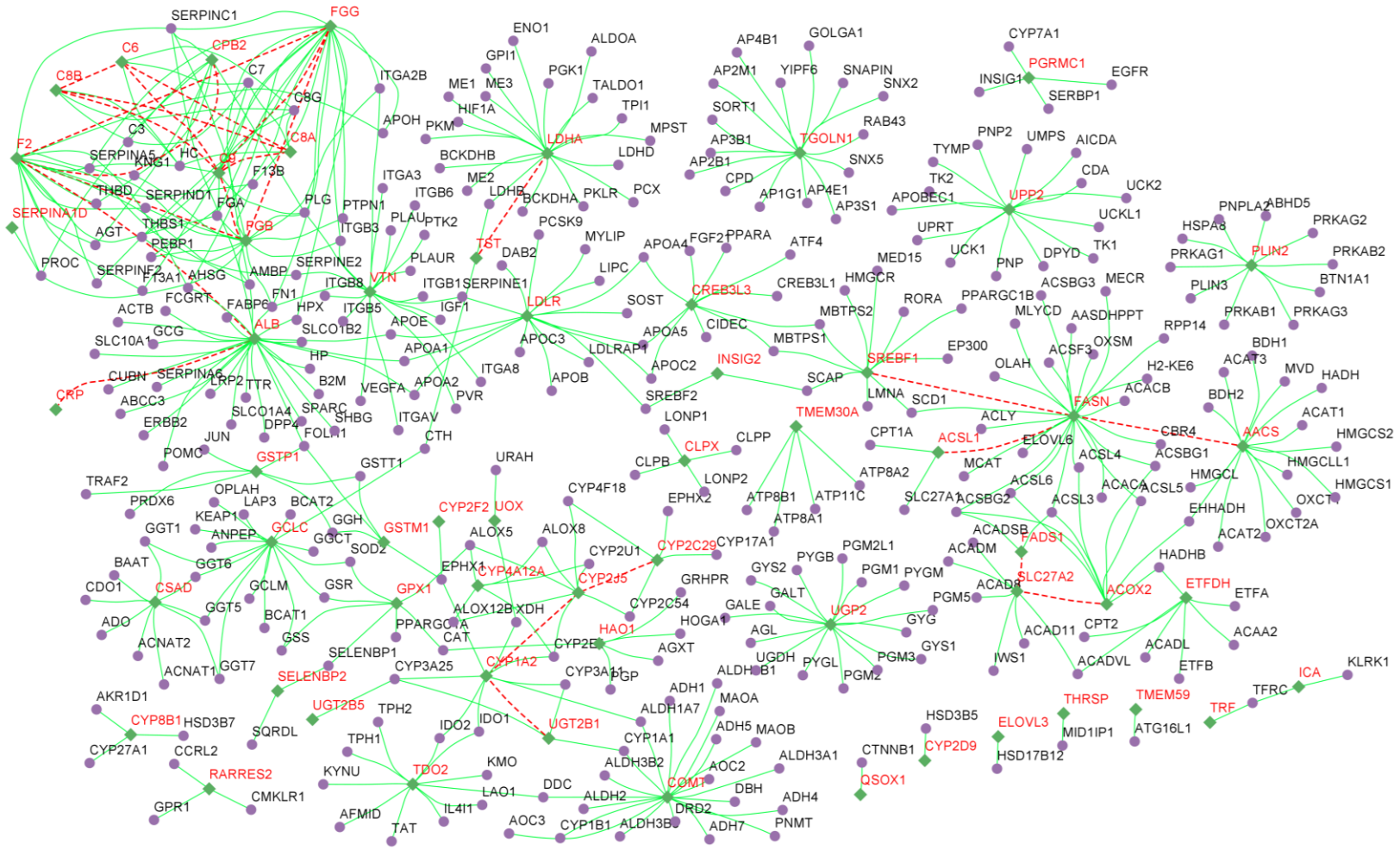


Figure 3.4: Network entropy-based analysis of the liver ageing network. Network of top 50 nodes with significant change in local entropy and their neighbours. Top nodes are shown as (♦) in red font, and the edges connecting them are shown with red dashed lines. Edges between neighbours are not shown.

underlying network are unique **(Figure 3.5A)**. We also compared the entropy-based candidate genes with the curated mouse immune genes[228] **(Figure 3.5B)**. The entropy-based analysis also identifies distinct immune-ageing genes compared to DEG analysis with a small overlap.

This suggests that ageing is characterized by global changes in the immune system. Non-overlapping 454 genes also include genes related to neurodegeneration (DNAHs) and protein digestion and absorption (Collagens). Immune markers unique to entropy-based analysis include genes VTN, FGB and FGG.

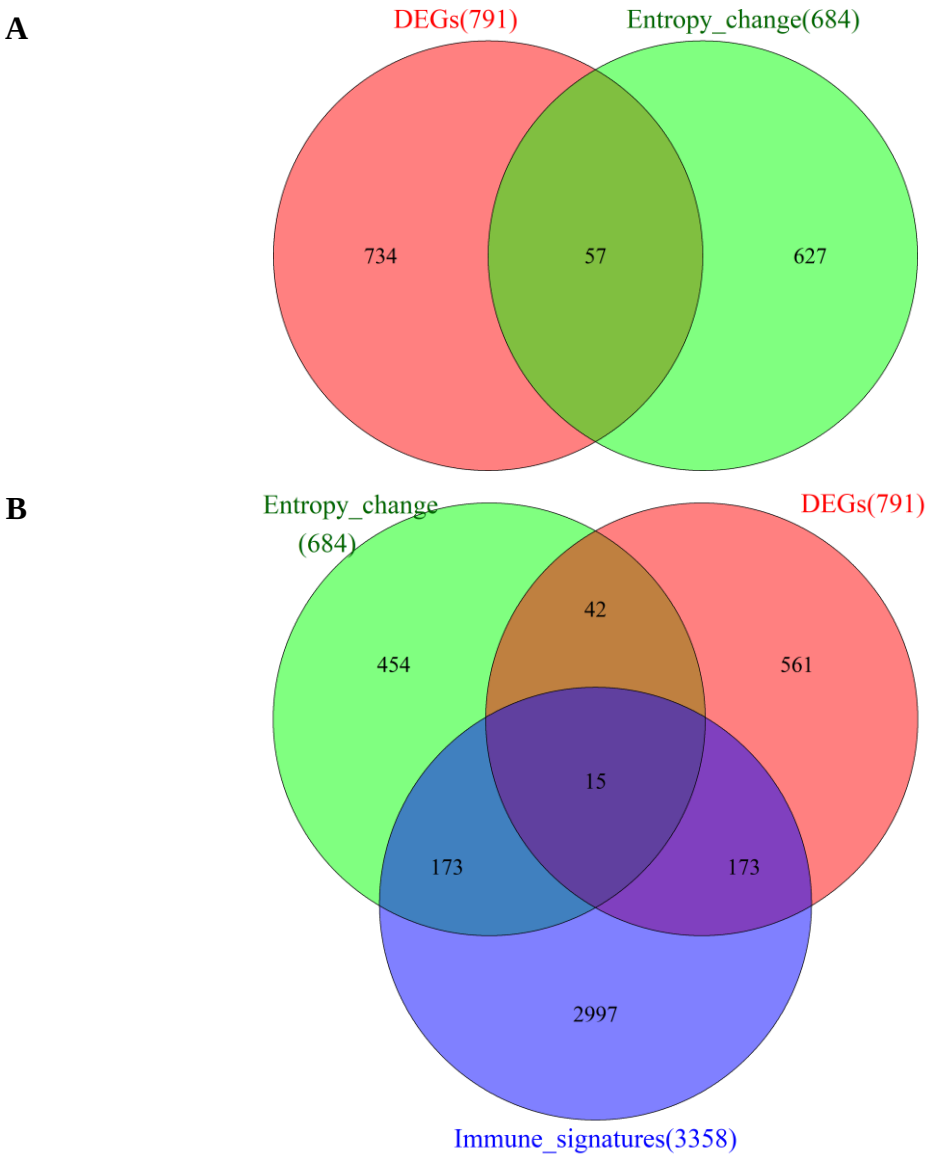


Figure 3.5: Venn diagrams showing the overlap of entropy-based candidates, ageing DEGs and immune signatures.

3.3.2 Core gene expression modules associated with ageing

We also alternatively explored the ageing gene expression changes at the network level by integrating DEGs and the PPI network. We expanded the network to include the first neighbours of DEGs. The PPI network built from DEGs comparing extreme age groups (3 and 27 months) and their first neighbours resulted in 38764 edges connecting 3770 nodes. Similarly, we also constructed an ageing network for other age groups (18, 21 and 24 months) for comparative analysis. We first clustered genes based on network topology to identify densely connected regions using MCODE.

The modular analysis of the liver ageing network (3 and 27 months) shows that genes corresponding to pathways such as ribosome, proteasome and oxidative phosphorylation are associated with top-scoring clusters **(Figure 3.6)**. These pathways are also found in the 18- and 24-months old mice ageing network **(Table B1 and Table B2)**. Signalling pathways (mainly Wnts) regulating the pluripotency of stem cells emerged as a significant pathway in the oldest 27-month age group. The clusters from 18 to 24-month networks are also associated with cell cycle, DNA repair, p53 signalling pathway and senescence. The enrichment of top clusters shows the relationship to NAFLD, basal cell carcinoma, neurodegenerative diseases, and viral infection.

We also identified critical nodes based on local and global network measures. Topological analysis of the ageing network based on local (MCC) and global (Bottleneck and Betweenness) metrics shows that RPS27A and TRP53 are critical nodes in the network **(Table 3.3)**. Other nodes of global importance in the network include AKT1, SRC, CTNNB1, and EGFR, while genes associated with proteosome (PSMB2, PSMA6, PSMB4, PSMA1, PSMB1, PSMA3, PSMD12, PSMC1, PSMD3, PSMA4, PSMD4) are locally important. It is also observed that RPS27A and TRP53 are not only the nodes of global and local importance nodes in the extreme age group but also form the early signs of ageing **(Table 3.3)**.

Table 3.3: Hub nodes based on cytoHubba metrics in the PPI network of DEGs comparing different age groups and their first neighbours.

Network measure	18 vs 3 months	24 vs 3 months	27 vs 3 months
MCC	NDUFB7, NDUFB9, NDUFAB1, NDUFB8, NDUFA5, NDUFA6, NDUFV2, NDUFB10, NDUFA12, NDUFB5, NDUFA8, NDUFS8, NDUFS7, NDUFA9, NDUFA10, NDUFV1, NDUFS1, NDUFS3, UQCRFS1, NDUFS2, UQCRC1, UQCRC2, COPS3, COPS4, COPS2, COPS5	PSMD1, PSMC3, PSMC6, PSMD11, PSMC5, PSMD12, PSMC1, PSMD3, PSMB7, PSMA5, PSMB5, PSMA2, PSMD6, PSMA1, PSMB3, PSMB2, PSMB10, PSMA3, PSMA6, PSMA4, PSMA7, PSMB4 PSMB6, PSMB1, PSMD4, PSMB8, PSMA8, PSMB9, CDC6 RELA, CCND1, UBA52, UBC, UBB, RPS27A, TRP53, CCNB1, CDK1	PSMB2, PSMA6, PSMB4, PSMA1, PSMA3, PSMD12, PSMC1, PSMD3, PSMA4, PSMD4, PTEN, RELA, UBB, UBC, UBA52, RPS27A, CDK1, TRP53
Bottleneck	AKT1, SRC, EGFR, CTNNB1, TRP53, RAC1, JUN	TRP53, ESR1, AKT1, CTNNB1	PTEN, UBA52, RPS27A, TRP53, AKT1, SRC, CTNNB1, ESR1
Betweenness	AKT1, SRC, EGFR, CTNNB1, TRP53, ESR1, RAC1	RPS27A, TRP53, ESR1, AKT1, CTNNB1, KRAS, SRC, RHOA	RPS27A, TRP53, AKT1, SRC, CTNNB1, EGFR, ESR1

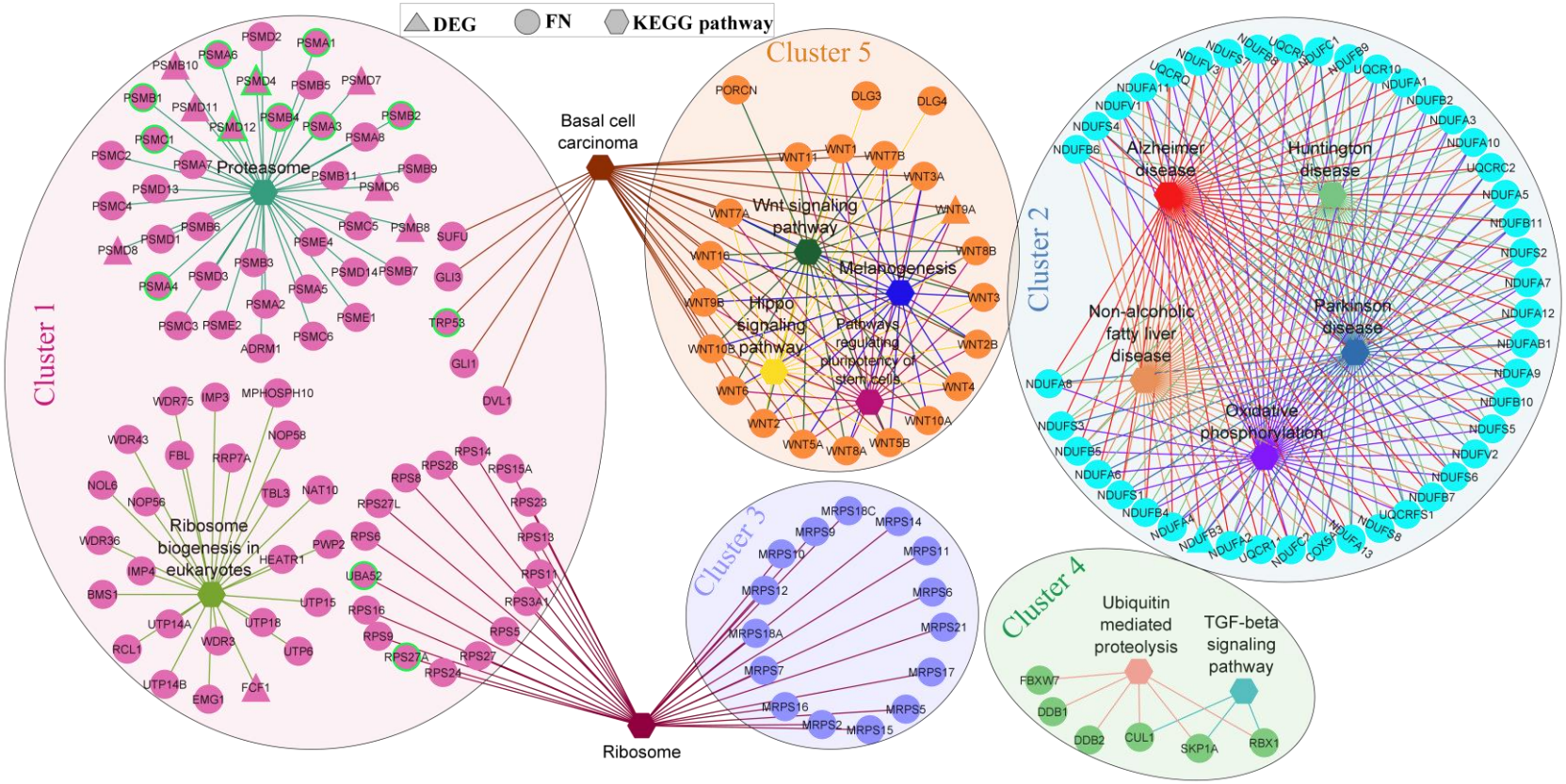


Figure 3.6: Top five MCODE clusters of the ageing network obtained using the DEGs between 3 and 27 month old mice and their first neighbours. Pathway enrichment of clusters is shown along with the gene information. Hub nodes/genes are highlighted with a green border.

3.3.3 Relationship between ageing and pathways associated with liver diseases and regeneration

Ageing can increase the susceptibility to liver diseases like HCC and NAFLD and affect the ability of the liver to regenerate after damage. We hypothesized that this might arise due to shared or related pathways associated with liver diseases and regeneration. We performed network proximity analysis using condition-specific DEGs to study the relationship between ageing and perturbations that influence liver function. The network distance was quantified using the mouse PPI network. We found significant proximities between ageing and liver-related pathologies such as NAFLD, HCC and acute and chronic damage by CCl_4 by integrating DEGs and mouse PPI network **(Figure 3.7)**. The proximity distance decreases with an increase in the age of mice.

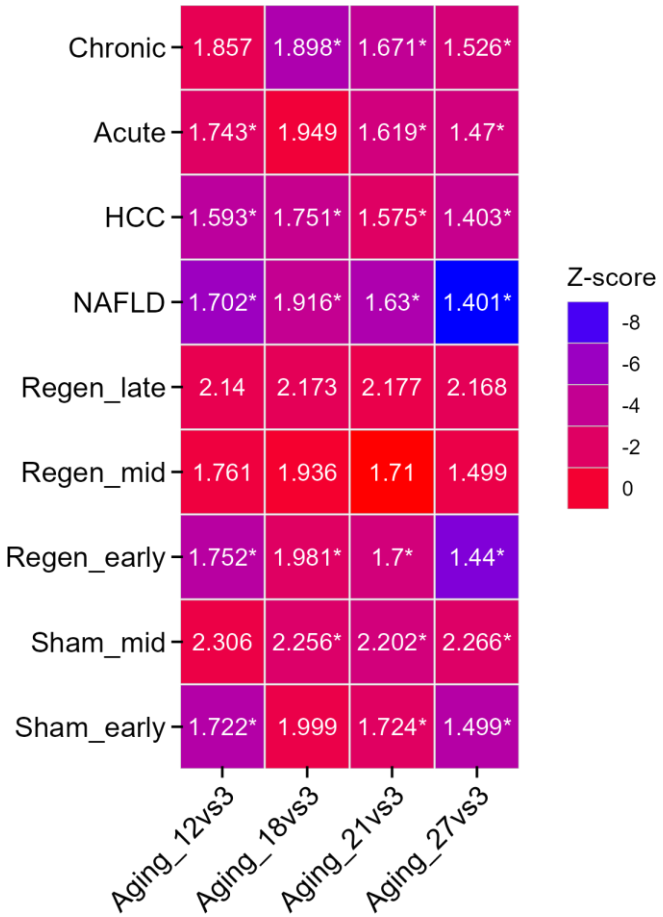


Figure 3.7: Network proximity of ageing with different liver-associated conditions: early (Regen_early), mid (Regen_mid) and late (Regen_late) phases of liver regeneration, early and mid-phases of sham-operated control, NAFLD, HCC, acute and chronic liver damages. The proximity is explored for different age groups (12, 18, 21 and 27 months). Text in the tiles represents proximity distance. *Indicates FDR < 0.05 and Z-score < − 1.5.

Proximity analysis between ageing and the early phase of liver regeneration (1, 4, and 10 hours post-PH) shows that older age groups are significantly proximal to the liver regeneration module. This proximity may influence the liver regeneration process in ageing. Ageing is shown to delay liver regeneration post-PH. However, the mid and late phases of liver regeneration associated with cell cycle and termination phases, respectively, are not significantly proximal to the ageing module **(Figure 3.7)**. Therefore, proximity analysis captures and quantifies the impact imposed by ageing on regeneration at the network level. The early phase of sham-operated control is also proximal to the older age compared to the young age. This is consistent with the observation that the early phase of sham-operated control and liver regeneration is similar[139]. Further, the proximity of the mid-phase of sham-operated control to the ageing network increases compared to the early phase.

To probe the qualitative picture of proximity analysis, we identified nodes falling in the shortest path between every ageing gene and all candidate genes of other conditions **(Figure 3.2)**. This resulted in 2101, 2112, 1791, 2075 and 2322 nodes in the pairwise comparisons: ageing and regeneration, ageing and NAFLD, ageing and HCC, ageing and acute damage, ageing and chronic damage, respectively. Nodes from each comparison were collectively projected onto the PPI network **(Figure 3.8A)**. The connectivity pattern suggests that ageing is connected to different conditions through intermediate nodes between condition-specific DEGs. We observed a common theme of 926 proximal molecular players connecting ageing with different liver conditions emerges **(Figure 3.8B)**. This converges on important KEGG pathways such as pathways in cancer, proteoglycans in cancer, Epstein-Barr virus infection, PI3K-Akt signalling pathway and MAPK signalling pathway **(Figure 3.9)**. GRB2, SOS, RAS, RAF and ERK1/2, are the important molecular players present in the top pathways associated with the common theme. GSK3B is another interesting candidate gene common across ageing, NAFLD and HCC **(Figure 3.10)**. It is upregulated in NAFLD and downregulated in HCC. GSK3B connects different conditions via CTNNB1. TRP53 signalling pathway also connects ageing to liver-associated conditions. This may control cell cycle entry by regulating genes such as CCND1, CDKN1A and GADD45A **(Figure 3.11)**. GSK3B and TRP53 interaction is also a part of the common theme. The overlap of 926 genes with curated mouse-specific immune genes shows that 366 genes are common **(Figure 3.12),** with NFKβ as a key transcriptional factor. NFKβ regulates innate and adaptive immunity and is the master regulator of inflammatory responses[229]. We also identified NLRP12 as a common candidate gene that plays the role of a mitigator of

inflammation. It is upregulated in the early phase of liver regeneration while downregulated in ageing, NAFLD, and acute and chronic liver damage.

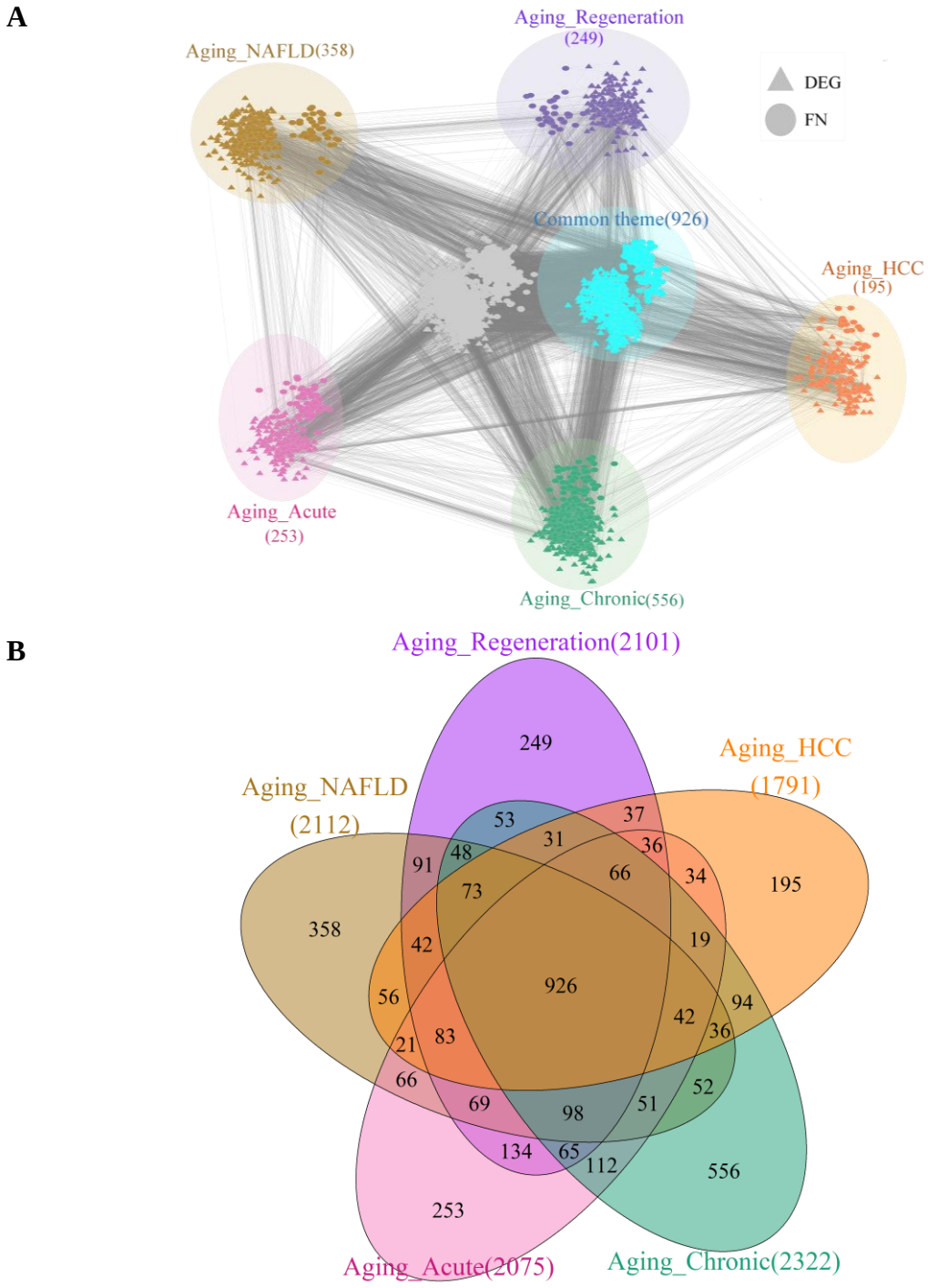


Figure 3.8: Overlap of proximity nodes obtained in the pairwise comparison of ageing and different liver-associated conditions. (A) Crosstalk (interactions) between nodes of different liver-associated conditions are shown using the PPI network. The common theme comprises nodes that are present in all comparisons. Nodes that are neither part of the common theme nor specific to a condition are shown in grey. A node that is a DEG in at least one condition is shown by a triangle, and the first neighbour (FN) of DEG is shown by a circle. (B) Venn diagram showing the number of nodes overlapping between different pairwise comparisons.

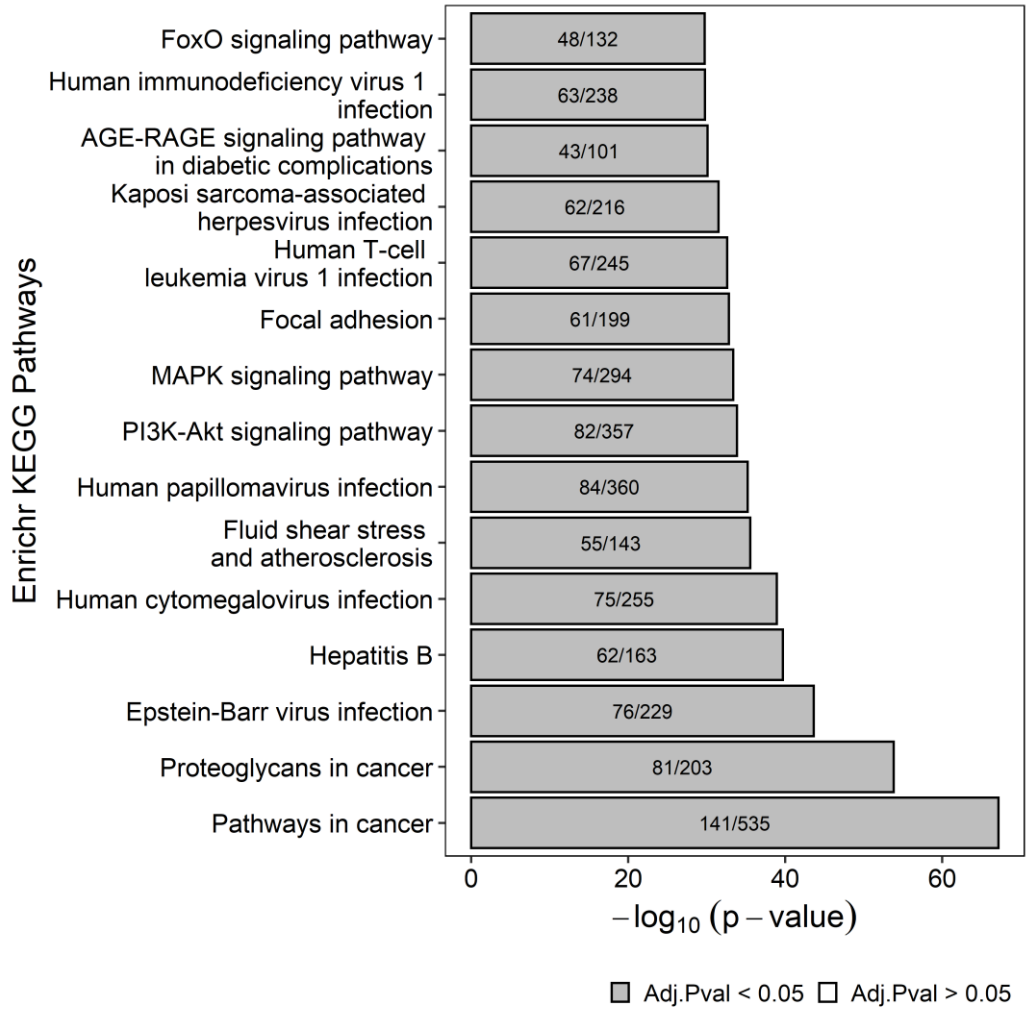


Figure 3.9: The KEGG pathway enrichment of 926 genes in the common theme of proximity nodes obtained in the pairwise comparison of ageing and different liver-associated conditions.

In addition to the immune system, we also explored other common relationships between ageing and different liver conditions. Lipid (FASN, HMGCR, SREBF1) and bile acid synthesis (CYP27A1) genes are differentially expressed in ageing and liver regeneration. FASN, HMGCR and SREBF1 are upregulated in ageing. Mitochondrial fatty acid β-oxidation (FADS1, HADHA, HADHB, ACSL1, ACADVL, CPT2, ECI2) also shows this differential pattern. CREB3L3 is differentially expressed across conditions. It is upregulated in liver regeneration and downregulated in ageing. PCSK9, which plays a role in cholesterol homeostasis, is downregulated in the early phase of regeneration and upregulated in ageing and NAFLD. It protects the liver against steatosis and liver injury. On the other hand, ANGPTL4, which facilitates the accumulation of TAG by inhibiting LPL, is downregulated in ageing while it is upregulated in liver regeneration.

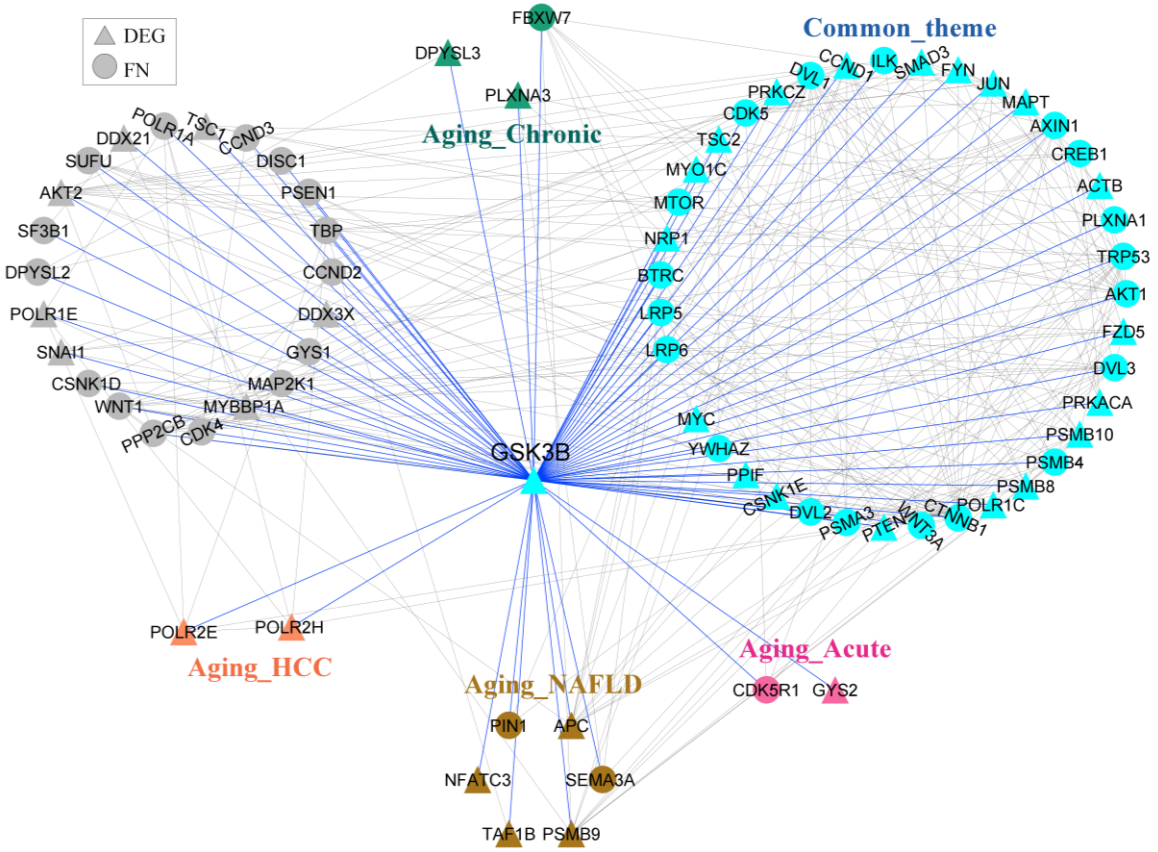


Figure 3.10: Interacting partners of candidate gene, GSK3β, present in the common theme

Ageing also influences amino acid metabolism. Genes of one-carbon metabolism (DHFR, MTHFD1, BHMT, SHMT1/2, MAT1A, MTR, TYMS) are affected across conditions. S-adenosyl-methionine (SAM) metabolism controlled by MAT1 is significantly upregulated in liver regeneration compared to ageing and is downregulated in HCC. MAT1 regulates the production of SAM from methionine, which is required for methylation reactions inside the cell. In NAD metabolism, NAMPT is upregulated in ageing, while NNMT is upregulated in regeneration and downregulated in NAFLD and HCC. Genes involved in BCAA catabolism, glutamine catabolism (GLS2), aspartate synthesis (ASS1), and Tryptophan metabolism (TDO2) are also affected across different liver conditions.

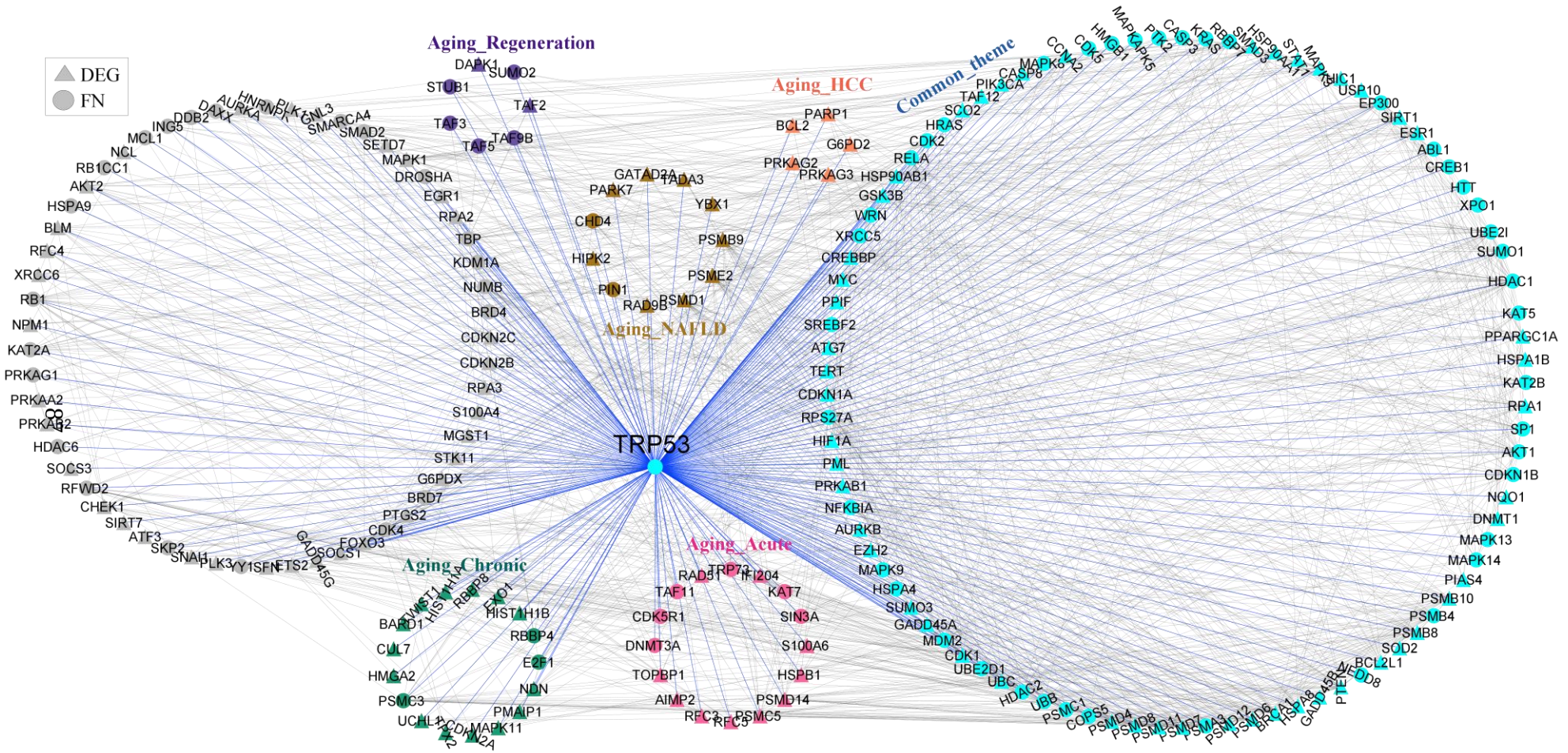


Figure 3.11: Interacting partners of the candidate gene, TRP53, present in the common theme.

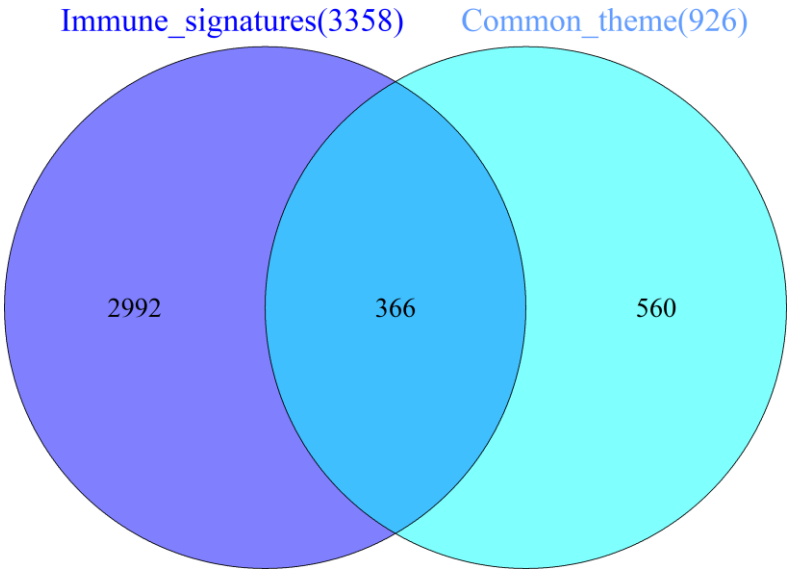


Figure 3.12: Venn diagram showing the immune signatures in common theme shared by different pathological conditions

3.4 Discussion

Ageing can lead to functional impairment of liver and predisposes the liver to NAFLD and HCC. The liver has a unique ability to regenerate itself post-injury and help in whole-body metabolic homeostasis and compound detoxification. Mapping the molecular changes of the liver with ageing may help to understand how ageing influences liver function and predisposes the liver to different pathological conditions. A systems-level analysis of the ageing-induced liver changes and its crosstalk with the pathology of liver diseases is lacking. In the present study, we performed a network-level analysis of liver ageing using transcriptomic data of ageing and the PPI network. We used network entropy measure to identify nodes and pathways that show significant entropy changes in ageing. Further, we also performed the topological analysis of the ageing network by considering the nodes differentially expressed in ageing and their first neighbours to identify the core modules of ageing. This framework was also used to study the proximity of the ageing network with liver regeneration and disease networks. We showed proximity measure provides insights into the interconnection between ageing and liver-associated conditions.

We observed an increase in entropy with ageing liver with the subtle difference between old and oldest groups **(Figure 3.3)**. The entropy-based approach captured the relevant pathway-level changes linked to ageing and helped identify novel candidate genes. The

entropy change is driven by the selected group of genes belonging to the immune, complement and coagulation cascade, lipid metabolism, cytochrome P450 and UDP-glucuronosyltransferases. Immune and lipid metabolism-related changes have been reported in ageing[135,209]. The candidate genes were filtered based on entropy changes. We provide experimental evidence available from the literature for the involvement of candidate genes in ageing or its related liver diseases. Top novel candidate genes with high entropy values include VTN, FGB and FGG, which are associated with changes observed in fibrosis under chronic liver damage condition[230]. We hypothesize that transcriptional remodelling of the liver during ageing can affect the integrity of the membrane and increase the susceptibility to fibrosis. Ageing is shown to increase the susceptibility to fibrosis in response to high-fat diet feeding[231]. We also found genes of the complement system (C6, C8, C8A, C8B) that are part of the membrane attack complex changing with ageing (**Figure 3.4**). There is increasing evidence that complement systems may play a role in ageing[232].

Our analysis also revealed the PGRMC1-INSIGs-SREBF1 axis in controlling the lipid levels in ageing **(Figure 3.4)**. PGRMC1 knockout leads to the buildup of fatty acids and predisposes mice to NAFLD[233]. PGRMC1 forms a complex with INSIG1 and is associated with the cleavage of SREBF1 via SCAP[233,234]. Deletion of INSIG2 also results in the activation of SREBF1 and *de novo* lipid synthesis[235]. Age-induced hepatic steatosis is alleviated in INSIG2 elevated condition[236]. Another candidate gene, PLIN2, also controls the activation of SREBP-1 and SREBP-2[237]. Its expression is shown to be altered in age-related diseases, including fatty liver[238,239]. Fat accumulation is negatively correlated with the decrease in mitochondrial mass with ageing[240]. Further, ageing is shown to affect lipid homeostasis by controlling the phosphorylation of CEBPα/β[241] and changing the nucleosome occupancy at the foci of PPAR targets[242]. We found that PPARA can also be affected through CREB3L3, the knockout of which results in severe fatty liver[243]. CEBPβ is implicated in the activation of SREBF1 transcription in liver[244]. RARRES2 (Chemerin) is another candidate ageing gene, which is also induced in NAFLD and Hepatitis B-related HCC. These observations suggest that ageing may increase susceptibility to liver diseases.

The network topology-based analysis of the ageing network revealed the involvement of ribosomes and proteasomes, which reflects the changes in the proteostasis capacity of cells with ageing[245]. The module associated with oxidative phosphorylation in the ageing network **(Figure 3.6)** reflects the change in mitochondrial metabolism with ageing[200]. We

found Wnt pathway as an ageing module, which controls cell renewal, tissue regeneration and the development of HCC[246]. Further, TRP53 was identified as a critical node based on local and global graph theoretical measures. It has relevance in ageing as it can promote repair, survival, or elimination of damaged cells[247]. TRP53 optimally balances tumour suppression and longevity[248]. The decline in the function of TRP53 is observed in various tissues of the mouse with ageing, which can contribute to increased mutation frequency and tumorigenesis[249]. Similarly, RPS27A was also found to be a critical node in the ageing network, which clustered with other ribosome machinery genes in MCODE cluster 1 **(Figure 3.6)** and is directly connected to TRP53. RPS27A is involved in the activation of p53[250,251]. The ribosome machinery genes RPS5 and RPS11 directly connected to RPS27A in the ageing network are shown to be repressed with ageing in mouse liver[252]. Other critical nodes include AKT1, SRC, CTNNB1 and EGFR, which are related to cancer signalling. CTNNB1 encodes a β-catenin protein responsible for controlling gene expression in the Wnt signalling pathway. EGFR also shows an increase in entropy, and its expression is correlated with liver steatosis in mice and humans[253].

The PPI network analysis of ageing and different liver conditions also shows the proximity of ageing genes to different liver conditions, including NAFLD, HCC, liver damage and repair **(Figure 3.7)**. The common theme shared between conditions maps to immune-related pathways, pathways in cancer and metabolic changes. MAPK, PI3K-AKT, Ras, Wnt and NFκB signalling are common pathways across conditions **(Figure 3.9)**. Studies on extended lifespan by pharmacological intervention suggested that anti-ageing effects are mediated by targeting the canonical MAPK pathway[254]. With ageing, there is an upregulation of MEK1, which triggers translation by phosphorylating its downstream target eIF4E[254]. Increased activity of eIF4E has been shown to promote tumorigenesis, thus implicating ageing effects on cancer[255]. GSK3β is a common node across conditions. Ageing is shown to inhibit GSK3β function[256] , and this, in turn, affects the liver regeneration potential[257]. Inhibition of GSK3β acts as a protective role against lipid accumulation in NAFLD. GSK3β can regulate cell proliferation by controlling the growth-inhibitory activity of CEBPα and negatively regulates many oncogenic signalling pathways, such as the Wnt/β-catenin pathway[258]. We found GSK3B-CTNBB1 interaction as part of the common theme **(Figure 3.10)**, which is linked to HCC development and NAFLD. There is also a mechanistic link between inflammation and the development of HCC mediated by NFKβ signalling[259]. NASH condition exhibits morphological conditions related to infiltration of lymphocytes

and neutrophils, hepatocyte death and activation of liver resident macrophages Kupffer cells, creating an environment favourable for compensatory hepatocyte proliferation that further drives hepatocarcinogenesis[260]. Further, the priming phase of liver regeneration after PH depends on the activation of NFKβ[261].

We observed lipid metabolism as a common theme across ageing and liver-associated conditions. Induced alteration in lipid metabolic genes in ageing may increase susceptibility to NAFLD and affect liver regeneration. Both lipid overloading and deficiency can affect liver regeneration ability. Fine-tuning lipid levels by transport, biosynthesis, and oxidation is crucial for liver regeneration[262]. A high-fat diet impairs liver regeneration through IKKβ overexpression and subsequent NFKβ inhibition[263]. The aberrant activation of FASN plays a major role in the development of HCC, and its level is also shown to increase during the induction of senescence[264].

In amino acid metabolism, one-carbon metabolism is altered across conditions, and it plays a crucial role in maintaining tissue homeostasis and longevity[265,266]. It generates various metabolites that are building blocks of nucleotide synthesis, methylation, and redox reactions. Oncogenic signalling hijacks the one-carbon metabolism to support proliferation and survival[267]. Genetic disruption of MAT1 inhibits liver regeneration[127]. MAT1 expression is reduced in different liver pathologies, including NAFLD and HCC. Hepatic methionine is depleted in mice that developed NAFLD, and administration of methionine and choline-deficient diet led to alterations in the expression of lipid metabolism genes[268,269]. Metabolomics analysis of ageing shows the levels of serine and methionine decrease in liver[270]. These highlight the importance of one-carbon metabolism in liver function and pathology. Further, BCAA is altered across conditions, and loss of BCAA catabolism promotes HCC development and progression[271]. However, this is not suppressed in liver regeneration[127]. BCAA metabolites are also altered in aged liver[200].

In summary, our study maps the network-level changes of ageing and dissects the crosstalk between different conditions, including regeneration and diseases in the mouse liver. We uncovered the local and global changes in immune response, cancer signalling and metabolism with ageing and identified novel candidate genes. We showed the proximity of the liver ageing network to liver-condition-specific networks and identified the interconnections through common pathways. This explains how ageing increases

susceptibility to different disease conditions and affects the capacity of the liver to regenerate.

As an initial study, we used the bulk sequencing data to generate a liver tissue-specific PPI network in different contexts for comparison. The bulk changes can be due to cell composition changes or alterations in the gene expression of each cell in the population. The single-cell data will further help to refine the interactions in a cell-type-specific manner. Nevertheless, our study provides the initial framework for single-cell network analysis of liver ageing and its related diseases. Our study is based on RNA-seq data of mouse liver available across different age groups and pathophysiological conditions from similar strains. Further, it will be useful to extend our analysis pipeline to human context to account for human specific regulation[272] and mouse-to-human translation.

Chapter 4 Gene expression signatures of stepwise progression of Hepatocellular Carcinoma

4.1 Introduction

HCC is the common form of primary liver tumour and the third-most leading cause of cancer-related deaths globally[273,274]. Major risk factors leading to HCC include viral infections (Hepatitis B - HBV and Hepatitis C - HCV), excessive alcohol and tobacco consumption, exposure to fungal toxins, and NAFLD, with 90% of cases arising from the underlying chronic liver disease[275]. While HBV-driven HCC is prevalent in East Asia and Africa, HCV infections are most common in the US and Europe. NAFLD is emerging as the leading risk factor of HCC, especially in the West, owing to the rise in obesity and diabetes[276]. Despite continuous advances and management strategies designed to mitigate the incidence of HCC, its mortality rates have been rising over the last two decades. The major caveat in reducing the incidence of HCC is the detection at an early stage since more than 50% of HCC cases are diagnosed at advanced stages[277]. Therefore, a better understanding of HCC pathogenesis and its molecular underpinnings will help reduce the rising cases.

Most HCC cases develop in the background of unresolved chronic inflammation[278], which triggers a persistent healing response[279]. The unbalanced healing response disturbs the architecture of the liver, leading to fibrosis, followed by cirrhosis. The regenerating nodules produced during cirrhosis fuel the transformation of hepatocytes to premalignant lesions called dysplastic nodules. These premalignant lesions develop into early HCC (eHCC) and progressive HCC (pHCC). Although this stepwise progression from chronic liver disease to tumour state is widely prevalent in HCC, about 20% of cases arise from a non-cirrhotic background[280]. While most non-cirrhotic HCCs develop from metabolic syndrome[281], HBV or HCV infection can also lead to HCC from accelerated fibrosis without cirrhosis[280,282]. Hence, it is crucial to consider the existence of multiple trajectories to HCC when developing diagnostic markers.

Due to the inherently complex nature of HCC development, managing patients is also quite challenging. Surgical resection is the primary treatment for HCC patients with preserved liver function but is prone to recurrence in about 70% of the cases within a few years[277]. Liver transplantation is another option for patients not eligible for resection but is limited

by the availability of donors[283]. Under the circumstances where resection or liver transplantation is not amenable, liver-directed medication fails, or recurrence is seen post-resection, systemic therapy is chosen[277]. Systemic therapy in the form of tyrosine kinase inhibitors and immunotherapy targeting immune checkpoints have been developed for treating advanced-stage HCC[284]. Further, treatment strategies must also consider underlying liver disease along with tumour stage[285,286], which may account for differences in the risk of HCC recurrence among patients[287].

The advancement in the high throughput techniques (next-generation sequencing) is helping to map the molecular changes of HCC at genomic, transcriptomic, and epigenetic levels[278,288]. This information provides insights into the various signalling pathways involved in hepatocarcinogenesis. These include differentiation and development pathways (Wnt/β-catenin, Notch Hedgehog signalling), genomic stability and cell cycle (TP53, RB1), telomerase (TERT), growth and cell proliferation (PI3K/AKT/mTOR, RAS/MAPK, EGF/EGFR), angiogenesis (VEGF/VEGFR, PDGF/PDFGR) and chromatin remodelling (ARID1A/ARID1B/ARID2 and MLL signalling)[278,289]. However, the current understanding of the interplay of various signalling pathways in HCC is far from complete.

Molecular profiling distinguishes diverse subgroups of HCC that are otherwise indistinguishable by conventional histological methods[290]. Gene expression changes in tumour samples are used to predict recurrence and stratify patients into high and low-risk groups[291]. In liver cancer, genes that show a fold change in expression between the normal and tumour samples are a better predictor of survival than considering candidates based on tumour samples alone[292]. Gene expression profile of tumour-adjacent normal tissue is also reported to predict HCC recurrence[293]. Prediction models proposed in these studies are based on differentially expressed genes in tumours or pre-defined gene signatures.

A recent study on a comprehensive analysis of tumour samples, tumour-adjacent normal samples, and normal healthy samples showed that tumour-adjacent normal samples represent an intermediate transcriptomic state between the other two[294]. Therefore, there is a need to explore the multi-step progression of HCC through different trajectories to gain further insights into the molecular mechanisms and develop predictive models. Network-based approaches provide a suitable platform to extract meaningful information from omics data, hypothesis generation, stratification of disease classes, and discovery of biomarkers[295]. In the present work, we aim to understand the molecular pathogenesis of

HCC sequentially from normal to tumour through different premalignant stages. A network-level analysis of the transcriptomic profile of tumour samples and tumour-adjacent normal samples in different liver damage conditions was performed to obtain insights into the transition from normal to precancerous to cancer state. The hierarchical changes: modules, pathways, and genes related to HCC progression and survival prediction were identified.

4.2 Methods

4.2.1 Dataset(s) description

Bulk RNA-seq transcriptomics data of HCC progression was obtained from GEO with accession number GSE148355. The dataset consists of tumour and non-tumour samples from HCC patients or patients with the chronic liver disease treated with either total hepatectomy (TH) or partial hepatectomy (PH) at Seoul National University Hospital. Clinical information is available for 54 tumour samples (35 are from TH patients and 19 from PH patients). The dataset also comprises 47 premalignant and 15 normal samples. The normal samples were from patients with metastatic cancer/polycystic liver disease/or cholangiocarcinoma. All non-tumour liver tissues have no evidence of hepatic fibrosis or viral hepatitis. Out of these 47 premalignant samples, 24 are tumour-adjacent normal samples. The premalignant stages include Fibrosis Low (FL) – 10 samples, Fibrosis High (FH) -10 samples, Cirrhosis (CS) – 10, Dysplastic nodule low grade (DL) – 10 samples, and Dysplastic nodule high grade (DH) – 7 samples. All samples were collected after receiving written informed consent from the patients, and the original study was approved by the Institutional Review Board of Seoul National University Hospital. This dataset is referred to as the Korean cohort. The plot summarising clinicopathological features of tumour samples is given in **Figure 4.1**.

In addition, we used HCC datasets from TCGA (TCGA-LIHC) and GEO (GSE14520) with available clinical information. TCGA gene expression data and clinical data were obtained from UCSC Xena (https://xena.ucsc.edu/). TCGA-LIHC comprises 316 tumour samples with clinical information, and 39 of them have paired normal samples. GSE14520 is a microarray-based (GPL3921 platform) gene expression profiling from HCC patients treated with surgical resection. The dataset includes gene expression data of 210 tumour and 210 adjacent normal samples with associated clinical information and is referred to as the Chinese cohort.

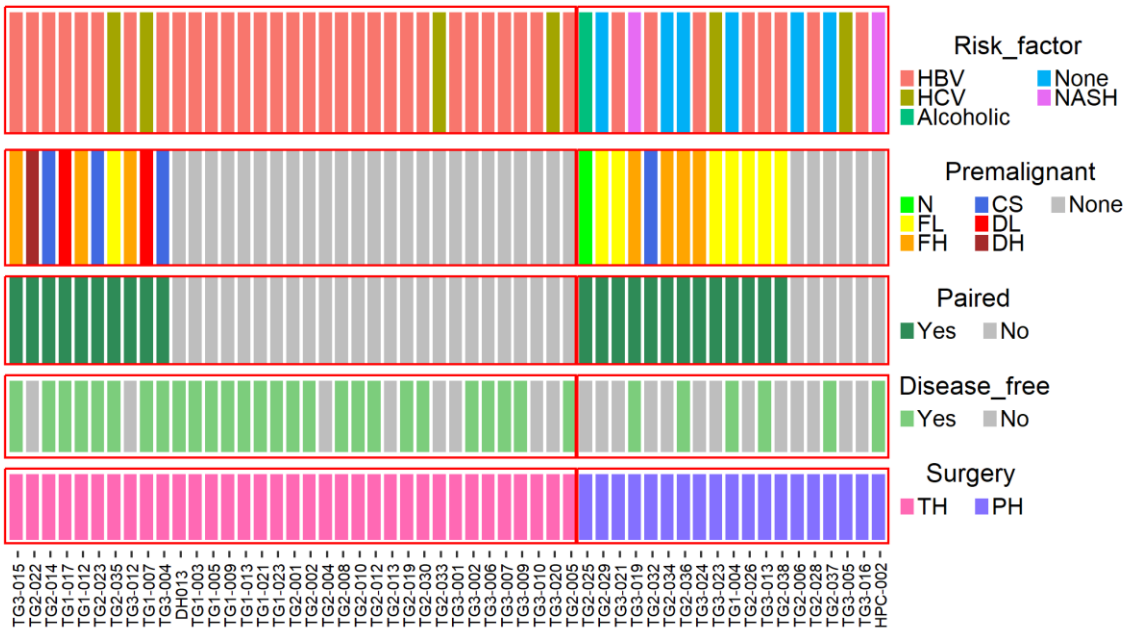


Figure 4.1: Clinicopathological features of tumour samples in the Korean cohort. This includes the surgery type (TH/PH), disease free (Yes/No), whether the tumour sample has an adjacent normal sample (Paired), the premalignant stage of the tumour-adjacent normal, and the risk factor (HBV, HCV, Alcoholic, None, NASH).

4.2.2 Workflow

A systems-level analysis was designed to study the pathogenesis of HCC at multiple levels: modules, pathways, and genes **(Figure 4.2)**. The analysis pipeline was applied to three groups of samples: (a) only tumour samples, (b) adjacent normal and tumour samples, and (c) all normal and premalignant samples. To identify modules, the co-expression network was constructed from gene expression data of the Korean cohort using weighted gene co-expression network analysis (WGCNA) in R[150,151]. FPKM values were transformed to $\log_2(\text{FPKM} + 1)$, and the top varying genes were selected using the rowVars function to construct the correlation (Pearson) matrix s_{ij} for WGCNA. A signed network was built by transforming the correlation matrix to an adjacency matrix (a_{ij}) using the power adjacency function and soft thresholding ($a_{ij} = s^{\beta}_{ij}$). Scale-free topology criteria was used to choose the power β. Subsequently, a Topological Overlap Matrix (TOM) was computed from the adjacency matrix, followed by dendrogram construction using 1 – TOM. Modules were identified from the dendrogram using the dynamic cut tree algorithm, and module eigengene expression (ME) for each module was calculated using singular value decomposition (SVD).

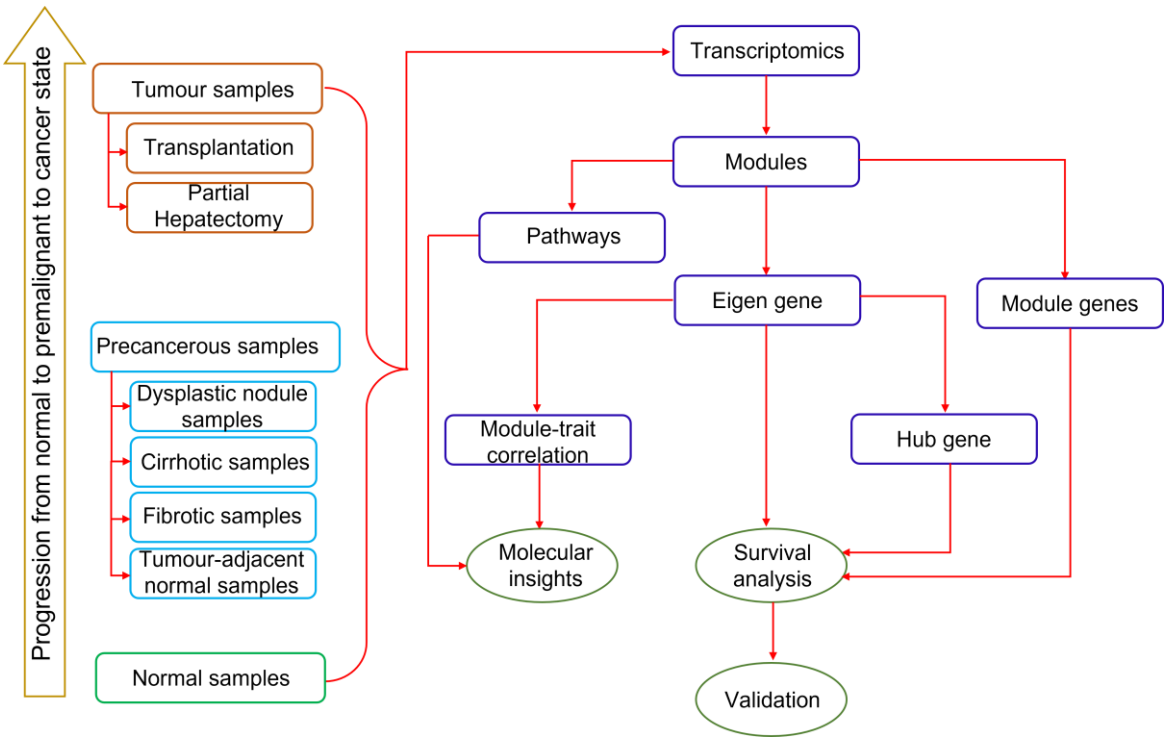


Figure 4.2: The workflow to study progression from normal to precancerous to cancer state in HCC. Gene co-expression analysis was performed on (a) only tumour samples, (b) adjacent normal and tumour samples, and (c) all normal and premalignant samples to identify modules, genes and pathways associated with HCC pathogenesis and survival.

Modules significantly correlating with disease-free survival (DFS) and other clinical traits were identified. Categorical traits such as surgery/treatment (TH and PH) and premalignant state were converted into continuous numerical values to compute correlation with different modules. For surgery, PH and TH were binarized as 1 and 2, respectively. The premalignant states were converted to numerical with 1, 2, 3, 4, 5, and 6 indicating normal, FL, FH, CS, DL, and DH, respectively. Hub genes from modules were extracted based on module membership (MM > 0.8) and intramodular connectivity.

Candidate genes were selected for univariate survival analysis based on the modules that correlated with DFS under each condition. Samples were dichotomised into two groups based on the median gene expression profile of candidate genes, and survival analysis was performed using the survival R package[296]. Further, module preservation analysis[297] was carried out using TCGA data as the test set to access the biological relevance of modules identified from the Korean cohort. The $Z_{summary}$ statistics was used to evaluate whether the module is preserved between the reference set (Korean cohort) and test set (TCGA) as:

$$module\ preservation(Z_{summary}) = \begin{cases} moderate, & 2 < Z_{summary} < 10 \\ strong, & Z_{summary} \geq 10 \end{cases} \quad (4.1)$$

4.2.3 Pathway enrichment analysis

The enrichment analysis of module genes was performed using Enrichr[152] to identify dysregulated pathways. The ClueGO Cytoscape plugin was used with default settings to visualise the interrelations of the GO biological terms associated with modules[298].

4.3 Results

4.3.1 Co-expressed modules of tumour samples

The co-expression pattern of genes within tumour samples (Korean cohort) was studied using top-varying genes. We found five modules (T1, T2, T3, T8, and T9) that significantly correlated with DFS **(Figure 4.3)**. Coincidentally, T1 and T9 modules are significantly correlated with the surgery/treatment (i.e., PH or TH) as well. An increase in DFS is associated with TH as the treatment. This is in agreement with the original study[299], which shows TH group has better DFS compared to patients undergoing PH treatment, although there were no differences in grade/stage of tumour in these two groups.

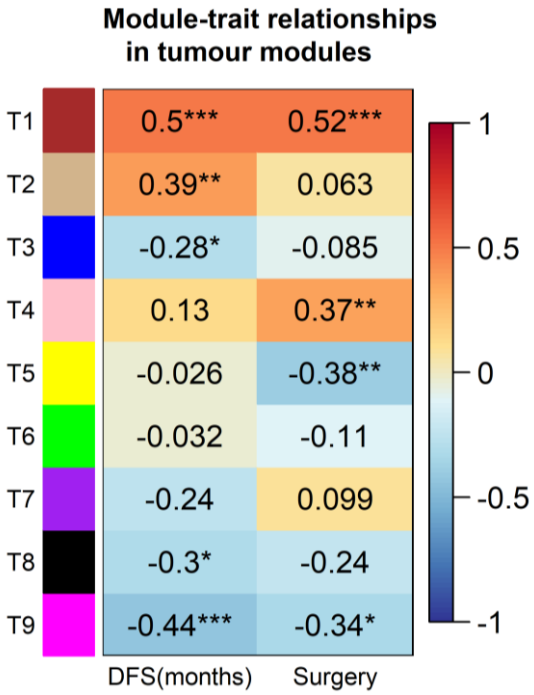


Figure 4.3: Module-trait correlations of tumour samples. DFS representing disease-free survival is a continuous variable. Surgery (treatment) is a binary variable with partial hepatectomy (PH) represented as 1 and total hepatectomy (TH) as 2. * indicates p-value < 0.001, ** indicates 0.001 ≤ p-value < 0.01, * indicates 0.01 ≤ p-value < 0.05.**

KEGG pathway enrichment of the T9 module showed that cell cycle-related pathways play an important role in governing the survival of a patient post-treatment **(Figure 4.4)**. The

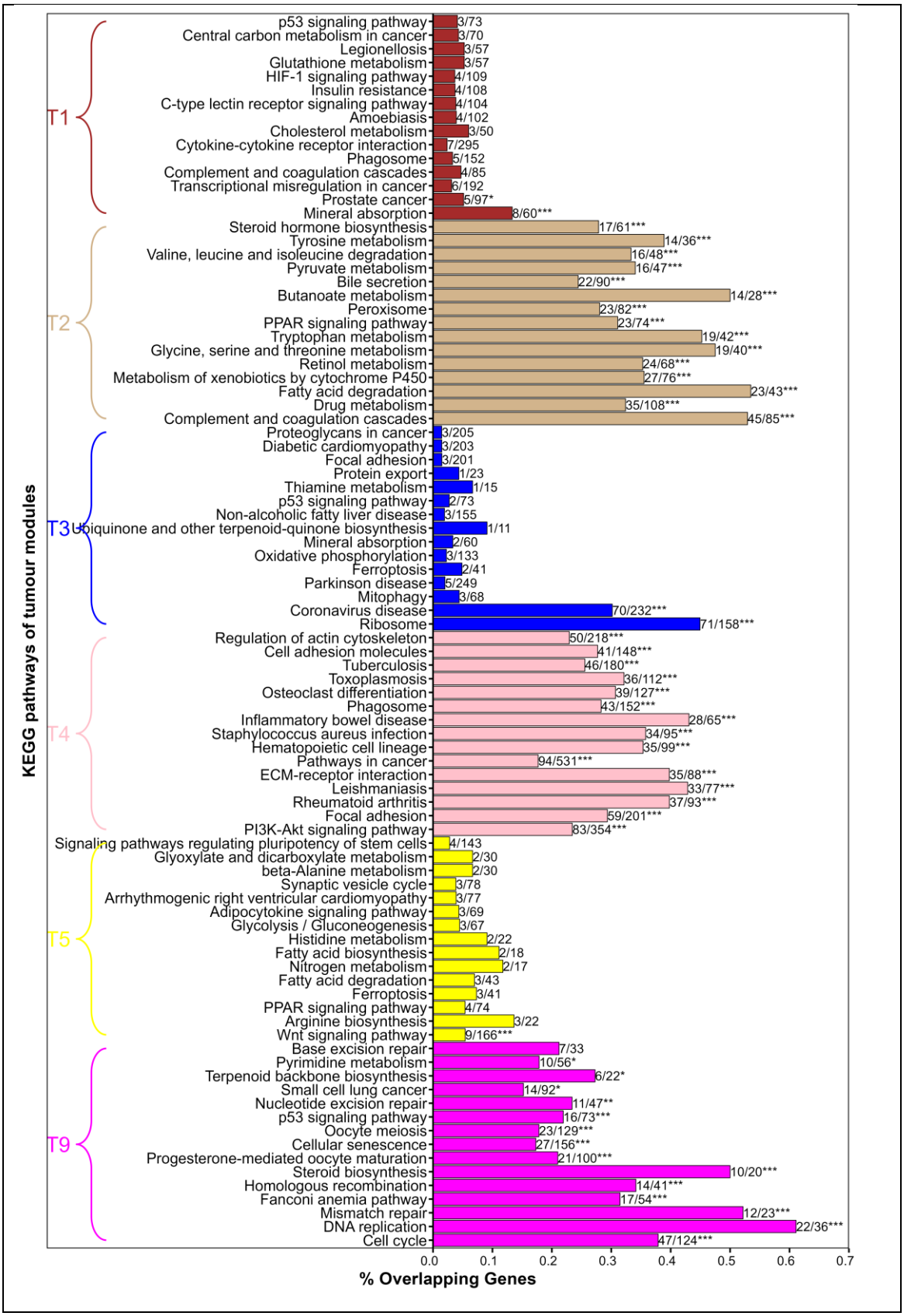


Figure 4.4: KEGG pathway enrichment of tumour modules. For each module, 15 most significant pathways sorted according to adjusted p-value are displayed (bottom to top within each module). * indicates adjusted p-value < 0.001, ** indicates 0.001 ≤ adjusted p-value < 0.01, * indicates 0.01 ≤ adjusted p-value < 0.05. To the right of each bar, the number of overlapping genes and the total number of genes of a pathway are shown.**

eigengene expression pattern of this module shows that DFS decreases with an increase in cell cycle activity. The T1 module includes cancer-related genes and pathways relevant for DFS prediction post-treatment. The T2 module that is positively correlated with DFS is enriched for amino acid metabolism, fatty acid degradation, and xenobiotic metabolism, indicating the capability of the liver to carry out its basic functions post-treatment, thus improving survival. It is also associated with complement and coagulation cascades. The T3 module is negatively correlated with DFS and is associated with ribosomes. T4 and T5 modules are associated with the treatment and are related to ECM and regulation of the Wnt signalling pathway, respectively.

Since some modules showed a significant correlation with DFS, we checked if their respective eigengene expression could be used to identify differences in survival probability **(Figure 4.5)**. We observed that the corresponding eigengene expression of DFS modules (median) also performs well in predicting the survival probability. Further, the hub genes of these modules also predicted the differences in survival probability and helped us to identify biomarkers. **Table 4.1** shows the list of hub genes in the T1 module and their association with the DFS of patients. The low expression of the macrophage scavenger receptor gene MARCO is associated with poor DFS in HCC patients. The evaluation of

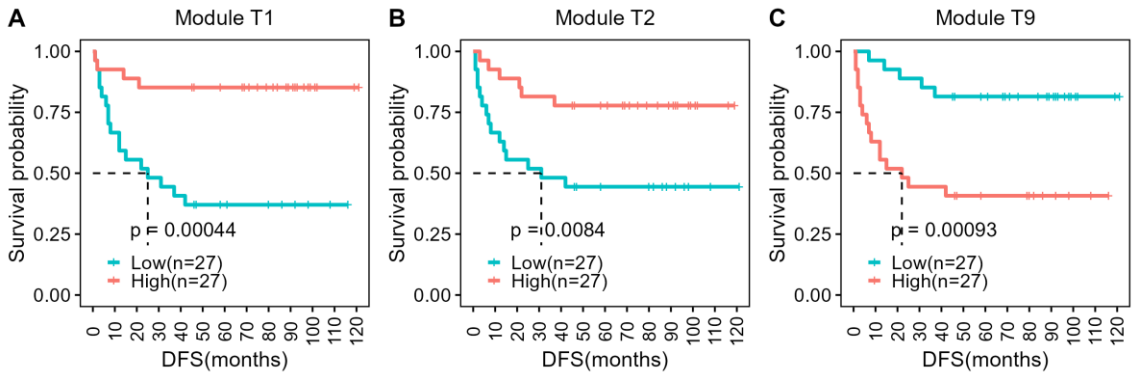


Figure 4.5: Survival analysis based on eigengene expression of tumour modules. Samples are classified into high and low-expression groups based on the median of eigengene expression of each module. 'p' indicates the p-value of survival analysis.

MACRO protein expression by immunostaining in HCC shows that its level decreases as the disease condition worsens[300]. CELC1B is a platelet-related gene, and its expression is related to immune cell infiltration[301]. CFP regulates the complement pathway, and its expression correlates with the infiltration of immune cells[302]. Genes related to the lectin pathway of complement activation (COLEC10, FCN2, FCN3) are also DFS hub genes of

Table 4.1: Survival analysis of hub genes of module T1 on Korean cohort tumour samples.

Gene	p-value	Hazard ratio
CLEC1B	0.00037	0.17
COLEC10	0.00044	0.18
MARCO	0.00044	0.18
CRHBP	0.00044	0.18
CFP	0.00060	0.18
HHIP	0.00064	0.18
FCN2	0.00220	0.23
FCN3	0.00330	0.25
DNASE1L3	0.00360	0.25
GDF2	0.00370	0.25
BMP10	0.00370	0.25
PLAC8	0.02000	0.34
STAB2	0.02300	0.35
CLEC4G	0.02300	0.35

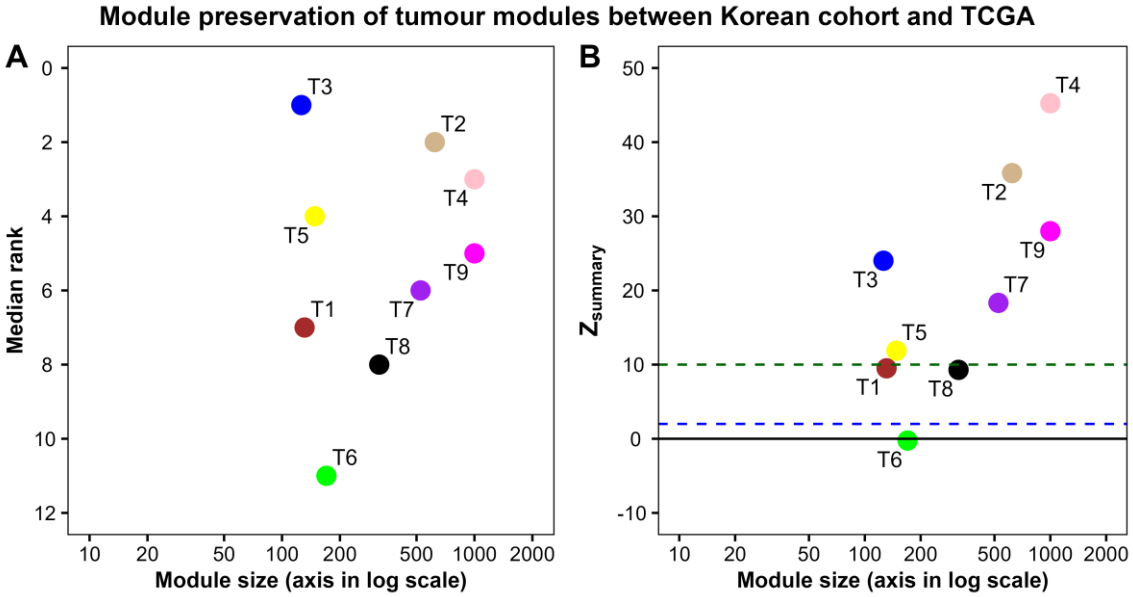


Figure 4.6: Module preservation statistics of Korean cohort tumour modules in TCGA tumour samples. Each point in the plots represents a module. (A) Median rank statistic as a function of module size. Low numbers on the y-axis indicate high preservation. (B) $Z_{summary}$ statistic as a function of module size. Thresholds for moderate (Z = 2) and high (Z = 10) preservation are shown as blue and green lines.

the T1 module. The expression of CRHBP, which mediates the reaction between the corticotropin-releasing hormone and its receptor, is also a predictor of DFS in HCC. The hub genes of the T2 module are related to metabolic processes, including TAT, a tumour suppressor gene. MTHDF1, involved in the interconversion of 1-carbon derivatives of THF, is also a DFS hub gene. Hub genes associated with microtubules and chromosomes from

the T9 module are also good predictors of DFS in HCC. The modules identified from tumour samples of the Korean cohort are also preserved in tumour samples of TCGA **(Figure 4.6)**. Further, the genes from the above modules also show significant survival differences in TCGA tumour samples **(Table C1)**.

4.3.2 Progression from precancerous to cancerous state in TH and PH samples

Module-trait correlation with tumour samples revealed that modules significantly correlated with DFS also captured the differences in surgery a patient has undergone. Based on these observations, we hypothesised that there could be differences in the mechanism of precancerous to cancerous progression between two groups of patients undergoing either TH or PH. Therefore, we investigated the differences in the progression by identifying co-expression modules in each group from both tumour and tumour-adjacent normal samples and correlating them with disease conditions **(Figure 4.7)**.

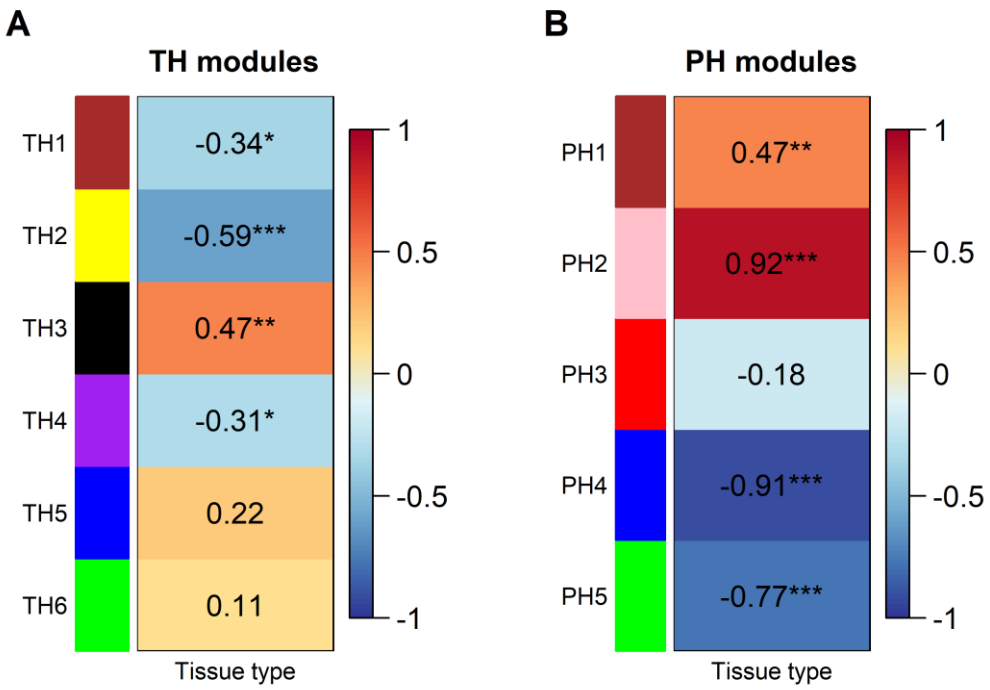


Figure 4.7: Module-trait correlations of precancerous-cancer samples in (A) TH and (B) PH treatment groups. Tissue type is a binary variable with the precancerous state as 1 and cancer state as 2. * indicates p-value < 0.001, ** indicates 0.001 ≤ p-value < 0.01, * indicates 0.01 ≤ p-value < 0.05.**

The progression from precancerous to cancerous state in both groups shows that liver function is affected in tumour samples. Tumour samples show a decrease in liver function (TH1 module in TH group, PH5 module in PH group) and compromised immune-related pathways (TH4 module in TH group, PH4 module in PH group) **(Figure 4.8)**. The

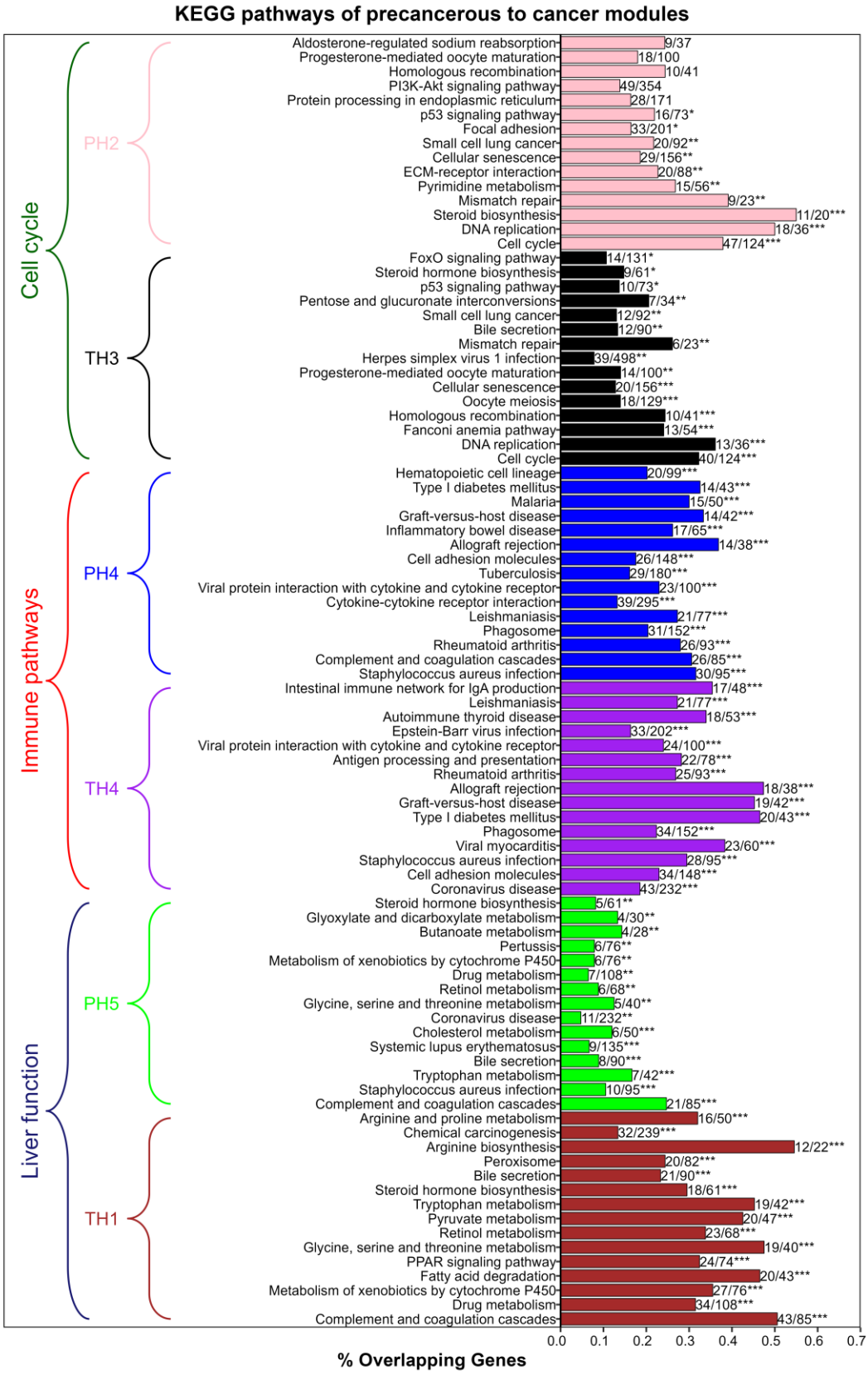


Figure 4.8: KEGG pathway enrichment of precancerous to cancer modules. For each module, 15 most significant pathways sorted according to adjusted p-value are displayed (bottom to top within each module). * indicates adjusted p-value < 0.001, ** indicates 0.001 ≤ adjusted p-value < 0.01, * indicates 0.01 ≤ adjusted p-value < 0.05. To the right of each bar, the number of overlapping genes and the total number of genes of a pathway are shown.**

transcription factor enrichment of TH1 and PH5 modules based on ENCODE data shows HNF4A as an associated transcription factor. Modules capturing cell cycle changes in both groups (TH3 module in TH, PH2 module in PH) show a positive correlation in tumour samples. We observed significant correlations to these biological processes in PH compared to TH **(Figure 4.7)**. PH4 module shows a higher negative correlation compared to the TH4 module with respect to tumour samples, suggesting that immune response genes are significantly downregulated in PH compared to TH. Another feature difference in the precancerous to cancer progression is that the cell cycle module shows a very high positive correlation in PH samples (PH2 module) compared to TH samples (TH3 module). A comparison of these modules in both groups shows some overlap, but the majority of genes are unique to a particular module in a group **(Figure 4.9)**. These sets of unique genes may account for the difference in the precancerous to cancer progression. This analysis gives a global picture of precancerous to cancer progression in both TH and PH groups fuelled by deviations in liver function, cell cycle, and immune response.

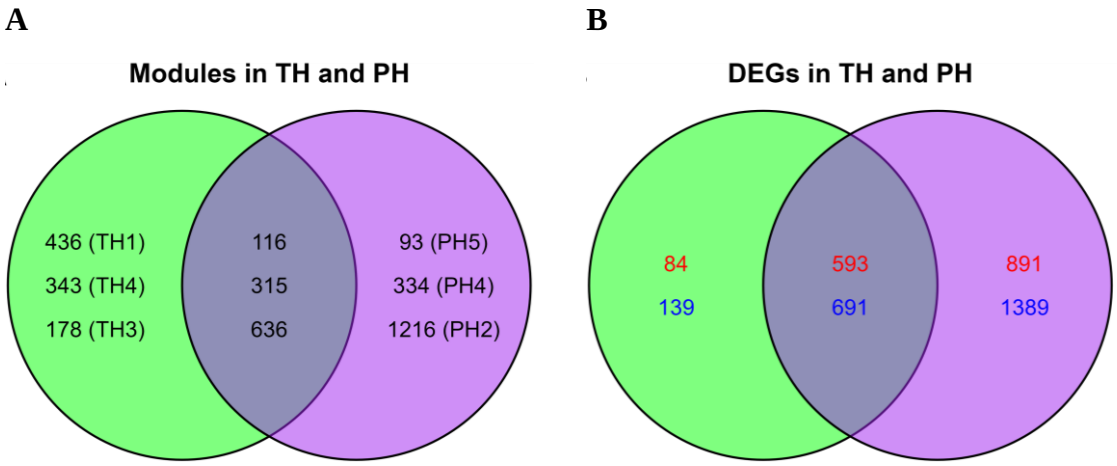


Figure 4.9: Comparison of precancerous to cancer progression in TH and PH treatment groups. (A) Venn diagram comparing precancerous-cancer modules in TH and PH. (B) Venn diagrams showing DEGs between cancer versus precancerous samples in TH and PH groups. Red and blue colour text indicates up and down-regulated DEGs.

In addition to these observations, DEGs comparing tumour versus adjacent normal in both the groups also supports this stark difference in cell cycle and immune response genes between the two groups **(Figure 4.10)**. Further, we also observed that oxidative phosphorylation genes are downregulated in the TH group, while genes of choline metabolism in cancer and arginine biosynthesis are upregulated in PH. Genes of Th1 and

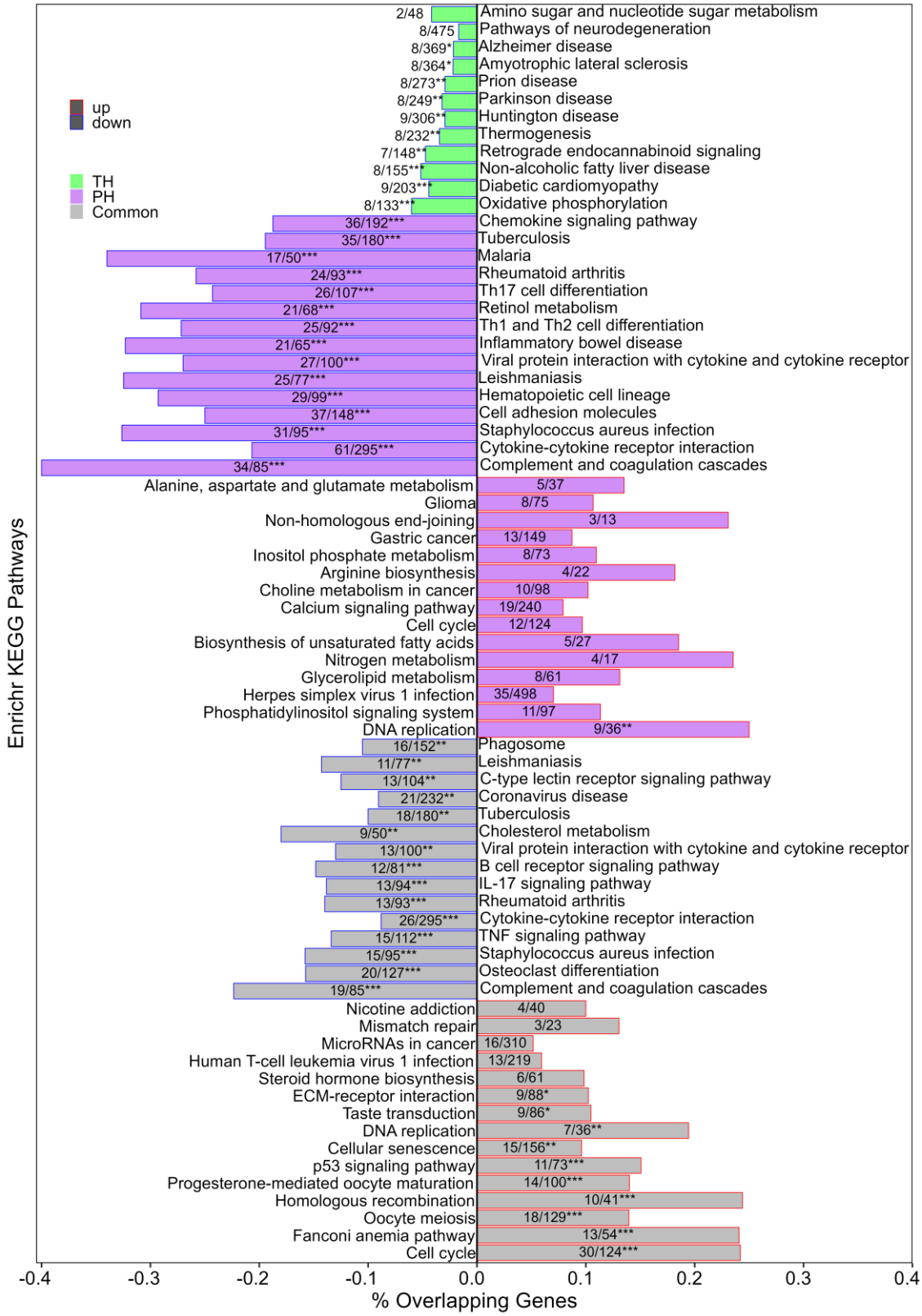


Figure 4.10: KEGG pathway enrichment of up and down regulated DEGs in different categories (TH only, PH only, Common to TH and PH). In each category, pathways with p-value < 0.05 within the top 15 pathways are shown. X-axis represents % of overlapping DEGs genes in each pathway. For down regulated pathways, negative value of % overlap is plotted. * indicates adjusted p-value < 0.001, ** indicates 0.001 ≤ adjusted p-value < 0.01, * indicates 0.01 ≤ adjusted p-value < 0.05.**

Th2 cell differentiation, Th17 differentiation, and complement and coagulation cascades are downregulated in the PH group. We also performed module preservation analysis using TCGA samples **(Figure 4.11)**. The modules from the PH and TH groups show medium to high preservation in TCGA samples.

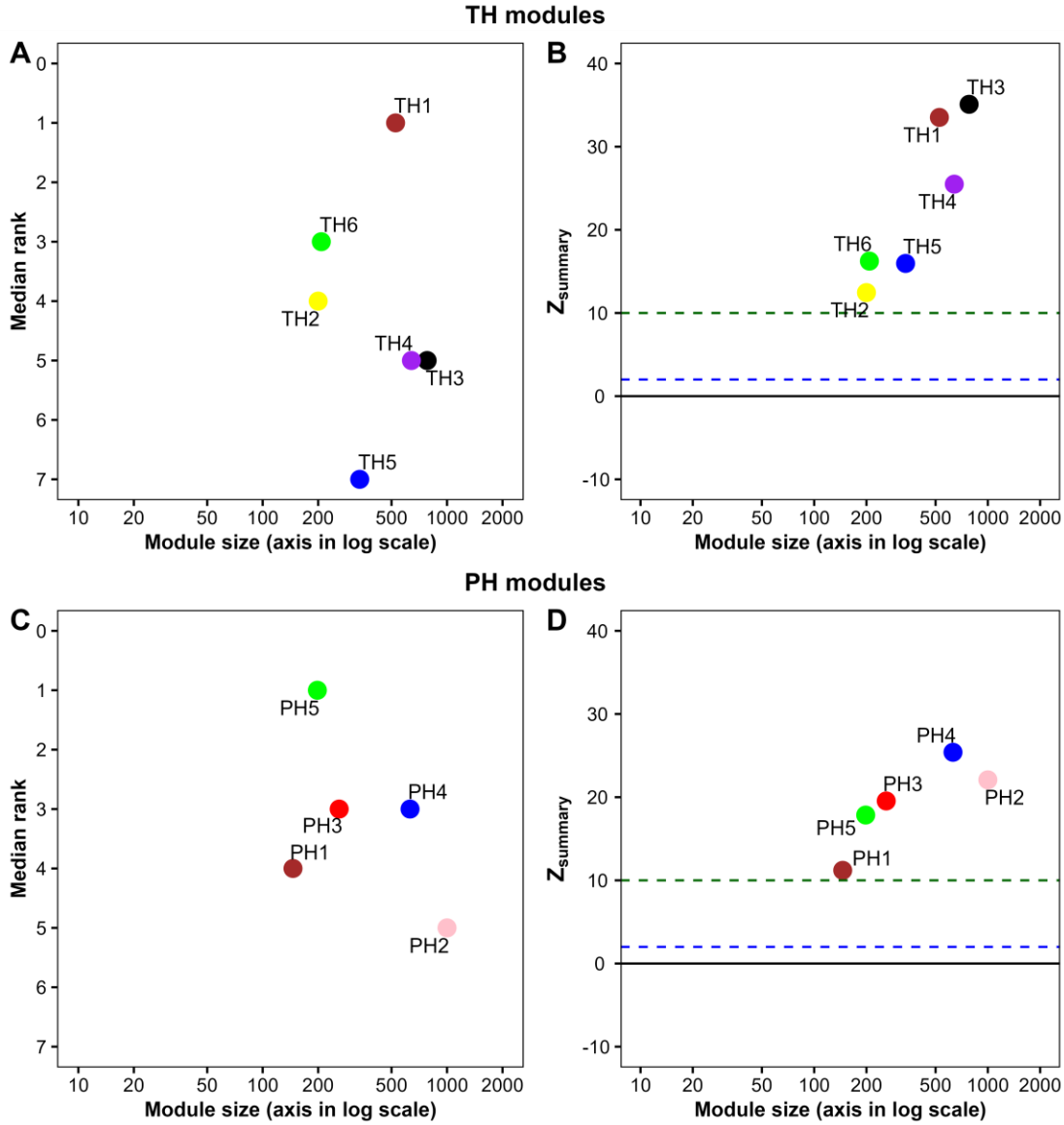


Figure 4.11: Module preservation statistics of Korean cohort precancerous-cancer modules in TCGA tumour and adjacent normal samples. Each point in the plot represents a module. (A), (C) Median rank statistic as a function of module size. Low numbers on the y-axis indicate high preservation. (B), (D) $Z_{summary}$ statistic as a function of module size. Thresholds for moderate (Z = 2) and high (Z = 10) preservation are shown as blue and green lines.

4.3.3 Co-expressed modules of normal and premalignant samples

The premalignant condition (47 samples) in the dataset ranges from fibrosis (low and high grade) to cirrhosis and dysplastic nodule (low and high). We also included 15 normal

samples to capture the changes from normal to premalignant lesions sequentially based on co-expression analysis. We found seven modules significantly correlating with premalignant stages **(Figure 4.12A)**. The N5 module showed a significantly high correlation to premalignant stages with low expression in normal and FL stages and high expression in FH, CS, DL, and DH stages **(Figure 4.13)**. This module is associated with cellular response to type 1 interferon, cytokine-mediated signalling, and defense response to the virus **(Table C2)**. The N7 module is related to neutrophil-mediated immunity and inflammatory response. The N10 module is also positively correlated to premalignant stages, capturing the changes in gene expression that occurred early in the FL stage. These early changes are associated with the complement coagulation cascade, ribosome machinery, and lipid metabolic process. The N11 module showed a gene expression pattern similar to the N10 module and is associated with mitochondrial oxidative phosphorylation.

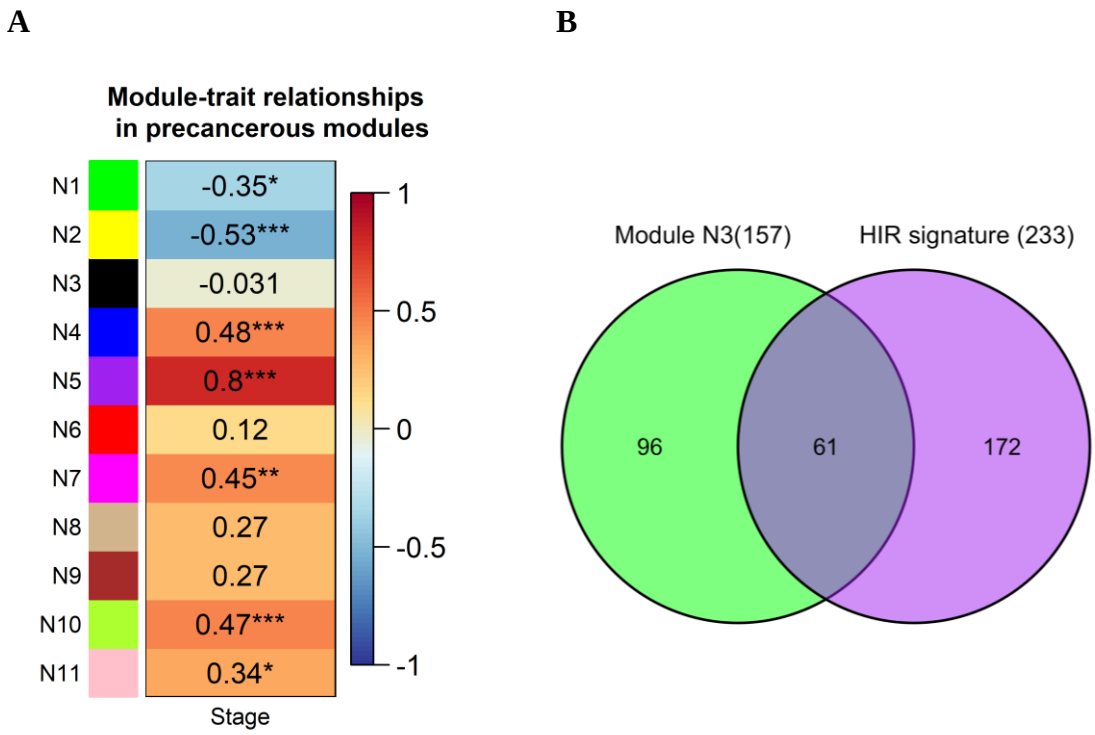


Figure 4.12: (A) Module-trait correlations of premalignant samples. Stage represents different premalignant conditions (N, FL, FH, CS, DL, DH). (B) Venn diagram showing the intersection of module N3 genes and genomic predictors of recurrence[303].

The N2 module is negatively correlated with the premalignant stage and is enriched for metabolic pathways linked to liver function and HNF4 transcriptional activity. The eigengene expression shows that liver function is compromised in the late premalignant stages **(Figure 4.13)**. Intriguingly, the N4 module that is positively correlated with

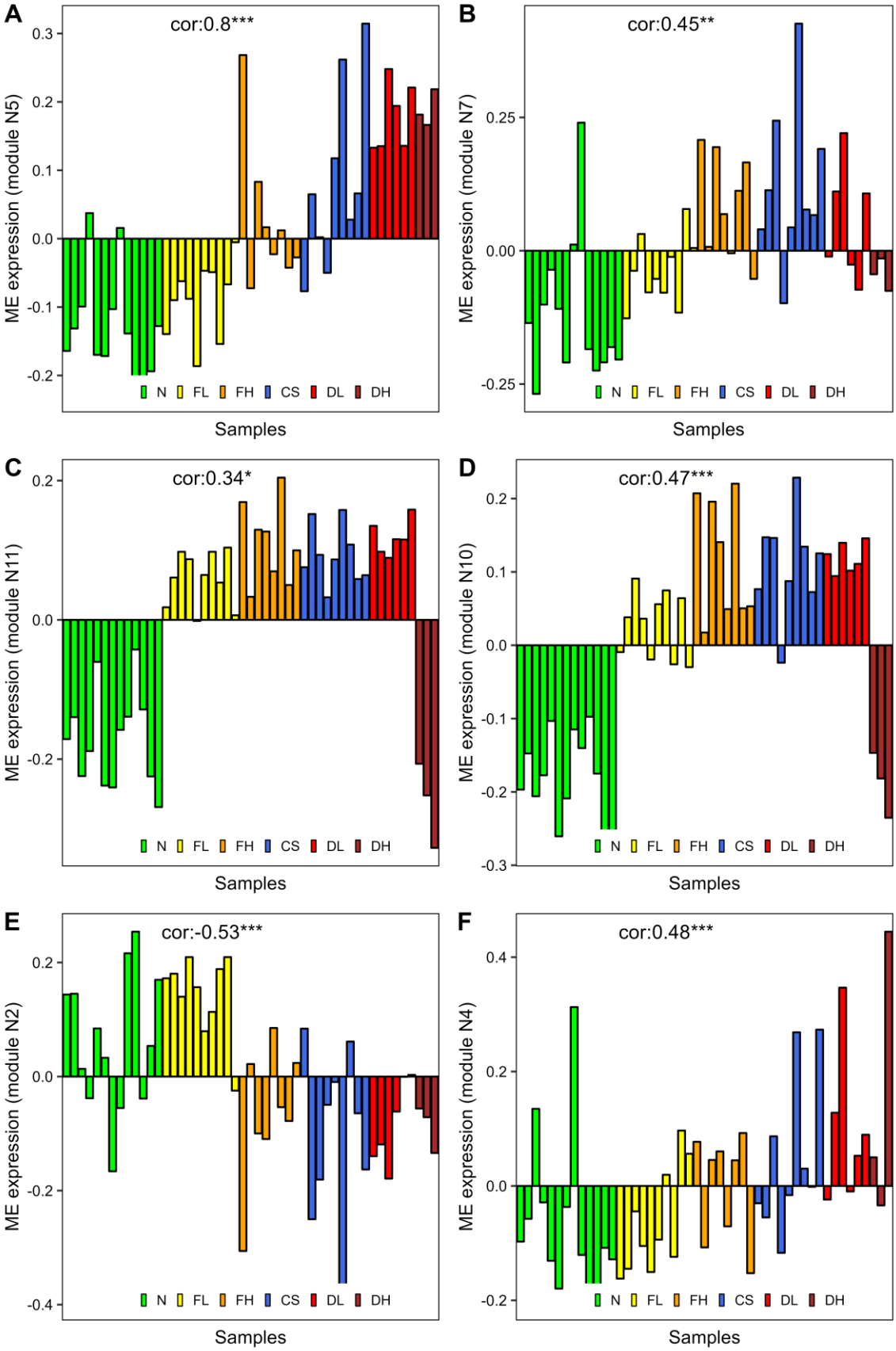


Figure 4.13: Eigengene plots for individual premalignant modules showing correlation with premalignant stage. * indicates p-value < 0.001, ** indicates 0.001 ≤ p-value < 0.01, * indicates 0.01 ≤ p-value < 0.05.**

premalignant stages is associated with cell cycle pathways, showing the onset of the tumourigenesis process.

A previous study on HCC showed that hepatic injury and regeneration (HIR) signature (233 genes) is a good predictor of DFS using premalignant samples from the Chinese cohort[303]. We verified the overlap of the HIR signature with premalignant modules identified through our analysis. We observed that only the N3 module, which is not associated with premalignant states, showed a significant overlap of 61 genes with the HIR signature **(Figure 4.12B)**. The N3 module is associated with immune pathways and cellular senescence.

We also tested the ability of individual modules to predict the DFS. For this purpose, the Chinese cohort was chosen due to the large sample size with clinical information compared to the Korean cohort. For each module (N1 – N11) identified in the Korean cohort, we calculated the corresponding eigengene from paired normal samples of the Chinese cohort, followed by survival analysis based on eigengene expression. The eigengene expression of the N3 module predicts DFS (p-value=0.004) with a high expression value associated with poor survival **(Figure 4.14)**. It was observed that 29 out of the 61 intersecting genes between the N3 module and the HIR signature performed well in predicting the DFS in univariate Cox regression analysis. These include genes, PLK2, ODC1, WWC1, MYC, DDX21, SOCS3 (**Table C3**). The high expression of these genes is associated with poor survival. 13 genes out of 96 non-intersecting genes also showed good predictability of DFS. Interestingly, we also observed that eigengene expression of modules associated with premalignant stages (N10, N7 and N5) predicted DFS based on normal/premalignant samples **(Figure 4.14).** The N5 module yielded the best p-value of 0.0022 in the DFS analysis. THBD (p-value=0.00035) and BCL2L1 (p-value=0.0007) are top candidate DFS genes from the N7 module **(Table C3)**. THBD is a classical marker for dendritic cells (DCs). Increased DCs are associated with early relapse of HCC[304]. BCL2L1 promotes invasion and inhibits apoptosis of liver cancer cells[305]. High expression of FOS (p-value=0.005) and JUN (p-value=0.0015) in the N5 module are also associated with poor DFS **(Table C3)**. Thus, by extracting modules of co-expressed genes from premalignant samples, we identified biomarkers for DFS prediction.

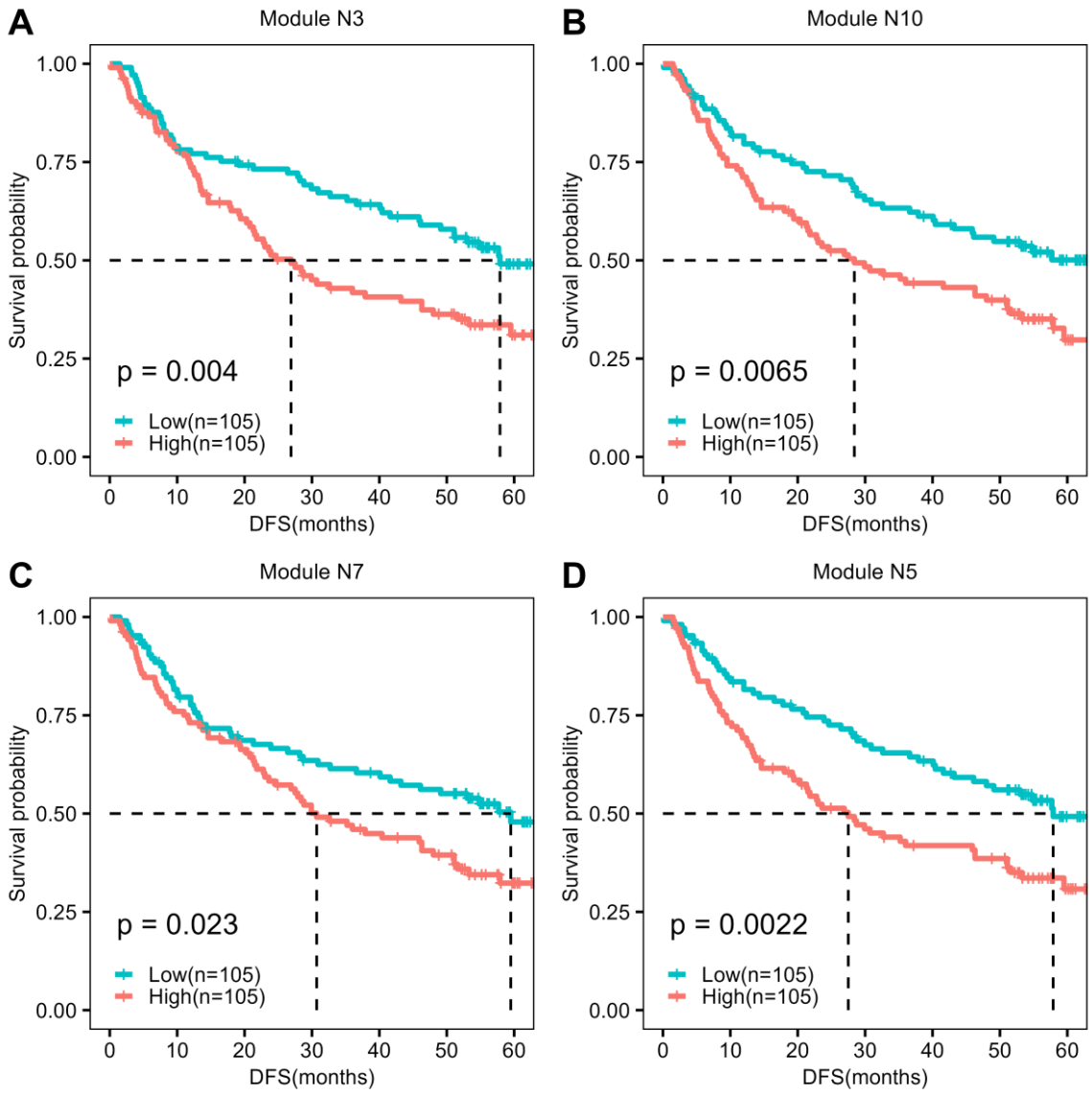


Figure 4.14: Eigengene-based survival analysis of tumour-adjacent normal samples in the Chinese cohort. For each of the premalignant modules from the Korean cohort that showed a significant correlation to the premalignant state, the corresponding eigengene was calculated in the Chinese cohort tumour-adjacent normal samples. Samples are classified into high and low-expression groups based on the median of eigengene expression of each module. 'p' indicates the p-value of survival analysis.

4.3.4 Cell cycle related pathways change in progression from normal to precancer to HCC

It was observed that cell cycle-related pathways were enriched in premalignant samples (N4 module). Similarly, TH3 and PH2 modules from the precancerous to the cancer stage of TH and PH samples were also associated with the cell cycle. The overlap of these module

genes with the cell cycle-related genes obtained from the GO term showed that 55 genes are common and found in precancerous stages **(Figure 4.15)**. We observed an increase in cell cycle-related genes with progression from precancer to cancer in TH3 and PH2 modules, having 222 and 365 genes, respectively. There are 169 cell cycle genes unique to the PH2 module.

To gain further insights into the cell cycle processes, genes of the individual modules (N4, TH3, PH2) overlapping with cell cycle genes (56, 222, 365 genes, respectively) were visualised using GO biological processes with ClueGO Cytoscape plugin. The 55 common genes map to biological processes related to the kinetochore, microtubule and chromosome **(Figure 4.16).** GO terms unique to TH3 and PH2 modules suggest the progression differences from precancerous to cancer state between TH and PH conditions. Checkpoint signalling, negative regulation of the cell cycle process, DNA repair process, and regulation of exit from mitosis are observed in the TH3 module but not in the PH2 module **(Figure 4.17 and Figure 4.18)**. On the other hand, the PH2 module shows positive regulation of cell cycle, proliferation, cell division, and cytokinesis, along with positive regulation of protein metabolic processes. There is an increase in the number of genes related to microtubule spindle organization compared to N4 and TH3 modules. Further, DFS cell cycle genes related to microtubules, kinetochores, and centromere also overlap with genes of the N4 module, suggesting some of these changes are associated with premalignant stages.

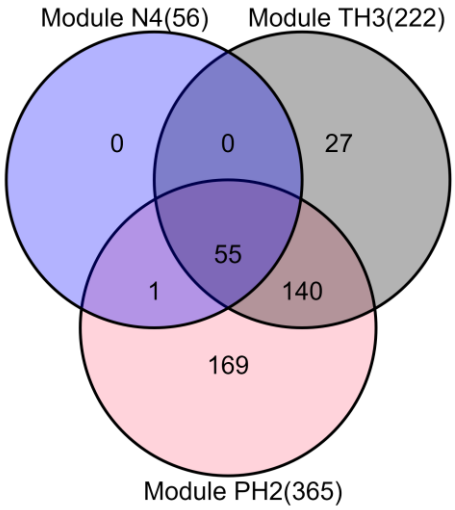


Figure 4.15: Venn diagram showing the overlap of cell cycle genes with modules significantly enriched for cell cycle related pathways.

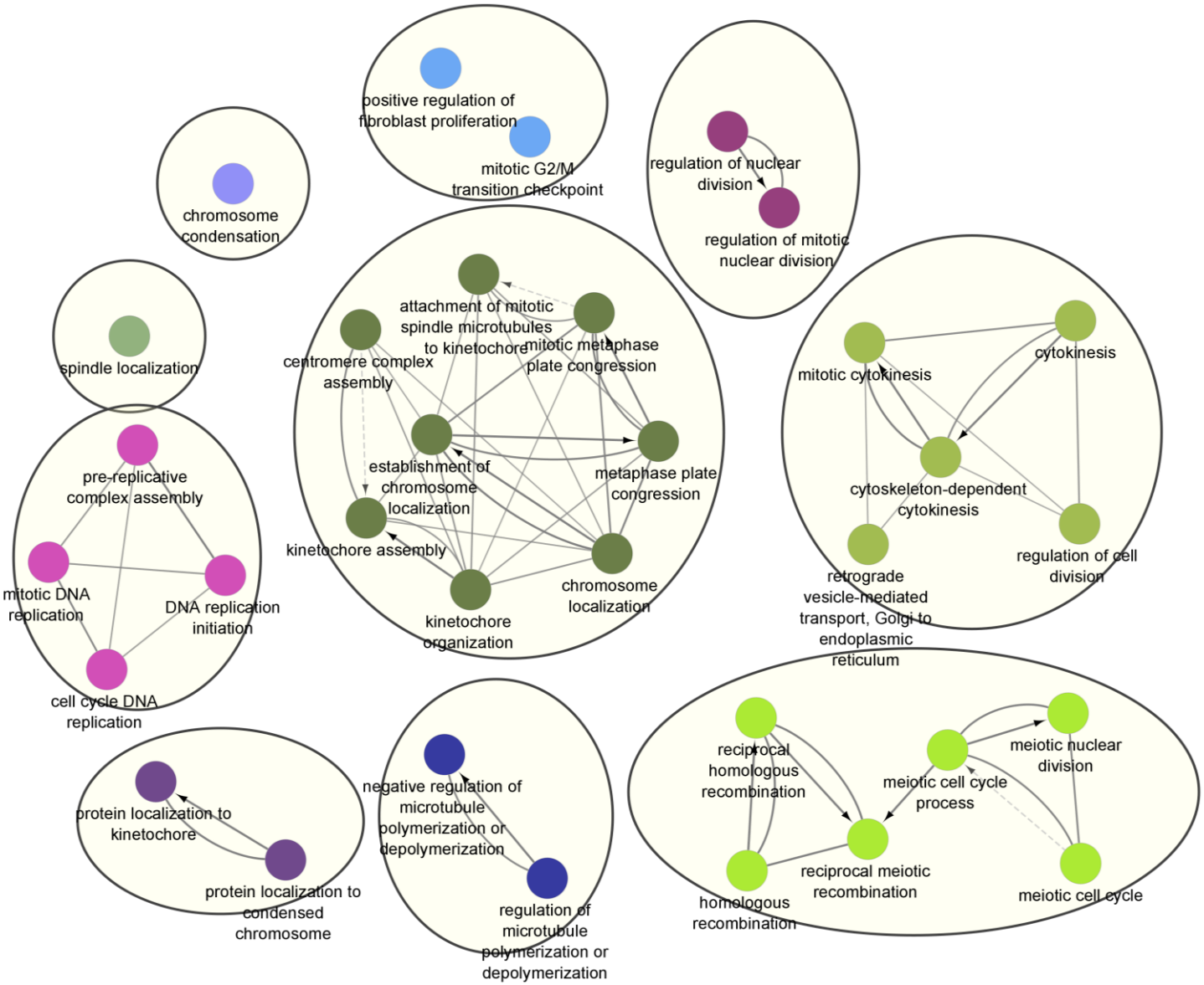


Figure 4.16: Network of GO biological Processes of Cell cycle genes in premalignant module N4

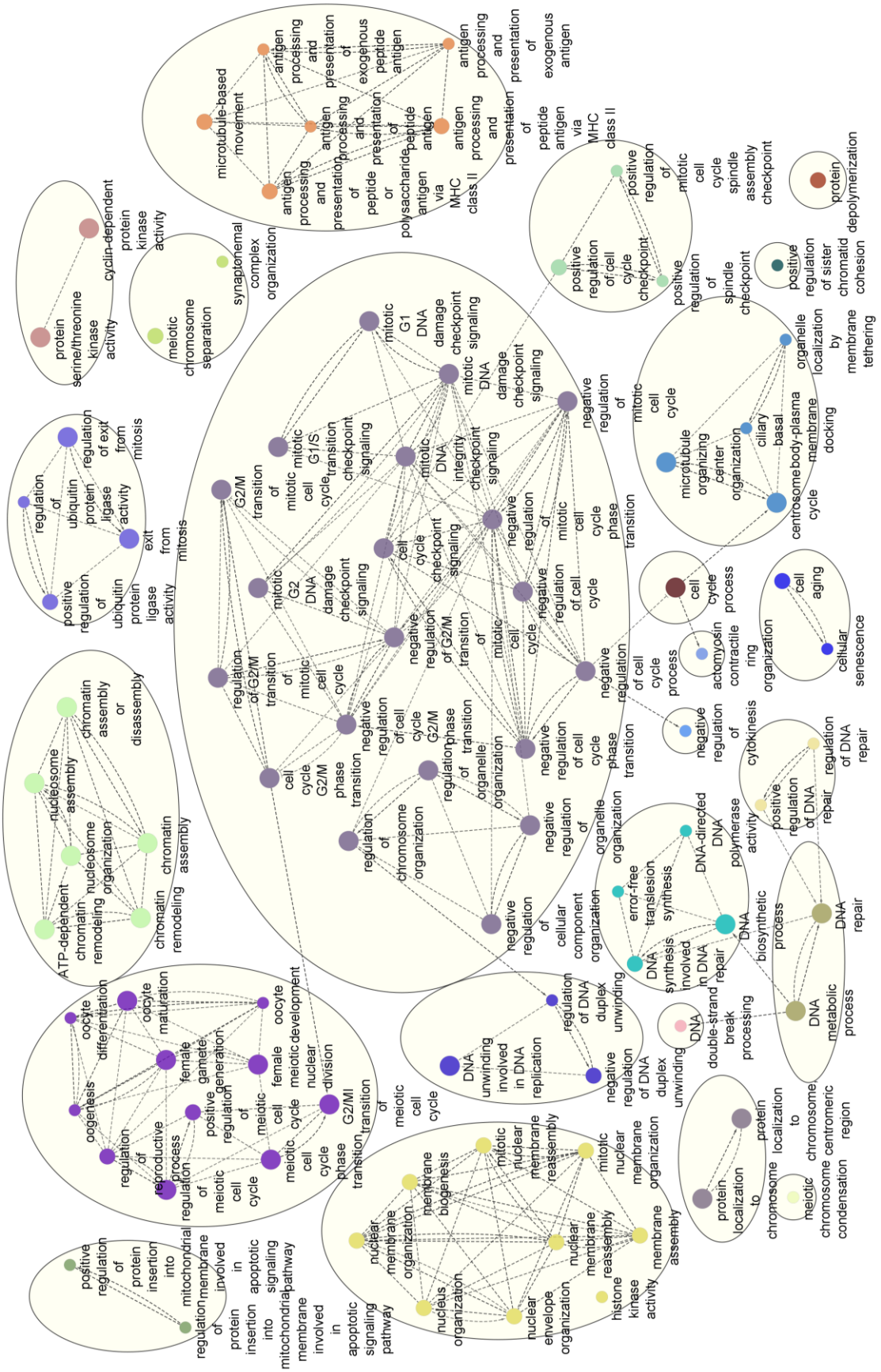


Figure 4.17: Network of GO biological Processes of Cell cycle genes in precancerous-cancer module TH3

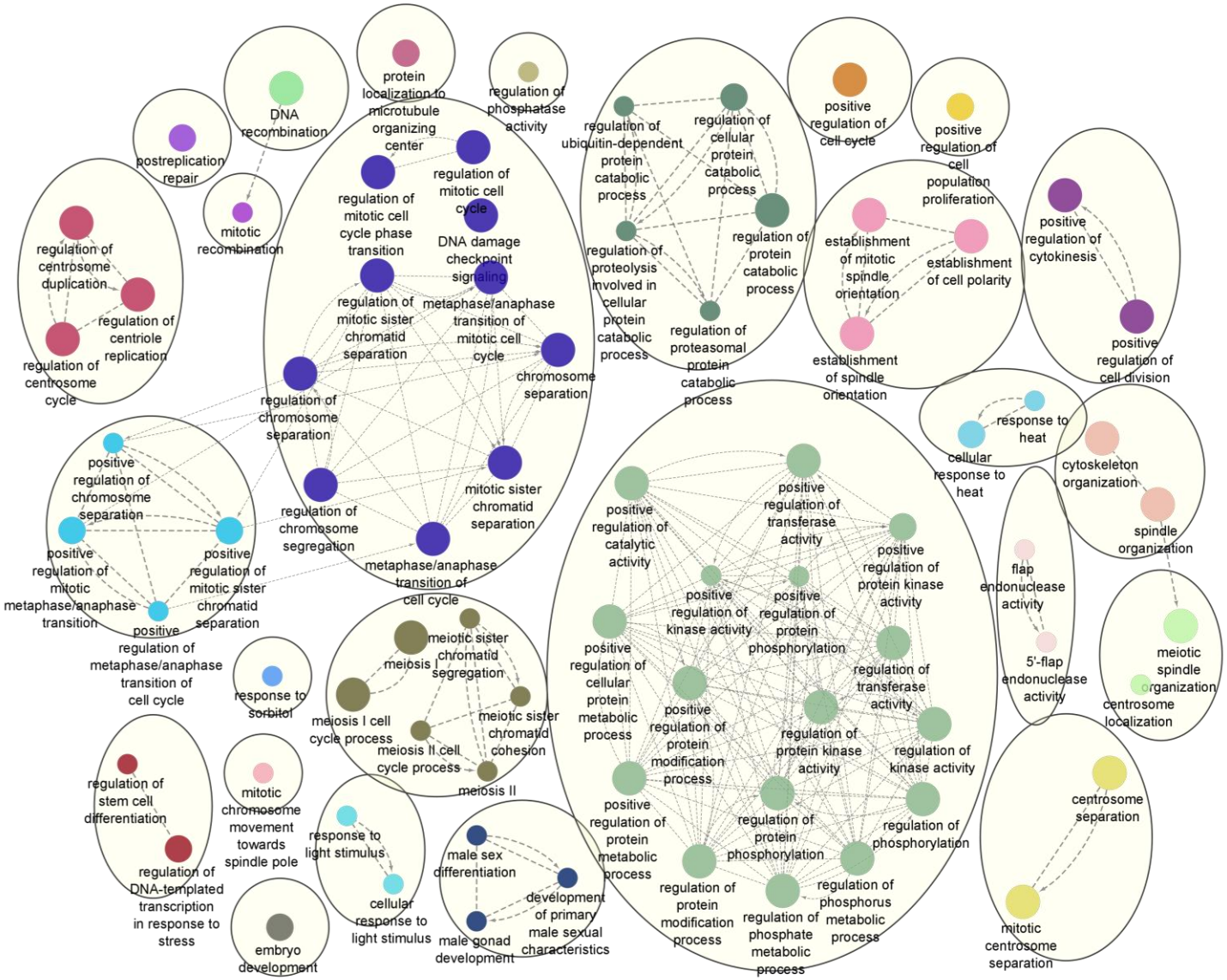


Figure 4.18: Network of GO biological Processes of Cell cycle genes in precancerous-cancer module PH2

4.4 Discussion

Understanding the molecular mechanisms involved in the progression of HCC through multiple trajectories is crucial for improved diagnosis, prognosis, and treatment. In this direction, we investigated publicly available transcriptomics data of HCC patients undergoing liver transplantation or resection treatment. Gene co-expression network-based framework was employed to get molecular insights from the transcriptomics data of tumour samples, tumour-adjacent normal samples in different premalignant states, and normal samples. This approach identified modules of co-expressed genes, pathways, and genes that characterize different trajectories and predict DFS based on premalignant and tumour samples.

Modules and genes related to the cell cycle, immune system, ribosome, and liver metabolic pathways were good predictors for DFS using tumour samples **(Figure 4.5 and Table C1)**. An increase in the ribosome and cell cycle activity and a decrease in the expression of immune (complement system) and liver metabolic genes are associated with poor DFS. Liver function and proliferation are shown to be mutually exclusive, and the transition to proliferation occurs with the inhibition of liver function[306]. HCC occurrence and progression are related to the interaction between viruses and ribosomes[307]. A decrease in the complement system also indicates a change in the immune infiltration patterns. DFS modules were also associated with the treatment (surgery) given to patients: PH and TH **(Figure 4.3)**. In addition, we also found a tumour module (T4) linked to ECM to be associated with treatment **(Figure 4.4)**.

The network analysis of patients who have undergone PH and TH was performed independently, including the tumour and corresponding tumour-adjacent normal samples, to understand the differences in progression. We observed that the same biological processes are affected to a different extent in TH and PH groups. Both groups show a decrease in liver function and immune system and an increase in cell cycle activity. However, the tumour samples in the PH group show a very high correlation to these biological processes **(Figure 4.7 and Figure 4.8)**. This indicates that the extent of immune suppression and decrease in liver function is related to cell cycle activity in tumour samples, bringing about the variability in the outcomes. This view contrasts with our observations from modules identified from normal and premalignant samples. We observed an increase in immune activity and cell cycle gene expression and a decrease in liver function. An

increase in immune activity may be associated with the antiviral mechanism (most patients have HBV infection) by interferon signalling. Genes of immune modules (N5, N7, N10) show some overlap with downregulated immune modules (TH4 and PH4) specific to tumour samples. FOS and JUN are part of the upregulated module in premalignant samples and downregulated modules in tumour samples. This suggests a shift in immune activity from a premalignant state to tumour state. Pro-inflammatory M1 marker CCL2 decreases in tumour modules TH4 and PH4 but increases in the premalignant state. The expression of fibrotic genes EGR1, JUND, KLF2 and TAGLN also decreases in tumour samples (PH4 module).

On the other hand, genes related to liver function decrease in premalignant and tumour samples. HNF4A, which controls liver function, is known to be inhibited by increased inflammation (immune activity) in liver fibrosis[308,309]. The expression of HNF4A leads to the restoration of metabolic function and reversing (attenuation) of liver fibrosis and cirrhosis via controlling macrophages and hepatic stellate cells[310]. HNF4 drives the transition of macrophages to the M2 phenotype. We hypothesise that the mutual antagonism between HNF4A and immune activity plays a role in HCC progression. An increase in inflammation may result in the inhibition of HNF4A with an increase in cell cycle activity. A progressive loss of HNF4A activity is observed in liver diseases (NAFLD and NASH) compared to HCC[311]. We observed that the transition from normal to premalignant to tumour state is also characterised by an increase in cell cycle activity. Genes related to mitotic spindle organisation are present in the premalignant state, and some of them are also DFS genes in tumour samples. An increase in the expression of genes involved in the maintenance of genomic integrity is associated with chromosomal instability (CIN), which is a prognostic factor in multiple cancers[312,313]. There is also an emerging link between CIN and tumour immunity.

We observed that multiple (N3, N7, N5, N10) modules from premalignant samples are good predictors of DFS **(Figure 4.14)**. The N3 module showed some overlap with the HIR signature, which was earlier proposed for DFS prediction. However, we identified three more modules that can be used for the prognostic task. These modules are associated with the immune system. We obtained the best performance (p value=0.0022) with the eigengene expression of the N5 module **(Figure 4.14D)**. These modules are associated with premalignant conditions (fibrosis, cirrhosis), and an increase in the eigengene expression is

associated with poor survival. This suggests that early relapse can also be predicted based on tumour-adjacent normal immune environment. Most studies on HCC relapse are based on immune cell recruitment in tumour samples. Early-relapse HCC cases have increased recruitment of dendritic cells (DC) and $CD8^+$ T cells compared with primary tumours[299,304]. However, our study showed that the gene expression of tumour-adjacent normal samples of HCC patients contains multiple signatures relevant to predicting DFS.

Chapter 5 Transcriptomic network analysis of liver transplantation: from rodent to human model of liver regeneration

5.1 Introduction

The incredible ability of the liver to acquire its lost mass (due to injury or partial resection) has been extensively investigated using rodent models. Investigation of liver regeneration in recent times has shifted to understanding the cellular processes that lead to restoration of liver mass post-transplantation or resection in humans[67]. In rodents, 2/3rd partial liver resection or a 30% partial liver graft restores to a size of 100% within 7-10 days and in humans, it takes around three months to restore liver mass after hepatectomy[314]. Despite enormous differences in the time scales, a study by Young and Periwal (2016) showed that the phenomenological model of rat liver regeneration could be adapted to humans with changes in parameters characterising metabolic load per hepatocyte. It is a theoretical measure of metabolic demand on the liver normalised to liver mass[315].

Animal models of liver regeneration have created hope of extrapolating the findings to translational impact on human health. The phylogenetic similarity between species argues for a unifying principle. However, a detailed understanding of the molecular mechanisms after liver donation or resection in the human liver regeneration context would be vital for clinical application. The major goal of the adult-to-adult donor liver transplantation cohort studies is to understand the clinical manifestations of liver regeneration. This involves characterising the growth pattern in donors and recipients and their final outcomes[316].

The transplantation procedure consists of three main stages: 1) donor surgery, 2) backbench preparation of the liver graft and 3) recipient surgery. Liver transplantation can be either from a deceased or live donor **(Figure 5.1)**. In the case of a live donor, a partial liver graft is removed, and the blood vessels are flushed with a storage solution to prevent thrombosis and maintain its functionality. During the backbench preparation, the graft is stored in a cold preservation solution until implantation. In the case of cadaveric liver transplantation, the donor's liver is resected with an adequate length of blood vessels, which are later connected to the blood vessels in the recipient[317]. During transplantation, the time from the perfusion of the donor's liver with storage solution to its removal from cold storage is defined as cold ischemia time. The time from the removal of the donor's liver from cold

storage to the implantation of the graft into the recipient (reperfusion) is called warm ischemia time. The restoration of circulation in the recipient via the portal vein and arterial anastomosis is called reperfusion[317,318]. The graft from a living donor helps to reduce dependency on dead donors' availability and reduces the wait time for liver transplantation.

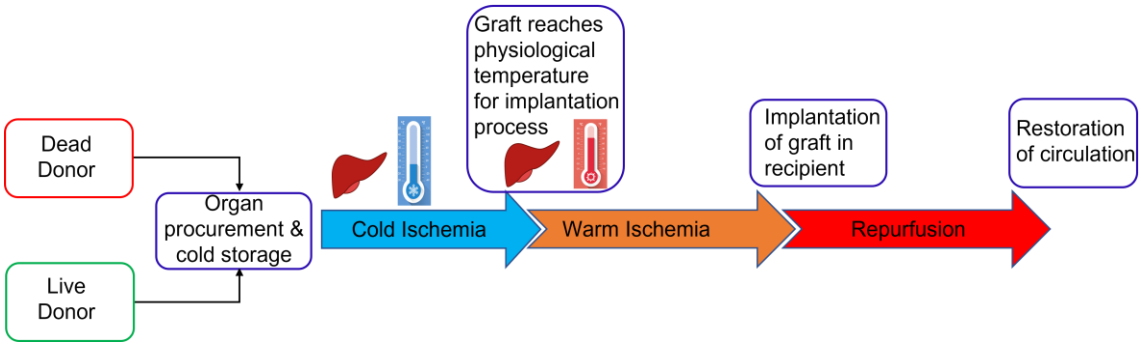


Figure 5.1: The process of liver transplantation from a donor to a recipient.

The regeneration from graft involves competition between cellular processes to restore the mass vs. the metabolic function of the liver, which can impact the graft function. The graft also undergoes cellular damage due to cold and warm ischemia, surgical stress, and implantation. In the case of deceased donors, grafts can also undergo damage due to brain death. Ischemia/reperfusion injury (IRI) occur during the restoration of blood flow after a period of ischemia. The early stage of IRI is characterised by energy depletion and metabolic stress followed by neutrophile and macrophage accumulation after reperfusion[319]. Complications arising due to IRI culminate in early allograft dysfunction (EAD) and impaired regeneration, leading to poor post-transplantation outcomes.

Apart from IRI affecting the outcome post-transplantation, the immune system poses challenges that may sometimes lead to graft rejection. Graft rejection is mostly seen within 1-2 weeks of organ transplantation unless the donor and recipient are identical twins or the recipient is supported by immunosuppressive therapy[320]. Unlike other organs, the liver is immunologically privileged due to its anatomical location, specialised dual-flow perfusion system and lobular microarchitecture, thereby rendering spontaneous acceptance of the grafts post-transplantation. Liver grafts have been shown to withstand rejection despite the absence of immunosuppression therapy, showing the tolerogenicity of the human liver[321]. An interplay between alloimmunity and inherent immune tolerance poses a challenge for clinicians. Dissecting the molecular mechanisms underlying the induction and maintenance of liver graft tolerance post-transplantation will be beneficial to identify biomarkers for

reliable prediction of tolerance or rejection and the safe window to withdraw immunosuppression **(Figure 5.2)**.

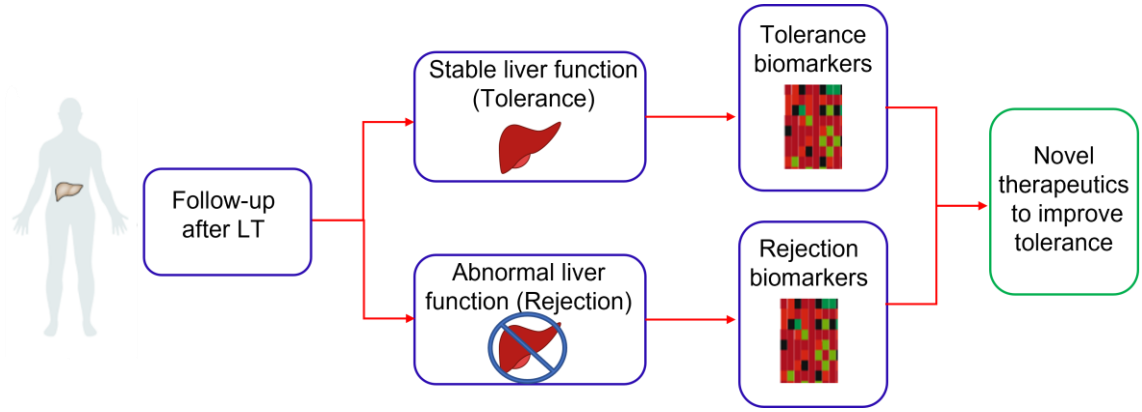


Figure 5.2: Identification of biomarkers for tolerance and rejection post-liver transplantation (LT) using transcriptomic data.

Employing the transcriptional biomarkers from liver biopsies is anticipated to be prognostically accurate compared to blood-based biomarkers in predicting the outcome of immunosuppression withdrawal[322]. Gene expression profiling has also been applied to investigate the role of IRI and correlate it to the clinical outcome of graft dysfunction[323]. Another study[324] analysed the gene expression profiles in the early hours of pre and post-reperfusion and reported the impact of graft type (living donor - LD, deceased donor - DD) on regeneration. These early transcriptome profiling-based studies laid the groundwork to address key issues in liver transplantation by harnessing the power of omics technology.

In the current chapter, we performed a comparison of liver regeneration in rodent models with the human transplantation model to understand cross-species differences/similarities. We also performed the differential coexpression analysis of liver post-transplantation biopsies to study the differences between patients with long-term stability (normal) and rejection.

5.2 Methods

5.2.1 Transcriptomic data

5.2.1.1 Different models of liver regeneration

For the mouse model of liver regeneration, we used the publicly available temporal RNA seq gene expression data (GSE95135), spanning ten different time-points post-PH (0, 1, 4, 10, 20, 28, 36, 44, 72, 168, and 672h). 3-4 replicates were available for each time point

(detailed description of the dataset in **section 2.2.1**). Based on temporal clusters found in mouse liver regeneration transcriptome (**Figure 2.5 and Figure 2.6**), 1, 4, and 10 h post-PH were considered early phase time points, and 36, 44, and 48 h post-PH mid-phase time points. Differential gene expression analysis was performed using the limma R package, comparing every time point with 0 hours as the reference. For the gene expression of human liver transplantation, we used the microarray profiles of liver biopsies from HCC patients undergoing liver transplantation from the GEO database (GSE12720). Liver biopsies were collected at serial time points of graft procurement (no manipulation), cold preservation (backbench) and 1-hour post-reperfusion from 13 patients. Post-reperfusion samples were compared with baseline samples (no manipulation) for differential gene expression. As another model of liver regeneration, we analysed the gene expression pattern of tumour samples from patients undergoing PH/TH for HCC treatment (GSE148355 – detailed description of the dataset in **section 4.2.1**). Further, differential gene expression analysis was carried out between tumour and normal samples.

5.2.1.2 Post-transplantation liver failure or success

To study the gene expression profile of post-transplantation liver biopsies and understand the differences between normal and rejection patients, we analysed the publicly available INTERLIVER study cohort with GEO accession number GSE145780. This is a microarray dataset with 235 biopsy samples collected from 217 liver transplantation patients. All biopsy samples were procured after transplantation with a median time of 962 days post-transplant and ranging between 0-11,676 days. The original INTERLIVER study[325] classified the samples into four different categories: normal (N = 129), T-cell mediated rejection (TCMR) (N = 37), early injury (N = 61), and fibrosis (N = 8) based on previously identified candidate genes known as rejection associated transcripts (RATs)[326]. Here, we attempt to compare the normal and TCMR (rejected) groups using differential coexpression networks.

5.2.2 Cross-species analysis of liver regeneration

To identify genes that are consistently deregulated in mouse and human models of liver regeneration, we performed the GSEA (Gene Set Enrichment Analysis) using fgsea R package[327]. GSEA is primarily used to interpret the expression pattern between two conditions/treatments and provide biological insights. Given a ranked list of genes L based on a differential expression test and a prior set of genes S, the goal of GSEA is to determine

if the members of a geneset S are overrepresented in the top or bottom of the ranked gene list L or are randomly distributed[328]. This behaviour is quantitively captured by the enrichment score (ES), which is calculated by moving down the ranked list L. Iteratively, the ES at every position 'i 'is given by

$$ES_i = \begin{cases} 0 \; if \; i = 0, \\ ES_{i-1} + \dfrac{|r_j|^p}{N_R} \; if \; 1 \leq i \leq N \; and \; i \in S \\ ES_{i-1} + \dfrac{1}{N - N_H} \; if \; 1 \leq i \leq N \; and \; i \notin S \end{cases} \tag{5.1}$$

The final ES is evaluated as

$$ES = ES_{i^*}, where \; i^* = arg \; max_i \; |ES_i| \tag{5.2}$$

i.e. ES is calculated by considering positively and negatively scoring gene sets separately (ES_+ and ES_- respectively), and finally, ES = ES_+ if $|ES_+| > |ES_-|$ otherwise, ES = ES_-. The core members of S contributing to the enrichment signal are extracted based on ES score. These constitute the subset of genes in S that are positioned in L at or before the running sum shoots its maximum deviation from zero and are called leading edges.

The top 500 upregulated genes (differential expression analysis) based on log fold change from each of the human models (transplantation and HCC) were enriched in different time points (1h vs 0h, 4h vs 0h, 10h vs 0h - early phase, 36h vs 0h, 44h vs 0h, 48h vs 0h - mid-phase) of mouse regeneration signatures. Similarly, enrichment scores were calculated for the top 500 downregulated genes. Leading edge genes were extracted from each of the enrichment analyses if the enrichment scores met the significance criteria of FDR < 0.05. Further, the sign of significant enrichment scores was used to identify whether the direction of gene regulation between two species is consistent or not. The upregulated genes from human signatures are consistently enriched in upregulated genes of mouse if the final enrichment score ES = ES_+ and vice versa for downregulated genes.

5.2.3 Differential gene co-expression analysis (DGCA)

Molecular mechanisms driving disease progression arise due to abnormalities in gene co-regulation. Not all changes in gene co-regulation are reflected as the up or down-regulation of individual genes. For instance, a group of genes exhibiting identical average expression in two different conditions may differ in the co-expression pattern between two groups[329].

DCGA captures such differences in co-expression patterns and provides a clue into the rewiring of transcriptional networks. To extract differential co-expression patterns between normal and rejected samples, we applied the DiffCoEx approach, which is based on WGCNA[330]. DiffCoEx identifies modules of differential co-expression by applying WGCNA on the adjacency matrix derived from correlation changes between conditions. This is followed by the calculation of the TOM matrix and module extraction, which are similar to WGCNA (discussed in **section 2.2.2**). In this way, the method clusters two genes together if they show correlation changes to the same sets of genes between different conditions.

Signed adjacency matrices with soft threshold power equal to 12 were separately calculated for normal and rejected samples based on 10,000 top varying genes (using rowVars function in R). The topological overlap matrix was derived from the absolute differences of powered adjacency matrices. Differentially co-expressed modules were extracted by average hierarchical clustering with the dynamic tree cut algorithm.

The statistical significance of differentially co-expressed modules was evaluated using dispersion statistic based on 1000 permutation tests[331]. If G is a differentially co-expressed module with g genes, then the dispersion statistic is given by

$$D\left(\rho_G^{S_1}, \rho_G^{S_2}\right) = \sqrt{\frac{1}{\rho_g}\sum_{m=1}^{\rho_g}(\rho_m^{S_1} - \rho_m^{S_2})^2} \qquad (5.3)$$

where $\rho_m^{S_1}$ and $\rho_m^{S_2}$ represent correlations between genes in a gene pair *m* in two conditions, S_1 and S_2. The summation runs over all the possible gene pairs within a gene set *G* of size *g,* given as $\rho_g = \binom{g}{2}$. The P-value is calculated from the null distribution of dispersion statistic generated by permuting the samples in two groups. Modules showing significant p-values were further considered for downstream analysis.

5.3 Results and discussion

5.3.1 Concordance pattern of gene expression profile in different models of liver regeneration

The cross-species analysis between mouse and human models of liver regeneration was performed using GSEA and leading edges. The GSEA of transplantation gene signatures with the early phase of mouse regeneration shows a consistent transcriptome profile in up-

and down-regulated directions **(Figure 5.3A)**. The pathway enrichment of leading edges contributing to significant enrichment scores is associated with TNF signalling, JAK-STAT signalling, NFKB signalling, MAPK signalling and metabolic pathways **(Figure 5.3B)**. This gene expression pattern and related biological pathways align with the transient downregulation of metabolism and upregulation of immune pathways during liver regeneration (see **Figure 2.5**). In contrast to the transplantation model, the transcriptomic responses in the HCC model depict opposite behaviour with genes in JAK-STAT, IL-17, NFKB signalling pathways downregulated. This can be seen as a significant but positive enrichment score of downregulated genes in HCC **(Figure 5.3A).**

Although the cell cycle pathway is not significantly enriched in the above comparisons between mouse and human regeneration models, cell cycle inhibitor CDKN1A is found in the set of leading edges. CDKN1A is upregulated in transplantation and between 1 and 4h post-PH in mouse regeneration, thus providing temporal separation between the preparatory and proliferation phases **(Figure 5.3C)**. Contrary to this, CDKN1A is downregulated in HCC, confirming that this regeneration model is already in the proliferative phase.

We also compared the human regeneration models with the mid-phase of mouse regeneration (36, 44, and 48 h post-PH). With the mid-phase of mouse regeneration, both transplantation and HCC regeneration models showed consistent upregulation of gene expression **(Figure 5.4A)**. The consistency in transcriptome upregulation between HCC and mouse regeneration is more pronounced than that of the transplantation. The leading edges between transplantation and mid-phase mouse regeneration are associated with immune signalling pathways (TNF and JAK-STAT), ribosome biogenesis, protein processing and PD-1 checkpoint **(Figure 5.4B)**. The core genes exhibiting concordant upregulation between HCC and mid-phase mouse regeneration are related to the cell cycle **(Figure 5.4C)**. Downregulated gene signatures of the HCC model also lead to significant enrichment with mid-phase mouse regeneration but in the opposite direction. Interesting candidate genes include GADD45 family of genes, which are tumour suppressors.

Overall, the early-phase of mouse regeneration (priming or preparatory phase) and human transplantation show consistent up regulation of TNF signalling, NFKB signalling, JAK-STAT signalling and consistent down regulation of metabolic pathways. The mid-phase of mouse liver regeneration (proliferatory phase) is upregulated in cell cycle pathways similar to HCC.

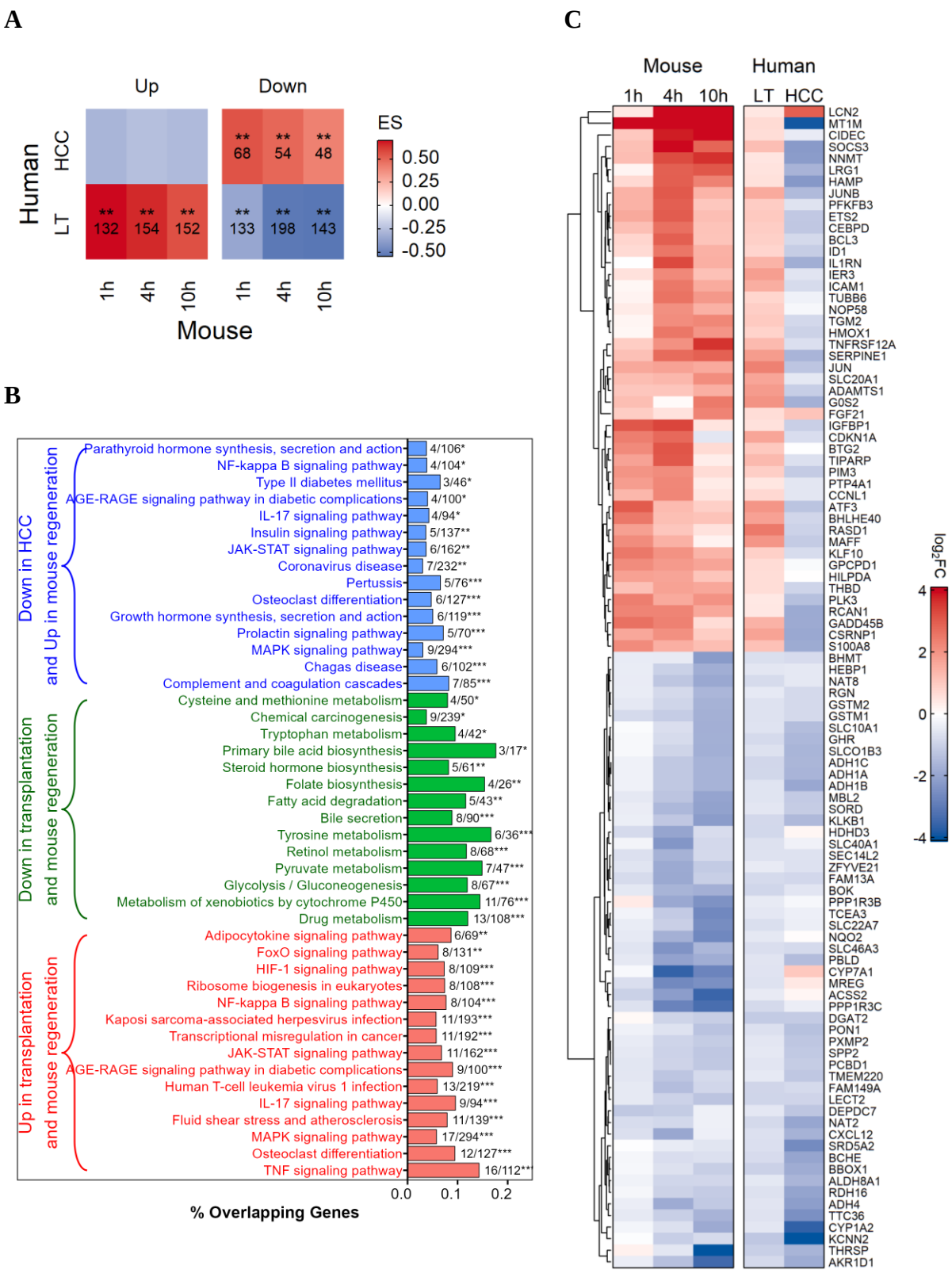


Figure 5.3: Cross-species analysis of transcriptome between human and mouse model (early phase) of regeneration. LT stands for liver transplantation. (A) GSEA showing enrichment score (ES) and the significance of enrichment. Text in the tile indicates the number of leading edges.* indicates adjusted p-value < 0.001, ** indicates 0.001 ≤ adjusted p-value < 0.01, * indicates 0.01 ≤ adjusted p-value < 0.05. (B) KEGG pathway enrichment of leading edges showing 15 most significant pathways sorted according to adjusted p-value (bottom to top within each category). To the right of each bar, the number of overlapping genes and the total**

number of genes of a pathway are shown. (C) Heatmap of top 100 candidates from the union of leading edges (edges with significant ES and consistent direction are only considered).

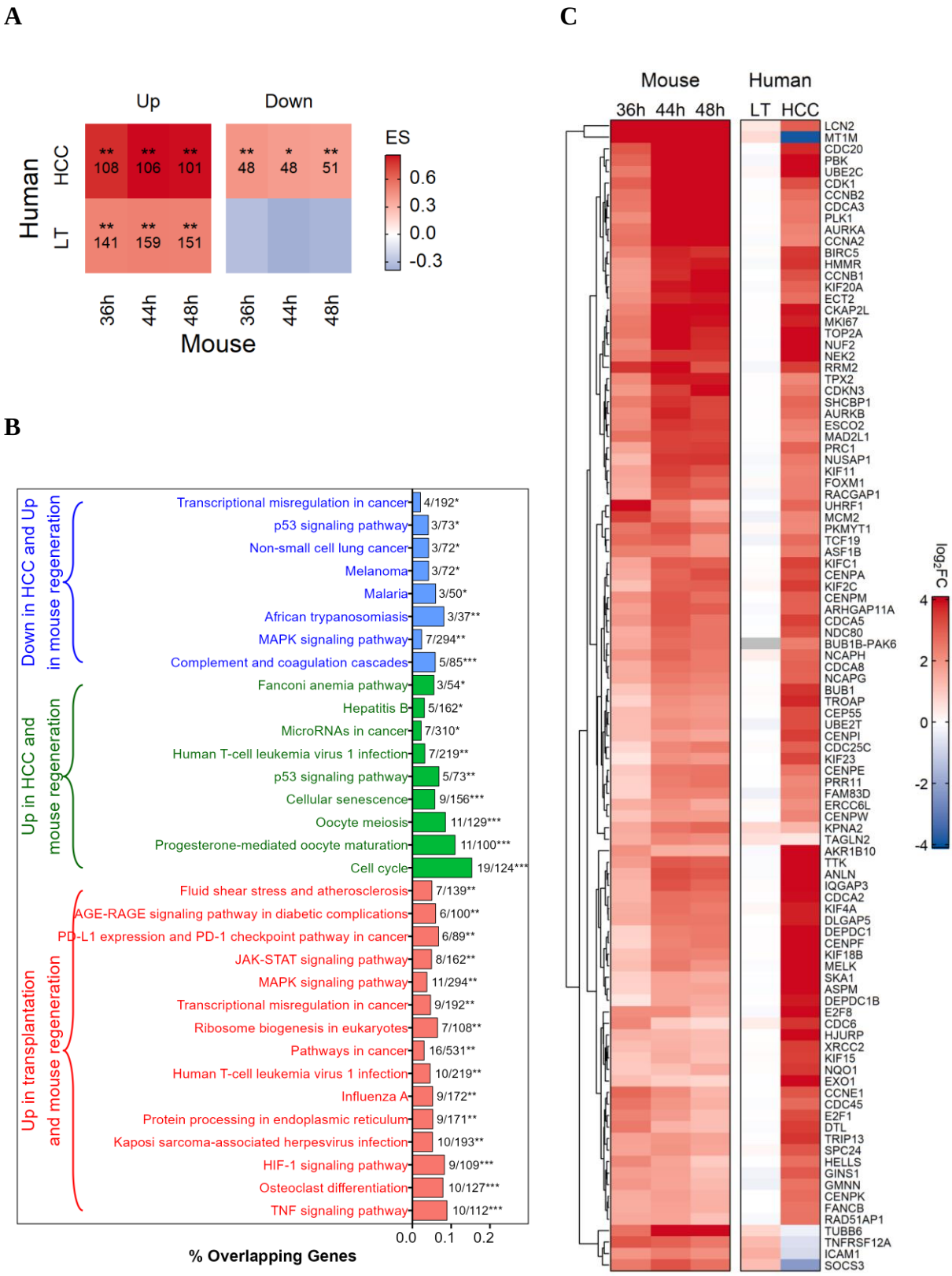


Figure 5.4:Cross-species analysis of transcriptome between human and mouse model (mid-phase) of regeneration. LT stands for liver transplantation. (A) GSEA showing enrichment score (ES) and the significance of enrichment. Text in the tile indicates the number of leading

edges.* indicates adjusted p-value < 0.001, ** indicates 0.001 ≤ adjusted p-value < 0.01, * indicates 0.01 ≤ adjusted p-value < 0.05. (B) KEGG pathway enrichment of leading edges showing 15 most significant pathways sorted according to adjusted p-value are displayed (bottom to top within each category). To the right of each bar, the number of overlapping genes and the total number of genes of a pathway are shown. (C) Heatmap of top 100 candidates from the union of leading edges (edges with significant ES and consistent direction are only considered).**

5.3.2 Gene co-expression differences between normal and rejection samples in post-transplantation

DGCA between normal and rejection samples identified eight modules of size ranging between 498 and 16 genes. Permutation test showed that modules D4, D7 and D8 are significantly rewired in their co-expression pattern. Visualisation of adjacency matrices revealed how different modules change their co-expression pattern in normal and rejection conditions **(Figure 5.5)**.

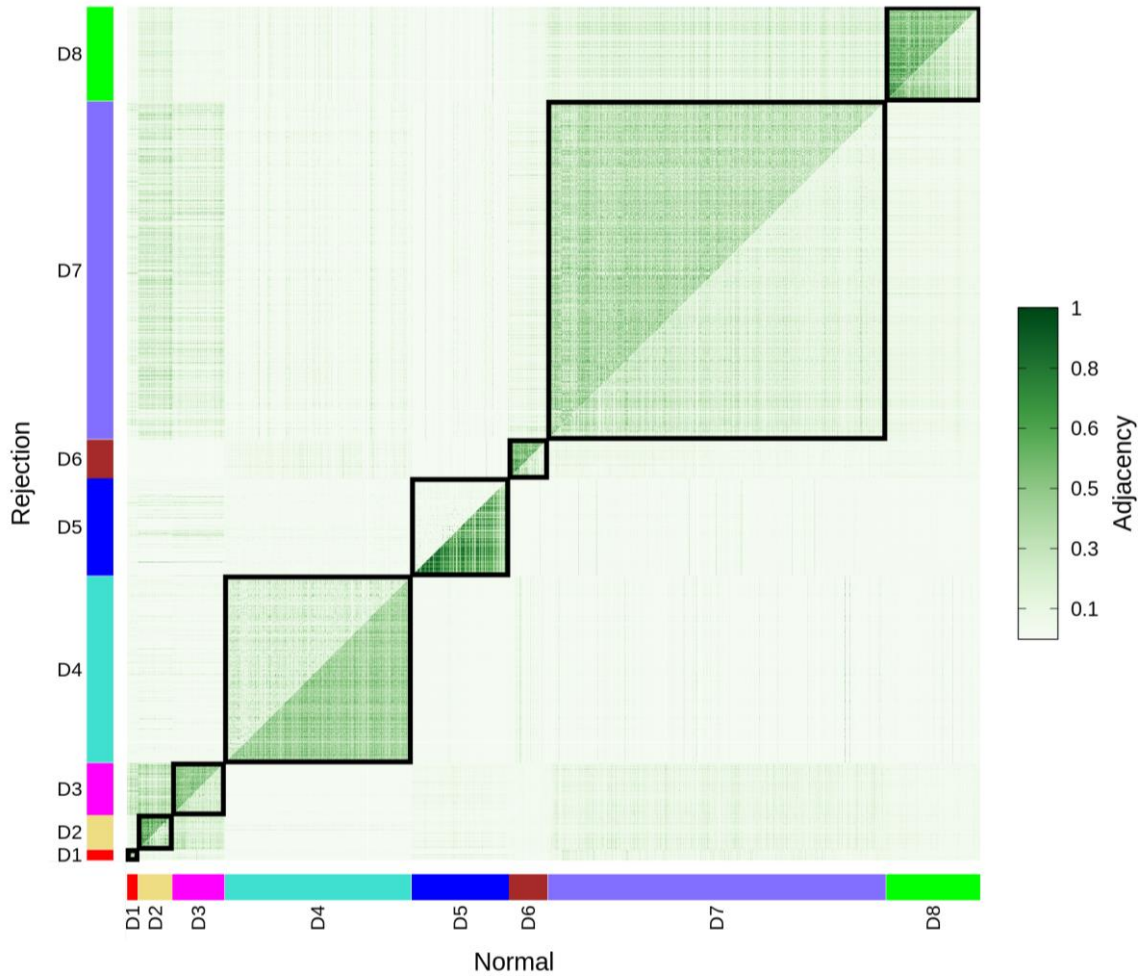


Figure 5.5:Differential gene co-expression analysis of normal and rejection samples. The heatmap showing adjacency matrices of rejection and normal conditions in the upper and lower triangle matrix, respectively.

The genes of modules D7 and D8 are coexpressed under rejection conditions, but their correlation strengths are weak in normal samples. The D4 module shows exactly the opposite behaviour. Pathway analysis of the D4 module shows that it is enriched for pathways specific to amino acid metabolism, complement coagulation cascade, and protein processing **(Figure 5.6A)**. This intact coexpression pattern may be attributed to the long-term stability of normal samples. The D7 module, being the largest and highly coexpressed in rejection samples, is associated with multiple immune pathways such as T-cell receptor signalling, B-cell receptor signalling, and PD-1 checkpoint pathway in cancer. This module comprises a few genes related to allograft rejection (CD86, IFNG, PRF1, GZMB, FASLG and TNF). The D8 module is enriched for cell cycle and DNA replication pathways.

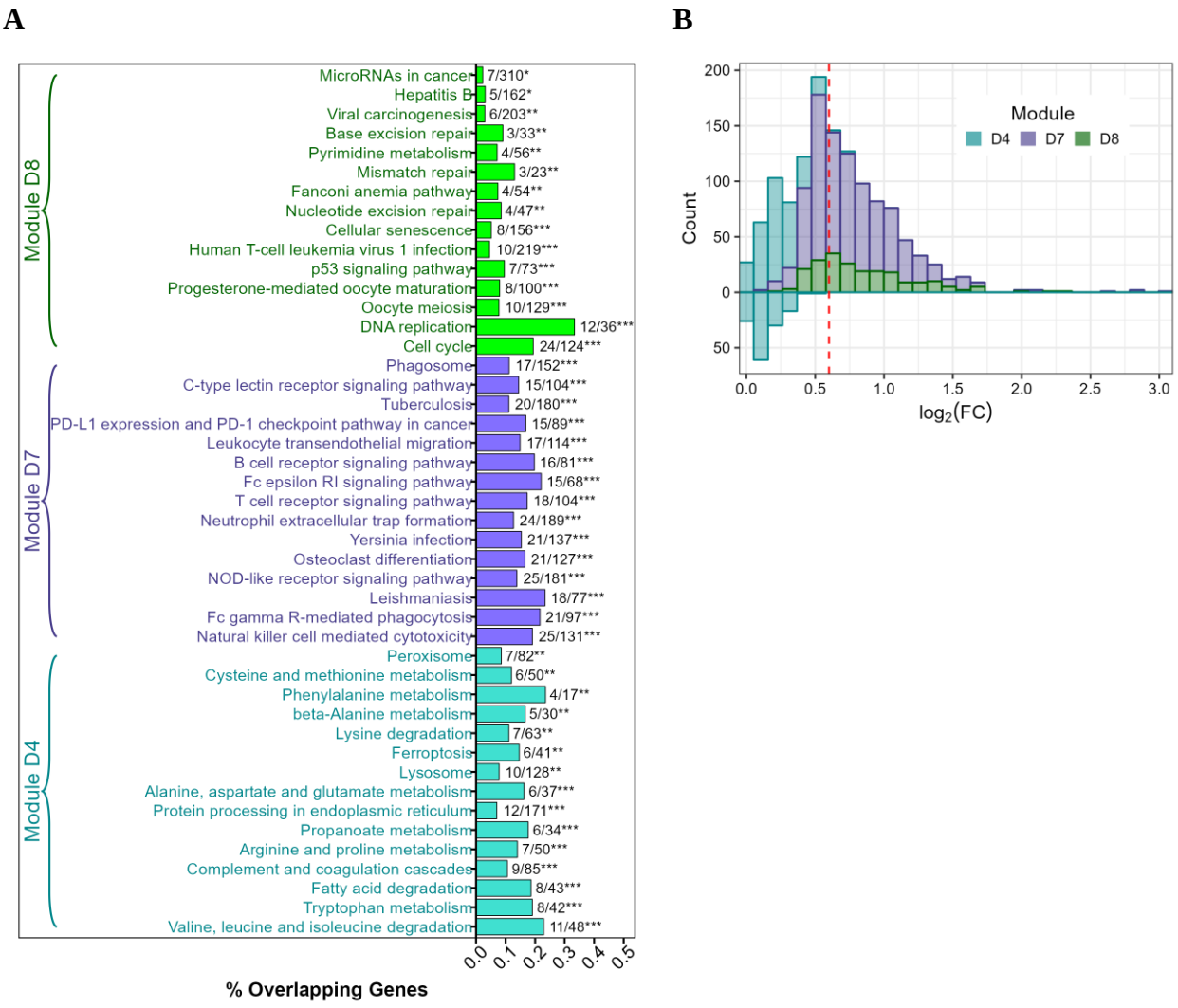


Figure 5.6: Biological pathways of differentially coexpressed modules. (A) KEGG pathway enrichment of modules D4, D7, and D8 genes showing 15 most significant pathways sorted according to adjusted p-value(bottom to top within each module).* indicates adjusted p-value < 0.001, ** indicates 0.001 ≤ adjusted p-value < 0.01, * indicates 0.01 ≤ adjusted p-value < 0.05. To the right of each bar, the number of overlapping genes and the total number of**

genes of a pathway are shown. (B) Histogram of log fold changes of genes in each of the three modules: D4, D7, and D8.

The genes of modules D7 and D8 are not only highly coexpressed in rejection samples but are also highly expressed under rejection conditions **(Figure 5.6B)**. Most genes exhibit large log fold changes greater than 0.6 between two conditions. All the genes in these two modules show positive log fold change. Unlike the other two modules, the D4 module is composed of genes with both positive and negative log fold changes between normal and rejection samples. The D4 module does not show a large difference in expression between normal and rejection samples.

The marked differences in the coexpression pattern of D7 and D8 modules between normal and rejection samples suggest a rewiring of gene expression between the two groups of samples. To confirm if these modules can segregate normal and rejection samples, we applied t-distributed Stochastic Neighbor Embedding (t-SNE) based on the gene expression of the top 20 features (ranked by log fold change) from each module separately. The t-SNE plots show that modules D7 **(Figure 5.7A)** and D8 **(Figure 5.7B)** harbour candidate genes that can distinguish between normal and rejection samples. We also compared how well these two modules overlap with the RATs, previously used marker genes to distinguish different groups of samples[325]. The D7 module shows good overlap with RATs, while the D8 module is unique **(Figure 5.7C and Figure 5.7D)**.

The top 20 features of the D7 module include CXCL11, FCGR1A, ANKRAD22, GBP5, GBP1, BCL2A1 and LILRB2, which are part of the previously reported RATs. In addition to these genes, our analysis identified new features EGR2, MTHFD2, CD52, and CD38 from the D7 module associated with graft rejection. CXCL11 and GBP5 are well-known TCMR-associated genes that are IFNG-induced. CXCL11 is expressed in Th1 effector cells, and it acts against foreign antigens[332]. LILRB2 is known to play a role in immune tolerance[333]. EGR2 acts as a negative regulator of T-cell activation[334] and MTHFD2 is involved in T-cell exhaustion[335]. On the other hand, CD52 and CD38 are glycoproteins expressed on the surface of immune cells and are involved in T cell proliferation and activation[336,337]. Alemtuzumab, an anti-CD52 mAb, is widely used for the treatment of patients undergoing organ transplantation[338]. T cell exhaustion can be due to persistent T cell signalling in rejection cases. We also identified RRM2, TOP2A, CDK1 and CDKN3 as top features from the D8 module associated with TCMR. The cell cycle changes may be related to clonal expansion required for TCMR. We also tested if the combined features of

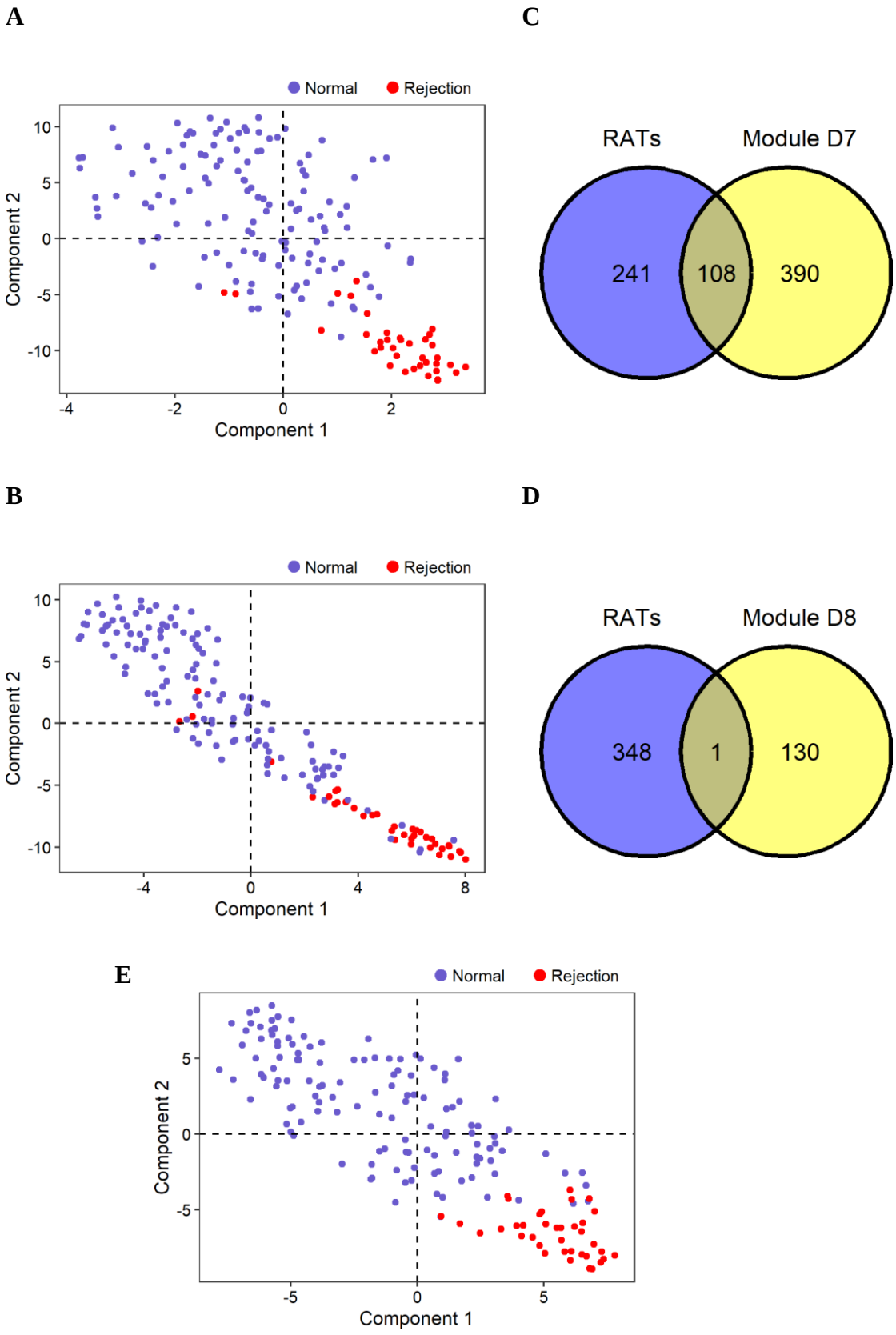


Figure 5.7: Clustering of normal and rejection samples. (A) & (B) t-SNE plot of normal and rejection samples based on the top 20 features of the D7 and D8 modules, respectively. (C) & (D) Venn diagram showing the intersection of RATs with D7 and

D8 modules, respectively. (E) t-SNE plot of normal and rejection samples based on the top 20 combined features of D7 and D8 modules.

the D7 and D8 modules can separate normal and rejection samples. The top 20 combined features comprising CXCL11, FCGR1A, ANKRD22, GBP5, GBP1, BCL2A1, EGR2, MTHFD2, CD52, CSF2RB, LILRB2, and TFEC from D7 module and RRM2, TOP2A, ZWINT, TYMS, MAD2L1, ANLN, PRC1, and CDKN3 from D8 module cluster normal and rejection samples with less overlap **(Figure 5.7E)**.

Overall, the detailed analysis of liver transplantation transcriptomic data provided insights into the cross-species overlap in liver regeneration and also helped to identify features relevant to understanding the difference between normal and rejection samples post-transplantation. We identified combined features of immune and cell cycle modules that can distinguish normal and rejection samples. The efficiency of combined immune and cell cycle features in classifying normal and rejection samples needs to be further validated in future.

Chapter 6 Conclusion

A systems-level investigation of a biological system provides novel insights into how interactions between molecular players coordinate to determine the phenotypic response. The identification of the liver's remarkable regenerative capacity has led to its clinical application as an intervention strategy in the form of partial resection or liver transplantation for hepatocellular carcinoma and other liver diseases. However, the major challenges impeding successful clinical intervention are normal regeneration and physiological function post-treatment. Translating the knowledge gained from molecular mechanisms to liver mass recovery may provide an opportunity to minimise post-treatment failure. This requires a detailed understanding of the molecular pathways and their interactions at the systems level. While successful liver regeneration is the key to attaining normal physiological function following resection or transplantation, the regenerative ability of the liver plays a crucial role in repairing the damage caused by acute or chronic damage under pathological conditions. Further, ageing also impairs the regenerative capacity of the liver and increases the susceptibility to pathological conditions. The work presented in the current thesis attempts to generate systems-level insights into the pathophysiology of the liver using transcriptomic data from experimental mouse models and data from HCC patients.

6.1 Major contributions

We explored how the liver achieves the dynamic balance between various cellular processes, including metabolism and cell cycle during liver regeneration, using mouse liver regeneration transcriptomic data (chapter 2). We showed that the transcriptome is reorganised into three core clusters capturing transient and sustained changes, categorised into specific biological pathways, including immune, cell cycle and metabolic pathways. While the dynamic reconstruction showed an overall downregulation of metabolic trajectories, temporal fine-tuning of specific metabolites such as SAM, GSH and NAD is essential for successful liver regeneration. The temporal expression pattern of main methyltransferases of the liver (GNMT, GAMT) and genes of methionine catabolism (MAT1A and MAT2A) may contribute to an increase in SAM levels in the early phase. This is followed by an upregulation of DNA methylation genes (DNMT1 and UHFR1) in the mid-phase that use SAM as a substrate for epigenetic modifications. Interestingly, DNMT1 and UHRF1 are coexpressed along with cell cycle genes. The temporal separation between

SAM accumulation in the early phase and its utilisation in the mid-phase may lead to crosstalk between the cell cycle and epigenetic regulation during liver regeneration. In addition to DNA methylation genes, PTBP1 and RBMX regulating alternative splicing are also coexpressed with cell cycle genes.

The proposed model of mutual antagonism between regulators of the cell cycle (CYCLIN D) and liver function (HNF4A) provided a mechanistic explanation for changes occurring during liver regeneration. We showed that the liver regeneration program can be seen as changes occurring around the bistable switch accounting for the co-existence of two populations of hepatocytes and the division of labour between them. While one population undergoes active cell division (high CYCLIN D), the other helps maintain liver function (high HNF4A). The model showed the transient inactivation of HNF4A with the dynamic change in the input signal and its reactivation upon successful termination of liver regeneration. The model also showed the condition for an irreversible transition from liver function to proliferation, accounting for the development of HCC. Further, coupling HNF4A with the EMT circuit also showed the emergence of tristability during liver regeneration, which explains the existence of a hybrid cluster of epithelial and mesenchymal cells. We also proposed an integrated circuit of liver regeneration by extending the core HNF4A-CYCLIN D circuit with interactions involving YAP1 and WNT. This extended circuit with additional regulators acting on HNF4A and CYCLIN D may provide underlying features of proliferation, compensatory metabolism and EMT states as observed in single-cell studies.

Chapter 3 showed the ageing-related changes and the crosstalk with liver regeneration and pathological conditions by integrating transcriptomics data from mouse models with the PPI network. The global network entropy analysis of the liver network showed an overall increase in network entropy with ageing. The nodes showing significant changes in local entropy with ageing are associated with complement and coagulation cascade, cytokine–cytokine receptor interaction, membrane remodelling and pathways in cancer. RARRES2 (Chemerin), a gene induced in NAFLD and HBV-related HCC, is captured as the top candidate gene by local entropy analysis. The local entropy analysis also revealed the PGRMC1-INSIGs-SREBF1 axis controlling the lipid levels in ageing. This network entropy analysis provided insights connecting ageing with pathological conditions of fibrosis, NAFLD and HCC. Further, the topological analysis of the ageing network revealed

TRP53, AKT1, SRC, CTNNB1, and EGFR related to cancer signalling as crucial nodes based on their connectivity.

The network proximity analysis was performed to obtain a quantitative picture (distance-based) of the crosstalk between liver ageing and its pathophysiological conditions (liver regeneration, NAFLD, HCC, acute and chronic liver damage). This analysis showed that the older age group is more proximal to pathophysiological conditions than the younger age group. The early phase of liver regeneration is significantly more proximal to the ageing network than the mid-phase, which explains delayed regeneration in ageing. The proximity analysis also revealed a common theme of molecular players connecting ageing and other pathophysiological conditions. This common theme maps to immune-related pathways and MAPK, PI3K-AKT, Ras, Wnt and NFKB signalling pathways. Interesting candidate genes in the common theme include GSK3B and TRP53. The common theme also includes the lipid metabolism genes central to the crosstalk between ageing and related pathophysiological conditions. In summary, our analysis showed that a network-based methodological framework could be used to map system-level changes in liver ageing and dissect the crosstalk between different liver-associated conditions, including regeneration.

To extend the liver regeneration scenario from mouse models to human models, we addressed the molecular pathogenesis of HCC development from the clinical data of patients. Gene coexpression network analysis was performed on tumour samples, tumour-adjacent normal samples (different premalignant stages) and normal samples. Our analysis revealed differences in progression from precancerous to cancer state between two groups of patients undergoing PH and TH (treatment) by targeting the same biological pathways related to cell cycle, immune and metabolism to a different extent. Both groups exhibited a decrease in liver function and immune system and an increase in cell cycle activity in the transition from precancerous to a cancer state, with the PH group showing a higher correlation compared to the TH group. The characterisation of cell cycle genes highlighted the presence of checkpoint signalling, negative regulation of cell cycle processes, and DNA repair processes in the TH group. On the contrary, these cell cycle pathways are not observed in the PH group, which showed positive regulation of cell cycle, cytokinesis, and protein metabolic processes. We also identified modules and genes related to DFS from tumour and premalignant samples. Gene coexpression network analysis of tumour samples showed that upregulation of ribosomal and cell cycle related gene expression and

downregulation of immune and metabolic processes are linked to poor patient survival. We also found multiple modules from premalignant stages predict DFS. Modules associated with cytokine signalling and defence response to virus yielded the best performance in predicting DFS. An increase in the expression of THBD, BCL2L1, FOS and JUN at the premalignant stage is significantly associated with DFS. Overall, our study showed that tumour-matched normal samples contain multiple signatures related to DFS that can be used for early diagnosis.

We also analysed the gene expression profile from patient biopsy samples collected post-liver transplantation, which serves as a human regeneration model. In this context, we performed cross-species analysis and showed the overlap in gene expression profiles between different regeneration models. To further explore the clinical aspect of liver regeneration, we used the transplantation gene expression profile to understand the differences between normal and T-cell-mediated rejection. Differential coexpression analysis showed intact coexpression of immune and cell cycle genes relating to T cell activation, proliferation and exhaustion in rejection samples, which is absent in normal samples. On the other hand, the liver metabolic genes are highly coexpressed in normal samples compared to rejection samples. We also showed that the top features of immune and cell cycle modules are capable of distinguishing normal and rejection samples. Few of these features from the immune module overlapped with rejection-associated transcripts from the heart and kidney. In addition, we identified a few novel candidate genes which need to be further studied and validated.

6.2 Future scope and directions

The work presented in this thesis showed the potential role of systems-level analysis in understanding the interplay of molecular pathways underlying liver regeneration, ageing and other pathological conditions. The conclusions drawn from the study are currently based on transcriptomics data that can be further refined by integrating with multi-omics data. Using the data-driven approach, we reconstructed pathways that formed the basis for proposing the minimalistic circuit to explain the changes occurring during liver regeneration. The model presented in the thesis showed the steady state changes in HNF4A with the regeneration input signal. The HNF4A-CYCLIN D circuit can be extended to build a detailed mechanistic model to capture the dynamics of other players in liver regeneration. Our analysis of HCC transcriptomic data showed co-regulation of various biological

processes: cell cycle, immune and metabolism. A dynamic gene regulatory network to depict the transition from precancerous to cancerous stage in HCC can be developed. This requires understanding the interconnections among the coregulated pathways. Based on existing literature, we proposed a minimalistic explanation of mutual antagonism between HNF4A and immune activity as a molecular mechanism for HCC progression, which needs to be further expanded. With single-cell data becoming increasingly available, the cell-type specific network can also be constructed to understand the heterogeneity at the cellular level. The analysis of HCC presented here is mostly based on the data from the HCC cohort driven by HBV infection. This analysis can be further carried out in NAFLD-driven HCC, given its growing incidence. Our analysis also showed that tumour modules from one cohort are preserved in another cohort of different ethnicities. The data analysed comes from mostly male samples in the old age group and does not have BMI information. Understanding the impact of ethnicity, age, gender, and BMI on liver pathophysiology is important. Further, identifying the molecular features that characterise the differences between normal and rejection in liver transplantation is crucial for developing predictive models. We identified novel features based on differential coexpression, separating the normal and T-cell-mediated rejected samples. Similarly, exploring the effect of ischemic reperfusion injury on post-transplantation liver failure is also clinically relevant. Machine learning approaches can be used for feature extraction and model development to predict liver failure post-transplantation. A validation of such models on multiple datasets is required for clinical application.

References

1. Legallais, C. *et al.* Bioengineering organs for blood detoxification. *Adv Healthcare Materials* **7**, 1800430 (2018).
2. Kampf, C. *et al.* The human liver-specific proteome defined by transcriptomics and antibody-based profiling. *FASEB J* **28**, 2901–2914 (2014).
3. Trefts, E., Gannon, M. & Wasserman, D. H. The liver. *Curr Biol* **27**, R1147–R1151 (2017).
4. Alamri, Z. Z. The role of liver in metabolism: an updated review with physiological emphasis. *Int J Basic Clin Pharmacol* **7**, (2018).
5. Ginguay, A. & Cynober, L. A. Amino Acids | Amino Acid Metabolism. *Encyclopedia of Biological Chemistry: Third Edition* **1**, 2–9 (2021).
6. Poole, L. G., Dolin, C. E. & Arteel, G. E. Organ–organ crosstalk and alcoholic liver disease. *Biomolecules* **7**, 62 (2017).
7. Winkler, S. *et al.* Immune-Deficient Pfp/Rag2−/− Mice Featured Higher Adipose Tissue Mass and Liver Lipid Accumulation with Growing Age than Wildtype C57BL/6N Mice. *Cells* **8**, 775 (2019).
8. Herrlich, A., Kefaloyianni, E. & Rose-John, S. Mechanisms of interorgan crosstalk in health and disease. *FEBS letters* vol. 596 529–533 Preprint at (2022).
9. López-Bermudo, L. *et al.* Contribution of Liver and Pancreatic Islet Crosstalk to β-Cell Function/Dysfunction in the Presence of Fatty Liver. *Front Endocrinol (Lausanne)* **13**, 888 (2022).
10. Hoehme, S. *et al.* Prediction and validation of cell alignment along microvessels as order principle to restore tissue architecture in liver regeneration. *Proc Natl Acad Sci U S A* **107**, 10371–10376 (2010).
11. Teutsch, H. F. The modular microarchitecture of human liver. *Hepatology* **42**, 317–325 (2005).
12. Wild, S. L. *et al.* The Canonical Wnt Pathway as a Key Regulator in Liver Development, Differentiation and Homeostatic Renewal. *Genes (Basel)* **11**, 1–20 (2020).
13. Zaefarian, F., Abdollahi, M. R., Cowieson, A. & Ravindran, V. Avian Liver: The Forgotten Organ. *Animals (Basel)* **9**, 4442 (2019).
14. Iismaa, S. E. *et al.* Comparative regenerative mechanisms across different mammalian tissues. *NPJ Regen Med* **3**, 6 (2018).
15. Fausto, N. Liver regeneration: From laboratory to clinic. *Liver Transplantation* **7**, 835–844 (2001).

16. Stanger, B. Z. Cellular Homeostasis and Repair in the Mammalian Liver. *Annu Rev Physiol* **77**, 179 (2015).

17. Loyer, P. *et al.* Growth Factor Dependence of Progression through G1 and S Phases of Adult Rat Hepatocytes in Vitro: EVIDENCE OF A MITOGEN RESTRICTION POINT IN MID-LATE G1. *Journal of Biological Chemistry* **271**, 11484–11492 (1996).

18. HIGGINS, G. M. Experimental pathology of the liver, Restoration of the liver of the white rat following surgical removal. *Arch Pathol* **12**, 186–202 (1931).

19. Gilgenkrantz, H. & de l'Hortet, A. C. Understanding liver regeneration: from mechanisms to regenerative medicine. *Am J Pathol* **188**, 1316–1327 (2018).

20. Yanger, K. *et al.* Adult hepatocytes are generated by self-duplication rather than stem cell differentiation. *Cell Stem Cell* **15**, 340–349 (2014).

21. Schaub, J. R., Malato, Y., Gormond, C. & Willenbring, H. Evidence against a stem cell origin of new hepatocytes in a common mouse model of chronic liver injury. *Cell Rep* **8**, 933–939 (2014).

22. Michalopoulos, G. K. Novel insights into liver homeostasis and regeneration. *Nat Rev Gastroenterol Hepatol* **18**, 369–370 (2021).

23. Huang, W. *et al.* A narrative review of liver regeneration—from models to molecular basis. *Ann Transl Med* **9**, 1705–1705 (2021).

24. Abu Rmilah, A. *et al.* Understanding the marvels behind liver regeneration. *Wiley Interdiscip Rev Dev Biol* **8**, e340 (2019).

25. Holland, C. H. *et al.* Transcriptomic Cross-Species Analysis of Chronic Liver Disease Reveals Consistent Regulation Between Humans and Mice. *Hepatol Commun* **6**, 161-177 (2021).

26. McCuskey, R. Anatomy of the liver. *Zakim and Boyer's Hepatology: a textbook of liver disease* **6**, 3–19 (2012).

27. Bhushan, B., Gunewardena, S., Edwards, G. & Apte, U. Comparison of liver regeneration after partial hepatectomy and acetaminophen-induced acute liver failure: A global picture based on transcriptome analysis. *Food and Chemical Toxicology* **139**, (2020).

28. Forbes, S. J. & Newsome, P. N. Liver regeneration—mechanisms and models to clinical application. *Nat Rev Gastroenterol Hepatol* **13**, 473–485 (2016).

29. Martins, P. N. A., Theruvath, T. P. & Neuhaus, P. Rodent models of partial hepatectomies. *Liver International* **28**, 3–11 (2008).

30. Stocker, E. Regeneration of liver parenchyma under normal and pathological conditions. *Beitr Path* **144**, 400–408 (1972).

31. Bucher, N. L. R. Regeneration of Mammalian Liver. *Int Rev Cytol* **15**, 245–300 (1963).

32. Michalopoulos, G. K. Liver regeneration. *J Cell Physiol* **213**, 286–300 (2007).

33. Andersen, K. J. *et al.* The natural history of liver regeneration in rats: description of an animal model for liver regeneration studies. *International Journal of Surgery* **11**, 903–908 (2013).

34. Chen, X. G. & Xu, C. S. Proteomic analysis of the regenerating liver following 2/3 partial hepatectomy in rats. *Biol Res* **47**, 1–11 (2014).

35. Haga, S. *et al.* Compensatory recovery of liver mass by Akt-mediated hepatocellular hypertrophy in liver-specific STAT3-deficient mice. *J Hepatol* **43**, 799–807 (2005).

36. Fausto, N., Campbell, J. S. & Riehle, K. J. Liver regeneration. *Hepatology* **43**, S45–S53 (2006).

37. Mao, S. A., Glorioso, J. M. & Nyberg, S. L. Liver regeneration. *Translational Research* **163**, 352–362 (2014).

38. Nejak-Bowen, K. N. & Monga, S. P. S. Developmental Pathways in Liver Regeneration-I. in *Liver Regeneration* 77–101 (Elsevier, 2015).

39. Hata, S., Namae, M. & Nishina, H. Liver development and regeneration: from laboratory study to clinical therapy. *Dev Growth Differ* **49**, 163–170 (2007).

40. Zimmermann, A. Regulation of liver regeneration. *Nephrology Dialysis Transplantation* **19**, iv6–iv10 (2004).

41. Li, W. *et al.* Global changes in interleukin-6-dependent gene expression patterns in mouse livers after partial hepatectomy. *Hepatology* **33**, 1377–1386 (2001).

42. Roy-Chowdhury, N. & Roy-Chowdhury, J. Liver physiology and energy metabolism. in *Sleisenger and Fordtran's Gastrointestinal and Liver Disease-2 Volume Set: Pathophysiology, Diagnosis, Management, Expert Consult Premium Edition-Enhanced Online Features and Print* 1207–1225 (Elsevier, 2010).

43. Mohammed, F. F. & Khokha, R. Thinking outside the cell: proteases regulate hepatocyte division. *Trends Cell Biol* **15**, 555–563 (2005).

44. Huh, C. G. *et al.* Hepatocyte growth factor/c-met signaling pathway is required for efficient liver regeneration and repair. *Proc Natl Acad Sci U S A* **101**, 4477–4482 (2004).

45. Borowiak, M. *et al.* Met provides essential signals for liver regeneration. *Proc Natl Acad Sci U S A* **101**, 10608–10613 (2004).

46. Paranjpe, S. *et al.* Combined systemic elimination of MET and EGFR signaling completely abolishes liver regeneration and leads to liver decompensation. *Hepatology* **64**, 1711 (2016).

47. Russell, J. O. & Monga, S. P. Wnt/β-Catenin Signaling in Liver Development, Homeostasis, and Pathobiology. *Annu Rev Pathol* **13**, 351–378 (2018).

48. Grijalva, J. L. *et al.* Dynamic alterations in Hippo signaling pathway and YAP activation during liver regeneration. *J Physiol Gastrointest Liver Physiol* **307**, 196–204 (2014).

49. Tao, Y., Wang, M., Chen, E. & Tang, H. Liver Regeneration: Analysis of the Main Relevant Signaling Molecules. *Mediators Inflamm* **2017**, (2017).

50. Michalopoulos, G. K. & Bhushan, B. Liver regeneration: biological and pathological mechanisms and implications. *Nat Rev Gastroenterol Hepatol* **18**, 40–55 (2021).

51. Michalopoulos, G. K. Hepatostat: Liver regeneration and normal liver tissue maintenance. *Hepatology* **65**, 1384–1392 (2017).

52. Liu, M. & Chen, P. Proliferation-inhibiting pathways in liver regeneration. *Mol Med Rep* **16**, 23–35 (2017).

53. Braun, L. *et al.* Transforming growth factor beta mRNA increases during liver regeneration: a possible paracrine mechanism of growth regulation. *Proc Natl Acad Sci U S A* **85**, 1539–1543 (1988).

54. Campbell, J. S. *et al.* Expression of suppressors of cytokine signaling during liver regeneration. *J Clin Invest* **107**, 1285–1292 (2001).

55. Luedde, T., Wuestefeld, T. & Trautwein, C. A new player in the team: SOCS-3 socks it to cytokine signaling in the regenerating liver. *Hepatology* **34**, 1254–1256 (2001).

56. Sakuda, S. *et al.* Activation of signal transducer and activator transcription 3 and expression of suppressor of cytokine signal 1 during liver regeneration in rats. *J Hepatol* **36**, 378–384 (2002).

57. Tschuor, C. *et al.* Yes-associated protein promotes early hepatocyte cell cycle progression in regenerating liver after tissue loss. *FASEB Bioadv* **1**, 51–61 (2019).

58. Furchtgott, L. A., Chow, C. C. & Periwal, V. A model of liver regeneration. *Biophys J* **96**, 3926–3935 (2009).

59. Pedone, E. *et al.* Modeling Dynamics and Function of Bone Marrow Cells in Mouse Liver Regeneration. *Cell Rep* **18**, 107 (2017).

60. Young, L. A. H. & Periwal, V. Metabolic scaling predicts posthepatectomy liver regeneration after accounting for hepatocyte hypertrophy. *Liver Transpl* **22**, 476–484 (2016).

61. Cook, D., Achanta, S., Hoek, J. B., Ogunnaike, B. A. & Vadigepalli, R. Cellular network modeling and single cell gene expression analysis reveals novel hepatic stellate cell phenotypes controlling liver regeneration dynamics. *BMC Syst Biol* **12**, 1–29 (2018).

62. Cook, D., Ogunnaike, B. A. & Vadigepalli, R. Systems analysis of non-parenchymal cell modulation of liver repair across multiple regeneration modes. *BMC Syst Biol* **9**, 1–24 (2015).

63. Christ, B. *et al.* Computational modeling in liver surgery. *Front Physiol* **8**, 906 (2017).

64. Zhao, J. *et al.* Identifying the Key Genes in Mouse Liver Regeneration After Partial Hepatectomy by Bioinformatics Analysis and in vitro/vivo Experiments. *Front Genet* **12**, 1083 (2021).

65. Lorenzo, C. S. F., Limm, W. M. L., Lurie, F. & Wong, L. L. Factors affecting outcome in liver resection. *HPB (Oxford)* **7**, 226 (2005).

66. Mann, D. V. *et al.* Human liver regeneration: hepatic energy economy is less efficient when the organ is diseased. *Hepatology* **34**, 557–565 (2001).

67. Verma, A., Manchel, A., Melunis, J., Hengstler, J. G. & Vadigepalli, R. From Seeing to Simulating: A Survey of Imaging Techniques and Spatially-Resolved Data for Developing Multiscale Computational Models of Liver Regeneration. *Frontiers in Systems Biology* **2**, 12 (2022).

68. Verma, B. K., Subramaniam, P. & Vadigepalli, R. Modeling the dynamics of human liver failure post liver resection. *Processes* **6**, (2018).

69. Ray, S., Mehta, N. N., Golhar, A. & Nundy, S. Post hepatectomy liver failure – A comprehensive review of current concepts and controversies. *Annals of Medicine and Surgery* **34**, 4 (2018).

70. Garcea, G. & Maddern, G. J. Liver failure after major hepatic resection. *J Hepatobiliary Pancreat Surg* **16**, 145–155 (2009).

71. Guglielmi, A., Ruzzenente, A., Conci, S., Valdegamberi, A. & Iacono, C. How much remnant is enough in liver resection? *Dig Surg* **29**, 6–17 (2012).

72. Borrello, M. T. & Mann, D. Chronic liver diseases: From development to novel pharmacological therapies: IUPHAR Review 37. *Br J Pharmacol* **180**, 2880–2897 (2023).

73. Kim, H. *et al.* Metabolic Spectrum of Liver Failure in Type 2 Diabetes and Obesity: From NAFLD to NASH to HCC. *Int J Mol Sci* **22**, (2021).

74. Llovet, J. M. *et al.* Hepatocellular carcinoma. *Nat Rev Dis Primers* **7**, 6 (2021).

75. New review elucidates potential contributors to non-alcoholic fatty liver disease. https://medicalxpress.com/news/2022-09-elucidates-potential-contributors-non-alcoholic-fatty.html#google_vignette.

76. What is functional genomics? | Functional genomics I. https://www.ebi.ac.uk/training/online/courses/functional-genomics-i-introduction-and-design/what-is-functional-genomics/.

77. Transcriptome: Connecting the Genome to Gene Function | Learn Science at Scitable. https://www.nature.com/scitable/topicpage/transcriptome-connecting-the-genome-to-gene-function-605/#.

78. Schulze, A. & Downward, J. Navigating gene expression using microarrays? a technology review. *Nat Cell Biol* **3**, (2001).

79. Trevino, V., Falciani, F. & Barrera-Saldaña, H. A. DNA microarrays: A powerful genomic tool for biomedical and clinical research. *Molecular Medicine* **13**, 527–541 (2007).

80. Patterson, T. A. *et al.* Performance comparison of one-color and two-color platforms within the MicroArray Quality Control (MAQC) project. *Nat Biotechnol* **24**, 1140–1150 (2006).

81. Microarrays | Functional genomics II. https://www.ebi.ac.uk/training/online/courses/functional-genomics-ii-common-technologies-and-data-analysis-methods/microarrays/.

82. Wang, Z., Gerstein, M. & Snyder, M. RNA-Seq: a revolutionary tool for transcriptomics. *Nat Rev Genet* **10**, 57–63 (2009).

83. Zeng, W. & Mortazavi, A. Technical considerations for functional sequencing assays. *Nat Immunol* **13**, 802–807 (2012).

84. Sims, D., Sudbery, I., Ilott, N. E., Heger, A. & Ponting, C. P. Sequencing depth and coverage: key considerations in genomic analyses. *Nat Rev Genet* **15**, 121–132 (2014).

85. Malone, J. H. & Oliver, B. Microarrays, deep sequencing and the true measure of the transcriptome. *BMC Biol* **9**, 1–9 (2011).

86. Zhao, S., Fung-Leung, W. P., Bittner, A., Ngo, K. & Liu, X. Comparison of RNA-Seq and microarray in transcriptome profiling of activated T cells. *PLoS One* **9**, (2014).

87. Wang, C. *et al.* The concordance between RNA-seq and microarray data depends on chemical treatment and transcript abundance. *Nat Biotechnol* **32**, 926–932 (2014).

88. Wainberg, M. *et al.* A genome-wide atlas of co-essential modules assigns function to uncharacterized genes. *Nat Genet* **53**, 638 (2021).

89. Mardinoglu, A., Boren, J., Smith, U., Uhlen, M. & Nielsen, J. Systems biology in hepatology: approaches and applications. *Nat Rev Gastroenterol Hepatol* **15**, 365–377 (2018).

90. Kotlyar, M., Wong, S. W. H., Pastrello, C. & Jurisica, I. Improving Analysis and Annotation of Microarray Data with Protein Interactions. *Microarray Data Analysis* 51–68 (2022).

91. Zogopoulos, V. L. *et al.* Approaches in Gene Coexpression Analysis in Eukaryotes. *Biology (Basel)* **11**, 1019 (2022).

92. Kruse, K., Chalancon, G. & Babu, M. M. Metabolic Networks, Applications. in *Encyclopedia of Systems Biology* (eds. Dubitzky, W., Wolkenhauer, O., Cho, K.-H. & Yokota, H.) 1252–1254 (Springer New York, New York, NY, 2013). doi:10.1007/978-1-4419-9863-7_562.

93. Patil, K. R. & Nielsen, J. Uncovering transcriptional regulation of metabolism by using metabolic network topology. *Proc Natl Acad Sci U S A* **102**, 2685–2689 (2005).

94. Becker, T., Beber, M. E., Windt, K., Hütt, M.-T. & Helbing, D. Flow control by periodic devices: a unifying language for the description of traffic, production, and metabolic systems. *Journal of Statistical Mechanics: Theory and Experiment* **2011**, P05004 (2011).

95. Roy, S., Bhattacharyya, D. K. & Kalita, J. K. Reconstruction of gene co-expression network from microarray data using local expression patterns. *BMC Bioinformatics* **15**, S10 (2014).

96. van Dam, S., Võsa, U., van der Graaf, A., Franke, L. & de Magalhães, J. P. Gene co-expression analysis for functional classification and gene–disease predictions. *Brief Bioinform* **19**, 575–592 (2018).

97. Ravasz, E., Somera, A. L., Mongru, D. A., Oltvai, Z. N. & Barabási, A. L. Hierarchical organization of modularity in metabolic networks. *Science* **297**, 1551–1555 (2002).

98. Sefik, E., Purcell, R. H., Walker, E. F., Bassell, G. J. & Mulle, J. G. Convergent and distributed effects of the 3q29 deletion on the human neural transcriptome. *Transl Psychiatry* **11**, 357 (2021).

99. Wang, R. S., Wang, Y., Zhang, X. S. & Chen, L. Inferring transcriptional regulatory networks from high-throughput data. *Bioinformatics* **23**, 3056–3064 (2007).

100. Chen, C. C. & Zhong, S. Inferring gene regulatory networks by thermodynamic modeling. *BMC Genomics* **9**, 1–7 (2008).

101. Karlebach, G. & Shamir, R. Modelling and analysis of gene regulatory networks. *Nat Rev Mol Cell Biol* **9**, 770–780 (2008).

102. Lee, W. P. & Tzou, W. S. Computational methods for discovering gene networks from expression data. *Brief Bioinform* **10**, 408–423 (2009).

103. Shmulevich, I., Dougherty, E. R. & Zhang, W. From Boolean to probabilistic Boolean networks as models of genetic regulatory networks. *Proceedings of the IEEE* **90**, 1778–1792 (2002).

104. Kim, H. D., Shay, T., O'Shea, E. K. & Regev, A. Transcriptional regulatory circuits: predicting numbers from alphabets. *Science (1979)* **325**, 429–432 (2009).

105. Chai, L. E. *et al.* A review on the computational approaches for gene regulatory network construction. *Comput Biol Med* **48**, 55–65 (2014).

106. Liu, Z.-P. Reverse Engineering of Genome-wide Gene Regulatory Networks from Gene Expression Data. *Curr Genomics* **16**, 3 (2015).

107. Albert, R. & Thakar, J. Boolean modeling: a logic-based dynamic approach for understanding signaling and regulatory networks and for making useful predictions. *WIREs Syst Biol Med* **6**, 353–369 (2014).

108. Kaderali, L. & Radde, N. Inferring gene regulatory networks from expression data. *Studies in Computational Intelligence* **94**, 33–74 (2008).

109. Tyson, J. J. & Novák, B. Functional Motifs in Biochemical Reaction Networks. *Annu Rev Phys Chem* **61**, 219 (2010).

110. Mjolsness, E., Sharp, D. H. & Reinitz, J. A connectionist model of development. *J Theor Biol* **152**, 429–453 (1991).

111. Sachs, K., Perez, O., Pe'er, D., Lauffenburger, D. A. & Nolan, G. P. Causal protein-signaling networks derived from multiparameter single-cell data. *Science (1979)* **308**, 523–529 (2005).

112. Luscombe, N. M. *et al.* Genomic analysis of regulatory network dynamics reveals large topological changes. *Nature* **431**, 308–312 (2004).

113. Ernst, J., Vainas, O., Harbison, C. T., Simon, I. & Bar-Joseph, Z. Reconstructing dynamic regulatory maps. *Mol Syst Biol* **3**, 74 (2007).

114. Zhu, S. & Wang, Y. Hidden Markov induced Dynamic Bayesian Network for recovering time evolving gene regulatory networks. *Sci Rep* **5**, 17841 (2015).

115. How to use BNT for DBNs. https://www.cs.ubc.ca/~murphyk/Software/BNT/usage_dbn.html.

116. Murphy, K. P. Dynamic Bayesian Networks: Representation, Inference and Learning, dissertation. *PhD thesis, UC Berkley, Dept. Comp. Sci* (2002).

117. Yoon, B.-J. Hidden Markov Models and their Applications in Biological Sequence Analysis. *Curr Genomics* **10**, 402 (2009).

118. Roy, S. Input Output HMMs for modeling network dynamics Goals for today.

119. Schulz, M. H. *et al.* DREM 2.0: Improved reconstruction of dynamic regulatory networks from time-series expression data. *BMC Syst Biol* **6**, 1–9 (2012).

120. Michalopoulos, G. K. Liver regeneration after partial hepatectomy: critical analysis of mechanistic dilemmas. *Am J Pathol* **176**, 2–13 (2010).

121. Su, A. I., Guidotti, L. G., Pezacki, J. P., Chisari, F. V. & Schultz, P. G. Gene expression during the priming phase of liver regeneration after partial hepatectomy in mice. *Proc Natl Acad Sci U S A* **99**, 11181–11186 (2002).

122. Taub, R. Liver regeneration: from myth to mechanism. *Nat Rev Mol Cell Biol* **5**, 836–847 (2004).

123. Natarajan, A., Wagner, B. & Sibilia, M. The EGF receptor is required for efficient liver regeneration. *Proc Natl Acad Sci U S A* **104**, 17081–17086 (2007).

124. Gkretsi, V. *et al.* Liver-specific ablation of integrin-linked kinase in mice results in abnormal histology, enhanced cell proliferation, and hepatomegaly. *Hepatology* **48**, 1932–1941 (2008).

125. Donthamsetty, S. *et al.* Liver-Specific Ablation of Integrin-Linked Kinase in Mice Results in Enhanced and Prolonged Cell Proliferation and Hepatomegaly after Phenobarbital Administration. *Toxicological Sciences* **113**, 358–366 (2010).

126. Caldez, M. J. *et al.* Metabolic Remodeling during Liver Regeneration. *Dev Cell* **47**, 425-438.e5 (2018).

127. Huang, J. & Rudnick, D. A. Elucidating the Metabolic Regulation of Liver Regeneration. *Am J Pathol* **184**, 309–321 (2014).

128. Rudnick, D. A. & Davidson, N. O. Functional Relationships between Lipid Metabolism and Liver Regeneration. *Int J Hepatol* **2012**, 1–8 (2012).

129. Weymann, A. *et al.* p21 is required for dextrose-mediated inhibition of mouse liver regeneration. *Hepatology* **50**, 207–215 (2009).

130. Newberry, E. P. *et al.* Altered hepatic triglyceride content after partial hepatectomy without impaired liver regeneration in multiple murine genetic models. *Hepatology* **48**, 1097–1105 (2008).

131. Shteyer, E., Liao, Y., Muglia, L. J., Hruz, P. W. & Rudnick, D. A. Disruption of hepatic adipogenesis is associated with impaired liver regeneration in mice. *Hepatology* **40**, 1322–1332 (2004).

132. Naugler, W. E. Bile Acid Flux Is Necessary for Normal Liver Regeneration. *PLoS One* **9**, e97426 (2014).

133. Otu, H. H. *et al.* Restoration of Liver Mass after Injury Requires Proliferative and Not Embryonic Transcriptional Patterns * □ S. *Journal of Biological Chemistry* **282**, 11197–11204 (2007).

134. Singh, P., Goode, T., Dean, A., Awad, S. S. & Darlington, G. J. Elevated Interferon Gamma Signaling Contributes to Impaired Regeneration in the Aged Liver. *The Journals of Gerontology: Series A* **66A**, 944–956 (2011).

135. Pibiri, M. *et al.* Global gene expression profile of normal and regenerating liver in young and old mice. *Age (Omaha)* **37**, 59 (2015).

136. Zimmers, T. A., Jin, X., Zhang, Z., Jiang, Y. & Koniaris, L. G. Epidermal growth factor receptor restoration rescues the fatty liver regeneration in mice. *Am J Physiol Endocrinol Metab* **313**, E440–E449 (2017).

137. Valdecantos, M. P. *et al.* Differential effects of a glucagon-like peptide 1 receptor agonist in non-alcoholic fatty liver disease and in response to hepatectomy. *Sci Rep* **8**, 16461 (2018).

138. Bangru, S. *et al.* Alternative splicing rewires Hippo signaling pathway in hepatocytes to promote liver regeneration. *Nat Struct Mol Biol* **25**, 928–939 (2018).

139. Rib, L. *et al.* Cycles of gene expression and genome response during mammalian tissue regeneration. *Epigenetics Chromatin* **11**, 1–19 (2018).

140. Wang, S. *et al.* Epigenetic Compensation Promotes Liver Regeneration. *Dev Cell* **50**, 43-56.e6 (2019).

141. White, P., Brestelli, J. E., Kaestner, K. H. & Greenbaum, L. E. Identification of transcriptional networks during liver regeneration. *Journal of Biological Chemistry* **280**, 3715–3722 (2005).

142. Caldez, M. J., Bjorklund, M. & Kaldis, P. Cell cycle regulation in NAFLD: when imbalanced metabolism limits cell division. *Hepatol Int* **14**, 463–474 (2020).

143. Halpern, K. B. *et al.* Single-cell spatial reconstruction reveals global division of labour in the mammalian liver. *Nature* **542**, 352–356 (2017).

144. Ben-Moshe, S. & Itzkovitz, S. Spatial heterogeneity in the mammalian liver. *Nat Rev Gastroenterol Hepatol* **16**, 395–410 (2019).

145. Pek, N. M. Q., Liu, K. J., Nichane, M. & Ang, L. T. Controversies Surrounding the Origin of Hepatocytes in Adult Livers and the in Vitro Generation or Propagation of Hepatocytes. *CMGH* **11**, 273–290 (2021).

146. Chen, T., Oh, S., Gregory, S., Shen, X. & Diehl, A. M. Single-cell omics analysis reveals functional diversification of hepatocytes during liver regeneration. *JCI Insight* **5**, (2020).

147. Chembazhi, U. V., Bangru, S., Hernaez, M. & Kalsotra, A. Cellular plasticity balances the metabolic and proliferation dynamics of a regenerating liver. *Genome Res* **31**, 576–591 (2021).

148. Wei, Y. *et al.* Liver homeostasis is maintained by midlobular zone 2 hepatocytes. *Science (1979)* **371**, (2021).

149. He, L. *et al.* Proliferation tracing reveals regional hepatocyte generation in liver homeostasis and repair. *Science (1979)* **371**, (2021).

150. Langfelder, P. & Horvath, S. WGCNA: An R package for weighted correlation network analysis. *BMC Bioinformatics* **9**, 1–13 (2008).

151. Zhang, B. & Horvath, S. A general framework for weighted gene co-expression network analysis. *Stat Appl Genet Mol Biol* **4**, (2005).

152. Kuleshov, M. V. *et al.* Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Res* **44**, W90–W97 (2016).

153. Shannon, P. *et al.* Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res* **13**, 2498–2504 (2003).

154. Zhang, Q., Bhattacharya, S. & Andersen, M. E. Ultrasensitive response motifs: basic amplifiers in molecular signalling networks. *Open Biol* **3**, 130031 (2013).

155. Apte, U. *et al.* Wnt/beta-catenin signaling mediates oval cell response in rodents. *Hepatology* **47**, 288–295 (2008).

156. Miettinen, T. P. *et al.* Identification of transcriptional and metabolic programs related to mammalian cell size. *Curr Biol* **24**, 598–608 (2014).

157. Zhao, Y., Tran, M., Wang, L., Shin, D. J. & Wu, J. PDK4-Deficiency Reprograms Intrahepatic Glucose and Lipid Metabolism to Facilitate Liver Regeneration in Mice. *Hepatol Commun* **4**, 504–517 (2020).

158. Pecqueur, C. *et al.* Uncoupling protein-2 controls proliferation by promoting fatty acid oxidation and limiting glycolysis-derived pyruvate utilization. *The FASEB Journal* **22**, 9–18 (2008).

159. Zhang, H., Liu, H., Zhou, L., Yuen, J. & Forman, H. J. Temporal changes in glutathione biosynthesis during the lipopolysaccharide-induced inflammatory response of THP-1 macrophages. *Free Radic Biol Med* **113**, 304–310 (2017).

160. Lu, S. C. Glutathione synthesis. *Biochimica et Biophysica Acta (BBA) - General Subjects* **1830**, 3143–3153 (2013).

161. Riehle, K. J. *et al.* Sustained glutathione deficiency interferes with the liver response to TNF-α and liver regeneration after partial hepatectomy in mice. *J Liver Disease Transplant* **1**, (2013).

162. Mukherjee, S. *et al.* Nicotinamide adenine dinucleotide biosynthesis promotes liver regeneration. *Hepatology* **65**, 616–630 (2017).

163. Hong, S. *et al.* Nicotinamide N-methyltransferase regulates hepatic nutrient metabolism through Sirt1 protein stabilization. *Nat Med* **21**, 887–894 (2015).

164. Roberti, A., Fernández, A. F. & Fraga, M. F. Nicotinamide N-methyltransferase: At the crossroads between cellular metabolism and epigenetic regulation. *Mol Metab* **45**, 101165 (2021).

165. Komatsu, M. *et al.* NNMT activation can contribute to the development of fatty liver disease by modulating the NAD+ metabolism. *Sci Rep* **8**, 8637 (2018).

166. Tian, B. & Manley, J. L. Alternative polyadenylation of mRNA precursors. *Nat Rev Mol Cell Biol* **18**, 18–30 (2017).

167. Sen, S., Jumaa, H. & Webster, N. J. G. Splicing factor SRSF3 is crucial for hepatocyte differentiation and metabolic function. *Nat Commun* **4**, 1336 (2013).

168. Kumar, D. *et al.* Degradation of splicing factor SRSF3 contributes to progressive liver disease. *J Clin Invest* **129**, 4477–4491 (2019).

169. Wu, P., Zhang, M. & Webster, N. J. G. Alternative RNA Splicing in Fatty Liver Disease. *Front Endocrinol (Lausanne)* **12**, 58 (2021).

170. Lee, S. E., Alcedo, K. P., Kim, H. J. & Snider, N. T. Alternative Splicing in Hepatocellular Carcinoma. *Cell Mol Gastroenterol Hepatol* **10**, 699–712 (2020).

171. Klochendler, A. *et al.* A Transgenic Mouse Marking Live Replicating Cells Reveals In Vivo Transcriptional Program of Proliferation. *Dev Cell* **23**, 681–690 (2012).

172. Walesky, C. *et al.* Hepatocyte-specific deletion of hepatocyte nuclear factor-4α in adult mice results in increased hepatocyte proliferation. *American Journal of Physiology-Gastrointestinal and Liver Physiology* **304**, G26–G37 (2013).

173. Dubois, V. *et al.* Endoplasmic reticulum stress actively suppresses hepatic molecular identity in damaged liver. *Mol Syst Biol* **16**, e9156 (2020).

174. Wu, H. *et al.* A negative reciprocal regulatory axis between cyclin D1 and HNF4α modulates cell cycle progression and metabolism in the liver. *Proc Natl Acad Sci U S A* **117**, 17177–17186 (2020).

175. Hanse, E. A. *et al.* Cyclin D1 inhibits hepatic lipogenesis via repression of carbohydrate response element binding protein and hepatocyte nuclear factor 4α. *Cell Cycle* **11**, 2681 (2012).

176. Sladek, F. M. The yin and yang of proliferation and differentiation: cyclin D1 inhibits differentiation factors ChREBP and HNF4α. *Cell cycle* **11**, 3156 (2012).

177. Walesky, C. & Apte, U. Role of Hepatocyte Nuclear Factor 4α (HNF4α) in Cell Proliferation and Cancer. *Gene Expr* **16**, 101 (2015).

178. Garibaldi, F. *et al.* An epistatic mini-circuitry between the transcription factors Snail and HNF4α controls liver stem cell and hepatocyte features exhorting opposite regulation on stemness-inhibiting microRNAs. *Cell Death Differ* **19**, 937–946 (2012).

179. Jolly, M. K. *et al.* Interconnected feedback loops among ESRP1, HAS2, and CD44 regulate epithelial-mesenchymal plasticity in cancer. APL Bioeng 2: 031908. Preprint at (2018).

180. Mardinoglu, A. *et al.* Personal model-assisted identification of NAD+ and glutathione metabolism as intervention target in NAFLD. *Mol Syst Biol* **13**, 916 (2017).

181. Fazia, M. A. Della & Servillo, G. Foie gras and liver regeneration: a fat dilemma. *Cell Stress* **2**, 162 (2018).

182. Argemí, J. *et al.* X-box Binding Protein 1 Regulates Unfolded Protein, Acute-Phase, and DNA Damage Responses During Regeneration of Mouse Liver. *Gastroenterology* **152**, 1203-1216.e15 (2017).

183. Liu, Y. *et al.* Role for the endoplasmic reticulum stress sensor IRE1α in liver regenerative responses. *J Hepatol* **62**, 590–598 (2015).

184. Hamano, M. *et al.* Lipid overloading during liver regeneration causes delayed hepatocyte DNA replication by increasing ER stress in mice with simple hepatic steatosis. *J Gastroenterol* **49**, 305–316 (2014).

185. Donati, G., Montanaro, L. & Derenzini, M. Ribosome biogenesis and control of cell proliferation: p53 is not alone. *Cancer Res* **72**, 1602–1607 (2012).

186. Walesky, C. M. *et al.* Functional compensation precedes recovery of tissue mass following acute liver injury. *Nat Commun* **11**, 5785 (2020).

187. Huck, I., Gunewardena, S., Espanol-Suner, R., Willenbring, H. & Apte, U. Hepatocyte Nuclear Factor 4 Alpha Activation Is Essential for Termination of Liver Regeneration in Mice. *Hepatology* **70**, 666–681 (2019).

188. Lazarevich, N. L. *et al.* Progression of HCC in mice is associated with a downregulation in the expression of hepatocyte nuclear factors. *Hepatology* **39**, 1038–1047 (2004).

189. Ning, B. F. *et al.* Hepatocyte nuclear factor 4α suppresses the development of hepatocellular carcinoma. *Cancer Res* **70**, 7640–7651 (2010).

190. Lv, D. D., Zhou, L. Y. & Tang, H. Hepatocyte nuclear factor 4α and cancer-related cell signaling pathways: a promising insight into cancer treatment. *Exp Mol Med* **53**, 8–18 (2021).

191. Jia, D. *et al.* Distinguishing mechanisms underlying EMT tristability. *Cancer Converg* **1**, (2017).

192. Noce, V. *et al.* YAP integrates the regulatory Snail/HNF4α circuitry controlling epithelial/hepatocyte differentiation. *Cell Death Dis* **10**, 768 (2019).

193. Yimlamai, D. *et al.* Hippo pathway activity influences liver cell fate. *Cell* **157**, 1324–1338 (2014).

194. Croci, O. *et al.* Transcriptional integration of mitogenic and mechanical signals by Myc and YAP. *Genes Dev* **31**, 2017–2022 (2017).

195. Monga, S. P. S. Role and regulation of β-catenin signaling during physiological liver growth. *Gene Expr* **16**, 51–62 (2014).

196. Yang, M. *et al.* A double-negative feedback loop between Wnt-β-catenin signaling and HNF4α regulates epithelial-mesenchymal transition in hepatocellular carcinoma. *J Cell Sci* **126**, 5692–5703 (2013).

197. Bisso, A. *et al.* Cooperation Between MYC and β-Catenin in Liver Tumorigenesis Requires Yap/Taz. *Hepatology* **72**, 1430–1443 (2020).

198. López-Otín, C., Blasco, M. A., Partridge, L., Serrano, M. & Kroemer, G. The Hallmarks of Aging. *Cell* **153**, 1194–1217 (2013).

199. Sahin, E. *et al.* Telomere dysfunction induces metabolic and mitochondrial compromise. *Nature* **470**, 359–365 (2011).

200. Houtkooper, R. H. *et al.* The metabolic footprint of aging in mice. *Sci Rep* **1**, 134 (2011).

201. Cuervo, A. M. *et al.* Autophagy and Aging: The Importance of Maintaining. *Clean" Cells, Autophagy* **1**, 131–140 (2005).

202. Russell, S. J. & Kahn, C. R. Endocrine regulation of ageing. *Nat Rev Mol Cell Biol* **8**, 681–691 (2007).

203. Harrison, D. E. *et al.* Rapamycin fed late in life extends lifespan in genetically heterogeneous mice. *Nature* **460**, 392–395 (2009).

204. Bergamini, E., Cavallini, G., Donati, A. & Gori, Z. The role of macroautophagy in the ageing process, anti-ageing intervention and age-associated diseases. *Int J Biochem Cell Biol* **36**, 2392–2404 (2004).

205. Shavlakadze, T. *et al.* Age-Related Gene Expression Signature in Rats Demonstrate Early, Late, and Linear Transcriptional Changes from Multiple Tissues. *Cell Rep* **28**, 3263-3273.e3 (2019).

206. Srivastava, A. *et al.* Tissue-specific Gene Expression Changes Are Associated with Aging in Mice. *Genomics Proteomics Bioinformatics* **18**, 430–442 (2020).

207. Schaum, N. *et al.* Ageing hallmarks exhibit organ-specific temporal signatures. *Nature* **583**, 596–602 (2020).

208. Lee, J. S. *et al.* Meta-analysis of gene expression in the mouse liver reveals biomarkers associated with inflammation increased early during aging. *Mech Ageing Dev* **133**, 467–478 (2012).

209. White, R. R. *et al.* Comprehensive transcriptional landscape of aging mouse liver. *BMC Genomics* **16**, 1–15 (2015).

210. Sheedfar, F., Biase, S. Di, Koonen, D. & Vinciguerra, M. Liver diseases and aging: Friends or foes? *Aging Cell* **12**, 950–954 (2013).

211. Uhlén, M. *et al.* Tissue-based map of the human proteome. *Science (1979)* **347**, 1260419 (2015).

212. Franceschi, C., Garagnani, P., Parini, P., Giuliani, C. & Santoro, A. Inflammaging: a new immune–metabolic viewpoint for age-related diseases. *Nat Rev Endocrinol* **14**, 576–590 (2018).

213. Teschendorff, A. E. & Severini, S. Increased entropy of signal transduction in the cancer metastasis phenotype. *BMC Syst Biol* **4**, 1–15 (2010).

214. West, J., Bianconi, G., Severini, S. & Teschendorff, A. E. Differential network entropy reveals cancer system hallmarks. *Sci Rep* **2**, 802 (2012).

215. Menichetti, G., Bianconi, G., Castellani, G., Giampieri, E. & Remondini, D. Multiscale characterization of ageing and cancer progression by a novel network entropy measure. *Mol Biosyst* **11**, 1824–1831 (2015).

216. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* **15**, 1–21 (2014).

217. Chin, C. H. *et al.* cytoHubba: Identifying hub objects and sub-networks from complex interactome. *BMC Syst Biol* **8**, 1–7 (2014).

218. Yu, H., Kim, P. M., Sprecher, E., Trifonov, V. & Gerstein, M. The Importance of Bottlenecks in Protein Networks: Correlation with Gene Essentiality and Expression Dynamics. *PLoS Comput Biol* **3**, e59 (2007).

219. Pržulj, N., Wigle, D. A. & Jurisica, I. Functional topology in a network of protein interactions. *Bioinformatics* **20**, 340–348 (2004).

220. Bader, G. D. & Hogue, C. W. V. An automated method for finding molecular complexes in large protein interaction networks. *BMC Bioinformatics* **4**, 2 (2003).

221. Zhou, Y. *et al.* Network medicine links SARS-CoV-2/COVID-19 infection to brain microvascular injury and neuroinflammation in dementia-like cognitive impairment. *Alzheimers Res Ther* **13**, 110 (2021).

222. Clemens, Z. *et al.* The biphasic and age-dependent impact of klotho on hallmarks of aging and skeletal muscle function. *Elife* **10**, (2021).

223. de Jong, S. *et al.* Implications of genetic variation in the complement system in age-related macular degeneration. *Prog Retin Eye Res* **84**, 100952 (2021).

224. Fuchs, D. *et al.* Tryptophan Metabolism in Inflammaging: From Biomarker to Therapeutic Target. *Frontiers in Immunology | www.frontiersin.org* **10**, 2565 (2019).

225. Schrader, M. *et al.* Metabolic Interplay between Peroxisomes and Other Subcellular Organelles Including Mitochondria and the Endoplasmic Reticulum. *Frontiers in Cell and Developmental Biology | www.frontiersin.org* **1**, 83 (2016).

226. Xu, S., Hou, D., Liu, J. & Ji, L. Age-associated changes in GSH S-transferase gene/proteins in livers of rats. *Redox Report* **23**, 213–218 (2018).

227. Jacenik, D. & Fichna, J. Chemerin in immune response and gastrointestinal pathophysiology. *Clinica Chimica Acta* **504**, 146–153 (2020).

228. Fabian, D. K., Fuentealba, M., Dönertaş, H. M., Partridge, L. & Thornton, J. M. Functional conservation in genes and pathways linking ageing and immunity. *Immunity and Ageing* **18**, 1–18 (2021).

229. Hayden, M. S., West, A. P. & Ghosh, S. NF-κB and the immune response. *Oncogene* **25**, 6758–6780 (2006).

230. Arteel, G. E. & Naba, A. The liver matrisome - looking beyond collagens. *J Hepatol* **2**, 100115 (2020).

231. Kim, I. H. *et al.* Aging increases the susceptibility of hepatic inflammation, liver fibrosis and aging in response to high-fat diet in mice. *Age (Omaha)* **38**, 291–302 (2016).

232. Zheng, R. *et al.* The complement system, aging, and aging-related diseases. *Int J Mol Sci* **23**, 8689 (2022).

233. Lee, S. R. *et al.* Loss of progesterone receptor membrane component 1 promotes hepatic steatosis via the induced de novo lipogenesis. *Sci Rep* **8**, 15711 (2018).

234. Yang, T. *et al.* Crucial Step in Cholesterol Homeostasis: Sterols Promote Binding of SCAP to INSIG-1, a Membrane Protein that Facilitates Retention of SREBPs in ER. *Cell* **110**, 489–500 (2002).

235. Wang, H. *et al.* Glucagon regulates hepatic lipid metabolism via cAMP and Insig-2 signaling: implication for the pathogenesis of hypertriglyceridemia and hepatic steatosis. *Sci Rep* **6**, 32246 (2016).

236. Yecies, J. L. *et al.* Akt Stimulates Hepatic SREBP1c and Lipogenesis through Parallel mTORC1-Dependent and Independent Pathways. *Cell Metab* **14**, 21–32 (2011).

237. Libby, A. E., Bales, E., Orlicky, D. J. & McManaman, J. L. Perilipin-2 Deletion Impairs Hepatic Lipid Accumulation by Interfering with Sterol Regulatory Element-binding Protein (SREBP) Activation and Altering the Hepatic Lipidome. *J Biol Chem* **291**, 24231 (2016).

238. Conte, M., Franceschi, C., Sandri, M. & Salvioli, S. Perilipin 2 and Age-Related Metabolic Diseases: A New Perspective. *Trends Endocrinol Metab* **27**, 893–903 (2016).

239. Najt, C. P. *et al.* Liver-specific loss of Perilipin 2 alleviates diet-induced hepatic steatosis, inflammation, and fibrosis. *Am J Physiol Gastrointest Liver Physiol* **310**, G726 (2016).

240. Lohr, K. *et al.* Reduced mitochondrial mass and function add to age-related susceptibility toward diet-induced fatty liver in C57BL/6J mice. *Physiol Rep* **4**, e12988 (2016).

241. Jin, J. *et al.* Increased Expression of Enzymes of Triglyceride Synthesis Is Essential for the Development of Hepatic Steatosis. *Cell Rep* **3**, 831–843 (2013).

242. Bochkis, I. M., Przybylski, D., Chen, J. & Regev, A. Changes in Nucleosome Occupancy Associated with Metabolic Alterations in Aged Mammalian Liver. *Cell Rep* **9**, 996–1006 (2014).

243. Nakagawa, Y. *et al.* CREB3L3 controls fatty acid oxidation and ketogenesis in synergy with PPARα. *Sci Rep* **6**, 39182 (2016).

244. Tian, J., Goldstein, J. L. & Brown, M. S. Insulin induction of SREBP-1c in rodent liver requires LXRα-c/EBPβ complex. *Proc Natl Acad Sci U S A* **113**, 8182–8187 (2016).

245. Steffen, K. K. & Dillin, A. A Ribosomal Perspective on Proteostasis and Aging. *Cell Metab* **23**, 1004–1012 (2016).

246. Perugorria, M. J. *et al.* Wnt–β-catenin signalling in liver development, health and disease. *Nat Rev Gastroenterol Hepatol* **16**, 121–136 (2019).

247. Wu, D. & Prives, C. Relevance of the p53–MDM2 axis to aging. *Cell Death Differ* **25**, 169–179 (2018).

248. Rodier, F., Campisi, J. & Bhaumik, D. Two faces of p53: aging and tumor suppression. *Nucleic Acids Res* **35**, 7475 (2007).

249. Feng, Z. *et al.* Declining p53 function in the aging process: A possible mechanism for the increased tumor incidence in older populations. *Proc Natl Acad Sci U S A* **104**, 16633–16638 (2007).

250. Sun, X.-X., DeVine, T., Challagundla, K. B. & Dai, M.-S. Interplay between ribosomal protein S27a and MDM2 protein in p53 activation in response to ribosomal stress. *Journal of Biological Chemistry* **286**, 22730–22741 (2011).

251. Wang, H. *et al.* PICT1 is critical for regulating the Rps27a-Mdm2-p53 pathway by microtubule polymerization inhibitor against cervical cancer. *Biochimica et Biophysica Acta (BBA)-Molecular Cell Research* **1868**, 119084 (2021).

252. Anisimova, A. S. *et al.* Multifaceted deregulation of gene expression and protein synthesis with age. *Proc Natl Acad Sci U S A* **117**, 15581–15590 (2020).

253. De L'Hortet, A. C. *et al.* GH Administration Rescues Fatty Liver Regeneration Impairment by Restoring GH/EGFR Pathway Deficiency. *Endocrinology* **155**, 2545–2554 (2014).

254. Wink, L., Miller, R. A. & Garcia, G. G. Rapamycin, Acarbose and 17α-estradiol share common mechanisms regulating the MAPK pathways involved in intracellular signaling and inflammation. *Immunity and Ageing* **19**, 1–20 (2022).

255. Siddiqui, N. & Sonenberg, N. Signalling to eIF4E in cancer. *Biochem Soc Trans* **43**, 763–772 (2015).

256. Jin, J., Wang, G. L., Timchenko, L. & Timchenko, N. A. GSK3β and aging liver. *Aging* **1**, 582–585 (2009).

257. Jin, J., Wang, G.-L., Shi, X., Darlington, G. J. & Timchenko, N. A. The Age-Associated Decline of Glycogen Synthase Kinase 3β Plays a Critical Role in the Inhibition of Liver Regeneration. *Mol Cell Biol* **29**, 3867–3880 (2009).

258. Emma, M. R. *et al.* GSK-3 in liver diseases: Friend or foe? *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research* **1867**, 118743 (2020).

259. Maeda, S., Kamata, H., Luo, J. L., Leffert, H. & Karin, M. IKKβ Couples Hepatocyte Death to Cytokine-Driven Compensatory Proliferation that Promotes Chemical Hepatocarcinogenesis. *Cell* **121**, 977–990 (2005).

260. Fausto, N. Mouse liver tumorigenesis: models, mechanisms, and relevance to human disease. *Semin Liver Dis* **19**, 243–252 (1999).

261. Iimuro, Y. & Fujimoto, J. TLRs, NF-B, JNK, and liver regeneration. *Gastroenterol Res Pract* **2010**, (2010).

262. Zhang, L. *et al.* mTORC2 Facilitates Liver Regeneration Through Sphingolipid-Induced PPAR-α-Fatty Acid Oxidation. *Cell Mol Gastroenterol Hepatol* **14**, 1311–1331 (2022).

263. DeAngelis, R. A., Markiewski, M. M., Taub, R. & Lambris, J. D. A high-fat diet impairs liver regeneration in C57BL/6 mice through overexpression of the NF-kappaB inhibitor, IkappaBalpha. *Hepatology* **42**, 1148–1157 (2005).

264. Fafián-Labora, J. *et al.* FASN activity is important for the initial stages of the induction of senescence. *Cell Death Dis* **10**, 318 (2019).

265. Suh, E., Choi, S.-W. & Friso, S. One-carbon metabolism: an unsung hero for healthy aging. in *Molecular basis of nutrition and aging* 513–522 (Elsevier, 2016).

266. Lionaki, E., Ploumi, C. & Tavernarakis, N. One-carbon metabolism: pulling the strings behind aging and neurodegeneration. *Cells* **11**, 214 (2022).

267. Rosenzweig, A., Blenis, J. & Gomes, A. P. Beyond the warburg effect: How do cancer cells regulate One-Carbon metabolism? *Front Cell Dev Biol* **6**, 90 (2018).

268. Da Silva, R. P., Eudy, B. J. & Deminice, R. One-Carbon Metabolism in Fatty Liver Disease and Fibrosis: One-Carbon to Rule Them All. *J Nutr* **150**, 994–1003 (2020).

269. Radziejewska, A., Muzsik, A., Milagro, F. I., Martínez, J. A. & Chmurzynska, A. One-Carbon Metabolism and Nonalcoholic Fatty Liver Disease: The Crosstalk between Nutrients, Microbiota, and Genetics. *Lifestyle Genom* **13**, 53–63 (2020).

270. Wilson, C. H. *et al.* Age-related proteostasis and metabolic alterations in Caspase-2-deficient mice. *Cell Death Dis* **6**, e1615–e1615 (2015).

271. Ericksen, R. E. *et al.* Loss of BCAA Catabolism during Carcinogenesis Enhances mTORC1 Activity and Promotes Tumor Development and Progression. *Cell Metab* **29**, 1151-1165.e6 (2019).

272. Jiang, C. *et al.* Comparative transcriptomics analyses in livers of mice, humans, and humanized mice define human-specific gene networks. *Cells* **9**, 2566 (2020).

273. Villanueva, A. Hepatocellular Carcinoma. *New England Journal of Medicine* **380**, 1450–1462 (2019).

274. Sung, H. *et al.* Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA Cancer J Clin* **71**, 209–249 (2021).

275. Llovet, J. M. *et al.* Hepatocellular carcinoma. *Nat Rev Dis Primers* **7**, 6 (2021).

276. Hepatocellular carcinoma. *Nat Rev Dis Primers* **7**, 7 (2021).

277. Leowattana, W., Leowattana, T. & Leowattana, P. T. Systemic treatment for unresectable hepatocellular carcinoma. *World J Gastroenterol* **29**, 1551 (2023).

278. Ho, D. W. H., Lo, R. C. L., Chan, L. K. & Ng, I. O. L. Molecular Pathogenesis of Hepatocellular Carcinoma. *Liver Cancer* **5**, 290 (2016).

279. Rico Montanari, N., Anugwom, C. M., Boonstra, A. & Debes, J. D. The role of cytokines in the different stages of hepatocellular carcinoma. *Cancers (Basel)* **13**, 4876 (2021).

280. Desai, A., Sandhu, S., Lai, J. P. & Sandhu, D. S. Hepatocellular carcinoma in non-cirrhotic liver: A comprehensive review. *World J Hepatol* **11**, 1 (2019).

281. Schütte, K. *et al.* Characterization and prognosis of patients with hepatocellular carcinoma (HCC) in the non-cirrhotic liver. *BMC Gastroenterol* **14**, 117 (2014).

282. Yoo, S. *et al.* A pilot systematic genomic comparison of recurrence risks of hepatitis B virus-associated hepatocellular carcinoma with low- and high-degree liver fibrosis. *BMC Med* **15**, 214 (2017).

283. Kow, A. W. C. Transplantation versus liver resection in patients with hepatocellular carcinoma. *Transl Gastroenterol Hepatol* **4**,33 (2019).

284. Sangro, B., Sarobe, P., Hervás-Stubbs, S. & Melero, I. Advances in immunotherapy for hepatocellular carcinoma. *Nat Rev Gastroenterol Hepatol* **18**, 525–543 (2021).

285. Maass, T. *et al.* Microarray-Based Gene Expression Analysis of Hepatocellular Carcinoma. *Curr Genomics* **11**, 261 (2010).

286. Alqahtani, A. *et al.* Hepatocellular Carcinoma: Molecular Mechanisms and Targeted Therapies. *Medicina 2019, Vol. 55, Page 526* **55**, 526 (2019).

287. Teufel, A. *et al.* Hepatocellular carcinoma in patients with autoimmune hepatitis. *World Journal of Gastroenterology : WJG* **15**, 578 (2009).

288. Ally, A. *et al.* Comprehensive and Integrative Genomic Characterization of Hepatocellular Carcinoma. *Cell* **169**, 1327-1341.e23 (2017).

289. Badwei, N. Hepatocellular Carcinoma Gene Expression: The New Era, Where It goes? *iLIVER* **2**, 36–40 (2023).

290. Wu, Y., Liu, Z. & Xu, X. Molecular subtyping of hepatocellular carcinoma: A step toward precision medicine. *Cancer Commun* **40**, 681 (2020).

291. Son, J. A. *et al.* Novel Gene Signatures as Prognostic Biomarkers for Predicting the Recurrence of Hepatocellular Carcinoma. *Cancers (Basel)* **14**, 865 (2022).

292. Huang, X., Stern, D. F. & Zhao, H. Transcriptional profiles from paired normal samples offer complementary information on cancer patient survival–evidence from TCGA pan-cancer data. *Sci Rep* **6**, 20567 (2016).

293. Hoshida, Y. *et al.* Gene Expression in Fixed Tissues and Outcome in Hepatocellular Carcinoma. *New England Journal of Medicine* **359**, 1995–2004 (2008).

294. Aran, D. *et al.* Comprehensive analysis of normal adjacent to tumor transcriptomes. *Nat Commun* **8**, 1077 (2017).

295. Galindez, G., Sadegh, S., Baumbach, J., Kacprowski, T. & List, M. Network-based approaches for modeling disease regulation and progression. *Comput Struct Biotechnol J* **21**, 780–795 (2023).

296. Therneau, T. A package for survival analysis in R. (2023).

297. Langfelder, P., Luo, R., Oldham, M. C. & Horvath, S. Is My Network Module Preserved and Reproducible? *PLoS Comput Biol* **7**, e1001057 (2011).

298. Bindea, G. *et al.* ClueGO: a Cytoscape plug-in to decipher functionally grouped gene ontology and pathway annotation networks. *Bioinformatics* **25**, 1091 (2009).

299. Yoon, S.-H., Choi, S.-W., Nam, S. W., Lee, K. B. & Nam, J.-W. Preoperative immune landscape predisposes adverse outcomes in hepatocellular carcinoma patients with liver transplantation. *NPJ Precis Oncol* **5**, 27 (2021).

300. Sun, H. *et al.* Association of decreased expression of the macrophage scavenger receptor MARCO with tumor progression and poor prognosis in human hepatocellular carcinoma. *J Gastroenterol Hepatol* **32**, 1107–1114 (2017).

301. Liang, X. *et al.* CLEC1B is a Promising Prognostic Biomarker and Correlated with Immune Infiltration in Hepatocellular Carcinoma. *Int J Gen Med* **15**, 5661 (2022).

302. Qian, X., Yang, Z., Gao, L., Liu, Y. & Yan, J. The role of complement in the clinical course of hepatocellular carcinoma. *Immun Inflamm Dis* **10**, e569 (2022).

303. Kim, J. H. *et al.* Genomic predictors for recurrence patterns of hepatocellular carcinoma: model derivation and validation. *PLoS Med* **11**, (2014).

304. Sun, Y. *et al.* Single-cell landscape of the ecosystem in early-relapse hepatocellular carcinoma. *Cell* **184**, 404-421.e16 (2021).

305. Zhang, N. *et al.* Molecular mechanism of the miR-7/BCL2L1/P53 signaling axis regulating the progression of hepatocellular carcinoma. *Ann Transl Med* **11**, 12 (2023).

306. Porukala, M. & Vinod, P. K. Systems-level analysis of transcriptome reorganization during liver regeneration. *Mol Omics* **18**, 315–327 (2022).

307. Xie, X., Guo, P., Yu, H., Wang, Y. & Chen, G. Ribosomal proteins: insight into molecular roles and functions in hepatocellular carcinoma. *Oncogene* **37**, 277–285 (2018).

308. Hatziapostolou, M. *et al.* An HNF4α-miRNA inflammatory feedback circuit regulates hepatocellular oncogenesis. *Cell* **147**, 1233–1247 (2011).

309. Goldman, O. *et al.* Early Infiltration of Innate Immune Cells to the Liver Depletes HNF4α and Promotes Extrahepatic Carcinogenesis. *Cancer Discov* OF1–OF20 (2023).

310. Yang, T. *et al.* Therapeutic HNF4A mRNA attenuates liver fibrosis in a preclinical model. *J Hepatol* **75**, 1420–1433 (2021).

311. Gunewardena, S. *et al.* Progressive loss of hepatocyte nuclear factor 4 alpha activity in chronic liver diseases in humans. *Hepatology* **76**, 372–386 (2022).

312. Carter, S. L., Eklund, A. C., Kohane, I. S., Harris, L. N. & Szallasi, Z. A signature of chromosomal instability inferred from gene expression profiles predicts clinical outcome in multiple human cancers. *Nat Genet* **38**, 1043–1048 (2006).

313. Zhang, W. *et al.* Centromere and kinetochore gene misexpression predicts cancer patient survival and response to radiotherapy and chemotherapy. *Nat Commun* **7**, 12619 (2016).

314. Yagi, S., Hirata, M., Miyachi, Y. & Uemoto, S. Liver Regeneration after Hepatectomy and Partial Liver Transplantation. *Int J Mol Sci* **21**, 1–23 (2020).

315. Young, L. A. H. & Periwal, V. Metabolic scaling predicts posthepatectomy liver regeneration after accounting for hepatocyte hypertrophy. *Liver Transpl* **22**, 476–484 (2016).

316. Olthoff, K. M. *et al.* Liver regeneration after living donor transplant: Adult to Adult Living Donor Liver Transplantation Cohort (A2ALL) Study. *Liver Transpl* **21**, 79 (2015).

317. Gurusamy, K. S., Naik, P., Abu-Amara, M., Fuller, B. & Davidson, B. R. Techniques of flushing and reperfusion for liver transplantation. *Cochrane Database of Systematic Reviews* (2012).

318. Lee, B. *et al.* Hepatic CEACAM1 expression indicates donor liver quality and prevents early transplantation injury. *J Clin Invest* **130**, 2689 (2020).

319. Dar, W. A., Sullivan, E., Bynon, J. S., Eltzschig, H. & Ju, C. Ischaemia reperfusion injury in liver transplantation: Cellular and molecular mechanisms. *Liver International* **39**, 788–801 (2019).

320. Benseler, V. *et al.* The liver: a special case in transplantation tolerance. *Semin Liver Dis* **27**, 194–213 (2007).

321. Thomson, A. W., Vionnet, J. & Sanchez-Fueyo, A. Understanding, predicting and achieving liver transplant tolerance: from bench to bedside. *Nat Rev Gastroenterol Hepatol* **17**, 719–739 (2020).

322. Clavien, P. A., Muller, X., de Oliveira, M. L., Dutkowski, P. & Sanchez-Fueyo, A. Can immunosuppression be stopped after liver transplantation? *Lancet Gastroenterol Hepatol* **2**, 531–537 (2017).

323. Berberat, P. O. *et al.* Differentially expressed genes in postperfusion biopsies predict early graft dysfunction after liver transplantation. *Transplantation* **82**, 699–704 (2006).

324. De Jonge, J. *et al.* Unique Early Gene Expression Patterns in Human Adult-to-Adult Living Donor Liver Grafts Compared to Deceased Donor Grafts. *American Journal of Transplantation* **9**, 758–772 (2009).

325. Madill-Thomsen, K. *et al.* The molecular diagnosis of rejection in liver transplant biopsies: First results of the INTERLIVER study. *American Journal of Transplantation* **20**, 2156–2172 (2020).

326. Halloran, P. F., Venner, J. M. & Famulski, K. S. Comprehensive Analysis of Transcript Changes Associated With Allograft Rejection: Combining Universal and Selective Features. *American Journal of Transplantation* **17**, 1754–1769 (2017).

327. Korotkevich, G. *et al.* Fast gene set enrichment analysis. *bioRxiv* 060012 (2021) doi:10.1101/060012.

328. Subramanian, A. *et al.* Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A* **102**, 15545–15550 (2005).

329. Kostka, D. & Spang, R. Finding disease specific alterations in the co-expression of genes. *Bioinformatics* **20**, i194–i199 (2004).

330. Tesson, B. M., Breitling, R. & Jansen, R. C. DiffCoEx: A simple and sensitive method to find differentially coexpressed gene modules. *BMC Bioinformatics* **11**, 1–9 (2010).

331. Choi, Y. & Kendziorski, C. Statistical methods for gene set co-expression analysis. *Bioinformatics* **25**, 2780–2786 (2009).

332. Dou, M. *et al.* Immune-Related Genes for Predicting Future Kidney Graft Loss: A Study Based on GEO Database. *Front Immunol* **13**, 859693 (2022).

333. Li, Q. & Lan, P. Activation of immune signals during organ transplantation. *Signal Transduct Target Ther* **8**, 110 (2023).

334. Safford, M. *et al.* Egr-2 and Egr-3 are negative regulators of T cell activation. *Nat Immunol* **6**, 472–480 (2005).

335. Hung, M. H. *et al.* Tumor methionine metabolism drives T-cell exhaustion in hepatocellular carcinoma. *Nat Commun* **12**, 1455 (2021).

336. Ramouz, A. *et al.* Preoperative CD52 Level Predicts Graft Survival following Kidney Transplantation. *Biomed Res Int* **2022**, 1–7 (2022).

337. Piedra-Quintero, Z. L., Wilson, Z., Nava, P. & Guerau-de-Arellano, M. CD38: An Immunomodulatory Molecule in Inflammation and Autoimmunity. *Front Immunol* **11**, 597959 (2020).

338. Zhao, Y. *et al.* The immunological function of CD52 and its targeting in organ transplantation. *Inflamm Res* **66**, 571–578 (2017).

Appendix A

Table A1: Significant KEGG pathways and biological processes (adj p-value < 0.05) associated with WGCNA modules. * represents uncorrected p-value.

Modules	KEGG pathways	Biological processes
Blue (M1)	DNA replication (6.02e-24), Cell cycle (1.02e-24), Homologous recombination (8.12e-11), Mismatch repair (2.26e-01), Cellular senescence (2.2e-07), Spliceosome (3.7e-07), Nucleotide excision repair (4.2e-06), p53 signaling pathway (8.7e-06), Pyrimidine metabolism (0.001), RNA transport (0.001), Glutathione metabolism (0.008), Steroid biosynthesis (0.01), Amino sugar and nucleotide sugar metabolism (0.01), Arginine and proline metabolism (0.01), Protein processing in endoplasmic reticulum (0.02)	DNA metabolic process (1.7e-31), mitotic cell cycle phase transition (1.85e-25), DNA replication (2.1e-23), G1/S transition of mitotic cell cycle (1.2e-17), DNA repair (2.1e-17), cellular response to DNA damage stimulus (8.2e-13), tRNA export from nucleus (6.8e-07), RNA splicing, via transesterification reactions with bulged adenosine as nucleophile (7.9e-07)
Green (M2)	Complement and coagulation cascades* (4e-04), Fatty acid degradation* (5e-03), Protein export* (0.002), PPAR signaling pathway* (0.008)	Platelet degranulation (5.2e-04), acute inflammatory response(0.001), positive regulation of protein secretion (0.01), cargo loading into COPII-coated vesicle (0.01), homotypic cell-cell adhesion (0.02), Golgi vesicle transport (0.02), fatty acid beta-oxidation* (0.003)
Red (M3)	Protein processing in endoplasmic reticulum (1.8e-17), Protein export (1.3e-06), N-Glycan biosynthesis (1e-04), Amino sugar and nucleotide sugar metabolism (0.01)	IRE1-mediated unfolded protein response (5.7e-13), protein N-linked glycosylation (3.7e-10), ERAD pathway (9.4e-06), ubiquitin-dependent ERAD pathway (6.9e-05), ATF6-mediated unfolded protein response (0.03)
Pink (M4)	TNF signaling pathway (0.01), Ribosome biogenesis in eukaryotes* (3e-04), JAK-STAT signaling pathway* (0.01)	Ribosome biogenesis (4.1e-07), rRNA processing (3.7e-04), maturation of 5.8S rRNA(0.001), protein localization to nucleus (0.001), ribosome assembly (0.001), response to cytokine (0.02), interleukin-6-mediated signaling pathway* (0.001)
Purple (M5)	PPAR signaling pathway (4.03e-08), Fatty acid degradation (1e-04), Peroxisome (0.001), Biosynthesis of unsaturated fatty acids* (0.001), TGF-beta signaling pathway* (0.002)	Regulation of lipid metabolic process (5.65e-07), fatty acid beta-oxidation (0.002), positive regulation of pathway-restricted SMAD protein phosphorylation* (0.003), peroxisomal membrane transport* (0.03)
Black (M6)	Steroid hormone biosynthesis (8.8e-12), PPAR signaling pathway (7.1e-09), Biosynthesis of unsaturated fatty acids (2.4e-05), Primary bile acid biosynthesis (1.4e-04), Arachidonic acid metabolism (3.6e-04), Cholesterol metabolism (3.2e-04), Retinol metabolism (3.1e-04), Tryptophan metabolism (0.002), Glutathione metabolism (0.01), Bile secretion (0.01), Fatty acid degradation (0.01), Complement and coagulation cascades (0.03), Arginine biosynthesis (0.04)	Alpha-amino acid catabolic process (8.3e-06), aromatic amino acid family catabolic process (2.7e-04), tryptophan metabolic process (0.002), tryptophan catabolic process (0.002), cholesterol homeostasis (0.004), sterol metabolic process (0.004), valine metabolic process (0.01)
Brown (M8)	Chemical carcinogenesis (2.2e-15), Steroid hormone biosynthesis (5.7e-14), Retinol metabolism (6.3e-14), Glutathione metabolism (2.2e-05), Pentose and glucuronate interconversions (6.03e-05), Tryptophan metabolism (9.1e-05), Glycine, serine and threonine metabolism (1.6e-04), Valine, leucine and isoleucine degradation (0.001), Glyoxylate and dicarboxylate metabolism (0.002), Histidine metabolism (0.005), Cysteine and methionine metabolism (0.02), Arachidonic acid metabolism (0.02), Phenylalanine metabolism (0.03), Tyrosine metabolism (0.04)	Alpha-amino acid catabolic process (3.8e-06), glutathione metabolic process (0.001), branched-chain amino acid metabolic process (0.002), histidine metabolic process (0.003)

Magenta (M9)	Retinol metabolism (2.3e-04), Glyoxylate and dicarboxylate metabolism (7e-04), Glycine, serine and threonine metabolism (0.03), One carbon pool by folate (0.02), Tryptophan metabolism (0.03), Complement and coagulation cascades (0.03), Terpenoid backbone biosynthesis (0.03), Folate biosynthesis (0.03)	Response to sterol (0.02), regulation of cholesterol biosynthetic process (0.02), serine family amino acid metabolic process (0.02), coenzyme metabolic process*(0.001)

Table A2: Significant KEGG pathways and biological processes (adj p-value < 0.05) associated with different paths in validation dataset. * represents uncorrected p-value.

Path	KEGG pathway	Biological processes
A	Cell cycle (4.5e-14), p53 signaling pathway (0.006), Cellular senescence (0.006), DNA replication, Homologous recombination(0.007), Base excision repair*(0.01), FoxO signaling pathway (0.04)*	Mitotic cell cycle phase transition (2.5e-11), DNA metabolic process (3.9e-11), cellular response to DNA damage stimulus(2.1e-06)
B	Circadian rhythm (0.005)*, Steroid hormone biosynthesis (0.04) *	Acute-phase response* (0.002), neutrophil degranulation* (0.002)
C	Cell cycle (2.7e-11), DNA replication (5.1e-06), Homologous recombination (1.1e-05), Base excision repair (0.03)	DNA metabolic process (5.2e-15), regulation of transcription involved in G1/S and G2/M transition of mitotic cell cycle (9e-07), signal transduction by p53 class mediator (0.02)
D	DNA replication (1.58e-10), Homologous recombination (3.38e-05), Mismatch repair (2.5e-04), Cell cycle (2.5e-03), Nucleotide excision repair (3.2e-03), Base excision repair (0.01), Cellular senescence* (0.005), One carbon pool by folate*(0.01)	DNA repair (6e-09), DNA replication (3.5e-10), tRNA export from nucleus (0.01)
E	Glycerophospholipid metabolism (0.001)*, Complement and coagulation cascades*(0.002), TNF signaling pathway*(0.002), Glycerolipid metabolism*(0.005), NF-kappa B signaling pathway*(0.005)	Cellular response to cytokine stimulus (2e-04), inflammatory response (0.003)
F	DNA replication*(0.001), RNA degradation*(0.005), RNA transport*(0.01), Pyrimidine metabolism*(0.01), Mismatch repair*(0.01), Spliceosome*(0.02), Nucleotide excision repair *(0.02)	DNA metabolic process* (1.6e-04), protein import* (5.1e-04), nucleobase-containing small molecule interconversion* (0.007)
G	p53 signaling pathway*(0.002), Non-homologous end-joining*(0.01), Homologous recombination*(0.01), One carbon pool by folate*(0.02), Cysteine and methionine metabolism*(0.02)	DNA recombination* (7e-04), DNA damage induced protein phosphorylation* (0.004)
H	Spliceosome (1.5e-05), Cell cycle (0.01), p53 signaling pathway (0.01), Protein processing in endoplasmic reticulum* (0.004), Amino sugar and nucleotide sugar metabolism* (0.01), Purine metabolism* (0.01), RNA transport* (0.02).	RNA splicing (1.9e-07), protein N-linked glycosylation (0.001), RNA transport (0.003)
I	NF-kappa B signaling pathway (2.6e-04), Platelet activation (0.005), TNF signaling pathway (0.02)	Cellular defense response (1.48e-06), inflammatory response (0.0001)
J	Fatty acid degradation* (0.001), Wnt signaling pathway* (0.01), PPAR signaling pathway* (0.02), Bile secretion* (0.03), Valine, leucine and isoleucine degradation* (0.03), Taurine and hypotaurine metabolism* (0.04)	Fatty acid beta-oxidation (0.02), regulation of canonical Wnt signaling pathway* (0.01)
K	Steroid hormone biosynthesis* (0.004), ECM-receptor interaction* (0.01), Metabolism of xenobiotics by cytochrome P450* (0.04)	Positive regulation of cytokine biosynthetic process* (0.001)
L	Steroid hormone biosynthesis (3.8e-07), Metabolism of xenobiotics by cytochrome P450 (1.57e-05), Chemical carcinogenesis (1.57e-05), Drug metabolism (2.2e-05), PPAR signaling pathway (7e-04), Primary bile acid biosynthesis (0.001), Retinol metabolism (0.005), Cholesterol metabolism (0.008), Glutathione metabolism (0.03), Tryptophan metabolism (0.03), Arachidonic acid metabolism (0.04)	Steroid biosynthetic process (3.1e-04), bile acid biosynthetic process* (5e-04)
M	TGF-beta signaling pathway* (0.01), Primary bile acid biosynthesis*(0.02), Glycine, serine and threonine metabolism* (0.02), Bile secretion*(0.03)	Bile acid metabolic process* (0.001)
N	Drug metabolism (0.04), Steroid hormone biosynthesis* (0.006), Glycolysis / Gluconeogenesis*(0.01), Biosynthesis of unsaturated fatty acids* (0.02), PPAR signaling pathway* (0.02)	Regulation of biosynthetic process* (0.001), regulation of gluconeogenesis* (0.01)

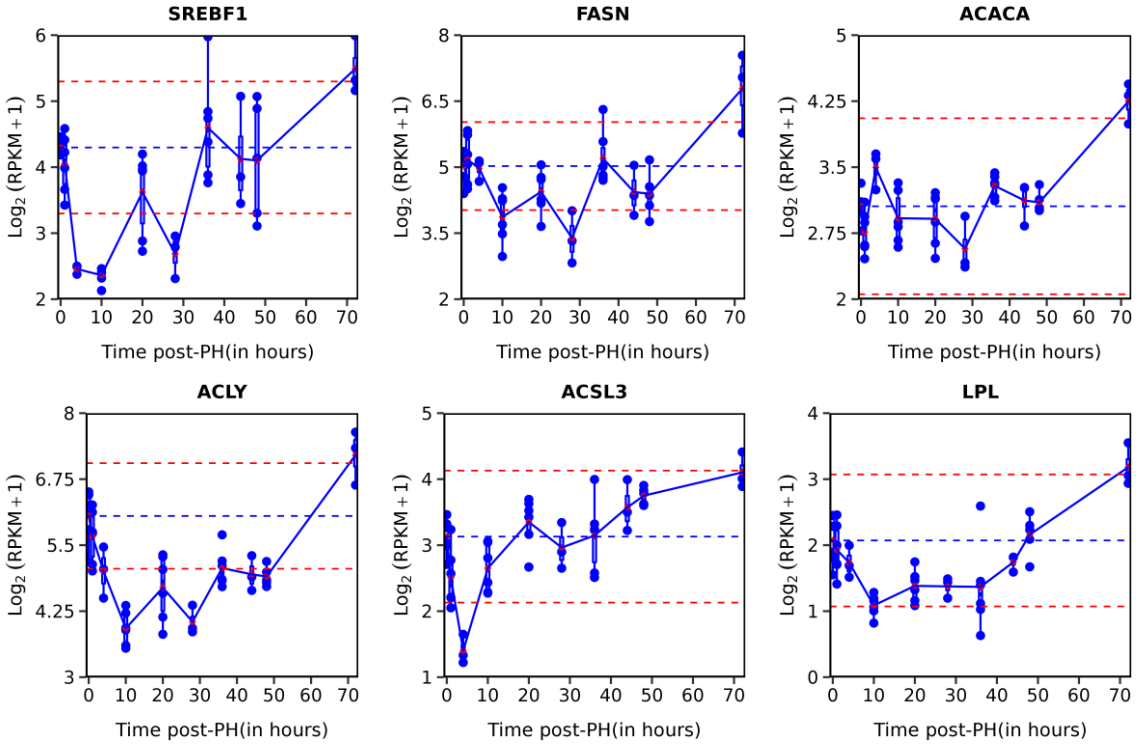


Figure A1: Expression profile of genes in lipid biosynthesis pathway. The blue line represents the baseline (0 hours), and the red line represents a two-fold change with respect to the baseline.

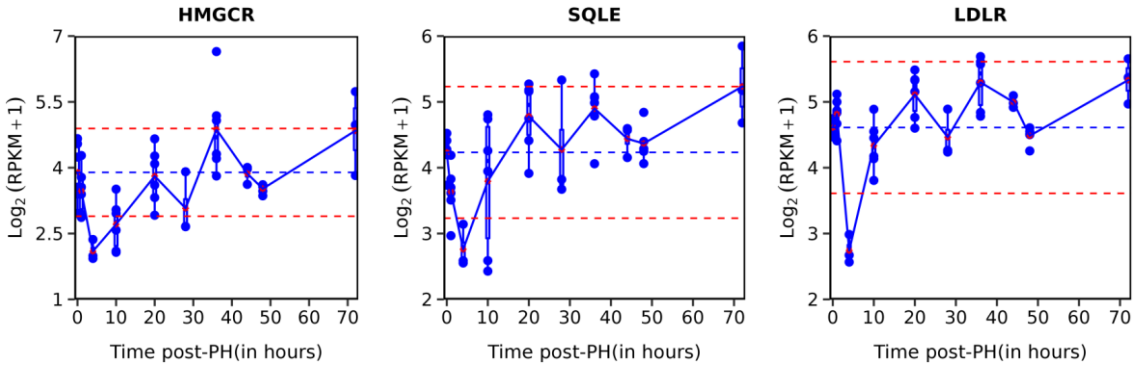


Figure A2: Expression profile of genes in cholesterol metabolism pathway. The blue line represents the baseline (0 hours), and the red line represents a two-fold change with respect to the baseline.

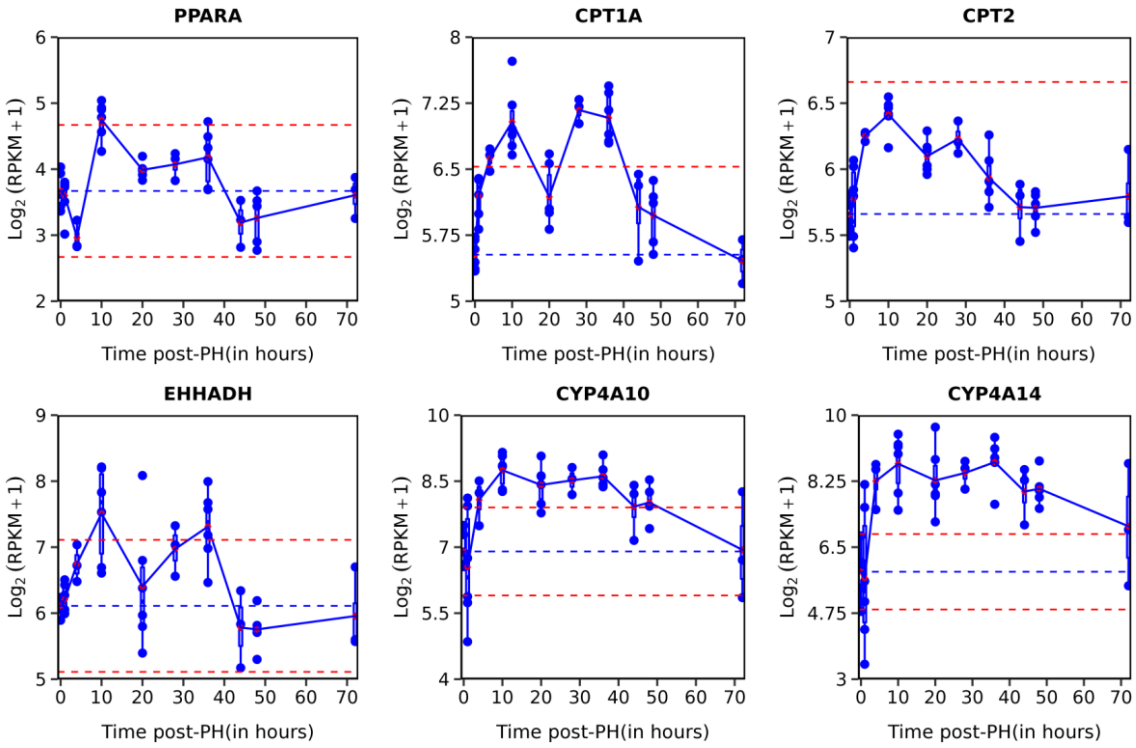


Figure A3: Expression profile of genes in β oxidation pathway. The blue line represents the baseline (0 hours), and the red line represents a two-fold change with respect to the baseline.

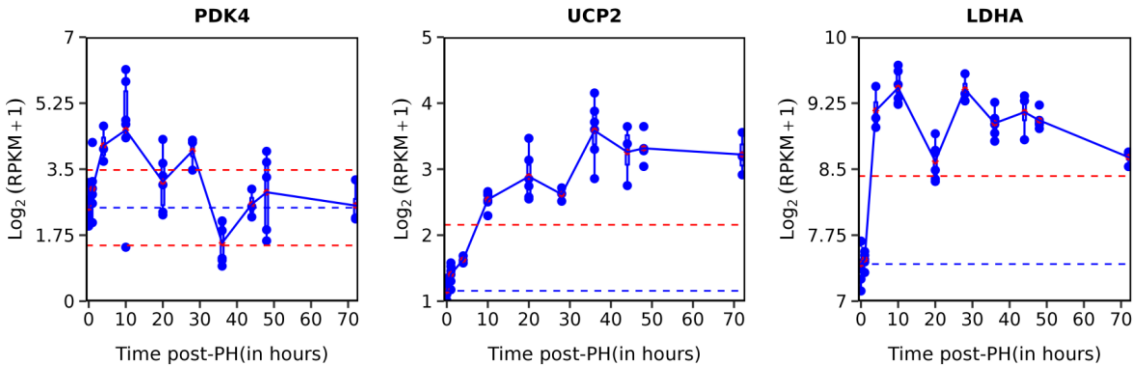


Figure A4: Expression profile of genes involved in the regulation of glucose and lipid metabolism and anaerobic glycolysis. The blue line represents the baseline (0 hours), and the red line represents a two-fold change with respect to the baseline.

Table A3: Discrete-continuous framework describing HNF4A regulatory circuit during liver regeneration

$\frac{d[HNF4A]}{dt} = \beta_{HNF4A}\left[\frac{1}{1 + e^{-\delta_{HNF4A} W_{HNF4A}}} - HNF4A\right]$	(A-1)
$\frac{d[CYCLIND]}{dt} = \beta_{CycD}\left[\frac{1}{1 + e^{-\delta_{CycD} W_{CycD}}} - CYCLIND\right]$	(A-2)
$W_{HNF4A} = w_{HNF4A\_0} + w_{S\_HNF4A} . S + w_{HNF4A\_CycD} . CYCLIND + w_{M\_HNF4A} . M$	(A-3)
$W_{CycD} = w_{CycD\_0} + w_{CycD\_Myc} . MYC$	(A-4)
$MYC = \frac{1}{1 + e^{-\delta_{Myc}(w_{Myc\_HNF4A} . HNF4A)}}$	(A-5)

Table A4: Discrete-continuous framework describing the balance between liver function and EMT during liver regeneration

$\frac{d[HNF4A]}{dt} = \beta_{HNF4A}\left[\frac{1}{1 + e^{-\delta_{HNF4A} W_{HNF4A}}} - HNF4A\right]$	(A-6)
$\frac{d[SNAIL]}{dt} = \beta_{SNAIL}\left[\frac{1}{1 + e^{-\delta_{SNAIL} W_{SNAIL}}} - SNAIL\right]$	(A-7)
$W_{HNF4A} = w_{HNF4A\_0} + w_{S\_HNF4A} . S + w_{HNF4A\_CycD} . CYCLIND + w_{HNF4A\_SNAIL} . SNAIL$	(A-8)
$W_{SNAIL} = w_{SNAIL\_0} + w_{SNAIL\_SNAIL} . SNAIL + w_{SNAIL\_HNF4A} . HNF4A$	(A-9)
$MYC = \frac{1}{1 + e^{-\delta_{Myc}(w_{Myc\_HNF4A} . HNF4A)}}$	(A-10)
$CYCLIND = \frac{1}{1 + e^{-\delta_{CycD}(w_{CycD\_0} + w_{CycD\_Myc} . MYC)}}$	(A-11)

Appendix B

Table B1: KEGG pathway enrichment using Enrichr for top 3 MCODE clusters of PPI network of DEGs comparing 18 months old mice with 3 months old mice

Cluster	Nodes	Edges	Score	KEGG Pathway	Overlap	Adjusted p-value
1	52	1240	48.627	Oxidative phosphorylation	45/134	3.66e-93
				Parkinson disease	45/144	8.99e-92
				Non-alcoholic fatty liver disease (NAFLD)	45/151	7.57e-91
				Alzheimer disease	45/175	1.30e-87
				Huntington disease	45/192	1.23e-85
2	55	710	26.296	Cell cycle	8/123	2.80e-08
				Progesterone-mediated oocyte maturation	7/90	4.52e-08
				Oocyte meiosis	6/116	4.61e-06
				p53 signaling pathway	3/71	4.43e-03
				Cellular senescence	4/185	6.02e-03
3	28	342	25.333	Ribosome	17/170	5.5e-29

Table B2: KEGG pathway enrichment using EnrichR for top 5 MCODE clusters of PPI network of DEGs comparing 24 months old mice with 3 months old mice

Cluster	Nodes	Edges	Score	KEGG Pathway	Overlap	Adjusted p-value
1	87	2875	66.860	Proteasome	40/46	8.08e-91
				Epstein-Barr virus infection	28/229	1.31e-31
				Basal cell carcinoma	5/63	2.83e-04
				Human T-cell leukemia virus 1 infection	8/245	2.83e-04
				p53 signaling pathway	5/71	2.83e-04
2	32	481	31.032	Thyroid hormone signaling pathway	11/115	1.62e-17
3	52	572	22.431	DNA replication	13/35	2.33e-24
				Cell cycle	16/123	2.23e-22
				Nucleotide excision repair	6/43	1.47e-08
				Ubiquitin mediated proteolysis	6/138	1.29e-05
				Circadian rhythm	3/30	4.35e-04
4	62	632	20.721	Collecting duct acid secretion	16/27	5.24e-33
				Synaptic vesicle cycle	18/77	1.54e-28
				Rheumatoid arthritis	18/84	5.89e-28
				Oxidative phosphorylation	18/134	3.84e-24
				Phagosome	18/180	7.76e-22
5	17	135	16.875	Peroxisome	12/84	1.13e-24

Appendix C

Table C1: Survival analysis of genes from modules T1, T2 and T9 on tumour samples of the Korean cohort and their validation on TCGA tumour samples. Genes with p-value ≤ 0.01 are shown.

	Gene	Korean		TCGA	
		p-value	Hazard ratio	p-value	Hazard ratio
Module T1	COLEC10	4.4e-04	0.18	4.2e-03	0.62
	CRHBP	4.4e-04	0.18	4.1e-04	0.55
	CFP	6.0e-04	0.18	3.3e-04	0.55
	CETP	1.8e-03	0.23	1.7e-03	0.59
	CD5L	1.8e-03	0.23	2.4e-03	0.60
	TIMD4	2.5e-03	0.24	1.1e-03	0.58
	FCN3	3.3e-03	0.25	5.0e-03	0.62
	ANKRD55	3.5e-03	0.25	1.2e-03	0.58
	DNASE1L3	3.6e-03	0.25	5.0e-05	0.51
Module T2	LPA	7.0e-04	0.21	4.4e-04	0.56
	ABCA9	1.4e-03	0.22	1.8e-04	0.53
	CFHR4	5.3e-03	0.28	3.4e-03	0.61
	HPX	7.4e-03	0.29	8.3e-03	0.64
	BDH1	7.7e-03	0.30	4.2e-03	0.62
	ASGR2	9.4e-03	0.30	4.8e-03	0.62
Module T9	CCNA2	1.3e-05	9.50	1.1e-03	1.70
	RAD51C	2.6e-04	6.00	4.7e-03	1.60
	DEPDC1B	8.0e-04	4.80	9.6e-03	1.50
	STMN1	9.3e-04	4.70	6.6e-04	1.80
	UBE2C	9.3e-04	4.70	8.4e-05	1.90
	UBE2S	9.3e-04	4.70	3.3e-04	1.80
	MCM3	9.3e-04	4.70	5.4e-03	1.60
	BIRC5	9.3e-04	4.70	7.8e-05	1.90
	STIL	9.3e-04	4.70	7.1e-03	1.60
	CENPM	9.3e-04	4.70	1.1e-03	1.70
	CHEK1	9.3e-04	4.70	4.6e-04	1.80
	CDK16	9.3e-04	4.70	5.1e-03	1.60
	ANGPT2	9.3e-04	4.70	9.7e-04	1.70
	MCM6	1.0e-03	4.70	1.9e-03	1.70
	TK1	1.2e-03	4.60	2.5e-04	1.80
	CDKN2C	1.2e-03	4.60	2.7e-03	1.70
	DLGAP5	1.2e-03	4.60	8.3e-03	1.60
	PLK4	1.2e-03	4.60	7.5e-04	1.80
	CDC20	1.5e-03	4.50	8.7e-03	1.60
	CDCA5	1.5e-03	4.50	2.6e-04	1.80
	MKI67	1.5e-03	4.40	1.6e-03	1.70
	FANCI	1.5e-03	4.40	7.8e-03	1.60
	KIF11	1.5e-03	4.40	5.0e-03	1.60
	TACC3	1.5e-03	4.40	8.1e-03	1.60
	CENPH	1.5e-03	4.40	2.8e-04	1.80
	MAD2L1	1.6e-03	4.40	5.2e-03	1.60
	PRKCA	2.2e-03	4.20	4.4e-03	1.60
	TTK	2.3e-03	4.20	2.1e-03	1.70
	KIF20A	2.3e-03	4.20	2.2e-03	1.70
	EFNA4	2.4e-03	4.20	3.2e-03	1.60
	EHMT2	2.4e-03	4.20	1.6e-03	1.70
	CDK1	5.1e-03	3.60	4.5e-03	1.60
	KIF2C	5.1e-03	3.60	6.6e-05	1.90
	SHCBP1	5.1e-03	3.60	5.7e-03	1.60
	NUP85	5.2e-03	3.60	3.7e-03	1.60
	CBX1	5.6e-03	3.50	4.7e-03	1.60
	CDK4	6.0e-03	3.50	3.1e-04	1.80
	WDR76	6.4e-03	3.50	1.1e-03	1.70

Table C2: Significant KEGG Pathways and GO Biological processes of premalignant modules

Module	KEGG Pathways	GO Biological Processes
N2	Glycine, serine and threonine metabolism (3.4e-07); Drug metabolism (7.4e-07); Tryptophan metabolism (1.7e-04); Complement and coagulation cascades (2.5e-04); Cysteine and methionine metabolism (2.5e-05); PPAR signaling pathway (0.001); Primary bile acid biosynthesis (0.002); Bile secretion (0.002); Steroid biosynthesis (0.002); Chemical carcinogenesis (0.005); Retinol metabolism (0.01); Metabolism of xenobiotics by cytochrome P450 (0.01); Tyrosine metabolism (0.01); Arginine and proline metabolism (0.02); Cholesterol metabolism (0.02); Steroid hormone biosynthesis (0.03); One carbon pool by folate (0.04); Purine metabolism (0.04)	cellular amino acid catabolic process (3.9e-08); steroid metabolic process (2.9e-05); fatty acid metabolic process (9.2e-05); epoxygenase P450 pathway (5e-04); cholesterol metabolic process (6e-04); drug catabolic process (8e-04); regulation of complement activation (8e-04); purine ribonucleoside monophosphate catabolic process (8e-04); regulation of immune effector process (0.001); regulation of humoral immune response (0.001); monocarboxylic acid metabolic process (0.001); dicarboxylic acid metabolic process (0.001)
N4	Cell cycle (1.4e-14); Oocyte meiosis (6.5e-06); Progesterone-mediated oocyte maturation (1.7e-05); Human T-cell leukemia virus 1 infection (1e-04); p53 signaling pathway (1e-04); Cellular senescence (0.01); DNA replication (0.03)	mitotic spindle organization (3.2e-27); microtubule cytoskeleton organization involved in mitosis (6.1e-24); mitotic sister chromatid segregation (7.9e-17); mitotic cytokinesis (8.5e-12); spindle assembly checkpoint signaling (3.2e-11); negative regulation of mitotic metaphase/anaphase transition (4.1e-11); cytoskeleton-dependent cytokinesis; mitotic nuclear division (6.1e-10); mitotic chromosome condensation (1.4e-08); kinetochore organization (1.8e-08); regulation of exit from mitosis (1.1e-07); regulation of G2/M transition of mitotic cell cycle (8.4e-07); cell cycle G2/M phase transition (5.5e-06); regulation of cyclin-dependent protein serine/threonine kinase activity (6e-06); anaphase-promoting complex-dependent catabolic process (6.7e-06)
N5	Protein processing in endoplasmic reticulum (5.1e-05); Hepatitis C (7.7e-05); Th17 cell differentiation(7.8e-05); IL-17 signaling pathway (1e-04); PPAR signaling pathway (1e-04); Antigen processing and presentation (0.001); Cytokine-cytokine receptor interaction (0.003); Th1 and Th2 cell differentiation(0.003); Biosynthesis of unsaturated fatty acids (0.005); Bile secretion (0.01); Viral protein interaction with cytokine and cytokine receptor(0.02); Arginine biosynthesis (0.02); TNF signaling pathway (0.04); Cholesterol metabolism (0.03); B cell receptor signaling pathway; Epstein-Barr virus infection(0.04); MAPK signaling pathway (0.04); PD-L1 expression and PD-1 checkpoint pathway in cancer (0.04)	cellular response to type I interferon (4.6e-16); cytokine-mediated signaling pathway (3.4e-14); defense response to virus (1.2e-10); negative regulation of viral process (3.6e-09); receptor-mediated endocytosis (2.7e-06); regulation of viral genome replication (5.6e-06); cellular response to interferon-gamma (2.2e-05); regulation of ribonuclease activity (3.2e-05); regulation of nuclease activity (8.9e-05); interleukin-27-mediated signaling pathway (2e-04); interferon-gamma-mediated signaling pathway (6e-04); response to unfolded protein (6e-04)
N7	Phagosome (4.6e-06); Complement and coagulation cascades (6.5e-05); Fructose	inflammatory response (2.2e-07); neutrophil degranulation (2.6e-07); neutrophil activation

	and mannose metabolism (1e-04); HIF-1 signaling pathway (2e-04); Viral protein interaction with cytokine and cytokine receptor (0.001); Human T-cell leukemia virus 1 infection (0.002); Cell adhesion molecules (0.002); Chemokine signaling pathway (0.003); Natural killer cell mediated cytotoxicity (0.004); PPAR signaling pathway (0.005); Antigen processing and presentation (0.01); Leukocyte transendothelial migration(0.01); Galactose metabolism(0.01)	involved in immune response (2.6e-07); neutrophil-mediated immunity (2.6e-07); cytokine-mediated signaling pathway (1.3e-06); regulation of T cell proliferation (7.8e-06); regulation of immune response (1.6e-05); chemokine-mediated signaling pathway (6e-04); negative regulation of lymphocyte activation (6e-04); positive regulation of MAPK cascade (7e-04); cellular response to chemokine (7e-04); dendritic cell differentiation (8e-04); positive regulation of ERK1 and ERK2 cascade (8e-04)
N10	Complement and coagulation cascades (0.02); Bladder cancer (0.02)	nuclear-transcribed mRNA catabolic process ,nonsense-mediated decay (0.008); cellular protein metabolic process (0.01); SRP-dependent cotranslational protein targeting to membrane (0.01); regulation of lipid metabolic process (0.01); protein targeting to ER (0.02); regulation of apoptotic process (0.02); response to interferon-gamma (0.02); glutamate metabolic process (0.02); regulation of actin cytoskeleton organization (0.03); cytoplasmic translation; peptide biosynthetic process (0.03); regulation of ERK1 and ERK2 cascade (0.04)
N11	Huntington disease (0.003); Prion disease (0.003); Ribosome (0.003); Oxidative phosphorylation (0.01); Non-alcoholic fatty liver diseas (0.01)e; Alzheimer disease (0.01); Parkinson disease (0.03); RNA polymerase (0.03)	mitochondrial translational elongation (7.9e-05); mitochondrial translational termination (7.9e-05); peptide biosynthetic process (8.5e-05); mitochondrial translation(1e-04); cellular macromolecule biosynthetic process (2e-04); aerobic electron transport chain (5e-04); mitochondrial ATP synthesis coupled electron transport (5e-04); translation (5e-04); cellular protein metabolic process (0.002); purine nucleotide metabolic process (0.005); lipoprotein catabolic process (0.01)

Table C3: Survival analysis of genes from premalignant modules N3, N10, N7, and N5 on Chinese cohort tumour-adjacent paired normal samples. Hub genes with p-value ≤ 0.01 are shown.

	Gene	p-value	Hazard ratio
Module N3	PLK2	8.5e-07	2.50
	PLSCR1	8.2e-04	1.90
	ODC1	1.4e-03	1.80
	WWC1	1.6e-03	1.80
	RANGAP1	3.2e-03	1.70
	MYC	4.0e-03	1.70
	AVPR1A	4.1e-03	0.58
	DDX21	4.5e-03	1.70
	MCL1	5.3e-03	1.70
	BAG3	5.6e-03	1.70
	PHLDA1	6.1e-03	1.70
	PPRC1	6.5e-03	1.70
	B4GALT5	7.2e-03	1.70
	CDC37L1	7.3e-03	1.70
	SPRY2	8.1e-03	1.60
	RAB20	9.4e-03	1.60
Module N10	CDC42EP1	0.00086	1.90
	GGT1	0.00150	1.80
	TAGLN2	0.00180	1.80
	SDF2L1	0.00250	1.80
	DUSP5	0.00270	1.80
	GDF15	0.00360	1.70
	WDR13	0.00410	1.70
	TSPAN4	0.00450	0.59
	CKS2	0.00570	1.70
	IER3	0.00600	1.70
	ID3	0.00800	1.60
	PCNA	0.00830	1.60
	GNB2	0.00830	1.60
	CTSA	0.00850	1.60
	IFITM2	0.00980	1.60
Module N7	THBD	0.00035	2.00
	NUCB2	0.00045	1.90
	BCL2L1	0.00072	1.90
	MVP	0.00130	1.80
	PFKFB3	0.00210	1.80
	EPS8L3	0.00260	0.57
	CD59	0.00290	1.80
	GPX3	0.00290	1.70
	RBPMS	0.00310	1.70
	MYH4	0.00390	0.58
	ME1	0.00400	0.58
	ALDOA	0.00470	1.70
	BIRC3	0.00630	1.70
	NCKAP1L	0.00930	1.60
	ASCC3	0.00960	1.60
Module N5	GABBR2	0.00044	0.52
	JUN	0.00150	1.80
	SPINK1	0.00250	1.80
	AKAP12	0.00290	1.70
	TIPARP	0.00460	1.70
	FOS	0.00470	1.70
	EGR2	0.00580	1.70
	HSP90AA1	0.00580	1.70
	CPD	0.00760	1.60
	ERBB3	0.00900	0.61
	CPT1A	0.00940	1.60
	EPO	0.00950	0.62