Innovative Approaches To Identify Regulators Of Liver Regeneration

Amber Weiching Wang

University of Pennsylvania

Follow this and additional works at: https://repository.upenn.edu/edissertations

Part of the Bioinformatics Commons, Genetics Commons, and the Systems Biology Commons

Recommended Citation

https://repository.upenn.edu/edissertations/3663

This paper is posted at ScholarlyCommons. https://repository.upenn.edu/edissertations/3663
For more information, please contact repository@pobox.upenn.edu.
Innovative Approaches To Identify Regulators Of Liver Regeneration

Abstract
The mammalian liver possesses a remarkable ability to regenerate after injury to prevent immediate organ failure. However, amid a rising global burden of liver disease, the only curative treatment for patients with end-stage liver disease is transplantation. Elucidating the mechanisms underlying tissue repair and regrowth will enable identification of therapeutic targets to stimulate native liver regeneration, thereby circumventing the great paucity of available transplant organs. Here, utilizing the Fah-/- mouse model of liver repopulation, I applied transcriptomic and epigenomic techniques to investigate the changes occurring as hepatocytes restore organ mass following toxic injury. By labeling ribosomal or nuclear envelope proteins, I performed the first extensive characterization of gene expression and chromatin landscape changes specifically in repopulating hepatocytes in response to injury. Transcriptomic analysis showed that repopulating hepatocytes highly upregulate Slc7a11, a gene that encodes the cystine/glutamate antiporter. I demonstrated that ectopic Slc7a11 expression promotes liver regeneration and Slc7a11 mutation inhibits hepatocyte replication. Integrative bioinformatics analyses of chromatin accessibility revealed dynamic changes at promoters and liver-enriched enhancer regions that correlate with the activation of proliferation-associated genes and the repression of transcripts expressed in mature, quiescent hepatocytes. Furthermore, changes in chromatin accessibility and gene expression are associated with increased promoter binding of CCCTC-binding factor (CTCF) and decreased enhancer occupancy of hepatocyte nuclear factor 4α (HNF4α). In summary, my thesis work identifies Slc7a11 as a potential driver of liver regeneration, and provides insights into the complex crosstalk between chromatin accessibility and transcription factor occupancy to regulate gene expression in repopulating hepatocytes.

Degree Type
Dissertation

Degree Name
Doctor of Philosophy (PhD)

Graduate Group
Pharmacology

First Advisor
Klaus H. Kaestner

Keywords
ATAC-seq, Fah, hepatocytes, liver regeneration, multiomics analysis, TRAP-seq

Subject Categories
Bioinformatics | Genetics | Systems Biology

This dissertation is available at ScholarlyCommons: https://repository.upenn.edu/edissertations/3663
DEDICATION

Dedicated to my parents, Dr. Chuhsing Kate Hsiao and Dr. Chih-Hao Wang.
ACKNOWLEDGMENT

I would like to thank my advisor Dr. Klaus Kaestner for providing me with the mentorship and support to explore the field of bioinformatics while performing bench experiments. This thesis work would not be possible without his vision and belief in implementing novel technologies to advance biomedical research. I am also grateful for all the Kaestner lab members for their friendship and advice. Drs. Kirk Wangensteen, Julia Yue Wang, and Adam Zahm have been integral in helping me advance my thesis work as well as inspiring me to persevere throughout the Ph.D. process. Thank you to my thesis committee, Drs. Mitchell Lazar, Michele Pack, Benjamin Voight, and Kirk Wangensteen, for their guidance and encouragement.

The tremendous help I received throughout training has enabled the progression of my thesis work. I would like to thank the Functional Genomics Core, Dr. Jonathan Schug, Shilpa Rao, Olga Smirnov, and Joseph Kutch, the Pharmacology Graduate Group, Dr. Julie Blendy, Sarah Squire, and my classmates, the Institute of Diabetes, Obesity, and Metabolism, especially Vesselina, for their assistance and company. I am blessed with the opportunity to meet the most intelligent and hardworking people at Penn, and I am particularly thankful to Drs. Kristy Ou, Julia Kieckhaefer, Yong Hoon Kim, Alexander Sakers, Ayano Kondo, Teguru Tembo, and Rebecca Myers for their invaluable friendship.

I am forever grateful to my family, my parents Kate and Peter, as well as my sister Leiya, for their unconditional love and endless support. My parents are my greatest inspiration to pursue a career in science and motivation to persist through any obstacles. I would also like to thank the Ho family, especially Yung and Cojen, for making me feel closer to home. I am thankful to Wesley for all the love and laughter he brought into my life. Finally, thank you to our dog Mochi, for keeping me where the light is.
ABSTRACT

INNOVATIVE APPROACHES TO IDENTIFY REGULATORS OF LIVER REGENERATION

Amber W. Wang
Klaus H. Kaestner

The mammalian liver possesses a remarkable ability to regenerate after injury to prevent immediate organ failure. However, amid a rising global burden of liver disease, the only curative treatment for patients with end-stage liver disease is transplantation. Elucidating the mechanisms underlying tissue repair and regrowth will enable identification of therapeutic targets to stimulate native liver regeneration, thereby circumventing the great paucity of available transplant organs. Here, utilizing the Fah<sup>−/−</sup> mouse model of liver repopulation, I applied transcriptomic and epigenomic techniques to investigate the changes occurring as hepatocytes restore organ mass following toxic injury. By labeling ribosomal or nuclear envelope proteins, I performed the first extensive characterization of gene expression and chromatin landscape changes specifically in repopulating hepatocytes in response to injury. Transcriptomic analysis showed that repopulating hepatocytes highly upregulate Slc7a11, a gene that encodes the cystine/glutamate antiporter. I demonstrated that ectopic Slc7a11 expression promotes liver regeneration and Slc7a11 mutation inhibits hepatocyte replication. Integrative bioinformatics analyses of chromatin accessibility revealed dynamic changes at promoters and liver-enriched enhancer regions that correlate with the activation of proliferation-associated genes and the repression of transcripts expressed in mature, quiescent hepatocytes. Furthermore, changes in chromatin accessibility and gene expression are associated with increased promoter binding of CCCTC-binding factor (CTCF) and decreased enhancer occupancy of hepatocyte nuclear factor 4α (HNF4α). In summary, my thesis work identifies Slc7a11 as a potential driver of liver regeneration, and provides insights into the complex crosstalk between chromatin accessibility and transcription factor occupancy to regulate gene expression in repopulating hepatocytes.
TABLE OF CONTENTS

DEDICATION II

ACKNOWLEDGMENT III

ABSTRACT IV

LIST OF TABLES VII

LIST OF ILLUSTRATIONS VIII

CHAPTER 1 INTRODUCTION 1
LIVER BIOLOGY 2
LIVER REGENERATION 12
LIVER TRANSCRIPTIONAL CONTROL 31
SPECIFIC AIMS 48
FIGURES 49
REFERENCES 52

CHAPTER 2 TRAP-SEQ IDENTIFIES CYSTINE/GLUTAMATE ANTIPORTER AS A DRIVER OF RECOVERY FROM LIVER INJURY 78
ABSTRACT 79
INTRODUCTION 80
RESULTS 82
TABLES 92
FIGURES 95
DISCUSSION 106
MATERIALS AND METHODS 109
REFERENCES 116
LIST OF TABLES

Table 2.1. Top ten abundant transcripts identified in quiescent livers. 92
Table 2.2. FPKM of cell type-specific transcripts detected by TRAP-seq. 93
Table 2.3. Upstream regulators predicted by Ingenuity Pathway Analysis. 94
Table 3.1. Materials for the TRAP-seq protocol. 135
Table 4.1. ATAC-seq library sequencing summary. 158
Table 4.2. Gene expression of enriched transcription factor motifs. 159
Table 4.3. Primer sequences used in this study. 160
LIST OF ILLUSTRATIONS

Figure 1.1. Schematic representation of a liver lobule. 49
Figure 1.2. Regulation of the cell cycle by cyclin proteins and cyclin-dependent kinases (CDK). 50
Figure 1.3. Liver-enriched transcription factors form a complex regulatory network. 51
Figure 2.1. Translating ribosome affinity purification (TRAP) enables cell type-specific isolation of RNA from quiescent and repopulating hepatocytes. 95
Figure 2.2. Mice in the 4-week regeneration after severe injury group exhibit significant weight loss. 97
Figure 2.3. TRAP-seq identifies differentially expressed genes specific to repopulating hepatocytes in the Fah⁻/- model. 98
Figure 2.4. Comparison of the Fah⁻/- TRAP-seq data with RNA-seq data from the PHx model identifies common and unique characteristics of liver repopulation paradigms. 99
Figure 2.5. Comparison of identified transcripts from single-cell RNA-seq (scRNA-seq) shows significant overlap between TRAP-seq and all nine layers of scRNA-seq. 100
Figure 2.6. Slc7a11 enhances hepatocyte repopulation. 101
Figure 2.7. Slc7a11 is activated at the transcript and protein levels under increased oxidative stress during liver regeneration. 103
Figure 2.8. No significant differences in the weight of mice with Slc7a11 inhibition compared to control after 4 weeks of repopulation. 104
Figure 2.9. Slc7a11 is activated by ATF4 during liver repopulation. 105
Figure 3.1. Implementation of translating ribosome affinity purification (TRAP) with Fah⁻/- to profile gene expression change of repopulating hepatocytes. 136
Figure 3.2. TRAP enables cell type-specific isolation of high-quality RNA. 137
Figure 3.3. TRAP-isolated RNA can be used for downstream gene expression analysis. 138
Figure 4.1. Implementation of the ‘isolation of nuclei tagged in specific cell types’ (INTACT) method with the Fah⁻/- mouse model allows isolation of repopulating hepatocyte nuclei. 161
Figure 4.2. Chromatin accessibility changes during liver repopulation are related to cell growth activation and metabolic inhibition. 162
Figure 4.3. Association of expression levels and chromatin accessibility implicates divergent regulatory mechanisms for gene activation and inhibition. 164
Figure 4.4. Enrichment analysis identifies transcription factor motifs overrepresented at differential accessible promoters and enhancers. 166
Figure 4.5. HNF4α binding is decreased in the repopulating liver. 167
Figure 4.6. CTCF binding is increased at promoters in the repopulating liver. 169
Figure 4.7. Decreased nucleosome density is associated with increased gene expression in repopulating hepatocytes. 171
Figure 4.8. Model of transcriptional regulation in repopulating hepatocytes. 173
CHAPTER 1

INTRODUCTION
LIVER BIOLOGY

I. Function

As the hub of various biological processes, the liver performs a multitude of functions that can be categorized as the following.

(1) Regulation of carbohydrate, protein, and lipid homeostasis. The liver undergoes gluconeogenesis to release glycogen as glucose in response to fasting, packages excess lipids for storage in other tissues, and processes amino acids via deamination to convert the non-nitrogenous carbon skeleton to glucose or lipids [1,2].

(2) Metabolism of nutrients, wastes, and xenobiotics. The metabolic process consists of phases I and II. Phase I involves direct modification including oxidation, reduction, and hydrolysis often achieved by cytochrome P450 (CYP450) proteins. Phase II is carried out by enzymes to conjugate large molecules of phase I metabolites to decrease activity and increase solubility [3].

(3) Synthesis of bile, amino acids, coagulation factors, and serum proteins. The liver performs the conversion of ammonia to urea through the urea cycle [4] and carries out the conjugation of bilirubin for secretion into the bile that drains into the intestine for degradation [5]. The liver also synthesizes non-essential amino acids, bile acid from cholesterol, clotting factors for blood coagulation, and various serum proteins such as albumin (ALB) and transferrin [2,6].

II. Cell types

The liver consists of cell types of divergent embryological origin – with two main parenchymal cell types, hepatocytes and cholangiocytes, and nonparenchymal cells including Kupffer cells, stellate cells, and sinusoidal epithelial cells.

Hepatocytes are the main cell type that performs the majority of liver functions mentioned above. They occupy 60% of the liver by cell number and 80% by cell mass [7]. Cholangiocytes, also known as biliary epithelial cells (BEC), are cuboidal cells that line the bile duct to allow passage of bile acid from the liver to the intestine. They are less metabolic active compared to hepatocytes but still exhibit functions such as bicarbonate synthesis and electrolyte secretion [8].
Cholangiocytes also play an important role in regulating immune and inflammatory responses [8]. Small ducts embedded deep within the liver are called intrahepatic bile ducts whereas large ducts that exit the liver are extrahepatic bile ducts [7].

Kupffer cells are resident hepatic macrophages that recognize stimuli introduced through the portal circulation to perform phagocytosis and secret pro- or anti-inflammatory mediators to defend liver against bacterial and viral infections [9]. Stellate cells exist in two states – under quiescent conditions, they store vitamin A in lipid droplets whereas other functions remain unclear [2]; upon liver damage, they are activated to become proliferative myofibroblasts [10]. The myofibroblasts derived from stellate cells deposit collagen that contributes to the fibrosis or scarring of the liver tissue, which could progress to cirrhosis with chronic liver injury [10]. Sinusoidal endothelial cells are specific nonparenchymal cells that line the capillaries of the liver, also known as sinusoids, to form fenestrated sieve plates that permit access of macromolecules from the space of Disse, an interstitial area that surrounds hepatocytes. This structure allows hepatocytes to extract a variety of protein-bound substrates and xenobiotics from the circulation [11].

III. Structure

The liver is composed of building blocks termed ‘lobules’, which contains parenchymal and nonparenchymal cells, the bile duct, and vessels of the circulatory system including the hepatic artery, portal vein, and central vein (Figure 1.1). The portal vein, hepatic artery, and the bile duct are often referred to as the portal triad due to their spatial proximity. A typical lobule is considered to be a hexagonal unit with the central vein in the middle and the portal triad at the six corners [12]. The portal vein and the hepatic artery are the two main sources of blood supply for the liver, with the portal vein providing two-thirds of the blood from the small intestine and the hepatic artery contributing to the remaining one-third of the blood from the celiac artery. The blood from these two sources mixes as it passes through sinusoids, and enters the central vein to exit the liver [7]. Owing to the direction of the blood supply, which moves from the portal triad to the central vein, the portal area is considered ‘zone one’ and the central area as ‘zone three’.
Hepatocytes synthesize and transport bile acids via a specialized channel formed by two adjacent cells, also referred to as the canaliculus. Bile is produced in the canaliculi as a mixture of bile acids, metabolites, and bilirubin secreted from hepatocytes; it enters the bile ducts that ultimately drain into the duodenum. Bile acids assist in lipid and cholesterol emulsification in the intestines, and are reabsorbed from the terminal ileum and transported back to the liver through the portal vein, recycling in a route known as the enterohepatic circulation. Contrary to the blood flow, bile flows from zone three to zone one [6]. Hepatocytes in proximity to the portal triad are referred to as periportal or zone one hepatocytes, those adjacent to the central vein are pericentral or zone three hepatocytes, and cells between zone one and three are referred to as zone two hepatocytes.

Apart from the spatial distribution, hepatocytes from different zones express divergent sets of genes to carry out various metabolic functions, a property known as metabolic zonation, which is tightly controlled by the Wnt/β-catenin signaling pathway [13]. Periportal hepatocytes perform gluconeogenesis, fatty acid oxidation, urea production, and glutathione (GSH) detoxification, whereas pericentral hepatocytes conduct glycolysis, lipogenesis, ketogenesis, and xenobiotic metabolism [14]. Multiple cell types coupled with metabolic zonation allow the liver to carry out diverse metabolic and biosynthetic functions central to homeostasis.

**IV. Development**

In mice, the parenchymal cells of the liver develop from the definitive endoderm from the anterior primitive streak of the gastrulating embryo on embryonic day (E) 7.5 [15]. By E8.5, endoderm patterning is complete and can be categorized from the anterior to posterior as the foregut, midgut, and hindgut regions [16]. Hepatic specification subsequently begins on E9.0 when hepatic endoderm cells extend off the posterior foregut [17] and continue to thicken to establish the liver diverticulum [18]. Early signals important for the initiation of liver specification include fibroblast growth factor (FGF) family members emanating from the developing heart [19] and bone morphogenic proteins (BMP) from the septum transversum mesenchyme [20] of the mesoderm.
Simultaneously, transcription factors of the GATA binding proteins [21] and forkhead box A (FOXA) subfamily [22] activate transforming growth factor β (TGFβ), WNT, and NOTCH signaling within the endoderm. Additionally, FOXA1 and GATA4 function as ‘pioneer factors’ that bind to heterochromatic DNA to establish transcriptional competence of downstream gene programs required for further differentiation [2,23]. The liver bud gives rise to hepatoblasts, bipotential progenitor cells that express α-fetoprotein (AFP) and ALB [24,25]. Prior to differentiation, hepatoblasts continue to migrate and proliferate into the septum transversum assisted by a gradient TGFβ signal from the portal vein mesenchyme starting on E13.5 [26]. Cells in proximity to the portal vein receive a higher TGFβ concentration that promotes expression of hepatocyte nuclear factors 1β (HNF1β) and HNF6, leading to the expression of cholangiocyte marker genes such as cytokeratin 19 (CK19) [26]. Hepatoblasts located away from the portal vein receive lower levels of TGFβ, resulting in elevated expression of HNF1α and HNF4α, which induce hepatocyte-specific gene expression such as ALB and CYP450 [27]. Beginning on E18, the differentiated hepatocytes continue to mature and undergo a metabolic switch from a glucose-consuming tissue to a glucose-producing organ. Developing hepatocytes at distinct locations in the liver also experience metabolic zonation to establish differential gene expression and protein production [2]. This process is regulated by complex crosstalk of signaling networks including the hepatocyte growth factor (HGF), glucocorticoids, HNF, and the Wnt/β-catenin pathway [27].

V. Homeostasis

Tissue turnover typically occurs through two models: replication of existing cells and differentiation of progenitor cells. While the presence of stem cells in the adult liver has long been contested, it is currently accepted that mature resident hepatocytes proliferate for homeostatic maintenance. Since differentiated hepatocytes are long-lived cells with a turnover rate of up to several months in vivo, under normal physiological conditions, fewer than 0.1% of the hepatocytes undergo replication at any time in the uninjured adult liver [28]. It is, therefore, questionable whether
stem cells are required at all for normal liver maintenance considering the long life span and low replication rate of mature hepatocytes.

Lineage-tracing in rodents is widely used to identify the source of hepatocyte homeostasis. The most debated model, the ‘streaming liver hypothesis’, implemented DNA radiolabeling assays in the rat liver and observed that newly-formed hepatocytes occur near the portal vein and flow towards the central vein to replenish the liver parenchyma [29]. The streaming liver hypothesis implies the existence of a stem cell compartment proximal to zone one and posits that periportal hepatocytes display a higher replication capacity, hence the ability to derive new cells from the portal vein to the central vein. Over time, it is found that the entire liver lobule contains hepatocytes derived from zone one [29]. However, evidence both for and against the streaming liver hypothesis has been provided. Genetic tracking that utilizes the cholangiocyte marker ‘sex-determining region Y (SRY)-box 9’ (SOX9) to label all BECs via a tamoxifen-inducible system in the adult liver of Sox9CreERT2;RosaLSL-LacZ mice observed LacZ spreading in a portal-to-central direction that eventually occupied the entire liver parenchyma within a year [30]. These studies suggest SOX9-positive cholangiocytes as a source of mature hepatocytes to maintain the homeostatic liver.

Nonetheless, studies using different lineage-tracing systems have not observed the same evidence [31,32]. Other radiolabeling assays failed to detect movement of marked periportal hepatocytes towards the pericentral area in the adult liver [32]. Another transgenic Sox9CreERT2;RosaLSL-YFP line generated with a bacterial artificial chromosome (BAC) found that YFP-positive cholangiocytes are restricted within the bile ducts and do not migrate to the central vein [33]. Labeling of adult hepatocytes of the RosaLSL-YFP mouse through injection of AAV8-TTR-Cre, a hepatotropic adeno-associated virus (AAV) serotype 8 that expresses the Cre recombinase under the hepatocyte-specific transthyretin (TTR) promoter, did not demonstrate any YFP-negative hepatocytes in the liver parenchyma or near the periportal region [34]. These results indicate that all newly-derived hepatocytes are from preexisting mature cells and exclude the possibility of progenitors contributing to adult liver homeostasis. More recently, a ‘reverse-streaming hypothesis’ in which WNT-enriched pericentral hepatocytes expand to the periportal region have been
proposed [35], albeit with much controversy. In summary, the current evidence does not support the notion of liver stem cells as a source of mature hepatocytes during normal homeostasis; instead, hepatocytes are likely maintained by replication of preexisting cells.

**VI. Diseases**

Liver disease accounts for roughly 2 million annual deaths worldwide, of which 15% of the mortality results from acute hepatitis, 35% from hepatocellular carcinoma (HCC), and 50% from complications related to cirrhosis [36]. Strikingly, the combination of cirrhosis and HCC constitutes 3.5% of global deaths [37]. Although accurate statistics are not available due to the scarcity of mortality data from developing countries [36] and the underestimation of liver diseases as a cause of death [38], there is a discernible increase in the global burden of both acute and chronic liver disease [36,37,39].

**A. Types of liver disease**

**Alcoholic liver disease**

Alcohol contributes to over 50% of cirrhosis-related mortality and heavy alcohol consumption is associated with the development of cirrhosis [40]. For heavy drinkers — daily ethanol consumption of over 30g — the incidence of cirrhosis ranges from 1-6%, depending on the dose [41]. Furthermore, alcohol use exacerbates preexisting liver injuries [37]. Approximately 20% of patients with alcoholism develop alcoholic hepatitis, a clinical representation of jaundice and liver failure after chronic alcohol abuse, with a daily mean ethanol consumption of 100g [42]. Coupled with the rising rate of obesity globally, the severity of alcoholic liver disease, especially alcoholic fatty liver disease, is expected to worsen [37].

The molecular mechanism of alcohol-induced liver injury involves the oxidative metabolism of ethanol that shifts the oxidative-reduction potential in the liver, preventing fatty acid oxidation and inhibiting the tricarboxylic acid cycle that normally promotes lipolysis [43]. Additionally, ethanol activates sterol regulatory element-binding protein 1 (SREBP-1) [44], prevents peroxisome
proliferator-activated receptor α (PPARα) binding to the DNA [45], and inhibits AMP-activated protein kinase (AMPK) activity [46], leading to activation of fatty acid synthesis and metabolic remodeling that contributes to the development of fatty liver [47].

Non-alcoholic fatty liver disease (NAFLD)

Fatty liver, also called steatosis, is defined as excess accumulation of triglycerides in over 5% of fat in hepatocytes [48]. NAFLD encompasses two distinct conditions, steatosis without liver injury and steatosis with hepatocyte necrosis, referred to as non-alcoholic steatohepatitis (NASH) [48]. The global prevalence of NAFLD is estimated to be 25.2% [49] and of NASH is between 2-7% [49–52]. Current epidemiological studies potentially underestimate due to the difficulty of detecting fatty liver unless through imaging or liver biopsies [49]. Moreover, the increasing rates of comorbid conditions associated with NASH such as obesity, type 2 diabetes mellitus, and hyperlipidemia contribute to the rapidly growing burden of NAFLD [49]. Particularly in the case of NASH, chronic liver injury followed by lobular and portal inflammation in the form of collagen deposition and scar tissue production often lead to progression to fibrosis, cirrhosis, and HCC [48].

The exact mechanisms of NAFLD and NASH development are not completely understood and likely involve extensive interactions between various pathways. Three leading sources have been identified to contribute to NAFLD. (1) Increased uptake of fatty acids via diet, activation of de novo lipogenesis, and increased adipose tissue lipolysis, resulting in the accumulation of hepatic fatty acids. (2) A combination of fatty acid-induced extrinsic cell death through upregulation of cell death receptors and their ligands [53], as well as intrinsic cell death via increased endoplasmic reticulum (ER) stress [54], leading to Jun N-terminal kinase (JNK) activation, reactive oxygen species (ROS) production, and mitochondrial uncoupling [55]. Fatty acid-induced cell death is followed by the release of damage-associated molecular patterns (DAMPs) into the extracellular space [56–58]. (3) Triggering of hepatic inflammation due to fatty acids, DAMPs released by dying hepatocytes, and endotoxin from the intestine [59], inducing the production of cytokines and chemokines with subsequent recruitment of immune cells [60] and activation of nonparenchymal
cells [61,62]. In particular, the transformation of stellate cells into collagen-producing myofibroblasts further exacerbates NASH and promotes progression to fibrosis [63,64].

**Drug- or toxin-induced liver injury**

The liver is the first organ to be perfused by blood through the portal vein from the intestine for first-pass metabolism and is thus the initial filter for molecules before they enter into the general circulation. Therefore, exposure to environmental toxins can be severely damaging to hepatocytes. Substances such as alcohol, acetaminophen, *Amanita phalloides* mushrooms, and idiosyncratically, common drugs including anabolic steroids and antibiotics can cause acute liver failure [65]. At a rate of 18%, drug-induced liver injury is the leading cause of post-market withdrawal during drug development [66]. Acetaminophen is the primary etiology for drug-induced liver injury in the US and the UK, whereas herbal and alternative medicine are the leading causes in the East [37]. The exact mechanisms of liver injury vary by the drug consumed or the toxin ingested, but generally involves oxidative stress accumulation, mitochondrial dysfunction, hepatocyte necrosis and apoptosis, immune response stimulation, and nonparenchymal cell activation [67].

**Viral hepatitis**

Viral hepatitis refers to liver inflammation induced by viral infections, routinely caused by five hepatotropic viruses, hepatitis viruses A, B, C, D, and E. While viral hepatitis affects individuals from all geographic locations, middle and low-income areas are disproportionately affected [37]. An estimated 1.34 million annual deaths are associated with hepatitis-related mortality [68]. Hepatitis A and E typically result in acute, self-contained illnesses, whereas hepatitis B and C lead to immune-mediated chronic liver disease. There is an increased risk of developing HCC and cholangiocarcinoma (CAA) among patients with hepatitis B and C, although the rate of cirrhosis progression and tumorigenesis display individual heterogeneity [69]. The pathogenesis of hepatitis
B and C virus includes complex crosstalk between the host and virus that involves immune activation of CD4+ helper T cells, CD8+ effector T cells, and natural killer cells [69,70].

B. Consequences of liver disease

Acute liver failure

Acute liver failure, also known as fulminant hepatic failure, is defined as the clinical presentation of severe liver injury and hepatic encephalopathy within 8 weeks of the first symptoms without preexisting liver disease [71]. Other than the liver and the brain, acute liver failure also affects organs including the heart, lungs, pancreas, and kidney [72]. Hepatitis A, B, and E infections are the predominant causes of acute liver failure worldwide [72]. In the US, drug-induced injury contributes to approximately 50% of acute liver failure cases [73].

Cirrhosis

Cirrhosis is the end-stage of all chronic liver disease that develops from an asymptomatic phase termed ‘compensated’ cirrhosis to a progressive ‘decompensated’ phase marked by complications of ascites, jaundice, encephalopathy, and variceal bleeding [74]. Features of cirrhosis include regenerative and nodular parenchyma, widespread deposition of fibrotic tissues, and hepatocyte necrosis [75]. Patients with cirrhosis have a 5-10-fold increased risk of mortality [76]; death occurs due to a variety of complications including infections, kidney failure, and gastrointestinal bleeding [77]. Common causes of cirrhosis include alcohol abuse (60-70%), chronic hepatitis B or C (10%), and NAFLD (10%) [78].

Liver cancer

Considering the late-stage at diagnosis in most cases, HCC exhibits a 5-year survival rate of 18.4% with an annual rise of mortality by 2.4% [79]. Currently, 40% of HCC results from hepatitis B, 40% from hepatitis C, 11% from alcohol use, and 9% from other causes [37]; however, the etiology varies significantly for different countries. Non-viral factors contribute to a larger pool of
HCC in regions with a low incidence of viral hepatitis. In the US, 25% of HCC patients present with alcoholic-liver disease and 20-30% could display metabolic syndrome or NAFLD [80]. In contrast, viral hepatitis contributes to approximately 90% of HCC cases in Vietnam [81]. The etiology is expected to change drastically due to the increasing prevalence of NAFLD and NASH [82].
I. Historical overview

Regeneration, defined as cell regrowth or repair, is widely represented in metazoa [83]. In lower organisms, whole body or tissue regeneration is easily achieved. Invertebrates such as planarians are capable of whole-body regeneration from as little as 1/279th of the original body [84] and lower vertebrates including amphibians are capable of regenerating complete appendages. In mammals, the regenerative ability is restricted to select tissues including the skin, cartilage, digits, muscle, intestinal epithelium, and liver [85]. The evolutionary importance of the liver in maintaining a regenerative ability likely stems from it being the largest metabolic organ in the mammalian system, as nutrient and xenobiotic metabolism subject the liver to frequent, unpredictable environmental insults. It is presumed that the regenerative ability has been retained in animals in order to recover from massive liver injuries following exposure to food toxins [86]. Human’s ability to regenerate the liver has long been known, even codified in Greek mythology by the story of Prometheus. After stealing fire and giving it to humanity, Prometheus is chained on Mount Atlas and punished to eternal torment by Zeus. An eagle would feed on his liver daily, only for it to grow back overnight just to be eaten again the next day [87].

Fewer than 0.1% of hepatocytes are in mitosis under physiological conditions [88] with the typical life span of 200-400 days [32]. The liver can regenerate upon loss of the parenchyma via toxin-induced liver injury or surgical tissue removal. Generally, mature hepatocytes replicate to repopulate the liver under normal conditions but hepatocyte progenitor cells (HPCs) can arise when hepatocyte DNA synthesis is severely impaired [89]. In fact, hepatocytes have an almost unlimited replicative capacity as serial partial heptectomy (PHx) has demonstrated the rat liver to be able to regenerate after 18 surgeries [90], and serial transplantation of mouse hepatocytes showed their ability to replicate at least 69 times without loss of function [91]. Given a stimulus, hepatocytes can rapidly re-enter the cell cycle to restore liver mass and function. Furthermore, the process of cell proliferation is terminated as soon as the liver mass is restored to its original liver-to-body weight-ratio [88,92]. In the event when over 90% loss of the parenchyma occurs, the liver can fail to
regenerate [93]. With the increasing global burden of liver disease, understanding the regulation of liver regeneration promises to identify potential therapeutic targets to promote tissue repair and regrowth after liver injury.

II. Rodent models of liver regeneration

Various rodent models have been developed to study the regenerative response and can be categorized by the stimulus to induce hepatocyte proliferation into surgical-, chemical-, or genetically-induced liver injury paradigms.

A. Surgical-induced liver regeneration

Partial hepatectomy

PHx is the most widely-used technique to study liver regeneration in rodents, in which two-thirds of the liver is surgically removed to induce cell growth and proliferation of the remnant lobes to restore liver mass and function within 10-14 days [94]. Due to the clean removal of the liver lobes, the majority (~95%) of the remaining mature hepatocytes enter the cell cycle in a synchronous fashion that is also species-specific [95]. In both rats and mice, hepatocytes enter G1 4 h post-PHx [96]; DNA synthesis initiates at 12-18 h [97] and proliferation peaks at 24 h in rats [98], while a 20 h lag is observed in mice due to a longer G1 [96]. At 36 h after PHx, approximately 40% of hepatocytes are in S phase [99]. The well-defined time periods for various cell cycle entry points provide the opportunity to study the regulatory mechanisms of adult hepatocyte transitioning from G0 to G1, and from G1 to S in vivo [100]. While PHx reflects what is seen in living-donor liver transplantation, which occurs in less than 5% of transplant cases in the US annually [101], this model does not recapitulate what is seen in human liver diseases that involve inflammatory responses and necrotic cell death [102]. Nonetheless, the high accuracy and reproducibility of the PHx model have allowed understanding of the signaling pathways and transcriptional control that take place during hepatocyte replication. It is worth noting that even though injury responses are
not seen in PHx, the majority of signaling networks underlying the regenerative process are similar within most liver repopulation models [95].

**Ischemia/reperfusion (I/R) injury**

I/R injury occurs due to prolonged low oxygen tension in tissues followed by normalized oxygen perfusion, leading to significant inflammatory responses that cause organ damage and dysfunction, as often seen during liver transplantation or organ hemorrhage [103]. In rodents, I/R injury is modeled via an artery clamp in rats [104] and a lobular clamp in mice [105] to temporarily block the blood supply to the liver. Recovery from I/R injury varies depending on the duration of ischemia, which generally lasts for 30-90 min. A significant injury is frequently observed 12 h later, followed by a peak proliferative response at 48 h, with complete recovery by 96 h [106].

The process of I/R injury can be separated into two phases. In the initial phase, complement triggers Kupffer cell activation followed by the release of ROS that induces oxidative damage to hepatocytes [104]. The hepatic architecture is unchanged in the initial phase, but injured and dying hepatocytes release signals to exacerbate inflammation that feedforward to complement and Kupffer cell activation, leading to the production of tumor necrosis factor α (TNFα) [107] and cytokines such as interleukin 12 (IL12) [108]. In the late phase, the combination of adhesion molecule expression on sinusoidal endothelial cells and secretion of CXC chemokines from nonparenchymal cells result in the recruitment of neutrophils to the liver followed by the release of oxidants and proteases to induce widespread destruction of the parenchyma [109,110]. The late phase of I/R injury thus induces significant changes in the hepatic architecture mainly through necrotic cell death [103].

**Bile duct ligation**

In bile duct ligation, the common bile duct, which drains bile produced in the liver to the small intestine, is irreversibly ligated, resulting in inflammatory responses, obstructive cholestasis, and fibrosis within the first 2 weeks followed by cirrhosis at 4 weeks [111]. Contrary to PHx,
cholangiocytes are the main cell type to regenerate after bile duct ligation as an extensive proliferation of bile duct cells are observed; DNA synthesis begins at 24 h and peaks at 48 h after the surgery [112]. However, when the proliferative capacity of cholangiocytes is impaired, as in the case of treatment with the biliary toxin methylene diamine, mature hepatocytes are able to transdifferentiate into cholangiocytes to rescue the biliary epithelium [113].

B. Chemical-induced liver regeneration

D-galactosamine (GalN)

Intraperitoneal injection of GalN contributes to hepatocyte death through three sources, (1) uridine 5'-diphospho (UDP)-galactosamine derivatives that inhibit RNA and protein synthesis [114], (2) endotoxin accumulation that leads to complement activation and necrotic cell death [115], and (3) mast cell degranulation that causes extensive inflammation [115]. Necrosis is scattered but more prevalent around the pericentral region [116]. Peak plasma alanine transferase (ALT) and aspartate aminotransferase (AST), as well as maximal necrosis, occur at 24 h followed by DNA synthesis that peaks at 48 h after GalN administration [117]. Hepatocyte proliferation, ensuing as early as 24 h post-GalN, is the main source of regeneration, but a contribution of oval cells 48-96 h after drug treatment is observed at higher GalN doses [116].

Carbon tetrachloride (CCl₄)

The main contributors of hepatotoxicity after CCl₄ administration are the zone 3 CYP450 enzymes [118]. CYP2B1, 2B2, and 2E1 [119,120] form reactive metabolites of CCl₄ including trichloromethyl (CCl₃*) and trichloromethyl peroxyl (CCl₃OO*) radicals that modify proteins, lipids, and DNA in hepatocytes, leading to necrotic cell death [121]. Parenchymal necrosis is most prominent in pericentral hepatocytes due to the high expression of CYP450 proteins [118]. The peak of cell death occurs at 24 h, followed by DNA synthesis highest at 36 h [122], and repopulation to replace lost liver mass that is completed within 7-10 days [106]. In addition, long-term CCl₄ exposure promotes fibrosis and even cirrhosis due to stellate cell activation that deposits collagen.
and matrix proteins [123,124]. Thus, repeated CCl₄ treatment is often used as a model for chronic liver injury.

**Thioacetamide (TA)**

TA was initially introduced as a fungicide but quickly realized to be hepatotoxic and carcinogenic [125]. A single-dose administration leads to acute injury [126], subchronic exposure induces fibrosis [127], and chronic use results in liver cancer [125]. Similar to CCl₄, TA toxicity stems from bioactivation through CYP450 enzymes, especially CYP2E1 [128,129] that metabolizes TA to TA-sulfoxide and TA-sulfone, active intermediates that cause necrotic cell death [130]. In addition, toxic TA metabolites induce oxidative stress and lipoperoxidation, leading to the destruction of cell membranes [131,132]. However, the exact mechanisms of TA-sulfone-induced cell death and replication response are not clear, since TA has not been widely used as a hepatotoxin for liver regeneration studies. Hepatic necrosis peaks at 24 h [133] with maximal DNA synthesis at 36-48 h after TA administration [134].

**Acetaminophen (APAP)**

APAP is a common analgesic and antipyretic drug due to its safety and efficacy. In the US, APAP overdose is the most common cause of acute liver failure, accounting for 46% of cases [135]. APAP is normally eliminated by phase II conjugation reactions including glucuronidation and sulfation followed by excretion through the kidneys [136]. However, at toxic doses, the phase II enzymes are saturated and excess APAP is metabolized instead by CYP2E1 to the reactive metabolite N-acetyl-p-benzoquinone imine (NAPQI) [137] followed by depletion of GSH, the cell's primary defense against oxidative damage. Subsequently, excess NAPQI forms covalent bonds with proteins to induce reactive oxygen and nitrogen species such as peroxynitrite [138]. Translocation of JNK [139] and the cell death protein BCL2-associated X protein (BAX) [140] to the mitochondrial outer membrane further induces membrane permeabilization and the release of mitochondrial proteins, eventually leading to severe centrilobular hepatic necrosis [141,142].
Necrosis begins 12 h after APAP overdose, peaks at 36 h \([143]\), and subsides gradually over 60 h \([144]\), whereas hepatocyte proliferation begins at 12 h and peaks at 24-36 h post-APAP administration \([144]\).

C. Genetically-induced liver regeneration

Fumarylacetoacetate hydrolase (FAH) deficiency

The \(\text{Fah}^\text{−}/\text{−}\) mouse is a model of hereditary tyrosinemia type I (HTI), an autosomal recessive disease \([145,146]\). HTI patients are deficient in the enzyme FAH, the last enzyme in the tyrosine catabolic pathway \([147]\), resulting in accumulation of the toxic metabolites succinylacetone (SA), succinylacetoacetate (SAA), fumarylacetoacetate (FAA), and maleylacetoacetate (MAA) that form DNA adducts \([147]\). In addition, FAA depletes intracellular GSH stores \([148]\) and triggers cell cycle arrest in G2/M followed by induction of apoptosis \([149]\). FAA also activates cyclin-dependent kinase 1 (CDK1) and caspase-1 (CASP-1) to induce cell cycle arrest and subsequent expression of CASP-3, resulting in mitochondrial dysfunction as demonstrated by cytochrome c release \([149]\). FAH is primarily expressed in the liver and kidney \([150]\), and at a lower level in endocrine glands and the gastrointestinal tract \([151]\).

The incidence of tyrosinemia is around 1 in 100,000 to 1 in 120,000 worldwide and is considered a rare disease \([152]\) except for Scandinavia and the province of Quebec, where the overall incidence is 1 in 16,000 \([152]\). In particular, the Saguenay-Lac-St-Jean area in Quebec has a prevalence of 1/1,846 \([153,154]\) due to a founder effect \([155]\). To date, more than 35 mutations in FAH have been described \([152]\), including missense mutations leading to 16 amino acid replacements, splice site mutations, and nonsense mutations \([156]\). Interestingly, the pathophysiological phenotype differs between humans and mice, and disease severity also varies greatly between individuals \([152]\).

While an oversimplification, liver phenotypes in HTI patients can be categorized into acute or chronic phases \([157]\). In the acute phase, morphological alterations can vary greatly and include an enlarged or shrunken liver, fibrosis with ductular proliferation in the surrounding region, and
different degrees of hepatocyte steatosis [158]. The most detrimental phenotype in the acute phase is liver crises, repeated episodes of liver insufficiencies due to liver decompensation, and generally manifests as hepatomegaly and coagulopathy [159]. Liver crises are typically present in early infancy before 2 years of age and historically speaking, around 80% of patients died before the age of 2 due to acute liver failure [159,160]. In the chronic phase, cirrhosis is often observed in HTI patients due to the prolonged hepatic injury [161]. There is also an increased risk of developing HCC in patients beyond 2 years of age, ranging from 15% [162] to 37% [163]. In addition, increased frequency of dysplasia, aneuploidy, and variable gene expression are observed in tyrosinemic livers [164]. The exact mechanism of elevated cancer risk is not completely understood but likely stems from the mutagenic environment in the liver due to the accumulation of reactive metabolites FAA, MAA, SA, and SAA, cultivating a milieu in which aberrant growth factors lead to altered gene and protein expression [164]. HTI patients also exhibit other organ dysfunctions including nephromegaly and renal failure [165], painful paresthesias and motor paralysis [166], and occasionally islet hypertrophy and hypoglycemia [167]. Orthotopic liver transplantation was considered the only curative treatment of HTI in the 1980s with an over 90% survival rate [168,169]. The development of 2-(2-nitro-4-trifluoromethylbenzoyl)-1,3-cyclohexanedione (NTBC) [170] gained widespread popularity as early treatment for HTI prevents HCC and circumvents the need for transplantation [171,172]. NTBC is a potent inhibitor of 4-hydroxyphenylpyruvate dioxygenase (HPPD) [173], the second enzyme in the tyrosine catabolic pathway, and thus prevents the production of the toxic intermediates SAA, SA, MAA, and FAA.

The mouse model of HTI was developed via targeted deletion of the FAH gene [145,146]. Fah<sup>−/−</sup> mice recapitulate the major biochemical and phenotypic alterations observed in HTI patients such as hypertyrosinemia, accumulation of SA, liver failure, renal tubular damage, and occasional tumorigenesis [174]. Interestingly, the Fah<sup>−/−</sup> mouse exhibits a much more severe liver phenotype than HTI patients as mice die within 12 h after birth from fulminant liver failure and hypoglycemia [145], likely attributed to higher levels of toxic metabolites, lower GSH concentrations, and increased sensitivity to FAA, MAA, and SAA in neonatal mice compared to humans [145,175]. Fah<sup>−/−</sup>
mice require early treatment with NTBC to survive beyond birth and a portion of FAH-deficient livers still develop tumors despite long-term NTBC administration, possibly resulting from suboptimal NTBC doses or other metabolic pathways that produce FAA and MAA found only in mice [145].

Fah–/– mice undergo liver repopulation by viral- or nonviral-mediated gene therapy to restore FAH function, transplantation with FAH-positive hepatocytes, or genome editing to correct Fah mutations. Liver injury is induced upon NTBC withdrawal, and FAH-negative hepatocytes experience inflammation, necrosis, or apoptosis [145,174]. Only hepatocytes with FAH expression are selected in vivo to proliferate and repopulate the injured liver parenchyma [176]. Transplantation with wild type hepatocytes revealed the competitive growth advantage of FAH-positive cells to repopulate the mutant liver, as injection of as few as 1,000 hepatocytes successfully rescued the phenotype of FAH deletion, requiring on average an estimated 16 cell doublings to restore liver mass [176]. Liver repopulation can also be carried out through retroviral induction of FAH expression ex vivo in Fah–/– hepatocytes followed by transplantation of the transduced cells [177]. Gene transfer via retrovirus [176], adenovirus [178], and AAV2 or 8 [179] also results in significant colonization of FAH-expressing hepatocytes. However, 9 out of 13 mice treated with adenovirus developed HCC after 9 months of liver repopulation from untransduced cells that constitute less than 10% of the liver [178]. Additionally, DNA-mediated transposition with the Sleeping Beauty (SB) transposable element is able to achieve permanent transgene expression through genomic integration from the plasmid containing FAH complementary DNA (cDNA) following hydrodynamic tail-vein injection [180]. Less than 0.1% of hepatocytes display integration and repopulate the liver to reverse the lethal phenotype after NTBC removal [180]. More recently, the clustered regularly interspaced short palindromic repeats (CRISPR) system has been utilized in Fah–/– mice that harbor a point mutation [181]. Expression of single guide RNAs (sgRNA) and a repair template was successful in directing the CRISPR-associated protein 9 (CAS9) nuclease to produce a targeted, double-stranded DNA break followed by homologous recombination to repair the Fah gene defect [181].
Since only FAH-positive hepatocytes undergo clonal expansion to repopulate the injured liver, the \( \text{Fah}^+ \) mouse also provides a remarkable tool to lineage-trace regenerating hepatocytes by tracking FAH-expressing cells. Coexpression of markers such as luciferase [182] and GFP [183] can be utilized to specifically trace and isolate repopulating hepatocytes for phenotypic studies. Furthermore, gene-activating or -silencing molecules can be tethered to FAH expression to functionally identify the significance of multiple genes during liver regeneration, including the use of small hairpin RNA (shRNA) [184], cDNA [185], gRNA [186], and tough decoy (TuD) microRNA (miRNA) inhibitors [187]. \( \text{Fah}^+ \) immunodeficient mice are also used to grow billions of human hepatocytes (Azuma et al. 2007).

**Urokinase plasminogen activator (uPA) overexpression**

uPA is a fibrinolytic enzyme that transforms plasminogen into plasmin to remove fibrin clots [188]. Hepatocyte-specific uPA overexpression regulated under the albumin enhancer and promoter leads to increased plasma uPA levels, fibrinogen depletion, followed by neonatal death within 3 days postpartum due to bleeding of the abdominal cavity and intestinal tract [189]. Hepatocytes that silenced uPA expression, mediated mainly by intrachromosomal recombination, are able to achieve complete regeneration of uPA transgenic mice [190]. Similar to the case in the \( \text{Fah}^+ \) model, loss of transgene expression in individual hepatocytes confers a selective advantage so that clonal expansion of the uPA-normal cells reconstitute the liver parenchyma [190]. The uPA transgenic mice can also be corrected with transplantation of wild-type hepatocytes that undergo an estimation of 12-18 rounds of replication [191]. Furthermore, xenogenic cell transplantation from rat [192] or human [193] hepatocytes can be performed in immunodeficient mice carrying the uPA transgene to generate chimeric livers. The uPA transgenic model, therefore, serves as an excellent paradigm to study liver regeneration after cell transplantation [190,191], drug metabolism in the chimeric human livers [194], and liver diseases including HBV [195] and HCV [193].
D. Models to study HPCs

As introduced above, liver regeneration is carried out by the proliferation of preexisting mature hepatocytes under normal physiological conditions [97]. Only when the replicative ability of resident liver cells is severely hindered, in the case of drug treatment or chronic liver injury, will HPCs be called into action to regenerate the injured parenchyma [89]. One source of HPC is the ‘oval cells’, small cells with oval nuclei that emerge during chemical hepatocarcinogenesis [89] or hepatotoxin-induced injury in rodents [196]. Since oval cells display intermediary phenotypes and histology between hepatocytes and cholangiocytes, their activation is also referred to as ‘ductular reaction’ [197]. The origin and contribution of oval cells have long been contested due to the lack of genetic lineage-tracing evidence; thus, the progenitor/descendant relationship was mostly inferred from their spatial proximity to hepatocytes or cholangiocytes [198,199]. Mature hepatic cells are other sources of HPC as hepatocytes and cholangiocytes are able to assume the role of facultative stem cells for one another and transdifferentiate into the other parenchymal cell type.

2-acetylaminofluorene (2-AAF) + PHx

2-AAF is a carcinogen that causes DNA damage and prevents DNA synthesis in hepatocytes, leading to the development of liver cancer [200]. Treatment of dietary 2-AAF for 2 weeks followed by PHx has been used as a model to induce ductular reactions in rats, as hepatocytes are unable to undergo cell replication [201]. Activation of 2-AAF is mediated by N-sulfotransferase to generate the active N-OH 2-AAF radical that translocates into the nucleus to induce DNA damage [202]. However, N-sulfotransferase is not expressed in mice, thus limiting the use of 2-AAF as an inhibitor of DNA synthesis in this species [202]. Interestingly, isotope labeling to track DNA synthesis observed labeled ovals cells but few hepatocytes, indicating that oval cells do not become hepatocytes after AAF-induced liver injury [201]. Another later experiment, however, identified labeled ovals cells and subsequently hepatocytes, establishing a precursor-product relationship between the two cell types during liver regeneration [203]. 2-AAF is also used in
combination with CCl₄ or allyl alcohol in rats to cause centrilobular or periportal damages that induce oval cell activation [204].

**3,5-diethoxycarbonyl-1,4-dihydro-collidine (DDC) diet**

DDC prevents heme biosynthesis and causes the accumulation of protoporphyrin, leading to severe porphyria and liver injury [205]. Treatment with 0.01% DDC for 2-4 weeks activates a ductular reaction and prolonged treatment results in liver cancer in mice [206]. Currently, there is no consensus on HPC markers in the DDC model but induction of HPC markers including A6 [207], CK19 [208], epithelial cell adhesion molecule (EpCAM) [209], and FOXL1 [210,211] have been reported.

**The choline-deficient, ethionine-supplemented (CDE) diet**

Another model to induce a ductular reaction is the CDE diet, in which a choline-deficient diet supplemented with 0.05-0.15% (w/v) ethionine mixture in the drinking water is provided to rats [212] or mice [213] for up to 4 weeks. Choline is a lipotropic factor important for the secretion of very-low-density lipoprotein (VLDL) [214]; a deficiency in choline causes intracellular lipid accumulation and hepatocyte membrane rupture, ultimately leading to steatosis followed by cirrhosis [215,216]. When combined with the potent hepatocarcinogen ethionine, activation of oval cells can occur followed by induction of liver cancer [212].

**III. Mechanisms of liver regeneration**

The majority of mature hepatocytes reside in the reversible, nonreplicative G0 phase under homeostasis [28]. Upon injury, liver cells re-enter the cell cycle and progress through G1, S (DNA synthesis), G2, and M (mitosis) phases. The cell cycle is tightly-controlled by cyclin proteins, in which the levels rise and fall to activate downstream target CDKs that control progression through various cell cycle checkpoints (Figure 1.2) [217,218]. The cyclin D-CDK4/CDK6 complex is the first to be detected followed by cyclin E-CDK2 formation to promote G1/S transition [219–221]. Next,
cyclin A-CDK1/CDK2 is activated to regulate S phase along with cyclin B-CDK1 assembly to modulate G2/M entry [222,223]. Regenerating hepatocytes are highly synchronous in the PHx model [224], whereas in other paradigms, hepatocytes traverse the cell cycle in a non-synchronous order [95]. While most studies regarding the mechanisms underlying liver regeneration are conducted in the PHx model due to its popularity, numerous signaling pathways have been shown to be important for other models as well. Regardless of the source of injury, liver regeneration is composed of three distinct phases, the initial ‘priming’ phase where hepatocytes acquire an enhanced ability to replicate [225], the second ‘progression’ phase that allows hepatocytes to proceed through the cell cycle to recreate an adequate cell number and mass [226–228], and the final ‘termination’ phase where liver cell proliferation is stopped once liver mass has returned to normal [86,228,229].

A. Priming

The priming phase is the initiating event in which terminally-differentiated hepatocytes acquire enhanced replicative ability that allows for the transition from a quiescent state (G0) to a competent state (G1) [225,226]. Cytokines released from non-parenchymal cells in the liver act as paracrine factors to promote signaling pathways in hepatocytes [228]: TNFα [230,231] and IL6 [232] are essential cytokines secreted by Kupffer cells in the early signaling phase. Pretreatment with TNFα increases the proliferative response to growth factors in rats [233], while administration of TNFα antibodies [231] as well as deletion of type I TNFα receptor (TNFR1) [234] inhibits DNA replication and impairs liver regeneration. TNFα activates nuclear factor kappa B (NF-κB) both in Kupffer cells to increase IL6 transcription [234] and in hepatocytes to activate cell proliferation [235].

IL6 is a proinflammatory cytokine that mediates the acute-phase response [236]. During the initial phase of liver injury and repopulation, IL6 is secreted from Kupffer cells due to stimulation by TNFα [231,234]. IL6 binds to its receptor glycoprotein 130 (gp130) [237] to activate transcription factors, usually within the first hour of PHx, including NF-κB, activator protein 1 (AP-1), signal transducers and activators of transcription 3 (STAT3), and CCAAT enhancer-binding protein β
(C/EBPβ), ultimately leading to the expression of immediate-early genes such as \textit{Jun}, \textit{Fos}, and \textit{Myc} \cite{238,239}. Later studies identified induction of as many as 73 immediate-early genes during the priming phase \cite{240} and another study reported that almost 40\% of immediate-early genes are induced via IL6 \cite{241}. Deletion of IL6 leads to a decrease of immediate-early gene expression followed by a 70\% reduction of DNA synthesis \cite{232}.

The importance of TNFα and IL6 as priming factors can be replicated in other regeneration paradigms including I/R injury \cite{242}, CCl4 \cite{243,244}, and APAP hepatotoxicity \cite{245,246}, in which downstream activation of pathways including NF-κB and STAT3 induces expression of immediate-early genes to promote cell cycle entry. Nonetheless, controversies regarding the role of cytokines during the priming phase arise when conflicting results were found. TNFα-deficient mice do not exhibit reduced DNA synthesis or delayed regeneration after PHx \cite{247,248}. Similarly, studies utilizing IL6- or gp130-deleted mice have demonstrated that IL6 does not mediate cell cycle entry but activates adaptive responses and apoptosis to fine-tune the regenerative process \cite{237,249,250}.

\textbf{B. Progression}

Hepatocytes acquire proliferative competence after priming and transition from G1 to S phase \cite{226–228}. Commitment to progress through the cell cycle is mediated through early G1 exposure to growth factors including epidermal growth factor (EGF) \cite{251}, TGFα \cite{252,253}, and HGF \cite{254}. These factors are also known as ‘direct mitogens’ due to their ability to independently induce cell growth in cultured hepatocytes. Furthermore, infusion or overexpression of EGF \cite{255}, HGF \cite{256}, or TGFα \cite{257} triggers hepatocyte proliferation and liver enlargement in normal rats. Both HGF and EGF stimulate liver regeneration as paracrine factors \cite{258,259}, while TGFα promotes hepatocyte replication in an autocrine fashion \cite{253}.

HGF is produced primarily by stellate cells \cite{260}, activated by urokinase \cite{261}, and released from the extracellular matrix during liver regeneration \cite{259}. Plasma levels of HGF are increased by 20-fold as early as 1 h after PHx in rats \cite{262}. HGF binds to the receptor tyrosine
kinase HGF receptor (MET) [263] to induce TGFα synthesis in hepatocytes [264], as well as to stimulate proliferation and survival pathways such as mitogen-activated protein kinase (MAPK) and phosphoinositide 3-kinase (PI3K) signaling directly, leading to the activation of transcription factors ETS domain-containing protein (ELK1), MYC, and C/EBPβ [265,266]. Inhibition of HGF activation, including deletion of urokinase [267], administration of an anti-HGF antibody [268], and conditional disruption of Met [269,270] causes an impaired regenerative response after PHx and hepatotoxin-induced liver injury.

EGF is produced by the Brunner’s gland in the duodenum [271] and continuously supplied to the liver through the portal circulation [258]. Interestingly, no significant change in plasma EGF concentration was detected after PHx [258], with another study suggesting activation of the sympathetic system could increase EGF production via norepinephrine [272]. Alternatively, the removal of two-thirds of the liver mass has been suggested to increase the load of EGF per hepatocyte by 3-fold [86,229].

TGFα is released from the hepatocyte plasma membrane by the TNFα-converting enzyme (TACE) that is activated through secreted TNFα during the priming phase [273]. The transcript level of TGFα increases after 2-3 h, peaks at 12-24 h, and persists for 48 h post-PHx [253]. Both EGF and TGFα activate the receptor tyrosine kinase EGF receptor (EGFR) [274] to induce proliferation and prevent apoptosis via activation of MAPK, PI3K, and STAT pathways [265,266,275]. Treatment with antisense oligonucleotides and antibodies against TGFα reduces the number of replicating hepatocytes [264,276]. However, the redundancy of EGFR ligands has made it difficult to definitively demonstrate the requirement of EGF and TGFα for regeneration. Tgfa−/− mice do not show a significant decrease in DNA synthesis or a delay in regeneration after PHx [277], possibly due to EGF compensation. However, mice with targeted EGFR deletion display a delayed G1 to S phase transition and a decreased expression of cyclin D1 after PHx [278], documenting the importance of EGFR signaling during the progression phase.
C. Termination

After cell growth and proliferation to restore the loss parenchyma are complete, liver regeneration stops through the activation of termination signals. Most research has focused on the mechanisms that induce hepatocyte replication and less is understood about the termination process. Additionally, the connections between the termination of liver regeneration and the pathogenesis of liver cancer have equally attracted attention.

TGFβ is the most well-known factor to repress hepatocyte proliferation. Produced by stellate cells [279], TGFβ is normally sequestered in the extracellular matrix [280]. Many mechanisms are implicated to release TGFβ during liver regeneration, but no direct evidence has been provided [281], suggesting multiple factors could be at play to exert the tight regulation of TGFβ localization. TGFβ mRNA is increased within 3-4 h and peaks at 48-72 h after PHx in rats [282]. Contrarily, all three TGFβ receptor subtypes are downregulated at the transcript and protein levels following PHx and only recover at 120 h [283], causing a decreased TGFβ sensitivity in regenerating hepatocytes isolated after 24-72 h post-PHx [284]. Resistance towards TGFβ via norepinephrine modulation could explain the observation that hepatocytes are able to continue DNA synthesis until 72 h post-PHx despite increased TGFβ levels [285]. Cascades of the ‘small mothers against decapentaplegic’ (SMAD) proteins are activated upon TGFβ receptor phosphorylation [286,287] to increase expression of CDK inhibitor p15 [288] and prevent assembly of cyclins and CDK complexes, including cyclin E-CDK2 [289,290] and cyclin D-CDK4 [290], leading to reversible cell cycle arrest at G1 [289]. TGFβ is a potent inhibitor of EGF-induced DNA synthesis in cultured rat hepatocytes [291], and infusion of TGFβ reversibly prevents hepatocyte replication by over 60% at 24 h that returns to normal by 72 h after PHx in rats [292]. Conflicting studies, however, have questioned the significance of TGFβ as a termination factor. Overexpression of TGFβ1 under control by the albumin promoter causes hepatic fibrosis due to collagen deposition but does not affect liver regeneration [293]. Similarly, conditional removal of the TGFβ type II receptor increases cyclin D and E expression, allowing accelerated S phase entry to enhance hepatocyte proliferation, but no effect on the termination of regeneration was observed.
Activin, another member of the TGFβ family, is also implicated in the termination phase as an autocrine agent that inhibits DNA replication, as its pharmacological inhibition with follistatin promotes DNA synthesis and leads to hepatomegaly following PHx in rats [295,296].

C/EBPα is a hepatocyte-enriched transcription factor that regulates the expression of multiple liver-specific genes [297,298]. C/EBPα is transcriptionally regulated during liver regeneration with a decrease in gene expression by 60-80% 1-3 h [299] and remains repressed until 24 h after PHx [300]. The drop of Cebpa transcripts during hepatocyte replication echoes previous observations of its antiproliferative quality in terminally-differentiated adipocytes [301]. C/EBPα induces cell-cycle arrest through various pathways including stabilization of the CDK inhibitor p21 to disrupt formation of cyclin-CDK complexes [302,303], modulation of growth-inhibiting E2F-RB complexes [304,305], direct inhibition of CDK2 and CDK4 [306], induction of proteasome-dependent degradation of CDK4 [307], and others [308]. Cebpa−/− hepatocytes exhibit increased DNA synthesis in culture and rapidly form proliferative nodules when inoculated into mice [309]. Similarly, Cebpa−/− mice display elevated transcript levels of Jun and Myc, as well as an increased hepatocyte proliferation [310].

Another important regulator of liver mass is the Hippo pathway, named after the Drosophila hippo gene, which encodes a kinase that regulates organ growth, cell proliferation, and developmental apoptosis [311,312]. The mammalian orthologs ‘mammalian sterile 20-like 1’ (MST1) and 2 kinases [311] activate the large tumor suppressor 1 (LATS1) and 2 proteins to phosphorylate and inhibit the activity of the transcriptional coactivator yes-associated protein (YAP), leading to its nuclear export and protein degradation [313,314]. The Hippo signaling pathway is altered after PHx, with increased nuclear localization of YAP by 4 h [315] and decreased kinase activity of MST1 and MST2 1-3 days post-PHx [316]. Upon activation, YAP increases the expression of genes involved in hepatocyte proliferation such as Ki67, Myc, and H19 [313]. Conditional YAP activation induces liver overgrowth by over 50% after 1 week, and persistent YAP elevation causes tumorigenesis [313,317]. Similarly, deletion of MST1 and MST2 in mouse livers leads to loss of YAP phosphorylation and nuclear retention, followed by hepatomegaly and HCC.
The mechanism of YAP inactivation at the end of liver regeneration is not fully understood but is hypothesized to include regulation by components of the extracellular matrix such as integrin-linked kinase (ILK) [319,320] and glypican-3 [321,322] to prevent nuclear localization of YAP in both PHx and toxin-induced liver injury.

D. Other factors to consider

Liver regeneration depends on a complex regulatory network that includes multiple additional soluble mediators, signaling pathways, and transcription factors not discussed above. (1) Growth factors including vascular endothelial growth factor (VEGF) [323,324], platelet-derived growth factor (PDGF) [325,326], FGF [327,328], and heparin-binding EGF-like growth factor (HB-EGF) [329]. (2) Extracellular signals such as bile acids [330], serotonin [331], insulin [332,333], norepinephrine [334,335], complement [336], and CXC chemokines [337,338]. (3) Pathways such as Wnt/β-catenin [339,340] and Notch/Jagged [341] signaling. (4) Growth-promoting nuclear receptors including [342] retinoid X receptor (RXR) [343], PPARα [344,345], farnesoid X receptor (FXR) [330], and pregnane X receptor (PXR) [346]. (5) Growth-inhibiting nuclear receptors [342] such as PPARγ [347,348] and HNF4α [349]. (6) Other factors such as microRNAs [350].

These signaling pathways have been implicated in multiple liver regeneration models and manipulation of the pathways generally shows consistent outcomes across various paradigms with some exceptions. For instance, CXC chemokines promote regeneration after PHx [337] but inhibit hepatocyte proliferation during I/R [351]. This could be due to the difference in CXC concentration — 10 times higher following I/R — suggesting that a moderate increase of CXC chemokines promotes but higher expression inhibits hepatocyte proliferation [352]. Another example is TNFα signaling, in which TNFR1 deletion prevents DNA synthesis and delays liver regeneration after PHx [353], but an overexpression cDNA screen identified TNFR1 to be a potent repressor of liver repopulation in the Fah-/- mouse [185]. The difference in TNFR1 expression levels, TNFα signaling activation, injury duration, or the inflammatory context could explain the divergent findings [185]. Finally, it is also worth noting that no single gene deletion results in complete abrogation of the
regenerative process, but at most causes a reduction of replicating hepatocytes and a delay of cell cycle progression, suggesting substantial redundancy of the diverse signaling pathways modulating liver regrowth [354].

IV. Source of regenerating hepatocytes

Regenerating hepatocytes can arise from three main sources in a context-dependent manner: expansion of preexisting hepatocytes, differentiation of oval cells, or transdifferentiation from cholangiocytes. As discussed above, resident hepatocytes are the first responders to replenish the hepatocyte pool but oval cells and cholangiocytes have been suggested more than sixty years ago as contributors of liver regeneration when DNA synthesis is severely impaired in hepatocytes. During the past decade, several lineage-tracing studies in mice have attempted to qualify the contribution of various proposed progenitor cells to hepatocyte regeneration. Using the Sox9\textsuperscript{CreERT2} system to trace all cholangiocytes, 1-2% of regenerating hepatocytes were lineage-labeled after CCl\textsubscript{4}, APAP, and DDC administration, indicating that only a small percentage of BECs contribute to liver regeneration under these settings [30,34]. A similar conclusion was drawn when oval cells are labeled with osteopontin (OPN) after CDE diet but not with CCl\textsubscript{4} or DDC treatment [355]. Other studies have suggested that FOXL1-expressing HPCs give rise to hepatocytes after a DDC-supplemented diet [211], or that LGR5-positive organoids can repopulate \(Fah^+\) mouse livers after transplantation [356]. Interestingly, using a Krt19\textsuperscript{CreERT2} system in conjunction with DDC or CDE treatment, no label-bearing hepatocytes arise after regeneration, suggesting that all new hepatocytes come from preexisting liver cells [357]. This was likely due to limited injury which did not completely block hepatocyte proliferation, thus deviating the need for replacement from cholangiocytes acting as facultative progenitors. However, when hepatocyte regeneration is completely abrogated by severe liver injury, inhibition of hepatocyte replication, or induction of hepatocyte senescence, a large contribution of cholangiocyte-derived hepatocytes to the regenerating liver has been documented [358–360]. As it turns out, mouse hepatocytes are more resilient than those from rats in terms of retaining proliferative potential in the face of liver injury,
possibly explaining some of the discrepancies in the literature regarding the importance of cholangiocytes and oval cells as facultative hepatocyte.

V. Clinical implications

To date, the only curative treatment for end-stage liver disease is liver transplant. While it is the second most common solid organ transplantation after kidney transplant, less than 10% of global liver transplantation needs are met at current rates [37]. Multiple strategies under active research for the treatment of liver disease [361,362] include (1) transplantation of primary hepatocytes or induced pluripotent stem cells (iPSCs), (2) induction of endogenous hepatocyte replication via pharmacological agents, cytokines, or growth factors, (3) bioartificial livers that incorporate hepatocytes into a dialysis-based artificial system to carry out the main metabolic functions of the liver, and (4) organ bioengineering that utilizes a xenogenic scaffold infused with mature hepatocytes to produce a functional liver graft. With the rising healthcare burden of chronic liver disease, a deeper understanding of the mechanisms underlying liver repair and regrowth will enable a broader utilization of regenerative medicine.
LIVER TRANSCRIPTIONAL CONTROL

I. Overview

Gene regulation is fundamental for all organisms; in particular, the complexity of eukaryotic transcriptional control allows intricate regulation of expression patterns to adapt to environmental queues. Transcriptional regulation occurs on several levels. (1) Cis-regulatory modules provide essential information for transcription factor binding and serve as a platform for the assembly of regulatory complexes. (2) Chromatin architecture including nucleosome patterning, histone modification, and DNA methylation impacts the accessibility of the transcriptional machinery. (3) Intra- and interchromosomal interactions establish topological hotspots for long-range regulation of gene expression [363].

A. Transcription factors

Transcription factors are trans-acting proteins that bind to cis-regulatory modules at the promoter or enhancer to activate or repress transcription [364]. Regulation of gene expression is achieved through various mechanisms including stabilization or blockade of RNA polymerase II [365], direct or indirect modification of chromatin structure [366], and recruitment of coactivator or corepressor proteins to the protein-DNA complex [367]. Recruitment of transcription factors to target sites is established mainly through the structure and sequence of the DNA-binding domains [363]. Evolutionarily-related transcription factors often share similar DNA recognition motifs and demonstrate binding redundancy. The specificity of transcription factors is determined by (1) cooperative or competitive binding with other regulatory proteins [368,369], (2) flexibility of the DNA-binding domain to recognize noncanonical motifs with mechanisms not yet clear [370], and (3) posttranslational modifications to affect subcellular localization, protein-protein interactions, and DNA binding activity of the transcription factors [371]. These mechanisms allow spatiotemporal binding of transcription factors to fine-tune eukaryotic gene transcription and establish distinct gene expression patterns.
B. Chromatin architecture

DNA is compacted into chromatin in the eukaryotic genome. The basic unit of chromatin is the nucleosome, which consists of 147 base pairs (bp) of DNA tightly wrapped around a histone octamer with two copies of core histones H2A, H2B, H3, and H4 each [372]. Chromatin is historically categorized into one of two states based on its accessibility to the transcriptional machinery. Heterochromatin is highly condensed, transcriptionally inactive, and associates with repressive histone modifications, whereas euchromatin is relatively accessible to transcriptional complexes, marked with active histone modifications, and contains actively-transcribed genes [373,374]. Diverse mechanisms contribute to the modification of chromatin structure and subsequent changes to DNA accessibility, including nucleosome positioning and occupancy via ATP-dependent chromatin remodelers [375,376], DNA methylation through DNA methyltransferases (DNMT) and demethylation via ten-eleven translocation (TET) proteins [377], and epigenetic modifications of the core histones such as methylation, acetylation, phosphorylation, and ubiquitination [378].

C. Three-dimensional structure

The three-dimensional structure of chromatin provides an additional layer of transcriptional control through the regulation of nuclear organization and chromosomal interactions.

(1) Nuclear organization. The nucleus is divided into functional domains in which chromosomes occupancy at different regions is associated with divergent transcriptional activity [379]. The non-uniform compartment of the nuclear interior enables highly-organized structures to establish chromatin territories based on gene activity and density; gene-rich regions are typically located in the nuclear center and gene-poor chromatin in the periphery [380].

(2) Chromosomal interactions. The identification of cis-regulatory modules located far from the promoters they regulate has led to the discovery of looping as the predominant mechanism for enhancer-promoter interactions [381]. Long-range chromatin communication includes interactions between regulatory sequences of a single locus [382], among elements within a gene complex.
Technological advances utilizing proximity-ligation followed by deep-sequencing and high-resolution microscopy methods will further allow the elucidation of inter- and intrachromosomal interactions to establish cell type-specific transcriptional regulation.

II. Discovery of liver transcriptional control

A. Liver-specific gene expression

Studies on transcriptional control were pioneered by the laboratory of James Darnell starting in the early 1980s. Rat liver nuclei became one of the first mammalian systems used to investigate gene expression regulation due to the large number of available cells and the relatively pure cell types, where hepatocytes constitute approximately 80% of the liver mass and 60% of liver cell number. Liver-specific gene sequences were isolated via the extraction of polyadenylated RNA, reverse transcription into cDNA, and recombination with antibiotic-resistant plasmids. DNA from individual colonies was used as a template to hybridize with nascent, radiolabeled mRNA isolated from rat liver nuclei. The hybridization signals from liver nuclear RNA is at least 10 times stronger than those from non-liver cells, suggesting a differential abundance of tissue-specific mRNA. Further analysis to compare the transcription rate of liver-enriched mRNA in liver and brain nuclei revealed that transcriptional regulation plays a primary role in establishing differential gene expression in various terminally-differentiated cell types.

B. Liver-specific regulatory regions

Analysis of liver-enriched genes in human and rodent ensued, unveiling tissue-specific regulatory regions including proximal promoters and distal enhancers that control the expression of cell type-specific genes. For instance, the 5' flanking sequences of rat Alb drives efficient expression of a reporter gene, chloramphenicol acetyltransferase, preferentially in ALB-expressing hepatoma cells. The promoter-proximal region of human SERPINA1 that encodes the α1-
antitrypsin (AAT) enzyme is sufficient for the transcription in Hep3B, a human hepatoma cell line, but not in HeLa cells [392]. Similarly, expression of distal enhancers of the mouse Ttr, encoding transthyretin (TTR), activates β-globulin transcription specifically in human hepatoma cells HepG2 but not in HeLa cells [393]. These observations suggest a combination of cis-regulatory sequences and trans-acting factors in particular cell types establishes tissue-specific expression regulation.

In addition, two hypotheses concerning cell type-specific transcriptional control through trans-acting proteins emerged [392,394]. It was proposed that activating factors expressed only in particular tissues govern gene expression. However, some liver-enriched genes are also expressed in other cell types, such as Serpina1 in macrophages [395], certain apolipoproteins in the gut [396], and Ttr in the choroid plexus [397]. This implicates that the activating mechanism in addition to the distribution of the activating factors exhibits liver specificity [394]. Furthermore, it was later shown that most, if not all, so-called 'liver-specific' transcription factors are also expressed in other cell types [398–401]. Another line of hypothesis suggested the presence of inhibiting factors to prevent gene expression in specific tissues [392,394]. This implies that a repressor is required for each gene to be not expressed in a certain cell type, suggesting a requirement for a large number of negative factors to restrict transcription in non-expressing tissues. Thus, a more plausible explanation of tissue-specific transcriptional control is that a unique combination of several liver-enriched positive and negative trans-acting factors modulate expression in a cell type-specific manner [402]. Various liver-enriched, but not necessarily liver-specific, activators induce expression [402,403] and repressors inhibit transcription to establish a liver-specific gene expression profile [404].

C. Liver-specific transcription factors

From the late 1980s to 1990s, the use of DNA sequence affinity chromatography with rat liver nuclear extracts enabled identification of several transcription factors highly expressed in the liver, collectively referred to as hepatocyte nuclear factors [398,403,405,406]. DNA sequences of liver-enriched genes suspected to encompass transcription factor binding sites were used as bait
to isolate protein-DNA complexes followed by high-performance liquid chromatography (HPLC) to purify the protein of interest. The partial amino acid sequences of the protein peptides were determined and used to design primers for PCR amplification from hepatoma cell lines. The amplified products were then used to screen rat cDNA libraries to obtain clones that encode the gene sequence of each liver-enriched transcription factor.

III. Hepatocyte nuclear factors

A. HNF1

HNF1 was identified as a nuclear protein that binds to the promoters of fibrinogen α and β chains, as well as AAT in hepatocytes [405]. The HNF1 subfamily contains two isoforms, HNF1α and HNF1β. Analysis of the ALB promoter established the requirement of the albumin proximal factor (APF), later found to be HNF1α, for Alb transcription [407]. HNF1α was initially only detected in differentiated rat hepatoma cells whereas HNF1β, originally identified as modified APF (vAPF), was observed in two dedifferentiated rat hepatoma cell lines [407]. Later studies confirmed expression of HNF1α and HNF1β in the liver, pancreas, intestine, and kidney [399,408], with HNF1α detected at much higher levels and HNF1β lower except for in the kidney [409,410]. HNF1β could exhibit broader physiological implications as it is also observed in the lung, testis, and ovary [401,411].

HNF1 is a member of the POU homeobox gene family [412]. The N-terminal contains the dimerization domain that allows the homeoproteins to dimerize [409], the DNA-binding domain consists of a bipartite POU homeodomain [412], and the C-terminal includes different transactivation domains less conserved within the subfamily, in which HNF1α demonstrates a higher transactivation potency than that of HNF1β [413]. Analysis of the promoters of various liver-enriched genes from rat, mouse, and human predicted the HNF1 consensus binding sequence as GTTAATNATTAAC [414]; both isoforms share the same DNA-binding motif with different transcriptional activity [415]. The homeoproteins form homo- and heterodimers within the subfamily [409,410], where HNF1α is able to dimerize without binding to the DNA recognition sequence.
Additionally, a dimerization cofactor of HNF1α, DCoH, selectively stabilizes the homodimers and assembles a tetrameric complex to enhance the trans-activating ability of HNF1α [416].

Sequence homology analysis determined additional HNF1α target genes including *Alb* [407], *Ttr* [414], *Afp* [414,417], *Apoa2* that produces apolipoprotein A-II (ApoA-II) [418], and *ApoB*, encoding the protein ApoB-100 [419]. Genome-wide analysis of HNF1α footprinting with chromatin immunoprecipitation (ChIP) followed by high-throughput sequencing (ChIP-seq) further identified HNF1α targets crucial for liver synthetic functions, such as carbohydrates, cholesterol, apolipoproteins, CYP450, and serum proteins [420]. Similar to HNF1α, HNF1β occupies the *Alb* proximal promoter to activate transcription [409]. Additional HNF1β targets determined through loss of HNF1β include *Slco1a1* that encodes the member 1a1 of the solute carrier organic anion transporter family (OATP-1) for bile acid reabsorption, and *Acadvl*, the very long-chain-acyl-Coenzyme A dehydrogenase (VCLAD) required for fatty acid oxidation [421].

During development, HNF1β is first detected on E4.5 in the endoderm of the foregut, while HNF1α expression is activated later on E8.5 in the yolk sac [422]. HNF1α and HNF1β are also present on E10.5 after the initiation of hepatocyte lineage in the liver primordia and continue to be expressed throughout embryonic development [423]. These observations suggest that HNF1β participates in the initial transcriptional activation of genes in the visceral endoderm, and the later activation of HNF1α could be required to maintain target gene expression for liver function [422]. *Hnf1a* transcript levels gradually decrease at the late period of embryonic liver development while *Hnf1b* increases [424].

*Hnf1a<sup>−/−</sup>* mice die around weaning due to hepatic dysfunction, phenylketonuria, and renal Fanconi syndrome [425]. HNF1α-deficient livers are enlarged with decreased *Alb* expression but a compensatory increase of HNF1β partially rescues the expression of ALB, AAT, and fibrinogen [425]. Since HNF1α deletion is not embryonically lethal, it is likely not required for specification of the hepatocyte cell lineage but important for the expression of differentiated liver function genes.

HNF1β-deficient mice die by E7.5 due to the lack of extraembryonic endoderm development [426]. *Hnf1b<sup>−/−</sup>* tetraploid complementation established that HNF1β activity is required
for visceral endoderm differentiation to direct the expression of HNF4 and other endoderm marker genes [408]. Further tetraploid embryo complement analysis showed that Hnf1b+/ mice do not form the hepatic bud and lack expression of liver-enriched genes [427]. Conditional HNF1β ablation using the Hnf1b<sup>F/F</sup>;Alfp<sup>Cα</sup> mouse causes abnormal gallbladder and intrahepatic bile duct formation, resulting in severe growth retardation and jaundice; liver metabolism was also affected, with downregulation of genes involved in bile acid sensing and fatty acid oxidation [421]. These studies implicate that HNF1β is required for endoderm commitment, hepatic specification, and bile duct morphogenesis during liver organogenesis.

Altogether, HNF1 proteins are important in establishing mature hepatic functions and appropriate bile duct differentiation. Interestingly, heterozygous human HNF1 mutations do not cause abnormalities in the liver but rather dysfunctions of the pancreatic islets; HNF1α mutations lead to autosomal dominant maturity-onset diabetes of the young type 3 (MODY3) [428] and HNF1β mutations result in MODY5 [429].

B. HNF3/FOXA3

HNF3 proteins were described due to their ability to occupy TTR and AAT promoters at sites distinct from HNF1 and C/EBP [403]. Ttr contains two recognition sequences for HNF3 within 150 bp upstream of the transcriptional start site. Mutation of the most 3' HNF3 binding site results in decreased Ttr expression, despite all other enhancer and promoter sequences being intact, indicating the importance of HNF3 for Ttr transcriptional activation [403]. The HNF3 family consists of three members identified from the purification of distinct protein-DNA complexes that bind to the mouse Ttr promoter, HNF3α, 3β, and 3γ [430–432]. The DNA-binding domains of all three members are highly conserved and share 90% amino acid similarity that matches the sequences of the Drosophila Fox nuclear protein [433]. Therefore, HNF3 proteins were renamed according to the nomenclature of all forkhead transcription factors to ‘FOXA’ [434]. The FOXA proteins are functionally redundant in the liver [435], of which FOXA3 exhibits the highest expression in adult hepatocytes [436].
FOXA proteins are members of the FOX family [430] and belong to the FOXA subfamily [434]. The forkhead box DNA binding domain is comprised of three α-helices flanked by two winged-like loops, thus the DNA recognition sequence is also referred to as the winged-helix domain [437]. All FOXA proteins share up to 95% sequence similarity in the DNA binding domain flanked by the nuclear localization sequence [438]. Outside of the FOX domain, the N and C termini also demonstrate high sequence conservation, and functional analysis of FOXA2 revealed their activity as transcriptional activators [432,438]. Within these domains, regions II, III, and IV contribute to transactivation, where the activity of region IV is dependent on the other two [432]. Analysis of FOXA binding sites in Ttr and Serpina1 regulatory regions indicated the consensus sequence as TATTAGAYTTWG, where Y is C/T and W is A/T [403]. FOXA proteins bind to DNA as monomers [439].

Other hepatocyte-specific genes regulated by FOXA proteins include the Alb enhancer [404] and promoter [440], the Afp distal enhancer [441], the Apob promoter [442], and the Pfkfb1 proximal promoter that controls the expression of 6-phosphofructo-2-kinase/fructose-2,6-bisphosphatase 1 (PFK/FBPase 1) [443]. Furthermore, the structure of the winged-helix domain is similar to that of linker histones H1 and H5, proteins that induce DNA compaction with the nucleosome core to repress gene expression [437]. FOXA proteins are able to bind DNA on the nucleosome core to displace linker histones and increase chromatin accessibility, leading to transcriptional activation [444]; hence, FOXA proteins are also known as ‘pioneer factors’.

FOXA2 is expressed on E6.5 in the node at the anterior primitive streak and is the first member of the FOXA subfamily to be expressed [445]. Its expression persists throughout the development of endoderm-derived tissues such as the liver, pancreas, and the intestine, and continues into adulthood [445]. FOXA1 displays a similar expression pattern as that of FOXA2 but is detected later on E7.0 in the primitive endoderm, whereas FOXA3 is expressed starting on E8.5 during hindgut differentiation [445].

Foxa1−/− embryos develop to term but have severe postnatal growth retardation and die between postnatal day 2 (P2) to P12 [446]. FOXA1-deficient mice experience hypoglycemia and
changes in islet glucagon gene expression, but no liver phenotype is observed prior to death [446], suggesting that FOXA1 is not required for early mouse development but is central to the regulation of glucose homeostasis.

FOXA2 deletion is embryonically lethal by E11 due to the lack of node and notochord formation, causing death prior to the formation of liver bud [447,448]. Conditional FOXA2 ablation with the Foxa2F/F;AlbCre mouse does not induce any significant disruption of the liver phenotype or cause apparent changes in gene expression, indicating that FOXA2 is dispensable in maintaining hepatocytes at a differentiated state [435]. However, Foxa2F/F;AlfpCre livers fail to integrate transcriptional response during prolonged fasting since expression of gluconeogenic enzymes typically activated during fasting is not induced, as seen in the cases of Pck1 that encodes the cytosolic phosphoenolpyruvate carboxykinase (PEPCK-C), Tat that produces tyrosine aminotransferase (TAT), and Igfbp1, encoding insulin-like growth factor-binding protein 1 (IGFBP-1) [449]. Whole-body FOXA1 deletion and endoderm-specific ablation of FOXA2 with the Foxa1-/-;Foxa2F/F;Foxa3Cre mouse showed a lack of liver bud formation and loss of hepatoblast marker Afp, indicating that FOXA1 and 2 are required for hepatogenesis during embryonic development [22]. Furthermore, combinatorial deletion of FOXA1 and 2 in the liver through Foxa1F/F;Foxa2F/F;AlfpCre mice causes bile duct expansion and proliferation, leading to biliary tree hyperplasia while liver differentiation is unaffected [450]. In short, these studies indicate that FOXA2 is required for early embryonic development prior to liver differentiation as well as bile duct maintenance, and FOXA1 and 2 are essential for liver bud specification.

FOXA3 ablation results in a 50-70% decrease in the expression of several hepatocyte-specific genes, including Pck1, Tat, and Tf (Trf) that encodes transferrin along with a compensatory increase of FOXA1 and 2 [451]. Additionally, Foxa3-/- mice exhibit hypoglycemia after prolonged fasting that associates with a decreased liver expression of Slc2a2 (Glut2), encoding the type 2 glucose transporter (GLUT-2) [452]. While FOXA3 deletion is not sufficient to cause severe liver function defects, it is required for mediating fasting glucose homeostasis.
During the acute phase after PHx, Foxa1 expression is dramatically decreased followed by the downregulation of its target gene Ttr, whereas Foxa3 level fluctuates minimally, suggesting that FOXA1, but not 3, is regulated by proliferative signals during liver regeneration [453]. On the other hand, both Foxa2 and Foxa3 expression are significantly reduced in CCl4-induced liver injury [454,455]. Furthermore, injection of AAV8-TBG-Cre into Foxa2<sup>F/F</sup> mice exacerbates CCl4-induced liver fibrosis while FOXA2 overexpression alleviates collagen deposition and reduces ER stress, indicating the hepatoprotective potential of FOXA2 during liver injury [455].

C. HNF4

HNF4 was identified as a nuclear protein with distinct recognition properties from C/EBP, HNF1, and FOXA in its binding to the promoters of TTR and AAT [403]. HNF4 proteins include three isoforms, HNF4α, 4β, and 4γ, but HNF4β is not detected in human or mouse [456]. HNF4α is expressed in the liver, kidney, pancreas, small intestine, colon, and testis [400], while HNF4γ is observed in all tissues mentioned above except for the liver [400]. HNF4α is transcriptionally regulated through two developmentally-controlled promoters P1 and P2 that are separated by more than 45 kilobases (kb) [457]. Differential promoter usage combined with alternative splicing produces six P1 isoforms, HNF4α1 to α6, and three P2 isoforms, HNFα7 to α12 [400,457,458]. However, the impact of HNF4α isoforms on the transcriptional control of downstream targets remains largely unknown.

HNF4 belongs to the nuclear hormone receptor superfamily that includes receptors for steroids, retinoids, thyroid hormones, and vitamin D [459,460]. Originally classified as an orphan member of the superfamily due to the lack of defined ligands, it was later observed that fatty acyl-CoA thioesters modulate HNF4 activity [461] and linoleic acid acts as the endogenous HNF4α ligand [462]. HNF4α displays the conventional modular structure of nuclear receptors that encompasses six functional regions A-F [460]. The N terminal contains the less conserved A/B region with the activation function 1 (AF-1) domain that acts as a constitutive autonomous transactivator [463]. Region C encodes the highly-conserved DNA-binding domain with two zinc
fingers [459] and is responsible for dimerization on the DNA [464]. Region D refers to a hinge that connects regions C and E [460], in which region E represents a conserved ligand-binding domain that contributes to protein dimerization in the absence of DNA binding [464]. The ligand-binding domain also prevents heterodimerization with RXRα and potentially other nuclear receptors that share similar DNA recognition sequences [465]. Additionally, region E consists of a second activation domain AF-2 with ligand-dependent transcriptional activity, providing an additional layer of HNF4α regulation [463]. Region F is located at the C-terminus and contains a unique repressive feature of the nuclear receptor superfamily to inhibit AF-2 [463]. Together, AF-1 and AF-2 activate transcription in a cell type-independent manner [463]. All HNF4α isoforms share the same DNA-binding domain and 90% of protein structure homology. The main difference between P1 and P2 classes is the lack of AF-1 at the N-terminus in P2 isoforms [466]. Finally, the zinc finger motifs bind to the hormone response elements located in promoters exclusively as homodimers to modulate transcription of HNF4α target genes [464,467].

Initial analysis of the HNF4 binding sites at regulatory regions of Ttr, Serpina1, and Apoc3, which encodes Apoc-III, suggested the consensus sequence as KGKWARGKCYCAY, where K is G/T, W is A/T, R is A/G, and Y is C/T [459]. Later analyses demonstrated that HNF4 recognizes repeats of half-site motifs AGGTCA separated by one or two nucleotides [465], as well as a sequence of NNNNCAAAGTCCA [468]. HNF4α regulates gene expression involved in glucose, cholesterol, and fatty acid metabolism through binding to promoters of apolipoproteins Apoa1 [469], Apob [470], and Apoc3 [470,471], as well as Hnf1a [459], Tf [472], and F7 that produces the human coagulation factor VII [473].

HNF4α mRNA is detected as early as E4.5 in the primitive endoderm of the blastocyte [474]; expression persists throughout liver development until adulthood to maintain hepatocytes at a differentiated state [466]. HNF4α ablation is embryonically lethal due to the lack of extraembryonic tissue development past E5.0 [475]. Tetraploid rescue of Hnf4a−/− embryos displays liver specification without full differentiation that lacks expression of a number of liver marker genes such as Alb, Afp, Tf, apolipoproteins Apoa1, Apoa4, Apob, Apoc2, and Apoc3, Nr1i2 (Pxr), Pah that
produces phenylalanine-4-hydroxylase (PAH), and liver-type fatty acid-binding protein (L-FABP) encoded by \textit{Fabp1} [476]. \textit{Hnf4a}^{F/F};\textit{Alfp}^{Cre} mice fail to undergo fetal liver epithelial transformation due to the lack of expression involved in cell adhesion and cell junction assembly [477]. On E18.5, HNF4α-deficient embryonic hepatocytes also demonstrate decreased gene expression related to glucose homeostasis including \textit{Pck1}, \textit{Gys2} that encodes the liver glycogen synthase, and \textit{G6pc} that produces glucose-6-phosphatase (G6Pase) [478]. Conditional HNF4α deletion with \textit{Hnf4a}^{F/F};\textit{Alb}^{Cre} results in lipid accumulation with reduced serum cholesterol and triglyceride, while serum bile acid level is increased, coinciding with the reduction of \textit{Apob}, \textit{Fapb1}, \textit{Slco1a1}, and \textit{Slc10a1} (\textit{Ntcp}) that encodes the sodium/bile acid cotransporter [479]. Furthermore, \textit{Hnf4a}^{F/F};\textit{Alb}^{Cre\text{ERT2}} mice treated with tamoxifen to remove HNF4α from mature hepatocytes exhibit elevated hepatocyte proliferation with increased expression of cell cycle genes [349]. Together, these studies utilizing transgenic mice suggest the requirement of HNF4α from early liver development for the establishment of epithelial morphology to maintenance of the mature hepatocyte phenotype including lipid, bile acid, cholesterol, and glucose homeostasis through gluconeogenesis and glycogen synthesis. Similar to that observed for HNF1, heterozygous HNF4α mutation affects the pancreas and causes autosomal dominant MODY1 in humans, with no obvious phenotypic deficiencies observed in the liver [480].

HNF4α activity is modulated through post-translational modifications. Phosphorylation of tyrosine residues is necessary for proper nuclear compartment localization, as well as the maintenance of DNA-binding activity and transactivation ability [481]. Additionally, cyclic AMP (cAMP) response element-binding protein (CBP) acetylation of lysine residues within the nuclear localization sequence is required for nuclear retention, DNA sequence binding, and target gene activation [482].

D. HNF6

HNF6 was identified via the protein-DNA complex formed with \textit{Pfkfb1} [483], which encodes a bifunctional enzyme that synthesizes and degrades the key regulator of glycolysis, fructose 2,6-
bisphosphate [484]. Analysis of the Pfkfb1 promoter identified two cis-acting sequences that account for approximately 50% of the transcriptional activity [406]. DNA-affinity labeling from rat liver nuclear proteins with the sequence from site IV of the Pfkfb1 promoter was utilized to extract and purify a liver-specific factor originally identified as LP4 and later renamed HNF6 [483]. The HNF6 subfamily includes two isoforms HNF6α and HNF6β that differ by the linker sequence between the cut domain and the homeodomain [483]. Both isoforms display transactivating abilities, but the DNA-binding affinity depends on the target gene sequence [485].

HNF6 belongs to the ONECUT homeodomain family that includes HNF6 (OC1), OC2, and OC3 [483]. CUT homeodomain proteins were initially described in the Drosophila cut gene [486] and the mammalian homologs mclox gene [487], both consisting of three CUT domains. Interestingly, HNF6 only exhibits a single CUT domain, hence the nomenclature as ‘ONECUT’ homeodomain [485]. The N terminus contains the STP box, a serine/threonine/proline-enriched region, that functions as a transcriptional activator [483]; the C terminus encompasses the bipartite DNA-binding domain formed by the CUT domain (CD) and the homeodomain (HD) [485]. Both HNF6 isoforms bind DNA as monomers and do not form heterodimers [485].

Comparison of HNF6 binding sites of liver-enriched genes including Afp, Hnf3b, Pck1, and Ttr determined the consensus sequence as DWRTCMATND, where D is not C, W is A/T, R is A/G, M is A/C [485]. In addition to controlling Foxa2 [488] and Hnf4a [489] expression, HNF6 activates the promoters of various liver function genes including Ttr [490], Afp [485], and Gck that encodes glucokinase (GCK) [491].

HNF6 is expressed at early developmental stages in the liver, pancreas, and neurons, suggesting its importance in regulating various differentiation programs [489]. HNF6 is detected on E9 prior to liver differentiation and continues to be expressed in the liver and the extrahepatic biliary system throughout development [489]. Hnf6−/− embryos lack the gallbladder primordium, resulting in abnormal morphology of extrahepatic bile ducts and perturbed development of intrahepatic bile ducts [492]. HNF6-deficient mice exhibit abnormal bile duct morphogenesis with increased mortality between P1-10 likely due to increased cholestasis that results in liver necrosis [492]. Hnf6−/−
mice also display reduced HNF1β expression in the biliary epithelial cells during development [492]. Hnf6<sup>F/F</sup>;Alb<sup>Cre</sup> mice demonstrate normal intrahepatic bile duct morphology with no indication of liver injury as measured by serum AST, ALT, and total bilirubin levels [493]. Nonetheless, a later study examining conditional Hnf6 ablation in the adult mouse liver through AAV8-TBG-Cre injection into Hnf6<sup>F/F</sup> mice showed severe hepatosteatosis with the induction of genes involved in oxidation-reduction and lipid metabolism [494]. Furthermore, ectopic HNF6 overexpression via adenovirus prior to PHx leads to an increased number of replicating hepatocytes entering S phase as well as upregulation of the mitogen TGFα, cell cycle regulator cyclin D1, and the transcription factor FOXM1 [495]. These observations implicate a crucial role of HNF6 during liver development and cholangiocyte differentiation, as well as its importance in transcriptional repression of lipid metabolic genes and the stimulation of hepatocyte proliferation during liver regeneration.

HNF6 expression can be elevated by growth hormone [496] through increased STAT5 and HNF4α occupancy [497] in conjunction with the displacement of C/EBPα at the HNF6 promoter [498].

E. C/EBP

C/EBP was discovered as a heat-stable nuclear protein to selectively bind to the CCAAT motif of several viral promoters [499] and viral enhancer core elements [500] in the rat liver nuclei [501]. The C/EBP subfamily consists of several isoforms including C/EBPα, C/EBPβ, C/EBPγ, C/EBPδ, C/EBPε, and C/EBPζ, but only the first three are enriched in the liver [502], with C/EBPα the most predominant isoform expressed in adult hepatocytes [503].

C/EBPα, originally named C/EBP, was identified through analysis of Alb, Ttr, and Serpina1 promoters and the simian virus 40 (SV40) core C enhancer element [297,298]. C/EBPβ was described as a nuclear factor to activate IL6 transcription after IL1 induction, hence its original nomenclature, NF-IL6 [504]. C/EBPβ also binds to regulatory regions of several acute-phase genes including TNF and IL8, indicating its importance to regulate acute inflammatory responses [504]. C/EBPγ was purified as a protein to bind the B cell-specific enhancer and promoter regions of the
immunoglobulin heavy chain (IgH) [505]. Subsequent structural analysis of C/EBPγ demonstrated the lack of transactivating domain observed in C/EBPα and β, suggesting the unlikelihood of C/EBPγ as a direct transcriptional activator or repressor [506]. Rather, C/EBPγ functions as a transdominant negative inhibitor and heterodimerizes with C/EBPα or β to repress their transcriptional activity [506].

C/EBP proteins belong to a larger structural category of the basic leucine zipper (bZIP) family of transcription factors [507], one of the most conserved groups of eukaryotic transcription factors that include JUN, FOS, and cAMP-responsive element-binding (CREB) proteins [502]. The C/EBP subfamily exhibits modular structures that contain an N-terminal transactivating region [508], a basic DNA-binding domain, and a C-terminal leucine zipper [508]; all C/EBP isoforms share over 90% of sequence homology at the bZIP domain [501,507,509]. C/EBP binds to the DNA as homo- or heterodimers and forms intramolecular heterodimers to recognize the same consensus sequence, with the exception of C/EBPζ [505,507,510].

The consensus motif for C/EBP proteins is RTTGCGYAAY, where R is A/G and Y is C/T [511]. Other than Alb and Ttr, C/EBPα also regulates the expression of liver-specific or -enriched genes including Pck1 [512], Tf [472], Slc2a2 [513], Igf1 that produces the insulin-like growth factor I (IGF-1) [514], F9 that encodes the coagulation factor IX [515], and several CYP450 genes [442]. C/EBPβ controls metabolic gene production such as Cyp2d5 [516], Pck1 [517], Aldh1a1 that encodes the cytosolic aldehyde dehydrogenase (RALDH 1) [518], Ca3 (Car3) that produces carbonic anhydrase 3 (CA-III) [518], and several other genes encoding acute-phase proteins during inflammation such as serum amyloid A (SAA) [519] and C-reactive protein (CRP) [520].

C/EBP transcription factors are pivotal for a variety of functions including cell proliferation, differentiation, metabolism, inflammation, tumorigenesis, and apoptosis, particularly in hepatocytes, adipocytes, and hematopoietic cells [521]. C/EBPα expression is detected on E9.5 in the mouse endoderm in the liver primordium, while C/EBPβ expression is detected between E13.5 and E14.5 in the liver [522]. Cebpa−/− mice fail to store hepatic glycogen and die within 8 hours after birth due to hypoglycemia associated with reduced or delayed gene expression of Gys2 and two
gluconeogenic enzymes, PEPCK and G6Pase [523]. Injection of AAV-Cre into $Cebpa^{F/F}$ mice to remove up to 90% of C/EBPα expression in the adult liver demonstrated decreased expression of bilirubin UDP-glucuronosyltransferase (UGT), an enzyme required for bilirubin conjugation and detoxification [524], leading to adult-onset jaundice [525]. Expression of $Pck1$, $Gys2$, and $F9$ was also decreased in the adult mouse liver with conditional C/EBPα ablation [525]. These experiments demonstrate the importance of C/EBPα as a central role for gluconeogenesis, glycogen synthesis, and bilirubin homeostasis in the liver.

The role of C/EBPβ in metabolic regulation is complex. Half of $Cebpb^{-/-}$ mice exhibit steadystate glucose homeostasis but demonstrate fasting hypoglycemia and impaired hepatic glucose production. The other half die at birth due to hypoglycemia attributed to the absence of PEPCK expression followed by the inability to mobilize glycogen stores [517]. In mice injected with concanavalin A to induce immune-mediated liver injury, C/EBPβ nuclear expression is increased as early as 1 h and mRNA levels increased 4 h after liver injury, but returns to normal before entering S phase [526]. C/EBPβ-deficient mice display decreased DNA synthesis and suppression of immediate-early growth response genes, $Mkp1$ and $Egr1$, 1 h post-PHx [239]. Furthermore, $Cebpb^{-/-}$ livers also show sustained hypoglycemia in conjunction with dysregulation of genes important for hepatic gluconeogenesis after PHx [239], suggesting the significance of C/EBPβ for glucose homeostasis after profound metabolic stress such as PHx.

IV. Regulatory circuits of liver-enriched transcription factors

The liver-enriched transcription factors form a cooperative network to establish transcriptional control and to synergistically interact with one another to maintain a hepatocyte-specific gene expression profile [527,528]. Of all transcription factors highly-expressed in hepatocytes, HNF1α and HNF4α deficiency correlate with the lack of liver-specific gene expression in dedifferentiated hepatomas and hepatocyte-fibroblast hybrids [527]. Furthermore, HNF1α and HNF4α reexpression correspond to the transcription of hepatocyte-specific genes in hybrid cells.
These observations led to the hypothesis that HNF1α and HNF4α are the primary transcriptional regulators to maintain the differentiated hepatic phenotype.

In particular, several independent observations suggest that HNF4α could function as the master regulator that sits atop the transcriptional cascade during hepatocyte differentiation [459,474]. (1) HNF4α mRNA is detected as early as E4.5 in the primitive endoderm of the blastocyst [474], preceding the expression of HNF1α on E8.5 [422]. (2) HNF4α is able to overcome the repression in dedifferentiated hepatoma cells to induce expression of epithelial marker genes [530,531]. (3) HNF4α transcriptionally activates HNF1α [459]. (4) Hnfa−/− mice are embryonically lethal due to the lack of extraembryonic tissue development [475] while Hnf1a−/− mice are viable at birth and die around weaning due to hepatic dysfunction [425]. (5) HNF4α occupies around 12% of the hepatocyte genome as determined with human DNA microarray, while HNF1α targets 1.6% and HNF6 1.4%, implying that HNF4α contributes to the regulation of a large portion of liver gene expression [420].

Later studies revealed the complex regulation between liver-enriched transcription factors and proposed that the interplay of hepatocyte nuclear factors presumably resembles a regulatory circuitry, rather than a linear hierarchy [527], in a context-dependent manner [420] summarized in Figure 1.3.
SPECIFIC AIMS

In summary, liver regeneration encompasses crosstalk from different cell types, interactions of various signaling pathways, and modulation of the chromatin architecture to initiate complex networks of transcriptional regulation. While the regenerative response is well described in PHx, it is less evident in injury models. Hence, the goal of this thesis is to utilize unbiased transcriptome- and epigenome-wide techniques to identify regulators of liver regeneration following acute injury. I hypothesize that investigating the modifications of gene expression and chromatin accessibility via cell type-specific analyses of regenerating hepatocytes in the Fah\(^{-}\) model will enable the identification of essential factors of liver repopulation.

In Specific Aim 1, I propose to perform transcriptomic profiling of regenerating hepatocytes to identify drivers of liver proliferation. With the combination of the translating ribosome affinity purification (TRAP) system [532] and the Fah\(^{-}\) model, regenerating hepatocytes will be explicitly isolated followed by high-throughput RNA-sequencing (TRAP-seq) to interrogate gene expression alterations during liver repopulation. Overexpression and inhibition studies will be carried out to investigate the functional significance of genes of interest as promoters of hepatocyte replication following acute liver injury.

In Specific Aim 2, I will assess the association of chromatin accessibility modification and gene expression regulation during the repopulation process. With the implementation of the 'isolation of nuclei tagged in specific cell types' (INTACT) method [533] in the Fah\(^{-}\) mouse, regenerating hepatocyte nuclei will be labeled and sorted followed by the 'assay for transposase accessible chromatin with high-throughput sequencing' (ATAC-seq) [534] to elucidate changes in the chromatin landscape. I propose to integrate multiomic datasets to identify crucial transcription factors and regulatory networks that underlie the regenerative process.

This thesis combines a mouse model reflective of human diseases, systematic in vivo analyses, and functional validation of genes and transcription factors during liver regeneration. By addressing these aims, I expect to identify novel therapeutic targets and critical regulators to enhance liver regeneration following acute injury.
The liver consists of various cell types including hepatocytes, cholangiocytes, stellate cells, Kupffer cells, and sinusoidal endothelial cells. The portal triad is located in zone one and contains the bile duct, hepatic artery, and portal vein, whereas the central vein resides in zone three. Together, the portal vein and the hepatic artery move through the sinusoid toward the central vein to provide blood supply to the liver. On the contrary, bile acids move from zone three to zone one in the bile duct.
The cell cycle is tightly controlled by the rise and fall of cyclin proteins that lead to the activation of CDKs to promote progression through cell cycle checkpoints.
Figure 0.3. Liver-enriched transcription factors form a complex regulatory network.

HNF4α activates HNF1α expression [459,535]. HNF1α negatively autoregulates its own expression [536] and inhibits HNF4α via suppression of the AF2- domain [537]. HNF1α also binds to a 3’ enhancer site to activate FOXA3 transcription [538]. HNF6 activates FOXA2 [488] and HNF4α [489], while FOXA2 is required for FOXA1 expression [539]. Additionally, FOXA1 and 2 compete for FOXA motifs on HNF1α and HNF4α, in which FOXA1 represses while FOXA2 induces HNF1α and HNF4α transcription [539]. Finally, C/EBPα binds to the FOXA2 promoter for transcriptional activation [540].
REFERENCES


42. You M, Crabb DW. Recent advances in alcoholic liver disease II. Mini review: molecular mechanisms of alcoholic fatty liver. Am J Physiol Gastrointest Liver Physiol. 2004;287: G1–6.


188. Kjeldgaard NO, Ploug J. Urokinase an activator of plasminogen from human urine. II. Mechanism of plasminogen activation. Biochim Biophys Acta. 1957;24: 283–289.


439. Marsden I, Jin C, Liao X. Structural changes in the region directly adjacent to the DNA-binding helix highlight a possible mechanism to explain the observed changes in the sequence-specific binding of winged helix proteins. J Mol Biol. 1998;278: 293–299.


CHAPTER 2

TRAP-SEQ IDENTIFIES CYSTINE/GLUTAMATE ANTIPORTER AS A DRIVER OF RECOVERY FROM LIVER INJURY

Parts of this chapter were adapted with permission from TRAP-seq identifies cystine/glutamate antiporter as a driver of recovery from liver injury. Wang AW*, Wangensteen KJ*, Wang YJ, Zahm AM, Moss NG, Erez N, Kaestner KH. The Journal of Clinical Investigation. 2018;128(6):2297-2309. doi:10.1172/JCI95120.
ABSTRACT

Understanding the molecular basis of the regenerative response following hepatic injury holds promise for improved treatment of liver diseases. Here, we report an innovative method to profile gene expression specifically in the hepatocytes that regenerate the liver following toxic injury. We used the Fah<sup>−/−</sup> mouse, a model of hereditary tyrosinemia, which conditionally undergoes severe liver injury unless fumarylacetoacetate hydrolase (FAH) expression is reconstituted ectopically. We used translating ribosome affinity purification followed by high-throughput RNA sequencing (TRAP-seq) to isolate mRNAs specific to repopulating hepatocytes. We uncovered upstream regulators and important signaling pathways that are highly enriched in genes changed in regenerating hepatocytes. Specifically, we found that glutathione metabolism, particularly the gene Slc7a11 encoding the cystine/glutamate antiporter (xCT), is massively upregulated during liver regeneration. Furthermore, we show that Slc7a11 overexpression in hepatocytes enhances, and its suppression inhibits, repopulation following toxic injury. TRAP-seq allows cell type-specific expression profiling in repopulating hepatocytes and identified xCT, a factor that supports antioxidant responses during liver regeneration. xCT has potential as a therapeutic target for enhancing liver regeneration in response to liver injury.
INTRODUCTION

The liver is the main metabolic organ in the body; it is the nexus for homeostasis of carbohydrates, proteins, and lipids, and it eliminates waste products by oxidation and reduction, conjugation, and excretion into the bile. As such, the liver is exposed to environmental toxins that can severely damage hepatocytes and cause acute liver failure [1]. Animals have conserved the ability to regenerate the liver parenchyma upon damage [2] and to restore full mass and function even with the loss of up to 75% of hepatocytes [3].

Liver cells in adult animals are normally quiescent and divide infrequently. With acute tissue damage, however, mature hepatocytes and cholangiocytes enter the cell cycle and divide [4]. In addition, hepatocyte proliferation occurs after partial hepatectomy (PHx), a noninflammatory liver regeneration model in which up to two-thirds of the liver is removed [5]. In rodents, this leads to cell division in most hepatocytes within hours and expansion of the remnant organ over the course of 1 to 2 weeks, until the entire mass of the liver is restored. Because PHx is relatively easily carried out in rodents, it has been used to study liver regeneration in mice for decades [5,6]. In fact, many studies have profiled changes in gene expression during regeneration, and a number of important genes and pathways have been identified [7–9]. The common theme from these studies is that cell-cycle genes are upregulated and metabolic genes are downregulated as hepatocytes divide to recover from PHx.

Other paradigms to study liver regeneration utilize injury models involving treatment of animals with hepatotoxins to examine the expression changes of injured liver tissue taken en bloc [10–12]. However, until now there has been no methodology to distinguish the responses of the healthy, repopulating liver cells from those of damaged hepatocytes and inflammatory cells. In clinically relevant hepatic injury, a minority of cells may be protected from the initial insult and thus poised to drive repopulation [13,14]. It is therefore important to establish which genes in the repopulating hepatocyte drive regeneration in the setting of widespread injury.

The mouse model of hereditary tyrosinemia, an inborn error of tyrosine metabolism caused by a deficiency of fumarylacetoacetate hydrolase (FAH) enzyme [14], is useful for studying the
mechanisms of liver regeneration, since repopulating hepatocytes can be labeled as they divide to
restore liver function after injury. Homozygous null (Fah−/−) mice die at birth with hepatic dysfunction
from toxic metabolites but can be maintained in a healthy state by the drug 2-(2-nitro-4-
trifluoromethylbenzoyl)-1,3-cyclohexanedione (NTBC) [14]. Alternatively, gene therapy that
restores FAH expression can normalize tyrosine catabolism within hepatocytes and allow liver
repopulation by the corrected cells upon NTBC removal [15]. Our previous work also demonstrated
that transgenes can be coexpressed with FAH and can be used to genetically trace repopulating
hepatocytes over time [15,16].

Here, we use translating ribosome affinity purification (TRAP) [17] followed by high-
throughput RNA sequencing (TRAP-seq) to profile the gene expression pattern specific to
repopulating hepatocytes. Slc7a11, encoding the cystine/glutamate antiporter (xCT), was
massively activated in regenerating hepatocytes. xCT imports cystine as a precursor for glutathione
(GSH) synthesis [18,19]. We show that ectopic expression of xCT promotes liver repopulation,
whereas CRISPR/Cas9-mediated mutation of Slc7a11 causes a decrease in replicating
hepatocytes. These findings indicate the functional significance of xCT and suggest that activation
of Slc7a11 could be used clinically to support therapeutic liver regeneration in the setting of acute
liver injury.
RESULTS

TRAP enables lineage-tracing of repopulating hepatocytes

With the goal of specifically isolating repopulating hepatocytes from the injured liver to perform RNA-seq, we initially set out to lineage trace repopulating hepatocytes with GFP and isolate tagged cells by FACS for expression analysis. However, we encountered several problems. First, the fragility of hepatocytes undergoing repopulation led to poor recovery following liver perfusion. Second, the large size of the repopulating hepatocytes hampered the yield and purity of isolated cells by sorting. Finally, the process from organ harvest to cell isolation took more than 2 h, which may have altered the expression profile.

Next, we turned to TRAP-seq (Figure 2.1A), which enables the immunoprecipitation of ribosome-bound, translating mRNA from cells that express a fusion protein of the ribosomal protein L10a and GFP (GFP-L10a) [17]. The fusion protein was subcloned into the coexpression vector pKT2/Fah-mCa//SB [15] to construct pKT2/Fah-Gfp-L10a//SB (TRAP vector), which expresses FAH together with GFP-L10a. The TRAP vector utilizes the Sleeping Beauty transposon system for stable plasmid integration into the hepatocyte genome [15]. The TRAP vector was hydrodynamically injected into Fah−/− mice, and NTBC was withdrawn to induce liver injury and create pressure for the selection of hepatocytes that stably express FAH to repopulate the liver. An estimated 0.1% to 1% of hepatocytes integrated the plasmid stably into their genomes [20]. Tissue was harvested 1 or 4 weeks after injection, and GFP-tagged polysomes were extracted to isolate translating mRNAs specifically from repopulating hepatocytes (Figure 2.1B). No RNA was recovered from mice that were not injected with the TRAP vector, indicating the specificity of TRAP isolation. Three mice in the four-week regeneration group had a greater degree of weight loss (Figure 2.2), which was suggestive of more severe injury. Indeed, livers from these mice had large areas lacking GFP staining, indicating a reduced level of initial plasmid uptake (Figure 2.1C). Hence, we grouped these mice into a separate category termed “4-week regeneration after severe injury.” Immunofluorescence (IF) analysis of liver sections confirmed that the majority of proliferating hepatocytes also expressed GFP (Figure 2.1C). Thus, TRAP allows for mRNA
isolation selectively from hepatocytes repopulating the injured liver, without contamination from
dying hepatocytes or inflammatory cells.

**TRAP allows hepatocyte-specific RNA isolation from the quiescent liver**

To obtain mRNA from quiescent hepatocytes as a reference for TRAP-seq, we used the
*Rosa<sup>LSL</sup>-GFP-L10a* mouse, in which expression of GFP-L10a can be activated following Cre expression
[21]. We injected *Rosa<sup>LSL</sup>-GFP-L10a* mice with hepatocyte-specific AAV8-TBG-Cre [22,23] and
performed TRAP to isolate hepatocyte mRNA 1 week later (Figure 2.1A). IHC of liver tissue from
these mice confirmed that GFP expression was only found in hepatocytes following AAV8-TBG-
Cre injection (Figure 2.1C). GFP and Ki67 colabeling revealed very few actively dividing
hepatocytes (Figure 2.1C), consistent with the quiescent liver state.

High-throughput sequencing of cDNA libraries derived from 16 samples of TRAP-isolated
mRNA obtained, on average, 5.8 million uniquely mapped reads (Supplementary Digital Table 2.1).
As expected in pure hepatic mRNA, the 10 most abundant transcripts in the quiescent animals
were specific to hepatocytes (Table 2.1) [24–26]. Hepatocyte-specific genes such as Alb and Ttr
were highly abundant in hepatocytes from all samples, whereas the biliary epithelium markers
CK19, CK7, CFTR, and PKD2, as well as transcripts from other cell types in the liver, were nearly
undetectable (Table 2.2) [27,28], demonstrating the exquisite specificity of the TRAP method.

**TRAP-seq detects differentially expressed genes in repopulating hepatocytes**

Differential gene expression analysis identified 6,745 genes that change in expression in
repopulating compared with quiescent hepatocytes (Supplementary Digital Table 2.2); 3,418 were
significantly upregulated and 3,380 downregulated (FDR ≤ 5%) (Figure 2.3A). Hierarchical
clustering of the differentially expressed genes showed a distinct separation between quiescent
and repopulating hepatocytes (Figure 2.3B). Notably, the 4-week regeneration group clustered
closer to the 1-week regeneration group, demonstrating that TRAP-seq allows identification of
different levels of liver regeneration. To establish whether the differentially expressed genes fall
into defined regulatory networks, we used pathway analysis and focused on the highly validated
Kyoto Encyclopedia of Genes and Genomes (KEGG) network collection [29,30] (Figure 2.3C). Pathways controlling replication and growth were overrepresented among the upregulated genes, including those regulating the cell cycle and DNA replication, indicating that genes involved in cell replication were activated during liver repopulation, as expected. Strikingly, the GSH metabolic pathway was strongly activated in regenerating hepatocytes, aligning with previous studies showing that control of oxidative stress plays a crucial role in the regenerative response following toxic liver injury [31]. Interestingly, metabolic pathways were enriched in both activated and inhibited genes, reflecting the important metabolic regulation of hepatocytes, although the genes at play were different in the 2 groups (Supplementary Digital Table 2.3). Upregulated metabolic genes included redox processes, whereas repressed genes regulate lipid biosynthesis, corroborating previous findings that hepatocytes limit the activity of metabolic networks to conserve energy for rapid cell replication and DNA synthesis during regeneration [6].

The key regulatory nodes enriched in differentially expressed genes were analyzed with Ingenuity Pathway Analysis, which takes into account the degree of change of each gene to generate putative regulatory networks and predict activation or inhibition of the pathways. We identified 227 upstream regulators, of which 24 met the following additional filters: (a) significant Z-scores (≥2 for predicted activation and ≤2 for predicted inhibition); (b) at least a 2-fold change in expression; and (c) congruence between the observed fold change and predicted state categories (Table 2.3). MYC, the most enriched regulator, is a proto-oncogene activated as early as 1 h after PHx [32] and is also upregulated in liver regeneration induced by carbon tetrachloride and galactosamine [33]. Previous work had identified MYC as the strongest driver of liver repopulation in Fah<sup>−/−</sup> mice in a cDNA overexpression screen of more than 40 genes [16], and its overexpression also induces spontaneous hepatocellular carcinoma (HCC) development in the Fah<sup>−/−</sup> mouse model within 8 weeks [34]. A second upstream regulator of the proliferative response is the transcription factor FOXM1, which was previously shown to enhance liver repopulation [35]. These results indicate that we were indeed able to profile the translating mRNA signature specifically in
repopulating hepatocytes and demonstrate that TRAP-seq is a robust methodology for identifying enriched pathways and upstream regulators.

**Fah**<sup>−/−</sup> and PHx regeneration models share common genetic pathways

Next, we set out to compare the transcriptional changes of regenerating hepatocytes in *Fah*<sup>−/−</sup> mice recovering from toxic injury with those occurring following PHx, a paradigm of noninjury regeneration. First, we reanalyzed previous RNA-seq data from whole-liver homogenates after PHx [36] and identified 2,321 differentially expressed genes, 1,449 of which were activated and 872 inhibited (Figure 2.3D). Hierarchical clustering showed a distinct separation of gene regulation at various time points after PHx (Figure 2.3E). Interestingly, gene expression at 1 h clustered closer with quiescent hepatocytes, indicating that at this very early time point only a few early-response genes were transcriptionally regulated. Pathway analysis [29,30] showed enrichment of genes regulating cell-cycle and DNA synthesis pathways among the upregulated genes and those regulating immune and metabolic pathways among the downregulated genes (Figure 2.3F).

We compared the gene expression changes between the *Fah*<sup>−/−</sup> and PHx models, defining congruent genes as those regulated in the same direction in both models for at least 1 time point. We identified a total of 1,236 congruent genes, 790 of which were activated and 446 repressed (Figure 2.4A). Gene expression changes that occurred at all time points in the *Fah*<sup>−/−</sup> repopulation mice were most similar to the changes observed in the PHx model at later time points (36 or 48 h after PHx), as shown by the high percentage of congruence. Additionally, we found that the percentage of congruence was higher among the upregulated genes, indicating a more similar gene activation pattern in the 2 regeneration models. We discovered that the top upregulated congruent genes — ranked by mean fold change in *Fah*<sup>−/−</sup> mice and subsequently retrieved from the PHx data set — were associated with GSH metabolism, including the genes *Slc7a11* and *Gsta1* (Supplementary Digital Table 2.4) [18]. This was confirmed by pathway analysis, in which GSH metabolism was highly enriched in the congruently upregulated genes, along with cell-cycle, DNA replication, and DNA repair pathways (Figure 2.4C). Immune response and metabolic pathways...
were enriched among the congruently downregulated genes (Figure 2.4D). Interestingly, the majority of the congruent genes did not show a significant change 1 h after PHx (Supplementary Digital Table 2.4), as at this stage, hepatocytes still resembled quiescent hepatocytes, with activation of only a few immediate early genes [2].

Of note, 2 of the top congruently upregulated genes, *Ly6d* and *Pbk*, are not typically expressed in hepatocytes. The average fragments per kilobase of transcript per million mapped (FPKM) reads for these genes in the quiescent hepatocytes were 2.7 and 0.01, but increased to 504.4 and 6.6 in regenerating hepatocytes, respectively (Supplementary Digital Table 2.2). *Ly6d* expression has been shown to be associated with HCC and liver regeneration after injury [37,38], while *Pbk* has been detected in HCC and cholangiocarcinoma [39,40]. This further demonstrates the sensitivity and specificity of TRAP-seq in detecting expression changes in a unique subpopulation of the liver — that of the regenerating hepatocytes.

Additionally, we identified genes that were only changed in 1 model but not the other (unique genes), of which 5,510 were unique to *Fah*−/− mice, and 1,033 were unique to the PHx model (Figure 2.4B). Of note, in both models, the percentage of unique genes compared with the total number of differentially expressed genes was approximately 81%. However, in the *Fah*−/− mice, up-and downregulated genes each constituted 50% of the unique genes, whereas in the PHx model, the upregulated and downregulated genes made up 64% and 36% of the unique genes, respectively. To further identify the biological pathways specific to each model, pathway enrichment analysis was performed on the unique genes [29,30], and overrepresented networks were identified (Figure 2.4, C and D). In *Fah*−/− mice, liver injury response categories such as alcoholism and viral carcinogenesis were uniquely activated, while immune response and metabolic pathways were uniquely inhibited. On the other hand, no significant pathway activation was unique to the PHx model, whereas the pancreatic secretion and protein and fat digestion/absorption pathways were uniquely inhibited. The striking difference in enriched pathways demonstrates the gene expression signatures that differentiate the 2 regeneration paradigms, in which injury response and immune modulation are unique to *Fah*−/− mice and nutrient redistribution is integral to the PHx model.
Recently, single-molecule RNA-FISH combined with single-cell RNA-seq (scRNA-seq) has been applied to reconstruct the spatial heterogeneity and identify novel zonal signature genes within the quiescent liver [41]. While TRAP-seq utilizes bulk RNA-seq and therefore cannot inform a spatial resolution of transcriptional changes during regeneration, we compared the expression profiles of quiescent hepatocytes from TRAP-seq with the scRNA-seq data. We reasoned that since all hepatocytes express GFP-L10a in the quiescent liver (Figure 2.1C), the isolated transcripts from TRAPseq should have an equal representation of the genes identified from the 9 different subpopulations by scRNA-seq. As expected, we found significant overlap between TRAP-seq and all 9 layers of scRNA-seq, with an average of 10,405 common genes, constituting 90.7% of the genes detected by TRAP-seq (Figure 2.5). Thus, TRAP-seq enables unbiased RNA isolation from all layers of hepatocytes.

**Slc7a11 is massively upregulated in regenerating hepatocytes**

The comparison of the Fah<sup>+</sup> and PHx models revealed Slc7a11 as the most significantly activated gene in both paradigms, with a remarkable increase of 900-fold in the former and 200-fold in the latter (Supplementary Digital Table 2.4). Slc7a11 encodes xCT, a sodium-independent transporter for cystine import and glutamate export [19]. After entering the cell, cystine is rapidly reduced to cysteine, a precursor for GSH synthesis necessary for cellular defense against oxidative stress [18]. Previous studies indicated that deficiency of glutamate-cysteine ligase, the rate-limiting enzyme in GSH synthesis, leads to decreased hepatocyte proliferation in vitro and delayed regeneration after PHx [42,43]. However, the role of xCT in liver regeneration has not been studied. We hypothesized that xCT upregulation supports actively repopulating hepatocytes to defend against increased oxidative stress during injury and regeneration (Figure 2.6A).

To evaluate the role of xCT in liver regeneration, we first validated our observations from RNA-seq with quantitative real-time reverse transcription PCR (qRT-PCR) on TRAP-purified mRNA and confirmed a significant upregulation of Slc7a11 and Gsta1 transcripts in repopulating hepatocytes (Figure 2.7A). Western blot analysis showed an increase in xCT protein in repopulating
livers (Figure 2.7B). Of note, there was low xCT expression in the quiescent liver, albeit no mRNA transcripts were present in hepatocytes. One possibility is that whole-liver homogenate was used for the protein analysis, and thus xCT protein from other cell types such as macrophages was detected [19]. Alternatively, the protein stability of Slc7a11 could exceed its RNA turnover rate. Regardless, the expression of Slc7a11 was significantly activated in the regenerating liver.

We next sought to investigate whether oxidative stress is increased in Fah−/− livers during regeneration. We used immunohistochemical methods to detect markers of lipid peroxidation (malondialdehyde and 4-hydroxynonenal) and protein nitration (nitrotyrosine). We observed an accumulation of redox metabolites in the injured livers compared with healthy, quiescent livers (Figure 2.7C). These results indicate that Slc7a11 mRNA expression and xCT protein levels are highly enriched in repopulating hepatocytes in the presence of increased reactive oxygen and nitrogen species, suggesting a functional role of Slc7a11 in the regulation of liver regeneration.

**Ectopic Slc7a11 expression promotes liver regeneration**

To examine the functional importance of xCT activation in regenerating hepatocytes, we constructed plasmids coexpressing Fah and overexpressing Slc7a11 (Fah-Slc7a11) or Gfp (Fah-Gfp). We performed a competition assay, in which equimolar amounts of Fah-Gfp and Fah-Slc7a11 were injected into Fah−/− mice, followed by NTBC withdrawal (Figure 2.6B). After 4 weeks of repopulation, we observed a 2.5-fold enrichment of Fah-Slc7a11 plasmid relative to the Fah-Gfp control plasmid by qPCR of extracted liver genomic DNA (Figure 2.6C) as well as overrepresentation of HA-tagged, xCT-expressing hepatocytes compared with GFP-expressing cells (Figure 2.6D). These results demonstrate a positive selection for hepatocytes overexpressing xCT, even above the already striking activation of endogenous Slc7a11.

To test whether Slc7a11 is required for liver regeneration, we used CRISPR/Cas9 to inactivate Slc7a11 specifically in the repopulating hepatocytes. We coexpressed FAH with either 10 single-guide RNAs (sgRNAs) targeting Slc7a11 exons (Fah-sgSlc7a11) or 10 control sgRNAs targeting luciferase (Fah-gCtrl) and performed hydrodynamic tail-vein injection of these sgRNAs,
together with adeno-associated virus 8 (AAV8) expressing Staphylococcus aureus Cas9 (SaCas9) to allow for hepatocyte-specific expression of the SaCas9 nuclease [44], which efficiently introduces indels comparable to those of Cas9 from S. pyogenes [45] (Figure 2.6E). Liver repopulation was then carried out for 4 weeks. To quantify and characterize the mutations induced by CRISPR/Cas9, we extracted genomic DNA from the repopulating livers, PCR amplified exon 1 for Sanger sequencing, and performed tracking of indels by decomposition (TIDE) analysis [46]. We found that the 2 sgRNAs targeting the first exon of Slc7a11 exhibited different mutation efficiency: 29.5% and 51.6%, respectively (Figure 2.8, A and B). Furthermore, the main mutation introduced by SaCas9 in either sgRNA was a 5-nucleotide deletion, with an efficacy of 27.9% and 51.6%, respectively. The difference in mutation rate could be due to the slight difference in the protospacer-associated motif (PAM) sequence (NNGRRT) of the 2 sgRNAs, CTGAGT and AAGGGT [44]. Nonetheless, TIDE analysis demonstrated that Slc7a11 was mutated through the expression of SaCas9 in the hepatocytes.

We measured weight changes over the 4-week period of liver repopulation and found no significant weight differences in mice treated with Slc7a11 sgRNAs compared with those treated with control sgRNAs (Figure 2.8C). Likewise, we detected no significant difference in the liver weight to body weight ratio by the end of the 4-week period (Figure 2.8D). However, sgSlc7a11-treated mice had smaller FAH repopulation nodules and fewer Ki67/FAH double-positive hepatocytes compared with sgCtl-treated mice (Figure 2.6F), indicating that Slc7a11 mutation inhibits replication of FAH-expressing cells during liver injury. It should be noted that these results are probably an underrepresentation of the true effect of Slc7a11 mutation, as only hepatocytes homozygous, not those that are heterozygous, for inactivation of Slc7a11 are expected to be at a growth disadvantage. Furthermore, redundant pathways could compensate for the loss of Slc7a11 [47]. Together, these studies demonstrate the functional importance of xCT during liver repopulation and show that Slc7a11 overexpression is sufficient to accelerate repopulation, whereas Slc7a11 inactivation, while not completely abrogating regeneration, hinders hepatocyte replication.
**Slc7a11 is transcriptionally activated by ATF4**

Finally, we investigated the mechanism of xCT activation during liver repopulation. Several transcription factors have been shown to regulate Slc7a11 expression in different contexts: nuclear factor E2-related factor 2 (NRF2) activates xCT during redox stress [48], activating transcription factor 4 (ATF4) upregulates xCT under ER stress, octamer-binding transcription factor (OCT1) disinhibits Slc7a11 following ethanol exposure [49], and p53 inhibits xCT under normal tumor suppression conditions [50]. Additionally, ATF4 is suggested to regulate the basal levels of Slc7a11 expression [51].

We first performed unbiased chromatin accessibility profiling to identify regulatory elements at the Slc7a11 locus in hepatocytes in the basal and repopulating state. We used the isolation of nuclei-tagged in specific cell types (INTACT) system to label the nuclei of regenerating hepatocytes [52]. Specifically, the nuclear envelope protein SUN1 was tagged with GFP [53], and the resulting fragment was subcloned into the FAH coexpression construct (Fah-Sun1-Gfp). One week after the Fah−/− mice were repopulated with Fah-Sun1-Gfp, livers were harvested and sorted for GFP-positive nuclei (Figure 2.9A). As a quiescent control, we injected RosaLSL-Sun1-GFP mice with AAV8-TBG-Cre and sorted hepatocytes after 1 week (Figure 2.9A). We used the assay for transposase accessible chromatin using sequencing (ATAC-seq) [54,55] to profile the chromatin landscape changes after 1 week of regeneration. Remarkably, the Slc7a11 promoter is highly accessible in the regenerating hepatocytes, as indicated by the strong peak present 1 week after regeneration (Figure 2.9B). In comparison, we observed no peak at the promoter in the quiescent liver, demonstrating a heterochromatic state in healthy liver cells. This observation coincides with our TRAP-seq analysis, in which no Slc7a11 transcripts were detected in quiescent hepatocytes, but became highly abundant in regenerating hepatocytes (Figure 2.7A).

Next, to determine how Slc7a11 is activated, we performed a motif search at the open chromatin region of the activated promoter and identified a potential NRF2-binding site 39 bases and 2 potential ATF4-binding sites 39 and 66 bases upstream of the transcriptional start site (Figure 2.9C). To assess whether ATF4 or NRF2 binds to the Slc7a11 promoter during liver repopulation,
we carried out ChIP-qPCR in quiescent and 4-week regenerating livers. We detected a significant 4-fold enrichment of bound ATF4 at the *Slc7a11* promoter in regenerating hepatocytes relative to that seen in quiescent controls. In contrast, NRF2 binding was undetected in either condition (Figure 2.9D), suggesting that ATF4, but not NRF2, activates *Slc7a11* transcription during liver repopulation.
### TABLES

**Table 2.1.** Top ten abundant transcripts identified in quiescent livers.

<table>
<thead>
<tr>
<th>Gene</th>
<th>Quiescent</th>
<th>1-week regeneration</th>
<th>4-week regeneration</th>
<th>4-week regeneration after severe injury</th>
</tr>
</thead>
<tbody>
<tr>
<td>Apoc3</td>
<td>22258.99</td>
<td>8284.08</td>
<td>16707.52</td>
<td>10378.8</td>
</tr>
<tr>
<td>Apoa2</td>
<td>18465.14</td>
<td>8906.01</td>
<td>8600.62</td>
<td>10388.86</td>
</tr>
<tr>
<td>Fabbp1</td>
<td>14824</td>
<td>15222.48</td>
<td>10529.91</td>
<td>1586.67</td>
</tr>
<tr>
<td>Apoc1</td>
<td>12632.39</td>
<td>14177.49</td>
<td>6945.49</td>
<td>15360.81</td>
</tr>
<tr>
<td>Apoe</td>
<td>12418.85</td>
<td>6753.43</td>
<td>5936.13</td>
<td>10505.54</td>
</tr>
<tr>
<td>Apoc1</td>
<td>11083.99</td>
<td>12446.11</td>
<td>6094.25</td>
<td>13496.84</td>
</tr>
<tr>
<td>Alb</td>
<td>10481.08</td>
<td>5141.19</td>
<td>2962.66</td>
<td>9602.19</td>
</tr>
<tr>
<td>Trf</td>
<td>6906.36</td>
<td>1267.84</td>
<td>1930.87</td>
<td>3292.12</td>
</tr>
<tr>
<td>Gpx1</td>
<td>6150.21</td>
<td>4540.07</td>
<td>5258.64</td>
<td>5562.72</td>
</tr>
<tr>
<td>Apoc4</td>
<td>5028.13</td>
<td>4193.01</td>
<td>4623.31</td>
<td>2596.7</td>
</tr>
</tbody>
</table>

Numbers represent the average fragments per kilobase of transcript per million (FPKM) reads in each regeneration group.
Table 2.2. FPKM of cell type-specific transcripts detected by TRAP-seq.

<table>
<thead>
<tr>
<th>Gene</th>
<th>Quiescent</th>
<th>1-week regeneration</th>
<th>4-week regeneration</th>
<th>4-week regeneration after severe injury</th>
<th>Cell type</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alb</td>
<td>10481.08</td>
<td>5141.19</td>
<td>2962.66</td>
<td>9602.19</td>
<td>Hepatocyte</td>
</tr>
<tr>
<td>Ttr</td>
<td>2639.6</td>
<td>3072.59</td>
<td>1212.62</td>
<td>5454.57</td>
<td>Hepatocyte</td>
</tr>
<tr>
<td>Cyp2e1</td>
<td>2424.01</td>
<td>588.72</td>
<td>1418.17</td>
<td>380.48</td>
<td>Hepatocyte</td>
</tr>
<tr>
<td>Asgr1</td>
<td>1190.04</td>
<td>865.69</td>
<td>1375.35</td>
<td>654.07</td>
<td>Hepatocyte</td>
</tr>
<tr>
<td>Krt19</td>
<td>0.18</td>
<td>0.16</td>
<td>0.15</td>
<td>1.97</td>
<td>Biliary epithelium</td>
</tr>
<tr>
<td>Pkd2</td>
<td>1.28</td>
<td>0.71</td>
<td>0.79</td>
<td>1.56</td>
<td>Biliary epithelium</td>
</tr>
<tr>
<td>Krt7</td>
<td>0.22</td>
<td>0</td>
<td>0.22</td>
<td>0.89</td>
<td>Biliary epithelium</td>
</tr>
<tr>
<td>Cfrt</td>
<td>0</td>
<td>0</td>
<td>0.02</td>
<td>0</td>
<td>Biliary epithelium</td>
</tr>
<tr>
<td>Des</td>
<td>1.09</td>
<td>0.33</td>
<td>0.43</td>
<td>0.86</td>
<td>Stellate cell</td>
</tr>
<tr>
<td>Acta2</td>
<td>0.41</td>
<td>0.35</td>
<td>0.27</td>
<td>0.34</td>
<td>Stellate cell</td>
</tr>
<tr>
<td>Col1a1</td>
<td>0.04</td>
<td>0.42</td>
<td>0.26</td>
<td>1.22</td>
<td>Stellate cell</td>
</tr>
<tr>
<td>Cd68</td>
<td>0.92</td>
<td>0.71</td>
<td>0.36</td>
<td>5.9</td>
<td>Kupffer cell</td>
</tr>
<tr>
<td>Emr1</td>
<td>0.67</td>
<td>0.28</td>
<td>0.27</td>
<td>0.84</td>
<td>Kupffer cell</td>
</tr>
<tr>
<td>Cd163l1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>Kupffer cell</td>
</tr>
<tr>
<td>Clec5a</td>
<td>0</td>
<td>0</td>
<td>0.02</td>
<td>0.05</td>
<td>Kupffer cell</td>
</tr>
</tbody>
</table>
Table 2.3. Upstream regulators predicted by Ingenuity Pathway Analysis.

<table>
<thead>
<tr>
<th>Upstream regulator</th>
<th>Fold change</th>
<th>Molecule type</th>
<th>Predicted state</th>
<th>Z-score</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>MYC</td>
<td>3.08</td>
<td>Transcription regulator</td>
<td>Activated</td>
<td>4.73</td>
<td>1.15E-24</td>
</tr>
<tr>
<td>SREBF1</td>
<td>0.26</td>
<td>Transcription regulator</td>
<td>Inhibited</td>
<td>-3.97</td>
<td>3.32E-16</td>
</tr>
<tr>
<td>THRB</td>
<td>0.50</td>
<td>Nuclear receptor</td>
<td>Inhibited</td>
<td>-2.28</td>
<td>4.26E-10</td>
</tr>
<tr>
<td>E2F1</td>
<td>2.53</td>
<td>Transcription regulator</td>
<td>Activated</td>
<td>2.59</td>
<td>2.06E-09</td>
</tr>
<tr>
<td>FOXM1</td>
<td>15.67</td>
<td>Transcription regulator</td>
<td>Activated</td>
<td>3.23</td>
<td>4.79E-09</td>
</tr>
<tr>
<td>EGR1</td>
<td>0.33</td>
<td>Transcription regulator</td>
<td>Inhibited</td>
<td>-2.09</td>
<td>4.33E-08</td>
</tr>
<tr>
<td>HBB-B1</td>
<td>0.11</td>
<td>Transporter</td>
<td>Inhibited</td>
<td>-2.07</td>
<td>7.41E-08</td>
</tr>
<tr>
<td>SPARC</td>
<td>0.30</td>
<td>Other</td>
<td>Inhibited</td>
<td>-3.62</td>
<td>1.43E-07</td>
</tr>
<tr>
<td>CSF1</td>
<td>0.48</td>
<td>Cytokine</td>
<td>Inhibited</td>
<td>-2.95</td>
<td>2.63E-07</td>
</tr>
<tr>
<td>HBB-B2</td>
<td>0.11</td>
<td>Other</td>
<td>Inhibited</td>
<td>-2.68</td>
<td>5.31E-07</td>
</tr>
<tr>
<td>ERF2</td>
<td>4.85</td>
<td>Transcription regulator</td>
<td>Activated</td>
<td>2.75</td>
<td>9.10E-07</td>
</tr>
<tr>
<td>USF2</td>
<td>0.35</td>
<td>Transcription regulator</td>
<td>Inhibited</td>
<td>-2.49</td>
<td>2.40E-06</td>
</tr>
<tr>
<td>AGTR1</td>
<td>0.40</td>
<td>GPCR</td>
<td>Inhibited</td>
<td>-2.70</td>
<td>6.04E-06</td>
</tr>
<tr>
<td>LMNB1</td>
<td>6.69</td>
<td>Other</td>
<td>Activated</td>
<td>2.56</td>
<td>6.26E-06</td>
</tr>
<tr>
<td>CCNE1</td>
<td>3.7</td>
<td>Transcription regulator</td>
<td>Activated</td>
<td>2.07</td>
<td>2.14E-05</td>
</tr>
<tr>
<td>MLXIPL</td>
<td>0.40</td>
<td>Transcription regulator</td>
<td>Inhibited</td>
<td>-3.70</td>
<td>3.19E-05</td>
</tr>
<tr>
<td>TFEB</td>
<td>0.35</td>
<td>Transcription regulator</td>
<td>Inhibited</td>
<td>-2.05</td>
<td>3.33E-04</td>
</tr>
<tr>
<td>S100A6</td>
<td>2.92</td>
<td>Transporter</td>
<td>Activated</td>
<td>2.85</td>
<td>1.20E-03</td>
</tr>
<tr>
<td>CTGF</td>
<td>0.44</td>
<td>Growth factor</td>
<td>Inhibited</td>
<td>-2.04</td>
<td>1.26E-03</td>
</tr>
<tr>
<td>TAS1R3</td>
<td>0.42</td>
<td>GPCR</td>
<td>Inhibited</td>
<td>-2.14</td>
<td>2.02E-03</td>
</tr>
<tr>
<td>IL15</td>
<td>0.37</td>
<td>Cytokine</td>
<td>Inhibited</td>
<td>-2.70</td>
<td>3.28E-03</td>
</tr>
<tr>
<td>FASN</td>
<td>0.14</td>
<td>Enzyme</td>
<td>Inhibited</td>
<td>-2.11</td>
<td>1.77E-02</td>
</tr>
<tr>
<td>TNK1</td>
<td>0.46</td>
<td>Kinase</td>
<td>Inhibited</td>
<td>-2.83</td>
<td>2.03E-02</td>
</tr>
<tr>
<td>MLYCD</td>
<td>0.50</td>
<td>Enzyme</td>
<td>Inhibited</td>
<td>-2.00</td>
<td>2.06E-02</td>
</tr>
</tbody>
</table>

Filter criteria: (a) significant Z-scores (≥2 for predicted activation and ≤2 for predicted inhibition); (b) at least 2-fold change in expression; and (c) congruence between the observed fold change and predicted state categories.
Figure 2.1. Translating ribosome affinity purification (TRAP) enables cell type-specific isolation of RNA from quiescent and repopulating hepatocytes.

(A) The approach for isolating repopulating hepatocyte RNA with the Fah−/− model involves use of the FAH expression construct to mediate liver repopulation and the GFP-tagged ribosomal protein L10a (GFP-L10a) to specifically isolate translating mRNAs with TRAP. Injection of the RosaLSL-GFP-L10a mouse with the AAV8-TBG-Cre virus, which has a tropism for hepatocytes and has a hepatocyte-specific promoter driving Cre expression in nearly all hepatocytes, allows for immunoprecipitation of translating mRNA from quiescent hepatocytes. (B) Bioanalyzer tracings of affinity-purified RNA from mice treated with or without the TRAP vector. FU, fluorescence units. (C) Representative (n = 3) IHC images of GFP show progressive repopulation over time in Fah−/− mice as well as complete labeling of quiescent hepatocytes in RosaLSL-GFP-L10a mice 1 week after injection.
of AAV8-TBG-Cre. No GFP expression was observed in livers from the uninjected mice. IF of Ki67 and GFP confirmed successful liver repopulation in Fah<sup>−/−</sup> mice injected with the TRAP vector, as all Ki67-positive hepatocytes express GFP. IF costaining also showed global GFP-expressing and rare Ki67-positive hepatocytes, indicating that the control tissue was truly quiescent. Note that a subset of mice showed only partial repopulation at 4 weeks (4-week regeneration after severe injury). Scale bars: 1 mm (top) and 100 μm (bottom).
Figure 2.2. Mice in the 4-week regeneration after severe injury group exhibit significant weight loss.

The proportion of weight loss was normalized to the initial weight prior to plasmid injection and NTBC removal. Bodyweight was monitored three times per week after induction of liver injury and regeneration. After four weeks of injury and regeneration, three mice lost ~30% of the starting weight (blue), significantly different from mice in the 4-week regeneration group that underwent initial weight loss but restored body weight after four weeks (red). A two-sided, two-tailed Student’s t-test was used to compare the proportion of body weight in the 4-week regeneration (n=6) and 4-week regeneration after severe injury (n=3) groups.
Figure 2.3. TRAP-seq identifies differentially expressed genes specific to repopulating hepatocytes in the \( \text{Fah}^{-/} \) model.

(A and D) Differential expression analysis identified 6,745 (3,418 upregulated and 3,380 downregulated) and 2,321 (1,449 upregulated and 872 downregulated) genes as being significantly altered in repopulating hepatocytes in the \( \text{Fah}^{-/} \) (A) and PHx (D) models (36), respectively, compared with quiescent controls. Red, 1-week \( \text{Fah}^{-/} \) regeneration and 1 h after PHx; blue, 4-week \( \text{Fah}^{-/} \) regeneration and 36 h after PHx; green, 4-week \( \text{Fah}^{-/} \) regeneration after severe injury and 48 h after PHx. (B and E) Hierarchical clustering of differentially expressed genes of quiescent and repopulating hepatocytes at different time points. (C and F) KEGG pathways significantly enriched for the sets of activated and repressed genes, respectively, in the \( \text{Fah}^{-/} \) (C) and PHx (F) data sets.
**Figure 2.4.** Comparison of the Fahn\textsuperscript{+-} TRAP-seq data with RNA-seq data from the PHx model identifies common and unique characteristics of liver repopulation paradigms.

(A) A total of 1,236 genes were significantly altered in the same direction in both models [36] for at least 1 time point (congruent genes). Of these genes, 790 were activated and 446 inhibited. Labels indicate the number of congruent genes at each time point. (B) A total of 5,510 and 1,033 genes were uniquely changed in the Fahn\textsuperscript{+-} and PHx models, respectively. (C and D) Comparison of the KEGG pathways enriched for genes upregulated (C) and downregulated (D) in the congruent (Cong) and unique gene sets.
Figure 2.5. Comparison of identified transcripts from single-cell RNA-seq (scRNA-seq) [41] shows significant overlap between TRAP-seq and all nine layers of scRNA-seq.

Genes identified in the quiescent samples from TRAP-seq was compared to that from scRNA-seq. Bar height indicates the number of overlapping genes identified in two techniques. Line and data points indicate the percentage overlap from each scRNA-seq layer compared to TRAP-seq. A hypergeometric test was used to calculate the significance of overlapping genes from the two sequencing methods.
Figure 2.6. *Slc7a11* enhances hepatocyte repopulation.

(A) The *Slc7a11* gene product (xCT) imports cystine, which is used for GSH synthesis to alleviate oxidative stress. Several GSH metabolic enzymes were significantly (FDR ≤ 5%) upregulated (red) in repopulating hepatocytes from *Fah*<sup>-/-</sup> mice. GSSG, glutathione disulfide; GCL, glutamate-cysteine ligase; GSS, glutathione synthetase; GST, glutathione S-transferase; GSR, glutathione reductase; GPX, glutathione peroxidase. (B) Schematic of the competition assay to determine the effects of *Slc7a11* overexpression on repopulation. (C) The Fah-Slc7a11 plasmid was significantly enriched after 4 weeks of repopulation. A 1-sample, 2-tailed Student’s t-test was used to compare the ratio of 2 plasmids before and after repopulation (n = 8). (D) Representative IF staining and quantification showing a significant increase in xCT-positive hepatocytes. A paired, 2-tailed Student’s t-test was used to compare HA- and GFP-expressing hepatocytes (n = 5). Scale bar: 100 μm. (E) Schematic of the CRISPR/Cas9 system used to inactivate *Slc7a11* in *Fah*<sup>-/-</sup> mice. sgCtl, sgRNAs targeting firefly luciferase. (F) Representative IHC and IF images and quantification showing a significant reduction in repopulation nodules and replicating hepatocytes in mice treated
with sgRNAs targeting *Slc7a11* (sgSlc7a11) compared with control mice treated with sgCtl. A 2-sample, 2-tailed Student’s t-test was used to compare groups (n = 4 each). Scale bars: 300 μm (top) and 100 μm (bottom).
**Figure 2.7.** *Slc7a11* is activated at the transcript and protein levels under increased oxidative stress during liver regeneration.

(A) Real-time reverse transcription PCR (qRT-PCR) analysis showed continuous upregulation of *Slc7a11* and *Gsta1*, both involved in GSH metabolism, in repopulating hepatocytes. A two-sample, two-tailed Student’s t-test was used to compare repopulating and quiescent hepatocytes. * p<0.05, *** p<0.001 (n=4, quiescent and 1-week regeneration; n=3, 4-week regeneration and 4-week regeneration after severe injury). (B) Western blot analysis confirmed the activation of xCT in the regenerating liver. (C) IHC staining of lipid peroxidation markers (malondialdehyde and 4-hydroxynonenal) and protein nitration (nitrotyrosine) showed accumulation of redox metabolites in the injured, repopulating liver compared to healthy, quiescent livers. Scale bar: 100µm.
**Figure 2.8.** No significant differences in the weight of mice with *Slc7a11* inhibition compared to control after 4 weeks of repopulation.

(A and B) Mutation analysis of *Slc7a11* exon one identified differential indel rates introduced by two single guide RNAs (sgRNA), sgSlc7a11-1 (A) and sgSlc7a11-2 (B). The x-axis indicates the number of nucleotides that were inserted or deleted and the y-axis indicates the percentage of mutation. (C) No weight differences during and after 4 weeks of repopulation and no changes in liver weight to body weight ratio (D) in mice treated with sgRNA against *Slc7a11* (n=4) compared to control mice (n=4). A two-sample, two-tailed Student’s t-test was used to compare mice treated with *Slc7a11* and control sgRNAs.
Figure 2.9. *Slc7a11* is activated by ATF4 during liver repopulation.

(A) Schematic of our approach utilizing the GFP-labeled nuclear envelope protein SUN1 to isolate hepatocyte nuclei [53], followed by ATAC-seq [54,55] analysis. (B) ATAC-seq identified an open chromatin state at the promoter region of *Slc7a11* specifically in regenerating hepatocytes (n = 2, quiescent; n = 4, 1-week regeneration). (C) The open chromatin region of the *Slc7a11* promoter contains binding motifs for NRF2 and ATF4. (D) ChIP-qPCR showed a 4-fold enrichment of ATF4 binding to the *Slc7a11* promoter after 4 weeks of liver regeneration, while no enrichment in NRF2 binding was observed. The Wilcoxon rank-sum test was used to compare the differential binding in regenerating and quiescent livers (n = 3, quiescent; n = 6, 4-week regeneration).
DISCUSSION

Here, we performed what we believe to be the first expression profile specific to repopulating hepatocytes by integrating the TRAP assay with the Fah<sup>−/−</sup> mouse model. We identified important signaling networks and regulators, including upregulation of the cell-cycle and GSH metabolic pathways, and several activated transcription factors such as MYC and FOXM1. Bioinformatics analysis comparing the gene expression of Fah<sup>−/−</sup> and PHx regeneration models identified pathways common to both models, i.e., cell cycle and GSH metabolism pathway genes among the congruently activated genes, and immune response pathway genes among the congruently inhibited genes. We also observed that liver damage pathways are uniquely upregulated in Fah<sup>−/−</sup> mice, while altered biosynthetic activity is a main theme in the PHx model.

A recent study utilizing single-cell technology to reconstruct the spatial heterogeneity of the liver had identified 9 distinct layers of gene expression profiles in quiescent hepatocytes [41]. We showed that transcripts identified from TRAP-seq significantly overlapped with those found in scRNA-seq, regardless of the layer, demonstrating the sensitivity and specificity of TRAP-seq in isolating transcripts from pure hepatocytes. Nonetheless, there are several differences between the 2 techniques. First, TRAP-seq utilizes bulk RNA-seq and therefore could not capture the zonal information by scRNA-seq. Second, TRAP-seq isolates mRNA bound to the ribosomal subunit L10a and hence only captures the actively translating mRNA. Third, TRAP-seq does not require cell sorting and therefore bypasses the time-consuming sample preprocessing required for scRNA-seq. Previous efforts to isolate intact regenerating hepatocytes after hydrodynamic injection has been unsuccessful, rendering TRAP-seq a valuable alternative. Future work could apply cell layer-specific expression of GFP-L10a to shed light on the zonal responses to liver injury and regeneration.

Previous work has demonstrated the importance of controlling oxidative stress during liver regeneration to allow hepatocyte replication, as an elevation of ROS induces a compensatory upregulation of GSH to inhibit irreversible cell damage and promote hepatic replication [56]. In support of the central role of GSH in liver regeneration, inhibition or deficiency of glutamate-cysteine
ligase, the rate-limiting enzyme in GSH synthesis, leads to downregulation of cyclin expression, decreased hepatocyte proliferation in vitro, and delayed regeneration after PHx [42,43]. Furthermore, GSH is depleted in acetaminophen-induced liver injury by the toxic metabolite NAPQI [57], pointing to the importance of GSH detoxification and ROS homeostasis in various regenerative paradigms.

Importantly, our results indicate that Slc7a11 becomes dramatically activated in repopulating hepatocytes, and we further show that ectopic expression of xCT concomitantly with the onset of liver injury promotes regeneration, probably by shielding repopulating hepatocytes from oxidative stress. These results highlight the therapeutic potential of activating Slc7a11 as a treatment for acute liver injury. We did not observe any health complications in mice overexpressing Slc7a11 during the 4-week period of repopulation in the Fah$^{-/-}$ mouse. However, determining whether this approach is beneficial in managing chronic liver injury and whether long term xCT activation is safe will require further examination.

Recent studies have found Slc7a11 to be highly expressed in HCC, breast cancer cells, and gastrointestinal tumors [58–60] and have shown that pharmacological xCT inhibition induces growth arrest in cancer cells and decreases tumor size in mouse models [59,60]. Therefore, it is possible that regenerating hepatocytes experience metabolic requirements similar to those seen in cancer cells to increase GSH availability. This raises the question of the safety of using xCT antagonists in patients with HCC, as both the growth of cancer cells and regenerating hepatocytes would be inhibited. Interestingly, in our gene inactivation studies, while a decrease in replicating hepatocytes during regeneration was observed, Slc7a11 inhibition did not completely abrogate liver repopulation, and the mice treated with sgRNAs against Slc7a11 were still able to restore full body weight after 4 weeks of regeneration. This observation is consistent with recent findings that xCT deficiency alone is not sufficient to induce liver injury but exacerbates injury when combined with secondary stress such as a high-iron diet [61] or inhibition of the transsulfuration pathway [62]. In addition, as discussed above, Slc7a11 was probably not inactivated for both alleles in all regenerating hepatocyte clones. Furthermore, genetic redundancy has been proposed to underlie
liver regeneration, as loss of any single gene rarely leads to complete inhibition of the regenerative process [3].

In conclusion, this study demonstrates the feasibility of TRAP-seq for cell type-specific mRNA isolation of hepatocytes and identifies Slc7a11 as a driver that promotes recovery after acute liver injury. Likewise, TRAP could be used to label other cell types in the liver to study their roles in acute liver injury. For instance, by combining the Rosa<sup>LSL-GFP-L10a</sup> mouse with a biliary-specific Cre or stellate cell-specific Cre transgene, it will be possible to profile the cell type-specific gene expression for these cells during injury and regeneration.
MATERIALS AND METHODS

All primer sequences are listed in Supplementary Digital Table 2.5.

Plasmid construction

The plasmid C2-EGFP-L10a was provided by Nathaniel Heintz (The Rockefeller University, New York, NY, USA). The GFP-L10a coding sequence was amplified by PCR using the primers L10a-R-BsiWI and MfeI-EGFP-F and subcloned into the vector pKT2/Fah-mCa//SB [15] at the EcoRI and BsiWI restriction sites. The vector utilizes the Sleeping Beauty (SB) transposon system to enable the integration of transgene sequences into the genome. The Slc7a11 cDNA was purchased (MG225346, OriGene) and amplified by PCR with the primers Slc7a11_clone_F1 and Slc7a11_psmd_bcd-R or Slc7a11-HA_bcd-R to include the HA tag. For the CRISPR/Cas9 studies, the vector pKT2/Fah-SpCas9//SB [34] was used to replace the SpCas9 with the SaCas9 sgRNA scaffold and introduce the subcloning site for further sgRNA subcloning using the oligonucleotides SaCas-9Ins-F and -R and the restriction enzymes SapI and EcoRI to generate the vector pKT2/Fah-SaCas9//SB. Next, 10 sgRNAs targeting the exon regions of Slc7a11 were designed with the online CRISPR RGEN Tools [63] and DESKGEN Cloud [64]. Ten sgRNAs against luciferase were designated as the control, and the oligonucleotides were subcloned into the pKT2/Fah-SaCas9//SB vector at the SapI restriction sites. For the ATAC-seq study, SUN1-GFP fragments with EcoRI and BsiWI restriction sites were amplified from the SUN1-GFP plasmid (a gift of Jeremy Nathans, Johns Hopkins University, Baltimore, MD, USA) with the primers MfeI-Sun1-F and BsiW1-Sun1-R and subcloned into the vector pKT2/Fah-mCa//SB to generate pKT2/Fah-Sun1-Gfp//SB. Endotoxin-free Maxi-scale DNA extraction and purification were performed with the GenElute HP Plasmid Maxiprep Kit (MilliporeSigma).

Mouse experiments

Fah−/− mice were maintained on NTBC (Swedish Orphan Biovitrum) in the drinking water (7.5 mg/l) until hydrodynamic tail-vein injection [15] of 10 μg plasmid, as specified below. For the TRAP-seq study, pKT2/Fah-Gfp-L10a//SB was injected, and the mice were euthanized 1 week (n
= 3) or 4 weeks (n = 9) after injection. Likewise, for the overexpression assay, mice were injected
with equimolar amounts of the plasmids pKT2/Fah-Gfp//SB and pKT2/Fah-Slc7a11//SB (n = 3) or
pKT2/Fah-Slc7a11-HA//SB (n = 5) and euthanized 4 weeks after injection. For the CRISPR/Cas9
studies, Mice were injected with either a mixture of 10 pKT2/Fah-sgSlc7a11//SB (n = 4) or
pKT2/Fah-sgCll//SB (n = 4) in conjunction with 110^{12} genome copies of AAV8.SaCas9 (Penn
Vector Core [65]) for 4 weeks of repopulation. For the ATAC-seq assay, mice were injected with
pKT2/Fah-Sun1-Gfp//SB (n = 4). One week after plasmid injection, the livers were harvested, and
GFP-positive nuclei were isolated by FACS. Mouse weights were measured 3 times per week over
the course of the repopulation period to ensure successful liver regeneration. \textit{Rosa}^{LSL-GFP-L10a} mice
were purchased from The Jackson Laboratory and used as a healthy control (n = 4) in the TRAP-
seq study, and the \textit{Rosa}^{LSL-GFP-L10a} mice were provided by Mitchell Lazar (University of
Pennsylvania, Philadelphia, PA, USA) as a quiescent control in the ATAC-seq (n = 2) experiments.
AAV8.TBG.PI.Cre.rBG (Penn Vector Core [65]) was injected into the tail vein of mice at 110^{11} virus
particles per mouse. Mice were euthanized after 1 week of injection, and the livers were harvested.
All animal studies were performed in 8- to 12-week-old female mice.

\textbf{Translating RNA isolation}

RNA specific for repopulating hepatocytes was isolated by TRAP [17]. Briefly, 200 mg liver
tissue was taken en bloc from mice injected with the TRAP construct and from \textit{Rosa}^{LSL-GFP-L10a}
mice, homogenized with lysis buffer, and incubated with magnetic beads that were conjugated with
anti-GFP antibodies (clones Htz-GFP-19F7 and Htz-GFP-19C8, Memorial Sloan-Kettering
Monoclonal Antibody Facility, New York, New York, USA) to affinity purify RNA that was bound by
the GFP-L10a fusion protein.

\textbf{IHC and IF}

Liver lobes were dissected from mice and fixed with 4% paraformaldehyde, embedded in
paraffin, and sectioned. For IHC, slides were rehydrated and subjected to antigen retrieval in
sodium citrate (pH 6.0). H$_2$O$_2$ (30%) was used for quenching endogenous peroxidases, and avidin D and biotin (Vector Laboratories) were used for blocking before incubation with primary antibodies overnight at 4 °C. The slides were then incubated with biotin-conjugated secondary antibody at 37 °C for 30 min. The avidin-peroxidase complex was incubated at 37 °C for 30 min (VECTASTAIN Elite Kit, Vector Laboratories). A DAB Substrate Kit for Peroxidase (Vector Laboratories) was used for development and hematoxylin for counterstaining. For IF, slides were prepared as described above. Incubation with primary antibodies was done overnight at 4 °C in a humid chamber, followed by secondary antibody incubation for 2 h at room temperature.

**Antibodies**

GFP was detected with goat anti-GFP antibody (ab6673, 1:100, Abcam) for IHC and chicken anti-GFP antibody (GFP-1020, 1:300, Aves Labs) for IF staining. We used rabbit anti-mouse Ki67 antibody (SP6, 1:300, Thermo Fisher Scientific) and anti-mouse Ki67 antibody (550609, 1:200, BD Biosciences) to detect proliferating cells, rabbit anti-HA antibody (sc-805, 1:100, Santa Cruz Biotechnology) for Slc7a11-HA-positive hepatocytes, rabbit anti-mouse FAH antibody (ab81087, 1:500 for IHC and 1:200 for IF, Abcam), and DAPI for nuclear staining.

**RNA-seq**

RNA integrity was measured using an Agilent RNA 6000 Bioanalyzer (Agilent Technologies). cDNA libraries were made from isolated RNA with a NEBNext Ultra RNA Library Prep Kit for Illumina (New England BioLabs) according to the manufacturer’s instructions. Library quality was measured with an Agilent High Sensitivity DNA Bioanalyzer, and cDNA libraries were purified and qPCR quantified (Kapa Biosystems). Twenty samples of equimolar libraries were pooled and sequenced with an Illumina HiSeq 2500.
RNA-seq data analysis

Fastq files of RNA-seq were processed using the RUM algorithm [66], with support from the University of Pennsylvania’s Next Generation Sequencing Core [67]. Differential gene expression analysis was performed using the package edgeR [68] in R software. Differentially expressed genes were identified with a cutoff of greater than 2-fold change and an FDR of less than 5%. Congruent genes in Fah+/ and PHx models were defined as genes regulated in the same direction for at least 1 time point in both models. Quantile-normalized reads were used for generating the heatmaps with the R package aheatmap, and Venn diagrams were created using Vennerable. Gene ontology was performed using the Database for Annotation, Visualization, and Integrated Discovery (DAVID) [29,30]. The top 3,000 upregulated and downregulated genes were uploaded to DAVID and analyzed using the functional annotation tool. A list of enriched KEGG pathways was obtained from the functional annotation chart report. The top-10 most significantly enriched KEGG pathways were selected and sorted according to the Bonferroni-corrected p-value. In addition, all differentially expressed genes, along with their corresponding fold change, were uploaded into the Ingenuity Pathway Analysis tool, and functional analysis was performed using the Core Analysis function. The upstream regulators predicted by Ingenuity Pathway Analysis were further filtered by (a) genes that were also changed in the RNA-seq analysis by at least 2-fold, (b) a significant Z-score (≥2 for predicted activation and ≤2 for predicted inhibition), and (c) congruence between the observed fold change and the predicted activation or inhibition.

Quantitative reverse-transcription with polymerase chain reaction (qRT-PCR)

Extracted RNAs were reverse transcribed to cDNA with SuperScript II Reverse Transcriptase (Invitrogen, Thermo Fisher Scientific), and qRT-PCR was performed with Slc7a11 primers (Slc7a11-qRTPCR-F and -R), Gsta1 primers (Gsta1-qRTPCR-F and -R), and Tbp primers (Tbp-qRTPCR-F and -R). Relative expression levels were normalized to Tbp.
Quantitative PCR (qPCR)

Genomic DNA was extracted from mice injected with equimolar amounts of pKT2/Fah-Gfp//SB (Fah-Gfp) and pKT2/Fah-Slc7a11-HA//SB (Fah-Slc7a11-HA) over a 4-week period with a DNeasy Blood and Tissue Kit (QIAGEN), followed by ethanol precipitation. qPCR was performed with PrimeTime primer sets (IDT DNA) Slc7a11 (Slc7a11-qPCR-F, -R, and -P) and Gfp (Gfp-qPCR-F, -R, and -P). Standard curves were generated by performing a serial dilution of the input plasmid with equimolar amounts of Fah-Gfp and Fah-Slc7a11-HA.

Western blotting

Proteins were extracted from whole-liver homogenate with lysis buffer containing 50 mM Tris, pH 7.5, 0.5 mM EDTA, 150 mM NaCl, 10% glycerol, 1% NP-40, and 1% SDS, supplemented with 1:100 Halt Protease Inhibitor Cocktail (Thermo Fisher Scientific). The lysates were sonicated at 30-second intervals for 5 min and electrophoresed on 4% to 12% NuPAGE Precast Gels (Life Technologies, Thermo Fisher Scientific). A nitrocellulose membrane was used for transfer, and 5% milk in TBST (TBS plus 0.1% Tween-20) was used to block the membrane at room temperature for 1 h. The anti-mouse xCT antibody (sc-79360, 1:200, Santa Cruz Biotechnology) was diluted in 5% BSA in TBST and incubated overnight at 4 °C. The membrane was washed with TBST 3 times for 10 min, followed by an HRP-conjugated secondary antibody in 5% milk in TBST for 1 h, and then exposed to film.

Hepatocyte nuclei isolation and sorting

Livers were harvested and nuclei isolation was performed as previously described [69]. Briefly, liver was dounced in a pestle tissue grinder in 10 ml hypotonic buffer (10 mM Tris-HCl, pH7.5, 2 mM MgCl2, 3 mM CaCl2) on ice. The homogenate was passed through a 100-μm filter and sedimented at 400 g at 4 °C for 10 min. The pellet was resuspended in 10 ml hypotonic buffer with 10% glycerol, and 10 ml lysis buffer (hypotonic buffer, 10% glycerol, 1% IGEPAL CA-630) was
added dropwise. After 5 min of incubation, the solution was centrifuged at 600 g for 5 min at 4 °C. The isolated nuclei were washed again in lysis buffer, and nuclei were counted in a hemocytometer.

Isolated hepatocyte nuclei were labeled with an Alexa Fluor 647 anti-GFP antibody (338006, clone FM264G, 1:25, BioLegend). Immediately before cell sorting, the nuclei suspension was stained with 2 μg/ml DAPI. GFP- and AF647–double-positive nuclei were sorted using a BD FACS Aria II, after gating for DAPI-positive nuclei. Because of the polyploidy state of the hepatocytes, only 4n nuclei were collected.

ATAC-seq

Sorted hepatocyte nuclei were tagmented and PCR amplified according to a previously published ATAC-seq protocol [54,55]. Briefly, 25,000 nuclei were aliquoted, and transposition was performed at 37 °C for 30 min. The transposition reaction was stopped by Buffer ERC (QIAGEN), and DNA was purified using the QIAGEN MinElute Reaction Cleanup Kit. Genomic fragments were preamplified for 5 cycles, and the final amplification cycle was determined by qPCR. The libraries were size selected with Agencourt AMPure XP beads (Beckman Coulter) and sequenced with an Illumina HiSeq 4000. ATAC-seq data analysis. Fastq files from ATAC-seq were analyzed with the pipeline developed by Anshul Kundaje (Stanford University, Stanford, CA, USA) [70]. Briefly, for each sample, adapters were trimmed and aligned to the genome mm9 with Bowtie. The aligned bam files of biological replicates were then merged and subjected to peak calling of open chromatin regions. The parameters for the analysis were -auto_detect_adapter-enable_idr -filt_bam-sample1 -filt_bam-sample2 … -filt_bam-sampleN.

ChIP-qPCR

The Slc7a11 promoter was analyzed, and potential NRF2- and ATF4-binding motifs were identified with JASPAR [71]. Liver chromatin was prepared as previously described [72]. Briefly, 100 mg liver was fixed and sonicated with a Bioruptor (Diagenode) for 2 rounds of 7.5 min each. Sheared DNA (10 μg) was then incubated with anti-ATF4 antibody (D4B8, 1:200, Cell Signaling
Technology) and NRF2 antibodies (D1Z9C, 1:100, Cell Signaling Technology). Immunoprecipitated DNA was then isolated with phenol:chloroform extraction and subjected to qPCR analysis with the primers S1c7a11-Nrf2/Atf4-ChIP-qPCR-2F and -2R. Fold enrichment was calculated by normalization to the average Ct value of Ins (Ins-ChIP-qPCR-F and -R) and Arbp (Arbp-ChIP-qPCR-F and -R) compared with input DNA. Sequencing data download. TRAP-seq and ATAC-seq data have been deposited according to MINSEQE standards in the NCBI’s Gene Expression Omnibus database (GEO GSE109466) [73]. RNA-seq data from PHx are available in the ArrayExpress database (accession no.E-MTAB-1612).

Statistics

Unless otherwise indicated, a 2-tailed, 2-sample Student’s t-test was used to analyze the experimental and control groups in all assays performed in this study. A 2-tailed, 1-sample Student’s t-test was used to compare the ratio of Fah-Slc7a11 to Fah-Gfp plasmids after liver repopulation with the injected plasmid mix. A hypergeometric test was used to analyze the overlapping genes in the scRNA-seq and TRAP-seq experiments. The Wilcoxon rank-sum test was used to compare the differential binding of NRF2 and ATF4 in the repopulating and quiescent livers. A P value or FDR of less than 0.05 was considered significant. Individual data are presented as dot plots, with the mean shown as a horizontal line. Study approval. All animal studies were reviewed and approved by the IACUC of the Penn Office of Animal Welfare (University of Pennsylvania).
REFERENCES


CHAPTER 3

CELL TYPE-SPECIFIC EXPRESSION PROFILING IN THE MOUSE LIVER

Parts of this chapter were adapted with permission from Cell type-specific gene expression profiling in the mouse liver. Wang AW, Zahm AM, Wangensteen KJ. The Journal of Visualized Experiments. 2019;151:e60242.doi:10.3791/60242.
ABSTRACT

Liver repopulation after injury is a crucial feature of mammals which prevents immediate organ failure and death after exposure to environmental toxins. A deeper understanding of the changes in gene expression that occur during the regenerative process could help identify therapeutic targets to promote the restoration of liver function in the setting of injuries. Nonetheless, methods to isolate specifically the repopulating hepatocytes are inhibited by a lack of cell markers, limited cell numbers, and the fragility of these cells. The development of the translating ribosome affinity purification (TRAP) method in conjunction with the $Fah^{-/-}$ mouse model to recapitulate repopulation in the setting of liver injury allows gene expression profiling of the repopulating hepatocytes. With TRAP, cell type-specific translating mRNA is rapidly and efficiently isolated. We developed a method that utilizes TRAP with affinity-based isolation of translating mRNA from hepatocytes that selectively express the green fluorescent protein (GFP)-tagged ribosomal protein (RP) L10A, GFP:RPL10A. TRAP circumvents the long time period required for fluorescence-activated cell sorting (FACS) that could change the gene expression profile. Furthermore, since only the repopulating hepatocytes express the GFP:RPL10A fusion protein, the isolated mRNA is devoid of contamination from the surrounding injured hepatocytes and other cell types in the liver. The affinity-purified mRNA is of high quality and enables downstream PCR- or high-throughput sequencing-based analysis of gene expression.
INTRODUCTION

As the main metabolic organ in vertebrates, the liver is responsible for glucose homeostasis, serum protein synthesis, bile acid secretion, and xenobiotic metabolism and detoxification. The liver possesses an extraordinary capacity to regenerate the injured parenchyma upon exposure to toxins to prevent immediate liver dysfunction [1]. However, failure of regeneration can occur in the setting of acetaminophen or alcohol overconsumption, which can lead to acute liver failure [2]. Furthermore, chronic liver injury caused by viral hepatitis infection, fatty liver disease, and steatohepatitis frequently result in liver fibrosis, cirrhosis, and hepatocellular carcinoma [3]. The only available curative treatment for end-stage liver disease is transplantation but is currently limited by organ shortage, preventing efficient treatment for all patients [4]. A better understanding of the recovery process after toxic liver injury is therefore crucial for the development of treatments to stimulate regeneration sufficient to rescue function in the diseased organ.

The most broadly-applied model system for the study of liver regeneration is partial hepatectomy in rodents, in which a large proportion of the liver is resected to stimulate rapid hepatocyte expansion [5]. However, partial hepatectomy does not recapitulate hepatocyte expansion following toxic liver injury due to the lack of immune cell infiltration and hepatocyte cell necrosis often observed in the setting of acute liver injury in humans [6]. A more suitable system to model this form of organ renewal is the Fah−/− mouse, which lacks functional fumarylacetoacetate hydrolase (FAH) required for proper tyrosine catabolism, and develops severe liver damage leading to death [7]. These mice can be maintained in a healthy state indefinitely by treatment with the drug nitisinone in the drinking water. Alternatively, FAH expression can be restored by transgene delivery to a subset of hepatocytes, which will expand to repopulate the liver upon nitisinone removal [8].

To profile the gene expression changes of repopulating hepatocytes, a tool to specifically isolate these cells in the Fah−/− mouse without contamination from the neighboring injured hepatocytes and other cell types is required. Unfortunately, fluorescence-assisted cell sorting (FACS) of hepatocytes is difficult since (1) the fragility of repopulating cells leads to poor recovery
after liver perfusion, (2) replicating hepatocytes are highly variable in size, making isolation of a pure population by FACS difficult, and (3) the procedure time from liver perfusion to RNA isolation is greater than 2 h, hence gene expression profiles may undergo substantial artificial changes prior to sample acquisition [9].

Alternatively, the expression of epitope-tagged ribosomes specifically in repopulating hepatocytes enables the rapid isolation of actively translating mRNA bound by ribosomes using affinity purification immediately after organ harvest with bulk liver tissue lysates. Here, we describe a protocol to perform translating-ribosome affinity purification (TRAP) [10] followed by high-throughput RNA-sequencing (TRAP-seq) to specifically isolate and profile mRNA in repopulating hepatocytes in the Fah−/− mouse [9]. Coexpression of green fluorescent protein (GFP)-tagged ribosomal protein (RP) L10A (GFP:RPL10A) with FAH allows affinity purification of translating mRNA bound by polysomes containing GFP:RPL10A. This method avoids any cell dissociation steps, such as liver perfusion to isolate fragile repopulating hepatocytes. Instead, TRAP utilizes whole organ tissue lysis and antibodies to rapidly extract the RNA specifically from target cells. Finally, isolation of abundant, high-quality mRNA via TRAP-seq enables downstream applications such as sequencing analysis to profile the dynamic change of gene expression during the repopulation process.
PROTOCOL

All methods that involve the use of mice are consistent with the guidelines provided by the Institutional Animal Care and Use Committee (IACUC) of the Penn Office of Animal Welfare at the University of Pennsylvania.

1. Reagent preparation

1.1. Cycloheximide. To make 500 μl of 0.1 g/ml cycloheximide, suspend 50 mg of cycloheximide in 500 μl of methanol. Cycloheximide can be stored at 4 °C for up to 1 day.

**NOTE:** Cycloheximide inhibits translation.

**CAUTION:** cycloheximide is extremely toxic to the environment and can cause congenital malformation. All wastes and buffers containing cycloheximide should be collected for proper disposal.

1.2. DTT. To make 1 ml of 1M DTT, suspend 0.15 g of DTT powder in RNase-free water. DTT can be stored at -20 °C. It is recommended to store 1M DTT in single-use aliquots of 50 μl.

**NOTE:** DTT is a detergent.

**CAUTION:** DTT can cause irritation to the skin, eye, and respiratory tract.

1.3. Deoxycholate (DOC). To make 10% DOC, suspend 1 g of DOC in a 50 ml conical tube and add RNase-free water up to 10 ml. Shake vigorously until the powder is dissolved. The 10% DOC solution is slightly yellow and can be stored at RT for up to 1 year.

**NOTE:** DOC is used for nuclear lysis.

1.4. GFP antibodies. Aliquot GFP antibodies when using for the first time. Snap freeze the aliquots and store at -80 °C. It is recommended to store 50 μg of GFP antibodies in single-use aliquots.

1.5. Biotinylated protein L. Resuspend biotinylated protein L in 1X PBS to make the final concentration 1 μg/μl. The resuspended solution can be stored at -20 °C for up to 6 months.
2. **Buffer preparation**

2.1. **BSA buffer.** To make 50 ml of 3% BSA buffer, add 1.5 g of IgG- and protease-free BSA powder into 40 ml of PBS followed by a quick vortex. After the BSA is dissolved, add PBS to a final volume of 50 ml. The BSA buffer can be stored at 4 °C for up to 6 months.

2.2. **Dissection buffer.** To make 50 ml of dissection buffer stock, combine 5 ml of 10X HBSS, 125 μl of 1M HEPES, 1750 μl of 1M glucose, and 200 μl of 1M NaHCO₃. Add RNase-free water to a final volume of 50 ml. The dissection buffer stock can be stored at 4 °C for up to 6 months. Immediately prior to use, add 100 μg/ml of 0.1 g/ml cycloheximide and keep on ice.

2.3. **High-salt buffer.** To make 50 ml of high-salt buffer stock, add 1 ml of 1M HEPES, 8.75 ml 2M KCl, 500 μl 1M MgCl₂, and 500 μl 100% branched octylphenoxy poly(ethyleneoxy)ethanol (IGEPAL) to RNase-free water. The high-salt buffer stock can be stored at 4 °C for up to 6 months. Immediately prior to use, add 0.5 μl/ml of 1M DTT and 1 μl/ml of 0.1 g/ml cycloheximide. Keep the fresh high-salt buffer on ice.

2.4. **Low-salt buffer.** To make 50 ml of low-salt buffer stock, add 1 ml of 1M 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES), 3.75 ml of 2M KCl, 500 μl of 1M MgCl₂, and 500 μl of 100% IGEPAL to 44.25 ml RNase-free water. The low-salt buffer stock can be stored at 4 °C up to 6 months. Add 0.5 μl/ml of 1M DTT and 1 μl/ml of 0.1 g/ml cycloheximide prior to use. Keep the fresh low-salt buffer on ice.

2.5. **Tissue lysis buffer.** To make 50 ml of tissue lysis buffer stock, combine 1 ml of 1M HEPES, 3.75 ml of 2M KCl, and 500 μl of 1M MgCl₂. Add RNase-free water to a final of 50 ml. The dissection buffer stock can be stored at 4 °C up to 6 months. Add 1 tab/ml of EDTA-free protease inhibitor, 1 μl/ml of 0.1 g/ml cycloheximide, 10 μl/ml of RNase inhibitors each immediately prior to use. Keep the fresh tissue lysis buffer on ice.
3. Conjugation of antibodies to magnetic beads

3.1. Antibodies

3.1.1. Calculate the amount of GFP antibodies required for all samples and prepare for one extra sample. For each sample, 50 μg of each GFP antibody is required.

3.1.2. Thaw GFP antibodies on ice and spin at maximum speed (> 13,000 x g) for 10 min at 4 °C and transfer supernatants to a new Eppendorf.

**NOTE:** The antibody preparation step can be performed prior to bead preparation and the thawed antibodies can be kept on ice. Alternatively, this step can be performed during incubation of magnetic beads with biotinylated protein L.

3.2. Resuspend magnetic beads

3.2.1. Resuspend magnetic beads by gentle pipetting. For each sample, 150 μl of magnetic bead is used. Calculate the volume of magnetic bead required for all samples and prepare one extra.

3.2.2. Transfer the resuspended magnetic beads to a 1.5 or 2 ml Eppendorf. If more than 1 ml is required for an experiment, split the total amount into equal volumes.

3.2.3. Collect beads on a magnetic stand for > 1 min and remove the supernatant. Remove Eppendorf from the magnetic stand and add 1 ml PBS followed by pipetting up and down to wash the beads. Collect beads on a magnetic stand for > 1 min and remove PBS.

3.3. Preparation of protein L-coated beads

3.3.1. Take the amount of biotinylated protein L required for all samples and prepare one extra. For each sample, 60 μl of biotinylated protein L is used. If protein L is previously resuspended and stored at -20 °C, thaw on ice.

3.3.2. Add the calculated volume of biotinylated protein L to the resuspended and washed magnetic beads. Add 1X PBS to make the final volume 1 ml if using a 1.5 ml Eppendorf, or 1.5 ml if using a 2 ml Eppendorf. Incubate magnetic beads with biotinylated protein L for 35 min at RT on a tube rotator.

**NOTE:** Antibodies can be prepared at this step during bead incubation with protein L.
3.3.3. Collect protein L-coated beads on a magnetic stand for > 1 min and remove the supernatant. Remove the Eppendorf tube from the magnetic stand and add 1 ml of 3% BSA buffer followed by gentle pipetting for 5 times to wash the protein L-coated beads.

3.3.4. Collect coated beads on a magnetic stand for > 1 min and remove the supernatant. Repeat the washing steps with 3% BSA for another 4 times (a total of 5 times).

3.4. Antibody binding

3.4.1. Add the calculated amount of GFP antibodies into the protein L-coated beads and incubate for 1 h at 4 °C on a tube rotator.

**NOTE:** After antibody incubation, take special care to not vortex or vigorously shake the affinity matrix as it could disrupt the binding of biotinylated protein L to the magnetic beads.

3.4.2. During incubation, prepare low-salt buffer by calculating the total volume required for all samples and add 0.5 μl/ml of 1M DTT and 1 μl/ml of 0.1 g/ml cycloheximide to low-salt buffer stock prior to use. 3 ml of low-salt buffer for washing each tube of GFP-conjugated beads and 200 μl/sample for resuspension of the GFP-conjugated beads are required. Fresh low-salt buffer can be kept on ice for a couple of hours.

3.4.3. Collect the affinity matrix on a magnetic stand for > 1 min and remove the supernatant. Add 1 ml of low-salt buffer and gently pipette up and down to wash the affinity matrix.

3.4.4. Collect the affinity matrix on a magnetic stand for > 1 min and remove low-salt buffer. Repeat the washing steps with low-salt buffer for another 2 times (a total of 3 times).

3.4.5. Resuspend the beads in low-salt buffer so that each sample has 200 μl of affinity matrix.

3.4.6. The affinity matrix can be stored in 0.02% NaN₃ at 4 °C for up to 2 weeks. The affinity matrix should be quickly washed in low-salt buffer 3 times and resuspended gently on a tube rotator at 4 °C for at least 10 min if the affinity matrix is prepared within 1 week or overnight if the affinity matrix is stored for over 1 week.

**CAUTION:** Sodium azide is extremely toxic to the environment. Contact with acids produces toxic gas. All wastes should be collected for proper disposal.

**NOTE:** The protocol can be paused after this step.
4. Liver tissue lysis

4.1. Buffer preparation and equipment setup

4.1.1. Calculate the number of Eppendorf tubes required, label and chill on ice. Usually, 7 1.5 ml Eppendorf tubes are required for each sample. 1 for the remaining dissected liver, 4 for 4 ml of homogenized liver lysate, and 2 for transferring supernatants.

4.1.2. Prepare fresh dissection buffer by calculating the total volume required for all samples and add 1 μl/ml of 0.1 g/ml cycloheximide. Place the fresh dissection buffer on ice to keep cold throughout the experiment. For each sample, 10 ml of dissection buffer is required.

4.1.3. Prepare fresh lysis buffer by calculating the total volume required for all samples and add 1 tab/10 ml of EDTA-free protease inhibitor, 1 μl/ml of 0.1 g/ml cycloheximide, and 10 μl/ml of RNase inhibitors each. Keep the lysis buffer on ice throughout the experiment. For each sample, 4 ml of lysis buffer is required.

4.1.4. Setup the homogenizer apparatus so that the Teflon-glass tubes can be placed on ice during homogenization of liver pieces. Put 4 ml of cold lysis buffer in the Teflon-glass tubes.

4.2. Repopulating liver homogenization

4.2.1. Euthanize 8-12-week-old Fah<sup>+/−</sup> mice injected with the TRAP vector and repopulated for one to four weeks with anesthesia and cervical dislocation according to approved animal experimental guidelines.

4.2.2. Place mice on a dissection board and spray the abdomen with 70% ethanol. Tent the skin and peritoneum using forceps and use scissors to make a transverse incision low in the abdomen and continue to cut with the scissors to make a wide U-shaped peritoneal flap, with care to not cut the viscera. Flip the peritoneal flap over the sternum to expose the liver.

4.2.3. Carefully remove the liver with scissors and forceps and quickly place the tissue in cold dissection buffer to rinse. To homogenize frozen tissues, quickly move the desired amount of liver tissue into Teflon-glass tubes with cold lysis buffer without the tissue thawing.

**NOTE:** The dissected tissue can be flash-frozen and stored at -80 °C after it is washed with dissection buffer. The protocol can be paused after this step.
4.2.4. Weigh the liver on a Petri dish, Isolate 200-500 mg of liver, and transfer to the Teflon-glass tubes. Place the remaining tissue into a pre-chilled microcentrifuge tube and flash freeze.

**NOTE:** The amount of tissue used is based on the abundance of the cell type of interest.

4.2.5. Homogenize the tissue in a motor-driven homogenizer starting at 300 rpm to dissociate hepatocytes from the liver structure for at least 5 strokes. Lower the glass tube each time but take care to not let the pestle rise above the solution to prevent aeration that could cause protein denaturation.

4.2.6. Raise the speed to 900 rpm to fully homogenize the liver tissues for at least 12 full strokes.

4.2.7. Transfer the lysate into labeled and pre-chilled Eppendorf tubes, with no more than 1 ml of lysate per 1.5 ml tube. If 4 ml of lysis buffer is used, keep 1 tube and flash freeze the remaining 3 tubes.

**NOTE:** The lysates can be kept on ice for up to 1 h while dissecting the next animal and preparing fresh lysates. The homogenized liver can be flash-frozen after the lysis step and stored at -80 °C. There could be a 50% decrease in isolated RNA if frozen lysates are used. The protocol can be paused after this step.

4.3. Nuclear lysis

4.3.1. Centrifuge the liver lysate at 2,000 x g at 4 °C for 10 min and transfer the supernatant to a new, prechilled Eppendorf on ice.

4.3.2. Add 1/9 of the supernatant volume of 10% IGEPAL to make a final concentration of 1% and mix by gently inverting the Eppendorf tubes.

4.3.3. Quickly spin down the Eppendorf tubes and add 1/9 of the sample volume of 10% DOC to make a final concentration of 1% and mix by gently inverting the Eppendorf tubes. Quickly spin down the Eppendorf tubes and incubate on ice for 5 min.

4.3.4. Centrifuge the nuclear lysate at 20,000 x g at 4 °C for 10 min and transfer the supernatant to a new, prechilled Eppendorf on ice.

**NOTE:** The mitochondria-depleted supernatant can be placed on ice for a couple of hours while the remaining samples are being collected.
5. **Immunoprecipitation**

5.1. For each sample, take out 1% of the total volume of the mitochondria-depleted supernatant as a pre-immunoprecipitation control to compare target enrichment after incubation with the affinity matrix. Place the pre-immunoprecipitation controls on a tube rotator at 4 °C overnight, the same way as the immunoprecipitated samples are processed.

5.2. Add 200 μl of affinity matrix to each sample. Take extra care to resuspend the beads by gentle pipetting prior to adding the affinity matrix to each sample. Incubate the lysates with affinity matrix at 4 °C overnight with gentle mixing on a tube rotator.

**NOTE:** The protocol can be paused for up to a day after this step.

6. **RNA isolation**

6.1. Removal of unbound background noise

6.1.1. Place the magnetic rack at 4 °C for at least 30 min to pre-chill and keep the rack on ice throughout the experiment.

6.1.2. Calculate the number of Eppendorf tubes required and pre-chill on ice or at 4 °C. Usually, each sample requires 1 Eppendorf tube for the final purified RNA.

6.1.3. Quickly spin down the supernatant incubated with the affinity matrix and collect the beads by placing on the magnetic rack for at least 1 min. Collect or discard the supernatant that contains the unbound fraction in additional Eppendorf tubes.

**NOTE:** The collected supernatant can be flash-frozen and stored at -80 °C to compare with the bound fraction for transcript enrichment after purification.

6.1.4. Prepare high-salt buffer by adding 0.5 μl/ml of 1M DTT and 1 μl/ml of 0.1 g/ml cycloheximide to high-salt buffer stock. 5 ml of high-salt buffer is required for each sample.

6.1.5. Add 1 ml of fresh high-salt buffer to each tube followed by gentle pipetting for at least 5 times without introducing bubbles.

**NOTE:** Insufficient washing could introduce backgrounds of unbound transcripts while the introduction of bubbles could accelerate RNA degradation.
6.1.6. Collect beads on a magnetic stand for > 1 min and remove the supernatant. Repeat the washing steps with high-salt buffer for another 4 times (a total of 5 times).

6.1.7. Remove remaining high-salt buffer and remove Eppendorf tubes from the magnetic stand and place at RT for 5 min to warm up.

6.2. RNA isolation with column-based kits

6.2.1. Resuspend the beads in 100 μl of lysis buffer with β-mercaptoethanol, both provided in the RNA isolation kit.

**NOTE:** Any RNA isolation and purification kit that contains the denaturant guanidine thiocyanate in the lysis buffer can be used to release bound RNA from the affinity matrix. RNA extraction should be processed at room temperature since guanidine thiocyanate can crystallize at low temperatures.

6.2.2. Vortex the beads and buffer for at least 5 sec at the highest speed, quickly spin down to collect the buffer on the side of the Eppendorf and incubate the beads at RT for 10 min to release the bead-bound RNA into the lysis buffer.

6.2.3. Collect beads on a magnetic stand for > 1 min and collect the supernatant to proceed immediately to RNA cleanup according to the RNA purification protocol as specified in the kit.

**NOTE:** The supernatant containing the eluted RNA in lysis buffer can also be stored at -80 °C for up to 1 month prior to cleanup. To proceed after storage, warm up the tubes to RT upon thawing.

6.2.4. To achieve maximum quality of the isolated RNA, perform all optional steps including DNase digestion and all RNA elution steps. Heat up the elution buffer provided by the RNA isolation kit or RNase-free water to 60 °C for maximum RNA recovery.

**NOTE:** The isolated RNA can be stored at -20 °C for up to 1 month or -80 °C for several years. The protocol can be paused after this step.
7. **Optional RNA quality analysis (recommended)**

7.1. Assess RNA quality using a Bioanalyzer and quantity with a Nanodrop to determine if repeating the immunoprecipitation process is required to obtain ample and high-quality RNA.

**NOTE:** The optimal RNA quality for high-throughput sequencing should follow protocols specified by individual library preparation kits and sequencing platforms.

8. **Downstream applications**

**NOTE:** Total RNA isolated by the TRAP protocol can be used in a number of standard downstream applications, including RNA-seq (TRAP-seq) and reverse transcription and quantitative PCR (RT-qPCR).

8.1. **RNA-seq.** Prepare cDNA sequencing libraries using commercial RNA-seq kits with oligo d(T)-based enrichment of polyadenylated (poly(A)) transcripts. Alternatively, if the total RNA quality is lower than recommended for poly(A) enrichment, use rRNA depletion modules. However, expect to see more rRNA alignment after sequencing.

8.2. **RT-qPCR.** Standard reverse transcription and quantitative PCR protocols can be used following TRAP.
**REPRESENTATIVE RESULTS**

To profile gene expression in repopulating hepatocytes of the $Fah^{-/-}$ mouse, $Gfp:Rpl10a$ fusion and $Fah$ transgenes are co-delivered within a transposon-containing plasmid [8] (TRAP vector) to livers by hydrodynamic injection (Figure 3.1A). The removal of nitroprusside induces a toxic liver injury that creates a selection pressure for hepatocytes stably expressing FAH to repopulate the injured parenchyma [9]. Immunofluorescence staining confirms the co-expression of FAH and the GFP:RPL10A fusion protein in repopulating hepatocytes after two weeks of liver repopulation (Figure 3.1B).

In the following representative experiment, TRAP-seq was performed using quiescent and repopulating mouse hepatocytes. First, to obtain GFP-tagged ribosomes from quiescent hepatocytes, transgenic Rosa$^{LSL-GFP-L10A}$ mice were injected with AAV8-TBG-Cre 7 days prior to sacrifice to induce GFP:RPL10A expression in all hepatocytes [11]. We also processed a liver sample collected from a wild type mouse as a negative control to ensure isolation of translating mRNA was specific, meaning RNA could only be extracted from mice expressing GFP:RPL10A. The concentration of isolated RNA correlated with the number of cells expressing the fusion protein; the quiescent sample displays the highest yield since all hepatocytes express GFP:RPL10A after AAV8-TBG-Cre injection (Figure 3.2A). Conversely, barely any RNA was detectable in wild type controls that did not possess the GFP:RPL10A transgene, indicating the TRAP procedure is highly specific and has a low background. When TRAP was used on liver tissues undergoing repopulation with GFP:RPL10A-transduced hepatocytes, abundant, high-quality RNA was obtained while no RNA trace was detected via Bioanalyzer for the negative control sample (Figure 3.2B).

Downstream gene expression analysis can be carried out via RT-qPCR or RNA-seq on TRAP-isolated RNA. $Gsta1$ encodes glutathione S-transferase that plays an important role in the metabolism of glutathione, the main detoxifying peptide to protect cellular oxidative stress damage [12]. $Gsta1$ expression is induced by over 10-fold in repopulating hepatocytes as compared to quiescent hepatocytes, while no CT cycle was detected with TRAP-isolated RNA from the wild type mouse due to the lack of input RNA (Figure 3.3A). Note that RNA quality can greatly impact gene
expression analysis. In the case of RNA-seq experiments, assessment of RNA quality should be performed according to the recommendations of the library preparation kit and the sequencing platform (Figure 3.3B). A Bioanalyzer is often used to determine the RNA integrity number (RIN), with a high RIN correlating with a higher rate of mRNA alignments to the genome (Figure 3.3B, left), whereas a lower RIN leading to a higher rate of ribosomal reads, indicating mRNA degradation (Figure 3.3B, right). Figures 3.3C and D demonstrate that TRAP-seq can identify differential gene expression in quiescent and repopulating hepatocytes. For instance, Alb expression is inhibited and Afp expression is activated during liver repopulation, reflecting that the regenerating hepatocytes assume a less differentiated state to inhibit liver metabolic functions during repopulation [9,13].
### Table 3.1. Materials for the TRAP-seq protocol.

<table>
<thead>
<tr>
<th>Name of Material/Equipment</th>
<th>Company</th>
<th>Catalog Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>10 ml Tissue Grinder, Potter-Elv, Coated</td>
<td>DWK Life Sciences (Wheaton)</td>
<td>358007</td>
</tr>
<tr>
<td>Absolutely RNA Miniprep Kit</td>
<td>Agilent</td>
<td>400800</td>
</tr>
<tr>
<td>Anti-GFP antibodies</td>
<td>Memorial Sloan-Kettering Antibody &amp; Bioresource Core</td>
<td>GFP Ab #19C8</td>
</tr>
<tr>
<td>Bovine Serum Albumin, IgG-Free, Protease-Free</td>
<td>Jackson</td>
<td>001-000-162</td>
</tr>
<tr>
<td>cComplete, Mini, EDTA-free Protease Inhibitor Cocktail</td>
<td>ImmunoResearch</td>
<td></td>
</tr>
<tr>
<td>Cycloheximide</td>
<td>Millipore Sigma</td>
<td>C7698</td>
</tr>
<tr>
<td>D-Glucose, Dextrose</td>
<td>Fisher Scientific</td>
<td>D16</td>
</tr>
<tr>
<td>Deoxycholic acid, DOC</td>
<td>Millipore Sigma</td>
<td>D2510</td>
</tr>
<tr>
<td>DL-Dithiothreitol</td>
<td>Millipore Sigma</td>
<td>D9779</td>
</tr>
<tr>
<td>Dynabeads MyOne Streptavidin T1</td>
<td>Thermo Fisher Scientific</td>
<td>65602</td>
</tr>
<tr>
<td>Fisherbrand Petri Dishes with Clear Lid</td>
<td>Fisher Scientific</td>
<td>FB0875712</td>
</tr>
<tr>
<td>HBSS (10X), calcium, magnesium, no phenol red</td>
<td>Thermo Fisher Scientific</td>
<td>14065-056</td>
</tr>
<tr>
<td>HEPES, 1M Solution, pH 7.3, Molecular Biology Grade, Ultrapure, Thermo Scientific</td>
<td>Thermo Fisher Scientific</td>
<td>AAJ16924AE</td>
</tr>
<tr>
<td>IGEPAL CA-630 (Octylphenoxy poly(ethyleneoxy)ethanol, branched)</td>
<td>Millipore Sigma</td>
<td>I8896</td>
</tr>
<tr>
<td>Magnesium chloride, MgCl2</td>
<td>Millipore Sigma</td>
<td>M8266</td>
</tr>
<tr>
<td>Methanol</td>
<td>Fisher Scientific</td>
<td>A452</td>
</tr>
<tr>
<td>NEBNext Poly(A) mRNA Magnetic Isolation Module</td>
<td>New England BioLabs</td>
<td>E7490S</td>
</tr>
<tr>
<td>NEBNext Ultra RNA Library Prep Kit for Illumina</td>
<td>New England BioLabs</td>
<td>E7530S</td>
</tr>
<tr>
<td>Nuclease-Free Water, not DEPC-Treated</td>
<td>Ambion</td>
<td>AM9932</td>
</tr>
<tr>
<td>Overhead Stirrer</td>
<td>DWK Life Sciences (Wheaton)</td>
<td>903475</td>
</tr>
<tr>
<td>PBS Buffer (10X), pH 7.4</td>
<td>Ambion</td>
<td>AM9625</td>
</tr>
<tr>
<td>Pierce Recombinant Protein L, Biotinylated</td>
<td>Thermo Fisher Scientific</td>
<td>29997</td>
</tr>
<tr>
<td>Potassium chloride, KCl</td>
<td>Millipore Sigma</td>
<td>P4504</td>
</tr>
<tr>
<td>RNaseZap RNase Decontamination Solution</td>
<td>Invitrogen</td>
<td>AM9780</td>
</tr>
<tr>
<td>RNasin Ribonuclease Inhibitors</td>
<td>Promega</td>
<td>N2515</td>
</tr>
<tr>
<td>RNA 6000 Pico Kit &amp; Reagents</td>
<td>Agilent</td>
<td>5067-1513</td>
</tr>
<tr>
<td>Sodium azide, NaN3</td>
<td>Millipore Sigma</td>
<td>S2002</td>
</tr>
<tr>
<td>Sodium bicarbonate, NaHCO3</td>
<td>Millipore Sigma</td>
<td>S6297</td>
</tr>
<tr>
<td>SUPERase·In RNase Inhibitor</td>
<td>Invitrogen</td>
<td>AM2694</td>
</tr>
</tbody>
</table>
FIGURES

Figure 3.1. Implementation of translating ribosome affinity purification (TRAP) with \( \textit{Fah}^{+/} \) to profile gene expression change of repopulating hepatocytes.

(A) Schematic of expressing the green fluorescent protein (GFP)-tagged ribosome protein (RP) subunit L10A (GFP:RPL10A) with FAH in the \textit{Sleeping Beauty} transposon system followed by injection into the \( \textit{Fah}^{+/} \) mouse. Green hexagons indicate repopulating hepatocytes with stable expression of FAH and GFP:RPL10A, whereas black hexagons represent injured, dying hepatocytes. (B) Representative immunofluorescence staining demonstrates coexpression of FAH (red) and GFP-tagged ribosomal protein L10A (green) in the repopulating hepatocytes. Scale bar, 50 μm.
**Figure 3.2.** TRAP enables cell type-specific isolation of high-quality RNA.

(A) The yield of RNA is positively correlated with the number of hepatocytes expressing GFP:RPL10A. The low yield of RNA from a wild type mouse demonstrates the specificity of TRAP from sources without the expression of GFP:RPL10A. (B) Bioanalyzer traces of total RNA isolated from repopulating livers expressing GFP:RPL10A and from wild type livers demonstrate the specificity of TRAP. Total RNA isolated from wild type liver tissue devoid of the GFP:RPL10A transgene shows that minimal RNA has been collected, whereas transgene-expressing tissues provide ample high-quality RNA. Note that ribosomal RNA peaks are present following successful TRAP [10]. FU, Fluorescence unit. RIN, RNA integrity number.
Figure 3.3. TRAP-isolated RNA can be used for downstream gene expression analysis.

(A) Representative reverse transcription and quantitative PCR (RT-qPCR) results of Gsta1 in quiescent and repopulating hepatocytes. No Ct value was detected with RNA isolated from wild type animals. (B) Alignment analysis of isolated RNA after high-throughput sequencing, demonstrating the importance of determining RNA integrity after isolation. High-quality RNA results in a higher percentage of mRNA reads (green), while low-quality RNA leads to a much higher percentage of ribosome reads (red), as most mRNA is degraded. RIN, RNA integrity number. (C) and (D) IGV tracks of RNA-sequencing reads of mRNA affinity-purified from quiescent and repopulating hepatocytes at the (C) Alb and (D) Afp loci. Note the 3' read bias is typical of a polyadenylated (poly(A)) selection pipeline.
DISCUSSION

TRAP-seq is a technique for cell type-specific isolation of translating mRNA via epitope-tagged ribosomes and presents an alternative to FACS approaches, as it circumvents limitations such as time requirements of FACS [9]. Instead, TRAP allows rapid and efficient isolation of RNA directly from bulk tissues, helping to avoid any alterations in gene expression. TRAP-seq is especially well-suited for use in the repopulating Fah<sup>−/−</sup> mouse liver, as hepatocyte expansion following removal of nitisinone is cell-autonomous and enables gene expression profiling of the subset of hepatocytes with integrated transgenes. The TRAP vector can also be coexpressed with gene-activating or -silencing molecules [14], including cDNA, short-hairpin RNA, and guide RNA, to study the effects on global gene expression of activation or inhibition of a specific gene. Alternatively, the Rosa<sup>LSL-GFP-L10A</sup> transgenic mouse provides the ability to profile gene expression in any cell with Cre recombinase activity. Since GFP:RPL10A can be specifically expressed in any cells that express Cre, the role of other cell types in the liver during liver injury and repopulation could be studied. For instance, crossing the CK19-Cre mouse with the TRAP transgenic mouse could be used to express GFP:RPL10A in cholangiocytes followed by TRAP-seq to study the change of gene expression in the biliary epithelium during the repopulation process.

To ensure accurate profiling of gene expression, it is critical to prepare all buffers and the affinity matrix prior to tissue dissection. All steps should be performed on ice with cold buffers unless otherwise specified to ensure polysome stabilization [10] and prevent RNA degradation. All buffers should be prepared with RNase-free reagents and the TRAP-seq protocol should be carried out in an RNase-free environment to prevent RNA degradation and low yield of immunoprecipitated RNA. The affinity matrix can be prepared up to 2 weeks prior to use with gentle resuspension on a tube rotator overnight. Special care should be taken to not vigorously shake the matrix to prevent disruption of the antibody-conjugated, protein L-coated magnetic beads. The methods to prepare the affinity matrix includes conjugation of magnetic beads to biotinylated protein L followed by incubation with anti-GFP antibodies. However, commercially available protein A/G magnetic beads can be substituted; if used, skip the initial conjugation step and proceed directly to antibody binding.
Furthermore, alternative epitope tags are presumably feasible with the above protocol with appropriate modification.

There are various points in which the RNA isolation and purification step can be paused (see protocol above). However, once liver samples have been harvested, continuing to the immunoprecipitation is recommended, as the yield of isolated RNA could drop by ~50% with freezing at this step [10]. Tissues should be quickly rinsed with dissection buffer that contains cycloheximide to inhibit mRNA translation. Insufficient tissue lysis could also contribute to low RNA yield. It is critical to homogenize tissues on ice until no tissue chunks are visible with the motor homogenizer while ensuring minimal aeration [10]. Additionally, sufficient washing with high-salt buffer is crucial to ensure removal of nonspecific binding of ribosomal proteins to the affinity matrix. Including a wild type mouse as a negative control helps to assess the specificity of the immunoprecipitation and the efficiency of the wash steps. Additionally, using a commercial RNA purification kit that includes RNase-free DNase treatment will increase RNA purity.

Moreover, it is recommended to verify the expression and abundance of the GFP:RPL10A fusion protein and assess the amount of tissue required to obtain ample RNA for downstream analysis. Tissue sections or lysates could be used for immuno-based detection methods to validate the expression of GFP:RPL10A. The amount of RNA isolated can vary by: (1) the number of cells expressing GFP:RPL10A, (2) the expression level of the transgene, and (3) the size and ploidy of the cells expressing the transgene. A pilot experiment using half and double the amount of the recommended amount of tissue could be useful in determining the optimal input lysate for TRAP-seq. In our hands, we could obtain ~150 ng of RNA with as little as 1-2% of hepatocytes expressing GFP:RPL10A from 200 mg of the repopulating Fah\(^{-}\) liver, representing ~2x10\(^5\) polyploid hepatocytes with transgene expression [9].

The TRAP-seq methodology isolates ribosome-bound mRNA to profile a cell’s translating mRNA pool. The resulting sequencing reads, therefore, correspond to the ‘translatome’ rather than the transcriptome. Note that translating ribosome footprints will not be collected, as TRAP is performed on native rather than cross-linked complexes. If footprinting analyses are desired, the
above protocol should be modified with relevant cross-linking followed by immunoprecipitation (CLIP) methodologies [15]. Another limitation of TRAP is the requirement of a sufficient amount of cells expressing the GFP:RPL10A fusion protein. For experiments in which the cell type of interest is small, combining multiple biological samples may be required to isolate sufficient RNA to enable RNA-seq [16]. Furthermore, TRAP-seq requires the presence of GFP:RPL10A in the cell type of interest. This could pose a challenge if there is no specific delivery system to the cells or if a cell-type specific promoter to drive Cre expression is not available.

The recent development of single-cell RNA-seq (scRNA-seq) technology has allowed direct sequencing followed by in silico identification of various cell types, enabling sequencing without sorting for specific cell types of interests [17–19]. However, scRNA-seq still requires dissociation of cells from the organ. In the case of the Fah−/− repopulation model, liver perfusion and hepatocyte isolation are extremely difficult and inefficient due to the fragility of both the injured and replicating hepatocytes. In fact, we have not yet been able to isolate sufficient hepatocytes from Fah−/− mice undergoing repopulation after hydrodynamic injection of FAH plasmids. Additionally, in the time it takes to process tissues, gene expression levels could change. Protocols for liver perfusion take up to 30 minutes of warm ischemia time. Future methodologies to optimize liver perfusion to decrease the processing time and increase isolation efficiency could allow scRNA-seq integration to the Fah−/− mouse model system and possibly to other injury and repopulation models. This would also support the study of all liver cell types.

In conclusion, the integration of TRAP-seq with the Fah−/− mouse enables specific isolation and gene expression profiling of regenerating hepatocytes to identify therapeutic targets that could promote liver repopulation. This method can be implemented to study other cell types in the liver and other organ systems for disease-specific identification of gene expression changes to identify potential drug targets or biomarkers. An analogous technique can be used to collect nuclei from repopulating hepatocytes using affinity purification, followed by epigenetic analysis of these specific cells [13].
REFERENCES

CHAPTER 4

THE DYNAMIC CHROMATIN ARCHITECTURE OF THE REGENERATING LIVER

Parts of this chapter were adapted with permission from The dynamic chromatin architecture of the regenerating liver. Wang AW, Yue YJ, Zahm AM, Morgan AR, Wangensteen KJ, Kaestner KH. Cellular and Molecular Gastroenterology and Hepatology. 2019. In press.
ABSTRACT

The adult liver is the main detoxification organ and is routinely exposed to environmental insults but retains the ability to restore its mass and function upon tissue damage. However, extensive injury can lead to liver failure, and chronic injury causes fibrosis, cirrhosis, and hepatocellular carcinoma (HCC). Currently, the transcriptional regulation of organ repair in the adult liver is incompletely understood. We isolated nuclei from quiescent as well as repopulating hepatocytes in a mouse model of hereditary tyrosinemia, which recapitulates the injury and repopulation seen in toxic liver injury in humans. We then performed the ‘assay for transposase accessible chromatin with high-throughput sequencing’ (ATAC-seq) specifically in repopulating hepatocytes to identify differentially accessible chromatin regions and nucleosome positioning. Additionally, we employed motif analysis to predict differential transcription factor occupancy and validated the in silico results with chromatin immunoprecipitation followed by sequencing (ChIP-seq) for hepatocyte nuclear factor 4α (HNF4α) and CCCTC-binding factor (CTCF). Chromatin accessibility in repopulating hepatocytes was increased in the regulatory regions of genes promoting proliferation and decreased in the regulatory regions of genes involved in metabolism. The epigenetic changes at promoters and liver enhancers correspond with the regulation of gene expression, with enhancers of many liver function genes displaying a less accessible state during the regenerative process. Our analysis of hepatocyte-specific epigenomic changes during liver repopulation identified CTCF and HNF4α as key regulators of hepatocyte proliferation and regulation of metabolic programs. Moreover, increased CTCF occupancy at promoters and decreased HNF4α binding at enhancers implicate these factors as key drivers of the transcriptomic changes in replicating hepatocytes that enable liver repopulation. Thus, liver repopulation in the setting of toxic injury makes use of both general transcription factors (CTCF) for promoter activation, and reduced binding by a hepatocyte-enriched factor (HNF4α) to temporarily limit enhancer activity.
INTRODUCTION

As the central metabolic organ in vertebrates, the liver regulates carbohydrate, protein, and lipid homeostasis, metabolizes nutrients, wastes, and xenobiotics, and synthesizes bile, amino acids, coagulation factors, and serum proteins [1]. To prevent acute liver failure upon exposure of harmful toxins, the liver has maintained an extraordinary ability to effectively restore its mass and function, in which the normally quiescent mature hepatocytes rapidly re-enter the cell cycle and divide [2]. Nonetheless, failure of regeneration can occur after exposure to harmful metabolites and environmental toxins, as often seen with the overconsumption of acetaminophen and alcohol [3]. Hence, understanding the genetic networks regulating the regenerative process can have an immense impact on the development of novel therapeutic strategies to treat acute liver failure.

The Fah null mouse model of human hereditary tyrosinemia type I provides a unique system to study the hepatocyte replication process after acute liver injury. Lack of the fumarylacetoacetate hydrolase (FAH) enzyme, essential for normal tyrosine catabolism, results in the accumulation of toxic intermediates followed by hepatocyte cell death [4,5]. Fah<sup>−/−</sup> mice can be maintained in a healthy state by supplementation with the drug 2-(2-nitro-4-trifluoromethylbenzoyl)-1,3-cyclohexanedione (NTBC) which inhibits an upstream enzymatic step that prevents toxin production [4]. Alternatively, gene therapy that utilizes hydrodynamic tail-vein injection and the Sleeping Beauty transposon system to restore Fah expression in hepatocytes can rescue these mice [6,7]. When a small fraction (0.1-1%) of hepatocytes express FAH following removal of NTBC, these hepatocytes competitively repopulate the liver in the context of injury through clonal expansion. Furthermore, this method allows lineage-tracing of repopulating hepatocytes since only those with stable FAH expression can expand and repopulate the injured parenchyma [7,8].

Eukaryotic DNA is highly organized and structured into compact chromatin to allow tight transcriptional control. Transcriptional regulation can be broadly categorized into two integrated layers: (1) transcription factors and the transcriptional machinery, and (2) chromatin structure and its regulatory proteins [9]. Expression of genes targeted by transcription factors depends on their binding affinity to specific target DNA recognition sequences, combinatorial assembly with other
cofactors, the concentration of the transcription factor, and post-translational modifications that affect protein localization [10]. The chromatin landscape is governed by DNA methylation, nucleosome properties, histone modifications, and intra- and interchromosomal interactions [10]. Establishing the relationship of chromatin structure, transcriptional regulators, and the effects on gene expression is therefore vital in elucidating the transcriptional control governing the regenerative process. To date, most studies have relied on transcriptomic studies to document gene expression changes in the regenerating liver [11–15] while two others focused on histone modifications [16,17]. However, these processes are downstream of chromatin reorganization and therefore do not capture the dynamic crosstalk of chromatin accessibility and transcriptional regulation. To identify transcriptomic changes specific to repopulating hepatocytes, we previously employed the translating ribosome affinity purification (TRAP) [18] to isolate translating mRNAs only from repopulating hepatocytes [15]. To discern the dynamic chromatin patterns that underlie liver repopulation, we now implement the ‘isolation of nuclei tagged in specific cell types’ (INTACT) [19] approach to isolate nuclei only from repopulating hepatocytes. This is achieved by expressing the GFP-tagged nuclear envelope protein SUN1-GFP together with FAH in Fah−/− mice, followed by the sorting of GFP-positive nuclei from repopulating hepatocytes and ATAC-seq [20]. We identify promoter accessibility changes corresponding to upregulation of cell cycle genes and downregulation of metabolic pathways, consistent with previous gene expression studies [12,15]. Integrative expression level and chromatin accessibility analysis suggests that gene activation is primarily associated with increased promoter accessibility, while inactivation is correlated with the closure of select promoters and enhancers. We propose a model in which a more accessible promoter allows increased transcription factor binding and gene activation, whereas decreased enhancer accessibility prevents binding of hepatocyte-enriched DNA binding proteins followed by inhibition of liver function genes so that the repopulating liver assumes a less differentiated state to promote cell growth and proliferation.
RESULTS

Adaptation of INTACT in the Fah−/− model allows for isolation of repopulating hepatocyte nuclei

Liver cells in humans and mice rarely undergo division in homeostatic conditions [2]. However, with injury and repopulation, hepatocytes become facultative stem cells and divide to replenish liver mass and restore liver function [2]. We hypothesized that this change from quiescence to replication is accompanied by substantial and specific changes to chromatin accessibility. To analyze the chromatin specific to repopulating hepatocytes, we adapted the INTACT [19] method to the Fah−/− model to label hepatocytes with the GFP-tagged nuclear envelope protein, SUN1-GFP, and performed fluorescence-activated cell sorting (FACS) to isolate nuclei from whole liver at selected time points (Figure 1). The SUN1-GFP fragment was subcloned into a FAH expression plasmid [7] so that all repopulating hepatocytes express GFP on the nuclear envelope. Following hydrodynamic injection of the FAH-SUN1-GFP plasmid into Fah−/− mice, NTBC was removed and liver repopulation was allowed to proceed for one or four weeks (Figure 1A). As a control for healthy, quiescent hepatocytes, RosaLSL-SUN1-GFP transgenic mice [19] were injected with AAV8-TBG-Cre [21] to label all hepatocytes. Nuclei were isolated from repopulating hepatocytes exclusively at the selected time points by FACS-sorting with an anti-GFP antibody (Figure 1B). ATAC-seq [20] was then performed on the sorted nuclei to profile the changes in the chromatin regulatory landscape that occur during liver repopulation.

Immunofluorescence labeling demonstrated expression of GFP-tagged nuclear envelopes in FAH-positive cells (Figure 1C), illustrating the specificity of using SUN1-GFP+ nuclei as a marker to identify repopulating hepatocytes. Interestingly, FAH and GFP signals were not homogeneous across all replicating cells, possibly due to the different copy numbers of plasmids taken in after hydrodynamic tail-vein injection of the SUN1-GFP construct [22]. In addition, since the Sleeping Beauty transposon system displays little insertion site preference [23], the loci in which the DNA fragments are integrated can affect expression levels of FAH and SUN1-GFP [24].
ATAC-seq detects differentially accessible chromatin regions

All ATAC-seq libraries were sequenced to ~100 million reads to ensure ample coverage across the genome followed by quality assessment to verify the robustness of the data (Table 4.1). We observed consistent ATAC-seq signals across various loci such as the *Alb* gene, which showed a progressive decrease in accessibility at the enhancer region during repopulation (Figure 4.1A). To identify differentially accessible chromatin regions, fragments below 150 bp, termed ‘nucleosome-free reads’, were used for peak calling. We identified 16,043 differentially accessible regions between quiescent and repopulating hepatocytes (Figure 4.1B, Supplementary Digital Table 4.1), of which 5,359 displayed increased accessibility in 1-week and 5,102 in 4-week repopulating hepatocytes, while 3,580 regions showed decreased accessibility in week 1 and 5,304 in week 4. Hierarchical clustering of the differentially accessible sites showed a clear separation of repopulating and quiescent hepatocytes (Figure 4.2C), corroborating previous transcriptome studies that 1-week and 4-week repopulating hepatocytes have a similar expression profile distinct from quiescent hepatocytes [15]. Replicates also clustered within the same condition, illustrating the reproducibility between biological replicates. Comparing accessibility regulated in the same direction in both time points (‘congruent’), 1,241 peaks were congruently increased and 2,033 congruently decreased (Figure 4.2B). Of note, only 28 regions exhibit accessibility changes in opposite directions in week 1 and week 4 (‘incongruent’), reflecting the similarity in the chromatin profile between the two repopulation time points.

Next, we focused on differentially accessible promoter elements. Differential ATAC-seq regions within 1 kb up- and downstream of the transcription start sites (TSS) were determined and KEGG pathway [25] analysis was performed (Figure 4.2D). As expected, pathways involved in cell growth and proliferation were enriched among the genes with increased accessibility in the promoter regions during repopulation, including MAPK signaling [26] and cancer pathways. Interestingly, purine and pyrimidine metabolism were only enriched in genes with increased promoter accessibility at week 1 but not at week 4, suggesting early activation of DNA synthesis immediately after liver injury in early stages of repopulation. This observation is consistent with
previous comparison of the $\text{Fah}^{-/-}$ and partial hepatectomy (PHx) models showing that the transcriptome of 1-week repopulating hepatocytes in the $\text{Fah}^{-/-}$ mouse is closest to that of 36 and 48 h post-PHx [15], at which the highest rate of DNA synthesis occurs in this model [27]. On the other hand, genes involved in hepatocyte functions such as complement and coagulation and metabolic pathways displayed significantly decreased promoter accessibility at both regeneration time points. Our pathway enrichment analysis substantiates prior studies of gene expression profiles and extends the findings to chromatin accessibility in that proliferation pathways are activated while liver functions are inhibited during repopulation [12,15].

**Integration of chromatin accessibility and gene activity infers regulatory mechanisms**

To evaluate the association of chromatin landscape and gene expression, we utilized our prior TRAP-seq study [15] as a dataset of transcriptomic changes in repopulating hepatocytes. Genes with ATAC-seq signals and TRAP-seq reads that changed in the same direction at the same time point were identified as 'concordant genes' (Figure 4.3A, Supplementary Digital Table 4.2). We observed significant overlap of the concordant genes with ATAC-seq and TRAP-seq ($p<1\times10^{-16}$ for all 1-week concordant genes and 4-week concordantly activated genes. $p=0.03$ for 4-week concordantly inhibited genes), while there was no significant overlap of genes with increased expression in 1 week and decreased chromatin accessibility at 4 weeks ($p=0.39$). KEGG pathway [25] analysis suggested enrichment of cell growth and replication in the week 1 concordantly activated genes, and overrepresentation of biosynthesis and metabolism in both week 1 and week 4 concordantly inhibited genes (Figures 3B, C). In addition, pathway enrichment supported previous observations that activation of the glutathione metabolic network is essential for reactive oxygen species removal after PHx or recovery following toxic liver injury [15,28,29]. We conclude that changes to the chromatin structure underlie the upregulation of genes involved in cell proliferation and downregulation of genes associated with metabolic processes.

Next, we sought to investigate co-regulatory networks in repopulating hepatocytes. All ATAC-seq peaks identified were first separated into increased, decreased, or unchanged
accessibility, with a cutoff of absolute fold change ≥1.5 and false discovery rate (FDR) ≤0.05, followed by subdivision into regulatory regions of promoters, liver-specific enhancers, or cerebellum-specific enhancers as a negative control [30]. Promoter peaks were annotated to the nearest genes and the corresponding transcript levels at the same time point were extracted from TRAP-seq data [15]. We then compared the gene expression levels in the differentially accessible promoters to those in the unchanged promoters (Figures 4.3D, E). The normalized log₂ fold change was positive (p=7.47E-03 in week 1 and 3.81E-02 in week 4) with increased and negative (p=1.06E-06 in week 1 and 1.38E-03 in week 4) with decreased promoter accessibility at both time points, demonstrating a significant association of promoter openness and transcriptional activity. Differentially accessible liver enhancer peaks were similarly categorized, putative enhancer-regulated genes extrapolated [30], corresponding target gene expression extracted [15], and the transcript level changes compared to those of genes with unchanged enhancer accessibility. Interestingly, decreased liver enhancer accessibility was highly correlated with decreased gene activity (p=1.89E-20 in week 1 and 1.19E-07 in week 4), while no significant expression changes (p=0.22 in week 1 and 0.88 in week 4) were associated with increased enhancer openness. While the exact mechanism explaining this lack of correlation requires further evaluation, we posit that target genes regulated by enhancers in the quiescent liver are already highly expressed in mature, differentiated hepatocytes [12,15]. An increase in liver enhancer accessibility hence does not further elevate the expression of these genes significantly. Another likely explanation for the lack of significant association between increased liver enhancer accessibility and activation of target genes could be the recruitment of repressors instead of activators to the regulatory elements to decrease expression [31–33]. Finally, refinement of the computationally predicted enhancer-promoter pairs with experimental approaches could result in a more accurate correlation of enhancer accessibility and transcriptional activity. Importantly, cerebellum enhancers exhibited no significant correlation with the changes in transcript levels and chromatin accessibility in the repopulating liver, as expected (Figures 4.3D, E, right). Our integrated ATAC-seq and TRAP-seq analysis reveal that gene activation is regulated by increased promoter accessibility, presumably
allowing recruitment of transcriptional activators and RNA polymerase II to the TSS, whereas gene inhibition may be governed by both decreased promoter and enhancer openness, preventing long-range enhancer-promoter interactions [34].

Differential chromatin accessibility predicts transcription factors involved in liver repopulation

Dynamic coordination of chromatin structure and transcription factors is required to fine-tune gene expression. Chromatin organization influences access of the transcriptional apparatus by regulating binding sequence accessibility [35] and transcription factor binding stability [36]; conversely, transcription factors affect access of remodelers to the chromatin [35] and histones [37]. To identify DNA binding transcription factors that connect differential chromatin accessibility and gene expression, we carried out de novo motif profiling at differentially accessible promoters and liver enhancers [30].

We found enrichment of the ETS transcription factor ELK1 motif in promoters with increased accessibility in both 1-week (FDR=1E-76) and 4-week (FDR=1E-41) repopulating hepatocytes (Figure 4.4A, B, Supplemental Digital Table 4.3). ELK1 binds to the serum response element upon MAPK phosphorylation [38] to activate immediate early genes such as Fos and components of the basal transcriptional machinery [39]. Furthermore, ELK1 supports cell cycle entry during liver regeneration as Elk1-/- mice show reduced hepatocyte proliferation after PHx [40]. We postulate that promoters became more accessible after acute liver injury to permit increased ELK1 occupancy, enabling hepatocyte repopulation.

Among the regions with increased accessibility during liver repopulation, surprisingly, the CTCF motif was highly enriched (FDR=1E-78 in week 1 and 1E-49 in week 4) (Figures 4.4C, D). CTCF plays numerous roles in transcriptional regulation to function as a transcriptional activator [41] or repressor [42], insulator to block enhancer-promoter interactions [43], chromatin structure organizer to form topologically-associated domains [44] modulator of long-range chromatin looping [45], and even mediator of local RNA polymerase II pausing to regulate alternative exon usage [46].
CTCF is recruited to the *Axin1* promoter as a transcriptional repressor by the ‘long noncoding RNA associated with liver regeneration’ (lncRNA-LALR1) after PHx, leading to activation of Wnt/β-catenin signaling to promote hepatocyte proliferation [47]. However, the function of CTCF in liver regeneration is not fully understood.

In addition, we found the HNF4α binding motif to be significantly associated with liver enhancers with decreased accessibility during liver regeneration (FDR=1E-146 in week 1 and 1E-186 in week 4) (Figures 4.4E, F). HNF4α is a master regulator atop the transcriptional cascade of hepatocyte differentiation [48,49] and a crucial factor that maintains hepatocytes in the differentiated state [50]. Importantly, HNF4α suppresses liver proliferation, as mice with conditional deletion of *Hnf4a* demonstrate increased hepatocyte BrdU incorporation and Ki67 expression [51]. HNF4α also directly inhibits cell growth and replication pathways, as illustrated by the upregulation of cell cycle and proliferation genes upon acute HNF4α loss [51,52]. Moreover, motifs of other liver-enriched transcription factors were also overrepresented at enhancers that became less accessible in repopulating hepatocytes, including hepatocyte nuclear factor 1β (HNF1β) and hepatocyte nuclear factor 6 (HNF6) [53] (Figures 4.4E, F). We examined the locations for CTCF and HNF4α motifs within regions of dynamic chromatin accessibility and found that they are present in the center of these regions with CTCF at those with increased (p=2.70E-04 in week 1 and 1.97E-13 in week 4), and HNF4α at those with decreased accessibility (p=0.59 in week 1 and 2.48E-03 in week 4) (Figure 4.4G, H).

In summary, *de novo* motif analysis of differentially accessible ATAC-seq regions suggests increased occupancy of ELK1 and CTCF at chromatin regions that become more accessible, and decreased binding of liver-enriched transcription factors at liver enhancers that become less accessible during repopulation.

**HNF4α occupancy is decreased in liver-specific enhancers during repopulation**

We postulated that decreased HNF4α binding allows repopulating hepatocytes to assume a less differentiated and pro-proliferative state and carried out ChIP-seq on quiescent and 4-week
repopulating livers to examine genome-wide HNF4α occupancy during the repopulation process. We observed 508 peaks with decreased and only 14 peaks with increased occupancy in repopulating livers (Figure 4.5A, Supplemental Digital Table 4.4). Remarkably, 42% (214) of lost HNF4α occupancy occurred within previously-defined liver enhancers [30], while 23% (119) fell into distal intergenic regions, and 10% (52) were within 1 kb up- and downstream of the TSS ('promoter') (Figure 4.5B). These data corroborate the differentially accessible chromatin analysis of transcription factor motifs that had identified enrichment of the HNF4α consensus sequence at enhancers with decreased accessibility in repopulating hepatocytes (Figure 4.4H).

Next, we integrated ATAC-seq, ChIP-seq, and TRAP-seq datasets [15], and identified hepatocyte-enriched genes crucial for establishing liver functions including complement and coagulation (Ctb, F2), biosynthesis (Itih1, Acsl1, Pgrmc1), and metabolism (Ugt1a5, Mthfs, Rdh10) [54] as correlated with decreased HNF4α enhancer occupancy during regeneration (Figure 4.5C, E). To explore the mechanism responsible for decreased HNF4α occupancy during liver repopulation, we next turned to the TRAP-seq dataset [15] to inspect Hnf4a expression levels in quiescent and replicating hepatocytes. Remarkably, we found a 50% reduction of Hnf4a transcripts in 4-week repopulating hepatocytes (FDR=4.16E-3) compared to the quiescent liver (Figure 4.5D, Table 4.2). Taken together, these results implicate decreased chromatin accessibility and reduced Hnf4a expression as contributors to the suppression of hepatocyte-specific genes and downregulation of liver biosynthetic functions during repopulation.

**CTCF promoter occupancy is increased in the repopulating liver**

In order to extend the computational finding of enriched CTCF motif at promoters with increased accessibility, we performed ChIP-seq in quiescent and 4-week repopulating livers. CTCF occupancy was increased at 1,382 sites in the repopulating liver, while only 2 peaks showed decreased binding (Figure 4.6A, Supplemental Digital Table 4.5). To characterize the role of increased CTCF occupancy during liver repopulation, we first evaluated its potential insulator function by calculating an ‘insulator strength score’ [55] at all gained binding sites. Genomic regions
with increased CTCF occupancy with divergent flanking promoters within 50 kb were identified and the normalized expression levels corresponding to the genes were extracted from our TRAP-seq data (Figure 4.6B) [15]. Surprisingly, gene pairs with increased CTCF binding were not significantly more enriched for differential gene expression than random gene pairs (p=0.9) (Figure 4.6C), suggesting that CTCF is unlikely to act as an insulator during liver repopulation.

Remarkably, the vast majority (1,026, 74%) of the gained CTCF peaks fell within 1 kb up- and downstream of the TSS (‘promoter’) (Figure 4.6D). To examine the targets of increased CTCF occupancy, all differentially bound peaks were annotated to the nearest genes and their corresponding expression changes were obtained from our TRAP-seq dataset [15,25,56]. We found 545 (39%) peaks associated with chromatin modification, transcription regulation, and cancer (Figure 4.6E), while 656 (47%) sites with increased CTCF binding were associated with inhibition of genes in cell death regulation, stress response, and morphogenesis. Together, our network analysis suggests a diverse role for CTCF in transcriptional regulation in which increased CTCF occupancy supports hepatocyte replication and prevents cell death during liver repopulation, possibly by enabling binding of both activating and repressing cofactors.

CTCF is known to exhibit divergent roles in activating and repressing transcription by recruiting various protein partners in a context-dependent manner [57]. To identify these cofactors, we performed motif analysis for the regions differentially bound by CTCF (Figure 4.6F). As expected, the CTCF motif was highly enriched (FDR=1E-26) at all differential binding sites, confirming the specificity of the anti-CTCF antibody for immunoprecipitation. At sites where CTCF binding corresponded to gene activation, we observed significant enrichment for the ‘zinc finger and BTB domain-containing protein 3’ (ZBTB3) (FDR=1E-10) and nuclear transcription factor Y (NF-Y) (FDR=1E-10) binding motifs (Figure 4.6F). ZBTB3 is considered a likely factor binding 5’ of CTCF due to its frequent enrichment ~10 bp upstream of CTCF motifs in the human genome [58]. Furthermore, expression of ZBTB3 is induced by the accumulation of reactive oxygen species to promote cancer cell growth and prevent apoptosis via the activation of antioxidant gene expression in cell lines [59]. Whether CTCF directly interacts with or indirectly recruits ZBTB3 is yet unclear,
but the proteins are likely to interact based on their close proximity at promoters. NF-Y binds to the CCAAT box present at ~30% of the promoters [60] and is required for cell cycle progression, DNA synthesis, and proliferation in mouse embryonic fibroblasts [61]. Additionally, reconstituted in vitro transcription reactions demonstrated that binding of NF-Y disrupts nucleosome structure at promoters containing the NF-Y recognition sequence [62]. Recruitment of NF-Y could hence induce local nucleosome repositioning to allow increased accessibility of the transcriptional apparatus to activate gene expression.

On the other hand, the Yin Yang 1 (YY1) binding motif was enriched (FDR=1E-13) at sites where increased CTCF occupancy corresponded with decreased gene expression (Figure 4.6F). YY1 regulates embryogenesis, cell differentiation, and tumorigenesis [63,64], as well as enhancer-promoter interactions analogous to long-range chromatin looping mediated by CTCF [65]. YY1 functions as a transcriptional repressor via recruitment of the polycomb repressor complex, resulting in trimethylation of histone H3 lysine 27 [66,67]. It is also a cofactor of CTCF in regulating X chromosome inactivation, although the exact mechanism remains unclear [68]. Given these observations, it is likely that direct or indirect co-binding of CTCF and YY1 at promoters induces transcriptional repression or disrupts enhancer to promoter interactions to downregulate target genes.

When examining gene expression, we found the levels of ZBTB3 and YY1 not significantly changed in repopulating hepatocytes (Table 4.2). Three NF-Y proteins exhibited varying changes in transcript levels, with unchanged NF-YA, downregulated NF-YB in 1-week, and downregulated NF-YC in 4-week repopulating hepatocytes, albeit all with modest changes of less than 2-fold. These observations do not rule out the possibility of post-translational modifications that might alter the abundance or localization of transcription factors.

To analyze if transcription factors colocalize to CTCF-occupied promoters with differential gene expression during liver regeneration, we performed ZBTB3 and YY1 ChIP-qPCR on quiescent and 4-week repopulating livers. We observed a significant increase of ZBTB3 occupancy at Ctnna2 (p=0.023) and Smad3 (p=0.025) promoters, two genes with increased promoter accessibility,
elevated CTCF binding, and upregulated expression during liver regeneration (Figure 4.6G, H). Regarding YY1 occupancy, there was a significant increase at the \textit{Bcl2l11} (p=0.029) promoter, a gene with increased promoter accessibility, enhanced CTCF occupancy, and decreased transcript levels (Figure 4.6I). With the limited loci tested, we conclude that ZBTB3 is recruited to open chromatin regions occupied by CTCF to activate gene expression during liver regeneration. On the other hand, increased YY1 binding to select promoters with elevated CTCF binding could regulate transcriptional repression in repopulating hepatocytes. These results suggest that increased chromatin accessibility correlates with enhanced CTCF occupancy that recruits coactivators or corepressors to fine-tune target gene expression to induce replication and prevent apoptosis during liver repopulation (Figure 4.6J). Future experiments that utilize co-immunoprecipitation and high-throughput sequencing technologies to analyze interactions between CTCF and cofactors as well as genome-wide binding patterns of the coregulators will aid in the understanding of mechanisms underlying CTCF modulation.

Liver regeneration is accompanied by nucleosome remodeling

Most eukaryotic DNA is packaged around histone protein octamers into nucleosomes to regulate chromatin organization and transcriptional control. Nucleosome properties such as positioning and turn-over rates can affect the binding of transcription factors and access of the transcriptional machinery [69]. The nucleosome landscape adjacent to the TSS is of particular interest, as nucleosomes adopt a specific phasing pattern immediately up- and downstream [70]. Hence, nucleosome organization could act as an additional layer of transcriptional regulation in repopulating hepatocytes.

We inferred nucleosome positioning from nucleosome-containing sequences by extracting ATAC-seq reads longer than 150 bp (Figure 4.7A). Nucleosomes surrounding the TSS were defined as ‘-1 nucleosomes’ within 350 bp upstream and ‘+1 nucleosomes’ within 250 bp downstream, and the distance between the +1 to -1 nucleosomes was defined as the ‘nucleosome-free region’. When compared to quiescent hepatocytes, there was a median downstream shift of 9
bp in 1-week (p=2.60E-13) and an upstream shift of 19 bp in 4-week (p<1E-15) repopulating hepatocytes for the -1 nucleosomes, while there was no significant shift in +1 nucleosome positioning (Figure 4.7B, Supplemental Digital Table 4.6). As a result, there was a global increase of promoter openness in 4-week repopulating hepatocytes as the distance between +1 to -1 nucleosomes increased, while the nucleosome-free region was shorter in 1-week regenerating liver compared to the quiescent state. The difference in genome-wide promoter openness in repopulating hepatocytes at various time points suggests that accessibility of divergent functional regions could be differentially regulated during liver regeneration. Indeed, the nucleosome-free region constitutes only 17.5% of regions with increased accessibility in week 1 but 45.6% in week 4 (Figure 4.7C), whereas 39.0% of week 1 and only 26.9% of week 4 regions that became more open fall into distal intergenic regions (Figure 4.7D). On the other hand, chromatin regions with decreased accessibility show a similar distribution between the nucleosome-free region and distal intergenic regions. These observations indicate that the increase of chromatin accessibility occurs mainly at distal genomic areas in 1-week and around the TSS in 4-week repopulating hepatocytes.

To evaluate the association of TSS accessibility and gene expression, we extracted the top 500 up- and downregulated genes in repopulation [15] and calculated the change in the length of the nucleosome-free region between quiescent and regenerating hepatocytes as a surrogate for differential TSS accessibility. We only observed a significant increase (p=1.15E-2) of +1 to -1 nucleosome distance in genes activated in week 4 when compared to quiescent hepatocytes, while no significant change in the nucleosome-free region was present in genes upregulated in week 1 or genes downregulated in week 1 and week 4 (Figure 4.7E, F). It is likely that eviction or repositioning of the -1 nucleosomes could expose transcription factor binding sequences and allow access of the transcriptional machinery to the TATA box for gene activation in regenerating hepatocytes [71]. Altogether, analysis of the nucleosome structure implies nucleosome reorganization could affect gene activation but not inhibition during liver repopulation.
# TABLES

**Table 4.1.** ATAC-seq library sequencing summary.

<table>
<thead>
<tr>
<th>Sample ID</th>
<th>Condition</th>
<th>Index</th>
<th>Cumulate reads</th>
</tr>
</thead>
<tbody>
<tr>
<td>SUN1-GFP-1</td>
<td>Quiescent</td>
<td>CGAGGCTG</td>
<td>119,120,180</td>
</tr>
<tr>
<td>SUN1-GFP-2</td>
<td>Quiescent</td>
<td>AAGAGGCA</td>
<td>111,970,248</td>
</tr>
<tr>
<td>#3603</td>
<td>1-week repopulation</td>
<td>AATTCGTT</td>
<td>97,320,484</td>
</tr>
<tr>
<td>#3604</td>
<td>1-week repopulation</td>
<td>GCCGTCGA</td>
<td>135,005,202</td>
</tr>
<tr>
<td>#2383</td>
<td>4-week repopulation</td>
<td>GTAGAGGA</td>
<td>186,365,116</td>
</tr>
<tr>
<td>#2385</td>
<td>4-week repopulation</td>
<td>TGCTGGGT</td>
<td>236,418,952</td>
</tr>
</tbody>
</table>
Table 4.2. Gene expression of enriched transcription factor motifs.

<table>
<thead>
<tr>
<th>Gene</th>
<th>Transcript</th>
<th>W1 log₂ fold change</th>
<th>W1 FDR</th>
<th>W4 log₂ fold change</th>
<th>W4 FDR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Elk1</td>
<td>NM_007922</td>
<td>0.23</td>
<td>0.83</td>
<td>0.05</td>
<td>0.96</td>
</tr>
<tr>
<td>Ctcf</td>
<td>NM_181322</td>
<td>-0.36</td>
<td>0.47</td>
<td>0.23</td>
<td>0.66</td>
</tr>
<tr>
<td>Hnf4a</td>
<td>NM_008261</td>
<td>-0.17</td>
<td>0.74</td>
<td>-1.12</td>
<td>0.00</td>
</tr>
<tr>
<td>Hnf1b</td>
<td>NM_009330</td>
<td>-0.15</td>
<td>0.85</td>
<td>-0.22</td>
<td>0.67</td>
</tr>
<tr>
<td>Hnf6</td>
<td>NM_008262</td>
<td>0.54</td>
<td>0.49</td>
<td>0.11</td>
<td>0.92</td>
</tr>
<tr>
<td>Zbtb3</td>
<td>NM_001098237</td>
<td>1.26</td>
<td>0.33</td>
<td>-0.64</td>
<td>0.73</td>
</tr>
<tr>
<td>Nfya</td>
<td>NM_001110832</td>
<td>0.23</td>
<td>0.85</td>
<td>0.03</td>
<td>1.00</td>
</tr>
<tr>
<td>Nfyb</td>
<td>NM_010914</td>
<td>-0.74</td>
<td>0.02</td>
<td>-0.20</td>
<td>0.64</td>
</tr>
<tr>
<td>Nfyc</td>
<td>NM_001048168</td>
<td>-0.67</td>
<td>0.10</td>
<td>-0.82</td>
<td>0.02</td>
</tr>
<tr>
<td>Yy1</td>
<td>NM_009537</td>
<td>-0.35</td>
<td>0.42</td>
<td>-0.12</td>
<td>0.79</td>
</tr>
</tbody>
</table>
Table 4.3. Primer sequences used in this study.

<table>
<thead>
<tr>
<th>Primer name</th>
<th>Sequence</th>
<th>Use</th>
</tr>
</thead>
<tbody>
<tr>
<td>MfeI-Sun1-F</td>
<td>GACTCAATTGGCGCCGCACTACTGGCC</td>
<td>Plasmid construction</td>
</tr>
<tr>
<td>BsiW1-Sun1-R</td>
<td>GCTACGTACGTTAACCAGCTACTATTAAGATC CTCTCGGATATTAACTTCGCC</td>
<td>Plasmid construction</td>
</tr>
<tr>
<td>ZBTB3-ChIP-Ctnna2-qPCR-F1</td>
<td>TTTGTTGATACAGTGCGG</td>
<td>ZBTB3 ChIP-qPCR</td>
</tr>
<tr>
<td>ZBTB3-ChIP-Ctnna2-qPCR-R1</td>
<td>TGGGAGCAACAGTGATGAA</td>
<td>ZBTB3 ChIP-qPCR</td>
</tr>
<tr>
<td>ZBTB3-ChIP-Smad3-qPCR-F2</td>
<td>AGACCTCGTGCCTTTTCTA</td>
<td>ZBTB3 ChIP-qPCR</td>
</tr>
<tr>
<td>ZBTB3-ChIP-Smad3-qPCR-R2</td>
<td>GGCCTGGTGGATTTCACAGAG</td>
<td>ZBTB3 ChIP-qPCR</td>
</tr>
<tr>
<td>YY1-ChIP-Bcl2I11-qPCR-F1</td>
<td>CTCTTGTAGCGATCCCTCCTT</td>
<td>YY1 ChIP-qPCR</td>
</tr>
<tr>
<td>YY1-ChIP-Bcl2I11-qPCR-R1</td>
<td>CTGCCGTCCCAATCAATGTT</td>
<td>YY1 ChIP-qPCR</td>
</tr>
<tr>
<td>40S-F2</td>
<td>AGCGAGCTGTGCTGAAGTTT</td>
<td>ChIP-qPCR control</td>
</tr>
<tr>
<td>40S-R2</td>
<td>AGGCTGCTGGATCTGTTTA</td>
<td>ChIP-qPCR control</td>
</tr>
</tbody>
</table>
FIGURES

Figure 4.1. Implementation of the ‘isolation of nuclei tagged in specific cell types’ (INTACT) [19] method with the Fah<sup>f</sup> mouse model allows isolation of repopulating hepatocyte nuclei.

(A) Schematic of coexpression of the GFP-tagged nuclear envelope protein SUN1 (SUN1-GFP) with FAH to label repopulating hepatocytes for fluorescence-activated cell sorting (FACS) followed by the ‘assay for transposase accessible chromatin with high-throughput sequencing’ (ATAC-seq).

(B) Representative images (n=2) of repopulating hepatocyte nuclei show specific isolation with anti-GFP antibody labeling. Gray boxes denote the sorting strategy to collect GFP<sup>+</sup> nuclei. (C) Representative images (n=2) of immunofluorescent staining of GFP and FAH show coexpression of SUN1-GFP and FAH in repopulating hepatocytes of the Fah<sup>f</sup> mouse after 1 week (left) and 4 weeks (middle), and global expression of SUN1-GFP and FAH in all hepatocytes of the Rosa<sup>LSL</sup>-SUN1-GFP<sup>+</sup> mouse 1 week after AAV8-TBG-Cre injection. FACS: fluorescence-activated cell sorting.
**Figure 4.2.** Chromatin accessibility changes during liver repopulation are related to cell growth activation and metabolic inhibition.

(A) ATAC-seq shows reproducible signals across biological replicates and a decrease of peak intensity in the proximal regulatory region [105] of the *Alb* locus. (B) 16,043 significantly differential accessible regions were identified in repopulating and quiescent hepatocytes (absolute fold change ≥1.5 and FDR ≤0.05). Comparison of differential accessible regions identified at different time
points during repopulation shows 3,273 that changed in the same direction ('congruent' peaks), of which 1,241 were congruently increased (red dots) and 2,033 congruently decreased (blue dots). (C) Hierarchical clustering of all differentially accessible regions shows that biological replicates have similar chromatin landscape. (D) KEGG pathway analysis of differentially accessible promoters with increased (left) and decreased (right) accessibility in repopulating hepatocytes. FDR: false discovery rate.
Figure 4.3. Association of expression levels and chromatin accessibility implicates divergent regulatory mechanisms for gene activation and inhibition.

(A) Differential gene expression data were obtained from a previous study that implemented translating ribosome affinity purification followed by RNA-sequencing (TRAP-seq) [15]. 'Upset' plot
demonstrates overlap of ATAC-seq regions and TRAP-seq genes that are significantly changed in the same direction at the same time points (‘concordant’ genes) in repopulating hepatocytes. Fisher’s exact test was performed to calculate the significance of overlapping targets. The horizontal black lines in the green bars of the top panel indicate the number of overlaps expected by chance. (B and C) KEGG pathway analysis of concordantly activated and repressed genes in (B) 1-week and (C) 4-week repopulating hepatocytes. Dashed lines denote FDR=0.05. (D and E) Association of changes in chromatin accessibility and gene expression in (D) 1-week and (E) 4-week repopulating hepatocytes indicates that promoter accessibility changes are related to both gene activation and inhibition, while only decreased liver enhancer accessibility is significantly correlated with decreased expression of putative target genes [30]. Cerebellum enhancers and their putative targets do not display any significant relationship to chromatin accessibility and gene expression changes in the liver. One-sample t-tests were carried out to identify the differences in normalized log$_2$ fold change in differentially accessible and unchanged chromatin regions. Vertical lines denote the 95% confidence interval of normalized log$_2$ fold change in peaks with increased and decreased accessibility.
**Figure 4.4.** Enrichment analysis identifies transcription factor motifs overrepresented at differential accessible promoters and enhancers [30].

(A and B) The ELK1 motif is enriched in promoter regions that became more open in repopulating hepatocytes. (C and D) The CTCF motif is overrepresented in liver enhancers with increased accessibility in both (C) 1-week and (D) 4-week repopulating hepatocytes, respectively. (E and F) Motifs of liver-enriched transcription factors HNF4α, HNF1β, and HNF6 are enriched in enhancers with decreased accessibility during (E) 1-week and (F) 4-week liver repopulation. (G and H) Motif frequency of the differential accessible peaks for (G) CTCF and (H) HNF4α display enrichment of the transcription factor motifs at the enhancer peak center in repopulating hepatocytes. Numbers presented in (A-F) denote FDR.
Figure 4.5. HNF4α binding is decreased in the repopulating liver.

(A) 508 genomic regions display decreased and only 14 display increased HNF4α occupancy in the regenerating liver (B) 40% of peaks with decreased HNF4α binding overlap with liver-enriched enhancers ('liver enhancer') [30], and 25% fall in distal intergenic regions that contain ubiquitous enhancers ('distal intergenic'). (C) Integrative analysis of chromatin accessibility (ATAC-seq), HNF4α binding (ChIP-seq), and gene expression (TRAP-seq) [15] changes suggests the suppression of liver functions including complement, biosynthesis, and metabolic pathways during liver regeneration is associated with reduced HNF4α occupancy. (D) HNF4α expression is downregulated in repopulating hepatocytes (n=4 for quiescent and n=6 for repopulating
hepatocytes) [15]. (E) Representative tracks (n=2 for ATAC-seq and ChIP-seq, n=4 for TRAP-seq) of chromatin accessibility, HNF4α occupancy, and transcript levels at *Itih1*, the locus with the strongest decrease of HNF4α occupancy. RPKM: reads per kilobase of transcript, per million mapped.
Figure 4.6. CTCF binding is increased at promoters in the repopulating liver. (A) 1,306 peaks show increased, while only 2 peaks show decreased CTCF occupancy during repopulation. (B) Schematic to test the insulator function of increased CTCF binding to differentially regulate expression of the flanking genes [55]. (C) Promoters flanking sites of increased CTCF
occupancy are not more enriched for differentially expressed genes compared to random gene pairs in the genome. A Fisher's exact test was used to examine the differentially expressed gene ratios from the two groups of gene pairs. (D) 75% of the genomic regions with increased CTCF binding are within 1 kb up- and downstream of the TSS ('promoter'), and only 13 peaks overlap with liver enhancers [30]. (E) Enriched pathways of increased chromatin accessibility, CTCF occupancy, and increased (red) or decreased (blue) gene expression during liver repopulation. (F) Motif enrichment analysis identifies an overrepresentation of CTCF motif in differentially-bound regions, the ‘zinc finger and BTB domain-containing protein 3’ (ZBTB3) and nuclear transcription factor Y (NF-Y) motifs at sites with increased CTCF occupancy associated with gene activation, and the Yin Yang 1 (YY1) motif at sites with increased CTCF occupancy associated with gene inhibition. Numbers denote FDR. (G and H) ZBTB3 occupancy is increased in the repopulating liver at the (G) Smad3 and (H) Ctnna2 promoters, two genes with increased CTCF occupancy and expression during regeneration. (H) YY1 occupancy is increased in the repopulating liver at the Bcl2l11 promoter, a gene with elevated CTCF binding and decreased expression during regeneration. (I) Representative tracks (n=2 for ATAC-seq and ChIP-seq, n=4 for TRAP-seq) of chromatin accessibility, CTCF occupancy, and transcript levels at Hells, the locus with the strongest increase in CTCF binding.
Figure 4.7. Decreased nucleosome density is associated with increased gene expression [15] in repopulating hepatocytes.

(A) Schematic for identifying nucleosome positioning information with NucleoATAC [104]. (B) Globally, -1 nucleosomes have an upstream shift away from the TSS in 4-week repopulating hepatocytes, while +1 nucleosomes positioning is constant during liver repopulation. (C and D) Distribution of regions with differential accessibility in (C) the nucleosome-free region that is within 350 bp upstream and 250 bp downstream of the TSS and (D) distal intergenic regions in 1- and 4-week repopulating hepatocytes. (E) The top 500 upregulated genes exhibit an increased +1 to -1 nucleosome distance in 4-week but not 1-week repopulating hepatocytes when compared to quiescent hepatocytes. (F) The top 500 downregulated genes are not significantly associated with changes in +1 to -1 nucleosome distance in repopulating compared to quiescent hepatocytes.
Permutation tests with 10,000 iterations were used to compare the nucleosome distance in repopulating and quiescent hepatocytes. NFR: nucleosome-free reads. NR: nucleosomal reads.
Figure 4.8. Model of transcriptional regulation in repopulating hepatocytes.

(A) Access to enhancers allows liver-enriched transcription factors to maintain quiescent hepatocytes in the differentiated state (top). In contrast, chromatin-dense enhancers and promoters prevent transcription factor binding to inhibit gene expression of cell cycle genes (bottom). (B) During liver repopulation, decreased accessibility of liver enhancers [30] in conjunction with more closed promoters prevents binding of transcription factor and assembly of the transcriptional machinery at hepatocyte-specific liver function genes, resulting in a less differentiated transcriptomic and epigenomic profile in the repopulating cells (top). Conversely, the promoter regions of cell cycle genes become more open, with increased +1 to -1 distance and increased CTCF occupancy at the promoter, allowing elevated expression of genes involved in the cell cycle and DNA synthesis pathways (bottom).
DISCUSSION

Gene regulation is tightly controlled by a complex network integrating transcription factor binding and transcriptional apparatus assembly, chromatin structure, epigenetic modifications, and even intra- and interchromosomal interactions [9,10]. In this study, we investigated the association of chromatin accessibility, nucleosome properties, transcription factor occupancy, and gene expression [15] to delineate the multidimensional framework of transcriptional regulation in the repopulating liver. By implementing the INTACT method [19] to express SUN1-GFP in the Fah⁻/⁻ model, we successfully performed cell type-specific isolation of only repopulating hepatocyte nuclei followed by ATAC-seq to identify changes of the chromatin landscape (Figures 4.1, 4.2). Integration of TRAP-seq [15] with ATAC-seq determined that gene activation corresponds with increased promoter openness, while gene inhibition is linked to a decreased promoter and enhancer accessibility (Figure 4.3C). We also corroborated previous findings that cell cycle, DNA synthesis, proliferation, and glutathione metabolism are activated whereas complement and coagulation, biosynthesis, and metabolic pathways are inhibited during liver repopulation (Figures 4.2D and 4.3B, C) [12,15]. In addition, de novo motif analysis identified enrichment of CTCF and HNF4α binding sequences in regions with increased and decreased accessibility in repopulating hepatocytes, respectively (Figure 4.4). We further validated the differential occupancy of both factors in the repopulating liver with ChIP-seq and observed decreased HNF4α binding at liver enhancers [30] (Figure 4.5) and increased CTCF binding at promoters (Figure 4.6). Integrated ATAC-seq, ChIP-seq, and TRAP-seq analysis suggests that CTCF recruits cofactors to activate genes involved in chromatin organization and replication and inhibit genes in the regulation of cell death (Figure 4.6E-J). On the other hand, loss of HNF4α occupancy at liver enhancers decreases the expression of hepatocyte-enriched genes crucial in establishing liver homeostasis and function (Figure 4.5C-E).

In general, 40% of CTCF binding sites occur in intergenic regions distant to the TSS, while 35% of CTCF sites are found in promoters [30,44]. Interestingly, the vast majority (75%) of sites with increased CTCF occupancy are located within promoters in the repopulating liver (Figure
4.6D). In fact, CTCF can function as a direct transcriptional repressor at the Myc promoter [72] and as an activator of the amyloid precursor protein promoter [73], strengthening the notion that CTCF plays a more localized role as a transcriptional regulator in the repopulating liver via recruitment of cofactors. Upregulation of CTCF in liver cancer is associated with poor survival, likely through the activation of forkhead box M1 (FOXM1) to stimulate cell growth and tumor metastasis [74]. The CTCF-FOXM1 axis could be triggered during liver regeneration to promote hepatocyte proliferation [75]. Increased CTCF activity at the Myc promoter [76] or decreased CTCF repression at the Myc enhancer [77] have both been observed in cancer cells that lead to increased MYC expression. The high tumor mutational burden of CTCF results in abnormal occupancy [78,79], and thus the cofactors and targets of CTCF could be different in the regenerating liver and liver cancer. The multitude of CTCF functions warrants further investigation to understand its contribution to mediating chromatin structure and organization in the context of liver repopulation. Specifically, CTCF also acts as an insulator to block enhancer-promoter interactions [43], a factor that promotes long-range chromatin looping [45], and a TAD boundary protein that defines expression domains for tight transcriptional control [44]. Future experiments to detect changes in chromatin interactions via chromosome conformation capture [80] would be valuable in determining whether differential CTCF occupancy affects three-dimensional chromatin organization during liver repopulation.

The mechanisms of increased CTCF and decreased HNF4α binding in the repopulating liver are also not fully understood. In the current study, we infer that a more open chromatin state at specific promoters correlates with the accessibility of CTCF to its binding sites; however, we have not assessed causality. Previous work found that enrichment of thymidine (T) at the 18th position in the CTCF motif reduces its affinity, where low-affinity sites are more sensitive to loss of CTCF binding during mouse embryonic stem cell differentiation [55]. Additionally, it is likely that changes in DNA methylation influence differential CTCF occupancy, as methylated CpGs in the CTCF recognition site can prevent its binding [81,82]. Demethylation at specific promoter regions could, therefore, increase CTCF occupancy during liver repopulation. In the case of reduced HNF4α occupancy at liver-specific enhancers in the regenerating liver, part of this effect can be
explained by reduced expression of HNF4α itself. Furthermore, HNF4α could be regulated post-
transcriptionally via phosphorylation by kinases such as protein kinase A and C, as well as AMP-
activated protein kinase to decrease its DNA binding activity or nuclear localization [83]. Activation
of the MAPK signaling pathway is also shown to inhibit Hnf4a expression via activation of the
transcription factor JUN [83,84]. The fact that enrichment of DNA synthesis pathways is only
observed in 1-week repopulating livers and that Hnf4a transcript level is unchanged in week 1 but
reduced in week 4 hepatocytes strengthens the notion that activation of cell growth and proliferation
occur early after the initiation of liver repopulation, followed by a later reduction of Hnf4a
transcription. Future studies using, for instance, targeted degradation of CTCF [85] or HNF4α could
be implemented to identify potential promoters and inhibitors of liver repopulation. Technologies
such as cDNA [8] or clustered regularly interspaced short palindromic repeats (CRISPR) [86,87]
screens could also be utilized to evaluate the effectors downstream of CTCF activation and HNF4α
inhibition.

In summary, we propose the following model to explain the transcriptional adaptations that
accompany liver repopulation (Figure 4.8): during hepatocyte replication, the promoters of selected
genes become more open due to an increased distance between histones at +1 to -1, increasing
accessibility for CTCF, transcription factor recruitment, and transcriptional machinery assembly to
activate genes that regulate cell cycle, DNA synthesis, and proliferation pathways. On the other
hand, decreased enhancer accessibility in conjunction with suppression of Hnf4a expression evicts
or prevents HNF4α binding, and possibly that of other hepatocyte nuclear factors, to liver
enhancers, resulting in repression of hepatocyte metabolic and biosynthetic function genes.
MATERIALS AND METHODS

All primer sequences are listed in Table 4.3.

Plasmid construction

The generation of the pKT2/Fah-Sun1-Gfp//SB plasmid was described previously [15]. The nuclear envelope SUN1-tagged GFP (SUN1-GFP) plasmid was a generous gift from Dr. Jeremy Nathans (Johns Hopkins University, Baltimore, MD, USA). We amplified the SUN1-GFP insert by PCR amplification with the primers MfeI-Sun1-F and BsiW1-Sun1-R and subcloned it into the vector pKT2/Fah-mCa//SB [7] to construct pKT2/Fah-Sun1-Gfp//SB. This construct utilizes the Sleeping Beauty (SB) transposase for stable transgene integration into the genome. The plasmid was prepared with the GenElute HP Plasmid Maxiprep Kit (NA0310-1KT, MilliporeSigma) for endotoxin-free maxi-scale DNA extraction and purification.

Mouse studies

Fah"-" mice were maintained on 7.5 mg/l 2-(2-nitro-4-trifluoromethylbenzoyl)-1,3-cyclohexanedione (NTBC) (Swedish Orphan Biovitrum) in the drinking water. Hydrodynamic tail-vein injection [86] of 10 μg of pKT2/Fah-Sun1-Gfp//SB was performed followed by NTBC withdrawal for 1 week (n=2) or 4 weeks (n=2) to induce liver repopulation [15]. The RosaLSL-Sun1-GFP mice [19,88] were kindly provided by Dr. Jeremy Nathans (Johns Hopkins University, Baltimore, MD, USA) and were tail-vein injected with AAV8.TBG.PI.Cre.rBG (Penn Vector Core [89]) at 1 x 10^{11} virus particles per mouse to ablate the loxP-stop-loxP cassette only in hepatocytes. Livers from these mice were harvested 1 week after viral injection and served as quiescent controls. All studies were performed in 8 to 12-week-old mice.

Immunofluorescence staining

Liver lobes were isolated, fixed in 4% paraformaldehyde overnight at 4 °C, embedded in paraffin, and sectioned. Tissue sections were deparaffinized with xylene and rehydrated with serial incubation of 100%, 95%, 80%, and 75% ethanol followed by PBS. Antigen retrieval was carried
out in Tris/EDTA buffer (10mM Tris, 1mM EDTA, pH 9.2) in a pressure cooker (2100 Antigen Retriever, Aptum Biologics Ltd.) and cooled to room temperature. Slides were then blocked with blocking buffer (PBS, 1% BSA) for 1 h followed by overnight incubation of antibodies in the blocking buffer at 4 °C in a humidified chamber. Three washes of PBS were carried out the next day followed by incubation with secondary antibodies at room temperature for 2 h. Goat anti-GFP antibody (ab6673, 1:300, Abcam) and rabbit anti-FAH antibody (ab81087, 1:600, Abcam) were used to label repopulating hepatocytes from Fah−/− mice after one and four weeks of repopulation and all hepatocytes from Rosa^LSL-GFP-L10a^ mice injected with AAV8-TBG-Cre. DAPI (B1098, 1:10,000, BioVision) was used to label nuclei.

**Hepatocyte nuclei isolation**

Liver was homogenized in 10 ml hypotonic buffer (10 mM Tris-HCl, pH 7.5, 2 mM MgCl₂, 3 mM CaCl₂) on ice. The homogenate was filtered with a 100 μm filter and sedimented at 400 g for 10 min at 4 °C. 10 ml of hypotonic buffer with 10% glycerol was used to resuspend the pellet followed by dropwise addition of 10 ml cell lysis buffer (hypotonic buffer, 10% glycerol, 1% IGEPAL CA-630). The homogenate was incubated for 5 min on ice and sedimented at 600 g for 5 min at 4 °C. Nuclei were washed with lysis buffer again and quantified in a hemocytometer. Isolated nuclei were labeled with an Alexa Fluor 647 anti-GFP antibody (338006, clone FM264G, 1:25, BioLegend, San Diego, CA) for 30 min and 2 μg/ml DAPI immediately prior to sorting. After gating for the DAPI-positive signal, nuclei double-positive for GFP and AF647 were sorted with a BD FACSria II, and only tetraploid hepatocyte nuclei were collected for further experiments.

**ATAC-seq library generation**

ATAC-seq libraries were generated as previously described [20]. Briefly, transposition was performed on 25,000 sorted tetraploid nuclei at 37 °C for 30 min followed by DNA purification with the MinElute Reaction Cleanup Kit (28206, QIAGEN). DNA fragments were PCR preamplified for 5 cycles initially, and 1/10 of the volume (5 μl) was removed for qPCR amplification for 20 cycles.
A ‘R vs Cycle Number’ plot was generated and the number of cycles required to reach ⅓ of the maximum R determined for each sample. The preamplified ATAC-seq libraries were then amplified for the calculated additional cycles. Agencourt AMPure XP beads (A63881, Beckman Coulter) were used for size selection to generate the final libraries [90]. Library quality was assessed with an Agilent High Sensitivity DNA Bioanalyzer (5067-4626, Agilent Technologies), and quantity measured with KAPA Library Quantification Kits (KK4835, KAPA Biosystems).

**ATAC-seq peak calling**

ATAC-seq libraries were paired-end sequenced on an Illumina HiSeq 4000 (Illumina, San Diego, CA, USA) with 50, 75, or 100 reads. Reads were then trimmed to 50 bp with Cutadapt [91] and peaks called with the ATAC-Seq/DNase-Seq pipeline [92]. Briefly, the trimmed fastq files were aligned to the mouse genome (mm10) with Bowtie2 [93] followed by removal of PCR duplicates and mitochondrial reads. Bam files of the same biological sample from various technical replicates were then merged with Samtools [94] and duplicated reads removed. The filtered reads were shifted 5 bp for + strands and 4 bp for - strands to adjust for the transposase binding sites [20]. Nucleosome-free reads were identified with the R package ATACseqQC using a random forest classifier [95] followed by peak calling with MACS2 [96]. Artifact signals were then removed according to the mm10 empirical blacklist regions [97]. The irreproducible discovery rate (IDR) framework was used to compare all pairs of biological replicates to identify reproducible peaks that passed a threshold of 10% for all pairwise analyses. The conservative peak set for each sample was identified by selecting the longest peak list from all pairs that passed the 10% IDR cutoff.

**ATAC-seq peak quality assessment**

To ensure the ATAC-seq peaks generated from the sorted nuclei are of high quality, The R package ATACseqQC [95] was employed for assessment. We first visualized the insert size distribution to confirm the presence of distinct periodicity of ~175 bp associated with nucleosome patterning in all samples, indicating the DNA fragments are protected by integer multiples of
nucleosomes [20]. The signal intensity of nucleosome-free reads and nucleosomal reads was also averaged across all TSS to examine evidence that no over-fragmentation was introduced during hepatocyte nuclei isolation, sorting, or ATAC-seq library preparation.

**ATAC-seq differential peak analysis**

The R package ATACseqQC [95] was used to split the aligned bam files into nucleosome-free reads and nucleosomal reads. The R package DiffBind [98] was used to identify differential accessible peaks from the nucleosome-free reads. The overlapping regions from the ATAC-seq peak sets for each sample were identified and merged into non-overlapping regions. Read counts for each region were quantified with dba.count (score=DBA_SCORE_TMM_READS_FULL, fragmentSize=0, bScaleControl=F, filter=0, bRemoveDuplicates=F, bUseSummarizeOverlaps=T). Peaks identified in both biological replicates in the same conditions were used for differential analysis with dba.analyze (method=DBA_EDGER, bSubControl=F, bTagwise=T) in conjunction with edgeR [99]. Peaks with an absolute fold change ≥1.5 and FDR ≤0.05 were identified as significant differentially accessible regions.

**Integrative analysis of TRAP-seq and ATAC-seq data**

To identify chromatin accessibility and gene expression that changed in the same direction at the same time point ('concordant genes'), the differentially accessible peaks were first annotated to the nearest TSS with the R package ChIPseeker [100]. Genes with differential expression during liver repopulation were obtained from a previous study that utilized translating-ribosome affinity purification followed by RNA-sequencing (TRAP-seq) [15]. The concordant ATAC-seq peaks and TRAP-seq genes were identified and the expected overlap and significance was calculated with a hypergeometric test. To evaluate the association of chromatin accessibility and gene expression changes, all chromatin regions were stratified into regions with increased, decreased, or unchanged accessibility, with the cutoff of an absolute fold change ≥1.5 and FDR ≤0.05. For promoter accessibility and gene activity association analysis, regions within 1 kb up- and
downstream of the TSS were identified and annotated to the nearest genes with the R package ChIPseeker [100]. The corresponding expression change at the same time point was extracted from TRAP-seq [15] and normalized by subtracting the mean log₂ fold change of the unchanged from the increased and decreased chromatin accessibility groups. The normalized expression fold change of the nearest genes in the differentially accessible promoters was compared to that in the unchanged accessibility promoters with a one-sample t-test. For enhancer accessibility and gene expression association studies, liver- and cerebellum-specific enhancers and their putative targets were obtained from a previous study [30]. Briefly, regions with the presence of H3K4me1 but the absence of H3K4me3 ChIP-seq peaks were identified as putative enhancers and refined with a chromatin-signature based enhancer predictor. Enhancer-promoter units were identified by calculating the correlation of H3K4me1 and RNA polymerase II ChIP-seq peak strength along each chromosome. All possible promoter and enhancer pairs with a >0.23 Spearman correlation coefficient were identified as linked enhancer-promoter units. Gene expression fold changes were normalized as described above, and the normalized gene expression fold-change of the enhancer target genes in the differentially accessible enhancers was compared to that in the unchanged accessibility enhancers with a one-sample t-test.

**Transcription factor motif enrichment analysis**

ATAC-seq peaks are separated into promoter and liver enhancer [30] regions and Homer [101] is used to identify enrichment of *de novo* motifs with the function findMotifsGenome.pl (mm10 -size given). Motifs with a p-value of lower than 1E-12 are considered significant to reduce the number of false positives. FDR is also calculated with each significant motif. To ensure the identified motifs are enriched in ATAC-seq peaks with different accessibility, motif frequency surrounding 500 up- and downstream of the peak center from all identified IDR peaks in quiescent hepatocytes and differentially accessible regions in repopulating cells is extracted. The difference in motif frequency distribution of regenerating and quiescent samples was then calculated with a Kolmogorov-Smirnov test.
**ChIP-seq library generation**

100 mg of quiescent (n=2) and repopulating (n=2) liver tissue was finely chopped with a razor blade and cross-linked in 1% formaldehyde for 10 min followed by addition of 2.5 M glycine and incubation for 5 min at room temperature. Tissues were sedimented, washed with cold PBS, and Dounce-homogenized in cold ChIP cell lysis buffer (10 mM Tris-HCl pH 8.0, 10 mM NaCl, 3 mM MgCl2, 0.5% IGEPAL CA-630, protease inhibitor) on ice. After incubation at 4 °C for 5 min, nuclei were pelleted and resuspended in nuclear lysis buffer (50 mM Tris-HCl pH 8.1, 1% SDS, 5 mM EDTA, protease inhibitor). Nuclei were sonicated with a Bioruptor (Diagenode) for 2 rounds of 7.5 min each. 10 μg of sheared DNA was incubated with anti-CTCF (2 μg, 07-729, Millipore) or anti-HNF4α (2 μg, ab181604, Abcam) antibodies in dilution buffer (16.7 mM Tris-HCl pH 8.1, 167 mM NaCl, 0.01% SDS, 1.1% Triton-X 100, protease inhibitor) at 4 °C overnight. Protein-A agarose beads were also washed with cold dilution buffer three times and incubated with blocking buffer (10 mg/ml BSA, ChIP dilution buffer, protease inhibitor) at 4 °C overnight. Sheared DNA incubated with antibody and blocked protein-A agarose were incubated at 4 °C for 1 h the next day and washed at room temperature with buffers TSEI (20 mM Tris-HCl pH 8.1, 150 mM NaCl, 2 mM EDTA, 0.1% SDS, 1% Triton X-100), TSE II (20 mM Tris-HCl pH 8.1, 500 mM NaCl, 2 mM EDTA, 0.1% SDS, 1% Triton X-100), ChIP buffer III (10 mM Tris-HCl pH 8.1, 0.25M LiCl, 1 mM EDTA, 1% NP-40, 1% deoxycholate), and TE (10 mM Tris-HCl pH 8.1, 1 mM EDTA). Chromatin was eluted with elution buffer (1% SDS, 0.1 M NaHCO₃) twice and incubated with 0.2 M NaCl at 65 °C overnight to reverse the cross-links. Digestion was carried out with 10 mg/ml proteinase K in 40 mM Tris-HCl pH 7.5 and 10 mM EDTA to purify CTCF- or HNF4α-bound and input DNA. ChIP-seq libraries were prepared with the NEBNext Ultra II DNA Library Prep Kit for Illumina (E7645S, New England BioLabs) and Agencourt AMPure XP beads were used for size selection to generate the final libraries. Library quality was assessed with an Agilent High Sensitivity DNA Bioanalyzer (5067-4626, Agilent Technologies), and quantity measured with KAPA Library Quantification Kits (KK4835, KAPA Biosystems).
ChIP-seq data analysis

ChIP-seq libraries were sequenced on an Illumina HiSeq 4000 (Illumina) with 100 single-end reads and aligned to the mm10 genome with STAR [102]. Bam files from various technical replicates of the same biological sample were merged with Samtools [94]. Peak calling was performed with Homer [101] and differential occupancy analysis was carried out with the R package DiffBind [98]. Read counts for each peak were quantified with dba.count (score=DBA_SCORE_TMM_MINUS_FULL, bUseSummarizeOverlaps=TRUE) and differential analysis were identified with dba.analyze (method=DBA_EDGER, bSubControl=T, bTagwise=F) in conjunction with edgeR [99].

ChIP-qPCR

ChIP was performed with 5 μg of anti-ZBTB3 (ab106536, Abcam) and 2 μg of YY1 (ab109237, Abcam) antibodies with 10 μg of sheared DNA from quiescent and 4-week repopulating livers as described above. Input and immunoprecipitated DNA were purified with phenol-chloroform extraction followed by qPCR with primer sets ZBTB3-ChIP-Ctnna2-qPCR-F1 and -R1, ZBTB3-ChIP-Smad3-qPCR-F1 and -R1, YY1-ChIP-Bcl2l11-qPCR-F1 and -R1, YY1-ChIP-Igf2r-qPCR-F1 and -R1, and 40S-F2 and -R2.

CTCF differential expression insulator analysis

Increased CTCF occupancy during liver repopulation could prevent distal regulatory regions to activate only one of the flanking promoters surrounding a CTCF binding site, and therefore leading to a larger difference in gene expression levels. We define this ‘differential expression insulator’ function, in which a gene pair is either highly or lowly expressed without the presence of CTCF, but only one flanking gene exhibits a decrease in gene expression after binding of CTCF. An insulator strength score was calculated for all significantly gained (fold change ≥1.5, FDR ≤0.05) CTCF peaks in the repopulating liver as previously described [55]. Briefly, CTCF sites with divergent flanking promoters within 50 kb were identified and the corresponding gene
expression levels from quiescent and 4-week repopulating hepatocytes were extracted from published TRAP-seq [15].

Low-expressors, in which RPKM-normalized read counts are 0 across all samples, were filtered followed by calculation of a rank percentile based on RPKM for each gene. Let \( x_Q \) and \( y_Q \) be the expression percentile in the quiescent hepatocytes; \( x_R \) and \( y_R \) be the expression percentile in the 4-week repopulating hepatocytes. The insulator strength score is calculated by taking the maximum value of \( x_Q \times y_Q \times x_R \times (1 - y_R) \) and \( x_Q \times y_Q \times (1 - x_R) \times y_R \). A differential expression insulator function will have one of the following effects: (1) Increased \( x_R \) and decreased \( y_R \); in this case, \( x_Q \times y_Q \times x_R \times (1 - y_R) \) will be the largest. (2) Decreased \( x_R \) and increased \( y_R \); in this case, \( x_Q \times y_Q \times (1 - x_R) \times y_R \) will be the largest. Gained CTCF sites with the top 25% insulator strength scores were categorized as strong insulators. Random gene pairs not flanked by CTCF within 50 kb were used as controls and a differential expression insulator score for each gene pair was calculated as described above. The number of significant (FDR≤0.05) and non-significant (FDR>0.05) differential expression of the flanking genes were identified for all strong insulators from increased CTCF binding and random genomic regions. Finally, we used Fisher’s exact test to examine the likelihood of gained CTCF sites to contain more significantly changed genes when compared to that of control regions.

**Nucleosome location analysis with ATAC-seq**

MAC2 (callpeak –keep-dup all, -B --SPMR, -q 0.05, --broad) [96] was used to identify broad peaks from all aligned bam files including nucleosome-free reads and nucleosome-containing reads from ATAC-seq. Broad peaks were then processed with BEDtools [103] to extend the peaks (bedtools slop -b 200), sorted by genomic positions (sort -k1,1 -k2,2n), and overlapping reads were merged (bedtools merge). Nucleosome position was identified with NucleoATAC [104] from the aligned bam and broad peak files. The closest nucleosomes with respect to TSS were identified, and those within 350 bp upstream and 250 bp downstream of the TSS were identified as the -1 and +1 nucleosomes, respectively.
Nucleosome positioning analysis

The distance of +1 to -1 nucleosomes was calculated for each transcript. We used the Kolmogorov-Smirnov test to compare the +1 and -1 nucleosome distribution differences between quiescent and repopulating hepatocytes, respectively. To analyze the association between gene activity and nucleosome positioning, transcriptomic changes in repopulating hepatocytes [15] were first stratified into three categories: top 500 upregulated (fold change ≥1.5, FDR ≤0.05), top 500 downregulated (fold change ≥1.5, FDR ≤0.05), and unchanged (absolute fold change <1.5 or FDR >0.05) genes. The distances between the +1 to -1 nucleosomes were calculated for each gene and differential positioning was carried out by comparing the distance in quiescent to regenerating hepatocytes in the upregulated, downregulated, and unchanged gene expression groups, respectively, with a permutation test (n=10,000).

Statistical analysis

EdgeR [99] was used for all high-throughput sequencing data analysis. For the integrative TRAP-seq and ATAC-seq analysis, a hypergeometric test was used for identifying the significance of overlapping gene sets, and a one-sample t-test was used to compare the difference between normalized gene expression fold change in differentially accessible promoter and enhancer regions, respectively. A Kolmogorov-Smirnov test was performed for global distribution change of +1 and -1 nucleosome positioning and a permutation test (n=10,000) was carried out to test the change in +1 to -1 nucleosome distance of genes with differential expression.

Study approval

The animal experiments carried out in this study were reviewed and approved by the IACUC of the Penn Office of Animal Welfare at the University of Pennsylvania.
REFERENCES


190


SUMMARY

This thesis work is the first to implement novel cell type-specific labeling technologies to mark repopulating hepatocytes in vivo to achieve unbiased profiling of transcriptomic and epigenomic alterations that occur during the regenerative process. With the utilization of innovative methodologies to exclusively track regenerating hepatocytes combined with extensive integrative multiomic analyses, I identified several factors and pathways with important biological implications followed by examining their functional significance in the regulation of liver regeneration. Additionally, the gene expression and chromatin accessibility datasets provide comprehensive information on the transcriptional regulation of repopulating hepatocytes.

By adopting translating ribosome affinity purification with high-throughput RNA-sequencing (TRAP-seq) for the isolation of mRNA from repopulating hepatocytes, I identified Slc7a11, encoding the cystine/glutamate antiporter, xCT, as a promoter of liver regeneration in the setting of acute liver injury. Slc7a11 is upregulated for over 600- and 250-fold in repopulating hepatocytes after 1 week and 4 weeks of liver injury and regeneration. Nonetheless, activation of Slc7a11 via ectopic expression at the time of injury still allowed hepatocytes with increased xCT expression to repopulate the injured liver more efficiently.

The implementation of hepatocyte nuclear isolation followed by the ‘assay for transposase-accessible chromatin using sequencing’ (ATAC-seq) allowed identification of the alterations in the chromatin landscape and investigation of the epigenomic regulation that occurs during liver repopulation. Multiomic data integration has enabled the detection of increased promoter accessibility that corresponds to enhanced CCCTC-binding factor (CTCF) occupancy followed by activation of proliferative genes. On the other hand, decreased liver-specific enhancer accessibility correlates with decreased hepatocyte nuclear factor 4α (HNF4α) binding and inhibition of liver function genes. These observations provide new insights into how mature hepatocytes assume a less differentiated state to enable cell growth and replication during acute injury followed by repopulation.
LIMITATIONS

Considerations of the quiescent and regeneration mouse models

In the studies aimed at understanding the genome-wide changes that occur during liver regeneration, I utilized two different transgenic mouse lines, i.e. \textit{Rosa}^{LSL-GFP-L10a} and \textit{Rosa}^{LSL-SUN1-GFP}, as sources for quiescent hepatocytes, whereas \textit{Fah}^{-/-} mice were used to induce liver injury and isolate repopulating liver cells. While no phenotypic deficiencies have been observed between the \textit{Rosa}^{LSL-GFP-L10a} [1] and \textit{Rosa}^{LSL-SUN1-GFP} [2] mice, studies have yet to demonstrate that hepatocytes isolated from these transgenic lines after injection of AAV8-TBG-Cre exhibit similar transcriptomic and epigenomic profiles compared to hepatocytes isolated from wild type mice. A small portion of differentially expressed genes and divergent accessible regions identified in the current investigations could result from the expression of GFP-tagged proteins.

In addition, the \textit{Fah}^{-/-} repopulation mouse could be perceived as an artificial model. In patients with hereditary tyrosinemia type I (HTI), all hepatocytes are exposed to toxic metabolites and subject to an injurious environment during tyrosine catabolism. However, cells that receive the \textit{Fah} transgene in the \textit{Fah}^{-/-} model are technically not injured. Moreover, mechanisms of injury specific to tyrosine metabolism could limit the interpretation and findings to expand upon other injury-induced liver repopulation conditions, further restricting the utility of potential therapeutic targets identified in the current work. Analysis of injured hepatocytes could elucidate the transcriptomic and epigenomic discrepancies between the liver cells in HTI patients and repopulating hepatocytes in \textit{Fah}^{-/-} mice. This comparison may also answer whether similar alterations in redox pathways are present in injured cells, similar to that observed in repopulating cells. Finally, investigation of injured hepatocytes could add to the knowledge on signaling from injured to repopulating hepatocytes to better understand the induction of liver regeneration.

TRAP-seq profiles the ‘translatome’

It is worth noting that TRAP isolates translating mRNA bound to the ribosomal protein L10a. Therefore, the sequencing reads represent the ‘translatome’ rather than the ‘transcriptome’ of
repopulating hepatocytes. Future methods to enable the isolation of nascent mRNAs particularly from repopulating hepatocytes is likely to generate divergent datasets and will allow the comparison of transcriptome and translatome, as well as the calculation of translation efficiency [3] during liver regeneration.

**Isolation of repopulating hepatocyte nuclei is time-consuming**

While affinity-purification was successful in isolating mRNA from repopulating hepatocytes, no enrichment was detected with immunoprecipitation of repopulating hepatocyte nuclei expressing the SUN1-GFP fusion protein. This was attempted several times with various anti-GFP antibodies according to methods described previously [2,4]. The exact mechanism of failure to immunoprecipitate SUN1-GFP-labeled nuclei is unclear but we postulate that a combination of the fragility of repopulating hepatocyte nuclei and the small amount of repopulating cells hinder the affinity-purification of SUN1-GFP expressing hepatocytes. To address the technical difficulty of lack of enrichment of target nuclei with the ‘isolation of nuclei-tagged in specific cell types’ (INTACT) method [2], we turned to fluorescence-activated cell sorting (FACS) to isolate repopulating hepatocyte nuclei. However, FACS is a time-consuming process and depending on the percentage of regenerating cells, took up to over 4 hours. The lengthy process could introduce cellular stress and cause chromatin fragmentation. In my hands, samples that required sorting for over 4 hours exhibited low-quality ATAC-seq reads that significantly reduced the signal to noise ratio, hindering peak calling in downstream analysis to identify open chromatin regions.

**Limitations of the ATAC-seq technology and bioinformatics analysis pipelines**

ATAC-seq makes use of the Tn5 transposase that accesses the relatively ‘open’ chromatin to fragment and tag accessible chromatin regions, a process referred to as ‘tagmentation’ [5]. However, the Tn5 transposase displays sequence-specific binding preferences that induce bias during tagmentation [6]. Computational methods to model and correct for the transposition bias
have been proposed [7,8], but these have not been widely-adopted or experimentally validated to date.

Furthermore, ATAC-seq is limited to only examining euchromatic areas and thus generates mainly short fragments under 200 bp constituted of nucleosome-free or mono-nucleosomal reads. This limits our ability to investigate the regulation of heterochromatic regions during liver repopulation. Other methods that utilize sonication-resistant heterochromatin followed by a gradient separation to discriminate subtypes of histone 3 lysine 9 trimethylation (H3K9me3) and histone 3 lysine 27 trimethylation (H3K27me3) [9] could be implemented to elucidate transcriptional silencing and domain repression during liver regeneration.

The development of ATAC-seq to examine open chromatin regions is a relatively new technology [5] and hence no pipelines have been proposed as the gold standard for data analysis. The analysis utilized in this thesis includes a combination of the ATAC-seq pipeline for ENCODE data developed by Anshul Kundaje and the ENCODE Data Analysis Center [10], as well as in-house scripts developed specifically for the analysis of the present study. The use of alternative bioinformatics programs could, therefore, generate distinct results. Nevertheless, I presume that the highly significant regions with differential accessibility should remain the same or at least similar across various analysis platforms.

Finally, chromatin regions with differential accessibility identified by ATAC-seq only refer to the state of openness but do not infer the activity of the gene. A chromatin region could become more accessible to allow the binding of transcriptional repressors, leading to suppression of its target genes. Hence, ATAC-seq only provides a broad overview of the modifications in the chromatin landscape rather than specific directional changes of gene activity. Integration of additional genome-wide experiments such as ChIP-seq, promoter-enhancer interaction mapping by chromatin capture methods, or other functional manipulation is necessary to guide the understanding of the effects of altered chromatin accessibility. For instance, the integration of TRAP-seq to inform transcriptomic modifications in this thesis provides additional information to assess the consequences of chromatin accessibility changes.
Whole repopulating livers are used for chromatin immunoprecipitation (ChIP)

In both studies to profile the transcriptomic and epigenomic changes that occur during liver regeneration, ChIP experiments were carried out on whole quiescent and repopulating livers to elucidate the mechanism of Slc7a11 activation as well as the occupancy of CTCF and HNF4α. Due to the large cell number required for typical ChIP assays (1-10 million) [11], the lack of sufficient repopulating hepatocytes isolated from the regenerating liver has prevented cell type-specific ChIP-seq. Thus, it is possible that the signals of increased activating transcription factor 4 (ATF4) binding at the Slc7a11 promoter are detected from hepatocytes undergoing injury and repopulation, as well as other cell types in the liver. The lack of significantly increased occupancy of nuclear factor erythroid 2-related factor 2 (NRF2) at the Slc7a11 locus could also be the result of chromatin dilution by injured hepatocytes in the regenerating liver. Similarly, CTCF and HNF4α ChIP-seq experiments were performed in whole-livers that underwent 4 weeks of repopulation in a non-cell type-specific manner. The changes in CTCF and HNF4α occupancy detected are therefore likely a mixture of signals from injured and regenerating hepatocytes. Changes in binding patterns specific to repopulating cells could also be diluted by other cell types and the surrounding dying hepatocytes.

Development of novel methods to utilize micrococcal nuclease-based native ChIP without cross-linking — including ‘Occupied Regions of Genomes from Affinity-purified Naturally Isolated Chromatin (ORGANIC)’ [12], ‘Ultra-Low-Input micrococcal nuclease-based Native ChIP (ULI-NChIP)’ [13], and ‘Cleavage Under Targets & Release Using Nuclease (CUNT&RUN)’ [14] — allows ChIP-seq from as few as 1,000 cells, depending on the abundance of the transcription factor of interest. Further optimization of these methods for hepatocyte nuclei will be useful in obtaining cell type-specific cistromes in repopulating hepatocytes.
FUTURE DIRECTIONS

The downstream effectors of Slc7a11 in regenerating hepatocytes

Slc7a11 overexpression promotes liver regeneration after acute injury, however, the effects of xCT activation in chronic injury has not been studied. Upregulation of Slc7a11 is observed in human gastrointestinal tumors [15], breast cancer cells [16], and hepatocellular carcinoma [17] to increase glutathione (GSH) synthesis for the defense of reactive oxygen species (ROS) and promote cell growth [15,16]. It is plausible that xCT activation in the setting of chronic liver injury plays a similar role in reducing oxidative stress to confer an advantage for hepatocyte survival or replication.

While Slc7a11 is not an oncogene, the safety of long-term xCT activation should be rigorously examined, especially in the inflammatory microenvironment often observed in chronic liver injury that could ultimately lead to tumorigenesis [18]. On the other hand, transient xCT induction could be considered as a treatment for acute and chronic liver injury. In APAP-induced liver failure, N-acetylcysteine is often used to restore intracellular GSH levels and prevent hepatic necrosis [19]. The efficacy of short-term upregulation of xCT via viral delivery of Slc7a11 or drug treatment can be assessed in the settings of liver injury to determine the extent of prevention of ROS-mediated cell death and promotion of hepatocyte survival.

To identify the mechanisms of Slc7a11 activation to enhance liver regeneration following acute injury, coexpression of Fah and Slc7a11 cDNA in conjunction with the GFP-L10a fusion protein in the Fah⁻/⁻ mouse would allow for specific isolation and expression profiling of repopulating hepatocytes with Slc7a11 induction. TRAP-seq of these cells could determine the effects of Slc7a11 overexpression on gene expression. Similarly, implementation of the clustered regularly interspaced short palindromic repeats (CRISPR) system to mutate Slc7a11 with the expression of FAH and GFP-L10a could establish downstream targets necessary to promote liver regeneration that are dependent on Slc7a11 activation. In particular, I hypothesize that genes involved in GSH metabolism, including Gsta and Gstm isoforms, as well as redox-sensitive transcription factors, such as NRF2 and AP1, will be activated following the upregulation of Slc7a11 and inhibited after
Slc7a11 mutation to support the induction of redox pathways [20]. Gain- and loss-of-function experiments of the top effectors of Slc7a11 activation could be carried out to further assess their functional significance during liver regeneration.

**Regulation of oxidative response during liver regeneration**

Expression profiling of regenerating hepatocytes identified massive induction of multiple redox pathway genes. Genes involved in the oxidation/reduction network have been implicated in liver injury and regeneration [21–23], and depletion of GSH availability or inhibition of GSH synthesis delays regeneration and exacerbates toxic hepatic injury [24,25]. What remains to be shown is the spatiotemporal regulation of redox balance in replicating hepatocytes during liver regeneration.

Recently, a method was established to determine the redox status of zebrafish cells *in situ* [26]. The effects of a biliary toxin were measured and determined to show that the toxin induces a more oxidized state in extrahepatic biliary cells [26]. The assay utilizes the redox-sensitive GFP biosensor, termed roGFP, that contains an engineered dithiol/disulfide switch sensitive to cytosolic redox states [27]. Different redox levels alter the state of cysteine amino acid residues, resulting in a shift in emission at two excitation wavelengths (405 and 488 nm). The coupling of roGFP to glutaredoxin (GRX1), an endogenous enzyme that catalyzes the GSH/GSSG equilibrium, allows for the specific determination of the cytoplasmic GSH redox potential by roGFP [27]. A high 405/488 signal indicates an oxidized sensor and conversely, a low signal reflects a reduced state.

The GRX1-roGFP biosensor could be adapted to the *Fah*<sup>−/−</sup> model to assess the intracellular redox potential of repopulating hepatocytes. Mapping of the spatial and temporal redox status in repopulating hepatocytes will enable the elucidation of the oxidative stress response during the regenerative process. Furthermore, the expression of GRX1-roGFP could be coupled with overexpression and inhibition of Slc7a11 to determine the effects of varying xCT levels on the redox environment of repopulating hepatocytes.
Regulation of HNF4α occupancy during liver regeneration

HNF4α is a crucial factor that maintains mature hepatocytes in a differentiated state [28] by suppressing cell cycle gene expression and inhibiting hepatocyte proliferation [29,30]. Increased hepatocyte BrdU incorporation and Ki67 staining were observed in mice with conditional Hnf4a deletion, demonstrating the requirement of HNF4α to inhibit quiescent hepatocytes from reentering the cell cycle.

In the current study, I observed a loss of HNF4α binding to liver enhancers during repopulation, but have not investigated the mechanisms that led to its decrease in occupancy. Several possibilities include a loss of liver enhancer accessibility that results in HNF4α eviction [31], downregulation of Hnf4a expression [20,32] or protein abundance [33], and decreased HNF4α nuclear localization [34] that prevents its binding to liver enhancers. In fact, TRAP-seq identified a 50% decrease of Hnf4a transcripts in 4-week repopulating hepatocytes [20]. Whether other mechanisms contribute to the alteration of HNF4α occupancy is currently not known.

Elucidating the functional significance of altered HNF4α binding in the regenerating liver could also inform the utility of inhibiting HNF4α as a strategy to promote repopulation. It is plausible that HNF4α occupancy is required to maintain a euchromatic conformation at enhancers regulating liver functions [35], and a loss of binding could cause nucleosomes to become less accessible during liver regeneration. Targeted HNF4α deletion in repopulating hepatocytes in conjunction with the expression of GFP-L10a and SUN1-GFP proteins would enable the understanding of the direct effects of modified HNF4α occupancy on gene expression and chromatin accessibility.

The functional significance of CTCF in repopulating hepatocytes

CTCF plays numerous roles in genome regulation as an activator [36] or repressor [37] to modulate transcriptional activities, as an insulator to prevent enhancer-promoter interactions [38], as an organizer of chromatin structures to form topologically-associated domains [39], and as a modulator of long-range chromatin interactions to mediate looping [40], to name a few.
Understanding the functional importance of CTCF during liver repopulation will shed light on its utility as a therapeutic target and enable identification of additional regulators of the regenerative process. Incorporation of targeted degradation of CTCF mediated by the auxin-inducible degron system [41] with the Fah<sup>−/−</sup> mouse and cell type-specific isolation technologies would provide a model to study the effects of CTCF deficiency in repopulating hepatocytes. Genomic strategies to analyze changes in chromatin conformation [42,43] after the induction of CTCF degradation including chromosome conformation capture (3C) [44], chromosome conformation capture-on-chip (4C) [45], circular chromosome conformation capture (4C) [46], chromosome conformation capture carbon copy (5C) [47], Hi-C [48], and chromatin interaction analysis by paired-end tag sequencing (ChIA-PET) [49] could provide insight to the functional role of CTCF in nuclear organization and genome topology during liver regeneration.

The ability of CTCF to interact with distinct proteins in a context-dependent manner enables its diverse functions for transcriptional regulation [50,51]. Transcriptional cofactors are recruited by CTCF to specific loci for transcriptional activation and repression include Y-box DNA/RNA-binding factor (YB1) that enhances Myc repression [52], YY1 for X chromosome inactivation [53], class II transactivator (CIITA) to induce expression of major histocompatibility complex class II (MHC-II) genes [54], and general transcription factor II-I (TFII-I) to promote metabolic gene transcription [55].

Chromatin proteins also cooperate with CTCF to mediate insulation, looping, and transcription. The cohesin complex coordinates transcriptional insulation [56,57], H2A and H2A.Z induce localization to the nucleolus for insulation [58], and Suz12 recruits the polycomb repressive complex 2 to suppress the maternal Igf2 promoter [59]. Other proteins demonstrated to interact with CTCF include Poly(ADP-ribose) polymerase 1 (PARP1) for post-translational modification of CTCF to modulate chromatin insulation properties [60], the nucleolar protein nucleophosmin to localize β globulin insulator sites to the nuclear periphery for insulation [58], and RNA polymerase II that induces pausing to regulate alternative exon usage [61].

In the current study, I identified increased CTCF occupancy at promoters with elevated accessibility during liver regeneration. Stratification of changes in transcript levels revealed several
likely protein partners that cooperate with CTCF to differentiallymediate gene activation or repression. The transcriptional activator ZBTB3 is colocalized with CTCF at upregulated genes involved in cell growth and proliferation, whereas the transcriptional repressor YY1 co-occupies CTCF-bound downregulated promoters related to cell death regulation in repopulating hepatocytes. Assays to coimmunoprecipitate CTCF and ZBTB3 or YY1 should be performed to evaluate the direct or indirect interactions between CTCF and its cofactors. Experiments that implement genome-wide methods to examine ZBTB3 and YY1 binding sites could also be utilized to identify additional loci of colocalization. Finally, studies to manipulate levels of ZBTB3 and YY1 in repopulating hepatocytes could inform the functional significance of these transcription factors in liver regeneration.
CONCLUSIONS

With the development of cell type-specific technologies to profile transcriptomic and epigenomic changes that occur specifically in the repopulating hepatocytes, I identified Slc7a11 as a potential therapeutic target to promote liver regeneration after acute injury. Future work to assess the utility of Slc7a11 in the setting of chronic liver injury, the safety of viral-mediated or drug-induced transient or long-term xCT activation, and the mechanism of Slc7a11 to support hepatocyte replication will allow a more extensive understanding of the regulation of the oxidative/reduction network during liver repopulation.

Furthermore, my work provides insights on the combinatorial modulation of increased promoter accessibility and decreased liver-enriched enhancer accessibility underlying liver repopulation. These chromatin changes enable the activation of cell growth pathways and repression of liver metabolic functions. The mechanism of decreased HNF4α occupancy in liver-enriched enhancers and its effects on the liver metabolic gene program will enable the evaluation of whether HNF4α inhibition could be used as a strategy to induce hepatocyte replication during liver repopulation. The effects of CTCF on chromatin modification, the regulation of differential CTCF occupancy, and the mediation of transcriptional activity with additional CTCF cofactors are intriguing questions that may reveal the importance of nuclear organization and genome topology to fine-tune gene expression during the regenerative process.
REFERENCES


