1 INTRODUCTION

1.1 Liver Structure

1.1.1 Gross and Cellular Anatomy of the Liver

The liver is an extremely complex organ with over 500 diverse functions in the human body. These functions range from secretion of blood components to xenobiotic detoxification and glucose homeostasis [1]. In order to facilitate these many diverse functions, the structure of the liver is extremely complex. The smallest functional unit of the liver is the liver lobule, a polygonal mass of tissue consisting of 3-6 portal triads surrounding a central vein [2]. Each portal triad consists of a branch of the portal vein, a branch of the hepatic artery, and a bile duct. Blood flows through hepatic sinusoids in each lobule from the portal vein to the central vein while bile flows in the opposite direction.

The hepatic sinusoids are blood vessels with a fenestrated, discontinuous endothelium composed of liver sinusoidal endothelial cells (LSECs) [2]. This discontinuous endothelium allows for bidirectional exchange between the blood and hepatocytes. The liver sinusoids as well as the space between the LSECs and hepatocytes, termed the Space of Disse, are home to several non-parenchymal hepatic cell types including Kupffer cells, the tissue resident macrophages, and hepatic stellate cells (HSCs), the liver resident fibroblasts. In addition, the Space of Disse contains several populations of antigen presenting, myeloid, and innate and adaptive lymphoid resident immune cells. These immune cells participate in maintaining organ homeostasis, innate and adaptive immunity, as well as regulating the inflammatory and fibrotic response during injury [3]. The location of these cells allows for efficient immune surveillance and clearance as well as facilitating metabolic functions.

Hepatocytes are the parenchymal cell of the liver and are responsible for the metabolic functions of the organ. As blood flows through the sinusoids, hepatocytes absorb oxygen, nutrients, xenobiotics, and toxins for metabolism, storage, and detoxification [4]. This creates an oxygen and nutrient gradient across the liver lobule which necessitates the zonation of different metabolic processes. The liver can be divided into three functional zones based upon these gradients: zone 1 (periportal), zone 2 (transition zone), zone 3 (pericentral). The periportal zone is where oxygen and nutrient rich blood enters the liver.

Thus, the oxygen and nutrient concentrations decrease along the spectrum from periportal to transition to pericentral. Metabolic processes with zonal organization include xenobiotic metabolism, lipid metabolism, and bile acid synthesis [5].

As bile flows in the opposite direction of blood in the liver, bile acid synthesis mainly occurs in the pericentral area. Bile is produced by hepatocytes before being secreted into bile canaliculi which are lined by cholangiocytes. Cholangiocytes line the entirety of the biliary system and are responsible for transporting and modifying bile. The unique structure of the liver maintains a barrier between the blood and bile using tight junctions between hepatocytes [4]. A schematic of the liver lobule can be found in Figure 1.1.

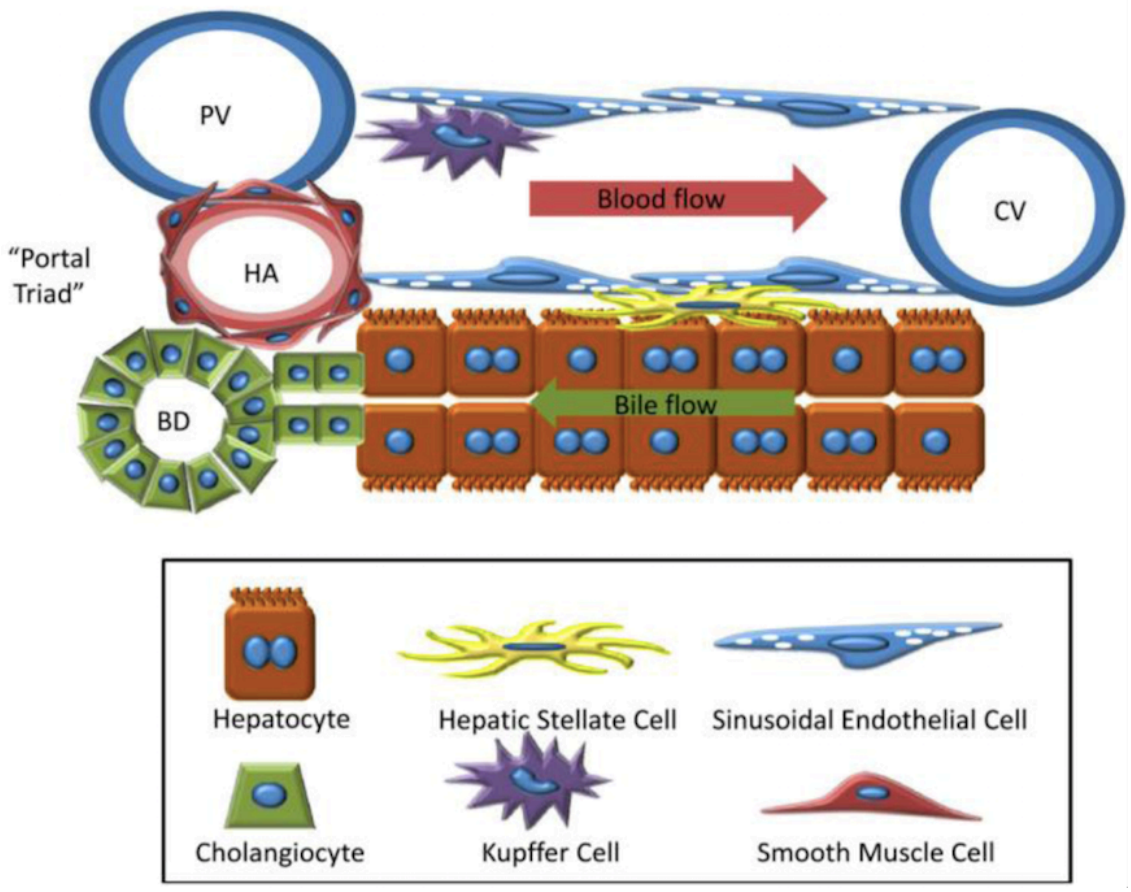

Figure 1.1 Organization of the liver lobule.

Cords of hepatocytes extend between the portal triad and the central vein. Blood flows through liver sinusoids lined by liver sinusoidal endothelial cells from the portal triad to the central vein. Kupffer cells reside within the sinusoid while hepatic stellate cells are located in the Space of Disse, the space between the LSECs and hepatocytes. Bile is secreted by hepatocytes and flows through bile canaliculi in the opposite direction as blood to a bile duct lined by cholangiocytes. (Figure adapted from Kang et al. Cells 2012) [1]

1.1.2 Hepatocytes

Hepatocytes are the most abundant cell type in the liver comprising approximately 60% of the total cell population and 80% of the total volume of the organ [6]. Hepatocytes are responsible for liver functions, including but not limited to, synthesis of serum proteins, storage of vitamins, intermediate metabolism of amino acids, lipids, and carbohydrates, xenobiotic detoxification, and bile acid synthesis.

In order to maintain their functionality, hepatocytes are polarized with three distinct membrane domains: sinusoidal (basal), lateral, and canalicular (apical). The sinusoidal membrane domain is in contact with the blood and non-parenchymal cells to facilitate export of serum proteins and import of oxygen, nutrients, and toxins from the blood. This surface has microvilli that extend into the Space of Disse, interact with HSCs, and occasionally extend through the fenestrae of the LSECs directly into the blood. These microvilli increase the surface area of the hepatocyte and enhance the capacity for receptor and transport functions. The lateral surface forms the blood-bile barrier by fusing hepatocytes together with tight junctions. The canalicular domain forms a bile canalicular network between adjacent hepatocytes to transport bile produced by hepatocytes to the bile ducts [4, 7]. Bile is a complex fluid that contains organic and inorganic substrates in an alkaline solution. Bile aids in digestion and absorption of lipids and lipid soluble vitamins while simultaneously eliminating waste products such as bilirubin and cholesterol.

The unique vascular organization of the liver exposes hepatocytes to a lobular gradient of oxygen, nutrients, toxins, and hormones. As a result, hepatocytes are heterogeneous in both their structure and function along this gradient. Hepatocytes in the periportal region are exposed to a high concentration of nutrients and oxygen, as a result, they are small, participate in gluconeogenesis and fatty acid degradation, and have a low capacity for xenobiotic detoxification. Alternately, pericentral hepatocytes are larger and are responsible for bile acid synthesis and xenobiotic metabolism. Transitional hepatocytes have structure and function intermediate between periportal and pericentral hepatocytes [5]. A schematic of metabolic zonation can be found in Figure 1.2. Hepatocytes are unique

among terminally differentiated epithelial cells as they maintain the ability to divide throughout adult life. This ability to re-enter the cell cycle, in part, grants the liver its unique ability to regenerate following injury [6].

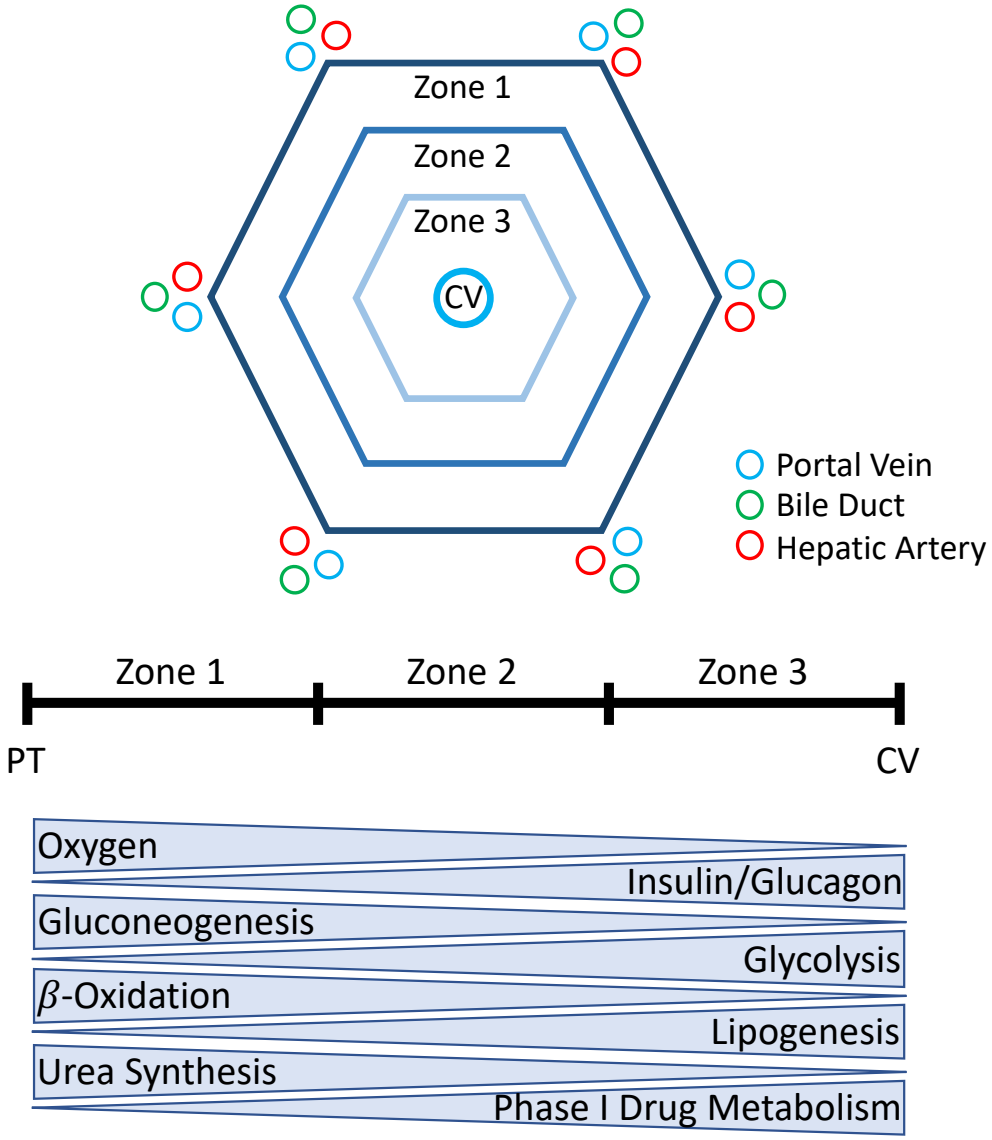

Figure 1.2 Metabolic Zonation.

The liver lobule can be divided into three functionally different metabolic zones. Zone 1 also known as the periportal region has high levels of oxygen and hepatocytes in this zone are responsible for gluconeogenesis, β-oxidation, and urea synthesis. Zone 2 is labelled the transition zone and hepatocytes in this zone represent a functional intermediate between Zone 1 and Zone 3 hepatocytes. Zone 3 is known as the pericentral region due to its proximity to the central vein. Hepatocytes in this region are exposed to low levels of oxygen and high levels of hormones and are responsible for glycolysis, lipogenesis, and phase one drug metabolism.

1.1.3 Cholangiocytes

Cholangiocytes, also known as biliary epithelial cells, represent only 3-5% of the cells in the liver. These cells line the biliary tree and are responsible for transporting and modifying bile. Cholangiocyte secretions account for approximately 40% of the total volume of bile. As the bile is transported from canaliculi, through progressively larger bile ducts, to the gallbladder, the bile is alkalinized and diluted by cholangiocytes in a series of secretory and absorptive processes. Cholangiocytes possess cilia on their apical plasma membrane that extends into the bile duct lumen to detect changes in bile flow, composition, and osmolality.

Cholangiocytes are heterogeneous throughout the biliary tree. Large cholangiocytes line the larger ducts and are responsible for mucin secretion and hormone regulated bile secretion while small cholangiocytes line the smaller ducts and have high proliferative capacity and considerable plasticity. Upon liver injury, cholangiocytes can participate in the regenerative process either by repairing bile ducts or transdifferentiating into hepatocytes in a process termed ductular reaction. Though this process can contribute to liver regeneration, in many chronic liver diseases, prolonged ductular reactions may contribute to disease progression and fibrosis. The study of these reactions have been hindered because the type and mechanism of ductular reaction is dependent on the injury, disease, and model system used; therefore, additional studies are needed to fully elucidate the role of these reactions in liver regeneration and disease [8-10]. Additionally, cholangiocytes stimulate angiogenesis in response to nutritional and functional demands of proliferating cells. It is believed that small cholangiocytes are responsible for this function. Cholangiocytes also participate in the immune system by secreting chemokines and cytokines to localize and coordinate immune responses [11, 12].

1.1.4 Hepatic Stellate Cells

Hepatic stellate cells (HSCs) represent approximately 5-8% of liver cells. These cells reside in the Space of Disse in close proximity to both hepatocytes and LSECs. These cells have long, branching cytoplasmic processes that allow them to cover the entire hepatic sinusoidal microcirculatory network [13]. In their quiescent state, HSCs are responsible for vitamin A storage in the form of retinyl esters. 50-80% of retinol within

the body is stored in the liver and 80-90% of those retinoids are stored in lipid droplets in HSCs [14].

When the liver faces a chronic injury that overwhelms the normal regenerative processes, HSCs become activated. These activated HSCs lose their retinoid stores, acquire a highly proliferative myofibroblast-like phenotype, and become the primary producer of extracellular matrix components. Activated HSCs produce mostly collagen 1 and are the cell type thought to be responsible for liver fibrosis. HSCs can be activated by mediators secreted by immune cells and necrotic hepatocytes. Additionally, activated HSCs can recruit immune cell transmigration into the parenchymal space thereby further enhancing hepatic injury [15]. A schematic of HSC activation can be found in Figure 1.3.

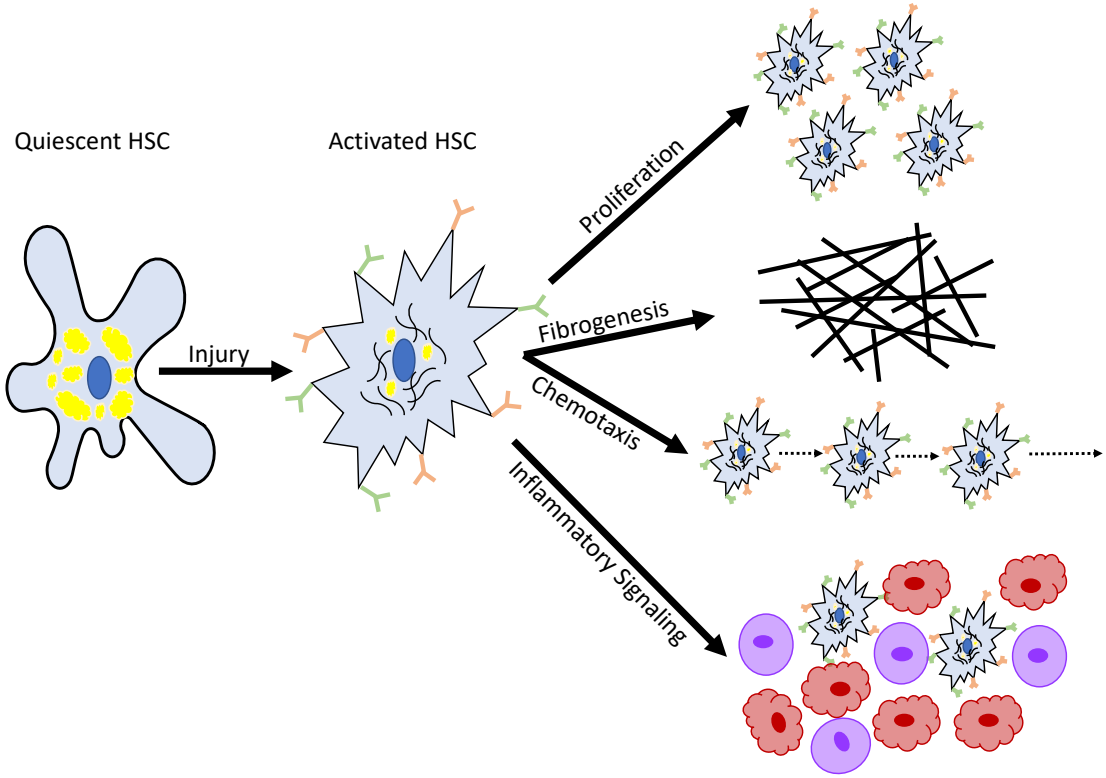

Figure 1.3 Activation of HSCs.

Quiescent HSCs are responsible for storing the majority of the body's retinol in lipid droplets. However, upon liver injury, these HSCs can become activated which causes several phenotypic changes. Activation of HSCs causes the depletion of intracellular retinol stores, increases proliferation and chemotaxis, stimulates fibrogenesis through production of irregular ECM components, and produces pro-inflammatory cytokines that modulate the inflammatory response to the injury.

1.1.5 Kupffer Cells

Kupffer cells are the resident liver macrophages. These cells are the largest tissue resident macrophage population in the body and make up approximately 12% of the cells within the liver. These cells reside in the sinusoidal lumen attached to LSECs. They function in the removal of gut- and environment-derived toxins, microorganisms, senescent and damaged red blood cells, and circulating neoplastic cells. The blood flow through sinusoids is slow which prolongs the interaction between Kupffer cells and portal blood which enhances the removal capacity of these cells. Kupffer cells can be activated by LPS and the complement system. Upon activation, Kupffer cells produce large quantities of inflammatory mediators which contribute to liver injury. The distribution of Kupffer cells is variable throughout the liver lobule with the highest number in the periportal region (~40%) and the lowest in the pericentral region (~25%). The Kupffer cells in the periportal region are the largest, perform the most phagocytosis, and produce the most cytokines of all the regions [16, 17].

1.1.6 Liver Sinusoidal Endothelial Cells

Liver sinusoidal endothelial cells (LSECs) are the second most abundant cell type of the liver comprising approximately 20% of all cells in the organ. LSECs are positioned between the blood and hepatocytes and comprise the liver sinusoids. The LSECs create a discontinuous endothelium with fenestrae that act as a sieve. The fenestrae measure approximately 150-175 nm and occupy 6-8% of the total endothelial surface. Porosity and fenestration diameter increase across the liver lobule from the periportal region to the pericentral region. LSEC protect parenchymal cells from direct contact with the blood by blocking passage of particles larger than 200 nm. Only smaller remnants can enter the Space of Disse and be metabolized by hepatocytes. The diameter of fenestrae is influenced by many factors and certain drugs can lead to permanent defenestration of LSECs which causes atherosclerosis and hyper-lipidemia by preventing the removal of lipoproteins from the blood [18].

These cells also possess endocytic capabilities and act as an important blood clearance system. They express at least 5 types of scavenger receptors to remove biomacromolecules, proteins, polysaccharides, lipids, and nucleic acids. LSECs regulate the migration of immune cells into the parenchymal space thereby mediating the immune

reaction to liver injury. As antigen presenting cells, LSECs are also a major contributor to the innate immunity in the liver [19].

1.2 Metabolic Functions of the Liver

1.2.1 Glucose Homeostasis

The liver is the metabolic hub of the body controlling the storage and availability of glucose to other tissues. The liver is the major source of post-prandial glucose uptake with 30-40% of portal vein glucose being absorbed on the first pass. This glucose can then be utilized by several different pathways. The glucose can be metabolized through glycolysis to meet the immediate energy needs of the organ. Alternately, the glucose may enter one of two storage pathways: de novo lipogenesis or glycogen synthesis. In addition to its role in post-prandial glucose storage, the liver is also capable of de novo synthesis of glucose from precursors such as lactate, amino acids, and glycerol in a process called gluconeogenesis [20].

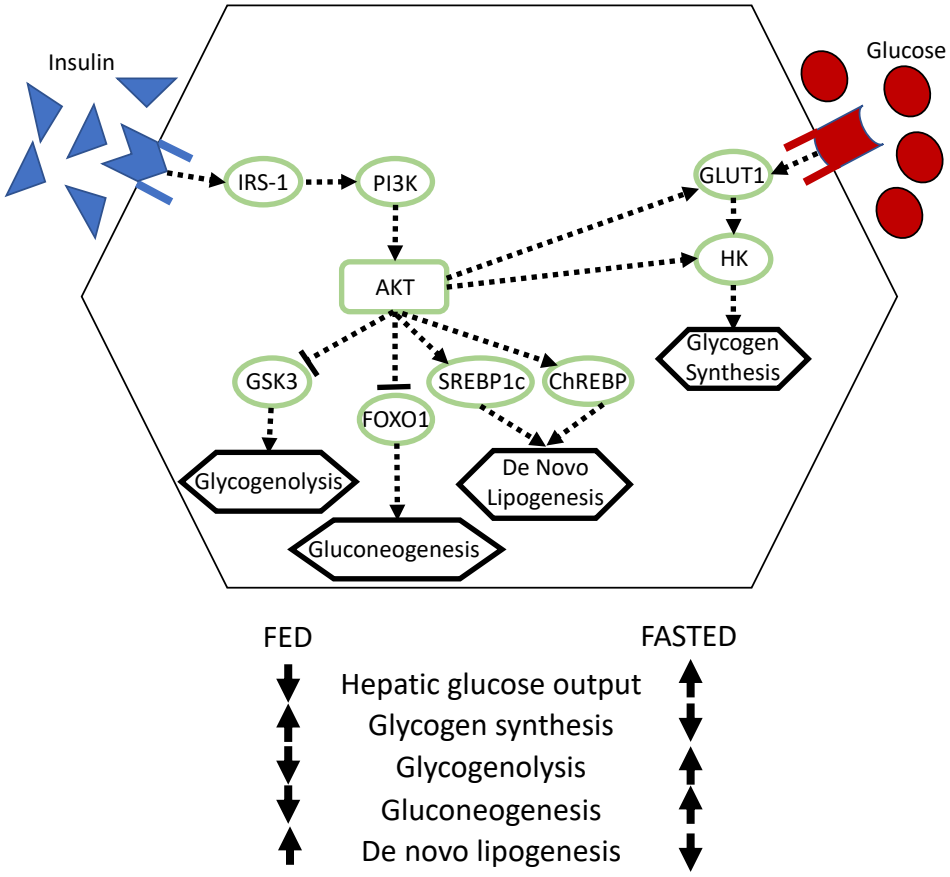

Figure 1.4 Glucose Homeostasis.

Schematic of AKT signalling in response to insulin and glucose. The binding of insulin to its receptor causes the phosphorylation of IRS-1 which in turn activates PI3K. PI3K activation then leads to the phosphorylation of AKT which has several downstream effects. AKT inhibits GSK3 and FOXO1 which leads to the downregulation of glycogenolysis and gluconeogenesis, respectively. Alternately, AKT activates SREBP1c and ChREBP which stimulates de novo lipogenesis. AKT also activates glycolysis and glycogen synthesis by upregulating the expression of GLUT1 and HK. In the fed state, glucose storage pathways such as glycogen synthesis and de novo lipogenesis are upregulated. In the fasted state, glucose production pathways such as glycogenolysis and gluconeogenesis are upregulated to meet the metabolic needs of the body.

Glucose homeostasis in the liver is under tight nutritional and hormonal control. In the post-prandial state, glucose and insulin levels are high which leads to the activation of the AKT signalling cascade. AKT activation leads to the upregulation of SREBP1c and ChREBP which are the master regulators of the de novo lipogenesis pathway. In addition, AKT inhibits glycogenolysis and gluconeogenesis through phosphorylation of GSK3 and FOXO1, respectively [21]. In the fasted state, insulin levels are low which releases the inhibitory effect on FOXO1. This transcription factor activates de novo glucose production by upregulating the gluconeogenic gene program and increasing the flux of gluconeogenic precursors to the liver. FOXO1 also inhibits glycolysis and de novo lipogenesis to prevent glucose utilization by the liver in this state [22, 23]. In the fasted state, glucagon levels are also very high. Glucagon activates glycogen phosphorylase which leads to glycogenolysis [24, 25]. In all, during times of high glucose availability, the liver uptakes and stores excess glucose while in times of low glucose availability, the liver generates and releases glucose for use in other tissues. A summary of the signalling pathways involved in glucose homeostasis can be found in Figure 1.4.

1.2.2 Lipid Metabolism

The liver is also a major site of lipid metabolism within the body. Lipids in the liver can be derived from three major sources: de novo lipogenesis, non-esterified fatty acids from the lipolysis of adipose tissues, and dietary fats transported to the liver in chylomicrons. Fatty acids within the liver have three major fates: storage as triglycerides in lipid droplets, secretion in very low-density lipoproteins (VLDL), or oxidation for energy utilization [26].

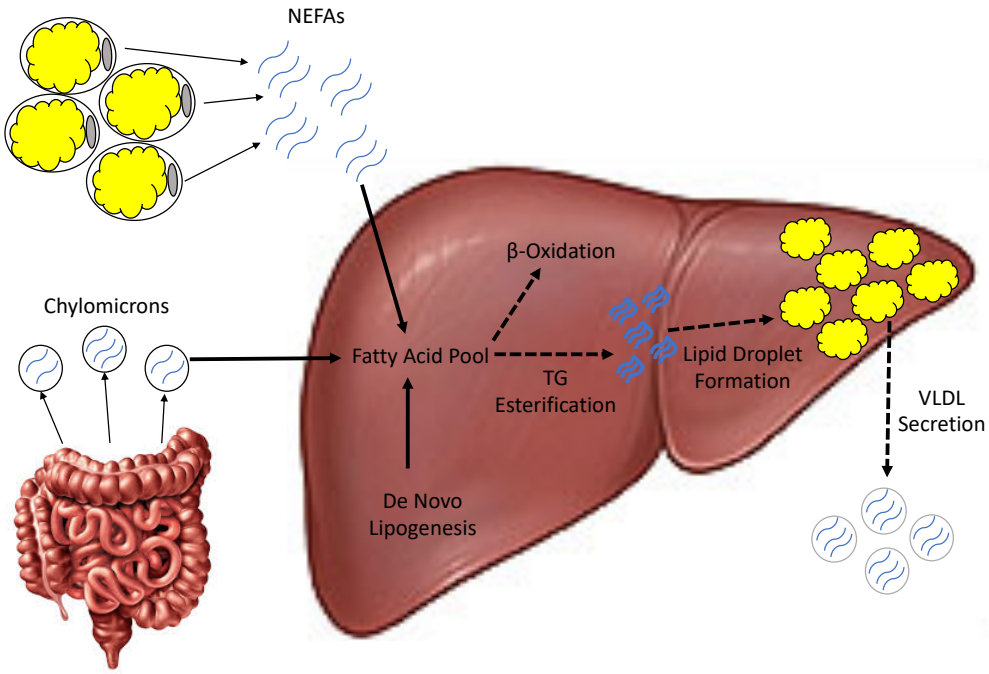

Figure 1.5 Fatty acids in the liver.

There are three primary sources of fatty acids within the liver: diet, adipose tissue, and hepatic de novo lipogenesis. Fatty acids from the diet are transported to the liver in chylomicrons while fatty acids from adipose tissue are transported in the form of non-esterified fatty acids. Upon entering a hepatocyte, the fatty acids can be metabolized through β-oxidation or esterified into triglycerides for storage in lipid droplets or secretion as VLDL.

1.2.2.1 De Novo Lipogenesis

De novo lipogenesis (DNL) converts excess carbohydrates to fatty acids for energy storage. Glucose is first converted to pyruvate before entering the Krebs Cycle to yield citrate. Citrate is then broken down into acetyl-CoA and oxaloacetate. Acetyl-CoA is then converted to malonyl-CoA by acetyl-CoA carboxylase (ACC) in the rate limiting step. Fatty acid synthase (FAS) then adds successive molecules of malonyl-CoA to an acetyl-CoA primer to generate palmitic acid. Palmitic acid is a saturated fatty acid with a 16-carbon chain. Palmitic acid can then be desaturated by stearoyl-CoA desaturase (SCD-1) to form a monounsaturated fatty acid. Alternately, it can be elongated by elongation of very long chain fatty acids protein 6 (ELOVL6) which catalyses the addition of 2 carbon units to fatty acid chains [26]. These fatty acids can then be esterified into triglycerides for storage or oxidized in the mitochondria for energy production.

De novo lipogenesis is tightly regulated by nutritional, hormonal, and genetic factors. The master regulators of de novo lipogenesis are the transcription factors SREBP1c and ChREBP [21, 27]. SREBP1c is regulated by intracellular sterol levels, insulin, and glucagon. SREBP1c is initially produced as a large precursor molecule that is bound to the ER membrane. SCAP is a sterol sensing protein that acts as an escort protein for SREBP1c from the ER to the Golgi Apparatus. When sterols are abundant, SCAP undergoes a conformational change and is unable to bind to SREBP; however, when sterols are absent, SCAP binds to SREBP1c and facilitates its translocation to the Golgi. Following this translocation, SREBP1c is proteolytically cleaved from the membrane by S1P and S2P before translocating to the nucleus. Nuclear SREBP1c activates lipogenic genes by binding the sterol response elements in their promoter/enhancer regions [27, 28]. ChREBP is a glucose responsive transcription factor that regulates the expression of lipogenic genes specifically induced by carbohydrates. Glucose availability controls both the nuclear translocation and functionality of ChREBP. Under fasting conditions, ChREBP is phosphorylated which sequesters it to the cytosol. ChREBP also has a glucose sensing module that controls its ability to bind to DNA. In low glucose concentrations, the low glucose inhibitory domain causes a conformational change that prevents ChREBP from binding to DNA. High glucose concentrations result in both dephosphorylation and restoration of DNA binding capacity. ChREBP and its functional partner MLX then form a heterodimer that binds to the ChoRE motifs in target lipogenic genes [29, 30]. Both SREBP1c and ChREBP activation leads to the upregulation of de novo lipogenesis genes including ACC, FAS, SCD1, and ELOVL6. An overview of this transcription regulation can be found in Figure 1.6.

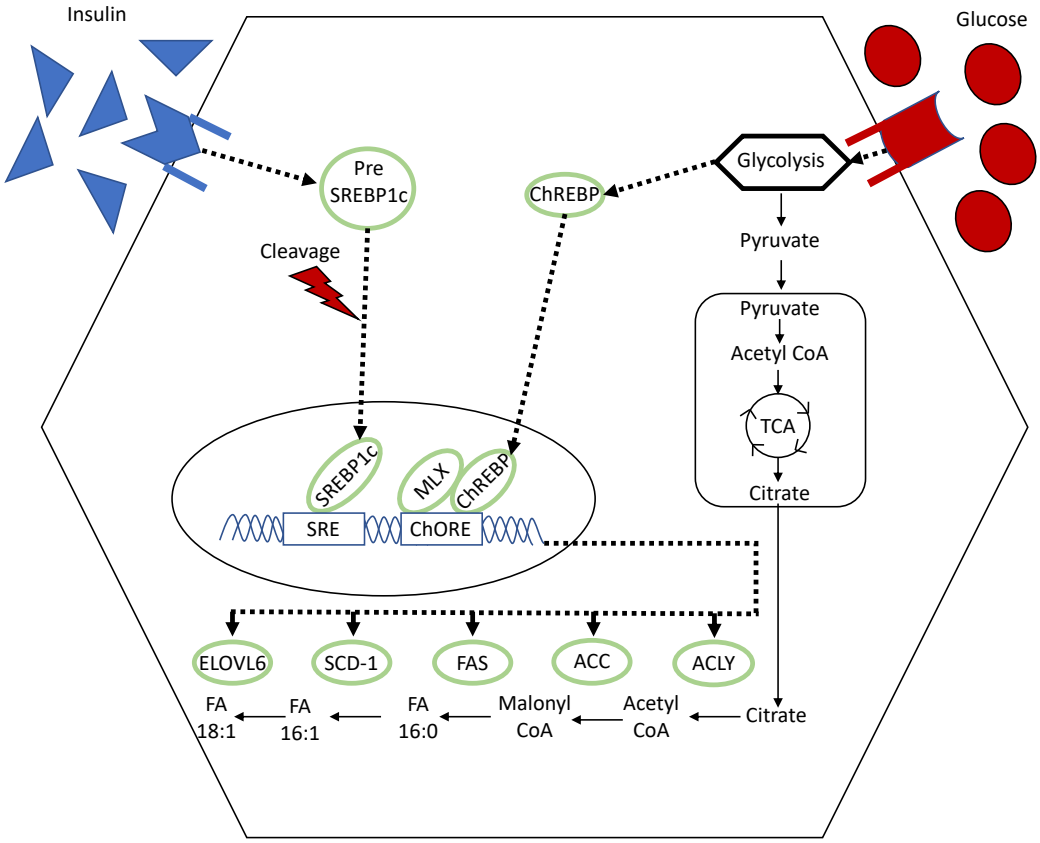

Figure 1.6 De Novo Lipogenesis.

Transcriptional regulation of de novo lipogenesis in hepatocytes. High levels of insulin cause increased transcription and nuclear translocation of SREBP1c where it binds to sterol response elements and promotes the transcription of downstream lipogenesis genes such as ACLY, ACC, FAS, SCD-1, and EVOVL6. High glucose concentrations have the dual effect of activating glycolysis to increase the availability of de novo lipogenesis precursors such as citrate as well as activating ChREBP which translocates to the nucleus and binds to carbohydrate response elements along with its functional partner MLX in order to stimulate the transcription of de novo lipogenesis genes.

1.2.2.2 Lipid Storage in Lipid Droplets

Lipids within hepatocytes derived from exogenous or endogenous pathways must often be stored for later use. This storage is accomplished using a specialized organelle called a lipid droplet. Lipid droplets contain a pool of neutral lipids surrounded by a phospholipid monolayer membrane decorated with various lipid droplet proteins. These droplets serve as an immediate energy source for the hepatocyte. However, lipid droplets are not inert storage receptacles for fat. These organelles are extremely dynamic and play a vital role in lipid homeostasis and intracellular signalling. The protein coating of the lipid droplet is tightly controlled and can be efficiently altered in order to regulate lipid

droplet biogenesis and balance the metabolic needs of the cell at any given time. Additionally, lipid droplets may function as a type of chaperone to transport specific lipid species to different cellular locations for specific functions [31-34]. Free fatty acids, especially saturated fatty acids, can be harmful to hepatocytes by inducing oxidative stress and mitochondrial dysfunction. Conversion of these metabolically harmful species to metabolically inert triglycerides in lipid droplets represents an adaptive mechanism to protect against lipotoxicity [35]. However, if acquisition of lipids outpaces their utilization and/or secretion, excessive numbers of lipid droplets can accumulate within the cell [26]. This accumulation of lipid droplets causes hepatic steatosis which can result in insulin resistance and metabolic dysfunction.

1.2.2.3 VLDL Secretion

VLDLs are lipoprotein particles that are assembled and secreted by the liver when plasma lipid levels are too low or the lipid content within the liver is too high. These lipoproteins contain triglycerides, cholesterol, and phospholipids. VLDLs are assembled in a two-step process that is dependent on the availability of the protein apolipoprotein B-100 (APOB-100). VLDL can be separated from the other major lipoprotein, chylomicrons, by the APOB protein that is used as a scaffolding. VLDL are built upon APOB-100 while chylomicrons are assembled on APOB-48 [36]. Following synthesis, ApoB-100 is translocated through the rough ER. During this translocation, ApoB-100 acquires a small number of triglycerides, phospholipids, and cholesterol esters and forms a small, dense VLDL-precursor. This precursor then goes through a maturation process in the smooth ER where it is fused with lipid droplets [37, 38]. The fully lipidated VLDL is then transported to the plasma membrane and secreted into circulation. These VLDL particles can then be hydrolysed in the peripheral vascular bed releasing the fatty acids for use by other peripheral tissues.

1.2.2.4 Fatty Acid β-Oxidation

Given that fatty acids produce the most energy per gram of any macronutrient, in times of low energy availability, they can be broken down and utilized for ATP production. β-oxidation is the catabolic process by which fatty acids are degraded into acetyl-CoA subunits that can then enter the citric acid cycle to produce energy. There are two major β-oxidation pathways. The mitochondrial β-oxidation pathway is the primary pathway

and is responsible for the degradation of short, medium, and long-chain fatty acids [39]. Whereas degradation of long and very long-chain fatty acids occurs via the peroxisomal β-oxidation pathway [40].

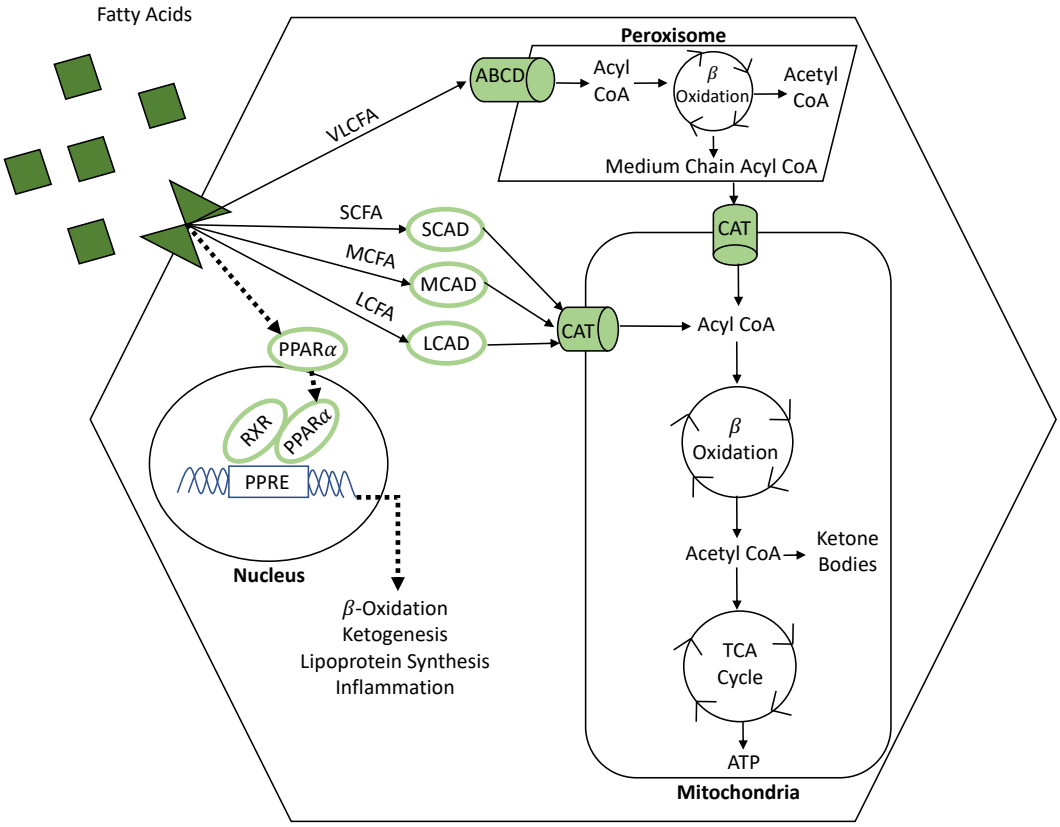

Figure 1.7 β-Oxidation.

Schematic of mitochondria and peroxisomal β-oxidation. Upon entering the cell, fatty acids activate PPARα which binds in conjunction with RXR to peroxisome proliferator response elements in target genes to upregulate the expression of genes involved in β-oxidation, ketogenesis, VLDL synthesis, and inflammation. The length of the fatty acid determines if it will first enter the peroxisome or the mitochondria for oxidation. Short, medium, and long chain fatty acids are activated into acyl-CoA esters by their respective acyl-CoA dehydrogenase enzymes before entering the mitochondria through the carnitine shuttle. In the mitochondria, the acyl CoA is slowly shortened by releasing two-carbon acetyl CoA subunits at a time. These acetyl CoA subunits are then released as ketone bodies or enter the TCA cycle to produce ATP. Alternately, very long chain fatty acids cannot be transported into the mitochondria. Therefore, these fatty acids are partially oxidized in the peroxisome until they are shortened enough to enter the mitochondria for full oxidation.

The first step in mitochondrial β-oxidation is the activation of fatty acids into acyl-CoA esters [41]. These acyl-CoA esters can then be transported across the inner mitochondrial

membrane by the carnitine shuttle. Following entry into the mitochondria, the acyl-CoA is progressively shortened by releasing two carboxy-terminal carbon atoms as acetyl-CoA subunits. These subunits can then be condensed into ketone bodies for export to other organs or enter the citric acid cycle to be further oxidized into water and carbon dioxide to produce ATP. The shortened acyl-CoA then re-enters the fatty oxidation cycle until it has been fully oxidized [42]. Peroxisomal β-oxidation is used to degrade very long chain fatty acids because the mitochondria lack the ability to convert these fatty acid species into acyl-CoA esters which prevents them from being imported into the mitochondria. The energy generated in peroxisomal β-oxidation dissipates as heat and therefore is less energy efficient than mitochondrial β-oxidation [43, 44]. For this reason, peroxisomal β-oxidation does not run to completion. Rather, acyl-CoA esters are shortened to the point they can be imported into the mitochondria for completion. An overview of the mitochondrial and peroxisomal β-oxidation pathway can be found in Figure 1.7.

Fatty acid β-oxidation is nutritionally and transcriptionally controlled. Peroxisome proliferator-activated receptor α (PPARα) is the master regulator of all the β-oxidation pathways [41, 45, 46]. PPARα is a nuclear receptor protein that, upon activation by peroxisome proliferators, functions as a transcription factor. When activated, PPARα heterodimerizes with the retinoid X receptor (RXR) and binds the peroxisome proliferator hormone response element in target genes to activate the β-oxidation gene program. PPARα is activated by the presence of excess saturated and polyunsaturated fatty acids. Under fasting conditions, there is an increase in flux of non-esterified fatty acids from the adipose tissue into the liver which results in the activation of PPARα signalling cascade [47-49]. In addition to activating β-oxidation, PPARα also plays a role in lipoprotein synthesis, inflammatory responses, and liver cancer.

1.2.3 Drug Metabolism

The liver plays an extremely crucial role in metabolizing xenobiotics and toxins. The liver metabolizes these drugs in order to either modify, bioactivate, detoxify, and/or excrete them. Hepatic drug metabolism is divided into three phases: modification, conjugation, and excretion [50]. Phase I reactions are generally facilitated by the cytochrome P450 (CYP) superfamily that function as monooxygenases. There are 57 CYP isoforms that

have been identified in humans, each with a unique substrate affinity and functionality [51]. Each CYP isoform can be induced or inhibited by various endogenous and exogenous substrates. This pronounced variability makes phase I drug metabolism particularly complex. The second phase of xenobiotic metabolism is conjugation. Enzymes reduce the reactivity and increase the polarity of the drug by adding charged and bulky groups such as methyl, acetyl, sulfate, glutathione, glycine, and glucuronic acid. This increased polarity primes the substrate for the third phase, excretion. The substrates are excreted from the cell through specialized transporters in the plasma membrane. These substrates can either be excreted into the serum or into bile for disposal. A schematic of hepatic drug metabolism is shown in Figure 1.8.

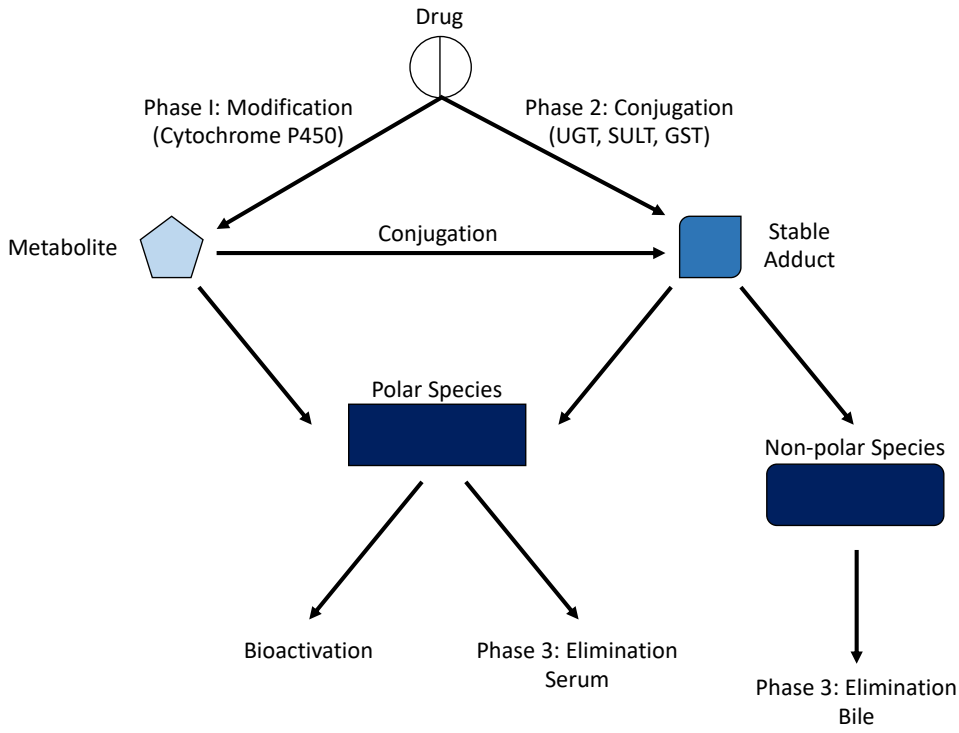

Figure 1.8 Hepatic drug metabolism.

Basic overview of hepatic drug metabolism. Upon entering a hepatocyte, a drug may be modified by Phase I enzymes such as CYP enzymes to produce a metabolite or conjugated by Phase 2 enzymes to form a stable adduct. Following the first two phases of metabolism, the resulting metabolite can be categorized as either polar or non-polar. Non-polar species are excreted into the bile for faecal elimination. Polar species can then undergo further metabolic processes that lead to its bioactivation or they can be secreted into the serum for elimination in the urine. Secretion into either the bile or the serum is facilitated by phase 3 enzymes.

1.3 Non-Alcoholic Fatty Liver Disease

Non-alcoholic fatty liver disease (NAFLD) is now the leading cause of chronic liver disease in the developed world [52-54]. In basic terms, NAFLD arises from an imbalance between triglyceride acquisition and removal. NAFLD is thought to be the hepatic manifestation of the metabolic syndrome and ranges from simple steatosis to non-alcoholic steatohepatitis (NASH). Simple steatosis is defined as the benign accumulation of fat within the liver. NASH is a progressive liver injury that is characterized by a profound inflammatory response that can develop into fibrosis, cirrhosis, and hepatocellular carcinoma. Only a small subset of patients will progress from simple steatosis to NASH and the mechanism causing this progression has yet to be fully elucidated [55]. Factors that lead to increased metabolic dysfunction and obesity such as high fat diet and sedentary lifestyle contribute to NAFLD development; however, there is a growing appreciation for the role of genetics and epigenetics in NAFLD disease progression [56].

1.3.1 Public Health Burden of NAFLD

NAFLD is defined as the presence of hepatic steatosis, defined as fat accumulation in at least 5% of hepatocytes, when no other cause for hepatic fat accumulation is present [52, 57]. The epidemiology of NAFLD is difficult to quantify given the lack of a reliable, affordable, non-invasive method to measure liver fat that can be used to precisely ascertain the number of affected individuals worldwide. Currently, several different techniques are used to quantify liver fat (ultrasound, elastography, H-MRS, MRI, biopsy, etc) with varying levels of accuracy and accessibility [57]. Given these discrepancies, the reported prevalence of NAFLD ranges from 6-35% [52]. The prevalence is highest in western industrialized countries with a median prevalence of 25-26% [52, 58]. The global prevalence is summarized in Figure 1.9.

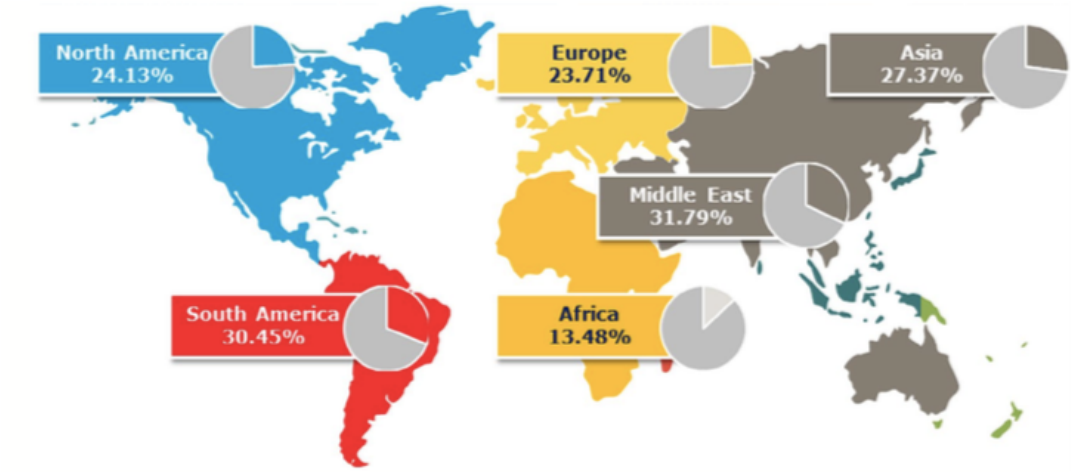

Figure 1.9 Global prevalence of NAFLD.

The prevalence of NAFLD is difficult to quantify but it is currently estimated to affect approximately one in four adults globally. The highest prevalence is in the Middle East and South America while the lowest prevalence is in Africa. (Figure adapted from Younossi et al. Hepatology 2018) [59]

Some patients with simple steatosis will go on to develop more severe forms of the disease such as NASH and fibrosis. It is difficult to estimate what percentage of NAFLD patients suffer from NASH given the invasive nature of liver biopsy which is necessary to confirm this diagnosis. However, studies estimate that up to one in three patients with NAFLD have biopsy confirmed NASH. Of the patients with NASH, 20-30% will progress to severe fibrosis within 10 years and 10-29% will develop cirrhosis. In patients with NASH-induced cirrhosis, 4-27% will go on to develop hepatocellular carcinoma [55, 56]. A schematic of NAFLD progression can be found in Figure 1.10. NASH is projected to become the leading indication for liver transplantation in the United States within the next 10 years [60]. However, 70% of patients that receive a liver transplantation for NASH go on to develop NAFLD in the transplanted liver [61].

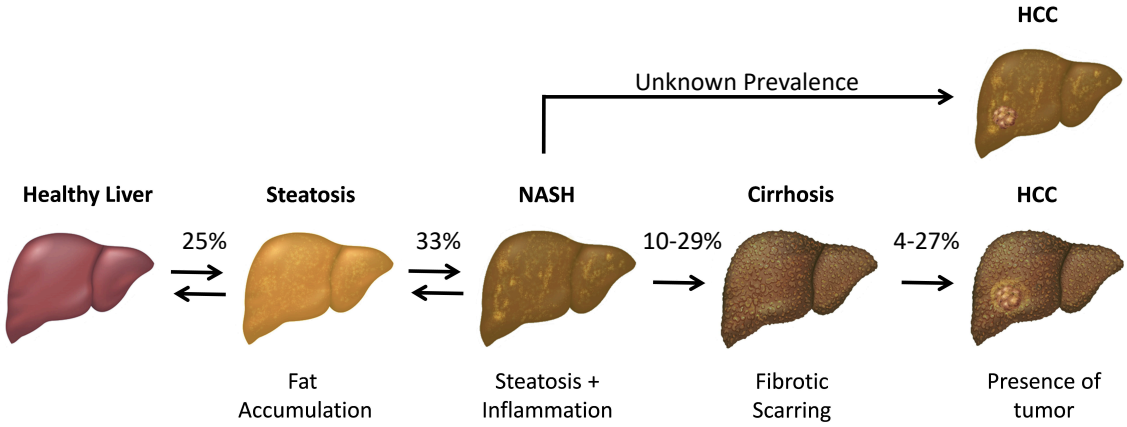

Figure 1.10 NAFLD progression.

The natural progression of NAFLD follows from steatosis to NASH to cirrhosis and finally to hepatocellular carcinoma. Some patients develop hepatocellular carcinoma without underlying cirrhosis and others fail to progress along the full spectrum of the disease. In the general population, approximately 25% of adults have hepatic steatosis. Of those patients, approximately one in three will develop NASH. In patients with NASH, 10-29% will go on to develop cirrhosis and 4-27% will develop hepatocellular carcinoma. (Figure adapted from Turchinovich et al. Frontiers in Physiology 2018) [62]

Patients with NAFLD have a 34%-69% increased risk of mortality over 15 years compared with the general population. Cardiovascular disease is the number one cause of death in NASH patients with hepatic complications accounting for approximately 5% of deaths [56]. In addition to cardiovascular disease, NASH patients have an increased risk of bowel and breast cancer. Despite this extremely high morbidity and mortality, there are currently no FDA approved pharmaceuticals for the treatment of NAFLD or NASH. At present, the only available treatment for this disease is lifestyle modification such as weight loss and exercise. A reduction in body weight by 5-10% has been shown to result in reversion of NASH and fibrosis. However, lifestyle modifications are difficult to achieve and maintain resulting in limited success of this treatment in the clinical setting [58].

1.3.2 Pathogenesis of NAFLD

The pathogenesis of NAFLD generally involves several individual insults that culminate in severe disease including severe lipid accumulation, insulin resistance, oxidative stress, and mitochondrial dysfunction leading to lipotoxicity and inflammation. Scientists have adopted the multiple-hit hypothesis to describe the pathogenesis of NAFLD, illustrated

in Figure 1.11. This hypothesis asserts that NAFLD progression requires a combination of multiple genetic, external, and intracellular events [63]. Some patients are predisposed to developing NAFLD due to genetic and epigenetic factors. In these patients, several environmental factors such as insulin resistance and oxidative stress lead to a state of chronic hepatic inflammation due to the activation of several different hepatocellular damage pathways. For this reason, the pathogenesis of NAFLD is extremely complex and the mechanism of disease progression is unique to each patient.

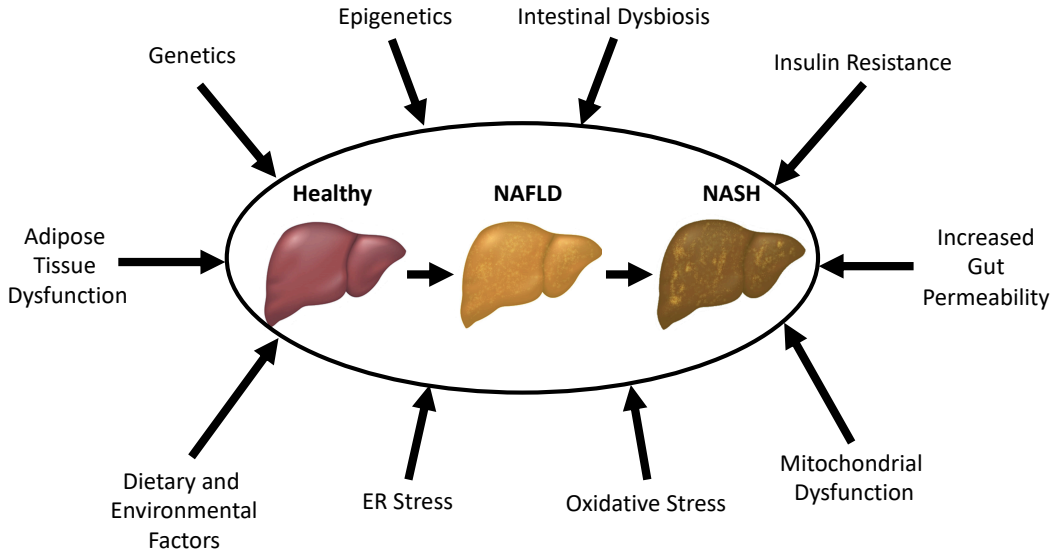

Figure 1.11 Multiple hit hypothesis of NAFLD progression.

The development and progression of NAFLD is mediated by several genetic, environmental, and intracellular factors. For example, genetic and epigenetic factors may make a person more susceptible to environmental insults such as diet which leads to insulin resistance, ER stress, and mitochondrial dysfunction which all contribute toward the inflammatory state and progression toward NASH. Other factors that may contribute to NAFLD progression are intestinal dysbiosis, increased gut permeability, oxidative stress, and adipose tissue dysfunction. Not all factors are necessary or sufficient for NAFLD progression to NASH making the progression extremely complex and unique to each patient.

The mechanisms that facilitate disease progression are still under investigation. Our understanding of the disease mechanisms has been severely limited by the lack of pre-clinical models that reflect the full spectrum of human disease. Mouse models using high-fat diets, methionine and choline deficient diets, and atherogenic diets have failed to reproduce the histology, pathophysiology, and/or the metabolic abnormalities seen in humans [64]. In order to fully elucidate the mechanisms of human disease and develop

drugs for its treatment, better pre-clinical models must be developed. In the following section, I will describe in general terms the current understanding of the pathogenesis of NAFLD from simple steatosis through NASH and fibrosis.

1.3.2.1 Lipid Accumulation

Hepatic fat accumulation or steatosis occurs when lipid acquisition outpaces lipid disposal in the liver. Steatosis can be divided into two categories: microsteatosis and macrosteatosis. Microsteatosis is defined as the accumulation of innumerable small lipid droplets with the nucleus remaining in its original location. Macrosteatosis occurs when larger lipid droplets displace the nucleus from its original location. In NAFLD, steatosis is generally centred around zone 3 hepatocytes, but this distribution can change as the disease progresses [56].

Several pathways are involved in the delicate balance of hepatic lipid homeostasis including uptake of circulating fatty acids, de novo lipogenesis, fatty acid oxidation, and VLDL secretion. Altered functionality of any of these pathways results in steatosis. Environmental factors such as diet and type 2 diabetes may also contribute to the development and severity of steatosis through insulin resistance as well as increased flux of fatty acids or their precursors to the liver.

Fatty acids used to synthesize hepatic triglycerides are derived from three sources. Approximately 15% of hepatic fat is derived from the diet, 26% is derived from de novo lipogenesis, and 59% is derived from adipose tissue lipolysis [65]. Therefore, the majority of hepatic TG is derived from fatty acids acquired from circulation. Hepatic lipid uptake is dependent on several fatty acid transporters (FATP and CD36) while their transport within the cell relies upon the availability of FABP1. All three of these transport proteins are upregulated in the early stages of NAFLD. This likely represents a compensatory mechanism to prevent hypertriglyceridemia as well as lipotoxicity within the liver. The transport proteins facilitate the incorporation of cytotoxic free fatty acids into metabolically inert triglycerides to prevent cellular injury. However, in later stages of the disease, this increased lipid flux could contribute to lipotoxicity and disease progression [64].

De novo lipogenesis is abnormally elevated in NAFLD patients and these patients fail to properly nutritionally control this process. This failure to regulate de novo lipogenesis is a central feature contributing to steatosis in NAFLD. NAFLD patients have higher expression of the master transcription factors that control de novo lipogenesis: SREBP1c and ChREBP [66]. This leads to the upregulation of downstream targets such as FAS and ACC and increased de novo lipogenesis of saturated fatty acids. The increased production of these cytotoxic fatty acids may contribute to lipotoxicity and disease progression.

One mechanism for disposal of excess fatty acids in the liver is oxidation. Fatty acid oxidation is increased in NAFLD patients as an adaptive mechanism to reduce the lipid burden in hepatocytes and the risk of lipotoxicity. This oxidation generally occurs in the mitochondria or peroxisomes to generate energy from the stored lipids. However, when there is a lipid overload within the hepatocyte, ω-oxidation by cytochromes is activated. ω-oxidation generates significantly more reactive oxygen species (ROS), oxidative stress, and toxic by-products. The increased fatty acid oxidation in NAFLD leads to high production of ROS that overwhelms the antioxidant capacity of the cell. This leads to oxidative stress and mitochondrial dysfunction which are precursors to lipotoxicity. So, although fatty acid oxidation is upregulated in NAFLD as a protective mechanism, this upregulation may ultimately contribute to disease progression [67, 68].

The major pathway for exporting fatty acids from the liver is VLDL secretion. These water-soluble particles are packed with triglycerides, cholesterol, phospholipids, and apolipoproteins before being secreted into circulation. In early stages of NAFLD, VLDL secretion is increased in an attempt to decrease the lipid burden of hepatocytes [69, 70]. However, as the disease progresses, ER stress results in the degradation of APOB-100 and reduced secretion of VLDLs [71, 72]. This diminished export results in steatosis, lipotoxicity, and disease progression.

1.3.2.2 Selective Insulin Resistance

Insulin resistance is a key mediator of hepatic fat accumulation, especially in NAFLD patients with metabolic syndrome. Insulin resistance occurs when insulin is unable to suppress gluconeogenesis but retains its ability to stimulate lipid production and secretion

[73-75]. The increased de novo production of glucose further stimulates de novo lipogenesis and metabolic dysfunction. The ultimate result of insulin resistance is hyperglycaemia and hypertriglyceridemia. Insulin exerts its effects on de novo lipogenesis by upregulating SREBP1c. SREBP1c then further exacerbates insulin resistance by increasing the accumulation of harmful lipid species such as diglycerides. Cytoplasmic diglycerides induce the expression and membrane translocation of protein kinase C which inhibits insulin receptor signalling [76-78]. The conversion of diglycerides to triglycerides prevents their deleterious effect on insulin signalling. Therefore, hepatocytes may be using steatosis as a protective mechanism against accumulation of diglycerides. An overview of this process can be found in Figure 1.12.

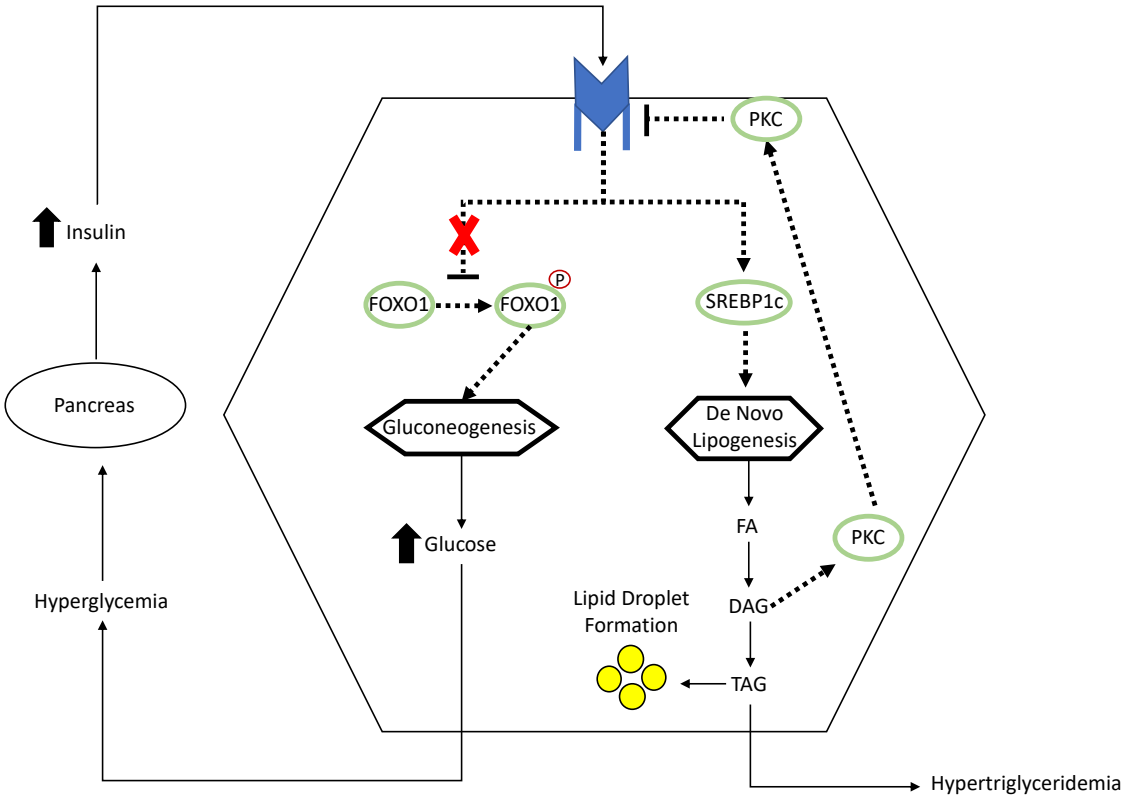

Figure 1.12 Selective insulin resistance.

Hepatic insulin resistance occurs when insulin loses the ability to suppress gluconeogenesis but retains its ability to induce de novo lipogenesis. In insulin resistant states, AKT signalling is disrupted so insulin loses the ability to suppress the activation of FOXO1 leading to constitutively active gluconeogenesis. This high hepatic glucose production leads to hyperglycaemia which in turn causes increased production of insulin in the pancreas and hyperinsulinemia. This hyperinsulinemia leads to upregulation of SREBP1c and downstream de novo lipogenesis genes which causes hepatic steatosis and hypertriglyceridemia. In

addition, high levels of de novo lipogenesis lead to accumulation of toxic intermediates such as diglycerides that further exacerbate insulin resistance through irregular protein kinase C signalling.

1.3.2.3 Lipotoxicity

Lipotoxicity is apoptosis caused by the harmful effects of high concentrations of lipids and lipid derivatives. Lipotoxicity may occur as a result of increased fatty acid oxidation and oxidative stress, alterations to cellular membrane and phospholipid composition, altered cellular cholesterol concentration, disturbed ceramide signalling, or direct toxicity of free fatty acids. These various mechanisms may trigger apoptosis via the intrinsic or extrinsic pathways. The intrinsic pathway of lipoapoptosis is triggered by intracellular ER stress and mitochondrial dysfunction whereas the extrinsic pathway is activated by the binding of death ligands FAS and TRAIL to receptors on the cell membrane. A schematic of the intrinsic lipotoxicity pathway can be found in Figure 1.13. Following activation of apoptosis, both the intrinsic and extrinsic pathways lead to mitochondrial permeabilization and activation of effector caspases. This lipoapoptosis is often accompanied by the release of proinflammatory cytokines which contribute to the progression from NAFLD to NASH [64, 79, 80].

The primary mediator of the intrinsic lipoapoptosis pathway is JNK activation. Activated JNK interacts with CHOP to upregulate the expression of PUMA [81]. PUMA and JNK then activate the pro-apoptotic protein BAX which is necessary for mitochondrial permeabilization. This permeabilization results in the release of cytochrome c and activation of the apoptosis effector caspases 3 and 7 [79]. JNK can be activated through several different cell stress pathways including ER stress, oxidative stress, and mitochondrial dysfunction. ER stress pathways are adaptive mechanisms to overcome stress stimuli such as accumulation of misfolded proteins or toxic lipid derivatives. However, under conditions of prolonged ER stress, apoptosis is triggered [80]. Oxidative stress occurs when the generation of ROS overcomes the antioxidant defences of the cell. This type of cell stress is caused in NASH by enhanced β- and ω-oxidation and CYP2E1 induction. This oxidative stress as well as increased mitochondrial permeability as a result of lipid accumulation results in mitochondrial dysfunction.

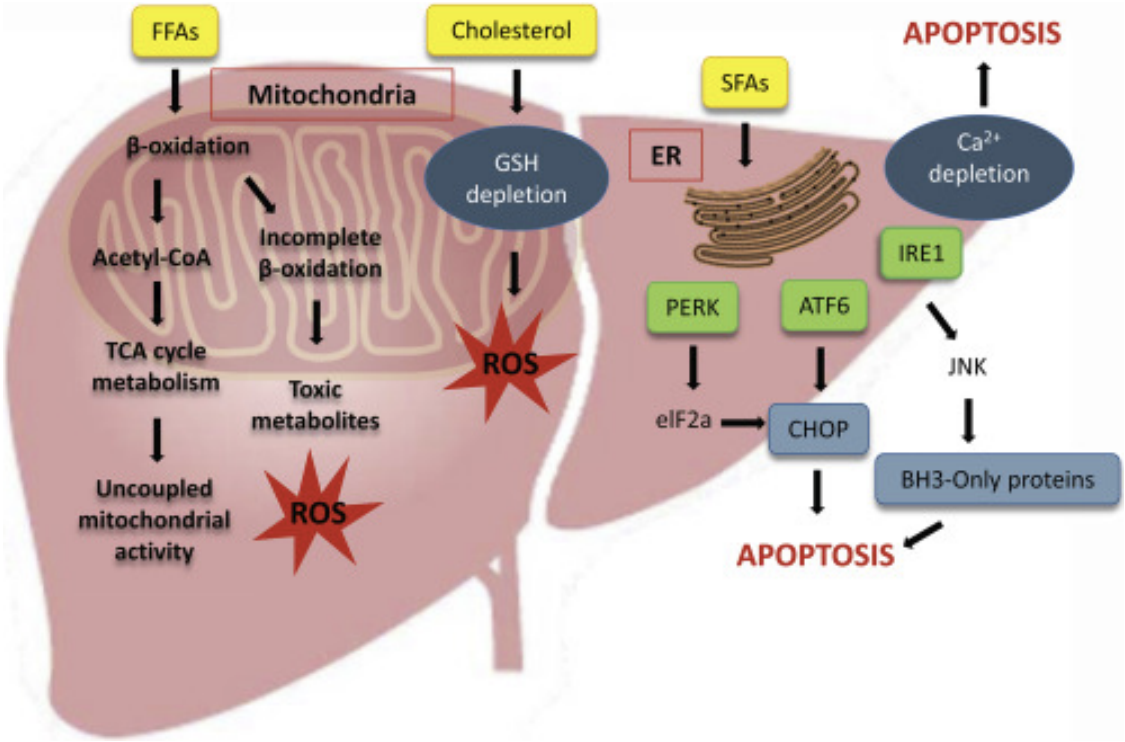

Figure 1.13 Mechanisms of lipotoxicity.

Lipids may induce toxicity through a variety of mechanisms including mitochondrial dysfunction, oxidative stress, and ER stress which ultimately lead to the activation of apoptosis. Excessive levels of free fatty acids may lead to mitochondrial dysfunction by uncoupling the TCA cycle from mitochondrial respiration or cause oxidative stress through incomplete β-oxidation. Cholesterol contributes to oxidative stress by depleting the hepatocyte's glutathione levels. SFAs cause ER stress by activating PERK, ATF6, and IRE1 which cause apoptosis through the activation of CHOP and JNK as well as disruption to calcium homeostasis. (Adapted from Svegliati-Baroni et al. Free Radical Biology and Medicine 2019) [82]

Several lipid species are considered lipotoxic including free fatty acids, especially saturated fatty acids, lysophosphatidyl choline (LPC), ceramides, and free cholesterol. Saturated fatty acids are increased in the hepatocytes of NAFLD patients due to increased fatty acid uptake and de novo lipogenesis. These fatty acids lead to lipotoxicity through several different mechanisms. Saturated fatty acids such as palmitic acid lead to the synthesis and accumulation of other cytotoxic lipid by-products such as ceramides and LPC [83]. Additionally, palmitic acid increases the synthesis of saturated phospholipids and glycerolipids [84-86]. When these saturated lipids are incorporated into plasma membranes, it triggers the activation of ER stress pathways which leads to JNK activation and ultimately lipoapoptosis. LPC is the phospholipid mediator of saturated fatty acid lipotoxicity in NASH. The production of this lipid causes a depletion of phosphatidyl

choline from membranes which results in loss of membrane integrity, extracellular release of lipotoxic lipids, inflammation, and apoptosis. Additionally, the accumulation of LPC independently induces ER stress [87].

Ceramides are cell membrane components that are biologically active in oxidative stress and inflammation. Accumulation of ceramides can induce ER stress by perturbing calcium homeostasis as well as promoting the release of ROS through its interaction with TNFα. The accumulation of ceramides is associated with insulin resistance and correlates with increased concentrations of proinflammatory cytokines [88-90].

Free cholesterol concentrations increase progressively from NAFLD to NASH. This cholesterol accumulates in the mitochondria and causes increased permeability and mitochondrial dysfunction. Additionally, cholesterol stimulates the activation of Kupffer cells and HSCs causing progression of disease toward inflammation and fibrosis [79, 81].

1.3.2.4 Inflammation and Fibrosis

Hepatocyte injury from lipotoxicity causes the release of damage associated molecular patterns (DAMPs) and proinflammatory cytokines that recruit and activate Kupffer cells and HSCs which cause disease progression from simple steatosis through inflammation and fibrosis. There are three inflammatory pathways in NAFLD: the JNK-AP-1 pathway, the NF-κB pathway, and the inflammasome pathway. Each of these pathways results in the production of proinflammatory cytokines such as IL1β, IL6, and TNFα. The accumulation of these proinflammatory cytokines leads to histological changes such as Kupffer cell activation, neutrophil chemotaxis, activation of HSCs, and production of Mallory bodies (inclusions of damaged cytoskeletal components in hepatocytes) [63]. The chronic inflammatory state of NASH increases the risk for fibrosis. The rate of fibrosis progression is higher in patients with obesity, diabetes, hypertension, and a high degree of inflammation found on biopsy [56]. Both inflammation and fibrosis have a degree of reversibility with weight loss and exercise [58]. However, some patients with severe inflammation and fibrosis will go on to develop cirrhosis and hepatic decompensation.

1.3.3 Genetics of NAFLD

Until recently, NAFLD was considered to merely be a consequence of the metabolic syndrome. However, recent studies have demonstrated a strong genetic component to the disease. The prevalence of hepatic steatosis varies among different ethnic groups. Hispanic Americans have the highest prevalence of hepatic steatosis at 45%, followed by European Americans at 33% and African Americans at 24% [91]. Severity of the disease also mirrors these ethnic differences with Hispanic Americans being at highest risk for severe NASH while African Americans appear to be protected from this disease. These ethnic differences are independent of adiposity, insulin resistance, and socioeconomic factors. Given these observations, the heritability of NAFLD is estimated to be approximately 52% while fibrosis is approximately 50% heritable among NAFLD patients. 78% of this heritability is shared between steatosis and fibrosis [92].

In order to identify candidate genes that are associated with the development and progression of NAFLD, several genome wide association studies (GWAS) were undertaken [91, 93, 94]. These GWAS identified single nucleotide polymorphisms (SNP), single base pair changes that are common in more than 1% of the population [95]. The first GWAS for NAFLD was performed by Romeo et al. on the Dallas Heart Study cohort. They identified a SNP (rs738409) in the gene coding for patatin-like phospholipase domain-containing protein 3 (*PNPLA3*). This SNP results in a missense mutation that replaces methionine for isoleucine at amino acid position 148. This variant has the strongest association with NAFLD and has been shown to increase risk for developing the full spectrum of NAFLD from simple steatosis through cirrhosis [91, 92, 95].

Several other SNPs have also been identified, though with lower effect sizes than *PNPLA3*, including *TM6SF2, MBOAT7*, and *GCKR*. The E167K (rs58542926) variant in TM6SF2 is the second most commonly associated variant with NAFLD. TM6SF2 is involved in the enrichment of triglycerides to APOB-100 for VLDL secretion and the E167K variant is loss of function. This results in retention of hepatic triglycerides which simultaneously increases a patient's risk for liver disease and reduces the risk of cardiovascular disease. The rs641738 SNP in *MBOAT7* increases the risk of NAFLD, inflammation, fibrosis, and hepatocellular carcinoma. The P446L (rs1260326) SNP in

GCKR is a loss of function mutation that results in constitutively active hepatic glucose uptake which stimulates de novo lipogenesis. Additionally, SNPs have been identified in several genes linked to insulin resistance (*ENPP1, IRS1*, and *TRIB1*). Some SNPs have also been found to be protective against NAFLD including *HSD17B13, LPIN1, UCP2, MnSOD*, and *MERTK* [92, 96]. A summary of SNPs that impact the NAFLD phenotype can be found in Figure 1.14. The effect size of the deleterious genetic variants is increased with increasing BMI indicating that genetic factors interact with environmental factors to determine disease severity [92, 97]. These genetic variants are also shared between other forms of chronic liver disease such as alcoholic fatty liver disease and viral hepatitis. This indicates that the steatosis in these diseases may share common pathogenic mechanisms.

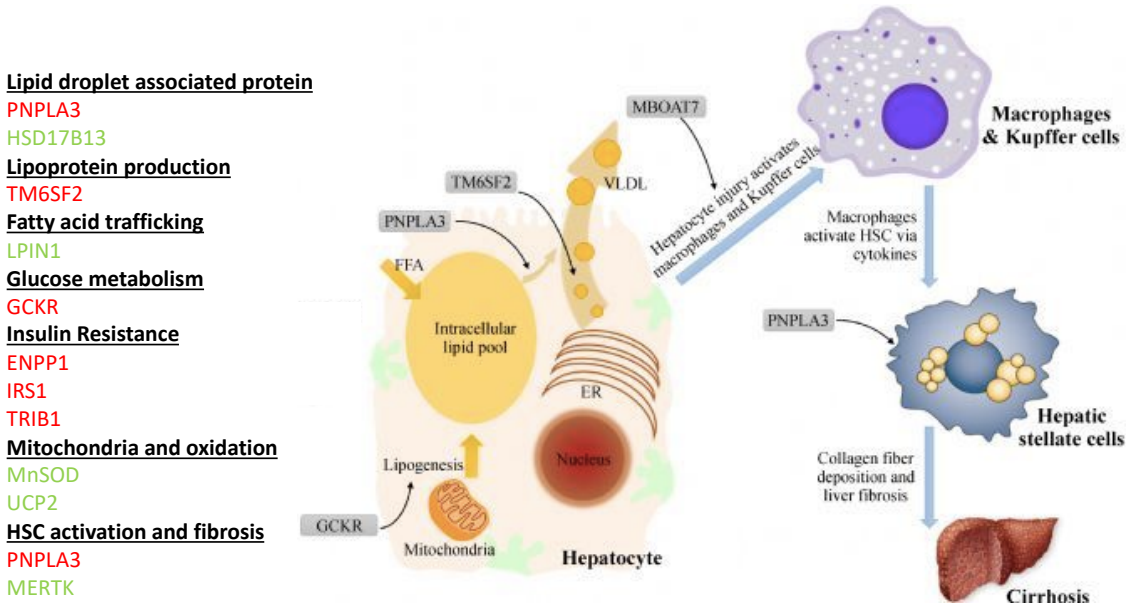

Figure 1.14 SNPs that modulate NAFLD phenotype.

List of genetic variants that have an impact on NAFLD progression and the cellular processes that they affect. Genes shown in red represent genetic variants that lead to worsening disease phenotype while genes shown in green represent genetic variants that are protective against NAFLD progression. The four genetic variants with the strongest effect sizes are *GCKR, MBOAT7, TM6SF2*, and *PNPLA3*. GCKR causes constitutively active de novo lipogenesis which causes hepatic steatosis. MBOAT7 plays a role in activation of the inflammatory response though its function remains to be fully elucidated. TM6SF2 causes hepatic steatosis by interfering with the secretion of VLDL. Finally, PNPLA3 plays a role in preventing the mobility of fatty acids from lipid droplets as well as activating HSCs to their pro-fibrotic state. (Figure adapted from Danford et al. The Journal of Biomedical Research 2018) [98]

In addition to the identified genetic risk of NAFLD, there is a growing appreciation for the role of epigenetics in disease progression. There is a general tendency toward demethylation of hepatic DNA in patients with NAFLD [99, 100]. The epigenetic regulation of genes involved in methylation, inflammation, and fibrogenesis is stage dependent. For example, *PPAR* genes are hypermethylated while fibrotic genes like $TGF\beta$ and $PDGF\alpha$ are hypomethylated in NAFLD patients with progressive disease [99]. This is evidence for the hypothesis that epigenetics may play a key role in the progression from NAFLD to NASH. Additionally, micro RNAs have also been linked to NAFLD pathogenesis. The most common micro RNA in the liver is miR-122 accounting for approximately 70% of hepatic mRNAs. It is downregulated in NAFLD patients leading to increased lipogenesis and activation of fibrosis through induction of HIF1α, MAPK1, and NIK [101-104]. However, it remains unclear if these epigenetic changes are causative agents in the disease progression from NAFLD to NASH or if they are mere by products of this progression.

Despite the growing appreciation for the genetic and epigenetic contributions to NAFLD pathogenesis the mechanistic role of the identified genetic variants remains largely obscure. This is due in large part to the lack of an appropriate pre-clinical model that recapitulates the full spectrum of human disease. This is especially true for the genetic variant with the highest effect size: *PNPLA3*. Despite the identification of the I148M variant over a decade ago, the mechanistic role of this protein both in healthy livers and NAFLD has yet to be fully elucidated. In the following section, I will summarize the state of knowledge about this gene and its contribution to NAFLD pathogenesis

1.4 PNPLA3 Gene

1.4.1 PNPLA3 Gene

PNPLA3, also known as adiponutrin or calcium-independent phospholipase A2ε, is a 52.8 kDa protein with 481 amino acids. The gene sequence is 9 exons long (2805 base pairs) and is located on the long arm of chromosome 22 at position 13.31 [105]. This protein is a member of the PNPLA family which has 9 members that demonstrate non-specific lipid acyl hydrolase activity [106-108]. The members of this family are soluble

proteins that are tightly associated with membranes and lipid droplets [109]. PNPLA3 shares the closest homology to PNPLA2 also known as ATGL [110].

1.4.1.1 Expression

In humans, PNPLA3 is expressed in several tissues including the liver, retina, skin, and adipose tissue. PNPLA3 is expressed at its highest levels in the liver and retina. Within the liver, PNPLA3 expression levels are higher in HSCs than hepatocytes [111-115].

1.4.1.2 Structure

The structure of PNPLA3 has yet to be fully resolved. However, several aspects of its structure have been elucidated using secondary structure prediction and homology modelling. The homology modelling of PNPLA3 resembles the crystal structure of the heartleaf horse nettle patatin. This includes a consensus serine lipase motif (Gly-X-Ser-X-Gly) that lies between a beta strand and an alpha helix. Like other members of the patatin family, PNPLA3 contains a catalytic dyad (Ser-Asp) as opposed to a catalytic triad that is common in most lipases. The patatin fold brings the catalytic serine (Ser47) in close proximity to Asp166 at the edge of a hydrophobic substrate binding groove to create the catalytic dyad [109, 116]. Controversy remains surrounding the presence of transmembrane domains within the protein. Studies of the secondary structure predicted 4 transmembrane domains that strongly associated with the ER and lipid droplets; however, tertiary structure modelling indicates that all alpha helices form a globular structure that does not span the membrane [105, 109]. Regardless, PNPLA3 is tightly associated with membranes, especially lipid droplet membranes.

1.4.1.3 Function

Considerable debate still remains surrounding the function of PNPLA3. *In vitro* studies have shown that the enzyme has triglyceride lipase activity against glycerolipids (triglycerides, diglycerides, and monoglycerides) as well as acylglycerol transacylase activity, and lysophosphatidic acid acyl transferase (LPAAT) activity in hepatocytes [117-121]. It has also been hypothesized that PNPLA3 may play a role in VLDL secretion [122]. Given this diverse range of functions, it has been difficult to ascertain the role of PNPLA3 in lipid metabolism. The LPAAT activity suggests a role in lipid anabolism while the lipase activity suggests a role in lipid catabolism. However, when the wild type

enzyme (derived from the reference allele) is overexpressed in mouse or human cells, there is no appreciable change in the intracellular triglyceride content [109, 119, 123]. Therefore, overexpression of PNPLA3 leads to neither a net increase nor decrease in triglyceride content which calls into question its role in either lipid anabolism or catabolism.

Recently, a new hypothesis has emerged to resolve these seemingly opposing results. Since PNPLA3 demonstrates both triglyceride lipase activity and acyl transferase activities, it has been hypothesized that PNPLA3 plays an active role in lipid droplet remodelling. Specifically, PNPLA3 seems to catalyse the transfer of polyunsaturated fatty acids (PUFA) from triglycerides to phospholipids [124-127]. In this way, PNPLA3 does not alter the total amount of triglycerides in the cell, rather the enzyme plays a role in maintaining the dynamic properties of lipid droplets by selective sorting of fatty acids between storage and membrane lipids. By sorting fatty acids between triglycerides and phospholipids, PNPLA3 plays a crucial role in determining the availability of fatty acids throughout the cell for signalling and other functional purposes. Additionally, by remodelling the phospholipid composition of the lipid droplet membrane, PNPLA3 may contribute to the recruitment of other lipid droplet proteins to further remodel the lipid content of the cell [128].

Additionally, in HSCs, PNPLA3 shows retinyl palmitate hydrolase activity which facilitates the release of retinol from HSCs [112, 129]. Due to its role in retinol homeostasis in these cells, it has been hypothesized that PNPLA3 contributes to HSC activation. This role has been confirmed by the upregulation of PNPLA3 expression in response to pro-fibrotic stimuli such as TGFβ and *in vitro* culture techniques. Overexpression of the reference PNPLA3 enzyme in HSCs revealed that upregulation of PNPLA3 leads to a reduced secretion of pro-fibrotic enzymes such as MMP2, TIMP1, and TIMP2 in a process that is mediated by retinoids [113, 114]. Therefore, PNPLA3-mediated retinol release is a protective mechanism against fibrosis.

1.4.1.4 Transcriptional and Nutritional Regulation

PNPLA3 is tightly controlled by nutritional status. In the fasted state, PNPLA3 expression is downregulated to near zero levels. Upon refeeding, especially a high carbohydrate diet,

PNPLA3 is quickly upregulated by insulin and glucose concentrations [111, 130, 131]. The pattern of induction of PNPLA3 is similar to proteins involved in glycolysis and lipogenesis. In humans, *PNPLA3* is transcriptionally controlled by SREBP1c whereas in mice, it is under the control of both SREBP1c and ChREBP [130]. Therefore, the upregulation of PNPLA3 in humans in response to a high carbohydrate diet is mediated entirely by insulin whereas the upregulation in mice is dependent on both glucose and insulin.

In addition to its transcriptional regulation, PNPLA3 is also regulated post-translationally by fatty acids. Fatty acids, especially saturated and monounsaturated fatty acids, inhibit the degradation of PNPLA3 protein. Addition of oleic acid extends the half-life of the enzyme by over four-fold. Since the products of SREBP1c-stimulated de novo lipogenesis are saturated and monounsaturated fatty acids, PNPLA3 appears to be controlled by a feed forward loop which allows for efficient upregulation of protein levels following feeding [111]. SREBP1c upregulates the transcription of *PNPLA3* and de novo lipogenesis genes, the products of this de novo lipogenesis then retard the degradation of PNPLA3 protein which causes the efficient postprandial rebound in PNPLA3 expression. This process is summarized in Figure 1.15.

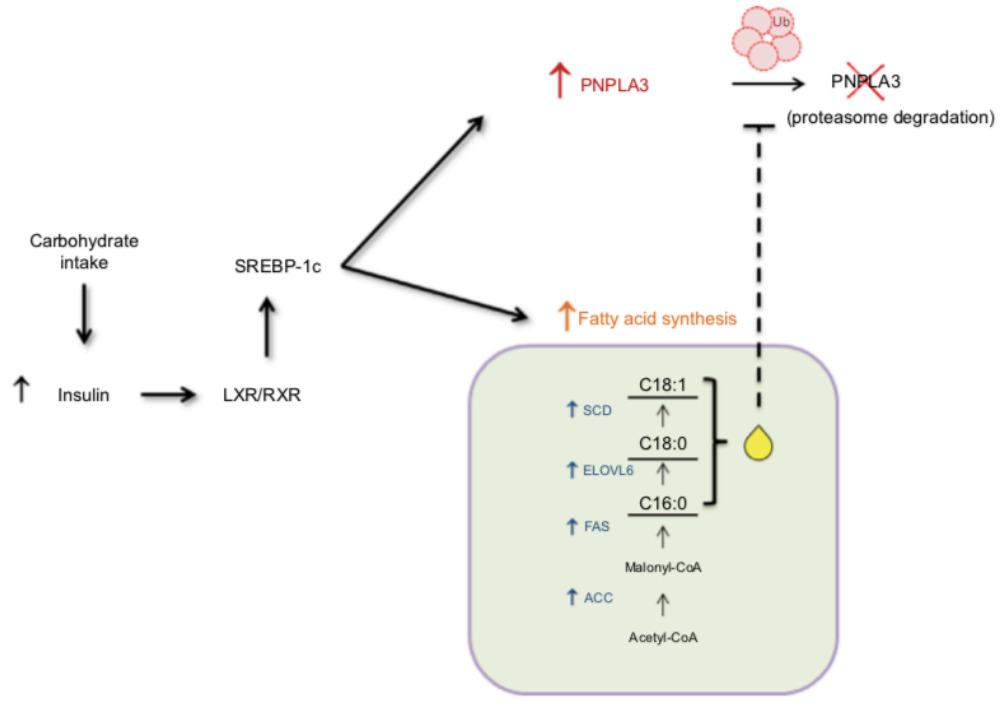

Figure 1.15 Nutritional regulation of PNPLA3.

Carbohydrate consumption leads to increased secretion of insulin from the pancreas. The binding of insulin to its receptor on the surface of hepatocytes leads to the activation of LXR/RXR which upregulates the expression of SREBP1c. SREBP1c then induces the expression of both PNPLA3 and de novo lipogenesis genes. The end-products of de novo lipogenesis, long chain fatty acids, prevent the proteasomal degradation of PNPLA3 thus creating a feed forward loop that allows for efficient post-prandial rebound of PNPLA3 expression. (Figure adapted from Bruschi et al. Hepatic Medicine: Evidence and Research 2017) [110]

PNPLA3 expression is not upregulated by a high fat diet despite induction of hepatic steatosis. In fact, the inclusion of lipids in a high carbohydrate diet fed to mice prevented the carbohydrate mediated induction of PNPLA3 in a dose dependent manner. The expression of PNPLA3 was especially reduced by the addition of PUFAs regardless of glucose concentration. These PUFAs reduce the nuclear abundance of SREBP1c by suppressing its gene transcription and enhancing mRNA decay [131]. The reduced expression of SREBP1c prevents the postprandial upregulation of *PNPLA3* transcription. Thus, PNPLA3 expression is not induced by all lipogenic stimuli and the presence of this enzyme is not necessary for the formation of steatosis.

In HSCs, *PNPLA3* is transcriptionally controlled by intracellular retinoid concentrations. When there is a high concentration of retinol, PNPLA3 expression is downregulated to allow the storage of retinyl palmitate in lipid droplets. When cells are deficient in retinol, PNPLA3 is upregulated to catalyse the hydrolysis of retinyl palmitate and the release of retinol. This reduction in lipid droplet size was also accomplished by supplementing HSCs with insulin indicating that the SREBP1c control of PNPLA3 is intact in HSCs [112]. In addition to nutritional regulation, PNPLA3 is also regulated by pro-fibrotic stimuli such as TGFβ [113].

1.4.2 I148M Variant

In 2009, Romeo et al. were the first to identify the rs738409 SNP in *PNPLA3* which causes a substitution of guanine for cytosine resulting in an amino acid change from isoleucine to methionine at position 148 (I148M) [91]. The authors found that the I148M variant (PNPLA3 risk allele) was strongly associated with high intrahepatic fat content ($P = 5.9 \times 10^{-10}$) as well as liver inflammation ($P = 3.7 \times 10^{-4}$). In this study, the minor allele frequency (MAF) varied between ethnic groups; Hispanics had the highest frequency of the risk allele (MAF = 0.49), followed by European Americans (MAF =

0.23), and African Americans had the lowest frequency of the risk allele (MAF = 0.17). The frequency of the risk allele in these ethnic populations mirrors the prevalence of hepatic steatosis in these groups which is estimated to be 45% in Hispanics, 33% in European Americans, and 24% in African Americans [54]. Since its initial discovery, the I148M variant has been associated with higher risk of developing the full spectrum of NAFLD from steatosis through inflammation, fibrosis, cirrhosis, and hepatocellular carcinoma [132-135]. Most recently, the correlation between the rs738409 SNP and NAFLD was confirmed in another GWAS ($P = 1.70 \times 10^{-20}$). This study found that the rs738409 SNP was correlated with both pediatric ($P = 9.92 \times 10^{-6}$) and adult ($P = 9.73 \times 10^{-15}$) NAFLD as well as increased disease severity measured by the NAFLD Activity Score ($P = 3.94 \times 10^{-8}$) [136]. The I148M variant confers increased risk of NAFLD development and progression independent of metabolic dysfunction [137, 138]. In addition to conferring increased risk of NAFLD, PNPLA3 I148M has been linked to worsening phenotypes in many other chronic liver diseases such as alcoholic fatty liver disease (ALD), hepatitis C virus (HCV), hepatitis B virus (HBV), hereditary hemochromatosis, and Wilson's Disease [139-144]. It should be noted that NAFLD and other chronic liver diseases are complex, multifactorial diseases that cannot be accounted for using standard inheritance models. Patients that are heterozygous for the PNPLA3 risk allele are at elevated risk of developing NAFLD and the addition of a second risk allele increases this probability in an additive manner [132]. However, the presence of this risk allele does not invariably lead to disease onset or progression; a patient's genetic predisposition interacts with environmental factors to determine disease severity.

1.4.2.1 Clinical Implications

The increased clinical risk of the I148M variant has been robustly confirmed in over 50 clinical studies [145]. Patients with the I148M variant tend to have 73% higher lipid content, 3.2 fold greater risk of high necro-inflammatory scores, 3.5 fold higher risk of NASH, 3.2 fold greater risk of developing fibrosis, and 3 fold increased risk of developing hepatocellular carcinoma [56, 57, 95, 133]. PNPLA3 I148M is associated with a higher risk of more severe NAFLD in several ethnicities (Hispanics, Argentinians, Germans, Italians, Japanese, Taiwanese, African Americans, Americans, Malaysians, etc.) [91, 137, 146-152]. The frequency of the risk allele in different ethnicities mirrors the relative

prevalence of NAFLD in those ancestry groups indicating that *PNPLA3* genotype is the major genetic determinant of NAFLD risk. The global prevalence of the PNPLA3 I148M variant is shown in Figure 1.16. It is estimated that the I148M variant accounts for 72% of the heritability of NAFLD [91]. The I148M variant is associated with increased fat content without altered glucose homeostasis or lipoprotein metabolism indicating that hepatic steatosis and insulin resistance are not inextricably linked. Additionally, the variant increases the risk of developing NASH and fibrosis independent of steatosis severity indicating that PNPLA3 I148M may differentially affect hepatic cell types involved in lipid metabolism and fibrosis [132, 145].

However, it is important to note that while NAFLD may develop in patients carrying the I148M variant in the absence of metabolic disease or insulin resistance, the presence of these comorbid conditions may exacerbate the NAFLD disease phenotype in these patients. It has been shown that the effect size of the PNPLA3 risk allele is increased in patients with a higher BMI and may be altered by dietary factors [97, 153, 154]. Additionally, studies show that insulin resistance in patients carrying the I148M variant amplifies the risk of developing steatosis. The reverse condition is also true. By improving the insulin response in patients carrying the risk allele, their genetic predisposition can be overcome leading to a reduction in steatosis and related disease factors [155]. Similarly, carriers of the I148M variant respond disproportionately well to lifestyle modification. The same amount of weight loss in these patients resulted in a larger reduction in liver fat and enzymes than non-carriers [156-158]. Thus, additional studies are needed to extricate the relationship between the I148M variant in PNPLA3 and the metabolic syndrome.

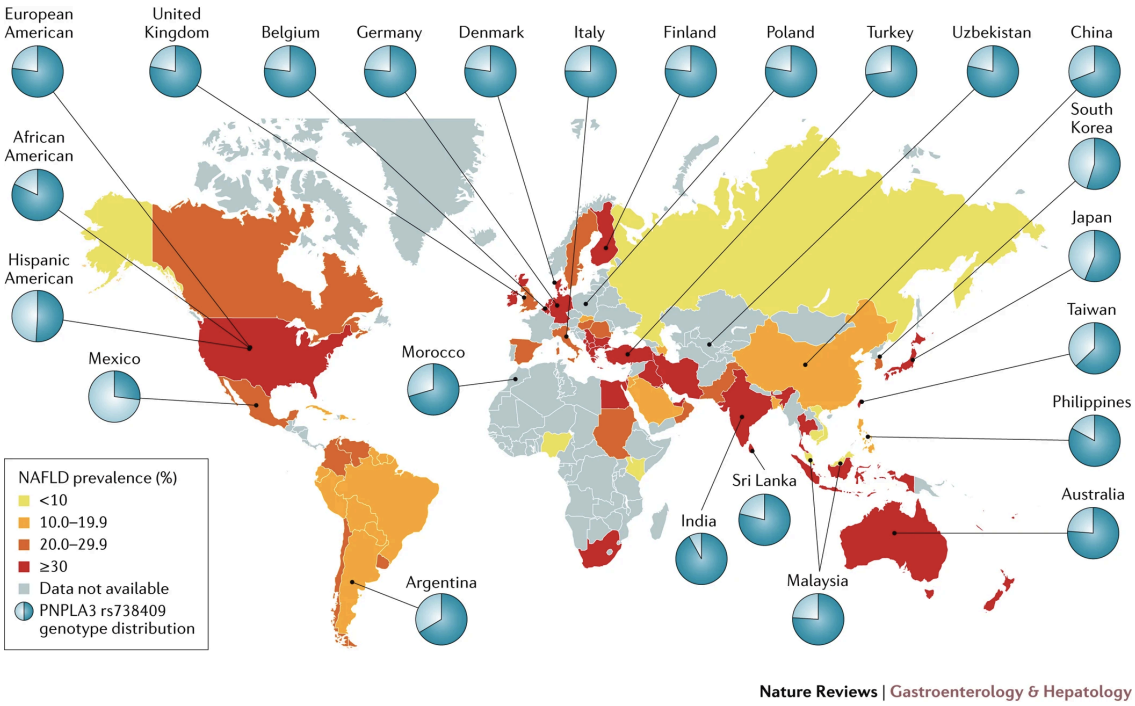

Figure 1.16 Global prevalence of PNPLA3 I148M variant.

The general prevalence of NAFLD in each country is denoted by the colour of the country with red being the most prevalent (>30%) and yellow being the least prevalent (<10%). The prevalence of the I148M variant in each country is denoted as the light blue section of the pie chart. The highest prevalence of the variant is found in Mexican and Hispanic American populations while the lowest prevalence is found in Indian populations. (Figure adapted from Younossi et al. Nature Reviews Gastroenterology & Hepatology 2017) [159]

In addition to NAFLD, PNPLA3 I148M has been linked to a worse prognosis in several other chronic liver diseases indicating a common genetic component to chronic liver diseases. In patients with ALD, the I148M variant has been linked to increased risk of cirrhosis, shorter time to hepatic decompensation, and increased mortality independent of alcohol consumption [139, 142]. In patients with chronic HCV infection, the I148M variant is associated with increased steatosis and fibrosis risk [141]. Whereas in chronic HBV infection, the variant is associated with worsening steatosis and inflammation but not necessarily fibrosis [144]. Additionally, I148M leads to increased steatosis and advanced fibrosis in patients with hereditary hemochromatosis [143]. This variant also confers and increased risk of developing hepatocellular carcinoma regardless of aetiology, although this correlation is stronger in metabolic liver diseases (NAFLD and ALD) than viral diseases [110].

1.4.2.2 Loss of Function or Gain of Function?

Given the fierce debate within the scientific community about the function of the PNPLA3 protein in a healthy liver, the role of the I148M variant in disease remains hotly contested. Several different hypotheses have been postulated to explain the pathogenesis of this variant, each of which is dependent on a different proposed function of the enzyme and the model system used.

There is general consensus that amino acid position 148 falls within the active site of the enzyme and that the substitution of isoleucine for the much bulkier methionine restricts the ability of substrates to bind to the active site of the enzyme [109]. This substitution results in an incomplete loss of lipase activity for both glycerolipids and retinyl esters [109, 112]. Additionally, the I148M variant was shown to decrease VLDL secretion *in vitro* [122]. Therefore, initial studies indicated that the pathogenic variant leads to reduced enzymatic activity which in turn causes steatosis through the reduced hydrolysis and secretion of triglycerides. However, when PNPLA3 was either overexpressed in or knocked-out of mouse livers, there was no effect on hepatic triglyceride content [160, 161]. This simultaneously argued against the role for PNPLA3 in triglyceride hydrolysis and the I148M variant being a loss of function variant.

Given that PNPLA3 also exhibited LPAAT activity *in vitro*, studies were undertaken to investigate the effect of the I148M variant on this function. The I148M variant appeared to increase the LPAAT activity *in vitro* which led to the hypothesis that the I148M variant is a gain of function that causes hepatic steatosis by upregulating triglyceride synthesis [118]. However, these results failed to be replicated in mouse models [162].

When the human or mouse I148M variant was overexpressed in mouse livers, a steatotic phenotype was observed. A similar phenotype was observed in mice overexpressing an obligate loss of function variant where the catalytic serine was mutated (S47A) [123, 163]. Both the I148M and S47A variants caused a significant increase in hepatic triglyceride content, in part, by evading degradation. These variants caused accumulation of I148M PNPLA3 proteins on the surface of lipid droplets which were hypothesized to reduce mobility of lipids from the lipid droplet [162]. Additionally, the high concentration of I148M PNPLA3 on the surface of lipid droplets has the potential to restrict access and

functionality of other lipid droplet proteins by displacing these proteins from the droplet surface or sequestering key cofactors [164]. Given that the full PNPLA3 knock-out produced no phenotype but the I148M variant had the same phenotype as the loss of function S47A variant, it was hypothesized that this variant may be a dominant-negative variant. Additional studies are needed to parse the pathogenic mechanism of action of the I148M variant.

1.4.2.3 Pathogenesis

Given the lack of clarity on the enzymatic activity and effect of the I148M variant, the full pathogenic mechanism of PNPLA3-induced NAFLD has yet to be elucidated. However, from what is known, the pathogenesis of the PNPLA3 I148M variant can be divided into two parts: steatosis and fibrogenesis.

The I148M variant disturbs the lipid homeostasis in hepatocytes by interfering with triglyceride hydrolysis, lipid anabolism, lipid droplet remodelling, and/or lipid mobility and secretion [109, 118, 122, 126, 162, 165]. The mechanism by which this variant interferes with lipid metabolism has yet to be fully elucidated. This metabolic dysfunction leads to triglyceride accumulation and steatosis which characterizes the early stages of NAFLD. It is unknown if this perturbed metabolism results in lipotoxicity or release of proinflammatory mediators that contribute to disease progression.

The progression from simple steatosis to NASH in patients with the I148M variant appears to be mediated by the effect of the variant on HSCs. PNPLA3 is known to play a role in activating HSCs to a profibrotic phenotype and the I148M variant appears to augment this function. HSCs with the I148M variant express significantly higher levels of the proinflammatory cytokines CCL2, CCL5, GM-CSF, CXCL8, CXCL1, and TGFβ which results in increased macrophage chemotaxis. HSCs harbouring the disease allele also have increased proliferation and migration rate, indicating that these cells are more prone to activation from their quiescent state [113, 114]. The role of the I148M variant in HSC activation is many-fold. In its role remodelling lipid droplets, PNPLA3 I148M alters the fatty acid composition of lipid droplets within HSCs leading to reduced retinol stores and increased concentrations of monounsaturated fatty acids and PUFAs. The

accumulation of these fatty acid species leads to the phosphorylation of JNK. Phospho-JNK then inactivates PPARγ. The inactivation of PPARγ has two consequences: decreased synthesis of oxysterols and increased expression of the pro-inflammatory and pro-fibrotic protein AP-1. The reduced synthesis of oxysterols prevents activation of LXR which causes the accumulation of free and total cholesterol in HSCs which further amplifies the pro-fibrotic response in these cells [113, 166]. An overview of this pathogenic mechanism can be found in Figure 1.17. The activation of HSCs by the I148M variant results in a profound inflammatory response as well as progressive fibrosis in patients carrying the disease allele.

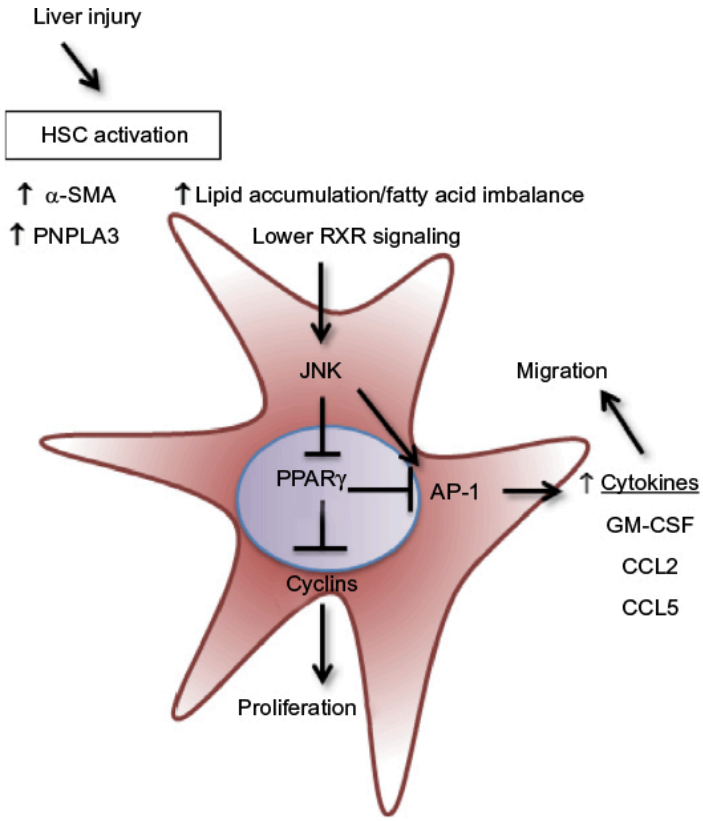

Figure 1.17 Pathogenic mechanism of I148M variant in HSCs.

Activation of HSCs leads to the upregulation of PNPLA3 and α-SMA. This leads to the accumulation of irregular fatty acid moieties such as MUFAs and sterols which lower RXR signalling and activate JNK. JNK simultaneously activates the pro-fibrotic gene AP-1 and inhibits PPARγ signalling. This has the downstream effect of increasing proliferation, migration, and secretion of pro-inflammatory cytokines such as GM-CSF, CCL2, and CCL5. (Figure adapted from Bruschi et al. Hepatic Medicine: Evidence and Research 2017) [110]

1.4.3 Limitations of Mouse Models for PNPLA3 Pathogenesis

The study of PNPLA3 and its effect on NAFLD has been extremely hindered by the lack of an appropriate human model of the disease. PNPLA3 function and mechanistic studies on the I148M variant have been performed almost exclusively in mice which has added a significant amount of confusion to the elucidation of this protein's function. We assert that the mouse is an inappropriate model to study PNPLA3 for three reasons. First, the homology between the two proteins is quite low. Second, the expression pattern of PNPLA3 is starkly different between the two species. Third, the transcriptional control of the protein differs between mice and humans indicating that the two proteins may function differently in each species.

1.4.3.1 Low homology

The protein sequence homology between mouse Pnpla3 and human PNPLA3 is approximately 68% [115]. This homology is extremely low when compared to the inter-species homology of other lipases which tend to have homologies greater than 90%. For reference, the sequence homology between human and mouse PNPLA3 is only approximately 20% higher than that sequence homology between human PNPLA3 and human PNPLA2 (46%). The human PNPLA3 protein is 97 amino acids longer than the mouse protein on the C-terminal end. In these additional 97 amino acids, the human protein contains 2 vesicle targeting motifs [167]. Given the low homology and large structural differences between human PNPLA3 and mouse Pnpla3, it cannot be guaranteed that the two proteins have identical functions in the two species. It is important to note that the majority of structural differences between human PNPLA3 and mouse Pnpla3 occur near the C-terminal end while the active site of the enzyme as well as exon 3 where the I148M variant resides remain largely conserved. The PNPLA3 protein sequence homology between humans and mice is shown in Figure 1.18.

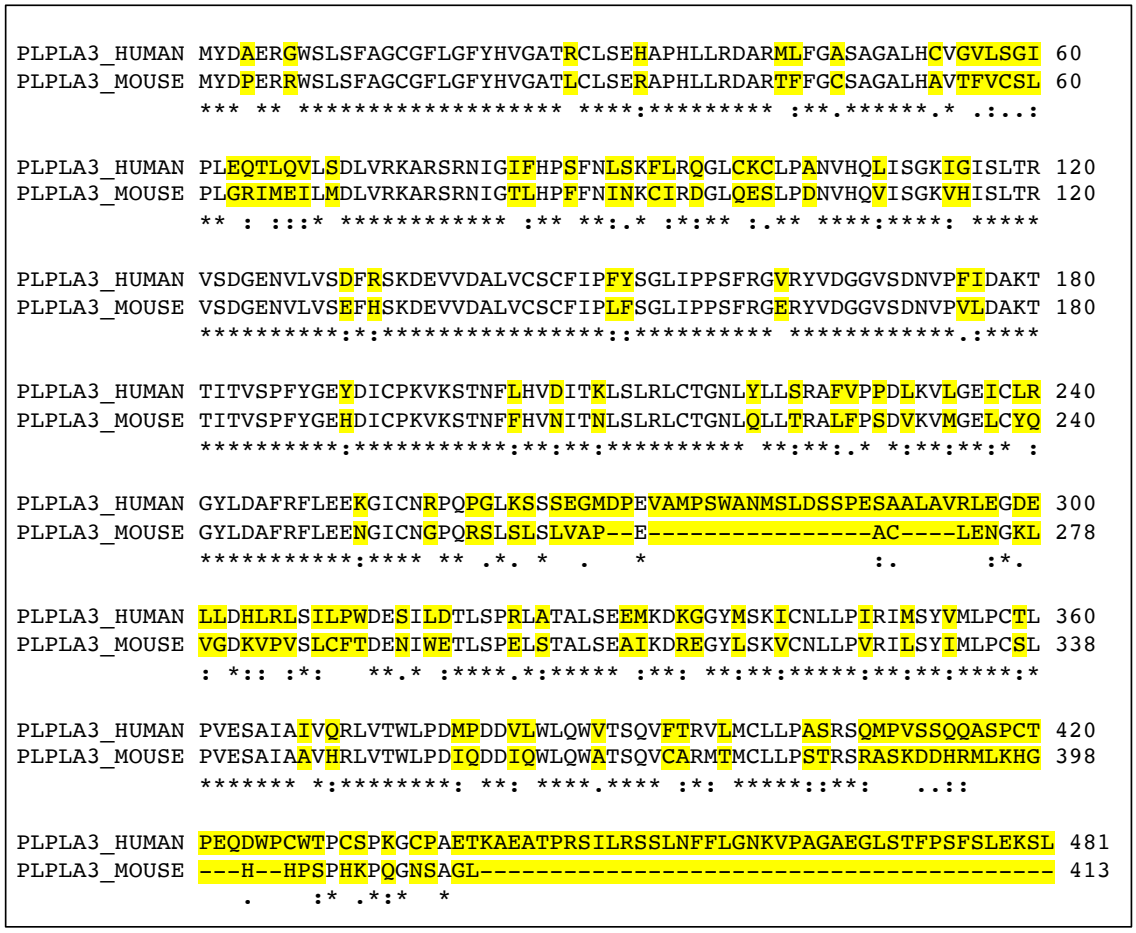

Figure 1.18 PNPLA3 protein sequence homology between humans and mice.

Amino acid sequence alignment of human and mouse PNPLA3 using UniProt. Identical amino acids are shown in white while non-identical amino acids are highlighted in yellow. Highly similar amino acids are denoted with (:), somewhat similar amino acids are denoted with (.), and entirely dissimilar amino acids are denoted with a blank space. The two proteins are extremely different, especially in the C-terminal end. The sequence homology as calculated by UniProt was only 57.6%. This low sequence homology makes the mouse protein an imperfect model to study the function and pathogenicity of human PNPLA3.

1.4.3.2 Differential tissue expression

The expression patterns of PNPLA3 are quite different between humans and mice. In humans, PNPLA3 is expressed at its highest levels in the retina, liver, skin, and adipose tissue. It is expressed at 10-fold higher levels in the liver than adipose tissue. Within the liver, it is expressed at significantly higher levels in HSCs than hepatocytes [111, 112]. Alternately, in mice, Pnpla3 is only expressed at appreciable levels in adipose tissue. The expression of Pnpla3 is 1000-fold higher in adipose tissue than liver tissue [105, 117, 119, 161]. Within mouse livers, Pnpla3 has the opposite expression pattern of humans with

higher expression levels in hepatocytes than HSCs [111]. This differential tissue expression is summarized in Figure 1.19.

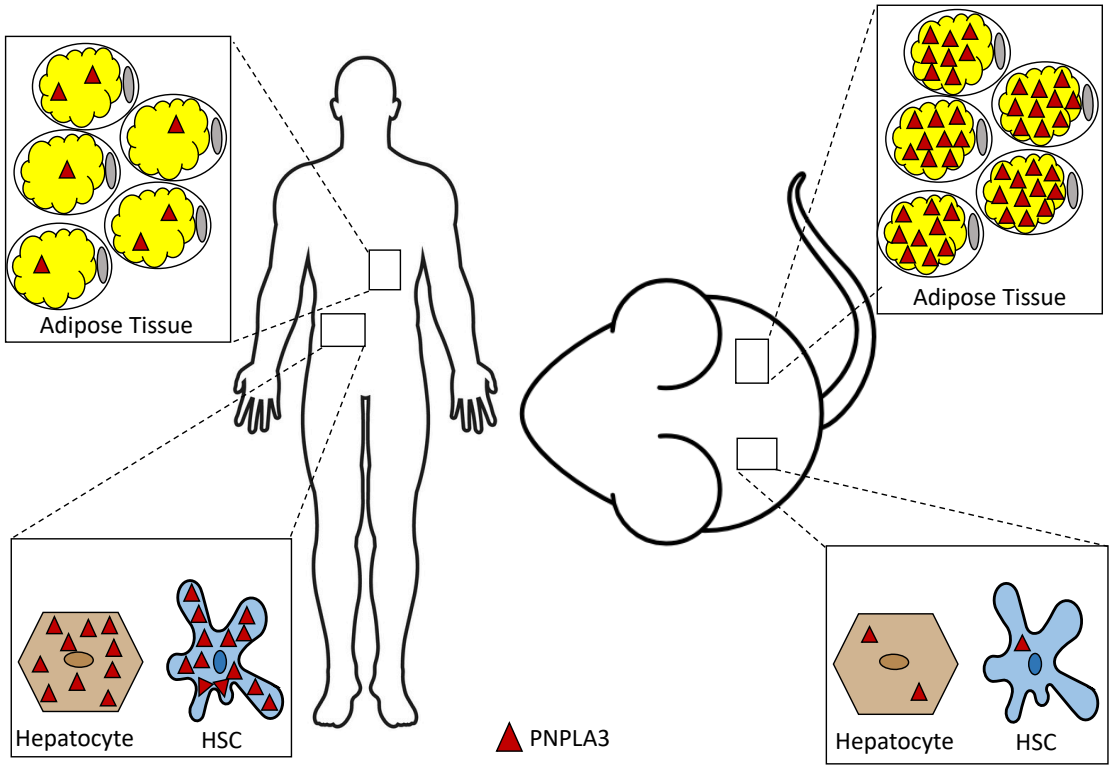

Figure 1.19 Differential tissue expression of PNPLA3 between humans and mice.

In humans, there is a very low expression of PNPLA3 in adipose tissue and the expression of PNPLA3 is 10-fold higher in the liver. In human livers, PNPLA3 is expressed at its highest levels in HSCs followed by hepatocytes. In mice, Pnpla3 is expressed at a 1000-fold higher level in adipose tissue than the liver. In mouse liver, Pnpla3 is expressed at higher levels in hepatocytes than HSCs. This differential tissue expression implies that PNPLA3 may play a different functional role in humans and mice.

Given that knocking-in the I148M variant to adipose tissue in mice has no effect while knocking the variant into liver tissue results in marked hepatic steatosis, it is clear that Pnpla3 exerts its detrimental effects on metabolism in the liver, not adipose tissue [123]. Since Pnpla3 is endogenously expressed at extremely low levels in mouse livers, it is unlikely that this protein plays the same role in hepatocyte metabolism that it does in human livers. Given this differential expression pattern, the murine model may not be the most appropriate model to study the function of PNPLA3 in hepatocytes and the I148M variant's effect on NAFLD disease progression.

1.4.3.3 Differential Transcriptional and Nutritional Control

The transcriptional regulation of human and mouse PNPLA3 is also different. In mice, *Pnpla3* transcription is regulated by both Srebp1c and Chrebp while human *PNPLA3* is controlled by SPREBP1c alone [111, 130]. This differential transcriptional control may indicate that human and mouse PNPLA3 possess different functions that require unique control mechanisms. To support this notion, mice are unable to nutritionally control the transcription of human *PNPLA3* when it is knocked-in to mouse livers [111].

The murine model of Pnpla3 differs in its sequence, expression pattern, and transcriptional control making it an inappropriate model to elucidate the mechanism of PNPLA3. In order to understand the function of human PNPLA3 in healthy livers as well as determine if the I148M variant is a loss of function or gain of function variant, a more appropriate model system is needed. This will require the development of a human model that endogenously expresses PNPLA3 at physiological levels. The human models that currently exist utilize either non-hepatic cell types or hepatocellular carcinoma cell lines. Often these cell lines do not express PNPLA3 at endogenous levels or at all so the protein must be overexpressed which may cause artefacts. I propose the use of human induced pluripotent stem cells and CRISPR/CAS9 to create a non-cancerous, hepatic, human model to study PNPLA3.

1.5 Human Induced Pluripotent Stem Cells

1.5.1 Introduction to hiPSCs

The first induced pluripotent stem cell (iPSC) line was established in 2006 by Takahashi and Yamanaka. They found that the overexpression of just four transcription factors, Oct4, Sox2, Klf4, and Cmyc, reprogrammed adult somatic cells to their pluripotent state. These iPSCs possessed a pluripotent morphology with a high nucleus to cytoplasm ratio and they formed densely packed colonies. Additionally, they had an extremely high proliferation rate as well as expression of key pluripotency genes [168, 169]. The original experiments were performed on mouse embryonic fibroblast cells but these results were quickly replicated in several somatic cell types from numerous species, indicating that this reprogramming method is robust and reproducible regardless of donor cell type or species [169-172].

The reprogrammed iPSCs possessed the key characteristics of pluripotent stem cells, namely they had self-renewal capacity and the ability to differentiate into cells from all three germ layers. These characteristics made them ideal candidates for disease modelling. These cells were capable of replicating indefinitely without undergoing growth arrest or developing chromosomal abnormalities. This overcomes one of the major limitations of using primary cells for disease modelling by providing a nearly endless supply of cells to work with. When given the correct stimuli, iPSCs can be differentiated into any cell type in the body by mimicking the embryonic development of a particular cell lineage. Therefore, multiple cells types can be generated in a single genetic background to model the complex intercellular interplay of many diseases. Human induced pluripotent stem cells (hiPSCs) can be generated quickly and economically from patient biopsies. This makes them suitable for use in personalized, regenerative medicine [168, 173]. They can be used to study both genetic diseases as well as the effect that genetic background has on a disease phenotype [174-176]. Importantly, these cells overcome the ethical questions that plagued their predecessors, human embryonic stem cells. Since hiPSCs are derived from consenting, adult, somatic cell samples rather than human embryos, they are not subject to the same funding restrictions and regulations [177, 178]. These hiPSCs can be generated from nearly any type of somatic cell which eliminates the need for expensive or invasive procedures to obtain a specific cell type. In addition, these cells can be grown *in vitro* for longer periods of time while primary adult cells often undergo spontaneous loss of function after being placed in culture [179]. All of these features make hiPSCs extremely useful tools for *in vitro* disease modelling. A summary of the potential uses for hiPSCs can be found in Figure 1.20.

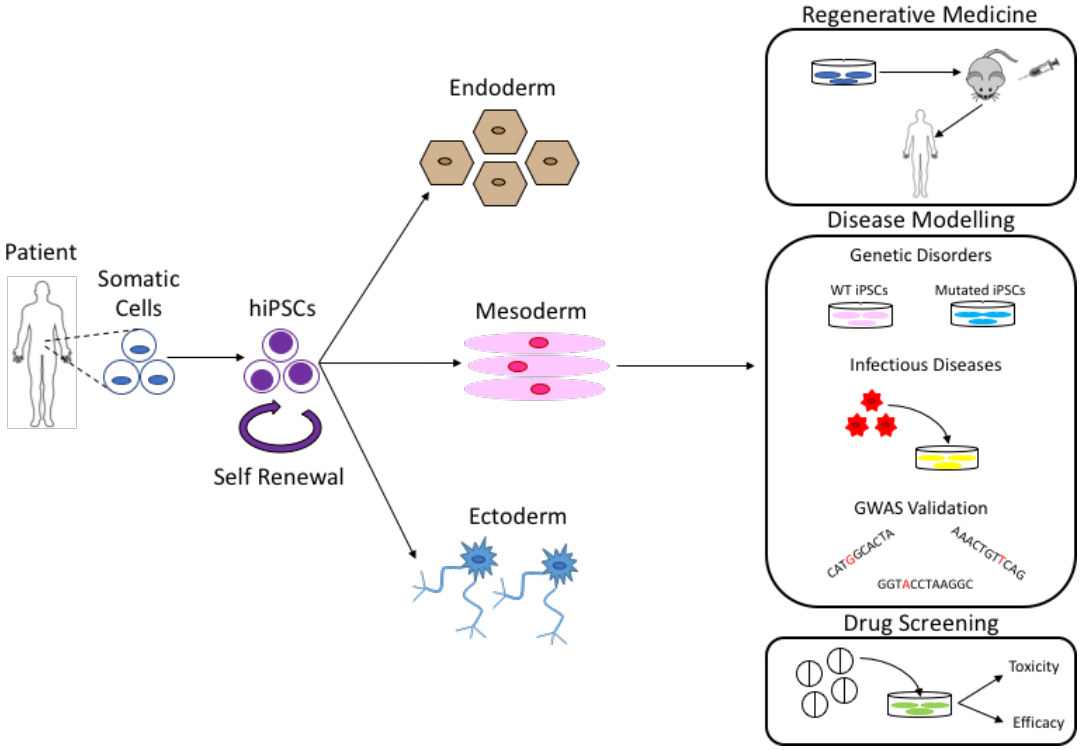

Figure 1.20 Potential of hiPSCs.

Schematic representation of the use of hiPSCs in disease modelling. hiPSCs are generated from human somatic cells. These hiPSCs possess the ability to self-renew indefinitely as well as differentiate into the three germ layers: endoderm, mesoderm, and ectoderm. Following differentiation, these cells can be used for regenerative medicine, modelling of genetic or infectious diseases, validation of GWAS hits, and drug screening. (Figure adapted from Yiangou et al. Cell Stem Cell 2018) [180]

1.5.2 Use of hiPSCs in Modelling Liver Disease

The *in vitro* modelling of liver diseases has been severely limited by the lack of access to primary tissue samples. These primary cells are extremely difficult to obtain due to the invasive nature of biopsies and the lack of availability of healthy donor tissues. Additionally, primary cells, especially hepatocytes, quickly lose their functional characteristics during *in vitro* culture. Primary human hepatocytes only remain differentiated and functional up to 7 days in 2D *in vitro* culture and thus have limited usefulness in disease modelling. Recent advancements in 3D culture as well as the development of co-culture systems have increased the viability and functionality of primary human hepatocyte cultures from days to weeks or months [181]. However, these hepatocytes do not replicate *in vitro* and therefore it remains difficult to obtain enough cells for disease modelling purposes.

To overcome these limitations, hiPSCs have been used to create hepatocyte-like cells (HLCs). hiPSCs undergo a directed differentiation that mimics the embryonic development of hepatocytes *in vivo* [182-184]. The cells are first differentiated into definitive endoderm before being specified into anterior foregut, hepatic endoderm, bipotent hepatoblasts, and ultimately "mature" HLCs. An overview of this differentiation protocol can be found in Figure 1.21. These HLCs express key markers of mature hepatocytes including HNF4α, HNF6, CEBPα, PROX1, GATA4, ALB, A1AT, and CYP3A4. Additionally, they demonstrate several hepatocellular functions including albumin secretion, glycogen synthesis, urea production, LDL uptake, and limited CYP activity [184]. However, these hepatocytes are not fully mature and more closely resemble the phenotype of foetal hepatocytes because, in addition to expressing adult hepatocyte markers such as ALB and A1AT, they express foetal markers such as alpha fetoprotein (AFP) and foetal isoforms of drug metabolizing enzymes such as CYP3A7 [185]. This is due in large part to the lack of knowledge regarding the signalling pathways and factors that control hepatocyte maturation *in vivo*. The liver is unique in that functional maturation occurs progressively after birth which makes these mechanisms difficult to study.

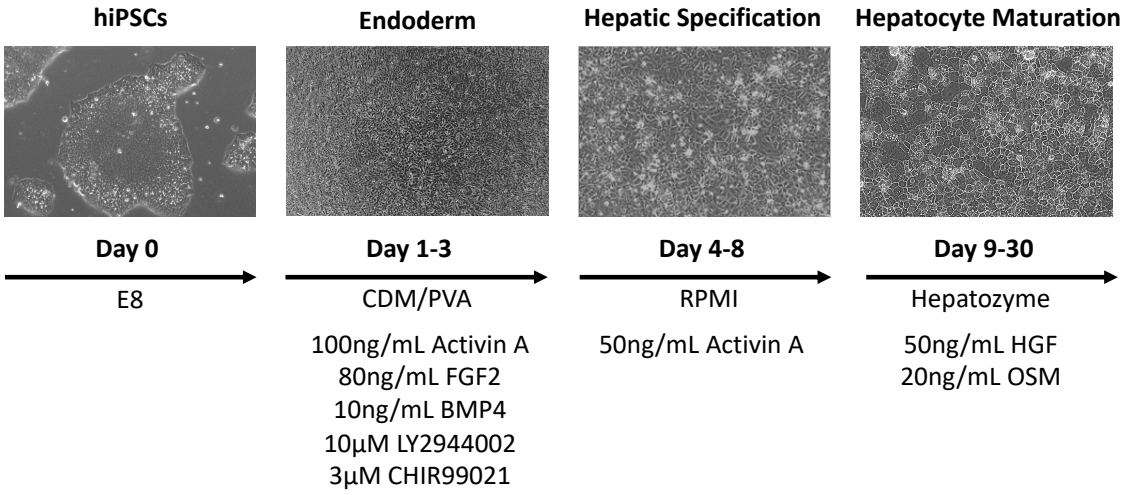

Figure 1.21 Overview of the HLC differentiation protocol.

The HLC differentiation protocol mimics embryonic development from pluripotency through endoderm, hepatic specification, and ultimately hepatocyte maturation. This process takes 30 days and is entirely chemically defined. hiPSCs are maintained in E8 medium until differentiation. Endoderm specification takes place over the first three days of differentiation and is achieved using a combination of several cytokines and small molecule inhibitors (Activin A, FGF2, BMP4, LY2944002, and CHIR99021). This is

followed by hepatic specification which is achieved over 4 days with Activin A signalling. Finally, hepatocyte maturation is the longest phase which takes place over days 9-30 with the supplementation of HGF and OSM.

Despite this functional immaturity, HLCs have been used extensively to model several different liver diseases *in vitro*. They have been used to study inherited metabolic disorders such as α-1-anti-trypsin deficiency, familial transthyretin amyloidosis, glycogen storage disease type 1a, Wilson's disease, and familial hypercholesterolemia [186-190]. In addition to using HLCs to gain mechanistic insights into disease progression, these cells have also been used as a platform for drug screening and validation for treatment of these diseases. In addition to these relatively rare genetic disease, HLCs have been used to model chronic liver diseases such as viral hepatitis and NAFLD [191-193]. Complex chronic liver diseases such as NAFLD are more difficult to model given the intricate inter-cellular interactions necessary for disease pathogenesis. In addition to HLCs, differentiation protocols have been published for several other hepatic cell types including cholangiocytes and HSCs [186, 194]. Several studies have been published that take advantage of the differentiation capacity of hiPSCs to create co-culture systems to model NAFLD *in vitro* and *in vivo [193, 195]*.

1.5.3 Limitations of hiPSCs

As with any model system, there are several limitations in using hiPSCs for disease modelling. These limitations include variation in differentiation capacity between hiPSC lines, limited maturity and functionality as well as short duration of culture for differentiated cells derived from hiPSCs. When hiPSCs are reprogrammed from somatic cells, these hiPSCs tend to retain an epigenetic signature that resembles their parental cell that is resistant to the reprogramming process. This "epigenetic memory" may cause cells to be more amenable to differentiation into the lineage of their parental cell type. This memory depends on several factors including the donor cell type, the reprogramming method used, and the culture conditions. Despite this major drawback, the epigenetic memory of hiPSCs is diluted by long term culture which reduces this inter-line variability [196]. Additionally, recent studies suggest that epigenetic memory is a rather minor contributor to inter-line variability. Primarily, this variation is driven by differences in genetic background between individuals [196-199]. Thus, it is important to include

isogenic hiPSC lines derived from multiple donors in experimental designs in order to overcome this limitation.

The major limitation of using hiPSCs for disease modelling, especially in metabolic organs such as the liver, is their functional immaturity. These cells more closely resemble foetal cells than adult cells due to the lack of understanding of hepatocellular development in the post-natal period [173, 180]. Given that these cells are not functionally mature, they may not faithfully recapitulate all aspects of adult disease. Additionally, hiPSC derived disease models have a short lifespan in culture [188]. Though the half-life of these cultures is significantly longer than primary cells, it remains on the order of weeks rather than months or years. This severely limits our ability to model chronic diseases that develop over long time scales. These limitations have been mitigated by recent developments in 3D culture technologies. Placing hiPSC derived hepatocytes into 3D culture has been shown to improve both their maturity and functionality by increasing the expression of mature hepatocyte markers as well as ameliorating their metabolic functionality [200, 201]. Not only do these 3D culture systems augment the maturity of hiPSC-derived hepatocytes, they also prolong their viability and functionality for much longer than traditional 2D culture systems. The use of 3D culture matrices also allows for co-cultures of hepatocytes and other non-parenchymal liver cells to be constructed. These co-culture systems have been shown to spontaneously self-assemble into physiologically relevant micro-structures with meaningful intercellular connections and communication [193, 195]. Co-culture of multiple hepatic cell types further improves metabolic functionality and extends culture time [202]. Unpublished data from our lab shows that co-culture of parenchymal and non-parenchymal liver cells in collagen matrices can maintain viability and functionality for up to 1 year of *in vitro* culture (data not shown). Similar advancements have been made in 3D and co-culture technologies for primary liver cells as well. Primary human hepatocytes cultured in 3D with or without other non-parenchymal liver cells can be cultured for long periods of time and faithfully recapitulate hepatic functions *in vitro [203-205]*. However, these primary cells remain extremely difficult to obtain in high quantities making them difficult to use for disease modelling and toxicity assay purposes. In addition, it is difficult to obtain several hepatic cell types from the same donor which may introduce unnecessary error into disease models utilizing

the co-culture of primary hepatic cells. Therefore, despite these minor limitations, hiPSCs remain the most ideal candidate to model complex diseases *in vitro*.

1.6 CRISPR/CAS9

CRISPR/CAS9 is an affordable, efficient, and simple method to achieve targeted genome editing. Originally described as an innate immune system in bacteria, this technology has since been utilized to precisely edit genomes in a variety of cell types and species. Unlike its predecessors zinc finger-nucleases (ZFNs) and transcription activator-like effector nucleases (TALENs), CRISPR/CAS9 relies on simple Watson and Crick base pairing rules rather than complex protein engineering [206, 207]. The CAS9 protein is guided to a specific point in the DNA sequence of a cell by a single guide RNA (sgRNA) where it causes a double stranded break in the DNA and activates one of two DNA repair pathways.

1.6.1 History of CRISPR/CAS9

CRISPR stands for clustered regularly interspaced palindromic repeats named after a series of repeated DNA sequences interspaced with short sequences derived from plasmid or viral origins within the genome of bacteria. These CRISPR regions serve as a type of memory of past invasions. Bacteria cleave a short sequence of the invading DNA and incorporate it into the CRISPR array. These sequences can then be transcribed into RNAs used to guide the cleavage of foreign DNA. The bacteria uses these antisense RNAs in combination with CRISPR associated (CAS) proteins as an adaptive immune system to interfere with virus proliferation [207]. The role of CRISPR/CAS9 in bacterial adaptive immunity is summarized in Figure 1.22. Given the simplicity and specificity of this system, researchers exploited the CRISPR/CAS9 system for the purposes of genome editing.

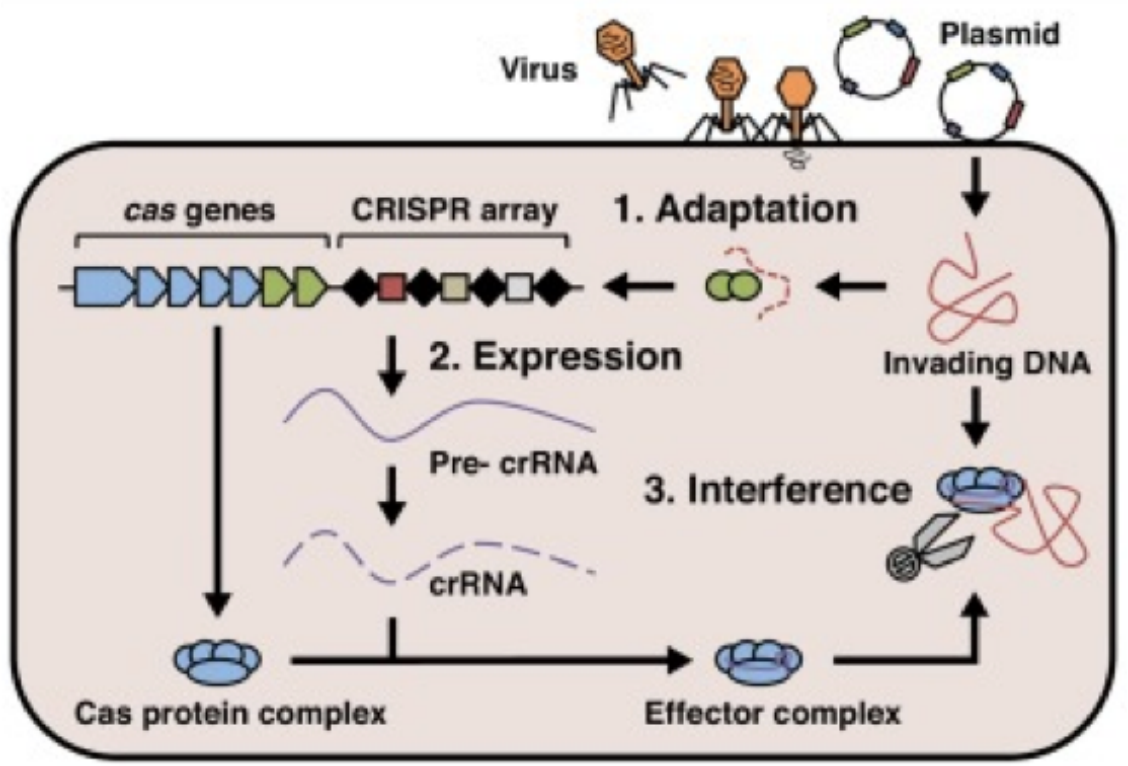

Figure 1.22 CRISPR/CAS9 adaptive immune system.

The CRISPR/CAS9 adaptive immunity can be divided into three steps. The first step is adaptation which involves incorporating a portion of invading DNA into a CRISPR array. In step two, expression, CRISPR arrays are transcribed and processed into crRNAs. In step 3, the crRNA in complex with CAS proteins binds to and degrades the invading DNA. (Figure adapted from Patterson et al. Current Opinion in Microbiology 2017) [208]

1.6.2 Mechanism of CRISPR/CAS9

CRISPR/CAS9 requires two functional partners: sgRNA and the CAS9 protein. The CAS9 protein is derived from Streptococcus pyogenes but can efficiently cleave DNA from any species. The sgRNA is designed to be complementary to a 20-nucleotide sequence in the genome. The target sequence must be adjacent to a protospacer adjacent motif (PAM). This PAM is usually NGG and is required for target site recognition by the CAS9 protein. Therefore, in theory, any DNA sequence that contains N_{20}-NGG could be used as a target site. However, additional aspects of the DNA landscape must be considered in order to design a highly specific and efficient sgRNA such as the repetitiveness of the DNA region of interest as well as the distance of the PAM site from the editing site. Following binding of the sgRNA to the target site, CAS9 generates a

double stranded break in the DNA. This double stranded break can be repaired by two mechanisms: non-homologous end joining (NHEJ) or homology directed repair (HDR). When a template sequence is not provided, the double stranded break is repaired by NHEJ which is error-prone and often results in short indels. These indels can cause non-sense mutations that generate a knock-out of a certain gene. In order to specifically edit a gene locus, a homologous repair template such as an ssODN can be provided to direct the cell toward HDR. Using HDR, the cell will replace the target site in a template-dependent manner [209, 210]. In this way, CRISPR/CAS9 can be used to either knock-out a gene of interest or knock-in a specific change to the DNA sequence. The mechanism of CRISPR/CAS9 genome editing is illustrated in Figure 1.23. CRISPR/CAS9 unlocked the potential for genome editing to be used in conjunction with hiPSCs for modelling genetic diseases *in vitro*.

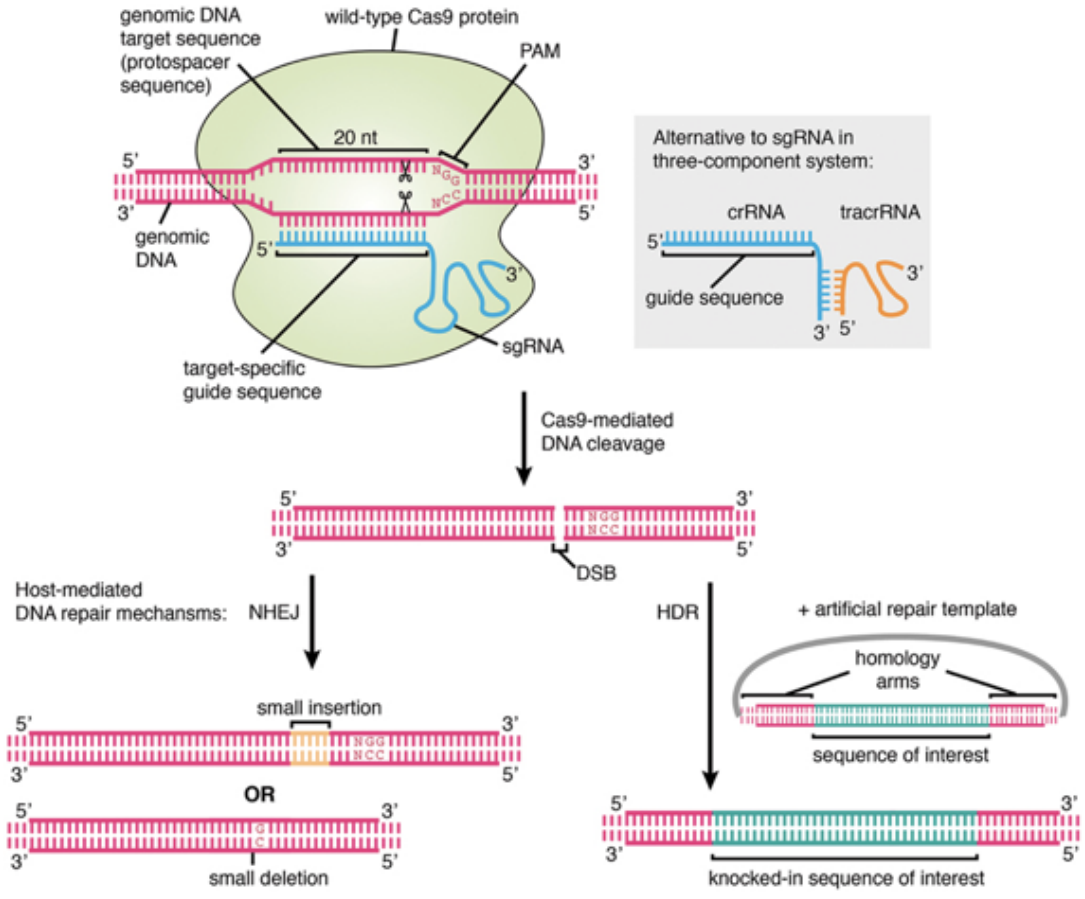

Figure 1.23 CRISPR/CAS9 genome editing mechanism.

The gRNA is designed against a 20 base pair genomic DNA target sequence that is adjacent to a PAM. This gRNA guides the CAS9 protein to the sequence of interest where it creates a double stranded break in the DNA. This double stranded break can then be repaired through NHEJ which is error prone and often results

in short indels or HDR which uses an exogenously supplied homologous sequence to specifically change the genomic sequence in a predictable manner. (Figure adapted from Agrotis et al. Frontiers in Genetics 2015) [211]

1.6.3 Limitations of CRISPR/CAS9

There are several limitations to the use of CRISPR/CAS9. The editing efficiency of CRISPR/CAS9 is inconsistent and often not very high. This editing efficiency is dependent on several factors including cell type, culture conditions, and sgRNA efficiency. This low editing efficiency can make the process of genome editing incredibly time consuming and expensive. Not all sgRNAs are equally efficient and it is difficult to determine the efficiency of any given gRNA without testing it. Several online search tools have been designed in an attempt to overcome this pitfall. These design tools are used to predict the activity of an sgRNA design *in silico*. However, despite the implementation of these design tools, several sgRNAs must still be tested experimentally to identify the sequences with the highest activity [206]. These experimental tests can be costly and time consuming.

The most concerning limitation of CRISPR/CAS9 is its off-target effects. Despite the specificity that is inherent to the design of sgRNAs, these sgRNAs may still hybridize with DNA sequences that have base-pair mismatches. This occasionally causes cleavage at off target sites. These off-target effects are often difficult to identify because they may not occur in coding regions and whole genome sequencing is not feasible for all cell lines that are generated [207, 209]. In order to mitigate these effects, web-based tools have been designed to identify potential off target effects by comparing sequence compositions. In addition, it is common practice to use “control” cell lines that have been exposed to the CRISPR/CAS9 gene editing process without editing at the site of interest. These cell lines are called untargeted control lines and would likely be subject to similar off target effects as their genetically edited brethren.

Despite these limitations, CRISPR/CAS9 remains the best genome editing tool available to biological researchers today. Several labs are working to improve the efficiency and specificity of CRISPR/CAS9 genome editing. Recent advances in base editing technology allows for the efficient and precise introduction of point mutations without

initiating a double stranded break in the DNA. These base editors directly convert one nucleotide into another and thus can be used in non-dividing cells without inducing off target effects [212]. The potential for the CRISPR/CAS9 editing technology and its derivatives for use in disease modelling as well as cell therapy and personalized medicine will only improve in the years to come.

1.7 Open Questions and Project Aims

NAFLD/NASH is one of the largest public health burdens facing the developed world. The genetic underpinnings of this disease have only recently been elucidated. The major genetic contributor to NAFLD/NASH is the I148M variant in PNPLA3. Despite over a decade of dedicated research into this enzyme and its pathogenic functions, very little is known about PNPLA3. PNPLA3 appears to play a key role in lipid homeostasis in several hepatic cell types but its function has yet to be fully clarified. Additionally, the I148M variant is a major determinant in NALFD onset and progression but the pathogenic mechanism of this variant remains incompletely understood. The study of PNPLA3 and its effect on NAFLD has been severely hindered by the lack of a human model of this disease. In order to overcome this limitation, we chose to take advantage of the unique properties of hiPSCs and CRISPR/CAS9 to develop a human model of PNPLA3-induced NAFLD.

The specific aims of this PhD project are:

1) Use CRISPR/CAS9 to create hiPSC lines with either a full knock-out of the PNPLA3 protein or with the I148M variant knocked in
2) Determine if the I148M variant represents a loss of function or gain of function variant in human cells
3) Investigate the effect of PNPLA3 knock-out and the I148M variant on lipid metabolism and gene transcription
4) Gain mechanistic insight into the pathophysiology of the PNPLA3 I148M variant in NAFLD