Investigating the Role of TM6SF2 in Lipid Metabolism

Sarah B Gibeley

Submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy under the Executive Committee of the Graduate School of Arts and Sciences

COLUMBIA UNIVERSITY

2022
Abstract

Investigating the Role of TM6SF2 in Lipid Metabolism

Sarah B Gibeley

A nonsynonymous, loss of function variant (rs58542926, E167K) located in the gene encoding TM6SF2 was identified in multiple genetic association studies as significantly correlating with increased risk for non-alcoholic fatty liver disease (NAFLD) and decreased risk for hyperlipidemia. Given the pivotal role that lipoproteins play at the juncture of these two conditions, researchers hypothesize that the ER-membrane spanning TM6SF2 protein regulates the degree of lipidation of VLDL particles synthesized in the liver. However, not all published data supports this theory and contradictions regarding many aspects of the mechanistic function of TM6SF2 remain. The inconsistencies observed in the literature are in part due to drawbacks of the models used to study TM6SF2 activity; thus, there is an obvious need for an improved hepatocyte model to better understand how TM6SF2 impacts lipid metabolism.

To address this need, we present an optimized protocol for the differentiation of inducible pluripotent stem cells (iPSCs) into hepatocyte-like cells (HLCs), created in collaboration with the Leong Lab. We provide extensive validation of HLC maturity and hepatic functionality, including prolonged albumin secretion, evidence of membrane polarity, and cytochrome P450 induction. We also demonstrate that HLCs express proteins essential for lipoprotein metabolism, secrete authentic VLDL particles, and respond to metabolic perturbations, supporting their value for modeling hepatosteatosis and VLDL metabolism in vitro.

To investigate the effect of TM6SF2 variant expression on hepatic lipid metabolism, we produced HLCs derived from 4 homozygous TM6SF2-carrier individuals (KK) and 4 age- and
sex-matched unaffected siblings (EE). We describe the variability in differentiation efficiency that we observed in our sibling-matched HLC model and present the gene editing strategy we developed using CRISPR/Cas9 technology and transgene-induced expression to create isogenic iPSCs differing only in their TM6SF2 genotype [EE, KK, or knockout (KO)]. After extensive confirmation of successful gene editing, we explore the effect of TM6SF2 on lipid metabolism in the edited iPSCs. RNA-sequencing and qPCR validation reveal that the Sterol Regulatory Element Binding Protein 2 (SREBP2)-mediated transcriptional program regulating cholesterol synthesis is significantly increased in TM6SF2 KO iPSCs. However, lipidomics analysis and de novo lipogenesis functional assays show that free cholesterol (FC) levels are unchanged. In TM6SF2 KO iPSCs, we further show a reduction in the activities of Acyl-Coenzyme A: Cholesterol Acyltransferase 1 (ACAT1) and Phosphatidylserine Synthase 1 (PSS1), two enzymes that display optimal function when specifically localized to cholesterol enriched ER lipid raft-like domains. Our findings suggest that TM6SF2 may impact cholesterol localization within ER subdomains, which regulate expression levels of cholesterol synthesis genes and activities of ER lipid-raft associated enzymes.

In summary, we present here methodological approaches for generating multiple cell culture models in which to investigate the function of TM6SF2, as well as novel evidence supporting a role for TM6SF2 in iPSC cholesterol metabolism.
# Table of Contents

List of Figures and Tables.............................................................................................................. iii

Acknowledgments.......................................................................................................................... vi

Chapter 1: Literature Review.......................................................................................................... 1

1.1 Lipoproteins in systemic lipid metabolism................................................................. 1

1.2 Dyslipidemia: A link between CVD and NAFLD risk...................................................... 5

1.3 Overview of hepatic lipid metabolism.............................................................................. 7

1.4 Regulation of VLDL synthesis and secretion................................................................. 11

1.5 NAFLD: clinical presentation, prevalence, and etiology................................................ 15

1.6 Human genetics implicate TM6SF2 Loss of Function in NAFLD, hypolipidemia, and reduced CVD risk .................................................................................................................. 20

1.7 Current understandings of TM6SF2 function................................................................. 23

1.8 Assessment of popular hepatocyte models for studying NAFLD ..................................... 34

1.9 Research goals .................................................................................................................... 41

Chapter 2: Characterizing Lipoprotein Metabolism in Hepatocyte-Like Cells ...................... 45

2.1 Reprogramming of human PBMCs to iPSCs................................................................. 45

2.2 Differentiation of iPSCs to hepatocyte-like cells (HLCs) ................................................ 47

2.3 Phenotypic and functional confirmation of HLC hepatic maturity ................................ 50

2.4 Lipoprotein synthesis and secretion in HLCs ................................................................. 56

2.5 Investigating the role of TM6SF2 rs58542926 in the regulation of lipid metabolism in patient-derived HLCs .................................................................................................................. 65

2.6 Materials and Methods...................................................................................................... 72
Chapter 3: CRISPR/Cas9-mediated Knockout of TM6SF2 in iPSCs ........................................ 85

3.1 CRISPR/Cas9 editing strategy and guide RNA screening ........................................ 85
3.2 iPSC transduction with Cas9/gRNA lentiviruses and initial screen .......................... 89
3.3 Monoclonal isolation of TM6SF2 KO iPSCs and confirmation of successful KO ...... 94
3.4 Phenotypic characterization of TM6SF2 KO HLCs .................................................... 99

Chapter 4: A New Role for TM6SF2 in Lipid Metabolism ................................................. 110

4.1 Overview of TM6SF2 function in cholesterol homeostasis ...................................... 110
4.2 RNA-sequencing Analysis of TM6SF2-KO in iPSCs .................................................. 123
4.3 Effect of TM6SF2-KO on the lipidome in iPSCs ......................................................... 126
4.4 TM6SF2 and Cholesterol-Rich ER Lipid Rafts ............................................................ 130
4.8 Detailed Methods ...................................................................................................... 141

Chapter 5: Summary and Discussion .............................................................................. 146

References ........................................................................................................................ 152
List of Figures and Tables

Chapter 1: Literature Review

Figure 1.1 Lipoproteins in systemic lipid metabolism

Figure 1.2 Overview of hepatic triglyceride metabolism

Figure 1.3 Assembly of VLDL

Figure 1.4 Depiction of the spectrum of Non-alcoholic fatty liver disease

Figure 1.5 TM6SF2 tissue distribution, subcellular localization and proposed structure

Figure 1.6 Hepatocyte membrane polarity and essential functions

Figure 1.7 Factors to consider when choosing hepatocyte sources for disease modeling

Chapter 2: Characterizing Lipoprotein Metabolism in Hepatocyte-Like Cells

Figure 2.1 PBMC-to-iPSC reprogramming

Table 2.1: iPSC donor information

Figure 2.2 Optimized protocol for the directed differentiation of iPSCs to mature HLCs using a combination of growth factors (GFs) and small molecules (SMs)

Figure 2.3 Monitoring of cell stages during iPSC differentiation to HLCs

Figure 2.4 Phenotypic characterizations in HLCs

Figure 2.5 Phenotypic and functional characterizations in HLCs

Figure 2.6 Assessment of basolateral membrane-localized functions in LG1 HLCs

Figure 2.7 Determination of xenobiotic metabolism in LG1 HLCs

Figure 2.8 Assessment of HLC apoB100 expression and secretion over time

Figure 2.9 MTP inhibition reduces VLDL-apoB and TG secretion in HLCs from donor LG1

Figure 2.10 Determination of the relative particle density of HLC-secreted lipoproteins
Table 2.2: Old Order Amish (OOA) donor information

Figure 2.11 Initial characterization of iPSCs and HLCs derived from Old Order Amish (OOA) sibling pairs

Figure 2.12 Comparison of hepatocyte surface marker ASGPR1 expression in detached HLCs and HepG2 cells

Table 2.3: List of Primary Antibodies

Table 2.4: List of qPCR Primers

Chapter 3: CRISPR/Cas9-mediated Knockout of TM6SF2 in iPSCs

Figure 3.1 Schematic illustrating the overall process of developing isogenic HLCs through genetic editing and cell screening in LG2-C6 iPSCs

Figure 3.2 Pre-screening of TM6SF2 targeting guide RNAs in HEK-293T cells

Figure 3.3 TM6SF2-gRNA lentivirus transduction in HepG2-Cas9-PuroR cells

Figure 3.4 Confirmation of Cas9-PuroR and TM6SF2-gRNA lentivirus (LV) transduction in LG2-C6 iPSCs

Figure 3.5 Schematic illustrating the process of visually screening iPSCs for the selection of LG2-C6-Cas9-PuroR-EGFP+ monoclones

Figure 3.6 First round of clonal screening for complete TM6SF2-KO in LG2-C6 iPSCs

Figure 3.7 Second round of clonal screening for complete TM6SF2-KO in LG2-C6 iPSCs

Figure 3.8 Screening for complete TM6SF2-KO in LG2-C6 iPSCs

Figure 3.9 Phenotypic characterization of HLCs derived from TM6SF2 KO iPSCs and isogenic (KK) controls

Table 3.1 gRNA sequences for CRISPR/Cas9-mediated knockout of human TM6SF2
Chapter 4: A New Role for TM6SF2 in Lipid Metabolism

Figure 4.1 Summary of the cholesterol biosynthesis pathway from acetyl-CoA

Figure 4.2 Schematic of SREBP2 regulation by ER cholesterol

Figure 4.3 Schematic of proposed TM6SF2 transmembrane topology in the ER and interactions with ERLIN1/2 and apoB

Table 4.1 Proposed TM6SF2 interacting partners, as determined by tandem affinity purification

Figure 4.4 Enrichment analysis of dysregulated genes in TM6SF2 KK vs. -/- iPSCs performed in Enrichr

Figure 4.5 Gene clustering determined in the upregulated gene set through building a large network in Cytoscape followed by analysis with MCODE

Figure 4.6 Lipidomics Analysis of TM6SF2-KO iPSCs

Figure 4.7 Impact of TM6SF2 on iPSC de novo lipogenesis

Figure 4.8 Protein expression of cholesterol synthesis regulators

Figure 4.9 Effect of TM6SF2 KO on MAM activity in iPSCs

Figure 4.10 Transgene expression of TM6SF2 WT (EE) and Mutant (KK) allele in isogenic iPSCs

Figure 4.11 Comparison of TM6SF2 exogenous gene expression vs cholesterol synthesis gene expression in iPSCs by linear regression analysis

Figure 4.12 Measurement of iPS cell cholesterol using Amplex Red fluorometric assay

Figure 4.13 Impact of TM6SF2 transgene expression on iPSC de novo lipogenesis

Figure 4.14 Impact of TM6SF2 transgene expression on ACAT and DGAT activity in iPSCs
Acknowledgments

I would like to acknowledge many people in my professional and personal life without whom the accomplishment of this doctoral work would not have been possible. First, I owe a tremendous debt of gratitude to my mentor, Dr. Henry Ginsberg, who accepted me as his student and provided continuous support and encouragement throughout this entire journey. He has provided me with a world-class education in the field of lipoprotein metabolism and exceptional opportunities for professional growth. This project was particularly challenging and I often found myself struggling to stay positive; Henry showed me endless patience and maintained confidence in my abilities as a scientist. He has been my strongest advocate and has gone above and beyond the measure of a mentor to see that I succeed.

I am grateful to my committee members Dr. Robert Schwabe, Dr. Rebecca Haeusler, and Dr. Anthony Ferrante Jr, for their time, dedication, and insights provided. Both Drs. Schwabe and Ferrante welcomed me as a rotating first year student and gave me the space to explore different aspects of lipid metabolism and learn essential skills I’d go on to apply to my thesis work. Dr. Haeusler always made herself available when I needed guidance and offered the invaluable advice to consider examining cholesterol metabolism when the project hit a snag. Finally, Dr. Kam Leong took a leap of faith and agreed to serve as co-mentor on this journey. I thank him for welcoming me so warmly into his lab, for introducing me to the fascinating world of bioengineering, and for his genuine investment in my success as a graduate student and beyond.

I am similarly thankful to Dr. Danielle Huang who was Dr. Leong’s mentee at the time that our collaboration began. Due to the nature of our overlapping projects, Danielle and I worked side-by-side on most of the experiments detailed in Chapters 2 and 3. This project would not have been possible without her significant contributions, including many of the initial HLC characterization
assays, the beautiful confocal images, the plasmid production, and much more. Danielle did not hesitate to share her protocols/supplies/bench, split time evenly (including many weekends) to care for our cells, and to motivate me when I felt discouraged.

I am grateful to everyone in the Ginsberg lab family who have contributed immensely to my dissertation work and beyond. Dr. Gissette Reyes-Soffer mentored me as a rotating student and continued to support me as a permanent member of the lab. As the principal scientist on another TM6SF2 project, Gissette was always willing to consult and to share a positive and encouraging outlook. I thank Dr. Tatyana Fedotova for the many hours of training on lab protocols when I first joined the lab and for providing general scientific and career advice. I also thank Dr. Jing Liu who taught me the TLC protocol and who also worked on studying aspects of TM6SF2 function and could therefore commiserate when the story often seemed more complicated than we had anticipated. We tragically lost our beloved lab manager Colleen Ngai at the end of 2019; I appreciate Colleen for everything she did to keep the lab running, for sharing her expertise at isolating lipoprotein fractions, and for swapping stories about our nephews while waiting for the ultracentrifuge. Although I had less time than I would have liked to hang out in with them in the lab, Anastasiya, James, and Anna kept lab meetings entertaining and were always willing to lend a helping hand or a sympathetic ear when needed. Finally, I must mention Dr. Antonio Hernandez-Ono: Antonio is a pillar of the Ginsberg lab and I am indebted to him for onboarding me into the lab and teaching me everything I needed to know; I’d to specifically thank him for putting a tremendous amount time into training me in general mouse handling and the complicated primary hepatocyte isolation surgery.

Next, I’d like to thank everyone in the Leong lab for warmly welcoming me and sharing their precious bench and hood space with me. I especially appreciate Drs. Sima Lao, Baoding
Huang, Cong Xu, Siyu He, Hyesung Kim, and Chai Hoon Quek who offered scientific advice, precious reagents, and an extra pair of hands on the weekends to care for our cells when we were both out of town. I am also grateful to Wei Leong and Ozgenur Celik for their assistance with running assays and maintaining cell lines that were crucial for this thesis work.

Many other scientists have helped to make this work possible that I would like to acknowledge. Dr. Rob Bauer offered guidance on CRISPR gene editing strategies. Dr. Barbara Corneo and the Columbia Stem Cell Core provided technical advice on the hepatocyte differentiation protocol. The Biomarkers Core, FACS facility in the Stem Cell Core, and the Genome Center processed samples for lipidomics, cell sorting, and RNA-sequencing, respectively.

I’d like to give special recognition to Dr. Estella Area-Gomez and Dr. Rishi Agrawal not just for sharing their insight and protocols related to the MAM, but for their generous time and encouragement. Estella provided bountiful conceptual advice when the project shifted to cholesterol, and shared her expertise in interpreting our lipidomics data. As my classmate, Rishi saw me struggling at times and refused to let me fail. I am truly indebted to both of you for your support and tough love.

I am grateful to the directors of the NMB program, both when I was admitted and now, for choosing to cultivate the potential they saw in me. The NMB administration, particularly Zachary Corter, have helped me in numerous ways through my graduate school career and I appreciate all of their hard work behind the scenes. I also would like to mention the TL1 program for providing financial support in the form of a 2 year fellowship and for challenging to see my project through the lens of precision medicine.

I would remiss if I did not acknowledge my pre-graduate school mentors in the Cardiovascular Nutrition Lab at the Tufts University Human Nutrition Research Center: Drs. Alice
Lichtenstein, Nirupa Matthan, and Stefania Lamon-Fava. Thank you for giving me my first opportunity to contribute to nutrition research, for serving as inspiring role models of successful women in academia, and for supporting me through the graduate school application process. I am also grateful to Jean and Joyce, my lab-mates in the CNL, who engendered a truly cooperative and fun lab environment.

Finally, I would like to acknowledge those who supported me through this project outside of academia. Thank you to my friends, classmates, and long list of roommates who were there to celebrate the achievements, console the disappointments, and stick around during the stretches of radio-silence. I would not have been able to complete this work without the constant encouragement from my family and extended relatives. My sister and brother-in-law showered me with pictures and videos of my nephews when I needed a little pick-me-up, and gave me a safe haven from NYC during the height of the pandemic. Lastly, I cannot thank my parents enough for instilling in me an intellectual curiosity, a tenacious work ethic, and a sense of integrity that has seen me through my academic endeavors. I am most thankful for their unending and unconditional love and support.
Chapter 1: Literature Review

1.1 Lipoproteins in systemic lipid metabolism

Cardiovascular disease (CVD) is a broad category of pathologies that affect the cardiac and vascular systems, which deliver blood to the body’s organs and peripheral tissues. Although tremendous advances in research and clinical practice have been made over the past decades, CVD remains the number one cause of death in the developed world.\(^1\) Atherosclerosis is a disease characterized by the formation of plaques composed of lipids, as well as fibrotic materials, inflammatory cells, and other debris, in the walls of the arteries.\(^2\) These plaques lead to the narrowing of the vessels and cause the majority of cardiovascular events such as myocardial infarction (MI), stroke, and carotid artery stenosis.\(^3\) In fact, atherosclerotic cardiovascular disease (ASCVD) starts as early as childhood and studies have found that by middle age, 71% of men and 43% of women show evidence of subclinical atherosclerosis.\(^2\)

Metabolic dysfunction is a major contributor to CVD risk and encompasses a myriad of etiologies and clinical manifestations. Dyslipidemia, or an imbalance in circulating lipids, is strongly associated with atherosclerosis. Plasma lipids, specifically triglycerides (TG) and cholesterol esters (CE), are very hydrophobic and must, therefore, be transported through the circulation as part of lipoprotein particles. The latter comprise a heterogeneous group of micelles differing by size, lipid composition, apolipoprotein content, and tissue of origin. TG and CE are in the core of the lipoprotein, covered by a surface monolayer of phosphatidylcholine (PC) and unesterified cholesterol, which act as an interface between the lipophilic core and the aqueous plasma. Lipoproteins also carry lipid soluble vitamins A, D, E, and K.\(^4\) The various associated apolipoproteins determine many of the characteristics of the lipoproteins, including their origin and sites of clearance from the circulation, as well as specific functions such as interactions with
enzymes and delivery of the core lipids to tissues and cells. Lipoproteins are classified by their density; those lipoproteins carrying a greater proportion of TG (such as chylomicrons and VLDL) are less dense, and those carrying a greater proportion of CE and apolipoproteins (such as LDL and HDL) are more dense.\textsuperscript{4}

Lipoproteins are also classified by their associated apolipoproteins, and specifically whether they contain the ~500kDa apolipoprotein B100 (apoB) as their primary structural protein. (ApoB will be discussed in more detail below). Chylomicrons, the least dense of the apoB containing lipoproteins, are synthesized in enterocytes in the post-prandial state from dietary lipids and apoB48 (a truncated isoform of apoB). These TG-rich particles are secreted into the lymphatic system before entering the peripheral circulation where they release free fatty acids via lipoprotein lipase (LPL)-dependent hydrolysis (\textbf{Figure 1.1}).\textsuperscript{5} VLDL are assembled with the full length apoB100 protein in hepatocytes and are also metabolized in the circulation by LPL to intermediate-density lipoproteins (IDL), and low-density lipoproteins (LDL).\textsuperscript{6} Together, chylomicrons and VLDL comprise the TG-rich lipoprotein (TRL) class; partial or inefficient lipolysis of these particles in the circulation by LPL leads to remnant formation.\textsuperscript{6} These TG depleted remnants are taken up by the liver via several pathways, including via LDL receptors (LDLRs).

The cholesterol-rich LDL particle may also be taken up via the LDL-receptor by peripheral cells, particularly in steroidogenic tissues, or by hepatocytes via the LDL-receptor. LDL and other apoB-containing lipoproteins can also be retained within the arterial wall; a key initiating step in atherogenesis; this process results from ionic interactions between negatively charged arterial proteoglycans and basic amino acids in apoB100.\textsuperscript{7} Notably, only one apoB molecule is associated per lipoprotein particle (chylomicron, VLDL, IDL, and LDL) and the apoB remains with the same
lipoprotein during lipolysis; thus, apoB concentration serves as a convenient surrogate measure for particle number.\textsuperscript{6}

High density lipoproteins (HDL) are characterized by having apolipoprotein A1 (apoA1) as their structural protein instead of apoB. In contrast to apoB-containing lipoproteins, which move lipids from the liver to the periphery, HDL particles take up cholesterol from the plasma membranes of cells within peripheral tissues for transport to the liver. ApoA1 is synthesized in hepatocytes and enterocytes and secreted with phospholipid into the circulation. There, apoA1-phospholipid discs stimulate efflux of free cholesterol from the plasma membrane from cells in all tissues, including from cholesterol-laden macrophages. This occurs through its interaction with the cell surface ATP-dependent transporter ABCA1, in a process known as reverse cholesterol transport.\textsuperscript{8} ApoA1 also activates the enzyme lecithin:cholesterol acyltransferase (LCAT), which generates cholesteryl esters from unesterified cholesterol and lysophosphatidic acid (lecithin). The cholesteryl esters enter the core of the apoA1 phospholipid disc, allowing for the growth of the nascent HDL particle as additional free cholesterol can move from cell membranes to the surface of the HDL particle.\textsuperscript{9} This particle can also receive additional unesterified cholesterol through interaction with ABCG1, another efflux transporter. This cholesterol transferred to HDL by ABCG1 also undergoes esterification by LCAT to generate cholesteryl esters. Although the metabolism of HDL particles and their component lipids is very complex, there is a consensus that the mature HDL selectively delivers CE to the liver either through interacting with scavenger receptor B1 (SR-B1) or by first transferring the CE to apoB-containing lipoproteins. The latter occurs through the action of plasma cholesteryl ester transfer protein (CETP), which mediates a biomolecular exchange of HDL (or LDL) CE for VLDL or chylomicron TG.\textsuperscript{4}
Figure 1.1: Lipoproteins in systemic lipid metabolism. Hepatic triglycerides derived from uptake of circulating FFA and TG-carrying remnant lipoproteins can be oxidized for energy, stored in lipid droplets, or used for the assembly of VLDL that are released into the circulation. Here, VLDL, together with intestinal-derived chylomicrons (CM), interact with the enzyme LPL in the capillaries of adipose tissue, skeletal muscle, and the heart. LPL hydrolyzes triglycerides and the product FFA that are taken up for storage as new synthesized TG or for oxidation to generate energy. The VLDL and chylomicrons become smaller and are thus transformed to cholesterol-enriched TRL remnants that can be internalized by hepatocytes through receptor and non-receptor endocytic pathways and digested in the endo-lysosomal compartment. All chylomicron remnants leave the circulations through hepatic uptake, but 25-75% of VLDL remnants avoid endocytosis and are further processed by other intravascular enzymes to become LDL. HDL particles are generated from lipid-poor APOA1 secreted from the liver and intestine (not shown) by accumulating PLs generated by TRL lipolysis. These nascent HDL mature by accumulating CE via the action of LCAT on HDL-surface FC. HDL exchange CE for TG in TRL in a reaction catalyzed by CETP. Abbreviations: FFA, free fatty acids; TG, triglyceride; VLDL, very low density lipoprotein; CM, chylomicron; LPL, lipoprotein lipase; TRL, triglyceride rich lipoprotein; LDL, low density lipoprotein; HDL, high density lipoprotein; APOA1, apolipoprotein A1; PL, phospholipid; CE, cholesterol ester;
1.2 Dyslipidemia: A link between CVD and NAFLD risk

Metabolic dysfunction is a major contributor to CVD risk and encompasses a myriad of etiologies and clinical manifestations. Dyslipidemia, or an imbalance in circulating lipids, is strongly associated with atherosclerosis. Extensive evidence from epidemiologic, genetic, and clinical studies has unequivocally shown that LDL is not only associated with, but is also causally linked to the development of atherosclerosis. LDL particles transcytose across the endothelium to enter the vascular intima and accumulate in arterial walls, where they can be internalized by macrophages that infiltrate the same areas, leading to cholesterol loading of the macrophages. This generates “foam cells,” which are the cellular hallmark of atherosclerotic plaques. While many pathophysiological components of this process are not fully understood, including factors modulating the intimal penetration and retention of LDL, lowering the circulating levels of cholesterol-rich apoB-containing lipoproteins results in clinically significant reductions in the risk for CVD.

Similarly, human clinical and epidemiological studies have observed a significant, inverse association of HDL cholesterol levels with CVD risk. It is thought that the cardioprotective effects of HDL are mediated by its ability to facilitate cholesterol efflux from macrophage foam cells in arterial plaque via reverse cholesterol transport, resulting in reductions in lesion size. However, the premature termination of CVD outcome trials that involved HDL raising due to lack of efficacy, has raised questions regarding the validity of the HDL cholesterol hypothesis. A growing body of evidence suggests, therefore, that HDL cholesterol may not be a causal risk mediator for CVD and that some or even most molecular mechanisms that raise HDL cholesterol may not consequently reduce the risk of CVD. Additionally, recent findings demonstrated that
although functional efflux data correlate with cardiovascular risk, HDL cholesterol genetics do not, further challenging the causal relationship of HDL cholesterol with CVD and suggesting that focus should be placed on developing strategies that enhance the efflux capacity, rather than quantity, of HDL.\textsuperscript{12}

During the past several decades, epidemiologic and genetic evidence has indicated that, in addition to the cholesterol-rich lipoproteins, levels of plasma TG, TRL, and TRL remnants are also causally related to risk for ASCVD, as well as all-cause mortality.\textsuperscript{6} TRL remnants are involved in both the initiation and progression of atherosclerotic lesions. Thus, retention of remnants in the subendothelial space leads to cholesterol deposition and macrophage foam cell formation followed by inflammation.\textsuperscript{6} In this manner, TRL and their remnants may contribute to residual CVD risk in patients on optimal low-density lipoprotein (LDL)-lowering therapy. As a result, the need for the development of innovative therapeutics to lower TG and TRL and their remnants persists.\textsuperscript{6}

Dyslipidemia is closely related to Non-Alcoholic Fatty Liver Disease (NAFLD), another metabolic disease that has reached epidemic proportions. The presence of dyslipidemia is reported in 20\% to 80\% of patients with NAFLD\textsuperscript{13} and in fact ASCVD represents the leading cause of mortality in NAFLD patients, independent of other metabolic comorbidities.\textsuperscript{14} Additionally, NAFLD is strongly associated with the phenotype of the metabolic syndrome (MetS), which encompasses an established group of risk factors for ASCVD.\textsuperscript{5} Diagnostic criteria for the metabolic syndrome include the presence of three of the following five conditions: central obesity, hypertriglycerideremia (TG level greater than 150 mg/dL), low HDL-C (less than 40 mg/dL in men, less than 50 mg/dL in women), hypertension (systolic BP greater than or equal to 130 mm Hg or diastolic BP greater than or equal to 85 mm Hg), or hyperglycemia (fasting plasma glucose greater than or equal to 100 mg/dL or a diagnosis of diabetes mellitus).\textsuperscript{15} NAFLD is also strongly
associated with type 2 diabetes mellitus and abdominal obesity and shares with MetS a common pathophysiological mechanism: insulin resistance (IR). In fact, features of these diseases overlap to such a great extent that NAFLD is now being considered the hepatic component of MetS.\(^5\) MetS is estimated to affect a quarter of the world’s population;\(^6\) thus, a deeper understanding of the mechanisms that regulate features of MetS can have a far-reaching, beneficial impact in reducing NAFLD progression, CVD events, and the development of other metabolic disease.

### 1.3 Overview of hepatic lipid metabolism

Hepatic lipid accumulation, the hallmark of NAFLD, occurs when one or more pathways regulating liver lipid metabolism becomes imbalanced. These pathways, summarized in Figure 1.2, are typically categorized into the following: I. lipid uptake, II. *de novo* lipogenesis, III. fatty acid oxidation, IV. lipid droplet dynamics and V. lipid secretion.

**I. Lipid uptake:** Lipids are taken up by the liver from the circulation primarily in the form of free fatty acids (FFAs) bound to albumin which are hydrolyzed from adipose tissue, or, to a lesser extent, as components of neutral lipids within lipoprotein remnant particles. FFAs are internalized via several transporters, including Fatty Acid Transport Proteins (FATPs) and Fatty acid translocase (CD36). Once in the cell, these fatty acids are activated with CoA by acyl CoA synthetase long-chain (ACSL) enzymes and can then be used by the hepatocyte to generate TG and phospholipids (PLs) for VLDL synthesis or incorporation into cytosolic lipid droplets. The acyl transferases MGAT and DGAT, catalyze the formation of the ester bonds between the glycerol backbone and the acyl-CoAs to form diacylglycerol (DG) and TG, respectively.\(^7\) Currently, two DGAT isoforms have been identified in eukaryotes; although both DGATs catalyze the same reaction, they are encoded by genes from distinct families and share no homology in primary amino acid sequences.\(^8\) DGAT1 is an ER integral membrane protein, is expressed mainly
in intestine, and plays a central role in fat absorption by esterification of exogenous FAs into TG destined for cytosolic lipid droplets (cLDs) and incorporation into chylomicrons. DGAT1 is also expressed in the liver, but its impact on hepatic lipid metabolism is less clear. DGAT2 is inserted into the outer leaflet of the ER membrane and can also be found on the surface of cLDs; it incorporates endogenous FAs into newly formed TG destined for VLDL secretion or cLDs. 19

Remnant lipoprotein particles taken up by hepatocytes from the circulation serve as another source of hepatic lipids, as described in a previous section. The LDLR mediates uptake of these particles by binding to either apoE on the chylomicron remnant or to apoB100 on the VLDL remnant. Other receptors, such as LDL-related receptors (LRP1) and VLDL receptor (VLDLR), as well as heparin sulfate proteoglycans, such as syndican 1, also facilitate remnant uptake, but their role is less clear. 2

II. De novo lipogenesis: De novo lipogenesis (DNL) is the pathway whereby acetyl-CoA, produced in the mitochondria during carbohydrate metabolism, is converted to FAs, which are then esterified mainly into TG for storage or secretion. Several enzymes, including ATP citrate lyase (ACLY), acetyl-CoA carboxylase (ACC), and FA synthase (FAS), are involved in the synthesis of the long-chain saturated FAs palmitate (C16:0) and stearate (C18:0) from acetyl-coA. 19 Stearoyl-CoA desaturase-1 (SCD1) introduces a double bond into these FAs to form their monounsaturated fatty acid (MUFA) counterparts, namely palmitoleate (C16:1) and oleate (C18:1). Further elongation of the fatty acid chain can occur by the ELOVL family of enzymes, which differ in their tissue distribution and FA affinity, based on degree of saturation.

Transcriptional control of these enzymes occurs mainly via SREBP-1c, which is the master regulator of FA synthesis, and the carbohydrate response element-binding protein (ChREBP), which responds to increased concentrations of glucose metabolism and products of glucose
metabolism to stimulate DNL. DNL is also highly subject to hormonal control: insulin acts on SREBP-1c, likely through liver X receptors (LXRs) and mammalian target of rapamycin (mTOR), and glucagon regulates ACC activity and malonyl-CoA levels. In healthy humans, DNL contributes 5–10% of TG that is either stored in the liver or is secreted as VLDL–TG. However, in individuals who are obese, insulin resistant, diabetic, or who consume diets high in fructose or sucrose, the contribution of DNL to hepatic and VLDL–TG can reach as much as 30% or more.

III. β-Oxidation: During times of energy depletion, the liver generates ATP from stored lipids via FA oxidation, which occurs primarily within the mitochondria (and in the peroxisome for select substrates, particularly in rodents). Carnitine palmitoyl transferase 1 (CPT1) converts fatty acyl-CoA to fatty acyl carnitine in the outer mitochondrial membrane, which is then transferred by carnitine translocase to Carnitine palmitoyl transferase 2 (CPT2). CPT2 regenerates fatty acyl-CoA and free carnitine within the mitochondrial membrane, allowing for β-oxidation to begin. The transcriptional factor PPARα regulates the expression of CPT1 and MCAD, the mitochondrial dehydrogenase enzyme that catalyzes the first step of oxidation. β-oxidation is also substrate driven, and thus CPT1 activity is inhibited by carnitine and malonyl-CoA concentrations.

IV. Cytosolic lipid droplets: Since free fatty acids and free cholesterol are cytotoxic, their neutral esterified counterparts, TG and CE, serve as an efficient means of storing these molecules safely when they exceed the needs of the cell. These neutral lipids are packaged as cLDs surrounded by an amphipathic monolayer of phospholipids. While the manner in which these structures form is still under intensive investigation, the prevailing theory asserts that the droplets bud off from the ER membrane (where lipid synthesis occurs) and expand in size through fusion with other LDs or by continued synthesis, particularly of TG, at the surface of the droplet. While
originally thought to only serve as a lipid depot, cLDs are now considered *bona fide* dynamic organelles with distinct surface proteins that regulate the composition and availability of the lipids within, as well as mediate interactions with other organelles. Several important proteins known to reside on cLDs include the CIDE family of proteins; lipases such as ATGL and HSL; SNARE proteins involved in trafficking and fusion; and members of the perilipin family.\textsuperscript{22}

**V. Secretion:** A key pathway for maintaining hepatic lipid homeostasis is VLDL secretion, which occurs exclusively in hepatocytes. The primary role of VLDL secretion is to transport TG to skeletal and cardiac muscle for energy production and deposit excess TG in adipose tissue for safe storage.\textsuperscript{6}

*Figure 1.2: Overview of hepatic triglyceride metabolism. Triglycerides (TG) are synthesized in the liver from fatty acyl CoA (FA-CoA), which can be derived from: I. the periphery (as lipolytic products from adipose or carried within lipoprotein remnants), II. *de novo* synthesis in the liver, or IV. hydrolysis of hepatic lipid droplets. Conversely, III. FA-CoA can also be oxidized in the mitochondria when additional energy is needed. Finally, excess TG can be IV. stored within neutral cytosolic lipid droplets, or V. secreted from the hepatocyte as a major component of VLDL for delivery to peripheral tissues. Created with Biorender.com*
1.4 Regulation of VLDL synthesis and secretion

VLDL are spherical aggregates comprised of a phospholipid amphipathic monolayer surrounding a neutral lipid core, predominantly composed of TG, but also significant CE. A single apoB100 molecule serves as the primary structural component of VLDL, and is integral to the proper assembly and secretion of these lipoproteins. The 4,536 amino acid-long apoB100 peptide contains long, very hydrophobic regions, which must bind to neutral lipid to ensure the stability of the protein. As apoB100 mRNA is translated on endoplasmic reticulum (ER)-associated ribosomes, the resulting peptide concurrently passes through a translocon into the lumen of the rough ER, where it is initially lipidated through its binding to microsomal triglyceride transfer protein (MTP) (Figure 1.3, 1). While MTP most efficiently transfers TG, it has also been shown to possess CE transfer ability between microsomes in vitro and possesses phospholipid transfer activity sufficient for lipoprotein assembly and secretion. In a process less understood, it appears that MTP also transfers lipid synthesized in the smooth ER membrane into the ER lumen independently of apoB. These ER lumenal lipid droplets are thought to fuse with nascent VLDL, producing larger, mature particles, although the mechanism for this process is unknown (Figure 1.3, 2).

Following synthesis and lipidation in the ER, TG-rich VLDL particles are packaged into specialized COPII coated vesicles, sometimes referred to as VLDL transport vesicles (or VTV) (Figure 1.3, 3). These transport vesicles are characterized by a number of unique proteins, including the GTPase Sar1B, which are thought to facilitate the expansion of the budding vesicle at ER exits sites (ERES) to accommodate VLDL, which are too large to fit in typical COPII vesicles. Carried in the VTV, VLDL are trafficked across the ER-Golgi Intermediate Compartment (ERGIC) (Figure 1.3, 4) and delivered to the trans-Golgi (Figure 1.3, 5), where
they become further modified by glycosylation.\textsuperscript{25} Controversy still exists as to whether the VLDL that exits the ER are completely lipidated, or if further lipidation occurs in the ERGIC or Golgi; evidence supporting both possibilities has been found.\textsuperscript{27,28} Eventually, the fully mature VLDL exit the cis-Golgi in vesicles destined for the plasma membrane, where they are released into the circulation.\textsuperscript{25}

The apoB100 gene is constitutively expressed in hepatocytes; therefore, VLDL assembly and secretion are initially regulated by energy balance and lipid availability.\textsuperscript{18} When insufficient substrate is available to lipidate apoB, translation pauses and the partial apoB peptide undergoes retrograde translocation through the translocon into the cytosol, where it is ubiquitinylated, and targeted for proteasomal degradation (\textbf{Figure 1.3, 0}).\textsuperscript{29} Alternatively, in cases where adequate energy is present, but components essential for apoB lipidation are compromised, the lipids synthesized in the ER membrane instead accumulate in cLDs. This is evidenced by the discovery that people bearing loss of function mutations in the apoB and MTP genes exhibit reduced assembly and secretion of VLDL, which clinically manifests as \textit{hypolipidemia} and (in some cases) hepatic steatosis due to the blockage of lipid secretion from the liver.\textsuperscript{30}
Figure 1.3: Assembly of VLDL. 1) MTP transfers lipids from the ER membrane onto the nascent VLDL or 2) onto ER luminal lipid droplets which then fuse with the nascent VLDL to form lipid-rich VLDL. 3) These particles are packaged into specialized COPII-coated vesicles at ER exit sites. The vesicles are trafficked across the ERGIC and release their cargo at the Golgi. 4) The VLDL cargo is delivered to the Golgi, where it undergoes further modifications, after which it is trafficked to the plasma membrane for release into the circulation. 5) When inadequate lipidation of apoB occurs, for example, due to lack of MTP activity, the nascent apoB becomes unstable, and is ultimately degraded by the proteasome. Created with BioRender.com

A variety of other conditions can lead to impaired lipidation of apoB as well. It has been well established that the majority of TG in VLDL (60%–80%) is derived from re-esterification of lipolytic products in hepatocytes. Therefore, impairments in the enzymes required for lipid re-esterification, such as DGAT2 (TG synthesis), ACAT2 (CE synthesis), and PEMT/CTα (PC synthesis) have been shown to reduce VLDL secretion and induce hepatosteatosis.
Additionally, proteins involved in the lipolysis of cLDs and transfer of lipolyzed products to the ER have been shown to be necessary for VLDL assembly. It has been suggested that a membrane-associated protein that can be found on cLDs, smooth ER, and Golgi, called Cell Death Inducing DFFA Like Effector B (CIDEB) facilitates lateral diffusion of TG from cLDs to the ER membrane at contact sites between these organelles, as well as within the ER membrane, from the site of lipid synthesis to apoB. CIDEB deficiency in mice reduces TG secretion no change in apoB secretion, leading to secretion of TG-poor lipoproteins and increased TG mass in cLDs. Ancient ubiquitous protein 1 (AUP1) is another protein associated with both the ER and cLDs; in contrast to CIDEB, AUP1 knockdown in HepG2 cells increased VLDL production and reduced the number of cLDs. AUP1 has been found to control protein content of cLDs via increasing the ubiquitination of ER- and cLD-associated proteins. AUP1 is thought to antagonize CIDEB activity during the early stages of apoB lipidation. Thus, relative changes in the cellular levels of these two proteins may regulate lipoprotein assembly, although data from additional models is needed to confirm these findings in vivo.

Proteins associated with other organelles also appear to be involved in the availability of lipid substrate for VLDL assembly. Work by Shin et al, in collaboration with the Ginsberg lab, uncovered an unanticipated relationship between the outer nuclear membrane protein Lamina Associated Polypeptide 1 (Lap1) and VLDL secretion. They showed that knockout of Lap1 was associated with reduced VLDL secretion, while lipid droplets accumulated in the nucleus as well as the cytosol. Further, when the ER luminal AAAatpase family member TorsinA, which requires Lap1 activation of its ATPase activity, was knocked out in mice, VLDL secretion was nearly completely ablated and lipid droplets appeared to accumulate in the ER rather than the cytoplasm, indicating an impairment in the lipidation of apoB or in the exit of assembled VLDL.
from the ER. In a similar manner, when apoB synthesis was inhibited by antisense oligonucleotides (ASOs) in mice, the continued function of MTP resulted in lipid accumulated in the ER, causing ER stress and eventually triggering ER autophagy. Much of this ER-localized lipid was targeted to the mitochondria for β-oxidation, and thus limited hepatosteatosis was observed, despite marked reductions in VLDL-apoB and -TG secretion.

Another point at which VLDL secretion is regulated occurs during the trafficking of the particles from the ER to the Golgi. Although the COPII vesicular transport system for ordinary secreted proteins has been extensively characterized, much uncertainty regarding the cellular itinerary of vesicles carrying VLDL cargo remains. Wang et al recently showed that SAR1B, which is known to play an early and critical role in the formation of very large COPII vesicles that transport chylomicrons from the ER to the Golgi in enterocytes, is also essential for VLDL vesicle transport in hepatocytes. They further identified SURF4, a multi-span trans-membrane cargo receptor, as a novel SAR1B partner and necessary component of COPII vesicles that carry VLDL out of the ER. SURF4 also seemed involved in the retrieval of these vesicles from the Golgi back to the ER. Liver-specific knockout of either of these proteins in mice resulted in significant defects in TG and apoB secretion, hepatic steatosis, and reduced plasma levels of TG, cholesterol, apoB, and apoA-I. Additionally, SURF4 was shown to mediate the protection of mice with loss of function of both SAR1B and SURF4 from PCSK9-induced atherosclerosis. Previous studies have also identified TANGO1 and KHLH12 as additional proteins that enable COPII vesicles to expand to a size needed for the transport of VLDL.

1.5 NAFLD: clinical presentation, prevalence, and etiology

Non-alcoholic fatty liver disease (NAFLD) is a generic name for a spectrum of progressive liver diseases ranging from simple steatosis, non-alcoholic steatohepatitis (NASH), to fibrosis, and
ultimately to cirrhosis (Figure 1.4).\textsuperscript{5} NAFLD is characterized by increased hepatic triglyceride content (HTGC) in the absence of secondary causes of hepatic fat accumulation, such as significant alcohol consumption, viral infection, or use of a steatogenic medication.\textsuperscript{14} The latter include the common antibiotic tetracycline or anti-arrhythmic agent amiodarone.\textsuperscript{41} The simplest and nonprogressive category of NAFLD, non-alcoholic fatty liver (NAFL), is defined by the presence of macrovesicular steatosis in $\geq$5\% of hepatocytes, as determined by imaging or histology, and the absence of hepatocellular injury. The more clinically significant category, NASH, includes hepatic steatosis as well as hepatocellular injury (in the form of ballooning) and lobular inflammation, which can progress to cirrhosis.\textsuperscript{14} Fibrosis may or may not be present in NASH, but is the defining feature of cirrhosis, which is mainly irreversible and, if untreated, can cause end stage liver disease and liver-related mortality.\textsuperscript{42} Additionally, patients with NASH are at greater risk of developing hepatocellular carcinoma (HCC), which is the fourth leading cause of cancer death worldwide; in fact, NAFLD is the fastest growing cause of HCC in many countries, including the USA and parts of Europe.\textsuperscript{43} The global prevalence of NAFLD is estimated at 25\% and is projected to become the leading cause of HCC and end-stage liver disease, for which the only current treatment is organ transplantation.\textsuperscript{44}
Figure 1.4: Depiction of the spectrum of Non-alcoholic fatty liver disease. Non-alcoholic fatty liver (NAFL) is defined by the presence of steatosis, or greater than 5% fat in the liver. This condition occurs in 15-30% of the population and is reversible. NAFL can progress to Non-Alcoholic Steatohepatitis (NASH), which is characterized by steatosis in addition to inflammation, hepatocyte ballooning, and fibrosis. Cirrhosis and Hepatocellular carcinoma encompass end-stage liver disease, which is irreversible, and for which the only current treatment is liver transplantation. Created with BioRender.com

Despite the widespread prevalence of NAFLD, uncertainty regarding its molecular basis has significantly hindered the development of additional effective therapies. The severity and progression of NAFLD can be affected by a number of factors, ranging from hereditary conditions such as lipodystrophy, abetalipoproteinemia, and cholesterol ester storage disease, to small intestinal bacterial overgrowth and jejunoileal bypass. The most common risk factors, however, are excess central adiposity and insulin resistance, and their associated factors, such as type 2 diabetes, excessive intake of highly processed fructose, and lack of physical activity.44

Although obesity is the major risk factor for hepatic steatosis, liver fat content varies widely among obese individuals; additionally, many individuals with simple hepatic steatosis do not progress to steatohepatitis or cirrhosis.45 Eslam et al. presented several lines of evidence from twin studies, multi-ethnic population studies, and family studies, all of which indicated that inherited features contribute to NAFLD, and that 25%–50% of hepatic fat variability could be accounted for
by heritable factors. Additionally, an interaction between major genetic risk variants and established environmental risk factors for NAFLD has been observed, suggesting that NAFLD should be considered a complex trait in which the disease susceptibility and phenotype are modulated by the interaction of environmental exposures and a polygenic background. This further implies that specific effective therapies will be needed to target NAFLD within relevant populations or individuals.

The advent of genome wide association studies (GWAS) more than 15 years ago allowed for the investigation into, and the ultimate discovery of, major genetic determinants of complex polygenic diseases. The first such GWAS to uncover major genetic determinants of NAFLD, conducted by Romeo et al, identified the non-synonymous SNP rs738409 (Ile148Met) in the patatin-like phospholipase domain containing 3 (PNPLA3) gene; the association between this SNP and increased risk for NAFLD has since been robustly validated across multiple patient cohorts. Further, carriage of this SNP has been reproducibly associated with clinical severity, of hepatic fibrosis/cirrhosis and the development of HCC. The incidence of the rs738409 SNP in the original study population was 0.49 in people of Hispanic ancestry, while only 0.23 in European Americans and 0.17 in African Americans. These frequencies were concordant with the relative prevalence of NAFLD in these ancestry groups, suggesting that carriage of this variant could account for much of the interethnic variability in NAFLD across different ethnic groups. Importantly, the I148M variant in PNPLA3 is the most clinically important genetic modifier of NAFLD.

Subsequent functional studies have elucidated the molecular function of PNPLA3 and characterized the mechanism by which it influences hepatic lipid accumulation, although some controversy remains. Hobbs, et al discovered that PNPLA3 is an enzyme localized to the surface
of cytosolic lipid droplets (cLDs) in hepatocytes and hepatic stellate cell, which demonstrates hydrolase activity for TG and retinyl esters; PNPLA3 may also possess transacetylase activity to incorporate polyunsaturated fatty acids into phospholipids.\textsuperscript{57} They also found that the I148M mutation interferes with the ubiquitinylation and degradation of PNPLA3 by the proteasome, causing it to instead accumulate on cLDs, where CGI-58, an essential cofactor for the adipose triglyceride lipase (ATGL), becomes sequestered, thus leading to an impairment in LD turnover.\textsuperscript{58} Genetic variation that results in lower PNPLA3 expression seems to protect against NAFLD in carriers of the I148M mutant.\textsuperscript{59} Furthermore, pharmacological downregulation of PNPLA3 ameliorates hepatic steatosis in both wild-type and mutant mice.\textsuperscript{60} In fact, PNPLA3 is a target for innovative therapies of NAFLD, demonstrating the clinical value of GWAS-identified genetic candidates.\textsuperscript{61} 

Given the success of the identification of PNPLA3 as a strong genetic modifier of NAFLD risk, the GWAS approach to finding additional genetic determinants of metabolic diseases was adopted by many researchers. Another SNP found to be highly associated with accumulation of hepatic fat, as well as the entire spectrum of NAFLD, including HCC, is the rs641738 loss-of-function variant in the \textit{MBOAT7} gene.\textsuperscript{62} This gene encodes the phospholipid remodeling enzyme lyso-phosphatidylinositol acyl-transferase1 (LPIAT1), which catalyzes the esterification of a fatty acyl-CoA to lysophosphatidylinositol (LPI) to generate phosphatidylinositol (PI). LPIAT1 primarily localizes to the surface of lipids droplets, the ER membrane, and mitochondria-ER associated membranes (MAMs), which are physical and functional bridges between the two organelles, at which lipid biosynthesis and lipid droplet formation are enriched.\textsuperscript{44} Importantly, LPIAT1 shows a strong substrate preference for the polyunsaturated arachidonyl-CoA, producing PI species with increased desaturation, as well as decreasing the overall available free arachidonic
acid (AA), which serves as a precursor to proinflammatory eicosanoids.\textsuperscript{62} Recent findings suggested that decreased hepatic LPIAT1 function resulted in the increased saturation of phospholipids (specifically PI), which in turn led to the increased synthesis of TG through a non-canonical pathway.\textsuperscript{63} While much remains to be learned about the pathways through which defects in LPIAT1 lead to NAFLD, these observations highlight an important relationship between the enzymes that impact membrane composition and the progression of NAFLD phenotypes.

1.6 Human genetics implicate TM6SF2 Loss of Function in NAFLD, hypolipidemia, and reduced CVD risk

Between 2008-2013, 6 separate GWAS identified a group of variants in a large region on chromosome 19 associated with plasma TG and cholesterol concentrations\textsuperscript{64-69}, as well as nonalcoholic fatty liver disease\textsuperscript{51,70}, coronary heart disease\textsuperscript{68,71}, and diabetes mellitus\textsuperscript{72,73}. This locus, termed 19p12 or NCAN-TM6SF2-SUGP1-CILP2, contains more than a dozen genes, none of which had previously been implicated in lipid metabolism. An effort to identify the gene(s) in the 19p12 locus driving the association with the observed phenotypes was undertaken independently by a number of groups.

To distinguish the causal variant at this locus, Kozlitina and colleagues used genotyping arrays to perform an exome-wide association study in a multi-ancestry cohort from the Dallas Heart Study (DHS).\textsuperscript{54} Over 130,000 sequence variants were tested in 2,736 DHS subjects for association with proton magnetic resonance spectroscopy (\textsuperscript{1}H-MRS)-quantified hepatic triglyceride content (HTGC). A non-synonymous SNP in coding nucleotide 499 of the \textit{TM6SF2} gene (rs58542926 c.449 C>T), which substitutes the glutamate at residue 167 with lysine (denoted E167K), emerged as the third most significant variant associated with HTGC; the top two most significant SNPs were located in the \textit{PNPLA3} gene. Additionally, they found that the TM6SF2
variant encoding E167K was associated with low plasma levels of TG and LDL-C, but not HDL-C, in the DHS, Dallas Biobank, and Copenhagen Study cohorts.

Holmen, et al, took a similar approach, systematically assessing genome-wide coding variation to fine map the NCAN-TM6SF2-SUGP1-CILP2 locus and identify the gene influencing the observed association with plasma lipid traits. They found that the same E167K variant in the TM6SF2 gene reached the threshold for genome-wide significant inverse association with total plasma cholesterol. Further, Holmen and colleagues observed a suggestive inverse association for the TM6SF2 rs58542926 coding variant with incidence of myocardial infarction, pointing to a clinically meaningful and protective effect from carriage of the SNP. Notably, they did not observe a significant association between any TM6SF2 variants and LDL-C, HDL-C, or plasma TG within their study parameters; however, an in silico analysis of a European database containing genotyped samples supported an inverse association between carriage of the rs58542926 TM6SF2 variant and total plasma cholesterol, LDL-C, and TG levels. In agreement with these findings, Mahdessian et al observed a significant, positive relationship between human hepatic expression of WT TM6SF2 and levels of plasma TG, suggesting that TM6SF2 is the putative functional gene in the 19p12 locus responsible for the observed relationships with plasma TG levels.

Since the publication of these landmark studies, many additional groups have corroborated the association of the TM6SF2 E167K substitution with liver steatosis and plasma lipids. In a bariatric cohort of individuals with extreme obesity, O’Hare et al found that participants with at least one copy of the rs58542926 variant exhibited a higher average steatosis grade than noncarriers. Additionally, they observed lower fasting serum levels of TC, TG, and LDL-C in carriers of the minor allele, although this association was only statistically significant in the subgroup also exhibiting NAFLD. Another study in patients undergoing laparoscopic bariatric
surgery confirmed these findings: based on histological evidence from liver biopsies, the macrovesicular steatosis was 2-fold higher in the TM6SF2 EK/KK carriers than in the TM6SF2 EE group. This finding correlated with the sum of all liver TGs measured by UHPLC-MS. Additionally, liver CE was 20% higher in the TM6SF2 EK/KK group than in controls. Serum lipidomic analysis in this cohort revealed lower plasma TG and PC levels in variant allele carriers as compared to non-carriers, while plasma CE levels were not different. Finally, Boren et al also detected (by proton magnetic spectroscopy) more than a 2-fold higher fat content in the livers of 10 homozygous minor allele carriers, compared to 10 WT matched controls.

A number of subsequent association studies, specifically within small groups of patients with histologically confirmed NAFLD, observed a significant correlation between carriage of the rs58542926 variant and severity of steatosis/NAFLD. Within one of these cohorts, patients with CT/TT genotypes, in comparison with those with the CC genotype, had lower levels of TC and a lower index of cardiovascular risk. Further, Liu et al sought to determine whether the impact of TM6SF2 genotype was limited to risk for the relatively benign steatosis or had broader clinical relevance. They found that carriage of the rs58542926 variant was also associated with liver fibrosis, severity of steatohepatitis, and risk of NAFLD-HCC. Although striking, more evidence is needed to confirm the association of TM6SF2 with risk for the full spectrum of NAFLD.

Taken together, the findings from these studies suggest that carriage of the minor TM6SF2 allele confers increased risk for hepatosteatosis and protection for hyperlipidemia and CVD more broadly. While for the majority of people, NAFLD is associated with an increased risk of cardiovascular disease, this relationship is dissociated in TM6SF2 minor allele carriers, who appear to be at greater risk for hepatic, rather than cardiovascular, associated morbidity and mortality. The broad body of initial clinical data led Kahali and colleagues to propose that wild
type TM6SF2 might promote hepatic VLDL secretion, protecting the liver at the expense of increased CVD risk, while mutant TM6SF2 would result in hepatic retention of TG and cholesterol, predisposing to progressive NAFLD.\textsuperscript{82} Thus, a better understanding of TM6SF2 could provide insights important for approaches to prevent both NAFLD and CVD.

1.7 Current understandings of TM6SF2 function

Many groups have subsequently carried investigations deeper into the mechanism by which TM6SF2 impacts hepatic lipid metabolism; the following section summarizes the key pieces of information that have been gleaned from these studies, as well as the inconsistencies that have arisen.

\textit{Discovery and proposed structure:}

In 2000, researchers published their discovery of a novel gene in the 15q24→q26 region lacking homology to any other known gene.\textsuperscript{83} Amino acid sequencing together with protein pattern and domain predictions revealed six indicatory transmembrane domains (TMDs), plus two additional TMDs identified with lower significance; therefore, it was named transmembrane 6 super-family 1 (TM6SF1). BLAST database searches using the TM6SF1 sequence produced a significant hit with a homologous putative gene on 19p12. This gene, predicted to encode a protein of 351 amino acids and 39.5-kDa, shared 68\% similarity and 52\% identity to TM6SF1 and exhibited the same number of predicted transmembrane domains, so it was named TM6SF2.

Other than these findings, no other protein features were detected that could help to elucidate the function of this pair of genes.\textsuperscript{83} Since its discovery, little else has been established regarding the higher order structure of the TM6SF2 peptide; it is currently predicted to contain 7-10 transmembrane domains, while NCBI denotes it as having nine transmembrane domains. In an effort to uncover clues to the function of TM6SF2, Ponting and Sanchez performed computational
peptide sequence analysis of the TM6SF protein family (1&2). They discovered a novel domain containing four transmembrane helices that is present twice in TM6SF proteins and that is also conserved in Emopamil Binding Protein (EBP) and Transmembrane Protein 97 (TMEM97) protein families (Figure 1.5, Right). The domain, which they termed “EXPERA” for EXPanded Emopamil binding protein superfamily, contains the active site of EBP, a Δ8, Δ7-sterol isomerase critical for the conversion of zymosterol to 5-a-cholesta-7,24-dien-3b-ol in the cholesterol biosynthetic pathway. Based on this sequence homology, the authors hypothesized that TM6SF2 may play a role in sterol metabolism; this hypothesis will be discussed in greater detail in Chapter 4.

Tissue Distribution:

Relative tissue distribution analysis of the TM6SF2 gene in mouse revealed highest expression in the small intestine, liver, and kidney; moderate expression in colon; and lower expression levels in other tissues. Endogenous TM6SF2 protein was also highly expressed in mouse liver, but was found to be 10-fold higher in jejunum than in liver by Smagris, et al. Evaluation of the expression of TM6SF2 in human tissues revealed substantial TM6SF2 mRNA levels in liver and intestine, while all other tissues analyzed showed low expression levels (Figure 1.5, Left). To date, evaluation of TM6SF2 protein distribution in human tissues has not been performed, although presence of TM6SF2 protein in human liver biopsies was confirmed by immunohistochemistry. Notably, the TM6SF2 tissue expression pattern mirrors that of apoB, which provides a clear functional link between these proteins.

Subcellular Localization:

Confocal microscopy was used to study the subcellular localization of TM6SF2 in human hepatoma Huh7 and HepG2 cells; these studies revealed significant colocalization of TM6SF2
with two ER proteins, as well as a marker for the ER-Golgi Intermediate Compartment (ERGIC). Notably, TM6SF2 did not significantly colocalize with the Golgi-resident protein Giantin, indicating that TM6SF2 localizes to the ER and ERGIC membranes, but not to the Golgi (Figure 1.5, Middle). These findings were validated by confocal imaging in primary mouse hepatocytes as well as by western blotting in Hepa1c1c7 cell lysate fractions. Additionally, no evidence was found for the co-localization of TM6SF2 in other membranous structures, including plasma membrane, mitochondria, and lipid droplets in Huh7 and HepG2 cells. Cell fractionation studies confirmed that mouse TM6SF2 protein also did not colocalize with lipid droplets. In a recent study, Hobbs and colleagues found that TM6SF2 localized to the smooth ER (SER) and ERGIC, but not the Golgi, in rat liver. To date, any impact of the K167 substitution on TM6SF2 subcellular localization has yet to be explored.

**Regulation of expression:**

Little information is available regarding the transcriptional regulation of the *TM6SF2* gene. Recently, however, ChIP-qPCR analysis of mouse liver identified associations between the *TM6SF2* promoter and two well-known transcription factors (TF). The first, Carbohydrate Response Element Binding Protein (ChREBP), is a glucose-activated TF that strongly regulates glycolytic and lipogenic pathways, impacts hepatic VLDL secretion, and has been implicated in the development of the MetS. The second, Hepatocyte Nuclear Factor 4α (HNF4α), belongs to the steroid/thyroid hormone receptor superfamily and regulates the expression of glycolytic enzymes, glucose transporters, and lipid homeostatic proteins, such as MTP, in the liver; additionally, HNF4α can bind medium to long chain FAs and their metabolites that are capable of modulating its transcriptional activity. In another study, Tm6sf2 mRNA abundance was found
to be about 50% higher in mouse female (WT) livers, compared to males, which might contribute to the dimorphic effects of Tm6sf2 deletion described later in this chapter.\textsuperscript{94}

The impact of TM6SF2 genotype on expression levels also remains ambiguous. Early studies in humans indicated that carriage of the minor allele was associated with reduced \textit{TM6SF2} transcript levels. In 206 human liver biopsies, Mahdessian observed that presence of the minor allele of rs10401969 (the GWAS index SNP in the \textit{TM6SF2} gene demonstrating near complete linkage disequilibrium with rs58542926 \( [r^2=0.97] \))\textsuperscript{74,75} was associated with lower \textit{TM6SF2} mRNA levels.\textsuperscript{75} Three other groups independently found a significant association between carriage of the rs58542926 minor allele (EK/KK versus EE) and lower \textit{TM6SF2} mRNA expression in human liver samples.\textsuperscript{77,79,86} Despite this evidence, overexpression models - in which equal doses of plasmids carrying either the WT or mutant \textit{TM6SF2} transgene were administered - did not show differences in relative mRNA levels, indicating that the SNP itself is unlikely to impact \textit{TM6SF2} transcriptional activity or mRNA stability;\textsuperscript{54,94–97} however, this does not eliminate the possibility that endogenous \textit{TM6SF2} expression may be altered depending on C/T genotype, perhaps by some sort of feedback mechanism, which is not applicable in exogenous expression systems.

Despite these questions, investigations into the impact of TM6SF2 genotype on its subsequent protein levels have consistently demonstrated that carriage of the rs58542926 minor allele is associated with decreased expression of the mutant protein (versus WT) in human liver biopsies\textsuperscript{79}; transiently transfected\textsuperscript{54,96,97} and stably transduced\textsuperscript{95} Huh7 cells; and transgenic mice\textsuperscript{94}. Kozlitina, \textit{et al} was the first to show that Huh7 cells transfected with the K167-transgene displayed 46\% less TM6SF2 protein, compared to those transfected with the WT-transgene.\textsuperscript{54} To explore the mechanism responsible for this observed reduction, Ehrhardt \textit{et al} used cycloheximide (CHX), an inhibitor of protein synthesis, to examine the degradation of luciferase-TM6SF2 fusion proteins in
Huh7 cells.\textsuperscript{95} Consistent with reduced steady-state expression, the K167 isoform exhibited a significantly increased rate of degradation compared to the WT isoform, establishing that the rs58542926 variant is indeed a loss-of-function SNP, which causes reduced TM6SF2 protein expression through increased turnover rate. Li \textit{et al} later confirmed that K167 TM6SF2 was subjected to increased degradation: they reported that the treatment of Huh7 cells, expressing the mutant transgene, with the proteasomal inhibitor MG132 rescued the reduced K167 TM6SF2 protein expression to WT levels.\textsuperscript{97} Hence, there appears to be a consensus that the K167 variant causes increased proteasomal degradation of the TM6SF2 protein, although the mechanism is unknown (\textbf{Figure 1.5}, Right); Ehrhardt \textit{et al} speculated that perhaps K167 disrupts the interaction of TM6SF2 with other protein(s) that are critically required for its own stability; TM6SF2 interacting partners will be discussed in further detail in Chapter 4. Whether the E167K substitution additionally impacts the inherent function of the protein (beyond altering its stability) remains to be investigated as well.

\textbf{Figure 1.5.} TM6SF2 tissue distribution, subcellular localization and proposed structure. Left: TM6SF2 is most highly expressed in the liver and small intestine. Middle: TM6SF2 localizes to the ER and ER-Golgi Intermediate Compartment (ERGIC). Right: The TM6SF2 polypeptide is predicted to contain 7-10 transmembrane domains and two instances of the EXPanded Emopamil binding protein superfamily (EXPERA) domain. Presence of lysine (K) at amino acid position 167 is thought to result in increased degradation of the TM6SF2 protein by the proteasome. Created with Biorender.com
In 2014, Holmen et al and Kozlitina et al performed the first transient knockdowns (KD) of mouse Tm6sf2 with liver-targeted virally delivered shRNA. While Kozlitina found a 3-fold increase in hepatic TG in KD mice (which was exacerbated when the mice were fed a high sucrose diet)\textsuperscript{54}, Holmen did not observe any evidence of TG accumulation in livers of KD mice on chow, compared to controls\textsuperscript{74}. Smagris and Fan both generated whole body Tm6sf2 knockout (KO) mice in 2016 and assessed liver steatosis via biochemical assays and oil Red O (ORO) staining for neutral lipids.\textsuperscript{45,85} Similarly, whereas Smagris found that levels of TG and CE were increased in livers of chow-fed KO male mice (which was further pronounced in female KO mice), Fan did not observe any changes in liver TG or cholesterol accumulation in mice fed a chow or high fat diet (HFD). Additionally, Smagris observed an increase in the number and the median size of LDs in the livers of KO animals. Finally, in 2021, Newberry et al published their findings from liver-specific Tm6sf2 KO (LKO) mice.\textsuperscript{94} They reported an increase in hepatic TG, noting a similar sexual dimorphism where the phenotype was more pronounced in female mice, but did not observe any difference in hepatic cholesterol levels; these observations were consistent between chow- and HFD-fed animals. The LKO mice also displayed increased numbers of large LDs and no change in small LDs, compared with controls.

Several groups overexpressed TM6SF2 in mice as well. Just as with their Tm6sf2 KD mice, Holmen et al did not observe any changes in hepatic TG when they transiently overexpressed the human TM6SF2 gene in WT mice.\textsuperscript{74} Fan et al also created mice stably overexpressing the human TM6SF2 gene under a liver-specific promoter; they observed that, although there were no obvious histological changes, the TG and cholesterol content in the livers of the transgenic mice were higher than controls on chow; this difference became more striking in mice on HFD.\textsuperscript{6} In agreement
with these findings, Ehrhardt et al also found that over expression of the mouse Tm6sf2 gene in mice via AAV8 vector promoted increased liver TG and ORO staining, but led to no changes in liver cholesterol.\textsuperscript{95} Alternatively, Newberry found that “rescuing” mouse Tm6sf2 (also via AAV8 vector) in their LKO mice led to a reduction in hepatosteatosis and the size of cLDs, compared to LKO mice; notably, this effect was observed when both the WT and E167K mutant Tm6sf2 transgenes were expressed.\textsuperscript{94}

These data suggest that exogenously expressed TM6SF2 causes a loss-of-function of the protein and parallels the phenotypic consequences of TM6SF2 deficiency, while introduction of the transgene into mice lacking endogenous TM6SF2 expression leads to a rescue of the KO phenotype. Ehrhardt et al noted that similar phenocopy in models of overexpression and deficiency can be seen with proteins that are part of multiprotein complexes where function is impaired by disruption of normal stoichiometry They further suggested that normal function of Tm6sf2 may therefore depend on interactions with other proteins, and both excess and deficiency of Tm6sf2 could disrupt the formation of a complex required for normal cellular function.\textsuperscript{95}

\textit{Mouse In Vivo Studies – plasma lipids}

Kozlitina et al observed significantly decreased plasma levels of TC and TG in transient knockdown mice on a chow diet after a 4-h fast. Further, fast-performance liquid chromatography (FPLC) fractionation of plasma lipoproteins isolated after a 4-h fast showed reduced content of cholesterol in the LDL and HDL fractions , as well as TG in the VLDL, in mice with knockdown of TM6SF2.\textsuperscript{54} In agreement with these findings, Holmen et al also reported that fasting TC levels were decreased by 18.2\% in Tm6sf2 transient knockdown mice compared to controls; however, they did not observe any differences in plasma TG, LDL-C, or HDL-C.\textsuperscript{74} Similarly, the whole body Tm6sf2 KO mice generated by Smagris, Fan, and Li all exhibited reduced plasma TC on
chow, reflecting reductions in LDL-C and HDL-C.\textsuperscript{45,85,97} However, whereas Li and Smagris respectively observed that Tm6sf2 KO mice had lower plasma TG and VLDL-TG compared with WT mice, Fan reported slightly \textit{increased} plasma TG levels in their KO mice; additionally, the increase in plasma TG was more prominent in mice fed a HFD. Finally, the L-KO mice generated by Newberry, \textit{et al} displayed reduced abundance of almost all TG species measured via mass spectrometry, as well as a 75\% reduction in the secretion of $^{3}$H-TG from primary hepatocytes isolated from these mice.\textsuperscript{94}

Conversely, mice overexpressing human Tm6sf2 consistently displayed increased plasma TC and LDL-C; further, Holmen also observed increased plasma TG, whereas Fan did not observe any change in plasma TG in chow-fed mice.\textsuperscript{74,85} In stark contrast to these findings, Ehrhardt \textit{et al} demonstrated that elevated hepatic expression of mouse Tm6sf2 suppressed plasma TG and cholesterol levels in their chow and high-cholesterol diet-fed mice, mirroring the phenotype observed in Tm6sf2 KO mice.\textsuperscript{95} Interestingly, when Newberry, \textit{et al} rescued the L-KO mice with a virally delivered mouse Tm6sf2 transgene, they observed only a partial restoration of VLDL-TG secretion, independent of Tm6sf2 genotype (E167K).\textsuperscript{94} Overall, these rodent \textit{in vivo} studies strongly suggest that lower TM6SF2 expression leads to reduced plasma cholesterol and possibly reduced plasma TG in mice, aligning with the human clinical phenotype, whereas induced TM6SF2 expression also impacts plasma lipid traits, though the directionality of this effect is thus far unclear.

Taken together, these findings demonstrate that mouse models of disrupted TM6SF2 expression recapitulate the main clinical phenotypes of rs58542926 carriage, namely hepatosteatosis and reduced plasma lipids. Investigators next sought to determine the molecular function of TM6SF2. Noting that the tissue expression pattern of TM6SF2 mirrors that of apoB,
and that models of reduced TM6SF2 expression phenocopy models of disrupted VLDL synthesis and secretion, the next studies aimed to uncover whether TM6SF2 plays a role in this pathway.

**Mouse In Vivo Studies – VLDL synthesis and secretion**

Smagris found no differences in steady state apoB-48 or apoB-100, detected by immunologic methods in liver lysates of KO and WT mice, as well as similar levels of plasma apoB100; however, the plasma levels of apoB-48 were significantly higher in Tm6sf2-/- mice.\textsuperscript{45} In contrast, Li et al found that hepatic apoB100 and apoB48 protein expression was decreased in mice with global deficiency or hepatic knockdown of Tm6sf2; they attribute this discrepancy to procedure differences between the studies.\textsuperscript{97}

Following injection with a detergent (Triton or Pluronic) to inhibit lipoprotein lipase activity, the rate of TG secretion was consistently lower in KD\textsuperscript{54}, L-KO\textsuperscript{94}, and whole body KO\textsuperscript{22} mice, when compared with WT littermates, indicative of defective VLDL-TG secretion. To interrogate the question of whether there was a defect in whole particle secretion in TM6SF2 KO mice, Smagris et al injected mice with \textsuperscript{35}S]methionine and Triton WR-1339 and measured the appearance of radiolabel apoB-48 and apoB-100 in the circulation. Similar rates of apoB-100 secretion were observed in WT and KO mice, whereas apoB48 secretion was modestly elevated in KO mice.\textsuperscript{45} Newberry et al reported that newly synthesized cell apoB100 from hepatocytes isolated from WT and L-KO mice, as determined by incorporation of the radiolabeled \textsuperscript{35}S-methionine, did not significantly differ.\textsuperscript{94}

**Human Cell Culture – hepatic steatosis**

The function of TM6SF2 has also been explored in a number of (human) hepatocyte models. One of the first functional studies, performed by Mahdessian, et al, found a significant increase in cellular TG in Huh7 and HepG2 hepatoma cell lines following transient knockdown of
TM6SF2, as well as a ~two-fold increase in lipid droplet area.\textsuperscript{75} These observations were confirmed later by Luukonen and colleagues, who also observed increased lipid droplet area in their transiently and stably silenced TM6SF2 KD Huh7 lines, as compared to controls\textsuperscript{77}, and by Ruhanen et al, who reported increased TG and CE concentrations in Huh7 hepatocytes with TM6SF2 stably knocked down\textsuperscript{98}. When PHHs were cultured as 3D spheroids and supplemented with high concentrations of FAs, Prill et al observed higher intracellular lipids in the spheroids derived from donors harboring the E167K mutation, as compared to age and gender matched WT controls.\textsuperscript{99} These \textit{in vitro} studies consistently suggest that disruption of TM6SF2 expression, either through E167K-mediated degradation or through genetic silencing, leads to an accumulation of lipid, in particular TG, in hepatocytes.

The data from TM6SF2 overexpression in cell culture models exhibit less consistency. When Mahdessian, \textit{et al} transiently overexpressed human \textit{TM6SF2} in Huh7 cells, they observed a decrease in the number and size of lipid droplets compared to control cells, suggestive of a gain of function.\textsuperscript{75} However, Ehrhardt et al observed a modest, but significant elevation of intracellular TG in Huh7 and HepG2 cells stably overexpressing TM6SF2, while cholesterol content remained unchanged.\textsuperscript{95} Recently, Pant et al showed that the overexpression of the E167K variant recapitulated the effects of knockdown of TM6SF2 on hepatic lipid accumulation and lipid droplets in Huh7 cells, suggesting that the E167K protein may have a dominant negative effect on hepatic lipid metabolism.\textsuperscript{100}

\textit{Human Cell Culture – VLDL synthesis and secretion}

Transient transfection with TM6SF2 siRNA allowed functional analysis of the protein in human hepatoma Huh7 and HepG2 cell lines. TM6SF2 siRNA inhibition led to a reduced secretion of TG by both Huh7 and HepG2 cells (as measured by the appearance of \textsuperscript{14}C-glycerol-labeled TG
in the media), whereas more modest reductions in the secretion of apoB (measured by ELISA) was observed.\textsuperscript{75} Using primary human hepatocyte spheroids from EK donors, Prill also found reduced immunodetectable apoB-100 in cells and media, compared to EE donors.\textsuperscript{99} These results contrast the observations of Ruhanen, \textit{et al}; in media collected from their Huh7 TM6SF2 KD cells, they observed a modest increase in ELISA-measured apoB100, compared to WT.\textsuperscript{98}

**Human In Vivo VLDL-secretion**

Few studies have directly assessed the effect of TM6SF2 genotype on apoB metabolism. Kim, \textit{et al} identified an independent association between carriage of the rs58542926 minor allele and lower concentration of total plasma apoB-100 in a cross-sectional study of in 6,929 Finnish men.\textsuperscript{101} Additionally, they observed a reduction in the mean diameter of VLDL particles in minor allele carriers. Boren and colleagues used compartmental modeling of stable-isotope tracer data to determine production and clearance of apoB100 and triglycerides in the 2 major VLDL subfractions: large triglyceride-rich (VLDL\textsubscript{1}) and smaller, less triglyceride-rich (VLDL\textsubscript{2}) in 10 homozygote TM6SF2 E167K carriers and 10 matched controls.\textsuperscript{78} They reported that VLDL\textsubscript{1}-apoB100 and VLDL\textsubscript{1}-TG production were lower in the TM6SF2 E167K carriers compared with controls, while the direct production rates for VLDL\textsubscript{2}-apoB100 and TG were similar in carriers and controls. They concluded that there was a specific defect in the assembly and secretion of large TG-rich VLDL\textsubscript{1} particles in individuals with the TM6SF2 E167K genetic variant.

VLDL particle size is an important metric for understanding the degree to which apoB-containing lipoproteins are lipidated. Despite contradictions in the literature regarding whether the \textit{number} of apoB100-containing particles secreted from hepatocytes is affected by TM6SF2, there appears to be a consensus thus far that carriage of the K167 variant or decreased expression of WT TM6SF2 result in a reduction in average VLDL \textit{particle size}. This observation was noted in a
variety of models, including human in vivo studies, Huh7 cultures, global KO mice, and L-KO mice. These findings suggest that, rather than compromising the secretion rate of VLDL particles, the absence of TM6SF2 reduces the lipidation of each particle, thus causing the observed hypolipidemia. This distinction indicates that TM6SF2 is not required for the assembly of the nascent, lipid poor VLDL, or for the transport of those lipid poor VLDL from the ER to the Golgi and then to the plasma membrane for secretion. Rather, it appears that TM6SF2 is required for the addition of neutral lipid to the nascent VLDL, allowing more TG to be transported out of the liver on each VLDL particle.

Although much progress has been made to understand the function of TM6SF2 and the mechanisms through which it impacts hepatic and plasma lipids, many questions remain unresolved. Of particular interest is the question of whether TM6SF2 alters the number and/or size of apoB-containing particles in humans. Additionally, the impact of TM6SF2 overexpression remains unclear, with some groups reporting that it resembles the phenotype resulting from TM6SF2 disruption, while others observed the opposite. The differences in models, including species, in study protocols, and in the degree of TM6SF2 knockdown or overexpression might be the basis for these inconsistent results. Therefore, successfully elucidating the function of TM6SF2 requires the use of an improved hepatic model.

### 1.8 Assessment of popular hepatocyte models for studying NAFLD

The liver is the center of metabolism in the human body, performing over 500 vital functions to support digestion, maintain metabolic homeostasis, detoxify xenobiotics, and regulate immune responses. Hepatocytes are the major parenchymal cells of the liver, making up 80% of the organ’s mass and appearing morphologically as irregular, polyhedral, often binucleated cells. Similar to other epithelial cells, hepatocytes are polarized, with distinct domain
structures and protein distributions within the plasma membrane.\textsuperscript{105} The basal surface of the hepatocyte membrane faces the hepatic sinusoid where macromolecule exchange between the cell and circulating blood takes place; the lateral surface engages in hepatocyte-hepatocyte contacts, which are crucial for maintaining proper liver architecture; finally, the apical surfaces are delineated from the basolateral membranes by barriers of tight junctions and form a narrow lumen, or canaliculus, into which bile and other metabolic products are secreted by the adjacent hepatocytes.\textsuperscript{104–106} This multipolar organization is essential for nutrient exchange, protein trafficking, and cytoskeletal dynamics, which in turn contribute to the highly diverse metabolic and synthetic functions carried out by hepatocytes (Figure 1.6).\textsuperscript{107,108}

Figure 1.6: Hepatocyte membrane polarity and essential functions: 1. Hepatocytes store excess glucose as glycogen, which serves as an energy depot. 2. Bile acids are synthesized from cholesterol and shuttled across the apical membrane into the bile canaliculus via ATP-dependent ABC transporters. 3. The liver synthesizes and secretes a host of essential proteins across the basolateral membrane, including albumin, the most abundant circulating protein. 4. Ammonium ions, products of amino acid catabolism, are converted by the liver into urea and secreted across the basolateral membrane for transport to the kidneys. 5. & 6. Hepatocytes take up most of the body’s LDL particles via receptor-mediated endocytosis following binding to the LDLR on the basolateral membrane. The cholesterol released from these particles can
have multiple fates, including excretion into bile and storage in neutral lipid droplets with excess triglyceride, surrounded by an amphipathic phospholipid monolayer. 7. Lipid from cytoplasmic droplets can be packaged into VLDL particles and secreted across the basolateral membrane for delivery to peripheral tissues. 8. Cytochrome P450 (CYP450) enzymes metabolize endo-and xeno-biotics; their polar products are transported across the basolateral membrane for renal elimination and their lipid soluble products are transported across the apical membrane for biliary elimination. Adapted from Huang, D. et al. Advanced Functional Materials (2020).

The diverse and complex nature of hepatocyte functions poses a challenge to the development of a suitable system for disease modelling (Figure 1.7). Primary human hepatocytes (PHHs) are widely accepted as the gold standard for in vitro investigations of liver physiology. However, they are expensive, difficult to source fresh, and typically come from donors who are undergoing surgeries for hepatobiliary diseases, which may affect the health of the cells. Furthermore, it has been well established that isolated PHHs are generally non-proliferative and rapidly undergo extreme morphological alterations when cultured as a monolayer, followed by sharp declines in many of their essential functional capabilities, in a process termed dedifferentiation. This makes PHHs challenging to genetically manipulate and maintain over time, and a poor choice for applications requiring a large number of cells.

Primary hepatocytes from nonhuman species, most commonly rodents, are popular due to their relative ease of accessibility, but often lack the ability to replicate human hepatic physiology. Specifically, rodents exhibit vastly different lipoprotein profiles from humans. For example, mice lack CETP to shuttle CE from HDL to apoB-containing lipoproteins; thus, mice carry the majority of plasma cholesterol in HDL and exhibit low levels of LDL, whereas humans carry considerable plasma cholesterol in LDL. This difference has been proposed as an explanation for why mice are inherently resistant to the development of atherosclerosis without genetic manipulation (eg LDLR-/-) and are thus often unsuitable for modeling CVD. Another significant difference is rodent
hepatocytes secrete both apoB100- and apoB48-containing lipoprotein particles, while apoB48 is not synthesized or secreted by human livers.

Alternative hepatocyte models, such as HepG2 and Huh7 - which were derived from human hepatocarcinoma tumors - are well characterized, and their rapid proliferation allows for affordable, consistent, and easily manipulated studies.\textsuperscript{102} However, these cell lines are phenotypically more similar to cancer cells than mature hepatocytes and secrete lipid-poor VLDL particles which distribute predominantly in the LDL density range.\textsuperscript{114} Additionally, the NAFLD-promoting I148M variant was recently identified in the PNPLA3 gene in both HepG2 and Huh7 lines; thus, these cells offer limited use for the dissection of the molecular pathways leading to NAFLD.\textsuperscript{115} Therefore, the development of an expandable, reliable, and functional alternative human hepatocyte cell source will provide great value to biomedicine.

Easily accessible adult somatic cells, such as skin fibroblasts or white blood cells can be virally reprogrammed, or induced, into pluripotent stem cells. These are rapidly proliferating cells that are characterized by their ability to differentiate into any cell lineage. When treated with special growth factor- and chemical- formulations, iPSCs can be differentiated \textit{in vitro} into Hepatocyte-like cells (HLCs). In the past decade, HLCs have been used for a wide range of applications, as they are more genetically relevant and physiologically translational than cancer cell lines and nonhuman hepatocytes and theoretically offer an infinite source of clonally identical cultures. A variety of protocols to generate HLCs that express liver-specific genes and proteins, as well as exhibit basic hepatic functions, such as albumin secretion and drug detoxification, have been published. Importantly, HLCs can be obtained from patient iPSCs harboring pathogenic mutations and have been shown to phenocopy hallmark disease characteristics \textit{in vitro}.\textsuperscript{116}
Despite their promising potential, HLCs possess their own set of disadvantages. Firstly, because the directed differentiation of HLCs from iPSCs is an artificially accelerated process, HLCs do not exhibit equivalent functional maturity to PHH; instead, HLCs are phenotypically closer to fetal hepatocytes.\textsuperscript{117–119} Additionally, an abundance of evidence confirms that in a conventional monolayer culture, HLCs, like primary hepatocytes (from any species), undergo dedifferentiation and therefore cannot be maintained for long term once they reach peak maturity.\textsuperscript{110–112}

Figure 1.7: Factors to consider when choosing hepatocyte sources for disease modeling. Created with BioRender.com

Hepatocyte function within the whole liver is critically dependent upon the incorporation of each highly specialized, non-parenchymal liver-resident cell type into the complex sinusoidal unit in which hepatocytes are embedded.\textsuperscript{120} Further, extracellular matrix (ECM) surrounding each cell controls the establishment and maintenance of membrane polarity. Thus, for hepatocytes to maximally function in culture, cellular polarity as closely reminiscent of the \textit{in vivo} state as possible must be achieved.\textsuperscript{121} This reliance of hepatocyte function on the intricate structural organization of their surrounding milieu indicates that the sub-optimal maturity and dedifferentiation exhibited by cultured HLCs are likely attributable to a common cause, namely
inadequate cell-cell and cell-ECM interactions. Consequently, many tissue engineering strategies have been developed to better mimic the hepatic microenvironment, and in turn, improve the maturation and sustain the survival of hepatocytes in vitro and ultimately expand their utility.

While polarized cells, including hepatocytes, were originally cultured in dishes coated with an ECM-protein, such as collagen, to mimic the substratum on which they reside in vivo, the collagen sandwich became widely adopted after its introduction in the early 1970’s to improve their functionality further. This technique calls for the seeding of cells onto collagen-coated dishes followed by the overlaying with either additional collagen or collagen-containing culture medium in order to fully encapsulate the cells; in hepatocyte cultures, the collagen sandwich was shown to improve ECM geometry, maintain membrane polarity, and extend long-term hepatic function, compared to collagen monolayers. Since then, commercially available non-adhesive hydrogels, such as Matrigel (which is derived from the matrix secreted by mouse sarcoma cells), have been developed; these formulations offer the benefit of containing multiple types of ECM proteins, which better mimic the in vivo basement membrane composition.

The observation that primary hepatocytes in suspension possess the intrinsic ability to self-organize into functional aggregates led to the approach of maintaining these cells in 3D-culture configurations called spheroids. It has since been demonstrated that independently self-assembled spheroids, composed of primary hepatocytes, hepatoma cells, or differentiated HLCs, all possessed tight cell-cell contacts, membrane polarity essential for functional bile canaliculi, and improved hepatic functions, when compared to their 2D counterparts. The use of microencapsulation in biocompatible scaffolds or droplet-based microfluidic systems was shown to further improve the structural and functional complexity of these microtissues.
Certain production methods for hepatic spheroids, including droplet-based microfluidics, allow for the incorporation and coculture of hepatocytes with nonparenchymal cells (NPCs) of the liver, which include Kupffer cells (KCs), hepatic stellate cells (HSCs), and liver sinusoidal endothelial cells (LSECs). Importantly, liver NPCs secrete growth factors and ECM proteins, which contribute to a micro-environment that facilitates both cell–cell and cell–matrix interactions. While the addition of these cells to primary hepatocyte cultures has been shown to provide substantial functional improvement over pure hepatocyte cultures, the benefit of incorporating non-hepatic cells into HLC cultures has not yet been extensively investigated. It should be noted that, in general, these co-culture systems are often complicated by the challenge of selecting an appropriate culture medium and ECM composition in which all of the cell lineages can be optimally co-maintained and will likely be even more difficult when adapting to HLC-containing cultures, given the strict environment in which those cells thrive.

A more recent microtissue engineering technique meant to address both cell-ECM and cell-cell improvements in iPSC-derived hepatic models is the production of liver organoids. Liver organoids by definition contain multiple liver-resident cell-types that exhibit spatial self-organization into an organ-like architecture and possess the ability to recapitulate hepatic-specific functions. Unlike traditional co-cultures, wherein multiple mature cell types are combined, this approach involves the co-differentiation of stem cells (embryonic, adult, or iPSC), while configured in 3D-culture, to parenchymal and supporting lineages. Takebe and colleagues developed a robust culture method to derive and maintain multi-cellular human liver organoids composed of hepatocyte-, stellate-, and Kupffer-like cells from donor iPSCs. These organoids exhibited transcriptomic patterns similar to those observed in whole liver, and, when stimulated
with fatty acids, displayed characteristics of steatohepatitis, including steatosis, inflammation, and fibrosis, offering a novel approach for modeling NAFLD in vitro.\textsuperscript{126}

1.9 Research goals

CVD and NAFLD are two of the leading causes of mortality globally. Their pathophysiologies are linked by the rates of hepatic production and clearance of apoB-containing lipoproteins; however, many details regarding these processes remain unknown and require additional investigations. A nonsynonymous, loss of function variant located in the gene encoding TM6SF2 was identified in multiple genetic association studies, which was shown to correlate with increased risk for NAFLD and decreased risk for hyperlipidemia. Given the pivotal role that lipoproteins play at the juncture of these two conditions, as well as the similarity between tissue expression patterns of TM6SF2 and apoB (i.e. liver and small intestine), researchers hypothesized, and later confirmed that TM6SF2 plays a role in the degree of lipidation of VLDL particles synthesized in the liver. The current prevailing theory as to TM6SF2 function asserts that TM6SF2 participates in the late-stage lipidation of the VLDL particle, and its absence results in the secretion of smaller, less lipilated particles, concomitant with the retention of excess lipid in the liver, and thus resulting in the observed hypolipidemia and steatosis. Recent work by Hobbs and colleagues suggests that TM6SF2, may act as a VLDL ‘gatekeeper’, ensuring that only adequately lipilated particles are trafficked to the Golgi for secretion. According to their theory, TM6SF2, through an unknown mechanism, blocks the exit from the ER of VLDL that are not fully lipilated, or facilitates their recycling from the ERGIC back to the smooth ER for additional lipidation.\textsuperscript{87} Thus, when TM6SF2 is impaired or absent, the underlipidated particles do not undergo recycling or further lipidation, and instead are secreted as is; this scenario rationalizes the phenotype in which a greater number of smaller VLDL particles are secreted.
However, not all published data supports this theory and disagreement regarding many aspects of the mechanistic function of TM6SF2 remain. For example, some groups have shown that disruptions of TM6SF2 cause decreases in apoB secretion \textit{in vivo} or from cultured hepatocytes, which would suggest that TM6SF2 may promote the synthesis, stability, lipidation, or trafficking through the secretory pathway of VLDL particles. In order to better understand how TM6SF2 functions, this discrepancy must be resolved. The inconsistencies observed in the literature may be due to drawbacks of the models used to study TM6SF2, including species differences, metabolic factors, and the degree of TM6SF2 expression, or lack of expression. Thus, there is an obvious need for an improved ex-vivo hepatocyte model to further elucidate the actions of TM6SF2. Therefore, the goals of this dissertation were two-fold: 1. To characterize lipoprotein metabolism in validated mature and functional human HLCs and 2. To investigate the role of TM6SF2 in lipoprotein metabolism in HLCs derived from donors carrying the E167K variant.

HLCs have been shown to be more translational than cancer cell lines and nonhuman hepatocytes, and can phenocopy hallmark disease characteristics \textit{in vitro}, and therefore seemed an appropriate model in which to explore TM6SF2. However, at the time of this study’s conception, many lipoprotein pathways had not yet been characterized in these cells. Therefore, the first aim of this dissertation was to generate functionally mature HLCs from healthy donors and assess their utility in modeling hepatic lipoprotein metabolism, compared to commonly used hepatocyte sources. To accomplish this goal, we collaborated with members of the Leong lab who specialize in the creation and differentiation iPSCs, as well as the engineering of 3D-culture systems. Briefly, peripheral blood mononuclear cells (PBMCs) were isolated from healthy donors’ peripheral blood and reprogrammed to iPSCs, which were then differentiated to HLCs with our optimized protocol. The resulting HLCs were rigorously validated for maturity and functionality, and further shown to
express proteins essential for lipoprotein metabolism, secrete authentic VLDL particles, and respond to metabolic perturbations. HLCs were also detached and microencapsulated into spheroids using high-throughput droplet microfluidics to improve the long-term maintenance of HLC cultures. These studies are beyond the scope of this dissertation and will not be discussed in further detail.

Once we confirmed the robustness of the iPSCs to HLCs differentiation protocol, our second aim sought to apply this technology to investigate the effect of TM6SF2 E167K variant expression on hepatic lipid metabolism in HLCs from 4 homozygous TM6SF2-carrier individuals (KK) and 4 age- and sex-matched unaffected siblings (EE). We reprogrammed PBMCs isolated from these 8 donors to iPSCs, which we then differentiated to HLCs to serve as an improved model in which to investigate the effect of the TM6SF2 E167K variant expression on hepatic lipid metabolism.

We planned to address some of the key questions that remain to be resolved regarding the relationship between TM6SF2 and hepatic VLDL metabolism. Firstly, we sought to determine how expression of the mutant TM6SF2 protein impacted HLC apoB and TG secretion, and whether the size of the secreted VLDL particles was affected. We also planned to test the effect of ALLN, a commonly used proteasomal inhibitor, on apoB degradation in mutant HLCs, compared to WT. Additionally, we planned to dissect the subcellular compartment in which TM6SF2 was acting on VLDL lipidation. To do so, we would utilize well-known chemical inhibitors to block the trafficking of VLDL particles from the ER to the Golgi (with Brefeldin A) and from the Golgi to the plasma membrane (with Nocodazole) and determine the effect of TM6SF2 expression on the size of VLDL particles isolated from each of these compartments. Unfortunately, as will be
discussed at the end of Chapter 2, our sibling-matched HLC model suffered from many unanticipated technical issues, and we ultimately chose to design alternative models.

In Chapter 2, we present our optimized protocol for differentiating HLCs from donor iPSCs and our characterization of lipoprotein metabolism in these cultures. We also discuss our attempts to explore the impact of the TM6SF2 rs58542926 SNP in EE and KK HLCs and the challenges we encountered in the process. In Chapter 3, we describe in detail the gene editing strategy we developed to address these constraints. Finally, in Chapter 4, we present our initial findings regarding the mechanism by which TM6SF2 impacts lipid metabolism.
Chapter 2: Characterizing Lipoprotein Metabolism in Hepatocyte-Like Cells

To date, many groups have verified the capability of HLCs to develop hepatosteatosis, typically through the addition of fatty acids to culture medium.\textsuperscript{129,130} Additionally, Liu, \textit{et al} demonstrated that HLCs derived from patients with abetalipoproteinemia (due to a loss-of-function mutation in MTP) displayed reduced VLDL-apoB and TG secretion, along with increased cytosolic lipid accumulation, compared to wild type HLCs, thus recapitulating the \textit{in vivo} phenotype.\textsuperscript{131} HLCs have also been shown to exhibit features of steatohepatitis with fatty acid stimulation, including inflammation and increased fibrosis,\textsuperscript{132,133} and have been employed to investigate mechanisms of mitochondrial dysfunction in hepatic steatosis\textsuperscript{134}. However, to our knowledge, lipoprotein metabolism has not been precisely characterized in HLCs from healthy donors to determine the utility of these cells in modeling human lipoprotein metabolism.

2.1 Reprogramming of human PBMCs to iPSCs

The protocol to generate specific donor-derived iPSCs is summarized in Figure 2.1A. PBMCs were first isolated from peripheral blood by Ficoll gradient and expanded in culture for 9 days in enrichment medium.\textsuperscript{135} Following expansion, PBMCs appeared visibly larger in size and formed small grape-like clusters, indicating active cell proliferation and readiness for reprogramming (Figure 2.1B, i). On Day 9, cells were transduced through spinfection with Sendai viruses carrying the four Yamanaka Factors: octamer-binding transcription factor 4 (Oct4), sex determining region Y(SRY)-box 2 (Sox2), Kruppel-like factor 4 (Klf4), and c-Myc (Figure 2.1B, ii). Cells were then transferred to a feeder layer of inactivated mouse embryonic fibroblasts (iMEFs), onto which successfully transduced PBMCs settled and formed small colonies (Figure 2.1B, iii), while cells that failed to be reprogrammed remained in suspension and were removed with
daily medium change. iPSC colonies were allowed to expand for the next 2 weeks (Figure 2.1B, iv-vi) until large enough for manual picking.

The reprogramming process was highly efficient and robust for all donors; most iPSC colonies exhibited defined and smooth edges, and dense, homogenous colonial features (Figure 2.1C). For each donor, 6 iPSC clones were selected for further propagation, adaptation to feeder-free culture on Matrigel, hepatic differentiation, and phenotypic analysis. The iPSC lines showing consistent proliferation, stable morphology, and high differentiation potential were chosen to represent the donor. Protocol optimizations were performed using iPSC lines from three healthy adults LG1-C7, LG2-C6, and LG4-C6 (Table 2.1). Strong expression of pluripotency markers Oct4, TRA-1-60, Sox2 and stage-specific embryonic antigen-4 (SSEA4) was observed in all of the chosen iPSC lines (Figure 2.1D).

Figure 2.1: PBMC-to-iPSC reprogramming. A) Schematic illustrating the process and timeline for generating donor-specific iPSC clones. Created by DH with BioRender.com B) Morphological changes of PBMCs undergoing reprogramming to mature iPSCs in 2 weeks (green arrow in Panel iii—a successfully transduced PBMC formed an iPSC colony planted on the feeder layer). Scale bar: 100 µm. C) PBMC-derived iPSCs were generated on feeder layer (iMEF) before transfer to Matrigel. D) Fluorescent staining of pluripotency markers (Oct4,
TRA-1-60, Sox2, and SSEA4) in iPSCs. Representative images collected from donor LG1; iPSCs from all 3 donors exhibited similar morphologies.

Table 2.1: iPSC donor information. Gender, age at the time of blood donation, ethnicity, and general health were reported. The clone selected for each donor is listed in the last column.

<table>
<thead>
<tr>
<th>Donor ID</th>
<th>Gender</th>
<th>Age</th>
<th>Ethnicity</th>
<th>Health Condition</th>
<th>Selected iPSC Clone</th>
</tr>
</thead>
<tbody>
<tr>
<td>LG1</td>
<td>Male</td>
<td>40</td>
<td>Caucasian</td>
<td>Good</td>
<td>LG1-C7</td>
</tr>
<tr>
<td>LG2</td>
<td>Male</td>
<td>58</td>
<td>Caucasian</td>
<td>Good</td>
<td>LG2-C6</td>
</tr>
<tr>
<td>LG3</td>
<td>Male</td>
<td>48</td>
<td>Caucasian</td>
<td>Good</td>
<td>LG3-C6</td>
</tr>
<tr>
<td>LG4</td>
<td>Female</td>
<td>59</td>
<td>Caucasian</td>
<td>Good</td>
<td>LG4-C6</td>
</tr>
<tr>
<td>LG5</td>
<td>Female</td>
<td>47</td>
<td>Caucasian</td>
<td>Good</td>
<td>LG5-C1</td>
</tr>
</tbody>
</table>

2.2 Differentiation of iPSCs to hepatocyte-like cells (HLCs)

The protocol we followed for human iPSC-to-HLC differentiation, which was adapted from the Cowan lab, is summarized in Figure 2.2. The differentiation process is carried out across four stages of maturity, each of which is achieved through the treatment of cells with a uniquely defined medium composition meant to mimic the embryonic developmental environment. The first stage directs the cells to germ layer specification, which, in the case of hepatocytes, is the definitive endodermal layer (DE). Germ layer specification plays a pivotal role in hepatic differentiation; therefore, rigorous assessment was carried out to determine the efficiency of DE induction in various iPSC lines. This was defined as the percent of cells expressing surface receptors C-X-C chemokine receptor type 4 (CXCR4 or CD184) and C-Kit (CD117), measured via flow cytometry. Since achieving a pure population at this stage was so crucial for the success of the subsequent differentiation, the efficiency of DE differentiation was used as a first-step assessment for hepatic differentiation potential among iPSC clones from the
same donor, for which only clones displaying >90% CXCR4+/C-Kit+ were selected for future differentiation (Figure 2.3A).

The morphological changes of differentiating cells were monitored daily and representative images at the end of each stage are shown in Figure 2.3C. Significant cell death was consistently observed within the first 24 hours of DE induction, followed by high proliferation in the next 48 hours. At the end of DE, cells typically reached 100% confluence, often displaying overcrowding. However, this density was too high for efficient hepatic endoderm (HE) differentiation, leading to contamination by cells from non-endodermal lineages and heterogeneously differentiated cultures. For example, areas of some cultures positively stained for cardiac myocyte markers, and at times were seen visibly “beating” (data not shown). Other cultures contained patches of small Sox2+ cells, which could either represent undifferentiated iPSCs, or cells committed to the ectoderm germ layer,142 rather than DE (data not shown). Hence, following 3 days in DE medium, cells were detached, counted, and reseeded onto new Geltrex-coated plates to commence the HE stage (Figure 2.3B), according to the method of Carpentier, et al.143 Optimal reseeding density was determined for each line based on the extent to which it continued to proliferate during HE (approximately 0.26 million cells/cm²). We further optimized a number of conditions for our specific iPSC lines (data not shown). As differentiation progressed through the immature hepatocyte (IMH) stage, to the final, mature hepatocyte (MH) stage, cells remained confluent and displayed morphological features unique to hepatocytes. MH cells formed a packed monolayer with distinct cell boundaries (Figure 2.3C, last panel), and exhibited polygonal shapes, prominent nucleoli, and occasional binucleation (blue arrows).
Figure 2.2: Optimized protocol for the directed differentiation of iPSCs to mature HLCs using a combination of growth factors (GFs) and small molecules (SMs). Created with BioRender.com

Figure 2.3: Monitoring of cell stages during iPSC differentiation to HLCs A) Flow cytometry of DE-specific surface receptors (CXCR4 and C-Kit). Control group was treated with isotype control antibodies. B) Replating of over-confluent DE cells on Geltrex enhanced downstream hepatic differentiation. Replating density is line-dependent. Yellow dashed circle indicates an aggregation of non-hepatic cells in un-replated culture at MH stage. C)
Representative images showing morphological changes on 2D culture as iPSCs progressed through 4 stages of differentiation (blue arrow—binucleation). Representative images and flow cytometry data collected from donor LG2; iPSCs from all 3 donors exhibited similar results.

### 2.3 Phenotypic and functional confirmation of HLC hepatic maturity

Once HLCs with appropriate morphology were successfully differentiated from donor-derived iPSCs, detailed phenotypic and functional characterizations were performed to assess their hepatic maturity. First, RNA was extracted from HLCs on Day 9 of maturation stage (MH Day 9) to determine expression levels of hepatic genes via quantitative reverse transcription PCR (RT-qPCR). Results are presented in a heatmap in Figure 2.4A showing fold-change relative to expression level in fresh adult liver extract, which is considered the gold standard reference for hepatic maturity.\(^\text{112}\) Compared to iPSCs, HLCs displayed significant enrichment of essential hepatic transcriptional regulators, including hepatocyte nuclear factor 4α (HNF4α) and liver x receptor α (LXRα), as well as specific markers of hepatocyte maturity, including albumin, alpha-1-antitrypsin (AAT), and apolipoprotein B100 (apoB). Although some cytochrome P450 (CYP) enzyme expression was detected in HLCs, these levels were much lower than those in liver extracts. Thus, our HLCs possessed an intermediate gene expression profile between adult liver extract and HepG2 cells, a hepatoma line popular for \textit{in vitro} hepatocyte, but still considered to exhibit an immature phenotype.\(^\text{144–146}\) Western blot analysis further confirmed these findings (Figure 2.4B), showing significant induction of CYP3A4, and HNF4α protein expression, compared to iPSCs, and similar levels of albumin and AAT, compared to fetal liver lysates.

To further assess the cellular composition of the HLC cultures, mature HLCs were dissociated into mono-dispersed suspensions, and stained for the hepatic markers AAT, AFP, and albumin, followed by analysis by flow cytometry (Figure 2.4C). AAT and AFP were expressed by almost all the cells, but the differentiated culture showed varied expression levels for the other
markers. 85.7% of HLCs were positive for albumin, about 40% of which could be characterized as highly mature, whereas 60% showed low to medium expression. Based on these initial assessments, and consistent with previously reported HLC differentiation protocols, the molecular phenotype of the HLCs produced from our modified protocol most closely resembled that of fetal hepatocytes.

![Figure 2.4: Phenotypic characterizations in HLCs. A) Heatmap comparing hepatic gene expression levels among adult liver extract, fetal liver extract, LG1 HLCs, HepG2, and iPSC. Values were normalized to adult liver extract. B) Western blot depicting LG4 HLC protein expression levels of common hepatic markers compared to iPSC and fetal liver extract levels. C) Flow cytometry analysis depicting percentages of LG2 HLCs positively labeled for AAT, AFP, and albumin.](image)
Next, immunocytochemistry and confocal imaging were performed to determine the cellular distribution of hepatic proteins. HLCs displayed extensive and uniform staining of hepatocyte markers, including HNF4α, AAT, albumin, alpha-fetoprotein (AFP) and hepatocyte nuclear factor 1α (HNF1α), suggesting efficient and homogenous differentiation (Figure 2.5A). CYP3A4 staining was observed in a clustered pattern, indicating some degree of functional heterogeneity and incomplete maturation in the HLC culture. As noted above, prominent cell boundaries were visible in bright-field images of HLC monolayers (Figure 2.3C), indicative of distinct cell-to-cell junctions, which are essential for hepatic structural integrity and function. Immunofluorescence staining for zonula occludens-1 (ZO-1, a tight junction-associated protein and apical membrane marker) and E-cadherin (a cell-surface epithelial protein) provide further evidence of epithelial polarity in the HLC culture (Figure 2.5A, inset) and are suggestive of the presence of a bile canalicular network.

To investigate whether these membranes possess biliary efflux activity, HLCs were treated with 5(6)-carboxy-2′,7′-dichlorofluorescein diacetate (CDFDA), which, when cleaved intracellularly, produces the fluorometric substrate CDF, which is subsequently excreted by the multidrug resistance-associated protein 2 (MRP2) on the canalicular membrane. Following 30 minutes of incubation, HLCs displayed substantial CDFDA uptake and cleavage (Figure 2.5D, left); after CDFDA removal and extensive washing, secreted CDF was seen along HLC junctions (Figure 2.5D, right), further suggesting the presence of active MRP2-dependent biliary efflux, polarized membranes, and bile canaliculi structures in HLCs.

Hepatocytes are a major site of glycogen storage in the body. Glycogen can be synthesized or mobilized based on the needs of extrahepatic cells. HLCs exhibited a bright pink color after
Periodic Acid-Schiff (PAS) staining (Figure 2.5D), indicating extensive glycogen storage throughout the culture.

Figure 2.5: Phenotypic and functional characterizations in HLCs. A) Immunofluorescence staining and confocal imaging of HNF4α, AAT, albumin, ZO-1, AFP, E-cadherin, CYP3A4, and HNF1α. B) Flow cytometry analysis depicting percentage of HLCs positively labeled for E-Cadherin. C) CDFDA uptake and CDF secretion for visualizing MRP2 bile transporter activity on the apical membrane. Insert with bright-field image overlap shows an example of CDF accumulation at one side of the cell boundaries. D) Glycogen storage (pink) in HLCs with PAS staining. Representative images collected from donor LG1.

Hepatocytes are the main producers of the body’s circulating proteins, which include carrier proteins, clotting factors, apolipoproteins, and hormones, and are additionally responsible for the conversion of ammonia (the toxic nitrogenous product of amino acid degradation) to urea for safe elimination.\textsuperscript{150,151} These products exit the hepatocyte into circulation across the basolateral
membrane, either through the canonical secretory pathway, or through specific membrane-bound transporters. To quantitatively determine the ability of HLCs to perform these liver-specific functions, albumin and urea secretions were measured over time, serving as surrogate indicators of protein synthesis and catabolism, respectively. Upon entering the maturation stage, the secretion of albumin, a marker of advanced hepatic maturity, steadily increased from 0 ng/day to >600 ng/day on MH Day 7, after which it stabilized (Figure 2.6A); a similar pattern was observed in the albumin secretion profiles of HLCs from two other donors (data not shown). Additionally, urea secretion increased 3-fold during the maturation stage, nearing plateauing levels at MH Day 9. Thus, HLCs at MH Day 7 were regarded as exhibiting peak hepatic maturity and phenotypic and functional characterizations were performed within the next 5 days. Normalization of secreted albumin and urea levels (Figure 2.6B) to the total protein content or cell number allowed for comparisons across different experiments and donors, as well as with other reported findings. With 1.9 µg of albumin and 4.7 µg of urea secreted per mg of total protein, our HLCs demonstrated comparable secretion levels to established protocols.

The liver plays a key role in modulating cholesterol homeostasis and this activity is also dependent upon hepatic membrane polarity. Hepatocytes clear excess cholesterol from the circulation via internalization of LDL particles, which are taken up at the basolateral membrane by LDL receptors (LDL-Rs) via receptor-mediated endocytosis. LDL-R activity can be tracked in vitro with the use of LDL particles conjugated to a pH-sensitive dye (pHodo Red-LDL), which only fluoresces after the particle is internalized. Following the treatment of HLCs with pHodo Red-LDL, the fluorescent intensity steadily increased over time in culture and more than 50% of the cells showed LDL uptake after 4 hours (Figure 2.6C), indicating functional LDL-R-mediated endocytosis for LDL clearance.
Another essential hepatic function is drug and xenobiotic metabolism through the activity of CYP enzymes. One of the major uses for hepatocyte models is the screening of new drug compounds for hepatotoxicity, so well-functioning CYP pathways are crucial for hepatocyte utility. To demonstrate this activity, cells were challenged with acetaminophen (N-acetyl-para-amino-phenol, APAP), a common analgesic that is oxidized to a hepatotoxic metabolite by CYP2E1,157 followed by the assessment of cytotoxicity. This was done by simultaneously staining cells with green fluorescent calcein-AM to indicate intracellular esterase activity (LIVE) and red-fluorescent ethidium homodimer-1 to indicate loss of plasma membrane integrity (DEAD). HLCs exposed to 10 and 20 mM of APAP for 48 hours displayed progressive apoptosis, as evidenced by the increased red fluorescence, and loss of structural features (Figure 2.7A), indicating CYP2E1 activity and dose-dependent cytotoxicity.

Next, we assessed the response of HLCs to well-characterized CYP clinical inducers. Omeprazole, phenobarbital, and rifampin are strong inducers of CYP1A2, 2B6 and 3A4 gene expression, respectively.158–160 HLCs were treated at two physiologically relevant concentrations of each drug for 48 hours prior to collection for RT-qPCR measurement of CYP levels. No
significant induction was observed at the lower concentration of each drug compared to the DMSO vehicle control (Figure 2.7B). At the higher concentration of 200 µM omeprazole, 2 mM phenobarbital and 50 µM rifampin, CYP1A2, 2B6 and 3A4 gene expression levels were upregulated by ~6-fold, ~3-fold and ~3-fold, respectively. Collectively, these results demonstrate the drug metabolizing capacity of HLCs in responding appropriately to pharmacological agents.

Figure 2.7: Determination of xenobiotic metabolism in LG1 HLCs. A) Assessment of hepatotoxicity in HLCs with APAP treatment at 10 mM and 20 mM, followed by LIVE/DEAD (green/red fluorescence) assay. B) RT-qPCR of CYP enzyme induction using omeprazole (CYP1A2), phenobarbital (CYP2B6), and rifampin (CYP3A4). HLCs were incubated with two concentrations of each drug for 48 hours. Scale bar: 100 µm in (A).

2.4 Lipoprotein synthesis and secretion in HLCs

While many groups have previously generated HLCs exhibiting the functions described above, at the time of this study, HLCs had not yet been confirmed to synthesize bona fide VLDL particles. Therefore, we next sought to fully characterize the ability of HLCs to mimic canonical components of the VLDL biogenesis pathway. First, the expression levels of proteins essential for this process were assessed via Western blot analysis. At MH Day 5, HLCs showed significantly induced protein expression of apoB100, MTP, and apoE compared to iPSCs; HLCs also displayed...
greater apolipoprotein expression than HepG2 cells (Figure 2.8A). Additionally, HLCs showed abundant expression of apoA1 (a crucial apolipoprotein for HDL biogenesis) as well as LDLR and SRB1, receptors that facilitate hepatic uptake of lipoproteins from the circulation (Figure 2.8B). Notably, expression levels of many of these proteins were reduced or no longer detectable by day 10 or 15 of MH stage, indicating that the HLCs reached a level of peak functionality and then began to deteriorate.

Given this unsteady expression over time, we next sought to characterize the daily HLC apoB100 secretion levels. Culture medium was collected and refreshed every 24 hours for 9 days and apoB100 levels were determined by ELISA. The results demonstrated that HLCs secreted greater than 1 ug/day of apoB100 during the first 9 days of the maturation stage (Figure 2.8C), which is comparable to or higher than other reports of apoB secretion in HLC, PHH, and HepG2 cultures (Figure 2.8D). It should be noted that, while albumin and urea secretion were undetectable at the commencement of the maturation stage and progressively increased over time, apoB100 secretion levels started out at a maximum during the immature hepatic stage, dipped slightly during the start of maturation, and peaked at approximately MH Day 5, before slowly tapering downward. Accordingly, RT-PCR analysis illustrated that, while albumin gene expression was nearly undetectable at the end of the IMH stage (Figure 2.8E), apoB expression was already significantly induced by the 3rd day of IMH and increased less than 2-fold by MH Day 4 (Figure 2.8F). Taken together with the protein expression above, these data suggest that apoB-containing lipoprotein secretion commences in early HLC maturation and HLCs exhibit peak lipoprotein profiles between MH days 4-8; thus, subsequent lipoprotein characterization studies were conducted within this timeframe.
A

B

C

D

E

F
Figure 2.8: Assessment of HLC apoB100 expression and secretion over time A&B) Western blot showing HLC (LG4) induced expression of apolipoproteins and associated proteins at the indicated timepoints during the mature hepatocyte (MH) stage. C) HLC (LG1) daily secretion of apoB100 during MH stage. n=3 wells per timepoint. D) Day 5 LG1 HLC apoB100 secretion compared to HepG2. n=3 wells per cell type. E&F) RT-qPCR analysis comparing HLC gene expression of E) albumin and F) apoB at various time points during HLC (LG1) differentiation.

The dependence of hepatic VLDL assembly on MTP-mediated lipidation of apoB has been widely established across numerous models. However, it was unknown if this was true in HLCs as well; therefore, we sought to determine if HLC VLDL secretion similarly requires MTP activity. HLCs were treated with CP-10447, a chemical inhibitor of MTP, and newly synthesized and secreted apoB100 were measured using established radio-isotope procedures for protein labeling with \(^{35}\)S-methionine. Following a 2-hour treatment with CP-10447, newly synthesized intracellular apoB100 ("Cell") was reduced by >80\% in HLCs compared to the vehicle control (Figure 2.9A), and newly secreted apoB100 ("Media") was undetectable in MTP-inhibited HLCs (Figure 2.9B). Overall, these results indicate degradation of apoB occurred due to poor lipidation. Notably, albumin synthesis and secretion remained intact following CP-10447 treatment, confirming that the inhibitor did not induce off-target defects in the protein secretory pathway.

HLCs were further treated with \(^{3}\)H-glycerol to determine the effect of MTP inhibition on the secretion of newly synthesized TG. When acutely stimulated with 0.4 mM oleic acid (OA) for 4 hours, HLCs secreted about 7 times as much labeled TG vs. control; however, TG secretion from HLCs treated with both OA and CP-10447 was reduced to levels similar to control (Figure 2.9D). This attenuation of OA-stimulated TG secretion by MTP inhibition further signifies that HLCs exhibit deficient TG transfer onto VLDL when MTP activity is impaired. To determine if this observed reduction in TG secretion resulted in a concomitant accumulation of intracellular lipid, cells were stained with BODIPY, a neutral lipid-binding fluorescent dye. HLCs stimulated
overnight with 0.4 mM OA accumulated significantly more BODIPY-stained lipid droplets, compared to untreated controls (Figure 2.9E, middle); when MTP was additionally inhibited in the presence of 0.4 mM OA, the lipid droplets appeared strikingly larger (Figure 2.9E, right), confirming that lipids were unable to be transferred to apoB100 in the ER and secreted in VLDL without MTP activity, and thus accumulated in cytosolic lipid droplets.
Figure 2.9: MTP inhibition reduces VLDL-apoB and TG secretion in HLCs from donor LG1. Radiographs and densitometric quantification of A) HLC newly synthesized apoB100 (cell) and B) secreted (media) apoB100 following 2 hours MTP inhibition with CP-10447. n=2 wells per condition. C) Quantification of newly synthesized TG from 3H-glycerol in HLCs and
HepG2 cells following treatment with 0.4 mM oleic acid and CP-10447. n=3 wells per condition. 
D) Quantification of \(^3\)H-glycerol labeled TG secreted from HLCs and HepG2 following 2 hour treatment with 0.4 mM oleic acid and CP-10447. n=2 wells per condition. E) Bodipy staining of neutral lipid droplets in HLCs following overnight treatment with 0.4 mM oleic acid and CP-10447. n=1 well per condition; multiple fields were visualized and photographed. Representative images are presented.

We next sought to determine the density distribution of apoB100 particles secreted by the HLCs. Using established radiolabeling and sucrose gradient ultracentrifugation methods in our group (Figure 2.10A), we found that the majority of apoB particles secreted by the HLCs were isolated in the 4 least dense fractions making up the VLDL density range (Figure 2.10B). In fact, ~60% of total HLC-secreted apoB came out in the first fraction, whereas less than 10% of HepG2-secreted apoB was observed in the first fraction (Figure 2.10C), indicating that the particles secreted by HLCs more closely resemble in vivo VLDL than those secreted by hepatoma cells.

Interestingly, when HLC cultures were acutely challenged with 0.4 mM OA for the duration of the radiolabel pulse, we did not see an increase in cell or total secreted apoB100, nor was the density distribution of secreted lipoproteins altered substantially (Figures 2.10D, E, F). This observation differed from the ~30% increase in cell apoB and nearly 4-fold increase in total secreted apoB100 compared to control seen in HepG2 cultures (Figures 2.10E, F), along with a shift toward more buoyant particles (Figure 2.10C, right), which is consistent with previous reports.\(^{164}\) There is evidence that OA treatment does not induce greater apoB secretion in primary human and rat hepatocytes,\(^{165,166}\) thus further suggesting that HLCs more closely mimic lipoprotein metabolism in vivo than do HepG2 cells. Remarkably, apoA1, the predominant lipoprotein found on high-density lipoprotein (HDL) particles, was distributed in the later, denser fractions, confirming that the HLCs produced other lipoproteins of hepatic origin at the appropriate density range (Figure 2.10G). As expected, the secretion of HDL was unaffected by OA treatment in HLCs and HepG2, as it primarily serves as a cholesterol carrying particle. Taken together, these
data prove that HLCs secrete highly lipidated, MTP-dependent, VLDL particles and appropriately exhibit steatosis when this secretion is impaired. They are therefore a highly valuable tool for modeling lipoprotein-related liver diseases in vitro.
Figure 2.10: Determination of the relative particle density of HLC-secreted lipoproteins. A) Schematic of lipoproteins separated by density gradient ultracentrifugation. Samples are collected in 12 fractions, starting from the top, of increasing densities. Lipoprotein graphic created with BioRender.com B) Representative radiographs showing the density distribution of apoB100-containing lipoproteins secreted by HLCs and HepG2 +/- 0.4mM oleic acid treatment. C) Scintillation counts of $^{35}$S-Met incorporation into apoB (CPM) per µg of cell protein following treatment with 0.4mM oleic acid, plotted against density fraction for HLC and HepG2. D) Radiographs and densitometric quantification of newly synthesized HLC or HepG2 apoB100 and Albumin in cells from density fractionation experiment, following treatment with 0.4mM oleic acid. Each cell type was standardized to its own Ctrl. E) Sum of scintillation counts of $^{35}$S-Met incorporation into apoB (CPM) across all density fractions, normalized to total cell protein (µg). F) Radiographs showing the density distribution of apoA1-lipoproteins secreted by HLCs and HepG2 +/- 0.4mM oleic acid treatment. Data presented from donor LG5, n=1 well per condition. Sucrose gradient ultracentrifugation was performed in three additional donors (LG1, LG2, and LG3) and displayed similar lipoprotein density distributions.

2.5 Investigating the role of TM6SF2 rs58542926 in the regulation of lipid metabolism in patient-derived HLCs

Once we confirmed the robustness of the protocol for directed differentiation from iPSCs to HLCs, we next sought to apply this technology to investigate the effect of TM6SF2 expression on hepatic lipid metabolism in HLCs bearing the EE or KK genotype. Fortunately, we had unique access to freshly isolated blood drawn from a subset of members of the Old Order Amish (OOA) cohort in Lancaster County, PA. The OOA is a founder population that currently numbers ~34,000 individuals and has been studied by the University of Maryland School of Medicine since 1993. These individuals are estimated to descend from 554 founders, with 128 founders contributing to 95% of the present-day gene pool; therefore, the OOA exhibits enriched frequencies for many minor alleles and greater numbers of homozygotes compared to outbred European populations and is thus particularly desirable as a model population for genetic studies. The cohort of 7,236 Amish adults from whom DNA samples were analyzed for one or more studies makes up the Amish Complex Disease Research Program (ACDRP). Extensive phenotype data in various areas,
including glucose homeostasis/diabetes and cardiovascular health have been collected from ~6,000 ACDRP participants.

Previously, O’Hare and colleagues identified 3,556 of these subjects for whom both \textit{TM6SF2} genotypic and relevant phenotypic information, specifically body mass index (BMI), serum LDL-cholesterol, HDL-cholesterol, and triglycerides, was available. Tests of association were carried out for the effect of E167K on serum lipids and hepatic fat content\cite{76}. They found that the TM6SF2 minor allele was strongly correlated with reduced circulating lipids (TC, LDL-C, and TG) and that homozygous (KK) carriers exhibited significantly lower fasting TG, total cholesterol, and LDL-C, higher HDL-C, compared to age and sex matched controls. However, no association was observed between the rs58542926 T allele and presence of steatosis (as measured by electron-beam computerized tomography). It should be noted, however, that the OOA are a relatively healthy population with low BMIs, less diabetes, more physical activity, and less alcohol consumption than the general US population, and thus present fewer confounding factors related to the study of systemic lipid homeostasis.

To accomplish the second aim of this dissertation work, we isolated PBMCs from four KK participants, as well as from four homozygous wild type (EE) siblings matched for age and sex, to serve as controls. Additional information on these subjects is summarized in Table 2.2, including genotyping for the NAFLD-associated PNPLA3 I148M variant, as well as the N352S variant in the B4GALT1 gene recently shown to be associated with decreased LDL cholesterol levels and risk for coronary artery disease\cite{167}.

\textbf{Table 2.2:} Old Order Amish (OOA) donor information. Age at the time of blood donation, Sex, and Genotype are reported for four sibling pairs. Subjects bearing the KK genotype appear in the darker blue rows.
PBMCs from these 8 donors were reprogrammed to iPSCs, as described above, and 6 iPSC clones were selected per donor for further propagation, adaptation to feeder-free culture on Matrigel, and screening for efficiency of DE induction. Figure 2.11A shows the results from flow cytometry analysis of DE markers in 6 iPSC clones derived from one sibling pair. The iPSC line showing the most consistent proliferation, stable morphology, and high differentiation potential was chosen to represent each donor. Similar results were observed for the other 3 sibling pairs (data not shown). Next, genomic DNA was collected from the selected iPSC lines and a small region containing the SNP responsible for the E167K variant was amplified and sequenced to confirm that the appropriate genotype was intact (Figure 2.11B).

While it is still uncertain whether carriage of the mutant TM6SF2 gene causes a reduction in transcript levels, most findings in the literature identify E167K as a loss-of-function variant mediated by reduced TM6SF2 protein levels (rather than the induction of a defect in protein activity).\(^\text{54,95,97}\) Therefore, iPSCs with confirmed genotypes were first differentiated according to the optimized protocol described above, and HLCs were collected to measure TM6SF2 gene and protein levels. No apparent difference in TM6SF2 gene expression was measured between HLCs from sibling pair 1, although a significant decrease in albumin gene expression was seen in the KK donor HLCs, compared to the EE sibling HLCs (Figure 2.11C). TM6SF2 protein levels were
assessed via western blot in HLCs from three sibling pairs (Figure 2.11D). While sibling pair 3 showed the anticipated decrease in TM6SF2 levels in the KK HLCs, pair 1 showed no difference and pair 2 actually displayed the opposite effect. Interestingly, MTP and Albumin levels were highly variable, even between samples from the same donor, indicating that the differentiation efficiency was not consistent between samples; therefore, no conclusion could be drawn from this analysis.

It was concerning that we did not observe a similar consistent reduction in TM6SF2 protein expression in the KK HLCs, compared to the sibling EE HLCs. Furthermore, it was clear that HLCs derived from different iPSC lines exhibited variable degrees of hepatic maturity. This variability became more striking when we attempted to assess the phenotypic impact of the KK variant on lipid accumulation and VLDL secretion in the sibling-matched HLCs. Unlike HLCs derived from previous donors, which exhibited small, evenly distributed lipid droplets under basal conditions, HLC cultures derived from both EE and KK donors displayed significant lipid droplet accumulation (as confirmed by BODIPY staining, Figure 2.11E) without fatty acid stimulation. Further, these lipid droplets often appeared in patches of densely packed cells that appeared smaller than typical HLCs. We also observed that HLC daily apoB100 secretion was highly inconsistent between passages and even sometimes within the same passage, varying within experimental replicates (Figure 2.11F). The level of apoB secretion appeared to be associated with the occurrence of these patches of tightly packed, lipid droplet-laden cells. These observations suggested that the cultures were heterogeneous, containing either cells from diverse lineages or cells from the hepatic lineage, but exhibiting varying degrees of hepatic maturation, despite the passaging of DE cells before HE induction, as described previously. This notion was supported by the fact that iPSCs derived from the OOA participants displayed greater survival and proliferation
during the DE phase, possibly indicating a lower overall DE induction efficiency. Such a case would result in the contamination of reseeded DE cells with greater numbers of cells more closely resembling iPSCs, which may then go on to mature more slowly into HLCs or differentiate to an alternate linage entirely.

Figure 2.11: Initial characterization of iPSCs and HLCs derived from Old Order Amish (OOA) sibling pairs A) Clonal screening of DE differentiation efficiency among 6 iPSC lines derived from sibling pair 1, among which Clone 6 was demonstrated to be the most robust.
n=3. B) Representative chromatogram of Sanger sequencing results, validating E167K genotype in iPSCs from OOA sibling pair 1. C) RT-qPCR analysis of TM6SF2 and Albumin gene expression in HLCs derived from sibling pair 1. D) Western blot comparing HLC protein expression of TM6SF2, MTP, and Albumin between EE and KK siblings. E) Bodipy staining of neutral lipid showing patches of lipid-laden cells in HLCs from both EE and KK donors of sibling pair 1. F) Variability in apoB secretion from 2 different HLC passages in sibling pair 1. Sample n=2 wells per passage.

At the time, we became aware that other groups had experienced similar issues with HLC cultures, and some had published new methodologies to attempt to resolve these issues. Therefore, we pursued several potential improvements to develop an HLC model that would demonstrate greater stability and reproducibility. Interestingly, the Cowan lab, the same group on whose differentiation protocol we had based ours, cited experiencing similar issues with HLC cultures and had published a new method to resolve these issues. Specifically, they detached HLCs at the mature hepatocyte stage and sorted them by FACS using an antibody targeting the hepatic specific ASGPR1 surface protein; this approach allowed for the isolation of a live cell population enriched for highly mature HLCs, which displayed increased expression of hepatic markers and improved hepatic-specific functions such as albumin secretion and CYP-enzyme activity. Following their protocol, we performed FACS on detached HLCs alongside HepG2 cells, as positive controls. We found that, compared to HepG2s, which comprised nearly 95% ASGPR1+ cells, our HLC cultures consisted of ~20% ASGPR1+ cells (Figure 2.12A); despite this low yield, these results were consistent with published findings. We confirmed these data by performing RT-PCR to assess ASGPR1 gene expression, which also demonstrated that HLCs contained about 20% ASGPR1 expressing cells, compared to HepG2s (Figure 2.12B). In order to conduct our desired experiments, sorted HLCs needed to be replated and maintained in adherent culture, which had been successfully demonstrated by Cowan, et al for up to 72 hrs. However, the
ASGPR1+ cells we collected following FACS did not reseed efficiently and rapidly underwent morphological changes and apoptosis (Figure 2.12C).

**Figure 2.12:** Comparison of hepatocyte surface marker ASGPR1 expression in detached HLCs and HepG2 cells A) FACs sorting of ASGPR1+ HepG2 and HLCs. B) RT-qPCR analysis of ASGPR1 gene expression in bulk (unsorted) HepG2 and HLC cultures. C) Representative bright field image of replated ASGPR1+ HLCs at 24 hours post-sorting.

Given the labor intensity of this approach and the inability for continued culture of the isolated ASGPR1+ HLCs, we decided that this approach was not viable. We further realized that, despite our best efforts, we could not reduce the variability in the OOA sibling HLC cultures enough to detect a significant TM6SF2 phenotype in this system in a reasonable amount of time. We determined that alternative methods would be required to enable the use of HLCs for detailed studies of the effect of TM6SF2 on hepatic lipid and lipoprotein secretion; we therefore chose to pursue alternative solutions to enhancing our HLC model, which will be detailed in the next chapter.

In summary, we presented our adapted protocols to reprogram donor PBMCs to iPSCs and to differentiate those iPSCs HLCs. We described the extensive characterizations we performed in these HLCs to confirm their use for modeling hepatosteatosis and VLDL metabolism *in vitro*. While these cells are not necessarily the ideal model for studying small differences between HLCs derived from different donors, they still show promise for the application of modeling genetic
mutations with robust phenotypes, as well as precision medicine approaches to testing drug hepatotoxicity.

2.6 Materials and Methods

**PBMC isolation and expansion**

Fresh peripheral blood was drawn from de-identified subjects at Columbia University Irving Medical Center (CUIMC), or from specific donors at the Amish Research Clinic in Lancaster, Pennsylvania, and shipped overnight at 4°C to New York City. The collection and usage of human blood were approved by the Institutional Review Boards at Columbia University Irving Medical Center and University of Maryland School of Medicine. PBMCs were immediately isolated from whole blood upon receipt through density gradient centrifugation with Ficoll-Paque Plus (GE Healthcare, 17144002), following the manufacturer’s instruction.

The culture and expansion of PBMCs for iPSC reprogramming were adapted from the protocol by Sommer et al. The enrichment medium consisted of QBSF-60 Stem Cell Medium (Quality Biological, 160-204-101), with 50 µg/mL ascorbic acid (Millipore Sigma, A4544), 50 ng/mL stem cell factor (SCF), 10 ng/mL interleukin 3 (IL-3), 2 U/mL erythropoietin (EPO), 40 ng/mL insulin-like growth factor (IGF-1), 1 µM dexamethasone (Dex; Millipore Sigma, 4902), and 25 µg/mL gentamicin (Thermo Fisher, 15750060). SCF, IL-3, EPO and IGF-1 were purchased from ProSpec (CYT-255, 210, 201, 216). 2 million freshly isolated PBMCs were seeded in a 35 mm dish, and the enrichment medium was replaced every 3 days.

**PBMC-to-iPSC reprogramming and iPSC culture**

250,000 PBMCs were used for CytoTune Sendai virus (Thermo Fisher, 16517) transduction, following the manufacturer’s instruction. Viruses were removed after 24 hours, and cells were given fresh enrichment medium. After 48 hours, reprogramming cells were transferred
to a 10 cm dish with a confluent feeder layer of iMEFs, which had been treated with 10 µg/mL mitomycin C (Tocris Bioscience, 50-07-7) for 2 hours to crosslink DNA. After 48 hours, the enrichment medium was replaced with iPSC culture medium, which consisted of KnockOut DMEM/F12 (Thermo Fisher, 12660012), 20% KnockOut Serum Replacement (KOSR; Thermo Fisher, 10828028), non-essential amino acids (NEAA; Thermo Fisher 11140050), β-mercaptoethanol (Thermo Fisher, 21985023), L-glutamine (Thermo Fisher, 35050061), 25 µg/mL gentamicin, and 10 ng/mL fibroblast growth factor 2 (FGF2; ProSpec CYT-085). iPSC medium was replenished daily while the colonies matured.

24 colonies were manually picked with a P20 pipet under an inverted microscope (Nikon, Eclipse TS100) and deposited onto new feeder layers. The 12 most successful clones were selected for the next passage, after which 6 clones were chosen for further propagation for each donor. iPSC colonies were detached from feeder layers by gentle scraping following incubation with 4 mg/mL collagenase type 4 (Worthington, LS004210) for 20 minutes at 37°C. 10 µM Y-27632 (Selleck Chemicals, S1049) was added to the medium for the first 16 hours post-passaging to support iPSC survival. The 6 clones chosen for propagation were adapted to feeder-free surfaces coated with Matrigel (Corning, 354277) at 1:100 dilution. Feeder-free iPSCs were maintained in mTeSR Plus (Stemcell Technologies, 05825) with daily medium change. Cells were passaged every 6 days by incubating with Accutase (Thermo Fisher, A1110501) for 5 minutes at 37°C. iPSCs were monitored daily for any morphological abnormalities and discarded if more than 10% of the colonies showed signs of differentiation. For banking iPSCs for future use, 1 to 5 million cells were cryopreserved in 1 mL freezing medium composed of 90% KOSR and 10% dimethyl sulfoxide (DMSO; Millipore Sigma, D8418), and supplemented with 10 µM Y-27632. iPSCs between passages 20-40 were used for analysis and differentiation.
Differentiation of iPSCs to HLCs

The optimized differentiation protocol was adapted from Peters et al.\textsuperscript{136} iPSCs were passaged as small colonies into 12-well plates pre-coated with 1:100 diluted reduced growth factor Geltrex (Thermo Fisher, A1413302). They were allowed to reach a confluency of 70-90% before endoderm induction, depending on the iPSC line. The DE medium consisted of RPMI (Thermo Fisher, 22400), NEAA, B27 minus insulin (Thermo Fisher, A1895601), 3 µM CHIR99021 (Stemcell Technologies, 72054), and 100 ng/mL Activin A. After 3 days of endoderm induction, cells were passaged by washing with PBS followed by incubation with Accumax (Millipore Sigma, A7089) for 5 minutes at 37°C. Cells were collected and resuspended in HE medium before seeding onto new Matrigel or Geltrex-coated plates at a concentration of ~0.26 million/cm\(^2\) (line-dependent). The HE medium consisted of RPMI, NEAA, B27 (Thermo Fisher, 17504044), 20 ng/mL bone morphogenetic protein 4 (BMP4), 10 ng/mL FGF2, and 0.5% DMSO. After 5 days of HE differentiation, cells were rinsed with DPBS and the medium was switched to IMH medium, which consisted of RPMI, NEAA, B27, 20 ng/mL hepatocyte growth factor (HGF), and 0.5% DMSO. After 5 days of IMH differentiation, cells were rinsed with DPBS and the medium was replaced with MH medium for hepatocyte maturation. The MH medium consisted of Hepatocyte Culture Medium (HCM) BulletKit (Lonza, CC-3198), 20 ng/mL HGF, 20 ng/mL Oncostatin M (OSM), 100 nM Dex, and 0.5% DMSO. All the growth factors for differentiation were purchased from ProSpec (CYP-145, 081, 085, 090, 231). Culture media were replenished daily during differentiation and maturation.

Primary cell culture and HepG2 culture

MEFs and HepG2 were purchased from ATCC (SCRC-1040 and HB-8065) and maintained in DMEM (Thermo Fisher, 11965118) supplemented with 10% FBS, 25 µg/mL
gentamicin, NEAA, β-mercaptoethanol, and sodium pyruvate (Thermo Fisher, 11360070). MEFs were cultured on surfaces pre-coated with 0.1% gelatin (Millipore Sigma, G1890), and HepG2 was cultured on 5 µg/cm² collagen I. Their media were replaced every 2 days and the cells were passaged every 4 to 6 days by incubating with 0.25% trypsin-EDTA (Thermo Fisher, 25200114) for 5 minutes at 37°C. MEFs were used up to passage 10 and HepG2 cells were used up to passage 30.

**Immunocytochemistry**

Cells grown in monolayers were fixed with 4% paraformaldehyde (PFA; Electron Microscopy Sciences, 15713; diluted with PBS) for 15 minutes and permeabilized with 0.1% Triton X-100 (Millipore Sigma, 93443) for 5 minutes at room temperature. Cells were first blocked with 10% goat serum or 10% donkey serum in PBS for 1 hour before incubating with any primary antibodies (Table 2.3). The secondary antibodies were conjugated with Alexa Fluor dyes 488 or 594, which were either derived from goat or donkey (Thermo Fisher, A11008, A11005, A32790, and A32758). Incubation with primary and secondary antibodies was performed at room temperature for 2 hours, after which cells were incubated with 1 µg/mL 4',6-diamidino-2-phenylindole dihydrochloride (DAPI; Thermo Fisher D1306) for 10 minutes to stain their nuclei. Fluorescent imaging was performed either on a regular inverted microscope (Nikon, Eclipse TE2000) or on a scanning confocal microscope (Nikon, Eclipse Ti), using the Nikon NIS-Elements software.
Table 2.3: List of Primary Antibodies

<table>
<thead>
<tr>
<th>Target Antigen</th>
<th>Manufacturer</th>
<th>Product #</th>
<th>Assay</th>
</tr>
</thead>
<tbody>
<tr>
<td>AAT</td>
<td>Thermo Fisher</td>
<td>PA5-16661</td>
<td>WB, IF, Flow</td>
</tr>
<tr>
<td>AFP</td>
<td>R&amp;D Systems</td>
<td>MAB1369</td>
<td>IF, Flow</td>
</tr>
<tr>
<td>Albumin</td>
<td>Sigma-Aldrich</td>
<td>A7544</td>
<td>IP</td>
</tr>
<tr>
<td>Albumin</td>
<td>Bethyl Laboratories</td>
<td>A80-129A</td>
<td>IF, Flow, ELISA</td>
</tr>
<tr>
<td>Albumin</td>
<td>R&amp;D Systems</td>
<td>MAB1455</td>
<td>WB</td>
</tr>
<tr>
<td>APC mouse IgG2a κ isotype</td>
<td>Biolegend</td>
<td>400219</td>
<td>Flow</td>
</tr>
<tr>
<td>apoA1</td>
<td>Calbiochem</td>
<td>178463</td>
<td>IP</td>
</tr>
<tr>
<td>apoA1</td>
<td>Thermo Fisher</td>
<td>PA5-19784</td>
<td>WB</td>
</tr>
<tr>
<td>apoB100</td>
<td>Calbiochem</td>
<td>178467</td>
<td>IP</td>
</tr>
<tr>
<td>apoB100</td>
<td>LS-Bio</td>
<td>LS-6180</td>
<td>WB</td>
</tr>
<tr>
<td>apoE</td>
<td>Genetex</td>
<td>gtx-27620</td>
<td>WB</td>
</tr>
<tr>
<td>ASGPR</td>
<td>BD</td>
<td>563655</td>
<td>Flow</td>
</tr>
<tr>
<td>C-Kit</td>
<td>Biolegend</td>
<td>313232</td>
<td>Flow</td>
</tr>
<tr>
<td>Collagen I</td>
<td>Abcam</td>
<td>ab34710</td>
<td>IF</td>
</tr>
<tr>
<td>CXCR4</td>
<td>Biolegend</td>
<td>306510</td>
<td>Flow</td>
</tr>
<tr>
<td>Cyclophilin B</td>
<td>abcam</td>
<td>ab178397</td>
<td>WB</td>
</tr>
<tr>
<td>CYP3A4</td>
<td>Novus Biologicals</td>
<td>NBP2-37502</td>
<td>WB, IF</td>
</tr>
<tr>
<td>E-cadherin</td>
<td>Cell Signaling Technology</td>
<td>3195</td>
<td>IF, flow</td>
</tr>
<tr>
<td>FAS</td>
<td>BD</td>
<td>610962</td>
<td>WB</td>
</tr>
<tr>
<td>FITC mouse IgG1 κ isotype</td>
<td>Biolegend</td>
<td>400109</td>
<td>Flow</td>
</tr>
<tr>
<td>HNF1α</td>
<td>Cell Signaling Technology</td>
<td>89670</td>
<td>IF</td>
</tr>
<tr>
<td>HNF4α</td>
<td>R&amp;D Systems</td>
<td>MAB4605</td>
<td>WB, IF</td>
</tr>
<tr>
<td>LDL-R</td>
<td>Cayman Chemical</td>
<td>10007665</td>
<td>WB</td>
</tr>
<tr>
<td>MTP</td>
<td>abcam</td>
<td>ab86759</td>
<td>WB</td>
</tr>
<tr>
<td>Oct4</td>
<td>Applied StemCell</td>
<td>ASK-3006</td>
<td>IF</td>
</tr>
<tr>
<td>PE mouse IgG1 κ isotype</td>
<td>BD</td>
<td>551436</td>
<td>Flow</td>
</tr>
<tr>
<td>PLIN2</td>
<td>Novus Biologicals</td>
<td>NBP2-23486</td>
<td>WB</td>
</tr>
<tr>
<td>Sox2</td>
<td>Applied StemCell</td>
<td>ASK-3006</td>
<td>IF</td>
</tr>
<tr>
<td>SRB1</td>
<td>Novus Biologicals</td>
<td>nb400-101</td>
<td>WB</td>
</tr>
<tr>
<td>SSEA4</td>
<td>Applied StemCell</td>
<td>ASK-3006</td>
<td>IF</td>
</tr>
<tr>
<td>TRA-1-60</td>
<td>Applied StemCell</td>
<td>ASK-3006</td>
<td>IF</td>
</tr>
<tr>
<td>Vinculin</td>
<td>Cell Signaling Technology</td>
<td>CS4650S</td>
<td>WB</td>
</tr>
<tr>
<td>ZO-1</td>
<td>Cell Signaling Technology</td>
<td>13663</td>
<td>IF</td>
</tr>
<tr>
<td>β-actin</td>
<td>Sigma-Aldrich</td>
<td>A5441</td>
<td>WB</td>
</tr>
</tbody>
</table>

Abbreviations: WB, western blot; IF, immunofluorescence; Flow, flow cytometry; IP, immunoprecipitation; ELISA, enzyme linked immunosorbent assay.
Flow cytometry

Flow cytometry of surface receptors (CXCR4, C-Kit, and ASGPR) was performed using live cells. After dissociation and blocking in 10% goat serum, DE cells or HLCs were incubated with the following antibodies conjugated with fluorophores (Table 2.3): anti-CXCR4, anti-C-Kit, and anti-ASGPR. Their respective isotype controls were included in each experiment: APC mouse IgG2a κ isotype, FITC mouse IgG1 κ Isotype, and PE mouse IgG1 κ Isotype. Live cells were incubated with test or control antibodies for 20 minutes at room temperature before being analyzed on the flow cytometer (BD, LSRFortessa). Flow cytometry of intracellular proteins (AAT, AFP, and albumin) and E-cadherin was performed using fixed cells. Immunostaining followed the same steps as discussed above and cells that were stained with secondary antibodies alone served as negative controls in the flow experiment. Results were analyzed using the FlowJo 7.6 software.

Gene expression analysis

Primers were designed using PrimerBank, NCBI Primer-BLAST or found through literature search, and required to span exon-exon junctions to avoid amplification of genomic DNA (Table 2.4). Total RNA was extracted using the RNeasy Mini Kit (Qiagen, 74104). Total RNA from adult liver and fetal liver were purchased from a commercial source (Takara Bio, 636531 and 636540; lot # 1402003 and 1402005). Reverse transcription was performed using the iScript cDNA Synthesis Kit (Biorad, 1708891) on the MyCycler thermocycler (Biorad), following the manufacturer’s instruction. Quantitative PCR (qPCR) reactions were made with SsoAdvanced Universal SYBR Green Supermix (Biorad, 1725271) and run on the StepOnePlus Real-Time PCR System (Applied Biosystems). Glyceraldehyde-3-phosphate dehydrogenase (GAPDH) was included in all RT-qPCR experiments as the housekeeping gene.
Table 2.4: List of qPCR Primers

<table>
<thead>
<tr>
<th>Target Gene</th>
<th>Gene ID</th>
<th>Forward Primer (5’-3’)</th>
<th>Reverse Primer (5’-3’)</th>
<th>Amplicon Size (bp)</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>AAT</td>
<td>5265</td>
<td>GATCAACGATTACGGAGAGAA</td>
<td>CCTAAACGCTTCATCATAGGCA</td>
<td>207</td>
<td>*</td>
</tr>
<tr>
<td>AFP</td>
<td>174</td>
<td>CTTGCGCTGCTGCTATGA</td>
<td>GCACTGATTTAACAAGCTGCT</td>
<td>131</td>
<td>*</td>
</tr>
<tr>
<td>Albumin</td>
<td>213</td>
<td>TTATGCCCGCGAAACTCTTT</td>
<td>AGTCCTGTTGGCAGAGCAA</td>
<td>148</td>
<td>*</td>
</tr>
<tr>
<td>ApoB</td>
<td>338</td>
<td>ACACACTGGACGCTAAGAGGA</td>
<td>ACTTGTTGCACATCCCATACT</td>
<td>106</td>
<td>*</td>
</tr>
<tr>
<td>CYP1A2</td>
<td>1544</td>
<td>GATGGCCGCTGCTCAGAGGA</td>
<td>CCCGGAACAGTGTCTTGCA</td>
<td>200</td>
<td>**</td>
</tr>
<tr>
<td>CYP2B6</td>
<td>1555</td>
<td>GCACCTCTACAGGACTCTTT</td>
<td>CCCAGGGGTACGGTGAGAC</td>
<td>185</td>
<td>*</td>
</tr>
<tr>
<td>CYP3A4</td>
<td>1576</td>
<td>AAGTCGCCCTGAGATACAACA</td>
<td>AAGGAGGAGAACACTGCTGAG</td>
<td>174</td>
<td>*</td>
</tr>
<tr>
<td>GAPDH</td>
<td>2597</td>
<td>GAGCCGGAGATCCGCCAAT</td>
<td>GGCGTTGTGCTACATTCTCATG</td>
<td>197</td>
<td>*</td>
</tr>
<tr>
<td>HNF4α</td>
<td>3172</td>
<td>GGCCACTGACCATCCGACTCTT</td>
<td>CAGCACCAGCTGCTAAGGG</td>
<td>220</td>
<td>*</td>
</tr>
<tr>
<td>LXRα</td>
<td>10062</td>
<td>GCAGATTTGGCGCTTTGCTCA</td>
<td>TCCGGAGGCTACCCAGGGTTTC</td>
<td>187</td>
<td>***</td>
</tr>
</tbody>
</table>

Source: *PrimerBank, **NCBI Primer-Blast, ***Hepatol Int. 2009 Sep; 3(3): 490–496.173

Cell secretion assays

The concentration of HLC-secreted albumin was measured using the ELISA Starter Accessory Kit and Human Albumin ELISA Quantitation Set (Bethyl Laboratories, E101 and E80-129), while the concentration of apoB was measured using the Human Apolipoprotein B ELISA development kit (Mabtech, 3715-1H-6), following the manufacturer’s instruction. The samples were measured on the FLUOstar OPTIMA microplate reader (BMG Labtech) with an absorbance of 430 nm. Because the limitation of detection of most commercial urea measurement kits is relatively high compared to the urea concentration in HLC medium, we first concentrated the HLC medium 10X through lyophilization on the FreeZone Freezer Dryer followed by reconstitution in water. The urea concentration was measured with the Urea Nitrogen (BUN) Test (Stanbio, 0580-250) and read on the FLUOstar OPTIMA microplate reader with an absorbance of 544 nm.
The measured amount of albumin, apoB and urea in the samples was divided by the number of days the cells were in that medium to calculate the daily average. These values were further normalized to the total protein or total cell number so that comparisons can be made across different experiments and with other reports from the literature.

*CDFDA transport, LDL uptake, and glycogen storage*

5(6)-carboxy-2',7'-dichlorofluorescein diacetate (CDFDA; Millipore Sigma, 21884) was reconstituted in DMSO at a concentration of 10 mM. To visualize the transport activity of MRP2 on the bile canaliculi, cells were first incubated with 5 µM CDFDA for 30 minutes at 37°C. Thereafter, CDFDA was removed, and the cells were washed thoroughly and returned to 37°C for 2 hours. Live cell imaging was performed on a regular, inverted fluorescent microscope.

Due to the sensitivity of HLCs to nutrient change, cells were not starved before treating with human LDL particles conjugated to a red fluorescent dye, pHrodo Red-LDL (Thermo Fisher, L34356). Instead, the treatment time was extended to 4 hours. The rest of the manufacturer’s instruction was followed, and cells were treated with 15 µg/mL LDL at 37°C. Live cell imaging was performed on a regular, inverted fluorescent microscope.

To visualize glycogen storage in HLCs, the cells were first fixed with 4% PFA for 15 minutes. The standard procedure of the Periodic Acid-Schiff (PAS) Kit (Millipore Sigma, 395B) was followed. Briefly, the cells were immersed in periodic acid for 5 minutes during which glycols were oxidized to aldehydes. After washing, cells were immersed in Schiff’s reagent for 15 minutes during which the glycol-containing components were stained with a pink color. The cells were counterstained with hematoxylin for 90 seconds. Color imaging was performed on an inverted microscope (Nikon, Eclipse TE2000) equipped with a color camera.
Drug treatments and CYP enzyme inductions

Acetaminophen (APAP) is a widely used hepatotoxin in liver injury studies; omeprazole, phenobarbital and rifampin are strong inducers of cytochrome P450 (CYP) enzymes 1A2, 2B6 and 3A4, respectively. APAP (Millipore Sigma, A7085) was prepared in complete MH medium. Stock solutions of omeprazole, phenobarbital (Millipore Sigma, O104 and P1636) and rifampin (Alfa Aesar, J60836) were prepared in DMSO at 200 mM, 2 M and 50 mM, respectively. For the hepatotoxicity test, cells were treated with 0, 10, or 20 mM APAP; for the CYP induction test, cells were treated with omeprazole (100 and 200 µM), phenobarbital (1 and 2 mM), rifampin (25 and 50 µM) or DMSO, which served as the vehicle control. All the drug treatments lasted for 48 hours before the cells were used for analysis.

Cell viability assays

The LIVE/DEAD Viability/Cytotoxicity Kit (Thermo Fisher, L3224) was used to qualitatively assess the health of the 2D cells. The reagents were kept at -20°C until they were ready to use. Calcein AM and ethidium homodimer-1 (EthD-1) were combined in sterile DPBS with calcium and magnesium at the recommended concentrations of 2 µM calcein AM and 4 µM EthD-1. Cells were immersed in this solution for 20 minutes at 37°C. After careful washing, live cell imaging was performed on a regular inverted fluorescent microscope.

Western Blot Analysis

Cells were collected in lysis buffer (62.5 mM sucrose, 0.5% sodium deoxycholate, 0.5% Triton X-100, 50 mM Tris-HCl, 150 mM NaCl, 1 mM benzamidine, 5 mM EDTA, 2 units/mL aprotinin, 50 µg/mL leupeptin, 50 µg/mL pepstatin A, and 10 mM HEPES, pH 8.0) with 0.86 mM freshly added PMSF (phenylmethylsulfonyl fluoride), and immediately frozen at -80°C. Cell lysates were thawed on ice and spun down at maximum speed, 4°C for 20 minutes to clear nuclear
material and debris. Supernatant protein extracts were diluted in equal volume of 2X reducing sample buffer [0.125 M Tris-HCl, pH 6.8, 10% SDS (w/v), 20% glycerol (v/v), 2% β-mercaptoethanol (v/v), and 0.01% bromophenol blue (w/v)] and boiled for 5 minutes. A small aliquot of undiluted protein extract was reserved for protein determination via Pierce BCA Protein Assay Kit (Thermo Scientific, 23225). Equal amounts of protein were loaded per well and separated by SDS-PAGE on appropriate percentage gels and electrotransferred to 0.2 μm nitrocellulose membranes (Biorad, 162-0150). Membranes were blocked in 5% milk for 1 hour at room temperature, then incubated with primary antibodies overnight at 4°C (Table 2.3). The following day, membranes were washed thoroughly with 0.05% Tween-PBS and incubated with species-specific HRP-conjugated secondary antibodies for 1 hour at room temperature. Protein bands were visualized using SuperSignal West Pico Chemiluminescent Substrate (Thermo Scientific, 34578).

Radiolabeling of newly synthesized proteins and immunoprecipitation

Cells were pre-treated for 1 hour in 1.5% FAF-BSA (MP Biomedicals, 0215240150) in methionine/cysteine/serum-free DMEM (Thermo Scientific, 21013-024) and 20 μM CP-10447 (a generous gift from Pfizer) or DMSO vehicle. Pre-treatment medium was discarded, and cells were pulsed for 2 hours in fresh 1.5% FAF-BSA in methionine/cysteine/serum-free DMEM medium containing 150 μCi/mL 35S-methionine/cysteine (PerkinElmer, NEG- 072014MC) and 20 μM CP-10447 or DMSO vehicle. Following labeling, media were transferred to tubes already containing a mixture of protease inhibitors (30 μl/mL protease inhibitors, 1 mM benzamidine, 5 mM EDTA, 100 units/mL aprotinin, 50 μg/mL leupeptin, 50 μg/mL pepstatin A, and 10 mM Hepes, pH 8.0) and 0.86 mM freshly added PMSF. Cells were rinsed thoroughly in ice cold PBS and collected in
lysis buffer containing 0.86 mM freshly added PMSF. Cell and medium samples were frozen immediately until further processing.

Immunoprecipitation of apoB, apoA1, or albumin in cells and medium was carried out as described previously. Briefly, cell lysates were thawed on ice and spun down as described above, reserving a small aliquot for BCA protein determination. 5XNET buffer was added to thawed medium samples. rProtein G Agarose Beads (Thermo Scientific, 20399) and appropriate antibody were added to each medium and cell tube, and mixed overnight at 4°C. The next day, supernatants were collected for the immunoprecipitation of additional proteins; beads were washed thoroughly with 1XNET buffer and proteins were released from beads in reducing sample buffer (described above) by boiling for 5 min. Equal amounts of samples (based on BCA determination) were resolved by SDS-PAGE. The gel was fixed in destaining solution (25% methanol, 20% acetic acid, in water), then treated with Autofluor intensifying solution (National Diagnostics) and, after drying, exposed to X-ray film (Thermo Scientific, 34091) at -80°C. Incorporation of radioactive label into immunoprecipitated proteins was quantified either by densitometric analysis of appropriate bands on the film, or, when indicated, the areas of the gels corresponding to the protein of interest were excised and radioactivity was quantified by liquid scintillation counting and expressed as counts per minute (CPM) per ug of cell protein.

Sucrose gradient ultracentrifugation of secreted lipoproteins

Sucrose gradient ultracentrifugation was conducted as described previously. Briefly, the sucrose gradient was formed by layering from the bottom of a 12 mL tube (Beckman Coulter): 2 mL of 47% sucrose, 2 mL of 25% sucrose, 5 mL of medium sample adjusted to 12.5% sucrose, and 3 mL of PBS. All solutions contained 0.1 mM leupeptin, 1 μM pepstatin A, 0.86 mM PMSF, 100 units/mL aprotinin, 5 μM ALLN, 5 μM EDTA, 150 mM NaCl, and 50 mM PBS, pH 7.4. The
gradients were spun at 35,000 rpm in a Beckman SW40 rotor for 65 hours at 12°C, then unloaded into 12X 1 mL fractions. Fractional density was determined as weight (g) per 1 mL. Fractions were subjected to immunoprecipitation, separation on SDS-PAGE, and exposure to X-ray film, as described above.

Determination of triglyceride secretion

Following 1 hour lipid fast in “starvation medium” consisted of 1.5% FAF-BSA (MP Biomedicals, 0215240150) in William’s Eagle Media (Thermo Fisher, A1217601), cells were pulsed with 20 μCi/mL 3H-glycerol (Perkin Elmer, NET022L001MC), +/- 0.4 mM OA, and 20μM CP-10447 or DMSO vehicle, in starvation medium. After 4 hours of incubation, pulse medium was collected directly into 20 volumes of Chloroform:Methanol (2:1, v/v) and lipids were extracted from pulse medium and cells according to previously published methods. Lipid extracts were evaporated under N₂ gas, resuspended in hexane, and spotted on Silica gel 60 TLC plates (Millipore Sigma, 1057480001). The lipid species were separated using a mobile phase of hexane/diethyl ether/acetic acid (70:30:1) and visualized in iodine vapor. The TG spots were scraped from the TLC plate into scintillation vials (Beckman Coulter, 592928), resuspended in 5 mL scintillation fluid (National Diagnostics, LS2754X4LTRCS), and the ³H counts in each spot were measured using a beta counter (Beckman Coulter). Total protein was recovered by solubilizing the delipidated cells in the plate in 0.1 N NaOH and measured by BCA assay. Lipid ³H counts were then normalized to total cell protein content.

Statistical analysis

Three biological replicates were used in each experiment and the results were presented as mean ± SEM, unless specified otherwise in the texts. Statistical significance was determined by Student t-test, where * represents P-value < 0.05, ** represents P-value < 0.01, *** represents P-
value < 0.001. In experiments with more than two independent groups, one-way ANOVA was first performed followed by multiple comparison t-test with Holm-Sidak correction. Microsoft Excel and GraphPad Prism 7 were used in data processing and presentation.
Chapter 3: CRISPR/Cas9-mediated Knockout of TM6SF2 in iPSCs

Due to the constraints described in the previous chapter to investigating the function of TM6SF2 in sibling matched HLCs, we decided to employ the use of CRISPR, the newly popular gene editing technique, to improve our experimental model. We sought to produce isogenic iPSCs, differing only in their expression of the intact or disrupted TM6SF2 gene. Successfully edited cells at the iPSC level offer the advantage of both a theoretically unlimited expansion without significant concern for cell senescence, as well as the capability to differentiate to many cell types, including HLCs, while minimizing the issue of donor variability and allowing for a more promising exploration of TM6SF2 function.\(^\text{175}\)

3.1 CRISPR/Cas9 editing strategy and guide RNA screening

Clustered regularly interspaced short palindromic repeat (CRISPR) and CRISPR-associated proteins 9 (CRISPR/Cas9) have become a popular and robust genetic engineering technology for generating stable and complete knockout cell lines. The commonly used CRISPR/Cas9 system from \textit{Streptococcus pyogenes} utilizes a 20 bp guide RNA (gRNA) that recognizes a specific DNA sequence upstream of a protospacer adjacent motif (PAM) sequence (5′-NGG-3′), and introduces a double-stranded break (DSB) in the target DNA.\(^\text{176}\) The DSB can be repaired either via nonhomologous end joining (NHEJ) or homology directed repair (HDR). NHEJ is error-prone and results in base insertions and/or deletions (indels) at the cleavage site, whereas HDR employs a template sequence which directs the repair of the cleavage site, for example, to introduce, or knock-in a specific point mutation or a premature termination codon (PTC) that halts translation of the peptide. However, high knockout efficiency can also be achieved in many cell types via the NHEJ pathway, since this mechanism often randomly induces frameshifts that lead to the generation of PTCs.\(^\text{177}\) Although HDR could theoretically be utilized
for the conversion of the TM6SF2 major variant (E167) to the minor variant (K167) following Cas9 cutting, this method would have been far less efficient than NHEJ, with no guarantee for success, and was beyond our technical capabilities at the time. Additionally, since the sibling-HLC experiments (and the in vivo kinetic data) did not display a strong observable phenotype, we chose to employ NHEJ repair to knockout TM6SF2, hypothesizing that a complete lack of TM6SF2 would magnify the pathological impact of the minor variant.

A schematic representation of the overall iPSC CRISPR editing strategy that we employed is summarized in Figure 3.1. Production of the CRISPR editing machinery began with the generation of two lentiviruses (LV) to deliver the Cas9 transgene and the gRNA sequence specifically designed to target the TM6SF2 gene. The plasmid carrying the Cas9 gene (which was fused to a FLAG-tag for easier detection) also harbored the gene conferring puromycin resistance, which would allow for the selection of successfully transduced cells following short-term puromycin treatment (Figure 3.1, inset top). To minimize off-target cutting, the Cas9 gene was positioned under the control of the doxycycline inducible Tet ON promoter; all other genetic components were positioned under constitutive promoters. The plasmid carrying the gRNA sequence was designed to additionally deliver the EGFP (Enhanced Green Fluorescent Protein) gene to allow for the separation and collection of successfully transduced cells via fluorescence activated cell sorting (FACS) (Figure 3.1, inset bottom). To maximize the chances for editing success, three different gRNAs (denoted SA, SB, and SC) targeting the TM6SF2 gene were designed using the Synthego Knockout Guide Design web tool. For creating stable expression cells, the purchased gRNAs were converted to their corresponding double-stranded DNA sequences and cloned into individual lentiviral plasmids. (Plasmid construction, lentivirus
production in HEK293T cells, and gRNA screening are detailed in the Detailed Methods section at the end of this chapter.)

**Figure 3.1** Schematic illustrating the overall process of developing isogenic HLCs through genetic editing and cell screening in LG2-C6 iPSCs. Inset top: Design of plasmid carrying FLAG-tagged Cas9 transgene under doxycycline-inducible promoter and puromycin resistance (PuroR). Inset bottom: Design of plasmid carrying gRNA and EGFP transgene. Created with Biorender.com.

Since iPSCs are known to be more resistant to both viral and non-viral transgene delivery methods compared to other cell types, the gRNA sequences were pre-screened in HEK293T cells to assess their targeting efficiency. To do so, native HEK293T were transduced with the first LV to generate a doxycycline-inducible Cas9-expressing line, which was followed by puromycin selection. The surviving successfully transduced cells were then treated with doxycycline to induce Cas9 transcription, which was confirmed by immunofluorescence microscopy (**Figure 3.2A**). HEK293T are easily transfected, so gRNAs were delivered (non-virally) via nanoparticle complexes composed of positively charged Lipofectamine 3000 and negatively charged gRNA. Following transfection, genomic DNA (gDNA) was collected and a 1000 bp DNA segment containing the Cas9 cut site was amplified; primers were designed in such a way that fragments of unequal length (300 and 700 bp) flanked the cleavage site. The 1000 bp amplicon was then subject to the T7E1 mismatch cleavage assay, in which the T7E1 enzyme recognizes and cuts heteroduplexes (which are present in locations where NHEJ has occurred). The products from the
mismatch assay were then resolved on agarose gel; bands intensities of the cleaved (300 and 700 bp) and un-cleaved (1000 bp) DNA fragments were measured to calculate the knockout efficiency of the three gRNAs (Figure 3.2B). Our results indicated that all three gRNAs were effective in guiding Cas9 to disrupt the TM6SF2 gene and non-specific bands were not detected. Although not significantly higher than the other gRNAs (following Tukey’s multiple comparisons test), SA gRNA showed the highest cleavage efficiency in HEK293T at 32% and was therefore chosen for subsequent experiments (Figure 3.2C).

Figure 3.2 Pre-screening of TM6SF2 targeting guide RNAs in HEK-293T cells. A) Representative images of FLAG-tagged Cas9 immunostaining in HEK293T-Cas9-PuroR cells. B) Agarose gel separation of T1E7 mismatch cleavage assay products to assess the knockout efficiency of three gRNA sequences, SA, SB, and SC versus UT (untreated). The top red arrowhead indicates the 700 BP fragment and the bottom red arrowhead indicates the 300 BP fragment. Uncleaved products are 1000 BP. C) Calculated cleavage efficiency in samples treated with gRNA SA, SB, and SC based on percentage of cleaved DNA bands. Scale bar: 100 μm. n=3.

Next, to test the viral delivery and editing efficiency of the gRNA LV, HepG2 cells already expressing Cas9-PuroR were transduced with the gRNA LV at a titer of 30 IU/cell. The HepG2
cells displayed nearly 100% transduction efficiency, as assessed by EGFP fluorescence, suggesting an appropriate titer was selected so FACS was not performed (Figure 3.3A). For Cas9 induction, HepG2-Cas9-PuroR_EGFP+ cells were treated with doxycycline either for 48 hours (KO-1), or 48 hours followed by an additional 48-hour exposure after passaging (KO-2). Following gDNA extraction and amplification of the 1000 bp fragment containing the Cas9 target site, the amplicon was analyzed with two methods: (1) T7E1 mismatch cleavage assay (Figure 3.3B), and (2) Sanger sequencing/trace decomposition using TIDE and ICE (see Detailed Methods) (Figure 3.3C). The results from both analyses were comparable, showing close to 40% knockout efficiency. Interestingly, the percentage of edited cells appeared to reach an upper limit in this HepG2 system, since further doxycycline treatment did not result in more detectable cleavage and only minimally increased indel generation in KO-2. As such, 48-hour doxycycline treatment was considered a suitable exposure time while also minimizing unwanted Cas9 activity.

Figure 3.3 TM6SF2-gRNA lentivirus transduction in HepG2-Cas9-PuroR cells. A) Strong EGFP expression after gRNA lentivirus transduction in HepG2 cells. B) T1E7 mismatch cleavage assay to assess the knockout efficiency in HepG2 WT, TM6SF2-KO-1, and TM6SF-KO-2. C) Sanger sequencing/trace decomposition using ICE (Synthego) to assess the knockout efficiency. Scale bar: 100 µm.

3.2 iPSC transduction with Cas9/gRNA lentiviruses and initial screen

Having tested all the genetic components in HEK293T and HepG2 and demonstrated precise and efficient targeting of the lentiviral system, we continued with our approach to create a
CRISPR-mediated TM6SF2-KO iPSC line. Unlike the cell lines used previously, iPSCs tend to exhibit lower transgene expression levels and are susceptible to damage from the DSBs generated by CRISPR/Cas9 gene cleavage, which can pose a challenge to iPSC genome editing.\textsuperscript{179} Therefore, care was taken to select for transduction a highly proliferative and dependably well-differentiating iPSC line, followed by confirmation of transgene expression. At the time that these studies were performed, the LG2-C6 iPSC line was the most reliable, so it was selected for all further CRISPR/Cas9 experiments; it should be noted that this line bore the minor TM6SF2 (KK) alleles.

First, the LG2-C6 iPSC line was transduced at three different titers, (5, 10, and 20 IU/cell) with doxycycline-inducible Cas9-carrying LV, followed by puromycin selection (\textbf{Figure 3.4A}). Upon doxycycline induction, the iPSCs exhibited dose-dependent Cas9 protein expression, as confirmed by Western blot (\textbf{Figure 3.4B}). Furthermore, no “leaky” Cas9 expression was detected in the control iPSCs that did not receive doxycycline, indicating tight regulation of Cas9 expression under the Tet ON promoter. iPSCs Next, LV carrying the SA gRNA sequence were transduced into the LG2-C6-Cas9-PuroR line (10 and 20 IU/cell, pre dox treatment). Most cells showed robust EGFP expression when viewed under a fluorescence microscope, indicating high transduction efficiency of the gRNA-carrying plasmid (\textbf{Figure 3.4C}). This was confirmed by FACS, which determined that ~75\% of LV-treated iPSCs were EGFP\(^{\text{+}}\) (data not shown). FACS was also used to collect a pure Cas9-PuroR-EGFP\(^{\text{+}}\) iPSC population, which was re-cultured on Matrigel-coated plates for subsequent use. Further visual assessment of these cells revealed complete EGFP expression throughout the culture (\textbf{Figure 3.4D}).
Figure 3.4 Confirmation of Cas9-PuroR and TM6SF2-gRNA lentivirus (LV) transduction in LG2-C6 iPSCs. A) Puromycin selection in iPSCs treated with 0 or 10 IU/cell of Cas9-PuroR LV. B) Western blot of FLAG-tagged Cas9 in iPSCs transduced with Cas9 lentiviruses at 5, 10, and 20 IU/cell (Dox = doxycycline treatment, 48 hours). C) EGFP expression in iPSCs after gRNA lentivirus transduction, before FACS and D) after FACS.

Previous experience had shown us that dox-induced Cas9 activity displayed nearly undetectable gene editing in the Cas9-PuroR-EGFP+ iPSCs when analyzed as a heterogeneous population (data not shown). Given the ability for a single iPSC to proliferate rapidly into a genetically homogenous culture, we elected to expand subclones from the transduced iPSC culture and screen those for successful TM6SF2 knockout. To do so, Cas9-PuroR-EGFP+ iPSCs were treated with doxycycline for 48 hours to induce Cas9 expression; iPSCs were then detached into a mono-dispersed suspension and seeded onto Matrigel-coated 96-well plates at a target concentration of one cell per well. Immediately following cell seeding, and for every day following until time of passage, wells were visually assessed under a bright field and fluorescent microscope; only those wells containing a single, uniform colony were maintained for further analysis (Figure 3.5). Any wells displaying more than one colony, or colonies with incomplete or inconsistent EGFP were discarded.
Figure 3.5 Schematic illustrating the process of visually screening iPSCs for the selection of LG2-C6-Cas9-PuroR-EGFP⁺ monoclones. 96-well plate graphic created with Biorender.com.

Twenty-one subclones were expanded, followed by genomic DNA (gDNA) collection and assessment for TM6SF2-KO efficiency by Sanger sequencing/ICE trace decomposition, as described above; results are summarized in Figure 3.6A, where M10 or M20 represents clones that had previously been treated with 10 or 20 IU/cell of Cas9 lentivirus, respectively. The ICE score represents the indel percentage calculated by comparing the edited sequence trace to that of the unedited control, and the KO score represents the proportion of cells exhibiting either a 1 or 2 bp frameshift or >21 bp indel, indicating the percentage of presumed TM6SF2 functional knockout. Only 1 out of the 21 subclones (5%) from the transduced iPSCs displayed any detectable gene editing (Figure 3.6B); within this clone, called E5, only a 15% KO score was determined by ICE, for an overall knockout efficiency of just 0.75% in these iPSCs. Additionally, while only ostensible monoclonal iPSCs were used for the screen, the E5 population exhibited at least 8 unique indels (Figure 3.6C), indicating that it was not, in fact, monoclonal (as bi-allelic cells would display a maximum of 2 unique indels per clone). This discovery led us to the hypothesis that the iPSCs had not been allowed ample time between doxycycline treatment and the terminal dilution for
monoclonal plating. As such, two scenarios likely occurred simultaneously: 1. Cas9 had not been expressed long enough prior to sorting/single cell-plating to effectively induce editing in a significant population of cells (hence the low editing efficiency), and 2. Cas9 was active in some of the post-sorted single cells as they proliferated, leading to inconsistent Cas9 editing in the daughter cells and thus incorporation of >2 types of indels. T7E1 mismatch cleavage assay confirmed that Cas9 editing was incomplete in E5, as evidenced by the remaining uncleaved 1000kb base band (Figure 3.6D).

![Figure 3.6 First round of clonal screening for complete TM6SF2-KO in LG2-C6 iPSCs.](image)

A) Transduced, puromycin-selected and EGFP-sorted cells were selected and analyzed
through Sanger sequencing/trace decomposition. B) Chromatograms of E5 (top) was compared to the unedited control (bottom) and used for trace decomposition to elucidate indels at the TM6SF2 cleavage site. C) Pie chart showing indel composition of E5 iPSCs. Eight indels were observed, accounting for 18% editing. D) T7E1 mismatch cleavage assay detected 29% TM6SF-KD in gDNA isolated from clone E5.

### 3.3 Monoclonal isolation of TM6SF2 KO iPSCs and confirmation of successful KO

An iPSC line with 15% functional knockout in *TM6SF2* was not sufficient for our study, especially given the mild phenotype previously observed. However, since it was clear that clone E5 contained some cells with the desired editing, we decided to further purify that line to obtain a complete and uniform TM6SF2-KO clone; therefore, a second round of iPSC clonal screening was performed. In order to increase the chances of growing colonies derived from a single iPSC, we changed our approach slightly: rather than attempting to plate one cell per well as before, we significantly diluted the mono-dispersed iPSC suspension and seeded the cells onto a 10 cm coated cell-culture dish at a concentration of ~100 cells/cm². After one week, 45 individual colonies were manually picked, expanded, and screened as before (Figure 3.7A); the ICE results are summarized in Figure 3.7B. The majority of the E5 subclones exhibited incomplete knockouts, with some indels present at the cleavage site. Three subclones had 97% indels compared to the unedited control, and 2 of these were identified by the program as functionally complete knockouts. The subclone E5-17 had a clean chromatogram at the *TM6SF2* cleavage site (Figure 3.7D, bottom), indicating it came from a purely monoclonal iPSC colony with one genotype. Trace decomposition demonstrated that it had only 2 indels, a 10 bp deletion in 50% of the sample and an 11 bp deletion in the other 50%, indicative of a compound-heterozygote (Figure 3.7C).
Figure 3.7 Second round of clonal screening for complete TM6SF2-KO in LG2-C6 iPSCs. A) Graphical summary of monoclonal isolation and screening of E5-subclones. Created using Biorender.com B) Subclone E5 was used to culture further subclones, which were expanded and analyzed through Sanger sequencing/trace decomposition. C) Pie chart showing different indel percentages in E5 and E5-17 iPSCs. Two indels, 10 and 11 bp deletions, were observed in E5-17. (D) Chromatograms of E5-17 (top) was compared to the unedited control (bottom) and used for trace decomposition to elucidate indels at the TM6SF2 cleavage site.

We sought to further corroborate the results from the trace decomposition analysis; since there was no robust and specific commercial antibody against the human TM6SF2 protein at the
time of this study, we evaluated the expression of TM6SF2 mRNA transcript levels. However, the effect of knocking out a gene may not be detectable at the transcript level because nonsense-mediated mRNA decay (NMD) has variable efficiency in targeting transcripts containing premature termination codons.\textsuperscript{180,181} As such, new RT-qPCR primers were designed following the method of Yu \textit{et al} to specifically overlap with the indel position on the TM6SF2 cDNA once reverse transcribed.\textsuperscript{182} Since it was known that clone E5-17 possessed only 10 and 11 bp deletions at the Cas9 cut site, primers directed to the intact sequence should not efficiently bind and amplify the edited transcripts, thus providing a quantitative method for assessing KO efficiency (\textbf{Figure 3.8A}).

Results from RT-PCR amplification showed that, while E5-17 demonstrated \textit{increased} expression compared to an unedited clone when primers targeting an area of TM6SF2 far downstream of the cut site were used, there was \textasciitilde80\% reduction in TM6SF2 expression evident with the cut-site targeting primers (\textbf{Figure 3.8B}). While this was a significant attenuation, we expected to see an even more robust result, given that TM6SF2 should have been completely ablated. We suspected that the cut-site targeting primers may have undergone some non-specific binding, which could account for the remaining \textasciitilde20\% of TM6SF2 expression. Therefore, RT-PCR products were resolved on agarose gel to visualize all amplified sequences; we observed that the expected 137 bp TM6SF2 product appeared in the unedited iPSCs, as well as in fetal liver explant, but was entirely missing in the E5-17 clone, confirming the knockout (\textbf{Figure 3.8C}). Furthermore, two additional bands were observed in the E5-17 clone (indicated by purple asterisks), suggesting that the residual 20\% TM6SF2 expression was likely an artifact of non-specific primer binding.
Finally, to ensure no functional TM6SF2 transcript were present in the knockout model, the polypeptide sequence predicted to result from the edited TM6SF2 mRNA were examined. The 10 and 11 bp deletions in E5-17 were expected to produce short polypeptide fragments before encountering stop codons within the first quarter of the WT sequence, which were highlighted in red in Figure 3.8D. Although there is a possibility that translation could be reinitiated at methionine codons other than the initial start codon, additional stop codons would soon be encountered and translation would terminate again, resulting in short polypeptides that likely have no functional significance. We therefore concluded that the E5-17 iPSC line was genetically homogenous and expressed no functional TM6SF2 gene.
Figure 3.8 Screening for complete TM6SF2-KO in LG2-C6 iPSCs. A) Schematic depicting the strategy for the design of primers targeting the Cas9 cut site of TM6SF2. Created
using Biorender.com. B) RT-qPCR results comparing expression levels determined using primer pairs targeting a region downstream of the cut site vs primers flanking the cut site. C) RT-PCR products run on agarose gel resolve cDNA products of differing sizes. The expected size of the product when using primers targeting the TM6SF2 cut site was 137 bp. This band is highlighted in the orange box. (*) denotes bands likely resulting from non-specific primer binding. D) Predicted TM6SF2 polypeptide sequences with 10 and 11 nucleotide (nt) deletions. The full TM6SF2 peptide is shown and only the translated regions are highlighted in pink. A dash indicates the presence of a stop codon; the following amino acid is presented in red. Sequence prediction by Expasy (SIB Swiss Institute of Bioinformatics).

3.4 Phenotypic characterization of TM6SF2 KO HLCs

Once it was confirmed that TM6SF2 had been successfully disrupted using the CRISPR/Cas9 system, the resulting iPSCs were subjected to HLC differentiation (following the protocol described in Chapter 2) in order to continue the exploration of TM6SF2 function. From this point forward, for the sake of clarity, the E5-17 clone will be referred to as TM6SF2 KO or -/- and the unedited isogenic line from which it was created will be referred to as KK.

To characterize the phenotype resulting from the TM6SF2 KO, we first investigated the impact of fatty acid stimulation on lipid droplet accumulation in HLCs as a measure of steatosis. HLCs produced from the KK and -/- iPSC lines were treated with 0.4mM oleic acid (OA) starting on Day 0 of MH stage and continued for 6 days, with medium collection and refresh every 24 hours. Cell images were captured in brightfield on Day 6 to assess degree of lipid droplet accumulation; Bodipy stain was not used because the HLCs retained EGFP fluorescence (from the gRNA plasmid) on the FITC channel. Following microscopy, the HLCs were collected for RNA extraction and gene expression analysis.

RT-qPCR analysis confirmed that TM6SF2 gene expression was knocked down by >95% in the TM6SF2/- HLCs, compared to the KK HLCs (Figure 3.9A). Brightfield images taken on maturation stage day 6 show that HLCs accumulated significant lipid droplets when treated with OA (Figure 3.9B). However, there was no apparent difference in the lipid droplet accumulation
between the KK and -/- HLCs. Daily apoB100 secretion was modestly reduced in the TM6SF2-/- HLCs compared to KK, and this trend was consistent in the OA treated HLCs as well (Figure 3.9C).
Figure 3.9 Phenotypic characterization of HLCs derived from TM6SF2 KO iPSCs and isogenic (KK) controls. A) RT-qPCR analysis of HLC TM6SF2 expression using primers targeting exon 6 and primers targeting Cas9 cut site. B) Representative bright field images of isogenic TM6SF2 KK and -/- HLCs with or without 0.4mM OA treatment for 6 days. Yellow arrows point to large lipid droplets. C) Daily apoB100 secretion from HLCs treated with 0.4mM OA starting at Day 0 of maturation stage. Values were standardized to total cell protein. For A) and C) n=2 wells per condition.
Although the successful differentiation of the CRISPR edited iPSCs to functional HLCs was encouraging, we were concerned that perhaps a robust phenotype would not be evident when comparing KO HLCs to mutant (KK), rather than WT, HLCs, particularly since increased hepatic lipid accumulation is the most consistent result of disrupted TM6SF2 function reported in the literature. Therefore, we next endeavored to create an isogenic WT iPSC line from which to differentiate HLCs and better explore the effect of both the TM6SF2 KK variant and the complete knockout on hepatic lipid metabolism. These efforts are described in the following chapter.

3.5 Detailed Methods

3.5.1 Lentivirus Production

Third generation lentiviral system was used to produce all the lentiviruses in Chapter 4. The packaging and envelope plasmids, psPAX2 and pMD2.G, were gifts from Didier Trono (Addgene plasmids # 12260 and 12259). The transfer plasmids containing genes of interest varied in molecular cloning experiments and are described in more details below. All the plasmids were transformed into the competent *E. coli* strain *stbl3* (Thermo Fisher, C737303), which was streaked on LB agar (Millipore Sigma, L2897) plates, from which specific colonies were selected for overnight culture in LB broth (Millipore Sigma, L3522). Plasmids were purified from the bacterial culture using the NucleoBond Xtra Midiprep Kit (Takara Bio, 740410), following the manufacturer’s instruction, and their concentrations were measured on a spectrophotometer (DeNovix, DS-11+). Plasmids were stored in -20°C until needed.

HEK293T (ATCC, CRL-3216) was used to produce lentiviruses. Cells were cultured on surfaces pre-coated with 0.1% gelatin, and maintained in DMEM supplemented with 10% FBS, 25 μg/mL gentamicin, NEAA, β-mercaptoethanol, and sodium pyruvate. 24 hours prior to transfection, cells were passaged by incubating with 0.25% trypsin-EDTA for 5 minutes at 37°C.
A confluency of 70% at the time of transfection was desired. Lentiviral plasmids were mixed at a molar ratio of 3:2:1 (transfer-to-packaging-to-envelope plasmids) prior to forming complexes with CalFectin (SignaGen Laboratories, SL100478), following the manufacturer’s instruction. The medium containing transfection reagents was replaced with fresh medium 16 hours post-transfection, and medium containing viruses were collected 40 and 64 hours post-transfection. Viruses were concentrated using the Amicon Ultra-15 Centrifugal Filter Unit (Millipore Sigma, UFC9100) and the final concentrations were measured using the qPCR Lentivirus Titer Kit (ABM, LV900). Lentiviruses were stored in small aliquots in -80°C until transduction.

3.5.2 Generation of Doxycycline-inducible Cas9-expressing Cell Lines

pCW-Cas9 was a gift from Eric Lander and David Sabatini (Addgene plasmid # 50661). It is a lentiviral plasmid containing the humanized S. pyogenes Cas9 gene under a Tet ON promoter, which allows the transcription of Cas9 upon doxycycline (dox) exposure. Cas9 is further fused with 3X FLAG sequences at the N-terminal and with a nucleoplasmin nuclear localization signal (NLS) at the C-terminal. On the same plasmid, a cell selection marker, puromycin resistance (PuroR), is expressed under the constitutive promoter, hPGK, along with rtTA, which binds doxycycline before activating the Tet ON promoter. Lentiviruses containing dox-inducible Cas9 were produced and titrated as described above.

Doxycycline-inducible Cas9-expressing cells included HEK293T, HepG2, and iPSCs (LG2-C6). HEK293T and HepG2 were passaged as single cells by incubating with 0.25% trypsin-EDTA for 5 minutes at 37°C, followed by vigorous pipetting, cell counting, and seeding on surfaces pre-coated with 0.1% gelatin and 5 µg/cm^2 collagen I, respectively. On the next day, HEK293T and HepG2 were transduced with lentiviruses carrying the Cas9 gene, at a titer of 10 IU/cell. Briefly, lentiviruses were thawed and mixed with 10 µg/mL polybrene (Millipore Sigma,
TR-1003) in complete growth medium before adding to the cells. Medium containing polybrene alone served as the negative control in transduction experiments. Viruses were removed 24 hours post-transduction. HEK293T and HepG2 were treated with 1 µg/mL puromycin (Thermo Fisher, A1113803) 48 hours post-transduction to select for successfully transduced cells. Puromycin concentration was increased to 3 µg/mL for the next two days, after which 100% cell death was observed in the negative control.

iPSCs were detached by incubating with Accutase for 5 minutes at 37°C, followed by gentle pipetting to break the small clusters into single cells, which were seeded on surfaces pre-coated with Matrigel (0.5 million cells per well of 6-well plates). For iPSC mono-dispersed single-cell passaging, culture medium was supplemented with 10% CloneR (Stemcell Technologies, 05888), which was designed to improve iPSC single cell survival and robust growth. On the next day, iPSCs were transduced with Cas9 lentiviruses at 5, 10 and 20 IU/cell, pre-mixed with 8 µg/mL polybrene. Cells were treated with 1 µg/mL puromycin 48 hours post-transduction for 3 days to select for successfully transduced cells. All the cells in the negative control had died after 3 days.

3.5.3 Pre-screening of gRNA Sequences Targeting TM6SF2

Single guide RNA (gRNA) sequences for CRISPR/Cas9-mediated knockout of human TM6SF2 were obtained using the Synthego Knockout Guide Design web tool (https://design.synthego.com/#/). The top three candidates, denoted SA, SB and SC, were selected based on early exon targeting, high on-target score, and low chance of off-targeting (Table 4). The gRNAs were purchased from Synthego in chemically modified, stable RNA forms.
Table 3.1 gRNA sequences for CRISPR/Cas9-mediated knockout of human TM6SF2.

<table>
<thead>
<tr>
<th>gRNA Name</th>
<th>Sequence</th>
<th>Cut Site on Chr. 19</th>
<th>Targeting Exon</th>
<th>On Target Score</th>
<th>Off Targets</th>
</tr>
</thead>
<tbody>
<tr>
<td>SA</td>
<td>UCAUCAAUUGCCACCCACAGG</td>
<td>19,271,120</td>
<td>Exon 2</td>
<td>0.650</td>
<td>0,0,0,7,128</td>
</tr>
<tr>
<td>SB</td>
<td>CCCUCUCACAGCCACCCUG</td>
<td>19,271,121</td>
<td>Exon 2</td>
<td>0.423</td>
<td>0,0,3,49,406</td>
</tr>
<tr>
<td>SC</td>
<td>UCCUCUCACAGCCACCCUG</td>
<td>19,271,122</td>
<td>Exon 2</td>
<td>0.496</td>
<td>0,0,3,50,507</td>
</tr>
</tbody>
</table>

Due to its ease of transfection, HEK293T-Cas9-PuroR was used to test the knockout efficiency of RNA sequences. Methods of cell transfection were adopted from Wang et al.\textsuperscript{183}. Cells were seeded at 0.1 million cells per well in a 12-well plate to reach a confluency of 50% after overnight culture. 1 µg of gRNA was mixed with 1.5 µL Lipofectimine 3000 (Thermo Fisher, L3000015) in Opti-MEM Reduced Serum Medium (Thermo Fisher, 31985070) and incubated for 10 minutes at room temperature. Cells were inoculated with the nanoparticles for 3 hours in Opti-MEM medium at 37°C, followed by inoculation in complete culture medium containing 4 µg/mL doxycycline to induce Cas9 expression. Genomic DNA (gDNA) was collected from the cells 48 hours post-transfection using the DNA Miniprep Kit (Zymo Research Corporation, D4068), following the manufacturer’s instruction.

To quantify the cleavage efficiency of the three gRNA sequences, an enzyme mismatch cleavage method was used to detect mutations that had occurred at the double-stranded breaks induced by CRISPR/Cas9.\textsuperscript{184} A pair of PCR primers (forward: 5’-CAGTGATCCT TTCTGGGGAG-3’; reverse: 5’-CGAAGACTGCAGTGAGTGG-3’) was used to amplify approximately 1000 bp of the gDNA on which the CRISPR/Cas9 target site was offset from the center (300 bp from the forward primer and 700 bp from the reverse primer). 100 ng of gDNA was used as the template to amplify the 1000 bp fragment. A 50 µL reaction was prepared for each sample using with the Q5 High-Fidelity DNA polymerase (New England BioLabs, M0491L) and allowed to proceed on the MyCycler thermocycler (Biorad) at an annealing temperature of 65°C.
Gel electrophoresis was performed to confirm the PCR product, using an ethidium bromide (Millipore Sigma, E1510) pre-stained, 2% agarose gel (GoldBio, A-201-1000). The PCR product was subsequently used to perform the enzyme mismatch cleavage assay, using the T7 endonuclease I (T7E1; New England BioLabs, M0302L). Heteroduplexes were first formed in NEBuffer 2 (New England BioLabs, B7002S) by denaturing the DNA at 95°C for 5 minutes, followed by reannealing them via a temperature gradient (a decrease of 2°C/s from 95°C to 85°C and 0.1°C/s from 85°C to 25°C). 200 ng of heteroduplexes were then digested with 10 units of T7E1 at 37°C for 30 minutes. The digested product was resolved on a 2% agarose gel, on which the intact product remained at 1000 bp and the digested product were separated into a 700 bp fragment and a 300 bp fragment.

To estimate the cleavage efficiency, the intensity of all three bands (1000 bp, 700 bp, and 300 bp) in each lane were measured using the ImageJ software. Calculations were based on the published protocol by Ran et al. The percentage of cleaved product ($P$) was first calculated by:

$$P = \frac{b + c}{a + b + c}$$

where $a$ represents the intensity of the intact product, and $b$ and $c$ are the intensities of each digested band. Cleavage efficiency, or indel occurrence, can be estimated by the following formula, assuming duplex formation has a binomial probability distribution:

$$E(\%) = \left(1 - \sqrt{1 - P}\right) \times 100\%$$

3.5.4 Generation of Lentiviruses Carrying TM6SF2-KO targeting gRNA

The optimal gRNA sequence, SA, was selected for insertion into the empty pAW12-lentiguide-EGFP (Addgene plasmid # 104374; a gift from Richard Young) lentiviral backbone. 5 µg of the intact pAW12-lentiguide-EGFP plasmids were digested with the BsmBI enzyme (New England BioLabs, R0580), following the manufacturer’s instruction. The digested backbone was resolved
through electrophoresis on a 1% agarose gel, extracted and purified using the Gel Extraction Kit (Qiagen, 28704). 1 µg of the purified product was dephosphorylated with Antarctic Phosphatase (New England BioLabs, M0289S), following the manufacturer’s instruction.

The SA sequences were synthesized as two complementary DNA oligomers, with the addition of sticky ends that anneal to the backbone and BsmBI restriction sequences:

Oligo 1: 5’-CGTCTCNCACC<sup>G</sup>TCATCAATGCCACCCACAGG<sup>GTTTNGAGACG-3’</sup>,
Oligo 2: 5’-CGTCTCNAAAC<sup>C</sup>CCTGTGGGTGGCATTGATGA<sup>GGTGNGAGACG–3’</sup>.

5’-CGTCTC(N)<sub>1</sub>-3’ and 3’-GCAGAG(N)<sub>5</sub>-5’ are the double-stranded recognition site of BsmBI and 5’-CACC-3’ and 5’-AAAC-3’ represent the 5’ and 3’ overhangs after BsmBI digestion. The pair of oligos was phosphorylated using T4 Polynucleotide Kinase (PNK; New England BioLabs, M0201S) and T4 ligase reaction buffer (New England BioLabs, B0202S) at 37°C for 30 minutes, followed by annealing via a temperature gradient (95°C for 5 minutes and decrease by 0.1°C/s from 95°C to 25°C) to produce 10 µM of double-stranded oligos.

gRNA oligos were ligated to the dephosphorylated pAW12-lentiguide-EGFP backbone using T7 DNA ligase (New England BioLabs, M0318S). The oligos were first diluted to 50 nM (1:200) and 2 µL of the diluted oligos were mixed with 100 ng of the plasmid and T7 DNA ligase, followed by incubation at 37°C for 60 minutes. Ligated product was transformed into stbl3 competent cells and four viable colonies were selected for liquid culture and plasmid purification using a Miniprep Kit (Macherey-Nagel, 740588). gRNA sequence from these four bacterial clones were verified with Sanger sequencing (Eton Bioscience), for which the human U6 promoter forward primer (hU6-F; 5’-GAGGGCCTATTTCCCATGATT-3’) was used. After verification, one clone was selected to prepare large liquid culture and Midiprep. The new plasmid, pAW12-TM6.gRNA-EGFP, was used to generate TM6SF2-KO lentiviruses, as described earlier.
3.5.5 Generation of TM6SF2-KO HepG2 and iPSC Lines

TM6SF2-KO lentiviruses were transduced into HepG2-Cas9-PuroR cells at a titer of 10 IU/cell, following the method described earlier for Cas9 lentivirus infection. Cells were treated with 4 µg/mL doxycycline for 48 hours, followed by inoculation in fresh medium for another 48 hours to allow gene cleavage to complete. Their gDNA was collected to amplify the 1000 bp fragment containing the KO region, as described earlier. Besides using the T7E1 mismatch cleavage assay to determine the KO efficiency, the fragment was sequenced using the amplification forward primer (5'-CAGTGATCCT TTCTGGGGAG-3'). Chromatograms of edited samples were analyzed by comparing to that of the unedited control in sequence trace decomposition programs TIDE (http://shinyapps.datacurators.nl/tide/) and ICE (https://ice.synthego.com/#/).

TM6SF2-KO lentiviruses were transduced into iPSC-Cas9-PuroR cells at a concentration of 30 IU/cell, following the method described earlier for Cas9 lentivirus infection. Cells were treated with 4 µg/mL doxycycline for 48 hours, followed by inoculation in fresh medium for another 48 hours to allow gene cleavage to complete before their gDNA was collected. Approximately 75% of the cells were positive for EGFP through flow cytometry analysis but gene editing was not detected through either T7E1 mismatch cleavage assay or Sanger sequencing/trace decomposition. Therefore, two rounds of single iPSC colony screening were performed to enrich the percentage of KO cells and to identify clones that are homozygous for TM6SF2-KO. iPSCs were first detached and sorted via fluorescence-activated cell sorting (FACS; BD, FACSaria), which deposited 10 cells per well in two plates of Matrigel-coated 96-well plates. Cells were collected in mTeSR Plus medium supplemented with 10 µM Y-27632 and 50% conditioned medium (mTeSR Plus), which was collected from unedited LG2-C6 cells and filtered with 0.22
µm Steriflip units (Millipore Sigma, SCGP00525). Cell survival was poor and only 21 wells were expanded for KO screening. 1 out of the 21 colonies had a KO score of 15 via Sanger sequencing/trace decomposition and was selected for the second round of screening. Instead of repeating FACS, these iPSCs were passaged as single cells and seeded at low density on a 10 cm dish pre-coated with Matrigel. After the single cells grew into mature colonies, they were incubated with 4 mg/mL collagenase type 4 at 37°C for 20 minutes, followed by manual colony picking. The subclones from the second round of screening were expanded in 48 well plates and split for cryopreservation or analysis for TM6SF2-KO through Sanger sequencing/trace decomposition.
Chapter 4: A New Role for TM6SF2 in Lipid Metabolism

4.1 Overview of TM6SF2 function in cholesterol homeostasis

4.1.1 Cholesterol and Lipid Rafts

Cholesterol is essential for the survival and proliferation of every mammalian cell. It serves as a precursor for bile acids, steroid hormones, and vitamin D.\textsuperscript{186} However, its most important function is to maintain the structure of cell membranes and regulate their fluidity and thickness. Cholesterol is not evenly distributed between the cell’s various membranes, nor is it distributed homogeneously within each membrane. It is widely accepted that cholesterol is highly enriched in lipid rafts, which are rigid membrane microdomains that determine the distribution and activity of a myriad of proteins involved in membrane-bound transport, ion channel formation, receptor signaling, and cell-to-cell contacts.\textsuperscript{187}

Lipid rafts are unique areas of a lipid membrane comprised of cholesterol and sphingolipids, wherein the relatively small cholesterol molecule occupies the spaces between the saturated hydrocarbon chains of sphingolipids. These associations result in a tightly packed structure termed the liquid-ordered phase, in contrast to the less structured liquid-disordered phase. The ability for raft proteins to diffusion is limited to the size of the raft, which, in turn, is determined by the concentration of sphingolipids and cholesterol in the membrane.\textsuperscript{188} Lipid rafts in the plasma membrane (PM) are enriched with distinct classes of proteins, including glycosylphosphatidylinositol-anchored proteins (GPI-APs), cholesterol-linked proteins, and some transmembrane (TM) proteins.\textsuperscript{189,190} Maintenance of the liquid-ordered phase is essential for the intercalation of the membrane-spanning domains of TM proteins within the raft lipids; removal of cholesterol from rafts (e.g., by cyclodextrin treatment) causes the TM raft proteins to dissociate from the lipids.\textsuperscript{189}
While the existence of lipid rafts in the PM is well-established, increasing data demonstrate that raft-like domains form within intracellular membranes as well.\textsuperscript{189,191–195} For example, growing evidence suggests that GPI-anchored proteins must partition into a permissive lipid raft environment in the ER and trans-Golgi network (TGN) for oligomerization and efficient transport.\textsuperscript{189,196} Additionally, it has been proposed that lipid rafts may be necessary for the organization of transport cargo and the segregation of specific proteins in sorting organelles (Golgi and endosomes) to facilitate their trafficking to the PM.\textsuperscript{197}

Given the breadth of processes that cholesterol impacts, it is not surprising that dysregulation of cholesterol homeostasis is associated with many human disease states. Excess cholesterol can be toxic to cells and high circulating cholesterol levels are strongly associated with increased risk for atherosclerosis; alternatively, abnormally low cholesterol can result in several developmental disorders, most notably of which is Smith-Lemli-Opitz syndrome.\textsuperscript{186} Therefore, it is essential that, not only total cellular cholesterol be tightly regulated, but its distribution between and within membranes be as well. Mammalian cells synthesize their own cholesterol, receive cholesterol by uptake from lipoproteins, and efflux cholesterol out of the cell to circulating lipoproteins.\textsuperscript{187} Since cholesterol is cytotoxic and cannot be effectively removed by catabolism, the cell has evolved intricate regulatory networks to sense intracellular cholesterol levels and regulate these three pathways in order to keep cellular cholesterol levels precisely controlled.

4.1.2 Regulation of cholesterol biosynthesis

Cholesterol synthesis starts with the combining of two acetyl-CoA units in the cytoplasm, and involves the coordinated actions of more than 20 enzymes (Figure 4.1), most of which are localized to the ER membrane.\textsuperscript{156} Notable steps include the reduction of HMG-CoA to mevalonate by HMG-CoA reductase (HMGCR), which is the primary rate-limiting enzyme in cholesterol...
Mevalonate is converted to farnesyl pyrophosphate (FPP) by a series of enzymatic reactions; FPP can serve as the precursor to sterols as well as non-sterol isoprenoids, which include dolichol, HemeA, ubiquinone, and Geranylgeranyl-PP. Two FPPs are condensed to squalene by squalene synthase (also known as farnesyl-diphosphate farnesyltransferase-1, FDFT1) and squalene monooxygenase (also known as squalene epoxidase, or SQLE); the latter commits the substrate to the sterol production pathway. Most of the synthesized squalene is converted to lanosterol, which undergoes several additional enzymatic reactions to become cholesterol.
Figure 4.1 Summary of the cholesterol biosynthesis pathway from acetyl-CoA. The schematic highlights several enzymes that are essential for the de novo synthesis of cholesterol, as well as the branchpoint of the common mevalonate pathway committing FPP to the production of sterol intermediates or non-sterol isoprenoids. Once synthesized, cholesterol can be further converted to essential biomolecules by tissue-specific enzymes. Also noted is the isomerization step catalyzed by Emopamil Binding Protein (EBP), the enzyme which shares homology with TM6SF2. *denotes rate-limiting enzyme. Created using Biorender.com
The regulation of *de novo* cholesterol synthesis involves the synergistic functioning of a growing network of proteins that sense and respond to fluctuations in cellular cholesterol levels. Additionally, the activity of these proteins may be regulated by cholesterol itself or by various cholesterol intermediates or derivatives.

i. **Sterol Regulatory Element Binding Proteins (SREBPs)**

Among these proteins are the sterol regulatory element-binding proteins 1 & 2 (SREBPs), the master transcription factors (TFs) in lipid homeostasis. SREBPs contain multiple transmembrane spanning domains which are inserted into the ER membrane. In the ER, SREBPs complex with SCAP (SREBP cleavage-activating protein), a polytopic membrane protein that functions as an ER cholesterol sensor. When the ER cholesterol level is low ([Figure 4.2 left](#)), SCAP escorts the SREBPs into COPII vesicles, facilitating their export from the ER to the Golgi complex ([Figure 4.2a](#)). At the Golgi, SREBPs are proteolytically processed by two cleavage enzymes, membrane-bound transcription factor site-1 protease (S1P, also known as MBTPS1) and membrane-bound transcription factor site-2 protease (S2P, also known as MBTPS2) ([Figure 4.2b](#)). This processing generates an active transcription factor, which can enter the nucleus and bind to sterol responsive element (SRE) sequences in the promoters of target genes and induce their expression ([Figure 4.2c](#)). SREBP1 specifically promotes the transcription of genes involved in fatty acid synthesis (ACC, FASN, SCD1, etc) and SREBP2 specifically promotes transcription of genes involved in cholesterol synthesis (HMGCR, FDFT1, SQLE, etc) and uptake (LDLR).

Conversely, when cholesterol in the ER membrane is abundant ([Figure 4.2, right](#)), the binding of cholesterol to SCAP induces/triggers conformational changes that promote SCAP to interact with the integral ER membrane protein Insulin Induced Gene 1 (Insig 1) ([Figure 4.2d](#)), which in turn blocks SCAP from being recruited to COPII vesicles ([Figure 4.2e](#)); thus, SCAP and
unprocessed SREBPs remain in the ER. Two other proteins, Erlins -1 and -2, have been shown to play an important role in this cholesterol-dependent ER retention of SCAP/SREBP under cholesterol-sufficient conditions. While their functions are not entirely clear, Erlins are hypothesized to act as scaffolds to support the Insig/SCAP/SREBP complex, or to help in the formation of the necessary ER raft subdomain required for the stability of the Insig/SCAP/SREBP complex.

Figure 4.2 Schematic of SREBP2 regulation by ER cholesterol. SREBP2 is processed to its cleaved (active) form under sterol depleted conditions (Left). Upon cholesterol repletion to the ER, the of inactive SREBP2 is retained in the ER (Right). (FC, free cholesterol). Created with BioRender.com

ii. HMG-CoA Reductase (HMGCR)
HMGCR activity is regulated at multiple points including transcription, translation, degradation, and phosphorylation; therefore, the enzyme’s activity may not always reflect its transcript or protein levels. Alternative splicing of the HMGCR gene to a shorter transcript produces an inactive protein; this splicing has been shown to be regulated by sterols. Insig1 and the E3 ligase GP78 are the major post-translational regulators of HMGCR protein levels and this regulation is dependent on ER sterol levels. When ER sterols are replete, Insig1 binds to HMGCR and recruits GP78 and the ATPase Valosin-containing protein (VCP) to promote the ubiquitination and degradation of HMGCR (Figure 4.2f). Under these conditions, binding of Insig1 to SCAP protects Insig1 from being ubiquitinated by GP78. On the other hand, when the cholesterol level drops, SCAP dissociates from Insig1, Insig1 is ubiquitinated by GP78 (Figure 4.2g), and HMGCR is stabilized, which allows for cholesterol synthesis to occur.

iii. Squalene Monooxygenase (SQLE)

The major substrate of Squalene Monooxygenase (SQLE) is squalene, a 30-carbon isoprenoid formed from the condensation of two farnesyl pyrophosphate (FPP) units by squalene synthase/FDFT1. SQLE converts squalene to 2,3(S)-monooxidosqualene, the precursor to lanosterol, which is the first sterol intermediate formed. SQLE is the second rate-limiting enzyme in the cholesterol synthesis pathway and its mRNA expression is subject to SREBP2-mediated, sterol-sensitive, transcriptional regulation. Since SQLE acts only within the sterol branch of the cholesterol synthesis pathway, the regulation of its activity impacts cholesterol synthesis downstream of HMGCR and independently of the common mevalonate pathway.

4.1.3 Regulation of cholesterol uptake, efflux, and storage

Dietary and biliary cholesterol can be absorbed by Niemann–Pick type C1 (NPC1)-like 1 (NPC1L1) protein on the apical surface of enterocytes in the intestine. These sources of cholesterol
are secreted as components of chylomicrons, which, after they deliver dietary triglycerides to adipose tissue and muscle via lipoprotein lipase (LPL)-mediated lipolysis (and become chylomicron remnants), are taken up by the liver. The liver — where the majority of cholesterol biosynthesis occurs — delivers both endogenously synthesized and exogenously absorbed cholesterol to the peripheral circulation as a component of very low density lipoproteins (VLDL). VLDL loses its triglyceride to adipose tissue and muscle via LPL-mediated lipolysis, generating, in the end, the cholesterol enriched low-density lipoproteins (LDL), which can be taken up by peripheral cells or returned to the liver.

Although most mammalian cells have the ability to synthesize endogenous cholesterol, this task is typically assigned to the highly experienced cholesterol-handling cells of the liver and small intestine. Peripheral tissues acquire the majority of their cholesterol (~80%) by receptor-mediated endocytosis of LDL particles. The low density lipoprotein receptor (LDLR), a canonical SREBP2 transcriptional target, mediates the uptake of exogenous cholesterol from LDL when cellular cholesterol levels are low. Following LDL internalization, LDL-derived cholesterol first enters the endo-/lysosomal system, where the esterified cholesterol is hydrolyzed, and the resultant free cholesterol is distributed to other cellular compartments by mechanisms that have not been entirely elucidated, but are known to involve the NPC1 and NPC2 proteins. NPC1 is a cholesterol-binding integral membrane protein that is essential for efficient export of cholesterol from the lysosome. When cholesterol levels are replete, PCSK9 is secreted out of the cell, where it binds to, and facilitates the degradation of, the LDLR. Additional regulation comes from Idol, an E3 ubiquitin ligase transcriptionally regulated by the nuclear liver X receptor (LXR), which mediates the degradation of LDLR via PCSK9.
LXR also upregulates the transcription of the genes encoding the ATP binding cassette (ABC) transporters ABCA1 and ABCG1.\textsuperscript{156} When cellular cholesterol levels are high, these transporters export excess cholesterol out of the cell to apoA1-phospholipid complexes produced by the liver, to generate high-density lipoproteins (HDLs). HDLs are taken up by the liver and intestine via the SRB1 scavenger receptor; HDL also delivers cholesterol to steroidogenic organs where the cholesterol serves as the precursor to steroid hormones.\textsuperscript{156}

Alternatively, excess cholesterol can also be esterified with a fatty acyl-CoA by the ER-transmembrane enzyme Acyl-Coenzyme A: Cholesterol Acyltransferase (ACAT) to form neutral cholesteryl ester (CE). CE, along with other lipids, can be stored, within cytosolic lipid droplets (cLDs) for later use, or transferred onto lipoproteins and secreted out of the cell. Two mammalian ACAT isoforms have been identified; ACAT1 is the predominant form in non-hepatic tissues, while ACAT2 appears to contribute most of the esterification activity in hepatocytes.\textsuperscript{207} Like the cholesterol synthesis pathway, cholesterol esterification is also subject to cellular cholesterol levels; ACAT has been shown to be allosterically activated by cholesterol\textsuperscript{208} and to exhibit increased activity when localized to cholesterol-rich raft-like domains of the ER.\textsuperscript{209,210}

\textit{4.1.4 Documented associations between TM6SF2 function and cholesterol metabolism}

Many published findings support the potential role of TM6SF2 in cholesterol homeostasis, although some results have been inconsistent. Smagris, \textit{et al} showed decreased HMGCR, HMGCS1, Insig1/2 gene expression in female TM6SF2 KO mice, as well as decreased expression of genes involved in DNL (FASN, SCD1, ELOVL6).\textsuperscript{45} The reduction in cholesterol genes was recapitulated by Fan \textit{et al} in liver-specific TM6SF2 KO mice.\textsuperscript{85} In another study by Newbury \textit{et al}, female mice with liver-specific deletion of TM6SF2 also exhibited a trend towards reduced SREBP2, HMGCS1, and HMGCR, as well as decreased expression of FASN, SCD1, ELOVL5;
however, no such differences were observed in the male mice.\textsuperscript{94} Conversely, in mice overexpressing TM6SF2 in the liver, Fan \textit{et al.} found increased expression of DHCR7, which catalyzes the last step in cholesterol synthesis.\textsuperscript{85}

Human models of disrupted TM6SF2 expression also displayed a cholesterol phenotype. RNAseq analysis in human WT and mutant liver biopsies showed dysregulation of genes involved in cholesterol metabolism, bile acid synthesis/transport, and DNL.\textsuperscript{77} In RNA-seq analysis performed on both liver tissue and primary hepatocyte spheroids by Prill \textit{et al}, it was found that cholesterol synthesis genes were upregulated in mutant samples, compared to WT.\textsuperscript{99} When TM6SF2 was overexpressed in the human hepatoma HepB3 cell line, increased cholesterol biosynthesis was observed, whereas overexpression of the E167K transgene attenuated this effect.\textsuperscript{85} Taken together, these published findings clearly pointing to a connection between TM6SF2 and cholesterol metabolism, although species and sex differences, as well as the inconsistent directionality of these findings have generated more questions than answers as to the exact role of TM6SF2.

An additional clue to TM6SF2 function comes from the peptide’s structure and its sequence homology to Emopamil Binding Protein (EBP), an enzyme in the cholesterol biosynthetic pathway with sterol isomerase activity. Sanchez-Pulido and Ponting found a novel 4 transmembrane region occurring in Emopamil Binding Protein (EBP), which they named EXPERA (EXPanded EBP superfamily), occurring twice in TM6SF1 and TM6SF2. EBP is an ER-resident protein and enzyme in the cholesterol biosynthetic pathway possessing sterol isomerase activity. The EXPERA domain was also identified in the peptide sequence of TMEM97\textsuperscript{84}, a protein recently identified as the Sigma 2 Receptor\textsuperscript{211}. The Sigma Receptors (1&2) are intracellular transmembrane chaperone proteins highly expressed in the central nervous system (CNS) that appear to play a role in CNS
pathologies, many of which are also associated with dysfunctional cholesterol metabolism, such as Parkinson’s and Alzheimer’s disease (AD). These receptors bind numerous drugs, which have been shown to alter cell proliferation/cell death pathways, and are highly expressed in many cancer cell types.

The Sigma 1 Receptor (σ1R) protein does not share homology with any known mammalian protein, but does share striking amino acid homology with the yeast homolog of EBP sterol isomerase ERG2 and has been shown to competitively bind steroids and cholesterol at its putative drug-binding site, as well as possess high affinity to inhibitors of mammalian cholesterol synthetic enzymes. The σ1R is highly enriched at mitochondria associated ER membranes (MAMs), and has been implicated in mediating ER lipid transport and the remodeling of cholesterol-enriched microdomains in the ER. Recently, Zhemkov et al identified a novel cholesterol-binding site in the σ1R sequence and demonstrated that the clustering of σ1R in lipid rafts is cholesterol-dependent.

Less is understood about the function of the Sigma 2 Receptor (σ2R), also called TMEM97, but its connection to cholesterol metabolism has been known for decades. The σ2R localizes to the ER membrane, but under sterol-depleted conditions, it becomes enriched in the endo-lysosomal membrane; there, it interacts with the cholesterol transporter NPC1 to regulate cholesterol trafficking out of these organelles. It has also been shown to be transcriptionally regulated by SREBP2 via the SRE sequence located upstream of the human TMEM97 gene. Recently, it was proposed that σ2R retains the inactive NPC1 in the ER when sterols are high, and chaperones it to the lysosome when ER sterols dip below some threshold, thus allowing NPC1 to traffic endocytosed cholesterol out of the lysosome and replenish depleted membranes.
Li and colleagues provided a recent link between TM6SF2 and cholesterol metabolism with their investigation into TM6SF2 binding partners. Firstly, they found that TM6SF2, through two of its ER luminal domains, directly binds to apoB; this interaction promotes the stability of both proteins (Figure 4.3). This is the first evidence to demonstrate that TM6SF2 physically associates with apoB. Secondly, they discovered that TM6SF2 shares the same mutual stabilizing relationship with ERLIN1 and ERLIN2, independent of its interaction with apoB.

![Figure 4.3 Schematic of proposed TM6SF2 transmembrane topology in the ER and interactions with ERLIN1/2 and apoB. Created with Biorender.com.](image)

Finally, Li et al performed tandem affinity purification and mass spectrometry analysis to determine the most abundant proteins associated with TM6SF2. Results from this study, summarized in Table 4.1, demonstrated that TM6SF2 interacts with enzymes essential for cholesterol biosynthesis, proteins involved in the ER to Golgi trafficking of specialized VLDL transport vesicles (VTVs), and proteins previously shown to associate with cholesterol-rich raft-like domains. Interestingly, Sec22B, VAPB, and SACM1L, three of the proteins reported by Li and colleagues, localize to ER-Golgi contact sites and are necessary for the cholesterol transfer
protein Oxysterol Binding Protein (OSBP)-mediated exchange of cholesterol from the ER for PI_{4}P from the Golgi; this exchange is essential for vesicle trafficking between the organelles.\textsuperscript{196,201,221–223} Most strikingly, recent findings by Venditti, \textit{et al} demonstrated that depletion of the SACM1L adapter protein FAPP1 resulted in an increase in the secretion of selected cargos, including apoB100.\textsuperscript{224}

\textbf{Table 4.1 Proposed TM6SF2 interacting partners, as determined by tandem affinity purification. Based on data from Li, \textit{et al}.}\textsuperscript{97}

<table>
<thead>
<tr>
<th>Gene Symbol</th>
<th>Name</th>
<th>*GO Term: ER Membrane</th>
<th>*UP Keyword: Cholesterol Metabolism</th>
<th>Lipid-Raft Associated</th>
<th>VLDL Transport Vesicle Cargo</th>
</tr>
</thead>
<tbody>
<tr>
<td>APOB</td>
<td>apolipoprotein B</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
</tr>
<tr>
<td>FDFT1</td>
<td>farnesyl-diphosphate farnesyltransferase 1</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SQLE</td>
<td>squalene epoxidase</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DHCR24</td>
<td>24-dehydrocholesterol reductase</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ERLN1</td>
<td>ER lipid raft associated 1</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ERLN2</td>
<td>ER lipid raft associated 2</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
</tr>
<tr>
<td>PIG5</td>
<td>phosphatidylinositol glycan anchor biosynthesis class S</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PIGX</td>
<td>phosphatidylinositol glycan anchor biosynthesis class X</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
</tr>
<tr>
<td>HMOX1</td>
<td>heme oxygenase 1</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
</tr>
<tr>
<td>VAPB</td>
<td>VAMP associated protein B and C</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>SEC22B</td>
<td>SEC22 homolog B, vesicle trafficking protein</td>
<td>x</td>
<td></td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>SACM1L</td>
<td>SAC1 suppressor of actin mutations 1-like</td>
<td>x</td>
<td></td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>FMO1</td>
<td>flavin containing monoxygenase 1</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*based on DAVID pathway analysis

Given the abundance of evidence to support a role for TM6SF2 in cholesterol homeostasis, and since we found detectable expression of TM6SF2 in iPSCs, which are a highly proliferative cell-type and therefore have high utilization rates of cholesterol, we chose to employ this model to further our investigation of TM6SF2 function. This strategy of characterizing a cholesterol-related phenotype in genetically/CRISSPR-edited undifferentiated iPSCs has been successfully demonstrated previously.\textsuperscript{175,225}
4.2 RNA-sequencing Analysis of TM6SF2-KO in iPSCs

In order to understand the global effect of TM6SF-KO on the iPSC transcriptome, we performed unbiased RNA-sequencing on mutant iPSCs (KK) and the isogenic CRISPR-disrupted line (-/-). Out of 35,838 genes sequenced, a total of 4,505 genes were identified as significantly differentially expressed in the -/- iPSCs, as compared to the KK line, after running the DESeq2 pipeline with the default Benjamini-Hochberg (BH) P-value of 0.1. These dysregulated genes were separated into an upregulated gene set (2,094) and a downregulated gene set (2,411). Enrichment analysis was first performed in Enrichr, an enrichment analysis web-based software tool for non-ranked gene lists based on Fisher’s exact test. Gene Ontology (GO) analysis reduces complexity in large gene sets and provides a first look at the biological processes in which these dysregulated genes are involved. The majority of GO biological processes (Figure 4.4A) identified in the upregulated genes related to lipid metabolism, specifically cholesterol biosynthesis. RNA processing and viral gene expression were identified as processes in the downregulated gene set, possibly indicative of the effect of the additional sgRNA lentivirus delivered in the KO cells. The molecular functions (Figure 4.4B) identified in both dysregulated gene sets were mostly associated with DNA maintenance, which could be attributed to the unwinding and cleavage activities of CRISPR/Cas9 in the KO cells. Lastly, Human metabolic pathway analysis was conducted through the HumanCyc database in Enrichr (Figure 4.4C). In the upregulated gene set, cholesterol biosynthetic genes were significantly enriched along with genes for the synthesis of its precursors mevalonate, geranyl diphosphate, and zymosterol. Carbohydrate metabolism pathways—glycolysis and gluconeogenesis—were enriched in the downregulated gene set.
Figure 4.4 Enrichment analysis of dysregulated genes in TM6SF2 KK vs. -/- iPSCs performed in Enrichr. Gene ontology analysis of the A) biological process, B) molecular function, and C) human metabolic pathway enrichment identified using the HumanCyc database. The upregulated gene set is represented with a magenta gradient and the downregulated gene set is represented with a teal gradient.

A network of the 2,094 upregulated genes was constructed in Cytoscape and the MCODE application was applied to find clusters, or highly interconnected regions within the network. A total of 106 genes were identified in a large cluster, which could be further broken down into 5 subclusters (data not shown). Pathway analysis was performed within each subcluster to reveal their various functions. Subcluster 1 was identified as regulating lipid metabolism, specifically cholesterol biosynthesis, further supporting that TM6SF2 knockout effects cholesterol homeostasis in iPSCs (Figure 4.5A). To confirm the significance of these findings, targeted RT-qPCR for key cholesterol synthesis genes was performed in the same KK and -/- iPSC lines, as
well as in sibling-matched iPSCs expressing the WT TM6SF2 gene (EE*). Similar to the phenotype observed in the RNA-seq experiment, we found increased levels of cholesterol synthesis genes in the -/- iPSCs compared to both the EE* and KK iPSCs (Figure 4.5B). However, when we measured total cellular cholesterol (TC), we found no significant difference between the three iPSC lines (Figure 4.5C). This was not entirely unexpected, as excess unesterified, or “free” cholesterol (FC), is cytotoxic due to its labile hydroxyl group; as such, FC levels are tightly regulated in the cell by several pathways, including esterification into CE, a neutral lipid which can be safely stored in cytosolic lipid droplets.\textsuperscript{186} We hypothesized that esterification of any excess synthesized cholesterol in the -/- iPSCs could explain the lack of additional cholesterol in these cells, despite the increased expression of cholesterol synthetic genes.

Figure 4.5 Gene clustering determined in the upregulated gene set through building a large network in Cytoscape followed by analysis with MCODE. A) Subcluster 1 consisted of 21 genes involved in lipid metabolism. B) RT-qPCR expression levels of select cholesterol
synthesis genes. Data are presented as mean +/- SEM, One-way ANOVA, *p<0.05, **p<0.01, n=2 per genotype C) Total cell cholesterol, determined by colorimetric assay and standardized to cell protein. Data are presented as mean +/- SEM, One-way ANOVA, *p<0.05, **p<0.01, n=3 per genotype

4.3 Effect of TM6SF2-KO on the lipidome in iPSCs

We next performed lipidomics analysis on EE*, KK, and -/- iPSCs, to determine the effect of TM6SF2 on cholesterol esterification, as well as to explore phenotypic changes in other lipid classes. Confirming our previous results, we found that TC levels determined by Mass Spec were not significantly different between the iPSC lines (Figure 4.6A); nor were FC levels different. Contrary to our hypothesis, we observed an incremental decrease in cell CE from EE* to KK, and from KK to -/- (Figure 4.6B). In fact, out of the 34 lipid classes detected, cholesterol ester (CE) displayed the most significant reduction in the mutant and KO lines, as compared to EE* (Figure 4.6C). Also unexpectedly, and in contrast to human and mouse data, we observed a decrease in cell TG in the -/- iPSCs (Figure 4.6D). Thus, contrary most hepatic models, TM6SF2 -/- iPSCs exhibit decreased neutral lipid accumulation at steady state, and no changes in FC content despite displaying increased expression of cholesterol synthesis genes.
Since there does not appear to be an increase in cholesterol esterification as a means of neutralizing excess FC, we next sought to determine if TM6SF2 -/- iPSCs do in fact exhibit increased *de novo* cholesterol synthesis as a result of elevated cholesterol synthesis gene
expression. To do so, sub-confluent EE*, KK, and -/- iPSCs were lipid starved in basal medium for 2 hours, followed by a 2-hour pulse with $^{14}$C-acetate, the ubiquitous precursor of lipid synthesis. Extracted lipids were separated by TLC and spots for free cholesterol (FC), cholesterol ester (CE), free fatty acid (FFA), and triglyceride (TG) were scraped for measurement of $^{14}$C-incorporation by scintillation counting. We found that de novo FC synthesis was not significantly different between any of the groups (Figure 4.7A), whereas $^{14}$C incorporation into CE was reduced in the KK and -/- iPSCs, compared to EE* (Figure 4.7 B,C). We found no differences in TG synthesis between the groups (Figure 4.7D). These data indicate that the aberrant increase in expression of cholesterol synthesis genes observed in the TM6SF2 -/- iPSCs did not result in greater cholesterol synthesis.

![Figure 4.7 Impact of TM6SF2 on iPSC de novo lipogenesis. $^{14}$C-acetate incorporation into A) Free Cholesterol B&C) Cholesterol Ester D) Triglyceride E) Free Fatty Acids. Data are presented as mean % species DPM standardized to total DPM counts +/- SEM, One-way ANOVA, *p<0.05, **p<0.01, ***p<0.001, n=3 per genotype.](image)

The possibility remains that, although the transcription of these genes is elevated in the -/- iPSCs, their protein products are not translated or are unstable, so the expected increase in cholesterol synthesis does not occur. However, we observed ~50% increase in HMGCR protein expression in -/- vs KK iPSCs and a trend towards increased SREBP2 cleavage (Figure 4.8), suggesting that this is unlikely to be the case. In addition, we saw a ~50% decrease in INSIG1 protein expression in -/- vs KK iPSCs (Figure 4.8); since INSIG1 is involved in retaining SREBP2
the ER and thus preventing its activation, reduced INSIG1 protein levels would also point to an increase, rather than an attenuation of SREBP2 transcriptional activity.\textsuperscript{228} Despite these observations, HMGCR activity may not always reflect transcript or protein level;\textsuperscript{202} therefore, direct measurement of the enzyme’s activity may help to elucidate the true impact of TM6SF2 knockout in iPSCs.

Another reasonable explanation for these discordant observations is a compensatory synthesis of alternative sterols, or non-sterol isoprenoids, such as Heme, CoQ10, and Vitamin K, which derive from the common mevalonate pathway (Figure 4.1).\textsuperscript{229} In fact, TM6SF2 was recently shown to physically interact with the enzyme FDFT1, which catalyzes the step in the mevalonate pathway committing FPP to the synthesis of sterols instead of isoprenoids. Given that knockdown of TM6SF2 was shown to destabilize other interacting partners (i.e. apoB, ERLINs),\textsuperscript{97} it stands to reason that lack of TM6SF2 could also result in the reduced stability of FDFT1, thereby causing a redirecting of the mevalonate pathway products into the synthesis of isoprenoids instead. If this occurred, increased activation of SREBP2, with its downstream effects on gene expression and protein synthesis, would result in ‘restoration’ of normal FC levels. Further work will need to be conducted to determine the effect of TM6SF2 on the stability of FDFT1 and to measure these metabolites and determine if this is indeed the case.
Figure 4.8 Protein expression of cholesterol synthesis regulators. Band intensities from western blots were determined by densitometry and standardized to β-actin.

4.4 TM6SF2 and Cholesterol-Rich ER Lipid Rafts

Given that there is no depletion of FC in TM6SF2 -/- iPSCs, the question remains: why is the SREBP2 transcriptional program aberrantly upregulated? Sandhu and colleagues recently reported observing a similar phenotype upon disrupting ER-anchored Aster proteins, which they came to discover bound cholesterol and trafficked it from the plasma membrane (PM) to the ER. They suggested that, under conditions in which the cholesterol sensing components of the SREBP2-Scap-Insig machinery in the ER have impaired access to certain regulatory cholesterol pools, SREBP2 processing can become aberrantly activated. Similarly, we hypothesized that loss of function of TM6SF2 causes a disruption in cholesterol localization to the ER domains where SREBP2-Scap-Insig reside. This proposition is supported by the consistent decrease in CE concentration and synthesis observed in the TM6SF2 -/- iPSCs (Figure 4.6B, 4.7B), which was also a phenotype demonstrated in Aster KO models. While it is possible that the reduced CE in our model is due to insufficient substrate (the cholesterol backbone or the fatty acid moiety), this is unlikely since neither de novo synthesis of FC nor FFAs is impaired in KK or -/- iPSCs (Figure
4.7 A,E). Instead, a more likely explanation is that reduced TM6SF2 expression alters the activity of Acyl-Coenzyme A: Cholesterol Acyltransferase 1 (ACAT1), the enzyme that catalyzes the esterification of the fatty acyl-CoA to cholesterol to form CE. ACAT1 is an ER-transmembrane protein which has been shown to be allosterically activated by cholesterol\(^{208}\) and to exhibit increased activity when localized to cholesterol-rich raft-like domains of the ER, often referred to as Mitochondria-Associated ER Membranes (MAM)\(^{209,210}\). Therefore, it would be expected that a defect in cholesterol localization in the ER due to TM6SF2 dysfunction could subsequently alter the activity of enzymes dependent upon the proper formation of those specific domains. (TM6SF2 direct effect on ACAT1 activity will be addressed later in the chapter).

Phosphatidylserine Synthase 1 (PSS1) is another enzyme whose optimal activity is dependent upon localization to cholesterol-enriched MAM.\(^{209}\) PSS1 catalyzes the replacement of the polar head group of phosphatidylcholine (PC) with L-serine to form phosphatidylserine (PS) in the ER. PS is then trafficked into the mitochondria where it can be converted to phosphatidylethanolamine (PE) by the inner mitochondrial enzyme Phosphatidylserine Decarboxylase (PSD). PSS1 activity can be tracked as a surrogate for MAM function by pulsing cells with \(^3\)H-serine and measuring its respective incorporation into PS and PE over time (Figure 4.9A).\(^{210}\) When we performed this assay in our TM6SF2 EE*, KK, and +/- iPSCs, we did not see any difference in \(^3\)H incorporation into PE at the 2-hour time point, indicating that the enzymatic decarboxylation of PS to PE in the mitochondria is intact in the +/- iPSCs, and therefore not effected by TM6SF2. However, the nearly 50% reduction of \(^3\)H incorporation into PS in the TM6SF2 +/- iPSCs suggests that PSS1-mediated synthesis of PS is defective, potentially due to improper MAM formation (Figure 4.9B). The reduction of \(^3\)H incorporation into PE seen at the 3- and 4-hour time points indicates that PS conversion into PE becomes delayed, further suggesting that movement
of PS across the MAM and into the mitochondria is impacted in TM6SF2 -/- iPSCs. Notably, total cell PS and PE concentrations determined through lipidomics do not show reduced levels in the TM6SF2 -/- iPSCs; in fact, PE levels were significantly higher in the KK and -/- iPSCs, compared to EE* (Figure 4.9C). However, the cell has multiple routes for synthesizing phospholipids, including the Kennedy (CDP) pathway and the Land’s (salvage) pathway. The RNAseq analysis showed a significant increase in PCYT2 and EPT (PE synthesis) as well as MBOAT1 (PS synthesis) gene expression in the -/- iPSCs, suggesting that alternative pathways may be upregulated in order to compensate for the defective MAM-dependent synthesis (Figure 4.9D).

Taken together, these data begin to suggest a role for TM6SF2 in modulating the activity of enzymes whose activity is dependent upon proper localization to cholesterol-enriched domains of the ER membrane.

Figure 4.9 Effect of TM6SF2 KO on MAM activity in iPSCs. A) Schematic depicting PS synthesis at MAM domain by PSS1/2 and its movement to the mitochondria for decarboxylation to PE by PSID. The measurement of radiolabeled Ser (serine) incorporation into PS and PE serves as a proxy assay for MAM activity. Created with Biorender.com B) ³H-
4.6 Generation of isogenic TM6SF2 WT, mutant, and knockout iPSC lines

While the information learned from the studies above was valuable in establishing an important role for TM6SF2 in iPSC cholesterol metabolism, the comparison of the mutant to the knockout and both of those lines to a non-isogenic WT is not the ideal model, as evidenced by the mild phenotype observed in both the iPSCs and differentiated HLCs (Chapter 3). Therefore, we next sought to engineer isogenic WT, mutant, and KO iPSCs which could be differentiated to HLCs in order to strengthen the phenotypic differences and thus better characterize the function of TM6SF2 in a hepatic context (Figure 4.10A). To do so, lentiviral plasmids containing either the TM6SF2 WT (encoding the native protein with glutamate at codon 167, denoted EE_OE) or the mutant (encoding the rs58542926 variant with lysine at codon 167, denoted KK_OE) transgene were generated, followed by confirmation via Sanger sequencing. TM6SF2 -/- iPSCs (described in the previous chapter) were then transduced with lentiviruses (LV) carrying either the EE or KK transgene at a titer of 2, 5, and 15 IU/cell, followed by antibiotic treatment with G418 sulfate to select for successfully transduced cells. A third line was created by transducing TM6SF2 -/- iPSCs with a LV carrying an empty backbone (i.e. no TM6SF2 transgene) at a titer of 15 IU/cell; these cells received only the antibiotic resistance gene (NeoR) and underwent the same course of antibiotic treatment as the experimental lines. This line (denoted KO_BB, Knockout_BackBone) served as the isogenic TM6SF2 knockout line for all further transgenic iPSC studies.

Measurement of transgene copy number (determined by commercial assay) confirmed successful incorporation of the lentivirally delivered transgenes (Figure 4.10B). RT-PCR analysis showed that TM6SF2 EE and KK expression was induced 30-50 times the level of endogenous serine incorporation into PS and PE over time in iPSCs. C) iPSC PS and PE concentration measured by lipidomics. Values presented as mol% of total lipid +/- SEM, One-way ANOVA, *p<0.05, **p<0.01, ***p<0.001, n=3 per genotype. D) Gene expression levels of PS and PE synthesis genes, as determined by RNA-sequencing.
TM6SF2 expressed in the parent iPSC line (data not shown). However, as it has been illustrated that TM6SF2 overexpression can induce a similar phenotype to that of the knockout, we repeated the iPSC lentiviral transduction (in TM6SF2 -/- iPSCs not previously transduced) using a lower range of titers (0.1-0.5 IU/cell), as well as a gentler antibiotic course, so as not to unintentionally select for only the iPSCs possessing the highest transgene number. Subsequent transduction experiments yielded iPSCs with TM6SF2 gene expression closer to endogenous levels (Figure 4.10C).

Figure 4.10 Transgene expression of TM6SF2 WT (EE) and Mutant (KK) allele in isogenic iPSCs A) Schematic illustrating the lentiviral delivery of TM6SF2 transgene to TM6SF2 KO iPSCs and antibiotic selection. B) Expression of TM6SF2 transgene copy number in transduced iPSCs at 3 viral titers. Values standardized to total genomic DNA. C) Expression of TM6SF2 mRNA levels in transgenic iPSCs at 3 viral titers. Values expressed as fold change vs levels of endogenous KK mRNA. n=3 per group.
Unfortunately, when we attempted to differentiate the three isogenic iPSC lines to HLCs, we found that, while the cells underwent robust definitive endodermal (DE) induction, they rapidly deteriorated during the hepatic endodermal (HE) stage and were no longer viable by Day 8 of differentiation (data not shown). We attempted to modify our differentiation protocol to maximize the potential for successful differentiation, for example, by supplementing media with antibiotic and antifungal agents to reduce possible contamination, adjusting the cell density during the passaging of DE cells before HE induction, and refraining from passaging DE cells entirely to ameliorate any added stress this step may have caused. Despite these efforts, we could not successfully generate HLCs from the TM6SF2-overexpressing iPSCs or KO controls. Given that the KO iPSCs had previously differentiated after CRISPR/Cas9 editing, but before transduction with the NeoR-carrying lentivirus and subsequent antibiotic selection, we believe that the process of introducing these additional agents somehow rendered the iPSCs incapable of differentiating towards the hepatic lineage.

Due to these setbacks, we decided to continue our investigation of the role of TM6SF2 in iPSC cholesterol metabolism in our newly generated isogenic model. First, to determine if the overexpressing iPSCs recapitulates the phenotype of altered expression of cholesterol synthesis genes previously observed from the RNA-sequencing experiment, we measured the gene expression of \textit{HMGCR} and \textit{SQLE}, the two rate limiting enzymes in the cholesterol synthesis pathway.\textsuperscript{231} Interestingly, we found that \textit{HMGCR} expression was lower in the OE\_KK iPSCs than in the KO\_BB, which mirrored the phenotype observed in the RNA-seq of the endogenous KK iPSCs. However, this relationship was reversed when \textit{TM6SF2} transgene expression was more highly induced (ie \textit{HMGCR} expression was higher in OE\_KK than in the KO\_BB) (\textbf{Figure 4.11A}). Correlation analysis of the combined experiments revealed a very strong linear relationship
between *TM6SF2* expression and *HMGCR/SQLE* expression, suggesting that the expression levels of these cholesterol synthesis genes increase as a function of *TM6SF2* expression. Additionally, comparison of the slopes between the EE and KK iPSC regression lines demonstrated that this linear relationship was independent of which *TM6SF2* variant was expressed, although the intercept for the KK regression line was significantly higher than that of the EE line, indicating that, for any *TM6SF2* expression level, the KK variant exhibits higher *HMGCR* expression than the EE variant. The same trend was present for *SQLE* expression, but did not reach statistical significance (p=0.1596) (*Figure 4.11B*). We chose to continue our experiments in the iPSC lines exhibiting lower *TM6SF2* transgene expression to mimic endogenous levels as closely as possible.
Figure 4.11 Comparison of TM6SF2 exogenous gene expression vs cholesterol synthesis gene expression in iPSCs by linear regression A) Linear regression analysis of TM6SF2 exogenous gene expression vs HMGCR gene expression in iPSCs. B) Linear regression analysis of TM6SF2 exogenous gene expression vs SQLE gene expression in iPSCs.

Given the predominant phenotype of reduced CE content observed in the -/- iPSCs in the previous model, we next measured levels of iPSC free and esterified cholesterol at steady state (using a fluorometric assay), as well as levels of *de novo* synthesis of several lipid classes (via $^{14}$C-acetate radiolabeling, as performed previously). Just as we saw in our lipidomics experiment, CE concentration was lower in the KO_BB iPSCs, compared to EE_OE cells (Figure 4.12A). We did
not observe any significant difference in CE levels between EE_OE compared to KK_OE, but FC was lower (Figure 4.12B).

![Figure 4.12 Measurement of iPS cell cholesterol using Amplex Red fluorometric assay](image)

Figure 4.12 Measurement of iPS cell cholesterol using Amplex Red fluorometric assay
A) Comparison of cell free cholesterol (FC), cholesterol ester (CE), total cholesterol (TC), and percent CE of TC for EE_OE iPSCs to +/-_BB iPSCs. B) Comparison of cell free cholesterol (FC), cholesterol ester (CE), total cholesterol (TC), and percent CE of TC for EE_OE iPSCs to KK_BB iPSCs. Data are presented as ng standardized to total cell protein. +/- SEM, One-way ANOVA, *p<0.05, **p<0.01, ***p<0.001, n=4 per genotype.

Surprisingly, when we measured de novo synthesis of CE, we found that it was increased in the KK_OE cells, compared to OE_EE; this difference was even more striking when comparing the +/-_BB to OE_EE (Figure 4.13A). Once again, FC synthesis did not appear to be impaired in any of the cell lines, as was previously observed in the endogenous TM6SF2 iPSCs (Figure
While these data at first seem to contradict our previous experiments, they do provide support for our theory that TM6SF2 impacts cellular CE levels; further experiments will be needed to clarify the directionality of this relationship.

Figure 4.13 Impact of TM6SF2 Transgene expression on iPSC de novo lipogenesis. $^{14}$C-acetate incorporation into A) Cholesterol Ester and B) Free Cholesterol. Data are presented as mean species CPM standardized to total cell protein. +/- SEM, One-way ANOVA, *p<0.05, **p<0.01, ***p<0.001, n=4 per genotype.

Finally, we measured the incorporation of $^{14}$C-oleic acid into TG and CE to determine whether the esterification activity of DGAT and ACAT are impacted by TM6SF2 expression, respectively. This assay serves as an indicator of MAM activity, so the results may indicate whether the formation of these cholesterol-rich domains is impacted as well. We found that cholesterol esterification was decreased in the KK_OE and -/-_BB cells over time, compared to EE_OE cells (Figure 4.14A), while TG synthesis from DAG did not differ between groups (Figure 4.14B). We measured ACAT protein levels to determine if the reduced cholesterol esterification was caused by less ACAT expression. We did not see any significant difference in ACAT protein levels between the EE_OE and KK_OE cells (Figure 4.14C), suggesting that ACAT was equally expressed, but functionally impaired in the KK_OE iPSCs. These results
support our hypothesis that ACAT activity is defective when TM6SF2 is disrupted, and provide further evidence that MAM formation may be compromised. ACAT protein expression did appear reduced in the -/-_BB cells, compared to EE_OE, which does not support this hypothesis.

Figure 4.14 Impact of TM6SF2 transgene expression on ACAT and DGAT activity in iPSCs. ¹⁴C-oleic acid incorporation into A) Cholesterol Ester (CE) and B) Triglyceride (TG). Data are presented as mean species CPM standardized to total cell protein +/- SEM, One-way ANOVA, *p<0.05, **p<0.01, ***p<0.001, n=4 per genotype. C) Western blot of ACAT1 protein expression in iPSCs (left) and quantification of band intensity (right) using ImageJ software and standardized to β-actin.

Taken together, the data presented in this chapter strongly support a role for TM6SF2 in cholesterol metabolism. The transcriptional program regulating cholesterol synthesis was
significantly increased in iPSCs which had undergone CRISPR/Cas9 disruption of the *TM6SF2* gene. Further, we showed that CE levels were reduced in these KO iPSCs, both at steady state and when newly esterified. These results mirror those observed in the earlier portion of the chapter when non-isogenic, endogenously expressing EE and KK iPSCs were compared. The observed reduction in ACAT activity in TM6SF2 KO iPSCs further supports the possibility that TM6SF2 is necessary for the proper formation of cholesterol-enriched ER subdomains.

Additional work is necessary to determine if TM6SF2 disruption indeed alters SREBP2 activation by increasing its retention in the ER and if this is mediated by a defect in cholesterol sensing. Firstly, the co-localization of SREBP2 with ER and Golgi markers can be determined via immunofluorescence microscopy and by western blot analysis of immature versus cleaved SREBP2 abundance. Secondly, cells can be treated with exogenous cholesterol complexed to methyl b-cyclodextrin (MBCD), which causes the ER cholesterol concentration to increase. Under normal conditions, this would trigger a conformational change in SCAP, promoting the association of Insig with SCAP-SREBP and blocking the binding of COPII coat proteins necessary for the delivery of SREBP2 to the Golgi to be activated; these events ultimately would result in an attenuation of SREBP2 target gene expression. However, if SREBP2 target genes remain elevated following cholesterol-MBCD treatment, this would indicate that the SCAP-SREBP2 complex does not sense an increase in the local ER cholesterol concentration and SCAP-SREBP2 was not retained in the ER.

### 4.8 Detailed Methods

#### 4.8.1 Bulk RNA-sequencing of TM6SF2-KO iPSCs

The LG2-C6 TM6SF2-KO iPSC line, E5-17, along with its parent line that had been transduced with doxycycline-inducible Cas9 but not gRNA, were passaged onto Matrigel-coated
surfaces as small clusters to reach a confluency of 10% by the next day. Once the cells reached 20% confluency (3 days post-seeding), iPSCs were treated with 400 µM oleic acid (OA) in mTeSR Plus medium for 72 hours. An equal amount of DPBS was added to mTeSR Plus medium in the untreated (UT) groups. 3 biological replicates were prepared for each of the 4 groups. Total RNA was extracted using the RNeasy Mini Kit (Qiagen, 74104) and DNA was removed with DNase I (Qiagen, 79254) through on-column digestion. RNA quality control was determined with the Agilent 2100 Bioanalyzer System.

Cas9 parent line, TM6SF2-KO and their respective OA-treated groups were submitted for bulk RNA-sequencing at the Columbia Genome Center. The poly-A pull down method was used to enrich mRNA from total RNA in the samples, followed by library construction using Illumina TruSeq chemistry. Libraries were sequenced on the Illumina NovaSeq 6000 system, with 20 million reads per sample and pair-end reading length of 100 bp. Base calling, file conversion and gene alignment were performed by staff at the Columbia Genome Center. Briefly, RTA (Illumina) was used for base calling and bcl2fastq2 (version 2.19) was used for converting BCL to fastq format, coupled with adaptor trimming. Pseudoalignment was performed with kallisto (0.44.0) based on the human transcriptome GRCh38. A count table summarizing all the gene reads was returned. Each row in the table represented a gene, each column represented a sequenced RNA library, and the values represented the raw numbers of sequencing reads that were mapped to the respective genes in each library.

Data from RNA-seq were accessed using an R integrated development environment, RStudio, and bioinformatics software packages were downloaded from Bioconductor. Differential expression of genes were analyzed using DESeq2 (version 1.26.0). Briefly, a DESeqDataSet was constructed from the count table and a sample information table (Table 4.2), before applying
the DESeq2 pipeline. The results for each gene consisted of the base mean, log2 fold change, standard error estimate of the fold change, test statistic, P-value, and Benjamini-Hochberg (BH) adjusted P-value. The base mean is the average of the normalized count values divided by size factors taken over all samples, and the log2 fold change is the effect size estimate. The statistic and P-value are reported after DESeq2 performs a hypothesis test for each gene to determine if the null hypothesis is true. The BH adjusted P-value takes into consideration the false discovery rate (FDR) in high-throughput experiments such as RNA-seq and is used as a cutoff in obtaining a set of genes of statistical significance.

Table 4.2. Sample information table used for running the DESeq2 pipeline.

<table>
<thead>
<tr>
<th>CellType</th>
<th>Treatment</th>
</tr>
</thead>
<tbody>
<tr>
<td>KD001</td>
<td>Cas9</td>
</tr>
<tr>
<td>KD002</td>
<td>Cas9</td>
</tr>
<tr>
<td>KD003</td>
<td>Cas9</td>
</tr>
<tr>
<td>KD004</td>
<td>Cas9</td>
</tr>
<tr>
<td>KD005</td>
<td>Cas9</td>
</tr>
<tr>
<td>KD006</td>
<td>Cas9</td>
</tr>
<tr>
<td>KD007</td>
<td>KO</td>
</tr>
<tr>
<td>KD008</td>
<td>KO</td>
</tr>
<tr>
<td>KD009</td>
<td>KO</td>
</tr>
<tr>
<td>KD010</td>
<td>KO</td>
</tr>
<tr>
<td>KD011</td>
<td>KO</td>
</tr>
<tr>
<td>KD012</td>
<td>KO</td>
</tr>
</tbody>
</table>

Results of differentially expressed genes (DEGs) were exported from RStudio and analyzed with published biological databases. Pathway enrichment was performed in Reactome233–
and Enrichr, and metabolic pathways were further examined using HumanCyc. Gene network analysis was performed in Cytoscape (software version 3.8.2), using the STRING database to identify a network relevant to each gene list. Cytoscape applications, MCODE and ClueGo, were used to detect densely connected regions in a large gene network and to visualize biological processes for large gene clusters, respectively. Transcriptional factor and target interactions were analyzed using the TRRUST database.

4.8.2 Design of Lentiviral Plasmid for exogenous TM6SF2 Expression in iPSCs

To construct the plasmid containing the human TM6SF2 gene sequence, double-stranded gene fragments of both the major and minor (rs58542926 variant) alleles were synthesized de novo and cloned into the lentiviral plasmid, pLV-EF1α-IRES-NeoR (Addgene # 85139, a gift from Tobias Meyer). 5 µg of the plasmids were digested with BamHI-HF and EcoRI-HF restriction enzymes (New England BioLabs, R3136M and R3101M) and resolved through electrophoresis on a 1% agarose gel. The linear backbone was extracted from the gel and purified, before dephosphorylation with Antarctic Phosphatase.

TM6SF2 mRNA transcript (~1100 bp) was used as the template for the WT gene fragment, and a G→A substitution was created at the rs58542926 variant site for the mutant gene fragment. Both gene fragments were designed to contain a Kozak sequence and a BamHI cleavage site at the 5’ end, as well as an EcoRI cleavage site at the 3’ end. They were synthesized as gBlocks by Integrated DNA Technologies. 1 µg of the gene fragments were digested with BamHI-HF and EcoRI-HF, followed by purification using the MinElute PCR Purification Kit (Qiagen, 28004). Gene fragments were ligated with the dephosphorylated pLV-EF1α-IRES-NeoR using T4 DNA ligase (New England BioLabs, M0202M) at a molar ratio of 3:1. Ligation was performed on an
overnight cyclic temperature gradient (gradual decrease of 1°C/min from 20°C to 10°C, followed by an increase of 1°C/min from 10°C to 20°C; on repeat for 99 cycles).

The ligated products were transformed into *stbl3* competent cells and several colonies were selected for small liquid culture and Miniprep. Plasmid sequences were confirmed through two methods: (1) restriction enzyme digestion using Xmal, which has one cleavage site on the backbone and a second on the *TM6SF2* insert; (2) Sanger sequencing using pre-designed primers that span the length of insert (forward primer 1: 5'-ACCAGTTGCGTGGACGGAAGATG-3'; forward primer 2: 5'-TTAGGCCAGCCTGGCAGTGATG-3'; forward primer 3: 5'-TATGACCCACTCTATGCTGTC-3'; forward primer 4: 5'-CTGCCTTTCTTCCTCACC-3'; forward primer 5: 5'-TTATGTGCTGGCTTTCTCGG-3'). Bacterial clones with the correct inserts were used to prepare large liquid culture and plasmid Midiprep. Purified pLV-EF1α-TM6SF2.WT-IRES-NeoR and pLV-EF1α-TM6SF2.Mu-IRES-NeoR were used to produce lentiviruses as described previously.

### 4.8.3 *TM6SF2* WT and Mutant Overexpression in *TM6SF2-KO* iPSCs

LG2-C6 TM6SF2-KO iPSCs were detached with Accutase, gently broken down into single cells, and plated on surfaces pre-coated with Matrigel (0.5 million cells per well of 6-well plates). The next day, iPSCs were transduced with lentiviruses carrying the TM6SF2 WT or mutant transgenes, or the empty backbone with NeoR only. Titers from 0.1 to 15 IU were used to produce a wide range of TM6SF2 expression. Cells were treated with 100 µg/mL G418 sulfate (Thermo Fisher, 10131035) 48 hours post-transduction for 3 days to select for successfully transduced cells. None of the cells in the negative control (no lentivirus) survived after 3 days.
Chapter 5: Summary and Discussion

Cardiovascular disease (CVD) and Non-Alcoholic Fatty Liver Disease (NAFLD) are two of the world’s leading causes of death and contribute significantly to the health care burden in the developed world. While environmental factors, such as poor diet and inadequate exercise, are well established as strong risk factors for these metabolic diseases, recent advances in sequencing technology have heralded the age of large-scale human genomic studies. These studies have led to the identification of numerous genomic regions that harbor variants which significantly associate with disease traits. However, the dissection of the processes that contribute to the observed signal and the identification of the specific genes driving the disease association require extensive functional investigations. Multiple such studies illustrated a significant association between carriage of a non-synonymous variant in the \textit{TM6SF2} gene (rs58542926) encoding the E167K amino acid substitution with increased hepatic triglyceride content and reduced plasma lipid levels. The results from these genomic studies clearly demonstrated that TM6SF2 plays a crucial role at the intersection of NAFLD and CVD; therefore, for this dissertation, we explored the role of TM6SF2 in lipid metabolism.

At its outset, this study aimed at expanding our understanding of the mechanism whereby TM6SF2 impacts hepatic VLDL metabolism. A growing body of literature exists regarding this topic, but the inconsistent models and techniques used from study to study have led to discrepant and often contradictory findings. Therefore, our first goal was to develop a model with which we could perform reliable and translational studies. We established a collaboration with the Leong lab to optimize a technique for generating hepatocyte like cells (HLCs) through the directed differentiation of inducible pluripotent stem cells (iPSCs). In Chapter 2, we summarized the protocol optimizations and the validation tests employed to produce functionally mature HLCs.
Since HLC lipoprotein metabolism had not yet been characterized in detail, we performed several specialized experiments to ensure our model met or exceeded the functional capabilities of alternative hepatocyte systems. Importantly, we demonstrated that the HLCs produced with our optimized protocol secreted apoB-containing lipoproteins exhibiting densities within the VLDL range, a trait which is notably lacking in popular hepatoma models. Ultimately, the work presented in Chapter 2 supports the translational relevance and value of HLCs for modeling human lipoprotein metabolism.

With the generous provision of blood samples drawn from 4 sibling donors within the Old Order Amish (OOA) community carrying the wild type (EE) or mutant (KK) homozygous TM6SF2 alleles, we aimed to apply our optimized technique to generate donor specific HLCs; within these HLCs, we planned to experimentally investigate the impact of the KK variant on hepatic lipoprotein pathways. However, the challenges we faced when attempting to study the function of TM6SF2 in HLCs derived from OOA sibling pairs highlighted several drawbacks of this model that we had not anticipated, including variability in hepatic function and differentiation potential from donor-to-donor and batch-to-batch. While we mitigated some of these issues with additional protocol optimizations, persistent variability diminished any observable phenotype between the EE and KK sibling HLCs. Therefore, we ultimately chose to follow the lead of other HLC researchers and applied CRISPR/Cas9 editing to create isogenic iPSCs from which to produce HLCs differing only in their \textit{TM6SF2} genotype.

In Chapter 3, we introduced our gene editing strategy and provided substantial evidence for the successful disruption of the \textit{TM6SF2} gene in iPSCs. Despite this accomplishment, the anticipated phenotypic differences between the HLCs differentiated from the TM6SF2 knockout iPSCs, compared to the HLCs differentiated from the parent isogenic KK line, were absent or
We presumed that an isogenic EE (wild type) iPSC line (which could be differentiated to EE HLCs) would be necessary to amplify the impact of TM6SF2 disruption and improve the utility of the model.

In Chapter 4, we first presented the mounting evidence supporting a role for TM6SF2 in cholesterol metabolism as a rationale for exploring the effect of TM6SF2 disruption in our iPSCs. Initially, we took a global approach to determine the broad relevance of TM6SF2 in iPSCs. Our RNA-sequencing and lipidomics analyses confirmed that TM6SF2 KO iPSCs exhibited increased expression of SREBP2-regulated cholesterol synthesis genes and decreased cholesterol ester (CE) content, compared to both isogenic KK iPSCs and sibling-matched EE iPSCs. Subsequent functional assays revealed that the activities of enzymes that specifically localize to cholesterol enriched ER lipid raft-like domains were reduced in KO iPSCs.

We concluded from these initial experiments that TM6SF2 does indeed alter cholesterol metabolism in iPSCs. Since steady state and newly synthesized cellular cholesterol were not increased in the KO iPSCs, despite the observed increase in cholesterol synthesis gene expression, we hypothesized that lack of TM6SF2 may prevent the retention of the SREBP2-Insig-SCAP complex in the ER by hindering its sensing of local cholesterol. Further, we speculated that perhaps TM6SF2 impacts cholesterol localization within ER subdomains which regulate cholesterol homeostasis.

At the end of Chapter 4, we described our strategy to generate an isogenic EE iPSC line from which to complete our HLC model. Briefly, we delivered plasmids carrying the E or K-encoding transgene via lentivirus to the KO iPSC line that we engineered previously. This created a TM6SF2 wild type (EE), mutant (KK), and KO iPSC line possessing the same genetic background. We had hoped that we could then produce HLCs from these three iPSC lines which would exhibit reduced
variability, and thus serve as an improved system in which to further investigate hepatic TM6SF2 function; however, we found the further engineered iPSCs to be incapable of differentiation past the initial DE stage. Therefore, we instead opted to continue our exploration of TM6SF2 and cholesterol metabolism in these 3 iPSC lines and to validate our previous findings.

We confirmed the impact of TM6SF2 KO on gene expression levels of HMGCR and SQLE, the rate-limiting enzymes in the cholesterol synthesis pathway. While we did not consistently observe the same impact of TM6SF2 disruption on CE content in this system, we did observe a reduction in the cholesterol esterification activity of the ACAT enzyme, which notably relies on localizing to ER lipid raft subdomains for optimal function. The data from these experiments supported the working hypothesis that we developed earlier in the chapter, namely that TM6SF2 may alter cholesterol localization within the ER membrane or impact the access of some ER proteins to regulatory cholesterol pools.

The recent work by Li et al, which demonstrated that ERLINs and TM6SF2 physically interact and stabilize one another, further supported our hypothesis. ERLINs are involved in the retention of the SREBP2-Insig-SCAP complex in the ER and are also established residents of cholesterol enriched ER lipid rafts. Given these relationships, it is possible to conjecture that when TM6SF2 is disrupted, ERLINs are destabilized, which in turn releases precursor SREBP2 from its anchor in the ER and ultimately leads to an increase in the activation of SREBP2-mediated transcription. It is yet unknown how TM6SF2 acts to stabilize ERLINs; perhaps TM6SF2 acts as a scaffold, allowing for the proper localization or interaction of ER lipid raft resident proteins. Thus, the association between TM6SF2 and ERLINs may mediate the observed effect of TM6SF2 on cholesterol access/localization in iPSCs.
Given that iPSCs do not synthesize apoB-containing lipoproteins, we can conclude that the impact of TM6SF2 on cholesterol homeostasis occurs either upstream or independently of changes in VLDL metabolism. It is well established that hepatic cholesterol plays a pivotal role in the development of NASH in humans; further, human-like NASH can be induced in mice through alterations in hepatic cholesterol homeostasis. The strong association we observed between TM6SF2 and cholesterol homeostasis in iPSCs therefore points to a possible additional mechanism whereby altered TM6SF2 expression contributes to the development of NASH (independent of TM6SF2-induced steatosis), although this theory would need to be tested further. Additional work is also needed to address whether TM6SF2 similarly affects cholesterol in hepatocytes, which are a more relevant cell-type than iPSCs, and what downstream functional impact it might have, particularly with regards to VLDL metabolism, which could not be determined in iPSCs.

There is evidence that disruptions to cholesterol-enriched microdomains leads to phenotypic changes reminiscent of those observed with altered TM6SF2 expression, namely the development of NAFLD with reduced plasma lipids, suggesting that the cholesterol effect could be contributing to, or even sufficient to account for, the VLDL impairment as well. The first evidence of this connection came from Vance and colleagues, who were the first to identify the MAM domain; they observed the main apoproteins of VLDL (apoB, apoE, and apoC) to be present in the MAM and demonstrated that enzyme necessary catalyze the biosynthesis of all major lipid classes in VLDLs are present in the MAM. Finally, they asserted that the MAM comprises a pre-Golgi compartment containing likely apoB-lipoproteins within the VLDL density range in its lumen. Based on these findings, Vance et al proposed that the MAM serves as a component of the secretory pathway that is involved in the assembly of nascent VLDLs. Interestingly, these findings are supported by the recent identification and characterization by Anastasia and colleagues of a
novel organelle contact site occurring between rough ER wrapped around mitochondria, termed wrappER. They demonstrated that the establishment of wrappER is essential for the synthesis and secretion of VLDL in the liver by providing physical support for MAM domains.\textsuperscript{245} They further discovered that wrappER plays a role in the storage of FFAs in and the provision of TG for VLDL lipidation.

Finally, in addition to the physical association of TM6SF2 with several known MAM-resident proteins (as discussed in Chapter 4), there is a substantial enrichment of mitochondria-resident proteins in the top TM6SF2-interacting proteins detected by Luo et al, including ACSL5 and several oxidative enzymes.\textsuperscript{87} Since TM6SF2 has not been detected in mitochondrial membrane, this finding suggests that it may physically interact very closely with mitochondria from the ER, further supporting the likelihood of an association between TM6SF2 and MAM homeostasis.


155


57. Mitsche, M. A., Hobbs, H. H. & Cohen, J. C. Patatin-like phospholipase domain–containing protein 3 promotes transfers of essential fatty acids from triglycerides to phospholipids in


80. Musso, G., Cassader, M., Paschetta, E. & Gambino, R. TM6SF2 may drive postprandial lipoprotein cholesterol toxicity away from the vessel walls to the liver in NAFLD. *J. Hepatol.* **64**, 979–981 (2016).


116. Hansel, M. C. *et al.* The use of induced pluripotent stem cells for the study and treatment of


120. Godoy, P. *et al.* Recent advances in 2D and 3D in vitro systems using primary hepatocytes, alternative hepatocyte sources and non-parenchymal liver cells and their use in investigating mechanisms of hepatotoxicity, cell signaling and ADME. *Archives of Toxicology* vol. 87 (2013).


135. Sommer, A. G. et al. Generation of Human Induced Pluripotent Stem Cells from Peripheral


209. Rusiñol, A. E., Cui, Z., Chen, M. H. & Vance, J. E. A unique mitochondria-associated


218. Moebius, F. F., Striessnig, J. & Glossmann, H. The mysteries of sigma receptors: new


