Specificity and Functional Conservation of COPII Components

by

Vi To Ba Tang

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Molecular and Integrative Physiology)
in the University of Michigan
2023

Doctoral Committee:

Professor David Ginsburg, Co-Chair
Associate Professor Sharlene M. Day, Co-Chair, University of Pennsylvania
Assistant Professor Brian T. Emmer
Assistant Professor Adam S. Helms
Professor Ling Qi
Associate Professor Andrew W. Tai

Vi To Ba Tang

tvi@umich.edu

ORCID iD: 0000-0001-6079-9756

© Vi To Ba Tang 2023

Dedication

To my dad, for telling me in middle school that I should be a scientist

even though I didn't think I would become one

Acknowledgements

First and foremost, I would like to thank my thesis advisor, Dr. David Ginsburg. I could not have asked for a more caring and supportive mentor. Thank you for a being a great role model as a scientist, a mentor, and a human being. I would also like to thank my dissertation committee for your time, scientific input, career advices, and support.

To Dr. Brian Emmer and Dr. Rami Khoriaty, thank you for welcoming me into the Ginsburg lab and being wonderful mentors and friends when I was rotating and first joined the lab. To Rami, thank you for setting a great example of kindness, hard work, and life outside of the lab. I will forever cherish our time working together in the lab, going to happy hours, traveling, and hosting football tailgates. I would also like to thank members of the Ginsburg lab, past and present. You all have been wonderful colleagues and friends. Thank you, Suzann, for being on top of everything administrative and for many chats and laughs outside of David's office. Special thanks to Dr. Prabhodh Abbineni, for being my lunch companion in most weeks, and for your many advices on lab projects, career development, and life.

Next, I would like to thank Dr. Dan Michele, Dr. Sharlene Day, and Dr. Adam Helms. Thank you Dan and Sharlene for taking a chance on me by offering me an opportunity to come and work in Sharlene's lab in the summer of 2012. Thank you Adam for being my first mentor in the lab, for teaching me all the molecular biology techniques, and for showing me how to mentor new students, which I have and will strive to follow. You three have been my constant support at Michigan ever since I first set foot on campus in 2012.

I would also like to thank the faculty, staff, and students of the Physiology department. You are truly my home away from home. Thank you Dr. Sue Moenter, Dr. Dan Michele, and Dr. Sue Brooks for being supportive graduate chairs. Many thanks to Michele Boggs for being the best Graduate Coordinator ever. To current and former MIP students, you are the best group of friends I could ask for. Special shout-out to: Kristy, Ben, and Rudi, for being constant pillars of support throughout the graduate school journey; Dr. Keita Uchida, Dr. Margo Emont, and Dr. Thomas Vigil, for hosting me during the interview weekend many years ago and for your advices since my first year up until now; Steve, Lloyd, Chris, Warren, Nick, Ryan, for joining me in countless pub nights, tailgates, and dinners. These are memories that I will forever treasure.

To Dr. Sara Saberi, Drs. Fredy and Ashleigh El Sakr, Dr. Dan Cronin, Dr. Paul Christine, Dr. Max Nagle, and Dr. Elizabeth Paluga, thank you for keeping me healthy through the years and occasionally giving me valuable life advices. You are a true model of holistic care.

Last but not least, my sincere thanks to my family and friends around the world. To my parents, thank you for your unconditional love and all your sacrifice for me to be where I am today. To Jenny, thank you for inspiring my dream of coming to the U.S. ever since I was three years old. I would definitely not be here without you. To my aunts, uncles, and many cousins, thank you for all the love and support. You have given me the most valuable gift in life: confidence that I will always have someone to fall back on.

Table of Contents

Dedication ..... **ii**

Acknowledgements ..... **iii**

List of Tables ..... **x**

List of Figures ..... **xi**

Abstract ..... **xiii**

Chapter 1 Cargo Selection in Endoplasmic Reticulum to Golgi Transport and Relevant Diseases **1**

1.1 Abstract 1

1.2 Introduction 1

1.3 ER to Golgi Transport by Coat Protein Complex II 2

1.3.1 COPII composition 2

1.3.2 Coat Assembly 2

1.4 Bulk flow and concentrative transport 3

1.4.1 Transport from the ER to the Golgi 4

1.5 Expansion of COPII paralogs in mammals 5

1.6 Specificity of COPII paralogs 7

1.6.1 SAR1 7

1.6.2 SEC23 9

1.6.3 SEC24 11

1.6.4 SEC31 14

1.6.5 SEC13 15

1.7 Specificity of cargo receptors 16

1.7.1 LMAN1 16

1.7.2 SURF4 18

1.8 Concluding remarks 20

Chapter 2 Overlap in Function between the SAR1A and SAR1B Paralogs *in vivo* 22

2.1 Abstract 22

2.2 Introduction 23

2.3 Methods 25

2.3.1 Animal care 25

2.3.2 Generation of Sar1a$^{+/gt}$ mice 25

2.3.3 Generation of Sar1b$^{+/-}$ mice 26

2.3.4 Generation of Sar1b$^{fl/+}$ Alb-Cre^{+} mice 26

2.3.5 Generation of Sar1b$^{+/a}$ mice 26

2.3.6 Mouse genotyping 27

2.3.7 Timed mating and embryo collection 27

2.3.8 Blood and tissue collection 27

2.3.9 Histologic analyses 28

2.3.10 Cholesterol assays 28

2.3.11 Adenovirus mediated expression of SAR1A or SAR1B 28

2.4 Results 28

2.4.1 Germline deletion of Sar1a in mice results in mid-embryogenic lethality 28

2.4.2 Germline homozygous Sar1b deletion results in perinatal lethality 31

2.4.3 Hypocholesterolemia resulting from hepatic SAR1B deficiency is rescued by overexpression of either SAR1A or SAR1B 32

2.4.4 Expression of Sar1a coding sequence at the Sar1b endogenous locus rescues the lethal Sar1b null phenotype in mice 34

2.5 Discussion 36

Chapter 3 Identification of Secreted Proteins by Comparison of Protein Abundance in Conditioned Media and Cell Lysates 42

3.1 Abstract 42

3.2 Introduction 42

3.3 Methods 44

3.3.1 Cell culture, and BFA treatment 44

3.3.2 Conditioned medium collection 44

3.3.3 Cell lysate collection 45

3.3.4 Immunoblotting 45

3.3.5 Mass spectrometry 45

3.3.6 Liquid chromatography-mass spectrometry analysis (LC-multinotch MS3): 46

3.3.7 Protein identification and quantification 47

3.4 Results 49

3.4.1 Identification of secreted proteins by comparison of protein abundance in the cell lysate and conditioned media 49

3.4.2 BFA inhibits ER-Golgi trafficking and further validates the M/L ratio analysis approach 53

3.4.3 Identification of unconventionally secreted proteins 57

3.5 Discussion 60

Chapter 4 Identification of LMAN1 and SURF4 Dependent Secretory Cargoes 64

4.1 Abstract 64

4.2 Introduction 64

4.3 Materials and Methods 66

4.3.1 Cell culture 66

4.3.2 CRISPR mediated inactivation of LMAN1 and SURF4 67

4.3.3 Conditioned media and cell lysate collection 67

4.3.4 Immunoblotting 68

4.3.5 Mass spectrometry, protein identification, and protein quantification 68

4.3.6 Statistical analyses 68

4.3.7 Analysis of ER-ESCAPE motifs 69

4.4 Results 69

4.4.1 Identification of bona fide secreted proteins by analysis of cell lysates and conditioned media 69

4.4.2 Few LMAN1-dependent cargoes identified in HuH7 cells 71

4.4.3 SURF4 traffics a wide range of secretory proteins in HuH7 cells 74

4.4.4 Weak correlation of ER-ESCAPE motif with SURF4 cargoes 75

4.5 Discussion 77

Chapter 5 Hepatic Inactivation of Murine *Surf4* Results in Marked Reduction in Plasma Cholesterol 80

5.1 Abstract 80

5.2 Introduction 81

5.3 Materials and Methods 82

5.3.1 Animal care and use 85

5.3.2 Generation of conditional Surf4 knockout mice 85

5.3.3 Genotyping assays 85

5.3.4 Blood and tissue collection 86

5.3.5 Analysis for sera from Surf4 liver knockout mice 86

5.3.6 Hepatic lipoprotein secretion assay 86

5.3.7 Oral fat tolerance test and lipid flux assay 87

5.3.8 Liver lipid extraction and quantification 87

5.3.9 Immunoblotting 88

5.3.10 Histology 88

5.3.11 Analysis of liver mRNA 88

5.3.12 Acute inactivation of hepatic Surf4 in adult mice 89

5.4 Results 90

5.4.1 Liver-specific deletion of Surf4 is compatible with normal development and survival in mice 90

5.4.2 Reduced circulating PCSK9 and increased LDLR levels in $Surf4^{fl/fl}$ Alb-Cre^{+} mice .. 93

5.4.3 Marked reduction of plasma cholesterol in $Surf4^{fl/fl}$ Alb-Cre^{+} mice 94

5.4.4 Intestinal lipid absorption and tissue lipid uptake are unaffected in $Surf4^{fl/fl}$ Alb-Cre^{+} mice 99

5.4.5 Loss of liver Surf4 expression results in mild lipid accumulation but no steatohepatitis or fibrosis 101

5.4.6 Dose-dependent reduction of plasma lipids in response to depletion of SURF4 by siRNA 103

5.5 Discussion 105

Chapter 6 Conclusions and Future perspectives 110

6.1 Functional conservation of COPII proteins 110

6.2 Cargo-cargo receptor relationship 113

6.3 SURF4 as a cargo receptor in mammals 114

Bibliography 117

List of Tables

Table 1-1: COPII paralogs in mammals and associated human disease or mouse phenotype 6

Table 2-1: Genotype distribution of *Sar1a*$^{gt/+}$ intercross 29

Table 2-2: Genotype distribution of *Sar1b*$^{+/-}$ intercross 31

Table 2-3: Genotype distribution of *Sar1b*$^{fl/+}$ *Alb-Cre*$^{+}$ intercross 33

Table 2-4: Genotype distribution of *Sar1b*$^{+/a}$ intercross 35

Table 2-5: Tukey's post-hoc test for comparison between AV type, time, and dose 40

Table 2-6: Genotyping primer sequences 41

Table 5-1: Key resources used in this study 83

Table 5-2: Genotype distribution of offspring of *Surf4*$^{fl/fl}$ *Alb-Cre*$^{+}$ and *Surf4*$^{fl/+}$ *Alb-Cre*$^{-}$ intercrosses 90

Table 5-3: Primers used in this chapter 109

Table 5-4: Guide RNA sequences for CRISPR-mediated *in vivo* hepatic *Surf4* inactivation 109

Table 5-5: Sequences and modifications of oligonucleotides for siRNA in mice 109

List of Figures

Figure 1-1: COPII coat assembly on the endoplasmic reticulum (ER) membrane. 3

Figure 1-2: ER to Golgi transport of secreted proteins. 5

Figure 2-1: Alignment of the mouse SAR1A and SAR1B protein sequences. 24

Figure 2-2: Complete loss of SAR1A results in mid-embryogenesis lethality. 30

Figure 2-3: Germline deletion of *Sar1b* results in perinatal death. 32

Figure 2-4: Hypocholesterolemia in hepatic SAR1B deficient mice can be rescued by overexpression of either SAR1A or SAR1B. 34

Figure 2-5: SAR1A expressed under control of the *Sar1b* endogenous locus can compensate for complete loss of SAR1B. 36

Figure 3-1: Comparison of protein abundance in the cell lysate and conditioned media enables identification of secreted proteins. 52

Figure 3-2: M/L ratio analysis combined with BFA treatment enables high-sensitivity identification of conventionally secreted proteins. 54

Figure 3-3: Proteins detected solely in the media fraction are sensitive to BFA. 56

Figure 3-4: Candidate proteins secreted via unconventional secretory pathways. 58

Figure 3-5: BFA sensitivity of transmembrane domain containing proteins. 60

Figure 4-1: Identifying proteins dependent on LMAN1 or SURF4 for efficient secretion. 70

Figure 4-2: Comparison of protein abundance in the media and lysates allows identification of bona fide secreted proteins. 71

Figure 4-3: Differential effects on protein secretion in HuH7 cells following LMAN1 or SURF4 deletion. 72

Figure 4-4: Comparison of proteins abundance in the lysates and media between LMAN1 deficient and control cells. 73

Figure 4-5: Efficient depletion of LMAN1 in cells treated with LMAN1g1 sgRNA. 74

Figure 4-6: Comparison of proteins abundance in the lysates and media between SURF4 deficient and control cells. 75

Figure 4-7: Most SURF4 cargoes do not contain an ER-ESCAPE motif [163]. 76

Figure 5-1: Generation of hepatocyte-specific *Surf4* deficient mice. 92

Figure 5-2: Read counts mapping to exon-exon junctions along the *Surf4* transcript based on RNA-sequencing data. 93

Figure 5-3: Deletion of hepatic *Surf4* results in decreased serum PCSK9 level and profound hypocholesterolemia in mice. 96

Figure 5-4: Normalized (Norm.) *Pcsk9* mRNA levels in $Surf4^{fl/fl}$ *Alb-*Cre^{-}.and $Surf4^{fl/fl}$ *Alb-*Cre^{+} livers. 97

Figure 5-5: Hepatic *Surf4* inactivation does not affect body mass while hypocholesterolemia is sustained through at least 1 year of age. 97

Figure 5-6: Hepatic lipoprotein and APOB secretion defect in $Surf4^{fl/fl}$ *Alb-*Cre^{+} mice. 98

Figure 5-7: Inactivation of hepatic *Surf4* does not impact dietary lipid absorption, incorporation, and oxidation. 100

Figure 5-8: Hepatic *Surf4* deletion results in mild hepatomegaly and an increase in liver lipid content, without apparent liver dysfunction or steatohepatitis. 102

Figure 5-9: Hepatic SURF4 deficiency does not result in liver injuries. 103

Figure 5-10: Depletion of hepatic *Surf4* by siRNA recapitulates the hypolipidemia seen in $Surf4^{fl/fl}$ *Alb-*Cre^{+} mice. 104

Figure 5-11: Acute inactivation of hepatic SURF4 results in decreased secretion of apolipoproteins. 105

Abstract

Protein secretion into the extracellular space is a fundamental process in all eukaryotic cells. In multicellular organisms, secreted proteins participate in critical processes such as intercellular communication, maintenance of metabolic homeostasis, immune reactions, and neurotransmission. The majority of secreted proteins follow the intracellular secretory pathway starting from the endoplasmic reticulum (ER). The first step in this pathway is the recruitment and transport of these proteins from the ER to the Golgi apparatus by coat protein complex II (COPII). Cargo recruitment is mediated by COPII protein in conjunction with cargo receptors – ER transmembrane proteins that bridge interactions between COPII proteins and soluble proteins restricted to the ER lumen. This dissertation examines the specificity of this process. We explored whether the two mammalian paralogs for the COPII component SAR1 are functionally equivalent and found that they exhibit extensive functional overlap such that SAR1A can compensate for the SAR1B deficiency in mice. We next focused on the two prototypical mammalian cargo receptors LMAN1 and SURF4. We developed a novel mass spectrometry-based analysis approach to identify bona fide secreted proteins with high sensitivity. We then applied this technique to define the cargo repertoire for LMAN1 and SURF4, finding that SURF4 has a much more extensive cargo range than LMAN1 in hepatocytes. Finally, we investigated the physiologic roles of hepatic SURF4 in mice, revealing that SURF4 mediates the secretion of lipoproteins from hepatocytes, with mice lacking hepatic SURF4 exhibiting profound hypocholesterolemia. Together, these data demonstrate both functional conservation among COPII paralogs and specificity of cargo receptors in mammals.

Chapter 1
Cargo Selection in Endoplasmic Reticulum to Golgi Transport and Relevant Diseases[1]

1.1 Abstract

Most proteins destined for the extracellular space or various intracellular compartments must traverse the intracellular secretory pathway. The first step is the recruitment and transport of cargoes from the endoplasmic reticulum (ER) lumen to the Golgi apparatus by coat protein complex II (COPII), consisting of five core proteins. Additional ER transmembrane proteins that aid cargo recruitment are referred to as cargo receptors. Gene duplication events have resulted in multiple COPII paralogs present in the mammalian genome. Here, we will review the functions of each COPII protein, human disorders associated with each paralog, and evidence for functional conservation between paralogs. We will also provide a summary of current knowledge for two prototypical cargo receptors in mammals, LMAN1 and SURF4, and their roles in human health and disease.

1.2 Introduction

Cells secrete proteins into the extracellular environment for multiple purposes, including intercellular communication, defense, and modulating the external environment. In eukaryotes, the majority of secreted proteins must transit through a series of membrane-bound organelles

[1]This chapter was published in the Journal of Clinical Investigation: Tang, VT, Ginsburg D. *Cargo selection in endoplasmic reticulum to Golgi transport and relevant diseases.* J Clin Invest. 2023 Jan 3;133(1):e163838.

before arriving at the plasma membrane for release into the extracellular environment [1-3]. This pathway is also shared by many proteins destined for the cell surface and various intracellular organelles, such as endosomes and lysosomes [2, 3]. Proteins enter the intracellular secretory pathway at the endoplasmic reticulum (ER) where they are cotranslationally embedded into the ER membrane or deposited into the ER lumen [4, 5]. Properly folded proteins are then transported from the ER to the Golgi for further processing and sorting before reaching their final extracellular or intracellular destinations [6, 7]. A few proteins have been shown to exit the cell via an ER/Golgi independent process referred to as unconventional protein secretion [8-11]. This review will focus on the early phase of conventional protein secretion, in which proteins are transported from the ER to the Golgi in a process mediated by coat protein complex II (COPII). Genetic disorders associated with deficiency in protein components of COPII will also be described.

1.3 ER to Golgi Transport by Coat Protein Complex II

1.3.1 COPII composition

The components of the COPII coat were first identified through a genetic screen by Randy Schekman and colleagues in 1980 for *Saccharomyces cerevisiae* mutants that were defective in protein secretion, designated “sec” mutants [12]. The core COPII components in *S. cerevisiae* are Sar1p, Sec23p, Sec24p, Sec13p, and Sec31p [13]. The corresponding vertebrate proteins are referred to as SAR1, SEC23, SEC24, SEC13, and SEC31. SAR1 is a small GTPase that recruits other coat proteins to the ER membrane. SEC23 and SEC24 form an inner coat complex that is proximal to the ER membrane. SEC13 and SEC31 form the outer coat complex (**Figure 1-1**).

1.3.2 Coat Assembly

COPII coat assembly begins in the membrane of the smooth ER (lacking ribosomes) where the localization of SEC12 defines the ER exit site (ERES). SEC12 is a type II ER transmembrane protein that functions as a guanine nucleotide exchange factor (GEF) for SAR1 while also recruiting SAR1 to the ER membrane [14, 15]. GTP-bound SAR1 inserts its hydrophobic N-terminus into the ER membrane and recruits SEC23-SEC24 heterodimers to the ERES by directly interacting with SEC23 [16, 17]. SEC23 stimulates SAR1 GTPase hydrolysis, thereby functioning as the GTPase-activating protein (GAP) for SAR1, while SEC24 mediates cargo recruitment and concentration in the ER lumen [18]. Finally, SEC13 and SEC31 are recruited as an outer coat to complete the coat assembly [19] (**Figure 1-1**).

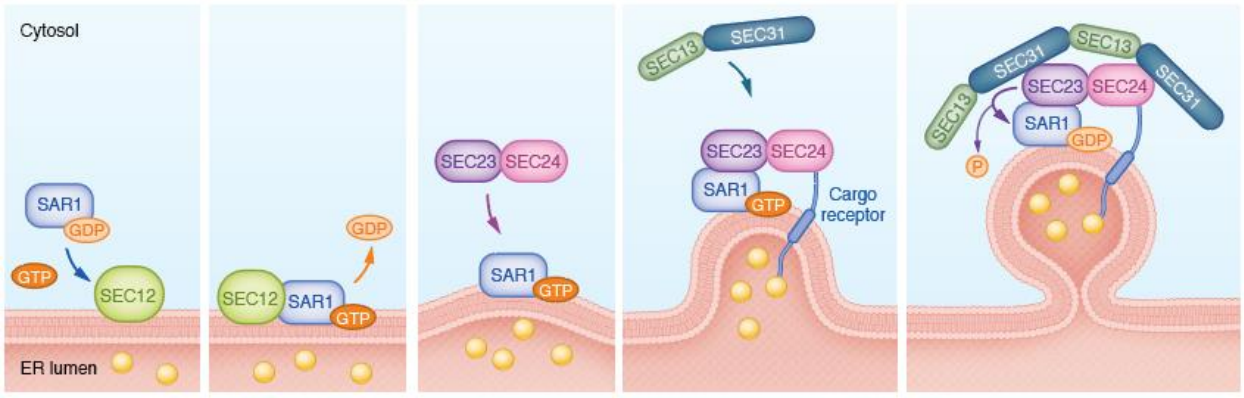

Figure 1-1: COPII coat assembly on the endoplasmic reticulum (ER) membrane. SEC12 recruits GDP-bound SAR1 to ER exit sites and acts as a guanine nucleotide exchange factor (GEF) for SAR1. GTP-bound SAR1 inserts its hydrophobic N-terminus into the ER membrane and recruits SEC23-SEC24 heterodimers to the ER exit site (ERES) via direct interaction with SEC23. SEC24 mediates cargo recruitment via direct physical interaction with transmembrane proteins through their cytoplasmic tails or with soluble cargoes via cargo receptors. SEC23 also functions as the GTPase-activating proteins (GAP) for SAR1 and stimulates SAR1 GTP-hydrolysis. Finally, SEC13-SEC31 heterotetramers are recruited as the outer coat to complete the coat assembly process. **Bulk flow and concentrative transport**Bulk flow is the default secretory pathway in which proteins in the ER lumen can freely and passively leave the ER [20, 21]. However, early immunoelectron microscopy and *in vitro* COPII reconstitution studies demonstrated that some secretory cargoes are selectively concentrated into COPII vesicles/tubules [22-25]. Active cargo concentration implies interaction between COPII subunits on the cytoplasmic face of the ER membrane and

secretory proteins within the ER lumen. While some transmembrane secretory proteins can directly interact with SEC23/SEC24 in the COPII coat [26-28], other transmembrane and soluble cargoes (proteins without a transmembrane domain that are entirely constrained within the ER lumen) achieve physical interaction with the COPII coat via transmembrane cargo receptors [23] (**Figure 1-1**). Several transmembrane cargo receptors and adaptors have been identified in yeast and mammals, including LMAN1, SURF4, lectins such as VIP36, and p24 proteins [29], though it is unclear if all or most secretory proteins require a cargo receptor for efficient secretion.

1.4.1 Transport from the ER to the Golgi

There are several models for how proteins are transported from one cellular compartment to another [30]. From the early 1960s, two distinct models for the identity of transport carriers between the ER and the Golgi have been proposed: discrete vesicles [31] and continuous tubules/tunnels [32, 33]. The foundation for the understanding of COPII-mediated ER to Golgi transport over the next few decades relied largely on the vesicular model, in which, following cargo recruitment, the outercoat (SEC13-SEC31) promotes ER membrane fission to generate discrete COPII vesicles of 60-90 nm in size [34-37]. These vesicles transport secretory proteins, along with the COPII coat proteins, to the ER-Golgi Intermediate Compartment (ERGIC) or the Golgi, where the vesicles fuse with the Golgi membrane, releasing their secretory cargo into the cis-Golgi network. The ERGIC is a stable membrane compartment between the rough ER and the Golgi that acts as the first post-ER sorting station for anterograde (to the Golgi) and retrograde (back to the ER) trafficking [38]. Challenges to this model include the discovery of COPII-free transport carriers [39] and explaining the mechanism for transport of large cargoes such as procollagen [40] or prechylomicrons [41] whose diameters exceed the typical size of COPII vesicles. Recent studies using super-resolution imaging techniques on intact mammalian

cells have visualized an interwoven ER network [42] suggesting that COPII coat proteins may remain on the ER membrane and function as a gatekeeper, restricting entry of secretory proteins into tubules rather than acting as an escort accompanying these proteins to the Golgi [43, 44] (**Figure 1-2**).

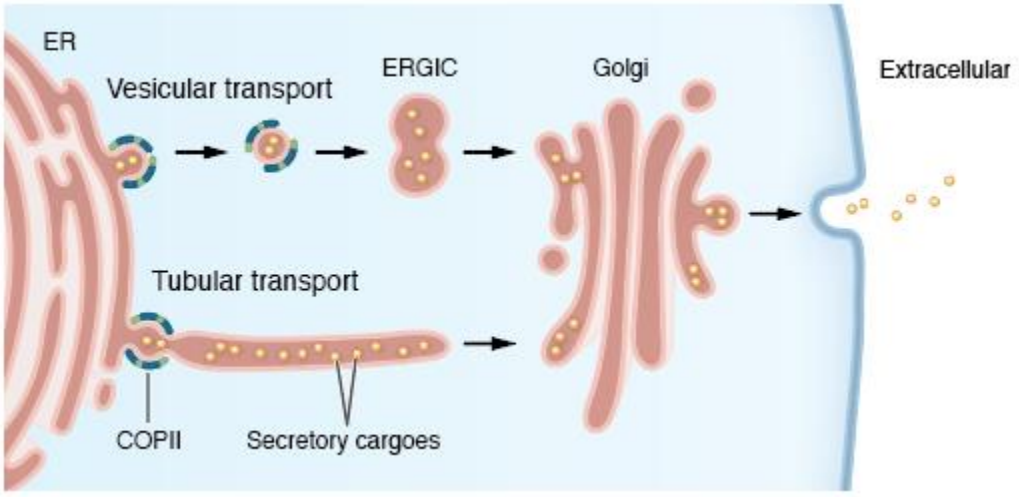

Figure 1-2: ER to Golgi transport of secreted proteins. Secretory proteins in the ER lumen are recruited into the COPII vesicle/tubule by COPII coat proteins. In the vesicular transport model, the vesicle buds from the ER and travels to the ERGIC/cis-Golgi network with COPII coat proteins accompanying the vesicle. In the tubular transport model, cargo proteins are transported in a continuous interwoven tubular network instead of discrete vesicles. COPII proteins remain on the ER membrane and function as gatekeeper restricting entry of secretory proteins into tubules.

1.5 Expansion of COPII paralogs in mammals

COPII proteins are highly conserved throughout eukaryotic evolution, and SEC23 and SEC24 orthologs have also been identified in Asgard, an archaea superphylum that is considered the closest prokaryotic relative of eukaryotes [45, 46]. Comparative genomic studies identify at least one paralog for each of the five core COPII proteins (SAR1, SEC23, SEC24, SEC13, and SEC31) in every eukaryote genome analyzed, suggesting that all 5 are present in the last eukaryotic common ancestor [47, 48]. Gene duplications have led to expansions of all five COPII proteins in multicellular plants and all but SEC13 in vertebrates (**Table 1-1**). Gene duplications are common evolutionary events. Typically, one duplicated copy accumulates loss-

of-function mutations, becoming a pseudogene, and is eventually lost from the genome. Occasionally, both gene copies are conserved through neofunctionalization, in which one paralog acquires a new function, or subfunctionalization, in which the paralogs divide the functions of the ancestral gene. In the latter case, the division of functions can occur at the protein level or at the transcription level [49]. A widely proposed explanation for the existence and conservation of multiple COPII paralogs in higher organisms is that each paralog has evolved unique functions to accommodate a more diverse secretome across different cell types. However, recent studies demonstrate that some paralogs have largely similar functions and can thus compensate for the loss of their partner paralog. Below, we will review and compare the reported human and mouse phenotypes associated with genetic deficiency for each COPII component as well as evidence for unique and/or redundant functions between paralogs.

Table 1-1: COPII paralogs in mammals and associated human disease or mouse phenotype

S. cerevisae	**Mammals**	**Human disease**	**Mouse phenotype**
Sar1p	SAR1A	None known	None reported
	SAR1B	CMRD (Anderson's disease) [50-52]	Late gestation lethality [53]
Sec23p	SEC23A	Cranio-lenticulo-sutural dysplasia (CLSD) [54, 55]	Embryonic lethality (~E12.5) [57]
	SEC23B	Congenital erythropoietic anemia type II (CDAII) [56]	Perinatal lethality, exocrine pancreas defect [58]
Sec24p	SEC24A	None known	Hypocholesterolemia [60]
Lst1p Iss1p	SEC24B	None known	Lethality (~E18.5), neural tube defect [61]
	SEC24C	None known	Embryonic lethality (~E7.5) [62]
	SEC24D	Syndromic form of osteogenesis imperfecta [59]	Embryonic lethality (before the 8-cell stage) [63]
Sec31p	SEC31A	None known	None reported
	SEC31B	None known	None reported
Sec13p	SEC13	None known	Lethality [64]

1.6 Specificity of COPII paralogs

1.6.1 SAR1

The first COPII protein recruited by SEC12 to the ERES is SAR1, which, in turn, recruits other COPII subunits to the ER membrane. Regulation of SAR1 GTPase kinetics by SEC23 and SEC31 is thought to be important for concentration of large cargo proteins [65]. Most invertebrate genomes contain a single *SAR1* gene, whereas most vertebrates, including mammals, contain 2 paralogs: *SAR1A* and *SAR1B*. The human SAR1A and SAR1B paralogs differ at only 20/198 amino acid residues. Though the average degree of amino acid sequence identity for orthologous proteins between the human and mouse genomes is ~85%, each of the human and mouse SAR1 ortholog pairs exhibit ~98% sequence identity, differing in only 3 amino acids for SAR1A and 2 amino acids for SAR1B.

Despite this high degree of similarity between the two SAR1 paralogs, human genetic data suggest distinct functions for SAR1A and SAR1B. Although both *SAR1* paralogs are highly expressed in the intestine, only mutations in *SAR1B* have been linked to the rare autosomal recessive disorder, chylomicron retention disease (CMRD, or Anderson's disease) [50-52], with no human disorder yet associated with mutations in *SAR1A*. Despite a ~1.5-fold increase in the expression of *SAR1A* in CMRD patients, this increase does not compensate for the loss of SAR1B [51]. CMRD is characterized by failure to thrive and chronic diarrhea in infants due to malabsorption of dietary lipids and fat-soluble vitamins [66, 67]. In CMRD patients, secretion of chylomicrons from the enterocytes is inhibited due to an inability to transport chylomicrons from the ER and/or abnormal fusion of prechylomicron transport vesicles to the Golgi – processes for which SAR1B appears to be required [68].

To explore the question of redundant/unique function between SAR1 paralogs at the molecular level, Melville et al [69] identified three clusters of at least three amino acid sequence differences between SAR1A and SAR1B. Two of these clusters are highly conserved among mammals, reptiles, and birds. One cluster is adjacent to the GTP-binding pocket and the other is on an α-helix near the known binding site of SEC31 on SEC23, which might influence SAR1 interaction with the SEC23/SEC31 complex. Further analyses demonstrated a faster GTPase exchange rate for SAR1A than for SAR1B, with a swap of the amino acid cluster adjacent to the GTP-binding pocket reversing this difference between the two paralogs. In contrast, SAR1B binds more strongly to SEC23 than does SAR1A and swapping the α-helix amino acid cluster reverses SEC23 affinity. These data indicate that SAR1A and SAR1B have some distinct biochemical properties at the molecular level.

However, limited *in vitro* data suggest some degree of functional overlap for these SAR1 paralogs. Inactivation of *SAR1B* in human Caco-2/15 cells, a chylomicron-secreting cell line, results in decreased chylomicron secretion, disrupted lipid homeostasis, and increased oxidative stress [70, 71]. SAR1A null Caco-2/15 cells exhibit a similar but less dramatic phenotype [71]. Combined deletion of SAR1A and SAR1B was required to recapitulate the more severe phenotype observed in patients with CMRD. This discrepancy between *in vivo* and *in vitro* could potentially be explained by relative expression levels for *SAR1A* and *SAR1B*. *SAR1B* is expressed at ~3-fold higher levels than *SAR1A* in human intestine [51], while *SAR1A* is expressed at slightly higher levels than *SAR1B* in Caco-2/15 cells, with a further ~1.3-fold increase in *SAR1A* expression following *SAR1B* deletion [70]. In mice, loss of expression from a single *Sar1b* allele is sufficient to recapitulate the reduced chylomicron secretion phenotype seen in humans, with homozygous *Sar1b* deletion in mice resulting in late-gestation lethality [53]. Taken together,

these data suggest at least some degree of functional overlap between SAR1A and SAR1B, with the total level of SAR1 expression providing an important determinant for CMRD manifestations.

1.6.2 SEC23

SEC23 is a cytosolic protein that forms a heterodimer with SEC24. The SEC23-SEC24 complex is recruited to the ER membrane by SAR1, and together with SAR1, forms the inner COPII coat, which recruits cargo proteins from the ER lumen [72, 73]. Though invertebrate genomes generally encode a single *SEC23* gene, most vertebrate genomes, including mammals, encode two *SEC23* paralogs, *SEC23A* and *SEC23B* [48]. The *SEC23* gene duplication is estimated to have occured ~615 million years ago. The two mammalian *SEC23* paralogs share ~85% sequence identity at the amino acid level, whereas the human and mouse SEC23A and SEC23B orthologs are ~98% and ~97% identical at the protein level, respectively [74].

Human genetic data again suggest that the *SEC23* paralogs have evolved divergent functions. Despite ubiquitous expression of both *SEC23A* and *SEC23B*, loss of function mutations in each paralog lead to different disorders affecting distinct cell types. Mutations in human *SEC23A* result in defective collagen secretion leading to the autosomal recessive condition cranio-lenticulo-sutural dysplasia (CLSD), which is characterized by abnormal cranial fontanel closures, facial dysmorphisms, skeletal abnormalities, and sutural cataracts [54, 55]. In contrast, mutations in *SEC23B* result in congenital dyserythropoietic anemia type II (CDAII), an autosomal recessive disorder characterized by anemia and increased numbers of bi/multinucleated red blood cell precursors in the bone marrow [56, 75]. Additionally, expression of human SEC23A and not SEC23B has been shown to rescue Sec23p null yeast, also consistent with unique functions for the two mammalian SEC23 paralogs [76].

Initial results from mouse models also suggested distinct, though potentially partially overlapping functions for SEC23A and SEC23B. *Sec23a* null mice exhibit mid-embryonic lethality associated with defective extraembryonic membrane development and neural tube closure in the midbrain, likely due to impaired secretion of multiple collagen types [57], and consistent with the collagen secretion defect observed in humans with CLSD. However, in contrast to human SEC23B deficiency, SEC23B-deficient mice exhibit an entirely normal red blood cell phenotype [58, 77], instead demonstrating perinatal lethality due to degeneration of the pancreas [58, 78]. There is also no evidence for altered collagen secretion in *Sec23b* null mice [58]. Of note, complete loss of one SEC23 paralog in combination with haploinsufficiency of the remaining paralog results in embryonic death at an earlier development time [57], suggesting some degree of overlapping functions between the paralogs.

Though it was previously demonstrated that only human SEC23A can rescue Sec23p deficiency in yeast, a more recent study shows that murine and human SEC23A and SEC23B are each individually sufficient to complement the loss of Sec23p in yeast, and delivery of a *sec23a* transgene rescues the lethality of Sec23b deficiency in zebrafish [74]. Additionally, substitution of the SEC23A coding sequence for that of SEC23B at the endogenous *Sec23b* locus fully rescued the perinatal lethal pancreatic degeneration seen in SEC23B-deficient mice, with no apparent abnormalities in these adult animals expressing only SEC23A sequences, but under the regulatory control of both the endogenous *Sec23a* and *Sec23b* genes [74]. Similarly, SEC23A has been demonstrated to overlap in functions with SEC23B in human erythroid cells with increased SEC23A expression being sufficient to rescue the erythroid differentiation defect in SEC23B-deficient cells [79]. Taken together, these data demonstrate that the two mammalian SEC23 paralogs exhibit highly overlapping and potentially identical functions, with the

discordant phenotypes observed between SEC23A and SEC23B deficiencies, and between humans and mice, resulting primarily from evolutionary differences in tissue-specific gene expression programs. Consistent with this hypothesis, murine *Sec23a* gene expression has been shown to be maintained throughout erythropoiesis, in contrast to human *SEC23A* expression, which declines rapidly upon induction of terminal erythroid differentiation [80], potentially explaining the absence of a red blood cell phenotype in mice with hematopoietic deficiency of SEC23B [77]. Indeed, inactivation of all four *Sec23* alleles in erythroid cells is required to reproduce the CDAII phenotype in mice [79]. While subtle differences in biochemical properties between the SEC23 paralogs (as between the SAR1 paralogs) cannot be excluded, these data suggest that mammalian SEC23A and SEC23B are largely functionally interchangeable.

1.6.3 SEC24

SEC24 is the primary COPII component responsible for cargo selection via either direct interaction with an ER exit signal on the cytoplasmic domain of the cargo protein itself (in the case of transmembrane proteins) or an indirect interaction mediated through a cargo receptor (for soluble cargoes restricted to the ER lumen) [26]. SEC24 has also been implicated in autophagy of the ER (ER-phagy) with SEC24A/B involved in bulk ER-phagy [81] whereas SEC24C is required for selective ER-phagy [82]. We refer readers to other reviews for a more detailed discussion of ER-phagy and crosstalk between the secretory and autophagy pathways [83, 84].

Several cargo binding sites on SEC24 have been mapped [27]. SEC24 is the only COPII component encoded by more than one paralogous gene in yeast (*Sec24, Lst1, and Iss1).* Yeast Sec24p shares 55% and 23% protein sequence identity with Iss1p and Lst1p, respectively, with overexpression of Iss1p, but not Lst1p, sufficient to rescue Sec24p deficiency [85-87]. Mammalian genomes encode four SEC24 paralogs, SEC24A-D, with the SEC24A and SEC24B

subgroup more similar to yeast Sec24p and Iss1p, and the SEC24C and SEC24D subgroup more similar to yeast Lst1p [88, 89]. Analysis of available eukaryotic SEC24 sequences suggests the presence of at least three SEC24 paralogs in the last common eukaryotic ancestor [48]. Human *SEC24* paralogs share ~50% sequence identity within and ~25% sequence identity across subgroups at the amino acid level. All SEC24 paralogs contain a highly conserved C-terminal region and a hypervariable N-terminal segment [90]. Given the role of SEC24 in cargo recruitment, the expansion of COPII paralogs is thought to have been driven to accommodate a greater diversity of secretory cargoes in mammals. Current evidence suggests that SEC24 paralogs within the same subgroup (SEC24A-B versus SEC24C-D) may exhibit largely but not entirely overlapping function, with larger differences in cargo-sorting signal affinity between the two subgroups [63, 91-93].

No human disorders have been associated with mutations in *SEC24A* or *SEC24C*. Compound heterozygosity for loss of function mutations in *SEC24D* has been reported to result in a syndromic form of osteogenesis imperfecta [59]. Though heterozygous *SEC24B* missense variants were reported in 4/163 cases of neural tube defects (NTDs) in one study [94], these and other missense variants are present in unaffected individuals in the Genome Aggregation Database (gnomAD) [95], arguing against a significant association of heterozygous *SEC24B* mutations with NTDs. In contrast, a wide range of phenotypes has been reported in mice genetically engineered to be deficient in each of the four SEC24 paralogs. SEC24A-deficient mice demonstrate normal development and survival with the only identifiable phenotype being moderate hypocholesterolemia due to impaired secretion of PCSK9, a plasma protein that negatively regulates low density lipoprotein (LDL) receptor abundance and, thereby, LDL clearance from circulation [60]. SEC24B-deficient mice exhibit late embryonic lethality at

~E18.5 due to a neural tube closure defect, likely resulting from decreased trafficking of VANGL2, a planar-cell polarity protein [61]. Notably, no NTD phenotype was observed in heterozygous *Sec24b*$^{+/-}$ mice, further arguing against an association between heterozygous mutations in human *SEC24B* and NTD. Ubiquitous loss of murine SEC24C results in early embryonic lethality at ~E7.5 due to abnormal gastrulation and ectoderm development [62] while mice with SEC24C deficiency restricted to neural progenitors demonstrate perinatal lethality and microcephaly due to widespread cell death [96]. Lastly, absence of SEC24D results in early embryonic death at or before the 8-cell stage [97]. A recent proteomic study using an *in vitro* vesicle reconstitution system demonstrate that SEC24C and SEC24D preferentially interact with ERGIC1 whereas SEC24A favors CNIH4 [98], a proposed cargo receptor for G-protein coupled receptors [99]. Several other cargo specific preferences for each SEC24 paralog have also been reported, including preference of the cargo PCSK9 for SEC24A/B [60]; VANGL2 for SEC24B [61]; SERT [100], SLC6A14 [101], and autotaxin [102] for SEC24C; and the GABA transporter 1 (GAT1) [103] for SEC24D. The glycine transporter (GLYT1) has also been demonstrated to physically interact with SEC24D, though it is unclear whether this interaction is exclusive to SEC24D, as interactions with the other SEC24 paralogs were not tested in this study [104].

In contrast to this evidence for cargo specificity, other studies suggest significant overlap in cargo repertoire for SEC24 paralogs, especially between those within the same subgroup. The vesicular stomatitis virus G glycoprotein (VSV-G) has been reported to interact strongly with mammalian SEC24A/B but not SEC24C/D, whereas Syntaxin 5 and membrin are specifically packaged by mammalian SEC24C/D [90]. Similarly, in *Sec24a* null mice, inactivation of an additional *Sec24b* allele results in a further ~25% reduction in plasma cholesterol, consistent with partial overlap in function between murine SEC24A and SEC24B [60]. In contrast, the

human recycling transmembrane protein p24-p23, which acts as a cargo receptor for GPI-anchored CD59, prefers SEC24C/D for ER export [105]. Lastly, replacement of the majority of the *Sec24c* coding sequence with *Sec24d* at the endogenous *Sec24c* locus partially rescues the embryonic lethal *Sec24c* null phenotype, again suggesting significant functional overlap between SEC24C and SEC24D [63]. Taken together, these data demonstrate partial functional overlap between SEC24 paralogs within the same subgroup and divergence between the two subgroups.

1.6.4 SEC31

Heterotetramers of SEC31 and SEC13 form the outer COPII coat [106-108]. The SEC13-SEC31 complex mediates membrane deformation. SEC31 also directly interacts with SEC23 to stimulate its GAP activity, thereby, triggering SAR1 GTP hydrolysis [109]. It has been suggested that SEC31 fine-tunes SEC23 GAP kinetics to accommodate large cargoes such as collagen [110, 111]. Indeed, downregulation of SEC31 as a result of SEC13 depletion in HeLa cells leads to impaired secretion of collagen but not of the temperature sensitive VSV-G-ts045 glycoprotein [112]. VSV-G is a viral protein that is evolutionarily optimized for ER-Golgi export and experimentation with this protein requires overexpression and culturing cells at high temperature for a prolonged period. It is unclear whether deletion of SEC31 would produce the same effect for physiologic cargoes. Mammalian genomes encode two paralogs of SEC31 (*SEC31A* and *SEC31B*) while yeast only contain a single *Sec31* gene. *SEC31A* is highly expressed in most human tissues except for the brain, whereas *SEC31B* is expressed at low levels in most tissues, except for the cerebellum and testis. Human SEC31A shares ~45% sequence identity with SEC31B, with these 2 paralogs sharing ~26% and ~19% sequence similarity, respectively with yeast Sec31p at the amino acid level [113-115]. Human *SEC31B* also appears to be alternatively spliced, producing a C-terminally truncated protein that is ~41% of the full length SEC31B,

though the function of this truncated SEC31B is unclear [115]. No human disorder has been associated with mutations in either *SEC31* paralog, and mouse models have also not yet been reported.

Though there are currently no published data to assess potential functional overlap among vertebrate SEC31 paralogs, there is some evidence for such overlap in plants. The *Arabidopsis* genome also encodes two *SEC31* paralogs, which share ~59% protein sequence identity with each other and ~25% protein sequence identity with their human SEC31 orthologs. *SEC31B* mutant *Arabidopsis* are infertile due to a defect in pollen development [116], whereas *SEC31A* deficient *Arabidopsis* exhibit normal vegetative and reproductive development. Inactivation of both *SEC31A* and *SEC31B* results in lethality due to impaired gametogenesis [117]. *SEC31B* is expressed at ~600-fold higher levels in most plant tissues than *SEC31A*. However, *SEC31B* null transgenic plants in which SEC31A expression is driven by the *SEC31B* promoter exhibit normal fertility [117], demonstrating significant functional overlap between these two plant paralogs, with their evolutionary conservation likely driven by their divergent gene expression programs.

1.6.5 SEC13

SEC13, together with SEC31, forms the outer COPII coat complex. SEC13 also interacts with several proteins of the nuclear pore complex and shuttles between the nucleus and the cytoplasm [118-120]. The single human *SEC13* gene shares ~97% and ~46% sequence identity with mouse and yeast SEC13, respectively. Though *SEC13* mutations are not associated with any known human disorder, mice with complete loss of SEC13 are not viable. However, mice with low levels of residual SEC13 appear grossly normal, though exhibiting aberrant expression of several genes involved in immune response and inflammation [64]. In zebrafish, deletion of *sec13* leads to impaired retinal and gut development associated with a procollagen secretion

defect [121, 122]. Consistent with these observations, depletion of SEC13 in human intestinal epithelial (Caco-2) cells results in aberrant cyst morphogenesis [123].

1.7 Specificity of cargo receptors

As previously noted, cargo receptors are ER transmembrane proteins that bridge the interaction between cargoes in the ER lumen and COPII proteins on the cytoplasmic face of the ER. Several cargo receptors have been described in mammals, including LMAN1 (ERGIC53) and SURF4 [124, 125]. The following discussion will focus on LMAN1 and SURF4, which have been extensively studied *in vitro* and *in vivo*. For a more comprehensive discussion of other putative ER cargo receptors, the reader is referred to [29].

1.7.1 LMAN1

LMAN1 (also known as ERGIC53) is a 53 KDa type I transmembrane protein that was originally identified as a marker for the ERGIC [126]. LMAN1 resides primarily in the ER/ERGIC lumen, with a short C terminal tail of 12 amino acid, including a dilysine diphenylalanine motif (KKFF), extending into the cytoplasm. The FF motif is required for ER export whereas the KK motif is necessary for ER retrieval [127, 128], facilitating LMAN1 cycling between the ER, ERGIC, and cis-Golgi. The LMAN1 luminal segment includes an L-type lectin domain that binds to mannose in a Ca^{2+} dependent manner [129, 130]. LMAN1 was initially suspected to function as a cargo receptor, recruiting secretory proteins in the ER lumen and escorting them to the Golgi, based on its homology to the well characterized yeast cargo receptor, Emp47p [131]. The identification of loss-of-function mutation in *LMAN1* as the cause for the autosomal recessive human bleeding disorder, combined factor V and factor VIII deficiency (F5F8D) [132, 133], identified these two proteins as putative cargoes for LMAN1. F5F8D patients exhibit plasma levels of coagulation factor V (FV) and factor VIII (FVIII)

reduced to 5-30% of normal [134]. Further studies identified other potential LMAN1 cargoes, including the lysosomal proteins cathepsin C (CTSC) [135] and cathepsin Z (CTSZ) [136], membrane protein γ-aminobutyric acid type A receptors (GABAARs) [137], and other secreted proteins including α_1-antitrypsin (A1AT) [138], Mac-2 binding protein (Mac-2BP) [139], and matrix metalloproteinase-9 (MMP-9) [140].

Mutations in *LMAN1* account for ~70% of F5F8D patients [141] with the remaining ~30% of cases due to inactivating mutations in *MCFD2* [142]. Together, mutations in *LMAN1* and *MCFD2* appear to account for all cases of F5F8D [143]. MCFD2 is a 16 kDa soluble EF-hand-containing protein that lacks an ER retrieval motif and is retained in the ER by forming a stable, Ca^{2+} dependent 1:1 stoichiometry complex with LMAN1 [142, 144]. In the absence of LMAN1, MCFD2 is efficiently secreted [145]. For efficient ER exit, dimerization of LMAN1 is required [146]; however, LMAN1 hexamers have also been observed [147, 148].

The recognition motif for LMAN1/MCDF2-dependent cargoes is unclear and the LMAN1 cargo repertoire appears to be limited [149]. The few known LMAN1/MCFD2 cargoes have been identified by various approaches including overexpression of an ER-trapped LMAN1 mutant in HeLa cells (CTSZ [135]), fluorescent-based protein fragment complementation assays (CTSC, A1AT, Mac-2BP, and MMP-9 [136, 138-140]), or mass spectrometry following coimmunoprecipitation (GABAARs [137]). Most of these putative cargoes are heavily glycosylated, with LMAN1 binding shown to be carbohydrate dependent for CTSC, CTSZ, A1AT, Mac-2BP, and MMP-9 [124, 136, 138-140], though not required for FVIII or GABAARs [137, 144]. In the absence of LMAN1, MCFD2 can still bind to FVIII; however, it is unclear whether LMAN1 can independently interact with FVIII in the absence of MCFD2 [144]. MCFD2 is required for efficient secretion of FV, FVIII, A1AT, and Mac-2BP [139, 144, 150]

but is dispensable for CTSC and CTSZ [145]. Dependence of FV, FVIII, and A1AT on LMAN1 has been confirmed *in vivo*, with *Lman1*-deficient mice exhibiting reduced plasma levels of FV and FVIII and ER accumulation of A1AT in hepatocytes. However, similar accumulation of CTSC and CTSZ was not observed [151].

1.7.2 SURF4

SURF4 is a 29 kDa protein with 5 transmembrane domains localizing to the ER and ERGIC [152]. Similar to LMAN1, SURF4 also displays a dilysine ER retrieval motif at its cytoplasm-facing C-terminus, facilitating its retention in the ER [153]. Erv29p, the SURF4 ortholog in yeast, is the well characterized cargo receptor for pro-α-factor, carboxypeptidase Y, and proteinase A [154-156]. SURF4 and Erv29p are highly conserved across eukaryotes, with orthologs also identified in *C. elegans* and *Drosophila* [157]. While Erv29p is dispensable in yeast, SFT-4 (the *C. elegans* ortholog of *SURF4*) and SURF4 are required for survival in *C. elegans* and mice, respectively [158, 159]. SURF4 was show to function in conjunction with LMAN1 to maintain ERGIC and Golgi structural integrity [152], and based on its cellular localization and sequence homology to Erv29p, was long suspected to function as a cargo receptor in mammalian cells, though putative cargoes have only recently been identified.

No human disorder has been associated with mutations in *SURF4.* However, genetic polymorphism at the *SURF4* gene is strongly associate with reduced plasma lipid levels and reduced cardiovascular risk in human populations [160]. As noted above, efficient secretion of PCSK9 is specifically dependent on SEC24A, with their localization to opposite sides of the ER membrane implying a requirement for a specific cargo receptor serving as the physical link [60]. A whole genome CRISPR screen in human embryonic kidney cells (HEK293T) identified SURF4 as this putative PCSK9 cargo receptor [125]. A recent report confirmed PCSK9 secretion

dependence on both SEC24A and SURF4 in HEK293T and human hepatic (HuH7) cells. In addition, the authors demonstrated that disruption of SEC24A-SURF4 binding by a small molecule, 4-phenylbutyrate, inhibits PCSK9 secretion, further supporting the role for SURF4 as a cargo receptor that bridges PCSK9 and SEC24A [161]. SFT-4 was also shown to be essential for the trafficking of the yolk protein VIT-2, a component of *C. elegans* lipoproteins that share ~22% sequence identity with human apolipoprotein B (APOB); this same group demonstrating that SURF4 also serves as a cargo receptor for APOB in human hepatic HepG2 cells [162]. Several other SURF4 cargoes have been identified, including growth hormone, dentin sialophosphoprotein (DSPP), and amelogenin [163], erythropoietin [164], pathogenic A1AT polymers [165], sonic hedgehog [166], proinsulin [167], and the lysosomal proteins progranulin and prosaposin [168].

Further analysis of multiple potential SURF4 cargoes across multiple vertebrate species led to the proposal for a highly conserved hydrophobic-proline-hydrophobic tripeptide (ER-ESCAPE) motif downstream of the signal peptide for many putative SURF4 cargoes [163]. Deletion of *SURF4* in human cells significantly reduced the secretion of some proteins with this ER-ESCAPE motif [161]. However, several putative SURF4 cargoes noted above do not carry this ER-ESCAPE motif, suggesting the presence of other recognition motifs that have not yet been identified. Proteomic analysis of conditioned media collected from SURF4 deficient cells revealed several SURF4 clients in HEK293T and HuH7 cells, half of which carry the ER-ESCAPE motif or a Cardin-Weintraub motif, which has also been shown to mediate interaction with SURF4 [161, 166]. Interestingly, while replacement of the ER-ESCAPE tripeptide motif with glutamic acids (EEE) significantly impairs secretion of several putative SURF4 cargoes [161, 163], APOB – arguably the most well studied SURF4 cargo in both *in vitro* and *in vivo*

models, carries the EEE motif downstream of the signal peptide. Taken together, these data suggest a more complex process governing cargo recognition by SURF4.

The above reports suggest that SURF4 may play a role in the trafficking of a much broader range of secretory cargoes compared to LMAN1. Proteomic analysis of *in vitro* reconstituted COPII vesicles are also consistent with this model, with SURF4 deletion resulting in the depletion of many more proteins from reconstituted COPII vesicles compared to LMAN1 deletion [149]. A recent study demonstrated that SURF4 traffics ER cargoes into an elongated tubular ERGIC that lacks LMAN1. This tubular ERGIC accelerates ER-to-Golgi trafficking of SURF4 cargoes, suggesting a distinct SURF4 trafficking route [169].

1.8 Concluding remarks

Although the conventional ER-Golgi secretory pathway was first identified over five decades ago, a number of key questions remain unanswered. Are the physical carriers that transport secreted proteins from the ER to the Golgi discrete vesicles, elongated tubules, or continuous tunnels that physically connect the two compartments? Early studies by Schekman and colleagues in yeast established the widely accepted COPII vesicle model. However, recent advances in super-resolution live-cell imaging demonstrate the presence of elongated tubules originating from the ERES, with the COPII proteins appearing to function as gatekeepers at the ERES collar rather than as escorts that travel with cargoes to the Golgi. While superficially conflicting, these models are not mutually exclusive, and multiple different pathways may exist, with the path that each specific protein takes depending on its size, relative abundance in the ER lumen, and requirement for a cargo receptor to facilitate its secretion.

Next, although a diverse set of secretory cargoes are transported from ER to Golgi, it remains unclear as to how many require a cargo receptor to facilitate this process. The thousands

of known secreted proteins and the very limited number of putative cargo receptors identified to date are suggestive of a role for a passive bulk flow mechanism, with only a small subset of secreted proteins requiring interaction with a cargo receptor. Alternatively, multiple additional cargo receptors may remain to be identified. Furthermore, even for the small subset of proteins clearly demonstrated to be dependent on a specific cargo receptor for secretion, deletion of the corresponding cargo receptor does not result in complete blockade of secretion, suggesting the presence of an alternative "back up" cargo receptor(s) or another mechanism for basal protein transport, with the cargo receptor only required for maximal secretion efficiency.

Finally, a better understanding of the functional conservation of COPII paralogs might have translational relevance for human diseases resulting from deficiency of a specific COPII protein paralog, such as CMRD (SAR1B) or CDAII (SEC23B). For COPII paralogs exhibiting extensive or complete functional overlap, therapies that upregulate expression of one paralog could potentially compensate for genetic deficiency of the other [79]. In addition, future insights into cargo-cargo receptor interactions and binding motifs could facilitate fine tuning of secretion for a specific cargo protein. Lastly, optimizing cargo/cargo receptor interactions could potentially be leveraged to improve efficiency of recombinant protein production.

Chapter 2

Overlap in Function between the SAR1A and SAR1B Paralogs *in vivo*

2.1 Abstract

Proteins carrying a signal peptide or a transmembrane domain typically traverse the intracellular secretory pathway before reaching the extracellular space, the plasma membrane, or other intracellular destinations. Transport of these proteins from the endoplasmic reticulum (ER) to the Golgi apparatus is mediated by COPII vesicles/tubules. SAR1 is the component of the COPII coat that initiates coat assembly by recruiting other coat proteins to the ER membrane. There are two SAR1 paralogs in mammals, SAR1A and SAR1B, which exhibit ~90% amino acid sequence identity. Consistent with previous reports, we observe that germline deletion of *Sar1b* results in perinatal lethality in mice. In contrast, we demonstrate that genetic deficiency for *Sar1a* results in mid-embryogenic lethality. Deletion of *Sar1b* restricted to hepatocytes is compatible with survival, though results in hypocholesterolemia that is rescued by adenovirus-mediated overexpression of either SAR1A or SAR1B. To test whether SAR1A can compensate for complete loss of SAR1B, we generated mice with the *Sar1a* coding sequence replacing *Sar1b* at the endogenous *Sar1b* locus. The resulting mice survive to adulthood and are phenotypically normal, demonstrating extensive or complete overlap in function between the two SAR1 protein paralogs in mice. These data suggest that the *SAR1A* gene could be therapeutically upregulated to treat chylomicron retention disease caused by SAR1B deficiency in humans.

2.2 Introduction

The intracellular secretory pathway transports proteins from the endoplasmic reticulum (ER) to the Golgi for subsequent secretion into the extracellular space, insertion into the plasma membrane, or transport to various intracellular organelles including lysosomes and endosomes. [170]. Following proper folding in the ER, cargo proteins are transported to the Golgi apparatus via coat protein complex II (COPII) vesicles/tubules [29]. The COPII coat is a highly conserved structure consisting of five core proteins that are present in all eukaryotes: SAR1, SEC23, SEC24, SEC13, and SEC31 [12, 35]. COPII coat formation begins with GTP-bound SAR1 recruiting the SEC23-SEC24 heterodimer complex to ER exit sites (ERES) on the cytoplasmic face of the ER membrane via direct interaction between SAR1 and SEC23. Following cargo recruitment and concentration in the ER lumen mediated by SEC24, SEC13 and SEC31 are recruited to form the outer coat and complete coat assembly [14-16, 18, 85, 171, 172].

Mammalian genomes contain multiple COPII paralogs as a result of gene duplication events, including two paralogs each for SAR1, SEC23, SEC31, and four paralogs for SEC24 [47]. The expansion of COPII paralogs through the course of eukaryotic evolution is thought to have accommodated increasing secretory demand. This hypothesis is further supported by observations that humans and mice deficient in one paralog exhibit distinct phenotypes from those deficient in the other [75, 172]. However, recent studies have shown that SEC23A can completely rescue the perinatal lethality exhibit by *Sec23b* null mice whereas SEC24D can only partially rescue the early embryonic lethality demonstrated by *Sec24c* null mice, allowing them to survive up to term though these neonates died shortly after birth [63, 74]. Together, these data demonstrate some degree of functional overlap between COPII paralogs in mammals though the extent of the overlap varies between SEC23A/B and SEC24C/D.

GDP-bound SAR1 is recruited to the ERES by membrane-bound SEC12 to initiate coat assembly. SEC12 also functions as the guanine exchange factor (GEF) for GDP-SAR1 [15], with GTP-bound SAR1 subsequently recruiting other coat proteins to the ER membrane to complete coat assembly. SEC23 acts as a guanine activating factor (GAP) for SAR1, resulting in GTP hydrolysis [16, 18, 171]. Most invertebrate genomes encode a single SAR1 paralog whereas most mammalian genomes encode two (*SAR1A* and *SAR1B*). The human SAR1A and SAR1B paralogs differ by only 20 out of 198 amino acids (**Figure 2-1**) and are both ubiquitously expressed in human and mouse tissues [173, 174]. Mutations in *SAR1B* are associated with the rare autosomal recessive disorder chylomicron retention disease (CMRD, or Anderson's disease) [50, 51, 175], whereas a disorder due to human mutations in *SAR1A* has not been reported. Melville and colleagues identified three amino acid clusters differing between SAR1A and SAR1B that alter GTPase kinetics and SEC23 affinity [69], suggesting unique biological functions for these 2 paralogs *in vivo*. Genetic deletion of either *SAR1A* or *SAR1B* in Caco-2/15 cells results in decreased chylomicron secretion with a less dramatic phenotype observed in SAR1A null cells. Combined inactivation of both paralogs is required to recapitulate the severe phenotype seen in humans with CMRD, suggesting synergism between the paralogs [70, 176].



Figure 2-1: Alignment of the mouse SAR1A and SAR1B protein sequences.

Germline deletion of *Sar1b* in mice results in late-gestation lethality in homozygous null mice, with haploinsufficient animals demonstrating a defect in chylomicron secretion similar to that observed in homozygous or compound heterozygous CRMD patients [177]. Liver specific deletion of *Sar1b* results in severe hypocholesterolemia due to a lipoprotein secretion defect [160]. To date, no mouse models for SAR1A deficiency have been reported.

We now report analysis of mice genetically deficient in *Sar1a* or *Sar1b* as well mice carrying a modified *Sar1b* locus at which the *Sar1a* coding sequence has been substituted for *Sar1b*. Our results demonstrate near complete overlap in functions between the SAR1A and SAR1B paralogous proteins.

2.3 Methods

2.3.1 Animal care

All animal care and use complied with the Principles of Laboratory and Animal Care established by the National Society for Medical Research. Mice were housed in a controlled lighting (12h light/dark cycle) and temperature (22°C) environment and had free access to food (5L0D, LabDiet, St. Louis, MO) and water. All animal protocols were approved by the Institutional Animal Care and Use Committee (IACUC) of the University of Michigan (protocol number PRO00011038). Both male and female mice were used in this study.

2.3.2 Generation of Sar1a$^{+/gt}$ mice

The Bay Genomics ES cell clone CSH949 [178] was obtained from the Mutant Mouse Resource & Research Centers (MMRRC). This clone is heterozygous for the *Sar1a^{gt}* allele in which a gene trap cassette is inserted into *Sar1a* gene intron 5 (**Figure 2-2A**). ES cell culture, expansion, and microinjection to generate ES cell chimera mice were performed as previously

described [62, 97]. Chimeric mice were bred with C57BL/6J (0006640, Jackson Laboratory, Bar Harbor ME) to obtain germ-line transmission. Mice carrying the *Sar1a*^{gt} allele were maintained by continuous backcrossing to C57BL/6J.

2.3.3 Generation of Sar1b$^{+/-}$ *mice*

Mice carrying the *Sar1b*$^{tm1a(EUCOMM)Wtsi}$ allele [179] were obtained from the European Conditional Mouse Mutagenesis (EUCOMM) program (Wellcome Trust Sanger Institute, Cambridge, UK). These mice carrying a conditional gene trap (*Sar1b*^{cgt}) allele in which the gene trap cassette flanked by FRT sites followed by a loxP site is inserted into intron 4 of the *Sar1b* gene. An additional loxP site is inserted downstream of exon 5 of *Sar1b*. Mice carrying the *Sar1b*^{cgt} allele were maintained by continuous backcrossing to C57BL/6J mice.

To generate the conditional allele, *Sar1b*$^{cgt/+}$ mice were crossed with mice carrying the *Flp1* recombinase gene under the control of *Actb* promoter (005703, Jackson Laboratory, Bar Harbor ME). Resulting mice carrying the conditional allele (*Sar1b*$^{fl/+}$) were then bred with mice carrying an *EIIa-Cre* transgene (003724, Jackson Laboratory, Bar Harbor ME) to generate the *Sar1b* null allele (*Sar1b*$^{-}$). *Sar1b*$^{fl/+}$ and *Sar1b*$^{+/-}$ mice were maintained by continuous backcrossing to C57BL/6J mice.

2.3.4 Generation of Sar1b$^{fl/+}$ *Alb-Cre*$^{+}$ *mice*

To generate *Sar1b*$^{fl/+}$ *Alb-Cre*$^{+}$ mice, we crossed *Sar1b*$^{fl/+}$ mice generated above with mice carrying an *Alb-Cre* transgene [62]. *Sar1b*$^{fl/+}$ *Alb-Cre*$^{+}$ mice were maintained by continuous backcrossing to C57BL/6J mice. *Sar1b*$^{fl/fl}$ *Alb-Cre*$^{+}$ mice were obtained by crossing *Sar1b*$^{fl/+}$ *Alb-Cre*$^{+}$ mice with *Sar1b*$^{fl/+}$ or *Sar1b*$^{fl/fl}$ mice.

2.3.5 Generation of Sar1b$^{+/a}$ *mice*

Mice carrying the *Sar1b*^{a} allele, in which the mouse *Sar1a* coding sequence is inserted into the *Sar1b* locus immediately downstream of the ATG codon in exon 2 of the gene, were generated by Biocytogen Boston Corp. (Wakefield MA) using CRISPR/Cas9 (**Figure 2-5A**). Cas9 and sgRNA targeting exon 2 of *Sar1b* along with the DNA repair template were injected into C57BL/6N zygotes. Founder mice with confirmed germ-line transmission were bred to C57BL/6N mice to obtain F1 (*Sar1b*$^{+/a}$) mice. *Sar1b*$^{+/a}$ mice were backcrossed to C57BL/6J mice for at least 3 generations before experimentation. *Sar1b*$^{+/a}$ mice were maintained by continuous backcrossing to C57BL/6J mice.

2.3.6 Mouse genotyping

Tail clips were obtained from 2-3 weeks old mice for genomic isolation and genotyping. PCR was performed using Go-Taq Green Master Mix (Promega, Madison, WI) and resulting products were resolved by 3% agarose gel electrophoresis. All primers used for genotyping are listed in **Table 2-6.**

2.3.7 Timed mating and embryo collection

Timed mating were performed by intercrossing heterozygous mice (*Sar1*$^{gt/+}$ or *Sar1b*$^{+/-}$). Embryos were collected at various time points, including E10.5, E11.0, E11.5, E18.5 for genotyping and histologic analyses. Genomic DNA was isolated from the embryonic yolk sacs of E10.5-E11.5 embryos or tail clips of E18.5 embryos. Genomic DNA was isolated and genotyped as described above. Embryos were immediately fixed in Z-fix solution (Anatech Ltd, Battle Creek MI).

2.3.8 Blood and tissue collection

Blood and tissue were collected and sera separated as previously described [180]. Tissues were immediately fixed in Z-fix solution or frozen in liquid nitrogen. Sera and tissues were stored at -80°C until experimentation.

2.3.9 Histologic analyses

Tissue processing, embedding, sectioning, hematoxylin and eosin (H&E) staining were performed at the University of Michigan In-Vivo Animal Core (IVAC). Slides were reviewed by investigators and a veterinarian pathologist blinded to genotypes.

2.3.10 Cholesterol assays

Total cholesterol in sera was determined using a colorimetric assay (SB-1010-225, Fisher Scientific, Hampton NH) according to the manufacturer's instructions.

2.3.11 Adenovirus mediated expression of SAR1A or SAR1B

Adenovirus (AV) expressing GFP, mouse SAR1A, or mouse SAR1B was produced by the University of Michigan Vector Core. AV was injected into *Sar1b*$^{fl/fl}$ *Alb-Cre*$^{+}$ mice by tail vein injection. Blood was collected on the day of injection, day 3, and day 7 post-injection by retro-orbital eye bleed and assayed for cholesterol levels as described above.

2.4 Results

2.4.1 Germline deletion of Sar1a in mice results in mid-embryogenic lethality

To investigate SAR1A function *in vivo,* we generated mice carrying a *Sar1a* targeted allele in which *Sar1a* expression is disrupted by insertion of a gene trap cassette into intron 5 [178] (**Figure 2-2A**). We designed a three-primer PCR assay to differentiate between the wild type and gene trap allele, with DNA sequence analysis confirming the expected configuration

(**Figure 2-2B**). Genotyping results for offspring from a heterozygous *Sar1a*$^{gt/+}$ intercross are shown in Error! Reference source not found.. No *Sar1a*$^{gt/gt}$ mice were identified among the 65 mice genotyped at weaning (P21). The results of genotyping for embryos harvested at embryonic ages of E10.5, E11, and E11.5 post coitus (pc) are shown in **Figure 2-2C** and Error! Reference source not found.. Though the expected number of *Sar1a*$^{gt/gt}$ embryos were observed at E10.5, no viable *Sar1a*$^{gt/gt}$ embryos remained by E11.5. Histologic evaluation of E10.5 embryos failed to identify any obvious abnormalities in the *Sar1a*$^{gt/gt}$ embryos (**Figure 2-2F**).

Table 2-1: Genotype distribution of *Sar1a*$^{gt/+}$ intercross

Mating	***Sar1a*$^{gt/+}$ X *Sar1a*$^{gt/+}$**			
Genotype	*Sar1a*$^{+/+}$	*Sar1a*$^{gt/+}$	*Sar1a*$^{gt/gt}$	
Expected	25%	50%	25%	**p-value (χ^2 test)**
E10.5	10 (28%)	16 (44%)	10 (28%)	0.800
E11.0	5 (26%)	10 (53%)	4 (21%)	0.924
E11.5	7 (29%)	15 (63%)	2[t] (8%)	0.167
Weaning (P21)	17 (26%)	48 (74%)	0 (0%)	7.22 x 10^{-6}

[t]Embryos appeared dead and were partially resorbed.

Heterozygous *Sar1a*$^{gt/+}$ mice were present at the expected Mendelian ratio at weaning and exhibit normal growth and fertility (**Figure 2-2C**). At 8 weeks of age, no significant differences in body mass and plasma cholesterol levels were observed between *Sar1a*$^{+/+}$ and *Sar1a*$^{gt/+}$ mice (**Figure 2-2D-E**). This is consistent with the presence of several individuals carrying heterozygous loss-of-function mutations in *SAR1A* in gnomAD [95], suggesting that SAR1A haploinsufficiency is also well tolerated in humans.

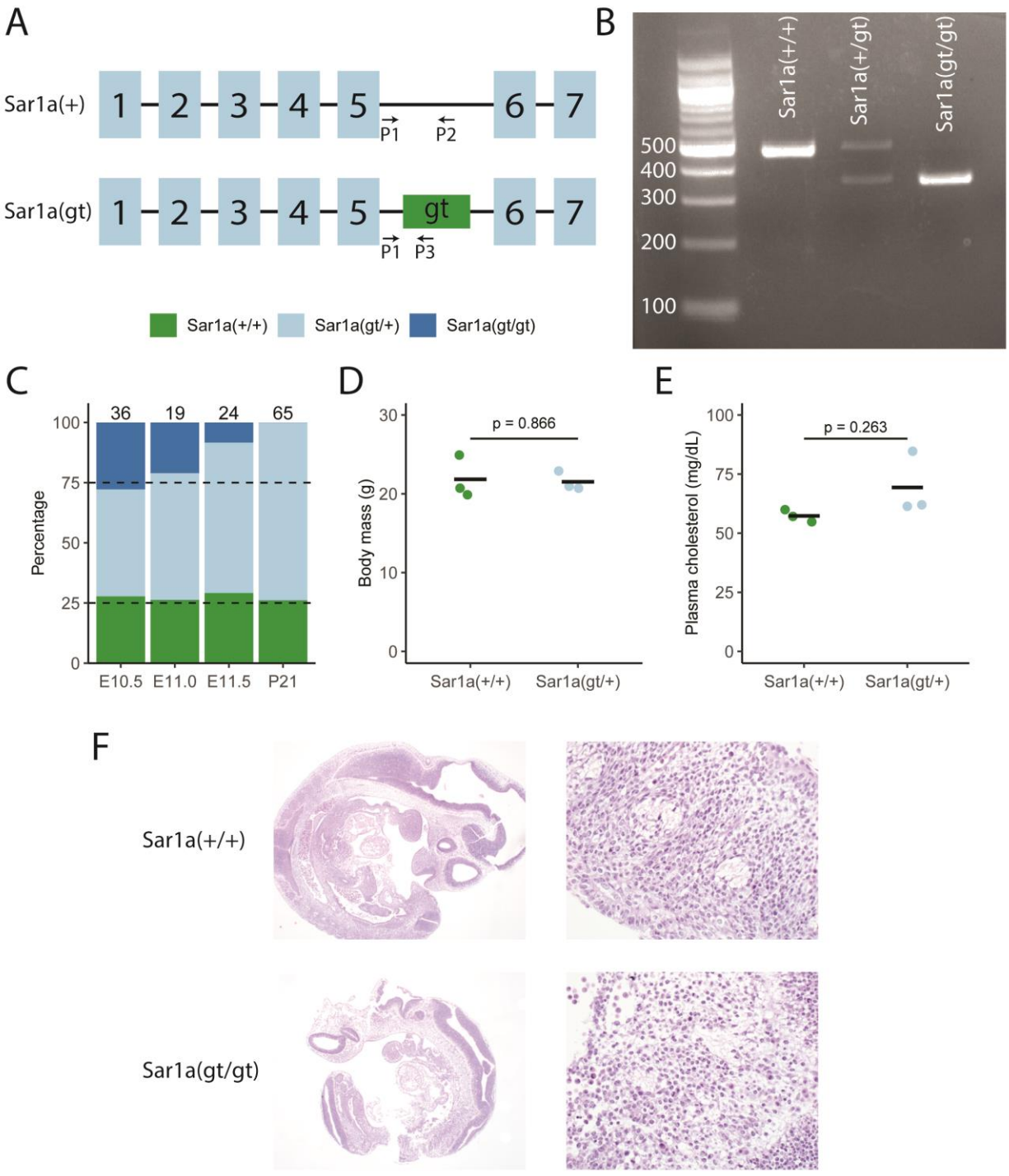

Figure 2-2: Complete loss of SAR1A results in mid-embryogenesis lethality. (A) Schematic of the wild type *Sar1a(+)* and gene trap *Sar1a(gt)* allele. The gene trap cassette is inserted into intron 5 of the *Sar1a* gene. P1, P2, P3 denote the binding sites of genotyping primers. Rectangles represent exons and solid line segments represent introns. Exons and introns are not drawn to scale. **(B)** Genotyping assay using primers denoted in (A). The *Sar1a(+)* allele produce a PCR amplicon of 489 bp whereas the *Sar1a(gt)* allele produce an amplicon of 352 bp. **(C)** Genotype distribution of offspring from $Sar1a^{gt/+}$ X $Sar1a^{gt/+}$ intercrosses at various developmental time points. Number at the top of each bar represents the total number of embryo/mice in that group. **(D-E)** Body mass and plasma cholesterol

levels of wild type and heterozygous *Sar1a*$^{gt/+}$ mice. Statistical significance was determined by two-sided t-test. **(F)** H&E staining of E10.5 wild type and *Sar1a*$^{gt/gt}$ embryos.

2.4.2 Germline homozygous Sar1b deletion results in perinatal lethality

We generated heterozygous *Sar1b* deficient mice (*Sar1b*$^{+/-}$) by crossing mice carrying a conditional *Sar1b* allele (*Sar1b*$^{fl/fl}$) with mice carrying a ubiquitously expressed Cre (*EIIa-Cre*) transgene. Cre-mediated DNA recombination resulted in the excision of exon 5 of the *Sar1b* gene (**Figure 2-3A**), with a three-primer PCR assay clearly differentiating the wild type and null alleles (**Figure 2-3B**). The results of heterozygous *Sar1b*$^{+/-}$ intercrosses are shown in **Figure 2-3C** and **Table 2-2**. Though no homozygous *Sar1b*$^{-/-}$ mice were observed at weaning (P21, among 83 mice genotyped), the expected number of *Sar1b*$^{-/-}$ embryos were present at E11.5 and E18.5 (**Figure 2-3C**). Most *Sar1b*$^{-/-}$ embryos appear to die around the time of birth with no viable *Sar1b*$^{-/-}$ neonates remaining at P1 (**Figure 2-3C**). Histologic analysis of E18.5 embryos failed to identify any obvious abnormalities in *Sar1b*$^{-/-}$ embryos compared to *Sar1b*$^{+/+}$ littermates.

Heterozygous *Sar1b*$^{+/-}$ mice were observed at the expected numbers at P21 and exhibited normal growth and fertility (**Figure 2-3C**). At 3-4 months of age, no differences in body mass or plasma cholesterol levels were observed between *Sar1b*$^{+/-}$ and *Sar1b*$^{+/+}$ mice (**Figure 2-3D-E**).

Table 2-2: Genotype distribution of *Sar1b*$^{+/-}$ intercross

Mating	***Sar1b*$^{+/-}$ X *Sar1b*$^{+/-}$**			
Genotype	*Sar1b*$^{+/+}$	*Sar1b*$^{+/-}$	*Sar1b*$^{-/-}$	
Expected	25%	50%	25%	**p-value (χ^2 test)**
E11.5	3 (19%)	8 (50%)	4 (31%)	0.779
E18.5	14 (26%)	24 (44%)	16 (30%)	0.665
P0	4 (31%)	7 (54%)	2 (15%)	0.707
P1	4 (57%)	3 (43%)	0 (0%)	0.095
Weaning (P21)	29 (35%)	54 (65%)	0 (0%)	9.21×10^{-7}

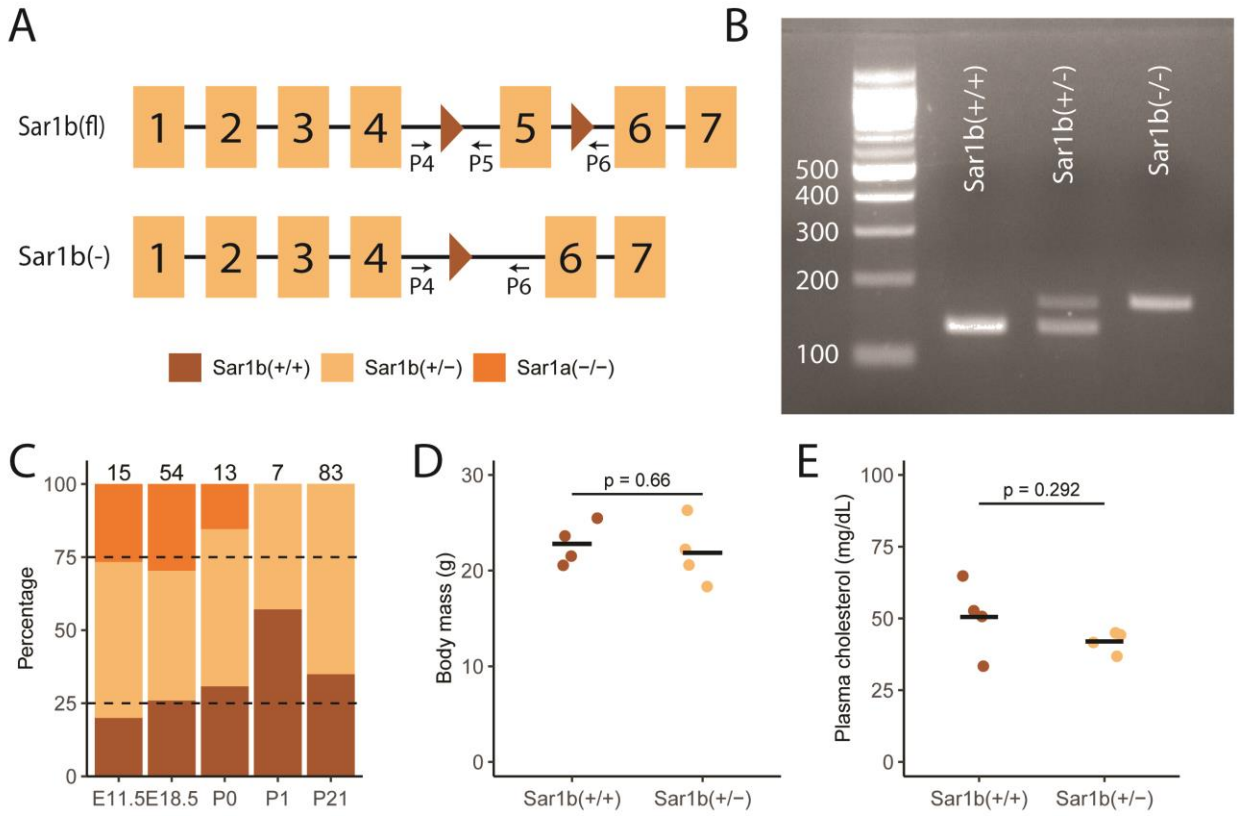

Figure 2-3: Germline deletion of *Sar1b* results in perinatal death. (A) Schematic of the *Sar1b(fl)* and the null *Sar1b(-)* allele. Exon 5 of the *Sar1b* gene is flanked by two floxP sites (triangles), which undergo recombination in the presence of the Cre recombinase leading to excision of exon 5 in the *Sar1b(-)* allele. P4, P5, P6 denote the binding sites of genotyping primers. Rectangles represent exons and solid line segments represent introns. Exons and introns are not drawn to scale. **(B)** Genotyping assay using primers denoted in (A). The *Sar1b(+)* allele generates a PCR product of 136 bp whereas the *Sar1b(-)* allele produce a PCR amplicon of 169 bp. **(C)** Genotype distribution of offspring from $Sar1b^{+/-}$ X $Sar1b^{+/-}$ intercrosses at various development times. Numbers at the top of each bar represent the total number of embryo/mice in that group. **(D-E)** Body mass and plasma cholesterol levels of wild type and heterozygous $Sar1b^{+/-}$ mice. Statistical significance was determined by two-sided t-test.

2.4.3 Hypocholesterolemia resulting from hepatic SAR1B *deficiency is rescued by overexpression of either* SAR1A *or* SAR1B

We next inactivated *Sar1b* in hepatocytes by combining an *Alb-Cre* transgene and the $Sar1b^{fl}$ allele. $Sar1b^{fl/fl}$ *Alb-Cre*$^{+}$ mice were viable (**Figure 2-4A** and **Table 2-3**), and exhibited marked reductions in plasma cholesterol levels to approximately 26% of those in littermate controls **(Figure 2-4B**). No reduction in cholesterol was observed in mice with hepatic

haploinsufficiency for *Sar1b* (*Sar1b*$^{fl/+}$ *Alb-Cre*$^{+}$), consistent with the normal plasma cholesterol level of *Sar1b*$^{+/-}$ mice (**Figure 2-4E**).

To examine potential overlap in function between SAR1A and SAR1B, we next tested the capacity of an adenovirus (AV) expressing either SAR1A, SAR1B, or GFP (as a control) to rescue the hypocholesterolemia of *Sar1b*$^{fl/fl}$ *Alb-Cre*$^{+}$ mice (**Figure 2-4C**). AV was injected by tail vein into *Sar1b*$^{fl/fl}$ *Alb-Cre*$^{+}$ mice and plasma cholesterol levels determined on day 0, 3, and 7 post-injection. Injection of the *Sar1b* AV resulted in a rapid increase in plasma cholesterol to wild type levels by day 3, with minimal increase observed with the control GFP AV, demonstrating efficacy of the AV delivery and efficient expression of *Sar1b* in the targeted hepatocytes. Injection of the *Sar1a* AV resulted in a similar rescue of the hypocholesterolemia phenotype, indistinguishable from that observed with the *Sar1b* AV (**Figure 2-4D**). These data demonstrate that SAR1A can compensate for SAR1B deficiency in this hepatocyte phenotype, suggesting significant overlap in function between the SAR1 paralogs, at least in hepatocytes.

Table 2-3: Genotype distribution of *Sar1b*$^{fl/+}$ *Alb-Cre*$^{+}$ intercross

Mating	***Sar1b*$^{fl/+}$ *Alb-Cre*$^{-}$ X *Sar1b*$^{fl/+}$ *Alb-Cre*$^{+}$**			
Genotype	Other[ǂ]	*Sar1b*$^{fl/+}$ *Alb-Cre*$^{+}$	*Sar1b*$^{fl/+}$ *Alb-Cre*$^{+}$	
Expected	62.5%	25%	12.5%	**p-value (χ^2 test)**
Weaning (P21)	26 (65%)	10 (25%)	4 (10%)	0.887

[ǂ]Other genotypes include *Sar1b*$^{+/+}$ *Alb-Cre*$^{+}$ (12.5%), *Sar1b*$^{+/+}$ *Alb-Cre*$^{-}$(12.5%), *Sar1b*$^{fl/+}$ *Alb-Cre*$^{-}$ (25%), and *Sar1b*$^{fl/fl}$ *Alb-Cre*$^{-}$ (12.5%)

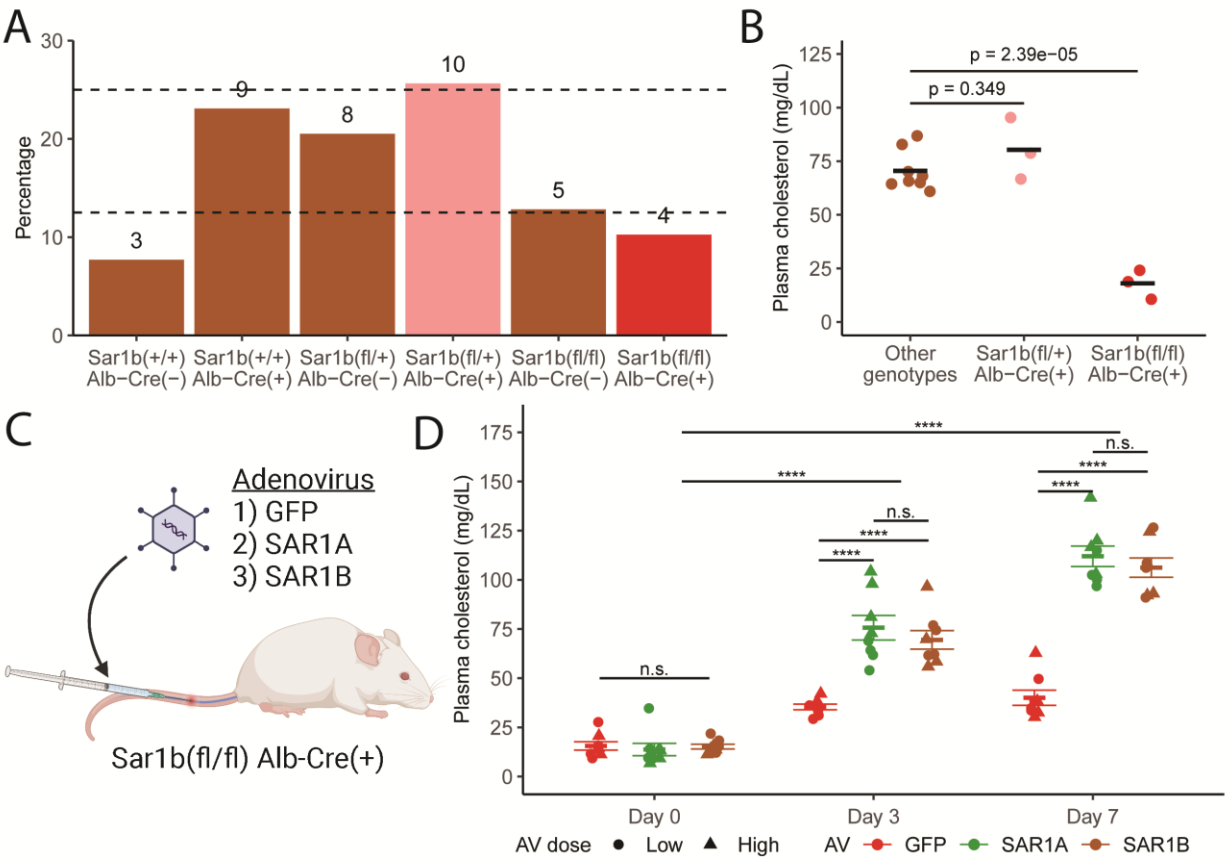

Figure 2-4: Hypocholesterolemia in hepatic SAR1B deficient mice can be rescued by overexpression of either SAR1A or SAR1B. (A) Genotype distribution of offspring from *Sar1b*$^{fl/+}$ *Alb-Cre*$^{+}$ X *Sar1b*$^{fl/+}$ intercrosses at weaning. Numbers at the top of each bar represent the total number of mice with that genotype. **(B)** Plasma cholesterol levels in other genotypes (*Sar1b*$^{fl/fl}$ *Alb-Cre*$^{+}$, *Sar1b*$^{fl/fl}$ *Alb-Cre*$^{-}$, *Sar1b*$^{fl/+}$ *Alb-Cre*$^{-}$, and *Sar1b*$^{fl/fl}$ *Alb-Cre*$^{-}$), *Sar1b*$^{fl/+}$ *Alb-Cre*$^{+}$, and *Sar1b*$^{fl/fl}$ *Alb-Cre*$^{+}$ mice. Statistical significance was determined by One-way ANOVA test followed by Tukey's post-hoc test. **(C)** Schematic of SAR1B deficiency rescue experiment. *Sar1b*$^{fl/fl}$ *Alb-Cre*$^{+}$ mice were injected with adenovirus (AV) expressing GFP, SAR1A, or SAR1B via the tail vein. Blood was sampled on day 0, 3, and 7 post-injection and assayed for cholesterol levels. **(D)** Plasma cholesterol levels of mice injected with AV as illustrated in (C). Two different AV doses were used. Statistical significance was determined by Two-way ANOVA test with interaction between AV type, time, and dose followed by Tukey's post-hoc test. **** indicates p-values of < 0.01; n.s., not significant. p-values for all comparisons are listed in **Table 2-5**.

2.4.4 Expression of Sar1a coding sequence at the Sar1b endogenous locus rescues the lethal Sar1b null phenotype in mice

To examine the potential for overlap in function between SAR1A and SAR1B more globally, we generated a *Sar1b*^{a} allele in which the *Sar1b* coding sequence is replaced with that of *Sar1a* at the *Sar1b* endogenous locus, beginning immediately downstream of the ATG codon in exon 2 of *Sar1b* (**Figure 2-5A**). Mice homozygous for this allele should be unable to express

the SAR1B protein, with SAR1A expressed under the transcriptional control of both the *Sar1a* and *Sar1b* endogenous regulatory elements. We designed a three-primer PCR genotyping assay to differentiate the wild type and *Sar1b*^{a} allele (**Figure 2-5B**).

As shown in **Figure 2-5C** and **Table 2-4**, intercrosses between *Sar1b*$^{+/a}$ mice generated homozygous *Sar1b*$^{a/a}$ mice that are present at weaning though at a lower percentage than the expected Mendelian ratio. Nevertheless, the surviving *Sar1b*$^{a/a}$ mice demonstrate normal growth (**Figure 2-5D**) and fertility and produce *Sar1b*$^{a/a}$ offspring at the expected ratio in *Sar1b*$^{+/a}$ x *Sar1b*$^{a/a}$ intercrosses (**Figure 2-5C**). Histologic examination of multiple organs did not identify any abnormalities in *Sar1b*$^{a/a}$ mice. *Sar1b*$^{a/a}$ mice also demonstrated plasma cholesterol levels indistinguishable from wild type *Sar1b*$^{+/+}$ and heterozygous *Sar1b*$^{+/a}$ mice (**Figure 2-5E**). These data demonstrate an unexpectedly high degree of functional overlap between the SAR1A and SAR1B paralogs *in vivo*.

Table 2-4: Genotype distribution of *Sar1b*$^{+/a}$ intercross

Mating	***Sar1b*$^{+/a}$ X *Sar1b*$^{+/a}$**			
Genotype	*Sar1b*$^{+/+}$	*Sar1b*$^{+/a}$	*Sar1b*$^{a/a}$	
Expected	25%	50%	25%	p-value (χ^2 test)
Weaning	78 (30%)	144 (55%)	41 (15%)	0.002
Mating	***Sar1b*$^{+/a}$ X *Sar1b*$^{a/a}$**			
Genotype	*Sar1b*$^{+/a}$		*Sar1b*$^{a/a}$	
Expected	50%		50%	
Weaning	13 (48%)		14 (52%)	0.847

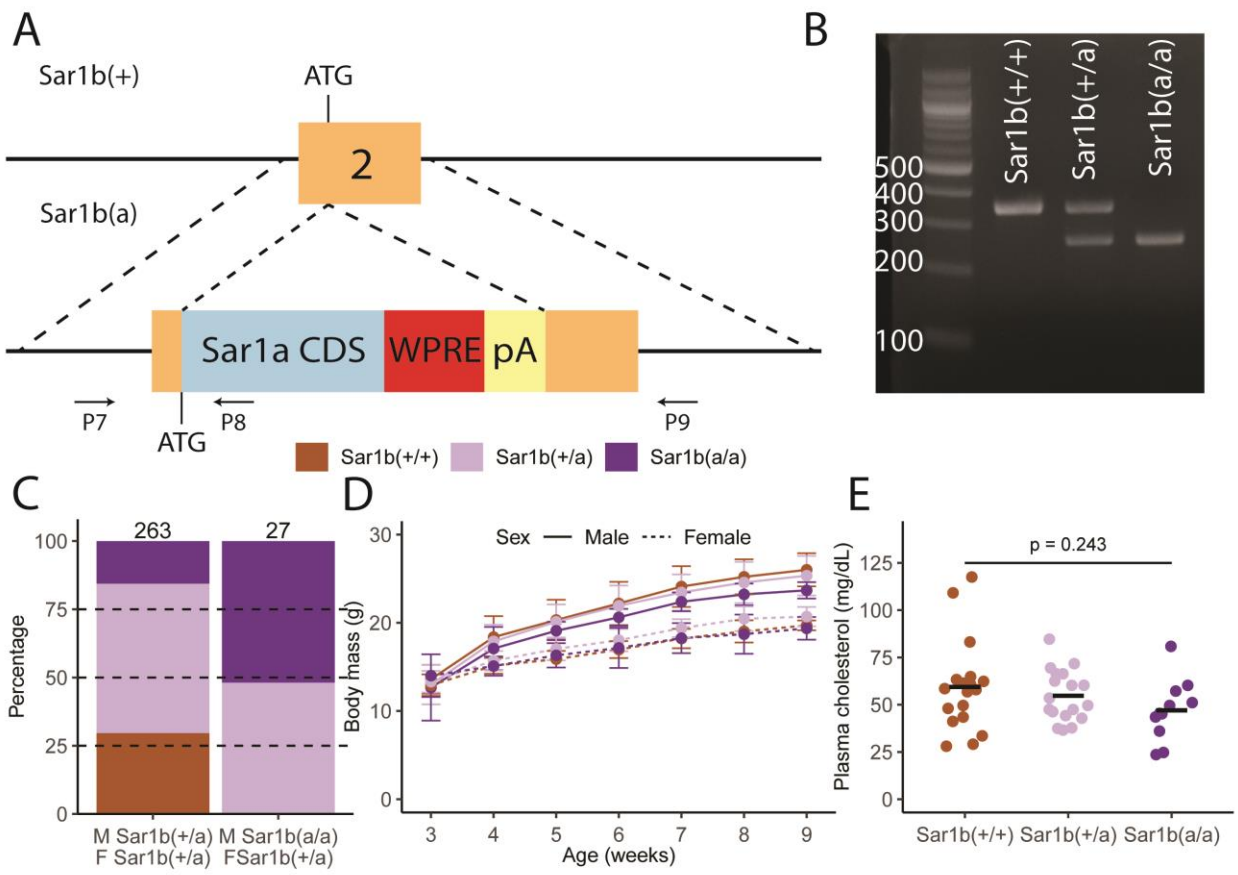

Figure 2-5: SAR1A expressed under control of the *Sar1b* endogenous locus can compensate for complete loss of SAR1B. (A) Schematic of the *Sar1b(a)* allele. Dashed lines represent the homology arms for DNA recombination. Exons and introns are not drawn to scale. CDS, coding sequence; WPRE; woodchuck hepatitis virus post-transcriptional regulatory element; pA, polyA sequence. P7, P8, P9 denote binding sites for genotyping primers. **(B)** Genotyping assay using primers depicted in (B). The wild type *Sar1b(+)* allele is expected to produce a PCR product of 358 bp whereas the *Sar1b(a)* allele generates a PCR amplicon of 268 bp. **(C)** Genotype distribution of offspring from $Sar1b^{+/a}$ X $Sar1b^{+/a}$ and $Sar1b^{+/a}$ X $Sar1b^{a/a}$ matings. Numbers at the top of each bar represent the number of mice in that group at weaning. **(D)** Body mass of male and female offspring from $Sar1b^{+/a}$ X $Sar1b^{+/a}$ intercrosses between 3 and 9 weeks of age. **(E)** Plasma cholesterol levels of $Sar1b^{+/+}$, $Sar1b^{+/a}$, and $Sar1b^{a/a}$ mice at 3-4 months of age. Statistical significance was determined by a One-way ANOVA test.

2.5 Discussion

SAR1 is a small GTPase that initiates the COPII coat formation process and plays a critical role in ER-Golgi protein transport. In this study, we investigated the roles of the two closely related SAR1 paralogs in mice, demonstrating a surprising degree of overlap in function *in vivo*.

Though both paralogs are required for survival to weaning, complete deficiency for SAR1A results in lethality at mid-embryogenesis, with *Sar1b* null mice surviving to term. This contrast with the phenotype seen in humans with homozygous *SAR1B* loss-of-function mutations who survive with CMRD. Despite the high degree of amino acid sequence similarity between the two paralogs, loss of each paralog in mice results in death at distinct developmental time point. This combined with data demonstrating biochemical differences between SAR1A and SAR1B [69] suggests a divergence in protein function.

We demonstrated in *Sar1b*$^{fl/fl}$ *Alb-Cre*$^{+}$ mice that overexpression of SAR1A can compensate for the loss of SAR1B in hepatocytes, demonstrating that SAR1A and SAR1B have overlapping functions in this cell type. However, in this system, SAR1A was expressed at a non-physiologic level and the AV could have confounded our observation, particularly given that there was a small increase in cholesterol levels of mice treated with GFP expressing AV. To address these issues, we generated mice that express SAR1A under the transcriptional control of both its endogenous locus and the *Sar1b* locus. The complete rescue exhibit by *Sar1b*$^{a/a}$ mice indicates that SAR1A is largely functionally equivalent to SAR1B and that the discordant in the *Sar1a*- and *Sar1b*-null mouse phenotype is due to differential gene expression patterns.

Gene duplications are common evolutionary events that contribute to the expansion of the genome. One of the two copies usually accumulates inactivating mutations, transitioning to a pseudogene, which is eventually lost from the genome. In a minority of cases, the duplicated copies undergo neofunctionalization (acquisition of new functions different from those of the ancestral gene) or subfunctionalization (division of the ancestral gene's functions among the paralogs) and remain in the genome. Subfunctionalization can occur at either the level of protein function or transcription [49]. Here, we demonstrated that subfunctionalization at the

transcription level likely explains the maintenance of two SAR1 paralogs in mammalian genomes. Expression of both paralogs is required to recapitulate the expression pattern of the ancestral *SAR1* gene.

SAR1B is expressed at approximately 3-fold higher level that *SAR1A* in human intestine [51] and 4-fold higher level than *Sar1a* in mouse liver [160]. In both mouse and human, the absence of SAR1B results in a compensatory increase in SAR1A expression [51, 160]. However, it appears that the approximately 1.5-fold increase in the expression of *SAR1A* in CMRD patients [51] and 2-fold increase in *Sar1a* expression in livers of $Sar1b^{fl/fl}$ $Alb\text{-}Cre^{+}$ mice [160] are insufficient to restore total SAR1 to the critical level required for normal function and prevent the manifestation of CMRD in human and hypocholesterolemia in mice. However, if the total level of SAR1 is restored to the critical level either by overexpression of SAR1A by a viral vector or additional expression of SAR1A under the control of *Sar1b* regulatory elements, the hypocholesterolemia and lethality observed in $Sar1b^{fl/fl}$ $Alb\text{-}Cre^{+}$ and $Sar1b^{-/-}$ mice can be rescued.

Melville et al recently reported that human SAR1A and SAR1B have different biochemical properties [69]. Our data suggest that even if these paralogs differ at the molecular level, they have largely similar functions such that SAR1A can compensate for the loss of SAR1B, though it is unclear if the reverse is also true. We also cannot exclude subtle phenotypic differences between $Sar1b^{+/+}$ and $Sar1b^{a/a}$ mice. While unlikely given that human and mouse SAR1A and SAR1B only differ in 3 and 2 amino acids, respectively, it is also possible that the mouse SAR1 paralogs do not exhibit biochemical differences that are observed between the human paralogs.

The phenotypic discrepancy between humans and mice with SAR1B deficiency resembles that seen between humans and mice that lack SEC23B [58, 181]. Divergence in gene expression patterns across tissues within the same species and across mammalian species could explain this phenomenon for both paralog pairs. Our data suggest that therapeutic upregulation of one paralog can rescue the deficiency of the other. Indeed, CRISPR-mediated activation of SEC23A can rescue the erythroid differentiation defects exhibit by SEC23B-deficient HUDEP-2, an erythroid cell line [79]. A similar strategy could potentially be leveraged to treat patients with CMRD due to SAR1B deficiency.

Table 2-5: Tukey's post-hoc test for comparison between AV type, time, and dose

Term	Group 1	Group 2	Null value	Estimate	Confidence Low	Confidence High	Adjusted p-value
av	GFP	SAR1A	0	36.79463	29.01538	44.57387	1.89E-11
av	GFP	SAR1B	0	33.31175	25.53251	41.09099	1.89E-11
av	SAR1A	SAR1B	0	-3.48288	-11.2621	4.296368	0.533
day	0	3	0	45.32062	37.54138	53.09987	1.89E-11
day	0	7	0	71.228	63.44876	79.00724	1.89E-11
day	3	7	0	25.90737	18.12813	33.68662	1.39E-10
dose	Low	High	0	4.275083	-1.01253	9.562695	0.111
av:day	GFP:0	SAR1A:0	0	-1.82725	-19.8608	16.20633	1
av:day	GFP:0	SAR1B:0	0	-0.32875	-18.3623	17.70483	1
av:day	GFP:0	GFP:3	0	19.80087	1.767292	37.83446	0.0211
av:day	GFP:0	SAR1A:3	0	60.08888	42.05529	78.12246	1.90E-11
av:day	GFP:0	SAR1B:3	0	53.91612	35.88254	71.94971	2.14E-11
av:day	GFP:0	GFP:7	0	24.48537	6.451792	42.51896	0.00153
av:day	GFP:0	SAR1A:7	0	96.4085	78.37492	114.4421	1.89E-11
av:day	GFP:0	SAR1B:7	0	90.63412	72.60054	108.6677	1.89E-11
av:day	SAR1A:0	SAR1B:0	0	1.4985	-16.5351	19.53208	1
av:day	SAR1A:0	GFP:3	0	21.62812	3.594542	39.66171	0.00797
av:day	SAR1A:0	SAR1A:3	0	61.91612	43.88254	79.94971	1.90E-11
av:day	SAR1A:0	SAR1B:3	0	55.74337	37.70979	73.77696	1.97E-11
av:day	SAR1A:0	GFP:7	0	26.31262	8.279042	44.34621	5.00E-04
av:day	SAR1A:0	SAR1A:7	0	98.23575	80.20217	116.2693	1.89E-11
av:day	SAR1A:0	SAR1B:7	0	92.46137	74.42779	110.495	1.89E-11
av:day	SAR1B:0	GFP:3	0	20.12962	2.096042	38.16321	0.0178
av:day	SAR1B:0	SAR1A:3	0	60.41763	42.38404	78.45121	1.90E-11
av:day	SAR1B:0	SAR1B:3	0	54.24487	36.21129	72.27846	2.09E-11
av:day	SAR1B:0	GFP:7	0	24.81412	6.780542	42.84771	0.00126
av:day	SAR1B:0	SAR1A:7	0	96.73725	78.70367	114.7708	1.89E-11
av:day	SAR1B:0	SAR1B:7	0	90.96287	72.92929	108.9965	1.89E-11
av:day	GFP:3	SAR1A:3	0	40.288	22.25442	58.32158	3.72E-08
av:day	GFP:3	SAR1B:3	0	34.11525	16.08167	52.14883	2.83E-06
av:day	GFP:3	GFP:7	0	4.6845	-13.3491	22.71808	0.995
av:day	GFP:3	SAR1A:7	0	76.60763	58.57404	94.64121	1.89E-11
av:day	GFP:3	SAR1B:7	0	70.83325	52.79967	88.86683	1.89E-11
av:day	SAR1A:3	SAR1B:3	0	-6.17275	-24.2063	11.86083	0.972
av:day	SAR1A:3	GFP:7	0	-35.6035	-53.6371	-17.5699	1.01E-06
av:day	SAR1A:3	SAR1A:7	0	36.31962	18.28604	54.35321	6.11E-07
av:day	SAR1A:3	SAR1B:7	0	30.54525	12.51167	48.57883	3.22E-05
av:day	SAR1B:3	GFP:7	0	-29.4308	-47.4643	-11.3972	6.74E-05
av:day	SAR1B:3	SAR1A:7	0	42.49238	24.45879	60.52596	7.78E-09
av:day	SAR1B:3	SAR1B:7	0	36.718	18.68442	54.75158	4.62E-07
av:day	GFP:7	SAR1A:7	0	71.92313	53.88954	89.95671	1.89E-11

av:day	GFP:7	SAR1B:7	0	66.14875	48.11517	84.18233	1.89E-11
av:day	SAR1A:7	SAR1B:7	0	-5.77438	-23.808	12.25921	0.981

Table 2-6: Genotyping primer sequences

Primer name	Target	Sequence (5' → 3')
P1	*Sar1a(+)*	GTTGTGAGCTGCCGTGTATG
P2	*Sar1a(+)*	GAGCCATTACCATGGAGGAA
P3	*Sar1a(gt)*	CCACAACGGGTTCTTCTGTT
P4	*Sar1b(fl)*	CCCATATTTAGCTCTATCTTGCTGGAA
P5	*Sar1b(fl)*	GCAGTTGTCCAGCCATCTTTGTGACC
P6	*Sar1b(fl)*	GCCGGAACCGAAGTTCCTA
P7	*Sar1b(+)*	GTATCCTGTGTATGTTGGCCATGC
P8	*Sar1b(a)*	TCAGTTCTTCTGAAGTGGGATGGAGT
P9	*Sar1b(+)*	TCGAGGTCTCAGATCACAGCTTCA

Chapter 3
Identification of Secreted Proteins by Comparison of Protein Abundance in Conditioned Media and Cell Lysates[2]

3.1 Abstract

Analysis of the full spectrum of secreted proteins in cell culture is complicated by leakage of intracellular proteins from damaged cells. To address this issue, we compared the abundance of individual proteins between the cell lysate and the conditioned medium, reasoning that secreted proteins should be relatively more abundant in the conditioned medium. Marked enrichment for signal-peptide-bearing proteins with increasing conditioned media to cell lysate ratio, as well loss of this signal following Brefeldin A treatment, confirmed the sensitivity and specificity of this approach. The subset of proteins demonstrating increased conditioned media to cell lysate ratio in the presence of Brefeldin A identified candidates for unconventional secretion via a pathway independent of ER to Golgi trafficking.

3.2 Introduction

Approximately 34% of genes in the mammalian genome encode proteins destined for the secretory pathway, with these secreted proteins serving a variety of physiological roles [182, 183]. Proteins bearing signal-peptides and/or transmembrane domains are co-translationally inserted into the endoplasmic reticulum (ER) where they are folded prior to undergoing

[2]This chapter has been published in the Journal of Analytical Biochemistry: Abbineni PS*, Tang VT*, da Veiga Leprevost F, Basrur V, Xiang J, Nesvizhskii AI, Ginsburg D. *Identification of secreted proteins by comparison of protein abundance in conditioned media and cell lysates.* Analytical Biochemistry. 2022 Oct 15;655:114846. *Equal Contributions.

anterograde transport to the Golgi via coatomer protein complex II (COPII) coated vesicles or tubular structures [184-186]. In the Golgi, these proteins are then packaged into mature vesicles that enter constitutive or regulated exocytosis pathways for delivery to various intracellular compartments, or fusion with the plasma membrane for secretion into the extracellular space or insertion into the plasma membrane. While the majority of secreted proteins utilize the ER-Golgi secretory vesicle pathway, a subset of cytoplasmic proteins has been shown to be secreted via ER-Golgi independent "unconventional" secretory routes [9, 10, 187]. Following synthesis in the cytoplasm, unconventionally secreted proteins, which include some cytokines and angiogenic growth factors, are secreted either through protein conducting pores that form in the plasma membrane, or via vesicular intermediates derived from the auto/endo-lysosomal system[10, 188, 189].

Characterizing the entire set of secreted proteins, i.e. the secretome, using cell culture-based model systems has been a long-standing challenge [183]. Several studies have focused on identifying changes to the secretome under pathophysiological conditions to identify biomarkers for disease, or following inhibition of specific components of the secretory pathway with the aim of identifying regulators of select secreted proteins [190-193]. However, a major challenge in studying the secretome of cultured cells is contamination by intracellular proteins released into the media due to cellular injury, which confounds identification of bona fide secreted proteins [194]. To circumvent this issue, we compared the abundance of proteins identified in the cell lysate and conditioned media. We reasoned that secreted proteins should be more abundant in the media relative to the cell lysates, whereas intracellular proteins that are released into the media due to cell injury should be present at a lower ratio in the media compared to the cell lysates. Previous studies have utilized a similar approach using label-free quantification-based mass

spectrometry whereby spectral counts from conditioned media and cell lysates are compared to identify secreted proteins, or have applied isobaric tags for relative and absolute quantitation (ITRAQ) reagents for label-based quantification of media and lysate proteins [195-197]. In order to independently test this approach, we utilized isobaric tandem-mass-tags (TMT) to label proteins in the cell lysate and conditioned media of HuH7 hepatoma cells, followed by high throughput mass spectrometry to compare the abundance of proteins in conditioned media relative to cell lysates. This approach distinguished between intracellular and secreted proteins with high specificity. Treatment with Brefeldin-A (BFA), an inhibitor of ER-Golgi trafficking, provided additional validation of this approach and identified candidate proteins trafficked via BFA-resistant unconventional secretion pathways. The approach described here is readily applicable to characterizing secretomes from a wide range of cell types.

3.3 Methods

3.3.1 Cell culture, and BFA treatment

HuH7 cells [198] were cultured in DMEM Glutamax media (ThermoFisher Scientific, Waltham MA, 10569-044) supplemented with 10% fetal bovine serum (MilliporeSigma, Burlington MA, F8067) and penicillin/streptomycin (ThermoFisher Scientific, Waltham MA, 15140-122). Cells were maintained at 20-85% confluence and passaged every 3-4 days. For BFA experiments, HuH7 cells were treated with 1 µg/ml of BFA (BioLegend, San Diego CA, 420601) and after incubation for 4 hours in serum-free, phenol red-free DMEM, conditioned medium and cell lysate were collected and processed as described below.

3.3.2 Conditioned medium collection

Conditioned media were centrifuged at 2500g at 4ºC for 20 minutes to pellet cell debris, followed by ultracentrifugation at 120,000g at 4ºC for 90 minutes to pellet exosomes [199].

Following ultracentrifugation, the supernatants were concentrated using a 3 kDa molecular weight cutoff concentrator (MilliporeSigma, Burlington MA, UFC900324). Protein concentrations were determined by DC protein assay (Bio-Rad, Hercules CA, 500-011).

3.3.3 Cell lysate collection

Following removal of conditioned media, cells were rinsed twice in cold (4ºC) PBS, followed by the addition of 2 ml of RIPA buffer (Thermo Scientific, 89900) containing a protease inhibitor cocktail (cOmplete™, Mini Protease Inhibitor Cocktail, 11836153001). Cells were then detached using a cell scraper and collected into 15 ml conical tubes. Cell suspensions were sonicated and rotated end-over-end for 1 hour. Next, lysates were centrifuged at 21,000g at 4ºC for 45 minutes. The supernatants were then transferred to a new Eppendorf tube and the protein concentrations determined by DC protein assay).

3.3.4 Immunoblotting

Cell lysates were diluted to 2 µg/µl in Laemmli buffer (Bio-Rad, Hercules CA, 1610747) and heated at 95ºC for 5 minutes and resolved using a 4-20% tris-glycine gel (ThermoFisher Scientific, Waltham MA, XP04200). Next, samples were transferred to nitrocellulose membranes (ThermoFisher Scientific, Waltham MA, LC2000) and probed with antibodies against PCSK9 (abcam, Cambridge UK, ab181142, 1:1000), and ß-actin (Santa Cruz Biotechnology, Dallas TX, sc-47778, 1:1000). Chemiluminescence-based quantification of PCSK9 abundance in HuH7 cell lysates and conditioned media following immunoblotting was performed using the Azure c600 imager (Azure Biosystems, CA, USA), and ImageJ for image analysis [200].

3.3.5 Mass spectrometry

Labeling of peptides with TMT reagents followed by TMT proteomics was performed as previously described [201]. 75 µg lysate and 50 µg media samples were proteolysed and labeled

with TMT 6-plex according to the manufacturer's protocol (ThermoFisher). Briefly, upon reduction (5 mM DTT, for 30 min at 45 C) and alkylation (15 mM 2-chloroacetamide, for 30 min at room temperature) of cysteines, the proteins were precipitated by adding 6 volumes of ice cold acetone followed by overnight incubation at -20° C. The precipitate was collected by centrifugation, and the pellet was allowed to air dry. The pellet was resuspended in 0.1M Triethylammonium bicarbonate (TEAB) and overnight (~16 h) digestion was performed with trypsin/Lys-C mix (1:25 protease:protein; Promega) at 37° C with constant mixing using a thermomixer. The TMT 6-plex reagents were dissolved in 41 µl of anhydrous acetonitrile and labeling was performed by transferring the entire digest to TMT reagent vial and incubating at room temperature for 1 h. Reaction was quenched by adding 8 µl of 5% hydroxyl amine and further 15 min incubation. Labeled samples were mixed together, and dried using a vacufuge. An offline fractionation of the combined sample (~200 µg) into 8 fractions was performed using high pH reversed-phase peptide fractionation kit according to the manufacturer's protocol (Pierce; Cat #84868). Fractions were dried and reconstituted in 9 µl of 0.1% formic acid/2% acetonitrile in preparation for LC-MS/MS analysis.

3.3.6 Liquid chromatography-mass spectrometry analysis (LC-multinotch MS3):

In order to obtain superior quantitation accuracy, we employed multinotch-MS3 (McAlister GC, ref below) which minimizes the reporter ion ratio distortion resulting from fragmentation of co-isolated peptides during MS analysis [202]. Orbitrap Fusion (Thermo Fisher Scientific) and RSLC Ultimate 3000 nano-UPLC (Dionex) was used to acquire the data. Two µl of the sample was resolved on a PepMap RSLC C18 column (75 µm i.d. x 50 cm; Thermo Scientific) at the flow-rate of 300 nl/min using 0.1% formic acid/acetonitrile gradient system (2-22% acetonitrile in 150 min; 22-32% acetonitrile in 40 min; 20 min wash at 90% followed by 50

min re-equilibration) and directly spray onto the mass spectrometer using EasySpray source (Thermo Fisher Scientific). Mass spectrometer was set to collect one MS1 scan (Orbitrap; 120K resolution; AGC target $2x10^5$; max IT 100 ms) followed by data-dependent, "Top Speed" (3 seconds) MS2 scans (collision induced dissociation; ion trap; NCE 35; AGC $5x10^3$; max IT 100 ms). For multinotch-MS3, top 10 precursors from each MS2 were fragmented by HCD followed by Orbitrap analysis (NCE 55; 60K resolution; AGC $5x10^4$; max IT 120 ms, 100-500 m/z scan range).

3.3.7 Protein identification and quantification

Raw mass spectrometry files were converted into open mzML format using the msconvert utility of the Proteowizard software suite, and analyzed using FragPipe computational platform (https://fragpipe.nesvilab.org/) using default TMT10-MS3 workflow. MS/MS spectra were searched using the database search tool MSFragger v3.4 [203, 204] against an *Homo sapiens* UniprotKB/SwissProt protein sequence database appended with an equal number of decoy sequences (downloaded on January 31, 2022). Whole cell lysate MS/MS spectra were searched using a precursor-ion mass tolerance of 20 ppm, and allowing C12/C13 isotope errors 0/1/2/3. Mass calibration and parameter optimization were enabled. Cysteine carbamidomethylation (+57.0215), and lysine TMT labeling (+229.1629) were specified as fixed modifications, and methionine oxidation (+15.9949), N-terminal acetylation (+42.01060), and TMT labeling (+229.1629) of peptide N terminus and serine residues were specified as variable modifications. The search was restricted to tryptic peptides, allowing up to two missed cleavage sites. Peptide to spectrum matches (PSMs) were further processed using Percolator [205] to compute the posterior error probability, which was then converted to posterior probability of correct identification for each PSM. The resulting files from Percolator were converted to

pep.xml format, and pep.xml files from both TMT experiments (the secretome and the whole lysate experiment) were then processed together to assemble peptides into proteins (protein inference) using ProteinProphet [206] run via the Philosopher toolkit v4.2.1 [207] to create a combined file (in prot.xml format) of high confidence protein groups. The combined prot.xml file and the individual PSM lists for each of the two TMT experiment were further processed using the Philosopher filter command as follows. Each peptide was assigned either as a unique peptide to a particular protein group or assigned as a razor peptide to a single protein group that had the most peptide evidence. The protein groups assembled by ProteinProphet were filtered to 1% protein-level False Discovery Rate (FDR) using the target-decoy strategy and the best peptide approach (allowing both unique and razor peptides). The PSM lists were filtered using a sequential FDR strategy, retaining only those PSMs that passed 1% PSM-level FDR filter and mapped to proteins that also passed the global 1% protein-level FDR filter. In addition, for all PSMs corresponding to a TMT-labeled peptide, TMT reporter ion intensities were extracted from the MS/MS scans (using 0.002 Da window) and the precursor ion purity scores were calculated using the intensity of the sequenced precursor ion and that of other interfering ions observed in MS1 data (within a 0.7 Da isolation window). The PSM output files were further processed using TMT-Integrator v3.2.1 to generate summary reports at the protein level. TMT-Integrator [208] used as input the PSM tables generated by the Philosopher pipeline as described above and created integrated reports with quantification across all samples. First, PSMs were filtered to remove all entries that did not pass at least one of the quality filters, such as PSMs with (a) no TMT label; (b) precursor-ion purity less than 50%; (c) summed reporter ion intensity (across all channels) in the lower 5% percentile of all PSMs in the corresponding PSM.tsv file. In the case of redundant PSMs (i.e., multiple PSMs in the same MS run sample corresponding the same

peptide ion), only the single PSM with the highest summed TMT intensity was retained for subsequent analysis. Both unique and razor peptides were used for quantification, while PSMs mapping to common external contaminant proteins (that were included in the searched protein sequence database) were excluded. Next, for each PSM the intensity in each TMT channel was converted into log2-based ratio to the reference using the virtual reference approach of TMT-Integrator. The PSMs were grouped to the protein level, and the protein ratios were computed as the median of the corresponding PSM ratios after outlier removal. Protein ratios were then converted back to absolute protein intensity in each sample by using the reference protein intensity estimated (separately for each experiment, the secretome and the whole lysate), using the sum of all MS2 reporter ions from all corresponding PSMs.

In order to calculate media/lysate abundance ratio, we used the absolute protein intensities values with log2(media/lysate) ratio as input. Statistical analysis of changes to protein abundance and media/lysate ratio following BFA treatment relative to untreated controls was performed using the limma statistical package [209].

3.4 Results

3.4.1 Identification of secreted proteins by comparison of protein abundance in the cell lysate and conditioned media

We reasoned that proteins secreted by cultured cells should exhibit significantly greater abundance in the conditioned media relative to the cell lysate when compared to non-secreted proteins. For this analysis, protein preparations from conditioned media and cell lysates of HuH7 cells were subjected to quantitative proteomic analysis using TMT and LC-MS/MS. Samples from different conditions were pooled after TMT labelling, with the lysates and media fractions analyzed in independent MS runs (**Figure 3-1A**).

A total of 5419 unique proteins were identified in the lysate and media fractions across all three independent replicates, with 48% of these proteins identified in both the media and lysate fractions (**Figure 3-1B**). Only 12.5% of the proteins detected in the media contained signal-peptides (**Figure 3-1B**), a 1.75-fold enrichment relative to the cell lysate. These data suggest that a large subset of the proteins detected in the media represent cytoplasmic/intracellular protein contamination, likely due to cell damage or lysis. We then calculated the media to lysate abundance ratio (M/L ratio) for 3 known cytoplasmic proteins (all lacking signal peptides) that were detected in the media, actin beta (ACTB), lactate dehydrogenase A (LDHA) and tubulin alpha-1B chain (TUBA1B), and 3 well documented signal peptide-containing, liver-specific secreted proteins, alpha-1-antitrypsin (SERPINA1), apolipoprotein B (APOB), and proprotein convertase subtilisin/kexin type 9 (PCKS9) (**Figure 3-1C**). As expected, all 3 signal peptide containing proteins exhibit significantly higher M/L ratios (~12-80, **Figure 3-1C M/L ratio**) relative to the intracellular proteins (~0.1-0.8, p = 0.0019). Calculation of media to lysate abundance ratios (M/L ratio) for all 2628 proteins identified in both the media and lysate fractions is shown in **Figure 3-1D.** The fraction of proteins with signal peptides steadily increased with greater M/L ratio, consistent with greater enrichment of secreted proteins in the media relative to the lysate, and validating the media/lysate ratio analysis approach.

Gene ontology analysis of proteins with M/L ratios greater than 10 revealed significant enrichment for proteins involved in multiple lipoprotein and cholesterol metabolism processes (**Figure 3-1E**), consistent with the hepatocyte origin of the HuH7 cell line. The majority (71.3%) of proteins detected solely in the media contained signal peptides (101 out of 147), with this subset largely composed of low abundance secreted proteins, likely explaining their failure to be

detected in the lysate fraction. The remaining proteins without signal peptides represent potential unconventionally secreted proteins (see below). Only ~5% of proteins detected solely in the lysate contained signal peptides, which are largely intracellular membrane associated proteins with transmembrane domains (91 out of 130 proteins), or soluble proteins contained within the lumen of intracellular organelles, including the lysosomal proteins, cathepsin L2 and galactocerebrosidase.

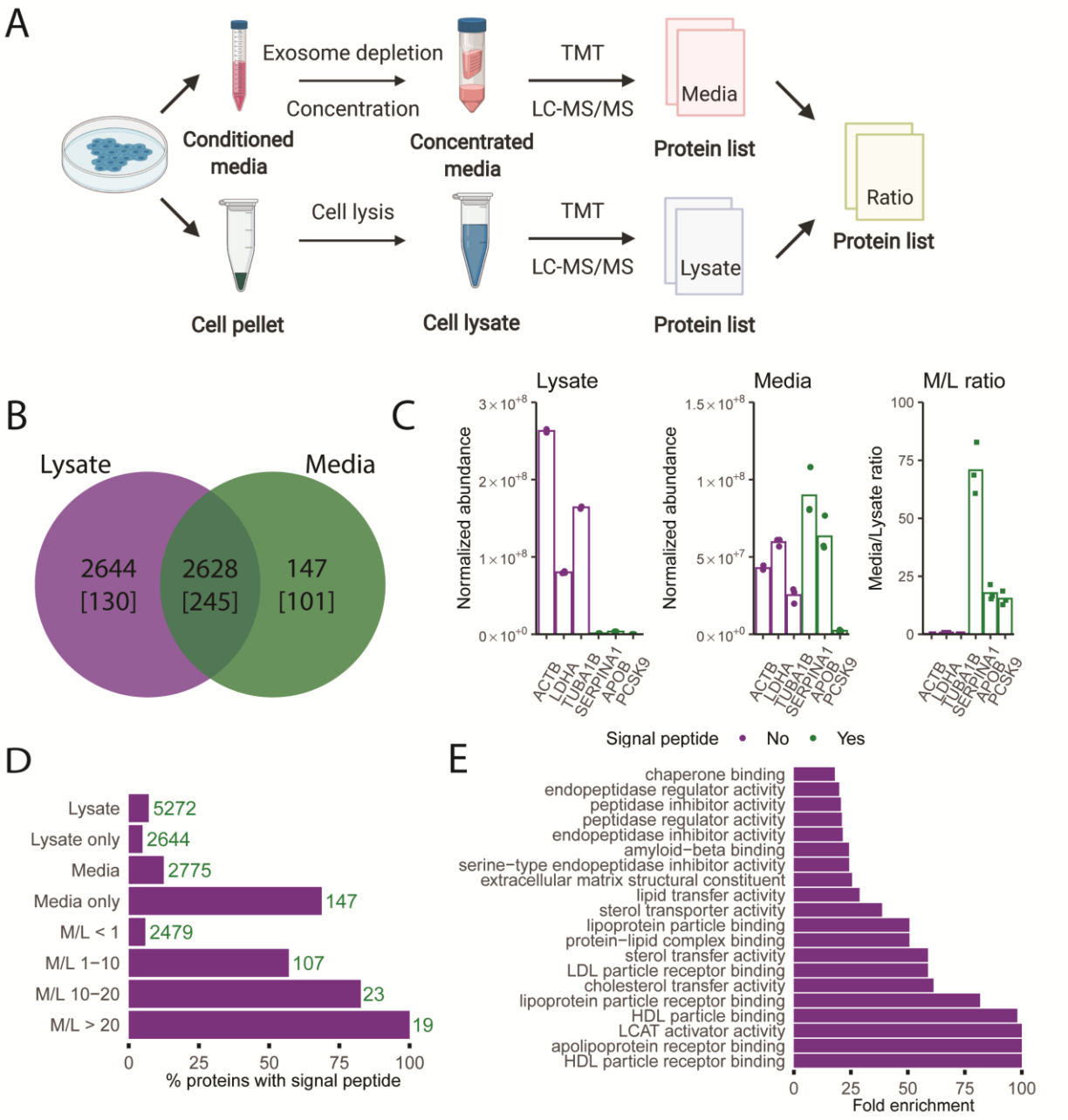

Figure 3-1: Comparison of protein abundance in the cell lysate and conditioned media enables identification of secreted proteins. (A) Experimental approach: Proteins isolated from cell lysates and conditioned media were labelled with tandem-mass-tag (TMT) reagents prior to LC-MS/MS analysis to quantify relative protein abundance. A media/lysate (M/L) abundance ratio was calculated for all proteins identified in both fractions. **(B)** Venn diagram showing the overlap in proteins detected in the cell lysates and conditioned media. The number of proteins with signal peptides present in each fraction is indicated in brackets. **(C)** The abundance of known cytoplasmic proteins (ACTB, LDHA, TUBA1B) and well documented secreted proteins (SERPINB1, APOB, PCSK9) in the lysate and media fraction, and their M/L ratio. **(D)** Proteins were binned into discreet M/L ratio intervals, and the percentage of proteins with signal peptides in each bin was calculated. The number of proteins present in each bin is indicated. **(E)** Gene ontology (GO) analysis for molecular functions for the set of proteins with M/L ratio greater than 10.

3.4.2 BFA inhibits ER-Golgi trafficking and further validates the M/L ratio analysis approach

We next treated HuH7 cells with BFA, an Arf1 GTPase inhibitor that blocks ER-Golgi transport and conventional protein secretion (**Figure 3-2A**). Western blotting for the conventionally secreted protein, PCSK9, with or without BFA treatment is shown in **Figure 3-2B.** As expected, PCSK9 accumulation in the cell lysate and depletion from the conditioned media is observed following BFA treatment (**Figure 3-2C**). Proteomic analysis of control and BFA treated cell lysates and conditioned media by TMT LC-MS/MS analysis is shown in **Figure 3-2D-F**.

In **Figure 3-2D-E**, the volcano plots show the log2 fold changes (l2fc) in the abundance of proteins in the cell lysate and conditioned media, respectively, following BFA treatment. A positive l2fc indicates increased abundance in response to BFA treatment, and a negative l2fc indicates decreased abundance. Following BFA treatment, most signal peptide containing proteins exhibit positive l2fc in the lysate (**Figure 3-2D**), and negative l2fc in the media (**Figure 3-2E**), consistent with BFA inhibition of conventional protein secretion resulting in depletion of secreted proteins from the conditioned media and accumulation in the cell lysate. Signal peptide containing proteins that do not change in abundance following BFA treatment are comprised primarily of either membrane proteins (38.6%, 93 out of 241 proteins), or are confined to the lumen of intracellular compartments as indicated by GO cellular component analysis.

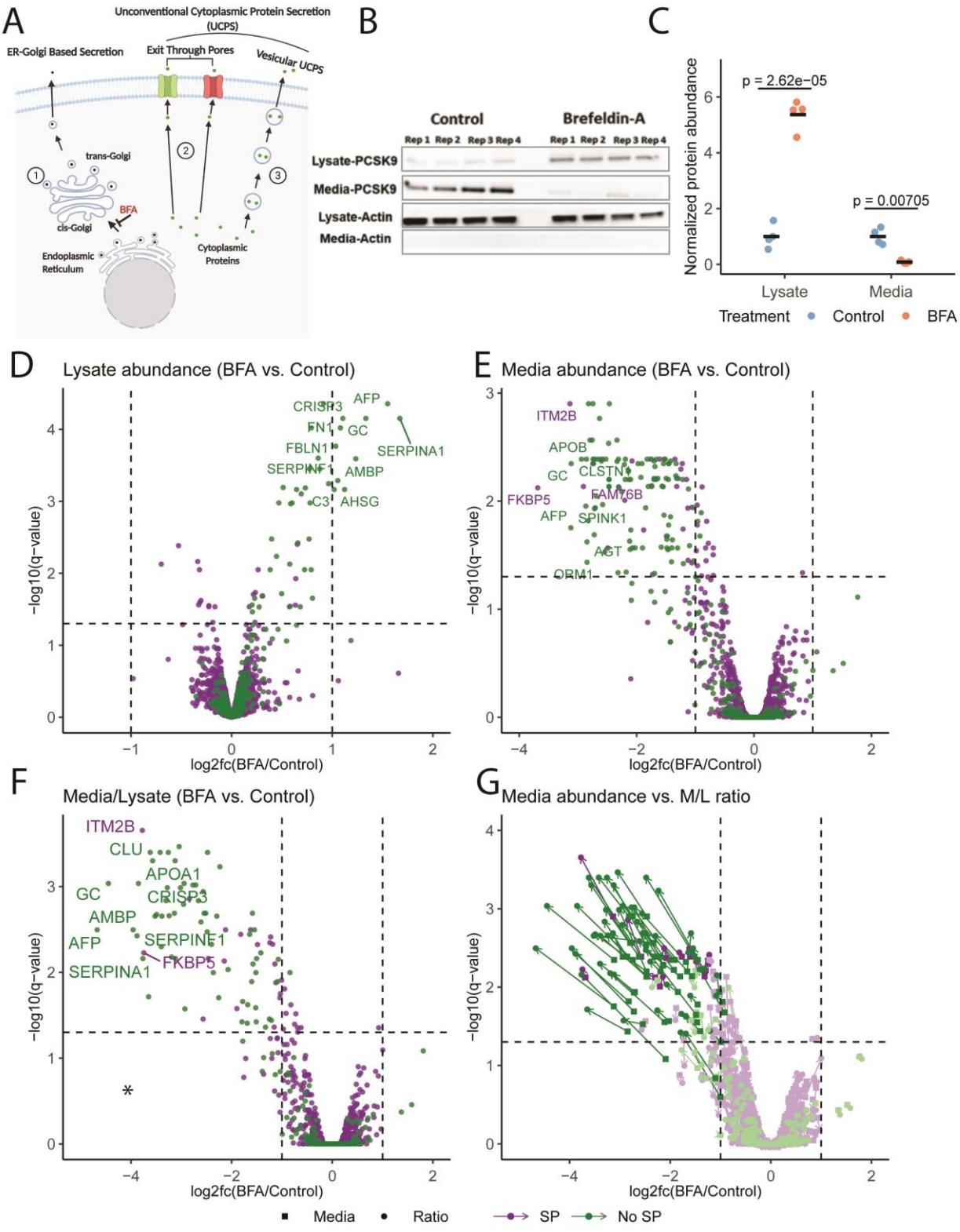

Figure 3-2: M/L ratio analysis combined with BFA treatment enables high-sensitivity identification of conventionally secreted proteins. (A) Proteins bearing signal peptides are co-translationally inserted into the ER and utilize the conventional ER-Golgi-Secretory vesicle/tubule route (pathway (1)) to exit the cell or to reach other

intracellular destinations. The process of ER-Golgi transport is blocked by BFA. A number of cytoplasmic proteins are known to be secreted unconventionally via ER-Golgi independent secretory routes by either translocating through plasma membrane pores that are formed by the secreted protein itself, or by other pore-forming proteins (pathway(2)), or by entering membrane bound organelles that fuse with the plasma membrane (pathway (3)). These latter 2 groups of proteins should be resistant to BFA. **(B)** Immunoblot analysis of PCSK9 abundance in the HuH7 cell lysate and conditioned media following treatment with 1 ug/ml BFA. Four biological replicates are labeled Rep1-4. **(C)** Chemiluminescence-based quantification of PCSK9 abundance in HuH7 cell lysates and conditioned media following immunoblotting. BFA treated samples were normalized to controls. **(D-G)** Volcano plots [45] showing changes to the abundance of proteins in the cell lysate **(D)** and conditioned media **(E)**, and changes to M/L ratio **(F)** following BFA treatment measured by TMT-based mass spectrometry. The l2fc and statistical significance are plotted on the x and y-axis, respectively. Signal peptide containing proteins are indicated in green and non-signal peptide containing in purple. **(G)** Comparison of l2fc following BFA treatment for individual proteins comparing protein abundance in the media alone (squares) to the M/L ratio (circles).

In order to account for protein abundance changes in both the lysate and media fractions as described above, we calculated l2fc of M/L ratios following BFA treatment **Figure 3-2F**). A negative l2fc indicates a decrease in M/L in response to BFA treatment (depletion from the media and/or accumulation in the lysate). 75% of proteins with negative l2fc contain signal peptides, further confirming the robustness of this analysis approach. **Figure 3-2G** compares the l2fc observed for protein abundance in the conditioned media to the l2fc for the M/L ratio. The M/L ratio analysis, which considers protein abundance in both the media and lysate fractions, provides a higher sensitivity for the identification of secreted proteins, with considerably larger (negative) l2fc compared to analysis of the conditioned media alone for the same protein. The 25% of proteins sensitive to BFA treatment but lacking signal peptides are limited to membrane proteins (e.g, integral membrane protein 2B (ITM2B)) with transmembrane domains and/or signal anchors that target them to the endomembrane system, or soluble proteins that potentially have cryptic internal signal peptides that target them to the ER (e.g, FKBP prolyl isomerase 5 (FKBP5)). The majority of proteins detected solely in the media contain signal peptides (71%) (**Figure 3-2D**), with most of these depleted from the media upon BFA treatment (**Figure 3-3B**).

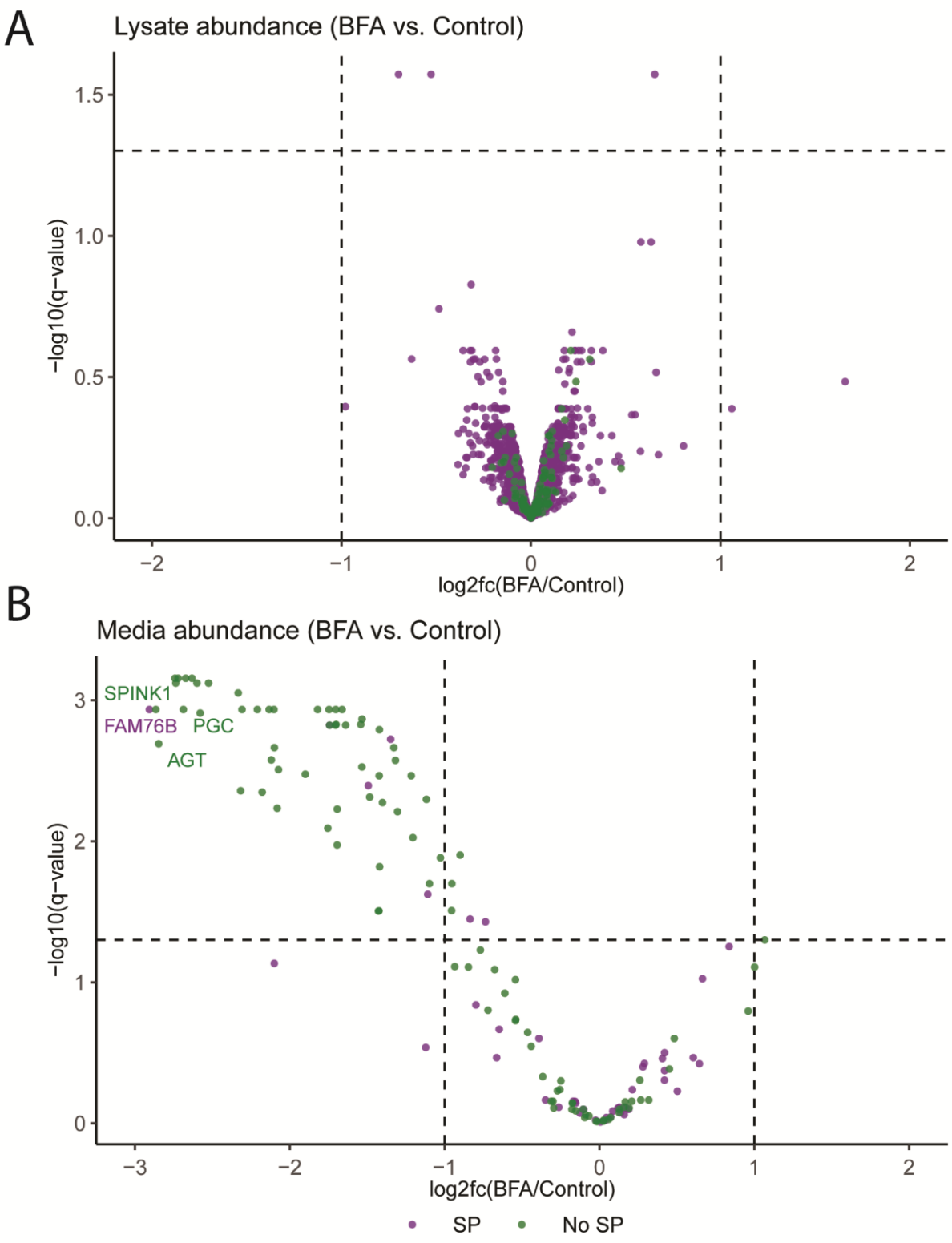

Figure 3-3: Proteins detected solely in the media fraction are sensitive to BFA. Volcano plots showing changes to the abundance of proteins detected solely in the cell lysate **(A)** and conditioned media **(B)** following BFA treatment measured by TMT-based mass spectrometry.

3.4.3 Identification of unconventionally secreted proteins

All proteins with an M/L ratio greater than 20 (19 proteins) contain a signal peptide, suggesting that the most efficiently secreted proteins in this cell type use the conventional ER-Golgi secretory pathway. In contrast, 17% of proteins with M/L ratios of 10 – 20 (4 out of 23 proteins) and 43% of proteins with M/L ratios of 1-10 (46 out of 107 proteins) lack a signal peptide, suggesting that a subset of these proteins may be secreted via unconventional, ER-Golgi independent pathways (**Figure 3-1D)**. Of the 10 proteins without signal peptides with the highest M/L ratios, nine are resistant to BFA treatment (**Figure 3-4A-B**). Among this group, the protein with the highest M/L ratio, metallothionein 2A (MT2A), has previously been shown to be secreted via ER-Golgi independent mechanisms [210, 211], and its identification in our dataset further confirms the utility of our approach to identify both conventionally and unconventionally secreted proteins. The one exception among the top 10 proteins without signal peptides with the highest M/L ratios that was BFA sensitive, FKBP5, is likely still dependent on the conventional secretory pathway, despite the absence of an annotated signal peptide. Six proteins in this list (gene names: LZTS1, ZBTB8OS, ANKRD40, FNTA, HECTD3, and GFUS) are resistant to BFA treatment (**Figure 3-4B**) and lack signal peptides and transmembrane domains. Although all have been reported to have intracellular roles in regulating mitosis [212], transcription [213], and posttranslational modifications [214-216], and are not known to be secreted, our findings raise the possibility that one or more of these proteins may indeed be secreted via unconventional secretory pathways with an as yet unknown extracellular function. Future work can examine the unconventional secretory routes taken by these proteins, and delineate whether they utilize a pore mediated (Type 1 unconventional secretion), or organelle-mediated unconventional secretory pathway (Type 3 unconventional secretion) [10].

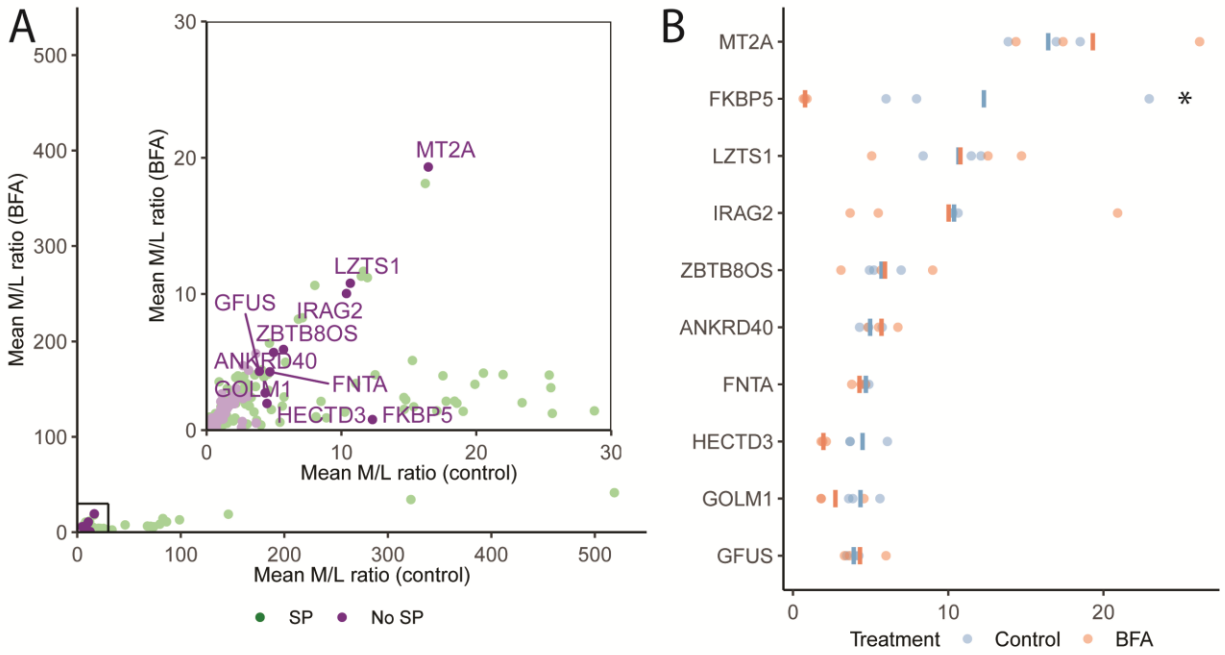

Figure 3-4: Candidate proteins secreted via unconventional secretory pathways. (A) M/L ratio of control and BFA treated cells are plotted on the x and y-axis, respectively. The inset shows proteins with a M/L ratio ranging from 0-30. Proteins sensitive to BFA have lower M/L ratios relative to untreated controls. The top 10 signal-peptide lacking proteins with the highest M/L ratios (candidate unconventionally secreted proteins) are identified by gene name. **(B)** M/L ratio of candidate unconventionally secreted proteins and sensitivity to BFA.

The remaining 2 of the 10 proteins, inositol 1,4,5-triphosphate receptor associated 2 (IRAG2) and Golgi membrane protein 1 (GOLM1), contain transmembrane domains that likely traffic them to the ER, but their resistance to BFA indicates that they may be stored in a post-Golgi compartment for secretion, as BFA inhibits ER-Golgi trafficking. The extracellular domains of transmembrane domain proteins can be proteolyzed by membrane bound proteases known as sheddases, and likely explains why these proteins were detected in the media [217]. We observed that the majority of transmembrane domain containing proteins with an M/L ratio > 1 were resistant to BFA (25 out of 27) (**Figure 3-5**), indicating secretion from a post-Golgi compartment or recycling between endocytic compartments and the plasma membrane. Alternatively, a subset of these proteins could bypass the Golgi en route to secretion and utilize unconventional secretory routes referred to as “Golgi-Bypass” (or Type 4 unconventional

secretion) [10]. However, the criteria for identification of Golgi-bypass cargoes are BFA resistance and absence of complex N-glycosylation that occurs in the Golgi [218]. As we are not able to assay for N-glycosylation in our high-throughput assay, we are not able to identify candidate Golgi-bypass cargoes.

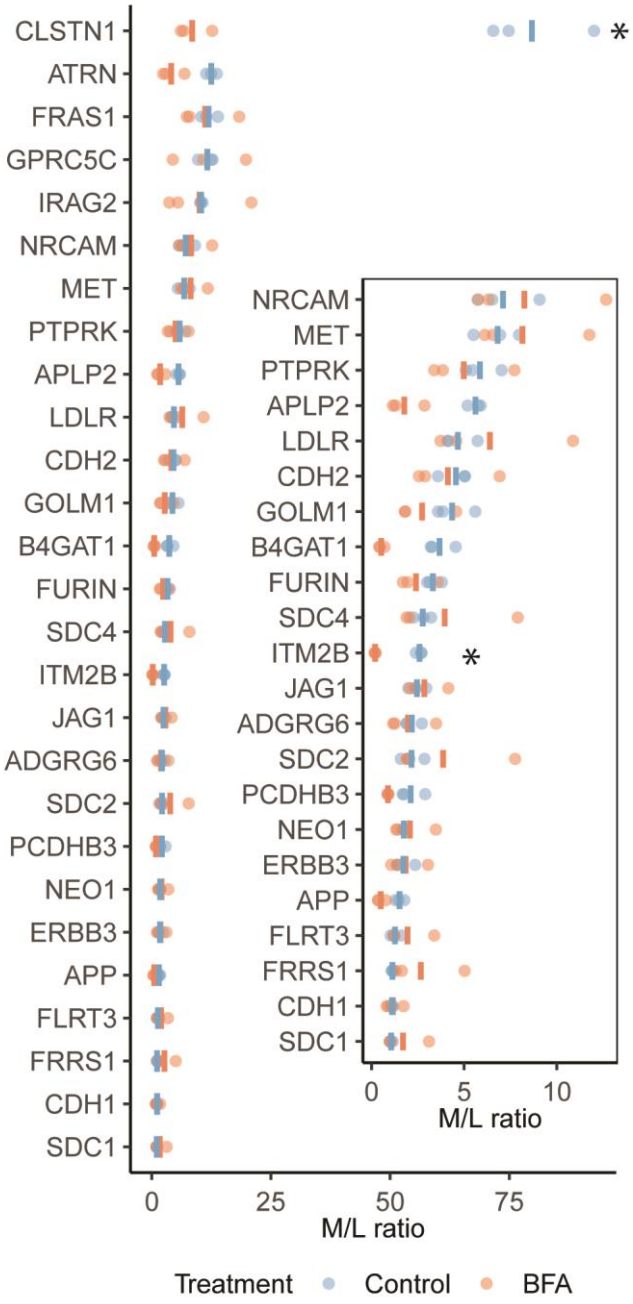

Figure 3-5: BFA sensitivity of transmembrane domain containing proteins. Transmembrane domain containing proteins detected in both the media and lysate fractions with a M/L ratio > 1. The inset shows a magnified view of the 0-15 M/L ratio range.

3.5 Discussion

Secretion of proteins into the extracellular space is a fundamental process that occurs in eukaryotic cells. In multicellular organisms, secreted proteins play multiple essential roles,

including intercellular communication, maintaining metabolic homeostasis, immune reactions, and neurotransmission. Efforts to define the set of proteins that are secreted, referred to as the secretome, have faced significant obstacles. Currently, the human secretome is defined primarily based on computational predictions of proteins that have a signal peptide and exclusion of transmembrane proteins [182, 183]. This approach has several shortcomings, including inaccuracy of algorithms for identification of proteins with a signal peptide or a transmembrane domain and omission of unconventionally secreted proteins. There are also significant challenges in defining a secretome experimentally. *In vivo* studies to identify circulating proteins in the plasma by mass spectrometry face the problem of variable protein abundance and the difficulty of detecting low abundance proteins [219]. This approach also excludes the large number of secreted proteins that function locally and do not enter the circulation. *In vitro* studies of cultured cells circumvent some of these problems and also allow selection of a specific cell type of interest. However, this approach is subject to a high false positive rate due to contamination by intracellular proteins that leak into the conditioned media from damaged cells.

Here, we propose an alternative approach that facilitates efficient and accurate identification of secreted proteins by analysis of both conditioned media and cell lysates. We show that in contrast to intracellular proteins, authentically secreted proteins are significantly more abundant in the conditioned media relative to the cell lysate. By analyzing an M/L ratio for each protein, we were able to clearly distinguish secreted proteins from those that typically remain intracellular. The efficacy of this ratiometric approach to identify secreted proteins was further confirmed by our observation that treatment with BFA significantly decreases the M/L ratio of proteins with a signal peptide compared to those without. Furthermore, as this approach does not require genetic manipulation or metabolic labeling, it can be widely applied to both

primary and immortalized cell culture systems.

In contrast to proteins traversing the conventional ER to Golgi secretory pathway, no definitive structural features analogous to the signal peptide have been identified that target cytoplasmic proteins for unconventional. Although algorithms such as SecretomeP have been designed to predict unconventionally secreted proteins based on shared physiochemical features (isoelectric point, number of charged residues, secondary structure, etc.), a recent analysis of SecretomeP and similar programs tested for known unconventionally secreted and non-secreted proteins demonstrated low sensitivity and a high false positive rate. Thus, while hundreds of proteins have been proposed to use unconventional secretory routes, few have been convincingly validated, and no standard methods are available to detect unconventionally secreted proteins. Our ratiometric analysis approach combined with the use of BFA to block ER-Golgi trafficking facilitated the identification of candidate unconventionally secreted cargoes. This method is also readily adaptable to identify unconventionally secreted proteins from diverse cell lines or cultured primary cells.

Stühler and colleagues recently pioneered the approach of comparing protein abundance in the cellular lysate fraction and conditioned media to identify secreted proteins [194, 195, 220]. We independently confirm the utility of this approach by using TMT-based labeling of lysate and media proteins, followed by label-based quantification of TMT data. Previous analyses have classified proteins as secreted if their mean abundance in the conditioned media was at least 1.5-fold higher relative to the cell lysate [196]. To identify candidate unconventionally secreted cargoes, we applied more stringent criteria based on prioritizing signal-peptide lacking, BFA-resistant proteins with the highest M/L ratios. Our observation of a strong concordance between increasing M/L ratio and higher fractions of signal peptide containing proteins suggests that the

candidates with the highest M/L ratios are most likely to represent true positives. Future studies should uncover extracellular roles for these novel unconventionally secreted cargoes and characterize molecular mechanisms that regulate their release from the cell.

The M/L ratio approach relies on the assumption that secreted proteins will be more abundant in the media relative to the lysate; thus, proteins that have dual intra-and extracellular roles that are present in similar quantities in the media and lysate may fail to be identified by this approach. Examples of such proteins include the unconventionally secreted proteins fibroblast growth factor 2 [221], galectin-3 [222], and histone-3A [223]. Furthermore, as this analysis is limited to proteins detected in the conditioned media, secreted proteins that reside on the cell surface or remain insoluble in the subcellular matrix will also be missed by this approach. Thus, the introduction of additional biochemical procedures to isolate these proteins will be required to gain a more comprehensive view of the full secretome.

Chapter 4
Identification of LMAN1 and SURF4 Dependent Secretory Cargoes

4.1 Abstract

Most secreted proteins traverse the intracellular secretory pathway, beginning with co-translational insertion into the endoplasmic reticulum (ER) lumen followed by transport to the Golgi and delivery to the cell surface, various intracellular organelles, or the extracellular space via coat protein complex II (COPII) vesicles/tubules. A subset of secreted proteins is actively recruited to COPII vesicles/tubules through interaction with specific cargo receptors, including LMAN1 or SURF4. However, only a limited number of cargoes have been identified for each of these cargo receptors. We now report mass spectrometry analysis of conditioned media and cell lysates from HuH7 cells CRISPR targeted to inactivate the *LMAN1* or *SURF4* gene to identify the broad secretory cargo repertoire for each of these cargo receptors. We found that LMAN1 has limited clients in HuH7 cells whereas SURF4 traffics a broad range of cargoes. Analysis of putative SURF4 cargoes suggests that cargo recognition is governed by complex mechanisms.

4.2 Introduction

Approximately a third of the proteins encoded by the mammalian genome contain a putative signal peptide that directs co-translational insertion into the endoplasmic reticulum (ER), from where they are subsequently trafficked to the plasma membrane, various intracellular organelles, or secreted into the extracellular space [183, 224]. Properly folded proteins destined

for secretion are transported from the ER to the Golgi via coat protein complex II (COPII) coated vesicles or tubular structures [42, 170, 225]. Entry into COPII vesicles is thought to occur passively via bulk flow or through active recruitment and concentrated into COPII vesicles [29]. Transmembrane cargoes can interact directly with the cargo selective COPII component SEC24 on the cytoplasmic face of the ER, whereas soluble proteins (those without a transmembrane domain) are restricted to the ER lumen and thus would require a cargo receptor to bridge this interaction. Relatively few cargo receptors have been discovered in mammalian cells to date and identifying the full cargo repertoire of these receptors is a major outstanding question.

LMAN1, also known as ERGIC-53, is a 53 kDa protein that localizes to the ER and ER-Golgi intermediate compartment (ERGIC) [126]. LMAN1 has been shown to function as a cargo receptor for coagulation factors V (F5) and VIII (F8) [133, 226], alpha-1-antitrypsin (SERPINA1) [138], Mac-2 binding protein (Mac-2BP) [139], matrix metalloproteinase-9 (MMP-9) [140], cathepsin C (CTSC) [135], cathepsin Z (CTSZ) [136], and membrane protein γ-aminobutyric acid type A receptors (GABAARs) [137]. *In vivo* studies in mice have confirmed the dependence of SERPINA1, but not CTSC and CTSZ, on LMAN1 for secretion [227]. No common LMAN1 binding motif has been identified. It is unclear whether there are other cargoes beyond those listed above that require LMAN1 for efficient secretion from the ER.

Another recently identified mammalian cargo receptor, SURF4, is a 29 kDa protein with multiple transmembrane domains that also localizes to the ER and ERGIC. SURF4 is highly conserved, with clear orthologs in yeast (Erv29p), *C. elegans* (SFT-4), and *Drosophila* [157]. Erv29p is a well characterized cargo receptor responsible for the secretion of yeast pro-α-factor, carboxypeptidase Y, and proteinase A [156, 228, 229]. Multiple SURF4 cargoes in mammals have been identified to date, including PCSK9 [125, 161, 180], apolipoprotein B (APOB) [160,

162, 180, 230], growth hormone [163], dentin sialophosphoprotein (DSPP) [163], amelogenin [163], erythropoietin [164], pathogenic SERPINA1 polymers [165], sonic hedgehog [166], proinsulin [167], and the lysosomal proteins progranulin and prosaposin [168]. Two SURF4 binding motifs on cargoes have been proposed, including an ER-ESCAPE tripeptide motif downstream of the signal peptide sequence [163] and a Cardin-Weintraub motif [166, 231], though not all of the putative cargoes listed above carry one of these motifs, suggesting the presence of additional determinants for recognition by SURF4 [161].

Previous attempts to define a comprehensive cargo repertoire for LMAN1 and SURF4 have revealed few additional cargoes. Using an *in vitro* vesicle formation assay and label-free mass spectrometry in cells depleted for LMAN1 or SURF4, Huang et al [149] described 4 and 17 novel cargoes, respectively for these receptors (out of 815 identified and quantified proteins). Using SILAC labeling and mass spectrometry analysis of conditioned media following SURF4 knockdown, Gomez-Navarro and colleagues [161] identified in 10/344 and 18/171 proteins as potential SURF4 cargoes in HEK293 and HuH7 cells, respectively.

We recently reported mass spec analysis of conditioned media and cell lysates from cultured cells, demonstrating that calculated media to lysate (M/L) ratios provided improved sensitivity and specificity for the identification of secreted proteins [232]. We now report the application of this approach to determine the cargo repertoire of LMAN1 and SURF4 in HuH7 cells.

4.3 Materials and Methods

4.3.1 Cell culture

HuH7 cells [233] were cultured in DMEM Glutamax media supplemented (ThermoFisher Scientific, Waltham MA, 10569-044) with 10% fetal bovine serum (MilliporeSigma, Burlington

MA, F8067) and penicillin/streptomycin (ThermoFisher Scientific, Waltham MA, 15140-122). Cells were passaged every 3-4 days and maintained between 20-80% confluence.

4.3.2 CRISPR mediated inactivation of LMAN1 and SURF4

On day 0, cells were seeded at 20% confluence and infected with the lentivirus, pLentiCRISPRv2 (Addgene #52961, a gift from Feng Zhang [234]) engineered to deliver Cas9 and a sgRNA targeting *LMAN1* (LMAN1g1: CCCCTTACACTATAGTGACG), *SURF4* (or SURF4g6: TCCGAGCTGCATGTACTGTT), or a nontargeting sgRNA (NTg1:GTTCATTTCCAAGTCCGCTG) as previously described [235]. On day 1, the medium was exchanged and 1 µg/ml puromycin (MilliporeSigma, Burlington MA, P8833) was added for 48 hours. Following selection, surviving cells were passaged every 3-4 days until day 14 to allow for gene editing and protein turnover. On day 14, cells were washed three times with phosphate-buffered saline (PBS, ThermoFisher Scientific, Waltham MA, 10010-023) prewarmed to 37ºC and switched to serum-free, phenol red-free DMEM (ThermoFisher Scientific, Waltham MA, 31053-036) for 12 hours.

4.3.3 Conditioned media and cell lysate collection

Conditioned media and cell lysates were collected as previously described [232]. Briefly, conditioned media were collected from the cell culture dish, centrifuged at 2,500g at 4ºC for 15 minutes to pellet cell debris, ultracentrifuged at 120,000g at 4ºC for 90 minutes to remove exosomes [199], and concentrated using a 3 kDa molecular weight cutoff concentrator (MilliporeSigma, Burlington MA, UFC900324). Cell lysates were collected in 2 ml of RIPA buffer (Thermo Scientific, 89900) containing a protease inhibitor cocktail (cOmplete™, Mini Protease Inhibitor Cocktail, 11836153001). Cell suspensions were sonicated, rotated end-over-

end for 1 hour, and centrifuged at 21,000g at 4ºC for 45 minutes. Supernatants were then transferred to a new Eppendorf tube. Protein concentration in the conditioned media and lysates were determined by DC protein assay (Bio-Rad, Hercules CA, 500-011).

4.3.4 Immunoblotting

Cell lysates (10 µg per sample) collected as above were resolved in 4-20% Tris-glycine gel as previously described [180]. Proteins were detected with antibodies against LMAN1 (Abcam, Cambridge UK, ab125006, 1:1000) or GAPDH (Abcam, Cambridge UK, ab181602, 1:10000).

4.3.5 Mass spectrometry, protein identification, and protein quantification

Mass spectrometry, protein identification, and protein quantification were performed as previously described [232]. Briefly, 75 µg of each lysate and 75 µg of each medium sample were proteolyzed, labeled with tandem mass tags (TMT) 10-plex according to manufacturer's protocol, and subjected to liquid chromatography-mass spectrometry analysis. Raw mass spectrometry files were converted into open mzML format and analyzed using the FragPipe (https://fragpipe.nesvilab.org/) computational platform with the default TMT10-MS3 workflow. For proteins that were identified and quantified in both the media and lysate fractions, a M/L ratio was calculated using absolute intensity values. Signal peptide and transmembrane domain annotations were obtained from the UniProt database [236].

4.3.6 Statistical analyses

The limma statistical package was used for comparison of protein abundance or M/L ratios between NTg1 and LMAN1g1 or SURF4g6 treated cells using log2-transformed protein abundance or M/L ratios as input [209]. P-values were adjusted for multiple hypothesis testing

using the Benjamini & Hochberg method. An adjusted p-value (q-value) of 0.05 or less is considered statistically significant.

4.3.7 Analysis of ER-ESCAPE motifs

The human proteome reference database was downloaded from Uniprot [236]. Proteins were filtered for the presence of a signal peptide and absence of a transmembrane domain(s). The tripeptide motif as proposed by Yin et al [163] was extracted from the protein sequence. For proteins with a conventional signal peptide, the tripeptide motif is defined as the first three amino acid residues downstream of the annotated signal peptide sequence. For proteins with a propeptide domain, the tripeptide motif is defined as the three amino acid residues downstream of the propeptide cleavage site. For proteins with an uncleaved signal peptide (based on Uniprot annotation), the tripeptide motif is assigned as the first three amino acid residues of the protein. Base on the classification system proposed by Yin et al [163], we assigned a score for each residue of the tripeptide motif ranging from -2 to 2: 2 for a “very good” amino acid, 1 for a “good” amino acid, 0 for a neutral amino acid, -1 for a “not good” amino acid, and -2 for a “bad” amino acid. The tripeptide score reported in this study is the sum of the scores of each individual residue. For this analysis, SURF4 cargoes include putative SURF4 cargoes identified in this study as well as all previously reported cargoes.

4.4 Results

4.4.1 Identification of bona fide secreted proteins by analysis of cell lysates and conditioned media

To identify secreted proteins that depend on either LMAN1 or SURF4 for secretion, *LMAN1* and *SURF4* deficient HuH7 cells were generated by CRISPR mediated gene editing. We

then collected conditioned media and cell lysates from *LMAN1* deficient cells, *SURF4* deficient cells (SURF4g6), or control cells (NTg1) for analysis by TMT mass spectrometry (n=3 per group) (**Figure 4-1A**).

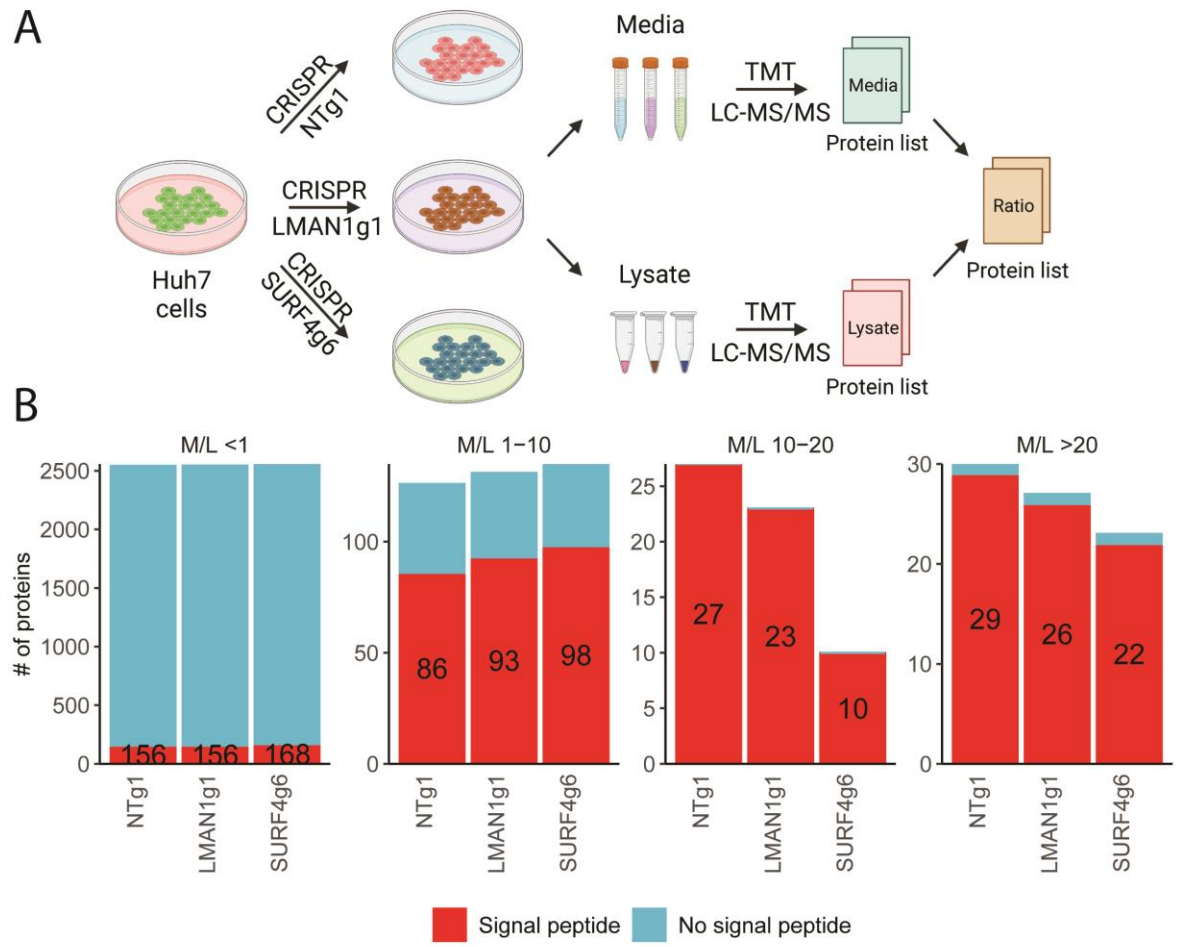

Figure 4-1: Identifying proteins dependent on LMAN1 or SURF4 for efficient secretion. (A) Experimental design to identify LMAN1 and SURF4 cargoes in the human hepatoma cell line (HuH7). HuH7 cells were infected with lentiviruses delivering CRISPR/Cas9 and guide RNAs targeting either LMAN1 (LMAN1g1), SURF4 (SURF4g6) or a nontargeting control (NTg1). Following selection, cells were cultured for 2 weeks before being switched to serum free media for 12 hours. Conditioned media and cell lysates were collected for protein identification and quantification by tandem mass tag (TMT) mass spectrometry (MS). A protein abundance ratio was calculated for each protein that was detected in both the media and lysate fractions. **(B)** Number of proteins with (red) and without (blue) a signal peptide identified in different M/L fractions. The number of signal peptide-containing proteins is each bar is indicated.

We identified and quantified 5858 and 2947 proteins in the lysate and media fractions, respectively, of which 2726 proteins were identified in both fractions (**Figure 4-2A**). Consistent with our previous observations [232], the majority of identified proteins lacked a signal peptide,

with increasing M/L correlating with an increased proportion of proteins carrying a signal peptide, consistent with the expected enrichment for secretory proteins in the media relative to the cell lysates (**Figure 4-2B**). Comparisons between NTg1, LMAN1g, and SURF4g6 samples revealed that there are fewer proteins with an M/L ratio greater than 10 in LMAN1g1 and SURF4g6 samples, consistent with inhibition of secretion (**Figure 4-1B**).

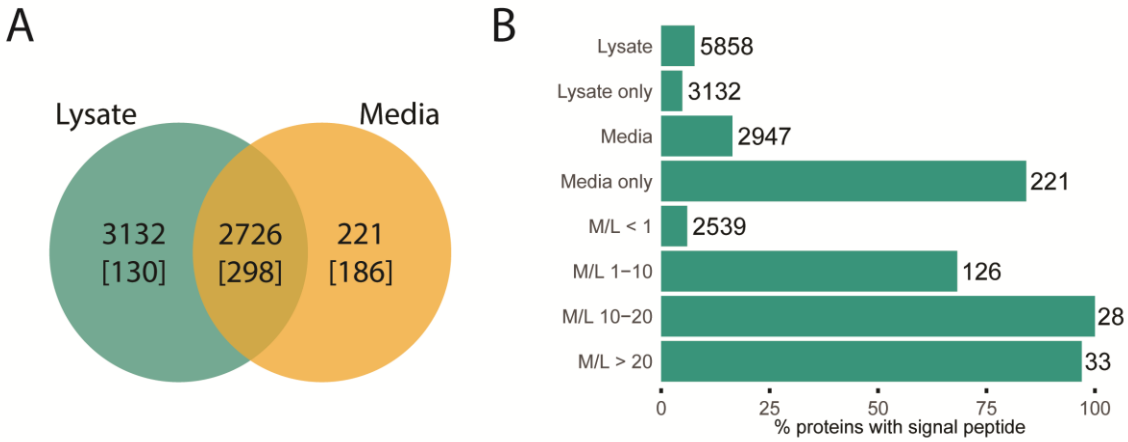

Figure 4-2: Comparison of protein abundance in the media and lysates allows identification of bona fide secreted proteins. (A) Venn diagram of proteins detected in conditioned media, cell lysates, or both fractions. Numbers in the bracket represents the number of proteins carrying an annotated signal peptide in each group. **(B)** Proteins were separated into discrete media/lysate (M/L) ratio intervals and the percentage of proteins carrying a signal peptide was calculated. Numbers at the end of each bar indicate the number of proteins in each bin.

4.4.2 Few LMAN1-dependent cargoes identified in HuH7 cells

To identify secretory proteins that depend on LMAN1 for secretion from the ER, we compared the M/L ratios of proteins collected from cells treated with an LMAN1-targeting sgRNA (LMAN1g1) to those from control cells treated with a non-targeting guide (NTg1). As shown in **Figure 4-3A**, the only protein demonstrating a significantly increased M/L ratio, following *LMAN1* deletion is MCFD2. MCFD2 is a 16 kDa soluble ER luminal protein that forms a complex with LMAN1 and is required for efficient secretion of factors V and VIII. MCFD2 lacks an ER retrieval motif and its retention in the ER and ERGIC depends on its interaction with LMAN1. As expected MCFD2 exhibited an increased M/L ratio in LMAN1

deficient cells (Figure 2A) as a result of decreased intracellular (**Figure 4-4A**) and increased extracellular levels (**Figure 4-4B**), consistent with increased secretion in the absence of LMAN1 as previously reported [145, 149].

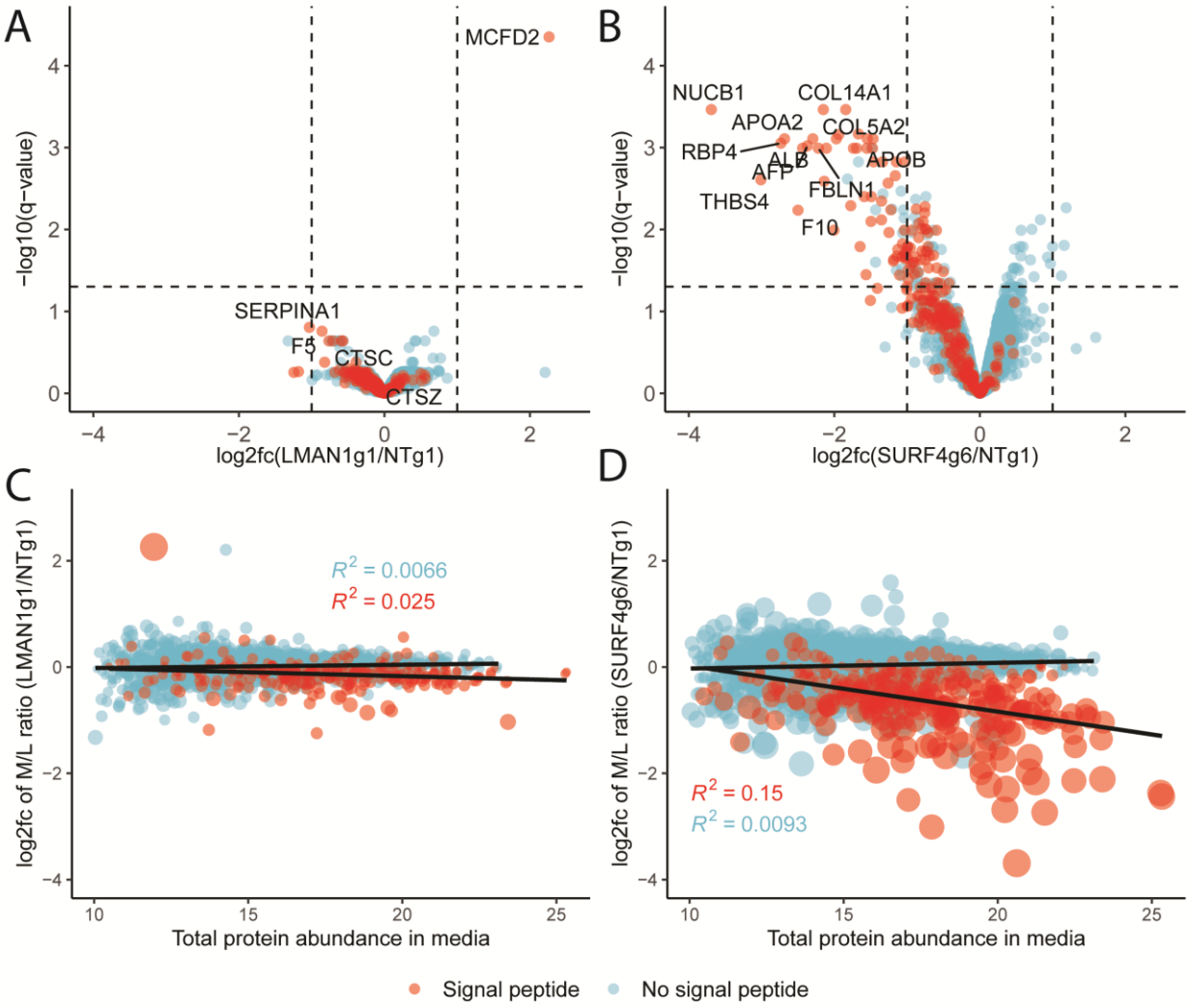

Figure 4-3: Differential effects on protein secretion in HuH7 cells following LMAN1 or SURF4 deletion. (A-B) Volcano plots comparing protein M/L ratios in LMAN1 **(A)** or SURF4 **(B)** deficient cells with those in controls. The log2 fold change (log2fc) and statistical significance are plotted on the x and y-axis, respectively. Proteins with a signal peptide are colored in red and proteins without a signal peptide are colored in blue. Dashed vertical lines represent the log2fc of 1 and -1. Dashed horizontal lines represent the log10(q-value of 0.05). **(C-D)** Comparison of total abundance (estimation from Fragpipe) in the media with LMAN1 **(C)** or SURF4 **(D)** dependency. Trend lines represent linear regression. Dot sizes are proportional to log10(q-value).

Consistent with previously reported *in vivo* data in mice, genetic deletion of LMAN1 in HuH7 cells results in intracellular accumulation of SERPINA1 [151] (**Figure 4-4B**), without

statistically significant reduction in media abundance (**Figure 4-4A**) or M/L ratio (**Figure 4-3A**). Though trends consistent with decreased secretion of additional LMAN1 cargoes including F5, CTSC, and CTSZ were observed, these changes did not reach statistical significance. Taken together, these data suggest that LMAN1 traffics a small number of secretory cargoes in HuH7 cells. Though we cannot exclude technical limitations including incomplete LMAN1 deletion, the marked LMAN1 depletion observed in the deleted cell lysates by mass spec (**Figure 4-4B**) and immunoblotting (**Figure 4-5**) suggests that this latter explanation unlikely.

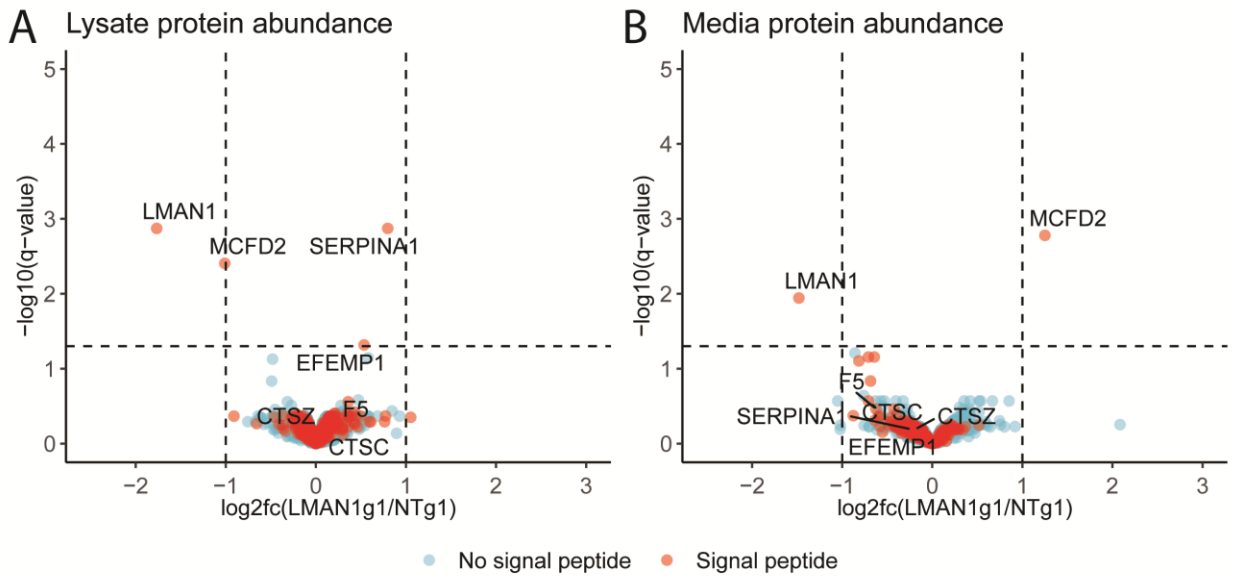

Figure 4-4: Comparison of proteins abundance in the lysates and media between LMAN1 deficient and control cells. Volcano plots representing changes in protein abundance in the cell lysates **(A)** or conditioned media **(B).** The log2 fold change (log2fc) and statistical significance are plotted on the x and y-axis, respectively. Proteins with a signal peptide are colored in red and proteins without a signal peptide are colored in blue. Dashed vertical lines represent the log2fc of 0.5 and -0.5. Dashed horizontal lines represent the log10 (q-value of 0.05).

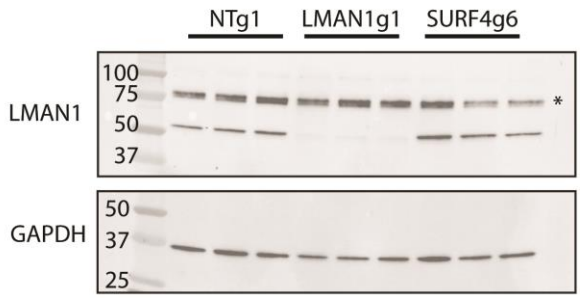

Figure 4-5: Efficient depletion of LMAN1 in cells treated with LMAN1g1 sgRNA. Immunoblots of cell lysates collected from controls (NTg1) and LMAN1 or SURF4 deleted cells. Numbers represent protein molecular weights in kDa. Asterisk indicates nonspecific binding of LMAN1 antibody.

4.4.3 SURF4 traffics a wide range of secretory proteins in HuH7 cells

In contrast to LMAN1 deleted cells, comparison of proteins collected from cells treated with a SURF4 targeting sgRNA (SURF4g6) to those from cells treated with NTg1 sgRNA, identified numerous signal peptide containing proteins exhibiting statistically significant (absolute log2 fold-change (log2fc) >1 and adjust p-value <0.05) increases in lysate abundance, decreases in media abundance (**Figure 4-6**), and associated reduction in M/L ratios in SURF4g6 treated cells (**Figure 4-3B**). These proteins include known SURF4 cargoes such as APOB, APOA1, and APOA2 [160, 230], several putative cargoes identified in other mass spectrometry based studies [149, 161], as well as a number of potential novel SURF4 cargoes (**Figure 4-3B**). We did not identify PCSK9 as a SURF4 cargo in HuH7 cells, which is likely due to upregulation of *PCSK9* expression following *SURF4* deletion [237]. Seven out of 2726 proteins in our dataset were among the known SURF4 cargoes from previously published studies, four of which showed significantly decreased M/L ratio. Comparison of protein abundance in the media with the effect of cargo receptor deletion suggests a greater dependency of highly abundant proteins on SURF4 but not LMAN1 (**Figure 4-3C-D**). A comparison of the putative SURF4 cargoes identified here with those in two other recent reports [149, 161] identifies four secreted proteins shared between

all datasets (**Figure 4-7A**) and ten between two or more datasets (**Figure 4-7A and Figure 4-7C**).

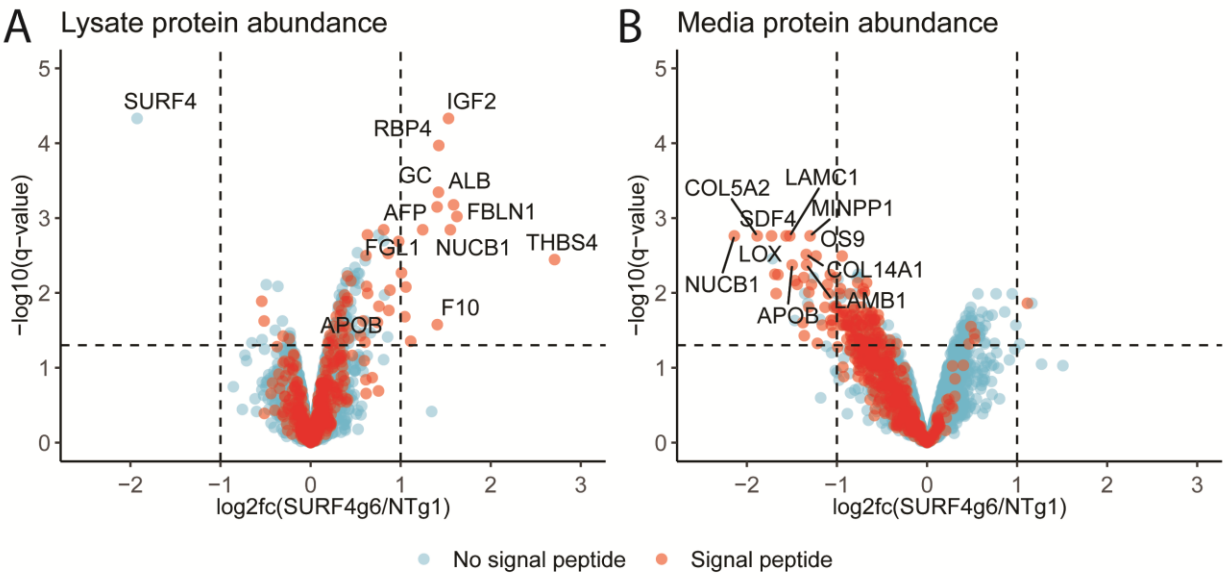

Figure 4-6: Comparison of proteins abundance in the lysates and media between SURF4 deficient and control cells. Volcano plots representing changes in protein abundance in the cell lysates **(A)** or conditioned media **(B).** The log2 fold change (log2fc) and statistical significance are plotted on the x and y-axis, respectively. Proteins with a signal peptide are colored in red and proteins without a signal peptide are colored in blue. Dashed vertical lines represent the log2fc of 0.5 and -0.5. Dashed horizontal lines represent the log10 (q-value of 0.05).

4.4.4 Weak correlation of ER-ESCAPE motif with SURF4 cargoes

Previously, Yin et al [163] proposed a tripeptide motif (ER-ESCAPE) identifying potential SURF4 cargoes. This motif is characterized by a proline residue flanked on either side by a hydrophobic amino acid located immediately downstream of the signal peptide (**Figure 4-7B**). Analysis of this region for candidate SURF4 cargoes identified in our and previous studies revealed a wide range of tripeptide motifs among these cargoes, from excellent fits for the proposed ER-ESCAPE motif such as NUCB1, NUCB2, and SDF4, to poor fits such as COL5A2 and APOB, the latter a SURF4 cargo that has been confirmed in multiple experimental systems [160, 162, 180, 230] (**Figure 4-7C**).

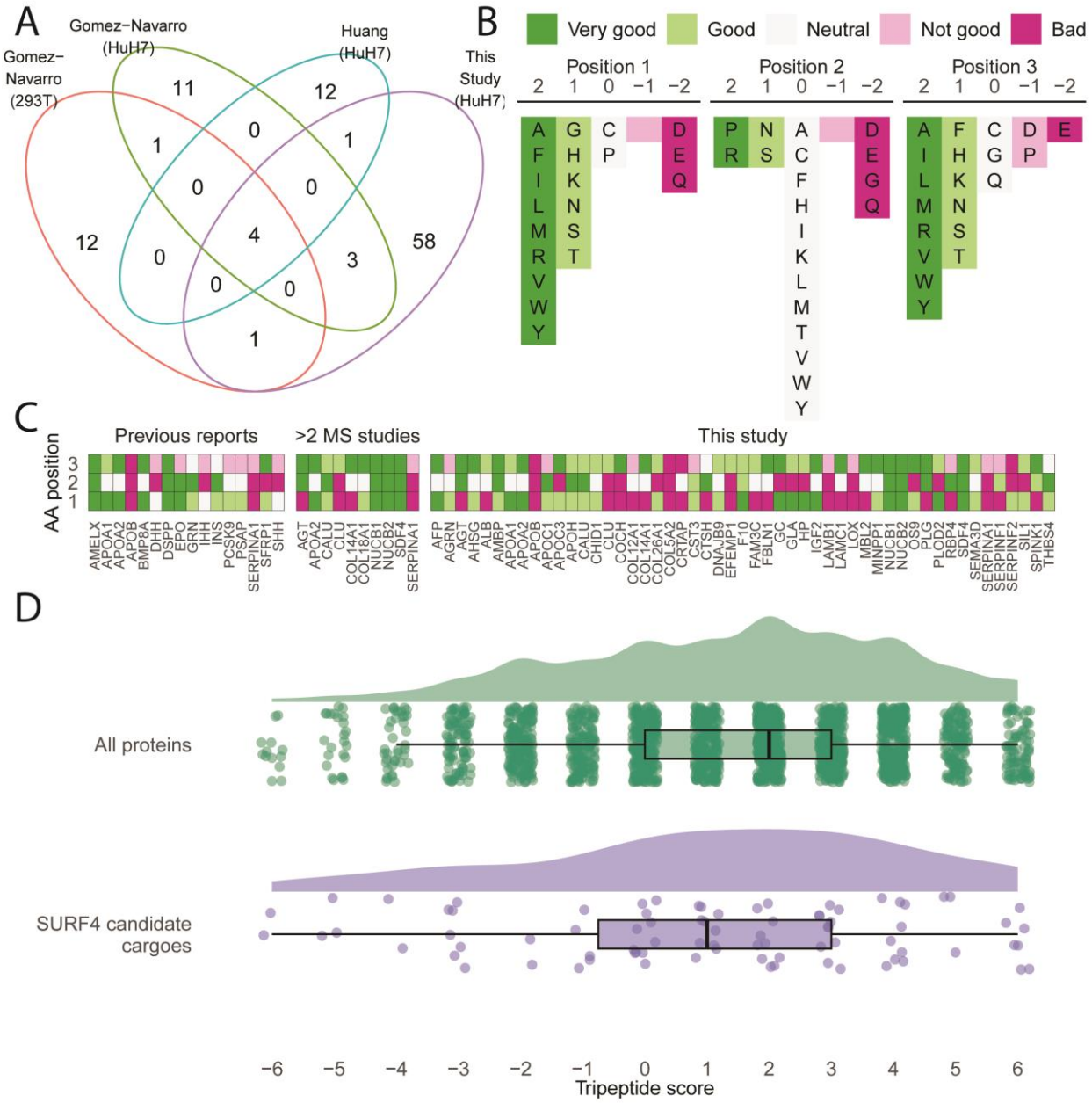

Figure 4-7: Most SURF4 cargoes do not contain an ER-ESCAPE motif [163]. (A) Venn diagram of putative SURF4 cargoes that were identified by our analysis, or previously reported by Gomez-Navarro et al [161] in HEK293T and HuH7 cells following analyses of conditioned media, or by Huang et al [149] using an *in vitro* COPII vesicle formation assay. **(B)** Yin et al [163] proposed a classification system for the three amino acid residues immediate downstream of the signal peptide cleavage site (ER-ESCAPE tripeptide motif). We assigned a score for each amino acid classification as proposed by Yin et al [163]: very good (2), good (1), neutral (0), not good (-1), and bad (-2). **(C)** Tripeptide motifs in previously reported SURF4 cargoes [125, 160-168, 180, 230], in cargoes that were identified in two or more MS-based datasets, or in this study color coded according to (B). **(D)** Distribution of tripeptide score (sum of amino acid scores at each position) in all proteins with an annotated signal peptide in the proteome (see Methods) and putative SURF4 cargoes.

Analysis of the potential ER-ESCAPE motif for the full set of candidate SURF4 cargoes identified in this study and previous reports relative to all proteins in the proteome with an annotated signal peptide is shown in **Figure 4-7**. Scoring was performed as described in Methods and illustrated in **Figure 4-7B.** Comparison of scores for all signal peptide containing proteins and those for putative SURF4 cargoes is shown in **Figure 4-7D**. These data suggest that the ER-ESCAPE motif may only be relevant for a limited subset of SURF4 cargoes. Recent studies also suggest a key role for the Cardin-Weintraub (CW) motif [K/R][K/R][K/R]XX[K/R][K/R] in recognition by SURF4 for cargoes such as the hedgehog family [166, 231]. Analysis of the human proteome [236] identified 267 proteins carrying a CW motif among 7327 proteins that carry a signal peptide or a transmembrane domain. Among the 115 SURF4 candidate cargoes identified by our and previous studies (**Figure 4-7C**), seven carry a CW motif though only one of these seven proteins was identified in our study.

4.5 Discussion

In this study, we applied a broad, whole proteome approach to profile the cargo repertoires for two well-characterized cargo receptors, LMAN1 and SURF4, in the human hepatocellular carcinoma cell line, HuH7. We used our recently reported method [232] to determine M/L ratios for LMAN1 and SURF4 deficient cells compared to controls, observing a striking difference between the cargo repertoire for LMAN1 and SURF4. This approach only identified a limited set of LMAN1-dependent cargoes in HuH7 cells, with SURF4 facilitating secretion for a broad range of proteins. Though a subset of putative SURF4 cargoes carry the previously described ER-ESCAPE [163] or CW motifs [166, 231] potentially mediating interaction with SURF4, many others do not, suggesting a considerably more complex mechanism governing SURF4 cargo recognition.

Surprisingly, our analysis failed to identify any novel LMAN1 cargoes. Though a number of established LMAN1 cargoes (F5, SERPINA1, CTSC, and CTSZ) were detected in our dataset, with several showing a trend toward reduced M/L ratio in LMAN1 deleted cells, none of these changes reached statistical significance. These findings suggest limited statistical power, potentially due to only partial secretion blockade in LMAN1-deficient HuH7 cells, and/or overlap with other cargo receptors. Indeed, humans and mice with loss of function LMAN1 mutations exhibit incomplete reduction in plasma levels for 2 key LMAN1 cargoes, F5 and F8, to only ~10% (humans) and 50% (mice) of those in wild-type controls [143, 151]. Similarly, levels for the LMAN1 cargo SERPINA1 are unchanged in the plasma of *Lman1*$^{-/-}$ mice, though modest accumulation is observed in hepatocytes [151]. Although the lysosomal proteins CTSC and CTSZ have also been proposed as cargoes for LMAN1 [135, 136], our experiment approach would not be expected to distinguish protein accumulation in different cellular compartments (i.e. the ER vs. the lysosome), Taken together, these data suggest that LMAN1 only traffics a small subset of proteins in HuH7 cells.

In contrast to LMAN1, depletion of SURF4 impedes secretion of over 60 proteins, suggesting a much broader cargo repertoire than observed for LMAN1. Notably, the majority of these putative cargoes are exclusive to SURF4 with their levels unchanged in LMAN1 deleted cells. SERPINA1, a previously reported LMAN1 cargo [138], also exhibited reduced M/L ratio following SURF4 deletion, suggesting dependence on SURF4 as well as LMAN1 for efficient secretion. These findings are consistent with a recent report demonstrating a role for SURF4 in the secretion of pathogenic SERPINA1 polymers as well as SERPINA1 monomers, albeit to a lesser extent than LMAN1 [161, 165]. The large difference in the number of potential clients for LMAN1 and SURF4 could also help explain the observed phenotypes in deficient animals, with

normal development and only a mild bleeding defect in LMAN1 deficient mice [151], in contrast to the early embryonic lethality observed in *Surf4*$^{-/-}$ mice [238].

Though the role of the ER-ESCAPE motif proposed by Yin et al [163] in the efficient SURF4-mediated trafficking of PCSK9 and NUCB1 has been confirmed by Gomez-Navarro and colleagues [161], our analysis suggests a more complex process for SURF4 cargo selection. We failed to confirm a general enrichment for the ER-ESCAPE motif among the broader repertoire of potential SURF4 cargoes. However, we cannot exclude the possibility that the tripeptide motif is shifted from the expected starting position in the majority of cargoes, or that interaction is required between the ER-ESCAPE motif and other structural features such as a CW domain [161, 166, 231]. Taken together, these observations suggest that SURF4 cargo recognition is a complex process requiring further study.

Chapter 5
Hepatic Inactivation of Murine *Surf4* Results in Marked Reduction in Plasma Cholesterol[3]

5.1 Abstract

PCSK9 negatively regulates low-density lipoprotein receptor (LDLR) abundance on the cell surface, leading to decreased hepatic clearance of LDL particles and increased levels of plasma cholesterol. We previously identified SURF4 as a cargo receptor that facilitates PCSK9 secretion in HEK293T cells [125]. Here, we generated hepatic SURF4-deficient mice (*Surf4*$^{fl/fl}$ *Alb-Cre*$^{+}$) to investigate the physiologic role of SURF4 *in vivo*. *Surf4*$^{fl/fl}$ *Alb-Cre*$^{+}$ mice exhibited normal viability, gross development, and fertility. Plasma PCSK9 levels were reduced by ~60% in *Surf4*$^{fl/fl}$ *Alb-Cre*$^{+}$ mice, with a corresponding ~50% increase in steady state LDLR protein abundance in the liver, consistent with SURF4 functioning as a cargo receptor for PCSK9. Surprisingly, these mice exhibited a marked reduction in plasma cholesterol and triglyceride levels out of proportion to the partial increase in hepatic LDLR abundance. Detailed characterization of lipoprotein metabolism in these mice instead revealed a severe defect in hepatic lipoprotein secretion, consistent with prior reports of SURF4 also promoting the secretion of apolipoprotein B. Despite a small increase in liver mass and lipid content, histologic evaluation revealed no evidence of steatohepatitis or fibrosis in *Surf4*$^{fl/fl}$ *Alb-Cre*$^{+}$ mice. Acute

[3] This chapter has been published in Elife: Tang VT, McCormick J, Xu B, Wang Y, Fang H, Wang X, Siemieniak D, Khoriaty R, Emmer BT, Chen XW, Ginsburg D. *Hepatic inactivation of murine Surf4 results in marked reduction in plasma cholesterol.* Elife. 2022 Oct 4;11:e82269.

depletion of hepatic SURF4 by CRISPR/Cas9 or liver-targeted siRNA in adult mice confirms these findings. Together, these data support the physiologic significance of SURF4 in the hepatic secretion of PCSK9 and APOB-containing lipoproteins and its potential as a therapeutic target in atherosclerotic cardiovascular diseases.

5.2 Introduction

An elevated plasma level of low-density lipoprotein (LDL) is a major risk factor for atherosclerotic cardiovascular disease [239], which is the leading cause of death worldwide. LDL is derived in the circulation by processing of very-low-density lipoprotein (VLDL) particles, which is synthesized and secreted by the liver. In humans, the major protein component of VLDL is APOB100, which is cotranslationally lipidated in the endoplasmic reticulum (ER) [240]. LDL is cleared from circulation by the LDL receptor (LDLR) on cell surfaces. Proprotein convertase subtilisin/kexin type 9 (PCSK9) is a soluble protein that is secreted by the liver and negatively regulates LDLR abundance by inducing its degradation [241].

Proteins destined for extracellular secretion are transported from the endoplasmic reticulum (ER) to Golgi by COPII coated vesicles/tubules [3, 242]. SEC24 is a key component of the COPII inner coat, which appears to play a primary role in selecting cargo proteins for export from the ER [26]. The mammalian genome encodes 4 paralogs of *Sec24* (*Sec24a-d*) [92]. Mice genetically deficient in SEC24A exhibit moderate hypocholesterolemia due to a selective block in PCSK9 secretion from the ER, resulting in an ~50% reduction in plasma PCSK9 levels [60].

We previously reported a whole genome CRISPR screen in HEK293T cells heterologously expressing PCSK9, identifying Surfeit locus protein 4 (SURF4) as the putative cargo receptor potentially linking PCSK9 within the ER lumen to SEC24A on the cytoplasmic face of the ER membrane [125]. SURF4 is a 29 kDa protein with multiple transmembrane

domains that localize to the ER and ER-Golgi intermediate compartment (ERGIC) [152]. *SURF4* is a homolog of *Erv29p,* a well-characterized cargo receptor in yeast that mediates the ER to Golgi trafficking of pro-α-mating factor [228]. The SURF4 ortholog in *C. elegans*, SFT-4, controls the ER export of the yolk protein VIT-2 [162]. Recent studies in human cells have also implicated SURF4 in the trafficking of other cargoes, including apolipoprotein B (APOB) [159, 162, 230], erythropoietin (EPO) [164], growth hormone, dentin sialophosphoprotein, and amelogenin [163]. A role for SURF4 in APOB secretion was further supported by a recent study in which acute deletion of hepatic *Surf4* in adult mice caused hypocholesterolemia and a reduction in hepatic lipoprotein secretion [160].

To investigate the physiologic significance of SURF4 in the secretion of PCSK9 and other putative cargoes, we previously generated mice with germline deletion of *Surf4*, which resulted in early embryonic lethality [159]. We now report the generation and characterization of mice with *Surf4* selectively inactivated in the liver by combining a conditional *Surf4* allele (*Surf4*^{fl}) with a Cre recombinase expressed under the control of the albumin promoter (*Alb-Cre*). *Surf4*$^{fl/fl}$ *Alb-Cre*$^{+}$ mice exhibit normal development, survival and fertility, with marked plasma hypocholesterolemia associated with a hepatic secretion defect for PCSK9 and APOB-containing lipoproteins without evidence for liver injury. Acute inactivation of hepatic *Surf4* by CRISPR/Cas9 or liver-targeted siRNA in adult mice further confirms these findings and the potential of hepatic SURF4 as a therapeutic target in atherosclerotic cardiovascular disease.

5.3 Materials and Methods

Table 5-1: Key resources used in this study

Reagent type (species) or resource	Designation	Source or reference	Identifiers	Additional information
strain, strain background (*M. musculus*)	$Surf4^{fl}$ mice (C57BL/6J)	[160]		
strain, strain background (*M. musculus*)	Albumin-Cre mice (C57BL/6J)	[62]	JAX 003574	
strain, strain background (*M. musculus*)	$Surf4^{fl/fl}$ Alb-Cre^{+} mice (C57BL/6J)	This paper		
strain, strain background (*M. musculus*)	SpCas9 mice	[243]	JAX 026556	
strain, strain background (*M. musculus*)	C57BL/6J mice	Jackson lab	JAX 0006640	
genetic reagent (include species here)	AAV-Cre-sgRNA	This paper		
antibody	anti-APOB - rabbit polyclonal	Fitzgerald Industries Internationa	70R-15771	WB(1:1000)
antibody	anti-PCSK9 – rabbit polyclonal	abcam	ab31762	WB (1:1000)
antibody	anti-LDLR – rabbit monoclonal	abcam	ab52818	WB (1:1000)
antibody	anti-HSP90 – rabbit monoclonal	Cell Signaling Technology	4877	WB (1:1000)
antibody	anti-SURF4 – rabbit polyclonal	[160]		WB (1:1000)

antibody	anti-APOA1 – rabbit polyclonal	Fitzgerald Industries Internationa	70R-15769	WB (1:1000)
antibody	anti-Albumin – mouse monoclonal	Proteintech	66051	WB (1:1000)
antibody	anti-Tubulin – rabbit polyclonal	Proteintech	10094-1-AP	WB (1:1000)
antibody	anti-GAPDH – rabbit monoclonal	abcam	ab181602	WB (1:1000)
sequence-based reagent	Primers	IDT		Sequences are listed in **Table 5-3**
sequence-based reagent	CRISPR-gRNA	IDT		Sequences are listed in **Table 5-4**
sequence-based reagent	SURF4-targeting siRNA	This paper		Sequences are listed in **Table 5-5**
commercial assay or kit	Cholesterol assay kit	SB-1010-225	Fisher Scientific	
commercial assay or kit	PCSK9 ELSIA	MPC900	R&D Systems	
commercial assay or kit	APOB ELISA	ab230932	abcam	
commercial assay or kit	Triglycerides assay	10010303	Cayman Chemical	
commercial assay or kit	RNeasy Plus Mini Kit	74134	Qiagen	

commercial assay or kit	Power SYBR Green PCR Master Mix	4367659	Invitrogen

5.3.1 Animal care and use

All animal care and use complied with the Principles of Laboratory and Animal Care established by the National Society for Medical Research. Mice were housed in a controlled lighting (12h light/dark cycle) and temperature (22°C) environment and had free access to food (5L0D, LabDiet, St. Louis, MO) and water. All animal protocols in this study have been approved by the Institutional Animal Care and Use Committee (IACUC) of the University of Michigan (protocol number PRO00009304) and the IACUC of Peking University. Both male and female mice were used in this study unless otherwise specified.

5.3.2 Generation of conditional Surf4 knockout mice

The generation of mice carrying a conditional *Surf4* allele in which exon 2 of the gene is flanked by 2 loxP sites (*Surf4*^{fl}) has been previously described [160]. *Surf4*$^{fl/+}$ mice were crossed with mice carrying an *Alb-Cre* transgene [62] to obtain *Surf4*$^{fl/+}$ *Alb-Cre*$^{+}$ mice. These mice were then crossed to *Surf4*$^{fl/fl}$ mice to generate *Surf4*$^{fl/fl}$ *Alb-Cre*$^{+}$ mice. The *Surf4*^{fl} and *Alb-Cre* alleles were maintained on the C57BL/6J background by continuous backcrosses to C57BL/6J mice (0006640, Jackson Laboratory, Bar Harbor ME).

5.3.3 Genotyping assays

Tail clips were obtained from 2 weeks old mice for genomic DNA isolation and genotyping. PCR was performed using Go-Taq Green Master Mix (Promega, Madison, WI) and resulting products were resolved by 3% agarose gel electrophoresis. All primers used for genotyping are listed in **Table 5-3**. Those used for genotyping the *Surf4* locus are also depicted

in **Figure 5-1A**. For the *Alb-Cre* transgene, parental mice were genotyped using promoter-specific *Cre* primers and offspring were genotyped with primers that detect the *Cre* transgene.

5.3.4 Blood and tissue collection

Mice were fasted overnight for up to 16 hours prior to sample collection. For non-terminal experiments, blood was collected from the superficial temporal vein using a 4 mm sterile lancet. For terminal experiments, mice were first euthanized by isoflurane inhalation and blood was drawn from the inferior vena cava using a 23G needle and a 1 ml syringe. Blood was collected into a serum separator tube (365967, BD, Franklin Lakes NJ), allowed to clot at room temperature for at least 10 minutes, and centrifuged at 15,000 g for 10 minutes to separate serum. Sera were aliquoted and stored at -80°C. Liver tissue were collected as previously described [159].

5.3.5 Analysis for sera from Surf4 liver knockout mice

Sera were analyzed by a colorimetric assay for total cholesterol (SB-1010-225, Fisher Scientific, Hampton NH) and by ELISAs for PCSK9 (MPC900, R&D Systems, Minneapolis MN) and APOB (ab230932, abcam, Cambridge UK). Serum lipoprotein fractionation assays were performed at the University of Cincinnati Mouse Metabolic Phenotyping Center. Sera were pooled from 5 mice for each genotype and fractionated by fast liquid protein chromatography (FPLC) into 50 fractions. Cholesterol (NC9343696, Fisher, Hampton NH) and triglyceride (TR213, Randox Laboratories, Crumlin UK) content in each fraction were determined using a microliter plate enzyme-based assay. Liver function tests were performed at the University of Michigan In-Vivo Animal Core (IVAC) with sera collected from individual mice using a Liasys analyzer (AMS Alliance).

5.3.6 Hepatic lipoprotein secretion assay

Hepatic triglyceride secretion assays were performed at the University of Michigan Mouse Metabolic Phenotyping Center as previously described [244]. Blood levels of glucose were measured using a glucometer (Acucheck, Roche, Basel Switzerland) and plasma levels of cholesterol (SB-1010-225, Fisher Scientific, Hampton NH) and triglycerides (10010303, Cayman Chemical, Ann Arbor MI) were determined using colorimetric assay kits. Plasma APOB levels were determined by ELISA (ab230932, abcam, Cambridge UK).

5.3.7 Oral fat tolerance test and lipid flux assay

Oral fat tolerance test and lipid flux assay were performed at the University of Michigan Mouse Metabolic Phenotyping Center. Mice were fasted overnight and ^{3}H triolein-labeled olive oil (0.026 µCi/µl) was given via oral gavage at 5 µl/g of body mass. Blood samples were collected at time 0, 30, 60, 120, and 240 minutes after the gavage via tail vein bleeding. Plasma levels of triglyceride (10010303, Cayman Chemical, Ann Arbor MI) and non-esterified fatty acid (NEFA-HR (H2), Wako Pure Chemical Industries, Ltd, Richmond VA) were determined using the colorimetric assays. Plasma radioactivity, reported as ^{3}H disintegration per minute (dpm), were determined from 2 µl of serum at each time point.

Tissues samples (liver, heart, gastrocnemius muscle, and perigonadal fat) were collected at the 240 minute time point, flash frozen in liquid nitrogen and stored at -80ºC. For the liver, tissue composition was measured using an NMR-based analyzer (EchoMRI^{TM}) immediately upon harvest and prior to freezing. ^{3}H-triolein flux was estimated as previously described [245, 246].

5.3.8 Liver lipid extraction and quantification

Liver lipid extraction and quantification were performed at the University of Cincinnati Mouse Metabolic Phenotyping Center. Lipids were extracted using the Folch's extraction

method as previously described [247]. Levels of cholesterol, triglycerides, free fatty acids, and phospholipids in each sample were quantified using specific colorimetric assays.

5.3.9 Immunoblotting

Lysates were prepared from snap frozen liver tissues and resolved on a 4-20% Tris-glycine gel as previously described [159]. Immunoblots were probed with antibodies against APOB (70R-15771, 1:1000, Fitzgerald Industries International, Acton MA), PCSK9 (ab31762, 1:1000, abcam, Cambridge UK), LDLR (ab52818, 1:1000, abcam, Cambridge UK), HSP90 (4877, 1:1000, Cell Signaling Technology, Danvers MA), SURF4 [160], APOA1 (70R-15769, 1:1000, Fitzgerald Industries International, Acton MA), Albumin (66051, 1:1000, Proteintech, Rosemont, IL), Tubulin (10094-1-AP, 1:1000, Proteintech, Rosemont, IL), and GAPDH (ab181602, 1:5000, abcam, Cambridge UK). For endoglycosidase H assays, 30 µg of lysate was analyzed as previously described using the above antibodies [125].

5.3.10 Histology

Tissue processing, embedding, sectioning, hematoxylin and eosin (H&E), and picrosirius red staining were performed at the University of Michigan In-Vivo Animal Core (IVAC). Slides were reviewed by an investigator blinded to the genotype.

5.3.11 Analysis of liver mRNA

Liver RNA was isolated from tissue using an RNeasy Plus Mini Kit according to the manufacturer's instructions (74134, Qiagen, Hilden Germany) and reverse transcription was performed using oligo$(dT)_{12\text{-}18}$ primers (18418012, Invitrogen, Waltham MA). Quantitative PCR reactions were performed using Power SYBR Green PCR Master Mix (4367659, Invitrogen, Waltham MA) and primers listed in **Table 5-3**. Total *Surf4* mRNA abundance was calculated using data from primers that bind to exon 5 and 6 on the *Surf4* transcript. Abundance of *Surf4*

mRNA that contains exon 2 was obtained using primers specific to exon 2 and exon 3 of the transcript. Normalized transcript abundance was calculated by the $2^{-\Delta\Delta Ct}$ method using *Gapdh* and *Rpl37* as housekeeping controls.

Library preparation and next generation sequencing was performed at the University of Michigan Advanced Genomics Core. Demultiplexed fastq files were aligned against the mouse reference genome (GRCm38.92) using STAR [248] and quantified with RSEM [249]. Differential expression analysis was performed by DESeq2 [250]. Sequencing coverage for the *Surf4* transcript was analyzed using the ggsashimi package [251]. Raw and processed sequencing data have been deposited to GEO (accession number GSE214393).

5.3.12 Acute inactivation of hepatic Surf4 in adult mice

Hepatic *Surf4* was selectively inactivated in adult mice by injection of adeno-associated virus (AAV) delivering a hepatocyte-specific Cre and a guide RNA targeting *Surf4* or *LacZ* (control) (**Table 5-4**) into a Cre-dependent spCas9 knockin mice [243] as previously described [160]. Blood samples were collected as previously described [160] and plasma PCSK9 concentrations were measured by the commercial kit (CY-8078 of MBL) according to the manufacturer's protocol.

To deplete *Surf4* mRNA, N-acetylgalactosamine (GalNAc) conjugated siRNA oligos are synthesized to ensure liver targeting. SiRNA targeting murine *Surf4* or GalNAc conjugated siRNA with scrambled sequence (siCTL) were injected into 6 weeks old male mice subcutaneously with concentrations indicated in the figures. Seven days after injection, blood samples were collected by tail vein, and then centrifuged at 6000 rpm, 4℃ for 10 min to harvest plasma. The plasma PCSK9 concentrations were measured by the commercial kit (CY-8078 of MBL) according to the manufacturer's protocol. SiRNA sequences are listed in **Table 5-5**.

5.4 Results

5.4.1 Liver-specific deletion of Surf4 is compatible with normal development and survival in mice

To investigate long term *Surf4* inactivation in hepatocytes *in vivo*, we generated mice with the *Surf4* gene genetically inactivated specifically in the liver by combining a previously reported conditional *Surf4* allele (in which *Surf4* exon 2 is flanked by loxP sites, denoted $Surf4^{fl}$) [160] with a *Cre* recombinase transgene under control of the *Albumin* promoter (*Alb-Cre*). $Surf4^{fl/fl}$ $Alb\text{-}Cre^{+}$ mice were observed at the expected Mendelian ratio (**Table 5-2**). Both male and female $Surf4^{fl/fl}$ $Alb\text{-}Cre^{+}$ mice were fertile and produce offspring of the predicted genotypes at expected Mendelian ratios (**Table 5-2**).

Table 5-2: Genotype distribution of offspring of $Surf4^{fl/fl}$ $Alb\text{-}Cre^{+}$ and $Surf4^{fl/+}$ $Alb\text{-}Cre^{-}$ intercrosses

Genotype (Expected)	$Surf4^{fl/+}$ $Alb\text{-}Cre^{-}$ (25%)	$Surf4^{fl/+}$ $Alb\text{-}Cre^{+}$ (25%)	$Surf4^{fl/fl}$ $Alb\text{-}Cre^{-}$ (25%)	$Surf4^{fl/fl}$ $Alb\text{-}Cre^{+}$ (25%)	p (χ^2)
Mating	♂ $Surf4^{fl/fl}$ $Alb\text{-}Cre^{-}$ X ♀ $Surf4^{fl/+}$ $Alb\text{-}Cre^{+}$				
	61 (25.8%)	71 (30.1%)	60 (25.4%)	44 (18.6%)	0.096
Mating	♂ $Surf4^{fl/+}$ $Alb\text{-}Cre^{-}$ X ♀ $Surf4^{fl/fl}$ $Alb\text{-}Cre^{+}$				
	8 (18.6%)	9 (20.9%)	12 (27.9%)	14 (32.6%)	0.549

Excision of exon 2 is predicted to result in a frameshift mutation and the generation of a premature termination codon 8 base pairs downstream of the new exon1-3 junction (**Figure 5-1A**). Analysis of genomic DNA collected from mouse tails and livers of $Surf4^{fl/fl}$ $Alb\text{-}Cre^{+}$ mice demonstrated efficient Cre-mediated excision of *Surf4* exon 2 only in the liver (**Figure 5-1B**), with the level of exon 2 containing *Surf4* transcripts in $Surf4^{fl/fl}$ $Alb\text{-}Cre^{+}$ livers reduced to ~5% of controls (**Figure 5-1C**). This residual unexcised *Surf4* mRNA is likely derived from nonhepatocyte cell types in the liver. Quantitative PCR of liver cDNA using primers outside of

exon 2 demonstrated a 38% reduction in total *Surf4* mRNA transcript in *Surf4*$^{fl/fl}$ *Alb-Cre*$^{+}$ mice compared to *Surf4*$^{fl/fl}$ *Alb-Cre*$^{-}$ littermates, likely due to nonsense mediated mRNA decay [252]. Analysis of *Surf4* mRNA transcripts by RNA sequencing confirmed the expected reduction of reads spanning the exon 1-2 and exon 2-3 junctions in *Surf4*$^{fl/fl}$ *Alb-Cre*$^{+}$ livers compared to controls (150 and 148 versus 2623 and 2490 reads, respectively, **Figure 5-1D** and **Figure 5-2**). Consistent with the qPCR data and incomplete nonsense mediated mRNA decay, we identified 928±51 reads mapping to the exon1-exon3 junction of the *Surf4* mRNA in liver from *Surf4*$^{fl/fl}$ *Alb-Cre*$^{+}$ mice and zero in *Surf4*$^{fl/fl}$ *Alb-Cre*$^{-}$ samples (**Figure 5-2**). This residual exon 2 excised mRNA in *Surf4*$^{fl/fl}$ *Alb-Cre*$^{+}$ liver contains a premature stop codon near the start of the SURF4 coding sequence (codon 23 of 270), which is expected to be translated into a nonfunctional, truncated protein (**Figure 5-1A**).

We also detected 404±39 reads (none in controls) mapping to the exon1-4 junction of an alternatively spliced *Surf4* mRNA in *Surf4*$^{fl/fl}$ *Alb-Cre*$^{+}$ samples (**Figure 5-1D**). Exclusion of exon 2 and 3 is predicted to restore the reading frame and result in the production of an internally deleted SURF4, missing ~1/3 of the full length sequence (**Figure 5-1A)**. Though also likely to be nonfunctional, residual activity and/or a dominant-negative effect of this internally deleted SURF4 cannot be excluded.

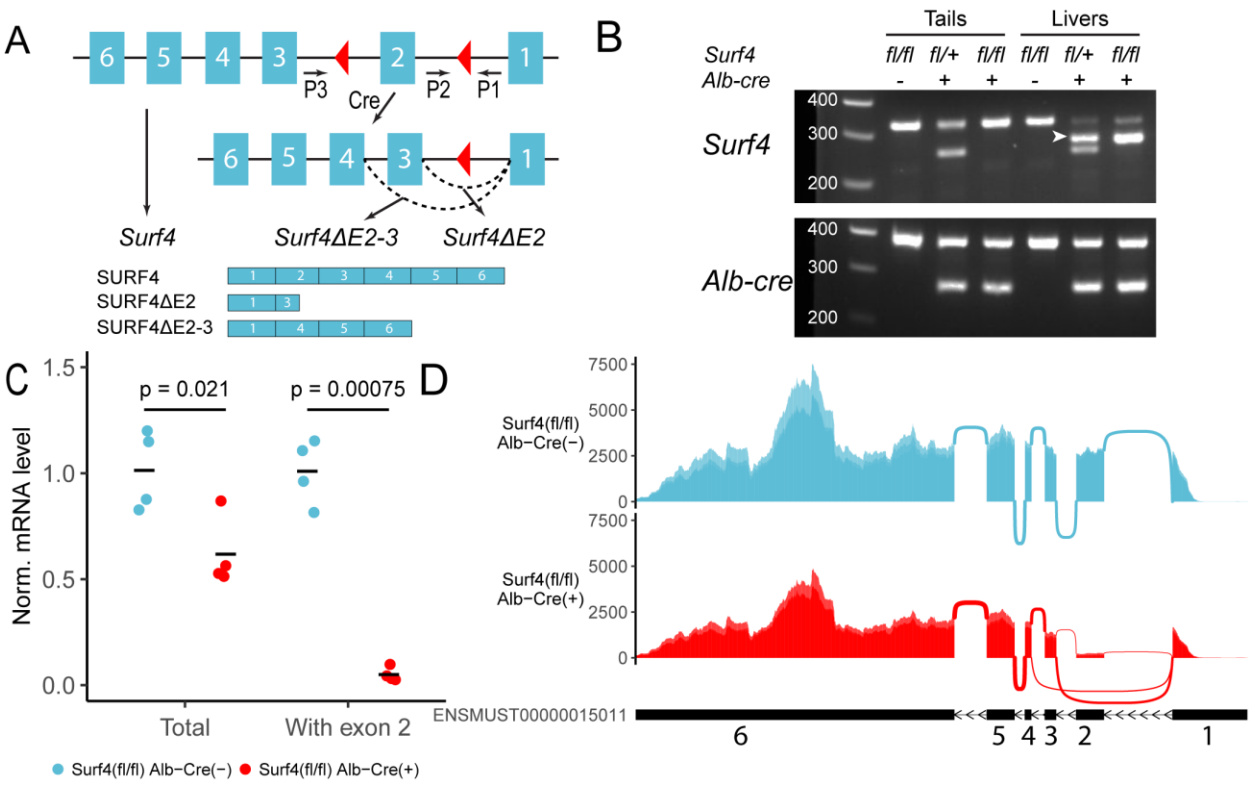

Figure 5-1: Generation of hepatocyte-specific *Surf4* deficient mice. (A) Schematic presentation of the *Surf4* conditional allele. Blue rectangles represent exons and black line segments represent introns. Red triangles denote loxP sites. Expression of a *Cre* recombinase leads to excision of exon 2, which results in the generation of a *Surf4* mRNA lacking exon 2 (*Surf4ΔE2)* or both exon 2 and 3 (*Surf4ΔE2-3*). *Surf4ΔE2* mRNA is translated into a truncated SURF4 that is only 22 amino acids in length. *Surf4ΔE2-3* mRNA restores the reading frame, producing an internally truncated protein missing the 88 amino acids encoded by exon 2 and 3. P1, P2, P3 indicate the positions for *Surf4* genotyping primers. Dashed lines represent splicing events. Exons and introns are not drawn to scale. **(B)** Agarose gel electrophoresis of PCR products generated using genomic DNA (gDNA) isolated from mouse tails and livers and primers P1-3 shown in **(A)**. For *Surf4* genotyping, the wild type allele produces a smaller PCR product whereas the conditional allele produces a larger amplicon. Excision of exon 2 results in the generation of a PCR product of intermediate size (white arrowheads) that is present in gDNA isolated from the livers of *Alb-Cre*$^{+}$ mice only. For *Alb-Cre* genotyping, presence of the *cre* transgene results in a smaller PCR product. The upper band represents the amplification of an internal control. **(C)** Quantification of Normalized (Norm.) *Surf4* mRNA abundance by quantitative PCR of liver cDNA from control (*Surf4*$^{fl/fl}$ *Alb-Cre*$^{-}$) and *Surf4*$^{fl/fl}$ *Alb-Cre*$^{+}$ mice. Crossbars represent the mean normalized abundance in each group. The denoted p-value was calculated by two-sided Student's t-test. **(D)** Density plots of RNA-seq reads mapping along exon and exon-exon junctions of *Surf4* mRNA. *Surf4*$^{fl/fl}$ *Alb-Cre*$^{+}$ samples have lower read counts due to incomplete nonsense mediated mRNA decay. Arcs between exons represent splicing events and line thickness is proportional to read count. Exact read counts are presented in **Figure 5-2**.

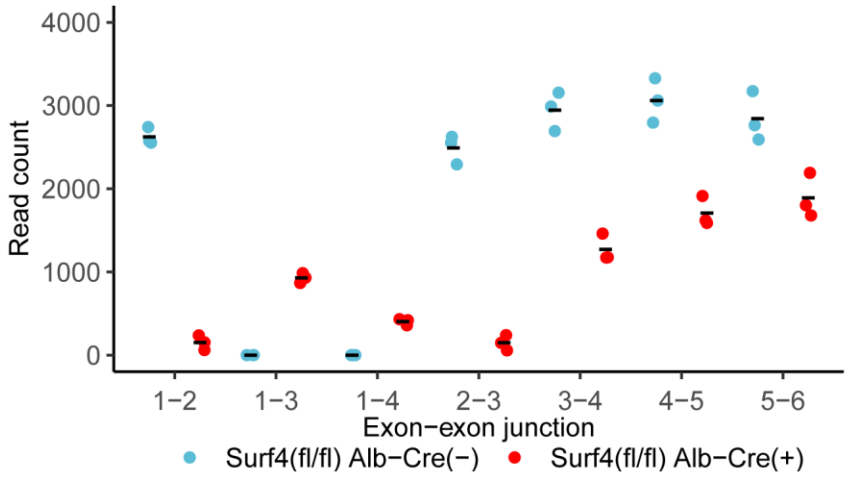

Figure 5-2: Read counts mapping to exon-exon junctions along the *Surf4* transcript based on RNA-sequencing data.

5.4.2 Reduced circulating PCSK9 and increased LDLR levels in Surf4$^{fl/fl}$ Alb-Cre^{+} mice

We previously demonstrated a key role of SURF4 in the efficient trafficking of PCSK9 heterologously expressed in HEK293T cells [125]. To test the dependence of PCSK9 secretion on SURF4 *in vivo*, we first examined steady state serum PCSK9 levels in *Surf4$^{fl/fl}$ Alb-Cre^{+}* and *Surf4$^{fl/fl}$ Alb-Cre^{-}* control mice. Serum PCSK9 levels were reduced by ~60% in *Surf4$^{fl/fl}$ Alb-Cre^{+}* compared to *Surf4$^{fl/fl}$ Alb-Cre^{-}* mice (from 46.0±19.0 ng/mL to 17.8±6.42 ng/mL, **Figure 5-3A**), though PCSK9 accumulation was not observed in liver lysates (**Figure 5-3-C**). Quantitative RT-PCR revealed that *Pcsk9* mRNA levels were also unchanged (**Figure 5-4**), consistent with a defect in PCSK9 protein secretion rather than gene expression as the cause for decreased plasma PCSK9 levels.

In contrast to the above findings, Wang et al reported no change in plasma PCSK9 levels in *Surf4$^{fl/fl}$ Alb-Cre^{+}* mice [230]. To address this issue, and to exclude complex adaptation to hepatic SURF4 deletion induced *in utero*, we acutely inactivated hepatic *Surf4* in adult mice using a previously reported Cas9 mouse system [160]. Analyses utilizing 3 different *Surf4* targeting sgRNAs demonstrated a reproducible ~40% reduction in plasma PCSK9 levels for all

mice receiving *Surf4* targeting sgRNA compared to control mice (**Figure 5-3D**), consistent with our findings in *Surf4*$^{fl/fl}$ *Alb-Cre*$^{+}$ mice.

Since PCSK9 is a negative regulator of LDLR, we next quantified LDLR levels in liver lysates collected from control and *Surf4*$^{fl/fl}$ *Alb-Cre*$^{+}$ mice. As shown in **Figure 5-3E-F**, *Surf4*$^{fl/fl}$ *Alb-Cre*$^{+}$ mice exhibit an ~1.5-fold increase in LDLR abundance in liver lysates compared to controls, consistent with the observed ~60% reduction in circulating PCSK9 level.

5.4.3 Marked reduction of plasma cholesterol in Surf4$^{fl/fl}$ *Alb-Cre*$^{+}$ *mice*

Humans with heterozygous loss of function mutations in *PCSK9* exhibit an ~28-40% reduction in circulating cholesterol [253-255] with a similar reduction observed in *Pcsk9*$^{+/-}$ mice [256]. Though *Surf4*$^{fl/fl}$ *Alb-Cre*$^{+}$ mice exhibit similar reductions in PCSK9, total serum cholesterol is markedly reduced (from 54.3±15.1 mg/dL in *Surf4*$^{fl/fl}$ *Alb-Cre*$^{-}$ mice to 9.51±2.6 mg/dL in *Surf4*$^{fl/fl}$ *Alb-Cre*$^{+}$ mice) (**Figure 5-3G**). No change in cholesterol was observed in the limited numbers of mice haploinsufficient for *Surf4* in the liver (*Surf4*$^{fl+l}$ *Alb-Cre*$^{+}$). Analysis of fractionated pooled sera demonstrated marked reductions in cholesterol and triglyceride content in *Surf4*$^{fl/fl}$ *Alb-Cre*$^{+}$ mice among all 3 major classes of lipoproteins – very low density lipoprotein (VLDL), low density lipoprotein (LDL), and high density lipoprotein (HDL) (**Figure 5-3H**). This striking hypocholesterolemia phenotype is sustained through at least 1 year of age (**Figure 5-5B**), with no difference in body mass between *Surf4*$^{fl/fl}$ *Alb-Cre*$^{-}$ and *Surf4*$^{fl/fl}$ *Alb-Cre*$^{+}$ littermates (**Figure 5-5A** and **Figure 5-5C**).

Consistent with a role of SURF4 in the ER export of APOB-containing lipoproteins [160], the mean serum APOB level in fasted *Surf4*$^{fl/fl}$ *Alb-Cre*$^{+}$ mice was 3.88±2.63 mg/mL, a >98% reduction compared to *Surf4*$^{fl/fl}$ *Alb-Cre*$^{-}$ mice (300±157mg/mL, **Figure 5-6A**). Western blotting demonstrated a trend towards an accumulation of APOB in the livers of *Surf4*$^{fl/fl}$ *Alb-*

Cre^{+} mice compared to littermate controls, predominantly in an endoglycosidase H (endo H) sensitive form (**Figure 5-6B-C)**, indicative of ER retention [257].

We next examined hepatic triglyceride secretion in *Surf4*$^{fl/fl}$ *Alb-Cre*$^{+}$ and control mice. For this experiment, mice were fasted to remove intestinal absorption of dietary fat and tissue lipid uptake was blocked by administration of a lipoprotein lipase inhibitor, with liver triglyceride output subsequently monitored by sampling of plasma triglycerides over 24 hours. Following fasting and inhibition of triglyceride hydrolysis, blood glucose levels fell equivalently between *Surf4*$^{fl/fl}$ *Alb-Cre*$^{+}$ and *Surf4*$^{fl/fl}$ *Alb-Cre*$^{-}$ littermates (**Figure 5-6D**). Although serum cholesterol and triglyceride levels steadily increased in both groups over time, both levels were consistently and significantly lower in *Surf4*$^{fl/fl}$ *Alb-Cre*$^{+}$ mice compared to littermate controls (**Figure 5-6E-F**). Following an initial decrease in the first hour after lipoprotein lipase inhibition, serum APOB levels steadily rose in control mice (**Figure 5-6G)**. In contrast, serum APOB levels were markedly reduced at baseline in *Surf4*$^{fl/fl}$ *Alb-Cre*$^{+}$ mice, and showed minimal increase after lipoprotein lipase inhibition.

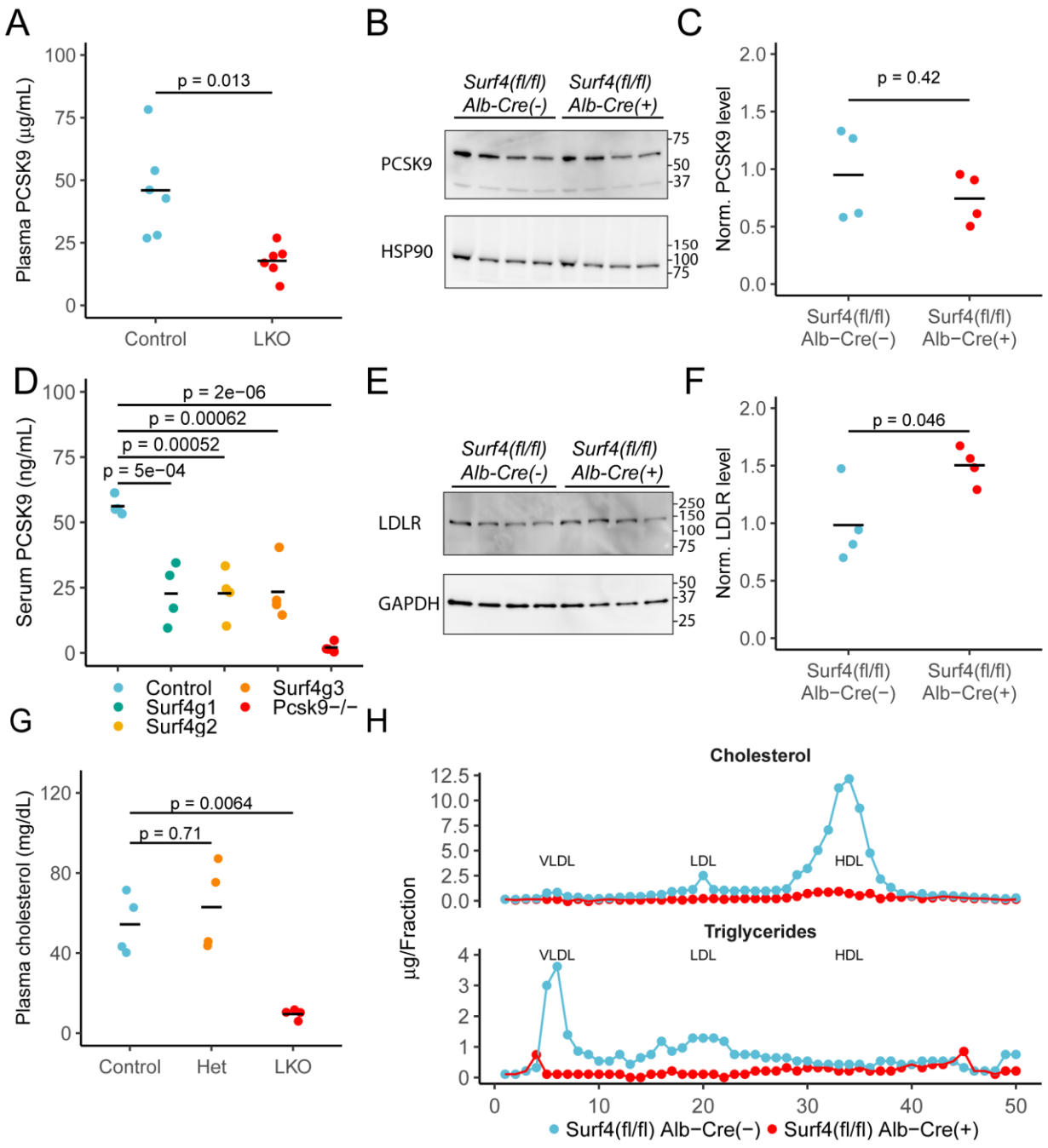

Figure 5-3: Deletion of hepatic *Surf4* results in decreased serum PCSK9 level and profound hypocholesterolemia in mice. **(A)** Serum PCSK9 levels measured by ELISA in *Surf4*$^{fl/fl}$ *Alb-Cre*$^{+}$ mice and *Surf4*$^{fl/fl}$ *Alb-Cre*$^{-}$ littermate controls. **(B)** Immunoblot for PCSK9 and HSP90 (loading control) in liver lysates collected from control and *Surf4*$^{fl/fl}$ *Alb-Cre*$^{+}$ mice. **(C)** Quantification of liver PCSK9 levels presented in **(B)**. **(D)** Serum PCSK9 levels in mice in which hepatic *Surf4* was acutely inactivated by CRISPR/Cas9. **(E)** Immunoblot of liver lysates collected from control and *Surf4*$^{fl/fl}$ *Alb-Cre*$^{+}$ mice for LDLR and GAPDH (loading control). **(F)** Quantification of liver LDLR levels presented in **(E)**. **(G)** Steady-state plasma cholesterol levels in 2 months old control (*Surf4*$^{fl/fl}$ *Alb-Cre*$^{-}$), heterozygous (*Surf4*$^{fl/+}$ *Alb-Cre*$^{+}$), and homozygous (*Surf4*$^{fl/fl}$ *Alb-Cre*$^{+}$) mice. **(H)** Fractionation of lipoproteins in mouse serum by fast protein liquid chromatography (FPLC). Cholesterol and triglyceride levels were measured in

each fraction. Each control and *Surf4*$^{fl/fl}$ *Alb-Cre*$^{+}$ sample was pooled from sera of 5 mice. Fractions corresponding to VLDL, LDL, and HDL are annotated. Crossbars represent the mean in all plots. For comparisons between control and *Surf4*$^{fl/fl}$ *Alb-Cre*$^{+}$, p-values were calculated by two-sided Student's t-test. For comparison between control, heterozygous, and *Surf4*$^{fl/fl}$ *Alb-Cre*$^{+}$, p-value were obtained by one-way ANOVA test followed by Tukey's post hoc test. Molecular weight markers notated are in kDa.

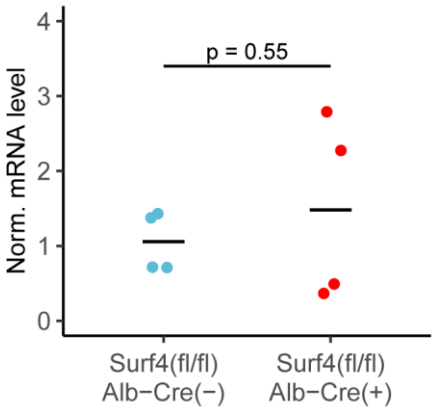

Figure 5-4: Normalized (Norm.) *Pcsk9* mRNA levels in *Surf4*$^{fl/fl}$ *Alb-Cre*$^{-}$.and *Surf4*$^{fl/fl}$ *Alb-Cre*$^{+}$ livers. Crossbars represent the mean, p-values were obtained from a two-sided Student's t-test.

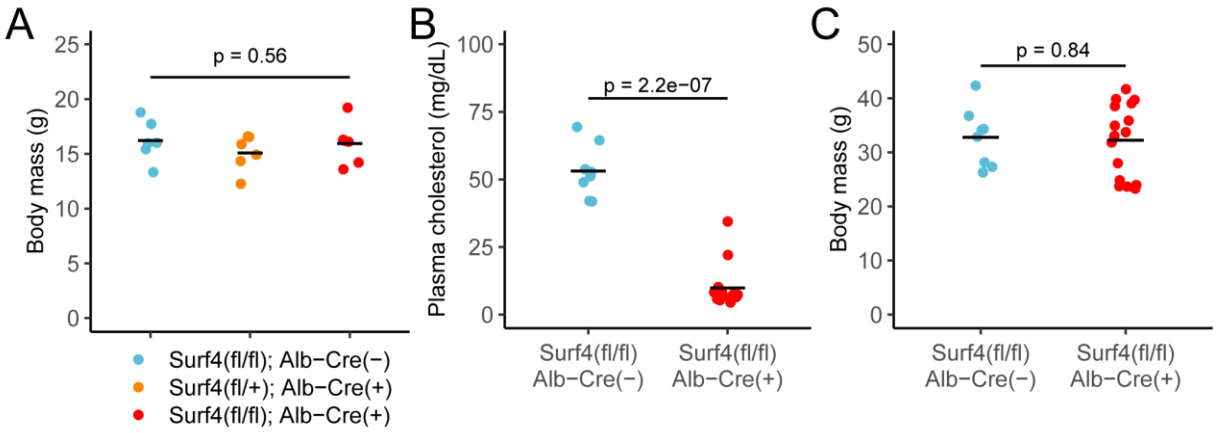

Figure 5-5: Hepatic *Surf4* inactivation does not affect body mass while hypocholesterolemia is sustained through at least 1 year of age. (A) Body mass of control (*Surf4*$^{fl/fl}$ *Alb-Cre*$^{-}$), heterozygous (Het, *Surf4*$^{fl/+}$ *Alb-Cre*$^{+}$), and homozygous (*Surf4*$^{fl/fl}$ *Alb-Cre*$^{+}$) mice at 2 months of age. P-value was obtained by a one-way ANOVA test. **(B-C)** Serum cholesterol and body mass of 1 year old control and *Surf4*$^{fl/fl}$ *Alb-Cre*$^{+}$ mice. P-values were obtained from a two-sided Student's t-test. For all panels: crossbars represent the mean.

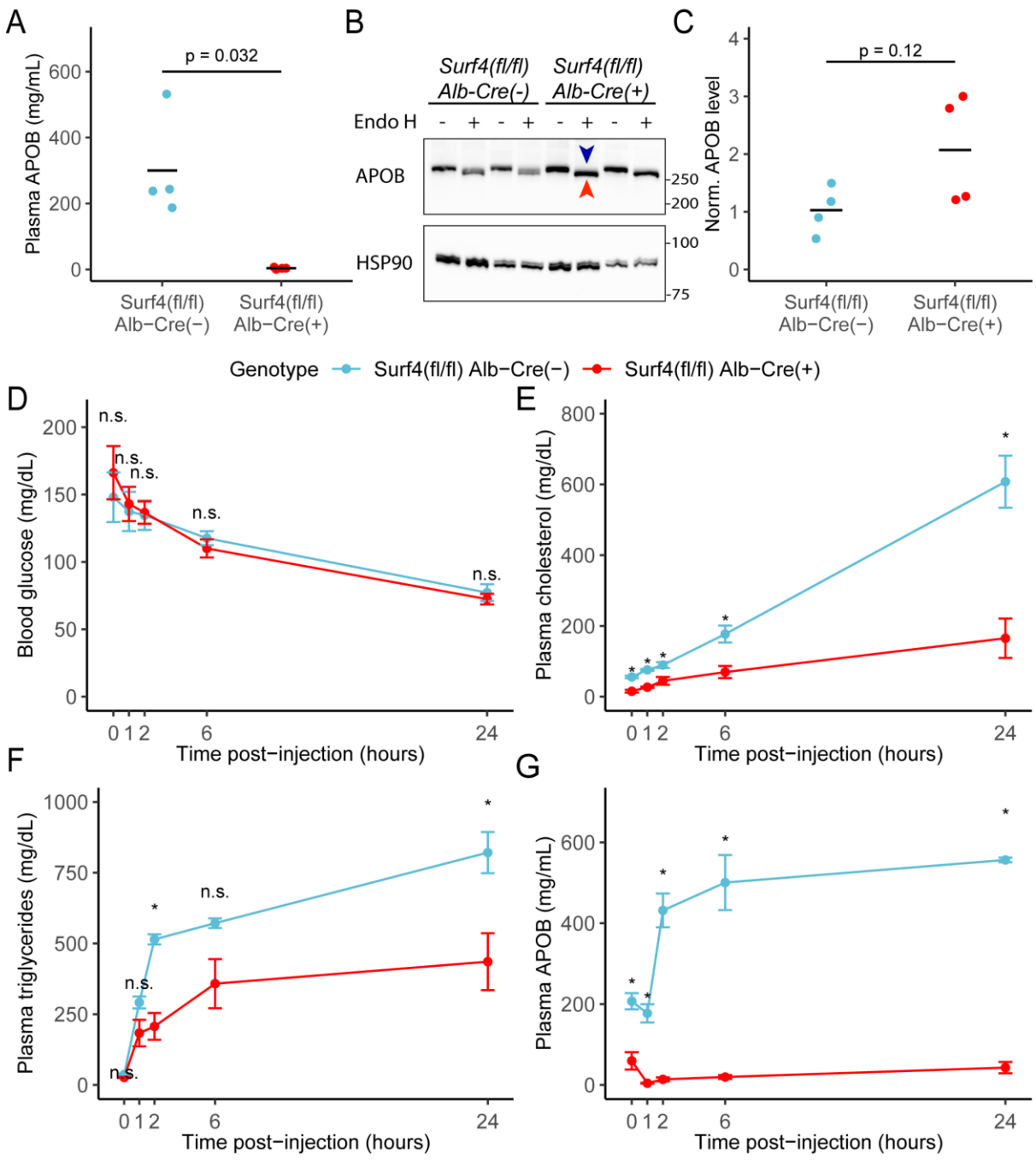

Figure 5-6: Hepatic lipoprotein and APOB secretion defect in *Surf4$^{fl/fl}$ Alb-Cre^{+}* mice. (A) Steady-state serum APOB levels in control and *Surf4$^{fl/fl}$ Alb-Cre^{+}* mice at 2 months of age. **(B)** Representative immunoblot for APOB and HSP90 in liver lysates with and without endoglycosidase H (endo H) treatment. Proteins in the pre-Golgi compartments are expected to be sensitive to endo H cleavage, resulting in an electrophoretic shift on an immunoblot. Blue arrowhead indicates the endo H resistant band whereas the red arrowhead indicates the endo H sensitive band. Molecular weight markers notated are in kDa. Accumulation of endo H sensitive APOB in the absence of SURF4 suggests accumulation in the ER. **(C)** Quantification of APOB abundance in control and *Surf4$^{fl/fl}$ Alb-Cre^{+}* liver lysates, without endo H treatment. For panel A and C, crossbars represent the mean, with statistical significance determined by two-sided Student's t-test. **(D-G)** *Surf4$^{fl/fl}$ Alb-Cre^{+}* and littermate control mice were

injected with a lipoprotein lipase inhibitor to block triglyceride hydrolysis. Blood was sampled prior to and following injection over 24 hours and assayed for **(D)** glucose, **(E)** cholesterol, **(F)** triglycerides, and **(G)** APOB levels. Data are presented as mean ± SEM for each time point (n=5 per group). Asterisks denotes $p < 0.05$ obtained from two-sided Student's t-test with Benjamini-Hochberg adjustment for multiple hypothesis testing, n.s., not significant.

5.4.4 Intestinal lipid absorption and tissue lipid uptake are unaffected in Surf4$^{fl/fl}$ Alb-Cre^{+} mice

To assess the potential role of hepatic *Surf4* gene expression on dietary lipid absorption, mice were fed a ^{3}H triolein-labelled lipid load following an overnight fast. No significant differences in blood glucose, serum triglycerides, non-esterified fatty acids, or total intestinal uptake of dietary lipids were observed between control and *Surf4$^{fl/fl}$ Alb-Cre^{+}* mice (**Figure 5-7A-D**). Similarly, no significant differences were observed in tissue lipid uptake or fatty acid oxidation between *Surf4$^{fl/fl}$ Alb-Cre^{+}* mice and littermate controls (**Figure 5-7E-F**).

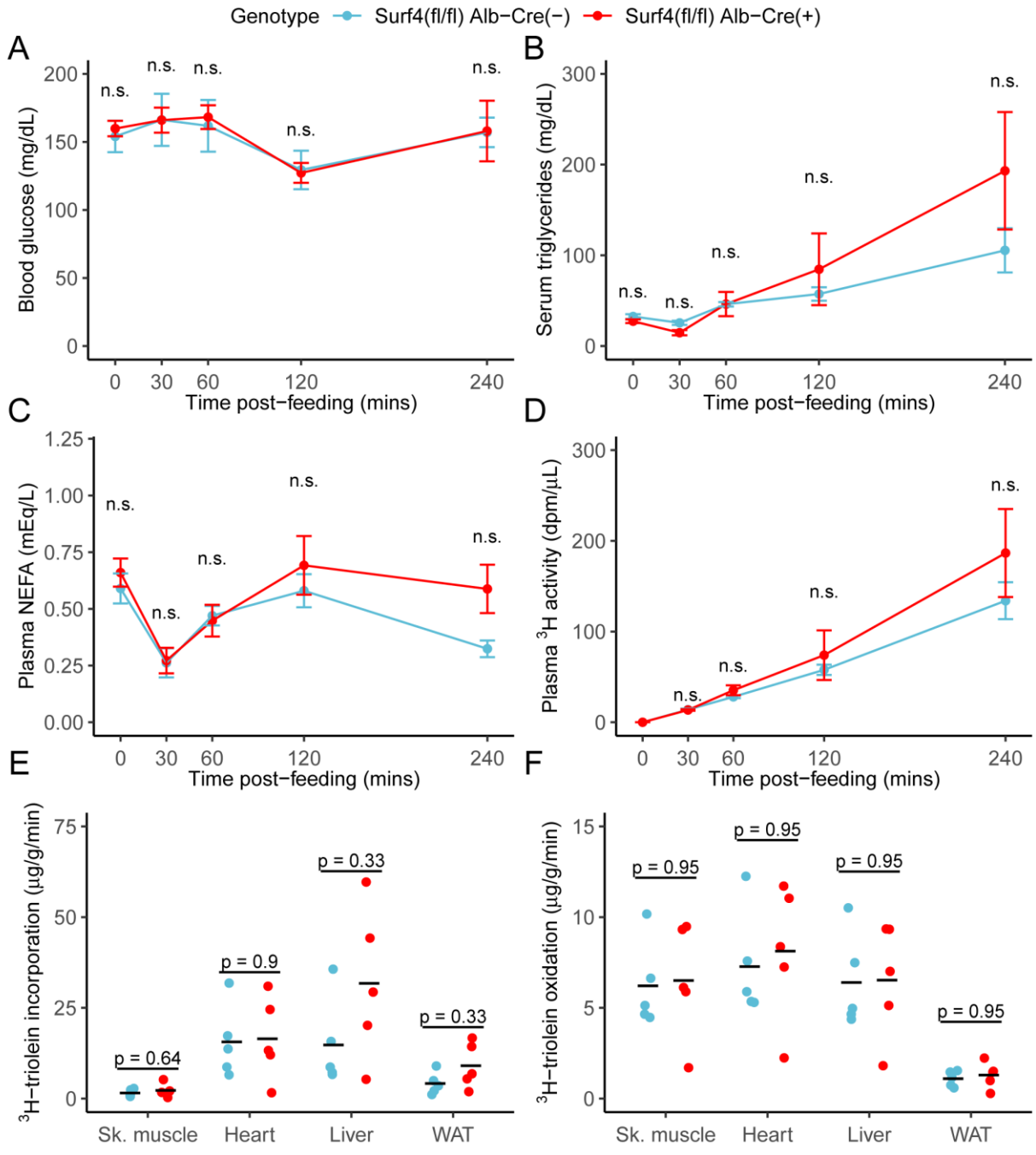

Figure 5-7: Inactivation of hepatic *Surf4* does not impact dietary lipid absorption, incorporation, and oxidation. Mice were administered ^{3}H-labelled triolein by oral gavage. Blood samples were collected over 4 hours and assayed for **(A)** glucose, **(B)** triglycerides, **(C)** non-esterified fatty acids (NEFA), and **(D)** ^{3}H radioactivity. Data are presented as mean ± SEM for each time point (n=5 per group), n.s., not significant. **(E-F)** Tissues were collected at the 4 hour time point and lipids were extracted by the Folch's method. ^{3}H radioactivity was measured in the hydrophobic phase, which represents incorporated triolein **(E)** and hydrophilic phase, which represents oxidized triolein **(F)**. All crossbars represent the mean. The denoted p-values were obtained by two-sided Student's t-test with Benjamini-Hochberg adjustment for multiple hypothesis testing.

5.4.5 Loss of liver Surf4 expression results in mild lipid accumulation but no steatohepatitis or fibrosis

Surf4$^{fl/fl}$ *Alb-Cre*$^{+}$ mice exhibited mildly enlarged livers (**Figure 5-8A**), with a small increase in hepatic fat content and a reduction in lean mass compared to littermate controls (**Figure 5-8B**). However, no differences were observed in fasting hepatic cholesterol, triglyceride, phospholipid, or nonesterified fatty acid content (**Figure 5-8C**). Hepatic lipid accumulation can lead to steatohepatitis and liver damage [258]. However, at 8-12 week of age, serum albumin, bilirubin, and liver function markers were indistinguishable between *Surf4*$^{fl/fl}$ *Alb-Cre*$^{+}$ and control mice (**Figure 5-9** and **Figure 5-8D**) and histologic analyses detected no evidence for steatohepatitis or fibrosis (**Figure 5-8E**). Finally, deep sequencing of liver mRNA identified only limited gene expression changes in response to *Surf4* deletion (**Figure 5-8F**). Significant downregulation was observed for several genes involved in fatty acid biosynthesis processes (**Figure 5-8H**). The most significantly upregulated gene in *Surf4*$^{fl/fl}$ *Alb-Cre*$^{+}$ liver was *Derl3*, a component of the ERAD pathway, which could be induced by protein accumulation in the ER of *Surf4*$^{fl/fl}$ *Alb-Cre*$^{+}$ mice. Genes involved in the unfolded protein response, such as *Ire1*, *Aft6*, and *Perk* [259], were not upgregulated in *Surf4*$^{fl/fl}$ *Alb-Cre*$^{+}$ mice (**Figure 5-8F**).

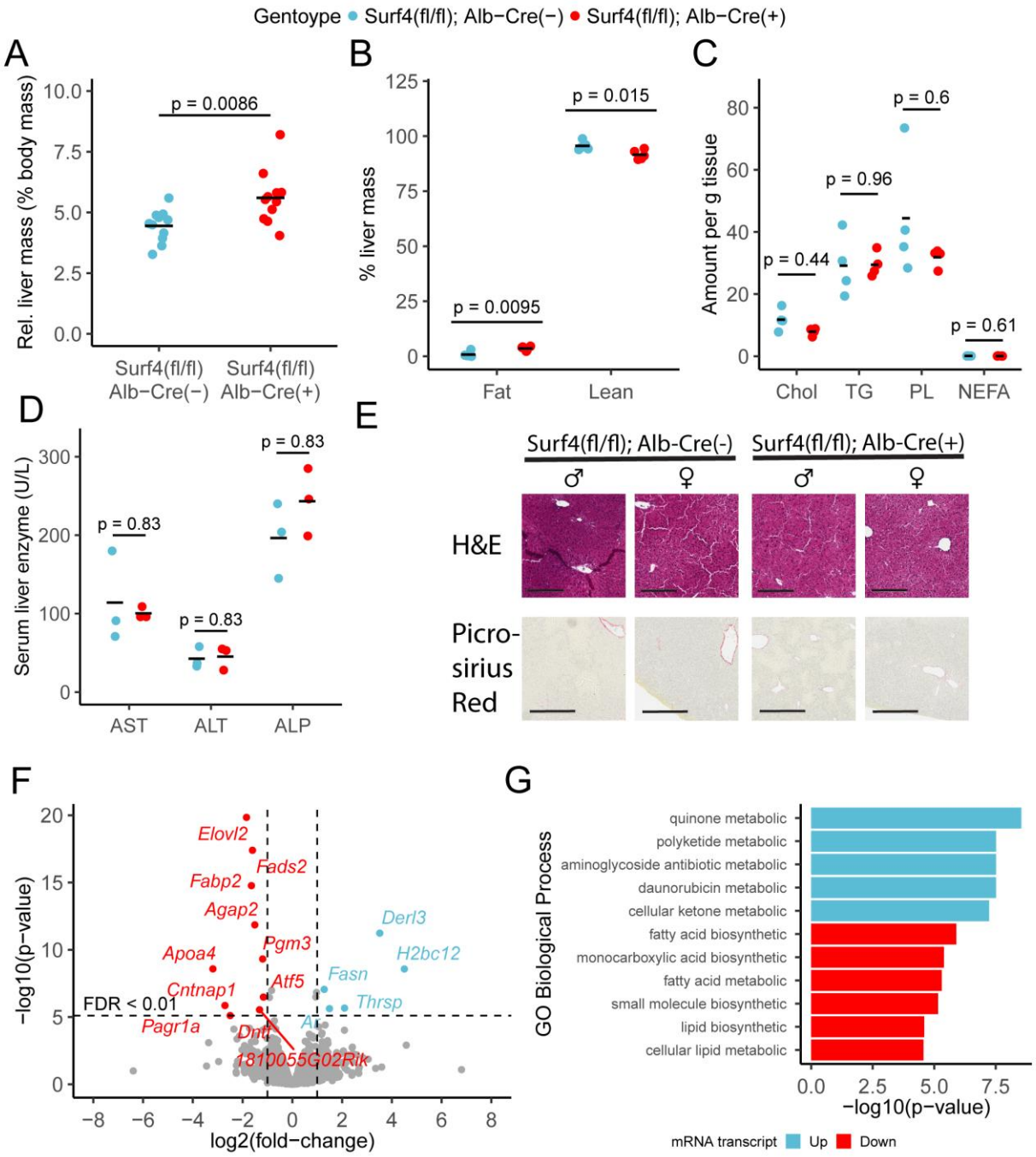

Figure 5-8: Hepatic *Surf4* deletion results in mild hepatomegaly and an increase in liver lipid content, without apparent liver dysfunction or steatohepatitis. (A) Relative liver mass in control and *Surf4*$^{fl/fl}$ *Alb-Cre*$^{+}$ mice presented as percentage of total body mass. **(B)** Relative fat and lean mass in the livers of control and *Surf4*$^{fl/fl}$ *Alb-Cre*$^{+}$ mice measured by EchoMRI™ and presented as percentage of liver mass. **(C)** Levels of cholesterol (Chol, mg/g tissue), triglycerides (TG, mg/g tissue), phospholipids (PL, mg/g tissue), and nonesterified fatty acid (NEFA, mEq/g tissue) in lipids extracted from the livers. **(D)** Serum levels of asparate aminotransferase (AST), alanine transaminase (ALT), and alkaline phosphatase (ALP). **(E)** H&E and picrosirius red stained liver sections from control and *Surf4*$^{fl/fl}$ *Alb-Cre*$^{+}$ mice. Scale bars represent 200 µm in H&E images and 300 µm in picrosirius red images. **(F)** Changes in mRNA transcript levels in *Surf4*$^{fl/fl}$ *Alb-Cre*$^{+}$ mice compared to littermate controls.

Horizontal line represents the p-value above which the false discovery rate (FDR) is less than 0.01. Significantly up (blue) or down (red) regulated transcripts are labelled. **(G)** Significantly overrepresented Gene Ontology (GO) terms for biological processes in up and down regulated gene lists. For panels A-D: All crossbars represent the mean. P-values were obtained by two-sided Student's t-test with Benjamini-Hochberg adjustment for multiple hypothesis testing where appropriate.

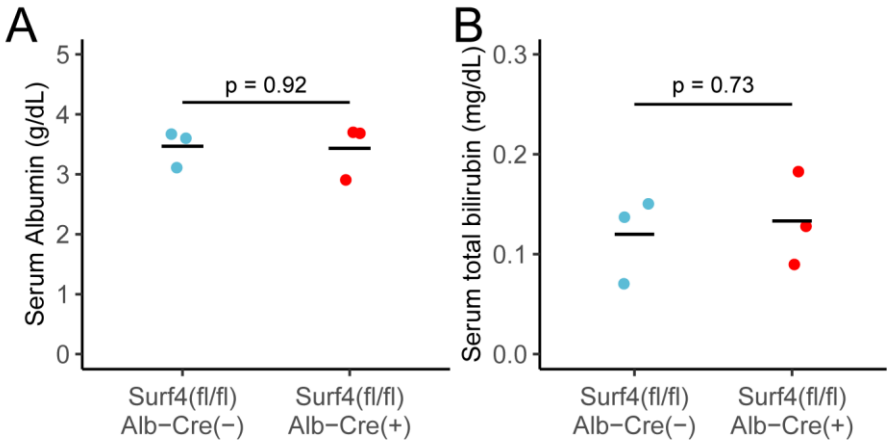

Figure 5-9: Hepatic SURF4 deficiency does not result in liver injuries. Serum **(A)** albumin and **(B)** bilirubin levels in *Surf4*$^{fl/fl}$ *Alb-Cre*$^{-}$ and *Surf4*$^{fl/fl}$ *Alb-Cre*$^{+}$ mice. Crossbars represent the mean. P-values were obtained from two-sided Student's t-test.

5.4.6 Dose-dependent reduction of plasma lipids in response to depletion of SURF4 by siRNA

To confirm the profound hypocholesterolemia with few adverse consequences observed in *Surf4*$^{fl/fl}$ *Alb-Cre*$^{+}$ mice, and to further explore SURF4 inhibition as a potential therapeutic approach, we next tested depletion of hepatic SURF4 using liver-targeted siRNA in adult mice. Mice were treated with control or *Surf4* targeting siRNA at multiple doses between 0.5-4 mg/kg. Mice treated with *Surf4* targeting siRNA demonstrated a dose-dependent reduction of liver *Surf4* mRNA and protein levels (**Figure 5-10A** and **Figure 5-11**). As expected, plasma PCSK9, cholesterol, triglycerides, APOB, and APOA1 levels were inversely correlated with siRNA dosage, with the highest siRNA dose (4 mg/kg) resulting in cholesterol levels similar to those observed in *Surf4*$^{fl/fl}$ *Alb-Cre*$^{+}$ mice (**Figure 5-10B-D** and **Figure 5-11**). Finally, no differences in plasma ALT and AST levels were observed between control and siRNA treated mice,

suggesting that siRNA treatment and *Surf4* depletion does not lead to liver injury, even at the highest siRNA dose (**Figure 5-10E-F**).

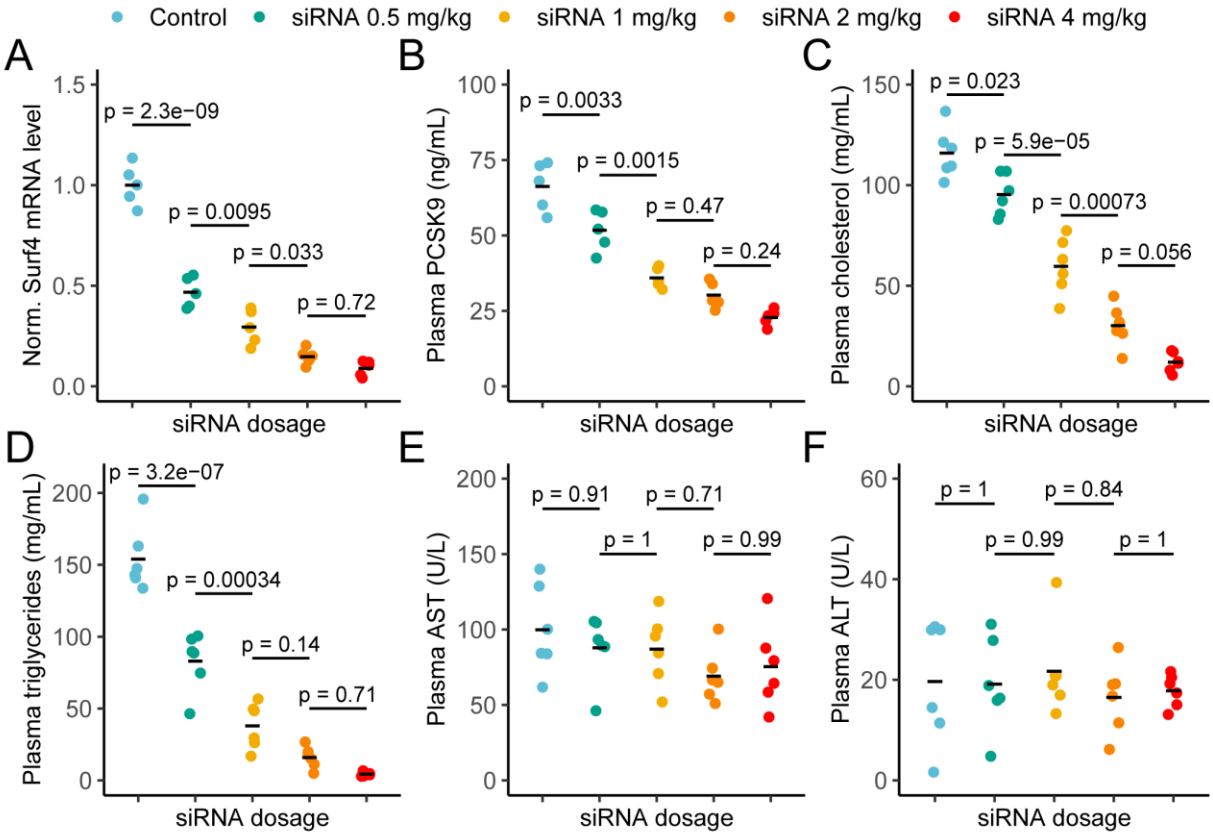

Figure 5-10: Depletion of hepatic *Surf4* by siRNA recapitulates the hypolipidemia seen in *Surf4*$^{fl/fl}$ *Alb-Cre*$^{+}$ mice. (A) Relative liver *Surf4* mRNA levels in mice treated with scrambled siRNA (control) or varying concentrations of *Surf4* targeting siRNA. **(B-D)** Plasma PCSK9, cholesterol, and triglyceride levels in control and siRNA treated mice. **(E-F)** Plasma levels of asparate aminotransferase (AST) and alanine transaminase (ALT) in mice treated with control or increasing doses of *Surf4* targeting siRNA. Statistical significance was computed by one-way ANOVA test followed by Tukey's post hoc test.

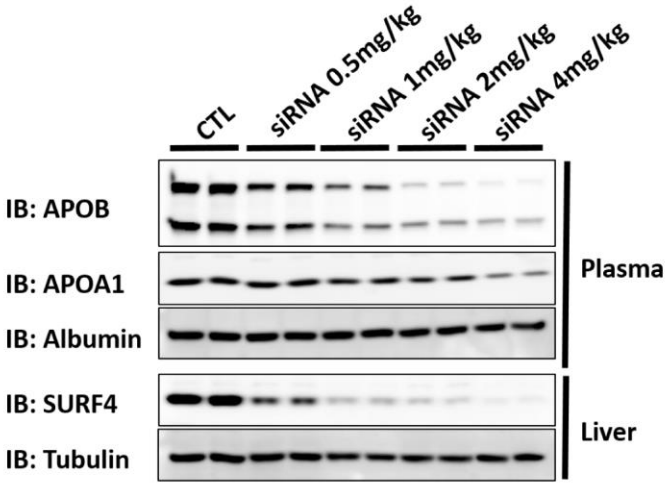

Figure 5-11: Acute inactivation of hepatic SURF4 results in decreased secretion of apolipoproteins. Immunoblot of plasma and liver lysates collected from mice treated with control or increasing doses of Surf4 targeting siRNA.

5.5 Discussion

We found that embryonic deletion of *Surf4* in hepatocytes results in profound hypocholesterolemia in mice associated with impaired hepatic lipoprotein secretion and normal dietary fat absorption. In addition, we also demonstrated that plasma PCSK9 levels are reduced in *Surf4*$^{fl/fl}$ *Alb-Cre*$^{+}$ mice. Hepatocyte specific *Surf4* deletion was well tolerated, with only modest increases in hepatic mass and lipid content, and no evidence of hepatic dysfunction or steatohepatitis. Finally, we confirm these findings by siRNA depletion of hepatic SURF4 in adult mice, which leads to reductions of plasma cholesterol, triglycerides, and PCSK9 in a dose dependent manner without apparent deleterious consequences in the liver.

We previously reported that PCSK9 is dependent on SURF4 for efficient secretion in cultured HEK293T cells [125]. In contrast, Shen et al reported that depletion of SURF4 by siRNA in cultured human hepatocytes leads to increased *Pcsk9* gene expression resulting in increased rather than decreased PCSK9 secretion [237]. The same group also recently reported analysis of *Surf4*$^{fl/fl}$ *Alb-Cre*$^{+}$ mice, observing no change in plasma PCSK9 levels, in contrast to

our findings in a similar genetic model. Our current findings, using three independent mouse models, are consistent with our previous *in vitro* data and support a physiologic role for SURF4 in facilitating the efficient transport of PCSK9 (as well as APOB) through the secretory pathway. The decrease in plasma PCSK9 in *Surf4*$^{fl/fl}$ *Alb-Cre*$^{+}$ mice is similar to that observed in SEC24A-deficient mice [60]. Additionally, we also detected a 1.5 fold increase in LDLR level in liver lysates collected from *Surf4*$^{fl/fl}$ *Alb-Cre*$^{+}$ mice, consistent with the reduction in circulating PCSK9. The increased hepatocyte LDLR levels in *Surf4*$^{fl/fl}$ *Alb-Cre*$^{+}$ mice are not accompanied by upregulation of *Ldlr* mRNA as measured by RNA-seq analysis, consistent with the increase in LDLR abundance being mediated by PCSK9 activity rather than an increase in gene expression. Furthermore, a recent report by Gomez-Navarro et al independently demonstrated that PCSK9 relies on both SEC24A and SURF4 for secretion and that chemical disruption of SEC24A-SURF4 interaction is sufficient to reduce PCSK9 secretion [161]. Taken together, these data are consistent with the proposed function of SURF4 as a cargo receptor linking PCSK9 in the ER lumen to the SEC24A component of the COPII coat on the cytoplasmic face of the ER [125]. The basis for the discrepancy between our findings and those of B. Wang and colleagues [230] is unclear but may be related to differences in mouse genetic or husbandry [260], or to technical differences in PCSK9 quantification.

The profound hypocholesterolemia we observed in *Surf4*$^{fl/fl}$ *Alb-Cre*$^{+}$ mice is in agreement with two other studies in which hepatic *Surf4* was acutely inactivated using an AAV/Cas9 mouse system [160] or a similar *Surf4*$^{fl/fl}$ and *Alb-Cre* model [230]. We also confirmed this observation in a third model using siRNA mediated knockdown of *Surf4* transcripts in mouse livers. The decrease in circulating cholesterol is likely due to impaired secretion of APOB containing lipoprotein particles from the liver, and is consistent with multiple

previous reports suggesting that APOB is a cargo for SURF4 [159, 160, 162, 230]. Despite remarkably low plasma cholesterol levels, *Surf4*$^{fl/fl}$ *Alb-Cre*$^{+}$ mice exhibit normal growth and fertility compared to littermate controls. Plasma cholesterol is an important precursor for steroid hormone synthesis. However, given that both male and female *Surf4*$^{fl/fl}$ *Alb-Cre*$^{+}$ mice are fertile, it is unlikely that sex hormone synthesis is significantly perturbed in these mice. Consistent with this conclusion, Chang et al recently reported that even though lipid droplets and cholesterol are depleted in the adrenal glands of *Surf4*$^{fl/fl}$ *Alb-Cre*$^{+}$ mice, circulating adrenal steroid hormone levels are unchanged under resting and stressed conditions [261].

Impaired protein secretion could lead to accumulation of proteins in the ER lumen, potentially triggering activation of UPR pathways and induction of ERAD. Indeed, livers from *Surf4*$^{fl/fl}$ *Alb-Cre*$^{+}$ mice exhibited upregulation of mRNA for *Derl3*, an ER transmembrane protein that is a functional component of the ER-associated degradation (ERAD) complex [262]. While *Derl3* is thought to be a target of the IRE1-XBP1 pathway, we did not detect upregulation of *Ire1* or *Xbp1*, or other ERAD components at the mRNA levels in *Surf4*$^{fl/fl}$ *Alb-Cre*$^{+}$ mice. Instead upregulation of *Derl3* could be an adaptive response to the protein accumulation in the ER leading to the rapid degradation of these proteins. This can also explain the lack of liver PCSK9 accumulation and a mild increase in liver APOB levels (relative to a significant reduction of plasma levels) in *Surf4*$^{fl/fl}$ *Alb-Cre*$^{+}$ mice.

Recently, Musunuru et al reported that *in vivo* CRISPR-mediated base editing of hepatic *PCSK9* leads to an ~60% reduction in plasma cholesterol in cynomolgus monkeys without overt hepatotoxicity [263]. Our data suggest that hepatic *Surf4* could be similarly targeted, potentially achieving an even more profound reduction in plasma cholesterol without deleterious consequences. Indeed, it has been shown that inactivation of hepatic *Surf4* is protective against

diet-induced atherosclerosis in mice with PCSK9 overexpression [160], LDLR deficiency [230], and APOE deficiency [264]. Furthermore, polymorphism and mild reduction of *SURF4* expression strongly associate with lower plasma lipid levels and reduced risks of cardiovascular disease in human populations [160]. Here, we further demonstrate that even more modest reductions in SURF4 induced by siRNA targeting are likely to confer significant lipid-lowering, though such benefits must be weighed against potential toxicity from disrupting the secretion of other SURF4-dependent cargoes.

Table 5-3: Primers used in this chapter

Primer name	Target	Sequence (5' → 3')
Surf4-fl-F1 (P1)	*Surf4*	TGAGAAAAGTATGGAGTCCACCTGC
Surf4-fl-R1 (P2)	*Surf4*	AACAGCTGTGCTATGGGCTGGAAAG
Surf4-fl-R2 (P3)	*Surf4*	CCTGCCTCTAAATCCCAAATGCTGTCG
Cre-genotyping-F1	*Cre transgene*	CCATCTGCCACCAGCCAG
Cre-genotyping-R1	*Cre transgene*	TCGCCATCTTCCAGCAGG
Cre-genotyping-F2	*Cpxm1* (Control)	ACTGGGATCTTCGAACTCTTTGGAC
Cre-genotyping-R2	*Cpxm1* (Control)	GATGTTGGGGCACTGCTCATTCACC
Alb-cre-F1	*Alb-Cre*	TGCAAACATCACATGCACAC
Alb-cre-R1	*Alb-Cre*	TTGGCCCCTTACCATAACTG
Alb-cre-R2	*Alb-Cre*	GAAGCAGAAGCTTAGGAAGATGG
Surf4-qPCR-F1	*Surf4* (exon 2)	CTGTTGGCCTCATCCTTCGT
Surf4-qPCR-R1	*Surf4* (exon 3)	GGCAATTGTCTGCAGTGCG
Surf4-qPCR-F2	*Surf4* (exon 5)	TTTGCTGGTGTCCCAACCAT
Surf4-qPCR-R2	*Surf4* (exon 6)	AGCTGTGCCCACAATGTTCT
Pcsk9-qPCR-F1	*Pcsk9*	TATAGCCGCATCCTCAACGC
Pcsk9-qPCR-R1	*Pcsk9*	CCCGACTGTGATGACCTCTG
Rpl37-F	*Rpl37*	CGGGACTGGTCGGATGAG
Rpl37-R	*Rpl37*	TCACGGAATCCATGTCTGAATC
Gapdh-F	*Gapdh*	ACCCAGAAGACTGTGGATGG
Gapdh-R	*Gapdh*	ACACATTGGGGGTAGGAACA

Table 5-4: Guide RNA sequences for CRISPR-mediated *in vivo* hepatic *Surf4* inactivation

	Guide RNA sequence (5'-3')
Control	TGCGAATACGCCCACGCGAT
Surf4g1	ACAGAACGACCTGATGGGCA
Surf4g2	GCATCCGCATGTGGTTCCAG
Surf4g3	CAGCAGGTTGAGGAACACGA

Table 5-5: Sequences and modifications of oligonucleotides for siRNA in mice

	Sense (5'-3') 3' GalNAc conjugated	Antiense (5'-3') 3' GalNAc conjugated
si*Surf4*-1	CmsCmsAmUmCfAmAfCfGfUmGmU mAmUmUmUmCmAmAm	UmsUfsGmAmAmAfUmAfCfAmCmG mUmUfGmAfUmGmGmsCmsAm
si*Surf4*-2	GmsGmsAmCmAfAmUfCfCfCmGmG mUmAmUmAmUmAmAm	UmsUfsAmUmAmUfAmCfCfGmGmG mAmUfUmGfUmCmCmsAmsGm
siCTL	mU*mU*mCmUfCmCfGfAfAmC mGmUmGmUmCmAmCmGmU	mA*fC*mGmUmGfAmCfAfCmGmU mUmCfGmGfAmGmAmA*mU*mU

Chapter 6
Conclusions and Future perspectives

The work presented in this dissertation explores the specificity of COPII mediated protein transport in the eukaryotic intracellular secretory pathway. Chapter 1 provides an overview on ER-to-Golgi transport and COPII components, current evidence for unique and overlapping functions between COPII paralogs, as well as background on the two prototypical cargo receptors in mammals, LMAN1 and SURF4. In Chapter 2, we investigated the functional conservation between two highly conserved SAR1 paralogs in mammals, SAR1A and SAR1B. Chapters 3-5 focus on cargo receptors and their roles in conferring specificity in cargo recruitment. We developed a novel analysis approach to enhance the sensitivity of detecting secreted proteins in cell culture systems (Chapter 3) and applied this technique to identify novel cargoes for LMAN1 and SURF4 – two prototypical cargo receptors in mammals (Chapter 4). In Chapter 5, we further explored the physiologic functions of SURF4 using mice with hepatic SURF4 deficiency. Together, we found a high degree of functional overlap between SAR1A/B in mice, contrasting cargo repertoire size between LMAN1 and SURF4, and profound hypocholesterolemia in mice that lack hepatic *Surf4*.

6.1 Functional conservation of COPII proteins

Four of the five COPII proteins have at least two paralogs in mammals, two each for SAR1, SEC23, and SEC31, and four for SEC24. It has been previously demonstrated that the two SEC23 paralogs, SEC23A and SEC23B, are largely functionally equivalent, with SEC23A

compensating for SEC23B deficiency in mice [74]. Loss of SEC24C results in lethality at E7.5 and loss of SEC24D leads to death before the 8-cell stage in mice. Our lab recently demonstrated that replacement of the C-terminal 90% of SEC24C with SEC24D protein sequence could partially rescue the early embryonic lethality exhibited by *Sec24c* null mice, allowing these embryos to survive to term though they still died shortly after birth. This suggests some degrees of functional overlap between SEC24C and SEC24D in early development stages. Thus, gene expression differences, rather than cargo specificity, are responsible for the early embryonic lethality of *Sec24c* null mice [63]. Continuing this line of investigation, we now demonstrate in Chapter 2 that the two SAR1 paralogs also exhibit significant functional overlaps in mice. Replacement of the *Sar1b* coding sequence with that of *Sar1a* can completely rescue SAR1B deficiency, again suggesting that differential gene expression accounts for the phenotypic discrepancy between *Sar1a*- and *Sar1b*-null mice. The remaining paralogous pairs to be explored are SEC24A/B and SEC31A/B. While there are no *in vivo* data to assess functional overlap between these paralogs in mammals, *in vitro* experiments with a human cell line and studies in plants provide evidences for partial functional overlap between SEC24A/B and SEC31A/B [60, 117].

Gene duplications are frequent evolutionary events. Typically, one of the duplicated copy accumulates deleterious mutations over time, becomes a pseudogene, and disappears from the genome. However, in some cases, both copies can be conserved due to neofunctionalization (acquisition of new functions different from those of the ancestral gene) or subfunctionalization (division of the ancestral gene's functions). The latter case can occurs at the protein level or the transcription level [49]. Together, the current data suggest that the primary evolutionary force driving fixation of the COPII paralog duplications in the genome is the subfunctionalization of

these genes at the transcription level. Differential gene expression patterns for each individual paralog results in the requirement for all paralogs to be present in order to recapitulate the expression pattern of the ancestral gene and thereby maintain normal organismal functions.

Mutations in COPII genes lead to several disorders in human, including chylomicron retention disease (CMRD) due to mutations in *SAR1B*, cranio-lenticulo-sutural dysplasia (CLSD) due to mutations in *SEC23A*, congenital dyserythropoietic anemia type II (CDAII) due to mutations in *SEC23B*, and a rare syndromic form of osteogenesis imperfeca (OI) due to mutations in *SEC24D*. Our findings that COPII proteins are largely functionally equivalent suggest that therapeutic upregulation of one paralog could be leveraged to compensate for deficiency of the other. This approach may be of limited value for CLSD and OI as they are developmental disorders and treatment after embryonic development is less likely to be beneficial. However, CMRD and CDAII could potentially be treated by such therapies. Indeed, a recent report demonstrated that CRISPR-mediated activation of *SEC24A* is sufficient to rescue the erythroid differentiation defects exhibited by SEC23B-deficiency in an erythroid cell line (HUDEP-2) [79]. Future studies focusing on upregulating *SAR1A* expression using either small molecules or activating-CRISPR could lead to development of novel treatments for patients with CMRD. More broadly, similar studies to investigate the functional conservation between paralogous proteins could provide valuable insights to develop novel therapies to treat disorders caused by deficiency of one paralog. For example, knowledge on the functional equivalency between β- and γ-hemoglobin has led to successful development of therapies that induce γ-hemoglobin expression to treat β-hemoglobinopathies [265-267].

6.2 Cargo-cargo receptor relationship

Cargo receptors are ER transmembrane proteins that bridge interactions between COPII proteins in the cytoplasm with soluble cargoes in the ER lumen. Even though cargo receptor mediated secretion has been demonstrated to be an important mechanism for protein transport from the ER to the Golgi, few cargo receptors have been identified. This suggests a role for passive bulk flow as the main mechanism to transport proteins from the ER with a small subset of secreted proteins requiring interaction with a cargo receptor to enhance secretion [29].

LMAN1 and SURF4 are prototypical cargo receptors in mammals. It is known that the cargo-cargo receptor is not a one-to-one relationship as each cargo receptor can traffic multiple proteins. However, the extent of the cargo repertoire for each receptor remains unclear. Both LMAN1 and SURF4 were identified as cargo receptors based on genetics studies [125, 142]. Given inconsistent data on the binding/recognition motifs for these two prototypical cargo receptors, it is challenging to extrapolate their features to search for other cargo receptors or to predict their clients. Recent advancement in whole genome CRISPR screens allow a relatively straightforward experiment to search for cargo receptors for a specific protein of interest. SURF4 was identified as a cargo receptor that is responsible for the secretion of PCSK9 and EPO using cell lines that overexpressed the cargo of interest fused to a GFP reporter [125, 164]. However, one caveat of this approach is the use of heterologously overexpressed reporter proteins, which might not reflect their physiologic abundance.

As for identifying the set of proteins that depend on a known cargo receptor for secretion, mass spectrometry analyses of conditioned media from cultured cells allows unbiased protein identification, though distinguishing truly secreted proteins from intracellular proteins released into the media as a result of cellular injury during the experiment process remains technically

challenging. In Chapter 3, we developed a novel analysis approach to identify bona fide secreted proteins in cell culture systems. By comparing a given protein's abundance in the conditioned media to that in the cell lysates, we were able to distinguish secreted proteins from intracellular protein contaminants. In Chapter 4, we applied this technique to identify cargoes that depend on LMAN1 or SURF4 for efficient secretion from the ER. We found that in hepatocytes, LMAN1 traffics a limited number of cargoes whereas SURF4 has a broad cargo range. It is unclear whether other cell types also exhibit a similar pattern. It has been previously reported that LMAN1 and SURF4 act in concert to maintain ERGIC and Golgi architecture [152]. It would be interesting to delete both LMAN1 and SURF4 and analyze the secretome of these cells in future studies.

6.3 SURF4 as a cargo receptor in mammals

Despite having long been suspected to function as a cargo receptor in mammals due to its high degree of protein sequence homology with Erv29p [152, 228], SURF4's clients have only been identified in the past few years. Chapter 4 revealed that SURF4 has an extensive cargo repertoire with over 60 putative cargoes identified in hepatocytes alone. Analysis of these cargo candidates failed to validate the previously proposed ER-ESCAPE motif for SURF4 recognition [163]. This suggests that SURF4-cargo interaction is governed by mechanisms that are more complex that a linear protein sequence. These might include representation of a SURF4 recognition motif only in the tertiary protein structure, presence of calcium in the environment [161], and post-translational modifications such as S-nitrosylation [268].

We had previously reported that germline deletion of *Surf4* results in early embryonic lethality in mice [159]. In Chapter 5, we extended our study to a mouse model with hepatic SURF4 deficiency. We found that apart from significant hypocholesterolemia due to a defect in

hepatic lipoprotein secretion, *Surf4*$^{fl/fl}$ *Alb-Cre*$^{+}$ mice exhibit normal growth and fertility and are grossly indistinguishable from control mice. In future studies, analysis of tissues and plasma from these mice could be used to validate *in vivo* the putative SURF4 cargoes identified in Chapter 4.

PCSK9 is a soluble plasma protein that is secreted by the liver. PCSK9 negatively regulates LDL receptor (LDLR) abundance by inducing its degradation. Therefore, high plasma PCSK9 level leads to reduced LDLR abundance, resulting in high circulating cholesterol level and vice versa. Genetic *PCSK9* deficiency results in hypocholesterolemia in human [241, 269]. Two monoclonal antibodies against PCSK9 (alirocumab and evolocumab) have been approved to treat familial hypercholesterolemia [270, 271]. Gene therapy targeting hepatic *PCSK9* is currently under development [272]. The key role of SURF4 in facilitating the secretion of PCSK9 coupled with the hypocholesterolemia observed in *Surf4*$^{fl/fl}$ *Alb-Cre*$^{+}$ mice suggests SURF4 as a potential therapeutic target for the treatment of hypercholesterolemia [125, 180]. Indeed, Gomez-Navarro et al recently demonstrated that disrupting SURF4 interaction with SEC24 – a COPII component that mediates cargo selection, using a small molecule is sufficient to inhibit the secretion of PCSK9 and other SURF4 clients [161]. However, the lethality observed in *Surf4* null embryos in addition to our finding of a wide range of SURF4 cargoes in HuH7 cells raises the concern for potential off target effects on other SURF4 cargoes aside from PCSK9, which could limit this approach. Nevertheless, the hypocholesterolemia observed in otherwise healthy *Surf4*$^{fl/fl}$ *Alb-Cre*$^{+}$ mice suggests that approaches specifically targeting SURF4 in hepatocytes could be viable. In addition, optimizing interaction between SURF4 and its cargoes such as insulin [167] and EPO [164] could potentially be leveraged to enhance protein secretion and improve the efficiency of recombinant protein production.

Bibliography

1. Jamieson, J.D. and G.E. Palade, *Synthesis, Intracellular Transport, and Discharge of Secretory Proteins in Stimulated Pancreatic Exocrine Cells.* Journal of Cell Biology, 1971. **50**(1): p. 135-158.
2. Dallner, G., P. Siekevitz, and G.E. Palade, *Biogenesis of Endoplasmic Reticulum Membranes.* Journal of Cell Biology, 1966. **30**(1): p. 97-117.
3. Palade, G., *Intracellular aspects of the process of protein synthesis.* Science, 1975. **189**(4200): p. 347-58.
4. Redman, C.M., P. Siekevitz, and G.E. Palade, *Synthesis and transfer of amylase in pigeon pancreatic micromosomes.* J Biol Chem, 1966. **241**(5): p. 1150-8.
5. Blobel, G. and D.D. Sabatini, *Controlled proteolysis of nascent polypeptides in rat liver cell fractions. I. Location of the polypeptides within ribosomes.* J Cell Biol, 1970. **45**(1): p. 130-45.
6. Jamieson, J.D. and G.E. Palade, *Intracellular transport of secretory proteins in the pancreatic exocrine cell. I. Role of the peripheral elements of the Golgi complex.* J Cell Biol, 1967. **34**(2): p. 577-96.
7. Jamieson, J.D. and G.E. Palade, *Intracellular transport of secretory proteins in the pancreatic exocrine cell. II. Transport to condensing vacuoles and zymogen granules.* J Cell Biol, 1967. **34**(2): p. 597-615.
8. Dimou, E. and W. Nickel, *Unconventional mechanisms of eukaryotic protein secretion.* Curr Biol, 2018. **28**(8): p. R406-R410.
9. Kim, J., H.Y. Gee, and M.G. Lee, *Unconventional protein secretion - new insights into the pathogenesis and therapeutic targets of human diseases.* J Cell Sci, 2018. **131**(12).
10. Rabouille, C., *Pathways of Unconventional Protein Secretion.* Trends Cell Biol, 2017. **27**(3): p. 230-240.
11. Filaquier, A., et al., *Roads and hubs of unconventional protein secretion.* Curr Opin Cell Biol, 2022. **75**: p. 102072.
12. Novick, P., C. Field, and R. Schekman, *Identification of 23 complementation groups required for post-translational events in the yeast secretory pathway.* Cell, 1980. **21**(1): p. 205-15.
13. Barlowe, C., et al., *COPII: a membrane coat formed by Sec proteins that drive vesicle budding from the endoplasmic reticulum.* Cell, 1994. **77**(6): p. 895-907.
14. Barlowe, C., C. d'Enfert, and R. Schekman, *Purification and characterization of SAR1p, a small GTP-binding protein required for transport vesicle formation from the endoplasmic reticulum.* J Biol Chem, 1993. **268**(2): p. 873-9.
15. Barlowe, C. and R. Schekman, *SEC12 encodes a guanine-nucleotide-exchange factor essential for transport vesicle budding from the ER.* Nature, 1993. **365**(6444): p. 347-9.
16. Huang, M., et al., *Crystal structure of Sar1-GDP at 1.7 A resolution and the role of the NH2 terminus in ER export.* J Cell Biol, 2001. **155**(6): p. 937-48.

17. Bi, X., R.A. Corpina, and J. Goldberg, *Structure of the Sec23/24-Sar1 pre-budding complex of the COPII vesicle coat.* Nature, 2002. **419**(6904): p. 271-7.
18. Yoshihisa, T., C. Barlowe, and R. Schekman, *Requirement for a GTPase-activating protein in vesicle budding from the endoplasmic reticulum.* Science, 1993. **259**(5100): p. 1466-8.
19. Lederkremer, G.Z., et al., *Structure of the Sec23p/24p and Sec13p/31p complexes of COPII.* Proc Natl Acad Sci U S A, 2001. **98**(19): p. 10704-9.
20. Wieland, F.T., et al., *The rate of bulk flow from the endoplasmic reticulum to the cell surface.* Cell, 1987. **50**(2): p. 289-300.
21. Thor, F., et al., *Bulk flow revisited: transport of a soluble protein in the secretory pathway.* Traffic, 2009. **10**(12): p. 1819-30.
22. Bendayan, M., et al., *Quantitative immunocytochemical localization of pancreatic secretory proteins in subcellular compartments of the rat acinar cell.* J Histochem Cytochem, 1980. **28**(2): p. 149-60.
23. Kuehn, M.J., J.M. Herrmann, and R. Schekman, *COPII-cargo interactions direct protein sorting into ER-derived transport vesicles.* Nature, 1998. **391**(6663): p. 187--190.
24. Malkus, P., F. Jiang, and R. Schekman, *Concentrative sorting of secretory cargo proteins into COPII-coated vesicles.* J Cell Biol, 2002. **159**(6): p. 915-21.
25. Balch, W.E., et al., *Vesicular stomatitis virus glycoprotein is sorted and concentrated during export from the endoplasmic reticulum.* Cell, 1994. **76**(5): p. 841-52.
26. Miller, E., et al., *Cargo selection into COPII vesicles is driven by the Sec24p subunit.* EMBO Journal, 2002. **21**(22): p. 6105--6113.
27. Miller, E.A., et al., *Multiple cargo binding sites on the COPII subunit Sec24p ensure capture of diverse membrane proteins into transport vesicles.* Cell, 2003. **114**(4): p. 497--509.
28. Mossessova, E., L.C. Bickford, and J. Goldberg, *SNARE selectivity of the COPII coat.* Cell, 2003. **114**(4): p. 483-495.
29. Barlowe, C. and A. Helenius, *Cargo Capture and Bulk Flow in the Early Secretory Pathway.* Annual Review of Cell and Developmental Biology, 2016. **32**(1): p. 197--222.
30. Mironov, A.A. and G.V. Beznoussenko, *Models of Intracellular Transport: Pros and Cons.* Front Cell Dev Biol, 2019. **7**: p. 146.
31. Caro, L.G. and G.E. Palade, *Protein Synthesis, Storage, and Discharge in the Pancreatic Exocrine Cell. An Autoradiographic Study.* J Cell Biol, 1964. **20**: p. 473-95.
32. Claude, A., *Growth and differentiation of cytoplasmic membranes in the course of lipoprotein granule synthesis in the hepatic cell. I. Elaboration of elements of the Golgi complex.* J Cell Biol, 1970. **47**(3): p. 745-66.
33. Morre, D.J., T.W. Keenan, and H.H. Mollenhauer, *Golgi apparatus function in membrane transformations and product compartmentalization: studies with cell fractions isolated from rat liver.* Adv Cytopharmacol, 1971. **1**: p. 159-82.
34. Kaiser, C.A. and R. Schekman, *Distinct sets of SEC genes govern transport vesicle formation and fusion early in the secretory pathway.* Cell, 1990. **61**(4): p. 723-33.
35. Barlowe, C., et al., *COPII: a membrane coat formed by Sec proteins that drive vesicle budding from the endoplasmic reticulum.* Cell, 1994. **77**(6): p. 895--907.
36. Bednarek, S.Y., et al., *COPI- and COPII-coated vesicles bud directly from the endoplasmic reticulum in yeast.* Cell, 1995. **83**(7): p. 1183-96.

37. Zeuschner, D., et al., *Immuno-electron tomography of ER exit sites reveals the existence of free COPII-coated transport carriers.* Nat Cell Biol, 2006. **8**(4): p. 377-83.
38. Appenzeller-Herzog, C. and H.P. Hauri, *The ER-Golgi intermediate compartment (ERGIC): in search of its identity and function.* J Cell Sci, 2006. **119**(Pt 11): p. 2173-83.
39. Mironov, A.A., et al., *ER-to-Golgi carriers arise through direct en bloc protrusion and multistage maturation of specialized ER exit domains.* Dev Cell, 2003. **5**(4): p. 583-94.
40. Bachinger, H.P., L.I. Fessler, and J.H. Fessler, *Mouse procollagen IV. Characterization and supramolecular association.* J Biol Chem, 1982. **257**(16): p. 9796-803.
41. Zilversmit, D.B., *Formation and transport of chylomicrons.* Fed Proc, 1967. **26**(6): p. 1599-605.
42. Weigel, A.V., et al., *ER-to-Golgi protein delivery through an interwoven, tubular network extending from ER.* Cell, 2021. **184**(9): p. 2412-2429 e16.
43. Shomron, O., et al., *COPII collar defines the boundary between ER and ER exit site and does not coat cargo containers.* J Cell Biol, 2021. **220**(6).
44. Westrate, L.M., et al., *Vesicular and uncoated Rab1-dependent cargo carriers facilitate ER to Golgi transport.* J Cell Sci, 2020. **133**(14).
45. Zaremba-Niedzwiedzka, K., et al., *Asgard archaea illuminate the origin of eukaryotic cellular complexity.* Nature, 2017. **541**(7637): p. 353-358.
46. Eme, L., et al., *Archaea and the origin of eukaryotes.* Nat Rev Microbiol, 2018. **16**(2): p. 120.
47. Neumann, N., D. Lundin, and A.M. Poole, *Comparative genomic evidence for a complete nuclear pore complex in the last eukaryotic common ancestor.* PLoS One, 2010. **5**(10): p. e13241.
48. Schlacht, A. and J.B. Dacks, *Unexpected ancient paralogs and an evolutionary model for the COPII coat complex.* Genome Biol Evol, 2015. **7**(4): p. 1098-109.
49. Innan, H. and F. Kondrashov, *The evolution of gene duplications: classifying and distinguishing between models.* Nat Rev Genet, 2010. **11**(2): p. 97-108.
50. Charcosset, M., et al., *Anderson or chylomicron retention disease: molecular impact of five mutations in the SAR1B gene on the structure and the functionality of Sar1b protein.* Mol Genet Metab, 2008. **93**(1): p. 74-84.
51. Georges, A., et al., *Molecular analysis and intestinal expression of SAR1 genes and proteins in Anderson's disease (Chylomicron retention disease).* Orphanet J Rare Dis, 2011. **6**: p. 1.
52. Jones, B., et al., *Mutations in a Sar1 GTPase of COPII vesicles are associated with lipid absorption disorders.* Nat Genet, 2003. **34**(1): p. 29-31.
53. Auclair, N., et al., *Sar1b mutant mice recapitulate gastrointestinal abnormalities associated with chylomicron retention disease.* J Lipid Res, 2021. **62**: p. 100085.
54. Boyadjiev, S.A., et al., *Cranio-lenticulo-sutural dysplasia is caused by a SEC23A mutation leading to abnormal endoplasmic-reticulum-to-Golgi trafficking.* Nat Genet, 2006. **38**(10): p. 1192-7.
55. Boyadjiev, S.A., et al., *Cranio-lenticulo-sutural dysplasia associated with defects in collagen secretion.* Clin Genet, 2011. **80**(2): p. 169-76.
56. Schwarz, K., et al., *Mutations affecting the secretory COPII coat component SEC23B cause congenital dyserythropoietic anemia type II.* Nat Genet, 2009. **41**(8): p. 936-40.
57. Zhu, M., et al., *Neural tube opening and abnormal extraembryonic membrane development in SEC23A deficient mice.* Sci Rep, 2015. **5**: p. 15471.

58. Tao, J., et al., *SEC23B is required for the maintenance of murine professional secretory tissues.* Proceedings of the National Academy of Sciences, 2012. **109**(29): p. E2001--9.
59. Garbes, L., et al., *Mutations in SEC24D, encoding a component of the COPII machinery, cause a syndromic form of osteogenesis imperfecta.* Am J Hum Genet, 2015. **96**(3): p. 432-9.
60. Chen, X.W., et al., *SEC24A deficiency lowers plasma cholesterol through reduced PCSK9 secretion.* eLife, 2013. **2013**(2): p. 1--23.
61. Merte, J., et al., *Sec24b selectively sorts Vangl2 to regulate planar cell polarity during neural tube closure.* Nature Cell Biology, 2010. **12**(1): p. 41--46.
62. Adams, E.J., et al., *Mammalian COPII coat component SEC24C is required for embryonic development in mice.* Journal of Biological Chemistry, 2014. **289**(30): p. 20858--20870.
63. Adams, E.J., et al., *Murine SEC24D can substitute functionally for SEC24C during embryonic development.* Sci Rep, 2021. **11**(1): p. 21100.
64. Moreira, T.G., et al., *Sec13 Regulates Expression of Specific Immune Factors Involved in Inflammation In Vivo.* Sci Rep, 2015. **5**: p. 17655.
65. Saito, K., M. Maeda, and T. Katada, *Regulation of the Sar1 GTPase Cycle Is Necessary for Large Cargo Secretion from the Endoplasmic Reticulum.* Front Cell Dev Biol, 2017. **5**: p. 75.
66. Anderson, C.M., et al., *Unusual causes of steatorrhoea in infancy and childhood.* Med J Aust, 1961. **48(2)**: p. 617-22.
67. Peretti, N., et al., *Chylomicron retention disease: a long term study of two cohorts.* Mol Genet Metab, 2009. **97**(2): p. 136-42.
68. Siddiqi, S.A., et al., *COPII proteins are required for Golgi fusion but not for endoplasmic reticulum budding of the pre-chylomicron transport vesicle.* J Cell Sci, 2003. **116**(Pt 2): p. 415-27.
69. Melville, D.B., S. Studer, and R. Schekman, *Small sequence variations between two mammalian paralogs of the small GTPase SAR1 underlie functional differences in coat protein complex II assembly.* Journal of Biological Chemistry, 2020: p. jbc.RA120.012964.
70. Sane, A.T., et al., *Understanding Chylomicron Retention Disease Through Sar1b GTPase Gene Disruption: Insight From Cell Culture.* Arterioscler Thromb Vasc Biol, 2017. **37**(12): p. 2243-2251.
71. Sane, A., et al., *SAR1B GTPase is necessary to protect intestinal cells from disorders of lipid homeostasis, oxidative stress, and inflammation.* J Lipid Res, 2019. **60**(10): p. 1755-1764.
72. Matsuoka, K., et al., *COPII-coated vesicle formation reconstituted with purified coat proteins and chemically defined liposomes.* Cell, 1998. **93**(2): p. 263-75.
73. Lee, M.C., et al., *Sar1p N-terminal helix initiates membrane curvature and completes the fission of a COPII vesicle.* Cell, 2005. **122**(4): p. 605-17.
74. Khoriaty, R., et al., *Functions of the COPII gene paralogs SEC23A and SEC23B are interchangeable in vivo.* Proc Natl Acad Sci U S A, 2018. **115**(33): p. E7748-E7757.
75. Khoriaty, R., M.P. Vasievich, and D. Ginsburg, *The COPII pathway and hematologic disease.* Blood, 2012. **120**(1): p. 31--38.
76. Paccaud, J.P., et al., *Cloning and functional characterization of mammalian homologues of the COPII component Sec23.* Mol Biol Cell, 1996. **7**(10): p. 1535-46.

77. Khoriaty, R., et al., *Absence of a red blood cell phenotype in mice with hematopoietic deficiency of SEC23B.* Mol Cell Biol, 2014. **34**(19): p. 3721-34.
78. Khoriaty, R., et al., *SEC23B is required for pancreatic acinar cell function in adult mice.* Mol Biol Cell, 2017. **28**(15): p. 2146-2154.
79. King, R., et al., *SEC23A rescues SEC23B-deficient congenital dyserythropoietic anemia type II.* Sci Adv, 2021. **7**(48): p. eabj5293.
80. Satchwell, T.J., et al., *Characteristic phenotypes associated with congenital dyserythropoietic anemia (type II) manifest at different stages of erythropoiesis.* Haematologica, 2013. **98**(11): p. 1788-96.
81. Jeong, Y.T., et al., *The ULK1-FBXW5-SEC23B nexus controls autophagy.* Elife, 2018. **7**.
82. Cui, Y., et al., *A COPII subunit acts with an autophagy receptor to target endoplasmic reticulum for degradation.* Science, 2019. **365**(6448): p. 53-60.
83. Anding, A.L. and E.H. Baehrecke, *Cleaning House: Selective Autophagy of Organelles.* Dev Cell, 2017. **41**(1): p. 10-22.
84. Davis, S., J. Wang, and S. Ferro-Novick, *Crosstalk between the Secretory and Autophagy Pathways Regulates Autophagosome Formation.* Dev Cell, 2017. **41**(1): p. 23-32.
85. Kurihara, T., et al., *Sec24p and Iss1p function interchangeably in transport vesicle formation from the endoplasmic reticulum in Saccharomyces cerevisiae.* Mol Biol Cell, 2000. **11**(3): p. 983-98.
86. Roberg, K.J., et al., *LST1 is a SEC24 homologue used for selective export of the plasma membrane ATPase from the endoplasmic reticulum.* J Cell Biol, 1999. **145**(4): p. 659-72.
87. Shimoni, Y., et al., *Lst1p and Sec24p cooperate in sorting of the plasma membrane ATPase into COPII vesicles in Saccharomyces cerevisiae.* J Cell Biol, 2000. **151**(5): p. 973-84.
88. Pagano, A., et al., *Sec24 proteins and sorting at the endoplasmic reticulum.* Journal of Biological Chemistry, 1999. **274**(12): p. 7833--7840.
89. Tang, B.L., et al., *A family of mammalian proteins homologous to yeast Sec24p.* Biochemical and biophysical research communications, 1999. **258**(3): p. 679--84.
90. Mancias, J.D., et al., *Structural basis of cargo membrane protein discrimination by the human COPII coat machinery.* The EMBO Journal, 2008. **27**(21): p. 2918--2928.
91. Peng, R., A. De Antoni, and D. Gallwitz, *Evidence for overlapping and distinct functions in protein transport of coat protein Sec24p family members.* J Biol Chem, 2000. **275**(15): p. 11521-8.
92. Wendeler, M.W., et al., *Role of Sec24 isoforms in selective export of membrane proteins from the endoplasmic reticulum.* EMBO reports, 2007. **8**(3): p. 258--64.
93. Chatterjee, S., A.J. Choi, and G. Frankel, *A systematic review of Sec24 cargo interactome.* Traffic, 2021. **22**(12): p. 412-424.
94. Yang, X.Y., et al., *Mutations in the COPII vesicle component gene SEC24B are associated with human neural tube defects.* Hum Mutat, 2013. **34**(8): p. 1094-101.
95. Karczewski, K.J., et al., *The mutational constraint spectrum quantified from variation in 141,456 humans.* Nature, 2020. **581**(7809): p. 434-443.
96. Wang, B., et al., *The COPII cargo adapter SEC24C is essential for neuronal homeostasis.* J Clin Invest, 2018. **128**(8): p. 3319-3332.
97. Baines, A.C., et al., *Disruption of the Sec24d Gene Results in Early Embryonic Lethality in the Mouse.* PLoS ONE, 2013. **8**(4): p. 1--10.

98. Adolf, F., et al., *Proteomic Profiling of Mammalian COPII and COPI Vesicles.* Cell Reports, 2019. **26**(1): p. 250--265.e5.
99. Sauvageau, E., et al., *CNIH4 interacts with newly synthesized GPCR and controls their export from the endoplasmic reticulum.* Traffic, 2014. **15**(4): p. 383-400.
100. Sucic, S., et al., *The serotonin transporter is an exclusive client of the coat protein complex II (COPII) component SEC24C.* J Biol Chem, 2011. **286**(18): p. 16482-90.
101. Kovalchuk, V., et al., *Trafficking of the amino acid transporter B(0,+) (SLC6A14) to the plasma membrane involves an exclusive interaction with SEC24C for its exit from the endoplasmic reticulum.* Biochim Biophys Acta Mol Cell Res, 2019. **1866**(2): p. 252-263.
102. Lyu, L., et al., *Selective export of autotaxin from the endoplasmic reticulum.* J Biol Chem, 2017. **292**(17): p. 7011-7022.
103. Farhan, H., et al., *Concentrative export from the endoplasmic reticulum of the gamma-aminobutyric acid transporter 1 requires binding to SEC24D.* J Biol Chem, 2007. **282**(10): p. 7679-89.
104. Fernandez-Sanchez, E., et al., *Mechanisms of endoplasmic-reticulum export of glycine transporter-1 (GLYT1).* Biochem J, 2008. **409**(3): p. 669-81.
105. Bonnon, C., et al., *Selective export of human GPI-anchored proteins from the endoplasmic reticulum.* J Cell Sci, 2010. **123**(Pt 10): p. 1705-15.
106. Salama, N.R., J.S. Chuang, and R.W. Schekman, *Sec31 encodes an essential component of the COPII coat required for transport vesicle budding from the endoplasmic reticulum.* Mol Biol Cell, 1997. **8**(2): p. 205-17.
107. Gurkan, C. and W.E. Balch, *Recombinant production in baculovirus-infected insect cells and purification of the mammalian Sec13/Sec31 complex.* Methods Enzymol, 2005. **404**: p. 58-66.
108. Stagg, S.M., et al., *Structure of the Sec13/31 COPII coat cage.* Nature, 2006. **439**(7073): p. 234-8.
109. Bi, X., J.D. Mancias, and J. Goldberg, *Insights into COPII coat nucleation from the structure of Sec23.Sar1 complexed with the active fragment of Sec31.* Dev Cell, 2007. **13**(5): p. 635-645.
110. Fromme, J.C. and R. Schekman, *COPII-coated vesicles: flexible enough for large cargo?* Curr Opin Cell Biol, 2005. **17**(4): p. 345-52.
111. Townley, A.K. and D.J. Stephens, *Vesicle coating and uncoating: controlling the formation of large COPII-coated carriers.* F1000 Biol Rep, 2009. **1**: p. 65.
112. Townley, A.K., et al., *Efficient coupling of Sec23-Sec24 to Sec13-Sec31 drives COPII-dependent collagen secretion and is essential for normal craniofacial development.* J Cell Sci, 2008. **121**(Pt 18): p. 3025-34.
113. Shugrue, C.A., et al., *Identification of the putative mammalian orthologue of Sec31P, a component of the COPII coat.* J Cell Sci, 1999. **112 (Pt 24)**: p. 4547-56.
114. Tang, B.L., et al., *Mammalian homologues of yeast sec31p. An ubiquitously expressed form is localized to endoplasmic reticulum (ER) exit sites and is essential for ER-Golgi transport.* J Biol Chem, 2000. **275**(18): p. 13597-604.
115. Stankewich, M.C., P.R. Stabach, and J.S. Morrow, *Human Sec31B: a family of new mammalian orthologues of yeast Sec31p that associate with the COPII coat.* J Cell Sci, 2006. **119**(Pt 5): p. 958-69.
116. Zhao, B., et al., *Secretory COPII Protein SEC31B Is Required for Pollen Wall Development.* Plant Physiol, 2016. **172**(3): p. 1625-1642.

117. Liu, X., et al., *COPII genes SEC31A/B are essential for gametogenesis and interchangeable in pollen development in Arabidopsis.* Plant J, 2021. **105**(6): p. 1600-1614.
118. Capelson, M., et al., *Chromatin-bound nuclear pore components regulate gene expression in higher eukaryotes.* Cell, 2010. **140**(3): p. 372-83.
119. Siniossoglou, S., et al., *A novel complex of nucleoporins, which includes Sec13p and a Sec13p homolog, is essential for normal nuclear pores.* Cell, 1996. **84**(2): p. 265-75.
120. Enninga, J., A. Levay, and B.M. Fontoura, *Sec13 shuttles between the nucleus and the cytoplasm and stably interacts with Nup96 at the nuclear pore complex.* Mol Cell Biol, 2003. **23**(20): p. 7271-84.
121. Niu, X., et al., *Sec13 safeguards the integrity of the endoplasmic reticulum and organogenesis of the digestive system in zebrafish.* Dev Biol, 2012. **367**(2): p. 197-207.
122. Schmidt, K., et al., *Early stages of retinal development depend on Sec13 function.* Biol Open, 2013. **2**(3): p. 256-66.
123. Townley, A.K., et al., *Epithelial organization and cyst lumen expansion require efficient Sec13-Sec31-driven secretion.* J Cell Sci, 2012. **125**(Pt 3): p. 673-84.
124. Appenzeller, C., et al., *The lectin ERGIC-53 is a cargo transport receptor for glycoproteins.* Nature cell biology, 1999. **1**(6): p. 330--4.
125. Emmer, B.T., et al., *The cargo receptor SURF4 promotes the efficient cellular secretion of PCSK9.* eLife, 2018. **7**: p. 1--18.
126. Schindler, R., et al., *ERGIC-53, a membrane protein of the ER-Golgi intermediate compartment, carries an ER retention motif.* Eur J Cell Biol, 1993. **61**(1): p. 1-9.
127. Itin, C., et al., *Recycling of the endoplasmic reticulum/Golgi intermediate compartment protein ERGIC-53 in the secretory pathway.* Biochem Soc Trans, 1995. **23**(3): p. 541-4.
128. Kappeler, F., et al., *The recycling of ERGIC-53 in the early secretory pathway. ERGIC-53 carries a cytosolic endoplasmic reticulum-exit determinant interacting with COPII.* J Biol Chem, 1997. **272**(50): p. 31801-8.
129. Velloso, L.M., et al., *Crystal structure of the carbohydrate recognition domain of p58/ERGIC-53, a protein involved in glycoprotein export from the endoplasmic reticulum.* J Biol Chem, 2002. **277**(18): p. 15979-84.
130. Velloso, L.M., et al., *The crystal structure of the carbohydrate-recognition domain of the glycoprotein sorting receptor p58/ERGIC-53 reveals an unpredicted metal-binding site and conformational changes associated with calcium ion binding.* J Mol Biol, 2003. **334**(5): p. 845-51.
131. Sato, K. and A. Nakano, *Oligomerization of a cargo receptor directs protein sorting into COPII-coated transport vesicles.* Mol Biol Cell, 2003. **14**(7): p. 3055-63.
132. Nichols, W.C., et al., *Linkage of combined factors V and VIII deficiency to chromosome 18q by homozygosity mapping.* J Clin Invest, 1997. **99**(4): p. 596-601.
133. Nichols, W.C., et al., *Mutations in the ER-Golgi intermediate compartment protein ERGIC-53 cause combined deficiency of coagulation factors V and VIII.* Cell, 1998. **93**(1): p. 61--70.
134. Giddings, J.C., U. Seligsohn, and A.L. Bloom, *Immunological studies in combined factor V and factor VIII deficiency.* Br J Haematol, 1977. **37**(2): p. 257-64.
135. Vollenweider, F., et al., *Mistargeting of the lectin ERGIC-53 to the endoplasmic reticulum of HeLa cells impairs the secretion of a lysosomal enzyme.* J Cell Biol, 1998. **142**(2): p. 377-89.

136. Nyfeler, B., S.W. Michnick, and H.P. Hauri, *Capturing protein interactions in the secretory pathway of living cells.* Proc Natl Acad Sci U S A, 2005. **102**(18): p. 6350-5.
137. Fu, Y.L., B. Zhang, and T.W. Mu, *LMAN1 (ERGIC-53) promotes trafficking of neuroreceptors.* Biochem Biophys Res Commun, 2019. **511**(2): p. 356-362.
138. Nyfeler, B., et al., *Identification of ERGIC-53 as an intracellular transport receptor of alpha1-antitrypsin.* Journal of Cell Biology, 2008. **180**(4): p. 705--712.
139. Chen, Y., et al., *Regulation of Mac-2BP secretion is mediated by its N-glycan binding to ERGIC-53.* Glycobiology, 2013. **23**(7): p. 904-16.
140. Duellman, T., et al., *LMAN1 (ERGIC-53) is a potential carrier protein for matrix metalloproteinase-9 glycoprotein secretion.* Biochem Biophys Res Commun, 2015. **464**(3): p. 685-91.
141. Nichols, W.C., et al., *ERGIC-53 gene structure and mutation analysis in 19 combined factors V and VIII deficiency families.* Blood, 1999. **93**(7): p. 2261-2266.
142. Zhang, B., et al., *Bleeding due to disruption of a cargo-specific ER-to-Golgi transport complex.* Nat Genet, 2003. **34**(2): p. 220-5.
143. Zhang, B., et al., *Combined deficiency of factor V and factor VIII is due to mutations in either LMAN1 or MCFD2.* Blood, 2006. **107**(5): p. 1903-7.
144. Zhang, B., R.J. Kaufman, and D. Ginsburg, *LMAN1 and MCFD2 form a cargo receptor complex and interact with coagulation factor VIII in the early secretory pathway.* J Biol Chem, 2005. **280**(27): p. 25881-6.
145. Nyfeler, B., et al., *Cargo selectivity of the ERGIC-53/MCFD2 transport receptor complex.* Traffic, 2006. **7**(11): p. 1473-81.
146. Nie, C., et al., *Dimeric sorting code for concentrative cargo selection by the COPII coat Cell Biology.* PNAS, 2018.
147. Nufer, O., et al., *ER export of ERGIC-53 is controlled by cooperation of targeting determinants in all three of its domains.* J Cell Sci, 2003. **116**(Pt 21): p. 4429-40.
148. Neve, E.P., U. Lahtinen, and R.F. Pettersson, *Oligomerization and interacellular localization of the glycoprotein receptor ERGIC-53 is independent of disulfide bonds.* J Mol Biol, 2005. **354**(3): p. 556-68.
149. Huang, Y., et al., *An in vitro vesicle formation assay reveals cargo clients and factors that mediate vesicular trafficking.* Proceedings of the National Academy of Sciences, 2021. **118**(35): p. e2101287118.
150. Zhang, Y., et al., *LMAN1-MCFD2 complex is a cargo receptor for the ER-Golgi transport of alpha1-antitrypsin.* Biochem J, 2022. **479**(7): p. 839-855.
151. Zhang, B., et al., *Mice deficient in LMAN1 exhibit FV and FVIII deficiencies and liver accumulation of alpha1-antitrypsin.* Blood, 2011. **118**(12): p. 3384--3391.
152. Mitrovic, S., et al., *The Cargo Receptors Surf4, Endoplasmic Reticulum-Golgi Intermediate Compartment (ERGIC)-53, and p25 Are Required to Maintain the Architecture of ERGIC and Golgi.* Molecular Biology of the Cell, 2008. **19**(5): p. 1976--1990.
153. Reeves, J.E. and M. Fried, *The surf-4 gene encodes a novel 30 kDa integral membrane protein.* Molecular membrane biology, 1995. **12**(2): p. 201--8.
154. Belden, W.J. and C. Barlowe, *Role of Erv29p in Collecting Soluble Secretory Proteins into ER-Derived Transport Vesicles.* Science, 2001. **294**(5546): p. 1528-1531.

155. Caldwell, S.R., K.J. Hill, and A.A. Cooper, *Degradation of endoplasmic reticulum (ER) quality control substrates requires transport between the ER and Golgi.* J Biol Chem, 2001. **276**(26): p. 23296-303.
156. Otte, S. and C. Barlowe, *Sorting signals can direct receptor-mediated export of soluble proteins into COPII vesicles.* Nature Cell Biology, 2004. **6**(12): p. 1189--1194.
157. Armes, N. and M. Fried, *Surfeit locus gene homologs are widely distributed in invertebrate genomes.* Mol Cell Biol, 1996. **16**(10): p. 5591-6.
158. Henricson, A., et al., *Functional characterization in Caenorhabditis elegans of transmembrane worm-human orthologs.* BMC Genomics, 2004. **5**: p. 85.
159. Emmer, B.T., et al., *Murine Surf4 is essential for early embryonic development.* PLoS ONE, 2020. **15**(1).
160. Wang, X., et al., *Receptor-Mediated ER Export of Lipoproteins Controls Lipid Homeostasis in Mice and Humans.* Cell Metabolism, 2021. **33**(2): p. 350--366.e7.
161. Gomez-Navarro, N., et al., *Selective inhibition of protein secretion by abrogating receptor-coat interactions during ER export.* Proc Natl Acad Sci U S A, 2022. **119**(31): p. e2202080119.
162. Saegusa, K., et al., *SFT-4 Surf4 control ER export of soluble cargo proteins and participate in ER exit site organization.* Journal of Cell Biology, 2018: p. 1--13.
163. Yin, Y., et al., *Surf4 (Erv29p) binds amino-terminal tripeptide motifs of soluble cargo proteins with different affinities, enabling prioritization of their exit from the endoplasmic reticulum.* PLoS biology, 2018. **16**(8): p. e2005140.
164. Lin, Z., et al., *The Endoplasmic Reticulum Cargo Receptor SURF4 Facilitates Efficient Erythropoietin Secretion.* Molecular and Cellular Biology, 2020. **40**(23).
165. Ordonez, A., et al., *Cargo receptor-assisted endoplasmic reticulum export of pathogenic alpha1-antitrypsin polymers.* Cell Rep, 2021. **35**(7): p. 109144.
166. Tang, X., et al., *A SURF4-to-proteoglycan relay mechanism that mediates the sorting and secretion of a tagged variant of sonic hedgehog.* Proc Natl Acad Sci U S A, 2022. **119**(11): p. e2113991119.
167. Saegusa, K., et al., *Cargo receptor Surf4 regulates endoplasmic reticulum export of proinsulin in pancreatic beta-cells.* Commun Biol, 2022. **5**(1): p. 458.
168. Devireddy, S. and S.M. Ferguson, *Efficient progranulin exit from the ER requires its interaction with prosaposin, a Surf4 cargo.* J Cell Biol, 2022. **221**(2).
169. Yan, R., et al., *SURF4-induced tubular ERGIC selectively expedites ER-to-Golgi transport.* Dev Cell, 2022. **57**(4): p. 512-525 e8.
170. Zanetti, G., et al., *COPII and the regulation of protein sorting in mammals.* Nature Cell Biology, 2012. **14**(1): p. 20--28.
171. Bi, X., R.A. Corpina, and J. Goldberg, *Structure of the Sec23/24-Sar1 pre-budding complex of the COPII vesicle coat.* Nature, 2002. **419**(6904): p. 271--7.
172. Tang, V.T. and D. Ginsburg, *Cargo selection in endoplasmic reticulum-to-Golgi transport and relevant diseases.* J Clin Invest, 2023. **133**(1).
173. Consortium, G.T., *The GTEx Consortium atlas of genetic regulatory effects across human tissues.* Science, 2020. **369**(6509): p. 1318-1330.
174. Li, B., et al., *A Comprehensive Mouse Transcriptomic BodyMap across 17 Tissues by RNA-seq.* Sci Rep, 2017. **7**(1): p. 4200.
175. Jones, B., et al., *Mutations in a Sar1 GTPase of COPII vesicles are associated with lipid absorption disorders.* Nature Genetics, 2003. **34**(1): p. 29--31.

176. San, *SAR1B GTPase is necessary to protect intestinal cells from disorders of lipid homeostasis, oxidative stress, and inflammation.* Journal of Lipid Research, 2019. **60**(10): p. 1755--1764.
177. Auclair, N., et al., *Sar1b mutant mice recapitulate gastrointestinal abnormalities associated with chylomicron retention disease.* Journal of Lipid Research, 2021. **62**(12).
178. Stryke, D., et al., *BayGenomics: a resource of insertional mutations in mouse embryonic stem cells.* Nucleic Acids Res, 2003. **31**(1): p. 278-81.
179. Skarnes, W.C., et al., *A conditional knockout resource for the genome-wide study of mouse gene function.* Nature, 2011. **474**(7351): p. 337-42.
180. Tang, V.T., et al., *Hepatic inactivation of murine Surf4 results in marked reduction in plasma cholesterol.* Elife, 2022. **11**.
181. Schwarz, K., et al., *Mutations affecting the secretory COPII coat component SEC23B cause congenital dyserythropoietic anemia type II.* Nature genetics, 2009. **41**(8): p. 936--40.
182. Uhlen, M., et al., *Proteomics. Tissue-based map of the human proteome.* Science, 2015. **347**(6220): p. 1260419.
183. Uhlen, M., et al., *The human secretome.* Sci Signal, 2019. **12**(609).
184. Hebert, D.N. and M. Molinari, *In and out of the ER: protein folding, quality control, degradation, and related human diseases.* Physiol Rev, 2007. **87**(4): p. 1377-408.
185. Sun, Z. and J.L. Brodsky, *Protein quality control in the secretory pathway.* J Cell Biol, 2019. **218**(10): p. 3171-3187.
186. Zanetti, G., et al., *COPII and the regulation of protein sorting in mammals.* Nat Cell Biol, 2011. **14**(1): p. 20-8.
187. Zhang, M. and R. Schekman, *Unconventional secretion, unconventional solutions.* Science, 2013. **340**(6132): p. 559-61.
188. Nickel, W., *Pathways of unconventional protein secretion.* Curr Opin Biotechnol, 2010. **21**(5): p. 621-6.
189. Stuhler, K., K. Schipper, and G. Poschmann, *Unconventional protein secretion: The hidden pathways.* Biochim Biophys Acta Proteins Proteom, 2019. **1867**(12): p. 140272.
190. Basisty, N., et al., *A proteomic atlas of senescence-associated secretomes for aging biomarker development.* PLoS Biol, 2020. **18**(1): p. e3000599.
191. Robinson, J.L., et al., *A Systematic Investigation of the Malignant Functions and Diagnostic Potential of the Cancer Secretome.* Cell Rep, 2019. **26**(10): p. 2622-2635 e5.
192. Leidal, A.M., et al., *The LC3-conjugation machinery specifies the loading of RNA-binding proteins into extracellular vesicles.* Nat Cell Biol, 2020. **22**(2): p. 187-199.
193. Ahat, E., et al., *GRASP55 regulates mutant huntingtin unconventional secretion and aggregation.* bioRxiv, 2021.
194. Schira-Heinen, J., et al., *Pitfalls and opportunities in the characterization of unconventionally secreted proteins by secretome analysis.* Biochim Biophys Acta Proteins Proteom, 2019. **1867**(12): p. 140237.
195. Grube, L., et al., *Mining the Secretome of C2C12 Muscle Cells: Data Dependent Experimental Approach To Analyze Protein Secretion Using Label-Free Quantification and Peptide Based Analysis.* J Proteome Res, 2018. **17**(2): p. 879-890.
196. Poschmann, G., et al., *Comparative Secretomics Gives Access to High Confident Secretome Data: Evaluation of Different Methods for the Determination of Bona Fide Secreted Proteins.* Proteomics, 2021. **21**(2): p. e2000178.

197. Loei, H., et al., *Mining the gastric cancer secretome: identification of GRN as a potential diagnostic marker for early gastric cancer.* J Proteome Res, 2012. **11**(3): p. 1759-72.
198. Nakabayashi, H., et al., *Growth of human hepatoma cells lines with differentiated functions in chemically defined medium.* Cancer Res, 1982. **42**(9): p. 3858-63.
199. Purushothaman, A., *Exosomes from Cell Culture-Conditioned Medium: Isolation by Ultracentrifugation and Characterization.* Methods Mol Biol, 2019. **1952**: p. 233-244.
200. Schneider, C.A., W.S. Rasband, and K.W. Eliceiri, *NIH Image to ImageJ: 25 years of image analysis.* Nat Methods, 2012. **9**(7): p. 671-5.
201. Tank, E.M., et al., *Abnormal RNA stability in amyotrophic lateral sclerosis.* Nat Commun, 2018. **9**(1): p. 2845.
202. McAlister, G.C., et al., *MultiNotch MS3 enables accurate, sensitive, and multiplexed detection of differential expression across cancer cell line proteomes.* Anal Chem, 2014. **86**(14): p. 7150-8.
203. Kong, A.T., et al., *MSFragger: ultrafast and comprehensive peptide identification in mass spectrometry-based proteomics.* Nat Methods, 2017. **14**(5): p. 513-520.
204. Yu, F., et al., *Identification of modified peptides using localization-aware open search.* Nat Commun, 2020. **11**(1): p. 4065.
205. Käll, L., et al., *Semi-supervised learning for peptide identification from shotgun proteomics datasets.* Nat Methods, 2007. **4**(11): p. 923-5.
206. Nesvizhskii, A.I., et al., *A statistical model for identifying proteins by tandem mass spectrometry.* Anal Chem, 2003. **75**(17): p. 4646-58.
207. da Veiga Leprevost, F., et al., *Philosopher: a versatile toolkit for shotgun proteomics data analysis.* Nat Methods, 2020. **17**(9): p. 869-870.
208. Djomehri, S.I., et al., *Quantitative proteomic landscape of metaplastic breast carcinoma pathological subtypes and their relationship to triple-negative tumors.* Nat Commun, 2020. **11**(1): p. 1723.
209. Ritchie, M.E., et al., *limma powers differential expression analyses for RNA-sequencing and microarray studies.* Nucleic Acids Res, 2015. **43**(7): p. e47.
210. De Lisle, R.C., et al., *Metallothionein is a component of exocrine pancreas secretion: implications for zinc homeostasis.* Am J Physiol, 1996. **271**(4 Pt 1): p. C1103-10.
211. Moltedo, O., et al., *Zinc transport and metallothionein secretion in the intestinal human cell line Caco-2.* J Biol Chem, 2000. **275**(41): p. 31819-25.
212. Ishii, H., et al., *FEZ1/LZTS1 gene at 8p22 suppresses cancer cell growth and regulates mitosis.* Proc Natl Acad Sci U S A, 2001. **98**(18): p. 10374-9.
213. Popow, J., et al., *Analysis of orthologous groups reveals archease and DDX1 as tRNA splicing factors.* Nature, 2014. **511**(7507): p. 104-7.
214. Zhou, J., et al., *The protein farnesyltransferase regulates HDAC6 activity in a microtubule-dependent manner.* J Biol Chem, 2009. **284**(15): p. 9648-55.
215. Yu, J., et al., *The E3 ubiquitin ligase HECTD3 regulates ubiquitination and degradation of Tara.* Biochem Biophys Res Commun, 2008. **367**(4): p. 805-12.
216. Tonetti, M., et al., *Synthesis of GDP-L-fucose by the human FX protein.* J Biol Chem, 1996. **271**(44): p. 27274-9.
217. Lichtenthaler, S.F., M.K. Lemberg, and R. Fluhrer, *Proteolytic ectodomain shedding of membrane proteins in mammals-hardware, concepts, and recent developments.* EMBO J, 2018. **37**(15).

218. Grieve, A.G. and C. Rabouille, *Golgi bypass: skirting around the heart of classical secretion.* Cold Spring Harb Perspect Biol, 2011. **3**(4).
219. Lee, P.Y., et al., *Plasma/serum proteomics: depletion strategies for reducing high-abundance proteins for biomarker discovery.* Bioanalysis, 2019. **11**(19): p. 1799-1812.
220. Poschmann, G., et al., *Secretomics-A Key to a Comprehensive Picture of Unconventional Protein Secretion.* Front Cell Dev Biol, 2022. **10**: p. 878027.
221. Nickel, W., *The unconventional secretory machinery of fibroblast growth factor 2.* Traffic, 2011. **12**(7): p. 799-805.
222. Popa, S.J., S.E. Stewart, and K. Moreau, *Unconventional secretion of annexins and galectins.* Semin Cell Dev Biol, 2018. **83**: p. 42-50.
223. Sulkowski, P.L., et al., *Horizontal Transfer of Histone H3 by Mammalian Cells.* bioRxiv, 2021.
224. Uhlen, M., et al., *Tissue-based map of the human proteome.* Science, 2015. **347**(6220): p. 1260419.
225. Markova, E.A. and G. Zanetti, *Visualizing membrane trafficking through the electron microscope: Cryo-tomography of coat complexes.* Acta Crystallographica Section D: Structural Biology, 2019. **75**(5): p. 467--474.
226. Neerman-Arbez, M., et al., *Molecular analysis of the ERGIC-53 gene in 35 families with combined factor V-factor VIII deficiency.* Blood, 1999. **93**(7): p. 2253-60.
227. Zhang, B., et al., *Mice deficient in LMAN1 exhibit FV and FVIII deficiencies and liver accumulation of alpha1-antitrypsin.* Blood, 2011. **118**(12): p. 3384-91.
228. Belden, W.J., *Role of Erv29p in Collecting Soluble Secretory Proteins into ER-Derived Transport Vesicles.* Science, 2001. **294**(5546): p. 1528--1531.
229. Caldwell, S.R., K.J. Hill, and A.A. Cooper, *Degradation of Endoplasmic Reticulum (ER) Quality Control Substrates Requires Transport between the ER and Golgi.* Journal of Biological Chemistry, 2001. **276**(26): p. 23296--23303.
230. Wang, B., et al., *Atherosclerosis-associated hepatic secretion of VLDL but not PCSK9 is dependent on cargo receptor protein Surf4.* Journal of Lipid Research, 2021. **62**: p. 100091.
231. Tang, X., T. Wang, and Y. Guo, *Export of polybasic motif-containing secretory proteins BMP8A and SFRP1 from the endoplasmic reticulum is regulated by surfeit locus protein 4 (SURF4): SURF4 regulates ER export of BMP8A and SFRP1.* J Biol Chem, 2022: p. 102687.
232. Abbineni, P.S., et al., *Identification of secreted proteins by comparison of protein abundance in conditioned media and cell lysates.* Anal Biochem, 2022. **655**: p. 114846.
233. Nakabayashi, H., et al., *Growth of human hepatoma cell lines with differentiated functions in chemically defined medium.* Cancer Research, 1982. **42**(9): p. 3858--3863.
234. Sanjana, N.E., O. Shalem, and F. Zhang, *Improved vectors and genome-wide libraries for CRISPR screening.* Nat Methods, 2014. **11**(8): p. 783-784.
235. Emmer, B.T., et al., *The cargo receptor SURF4 promotes the efficient cellular secretion of PCSK9.* Elife, 2018. **7**.
236. UniProt, C., *UniProt: the universal protein knowledgebase in 2021.* Nucleic Acids Res, 2021. **49**(D1): p. D480-D489.
237. Shen, Y., et al., *Surf4 regulates expression of proprotein convertase subtilisin/kexin type 9 (PCSK9) but is not required for PCSK9 secretion in cultured human hepatocytes.*

Biochimica et Biophysica Acta (BBA) - Molecular and Cell Biology of Lipids, 2020. **1865**(2): p. 158555.
238. Emmer, B.T., et al., *Murine Surf4 is essential for early embryonic development.* bioRxiv, 2019.
239. Chapman, M.J., et al., *Triglyceride-rich lipoproteins and high-density lipoprotein cholesterol in patients at high risk of cardiovascular disease: Evidence and guidance for management.* European Heart Journal, 2011. **32**(11): p. 1345--1361.
240. Kolovou, G., et al., *Microsomal Transfer Protein Inhibitors, New Approach for Treatment of Familial Hypercholesterolemia, Review of the Literature, Original Findings, and Clinical Significance.* Cardiovascular Therapeutics, 2015. **33**(2): p. 71--78.
241. Benjannet, S., et al., *NARC-1/PCSK9 and its natural mutants: Zymogen cleavage and effects on the low density lipoprotein (LDL) receptor and LDL cholesterol.* Journal of Biological Chemistry, 2004. **279**(47): p. 48865--48875.
242. Bonifacino, J.S. and B.S. Glick, *The mechanisms of vesicle budding and fusion.* Cell, 2004. **116**(2): p. 153-66.
243. Platt, R.J., et al., *CRISPR-Cas9 knockin mice for genome editing and cancer modeling.* Cell, 2014. **159**(2): p. 440-55.
244. Millar, J.S., et al., *Determining hepatic triglyceride production in mice: comparison of poloxamer 407 with Triton WR-1339.* Journal of Lipid Research, 2005. **46**(9): p. 2023--2028.
245. Kusminski, C.M., et al., *MitoNEET-driven alterations in adipocyte mitochondrial activity reveal a crucial adaptive process that preserves insulin sensitivity in obesity.* Nature Medicine, 2012. **18**(10): p. 1539--1551.
246. Ye, R., et al., *Adiponectin is essential for lipid homeostasis and survival under insulin deficiency and promotes β-cell regeneration.* eLife, 2014. **3**: p. 1--21.
247. Folch, J., M. Lees, and G.H.S. Stanley, *A Simple Method for the Isolation and Purification of Total Lipides from Animal Tissues.* Journal of Biological Chemistry, 1957. **226**(1): p. 497--509.
248. Dobin, A., et al., *STAR: Ultrafast universal RNA-seq aligner.* Bioinformatics, 2013. **29**(1).
249. Li, B. and C.N. Dewey, *RSEM: Accurate transcript quantification from RNA-Seq data with or without a reference genome.* BMC Bioinformatics, 2011. **12**.
250. Love, M.I., W. Huber, and S. Anders, *Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2.* Genome Biology, 2014. **15**(12).
251. Garrido-Martn, D., E. Palumbo, and Guig, *ggsashimi: Sashimi plot revised for browser- and annotation-independent splicing visualization.* PLoS Computational Biology, 2018. **14**(8): p. 1--6.
252. Popp, M.W. and L.E. Maquat, *Organizing principles of mammalian nonsense-mediated mRNA decay.* Annu Rev Genet, 2013. **47**: p. 139-65.
253. Cohen, J., et al., *Low LDL cholesterol in individuals of African descent resulting from frequent nonsense mutations in PCSK9.* Nat Genet, 2005. **37**(2): p. 161-5.
254. Hooper, A.J., et al., *The C679X mutation in PCSK9 is present and lowers blood cholesterol in a Southern African population.* Atherosclerosis, 2007. **193**(2): p. 445-8.
255. Cohen, J.C. and H.H. Hobbs, *Simple genetics for a complex disease.* Science, 2013. **340**(6133): p. 689-90.

256. Rashid, S., et al., *Decreased plasma cholesterol and hypersensitivity to statins in mice lacking Pcsk9.* Proceedings of the National Academy of Sciences, 2005. **102**(15): p. 5374--5379.
257. Freeze, H.H. and C. Kranz, *Endoglycosidase and glycoamidase release of N-linked glycans*. Vol. 2010. 2010. 1--33.
258. Ipsen, D.H., J. Lykkesfeldt, and P. Tveden-Nyborg, *Molecular mechanisms of hepatic lipid accumulation in non-alcoholic fatty liver disease.* Cellular and Molecular Life Sciences, 2018. **75**(18): p. 3313--3327.
259. Hetz, C., K. Zhang, and R.J. Kaufman, *Mechanisms, regulation and functions of the unfolded protein response.* Nat Rev Mol Cell Biol, 2020. **21**(8): p. 421-438.
260. Tang, A.T., et al., *Endothelial TLR4 and the microbiome drive cerebral cavernous malformations.* Nature, 2017. **545**(7654): p. 305--310.
261. Chang, X., et al., *Loss of Hepatic Surf4 Depletes Lipid Droplets in the Adrenal Cortex but Does Not Impair Adrenal Hormone Production.* Frontiers in Cardiovascular Medicine, 2021. **0**: p. 1613.
262. Oda, Y., et al., *Derlin-2 and Derlin-3 are regulated by the mammalian unfolded protein response and are required for ER-associated degradation.* Journal of Cell Biology, 2006. **172**(3): p. 383--393.
263. Musunuru, K., et al., *In vivo CRISPR base editing of PCSK9 durably lowers cholesterol in primates.* Nature, 2021. **593**(7859): p. 429--434.
264. Shen, Y., et al., *The role of hepatic Surf4 in lipoprotein metabolism and the development of atherosclerosis in apoE(-/-) mice.* Biochim Biophys Acta Mol Cell Biol Lipids, 2022. **1867**(10): p. 159196.
265. Steinberg, M.H., *Fetal hemoglobin in sickle cell anemia.* Blood, 2020. **136**(21): p. 2392-2400.
266. Drysdale, C.M., et al., *Hematopoietic-Stem-Cell-Targeted Gene-Addition and Gene-Editing Strategies for beta-hemoglobinopathies.* Cell Stem Cell, 2021. **28**(2): p. 191-208.
267. Bou-Fakhredin, R., et al., *Pharmacological Induction of Fetal Hemoglobin in beta-Thalassemia and Sickle Cell Disease: An Updated Perspective.* Pharmaceuticals (Basel), 2022. **15**(6).
268. Stomberski, C.T., et al., *A multienzyme S-nitrosylation cascade regulates cholesterol homeostasis.* Cell Rep, 2022. **41**(4): p. 111538.
269. Abifadel, M., et al., *Mutations in PCSK9 cause autosomal dominant hypercholesterolemia.* Nature genetics, 2003. **34**(2): p. 154--6.
270. Robinson, J.G., et al., *Efficacy and Safety of Alirocumab in Reducing Lipids and Cardiovascular Events.* New England Journal of Medicine, 2015. **372**(16): p. 1489--1499.
271. Sabatine, M.S., et al., *Efficacy and Safety of Evolocumab in Reducing Lipids and Cardiovascular Events.* New England Journal of Medicine, 2015. **372**(16): p. 1500--1509.
272. Musunuru, K., et al., *In vivo CRISPR base editing of PCSK9 durably lowers cholesterol in primates.* Nature, 2021. **593**(7859): p. 429--434.