Novel Genes for Dyslipidemias: Genome-Wide Association Studies

Kiran Musunuru

Introduction

As outlined in some of the preceding chapters of this book, a number of genes responsible for dyslipidemias have been discovered because they harbor rare mutations that cause a very small number of individuals—typically in one or a few families—to have profoundly dysregulated blood lipid concentrations. In other words, these few individuals have monogenic or "Mendelian" dyslipidemias caused by the aberrant action of single genes. Although researchers have learned much about cholesterol metabolism by studying these genes over the past few decades, this work has left unanswered the question of what causes a significant proportion of the general population to have what might be thought of as garden variety dyslipidemias—high but not exceptionally high blood low-density lipoprotein cholesterol (LDL-C) concentrations, low but not exceptionally low blood high-density lipoprotein cholesterol (HDL-C) concentrations, and/or high but not exceptionally high blood triglyceride (TG) concentrations. It has been a popular belief that in these patients, in whom there do not seem to be single aberrant genes driving the abnormal lipid levels, the dyslipidemias are polygenic in nature, i.e., caused by the actions of multiple genes in tandem. By this reasoning, if the functions of 5, 10, 20, or 50 genes were slightly dysregulated, the combined effect would result in abnormal lipid levels.

Although the notion of polygenic dyslipidemias is attractive to many researchers, it had not been possible to test this model by detecting and measuring the slight dysregulation of each of the numerous genes that would be involved—indeed, it was not even clear *which* genes out of the roughly 20,000 genes in the human genome might be involved. In the past few years, the completion of the Human Genome Project and advances in genotyping and sequencing technologies have made it possible for the first time to do unbiased searches for genes that make small contributions to blood lipid levels in dyslipidemia patients. This chapter focuses on the methodology known as the genome-wide association study (GWAS) and summarizes the advances in knowledge regarding cholesterol metabolism that have emerged from the application of this methodology to tens of thousands of people and subsequent work to identify and characterize novel genes involved in dyslipidemias.

A Primer on Genome-Wide Association Studies

The human genome is roughly three billion DNA bases in size, spanning 23 chromosome pairs; the vast majority of the sequence is identical across

K. Musunuru (✉)
Brigham and Women's Hospital, Harvard University, Sherman Fairchild Biochemistry Bldg 160 7 Divinity Ave, 02138 Cambridge, MA, USA
e-mail: kiranmusunuru@gmail.com

the human species. What makes each individual unique is a large number of DNA variants distributed throughout the genome. Some of these DNA variants are extremely rare and have large effects on gene function; as described above, these variants can be responsible for monogenic disorders. Other DNA variants are quite common, occurring in >1% of the general population and, in some cases, the majority of the people in a population. Most of these common DNA variants are of no functional consequence, but some lie either within or near genes and have small effects on gene function. These variants do not alter gene activity enough to cause disease by themselves, but instead need to be combined with other gene variants or with environmental factors in order for disease to occur.

All of these common DNA variants are termed polymorphisms, of which there are several varieties. The most relevant to the use of GWAS to study polygenic disorders is the single nucleotide polymorphism (SNP), in which a single base pair in the DNA differs from the usual base pair at that position. There are an estimated 11 million SNPs across the human genome, occurring on an average every few hundred base pairs. A local area on a chromosome around an SNP is termed a locus. Each person has two copies of each locus because of the pairing of chromosomes (the exceptions are loci on the X or Y chromosome in men, who have only one of each). A person's genotype at an SNP is the identity of the base pair position for each of the two copies—also termed alleles—of the SNP on paired chromosomes; thus, a genotype is typically a combination of two alleles. These two alleles may be identical (termed homozygosity) or different (termed heterozygosity).

Groups of SNP alleles near genes tend to stay together with the genes as they are passed along from parents to children for generation after generation, over thousands of years. Thus, even if it is not known which gene contributes to a disease, one can use an SNP that is not in the gene—but is near to and therefore linked to the gene—as a "tag" for the gene. In the past decade, the technology has become available to determine the genotypes at hundreds of thousands of "tag" SNPs in a person's DNA in a single experiment

using a "gene chip." By applying the gene chips to thousands of individuals, some with a disease and some without a disease, researchers are able to identify tag SNPs that are associated with disease. This is the principle underlying GWAS.

As an example of how a GWAS might be performed, imagine a study in which DNA samples are collected from several thousand individuals with high blood LDL-C levels (e.g., >200 mg/dL) and several thousand individuals with low blood LDL-C levels (e.g., <60 mg/dL in the absence of lipid-lowering medications). For each study participant, a gene chip would be used to determine the genotypes of more than 1 million SNPs across the genome. Although the use of gene chips on thousands of people would yield billions of pieces of data, the statistical methods to analyze these data are conceptually straightforward. Computer software is used to analyze each of the 1 million SNPs separately; for each SNP, the question is asked whether allele "A" and allele "B" of the SNP occur in equal proportions in the high LDL-C cohort and in the low LDL-C cohort. For the vast majority of the 1 million SNPs, no difference in the allele proportions would be observed. For a particular SNP, however, there might be a statistically significant difference in the allele proportions such that allele "A" occurs more commonly in the high LDL-C cohort than in the low LDL-C cohort. Because the SNP tags any nearby genes, the conclusion would be that there is a DNA variant in one of the local genes that influences that gene's function in such a way as to influence blood LDL-C levels. From a researcher's perspective, the SNP acts as a "signpost" indicating that somewhere in the locus lies the key to a biological mechanism that contributes to dyslipidemia in the general population. (In fact, some of the million tested SNPs are very close to one another and effectively tag the same locus, so it is the locus rather than any individual SNP that is considered to be a GWAS discovery.)

In practice, it would be difficult to recruit thousands of study participants with either very high or very low LDL-C levels due to their relative scarcity in the population. In an alternative study design, thousands of people from the gen-

eral population at large would be recruited, and computer analysis would be performed using the blood LDL-C levels as a continuous rather than a categorical variable. The question would be framed in a different way: For a given SNP, what are the average LDL-C levels for individuals who are homozygous for allele "A" versus individuals who are heterozygous for alleles "A" and "B" versus individuals who are homozygous for allele "B"? If there are statistically significant differences among the three genotype groups, the SNP/locus would be considered to be associated with blood LDL-C levels.

It is worth noting that in either of these GWAS study designs, there are effectively a million experiments being performed, one for each individual SNP. As such, the traditional statistical significance threshold of $P < 0.05$ is inappropriate, since by that threshold 50,000 SNPs (5 % of 1,000,000) would be associated with LDL-C levels by chance alone, with most if not all being false positives. Accordingly, GWAS researchers insist on the statistical significance threshold being much more stringent, e.g., by adjusting for the number of experiments (known as the Bonferroni correction) such that P should be less than $0.05 \div 1,000,000$, or $P < 5 \times 10^{-8}$, for the SNP/locus to be regarded as having a true association.

GWAS on Blood Lipid Traits

One consequence of the GWAS study design is that it becomes increasingly powered to detect associations as the number of study participants grows. Accordingly, the past several years have seen successive reports of increasingly larger GWAS studies of blood lipid concentrations, beginning with a few thousand individuals and culminating in more than 100,000 individuals. As such, the list of reported lipid-associated GWAS loci has substantially grown over that time period.

The first published high-density GWAS on blood lipid concentrations was performed with data from about 3000 individuals of European descent in the Diabetes Genetics Initiative. This study identified one statistically significant locus each for three lipid traits—LDL-C, HDL-C, and TG [1]. The LDL-C locus contained the apolipoprotein E (*APOE*) gene, and the HDL-C locus contained the cholesteryl ester transfer protein (*CETP*) gene—both well-established regulators of lipoprotein metabolism. The TG locus contained no previously identified lipid regulators; the single gene in the locus is glucokinase regulatory protein (*GCKR*). Subsequent functional experiments pointed to a coding missense variant in the *GCKR* gene as being responsible for the TG association [2, 3]. The fact that this first GWAS identified two known lipid genes provided strong validation of the study design as well as giving confidence that the *GCKR* locus was a true positive (and novel) finding.

A second set of published GWAS studies for blood lipid concentrations added the Finland–US investigation of NIDDM (noninsulin-dependent diabetes mellitus) genetics study (FUSION) and SardiNIA cohorts to the Diabetes Genetics Initiative for a cohort of about 9000 individuals of European descent [4, 5]. As a method of increasing the studies' power to detect statistically significant associations, the researchers used a staged approach: They selected the SNPs with the best P values from the analysis of data from the initial 9000 individuals and genotyped just those SNPs in an additional 18,000 individuals of European descent from several other cohorts. This approach yielded a total of 19 statistically significant lipid-associated loci. In addition to the three loci identified by the first GWAS (*APOE, CETP, GCKR*), the list now included many more well-established lipid regulators, including apolipoprotein A-I (*APOA1*), apolipoprotein B (*APOB*), LDL receptor (*LDLR*), lipoprotein lipase (*LPL*), proprotein convertase subtilisin/kexin type 9 (*PCSK9*), and 3-hydroxy-3-methyl-glutaryl-coenzyme A reductase (*HMGCR*). The last is noteworthy because it encodes the enzyme that is targeted by the statin class of LDL-C-lowering drugs. These GWAS studies also identified six novel loci, two of which were also discovered in simultaneously published GWAS studies that focused solely on LDL-C (a locus on chromosome 1p13) or TG (a locus on chromosome 7q11) [6–8].

A third set of GWAS studies on blood lipid concentrations, analyzing data from up to 40,000 individuals of European descent, identified more than 30 lipid-associated loci, half of them harboring well-established lipid regulators, the other half novel, continuing the trend observed from the first two sets of studies [9–11].

Finally, a definitive GWAS combining data from all of the cohorts with which the previous three sets of GWAS studies had been performed—collectively termed the Global Lipids Genetics Consortium (GLGC)—used more than 100,000 individuals of European descent to identify a total of 95 loci associated with one or more blood lipid concentrations—LDL-C, HDL-C, TG, and/or total cholesterol (Table 17.1) [12]. Many of these loci were also shown to be associated with lipids in other ethnic groups, including East Asians, South Asians, and African Americans. About two thirds of these loci are novel. Among the remaining one third of the loci are 16 genes that have been implicated in familial lipid disorders (Table 17.2), highlighting that the same gene can contribute to both monogenic and polygenic dyslipidemias, with a rare variant in the gene greatly perturbing its function and causing disease, and with a common variant in the gene mildly perturbing its function and combining with common variants in other genes to collectively produce disease.

In support of the polygenic model of disease, the GLGC study calculated SNP "risk scores" that summarized the number of LDL-C-raising, HDL-C-raising, or TG-raising SNP alleles in each individual who had high or low LDL-C levels (mean 219 mg/dL vs. mean 110 mg/dL), high or low HDL-C levels (mean 90 mg/dL vs. mean 36.2 mg/dL), or high or low TG levels (mean 1,079 mg/dL vs. mean 106 mg/dL). Individuals with LDL-C risk scores in the top quartile of the combined GLGC cohort were 13 times as likely to have a high-LDL-C level than individuals with scores in the bottom quartile; individuals with HDL-C risk scores in the top quartile of the combined GLGC cohort were four times as likely to have a high-HDL-C level than individuals with scores in the bottom quartile; and individuals with TG risk scores in the top quartile

of the combined GLGC cohort were 44 times as likely to have a high-TG level than individuals with scores in the bottom quartile [12]. These results confirm that the additive effects of multiple common variants do indeed contribute to dyslipidemias in many individuals.

In spite of these data, there have continued to be criticisms that the common variants discovered by GWAS have little clinical relevance, since the effects on gene function are small. Besides disregarding the polygenic model of disease, such arguments ignore the possibility that a GWAS gene can turn out to be clinically important if its activity is modulated by a large degree with pharmacological intervention, with *HMGCR* being the prototypic example. If the discovery of statins had not predated the GWAS era, the finding that *HMGCR* is in a locus associated with LDL-C levels would have suggested to researchers that pharmacological inhibition of *HMGCR* might be a viable therapeutic strategy. By this reasoning, some of the novel lipid GWAS genes may emerge as clinically useful drug targets, as described below.

Novel Genes That Have Emerged from GWAS

Although the 95 lipid-associated loci identified by the GLGC study potentially point to dozens of novel genes, to date only a few such candidates have been studied with functional experimentation. We focus on three genes from which novel insights into lipoprotein metabolism have started to be gleaned.

SORT1. Perhaps the most compelling of the novel lipid GWAS loci lies on chromosome 1p13. SNPs in the 1p13 locus have among the strongest associations with LDL-C of any loci in the genome. Individuals who are homozygous for the more common allele of one of these SNPs have an average 16 mg/dL higher blood LDL-C concentration than individuals who are homozygous for the less common allele [4]. The former also have about a 40 % increase in risk of myocardial infarction compared to the latter [13]. Thus, the 1p13 locus is strongly associated with both blood

Table 17.1 GWAS loci associated with blood lipid traits

Locus/gene	Chr	Best SNP	Best trait	Other traits	P
LDLRAP1	1	rs12027135	TC	LDL	4E-11
PABPC4	1	rs4660293	HDL		4E-10
PCSK9	1	rs2479409	LDL	TC	2E-28
ANGPTL3	1	rs2131925	TG	TC, LDL	9E-43
EVI5	1	rs7515577	TC		3E-8
SORT1	1	rs629301	LDL	TC	1E-170
ZNF648	1	rs1689800	HDL		3E-10
MOSC1	1	rs2642442	TC	LDL	6E-13
GALNT2	1	rs4846914	HDL	TG	4E-21
IRF2BP2	1	rs514230	TC	LDL	5E-14
APOB	2	rs1367117	LDL	TC	4E-114
		rs1042034	TG	HDL	1E-45
GCKR	2	rs1260326	TG	TC	6E-133
ABCG5/8	2	rs4299376	LDL	TC	2E-47
RAB3GAP1	2	rs7570971	TC		2E-8
COBLL1	2	rs10195252	TG		2E-10
		rs12328675	HDL		3E-10
IRS1	2	rs2972146	HDL	TG	3E-9
RAF1	3	rs2290159	TC		4E-9
MSL2L1	3	rs645040	TG		3E-8
KLHL8	4	rs442177	TG		9E-12
SLC39A8	4	rs13107325	HDL		7E-11
ARL15	5	rs6450176	HDL		5E-8
MAP3K1	5	rs9686661	TG		1E-10
HMGCR	5	rs12916	TC	LDL	9E-47
TIMD4	5	rs6882076	TC	LDL, TG	7E-28
MYLIP	6	rs3757354	LDL	TC	1E-11
HFE	6	rs1800562	LDL	TC	6E-10
HLA	6	rs3177928	TC	LDL	4E-19
		rs2247056	TG		2E-15
C6orfTO6	6	rs2814944	HDL		4E-9
		rs2814982	TC		5E-11
ABCA1	9	rs1883025	HDL	TC	2E-33
ABO	9	rs9411489	LDL	TC	6E-13
JMJD1C	10	rs10761731	TG		3E-12
CYP26A1	10	rs2068888	TG		2E-8
GPAM	10	rs2255141	TC	LDL	2E-10
AMPD3	11	rs2923084	HDL		5E-8
SPTY2D1	11	rs10128711	TC		3E-8
LRP4	11	rs3136441	HDL		3E-18
FADS1-2-3	11	rs174546	TG	HDL, TC, LDL	5E-24
APOA1	11	rs964184	TG	TC, HDL, LDL	7E-240
UBASH3B	11	rs7941030	TC	HDL	2E-10
ST3GAL4	11	rs11220462	LDL	TC	1E-15
PDE3A	12	rs7134375	HDL		4E-8
LRP1	12	rs11613352	TG	HDL	4E-10
MVK	12	rs7134594	HDL		7E-15
BRAP	12	rs11065987	TC	LDL	7E-12
HNF1A	12	rs1169288	TC	LDL	1E-14
SBNO1	12	rs4759375	HDL		7E-9
ZNF664	12	rs4765127	HDL	TG	3E-10
SCARB1	12	rs838880	HDL		3E-14
NYNRIN	14	rs8017377	LDL		5E-11
CAPN3	15	rs2412710	TG		2E-8
FRMD5	15	rs2929282	TG		2E-11
LIPC	15	rs1532085	HDL	TC, TG	3E-96
LACTB	15	rs2652834	HDL		9E-9
CTF1	16	rs11649653	TG		3E-8
CETP	16	rs3764261	HDL	TC, LDL, TG	7E-380
LCAT	16	rs16942887	HDL		8E-33
HPR	16	rs2000999	TC	LDL	3E-24
CMIP	16	rs2925979	HDL		2E-11
STARD3	17	rs11869286	HDL		1E-13
OSBPL7	17	rs7206971	LDL	TC	2E-8

Table 17.1 (continued)

Locus/gene	Chr	Best SNP	Best trait	Other traits	P	Locus/gene	Chr	Best SNP	Best trait	Other traits	P
FRK	6	rs9488822	TC	LDL	2E-10	ABCA8	17	rs4148008	HDL		2E-10
CITED2	6	rs605066	HDL		3E-8	PGS1	17	rs4129767	HDL		8E-9
LPA	6	rs1564348	LDL	TC	2E-17	LIPG	18	rs7241918	HDL	TC	3E-49
		rs1084651	HDL		3E-8	MC4R	18	rs12967135	HDL		7E-9
DNAH11	7	rs12670798	TC	LDL	9E-10	ANGPTL4	19	rs7255436	HDL		3E-8
NPC1L1	7	rs2072183	TC	LDL	3E-11	LDLR	19	rs6511720	LDL	TC	4E-117
TYW1B	7	rs13238203	TG		1E-9	LOC55908	19	rs737337	HDL		3E-9
MLXIPL	7	rs17145738	TG	HDL	6E-58	CILP2	19	rs10401969	TC	TG, LDL	3E-38
KLF14	7	rs4731702	HDL		1E-15	APOE	19	rs4420638	LDL	TC, HDL	9E-147
PPP1R3B	8	rs9987289	HDL	TC, LDL	6E-25	FLJ36070	19	rs492602	TC		2E-10
PINX1	8	rs11776767	TG		1E-8	LILRA3	19	rs386000	HDL		4E-16
NAT2	8	rs1495741	TG	TC	5E-14	ERGIC3	20	rs2277862	TC		4E-10
LPL	8	rs12678919	TG	HDL	2E-115	MAFB	20	rs2902940	TC	LDL	6E-11
CYP7A1	8	rs2081687	TC	LDL	2E-12	TOP1	20	rs6029526	LDL	TC	4E-19
TRPS1	8	rs2293889	HDL		6E-11	HNF4A	20	rs1800961	HDL	TC	1E-15
		rs2737229	TC		2E-8	PLTP	20	rs6065906	HDL	TG	2E-22
TRIB1	8	rs2954029	TG	TC, LDL, HDL	3E-55						
PLEC1	8	rs11136341	LDL	TC	4E-13						
TTC39B	9	rs581080	HDL	TC	3E-12						

The listed gene is either an established lipid-related gene in the locus or the gene closest to the best SNP with each best lipid trait for each locus. *P* values are given for the associations of each best SNP with each best trait closest to the best SNP

HDL high-density lipoprotein cholesterol, *LDL* low-density lipoprotein cholesterol, *SNP* single-nucleotide polymorphism, *TG* triglycerides, *TC* total cholesterol

Table 17.2 Monogenic dyslipidemias caused by genes with nearby common DNA variants identified in lipid GWAS

Gene	Chr	GWAS SNP	Monogenic dyslipidemia	Associated traits
ABCA1	9	rs1883025	Tangier disease	Low HDL
ABCG5	2	rs4299376	Sitosterolemia	High LDL
ABCG8	2	rs4299376	Sitosterolemia	High LDL
ANGPTL3	1	rs2131925	Familial combined hypolipidemia	Low LDL, low HDL, low TG
APOA1	11	rs964184	ApoA-I deficiency	Low HDL
APOA5	11	rs964184	ApoA-V deficiency	High VLDL, high chylomicrons
APOB	2	rs1367117	Familial hypobetalipoproteinemia	Low LDL
			Familial defective ApoB-100	High LDL
APOC2	19	rs4420638	Familial ApoC-II deficiency	High chylomicrons
APOE	19	rs4420638	Familial dysbetalipoproteinemia	High VLDL, high chylomicrons
CETP	16	rs3764261	Cholesteryl ester transfer protein deficiency	High HDL
LCAT	16	rs16942887	LCAT deficiency (fish-eye disease)	Low HDL
LDLR	19	rs6511720	Familial hypercholesterolemia	High LDL
LDLRAP1	1	rs12027135	Autosomal recessive hypercholesterolemia	High LDL
LIPC	15	rs1532085	Familial hepatic lipase deficiency	High VLDL remnants
LPL	8	rs12678919	Lipoprotein lipase deficiency	High chylomicrons
PCSK9	1	rs2479409	PCSK9 deficiency	Low LDL
			Autosomal dominant hypercholesterolemia	High LDL

GWAS genome-wide association study, *HDL* high-density lipoprotein, *LDL* low-density lipoprotein, *SNP* single-nucleotide polymorphism, *TG* triglycerides, *VLDL* very low-density lipoprotein

lipids and the most serious clinical phenotype resulting from dyslipidemias.

The 1p13 locus harbors several genes including *CELSR2, PSRC1,* and *SORT1*, none of which had previously been linked to lipid metabolism in the pre-GWAS era. Functional experimentation with these genes in mouse models revealed that *SORT1* (which encodes the sortilin protein) modulated blood lipid levels when its expression was either increased or decreased in mouse liver, or when it was deleted in mice altogether [14, 15]. Cell-based experiments have discovered two roles for sortilin in lipoprotein metabolism. First, it regulates blood LDL-C levels by reducing the secretion of VLDL particles from the liver into the bloodstream, where the VLDL particles are ultimately converted to LDL particles [14]. It appears to do this in hepatocytes by directly binding apolipoprotein B (apoB)—the core protein of VLDL/LDL particles—in the endoplasmic reticulum/Golgi apparatus and trafficking apoB to the endolysosomal compartment for degradation, thereby reducing the number of VLDL particles produced and, ultimately, secreted [14, 16].

Second, sortilin appears to be able to function as an alternative LDL receptor by binding to apoB-carrying LDL particles in the bloodstream and facilitating the endocytosis of the particles into the cell, followed by their degradation in the endolysosmal compartment [16]. Both mechanisms should have the effect of lowering blood LDL-C concentrations (Fig. 17.1), consistent with the association of the 1p13 locus with LDL-C in human populations.

The strong association of the 1p13 locus with myocardial infarction suggests *SORT1* as a plausible clinical drug target, as modulation of the gene would be expected to not only reduce blood LDL-C levels but also the risk of myocardial infarction. However, the biological evidence indicates that increasing sortilin activity in liver would be a therapeutically useful intervention, which may be difficult to achieve with traditional therapies (in contrast to inhibiting an enzyme's activity, as is the case with statin drugs and *HMGCR*).

TRIB1. Another compelling novel lipid GWAS locus lies on chromosome 8q24. SNPs in
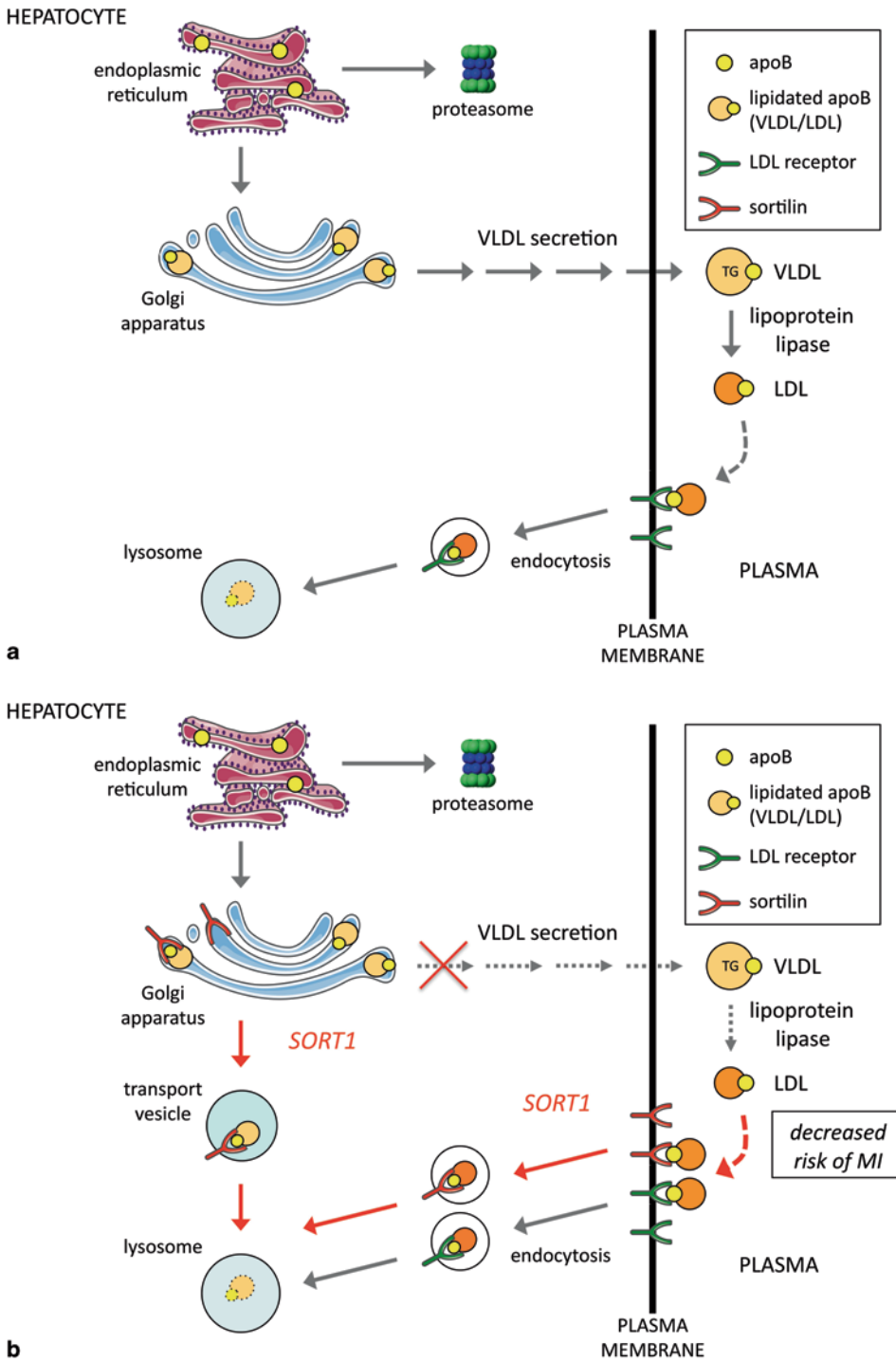
Fig. 17.1 Model of sortilin actions in hepatocytes **a** Lipoprotein production, secretion, uptake, and degradation in hepatocytes. **b** Sortilin decreases very low-density lipoprotein (VLDL) particle secretion by trafficking nascent particles to the endolysosomal compartment. It also acts as an alternative LDL receptor to facilitate endocytosis of LDL particles into the cell and degradation in the endolysosomal compartment. The consequence of both actions is to reduce LDL cholesterol levels in the blood and thereby reduce the risk of myocardial infarction (MI)

this locus have a distinctive pattern of association with multiple lipid traits, with the minor allele conferring decreased LDL-C levels, increased HDL-C levels, and decreased TG levels—all changes that are epidemiologically associated with lower risk of coronary artery disease (CAD) [4, 5]. Perhaps not surprisingly, then, the same SNPs are themselves associated with CAD risk [12].

There is a single gene in the locus, *TRIB1* (tribbles homolog 1), which had not previously been linked to lipid metabolism in the pre-GWAS era. Two types of functional experimentation have been undertaken in mice [17]. First, *Trib1* knockout mice (in which the gene has been deleted) were observed to have increased blood levels of cholesterol and TG. Second, mice in which *Trib1* was overexpressed in liver showed the opposite effect—decreased blood levels of cholesterol and TG. The mechanism appears in part to be related to hepatic production of VLDL particles; *Trib1* overexpression resulted in decreased particle production and secretion, whereas the knockout mice displayed the opposite effect [17]. This was reproduced in cell-based experiments in which *TRIB1* was overexpressed in cultured human hepatoma cells, resulting in decreased apoB particle secretion. Furthermore, *Trib1* appears to reduce the hepatic expression of genes involved in lipogenesis, including *Acc1, Fasn*, and *Scd1* [17]. The mechanism(s) through which the protein produces these effects remains to be defined.

As with *SORT1*, it would appear that increasing *TRIB1* expression in the liver would be a therapeutically useful intervention. Indeed, a *TRIB1*-targeting strategy might be of even greater clinical benefit than a *SORT1*-targeting strategy since GWAS data suggest that *TRIB1* is linked to decreased LDL-C, increased HDL-C, and decreased TG in the blood, whereas *SORT1* appears to be linked solely to decreased LDL-C. Moreover, the genetic association of SNPs in the *TRIB1* locus with CAD risk offers reassurance that targeting *TRIB1* would have the desired clinical outcome—a reduction in CAD rather than just modulation of blood lipid levels.

GALNT2. A novel lipid GWAS locus on chromosome 1q42 that is associated with both HDL-C and TG levels in the blood harbors a single gene, *GALNT2,* which encodes UDP-N-acetyl-alpha-D-galactosamine:polypeptide N-acetylgalactosaminyltransferase-2, a member of a family of proteins that are involved in the initiation of mucin-type O-linked glycosylation on various proteins. As with *SORT1* and *TRIB1*, no prior role of GALNT2 in lipoprotein metabolism was known. Unlike *SORT1* and *TRIB1,* SNPs in the locus do not have any significant association with CAD.

Functional experimentation with *GALNT2* in mice in which the gene was either overexpressed or knocked down in the liver indicated that the gene negatively regulates blood HDL-C levels [12]. The connection of *GALNT2* with blood HDL-C and TG concentrations in humans was confirmed by the identification of two families with individuals with dyslipidemia—very high HDL-C and low TG levels—who were heterozygous for a loss-of-function missense mutation in *GALNT2* [18]. Physiological studies of these individuals suggested that they had improved postprandial TG clearance due to impaired glycosylation of apolipoprotein C-III (apoC-III), which normally inhibits LPL, which itself hydrolyzes and thus decreases TG in the blood. Thus, *GALNT2* represents another example of a gene for which common DNA variants produce a small effect on blood lipid levels and rare DNA variants can single-handedly produce dyslipidemias. However, to date there is no evidence that *GALNT2* is associated with a change in CAD risk, and thus its relevance as a therapeutic target remains in question.

Conclusion

The past few years have witnessed remarkable progress in human genetics towards the understanding of polygenic disorders. Blood lipid concentrations represent some of the most successfully studied clinical traits with the use of GWAS. Whereas our prior knowledge of the genetics of

dyslipidemias was limited to rare variants causing monogenic disorders, GWAS studies have now identified dozens of novel loci that appear to contribute to polygenetic dyslipidemias. A key challenge will be to determine the molecular mechanisms by which these loci influence blood lipid concentrations. Although functional studies of the novel loci are starting to yield new insights into lipoprotein metabolism, as demonstrated by the examples of the genes *SORT1, TRIB1,* and *GALNT2,* there are undoubtedly as many new discoveries to be made with respect to lipoprotein metabolism as have been made in the past few decades of investigation.

References

1. Diabetes Genetics Initiative of Broad Institute of Harvard and MIT, Lund University, and Novartis Institutes of BioMedical Research, Saxena R, Voight BF, Lyssenko V, Burtt NP, de Bakker PI, et al. Genome-wide association analysis identifies loci for type 2 diabetes and triglyceride levels. Science. 2007;316(5829):1331–6.

2. Beer NL, Tribble ND, McCulloch LJ, Roos C, Johnson PR, Orho-Melander M, et al. The P446 L variant in GCKR associated with fasting plasma glucose and triglyceride levels exerts its effect through increased glucokinase activity in liver. Hum Mol Genet. 2009;18(21):4081–8.

3. Orho-Melander M, Melander O, Guiducci C, Perez-Martinez P, Corella D, Roos C, et al. Common missense variant in the glucokinase regulatory protein gene is associated with increased plasma triglyceride and C-reactive protein but lower fasting glucose concentrations. Diabetes. 2008;57(11):3112–21.

4. Kathiresan S, Melander O, Guiducci C, Surti A, Burtt NP, Rieder MJ, et al. Six new loci associated with blood low-density lipoprotein cholesterol, high-density lipoprotein cholesterol or triglycerides in humans. Nat Genet. 2008;40(2):189–97.

5. Willer CJ, Sanna S, Jackson AU, Scuteri A, Bonnycastle LL, Clarke R, et al. Newly identified loci that influence lipid concentrations and risk of coronary artery disease. Nat Genet. 2008;40(2):161–9.

6. Kooner JS, Chambers JC, Aguilar-Salinas CA, Hinds DA, Hyde CL, Warnes GR, et al. Genome-wide scan identifies variation in MLXIPL associated with plasma triglycerides. Nat Genet. 2008;40(2):149–51.

7. Sandhu MS, Waterworth DM, Debenham SL, Wheeler E, Papadakis K, Zhao JH, et al. LDL-cholesterol concentrations: a genome-wide association study. Lancet. 2008;371(9611):483–91.

8. Wallace C, Newhouse SJ, Braund P, Zhang F, Tobin M, Falchi M, et al. Genome-wide association study identifies genes for biomarkers of cardiovascular disease: serum urate and dyslipidemia. Am J Hum Genet. 2008;82(1):139–49.

9. Sabatti C, Service SK, Hartikainen AL, Pouta A, Ripatti S, Brodsky J, et al. 2009. Genome-wide association analysis of metabolic traits in a birth cohort from a founder population. Nat Genet. 2009;41(1):35–46.

10. Aulchenko YS, Ripatti S, Lindqvist I, Boomsma D, Heid IM, Pramstaller PP, et al. Loci influencing lipid levels and coronary heart disease risk in 16 European population cohorts. Nat Genet. 2009;41(1):47–55.

11. Kathiresan S, Willer CJ, Peloso GM, Demissie S, Musunuru K, Schadt EE, et al. Common variants at 30 loci contribute to polygenic dyslipidemia. Nat Genet. 2009;41(1):56–65.

12. Teslovich TM, Musunuru K, Smith AV, Edmondson AC, Stylianou IM, Koseki M, et al. Biological, clinical, and population relevance of 95 loci for blood lipids. Nature. 2010;466(7307):707–13.

13. Myocardial Infarction Genetics Consortium, Kathiresan S, Voight BF, Purcell S, Musunuru K, Ardissino D, et al. Genome-wide association of early-onset myocardial infarction with single nucleotide polymorphisms and copy number variants. Nat Genet. 2009;41(3):334–41.

14. Musunuru K, Strong A, Frank-Kamenetsky M, Lee NE, Ahfeldt T, Sachs KV, et al. From noncoding variant to phenotype via SORT1 at the 1p13 cholesterol locus. Nature. 2010;466(7307):714–19.

15. Kjolby M, Andersen OM, Breiderhoff T, Fjorback AW, Pedersen KM, Madsen P, et al. Sort1, encoded by the cardiovascular risk locus 1p13.3, is a regulator of hepatic lipoprotein export. Cell Metab. 2010;12(3):213–23.

16. Strong A, Ding Q, Edmondson AC, Millar JS, Sachs KV, Li X, et al. Hepatic sortilin regulates both apolipoprotein B secretion and LDL catabolism. J Clin Invest. 2012;122(8):2807–16

17. Burkhardt R, Toh SA, Lagor WR, Birkeland A, Levin M, Li X, et al. Trib1 is a lipid- and myocardial infarction-associated gene that regulates hepatic lipogenesis and VLDL production in mice. J Clin Invest. 2010;120(12):4410–4.

18. Holleboom AG, Karlsson H, Lin RS, Beres TM, Sierts JA, Herman DS, et al. Heterozygosity for a loss-of-function mutation in GALNT2 improves plasma triglyceride clearance in man. Cell Metab. 2011;14(6):811–8.