



OPEN

Inclusion of endophenotypes in a standard GWAS facilitate a detailed mechanistic understanding of genetic elements that control blood lipid levels

Qianqian Zhang^{1,6}, Zexi Cai^{2,6}, Marie Lhomme³, Goutam Sahana², Philippe Lesnik⁴, Maryse Guerin⁴, Merete Fredholm⁵ & Peter Karlskov-Mortensen⁵✉

Dyslipidemia is the primary cause of cardiovascular disease, which is a serious human health problem in large parts of the world. Therefore, it is important to understand the genetic and molecular mechanisms that regulate blood levels of cholesterol and other lipids. Discovery of genetic elements in the regulatory machinery is often based on genome wide associations studies (GWAS) focused on end-point phenotypes such as total cholesterol level or a disease diagnosis. In the present study, we add endophenotypes, such as serum levels of intermediate metabolites in the cholesterol synthesis pathways, to a GWAS analysis and use the pig as an animal model. We do this to increase statistical power and to facilitate biological interpretation of results. Although the study population was limited to ~300 individuals, we identify two genome-wide significant associations and ten suggestive associations. Furthermore, we identify 28 tentative associations to loci previously associated with blood lipids or dyslipidemia associated diseases. The associations with endophenotypes may inspire future studies that can dissect the biological mechanisms underlying these previously identified associations and add a new level of understanding to previously identified associations.

Blood lipid levels are routinely measured to evaluate risk of cardiovascular disease (CVD), which is a devastating disease in humans. According to WHO, 17.5 million people died from CVD in 2012. This was more than twice the number of deaths due to cancers. Furthermore, six million people were under age 70 as CVD caused a premature end of life¹. This places CVD as the leading cause of death and early demise in the world. The culprit of CVD is dyslipidemia, and this emphasizes why it is of crucial importance to understand the mechanisms that regulate serum levels of blood lipids.

Whole body cholesterol content and serum levels of cholesterol are regulated by a balance in uptake of cholesterol from food, excretion of cholesterol via the bile, de novo biosynthesis of cholesterol in the body as well as microbial metabolism and reduction of cholesterol to coprostanol^{2,3}. Several genes and molecules that facilitate and regulate uptake, excretion and fluxes of cholesterol between compartments and over cell membranes have been described⁴⁻⁸. De novo biosynthesis of cholesterol can take place in any cell in the body, but the great majority of the molecule is synthesized in the liver in a tightly regulated process⁹. The homeostatic mechanisms are however only partly understood.

Two separate pathways for cholesterol biosynthesis have been described; the Bloch pathway (BPW)¹⁰ and the Kandutsch-Russel pathway (KRPW)¹¹. The initial steps in cholesterol synthesis are shared by both pathways.

¹Bioinformatics Research Centre (BiRC), Aarhus University, C.F.Møllers Allé 8, 8000 Aarhus C, Denmark. ²Center for Quantitative Genetics and Genomics, Aarhus University, Blichers Allé 20, 8830 Tjele, Denmark. ³ICANalytics, Institute of Cardiometabolism and Nutrition (ICAN), 47-83 boulevard de l'hôpital, 75013 Paris, France. ⁴Unité de Recherche sur les maladies cardiovasculaires, le métabolisme et la nutrition, INSERM UMR_S 1166, ICAN Institute of Cardiometabolism & Nutrition, Faculté de Médecine Sorbonne Université, Sorbonne Université, 4ème étage, Bureau 421,91, boulevard de l'Hôpital, 75634 Paris Cedex 13, France. ⁵Animal Genetics, Bioinformatics and Breeding, Department of Veterinary and Animal Sciences, University of Copenhagen, Grønnegaardsvej 3, 1870 Frederiksberg C, Denmark. ⁶These authors contributed equally: Qianqian Zhang and Zexi Cai. ✉email: pkm@sund.ku.dk

HMG-CoA reductase (the target of current cholesterol-lowering statin therapy) is central in the early step of the cholesterol synthesis where mevalonate is formed from acetyl-CoA. Mevalonate is hereafter converted to squalene (Sq), which again is turned into lanosterol (Lan), which has the cyclic nature of cholesterol. Lan is the starting point for both the BPW and KRPW. In the BPW, Lan is converted to cholesterol via an intermediate called desmosterol (Des), whereas cholesterol is synthesized from Lan via an intermediate called lathosterol (Lat) in the KRPW.

A fundamental regulatory mechanism for de novo synthesis involves sensing of intracellular cholesterol levels by the SREBP proteins (sterol regulatory element binding protein 1 and 2). If cholesterol is present, SREBP binds to SCAP and INSIG-1. Else, INSIG-1 dissociates from the complex and allows cleavage of SREBP, which is then free to migrate to the nucleus where it acts as a transcription factor for a number of genes including the LDL-receptor gene (*LDLR*) and the HMG-CoA reductase gene (reviewed by REF¹²). Up-regulation of the LDL-receptor increases uptake of low-density lipoprotein cholesterol (LDL-C) from the blood and up-regulation of the HMG-CoA reductase gene increases de novo biosynthesis of cholesterol. The SREBP proteins regulate the initial steps of cholesterol synthesis. Regulatory mechanisms later in the cholesterol biosynthesis are not as well characterized. It has been observed in mice that relative use of the BPW or the KRPW is tissue specific, but the mechanisms, that regulate relative use of the two pathways, are unknown¹³.

The present genome-wide association study (GWAS) is designed to dig deeper into the uncharacterized regulatory mechanisms by the use of intermediate phenotypes—also known as endophenotypes.

Use of endophenotypes in genetic mapping have been advocated since the beginning of the century¹⁴. The rationale is that an intermediate phenotype is more proximal to the direct effect of a single gene. Hence, an endophenotype is expected to be genetically less complex than an end-point phenotype. This has implications both for the power of a GWAS study and for the interpretation of GWAS results into biological function. A GWAS detects the effect of a variation in the genome—most often a single nucleotide polymorphism (SNP). Associating a SNP with an appropriate endophenotype determines the direct effect of the given SNP, whereas associating the SNP with an end-point phenotype determines a derived effect of the SNP. The derived effect may be influenced by other genes and/or environmental factors and the resulting noise cause a reduction in statistical power. The biological complexity of an end-point phenotype furthermore hampers the biological interpretation of an identified association because the associated SNP, or genes near the genetic variant, may affect one of many phenotype-determining factors. For the endophenotype, on the other hand, the associated SNP, or genes nearby, may affect one of few endophenotype-determining factors.

There has been a general realization that endophenotypes are required in order to resolve genetic heterogeneity¹⁵ and to enable a functional characterization of newly discovered genetic variants¹⁶. Nevertheless, efforts to increase statistical power in GWAS has until now mainly been focused on increasing sample sizes¹⁷. This has also been the case for GWAS studies on blood lipid levels and CVD in humans. Here, sample sizes have increased to several hundred thousand individuals^{18–20}. However, the measured phenotypes in these studies are still the crude end-point phenotypes, LDL-C, HDL-C, TG levels, or a CVD diagnosis.

In the present study, we investigate the power of including endophenotypes in the analyses instead of increasing sample size. Furthermore, we use the pig, which is an established model for the highly prevalent human diseases involving lipid metabolism (metabolic syndrome, diabetes, obesity, and cardiovascular diseases)^{21–23}. The pig offers an attractive alternative model to rodents because of their anatomical and physiological similarities to humans and the availability of genomic, transcriptomic, metagenomic, proteomic, and metabolomic tools for analysis of this species²⁴. Regarding blood lipids, pigs transport most of their cholesterol in LDL-C, as do humans, while rodents carry the majority of cholesterol in high-density lipoprotein cholesterol (HDL-C) rather than LDL-C, making them less than desirable as models²⁵. Additionally, high-cholesterol diets can induce human-like changes in the plasma lipoprotein profile of pigs, with ~60% of plasma cholesterol distributed in LDL particles. Currently, there is no single golden standard animal model of atherosclerosis and CVD development, but the pig is probably the best way to recreate human plaque instability that can be linked to hypercholesterolemia, obesity, and diabetes, which contributes to accelerated atherosclerosis^{26,27}.

The goal of the present study is, like in other GWAS studies, to identify loci with an ultimate effect on cholesterol levels, TG levels and CVD. However, unlike previous studies, we include levels of intermediate metabolites in the cholesterol synthesis pathways as endophenotypes in the analyses. This includes serum levels of Lan as well as the BPW and KRPW specific intermediate metabolites, Des and Lat. Furthermore, serum levels of phytosterols and microbiota-derived sterols are included in the analyses. These sterols are absorbed from the gut by the same mechanisms as cholesterol. Hence, serum levels of these lipids provide an estimate for the efficiency of cholesterol absorption from the gut^{28–30}. Our goal is to dissect the cholesterol phenotypes and thereby identify loci controlling specific steps and pathways for biosynthesis, catabolism, and cholesterol transport mechanisms. Hereby we will add to a deeper understanding of biological and genetic mechanisms in cholesterol regulation.

Results

The examined endophenotypes and end-point phenotypes are listed in Table 1 and results of the GWAS analysis are summarized in Table 2. We identified two genome-wide ($p < 4.8E-08$) and 10 suggestive associations ($p < 4.8E-07$). As a general overview, six loci were found to be associated with levels of different precursor molecules in the cholesterol synthesis pathway (Lan, Des, Lat, and Sum of intermediates (Sint)). One of these was genome-wide significant. Two suggestive associations were identified for end-point phenotypes LDL-C and total cholesterol (TC). Three associations (1 genome-wide significant) were identified for levels of phytosterols; campesterol (Cste), betasitosterol (Bsit), and sum of phytosterols (Sphy). One suggestive association were identified for levels of the microbiota-derived sterol, coprostanol (Csta).

Phenotype	Abbreviation	n	Min	Max	Mean	SD
Endophenotypes						
<i>Intermediates in cholesterol biosynthesis</i>						
Squalene	Sq	245	0.70	17.30	4.96	3.69
Lanosterol	Lan	284	1.43	5.89	3.42	0.91
Lathosterol	Lat	299	13.06	96.99	49.85	17.15
Desmosterol	Des	300	22.33	123.63	68.35	20.06
Sum of intermediates	Sint	300	36.15	222.51	129.05	35.00
<i>Phytosterols</i>						
Betasitosterol	Bsit	297	45.36	470.17	186.60	96.30
Campesterol	Cste	288	106.68	862.77	392.42	158.37
Stigmasterol	Stig	249	0.16	2.56	1.14	0.48
Sum of phytosterols	Sphy	290	187.96	1253.79	577.60	232.74
<i>Microbiota-derived sterols</i>						
Coprostanol	Csta	298	1.95	31.12	13.66	6.35
Epicoprostanol	Esta	276	0.23	5.41	2.23	1.11
Sum of microbiota-derived sterols	Sste	303	2.17	35.01	15.84	7.28
End-point phenotypes						
LDL-Cholesterol	LDL-C	296	30.82	86.53	57.70	10.67
HDL-Cholesterol	HDL-C	317	13.00	57.00	36.19	7.92
Total Cholesterol	TC	297	67.78	137.84	100.88	13.88
Triglycerides	TG	295	42.57	200.12	119.68	31.02

Table 1. Phenotypes. All endophenotypes were measured in $\mu\text{g}/\text{dl}$. End-point phenotypes were measured in mg/dl . SD = standard deviation of the mean.

Six of the genome-wide or suggestively associated SNPs were located in introns of protein-coding genes and six were within intergenic regions. Strong functional candidate genes were found for seven of the associated SNPs and one locus appeared to have a pleiotropic effect.

The association between Des and a SNP on chromosome (Chr) 6, 157,486,305 bp was the most significant association identified ($p = 1.06\text{E}-11$) (Fig. 1a). The markers in strong linkage disequilibrium (LD) with this SNP position encompass a region of 4397 bp entailing exon 1 of *DHCR24* (Fig. 2a). This gene encodes the enzyme 24-dehydrocholesterol reductase that converts desmosterol to cholesterol and different mutations in this gene has been shown to cause desmosterolosis (OMIM:602398; Refs.^{32,33}). Our finding shows that the limiting step for Des serum levels in the examined pigs are not synthesis of the molecule but rather conversion of the molecule to cholesterol by *DHCR24*.

The four phenotypes Lat, Des, Sint and TC were found to be suggestively associated with four different SNPs on Chr 13 in strong LD with each other. The Des-associated SNP was located in an intron of the *HPS3* gene but the remaining three SNPs were located in intergenic regions. All of them were located close to the *CP* gene, which encodes ceruloplasmin. Ceruloplasmin concentrations has previously been shown to be strongly correlated with serum triglyceride and cholesterol levels³⁷. Furthermore, ceruloplasmin administration has been shown to produce a partial correction of dyslipidemia, manifested by normalization of levels of TG, TC, LDL-C and HDL-C³⁸. A mechanistic explanation for this may be found in the ability of ceruloplasmin to oxidize LDL particles (reviewed by Ref.³⁹). The oxidized particles are scavenged and degraded by macrophages (reviewed by Ref.⁴⁰) and thereby ceruloplasmin has a LDL-C lowering effect.

LDL-C level was furthermore suggestively associated with a SNP on Chr 2 in an intron of *FCHO2*. The FCH and mu domain containing endocytic adaptor 2 encoded by *FCHO2* plays a direct role in clathrin-mediated endocytosis of LDL-C by organizing clathrin-coated structures for LDLR endocytosis³¹. Hence, this gene plays a role in clearance of LDL-C from the blood stream.

Sint and Lan were found to be suggestively associated with positions in introns of *LINGO1* on Chr 7 and *NYAP2* on Chr 15, respectively. To our knowledge, this is the first time these two genes has been linked to levels of cholesterol precursors in the blood, or blood lipid levels in general.

Three SNPs were associated with levels of different phytosterols. These results establish the first identified loci associated with serum levels of phytosterols. The three SNPs were located on Chr 13 but not close to each other. The Sphy associated SNP at 75.59 Mb on Chr 13 was genome-wide significant ($p = 1.3\text{E}-08$) (Fig. 1b) and the *ANAPC13* gene, encoding anaphase promoting complex subunit 13, was located within a region in strong LD with the lead-SNP (Fig. 2b). This gene has not previously been associated with serum levels of phytosterols or other lipids. A SNP suggestively associated with Cste was located close to the genes *FAIM* and *FOXL2*. The two genes are not obvious candidate genes for phytosterol levels but they are involved in lipid metabolism. *FAIM* deficiency enhances SREBP signaling and promotes lipogenesis in liver³⁴. *FOXL2* represses expression of Star, a protein that controls cholesterol transport from the outer to the inner mitochondrial membranes^{35,36}.

Pheno-type	Chr	Lead-SNP BP	MAF	P	B	SE	Genomic region	Candidate gene	Position of functional candidate gene	Functional link	Region in strong LD with lead-SNP, $r^2 > 0.8$
LDL-C	2	82,111,387	0.10	4.60E-07	-6.78	1.34	Intron, <i>FCHO2</i>	<i>FCHO2</i>	82,117,891-82,248,407	The <i>FCHO2</i> gene product is involved in clathrin-mediated endocytosis of LDL-C ³¹	81,958,868-82,859,296
Csta	3	15,599,531	0.08	7.30E-07	-4.24	0.86	Intron, <i>CALN1</i>	n/a		n/a	15,580,136-15,637,077
Des	6	157,486,305	0.19	1.06E-11	-13.58	2.00	Intron, <i>DHCR24</i>	<i>DHCR24</i>	157,483,794-157,519,059	<i>DHCR24</i> encodes the enzyme that converts desmosterol to cholesterol ^{32,33}	157,481,935-157,486,331
Sint	7	57,828,885	0.22	3.62E-07	18.54	3.64	Intron, <i>LINGO1</i>	n/a		n/a	57,707,845-60,000,893
Bsit	13	53,636,117	0.39	2.57E-07	-41.50	8.05	Intergenic	n/a		n/a	53,062,778-54,254,135
Sphy	13	75,586,650	0.32	1.30E-08	-106.55	18.74	Intergenic	n/a		n/a	75,586,650-75,723,225
Cste	13	79,228,431	0.47	2.58E-07	-65.28	12.67	Intergenic	<i>FAIM</i> <i>FOXL2</i>	79,398,049-79,411,772 79,708,693-79,709,825	<i>FAIM</i> deficiency enhances SREBP signaling and promotes lipogenesis in liver ³⁴ . <i>FOXL2</i> represses expression of <i>Star</i> , which controls cholesterol transport in mitochondria ^{35,36}	79,109,893-79,247,374
Sint	13	88,500,932	0.53	2.37E-07	15.30	2.96	Intergenic	<i>CP</i>	89,398,185-89,463,223	<i>CP</i> encodes ceruloplasmin. Ceruloplasmin concentrations are correlated with serum triglyceride and cholesterol levels ³⁷	88,087,429-89,821,031
Lat	13	88,852,202	0.50	3.38E-07	8.02	1.57	Intergenic	<i>CP</i>	89,398,185-89,463,223	Do	88,084,836-90,249,783
Des	13	89,367,223	0.48	4.67E-07	8.48	1.68	Intron, <i>HPS3</i>	<i>CP</i>	89,398,185-89,463,223	Do	88,084,836-91,010,648
TC	13	89,563,053	0.47	2.84E-07	4.65	0.91	Intergenic	<i>CP</i>	89,398,185-89,463,223	Do	88,084,836-89,993,057
Lan	15	127,265,666	0.20	8.43E-08	0.57	0.11	Intron, <i>NYAP2</i>	n/a	127,032,688-127,295,095	n/a	127,263,233-127,267,733

Table 2. Genome-wide and suggestively significant GWAS results. Summary data of lead-SNPs surpassing the genome-wide ($p < 4.8E-08$) or suggestive ($p < 4.8E-07$) significance level. Results are listed in chromosomal and positional order. Chromosome number (Chr) and base-pair (BP) position refer to pig genome assembly Sscrofa 11.1. MAF = minor allele frequency. $P = p$ value. B = regression coefficient. SE = standard error for B. r^2 = The squared correlation coefficient between pairs of SNPs as a measure for linkage disequilibrium (LD). Phenotype abbreviations: Serum levels of the following lipids, Lanosterol (Lan), Lathosterol (Lat), Desmosterol (Des), Sum of intermediates in the cholesterol synthesis pathway (Sint) i.e. Lan + Lat + Des, Betasitosterol (Bsit), Campesterol (Cste), Stigmasterol (Stig) Sum of phytosterols (Sphy) i.e. Bsit + Cste + Stig, Coprostanol (Csta), low-density lipoprotein cholesterol (LDL-C), total cholesterol (TC).

The microbiota-derived sterol, Csta, was suggestively associated with a SNP on Chr 3. The SNP was located within an intron of the *CALN1* gene. To our knowledge, it is the first time this locus has been associated with lipid absorption or lipid metabolism and the result needs to be confirmed by further studies.

We furthermore identified 52 SNPs with a p -value between $1E-05$ and $4.8E-07$. Associations at this significance level will inevitably contain many false positive results. Table 3 lists 28 of the 52 SNPs. These SNPs are located in loci, which previously have been linked to blood lipid levels and/or dyslipidemia associated comorbidities. We classify these SNPs as tentative associations. Our results suggest that the previously identified effects may be mediated via specific molecular mechanisms identified by the tentative associations with specific endophenotypes in our study. For example, the *ARNT2* (Chr 7) and the *TSPYL5* (Chr 4) genes have previously been associated with blood lipid levels, regulation of lipid metabolism and fatty acid synthesis⁴¹⁻⁴³. In our results, these two loci are tentatively associated with Lan. Hence, our results point more precisely towards a function of

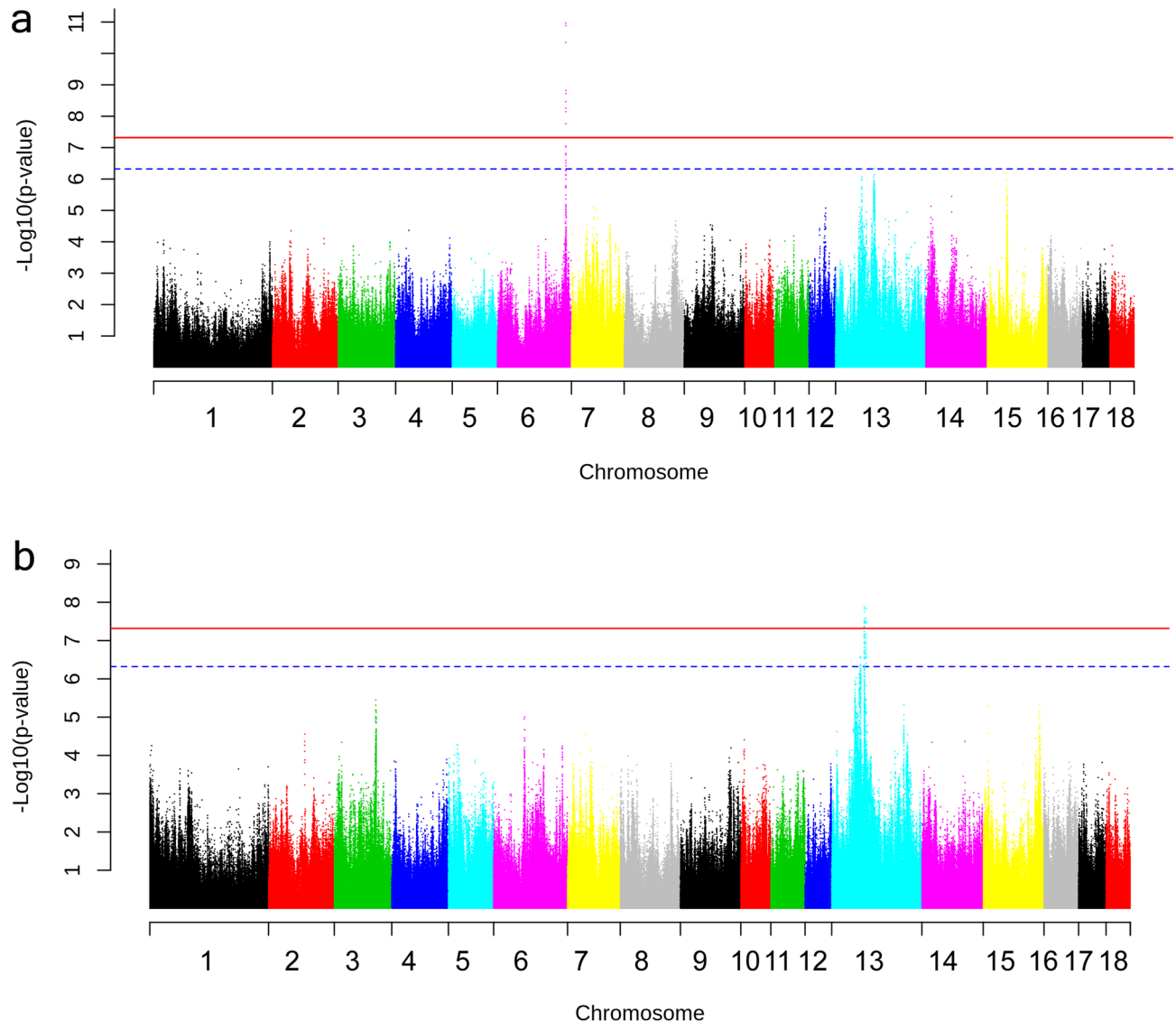
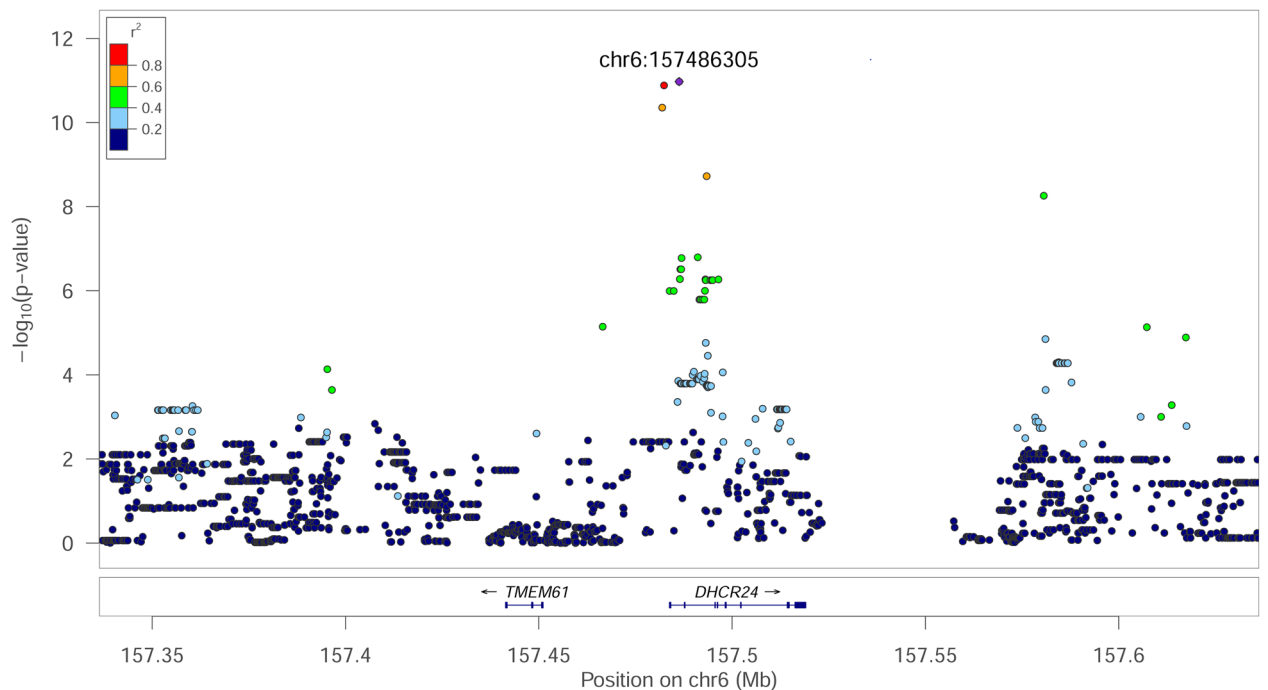


Figure 1. Manhattan plot illustrating results of the genome-wide association study for the phenotypes (a) Desmosterol (Des) and (b) Sum of phytosterols (Sphy). Chromosome numbers are indicated on the X-axis and $-\log(p)$ values on the Y-axis. Horizontal red line: Genome-wide significance threshold ($-\log(4.8E-08)$). Horizontal dashed blue line: Suggestive significance threshold ($-\log(4.8E-07)$).

ARNT2 and *TSPYL5* in the early steps of cholesterol synthesis where Lan is formed from Sq. Similarly, previous studies have linked *PID1* (Chr 15) to HDL-C level⁴⁴ and *PEX2* (Chr 4) to levels of HDL-C and TC⁴⁵, whereas the tentative associations in the present study point more precisely towards a function of these loci in the KRPW for cholesterol biosynthesis because the two loci are associated specifically with Lat levels. On the other hand, the associations between Des levels and the two genes *ARV1* (Chr 14) and *ACSL1* (Chr 15) identify functions of these loci in the BPW of cholesterol biosynthesis. These loci were previously associated with levels of LDL-C, free cholesterol in circulation and accumulations of cholesterol in the liver^{46,47}, whereas our tentative results point more precisely towards an effect on BPW cholesterol synthesis. More examples like these are described in Table 3. Overall, the tentative results in the present study suggest new hypotheses about the fundamental factors and mechanisms causing previously identified end-point phenotype associations. The tentative results of the present study may therefore form a foundation for future studies to clarify these mechanisms.

Correlations between phenotypes are listed in Supplementary Table 1. In general, serum levels of intermediate molecules in the cholesterol synthesis pathway were correlated with each other, phytosterol levels were correlated with each other and levels of microbiota-derived sterols were correlated with each other. Despite the close functional relationship between endophenotypes and end-point phenotypes, correlations between these phenotypes were generally weak and non-significant. However, a moderate correlation was observed between Des and TC. Additionally, we observed a moderate correlation between Lan and the two phytosterols Bsit and Stig. LDL-C was more strongly correlated with TC than HDL-C.

a



b

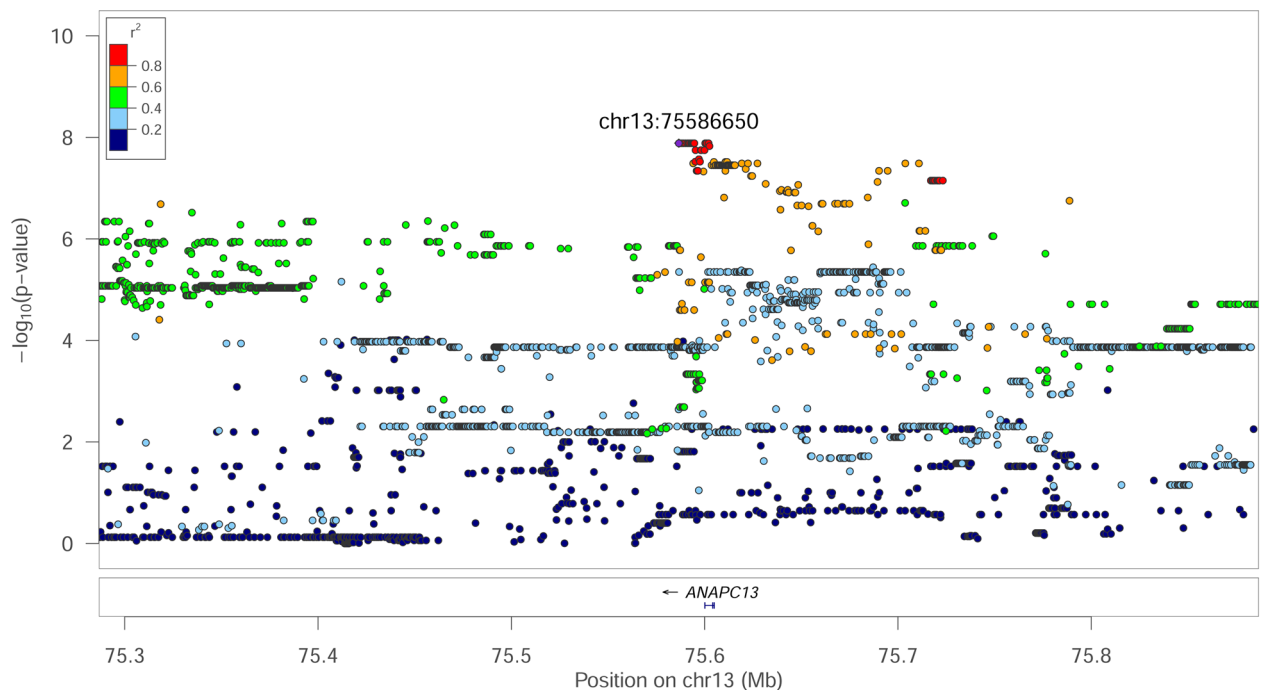


Figure 2. LocusZoom plots of regional genome-wide association results, linkage disequilibrium and gene annotation for the two genome-wide significant associations. (a) Cover the region on chromosome 6 associated with serum levels of desmosterol. (b) cover the region on chromosome 13 associated with sum of phytosterols in serum. Genome positions are indicated on the X-axes. Annotated genes in the regions are indicated above the X-axes; exons and introns are indicated by boxes and lines, respectively. $-\log(p)$ values are indicated on Y-axes. The squared correlation coefficients (r^2) as a measure of linkage disequilibrium between SNPs are indicated by color. SNPs with $r^2 > 0.8$ are considered in strong linkage disequilibrium.

Pheno-type	Chr	Lead-SNP BP	MAF	P	B	SE	Genomic region	Candidate gene	Position of functional candidate gene	Functional link	Region in strong LD with lead-SNP, $r^2 > 0.8$
Sste	1	5,288,352	0.30	7.64E-06	-3.20	0.72	Intron, <i>PACRG</i>	<i>PACRG</i> <i>PRKN</i>	4,967,800-5,465,157 5,698,508-6,731,132	<i>PACRG</i> and <i>PRKN</i> are located head to head and are co-regulated (Entrez gene). <i>PRKN</i> encodes parkin, which is a lipid-responsive regulator of cellular fat uptake ⁴⁸ and has an effect on lipid absorption from the gut ⁴⁹	5,288,352-5,288,352
TC	1	26,458,317	0.12	2.56E-06	5.29	1.12	Intergenic	<i>TNFAIP3</i>	26,473,985-26,489,555	The product of <i>TNFAIP3</i> is a suppressor of the ASK1 activation that plays a key role in development of non-alcoholic steatohepatitis ^{50,51}	26,313,479-26,469,377
LDL-C	1	26,864,925	0.19	6.60E-06	4.76	1.06	Intergenic	<i>TNFAIP3</i>	26,473,985-26,489,555	Do	26,864,925-26,875,955
Bsit	1	97,876,982	0.01	8.22E-06	-161.68	36.26	Intron, <i>ZBTB7C</i>	<i>ZBTB7C</i>	97,610,258-98,018,149	The product of <i>ZBTB7C</i> changes transcription factor binding dynamics of SREBP-1C in mice ⁵²	97,872,191-97,877,224
HDL-C	1	261,632,218	0.10	5.24E-06	-4.37	0.96	Intergenic	<i>DAB2IP</i>	261,659,799-261,859,833	<i>DAB2IP</i> is associated with risk of coronary heart disease in humans ⁵³ that is correlated with HDL-C levels ^{54,55}	261,632,218-261,653,928
TC	2	80,066,762	0.07	2.60E-06	-7.83	1.67	Intron, <i>COL23A1</i>	<i>COL23A1</i>	79,766,160-80,137,598	<i>COL23A1</i> encode a collagen variant found in lipid rafts in cell membranes ⁵⁶ , which are involved in cholesterol transport ⁵⁷	80,066,184-80,075,271
Sint	4	6,495,094	0.08	2.29E-06	26.14	5.53	Intergenic	<i>MIR30D</i>	6,948,669-6,948,747	<i>MIR30D</i> is negatively correlated with HDL-C levels and positively correlated with LDL-C levels ⁵⁸	6,198,095-6,774,519
Lan	4	39,414,823	0.04	7.52E-06	0.85	0.19	Intergenic	<i>TSPYL5</i>	39,362,978-39,367,404	Down-regulation of <i>TSPYL5</i> occur concomitantly with a reduction in cellular cholesterol and fatty acid synthesis and a decrease in total cholesterol and free fatty acid levels ^{42,43}	39,414,823-39,414,823
Lat	4	58,499,547	0.05	5.38E-06	14.81	3.26	Intergenic	<i>PEX2</i>	59,252,313-59,270,099	<i>PEX2</i> controls levels of HDL-C and TC ⁴⁵	58,499,547-58,515,494
TG	4	108,751,861	0.03	1.42E-06	-34.01	7.05	Upstream variant, <i>RAP1A</i>	<i>RAP1A</i>	108,673,364-108,694,720	<i>RAP1A</i> regulates hepatic and plasma PCSK9, plasma TC and LDL-C in mice ⁵⁹	108,751,861-108,751,861
TC	4	119,813,955	0.13	9.39E-06	5.05	1.14	Intergenic	<i>SNX7</i>	119,162,676-119,255,839	<i>SNX7</i> is associated with LDL-C levels ⁶⁰	119,765,369-119,828,033
Csta	5	90,450,047	0.14	5.80E-06	-3.24	0.71	Intergenic	<i>EEA1</i>	90,131,742-90,259,208	<i>EEA1</i> encodes an early endosome antigen, which is a core component of endosome docking ⁶¹ and may play a role in transport of lipids between membrane compartments ⁶²	90,344,197-90,566,508
Sste	5	90,450,047	0.14	9.69E-06	-3.63	0.82	Intergenic	<i>EEA1</i>	90,131,742-90,259,208	Do	90,344,197-90,566,508
Sphy	6	71,581,677	0.10	9.99E-06	-136.25	30.84	Upstream variant, <i>DISP3</i>	<i>DISP3</i> <i>UBIAD1</i> <i>ANGPTL7</i> <i>MTOR</i> <i>MFN2</i>	71,584,051-71,640,673 71,419,574-71,431,237 71,350,921-71,356,683 71,286,991-71,412,884 72,027,996-72,056,434	<i>DISP3</i> encodes a sterol-sensing-domain-containing protein ⁶³ <i>UBIAD1</i> may be involved in cholesterol metabolism (Entrez Gene) <i>ANGPTL7</i> plays a role in lipid trafficking and metabolism ⁶⁴ <i>MTOR</i> encodes a SREBP1 regulator ⁶⁵ <i>MFN2</i> plays a role in regulation of cholesterol synthesis ⁶⁶	71,549,824-71,591,026
TG	7	8,862,047	0.29	1.31E-06	11.88	2.46	Intergenic	<i>EDN1</i>	8,752,082-8,758,348	Endothelin-1 encoded by <i>EDN1</i> inhibits <i>IRS-1</i> expression and IRS-1 activity ⁶⁷ . <i>IRS-1</i> is a determinant for HDL-C and TG levels ⁶⁸	8,862,047-8,863,301

Continued

Pheno-type	Chr	Lead-SNP BP	MAF	P	B	SE	Genomic region	Candidate gene	Position of functional candidate gene	Functional link	Region in strong LD with lead-SNP, $r^2 > 0.8$
Lan	7	49,354,803	0.22	4.27E-06	-0.41	0.09	Intron, <i>ARNT2</i>	<i>ARNT2</i>	49,259,648-49,450,466	<i>ARNT2</i> is associated with blood lipid levels ⁴¹	49,347,252-49,354,803
Cste	7	53,150,849	0.14	6.45E-06	-83.75	18.57	Intron, <i>CRTC3</i>	<i>CRTC3</i>		<i>CRTC3</i> is associated with total cholesterol plasma levels ⁶⁹	53,038,983-53,169,774
LDL-C	8	131,059,165	0.13	6.10E-06	5.88	1.30	Upstream variant, <i>PKD2</i>	<i>SPP1</i>	131,077,825-131,085,327	<i>SPP1</i> regulates CYP7A1, which converts cholesterol to hydroxyl-cholesterol in the first step of bile acid synthesis ⁷⁰	131,057,567-131,064,130
TG	9	121,527,397	0.14	8.42E-06	16.21	3.64	Intron, <i>TOR1AIP1</i>	<i>TOR1AIP1</i>	121,515,664-121,566,795	<i>TOR1AIP1</i> encodes Torsin 1A interacting protein 1, which regulates hepatic VLDL secretion ⁷¹	121,527,397-121,527,397
Sint	12	38,331,125	0.19	3.98E-06	-16.31	3.54	Intergenic	<i>LHX1</i>	38,461,948-38,563,314	<i>LHX1</i> is associated with circulating lipid levels ⁷²	38,331,125-38,331,452
Des	12	38,656,438	0.16	8.52E-06	-10.69	2.40	Intron, <i>ACACA</i>	<i>ACACA</i>	38,581,541-38,875,025	<i>ACACA</i> is involved in fatty acid synthesis ⁷³	38,656,438-38,723,182
TG	12	40,823,266	0.16	5.54E-07	-17.69	3.53	Intergenic	<i>CCL2</i>	40,740,308-40,799,969	The <i>CCL2</i> gene has been associated with TG level, atherosclerosis and myocardial infarction ⁷⁴ and knock-out of <i>CCL2</i> in mice results in lower levels of cholesterol and TG ⁷⁵	40,823,120-40,850,200
Sint	14	12,003,949	0.27	1.02E-06	-18.11	3.71	Intergenic	<i>PBK</i>	11,543,665-11,561,923	<i>PBK</i> interacts with cholesterol and modulate cell-signaling ⁷⁶	12,001,579-12,021,393
Des	14	59,484,064	0.09	3.56E-06	13.54	2.92	Intron, <i>TTC13</i>	<i>ARV1</i>	59,397,433-59,408,878	Decreased expression of <i>ARV1</i> cause hypercholesterolemia ⁴⁷	59,389,882-59,541,999
HDL-C	14	64,268,082	0.03	2.85E-06	7.58	1.62	Intergenic	<i>CDK1</i>	64,233,839-64,249,546	<i>CDK1</i> encodes a kinase, which stabilized SREBP1 ⁷⁷	61,776,948-64,843,717
Des	15	45,876,346	0.04	6.44E-07	-21.42	4.30	Intron, <i>ACSL1</i>	<i>ACSL1</i>	45,867,356-45,933,112	Hepatic <i>ACSL1</i> depletion causes a hypercholesterolemic phenotype in mice ⁴⁶	45,665,239-46,524,879
Bsit	15	93,370,432	0.02	3.23E-06	-116.95	25.12	Intergenic	<i>COL3A1</i>	93,556,914-93,595,678	Expression of <i>COL3A1</i> is affected by plant sterols ⁷⁸	93,370,432-93,370,432
Lat	15	130,395,782	0.12	1.29E-06	9.42	1.95	Intergenic	<i>PID1</i>	130,079,133-130,332,658	<i>PID1</i> controls HDL-C level ⁴⁴	130,384,947-130,395,782

Table 3. Tentative associations in regions previously associated with blood lipid levels or dyslipidemia associated comorbidities. Tentatively significant results of the genome wide association analyses for loci, which has been associated with blood lipid levels and/or dyslipidemia associated comorbidities in previous studies and have a p value $< 10E-5$ in the present study. All associations are listed in chromosomal and positional order. Chromosome number (Chr) and base-pair (BP) position refer to pig genome assembly Sscrofa 11.1. MAF = minor allele frequency. $P = p$ value. B = regression coefficient. SE = standard error for B. r^2 = The squared correlation coefficient between pairs of SNPs as a measure for linkage disequilibrium (LD). Phenotype abbreviations: Serum levels of the following lipids, Lanosterol (Lan), Lathosterol (Lat), Desmosterol (Des), Sum of intermediates in the cholesterol synthesis pathway (Sint) i.e. Lan + Lat + Des, Betasitosterol (Bsit), Campesterol (Cste), Stigmasterol (Stig) Sum of phytosterols (Sphy) i.e. Bsit + Cste + Stig, Coprostanol (Csta), Epicoprostanol (Esta), Sum of microbiota-derived sterols (Sste) i.e. Csta + Esta, low-density lipoprotein cholesterol (LDL-C), high-density lipoprotein cholesterol (HDL-C), total cholesterol (TC), Triglycerides (TG). The tentative associations with endophenotypes in the present study point towards specific biological mechanisms, which may be the underlying causes for the observed associations with end-point phenotypes in previous studies.

Discussion

Larger and larger cohorts has been the mantra for increasing power in GWAS for many years, and for a good reason of course. Results obtained repeatedly in large cohorts are more trustworthy than those, which are only supported by a few more or less coincidental observations in a small sample. However, in the incessant race for larger studies and big consortia, it seems the value of meticulous phenotyping has been comparatively neglected. The power of a study depends on the statistical significance criterion, the sample size and the magnitude of the effect that we want to detect. In a GWAS, the phenotypic effect of any nucleotide variant may be fixed but the accurate ascertainment of the phenotypic effect has consequences for the power of the study. At least two factors play a critical role for accurate measurement of the phenotypic effect. Firstly, any confounding environmental

factor must be minimized. Secondly, the direct effect must be measured rather than a secondary or a derived effect, which may be genetically more complex and hence biased by other loci in the genome. In the present study, we minimized environmental effects by using the pig as an animal model and by raising the animals in a highly controlled environment including a uniform diet for all animals. Furthermore, in order to detect a more direct effect of genetic variants, we measured a number of endophenotypes, for example intermediate metabolites in the cholesterol synthesis pathway. On the other hand, opposing the trend towards larger studies, we reduced the number of animals in the study to a number, which must be considered an absolute minimum for a study of this kind.

Despite the very small sample size, our approach resulted in detection of two genome-wide significant associations with p -values of $1.06E-11$ and $1.30E-08$. While p -values like those are remarkable for a GWAS with this sample size, the many suggestive and tentative associations also show that the number of animals was too restrictive and statistical power suffered from that. Nevertheless, strong functional candidate genes were found for a large fraction of the suggestively and tentatively associated loci. Furthermore, many of those demonstrated how inclusion of endophenotypes could be used to dissect genetic and molecular mechanisms underlying more complex end-point phenotypes. Many of the loci associated with endophenotypes in the present study have previously been associated with the more complex end-point phenotypes HDL-C, LDL-C or a dyslipidemia-associated disease. In the previous studies, detection of the relatively weak effect of these loci on the end-point phenotypes was only possible because large cohorts were studied. The associations to end-point phenotypes were not confirmed for these loci in the present study, because it was underpowered to detect loci with weak effects on complex end-point phenotypes. Instead, by including endophenotypes we found these loci associated with some of the genetically less complex but biologically more fundamental mechanisms, which ultimately cause a change in levels of LDL-C, HDL-C or disease risk as documented by the previous large cohort studies. That is, we detected the same loci in a simpler setup and at a lower cost, and at the same time, we obtain a more detailed understanding of the biological mechanism by which the loci have an effect on the complex end-point phenotypes.

Inclusion of endophenotypes furthermore enabled identification of one pleiotropic locus on Chr 13 associated with Lat, Des, Sint, and TC. The pleiotropic effect indicates either a direct effect at several levels of cholesterol biosynthesis or an early and strong fundamental effect reflected in levels of intermediates in all later stages of cholesterol synthesis. Hence, due to inclusion of endophenotypes, this locus can be identified as a master regulator in cholesterol biosynthesis.

In addition to loci with effect on endophenotypes, the present study also identifies a number of loci with an effect on the end-point phenotypes, HDL-C, LDL-C, TC and TG. In several cases, these results point towards genes with a very central role in cholesterol synthesis or to loci with an effect on cholesterol clearance and/or excretion rather than synthesis. HDL-C level was tentatively associated with the *CDK1* gene, the product of which stabilizes members of the SREBP family⁷⁷. As mentioned above, SREBPs play a key role in the biosynthesis of cholesterol⁷⁹. LDL-C level was suggestively associated with a key mechanism for LDL-C clearance from the blood, namely LDLR endocytosis, via an association with the *FCHO2* gene. TC was tentatively associated with another transmembrane transport mechanism by its association with *COL23A1*. Full-length *COL23A1* molecules are found in lipid rafts⁵⁶, which are tightly packed microdomains of the cell membrane. Evidence suggests that these rafts are directly involved in reverse cholesterol transport⁵⁷. Our results lend support to the cholesterol transporting capacity of *COL23A1*-containing lipid rafts and point to an important role in regulation of overall cholesterol levels. TG level was tentatively associated with *TOR1AIP1*, which plays an essential role in regulation of secretion of hepatic VLDL⁷¹, which is the principal carrier of TG in the blood. All these results are in agreement with the expectation that a study of this scale will only be able to detect the loci with strongest effect for the genetically more complex end-point phenotypes.

The serum levels of phytosterols and microbiota-derived sterols can be used as a proxy for the efficiency of sterol absorption from the gastrointestinal canal. Our results confirm previous observations of a role for the PACRG/PRKN locus in lipid absorption from the gut. Besides that, our results indicate that loci with a more general effect on lipid transport and biosynthesis also have a role to play in sterol absorption.

Overall, the present study corroborate many previous results from studies in human and mice and the great overlap between results affirm the quality of the pig as an excellent animal model for human blood lipid metabolism. Furthermore, we identify new loci associated with different blood lipid levels. These results must be further evaluated in future studies in humans and animal models. Most importantly, the study identifies suggestive and tentative associations between endophenotypes and genes, which previously have been associated with end-point phenotypes. These results suggest hypotheses about more fundamental molecular mechanisms underlying the previously identified associations with end-point phenotypes. We propose that these hypotheses should be evaluated in future studies in humans and in animal models. The study demonstrates how inclusion of endophenotypes has the power to detect biologically important loci even in a small-scale study. This is a cost-effective approach compared to larger GWAS with complex end-point phenotypes. At the same time, the results demonstrate how the inclusion of endophenotypes facilitates elucidation of specific details in biological mechanisms underlying variation in end-point phenotypes.

Material and methods

Animal material and sample collection. All animals used in the present study were a three-way cross between Duroc, Landrace and Yorkshire used in the Danish pig production system. The sows from crossings between Landrace and Yorkshire were inseminated with mixed semen from Duroc boars to produce the pigs used in the study. All parental animals were provided by Danbred (Herlev, Denmark). The pigs were produced in a production farm and raised under the conditions for production pigs in Denmark observing guidelines in the

Danish “Animal Maintenance Act” (Act 432 dated 09/06/2004) and the “Order regarding animal experimentation” (BEK nr 12 af 07/01/2016) and approved by the Danish Veterinary and Food Administration. All pigs were ear-tagged with individual ID at weaning. Both female and male pigs were used in the study. All pigs were fed the same diet. None of the pigs was subjected to any treatment. At an age of approximately 6 months and a body weight of approximately 100 kg, animals were sent to an approved commercial abattoir where they were slaughtered in the morning after overnight fasting. Blood and serum samples were collected immediately after exsanguination in BD K2E (EDTA) tubes and BD Vacutainer SST tubes, respectively, from Thermo Fisher Scientific.

DNA isolation, genotyping and imputation to whole genome sequence variants. High quality DNA was isolated from EDTA stabilized blood using a classic salting out procedure⁸⁰. SNP genotyping was performed by Edinburg Genomics, Ashworth Laboratories (Edinburgh) using the 700 K Affymetrix Axiom PigHD chip. To establish marker positions in the newest assembly of the pig genome (Sscrofa11.1⁸¹), sequences of the probes for Affymetrix Axiom PigHD SNPs were mapped to the new assembly using BWA⁸². Markers with a unique map position were retained for further analyses. Whole genome sequence (WGS) from a reference population of 217 animals from three pig breeds (89 Duroc, 61 Landrace and 67 Yorkshire) was used to impute genotypes at a WGS variant level based on HD chip genotypes. Non-autosomal markers and indels with a position coinciding with HD chip markers were removed. To phase the haplotype for the HD marker set and the WGS marker set, we used Eagle⁸³ with default parameters. The WGS dataset comprised 26,581,741 bi-allelic markers on 18 autosomes. Finally, Minimac3⁸⁴ was used to impute the HD marker set to WGS level. SNPs with a minor allele frequency below 2%, SNPs with large deviation from Hardy–Weinberg proportions ($P < 1.0E - 4$) and SNPs with imputation accuracy below 0.4, reported by Minimac3, were removed. After this, 14,763,710 markers were retained for association analyses.

Phenotypes. Serum levels of all lipids and sterols listed in Table 1 were measured by Gas Chromatography–Mass Spectrometry (GC–MS) using a method adapted from Heuillet, et al.⁸⁵ and Quehenberger, et al.⁸⁶. Briefly, serum were supplemented with deuterium-labelled internal standards and esterified sterols were saponified with 500 μ l 0.5 N KOH for 20 min at 60 °C. Sterols were extracted twice with 450 μ l water and 900 μ l hexane and derivatized with 60 μ l BSTFA:TMCS (90:10), 1 h at 80 °C. Samples were dried and resuspended in 60 μ l cyclohexane containing 1% BSTFA for GC/MS injection. Samples were analyzed using a Trace 1310-ISQ LT GC–MS instrument (Thermo Fisher Scientific). Sterols were injected at 250 °C in split mode and separated on a 30 m \times 0.25 mm, 0.25 μ m DB-5MS column (Agilent). Sterols were ionized using electronic impact (EI) and analyzed in SIM mode.

Outliers for each of the phenotypes were removed. A phenotypic value was considered an outlier if it was either below the first quantile – 1.5 IQR (interquartile range, which is the difference between third and first quantile), or above the third quantile + 1.5 IQR. The following linear mixed model was used to adjust the phenotypes for the fixed effects:

$$y = Zb + Xg + e$$

where y is the vector of phenotypes; b is the vector of fixed effects (i.e. batch, pen, sex, group); g is the vector of random polygenic effect estimated using a genomic relationship matrix constructed using the markers; e is the random residual. It was assumed that g follows a normal distribution $N(0, G\sigma_g^2)$, in which G was the matrix of genomic relationship between individuals estimated using HD marker genotypes following VanRaden⁸⁷, and σ_g^2 was the genetic variance. For random residuals, it was assumed that $e \sim N(0, I\sigma_e^2)$, where σ_e^2 was the residual variance and I was an identity matrix. The corrected phenotypes were calculated as estimated genetic values plus residuals ($y_c = \hat{g} + \hat{e}$). These corrected phenotypes were used as the response variables in the association analysis.

Correlations between corrected phenotypes were calculated using the Hmisc R package (<https://www.R-project.org>). The False discovery rates (FDR) was calculated based on p -values using the $p.adjust$ function of the R package.

Genome wide association analysis. The method for identifying possibly closely linked QTL has been described previously⁸⁸. We estimated the genomic relationship matrix (GRM) using GCTA⁸⁹ with the 700 K HD marker set. We estimated the GRM for each chromosome by leaving this chromosome out. This GRM was used for the following GWAS analysis. SimpleM⁹⁰ was used to estimate the number of independent tests and set a genome wide significance threshold p -value of $4.8e-08$ ($-\log_{10}P = 7.32$). A threshold for suggestive significant was set to 10 times the genome wide significance level ($4.8E-07$). Furthermore, SNPs with $-\log_{10}(p \text{ value}) < 5$ in regions, which previously had been associated with blood lipid levels or dyslipidemia associated comorbidities, are reported.

The GWAS was performed in several rounds. In a first round, single SNP GWAS analysis using GCTA⁸⁹ was performed for each chromosome. Then all SNPs were ranked based on their $-\log_{10}P$ value and the largest $-\log_{10}P$ value within each chromosome was identified. If the $-\log_{10}P$ value of a SNP exceeded five, and there were at least one SNP with $-\log_{10}P > 4$ within a 2 Mb region (1 Mb up and down stream of the lead SNP), this SNP was retained as lead-SNP. Then, for each lead-SNP we extracted the lead SNP’s genotype dosage, fitted it as a covariate, and scanned the whole chromosome again in a second round of the GWAS. If the result of second round detected another SNP that fulfilled lead-SNP criteria, and if this SNP had been a significant ($-\log_{10}P > 5$) non-lead SNP in the first round, then this SNP was added to the lead-SNP list. We then extracted the allele dosage of this SNP,

fixed it as another covariate, and scanned the chromosome in a third GWAS round. This procedure was iterated until no additional SNP remained significant.

Genomic heritability (the proportion of phenotypic variance explained by all HD markers) was estimated for each phenotype using GCTA *--reml* function⁸⁹ and a GRM with all HD markers on autosomes. LD was calculated between each lead-SNP and all other SNPs with $-\log_{10}P > 5$ on the same chromosome as the lead-SNP. Pairwise LD (r^2) was calculated using Plink⁹¹. LocusZoom⁹² was used to illustrate regional GWAS and LD results for genome-wide significant SNPs.

Identification of candidate genes and comparative analysis with previous studies. For each identified association, a 1 Mb region of the porcine genome (Sscrofa11.1) centered on associated SNP were analyzed. Even though strong LD in many cases extended considerably less than 0.5 Mb on each side of the associated SNP, a 1 Mb region were analyzed for all associated positions. All protein coding as well as non-coding genes in the analyzed regions were catalogued. Previous knowledge about each gene was ascertained by searches in the PubMed database, the Entrez-Gene database and the database of www.genecards.org. Searches were performed with search terms combining each gene name with other relevant terms such as the specific endophenotype or end-point phenotype associated with the locus. For each gene, searches were also performed with a combination of gene name and the general search terms “sterols”, “cholesterol” and “blood lipids”.

Received: 26 March 2020; Accepted: 15 October 2020

Published online: 28 October 2020

References

1. WHO Press, W. H. O. World Health Statistics 2016: Monitoring Health for the SDGs, Sustainable Development Goals (Geneva, 2016).
2. Villette, R. *et al.* Unraveling host-gut microbiota dialogue and its impact on cholesterol levels. *Front. Pharmacol.* **11**, 278. <https://doi.org/10.3389/fphar.2020.00278> (2020).
3. Kenny, D. J. *et al.* Cholesterol metabolism by uncultured human gut bacteria influences host cholesterol level. *Cell Host Microbe* **28**, 245–257.e246. <https://doi.org/10.1016/j.chom.2020.05.013> (2020).
4. Sugizaki, T. *et al.* The Niemann-Pick C1 like 1 (NPC1L1) inhibitor ezetimibe improves metabolic disease via decreased liver X receptor (LXR) activity in liver of obese male mice. *Endocrinology* **155**, 2810–2819. <https://doi.org/10.1210/en.2013-2143> (2014).
5. Yu, L. *et al.* Stimulation of cholesterol excretion by the liver X receptor agonist requires ATP-binding cassette transporters G5 and G8. *J. Biol. Chem.* **278**, 15565–15570. <https://doi.org/10.1074/jbc.M301311200> (2003).
6. Altmann, S. W. *et al.* Niemann-Pick C1 Like 1 protein is critical for intestinal cholesterol absorption. *Science* **303**, 1201–1204. <https://doi.org/10.1126/science.1093131> (2004).
7. Yu, L. *et al.* Disruption of *Abcg5* and *Abcg8* in mice reveals their crucial role in biliary cholesterol secretion. *Proc. Natl. Acad. Sci. U.S.A.* **99**, 16237. <https://doi.org/10.1073/pnas.252582399> (2002).
8. Nguyen, T. M., Sawyer, J. K., Kelley, K. L., Davis, M. A. & Rudel, L. L. Cholesterol esterification by ACAT2 is essential for efficient intestinal cholesterol absorption: evidence from thoracic lymph duct cannulation. *J. Lipid Res.* **53**, 95–104. <https://doi.org/10.1194/jlr.M018820> (2012).
9. van der Wulp, M. Y., Verkade, H. J. & Groen, A. K. Regulation of cholesterol homeostasis. *Mol. Cell. Endocrinol.* **368**, 1–16. <https://doi.org/10.1016/j.mce.2012.06.007> (2013).
10. Bloch, K. The biological synthesis of cholesterol. *Science* **150**, 19–28 (1965).
11. Kandutsch, A. A. & Russell, A. E. Preputial gland tumor sterols. 3. A metabolic pathway from lanosterol to cholesterol. *J. Biol. Chem.* **235**, 2256–2261 (1960).
12. Afonso, M. S. *et al.* Molecular pathways underlying cholesterol homeostasis. *Nutrients* **10**, 760. <https://doi.org/10.3390/nu10060760> (2018).
13. Mitsche, M. A., McDonald, J. G., Hobbs, H. H. & Cohen, J. C. Flux analysis of cholesterol biosynthesis in vivo reveals multiple tissue and cell-type specific pathways. *Elife* **4**, e07999. <https://doi.org/10.7554/eLife.07999> (2015).
14. Gottesman, I. I. & Gould, T. D. The endophenotype concept in psychiatry: etymology and strategic intentions. *Am. J. Psychiatry* **160**, 636–645. <https://doi.org/10.1176/appi.ajp.160.4.636> (2003).
15. MacRae, C. A. & Vasan, R. S. Next-generation genome-wide association studies. *Circ. Cardiovasc. Genet.* **4**, 334–336. <https://doi.org/10.1161/CIRCGENETICS.111.960765> (2011).
16. Hall, M.-H. & Smoller, J. W. A new role for endophenotypes in the GWAS era: functional characterization of risk variants. *Harv. Rev. Psychiatry* **18**, 67–74. <https://doi.org/10.3109/10673220903523532> (2010).
17. Kulminski, A. M. *et al.* Explicating heterogeneity of complex traits has strong potential for improving GWAS efficiency. *Sci. Rep.* **6**, 35390. <https://doi.org/10.1038/srep35390> (2016).
18. Klarin, D. *et al.* Genetics of blood lipids among ~300,000 multi-ethnic participants of the million veteran program. *Nat. Genet.* **50**, 1514–1523. <https://doi.org/10.1038/s41588-018-0222-9> (2018).
19. Willer, C. J. *et al.* Discovery and refinement of loci associated with lipid levels. *Nat. Genet.* **45**, 1274–1283. <https://doi.org/10.1038/ng.2797> (2013).
20. Shah, S. *et al.* Genome-wide association and Mendelian randomisation analysis provide insights into the pathogenesis of heart failure. *Nat. Commun.* **11**, 163. <https://doi.org/10.1038/s41467-019-13690-5> (2020).
21. Zhang, X. & Lerman, L. O. Investigating the metabolic syndrome: contributions of swine models. *Toxicol. Pathol.* **44**, 358–366. <https://doi.org/10.1177/0192623316630835> (2016).
22. Kleinert, M. *et al.* Animal models of obesity and diabetes mellitus. *Nat. Rev. Endocrinol.* **14**, 140–162. <https://doi.org/10.1038/nrendo.2017.161> (2018).
23. Poldne, R. & Jurčiková-Novotná, L. Experimental models of hyperlipoproteinemia and atherosclerosis. *Physiol. Res.* **66**, S69–s75. <https://doi.org/10.33549/physiolres.933585> (2017).
24. Bassols, A. *et al.* The pig as an animal model for human pathologies: A proteomics perspective. *Proteomics Clin. Appl.* **8**, 715–731. <https://doi.org/10.1002/prca.201300099> (2014).
25. Kaabia, Z. *et al.* Plasma lipidomic analysis reveals strong similarities between lipid fingerprints in human, hamster and mouse compared to other animal species. *Sci. Rep.* **8**, 15893. <https://doi.org/10.1038/s41598-018-34329-3> (2018).

26. Daugherty, A. *et al.* Recommendation on design, execution, and reporting of animal atherosclerosis studies: a scientific statement from the American Heart Association. *Arterioscler. Thromb. Vasc. Biol.* **37**, e131–e157. <https://doi.org/10.1161/atv.0000000000000062> (2017).
27. Zaragoza, C. *et al.* Animal models of cardiovascular diseases. *J. Biomed. Biotechnol.* **2011**, 497841. <https://doi.org/10.1155/2011/497841> (2011).
28. Kesäniemi, Y. A. & Miettinen, T. A. Cholesterol absorption efficiency regulates plasma cholesterol level in the Finnish population. *Eur. J. Clin. Invest.* **17**, 391–395. <https://doi.org/10.1111/j.1365-2362.1987.tb01132.x> (1987).
29. Miettinen, T. A., Tilvis, R. S. & Kesäniemi, Y. A. Serum plant sterols and cholesterol precursors reflect cholesterol absorption and synthesis in volunteers of a randomly selected male population. *Am. J. Epidemiol.* **131**, 20–31. <https://doi.org/10.1093/oxfordjournals.aje.a115479> (1990).
30. Kuksis, A. Plasma non-cholesterol sterols. *J. Chromatogr. A* **935**, 203–236. [https://doi.org/10.1016/s0021-9673\(01\)01226-2](https://doi.org/10.1016/s0021-9673(01)01226-2) (2001).
31. Mulkearns, E. E. & Cooper, J. A. FCH domain only-2 organizes clathrin-coated structures and interacts with Disabled-2 for low-density lipoprotein receptor endocytosis. *Mol. Biol. Cell* **23**, 1330–1342. <https://doi.org/10.1091/mbc.E11-09-0812> (2012).
32. Zolotushko, J. *et al.* The desmosterolosis phenotype: spasticity, microcephaly and micrognathia with agenesis of corpus callosum and loss of white matter. *Eur. J. Hum. Genet.* **19**, 942–946 (2011).
33. Waterham, H. R. *et al.* Mutations in the 3 β -hydroxysterol Δ 24-reductase gene cause desmosterolosis, an autosomal recessive disorder of cholesterol biosynthesis. *Am. J. Hum. Genet.* **69**, 685–694 (2001).
34. Huo, J. *et al.* Loss of Fas apoptosis inhibitory molecule leads to spontaneous obesity and hepatosteatosis. *Cell Death Dis.* **7**, e2091–e2091. <https://doi.org/10.1038/cddis.2016.12> (2016).
35. Batista, F., Vaiman, D., Dausset, J., Fellous, M. & Veitia, R. A. Potential targets of FOXL2, a transcription factor involved in craniofacial and follicular development, identified by transcriptomics. *Proc. Natl. Acad. Sci. U.S.A.* **104**, 3330–3335. <https://doi.org/10.1073/pnas.0611326104> (2007).
36. Pisarska, M. D., Bae, J., Klein, C. & Hsueh, A. J. W. Forkhead L2 Is expressed in the ovary and represses the promoter activity of the steroidogenic acute regulatory gene. *Endocrinology* **145**, 3424–3433. <https://doi.org/10.1210/en.2003-1141> (2004).
37. Cignarelli, M. *et al.* Relationship of obesity and body fat distribution with ceruloplasmin serum levels. *Int. J. Obes. Relat. Metab. Disord.* **20**, 809–813 (1996).
38. Ermolaeva, E. N. & Krivokhizhina, L. V. Ceruloplasmin correction of dyslipidemia caused by chronic physical load in experiment at submaximal power. *Eksperimental'naiia i klinicheskaia farmakologiya* **79**, 9–11 (2016).
39. Fox, P. L., Mazumder, B., Ehrenwald, E. & Mukhopadhyay, C. K. Ceruloplasmin and cardiovascular disease. *Free Radic. Biol. Med.* **28**, 1735–1744. [https://doi.org/10.1016/s0891-5849\(00\)00231-8](https://doi.org/10.1016/s0891-5849(00)00231-8) (2000).
40. Parthasarathy, S., Raghavamenon, A., Garelnabi, M. O. & Santanam, N. Oxidized low-density lipoprotein. *Methods Mol. Biol. (Clifton, NJ)* **610**, 403–417. https://doi.org/10.1007/978-1-60327-029-8_24 (2010).
41. Dong, C. *et al.* Genetic loci for blood lipid levels identified by linkage and association analyses in Caribbean Hispanics. *J. Lipid Res.* **52**, 1411–1419. <https://doi.org/10.1194/jlr.P013672> (2011).
42. Lu, M. *et al.* ACOT12-dependent alteration of Acetyl-CoA drives hepatocellular carcinoma metastasis by epigenetic induction of epithelial-mesenchymal transition. *Cell Metab.* **29**, 886–900.e885. <https://doi.org/10.1016/j.cmet.2018.12.019> (2019).
43. Pampalakis, G. *et al.* The KLK5 protease suppresses breast cancer by repressing the mevalonate pathway. *Oncotarget* **5**, 2390–2403. <https://doi.org/10.18632/oncotarget.1235> (2014).
44. Mu, Q. *et al.* Overexpression of PID1 reduces high density lipoprotein level and functionality in swine. *IUBMB Life* <https://doi.org/10.1002/iub.2135> (2019).
45. Kovacs, W. J. *et al.* Disturbed cholesterol homeostasis in a peroxisome-deficient PEX2 knockout mouse model. *Mol. Cell. Biol.* **24**, 1–13. <https://doi.org/10.1128/mcb.24.1.1-13.2004> (2004).
46. Singh, A. B., Dong, B., Xu, Y., Zhang, Y. & Liu, J. Identification of a novel function of hepatic long-chain acyl-CoA synthetase-1 (ACSL1) in bile acid synthesis and its regulation by bile acid-activated farnesoid X receptor. *Biochim. Biophys. Acta Mol. Cell Biol. Lipids* **358–371**, 2019. <https://doi.org/10.1016/j.bbalip.2018.12.012> (1864).
47. Tong, F. *et al.* Decreased expression of ARV1 results in cholesterol retention in the endoplasmic reticulum and abnormal bile acid metabolism. *J. Biol. Chem.* **285**, 33632–33641. <https://doi.org/10.1074/jbc.M110.165761> (2010).
48. Kim, K. Y. *et al.* Parkin is a lipid-responsive regulator of fat uptake in mice and mutant human cells. *J. Clin. Invest.* **121**, 3701–3712. <https://doi.org/10.1172/jci44736> (2011).
49. Costa, D. K. *et al.* Reduced intestinal lipid absorption and body weight-independent improvements in insulin sensitivity in high-fat diet-fed Park2 knockout mice. *Am. J. Physiol. Endocrinol. Metab.* **311**, E105–E116. <https://doi.org/10.1152/ajpendo.00042.2016> (2016).
50. Zhang, P. *et al.* The deubiquitinating enzyme TNFAIP3 mediates inactivation of hepatic ASK1 and ameliorates nonalcoholic steatohepatitis. *Nat. Med.* **24**, 84. <https://doi.org/10.1038/nm.4453> (2017).
51. Huang, Z. *et al.* Dual specificity phosphatase 12 regulates hepatic lipid metabolism through inhibition of the lipogenesis and apoptosis signal-regulating kinase 1 pathways. *Hepatology* **70**, 1099–1118. <https://doi.org/10.1002/hep.30597> (2019).
52. Jeon, B. N. *et al.* Kr-pok increases FASN expression by modulating the DNA binding of SREBP-1c and Sp1 at the proximal promoter. *J. Lipid Res.* **53**, 755–766. <https://doi.org/10.1194/jlr.M022178> (2012).
53. Harrison, S. C. *et al.* Association of a sequence variant in DAB2IP with coronary heart disease. *Eur. Heart J.* **33**, 881–888. <https://doi.org/10.1093/eurheartj/ehr075> (2011).
54. Gordon, D. J. *et al.* High-density lipoprotein cholesterol and cardiovascular disease Four prospective American studies. *Circulation* **79**, 8–15. <https://doi.org/10.1161/01.cir.79.1.8> (1989).
55. van der Steeg, W. A. *et al.* High-density lipoprotein cholesterol, high-density lipoprotein particle size, and apolipoprotein A-I: significance for cardiovascular risk: the IDEAL and EPIC-Norfolk studies. *J. Am. Coll. Cardiol.* **51**, 634–642. <https://doi.org/10.1016/j.jacc.2007.09.060> (2008).
56. Veit, G. *et al.* Shedding of collagen XXIII is mediated by furin and depends on the plasma membrane microenvironment. *J. Biol. Chem.* **282**, 27424–27435. <https://doi.org/10.1074/jbc.M703425200> (2007).
57. Storey, S. M. *et al.* Selective cholesterol dynamics between lipoproteins and caveolae/lipid rafts. *Biochemistry* **46**, 13891–13906. <https://doi.org/10.1021/bi700690s> (2007).
58. Han, Y.-L., Cao, X.-E., Wang, J.-X., Dong, C.-L. & Chen, H.-T. Correlations of microRNA-124a and microRNA-30d with clinicopathological features of breast cancer patients with type 2 diabetes mellitus. *SpringerPlus* **5**, 2107. <https://doi.org/10.1186/s40064-016-3786-9> (2016).
59. Spolitu, S. *et al.* Hepatic glucagon signaling regulates PCSK9 and low-density lipoprotein cholesterol. *Circ. Res.* **124**, 38–51. <https://doi.org/10.1161/circresaha.118.313648> (2019).
60. Rotroff, D. M. *et al.* Genetic variants in HSD17B3, SMAD3, and IPO11 impact circulating lipids in response to fenofibrate in individuals with type 2 diabetes. *Clin. Pharmacol. Ther.* **103**, 712–721. <https://doi.org/10.1002/cpt.798> (2018).
61. Christoforidis, S., McBride, H. M., Burgoyne, R. D. & Zerial, M. The Rab5 effector EEA1 is a core component of endosome docking. *Nature* **397**, 621–625. <https://doi.org/10.1038/17618> (1999).
62. Liu, P. *et al.* Rab-regulated interaction of early endosomes with lipid droplets. *Biochim. Biophys. Acta* **1773**, 784–793. <https://doi.org/10.1016/j.bbamcr.2007.02.004> (2007).

63. Zikova, M., Corlett, A., Bendova, Z., Pajer, P. & Bartunek, P. DISP3, a sterol-sensing domain-containing protein that links thyroid hormone action and cholesterol metabolism. *Mol. Endocrinol. (Baltim., MD)* **23**, 520–528. <https://doi.org/10.1210/me.2008-0271> (2009).
64. Quagliarini, F. *et al.* Atypical angiopoietin-like protein that regulates ANGPTL3. *Proc. Natl. Acad. Sci. U.S.A.* **109**, 19751–19756. <https://doi.org/10.1073/pnas.1217552109> (2012).
65. Li, S., Oh, Y. T., Yue, P., Khuri, F. R. & Sun, S. Y. Inhibition of mTOR complex 2 induces GSK3/FBXW7-dependent degradation of sterol regulatory element-binding protein 1 (SREBP1) and suppresses lipogenesis in cancer cells. *Oncogene* **35**, 642–650. <https://doi.org/10.1038/onc.2015.123> (2016).
66. Chung, K.-P. *et al.* Mitofusins regulate lipid metabolism to mediate the development of lung fibrosis. *Nat. Commun.* **10**, 3390–3390. <https://doi.org/10.1038/s41467-019-11327-1> (2019).
67. Chai, S.-P. & Fong, J. C. Synergistic induction of insulin resistance by endothelin-1 and cAMP in 3T3-L1 adipocytes. *Biochim. Biophys. Acta (BBA) Mol. Basis Dis.* **1852**, 2048–2055. <https://doi.org/10.1016/j.bbadis.2015.06.026> (2015).
68. Sharma, R. *et al.* The type 2 diabetes and insulin-resistance locus near IRS1 is a determinant of HDL cholesterol and triglycerides levels among diabetic subjects. *Atherosclerosis* **216**, 157–160. <https://doi.org/10.1016/j.atherosclerosis.2011.01.022> (2011).
69. Ou, Z. *et al.* CRTC3 polymorphisms were associated with the plasma level of total cholesterol and the risks of overweight and hypertriglyceridemia in a Chinese Han population. *Mol. Biol. Rep.* **41**, 125–130. <https://doi.org/10.1007/s11033-013-2844-4> (2014).
70. Nunez-Garcia, M. *et al.* Osteopontin regulates the cross-talk between phosphatidylcholine and cholesterol metabolism in mouse liver. *J. Lipid Res.* **58**, 1903–1915. <https://doi.org/10.1194/jlr.M078980> (2017).
71. Shin, J.-Y. *et al.* Nuclear envelope-localized torsinA-LAP1 complex regulates hepatic VLDL secretion and steatosis. *J. Clin. Investig.* **129**, 4885–4900. <https://doi.org/10.1172/JCI129769> (2019).
72. Kumpulainen, T. O. *et al.* Multi-ancestry study of blood lipid levels identifies four loci interacting with physical activity. *Nat. Commun.* **10**, 376. <https://doi.org/10.1038/s41467-018-08008-w> (2019).
73. Tong, L. Acetyl-coenzyme A carboxylase: crucial metabolic enzyme and attractive target for drug discovery. *Cell. Mol. Life Sci. (CMLS)* **62**, 1784–1803. <https://doi.org/10.1007/s00018-005-5121-4> (2005).
74. McDermott, D. H. *et al.* CCL2 polymorphisms are associated with serum monocyte chemoattractant protein-1 levels and myocardial infarction in the Framingham heart study. *Circulation* **112**, 1113–1120. <https://doi.org/10.1161/circulationaha.105.543579> (2005).
75. Rull, A., Camps, J., Alonso-Villaverde, C. & Joven, J. Insulin resistance, inflammation, and obesity: role of monocyte chemoattractant protein-1 (orCCL2) in the regulation of metabolism. *Mediators Inflamm.* **2010**, 11. <https://doi.org/10.1155/2010/326580> (2010).
76. Sheng, R. *et al.* Cholesterol modulates cell signaling and protein networking by specifically interacting with PDZ domain-containing scaffold proteins. *Nat. Commun.* **3**, 1249–1249. <https://doi.org/10.1038/ncomms2221> (2012).
77. Bengoechea-Alonso, M. T. & Ericsson, J. Cdk1/cyclin B-mediated phosphorylation stabilizes SREBP1 during mitosis. *Cell Cycle* **5**, 1708–1718. <https://doi.org/10.4161/cc.5.15.3131> (2006).
78. Tanaka, M., Misawa, E., Yamauchi, K., Abe, F. & Ishizaki, C. Effects of plant sterols derived from Aloe vera gel on human dermal fibroblasts in vitro and on skin condition in Japanese women. *Clin. Cosmet. Investig. Dermatol.* **8**, 95–104. <https://doi.org/10.2147/CCID.S75441> (2015).
79. Espenshade, P. J. & Hughes, A. L. Regulation of sterol synthesis in eukaryotes. *Annu. Rev. Genet.* **41**, 401–427. <https://doi.org/10.1146/annurev.genet.41.110306.130315> (2007).
80. Miller, S. A., Dykes, D. D. & Polesky, H. F. A simple salting out procedure for extracting DNA from human nucleated cells. *Nucleic Acids Res.* **16**, 1215 (1988).
81. Groenen, M. A. M. *et al.* Analyses of pig genomes provide insight into porcine demography and evolution. *Nature* **491**, 393–398. <https://doi.org/10.1038/nature11622> (2012).
82. Li, H. Aligning sequence reads, clone sequences and assembly contigs with **BWA-MEM. *arXiv:1303.3997* (2013).
83. Loh, P.-R., Palamara, P. F. & Price, A. L. Fast and accurate long-range phasing in a UK Biobank cohort. *Nat. Genet.* **48**, 811–816. <https://doi.org/10.1038/ng.3571> (2016).
84. Howie, B., Fuchsberger, C., Stephens, M., Marchini, J. & Abecasis, G. R. Fast and accurate genotype imputation in genome-wide association studies through pre-phasing. *Nat. Genet.* **44**, 955–959. <https://doi.org/10.1038/ng.2354> (2012).
85. Heuillet, M. *et al.* Validation of a reference method for total cholesterol measurement in human serum and assignment of reference values to proficiency testing samples. *Clin. Biochem.* **46**, 359–364. <https://doi.org/10.1016/j.clinbiochem.2012.11.026> (2013).
86. Quehenberger, O. *et al.* Lipidomics reveals a remarkable diversity of lipids in human plasma. *J. Lipid Res.* **51**, 3299–3305. <https://doi.org/10.1194/jlr.M009449> (2010).
87. VanRaden, P. M. Efficient methods to compute genomic predictions. *J. Dairy Sci.* **91**, 4414–4423. <https://doi.org/10.3168/jds.2007-0980> (2008).
88. Cai, Z., Gulbrandsen, B., Lund, M. S. & Sahana, G. Dissecting closely linked association signals in combination with the mammalian phenotype database can identify candidate genes in dairy cattle. *BMC Genet.* **20**, 15. <https://doi.org/10.1186/s12863-019-0717-0> (2019).
89. Yang, J., Lee, S. H., Goddard, M. E. & Visscher, P. M. GCTA: a tool for genome-wide complex trait analysis. *Am. J. Hum. Genet.* **88**, 76–82. <https://doi.org/10.1016/j.ajhg.2010.11.011> (2011).
90. Gao, X., Starmer, J. & Martin, E. R. A multiple testing correction method for genetic association studies using correlated single nucleotide polymorphisms. *Genet. Epidemiol.* **32**, 361–369. <https://doi.org/10.1002/gepi.20310> (2008).
91. Purcell, S. *et al.* PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81**, 559–575. <https://doi.org/10.1086/519795> (2007).
92. Pruim, R. J. *et al.* LocusZoom: regional visualization of genome-wide association scan results. *Bioinformatics (Oxf., Engl.)* **26**, 2336–2337. <https://doi.org/10.1093/bioinformatics/btq419> (2010).

Acknowledgements

Anni Øyan Pedersen is acknowledged for all her work related to raising the pigs used in the present study. Additionally, we are thankful to Tina N. Mahler, Minna Jakobsen and Charlotte B. Larsen for excellent work in the laboratories. The work was funded by Innovation Fund Denmark, Grant ID: 4105-00011B.

Author contributions

Q.Z.: Statistical analyses, GWAS. Z.C.: Genotype imputation, GWAS. M.L.: Blood lipid measurements. G.S.: Main responsible for GWAS. P.L.: Blood lipid measurements. M.G.: Main responsible for blood lipid measurements. M.F.: Study design, sample collection. P.K.M.: Study design, sample collection, candidate gene analyses and comparative analyses, drafted the manuscript. Q.Z. and Z.C. contributed equally. All authors have reviewed and approved the final manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at <https://doi.org/10.1038/s41598-020-75612-6>.

Correspondence and requests for materials should be addressed to P.K.-M.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2020