ARTICLES

From noncoding variant to phenotype via *SORT1* at the 1p13 cholesterol locus

Kiran Musunuru[1,2,3]*, Alanna Strong[4]*, Maria Frank-Kamenetsky[5], Noemi E. Lee[1], Tim Ahfeldt[1,6], Katherine V. Sachs[4], Xiaoyu Li[4], Hui Li[4], Nicolas Kuperwasser[1], Vera M. Ruda[1], James P. Pirruccello[1,2], Brian Muchmore[7], Ludmila Prokunina-Olsson[7], Jennifer L. Hall[2,8], Eric E. Schadt[9], Carlos R. Morales[10], Sissel Lund-Katz[11], Michael C. Phillips[11], Jamie Wong[5], William Cantley[5], Timothy Racie[5], Kenechi G. Ejebe[1,2], Marju Orho-Melander[12], Olle Melander[12], Victor Koteliansky[5], Kevin Fitzgerald[5], Ronald M. Krauss[13], Chad A. Cowan[1,2], Sekar Kathiresan[1,2]* & Daniel J. Rader[4]*

Recent genome-wide association studies (GWASs) have identified a locus on chromosome 1p13 strongly associated with both plasma low-density lipoprotein cholesterol (LDL-C) and myocardial infarction (MI) in humans. Here we show through a series of studies in human cohorts and human-derived hepatocytes that a common noncoding polymorphism at the 1p13 locus, rs12740374, creates a C/EBP (CCAAT/enhancer binding protein) transcription factor binding site and alters the hepatic expression of the *SORT1* gene. With small interfering RNA (siRNA) knockdown and viral overexpression in mouse liver, we demonstrate that *Sort1* alters plasma LDL-C and very low-density lipoprotein (VLDL) particle levels by modulating hepatic VLDL secretion. Thus, we provide functional evidence for a novel regulatory pathway for lipoprotein metabolism and suggest that modulation of this pathway may alter risk for MI in humans. We also demonstrate that common noncoding DNA variants identified by GWASs can directly contribute to clinical phenotypes.

MI is the leading cause of death in the developed world. LDL-C is a causal risk factor for the disease, as demonstrated by the increased and early burden of MI in individuals with the Mendelian disorder of familial hypercholesterolemia[1] and the success of LDL-C-lowering medications in reducing the incidence of MI in clinical trials in many populations[2]. Despite aggressive use of statin drugs, many individuals do not achieve the LDL-C levels recommended by clinical guidelines[3]. There remains a need for additional methods of reducing LDL-C.

GWASs for plasma lipoprotein traits have identified a number of common single nucleotide polymorphism (SNP) variants that are strongly associated with plasma LDL-C[4–10]. Many of these SNPs are in or near genes known to cause Mendelian dyslipidaemias (*LDLR*, *APOB* and *PCSK9*) or established molecular targets for LDL-C-lowering therapies (*HMGCR*). However, several of the LDL-C loci contain genes not previously implicated in lipoprotein metabolism. Of the newly mapped loci, the novel SNPs most strongly associated with LDL-C all lie on chromosome 1p13; indeed, in a meta-analysis of ~100,000 individuals (reported in the accompanying paper[10]) this locus has the strongest association with LDL-C of any locus in the genome ($P = 1 \times 10^{-170}$). The same 1p13 SNPs have also been independently linked to coronary artery disease and MI in GWASs[10–12]. Individuals of European descent who are homozygous for the major alleles of these SNPs have up to 16 mg dl^{-1} higher LDL-C as well as ~40% increased risk of MI[11,12] when compared with minor allele homozygotes. Thus, the same genetic locus is linked to both a recognized intermediate phenotype as well as a hard clinical disease endpoint.

As compelling as these associations are, they do not explain how human genetic variation at the 1p13 locus confers change in plasma LDL-C and thereby alters risk of MI. We therefore sought to identify (1) the causal DNA variant in the 1p13 locus, (2) the gene regulated by the locus, (3) the mechanism by which the DNA variant affects the gene, and (4) the mechanism by which the gene influences lipoprotein metabolism.

1p13 SNPs associated with LDL particles

LDL-C comprises a variety of lipoprotein particles that range in size and density, and it has been hypothesized that smaller LDL particles are more atherogenic than larger LDL particles[13]. To determine whether the 1p13 locus selectively affects certain LDL subclasses, we used different methodologies—ion mobility and gradient gel electrophoresis—to measure lipoprotein subclasses in two different cohorts—the Malmö Diet and Cancer Study – Cardiovascular Cohort (MDC-CC)[14] and the Pharmacogenomics and Risk of Cardiovascular Disease (PARC) study[15]. We found that an index SNP in the 1p13 locus, rs646776, was most highly associated with changes in the very small LDL (LDL-VS) lipoprotein subclass (20% increase in major allele homozygotes versus minor allele homozygotes with $P = 1.1 \times 10^{-11}$ in MDC-CC; 37% increase with $P = 8.0 \times 10^{-11}$ in PARC); progressively smaller changes were seen with larger LDL subclasses (Fig. 1a; Supplementary Fig. 1a, b).

[1]Cardiovascular Research Center and Center for Human Genetic Research, Massachusetts General Hospital and Harvard Medical School, Boston, Massachusetts 02114, USA. [2]Broad Institute, Cambridge, Massachusetts 02142, USA. [3]Division of Cardiology, Johns Hopkins University School of Medicine, Baltimore, Maryland 21287, USA. [4]Institute for Translational Medicine and Therapeutics, Institute for Diabetes, Obesity and Metabolism, and Cardiovascular Institute, University of Pennsylvania School of Medicine, Philadelphia, Pennsylvania 19104, USA. [5]Alnylam Pharmaceuticals, Inc., Cambridge, Massachusetts 02142, USA. [6]Department of Biochemistry and Molecular Biology II: Molecular Cell Biology, University Medical Center Hamburg-Eppendorf, Hamburg 20246, Germany. [7]Laboratory of Translational Genomics, Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health, Bethesda, Maryland 20892, USA. [8]Program in Cardiovascular Translational Genomics, Lillehei Heart Institute, University of Minnesota, Minneapolis, Minnesota 55455, USA. [9]Sage Bionetworks, Seattle, Washington 98109, USA. [10]Department of Anatomy and Cell Biology, McGill University, Montreal, Quebec H3A 2B2, Canada. [11]The Children's Hospital of Philadelphia, University of Pennsylvania School of Medicine, Philadelphia, Pennsylvania 19104, USA. [12]Department of Clinical Sciences, Skania University Hospital, Lund University, SE-20502 Malmö, Sweden. [13]Children's Hospital Oakland Research Institute, Oakland, California 94609, USA.
*These authors contributed equally to this work.

1p13 SNPs and liver-specific expression

The SNPs in the 1p13 locus reported previously to be most highly associated with LDL-C—rs646776, rs599839, rs12740374, and rs629301—lie in a noncoding DNA region between two genes, *CELSR2* and *PSRC1*, whose functions are unknown (Figs 1b and 2a)[4–10]. As noncoding DNA variants may alter gene expression, we previously used expression quantitative trait locus (eQTL) analyses to explore whether 1p13 SNPs are *cis*-acting regulators of nearby genes in human liver[4,7]. We have now extended these studies by measuring expression of genes in or near the 1p13 locus in three types of human tissue samples: liver (960 samples), subcutaneous fat (433 samples) and omental fat (520 samples).

In liver, presence of the minor allele of rs646776 was highly associated with elevated transcript levels of three genes: *CELSR2*, *PSRC1* and *SORT1* (Fig. 1c). *SORT1* displayed the largest expression change. We replicated these liver eQTL findings in an independent cohort of 62 human liver samples, from which rs12740374 (the putative causal 1p13 SNP, see below) was directly genotyped and *SORT1*, *PSRC1* and *CELSR2* expression were individually measured. Minor allele homozygotes displayed more than 12-fold higher *SORT1* and *PSRC1* expression than major allele homozygotes, with no significant change for *CELSR2* (Fig. 1d). Immunoblot analysis of liver lysates demonstrated a significant increase in abundance of the *SORT1* protein product (sortilin) in heterozygotes compared to major allele homozygotes (Fig. 1d and Supplementary Fig. 1c).

Notably, none of the gene expression changes in liver were seen in the two adipose tissue types (Fig. 1c), and minimal changes were reported in lymphocytes[16], suggesting that the regulatory mechanism underlying the allele-specific gene expression is liver-specific.

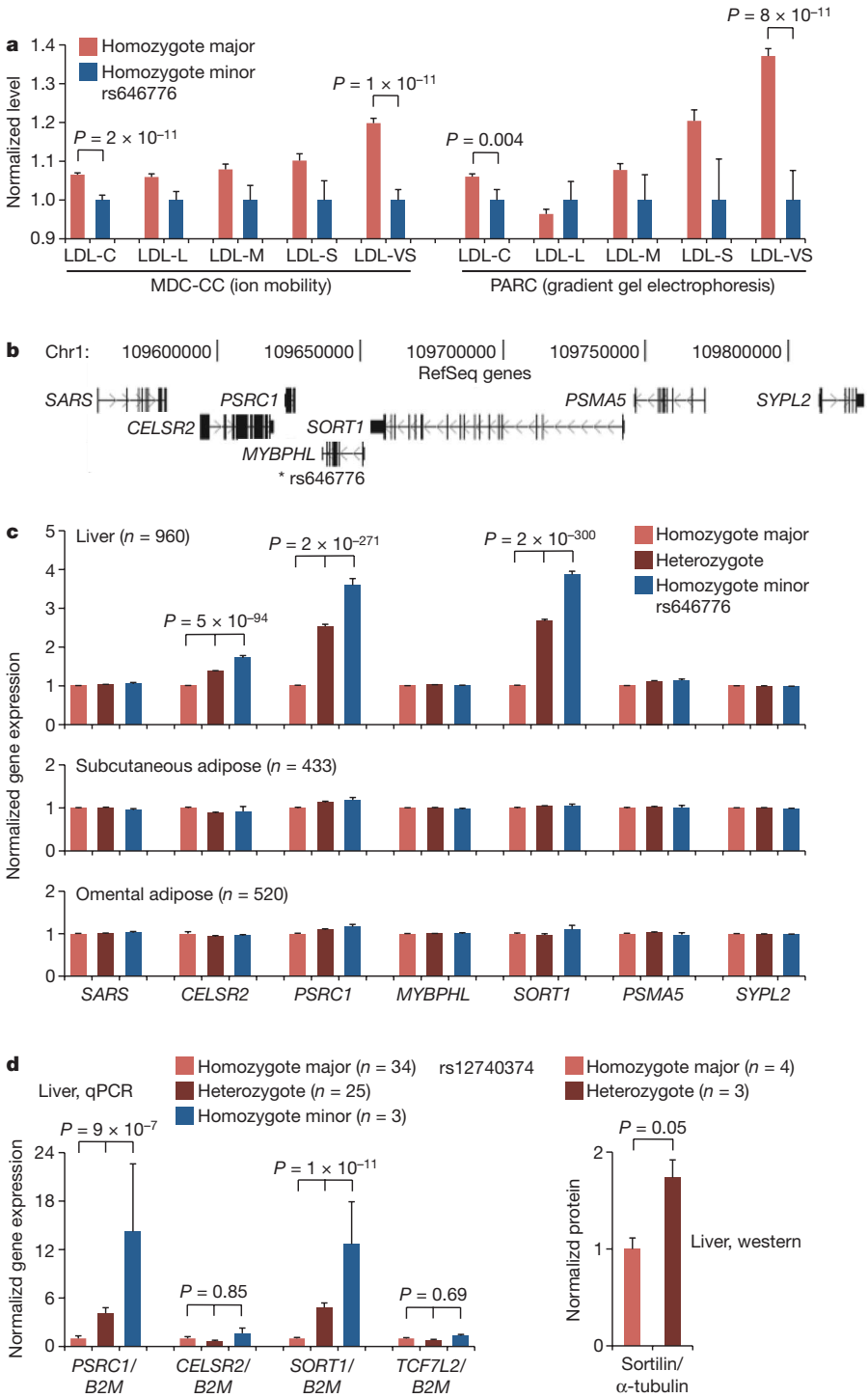

Figure 1 | Human chromosome 1p13 locus is preferentially associated with very small LDL and liver gene expression. **a**, Mean plasma lipid and lipoprotein particle levels in homozygotes for the minor haplotype of the 1p13 locus (minor allele of rs646776) versus homozygotes for the major haplotype (major allele of rs646776), normalized to the mean level in minor haplotype homozygotes, in the MDC-CC cohort (measured by ion mobility) and the PARC cohort (measured by gradient gel electrophoresis). LDL-L, large LDL; LDL-M, medium LDL; LDL-S, small LDL; LDL-VS, very small LDL. **b**, Relative gene positions in and around the 1p13 locus; * indicates position of rs646776. **c**, Mean expression of local genes in homozygotes for the major 1p13 haplotype (major allele of rs646776) versus heterozygotes versus homozygotes for the minor 1p13 haplotype (minor allele of rs646776), normalized to the mean level in major haplotype homozygotes, in samples of human liver, human subcutaneous adipose and human omental adipose. **d**, Mean expression of *PSRC1*, *CELSR2*, *SORT1* and *TCF7L2* (negative control) mRNA, standardized to *B2M* expression, and sortilin protein, standardized to α-tubulin, in samples of human liver from homozygotes for the major 1p13 haplotype (major allele of rs12740374) versus heterozygotes versus homozygotes for the minor 1p13 haplotype (minor allele of rs12740374) if available, normalized to the mean level in major haplotype homozygotes. *P* values derived from linear regression analyses or unpaired *t*-test. Error bars show s.e.m.

A causal 1p13 noncoding DNA variant

We performed fine mapping of the 1p13 locus to define the minimal DNA region responsible for the LDL-C association. Because rs646776, rs599839, rs12740374 and rs629301 lie between *CELSR2* and *PSRC1*, we used data from a recent GWAS of ~20,000 individuals of European descent[7] to perform association analyses with LDL-C on these and other SNPs spanning the two genes. Out of 18 other SNPs, we identified two SNPs with *P* values comparable to rs646776, rs599839, rs12740374 and rs629301 (*P* values ranging from 1.8×10^{-42} to 8.3×10^{-41}) and no SNPs with lower *P* values (Supplementary Fig. 2a). Together these six best SNPs cluster in a noncoding DNA region that is 6.1 kilobases in size, spanning the 3′ untranslated region (3′UTR) of *CELSR2*, the intergenic region, and the *PSRC1* 3′UTR oriented in the opposite direction (Fig. 2a). The six SNPs are in high linkage disequilibrium (LD) and comprise two predominant haplotypes in HapMap Europeans (CEU), with the 'major' haplotype present on 68% of chromosomes 1 and the 'minor' haplotype on 29% (Supplementary Fig. 3).

We identified two human bacterial artificial chromosomes (BACs) harbouring the major and minor haplotypes of the 6.1 kb region. We sequenced the region on each of the BACs in full and identified 16 polymorphisms (Supplementary Fig. 2b). From each BAC, the region spanning precisely between the stop codon of *CELSR2* and the stop codon of *PSRC1* was subcloned into firefly luciferase expression constructs just distal to the stop codon of the luciferase gene in either the 'forward' (*CELSR2*) or 'reverse' (*PSRC1*) orientation. On transfection of the constructs into Hep3B cultured human hepatocellular carcinoma cells, we found that in both orientations, the minor haplotype produced significantly greater luciferase expression than the major haplotype, consistent with the human liver eQTL analyses (Fig. 2b, compare to Fig. 1c). After localizing the haplotype-specific effect to the proximal 2.1 kb of the region (Supplementary Fig. 4), we tested an array of constructs in which single polymorphisms in the minor haplotype were switched to major alleles. We identified the SNP rs12740374 as being sufficient to confer the haplotype-specific effect (Figs 2c and 3c).

We genotyped rs12740374 and 15 other SNPs in or near the 6.1 kb noncoding region in ~9,000 African Americans. Whereas six SNPs have indistinguishable evidence for association with LDL-C in Europeans, we found that, in African Americans, rs12740374 alone had the strongest evidence for association ($P = 2.3 \times 10^{-20}$ for rs12740374 versus 9.2×10^{-15} at the next best 1p13 SNP) (Supplementary Fig. 2a). This is consistent with rs12740374 being in high LD with nearby SNPs in HapMap Europeans (CEU), but not so in HapMap Africans (YRI) (Supplementary Fig. 3).

We observed that rs12740374 alters a predicted binding site for C/EBP transcription factors, with the minor allele creating the site and the major allele disrupting it; the binding site is not present in the orthologous DNA region in mice (Fig. 3a). C/EBPα (also known as CEBPA) is a liver-enriched transcriptional factor that regulates the expression of numerous hepatic genes involved in a variety of metabolic processes[17]. We tested binding of the rs12740374 minor and major allele sequences by C/EBP with electrophoretic mobility shift assays and found the minor allele sequence to be shifted as much as a classic C/EBP binding sequence[18], with minimal shifting of the major allele sequence; addition of either of two C/EBPα antibodies impaired the binding (Fig. 3b).

We tested luciferase constructs in Hep3B cells expressing a dominant negative C/EBP protein (A-C/EBP)[19,20] and found significantly reduced differences in haplotype-specific expression (Fig. 3d and Supplementary Fig. 5b). We also tested luciferase constructs in NIH 3T3 cultured mouse fibroblast cells (Fig. 3e) and found no haplotype-specific expression difference, consistent with liver specificity. Addition of C/EBPα to the 3T3 cells restored the haplotype-specific effect (Fig. 3e). We altered other nucleotides besides rs12740374 in the consensus binding site predicted to be critical for C/EBP protein–DNA interactions[21] and found that they were needed for transcriptional activation by the minor

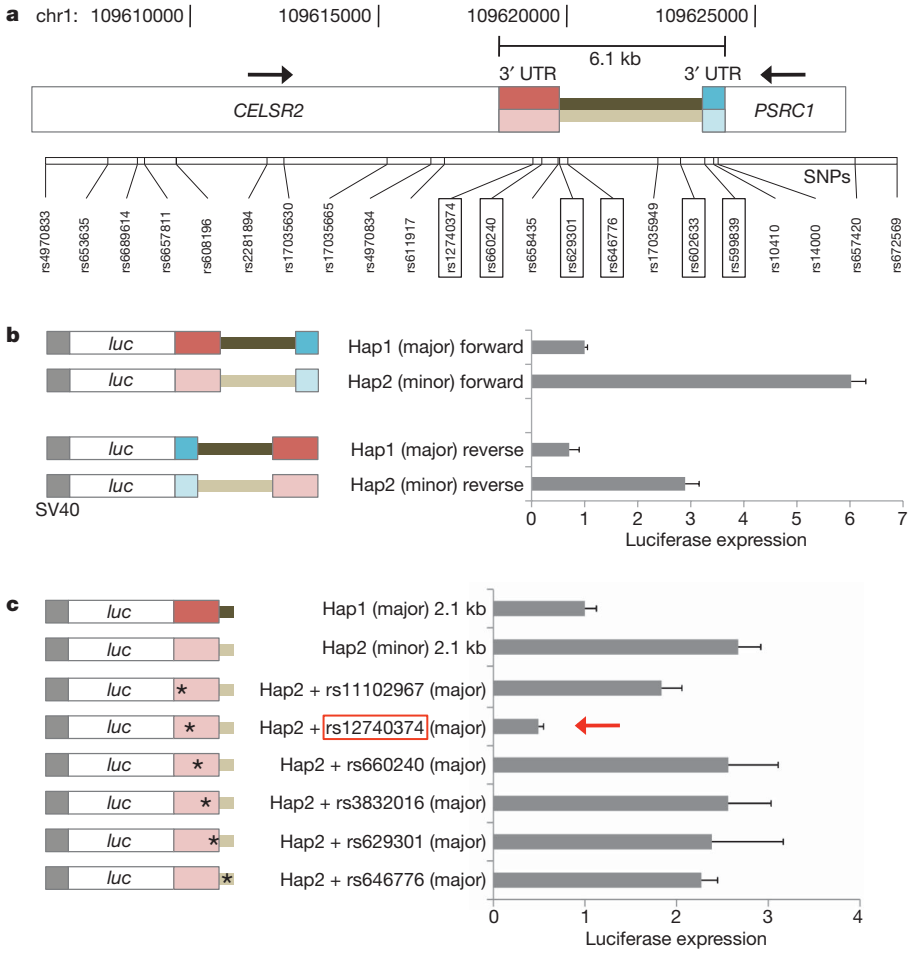

Figure 2 | rs12740374 is responsible for haplotype-specific difference in transcriptional activity. a, Map of 1p13 SNPs genotyped in ~20,000 individuals of European descent relative to *CELSR2* and *PSRC1* genes. The six SNPs with strongest association with LDL-C (indicated with boxes), comprising a single haplotype, define the 6.1 kb region between the stop codons of the two genes. **b**, Firefly luciferase expression from constructs transfected into Hep3B human hepatoma cells. Both the major (darker colours) and minor (lighter colours) haplotypes of the 6.1 kb region were subcloned in forward and reverse orientations into a basal firefly luciferase construct with the SV40 promoter. Shown are ratios of firefly luciferase expression to *Renilla* luciferase expression (expressed from cotransfected plasmid), measured 48 h after transfection, normalized to the mean ratio from the major haplotype, forward orientation construct. Error bars show s.e.m., $n = 2$. **c**, Both the major and minor haplotypes of a minimal 2.1 kb region were subcloned into the basal construct. Single nucleotide alterations were introduced individually into the minor haplotype, changing minor alleles of SNPs into major alleles. Shown are ratios of firefly luciferase expression to *Renilla* luciferase expression normalized to the mean ratio from the major haplotype construct. Error bars show s.e.m., $n = 4$.

haplotype (Fig. 3c). Furthermore, we determined that C/EBPα binds to the site of rs12740374 in homozygous minor allele cells by chromatin immunoprecipitation (Supplementary Fig. 5c).

We tested whether C/EBP proteins can influence *SORT1* expression via rs12740374. When we added A-C/EBP to Hep3B cells that are homozygous for the major allele, there was no difference in *SORT1* expression (Fig. 3f). In contrast, when we added A-C/EBP to SK-HEP-1 cultured human hepatoma cells that are heterozygous (one minor allele), we observed a threefold reduction in *SORT1* expression (Fig. 3f). When we added C/EBPα to human embryonic stem (ES) cells that are homozygous for the minor allele (HUES-1), there was no change in *SORT1* expression, presumably because ES cells do not harbour cofactors needed for transcriptional activation (Supplementary Fig. 5d). When we differentiated HUES-1 cells into endoderm, the first step towards hepatocyte differentiation[22], addition of C/EBPα resulted in significantly increased *SORT1* expression; in contrast, human ES cells homozygous for the major allele (HUES-9), when differentiated into endoderm, showed no expression difference (Supplementary Fig. 5d).

Together, these findings indicate that rs12740374 is the causal variant responsible for the liver-specific association between the 1p13 locus and gene expression and, by extension, the associations with LDL-C and MI risk.

Sort1 in mouse liver alters plasma lipids

Of the genes differentially expressed in human liver by 1p13 genotype, the *SORT1* gene showed the largest difference (Fig. 1c). *SORT1* encodes the sortilin protein[23], also known as neurotensin receptor 3, a protein that functions as a multiligand sorting receptor. Sortilin localizes to various intracellular compartments including the Golgi apparatus and has roles in both endocytosis and intracellular trafficking of other proteins[24]. To model the functional effects of altered *SORT1* expression on lipids and lipoproteins, we performed knockdown and overexpression studies of *Sort1* in the livers of mice. Importantly, we chose approaches to specifically alter gene expression in liver, because variation at the 1p13 locus results in a large *SORT1* expression change in liver, but no change in adipose tissues (Fig. 1c). Because sortilin is known to be highly expressed and have important physiological roles in adipocytes and neurons[25,26], we felt it was most appropriate to restrict knockdown and overexpression to liver to model the effects of the 1p13 locus on phenotype. Because wild-type mice have very low levels of plasma LDL-C compared to humans, we used 'humanized' mice of various genetic backgrounds for our studies (Supplementary Fig. 6).

Adeno-associated virus serotype 8 (AAV8) has been demonstrated to appropriately target genes for specific expression in liver[27,28]. AAV8 vector encoding the murine *Sort1* gene driven by a liver-specific promoter (thyroglobulin) was delivered to mouse liver via intraperitoneal injection. A null AAV8 vector was used as a control. The *Sort1* AAV resulted in increased sortilin levels in liver with no change in adipose tissue (Supplementary Fig. 7a). Use of these viral vectors did not result in elevated alanine aminotransferase (ALT) levels (Supplementary Fig. 7b).

When compared with mice injected with null virus, *Sort1*-overexpressing *Apobec1*$^{-/-}$; *APOB* Tg mice showed a marked decrease

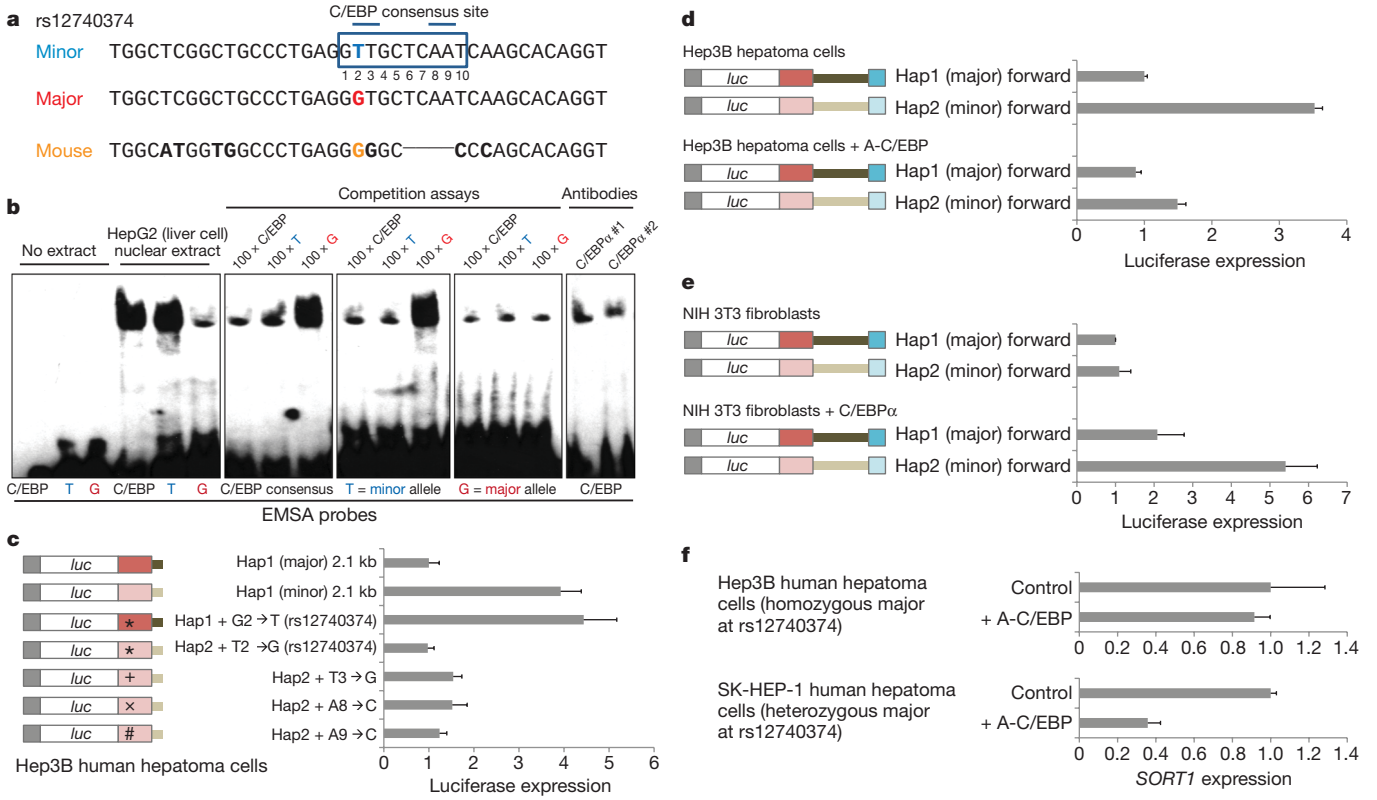

Figure 3 | rs12740374 alters a C/EBP transcription factor binding site. **a**, The human DNA sequence surrounding rs12740374, major and minor alleles, and orthologous DNA sequence in mouse. The major allele of rs12740374 disrupts one of two core elements (position 2, 3 and 8, 9) in the predicted consensus binding site on which a C/EBP dimer binds[21]. **b**, Electrophoretic mobility shift assays (EMSA) with labelled probes matching the C/EBP consensus binding site[18], the rs12740374 minor allele (T) sequence, and the rs12740374 major allele (G) sequence. Competition assays were performed with 100-fold excess of cold probe. Either of two C/EBPα antibodies was used to compete for binding and/or shift the protein–DNA complex. **c**, Relative firefly luciferase expression from constructs with haplotypes of 2.1 kb region transfected into Hep3B cells. Single nucleotide alterations were introduced into constructs as indicated, altering rs12740374 and the three other core recognition nucleotides in the predicted C/EBP binding site. **d**, **e**, Relative firefly luciferase expression from constructs with haplotypes of 6.1 kb region transfected into (**d**) Hep3B human hepatoma cells with or without concomitant transduction with A-C/EBP (dominant negative C/EBP) cDNA via lentivirus and (**e**) NIH 3T3 fibroblasts with or without concomitant transduction with C/EBPα cDNA via lentivirus. **f**, Relative *SORT1* expression, determined as a ratio with *B2M* expression by qRT–PCR, in Hep3B cells (homozygous major (GG) at rs12740374) or SK-HEP-1 human hepatoma cells (heterozygous (GT) at rs12740374) with or without concomitant transduction with A-C/EBP cDNA via lentivirus. Error bars show s.e.m., $n = 3$ for each experiment.

in total plasma cholesterol (70% reduction at 2 weeks, 46% reduction at 6 weeks) and LDL-C (73% reduction at 2 weeks) (Fig. 4a, d); consistent results were seen in three other mouse backgrounds (Supplementary Figs 6 and 7c–f). At 6 weeks the mice had a 73% reduction in very small LDL particles and an 88% reduction in medium small LDL particles (Fig. 4b), resulting in increased LDL peak particle size (22.0 nm versus 20.9 nm, $P = 0.05$). These gain-of-function studies in mice were concordant with the genetic findings in human cohorts, in whom the 1p13 minor haplotype was associated with increased liver *SORT1* expression as well as decreased LDL-C and, especially, very small LDL particles (Fig. 1a, c).

To study hepatic VLDL secretion, we administered Pluronic F-127 detergent to the AAV-injected mice and measured lipoproteins at serial time points. We found a 57% decrease in the rate of VLDL secretion (Fig. 4c) and a similar decrease in the rate of triglyceride secretion (data not shown) in *Sort1*-overexpressing mice.

Chemically synthesized small interfering RNA (siRNA)-mediated knockdown of *Apob* or *Pcsk9* in liver has been successful in determining the effects of these genes on plasma lipid levels[29,30]. We used a similar approach to reduce *Sort1* expression in mouse liver. We identified siRNA duplexes that effected >90% knockdown of *Sort1* expression in cells (Supplementary Fig. 8a). We selected one chemically modified duplex with a low half-maximal inhibitory concentration (IC_{50}) that did not induce cytokines in a human peripheral blood mononuclear cell assay (Supplementary Fig. 8b; data not shown) for large-scale preparation in a lipidoid formulation optimized for liver-specific delivery[31] and injection into mouse tail veins. As a negative control in some experiments, we used a chemically modified, non-immunostimulatory siRNA duplex specific for the firefly luciferase gene. *Sort1* siRNA achieved 70–80% reduction in *Sort1* expression in liver, confirmed to be due to siRNA-mediated cleavage, as well as reduced sortilin levels in liver with no change in adipose tissue (Supplementary Fig. 9a–c).

Sort1 knockdown in *Apobec1*$^{-/-}$; *APOB* Tg mice resulted in a 46% increase in total cholesterol compared to control mice at 2 weeks, with a more than twofold increase in LDL-C (Fig. 4e, f). Consistent results were seen in two other mouse backgrounds, as well as a significant increase in the plasma VLDL level (Supplementary Figs 6 and 9e–g). We also compared plasma lipid levels in *Sort1* knockout mice[32] and wild-type mice and observed significantly higher total cholesterol and LDL-C levels in the knockout mice (Supplementary Fig. 9h), consistent with the results of liver-specific knockdown.

To confirm that the altered plasma VLDL levels in the overexpression and knockdown mice were due specifically to altered VLDL secretion from hepatocytes, we performed labelling experiments using primary hepatocytes isolated from these mice. With *Sort1* knockdown, we observed a significant increase in labelled apoB-100 secretion; with *Sort1* overexpression, there was decreased apoB-100 secretion (Supplementary Fig. 10).

Besides *SORT1*, *PSRC1* displayed the greatest differential expression in human liver by 1p13 genotype (Fig. 1c). We used an AAV8 vector encoding the murine *Psrc1* gene for mouse liver overexpression and did not observe any significant changes in total cholesterol or LDL-C levels (Fig. 4g, h).

A novel lipoprotein regulatory pathway

Through a series of studies in human cohorts, mice and hepatocytes, we provide evidence that a single noncoding DNA variant at the chromosome 1p13 locus, rs12740374, influences LDL-C and MI risk via liver-specific transcriptional regulation of the *SORT1* gene by C/EBP transcription factors. The clinical importance of this novel

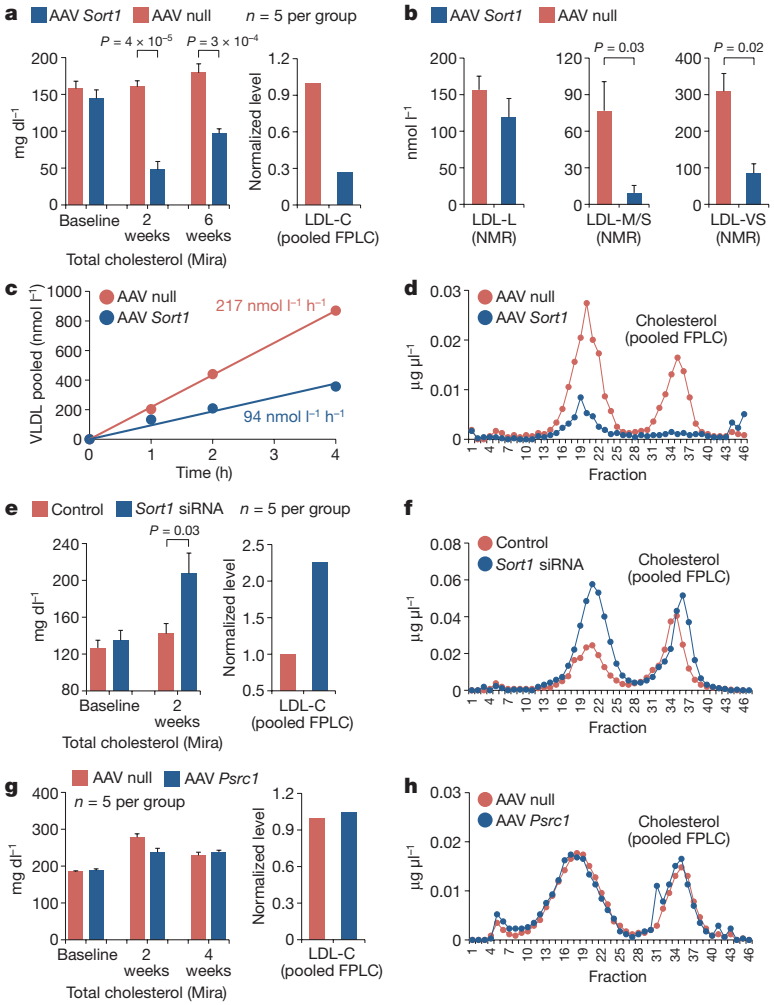

Figure 4 | Overexpression or knockdown of *Sort1* in mouse liver alters plasma lipids and lipoproteins. Adeno-associated virus 8 (AAV8) vectors either containing no gene, murine *Sort1* cDNA or murine *Psrc1* cDNA were administered via intraperitoneal injection; phosphate-buffered saline or siRNA duplex targeting firefly luciferase or mouse *Sort1* and prepared in lipidoid formulation was administered weekly via tail vein injection at 2.0 mg kg^{-1}. Plasma samples were collected before injection and at various time points after injection, and were subjected: individually to analytical chemistry (Mira autoanalyser) to measure total cholesterol (**a**, **e**, **g**); as pooled samples to FPLC (**d**, **f**, **h**), from which fractions 10 to 26 were used to calculate LDL-C levels (**a**, **e**, **g**); individually to NMR to measure LDL particle concentrations (**b**). *P* values calculated with unpaired *t*-test, shown if $P < 0.05$. Error bars show s.e.m. **a–d**, *Apobec1*$^{-/-}$; *APOB* Tg mice (five mice per group). **b**, NMR measurements at 6 weeks. **c**, Mice were injected intraperitoneally with Pluronic F-127 detergent to block VLDL triglyceride lipolysis and permit assessment of the rate of VLDL secretion. Plasma samples were collected at baseline, 1 h, 2 h and 4 h after injection. VLDL particle concentrations were measured from pooled samples with NMR. **e**, **f**, *Apobec1*$^{-/-}$; *APOB* Tg mice (five mice per group). **g**, **h**, *Ldlr*$^{-/-}$ mice (five mice per group).

pathway is defined by the ~40% difference in MI risk between alternative 1p13 homozygotes, an effect comparable to those of common variants of *LDLR* and *PCSK9* and larger than the effects of common variants in *HMGCR* (the target of statin drugs)[11,12]. As the 1p13 minor allele frequency is about 30% in Europeans and is also common in other ethnicities including African Americans, Hispanics, Asian Indians and Chinese[8,33], this locus is an important global genetic determinant of MI risk. We note that among lipid-regulating genes related to MI, *SORT1* is unique for having been identified by GWAS mapping of common DNA variants, rather than by discovery of rare gene variants underlying Mendelian disorders.

In conclusion, our results nominate *SORT1* as the causal gene at the 1p13 locus for LDL-C and MI and the sortilin pathway as a promising new target for therapeutic intervention in the reduction of LDL-C and prevention of MI. They also provide insights into mechanisms by which common noncoding genetic variants can lead to clinical phenotypes, rather than simply being markers for disease.

METHODS SUMMARY

The full Methods provides information about all experimental procedures: (1) description of association analyses in the population cohorts; (2) description of genotype-expression analyses in human liver, subcutaneous adipose, and omental adipose samples; (3) details for generation of luciferase expression constructs; (4) details for conducting luciferase expression assays; (5) details for conducting *SORT1* expression assays; (6) details for performing electrophoretic mobility shift assays; (7) details for performing chromatin immunoprecipitation assays; (8) description of siRNA screening and validation; (9) details for performing gene knockdown studies in mouse liver; (10) details for performing gene overexpression studies in mouse liver; (11) details for measuring lipids and lipoproteins by analytic chemistry, fast protein liquid chromatography, and NMR; (12) details for performing VLDL secretion studies; and (13) details for performing hepatocyte apoB studies.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 3 August 2009; accepted 9 June 2010.

1. Rader, D. J. *et al.* Monogenic hypercholesterolemia: new insights in pathogenesis and treatment. *J. Clin. Invest.* **111,** 1795–1803 (2003).
2. Brown, M. S. & Goldstein, J. L. Heart attacks: gone with the century? *Science* **272,** 629 (1996).
3. Waters, D. D. *et al.* Lipid treatment assessment project 2: a multinational survey to evaluate the proportion of patients achieving low-density lipoprotein cholesterol goals. *Circulation* **120,** 28–34 (2009).
4. Kathiresan, S. *et al.* Six new loci associated with blood low-density lipoprotein cholesterol, high-density lipoprotein cholesterol or triglycerides in humans. *Nature Genet.* **40,** 189–197 (2008).
5. Willer, C. J. *et al.* Newly identified loci that influence lipid concentrations and risk of coronary artery disease. *Nature Genet.* **40,** 161–169 (2008).
6. Wallace, C. *et al.* Genome-wide association study identifies genes for biomarkers of cardiovascular disease: serum urate and dyslipidemia. *Am. J. Hum. Genet.* **82,** 139–149 (2008).
7. Kathiresan, S. *et al.* Common variants at 30 loci contribute to polygenic dyslipidemia. *Nature Genet.* **41,** 56–65 (2009).
8. Aulchenko, Y. S. *et al.* Loci influencing lipid levels and coronary heart disease risk in 16 European population cohorts. *Nature Genet.* **41,** 47–55 (2009).
9. Sabatti, C. *et al.* Genome-wide association analysis of metabolic traits in a birth cohort from a founder population. *Nature Genet.* **41,** 35–46 (2009).
10. Teslovich, T. M. *et al.* Biological, clinical and population relevance of 95 loci for blood lipids. *Nature* doi:10.1038/nature09270 (this issue).
11. Samani, N. J. *et al.* Genomewide association analysis of coronary artery disease. *N. Engl. J. Med.* **357,** 443–453 (2007).
12. Myocardial Infarction Genetics Consortium. Genome-wide association of early-onset myocardial infarction with single nucleotide polymorphisms and copy number variants. *Nature Genet.* **41,** 334–341 (2009).
13. Berneis, K. K. & Krauss, R. M. Metabolic origins and clinical significance of LDL heterogeneity. *J. Lipid Res.* **43,** 1363–1379 (2002).
14. Musunuru, K. *et al.* Ion mobility analysis of lipoprotein subfractions identifies three independent axes of cardiovascular risk. *Arterioscler. Thromb. Vasc. Biol.* **29,** 1975–1980 (2009).
15. Siri-Tarino, P. W., Williams, P. T., Fernstrom, H. S., Rawlings, R. S. & Krauss, R. M. Reversal of small, dense LDL subclass phenotype by normalization of adiposity. *Obesity (Silver Spring)* **17,** 1768–1775 (2009).
16. Linsel-Nitschke, P. *et al.* Genetic variation at chromosome 1p13.3 affects sortilin mRNA expression, cellular LDL-uptake and serum LDL levels which translates to the risk of coronary artery disease. *Atherosclerosis* **208,** 183–189 (2010).
17. Darlington, G. J., Wang, N. & Hanson, R. W. C/EBP alpha: a critical regulator of genes governing integrative metabolic processes. *Curr. Opin. Genet. Dev.* **5,** 565–570 (1995).
18. Osada, S., Yamamoto, H., Nishihara, T. & Imagawa, M. DNA binding specificity of the CCAAT/enhancer-binding protein transcription factor family. *J. Biol. Chem.* **271,** 3891–3896 (1996).
19. Olive, M., Williams, S. C., Dezan, C., Johnson, P. F. & Vinson, C. Design of a C/EBP-specific, dominant-negative bZIP protein with both inhibitory and gain-of-function properties. *J. Biol. Chem.* **271,** 2040–2047 (1996).
20. Ahn, S. *et al.* A dominant-negative inhibitor of CREB reveals that it is a general mediator of stimulus-dependent transcription of c-fos. *Mol. Cell. Biol.* **18,** 967–977 (1998).
21. Miller, M., Shuman, J. D., Sebastian, T., Dauter, Z. & Johnson, P. F. Structural basis for DNA recognition by the basic region leucine zipper transcription factor CCAAT/enhancer-binding protein alpha. *J. Biol. Chem.* **278,** 15178–15184 (2003).
22. Si-Tayeb, K. *et al.* Highly efficient generation of human hepatocyte-like cells from induced pluripotent stem cells. *Hepatology* **51,** 297–305 (2010).
23. Petersen, C. M. *et al.* Molecular identification of a novel candidate sorting receptor purified from human brain by receptor-associated protein affinity chromatography. *J. Biol. Chem.* **272,** 3599–3605 (1997).
24. Nielsen, M. S. *et al.* The sortilin cytoplasmic tail conveys Golgi-endosome transport and binds the VHS domain of the GGA2 sorting protein. *EMBO J.* **20,** 2180–2190 (2001).
25. Nykjaer, A. *et al.* Sortilin is essential for proNGF-induced neuronal cell death. *Nature* **427,** 843–848 (2004).
26. Shi, J. & Kandror, K. V. Sortilin is essential and sufficient for the formation of Glut4 storage vesicles in 3T3–L1 adipocytes. *Dev. Cell* **9,** 99–108 (2005).
27. Kitajima, K. *et al.* Complete prevention of atherosclerosis in apoE-deficient mice by hepatic human apoE gene transfer with adeno-associated virus serotypes 7 and 8. *Arterioscler. Thromb. Vasc. Biol.* **26,** 1852–1857 (2006).
28. Tanigawa, H. *et al.* Expression of cholesteryl ester transfer protein in mice promotes macrophage reverse cholesterol transport. *Circulation* **116,** 1267–1273 (2007).
29. Soutschek, J. *et al.* Therapeutic silencing of an endogenous gene by systemic administration of modified siRNAs. *Nature* **432,** 173–178 (2004).
30. Frank-Kamenetsky, M. *et al.* Therapeutic RNAi targeting PCSK9 acutely lowers plasma cholesterol in rodents and LDL cholesterol in nonhuman primates. *Proc. Natl Acad. Sci. USA* **105,** 11915–11920 (2008).
31. Akinc, A. *et al.* A combinatorial library of lipid-like materials for delivery of RNAi therapeutics. *Nature Biotechnol.* **26,** 561–569 (2008).
32. Zeng, J., Racicott, J. & Morales, C. R. The inactivation of the sortilin gene leads to a partial disruption of prosaposin trafficking to the lysosomes. *Exp. Cell Res.* **315,** 3112–3124 (2009).
33. Keebler, M. E. *et al.* Association of blood lipids with common DNA sequence variants at 19 genetic loci in the multiethnic United States National Health and Nutrition Examination Survey III. *Circ. Cardiovasc. Genet.* **2,** 238–243 (2009).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank D. Altshuler, E. Fisher and J. Maraganore for advice and guidance, and A. Akinc, J. Billheimer, R. Brown, R. Camahort, D. Cromley, E. Eduoard, I. Fuki, C. Geaney, G. Hinkle, I. Kohaar, S. Kuchimanchi, W. Lagor, F. Lau, D. Lum, M. Maier, D. Marchadier, R. Meyers, J. Millar, S. Milstein, D. Nguyen, D. Perez, D. Peters, V. Redon, A. Rigamonti, R. Schinzel, M.-S. Sun, S.-A. Toh, A. Wilson and K. Wojnoonski for assistance and suggestions. We acknowledge the National Heart, Lung, and Blood Institute (NHLBI) Gene Therapy Resource Program for providing support for viral vector production as well as the Vector Core laboratory of the University of Pennsylvania for producing the vectors. We acknowledge the members of the NHLBI Candidate Gene Association Resource (CARe) lipids working group for the contribution of association data in African Americans. This work was supported in part by a T32 grant in Cell and Molecular Training for Cardiovascular Biology from the United States National Institutes of Health (NIH), K99-HL098364 from the NIH, and the Clinician Scientist Program of the Harvard Stem Cell Institute (K.M.); a Medical Scientist Training Program grant from the NIH (A.S.); the intramural research program of the Division of Cancer Epidemiology & Genetics, National Cancer Institute, NIH (L.P.-O.); the Swedish Medical Research Council, Heart-Lung Foundation, and Påhlsson Foundation (M.O.-M., O.M.); U01-HL069757 from the NIH and research support from Quest Diagnostics, Inc. (R.M.K.); RC2-HL101864 from the NIH (S.K.); and P01-HL059407 and RC2-HL101864 from the NIH and a "Freedom to Discover" Unrestricted Cardiovascular Research Grant from Bristol-Myers Squibb (D.J.R.).

Author Contributions K.M., A.S., M.F.-K., N.E.L., T.A., K.V.S., X.L., H.L., N.K., V.M.R., J.J.P., B.M., L.P.-O., J.L.H., E.E.S., C.R.M., S.L.-K., M.C.P., J.W., W.C., T.R., K.G.E., M.O.-M., O.M. and R.M.K. carried out experimental work and/or performed data analysis. V.K., K.F., C.A.C., S.K. and D.J.R. supervised the study. K.M., A.S., S.K. and D.J.R. conceived and designed the study. K.M. wrote the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare competing financial interests: details accompany the full-text HTML version of the paper at www.nature.com/nature. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to D.J.R. (rader@mail.med.upenn.edu) or S.K. (skathiresan@partners.org).

METHODS

Association analyses. The Malmö Diet and Cancer Study – Cardiovascular Cohort (MDC-CC) is a prospective, community-based epidemiological cohort of 6,103 residents of Malmö, Sweden, for whom a comprehensive analysis of cardiovascular risk factors has been performed. The 1p13 SNP rs646776 was genotyped as described previously[14]. The ion mobility method of lipoprotein measurement was applied to archived baseline blood samples from these individuals to quantify directly the full spectrum of lipoprotein particles, as described previously[14]. Multivariable linear regression analyses were used to test whether each of the lipid or lipoprotein measures differed according to an increasing copy number of the SNP minor allele, adjusted for age, gender and diabetes status. SPSS (version 16.0) was used for the analyses.

The Pharmacogenomics and Risk of Cardiovascular Disease (PARC) study is a two-stage genome-wide association (GWA) study[34]. In stage 1, 980 subjects were typed for 317,000 SNPs with the Illumina Human-1 BeadChip. In stage 2, 930 additional subjects were typed for a subset of 13,680 SNPs with the Illumina iSelect platform. All subjects were of self-reported European ancestry. The gradient gel electrophoresis method of lipoprotein measurement was applied to blood samples from these individuals to quantify directly the full spectrum of lipoprotein particles, as described previously[15]. Multivariable linear regression analyses were used to test whether each of the lipid or lipoprotein measures differed according to an increasing copy number of the SNP minor allele. JMP (SAS Institute) was used for the analyses.

Roughly 20,000 individuals of European descent were genotyped on various array platforms, and roughly 9,000 African American individuals were genotyped on the ITMAT-Broad-CARe Array (Illumina). Association analyses for LDL-C and meta-analyses were performed as described previously[7,35].

Genotype-expression analyses. To evaluate whether SNPs serve as eQTLs with putative *cis* regulatory effects on liver and adipose gene expression traits, 782,476 SNPs had been genotyped and expression levels of 39,280 transcripts profiled in 960 human liver samples, 433 human subcutaneous adipose samples, and 520 human omental adipose samples. Tissue samples were either post-mortem or surgical resections from organ donors or elective cases. Methods for tissue collection, RNA and DNA isolation, expression profiling and DNA genotyping have been described previously[36]. The correlation of rs646776 minor allele count with each of the profiled transcripts was determined using linear regression analysis.

For the replication study in 62 liver samples, de-identified histopathologically normal human liver samples were provided by the University of Minnesota Academic Health Center's Biological Materials Procurement Facility (BioNet; www.bionet.umn.edu). For each sample, 1 μg of DNAase-treated total RNA was converted into cDNA with random hexamers and SuperScript III reverse transcriptase (Invitrogen). cDNA samples were diluted with water and 2 ng of total RNA was used for each quantitative reverse transcriptase-polymerase chain reaction (qRT–PCR), performed with TaqMan Gene Expression Assays for *SORT1*, *PSRC1*, *CELSR2*, *TCF7L2* and *B2M* (beta-2-microglobulin) and associated reagents (Applied Biosystems) according to the manufacturer's protocol. Expression of all assays was measured in technical duplicates and average values of the duplicates were used for the analysis. The *SORT1*, *PSRC1*, *CELSR2* and *TCF7L2* expression values for each target were normalized by *B2M* expression values (ΔC_t method) and were tested for normality of distribution before analysis. A pre-developed TaqMan genotyping allelic discrimination assay for SNP rs12740374 was used according to the manufacturer's protocol (Applied Biosystems). A univariate linear regression analysis was used to test the associations between mRNA expression and the SNP coded by the number of major alleles and was performed with SPSS 16.0. Information on age and sex was tested as a covariate but was not included in the final analysis as it was not available for all samples.

Protein extracts from liver tissue samples were prepared by homogenization of ~ 30 mg of tissue with Tissue Lyser (Qiagen) in RIPA buffer (Invitrogen) in the presence of complete cocktail of proteinase inhibitors (Roche). Samples were subjected to immunoblotting and probed with anti-sortilin antibody (AF2934, R&D Systems) or anti-α-tubulin antibody as loading control (ab-7291-100, Abcam).

Luciferase expression constructs. To characterize the intergenic region between *CELSR2* and *PSRC1*, the major (Hap1) and minor (Hap2) haplotypes from two bacterial artificial chromosomes (CTD-2068B15 and RP11-463O24, respectively; Invitrogen) were cloned into the pGL3-Promoter vector (Promega) in both the 5′-to-3′ and 3′-to-5′ orientations just downstream of the stop codon of the firefly luciferase gene. A naturally occurring BamHI site was used to generate constructs with truncations and composites of the two haplotypes. PCR was used to generate smaller truncations. The QuikChange Site-Directed Mutagenesis Kit (Stratagene) was used to alter single nucleotides (that is, SNP alleles). All constructs were verified by DNA sequencing.

Luciferase expression assays. Hep3B cultured human hepatoma cells, BNL CL.2 cultured mouse embryonic liver cells or NIH 3T3 cultured mouse fibroblast cells were transfected at roughly 50% confluence and maintained in DMEM with 10% FBS. In some experiments, cells were infected with a lentivirus encoding the C/EBPα cDNA or a lentivirus encoding the A-C/EBP[19,20] (dominant negative C/EBP) cDNA 24 h before transfection. The firefly luciferase constructs were co-transfected with the *Renilla* luciferase pRL-CMV Vector (Promega) using the FuGENE 6 transfection reagent (Roche) in the ratio 1 μg:100 ng:3 μl mixed with Opti-MEM I Reduced Serum Medium (Invitrogen) for a 100 μl mix, of which 20 μl was used for each well of 24-well plates. Forty-eight hours after transfection, firefly and *Renilla* luciferase activities were measured using the Dual-Luciferase Reporter Assay System (Promega) according to the manufacturer's protocol, using untransfected cells to adjust for background activity.

***SORT1* expression assays.** Hep3B cultured human hepatoma cells or SK-HEP-1 cultured human hepatoma cells were seeded at roughly 50% confluence in 24-well plates and infected with a lentivirus encoding the A-C/EBP (dominant negative C/EBP) cDNA or mock-infected; virus was removed after 24 h. The cells were maintained in DMEM with 10% FBS. Total RNA was isolated with the RNeasy Mini Kit (Qiagen) 72 h after infection.

HUES-1 or HUES-9 human embryonic stem cells were seeded at roughly 50% confluence on Geltrex matrix (Invitrogen) and initially maintained on mTeSR1 medium (StemCell Technologies). The cells were then switched to and maintained on EndoMedia (RPMI-B27 medium, supplemented with 100 ng ml^{-1} human recombinant activin A, Invitrogen) for 7 days to induce differentiation into definitive endoderm. Successful differentiation was confirmed in parallel experiments by monitoring morphological changes and detecting expression of endoderm-specific markers *SOX17* and *GATA4*. C/EBPα was expressed through lentiviral infection during the last 2 days of differentiation, followed by isolation of total RNA with the RNeasy Mini Kit (Qiagen).

For each sample, 2 μg of total RNA was converted into cDNA with the High Capacity cDNA Reverse Transcription Kit (Applied Biosystems). qRT–PCR was performed with TaqMan Gene Expression Assays for *SORT1* and *B2M* and associated reagents (Applied Biosystems) according to the manufacturer's protocol. The *SORT1* expression values for each target were normalized by *B2M* expression values (ΔC_t method).

Electrophoretic mobility shift assays (EMSA). Primers with the consensus C/EBPα binding site were described previously[18]: C/EBPα-F, 5′-CTAGGCATATTGCGCAATATGC-3′; C/EBPα-R, 5′-GCATATTGCGCAATATGCCTAG-3′. Primers for rs12740374 were designed based on genomic sequences surrounding the SNP (http://www.ncbi.nlm.nih.gov/projects/SNP/): rs12740374_T-F, 5′-TGCCCTGAGG**T**TGCTCAATCA-3′; rs12740374_T-R, 5′-TGATTGAGCAA**C**CTCAGGGCA-3′; rs12740374_G-F, 5′-TGCCCTGAGG**G**TGCTCAATCA-3′; rs12740374_G-R: 5′-TGATTGAGCA**C**CCTCAGGGCA-3′. The variable nucleotide is shown in bold. All primers were ordered from Invitrogen. Individual primers were labelled with a biotin 3′ end DNA labelling kit (Pierce) according to instructions, and the efficiency of labelling was tested by a dot-test that confirmed that all the primers were labelled similarly. Corresponding forward and reverse primers were annealed to create 3′-end biotin-labelled double-stranded probes. EMSA reactions were performed with the biotin 3′-end DNA labelling kit (Pierce) according to instructions, with 8 μg of nuclear extract from HepG2 cultured human hepatoma cells per reaction (Active Motif). For competition assays, we used 100-fold excess of unlabelled probe. To test for involvement of CEBP/α in interaction with the probes, we preincubated the HepG2 nuclear extract for 15 min at room temperature with either of two antibodies for CEBP/α (39306, Active Motif; 2295, Cell Signaling). The protein complexes were resolved on 6% DNA retardation gels (Invitrogen) for 1 h at 100 V, transferred to Biodyne B Nylon Membranes (Pierce), crosslinked, and processed with the Chemiluminescent Nucleic Acid Detection Module (Pierce).

Chromatin immunoprecipitation assays. HUES-1 human embryonic stem cells were seeded at roughly 50% confluence on Geltrex matrix (Invitrogen) and maintained on mTeSR1 medium (StemCell Technologies). The cells were infected with a lentivirus encoding C/EBPα; virus was removed after 24 h. Seventy-two hours after infection, the cells were harvested and cross-linked with 4% paraformaldehyde at 37 °C for 10 min followed by quenching with glycine and flash freezing. After thawing, the lysates were sonicated in RIPA buffer 25 times for 10 s at 4 °C. The lysates were precipitated with anti-C/EBPα antibody (2295, Cell Signaling) at 1:40 dilution overnight versus no antibody. After incubation with Protein G Sepharose beads (GE Healthcare) for 2 h at room temperature, serial washes, and elution, DNA was recovered by addition of sodium chloride and incubation overnight at 65 °C, followed by treatment with proteinase K and RNase A for 2 h at 42 °C. DNA was purified with the QIAquick PCR Purification Kit (Qiagen). The presence of immunoprecipitated DNA sequence around rs12740374 was assayed by quantitative PCR using the primers 5′-CTGAGGTTGCTCAATCAAGCGCTTGATTGAGCAACCTCAG-3′

and 5′-CTGAGGGTGCTCAATCAAGCGCTTGATTGAGCACCCTCAG-3′ and the probe 5′-FAM-AGCCAGCACTGTGTTTACTCTTCCTC-Iowa Black-3′ (Integrated DNA Technologies). The values for the immunoprecipitated target were normalized by values for the target from 1:30 dilution of input chromatin.

siRNA screening and validation. siRNA design was carried out to identify siRNAs targeting both homologues of the gene *SORT1* from human (symbol *SORT1*) and mouse (symbol *Sort1*). The design used the *SORT1* transcripts NM_002959.4 (human) and NM_019972.2 (mouse) from the NCBI RefSeq collection. siRNA duplexes were designed with 100% identity to both respective *SORT1* genes. To select appropriate candidate target sequences and their corresponding siRNAs, their predicted potentials for interacting with irrelevant targets (off-target potentials) were used as a ranking parameter. siRNAs with low off-target potentials were defined as preferable and assumed to be more specific *in vivo*. To identify potential off-target genes, 19-mer candidate sequences were subjected to a homology search against the human and mouse RefSeq mRNA databases. The following off-target properties for each 19-mer input sequence were extracted for each off-target gene to calculate the off-target score: number of mismatches in non-seed region, number of mismatches in seed region, and number of mismatches in cleavage site region. The 29 siRNAs with best off-target scores were selected for synthesis and screening.

Single-stranded RNAs were produced at Alnylam Pharmaceuticals as described previously[31,30]. Deprotection and purification of the crude oligoribonucleotides by anion exchange high performance liquid chromatography (HPLC) were carried out according to established procedures. siRNAs were generated by annealing equimolar amounts of complementary sense and antisense strands.

For screening transfection experiments, BNL CL.2 cultured mouse embryonic liver cells were seeded at 4×10^4 cells per well in 24-well plates and reverse transfected with the siRNAs using Lipofectamine RNAiMAX (Invitrogen) according to the manufacturer's protocol. Total RNA was isolated with the RNeasy Mini Kit (Qiagen) 48 h after transfection. For each sample, 2 μg of total RNA was converted into cDNA with the High Capacity cDNA Reverse Transcription Kit (Applied Biosystems). qRT–PCR was performed with TaqMan Gene Expression Assays for *Sort1* and 18S rRNA and associated reagents (Applied Biosystems) according to the manufacturer's protocol. The *Sort1* expression values for each target were normalized by 18S rRNA expression values (ΔC_t method). A half-maximal inhibitory concentration (IC_{50}) curve was determined for the *Sort1* duplex yielding the greatest degree of knockdown. The sequences for this duplex were: 5′-uGucAGAAuGGucGAGAcudTsdT-3′ and 5′-AGUCUCGACcAUUCUGAcAdTsdT-3′ (2′-OMe modified nucleotides are in lower case, and phosphorothioate linkages are indicated by 's').

A previously validated siRNA duplex targeting the luciferase gene was used[31].

Gene knockdown studies in mouse liver. Lipidoid formulations of siRNAs were prepared as described previously[31]. Mice received either phosphate-buffered saline (PBS) or formulated siRNAs via weekly tail vein injection at dosages of 2.0 mg kg^{-1}. At various time points (including before injection), animals were anaesthetized by isoflurane inhalation, and blood was collected by retro-orbital bleed followed by centrifugation to isolate plasma. Mice were killed at 5 days or at 2 weeks after 4 h of fasting. After killing the mice, terminal bleeds were collected, and livers and adipose were collected and snap frozen in liquid nitrogen. Frozen tissue was ground, and tissue lysates were prepared. *Sort1* mRNA levels relative to those of *GAPDH* mRNA were determined in the liver lysates by using the branched-DNA-technology-based QuantiGene Reagent System (Panomics), according to the manufacturer's protocols. Sortilin and actin expression in liver and adipose was determined by immunoblotting (612100, BD Transduction Laboratories; ab20272, Abcam).

5′-RACE (rapid amplification of cloned/cDNA ends) was conducted as described previously[30]. In brief, an oligonucleotide adaptor was ligated to total liver RNA, and the ligation mixture was reverse transcribed using the *Sort1*-specific oligonucleotide 5′- TATTCCAGGAGGTCCTCATCTGAGTCGTC-3′, followed by cDNA amplification with the oligonucleotides 5′-CGACTGGAG CACGAGGACACTGACATGG-3′ and 5′- GGATTCATCCCACCTTGGCATTT GTCTC-3′. Nested PCR was performed using the oligonucleotides 5′- GGACA CTGACATGGACTGAAGGAGTAG-3′ and 5′- GAAGTAGCCAAAGTCACAG AGGAAGTC-3′. PCR products were examined by gel electrophoresis, purified, and subcloned for sequencing.

Sort1$^{-/-}$ mice were generated as described previously[32] and outbred to the C57BL/6 strain. Matched wild-type C57BL/6 mice were used as controls.

All mice were fed ad libitum with regular rodent chow. All procedures used in animal studies were approved by the pertinent Institutional Animal Care and Use Committee and were consistent with local, state and federal regulations as applicable.

Gene overexpression studies in mouse liver. The murine *Sort1* cDNA (Origene, MR210834) was subcloned into a specialized vector for use by the University of Pennsylvania's Penn Vector Core for production of AAV8 viral particles expressing *Sort1*. Viruses were produced with a chimaeric packaging construct in which the AAV2 rep gene was fused with the cap gene of AAV serotype 8 (ref. 27). Empty AAV8 viral particles were also provided by the Penn Vector Core.

Mice received either 1×10^{12} viral particles of null AAV or 1×10^{12} viral particles of AAV-encoding *Sort1* in PBS via intraperitoneal injection. At various time points (including before injection), animals were anaesthetized by isoflurane inhalation and blood was collected by retro-orbital bleed followed by centrifugation to isolate plasma. Mice were killed at 6 weeks after 4 h of fasting. After killing the mice, terminal bleeds were collected and livers and adipose were collected and analysed for protein and gene expression as described above.

Measurement of mouse plasma lipids and lipoproteins. Collected mouse plasma samples were analysed for lipids by analytical chemistry and fast protein liquid chromatography (FPLC) and for lipoproteins by nuclear magnetic resonance (NMR). Total plasma cholesterol and alanine aminotransferase (ALT) were measured enzymatically on a Cobas Mira autoanalyser (Roche Diagnostic Systems). Pooled plasma from each experimental group (140 μl) was separated by FPLC gel filtration. Cholesterol and triglyceride plate assays were performed on FPLC fractions using the Infinity cholesterol and triglyceride reagents, respectively. Individual plasma samples were sent for NMR lipoprotein measurement (LipoScience).

VLDL secretion studies. To study hepatic VLDL secretion, mice were prebled by retro-orbital bleeding followed by intraperitoneal injection of 400 μl of 1 mg g^{-1} Pluronic F-127 detergent resuspended in PBS. The mice were fasted for 4 h before injection and through the study. We performed serial retro-orbital bleeds at 1, 2, and 4 h after injection of the detergent. Plasma samples were individually subjected to triglyceride measurements by analytical chemistry (plate assays with the Infinity triglyceride reagent) and pooled together by experimental condition and sent for NMR analysis for VLDL measurement (LipoScience).

Primary hepatocyte apoB studies. Mice of the *Apobec*$^{-/-}$; *APOB* Tg; *Ldlr*$^{+/-}$ or *Apobec*$^{-/-}$; *Ldlr*$^{-/-}$ background that had been administered AAV vectors or siRNAs were used as the source of primary hepatocytes for all experiments. Mice were anesthetized with 2,2,2-tribromoethanol and then dissected to expose the liver, portal vein, and inferior vena cava. A catheter was inserted into the portal vein and sutured in place. The livers were perfused with buffer for 5 min to remove all red blood cells, followed by digestion *in situ* by running digestion media through the catheter for 15 min. The livers were transferred to 10 mm dishes with 15 ml of hepatocyte wash media and run through a mesh into 50 ml conical tubes to separate the cells. The cells were centrifuged at 50*g* at 4 °C to remove Kupffer cells. The hepatocyte pellets were washed twice with hepatocyte wash media and resuspended in 25 ml PBS + 25 ml of Percoll solution (45 ml Percoll + 5 mL 10× PBS + 100 μl of 1 M HEPES). The cells were then centrifuged at 115*g* for 5 min at 4 °C to pellet the viable hepatocytes. The hepatocytes were resuspended in Hepatozyme medium + 10% FBS + 1% amino acids and plated at one million cells per well. A subset of the cells was analysed for sortilin and actin protein expression as described above.

For labelling experiments, cells were switched to cystine/methionine-free DMEM with 1% FBS, 1% antibiotics/antimycotics, and 0.4 mM oleic acid for 1 h, followed by addition of 200 μCi per well of ^{35}S-methionine/cysteine. After 3 h, media from the cells were harvested, and apoB was immunoprecipitated with the antibody ab20737 (Abcam). The immunoprecipitate was subjected to SDS–PAGE, and the gel was exposed to film at –80 °C for 3 days to 2 weeks. Relative secreted apoB-100 levels were determined by quantification of appropriately sized bands by densitometry.

To determine relative total secreted protein levels, 50 μl of 2 mg ml^{-1} BSA and 25 μl of 50% trichloroacetic acid (TCA) were added to 50 μl of harvested media, followed by incubation on ice for 20 min. The samples were centrifuged for 15 min, and the pellets were washed with 1 ml of 50% TCA and resuspended by boiling in 1 ml of 0.2 M NaOH. The NaOH suspension (200 μl) was analysed in a scintillation counter for ^{35}S counts.

34. Reiner, A. P. *et al.* Polymorphisms of the *HNF1A* gene encoding hepatocyte nuclear factor-1α are associated with C-reactive protein. *Am. J. Hum. Genet.* **82**, 1193–1201 (2008).
35. Musunuru, K. *et al.* Candidate Gene Association Resource (CARe): design, methods, and proof of concept. *Circ. Cardiovasc. Genet.* **3**, 267–275 (2010).
36. Schadt, E. E. *et al.* Mapping the genetic architecture of gene expression in human liver. *PLoS Biol.* **6**, e107 (2008).