Multiple Hepatic Regulatory Variants at the *GALNT2* GWAS Locus Associated with High-Density Lipoprotein Cholesterol

Tamara S. Roman,[1,2] Amanda F. Marvelle,[1] Marie P. Fogarty,[1] Swarooparani Vadlamudi,[1] Arlene J. Gonzalez,[1] Martin L. Buchkovich,[1] Jeroen R. Huyghe,[3] Christian Fuchsberger,[3] Anne U. Jackson,[3] Ying Wu,[1] Mete Civelek,[4,5] Aldons J. Lusis,[4,6,7] Kyle J. Gaulton,[1,8] Praveen Sethupathy,[1] Antti J. Kangas,[9] Pasi Soininen,[9,10] Mika Ala-Korpela,[9,10,11,12] Johanna Kuusisto,[13] Francis S. Collins,[14] Markku Laakso,[13] Michael Boehnke,[3] and Karen L. Mohlke[1,*]

Genome-wide association studies (GWASs) have identified more than 150 loci associated with blood lipid and cholesterol levels; however, the functional and molecular mechanisms for many associations are unknown. We examined the functional regulatory effects of candidate variants at the *GALNT2* locus associated with high-density lipoprotein cholesterol (HDL-C). Fine-mapping and conditional analyses in the METSIM study identified a single locus harboring 25 noncoding variants ($r^2 > 0.7$ with the lead GWAS variants) strongly associated with total cholesterol in medium-sized HDL (e.g., rs17315646, $p = 3.5 \times 10^{-12}$). We used luciferase reporter assays in HepG2 cells to test all 25 variants for allelic differences in regulatory enhancer activity. rs2281721 showed allelic differences in transcriptional activity (75-fold [T] versus 27-fold [C] more than the empty-vector control), as did a separate 780-bp segment containing rs4846913, rs2144300, and rs6143660 (49-fold [AT^- haplotype] versus 16-fold [CC^+ haplotype] more). Using electrophoretic mobility shift assays, we observed differential CEBPB binding to rs4846913, and we confirmed this binding in a native chromatin context by performing chromatin-immunoprecipitation (ChIP) assays in HepG2 and Huh-7 cell lines of differing genotypes. Additionally, sequence reads in HepG2 DNase-I-hypersensitivity and CEBPB ChIP-seq signals spanning rs4846913 showed significant allelic imbalance. Allelic-expression-imbalance assays performed with RNA from primary human hepatocyte samples and expression-quantitative-trait-locus (eQTL) data in human subcutaneous adipose tissue samples confirmed that alleles associated with increased HDL-C are associated with a modest increase in *GALNT2* expression. Together, these data suggest that at least rs4846913 and rs2281721 play key roles in influencing *GALNT2* expression at this HDL-C locus.

Introduction

Genome-wide association studies (GWASs) have identified more than 150 loci associated with blood lipid and cholesterol levels.[1–5] One of the first novel GWAS signals for high-density lipoprotein cholesterol (HDL-C) levels in Europeans was reported for variants rs2144300 and rs4846914, located within intron 1 of *GALNT2*[6,7](MIM: 602274). These two lead GWAS variants are in perfect linkage disequilibrium (LD; $r^2 = 1$). The association signal for rs4846914 has been replicated ($n = 187{,}000$ individuals, $p = 4 \times 10^{-41}$) in subsequent studies with larger sample sizes[2,4] and in Japanese and Mexican populations.[8,9] HDL-C-associated variant rs4846914 is also associated ($n = 178{,}000$, $p = 7 \times 10^{-31}$) with triglycerides[4] and nominally associated ($n = 2{,}744$–$3{,}481$, $p < 0.05$) with large HDL concentration, low-density lipoprotein (LDL) size, HDL size, HDL-2 subfraction, and the ratio of total cholesterol to HDL-C.[10] The alleles associated with increased HDL-C are also nominally associated ($n = 84{,}068$, $p = 0.04$) with decreased risk of coronary artery disease.[4]

According to 1000 Genomes phase 1 version 3 European (EUR) data,[11] 24 variants exhibit strong LD ($r^2 > 0.7$) with the lead *GALNT2* HDL-C-associated SNP, rs4846914, and all 25 variants are noncoding. According to available chromatin data from the ENCODE Project[12] and Human Epigenome Atlas,[13] these variants overlap many candidate regulatory regions. Therefore, we hypothesized that one or more of these variants regulate gene expression. Many variants identified through GWASs are located within noncoding or intergenic regions,[14,15] and variants at the *GALNT2* locus might also alter regulatory elements.

GALNT2, encoding UDP-N-acetylgalactosamine:polypeptide N-acetylgalactosaminyltransferase (GALNT2), is a

[1]Department of Genetics, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599, USA; [2]Curriculum in Genetics and Molecular Biology, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599, USA; [3]Department of Biostatistics and Center for Statistical Genetics, School of Public Health, University of Michigan, Ann Arbor, MI 48109, USA; [4]Department of Medicine, David Geffen School of Medicine, University of California, Los Angeles, Los Angeles, CA 90095, USA; [5]Department of Biomedical Engineering, Center for Public Health Genomics, University of Virginia, Charlottesville, VA 22908, USA; [6]Deparment of Microbiology, Immunology, and Molecular Genetics, David Geffen School of Medicine, University of California, Los Angeles, Los Angeles, CA 90095, USA; [7]Department of Human Genetics, David Geffen School of Medicine, University of California, Los Angeles, Los Angeles, CA 90095, USA; [8]Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford OX3 7BN, UK; [9]Computational Medicine, Institute of Health Sciences, University of Oulu, 90014 Oulu, Finland; [10]Nuclear Magnetic Resonance Metabolomics Laboratory, School of Pharmacy, University of Eastern Finland, 70211 Kuopio, Finland; [11]Oulu University Hospital, 90220 Oulu, Finland; [12]Computational Medicine, School of Social and Community Medicine and Medical Research Council Integrative Epidemiology Unit, University of Bristol, Bristol BS8 2BN, UK; [13]Department of Medicine, University of Eastern Finland and Kuopio University Hospital, 70210 Kuopio, Finland; [14]National Human Genome Research Institute, NIH, Bethesda, MD 20892, USA
*Correspondence: mohlke@med.unc.edu
http://dx.doi.org/10.1016/j.ajhg.2015.10.016.

reasonable positional candidate within the HDL-C association signal. GALNT2 is an enzyme that transfers an N-acetylgalactosamine to serine or threonine residues in target proteins in the initial step of O-linked glycosylation.[16] *GALNT2* is expressed in many tissues, including liver, heart, lung, muscle, pancreas, ovary, and colon.[17–19] GALNT2 might influence HDL-C levels by catalyzing O-glycosylation on target proteins that play a role in lipid metabolism. In vitro, GALNT2 has been shown to O-glycosylate lecithin-cholesterol acyltransferase (LCAT), phospholipid transfer protein (PLTP), and angiopoietin-like protein 3 (ANGPTL3), and O-glycosylation of ANGPTL3 was shown to inhibit activation of this protein.[20] Individuals heterozygous for a *GALNT2* missense variant (c.941A>C [p.Asp314Ala]) shown to decrease GALNT2 function in vitro exhibit decreased glycosylation of apoC-III and high (>95th percentile for age and gender) plasma HDL levels.[21] In mice, liver-specific *Galnt2* overexpression and knockdown have been shown to decrease and increase HDL-C levels, respectively.[2]

We aimed to identify the functional regulatory variant(s) responsible for the *GALNT2* HDL-C GWAS signal by fine mapping the association with lipoprotein traits in the METSIM (Metabolic Syndrome in Men) study[22] and by examining a comprehensive set of candidate variants for evidence of allelic differences in enhancer function. We identified a single signal driven by at least two regulatory variants, rs4846913 and rs2281721, that exhibited binding of transcription factors and allelic differences in enhancer activity, as well as additional variants that might contribute to enhancer function. In human hepatocyte and subcutaneous adipose tissue samples, we observed an association between this GWAS signal and *GALNT2* expression. Together, these data show a consistent direction of regulatory effect in which increased expression of *GALNT2* is implicated in increased HDL-C.

Material and Methods

Defining the Candidate Set of Variants

We used the 1000 Genomes Project Phase 1 version 3 EUR dataset,[11] including 24 variants in strong LD ($r^2 > 0.7$) with the lead HDL-C GWAS SNP rs4846914, to calculate LD. We used ENCODE data[12] available through the UCSC Genome Browser to determine which of the 25 total variants overlapped open-chromatin peaks and chromatin-immunoprecipitation-sequencing (ChIP-seq) peaks of histone modifications H3K4me1, H3K4me2, H3K4me3, H3K9ac, and H3K27ac and transcription factors in liver cell types (HepG2 cells, Huh-7 cells, and primary human hepatocytes) and ChromHMM chromatin states[23] in multiple cell types. We used the Human Epigenome Atlas[13] to determine overlap with peaks of H3K4me1 and H3K9ac in primary adult liver and with ChromHMM chromatin states in multiple cell types and tissues.

Genotyping and Imputation

We used the Illumina HumanOmniExpress and HumanCoreExome[24] Beadchips to genotype 10,134 Finnish men from the METSIM study.[22] Sample-level and SNP-level quality control included detecting sample contamination,[25] confirming sex and relationships, and using principal-component analysis to detect population outliers. After we filtered SNPs to retain those with a call rate of >95% and Hardy-Weinberg Equilibrium $p > 10^{-6}$, we successfully analyzed 10,082 individuals and 681,803 SNPs. The METSIM study was approved by the ethics committee of the University of Kuopio and Kuopio University Hospital, and informed consent was obtained from all study participants. To impute ungenotyped SNPs, we used a panel of 5,474 reference haplotypes derived from genome sequences of 2,737 central-northern European individuals sequenced as part of the Genetics of Type 2 Diabetes study (C.F., J. Flannick, K.J.G., H. Kang, and the GoT2D Consortium, unpublished data). The minimum MaCH imputation quality score for the imputed variants was $R^2 = 0.971$. We used a two-step imputation strategy wherein individuals were pre-phased with ShapeIT version 2 before imputation using Minimac.[26,27] We also used these data to verify LD proxies for allelic-expression-imbalance (AEI) assays.

Fine-Mapping and Conditional Analyses

We analyzed 72 measures of lipid and lipoprotein particle-serum concentration obtained via nuclear-magnetic-resonance (NMR) metabolomics (65 traits) or enzymatic assays (7 traits) in up to 10,079 Finnish men. The NMR platform has been described previously.[28,29] The methodology for measuring lipoprotein subclasses has been described previously,[30] and subclasses are defined in Table S1. Trait values with skewed distributions were log transformed, and all traits were Winsorized at 5 SDs from the mean. After adjustment for age, squared age, smoking status, and lipid-lowering-medication status, we transformed residuals to a standard normal. We tested for association between normalized residuals and SNPs with a minor allele frequency (MAF) > 0.0005 (minor allele count > 10) by assuming an additive genetic model and using a linear mixed model with an empirical kinship matrix to account for relatedness, as implemented in EMMAX.[31] We repeated the analysis while excluding individuals on lipid-lowering medication (2,844 participants), individuals with type 1 or 2 diabetes (1,420 participants), or both (3,560 participants) to assess sensitivity to these exclusions, and in each instance we obtained qualitatively similar results. To identify any additional independent signals in the region, we performed a conditional analysis by using the rs17315646 allele count as an additional covariate in the model. LocusZoom[32] plots were generated to include 2,079 variants in a 350-kb region surrounding the lead variant rs17315646 and spanning *GALNT2*.

Cell Culture

Human HepG2 hepatocellular carcinoma cells (ATCC, HB-8065) were grown in Eagle's minimum essential medium (MEM) alpha supplemented with 10% fetal bovine serum (FBS) and 1 mM sodium pyruvate at 37°C and 5% CO_2. Human Huh-7 hepatocellular carcinoma cell lines (JCRB0403, Japanese Collection of Research Bioresources Cell Bank, National Institute of Biomedical Innovation) were grown in DMEM supplemented with 10% FBS, 1 mM sodium pyruvate, 1× MEM non-essential amino acids, and 2 mM L-glutamine. HepG2 and Huh-7 cells were seeded into 24-well plates (100,000 cells per well) 1 day prior to transfection experiments.

Transcriptional Reporter Assays

To test candidate variants and haplotypes for allele-specific effects on transcriptional activity, we amplified segments of 109–262 bp

(for single variants) and 349–780 bp (for multiple variants) from DNA of individuals homozygous for each allele or haplotype. Segment size was chosen to include the ~147 nucleotides of DNA spanning one nucleosome. However, because of the proximity of other candidate variants, larger segments of 349–780 bp were designed to include multiple variants. Primer sequences are listed in Table S2. Amplicons were cloned into the KpnI and XhoI restriction sites of the firefly luciferase transcriptional reporter vector pGL4.23 in both forward and reverse orientations with respect to the minimal promoter (Promega). Three to seven independent plasmids for each allele or haplotype were isolated and confirmed by sequencing. When additional variants were identified, we selected clones for which the alleles matched at these variants. For simplicity of presentation, Figures 3 and S2 do not show the differing alleles of rs1555290; these data are provided in Figures 4 and S5. We then transfected each purified clone into HepG2 or Huh-7 cells in duplicate (720 ng in each well) by using FuGENE 6 (Promega) and Opti-MEM (Life Technologies). To control for transfection efficiency, we co-transfected a phRL-TK *Renilla* luciferase reporter vector (80 ng in each well) into cells. For empty-vector controls, two independent preparations of empty vector were each transfected into HepG2 or Huh-7 cells in duplicate. After 48 hr, cell-lysate luciferase activity was measured with the Dual-Luciferase Reporter Assay System (Promega), normalized, and compared to readings for empty-vector controls. These control readings were very similar for the two independent preparations. We performed two-tailed t tests to compare the luciferase activities between variant alleles. For comparisons of multiple haplotypes, we performed ANOVA and Tukey's post hoc tests by using JMP 10.0.1 software (SAS Institute). We also assessed the transcriptional activity of the 780-bp segment by cloning the haplotype into the promoterless pGL4.10 firefly luciferase reporter vector (Promega). To examine individual variant effects within the 780-bp segment, we altered alleles by using the QuikChange Site-Directed Mutagenesis Kit (Agilent Technologies) and confirmed them by sequencing.

Electrophoretic Mobility Shift Assays

Complementary DNA oligonucleotides (17–19 bp) centered on variant alleles (Table S2) were synthesized by Integrated DNA Technologies. The rs6143660 insertion allele consisted of 39-bp complementary oligonucleotides containing the 21-bp insertion (9 bp flanking each side of the insertion), and the rs6143660 deletion allele consisted of 18-bp complementary oligonucleotides (9 bp flanking each side of the deleted sequence). Labeled oligonucleotides included biotin on the 5′ end. We performed assays as previously described[33] by using 3.5–6 μg of HepG2 nuclear lysate and 30- to 300-fold excess unlabeled probe. For supershift reactions, 4–8 μg CEBPB antibody (sc-150X), USF1 antibody (sc-229X), and FOXO3 antibody (sc-34895X), all from Santa Cruz Biotechnology, were incubated with binding buffer, poly(dI·dC), and HepG2 nuclear lysate for 20 min at room temperature before the addition of labeled DNA probes and incubation. Additional control antibodies (4–6 μg, all from Santa Cruz Biotechnology) were chosen on the basis of transcription-factor binding motifs, ENCODE ChIP-seq peaks, expression of transcription factors in liver, or plausible roles of the factors in cholesterol metabolism. These included antibodies to ARNT (sc-271801X), SF1 (sc-10976X), HNF4A (sc-6556X), RXRA (sc-553X), CEBPA (sc-61X), CEBPB (sc-150X), NR1H3 (sc-1202X), and MAX (sc-765X). Reactions were loaded into a 6% DNA retardation gel (Life Technologies), subjected to electrophoresis, transferred to Biodyne B nylon membranes (Life Technologies), and UV crosslinked. Wash and detection steps were performed according to instructions in the Chemiluminescent Nucleic Acid Detection Module (Life Technologies). Experiments involving electrophoretic mobility shift assays (EMSAs) were repeated on 2–7 separate days, and all had consistent results.

To predict transcription-factor binding sites, we searched databases for transcription-factor binding-site motifs in 17- to 21-bp genomic sequences containing each allele of candidate variants. For JASPAR,[34] we searched all available matrix models with a relative profile-score threshold of 80%. We also searched position-weight matrices (PWMs) from vertebrates in TRANSFAC by using default parameters in the TRANSFAC Professional's Match tool and by using PWM-SCAN.[35]

ChIP Assays

We used a TaqMan SNP Genotyping Assay (Life Technologies) to genotype HepG2 and Huh-7 cells at rs4846913. Cells were crosslinked with 1% formaldehyde, and glycine was added to stop fixation. Fixed cells were resuspended in SDS lysis buffer (1% SDS, 10 mM EDTA, and 50 mM Tris [pH 8.1]), diluted with immunoprecipitation (IP) buffer (0.01% SDS, 1.1% Triton X-100, 1.2 mM EDTA, 16.7 mM Tris [pH 8.1], and 167 mM NaCl), and sonicated on ice with a Branson sonifier for the generation of 100- to 500-bp DNA fragments. Each CEBPB IP or immunoglobulin G (IgG) reaction used two to three million cells. After preclearing with Protein A Agarose beads (sc-2001, Santa Cruz Biotechnology), 10 μg of CEBPB antibody (sc-150X, Santa Cruz Biotechnology) or 10 μg of normal rabbit IgG (sc-2027, Santa Cruz Biotechnology) was added to HepG2 and Huh-7 cell lysates and incubated overnight. Protein A Agarose beads were then added for 3 hr and washed separately with low-salt buffer, high-salt buffer, LiCl buffer, 10 mM Tris-EDTA, and elution buffer (1% SDS and 0.1 M $NaHCO_3$). The crosslinks were then reversed by the addition of 5 M NaCl and overnight incubation. Samples were incubated with 20 μl 1M Tris, 10 μl 0.5 M EDTA, and 0.03 mg proteinase K. DNA was purified by phenol-chloroform extraction and ethanol precipitation. For qPCR, a 120-bp region spanning rs4846913 was amplified with FAST SYBR Green Master Mix (Life Technologies). qPCR was performed in triplicate and quantified with respect to a standard curve generated from sonicated HepG2 DNA standards. The mean quantities of CEBPB IP and IgG control sample were normalized to input HepG2 or Huh-7 DNA. CEBPB and IgG ChIP experiments with HepG2 and Huh-7 cells were each performed twice and had similar results. We used two-tailed t tests to compare CEBPB enrichment of the rs4846913 region in HepG2 and Huh-7 cells.

We performed USF1 ChIP assays similarly, except that we used a Diagenode Bioruptor Standard sonicator. We used 10 μg of USF1 antibody (sc-229X, Santa Cruz Biotechnology) or 10 μg IgG (sc-2027, Santa Cruz Biotechnology) and purified DNA with the QIAquick PCR Purification Kit (QIAGEN) to amplify and quantify a 164-bp region spanning rs2281721. We performed USF1 and IgG ChIP experiments with HepG2 and Huh-7 cells twice each, and they had similar results.

Allelic-Imbalance Analysis of Sequence Reads from CEBPB ChIP-Seq and DNase-Seq

Sequence reads from HepG2 for DNase-I-hypersensitivity-site sequencing (DNase-seq) and CEBPβ ChIP-seq experiments generated by the ENCODE Consortium[12] were aligned to UCSC Genome Browser build hg19 with AA-ALIGNER.[36] In brief, HepG2

genotypes were obtained from the Illumina Human-1M-Duo BeadChip array genotyped at HusdonAlpha Institute of Biotechnology, and imputation was performed with the 1000 Genomes Phase 1 EUR reference panel.[11] Using this data, we verified that HepG2 is diploid in the chromosome 1 region containing *GALNT2*, and we created a personalized HepG2 reference genome containing non-reference alleles for sites at which HepG2 is homozygous for the non-reference allele. We aligned sequence reads to the personalized genome by using HepG2 heterozygous sites identified by imputation and GSNAP[37] with the following parameters: -m 1, -k 11, –basesize = 11, -sampling = 1, -terminal-threshold = 10, -n 1, -query-unk-mismatch = 1, -genome-unk-mismatch = 1, -trim-mismatch-score = 0, -t 7, and –A sam. We filtered the alignments to remove sequences aligned to more than one genomic location, sequences aligned to regions underrepresented in the reference sequence (ENCODE blacklisted[12] regions), and duplicate reads that might represent PCR artifacts. We determined the significance of allelic imbalance at rs4846913 by using an exact binomial test, based on the number of reads containing the reference allele and the total number of reads at the heterozygous site. For the HepG2 DNase-seq data and USF1 HepG2 ChIP-seq data, insufficient reads were available for allelic imbalance analysis at rs2281721.

Preparation of cDNA for *GALNT2* mRNA-Expression and AEI Assays

RNA and DNA were isolated from hepatocyte samples of 50 individuals as described previously.[38] RNA for each sample was treated with DNase I with the DNA-free Kit (Life Technologies) and added to a final concentration of 24 ng/μl in an RT-PCR reaction in the SuperScript III First-Strand Synthesis System (Life Technologies), which includes both oligo$(dT)_{20}$ and random hexamer primers. The synthesized cDNA for each sample was then diluted with diethylpyrocarbonate (DEPC)-treated water for use in mRNA-expression and AEI assays.

AEI Assays

We used a TaqMan SNP genotyping assay (Life Technologies) to genotype human hepatocyte genomic DNA (gDNA) from ADMET Technologies for the HDL-C index SNP rs4846914. To quantify allele-specific expression, we diluted hepatocyte cDNA and gDNA from 36 individuals heterozygous for rs4846914 and performed subsequent qPCR reactions in triplicate. To generate a standard curve, we mixed gDNAs from samples homozygous for each rs4846914 allele in the following ratios: 95:5, 72.5:27.5, 61.25:38.75, 50:50, 38.75:61.25, 27.5:72.5, and 5:95. We generated a standard curve by plotting the quantity of one allele against the difference between the cycle-threshold (Ct) values of the two alleles. For each heterozygous sample, we estimated the expression percentage of one allele by using the difference between the mean Ct values of the alleles and the standard curve.[38] We used two-tailed t tests to compare gDNA and cDNA values and used F-tests to determine equal or unequal variance between gDNA and cDNA samples.[38,39]

GALNT2 Hepatocyte mRNA Expression

We measured expression of *GALNT2* in 50 human hepatocyte samples by qPCR with a standard curve and FAST SYBR Green Master Mix (Life Technologies). We performed triplicate qPCR reactions including 9 ng of total cDNA in each well, Taqman Gene Expression Master Mix (Life Technologies), and primer-set sequences within *GALNT2* exons (Table S2). *GALNT2* expression values were normalized to the expression of beta-2-microglobulin (*B2M* [MIM: 109700]), natural-log transformed, and plotted according to the rs4846914 genotype (AA, AG, and GG). Using linear regression and an additive model including sex, ancestry, and age as covariates, we tested for association between the level of *GALNT2* mRNA and the rs4846914 genotype.

Lookup of Expression Quantitative Trait Loci in Subcutaneous Adipose Tissue

We looked for evidence of association between the HDL-C GWAS variant region and gene expression by using preliminary microarray expression data from subcutaneous adipose tissue from the METSIM study. In brief, we used EPACTS-multi (with EMMAX implemented to account for family relatedness) Affymetrix Human Genome U219 Array data for 1381 individuals was adjusted for 40 confounding factors,[40] inverse-normal transformed, and tested for association with variants. We analyzed variants within 1 Mb of rs4846914 and identified one probe set for *GALNT2* and other genes located in this region. Subsequent reciprocal conditional analyses included two variants in each model.

siRNA-Mediated Knockdown of *CEBPB*

HepG2 cells (80,000 per well) were plated into 24-well collagen-coated plates and then treated with 50 nM *CEBPB* Silencer Select small interfering RNA (siRNA; s2893, Thermo Fisher Scientific) or Silencer Select Negative Control No. 1 siRNA (4390843, Thermo Fisher Scientific) with Opti-MEM (Life Technologies) and Dharma-FECT1 transfection reagent (GE Healthcare Life Sciences) on the following day. Cells were incubated at 37°C with 5% CO_2 for 48 hr, and HepG2 medium was changed the day following siRNA transfection. RNA was harvested with the RNeasy Plus Mini Kit (QIAGEN), and cDNA was prepared with the SuperScript III First-Strand Synthesis System (Life Technologies). *CEBPB* and *GALNT2* expression was measured by qPCR with a standard curve. Raw expression values were normalized to the expression of *B2M*, and expression percentages were calculated by comparison to expression values in HepG2 cells transfected with the negative control siRNA. Primers for gene expression are listed in Table S2. We used two-tailed t tests to compare normalized *GALNT2* expression between HepG2 cells treated with *CEBPB* siRNA and HepG2 cells treated with negative control siRNA.

Results

Fine Mapping Shows Evidence of a Single Association Signal Strongest for Total Cholesterol in Medium HDL

We tested one of the HDL-C-associated lead GWAS variants, rs2144300 ($r^2 = 1$ with rs4846914), for association with 72 lipid and cholesterol traits in up to 10,079 Finnish men from the METSIM study. The strongest evidence of association across all 72 traits was with total cholesterol in medium HDL ($n = 9{,}810$, $\beta = 0.10$, $p = 5.3 \times 10^{-12}$; Table S3). Using this trait and variants that were directly genotyped or imputed from a reference panel of 2,737 genomes of central-northern European individuals, we performed fine-mapping association and conditional analyses. Among 2,079 total variants in the region spanning 350 kb surrounding the lead variant, the strongest evidence of association was observed for rs17315646 ($n = 9{,}810$, $p = 3.5 \times 10^{-12}$), which is in perfect LD ($r^2 = 1$) with

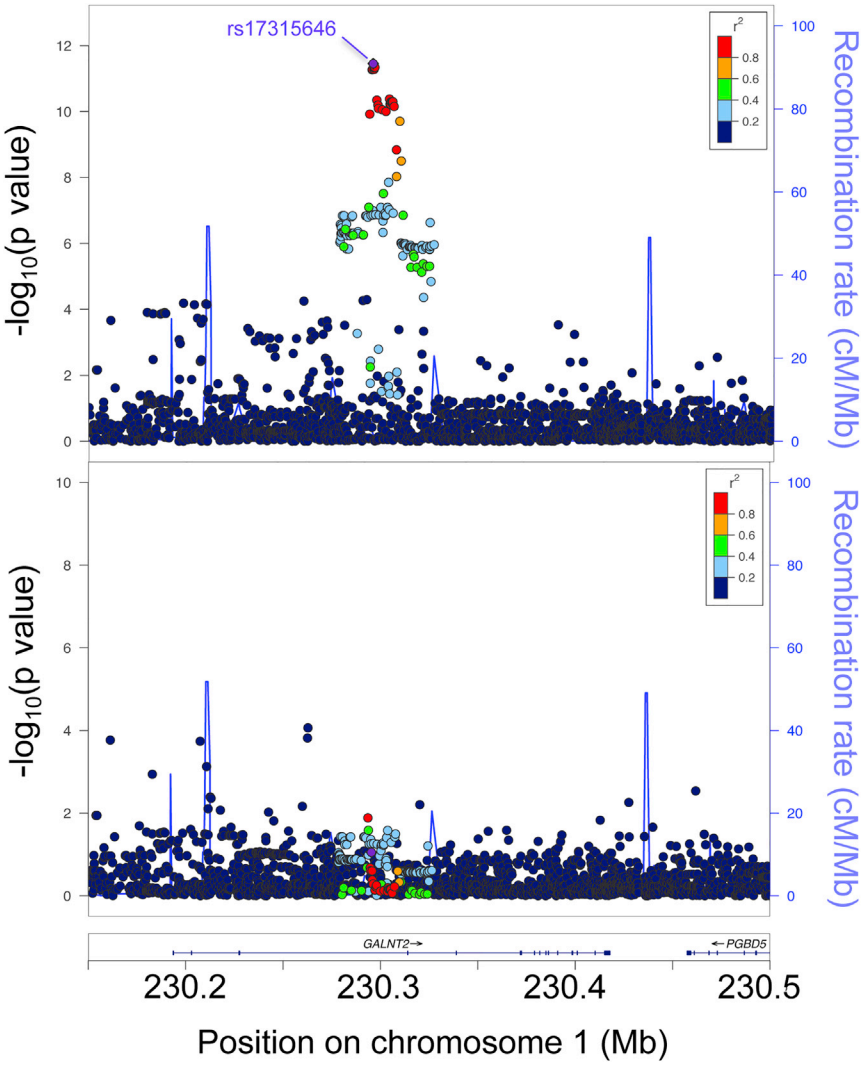

Figure 1. Non-coding Variants at *GALNT2* Are Associated with Total Cholesterol in Medium HDL in the METSIM Study
The entire initial association signal (upper panel) was reduced after conditioning on lead SNP rs17315646 (lower panel). Circles represent genotyped and imputed DNA variants and their LD r^2 values with rs17315646 in the METSIM study (2,079 variants are shown). Chromosome coordinates correspond to UCSC Genome Browser build hg19. The left y axis indicates the $-\log_{10}$(p value), the right y axis indicates the recombination rate (cM/Mb), and the x axis indicates position on chromosome 1 (Mb).

HDL-C GWAS lead SNPs rs4846914 (n = 9,810, p = 5.3 × 10^{-12}) and rs2144300 (n = 9,810, p = 5.3 × 10^{-12}). Conditioning on rs17315646 attenuated the 13.7-kb association signal (all p > 0.01; Figure 1 and Table S4). The new lead variants after conditional analysis (p_{cond} ~ 10^{-4}–10^{-6}) showed weak initial evidence of association (p ~ 10^{-2}–10^{-4}) and have a MAF < 0.04 (Table S4). These data provide evidence that common variants in strong LD with rs4846914 and rs17315646 are most likely responsible for the HDL-C association signal. All 25 candidate variants (r^2 > 0.7 with HDL-C GWAS lead variant rs4846914) are located within intron 1 of *GALNT2*. These variants all exhibited strong evidence of association with total cholesterol in medium HDL (p ≤ 1.45 × 10^{-9}; Table S4).

Open-Chromatin, Histone-Modification, and Transcription-Factor Marks Indicate Potential Regulatory Elements Overlapping the *GALNT2* Association Signal

We hypothesized that one or more of the 25 HDL-C-associated non-coding variants affect transcriptional activity. We asked whether regulatory datasets from the ENCODE Project and Human Epigenome Atlas could help us identify variants that exhibit regulatory activity. We compared the location of the 25 candidate variants to regions of open chromatin depicted by DNase I hypersensitivity and formaldehyde-assisted isolation of regulatory elements (FAIRE); histone-modification ChIP-seq peaks H3K4me1, H3K4me2, H3K4me3, H3K9ac, and H3K27ac, which often mark enhancer or promoter elements; and transcription-factor ChIP-seq peaks. On the basis of the liver's key role in HDL synthesis and transport, and the enrichment of lipid GWAS signals in liver,[4] we focused on datasets from human liver cell lines, primary human hepatocytes, and human adult liver cells. All 25 variants overlap broad patterns of H3K4me1-enriched domains in HepG2 cells (Table S5), and 13 overlap narrower H3K4me1-enriched peaks.[41] Sixteen variants overlap two or more histone-modification peaks, and six variants (rs4631704, rs4846913, rs2144300, rs6143660, rs2281721, and rs11122453) overlap at least one transcription-factor ChIP-seq peak (Figure 2 and Table S5). Three variants (rs4846913, rs2144300, and rs6143660) overlap the most marks of potential regulatory function (at least 22 peaks, including open-chromatin, histone-modification, and transcription-factor peaks). According to ChromHMM data[23] from ENCODE[12] and the Roadmap Epigenomics Project,[13] all 25 variants are located within predicted enhancer or transcribed-region chromatin states in HepG2 cells and primary liver cells (Figure S1). The regulatory peaks are not specific to liver cells, given that open-chromatin, histone-modification, and transcription-factor peaks were also observed in GM12878 lymphoblastoid cells, human umbilical-vein endothelial cells (HUVECs), K562 leukemia cells, $CD20^+$ B cells, $CD14^+$

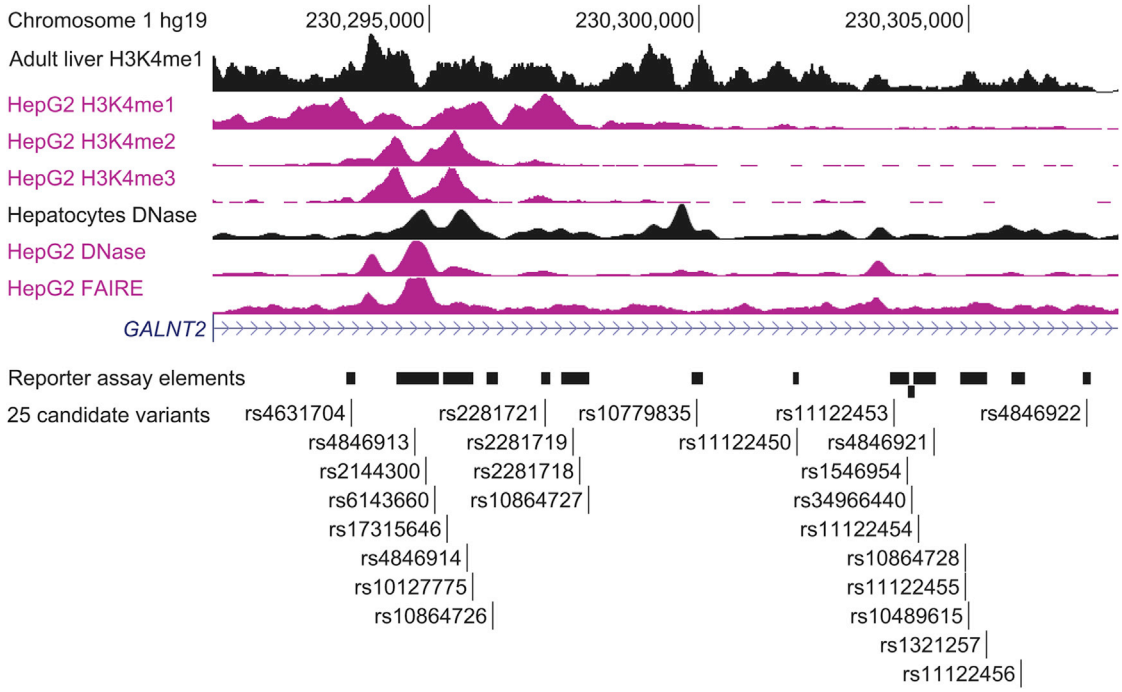

Figure 2. HDL-C-Associated Variants Overlap Open Chromatin and Histone Modifications Indicating Potential Regulatory Regions in *GALNT2* Intron 1

A 13.7-kb region includes all 24 variants in strong LD ($r^2 > 0.7$) with the HDL-C-associated index SNP rs4846914 (25 total candidate variants). Selected Human Epigenome Atlas and ENCODE open-chromatin and histone-modification tracks are shown. Rectangular bars represent elements containing the variant(s) that were tested in luciferase reporter assays.

monocytes, human skeletal-muscle myoblasts (HSMMs), normal human astrocytes (NHAs), normal human lung fibroblasts (NHLFs), and osteoblasts (Figure S1).

All 25 Candidate Regulatory Variants Were Evaluated for Allelic Differences in Luciferase Activity

Given the limited resolution of open-chromatin peaks and histone-modification and transcription-factor ChIP-seq peaks, and the knowledge that these marks do not predict allelic differences in regulatory activity, we tested all 25 variants in transcriptional reporter assays in luciferase vectors containing a minimal promoter (Figures S2 and S3). Tested elements contained one to four variants, and three to seven independent clones for each allele or haplotype in the element were tested. We considered elements exhibiting >1.5-fold more activity than empty-vector controls as enhancers. We focused on elements whose enhancer activity in both forward and reverse orientations was >1.5-fold higher than that in the luciferase reporter gene alone, as well as differences ($p < 0.05$) between the alleles or haplotypes in both orientations.

rs2281721 Exhibits Allelic Differences in Transcriptional Activity

Among all the segments tested, a 174-bp DNA segment containing rs2281721 showed the strongest enhancement of luciferase activity, and this activity also differed between the alleles (Figure 3A). In the forward orientation, the DNA segment containing the rs2281721 T allele, associated with increased HDL-C, showed 75-fold more luciferase activity than the empty-vector control, whereas the segment containing the rs2281721 C allele showed 27-fold more, and significant differences were observed between the alleles ($p = 2.4 \times 10^{-6}$). In the reverse orientation, the T and C alleles exhibited 37-fold and 14-fold, respectively, more luciferase activity than did the control, and there were significant differences between the alleles ($p = 5.8 \times 10^{-5}$). The rs2281721 T allele also showed stronger enhancer activity than did the rs2281721 C allele in both forward (205-fold versus 59-fold more than in the control) and reverse (49-fold versus 19-fold more than in the control) orientations in a second hepatocellular carcinoma cell line, Huh-7 (both $p < 3 \times 10^{-4}$; Figure S4A). The segment and the position of rs2281721 overlaps H3K4me1 peaks in HepG2 cells and adult liver; H3K4me2, H3K9ac, and H3K27ac peaks in HepG2 cells; and a USF1 transcription-factor ChIP-seq peak in HepG2 cells (Figure 2 and Table S5).

A Segment Containing rs4846913, rs2144300, and rs6143660 Shows Haplotype Differences in Transcriptional Activity

A 780-bp DNA segment containing three variants exhibited significant haplotype differences in luciferase activity (Figure 3B). We analyzed two haplotypes of rs4846913, rs2144300, and rs6143660 (a 21-bp indel): AT^- (containing the alleles associated with increased HDL-C) and CC^+ (containing the alleles associated with decreased HDL-C). In the forward orientation, the AT^- and CC^+ haplotypes showed

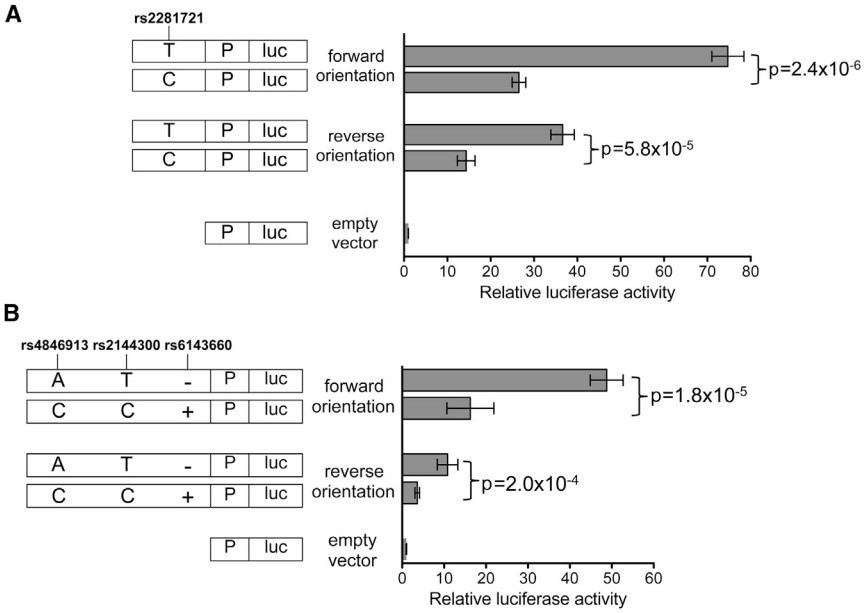

Figure 3. Haplotype and Allelic Differences in Transcriptional Activity at the *GALNT2* Locus

Segments containing each haplotype or allele were cloned into a pGL4.23 luciferase reporter vector upstream of the minimal promoter in both orientations. The vectors were transfected into HepG2 cells, and luciferase expression normalized to that of an empty vector control is shown. Error bars represent the pairwise SD of three to six independent clones per allele or haplotype (t tests). Abbreviations are as follows: P, promoter; and luc, luciferase.

(A) Luciferase activity of 174-bp DNA segments containing rs2281721 alleles.

(B) Luciferase activity of 780-bp DNA segments of two different haplotypes. The haplotypes contained three candidate variants in strong LD: rs4846913, rs2144300, and 21-bp indel, rs6143660. An additional variant, rs1555290, was detected in the segment, as shown in Figures 4 and S5.

49-fold and 16-fold, respectively, more luciferase activity than did the empty-vector control, and significant differences were observed between the haplotypes ($p = 1.8 \times 10^{-5}$). In the reverse orientation, the AT^{-} and CC^{+} haplotypes exhibited 11-fold and 4-fold, respectively, more luciferase activity than did the empty-vector control, and significant differences were observed between the haplotypes ($p = 2.0 \times 10^{-4}$). The three HDL-C-associated variants rs4846913, rs2144300, and rs6143660 (a 21-bp indel) each overlapped ≥22 open-chromatin, histone-modification, and transcription-factor peaks (Figure 2 and Table S5).

Haplotype effects on transcriptional activity were similar in Huh-7 cells. In the forward orientation, haplotypes AT^{-} and CC^{+} showed 36-fold and 13-fold, respectively, more luciferase activity than did the empty-vector control; the differences observed between the haplotypes were significant ($p = 1.4 \times 10^{-5}$ Figure S4B). In the reverse orientation, haplotypes AT^{-} and CC^{+} showed 12-fold and 6-fold, respectively, more luciferase activity than did the empty-vector control ($p = 0.08$, Figure S4B). The direction of effect was the same as that of rs2281721; alleles associated with increased HDL-C showed increased luciferase activity.

Taken together, our data show consistent haplotype differences in luciferase activity for a 780-bp segment consisting of rs4846913, rs2144300, and rs6143600. The results suggest that these variants are located within an enhancer element that can affect transcription and that one or more of them might have an allelic effect on transcriptional activity.

The Haplotype Variants Act Together to Regulate Enhancer Activity

We then created additional haplotypes of the 780-bp segment by performing site-directed mutagenesis to investigate the variant responsible for allelic differences in transcriptional activity. In addition to including the candidate variants rs4846913, rs2144300, and rs6143660, the segment included an additional common variant, rs1555290, in moderate LD with rs4846914 ($D' = 1$, $r^2 = 0.26$). We analyzed and tested natural haplotypes of rs4846913, rs2144300, rs1555290, and rs6143660 (ATA^{-}, CCC^{+}, and CCA^{+}), as well as constructed haplotypes CTA^{-}, ACA^{-}, CCA^{-}, and CCC^{-}. These seven haplotypes were then tested separately in luciferase assays (Figure 4). ATA^{-}, CTA^{-}, ACA^{-}, and CCA^{-} haplotypes showed 44-fold, 38-fold, 31-fold, and 23-fold, respectively, more luciferase activity than did the empty-vector control, suggesting that both of the first two variants, rs4846913 and rs2144300, contribute to haplotype differences in transcriptional activity. Specifically, significant differences were observed between the ATA^{-} and ACA^{-} haplotypes ($p = 0.04$), between the CTA^{-} and CCA^{-} haplotypes ($p = 0.04$), and between the ATA^{-} versus CCA^{-} haplotypes ($p = 0.001$), supporting a contribution from both rs4846913 and rs2144300 to haplotype differences in luciferase activity. Compared to the control, the CCA^{+}, CCC^{-}, and CCC^{+} haplotypes showed similar 12- to 14-fold increases in luciferase activity ($p > 0.05$; an intermediate between the CCA^{-} haplotype and the empty vector), suggesting that rs1555290 and/or rs6143660 might also contribute to increased transcriptional activity. We observed similar results in Huh-7 cells: the ATA^{-}, ACA^{-}, and CCA^{-} haplotypes showed 36-fold, 30-fold, and 27-fold, respectively, more luciferase activity than did the empty-vector control, and significant differences were observed between the ATA^{-} and CCC^{-} haplotypes ($p = 0.0001$). Compared to the empty-vector control, the CCA^{+}, CCC^{-}, and CCC^{+} haplotypes showed similar ($p > 0.05$) 13- to 15-fold increases in luciferase activity (Figure S5). Taken together, these data suggest a role for

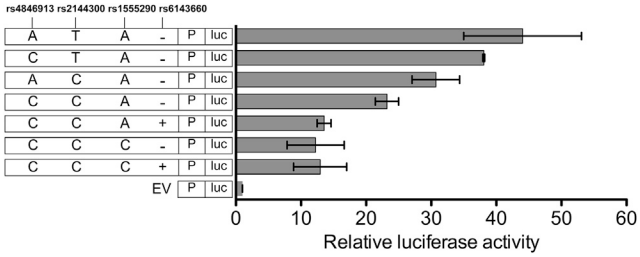

Figure 4. Haplotype Variants Act Together to Increase Transcriptional Activity
Additional haplotypes were created by site-directed mutagenesis of haplotypes cloned into a pGL4.23 luciferase vector in the forward orientation. All constructs were transfected separately into HepG2 cells, and this experiment was performed separately from the experiment presented in Figure 3. Luciferase activity was measured and normalized to that of an empty vector control. Error bars represent the pairwise SD of three to four independent clones per haplotype (ANOVA and Tukey's post hoc tests). Abbreviations are as follows: P, promoter; and luc, luciferase.

at least two variants, rs4846913 and rs2144300, in haplotype differences in enhancer activity.

Because the SNPs in the 780-bp segment overlap H3K1me1 and H3K4me2 peaks, which are frequently present in enhancer regions, as well as a H3K4me3 peak, which is often found at promoters, we also evaluated the haplotypes in a promoterless vector (Figure S6). In the forward orientation, haplotypes ATA^-, CCA^+, and CCC^+ showed 44-fold, 27-fold, 16-fold, respectively, more luciferase activity than did the empty-vector control, and significant differences were observed between all haplotype comparisons ($p < 0.001$). In the reverse orientation, haplotypes ATA^-, CCA^+, and CCC^+ exhibited 12-fold, 9-fold, and 5-fold, respectively, more luciferase activity than did the empty-vector control, and significant differences were observed between all the haplotype comparisons ($p < 0.04$).

We subsequently analyzed the four variants individually in 100- to 200-bp DNA segments. Of the four variants tested individually, only the element containing rs2144300 exhibited enhanced luciferase activity (average 6.5-fold more than in the empty-vector control). Almost none of these segments showed allelic differences ($p > 0.1$; Figure S7), but rs4846913 did show allelic differences in only the forward orientation ($p = 0.03$). These data suggest that the larger segment is necessary for observing allelic differences in enhancer activity in this assay.

USF1 Binds to rs2281721

To investigate whether transcription factors bind differentially to rs2281721, rs4846913, rs2144300, rs1555290, and rs6143660, we performed EMSAs with HepG2 nuclear lysate. The rs2281721 C probe showed stronger protein binding than the rs2281721 T probe (lane 2 versus 7, arrow, Figure 5). The addition of 40-fold excess unlabeled rs2281721 C probe competed away the signal more effectively than did unlabeled rs2281721 T probe (lane 3 versus 4). rs2281721 overlaps a USF1 ChIP-seq peak from ENCODE data in HepG2 cells, as well as a predicted USF1 motif. In EMSAs, the addition of a USF1 antibody resulted in a disruption of the band observed with the C allele (lane 5). As negative controls, we tested ARNT and SF1 antibodies; we did not observe disruption of the rs2281721 C allele band (Figure S8). These data provide evidence supporting USF1 binding to the C allele of rs2281721.

To validate the USF1 binding in a native chromatin context, we performed ChIP assays in both HepG2 (rs2281721 genotype T/C) and Huh-7 (rs2281721 genotype C/C) cells. The ChIP assays provided evidence supporting USF1 binding to a 164-bp DNA region spanning rs2281721; however, this binding was not allele specific (Figure 5B).

CEBPB Binds Differentially to the Alleles of rs4846913

We also observed more protein binding to the A allele than to the C allele of rs4846913 (lane 2 versus 7, arrow, Figures 6A and S9). We observed a greater decrease in band intensity upon addition of 63-fold excess unlabeled rs4846913 A probe than upon addition of 63-fold excess unlabeled rs4846913 C probe, suggesting that competition of the lane 2 band with unlabeled rs4846913 A probe is more effective than competition with the unlabeled rs4846913 C probe (lane 3 versus 4). rs4846913 overlaps CEBPB ChIP-seq peaks from ENCODE data in HepG2 cells and a predicted CEBPB binding motif. Incubation of the EMSA reactions with a CEBPB antibody generated a strong supershift band for the A allele and a detectable supershift band for the C allele (lane 5 versus 10, Figure 6A). As negative controls, we tested HNF4A or RXRA antibodies; we did not observe evidence of supershifts (Figure S8).

To validate the differential CEBPB binding in a native chromatin context, we performed ChIP assays in both HepG2 (rs4846913 genotype A/C) and Huh-7 (rs4846913 genotype C/C) cells. A 120-bp region of DNA containing rs4846913 showed 31-fold more CEBPB binding in HepG2 cells than in Huh-7 cells ($p = 0.006$; Figure 6B). These results are consistent with the EMSA result of increased CEBPB binding to rs4846913 A.

The CEBPB ChIP-seq signal at rs4846913 was sufficiently strong to permit an analysis of allelic imbalance reflecting in vivo CEBPB binding. We re-aligned HepG2 CEBPB ChIP-seq reads and HepG2 DNase-seq reads by using an allele-aware approach to avoid reference-allele bias (see Material and Methods). Among CEBPB reads spanning rs4846913, 57 of 78 (73%) contained the rs4846913 A allele (binomial $p = 5.6 \times 10^{-5}$; Figure 6C). Among DNase-seq reads spanning rs4846913, 31 of 83 (37%) contained the rs4846913 A allele (binomial $p = 0.03$; Figure 6C), suggesting that CEBPB might help protect the DNA sequence containing the A allele from being accessible to the DNase I enzyme. Taken together, the EMSA, ChIP, and allelic imbalance in ChIP-seq and DNase-seq reads all show consistent evidence suggesting stronger CEBPB binding to the A allele of rs4846913.

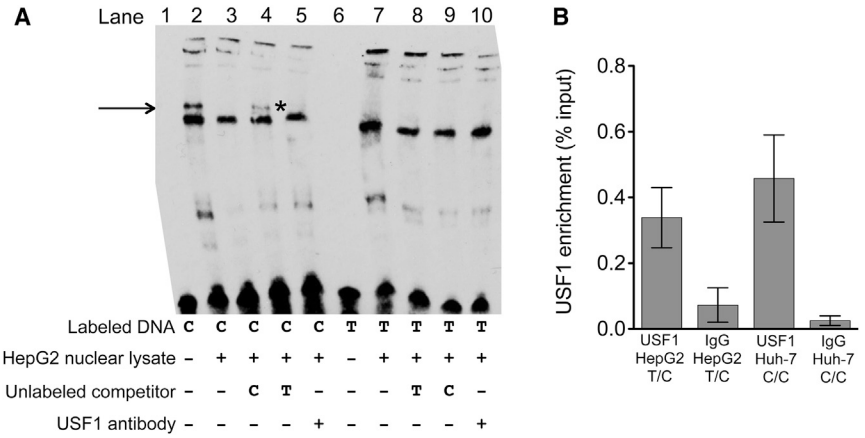

Figure 5. The rs2281721 C Allele Shows Binding to USF1

(A) EMSAs with biotin-labeled probes containing either the T or C allele of rs2281721 and incubated with 5 μg HepG2 nuclear lysate. The arrow indicates an allele-specific band (lanes 2 and 7), and the asterisk indicates the C-allele-specific band that was disrupted upon incubation with USF1 antibody (lane 5). For competition reactions, 40-fold excess unlabeled probe was added.

(B) ChIP experiments were performed in HepG2 cells (T/C at rs2281721) and Huh-7 cells (C/C at rs2281721) with USF1 antibody or rabbit IgG control, and a 164-bp DNA region containing rs2281721 was amplified via qPCR and quantified with a standard curve. Results are shown as percentages of input DNA. Error bars represent the SEM of two independent USF1 and IgG ChIP experiments each in HepG2 and Huh-7 cells.

rs2144300, rs1555290, and rs6143660 Also Show Suggestive Evidence of Protein Binding

EMSAs using HepG2 nuclear lysate for the other variants in the four-variant haplotype (rs2144300, rs1555290, and rs6143660) showed suggestive evidence of differential allelic protein binding (Figures S10–S12). The rs1555290 C probe showed a reproducible band (arrow, Figure S10), and the addition of 30-fold excess unlabeled rs1555290 C probe competed away the signal (lane 7) more effectively than did the addition of 30-fold excess unlabeled rs1555290 A probe (lane 8). The rs6143660 deletion allele showed multiple protein-binding bands (lane 2, Figure S11) that were altered by incubation with FOXO3 antibody (Figure S13), although it is challenging to interpret results when using EMSA probes of different sequence lengths to compare the 21-bp-insertion and 21-bp-deletion alleles. However, we did not observe any disruption of the band corresponding to the rs6143660 deletion allele when we tested additional antibodies CEBPA, CEBPB, HR1H3, MAX, and RXRA (Figure S13). A weak, reproducible, allele-specific band for rs2144300 C was observed, but it was not fully competed away with excess unlabeled rs2144300 C probe (Figure S12). Overall, these EMSA data suggest that these other three variants in the haplotype might exhibit differential transcription-factor binding.

Variants Associated with *GALNT2* Expression

On the basis of the location of the HDL-C-associated variants within *GALNT2* intron 1, we hypothesized that these variants might be acting in *cis* to regulate *GALNT2* expression. We assessed *GALNT2* expression stratified by genotype of rs4846914 ($r^2 = 1$ with rs4846913) in 50 primary human hepatocyte samples, and we observed a trend toward association between genotype and *GALNT2* expression ($p = 0.08$; Figure S14). To increase the sensitivity by removing the sample-specific contribution of environmental factors, we performed AEI assays in 36 human hepatocyte samples heterozygous for rs4846914. Of 142 variants in $r^2 > 0.2$ with rs4846914 in METSIM, none are transcribed, so we used qPCR to measure relative allelic cDNA levels from *GALNT2* pre-mRNA at rs4846914. Previous studies have similarly used intronic SNPs in unspliced RNA to measure allelic expression.[42,43] The rs4846914 A allele, which is associated with increased HDL-C, showed higher *GALNT2* cDNA expression ($p = 5.4 \times 10^{-7}$; Figure 7A). When normalized to the allelic difference detected in heterozygous genomic DNA, the A allele of rs4846914 showed a 7.4% increase in expression. Finally, we examined rs4846914 in preliminary expression-quantitative-trait-locus (eQTL) data from subcutaneous adipose tissue from 1,381 individuals in the METSIM study (M.C. and Y.W., unpublished data) and observed consistent association between the rs4846914 A allele and increased expression of *GALNT2* ($p = 2.2 \times 10^{-14}$; Figure 7B), but not expression of 16 other potential target genes ($p > 0.05/17 = 0.003$ on the basis of 17 tests; Figure S15 and Table S6). Conditioning on the lead eQTL variant, rs4846922, attenuated the association signal with *GALNT2* expression ($p > 0.2$; Figure S15 and Table S7).

Discussion

We identified multiple functional regulatory variants that contribute to the single strong HDL-C GWAS signal at *GALNT2*. Variants rs2281721 and rs4846913, located 2 kb apart, showed strong and consistent allelic and haplotype differences in enhancer activity and exhibited transcription-factor binding of USF1 and CEBPB, respectively. Variants rs2144300 and rs6143660 showed moderate effects on transcriptional activity and suggestive allelic differences in binding of nuclear proteins, although we did not confirm the identity of these potential regulatory proteins. In addition, rs1555290, in moderate LD ($r^2 = 0.26$) with the GWAS variants, also showed suggestive haplotype differences in enhancer activity and allelic differences in binding of a nuclear protein. For all of these variants, the alleles associated with increased HDL-C showed higher transcriptional activity in reporter assays and were

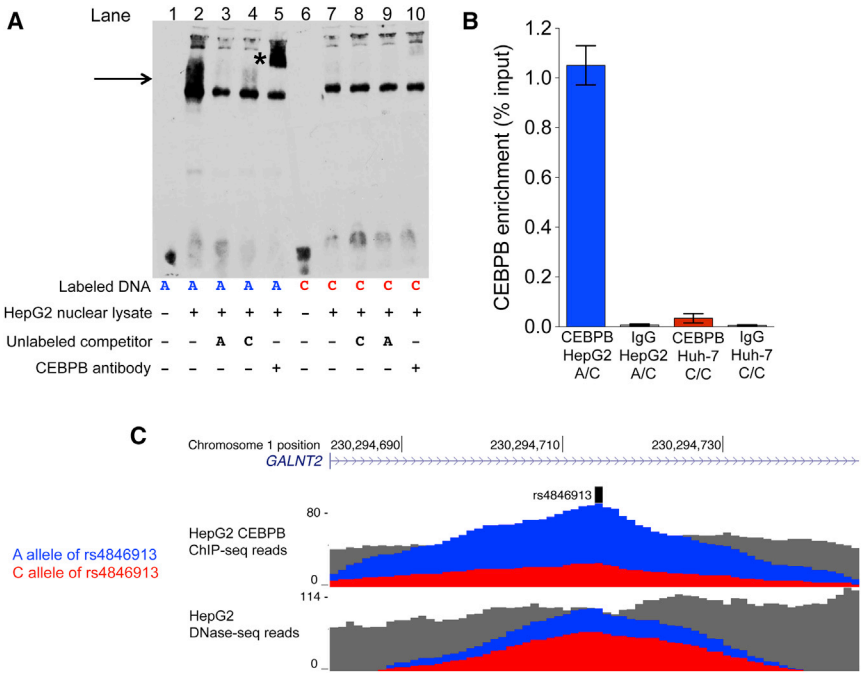

Figure 6. CEBPB Binds Differentially to the Alleles of rs4846913

(A) EMSAs with biotin-labeled probes containing either the A or C allele of rs4846913 and incubated with 6 μg HepG2 nuclear lysate. The arrow indicates increased protein binding to the A allele (lanes 2 versus 7). The asterisk indicates evidence of a CEBPB supershift. For competition reactions, 63-fold excess unlabeled probe was added.

(B) ChIP experiments were performed in HepG2 cells (A/C at rs4846913) and Huh-7 cells (C/C at rs4846913) with CEBPB antibody or rabbit IgG control, and a 120-bp DNA region containing rs4846913 was amplified via qPCR and quantified with a standard curve. Results are shown as percentages of input DNA. Error bars represent the SEM of two independent CEBPB and IgG ChIP experiments each in HepG2 and Huh-7 cells.

(C) CEBPB ChIP-seq reads and DNase-seq reads from ENCODE data in a region containing rs4846913 (UCSC Genome Browser hg19 chromosome coordinates). Blue indicates reads that contain the A allele of rs4846913, red indicates reads that contain the C allele of rs4846913, and gray indicates reads that do not overlap rs4846913 in the peak.

associated with higher expression of *GALNT2* in human hepatocyte and human subcutaneous adipose tissue samples.

By testing all variants in strong LD ($r^2 > 0.7$) with lead HDL-C GWAS variant rs4846914 in luciferase reporter assays in both forward and reverse orientations, we identified the most likely functional regulatory candidates. Although we used a threshold of $r^2 > 0.7$ to identify the most likely functional candidates, we cannot rule out the possibility that we could have missed additional functional variants in weaker LD with the lead HDL-C-associated variant. Of 14 DNA segments tested, we observed that a segment containing four variants and a separate segment containing rs2281721 showed up to 49-fold and 75-fold, respectively, more enhancer activity than did the empty-vector control. One additional element containing one SNP, rs10864726, showed more-moderate allelic differences in enhancer activity (2.5-fold [T allele] versus 1.8-fold [C allele] more than the empty vector, $p = 0.003$) in only the forward orientation (Figure S16). Compared to a vector control, the remaining 11 segments did not exhibit enhancer activity in both orientations, nor did they exhibit significant differences between alleles. These 11 segments included one of the lead GWAS SNPs, rs4846914, and the lead SNP from our fine-mapping analysis, rs17315646. These two SNPs are located in regions containing H3K4me1, H3K4me2, H3K4me3, H3K9ac, and H3K27ac peaks in HepG2 cells and H3K4me1 peaks in adult liver cells, demonstrating that not all GWAS variants located in regulatory elements contribute to regulatory function.

Open-chromatin peaks and histone-modification and transcription-factor ChIP-seq data in liver cell types partially predicted the variants that showed regulatory activity. Variants rs4846913, rs2144300, rs1555290, and rs6143660 overlapped the most peaks (≥22 each) of any candidate variant, but rs2281721 did not overlap any DNase or FAIRE peaks; instead, it overlapped only four histone-modification peaks (H3K4me1, H3K4me2, H3K9ac, and H3K27ac) and one transcription-factor ChIP-seq peak (USF1). We compared our results to the Probabilistic Identification of Causal SNPs (PICS) algorithm's[44] prediction of candidate causal variants in this region. The variants that overlapped the most marks of open chromatin and histone modification in liver cell types (rs4846914, rs10127775, rs4846913, rs2144300, rs17315646, and rs10864726) showed the highest prediction scores, whereas the remaining 19 variants (including rs2281721) showed a prediction score of 0. Thus, testing many candidate variants in experimental assays was necessary for identifying all variants that showed strong allelic differences in enhancer activity. We also conclude that testing all candidate variants in strong LD in combination with regulatory datasets provides the best chance of identifying regulatory variants that exhibit allelic differences in enhancer activity.

One or more of these regulatory variants might also influence gene expression in other cell types. In addition to being present in HepG2 cells, rs4846913, rs2144300, rs1555290, rs6143660, and rs2281721 are also located within predicted ChromHMM[23] strong enhancer chromatin states in K562 leukemia cells from ENCODE data[12] and ChromHMM active enhancer chromatin states in multiple cell types and tissues from Roadmap Epigenomics

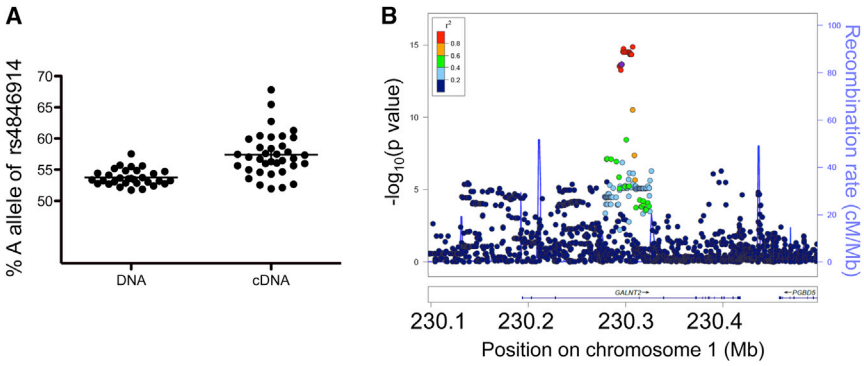

Figure 7. The rs4846914 A Allele Associated with Increased HDL-C Is Associated with Higher *GALNT2* RNA Expression in Primary Human Hepatocytes and Subcutaneous Adipose Tissue

(A) AEI assays were performed in primary human hepatocytes from 36 individuals heterozygous for the intronic HDL-C-associated lead SNP rs4846914 ($p = 5.4 \times 10^{-7}$). rs4846914 was used as a proxy for SNP rs4846913. DNA results from genomic DNA are shown as a control, and cDNA results in this intronic region represent pre-mRNA.

(B) eQTL queries were performed for the noncoding *GALNT2* variants and *GALNT2* expression in subcutaneous adipose tissue samples from 1,381 individuals from the METSIM study. Circles represent genotyped and imputed DNA variants and the LD r^2 values with lead eQTL variant rs4846922. LD was calculated from METSIM genotypes imputed from 1000 Genomes Phase 1 EUR dataset. Chromosome coordinates correspond to UCSC Genome Browser build hg19. The left y axis indicates the $-\log_{10}$(p value), the right y axis indicates the recombination rate (cM/Mb), and the x axis indicates position on chromosome 1 (Mb).

data,[13] including adipose nuclei, CD34 cells, brain, pancreas, and skeletal muscle (Figure S1).

One or more of the 21 candidate variants that did not show haplotype or allelic differences in enhancer activity in both orientations, especially those variants tested individually in smaller segments, might also contribute to the *GALNT2* regulatory mechanism. Testing smaller segments (~100–200 bp) in reporter assays can help with focusing on individual variants; however, these segments might fail to capture full regulatory elements. Testing larger segments can be beneficial because they could encompass a regulatory element, although segment size must be balanced with the fact that larger elements might contain both enhancer and repressor regions and mask modest allelic effects.

The observation that the 780-bp segment showed similar haplotype differences in enhancer activity in both a luciferase vector with a minimal promoter and a promoterless luciferase vector further confirms that rs4846913, rs2144300, rs1555290, and rs6143660 are located within a regulatory element that drives expression. We did not find annotated evidence of an alternative promoter or alternative splicing in this region; however, we cannot rule out alternate mechanisms. Using qRT-PCR and strand-specific primers, we detected hepatocyte and HepG2 RNA transcribed from both strands within *GALNT2* intron 1 (T.R., unpublished data). This result is not unexpected, given that transcription initiation has been found to occur in both directions at both promoters and enhancers.[45]

The functional regulatory variants at this locus might bind multiple transcription factors and act together in a complex to influence transcriptional enhancer activity. Increased binding of CEBPB to the rs4846913 A allele, associated with increased HDL-C, is consistent with the fact that CEBPB acts as a transcriptional activator.[46] CEBPB is a CCAAT/enhancer binding transcription factor that plays diverse roles in cell proliferation, development, adipocyte differentiation, immune response, and liver gene expression.[47–51] Knockdown of *CEBPB* (71% of negative control) by siRNA in HepG2 cells resulted in a modest 22% decrease in *GALNT2* expression ($p = 0.058$; Figure S17), suggesting that other transcription factors within a complex might bind to rs4846913 and be sufficient to preserve substantial enhancer activity and *GALNT2* expression in HepG2 cells. Nonetheless, our data suggest that differential binding of CEBPB to rs4846913 is one potential molecular mechanism underlying the HDL-C association.

The other candidate regulatory variants, rs2281721, rs2144300, rs1555290, and rs6143660, might also bind one or more transcription factors. Binding of USF1 to rs2281721 was allele specific in EMSAs; however, ChIP assays in HepG2 (C/T at rs2281721) and Huh-7 (C/C at rs2281721) cells suggest that USF1 binds to both alleles of rs2281721 in a native chromatin context. Therefore, USF1 might not drive the allelic differences observed in EMSAs and in HepG2 and Huh-7 cell transcriptional reporter assays. USF1 is an upstream stimulatory factor and has been shown to bind to lipid- and glucose-metabolism-related genes[52–54] and affect cholesterol homeostasis, insulin sensitivity, and body-fat mass.[55] The other three potential regulatory variants, rs2144300, rs1555290, and rs6143660, each overlap ≥ 13 transcription-factor ChIP-seq peaks in HepG2 cells. Future experiments will be valuable for confirming the role of CEBPB and USF1 in *GALNT2* expression, identifying other transcription factors contributing to the haplotype differences in transcriptional enhancer activity, detecting physical interactions of this enhancer region, and fully characterizing this complex molecular mechanism at *GALNT2*.

The transcriptional reporter assays all exhibit the same direction of effect; alleles associated with increased HDL-C are also associated with increased enhancer activity. In the assay of liver AEI (Figure 7A) and the study on adipose eQTLs (Figure 7B and Figure S15), the rs4846914 A allele associated with increased HDL-C was associated with higher *GALNT2* expression. Although the HDL-C GWAS variants were not associated with *GALNT2* expression in an eQTL dataset of more than 400 liver samples

($p < 3.95 \times 10^{-8}$)[56] or in our eQTL analysis of primary hepatocyte samples from 50 individuals (Figure S14), significant associations were previously observed[57] in 146 human liver biopsy samples ($p = 0.002$) and 105 carotid-atherosclerotic-plaque biopsy samples ($p = 0.001$), consistent with our observed direction of effect. The evidence of variant association with the level of *GALNT2* expression in both liver and adipose tissue is consistent with the frequent observation that eQTLs are shared across multiple tissues.[58] This direction of effect is opposite to the effect on HDL-C of a previous study that used adeno-associated viral vectors to overexpress and knock down mouse *Galnt2*[2] and a study that observed a rare *GALNT2* missense variant in humans.[21] Notably, our observed direction of effect of the human GWAS variants is consistent with recent unpublished studies demonstrating that total loss of function of GALNT2 in humans, mice, rats, and cynomolgus monkeys consistently results in lower HDL-C (S. Khetarpal, D. Rader, personal communication). *GALNT2* remains a likely candidate gene, considering the association between the GWAS variants and *GALNT2* expression, but not expression of 16 other potential target genes within 1 Mb in samples of subcutaneous adipose tissue. However, the regulatory variants we identified might also act to increase expression of other nearby genes that might contribute to the HDL-C association signal. Identification of the functional variants responsible for the human GWAS signal can provide further understanding of the direction of effect in humans and lead to a greater insight into the molecular mechanisms of how these variants might influence gene expression and function.

In the METSIM cohort, the studied variants were most strongly associated with traits correlated with total cholesterol in medium HDL, including cholesterol esters in medium HDL ($r^2 = 1.0$), concentrations of medium HDL particles (>0.9), phospholipids in medium HDL (>0.9), and free cholesterol in medium HDL (>0.9). Sub-phenotype associations can provide additional insight into the function of GWAS loci.[59] GALNT2 might act during specific steps of HDL-particle formation or remodeling by directly glycosylating lipid modifiers or enzymes, such as ANGPTL3[20] or LIPG, an enzyme that has been shown to hydrolyze phospholipids in HDL[60,61] and remodel HDL particles.[62] More work is necessary for determining the mechanism(s) by which GALNT2 might influence the size of HDL particles.

Overall, this study demonstrates multiple lines of evidence suggesting that at least two regulatory variants might act to regulate expression of *GALNT2*, a gene involved in HDL-C metabolism (Figure S18). The eQTL and AEI data clarify the direction of effect by which the GWAS variants act. Our study joins a growing set of studies that implicate multiple functional regulatory variants at a GWAS locus,[63,64] highlighting the complexity of molecular mechanisms underlying GWAS loci, and emphasizes that multiple common functional regulatory variants might work in concert.

Supplemental Data

Supplemental Data include 18 figures and 7 tables and can be found with this article online at http://dx.doi.org/10.1016/j.ajhg.2015.10.016.

Conflicts of Interest

A.J.K., P.S., and M.A.-K. are shareholders of Brainshake Ltd., which offers NMR-based metabolite profiling.

Acknowledgments

This study was supported by NIH grants T32HL069768 (T.S.R.), T32GM007092 (T.S.R.), R01DK072193 (K.L.M.), R21DA027040 (K.L.M.), R01DK093757 (K.L.M.), T32GM067553 (M.L.B), R01DK062370 (M.B.), K99HL121172 (M.C.), and P01HL28481 (A.J.L.) and National Human Genome Research Institute Division of Intramural Research project number Z01HG000024 (F.S.C.); American Heart Association Predoctoral Fellowship 13PRE16930025 (M.L.B.); Academy of Finland grants 77299 and 124243 (M.L.); the Finnish Heart Foundation (M.L.); the Finnish Diabetes Foundation (M.L.); Finnish Funding Agency for Technology and Innovation (TEKES) contract 1510/31/06 (M.L.); and Commission of the European Community HEALTH-F2-2007-201681 (M.L.). Applications and development of the quantitative-serum nuclear-magnetic-resonance (NMR) metabolomics platform are supported by the Academy of Finland, TEKES, Sigrid Juselius Foundation, Novo Nordisk Foundation, and Finnish Diabetes Research Foundation. A.J.K., P.S., and M.A.-K. are supported by the University of Oulu, British Heart Foundation, Wellcome Trust, and Medical Research Council. Bristol Myers Squibb supported the generation of METSIM microarray data. The authors acknowledge the GoT2D Consortium for access to the reference panel used for imputation; human organ donors whose liver samples were used in this study; the ENCODE and NIH Roadmap Epigenomics Consortia, Data Analysis and Coordination Centers, and production laboratories that generated the chromatin, histone-modification, and chromatin-immunoprecipitation (ChIP) sequencing data used for variant annotation. We thank Terry Furey and Greg Crawford for interpretating ENCODE data, Jennifer Kulzer for editing the manuscript, Yun Li and Qing Duan for providing 1000 Genomes linkage-disequilibrium data, Scott Bultman and Dallas Donohoe for providing advice on ChIP experiments, and Sumeet Khetarpal and Daniel Rader for sharing unpublished results.

Received: May 4, 2015
Accepted: October 28, 2015
Published: December 3, 2015

Web Resources

The URLs for data presented herein are as follows:

Human Epigenome Atlas, http://www.genboree.org/epigenomeatlas/index.rhtml
OMIM, http://www.omim.org
PWM-SCAN, http://www.cbcb.umd.edu/~sridhar/software.html
Roadmap Epigenomics Project, http://www.roadmapepigenomics.org/data/
UCSC Genome Browser, http://genome.ucsc.edu

References

1. Kuivenhoven, J.A., and Hegele, R.A. (2014). Mining the genome for lipid genes. Biochim. Biophys. Acta *1842*, 1993–2009.
2. Teslovich, T.M., Musunuru, K., Smith, A.V., Edmondson, A.C., Stylianou, I.M., Koseki, M., Pirruccello, J.P., Ripatti, S., Chasman, D.I., Willer, C.J., et al. (2010). Biological, clinical and population relevance of 95 loci for blood lipids. Nature *466*, 707–713.
3. Dumitrescu, L., Carty, C.L., Taylor, K., Schumacher, F.R., Hindorff, L.A., Ambite, J.L., Anderson, G., Best, L.G., Brown-Gentry, K., Bůžková, P., et al. (2011). Genetic determinants of lipid traits in diverse populations from the population architecture using genomics and epidemiology (PAGE) study. PLoS Genet. *7*, e1002138.
4. Willer, C.J., Schmidt, E.M., Sengupta, S., Peloso, G.M., Gustafsson, S., Kanoni, S., Ganna, A., Chen, J., Buchkovich, M.L., Mora, S., et al.; Global Lipids Genetics Consortium (2013). Discovery and refinement of loci associated with lipid levels. Nat. Genet. *45*, 1274–1283.
5. Weissglas-Volkov, D., Aguilar-Salinas, C.A., Nikkola, E., Deere, K.A., Cruz-Bautista, I., Arellano-Campos, O., Muñoz-Hernandez, L.L., Gomez-Munguia, L., Ordoñez-Sánchez, M.L., Reddy, P.M., et al. (2013). Genomic study in Mexicans identifies a new locus for triglycerides and refines European lipid loci. J. Med. Genet. *50*, 298–308.
6. Willer, C.J., Sanna, S., Jackson, A.U., Scuteri, A., Bonnycastle, L.L., Clarke, R., Heath, S.C., Timpson, N.J., Najjar, S.S., Stringham, H.M., et al. (2008). Newly identified loci that influence lipid concentrations and risk of coronary artery disease. Nat. Genet. *40*, 161–169.
7. Kathiresan, S., Melander, O., Guiducci, C., Surti, A., Burtt, N.P., Rieder, M.J., Cooper, G.M., Roos, C., Voight, B.F., Havulinna, A.S., et al. (2008). Six new loci associated with blood low-density lipoprotein cholesterol, high-density lipoprotein cholesterol or triglycerides in humans. Nat. Genet. *40*, 189–197.
8. Nakayama, K., Bayasgalan, T., Yamanaka, K., Kumada, M., Gotoh, T., Utsumi, N., Yanagisawa, Y., Okayama, M., Kajii, E., Ishibashi, S., and Iwamoto, S.; Jichi Community Genetics Team (JCOG) (2009). Large scale replication analysis of loci associated with lipid concentrations in a Japanese population. J. Med. Genet. *46*, 370–374.
9. Weissglas-Volkov, D., Aguilar-Salinas, C.A., Sinsheimer, J.S., Riba, L., Huertas-Vazquez, A., Ordoñez-Sánchez, M.L., Rodriguez-Guillen, R., Cantor, R.M., Tusie-Luna, T., and Pajukanta, P. (2010). Investigation of variants identified in caucasian genome-wide association studies for plasma high-density lipoprotein cholesterol and triglycerides levels in Mexican dyslipidemic study samples. Circ Cardiovasc Genet *3*, 31–38.
10. Kathiresan, S., Willer, C.J., Peloso, G.M., Demissie, S., Musunuru, K., Schadt, E.E., Kaplan, L., Bennett, D., Li, Y., Tanaka, T., et al. (2009). Common variants at 30 loci contribute to polygenic dyslipidemia. Nat. Genet. *41*, 56–65.
11. Abecasis, G.R., Auton, A., Brooks, L.D., DePristo, M.A., Durbin, R.M., Handsaker, R.E., Kang, H.M., Marth, G.T., and McVean, G.A.; 1000 Genomes Project Consortium (2012). An integrated map of genetic variation from 1,092 human genomes. Nature *491*, 56–65.
12. ENCODE Project Consortium (2012). An integrated encyclopedia of DNA elements in the human genome. Nature *489*, 57–74.
13. Kundaje, A., Meuleman, W., Ernst, J., Bilenky, M., Yen, A., Heravi-Moussavi, A., Kheradpour, P., Zhang, Z., Wang, J., Ziller, M.J., et al.; Roadmap Epigenomics Consortium (2015). Integrative analysis of 111 reference human epigenomes. Nature *518*, 317–330.
14. Visscher, P.M., Brown, M.A., McCarthy, M.I., and Yang, J. (2012). Five years of GWAS discovery. Am. J. Hum. Genet. *90*, 7–24.
15. Manolio, T.A. (2013). Bringing genome-wide association findings into clinical use. Nat. Rev. Genet. *14*, 549–558.
16. Ten Hagen, K.G., Fritz, T.A., and Tabak, L.A. (2003). All in the family: the UDP-GalNAc:polypeptide N-acetylgalactosaminyltransferases. Glycobiology *13*, 1R–16R.
17. White, T., Bennett, E.P., Takio, K., Sørensen, T., Bonding, N., and Clausen, H. (1995). Purification and cDNA cloning of a human UDP-N-acetyl-alpha-D-galactosamine:polypeptide N-acetylgalactosaminyltransferase. J. Biol. Chem. *270*, 24156–24165.
18. Bennett, E.P., Hassan, H., and Clausen, H. (1996). cDNA cloning and expression of a novel human UDP-N-acetyl-alpha-D-galactosamine. Polypeptide N-acetylgalactosaminyltransferase, GalNAc-t3. J. Biol. Chem. *271*, 17006–17012.
19. Bennett, E.P., Mandel, U., Clausen, H., Gerken, T.A., Fritz, T.A., and Tabak, L.A. (2012). Control of mucin-type O-glycosylation: a classification of the polypeptide GalNAc-transferase gene family. Glycobiology *22*, 736–756.
20. Schjoldager, K.T., Vester-Christensen, M.B., Bennett, E.P., Levery, S.B., Schwientek, T., Yin, W., Blixt, O., and Clausen, H. (2010). O-glycosylation modulates proprotein convertase activation of angiopoietin-like protein 3: possible role of polypeptide GalNAc-transferase-2 in regulation of concentrations of plasma lipids. J. Biol. Chem. *285*, 36293–36303.
21. Holleboom, A.G., Karlsson, H., Lin, R.S., Beres, T.M., Sierts, J.A., Herman, D.S., Stroes, E.S., Aerts, J.M., Kastelein, J.J., Motazacker, M.M., et al. (2011). Heterozygosity for a loss-of-function mutation in GALNT2 improves plasma triglyceride clearance in man. Cell Metab. *14*, 811–818.
22. Stančáková, A., Javorský, M., Kuulasmaa, T., Haffner, S.M., Kuusisto, J., and Laakso, M. (2009). Changes in insulin sensitivity and insulin release in relation to glycemia and glucose tolerance in 6,414 Finnish men. Diabetes *58*, 1212–1221.
23. Ernst, J., Kheradpour, P., Mikkelsen, T.S., Shoresh, N., Ward, L.D., Epstein, C.B., Zhang, X., Wang, L., Issner, R., Coyne, M., et al. (2011). Mapping and analysis of chromatin state dynamics in nine human cell types. Nature *473*, 43–49.
24. Huyghe, J.R., Jackson, A.U., Fogarty, M.P., Buchkovich, M.L., Stančáková, A., Stringham, H.M., Sim, X., Yang, L., Fuchsberger, C., Cederberg, H., et al. (2013). Exome array analysis identifies new loci and low-frequency variants influencing insulin processing and secretion. Nat. Genet. *45*, 197–201.
25. Jun, G., Flickinger, M., Hetrick, K.N., Romm, J.M., Doheny, K.F., Abecasis, G.R., Boehnke, M., and Kang, H.M. (2012). Detecting and estimating contamination of human DNA samples in sequencing and array-based genotype data. Am. J. Hum. Genet. *91*, 839–848.
26. Delaneau, O., Zagury, J.F., and Marchini, J. (2013). Improved whole-chromosome phasing for disease and population genetic studies. Nat. Methods *10*, 5–6.
27. Howie, B., Fuchsberger, C., Stephens, M., Marchini, J., and Abecasis, G.R. (2012). Fast and accurate genotype imputation in genome-wide association studies through pre-phasing. Nat. Genet. *44*, 955–959.

28. Soininen, P., Kangas, A.J., Würtz, P., Tukiainen, T., Tynkkynen, T., Laatikainen, R., Järvelin, M.R., Kähönen, M., Lehtimäki, T., Viikari, J., et al. (2009). High-throughput serum NMR metabonomics for cost-effective holistic studies on systemic metabolism. Analyst (Lond.) *134*, 1781–1785.
29. Soininen, P., Kangas, A.J., Würtz, P., Suna, T., and Ala-Korpela, M. (2015). Quantitative serum nuclear magnetic resonance metabolomics in cardiovascular epidemiology and genetics. Circ Cardiovasc Genet *8*, 192–206.
30. Inouye, M., Kettunen, J., Soininen, P., Silander, K., Ripatti, S., Kumpula, L.S., Hämäläinen, E., Jousilahti, P., Kangas, A.J., Männistö, S., et al. (2010). Metabonomic, transcriptomic, and genomic variation of a population cohort. Mol. Syst. Biol. *6*, 441.
31. Kang, H.M., Sul, J.H., Service, S.K., Zaitlen, N.A., Kong, S.Y., Freimer, N.B., Sabatti, C., and Eskin, E. (2010). Variance component model to account for sample structure in genome-wide association studies. Nat. Genet. *42*, 348–354.
32. Pruim, R.J., Welch, R.P., Sanna, S., Teslovich, T.M., Chines, P.S., Gliedt, T.P., Boehnke, M., Abecasis, G.R., and Willer, C.J. (2010). LocusZoom: regional visualization of genome-wide association scan results. Bioinformatics *26*, 2336–2337.
33. Kulzer, J.R., Stitzel, M.L., Morken, M.A., Huyghe, J.R., Fuchsberger, C., Kuusisto, J., Laakso, M., Boehnke, M., Collins, F.S., and Mohlke, K.L. (2014). A common functional regulatory variant at a type 2 diabetes locus upregulates ARAP1 expression in the pancreatic beta cell. Am. J. Hum. Genet. *94*, 186–197.
34. Sandelin, A., Alkema, W., Engström, P., Wasserman, W.W., and Lenhard, B. (2004). JASPAR: an open-access database for eukaryotic transcription factor binding profiles. Nucleic Acids Res. *32*, D91–D94.
35. Levy, S., and Hannenhalli, S. (2002). Identification of transcription factor binding sites in the human genome sequence. Mamm. Genome *13*, 510–514.
36. Buchkovich, M.L., Eklund, K., Duan, Q., Li, Y., Mohlke, K.L., and Furey, T.S. (2015). Removing reference mapping biases using limited or no genotype data identifies allelic differences in protein binding at disease-associated loci. BMC Med. Genomics *8*, 43.
37. Wu, T.D., and Nacu, S. (2010). Fast and SNP-tolerant detection of complex variants and splicing in short reads. Bioinformatics *26*, 873–881.
38. Fogarty, M.P., Xiao, R., Prokunina-Olsson, L., Scott, L.J., and Mohlke, K.L. (2010). Allelic expression imbalance at high-density lipoprotein cholesterol locus MMAB-MVK. Hum. Mol. Genet. *19*, 1921–1929.
39. Xiao, R., and Scott, L.J. (2011). Detection of cis-acting regulatory SNPs using allelic expression data. Genet. Epidemiol. *35*, 515–525.
40. Stegle, O., Parts, L., Durbin, R., and Winn, J. (2010). A Bayesian framework to account for complex non-genetic factors in gene expression levels greatly increases power in eQTL studies. PLoS Comput. Biol. *6*, e1000770.
41. Bailey, T., Krajewski, P., Ladunga, I., Lefebvre, C., Li, Q., Liu, T., Madrigal, P., Taslim, C., and Zhang, J. (2013). Practical guidelines for the comprehensive analysis of ChIP-seq data. PLoS Comput. Biol. *9*, e1003326.
42. Verlaan, D.J., Ge, B., Grundberg, E., Hoberman, R., Lam, K.C., Koka, V., Dias, J., Gurd, S., Martin, N.W., Mallmin, H., et al. (2009). Targeted screening of cis-regulatory variation in human haplotypes. Genome Res. *19*, 118–127.
43. Qu, H.Q., Verlaan, D.J., Ge, B., Lu, Y., Lam, K.C., Grabs, R., Harmsen, E., Hudson, T.J., Hakonarson, H., Pastinen, T., and Polychronakos, C. (2009). A cis-acting regulatory variant in the IL2RA locus. J. Immunol. *183*, 5158–5162.
44. Farh, K.K., Marson, A., Zhu, J., Kleinewietfeld, M., Housley, W.J., Beik, S., Shoresh, N., Whitton, H., Ryan, R.J., Shishkin, A.A., et al. (2015). Genetic and epigenetic fine mapping of causal autoimmune disease variants. Nature *518*, 337–343.
45. Core, L.J., Martins, A.L., Danko, C.G., Waters, C.T., Siepel, A., and Lis, J.T. (2014). Analysis of nascent RNA identifies a unified architecture of initiation regions at mammalian promoters and enhancers. Nat. Genet. *46*, 1311–1320.
46. Cui, T.X., Lin, G., LaPensee, C.R., Calinescu, A.A., Rathore, M., Streeter, C., Piwien-Pilipuk, G., Lanning, N., Jin, H., Carter-Su, C., et al. (2011). C/EBPβ mediates growth hormone-regulated expression of multiple target genes. Mol. Endocrinol. *25*, 681–693.
47. Tanaka, T., Yoshida, N., Kishimoto, T., and Akira, S. (1997). Defective adipocyte differentiation in mice lacking the C/EBP-beta and/or C/EBPdelta gene. EMBO J. *16*, 7432–7443.
48. Sterneck, E., Tessarollo, L., and Johnson, P.F. (1997). An essential role for C/EBPbeta in female reproduction. Genes Dev. *11*, 2153–2162.
49. Robinson, G.W., Johnson, P.F., Hennighausen, L., and Sterneck, E. (1998). The C/EBPbeta transcription factor regulates epithelial cell proliferation and differentiation in the mammary gland. Genes Dev. *12*, 1907–1916.
50. Huber, R., Pietsch, D., Panterodt, T., and Brand, K. (2012). Regulation of C/EBPβ and resulting functions in cells of the monocytic lineage. Cell. Signal. *24*, 1287–1296.
51. Rahman, S.M., Choudhury, M., Janssen, R.C., Baquero, K.C., Miyazaki, M., and Friedman, J.E. (2013). CCAAT/enhancer binding protein β deletion increases mitochondrial function and protects mice from LXR-induced hepatic steatosis. Biochem. Biophys. Res. Commun. *430*, 336–339.
52. Botma, G.J., Verhoeven, A.J., and Jansen, H. (2001). Hepatic lipase promoter activity is reduced by the C-480T and G-216A substitutions present in the common LIPC gene variant, and is increased by Upstream Stimulatory Factor. Atherosclerosis *154*, 625–632.
53. Weigert, C., Brodbeck, K., Sawadogo, M., Häring, H.U., and Schleicher, E.D. (2004). Upstream stimulatory factor (USF) proteins induce human TGF-beta1 gene activation via the glucose-response element-1013/-1002 in mesangial cells: up-regulation of USF activity by the hexosamine biosynthetic pathway. J. Biol. Chem. *279*, 15908–15915.
54. Kim, J.W., Monila, H., Pandey, A., and Lane, M.D. (2007). Upstream stimulatory factors regulate the C/EBP alpha gene during differentiation of 3T3-L1 preadipocytes. Biochem. Biophys. Res. Commun. *354*, 517–521.
55. Wu, S., Mar-Heyming, R., Dugum, E.Z., Kolaitis, N.A., Qi, H., Pajukanta, P., Castellani, L.W., Lusis, A.J., and Drake, T.A. (2010). Upstream transcription factor 1 influences plasma lipid and metabolic traits in mice. Hum. Mol. Genet. *19*, 597–608.
56. Schadt, E.E., Molony, C., Chudin, E., Hao, K., Yang, X., Lum, P.Y., Kasarskis, A., Zhang, B., Wang, S., Suver, C., et al. (2008). Mapping the genetic architecture of gene expression in human liver. PLoS Biol. *6*, e107.
57. Folkersen, L., van't Hooft, F., Chernogubova, E., Agardh, H.E., Hansson, G.K., Hedin, U., Liska, J., Syvänen, A.C., Paulsson-Berne, G., Franco-Cereceda, A., et al.; BiKE and ASAP study groups (2010). Association of genetic risk variants with expression of proximal genes identifies novel susceptibility genes for cardiovascular disease. Circ Cardiovasc Genet *3*, 365–373.

58. GTEx Consortium (2015). Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. Science *348*, 648–660.
59. Kettunen, J., Tukiainen, T., Sarin, A.P., Ortega-Alonso, A., Tikkanen, E., Lyytikäinen, L.P., Kangas, A.J., Soininen, P., Würtz, P., Silander, K., et al. (2012). Genome-wide association study identifies multiple loci influencing human serum metabolite levels. Nat. Genet. *44*, 269–276.
60. Jaye, M., Lynch, K.J., Krawiec, J., Marchadier, D., Maugeais, C., Doan, K., South, V., Amin, D., Perrone, M., and Rader, D.J. (1999). A novel endothelial-derived lipase that modulates HDL metabolism. Nat. Genet. *21*, 424–428.
61. McCoy, M.G., Sun, G.S., Marchadier, D., Maugeais, C., Glick, J.M., and Rader, D.J. (2002). Characterization of the lipolytic activity of endothelial lipase. J. Lipid Res. *43*, 921–929.
62. Jahangiri, A., Rader, D.J., Marchadier, D., Curtiss, L.K., Bonnet, D.J., and Rye, K.A. (2005). Evidence that endothelial lipase remodels high density lipoproteins without mediating the dissociation of apolipoprotein A-I. J. Lipid Res. *46*, 896–903.
63. Corradin, O., Saiakhova, A., Akhtar-Zaidi, B., Myeroff, L., Willis, J., Cowper-Sal lari, R., Lupien, M., Markowitz, S., and Scacheri, P.C. (2014). Combinatorial effects of multiple enhancer variants in linkage disequilibrium dictate levels of gene expression to confer susceptibility to common traits. Genome Res. *24*, 1–13.
64. He, H., Li, W., Liyanarachchi, S., Srinivas, M., Wang, Y., Akagi, K., Wang, Y., Wu, D., Wang, Q., Jin, V., et al. (2015). Multiple functional variants in long-range enhancer elements contribute to the risk of SNP rs965513 in thyroid cancer. Proc. Natl. Acad. Sci. USA *112*, 6128–6133.