Enhancer deletion and allelic effects define a regulatory molecular mechanism at the *VLDLR* cholesterol GWAS locus

James P. Davis, Swarooparani Vadlamudi, Tamara S. Roman, Monica Zeynalzadeh, Apoorva K. Iyengar and Karen L. Mohlke*

Department of Genetics, University of North Carolina, Chapel Hill, NC 27599, USA

*To whom correspondence should be addressed at: Department of Genetics, University of North Carolina, 120 Mason Farm Road, Chapel Hill, NC 27599-7264, USA. Tel: +919 9662913; Fax: +919 8430291; Email: mohlke@med.unc.edu

Abstract

Total cholesterol (TC) and low-density lipoprotein cholesterol (LDL-C) are heritable risk factors for cardiovascular disease, yet the molecular mechanisms underlying the majority of blood lipid-associated genome-wide association studies signals remain elusive. One association signal is located in intron 3 of *VLDLR*; rs3780181-A is a risk allele associated ($P \leq 2 \times 10^{-9}$) with increased TC and LDL-C. We investigated variants, genes and mechanisms underlying this association signal. We used a functional genetic approach to show that the intronic region spanning rs3780181 exhibited 1.6–7.6-fold enhancer activity in human HepG2 hepatocyte, THP-1 monocyte and Simpson-Golabi-Behmel Syndrome (SGBS) preadipocyte cells and that the rs3780181-A risk allele showed significantly less enhancer activity compared with the G allele, consistent with the direction of an expression quantitative trait locus in liver. In addition, rs3780181 alleles showed differential binding to multiple nuclear proteins, including stronger IRF2 binding to the rs3780181 G allele. We used a CRISPR-cas9 approach to delete 475 and 663 bp of the putative enhancer element in HEK293T kidney cells; compared to expression of mock-edited cell lines, the homozygous enhancer deletion cell lines showed 1.2-fold significantly ($P < 0.04$) decreased expression of *VLDLR*, as well as 1.5-fold decreased expression of *SMARCA2*, located 388 kb away. Together, these results identify an enhancer of *VLDLR* expression and suggest that altered binding of one or more factors bound to rs3780181 alleles decreases enhancer activity and reduces at least *VLDLR* expression, leading to increased TC and LDL-C.

Introduction

Low-density lipoprotein cholesterol (LDL-C) and total cholesterol (TC) are heritable risk factors for cardiovascular disease (1,2), and genome-wide association studies (GWASs) for lipid traits have identified association signals at least 240 loci (3,4). At one locus identified in Europeans, non-coding variant rs3780181 A allele (Minor allele frequency, MAF = 0.05), located in *VLDLR* intron 3, is significantly associated with increased LDL-C ($P = 2.0 \times 10^{-9}$; $n = 172\,000$; Supplementary Material, Fig. S1) and TC ($P = 7.0 \times 10^{-10}$, $n = 186\,000$) (5). rs3780181 showed the same direction of association with LDL-C in East Asians (MAF = 0.08) and a meta-analysis of European and East Asian data increased the evidence for association ($P = 5.1 \times 10^{-10}$, $n = 203\,746$) (6). In a Chinese population association study, rs3780181 shared the same direction of effect, and was also associated ($P < 0.05$) with increased TC and LDL-C (7). Among GWAS data for 68 lipoprotein subclasses, rs3780181 showed the strongest association with concentration of chylomicrons and largest VLDL particles ($P = 0.05$) and LDL diameter ($P = 0.09$) (8). This association signal

is independent ($r^2 < 0.05$) from other variants within 1 Mb of the *VLDLR* gene that have been associated with circulating plasma levels of vascular endothelial growth factor (VEGF), interleukin levels and systolic blood pressure (9,10).

Association studies do not identify the underlying molecular mechanisms, including the causal genes and variants, that can be used to understand biological processes and detect or validate therapeutic targets. For the majority of associated variants located in non-coding intergenic and intronic regions, risk variants may affect one or more genes located up to hundreds of kilobases (kb) away (11–13). Two strategies to link non-coding GWAS association signals to candidate genes are identification of a colocalized association of the same variants with expression level of a gene (expression quantitative trait locus, eQTL) and observation of more physical interaction/contact between the regulatory variants and the promoter of a target gene than expected by chance (14,15). These association-based methods are valuable to suggest plausible candidate genes but do not show that the associated variants have a causal effect on expression.

A more rigorous conclusion about the target gene(s) of a GWAS signal requires identifying a mechanistic link between a causal variant and an increased or decreased gene expression or function. Evidence for a causal role of a non-coding variant includes allelic differences in transcription factor binding and regulatory activity of gene expression (16,17). Causal connections between a regulatory variant and gene expression detected in extrachromosomal assays can be validated by assaying regulatory effects in the context of the genome, such as through the use of genome editing (18,19).

In this study, we conducted a systematic analysis of the variants and genes at the LDL-C and TC signal at the *VLDLR* GWAS locus. We identify allelic differences in transcription factor binding and transcriptional activity in three cell types and show that these differences are consistent with associations observed in human liver. We demonstrate that deletion of the intronic enhancer containing the risk allele affects *VLDLR* expression. Together these data provide a molecular mechanism for this GWAS locus.

Results

Characteristics of candidate genes

To identify candidate genes at the rs3780181 signal, we examined gene expression patterns, eQTL and physical interactions in available data sets. The region within 500 kb of rs3780181 includes four protein-coding genes (*SMARCA2*, *VLDLR*, *KCNV2* and *KIAA0020-PUM3*; Supplementary Material, Fig. S1). *VLDLR* and *SMARCA2* are expressed across multiple tissues including liver and adipose, while *KCNV2* is primarily expressed in testes (Supplementary Material, Fig. S2). While rs3780181 did not show association with expression level of any gene in Genotype-Tissue Expression (GTEx) v6 ($P < 0.05$), rs3780181 was the lead variant associated with *VLDLR* expression level in two human liver eQTL data sets of 966 ($P = 1.4 \times 10^{-7}$) and 225 ($P = 2.0 \times 10^{-7}$) individuals (20,21). In both data sets, the rs3780181-A risk allele was associated with less *VLDLR* expression. The study by Giambartolomei *et al.* (20) further demonstrated that the GWAS and eQTL signals were colocalized using COLOC (posterior probability of colocalization, 91%). Neither of the liver eQTL data sets showed suggestive evidence of association with other genes within 500 kb. Chromatin interaction data from mesenchymal stem cells, human embryonic stem cells and human liver tissue showed the most significant interactions ($P < 2.6 \times 10^{-13}$) between the region containing rs3780181 and the promoter regions of *VLDLR* and *KCNV2* (Supplementary Material, Fig. S3). These data suggest that expression of *VLDLR* and/or *KCNV2* may be affected by the GWAS variant(s).

Characterization and functional annotation of GWAS variants

To identify the candidate variants at the GWAS signal, we evaluated variant linkage disequilibrium (LD) with rs3780181, which is the lead GWAS variant for TC and LDL-C and is located in intron 3 of *VLDLR* (Supplementary Material, Fig. S4). rs3780181 is a common variant (minor allele frequency = 0.05) and not in high LD ($r^2 \geq 0.7$) with any other variants in Europeans (Supplementary Material, Fig. S1), suggesting that rs3780181 is the strongest candidate functional variant for the GWAS signal.

To identify appropriate cell types to consider for regulatory activity, we evaluated epigenomic evidence of potential regulatory activity across the full set of available cell and tissue types. Based on Encyclopedia of DNA Elements (ENCODE), Roadmap and Regulome database (RegulomeDB) data sets, rs3780181 is located in a predicted enhancer region in HepG2 liver carcinoma cells, human liver, ovary and fetal adrenal gland tissues (Supplementary Material, Fig. S4) (22). rs3780181 also overlaps a region of accessible chromatin in HepG2 and Huh7 cells, and the genomic sequence around the variant shows sequence similarity to a consensus DNA-binding motif for HMG, RXRA, Pax4, Pax5, NKX3.2, STAT, BCL, Isgf3g, PRDM1 and multiple IRF proteins including IRF1, IRF2 and IRF7 (Supplementary Material, Fig. S5).

rs37801810 exhibits allelic differences in enhancer activity

Based on the eQTL for *VLDLR* in human liver (20,21), its role in adipocyte–macrophage interactions and inflammation (23) and variant annotations, we evaluated the allelic effects of rs3780181 using transcriptional reporter assays in HepG2 hepatocytes, THP-1 monocytes and Simpson-Golabi-Behmel Syndrome (SGBS) preadipocyte cells (Fig. 1A–C). A 223 bp element containing the G allele showed enhancer activity as high as 4.1-fold in HepG2, 7.6-fold in THP-1 and 1.7-fold in SGBS. In all three cell types and in both orientations with respect to a minimal promoter, elements containing the risk A allele displayed less transcriptional activity than elements containing the risk G allele. The alleles showed the largest magnitude of differences in THP-1 cells (Fig. 1A). In the forward orientation, the risk A allele had 2.1-fold higher enhancer activity and the G allele had 5.1-fold higher enhancer activity than the empty vector ($P = 0.02$), and in the reverse orientation, the risk A allele had 4.6-fold higher and the G allele had 6.6-fold higher activity than the empty vector ($P = 0.1$). These data demonstrate that rs3780181 is present in an enhancer element and that the non-risk G allele has significantly higher enhancer activity than the risk A allele, consistent with the direction of the rs3780181 association with *VLDLR* expression level in the liver eQTL studies.

IRF2 and other proteins are differentially bound to rs3780181 alleles

To test whether rs3780181 affects transcription factor binding, we performed electrophoretic mobility shift assays (EMSAs) using THP-1 and HepG2 nuclear protein extract. With THP-

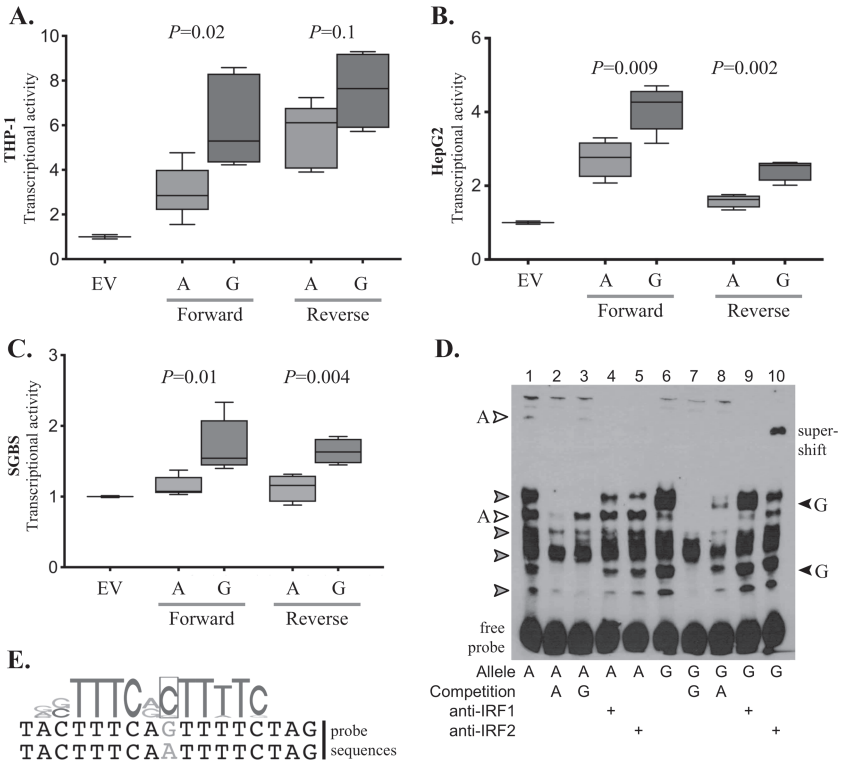

Figure 1. GWAS lead variant rs3780181 displays allelic differences in transcriptional activity and protein binding. A 222 bp element containing the A allele showed less enhancer activity when compared to the G allele in **(A)** THP-1, **(B)** HepG2 and **(C)** SGBS cells. We tested 3–5 clones per allele in both orientations with respect to the minimal promoter and normalized activity to the EV. **(D)** EMSA with THP-1 NE and oligonucleotide probes centered on rs3780181 revealed at least two protein complexes specific to the G allele (black arrows, lane 6) and two specific to the A allele (white arrows, lane 1). Protein complexes non-specifically bound to both allele probes are denoted by gray arrows. The addition of anti-IRF2 antibody (lanes 5 and 10) caused a supershift of the larger G allele-specific band (lane 10). **(E)** The motif for IRF-2 shows homology to the DNA region around rs3780181. The variant location is indicated by a rectangle and the variant alleles are shown in the probe sequences below.

1 extract, we resolved four protein–probe complexes (Fig. 1D, black and white arrows) that were differentially bound to the alleles (lanes 1 and 6). The two proteins differentially bound to the G allele were reduced more by including excess (35×) unlabeled competitor G-allele probe than competitor A-allele probe, providing further evidence of allele specificity. Similarly, one protein differentially bound to the A allele was reduced more by including excess unlabeled competitor A-allele probe. EMSAs with HepG2 nuclear extract (NE) similarly showed two protein complexes differentially bound to each allele, although differences exist, especially for proteins bound to both alleles (Supplementary Material, Fig. S6). To identify the factors bound to the rs3780181 alleles, we conducted a DNA affinity pull-down assay by using rs3780181 probes and THP-1 nuclear extract. Among the transcription factors identified by the pull-down assay (Supplementary Material, Table S1), the IRF2 consensus-binding site motif matched the genomic sequence surrounding rs3780181 best, and while the motif includes a C allele most often at the site of the variant, G is observed more frequently than A in the motif (Fig. 1E). Supershift assays with antibodies to IRF2 and nuclear extracts from THP-1 and HepG2 confirmed that IRF2 was specifically bound to the G allele (Fig. 1D, lane 10, and Supplementary Material, Fig. S6, lane 10). Antibodies to five other transcription factors suggested by ENCODE and TRANScription FACtor database (TRANSFAC) motifs (see Materials and Methods section) did not result in a supershift. These data show that the risk A allele binds more strongly than the G allele to at least one protein and disrupts the binding of IRF2 and one other G allele-specific protein, suggesting that one or more of these transcription factors mediates the effect of rs3780181 alleles on transcriptional activity.

Enhancer deletion reduces *VLDLR* and *SMARCA2* expression

To directly test the role of the enhancer containing rs3780181 on expression of the candidate genes, we genetically engineered two deletions of the enhancer element in human HEK293T cells. We isolated 6 clonal cell lines homozygous for deletion of 475 bp (ΔEnh1) and 6 lines homozygous for a deletion of 663 bp (ΔEnh2). Both deletions span rs3780181 and the 223 bp element that showed allelic differences in enhancer activity (Fig. 2 and Supplementary Material, Fig. S7). For controls, we used mock-edited (control) clonal cell lines and wild-type (WT) HEK293T cells with an intact enhancer containing rs3780181. We quantified the expression of all genes located within 500 kb, *VLDLR*, *SMARCA2* and *KCNV2*, and normalized to the housekeeping gene *ACTB*. Compared to the control cell lines, the ΔEnh2 cell lines showed 50% reduced expression ($P < 0.05$) of both *VLDLR* and *SMARCA2* (Fig. 2B and C) and the ΔEnh1 cell lines showed 14 and 27% reduced expression for the same genes, respectively. *KCNV2* showed only nominal expression in all the tested cell lines. A control gene located on a different chromosome, *VEGFA*, did not have significantly reduced expression ($P > 0.05$) in the engineered cell lines (Supplementary Material, Fig. S8). These data show that deletion of the DNA element containing rs3780181 can decrease expression of *VLDLR* and *SMARCA2* consistent with its putative role as an enhancer.

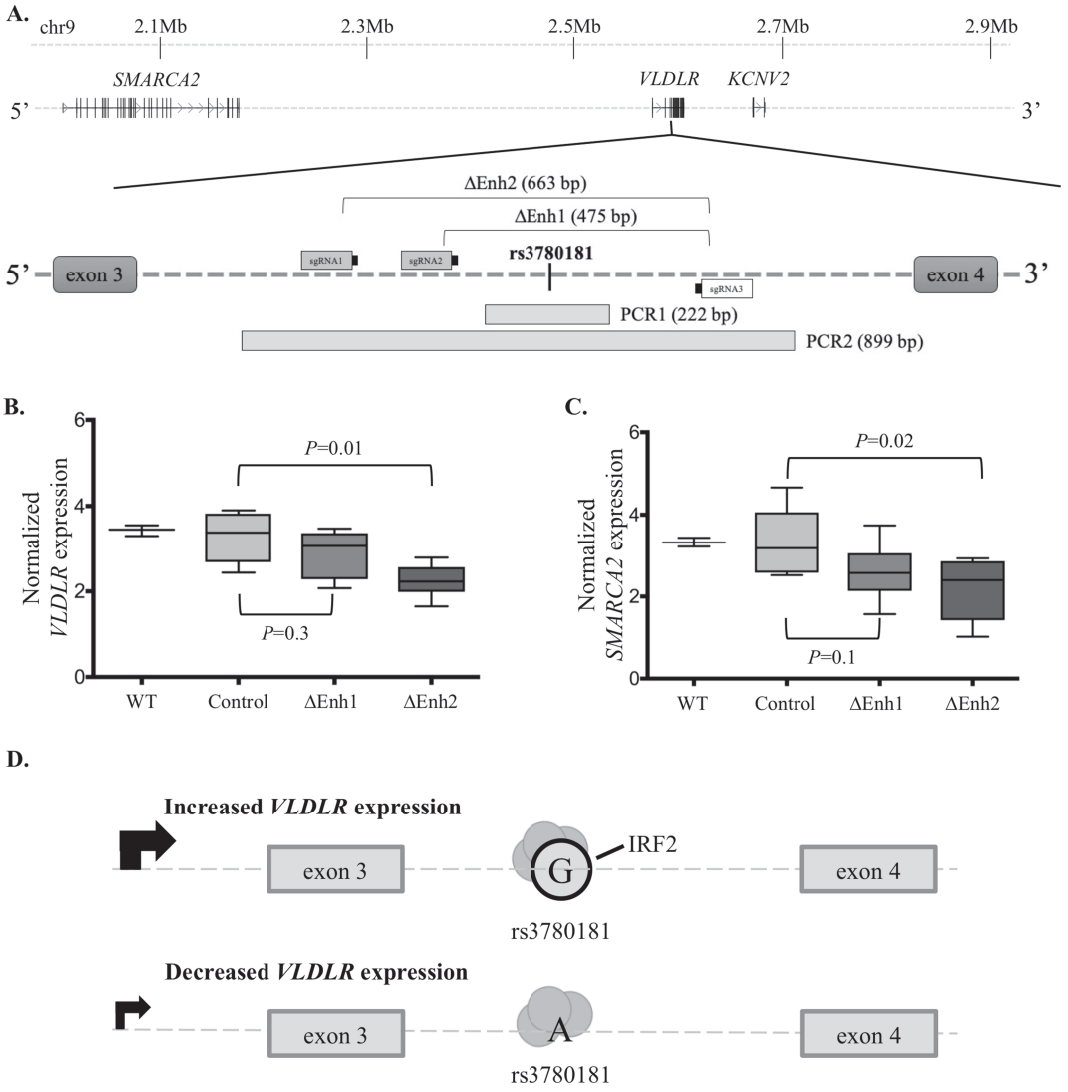

Figure 2. Deletion of an enhancer containing rs3780181 leads to decreased *VLDLR* and *SMARCA2* gene expression. **(A)** The DNA region (chr9:1 977 106–2 919 542; hg19) from rs3780181 with the three candidate functional genes. The DNA region surrounding rs3780181 in intron 3 of *VLDLR* was used to design guide RNAs upstream (dark gray bars) and downstream (white bar) and PCR elements (light gray bars) for the transcriptional reporter assays and to screen for deletion. **(B** and **C)** mRNA gene expression in the unedited (WT and sort control) cell lines and edited (ΔEnh1 and ΔEnh2) cell lines for *VLDLR* and *SMARCA2* was normalized to *ACTB*. **(D)** A proposed molecular mechanism underlying the LDL-C and TC GWAS signal. The rs3780181 G allele differentially binds IRF2 and one other factor, increasing expression of *VLDLR*, which leads to decreased LDL-C and TC.

Discussion

This study describes a molecular mechanism underlying the LDL-C and TC GWAS signal at *VLDLR* (Fig. 2D). We showed that evidence of association, LD and epigenomic annotation support rs3780181 as the best candidate functional variant for this signal. When compared to the non-risk rs3780181 G allele, the A allele showed significantly decreased enhancer activity in hepatocytes, monocytes and preadipocytes consistent with the association of rs3780181-A with lower *VLDLR* expression level in human liver. We identified IRF2 as one of the nuclear proteins that exhibited allelic differences in rs3780181 binding. Based on eQTL, HiC interaction and gene expression after genome editing to delete the rs3780181 enhancer, *VLDLR* showed the most consistent evidence of a functional candidate gene. We propose that differential binding of IRF2 and/or other transcription factors at rs3780181 causes reduced expression of at least *VLDLR*, leading to higher plasma LDL-C and TC levels.

VLDLR functions as a transporter for LDL family molecules by binding and transporting apolipoprotein E and triglyceride-rich lipoproteins into the cell (24,25). By regulating lipoprotein lipase, VLDLR is involved in the catabolism of triglyceride-rich lipoproteins (24). VLDLR stimulates the ERK1/2 pathway and could have a role in macrophage foam cell formation and atherosclerosis pathogenesis (26). *VLDLR* knockout mice have at least 69–89% increased plasma lipids, consistent with the direction we observed in humans by connecting the GWAS variants to gene expression level (23,24). Given the known function of *VLDLR* and the eQTL data connecting rs3780181 to *VLDLR* expression in human liver tissues, the data from this study indicate that *VLDLR* is the most compelling biological candidate gene in this region.

Deletion of the intronic enhancer also decreased expression level of *SMARCA2*, which encodes a chromatin remodeler. SMARCA2 has been shown to down-regulate *CYP7A1* (27) and CYP7A1 deficiency in humans leads to increased

LDL-C (28). The rs3780181 GWAS signal has not yet been found to be colocalized with expression level of *SMARCA2*, but we cannot rule out that the eQTL studies are underpowered. Thus, while *VLDLR* is a stronger biological candidate gene for this GWAS signal, the enhancer containing rs3780181 may also affect *SMARCA2*, which could also play a role in LDL-C and TC levels.

Based on the EMSAs, several transcription factors showed allelic differences in binding at rs3780181, including both putative activators and repressors. With THP-1 and HepG2 nuclear extract, two protein complexes, including IRF2 or a family member bound more strongly to the G allele, suggesting roles as activators, and two unidentified protein complexes bound more strongly to the A allele, suggesting roles as repressors. IRF genes encode interferon regulatory factors, and levels of IRF1 are positively correlated with oxidized-LDL in human THP-1 cell lines (29). Interleukin 6 is a known regulator of IRF1 and IRF2, and increased interleukin 6 led to increased *VLDLR* expression in liver and adipose tissue, which led to decreased blood TC and triglyceride levels in mice (30). The *in vivo* effects of this and other transcriptional regulators that alter *VLDLR* expression through this enhancer remain to be identified.

One limitation of this study is that the most important cell types for the molecular mechanism remain to be resolved. The evidence supporting a role in liver includes the eQTL data, the HepG2 and human liver chromatin marks at rs3780181 and the allelic differences in transcriptional activity and transcription factor binding observed with HepG2. However, *VLDLR* expression level in liver is low. The eQTL associations could reflect other cell types present in human liver, including blood cells and the allelic effects may exist in other cell types not examined here. Another limitation of this study is the cell type (HEK293T) used for the *in vivo* genome-editing experiments. A cell type more biologically relevant to cholesterol levels may implicate different genes and/or provide biological evidence to link to metabolic phenotypes. Further work is needed to understand the key cell types relevant for the enhancer's effect on *VLDLR* expression level.

Together these data define an enhancer and suggest a molecular mechanism by which a human GWAS signal affects *VLDLR* and LDL-C and TC levels, thereby influencing cardiovascular risk. Identifying and validating the connections between genetic variants and the effect on a broadly selected set of candidate genes improves understanding of the molecular basis of cardiovascular disease.

Materials and Methods

Candidate variant annotation

We considered any variants in high LD ($r^2 \geq 0.7$; 1000 Genomes Project, phase 3) with rs3780181 in Europeans as a candidate functional variant (31). We annotated the candidate functional variants for overlap with regulatory data sets from the ENCODE, the Roadmap Epigenomics Consortium and the RegulomeDB (22,32,33). In these data sets, we considered histone epigenomic marks H3K4me1 and H3K27ac as marks of enhancer activity and DNase hypersensitivity, formaldehyde-assisted isolation of regulatory elements-seq and assay for transposase-accessible chromatin-seq as marks of accessible chromatin. We used chromatin immunoprecipitation data and TRANSFAC and ENCODE motifs to identify candidate transcription factors at variant sites (34,35).

Candidate gene annotation

We identified and visualized chromatin interaction between candidate variant locations with Hi-C data sets using the Hi-C unifying genomic interrogation (HUGIn) tool (12,36,37). We evaluated the human primary tissues and assessed the HUGIn significance *P*-values generated using Fit-Hi-C at chromosomal segments overlapping transcription start sites to identify significant excess of chromatin interactions (38). We used the GTEx Project database to compare candidate gene expression across tissues (14). To identify eQTL with the candidate variants, we evaluated liver eQTL data from the literature and searched the GTEx v6 eQTL data using the lead GWAS variant(s), requiring an association of $P < 0.05$ (20,21).

Transcriptional reporter and EMSAs

We tested allelic differences in transcriptional activity with dual-luciferase reporter assays as previously described (18). For the reporter assays, we designed polymerase chain reaction (PCR) primers (5′-AGAGGCCTGCACTGGTAGAA and 5′-CTCCACTAAAAGTTGAGGCTG) to amplify a 223 bp DNA element surrounding rs3780181 from individuals homozygous for both alleles. We cloned PCR products for each allele into the firefly luciferase pGL4.23 reporter vector (Promega). We co-transfected the constructs with phRL-TK renilla luciferase reporter vector (Promega, Fitchburg, WI) into human THP-1 monocyte, SGBS preadipocyte and HepG2 hepatocyte cells. At 24–48 h post-transfection, we measured the luciferase activity from cell extracts, normalized the firefly activity to renilla activity and report activity relative to an empty pGL4.23 vector control.

To conduct the EMSAs, we designed biotin-labeled and unlabeled oligonucleotide probes (5′-TACTTTCA[A/G]TTTTCTAG-3′) for both rs3780181 alleles (A/G), as previously described (18). EMSAs were conducted using the LightShift Chemiluminescent EMSA Kit (Thermo Fisher Scientific, Waltham, MA), following the manufacturer's protocol. Briefly, the oligonucleotides were incubated with NE (4–5 µg, NE-PER kit, Thermo Fisher Scientific) from THP-1 and HepG2 cell lines. To test the specificity of the protein complexes to the specific allele, we added 45× excess competitive unlabeled probes to the binding reactions. Protein–DNA complexes were resolved by gel electrophoresis and transfer, wash and detection steps were performed as described (18). To conduct supershift assays, we added 4 µg of antibody to the binding reactions for the following antibodies: IRF1 (sc-514544X, Santa Cruz), IRF2 antibody (sc-374327X), IRF3 (sc-9082X), HMG1 (sc-26348X), RXRA (sc-553X) and NKX3.2 (sc-25066X).

Identification of protein complexes binding to rs3780181

To identify factors in protein complexes binding rs3780181, we perform a DNA-affinity capture assay using the micro MACS FactorFinder Kit (Miltenyi Biotec, Bergisch Gladbach, Germany), as described in (39). The binding reactions consisted of 200 µg THP-1 NE, 50 pmol biotin probe for rs3780181 (as in EMSAs), 1× protease inhibitor, 10 ng poly-dI:dC and 1× binding buffer. Binding reactions were incubated for 25 min and an additional 10 min with 75 µl streptavidin microbeads. The binding reaction beads were washed with the MACS columns following the kit protocol. The column eluate was added to 80 µl 1× Nupage LDS sample buffer (Life Technologies,

Carlsbad, CA) with 1× NuPAGE reducing agent. The captured protein for both alleles was subjected to electrophoresis on a 4–12% NuPAGE Bis-Tris Gel to a depth of 1 cm from the bottom of the well. The proteins were identified by nano-liquid chromatography followed by tandem mass spectrometry analysis at the University of North Carolina (UNC) Proteomic Core Facility (Supplementary Material, Table S1). We analyzed the motifs of the identified proteins using TRANSFAC and ENCODE matrices (34,35).

In vivo enhancer deletion

To conduct the CRISPR-cas9 assays, we designed pairs of single-guide RNA (sgRNA) oligonucleotides in the genomic region upstream and downstream of rs3780181 to create deletion of an enhancer region in intron 3 of *VLDLR* (Supplementary Material, Table S2). Two guide combinations (sgRNA1 + sgRNA3 and sgRNA2 + sgRNA3) were used to create 663 bp (ΔEnh2) and 475 bp (ΔEnh1) deletions, respectively. The sgRNA guides were cloned as previously described (40). Briefly, we cloned the upstream sgRNAs into the pSpCas9(BB)-2A-GFP (PX458, Addgene plasmid # 48138, a gift from Feng Zhang) and the downstream sgRNA into a modified PX458 with a mCherry reporter in place of the green fluorescent protein (GFC) (a gift from Michael Stitzel, Jackson Laboratory for Genomic Medicine) (18). We used native PX458 vectors without a sgRNA to generate five non-edited clonal cell lines to control for cell manipulation. We transfected the vectors into HEK293T cells using Lipofectamine 2000 following the manufacturer's recommendations, sorted doubly fluorescent cells into 96-well plates using a Becton Dickinson (Franklin Lakes, NJ) FACSAria II and grew cell lines. We screened clonal cell lines for homozygous deletions using PCR primers (5′-GGTCCTGGAAGCTAAGTCA and 5′-AAGCTGTGCCAACCTATGCT) to amplify an 899 bp region surrounding all three sgRNAs; single-cell originating clonal cell lines that yielded single PCR bands were considered candidate homozygous deletions, and the PCR products were Sanger-sequenced to confirm the deletion. We confirmed that six clonal cell lines each for ΔEnh1 and ΔEnh2 have homozygous deletions of the enhancer containing rs3780181 (Supplementary Material, Fig. S7).

mRNA gene expression assays

We extracted whole-cell RNA (RNeasy kit, Qiagen, Venlo, Netherlands) from unedited HEK293T WT (3 replicates), mock-edited control (control, 5 cell lines) and both sizes of homozygous enhancer deletions (ΔEnh1 and ΔEnh2, 6 cell lines each) from >80% confluent wells from 12-well plates. We prepared cDNA libraries as previously described (18) and conducted real-time PCR with TaqMan gene expression probes (Applied Biosystems, Foster City, CA) for the following genes: *VLDLR* (Hs01045921_m1), *SMARCA2* (Hs01030846_m1), *ACTB* (Hs01060665_g1), *KCNV2* (Hs00377936_m1) and *VEGFA* (Hs00900055_m1), following the TaqMan assay protocol. Real-time PCR assays were conducted in triplicate. To determine the significance of the deletions on gene expression, the gene expression of ΔEnh1 and ΔEnh2 cell lines were compared to the control cell lines using unpaired *t*-tests.

Supplementary Material

Supplementary Material is available at *HMG* online.

Acknowledgements

We thank Michael Stitzel, Jackson Laboratory of Genomic Medicine, for providing the modified PX458-mCherry vector and Martin Wabitsch, University of Ulm, for providing SGBS cells. We thank Amy Swift and Lori Bonnycastle for technical assistance with the genome-editing methods. We thank Evan Trudeau at the UNC Flow Cytometry Core Facility and Laura Herring and Nely Dicheva from the UNC Michael Hooker Proteomics Core facility for technical assistance. The UNC Flow Cytometry Core Facility is supported in part by P30 CA016086 Cancer Center Core Support Grant to the UNC Lineberger Comprehensive Cancer Center.

Conflict of Interest statement. None declared.

Funding

National Institutes of Health (T32 HL129982 to J.P.D., T32 HL069768 to T.S.R., T32 GM007092 to T.S.R., R01 DK072193 to K.L.M., U01 DK105561 to K.L.M.).

References

1. Kannel, W.B., Dawber, T.R., Kagan, A., Revotskie, N. and Stokes, J. (1961) Factors of risk in the development of coronary heart disease—six year follow-up experience. The Framingham Study. *Ann. Intern. Med.*, **55**, 33–50.
2. Pilia, G., Chen, W.-M., Scuteri, A., Orrú, M., Albai, G., Dei, M., Lai, S., Usala, G., Lai, M., Loi, P. *et al.* (2006) Heritability of cardiovascular and personality traits in 6,148 Sardinians. *PLoS Genet.*, **2**, e132.
3. Hoffmann, T.J., Theusch, E., Haldar, T., Ranatunga, D.K., Jorgenson, E., Medina, M.W., Kvale, M.N., Kwok, P.Y., Schaefer, C., Krauss, R.M. *et al.* (2018) A large electronic-health-record-based genome-wide study of serum lipids. *Nat. Genet.*, **50**, 401–413.
4. Liu, D.J., Peloso, G.M., Yu, H., Butterworth, A.S., Wang, X., Mahajan, A., Saleheen, D., Emdin, C., Alam, D., Alves, A.C. *et al.* (2017) Exome-wide association study of plasma lipids in >300,000 individuals. *Nat. Genet.*, **49**, 1758–1766.
5. 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.* (2013) Discovery and refinement of loci associated with lipid levels. *Nat. Genet.*, **45**, 1274–1283.
6. Spracklen, C.N., Chen, P., Kim, Y.J., Wang, X., Cai, H., Li, S., Long, J., Wu, Y., Wang, Y.X., Takeuchi, F. *et al.* (2017) Association analyses of East Asian individuals and trans-ancestry analyses with European individuals reveal new loci associated with cholesterol and triglyceride levels. *Hum. Mol. Genet.*, **26**, 1770–1784.
7. Gao, H., Yin, R., Zhang, Q., Qiu, L., Khounphinith, E., Wang, D. and Li, K. (2017) Association of the VLDLR rs3780181 SNP and serum lipid levels in the Jing and Han populations. *Int. J. Clin. Exp. Pathol.*, **10**, 5636–5654.
8. Davis, J.P., Huyghe, J.R., Locke, A.E., Jackson, A.U., Sim, X., Stringham, H.M., Teslovich, T.M., Welch, R.P., Fuchsberger, C., Narisu, N. *et al.* (2017) Common, low-frequency, and rare genetic variants associated with lipoprotein subclasses and triglyceride measures in Finnish men from the METSIM study. *PLoS Genet.*, **13**, e1007079.
9. Ahola-Olli, A.V., Würtz, P., Havulinna, A.S., Aalto, K., Pitkänen, N., Lehtimäki, T., Kähönen, M., Lyytikäinen, L.P., Raitoharju, E., Seppälä, I. *et al.* (2017) Genome-wide associ-

ation study identifies 27 loci influencing concentrations of circulating cytokines and growth factors. *Am. J. Hum. Genet.*, **100**, 40–50.
10. Parmar, P.G., Taal, H.R., Timpson, N.J., Thiering, E., Lehtimäki, T., Marinelli, M., Lind, P.A., Howe, L.D., Verwoert, G., Aalto, V. *et al.* (2016) International genome-wide association study consortium identifies novel loci associated with blood pressure in children and adolescents. *Circ. Cardiovasc. Genet.*, **9**, 266–278.
11. Gupta, R.M., Hadaya, J., Trehan, A., Zekavat, S.M., Roselli, C., Klarin, D., Emdin, C.A., Hilvering, C.R.E., Bianchi, V., Mueller, C. *et al.* (2017) A genetic variant associated with five vascular diseases is a distal regulator of endothelin-1 gene expression. *Cell*, **170**, 522–533 e15.
12. Martin, J.S., Xu, Z., Reiner, A.P., Mohlke, K.L., Sullivan, P., Ren, B., Hu, M. and Li, Y. (2017) HUGIn: Hi-C unifying genomic interrogator. *Bioinformatics*, **33**, 3793–3795.
13. Smemo, S., Tena, J.J., Kim, K.H., Gamazon, E.R., Sakabe, N.J., Gómez-Marín, C., Aneas, I., Credidio, F.L., Sobreira, D.R., Wasserman, N.F. *et al.* (2014) Obesity-associated variants within FTO form long-range functional connections with IRX3. *Nature*, **507**, 371–375.
14. Lonsdale, J., Thomas, J., Salvatore, M., Phillips, R., Lo, E., Shad, S., Hasz, R., Walters, G., Garcia, F., Young, N. *et al.* (2013) The Genotype-Tissue Expression (GTEx) project. *Nat. Genet.*, **45**, 580–585.
15. Tong, P., Monahan, J. and Prendergast, J.G.D. (2017) Shared regulatory sites are abundant in the human genome and shed light on genome evolution and disease pleiotropy. *PLoS Genet.*, **13**, e1006673.
16. Roman, T.S. and Mohlke, K.L. (2018) Functional genomics and assays of regulatory activity detect mechanisms at loci for lipid traits and coronary artery disease. *Curr. Opin. Genet. Dev.*, **50**, 52–59.
17. Edwards, S.L., Beesley, J., French, J.D. and Dunning, M. (2013) Beyond GWASs: illuminating the dark road from association to function. *Am. J. Hum. Genet.*, **93**, 779–797.
18. Roman, T.S., Cannon, M.E., Vadlamudi, S., Buchkovich, M.L., Wolford, B.N., Welch, R.P., Morken, M.A., Kwon, G.J., Varshney, A., Kursawe, R. *et al.* (2017) A type 2 diabetes-associated functional regulatory variant in a pancreatic islet enhancer at the ADCY5 locus. *Diabetes*, **66**, 2521–2530.
19. Sanjana, N.E., Wright, J., Zheng, K., Shalem, O., Fontanillas, P., Joung, J., Cheng, C., Regev, A. and Zhang, F. (2016) High-resolution interrogation of functional elements in the non-coding genome. *Science*, **353**, 1545–1549.
20. Giambartolomei, C., Vukcevic, D., Schadt, E.E., Franke, L., Hingorani, A.D., Wallace, C. and Plagnol, V. (2014) Bayesian test for colocalisation between pairs of genetic association studies using summary statistics. *PLoS Genet.*, **10**, e1004383.
21. Innocenti, F., Cooper, G.M., Stanaway, I.B., Gamazon, E.R., Smith, J.D., Mirkov, S., Ramirez, J., Liu, W., Lin, Y.S., Moloney, C. *et al.* (2011) Identification, replication, and functional fine-mapping of expression quantitative trait loci in primary human liver tissue. *PLoS Genet.*, **7**, e1002078.
22. Boyle, A.P., Hong, E.L., Hariharan, M., Cheng, Y., Schaub, M.A., Kasowski, M., Karczewski, K.J., Park, J., Hitz, B.C., Weng, S. *et al.* (2012) Annotation of functional variation in personal genomes using RegulomeDB. *Genome Res.*, **22**, 1790–1797.
23. Nguyen, A., Tao, H., Metrione, M. and Hajri, T. (2014) Very low density lipoprotein receptor (VLDLR) expression is a determinant factor in adipose tissue inflammation and adipocyte-macrophage interaction. *J. Biol. Chem.*, **289**, 1688–1703.
24. Yagyu, H., Lutz, E.P., Kako, Y., Marks, S., Hu, Y., Choi, S.Y., Bensadoun, A. and Goldberg, I.J. (2002) Very low density lipoprotein (VLDL) receptor-deficient mice have reduced lipoprotein lipase activity. Possible causes of hypertriglyceridemia and reduced body mass with VLDL receptor deficiency. *J. Biol. Chem.*, **277**, 10037–10043.
25. Takahashi, S., Suzuki, J., Kohno, M., Oida, K., Tamai, T., Miyabo, S., Yamamoto, T. and Nakai, T. (1995) Enhancement of the binding of triglyceride-rich lipoproteins to the very low density lipoprotein receptor by apolipoprotein E and lipoprotein lipase. *J. Biol. Chem.*, **270**, 15747–15754.
26. Liu, Z., Li, H., Li, Y., Wang, Y., Zong, Y., Feng, Y., Feng, Z., Deng, Y. and Qu, S. (2009) Up-regulation of VLDL receptor expression and its signaling pathway induced by VLDL and beta-VLDL. *J. Huazhong Univ. Sci. Technolog. Med. Sci.*, **29**, 1–7.
27. Smith, Z., Ryerson, D. and Kemper, J.K. (2013) Epigenomic regulation of bile acid metabolism: emerging role of transcriptional cofactors. *Mol. Cell. Endocrinol.*, **368**, 59–70.
28. Pullinger, C.R., Eng, C., Salen, G., Shefer, S., Batta, A.K., Erickson, S.K., Verhagen, A., Rivera, C.R., Mulvihill, S.J., Malloy, M.J. *et al.* (2002) Human cholesterol 7α-hydroxylase (CYP7A1) deficiency has a hypercholesterolemic phenotype. *J. Clin. Invest.*, **110**, 109–117.
29. Holvoet, P., Davey, P.C., De Keyzer, D., Doukouré, M., Deridder, E., Bochaton-Piallat, M.L., Gabbiani, G., Beaufort, E., Bishay, K., Andrieux, N. *et al.* (2006) Oxidized low-density lipoprotein correlates positively with toll-like receptor 2 and interferon regulatory factor-1 and inversely with superoxide dismutase-1 expression: studies in hypercholesterolemic swine and THP-1 cells. *Arterioscler. Thromb. Vasc. Biol.*, **26**, 1558–1565.
30. Hashizume, M., Yoshida, H., Koike, N., Suzuki, M. and Mihara, M. (2010) Overproduced interleukin 6 decreases blood lipid levels via upregulation of very-low-density lipoprotein receptor. *Ann. Rheum. Dis.*, **69**, 741–746.
31. Machiela, M.J. and Chanock, S.J. (2018) LDassoc: an online tool for interactively exploring genome-wide association study results and prioritizing variants for functional investigation. *Bioinformatics*, **34**, 887–889.
32. The ENCODE Project Consortium (2012) An integrated encyclopedia of DNA elements in the human genome. *Nature*, **489**, 57–74.
33. Roadmap Epigenomics Consortium (2015) Integrative analysis of 111 reference human epigenomes. *Nature*, **518**, 317–330.
34. Matys, V. (2006) TRANSFAC and its module TRANSCompel: transcriptional gene regulation in eukaryotes. *Nucleic Acids Res.*, **34**, D108–D110.
35. Kheradpour, P. and Kellis, M. (2014) Systematic discovery and characterization of regulatory motifs in ENCODE TF binding experiments. *Nucleic Acids Res.*, **42**, 2976–2987.
36. Jin, F., Li, Y., Dixon, J.R., Selvaraj, S., Ye, Z., Lee, A.Y., Yen, C.-A., Schmitt, A.D., Espinoza, C.A. and Ren, B. (2013) A high-resolution map of the three-dimensional chromatin interactome in human cells. *Nature*, **503**, 290–294.

37. Schmitt, A.D., Hu, M., Jung, I., Xu, Z., Qiu, Y., Tan, C.L., Li, Y., Lin, S., Lin, Y., Barr, C.L. *et al.* (2016) A compendium of chromatin contact maps reveals spatially active regions in the human genome. *Cell Rep.*, **17**, 2042–2059.
38. Ay, F., Bailey, T.L. and Noble, W.S. (2014) Statistical confidence estimation for Hi-C data reveals regulatory chromatin contacts. *Genome Res.*, **24**, 999–1011.
39. Lu, X., Zoller, E.E., Weirauch, M.T., Wu, Z., Namjou, B., Williams, A.H., Ziegler, J.T., Comeau, M.E., Marion, M.C., Glenn, S.B. *et al.* (2015) Lupus risk variant increases pSTAT1 binding and decreases ETS1 expression. *Am. J. Hum. Genet.*, **96**, 731–739.
40. Ran, F.A., Hsu, P.D., Wright, J., Agarwala, V., Scott, D.A. and Zhang, F. (2013) Genome engineering using the CRISPR-Cas9 system. *Nat. Protoc.*, **8**, 2281–2308.