



Published in final edited form as:

Nat Commun. 2013 ; 4: 1973. doi:10.1038/ncomms2973.

A common functional promoter variant links *CNR1* gene expression to HDL cholesterol level

Q Feng¹, KC Vickers^{1,2}, MP Anderson², MG Levin², W Chen¹, DG Harrison¹, and RA Wilke¹

¹Dept. of Medicine, Vanderbilt University School of Medicine, Nashville, TN, USA

²National Heart, Lung and Blood Institute, National Institutes of Health, Bethesda, MD, USA

Abstract

CB₁ receptor blockers increase HDL-C levels. Although genetic variation in the CB₁ receptor – encoded by the *CNR1* gene – is known to influence HDL-C level as well, human studies conducted to date have been limited to genetic markers such as haplotype tagging SNPs. Here we identify rs806371 in the *CNR1* promoter as the causal variant. We resequenced the *CNR1* gene and genotype all variants in a DNA biobank linked to comprehensive electronic medical records. By testing each variant for association with HDL-C level in a clinical practice-based setting, we localize a putative functional allele to a 100bp window in the 5'-flanking region. Assessment of variants in this window for functional impact on electrophoretic mobility shift assay identified rs806371 as a novel regulatory binding element. Reporter gene assays confirm that rs806371 reduces HDL-C gene expression, thereby linking *CNR1* gene variation to HDL-C level in humans.

INTRODUCTION

Clinical lipid disorders have enormous public health significance and increasing societal burden in developed countries¹. High-density lipoprotein cholesterol (HDL-C) levels are inversely correlated with cardiovascular disease (CVD), and each 1 mg/dL decrease in HDL-C is associated with a 6% increase in adverse clinical events². Because HDL-C levels are highly heritable (H^2 ranging from ~0.4 to 0.7), there is great interest in characterizing the genetic architecture underlying this important complex trait³⁻⁸.

The type 1 cannabinoid receptor (CB₁) is a novel therapeutic target for controlling lipoprotein metabolism. Because CB₁ receptors within the brain influence eating behavior, rimonabant, a selective CB₁ antagonist, was initially designed to correct weight gain^{9,10}.

Users may view, print, copy, download and text and data- mine the content in such documents, for the purposes of academic research, subject always to the full Conditions of use: http://www.nature.com/authors/editorial_policies/license.html#terms

Author Contribution

Q.F. and R.A.W. designed the study and drafted the manuscript. Q.F., K.C.V., M.P.A., M.G.L. and W.C. performed the experiments and statistical analyses. Q.F., K.C.V., M.P.A., M.G.L., W.C., D.G.H., and R.A.W. helped with data interpretation, and provided critical revisions to the final manuscript.

Conflict of interest

The authors declare no conflict of interest.

Accession codes

CNR1 SNP data have been deposited at Seattle SNPs on Genome Variation Server 134 (<http://gvs.gs.washington.edu/GVS134/>) under accession code XXXX.

However, in large clinical trials, rimonabant improved HDL-C levels far greater than originally anticipated^{9–12}. Likewise, variation in the *CNR1* gene, which transcribes the CB₁ receptor, has previously been associated with HDL-C levels in several independent study cohorts^{13–15}. We have previously reported that a common *CNR1* haplotype (H4, frequency ~15% in the general population) is associated with decreased HDL-C levels, independent of body mass index (BMI)¹³.

Interestingly, common non-synonymous SNPs have not been observed within the *CNR1* coding region, and all studies assessing the impact of *CNR1* gene variation on clinically recognizable traits have been conducted using “markers” (haplotype tagging SNPs). Therefore, the causal allele has not been identified. As such, we present here a series of polymorphisms in the *CNR1* gene identified through deep re-sequencing, and we re-genotype these variants testing them for association with HDL-C levels *in vivo*, using the largest clinical practice-based DNA biobank in United States^{16,17}. Variants associated with HDL-C level were then further characterized experimentally *in vitro*, using electrophoretic mobility shift assays (EMSA) and gene promoter reporter (luciferase) assays. We now report that rs806371 is the likely causal variant linking *CNR1* gene expression to HDL-C level.

RESULTS

Identification of putative functional variants

BioVU is a clinical practice-based biobank linked to comprehensive electronic medical records (EMRs)¹⁷. The largest resource of its kind (n = 157,719 on January 10, 2013), BioVU is robust in its ability to replicate genetic associations previously identified in disease-based cohorts¹⁶. Our previous analysis identified a *CNR1* haplotype associated with HDL-C levels in extended families of Northern European ancestry¹³. To assess this relationship within the community, we sampled 1% of 100,000 BioVU subject records claiming European Ancestry for further study (50% females; 50% males). The precision of ancestry data within BioVU has previously been validated using a panel of 360 ancestry informative markers (AIMs)¹⁸. Clinical characteristics of the current BioVU sub-sample are shown in Table 1.

To quantify variability across our locus of interest, the entire *CNR1* gene (15kb of genomic DNA) was re-sequenced in 95 individuals selected from the Utah Centre d’Etude du Polymorphisme Humain (CEPH) reference panel within HapMap (Coriell Cell Repository) (Figure 1, Table 2). A total of 65 polymorphisms were identified, including 62 SNPs and three insertion/deletions. Thirty-seven of these polymorphisms had not been reported previously (i.e., did not have existing rsNumbers). All observed variants were re-genotyped in the BioVU sub-cohort from Table 1 (n = 1006) using our high-throughput Sequenom platform (see **Methods**). As described in our previous study using clinical data derived from EMRs¹⁹, the primary endpoint in the BioVU sub-cohort was median outpatient HDL-C levels, adjusted for age and gender¹⁴. Other important clinical covariates were either extracted directly from the EMRs (e.g., BMI nearest to the date that each median HDL-C level was extracted) or defined using time stamps through natural language processing (e.g., exposure to medications known to alter HDL-C levels)²⁰.

Median HDL-C levels were then tested for association with each variant genotyped across the *CNR1* locus (Table 2). Using this strategy, we found three common *CNR1* variants to be nominally associated with HDL-C level ($p < 0.05$, additive model): two in the 5'-flanking region (5'-FR) and one located in the 3' untranslated region (3'-UTR). Both variants in the 5'-FR remained significant after adjustment for BMI (Table 2): CNR1-8727 (dbSNP designation rs806370) and CNR1-8695 (dbSNP designation rs806371). None of the synonymous variants in the *CNR1* coding region were found to be associated with HDL-C level. Although rare, two additional *CNR1* variants were associated with HDL-C level after adjustment for BMI (Table 2): (CNR1-11675 and CNR1-19154). These rare 3' variants were tightly linked in our pair-wise allelic association analyses ($r^2 = 1.0$). However, due to their extremely low minor allele frequencies (2 alleles in 1006 study subjects), these variants were not pursued further in our comparison with prior *CNR1* risk haplotypes, or in our functional assessment of *CNR1* gene expression *in vitro*.

rs806371 is associated with HDL-C levels

We have previously reported a common *CNR1* haplotype associated with HDL-C level in families of Northern European ancestry¹³. Because our re-genotyping of the *CNR1* locus was comprehensive in the current study, we were able to reconstruct the previously reported risk haplotype, H4, in this BioVU sub-sample (Table 3)¹³⁻¹⁵. Strikingly, the effect size and level of significance for association with HDL-C was identical for the H4 risk haplotype and for rs806371, our putative functional variant harbored within the 5'-FR. (HDL-C mean \pm SD was 53.47 ± 16.30 , 50.85 ± 16.76 , and 48.03 ± 12.52 mg/dl for carriers of 0, 1, and 2 copies of the H4 risk haplotype; and HDL-C mean \pm SD was 53.47 ± 16.45 , 51.13 ± 16.64 , and 48.23 ± 12.60 mg/dl for subjects with 0, 1, and 2 copies of the minor allele at **rs806371**.)

We therefore conditioned our findings for rs806371 and HDL-C on the tagging SNPs used previously to define the H4 risk haplotype. When we conditioned our findings for rs806371 on genotype at the adjacent variant rs806370, the significance of this relationship was attenuated (from $p = 0.026$ to $p = 0.90$, additive model) implying that the association was not solely driven by rs806371. Furthermore, when we did the converse (conditioned the analyses for rs806370 on rs806371), the association of HDL-C with rs806370 was also attenuated (from $p = 0.017$ to $p = 0.54$, additive model). Because these observations suggest that *both* variants may contribute to the association, we screened all 5' variants to quantify the strength of their linkage disequilibrium with the H4 haplotype using Genome Variation Server 134 at *Seattle SNPs* (<http://gvs.gs.washington.edu/GVS134/>) (Figure 2), and we subsequently tested rs806370 and rs806371 for binding interactions with nuclear proteins using EMSA (Figure 3).

To begin assessing the causality of rs806371, we plotted genotypes for each *CNR1* SNP in all 16 individuals currently found to be homozygous for the H4 haplotype described in our prior work (Figure 2); in this analysis rs806371 was found to only occur on the H4 background. Although CNR1-7738, another *CNR1* 5'-FR variant, was also found to have increased prevalence on the H4 background (Figure 2), this variant was frequently present on other haplotypes (not shown), and its correlation with H4 was not robust ($R^2 = 0.139$). As such, we did not pursue CNR1-7738 in our subsequent functional studies.

rs806371 creates a new regulatory element for *CNR1*

Electrophoretic mobility shift assays (EMSA) were then conducted to determine if nucleotide substitutions disrupt or induce any regulatory elements in the *CNR1* promoter (Figure 3). As noted, rs806370 and rs806371 in the 5'-FR of *CNR1* are both associated with HDL-C *in vivo*. These variants are tightly linked ($r^2=0.925$), and their minor alleles are commonly inherited together (as a diplotype, frequency ~15%). To determine their impact in EMSA, we first tested their collective impact on DNA binding using nuclear protein lysates from embryonic kidney cells (HEK293) (Figure 3A). The presence of both rs806370 and rs806371 induced a super-shift band compared to combined major alleles when characterized by EMSA. Although our *in vivo* studies indicated that rs806370 and rs806371 are tightly linked, our *in vitro* studies indicate that only rs806371 introduces a novel DNA binding site for nuclear proteins, based upon subsequent EMSA analyses conducted with probes for single variants (Figure 3A).

rs806371 decreases reporter gene expression

To further characterize the functional role of rs806371 in the context of gene expression, and to quantify the direction of the effect, gene promoter reporter (luciferase) assays were conducted using site directed mutagenesis. The *CNR1* 5'-FR (~3kb immediately 5' from the *CNR1* transcription start site) was cloned upstream of *Gaussian* luciferase (secreted) in the the pEZX-PG04 reporter construct which also contains secretory alkaline phosphatase for transfection efficiency control analysis (see **Methods**). Due to the observation that rs806371, not rs806370, was found to be the functional variant on EMSA, site-directed mutagenesis was used to create a representative rs806371 novel regulatory DNA binding site. After transfection into human hepatoma cells (Huh7 cells), promoter reporter luciferase and control alkaline phosphatase activities were quantified serially in cultured media, over three days (Figure 3B). Reporter gene expression, quantified by secreted luciferase activity, was markedly reduced by the presence of rs806371: ~25% reduction in activity at 48 h, and ~50% reduction at 72 and 96 h.

DISCUSSION

There is growing evidence that genetic variation in *CNR1* directly influences dysmetabolic traits in humans by altering the activity of CB₁ receptor-dependent signaling in peripheral tissues. We provide evidence in support of this claim through a novel series of experiments involving deep sequencing and comprehensive re-genotyping in a clinical practice-based biobank, followed by functional validation in two separate *in vitro* assay systems. Our results show that rs806371 in the *CNR1* promoter alters HDL-C level in humans by generating a novel regulatory DNA binding site capable of reducing *CNR1* expression.

Our current approach is unique. We had previously reported an association between *CNR1* haplotype and circulating HDL-C levels in multi-generational families¹³. To narrow our search for the causal variant underlying this association, a 15kb region of genomic DNA (containing the promoter, 5'-FR, coding region, and 3'-UTR of the *CNR1* gene) was sequenced, and all polymorphisms discovered in this region were re-genotyped in a biobank linked to electronic medical records (EMRs)¹⁶. In recent years, the use of EMRs has

expanded rapidly (in response to the Affordable Care Act) creating huge longitudinal datasets ideal for observational research^{20–22}, and BioVU at Vanderbilt University is currently the nation's largest collection of DNA samples linked to EMRs (n = 157,719). Because of its unprecedented sample size and its unique design, we were able to leverage BioVU to identify a narrow region (~100bp) within the *CNR1* promoter associated with median HDL-C levels independent of BMI. The strength of this association did not change when our clinical practice-based lipid data were right-censored by exposure to lipid-modifying therapy (*before* censoring, mean HDL-C \pm SD was 53.47 \pm 16.45, 51.13 \pm 16.64, and 48.23 \pm 12.60 mg/dl for subjects with AA, AC, and CC at rs806371; *after* censoring, mean HDL-C \pm SD was 52.71 \pm 16.80, 52.70 \pm 17.05, and 48.13 \pm 12.84 mg/dl for AA, AC, and CC).

In our primary analysis, two common variants within the 5'-FR, rs806370 and rs806371, were associated with median HDL-C level extracted from clinical data, after adjustment for BMI. Because the effect size for each of these associations was clinically significant (~5 mg/dL change in HDL-C level), and nearly identical to the effect of our previously reported *CNR1* risk haplotype (Table 3), rs806370 and rs806371 were selected for further functional characterization *in vitro*. Even though these two SNPs are co-inherited, our data clearly indicate that rs806371, not rs806370, alters nuclear protein binding in an EMSA. Because rs806371 markedly reduces reporter gene expression when engineered into a vector containing the *CNR1* promoter, further studies are needed to define the transcription factors involved.

Tissue specific differences in transcription factors regulating the *CNR1* locus (e.g., within brain versus adipose tissue) may explain the fact that CB₁ receptor antagonists increase HDL-C, whereas *CNR1* loss of function variants decrease HDL-C²³. For example, insulin-dependent signaling in the brain engages different second messengers than peripheral tissues^{24,25}. The degree to which these differences impact *CNR1* expression remain uncharacterized. Physical interaction between our 5' loss of function variant and *CNR1* elements located more distally may also differ tissue by tissue. Within the current study, we also observed that HDL-C was associated with a rare variant located on the 3' end of the gene. While this variant, CNR1-11675, was only observed in 2 of 1000 study subjects, the de-identified clinical data linked to these 2 samples revealed a marked elevation in HDL-C level (70.2 \pm 17.3 mg/dl, mean \pm SD) in the absence of an obvious clinical explanation for abnormal lipid homeostasis. Thus, additional work is needed to clarify the impact of these *CNR1* variants on chromatin structure. Chromatin state predictions now available on line for ENCODE indicate the presence of an insulator regulated by CTCF near *CNR1* (<http://genome.ucsc.edu/cgi-bin/hgTracks?position=chr6:88847472-88881527&hgside=325405609&wgEncodeUwHistoneViewHot=full>). Bioinformatic algorithms that predict transcription factor binding suggest additional regulatory mechanisms on the 5'end; for example, the minor allele at rs806371 may disrupt a non-canonical glucocorticoid receptor (TRANSFAC V8.3 at http://algggen.lsi.upc.es/cgi-bin/promo_v3/promo/promo.cgi?dirDB=TF_8.3&calledBy=algggen). Publically available eQTL datasets indicate that this locus also regulates the nearby *RNGTT* gene (p = 0.0273), in HapMap Europeans²³. *RNGTT* is located just upstream of *CNR1* (~400kb), and it encodes an

RNA guanylyltransferase recently associated with subcutaneous adipose tissue volume in women from the Framingham Offspring Study²⁸.

Clearly, the pathophysiological mechanism linking rs806371 to HDL-C level warrants further investigation. Endocannabinergic signaling alters HDL homeostasis through mechanisms that are both direct (*CNR1* expression in hepatocytes or adipocytes)^{29–31} and indirect (*CNR1* expression in the brain)³². Non-brain-penetrant CB₁ receptor antagonists directly increase plasma HDL-C levels in animal models²⁶, and CB₁ blockade modulates the release of adipocytokines from human adipocytes^{31–34}. Because these processes influence the intravascular remodeling of HDL particles *in vivo*^{34–36}, *CNR1* gene variability may influence public health by altering cardiovascular risk in the context of the current obesity epidemic³⁷.

METHODS

Study Population

The current study was conducted in accordance with the Principles outlined within the Declaration of Helsinki. Approval was obtained from the Institutional Review Board of Vanderbilt University. The Vanderbilt DNA biobank (BioVU) currently contains electronic medical records from 157,719 subjects (updated January 10, 2013). BioVU accrues DNA samples extracted from blood drawn for routine clinical testing after these samples have been retained for 3 days and scheduled to be discarded. The DNA samples in BioVU are linked to a *de-identified* mirror image of each individual's EMR. The current study cohort was randomly selected from approximately 100,000 unique individuals with European-ancestry in BioVU. In order to reduce the potential data fragmentation caused by multiple health care providers, we restricted study subjects to those listing Vanderbilt University Medical Center as their primary health care provider^{38,39}. To do so, at least one note from general internal medicine was required. To enrich the dataset for de-identified patient records containing dense longitudinal lipid data, at least three clinical lipid panels were also required for each subject. Using this approach, a total of 1006 representative subjects were selected, including 509 females and 497 males.

Phenotyping

Outpatient median HDL-C levels represented our primary endpoint, available on all 1006 subjects. Clinical lipid data were extracted from EMRs. These data reflected longitudinal lipid data (more than 10 years) collected during the course of routine clinical care. On average, each subject has 8 HDL-C values (ranging from 3–38 tests). All lipid data from inpatients were excluded since acute illness typically influences circulating lipid levels in most inpatients⁴⁰. Data on lipid medications and related events (drug, dose, date and time) were also obtained from EMRs by applying our nature language processing algorithms^{41,42}. Relevant clinical covariates were also obtained from EMRs. On average, each subject record had 18 glucose values (ranging from 1–196 tests) and 25 blood pressure values (ranging from 1–229 tests). Body mass index (BMI) was calculated for each subject using median height and weight nearest the time stamp for median lipid values. When comparing our study cohort to the community of 1.8 million unique individuals from which this sample set

was derived (Table 1), no statistically significant difference was observed for BMI. However, there was a significant difference in median glucose levels ($p < 0.001$ additive model). Because our inclusion criteria required dense longitudinal data (i.e., at least three lipid panels per subject), our sampling process enriched for study subjects with regularly monitored cardiovascular risk factors. Thus, our sample was more likely to contain subjects with well-controlled diabetes mellitus. As anticipated, 250 of our 1006 study subjects were diabetic, and because they were well-controlled, the mean glucose level in our overall sample was near normal (103.2 ± 26.6 mg/dl, mean \pm SD) (Table 1).

Sequencing

The *CNRI* gene was completely sequenced in 95 individuals selected from Utah Centre d'Etude du Polymorphisme Humain (CEPH) HapMap reference panel (available from the Coriell Cell Repository, Camden, NJ). A total of 15kb genomic region was sequenced, including 5 kb upstream of the gene, 5kb around the coding exon, and 5 kb downstream. Briefly, 5'- M13 tailed-gene specific PCR primers were designed to cover the target region with amplicon sizes ranging from 500–750 bp with a minimum of 100 bp overlap between adjacent amplicons, where applicable, resulting in double-stranded coverage of all targeted regions. Overlapping amplicons were used to validate gene-specific primer sequences in independent experiments and rule out the possibility of allele-specific PCR amplifications. All primer sequences were compared to the whole genome assembly to verify uniqueness against pseudogenes and gene families. Following temperature gradient optimization of small-scale reactions to determine optimal thermal cycling conditions, production level PCR amplifications were performed in 96-well plates in a volume of 7 μ l comprising 0.2 μ l each of 7 μ M forward and reverse primers, 2.8 μ l DNA (5 ng/ μ l), and 0.4 μ l Elongase Enzyme (Invitrogen, Grand Island, NY, USA) or iProof polymerase (Bio-Rad, Hercules, CA, USA) per well. Sequencing reactions were performed in MJ Tetrad PTC 225 thermal cyclers in 384-well format by using 5% BDT v3.1 sequencing chemistry (ABI, Foster City, CA, USA). Chromatograms were generated from sequence reaction on an Applied Biosystems ABI 3730XL capillary sequencer. Data flow was tracked by using a custom-designed LIMS system at University of Washington. All chromatograms were base-called by using Phred, assembled into contigs by using Phrap, and scanned for SNPs with PolyPhred (v6.02) to identify polymorphic sites. Each chromatogram was trimmed to remove low-quality sequence (Phred score < 25), resulting in analyzed reads averaging > 450 bp with an average Phred quality of 40. Following assembly of all chromatograms onto an initial reference sequence, putative polymorphic sites were selectively reviewed by sequence analysts using Consed. Individual polymorphic sites in regions with lower quality data, ambiguous base calls, deviations from Hardy-Weinberg equilibrium or those identified using laboratory quality control tools were reviewed to eliminate potential false positive positions. Outlier genotypes (i.e., deviations from Hardy-Weinberg equilibrium) were scrutinized by data analysts and removed from the dataset if ambiguous. This approach generates sequence-based SNP genotypes with accuracy $> 99.9\%$. The results are publically available at Seattle SNPs on Genome Variation Server 134 at <http://gvs.gs.washington.edu/GVS134/>.

Genotyping

Genotyping was performed on 1006 DNA samples from BioVU using both common and rare variants identified from re-sequencing. Because our sequencing efforts revealed no non-synonymous SNPs, three potential non-synonymous SNPs (rs78783387, rs75770301 and rs77016054) were included from existing databases. Two platforms were employed: six tagSNPs from our previous report¹³ were genotyped using Taqman assay (ABI); all other polymorphisms were genotyped using iPLEX Gold assay (Sequenom, San Diego, CA, USA). MassARRAY[®] Designer software was used to design iPLEX single base extension primers for multiplexed assays. Fifteen polymorphisms were rejected by the software and eliminated from genotyping. As a result, 54 polymorphisms were genotyped in the *CNR1* genomic region on 1006 subjects. All polymorphisms met a call rate threshold of 94%. Except for rs806368 on the Taqman platform, all other SNPs had a call rate greater than 97%. Except for CNR1-4902 on the iPLEX Gold platform, all other SNPs had call rate greater than 99%. Nine of these SNPs were monomorphic in BioVU, and they were removed from our analyses. All remaining SNPs followed H-W equilibrium.

Electrophoretic mobility shift assays

Double stranded DNA probes (with or without rs806370 and rs806371) were prepared from biotin-labeled DNA oligos synthesized by IDT (Coralville, Iowa, USA). Nuclear protein was extracted with NE-PER nuclear extraction reagents (Pierce, Rockford, IL, USA), and each electrophoretic mobility shift assays (EMSA) was performed with a Light Shift chemiluminescent EMSA kit (Pierce). For EMSA, the binding reactions were performed for 20 min in 1× binding buffer, 5 mM MgCl₂, 50 ng/μl poly(dI-dC), 0.05% Nonidet P-40, 2.5% glycerol, biotin-labeled probe, and nuclear protein extracts. Samples were electrophoresed on a native 6% polyacrylamide gel in 1× Tris-borate-EDTA buffer and then transferred to a Biodyne membrane according to the manufacturer's recommendation. For the competitive binding assay, non-labeled probes were added to the binding reaction at 200-fold excess over labeled probe⁴³.

Gene promoter reporter (luciferase) assays

Approximately 3kb of *CNR1* 5'-flanking region was amplified from human genomic DNA to create a reporter gene construct in pEZX-PG04 (GeneCopoeia, Rockville, MD, USA). Each genomic variant was introduced by circular PCR (QuikChange II XL Site-Directed Mutagenesis Kit, Agilent Technologies, Santa Clara, CA, USA) and each insert was sequenced on both strands to avoid any amplification artifact or inadvertent inclusion of flanking polymorphisms. Constructs were transiently transfected into Huh-7 cells (500 ng). The assay was performed at 48h, 72h or 96h after transfection, and promoter reporter gene activity in cell lysates was measured using the Secrete-Pair[™] Dual Luminescence Assay Kit (GeneCopoeia). Results were reported as the ratio of luciferase light units to secreted alkaline phosphatase, and the mean ± standard error (SE) has been illustrated in Figure 3B.

Statistics

Each reporter gene assay was performed in triplicate. A student's t test was performed to evaluate the difference between groups. In BioVU, genotype-phenotype association tests

were conducted using median lipid values as the primary endpoint, and all tests were performed using PLINK, a free, open-source genetic analysis toolset (<http://pngu.mgh.harvard.edu/~purcell/plink/>). This platform was selected based on its efficiency, flexibility and ease of application. The **-freq** option was used to calculate minor allele frequency (MAF), and the **-hardy** option was used to calculate Hardy Weinberg equilibrium (HWE). No SNPs in this study significantly deviated from HWE. A linear regression model was applied to test the relationship between each CNR1 variant and HDL-C level using the **-linear** option in PLINK. Asymptotic p-values were reported. An additive model, which fit our data best, was chosen for all analyses. We estimated the influence of covariates, including them in the linear regression model using the **-covar** option. We also reconstructed haplotypes using previously reported tag SNPs¹³. Haplotypes were estimated by applying standard E-M algorithms in the **-hap** option of PLINK. A linear regression model was applied to test each haplotype (versus all others), using the **-hap-linear** option. Only haplotypes with frequency > 5% were included in these analyses.

Acknowledgments

This work was funded by UL1 RR024975 and R01 DK080007 (Dr. Wilke). Dr. Vickers was also supported in part by NIH NHLBI Intramural Research Funds. The authors wish to thank Dr. Alan Remaley at NIH for helpful comments during the preparation of the manuscript, and Dr. Mark Rieder, in the Department of Genome Sciences at University of Washington, for oversight during the sequencing of the *CNR1* gene (Contract# 20100707). The authors also wish to thank Dr. Wei-Qi Wei, in the Department of Biomedical Informatics at Vanderbilt University, for his demographic summary of the entire EMR-linked biobank.

References

1. Degoma EM, Rader DJ. Novel HDL-directed pharmacotherapeutic strategies. *Nat Rev Cardiol.* 2011; 8:266–277. [PubMed: 21243009]
2. Gordon DJ, Rifkind BM. High-density lipoprotein—the clinical implications of recent studies. *N Engl J Med.* 1989; 321:1311–6. [PubMed: 2677733]
3. Knoblauch H, et al. Haplotypes and SNPs in 13 lipid-relevant genes explain most of the genetic variance in high-density lipoprotein and low-density lipoprotein cholesterol. *Hum Mol Genet.* 2004; 13:993–1004. [PubMed: 15044381]
4. Macgregor S, Cornes BK, Martin NG, Visscher PM. Bias, precision and heritability of self-reported and clinically measured height in Australian twins. *Hum Genet.* 2006; 120:571–80. [PubMed: 16933140]
5. Makowsky R, et al. Beyond missing heritability: prediction of complex traits. *PLoS Genet.* 2011; 7:e1002051. [PubMed: 21552331]
6. Silventoinen K, et al. Heritability of adult body height: a comparative study of twin cohorts in eight countries. *Twin Res.* 2003; 6:399–408. [PubMed: 14624724]
7. Zhang Y, et al. Obesity-related dyslipidemia associated with FAAH, independent of insulin response, in multigenerational families of Northern European descent. *Pharmacogenomics.* 2009; 10:1929–39. [PubMed: 19958092]
8. Vickers KC, Palmisano BT, Shoucri BM, Shamburek RD, Remaley AT. MicroRNAs are transported in plasma and delivered to recipient cells by high-density lipoproteins. *Nat Cell Biol.* 2011; 13:423–33. [PubMed: 21423178]
9. Pi-Sunyer FX, Aronne LJ, Heshmati HM, Devin J, Rosenstock J. Effect of rimonabant, a cannabinoid-1 receptor blocker, on weight and cardiometabolic risk factors in overweight or obese patients: RIO-North America: a randomized controlled trial. *JAMA.* 2006; 295:761–75. [PubMed: 16478899]
10. Van Gaal LF, Rissanen AM, Scheen AJ, Ziegler O, Rossner S. Effects of the cannabinoid-1 receptor blocker rimonabant on weight reduction and cardiovascular risk factors in overweight

- patients: 1-year experience from the RIO-Europe study. *Lancet*. 2005; 365:1389–97. [PubMed: 15836887]
11. Despres JP, Golay A, Sjoström L. Effects of rimonabant on metabolic risk factors in overweight patients with dyslipidemia. *N Engl J Med*. 2005; 353:2121–34. [PubMed: 16291982]
 12. Scheen AJ, Finer N, Hollander P, Jensen MD, Van Gaal LF. Efficacy and tolerability of rimonabant in overweight or obese patients with type 2 diabetes: a randomised controlled study. *Lancet*. 2006; 368:1660–72. [PubMed: 17098084]
 13. Baye TM, et al. Genetic variation in cannabinoid receptor 1 (CNR1) is associated with derangements in lipid homeostasis, independent of body mass index. *Pharmacogenomics*. 2008; 9:1647–56. [PubMed: 19018721]
 14. Feng Q, et al. A common CNR1 (cannabinoid receptor 1) haplotype attenuates the decrease in HDL cholesterol that typically accompanies weight gain. *PLoS ONE*. 2010; 5:e15779. [PubMed: 21209828]
 15. Silver HJ, et al. CNR1 Genotype Influences HDL-Cholesterol Response to Change in Dietary Fat Intake. *PLOS ONE*. 2012; 7:e36166. [PubMed: 22567136]
 16. Ritchie MD, et al. Robust replication of genotype-phenotype associations across multiple diseases in an electronic medical record. *Am J Hum Genet*. 2010; 86:560–72. [PubMed: 20362271]
 17. Roden DM, et al. Development of a large-scale de-identified DNA biobank to enable personalized medicine. *Clin Pharmacol Ther*. 2008; 84:362–9. [PubMed: 18500243]
 18. Dumitrescu L, et al. Assessing the accuracy of observer-reported ancestry in a biorepository linked to electronic medical records. *Genetics in Medicine*. 2010; 12:648–650. [PubMed: 20733501]
 19. Turner SD, et al. Knowledge-Driven Multi-Locus Analysis Reveals Gene-Gene Interactions Influencing HDL Cholesterol Level in Two Independent EMR-Linked Biobanks. *PLoS ONE*. 2011; 6:e19586. [PubMed: 21589926]
 20. Wilke RA, et al. The emerging role of electronic medical records in pharmacogenomics. *Clin Pharmacol Ther*. 2011; 89:379–386. [PubMed: 21248726]
 21. Denny JC, et al. PheWAS: demonstrating the feasibility of a phenome-wide scan to discover gene-disease associations. *Bioinformatics*. 2010; 26:1205–10. [PubMed: 20335276]
 22. Pakhomov S, Bjornsen S, Hanson P, Smith S. Quality performance measurement using the text of electronic medical records. *Med Decis Making*. 2008; 28:462–70. [PubMed: 18480037]
 23. DiPatrizio NV, Piomelli D. The thrifty lipids: endocannabinoids and the neural control of energy conservation. *Trends in Neurosciences*. 2012; 35:403–411. [PubMed: 22622030]
 24. Niswender KD. Basal insulin: beyond glycemia. *Postgrad Med*. 2011; 123:27–37. [PubMed: 21680986]
 25. Niswender KD. Basal insulin: physiology, pharmacology, and clinical implications. *Postgrad Med*. 2011; 123:17–26. [PubMed: 21680985]
 26. Stranger BE, et al. Patterns of Cis Regulatory Variation in Diverse Human Populations. *PLoS Genetics*. 2012; 8:e1002639. [PubMed: 22532805]
 27. Kent JW. Analysis of multiple phenotypes. *Genetic Epidemiology*. 2009; 33:S33–S39. [PubMed: 19924720]
 28. Fox CS, et al. Genome-Wide Association for Abdominal Subcutaneous and Visceral Adipose Reveals a Novel Locus for Visceral Fat in Women. *PLoS Genet*. 2012; 8:e1002695. [PubMed: 22589738]
 29. Tam J, et al. Peripheral CB1 cannabinoid receptor blockade improves cardiometabolic risk in mouse models of obesity. *J Clin Invest*. 2010; 120:2953–66. [PubMed: 20664173]
 30. Tam J, et al. Peripheral Cannabinoid-1 Receptor Inverse Agonism Reduces Obesity by Reversing Leptin Resistance. *Cell Metabolism*. 2012; 16:167–179. [PubMed: 22841573]
 31. Crunkhorn S. Metabolic disorders: Safe cannabinoid receptor modulators in sight? *Nature Reviews Drug Discovery*. 2012; 11:749–749.
 32. Kunos G, Tam J. The case for peripheral CB1 receptor blockade in the treatment of visceral obesity and its cardiometabolic complications. *British Journal of Pharmacology*. 2011; 163:1423–1431. [PubMed: 21434882]

33. Gary-Bobo M, et al. Rimonabant reduces obesity-associated hepatic steatosis and features of metabolic syndrome in obese Zucker fa/fa rats. *Hepatology*. 2007; 46:122–9. [PubMed: 17526015]
34. Bensaid M, et al. The cannabinoid CB1 receptor antagonist SR141716 increases Acp30 mRNA expression in adipose tissue of obese fa/fa rats and in cultured adipocyte cells. *Mol Pharmacol*. 2003; 63:908–14. [PubMed: 12644592]
35. Badellino KO, Wolfe ML, Reilly MP, Rader DJ. Endothelial lipase concentrations are increased in metabolic syndrome and associated with coronary atherosclerosis. *PLoS Med*. 2006; 3:e22. [PubMed: 16354105]
36. Edmondson AC, et al. Loss-of-function variants in endothelial lipase are a cause of elevated HDL cholesterol in humans. *J Clin Invest*. 2009; 119:1042–50. [PubMed: 19287092]
37. de Miguel-Yanes JM, et al. Variants at the Endocannabinoid Receptor CB1 Gene (CNR1) and Insulin Sensitivity, Type 2 Diabetes, and Coronary Heart Disease. *Obesity*. 2011; 19:2031–2037. [PubMed: 21633404]
38. Wei WQ, et al. Impact of data fragmentation across healthcare centers on the accuracy of a high-throughput clinical phenotyping algorithm for specifying subjects with type 2 diabetes mellitus. *J Am Med Inform Assoc*. 2012; 19:219–24. [PubMed: 22249968]
39. Wei WQ, Leibson CL, Ransom JE, Kho AN, Chute CG. The absence of longitudinal data limits the accuracy of high-throughput clinical phenotyping for identifying type 2 diabetes mellitus subjects. *Int J Med Inform*. 2012 S1386-5056(12)00109-8 [pii] 10.1016/j.ijmedinf.2012.05.015.
40. Wilke RA, et al. Quantification of the clinical modifiers impacting high-density lipoprotein cholesterol in the community: Personalized Medicine Research Project. *Prev Cardiol*. 2010; 13:63–8. [PubMed: 20377807]
41. Peissig P, et al. Construction of atorvastatin dose-response relationships using data from a large population-based DNA biobank. *Basic Clin Pharmacol Toxicol*. 2007; 100:286–8. [PubMed: 17371534]
42. Wilke RA, et al. Characterization of low-density lipoprotein cholesterol-lowering efficacy for atorvastatin in a population-based DNA biorepository. *Basic Clin Pharmacol Toxicol*. 2008; 103:354–9. [PubMed: 18834356]
43. Feng Q, et al. Human S-adenosylhomocysteine hydrolase: common gene sequence variation and functional genomic characterization. *J Neurochem*. 2009; 110:1806–17. [PubMed: 19619139]

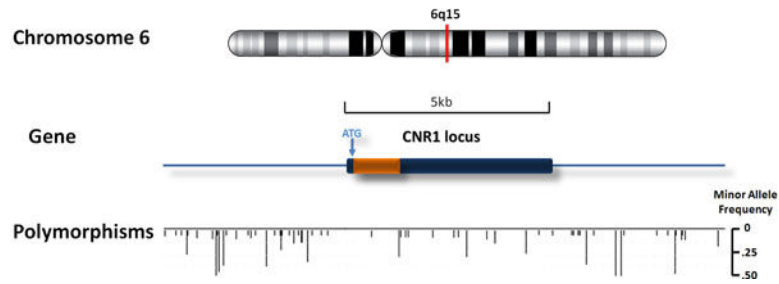


Figure 1. Location of polymorphisms identified from re-sequencing

The figure illustrates chromosome location of the *CNR1* gene (upper panel), the region of re-sequencing (middle panel), and the locations as well as minor allele frequencies of all polymorphisms (lower panel).

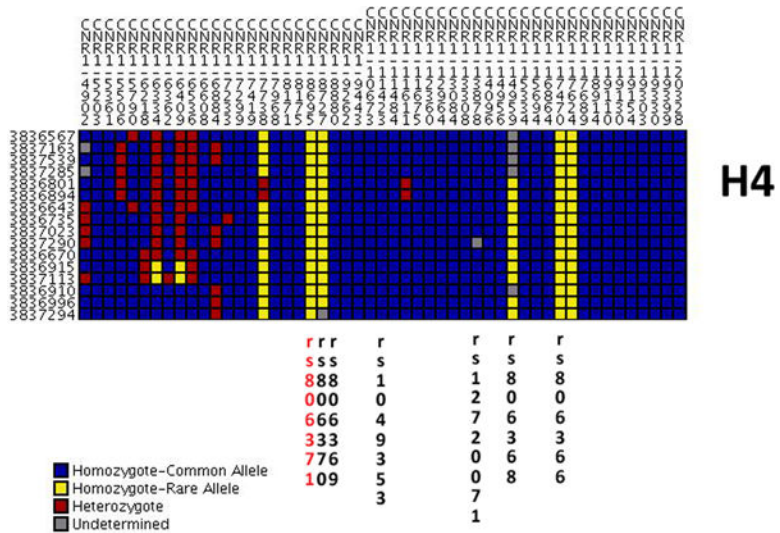


Figure 2. Individual CNR1 genotypes display visually for all H4 homozygotes
 The figure illustrates individual genotypes of all H4 homozygotes in the BioVU study cohort. The figure was generated using publicly available data on Genome Variation Server 134 (<http://gvs.gs.washington.edu/GVS134/>).

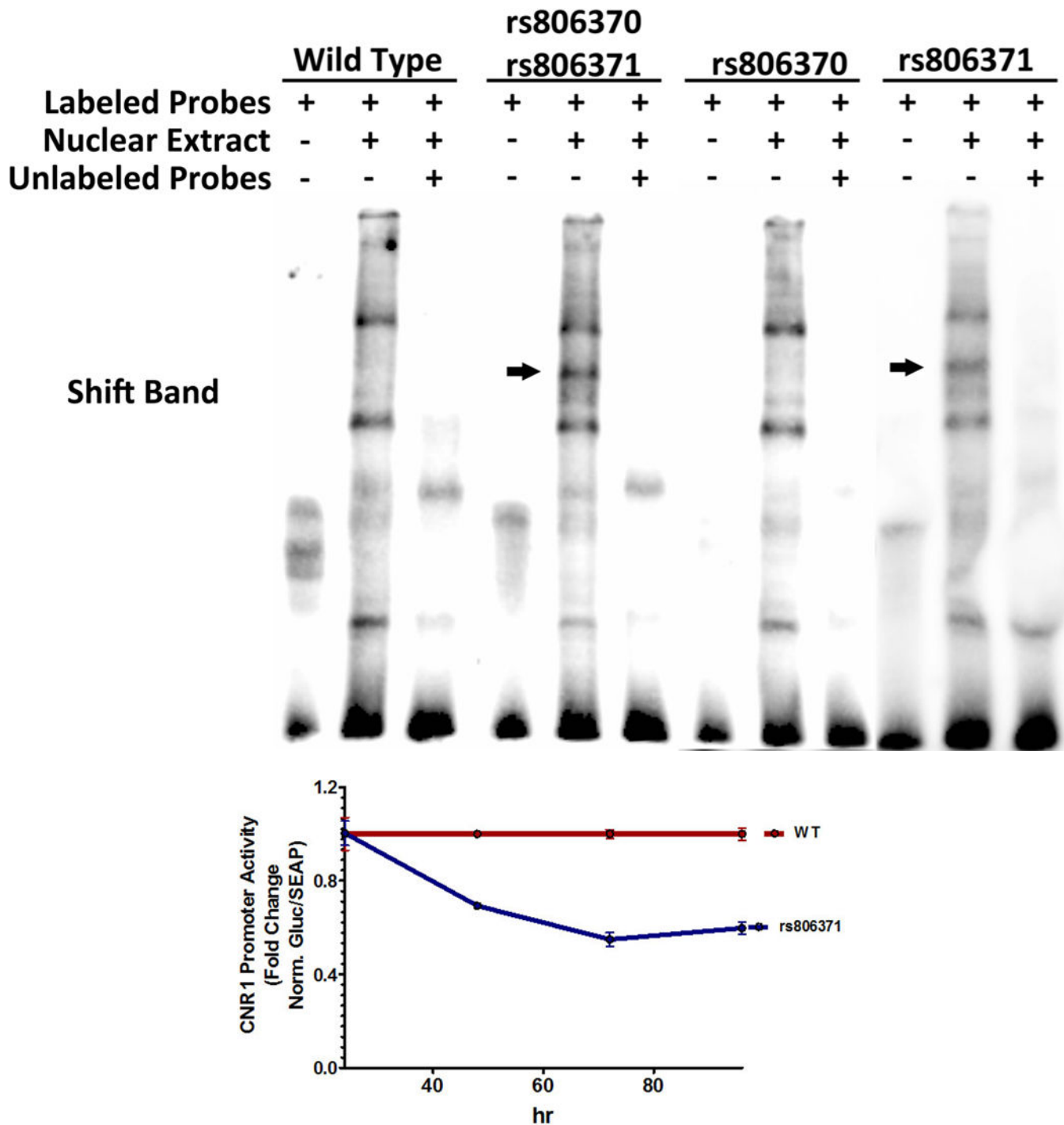


Figure 3. rs806371 is a functional variant in the *CNR1* promoter

(A) The figures illustrate DNA-protein binding patterns analyzed by electrophoretic mobility shift assay. Biotin-labeled double stranded probes were incubated with nuclear extract from HEK293 cells. (Left) A diplotype probe carrying both rs806370 and rs806371 was compared to wild type (WT) probe. (Right) DNA-protein binding patterns were compared between WT probe, and single variant probes carrying only rs806370 or

rs806371. The additional shift bands are indicated by arrows. (B) The figure illustrates normalized fold change for *CNR1* promoter activity *in vitro*. Reporter gene constructs, with or without rs806371, were transiently transfected into Huh-7 cells. Luciferase activities were measured at 48h, 72h and 96h. Results represent mean \pm S.E. for experiments run in triplicate.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 1

Demographic characteristics of BioVU study cohort.

	Mean \pm SD (units)	
	Study Cohort	Entire EMR
	N=1006	N>180,000*
Age	57.7 \pm 6.6 (years)	56.5 \pm 16.7 (years)
BMI	29.5 \pm 6.7 (kg/m ²)	30.2 \pm 13.0 (kg/m ²)
HDL-C	52.9 \pm 16.4 (mg/dl)	50.8 \pm 18.5 (mg/dl)
total chol	196.2 \pm 32.5 (mg/dl)	191 \pm 49.5 (mg/dl)
LDL-C	111.7 \pm 28.4 (mg/dl)	108.5 \pm 38.9 (mg/dl)
TG	150.0 \pm 84.2 (mg/dl)	171.5 \pm 148.4 (mg/dl)
Glucose	103.2 \pm 26.6 (mg/dl)	121.4 \pm 61.9 (mg/dl)
SBP	127.2 \pm 10.7 (mmHg)	127.2 \pm 18.9 (mmHg)
DBP	77.7 \pm 6.7 (mmHg)	75.2 \pm 14.3 (mmHg)

* Demographics for 180,000 adults (age \geq 18 years) with at least one HDL-C record, within an electronic medical record representing 1,600,000 unique individuals

Table 2

CNR1 single SNP associations with HDL-C in the Bio VU study cohort.

SNP	Variant (Minor/Major) (deletion)/AG	Minor Allele Frequency	Reference SNP ID**	P-value***	
				unadjusted	adjusted by BMI
CNR1-4902	(deletion)/AG	0.06631	N/A	0.6068	0.4279
CNR1-5203	G/A	0.001998	N/A	0.6533	0.7652
CNR1-5361	A/G	0.01798	N/A	0.5942	0.5466
CNR1-5506	A/G	0.2563	rs806378	0.3561	0.2777
CNR1-5790	A/G	0.08741	N/A	0.3753	0.6704
CNR1-6218	A/G	0.118	N/A	0.5129	0.6457
CNR1-6334	A/G	0.493	rs806377	0.7767	0.9658
CNR1-6362	G/A	0.004496	N/A	0.9105	0.6698
CNR1-6409	G/A	0.483	rs806376	0.624	0.8088
CNR1-6536	A/T	0.4081	rs806375	0.6623	0.876
CNR1-6608	C/A	0.006006	N/A	0.8088	0.7562
CNR1-6884	A/(deletion)	0.08184	rs12720072	0.1351	0.1718
CNR1-7233	C/A	0.00249	rs12195101	0.8439	0.7408
CNR1-7299	A/G	0.09114	N/A	0.6155	0.6878
CNR1-7419	C/T	0.007493	N/A	0.7738	0.9901
CNR1-7738	G/A	0.3544	rs806374	0.6103	0.7422
CNR1-8695	C/A	0.1329	rs806371	0.02598	0.01631
CNR1-8727	T/C	0.1271	rs806370	0.01654	0.01217
CNR1-8880	T/C	0.2734	rs806369	0.3119	0.2901
CNR1-9262	T/C	0.001505	N/A	0.9209	0.8586
CNR1-9443	T/C	0.000498	N/A	0.7634	0.5541
CNR1-11423	T/C	0.2936	rs1049353	0.7811	0.9505
CNR1-11484	C/G	0.001996	N/A	0.3331	0.7278
CNR1-11611	T/C	0.006993	N/A	0.09541	0.05483
CNR1-11675	T/G	0.000999	rs16880260	0.05435	0.03001

SNP	Variant (Minor/Major)	Minor Allele Frequency	Reference SNP ID**	P-value***	
				unadjusted	adjusted by BMI
CNR1-12964	A/(deletion)	0.08392	N/A	0.8841	0.83
CNR1-13084	T/A	0.003984	N/A	0.3	0.3883
CNR1-13308	T/C	0.2922	rs4707436	0.7135	0.9288
CNR1-13878	C/T	0.08026	rs12720071	0.6458	0.575
CNR1-14096	A/C	0.1289	rs45516291	0.4158	0.5551
CNR1-14956	T/C	0.005994	N/A	0.7128	0.507
CNR1-14959	C/T	0.1933	rs806368	0.1228	0.06797
CNR1-15334	C/A	0.00201	rs7738931	0.2418	0.3015
CNR1-15694	G/A	0.006993	rs12189668	0.591	0.3195
CNR1-16864	C/A	0.001002	N/A	0.6556	0.3913
CNR1-17470	C/T	0.4886	rs806366	0.03785	0.05556
CNR1-17624	A/G	0.4461	rs7766029	0.5198	0.3788
CNR1-17689	T/G	0.02806	N/A	0.4653	0.3185
CNR1-18914	G/A	0.001998	rs16880218	0.2435	0.3031
CNR1-19110	A/G	0.3965	rs806365	0.2964	0.3684
CNR1-19130	T/C	0.001002	N/A	0.7845	0.4773
CNR1-19154	C/T	0.000999	N/A	0.05435	0.03001
CNR1-19303	T/C	0.07934	N/A	0.8671	0.9349
CNR1-19399	G/A	0.07958	N/A	0.9376	0.9816
CNR1-20328	G/C	0.1614	rs35951010	0.2881	0.2428

* All variants have been named according to their nucleotide position within our region re-sequenced. SNPs showing monomorphism have been removed from this table.

** When available, we present rsNumber from dbSNP.

*** An additive model was used to calculate P-values in PLINK. P-values less than 0.05 are highlighted in red.

Effect size for HDL-C (mg/dl) based on haplotypes reconstructed using tag SNPs from our prior work in families⁽¹³⁾.

Table 3

Haplotype*	Frequency in BioVU cohort	Association with HDL-C (<i>p-value</i> **)	Haplotype copy number (mean \pm SD, mg/dl)		
			0	1	2
H1	0.29	0.7273	52.80 \pm 16.47	53.02 \pm 16.58	51.92 \pm 14.76
H2	0.27	0.2699	52.65 \pm 15.68	52.38 \pm 17.34	56.51 \pm 15.47
H3	0.22	0.3366	52.42 \pm 16.26	53.46 \pm 16.85	53.35 \pm 14.00
H4	0.12	0.01704	53.47 \pm 16.30	50.85 \pm 16.76	48.03 \pm 12.52
H5	0.07	0.8467	52.80 \pm 16.33	52.83 \pm 16.85	54.30 \pm 15.76

* TagSNPs used to defined haplotype from 5' to 3': rs806370|rs806369|rs1049353|rs1272007|rs806368|rs806366

** An additive model was used to calculate P-values in PLINK. P-values less than 0.05 are highlighted in red.