A common functional promoter variant links \textit{CNR1} gene expression to HDL cholesterol level

Q. Feng\textsuperscript{1}, K.C. Vickers\textsuperscript{1,2}, M.P. Anderson\textsuperscript{2}, M.G. Levin\textsuperscript{2}, W. Chen\textsuperscript{1}, D.G. Harrison\textsuperscript{1} & R.A. Wilke\textsuperscript{1}

Type 1 cannabinoid receptor blockers increase high-density lipoprotein cholesterol levels. Although genetic variation in the type 1 cannabinoid receptor—encoded by the \textit{CNR1} gene—is known to influence high-density lipoprotein cholesterol level as well, human studies conducted to date have been limited to genetic markers such as haplotype-tagging single nucleotide polymorphisms. Here we identify rs806371 in the \textit{CNR1} promoter as the causal variant. We re-sequence the \textit{CNR1} gene and genotype all variants in a DNA biobank linked to comprehensive electronic medical records. By testing each variant for association with high-density lipoprotein cholesterol level in a clinical practice-based setting, we localize a putative functional allele to a 100-bp window in the 5\textsuperscript{\textprime}-flanking region. Assessment of variants in this window for functional impact on electrophoretic mobility shift assay identifies rs806371 as a novel regulatory binding element. Reporter gene assays confirm that rs806371 reduces gene expression, thereby linking \textit{CNR1} gene variation to high-density lipoprotein cholesterol level in humans.
Clinical lipid disorders have enormous public health significance and increasing societal burden in developed countries. High-density lipoprotein cholesterol (HDL-C) levels are inversely correlated with cardiovascular disease, and each 1 mg dl$^{-1}$ decrease in HDL-C is associated with a 6% increase in adverse clinical events. As HDL-C levels are highly heritable (H$^2$ ranging from $\sim 0.4$–$0.7$), there is great interest in characterizing the genetic architecture underlying this important complex trait.

The type 1 cannabinoid receptor (CB$_1$) is a novel therapeutic target for controlling lipoprotein metabolism. As CB$_1$ receptors within the brain influence eating behavior, rimonabant, a selective CB$_1$ antagonist, was initially designed to correct weight gain. However, in large clinical trials, rimonabant improved HDL-C within the brain influence eating behavior, rimonabant, a selective CB$_1$ antagonist, was initially designed to correct weight gain. However, in large clinical trials, rimonabant improved HDL-C levels far greater than originally anticipated. Likewise, variation in the CNR1 gene, which transcribes the CB$_1$ receptor, has previously been associated with HDL-C levels in several independent study cohorts. We have previously reported that a common CNR1 haplotype (H4, frequency $\sim 15\%$ in the general population) is associated with decreased HDL-C levels, independent of body mass index (BMI).

Interestingly, common non-synonymous single nucleotide polymorphisms (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 their association with HDL-C levels in vivo, using the largest clinical practice-based DNA biobank in the United States.

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.

### Table 1 | Demographic characteristics of BioVU study cohort.

| Mean ± s.d. (units) | Study cohort (N = 1,006) | Entire EMR (N = 180,000)* |
|---------------------|--------------------------|---------------------------|
| Age                | 57.7 ± 6.6 (years)       | 56.5 ± 16.7 (years)       |
| BMI                 | 29.5 ± 6.7 (kg m$^{-2}$) | 30.2 ± 13.0 (kg m$^{-2}$) |
| HDL-C              | 52.9 ± 16.4 (mg dl$^{-1}$) | 50.8 ± 18.5 (mg dl$^{-1}$) |
| Total chol         | 196.2 ± 32.5 (mg dl$^{-1}$) | 191 ± 49.5 (mg dl$^{-1}$) |
| LDL-C              | 111.7 ± 28.4 (mg dl$^{-1}$) | 108.5 ± 38.9 (mg dl$^{-1}$) |
| TG                 | 150.0 ± 84.2 (mg dl$^{-1}$) | 171.5 ± 148.4 (mg dl$^{-1}$) |
| Glucose            | 103.2 ± 26.6 (mg dl$^{-1}$) | 121.4 ± 61.9 (mg dl$^{-1}$) |
| SBP                | 127.2 ± 10.7 (mm Hg)     | 127.2 ± 18.9 (mm Hg)      |
| DBP                | 77.7 ± 6.7 (mm Hg)       | 75.2 ± 14.3 (mm Hg)       |

*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.

### Results

#### Identification of putative functional variants.

BioVU is a clinical practice-based biobank linked to comprehensive electronic medical records (EMRs). The largest resource of its kind (n = 157,719 on 10 January 2013), BioVU is robust in its ability to replicate genetic associations previously identified in disease-based cohorts. Our previous analysis identified a CNR1 haplotype associated with HDL-C levels in extended families of Northern European ancestry. To assess this relationship within the community, we sampled 1% of 100,000 BioVU subject records claiming European Ancestry for further study (50% female and 50% male subjects). The precision of ancestry data within BioVU has previously been validated using a panel of 360 ancestry informative markers. 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 (15 kb 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) (Fig. 1, Table 2). A total of 65 polymorphisms were identified, including 62 SNPs and 3 insertion/deletions. Thirty-seven of these polymorphisms had not been reported previously (that is, did not have existing rsNumbers). All observed variants were re-genotyped in the BioVU sub-cohort from Table 1 (n = 1,006) using our high-throughput Sequenom platform (see Methods). As described in our previous study using clinical data derived from EMRs, the primary endpoint in the BioVU sub-cohort was median outpatient HDL-C levels, adjusted for age and gender. Other important clinical covariates were either extracted directly from the EMRs (for example, BMI nearest to the date on which each median HDL-C level was extracted) or defined using time stamps through natural language processing (for example, exposure to medications known to alter HDL-C levels).

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 30 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 1,006 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. As 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 ± s.d. was 53.47 ± 16.30, 50.85 ± 16.76 and 48.03 ± 12.52 mg dl$^{-1}$ for carriers of 0, 1 and 2 copies of the H4 risk haplotype; and HDL-C mean ± s.d. was 53.47 ± 16.45, 51.13 ± 16.64 and 48.23 ± 12.60 mg dl$^{-1}$ 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

| SNP* | Variant (minor/major) | 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 | 0.18 | 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 | 0.6998 |
| 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 | 0.0217 |
| CNR1-8880 | T/C | 0.2734 | rs806369 | 0.319 | 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 | 0.781 | 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 |
| CNR1-12964 | A/(deletion) | 0.08392 | N/A | 0.8841 | 0.83 |
| CNR1-13084 | T/A | 0.003984 | N/A | 0.3 | 0.3883 |
| 130308 | T/C | 0.2922 | rs4707436 | 0.7153 | 0.9288 |
| 13878 | T/C | 0.08026 | rs12720071 | 0.6458 | 0.575 |
| 14096 | A/C | 0.1289 | rs45516291 | 0.4158 | 0.5551 |
| 14956 | T/C | 0.005994 | N/A | 0.7128 | 0.507 |
| 14959 | C/T | 0.193 | rs806368 | 0.1228 | 0.06797 |
| 15334 | C/A | 0.00201 | rs7738931 | 0.2418 | 0.3015 |
| 15694 | G/A | 0.006993 | rs12189668 | 0.591 | 0.3195 |
| 16864 | C/A | 0.001002 | N/A | 0.6556 | 0.3913 |
| 17470 | C/T | 0.4886 | rs806366 | 0.03785 | 0.05556 |
| 17624 | A/G | 0.4461 | rs7766029 | 0.5198 | 0.3788 |
| 17689 | T/G | 0.02806 | N/A | 0.4653 | 0.3185 |
| 18914 | G/A | 0.001998 | rs16880218 | 0.2435 | 0.3031 |
| 19110 | A/G | 0.3965 | rs806365 | 0.2964 | 0.3684 |
| 19130 | T/C | 0.001002 | N/A | 0.7845 | 0.4773 |
| 19154 | C/T | 0.000999 | N/A | 0.05435 | 0.03001 |
| 19303 | T/C | 0.07934 | N/A | 0.8671 | 0.9349 |
| 19399 | G/A | 0.07958 | N/A | 0.9376 | 0.9816 |
| 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.

**The variant found to be functional is highlighted in bold type.

1When available, we present rsNumber from dbSNP.

2An additive model was used to calculate P-values in PLINK. P-values < 0.05 are in italics.
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). As these observations suggest that both variants may contribute to the association, we screened all 5 variants in the CNR1 gene to determine if nucleotide substitutions disrupt or induce any regulatory elements in the CNR1 promoter (Fig. 2). This variant was frequently present on the H4 background. Although CNR1-7738, another CNR1 5′-FR variant, was also found to have increased prevalence and not robust \( (r^2 = 0.139) \). As such, we did not pursue CNR1-7738 in our subsequent functional studies.

Table 3 | Effect size for HDL-C (mg dl \(^{-1}\)) based on haplotypes reconstructed using tag SNPs from our prior work in families\(^{13}\).

| Haplotype* | Frequency in BioVU cohort | Association with HDL-C (P-value) | Haplotype copy number (mean ± s.d., mg dl \(^{-1}\)) |
|------------|--------------------------|---------------------------------|-----------------------------------------------|
| H1         | 0.29                     | 0.7273                          | 52.80 ± 16.47                                  |
| H2         | 0.27                     | 0.2699                          | 52.65 ± 15.68                                  |
| H3         | 0.22                     | 0.3366                          | 52.42 ± 16.26                                  |
| H4         | 0.12                     | 0.0704                          | 53.47 ± 16.30                                  |
| H5         | 0.07                     | 0.8467                          | 52.80 ± 16.33                                  |

*TagSNPs used to define haplotype from 5′-3′: rs806370|rs806369|rs1049353|rs12720071|rs806368|rs806366.

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 in vitro studies that indicate only rs806371 introduces a novel DNA-binding site for nuclear proteins, based upon subsequent EMSA analyses conducted with probes for single variants (Fig. 2).
**Figure 3 | rs806371 is a functional variant in the CNR1 promoter.** (a) The figures illustrate DNA-protein binding patterns analyzed by EMSA. Biotin-labeled double-stranded probes were incubated with nuclear extract from HEK293 cells. (Left) A diplotype probe carrying both rs806370 and rs806371 was compared with 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 48, 72 and 96 h. Results represent mean ± s.e. for experiments run in triplicate.

**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. 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±17.3 mg dl⁻¹, mean ± s.d.) in the absence of an obvious clinical explanation for abnormal lipid homeostasis. HDL-C, whereas CNR1 directly influences HDL-C levels in multi-generational families. To narrow our search for the causal variant underlying this association, a 15-kb 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 EMRs. In recent years, the use of EMRs has expanded rapidly (in response to the Affordable Care Act) creating huge longitudinal data sets ideal for observational research. Even though these two SNPs are co-inherited, our data clearly indicate that rs806371, not rs806370, alters nuclear protein binding in an EMSA. As 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 (for example, within brain versus adipose tissue) may explain the fact that CB1 receptor antagonists increase HDL-C, whereas CNR1 loss-of-function variants decrease HDL-C. For example, insulin-dependent signaling in the brain engages different second messengers than peripheral tissues. The 5′-end; for example, the minor allele at rs806371 may disrupt a non-canonical guanylyltransferase receptor (TRANSFAC V8.3 at http://alggen.lsi.upc.es/cgi-bin/promo_v3/promo/promo.cgi?dirDB=TF_8.3&alledBy=alggen). Publically available eQTL data sets indicate that this locus also regulates the nearby RNGTT gene (P = 0.0273), in HapMap Europeans. RNGTT is located just upstream of CNR1 (~400 kb), and it encodes an RNA guanylyltransferase recently associated with
subcutaneous adipose tissue volume in women from the Framingham Offspring Study26.

Clearly, the pathophysiologic 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)27–29 and indirect (CNR1 expression in the brain)30. Non-brain-penetrant CB1 receptor antagonists directly increase plasma HDL-C levels in animal models26, and CB1 blockade modulates the release of adipokines from human adipocytes27. As these processes influence the intravascular remodeling of HDL particles in vivo23–35, CNR1 gene variability may influence public health by altering cardiovascular risk in the context of the current obesity epidemic36.

Methods

Study population. The current study was conducted in accordance with the Principles outlined in the Declaration of Helsinki. Approval was obtained from the Institutional Review Board of Vanderbilt University. The Vanderbilt DNA biobank (BioVU) currently contains EMRs from 157,719 subjects (updated 10 January 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 ~ 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 provider37. 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 1,006 representative subjects were selected, including 509 females and 497 males.

Phenotyping. Outpatient median HDL-C levels represented our primary endpoint, available on all 1,006 subjects. Clinical lipid data were extracted from EMRs. These data reflected longitudinal lipid data (> 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 as acute illness typically influences circulating lipid levels in most inpatients38. Data on lipid medications and related events (drug, dose, date and time) were also obtained from EMRs by applying our nature language processing algorithms39,40. 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 was calculated for each subject using median height and weight at the time stamp for median lipid values. When computing 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). As our inclusion criteria required dense longitudinal data (that is, 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 1,006 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/dl1), and the mean ± s.d. has been illustrated in Fig. 3b.

Sequencing. The CNR1 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, New Jersey, USA)42. A total of 15 kb genomic region was sequenced, including 5 kb upstream of the gene, 5 kb 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 300–750 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 with the whole-genome sequences in independent experiments and rule out the possibility of allele-specific 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 with the whole-genome sequences in independent experiments and rule out the possibility of allele-specific 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 with the whole-genome sequences in independent experiments and rule out the possibility of allele-specific regions.
References

1. Degoma, E. M. & Rader, D. J. Novel HDL-directed pharmacotherapeutic strategies. Nat. Rev. Cardiol. 8, 266–277 (2011).
2. Gordon, D. J. & Rifkind, B. M. High-density lipoprotein—the clinical implications of recent studies. N. Engl. J. Med. 321, 1311–1316 (1989).
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. Hepat. Mol. Genet. 13, 993–1004 (2004).
4. Macgregor, S., Cornes, B. K., Martin, N. G. & Visscher, P. M. Bias, precision and heritability of self-reported and clinically measured height in Australian twins. Hum. Genet. 120, 571–580 (2006).
5. Makowsky, R. et al. Beyond missing heritability: prediction of complex traits. PLoS Genet. 7, e1002051 (2011).
6. Silverstein, K. E. et al. Heritability of adult body height: a comparative study of twin cohorts in eight countries. Twin Res. 6, 399–408 (2003).
7. Zhang, Y. et al. Obesity-related dyslipidemia associated with FAAH, independent of insulin response, in multigenerational families of Northern European descent. Pharmacogenomics 10, 1929–1939 (2009).
8. Vickers, K. C., Palmisano, B. T., Shoucr, B. M., Shamburek, R. D. & Remaley, A. T. MicroRNAs are transported in plasma and delivered to recipient cells by high-density lipoproteins. Nat. Cell Biol. 13, 423–433 (2011).
9. Pi-Sunyer, F. X., Aronne, L. J., Heshmati, H. M., Devin, J. & Rosenstock, J. Effect of rimonabant, a cannabinoid-1 receptor blocker, on weight and cardio metabolic risk factors in overweight or obese patients: RIO-North America: a randomized controlled trial. JAMA. 295, 761–775 (2006).
10. Van Gaal, L. F., Rissanen, A. M., Scheen, A. J., 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 365, 1389–1397 (2005).
11. Despres, J. P., Golay, A. & Sjostrom, L. Effects of rimonabant on metabolic risk factors in overweight patients with dyslipidemia. N. Engl. J. Med. 353, 2121–2134 (2005).
12. Scheen, A. J., Finer, N., Hollandier, P., Jensen, M. D. & Van Gaal, L. F. Efficacy and tolerability of rimonabant in overweight or obese patients with type 2 diabetes: a randomised controlled study. Lancet 368, 1660–1672 (2006).
13. Baye, T. M. et al. Genetic variation in cannabinoid receptor 1 (CNR1) is associated with derangements in lipid homeostasis, independent of body mass index. Pharmacogenomics 9, 1647–1656 (2008).
14. Feng, Q. et al. A common CNR1 (cannabinoid receptor 1) haplotype attenuates the decrease in HDL cholesterol that typically accompanies weight loss. PLoS One 5, e15779 (2010).
15. Silver, H. J. et al. CNR1 genotype influences HDL-cholesterol response to change in dietary fat intake. PLOS One 7, e36166 (2012).
16. Ritchie, M. D. et al. Robust replication of genotype-phenotype associations across multiple diseases in an electronic medical record. Am. J. Hum. Genet. 86, 560–572 (2010).
17. Roden, D. M. et al. Development of a large-scale de-identified DNA biobank to enable personalized medicine. Clin. Pharmacol. Ther. 84, 362–369 (2008).
18. Dumitrescu, L. et al. Assessing the accuracy of observed-ancestry ancestry in a biorepository linked to electronic medical records. Genet. Med. 12, 648–650 (2010).
19. Turner, S. D. et al. Knowledge-driven multi-locus analysis reveals gene-gene interactions influencing HDL cholesterol level in two independent emr-linked biobanks. PLoS One 6, e19586 (2011).
20. Wilke, R. A. et al. The emerging role of electronic medical records in phenotyping for identifying type 2 diabetes mellitus subjects. Int. J. Med. Inform. 82, 239–247 (2012).
21. Wilke, R. A. et al. Quantification of the clinical modifiers impacting high-density lipoprotein cholesterol in the community: Personalized Medicine Research Project. Prev. Cardiol. 13, 63–68 (2010).
22. Peissig, P. et al. Construction of atovastatin dose-response relationships using data from a large population-based DNA biobank. Basic Clin. Pharmacol. Toxicol. 100, 286–288 (2007).
23. Wilke, R. A. et al. Characterization of low-density lipoprotein cholesterol-lowering efficacy for atorvastatin in a population-based DNA biorepository. Basic Clin. Pharmacol. Toxicol. 103, 354–359 (2008).
24. Rieder, M. J. et al. NHLBI Resequecing and Genotyping Service (RSG). Department of Genome Sciences, Seattle, Washington, USA. UW HV4814 (2010).
25. Feng, Q. et al. Human S-adenosylhomocysteine hydrolase: common gene sequence variation and functional genomic characterization. J. Neurochem. 110, 1806–1817 (2009).

Acknowledgements

This work was funded by U11 RO024957 and ROI DK080007 (Dr Wilke). Dr Vickers was also supported in part by NIH NHLBI Intramural Research Funds. We 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 UW HV4814. We 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.

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.

Additional information

Accession codes: CNR1 SNP data have been deposited in Genbank under accession code KF042592.

Competing financial interests: The authors declare no competing financial interests.

Reprints and permission information is available online at http://npg.nature.com/reprintsandpermissions/

How to cite this article: Feng, Q. et al. A common functional promoter variant links CNR1 gene expression to HDL cholesterol level. Nat. Commun. 4:1973 doi: 10.1038/ncomms2973 (2013).