

# Novel gene-by-environment interactions: *APOB* and *NPC1L1* variants affect the relationship between dietary and total plasma cholesterol<sup>[S]</sup>

Daniel S. Kim,<sup>\*,†</sup> Amber A. Burt,<sup>\*</sup> Jane E. Ranchalis,<sup>\*</sup> Ella R. Jarvik,<sup>\*</sup> Elisabeth A. Rosenthal,<sup>\*</sup> Thomas S. Hatsukami,<sup>§</sup> Clement E. Furlong,<sup>\*,†</sup> and Gail P. Jarvik<sup>1,\*,†</sup>

Department of Medicine, Division of Medical Genetics,<sup>\*</sup> Department of Genome Sciences,<sup>†</sup> and Department of Surgery, Division of Vascular Surgery,<sup>§</sup> University of Washington School of Medicine, Seattle, WA

**Abstract** Cardiovascular disease (CVD) is the leading cause of death in developed countries. Plasma cholesterol level is a key risk factor in CVD pathogenesis. Genetic and dietary variation both influence plasma cholesterol; however, little is known about dietary interactions with genetic variants influencing the absorption and transport of dietary cholesterol. We sought to determine whether gut expressed variants predicting plasma cholesterol differentially affected the relationship between dietary and plasma cholesterol levels in 1,128 subjects (772/356 in the discovery/replication cohorts, respectively). Four single nucleotide polymorphisms (SNPs) within three genes (*APOB*, *CETP*, and *NPC1L1*) were significantly associated with plasma cholesterol in the discovery cohort. These were subsequently evaluated for gene-by-environment (GxE) interactions with dietary cholesterol for the prediction of plasma cholesterol, with significant findings tested for replication. Novel GxE interactions were identified and replicated for two variants: rs1042034, an *APOB* Ser4338Asn missense SNP and rs2072183 (in males only), a synonymous *NPC1L1* SNP in linkage disequilibrium with SNPs 5' of *NPC1L1*. This study identifies the presence of novel GxE and gender interactions implying that differential gut absorption is the basis for the variant associations with plasma cholesterol. These GxE interactions may account for part of the "missing heritability" not accounted for by genetic associations.—Kim, D. S., A. A. Burt, J. E. Ranchalis, E. R. Jarvik, E. A. Rosenthal, T. S. Hatsukami, C. E. Furlong, and G. P. Jarvik. **Novel gene-by-environment interactions: *APOB* and *NPC1L1* variants affect the relationship between dietary and total plasma cholesterol.** *J. Lipid Res.* 2013. 54: 1512–1520.

**Supplementary key words** dietary cholesterol • total cholesterol • cardiovascular disease • Niemann-Pick C1-like 1

Cardiovascular disease (CVD), including coronary heart disease and carotid artery disease (CAAD), is the leading cause of death in developed countries, accounting for approximately one-third of all deaths in adults over age 35 (1). One of the key risk factors for CVD is hypercholesterolemia (2), with epidemiologic studies demonstrating a graded relationship between total plasma cholesterol and CVD risk (3). Levels of total cholesterol and other lipids and lipoproteins are both heritable and targets for pharmacologic and lifestyle interventions (4). One of the primary methods of lifestyle intervention to alter lipid levels is through dietary changes, such as decreasing total fat intake (5, 6).

Genome-wide association studies (GWASs) have implicated numerous genetic loci for association with total cholesterol levels. A recent meta-analysis by Teslovich et al. (7) of over 100,000 individuals of European ancestry confirmed numerous previous GWAS findings and identified novel variants and loci for prediction of total lipid phenotypes, including total cholesterol levels.

While genetic evidence is accumulating for many of these loci, less is known about the specific mechanisms through which these genetic variants affect cholesterol levels. We hypothesized that one pathway through which these genetic variants could alter cholesterol homeostasis is through differential absorption or processing of dietary

This work was funded by National Institutes of Health Grant RO1 HL-67406 and a State of Washington Life Sciences Discovery Award to the Northwest Institute of Genetic Medicine. D.S.K. was supported in part by a Sarnoff Cardiovascular Research Fellowship for Medical Students Award, the Benjamin and Margaret Hall Endowed Fellowship in Genome Sciences, and National Institutes of Health Grant 5T31HG-000035-18. This work utilized resources of SeattleSNPs; National Heart, Lung, and Blood Institute Program for Genomic Applications, SeattleSNPs, Seattle, WA (<http://pga.gs.washington.edu>). The authors declare no conflicts of interest.

Manuscript received 19 December 2012 and in revised form 9 March 2013.

Published, JLR Papers in Press, March 11, 2013  
DOI 10.1194/jlr.P035238

Abbreviations: CAAD, carotid artery disease; CLEAR, Carotid Lesion Epidemiology and Risk; CVD, cardiovascular disease; eQTL, expression quantitative trait locus; GWAS, genome-wide association study; GxE, gene-by-environment; LD, linkage disequilibrium; NPC1L1, Niemann-Pick C1-like 1; SNP, single nucleotide polymorphism; TC, total cholesterol.

<sup>1</sup>To whom all correspondence should be addressed.

e-mail: pair@u.washington.edu

[S] The online version of this article (available at <http://www.jlr.org>) contains supplementary data in the form of six figures and one table.

cholesterol, thereby influencing the potential benefit of lifestyle changes in the prevention of CVD. Utilizing the Carotid Lesion Epidemiology and Risk (CLEAR) CAAD case-control cohort, we investigated whether there was any evidence of single nucleotide polymorphism (SNP)-dietary cholesterol interactions on prediction of total cholesterol levels, to better understand the potential influence of the genetic variation on the handling of dietary cholesterol and the potential downstream pathogenesis of CVD.

## METHODS

### Ethics statement

Institutional Review Boards at the University of Washington, Virginia Mason Medical Center, and Veterans Affairs Puget Sound Health Care approved the study. Written informed consent was obtained from all participants.

### Sample

The study population for these analyses consisted of 1,128 subjects (772 subjects in the discovery subset and 356 subjects in the replication subset) from the previously described CLEAR study (8–13), who had both diet and genetic data. The discovery set had Illumina HumanCVD BeadChip (14) genotypes, while the replication cohort was genotyped later for SNPs of interest, as outlined in the genotyping methods section. This cohort includes 358 CAAD cases, 639 controls, and 131 subjects of other phenotypes, including moderate carotid artery obstruction (15–49% by ultrasound), as well as coronary artery and peripheral artery disease. Only Caucasian subjects were analyzed due to under-representation of minority samples in this primarily Seattle Veterans-based cohort. Ancestry was confirmed using STRUCTURE with three ancestral groups (15). Medication use, including statins, oral hypoglycemics, and insulin injection were ascertained from pharmacy and medical records, as well as subject report. Diabetes status was determined by *a*) if a subject was on an oral hypoglycemic or insulin, or *b*) if their hemoglobin A1C was  $\geq 6.5\%$ . CLEAR cohort exclusion criteria included familial hypercholesterolemia, total fasting cholesterol  $>400$  mg/dl, hypocoagulable state and/or the use of Coumadin, post-organ transplant, or the inability to consent. Descriptive statistics for the cohort are presented in Table 1.

### Survey methods

Subjects were asked to complete the standardized Harvard food frequency questionnaire developed by The Health Professionals Follow-Up Study (<https://regepi.bwh.harvard.edu/health/nutrition.html>) at enrollment. The survey asked about: *a*) the average frequency of intake over the previous year of specified portions of 131 foods; and *b*) the use of vitamins and mineral supplements, including the dose and duration of use. Questions regarding brand of multivitamins and cereal used were asked to clarify the quantities of specific vitamin supplementation. Surveys were returned to Harvard School of Public Health and Brigham and Women's Hospital where they underwent quantitative analyses to return average daily intakes of individual micro-nutrients (e.g., total fat, total cholesterol, etc.) and vitamins calculated from food frequency data. Natural log-transformed dietary cholesterol intake data from the food frequency questionnaire was used as the variable for interaction with SNP genotypes in this study. Subjects were excluded from analyses if: *a*) their caloric intake was not between 800 and 4,200 kcal/day; or *b*)

their surveys had  $\geq 70$  blank items of a total of 131 questions. This food frequency survey has been validated against two 1 week diet records taken approximately six months apart (16). As well, the inferred intake of dietary fats has been validated against total lipid levels (17, 18).

### Total lipid levels

Lipid measurements were performed on fasting plasma. Standard enzymatic methods were utilized to determine the levels of total cholesterol (19, 20). We estimated “pre-therapy” total cholesterol levels in statin users. We based these values on 41 individuals in the CLEAR cohort that had repeat lipid measures before and after initiation of statin pharmacotherapy, and found that total cholesterol decreased by 25.1% (10). Using these previously calculated values, we imputed pretreatment total cholesterol levels in statin users ( $n = 425$ ) within the CLEAR cohort by increasing measured cholesterol levels by 25.1%. A comparison of the pre- and poststatin adjustment total cholesterol levels is presented in Table 1. These values from a subset of the CLEAR cohort are consistent with those found in a recent meta-analysis of 134,537 subjects on statins (21).

### Genotyping

SNPs for the 772 subjects comprising the discovery set were genotyped using the Illumina HumanCVD BeadChip (14). Duplicate genotyping for 34 individuals showed 99.7% consistency in calls. An additional 356 subjects comprising the replication set were genotyped by TaqMan for the SNPs in Table 3, with duplicate genotyping in 12 individuals showing 99.9% consistency in calls. SNPs were filtered with cutoffs of: minor allele frequency  $<1\%$ , Hardy-Weinberg equilibrium rejected at  $P < 10^{-4}$ , or call-rate  $<97\%$ .

### SNP selection

Due to the power considerations related to multiple testing, we limited our testing for gene-by-environment (GxE) interactions to SNPs: *a*) identified as predictive of total cholesterol by Teslovich et al. (7); *b*) also predictive for total cholesterol in our cohort; *c*) expressed in the stomach, small intestines, or pancreas; and *d*) involved in cholesterol handling (this is tracked by SNP in supplementary Table 1 and supplementary Fig. IV). Teslovich et al. (7) identified SNPs at 53 genetic loci that were significant at the genome-wide level for association with total cholesterol levels (these SNPs are hereafter referred to as “lead SNPs”). We identified relevant SNPs from our Illumina HumanCVD BeadChip genotyped data using SNAP Proxy Search to parse for SNPs that were in strong linkage disequilibrium (LD) ( $r^2 > 0.80$ ) (22) in the 1000 Genomes database with predictive SNPs reported by Teslovich et al. We identified such SNPs in our data for only 19 of their 53 identified loci. Due to the inclusion of multiple *APOB* SNPs and its physiologic relevance to cholesterol absorption and total cholesterol, we followed up on one *APOB* SNP that was primarily reported to be associated with variation in high-density lipoprotein (HDL) and triglycerides. While the effect of this SNP is not addressed in their paper, Teslovich and Kathiresan (personal communication) provided us with the *P* value for the total cholesterol effect, which was  $1.75 \times 10^{-19}$ . This added *APOB* SNP (rs1042034) was not in LD with the other identified *APOB* SNP ( $r^2 = 0.094$ ). We next utilized Tissue-Specific Gene Expression and Regulation (TIGER) (23) to examine the tissue expression of these 22 loci to exclude genes not expressed in the stomach, small intestine, or pancreas, with separate confirmation of nonhepatic gastrointestinal gene expression through the total European Molecular Biology Laboratory (EMBL) Nucleotide Sequence Database (24). We excluded genes for transcription

factors. In sum, seven SNPs from six genes [*ABCG8*, *APOB*, *APOE*, *CETP*, Niemann-Pick C1-like 1 (*NPC1L1*), and *PCSK9*] were considered (Table 2). We then tested these seven SNPs for their main effects on total cholesterol in the discovery cohort. Four SNPs (two from *APOB*, one each from *CETP* and *NPC1L1*) were significant at the  $P = 0.05$  level and were carried forward for GxE interaction testing. For SNPs for which a GxE interaction was detected with a  $P \leq 0.05$ , we performed sensitivity analyses by gender, CAAD case-control status, and statin use. We then considered other SNPs in LD with these SNPs to determine if they might be better predictors than the original SNP considered.

### Statistical analyses

Statistical analyses were conducted using the R statistical language (<http://www.r-project.org/>). Genotypes were coded in an additive model. Extreme outliers for both dietary and total cholesterol were Winsorized to values three standard deviations from the mean (25, 26). The positively skewed distribution for dietary cholesterol was adjusted by using a natural log transformation. Hereafter, "dietary cholesterol" will refer to  $\ln(\text{dietary cholesterol})$ .

In order to adjust for the effects of nonstatin confounding covariates on total cholesterol, we first used linear regression to adjust total cholesterol levels for age, sex, and diabetes status, thereby obtaining a residual value of total cholesterol. Residual values of adjusted total cholesterol from this analysis were then utilized as the outcome variable to test for interactions between the four SNPs tested and dietary cholesterol intake through further linear regression. SNPs were considered under an additive genetic model, coded as 0, 1, and 2. Multiplicative interactions were modeled. Due to power considerations, no Bonferroni correction was utilized. Rather, we accepted a nominal  $P \leq 0.05$  and relied on replication to avoid false positives. Plotting of genotype-dependent relationships between dietary and adjusted total cholesterol levels was performed using the R statistical language.

The equations in the Fig. 1 legend represent the genotype-specific regression equations for predicted residual total cholesterol level for a given value of dietary cholesterol. These were obtained via simplification of our interaction equation [Predicted residual total cholesterol =  $\alpha + \beta_1$  (SNP) +  $\beta_2$  (dietary cholesterol) +  $\beta_3$  (SNP  $\times$  dietary cholesterol)] such that genotype classes 0, 1, and 2 have the intercept and regression coefficients  $\alpha + \beta_2$  (dietary cholesterol),  $(\alpha + \beta_1) + (\beta_2 + \beta_3)$  (dietary cholesterol), and  $(\alpha + 2\beta_1) + (\beta_2 + 2\beta_3)$  (dietary cholesterol), respectively. Total cholesterol residuals were relocated to the mean value of total cholesterol by the addition of 207.97 to aid in interpretation of the interaction plots.

## RESULTS

Demographic, clinical, dietary, and total cholesterol measures for the cohort are presented in Table 1. The pooled sample was composed of 1,128 subjects with a mean age of 66.1 years, of which 67.4% were males. Subjects with existing Illumina CVD chip genotype data defined the discovery subset; the replication subset consisted of the remaining subjects with food frequency questionnaire data, which we then genotyped for SNPs that were significant in GxE discovery analyses. The independent discovery ( $n = 772$  subjects) and replication ( $n = 356$  subjects) subsets differed significantly, as the replication subset was composed of 65.7% female subjects whereas the discovery subset was only 17.4% female ( $P < 0.001$ ). The overall

sample included 358 cases with  $>50\%$  carotid stenosis, 57 subjects with 15–49% carotid stenosis, and 639 controls with  $<15\%$  stenosis bilaterally. Within the discovery subset, there were a significantly higher proportion of cases (39.9% vs. 14.0%) compared with the replication subset ( $P < 0.001$ ). The rates of diabetes and statin use were 12.4% and 37.7% overall. The rates of statin use were significantly higher in the discovery (42.7% vs. 25.0%) compared with the replication subset ( $P < 0.001$ ). Dietary and total cholesterol were highly correlated in the pooled cohort, with a Pearson pair-wise correlation coefficient,  $r = 0.80$  and  $P < 0.001$ . The outcome of residual total cholesterol differed by statin use (supplementary Fig. I), but not by CAAD case-control status or sex (supplementary Figs. II and III, respectively).

The SNP selection process is described in supplementary Table I and supplementary Fig. IV. Seven SNPs from six genes (*ABCG8*, *APOB*, *APOE*, *CETP*, *NPC1L1*, and *PCSK9*) were considered. Descriptions of the seven lead SNPs [identified by Teslovich et al. (7) and then selected for utilizing the criteria outlined in the "SNP selection" section of the Methods] spanning the six candidate genes studied are provided in Table 2. The two lead SNPs within *APOB* are not in strong LD with each other ( $r^2 = 0.094$ ). *APOB*, the apolipoprotein B structural locus, and *NPC1L1* each had multiple proxy SNPs genotyped by the Illumina HumanCVD BeadChip that were in strong LD with the lead SNPs identified by Teslovich et al. [(7) and unpublished data] within the discovery subset of CLEAR (with *APOB* SNP rs1042034: rs676210 and rs2678379  $r^2 = 0.996$ , rs673548  $r^2 = 0.992$ ; with *NPC1L1* SNP rs2072183: rs17725246  $r^2 = 0.848$ ).

To investigate the association between the seven candidate SNPs and total cholesterol, we utilized linear regression adjusting for age, sex, and diabetes status. Four lead SNPs spanning three genes (*APOB*, *CETP*, and *NPC1L1*) were significant at the  $P < 0.05$  level for association with adjusted total cholesterol levels (Table 2). Because our hypothesis test relates not to the marginal SNP effect, but to the GxE interaction and because these SNPs were selected based on established associations with total cholesterol, we do not apply a Bonferroni correction to the inclusion of these SNPs for interaction testing. Only these SNPs with marginal effects on total cholesterol were carried forward to GxE interaction testing due to power concerns.

We used linear regression to investigate the possibility of GxE interactions adjusting the outcome of total cholesterol for age, sex, and diabetes status. Another linear regression then modeled the interaction between the genotypes of the four SNPs and dietary cholesterol intake on the outcome of residual total cholesterol levels (see Statistical analyses). Results of the SNP interactions with dietary cholesterol levels on the outcome of total cholesterol in the discovery subset of CLEAR are presented in Table 2. Of the four lead SNPs tested, only an *APOB* missense SNP, rs1042034 ( $P = 0.0174$ ), and an *NPC1L1* synonymous SNP, rs2072183 ( $P = 0.0125$ ) had GxE interactions with dietary cholesterol that were significant at the  $P = 0.05$  level for prediction of adjusted total cholesterol.

TABLE 1. Baseline characteristics of the CLEAR cohort

Baseline Characteristics	Discovery CLEAR Subset (n = 772)	Replication CLEAR Subset (n = 356)	Pooled CLEAR Cohort (n = 1,128)
Gender			
Female, n (%)	134 (17.4)	234 (65.7)	368 (32.6)
Male, n (%)	638 (82.6)	122 (34.3)	760 (67.4)
CAAD case-control status			
Case, n (%)	308 (39.9)	50 (14.0)	358 (31.7)
Control, n (%)	422 (54.7)	217 (61.0)	639 (56.6)
Moderate obstruction, n (%)	27 (3.5)	30 (8.4)	57 (5.1)
Other, n (%)	15 (1.9)	59 (16.6)	74 (6.6)
Age, mean ± SE, years	66.19 ± 9.43	66.01 ± 9.17	66.13 ± 9.34
Diabetes status, n (%)	110 (14.0)	30 (8.4)	140 (12.4)
Medication use:			
Statin use, n (%)	336 (42.7)	89 (25.0)	425 (37.7)
Oral hypoglycemic use, n (%)	90 (11.5)	23 (6.5)	113 (10.0)
Body mass index, kg/m <sup>2</sup> , mean ± SE	28.49 ± 5.06	28.01 ± 5.40	28.34 ± 5.34
Total caloric intake, kcal/day, mean ± SE	1,844.86 ± 737.46	1,800.43 ± 629.56	1,830.84 ± 705.54
Ln (dietary cholesterol), mg/day, mean ± SE	5.21 ± 0.27	5.29 ± 0.20	5.24 ± 0.21
Unadjusted total cholesterol, mg/dl, mean ± SE	191.78 ± 38.97	192.01 ± 38.47	191.85 ± 38.83
Statin-use adjusted total cholesterol, mg/dl, mean ± SE <sup>a</sup>	206.27 ± 41.05	211.67 ± 38.75	207.97 ± 40.42

<sup>a</sup>Statin use adjusted total cholesterol was calculated through imputing pretreatment lipid levels for the 425 subjects taking statins, see Total lipid levels section.

These two SNPs and their proxy SNPs in high LD were carried forward for sensitivity and replication testing.

Due to the expected limited power to detect higher order interactions with CAAD case-control status, statins, or sex, we instead performed sensitivity analyses stratified by CAAD case-control status, statin use, and sex. These analyses did not show differences in direction or general magnitude of effects ( $\beta$  coefficients) for the SNP genotypes, dietary cholesterol, or the dietary cholesterol-by-SNP interaction for the *APOB* SNPs rs1042034. However, in the discovery subset of CLEAR, the magnitude of *NPCIL1* SNP effects in males and females were strikingly different and the GxE interaction was not significant for females alone. For rs17725246, the interaction  $\beta = 26.06$  in males and 0.742 in females and the GxE interaction test  $P = 0.982$  in females versus  $P = 0.0009$  for males (Table 3). We reanalyzed

the marginal effect of the SNP on total cholesterol by gender and found that *NPCIL1* SNP rs17725246 marginally predicted total cholesterol in males only (SNP effect  $\beta$  coefficient = 5.2,  $P = 0.054$ ) and not in females ( $\beta = -0.243$ ,  $P = 0.941$ ) in this cohort. We formally tested for a SNP-by-sex interaction effect in the prediction of total cholesterol and found it not significant ( $\beta = -5.258$ ,  $P = 0.24$ ); however, we identified a report of this gender interaction in the literature (27). Although underpowered, we modeled a SNP-by-dietary cholesterol-by-sex interaction in the pooled analyses and found that the rs17725246-sex-dietary cholesterol interaction term was suggestive given limited power (SNP-by-dietary cholesterol-by-sex interaction beta coefficient =  $-25.318$ ,  $P = 0.096$ ). For rs2072183, which is in high LD with rs17725246 ( $r^2 = 0.72$  in the pooled CLEAR cohort), SNP-by-dietary cholesterol-by-sex interaction test

TABLE 2. Candidate SNP associations with total cholesterol levels and GxE interaction in the CLEAR cohort

Gene	Gut Tissue Expression <sup>a</sup>	Lead SNP <sup>b</sup>	Type	Other SNPs in LD <sup>c</sup>	Minor Allele	MAF <sup>d</sup>	Beta <sup>e,f</sup>	Association with TC, $P$	Discovery Interaction, $P$
<i>ABCG8</i>	SI	rs4299376	Intronic	0	C	0.325	0.165	0.938	—
<i>APOB</i>	SI	rs1367117 <sup>g</sup>	nsSNP	1	A	0.288	4.489	0.0382	0.384
<i>APOB</i>	SI	rs1042034 <sup>g</sup>	nsSNP	3	G	0.199	$-8.78^h$	$3.63 \times 10^{-4}$	0.0174
<i>APOE</i>	Pancreas, SI	rs4420638	3'NC	0	G	0.183	5.38	0.171	—
<i>CETP</i>	SI, pancreas	rs3764261	5'NC	0	T	0.289	5.095	0.0227	0.661
<i>NPCIL1</i>	SI, pancreas	rs2072183	sSNP	1	G	0.227	5.524	0.0229	0.0125
<i>PCSK9</i>	SI, stomach	rs2479409	5'NC	0	G	0.328	$-2.67$	0.208	—

3'NC, 3' noncoding variant; 5'NC, 5' noncoding variant; MAF, minor allele frequency; nsSNP, nonsynonymous variant (i.e., missense variant); SI, small intestine; sSNP, synonymous variant.

<sup>a</sup>Major tissues expressed.

<sup>b</sup>Lead SNP refers to primary SNP tagging the specific loci identified in meta-analyses by Teslovich et al. [(7) and unpublished data].

<sup>c</sup>SNPs in LD genotyped in the CLEAR cohort.

<sup>d</sup>Minor allele frequencies calculated from the CLEAR study cohort,  $n = 1,128$ .

<sup>e</sup>Beta coefficients and  $P$  values from the linear model of an additive genotype predicting adjusted total cholesterol.

<sup>f</sup>All SNP  $\beta$  coefficients are in the same direction as indicated by Teslovich et al. [(7) and unpublished data] with the exception of *PCSK9* SNP rs2479409, whose direction of effect ( $-2.67$ ) is opposite of what was reported for total cholesterol (Teslovich  $\beta$  coefficient = 1.96).

<sup>g</sup>These two *APOB* SNPs are not in strong LD with each other ( $r^2 = 0.094$ ).

TABLE 3. Gene-by-dietary cholesterol interactions on total cholesterol levels in the CLEAR cohort

SNP	Distance from Lead SNP <sup>a</sup>	LD $r^2$ with Lead SNP	Type	Protein Change	Discovery Interaction PValue (n = 772)	Replication Interaction PValue (n = 356)	Pooled Interaction Beta Coefficient	Pooled Variation Explained	Pooled Interaction PValue
Interactions with <i>APOB</i> SNPs									
rs2678379	1,279b	0.996	I <sup>b</sup>	—	0.0175	0.0132	-28.10	0.654%	$2.92 \times 10^{-5}$
rs676210	6,243b	0.996	nsSNP	Pro27391Leu	0.0176	0.0217	-27.67	0.635%	$3.70 \times 10^{-5}$
rs673548	12,263b	0.992	I <sup>b</sup>	—	0.0178	0.0476	-26.04	0.592%	$6.69 \times 10^{-5}$
rs1042034	—	—	nsSNP	Ser4338Asn	0.0174	0.0442	-25.51	0.570%	$9.15 \times 10^{-5}$
Interactions with <i>NPC1L1</i> SNPs									
rs17725246	2,806b	0.727	5'NC	—	0.0010	—	—	—	—
Males	—	—	—	—	0.0009	0.044	26.06	0.507%	$2.20 \times 10^{-3}$
Females	—	—	—	—	0.982	0.804	0.742	0.00039%	0.948
rs2072183	—	—	sSNP	—	0.0125	—	—	—	—
Males	—	—	—	—	0.0047	0.080	20.80	0.385%	$7.70 \times 10^{-3}$
Females	—	—	—	—	0.512	0.511	1.332	0.0015%	0.899

5'NC, 5' noncoding variant; b, nucleotide bases; I, intronic bases; nsSNP, nonsynonymous variant (i.e., missense variant); sSNP, synonymous variant.

<sup>a</sup>Lead SNP refers to SNP identified in meta-analyses by Teslovich, et al. [17] and unpublished data.

<sup>b</sup>rs2678379 is located in 28th of 28 introns of *APOB* and is not part of the apoB-48 protein and rs673548 is located in the 23<sup>rd</sup> of 28 introns of *APOB* and is within the gene region included in the apoB-48 protein.

Supplemental Material can be found at:  
<http://www.jlr.org/content/suppl/2013/03/11/jlr.P035238.DC1.html>

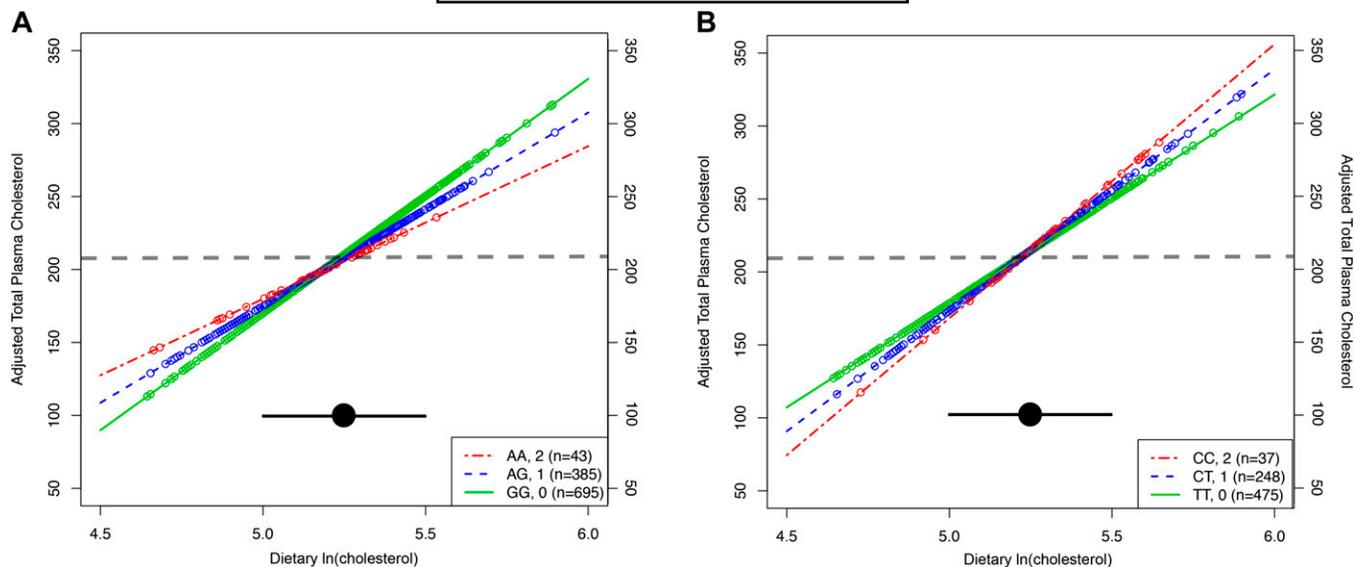
$P = 0.17$ . Given the indication and prior report of sex interactions and the LD between *NPC1L1* SNPs, we decided a priori to perform a sex-stratified replication analyses for these *NPC1L1* SNPs.

Results of the tests of the two lead SNPs' (rs1042034 and rs2072183, and their proxy SNPs in high LD) interactions with dietary cholesterol levels on adjusted total cholesterol within the discovery, replication, and pooled analyses are presented in Table 3. The *APOB* rs1042034 SNP GxE effect did replicate in the additional independent cohort ( $P = 0.04$ ) as did the GxE interaction effect of *NPC1L1* SNP rs17725246 in males ( $P = 0.04$ ), while *NPC1L1* SNP rs2072183 had a marginally replicated effect ( $P = 0.08$ ).

Both of the two lead SNPs identified also had other proxy SNPs in strong LD that were genotyped in our cohort. To investigate whether any of these proxies were superior at predicting interactions with dietary cholesterol on the outcome of adjusted total cholesterol levels, we replaced rs1042034 (*APOB* missense SNP) with rs673548, rs2678379, or rs676210 (Table 3). As well, we replaced rs2072183 (*NPC1L1* synonymous SNP) with rs17725246 in the model for GxE interactions. In pooled analyses considering both the discovery and replication sets, *APOB* SNP rs2678379 was only modestly more significant than rs1042034 (GxE  $P = 2.92 \times 10^{-5}$  vs.  $9.15 \times 10^{-5}$ ); of the two *NPC1L1* SNPs rs17725246 was also modestly more significant than rs2072183 (GxE  $P = 2.20 \times 10^{-3}$  vs.  $7.70 \times 10^{-3}$ ). We performed an ANOVA to estimate the percentage of total cholesterol attributed to each GxE interaction (Table 3). The top *APOB* SNP rs2678378 accounted for approximately 0.654% of total cholesterol variance, while the top *NPC1L1* SNP rs17725246 was attributed 0.507% of the variance in total cholesterol levels in a male-only subset of the cohort.

To consider the potential confounding effects of other cardiovascular covariates, we performed additional analyses where body mass index (BMI), total caloric intake, and CAAD case-control status were added individually and together to the base model (age, sex, and diabetes status) for analyses of the SNP and SNP-by-dietary cholesterol effects for *APOB* SNP rs2678379 and *NPC1L1* SNP rs17725246 on the outcome of total plasma cholesterol. Even when considering all three additional covariates in conjunction with the base model, both SNP-by-dietary cholesterol interactions remained test-wise significant (all covariates model *NPC1L1* rs17725246,  $P = 2.44 \times 10^{-3}$ ; *APOB* rs2678379,  $P = 2.44 \times 10^{-4}$ ).

The effects of the interaction of dietary cholesterol intake and the genotype of *APOB* SNP rs2678379 on residual total cholesterol, in the pooled data, are presented in Fig. 1A. The common allele was associated with a dose-dependent steeper increase in total cholesterol levels with increased dietary cholesterol intake. For *NPC1L1* SNP rs17725246, the common alleles were associated with a less steep increase in total cholesterol per allele substitution, considering only the males in the pooled CLEAR analyses ( $n = 760$ ) (Fig. 1B). When not considering the interaction term and looking solely at the outcome of residual total cholesterol for the genotypes of these two SNPs, we observe a similar



**Fig. 1.** *APOB* and *NPC1L1* SNP genotype dependent interaction with dietary cholesterol intake on the outcome of total cholesterol levels in the pooled analyses of the CLEAR cohort. Panel (A) shows the relationship between genotypes for *APOB* SNP rs2678379 and dietary cholesterol on the outcome of total cholesterol. Panel (B) shows the relationship between genotypes for *NPC1L1* SNP rs17725246 and dietary cholesterol on the outcome of total cholesterol in males only ( $n = 760$ ). Predicted values of total cholesterol for the pooled analyses' three genotypes and the natural log of dietary cholesterol (DC) intake (ln): rs2678379<sub>GG</sub> =  $-632.36 + 160.49$  (DC), rs2678379<sub>AG</sub> =  $-492 + 132.59$  (DC), and rs2678379<sub>AA</sub> =  $-343.64 + 104.69$  (DC); rs17725246<sub>TT</sub> =  $-535.85 + 142.90$  (DC), rs17725246<sub>CT</sub> =  $-653.26 + 165.34$  (DC), and rs17725246<sub>CC</sub> =  $-770.66 + 187.78$  (DC). Dashed lines represent the mean and standard deviation, respectively, for dietary cholesterol. Colored circles represent the predicted value of total cholesterol for each participant, according to their genotype class. Green circles and lines represent the major allele homozygous genotypes, the blue represent the heterozygous genotypes, and the red represent the minor allele homozygous genotypes.

pattern: for *APOB* SNP rs2678379 (supplementary Fig. V), the common allele is associated with a slight increase in total cholesterol, while the converse is true for *NPC1L1* SNP rs17725246 (supplementary Fig. VI).

## DISCUSSION

Total cholesterol level is perhaps the best established and most modifiable risk factor for CVD (2, 3). A recent, large GWAS identified over 95 loci for blood lipid phenotypes (4, 7). However, together the variants identified for total cholesterol accounted for only 12.4% of its heritability (7), underscoring concerns regarding the missing heritability of complex traits such as this. In addition to unidentified rare variants (28), genetic interactions, both gene-by-gene between loci and GxE have been proposed to explain some of this missing heritability (28). Recent examples of genes interacting with dietary intake have been identified for obesity traits (29, 30). Accordingly, in this study we have demonstrated a novel and replicated finding that common variants at *APOB* in males and females (rs2678379,  $P = 2.92 \times 10^{-5}$ ) and *NPC1L1* in males only (rs17725246,  $P = 2.20 \times 10^{-3}$ ) demonstrate GxE interactions with dietary cholesterol intake in the prediction of total cholesterol. *NPC1L1* appears to have an interaction with gender as well, with dietary cholesterol predicting higher total cholesterol, per genotype, in males. These replicated GxE interactions may be related to allelic differences in the absorption or processing of dietary cholesterol. In short, the extent to which "you are what you eat"

depends on your genotype. A recent study reported that overweight adolescents reported eating fewer calories than slimmer peers (31). While the authors hypothesized that this is due to differences in energy expenditures, and misreporting is also possible, a third possibility is that some of the differences are due to genetic factors in food absorption. We are not aware that in vivo genotype-dependent differences in the relationship between dietary and total cholesterol have previously been reported.

ApoB is one of the primary lipoproteins of chylomicrons and all liver-produced lipoproteins, including very low-density lipoprotein (VLDL) and LDL. There are two forms of apoB: apoB48, which is synthesized exclusively in the small intestine, and apoB100, which is primarily synthesized in the liver (32) and is a ligand for the LDL receptor; deficiencies in apoB100 or its receptor are etiological in familial hypercholesterolemia (33). ApoB48 is a component of chylomicrons produced in the small intestines to transport lipid to the liver, and is involved in the absorption of lipids (34). ApoB48 is generated through a single base substitution of a cytosine to uracil at nucleotide 6538 that results in a premature stop codon at protein position 2153 (32). However, VLDLs associated with apoB100 are also found in the small intestine (35). Apobec-1 knockout mice containing apoB100 and not apoB48 still form chylomicrons and can absorb dietary fat without noticeable deficiencies versus wild-type mice on normal diets. However, in low-fat conditions, the lack of apoB48 results in a decrease in the rate of chylomicron assembly (36). Therefore, one potential pathway whereby the SNPs identified

in this study could affect the relationship between dietary and total cholesterol could be at the stage of absorption or chylomicron assembly in the small intestine. Further molecular follow-up is necessary to elucidate the specific mechanisms by which these SNPs influence the relationship of dietary to total cholesterol levels.

*NPC1L1* is enriched in the lumen of the small intestine and specifically in the brush border of enterocytes (37) and is crucially involved in the early steps of cholesterol absorption by binding nonplant sterols with its N-terminal domain (38). Mouse knockout studies of *NPC1L1* resulted in a significant reduction in cholesterol absorption that was unaffected by bile acid supplementation (37, 39). In humans, in vitro experiments have demonstrated the potential molecular mechanisms through which nonsense variants affect cholesterol absorption: namely, through impairment of *NPC1L1* recycling, localization, glycosylation, or stability (40). Recently, Miao et al. (27) reported a sex-dependent difference for *NPC1L1* SNP rs2072183 minor allele association with total cholesterol levels in a Mulao Chinese cohort of 688 subjects and 288 males (SNP  $\times$  sex interaction,  $P = 0.03$ ); while this interaction test was not significant in a Han Chinese cohort of 738 subjects (274 males), the rs2072183 SNP effect in the Han subset was significant in males, but not in the females. In both the Mulao and Han Chinese subsets, as well as in our data, the rs2072183 minor G allele was associated with an increase in total cholesterol in males. Although Miao et al. (27) did not correct for multiple comparisons, we did find a similar effect in our data for *NPC1L1* SNP rs17725246, with the SNP effect on total cholesterol (TC) marginally significant in males ( $P = 0.054$ ) and not females ( $P = 0.94$ ), and the  $\beta$  coefficient indicating an increase in cholesterol levels in males only. Thus, we have identified a potential explanation for the reported lack of this SNP's association with total cholesterol in females: differential absorption or processing of dietary cholesterol based on *NPC1L1* genotype in males, which is not detectable in females.

Previous genetic investigations have demonstrated other functions for SNPs in high LD of those identified in this study in *APOB* and *NPC1L1*. *APOB* SNP rs676210 ( $r^2 = 0.996$ ) was associated with a genotype-dependent response to fenofibrate (41). This is consistent with the GxE interaction we observed, and based on differential dietary cholesterol absorption or processing by an *APOB* variant in LD with this SNP. In addition, two SNPs further 5' in the *APOB* gene but not in LD with the SNPs of interest here ( $r^2 < 0.15$ ) have been associated with an increased risk of biliary tract stones and cancer (42).

With association testing, it is often not possible to distinguish between the effects of SNPs in high LD to determine which is most likely the functional genetic variant driving the observed associations. The *APOB* SNP rs1042034 is a missense resulting in Ser4338Asn. It is in strong LD with a second missense, rs676210, which results in Ser4338Asn as well as intronic SNP rs2678379. While rs2678379 is intronic, it is in the last of 28 total introns of *APOB*, a position that can be regulatory. The latter SNP has a modestly smaller  $P$  value for our pooled data; however, any of the

tested variants could be the functional SNP. Rs1042034 is also in LD with seven intergenic SNPs in the 1000 Genomes data, which are less likely to be the causative SNP. The *NPC1L1* SNP rs2072183 is synonymous and less likely to be functional than the 5' SNP in LD, rs17725246, which was modestly more significant for GxE interaction in both the discovery and replication subsets. However, these are also in LD with an untyped 5' SNP, rs2074547. Recent reports have identified a rare *NPC1L1* missense mutation which causes a differential response to ezetimibe pharmacotherapy, *NPC1L1*<sub>V55M</sub> (rs119457968, located 653 bases from rs2072183, but with no LD information available in the 1000 Genomes database) (43). These and other rare coding variants nearby the SNPs identified in both *APOB* and *NPC1L1* may underlie the findings of GxE interactions in this study.

Another explanation for our findings may come from our studied SNPs being in close proximity to sites that alter gene expression. For example, recent studies into expression quantitative trait loci (eQTL) that influence hepatocyte gene expression found an enrichment of reproducible eQTL SNPs at gene starts and ends (44), including *APOB* SNP rs3923672, which is located in the last exon of *APOB*. Despite the proximity of rs3923672 to the studied *APOB* SNPs, it is not reported to be in strong LD ( $r^2 > 0.8$ ) in the 1000 Genomes data, and we lack the genotype of this SNP to assess its relationship to the studied SNPs in our cohort. For *NPC1L1*, recent publications of the Encyclopedia of DNA Elements (ENCODE) Consortium identified a region extremely close to rs17725246 as being hypersensitive to DNase I enzymatic activity and also being involved in chromatin structure and histone modifications (45–47). In addition, separate reports using chromatin immunoprecipitation sequencing have reported the genomic 5' untranslated region around rs17725246 to bind the proteins GATA2 (48) and CDX2 (49). These results in conjunction with the eQTL study findings of enrichment at gene starts and ends may implicate altered gene expression as a possible explanation for our findings of SNP-by-dietary cholesterol interactions for *APOB* and *NPC1L1*. Further functional studies will be required to determine which of these close-proximity variants, both coding and regulatory, are functionally responsible for the observed *APOB* and *NPC1L1* GxE interactions.

Strengths of the present study include a large well-characterized community-based cohort with dietary intake information, genetic data, and total lipid phenotypes. In addition, through separate genotyping of SNPs identified as significant for interactions with dietary cholesterol in the discovery subset of the cohort, we were able to replicate our findings in an independent subset of the CLEAR study. Limitations include the lack of ethnic diversity, which limits the generalizability of these findings, and the sample size. Due to power considerations inherent in GxE interaction studies, this investigation only considered seven SNPs in six candidate loci. Larger studies could assess a wider range of variation at loci for GxE interactions and further explore the sex interaction seen for *NPC1L1*. In this case, to reduce contrasts we utilized a candidate

gene approach to narrow our genetic loci to only those that are expressed in the gastrointestinal system (excluding the liver and large intestine) and only tested SNPs for GxE interactions if they were at least marginally significant for association with total cholesterol in these data. It is entirely possible that GxE interactions could occur when no marginal TC effect is identified.

In conclusion, our study has identified and replicated novel GxE interactions between SNPs in *APOB* in both sexes and also in *NPC1L1* in males, and dietary cholesterol intake in the prediction of total cholesterol level. Such replicated interactions may account for part of the missing heritability of complex traits such as total cholesterol levels. Further work is needed to elucidate the mechanism, whether through altered gene expression or through protein function changes, by which these variants, or variants in LD, alter dietary cholesterol absorption or processing and hence alter the relationship between dietary and total cholesterol. 

The authors thank the study participants. They also thank the following people for their technical assistance: Tamara J. Bacus, Edward Boyko, Julieann Marshall, Laura McKinstry, Karen Nakayama, and Jeff Rodenbaugh. They thank Drs. Tanya Teslovich and Sekar Kathiresan for providing their unpublished rs1042034 data.

## REFERENCES

- Roger, V. L., A. S. Go, D. M. Lloyd-Jones, E. J. Benjamin, J. D. Berry, W. B. Borden, D. M. Bravata, S. Dai, E. S. Ford, C. S. Fox, et al. 2012. Heart disease and stroke statistics—2012 update: a report from the American Heart Association. *Circulation*. **125**: e2–e220.
- Verschuren W. M., D. R. Jacobs, B. P. Bloemberg, D. Kromhout, A. Menotti, C. Aravanis, H. Blackburn, R. Buzina, A. S. Dontas, F. Fidanza, et al. 1995. Serum total cholesterol and long-term coronary heart disease mortality in different cultures: Twenty-five-year follow-up of the seven countries study. *JAMA*. **274**: 131–136.
- Wallis, E. J., L. E. Ramsay, I. U. Haq, P. Ghahramani, P. R. Jackson, K. R. Yeo, and W. W. Yeo. 2000. Coronary and cardiovascular risk estimation for primary prevention: validation of a new Sheffield table in the 1995 Scottish health survey population. *BMJ*. **320**: 671–676.
- Kathiresan, S., A. K. Manning, S. Demissie, R. B. D'Agostino, A. Surti, C. Guiducci, L. Gianniny, N. P. Burt, O. Melander, M. Orholm-Melander, et al. 2007. A genome-wide association study for blood lipid phenotypes in the Framingham Heart Study. *BMC Med. Genet.* **8**(Suppl 1): S17.
- Varady, K. A., and P. J. Jones. 2005. Combination diet and exercise interventions for the treatment of dyslipidemia: an effective preliminary strategy to lower cholesterol levels? *J. Nutr.* **135**: 1829–1835.
- Vogel, J. H., S. F. Bolling, R. B. Costello, E. M. Guarneri, M. W. Krucoff, J. C. Longhurst, B. Olshansky, K. R. Pelletier, C. M. Tracy, R. A. Vogel, et al. 2005. Integrating complementary medicine into cardiovascular medicine. A report of the American College of Cardiology Foundation Task Force on Clinical Expert Consensus Documents (Writing Committee to Develop an Expert Consensus Document on Complementary and Integrative Medicine). *J. Am. Coll. Cardiol.* **46**: 184–221.
- Teslovich, T. M., K. Musunuru, A. V. Smith, A. C. Edmondson, I. M. Stylianou, M. Koseki, J. P. Pirruccello, S. Ripatti, D. I. Chasman, C. J. Willer, et al. 2010. Biological, clinical and population relevance of 95 loci for blood lipids. *Nature*. **466**: 707–713.
- Jarvik, G. P., T. S. Hatsukami, C. Carlson, R. J. Richter, R. Jampsa, V. H. Brophy, S. Margolin, M. Rieder, D. Nickerson, G. D. Schellenberg, et al. 2003. Paraoxonase activity, but not haplotype utilizing the linkage disequilibrium structure, predicts vascular disease. *Arterioscler. Thromb. Vasc. Biol.* **23**: 1465–1471.

- Jarvik, G. P., L. S. Rozek, V. H. Brophy, T. S. Hatsukami, R. J. Richter, G. D. Schellenberg, and C. E. Furlong. 2000. Paraoxonase (PON1) phenotype is a better predictor of vascular disease than is PON1(192) or PON1(55) genotype. *Arterioscler. Thromb. Vasc. Biol.* **20**: 2441–2447.
- Ronald, J., R. Rajagopalan, J. Ranchalis, J. Marshall, T. Hatsukami, P. Heagerty, and G. Jarvik. 2009. Analysis of recently identified dyslipidemia alleles reveals two loci that contribute to risk for carotid artery disease. *Lipids Health Dis.* **8**: 52.
- Kim, D. S., A. A. Burt, J. E. Ranchalis, R. J. Richter, J. K. Marshall, J. F. Eintracht, E. A. Rosenthal, C. E. Furlong, and G. P. Jarvik. 2012. Additional common polymorphisms in the PON gene cluster predict PON1 activity but not vascular disease. *J. Lipids*. **2012**: 476316.
- Kim, D. S., A. A. Burt, J. E. Ranchalis, R. J. Richter, J. K. Marshall, K. S. Nakayama, E. R. Jarvik, J. F. Eintracht, E. A. Rosenthal, C. E. Furlong, et al. 2012. Dietary cholesterol increases paraoxonase 1 enzyme activity. *J. Lipid Res.* **53**: 2450–2458.
- Kim, D. S., A. A. Burt, D. R. Crosslin, P. D. Robertson, J. E. Ranchalis, E. J. Boyko, D. A. Nickerson, C. E. Furlong, and G. P. Jarvik. 2013. Novel common and rare genetic determinants of paraoxonase activity: FTO, SERPINA12, and ITGAL. *J. Lipid Res.* **54**: 552–560.
- Illumina. 2010. HumanCVD Genotyping BeadChip. Retrieved July 10, 2012, from [http://www.illumina.com/documents/products/datasheets/datasheet\\_human\\_cvd.pdf](http://www.illumina.com/documents/products/datasheets/datasheet_human_cvd.pdf).
- Pritchard, J. K., M. Stephens, and P. Donnelly. 2000. Inference of population structure using multilocus genotype data. *Genetics*. **155**: 945–959.
- Rimm, E. B., E. L. Giovannucci, M. J. Stampfer, G. A. Colditz, L. B. Litin, and W. C. Willett. 1992. Reproducibility and validity of an expanded self-administered semiquantitative food frequency questionnaire among male health professionals. *Am. J. Epidemiol.* **135**: 1114–1126; discussion 1127–1136.
- Willett, W., M. Stampfer, N. F. Chu, D. Spiegelman, M. Holmes, and E. Rimm. 2001. Assessment of questionnaire validity for measuring total fat intake using plasma lipid levels as criteria. *Am. J. Epidemiol.* **154**: 1107–1112.
- Hunter, D. J., E. B. Rimm, F. M. Sacks, M. J. Stampfer, G. A. Colditz, L. B. Litin, and W. C. Willett. 1992. Comparison of measures of fatty acid intake by subcutaneous fat aspirate, food frequency questionnaire, and diet records in a free-living population of US men. *Am. J. Epidemiol.* **135**: 418–427.
- Bachorik, P. S., and J. J. Albers. 1986. Precipitation methods for quantification of lipoproteins. *Methods Enzymol.* **129**: 78–100.
- Warnick, G. R. 1986. Enzymatic methods for quantification of lipoprotein lipids. *Methods Enzymol.* **129**: 101–123.
- Mihaylova, B., J. Emberson, L. Blackwell, A. Keech, J. Simes, E. H. Barnes, M. Voysey, A. Gray, R. Collins, and C. Baigent. 2012. The effects of lowering LDL cholesterol with statin therapy in people at low risk of vascular disease: meta-analysis of individual data from 27 randomised trials. *Lancet*. **380**: 581–590.
- Johnson, A. D., R. E. Handsaker, S. L. Pulit, M. M. Nizzari, C. J. O'Donnell, and P. I. W. de Bakker. 2008. SNAP: a web-based tool for identification and annotation of proxy SNPs using HapMap. *Bioinformatics*. **24**: 2938–2939.
- Liu, X., X. Yu, D. J. Zack, H. Zhu, and J. Qian. 2008. TiGER: a database for tissue-specific gene expression and regulation. *BMC Bioinformatics*. **9**: 271.
- The EMBL Nucleotide Sequence Database. 2012. <http://www.ebi.ac.uk/embl/>.
- Dixon, W. J., and J. W. Tukey. 1968. Approximate behavior of distribution of Winsorized t (trimming/Winsorization 2). *Technometrics*. **10**: 83–98.
- Rivest, L. P. 1994. Statistical properties of Winsorized means for skewed distributions. *Biometrika*. **81**: 373–383.
- Miao, L., R. X. Yin, X. J. Hu, D. F. Wu, X. L. Cao, Q. Li, T. T. Yan, L. H. Aung, J. Z. Wu, and W. X. Lin. 2012. Association of rs2072183 SNP and serum lipid levels in the Mulao and Han populations. *Lipids Health Dis.* **11**: 61.
- Zuk, O., E. Hechter, S. R. Sunyaev, and E. S. Lander. 2012. The mystery of missing heritability: genetic interactions create phantom heritability. *Proc. Natl. Acad. Sci. USA*. **109**: 1193–1198.
- Velez Edwards, D. R., A. C. Naj, K. Monda, K. E. North, M. Neuhouser, O. Magvanjav, I. Kusimo, M. Z. Vitolins, J. E. Manson, M. J. O'Sullivan, et al. 2013. Gene-environment interactions and obesity traits among postmenopausal African-American and Hispanic women in the Women's Health Initiative SHARe Study. *Hum. Genet.* **132**: 323–336.

30. Qi, Q., A. Y. Chu, J. H. Kang, M. K. Jensen, G. C. Curhan, L. R. Pasquale, P. M. Ridker, D. J. Hunter, W. C. Willett, E. B. Rimm, et al. 2012. Sugar-sweetened beverages and genetic risk of obesity. *N. Engl. J. Med.* **367**: 1387–1396.
31. Skinner, A. C., M. J. Steiner, and E. M. Perrin. 2012. Self-reported energy intake by age in overweight and healthy-weight children in NHANES, 2001–2008. *Pediatrics.* **130**: e936–e942.
32. Higuchi, K., A. V. Hospattankar, S. W. Law, N. Meglin, J. Cortright, and H. B. Brewer. 1988. Human apolipoprotein B (apoB) mRNA: identification of two distinct apoB mRNAs, an mRNA with the apoB-100 sequence and an apoB mRNA containing a premature in-frame translational stop codon, in both liver and intestine. *Proc. Natl. Acad. Sci. USA.* **85**: 1772–1776.
33. Vega, G. L., and S. M. Grundy. 1986. In vivo evidence for reduced binding of low density lipoproteins to receptors as a cause of primary moderate hypercholesterolemia. *J. Clin. Invest.* **78**: 1410–1414.
34. Kane, J. P., D. A. Hardman, and H. E. Paulus. 1980. Heterogeneity of apolipoprotein B: isolation of a new species from human chylomicrons. *Proc. Natl. Acad. Sci. USA.* **77**: 2465–2469.
35. Tamura, M., A. Tanaka, Y. Kobayashi, Z. Nihei, and F. Numano. 2000. Expression of apolipoprotein B-100 in isolated human small intestine epithelium. *Horm. Metab. Res.* **32**: 343–349.
36. Kendrick, J. S., L. Chan, and J. A. Higgins. 2001. Superior role of apolipoprotein B48 over apolipoprotein B100 in chylomicron assembly and fat absorption: an investigation of apobec-1 knock-out and wild-type mice. *Biochem. J.* **356**: 821–827.
37. Altmann, S. W., H. R. Davis, Jr., L. J. Zhu, X. Yao, L. M. Hoos, G. Tetzloff, S. P. Iyer, M. Maguire, A. Golovko, M. Zeng, et al. 2004. Niemann-Pick C1 Like 1 protein is critical for intestinal cholesterol absorption. *Science.* **303**: 1201–1204.
38. Zhang, J. H., L. Ge, W. Qi, L. Zhang, H. H. Miao, B. L. Li, M. Yang, and B. L. Song. 2011. The N-terminal domain of NPC1L1 protein binds cholesterol and plays essential roles in cholesterol uptake. *J. Biol. Chem.* **286**: 25088–25097.
39. Davies, J. P., C. Scott, K. Oishi, A. Liapis, and Y. A. Ioannou. 2005. Inactivation of NPC1L1 causes multiple lipid transport defects and protects against diet-induced hypercholesterolemia. *J. Biol. Chem.* **280**: 12710–12720.
40. Wang, L. J., J. Wang, N. Li, L. Ge, B. L. Li, and B. L. Song. 2011. Molecular characterization of the NPC1L1 variants identified from cholesterol low absorbers. *J. Biol. Chem.* **286**: 7397–7408.
41. Wojczynski, M. K., G. Gao, I. Borecki, P. N. Hopkins, L. Parnell, C. Q. Lai, J. M. Ordovas, B. H. Chung, and D. K. Arnett. 2010. Apolipoprotein B genetic variants modify the response to fenofibrate: a GOLDN study. *J. Lipid Res.* **51**: 3316–3323.
42. Andreatti, G., J. Chen, Y. T. Gao, A. Rashid, B. E. Chen, P. Rosenberg, L. C. Sakoda, J. Deng, M. C. Shen, B. S. Wang, et al. 2008. Polymorphisms of genes in the lipid metabolism pathway and risk of biliary tract cancers and stones: a population-based case-control study in Shanghai, China. *Cancer Epidemiol. Biomarkers Prev.* **17**: 525–534.
43. Simon, J. S., M. C. Karnoub, D. J. Devlin, M. G. Arreaza, P. Qiu, S. A. Monks, M. E. Severino, P. Deutsch, J. Palmisano, A. B. Sachs, et al. 2005. Sequence variation in NPC1L1 and association with improved LDL-cholesterol lowering in response to ezetimibe treatment. *Genomics.* **86**: 648–656.
44. Innocenti, F., G. M. Cooper, I. B. Stanaway, E. R. Gamazon, J. D. Smith, S. Mirkov, J. Ramirez, W. Liu, Y. S. Lin, C. Moloney, et al. 2011. Identification, replication, and functional fine-mapping of expression quantitative trait loci in primary human liver tissue. *PLoS Genet.* **7**: e1002078.
45. Neph, S., J. Vierstra, A. B. Stergachis, A. P. Reynolds, E. Haugen, B. Vernot, R. E. Thurman, S. John, R. Sandstrom, A. K. Johnson, et al. 2012. An expansive human regulatory lexicon encoded in transcription factor footprints. *Nature.* **489**: 83–90.
46. Maurano, M. T., R. Humbert, E. Rynes, R. E. Thurman, E. Haugen, H. Wang, A. P. Reynolds, R. Sandstrom, H. Qu, J. Brody, et al. 2012. Systematic localization of common disease-associated variation in regulatory DNA. *Science.* **337**: 1190–1195.
47. Dunham, I., A. Kundaje, S. F. Aldred, P. J. Collins, C. A. Davis, F. Doyle, C. B. Epstein, S. Fietze, J. Harrow, R. Kaul, et al. 2012. An integrated encyclopedia of DNA elements in the human genome. *Nature.* **489**: 57–74.
48. Hu, G., D. E. Schones, K. Cui, R. Ybarra, D. Northrup, Q. Tang, L. Gattinoni, N. P. Restifo, S. Huang, and K. Zhao. 2011. Regulation of nucleosome landscape and transcription factor targeting at tissue-specific enhancers by BRG1. *Genome Res.* **21**: 1650–1658.
49. Verzi, M. P., H. Shin, H. H. He, R. Sulahian, C. A. Meyer, R. K. Montgomery, J. C. Fleet, M. Brown, X. S. Liu, and R. A. Shivdasani. 2010. Differentiation-specific histone modifications reveal dynamic chromatin interactions and partners for the intestinal transcription factor CDX2. *Dev. Cell.* **19**: 713–726.