Published in final edited form as:
Lipids. 2013 February ; 48(2): 155–165. doi:10.1007/s11745-012-3740-8.

Genetic Analysis of 16 NMR-Lipoprotein Fractions in Humans, the GOLDN Study

Aldi T. Kraja,
Division of Statistical Genomics, Washington University School of Medicine, 4444 Forest Park Ave, Campus Box 8506, St. Louis, MO 63108, USA aldi@wustl.edu

Ingrid B. Borecki,
Division of Statistical Genomics, Washington University School of Medicine, 4444 Forest Park Ave, Campus Box 8506, St. Louis, MO 63108, USA

Michael Y. Tsai,
Laboratory Medicine and Pathology, University of Minnesota, Minneapolis, MN, USA

Jose M. Ordovas,
Nutrition and Genomics Laboratory, Tufts University, Boston, MA, USA

Centro Nacional Investigaciones Cardiovasculares (CNIC), Madrid, Spain

Instituto Madrileno de Estudios Avanzados (IMIDEA) Alimentacion, Madrid, Spain

Paul N. Hopkins,
Cardiovascular Genetics Research, University of Utah, Salt Lake City, UT, USA

Chao-Qiang Lai,
Nutrition and Genomics Laboratory, Tufts University, Boston, MA, USA

Alexis C. Frazier-Wood,
Department of Epidemiology, University of Alabama, Birmingham, AL, USA

Robert J. Straka,
Department of Experimental and Clinical Pharmacology, University of Minnesota, Minneapolis, MN, USA

James E. Hixson,
Human Genetics Center, School of Public Health, University of Texas Health Science Center at Houston, Houston, TX, USA

Michael A. Province, and
Division of Statistical Genomics, Washington University School of Medicine, 4444 Forest Park Ave, Campus Box 8506, St. Louis, MO 63108, USA

Donna K. Arnett
Department of Epidemiology, University of Alabama, Birmingham, AL, USA

Ingrid B. Borecki: iborecki@wustl.edu; Michael Y. Tsai: tsaix001@umn.edu; Jose M. Ordovas: jose.ordovas@tufts.edu; Paul N. Hopkins: paul.hopkins@utah.edu; Chao-Qiang Lai: chao.lai@tufts.edu; Alexis C. Frazier-Wood: lekki.wood@gmail.com; Robert J. Straka: strak001@tc.umn.edu; James E. Hixson: James.E.Hixson@uth.tmc.edu; Michael A. Province: mprovince@wustl.edu; Donna K. Arnett: Arnett@uab.edu

Correspondence to: Aldi T. Kraja.

Electronic supplementary material The online version of this article (doi:10.1007/s11745-012-3740-8) contains supplementary material, which is available to authorized users.

Conflict of interest None.

Abstract

Sixteen nuclear magnetic resonance (NMR) spectroscopy lipoprotein measurements of more than 1,000 subjects of GOLDN study, at fasting and at 3.5 and 6 h after a postprandial fat (PPL) challenge at visits 2 and 4, before and after a 3 weeks Fenofibrate (FF) treatment, were included in 6 time-independent multivariate factor analyses. Their top 1,541 unique SNPs were assessed for association with GOLDN NMR-particles and classical lipids. Several SNPs with $-\log_{10} p > 7.3$ and MAF ≥ 0.10, mostly intergenic associated with NMR-single traits near genes *FAM84B* (8q24.21), *CRIPT* (2p21), *ACOXL* (2q13), *BCL2L11* (2q13), *PCDH10* (4q28.3), *NXPH1* (7p22), and *SLC24A4* (14q32.12) in association with NMR-LDLs; *HOMER1* (5q14.2), *KIT* (4q11–q12), *VSNL1* (2p24.3), *QPRT* (16p11.2), *SYNPR* (3p14.2), *NXPH1* (7p22), *NELL1* (11p15.1), and *RUNX3* (1p36) with NMR-HDLs; and *DOK5-CBLN4-MC3R* (20q13), *NELL1* (11p15.1), *STXBP6* (14q12), *APOB* (2p24-p23), *GPR133* (12q24.33), *FAM84B* (8q24.21) and *NR5A2* (1q32.1) in association with NMR-VLDLs particles. NMR single traits associations produced 75 % of 114 significant candidates, 7 % belonged to classical lipids and 18 % overlapped, and 16 % matched for time of discovery between NMR- and classical traits. Five proxy genes, (*ACOXL, FAM84B, NXPH1, STK40* and *VAPA*) showed pleiotropic effects. While tagged for significant associations in our study and with some extra evidence from the literature, candidates as *CBNL4, FAM84B, NXPH1, SLC24A4* remain unclear for their functional relation to lipid metabolism. Although GOLDN study is one of the largest in studying PPL and FF treatment effects, the relatively small samples (over 700–1,000 subjects) in association tests appeals for a replication of such a study. Thus, further investigation is needed.

Keywords

Nuclear magnetic resonance particles; Lipoproteins; Fenofibrate; Postprandial challenge; Genome-wide association

Introduction

Traditional lipid measurements, including low- and high density lipoprotein cholesterol (LDLC, HDLC), triglycerides (TAG) and total cholesterol (TC) are used to guide preventive therapy [1, 2]. Nevertheless, better understanding of atherogenic particles and their role in disease development might deliver improved treatment as well as prediction. Nuclear magnetic resonance (NMR) spectroscopy can be used to measure lipoprotein particles at various sizes, with the hope of conveying such improved understanding [3-8]. Lipoprotein particle concentrations within specified size subfractions may provide useful information beyond traditional lipid measurements [9-14].

The GOLDN (Genetics of Lipid Lowering Drugs and Diet Network) study considered NMR measurements of the lipoprotein subfraction concentration at 6 time points. Blood was obtained while fasting and at 3.5 and 6 h after an oral lipid challenge. The pre- and post-prandial lipids were measured both before and after 3 weeks Fenofibrate (FF) treatment (Supplemental Fig. 1). This report expands on our previous findings with candidate SNPs [15, 16], on quantified lipid profiles changes by way of NMR particle sizes and concentrations as well as classical lipids to a large array of “hybrid” genotypes in 2,543,887 SNPs of GOLDN participants.

Methods

Sampled Population

Approximately 200 families were recruited from the Minnesota and Utah field centers of the Family Heart Study. The numbers of participants at each of the time-points vary and are reported in the Supplemental Table 1. The GOLDN study examined metabolic responses of an oral fat challenge prior to and after 3 weeks of oral FF therapy. The oral fat (PPL) challenge consisted of a cream shake prepared according to the protocol of Patsch et al. [17]. Details of the shake are provided in Tsai et al. [18]. The drug prescribed as open-label to all participants at visit 2 was micronized Fenofibrate 160 mg pill once daily for 3 weeks period (Supplemental Fig. 1). The study was approved by the Institutional Review Boards of each institution and informed consent was received from all GOLDN participants.

NMR- and Classical Lipid Fractions

Sixteen NMR lipid particle concentrations and size measurements were produced by the following procedure. Fasting blood samples for lipoprotein analysis were collected in EDTA-containing tubes. The fasting, 3.5, and 6 h postprandial blood samples were centrifuged within 20 min of collection at 2,000*g* for 15 min at 4 °C. Plasma samples were stored at 4 °C, and all analyses were performed within 4 days of collection. NMR lipid fractions were measured by proton NMR spectroscopy on non-fractionated plasma (Liposcience). This method quantified particle concentration by converting the characteristic signal amplitude generated by the methyl group NMR signal of each lipoprotein subclass particle [19].

Very low density lipoprotein (VLDL) were represented by 5 types of measurements: mean VLDL particle size (nm), and concentrations of small, medium, large, and total VLDL all expressed in mg/dl. For low density lipoprotein (LDL) 4 measurements were assessed: mean LDL size (nm), and concentrations of small, large and total LDL expressed in mg/dl. Concentrations of chylomicrons (CHYL, mg/dl) and intermediate density lipoproteins (IDL) (mg/dl) were also provided. For high density lipoprotein (HDL) 5 measurements were analyzed: mean HDL size (nm), and concentrations of small, medium, large, and total HDL expressed in mg/dl.

For this study we examined 16 NMR lipoprotein measurements and classical lipids at visits 2 and 4 (before and after FF treatment) for blood drawn at fasting and at 3.5 and 6 h after dietary fat challenge, labeled respectively with **v**isit and blood **d**raw as v2d1, v2d2, v2d3 and v4d1, v4d2 and v4d3. They correspond to 6 time points studied, (NMR, 16 × 6 + TAG × 6 + HDLC × 2 + LDLC × 2) for a total of 106 trait measurements.

Genotyping

The original genotype data were produced using the Affymetrix 6.0 platform. The genotypes were QC-d in families and any Mendelian errors were set to missing. Based on the cleaned original genotypes of tag SNPs for a LD with $r^2 = 0.8$, 10 principal components were produced using EIGENSTRAT software. These 10 PCs were used as covariates in the association tests models for accounting any population stratification. Finally, the clean original genotyped data were recoded to an additive model as 0, 1 and 2. To increase the number of SNPs for testing to 2,543,887 SNPs in this study we employed imputations of additional markers compared to original genotypes, based on a reference panel of phased haplotypes of CEPH subjects from HapMap (release 22, build 36). From a total of 718,542 original genotyped autosomal SNPs in GOLDN, we used 584,063 SNPs in the imputation after filtering out SNPs for Hardy–Weinberg p value <1e-6, SNPs with minor allele frequency (MAF) <1 % and SNPs that were not present in HapMap. The imputations were

performed by using MACH software, v. 1.0.16. Then the imputed data were overwritten with the clean original (genotyped) dosage data, in the locations where SNPs matched. This process produced a "hybrid" dosage data (a combined genotyped and imputed genetic data) with a total of 2,543,887 SNPs for autosomal chromosomes.

Factor Scores, Single Traits Data Preparation and Statistical Analysis

All variables were checked and, where necessary, transformations were performed to render them with a normal distribution. For example, classical HDLC and TAG were natural log transformed. Of NMR-fractions Total, Large, and Size of VLDL, Total HDL, CHYL were log transformed; Large LDL, Small and Medium HDL, and Small VLDL were square root transformed; IDL was BLOM transformed; and power transformations (based on Box-Cox maximum likelihood estimation) were applied to Large and Medium HDL (a power coefficient of 0.37), and for Medium VLDL (a power coefficient of 0.34). Each trait was adjusted for a polynomial trend for age, age^2 and age^3 within gender by the use of a stepwise regression analysis. Residuals from the stepwise regression for each NMR-trait with a normal distribution N(0,1) were included (arrangement by time-points) in multivariate factor analyses. Factor scores were produced using the procedure FACTOR with *VARIMAX* rotation for selecting orthogonal sets of factors, achieved with SAS 9.2 for Linux OS. The statistical formulation of factor analysis is:

$$y_{ij}=l_{i1}\times f_{1j}+l_{i2}\times f_{2j}+\cdots+l_{ik}+f_{kj}+e_{ij},$$

where y is an observed variable i for subject j which can be expressed as a weighted composite of latent factor k scores (f_{kj}) for subject j weighted by the loading (l_{ik}) of variable i in latent factor k. The criteria to select factors were two, the factor model had to be significant ($p < 0.05$), and at least two traits contributed significantly in a factor (with a loading coefficient ≥0.4 each). Because the NMR-traits were standardized in advance, their loading coefficients represent correlation coefficients of NMR-traits to factor scores. The number of factors accepted for a model was four because of the above two criteria, and a summary of those contributions for each time-point is provided in Supplemental Figs. 2–7. The factor scores were utilized as response variables for the exploratory association tests. The same process of adjustments was implemented also for each single classical lipid trait. After exploring genome wide factor score discoveries, the adjusted single NMR-fractions and also single classical lipid measurements were used as response variables to test only the top selected associations.

The association tests among GOLDN SNPs and traits followed these steps. Factor scores (4 at each time point) were tested for association in a mixed model with each SNP genome-wide, where the SNP was included as a fixed effect. Furthermore, in the model as fixed covariates, were included sex and 10 PCs produced by Eigenstrat, the last ones with the purpose of correcting for any population stratification. To account for the correlation/similarity among members of the same family, we introduced a random effect in the mixed model by using the following statement: repeated/subject=pedid type=un, where *pedid* represents a familial unique pedigree id membership assigned to members of each family. In the MIXED PROCEDURE of SAS subject= is a keyword for identifying the repeated units. The repeated statement is used to specify the random covariance $\boldsymbol{R}$ matrix in the mixed model for the ε term. This statement produces blocks with repeated levels (subjects within a pedigree) for each pedigree by specifying the covariance structure of the familial observations. If no repeated statement was specified then the $\boldsymbol{R}$ matrix would be equal to $\sigma^2 I$ (one general random error for all data), without accounting for familial relationships. While subject=defines the blocks of $\boldsymbol{R}$, the type=unstructured specifies the structure of these blocks, in this case dependent on the GOLDN data. In summary, the linear mixed effects

(LME) model is given by: $\boldsymbol{Y} = \boldsymbol{XB} + \boldsymbol{ZU} + \varepsilon$, where $\boldsymbol{Y}$ is an m observation of a variable (a latent factor, or an original NMR-trait or a classical lipid measure) ($m \times 1$) vector serving as a response; $\boldsymbol{X}$ is an $m \times p$ design matrix of the fixed effects with column vectors for the intercept, followed by SNP, sex and 10 PCs; $\boldsymbol{B}$ is the parameter $p \times 1$ vector of fixed effects, where p is the number of parameters fit in the model (an intercept + SNP effect + sex + 10 PCs, forming a total of 13 parameters fit in the model); $\boldsymbol{Z}$ is an $m \times q$ incidence matrix of random effects, and $\boldsymbol{U}$ is a $q \times 1$ vector of random effects with $\boldsymbol{E}(\boldsymbol{U}) = \boldsymbol{0}$, and covariance matrix $\boldsymbol{G}$; ε is an $m \times 1$ vector of random effects with $\boldsymbol{E}(\varepsilon) = \boldsymbol{0}$ and covariance matrix $\boldsymbol{R}$. Because of unequal variances the Kenward–Roger approximation for degrees of freedom for the reference distribution was used by defining the option ddfm=KR in the SAS MIXED procedure [20]. A SNP' additive genetic effect was tested to determine whether it was significantly different from zero. The same statistical model for association testing was implemented for top SNPs only, when tested for their association to NMR-single traits (corresponding to those contributing to a factor) and to single trait classical lipids.

The residuals from covariate adjustment stepwise regression for all single NMR-traits and for each classical lipids at a time-point were tested individually with only the best unique selected SNPs (one per genes/ intergenic) sourced from the association tests with factor scores with a $-\log_{10} p > 6$. Finally, although in tables reported are results with $-\log_{10} p > 6$, significant associations were considered only the ones that passed a threshold $-\log_{10} p > 7.3$ which corresponds to a p value $< 5 \times 10^{-8}$.

Results

The participants in the GOLDN study had a mean age of 48 (±16) years, a body mass index mean of 28–29 (±6) kg/m^2 among visits and a fasting glucose mean of 101 (±19) mg/dl. At different time points NMR-TAG and VLDL particularly, show modifications reflecting effects of FF and PPL challenge (Supplemental Table 1). The GOLDN NMR measurements correlated well with classical ultra-centrifugation lipid measurements. NMR-VLDL × classical TAG displayed correlations ranging from 0.7 to 1.0; NMR-LDL × classical LDL from 0.7 to 0.9; and NMR-HDL × classical HDL from 0.7–0.9. Such concordance was also identified in advance in a pilot sample of 78 subjects of the GOLDN study by Tsai et al. [18].

Factor Scores and NMR-Single Lipoprotein Association Tests Results

Sixteen NMR variables were employed in multivariate factor analyses only for the subjects with no missing observations. Each multivariate factor analysis at **v**isit 2 and blood **d**raw 1,v2d2,v2d3, v4d1,v4d2 and v4d3 produced 4 factors scores each (for statistical details see "Methods"), that captured up to 12 % of the original variables variance at a time point. Kraja et al. [21] reported factorial analyses for metabolic syndrome in the Family Blood Pressure Program, a large multi-ethnic network for studying genetic causes of blood pressure. The classical lipids (LDLC, HDLC and TAG) clustered with insulin and glucose in a separate factor out of 10 traits studied. The lipid-insulin factors explained in GENOA (The Genetic Epidemiology Network of Atherosclerosis) 11.9 % in African Americans, 9 % in whites, and 25.6 % in Hispanics; in HyperGEN (The Hypertension Genetic Epidemiology Network) 15.5 % in African Americans and 12.2 % in whites; and in the SAPPHIRe (The Stanford Asian Pacific Program in Hypertension and Insulin Resistance) study 16.8 % in Chinese and 13.2 % in Japanese of the total variance of 10 traits. The total variance explained by all factors representing domains of obesity-insulin, lipids-insulin, blood pressure and central obesity was respectively 58, 57, 56 % for GENOA, 59 and 55 % for HyperGEN and 58 and 57 % for SAPPHIRe studies [21]. This leads us to believe that GOLDN-measured NMRs are probably noisier than classical lipids, otherwise NMRs capture more individual variability of

the lipid fractions than commonality, which is not accounted in full by clusters of latent factors.

The GOLDN NMR factor scores were tested for association with all SNPs available genome-wide (for space limitation results not shown). The results of factor score associations with SNPs were utilized as a preliminary screening of the most important SNPs in relation to NMR-lipoprotein variables. Of these tests of 2,543,887 SNPs with factor scores, only 1,541 unique SNPs of all times and all factor associations were selected as top SNPs (unique one per gene/ intergenic with a threshold $-\log_{10} p > 6$). Because factor scores per se as phenotypes represent compound contributions of several NMR-single traits, the next step applied, was to test the association of 16 NMR-single lipoproteins and the LDLC, HDLC and TAG classical lipids with this reduced list of 1,541 unique SNPs.

NMR-LDL Traits (Supplemental Tables 2–5)

Among the top selected variants in association with NMR-Total LDL were rs13406427 (MAF = 0.27) near *CRIPT* gene (2p21), with a significant association at v2d1 ($-\log_{10} p = 7.27$), v2d3 ($-\log_{10} p = 7.92$) and less significant at v2d2 ($-\log_{10} p = 6.81$); rs350961 (MAF = 0.17) near *PCDH10* (4q28.3) with $-\log_{10} p = 8.37$ at v2d1, 6.21 at v2d2, 6.45 at v2d3, and 6.73 at v4d1; rs7842598 (MAF = 0.40) near *FAM84B* (8q24.21) with $-\log_{10} p = 8.33$ at v2d1, 6.36 at v4d1, and rs6987948 with $-\log_{10} p = 7.61$ at v2d2 and 7.06 at v2d3; rs1108161 (MAF = 0.12) an intron of *SLC24A4* (14q32.12) showed associations after FF treatment at times v4d1 ($-\log_{10} p = 9.72$), v4d2 ($-\log_{10} p = 9.89$) and at v4d3 ($-\log_{10} p = 9.68$). A few other associations were also of interest, rs13116085 (MAF = 0.49) an intron of *UNC5C* (4q21–23) with $-\log_{10} p = 9.4$ at v4d1 and relatively weaker associations at v2d1, v2d2, v2d3; and rs4415681 (MAF = 0.13) near *ADARB2* (10p15.3) with highest $-\log_{10} p = 7.25$ at v2d2 and weaker at times v2d1, v2d3. NMR-Large LDL captured rs1837369 (MAF = 0.48) an intron of *ACOXL* (2q13) with a $-\log_{10} p = 7.87$ at v2d3 and rs2241842 (MAF = 0.50) an intron of *BCL2L11* (2q13) also at v2d3, and rs701078 (MAF = 0.13) near *NCOR2* (12q24) with $-\log_{10} p = 7.73$ at v3d3. In contrast, NMR-Small LDL and NMR-LDL Size among others captured significant associations with rs10952132 (MAF = 0.20) near *NXPH1* (7p22) at times v2d1 and v2d3, respectively with $-\log_{10} p = 7.61$, 7.47 and $-\log_{10} p = 9.05$, 8.3.

NMR-HDL Traits (Supplemental Tables 6–10)

A number of SNPs associated with NMR-Total HDL including rs12478869 (MAF = 0.12) an intron of *VSNL1* (2p24.3) with $-\log_{10} p = 8.08$ in v2d1; rs6831020 (MAF = 0.30) near *KIT* (4q11–q12) with $-\log_{10} p = 8.06$ in v2d1 and 7.88 in v4d1; rs736201 (MAF = 0.36) near *HOMER1* (5q14.2) with $-\log_{10} p = 7.86$ in v2d1, 8.54 in v2d2 and 7.55 in v2d3; rs9922666 (MAF = 0.25) an intron of *QPRT* (16p11.2) with $-\log_{10} p = 7.46$ at v2d1; rs17015072 (MAF = 0.41) an intron of *SYNPR* (3p14.2) with $-\log_{10} p = 7.27$ at v2d2, 6.64 at v2d3, 6.92 at v4d2 and 7.01 at v4d3; rs11213927 (MAF = 0.06) an intron of *BTG4* (11q23) with $-\log_{10} p = 7.97$ in v2d2, and 7.32 in v2d3; and rs7203413 (MAF = 0.05) an intron of *NECAB2* (16q2.3) with $-\log_{10} p = 6.58$ in v2d1, 8.35 in v2d2, 7.91 in v2d3, 6.84 in v4d2, and 7.42 in v4d3. The variants near *QPRT, NECAB2, BTG4*, replicated for NMR-Large HDL also. In addition, rs10952132 (MAF = 0.20) near *NXPH1* (7p22) showed significant associations with NMR-Large HDL (8.08, 6.81, 7.98 at visit 2), and NMR-HDL Size (8.25, 7.35, 8.92 at visit 2); rs4244549 (MAF = 0.18) an intron of *NELL1* (11p15.1) with $-\log_{10} p = 8.85$ at v2d3 with NMR-HDL Size and 7.14 with NMR-Large HDL; rs380337 (MAF = 48 %) near *RUNX3* (1p36) with $-\log_{10} p = 8.98$, 8.24, and 6.84 in visit 2 with NMR-HDL Size.

NMR-VLDL Traits (Supplemental Tables 11–15)

Several SNPs associated significantly with NMR-Total VLDL including rs212600 (MAF = 0.22) near *CBLN4* (20q13) with a $-\log_{10} p = 7.61$ at v2d2, 7.29 at v2d3 and 8.22 at v4d2; rs7980440 (MAF = 0.46) near *GPR133* (12q24.33) with a $-\log_{10} p = 8.54$ at v2d3 and relatively weaker associations at v2d1 and v2d2; rs1955851 (MAF = 0.37) near *STXBP6* (14q12) with $-\log_{10} p = 8.35$ at v4d2, and 7.53 at v4d3; rs12069549 (MAF = 0.12) an intron of *NR5A2* (1q32.1) with a $-\log_{10} p = 7.83$ at v4d2; rs13001457 (MAF = 0.07) an intron of *SLC8A1* (2p23–22) with $-\log_{10} p = 7.07$; rs4796870 (MAF = 0.20) near *VAPA* (18p11.22) gene with a $-\log_{10} p = 7.87$; and a number of SNPs with significant associations in v4d3, rs7652652 (MAF = 0.10, $-\log_{10} p = 7.40$), rs9693089 (MAF = 0.35) an intron of *FAM167A* (8p23–22) ($-\log_{10} p = 8.16$), rs12479885 (MAF = 0.10) near *DEFB119* (20q11.21) ($-\log_{10} p = 7.52$) and rs17328705 (MAF = 0.10) near *DEFB123* (20q11.1) ($-\log_{10} p = 7.41$). The SNPs with gene proxies *NELL1* ($-\log_{10} p = 8.64$), *STXBP6* ($-\log_{10} p = 8.41$), and *CBLN4* ($-\log_{10} p = 7.96$) replicated at v4d3 for NMR-Large VLDL and for NMR-Medium VLDL the SNPs with gene proxies *GPR133* at v2d2 and v2d3 respectively with $-\log_{10} p = 8.62$ and $-\log_{10} p = 8.23$; *SLC8A1* at v4d1 ($-\log_{10} p = 7.81$); *DEFB119* ($-\log_{10} p = 8.23$) and *DEFB123* ($-\log_{10} p = 8.09$) at v4d3. The NMR-Small VLDL captured one significant association for rs6470475 (MAF = 0.42) near *FAM84B* ($-\log_{10} p = 7.60$), whereas for NMR-VLDL size rs11902417 (MAF = 0.24) near *APOB* (2p24–p23) showed significance at v2d1 ($-\log_{10} p = 7.01$), at v2d3 ($-\log_{10} p = 7.63$), at v4d3 ($-\log_{10} p = 7.09$), and rs6061612 (MAF = 22) an intron of *CDH4* ($-\log_{10} p = 7.77$).

NMR-IDL (Supplemental Table 16)

A few SNPs showed significant associations with NMR-IDL, rs11578216 (MAF = 0.20) a utr-variant of *RHOU* (1q42.11–q42.3, $-\log_{10} p = 7.44$) at v2d1; rs12874914 (MAF = 0.24) near LOC100652869 ($-\log_{10} p = 7.59$) and rs1182524 (MAF = 0.08) an intron of *PHACTR3* (20q13.32-q13.33, $-\log_{10} p = 7.01$) at v2d2; and rs10023694 (MAF = 0.25) an intron of *MARCH1* (4q32.2, $-\log_{10} p = 7.95$) at v4d1.

NMR-CHYL (Supplemental Table 17)

Five SNPs, rs7787327 (MAF = 0.07, $-\log_{10} p = 7.09$) and rs7189621 (MAF = 0.42, $-\log_{10} p = 7.21$) respectively near genes *GLI3* (7p13) and *CDH8* (16q22.1) at v4d2; and rs4640436 (MAF = MAF = 0.33, $-\log_{10} p = 7.55$), rs6573406 (MAF = 0.06, $-\log_{10} p = 7.16$), rs8015534 (MAF = 0.09, $-\log_{10} p = 7.49$) respectively near genes *HAAO* (2p21), *SYT16* (14q23.2) and *C14orf177* (14q32.2) associated with CHYL.

Classical Lipids Single-Trait Association Test Results

Next, the same list of 1,541 selected unique SNPs (the best per/near gene) extracted from GWAS with factor scores, were tested also for associations with classical ultra-centrifuged lipids single-traits. Results for TAG, LDLC and HDLC are shown respectively in Supplemental Tables 18, 19 and 20. A number of SNPs exhibited significant associations with classical TAG, rs212600 near *CBLN4* ($-\log_{10} p = 7.68$, v2d2, 8.16, v2d3 and 8.19, v4d2); rs7980440 near *GPR133* ($-\log_{10} p = 8.53$, v2d3); rs12069549 an intron of *NR5A2* ($-\log_{10} p = 7.38$, v4d2); rs1955851 near *STXBP6*, ($-\log_{10} p = 8.40$, v4d2 and 7.97 at v4d3); rs10179753 near *SLC8A1* ($-\log_{10} p = 7.80$, v4d3) and rs12479885 near *DEFB119* and rs17328705 near *DEF123* respectively with $-\log_{10} p = 7.40$ and 7.26 both at v4d3.

The classical HDLC was measured only at two time points, v2d1 and v4d1. Several SNPs showed association with classical HDLC including rs3007220 an intron of *STK40* ($-\log_{10} p = 7.45$ and 6.37); rs12478869 an intron of *VSNL1* ($-\log_{10} p = 8.47$ and 6.71); rs6831020 near *KIT* ($-\log_{10} p = 8.38$ and 7.16); rs736201 near *HOMER1* ($-\log_{10} p = 7.42$ and 6.80);

rs9922666 an intron of *QPRT* ($-\log_{10} p = 7.83$ and 7.71); rs173539 near *CETP* ($-\log_{10} p = 8.47$ and 8.00); and rs9937660 an intron of *NECAB2* ($-\log_{10} p = 7.99$ and 7.69).

Also classical LDLC was measured at baseline time points, v2d1 and v4d1. Among significant associations were rs350961 near *PCDH10* ($-\log_{10} p = 7.52$) and rs7842598 near *FAM84B* ($-\log_{10} p = 7.96$) at v2d1; and rs10179753 near *SLC8A1* ($-\log_{10} p = 7.02$), rs1108161 an intron of *SLC24A4* at v4d1.

Discussion

The GOLDN Study is one of the largest of its kind studying a two arm treatment, uptake and clearance of PPL and FF effects. As the first step we chose to analyze the NMR data cross-sectionally, which produced snapshots at 6 time points of NMR traits associated with DNA variants/genes. Other colleagues of the GOLDN study are studying slopes of uptake and clearance and the area under the curve. Although the results of our study are compelling, because of the relatively small sample size available in the association tests (more than 700 to 1,000 subjects analyzed), a need for replication is warranted. Following the discussion we provide several examples where our findings replicate previous reported findings, as well as underline some interesting findings with less or missing prior evidence.

Factor scores of association test results were explored for reducing the number of tests at each time-point, from 16 NMR-lipoprotein single traits and 3 classical lipid measurements to only 4 composite factor scores, thus protecting against random false results genome-wide. Based on a $-\log_{10} p > 6$, only 1,541 unique SNPs (the best per gene/intergenic) were selected for further testing with NMR-lipoprotein single traits and classical lipid measurements. Of these, 67 single unique significant SNP-NMR single traits showed significant associations at the PPL uptake/clearance, 20 at baseline and 11 at baseline and uptake/clearance of fat. By design (see Supplemental Fig. 1) there is a double chance to have uptake/clearance findings compared to baseline. Yet, half the number of uptake/clearance significant findings (67/2) were higher than the one at baseline. These differences, quite possibly, reflect differential genetic individual responses on PPL and FF treatments. For example, the significant overlap (NMR-single traits vs. classical lipids) findings for *GPR133, MAFB*, and *STK40* represent associations mainly responding to PPL treatment. In contrast, several associations were deemed significant after the PPL and FF treatments (*CBLN4, DEFB119, DEFB123, NECAB2, NR3A2, SLC24A4* and *STXBP6*) and only two at baseline after the FF treatment (*PLXDC2* and *SLC8A1*). Of the overlapping results between NMR-single traits and classical lipids, the following proxy genes *CACNG3, EEPD1, KIT, FAM84B, PCDH10, QPRT*, and *VSNL1* were tagged with significant SNP associations at the baseline. In these findings there are a number of new interesting gene discoveries. For example, the rs6987948 (MAF = 0.38) near *FAM84B* showed significant associations with NMR-Total LDL in v2d2 and v2d3 (Supplemental Table 2) and rs6470475 (MAF = 0.42) in v2d1 with NMR-Small VLDL (Supplemental Table 14). Variants near *FAM84B* gene represent new findings to lipid metabolism, which for GOLDN data do not remain significant after FF treatment. Recent information in the dbGAP Association Viewer (Supplemental Table 21) shows two SNPs, rs4345522 and rs2385686 intergenic to *FAM84B*, associated with heart failure ($-\log_{10} p = 3.28$) and myocardial infarction ($-\log_{10} p = 4.2$). About 1.3 Mbps near the above findings, rs11987875 near *TRIB1* gene shows similar patterns of significance with NMR-Total LDL. Deliberately, we did not mention this SNP in the results because it represents an MAF of 0.01. Very close to this variant, rs2001945 associates with TAG ($-\log_{10} p = 20$) [22], rs2954021 with LDLC ($-\log_{10} p = 7$) [23] and with TAG × HDLC ($-\log_{10} p = 8.1$) [24], rs17321515 with TAG ($-\log_{10} p = 16.4$) [25]. In mice, *TRIB1* is reported as a regulator of hepatic lipogenesis and VLDL production

and associated with myocardial infarction [26]. Thus, variants neighboring *FAM84B* yield evidence of pleiotropic effects to lipid fractions and are influenced by FF treatment.

Factor scores served also in discovering pleiotropic effects, by combining in a factor the correlated NMR-single traits. Seven proxy genes showed pleiotropic effects based on NMR traits, *ACOXL* for large LDL and HDL-size; *BCL2L11* for large LDL and total VLDL; *FAM84B* for total LDL and small VLDL; *NELL1* for large HDL, HDL size and large VLDL; *NXPH1* for small LDL, LDL size, large HDL, and HDL size; *STK40* for large LDL and large HDL; and *VAPA* for total HDL and total VLDL. Although *ACOXL* (acyl-Coenzyme A oxidase-like) does not have any publication for its association with lipids, GO terms connect it with fatty-acid beta oxidation, a process of energy production. Moreover, rs2015454 of the same gene showed a $-\log_{10} p$ of 3.6 in association with myocardial infarction (dbGaP Viewer results). In contrast, *NELL1* has associated with LDLC ($-\log_{10} p = 4.5$) [27] and *VAPA* with HDLC with a $-\log_{10} p = 4.5$ [27]. A special case is represented by *NXPH1* associated with LDLC, HDLC and TAG [27], and protein quantitative trait loci [28], T2D [29], CRP [30] and blood pressure (Supplemental Table 21). Although it is known that Neurexophilin forms a tight complex with alpha neurexins that promote adhesion between dendrites and axons, there is not yet biological information that connects *NXPH1* and lipids.

Different from several variants that associated with VLDL particles at baseline, the rs6069250 (MAF = 0.15) displayed significant association near *CBLN4*, contributing more importantly in the uptake of lipid particles. Cerebellin is found mainly in the adrenal medulla, where it has been shown to have a neuromodulatory function. Another finding with similar association trend to lipid uptake was rs7980440 (MAF = 0.46) near *GPR133. GPR133* has been reported for its association with electrocardiography measurements [31]. At least three variants associated with VLDL particles during the lipid clearance, rs4640436 near *HAAO*, rs4244549 an intron of *NELL1* at v4d3, and rs1955851 near *STXBP6* showed most of the significant associations with NMR-Total VLDL during PPL challenge (v4d2 and v4d3). Another important finding was association of rs11902417 near *APOB* at v2d1, v2d3 and v4d3 with NMR-VLDL Size (Supplemental Table 15). This gene's product is the main apolipoprotein of chylomicrons and low density lipoproteins and is well studied to associate with HDLC, TAG and LDLC [32].

Of course, rs173539 near *CETP* (16q21) and rs397923 near *LIPC* (15q21–q23) were quite of interest in association with HDLC and NMR-HDL particles. These SNPs have been reported already for associations with HDLC [24, 33]. However, in our study there are a number of other SNPs that tagged *HOMER1, KIT, VSNL1, QPRT, SYNPR, NXPH1, NELL1* and *RUNX3* distinctively associated with NMR-HDL particles.

Some of the findings replicate in a few time points. For example, the intron rs1108161 of *SLC24A4* (14q32.12) a sodium/potassium/calcium exchanger associated with NMR- total LDL at v4d1, v4d2 and v4d3, and with classical LDLC at v4d1. In the vicinity of this SNP, the rs4904947 (intergenic to *SLC24A4* and about 110 K distant from rs1108161) is reported to be associated with all-cause death on dialysis in a sample of African Americans with T2D and with high end-stage renal disease ($n = 619$, $-\log_{10} p = 5.1$) [34]. But it is not clear why lipoprotein fractions associate with an intron of solute carrier 24. Other variants were significant only at visit 4, as for example rs13001457 an intron of *SLC8A1* (2p23–p22) a sodium/ calcium exchanger, that significantly associated with NMR-Medium VLDL at blood draw 1. The *SLC8A1* is an important contributor in returning the cardiac myocyte to its resting state following excitation. Previous research has also shown these captivating associations between lipids and *SLC8A1*. For example, rs10490369 an intergenic variant of

SLC8A1 is reported to associate with cholesterol ($-\log_{10} p = 10.04$) and with LDLC ($-\log_{10} p = 8.51$) to mention a few (Supplemental Table 21).

A gene new to the literature of associations with NMR lipoproteins was *CBLN4* flanked by *DOK5* and *MC3R. DOK5* is a membrane protein involved in signal transduction and reported to be associated with T2D in Northern Indians [35], and *MC3R* (melanocortin 3 receptor), with a probable role in the regulation of energy homeostasis, because mice deficient in this gene have increased fat mass despite decreased food intake [36, 37]. There is no clear function for *CBLN4*, tagged by rs212600. One can see (Supplemental Table 1) the effect of FF displayed as a lower level of NMR-TAG and VLDL particles in v4d1 than v2d1, and higher levels at blood draws 2 and 3, but still remaining lower at visit 4 compared to visit 2 counterparts. Especially levels of NMR-TAG and Large and Total VLDL show the trend of fat clearance at v4d3. This setting, of TAG and VLDL differences in response to FF and being at the process of clearing fat, we reason delivered the identification of such variant that may point toward a haplotype of genes (*DOK5-CBLN4-MC3R*) probably involved in cholesterol clearance.

One of the largest NMR studies was published by Chasman et al. (2009) [10]. They studied 17,296 subjects of Women's Genome Health Study, 12,489 of them had complied with fasting requirements. Twenty two NMR-based lipoprotein fractions and size were studied implementing an additive model in association tests with 335,603 SNPs genome-wide. Of their 43 reported significant associations, findings for *APOB, LIPC, CETP*, and *TRIB1* overlap with our study. A number of other findings are in about a 1-12 MB distance apart. For example, our finding for *BTG4* vs. their finding for the cluster of *APOA1-5* (~5 MB), *NCOR2* vs. *HNF1A* (~3 MB), *TCF4* vs. *LIPG* (~5 MB), *TTC32* vs. *APOB* (~1 MB), *TANC1* vs. *COBLL1* (~6 MB), *CBLN4* vs. *PLTL* (~10 MB), *CPNE4* vs. *PCCB* (~5 MB), *HOMER1* vs. *HMGCR* (~4 MB), *CDKAL1* vs. *BTNL2* (~12 MB), *FAM167A* vs. *PPP1R3B* (~3 MB), and *ATP6V1G1* vs. *ABCA1* (~10 MB). The two studies also have differences in the power of detecting new discoveries. In the time of Chasman et al. publication the WGHS had strength in the large sample (N~17,296) compared to GOLDN (N~1,000–1,300 with phenotypes depending on the visits), but GOLDN has a larger "hybrid" genotyping (~2,5 M SNPs) compared to WGHS (0.34 M SNPs). In addition, WGHS studied 22 NMR-measurements at baseline, whereas GOLDN studies 16 NMR-phenotypes at baseline (time 0), at uptake of PPL (3.5 h after taking the fat shake) and clearance of PPL (6 h) before (visit 2) and after FF (visit 4) treatment. These treatments create the opportunity for GOLDN study to see a larger number of associations under more differentiating response to treatment. Moreover, in the WGHS paper the analyses were performed as single trait-single SNP associations, whereas in our study a multivariate method based on factor analysis is used as a primary screen to identify the most important associations. At least two are advantages of factor analysis in this case, (a) it reduces the number of test-comparisons out of 16 traits per time point to only 4 latent factors in the primary screening and after it we test only the top significant SNPs with the corresponding original traits that contribute to a particular factor (see Supplemental Figs. 2–7 and Methods); (b) A latent NMR-factor is a stronger summary of a number of correlated contributing variables than a single NMR-trait, which helps also discovering gene pleiotropic effects. As a result, out of these comparisons a number of differences remain, but also 35 % of the 43 loci reported in the WGHS study compared to GOLDN NMR-spectroscopy phenotypes do match exact or they had a relatively close distance with our findings. They may point to clusters of genes that effect lipid metabolism.

NMR lipoprotein measurements identified more associations than those with classical lipid measurements alone. The fact that 18 % of significant variants for NMR-lipoprotein particles can be found significant simultaneously with classical lipids in the GOLDN study

is encouraging. Of the overlapped significant findings between classical lipids and NMR particles, 4 SNPs for classic LDLC matched 1 to 1 and 2 SNPs in association with classic LDLC associated with 2 NMR particles (a ratio of 4/2); for HDL 4 SNPs matched 1 on 1 and 3 SNPs from 1 to 2 (a ratio of 4/3); for TAG 3 SNPs matched 1 on 1 from classic to NMR VLDL particles and 6 SNPs from 1 to 2 (a ratio 3/6). This indicates that NMR particles delivered additional information about associations of a particular variant than classical lipid measurements. In addition, the remaining 75 % were only NMR-single traits discoveries and 7 % were significant only with classical lipid traits. Sixteen percent of unique significant associations based on NMR-particles matched for the time of discovery with classical lipids (shown with red letters in Table 1). It is encouraging that several findings related with genes known as contributors to lipid metabolism and downstream to T2D and CVD outcomes (Supplemental Table 21). Furthermore, by searching the GeneGO database of the top 50 GO processes captured by the list of 21 overlapping genes (between NMR-traits and classical lipids), 35 processes were related to lipids (Supplemental Table 22).

This study also has limitations. NMR-spectroscopy non-destructively determines a substance's composition and as a technique has been improving its accuracy. However, our factor analyses capturing up to 12 % of variation of 16 NMR-variables, may point out that these NMR measurements are more variable (containing more unexplained variability) than classical lipid measurements. Moreover, although we were expecting IDL and CHYL variables to have produced new and important information, these particles were less informative compared to NMR- LDL, HDL and VLDL association results. This also may point to a weakness of NMR technology. The other weakness of our study is the fact, that although GOLDN Study represents one of the largest for studying FF effects and PPL challenge in humans, its sample size for association tests is relatively small when considering that some of the results reported in the tables and figures are with MAF less than 5 %, which may limit its generalizability, unless replicated. It was positive that our study showed, that discerning atherogenic particles is of interest and a data-driven systems biology approach to biomedical research, necessitates the study of NMR metabolites, for which at present there is a great need for more affordable assays costs [38]. The GOLDN study is preparing for a metabolomics expansion and exome sequencing-level data, which will improve further our understanding of PPL and FF effects and lipid metabolism and help in clarifying the potential value of the most important genetic findings. Follow up at the molecular, physiological, biochemical and model organism levels especially in understanding functions of the important variants/genes in lipid metabolism is required.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

The authors are thankful to Jim Otvos of Liposciences Inc. for providing comments on a preliminary draft of the manuscript. This work was supported in part by the GOLDN NIH grant R01 HL09135701.

References

1. Third Report of the National Cholesterol Education Program (NCEP) Expert Panel on Detection, Evaluation, and Treatment of High Blood Cholesterol in Adults (Adult Treatment Panel III) final report. Circulation. 2002; 106(25):3143–3421. [PubMed: 12485966]
2. Carmena R, Duriez P, Fruchart JC. Atherogenic lipoprotein particles in atherosclerosis. Circulation. 2004; 109(23 Suppl 1):III2–III7. [PubMed: 15198959]

3. Otvos JD, Jeyarajah EJ, Bennett DW, Krauss RM. Development of a proton nuclear magnetic resonance spectroscopic method for determining plasma lipoprotein concentrations and subspecies distributions from a single, rapid measurement. Clin Chem. 1992; 38(9):1632–1638. [PubMed: 1326420]
4. Garvey WT, Kwon S, Zheng D, Shaughnessy S, Wallace P, Hutto A, Pugh K, Jenkins AJ, Klein RL, Liao Y. Effects of insulin resistance and type 2 diabetes on lipoprotein subclass particle size and concentration determined by nuclear magnetic resonance. Diabetes. 2003; 52(2):453–462. [PubMed: 12540621]
5. Blake GJ, Albert MA, Rifai N, Ridker PM. Effect of pravastatin on LDL particle concentration as determined by NMR spectroscopy: a substudy of a randomized placebo controlled trial. Eur Heart J. 2003; 24(20):1843–1847. [PubMed: 14563343]
6. Nofer JR. Hyperlipidemia and cardiovascular disease: triglycerides—a revival of cardiovascular risk factor? Curr Opin Lipidol. 2011; 22(4):319–321. [PubMed: 21743309]
7. Lada AT, Rudel LL. Associations of low density lipoprotein particle composition with atherogenicity. Curr Opin Lipidol. 2004; 15(1):19–24. [PubMed: 15166804]
8. Kettunen J, Tukiainen T, Sarin AP, Ortega-Alonso A, Tikkanen E, Lyytikainen LP, Kangas AJ, Soininen P, Wurtz P, Silander K, et al. Genome-wide association study identifies multiple loci influencing human serum metabolite levels. Nat Genet. 2012; 44(3):269–276. [PubMed: 22286219]
9. Mora S, Otvos JD, Rosenson RS, Pradhan A, Buring JE, Ridker PM. Lipoprotein particle size and concentration by nuclear magnetic resonance and incident type 2 diabetes in women. Diabetes. 2010; 59(5):1153–1160. [PubMed: 20185808]
10. Chasman DI, Pare G, Mora S, Hopewell JC, Peloso G, Clarke R, Cupples LA, Hamsten A, Kathiresan S, Malarstig A, et al. Forty-three loci associated with plasma lipoprotein size, concentration, and cholesterol content in genome-wide analysis. PLoS Genet. 2009; 5(11):e1000730. [PubMed: 19936222]
11. Jenkins AJ, Lyons TJ, Zheng D, Otvos JD, Lackland DT, McGee D, Garvey WT, Klein RL. Lipoproteins in the DCCT/EDIC cohort: associations with diabetic nephropathy. Kidney Int. 2003; 64(3):817–828. [PubMed: 12911531]
12. Otvos JD, Collins D, Freedman DS, Shalaurova I, Schaefer EJ, McNamara JR, Bloomfield HE, Robins SJ. Low-density lipoprotein and high-density lipoprotein particle subclasses predict coronary events and are favorably changed by gemfibrozil therapy in the Veterans Affairs High-Density Lipoprotein Intervention Trial. Circulation. 2006; 113(12):1556–1563. [PubMed: 16534013]
13. Freedman DS, Otvos JD, Jeyarajah EJ, Shalaurova I, Cupples LA, Parise H, D'Agostino RB, Wilson PW, Schaefer EJ. Sex and age differences in lipoprotein subclasses measured by nuclear magnetic resonance spectroscopy: the Framingham Study. Clin Chem. 2004; 50(7):1189–1200. [PubMed: 15107310]
14. Kaess BM, Tomaszewski M, Braund PS, Stark K, Rafelt S, Fischer M, Hardwick R, Nelson CP, Debiec R, Huber F, et al. Large-scale candidate gene analysis of HDL particle features. PLoS ONE. 2011; 6(1):e14529. [PubMed: 21283740]
15. Lai CQ, Arnett DK, Corella D, Straka RJ, Tsai MY, Peacock JM, Adiconis X, Parnell LD, Hixson JE, Province MA, et al. Fenofibrate effect on triglyceride and postprandial response of apolipoprotein A5 variants: the GOLDN study. Arterioscler Thromb Vasc Biol. 2007; 27(6):1417–1425. [PubMed: 17431185]
16. Kraja AT, Province MA, Straka RJ, Ordovas JM, Borecki IB, Arnett DK. Fenofibrate and metabolic syndrome. Endocr Metab Immune Disord Drug Targets. 2010; 10(2):138–148. [PubMed: 20406163]
17. Patsch JR, Miesenbock G, Hopferwieser T, Muhlberger V, Knapp E, Dunn JK, Gotto AM Jr, Patsch W. Relation of triglyceride metabolism and coronary artery disease. Studies in the postprandial state. Arterioscler Thromb. 1992; 12(11):1336–1345. [PubMed: 1420093]
18. Tsai MY, Georgopoulos A, Otvos JD, Ordovas JM, Hanson NQ, Peacock JM, Arnett DK. Comparison of ultracentrifugation and nuclear magnetic resonance spectroscopy in the quantification of triglyceride-rich lipoproteins after an oral fat load. Clin Chem. 2004; 50(7):1201–1204. [PubMed: 15142979]

19. Otvos JD. Measurement of lipoprotein subclass profiles by nuclear magnetic resonance spectroscopy. Clin Lab. 2002; 48(3–4):171–180. [PubMed: 11934219]
20. Kenward MG, Roger JH. Small sample inference for fixed effects from restricted maximum likelihood. Biometrics. 1997; 53(3):983–997. [PubMed: 9333350]
21. Kraja AT, Rao DC, Weder AB, Mosley TH, Turner ST, Hsiung CA, Quertermous T, Cooper R, Curb JD, Province MA. An evaluation of the metabolic syndrome in a large multi-ethnic study: the Family Blood Pressure Program. Nutr Metab (Lond). 2005; 2:17. [PubMed: 16076393]
22. Kim YJ, Go MJ, Hu C, Hong CB, Kim YK, Lee JY, Hwang JY, Oh JH, Kim DJ, Kim NH, et al. Large-scale genome-wide association studies in East Asians identify new genetic loci influencing metabolic traits. Nat Genet. 2011; 43(10):990–995. [PubMed: 21909109]
23. Waterworth DM, Ricketts SL, Song K, Chen L, Zhao JH, Ripatti S, Aulchenko YS, Zhang W, Yuan X, Lim N, et al. Genetic variants influencing circulating lipid levels and risk of coronary artery disease. Arterioscler Thromb Vasc Biol. 2010; 30(11):2264–2276. [PubMed: 20864672]
24. Kraja AT, Vaidya D, Pankow JS, Goodarzi MO, Assimes TL, Kullo IJ, Sovio U, Mathias RA, Sun YV, Franceschini N, et al. A bivariate genome-wide approach to metabolic syndrome: STAMPEED consortium. Diabetes. 2011; 60(4):1329–1339. [PubMed: 21386085]
25. Kathiresan S, Melander O, Guiducci C, Surti A, Burtt NP, Rieder MJ, Cooper GM, Roos C, Voight BF, Havulinna AS, et al. Six new loci associated with blood low-density lipoprotein cholesterol, high-density lipoprotein cholesterol or triglycerides in humans. Nat Genet. 2008; 40(2):189–197. [PubMed: 18193044]
26. Burkhardt R, Toh SA, Lagor WR, Birkeland A, Levin M, Li X, Robblee M, Fedorov VD, Yamamoto M, Satoh T, et al. Trib1 is a lipid- and myocardial infarction-associated gene that regulates hepatic lipogenesis and VLDL production in mice. J Clin Invest. 2010; 120(12):4410–4414. [PubMed: 21084752]
27. Kathiresan S, Manning AK, Demissie S, D'Agostino RB, Surti A, Guiducci C, Gianniny L, Burtt NP, Melander O, Orho-Melander M, et al. A genome-wide association study for blood lipid phenotypes in the Framingham Heart Study. BMC Med Genet. 2007; 8(Suppl 1):S17. [PubMed: 17903299]
28. Melzer D, Perry JR, Hernandez D, Corsi AM, Stevens K, Rafferty I, Lauretani F, Murray A, Gibbs JR, Paolisso G, et al. A genome-wide association study identifies protein quantitative trait loci (pQTLs). PLoS Genet. 2008; 4(5):e1000072. [PubMed: 18464913]
29. Meigs JB, Manning AK, Fox CS, Florez JC, Liu C, Cupples LA, Dupuis J. Genome-wide association with diabetes-related traits in the Framingham Heart Study. BMC Med Genet. 2007; 8(Suppl 1):S16. [PubMed: 17903298]
30. Benjamin EJ, Dupuis J, Larson MG, Lunetta KL, Booth SL, Govindaraju DR, Kathiresan S, Keaney JF Jr, Keyes MJ, Lin JP, et al. Genome-wide association with select biomarker traits in the Framingham Heart Study. BMC Med Genet. 2007; 8(Suppl 1):S11. [PubMed: 17903293]
31. Marroni F, Pfeufer A, Aulchenko YS, Franklin CS, Isaacs A, Pichler I, Wild SH, Oostra BA, Wright AF, Campbell H, et al. A genome-wide association scan of RR and QT interval duration in 3 European genetically isolated populations: the EUROSPAN project. Circ Cardiovasc Genet. 2009; 2(4):322–328. [PubMed: 20031603]
32. Teslovich TM, Musunuru K, Smith AV, Edmondson AC, Stylianou IM, Koseki M, Pirruccello JP, Ripatti S, Chasman DI, Willer CJ, et al. Biological, clinical and population relevance of 95 loci for blood lipids. Nature. 2010; 466(7307):707–713. [PubMed: 20686565]
33. Kathiresan S, Willer CJ, Peloso GM, Demissie S, Musunuru K, Schadt EE, Kaplan L, Bennett D, Li Y, Tanaka T, et al. Common variants at 30 loci contribute to polygenic dyslipidemia. Nat Genet. 2009; 41(1):56–65. [PubMed: 19060906]
34. Murea M, Lu L, Ma L, Hicks PJ, Divers J, McDonough CW, Langefeld CD, Bowden DW, Freedman BI. Genome-wide association scan for survival on dialysis in African-Americans with type 2 diabetes. Am J Nephrol. 2011; 33(6):502–509. [PubMed: 21546767]
35. Tabassum R, Mahajan A, Chauhan G, Dwivedi OP, Ghosh S, Tandon N, Bharadwaj D. Evaluation of DOK5 as a susceptibility gene for type 2 diabetes and obesity in North Indian population. BMC Med Genet. 2010; 11:35. [PubMed: 20187968]

36. Santos JL, De la Cruz R, Holst C, Grau K, Naranjo C, Maiz A, Astrup A, Saris WH, MacDonald I, Oppert JM, et al. Allelic variants of melanocortin 3 receptor gene (MC3R) and weight loss in obesity: a randomised trial of hypo-energetic high-versus low-fat diets. PLoS ONE. 2011; 6(6):e19934. [PubMed: 21695122]

37. Mencarelli M, Dubern B, Alili R, Maestrini S, Benajiba L, Tagliaferri M, Galan P, Rinaldi M, Simon C, Tounian P, et al. Rare melanocortin-3 receptor mutations with in vitro functional consequences are associated with human obesity. Hum Mol Genet. 2011; 20(2):392–399. [PubMed: 21047972]

38. Soininen P, Kangas AJ, Wurtz P, Tukiainen T, Tynkkynen T, Laatikainen R, Jarvelin MR, Kahonen M, Lehtimaki T, Viikari J, et al. High-throughput serum NMR metabonomics for cost-effective holistic studies on systemic metabolism. Analyst. 2009; 134(9):1781–1785. [PubMed: 19684899]

Abbreviations

CHYL	Chylomicrons
dbGaP	The database of Genotypes and Phenotypes
FF	Fenofibrate
GOLDN	Genetics of Lipid Lowering Drugs and Diet Network study
HDL	High density lipoprotein
IDL	Intermediate density lipoproteins
LDL	Low density lipoprotein
MAF	Minor allele frequency
NMR	Nuclear magnetic resonance spectroscopy lipoprotein measurements
NCBI	The National Center for Biotechnology Information
PPL	Oral fat challenge
SNPs	Single nucleotide polymorphisms
TC	Total cholesterol
TAG	Triglycerides
VLDL	Very low density lipoprotein

Lipids. Author manuscript; available in PMC 2014 February 01.

Table 1

A summary of significant NMR-single traits and classical lipids SNPs associations tagging proxy genes, with details analogous to the ones produced in Supplemental Tables 2–20

No	GENE	TOTAL LDL	LARGE LDL	SMALL LDL	LDL SIZE	TOTAL HDL	LARGE HDL	MEDIUM HDL	SMALL HDL	HDL SIZE	TOTAL VLDL	LARGE VLDL	MEDIUM VLDL	SMALL VLDL	VLDL SIZE	IDL	CHYL	TG	HDLC	LDLC	GeneName	Position
1	ABCC11		D																		ATP-binding cassette, sub-family C (CFTR/MRP), member 11	chr16:q12.1
2	ABCC12		D																		ATP-binding cassette, sub-family C (CFTR/MRP), member 12	chr16:q12.1
3	ACOXL		C							C											acyl-CoA oxidase-like	chr2:q13
4	ADARB2	B																			adenosine deaminase, RNA-specific, B2	chr10:p15.3
5	APOB														A, C, F						apolipoprotein B (including Ag(x) antigen)	chr2:p24.1
6	ASPSCR1																			D	alveolar soft part sarcoma chromosome region, candidate 1	chr17:q25.3
7	ATP6V1G1																	E			ATPase, H+ transporting, lysosomal 13kDa, V1 subunit G1	chr9:q32
8	B3GAT2						D, F			D, E, F											beta-1,3-glucuronyltransferase 2 (glucuronosyltransferase S)	chr6:q13
9	BCL2L11		C							C											BCL2-like 11 (apoptosis facilitator)	chr2:q13
10	BTG4					B, C	A														B-cell translocation gene 4	chr11:q23.1
11	BZW2																	E, F			basic leucine zipper and W2 domains 2	chr7:p21.1
12	C10ORF47										F										chromosome 10 open reading frame 47	chr10:p14
13	C14ORF177																F				chromosome 14 open reading frame 177	chr14:q32.2
14	C1ORF174						D			D											chromosome 1 open reading frame 174	chr1:p36.32
15	C1ORF52						F														chromosome 1 open reading frame 52	chr1:p22.3
16	C5ORF48										F										chromosome 5 open reading frame 48	chr5:q23.2
17	C9ORF79	B																			chromosome 9 open reading frame 79	chr9:q22.1
18	CACNG3	A, B																		A	calcium channel, voltage-dependent, gamma subunit 3	chr16:p12.1
19	CBLN4										B, C, E	F						B, C, E			cerebellin 4 precursor	chr20:q13.2
20	CCDC41						E														coiled-coil domain containing 41	chr12:q22
21	CCDC79		D																		coiled-coil domain containing 79	chr16:q22.1
22	CDH4														E						cadherin 4, type 1, R-cadherin (retinal)	chr20:q13.33
23	CDH8																E				cadherin 8, type 2	chr16:q21
24	CDKAL1												B, F								CDK5 regulatory subunit associated protein 1-like 1	chr6:p22.3
25	CETP																		A, D		cholesteryl ester transfer protein, plasma	chr16:q13
26	COL27A1		B																		collagen, type XXVII, alpha 1	chr9:q32
27	CPEB2																		D		cytoplasmic polyadenylation element binding protein 2	chr4:4p15.33
28	CPNE4										F	F									copine IV	chr3:q22.1
29	CSMD1																	F			CUB and Sushi multiple domains 1	chr8:p23.2
30	CRIPT	A, C																			cysteine-rich PDZ-binding protein	chr2:p21
31	DEFB119										F		F					F			defensin, beta 119	chr20:q11.21
32	DEFB123										F		F					F			defensin, beta 123	chr20:q11.21
33	DEFB133											B									defensin, beta 133	chr6:p12.3
34	DISC1		B, C																		disrupted in schizophrenia 1	chr1:q42.2
35	DYNC1LI2		D																		dynein, cytoplasmic 1, light intermediate chain 2	chr16:q22.1
36	EEPD1						A												A		endonuclease/exonuclease/phosphatase family domain containing 1	chr7:p14.2
37	EFCAB4B						E			E, F											EF-hand calcium binding domain 4B	chr12:p13.32

No	GENE	TOTAL LDL	LARGE LDL	SMALL LDL	LDL SIZE	TOTAL HDL	LARGE HDL	MEDIUM HDL	SMALL HDL	HDL SIZE	TOTAL VLDL	LARGE VLDL	MEDIUM VLDL	SMALL VLDL	VLDL SIZE	IDL	CHYL	TG	HDLC	LDLC	GeneName	Position
38	EGLN3						C			C											egl nine homolog 3 (C. elegans)	chr14:q13.1
39	EMCN																E				endomucin	chr4:q24
40	FAM167A										F										family with sequence similarity 167, member A	chr8:p23.1
41	FAM63B	B																			family with sequence similarity 63, member B	chr15:q21.3
42	FAM84B	A, B, C												A						A	family with sequence similarity 84, member B	chr8:q24.21
43	FLRT2											F									fibronectin leucine rich transmembrane protein 2	chr14:q31.3
44	FSHR									B, F											follicle stimulating hormone receptor	chr2:p16.3
45	FTMT						A														ferritin mitochondrial	chr5:q23.1
46	GDF10												D								growth differentiation factor 10	chr10:q11.22
47	GDF2												D								growth differentiation factor 2	chr10:q11.22
48	GLI3																E				GLI family zinc finger 3	chr7:p14.1
49	GPR133										C		B, C					C			G protein-coupled receptor 133	chr12:q24.33
50	HAAO																F				3-hydroxyanthranilate 3,4-dioxygenase	chr2:p21
51	HOMER1				A, B, C														A		homer homolog 1 (Drosophila)	chr5:q14.1
52	IQCJ-SCHIP1											D									IQ motif containing J-schwannomin interacting protein 1 fusion protein	chr3
53	KIAA1217							F													KIAA1217	chr10:p12.2
54	KIT				A, D														A, D		v-kit Hardy-Zuckerman 4 feline sarcoma viral oncogene homolog	chr4:q12
55	KRTAP5-6																	A		D	keratin associated protein 5-6	chr11:p15.5
56	LIPC					A															lipase, hepatic	chr15:q21.3
57	LMO7													E			E				LIM domain 7	chr13:q22.2
58	LOC100129636											F									NA	chr6:p22.1
59	LOC100132713										E										NA	chr18:q23
1	ABCC11		D																		ATP-binding cassette, sub-family C (CFTR/MRP), member 11	chr16:q12.1
2	ABCC12		D																		ATP-binding cassette, sub-family C (CFTR/MRP), member 12	chr16:q12.1
3	ACOXL		C							C											acyl-CoA oxidase-like	chr2:q13
4	ADARB2	B																			adenosine deaminase, RNA-specific, B2	chr10:p15.3
5	APOB														A, C, F						apolipoprotein B (including Ag(x) antigen)	chr2:p24.1
6	ASPSCR1																			D	alveolar soft part sarcoma chromosome region, candidate 1	chr17:q25.3
7	ATP6V1G1																	E			ATPase, H+ transporting, lysosomal 13kDa, V1 subunit G1	chr9:q32
8	B3GAT2						D, F			D, E, F											beta-1,3-glucuronyltransferase 2 (glucuronosyltransferase S)	chr6:q13
9	BCL2L11		C							C											BCL2-like 11 (apoptosis facilitator)	chr2:q13
10	BTG4					B, C	A														B-cell translocation gene 4	chr11:q23.1
11	BZW2																	E, F			basic leucine zipper and W2 domains 2	chr7:p21.1
12	C10ORF47										F										chromosome 10 open reading frame 47	chr10:p14
13	C14ORF177																F				chromosome 14 open reading frame 177	chr14:q32.2
14	C1ORF174						D			D											chromosome 1 open reading frame 174	chr1:p36.32
15	C1ORF52						F														chromosome 1 open reading frame 52	chr1:p22.3
16	C5ORF48										F										chromosome 5 open reading frame 48	chr5:q23.2
17	C9ORF79	B																			chromosome 9 open reading frame 79	chr9:q22.1
18	CACNG3	A, B																		A	calcium channel, voltage-dependent, gamma subunit 3	chr16:p12.1

No	GENE	TOTAL LDL	LARGE LDL	SMALL LDL	LDL SIZE	TOTAL HDL	LARGE HDL	MEDIUM HDL	SMALL HDL	HDL SIZE	TOTAL VLDL	LARGE VLDL	MEDIUM VLDL	SMALL VLDL	VLDL SIZE	IDL	CHYL	TG	HDLC	LDLC	GeneName	Position
19	CBLN4										B, C, E	F						B, C, E			cerebellin 4 precursor	chr20:q13.2
20	CCDC41						E														coiled-coil domain containing 41	chr12:q22
21	CCDC79		D																		coiled-coil domain containing 79	chr16:q22.1
22	CDH4														E						cadherin 4, type 1, R-cadherin (retinal)	chr20:q13.33
23	CDH8																E				cadherin 8, type 2	chr16:q21
24	CDKAL1												B, F								CDK5 regulatory subunit associated protein 1-like 1	chr6:p22.3
25	CETP																		A, D		cholesteryl ester transfer protein, plasma	chr16:q13
26	COL27A1		B																		collagen, type XXVII, alpha 1	chr9:q32
27	CPEB2																		D		cytoplasmic polyadenylation element binding protein 2	chr4:4p15.33
28	CPNE4										F	F									copine IV	chr3:q22.1
29	CSMD1																	F			CUB and Sushi multiple domains 1	chr8:p23.2
30	CRIPT	A, C																			cysteine-rich PDZ-binding protein	chr2:p21
31	DEFB119										F		F					F			defensin, beta 119	chr20:q11.21
32	DEFB123										F		F					F			defensin, beta 123	chr20:q11.21
33	DEFB133											B									defensin, beta 133	chr6:p12.3
34	DISC1		B, C																		disrupted in schizophrenia 1	chr1:q42.2
35	DYNC1LI2		D																		dynein, cytoplasmic 1, light intermediate chain 2	chr16:q22.1
36	EEPD1						A												A		endonuclease/exonuclease/phosphatase family domain containing 1	chr7:p14.2
37	EFCAB4B						E			E, F											EF-hand calcium binding domain 4B	chr12:p13.32
38	EGLN3						C			C											egl nine homolog 3 (C. elegans)	chr14:q13.1
39	EMCN																E				endomucin	chr4:q24
40	FAM167A										F										family with sequence similarity 167, member A	chr8:p23.1
41	FAM63B	B																			family with sequence similarity 63, member B	chr15:q21.3
42	FAM84B	A, B, C												A						A	family with sequence similarity 84, member B	chr8:q24.21
43	FLRT2											F									fibronectin leucine rich transmembrane protein 2	chr14:q31.3
44	FSHR									B, F											follicle stimulating hormone receptor	chr2:p16.3
45	FTMT						A														ferritin mitochondrial	chr5:q23.1
46	GDF10												D								growth differentiation factor 10	chr10:q11.22
47	GDF2												D								growth differentiation factor 2	chr10:q11.22
48	GLI3																E				GLI family zinc finger 3	chr7:p14.1
49	GPR133										C		B, C					C			G protein-coupled receptor 133	chr12:q24.33
50	HAAO																F				3-hydroxyanthranilate 3,4-dioxygenase	chr2:p21
51	HOMER1					A, B, C													A		homer homolog 1 (Drosophila)	chr5:q14.1
52	IQCJ-SCHIP1											D									IQ motif containing J-schwannomin interacting protein 1 fusion protein	chr3
53	KIAA1217								F												KIAA1217	chr10:p12.2
54	KIT					A, D													A, D		v-kit Hardy-Zuckerman 4 feline sarcoma viral oncogene homolog	chr4:q12
55	KRTAP5-6																	A		D	keratin associated protein 5-6	chr11:p15.5
56	LIPC						A														lipase, hepatic	chr15:q21.3
57	LMO7													E			E				LIM domain 7	chr13:q22.2
58	LOC100129636											F									NA	chr6:p22.1

No	GENE	TOTAL LDL	LARGE LDL	SMALL LDL	LDL SIZE	TOTAL HDL	LARGE HDL	MEDIUM HDL	SMALL HDL	HDL SIZE	TOTAL VLDL	LARGE VLDL	MEDIUM VLDL	SMALL VLDL	VLDL SIZE	IDL	CHYL	TG	HDLC	LDLC	GeneName	Position
59	LOC100132713										E										NA	chr18:q23

A–F letters used in Table 1 point to results corresponding to the following times, visit 2 blood draw 1, v2d2, v2d3, v4d1, v4d2, and v4d3. A yellow color at the gene name column annotates findings that overlap as significant proxy genes between NMR-single traits and classical lipids; purple background annotates significant findings after FF and most of the time also after PPL; tan background color annotates mainly baseline discoveries; red color letters annotate mainly NMR-single traits and classical lipids gene discoveries that concord between the two groups in time