A Genome-Wide Association Study For Blood Lipid Phenotypes in the Framingham Heart Study

Kathiresan, Sekar, Alisa K Manning, Serkalem Demissie, Ralph B D'Agostino, Aarti Surti, Candace Guiducci, Lauren Gianniny, Nöel P Burtt, Olle Melander, Marju Orho-Melander, Donna K Arnett, Gina M Peloso, Jose M Ordovas, L Adrienne Cupples.

"A genome-wide association study for blood lipid phenotypes in the Framingham Heart Study" BMC Medical Genetics 8 (Suppl 1): S17. (2007)
https://hdl.handle.net/2144/2562

Boston University
A genome-wide association study for blood lipid phenotypes in the Framingham Heart Study

Sekar Kathiresan1,2, Alisa K Manning1,4, Serkalem Demissie1,4, Ralph B D’Agostino1,5, Aarti Surti2, Candace Guiducci2, Lauren Gianniny2, Nöel P Burtt2, Olle Melander6, Marju Orho-Melander7, Donna K Arnett8, Gina M Peloso1,4, Jose M Ordovas9 and L Adrienne Cupples*1,4

Address: 1National Heart Lung and Blood Institute’s Framingham Heart Study, Framingham, MA, USA, 2Broad Institute of Massachusetts Institute of Technology and Harvard University, Cambridge, MA, USA, 3Cardiovascular Disease Prevention Center, Cardiology Division, Massachusetts General Hospital, Harvard Medical School, Boston, MA, USA, 4School of Public Health, Boston University, Boston, MA, USA, 5Department of Mathematics and Statistics, Boston University, Boston, MA, USA, 6Department of Clinical Sciences, Hypertension and Cardiovascular Diseases, University Hospital Malmö, Lund University, Malmö, Sweden, 7Diabetes and Endocrinology, University Hospital Malmö, Lund University, Malmö, Sweden, 8Department of Epidemiology, University of Alabama at Birmingham, Birmingham, AL, USA and 9Nutrition and Genomics Laboratory, Jean Mayer-USDA Human Nutrition Research Center at Tufts University, Boston, MA, USA

Email: Sekar Kathiresan - skathiresan@partners.org; Alisa K Manning - amanning@bu.du; Serkalem Demissie - demissie@bu.edu; Ralph B D’Agostino - Ralph@bu.edu; Aarti Surti - asurti@broad.mit.edu; Candace Guiducci - candace@broad.mit.edu; Lauren Gianniny - gianniny@broad.mit.edu; Nöel P Burtt - burtt@broad.mit.edu; Olle Melander - olle.melander@med.lu.se; Marju Orho-Melander - marju.orho@med.lu.se; Donna K Arnett - arnett@ms.soph.uab.edu; Gina M Peloso - gpeloso@bu.edu; Jose M Ordovas - jose.ordovas@tufts.edu; L Adrienne Cupples* - adrienne@bu.edu

* Corresponding author

Abstract

Background: Blood lipid levels including low-density lipoprotein cholesterol (LDL-C), high-density lipoprotein cholesterol (HDL-C), and triglycerides (TG) are highly heritable. Genome-wide association is a promising approach to map genetic loci related to these heritable phenotypes.

Methods: In 1087 Framingham Heart Study Offspring cohort participants (mean age 47 years, 52% women), we conducted genome-wide analyses (Affymetrix 100K GeneChip) for fasting blood lipid traits. Total cholesterol, HDL-C, and TG were measured by standard enzymatic methods and LDL-C was calculated using the Friedewald formula. The long-term averages of up to seven measurements of LDL-C, HDL-C, and TG over a ~30 year span were the primary phenotypes. We used generalized estimating equations (GEE), family-based association tests (FBAT) and variance components linkage to investigate the relationships between SNPs (on autosomes, with minor allele frequency ≥10%, genotypic call rate ≥80%, and Hardy-Weinberg equilibrium p ≥ 0.001) and multivariable-adjusted residuals. We pursued a three-stage replication strategy of the GEE association results with 287 SNPs (P < 0.001 in Stage I) tested in Stage II (n ~1450 individuals) and 40 SNPs (P < 0.001 in joint analysis of Stages I and II) tested in Stage III (n~6650 individuals).

Results: Long-term averages of LDL-C, HDL-C, and TG were highly heritable (h² = 0.66, 0.69, 0.58, respectively; each P < 0.0001). Of 70,987 tests for each of the phenotypes, two SNPs had p < 10^-5 in GEE results for LDL-C, four for HDL-C, and one for TG. For each multivariable-adjusted phenotype, the number of SNPs with association p < 10^-4 ranged from 13 to 18 and with p < 10^-3, from 94 to 149. Some results confirmed previously reported associations with candidate genes including variation in the lipoprotein lipase gene (LPL) and HDL-C and TG (rs7007797; P = 0.0005 for HDL-C and 0.002 for TG). The full set of GEE, FBAT and linkage results are posted at the database of Genotype and Phenotype

Published: 19 September 2007

BMC Medical Genetics 2007, 8(Suppl 1):S17 doi:10.1186/1471-2350-8-S1-S17

This article is available from: http://www.biomedcentral.com/1471-2350/8/S1/S17

© 2007 Kathiresan et al; licensee BioMed Central Ltd.

This is an open access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/2.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.
Introduction
Blood lipid levels are a major contributor to atherosclerotic cardiovascular disease [1]. Current evidence suggests that blood lipids are complex genetic phenotypes, influenced by both environmental and genetic factors. Heritability estimates for blood lipids are high, including ~40–60% for high-density lipoprotein cholesterol (HDL-C), ~40–50% for low-density lipoprotein cholesterol (LDL-C), and ~35–48% for triglycerides (TG) [2]. These estimates indicate that DNA sequence variation plays an important role in explaining inter-individual variation in blood lipid levels. Indeed, sequence variants in individual genes have been consistently related to blood lipid phenotypes, including APOE/PCSK9 with LDL-C [3-5], CETP/LIPC/LPL with HDL-C [6-9], and APOA5/LPL with TG [10,11], among others. However, the extent to which common genetic variants across the genome account for total variation in blood lipid levels is unknown.

Recent advances in genomics enable a genome-wide association study (GWAS), an approach in which a substantial fraction of common genetic variation is tested for a role in determining phenotypic variation [12]. These advances include a map of the correlation structure for approximately 4 million common genetic variants (minor allele frequency >5%) and whole-genome genotyping technologies capable of assaying 100,000–500,000 single nucleotide polymorphisms (SNPs) in an individual [13]. Utilizing a fixed genotyping marker set such as the Affymetrix 100K GeneChip in an association study tests a substantial fraction of the genome in whites, ~30–45% in some estimates [14]. GWAS has been successfully applied to identify novel genetic loci related to several medical phenotypes including age-related macular degeneration [15], inflammatory bowel disease [16], and electrocardiographic QT interval [17]. Identifying novel genetic variants related to blood lipid phenotypes may provide new drug targets to alter blood lipid levels and may aid in the prediction of cardiovascular disease.

We hypothesized that common genetic variants explain a proportion of the inter-individual variability in LDL-C, HDL-C, and TG. Accordingly, we conducted genome-wide linkage and association studies for these three phenotypes in Framingham Heart Study (FHS) participants.

Materials and methods
GWAS sample
Of the 1345 FHS participants who are part of the family plate set (see Executive Summary), we focused our analyses on the 1087 participants from the Offspring cohort who had Affymetrix 100K genotypes. Lipid phenotypes were measured at various examinations as described in Table 1. Each study participant provided written informed consent for genetic analyses and the study was approved by Boston University’s Institutional Review Board.

Phenotype definition and methods
Blood lipids were measured from fasting venous blood collected at each of seven clinical examination time points extending from 1971 to 2001. Total cholesterol, HDL-C, and TG were measured by standard enzymatic methods. LDL-C was calculated using the Friedewald formula, with a missing value assigned for participants with a measured TG > 400 mg/dL. Clinical covariates utilized in phenotypic regression modeling included age at the time of blood lipid measurement, age², body mass index (weight in kg divided by the height in m²), alcohol intake (drinks per week), current cigarette smoking (yes, no), menopausal status (postmenopausal yes, no), and hormone replacement therapy (yes, no).

Commonly-used lipid lowering therapies affect total cholesterol and TG. To account for treatment effect, we imputed total cholesterol and TG values for those treated with lipid-lowering therapy. The imputation procedure was modeled after prior work on imputing blood pressure values for those on antihypertensive medication [18]. For each treated individual, a correction factor was added to the observed [treated] lipid value (total cholesterol or TG). This correction factor consisted of the difference between an "expected" residual and the "calculated" residual. The "calculated" residual for each individual was generated in a sex-specific manner after adjustment for age, age², age³, and examination year (by decade). The "expected" residual was generated within each sex and 10 year-age-group as the average of "calculated" residuals.
Table 1: Lipid Phenotypes Examined Using Affymetrix 100K GeneChip Scan

| Phenotype                        | Acronym | Phenotypes | N    | h² * | Offspring Exam | Adjustment† Multivariable model |
|---------------------------------|---------|------------|------|------|----------------|---------------------------------|
| Total cholesterol                | TC      | 7          | 1069 | 0.57†| 1,2,3,4,5,6,7 | Age, age², smoking, body mass index, alcohol consumption, menopausal status, hormone replacement therapy; Covariate-adjusted residuals created separately by gender |
| Low-density lipoprotein cholesterol | LDL-C  | 7          | 1056 | 0.59†| 1,2,3,4,5,6,7 |                   |
| High-density lipoprotein cholesterol | HDL-C  | 7          | 1062 | 0.52†| 1,2,3,4,5,6,7 |                   |
| Triglycerides                    | TG      | 7          | 1068 | 0.48†| 1,2,3,4,5,6,7 |                   |
| Mean total cholesterol           | MeanTC  | 1          | 1087 | 0.61 | Avg of Exams 1,2,3,4,5,6,7 |                   |
| Mean low-density lipoprotein cholesterol | MeanLDL-C  | 1          | 1086 | 0.66 | Avg of Exams 1,2,3,4,5,6,7 |                   |
| Mean high-density lipoprotein cholesterol | MeanHDL-C  | 1          | 1087 | 0.69 | Avg of Exams 1,2,3,4,5,6,7 |                   |
| Mean triglycerides               | MeanTG  | 1          | 1087 | 0.58 | Avg of Exams 1,2,3,4,5,6,7 |                   |
| Apolipoprotein A-I               | ApoA-I  | 1          | 997  | 0.42 | 4               |                   |
| Apolipoprotein B                 | ApoB    | 1          | 997  | 0.47 | 4               |                   |
| Apolipoprotein C3                | ApoC3   | 1          | 767  | 0.38 | 4               |                   |
| Apolipoprotein E                 | ApoE    | 1          | 744  | 0.54 | 5               |                   |
| High-density lipoprotein 2 cholesterol | HDL2-C  | 2          | 955  | 0.50 | 4,5             |                   |
| High-density lipoprotein 3 cholesterol | HDL3-C  | 2          | 984  | 0.48 | 4,5             |                   |
| Large high-density lipoprotein by NMR | HLDNMRlg | 1          | 851  | 0.53 | 4               |                   |
| Intermediate high-density lipoprotein by NMR | HLDNMRint | 1          | 851  | 0.22 | 4               |                   |
| Small high-density lipoprotein by NMR | HLDNMRsm | 1          | 851  | 0.24 | 4               |                   |
| High-density lipoprotein particle size by NMR | HLDNMRsz | 1          | 851  | 0.50 | 4               |                   |
| Intermediate-density lipoprotein particle by NMR | IDLNMR  | 1          | 851  | 0.20 | 4               |                   |
| Large low-density lipoprotein by NMR | LDLNMRlg | 1          | 851  | 0.39 | 4               |                   |
| Small low-density lipoprotein by NMR | LDLNMRsm | 1          | 851  | 0.40 | 4               |                   |
| Low-density lipoprotein particle size by NMR | LDLNMRsz | 1          | 851  | 0.40 | 4               |                   |
| Large very low-density lipoprotein particle by NMR | VLDLNMRlg | 1          | 851  | 0.34 | 4               |                   |
| Intermediate very low-density lipoprotein particle by NMR | VLDLNMRint | 1          | 878  | 0.40 | 4               |                   |
| Small very low-density lipoprotein particle by NMR | VLDLNMRsm | 1          | 851  | 0.24 | 4               |                   |
| Very low-density lipoprotein particle size by NMR | VLDLNMRsz | 1          | 851  | 0.42 | 4               |                   |
| Triglyceride/high-density lipoprotein cholesterol ratio | tghdl | 7          | 1060 | 0.56†| 1,2,3,4,5,6,7 |                   |
| Total cholesterol/high-density lipoprotein cholesterol ratio | cholhdl | 7          | 1060 | 0.58†| 1,2,3,4,5,6,7 |                   |
| Lipoprotein(a)                   | Lp(a)   | 1          | 763  | 0.90 | 3               |                   |
| Remnant lipoprotein cholesterol  | RLP-C   | 1          | 746  | 0.34 | 4               |                   |
| Remnant lipoprotein triglycerides | RLP-TG  | 1          | 715  | 0.33 | 4               |                   |

Note: TG, MeanTG, cholhdl, and tghdl were natural log transformed due to skewed distribution;
*Heritability (h²) estimates presented are those after multivariable-adjustment; P < 0.0001 for all heritability estimates.
†Each phenotype had 2 adjustment schemes: age- and sex-adjusted and multivariable-adjusted. Both age- and sex-adjusted and multivariable-adjusted model results are web posted.
‡Heritability estimates were derived from lipid phenotypes at a single time-point, that of FHS Examination 1.
equal or greater than the treated individual’s "calculated" residual.

Lipoprotein subclass profiles were measured by a commercially available proton NMR spectroscopic assay (LipoScience, Raleigh, NC) on plasma samples stored at -70 °C as described previously [19]. The particle concentration of the following 9 lipoprotein subclasses were determined: 3 VLDL subclasses [large, >60 nm (including chylomicrons); intermediate, 35–60 nm; small, 27–35 nm]; 3 LDL subclasses (IDL, 23–27 nm; large LDL 21.3–23 nm; small LDL, 18.3–21.2 nm); and 3 HDL subclasses (large, 8.8–13 nm; intermediate, 8.2–8.8 nm; small, 7.3–8.2 nm). The small LDL subclass comprises the sum of subclasses formerly labeled "intermediate" (19.8–21.2 nm) and "small" (18.3–19.7 nm) [19], since concentrations of both have very similar relations to lipid levels.

**Genotyping methods**

All analyses were based on the Affymetrix 100K GeneChip genotyping data generated in Framingham Heart Study participants as described previously [20]. In order to minimize false positive associations due to genotyping artifact, we limited our analyses to SNPs with a genotyping call rate ≥80% and a Hardy-Weinberg Equilibrium P ≥ 0.001. Given lower statistical power to detect associations with rarer SNPs, we limited our results to SNPs with a minor allele frequency ≥10%.

**Statistical analysis methods**

TG levels were log-transformed to approximate a normal distribution. For each blood lipid phenotype, the long-term average of 4 to 7 serial measurements was used as the primary phenotype. Participants contributing fewer than 4 of 7 measures of a given phenotype were excluded from that analysis. MeanLDL-C, MeanHDL-C, and MeanTG were adjusted for covariates in sex-specific linear regression models. Two sets of phenotypic models were created: Model 1 (age, age²) and Model 2 (age, age², body mass index, alcohol intake, cigarette smoking, menopausal status, and hormone replacement therapy). For quantitative covariates (age, body mass index, and alcohol intake), the mean value across examinations was used as a covariate. For categorical covariates, the proportion of exams scored as 'yes' was used. The residual MeanLDL-C, MeanHDL-C, and MeanLogTG values from Model 1 and Model 2 served as the primary phenotypes.

For genotype-phenotype association analyses, we assumed an additive model of inheritance. We conducted multivariable linear regression using GEE, family-based association testing using FBAT, and linkage using Merlin for computation of IBDs and SOLAR for variance component models as described in the Executive Summary.

**Heritability analyses**

Heritability estimates for the lipid phenotypes were obtained from extended families with at least two members by variance-components methods using the Sequential Oligogenic Linkage Analysis Routines (SOLAR) package [21]. Using this approach, maximum-likelihood estimation was applied to a mixed-effects model that incorporated fixed covariate effects, additive genetic effects, and residual error. The additive genetic effects and residual errors were assumed to be normally distributed and to be mutually independent. The analyses were performed using residuals from the multivariable models (Model 1 and Model 2) mentioned above. For phenotypes with kurtosis > 1, heritability estimates were computed on ranked normalized deviates.

**Replication samples**

Replication genotyping was attempted in three independent sample sets: a) the FHS unrelated plate set; b) Genetics of Lipid Lowering Drugs and Diet Network (GOLDN); and c) Malmö Diet and Cancer Study – Cardiovascular Cohort (MDC-CC).

The second stage consisted of ~1450 biologically unrelated individuals from the FHS unrelated plate set. The third stage consisted of ~1450 participants from GOLDN and ~5200 participants from MDC-CC. GOLDN is a family-based sample recruited from two National Heart, Lung, and Blood Institute’s Family Heart Study field centers (Minneapolis, MN and Salt Lake City, UT). The Family Heart Study is a multi-center, population-based cohort designed to study the genetic and environmental determinations of cardiovascular disease.

The MDC study is a community-based prospective epidemiologic cohort of 28,098 persons recruited for a baseline examination between 1991 and 1996. From this cohort, 6103 persons were randomly selected to participate in the MDC-CC which sought to investigate risk factors for cardiovascular disease. Of the MDC-CC participants, 5466 had DNA and lipid phenotypes available. Individuals on lipid lowering therapy and with outlier values of LDL-C, HDL-C, or TG (top 0.5% of the distribution) were excluded, leaving 5212 individuals available for the SNP-lipid association analyses.

**Staged replication strategy**

For follow-up into Stage II (the FHS unrelated plate set), we selected all SNPs in the GWAS with an association P < 0.001 for the MeanLDL-C, MeanHDL-C, or MeanTG phenotypes from the minimally-adjusted phenotypic model (Model 1, adjustment for age, age² only). We next conducted a joint analysis of Stage I (GWAS 100K data) and Stage II (FHS unrelated plate set). The joint analysis consisted of a weighted average of the beta estimates and...
standard errors from Stages I and II and used the inverse of the variance in each stage as weights.

For follow-up into Stage III (GOLDN and MDC-CC), we selected for genotyping all SNPs with a P < 0.001 in the joint analysis of Stages I and II. For genotype-phenotype association analyses in MDC-CC and GOLDN, we assumed an additive model of inheritance. In MDC-CC, we conducted multivariable linear regression analyses to test the null hypothesis that LDL-C, HDL-C, or TG residuals (sex-specific residuals adjusted for age and age^2) did not differ by increasing minor allele copy number. In GOLDN, to account for correlated observations due to family relationships we used linear mixed-effects methods in SOLAR.

To summarize the statistical evidence for association for each SNP across all three stages, we reiterated the weighted average beta-estimates and standard errors as described above.

Results

Clinical characteristics of the FHS sample of 1345 subjects are presented in the Executive Summary. Table 1 displays the variables that were studied in our analyses of lipid phenotypes. Further information on these phenotypes can be found at http://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?id=phs000007. Since Original cohort members were non-fasting at examination, our analyses considered only the 1087 Offspring Study participants with fasting lipid measurements and Affymetrix 100K SNP genotypes. For this paper we focus on longitudinal mean levels of serially measured values (minimum of 4, maximum of 7) of LDL-C, HDL-C, and TG (labeled MeanLDL-C, MeanHDL-C, and MeanTG).

Heritability estimates for long-term average lipid phenotypes (Mean LDL-C, MeanHDL-C, and MeanTG) were greater than those from single time-point measurements (Table 1). For example, the heritabilities of MeanLDL-C, MeanHDL-C, and MeanTG were 0.66, 0.69, and 0.58, respectively, whereas heritabilities for LDL-C, HDL-C, and TG measured at FHS Examination 1 (a single time-point) were 0.59, 0.52, and 0.48, respectively. The highest heritability estimate for any available lipid phenotype was that for lipoprotein (a) at 0.90.

From the GEE analyses, the strongest associations for MeanLDL-C, MeanHDL-C, and MeanTG were for SNPs rs287474 (p = 6.3*10^-9), rs524802 (p = 7.6*10^-7), and rs7007075 (p = 7.7*10^-6), respectively (Table 2a). From the FBAT analyses, the strongest associations for MeanLDL-C, MeanHDL-C, and MeanTG were for SNPs rs287474 (p = 1.4*10^-8), rs10495594 (p = 5.1*10^-5), and rs1449866 (p = 1.8*10^-5), respectively (Table 2b). For each multivariable-adjusted phenotype, the number of SNPs with a GEE association p < 10^-4 ranged from 13 to 18 and with p < 10^-3, from 94 to 149. The number of SNPs with FBAT association p < 10^-4 ranged from 2 to 5 and with p < 10^-3, from 74 to 79.

Linkage LOD scores > 2.0 are presented in Table 2c. The best evidence for linkage was a peak LOD score of 3.3 on chromosome 7 for the MeanHDL-C phenotype.

Because the prior probability of any SNP relating to a phenotype is low and given the number of tests, the P value distribution in a GWAS should approach a null distribution. Any strong departure from this expectation might suggest artifacts in genotyping or analysis. For the 70,987 SNPs that passed quality-control filters, the distribution of association P values (generated by the GEE methodology) approached a null distribution but with a slight excess of low P values. For example, for the MeanLDL-C, whereas one would expect 1% of SNPs to demonstrate a P < 0.01 by chance, we found that 1.34% of SNPs displayed a P < 0.01. Similar results were seen for meanHDL-C and meanTG (data not shown).

We evaluated the association results for a SNP and each of a set of four correlated phenotypes – ApoA-I, LDLNMRsm, MeanHDL-C, and MeanTG (Table 3). Several SNPs were associated with P < 0.01 for 3 of the 4 phenotypes.

Among the GEE association results, a SNP (rs7007797) in the lipoprotein lipase (LPL) was associated with MeanHDL-C (p = 0.0005) and MeanTG (p = 0.002) (Table 4). This SNP is a perfect proxy (r^2 = 1) to the previously studied rs328 (also known as S447X) [22]. The minor allele of rs328 has been consistently related to higher HDL-C and lower TG. The direction of effect for SNP rs7007797 in our dataset was consistent with previous observations. Due to a lack of SNPs in the Affymetrix 100K GeneChip correlated with previously reported variants (at r^2 > 0.5 threshold) in the APOE, PCSK9, CETP, LIPC, and APOA5 genes, we were unable to confirm these other previously reported associations (Table 4).

Replication is critical to distinguish true positives from false ones in a GWAS. We pursued a three-stage replication strategy with 287 SNPs (P < 0.001 in Stage I) tested in Stage II (n=1450 individuals) and 40 SNPs (P < 0.001 in joint analysis of Stages I and II) tested in Stage III (n=6650 individuals). Results are displayed in Table 5. After three stages of replication, there was no convincing statistical evidence for association (i.e. joint analysis stages I, II & III P < 10^-5) between any of the tested SNPs and lipid phenotypes.
Table 2: Overview of Top Association and Linkage Results for MeanLDL-C, MeanHDL-C, and MeanTG

2a. Top 25 SNPs for association with MeanLDL-C, MeanHDL-C, or MeanTG based on the lowest p values of the GEE tests

| Phenotype | SNP rs ID* | Chr | Physical location (bp) | GEE P-value | FBAT P-value | gene (IN or NEAR) |
|-----------|-----------|-----|------------------------|-------------|--------------|-------------------|
| MeanLDL-C | rs287474  | 13  | 68173961               | 6.3*10^-9   | 1.4*10^-8    |                   |
| MeanLDL-C | rs287354  | 13  | 68137953               | 3.4*10^-8   | 6.8*10^-8    |                   |
| MeanHDL-C | rs524802  | 19  | 42138787               | 7.6*10^-4   | 0.04         |                   |
| MeanHDL-C | rs505717  | 19  | 42122156               | 1.2*10^-4   | 0.07         | ZNF345            |
| MeanHDL-C | rs544543  | 19  | 42122663               | 2.5*10^-4   | 0.07         |                   |
| MeanTG    | rs700707  | 8   | 13816848               | 7.7*10^-5   | 0.04         |                   |
| MeanHDL-C | rs3734678 | 6   | 107639853              | 8.2*10^-4   | 0.09         | C6orf210          |
| MeanTG    | rs1048882 | 11  | 81678287               | 1.4*10^-2   | 0.03         |                   |
| MeanTG    | rs2142136 | 10  | 89753050               | 1.5*10^-2   | 0.06         | BC005821| PTEN |
| MeanHDL-C | rs1555173 | 6   | 85121177               | 1.7*10^-5   | 0.09         |                   |
| MeanTG    | rs88993   | 9   | 70902713               | 1.8*10^-5   | 0.38         | TRPM3(1)| A550526| AL136545 |
| MeanTG    | rs314474  | 11  | 11482634               | 1.8*10^-5   | 0.09         | AB094146| ISGF4| BC047021 |
| MeanLDL-C | rs10488779| 11  | 79494759               | 3.6*10^-5   | 0.07         |                   |
| MeanHDL-C | rs2415621 | 14  | 4014277               | 3.8*10^-5   | 0.05         |                   |
| MeanHDL-C | rs10508518| 10  | 17081479               | 3.9*10^-5   | 0.09         | CUBN              |
| MeanHDL-C | rs1245058 | 10  | 69760085               | 4.1*10^-5   | 0.04         | LRR6C             |
| MeanHDL-C | rs1048780 | 11  | 79490000               | 4.4*10^-5   | 0.06         |                   |
| MeanHDL-C | rs541326  | 9   | 70936500               | 4.5*10^-5   | 0.39         | TRPM3(1)| A550526| AL136545 |

2b. Top 25 SNPs for association with MeanLDL-C, MeanHDL-C, or MeanTG based on the lowest p values of the FBAT tests

| Phenotype | SNP rs ID* | Chr | Physical location (bp) | GEE P-value | FBAT P-value | gene (IN or NEAR) |
|-----------|-----------|-----|------------------------|-------------|--------------|-------------------|
| MeanLDL-C | rs287474  | 13  | 68173961               | 1.4*10^-4   | 6.3*10^-9    |                   |
| MeanLDL-C | rs287354  | 13  | 68137953               | 6.8*10^-4   | 3.4*10^-8    |                   |
| MeanTG    | rs1449866 | 3   | 14435931               | 1.8*10^-4   | 0.002        | CHST2(1)| AK131346 |
| MeanTG    | rs10506354| 12  | 57759239               | 2.4*10^-5   | 0.007        |                   |
| MeanTG    | rs1049499 | 11  | 21576634               | 4.7*10^-4   | 2.4*10^-5    |                   |
| MeanTG    | rs2371978 | 12  | 57759239               | 4.7*10^-5   | 2.4*10^-5    |                   |
| MeanHDL-C | rs1049559 | 2   | 12280997               | 1.8*10^-4   | 5.1*10^-5    |                   |
| MeanHDL-C | rs1049559 | 2   | 12279945               | 1.8*10^-4   | 5.1*10^-5    |                   |
| MeanHDL-C | rs1411931 | 1   | 37451967               | 1.8*10^-5   | 5.8*10^-6    |                   |
| MeanLDL-C | rs10516434| 4   | 99800704               | 1.6*10^-4   | 7.7*10^-3    |                   |
| MeanLDL-C | rs2182114 | 1   | 111496269              | 2.1*10^-5   | 8.2*10^-2    |                   |
| MeanTG    | rs1835353 | 2   | 124880057              | 1.6*10^-5   | 8.3*10^-3    |                   |
| MeanLDL-C | rs10516434| 4   | 99800704               | 1.6*10^-4   | 8.6*10^-5    |                   |
| MeanLDL-C | rs10507755| 13  | 68355991               | 1.6*10^-4   | 1.1*10^-4    |                   |
| MeanLDL-C | rs6853079 | 4   | 99800789               | 1.6*10^-4   | 1.3*10^-4    |                   |
| MeanLDL-C | rs1508116 | 2   | 41437773               | 1.6*10^-4   | 1.3*10^-4    |                   |
| MeanLDL-C | rs10518072| 4   | 71271418               | 1.6*10^-4   | 1.4*10^-4    |                   |
| MeanLDL-C | rs715260  | 4   | 9804415                | 1.6*10^-4   | 1.5*10^-4    |                   |
| MeanLDL-C | rs7784056 | 7   | 77762038               | 1.6*10^-4   | 1.3*10^-4    |                   |
| MeanHDL-C | rs6500623 | 18  | 1115057                | 1.6*10^-4   | 1.6*10^-4    |                   |
| MeanHDL-C | rs1351296 | 18  | 59173095               | 1.6*10^-4   | 1.6*10^-4    |                   |
| MeanHDL-C | rs6821328 | 4   | 54058427               | 1.6*10^-4   | 1.7*10^-4    |                   |
| MeanTG    | rs1501572 | 9   | 2894794                | 1.6*10^-4   | 1.7*10^-4    |                   |
| MeanTG    | rs4684343 | 3   | 2366098                | 1.6*10^-4   | 1.8*10^-4    |                   |
| MeanTG    | rs225634  | 6   | 14251832               | 1.6*10^-4   | 1.9*10^-4    |                   |

2c. Magnitude and Location of Peak LOD scores > 2.0 for MeanLDL-C, MeanHDL-C, and MeanTG

| Phenotype | Exam | Chr | Physical location (Mb) | Maximum LOD | LOD-1.5 Interval | LOD+1.5 Interval |
|-----------|------|-----|------------------------|-------------|------------------|-----------------|
| MeanHDL-C | 1–7  | 7   | 33485983               | 3.30        | 27810820         | 36074331        |
| MeanLDL-C | 1–7  | 9   | 94130181               | 2.83        | 88057135         | 98516033        |
| MeanTG    | 1–7  | 1   | 153448398              | 2.73        | 151582274        | 155505440       |
| MeanLDL-C | 1–7  | 3   | 196389998              | 2.12        | 187706181        | 199138789       |

Chr = chromosome
Discussion
We examined associations of Affymetrix 100K SNPs and lipid traits in FHS and identified putative associations with lipid phenotypes. We studied the long-term average of up to 7 measurements each of LDL-C, HDL-C, and TG as the primary phenotypes and for one phenotype, the MeanLDL-C, we observed a nominal P that exceeded genome-wide significance [13]. However, validation of selected hypotheses in additional samples did not identify any new loci underlying variability in blood lipids.

GWAS offers the potential to identify novel genetic variants/loci that are associated with blood lipid variation, unlimited by our current knowledge of lipoprotein biology. However, a central limitation of GWAS is that the true signals are mixed amidst a large number of false positive results. Validation in additional samples is required to distinguish the true positives from the false ones.

Replication of initial GWAS findings using a staged design has been suggested to minimize genotyping cost and maximize statistical power [23,24]. An important consideration in such a design is the proportion of markers taken forward to a second stage. We estimated the statistical power for our three-stage GWAS strategy. Assuming a modest number of markers (all SNPs with P < 0.001 for each phenotype, ~0.1% of markers) are taken forward to Stage II, a second stage sample size of 1450, that SNPs with P < 0.001 are taken forward from Stage II to Stage III, a stage III sample size of 6650, and that the final alpha (after Stages I, II, & III) is set at a conservative 5*10^{-8}, we estimated that we had 89% power to detect a quantitative

Table 4: Comparison with the prior literature

| Gene Symbol | Phenotype | Selected SNPs previously associated with phenotype | # SNPs in Affymetrix 100K within 60 kb of gene | # SNPs in Affymetrix 100K within 60 kb of Gene Locus and with r² > 0.5 to previously reported SNP | SNP in Affymetrix 100K (r² to previously reported SNP) | p for association in FHS 100K |
|-------------|-----------|-----------------------------------------------|---------------------------------------------|------------------------------------------------------------------------------------------------|-----------------------------------------------|-----------------------------|
| APOE        | LDL-C     | rs429358, rs7412                               | 1                                           | 0                                                                                              | -                                             | -                           |
| PCSK9       | LDL-C     | rs11591147                                     | 7                                           | 0                                                                                              | -                                             | -                           |
| CETP        | HDL-C     | rs1800775                                      | 2                                           | 0                                                                                              | -                                             | -                           |
| LIPC        | HDL-C     | rs1800588                                      | 18                                          | 0                                                                                              | -                                             | -                           |
| LPL         | HDL-C     | rs328                                          | 8                                           | 2                                                                                              | rs10503669 (r² = 1.0)                         | rs10503669 – 9.1*10^{-4}   |
|             | TG        | rs328                                          | 8                                           | 2                                                                                              | rs10503669 (r² = 1.0)                         | rs10503669 – 0.02          |
| APOA5       | TG        | rs662799, rs3135506                            | 4                                           | 0                                                                                              | -                                             | -                           |
Table 5: Association Results for 40 SNPs Attempted for Replication in Three Stages

| SNP      | Chr | Gene     | Allele* | Trait     | Stage I – FHS 100K | Stage II – FHS Unrelated | Joint P | Stage III – MDC | Stage III – GOLDN | Joint Stages I, II, & III |
|----------|-----|----------|---------|-----------|---------------------|--------------------------|--------|-----------------|-------------------|--------------------------|
| rs7231460| 18  | MeanHDL-C| C       | MeanHDL-C | 4.3*10^{-4}        | -0.145                    | 0.041  | 2.4*10^{-4}     | -0.131            | 0.036                    | 3.7*10^{-4} | 0.50 | 0.013 | 0.020 | 0.60 | 0.023 | 0.044 | 0.03 | -0.033 | 0.015 |
| rs966376 | 18  | MeanHDL-C| C       | MeanHDL-C | 2.5*10^{-4}        | -0.148                    | 0.040  | 4.8*10^{-4}     | -0.125            | 0.036                    | 4.5*10^{-4} | 0.43 | 0.016 | 0.020 | 0.56 | 0.026 | 0.044 | 0.04 | -0.031 | 0.015 |
| rs744134 | 13  | L10374   | C       | MeanHDL-C | 5.4*10^{-4}        | 0.168                     | 0.048  | 4.7*10^{-4}     | -0.124            | 0.036                    | 1.1*10^{-4} | 0.67 | 0.009 | 0.020 | 0.20 | 0.062 | 0.048 | 0.007 | 0.053 | 0.016 |
| rs7233386| 18  | MeanHDL-C| T       | MeanHDL-C | 2.4*10^{-4}        | 0.149                     | 0.041  | 0.005           | 0.100            | 0.036                    | 6.3*10^{-4} | 0.32 | -0.020 | 0.020 | 0.61 | 0.023 | 0.046 | 0.11 | 0.024 | 0.015 |
| rs1428445| 5   | A        | MeanHDL-C| C       | 3.0*10^{-4}        | -0.187                    | 0.052  | 0.01           | -0.113            | 0.044                    | 1.8*10^{-4} | 0.24 | -0.028 | 0.024 | 0.77 | 0.016 | 0.054 | 0.002 | -0.058 | 0.018 |
| rs2304589| 2   | LOC130502 | A       | MeanHDL-C | 6.4*10^{-4}        | 0.187                     | 0.055  | 0.006           | 0.124            | 0.045                    | 1.8*10^{-4} | 0.47 | -0.020 | 0.027 | 0.09 | -0.101 | 0.060 | 0.18 | 0.027 | 0.020 |
| rs2278528| 18  | LOC130502 | G       | MeanHDL-C | 6.7*10^{-4}        | 0.180                     | 0.053  | 0.007           | 0.122            | 0.045                    | 2.0*10^{-4} | 0.51 | -0.018 | 0.027 | 0.12 | -0.092 | 0.060 | 0.14 | 0.029 | 0.020 |
| rs2142136| 10  | PTEN     | G       | MeanTG    | 8.1*10^{-4}        | 0.228                     | 0.051  | 0.14           | 0.077            | 0.052                    | 2.4*10^{-4} | 0.38 | -0.026 | 0.029 | 0.85 | 0.011 | 0.061 | 0.06 | 0.040 | 0.021 |
| rs1555173| 6   | C        | MeanLDL-C| T       | 1.0*10^{-4}        | -0.243                    | 0.055  | 0.10           | -0.083            | 0.050                    | 1.1*10^{-6} | 0.67 | 0.009 | 0.020 | 0.20 | 0.062 | 0.048 | 0.007 | 0.053 | 0.016 |
| rs6053754| 2   | LOC130502 | A       | MeanLDL-C | 6.4*10^{-4}        | 0.171                     | 0.048  | 0.08           | 0.061            | 0.034                    | 5.5*10^{-4} | 0.09 | 0.033 | 0.020 | 0.60 | 0.026 | 0.046 | 0.02 | 0.044 | 0.016 |
| rs2278528| 18  | LOC130502 | G       | MeanLDL-C | 6.7*10^{-4}        | 0.180                     | 0.053  | 0.007           | 0.122            | 0.045                    | 2.0*10^{-4} | 0.51 | -0.018 | 0.027 | 0.12 | -0.092 | 0.060 | 0.14 | 0.029 | 0.020 |

*The allele on the positive strand of the reference genome was modeled in all analyses.

†Beta refers to the proportion of 1 standard deviation unit change in phenotype (phenotype is sex-specific residual adjusted for age and age2) per copy of the allele modeled.

‡"Failed" refers to SNP genotype failure in the sample.
trait locus explaining 2% of phenotypic variance, 48% power to detect a locus explaining 1% of the variance, and 13% power to detect a locus explaining 0.5% of the variance.

With our replication effort, we failed to identify any novel loci related to blood lipids. At least two potential explanations are possible. First, our study design had limited statistical power to detect common SNPs that explain ≤1% of trait variance. In the Diabetes Genetics Initiative genome-wide association study for blood lipid traits, we recently showed that for lipid traits, there are few common variants that explain >2% of the variance and most SNPs explain <1% of trait variance [25]. To have adequate statistical power to detect these effects given an initial GWAS sample size of ~1000, many more markers (i.e., hundreds of SNPs) will need to be taken to the second and third stages. Second, the limited genomic coverage of the Affymetrix 100K array may have limited our ability to replicate previously reported loci and discover novel loci. For example, using the Affymetrix 500 K array, we recently identified glucokinase regulatory protein (GCKR) as a novel locus associated with TG [25]. Of any SNP on the 500 K array, an intronic GCKR SNP (rs7800994) explained the greatest proportion of blood TG variance in the Diabetes Genetics Initiative study. However, on the Affymetrix 100K array, there are no SNPs within the 60 kb spanning GCKR.

Strengths and limitations
This study is distinguished by the availability of serial lipid phenotypes over a 30-year time span, the community-based nature of the collection, and the routine ascertainment of covariates in a standardized clinical examination. We acknowledge several limitations. These include the lack of validation for the imputation methodology used to address lipid lowering therapy, limited statistical power due to sample size, and confinement to a single ancestral group – whites of European ancestry.

Conclusions & future directions
Using a 100K genome-wide scan, we present association and linkage results for a rich set of lipid phenotypes in FHS. This resource may be useful for comparisons with other GWAS currently in progress. GWAS in FHS using a denser genome-wide genotyping platform and a better-powered replication strategy may identify novel loci underlying blood lipids.

Abbreviations
FBAT = family-based association test; GEE = generalized estimating equations.

Competing interests
The authors declare that they have no competing interests.

Authors’ contributions
SK, AM, SD, GP, JMO, and LAC participated in the design of the study and the interpretation of the data. AM, GP, and SD conducted the statistical analyses. SK drafted the manuscript. AS, CG, LG, and NPB generated replication genotype data and analyses. OM and MOM provided replication samples and conducted association analyses in the Malmo Diet and Cancer Study. DKA and JMO provided replication samples and led the generation of lipid phenotypes in the GOLDN study. SD, SK, RD, JMO, and LAC revised the manuscript critically for important intellectual content. All authors read and approved the above manuscript.

Acknowledgements
We thank the Framingham Heart Study participants for their long-term voluntary commitment to this study. The Framingham Heart Study is supported by a contract from the National Heart, Lung and Blood Institute (contract No. N01-HC-25195). We acknowledge Dr. Michael Christman, Dr. Alan Herbert and colleagues at Boston University who conducted the Affymetrix 100K genotyping and have made these data publicly available. A portion of the research was conducted using the Boston University Linux Cluster for Genetic Analysis (LinGA) funded by the NIH NCCR (National Center for Research Resources) Shared Instrumentation grant (S10RR16376-01A1). Dr. Ordovas is supported by contracts 53-K06-5-10 and 58-1950-9-001 from the US Department of Agriculture Research Service. Dr. Kathiresan is funded by the Doris Duke Charitable Foundation Clinical Scientist Development Award, the Fannie E. Rippel Foundation, and NIH K23 HL083102.

This article has been published as part of BMC Medical Genetics Volume 8 Supplement 1, 2007: The Framingham Heart Study 100,000 single nucleotide polymorphisms resource. The full contents of the supplement are available online at http://www.biomedcentral.com/1471-2350/8?issue=S1.

References
1. Kannel WB, Dawber TR, Kagan A, Revotskie N, Stokes J 3rd: Factors of risk in the development of coronary heart disease – six year follow-up experience. The Framingham Study. Annals of internal medicine 1961, 55:23-50.
2. Weiss LA, Pan L, Abney M, Ober C: The sex-specific genetic architecture of quantitative traits in humans. Nature genetics 2006, 38(2):218-222.
3. Cohen J, Pershagen H, Karlsson H, Weman O, Milunsky A, Nejfelt L, Karasov H, Reiner P, Lyytinen H, Vaheri A, et al: The Framingham Heart Study: genetic variation and the risk of cardiovascular disease. Journal of the American Medical Association 2005, 293(7):820-827.
4. Ordovas JM, Liwacki-Klein L, Wilson PW, Schaefer MM, Schaefer EJ: Apolipoprotein E isoform phenotyping methodology and population frequency with identification of apoE1 and apoE5 isoforms. Journal of lipid research 1998, 39(2):268-285.
5. Boekholdt SM, Sacks FM, Jukema JW, Shepherd J, Freeman DJ, McManus AD, Cambien F, Nicaud V, de Grooth GJ, Talmud PJ, et al: Cholesterol ester transfer protein TaqIB variant, high-density lipoprotein cholesterol levels, cardiovascular risk, and efficacy of pravastatin treatment: individual patient meta-analysis of 13,677 subjects. Circulation 2005, 111(3):278-287.
6. Guerra R, Wang J, Grundy SM, Cohen JC: A hepatic lipase (LIIPC) allele associated with high plasma concentrations of high density lipoprotein cholesterol. Proceedings of the National Academy of Sciences of the United States of America 1995, 92(9):4532-4537.
7. Ordovas JM, Cupples LA, Corella D, Otvos JD, Osgood D, Martinez A, Lahoz C, Coltell O, Wilson PW, Schaefer EJ: Association of
cholesteryl ester transfer protein-TaqI B polymorphism with variations in lipoprotein subclasses and coronary heart disease risk: the Framingham study. Arteriosclerosis, thrombosis, and vascular biology 2000, 20(5):1323-1329.

9. Wittrup HH, Tybjærg-Hansen A, Nordestgaard BG: Lipoprotein lipase mutations, plasma lipids and lipoproteins, and risk of ischemic heart disease. A meta-analysis. Circulation 1999, 99(22):2901-2907.

10. Lai CQ, Demissie S, Cupples LA, Zhu Y, Adiconis X, Parnell LD, Corella D, Ordovalas JM: Influence of the APOA5 locus on plasma triglyceride, lipoprotein subclasses, and CVD risk in the Framingham Heart Study. Journal of lipid research 2004, 45(11):2096-2105.

11. Pennacchio LA, Olivier M, Hubacek JA, Krauss RM, Rubin EM, Cohen JC: Two independent apolipoprotein A5 haplotypes influence human plasma triglyceride levels. Human molecular genetics 2002, 11(24):3031-3038.

12. Hirschhorn JN, Daly MJ: Genome-wide association studies for common diseases and complex traits. Nature reviews 2005, 6(2):95-108.

13. A haplotype map of the human genome. Nature 2005, 437(7063):1299-1320.

14. Pe'er I, de Bakker PJ, Maller J, Daly MJ: Evaluating and improving power in whole-genome association studies using fixed marker sets. Nature genetics 2006, 38(6):663-667.

15. Klein RJ, Zeiss C, Chew EY, Tsai JY, Sackler RS, Haynes C, Henning AK, SanGiovanni JP, Mane SM, Mayne ST, et al.: Complement factor H polymorphism in age-related macular degeneration. Science (New York, NY) 2005, 308(5720):385-389.

16. Duerr RH, Taylor KD, Brant SR, Riou JJ, Kathiresan S, Hirschhorn JN, Daly MJ, Steinhardt AH, Abraham C, Regueiro M, Griffiths A, et al.: A genome-wide association study identifies IL23R as an inflammatory bowel disease gene. Science (New York, NY) 2006, 314(5804):1461-1463.

17. Arking DE, Pfeufer A, Post W, Kao WH, Newton-Cheh C, Ikeda M, West K, Kashuk C, Akylol M, Perz S, et al.: A common genetic variant in the NOS1 regulator NOS1AP modulates cardiac repolarization. Nature genetics 2006, 38(6):644-651.

18. Levy D, De Stefano AL, Larson MG, O'Donnell CJ, Lifton RP, Gavras H, Cupples LA, Myers RH: Evidence for a gene influencing blood pressure on chromosome 17. Genome scan linkage results for longitudinal blood pressure phenotypes in subjects from the Framingham heart study. Hypertension 2000, 36(4):477-483.

19. Freedman DS, Otos J, 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. Clinical chemistry 2004, 50(7):1189-1200.

20. Herbert A, Gerry NP, McQueen MB, Heid IM, Pfeufer A, Illig T, Wichmann HE, Pleitinger T, Hunter D, Hu FB, et al.: A common genetic variant is associated with adult and childhood obesity. Science (New York, NY) 2006, 312(5771):279-283.

21. Almasy L, Blangero J: Multipoint quantitative-trait linkage analysis in general pedigrees. American journal of human genetics 1999, 65(2):1198-1211.

22. Rip J, Nierman MC, Ross CJ, Juema JW, Hayden MR, Kastelein JJ, Stroes ES, Kuivenhoven JA: Lipoprotein lipase 5447T: a naturally occurring gain-of-function mutation. Arteriosclerosis, thrombosis, and vascular biology 2006, 26(6):1236-1245.

23. Satagopan JM, Verbel DA, Venkataraman ES, O'ferr KE, Begg CB: Two-stage designs for gene-disease association studies. Biometrics 2002, 58(1):163-170.

24. Skol AD, Scott LJ, Abecasis GR, Boehnke M: Joint analysis is more efficient than replication-based analysis for two-stage genome-wide association studies. Nature genetics 2006, 38(2):209-213.

25. Saxena R, Voight BF, Lyssenko V, Burtt NP, de Bakker PI, Chen H, Roix JJ, Kathiresan S, Hirschhorn JN, Daly MJ, et al.: Genome-wide association analysis identifies loci for type 2 diabetes and triglyceride levels. Science (New York, NY) 2007, 316(5829):1331-1336.

Publish with BioMed Central and every scientist can read your work free of charge

"BioMed Central will be the most significant development for disseminating the results of biomedical research in our lifetime."

Sir Paul Nurse, Cancer Research UK

Your research papers will be:
- available free of charge to the entire biomedical community
- peer reviewed and published immediately upon acceptance
- cited in PubMed and archived on PubMed Central
- yours — you keep the copyright

Submit your manuscript here:
http://www.biomedcentral.com/info/publishing_adv.asp