

ARIC Manuscript Proposal # 1482

PC Reviewed: 03/17/09
SC Reviewed: _____

Status: A
Status: _____

Priority: 2
Priority: _____

1.a. Full Title: Relation of a lipid gene score to longitudinal trends in lipid levels, and to statin therapy response in Caucasians: The ARIC study.

b. Abbreviated Title (Length 26 characters): Lipid gene score and lipid trends

2. Writing Group: Pamela L. Lutsey, Weihong Tang, Aaron R. Folsom, James S. Pankow, Alvaro Alonso, Kelly A. Volcik, Christie M. Ballantyne, Eric Boerwinkle

I, the first author, confirm that all the coauthors have given their approval for this manuscript proposal. PLL

First author: Pamela L. Lutsey, PhD, MPH
Address: 1300 South 2nd Street, Suite 300
Minneapolis, MN 55454

Phone: (612) 624-5812 Fax: (612) 624-0315
E-mail: lutsey@umn.edu

ARIC author to be contacted if there are questions about the manuscript and the first author does not respond or cannot be located (this must be an ARIC investigator).

Name: Aaron R. Folsom, MD, MPH
Address: 1300 South 2nd Street, Suite 300
Minneapolis, MN 55454

Phone: (612) 626-8862 Fax: (612) 624-0315
E-mail: folso001@umn.edu

3. Timeline: Literature review – 2 months
Data analysis – 3 months
Writing the manuscript – 5 months
Coauthor review and revisions – 3 months

4. Rationale:

Seven recent genome-wide association studies (GWASs) in Caucasians have successfully localized common DNA sequence variants that are associated with blood lipoprotein levels¹⁻⁷. Given that individual loci only have modest effects, for three of the GWASs the authors computed summary lipid gene scores, to assess whether the cumulative allelic dosage of risk alleles at these loci contributed to quantitative variation in lipoprotein levels seen in their respective populations. The genotype score by Sabatti *et al* explained 6% of the variance in LDL levels, 6% of the variance in HDL, and 4.2% of the variance in triglycerides². The variance explained by Aulchenko *et al*'s genotype score was 3.9% for total cholesterol, 3.4% for LDL, 4.8% for HDL, and 3.0% for triglycerides³. Kathiresan *et al* did not report the proportion of variance explained, but did demonstrate a linear association between genotype score and the proportion of individuals exceeding clinical cut-points for high LDL, low LDL, and high triglycerides¹. Aulchenko *et al* found their total cholesterol genotype score was positively associated with increased intima-media thickness (IMT), even after adjustment for circulating levels of total cholesterol³. A positive relation was also observed with incident coronary heart disease (CHD), though it did not remain significant after further adjustment for circulating total cholesterol levels³.

Etiologically, it is of interest to know the timeframe wherein the effects of these genes are manifested. Presently, it is unclear whether the genotypic effects are stable, or if they vary across the lifespan. Though largely unexplored, there is some evidence suggesting that the effects of some SNPs may vary across the life-course^{3, 8, 9}. It has been hypothesized that the differential effects may be due to age-dependent gene expression, changes in the penetrance of underlying genes, and/or gene-environment interactions related to the effect of cumulative exposure to certain environmental factors⁹.

Further, it is unknown whether sex interactions exist in the effect of lipid genotype scores on circulating lipid levels. Lipid levels are known to vary by sex¹⁰, and sex-differences in the effects of some alleles have been noted^{2, 3, 11, 12}.

Likewise, response to statins, which according to NHANES 1999-2004 data are taken by 24.4% of older (men ≥ 50 years and women ≥ 60 years) Americans¹³, may differ according to lipid gene score. It has been hypothesized that genetic factors may explain some of the large interindividual variation in statin therapy response^{14, 15}. Research on this topic is mixed, and thus far has centered on individual SNPs. A recent review of this literature suggested that, as statin response results from complex interactions among various biologic pathways, future studies should consider "a set of candidate genes and/or a genome wide scan, rather than addressing a single or small number of SNPs."¹⁵

5. Main Hypothesis/Study Questions:

Using a lipid gene score created from GWAS-identified genes, we will first evaluate whether overall lipid values across visits differ according to lipid gene score in ARIC Caucasians. We will also assess total change between baseline and exam 4, by lipid gene score. Then, we will investigate age- and sex-interactions in longitudinal trends in lipid levels across the four ARIC exams. Lastly, we will explore whether statin therapy response differs according to gene score.

Our hypotheses are as follows:

- 1) Over 9 years of follow-up, lipid trajectories will reveal higher overall lipid levels among participants with more unfavorable lipid gene scores.
- 2) Change in lipid levels between baseline and exam 4 will differ by lipid gene score.
- 3) The relation of lipid gene scores to lipid trajectories may differ by time and sex.
- 4) Statin therapy response may also differ according to lipid gene score.

6. Design and analysis (study design, inclusion/exclusion, outcome and other variables of interest with specific reference to the time of their collection, summary of data analysis, and any anticipated methodologic limitations or challenges if present).

Study Design

Hypotheses 1 and 3 will assess lipid level trajectories using data from exams 1 through 4, as well as genetic data. Hypothesis 2 will use data from baseline and exam 4.

Hypothesis 4 will also use data from all exams. However, these analyses will be restricted to participants who used statins. Only data from the visit they first reported statin use, and that immediately preceding, will be used to test this hypothesis.

Inclusions/Exclusion

Only ARIC Caucasians will be included. Participants who did not consent for genetic analyses will be excluded, as will those who attended only the baseline exam. Further, for LDL and triglyceride analyses datapoints will be excluded in instances when the participant did not fast for 8 hours prior to venipuncture. Hypothesis 4 analyses will further exclude participants on statins at baseline.

Variables

Outcome variables: Trends in levels of total cholesterol, HDL cholesterol, LDL cholesterol, and triglycerides. Also, total change in lipid levels from baseline to exam 4.

As recommended by Tobin, *et al*, we will add a constant to lipid values of participants on lipid lowering medications¹⁶. The choice of constant will be based on existing trial data, and sensitivity analyses will be used to explore the robustness of our findings assuming different reasonable constants.

Exposure variables: Loci identified in the seven lipid GWASs will be utilized to create gene scores. Depending on availability of the data, we may also incorporate new ARIC lipid genes identified in CHARGE. Of the 104 different SNPs identified, 55 were measured in ARIC, 45 have been imputed, and 4 are unavailable. Linkage disequilibrium will be evaluated using HaploView, and only one of SNPs correlated at $r^2 > 0.9$ will contribute to the genotype score.

Lipid genotype scores will be computed as the sum of:

- 1) Measured SNPs: Each unfavorable allele will be counted as +1, and
- 2) Imputed SNPs: The probability (dosage) of the SNP will be multiplied by the number of unfavorable alleles.

All SNPs with ARIC information will be used to compute the total cholesterol lipid gene score. Genotype scores for separate lipoprotein classes will be calculated using the SNPs identified for each class. The genotype scores will be categorized into quintiles for analyses.

Potential interacting factors: Age, sex, statin therapy response.

Potential confounding factors: Center, education, income, smoking status (former, current, never), pack-years (continuous), physical activity (Braecke score), dietary indicators (Key's score, intake of whole grains, alcohol, and total calories), BMI, diabetes, and hormone replacement therapy use.

Data Analysis Summary

Aim 1 – Main effects of genotype score on overall lipid levels

Associations of genotype scores and overall lipid levels across the ARIC exams will be assessed using repeated measures regression implemented in PROC MIXED, which takes into account inherent within-person correlation in lipid values across exams. The dependent variable will be lipid levels across visits, and adjustments will be made for age, sex, center, education, and income.

For this aim, and all that follow, we will also explore a second, behavioral- and physiologic-adjusted model, which will include the following covariates: smoking status, pack-years, physical activity, dietary indicators, BMI, diabetes, and hormone replacement therapy use. Given that this is a genetic analysis, however, it is unlikely that covariates will confound the associations of interest, and if so, only reduced models will be presented in the final paper.

Aim 2 – Association between genotype score and total change in lipid levels

General linear regression will be used to assess the relation of genotype score to change in lipid levels between baseline and exam 4. Change in lipid levels will be the dependent variable. Independent variables will include genotype score quintile, age, sex, center, education, and income.

Aim 3 – Interactions in lipid trajectories by genotype score

These analyses will be similar to those of Aim 1, except that, gene score-by-time, and gene score-by-sex interactions will be considered. Quadratic terms will also be explored.

Aim 4 – Statin therapy response by genotype score

This analysis will be restricted to statin users. Linear regression will be used to evaluate genotype score differences in statin response. Lipid values at the exam where statin use was first reported will be the dependent variables. Independent variables will include genotype score, lipid values at the exam immediately preceding the exam at which statin use was first reported, age, sex, center, education, and income.

Sensitivity Analyses

All analyses will be re-run, using genotype scores created from only the genotyped SNPs (omitting imputed SNPs).

7.a. Will the data be used for non-CVD analysis in this manuscript? ___ Yes
 X No

b. If Yes, is the author aware that the file ICTDER03 must be used to exclude persons with a value RES_OTH = “CVD Research” for non-DNA analysis, and for DNA analysis RES_DNA = “CVD Research” would be used? ___

Yes ___ No

(This file ICTDER03 has been distributed to ARIC PIs, and contains the responses to consent updates related to stored sample use for research.)

8.a. Will the DNA data be used in this manuscript? X Yes ___ No

8.b. If yes, is the author aware that either DNA data distributed by the Coordinating Center must be used, or the file ICTDER03 must be used to exclude those with value RES_DNA = “No use/storage DNA”?

 X Yes ___ No

8.c. If yes, is the author aware that the participants with RES_DNA = ‘not for profit’ restriction must be excluded if the data are used by a for profit group?

 X Yes ___ No

9. The lead author of this manuscript proposal has reviewed the list of existing ARIC Study manuscript proposals and has found no overlap between this proposal and previously approved manuscript proposals either published or still in active status.

ARIC Investigators have access to the publications lists under the Study Members Area of the web site at: <http://www.csc.unc.edu/ARIC/search.php>

 X Yes ___ No

10. What are the most related manuscript proposals in ARIC (authors are encouraged to contact lead authors of these proposals for comments on the new proposal or collaboration)?

There are no ARIC proposals using lipid gene scores. Further, only one proposal (#1306) exists which assesses the longitudinal association between a SNP and lipid levels. Specifically, #1306 explores the relation of ANGPTL4 to triglycerides. Thus, there is no overlap with existing ARIC proposals, or to our knowledge, any CHARGE or CARE working group.

11. a. Is this manuscript proposal associated with any ARIC ancillary studies or use any ancillary study data? X Yes ___ No

SNPS data will come from the ARIC Caucasian GWAS. Ancillary: 2006.03

11.b. If yes, is the proposal

- ___ **A. primarily the result of an ancillary study (list number* _____)**
___ **B. primarily based on ARIC data with ancillary data playing a minor
role (usually control variables; list number(s)* _____
_____)**

Neither of the above criteria fit. The proposal is not a primary result of the ancillary, but the data we will use from the ancillary are essential to this proposal.
Ancillary #2006.03

*ancillary studies are listed by number at <http://www.csc.unc.edu/aric/forms/>

12. Manuscript preparation is expected to be completed in one to three years. If a manuscript is not submitted for ARIC review at the end of the 3-years from the date of the approval, the manuscript proposal will expire.

References

1. Kathiresan S, Willer CJ, Peloso GM, et al. Common variants at 30 loci contribute to polygenic dyslipidemia. *Nat Genet* 2009;41:56-65.
2. Sabatti C, Service SK, Hartikainen AL, et al. Genome-wide association analysis of metabolic traits in a birth cohort from a founder population. *Nat Genet* 2009;41:35-46.
3. Aulchenko YS, Ripatti S, Lindqvist I, et al. Loci influencing lipid levels and coronary heart disease risk in 16 European population cohorts. *Nat Genet* 2009;41:47-55.
4. Kathiresan S, Melander O, Guiducci C, et al. Six new loci associated with blood low-density lipoprotein cholesterol, high-density lipoprotein cholesterol or triglycerides in humans. *Nat Genet* 2008;40:189-97.
5. Kooner JS, Chambers JC, Aguilar-Salinas CA, et al. Genome-wide scan identifies variation in MLXIPL associated with plasma triglycerides. *Nat Genet* 2008;40:149-51.
6. Wallace C, Newhouse SJ, Braund P, et al. Genome-wide association study identifies genes for biomarkers of cardiovascular disease: serum urate and dyslipidemia. *Am J Hum Genet* 2008;82:139-49.
7. Willer CJ, Sanna S, Jackson AU, et al. Newly identified loci that influence lipid concentrations and risk of coronary artery disease. *Nat Genet* 2008;40:161-9.
8. Webster RJ, Warrington NM, Weedon MN, et al. The association of common genetic variants in the APOA5, LPL and GCK genes with longitudinal changes in metabolic and cardiovascular traits. *Diabetologia* 2009;52:106-14.
9. Chen W, Srinivasan SR, Elkasabany A, Ellsworth DL, Boerwinkle E, Berenson GS. Influence of lipoprotein lipase serine 447 stop polymorphism on tracking of triglycerides and HDL cholesterol from childhood to adulthood and familial risk of coronary artery disease: the Bogalusa heart study. *Atherosclerosis* 2001;159:367-73.
10. Ghandehari H, Kamal-Bahl S, Wong ND. Prevalence and extent of dyslipidemia and recommended lipid levels in US adults with and without cardiovascular comorbidities: the National Health and Nutrition Examination Survey 2003-2004. *Am Heart J* 2008;156:112-9.
11. Wu FC, von Eckardstein A. Androgens and coronary artery disease. *Endocr Rev* 2003;24:183-217.
12. Weiss LA, Pan L, Abney M, Ober C. The sex-specific genetic architecture of quantitative traits in humans. *Nat Genet* 2006;38:218-22.
13. Spatz EC, ME, Desai, MM. From here to Jupiter. Identifying new patients for statin therapy using data from the 1999-2004 National Health and Nutrition Examination Survey. *Circ Cardiovasc Qual Outcomes* 2009:41-8.
14. Chasman DI, Posada D, Subrahmanyam L, Cook NR, Stanton VP, Jr., Ridker PM. Pharmacogenetic study of statin therapy and cholesterol reduction. *JAMA* 2004;291:2821-7.
15. Peters B, Maitland-van der Zee AH. Pharmacogenomic importance of pravastatin. *Pharmacogenomics* 2008;9:1207-10.
16. Tobin MD, Sheehan NA, Scurrah KJ, Burton PR. Adjusting for treatment effects in studies of quantitative traits: antihypertensive therapy and systolic blood pressure. *Stat Med* 2005;24:2911-35.