ARIC Manuscript Proposal # 1526

PC Reviewed: 6/2/09	**Status: A**	**Priority: 2**
SC Reviewed: _____	**Status: _____**	**Priority: ____**

1a. Full Title: Genotype-by-Statin Interactions & Lipids: the CHARGE Drug-Gene GWAS Consortium

b. Abbreviated Title: CHARGE Drug-Gene GWAS of Lipids

2. Writing Group: Eric A. Whitsel, Christy L. Avery, Til Stürmer, Christie Ballantyne (and attempting to maintain symmetry across contributing cohorts), other members of the CHARGE Drug-Gene GWAS Consortium, as well as interested members of the ARIC Lipid Phenotype Working Group (all of whom have been contacted and invited to participate).

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

Eric A. Whitsel
University of North Carolina at Chapel Hill
Departments of Epidemiology and Medicine
Cardiovascular Disease Program
Bank of America Center, Suite 306-E
137 East Franklin Street
Chapel Hill, NC 27514
(T) 919-966-3168 or 1967
(F) 919-966-9800
ewhitsel@email.unc.edu

Corresponding/senior author (if different from first author correspondence will be sent to both the first author & the corresponding author):

3. Timeline:

Statistical analyses: August 2009 – September, 2009 (or on availability of Freeze 3 data)
Manuscript preparation: October, 2009 – November, 2009
Manuscript revision: December, 2009 – February, 2010
Manuscript submission: March, 2010

4. Rationale:

Blood lipid concentrations are complex phenotypes with a strong genetic component. Heritability estimates for high-density lipoprotein cholesterol (HDL-C), low-density lipoprotein cholesterol (LDL-C) and triglycerides (TG) range from 40-60% (1) and several rare Mendelian forms of dyslipidemia have been identified (2) (3-5). Candidate gene studies also have reported several common loci in multiple genes including *APOE* (6, 7), *CTEP* (8, 9), and *LPL* (10, 11), although these polymorphisms explain only a small fraction of the population-level variation in lipid concentrations. Recent genome-wide association studies (GWA) have identified additional

genes not previously considered by candidate gene studies of lipoprotein concentrations (e.g. *ABCG5, TMEM57, CTCF-PRMT8,* and *DNAH11*) (12).

Several studies have also examined whether genetic variants modify the association between 3-hydroxy-3-methylglutaryl-coenzyme A (HMG-CoA) reductase inhibitors ("statins") and LDL-C concentrations. Lai and colleagues evaluated the association of two *APOA5* gene variants with lipoprotein responses in Genetics of Lipid Lowering Drugs and Diet Network (GOLDN) participants treated with fenofibrate for three weeks and reported that carriers of the *APOA5* 56G (W19) allele had enhanced changes in plasma TG and HDL concentrations(13). Voora et al., (2007) also evaluated 2,361 SNPs in the STRENGTH statin response trial and identified interactions between *ABCA1* and *CETP* genetic variants and LDL reduction(14).

Extant pharmacogenetic studies of lipoprotein responses to statin treatment rely largely on candidate-gene studies because drug receptors, transporters, and metabolizing enzymes are such obvious candidates. In practice, however, many candidate gene pathways overlap and interact. In the renin-angiotensin system, for instance, the angiotensin II receptor type 1 is coupled to Gq/11, and signal conduction occurs through several pathways, including activated phospholipases C, D and A2, adenylyl cyclase, L-type voltage-gated calcium channels, catecholamine release, and eventually gene transcription for proteins that control the accumulation of extracellular matrix, growth factor production, catecholamine release, and inflammatory response (15). These biologic complexities make it difficult to identify the appropriate candidate genes (16). Moreover, replication in candidate-gene studies has been only modestly successful (17). Even after a decade or more of research, clinical applications are few or far off.

Recently, genome-wide association studies (GWAS) have identified and replicated many genetic variants associated with conditions such as type 2 diabetes, dyslipidemia, and inflammatory bowel disease. The large number of statistical tests required in GWAS poses a special challenge because few studies that have DNA and high-quality phenotype data are sufficiently large to provide adequate statistical power for detecting small to modest effect sizes. The requirement for large sample sizes and the importance of replication have served as powerful incentives for scientific collaboration. The Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) Consortium was formed to facilitate GWAS meta-analyses and replication opportunities among multiple large population-based prospective cohort studies, including the Age, Gene/Environment Susceptibility (AGES) -- Reykjavik Study, the Atherosclerosis Risk in Communities Study (ARIC), the Cardiovascular Health Study (ARIC), the Framingham Heart Study (FHS), and the Rotterdam Study (RS). HealthABC (HABC), which will have GWAS data by June 2009, has just joined the effort. With genome-wide data on more than 40,000 participants (>5000 of them African Americans), this collaboration represents a unique resource for evaluating statin-gene interactions in lipoprotein response in the “real world” of community-based studies.

5. Main Hypotheses/Study Questions:

1) To evaluate evidence for genotype-by-statin interaction as it relates to concentrations of:
    i. LDL-C
    ii. HDL-C
    iii. Total cholesterol (TC)

6. Design and Analysis:

General Analysis Plan: The approach is first to conduct within-study analyses of the association between phenotype and genotype for each SNP and then to combine the findings

from the within-study analyses by the method of meta-analysis. The CHARGE Analysis Committee developed for all working groups a set of general guidelines about quality control of genotype data, decisions about what results to share across cohorts, formats for sharing data, strand alignments, coding of alleles, choice of covariates for adjustments, detection of and correction for population structure, within-study phenotype analysis plans, between-study meta-analysis methods, and the development of written analysis plans prior to sharing the results. For most traits, the additive or the 1 degree-of-freedom regression model is used to assess the association between the phenotype and the number of copies of a specified allele. For many underlying patterns of 'true' associations, tests derived from this model have good power compared with other approaches. The single regression coefficient is readily interpreted and easily used in meta-analysis. When imputed genotypes are used, the observed allele count is simply replaced by the imputation's "estimated dosage." Standard errors for the regression estimates are usually calculated with model-robust ('sandwich') methods. Routine adjustment is anticipated for age and sex though specific studies may also adjust for site (CHS, ARIC, HABC), for family relationships (FHS), or for cohort (FHS, RS). Longitudinal methods with repeated measures can take advantage of the data structures available in these cohort studies. Between-study heterogeneity will be examined, but with so few cohorts, this analysis will lack power. Analyses will be conducted separately for the major ethnic groups (European and African-Americans). Use of GWAS data in African-Americans will follow CARE procedures. When necessary, principal components analysis will be used to correct for within-study population structure. Additionally, the method of genomic control will be used to correct both within-study and meta-analyzed GWAS results for possible stratification. The Pharmacogenetics Working Group plans to take advantage of the successful and productive experience embodied in these recommendations.

Analysis Methods: The Pharmacogenetics Working Group is currently using simulation to compare two candidate analysis strategies that assume an additive model of inheritance, the first based on an ordinary least squares regression (OLS) of the pre- minus post-treatment difference in interval-scale phenotypes restricted to new users of a given medication and the second, based on generalized estimation equations (GEE) applied to all participants with genotype data and at least one measurement of the phenotype.

The OLS model is given by $Y_{ij} - Y_{ij-1} = \beta_0 + \beta_1 SNP_i + \beta_2 C_i$, where $Y_{ij} - Y_{ij-1}$ is the pre-post treatment difference for the ith new user across the jth and prior visit, β_0 is the intercept, SNP_i is the genetic variant of interest, and C_i is a vector of covariables including study center and principal components to account for population substructure. The parameter of interest is β_1, which represents the effect of a one-unit increase in the genetic variant on pre-post-treatment difference in the phenotype among the newly treated.

The GEE model is given by $Y_{ij} = \beta_0 + \beta_1 I_{ij} + \beta_2 SNP_i + \beta_3 I_{ij} \times SNP_i + \beta_4 C_{ij}$, where Y_{ij} is the interval-scale phenotype for the ith participant at the jth visit, β_0 is the intercept, I_{ij} is an indicator of medication use (1,0), SNP_i is the genetic variant of interest, and C_{ij} is a vector of covariables including study center, principal components to account for population substructure, and several potential sources of confounding by indication. The parameter of interest is β_3, the multiplicative interaction term.

Simulation data regarding the bias, efficiency and speed of the models will be used to help select final analysis strategies with input from the CHARGE Analysis Committee. We recognize that sample sizes, effect sizes, and power may be modest in some situations. Adopting longitudinal methods, using repeated measures, and establishing additional

collaborations with other studies or consortia that have comparable data may be helpful under such circumstances. The latter approach has worked well in CHARGE. For instance, both the CHARGE Blood Pressure and QT Working Groups, after completing their main papers, now plan joint analyses with other consortia that had submitted parallel papers. In other situations, sensitivity analyses may be required. For example, only three statins (atorvastatin, lovastatin, and simvastatin) are metabolized by CYP3A4; and selected drug-gene interactions may depend importantly on metabolic pathways. In the latter case, sensitivity analysis would be limited to users of the three statins. In all cases, the Pharmacogenetics Working Group will monitor progress during bimonthly conference calls (ongoing) and direct resources to the most promising efforts.

Design of an illustrative analysis: For purposes of illustration, consider a GWAS study of the association between genetic variants and statin treatment. Several variants that affect statin response, including one splice variant, in the HMG-Co-A reductase gene have been identified. In this analysis, investigators from each cohort will identify all new users of statins during follow-up; the response of interest is the change in LDL-C between baseline and the first examination when a statin was used. In a linear regression model, the outcome is change in LDL-C adjusted for age, sex, and statin use, and the primary exposure of interest is the genetic variant, coded as 0, 1 or 2 variant alleles. For this additive model, the regression coefficients estimate the difference in LDL-C associated with each extra copy of the minor allele. The association results from within each cohort are then combined in a fixed-effects meta-analysis to produce CHARGE-wide combined estimates of the beta estimates, standard errors, and p-values. This model, which is analogous to the case-only design for dichotomous traits, assumes that the genetic variant is not associated with short-term changes in LDL.

Genotyping methods and imputation. The CHARGE consortium was developed after each cohort study had contracted for their genotyping platforms and decided on the selection of the individuals to be included in the GWAS. Indeed, the 6 cohorts used 4 different platforms that have fewer than about 60,000 SNPs in common. AGES and CHS used the Illumina 370CNV; ARIC and Health ABC, the Affymetrix 6.0; FHS, the Affymetrix, 500k with MIPS 50K; the RS, the Illumina 550. To maximize the availability of comparable genetic data and coverage of the genome, each cohort used recently developed methods to impute for Europeans and European Americans their genotypes at each of the 2.5 million autosomal CEPH HapMap SNPs. Imputation for the African American populations requires data just becoming available through the extended HapMap project.

Prior to imputation, individuals were excluded for low call rates or sex mismatches. High levels of missingness, significant departures from Hardy-Weinberg equilibrium, or low minor allele frequencies (MAF) were used to determine which SNPs to exclude in the imputation. All the remaining individuals and SNPs entered the imputation process, which provided estimates for all the HapMap SNPs, including any that may have failed the data-cleaning criteria.

Multiple testing The analysis of 2.5 million SNPs across the genome poses an obvious multiple-testing problem. The two primary defenses against reporting false-positive findings are the selection of an appropriate genome-wide threshold for declaring statistical significance and additional efforts at replication. For the GWAS threshold, the Analysis Working Group favors one of two choices. With 2.5 million tests, the use of a Bonferroni correction to control the Family-Wise Error rate (FWER) at 0.05 yields a threshold p-value of $2x10^{-8}$. Another way to interpret this threshold is to estimate the expected number of false-positive (EFP) tests: if there

are no true associations, each test contributes on average $2x10^{-8}$ false positives and, across the genome, yields an expected total of 0.05 false-positive results. Similarly, a threshold of 1/2.5 million, which equals $4x10^{-7}$, gives an expectation of one false-positive result for all tests. Unlike the FWER interpretation, the control of EFP is not "conservative" for correlated tests.

Power For illustration, more than 5,000 participants in the consortia use statins, and about 2000 have data on change in LDL-C concentrations. We assumed that the standard deviation of the within-individual differences is 10% of the baseline concentration (corresponding to a measurement error of 7% per measurement) and that the two-sided alpha was $4x10^{-7}$. For an effect size of 5% change in LDL-C (one sixth of the effect of a "statin" equivalent or about 8 mg/dL at an LDL-C of 160 mg/dL), power is equal to or greater than 97% for MAFs >= 0.05. For an effect size of 7.5% change in LDL, power is 94% or greater for MAFs > 0.02.

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 ICTDER04 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 ICTDER04 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 ICTDER04 must be used to exclude those with value RES_DNA = "No use/storage DNA"?

__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.cscc.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)?

Manuscript proposals #1482 (Lutsey, "Relation of lipid gene score to longitudinal trends in lipid levels, and to statin therapy response in Caucasians: The ARIC Study") and #1391 (Volcik, "A genome-wide association study for HDL-cholesterol, LDL-cholesterol, and triglycerides in ~11,000 African Americans") are related to the proposal presented herein. However, #1482 does not consider a genome-wide approach and does not propose collaborating with other CHARGE consortium members. Proposal #1391 examines the main effect of genetic variants among

African Americans in the CARe consortium. Although our proposal is distinct from these, we welcome collaboration with interested MS #1482 and #1391 investigators.

11. a. Is this manuscript proposal associated with any ARIC ancillary studies or use any ancillary study data?

__X__ Yes

_____ No

11.b. If yes, is the proposal

_X_ A. primarily the result of an ancillary study (AS #2009.10; #2007.02; #2006.03)

___ B. primarily based on ARIC data with ancillary data playing a minor role (usually control variables; list number(s)* _____ _____)

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

The following acknowledgment will appear in the published manuscript

The Atherosclerosis Risk in Communities Study is carried out as a collaborative study supported by National Heart, Lung, and Blood Institute contracts N01-HC-55015, N01-HC-55016, N01-HC-55018, N01-HC-55019, N01-HC-55020, N01-HC-55021, N01-HC-55022, R01HL087641, R01HL59367 and R01HL086694; National Human Genome Research Institute contract U01HG004402; and National Institutes of Health contract HHSN268200625226C. The authors thank the staff and participants of the ARIC study for their important contributions. Infrastructure was partly supported by Grant Number UL1RR025005, a component of the National Institutes of Health and NIH Roadmap for Medical Research. Support for genotyping ARIC participants to facilitate interaction studies was provided by NHGRI through the GENEVA study.

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. Weiss LA, Pan L, Abney M, et al. The sex-specific genetic architecture of quantitative traits in humans. Nat Genet 2006;38:218-22.
2. Miller NE, Miller GJ. Letter: High-density lipoprotein and atherosclerosis. Lancet 1975;1:1033.
3. Jensen HK. The molecular genetic basis and diagnosis of familial hypercholesterolemia in Denmark. Dan Med Bull 2002;49:318-45.
4. Myant NB. Familial defective apolipoprotein B-100: a review, including some comparisons with familial hypercholesterolaemia. Atherosclerosis 1993;104:1-18.
5. Ouguerram K, Chetiveaux M, Zair Y, et al. Apolipoprotein B100 metabolism in autosomal-dominant hypercholesterolemia related to mutations in PCSK9. Arterioscler Thromb Vasc Biol 2004;24:1448-53.
6. Cohen J, Pertsemlidis A, Kotowski IK, et al. Low LDL cholesterol in individuals of African descent resulting from frequent nonsense mutations in PCSK9. Nat Genet 2005;37:161-5.
7. Ordovas JM, Litwack-Klein L, Wilson PW, et al. Apolipoprotein E isoform phenotyping methodology and population frequency with identification of apoE1 and apoE5 isoforms. J Lipid Res 1987;28:371-80.
8. Boekholdt SM, Sacks FM, Jukema JW, et al. Cholesteryl 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;111:278-87.
9. Guerra R, Wang J, Grundy SM, et al. A hepatic lipase (LIPC) allele associated with high plasma concentrations of high density lipoprotein cholesterol. Proc Natl Acad Sci U S A 1997;94:4532-7.
10. Lai CQ, Demissie S, Cupples LA, et al. Influence of the APOA5 locus on plasma triglyceride, lipoprotein subclasses, and CVD risk in the Framingham Heart Study. J Lipid Res 2004;45:2096-105.
11. Pennacchio LA, Olivier M, Hubacek JA, et al. Two independent apolipoprotein A5 haplotypes influence human plasma triglyceride levels. Hum Mol Genet 2002;11:3031-8.
12. 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.
13. Lai CQ, Arnett DK, Corella D, et al. Fenofibrate effect on triglyceride and postprandial response of apolipoprotein A5 variants: the GOLDN study. Arterioscler Thromb Vasc Biol 2007;27:1417-25.
14. Voora D, Reed CR, Zhai J, et al. Polymorphisms In ABCA1 Predict Statin Mediated LDL Cholesterol Lowering And Suggest An Interaction With CETP. *Circulation* 2007;116:II_178.
15. de Gasparo M, Catt KJ, Inagami T, et al. International union of pharmacology. XXIII. The angiotensin II receptors. Pharmacol Rev 2000;52:415-72.
16. Luft FC. Geneticism of essential hypertension. Hypertension 2004;43:1155-9.
17. Hirschhorn JN, Daly MJ. Genome-wide association studies for common diseases and complex traits. Nat Rev Genet 2005;6:95-108.