1a. Full Title: Heterogeneity of Variance for Lipid Traits

b. Abbreviated Title: Heterogeneity of Variance for Lipid Traits

2. Writing Group:

Mariaelisa Graff, Kristin Young, Lekki Frazier-Wood, Alana Morrison, Eric Boerwinkle, Christie Ballantyne, Kari E. North, Paul Franks and other Authors for the Giant and Global Lipids Genetics consortium

Other investigators welcome

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

First author: Mariaelisa Graff

Address: Department of Epidemiology

University of North Carolina at Chapel Hill

137 E. Franklin St, Suite 306

CB #8050

Chapel Hill, NC 27514

Phone: 919-966-8491 Fax: 919-966-9800

E-mail: migraff@unc.edu
Corresponding/senior author (if different from first author correspondence will be sent to both
the first author & the corresponding author): Kari North

Address: Department of Epidemiology
University of North Carolina at Chapel Hill
137 E. Franklin St, Suite 306
CB #8050
Chapel Hill, NC 27514
Phone: 919-966-2148 Fax: 919-966-9800
E-mail: kari_north@unc.edu
3. Timeline:

- Individual cohort statistical analyses: September, 2012
- Consortium meta-analyses: March 2013
- Manuscript preparation: September 2013
- Manuscript submission: March 2014

4. Rationale:

Obesity, often a prelude to diabetes and cardiovascular disease, is becoming pandemic worldwide, particularly in developed countries. Plasma concentrations of lipids and lipoproteins [low-density lipoprotein cholesterol (LDL), high-density lipoprotein cholesterol (HDL-C), triglycerides (TG)] are heritable risk factors for cardiovascular disease [1, 2]. Recently, genome-wide association studies have identified common variants (>5% minor allele frequency) ancestry related to plasma lipids [3, 4] in individuals of European descent. Roughly one third of these loci harbor genes with previously recognized involvement in lipid metabolism, many by virtue of having rare variants that result in Mendelian disorders. Accurate prediction of an individual’s risk will provide effective strategies prevention and treatment or cardiovascular disease and related morbidities. While GWAS have implicated several gene loci in cardiovascular disease and related lipid traits have shed light on biological mechanisms, there remains a large proportion of unexplained variance, partly because of the complexity of gene-gene and gene-environment interactions.

Although we expect most complex traits to be caused by both genetic and environmental factors, there has been little in terms of methodological advancements in the field modeling complex relationships causing common complex disease. While several novel methods addressing the degree and nature of interaction between two or more genetic loci [5-7] have been developed, genome wide SNP-SNP interaction is still impractical, as the multiple
hypothesis testing incurred when searching for interactions between hundreds of thousands of genetic variants is too large.

In a recent paper Pare et al develop a novel approach to prioritize SNPs with interactive effects (gene-gene, gene-environment), termed, variance prioritization. In short, Pare and colleagues demonstrate that the variance of a quantitative trait will differ between the three possible genotypes of a SNP in the presence of genetic interactions. Variance prioritization entails two steps. In the first step, Levene’s test of equality of variance is used to prioritize SNPs for further interaction testing. In the second step, prioritized SNPs are tested for interaction effects against environmental covariates or other SNPs using linear regression. Therefore, we propose to use this novel method to prioritize SNPs that likely contain complex interactions. The proposed project will investigate the inequality of variance of lipid traits, including fasting glucose, fasting insulin, 2-hour glucose tolerance, BMI and waist-hip ratio adjusted for BMI, among the three possible genotypes of a biallelic SNPs. We will use the method demonstrated by Pare et al [8] that incorporates Levene’s test of equality of variance, to a list of prioritized SNPs for subsequent gene–gene and gene–environment testing. This method has the advantageous characteristic that the interacting covariate need not be known or measured for a SNP to be prioritized.

5. Main Hypotheses/Study Questions:

The aim of this proposed project is to investigate the inequality of variance of obesity and lipid traits, including total cholesterol, low-density lipoprotein cholesterol (LDL-C), high-density lipoprotein cholesterol (HDL-C), and triglycerides (TG) among the three possible genotypes of a biallelic SNPs. This may potentially lead to the development of novel methods for the prediction of an individual’s risk of dyslipidemia, potentially informing new guidelines for personalized health and effective strategies for prevention and treatment of cardiovascular disease.
6. Design and Analysis:

Subjects and Sample:
We will use ARIC participants of European ancestry with available anthropometric, genome-wide genotyped or imputed SNPs from freeze 3, and covariate measures. Exclusion criteria include the following: study specific standard exclusion criteria for sample call rate, gender checks, and sample heterogeneity. In addition, individuals who are related, pregnant, or on lipid-lowering therapy will be excluded. If lipid lowering therapy is unknown individuals will be left in analyses. We will use ARIC baseline data where the number of medicated subjects is less than 5%. Therefore, we do not expect that the exclusion of medicated subjects will introduce a large bias.

Definitions and treatment of variables:

**Obesity and Lipid traits**
1. LDL-C will be calculated using Friedewald’s formula \[\text{LDL-C} = \text{TC} - \text{HDL-C} - (\text{TG} \div 5)\] with assignment of missing value to all individuals with TG > 400 mg/dL. For samples with directly measured LDL-C, this should be utilized. Units for LDL-C will be converted to mmol/L.
2. TG will be natural log transform before further adjustment for covariates.
3. untransformed TC. Units for TC will be converted to mmol/L
4. untransformed HDL-C; . Units for HDL-C will be converted to mmol/L.

**Genetic data**
As a priority we will use genotyped and “best guess” imputed SNPs (not imputed dosages)

**Covariates**
Models will be minimally adjusted for age, sex, field center, and ancestry informative markers (10 principal components) to account for population substructure within each race. Because
there are more than 3 centers, we will create N-1 center covariates, each having (0,1) showing if a participant is from that center or not.

**Statistical Analyses**

The first part of the analysis is to generate pre-adjusted and standardized outcomes for each lipid trait. This process will give us three outcomes for each lipid trait: 1) residual, 2) z-score, 3) absolute value of z-score that we can use to associate with the genetic data. For each trait outcome, we will use plink to obtain the mean and standard deviation, per genotype and N (genotype count) as outlined in the following steps:

Step 1 – to obtain the residual “RESID”:

In gender combined and gender stratified models we will first obtain residuals for each obesity and lipid trait (i.e. calculated or directly measured LDL-C, untransformed TC and HDL-C, and natural log transformed TG) after adjusting for age, gender (in gender-combined model), principal components, and field center.

=> residuals -> report: mean (RESID_MEAN_PHENO) and SD (RESID_SD_PHENO) per genotype, and genotype counts (N_) for Levene’s test

Step 2– to obtain the residual “ZSCORE”

-> residuals -> Z-score (standardized residuals) -> report: SD (ZSCORE_SD_PHENO) per genotype for Levene’s test

Step 3– to obtain the residual “ABSLT”

-> residuals-> Z-score (standardized residuals) -> absolute values of the Z-score (change each negative value of the Z-score to positive, keep positive values as they are) -> report: mean (ABSLT_MEAN_PHENO) and SD (ABSLT_SD_PHENO) per genotype for Levene’s test.

**Uploading Results for ARIC Data**
Results from above analyses for ARIC will be posted to the CHARGE ftp server by 1st September in the GWAS/MChip sub-folders for each trait including: TC, TG, LDL-C, and HDL-C BMI for further meta-analyses with other cohorts collaborating on this project.

**Meta-analyses**

Meta-analysis will be implemented centrally by the “Heterogeneity of Variance group” using the custom written software performing Levene’s test meta-analysis, including the option to identify optimal heterogeneity of variance P-value prioritization threshold for each SNP.
1. 7.a. Will the data be used for non-CVD analysis in this manuscript?
   ___ Yes
   ___ No

   b. If Yes, is the author aware that the file ICTDER02 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 ICTDER02 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?
    ___ 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 ICTDER02 must be used to exclude those with value RES_DNA = “No use/storage DNA”?
     ___ 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 website at: http://www.cscc.unc.edu/ARIC/search.php

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)?

This manuscript does not overlap any proposal other than Dr. North’s own proposals.

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

Yes

No

11.b. If yes, is the proposal

A. primarily the result of an ancillary study (AS #2006.03 & 2007.02)

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

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


