1a. Full Title: Heterogeneity of Variance for Obesity and Related Glycemic Traits

b. Abbreviated Title: Heterogeneity of Variance for Glycemic Traits

2. Writing Group:

Mariaelisa Graff, Kristin Young, Lekki Frazier-Wood, Alana Morrison, Eric Boerwinkle, Jim Pankow, Kari E. North, Paul Franks and other Authors for the Giant and MAGIC 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. Impaired glucose tolerance (IGT) is a pre-diabetic state of hyperglycemia that is associated with insulin resistance and increased risk of cardiovascular pathology [1, 2]. However, prevention of obesity and related comorbidities such as glucose intolerance and T2D faces great challenges. Accurate prediction of individual risk will provide effective strategies for obesity prevention and treatment. The recent proliferation of whole-genome population studies provides a vast amount of genotype data that could facilitate the identification of genetic variants [3-6], dietary factors and lifestyle choices [7-9] contributing to the risk of obesity and glucose intolerance. However, even with the most recent GWAS data a large proportion of unexplained variance remains for obesity and glucose related traits.

   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 (PMID:12548676; PMID: 21158747; PMID:12351582) 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 obesity and glycemic 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 [10] 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 glycemic traits, including fasting glucose fasting glucose (FG), fasting insulin (FI), 2-hour glucose (2hrG) tolerance, BMI and waist-hip ratio (WHR) adjusted for BMI, 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 risk for poor glycemic control, potentially informing new guidelines for personalized health and effective strategies for prevention and treatment of obesity.

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, non-fasting, have diabetes mellitus ('diagnosed', on diabetes treatment, oral and insulin), and/or $FG \geq 7 \text{ mmol/L}$, and/or $2hrG \geq 11.1 \text{ mmol/L}$) will also be excluded.

Definitions and treatment of variables:

**Obesity and Glycemic traits**
- untransformed **fasting PLASMA glucose**, $FG$ (in mmol/L). Check units: multiply mg/dl by 0.0555 to get mmol/L. If glucose measurement was made in BLOOD, you should adjust it by multiplying by 1.13, because glucose concentration in blood is lower than in plasma. Check the laboratory method carefully and ensure, you use PLASMA values for the analysis.
- untransformed **fasting insulin**, $FI$ (in pmol/L). Check units: 1pmol/L=6 mg/dL. This trait will be log transformed before further adjustment for covariates.
- untransformed **2hr Glucose**, $2hrG$ (in mmol/L). Check units: multiply mg/dl by 0.0555 to get mmol/L.
- untransformed **BMI** (in kg/m$^2$)
- untransformed **Waist to Hip Ratio**, WHR. (WHR will be additionally adjusted for BMI in the current analysis.)

**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 glycemic trait. This process will give us three outcomes for each glycemic 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 glycemic trait (i.e. FG, log-transformed FI, and 2hrG, BMI and WHR adjusted for BMI) 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: FG, FI, 2hrG, WHR(adjusted for BMI), and 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

___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)?

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
___x___ 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