ARIC Manuscript Proposal

1a. Full Title: Analysis of single nucleotide polymorphisms from genome-wide association data for adiposity traits

b. Abbreviated Title: GWAS and adiposity

2. Writing Group:
Writing group members:
   Kari North
   Anne Justice
   Mariaelisa Graff
   Yujie Wang

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
123 E. Franklin St, #4216
CB #8050
Chapel Hill, NC 27514
Phone: 919-966-2148 Fax: 919-966-9800
E-mail: migraff@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:
   Data arrival: June 1, 2018
   Statistical analyses (includes meta-analyses with other cohorts): June – Dec, 2018
   Manuscript preparation: Jan – April, 2019
   Manuscript revision: April – August 2019
   Manuscript submission: September – September 2019

4. Rationale:
Several lines of evidence support the role of genetics in the regulation of body mass, which include both longitudinal family and twin studies that show that body mass index (BMI), weight, and weight change are all heritable traits (1-7). However, most forms of obesity do not follow simple Mendelian modes of inheritance and thus investigating potential genetic variants that contribute to common forms of obesity require large population-based studies. To date, researchers have had great success using genome-wide association studies (GWAS) of obesity-related anthropometric traits (8, 9).
Despite the growing body of literature on the genetic architecture of obesity-related traits, several questions remain that result in a poor understanding of the mechanisms by which obesity leads to the cascade of negative cardiometabolic outcomes (10). For example, few studies have focused investigations on key life stages with accelerated body mass change, including the period following menopause, but research suggests that menopause is a critical life stage, separate from earlier adulthood and later senescence, during which women are at increased risk for adiposity gain (11-23). A number of recent genetic epidemiological studies have utilized longitudinal data in order to take advantage of repeat measurements of time-varying variables to increase power to detect genetic effects on complex traits, and identify unique genetic components to changes in these phenotypes across time (24).

We aim to look for genetic loci associated with adiposity change, including BMI and weight change, as well as waist circumference and waist-to-hip ratio change in men and women between the ages of 20 and 65 years. Further, we will conduct separate association analyses stratified by those that gain or lose weight.

We will meta-analyze the result from ARIC with several other cohorts listed below.

5. Main Hypotheses/Study Questions:
To complete analyses on the genome-wide SNP data available imputed to 1000 genomes, phase 3 and on whole-exome sequencing data in the ARIC sample. Assessed phenotypes will include the following (in order of priority):

- Weight and/or BMI change
- Waist circumference change
- Waist-to-hip ratio (WHR) change

For these phenotypes we will use data from visits 1-4. We plan also to investigate only those who gain weight over time and or decrease weight
6. Design and Analysis:
Subjects and Sample size:
The usual DNA consent restriction and missing data exclusion criteria will be used. Mean (SD) levels and sample sizes by race for the main phenotypes to be assessed are in the table below.

<table>
<thead>
<tr>
<th></th>
<th>Whites</th>
<th>African Americans</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Body mass index (kg/m^2)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>v1</td>
<td>27.0 (4.9) n=11468</td>
<td>29.6 (6.2) n=4196</td>
</tr>
<tr>
<td>v2</td>
<td>27.3 (4.9) n=10720</td>
<td>30.0 (6.3) n=3509</td>
</tr>
<tr>
<td>v3</td>
<td>27.9 (5.2) n=9838</td>
<td>30.4 (6.4) n=2951</td>
</tr>
<tr>
<td>v4</td>
<td>28.3 (5.2) n=8946</td>
<td>30.6 (6.4) n=2603</td>
</tr>
<tr>
<td><strong>Waist circumference (cm)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>v1</td>
<td>96.3 (13.4) n=11464</td>
<td>99.2 (15.2) n=4198</td>
</tr>
<tr>
<td>v2</td>
<td>97.0 (13.9) n=10722</td>
<td>101.0 (15.2) n=3517</td>
</tr>
<tr>
<td>v3</td>
<td>100.0 (13.9) n=9841</td>
<td>102.8 (15.7) n=2948</td>
</tr>
<tr>
<td>v4</td>
<td>101.4 (14.1) n=8951</td>
<td>104.1 (15.8) n=2603</td>
</tr>
<tr>
<td><strong>WHR (waist-to-hip ratio)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>v1</td>
<td>77.0 (16.3) n=11437</td>
<td>83.2 (17.4) n=4183</td>
</tr>
<tr>
<td>v2</td>
<td>77.8 (16.4) n=10691</td>
<td>84.2 (17.7) n=3502</td>
</tr>
<tr>
<td>v3</td>
<td>79.0 (16.8) n=9807</td>
<td>85.0 (18) n=2938</td>
</tr>
<tr>
<td>v4</td>
<td>79.7 (16.9) n=8918</td>
<td>85.5 (18) n=2591</td>
</tr>
<tr>
<td><strong>Weight (kg)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>v1</td>
<td>0.9 (0.1) n=11432</td>
<td>0.9 (0.1) n=4185</td>
</tr>
<tr>
<td>v2</td>
<td>0.9 (0.1) n=10692</td>
<td>0.9 (0.1) n=3502</td>
</tr>
<tr>
<td>v3</td>
<td>0.9 (0.1) n=9810</td>
<td>0.9 (0.1) n=2934</td>
</tr>
<tr>
<td>v4</td>
<td>1.0 (0.1) n=8920</td>
<td>0.9 (0.1) n=2590</td>
</tr>
</tbody>
</table>

Quality control of genotyping data
Before imputation genetic data was cleaned using standard procedures (e.g. analyses of blind duplicates, remove individuals and SNPs with excessive missing data, exclusion of SNPs without chromosomal location, removal of SNPs that fail Hardy-Weinberg equilibrium p<10^{-6}, and exclusion of individuals with outlying heterozygosity). Subsequently, we imputed to 1000 genomes, phase 3 as well as the Human Reference panel for Europeans only. Imputed data to 1000 genomes or the HRC (Europeans only) is available for 9489 Europeans and 2906 African Americans in ARIC. We also have whole exome sequencing data for 2817 African Americans in ARIC and 5719 European Americans in ARIC which we will also plan to use for association analyses, and further meta-analyses with other cohorts.
Publication strategy
We plan to meta-analyze the results in ARIC with other studies. Some studies will also be used for replication efforts. These additional studies include: Women’s health initiative (WHI), Hispanic Communities Health Study, Study of Latinos (HCHS/SOL), Multi-Ethnic Study of Atherosclerosis (MESA), Framingham (FHS), The Coronary Artery Risk Development in Young Adults Study (CARDIA), Santiago Longitudinal Study, Jerusalem Perinatal Study, Kibbutzim Family Study, UK Biobank, DiscovEHR’ partnership, Geisinger and Regeneron Genetics Center (Geisinger/MyCode/DiscovEHR), CoLaus study, BioMe Biobank, METabolic Syndrome In Men (METSIM) study, Prostate, Lung, Colorectal and Ovarian (PLCO), and Netherlands Epidemiology of Obesity’ (NEO).

All of these studies have data collected at multiple time points in adults (ages 20 to 65), as well as genotyped data imputed to 1000 genomes, phase 3, HRC imputed data, and/or whole exome sequencing data.

Definitions and treatment of variables
Change over time for BMI, weight, waist, and WHR will be defined by a slope using multiple time points for each phenotype, but sex and ancestry. The slope or growth curve analysis as follows: There will be multiple observations per subject. Keep only those subjects with at least 2 measures of weight after menopause. Each observation should have weight and corresponding AGE. A growth curve model is a mixed model that has fixed effects for intercept and age and also random effects (for each person) for intercept and age.

Using slope from above, we will next obtain the residuals by performing linear regression analysis in each strata and adjusting for age at first measure, principal components, years since menopause (for women, if available), smoking status at first measure.

\[ \text{slopetwcg} = \text{age} + \text{age}^2 + \text{height} + \text{PC1}...\text{PCN} + \text{smoker (y/n)} + \text{study-specific} \]

*Slopetwcg residuals

Analysis strategy / statistical analysis
Modeling strategy: Prior to running genetic models, sex- and race-specific slope residuals will be calculated for each phenotype controlling for age, age-squared, ARIC field center, up to 10 principal components, and current smoking status (yes/no). For the slope residuals, we will test for normality by examining histogram plots of residuals to determine if outcome meets assumptions of normality, and apply appropriate trait transformation if distribution of the slope for the particular trait statistically deviate from normality. Then we will conduct a SNP- trait change slope association analysis for total sample by ancestry, in the subsample of weight gainers only. All analyses will be run sex-combined and sex-stratified.

Meta-analytic strategy: Meta-analysis will be done separately in studies based on the genetic panels imputed to and for those with whole exome-sequencing data. Meta-analyses with a total of 3 types of genetic data are possible: 1) meta-analyses with European ancestry only imputed to the HRC panel, 2) meta-analyses with all ancestries imputed to 1000 genomes, phase 3 panels, and 3) whole-exome sequencing data for each adiposity trait change. We would just focus on the 1000 genomes imputed data, except that the HRC imputed data includes many more variants and might harbor new loci, or interesting details about the architecture of known loci. Whole-exome sequenced data allows us to focus on the rare, but coding variants of the genomes, which could also be beneficial as these variants mainly contribute to coding variation across the genome.

Population stratification: Principal component scores will be incorporated into genetic models to account for population stratification in each of the samples.

Multiple testing: For single-variant analyses of imputed data, we plan to use a genome-wide cut-off of 4e-9 for imputed variants to HRC or 1000 genomes, phase 3 and p<5e-7 for whole-exome sequencing data.
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 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?
   _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 ICTDER02 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)?  
#1368: Analysis of single nucleotide polymorphisms from genome-wide association data for adiposity traits  
The above ARIC manuscript proposal focused on cross sectional measures of adiposity traits to look at genome-wide associations.

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 (list number __AS#1995.07__)  
__ 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/

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


