ARIC Manuscript Proposal # 1368

PC Reviewed: 05/13/08 **Status: A** **Priority: 2**
SC Reviewed: _____ **Status: _____** **Priority: ____**

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
Keri Monda
Ellen Demerath
Linda Kao
Eric Boerwinkle
Braxton Mitchell (with the OOA Study)
Caroline Fox (with the FHS)
Tamara Harris (with the AGES Study)

Other investigators welcome

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

First 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

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, 2008
Statistical analyses: June – July, 2008
Manuscript preparation: July – August, 2008
Manuscript revision: August – September 2008
Manuscript submission: September – October 2008

4. Rationale:

Several lines of evidence support the role of genetics in the regulation of body mass, including longitudinal family and twin studies which show that BMI, weight, and weight change are all heritable traits (Adams, Hunt et al. 1993; Austin, Friedlander et al. 1997; Lee, Reed et al. 1997; Bouchard, Perusse et al. 1998; Comuzzie and Allison 1998; Hunt, Katzmarzyk et al. 2002; Loos and Bouchard 2003). 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 will require large population-based studies. Linkage analyses of family-based data have identified areas of the human genome that are associated with adiposity traits (Golla, Strauch et al. 2003; Fox, Heard-Costa et al. 2005). In fact, according to the annually updated 'Obesity Gene Map' (Rankinen, Zuberi et al. 2006) 253 quantitative trait locus (QTL) regions for obesity-related phenotypes have been identified in 61 genome-wide scans, and a total of 52 genomic regions that harbor QTLs replicated in two or more studies. Despite this, no specific genetic variants clearly responsible for any of the linkage signals have been identified. It is only with recent major technological advances that we have rapidly expanded options for the evaluation of genetic variation at the level of the single nucleotide polymorphism (SNP).

Genome-wide Association (GWA) studies interrogate whether variation across the human genome in the form of SNPs is associated with given phenotypes. GWAS are now widely recognized as powerful data-driven tools for identifying genetic variants related to common complex diseases such as obesity. GWA was used to identify variants near *INSIG2* reported to be associated with obesity in the Framingham Heart Study and replicated in four of five studies of adults and children (Herbert, Gerry et al. 2006). Just last year, an association between *FTO* and obesity was reported by Frayling et al. (Frayling, Timpson et al. 2007) and has since been widely replicated in multiple race/ethnic populations as well as adults and children (Dina, Meyre et al. 2007; Scuteri, Sanna et al. 2007; Grant, Li et al. 2008; Hunt, Stone et al. 2008; Lopez-Bermejo, Petry et al. 2008). Replication of findings is a key ingredient in genetic epidemiology studies and investigators are encouraged to set up collaborations to facilitate this. Failure to replicate could be due to many reasons including sample differences, lack of power to find an effect, incomplete phenotype harmonization, among others.

5. Main Hypotheses/Study Questions:

To complete analyses on the genome-wide SNP data available (~1,000,000 SNPs) on the ARIC sample through its collaboration with the Broad Institute. Assessed phenotypes will include the following (in order of priority):

- Baseline BMI and waist circumference measures
- Mean and maximum BMI and waist circumference measurements
- Height
- Alternative measures of adiposity (skinfold measures)
- Weight and/or BMI change

We plan also to investigate qualitative traits based on the quantitative traits listed above. For example, using standard cutoffs, we will investigate the phenotypes overweight

(BMI≥25 kg/m^2), obesity (BMI≥30 kg/m^2), and high waist circumference (≥88 cm women, ≥102 cm men).

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.

	Whites	**African Americans**
Body mass index (kg/m^2)		
v1	27.0 (4.9) n=11468	29.6 (6.2) n=4196
v2	27.3 (4.9) n=10720	30.0 (6.3) n=3509
v3	27.9 (5.2) n=9838)	30.4 (6.4) n=2951
v4	28.3 (5.2) n=8946	30.6 (6.4) n=2603
Waist circumference (cm)		
v1	96.3 (13.4) n=11464	99.2 (15.2) n=4198
v2	97.0 (13.9) n=10722	101.0 (15.2) n=3517
v3	100.0 (13.9) n=9841	102.8 (15.7) n=2948
v4	101.4 (14.1) n=8951	104.1 (15.8) n=2603
Skinfold (mm)*		
v1	46.4 (16.7) n=11436	59.5 (25.1) n=4179
v2	42.2 (15.4) n=10723	51.8 (22.2) n=3513

*Sum of triceps and subscapular skinfolds

Quality control of genotyping data

The following quality control analyses of genotype data will take place prior to our receiving it: analyses of blind duplicates, individuals analyzed for excessive missing data, SNPs analyzed for excessive missing data, exclusion of SNPs without chromosomal location, exclusion of monomorphic SNPs, analysis of Hardy-Weinberg equilibrium and exclusion of autosomal SNPs with $p<10^{-6}$, and exclusion of individuals with outlying heterozygosity.

Publication strategy

Because replication of findings is a necessary component of this work, we have already set up a collaborative agreement with investigators from on other studies to replicate any significant results we might find. Collaborators include investigators from the Framingham Heart Study (FHS), studies of the Old Order Amish (OOA), and the Age Gene/Environment Susceptibility (AGES)-Reykjavik Study (investigators' names included in the writing group). Brief descriptions of these studies are below. This group will facilitate replication for findings in the white sample; we are still working to set up collaborations to replicate any findings from the African American sample.

FHS: The Framingham Heart Study is a longitudinal study of cardiovascular disease with the original participants (aged 30-62) recruited in 1948. A second-generation group composed of the original participants' adult children and their spouses was enrolled in 1971. Finally, a third-generation was enrolled in 2002. Genotyping data is available via

the Affymetrix 500K SNP chip on a sample size of approximately 9300 individuals. The FHS is a joint project of the NHLBI and Boston University.

OOA: Approximately 1,600 Old Order Amish individuals from Lancaster, PA will also be included. These individuals, aged 18 years and older, were participants of studies carried out by investigators at the University of Maryland, Baltimore (the HAPI Heart Study, the Amish Family Calcification Study, and the Amish Longevity Study). Genotyping data (via the Affymetrix 500K or 1M SNP chip) is available on approximately 1600 individuals.

AGES-Reykjavik: The Age Gene/Environment Susceptibility-Reykjavik Study was initiated in 2002 to examine risk factors in relation to disease and disability in old age. The sample is drawn from an established population-based cohort, the Reykjavik Study, of men and women born between 1907 and 1935 in Iceland. Genotyping data will be available in June via the Affymetrix chip.

Definitions and treatment of variables

Genotype: In our large-scale analysis we will use an additive model to estimate the association between SNP and adiposity traits. In an additive model the SNP is coded as a continuous variable (0=major homozygote, 1=heterozygote, 2=minor homozygote), thus the heterozygote is forced to have an effect midway between the two homozygotes. While less flexible than the codominant (or general) model, the additive model is often seen as biologically plausible. Nonetheless, it was selected primarily for its usefulness in meta-analytic procedures. The additive model, while not as flexible as the co-dominant (general) model, has been shown to perform well when the underlying mode of inheritance is additive or dominant, but less well when recessive (Lettre, Lange et al. 2007).

Phenotype measures: Measures used in phenotype characterization in ARIC (BMI, waist circumference, height, and sum of triceps and subscapular subcutaneous skinfolds) will be used primarily as continuous variables. If these variables exhibit significant skewness or kurtosis, a variety of transformations will be tested to achieve normality prior to analysis. As stated above, the qualitative traits overweight (BMI$\geq$25 kg/m^2), obesity (BMI$\geq$30 kg/m^2), and high waist circumference ($\geq$88 cm men, $\geq$102 cm women) will be created from their quantitative versions.

Analysis strategy / statistical analysis

Modeling strategy: Prior to running genetic models, sex- and race-specific residuals will be calculated for each phenotype controlling for age, age-squared, ARIC field center, and current smoking status (yes/no). Next, as mentioned above, we will run additive models to estimate the association between the SNP and the sex- and race-specific residual. Additive models were selected as the genetic model of choice to facilitate meta-analyses with our collaborators with whom we will replicate results. Linear regression models will be used to assess the relationship between quantitative traits and the genetic factors and logistic regression models will be used for assessing the relationship with qualitative traits.

Meta-analytic strategy: Meta-analysis will be done on any SNPs found to be informative in any of the four studies. With our collaborators we are discussing meta-analytic procedures and will run analyses based on p-values or based on both effect estimates and p-values. These considerations are notable due to differences in analytic strategy between cohorts. Because ARIC analyses have yet to proceed, we are fortunate to have some flexibility in our analysis strategy to facilitate replication. As a final point, we mention that while ARIC has data from the 1 million SNP chip, some of our collaborators have data on the 500K SNP chip; thus plan first to look at the intersection of SNPs directly typed. As a secondary aim we will consider imputation strategies to assess the SNPs not directly genotyped.

Population stratification: Because systematic differences in ancestry can produce spurious associations, all analyses will be stratified by race to account for systematic allele frequency differences between racial groups. However, we will also need to account for population substructure within racial groups. While there are a number of methods available to the analyst, we can take advantage of GWAS data most effectively using the principal components analysis method developed by Price and colleagues (Price, Patterson et al. 2006), implemented in the software EIGENSOFT. This method explicitly models ancestry and has higher power to detect true associations than other methods. Principal component factor scores will be incorporated into genetic models to account for population stratification in each of the samples.

Multiple testing: The large number of statistical tests these analyses entail will yield false positive results unless appropriate corrections are made for multiple testing. We will control for this using the Bonferroni correction on an overall α=0.05, a standard approach in GWA analyses, resulting in a significant p-value of approximately 0.05×10^{-6}. Secondarily, we will use the false discovery rate (FDR) (Benjamini, Drai et al. 2001; van den Oord and Sullivan 2003) and permutation-based procedures (Nichols and Holmes 2002).

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)?

#795: Resistin gene polymorphisms and association with obesity and body size measures in African Americans, Mexican Americans, and non-hispanic Whites from two community-based studies

#814: Association of beta2-adrenergic receptor polymorphisms with asthma and obesity in the Atherosclerosis Risk in Communities Study

#1041: Obesity resistance in an aging population and the effects of two obesity candidate genes in the Atherosclerosis Risk in Communities Study.

#1269r: *FTO*, Obesity and Diabetes.

The above ARIC manuscript proposals all describe studies in which candidate obesity genes will be investigated with adiposity traits. None of these proposals include the use of GWAS data.

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

Adams, T. D., S. C. Hunt, et al. (1993). "Familial aggregation of morbid obesity." Obes Res 1(4): 261-70.

Austin, M. A., Y. Friedlander, et al. (1997). "Genetic influences on changes in body mass index: a longitudinal analysis of women twins." Obes Res 5(4): 326-31.

Benjamini, Y., D. Drai, et al. (2001). "Controlling the false discovery rate in behavior genetics research." Behav Brain Res 125(1-2): 279-84.

Bouchard, C., L. Perusse, et al. (1998). The Genetics of Human Obesity. Handbook of Obesity. G. Bray, C. Bouchard and W. James. New York, Marcel Dekker, Inc.: 157-190.

Comuzzie, A. G. and D. B. Allison (1998). "The search for human obesity genes." Science 280(5368): 1374-7.

Dina, C., D. Meyre, et al. (2007). "Variation in FTO contributes to childhood obesity and severe adult obesity." Nat Genet 39(6): 724-6.

Fox, C. S., N. L. Heard-Costa, et al. (2005). "Genomewide linkage analysis of weight change in the Framingham Heart Study." J Clin Endocrinol Metab 90(6): 3197-201.

Frayling, T. M., N. J. Timpson, et al. (2007). "A common variant in the FTO gene is associated with body mass index and predisposes to childhood and adult obesity." Science 316(5826): 889-94.

Golla, A., K. Strauch, et al. (2003). "Quantitative trait linkage analysis of longitudinal change in body weight." BMC Genet 4 Suppl 1: S7.

Grant, S. F., M. Li, et al. (2008). "Association analysis of the FTO gene with obesity in children of Caucasian and African ancestry reveals a common tagging SNP." PLoS ONE 3(3): e1746.

Herbert, A., N. P. Gerry, et al. (2006). "A common genetic variant is associated with adult and childhood obesity." Science 312(5771): 279-83.

Hunt, M. S., P. T. Katzmarzyk, et al. (2002). "Familial resemblance of 7-year changes in body mass and adiposity." Obes Res 10(6): 507-17.

Hunt, S. C., S. Stone, et al. (2008). "Association of the FTO Gene With BMI." Obesity (Silver Spring) 16(4): 902-4.

Lee, J. H., D. R. Reed, et al. (1997). "Familial risk ratios for extreme obesity: implications for mapping human obesity genes." Int J Obes Relat Metab Disord 21(10): 935-40.

Lettre, G., C. Lange, et al. (2007). "Genetic model testing and statistical power in population-based association studies of quantitative traits." Genet Epidemiol 31(4): 358-62.

Loos, R. J. and C. Bouchard (2003). "Obesity--is it a genetic disorder?" J Intern Med 254(5): 401-25.

Lopez-Bermejo, A., C. J. Petry, et al. (2008). "The Association between the FTO Gene and Fat Mass in Humans Develops by the Postnatal Age of Two Weeks." J Clin Endocrinol Metab 93(4): 1501-5.

Nichols, T. E. and A. P. Holmes (2002). "Nonparametric permutation tests for functional neuroimaging: a primer with examples." Hum Brain Mapp 15(1): 1-25.

Price, A. L., N. J. Patterson, et al. (2006). "Principal components analysis corrects for stratification in genome-wide association studies." Nat Genet 38(8): 904-909.

Rankinen, T., A. Zuberi, et al. (2006). "The human obesity gene map: the 2005 update." Obesity (Silver Spring) 14(4): 529-644.

Scuteri, A., S. Sanna, et al. (2007). "Genome-Wide Association Scan Shows Genetic Variants in the FTO Gene Are Associated with Obesity-Related Traits." PLoS Genet 3(7): e115.

van den Oord, E. J. and P. F. Sullivan (2003). "False discoveries and models for gene discovery." Trends Genet 19(10): 537-42.