ARIC Manuscript Proposal # 1274

PC Reviewed: __8/_21/07	**Status: __A__**	**Priority: 2___**
SC Reviewed: _____	**Status: _____**	**Priority: ____**

1.a. Full Title: Genome-wide association analysis identifies region at 20q11.2 with consistent effect on human height.

b. Abbreviated Title (Length 26 characters): GWA of height

2. Writing Group:
S. Sanna, A.U. Jackson, G. Usala, C.J. Willer, M. Dei, L.L. Bonnycastle, S. Lai, Y. Li, M. Uda, M.R. Erdos, H. Shen, A. Shuldiner, E. Boerwinkle, J. Coresh, A. Cao, R.M. Bergman, D. Schlessinger, F.S. Collins, M. Boehnke, G.R. Abecasis, R. Nagaraja, K.L. Mohlke

I, the first author, confirm that all the coauthors have given their approval for this manuscript proposal. __SS__ **[please confirm with your initials electronically or in writing]**

First author: Serena Sanna
Address: University of Michigan
Department of Biostatistics
Center for Statistical Genetics
1420 Washington Heights
Ann Arbor, MI 48109-2029

Phone: (734) 936 1017 Fax:
E-mail: sannas@umich.edu

Communicating author for ARIC (if different from first author correspondence will be sent to both the first author & the communicating author):
Eric Boerwinkle
Address: Human Genetics Center
University of Texas Health Science Center at Houston
1200 Herman Pressler, Suite E-447
Houston, TX 77030

Phone: 713-500-9816 Fax: 713-500-0900
E-mail: eric.boerwinkle@uth.tmc.edu

3. Timeline:
Genotyping of the whole cohort can commence immediately, and be completed in three days.

Manuscript is already partially written including results from the other cohorts.
Following analysis, results can be added quickly.
Manuscript should be ready August 15, 2007

4. Rationale:

Height is a highly heritable quantitative trait. Family studies have repeatedly shown that genetic variation explains >80% of the inter-individual variation in height. Nevertheless, with the exception of a few rare Mendelian syndromes, identification of the underlying genes has proven difficult. Genetic influences on height are probably due to the contribution of multiple loci with common allelic variation each with small to moderate effects. We have carried out a meta-analysis of genome-wide association results from two different groups, ProgeNIA and FUSION. The first sample consist of 4,305 individuals from 570 families from Sardinia, the second includes 2,366 mostly unrelated Finnish individuals. In both GWA scans, we evaluated the additive effect of each SNP, adjusting the model for familiality (i.e. non-independence of the observations) and covariates. In the combined results, the top associated SNP ($p=4.0 \times 10^{-7}$) maps to a region on chromosome 20q11.2 containing several genes, including one previously implicated in growth. Replication is ongoing, but preliminary results on 2017 Finnish and 858 Amish samples support our initial finding ($p=1.7 \times 10^{-3}$), with the same direction of effect.

The ARIC study provides a good opportunity to validate the above-mentioned finding and test its generalizability in the US population. The ARIC study's large sample size provided good power for this study and its biethnic characteristics provide an opportunity not possible in most European populations. Finally, ARIC measured both height and sitting height and the baseline examination, and ARIC's age range dictates that adult height would have been reached by the baseline examination.

Note: The SNP currently does not have an rs#. However, it is the intention of the research group to submit the necessary information to dbSNP to have an rs# assigned. By the time of publication, it should have an rs#.

5. Main Hypothesis/Study Questions:

1. The 20q11.2 height-associated SNP is (is not) associated with baseline height in ARIC study participants. Separate analyses will be carried out for whites and African-Americans.
2. The 20q11.2 height-associated SNP is (is not) associated with baseline sitting height in ARIC study participants. Separate analyses will be carried out for whites and African-Americans.

6. Design and analysis (study design, inclusion/exclusion, outcome and other variables of interest with specific reference to the time of their collection, summary of data analysis, and any anticipated methodologic limitations or challenges if present).

All analyses will be carried-out at the ARIC DNA laboratory in Houston, under the direction of Dr. Eric Boerwinkle. No data will be transferred to non-ARIC investigators as part of this manuscript.

All consenting ARIC participants will be genotyped. Exclusions will include those that have restricted the use of their DNA or limited use of their data in non-CVD research. In addition, those reporting to be non-black and non-white (e.g. Asian) will be excluded.

Height will be considered as a quantitative trait. Prior to the primary analyses, height and sitting height will be investigated for obvious outliers and agreement to normality. Anthropomtrics from visit 1 (i.e. baseline) will be used since height is measured with little error. Consistency across visits will be examined and inconsistent values will be scrutinized. Measurements greater than four standard deviations from the mean will be excluded. (**Note**: In very preliminary analyses, 2 individuals may be excluded because of outlying values.) Goodness of fit to a normal distribution, and, if necessary, normality (or near normality) will be achieved via transformation, such as log or square-root.

The primary analysis tool with be the analysis of variance with covariates. All analyses will be done race-specific. The covariates to be included are sex and age. An additive model for the single SNP will be assumed.

Note: The results will be part of a multi-cohort paper.

7.a. Will the data be used for non-CVD analysis in this manuscript? __X__ 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?

_X___ 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)? There are no manuscript proposals that are related to this one.

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 (list number* _2006.03__)

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

Note: The title of 2006.03 is GWA for incident CHD and other HLB phenotypes.

*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. Agreed.