ARIC Manuscript Proposal # 1309

PC Reviewed: 11/13/07	**Status: __A__**	**Priority: __2_**
SC Reviewed: _____	**Status: _____**	**Priority: ____**

1.a. Full Title: Genome-wide admixture mapping analyses of cardiovascular and related metabolic traits

b. Abbreviated Title (Length 26 characters): admixture mapping

2. Writing Group:
Writing group members: Linda Kao, Ching-Yu Cheng, Man Li, David Reich, Jim Wilson, Joe Coresh, Eric Boerwinkle, and anyone else interested in a particular phenotype for admixture. In particular, we would welcome an interested ARIC investigator from Jackson since the proposal focuses on African-Americans

We would like to submit a general proposal for admixture mapping analyses for various phenotypes. The ones that we are definitely working as of the date of this submission are: white blood cell counts, type 2 diabetes, obesity-related traits. We welcome any other investigators who would like to perform admixture mapping analyses for other phenotypes to use these data (already returned to the Coordinating Center in October 2007).

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

First author: Linda Kao (may change depending on the phenotype being examined)
Address: 615 N. Wolfe St., Room W6513
Baltimore, MD 21205

Phone: 410-614-0945 Fax: 410-955-0863
E-mail: wkao@jhsph.edu

Corresponding/senior author (if different from first author correspondence will be sent to both the first author & the corresponding author):

Linda Kao
Address: 615 N. Wolfe St., Room W6513
Baltimore, MD 21205

Phone: 410-614-0945 Fax: 410-955-0863
E-mail: wkao@jhsph.edu

3. Timeline:

Starting Analyses: June 2007

First Draft: February 2008 for white blood cell and diabetes, March 2008 for obesity related traits

Submission for Publication: June 2008

4. Rationale:

Traditional methods for identification of disease-causing variants, such as family-based linkage analysis, have been very effective in mapping rare disorders for which single mutations are sufficient to cause disease, but have been unsuccessful in identification of genes for complex common conditions, such as type 2 diabetes. Association studies have been shown to provide greater power for identifying variants responsible for common complex conditions[1, 2].

Mapping by Admixture Linkage Disequilibrium (MALD) is a specialized form for genome-wide association design. It is a highly-efficient (needs only ~1% of markers required in a genome-wide direct association mapping and innovative approach to studying the genome systematically for detecting disease susceptibility genes[3-13].

The basic principle of mapping by admixture linkage disequilibrium is straight forward. Although most human genetic variations exist across populations, some vary substantially across populations. Thus, when mixing occurs between genetically heterogeneous populations, the admixed offspring population inherits chromosomal regions of distinct ancestry. This generates association among alleles that are informative for ancestry. Over successive generations of random mating, initial association between linked loci, or linkage disequilibrium, persists longer than association between unlinked loci. MALD exploits admixture-generated linkage disequilibrium to map loci that explain phenotypic variation between the ancestral populations.

Figure 1 Conceptual framework of Mapping by Admixture Linkage Disequilibrium (MALD) analysis using a case-control design

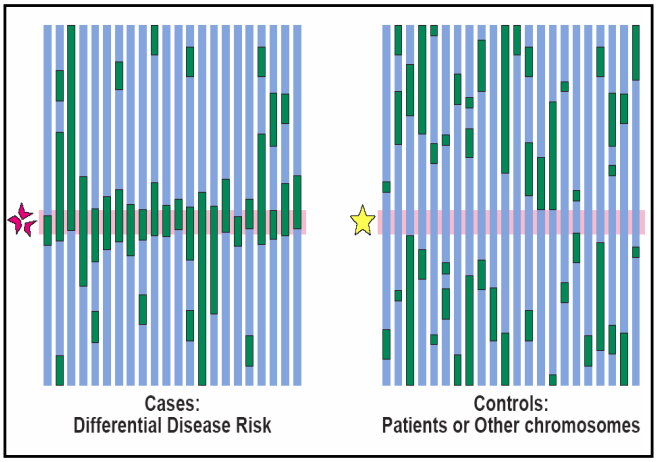

Figure 1 presents a schematic representation of MALD analysis. The dark areas represent chromosomal segments from the parental population with the higher frequency of the risk allele, e.g. African population. Due to mixing of populations and recombination over several generations, originally large blocks of DNA from the African ancestry have become incorporated in smaller segments throughout the chromosome. Risk alleles that are inherited from the ancestral population are closely related to nearby ancestral markers (i.e., in linkage disequilibrium). A comparison of cases (affected) and controls (unaffected) will show an excess number of ancestral markers in cases near the disease locus.

The principal requirement for the MALD strategy to work is the existence of a gene or genes that are both associated with disease risk and have substantial allele frequency difference between the two populations ancestral to the admixed population. The greater relative risk of T2DM in African Americans in comparison to Caucasians, the higher incidence in related individuals, and the recent history of ongoing admixture between Africans and Caucasians make T2DM in African-Americans an ideal problem to study with the MALD approach.

5. Main Hypothesis/Study Questions:

The central hypothesis of the present study is that some susceptibility alleles for type 2 diabetes (T2DM), and other cardiovascular traits, are present at higher frequency in African Americans than in whites and that specific regions of the genome in African Americans contain marker alleles that are in admixture linkage disequilibrium with T2DM (and other cardiovascular traits) susceptibility alleles. This hypothesis will be tested by comprehensively evaluating MALD markers in ARIC African-American participants.

6. Data (variables, time window, source, inclusions/exclusions):

Inclusion:

All African-American participants

1) giving consent for use of DNA by for-profit investigators
2) have sufficient DNA
3) are from either Jackson or Forsyth County

Outcome: type 2 diabetes, white blood cell count, obesity-related traits, and other cardiovascular traits

Exposure: estimated ancestry using ancestry informative markers

Covariates include, but are not limited to, risk factors for diabetes including age, sex, race, lipid levels, hypertension, physical activity, and (in some analyses) adiposity

Analysis Plan

We will use a novel Bayesian method, developed by Reich et al (2004)[14], to perform the genome-wide association analyses of African-American MALD markers and ESRD case-control status. There are two general concepts to this analysis: (1) local estimates of ancestry state, that is, whether an individual has 0, 1, or 2 alleles from the African parental population at a given locus, are made along the genome are made by combining information from multiple, observable, closely linked MALD markers and (2) the ancestry state estimates are averaged across individuals to identify a higher-than-expected amount of ancestry from one parental population, thus indicating nearby disease gene. The estimates of ancestry state are obtained by the Markov Chain Monte Carlo (MCMC) method, which is an iterative process that randomly samples from the posterior likelihood distribution of the unknown parameters to obtain updated values for subsequent iterations and is implemented in the software, ANCESTRYMAP[14].

Two approaches will be used to test for association between ancestry state and presence of T2DM locus. Both statistics are based on the (values, ancestry estimates, from the MCMC. We will construct a "locus-genome statistic," which is a likelihood ratio statistic that compares the likelihood of the data if a disease locus is present [P(data|disease)] to the likelihood if no disease locus is present [P(data|no disease0]. We will also employ a "case-control" statistic, which directly compares cases with controls at every marker in the genome with respect to ancestry state estimate.

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

None

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 (2004.10)
___ 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.**

Reference List

1. Merikangas K, Chakravarti A, Moldin S et al. Future of genetics of mood disorders research. Biol Psychiatry 2002; 52(6):457.

2. Risch N, Merikangas K. The future of genetic studies of complex human diseases. Science 1996; 273(5281):1516-1517.

3. Chakraborty R, Weiss KM. Admixture as a tool for finding linked genes and detecting that difference from allelic association between loci. Proc Natl Acad Sci U S A 1988; 85(23):9119-9123.

4. Stephens JC, Briscoe D, OBrien SJ. Mapping by admixture linkage disequilibrium in human populations; limits and guidelines. Am J Hum Genet 1994; 55(4):809-824.

5. Stephens JC, Briscoe D, O'Brien SJ. Mapping by admixture linkage disequilibrium in human populations: limits and guidelines. Am J Hum Genet 1994; 55(4):809-824.

6. McKeigue PM. Mapping genes underlying ethnic differences in disease risk by linkage disequilibrium in recently admixed populations. Am J Hum Genet 1997; 60(1):188-196.

7. McKeigue PM, Carpenter JR, Parra EJ, Shriver MD. Estimation of admixture and detection of linkage in admixed populations by a Bayesian approach: application to African-American populations. Ann Hum Genet 2000; 64(Pt 2):171-186.

8. Zheng C, Elston RC. Multipoint linkage disequilibrium mapping with particular reference to the African-American population. Genet Epidemiol 1999; 17:79-101.

9. Lautenberger JA, Stephens JC, O'Brien SJ, Smith MW. Significant admixture linkage disequilibrium across 30 cM around the FY locus in African Americans. Am J Hum Genet 2000; 66:969-978.

10. Pfaff CL, Parra EJ, Ye J, Massac A, Kittles RA, Shriver MD. Using estimates of individual admixture to study the genetics of phenotypic traits: skin pigmentation in African Americans. American Journal of Human Genetics 2001; 69(4):410.

11. Smith MW, Patterson N, Lautenberger JA et al. A high-density admixture map for disease gene discovery in african americans. Am J Hum Genet 2004; 74(5):1001-1013.

12. Devlin B, Risch N. A note on Hardy-Weinberg equilibrium of VNTR data by using the Federal Bureau of Investigation's fixed-bin method. Am J Hum Genet 1992; 51(3):549-553.

13. Shriver MD, Parra EJ, Dios S et al. Skin pigmentation, biogeographical ancestry and admixture mapping. Hum Genet 2003; 112(4):387-399.

14. Patterson N, Hattangadi N, Lane B et al. Methods for high-density admixture mapping of disease genes. Am J Hum Genet 2004; 74(5):979-1000.