ARIC Manuscript Proposal # 1828

PC Reviewed: 8/9/11 Status: A Priority: 2
SC Reviewed: ________ Status: _____ Priority: ____

1.a. Full Title: Transnational Asthma Genetics Consortium (TAGC) – Asthma GWAS Meta-Analysis

b. Abbreviated Title (Length 26 characters): Asthma GWAS Meta-analysis (TAGC)

2. Writing Group:
Writing group members: Stephanie J. London, Laura Loehr, Grace Y Chiu (analyst)

TAGC has indicated that each cohort should include only two authors. Stephanie London is also participating with another study and thus we propose that the two unique ARIC authors will be Laura Loehr and Grace Chiu for the Atherosclerosis Risk in Communities Study.

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

First author: Stephanie J London
Address: NIEHS, PO Box 12233, MD A3-05, RTP NC 27709

Phone: 919-541-5772 Fax: 919-541-2511
E-mail: london2@niehs.nih.gov

ARIC author to be contacted if there are questions about the manuscript and the first author does not respond or cannot be located (this must be an ARIC investigator).

Name: Laura Loehr
Address: 137 E. Franklin Street, Ste 306, CB #8050, Chapel Hill, NC, 27514

Phone: (919) 966-8275 Fax: 919-966-9800
E-mail: lloehr@email.unc.edu

3. Timeline: Single cohort analyses August 2011, consortium meta-analysis Sept 2011, submission of manuscript to ARIC committee December 2011
4. **Rationale:** Asthma is a complex trait. Genome wide association studies have identified at a few replicated loci although only a few novel ones that had not been suggested by candidate gene studies. It is clear that large sample sizes are needed. The NHLBI has made a priority to encourage collaboration among investigators with both asthma phenotype and GWAS data. NHLBI has previously funded the EVE Consortium to encourage asthma GWAS meta-analyses and the first paper is in press (Torgerson et al. Nature Genetics, in press). The European GABRIEL consortium had previously published an asthma GWAS meta-analysis (Moffatt et al., NEJM 2010). NHLBI encourage EVE and GABRIEL to collaborate on a larger meta-analysis and also make an effort to include as many cohorts with asthma phenotype and GWAS data as possible. To this end a meeting was held in London in Sept 2010 and the Transnational Asthma Genetics Consortium (TAGC) was formed. It was decided to start with a meta-analysis of asthma and two leaders were decided upon – Dan Nicolae from EVE and Florence Demenais from GABRIEL. In late May 2011 they sent an analysis plan to all identified investigators with asthma and GWAS data and a conference call was held to discuss the plan in June 2011. The finalized analysis plan is attached to the end of this document.

Although the TAGC leadership had asked CHARGE to participate as a single meta-analyzed entity, it became clear that this would not be optimal for the CHARGE pulmonary group because for the asthma phenotype, FHS is part of another consortium lead by an investigator at the BROAD. So instead it was decided that individual CHARGE cohorts that wish to contribute will do so separately.

We propose to contribute as ARIC to this international asthma meta-analysis following the analysis plan agreed upon by TAGC.

5. **Main Hypothesis/Study Questions:**

Can novel loci for asthma be identified by doing GWAS meta-analysis with a larger sample size than has been previously attempted?

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).**

The plan was developed by TAGC and is a consensus plan. The definition of asthma is left to the cohorts. We propose to use doctor diagnosis of asthma in ARIC. Also, TAGC has requested subgroup analyses with age at onset prioritized. Only Caucasians will be included. Joel Hirschorn at the BROAD is planning to include CARe.

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 ICTDER03 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 ICTDER03 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 ICTDER03 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.cscn.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)?

11.a. Is this manuscript proposal associated with any ARIC ancillary studies or use any ancillary study data?  ____ Yes  ____ No

11.b. If yes, is the proposal  ____ A. primarily the result of an ancillary study (list number* _________)
____ 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.cscn.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.
Proposal for Information Requested to Be Shared
TAGC Consortium

General Overview. The following details the information that will be shared among the groups involved in TAGC. There will be a file (or set of files) for each homogeneous subgroup of each participating consortium (e.g. EVE African Americans). It is assumed that each group will contribute QC-ed data, including adjustments for population structure. It is also required that imputation is performed so summary data is shared on a common set of SNPs (proposal: HapMap Phase 2, Release 21).

1. The summary files for consortium GWASs will contain the following information:

General Information on SNPs (mostly used for checks on consistency):

1. rs number
2. Chromosome
3. Position (bp; build 36.3)
4. Alleles (in letters, forward strand, and in the order baseline allele, risk allele as used in analysis).

Information on genotyping (some will be used in filtering SNPs):

5. Genotyped (Yes/No)
6. SNP Call Rate
7. HWE p-value (if applicable)
8. Imputation R² (as defined in MaCH program) or other quality score
9. QC filter for genotyped SNPs (1=kept, 0=filtered, NA for imputed SNPs)
10. QC exclusion criterion for genotyped SNPs (needed for the SNPs with code 0 in previous column).

Information on association variables (needed for meta-analysis):

11. Controls allele frequency
12. Cases allele frequency
13. Test statistic – a Z-score showing direction of the effect
14. Association p-value
15. Odds Ratio (or another measure of effect size: regression coefficient…)
16. Odds Ratio standard error

2. Additional information for each study should be provided in a separate file (eg, Word file)
1. Study identification (2 variables: consortium name, study name)
2. PI (or PIs) of that study (names, institutions, e-mails)
3. Ethnic origin of population studied
4. Country of living of population studied (indicate several countries if necessary)
5. Type of samples (cases/controls, trios, nuclear families, pedigrees...)
6. Mode of ascertainment of studied sample
7. Sample size (no of cases/no of controls, no of trios,...)
8. Genotyping platform used (Illumina, Affymetrix; chips used)
9. QC criteria used for samples (call rate, IBS, PCs, heterozygosity on sex chromosomes, etc.)
10. QC criteria used for SNPs (HWE, call rate, Mendelian errors etc.)
11. Asthma definition used
12. Other phenotypic information available / used in analysis (2 answers per variable)
   - gender (Yes, No)
   - age at examination (Yes, No)
   - age of onset of asthma (Yes, No)
13. Imputation methodology: Program and reference population used
   (recommended: MACH / Hapmap2)
14. Statistical method used for analysis: logistic regression, TDT....
15. Subgroup analysis done (Yes / No) and if yes: definition of the subgroups

These files will be up-loaded by TAGC participants on an ftp site (that will be discussed at the Conference Call of May 24, 2011)

General Analysis Plan
- Description of the data available and information sent by each group
- Analysis of all samples using two main statistical approaches (see below)
- Sub-group analysis
  define sub-groups
  - by ethnicity
  - by geographic location
  - by age of onset of asthma
  - by gender...
  test heterogeneity between sub-groups

Statistical Methods

1) Combining Z scores (using appropriate weights)

2) Meta-analysis combining effect size estimates: random and fixed effects models
   + tests of heterogeneity across studies
The methodology behind the two approaches was discussed at the May 24 Conference Call by Dan (Z scores) and Florence (fixed effects/random effects meta-analysis)