1.a. Full Title: Regulators of *CHRNA5* Gene Expression and Their Association with Nicotine Dependence: A Systems Biology Approach

b. Abbreviated Title (Length 26 characters): CHRNA5 regulators

2. Writing Group:
   Writing group members: Dana B. Hancock, Robert F. Clark, Laura J. Bierut, Eric O. Johnson, Nora Franceschini, Eric Boerwinkle, David Couper, Rebecca Rohde, Gerardo Heiss, Kari E. North, and other interested ARIC and COGEND investigators

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

First author: Dana B. Hancock
Address: RTI International
3040 Cornwallis Road
PO Box 12194
Research Triangle Park, NC 27709

Phone: 919-541-6701 Fax: 919-485-5555
E-mail: dhancock@rti.org

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: Nora Franceschini
Address: 137 E. Franklin St., Suite 306
Campus Box 8050
Chapel Hill, NC 27514

Phone: 919-966-1305 Fax: 919-966-9800
E-mail: noraf@unc.edu

3. Timeline:
We will begin analyses soon after obtaining approval, and we aim to have analyses completed within six months (by August 2012). We plan to have the manuscript drafted, reviewed by co-authors, and submitted for publication within a year (by February 2012).
4. **Rationale:**
Nicotinic acetylcholine receptor (nAChR) gene upregulation has been reported in various *in vitro* and *in vivo* systems in the hippocampus and other brain areas after exposure to nicotine. Single nucleotide polymorphisms (SNPs) upstream of nAChR genes have been associated with nicotine dependence, and these SNPs may alter their expression. More specifically, SNPs in the α5 cholinergic nicotine receptor subunit gene (*CHRNA5*) on chromosome 15q25 have been identified and confirmed in genome-wide association studies of smoking phenotypes, including the study of nicotine dependence conducted in ARIC (by N.F. and E.B.) as part of the Tobacco and Genetics consortium [*Nat Genet* 2010; 42(5): 441-7]. Expression variation of *CHRNA5* mRNA levels has also been associated with nicotine dependence.

In mice, strain differences in nicotine consumption and sensitivity have been reported. C57BL/6J (B6) mice show a greater preference for nicotine consumption than DBA/2 (D2) mice, and quantitative trait loci (QTLs) that account for this nicotine difference map to several loci including nAChR genes. Polymorphisms in several nAChR genes have been shown to influence the effects of nicotine in mice.

Investigators from the Collaborative Genetic Study of Nicotine Dependence (COGEND) (R.F.C., L.J.B., and E.O.J.) used *in silico* systems biology tools to identify genes that are co-regulated with *CHRNA5* in mice and to evaluate the associations of nominated regulatory candidate genes with nicotine dependence in humans. In particular, they used expression QTL mapping to investigate genetic variation in expression levels of >39,000 transcripts from mouse hippocampus. Two related mouse panels of B6 x D2 were used to identify candidate regulatory genes that alter *CHRNA5* transcription in a normal physiological state.

Over 25 genes were implicated as cis-regulators and two genes were implicated as trans-regulators for *CHRNA5*. In COGEND, SNPs from the human homologues of the two trans-regulating genes had significant evidence for association with nicotine dependence in ~2,000 nicotine dependent cases and controls. These analyses implicate two novel candidate genes (*AMPH* and *PDE10A*) in nicotine addiction pathways that regulate hippocampal expression of *CHRNA5*, but the SNP coverage in these gene regions are limited in COGEND. Given that ARIC has more extensive coverage in these regions, we wish to test the candidate gene SNPs for association with nicotine dependence in the ARIC study, which will likely be used as the primary cohort for association testing (and not merely replication testing).

5. **Main Hypothesis/Study Questions:**
Do SNPs from *AMPH* and *PDE10A* have significant evidence for association with nicotine dependence in the ARIC study? Cigarettes per day will be used to measure nicotine dependence in ARIC.
Genotyped and imputed SNPs spanning the two candidate regulatory genes would be tested for association with nicotine dependence in a nested case-control study of white ARIC study participants.

Outcome variable
COGEND relied on the Fagerström Test for Nicotine Dependence (Heatherton et al., Br J Addict 1991;86:119-27) to distinguish dependent and non-dependent smokers. In addition to knowing cigarettes smoked per day, calculation of the Fagerström Test score requires questionnaire responses on smoking habits after awakening, refraining from smoking in forbidden public places, and smoking during illnesses. Scores on the following scale assess the intensity of physical addiction to nicotine with higher scores indicating more intense physical dependence: <4 indicate mild dependence, 4-6 indicate moderate dependence, and 7-10 indicate severe dependence. In COGEND, participants with a Fagerström score=0 were classified as controls, and those with a Fagerström score≥4 were classified as cases. Since questionnaire data are not available in ARIC to compute the Fagerström score, we would use cigarettes per day (collected at baseline) as a proxy for measuring nicotine dependence. Heavy smoking (>30 cigarettes per day) has moderate sensitivity (~60%) and high specificity (~95%) when compared to Fagerström scores ≥6 (de Leon et al., Addict Behav 2003;28:1481-6).

ARIC participants in the upper and lower quartiles of cigarettes per day (among ever smokers) would be classified as cases and controls, respectively, to maximize comparability with the COGEND case/control classification scheme. Among white ARIC participants, there are ~6,800 ever smokers, so we estimate that ~3,400 participants would be included in this analysis (~1,700 cases and ~1,700 controls). Several studies have demonstrated that such extreme sampling can increase power to detect genetic effects, despite the reduction in sample size due to excluding individuals with a moderate phenotype (Schork et al., Am J Hum Genet 2000; 67: 1208-1218).

Predictor variables
Age, gender, and ascertainment center would be included as covariates in models to test SNP associations with nicotine dependence. All predictor variables were collected at the baseline examination. To adjust for potential population stratification, the regression models will also include significantly associated principal component eigenvectors, selected from a base model containing the first 10 eigenvectors and the non-genetic variables (age, gender, and center) as predictors and nicotine dependence case/control status as the outcome.
Power analysis
Power calculations for the SNP association testing were conducted using QUANTO. In 1,700 cases and 1,700 controls, we assumed a log-additive genetic effect with odds ratios (ORs) varying from 1.1 to 1.5, varying minor allele frequencies from 10% and 50%, and \( \alpha = 0.001 \) (based on a Bonferroni correction for 49 tests corresponding to the number of haplotype blocks). The minimum detectable effect sizes to achieve 80% power ranged from OR=1.15 for the lowest frequency SNPs (10%) to OR=1.25 for higher frequency SNPs (>40%).

Data analysis
In the primary analysis, SNPs in or within 20 kb of \( AMPH \) and \( PDE10A \) would be tested for association with nicotine dependent case/control status. The logistic regression models would include additive SNP genotype, age, gender, center, and significantly associated principal components as predictors along with adjustment for known associations of two distinct mechanisms conferring risk for nicotine dependence within \( CHRNA5 \): altered receptor function caused by a D398N amino acid variant (rs16969968) and variability in mRNA expression (rs588765).

Given the large number of imputed SNPs that span the candidate genes in ARIC (>100 for each), a multiple testing correction would be needed. Using Bonferroni correction based on 49 tests from 16 haplotype blocks spanning \( AMPH \) and 33 haplotype blocks spanning \( PDE10A \), \( P < 0.001 \) would be considered statistically significant evidence for association.

As a secondary analysis, we would also assess these associations in linear regression models using the full range of cigarettes per day in all white ARIC ever smokers (N~6,800).

Limitation and Strengths
Some of the anticipated methodological limitations involve the lack of a completely harmonious outcome variable. The strengths of this proposal include the large sample size and thus statistical power for candidate gene association testing and the extensive experience in conducting candidate gene and genome-wide association studies by the COGEND and ARIC investigators.

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.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)?
MS 1382: Genome-wide association study of smoking initiation, intensity, and cessation in African American and white ARIC participants, and meta-analysis of smoking within the Tobacco & Genetics (TAG) Consortium

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* __________)
___ 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.