

ARIC Manuscript Proposal # 1382

PC Reviewed: 06/10/08
SC Reviewed: _____

Status: A
Status: _____

Priority: 2
Priority: _____

1.a. Full Title: 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.

b. Abbreviated Title (Length 26 characters): Smoking GWAS in ARIC

2. Writing Group: Helena Furberg, Nora Franceschini, Thomas Payne, Eric Boerwinkle.

Other ARIC investigators are welcome.

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

First author: Helena Furberg, PhD

Assistant Professor
Department of Genetics, School of Medicine
University of North Carolina Chapel Hill
CB#7264, 115 Mason Farm Road
Medical Biomolecular Research Building, 4th floor
Chapel Hill, NC, 27599-7264, USA
v: 919-966-1764
f: 919-966-3630

Corresponding/senior author (must be an ARIC investigator for the proposal but can be different in the published paper; correspondence will be sent to both the first author & the corresponding author):

Senior author: Eric Boerwinkle, PhD

Address: Professor, and Center and Division Director
Human Genetics Center
University of Texas Health Science Center at Houston
PO Box 20334
Houston, Texas 77225-0334
(713) 500-9816

3. Timeline: 1-2 years for the ARIC GWAS and 1-3 years for meta-analyses

4. Rationale:

Epidemiology of Smoking. To reduce the impact of tobacco use on health, tobacco users have been urged to quit for decades. These efforts have been at least partially successful: smoking rates have declined dramatically in the US in the past 40 years, but have recently stalled among adults and adolescents (1). The majority of smokers are aware of deleterious health effects of smoking and ~70% of smokers want to quit (2). However, while over 46% make an active attempt to quit each year, only 2.3% achieve sustained abstinence (3). The astonishing addictive capacity of nicotine, manifest in the behavioral and cognitive constellation of signs and symptoms termed nicotine dependence (ND) (4), is a prime factor in thwarting attempts at cessation (5).

ND is considered a psychoactive substance use disorder by the American Psychiatric Association and is regarded as a multidimensional construct (4, 6). The most commonly used measure to assess ND is the Fagerström Test for Nicotine Dependence (FTND) which consists of six questions that capture physiologic dependence on nicotine (range 0-10) (7, 8). Smokers with higher FTND scores are considered more nicotine dependent (7) and less likely to quit smoking. Results from large, population surveys estimate the prevalence of ND among smokers is ~50% (9-11).

The Genetic Basis of Smoking. There is strong and consistent evidence from twin, family and adoption studies that genetic factors account for a sizeable portion of individual differences in smoking. Fisher was the first to report that concordance for smoking was higher among MZ than DZ twins (12, 13). These findings have been confirmed by numerous investigations in different populations (14-26). The heritability estimates for smoking initiation, FTND score, and smoking cessation average range from 0.50-0.80 (27, 28). Of the variables that comprise the FTND score, number of cigarettes smoked per day, and time to first cigarette upon waking exhibit the highest heritability estimates. Although genome-wide linkage studies (29-32) have identified some genomic regions that predispose or protect against ND, the data are not consistent. Similarly, candidate gene association studies of smoking have not been successful in identifying replicable associations (33-35) and are dependent upon our limited knowledge of addiction biology. To identify unsuspected genetic variants of relevance to ND, an agnostic genomic search is essential. The first published GWAS of ND (36) unfortunately used DNA pooling, however the most recent GWAS for ND (37) and two others for lung cancer (38, 39) identified the same region on chromosome 15 which contains several nicotinic acetylcholine receptor genes.

Era of GWAS. Recent technological advances and recognition that we need to engage in “aggressive data sharing”(40) provide an unprecedented opportunity to assess how genetics contribute to smoking. Tens of billions of genotypes are being generated from GWAS for many different diseases and smoking-related measures are routinely collected. This presents an incredible opportunity to analyze very large GWAS for smoking behavior. Large collaborative consortia are powerful and efficient ways to address

important research questions (41) and have already produced robust and replicated genetic associations for diabetes (42), and several cancers (43, 44). Furthermore, *secondary* analysis of GWAS for additional heritable complex traits is a sound and successful approach, as evidenced by recent findings for height and body mass index (BMI). Two primary T2DM studies meta-analyzed GWAS data (4,921 subjects) for height (45) found that a common variant in *HMGA2* was associated with childhood and adult height and was confirmed in replication samples. Similarly, Frayling et al. found an association between SNPs in *FTO* and BMI which replicated in >38,000 individuals from 13 samples (46).

We propose to analyze the ARIC smoking GWAS data independently and then incorporate the ARIC findings in a meta-analysis with ~20 other studies within the Tobacco & Genetics (TAG) Consortium (Ancillary ARIC Study approval # 2008.04; PIs N Franceschini & H Furberg).

5. Main Hypothesis/Study Questions:

1. We will examine whether there are genetic variants associated with three aspects of smoking behavior (smoking initiation, smoking intensity, and smoking cessation), among ARIC participants of Northern European ancestry and African ancestry.
2. We will examine whether there are genetic variants associated with smoking behavior across multiple ancestry samples in the context of a GWAS meta-analysis that includes ARIC and ~20 other studies within the TAG Consortium.

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).

Study Population: 4,266 African American and 11,478 Caucasian ARIC participants with available smoking phenotype data and genomewide marker data who consented to future DNA use. The smoking phenotype variables were derived from participant responses during visits 1-4.

For the TAG Consortium meta-analyses, ARIC data will be integrated with data from ~20 other GWAS studies, for a total approaching n=100,000 participants (see details in the Ancillary Study – Franceschini & Furberg PIs).

Study design: Cross-sectional analysis stratified by ancestry. Traditional ancestry-stratified meta-analyses for the TAG Consortium.

Exposure: Whole genome genotyping and imputation for the ARIC study has been performed by the ARIC Central Laboratory using Taqman[®] genotyping assays under direction of Dr Eric Boerwinkle. Imputation will result in 2.55 million SNPs for each ARIC participant, which produces a common set of genetic markers for all participants. For the TAG Consortium, imputed genotypes from each study enable us to analyze all studies together regardless of genotyping platform used in the initial study.

Outcome: In ancestry-stratified analyses, we will investigate three aspects of smoking behaviors as six separate dependent variables:

1. Smoking Initiation (Ever/Never Smoker & Age at onset of Smoking)
2. Heaviness of Smoking (Average number of Cigarettes Smoked/Day, Total number of Years Smoked)
3. Smoking Cessation (Former/Current Smoker & Age quit Smoking)

Statistical analyses: We will fit logistic regression models in ancestry-stratified samples using PLINK. The Cochran-Armitage trend test will produce genetic relative risks (GRR) describing the association between each genotype and the three aspects of smoking behavior separately under a log-additive model. A conservative Bonferroni correction(47) will account for multiple testing. All analyses will be adjusted for the effects of age, sex, and study center, within each ancestry-stratified population sample. We will also consider the effects of variables which affect the likelihood of smoking, depending on the prevalence of these factors in these data.

The TAG meta-analysis will be performed on the Genetic Computing Cluster at vrije University in Amsterdam (www.geneticcomputingcluster.org) using standard meta-analysis methods (48, 49). The I^2 statistic will be used to estimate between study heterogeneity. The Cochran-Armitage trend test will produce GRR describing the association between each genotype and the three aspects of smoking behavior under a log-additive model. A conservative Bonferroni correction (47) will account for multiple testing. As the smoking phenotypes were collected in studies of other traits, we will account for potential confounding introduced if a SNP is associated with smoking *and* the primary phenotype.

Limitations: Smoking phenotype information was collected as a covariate in the ARIC study and does not include direct assessment of nicotine dependence by the FTND scale. However, the variable ‘average number of cigarettes smoked per day’ is available, highly heritable and regarded as a reasonable approximation of ND. Although the sample size of ARIC is very large, it may not be large enough to detect the modest effect sizes expected in a genome-wide study of smoking behavior. Therefore, we intend to replicate findings from the initial ARIC analyses into the TAG Consortium meta-analysis (n=100,000) where we will have outstanding statistical power: for minor allele frequencies of 0.1, 0.2, 0.3, 0.4, and 0.5, there will be 80% power to detect genetic relative risks of 1.14, 1.105, 1.09, 1.085, and 1.082.

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

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? 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”?

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.csc.unc.edu/ARIC/search.php>

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)?

Manuscript proposal #1045 “Gene-by-smoking, subclinical atherosclerosis, incident CHD and stroke in ARIC” North (PI)

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* 2006-12)**
 B. primarily based on ARIC data with ancillary data playing a minor role (usually control variables; list number(s)* _____)

Ancillary ARIC Study approval # 2008.04; PIs N Franceschini & H Furberg.

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

1. CDC. Cigarette smoking among adults-United States, 2006. MMWR CDC Surveillance Summaries 2007;56:1157-1161.
2. Cigarette Smoking among adults-United States, 1993. MMWR. Morbidity and Mortality Weekly Report 1994;43:925-930.
3. Prevalence of cigarette use among 14 racial/ethnic populations--United States, 1999-2001. MMWR. Morbidity and Mortality Weekly Report 2004;53(3):49-52.
4. Association AP. Diagnostic and Statistical Manual of Mental Disorders. Fourth Edition ed. Washington, DC; 1994.

5. Hogg RC, Bertrand D. Neuroscience. What genes tell us about nicotine addiction. *Science* 2004;306(5698):983-5.
6. Hudmon KS, Marks JL, Pomerleau CS, Bolt DM, Brigham J, Swan GE. A multidimensional model for characterizing tobacco dependence. *Nicotine Tob Res* 2003;5(5):655-64.
7. Heatherton TF, Kozlowski LT, Frecker RC, Fagerstrom KO. The Fagerstrom Test for Nicotine Dependence: a revision of the Fagerstrom Tolerance Questionnaire. *British Journal of Addiction* 1991;86(9):1119-27.
8. Fagerstrom K-O. Measuring degree of physical dependence to tobacco smoking with reference to individualization of treatment. *Addictive Behaviors* 1978;3:235-241.
9. Breslau N, Johnson EO, Hiripi E, Kessler R. Nicotine dependence in the United States: prevalence, trends, and smoking persistence. *Archives of General Psychiatry* 2001;58(9):810-6.
10. Hoch E, Muehlig S, Hofler M, Lieb R, Wittchen HU. How prevalent is smoking and nicotine dependence in primary care in Germany? *Addiction* 2004;99(12):1586-98.
11. Fagerstrom KO, Kunze M, Schoberberger R, Breslau N, Hughes JR, Hurt RD, et al. Nicotine dependence versus smoking prevalence: comparisons among countries and categories of smokers. *Tobacco Control* 1996;5(1):52-6.
12. Fisher RA. Cancer and smoking. *Nature* 1958;182(4635):596.
13. Fisher RA. Lung cancer and cigarettes. *Nature* 1958;182(4628):108.
14. Sullivan PF, Kendler KS. The genetic epidemiology of smoking. *Nicotine Tob Res* 1999;1 Suppl 2:S51-7; discussion S69-70.
15. Li MD, Cheng R, Ma JZ, Swan GE. A meta-analysis of estimated genetic and environmental effects on smoking behavior in male and female adult twins. *Addiction* 2003;98(1):23-31.
16. Maes HH, Woodard CE, Murrelle L, Meyer JM, Silberg JL, Hewitt JK, et al. Tobacco, alcohol and drug use in eight- to sixteen-year-old twins: the Virginia Twin Study of Adolescent Behavioral Development. *Journal of Studies on Alcohol* 1999;60(3):293-305.
17. Heath AC. Persist or quit? Testing for a genetic contribution to smoking persistence. *Acta Geneticae Medicae et Gemellologiae* 1990;39(4):447-58.
18. Raaschou-Nielsen E. Smoking habits in twins. *Danish Medical Bulletin* 1960;7:82-8.
19. Crumpacker DW, Cederlof R, Friberg L, Kimberling WJ, Sorensen S, Vandenberg SG, et al. A twin methodology for the study of genetic and environmental control of variation in human smoking behavior. *Acta Geneticae Medicae et Gemellologiae* 1979;28(3):173-95.
20. Eaves LJ, Silberg JL, Hewitt JK, Rutter M, Meyer JM, Neale MC, et al. Analyzing twin resemblance in multisymptom data: genetic applications of a latent class model for symptoms of conduct disorder in juvenile boys. *Behavior Genetics* 1993;23(1):5-19.
21. Carmelli D, Swan GE, Robinette D, Fabsitz R. Genetic influence on smoking--a study of male twins. *New England Journal of Medicine* 1992;327(12):829-33.
22. Kaprio J, Koskenvuo M, Langinvainio H. Finnish twins reared apart. IV: Smoking and drinking habits. A preliminary analysis of the effect of heredity and environment. *Acta Geneticae Medicae et Gemellologiae* 1984;33(3):425-33.
23. Edwards KL, Austin MA, Jarvik GP. Evidence for genetic influences on smoking in adult women twins. *Clinical Genetics* 1995;47(5):236-44.
24. Hannah MC, Hopper JL, Mathews JD. Twin concordance for a binary trait. II. Nested analysis of ever-smoking and ex-smoking traits and unnested analysis of a "committed-smoking" trait. *American Journal of Human Genetics* 1985;37(1):153-65.
25. Heath AC, Cates R, Martin NG, Meyer J, Hewitt JK, Neale MC, et al. Genetic contribution to risk of smoking initiation: comparisons across birth cohorts and across cultures. *Journal of Substance Abuse* 1993;5(3):221-46.
26. Vink JM, Willemsen G, Boomsma DI. Heritability of smoking initiation and nicotine dependence. *Behavior Genetics* 2005;35(4):397-406.
27. Lessov CN, Martin NG, Statham DJ, Todorov AA, Slutske WS, Bucholz KK, et al. Defining nicotine dependence for genetic research: evidence from Australian twins. *Psychological Medicine* 2004;34(5):865-79.
28. Broms U, Silventoinen K, Madden PA, Heath AC, Kaprio J. Genetic architecture of smoking behavior: a study of Finnish adult twins. *Twin Res Hum Genet* 2006;9(1):64-72.
29. Ott J. *Analysis of Human Genetic Linkage*. Baltimore, MD: The Johns Hopkins University Press; 1999.

30. Straub RE, Sullivan PF, Ma Y, Myakishev MV, Harris-Kerr C, Wormley B, et al. Susceptibility genes for nicotine dependence: a genome scan and followup in an independent sample suggest that regions on chromosomes 2, 4, 10, 16, 17 and 18 merit further study. *Molecular Psychiatry* 1999;4(2):129-44.
31. Sullivan PF, Neale BM, van den Oord E, Miles MF, Neale MC, Bulik CM, et al. Candidate genes for nicotine dependence via linkage, epistasis, and bioinformatics. *Am J Med Genet B Neuropsychiatr Genet* 2004;126(1):23-36.
32. Sullivan PF, Kuo P-H, Webb BT, Neale MC, Vittum J, Furberg H, et al. Genomewide linkage survey of nicotine dependence phenotypes. *Drug and Alcohol Dependence* in press.
33. Hirschhorn JN, Lohmueller K, Byrne E, Hirschhorn K. A comprehensive review of genetic association studies. *Genet Med* 2002;4(2):45-61.
34. Lohmueller KE, Pearce CL, Pike M, Lander ES, Hirschhorn JN. Meta-analysis of genetic association studies supports a contribution of common variants to susceptibility to common disease. *Nature Genetics* 2003;33(2):177-82.
35. Munafò M, Clark T, Johnstone E, Murphy M, Walton R. The genetic basis for smoking behavior: a systematic review and meta-analysis. *Nicotine Tob Res* 2004;6(4):583-97.
36. Bierut LJ, Madden PA, Breslau N, Johnson EO, Hatsukami D, Pomerleau OF, et al. Novel genes identified in a high-density genome wide association study for nicotine dependence. *Human Molecular Genetics* 2007;16(1):24-35.
37. Thorgeirsson TE, Geller F, Sulem P, Rafnar T, Wiste A, Magnusson KP, et al. A variant associated with nicotine dependence, lung cancer and peripheral arterial disease. *Nature* 2008;452(7187):638-42.
38. Amos CI, Wu X, Broderick P, Gorlov IP, Gu J, Eisen T, et al. Genome-wide association scan of tag SNPs identifies a susceptibility locus for lung cancer at 15q25.1. *Nature Genetics* 2008;40(5):616-22.
39. Hung RJ, McKay JD, Gaborieau V, Boffetta P, Hashibe M, Zaridze D, et al. A susceptibility locus for lung cancer maps to nicotinic acetylcholine receptor subunit genes on 15q25. *Nature* 2008;452(7187):633-7.
40. Zeggini E, Weedon MN, Lindgren CM, Frayling TM, Elliott KS, Lango H, et al. Replication of genome-wide association signals in UK samples reveals risk loci for type 2 diabetes. *Science* 2007;316(5829):1336-41.
41. Seminara D, Khoury MJ, O'Brien TR, Manolio T, Gwinn ML, Little J, et al. The emergence of networks in human genome epidemiology: challenges and opportunities. *Epidemiology* 2007;18(1):1-8.
42. Frayling TM. Genome-wide association studies provide new insights into type 2 diabetes aetiology. *Nat Rev Genet* 2007;8(9):657-62.
43. Easton DF, Pooley KA, Dunning AM, Pharoah PD, Thompson D, Ballinger DG, et al. Genome-wide association study identifies novel breast cancer susceptibility loci. *Nature* 2007;447(7148):1087-93.
44. Yeager M, Orr N, Hayes RB, Jacobs KB, Kraft P, Wacholder S, et al. Genome-wide association study of prostate cancer identifies a second risk locus at 8q24. *Nature Genetics* 2007;39(5):645-9.
45. Weedon MN, Lettre G, Freathy RM, Lindgren CM, Voight BF, Perry JR, et al. A common variant of HMGA2 is associated with adult and childhood height in the general population. *Nature Genetics* 2007;39(10):1245-1250.
46. Frayling TM, Timpson NJ, Weedon MN, Zeggini E, Freathy RM, Lindgren CM, et al. A common variant in the FTO gene is associated with body mass index and predisposes to childhood and adult obesity. *Science* 2007;316(5826):889-94.
47. Bonferroni CE, 3--62. Teoria statistica delle classi e calcolo delle probabilit `a. Pubblicazioni del R Istituto Superiore di Scienze Economiche e Commerciali di Firenze 1936;8:3-62.
48. Whitehead A. Meta-analysis of controlled clinical trials. Hoboken: John Wiley & Sons; 2002.
49. Hedges LV, Olkin I. Statistical Methods for Meta-Analysis. New York: Academic Press; 1985.