Candidate Gene Association Resource (CARe) Project
Database Application

NHLBI CARe Project Number _____________________
Submission Date ___October 9, 2008_________________

All proposals must be submitted by email to the CARe DAC COORDINATOR at care-
dac@nhlbi.nih.gov

All sections of this application must be completed. Incomplete applications will be returned.

Sections I-IV should be no more than 3 pages in length, not including references.

I. INVESTIGATOR INFORMATION:

Title of Proposed Project: Candidate gene association study of sleep apnea and snoring: The CARE Study

Name of proposed *Lead Author: Sanjay R. Patel
Email Address: srp20@case.edu
Telephone Number: 216-844-6258
Fax Number: 216-844-6265
Mailing Address: 11400 Euclid Avenue, Room 290-D, Cleveland, OH 44106

Institution/Company of proposed Lead Author: Case Western Reserve University

Name of proposed *Senior Author: Susan Redline
Email Address: susan.redline@gmail.com
Telephone Number: 216-844-6262
Fax Number: 216-844-2580
Mailing Address: 11400 Euclid Avenue, Room 290-D, Cleveland, OH 44106

Institution/Company of proposed Senior Author: Case Western Reserve University

*Author sections may be repeated as needed for starred Lead or Senior Authors.

Name of Lead Analyst / Statistician: Emma K. Larkin
Email Address: ekl3@case.edu
Telephone Number: 216-844-6257
Fax Number: 216-844-2580
Mailing Address: 11400 Euclid Avenue, Room 260, Cleveland, OH 44106

Names (institutions) of all other participating CARe Investigators:

Daniel Gottlieb, Boston University
Diane Lauderdale, University of Chicago
Naresh Punjabi, Johns Hopkins University
II. SCIENTIFIC RATIONALE (~250 WORDS)

A. Please provide an abstract describing the rationale and design of the proposed research project. The abstract must include major hypotheses, an outline of the research methods and analytical approach, and phenotypes to be studied. It should state clearly the objectives of the proposed project and provide the background rationale that would justify them. It should also address why the CARe database is appropriate for answering the research question. (Abstracts for approved projects will be posted, with the name(s) of the proposed Lead Author(s), on the NHLBI/NCBI CARe website.)

Background
With the increasing prevalence of obesity, obstructive sleep apnea (OSA) has emerged as an important public health issue. The sleep disruption occurring with OSA leads to daytime sleepiness, reduced quality of life and increased risk for motor vehicle accidents.(1) More recently, OSA has been identified as a risk factor for hypertension, heart failure, stroke, and mortality.(2-4) A familial basis for OSA has been well established,(5-7) and several studies have demonstrated that this familial basis exists independent of obesity. (7-8) Three whole genome linkage scans have been performed identifying chromosomal regions of interest,(9-11) and multiple small case control studies have identified candidate genes.(12-13) However, no susceptibility gene for OSA has yet been definitively identified. The CARe project provides a mechanism to perform the first large association study for OSA-related traits. Detailed sleep apnea phenotyping with overnight polysomnography has been performed in all participants of two of the cohorts – the Cleveland Family Study and Sleep Heart Health Study. The same reading center (Case Western) was used for both studies minimizing issues of harmonization of data for such a combined analysis. In addition, information about snoring status, a good surrogate for OSA, is available in nearly all of the cohorts.

Analytic Plan
The primary phenotypes to be assessed will be the apnea hypopnea index (AHI) as a continuous variable and OSA (defined as an AHI ≥ 30) as a dichotomous variable for analysis of the CFS and SHHS cohorts. A second analysis using the broader cohort data will use loud snoring as a dichotomous outcome variable. Linear and logistic regression modeling will be performed testing for an association between AHI and OSA or loud snoring respectively with each individual SNP on the iSelect Candidate Gene Panel assuming both an additive penetrance model and under a generalized model (2 df). Covariates in the primary model will include age, sex, and menopausal status and standardized residuals will be generated adjusting for these covariates within each cohort for each race. Joint association analysis across cohorts will be performed using inverse variance weights. Secondary analyses will adjust for alcohol and tobacco use. Additional analyses will be conducted assessing the effect on outcomes of adjusting for adiposity measures including BMI, waist circumference, and neck circumference.

References

III. PRIOR EXPERIENCE OF THE PI AND ASSOCIATES (~250 WORDS)
Please describe the experience and expertise of your team to complete the research project.

Dr. Patel is a sleep specialist and epidemiologist who has worked with the Cleveland Family Study since 2002. His research has focused on the genetic epidemiology of sleep disorders. He has a Career Development Award (K08) to study the genetic basis of sleep apnea consequences.

Drs. Redline and Gottlieb have published extensively on the genetics of sleep apnea and other sleep phenotypes. Dr. Redline performed the first genome wide linkage analysis of sleep apnea and Dr. Gottlieb performed the first GWAS of sleep phenotypes including sleep apnea.

Drs. Redline, Gottlieb, Pack, Lauderdale, and Punjabi are all established researchers in the field of sleep disorders.

Drs. Larkin and Buxbaum are genetic epidemiologists who have experience studying the genetics of sleep apnea. Both have been involved with performing linkage analyses with the Cleveland Family Study.

IV. DATA REQUESTED FOR THE PROPOSED ANALYSES (Provide rationale for any requested data whose relevance to these analyses is not obvious):
Genotype data (check all that apply):

[x] CARe IBC Candidate Gene array
[ ] Affymetrix 6.0
[ ] Both IBC array and Affymetrix 6.0
[ ] Other (please specify; e.g. SNPs within candidate genes, specific regions, etc.)

Specification of “other” genotype data: _______________________________________

Phenotype data:

Sleep apnea severity (from polysomnography), snoring status
Age, sex, race, body mass index, waist circumference, neck circumference, menopausal status
Alcohol and tobacco consumption

V. *ORGANIZATION APPLYING FOR DATA ACCESS:

Name of Applicant Organization: Case Western Reserve University

Address of Applicant Organization: 10900 Euclid Avenue, Cleveland, OH 44106

Name and Title of Applicant Organization’s Authorized Institutional Business Official:
Cynthia O. Case, Director of Grants and Contracts

Name of Each Institution/Company whose investigators will receive access to the requested data:
Case Western Reserve University

I __Sanjay R. Patel_________, Data Recipient for the proposed project, request the following data:

*Please provide information for each Applicant Organization and Data Recipient to whom data are to be released

a. [x] Yes [ ] No: Genotype Data
b. [x] Yes [ ] No: Phenotype Data

Please answer the following questions:

a. [ ] Yes [x] No: This research using the CARe database may be used for development of a commercial product or for commercial purposes.

b. [ ] Yes [x] No: Data accessed through this application will be used by or shared with individuals from a for-profit company.
NOTE: Some participants in the CARe Study have not provided consent for use of their CARe data for development of a commercial product or to be accessed by a for-profit entity. For each approved project, a dataset will be generated by a computer algorithm that incorporates the consent options of each of the CARe Cohorts, such that data access will only be provided as specified by the informed consent document of each study participant.

VI. ANNUAL REPORTING OF RESULTS

Continued access to the CARe database after one year will only be permitted if an annual report is submitted to the DAC Coordinator (care-dac@nhlbi.nih.gov) that describes the product of your research using the CARe database, and includes a listing of presentations and publications resulting from that research. Please complete the following:

[x] Yes  [ ] No: I understand that use of the CARe dataset includes the timely completion of an annual report.

Complete results of CARe analyses are being web-posted by the National Center for Biotechnology Information (NCBI). Please complete the following:

[x] Yes  [ ] No: I am willing to provide my completed analysis results for such web-posting.

VII. ADDITIONAL DOCUMENTS:

Return by email to the DAC Coordinator: care-dac@nhlbi.nih.gov
(Signed or other non-electronic documents can be faxed.)

Please include:

(1) Application Form

(2) Supporting documentation, including current human studies training certification for all key personnel*

(3) IRB Approval (Note: Full board or expedited review is required; “exempt” status is not acceptable.)*

(4) Data Distribution Agreement*

*Can be submitted after DAC Committee approval

Applications can be submitted at any time. Applicants will be notified of the DAC Committee decision within four weeks.

Upon approval, a Data Recipient at each institution that will receive data will be required to submit a signed Data Distribution Agreement to the CARe DAC Coordinator. The CARe Data Distribution Agreement can be obtained from Website to be added. The Data Recipient(s) will be responsible for obtaining signatures on behalf of the Recipient Entities and returning the signed Data Distribution Agreement to the CARe DAC Coordinator before access can be provided to the CARe database.
VII. ADDITIONAL INFORMATION (Required only for “CARe Investigators”)

Name and CARe Cohort Affiliation(s) of proposed Lead Author(s): Sanjay Patel (CFS)

Names and CARe Cohort Affiliation(s) of proposed Senior Author(s): Allan Pack (U.Penn)

Table of CARe Cohort Representation:

<table>
<thead>
<tr>
<th>Cohort</th>
<th>Investigator(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ARIC</td>
<td>Naresh Punjabi</td>
</tr>
<tr>
<td>BROAD</td>
<td>Deb Farlow</td>
</tr>
<tr>
<td>CSSCD</td>
<td></td>
</tr>
<tr>
<td>CARDIA</td>
<td>Diane Lauderdale</td>
</tr>
<tr>
<td>CFS</td>
<td>Sanjay Patel, Emma Larkin, Susan Redline</td>
</tr>
<tr>
<td>CHS</td>
<td>Naresh Punjabi</td>
</tr>
<tr>
<td>FHS</td>
<td>Dan Gottlieb</td>
</tr>
<tr>
<td>JHS</td>
<td>Tibor Fulop, Sarah Buxbaum</td>
</tr>
<tr>
<td>MESA</td>
<td>Brandon Lu</td>
</tr>
<tr>
<td>SHHS</td>
<td>Susan Redline, Naresh Punjabi, Dan Gottlieb</td>
</tr>
<tr>
<td>U.PENN</td>
<td>Allan Pack</td>
</tr>
</tbody>
</table>

Reason(s) for CARe cohorts not represented:

Attempts have been made to include representatives from CSSCD but no interested individuals have been identified.

[x] I, _Sanjay R. Patel____, affirm that this proposal has been reviewed and approved by the _Sleep__________________ working group and by all listed investigators. I further affirm that the working group convener, _Susan Redline_______, concurs in this statement.
Association of genomic loci with sleep apnea in European Americans and African-Americans: The Candidate Gene Association Resource (CARe)

Sanjay R. Patel¹, Robert Goodloe, Gourab De, Matthew Kowgier, Emma K Larkin², Taylor Young, Sarah Buxbaum, Tibor Fulop³, Sina A. Gharib⁴, Daniel J. Gottlieb, Craig Johnson, Diane S. Lauderdale⁵, Naresh M. Punjabi, Phyllis Zee, Brian Cade, Nan Laird, Sutapa Mukherjee⁶-⁸, Lyle Palmer, Xiaofeng Zhu¹⁰, Susan Redline¹

¹Division of Sleep Medicine, Brigham and Women’s Hospital, Boston MA; ²Division of Allergy, Pulmonary and Critical Care, Department of Medicine, Vanderbilt University Medical Center, Nashville TN; ³Department of Medicine, University of Mississippi Medical Center, Jackson MS; ⁴Center for Lung Biology and University of Washington Medicine Sleep Center, Department of Medicine, University of Washington, Seattle WA; ⁵Department of Health Studies, University of Chicago, Chicago IL; ⁶Department of Medicine, University of Toronto, Toronto ON; ⁷Women’s College Research Institute, Women’s College Hospital, Toronto ON; ⁸Western Australia Sleep Disorders Research Institute, Perth Australia; ¹⁰Department of Epidemiology and Biostatistics, School of Medicine, Case Western Reserve University, Cleveland OH

Word Counts: 3580 text; 238 abstract

Figure/Table Count: 4 figures, 4 supplemental figures, 3 tables

Reference Count: 30

Corresponding Author:
Sanjay R. Patel MD, MS
Division of Sleep Medicine
Brigham and Women’s Hospital
221 Longwood Avenue, Room 225-C
Boston, MA 02115
Email Address: spatel@partners.org
Telephone: 857-307-0347
Fax: 617-278-6946
ABSTRACT

Introduction: Although obstructive sleep apnea (OSA) is known to have a strong familial basis, no genetic polymorphisms influencing apnea risk have been identified in cross-cohort analyses. We utilized the National Heart, Lung, and Blood Institute (NHLBI) Candidate Gene Association Resource (CARe) to identify sleep apnea susceptibility loci.

Methods: Using a panel of 46,449 polymorphisms from roughly 2,100 candidate genes on a customized Illumina iSelect chip, we tested for association with the apnea hypopnea index (AHI) as well as moderate to severe OSA (AHI ≥ 15) in 3,551 participants of the Cleveland Family Study and three cohorts within the Sleep Heart Health Study. Ethnicity and cohort-specific analyses were performed adjusting for age, sex and body mass index as well as the first 10 principal components to account for population stratification. Meta-analysis was performed to summarize results across cohorts.

Results: Among 2,904 individuals of European ancestry, 1 SNP in the prostaglandin E2 receptor, PTGER3, was associated with OSA at a chip-wide significance level of p-value < 2x10^{-6}. Among 647 African-Americans, 1 SNP in pleckstrin (PLEK) was associated with OSA while 1 SNP in the lysophosphatidic acid receptor 1 (LPAR1) gene was associated with AHI. Consistency of effects between the SNP in PTGER3 and AHI was observed in an independent clinic-based cohort.

Conclusion: Novel genetic loci for OSA were identified through the use of customized gene chips and meta-analyses of cohort data. This approach is also conducive for future replication studies.
INTRODUCTION

Obstructive sleep apnea (OSA) is a common disorder characterized by collapse of the upper airway during sleep leading to recurrent arousals, intermittent hypoxia, and surges in sympathetic activation. By narrowing the upper airway lumen, obesity is one of the strongest risk factors for OSA.[1] Independent of obesity, OSA has been implicated as an independent risk factor in the development of insulin resistance, hypertension and cardiovascular disease.[2-4] In addition, the fragmentation of sleep caused by OSA increases the risk of excessive daytime sleepiness and as a result, risk of motor vehicle accidents.[5, 6] In total, OSA has been associated with substantial increases in health care spending.[7]

Numerous studies have established that OSA aggregates within families suggesting the presence of a genetic predisposition.[8-10] Those with one affected relative are approximately 50% more likely to have OSA themselves.[8] Though obesity itself has a strong genetic basis, the familial aggregation of OSA persists even after accounting for obesity.[11] This may be due to the role of craniofacial morphology, ventilatory drive or other heritable traits important in OSA pathogenesis. The overall health impact of OSA and limited treatment options currently available for this disease underscore the need to better understand its molecular basis.

While prior genetic studies of OSA have suggested novel susceptibility loci, these studies have been relatively limited in terms of sample size. In this work, we sought to combine data from two of the largest sleep apnea epidemiologic cohorts, the Cleveland Family Study and the Sleep Heart Health Study, to identify genetic variants that predict OSA risk in both European ancestry and African-American populations.

These two cohorts participated in the NHLBI sponsored Candidate gene Association Resource (CARe) project wherein a custom candidate gene array carrying single nucleotide
polymorphisms (SNPs) for over 2000 candidate regions relevant to heart, lung, blood, and sleep disorders was genotyped to provide the ability to combine information from multiple cohorts in the search for genetic variants that influence risk for common diseases.

MATERIALS AND METHODS

The CARe Consortium is a NHLBI supported resource for analyses of the association of genotypes with heart, lung, blood, and sleep phenotypes. CARe is composed of 9 large cohort studies of which four have collected polysomnographic data in at least a subset. These are the Cleveland Family Study (CFS), and subsets of three cohorts – the Atherosclerosis Risk in Communities (ARIC) study, the Cardiovascular Health Study (CHS), and the Framingham Heart Study (FHS) – which participated in the Sleep Heart Health Study (SHHS). Because the OSA phenotype changes across the lifespan with differing risk factors in pediatric, middle-aged, and elderly populations, we focused on the middle-aged. Thus, children (age < 18 yrs) from the CFS cohort and the CHS cohort where the mean age at the time of OSA phenotyping was 72 yrs were excluded from the primary analyses.

CFS is a family-based longitudinal cohort study designed to study the genetic basis of OSA.[8] Index probands with a laboratory confirmed diagnosis of OSA, and at least two first-degree relatives available to be studied were recruited along with family members. Initially, neighborhood controls and their relatives were also recruited. Overall, a total of 2284 individuals from 361 families were recruited from the Cleveland metropolitan area. DNA was available in a subset of 1665 individuals for CARe genotyping. Phenotype data was used from the last available exam.

SHHS is a prospective cohort of 6,441 subjects who underwent polysomnography recruited from six established cohorts to study the effects of OSA on cardiovascular disease.[12] Three of the
six parent cohorts contributing to SHHS, ARIC, FHS, and CHS participated in CARe and so contributed genetic material. SHHS recruited subjects over the age of 40 with oversampling of those reporting a history of snoring. Full details have been previously published.[12]

ARIC is a longitudinal cohort study of atherosclerosis and its clinical sequelae. From 1987 to 1989, a population-based sample of 15,792 men and women aged 45 to 64 years were recruited from 4 US communities (Forsyth County NC, Jackson MS, suburban Minneapolis MN, and Washington County MD). SHHS recruited 1000 ARIC participants in Minneapolis and 750 participants from Washington County.

The FHS started in 1948 with 5209 randomly ascertained participants from Framingham, MA, who had undergone biannual examinations to investigate cardiovascular disease and its risk factors. In 1971, the offspring cohort (comprised of 5124 children of the original cohort and the children’s spouses) and in 2002, the third generation (consisting of 4095 children of the offspring cohort), were recruited. FHS participants in this study are of European ancestry. SHHS recruited 1000 FHS participants from the offspring cohort.

The Western Australia Sleep Health Study (WASHS) is a prospective cohort study of patients referred to the sole public sleep clinic in Western Australia. It has been designed to identify the genetic basis of sleep disorders and co-morbidities.[13] The vast majority (91%) of participants were referred for sleep apnea. Recruitment began in 2006. The cohort is predominantly of European ancestry. Analyses are limited to the first 1795 individuals of European ancestry with both DNA and sleep phenotyping available.

Sleep Apnea Phenotyping

Quantification of sleep apnea severity was done using polysomnography as part of either the SHHS or the CFS, using techniques that have been previously described.[14, 15] In brief,
SHHS participants underwent in-home polysomnography using the Compumedics P Series System (Abbotsville, Australia) while CFS participants underwent in laboratory polysomnography using the Compumedics E series system or in-home sleep studies with measurement of oximetry, effort, thermistry, and body position (Edentec, Eden Prairie, MN). Indices derived from either technique were highly correlated.[16] Scoring of both sleep and breathing was done in a standardized fashion across SHHS and CFS by the same Reading Center at Case Western Reserve University.[14] Apneas were defined as no airflow for 10 seconds while hypopneas were defined as a 30% reduction in airflow accompanied by a 3% desaturation or arousal. Overnight in-laboratory polysomnography was performed in WASHS and the AHI was computed using similar scoring criteria as recommended by the American Academy of Sleep Medicine.[17] Two phenotypes were assessed for association with candidate SNPs. A quantitative trait, apnea hypopnea index, was log-transformed (after adding 1) to approximate a normal distribution. In addition, the dichotomous trait, OSA defined as an AHI≥15, was assessed.

**Covariate Assessment**

Age, gender, and race were obtained by self-report. Height and weight were measured in a standardized fashion in each cohort and body mass index (BMI) calculated as the ratio of weight to height squared.

**Genotyping**

Genotyping was performed at the Broad Institute utilizing an IBC 50K SNP array which was custom designed as a gene-centric single nucleotide polymorphism (SNP) genotyping array that contains greater SNP marker density and linkage disequilibrium coverage for over 2000 candidate regions relevant to heart, lung, blood and sleep disorders than current genome-wide arrays, including approximately 7800 SNPs not present in the HapMap.[18] A total of 45,237
candidate SNPs were genotyped on the IBC array. The tagging approach utilized on the IBC array was designed to capture maximal genetic information from the HapMap populations as well as European and African American individuals from the SeattleSNPs and NIEHS sequencing programs.[18] Only SNPs with a MAF of 5% or greater were included in these analyses due to concerns for unstable point estimates and inflated type 1 error. SNPs were clustered into genotypes using the Illumina Beadstudio software and subjected to quality control filters at the sample and SNP level, separately within each cohort. Samples were excluded for individual call rates <90%, gender mismatch and duplicate discordance. SNPs were removed for call rates <95% or Hardy-Weinberg equilibrium (HWE) \( p < 10^{-7} \) in analyses of individuals of European ancestry (EAs). Because of expected admixture, no HWE filtering was used for African-Americans (AAs).

**Statistical Analysis**

*Principal Components Analysis (PCA)*

Principal components were generated using EIGENSTRAT within each cohort using the CARe genotype data in order to control for population stratification.[19, 20] Two reference populations were included in the principal component analysis of AAs: 1,178 European Americans from a multiple sclerosis GWA study (from Dr. Phil de Jager and colleagues), and 756 Nigerians from the Yoruba region from a hypertension GWA (provided by Dr. Richard Cooper and colleagues). Importantly, these two samples underwent extensive quality control procedures to remove population outliers using PCA. Ten principal components were generated for each cohort.

*Within Cohort Analysis*

All data analyses were performed separately for EAs and AAs. To further adjust for population stratification, the first 10 principal components were incorporated as covariates in all
analyses. Primary analyses adjusted for BMI to identify variants whose actions were independent of obesity. Secondary analyses assessed the strength of association without BMI adjustment.

Analyses were carried out using a linear or logistic statistical framework for logAHI and OSA respectively in PLINK V 1.0.7,[21] or using R scripts that model family structure for related cohorts (CFS and FHS).[22] To obtain interpretable parameter estimates, beta coefficients from linear regression models were multiplied by the standard deviation. Consistency of findings was evaluated by comparing results in secondary analyses stratified on obesity status as defined by a BMI $\geq 30$ kg/m$^2$.

Meta-analysis

For EAs, results from each of the three cohorts (CFS, ARIC, FHS) were combined using a random effects meta-analysis with an inverse-variance weighted approach as implemented in METAL.[23] Heterogeneity was assessed using the $I^2$ inconsistency metric. No meta-analysis was required for AAs as only CFS had sufficient AAs for analysis.

Genomic Control

Genomic control correction was applied after calculating the genomic control inflation factor ($\lambda$) based on the results of the meta-analysis performed in EAs or in the CFS cohort alone for AAs.

Significance Criterion

To correct for multiple testing, we determined the effective number of independent tests as 26,482 for AA and 20,544 for EA based on the correlation structure among the SNPs estimated using the spectral method.[24, 25] To maintain an overall type 1 error rate of 5%, a statistical threshold of $\alpha = 2 \times 10^{-6}$ was used to declare array-wide significance.

Replication Analysis
Replication of EA findings was assessed using the WASHS cohort. Genotyping was performed using the Taqman platform (Applied Biosystems). Statistical analyses were done using linear and logistic regression for logAHI and OSA respectively adjusting for age, sex, and BMI and one-sided p-values generated.

RESULTS

The sample size and participant characteristics from each study are shown in Table 1. In total, 3,551 individuals were included in the primary analyses – 2,904 EAs and 647 AAs with an additional 1795 WASHS participants serving as the replication sample for the EA findings. There was a fairly even gender distribution in the population-based cohorts with a greater proportion of men in WASHS as one would expect for a clinical cohort. Mean BMI across cohorts ranged from 28 to 33 kg/m² suggesting a high prevalence of obesity. The prevalence of moderate to severe OSA as defined by an AHI ≥ 15 was 29-36% across population-based cohorts and 77% in the clinical cohort selected on suspected OSA.

Overall, there was little evidence for population stratification in the EAs with genomic control inflation factors ($\lambda$) of 1.05 and 1.04 for the OSA and logAHI analyses respectively. Slightly greater $\lambda$s were obtained for AAs (1.10 for OSA and 1.08 for logAHI). However, after using genomic control correction, the resulting quantile-quantile plots (Supplemental Figures S1-S4) showed no evidence of type I error inflation in any of the analyses.

African-American Findings

Manhattan plots displaying the strength of association between logAHI and OSA with each genotyped SNP in AAs are displayed in Figures 1-2 respectively. One SNP, rs7030789 in an intronic region of the LPAR1 gene was significantly associated with logAHI in AAs (Table 2)
while a second SNP, rs7972342, in an intronic region of the ITPR2 gene, just missed the threshold for statistical significance with a p-value of $2.3 \times 10^{-6}$. The effects of these SNPs were reduced slightly in BMI-adjusted models compared to models not including BMI. However, stratified analyses revealed the effect of rs7030789 was actually greater in the non-obese than the obese ($\beta = 0.18$ vs. $0.09$) and the effect of rs7972342 was similar in both groups ($\beta = -0.12$ vs. -0.15). Both SNPs were nominally associated with the dichotomous OSA phenotype ($p=0.007$ for rs7030789 and $p=0.003$ for rs7972342).

One SNP, rs11126184 downstream of the PLEK gene on chromosome 2 was significantly associated with the dichotomous OSA phenotype (Table 3). The minor allele in this SNP was associated with a reduced risk of OSA at an OR of 0.43. The effect of this SNP was not reduced with BMI adjustment and in fact, stratified analyses suggested a stronger association in the non-obese (OR=0.26) compared to the obese (OR=0.47) subgroup. This suggests this locus may influence OSA through obesity-independent pathways. This SNP was nominally associated with logAHI ($p=0.004$).

**European Ancestry Findings**

Manhattan plots displaying the strength of association between logAHI and OSA with each genotyped SNP in EAs are displayed in Figures 3-4 respectively. Among EAs, no SNP met criteria for a statistically significant association with logAHI. One SNP, rs1409986, located in an intronic region of the PTGER3 gene which codes for a prostaglandin E2 receptor was significantly associated with OSA in EAs ($p=1.0 \times 10^{-6}$), with an OR of 2.1 for each additional risk allele (Table 3). There was no evidence of heterogeneity across the cohorts ($I^2 = 0.32$, $p=0.22$ for heterogeneity). The association strengthened with adjustment for BMI suggesting that this association was not mediated through obesity. In support of this interpretation, analyses stratified by obesity status
found similar effects for rs1409986 among those with and without obesity (OR = 2.36 vs. 2.37). This SNP was nominally associated with logAHI (p=0.02).

Replication

The minor allele frequency for rs1409986 in WASHS was 7.3% with no evidence of deviation from HWE (p=0.38). The OR for OSA for each risk allele in this cohort adjusting for age, sex, and BMI was 1.20 (p=0.13). Under an additive model, there was no association with logAHI (p=0.35). However, under a co-dominant model, an association was found (p=0.05). While no difference was found between CT and CC genotypes, the TT genotype was associated with a 50% greater AHI than the wild type CC genotype (p=0.02).

Sensitivity Analyses

No change in results for either EAs or AAs was found by including data from the 354 children in the CFS cohort. In contrast, results from the CHS cohort did not support findings for either race. Note that this cohort was on average more than 10 years older than both ARIC and FHS with an even greater age difference with CFS.

Comparing results across races, the prevalence of the minor allele for rs1409986 (PTGER3) in AAs was only 2% precluding accurate estimation of its effect in this population. Among the top three SNPs for apnea-related traits identified in AAs, only rs7030789 (LPAR1) showed any evidence for association to an apnea phenotype in EAs with p=0.01 for association to OSA and p=0.06 to logAHI.

DISCUSSION

This study represents one of the first systematic evaluations of a large number of genetic loci identified as relevant to heart, lung, blood, and sleep phenotypes in relationship to sleep apnea.
OSA represents one of the most common sleep disorders with a prevalence of close to 5% in both pediatric and adult populations. The substantial neurocognitive and cardiovascular morbidity attributed to OSA as well as the likely rising prevalence of this condition with the growing obesity epidemic make understanding the genetic basis for this disease an important priority. Using genetic studies to identify novel molecular pathways involved in OSA pathogenesis may allow for the development of new treatments for this disorder – an important goal given the relatively poor tolerance of current treatment options.

Using data collected from three large cohorts and a genotyping platform that selected for candidate genes relevant to cardiovascular, respiratory and sleep physiology, we identified several novel SNPs associated with sleep apnea related phenotypes. In EAs, a polymorphism in PTGER3, a prostaglandin E2 receptor, was significantly associated with OSA. In a replication analysis using data from a clinic-based cohort which had a skewed distribution of AHI due to high representation of patients with sleep apnea or sleep apnea symptoms, this SNP was associated with an incremental increase in disease severity as measured by the AHI level. Although association with the dichotomous trait OSA was not replicated, this may be due to limited power for this trait since nearly 80% of subjects in the WASHS cohort met OSA criteria. The PTGER3 gene is expressed in neuronal tissue and modulates neurotransmitter release in both central and peripheral neurons. A haplotype analysis of PTGER3 recently suggested this gene may represent a risk factor for hypertension.[26] Unfortunately, sleep apnea was not characterized in that study.

Among AAs, the implicated susceptibility genes included the lysophosphatidic acid receptor 1 (LPAR1) and plekstrin (PLEK). A GWA study identified a locus close to LPAR1 as predictive of monocyte counts.[27] suggesting a pro-inflammatory role for this gene. PLEK is a substrate for protein kinase C in platelets and a wide range of leukocytes including monocytes and
macrophages. This gene plays a role in actin assembly; knock out of PLEK in mouse models results in defective exocytosis.\[28\] In addition, there was suggestive evidence for an association with a locus in the gene for the inositol triphosphate receptor 2 (ITPR2). The ITPR2 gene plays a central role in intra-cellular calcium regulation, which is vital to cellular stability, cellular adhesion, and second messenger activity. Prior candidate gene studies have associated SNPs in ITPR2 with markers of inflammation and endothelial dysfunction as well as blood pressure phenotypes.\[29, 30\] Together, these results suggest alterations in inflammatory pathways including monocyte/macrophage function may be particularly important in OSA pathogenesis among AAs as well as the possibility of overlapping pathways that may mediate both hypertension and OSA.

It is important to note that our findings were not replicated in the Cardiovascular Health Study. This cohort is substantially older than the cohorts from which the primary results are based and the mechanisms underlying OSA in the elderly may differ substantially from OSA in middle aged populations. For example, obesity plays a much weaker role in the elderly while issues related to ventilatory control are much more prominent as evidenced by the greater number of central events in older individuals. Only one of the associations (rs7030789 in LPAR1) could be replicated across races. This may be due to differences in allele frequencies impacting power (as for rs1409986 in PTGER3), differences in linkage disequilibrium structure or other ancestry related effects.

Of interest, although the African American sample was relatively small, it was derived from a single cohort that was assembled explicitly to enhance the power to detect genetic associations through recruitment of affected probands and multiplex families. OSA is common in minority populations and prior segregation analyses suggested that there may be stronger transmission of one or several alleles in AAs compared to EA families.\[31\] Although genetic studies in AAs may
be limited by the coverage patterns of existing chips, the genetic diversity of this admixed population also provides opportunities to identify novel variants and to perform in silico “fine mapping”.[32] Despite the sample size, we identified two loci significantly associated with sleep apnea phenotypes in AAs and a third nearly significant association. Unfortunately, we could not locate other AA samples with OSA phenotyping to perform replication studies. This is a large gap that will need to be addressed for further development of minority cohorts who have had sleep apnea phenotyping and genotyping.

Strengths of this study include its use of standardization in both phenotyping and genotyping across several cohorts. The use of a broadly defined candidate gene approach (i.e., with genes selected for heart, lung, blood and sleep traits) provides balance in examining relevant but not narrowly defined susceptibility loci without the statistical penalties incurred by using a genome-wide association approach. This is particularly relevant for conditions such as OSA that are expensive to phenotype. On the other hand, this approach is not amenable to identifying completely novel loci.

We also conducted analyses in parallel to identify loci important in two racial groups – those of an African-American background and those of European ancestry. Demonstrating a similar pattern of association between apnea phenotypes and rs7030789 across races provides additional evidence as to the presence of a true association. Another strength of this work is the use of both principal component adjustment and genomic control correction to prevent inflation of the false positive rate due to cryptic population stratification or other sources of bias. Nevertheless, additional analyses from independent cohorts are needed to replicate the findings presented in this study.
Limitations of the work include the relatively small number of cohorts included and modest sample size compared to other cross-cohort genetic analyses. Unfortunately, few cohorts with detailed sleep apnea phenotyping as well as genotyping currently exist, particularly in minority populations. This limited our opportunity to conduct replication studies which will be needed to definitively confirm our findings.

In summary, our results identify three novel loci, each of which has been implicated in inflammation, associated with a sleep apnea phenotype. While much work has focused on the potential pro-inflammatory effect of OSA, inflammation may play a causal role in OSA pathogenesis as well. Inflammation is prominent in both the mucosal and muscular layers of upper airway tissues of patients with OSA.[33, 34] In addition, inflammation may also impact ventilatory or sleep-wake control mechanisms relevant to OSA pathogenesis.

Of note, though all of the cohorts studied had a high prevalence of obesity, the identified associations were independent of obesity supporting the importance of identifying pathways leading to OSA within an obese population. Although further replication is needed, their association with OSA provides further support for a role for inflammation in the pathogenesis of OSA. Additional studies are needed to better understand the molecular pathways that increase susceptibility to this highly prevalent disorder, the overlap of such pathways with ones that also mediate cardiopulmonary disorders, and the extent to which the associations are moderated by race and ethnicity.

ACKNOWLEDGEMENTS:

We thank the investigators, staff, and participants of ARIC, CFS, FHS, and WASHS for their valuable contributions. The authors additionally thank Deborah Farlow (The Broad Institute,
Cambridge, MA), and Guillaume Lettre (Montreal Heart Institute/Universite de Montreal, Montreal, Canada) for their significant efforts and contributions to the CARe consortium.
Table 1. Characteristics of the Study Participants

<table>
<thead>
<tr>
<th>Cohorts</th>
<th>European Ancestry</th>
<th></th>
<th></th>
<th>African-Americans</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>ARIC (n=1673)</td>
<td>FHS (n=567)</td>
<td>CFS (n=664)</td>
<td>WASHS (n=1795)</td>
</tr>
<tr>
<td>Age</td>
<td>62.5 ± 5.7</td>
<td>59.2 ± 9.0</td>
<td>41.3 ± 19.6</td>
<td>51.5 ± 13.2</td>
</tr>
<tr>
<td>Male</td>
<td>47%</td>
<td>48%</td>
<td>47%</td>
<td>63%</td>
</tr>
<tr>
<td>BMI</td>
<td>28.9 ± 5.1</td>
<td>28.3 ± 5.1</td>
<td>30.1 ± 8.7</td>
<td>32.9 ± 7.8</td>
</tr>
<tr>
<td>AHI</td>
<td>9.1 (3.7, 19.0)</td>
<td>8.0 (3.1, 17.7)</td>
<td>4.7 (1.4, 18.3)</td>
<td>28.5 (16.0, 50.7)</td>
</tr>
<tr>
<td>AHI ≥ 15</td>
<td>33%</td>
<td>29%</td>
<td>36%</td>
<td>77%</td>
</tr>
</tbody>
</table>

Values displayed are means ± standard deviation or median (interquartile range) for continuous variables and percentages for dichotomous variables.
Table 2. SNPs associated with AHI

<table>
<thead>
<tr>
<th>Race</th>
<th>SNP</th>
<th>Gene</th>
<th>Chr</th>
<th>Position</th>
<th>Minor Allele</th>
<th>Major Allele</th>
<th>Minor Allele Frequency</th>
<th>Hardy-Weinberg P-value</th>
<th>Beta (SE)</th>
<th>P-value</th>
<th>BMI-Adjusted</th>
<th>BMI-Unadjusted</th>
</tr>
</thead>
<tbody>
<tr>
<td>African-American</td>
<td>rs7030789</td>
<td>LPAR1</td>
<td>9</td>
<td>112775333</td>
<td>A</td>
<td>G</td>
<td>0.319</td>
<td>0.10</td>
<td>0.109 (0.023)</td>
<td>1.50 x 10^{-6}</td>
<td>0.129 (0.028)</td>
<td>4.58 x 10^{-6}</td>
</tr>
<tr>
<td>African-American</td>
<td>rs7972342</td>
<td>ITPR2</td>
<td>12</td>
<td>26681785</td>
<td>A</td>
<td>G</td>
<td>0.290</td>
<td>1.00</td>
<td>-0.113 (0.024)</td>
<td>2.33 x 10^{-6}</td>
<td>-0.134 (0.030)</td>
<td>6.97 x 10^{-6}</td>
</tr>
</tbody>
</table>
Table 3. SNPs associated with OSA

<table>
<thead>
<tr>
<th>Race</th>
<th>SNP</th>
<th>Gene</th>
<th>Chr</th>
<th>Position</th>
<th>Minor Allele</th>
<th>Major Allele</th>
<th>Minor Allele Frequency</th>
<th>P-value*</th>
<th>BMI-Adjusted OR (95% CI)</th>
<th>P-value</th>
<th>BMI-Unadjusted OR (95% CI)</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>African-American</td>
<td>rs11126184</td>
<td>PLEK</td>
<td>2</td>
<td>68505678</td>
<td>A</td>
<td>C</td>
<td>0.378</td>
<td>---</td>
<td>0.43 (0.31, 0.60)</td>
<td>1.54 x 10^{-6}</td>
<td>0.45 (0.33, 0.62)</td>
<td>1.41 x 10^{-6}</td>
</tr>
<tr>
<td>European Ancestry</td>
<td>rs1409986</td>
<td>PTGER3</td>
<td>1</td>
<td>71104086</td>
<td>A</td>
<td>G</td>
<td>0.075</td>
<td>32%</td>
<td>2.14 (1.58, 2.90)</td>
<td>1.01 x 10^{-6}</td>
<td>1.77 (1.30, 2.41)</td>
<td>2.21 x 10^{-4}</td>
</tr>
</tbody>
</table>

* Heterozygosity P-value for European ancestry cohorts and Hardy-Weinberg P-value for African-Americans
Figure Legends

Figure 1. Manhattan plot for apnea hypopnea index in African-Americans. This figure plots the association results for all SNPs against log(apnea hypopnea index + 1) among African-Americans. The y-axis displays the -log(p-value), the x-axis the SNP position on each chromosome. The dotted line represents the threshold for statistical significance based on the number of independent comparisons being made.

Figure 2. Manhattan plot for obstructive sleep apnea in African-Americans. This figure plots the association results for all SNPs against an apnea hypopnea index of 15 or greater among African-Americans. The y-axis displays the -log(p-value), the x-axis the SNP position on each chromosome. The dotted line represents the threshold for statistical significance based on the number of independent comparisons being made.

Figure 3. Manhattan plot for apnea hypopnea index in the European ancestry population. This figure plots the association results for all SNPs against log(apnea hypopnea index + 1) among those of European ancestry. The y-axis displays the -log(p-value), the x-axis the SNP position on each chromosome. The dotted line represents the threshold for statistical significance based on the number of independent comparisons being made.

Figure 4. Manhattan plot for obstructive sleep apnea in the European ancestry population. This figure plots the association results for all SNPs against an apnea hypopnea index of 15 or greater among those of European ancestry. The y-axis displays the -log(p-value), the x-axis the
SNP position on each chromosome. The dotted line represents the threshold for statistical significance based on the number of independent comparisons being made.

**Figure S1. Quantile-quantile plot for apnea hypopnea index in African-Americans.** This figure plots expected versus observed p-values from the association analyses of all SNPs against log(apnea hypopnea index+1) in African-Americans. The plotted observed p-values are after accounting for genomic control.

**Figure S2. Quantile-quantile plot for obstructive sleep apnea in African-Americans.** This figure plots expected versus observed p-values from the association analyses of all SNPs against an apnea hypopnea index of 15 or greater in African-Americans. The plotted observed p-values are after accounting for genomic control.

**Figure S3. Quantile-quantile plot for apnea hypopnea index in European ancestry individuals.** This figure plots expected versus observed p-values from the association analyses of all SNPs against log(apnea hypopnea index+1) in those of European ancestry. The plotted observed p-values are those from the meta-analysis after accounting for genomic control.

**Figure S4. Quantile-quantile plot for obstructive sleep apnea in European ancestry individuals.** This figure plots expected versus observed p-values from the association analyses of all SNPs against an apnea hypopnea index of 15 or greater in those of European ancestry. The plotted observed p-values are those from the meta-analysis after accounting for genomic control.
Figure 1: Manhattan plot for apnea hypopnea index in African-Americans
Figure 2: Manhattan plot for obstructive sleep apnea in African-Americans
Figure 3: Manhattan plot for apnea hypopnea index in those of European ancestry
Figure 4: Manhattan plot for obstructive sleep apnea in those of European ancestry
Figure S1: Quantile-quantile plot for apnea hypopnea index in African-Americans
Figure S2: Quantile-quantile plot for obstructive sleep apnea in African-Americans
Figure S3: Quantile-quantile plot for apnea hypopnea index in those of European ancestry
Figure S4: Quantile-quantile plot for obstructive sleep apnea in those of European ancestry
References


