Whole-genome association analyses of sleep-disordered breathing phenotypes in the NHLBI TOPMed program

Brian E. Cade1,2,3*, Jiwon Lee1, Tamar Sofer1,2, Heming Wang1,2,3, Man Zhang4, Han Chen5,6, Sina A. Gharib7, Daniel J. Gottlieb1,2,8, Xiuxing Guo9, Jacqueline M. Lane1,2,3,10, Jingjing Liang11, Xihong Lin12, Hao Mei13, Sanjay R. Patel14, Shaun M. Purcell1,2,3, Richa Saxena1,2,3,10, Neomi A. Shah15, Daniel S. Evans16, Craig L. Hanis5, David R. Hillman17, Sutapa Mukherjee18,19, Lyle J. Palmer20, Katie L. Stone16, Gregory J. Tranah16, NHLBI Trans-Omics for Precision Medicine (TOPMed) Consortium, Gonçalo R. Abecasis21, Eric A. Boerwinkle5,22, Adolfo Correa23,24, L. Adrienne Cupples25,26, Robert C. Kaplan27, Deborah A. Nickerson28,29, Kari E. North30, Bruce M. Psaty31,32, Jerome I. Rotter9, Stephen S. Rich33, Russell P. Tracy34, Ramachandran S. Vasan26,35,36, James G. Wilson37, Xiaofeng Zhu11, Susan Redline1,2,38 and TOPMed Sleep Working Group

Abstract

Background: Sleep-disordered breathing is a common disorder associated with significant morbidity. The genetic architecture of sleep-disordered breathing remains poorly understood. Through the NHLBI Trans-Omics for Precision Medicine (TOPMed) program, we performed the first whole-genome sequence analysis of sleep-disordered breathing.

Methods: The study sample was comprised of 7988 individuals of diverse ancestry. Common-variant and pathway analyses included an additional 13,257 individuals. We examined five complementary traits describing different aspects of sleep-disordered breathing: the apnea-hypopnea index, average oxyhemoglobin desaturation per event, average and minimum oxyhemoglobin saturation across the sleep episode, and the percentage of sleep with oxyhemoglobin saturation < 90%. We adjusted for age, sex, BMI, study, and family structure using MMSKAT and EMMAX mixed linear model approaches. Additional bioinformatics analyses were performed with MetaXcan, GIGSEA, and ReMap.

* Correspondence: bcade@bwh.harvard.edu

Full lists of consortium and working group authors are provided in Additional file 1: Tables S1 and S2.

© The Author(s). 2021 Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/. The Creative Commons Public Domain Dedication waiver (http://creativecommons.org/publicdomain/zero/1.0/) applies to the data made available in this article, unless otherwise stated in a credit line to the data.
Results: We identified a multi-ethnic set-based rare-variant association ($p = 3.48 \times 10^{-5}$) on chromosome X with ARMCX3. Additional rare-variant associations include ARMCX3-AS1, MRPS33, and C16orf90. Novel common-variant loci were identified in the NRG1 and SLC45A2 regions, and previously associated loci in the IL18RAP and ATP2B4 regions were associated with novel phenotypes. Transcription factor binding site enrichment identified associations with genes implicated with respiratory and craniofacial traits. Additional analyses identified significantly associated pathways.

Conclusions: We have identified the first gene-based rare-variant associations with objectively measured sleep-disordered breathing traits. Our results increase the understanding of the genetic architecture of sleep-disordered breathing and highlight associations in genes that modulate lung development, inflammation, respiratory rhythmogenesis, and HIF1A-mediated hypoxic response.

Keywords: Sleep-disordered breathing, Sleep apnea, Whole-genome sequencing, WGS, Genome-wide association study, GWAS

Background
Sleep-disordered breathing (SDB) is a prevalent disorder associated with increased sleepiness, mortality, and morbidity from a wide range of cardiometabolic and other diseases [1, 2]. The most common type of SDB is obstructive sleep apnea (OSA), characterized by repeated airway collapse leading to intermittent hypoxemia and sleep disruption, that is increased in prevalence with older age and male sex [2]. An estimated 936 million adults aged 30–69 have mild to severe OSA worldwide [3]. The disease is heritable and appears to be multifactorial, reflecting variable contributions of abnormalities in ventilatory control, craniofacial anatomy, and adiposity [2, 4–7]. Sleep-related hypoxemia can also be due to central sleep apnea, a less common disorder, due to a lack of respiratory drive [8]. OSA is typically measured clinically using the apnea-hypopnea index, which counts the number of total (apnea) and partial (hypopnea) breathing cessations per hour of sleep. Due to an incomplete understanding of its molecular basis, the standard OSA treatment of continuous positive airway pressure (CPAP) only addresses the downstream manifestations of airway collapse through nightly use of pressurized air to the nasopharynx, a therapy that often is poorly tolerated. Therefore, there is a critical need to identify molecular pathways that could provide specific therapeutic targets. The need for overnight studies to phenotype SDB traits has limited the available sample size for genetic analyses, and only several common-frequency genome-wide analysis studies have been reported [9–11]. Increased statistical power may increase the genetic resolution of regions that may not be adequately tagged by current genotyping arrays due to population differences and/or reduced linkage disequilibrium with biologically relevant regions.

The Trans-Omics for Precision Medicine (TOPMed) program is an NIH National Heart, Lung, and Blood Institute program designed to improve the understanding of the biological processes that contribute to heart, lung, blood, and sleep disorders [12]. TOPMed has generated whole-genome sequencing (WGS) data on over 100,000 individuals from multiple cohorts at >30× depth, including seven studies with objective assessment of SDB. A variant imputation server using TOPMed data also allows for high-quality imputation of non-sequenced genotype chip data [13]. A complementary initiative sponsored by the Centers for Common Disease Genomics (CCDG) of the NIH National Human Genome Research Institute has generated sequencing data from additional individuals in two TOPMed cohorts. These initiatives provide the ability to examine the genetics of SDB at unprecedented detail in African-Americans (AA), Asian-Americans (AsA), European-Americans/Australians (EA), and Hispanic/Latino-Americans (HA).

In this first genome-wide sequencing analysis of SDB, we examine the apnea-hypopnea index (AHI), the standard clinic metric of SDB, and four complementary measurements of overnight hypoxemia: average and minimum oxyhemoglobin saturation ($\text{SpO}_2$) during sleep and the percent of the sleep recording with $\text{SpO}_2 < 90\%$ (Per90), and the average desaturation per hypopnea event. These indices were chosen because of clinical relevance, high heritability, or prior significant GWAS findings [9, 11, 14]. We examined 7988 individuals with objectively measured SDB and WGS data in conjunction with data from 13,257 individuals with imputed genotype data.

Methods
Each study had a protocol approved by its respective Institutional Review Board and participants provided informed consent. A study overview is provided in Additional file 2: Figure S1. There were two classes of data: “WGS studies” had WGS performed by the TOPMed...
program and, in some cases, in additional participants by the CCDG program (referred to as “WGS” studies); “Imputed studies” had array-based genotyping later imputed using the TOPMed imputation server (as described below). Some studies with WGS contributed imputed study data from additional array-based genotyped individuals. Ten studies were analyzed (Tables 1 and 2).

**WGS studies**

The Atherosclerosis Risk in Communities Study (ARIC), the Cardiovascular Health Study (CHS), and the Framingham Heart Study Offspring Cohort (FHS) included individuals who participated in the Sleep Heart Health Study (SHHS), who underwent polysomnography (PSG) between 1995 and 1998 using the Compumedics PS-2 system [15–18]. These samples included 1028 EAs from ARIC, 151 AAs and 557 EAs from CHS, and 478 EAs from FHS.

The Multi-Ethnic Study of Atherosclerosis (MESA) is investigating the familial basis of SDB, with four visits occurring from 1990 to 2006 [21]. PSG was obtained between 2010 and 2013 using the Compumedics Somte system [20]. This analysis includes data from 698 EAs, 486 AAs, 456 HAs, and 229 AsAs.

The Cleveland Family Study (CFS) was designed to investigate the familial basis of SDB, with four visits occurring from 1990 to 2006 [21]. Sleep was assessed either in a clinical research center using full PSG (Compumedics Somte system) (visit 4) or in the latest available prior examination using an in-home sleep apnea testing device (Edwards-Embletta Gold) [26, 27]. Five hundred seventy-five individuals contributed data.

**Imputed genotype studies**

The Osteoporotic Fractures in Men Study (MrOS) is a multi-center cohort study initially designed to examine the risk factors for osteoporosis, fractures, and prostate cancer in older males [28, 29]. An ancillary study (MrOS Sleep; 2003–2005) focused on outcomes of sleep disturbances used PSG and nearly identical procedures as in MESA (Compumedics Sario system) [30]. Two thousand one hundred eighty-one EA individuals were included, with genotyping performed using the Illumina Human Omni 1 Quad v1-0 H array.

The Jackson Heart Study (JHS) is investigating cardiovascular disease in AAs [25]. An in-home sleep study was performed from 2012 to 2016 using a validated device 3 sleep apnea testing device (Embla Embletta Gold) [26, 27]. Five hundred seventy-five individuals contributed data.

The Hispanic Community Health Study/Study of Latinos (HCHS/SOL) is studying multiple health conditions in HAs [22, 23]. Home sleep apnea testing was performed during the baseline examination (2008–2011) using the ARES Unicorder 5.2, a validated device including a forehead-based reflectance oximeter, a nasal pressure cannula and pressure transducer, an accelerometer, and a microphone [24]. Two thousand three hundred thirty-nine individuals provided data.

The Jackson Heart Study (JHS) is investigating cardiovascular disease in AAs [25]. An in-home sleep study was performed from 2012 to 2016 using a validated type 3 sleep apnea testing device (Embla Embletta Gold) [26, 27]. Five hundred seventy-five individuals contributed data.

### Table 1 Sample description for WGS cohorts

| Population          | Cohort | N       | Age     | Percent female | BMI     | Apnea-hypopnea index | AHI (percent < 5, 5–15, ≥ 15) | Average desaturation | Average SpO2 | Minimum SpO2 | Percent sleep under 90% SpO2 |
|---------------------|--------|---------|---------|----------------|---------|----------------------|-------------------------------|----------------------|--------------|--------------|-----------------------------|
| African-American    | CFS*   | 505     | 38.65 (18.96) | 56.4 | 32.44 (9.48) | 6.85 (22.48) | 43.4, 20.5, 36.0 | 3.62 (1.99) | 94.49 (3.91) | 84.76 (9.83) | 4.79 (13.15) |
|                     | CHS    | 151     | 75.39 (43.3)  | 60.3 | 29.02 (5.08) | 9.60 (16.96) | 28.5, 36.4, 35.1 | 2.70 (1.74) | 94.82 (2.19) | 85.74 (5.35) | 3.99 (9.63)  |
|                     | JHS    | 575     | 63.47 (10.94) | 64.9 | 31.8 (6.88)  | 10.69 (14.42) | 24.7, 39.5, 35.8 | 3.54 (1.72) | 94.77 (2.02) | 84.30 (6.57) | 2.97 (8.91)  |
|                     | MESA   | 486     | 68.81 (9.07)  | 53.7 | 30.23 (5.68) | 12.67 (20.56) | 22.4, 32.9, 44.7 | 3.42 (2.10) | 94.46 (1.99) | 83.32 (7.98) | 3.89 (9.49)  |
| East Asian-American | MESA   | 229     | 67.89 (9.11)  | 49.8 | 24.28 (3.30) | 14.96 (24.8)  | 21.8, 28.4, 49.8 | 3.72 (1.79) | 94.92 (1.22) | 83.23 (7.58) | 2.25 (4.46)  |
| European-American   | ARC    | 1028    | 62.28 (5.67)  | 53.1 | 28.72 (5.06) | 8.64 (15.62)  | 34.6, 32.4, 33.0 | 2.35 (1.29) | 94.57 (1.84) | 85.95 (5.93) | 2.92 (9.24)  |
|                     | CFS*   | 485     | 43.23 (19.49) | 50.5 | 30.81 (8.83) | 7.09 (21.90)  | 44.7, 19.4, 35.9 | 3.29 (1.86) | 93.67 (3.59) | 85.55 (9.33) | 4.66 (11.87) |
|                     | CHS    | 557     | 77.93 (4.34)  | 54.2 | 27.25 (4.44) | 11.42 (15.54) | 23.2, 38.1, 38.8 | 2.58 (1.34) | 94.00 (2.00) | 84.99 (5.67) | 4.77 (12.28) |
|                     | FHS*   | 478     | 60.09 (8.54)  | 49.8 | 28.40 (9.06) | 10.18 (14.28) | 35.1, 35.1, 29.7 | 2.35 (1.27) | 94.68 (2.04) | 87.85 (6.25) | 2.96 (9.18)  |
|                     | MESA   | 698     | 68.53 (9.06)  | 33.2 | 27.91 (5.10) | 12.18 (20.45) | 21.6, 35.0, 43.4 | 3.11 (1.74) | 93.96 (1.75) | 83.49 (7.50) | 4.27 (10.82) |
| Hispanic/Latino-American | HCHS/SOL | 2339 | 46.27 (13.86) | 60.5 | 30.23 (6.44) | 20.3 (6.30) | 68.9, 19.5, 11.6 | N/A | 96.42 (0.99) | 87.04 (5.92) | 0.88 (3.63)  |
|                     | MESA   | 456     | 68.49 (9.27)  | 53.3 | 30.08 (5.46) | 16.31 (22.53) | 17.1, 28.3, 54.6 | 3.62 (2.12) | 94.33 (1.60) | 81.59 (9.32) | 3.80 (7.64)  |

Seven studies contributed 7988 individuals with WGS in TOPMed Freeze 6a and objectively measured phenotypes (1717 African-Americans, 229 Asian-Americans, 3246 European-Americans, 2796 Hispanic/Latino-Americans). The overall sample had a mean age of 57.7 and was 56.1% female. Values are displayed as mean (SD), except for the skewed apnea-hypopnea index, which is displayed as median (IQR). Sample size N reflects individuals with non-missing AHI and covariate values. *Family cohort
Table 2 Sample description for imputed genotype chip cohorts

| Population            | Cohort | N      | Age       | Percent female | BMI       | AHI (percent < 5, 5–15, ≥ 15) | Average desaturation | Average SpO2 | Minimum SpO2 | Percent sleep under 90% SpO2 |
|-----------------------|--------|--------|-----------|----------------|-----------|-------------------------------|-----------------------|--------------|--------------|-------------------------------|
| African-American      | CFS*   | 225    | 35.46 (20.32) | 56.4          | 29.97 (10.09) | 3.99 (10.55) | 55.1, 23.1, 21.8 | 2.90 (1.09) | 94.65 (4.01) | 88.17 (9.60) | 5.20 (16.01) |
| European-American,    | ARIC   | 631    | 62.74 (5.72)  | 49.4          | 29.15 (5.23)  | 9.15 (15.02) | 29.3, 37.9, 32.8 | 2.50 (1.73) | 94.32 (2.15) | 85.17 (6.17) | 4.12 (11.76) |
| Australian            | CFS*   | 218    | 37.57 (18.66) | 56.9          | 28.76 (8.81)  | 3.4 (10.59)  | 57.8, 22.5, 19.7 | 2.30 (1.11) | 94.09 (3.35) | 88.81 (7.80) | 3.26 (12.79) |
| CHS                   | 365    | 77.44 (4.65)  | 64.9          | 27.10 (4.14)  | 10.50 (15.14) | 25.8, 39.2, 35.1 | 2.63 (1.57) | 94.41 (1.91) | 84.87 (5.96) | 3.93 (11.89) |
| FHS*                  | 192    | 57.45 (9.68)  | 51.0          | 28.87 (5.16)  | 7.30 (14.38)  | 38.0, 31.8, 30.2 | 2.42 (1.51) | 94.73 (1.80) | 85.76 (5.46) | 2.82 (8.38)  |
| MiOS                  | 2181   | 76.65 (5.60)  | 6.0           | 27.21 (3.75)  | 13.00 (18.00) | 18.9, 36.1, 45.0 | 3.54 (1.48) | 93.85 (1.73) | 84.39 (5.88) | 4.40 (9.95)  |
| WASHS                 | 1508   | 52.29 (13.71) | 40.9          | 31.84 (7.93)  | 7.24 (15.37)  | 40.1, 31.1, 28.8 | 3.56 (2.00) | 94.56 (2.38) | 84.61 (7.86) | 5.44 (13.82) |
| Hispanic, Latino-American | HCHS, SOL | 7155 | 46.10 (13.81) | 57.8          | 29.68 (5.86)  | 2.00 (6.15)  | 69.1, 19.3, 11.6 | N, A          | 96.46 (0.95) | 87.06 (6.11) | 0.83 (2.99)  |
| Starr                 | 782    | 52.34 (11.29) | 71.9          | 32.15 (6.78)  | 10.35 (17.18) | 31.5, 31.5, 37.1 | N, A          | 94.65 (2.09) | 85.78 (7.50) | 2.83 (8.79)  |

Eight studies contributed 13,257 individuals with genomic data imputed with a TOPMed Freeze 5b reference panel and objectively measured phenotypes (225 African-Americans, 5095 European-Americans, 7937 Hispanic/Latino-Americans). ARIC, CFS, CHS, FHS, and HCHS/SOL imputed genomic data reflect individuals without available sequencing in TOPMed Freeze 6. The overall sample had a mean age of 53.7 and was 46.9% female. Values are displayed as mean (SD), except for the skewed apnea-hypopnea Index, which is displayed as median (IQR). Sample size N reflects individuals with non-missing AHI and covariate values.

*Family cohort

WatchPAT-200) [33]. Seven hundred eighty-two HA individuals were studied, using Affymetrix 6.0 genotyping data. The Western Australian Sleep Health Study (WASHS) is a clinic-based study focused on the epidemiology and genetics of SDB [34]. PSG was obtained from 1508 European-ancestry patients (91% referred for SDB evaluation) from 2006 to 2010 (Compumedics Series E). Genotyping was performed using the Illumina Omni 2.5 array.

Imputed genotype data were available for additional members of the TOPMed cohorts described above. Study/population combinations with fewer than 100 individuals were excluded. ARIC contributed an additional 631 EA individuals (Affymetrix 6.0; dbGaP phg000035.v1.p1). CFS contributed 225 AA and 218 EA individuals (Affymetrix 6.0; Illumina OmniExpress+Exome, Exome, and IBC). CHS contributed 365 individuals (Illumina CNV370 and IBC; phg000135.v1.p1 and phg000077.v1.p1). FHS contributed 192 EA individuals (Affymetrix 500 k; phg000066.v7). HCHS/SOL contributed 7155 HA individuals (Illumina Omni 2.5; phg000663.v1).

Phenotype and covariate definitions

We examined several SDB measures, including specific measures of OSA: AHI (number of apneas plus hypopneas per hour of sleep, with a minimum 3% desaturation per event) and average oxyhemoglobin desaturation per apnea or hypopnea, and measures of SDB severity [14]: average and minimum SpO2 and the percentage of the night with SpO2 < 90% (Per90). Apart from WASHS, all sleep data were scored by blinded scorers at one central Sleep Reading Center with high levels of scorer reliability using well-defined procedures [35]. The AHI reflected all events. We did not attempt to disentangle the apnea-hypopnea index from central versus obstructive sleep apnea events, due to the relatively low prevalence of central sleep apnea (< 2%) in these largely community-based studies [36, 37] (some of which are enriched with snorers) and the complexities of classifying mixed events. We adjusted for age, age², sex, age × sex, body mass index (BMI), and BMI² due to known age and sex effects, some of which are non-linearly associated with outcomes, and our goal of identifying obesity-independent loci. Age and BMI were obtained at the time of the sleep recording. We adjusted for BMI as over half of the AHI trait heritability is attributable to factors other than obesity as measured by the BMI and our goal was to identify associations with other mechanistic pathways (e.g., ventilatory control) that could indicate novel future targets. Phenotype analyses were pooled within populations to aggregate very rare variants for testing and therefore further adjusted for study. Population assignments were based on self-report, in accordance with other research from TOPMed and other consortia. ASA and EA-identifying individuals with population principal components > 5 standard deviations [38] from applicable 1000 Genomes and Human Genome Diversity Project super-populations were excluded. We used a two-stage procedure to rank-normalize the phenotypes adjusted for covariates [39]. Cryptic relatedness and population substructure were controlled for using linear mixed models. Genomic control was applied to population-specific results (or cohort-specific imputed genotype results).

WGS and genotyping

Sequence data were derived from the TOPMed Freeze 6a release, jointly called by the TOPMed Informatics Research Center at the University of Michigan (http://github.com/statgen/topmed_variant_calling). The methodology was described elsewhere [12]. In brief, WGS
was performed at the Broad Institute (ARIC, FHS, MESA), Baylor College of Medicine (ARIC, CHS, HCHS/SOL), and the University of Washington (CFS, JHS). Additional ARIC and HCHS/SOL WGS funded by CCDBG (https://www.genome.gov/27563570) and performed at Baylor College of Medicine were included in the jointly called data. TOPMed and CCDBG calling pipelines have functionally equivalent outcomes despite data processing differences (as detailed in [40]). WGS data were merged and normalized; inferred sequence contamination was identified; and SNPs and small indels were detected (structural variants are not currently available). Lower quality variants were excluded using Mendelian consistency checks. Variants were aligned to Build 38 and annotated using snpEff 4.3 t [41]. We excluded variants with < 10× depth or > 5% missingness, leaving 152.7 million polymorphic variants in 7988 individuals with SDB phenotypes. Up to 22,030,888 variants from individuals with sequencing were tested in the GWAS analyses, following filtering for quality control and minor allele frequencies.

Genotype data were imputed using the TOPMed Imputation Server [13] using a Freeze 5b (Build 38) template. Forward strand checks were performed using the Strand database and the Haplotype Reference Consortium imputation preparation script (https://www.well.ox.ac.uk/~wrayner/tools/) and confirmed using Ensembl variant allele checks and internal QC performed on the server. Study-level data were imputed separately. Analyses on variants with $r^2$ score > 0.5 were therefore performed separately for each study. Up to 22,105,437 variants from individuals with imputed data were tested in the GWAS analyses, following filtering for quality control, imputation $r^2$, and minor allele frequencies.

**Statistical analyses**

Single and grouped variant analyses were performed using EMMAX and MMSKAT, both within the EPAC TS suite (v3.3) [42]. WGS genetic relatedness matrices (GRM) were constructed using autosomal variants (MAF > 0.1%) following a comparison of EAPCTS point-wise heritability estimates of the AHI using different minimal MAFs. A grid search identified optimal GRM parameters with imputed data (MAF > 0.5%, $r^2 > 0.90$) using 929 ARIC individuals with imputation and WGS data. Log_{10} $p$-values using identical association test parameters had a Spearman’s $r$ correlation of 0.951 between WGS and imputed data. Matrices were constructed separately for each study + population combination (due to potentially differential imputation coverage).

Gene-based group sets considered Ensembl-defined non-pseudogenes expressed in any GTEx v7 tissue. Variants needed to clear a series of frequency, regional, functional class, and presumed functionality score filters in order to test a gene using its most biologically plausible variants. Variants could have a maximum minor allele frequency of 5%. Regions were largely exon-based. We also included variants located within experimentally derived promoter regions and Ensembl-derived Tarbase miRNA binding sites; and regulatory variants located within 1000 bases of a particular gene, including ChIP-seq determined transcription factor binding sites (TFBS), and Ensembl-derived CTCF, TFBS, and promoter sites [43–45]. Variants from a subset of 19 snpEff gene-based annotation functional classes (e.g., missense or nonsense, but not synonymous mutations) were considered. Finally, group set variants passing these prior filters were additionally filtered for the plausibility of biological function by requiring either a FATHMM-XF score > 0.5 or a CDTS < 1% constrained region score [46, 47]. Exonic variants could alternatively have a PrimateAI score > 0.803 or a Havrilla et al. < 1% constrained coding region score [48, 49].

Gene-based tests considered variants in WGS-only data. Pooled (across cohort) analyses were performed within each population in order to aggregate information on very rare variants across studies. Combined population results were obtained through meta-analysis of $p$-values weighted by sample size (due to potentially different MAF spectra driven by population demography). A significance level of $p < 4.51 \times 10^{-8}$ was used, reflecting a Bonferroni adjustment for all genes tested across all phenotype and population configurations.

A second set-based analysis was designed to query for TFBS annotation enrichment [50]. We performed 250-base pair sliding window analyses (to improve power by aggregating additional variants beyond an approximate ChIP-seq peak width of 100 base pairs). We filtered for variants with either a FATHMM-XF score > 0.5 or a CDTS 1% score with no MAF cut-offs and meta-analyzed MMSKAT results across the 4 populations, noting windows with $p$-values < 0.01. These intervals were tested for enrichment of ChIP-seq coordinates with at least 50% physical overlap for up to 437 transcription factors using ReMap 2018 v1.2 [51].

Single-variant EMMAX tests examined common variants (MAF > 0.5%). Meta-analysis across populations (and imputed genotype studies) used METAL with genomic control [52]. We performed bidirectional discovery and replication using the WGS and imputed samples (noting the high genomic resolution in the WGS samples and the higher sample size in the imputed data). We report results including at least 1000 individuals in discovery analyses, discovery association $p$-values < 1 x 10^{-5} and replication association $p$-values < 0.05. Therefore, no population-specific discovery analyses of Asian-Americans were performed. Multi-ethnic analyses included a minimum of two populations where a variant
cleared minimum MAF and imputation quality (for chip-based results) criteria. Significance was defined as \( p < 1 \times 10^{-8} \) in joint analyses, reflecting adjustment for five correlated phenotypes (Additional file 1: Table S3). We performed MetaXcan imputed GTEx gene expression analyses using joint EA results in selected tissues relevant to SDB and GIGSEA pathway analyses of MetaXcan output in whole blood (to maximize power), with empirical \( p \)-values incorporating 10,000 permutations [53, 54]. Bioinformatics annotations of single-variant results (Additional file 1: Table S7) include significant eQTL associations from GTEx v7, and overlapping promoter and enhancer coordinates derived from Roadmap Epigenomics, BLUEPRINT, and Vermunt et al. brain tissues (enhancers only) [55–58]. Lookups of potentially druggable genes as defined within DGIdb, a database of 56,000 drug-gene interactions from over 30 literature sources, were performed using the GeneCards suite [59, 60].

**Results**

**Study sample**

A study overview is provided in Additional file 2: Figure S1. Tables 1 and 2 provide a summary of the study samples and SDB traits analyzed using WGS and imputed genotypes, respectively. In total, there were 21,244 individuals (1942 AAs, 229 AsAs, 8341 EAs, and 10732 HAs). Median AHI levels ranged from mildly to moderately elevated, reflecting the age range and sex distribution of each cohort. Pairwise correlations of phenotypes and covariates are provided in Additional file 1: Table S3.

**Gene-based results**

Gene-based rare-variant results are presented in Table 3 (for meta-analyzed results across multiple populations) and in Table 4 (for secondary population-specific results). Collectively, we identified four significantly associated genes (Bonferroni \( p < 4.51 \times 10^{-8} \)). ARMCX3, identified in the multiple-population analysis, is an X-linked protein-coding that was associated with average desaturation (\( p = 5.29 \times 10^{-8} \)). Two protein-coding genes were identified in population-specific analyses of Per90: MRPS33 \( (p = 1.22 \times 10^{-5}) \) and C16orf90 \( (p = 1.36 \times 10^{-8}) \). We identified 12 suggestively associated genes (\( p \leq 4.22 \times 10^{-7} \)). Three genes are druggable [59, 60]. Nominally significant results \( (p < 0.01) \) and additional details are presented in Additional file 1: Tables S4 and S5. A list of individual variants comprising each gene is provided in Additional file 1: Table S6.

**Single-variant results**

We identified four genome-level significant loci in single-variant analyses (MAF > 0.5%; \( p < 1.0 \times 10^{-8} \); Table 5). In multiple-population analyses, the 2q12 locus \((rs77375846; IL18RAP)\) was associated with average event desaturation in a multiple-population analysis (combined \( p = 1.57 \times 10^{-9} \)) and minimum \( \text{SpO}_2 \) (consistent with a previous report [10]). Two novel population-specific loci were identified. The 8p12 locus \((rs35447033, NRG1)\) was associated with AHI in EAs (combined \( p = 3.02 \times 10^{-9} \), Fig. 1). The 5p13 locus \((rs287777; SLC45A2)\) was associated with average \( \text{SpO}_2 \) in EAs (combined \( p = 8.08 \times 10^{-10} \), Fig. 2). In HAs, the 1q32 locus \((rs116133558; ATP2B4)\) was associated with Per90 (combined \( p = 3.51 \times 10^{-10} \) and with average \( \text{SpO}_2 \) (as previously identified [9]). Twelve additional regions were suggestively associated (\( p < 1.0 \times 10^{-7} \)). Additional file 1: Table S7 provides additional context for all variants in these loci (\( p < 1.0 \times 10^{-7} \)), including imputation quality, significant eQTLs, and overlap with epigenetic regions. Lookups of loci that we have identified in prior publications [9–11] are provided in Additional file 1: Table S8. Manhattan and QQ plots corresponding to the significant associations are provided in Additional file 2: Figures S2–S5. GWAS summary statistics have been posted to the Broad Institute Sleep Disorders Research Portal (https://sleep.hugeamp.org/).

**MetaXcan imputed gene expression and GIGSEA pathway analyses**

We used joint WGS and imputed EA results to impute associations with gene expression levels using a MetaXcan framework for six tissues (subcutaneous and visceral omentum adipose, lung, monocytes, skeletal muscle, and whole blood). No individual tests reached Bonferroni significance (\( p < 2.60 \times 10^{-7} \); Additional file 1: Table S9). Genes that were observed in the top 10 results across the varied analyses (Additional file 1: Table S10) included ZNF83 (15 instances) and CHRNAE (13 instances).

Whole blood MetaXcan results (with the largest sample size) were further evaluated in GIGSEA-based pathway analyses. KEGG pathway results are shown in Additional file 1: Table S11. The most significantly associated pathway was KEGG_STEROID_HORMONE_BIOSYNTHESIS (average \( \text{SpO}_2 \) empirical \( p\)-value = 7.00 \( \times 10^{-6} \)). KEGG_RIG_I_LIKE_RECEPTOR_SIGNALING_PATHWAY was observed in the top 10 results for four of the five phenotypes. Gene-centric transcription factor binding site (TFBS) enrichment analysis results are presented in Additional file 1: Table S12. VSPEA3_Q6 (ETV4) was the most significantly associated TFBS (average desaturation empirical \( p\)-value = 3.00 \( \times 10^{-5} \)) and was the strongest association for AHI and minimum \( \text{SpO}_2 \) (empirical \( p\)-values 0.002 and 0.001, respectively). The most significant miRNA binding site enrichment analysis association was GCATTTC,MIR-105 (average \( \text{SpO}_2 \), \( p = 0.002 \); Additional file 1: Table S13). AGGCACT,MIR-515-3P (the strongest AHI association, \( p = 0.009 \)) was observed in the top ten results for four phenotypes.
Table 3 Lead gene-based multiple-population results

| Phenotype | Sex | Gene      | B38 positions | P      | N  | Variants | Population P | Population N | Population variants |
|-----------|-----|-----------|---------------|--------|----|----------|--------------|--------------|---------------------|
| Avg desaturation | All | ARMCX3 | X:101,623,082–101,625,765 | 3.48 × 10⁻⁹ | 5222 | 41 | 0.220, 0.179, 2.17 × 10⁻⁵, 8.93 × 10⁻⁴ | 1545; 227; 2994; 456 | 8, 5, 24, 9 |
| All | ARMCX3-AS1 | X:101,623,082–101,625,153 | 3.49 × 10⁻⁹ | 5222 | 38 | 0.225, 0.179, 2.19 × 10⁻⁵, 8.20 × 10⁻⁴ | 1545; 227; 2994; 456 | 7, 5, 23, 8 |
| Per90 | All | OR5K2 | 3:98,497,633–98,498,634 | 2.55 × 10⁻⁷ | 7986 | 7 | 0.143, 0.440, 4.14 × 10⁻⁷, 2.74 × 10⁻⁷ | 1712; 229; 3,242; 2803 | 4, 2, 1, 1 |
| Per90 | Females | ZZF1 | 174,004,409–4,144,018 | 4.22 × 10⁻⁷ | 4485 | 236 | 0.634, 0.337, 5.03 × 10⁻⁵, 3.05 × 10⁻⁵ | 1009; 114; 1702; 1600 | 85, 16, 87, 131 |

Lead MMSKAT gene-based results meta-analyzed across populations within one order of magnitude of significance (p < 4.51 × 10⁻⁵) are shown. Population-specific information for each gene is displayed in the latter columns for AA, AS, EA, and HA, respectively. Individual populations varied in the number of polymorphic variants available for testing (e.g., due to singletons or excessively common variants). ARMCX3-AS1 is a RNA gene that is anti-sense to the protein-coding ARMCX3 gene. Full results for genes with p < 0.01, including Ensembl-derived gene biotypes and descriptions, are provided in Additional file 1: Table S6. A list of individual variants comprising each gene is provided in Additional file 1: Table S6.

ChIP-seq transcription factor binding site interval enrichment

We performed a sliding window analysis to examine enriched intervals containing ChIP-seq derived coordinates for up to 437 transcription factors (Table 6, Additional file 1: Table S14). FOXP2 TFBS were consistently the most enriched for all phenotypes. Other notable transcription factors in the top 5 included EGR1, KDM4B, KDM6B, and TP63. KDM4B and KDM6B are druggable [59, 60]. Leading sliding window results are provided in Additional file 1: Table S15.

Discussion

Sleep-disordered breathing is associated with increased risk of a wide range of disorders, including cardiometabolic disease, cancer, cognitive impairment, and interstitial lung diseases, as well as premature mortality [2, 61]. Treatment options, however, are limited by a lack of knowledge of molecular pathways, including those that may be “druggable.” Recent analyses of SDB traits have focused on common variants and identified several preliminary genome-level significant associations [9–11], but did not address gene-based or rare-variant effects. Ten studies and over 21,000 individuals of multiple ancestries with WGS data at unprecedented resolution from the NHLBI TOPMed program combined with densely imputed data from other sources contributed to these results. We identified several variant, gene-based, and pathway-level associations. Analyses adjusted for obesity, a major SDB risk factor, identified loci and genes implicated in pulmonary, inflammatory, and craniofacial pathways. Some associations were population-specific, while others were sex-specific, consistent with population differences and strong sex differences for SDB [20, 62]. Notably, across multiple ancestral groups, we identified a set-based rare-variant association (p = 3.48 × 10⁻⁸) on chromosome X with ARMCX3.

Table 4 Lead gene-based population-specific results

| Phenotype | Model | Gene     | B38 positions | N  | Variants | Singletons | P         |
|-----------|-------|----------|---------------|----|----------|------------|-----------|
| Per90     | HA    | LIN01277 | 6:142,985,371–143,010,415 | 2803 | 2 | 0 | 5.02 × 10⁻⁸ |
|           |       | OR5K2   | 3:98,497,633–98,498,634 | 2803 | 1 | 0 | 2.74 × 10⁻⁷ |
| AA females |       | S100A16* | 1:153,607,528–153,616,353 | 1009 | 1 | 1 | 2.07 × 10⁻⁷ |
| EA females |       | CSMD2-AS1 | 1:33,867,977–33,885,456 | 1009 | 1 | 1 | 2.07 × 10⁻⁷ |
|           |       | MRPS33  | 7:141,006,422–141,014,911 | 1702 | 9 | 8 | 1.22 × 10⁻⁹ |
|           |       | LINC01811 | 3:34,170,921–34,558,474 | 1702 | 6 | 5 | 9.71 × 10⁻⁸ |
|           |       | NELFCD* | 20:58,980,722–58,995,761 | 1702 | 12 | 10 | 3.32 × 10⁻⁷ |
|           |       | SLC22A8* | 11:62,988,399–63,015,986 | 1702 | 3 | 3 | 3.58 × 10⁻⁷ |
| HA females |       | AL132709.1 | 14:101,077,452–101,077,758 | 1660 | 2 | 0 | 1.41 × 10⁻⁷ |
|           |       | EPHX4   | 19:202,944–92,063,474 | 1660 | 12 | 10 | 3.48 × 10⁻⁷ |
| HA males  |       | C16orf90 | 16:3,493,483–3,496,479 | 1143 | 6 | 3 | 1.36 × 10⁻⁸ |
|           |       | TVP2B3  | 17:18,781,270–18,806,714 | 1143 | 4 | 4 | 2.53 × 10⁻⁷ |
|           |       | IPCEF1  | 6:154,154,536–154,356,890 | 1143 | 10 | 8 | 4.07 × 10⁻⁷ |

Lead MMSKAT gene-based population-specific associations within one order of magnitude of significance (p < 4.51 × 10⁻⁵) are shown. The Variants column indicates the number of filtered polymorphic variants with minor allele frequency < 5% available for testing, a portion of which were singletons. *Druggable gene [59, 60]. Full results for genes with p < 0.01, including descriptions, are provided in Additional file 1: Table S5. A list of individual variants comprising each gene is provided in Additional file 1: Table S6.
## Table 5 Lead single-variant analysis results

| Region | Phenotype | Model | SNP | WGS/Chip N | CAF | WGS beta (SE) | WGS P | Chip beta (SE) | Chip P | Combined beta (SE) | Combined P |
|--------|-----------|-------|-----|------------|-----|---------------|-------|--------------|-------|-----------------|-----------|
| 2q12.1: IL18RAP | Avg desaturation | All | rs77375846 C | 4995, 4838 | 0.028-0.019 | -0.152 (0.049) | 1.87 × 10^-3 | -0.264 (0.049) | 5.97 × 10^-4 | -0.208 (0.035) | 1.57 × 10^-3 |
| 2q33.3: PRAF68 | Avg desaturation | All | rs60132122 T | 5222, 4838 | 0.308-0.037 | 0.062 (0.031) | 0.043 | 0.195 (0.034) | 6.26 × 10^-4 | 0.122 (0.023) | 6.49 × 10^-4 |
| 11q12.2: M54A15 | Avg SpO2 | All | rs4939452 C | 7929, 1319 | 0.347-0.024 | 0.066 (0.023) | 4.34 × 10^-3 | 0.066 (0.014) | 3.29 × 10^-6 | 0.064 (0.012) | 4.87 × 10^-8 |
| 18q12.3: UNC00907 | Avg SpO2 | All | rs187880354 G | 4500, 7391 | 0.006-0.022 | 0.442 (0.146) | 2.36 × 10^-3 | 0.432 (0.097) | 8.53 × 10^-4 | 0.496 (0.081) | 7.04 × 10^-4 |
| 2q12.1: IL18RAP | Min SpO2 | All | rs13899820 G | 7705, 1319 | 0.025-0.131 | 0.510 (0.184) | 5.58 × 10^-3 | 0.654 (0.128) | 3.36 × 10^-5 | 0.607 (0.105) | 7.93 × 10^-9 |
| 1p12.23: DAB2 | Min SpO2 | Females | rs1145307 CA | 4450, 6202 | 0.138-0.014 | 0.651 (0.140) | 3.34 × 10^-6 | 0.338 (0.102) | 8.63 × 10^-4 | 0.446 (0.082) | 5.73 × 10^-8 |
| 1q21.12: UNC024064 | Min SpO2 | Females | rs2176909 T | 4450, 6202 | 0.724-0.090 | 0.828 (0.157) | 1.38 × 10^-7 | 0.319 (0.116) | 5.77 × 10^-3 | 0.498 (0.093) | 9.06 × 10^-8 |
| 5p13.3: C5orf22 | Per90 | Males | rs10940956 A | 3502, 7043 | 0.470-0.079 | 0.930 (0.422) | 2.74 × 10^-2 | 1.430 (0.269) | 1.09 × 10^-7 | 1.285 (0.227) | 1.48 × 10^-8 |
| 9q22.1: DEND4AC | AHI | AA | rs11165400 A | 1717, 225 | 0.016-0.018 | -1.124 (2.268) | 7.18 × 10^-7 | -18.10 (6.724) | 7.07 × 10^-1 | -11.94 (2.149) | 2.74 × 10^-8 |
| 1q31.2: ADH5-4650.1 | AHI | AA | chr1:11919650_H/G/A | 1717, 225 | 0.286-0.031 | 3.078 (0.641) | 1.56 × 10^-6 | 5.080 (1.759) | 3.88 × 10^-3 | 3.313 (0.602) | 3.75 × 10^-8 |
| 8q12.2: ACO686721.1, NRG1 | AHI | EA | rs35447033 T | 3266, 5095 | 0.060-0.094 | 2.247 (0.621) | 2.95 × 10^-4 | 2.453 (0.521) | 2.54 × 10^-6 | 2.368 (0.399) | 3.02 × 10^-9 |
| 3p13.2: SLC45A2 | Avg SpO2 | EA | rs28777 A | 3201, 5024 | 0.885-0.069 | -0.526 (0.133) | 8.00 × 10^-5 | -0.454 (0.096) | 2.23 × 10^-5 | -0.478 (0.078) | 8.08 × 10^-10 |
| 1q32.1: ATP2B4 | Avg SpO2 | HA | rs116133558 T | 2803, 7956 | 0.006-0.014 | 0.371 (0.120) | 2.08 × 10^-3 | 0.294 (0.062) | 2.15 × 10^-6 | 0.310 (0.055) | 1.88 × 10^-8 |
| 1q33.3: Intergenic (RNU6-755P) | Min SpO2 | HA | rs140743827 A | 2803, 7174 | 0.017-0.030 | 0.502 (0.539) | 1.13 × 10^-2 | -1.770 (0.367) | 1.42 × 10^-2 | -1.696 (0.312) | 5.51 × 10^-8 |
| 1q22.1: ATP2B4 | Per90 | HA | rs116133558 T | 2803, 7956 | 0.006-0.014 | -1.005 (0.450) | 2.54 × 10^-2 | -1.218 (0.207) | 4.15 × 10^-5 | -1.181 (0.188) | 3.51 × 10^-5 |
| 11p12.1: Intergenic (AC104010.1) | Avg SpO2 | HA males | chr11:446529795_CT/T | 1143, 3024 | 0.007-0.008 | 0.686 (0.248) | 5.65 × 10^-3 | 0.710 (0.154) | 3.83 × 10^-10 | 0.703 (0.131) | 7.25 × 10^-8 |
| 10q22.1: HK1 | Min SpO2 | EA males | rs1740636 C | 1523, 3560 | 0.072-0.115 | 1.215 (0.392) | 1.94 × 10^-3 | 1.099 (0.235) | 2.81 × 10^-6 | 1.120 (0.201) | 2.01 × 10^-8 |
| 8q23.2: KCN1 | Min SpO2 | EA males | rs58365105 A | 1523, 3560 | 0.007-0.026 | -2.878 (0.864) | 8.65 × 10^-4 | -2.406 (0.540) | 8.36 × 10^-4 | -2.539 (0.458) | 2.96 × 10^-8 |
| 2q35: AC019211.1 | Per90 | EA males | chr22:103096683_G/A | 1540, 187 | 0.005-0.006 | 12.280 (2.431) | 4.38 × 10^-7 | 17.505 (7.989) | 2.85 × 10^-2 | 12.723 (2.326) | 4.48 × 10^-8 |

Lead EMMAX single-variant associations within one order of magnitude of significance (combined \( p < 1.00 \times 10^{-8} \)) and with replication evidence (\( p < 0.05 \)) are shown. Full results for all variants in each locus with \( p < 1.00 \times 10^{-8} \), including additional associations with secondary models, and metadata and annotations, are provided in Additional file 1: Table S7.
Fig. 1 Regional plot of the rs35447033 association with AHI in European-ancestry individuals. Joint WGS and imputed results are shown, using Build 38 coordinates on the X-axis. Log-transformed \( p \)-values are shown on the Y-axis. Variant colors indicate the degree of linkage disequilibrium with the lead variant rs35447033.

Fig. 2 Regional plot of the rs28777 association with average \( \text{SpO}_2 \) in European-ancestry individuals. Joint WGS and imputed results are shown, using Build 38 coordinates on the X-axis. Log-transformed \( p \)-values are shown on the Y-axis. Variant colors indicate the degree of linkage disequilibrium with the lead variant rs28777.
Gene-based results

Across multiple populations, ARMCX3 (ALEX3) and the RNA anti-sense gene ARMCX3-AS1 were associated with apnea-hypopnea triggered intermittent hypoxia. ARMCX3 regulates mitochondrial aggregation and trafficking in multiple tissues and facilitates neuronal survival and axon regeneration [63–65]. Wnt signaling regulates reactive oxygen species (ROS) generation and ARMCX3-associated mitochondrial aggregation [64, 66]. Potential mechanisms for further study include sensitized carotid body chemoreflexes, interaction with inflammatory mechanisms, and neuronal dysfunction within respiratory centers. Sleep apnea and reduced ventilatory drive are enriched in individuals with a primary mitochondrial disorder [67]. Mitochondria are an important source of ROS, which modulate the acute hypoxic ventilatory response. Mitochondria impact HIF1A signaling and may contribute to oxygen sensing [68, 69]. ROS are required for intermittent hypoxia-induced respiratory long-term facilitation [70]. These effects may mitigate the level of hypoxia resulting from recurrent apneas, or conversely, lead to ventilatory instability, promoting apnea occurrence. Mitochondrial ROS also activate the NLRP3 inflammasome in multiple pulmonary diseases, consistent with an inflammation model that includes our IL18-pathway and HK1 results, ROS-related proinflammatory responses to lung capillary pressure, and evidence of alveolar epithelial injury/SDB interactions [10, 69, 71–73]. Our findings suggest value in investigating the mechanisms by which ARMCX3 predisposes to SDB, and whether these associations are mediated by neuronal dysfunction and/or ROS and carotid body sensitization, and interact with the inflammasome.

Additional genes were significantly associated in population-specific analyses, including the mitochondrial ribosomal gene MRPS33. Mitoribosomes are responsible

| Phenotype                | Transcription factor | # Observed overlap | # Expected overlap | −log10 (E-value) |
|--------------------------|----------------------|--------------------|--------------------|------------------|
| AHI                      | FOXP2                | 588                | 36.20              | 473.99           |
|                          | KDM6B                | 630                | 51.58              | 435.29           |
|                          | THAP1                | 505                | 31.89              | 402.07           |
|                          | KLF9                 | 745                | 91.81              | 395.52           |
|                          | TP63                 | 997                | 182.22             | 383.85           |
| Average desaturation     | FOXP2                | 493                | 22.32              | 460.00           |
|                          | THAP1                | 439                | 19.55              | 412.76           |
|                          | UBTF                 | 489                | 28.20              | 407.50           |
|                          | TP63                 | 788                | 109.36             | 382.89           |
|                          | KDM6B                | 482                | 30.98              | 380.39           |
| Average SpO2             | FOXP2                | 582                | 35.87              | 468.89           |
|                          | KDM6B                | 613                | 51.21              | 418.65           |
|                          | EGR1                 | 664                | 66.76              | 404.83           |
|                          | UBTF                 | 574                | 46.35              | 399.91           |
|                          | KDM4B                | 489                | 29.56              | 398.10           |
| Min SpO2                 | FOXP2                | 561                | 35.57              | 445.57           |
|                          | THAP1                | 515                | 31.32              | 417.89           |
|                          | KDM6B                | 569                | 50.87              | 373.41           |
|                          | UBTF                 | 536                | 45.99              | 360.56           |
|                          | EGR1                 | 602                | 66.25              | 346.03           |
| Per90                    | FOXP2                | 689                | 39.05              | 578.42           |
|                          | KDM6B                | 739                | 54.79              | 539.69           |
|                          | TP63                 | 1199               | 193.28             | 515.44           |
|                          | THAP1                | 607                | 34.47              | 509.33           |
|                          | EGR1                 | 786                | 72.09              | 507.27           |

Two-hundred-fifty-base pair sliding window coordinates with association $p < 0.01$ were queried for interval enrichment of ChIP-seq-derived transcription factor binding sites using the ReMap annotation tool. ChIP-seq coordinates were required to have >50% overlap with a sliding window interval. ReMap-derived expected overlaps are obtained from the equivalent number of similarly sized random regions. E-value indicates the expected value, with a higher log-transformed value indicating greater enrichment. Full results are provided in Additional file 1: Table S14.
for the expression of the 13 essential components of the oxidative phosphorylation system, and a majority of the small subunit proteins have been implicated in disease [74]. The expression of several small and large subunit proteins are altered in a hypoxic environment [75]. MRPS33 expression varies with oxygen treatment in COPD [76].

Single-variant results
We identified four common frequency associated loci, including multiple-population associations with the IL18RAP region. The IL18RAP region has been associated with minimum SpO2 [10], and here we further identify an association with average event desaturation, highlighting a role in an OSA-specific trait. Multiple variants in this region are also GTEx eQTL variants for both interleukin-18 receptor subunits IL18RAP and IL18R1 (Additional file 1: Table S7) and experimental studies support a role for IL18 signaling in mediating this association, possibly through effects of pulmonary inflammation on gas exchange (reviewed in [10]).

We identified three population-specific loci, including two novel associations in individuals of European ancestry (Figs. 1 and 2). Sixty-five variants in the NRG1 region were associated with the AHI (p < 1.0 × 10⁻⁸, Additional file 1: Table S7). This region was suggestively associated with sleep apnea in a Korean population [77]; however, the lead signals appear to be independent (rs10097555 Korean p = 2.6 × 10⁻⁶, EA p = 0.91). NRG1 is associated with lung development and acute lung injury and mediates inflammasome-induced alveolar cell permeability [78–80]. NRG1 promotes accumulation of HIF1A and has increased expression in vascular smooth muscle cells following exposure to intermittent hypoxia [81, 82]. The lead SLC45A2 region variant rs28777 (average SpO2 p = 8.08 × 10⁻¹⁰) has been associated with multiple traits and is in a splicing regulatory element with extreme population differentiation [83]. An association in the ATP2B4 region with average SpO2 in HAs [9] has been extended to a second hypoxemia trait at the same variant (Per90 p = 3.31 × 10⁻¹⁰). This gene is the main cellular membrane calcium pump in erythrocytes and also regulates vascular tone [84, 85].

Pathway analyses
Several gene pathways were identified in EA individuals using imputed gene expression in whole blood (Additional file 1: Table S11). KEGG_RIG_I_LIKE_RECEPTOR_SIGNALING_PATHWAY (retinoic acid-inducible gene I-like) was the most commonly observed, occurring in the top 10 results for 4 of the 5 phenotypes. This pathway initiates the immune response to RNA virus infection [86], consistent with a role for inflammation at the NRG1 and IL18RAP loci. Steroid hormone biosynthesis (the most significantly associated pathway), PPAR signaling, and metabolism (via “starch and sucrose metabolism”) suggest the importance of biological pathways modulating energy homeostasis and balance and metabolic function [87]. In the gene-centric GIGSEA TFBS analysis, VSPEA3_Q6 (ETV4) was the lead association for three phenotypes. ETV4 influences branching in the developing lung and regulates hypoxia-inducible factor signaling [88, 89], a major mechanism influencing ventilatory control.

Transcription factor binding site enrichment
Several transcription factors were identified through interval enrichment of observed TFBS across the genome (Table 6). FOXP2 was consistently the most enriched transcription factor and is known to regulate gene expression in epithelial lung tissue and response to lung injury through an inflammatory mechanism [90, 91]. FOXP2 is also expressed in brainstem respiratory areas including the pre-Bötzing complex (which is essential for respiratory rhythemogenesis) and impacts airway morphology [92, 93]. Two lysine demethylases (KDM4B and KDM6B) were also identified. KDM6B (JMJD3) is required for a functional pre-Bötzing complex [94, 95] and reduced KDM6B protein expression was reported in hypoxic OSA patients [96]. Kdm6b also plays roles in immune function and lung development [97–99]. Drosophila Kdm6b knock-outs have increased sleep [100]. KDM4B (JMJD2B) and KDM6B are both members of the JmJ-C protein domain family and are regulated by HIF1A, require oxygen as a cofactor, and act as oxygen sensors for chromatin in hypoxia [101, 102]. EGR1 mediates hypoxia-induced pulmonary fibrosis [103]. TP63 is associated with cleft palate in Tp63 deficient mice, which is associated with an increased prevalence of OSA [104, 105], suggesting that its relationship to OSA may be through pathways influencing craniofacial development. Among the leading 250-base pair sliding window results (Additional file 1: Table S15), 4:105708751-105709001 (Per90 HA p = 2.72 × 10⁻⁶) is of note due to regional associations with lung function and expression in the human lung [106].

Strengths and weaknesses
This study is the first genome-wide analysis of objectively measured SDB traits using deep sequencing. Together with improved imputation quality, the TOPMed resource has enabled unprecedented genetic resolution. We examined clinically relevant phenotypes measured using rigorous methodology [2, 14]. We analyzed data from 10 studies of individuals from four population groups that used different ascertainment strategies, which may potentially improve the generalization of our results. While this analysis is among the largest performed for SDB traits to date, our moderate sample size has lower power to detect weaker associations, and data were not available to replicate these first rare-variant associations. We did not specifically study the central apnea-hypopnea index due to the relatively low prevalence of central sleep apnea (< 2%)
in these largely community-based studies [36, 37]. While there are multiple lines of evidence in the literature to support our findings, additional experimental follow-up analyses are required.

Conclusions

We have identified the first rare-variant and additional common-variant associations at genome-level significance with objectively measured SDB traits in humans. The results point to biologically relevant pathways for sleep-disordered breathing and related comorbidities.

Abbreviations

AA: African-American; AsA: Asian-American; BMI: Body mass index; CCGD: Centers for Common Disease Genomics; EA: European-American; Australian; GWAS: Genome-wide association study; HA: Hispanic/Latino-American; MAF: Minor allele frequency; OSA: Obstructive sleep apnea; Per90: Percent of the sleep recording with oxyhemoglobin saturation < 90%; ROS: Reactive oxygen species; SDB: Sleep-disordered breathing; SpO2: Oxyhemoglobin saturation; TFBS: Transcription factor binding site; TOPMed: Trans-Omics for Precision Medicine; WGS: Whole-genome sequencing

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s13073-021-00917-8.

Additional file 1: Table S1. NHLBI TOPMed Consortium. Table S2. NHLBI TOPMed Consortium Sleep Working Group. Table S3. Pairwise Phenotype and Covariate Correlations. Table S4. MMKSAT gene-based multiple-population results (p < 0.01). Table S5. MMKSAT gene-based population-specific results (p < 0.01). Table S6. Lead MMKSAT result variants. Table S7. Single-variant analysis results for lead loci. Table S8. Lookups of previously reported GWAS results. Table S9. MetaXcan imputed gene expression results. Table S10. Lead genes in multiple MetaXcan results. Table S11. GIGSEA KEGG pathway results. Table S12. GIGSEA MethDB transcription factor binding site enrichment results. Table S13. GIGSEA MethDB mirRNA binding site enrichment results. Table S14. Sliding window analysis transcription factor binding analysis enrichment. Table S15. Leading sliding window analysis results.

Additional file 2: Figure S1. Study Overview. Figure S2. NRG1 Locus Models Manhattan and QQ Plots. Figure S3. SLCO4A2 Locus Models Manhattan and QQ Plots. Figure S4. IL1RAP Locus Models Manhattan and QQ Plots. Figure S5. ATP2B4 Locus Models Manhattan and QQ Plots.

Acknowledgements

We acknowledge our TOPMed Consortium and TOPMed Sleep Traits Working Group collaborators, who are listed in Additional file 1: Tables S1 and S2. The authors wish to thank the participants and study staff of all of our cohorts for their important contributions. We gratefully acknowledge the studies and participants who provided biological samples and data for TOPMed. The authors thank the staff and participants of the ARIC study for their important contributions. A full list of principal CHS investigators and institutions can be found at CHS-NHLBI.org. The Framingham Heart Study thanks the study participants and the multitude of investigators who over its 70-year history continue to contribute so much to further our knowledge of heart, lung, blood, and sleep disorders and associated traits. The authors thank the staff and participants of HCHS/SOL for their important contributions. The views expressed in this manuscript are those of the authors and do not necessarily represent the views of the National Heart, Lung, and Blood Institute; the National Institutes of Health; or the US Department of Health and Human Services. This manuscript was not approved by the HCHS/SOL publications committee. Investigator’s website—http://www.cscce.unc.edu/hchs/. The authors also wish to thank the staff and participants of the JHS. We thank the field staff in Starr County for their careful collection of these data and are especially grateful to the participants who so graciously cooperated and gave of their time.

Authors’ contributions

Conception and design: B.E.C., E.A.B., L.A.C., R.C.K., K.E.N., B.M.P., J.I.R., S.S.R., R.P.T., R.S.V., and S.R. Data acquisition: B.E.C., J.L., T.S., M.Z., H.C., S.A.G., D.J.G., J.M.L., J.L., X.L., H.M., S.R.P., S.M.P., R.S., N.A.S., H.W., X.Z., D.S.E., C.L.H., D.R.H., S.M., L.J.P., K.L.S., G.J.T., and S.R. Analysis: all authors. Interpretation, draft and review, and final approval: all authors. B.E.C. and S.R. had full access to the study data and take responsibility for the integrity of the data and accuracy of analyses.

Funding

Whole-genome sequencing (WGS) for the Trans-Omics in Precision Medicine (TOPMed) program was supported by the National Heart, Lung, and Blood Institute (NHLBI). WGS for “NHLBI TOPMed: Trans-Omics for Precision Medicine Whole Genome Sequencing Project: ARIC” (phs001211.v1.p1) was performed at Baylor College of Medicine Human Genome Sequencing Center (HHSN268201500015C and 3U54HG003273-1252) and the Broad Institute of MIT and Harvard (3R01HL092577-06). WGS for “NHLBI TOPMed: The Cleveland Family Study (WGS)” (phs000954.v2.p1) was performed at the University of Washington Northwest Genomics Center (3R01HL098433-05S1). WGS for “NHLBI TOPMed: Cardiovascular Health Study” (phs001368.v1.p1) was performed at Baylor College of Medicine Human Genome Sequencing Center (HHSN268201500015C). WGS for “NHLBI TOPMed: Whole Genome Sequencing and Related Phenotypes in the Framingham Heart Study” (phs000974.v3.p2) was performed at the Broad Institute of MIT and Harvard (3R01HL092577-06S1). WGS for “NHLBI TOPMed: Hispanic Community Health Study/Study of Latinos (HCHS/SOL)” (phs001395) was performed at the Baylor College of Medicine Human Genome Sequencing Center (HHSN268201500015C and 3U54HG003273-1252). WGS for “NHLBI TOPMed: The Jackson Heart Study” (phs000964.v3.p1) was performed at the University of Washington Northwest Genomics Center (HHSN268201100037C). WGS for “NHLBI TOPMed: NHLBI TOPMed: MESA” (phs001416.v1.p1) was performed at the Broad Institute of MIT and Harvard (3U54HG003067-1). Centralized read mapping and genotype calling, along with variant quality metrics and filtering were provided by the TOPMed Informatics Research Center (3R01HL-117672-02S1). Phenotype harmonization, data management, sample-identity QC, and general study coordination were provided by the TOPMed Data Coordinating Center (3R01HL-120393-02S1). The Genome Sequencing Program (GSP) was funded by the National Human Genome Research Institute (NHGRI), the National Heart, Lung, and Blood Institute (NHLBI); and the National Eye Institute (NEI). The GSP Coordinating Center (U24 HG008956) contributed to cross-program scientific initiatives and provided logistical and general study coordination. The Centers for Common Disease Genomics (CCDG) program was supported by NHGRI and NHLBI, and CCDG-funded whole-genome sequencing of the ARIC and HCHS/SOL studies was performed at the Baylor College of Medicine Human Genome Sequencing Center (UM1 HG008898 and R01HL059367). Brian Cade is supported by grants from the National Institutes of Health [K01-HL135405-01, R01-HL113338-04, R35-HL135818-01] and the American Thoracic Society Foundation (http://foundation.thoracic.org). Susan Redline is partially supported by grants from the National Institutes of Health [R35-HL135818-01, R01-HL113338-04]. Sanjay Patel has had grant support through his institution from the ResMed Foundation, the American Sleep Medicine Foundation, Bayer Pharmaceuticals, and Philips Respironics. James Wilson is supported by US4GM113528 from the National Institute of General Medical Sciences. The Atherosclerosis Risk in Communities (ARIC) study has been funded in whole or in part with Federal funds from the National Heart, Lung, and Blood Institute, National Institutes of Health, Department of Health and...
Human Services (contract numbers HHSN268201700001I, HHSN268201700002I, HHSN268201700003I, HHSN268201700004I, and HHSN268201700005I), H101HL087641, R01HL095367, and R01HL086694; National Human Genome Research Institute contract U10HG004402; and National Institutes of Health contract HHSN2682012000036C. Infrastructure was partly supported by Grant Number UL1RR025005, a component of the National Institutes of Health and NIH Roadmap for Medical Research.

This Cardiovascular Health Study (CHS) research was supported by NHLBI contracts R01HL080295, R01HL103621, R01HL085251, and R01HL120393 with additional contribution from the National Institute of Neurological Disorders and Stroke (NINDS). Additional support was provided through R01AG023629 from the National Institute on Aging (NIA). The Cleveland Family Study has been supported by National Institutes of Health grants (R01-HL046380, KL2-RR024990, R35-HL33518, and R01-HL113330). The Framingham Heart Study (FHS) has been supported by contracts N01-HC-55222, N01-HC-55223, N01-HC-55224, N01-HC-55225, N01-HC-55226, N01-HC-55227, N01-HC-55228, N01-HC-55229, N01-HC-55230, N01-HC-55231, N01-HC-55232, N01-HC-55233, N01-HC-55234, N01-HC-55235, N01-HC-55236, N01-HC-55237, N01-HC-55238, N01-HC-55239, N01-HC-55240, N01-HC-55241, N01-HC-55242, N01-HC-55243, N01-HC-55244, and N01-HC-55245.

The Genetic Analysis Center at Washington University was supported by NHLBI and NIDCR contracts (HHSN268201300005C AM03 and MOD03). The Jackson Heart Study (JHS) is supported and conducted in collaboration with Jackson State University (HHSN2682018000138), Tougaloo College (HHSN26820180000148), the Mississippi State Department of Health (HHSN2682018000015H/HHSN26820000001), and the University of Mississippi Medical Center (HHSN2682018000101/HHSN268201800111). The MESA Project, funded by Johnson & Johnson. KLS has grant funding from Merck. BMP serves on the Steering Committee of the Yale Open Data Access Project, funded by Johnson & Johnson. KLS has grant funding from Merck.

The remaining authors declare that they have no competing interests.}

Availability of data and materials
Variant-level meta-analysis data are available for visualization and download at the Sleep Disorders Knowledge Portal: https://sleephupeamp.org/.

Declarations
Ethics approval and consent to participate
This study was approved by the Mass General Brigham Institutional Review Board (IRB) (2010P001765) and the Institutional Review Boards from the institutions of all participating studies. All patients have given their written informed consent. The ARC study was approved by the IRBs of the University of North Carolina at Chapel Hill, University of Minnesota, and Johns Hopkins University. Cleveland Family Study was approved by the IRBs of Case Western Reserve University and Mass General Brigham (formerly Partners HealthCare). The CHS study was approved by the IRBs [or ethics review committee] of University Washington, University of Pittsburgh, and University of California, Sacramento. The Framingham Heart Study was approved by the IRBs of the Boston University Medical Center. The JHS study was approved by Jackson State University, Tougaloo College, and the University of Mississippi Medical Center IRBs. The HCHS/SOL study was approved by the IRBs at each field center, and by the Non-Biomedical IRB at the University of North Carolina at Chapel Hill, to the HCHS/SOL Data Coordinating Center. All IRBs approving the study are Non-Biomedical IRB at the University of North Carolina at Chapel Hill, Chapel Hill, NC; Einstein IRB at the Albert Einstein College of Medicine of Yeshiva University, Bronx, NY; IRB at Office for the Protection of Research Subjects (OPRS), University of Illinois at Chicago, Chicago, IL; Human Subject Research Office, University of Miami, Miami, FL; and Institutional Review Board of San Diego State University, San Diego, CA. The MESA study was approved by the IRBs at The Lundquist Institute (formerly Los Angeles BioMedical Research Institute) at Harbor-UCLA Medical Center, University of Washington, Wake Forest School of Medicine, Northwestern University, University of Minnesota, Columbia University, and Johns Hopkins University. The MoS study was approved by the IRBs of the San Francisco Coordinating center (UCSF and California Pacific Medical Center), University of Alabama at Birmingham, University of Minnesota, Stanford University, University of Pittsburgh, Oregon Health and Science University, and University of California, San Diego. The Starr County study was approved by the IRBs at the University of Texas Health Science Center at Houston. The WASHS study was approved by the Human Research Ethics Committee of Sir Charles Gairdner Hospital, as well as The University of Western Australia, the Confidentiality of Health Information Committee, and the Human Research Ethics Committee of the Western Australian Department of Health. The research was conducted in strict compliance with the Declaration of Helsinki.

Consent for publication
Not applicable.

Competing interests
The authors disclose the following industry funding, which they believe is unrelated to this study. SRP has had grant support through his institution from the ResMed Foundation, Bayer Pharmaceuticals, and Philips Respironics. BMP serves on the Steering Committee of the Yale Open Data Access Project. Funded by Johnson & Johnson. KLS has grant funding from Merck. The remaining authors declare that they have no competing interests.
Author details
1Division of Sleep and Circadian Disorders, Brigham and Women’s Hospital, Harvard Medical School, 221 Longwood Avenue, Boston, MA 02115, USA.
2Division of Sleep Medicine, Harvard Medical School, Boston, MA 02115, USA.
3Program in Medical and Population Genetics, Broad Institute, Cambridge, MA 02142, USA.
4Department of Medicine, University of Maryland School of Medicine, Baltimore, MD 21201, USA.
5Human Genetics Center, Department of Epidemiology, Human Genetics and Environmental Sciences, School of Public Health, The University of Texas Health Science Center at Houston, Houston, TX 77030, USA.
6Center for Precision Health, School of Public Health and School of Biomedical Informatics, The University of Texas Health Science Center at Houston, Houston, TX 77030, USA.
7Computational Medicine Core, Center for Lung Biology, UW Medicine Sleep Center, Division of Pulmonary, Critical Care and Sleep Medicine, University of Washington, Seattle, WA 98195, USA.
8VA Boston Healthcare System, Boston, MA 02132, USA.
9The Institute for Translational Genomics and Population Sciences, Department of Pediatrics, The Lundquists Institute for Biomedical Innovation at Harbor-UCLA Medical Center, Torrance, CA 90502, USA.
10Center for Genomic Medicine and Department of Anesthesia, Pain, and Critical Care Medicine, Massachusetts General Hospital, Boston, MA 02114, USA.
11Department of Population and Quantitative Health Sciences, School of Medicine, Case Western Reserve University, Cleveland, OH 44106, USA.
12Department of Biostatistics, Harvard T.H. Chan School of Public Health, Boston, MA 02115, USA.
13Department of Data Science, University of Mississippi Medical Center, Jackson, MS 39216, USA.
14Division of Pulmonary, Allergy, and Critical Care Medicine, University of Pittsburgh, Pittsburgh, PA 15213, USA.
15Division of Pulmonary, Critical Care and Sleep Medicine, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA.
16California Pacific Medical Center Research Institute, San Francisco, CA 94107, USA.
17Department of Pulmonary Physiology and Sleep Medicine, Sir Charles Gairdner Hospital, Perth, Western Australia 6009, Australia.
18Sleep Health Service, Respiratory and Sleep Services, Southern Adelaide Local Health Network, Adelaide, South Australia, Australia.
19Adelaide Institute for Sleep Health, Flinders University, Adelaide, South Australia, Australia.
20School of Public Health, University of Adelaide, South Australia 5000, Australia.
21Department of Biostatistics and Center for Statistical Genetics, University of Michigan School of Public Health, Ann Arbor, MI 48109, USA.
22Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX 77030, USA.
23Department of Medicine, University of Mississippi Medical Center, Jackson, MS 39216, USA.
24Jackson Heart Study, Jackson, MS 39216, USA.
25Department of Biostatistics, Boston University School of Public Health, Boston, MA 02218, USA.
26Framingham Heart Study, Framingham, MA 01702, USA.
27Department of Epidemiology and Population Health, Albert Einstein College of Medicine, Bronx, New York 10461, USA.
28Department of Genomics Sciences, University of Washington, Seattle, WA 98195, USA.
29Northwest Genomics Center, Seattle, WA 98105, USA.
30Department of Epidemiology and Carolina Center of Genome Sciences, University of North Carolina, Chapel Hill, NC 27514, USA.
31Cardiovascular Health Study, Departments of Medicine, Epidemiology, and Health Services, University of Washington, Seattle, WA 98101, USA.
32Kaiser Permanente Washington Health Research Institute, Seattle, WA 98101, USA.
33Center for Public Health Genomics, University of Virginia, Charlottesville, VA 22908, USA.
34Department of Pathology, University of Vermont, Colchester, VT 05405, USA.
35Sections of Preventive Medicine and Epidemiology and Cardiology, Department of Medicine, Boston University School of Medicine, Boston, MA 02118, USA.
36Department of Epidemiology, Boston University School of Public Health, Boston, MA 02118, USA.
37Department of Physiology and Biophysics, University of Mississippi Medical Center, Jackson, MS 39216, USA.
38Division of Pulmonary, Critical Care, and Sleep Medicine, Beth Israel Deaconess Medical Center, Boston, MA 02215, USA.

Received: 7 January 2020 Accepted: 28 May 2021
Published online: 26 August 2021

References
1. Ancoli-Israel S, Yaffe K, Kamerow DB. The familial aggregation of obstructive sleep apnea: a decade-long historical cohort study. PLoS Med. 2014;11(2):e1001599. https://doi.org/10.1371/journal.pmed.1001599.
2. Cade BE, Chen H, Stilp AM, Gleason KJ, Sofer T, Ancoli-Ills, S et al. Multiethnic meta-analysis identifies RA1 as a possible obstructive sleep apnea-related quantitative trait locus in men. Am J Respir Cell Mol Biol. 2018;58(3):391–401. https://doi.org/10.1165/rcmb.2017-0237OC.
3. Deng S, Forer L, Schönherr S, Sidore C, Locke A, Kwong A, et al. Next-generation genotype imputation service and methods. Nat Genet. 2016;48:1284–7. https://doi.org/10.1038/ng.3515.
4. Feinleib M. The Framingham Study: sample selection, follow-up, and methods of analyses. Nat Cancer Inst Monogr. 1985;67:59–64.
5. Quan SF, Howard BV, Iber C, Kiley JP, Nieto FJ, O’Connor GT, et al. The Sleep Heart Health Study: design, rationale, and methods. Sleep. 1995;18(9):673–80. https://doi.org/10.1093/oxfordjournals.scp.a115184.
6. Fried LP, Borhani NO, Enright P, Furberg CD, Gardin JM, Kronmal RA, et al. The Atherosclerosis Risk in Communities (ARIC) study: design and implementation of the Hispanic Community Health Study/Study of Latinos. Am J Epidemiol. 2002;156(9):871–9. https://doi.org/10.1093/aje/kwf113.
7. Chen H, Cade BE, Gleason KJ, Bjonnes AC, Stilp AM, Sofer T, et al. Genome-wide association studies of obstructive sleep apnea-related quantitative trait loci. Am J Respir Crit Care Med. 2018;197(3):687–98. https://doi.org/10.1164/rccm.201712-2431OC.
8. Chen H, Cade BE, Gleason KJ, Bjonnes AC, Stilp AM, Sofer T, et al. Genome-wide association studies of obstructive sleep apnea-related quantitative trait loci. Am J Respir Crit Care Med. 2018;197(3):687–98. https://doi.org/10.1164/rccm.201712-2431OC.
102. Hancock RL, Masson N, Dunne K, Flashman E, Kawamura A. The activity of JmjC histone lysine demethylase KDM4A is highly sensitive to oxygen concentrations. ACS Chem. Biol. 2017;12(4):1011–9. https://doi.org/10.1021/acscchembio.6b00958.

103. Yan SF, Zou YS, Gao Y, Zhai C, Mackman N, Lee SL, et al. Tissue factor transcription driven by Egr-1 is a critical mechanism of murine pulmonary fibrin deposition in hypoxia. Proc Natl Acad Sci U S A. 1998;95:8298–303.

104. Thomason HA, Dixon MJ, Dixon J. Facial clefting in Tp63 deficient mice results from altered Bmp4, Fgf8 and Shh signaling. Dev Biol. 2008;321(1):273–82. https://doi.org/10.1016/j.ydbio.2008.06.030.

105. Robison JG, Otteson TD. Increased prevalence of obstructive sleep apnea in patients with cleft palate. Arch. Otolaryngol. Head Neck Surg. 2011;137(3):259–74. https://doi.org/10.1001/archoto.2011.8.

106. Obeidat M, Miller S, Probert K, Billington CK, Hodge E, et al. GSTCD and INTS12 regulation and expression in the human lung. PLoS One. 2013;8(9):e74630. https://doi.org/10.1371/journal.pone.0074630.

Publisher’s Note
Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.