Genome-wide analyses in 1,987,836 participants identify 39 genetic loci associated with sleep apnoea

Adrian I. Campos[1,2], Nathan Ingold[1,3], Yunru Huang[4,¶], Pik-Fang Kho[1,3], Xikun Han[1,2], Jue-Sheng Ong[1], Luis M. García-Marín[1], 23andMe Research Team[4], Matthew H. Law[1], Nicholas G. Martin[1], Xianjun Dong[5], Gabriel Cuellar-Partida[2,§], Stuart MacGregor[1], Stella Aslibekyan[4] and Miguel E. Rentería[1,2,3,5,6,*]

1. Department of Genetics & Computational Biology, QIMR Berghofer Medical Research Institute, Brisbane, QLD, Australia

2. Faculty of Medicine, The University of Queensland, Brisbane, QLD, Australia

3. School of Biomedical Sciences, Faculty of Health, Queensland University of Technology, Brisbane, Queensland, Australia.

4. 23andMe, Inc., Sunnyvale, CA, USA

5. Department of Neurology, Brigham & Women's Hospital and Harvard Medical School, Boston, MA, USA

6. Global Brain Health Institute, University of California, San Francisco, San Francisco, CA, USA

¶. Current address: Genentech Inc., South San Francisco, CA, USA

§. Current address: 23andMe, Inc., Sunnyvale, CA, USA

* **Correspondence:** Dr Miguel E. Rentería, QIMR Berghofer Medical Research Institute, Locked Bag 2000, Royal Brisbane Hospital, Herston, QLD 4029, Australia. Email: miguel.renteria@qimrberghofer.edu.au

Author Contributions: MER, and SM designed the study with aid from AIC, NI, XH, ML,JSO, GCP and NGM. MER, ML and SM supervised the study. NI performed sleep apnoea and Snoring GWAS for UK-B and CLSA. PFK and GCP performed the COLOC expression analyses. MER and XD performed the Partner's Biobank GWAS. AIC performed the AGDS GWAS, the meta and MTAG analyses, as well as post-GWAS gene based tests, genetic correlation and PRS analyses with aid from GCP and LMGM. YH and SA performed the replication GWAS. All co-authors contributed to drafting and provided relevant input for this manuscript.

Funding: National Health and Medical Research Council, Australia.

Short Title: GWAS of sleep apnoea.

Subject Category: 15.3 Genetics of Sleep Disorders.

Word Count: 3,475

Supplementary Material: This article has an online data supplement, which is accessible from this issue's table of content online at www.atsjournals.org

ABSTRACT

Rationale: Sleep apnoea is a complex disorder characterised by periods of halted breathing during sleep. Despite its association with serious health conditions such as cardiovascular disease, the aetiology of sleep apnoea remains understudied, and previous genetic studies have failed to identify replicable genetic risk factors.

Objective: To advance our understanding of factors that increase susceptibility to sleep apnoea by identifying novel genetic associations.

Methods: We conducted a genome-wide association study (GWAS) meta-analysis of sleep apnoea across five cohorts, and a previously published GWAS of apnoea-hypopnea index (N_{Total}=510,484). Further, we used multi-trait analysis of GWAS (MTAG) to boost statistical power, leveraging the high genetic correlations between apnoea, snoring and body mass index. Replication was performed in an independent sample from 23andMe, Inc (N_{Total}=1,477,352; N_{cases}=175,522)

Measurements and Main Results: Our results revealed 39 independent genomic loci robustly associated with sleep apnoea risk, and significant genetic correlations with multisite chronic pain, sleep disorders, diabetes, high blood pressure, osteoarthritis, asthma and BMI-related traits. We also derived polygenic risk scores for sleep apnoea in a leave-one-out independent cohort and predicted probable sleep apnoea in participants (OR=1.15 to 1.22; variance explained = 0.4 to 0.9%).

Conclusions: We report novel genetic markers robustly associated with sleep apnoea risk and substantial molecular overlap with other complex traits, thus advancing our understanding of the underlying biological mechanisms of susceptibility to sleep apnoea.

Abstract Word Count: 234.

Keywords: Sleep apnoea, genetics, CLSA, UK Biobank, FinnGen.

INTRODUCTION

Sleep apnoea is a sleep disorder characterised by episodes of halted breathing during sleep, which in turn leads to frequent arousal and intermittent hypoxia(1). The most common type of sleep apnoea is *obstructive sleep apnoea*, which affects 5-47% of adults and 1% of children(2, 3). *Central sleep apnoea* affects ~0.9% of adults and is also characterised by a lack of breathing and oxygenation during sleep, but without the characteristic respiratory efforts observed in obstructive sleep apnoea(3, 4). Remarkably, the prevalence of sleep apnoea is often underestimated because those affected are unable to gain awareness about their condition or may confuse it with habitual snoring.

Obstructive sleep apnoea is caused by an augmented function of pharyngeal dilator muscles, brought about by the onset of sleep, causing the collapse of the upper airways and subsequent hypopnea or apnoea(5). Central sleep apnoea can be caused by a lower respiratory drive, opioid use, obesity, lesions to the central nervous system, and impaired reactivity to CO2 among others(4). A high degree of overlap between the sleep apnoea subtypes makes them hard to distinguish. Sleep apnoea has been linked to mental and physical fatigue, increased risk of motor accidents(6), and decreased mental wellbeing and overall quality of life(7). In addition, sleep apnoea has also been associated with increased risk of hypertension(8), stroke(9) and increased reactive oxygen species in blood, which in turn increases oxidative stress in the body(10, 11).

The proportion of risk for sleep apnoea that can be attributed to genetic factors (i.e. heritability) has been estimated between 35 and 75%(12, 13). Familial aggregation seems to be partially

independent from weight(14). Understanding of the genetic architecture of sleep apnoea could help generate risk prediction models and provide important groundwork for developing interventions and informing the development of therapies. Similarly, having information on the effect of genetic variants on sleep apnoea risk might enable us to evaluate the association between several postulated risk factors and sleep apnoea through Mendelian Randomisation (MR) analysis(15). While some candidate gene studies for sleep apnoea have reported some potentially associated genes(16, 17), genome-wide association studies (GWAS) have failed to replicate those associations(18–20). In fact, GWAS have identified few genetic loci robustly associated (i.e. with a replication in an independent cohort) with sleep apnoea to date.

Sleep apnoea is a highly polygenic trait, with hundreds to thousands of variants of small effect contributing to its genetic liability, and uncovering its genetic architecture has been challenging, due in part to the fact that sleep apnoea is recognised as an underdiagnosed condition(21, 22). The number of sleep apnoea cases in existing large population datasets is low. In a sample of 500,000 individuals the expected number of sleep apnoea cases (assuming a prevalence of ~5%) would be ~25,000. However, in the UK Biobank (~500,000 individuals), only ~8000 sleep apnoea cases have been recorded. Underdiagnosis further reduces statistical power because a proportion of cases would be included in the control group in a standard analysis.

In this study, we perform a GWAS meta-analysis of sleep apnoea across five cohorts and a previously published apnoea-hypopnea index GWAS(19). Moreover, we employ Multi-trait Analysis of Genome-Wide Association Summary Statistics (MTAG) to combine our sleep

apnoea meta-analysis, a Snoring meta-analysis across 5 cohorts and BMI to boost power by leveraging the high genetic correlations between sleep apnoea, snoring and BMI(23). We characterise the genetic architecture of sleep apnoea by performing gene based tests and genetic correlation analyses. Finally, the different GWAS results were validated by polygenic prediction using a leave-one cohort out (LOO) scheme, and by assessing replication in an independent sample from 23andMe, Inc. We refer to our results broadly as *sleep apnoea*, although they likely represent a closer proxy to obstructive sleep apnoea given its higher prevalence.

METHODS

Sample information and phenotype ascertainment

This study analysed GWAS from five cohorts from the UK (UK Biobank; UKB), Canada (Canadian Longitudinal Study of Aging; CLSA)(24, 25), Australia (Australian Genetics of Depression Study; AGDS), USA (F) and Finland (FinnGen). A previously published apnoea-hypopnea index (AHI) GWAS(19) was also included. Sample sizes, number of cases and number of controls are available in **Table 1**. For each cohort, sleep apnoea cases were defined using either self-reported diagnostic items, a self-reported diagnosis or a general practitioner diagnosis based on ICD10 codes (**Supplementary Methods**). To increase power, and to better define sleep apnoea cases from controls, snoring cases were excluded from the analyses for the UKB, CLSA and AGDS cohorts. An overview of the analysis pipeline used for sleep apnoea loci discovery is available in **Supplementary Figure 1**.

GWAS

For each cohort a GWAS was performed using logistic regression models, and including random effects to account for cryptic relatedness where relevant (**Supplementary Methods**). All studies were adjusted for age, sex, batch (where relevant), and genetic ancestry principal components derived from genotype data. Standard sample- and variant-level quality control filters were applied. Variants were excluded from the analyses if they had a low minor allele frequency (MAF<0.01) or low imputation quality score (INFO<0.6). Individuals were excluded based on excess missingness, heterozygosity, or evidence from a deviation from European ancestry based on genetic principal components. For the UKB, we used the summary statistics from a GWAS

for snoring we previously published(26). FinnGen GWAS results for sleep apnoea and snoring were obtained from the open access FinnGen resource (http://r3.finngen.fi/). Apnoea-hypopnea index GWAS(19) summary statistics on a subset of European individuals were downloaded from a publicly available repository(27).

GWAS meta analyses

Sample-size weighted (P-value based) meta-analyses for sleep apnoea and snoring were performed across the five cohorts described above using METAL (v2020-05-05). Studies were weighted according to their effective sample size as described by the formula: $Neff = 4/(1/Ncases + 1/Ncontrols)$ as recommended for studies with different levels of ascertainment. The sleep apnoea meta-analysis also included the summary statistics for a published apnoea-hypopnea index on a subset of participants of European descent(19) which was weighted by the total sample size (**Supplementary Methods**).

Multi-trait GWAS analyses

Multi-trait Analysis of Genome-Wide Association Summary Statistics (MTAG) was used to boost power for discovering sleep apnoea associated loci. MTAG performs a generalized meta-analysis of GWAS summary statistics for different, but genetically correlated traits(28). We performed two MTAG analyses: the first one combining our sleep apnoea and snoring meta-analyses, and the second one including both sleep apnoea and snoring meta-analyses and a body-mass index (BMI) GWAS from the UK-Biobank.

Gene-based tests and eQTL colocalisation

Gene-based association analyses were conducted on both the sleep apnoea meta and MTAG GWAS using the "set-based association analysis for human complex traits" fastBAT method. FastBAT performs a set-based enrichment analysis based on the GWAS summary statistics while accounting for LD between SNPs. Statistical significance was defined using the Bonferroni method for multiple testing correction. Genes identified as statistically significant were further assessed for eQTL colocalisation. Briefly, we integrated our GWAS summary data with *cis*-eQTL data from whole blood, esophagus, adipose and lung tissue from GTEx V8 to estimate the posterior probability that GWAS signals co-occur with eQTL signals while accounting for LD structure. This method estimates the posterior probabilities for five different scenarios. The scenario of interest is colocalistion due to associations with both traits through the same SNPs (PP4). A threshold of PP4>=0.8 was considered as evidence for co-occurrence of GWAS signals and eQTL signals at the region of interest (**Supplementary Methods**).

Heritability and genetic correlations

We used LDSC to estimate the SNP-based heritability (h_{SNP}^2) for the sleep apnoea meta-analysis. Given that samples were not particularly ascertained for sleep apnoea, we assumed the overall sample and population prevalence for sleep apnoea to be the sleep apnoea prevalence estimated across cohorts (0.05) which is consistent with reported epidemiological estimates(2). Genetic correlations (rG) between sleep apnoea and 1,522 phenotypes (with available GWAS summary statistics) were estimated using bivariate LDSC regression in CTG-VL(29) based on a common

set of HapMap3 variants. The Benjamini-Hochberg FDR at 5% was used to define statistical significance.

Polygenic risk scoring

To assess the external validity of the GWAS, we performed polygenic based prediction on a target sample of 9,221 unrelated Australian Adults from the Australian Genetics of Depression Study(30) (AGDS) with complete data. Briefly, the meta and MTAG analyses were repeated leaving out the AGDS cohort to avoid sample overlap. We employed a recently developed method, SBayesR, to obtain the conditional effects of the studied variants, thus avoiding inflation arising from using correlated SNPs due to LD(31). Sleep apnoea polygenic risk scores (SA_{PRS}) were calculated using PLINK 1.9. The association between PRS and sleep apnoea in AGDS was assessed using a logistic regression model (python *statsmodels*). SA_{PRS} was the predictive variable of interest, with age, sex and the first 10 genetic principal components included as covariates. Finally, binary classifiers based on a logistic regression were built including either age and sex (base model) or age, sex and the PRS of interest (SA_{PRS}, $SAmtagSnoring_{PRS}$ or $SAmtagSnoringBMI_{PRS}$). These classifiers were used to further assess the polygenic predictive ability. The sample was divided randomly into training and testing (left out of training) datasets of equal sizes. Then, the classifier's ability to predict sleep apnoea was assessed using the area under the receiver operating characteristic (ROC) curve. To avoid potential biases from the random division of training and testing datasets, the procedure was repeated 100 times to estimate a mean area under the curve (**Supplementary Methods**).

23andMe replication GWAS

We sought replication of variants identified in the discovery phase in the 23andMe sample (N=1,477,352 participants). Cases were ascertained based on the question "*Have you ever been diagnosed with, or treated for any of the following conditions*?" with one of the choices being "*Sleep apnoea*" (Yes= 175,522; No = 1,301,830). Methods and results from this GWAS were presented at the 2018 American Society for Human Genetics annual conference(32). Briefly, logistic regressions using sleep apnoea as the dependent variable while adjusting for sex, age, BMI, genetic principal components and genotype platform were used for the GWAS. Participants provided informed consent and participated in the research online, under a protocol approved by the external AAHRPP-accredited IRB, Ethical & Independent Review Services (E&I Review). Only unrelated participants of European ancestry who provided informed consent were included in the analysis. The full GWAS summary statistics for the 23andMe discovery data set will be made available through 23andMe to qualified researchers under an agreement with 23andMe that protects the privacy of the 23andMe participants. Please visit research.23andme.com/collaborate/#publication for more information and to apply to access the data.

RESULTS

Phenotype prevalence

After data quality control, the prevalence of both sleep apnoea and snoring showed some variation between the different cohorts included in this study. sleep apnoea prevalence ranged from 2 to 18% while snoring affected 6-43% of the samples (**Table 1**). The prevalence of sleep apnoea and snoring across all cohorts was 5 and 34% respectively.

GWAS meta-analysis

Our meta-analysis identified six independent (LD $r^2<0.05$) genome-wide significant loci associated with sleep apnoea (**Figure 1a**). The signals spanned chromosomes 5, 11, 12 and 16 near genes *BDNF*, *AC091980.1, ANKRD31*, *KDM2B*, *FTO* and *PRIM1* (**Supplementary Figure 2**). The SNP based heritability on the liability scale was 0.10 (S.E.=0.0067). LD-score regression suggested inflation ($\boldsymbol{\lambda}_{GC}$=1.21) was due to polygenic signal (intercept= 1.013, S.E.=0.008) rather than population stratification.

MTAG

We performed two MTAG analyses to boost statistical power. The first one (SAmtagSnoring) included a snoring meta-analysis performed across the same five cohorts used for the sleep apnoea meta-analysis. The results identified 40 independent genome-wide significant loci associated with sleep apnoea (**Figure 1b**). The second MTAG (SAmtagSnoringBMI) included BMI in addition to sleep apnoea and snoring. BMI is robustly correlated with sleep apnoea (rG=0.56) before MTAG. This analysis identified 154 independent genome-wide significant loci

(**Figure 1c**). Effect size and direction of the independent hits were highly consistent across the meta-analysis and the mtag analyses (R^2>0.85 **Supplementary Figure 3**).

Gene-based tests and colocalisation

The gene-based association analyses identified 30, 146 and 529 genes beyond genome-wide significance threshold for the sleep apnoea meta-analysis, the SAmtagSnoring, and the SAmtagSnoringBMI respectively (**Supplementary Figure 4** and **Supplementary Data**). Some of these genes were located within the same locus and in high LD. Thus, to identify genes linked to sleep apnoea through potential changes in gene expression, we performed eQTL colocalisation analyses for the genes identified above. Of the 458 genes with available eQTL data, 74 showed strong evidence of eQTL colocalisation with sleep apnoea, SAmtagSnoring or SAmtagSnoringBMI (**Figure 2**). Of the genes associated with the sleep apnoea meta-analysis, three (*PRIM1*, *KDM2B* and *SPTBN2*) showed evidence for eQTL colocalisation with sleep apnoea. Eighteen genes, including *ABO*, *SKAP1* and *ACTN3* among others showed evidence for eQTL colocalisation with SAmtagSnoring (**Figure 2** and **Supplementary Data**).

Genetic correlations

Bivariate LDSC was used to assess the genetic correlation between sleep apnoea and other complex traits. The trait with the highest genetic correlation with sleep apnoea was a sleep apnoea GWAS performed on the UK-Biobank from a public GWAS repository (http://www.nealelab.is/uk-biobank/); this is essentially a subset of the UK-Biobank GWAS used in our meta-analysis. Other genetically correlated traits included respiratory diseases, type 2

diabetes, obesity, stroke, depression and smoking history among others (**Supplementary Data**). The sleep apnoea meta-analysis and the SAmtagSnoring analyses showed a highly concordant pattern of genetic correlations. While also showing overall agreement, the SAmtagSnoringBMI results showed higher genetic correlations with BMI derived traits such as obesity and stroke (**Figure 3**).

Polygenic risk scoring

PRS based on either of our results were significantly associated with sleep apnoea in a leave one out polygenic prediction analysis (**Table 2**). Odds ratios (OR) per standard deviation of PRS increased with the number of hits. For example, the meta-analysis based PRS (SA_{PRS}) showed an OR = 1.15 (1.08-1.21), whereas the PRS based on the SAmtagSnoringBMI showed an OR=1.22 (1.15-1.29). A similar pattern was observed for variance explained and significance (**Table 2**). Participants on the highest PRS quintile showed between 50 to 75% higher odds of reporting sleep apnoea depending on the PRS used (**Figure 4a**). Classifier models showed higher area under the curve, and thus higher polygenic prediction ability, for the MTAG GWAS (**Figure 4b** and **c** and **Supplementary Figure 5**) compared to the meta-analysis alone.

Independent sample replication

We sought replication of our GWAS results using an independent sample (N= 1,477,352) from 23andMe (**see Methods**). Overall, sixteen of the independent variants identified by our analyses showed evidence of association beyond the genome-wide significance threshold (p<5e-8; **Supplementary Data**) in the replication sample. Moreover, 39 variants replicated after multiple

testing correction (**Table 3**). A high replication rate was observed for the sleep apnoea meta analysis; five out of the six loci (80%) were replicated. For SAmtagSnoring, 30 variants out of 39 (with replication data) were replicated. A smaller replication rate was observed for the SAmtagSnoringBMI analysis, as only 21 out of 152 SNPs with data, were replicated. This was not surprising given that the 23andMe GWAS had been adjusted for BMI (see **Discussion**). Overall, there was a large concordance in the direction and magnitude of effect sizes (R^2=0.86, 0.89 and 0.35 for sleep apnoea, SAmtagSnoring and SAmtagSnoringBMI respectively) of our analyses and the 23andMe replication results (**Supplementary Figure 6**).

DISCUSSION

This study combined data across five different cohorts and a published study. Our sleep apnoea GWAS meta-analysis identified six independent loci and 30 genes robustly associated with sleep apnoea. We used MTAG to boost statistical power and identify more genetic loci associated with sleep apnoea by combining our meta-analysis results with i) a snoring meta-analysis also performed herein, and ii) our snoring meta-analysis and a GWAS on BMI, both of which are genetically correlated with sleep apnoea.

Our results did not replicate any of the previously reported sleep apnoea genome-wide hits(18–20). In fact, all of them showed $p>0.2$ for sleep apnoea. Two of them (rs12936587 and rs4731117) showed a nominally significant association ($p\sim0.046$) with SAmtagSnoring. This is not surprising given substantially lower power of previous studies. Furthermore, there are several methodological differences that could explain the lack of replication. The previous studies used different apnoea related phenotypes, distinguishing between sleep phases and on samples of distinct ancestries. Despite not replicating previous results, the evidence for polygenic prediction as well as the replication rate of our results in the 23andMe sample suggest our study identified loci robustly associated with sleep apnoea.

We identified thirty genes associated with sleep apnoea through summary-based gene association analyses using our sleep apnoea meta-analysis results. The same approach yielded 146 and 529 genes for the MTAG analyses. We could not replicate previously reported candidate gene associations such as *TNFA*, *APOE*, *PTGER3* and *LPAR1(16)*. This could be explained by

methodological differences between our analysis and those reporting the candidate genes. For example, the *LPAR1* association was observed in participants of African ancestry(33). Nonetheless, studies assessing the support for candidate gene associations using GWAS have found poor consistency on disorders such as depression and schizophrenia. Our results suggest a similar trend for candidate gene-studies for sleep apnoea. Our study should be powered enough to detect effect sizes as big as OR=2.01 which has been reported for *TNFA(34)*. Future studies should perform a systematic evaluation of candidate gene studies and GWAS concordance in sleep apnoea, an objective that was outside of the scope of the current study.

Consistent with a well-defined link between BMI and sleep apnoea, our meta-analysis identified genes such as *FTO, GCNT4, KDM2B, LGR4, LINC00678, TRIM66, ANKRD31, TEF* and *BDNF-AS* that have been previously linked to BMI. We also identified overlapping genes for height (PRIM1, *HSD17B6*, *KDM2B* and *ZC3H7B*), osteoarthritis (*CHADL*, *SPTBN2* and *C11orf80*), type II diabetes (*KDM2B*, *ANKRD31* and *FTO*), smoking (*HSD17B6*, *BDNF-AS*, *TEF* and *USP4*), and alcohol consumption (*LINC00678*, *TRIM66*, *BDNF-AS*). In line with our gene-based tests, we identified sleep apnoea to be genetically correlated with complex traits such BMI, Hiatus hernia, type II diabetes, obesity, stroke, hypertension and depression among others. These results are consistent with known risk factors and known comorbidities of sleep apnoea.

This study was performed using cohorts of European ancestry. Thus generalisations to and comparisons with other ancestry groups should be performed with caution. In order to maximize sample size, we included cohorts with different definitions of sleep apnoea, including

self-reported diagnosis and self-reported diagnostic items. While this is warranted to achieve the required sample sizes for GWAS, heterogeneity might be introduced by minimal phenotyping. Ideally, any sleep apnoea study would ascertain cases employing a robust measure such as the apnoea-hypopnea index or oxygen saturation; GWAS of complex traits require very large sample sizes, making such an approach challenging. Although MTAG has proven successful for boosting loci discovery, combining traits with extreme power differences might inflate signals related to the most powered phenotype(28). In our study, performing an MTAG including BMI seemed to affect the pattern of genetic correlations, particularly increasing the correlations with BMI related traits such as stroke, obesity and BMI. Replication of the SAmtagSnoringBMI results was lower than our other two analyses. This result is not surprising given two reasons: First, the GWAS performed by 23andMe adjusted for BMI as a covariate. As such, it is identifying hits that are independent to those for which the SAmtagSnoringBMI is boosting power. Second, the new hits identified by the SAmtagSnoringBMI analysis had small effect sizes, which would not have been identified without boosting power using BMI. These associations are naturally harder to replicate. Overall we identified highly concordant effect sizes between our analyses and the 23andMe GWAS, and provided evidence of replication for 39 independent hits associated with sleep apnoea. Finally, polygenic prediction of sleep apnoea was the highest for the PRS derived from SAmtagSnoringBMI GWAS. Thus suggesting the SAmtagSnoringBMI results are a reliable proxy for sleep apnoea as long as the relationship with BMI is accounted for.

In summary, we performed a meta-analysis for sleep apnoea across five European-ancestry cohorts and a published AHI GWAS identifying six independent genome-wide significant loci. Multi-trait analyses including either snoring, or snoring and BMI, identified 40 and 154 independent genome-wide significant loci respectively. Replication in an independent sample was achieved for 39 variants. All analyses showed significant polygenic prediction of sleep apnoea in a leave-one-out polygenic risk-scoring analysis. The MTAG analyses showed greater polygenic prediction than the meta-analysis suggesting an increase in polygenic signal for sleep apnoea. The associations identified here can be used as instrumental variables in targeted Mendelian randomisation studies aiming at understanding the relationship between sleep apnoea and hypothesised causally related traits. Identifying robust loci associated with sleep apnoea is an important step towards a deeper biological understanding, which can translate into novel treatments, and risk assessment strategies.

ACKNOWLEDGEMENTS

The opinions expressed in this manuscript are the author's own and do not reflect the views of the Canadian Longitudinal Study on Aging or any affiliated institution. This project used data from the UK-Biobank under application number 25331. We want to acknowledge Partners HealthCare Biobank for providing samples, genomic data, and health information data. We want to acknowledge the participants and investigators of the FinnGen study. This research was made possible using the data/biospecimens collected by the Canadian Longitudinal Study on Aging (CLSA). Funding for the Canadian Longitudinal Study on Aging (CLSA) is provided by the Government of Canada through the Canadian Institutes of Health Research (CIHR) under grant reference: LSA 94473 and the Canada Foundation for Innovation. This research has been conducted using the CLSA dataset [Baseline Comprehensive Dataset version 4.0, Follow-up 1 Comprehensive Dataset version 1.0], under Application Number 190225. The CLSA is led by Drs. Parminder Raina, Christina Wolfson and Susan Kirkland. Data collection for the Australian Genetics of Depression Study was possible, thanks to funding from the Australian National Health & Medical Research Council (NHMRC) to N.G.M. (GNT1086683). A.I.C. is supported by UQ Research Training Scholarships from The University of Queensland (UQ). P.F.K. is supported by an Australian Government Research Training Program Scholarship from Queensland University of Technology (QUT). M.E.R. thanks support of the NHMRC and Australian Research Council (GNT1102821). SM is supported by a research fellowship from the Australian National Health and Medical Research Council (NHMRC). This work was supported by an NHMRC project grant [1123248]. We would like to thank the research participants and employees of 23andMe for making this work possible. The following members of the 23andMe

Research Team contributed to this study: Michelle Agee, Stella Aslibekyan, Adam Auton, Elizabeth Babalola, Robert K. Bell, Jessica Bielenberg, Katarzyna Bryc, Emily Bullis, Briana Cameron, Daniella Coker, Devika Dhamija, Sayantan Das, Sarah L. Elson, Teresa Filshtein, Kipper Fletez-Brant, Pierre Fontanillas, Will Freyman, Pooja M. Gandhi, Karl Heilbron, Barry Hicks, David A. Hinds, Karen E. Huber, Ethan M. Jewett, Yunxuan Jiang, Aaron Kleinman, Katelyn Kukar, Keng-Han Lin, Maya Lowe, Marie K. Luff, Jennifer C. McCreight, Matthew H. McIntyre, Kimberly F. McManus, Steven J. Micheletti, Meghan E. Moreno, Joanna L. Mountain, Sahar V. Mozaffari, Priyanka Nandakumar, Elizabeth S. Noblin, Jared O'Connell, Aaron A. Petrakovitz, G. David Poznik, Anjali J. Shastri, Janie F. Shelton, Jingchunzi Shi, Suyash Shringarpure, Chao Tian, Vinh Tran, Joyce Y. Tung, Xin Wang, Wei Wang, Catherine H. Weldon, Peter Wilton.

REFERENCES

1. Lévy P, Kohler M, McNicholas WT, Barbé F, McEvoy RD, Somers VK, Lavie L, Pépin J-L. Obstructive sleep apnoea syndrome. *Nat Rev Dis Primers* 2015;1:15015.
2. Veasey SC. *Encyclopedia of Neuroscience*. In: Squire LR, editor. Elsevier; 2009.
3. Donovan LM, Kapur VK. Prevalence and Characteristics of Central Compared to Obstructive Sleep Apnea: Analyses from the Sleep Heart Health Study Cohort. *Sleep* 2016;39:1353–1359.
4. Eckert DJ, Jordan AS, Merchia P, Malhotra A. Central sleep apnea: Pathophysiology and treatment. *Chest* 2007;131:595–607.
5. Mezzanotte WS, Tangel DJ, White DP. Influence of sleep onset on upper-airway muscle activity in apnea patients versus normal controls. *Am J Respir Crit Care Med* 1996;153:1880–1887.
6. Teran-Santos J, Jimenez-Gomez A, Cordero-Guevara J, Burgos--Santander CG. The association between sleep apnea and the risk of traffic accidents. *N Engl J Med* 1999;340:847–851.
7. Engleman HM, Douglas NJ. Sleep\textperiodcentered 4: Sleepiness, cognitive function, and quality of life in obstructive sleep apnoea/hypopnoea syndrome. *Thorax* 2004;59:618–622.
8. Lavie P, Herer P, Hoffstein V. Obstructive sleep apnoea syndrome as a risk factor for hypertension: population study. *BMJ* 2000;320:479–482.
9. Munoz R, Duran-Cantolla J, Martínez-Vila E, Gallego J, Rubio R, Aizpuru F, De La Torre G. Severe sleep apnea and risk of ischemic stroke in the elderly. *Stroke* 2006;37:2317–2321.
10. Simiakakis M, Kapsimalis F, Chaligiannis E, Loukides S, Sitaras N, Alchanatis M. Lack of

effect of sleep apnea on oxidative stress in obstructive sleep apnea syndrome (OSAS) patients. *PLoS One* 2012;7:e39172.

11. Alonso-Fernández A, García-Río F, Arias MA, Hernanz A, de la Peña M, Piérola J, Barceló A, López-Collazo E, Agustí A. Effects of CPAP on oxidative stress and nitrate efficiency in sleep apnoea: a randomised trial. *Thorax* 2009;64:581–586.
12. Szily M, Tarnoki AD, Tarnoki DL, Kovacs DT, Forgo B, Lee J, Kim E, Sung J, Kunos L, Meszaros M, Muller V, Bikov A. Genetic influences on the onset of obstructive sleep apnoea and daytime sleepiness: a twin study. *Respir Res* 2019;20:125.
13. Casale M, Pappacena M, Rinaldi V, Bressi F, Baptista P, Salvinelli F. Obstructive sleep apnea syndrome: from phenotype to genetic basis. *Curr Genomics* 2009;10:119–126.
14. Redline S, Tosteson T, Tishler PV, Carskadon MA, Millman RP. Studies in the genetics of obstructive sleep apnea. Familial aggregation of symptoms associated with sleep-related breathing disturbances. *Am Rev Respir Dis* 1992;145:440–444.
15. Burgess S, Butterworth A, Thompson SG. Mendelian randomization analysis with multiple genetic variants using summarized data. *Genet Epidemiol* 2013;37:658–665.
16. Mukherjee S, Saxena R, Palmer LJ. The genetics of obstructive sleep apnoea. *Respirology* 2018;23:18–27.
17. Kalra M, Pal P, Kaushal R, Amin RS, Dolan LM, Fitz K, Kumar S, Sheng X, Guha S, Mallik J, Deka R, Chakraborty R. Association of ApoE genetic variants with obstructive sleep apnea in children. *Sleep Med* 2008;9:260–265.
18. Farias Tempaku P, Leite Santoro M, Bittencourt L, D'Almeida V, Iole Belangero S, Tufik S. Genome-wide association study reveals two novel risk alleles for incident obstructive

sleep apnea in the EPISONO cohort. *Sleep Med* 2020;66:24–32.

19. Chen H, Cade BE, Gleason KJ, Bjonnes AC, Stilp AM, Sofer T, Conomos MP, Ancoli-Israel S, Arens R, Azarbarzin A, Bell GI, Below JE, Chun S, Evans DS, Ewert R, Frazier-Wood AC, Gharib SA, Haba-Rubio J, Hagen EW, Heinzer R, Hillman DR, Johnson WC, Kutalik Z, Lane JM, Larkin EK, Lee SK, Liang J, Loredo JS, Mukherjee S, *et al.* Multiethnic Meta-Analysis Identifies RAI1 as a Possible Obstructive Sleep Apnea-related Quantitative Trait Locus in Men. *Am J Respir Cell Mol Biol* 2018;58:391–401.
20. Cade BE, Chen H, Stilp AM, Gleason KJ, Sofer T, Ancoli-Israel S, Arens R, Bell GI, Below JE, Bjonnes AC, Chun S, Conomos MP, Evans DS, Johnson WC, Frazier-Wood AC, Lane JM, Larkin EK, Loredo JS, Post WS, Ramos AR, Rice K, Rotter JI, Shah NA, Stone KL, Taylor KD, Thornton TA, Tranah GJ, Wang C, Zee PC, *et al.* Genetic Associations with Obstructive Sleep Apnea Traits in Hispanic/Latino Americans. *Am J Respir Crit Care Med* 2016;194:886–897.
21. Fuhrman C, Fleury B, Nguyên X-L, Delmas M-C. Symptoms of sleep apnea syndrome: high prevalence and underdiagnosis in the French population. *Sleep Med* 2012;13:852–858.
22. Costa LE, Uchôa CHG, Harmon RR, Bortolotto LA, Lorenzi-Filho G, Drager LF. Potential underdiagnosis of obstructive sleep apnoea in the cardiology outpatient setting. *Heart* 2015;101:1288–1292.
23. Main C, Liu Z, Welch K, Weiner G, Jones SQ, Stein K. Surgical procedures and non-surgical devices for the management of non-apnoeic snoring: a systematic review of clinical effects and associated treatment costs. *Health Technol Assess* 2009;13:iii, xi–xiv, 1–208.

24. Raina PS, Wolfson C, Kirkland SA, Griffith LE, Oremus M, Patterson C, Tuokko H, Penning M, Balion CM, Hogan D, Wister A, Payette H, Shannon H, Brazil K. The Canadian longitudinal study on aging (CLSA). *Can J Aging* 2009;28:221–229.

25. Raina P, Wolfson C, Kirkland S, Griffith LE, Balion C, Cossette B, Dionne I, Hofer S, Hogan D, van den Heuvel ER, Liu-Ambrose T, Menec V, Mugford G, Patterson C, Payette H, Richards B, Shannon H, Sheets D, Taler V, Thompson M, Tuokko H, Wister A, Wu C, Young L. Cohort Profile: The Canadian Longitudinal Study on Aging (CLSA). *Int J Epidemiol* 2019;48:1752–1753j.

26. Campos AI, García-Marín LM, Byrne EM, Martin NG, Cuéllar-Partida G, Rentería ME. Insights into the aetiology of snoring from observational and genetic investigations in the UK Biobank. *Nat Commun* 2020;11:817.

27. *Sleep Disorder Knowledge Portal* at <http://sleepdisordergenetics.org/home/portalHome>.

28. Turley P, Walters RK, Maghzian O, Okbay A, Lee JJ, Fontana MA, Nguyen-Viet TA, Wedow R, Zacher M, Furlotte NA, 23andMe Research Team, Social Science Genetic Association Consortium, Magnusson P, Oskarsson S, Johannesson M, Visscher PM, Laibson D, Cesarini D, Neale BM, Benjamin DJ. Multi-trait analysis of genome-wide association summary statistics using MTAG. *Nat Genet* 2018;50:229–237.

29. Cuellar-Partida G, Lundberg M, Kho PF, D'Urso S, Gutierrez-Mondragon LF, Hwang L-D. Complex-Traits Genetics Virtual Lab: A community-driven web platform for post-GWAS analyses. *Genetics* 2019;

30. Byrne EM, Kirk KM, Medland SE, McGrath JJ, Colodro-Conde L, Parker R, Cross S, Sullivan L, Statham DJ, Levinson DF, Licinio J, Wray NR, Hickie IB, Martin NG. Cohort

profile: the Australian genetics of depression study. *BMJ Open* 2020;10:e032580.

31. Lloyd-Jones LR, Zeng J, Sidorenko J, Yengo L, Moser G, Kemper KE, Wang H, Zheng Z, Magi R, Esko T, Metspalu A, Wray NR, Goddard ME, Yang J, Visscher PM. Improved polygenic prediction by Bayesian multiple regression on summary statistics. *Nat Commun* 2019;10:5086.
32. Yunru Huang, The 23andMe Research Team, Robert Gentleman. Genetic Association with Sleep Apnea in European Population. ASHG; .
33. Patel SR, Goodloe R, De G, Kowgier M, Weng J, Buxbaum SG, Cade B, Fulop T, Gharib SA, Gottlieb DJ, Hillman D, Larkin EK, Lauderdale DS, Li L, Mukherjee S, Palmer L, Zee P, Zhu X, Redline S. Association of genetic loci with sleep apnea in European Americans and African-Americans: the Candidate Gene Association Resource (CARe). *PLoS One* 2012;7:e48836.
34. Zhong A, Xiong X, Xu H, Shi M. An updated meta-analysis of the association between tumor necrosis factor-α -308G/A polymorphism and obstructive sleep apnea-hypopnea syndrome. *PLoS One* 2014;9:e106270.

TABLES

Table 1 Cases and controls across included cohorts			
Cohort	Total sample size	Apnoea cases	Snoring cases
UK-Biobank	408,317	7,902	152,303
FinnGen	66,216	9,096	4,270
Partners Biobank	20,047	3,102	4,175
CLSA	18,427	3,391	6,852
AGDS	10,359	1,517	4,450
Chen H. et al. (2018)*(19)	5,118	NA	NA
Total	510,484	25,062	172,050
*Apnoea-hypopnea index (quantitative) GWAS.			

Table 2. Sleep apnoea polygenic prediction comparison				
Model	**OR (95%C.I.)**	**P**	**Nagelkerke R2**	**Variance explained (%)***
SAprs	1.15 (1.08-1.21)	5.6e-06	0.0040	0.45
SAmtagSnoringprs	1.20 (1.13-1.27)	1.9e-09	0.0071	0.79
SAmtagSnoringBMIprs	1.22 (1.15-1.29)	1.2e-11	0.0090	0.98
Results from three logistic regressions predicting sleep apnoea adjusting for age, sex and 10 genetic principal components. *Variance explained on the logistic liability scale.				

Table 3. Independent hits associated with sleep apnoea and replicated in 23andMe

SNP	Nearest gene(s)	Chr	BP	A1/A2*	Beta*	SE*	P_23&me	P_SA	P_SAmtag Snoring	P_SAmtag SnoringBMI
rs9783497	MSRB3	12	65830349	G/A	-0.04	0.004	1.23E-25	5.09E-07	2.89E-16	3.62E-08
rs592333	DLEU7	13	51340315	G/A	0.04	0.004	9.04E-23	8.33E-06	6.80E-15	3.55E-08
rs1444789	GATA3	10	9064361	T/C	-0.037	0.005	2.40E-13	5.10E-03	3.66E-10	1.54E-05
rs2958153	PTGES3	12	57081517	G/A	-0.031	0.004	1.29E-12	5.93E-06	7.72E-10	1.35E-05
rs2277339	PRIM1	12	57146069	T/G	0.045	0.006	2.25E-12	5.73E-09	5.19E-14	2.21E-10
rs7107532	METTL15	11	28480924	G/A	-0.027	0.004	2.27E-12	1.23E-05	4.67E-09	1.23E-04
rs35445111	TSHZ3,ZNF507	19	32172047	G/A	-0.048	0.007	2.62E-12	1.43E-05	3.97E-11	2.49E-06
rs227731	NOG,C17orf67	17	54773238	T/G	-0.026	0.004	2.40E-11	1.84E-03	1.11E-08	2.98E-04
rs1436047	RP11-457D20.2	16	60618247	G/A	-0.025	0.004	1.19E-10	3.60E-06	9.09E-09	4.40E-06
rs11634019	ISL2	15	76634680	T/C	0.027	0.004	4.44E-10	1.42E-06	1.19E-09	4.86E-06
rs72904209	NR4A2	2	1.57E+08	T/C	-0.036	0.006	9.20E-10	3.84E-05	7.80E-14	1.31E-14
rs806292	DLEU1,DLEU7	13	50811031	G/A	-0.022	0.004	9.93E-09	9.19E-03	7.33E-09	NA
rs4987719	BCL2	18	60960310	T/C	0.061	0.011	1.28E-08	1.18E-03	4.47E-09	6.13E-07
rs7715167	TLX3,NPM1	5	1.71E+08	T/C	0.024	0.004	1.43E-08	4.50E-08	4.29E-08	4.86E-07
rs543874	FAM5B,SEC16B	1	1.78E+08	G/A	-0.027	0.005	6.40E-08	6.38E-01	4.91E-01	1.00E-15

rs6567160	PMAIP1,MC4R	18	57829135	T/C	0.024	0.005	1.31E-07	4.94E-03	6.51E-02	1.19E-31
rs2735309	HMGN1	21	40715313	T/C	-0.021	0.004	3.68E-07	2.50E-05	8.25E-07	9.10E-09
rs1136070	HN1L	16	1751935	T/C	0.037	0.007	3.88E-07	6.13E-05	4.06E-09	3.28E-08
rs6113592	PAX1,FOXA2	20	22229505	G/A	-0.02	0.004	6.42E-07	1.89E-07	5.51E-11	2.37E-08
rs11205802	MACF1	1	39699114	T/C	-0.02	0.004	1.22E-06	8.08E-06	7.30E-10	2.57E-09
rs12805133	SPTBN2	11	66483265	G/A	0.019	0.004	1.85E-06	1.84E-06	1.57E-08	1.44E-09
rs1554654	ABHD5,TOPAZ1	3	44044344	T/C	-0.018	0.004	3.46E-06	1.23E-03	1.22E-06	8.12E-10
rs12603115	SKAP1	17	46248994	T/C	-0.018	0.004	3.95E-06	6.12E-04	5.67E-10	1.23E-10
rs6265	BDNF	11	27679916	T/C	-0.022	0.005	1.12E-05	3.65E-09	3.05E-14	1.29E-36
rs11075985	FTO	16	53805207	C/A	-0.017	0.004	1.13E-05	3.73E-19	1.68E-23	NA
rs8176749	ABO	9	1.36E+08	T/C	-0.032	0.007	1.47E-05	1.10E-05	1.23E-09	1.30E-06
rs17794954	ZFP64,TSHZ2	20	50966307	T/C	-0.022	0.005	1.98E-05	8.37E-05	4.79E-06	3.11E-14
rs13021737	FAM150B,TMEM18	2	632348	G/A	-0.022	0.005	2.08E-05	1.98E-01	8.88E-01	7.88E-22
rs9526702	DLEU7,RNASEH2B	13	51475452	G/A	0.024	0.006	2.80E-05	8.92E-06	4.55E-08	NA
rs79932406	XKR9	8	71607667	T/G	0.016	0.004	4.05E-05	1.19E-01	2.21E-08	NA
rs7005777	PEX2	8	78233600	T/G	0.018	0.004	5.18E-05	2.52E-02	3.68E-08	1.36E-05
rs11176018	AC090673.2,LLPH	12	66451720	T/G	0.018	0.004	5.92E-05	1.43E-04	2.47E-08	8.81E-06
rs1403848	ROBO2	3	77609655	C/A	0.016	0.004	7.51E-05	3.35E-03	3.87E-09	2.27E-10
rs879620	ADCY9	16	4015729	T/C	-0.016	0.004	9.67E-05	8.24E-02	7.15E-02	1.04E-09
rs7138383	C12orf42	12	1.04E+08	G/A	0.017	0.004	1.08E-04	9.60E-03	1.43E-02	4.67E-08

rs11821161	GRM5,TYR	11	88888262	T/C	-0.015	0.004	2.16E-04	9.82E-04	1.15E-08	NA
rs6038517	FERMT1,BMP2	20	6458205	G/A	0.017	0.005	2.17E-04	3.85E-04	1.44E-08	3.86E-06
rs12683343	HSPA5,GAPVD1	9	1.28E+08	G/A	0.015	0.004	2.28E-04	1.57E-04	2.22E-07	1.62E-08
rs10075809	CEP120	5	1.23E+08	C/A	0.016	0.004	2.36E-04	4.76E-04	7.39E-09	2.91E-11

GWAS and replication significant hits. Results are shown for variants with genome-wide evidence of association ($p<5e-8$) in at least one of the main analyses, and evidence of replication ($p<2.6e-4$) in 23andMe. *Information based on the 23andMe result. For full results please refer to the publically available summary statistics.

FIGURES

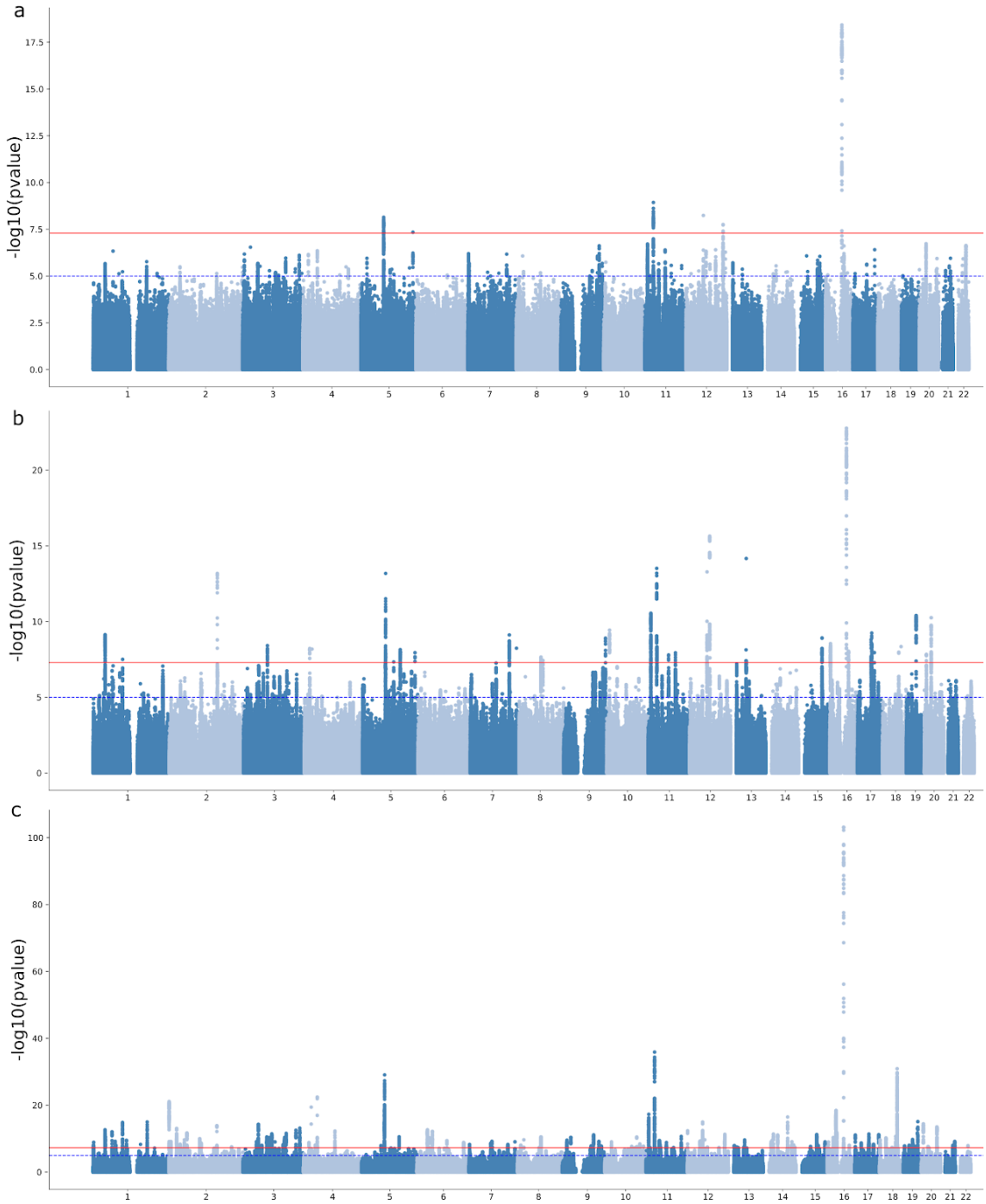

Figure 1. Sleep apnoea genetic associations

Manhattan plots depict the results of a) the meta-analysis for sleep apnoea, b) the MTAG for sleep apnoea increasing power using the snoring meta-analysis and c) MTAG for sleep apnoea increasing power using snoring and BMI. Each dot represents a genetic variant. The x-axis represents the variant genomic position (chromosome labels are shown). The y-axis depicts the significance of the association with sleep apnoea.

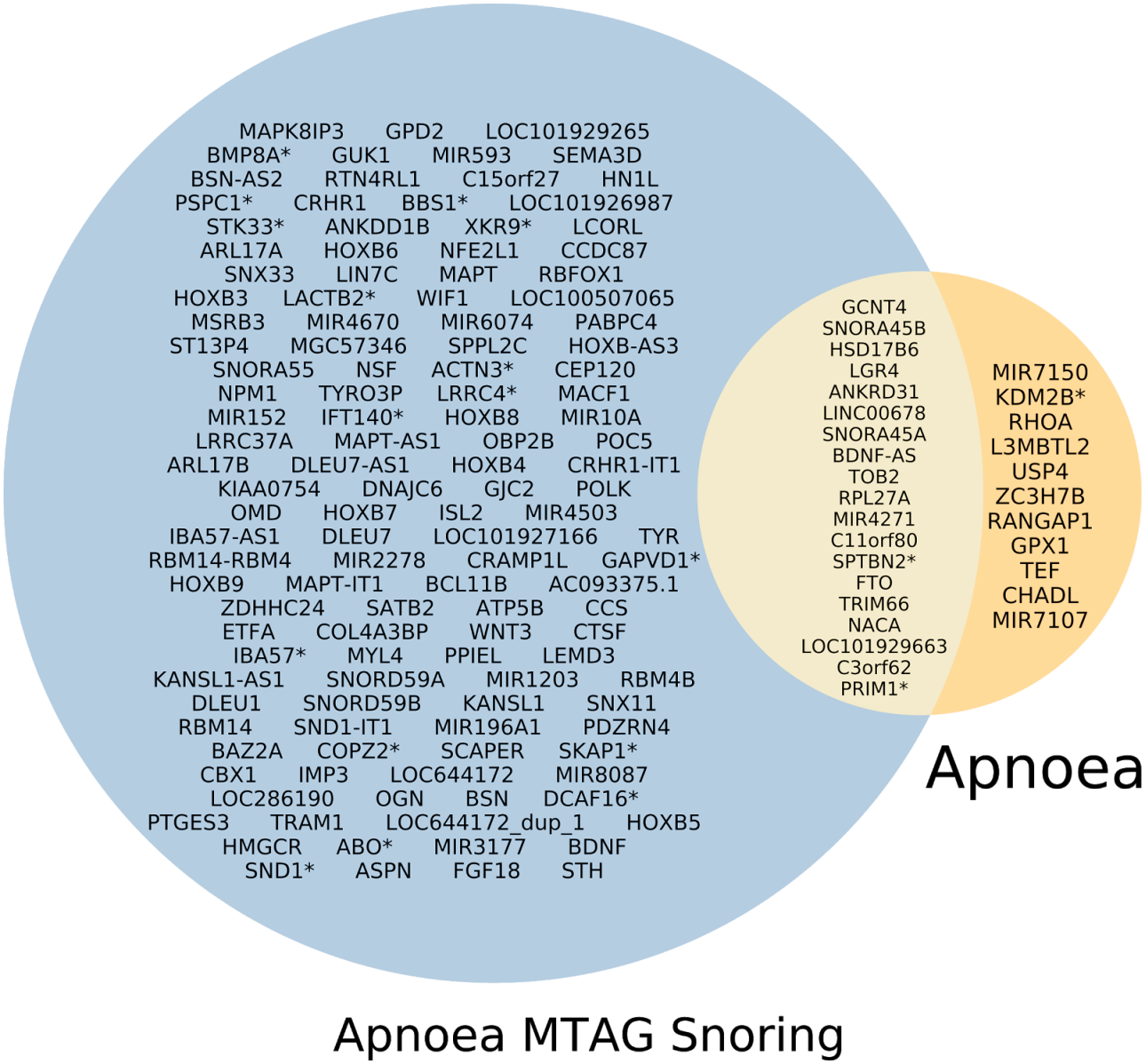

Figure 2. Gene based association and eQTL colocalisation results

Venn diagrams show the genes with a genome-wide significant evidence of association with either sleep apnoea (right) or the sleep apnoea MTAG with snoring (left). Genes with an asterisk (*) showed evidence of eQTL colocalisation in at least one of the tissues and GWAS studied (**see methods**). The results for the Apnoea MTAG boosting power with Snoring and BMI are not shown due to size constraints. Gene-based Manhattan plots are available in **Supplementary**

Figure 4. Full results for the gene-based tests and colocalisation analyses are available in Supplementary Data.

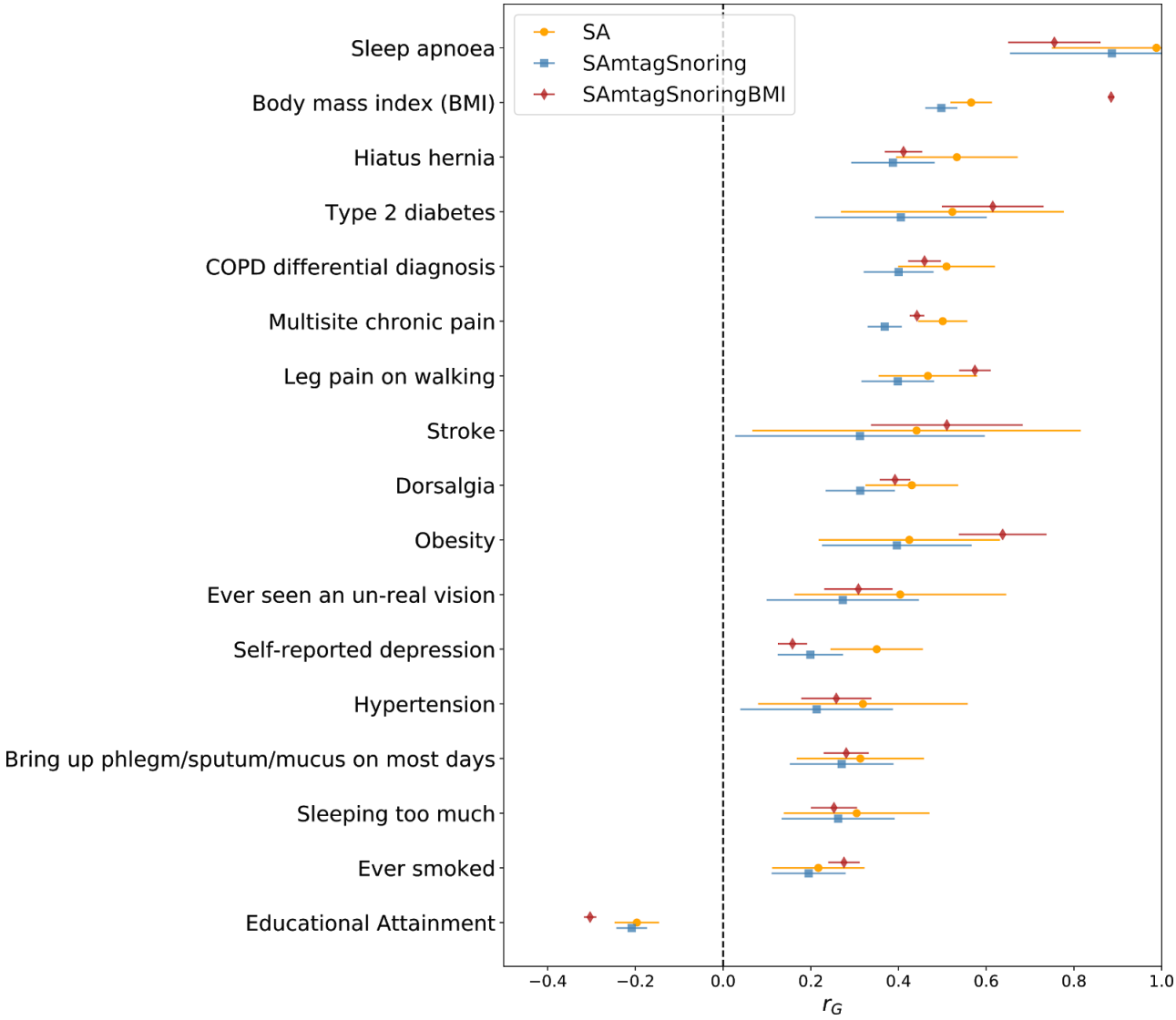

Figure 3. SA is genetically correlated with psychiatric, behavioural and cardiorespiratory traits

Forest plots showing the genetic correlations from CTG-VL(29) between sleep apnoea meta-analysis, MTAG between sleep apnoea meta-analysis and snoring meta-analysis (SAmtagSnoring) and MTAG between sleep apnoea meta-analysis, snoring meta-analysis and BMI (SAmtagSnoringBMI). Markers depict rg whereas lines represent 95% confidence intervals

derived from the rg standard error. Not all traits with a significant association are shown, see supplementary data.

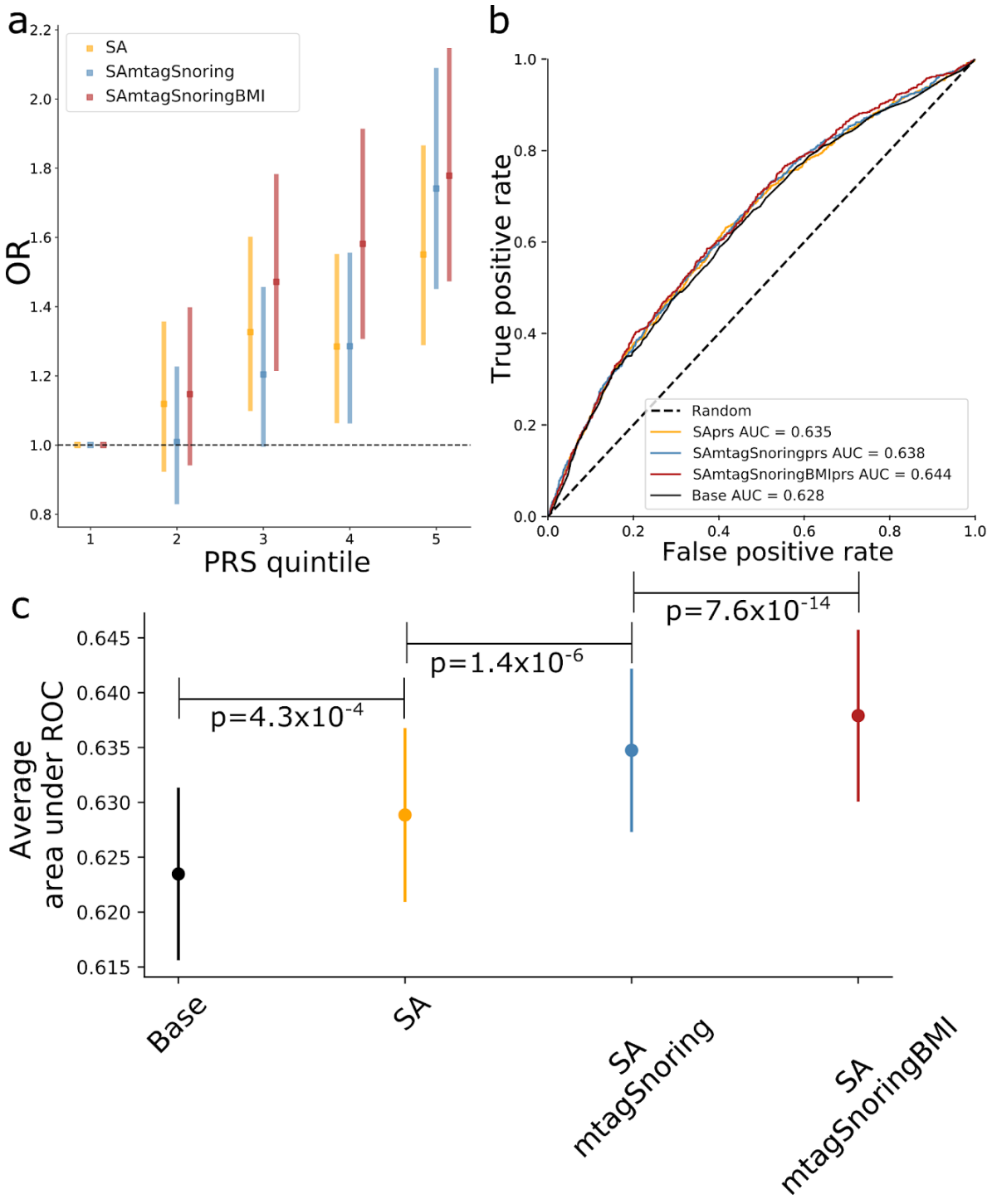

Figure 4. Sleep Apnoea polygenic prediction

a) Plot showing the odds ratio (OR) per change in polygenic risk score (PRS) quintile. Error bars depict the 95% confidence intervals b) Example of a receiver operating characteristic (ROC)

curve derived from assessing the ability of a logistic regression to predict sleep apnoea using either a base model (age+sex), or the base model plus the PRS of interest. The higher the area under the curve, the higher the predictive power of the model. c) Average area under ROC curve after 100 iterations of leave out validation randomly assigning training and testing subsamples. Error bars depict the standard deviation of the mean. Full results (100 ROC curves per model) are available in **Supplementary Figure 5**.

AUTHOR DISCLOSURE

Stella Aslibekyan and Yunru Huang are current or former employees of 23andMe, Inc. and may hold stock or stock options. All other authors declare no conflicts of interest.

ROLE OF FUNDING SOURCES

The funding sources had no input or involvement in the design, execution or reporting of the present study.

ETHICS STATEMENT

This study was approved by the QIMR Berghofer Human Research Ethics Committee. In addition, relevant details for each of the participating cohorts are provided below:

UK Biobank. The UK Biobank study was approved by the National Health Service National Research Ethics Service (ref. 11/NW/0382) and all participants provided written informed consent to participate in the UK Biobank study. Information about ethics oversight in the UK Biobank can be found at https://www.ukbiobank.ac.uk/ethics/.

CLSA. The CLSA abides by the requirements of the Canadian Institutes of Health Research (CIHR). The protocol of the CLSA has been reviewed and approved by 13 research ethics boards across Canada. A complete and detailed list is available at: https://www.clsa-elcv.ca/participants/privacy/who-ensures-high-ethical-standards/research-ethics-boards.

FinnGen. The Ethical Review Board of the Hospital District of Helsinki and Uusimaa approved the FinnGen study protocol (HUS/990/2017)

Partners Healthcare Biobank: Participants in the Partners Healthcare Biobank provided informed consent at sign up and ethics approval was obtained from the Partners Human Research Committee.

AGDS. all participants provided informed consent prior to participating in the study. This study and all questionnaires used for AGDS were approved by the QIMR Berghofer Human Research Ethics Committee.

23andMe Inc. Participants provided informed consent and participated in the research online, under a protocol approved by the external AAHRPP-accredited IRB, Ethical & Independent Review Services (E&I Review).

DATA AVAILABILITY

The full GWAS summary statistics for this study will be made available through the NHGRI-EBI GWAS Catalogue (https://www.ebi.ac.uk/gwas/downloads/summary-statistics). Data are available from the Canadian Longitudinal Study on Aging (www.clsa-elcv.ca) for researchers who meet the criteria for access to de-identified CLSA data. UK Biobank and FinnGen data are also accessible through their respective application procedures.